Numerical Algorithms (2023) 92:1655–1677
https://doi.org/10.1007/s11075-022-01359-y

ORIGINAL PAPER

Unconditionally optimal H^1-error estimate of a fast nonuniform L2-1_σ scheme for nonlinear subdiffusion equations

Nan Liu[1] · **Yanping Chen**[2] · **Jiwei Zhang**[3] · **Yanmin Zhao**[4]

Received: 2 January 2022 / Accepted: 13 June 2022 / Published online: 14 July 2022

Abstract

This paper is concerned with the unconditionally optimal H^1-error estimate of a fast second-order scheme for solving nonlinear subdiffusion equations on the nonuniform mesh. We use the Galerkin finite element method (FEM) to discretize the spacial direction, the Newton linearization method to approximate the nonlinear term and the sum-of-exponentials (SOE) approximation to speed up the evaluation of Caputo derivative. Our analysis of the unconditionally optimal H^1-error estimate involves the temporal-spatial error splitting approach, an improved discrete fractional Grönwall inequality and error convolution structure. In order to find a suitable test function to estimate H^1-error, we here consider two cases: linear and high-order FEM space, using the time-discrete operator and Laplace operator in the test function respectively. Numerical tests are provided demonstrate the effectiveness and the unconditionally optimal H^1-error convergence of our scheme.

✉ Jiwei Zhang
jiweizhang@whu.edu.cn

✉ Yanmin Zhao
zhaoym@lsec.cc.ac.cn

Nan Liu
liunan@whu.edu.cn

Yanping Chen
yanpingchen@scnu.edu.cn

[1] School of Mathematics and Statistics, Wuhan University, Wuhan, 430072, China

[2] School of Mathematical Sciences, South China Normal University, Guangzhou, 510631, China

[3] School of Mathematics and Statistics, and Hubei Key Laboratory of Computational Science, Wuhan University, Wuhan, 430072, China

[4] School of Science, Xuchang University, Xuchang, 461000, China

Keywords H^1-error estimate · Unconditional convergence · The fast scheme · The nonuniform L2-1_σ approximation

Mathematics Subject Classification 2010 11T23 · 26A33 · 35R11 · 65N30 · 65M12 · 35K61

1 Introduction

The subdiffusion equations are widely used to describe various phenomena of anomalous diffusion in control systems, physics, biology [1–3], and attract lots of researchers in theoretical and numerical analysis. This paper focuses on the unconditionally optimal H^1-error estimate of a fully discrete scheme for the following nonlinear subdiffusion problem on a bounded convex domain $\Omega \subset \mathbb{R}^d (d = 1, 2, 3)$:

$$\begin{aligned} {}^{\mathrm{C}}\mathcal{D}_t^\alpha u(\mathbf{x}, t) &= \Delta u(\mathbf{x}, t) + f(u), && \mathbf{x} \in \Omega, t \in (0, T], && (1.1)\\ u(\mathbf{x}, 0) &= u_0(\mathbf{x}), && \mathbf{x} \in \bar{\Omega}, && (1.2)\\ u(\mathbf{x}, t) &= 0, && \mathbf{x} \in \partial\Omega, t \in [0, T], && (1.3) \end{aligned}$$

where the Caputo derivative ${}^{\mathrm{C}}\mathcal{D}_t^\alpha$ $(0 < \alpha < 1)$ acting on u is defined as

$${}^{\mathrm{C}}\mathcal{D}_t^\alpha u(\mathbf{x}, t) = \int_0^t \partial_s u(\mathbf{x}, s)\omega_{1-\alpha}(t - s)\,\mathrm{d}s \quad \text{with} \quad \omega_\beta(t) = t^{\beta-1}/\Gamma(\beta). \tag{1.4}$$

There are three features for problem (1.1)–(1.3):

- the solution u has the initial time singularity;
- the Caputo derivative is nonlocal;
- the problem is nonlinear.

The first feature of problem (1.1)–(1.3) is ubiquitous in nature that the solution u is weakly singular near the initial time $t = 0$. Generally, the regularity of the solution has the following property [4–10]:

$$\|\partial_t^m u(t)\|_{H^2(\Omega)} \le Ct^{\alpha-m}, \qquad \text{for } m = 1, 2. \tag{1.5}$$

We point out that here C generally means a constant, which is independent of mesh sizes h and τ, but it may depend on the given data (such as f, u_0, α, Ω, T). Thus, the initial regularity will become an important consideration for any numerical method to solve the subdiffusion problems. To achieve the optimal convergence rate, the nonuniform/adaptive time step is required, which also brings more complicated and difficult theoretical analysis of numerical schemes comparing with the uniform mesh. A typical nonuniform mesh is the following general graded mesh

$$\tau_k \le C_\gamma \tau t_k^{1-1/\gamma}\ (1 \le k \le N), \quad t_k \le C_\gamma t_{k-1} \ \text{and}\ \tau_k/t_k \le C_\gamma \tau_{k-1}/t_{k-1}\ (2 \le k \le N), \tag{1.6}$$

where N represents the total number of time steps, $\gamma \ge 1$ is a parameter and $\tau = \max_{1\le k\le N} \tau_k$ is the maximum step size. The general graded mesh has been successfully applied to various time fractional PDEs, see [8–18]. For instance, Liao et al.

[8] obtain the optimal $\mathcal{O}\big(\tau^{\min\{\gamma\alpha,2-\alpha\}}\big)$ convergence rate of L1 scheme on the nonuniform mesh. In which, a theoretical framework is proposed by presenting a discrete complementary convolution (DCC) kernel, a discrete fractional Grönwall inequality and an error convolution structure (ECS). Based on this framework, the optimal convergence order of several widely used numerical schemes on the nonuniform mesh is obtained successively, such as the optimal $\mathcal{O}\big(\tau^{\min\{\gamma\alpha,2\}}\big)$ convergence rate of L2-1_σ scheme in L^2-norm [9] and the optimal $\mathcal{O}\big(\tau^{\min\{\gamma\alpha,2-\alpha\}}\big)$ convergence rate of the two-level fast L1 scheme in L^2-norm [13].

The second nonlocal feature will lead to the huge computational storage and cost for the long-time or small-mesh simulations, which is especially prohibitive to compute the high-dimensional problem. To circumvent this difficulty of computational complexity, there are generally two alternatives: one is to introduce the fast algorithms to significantly reduce the computational storage and cost; another is to use the high-order schemes to obtain the same accuracy with less time steps. For the fast algorithms, one can refer to [19–26]. For instance, Jiang et al. [20] present the sum-of-exponentials (SOE) approximation to speed up the efficient evaluation of Caputo derivative, which significantly reduces the computational storage and cost. Late, Yan et al. [21] apply the idea of SOE approximate to the second-order L2-1_σ scheme. Baffet et al. [22] combine the kernel compression scheme with a time stepping method. Zhu et al. [23] present a fast L2 scheme with (3-α)-order with the application of SOE approximation. Guo et al. [24] apply the fractional linear multistep method to deal with the tempered fractional integral and derivative operators. For the high-order schemes, one can refer to [27–30] for the widely used L2 scheme [28] and L2-1_σ scheme [29]. In this paper, we will use the L2-1_σ scheme with the corresponding fast algorithm presented in [20, 21] to speed up the computation of the Caputo derivative on general graded mesh (1.6).

The third feature involves the nonlinearity of the problem itself. The typical methods to numerically deal with the nonlinearity involves pure explicit scheme, fully implicit scheme, and implicit-explicit (or semi-implicit) scheme and so on. The pure explicit scheme is the most easy implementation, but suffers from a CFL condition for the stability. The fully implicit scheme is generally unconditionally stable, but needs extra computational cost for iteratively solving a nonlinear algebraic system. A popular alternative is semi-implicit scheme which discretizes the linear term by implicit scheme and the nonlinear term by a linearized or an explicit scheme. The resulting semi-implicit scheme circumvents the iteration for the fully implicit scheme, but also brings the difficulty of theoretical analysis of the unconditional convergence. The so-called unconditional convergence here means the optimal convergence does not suffer from any restrictions of ratios between temporal and spacial mesh sizes. In this paper, we use the implicit scheme to discretize the linear dispersive term and use the Newton linearization method to approximate the nonlinear term for the time direction, and employ the Galerkin FEM for the spacial direction.

The focus of this paper is on the unconditionally optimal H^1-error estimate for the proposed second order fast scheme to numerically solve problem (1.1) on the nonuniform mesh. To do so, we first carefully use the SOE-based fast L2-1_σ approximation of Caputo derivative [20, 21], which significantly reduce the computational complexity when the total number of the time step is large enough. After that, we use the spatial-temporal splitting approach introduced in [31, 32] to prove our linearization scheme is unconditionally convergent. The idea of the spatial-temporal splitting approach is beginning with proving the boundedness of the solution to the temporal semi-discrete scheme in the infinity norm, and then prove the optimal error estimate of the fully discrete scheme, which successfully evade the ratio between time and space mesh sizes. In the proof process, we consider u is smooth enough in spatial directions. Combined with the original (1.1), it further implies that Δu is zero on $\partial\Omega$. Finally, we use the theoretical framework developed in [8–10] to present the error estimate for subdiffusion equations, which involves the discrete time fractional Grönwall inequality, DCC kernel and ECS. This framework can effectively deal with the nonuniform temporal scheme when the initial regularity is considered. Specially for H^1-error estimate, we divide the FEM space into two cases of linear and high-order. Namely, when $r = 1$ (here r represents the degree of continuous piecewise polynomials), the time-discrete operator is taken in the test function like [33, 34]; when $r \geq 2$, the Laplace operator is taken in the test function.

The paper is organized as follows. In Section 2, we introduce the fast L2-1_σ scheme and fully discrete scheme. In Section 3, we first give some necessary lemmas, then split the error into spatial and temporal components for detailed analysis respectively, and present the unconditionally optimal H^1-error estimate. In Section 4, some numerical results are provided to verify our theoretical analysis.

2 Fast L2-1_σ and fully discrete schemes

We take the general nonuniform by $0 = t_0 < t_1 < t_2 < \ldots < t_N = T$ with time steps $\tau_k = t_k - t_{k-1}$. Set $t_{k-\sigma} = (1-\sigma)t_k + \sigma t_{k-1}$, $\sigma \in [0, 1]$ and the step ratios $\rho_k = \tau_k/\tau_{k+1}$. There is a constant $\rho > 0$ such that the step ratios $\rho_k < \rho$ for $1 \leq k \leq N-1$. Set $u^k = u(\cdot, t_k)$, $u^{k-\sigma} = (1-\sigma)u^k + \sigma u^{k-1}$ and the difference operator $\nabla_\tau u^k = u^k - u^{k-1}$ for $1 \leq k \leq N$. Taking $\sigma = \frac{\alpha}{2}$ here and after, the L2-1_σ scheme is defined by

$$\begin{aligned} {}^{\mathrm{C}}\mathcal{D}_\tau^\alpha u^{n-\sigma} &= \sum_{k=1}^{n-1}\int_{t_{k-1}}^{t_k} \partial_s\big(\Pi_{2,k}u(x,s)\big)\omega_{1-\alpha}(t_{n-\sigma}-s)\,\mathrm{d}s + \int_{t_{n-1}}^{t_{n-\sigma}} \partial_s\big(\Pi_{1,n}u(x,s)\big)\omega_{1-\alpha}(t_{n-\sigma}-s)\,\mathrm{d}s \\ &:= \sum_{k=1}^{n} A_{n-k}^{(n)}\nabla_\tau u^k, \end{aligned} \tag{2.7}$$

where $\Pi_{1,k}u(x,s)$ and $\Pi_{2,k}u(x,s)$ represent the linear interpolation with the nodes t_{k-1}, t_k and the quadratic interpolation at t_k, t_{k-1}, t_{k+1} for the variable s, and the discrete convolution kernel $A_{n-k}^{(n)}$ can be calculated by

$$A_{n-k}^{(n)} = \begin{cases} a_0^{(n)} + \rho_{n-1}b_1^{(n)}, & k = n, \\ a_{n-k}^{(n)} + \rho_{k-1}b_{n-k+1}^{(n)} - b_{n-k}^{(n)}, & 2 \leq k \leq n-1, \\ a_{n-1}^{(n)} - b_{n-1}^{(n)}, & k = 1, \end{cases} \tag{2.8}$$

with

$$a_0^{(n)} = \frac{1}{\tau_n}\int_{t_{n-1}}^{t_{n-\sigma}} \omega_{1-\alpha}(t_{n-\sigma}-s)\,\mathrm{d}s, \quad a_{n-k}^{(n)} = \frac{1}{\tau_k}\int_{t_{k-1}}^{t_k} \omega_{1-\alpha}(t_{n-\sigma}-s)\,\mathrm{d}s,$$
$$b_{n-k}^{(n)} = \frac{2}{\tau_k(\tau_k+\tau_{k+1})}\int_{t_{k-1}}^{t_k} (s-t_{k-\frac{1}{2}})\omega_{1-\alpha}(t_{n-\sigma}-s)\,\mathrm{d}s.$$

It is known that the direct algorithm of the L2-1_σ scheme (2.7) has the computational complexity with the storage $\mathcal{O}(N)$ and cost $\mathcal{O}(N^2)$, respectively. This computational complexity will be huge and inadmissible for small time size or long time simulations. It motivates us to consider a fast algorithm of L2-1_σ scheme based on the SOE technique developed in [20, 21] to approximate the kernel $t^{-\alpha}$. The resulting fast L2-1_σ scheme only has the complexity of the storage $\mathcal{O}(\log^2 N)$ and cost $\mathcal{O}(N\log^2 N)$, which significantly reduces the computational complexity for large N. The main methodology of the fast algorithm in [20, 21] is presented as follows.

We first split ${}^{\mathrm{C}}\mathcal{D}_t^{\alpha}u^{n-\sigma}$ into a local part I and a history part II, say

$${}^{\mathrm{C}}\mathcal{D}_t^{\alpha}u^{n-\sigma} = \int_{t_{n-1}}^{t_{n-\sigma}} \partial_s u(x,s)\omega_{1-\alpha}(t_{n-\sigma}-s)\,\mathrm{d}s + \int_0^{t_{n-1}} \partial_s u(x,s)\omega_{1-\alpha}(t_{n-\sigma}-s)\,\mathrm{d}s = I + II. \quad (2.9)$$

The local part I is directly approximated by

$$I \approx \int_{t_{n-1}}^{t_{n-\sigma}} \partial_s\big(\Pi_{1,n}u(x,s)\big)\omega_{1-\alpha}(t_{n-\sigma}-s)\,\mathrm{d}s = \frac{\nabla_\tau u^n}{\tau_n}\int_{t_{n-1}}^{t_{n-\sigma}} \omega_{1-\alpha}(t_{n-\sigma}-s)\,\mathrm{d}s := a_0^{(n)}\nabla_\tau u^n. \quad (2.10)$$

To speed up the evaluation of II, we use the following SOEs approximation for the kernel $t^{-\alpha}$.

Lemma 2.1 ([20, 21]) *For the given parameters α, ϵ, δ and T, one can find a family of points s_i and weights ω_i ; $(i = 1, 2, \ldots, \mathcal{P})$ such that*

$$\left| t^{-\alpha} - \sum_{i=1}^{\mathcal{P}} \omega_i e^{-s_i t} \right| \le \epsilon, \quad \forall t \in [\delta, T], \quad (2.11)$$

where the total number $\mathcal{P}$ of exponentials needed is of order

$$\mathcal{P} = \mathcal{O}\Big(\big(\log(T/\delta) + \log(\log\epsilon^{-1})\big)\log\epsilon^{-1} + \big(\log\delta^{-1} + \log(\log\epsilon^{-1})\big)\log\delta^{-1}\Big). \quad (2.12)$$

Remark 1 In the practical simulations, we generally fix the tolerance error ϵ as the machine precision. Once fixing ϵ, taking $\delta = \min_{1\le k\le N}\tau_k$ and noting $N = \mathcal{O}(T/\tau)$ in (2.12), we have $\mathcal{P} = \mathcal{O}(\log N)$ for $T \gg 1$ and $\mathcal{P} = \mathcal{O}(\log^2 N)$ for $T = \mathcal{O}(1)$.

By using the SOEs approximation of $t^{-\alpha}$ in (2.11), the history part II can be written as

$$II \approx \frac{1}{\Gamma(1-\alpha)}\sum_{i=1}^{\mathcal{P}}\int_0^{t_{n-1}} \partial_s\big(\Pi_{2,n-1}u(x,s)\big)\omega_i e^{-s_i(t_{n-\sigma}-s)}\,\mathrm{d}s := \sum_{i=1}^{\mathcal{P}} H_i(t_{n-1}) \quad (2.13)$$

with the history integral $H_i(t_0) = 0$ and the following recurrence formula

$$H_i(t_{n-1}) = e^{-s_i \tau_{n-\sigma}} H_i(t_{n-2}) + \frac{1}{\Gamma(1-\alpha)} \int_{t_{n-2}}^{t_{n-1}} \partial_s \big(\Pi_{2,n-1} u(x,s)\big) \omega_i e^{-s_i(t_{n-\sigma}-s)} \, \mathrm{d}s. \tag{2.14}$$

Combining (2.9), (2.10) and (2.13), we have the fast L2-1_σ scheme given as

$$^{\mathrm{F}}\mathcal{D}_\tau^\alpha u^{n-\sigma} = a_0^{(n)} \nabla_\tau u^n + \sum_{i=1}^{\mathcal{P}} H_i(t_{n-1}), \tag{2.15}$$

where $H_i(t_{n-1})$ can be calculated by the recurrence formula (2.14). For further theoretical analysis, we can equivalently reformulate (2.15) into the following convolution form

$$^{\mathrm{F}}\mathcal{D}_\tau^\alpha u^{n-\sigma} = \sum_{k=1}^{n} B_{n-k}^{(n)} \nabla_\tau u^k, \tag{2.16}$$

where

$$B_{n-k}^{(n)} = \begin{cases} a_0^{(n)} + \sum\limits_{i=1}^{\mathcal{P}} \rho_{n-1} \tilde{b}_1^{(n)}, & k = n, \\ \sum\limits_{i=1}^{\mathcal{P}} e^{-s_i(t_{n-\sigma}-t_{k+1-\sigma})} (\tilde{a}_{n-k}^{(k+1)} + e^{-s_i \tau_{k+1-\sigma}} \rho_{k-1} \tilde{b}_{n-k+1}^{(k)} - \tilde{b}_{n-k}^{(k+1)}), & 2 \le k \le n-1, \\ \sum\limits_{i=1}^{\mathcal{P}} e^{-s_i(t_{n-\sigma}-t_{2-\sigma})} (\tilde{a}_{n-1}^{(2)} - \tilde{b}_{n-1}^{(2)}), & k = 1, \end{cases} \tag{2.17}$$

where

$$\begin{aligned} \tilde{a}_{n-k}^{(k+1)} &= \frac{\omega_i}{\tau_k \Gamma(1-\alpha)} \int_{t_{k-1}}^{t_k} e^{-s_i(t_{k+1-\sigma}-s)} \, \mathrm{d}s, \\ \tilde{b}_{n-k}^{(k+1)} &= \frac{2\omega_i}{\tau_k(\tau_k + \tau_{k+1})\Gamma(1-\alpha)} \int_{t_{k-1}}^{t_k} (s - t_{k-\frac{1}{2}}) e^{-s_i(t_{k+1-\sigma}-s)} \, \mathrm{d}s. \end{aligned}$$

The discrete convolution kernel $B_{n-k}^{(n)}$ has the following properties [35]:

M1. $B_{n-k-1}^{(n)} - B_{n-k}^{(n)} > 0$, for $1 \le k \le n-1$,

M2. $B_{n-k}^{(n)} \ge \frac{1}{\pi_A} \int_{t_{k-1}}^{t_k} \frac{\omega_{1-\alpha}(t_n - s)}{\tau_k} \mathrm{d}s, \pi_A = 2,$ for $1 \le k \le n$,

M3. $B_0^{(n)} \le \frac{26}{11} \int_{t_{n-1}}^{t_n} \frac{\omega_{1-\alpha}(t_n - s)}{\tau_n} \mathrm{d}s \le \frac{2\tau_n^{-\alpha}}{\Gamma(2-\alpha)}$,

where $\epsilon \le \epsilon_1 = \min\{\frac{\alpha}{2(1-\alpha)} \omega_{1-\alpha}(T), \frac{1}{26}\omega_{1-\alpha}(T)\}$. We point out that we only consider α is a given constant throughout the paper, and does not consider the case of $\alpha \to 0$.

For the discretization in space, the continuous Galerkin FEM is used. For brevity, we denote $\|\cdot\|_\infty$ as $\|\cdot\|_{W^{0,\infty}}$ and $\|\cdot\|_m$ as $\|\cdot\|_{W^{m,2}}$, where $\|\cdot\|_{W^{m,p}}$ represent the norms for the Sobolev space $W^{m,p}(\Omega)$ [36]. Let $V_h \subset H_1^0(\Omega)$ be the space of piecewise polynomials of degree $\le r$ corresponding to a conforming (quasi-uniform) triangulation of Ω with maximum element size h. We can get the following fully discrete scheme based on the Newton linearization method for $n = 1, 2, \ldots, N$, namely

$$\left({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha U_h^{n-\sigma}, v\right) = -(\nabla U_h^{n-\sigma}, \nabla v) + \left(f(U_h^{n-1}) + (1-\sigma) f'(U_h^{n-1}) \nabla_\tau U_h^n, v\right), \quad v \in V_h. \tag{2.18}$$

We first report the unconditionally optimal error estimate of scheme (2.18) as follows.

Theorem 2.1 *Assume the problem (1.1)–(1.3) has a unique solution u^n which satisfies (1.5) and is smooth enough in spatial directions. Then the fully discrete scheme defined in (2.18) has a unique solution U_h^n. If $f \in C^4(\mathbb{R})$, there exist τ^*, h^*, ϵ^*, such that when $\tau < \tau^*$, $h < h^*$, $\epsilon < \epsilon^*$ and $\gamma\alpha \le 2\,(\gamma \ge 1)$, satisfying*

$$\|u^n - U_h^n\|_1 \le C^*(\tau^{\min\{2,\gamma\alpha\}} + h^r + \epsilon), \tag{2.19}$$

where C^* is a positive constant independent of h, τ and ϵ.

The proof of Theorem 2.1 is presented in the next section.

3 Unconditionally optimal H^1-error estimate

We here consider the truncation errors caused by the Taylor expansion at $t_{n-\sigma}$ and the Newton linearization method, and introduce a discrete fractional Grönwall inequality. After that, we use the temporal-spatial error splitting approach developed in [31] to obtain the unconditionally optimal H^1-error estimate.

3.1 Preliminaries

It is known that the continued kernel ω_α holds the semigroup property $\omega_\alpha * \omega_\beta = \omega_{\alpha+\beta}$, namely

$$\int_s^t \omega_\alpha(t-\mu)\omega_\beta(\mu - s)\,\mathrm{d}\mu = \omega_{\alpha+\beta}(t-s), \quad \forall\ 0 < s < t < \infty, \tag{3.20}$$

thus we have $\omega_\alpha * \omega_{1-\alpha} = \omega_1 = 1$. For the discrete kernel $B_{n-j}^{(n)}$ in (2.17), it generally does not hold the same semigroup property (3.20) as the continued kernel. To preserve the same property, a complementary discrete kernel $P_{n-k}^{(n)}$ proposed in [9] is introduced such that

$$\sum_{j=k}^{n} P_{n-j}^{(n)} B_{j-k}^{(j)} = 1, \quad 1 \le k \le n. \tag{3.21}$$

For the given value $B_{j-k}^{(j)}$, we can calculate $P_{n-j}^{(n)}$ from (3.21) by using the following recursive formula:

$$P_{n-j}^{(n)} = \frac{1}{B_0^{(j)}} \begin{cases} 1, & j = n, \\ \sum_{k=j+1}^{n} (B_{k-j-1}^{(k)} - B_{k-j}^{(k)}) P_{n-k}^{(n)}, & 1 \le j \le n-1. \end{cases} \tag{3.22}$$

Next, we present several useful Lemmas.

Lemma 3.1 ([10]) *Let $B_{n-k}^{(n)}$ has properties **M1** and **M2**. For any sequence $(v^n)_{n=1}^{N}$, it holds*

$$\frac{1}{2} \sum_{k=1}^{n} B_{n-k}^{(n)} \nabla_\tau ||v^k||_0^2 \le \left({}^F\mathcal{D}_\tau^\alpha v^{n-\sigma}, v^{n-\sigma} \right), \quad for \quad 1 \le n \le N. \tag{3.23}$$

Lemma 3.2 ([34] An improved discrete fractional Grönwall inequality) *Assume $B_{n-k}^{(n)}$ holds the properties of **M1** and **M2**, and $(\xi^n)_{n=1}^N$, $(\eta^n)_{n=1}^N$ and $(\lambda_l)_{l=0}^{N-1}$ are three nonnegative sequences. If the nonnegative sequence $(v^k)_{k=0}^N$ satisfies*

$$\sum_{k=1}^{n} B_{n-k}^{(n)} \nabla_\tau (v^k)^2 \le \sum_{k=1}^{n} \lambda_{n-k} (v^{k-\sigma})^2 + v^{n-\sigma} \xi^n + (\eta^n)^2, \qquad 1 \le n \le N \tag{3.24}$$

and the maximum step size $\tau \le 1/\sqrt[\alpha]{2\pi_A \Lambda \Gamma(2-\alpha)}$, then there exists a constant $\Lambda \ge \sum_{l=0}^{N-1} \lambda_l$ such that

$$\begin{aligned} v^n &\le 2E_\alpha(2\max(1,\rho)\pi_A \Lambda t_n^\alpha)\Big(v^0 + \max_{1\le k\le n} \sum_{j=1}^{k} P_{k-j}^{(k)} \xi^j + \sqrt{\pi_A \Gamma(1-\alpha)} \max_{1\le k\le n} \{t_k^{\alpha/2} \eta^k\}\Big) \\ &\le 2E_\alpha(2\max(1,\rho)\pi_A \Lambda t_n^\alpha)\Big(v^0 + \pi_A \Gamma(1-\alpha) \max_{1\le j\le n} \{t_j^\alpha \xi^j\} + \sqrt{\pi_A \Gamma(1-\alpha)} \max_{1\le k\le n} \{t_k^{\alpha/2} \eta^k\}\Big). \end{aligned} \tag{3.25}$$

Lemma 3.3 ([35]) *Assume that $v \in C^3((0,T])$ and satisfies (1.5). Denote*

$$R_t^{k-\sigma} = {}^C\mathcal{D}_t^\alpha v(t_{k-\sigma}) - {}^F\mathcal{D}_\tau^\alpha v^{k-\sigma} \tag{3.26}$$

as the local consistency error of fast L2-1_σ scheme. Then the global consistency error can be bounded by

$$\sum_{k=1}^{n} P_{n-k}^{(n)} |R_t^{k-\sigma}| \le C_v \Big(\frac{\tau_1^\alpha}{\alpha} + \frac{1}{1-\alpha} \max_{2\le k\le n} t_k^\alpha t_{k-1}^{\alpha-3} \tau_k^3 \tau_{k-1}^{-\alpha} + \frac{\epsilon}{\alpha} t_n^\alpha \hat{t}_{n-1}^2 \Big), \tag{3.27}$$

where $\hat{t}_n = \max\{1, t_n\}$.

Lemma 3.4 *Assume that $v \in C^2((0,T])$ and satisfies (1.5), and the nonlinear function $f(v) = f \in C^4(\mathbb{R})$. Denote the local truncation error by*

$$R_f^{k-\sigma} = f\big(v(t_{k-\sigma})\big) - f(v^{k-1}) - (1-\sigma) f'(v^{k-1}) \nabla_\tau v^k. \tag{3.28}$$

Then the global consistency error can bounded by

$$\sum_{k=1}^{n} P_{n-k}^{(n)}\left|R_f^{k-\sigma}\right| + \sum_{k=1}^{n} P_{n-k}^{(n)}\left|\nabla R_f^{k-\sigma}\right| + \sum_{k=1}^{n} P_{n-k}^{(n)}\left|\Delta R_f^{k-\sigma}\right| \le C_v\big(\tau_1^{3\alpha} + t_n^{\alpha}\max_{2\le k\le n}\tau_k^2 t_{k-1}^{2(\alpha-1)}\big). \tag{3.29}$$

The proof is presented in the Appendix for brevity.

Lemma 3.5 ([37]) *Assume* $v \in C^2((0,T])$ *and satisfies* (1.5)*. Denote by*

$$R_\sigma^{k-\sigma} = \Delta v(t_{k-\sigma}) - \Delta v^{k-\sigma}.$$

Then it holds

$$\sum_{k=1}^{n} P_{n-k}^{(n)}\left|R_\sigma^{k-\sigma}\right| \le C_v\Big(\frac{\tau_1^{2\alpha}}{\alpha} + t_n^{\alpha}\max_{2\le k\le n} t_{k-1}^{\alpha-2}\tau_k^2\Big). \tag{3.30}$$

3.2 Analysis of the semi-discrete scheme

We now consider the following semi-discrete problem at time $t_{n-\sigma}$, namely

$$^{\mathrm{F}}\mathcal{D}_\tau^\alpha U^{n-\sigma} = \Delta U^{n-\sigma} + f(U^{n-1}) + (1-\sigma)f'(U^{n-1})\nabla_\tau U^n, \quad n = 1,\dots,N \tag{3.31}$$

with the initial and boundary conditions

$$U^0(x) = u_0(x), \qquad x \in \bar{\Omega}, \tag{3.32}$$

$$U^n(x) = 0, \qquad x \in \partial\Omega, \quad n = 1,\dots,N. \tag{3.33}$$

Set $e^n = u^n - U^n$, $n = 0, 1, \dots, N$. Subtracting (3.31) from (1.1) produces

$$^{\mathrm{F}}\mathcal{D}_\tau^\alpha e^{n-\sigma} = \Delta e^{n-\sigma} + E_1^{n-\sigma} + R_\sigma^{n-\sigma} + R_t^{n-\sigma} + R_f^{n-\sigma}, \tag{3.34}$$

where

$$E_1^{n-\sigma} = f(u^{n-1}) + (1-\sigma)f'(u^{n-1})\nabla_\tau u^n - f(U^{n-1}) - (1-\sigma)f'(U^{n-1})\nabla_\tau U^n. \tag{3.35}$$

Next, we consider the boundedness of U^n and the error estimate of e^n.

Theorem 3.1 *The semi-discrete problem* (3.31)–(3.33) *has a unique solution* U^n*. Moreover, if* $f \in C^4(\mathbb{R})$*, there exist* $\epsilon^* > 0$ *and* $\tau^{**} > 0$ *such that when* $\epsilon \le \epsilon^*$ *and* $\tau \le \tau^{**}$*, it holds*

$$\|e^n\|_2 \le C_1\tau^{\min\{\gamma\alpha,2\}} + C_2\epsilon, \tag{3.36}$$

$$\|U^n\|_\infty \le Q_1, \tag{3.37}$$

$$\|^{F}\mathcal{D}_\tau^\alpha U^{n-\sigma}\|_2 \le C_3, \tag{3.38}$$

where $\gamma\alpha \le 2$*,* $Q_1 = \max_{1\le n\le N}\|u^n\|_\infty + 1$*,* C_1, C_2 *and* C_3 *are constants independent of* τ *and* ϵ*.*

Proof At each time level, (3.31) is a linear elliptic problem. So it is easy to obtain the existence and uniqueness of the solution U^n. We now use the induction to prove

(3.36) and (3.37). For $n = 0$, it is obvious that (3.36) and (3.37) hold. Then, we assume that (3.36) holds for $0 \le k \le n-1$, and have

$$\begin{aligned} \|U^k\|_\infty &\le \|u^k\|_\infty + \|e^k\|_\infty \le \|u^k\|_\infty + C_\Omega \|e^k\|_2 \\ &\le \|u^k\|_\infty + C_\Omega (C_1 \tau^{\min\{\gamma\alpha,2\}} + C_2 \epsilon) \le Q_1, \end{aligned} \tag{3.39}$$

whenever $\epsilon \le \epsilon_2 = 1/(2C_\Omega C_2)$ and $\tau < \tau_a = (2C_\Omega C_1)^{-\frac{1}{\min\{\gamma\alpha,2\}}}$. Noting $\|U^k\|_\infty$ and $\|u^k\|_\infty$ are bounded for all $0 \le k \le n-1$, we further have

$$\begin{aligned} \|E_1^{k-\sigma}\|_2 &\le \|f(u^{k-1}) + (1-\sigma) f'(u^{k-1}) \nabla_\tau u^k - f(U^{k-1}) - (1-\sigma) f'(U^{k-1}) \nabla_\tau U^k\|_2 \\ &\le \|f(u^{k-1}) - f(U^{k-1})\|_2 + (1-\sigma) \|\big(f'(u^{k-1}) - f'(U^{k-1})\big) u^k\|_2 \\ &\quad +(1-\sigma)\|f'(U^{k-1})(u^k - U^k)\|_2 + (1-\sigma)\|\big(f'(u^{k-1}) - f'(U^{k-1})\big) u^{k-1}\|_2 \\ &\quad +(1-\sigma)\|f'(U^{k-1})(u^{k-1} - U^{k-1})\|_2 \\ &\le C_4 \|e^{k-1}\|_2 + C_4 \|e^k\|_2, \end{aligned} \tag{3.40}$$

where C_4 is a constant dependent on u, σ, f, Q_1.

Taking the inner production with $e^{k-\sigma}$ both sides of (3.34) for $k = n$, we have

$$\begin{aligned} \big({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha e^{k-\sigma}, e^{k-\sigma}\big) &= (\Delta e^{k-\sigma}, e^{k-\sigma}) + (E_1^{k-\sigma}, e^{k-\sigma}) + (R_\sigma^{k-\sigma}, e^{k-\sigma}) + (R_t^{k-\sigma}, e^{k-\sigma}) + (R_f^{k-\sigma}, e^{k-\sigma}) \\ &= -\|\nabla e^{k-\sigma}\|_0^2 + (E_1^{k-\sigma}, e^{k-\sigma}) + (R_\sigma^{k-\sigma} + R_t^{k-\sigma} + R_f^{k-\sigma}, e^{k-\sigma}) \\ &\le (E_1^{k-\sigma}, e^{k-\sigma}) + (R_\sigma^{k-\sigma} + R_t^{k-\sigma} + R_f^{k-\sigma}, e^{k-\sigma}) \\ &\le \frac{C_5}{2}\|e^{k-1}\|_0^2 + \frac{C_5}{2}\|e^k\|_0^2 + \|R_\sigma^{k-\sigma} + R_t^{k-\sigma} + R_f^{k-\sigma}\|_0 \|e^{k-\sigma}\|_0, \end{aligned} \tag{3.41}$$

where C_5 is a constant dependent on u, σ, f, Q_1. By Lemma 3.1 (3.41) can be rewritten as

$$\frac{1}{2}\sum_{i=1}^{k} B_{k-i}^{(k)} \nabla_\tau ||e^i||_0^2 \le \frac{C_5}{2}\|e^{k-1}\|_0^2 + \frac{C_5}{2}\|e^k\|_0^2 + \|R_\sigma^{k-\sigma} + R_t^{k-\sigma} + R_f^{k-\sigma}\|_0 \|e^{k-\sigma}\|_0. \tag{3.42}$$

Recalling Lemma 3.2 and taking $\tau < \tau_b = 1/\sqrt[\alpha]{8C_5\Gamma(2-\alpha)}$, it holds

$$\|e^k\|_0 \le 4E_\alpha(8C_5 \max\{1, \rho\} t_k^\alpha)\Big(\max_{1\le j\le k}\sum_{i=1}^{j} P_{j-i}^{(j)} \|R_\sigma^{i-\sigma} + R_t^{i-\sigma} + R_f^{i-\sigma}\|_0\Big). \tag{3.43}$$

Similarly, taking the inner production with $-\Delta e^{k-\sigma}$ both sides of (3.34) yields

$$\begin{aligned} &\big({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha e^{k-\sigma}, -\Delta e^{k-\sigma}\big) \\ &= (\Delta e^{k-\sigma}, -\Delta e^{k-\sigma}) + (E_1^{k-\sigma}, -\Delta e^{k-\sigma}) + (R_\sigma^{k-\sigma} + R_t^{k-\sigma} + R_f^{k-\sigma}, -\Delta e^{k-\sigma}) \\ &= -\|\Delta e^{k-\sigma}\|_0^2 + (\nabla E_1^{k-\sigma}, \nabla e^{k-\sigma}) + (\nabla(R_\sigma^{k-\sigma} + R_t^{k-\sigma} + R_f^{k-\sigma}), \nabla e^{k-\sigma}) \\ &\le (\nabla E_1^{k-\sigma}, \nabla e^{k-\sigma}) + \big(\nabla(R_\sigma^{k-\sigma} + R_t^{k-\sigma} + R_f^{k-\sigma}), \nabla e^{k-\sigma}\big) \\ &\le \frac{C_6}{2}\|\nabla e^{k-1}\|_0^2 + \frac{C_6}{2}\|\nabla e^k\|_0^2 + \|\nabla(R_\sigma^{k-\sigma} + R_t^{k-\sigma} + R_f^{k-\sigma})\|_0 \|\nabla e^{k-\sigma}\|_0, \end{aligned} \tag{3.44}$$

where C_6 is a constant dependent on u, σ, f, Q_1.

Recalling Lemma 3.2 again and taking $\tau < \tau_c = 1/\sqrt[\alpha]{8C_6\Gamma(2-\alpha)}$, it holds

$$\|\nabla e^k\|_0 \le 4E_\alpha(8C_6 \max\{1,\rho\}t_k^\alpha)\Big(\max_{1\le j\le k}\sum_{i=1}^{j} P_{j-i}^{(j)}\|\nabla(R_\sigma^{i-\sigma}+R_t^{i-\sigma}+R_f^{i-\sigma})\|_0\Big). \tag{3.45}$$

Next, multiplying (3.34) by $\Delta^2 e^{k-\sigma}$ and integrating the result over Ω, we get

$$\begin{aligned}\big({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha e^{k-\sigma}, \Delta^2 e^{k-\sigma}\big) &= (\Delta e^{k-\sigma}, \Delta^2 e^{k-\sigma}) + (E_1^{k-\sigma}, \Delta^2 e^{k-\sigma}) + (R_\sigma^{k-\sigma}+R_t^{k-\sigma}+R_f^{k-\sigma}, \Delta^2 e^{k-\sigma})\\ &= -\|\Delta\nabla e^{k-\sigma}\|_0^2 + (\Delta E_1^k, \Delta e^{k-\sigma}) + (\Delta(R_\sigma^{k-\sigma}+R_t^{k-\sigma}+R_f^{k-\sigma}), \Delta e^{k-\sigma})\\ &\le (\Delta E_1^{k-\sigma}, \Delta e^{k-\sigma}) + \big(\Delta(R_\sigma^{k-\sigma}+R_t^{k-\sigma}+R_f^{k-\sigma}), \Delta e^{k-\sigma}\big)\\ &\le \frac{C_7}{2}\|\Delta e^{k-1}\|_0^2 + \frac{C_7}{2}\|\Delta e^k\|_0^2 + \|\Delta(R_\sigma^{k-\sigma}+R_t^{k-\sigma}+R_f^{k-\sigma})\|_0\|\Delta e^{k-\sigma}\|_0, \end{aligned}\tag{3.46}$$

where C_7 is a constant dependent on u, σ, f, Q_1.

Recalling Lemma 3.2 again and taking $\tau < \tau_d = 1/\sqrt[\alpha]{8C_7\Gamma(2-\alpha)}$, it holds

$$\|\Delta e^k\|_0 \le 4E_\alpha(8C_7 \max\{1,\rho\}t_k^\alpha)\Big(\max_{1\le j\le k}\sum_{i=1}^{j} P_{j-i}^{(j)}\|\Delta(R_\sigma^{i-\sigma}+R_t^{i-\sigma}+R_f^{i-\sigma})\|_0\Big). \tag{3.47}$$

Based on Lemmas 3.3, 3.4 and 3.5, one has

$$\begin{aligned}&\sum_{i=1}^{j} P_{j-i}^{(j)}\|R_\sigma^{i-\sigma}+R_t^{i-\sigma}+R_f^{i-\sigma}\|_2\\ \le\ &\sum_{i=1}^{j} P_{j-i}^{(j)}\|R_\sigma^{i-\sigma}\|_2 + \sum_{i=1}^{j} P_{j-i}^{(j)}\|R_t^{i-\sigma}\|_2 + \sum_{i=1}^{j} P_{j-i}^{(j)}\|R_f^{i-\sigma}\|_2\\ \le\ &C_v\Big(\frac{\tau_1^{2\alpha}}{\alpha} + t_j^\alpha \max_{2\le i\le j} t_{i-1}^{\alpha-2}\tau_i^2 + \frac{\tau_1^\alpha}{\alpha} + \frac{1}{1-\alpha}\max_{2\le i\le j} t_i^\alpha t_{i-1}^{\alpha-3}\tau_i^3\tau_{i-1}^{-\alpha} + \frac{\epsilon}{\alpha}t_j^\alpha \hat{t}_{j-1}^\alpha + \tau_1^{3\alpha} + t_j^\alpha \max_{2\le i\le j}\tau_i^2 t_{i-1}^{2(\alpha-1)}\Big)\\ \le\ &C_v\Big(\frac{\tau_1^\alpha}{\alpha} + t_j^\alpha \max_{2\le i\le n} t_{i-1}^{\alpha-2}\tau_i^2 + \frac{1}{1-\alpha}\max_{2\le i\le j} t_i^\alpha t_1^{-3}\tau^3 + \frac{\epsilon}{\alpha}t_j^\alpha \hat{t}_{j-1}^\alpha + t_j^\alpha \max_{2\le i\le j}\tau_i^2 t_{i-1}^{2(\alpha-1)}\Big)\\ \le\ &C\big(\tau^{\min\{\gamma\alpha,2\}} + \frac{\epsilon}{\alpha}(T\hat{T})^\alpha\big), \end{aligned}\tag{3.48}$$

where $\hat{T} = \max\{1, T\}$. Then putting (3.48) into (3.43), (3.45) and (3.47), we can obtain

$$\|e^k\|_2 \le C_1\big(\tau^{\min\{\gamma\alpha,2\}} + \frac{\epsilon}{\alpha}(T\hat{T})^\alpha\big) = C_1\tau^{\min\{\gamma\alpha,2\}} + C_2\epsilon, \tag{3.49}$$

where $C_2 = \frac{(T\hat{T})^\alpha}{\alpha}C_1$ and

$$C_1 = 4C_u\sqrt{E_\alpha^2(8C_5\max\{1,\rho\}T^\alpha) + E_\alpha^2(8C_6\max\{1,\rho\}T^\alpha) + E_\alpha^2(8C_7\max\{1,\rho\}T^\alpha)}.$$

Then, whenever $\epsilon \le \epsilon_2$ and $\tau < \tau_a$, we have

$$\|U^k\|_\infty \le \|u^k\|_\infty + \|e^k\|_\infty \le \|u^k\|_\infty + C_\Omega\big(C_1\tau^{\min\{\gamma\alpha,2\}} + C_2\epsilon\big) \le Q_1. \tag{3.50}$$

Thus, the estimates (3.36) and (3.37) hold for $n = k$ by taking $\tau^{**} = \min\{\tau_a, \tau_b, \tau_c, \tau_d\}$ and $\epsilon \leq \epsilon_2$. The mathematical induction is finished.

Based on the above results, we now consider the proof of (3.38). By the definition, we have

$$\begin{aligned}
\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha e^{n-\sigma}\|_2 &= \|B_0^{(n)} e^n - \sum_{k=1}^{n-1}(B_{n-k-1}^{(n)} - B_{n-k}^{(n)})e^k - B_{n-1}^{(n)} e^0\|_2 \\
&= B_0^{(n)}\|e^n\|_2 + \sum_{k=1}^{n-1}(B_{n-k-1}^{(n)} - B_{n-k}^{(n)})\|e^k\|_2 + B_{n-1}^{(n)}\|e^0\|_2 \\
&\leq \Big(B_0^{(n)} + \sum_{k=1}^{n-1}(B_{n-k-1}^{(n)} - B_{n-k}^{(n)}) + B_{n-1}^{(n)}\Big)\big(C_1\tau^{\min\{\gamma\alpha,2\}} + C_2\epsilon\big) \\
&\leq 2B_0^{(n)}\big(C_1\tau^{\min\{\gamma\alpha,2\}} + C_2\epsilon\big) \leq \frac{48\tau_n^{-\alpha}}{11\Gamma(2-\alpha)}\big(C_1\tau^{\min\{\gamma\alpha,2\}} + C_2\epsilon\big) \\
&\leq \frac{48}{11\Gamma(2-\alpha)}\big(C_1 + C_2\big), \qquad (3.51)
\end{aligned}$$

where we apply the properties **M3** in the penultimate inequality and take $\epsilon < \epsilon_3 = \tau^{\min\{\gamma\alpha,2\}}$ and $\gamma\alpha \leq 2$ in the last inequality. Therefore,

$$\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha U^{n-\sigma}\|_2 \leq \|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha u^{n-\sigma}\|_2 + \|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha e^{n-\sigma}\|_2 \leq C_3. \qquad (3.52)$$

The proof is completed after taking $\epsilon^* = \min\{\epsilon_1, \epsilon_2, \epsilon_3\}$. □

3.3 Analysis of the fully discrete scheme

We now consider the boundedness of the fully discrete solution U_h^n. To analyze the fully discrete scheme (2.18), we introduce the Ritz projection operator $R_h : H_0^1(\Omega) \to V_h \subset H_0^1(\Omega)$ by

$$\big(\nabla(v - R_h v), \nabla\omega\big) = 0, \quad \forall\omega \in V_h.$$

For $\forall v \in H^s(\Omega) \cap H_0^1(\Omega)$, it holds

$$\|v - R_h v\|_0 + h\|\nabla(v - R_h v)\|_0 \leq C_\Omega h^s\|v\|_s, \quad 1 \leq s \leq r + 1. \qquad (3.53)$$

Let

$$U^n - U_h^n = U^n - R_h U^n + R_h U^n - U_h^n = \rho_h^n + \theta_h^n, \quad n = 0, 1, \ldots, N. \qquad (3.54)$$

The weak form of the semi-discrete (3.31) is

$$\big({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha U^{n-\sigma}, v\big) = -(\nabla U^{n-\sigma}, \nabla v) + \big(f(U^{n-1}) + (1-\sigma)f'(U^{n-1})\nabla_\tau U^n, v\big), \ \forall\, v \in V_h. \quad (3.55)$$

Subtracting (2.18) from (3.55), we get

$$\big({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha \theta_h^{n-\sigma}, v\big) = -(\nabla\theta_h^{n-\sigma}, \nabla v) + (E_2^{n-\sigma}, v) - \big({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha \rho_h^{n-\sigma}, v\big), \qquad (3.56)$$

where

$$E_2^{n-\sigma} = f(U^{n-1}) + (1-\sigma) f'(U^{n-1}) \nabla_\tau U^n - f(U_h^{n-1}) - (1-\sigma) f'(U_h^{n-1}) \nabla_\tau U_h^n. \tag{3.57}$$

Then, based on the boundedness of $||U^n||_\infty$ in Theorem 3.1, we present the following result.

Theorem 3.2 *Suppose the semi-discrete scheme* (3.55) *has a unique solution* U^n *and the fully discrete scheme defined in* (2.18) *has a unique solution* U_h^n, $n = 1, ..., N$. *If* $f \in C^4(\mathbb{R})$, *there exist* $h^* > 0$ *and* $\tau^{***} > 0$ *such that when* $h < h^*$ *and* $\tau < \tau^{***}$, *it holds*

$$\|\theta_h^n\|_0 \le h^{\frac{7}{4}}, \tag{3.58}$$

$$\|U_h^n\|_\infty \le Q_2, \tag{3.59}$$

where $Q_2 = \max\limits_{1\le n\le N} \|R_h U^n\|_\infty + 1$.

Proof Noting the coefficient matrix of the linear system arising from (2.18) is diagonally dominant, the solution U_h^n of (2.18) exists and is unique. Here, we also apply the mathematical induction to prove (3.58). It is easy to show (3.58) hold for $n = 0$. Next, we assume (3.58) holds for $1 \le k \le n-1$, and have

$$\begin{aligned} \|U_h^k\|_\infty &\le \|R_h U^k\|_\infty + \|\theta_h^k\|_\infty \le \|R_h U^k\|_\infty + C_\Omega h^{-\frac{d}{2}} \|\theta_h^k\|_0 \\ &\le \|R_h U^k\|_\infty + C_\Omega h^{-\frac{d}{2}} h^{\frac{7}{4}} \le \|R_h U^k\|_\infty + 1 \le Q_2, \end{aligned} \tag{3.60}$$

whenever $h < h_1 = C_\Omega^{-\frac{4}{7-2d}}$.

Similar to the estimate (3.40) of $E_1^{k-\sigma}$, we use the boundedness of $||U^k||_\infty$ and $||U_h^k||_\infty$ for all $0 \le k \le n-1$ to obtain

$$\begin{aligned} \|E_2^{k-\sigma}\|_0 &\le \|f(U^{k-1}) + (1-\sigma) f'(U^{k-1}) \nabla_\tau U^k - f(U_h^{k-1}) - (1-\sigma) f'(U_h^{k-1}) \nabla_\tau U_h^k\|_0 \\ &\le \|f(U^{k-1}) - f(U_h^{k-1})\|_0 + (1-\sigma)\|\big(f'(U^{k-1}) - f'(U_h^{k-1})\big) U^k\|_0 \\ &\quad + (1-\sigma)\|f'(U_h^{k-1})(U^k - U_h^k)\|_0 + (1-\sigma)\|\big(f'(U^{k-1}) - f'(U_h^{k-1})\big) U^{k-1}\|_0 \\ &\quad + (1-\sigma)\|f'(U_h^{k-1})(U^{k-1} - U_h^{k-1})\|_0 \\ &\le C_8\|U^{k-1} - U_h^{k-1}\|_0 + C_8\|U^k - U_h^k\|_0 \\ &\le C_8\|\theta_h^{k-1}\|_0 + C_8\|\theta_h^k\|_0 + 2C_8 C_\Omega h^2, \end{aligned} \tag{3.61}$$

where C_8 is a constant dependent on σ, f, Q_2 and (3.53) is used in the last inequality. Setting $k = n$ and $v = \theta_h^{k-\sigma}$ in (3.56), we obtain

$$\begin{aligned} \left({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha \theta_h^{k-\sigma}, \theta_h^{k-\sigma}\right) &= -(\nabla\theta_h^{k-\sigma}, \nabla\theta_h^{k-\sigma}) + (E_2^{k-\sigma}, \theta_h^{k-\sigma}) - \left({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha \rho_h^{k-\sigma}, \theta_h^{k-\sigma}\right) \\ &\le -\|\nabla\theta_h^{k-\sigma}\|_0^2 + \frac{C_9}{2}\|\theta_h^k\|_0^2 + \frac{C_9}{2}\|\theta_h^{k-1}\|_0^2 + \big(2C_8 C_\Omega h^2 + \|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha \rho_h^{k-\sigma}\|_0\big)\|\theta_h^{k-\sigma}\|_0 \\ &\le \frac{C_9}{2}\|\theta_h^k\|_0^2 + \frac{C_9}{2}\|\theta_h^{k-1}\|_0^2 + \big(2C_8 C_\Omega h^2 + C_\Omega\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha U^{k-\sigma}\|_2 h^2\big)\|\theta_h^{k-\sigma}\|_0, \end{aligned} \tag{3.62}$$

where C_9 is a constant dependent on σ, f, Q_2. Applying Lemma 3.2, we have

$$\begin{aligned}\|\theta_h^k\|_0 &\le 2E_\alpha(8C_9\max\{1,\rho\}t_k^\alpha)\Big(\|\theta_h^0\|_0+4t_k^\alpha\Gamma(1-\alpha)\big(2C_8C_\Omega h^2+\max_{1\le i\le k}C_\Omega\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha U^{i-\sigma}\|_2h^2\big)\Big)\\&\le 2E_\alpha(8C_9\max\{1,\rho\}t_k^\alpha)\Big(C_\Omega+4T^\alpha\Gamma(1-\alpha)\big(2C_8C_\Omega+\max_{1\le i\le k}C_\Omega\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha U^{i-\sigma}\|_2\big)\Big)h^2\\&\le h^{\frac{7}{4}},\end{aligned}\tag{3.63}$$

when $\tau<\tau_e=1/\sqrt[\alpha]{8C_9\Gamma(2-\alpha)}$ and

$h<h_2=\Big(2E_\alpha(8C_9\max\{1,\rho\}t_k^\alpha)\big(C_\Omega+4T^\alpha\Gamma(1-\alpha)(2C_8C_\Omega+\max_{1\le i\le k}C_\Omega\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha U^{i-\sigma}\|_2)\big)\Big)^{-4}$.

Then, when $h<h_1$, we can verify that

$$\|U_h^k\|_\infty\le\|R_hU^k\|_\infty+\|\theta_h^k\|_\infty\le\|R_hU^k\|_\infty+C_\Omega h^{-\frac{d}{2}}\|\theta_h^k\|_0\le Q_2.\tag{3.64}$$

Thus, the estimates (3.58) and (3.59) hold for $n=k$ by taking $h^*=\min\{h_1,h_2\}$ and $\tau^{***}=\tau_e$. The proof is completed. □

3.4 The proof of Theorem 2.1

Noting the boundedness of $||U^n||_\infty$ in Theorem 3.1, we can use the method in [33, 34] to get the estimate of $\|u^n-U_h^n\|_1$ for the linear element (i.e., $r=1$). The method in [33, 34] will become too complicated to be used for high-order elements (i.e., $r\ge2$). Hence, the following proof will be split into two cases: one is for $r=1$ and another one is for $r\ge2$.

Proof We first prove the case of $r=1$ by taking $v={}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\theta_h^{n-\sigma}$ in (3.56) to get

$$\begin{aligned}\big({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\theta_h^{n-\sigma},{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\theta_h^{n-\sigma}\big)&=-\big(\nabla\theta_h^{n-\sigma},{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\nabla\theta_h^{n-\sigma}\big)+\big(E_2^{n-\sigma},{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\theta_h^{n-\sigma}\big)-\big({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\rho_h^{n-\sigma},{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\theta_h^{n-\sigma}\big)\\&\le-\big(\nabla\theta_h^{n-\sigma},{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\nabla\theta_h^{n-\sigma}\big)+\frac{C_{10}}{2}\|\theta_h^n\|_0^2+\frac{C_{10}}{2}\|\theta_h^{n-1}\|_0^2+\frac{C_{10}}{2}h^4\\&\quad+\frac12\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\theta_h^{n-\sigma}\|_0^2+\frac12\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\rho_h^{n-\sigma}\|_0^2+\frac12\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\theta_h^{n-\sigma}\|_0^2,\end{aligned}\tag{3.65}$$

where C_{10} is a constant dependent on σ, f, Q_2. Rearranging (3.65) produces

$$\begin{aligned}\big({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\nabla\theta_h^{n-\sigma},\nabla\theta_h^{n-\sigma}\big)&\le\frac{C_{10}}{2}\|\theta_h^n\|_0^2+\frac{C_{10}}{2}\|\theta_h^{n-1}\|_0^2+\frac{C_{10}}{2}h^4+\frac12\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\rho_h^{n-\sigma}\|_0^2\\&\le\frac{C_{10}}{2}\|\theta_h^n\|_0^2+\frac{C_{10}}{2}\|\theta_h^{n-1}\|_0^2+\frac{C_{10}}{2}h^4+\frac12(C_\Omega\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha U^{n-\sigma}\|_2h^2)^2\\&\le\frac{C_{10}C_\Omega}{2}\|\nabla\theta_h^n\|_0^2+\frac{C_{10}C_\Omega}{2}\|\nabla\theta_h^{n-1}\|_0^2+\frac{C_{10}}{2}h^4+\frac12(C_\Omega\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha U^{n-\sigma}\|_2h^2)^2.\end{aligned}$$

From Lemma 3.2, it holds

$$\begin{aligned}\|\nabla\theta_h^n\|_0&\le 2E_\alpha(8C_{10}C_\Omega\max\{1,\rho\}t_n^\alpha)\\&\qquad\Big(\|\nabla\theta_h^0\|_0+\sqrt{2\Gamma(1-\alpha)}t_n^{\frac{\alpha}{2}}\big(\max_{1\le k\le n}C_\Omega\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha U^{k-\sigma}\|_2+\sqrt{C_{10}}\big)h^2\Big),\end{aligned}\tag{3.66}$$

when $\tau \leq \tau_f = 1/\sqrt[\alpha]{8C_{10}C_\Omega\Gamma(2-\alpha)}$. Therefore, we have

$$\begin{aligned}\|u^n - U_h^n\|_1 &\leq \|u^n - U^n\|_1 + \|U^n - R_hU^n\|_1 + \|R_hU^n - U_h^n\|_1 \\ &= \|e^n\|_1 + \|\rho_h^n\|_1 + \|\theta_h^n\|_1 \leq C^*\big(\tau^{\min\{\gamma\alpha,2\}} + h + \epsilon\big). \end{aligned} \tag{3.67}$$

Then, we prove the case of $r \geq 2$ by considering the exact solution u^n satisfies

$$\big({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha u^{n-\sigma}, v\big) = -(\nabla u^{n-\sigma}, \nabla v) + \big(f(u^{n-1}) + (1-\sigma)f'(u^{n-1})\nabla_\tau u^n, v\big), \quad \forall\, v \in V_h. \tag{3.68}$$

Set

$$u^n - U_h^n = u^n - R_hu^n + R_hu^n - U_h^n = \xi_h^n + \eta_h^n, \quad n = 0, 1, \ldots, N. \tag{3.69}$$

Subtracting (2.18) from (3.68), we get

$$\big({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha \eta_h^{n-\sigma}, v\big) = -(\nabla\eta_h^{n-\sigma}, \nabla v) + (E_3^{n-\sigma}, v) + (R_t^{n-\sigma} + R_\sigma^{n-\sigma} + R_f^{n-\sigma}, v) - \big({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha \xi_h^{n-\sigma}, v\big), \tag{3.70}$$

where

$$E_3^{n-\sigma} = f(u^{n-1}) + f'(u^{n-1})\nabla_\tau u^n - f(U_h^{n-1}) - (1-\sigma)f'(U_h^{n-1})\nabla_\tau U_h^n. \tag{3.71}$$

It is obvious that $||u^n||_\infty$ and $||U_h^n||_\infty$ are bounded for $1 \leq n \leq N$. So we obtain

$$\begin{aligned}\|\nabla E_3^{n-\sigma}\|_0 &\leq \|\nabla\big(f(u^{n-1}) + (1-\sigma)f'(u^{n-1})\nabla_\tau u^n\big) - \nabla\big(f(U_h^{n-1}) + (1-\sigma)f'(U_h^{n-1})\nabla_\tau U_h^n\big)\|_0 \\ &\leq \|\nabla\big(f(u^{n-1}) - f(U_h^{n-1})\big)\|_0 + (1-\sigma)\|\nabla\big((f'(u^{n-1}) - f'(U_h^{n-1}))u^n\big)\|_0 \\ &\quad +(1-\sigma)\|\nabla\big(f'(U_h^{n-1})(u^n - U_h^n)\big)\|_0 + (1-\sigma)\|\nabla\big((f'(u^{n-1}) - f'(U_h^{n-1}))u^{n-1}\big)\|_0 \\ &\quad +(1-\sigma)\|\nabla\big(f'(U_h^{n-1})(u^{n-1} - U_h^{n-1})\big)\|_0 \\ &\leq C_{11}\|\nabla(u^{n-1} - U_h^{n-1})\|_0 + C_{11}\|\nabla(u^n - U_h^n)\|_0 \\ &\leq C_{11}\|\nabla\eta_h^{n-1}\|_0 + C_{11}\|\nabla\eta_h^n\|_0 + C_{11}\|\nabla\xi_h^{n-1}\|_0 + C_{11}\|\nabla\xi_h^n\|_0 \\ &\leq C_{11}\|\nabla\eta_h^{n-1}\|_0 + C_{11}\|\nabla\eta_h^n\|_0 + 2C_{11}C_\Omega h^r, \end{aligned} \tag{3.72}$$

where C_{11} is a constant dependent on u, σ, f and Q_2. Next, taking $v = -\Delta\eta_h^{n-\sigma}$ in (3.68) , we get

$$\begin{aligned}\big({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha \eta_h^{n-\sigma}, -\Delta\eta_h^{n-\sigma}\big) &= -(\nabla\eta_h^{n-\sigma}, -\nabla\Delta\eta_h^{n-\sigma}) + (E_3^{n-\sigma}, -\Delta\eta_h^{n-\sigma}) \\ &\quad +(R_t^{n-\sigma} + R_\sigma^{n-\sigma} + R_f^{n-\sigma}, -\Delta\eta_h^{n-\sigma}) - \big({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha \xi_h^{n-\sigma}, -\Delta\eta_h^{n-\sigma}\big) \\ &= -\|\Delta\eta_h^{n-\sigma}\|_0^2 + (\nabla E_3^{n-\sigma}, \nabla\eta_h^{n-\sigma}) \\ &\quad +(\nabla R_t^{n-\sigma} + \nabla R_\sigma^{n-\sigma} + \nabla R_f^{n-\sigma}, \nabla\eta_h^{n-\sigma}) - \big({}^{\mathrm{F}}\mathcal{D}_\tau^\alpha \nabla\xi_h^{n-\sigma}, \nabla\eta_h^{n-\sigma}\big) \\ &\leq \frac{C_{12}}{2}\|\nabla\eta_h^k\|_0^2 + \frac{C_{12}}{2}\|\nabla\eta_h^{n-1}\|_0^2 + \big(2C_{11}C_\Omega h^r \\ &\quad +\|\nabla(R_t^{n-\sigma} + R_\sigma^{n-\sigma} + R_f^{n-\sigma})\|_0 + \|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\nabla\xi_h^{n-\sigma}\|_0\big)\|\nabla\eta_h^{n-\sigma}\|_0 \\ &\leq \frac{C_{12}}{2}\|\nabla\eta_h^k\|_0^2 + \frac{C_{12}}{2}\|\nabla\eta_h^{n-1}\|_0^2 + \big(2C_{11}C_\Omega h^r + \|\nabla(R_t^{n-\sigma} + R_\sigma^{n-\sigma} \\ &\quad +R_f^{n-\sigma})\|_0 + C_\Omega\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\nabla u^{n-\sigma}\|_{r+1}h^r\big)\|\nabla\eta_h^{n-\sigma}\|_0, \end{aligned} \tag{3.73}$$

where C_{12} is a constant dependent on u, σ, f and Q_2. By Lemma 3.2 and (3.48), it holds

$$\begin{aligned}\|\nabla\eta_h^n\|_0 &\le 2E_\alpha(8C_{12}t_n^\alpha)\Big(\|\nabla\eta_h^0\|_0^2+4t_n^\alpha\Gamma(1-\alpha)\big(2C_{11}C_\Omega h^r+\max_{1\le i\le n}C_\Omega\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\nabla u^i\|_{r+1}h^r\big)\\ &\quad+\max_{1\le j\le n}\sum_{i=1}^{j}P_{j-i}^{(j)}\|\nabla(R_t^{i-\sigma}+R_\sigma^{i-\sigma}+R_f^{i-\sigma})\|_0\Big)\\ &\le 2E_\alpha(8C_{12}t_n^\alpha)\Big(C_\Omega+4T^\alpha\Gamma(1-\alpha)\big(2C_{11}C_\Omega h^r+\max_{1\le i\le n}C_\Omega\|{}^{\mathrm{F}}\mathcal{D}_\tau^\alpha\nabla u^i\|_{r+1}h^r\big)\\ &\quad+C\big(\tau^{\min\{\gamma\alpha,2\}}+\frac{\epsilon}{\alpha}(T\hat{T})^\alpha\big)\Big),\end{aligned}\tag{3.74}$$

when $\tau\le\tau_g=1/\sqrt[\alpha]{8C_{12}\Gamma(2-\alpha)}$. Therefore, we have

$$\|u^n-U_h^n\|_1\le\|u^n-R_hu^n\|_1+\|R_hu^n-U_h^n\|_1=\|\xi_h^n\|_1+\|\eta_h^n\|_1\le C^*\big(\tau^{\min\{\gamma\alpha,2\}}+h^r+\epsilon\big),\tag{3.75}$$

whenever $\gamma\alpha\le 2$, $\tau^*=\min\{\tau^{**},\tau^{***},\tau_f,\tau_g\}$, $h^*=\min\{h_1,h_2\}$ and $\epsilon^*=\min\{\epsilon_1,\epsilon_2,\epsilon_3\}$. Thus the proof of Theorem 2.1 is completed. □

4 Numerical examples

We now provide two examples to demonstrate the unconditionally optimal convergence orders with $\mathcal{O}(\tau^{\min\{\gamma\alpha,2\}})$ in time and $\mathcal{O}(h^r)$ in space as presented in Theorem 2.1. Here we consider the graded meshes $t_k=(k/N)^\gamma$ $(k=1,\ldots,N)$, and divide the space Ω into M parts with quasi-uniform quadrilateral partition $\mathcal{T}_i$ $(i=1,\cdots,M)$ with maximum mesh size $h=\max_{1\le i\le M}\{\mathrm{diam}\,\mathcal{T}_i\}$. The error is calculated by H^1-norm in space and the maximum norm in time.

Example 4.1 We first consider the following nonlinear subdiffusion equation

$${}^{\mathrm{C}}\mathcal{D}_t^\alpha u=\Delta u+\sin(u)+g(x,y,t),\quad (x,y)\in(0,1)^2,\quad t\in(0,1].$$

As a benchmark solution to investigate the convergence orders in time and space by using the linear and quadratic elements, the exact solution is constructed by $u(x,y,t)=(t^2+t^\alpha)x(1-x)y(1-y)$.

Tables 1 and 2 show the temporal errors and convergence orders by taking $N=8,16,32,64$ and $M=\lceil N^{\gamma\alpha}\rceil$ with linear element for $\alpha=0.5$ and $\alpha=0.8$, respectively. Table 3 presents the numerical results of the linear and quadratic elements for $\alpha=0.5$, $\gamma=2$, $N=10^4$, $M=12,24,48,96$, which illustrates the r-degree finite element method has r-order accuracy.

The CPU time of the direct algorithm (2.7) and the fast algorithm (2.16) is given in Table 4, which is calculated by using the linear finite element with $M=15$, $\alpha=0.5$, $\gamma=2$ and changing N from 1000 to 16000. One can see that the fast L2-1_σ scheme can speed up the simulations significantly as N is larger.

If a numerical method is unconditionally convergent, given a N, the error should be tended to a constant as M is taken larger and larger. The phenomenon is displayed in Fig. 1 by respectively taking the linear and quadratic elements with $\alpha=0.5$, $\gamma=2$

Table 1 Errors and convergence orders of temporal directions for Example 4.1

α	N	$\gamma = 2$		$\gamma = 4$	
		Error	Order	Error	Order
0.5	8	3.72e-02	–	6.50e-03	–
	16	1.87e-02	1.00	1.74e-03	1.90
	32	9.32e-03	1.00	4.51e-04	1.95
	64	4.66e-03	1.00	1.15e-04	1.98
	$\gamma\alpha$		1		2

Table 2 Errors and convergence orders of temporal directions for Example 4.1

α	N	$\gamma = 1.25$		$\gamma = 2.5$	
		Error	Order	Error	Order
0.8	8	3.74e-02	–	5.50e-03	–
	16	1.86e-02	1.00	1.39e-03	1.99
	32	9.32e-03	1.00	3.48e-04	1.99
	64	4.66e-03	1.00	8.7037e-05	2.00
	$\gamma\alpha$		1		2

Table 3 Errors and convergence orders of spatial directions when $\alpha = 0.5$, $\gamma = 2$ and $N = 10^4$ for Example 4.1

M	Linear element		Quadratic element	
	Error	Order	Error	Order
12	2.49e-02	–	1.87e-04	–
24	1.24e-02	1.00	3.52e-05	2.41
48	6.21e-03	1.00	6.44e-06	2.45
96	3.11e-03	1.00	1.42e-06	2.18

Table 4 Computational time with $M = 15$, $\alpha = 0.5$, $\gamma = 2$ for Example 4.1

N	1000	2000	4000	8000	16000
fast algorithm	18.7s	59.8s	138s	289s	623s
direct algorithm	98.1s	472s	2104s	7546s	24920s

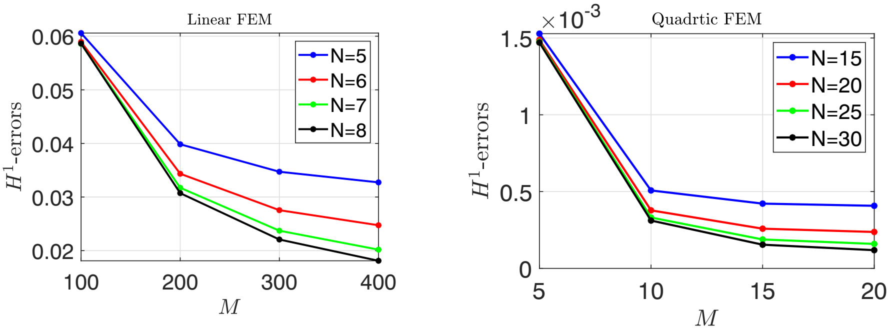

Fig. 1 H^1-errors of linear and quadratic finite elements with $\alpha = 0.5$ and $\gamma = 2$ for Example 4.1

for different M and N. Figure 1 shows that our error estimates are unconditionally convergent.

Example 4.2 Consider the following two-dimensional nonlinear subdiffusion equation

$$ {}^{\mathrm{C}}\mathcal{D}_t^{\alpha} u = \Delta u + u(1 - u^2) + g(x, y, t), \quad (x, y) \in (0, 1)^2, \quad t \in (0, 1], $$

with $u(x, y, t) = (1 + t^{\alpha}) \sin(\pi x) \sin(\pi y)$ to investigate the convergence order in time and space by using the linear and quadratic elements.

Tables 5 and 6 show the errors and convergence orders of temporal directions by taking $N = 8, 16, 32, 64$ and $M = \lceil N^{\gamma\alpha} \rceil$ with linear element for $\alpha = 0.5$ and $\alpha = 0.8$, respectively. Table 7 presents the numerical results of the linear and quadratic elements for $\alpha = 0.5$, $\gamma = 2$, $N = 10^4$, $M = 8, 16, 32, 64$, which illustrates the r-degree finite element method has r-order accuracy again.

In Table 8, the computational time of the direct algorithm (2.7) is compared with the fast algorithm (2.16) with $M = 10$, $\alpha = 0.3$ and $\gamma = 2$ for N from 1000 to 16000. One can see that the smaller the time step, the more effective the fast algorithm.

In Fig. 2, the errors of the linear (on the left) and quadratic (on the right) elements are shown with a fixed N and increasing M for $\alpha = 0.3$, $\gamma = 2$. The figure shows

Table 5 Errors and convergence orders of temporal directions for Example 4.2

α	N	$\gamma = 2$		$\gamma = 4$	
		Error	Order	Error	Order
0.5	8	5.03e-01	–	6.64e-02	–
	16	2.52e-01	1.00	1.67e-02	1.99
	32	1.26e-01	1.00	4.18e-03	2.00
	64	6.30e-01	1.00	1.05e-03	2.00
	$\gamma\alpha$		1		2

Table 6 Errors and convergence orders of temporal directions for Example 4.2

α	N	$\gamma = 1.25$		$\gamma = 2.5$	
		Error	Order	Error	Order
0.8	8	5.03e-01	–	6.38e-02	–
	16	2.52e-01	1.00	1.60e-02	2.00
	32	1.26e-01	1.00	3.99e-03	2.00
	64	6.30e-01	1.00	9.98e-04	2.00
	$\gamma\alpha$		1		2

Table 7 Errors and convergence orders of spatial directions when $\alpha = 0.5$, $\gamma = 2$ and $N = 10^4$ for Example 4.2

M	Linear element		Quadratic element	
	Error	Order	Error	Order
8	5.03e-01	–	2.66e-02	–
16	2.52e-01	1.00	6.56e-03	2.02
32	1.26e-01	1.00	1.62e-03	2.02
64	6.30e-02	1.00	4.02e-04	2.01

Table 8 Computational time with $M = 10$, $\alpha = 0.3$ and $\gamma = 2$ for Example 4.2

N	1000	2000	4000	8000	16000
Fast algorithm	10.7s	35.6s	85.8s	201s	562s
Direct algorithm	77.1s	399s	1980s	7063s	24115s

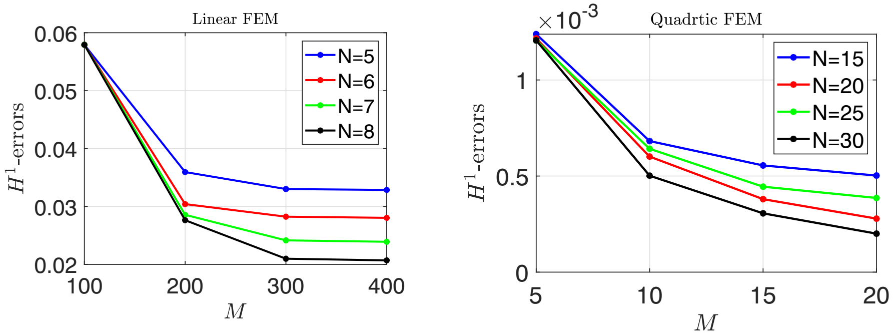

Fig. 2 H^1-errors of linear and quadratic finite elements with $\alpha = 0.3$ and $\gamma = 2$ for Example 4.2

that the errors tend to different constants which illustrates our theoretical analysis is unconditionally convergent.

5 Conclusion

The unconditionally optimal H^1-error estimate of the SOE-based fast L2-1_σ scheme is presented to numerically solve the nonlinear subdiffusion problem (1.1)–(1.3) on the nonuniform mesh. The SOE approximation for Caputo derivative can efficiently reduce the computational storage and cost when time steps are large. To deal with the initial singularity of the solution, a nonuniform mesh is used to have the globally optimal convergence, which also bring the complication of theoretical analysis. Thus, we used a modified discrete fractional Grönwall inequality to present the stability analysis, and introduced the ECS to express the truncation error of SOE-based fast L2-1_σ scheme. Combining with DCC kernels and ECS, the global error analysis is significantly simplified. For the nonlinear term, we consider a linearized scheme by approximating it with a Newton linearization method and approximate the dissipative term with implicit scheme. Then, we use the spatial-temporal splitting approach to prove that the proposed scheme is unconditionally convergent. Numerical tests are given to verify the effectiveness and optimal convergence of our scheme.

Appendix. The proof of Lemma 3.4

Proof By the Taylor expansion, we obtain

$$
\begin{aligned}
R_f^{k-\sigma} &= \int_{v^{k-1}}^{v^{k-\sigma}} (v^{k-\sigma}-\mu) f''(\mu)\,\mathrm{d}\mu \\
&= \int_0^1 \left[v^{k-\sigma}-v^{k-1}-s(v^{k-\sigma}-v^{k-1})\right] f''\left(v^{k-1}+s(v^{k-\sigma}-v^{k-1})\right)(v^{k-\sigma}-v^{k-1})\,\mathrm{d}s \\
&= (1-\sigma)^2(v^k-v^{k-1})^2\int_0^1 (1-s) f''\left(v^{k-1}+s(v^{k-\sigma}-v^{k-1})\right)\mathrm{d}s, \\
\nabla R_f^{k-\sigma} &= (1-\sigma)^2(v^k-v^{k-1})^2\int_0^1 (1-s) f'''\left(v^{k-1}+s(v^{k-\sigma}-v^{k-1})\right)\nabla\left(v^{k-1}+s(v^{k-\sigma}-v^{k-1})\right)\mathrm{d}s \\
&\quad +2(1-\sigma)^2(v^k-v^{k-1})\nabla(v^k-v^{k-1})\int_0^1 (1-s) f''\left(v^{k-1}+s(v^{k-\sigma}-v^{k-1})\right)\mathrm{d}s, \\
\Delta R_f^{k-\sigma} &= 2(1-\sigma)^2(v^k-v^{k-1})\nabla(v^k-v^{k-1})\int_0^1 (1-s) f'''\left(v^{k-1}+s(v^{k-\sigma}-v^{k-1})\right) \\
&\qquad \nabla\left(v^{k-1}+s(v^{k-\sigma}-v^{k-1})\right)\mathrm{d}s \\
&\quad +(1-\sigma)^2(v^k-v^{k-1})^2\int_0^1 (1-s) f^{(4)}\left(v^{k-1}+s(v^{k-\sigma}-v^{k-1})\right)\left(\nabla(v^{k-1}+s(v^{k-\sigma}-v^{k-1})))\right)^2\mathrm{d}s \\
&\quad +(1-\sigma)^2(v^k-v^{k-1})^2\int_0^1 (1-s) f'''\left(v^{k-1}+s(v^{k-\sigma}-v^{k-1})\right)\Delta\left(v^{k-1}+s(v^{k-\sigma}-v^{k-1})\right)\mathrm{d}s \\
&\quad +2(1-\sigma)^2\left(\nabla(v^k-v^{k-1})\right)^2\int_0^1 (1-s) f''\left(v^{k-1}+s(v^{k-\sigma}-v^{k-1})\right)\mathrm{d}s
\end{aligned}
$$

$$
\begin{aligned}
&+2(1-\sigma)^2(v^k-v^{k-1})\Delta(v^k-v^{k-1})\int_0^1(1-s)f''\left(v^{k-1}+s(v^{k-\sigma}-v^{k-1})\right)\mathrm{d}s\\
&+2(1-\sigma)^2(v^k-v^{k-1})\nabla(v^k-v^{k-1})\int_0^1(1-s)f'''\left(v^{k-1}+s(v^{k-\sigma}-v^{k-1})\right)\\
&\quad\nabla\left(v^{k-1}+s(v^{k-\sigma}-v^{k-1})\right)\mathrm{d}s.
\end{aligned}
$$

By the condition (1.5) of v, we have

$$
\left|R_f^{k-\sigma}\right| \le C_v\left(\int_{t_{k-1}}^{t_k}\left|v'(t)\right|dt\right)^2 \le C_v\left(\int_{t_{k-1}}^{t_k}(1+t^{\alpha-1})dt\right)^2 \le \begin{cases} C_v(\tau_1^2+\frac{\tau_1^{2\alpha}}{\alpha^2}), & k=1,\\ C_v(\tau_k^2+\tau_k^2t_{k-1}^{2(\alpha-1)}), & 2\le k\le n,\end{cases}
$$

$$
\left|\nabla R_f^{k-\sigma}\right| \le C_v\left(\int_{t_{k-1}}^{t_k}|v'(t)|dt\right)^2+C_v\int_{t_{k-1}}^{t_k}|v'(t)|dt\int_{t_{k-1}}^{t_k}|\nabla v'(t)|dt \le \begin{cases} C_v(\tau_1^2+\frac{\tau_1^{2\alpha}}{\alpha^2}), & k=1,\\ C_v(\tau_k^2+\tau_k^2t_{k-1}^{2(\alpha-1)}), & 2\le k\le n,\end{cases}
$$

$$
\begin{aligned}
\left|\Delta R_f^{k-\sigma}\right| \le\ & C_v\int_{t_{k-1}}^{t_k}|v'(t)|dt\int_{t_{k-1}}^{t_k}|\nabla v'(t)|dt+C_v\left(\int_{t_{k-1}}^{t_k}|v'(t)|dt\right)^2+C_v\left(\int_{t_{k-1}}^{t_k}|v'(t)|dt\right)^2\\
&+C_v\left(\int_{t_{k-1}}^{t_k}|\nabla v'(t)|dt\right)^2+C_v\int_{t_{k-1}}^{t_k}|v'(t)|dt\int_{t_{k-1}}^{t_k}|\Delta v'(t)|dt+C_v\int_{t_{k-1}}^{t_k}|v'(t)|dt\int_{t_{k-1}}^{t_k}|\nabla v'(t)|dt\\
\le\ & \begin{cases} C_v(\tau_1^2+\frac{\tau_1^{2\alpha}}{\alpha^2}), & k=1,\\ C_v(\tau_k^2+\tau_k^2t_{k-1}^{2(\alpha-1)}), & 2\le k\le n.\end{cases}
\end{aligned}
$$

It can be further obtained that

$$
\begin{aligned}
&\sum_{k=1}^n P_{n-k}^{(n)}\left|R_f^{k-\sigma}\right|+\sum_{k=1}^n P_{n-k}^{(n)}\left|\nabla R_f^{k-\sigma}\right|+\sum_{k=1}^n P_{n-k}^{(n)}\left|\Delta R_f^{k-\sigma}\right|\\
\le\ & P_{n-1}^{(n)}\left|R_v^{1-\sigma}\right|+\max_{2\le k\le n}\left|R_f^{k-\sigma}\right|\sum_{k=2}^n P_{n-k}^{(n)}+P_{n-1}^{(n)}\left|\nabla R_v^{1-\sigma}\right|+\max_{2\le k\le n}\left|\nabla R_f^{k-\sigma}\right|\sum_{k=2}^n P_{n-k}^{(n)}\\
&+P_{n-1}^{(n)}\left|\Delta R_v^{1-\sigma}\right|+\max_{2\le k\le n}\left|\Delta R_f^{k-\sigma}\right|\sum_{k=2}^n P_{n-k}^{(n)}\\
\le\ & C_v\left[\tau_1^{3\alpha}+\max_{2\le k\le n}t_n^\alpha\left(\tau_k^2+\tau_k^2t_{k-1}^{2\alpha-2}\right)\right],
\end{aligned}
$$

where C_v in different places represents different constant. The proof is completed. □

Acknowledgements The numerical simulations in this work have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

Funding Jiwei Zhang is partially supported by NSFC under grants Nos. 11771035 and 12171376, 2020-JCJQ-ZD-029. Yanping Chen is partially supported by the State Key Program of National Natural Science Foundation of China (No. 11931003) and National Natural Science Foundation of China (No. 41974133). Yanmin Zhao is partially supported by NSFC (No. 11971416) and the Scientific Research Innovation Team of Xuchang University (No. 2022CXTD002).

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Agarwal, P., Berezansky, L., Braverman, E., Domoshnitsky, A.: Nonoscillation Theory of Functional Differential Equations with Applications. Springer, New York (2012)
2. Yuste, S., Acedo, L., Lindenberg, K.: Reaction front in an $A + B \rightarrow C$ reaction-subdiffusion process. Phys. Rev. E. **69**, 036126 (2004)
3. Bouchaud, J., Georges, A.: Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications. Phys. Rep. **195**, 127–293 (1990)
4. Jin, B., Li, B., Zhou, Z.: Numerical analysis of nonlinear subdiffusion equations. SIAM J. Numer. Anal. **56**(1), 1–23 (2018)
5. Li, D., Sun, W., Wu, C.: A novel numerical approach to time-fractional parabolic equations with nonsmooth solutions. Numer. Math. Theor. Meth. Appl. **14**(2), 355–376 (2021)
6. Jin, B., Li, B., Zhou, Z.: Correction of high-order BDF convolution quadrature for fractional evolution equations. SIAM J. Sci. Comput. **39**(6), A3129–A3152 (2017)
7. Kopteva, N.: Error analysis of the l1 method on graded and uniform meshes for a fractional-derivative problem in two and three dimensions. Math. Comp. **88**, 2135–2155 (2019)
8. Liao, H., Li, D., Zhang, J.: Sharp error estimate of the nonuniform l1 formula for linear reaction-subdiffusion equations. SIAM J. Numer. Anal. **56**, 1112–1133 (2018)
9. Liao, H., Mclean, W., Zhang, J.: A second-order scheme with nonuniform time steps for a linear reaction-subdiffusion problem. Commu. Comput. Phys. **30**(2), 567–601 (2021)
10. Liao, H., McLean, W., Zhang, J.: A discrete Grönwall inequality with applications to numerical schemes for subdiffusion problems. SIAM J. Numer. Anal. **57**, 218–237 (2019)
11. Brunner, H.: The numerical solution of weakly singular Volterra integral equations by collocation on graded meshes. Math. Comput. **45**, 417–437 (1985)
12. McLean, W., Mustapha, K.: A second-order accurate numerical method for a fractional wave equation. Numer. Math. **105**, 481–510 (2007)
13. Liao, H., Yan, Y., Zhang, J.: Unconditional convergence of a fast two-level linearized algorithm for semilinear subdiffusion equations. J. Sci. Comput. **80**, 1–25 (2019)
14. Li, D., Wu, C., Zhang, Z.: Linearized Galerkin FEMs for nonlinear time fractional parabolic problems with non-smoooth solutions in time direction. J. Sci. Comput. **80**, 403–419 (2019)
15. Li, D., Wang, J.: Unconditionally optimal error analysis of Crank-Nicolson Galerkin FEMs for a strongly nonlinear parabolic system. J. Sci. Comput. **72**, 892–915 (2017)
16. Li, D., Zhang, J., Zhang, Z.: Unconditionally optimal error estimates of a linearized Galerkin method for nonlinear time fractional reaction-subdiffusion equations. J. Sci. Comput. **76**, 848–866 (2018)
17. Ji, B., Liao, H., Gong, Y.: Adaptive second-order Crank-Nicolson time-stepping schemes for time-fractional molecular beam epitaxial growth models. SIAM J. Sci. Comput. **42**(3), B738–B760 (2020)
18. Liao, H., Tang, T., Zhou, T.: An energy stable and maximum bound preserving scheme with variable time steps for time fractional Allen-Cahn equation. SIAM J. Sci. Comput. **43**(5), A3503–A3526 (2021)
19. Liao, H., Tang, T., Zhou, T.: A second-order and nonuniform time-stepping maximum-principle preserving scheme for time-fractional Allen-Cahn equations. J. Comput. Phys. **414**, 109473 (2020)
20. Jiang, S., Zhang, J., Zhang, Q., Zhang, Z.: Fast evaluation of the Caputo fractional derivative and its applications to fractional diffusion equations. Commun. Comput. Phys. **21**, 650–678 (2017)
21. Yan, Y., Sun, Z., Zhang, J.: Fast evaluation of the Caputo fractional derivative and its applications to fractional diffusion equations a second-order scheme. Commun. Comput. Phys. **22**, 1028–1048 (2017)
22. Baffet, D., Hesthaven, J.: A kernel compression scheme for fractional differential equations. SIAM J. Numer. Anal. **55**, 496–520 (2017)
23. Zhu, H., Xu, C.: A fast high order method for the time-fractional diffusion equation. SIAM J. Numer. Anal. **57**, 2829–2849 (2019)
24. Guo, L., Zeng, F., Turner, I., Burrage, K., Karniadakis, G.: Effcient multistep methods for tempered fractional calculus: algorithms and simulations. SIAM J. Sci. Comput. **41**, A2510–A2535 (2019)

25. Banjai, L., Lopez-Fernandez, M.: Effcient high order algorithms for fractional integrals and fractional differential equations. Numer. Math. **141**, 289–317 (2019)
26. Sun, J., Nie, D., Deng, W.: Fast algorithms for convolution quadrature of Riemann-Liouville fractional derivative. Appl. Numer. Math. **145**, 384–410 (2019)
27. Mustapha, K., Abdallah, B., Furati, K.: A discontinuous Petrov-Galerkin method for time-fractional diffusion equations. SIAM J. Numer. Anal. **52**, 2512–2529 (2014)
28. Lv, C., Xu, C.: Error Analysis of a high order method for time-fractional diffusion equations. SIAM J. Sci. Comput. **38**(5), 2699–2724 (2016)
29. Alikhanov, A.: A new difference scheme for the time fractional diffusion equation. J. Comput. Phys. **280**, 424–438 (2015)
30. Cao, J., Xu, C., Wang, Z.: A high order finite difference/spectral approximations to the time fractional diffusion equations. Adv. Mater. Res. **875**, 781–785 (2014)
31. Li, B., Gao, H., Sun, W.: Unconditionally optimal error estimate of a Crank-Nicolson Galerkin method for nonlinear thermistor equations. SIAM J. Numer. Anal. **52**, 933–954 (2014)
32. Li, D., Wang, J., Zhang, J.: Unconditionally convergent L1-Galerkin FEMs for nonlinear time-fractional Schrödinger equations. SIAM J. Sci. Comput. **39**(6), A3067–A3088 (2017)
33. Ren, J., Liao, H., Zhang, Z.: Superconvergence error estimate of a finite element method on nonuniform Time Meshes for reaction-subdiffusion equations. J. Sci. Comput. **84**(2), 38 (2020)
34. Ren, J., Liao, H., Zhang, J., Zhang, Z.: Sharp H^1-norm error estimates of two time-stepping schemes for reaction-subdiffusion problems. J. Comput. Appl. Math. **389**, 113352 (2021)
35. Li, X., Liao, H., Zhang, L.: A second-order fast compact scheme with unequal time-steps for subdiffusion problems. Numer. Algo. **86**, 1011–1039 (2021)
36. Thomee, V.: Glalerkin Finite Element Methods for Parabolic Problems. Springer, Berlin (1997)
37. Zhou, B., Chen, X., Li, D.: Nonuniform Alikhanov linearized Galerkin finite element methods for nonlinear time-fractional parabolic equations. J. Sci. Comput. **85**(2), 39 (2020)