A DISCRETE GRÖNWALL INEQUALITY WITH APPLICATIONS TO NUMERICAL SCHEMES FOR SUBDIFFUSION PROBLEMS[*]

HONG-LIN LIAO[†], WILLIAM MCLEAN[‡], AND JIWEI ZHANG[§]

Abstract. We consider a class of numerical approximations to the Caputo fractional derivative. Our assumptions permit the use of nonuniform time steps, such as is appropriate for accurately resolving the behavior of a solution whose temporal derivatives are singular at $t = 0$. The main result is a type of fractional Grönwall inequality and we illustrate its use by outlining some stability and convergence estimates of schemes for fractional reaction-subdiffusion problems. This approach extends earlier work that used the familiar L1 approximation to the Caputo fractional derivative, and will facilitate the analysis of higher order and linearized fast schemes.

Key words. fractional subdiffusion equations, nonuniform time mesh, discrete Caputo derivative, discrete Grönwall inequality

AMS subject classifications. 65M06, 35B65

DOI. 10.1137/16M1175742

1. Introduction. This paper builds on earlier results [15] for the nonuniform L1 method applied to the time discretization of a fractional reaction-subdiffusion problem [22] in a spatial domain Ω,

$$
\begin{aligned}
\mathcal{D}_t^\alpha u + \mathcal{L}u &= f(x,t,u) &&\text{for } x \in \Omega \text{ and } 0 < t \le T, \\
(1.1) \qquad u &= u_0(x) &&\text{for } x \in \Omega \text{ when } t = 0, \\
u &= 0 &&\text{for } x \in \partial\Omega \text{ and } 0 < t < T.
\end{aligned}
$$

Here, $\mathcal{D}_t^\alpha = {}_0^C\mathcal{D}_t^\alpha$ denotes the Caputo fractional derivative of order α with respect to time t with $0 < \alpha < 1$, and $\mathcal{L}$ is a linear, second-order, strongly elliptic partial differential operator in the spatial variable(s) x. We establish a discrete Grönwall inequality intended for the error analysis of higher-order time discretizations [16] and linearized fast algorithms [17] for solving (1.1) that employ nonuniform step sizes.

In any numerical methods for solving the reaction-subdiffusion problem (1.1), a key consideration is that the solution $u(x,t)$ is typically less regular than would be the case for a classical parabolic PDE (which arises as the limiting case when $\alpha \to 1$). For example, in the simplest case $f(x,t,u) \equiv 0$ when (1.1) is linear and homogeneous, let $\varphi_\mathcal{L}$ be a Dirichlet eigenfunction of $\mathcal{L}$ on Ω, with eigenvalue $\lambda_\mathcal{L} > 0$, so that

[†]Department of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, People's Republic of China (liaohl@csrc.ac.cn).

[‡]School of Mathematics and Statistics, University of New South Wales, Sydney 2052, Australia (w.mclean@unsw.edu.au).

[§]School of Mathematics and Statistics, and Hubei Key Laboratory of Computational Science, Wuhan University, Wuhan 430072, People's Republic of China (jiweizhang@whu.edu.cn).

$\mathcal{L}\varphi_{\mathcal{L}} = \lambda_{\mathcal{L}}\varphi_{\mathcal{L}}$. Let E_α denote the Mittag–Leffler function,

$$(1.2) \qquad E_\alpha(z) := \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(1+k\alpha)},$$

and choose as the initial data $u_0(x) = \varphi_{\mathcal{L}}(x)$. Term-by-term differentiation shows that the solution is $u(x,t) = E_\alpha(-\lambda_{\mathcal{L}}t^\alpha)\varphi_{\mathcal{L}}(x)$, and so $\partial u/\partial t = O(t^{\alpha-1})$ as $t \to 0$, whereas the solution of the classical parabolic equation, $u(x,t) = e^{-\lambda_{\mathcal{L}}t}\varphi_{\mathcal{L}}(x)$, is a smooth function of t. Sakamoto and Yamamoto [23] study the (lack of) regularity of u for more general initial data u_0 and a linear source term $f = f(x,t)$. In fact, u can only be a smooth function of t if the initial data and source term satisfy some restrictive compatibility conditions [24].

The nature of polynomial interpolation means that the convergence rate of the L1 or similar approximations to $\mathcal{D}_t^\alpha u$ is limited by the smoothness of the solution u. In the presence of a fixed singularity at $t = 0$ of the type described above, an established technique to restore an optimal convergence rate is to employ a graded mesh

$$(1.3) \qquad t_n := (n/N)^\gamma T \quad \text{for } 0 \le n \le N,$$

where the parameter $\gamma \ge 1$ must be adapted to the strength of the singularity. Choosing $\gamma = 1$ results in a uniform mesh, and the larger the value of γ the more strongly the grid points are concentrated near $t = 0$. For example, such meshes have long been used in the numerical solution of Fredholm [8] and Volterra [4] integral equations, and their use for time-fractional PDEs [20] is now well established.

Early papers on L1 schemes [18, 26] assumed a uniform step size τ, and showed that if u is smooth, then the time discretization error is $O(\tau^{2-\alpha})$. Recently, Jin, Lazarov, and Zhou [10] presented a new analysis, based on generating functions, that permitted nonsmooth initial data u_0. They showed that if $f \equiv 0$ and $u_0 \in L_2(\Omega)$, then the error in the norm of $L_2(\Omega)$ due to the time discretization is $O(\tau t_n^{-1})$. Thus, for t_n bounded away from zero, the method achieves first-order accuracy in time. Yan, Khan, and Ford [28] proposed a modified L1 scheme and obtained error estimates for smooth and nonsmooth initial data. It was shown that the modified L1 scheme on a uniform mesh has a convergence rate of $O(\tau^{2-\alpha})$. Alikhanov [2] introduced the L2-1$_\sigma$ formula, a modification of the L1 method that uses piecewise-quadratic instead of piecewise-linear interpolation, and approximates $\mathcal{D}_t^\alpha u$ at an offset grid point $t_{j+\sigma} = (j + \sigma)\tau$. He showed that if u is sufficiently smooth, then the time discretization error is $O(\tau^2)$ for the special choice $\sigma = 1 - \alpha/2$.

Although nonuniform meshes are flexible and reasonably convenient for practical implementation, they can significantly complicate the numerical analysis of schemes, both with respect to stability and consistency. Stynes, O'Riordan, and Gracia [25] considered the L1 method on a graded mesh of the form (1.3) applied to (1.1) for the case $\mathcal{L}u = -u_{xx}$ and a linear reaction term $f(x,t,u) = -c(x)u + g(x,t)$. They showed that, given the typical singular behavior of u, the maximum error in the fully discrete solution is of order $N^{-\min\{2-\alpha,\gamma\alpha\}}$. (Here we ignore the additional error due to the spatial discretization.) Thus, for a uniform mesh the error is $O(N^{-\alpha})$, but if $\gamma = (2-\alpha)/\alpha$, then the error is $O(N^{\alpha-2})$. Their stability analysis requires $c(x) \ge 0$, which prevents extending the approach to deal with a reaction term that is nonlinear but uniformly Lipschitz in u. This limitation was overcome recently in the precursor [15] to the present work by exploiting a novel discrete fractional Grönwall inequality for the L1 method.

Nonetheless, practical applications of the discrete Grönwall inequality in its basic form [15] are still limited because it does not apply to other numerical approximation schemes for the Caputo derivative and excludes certain adaptive time meshes required to resolve complex behaviors (physical oscillations, blowup, and so on) in nonlinear fractional differential equations. Also, the proof relies on specific properties of the L1 kernels $a_{n-k}^{(n)}$ and their complementary discrete kernels $P_{n-k}^{(n)}$, with a key step [15, Lemma 2.1] employing rough estimates of the truncation error that, to a large extent, rely on the simple form of the $a_{n-k}^{(n)}$. In summary, the main novel contributions of the present work are threefold:

(i) to generalize the discrete Grönwall inequality, permitting its use with a variety of discretizations of the Caputo derivative, not just the L1 scheme;

(ii) to provide a concise proof based on two simple assumptions on the discrete kernels, independent of their precise form;

(iii) to permit a more general class of nonuniform meshes or adaptive time grids, not just the graded meshes for resolving the initial singularity.

In more detail, section 2 defines a discrete fractional derivative (2.2) having the form of the classical L1 approximation but with general discrete kernels. We formulate three assumptions required for our theory. The first two impose a monotonicity property (A1) and a lower bound (A2) on the discrete kernels, and the third (A3) places a mild restriction on the local step size ratio. We give some examples of schemes satisfying these assumptions, and define a family of complementary discrete kernels, generalizing those introduced in the earlier paper [12, 15]. Lemma 2.3 establishes a key estimate involving the discrete kernels and the Mittag–Leffler function (1.2). In section 3, we prove our main result, a discrete fractional Grönwall inequality stated as Theorem 3.1, and provide, in Remark 6, a strategy to treat cases where the monotonicity assumption breaks down. Section 4 illustrates the use of the Grönwall inequality in conjunction with an abstract Galerkin method for the spatial discretization. Finally, a short appendix proves two technical inequalities needed for the stability analysis of section 4.

The generalized results proved below will allow us to show, in two companion papers [16, 17], that Alikhanov's L2-1_σ formula can achieve second-order accuracy on certain nonuniform time grids and that a linearized fast algorithm is unconditionally convergent for nonlinear subdiffusion equations.

2. Discrete fractional derivative. Recall that the Riemann–Liouville fractional integral operator of order $\beta > 0$ is defined by [21, 22]

$$(\mathcal{I}^\beta v)(t) := \int_0^t \omega_\beta(t-s)v(s)\,\mathrm{d}s \quad \text{for } t > 0, \quad \text{where } \omega_\beta(t) := \frac{t^{\beta-1}}{\Gamma(\beta)},$$

and, in turn, the Caputo fractional derivative is defined by

$$(2.1) \qquad (\mathcal{D}_t^\alpha v)(t) := (\mathcal{I}^{1-\alpha} v')(t) = \int_0^t \omega_{1-\alpha}(t-s)v'(s)\,\mathrm{d}s \quad \text{for } t > 0.$$

For (possibly nonuniform) time levels $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$, we denote the nth step size by $\tau_n := t_n - t_{n-1}$, fix an offset parameter $\theta \in [0,1)$, and define

$$t_{n-\theta} := \theta t_{n-1} + (1-\theta)t_n \quad \text{and} \quad v^{n-\theta} := \theta v^{n-1} + (1-\theta)v^n,$$

where v^k may be any sequence. Letting $v^k \approx v(t_k)$ and $\nabla_\tau v^k := v^k - v^{k-1}$, we consider a discrete Caputo derivative (not necessarily a direct approximation of (2.1);

see Remark 5) given by a convolution-like sum, as follows:

$$(2.2) \qquad (\mathcal{D}_\tau^\alpha v)^{n-\theta} := \sum_{k=1}^{n} A_{n-k}^{(n)} \nabla_\tau v^k \quad \text{for } 1 \le n \le N.$$

Here, the corresponding discrete convolution kernels are written as $A_{n-k}^{(n)}$ instead of A_{nk} to reflect the convolution structure of the fractional derivative. Our theory requires the following three assumptions:

A1. The discrete kernels are positive and monotone, that is,

$$A_0^{(n)} \ge A_1^{(n)} \ge A_2^{(n)} \ge \cdots \ge A_{n-1}^{(n)} > 0 \quad \text{for } 1 \le n \le N.$$

A2. There is a constant $\pi_A > 0$ such that the discrete kernels satisfy the lower bound

$$A_{n-k}^{(n)} \ge \frac{1}{\pi_A \tau_k} \int_{t_{k-1}}^{t_k} \omega_{1-\alpha}(t_n - s)\, ds \quad \text{for } 1 \le k \le n \le N.$$

A3. There is a constant $\rho > 0$ such that the step size ratios $\rho_k := \tau_k/\tau_{k+1}$ satisfy

$$\rho_k \le \rho \quad \text{for } 1 \le k \le N-1.$$

The boundedness and monotonicity assumptions A1 and A2 on the discrete convolution kernels $A_{n-k}^{(n)}$ are valid for several frequently used discrete Caputo derivatives, at least if assumption A3 is satisfied for appropriate ρ. Included are the well-known L1 formula [15, 18, 21, 25, 26], the fast L1 formula [17], the Alikhanov approximation [2, 13, 16], and their applications for multiterm and distributed-order Caputo derivatives (see Remark 5). Here we list three examples on nonuniform grids. Note that the local mesh parameter ρ from A3 will always appear in our discrete fractional Grönwall inequality and our stability estimates.

Example 1 (nonuniform L1 formula). The widespread L1 formula [21, p. 140] uses $\theta = 0$ and $v'(s) \approx \nabla_\tau v^k / \tau_k$ (linear interpolation) to obtain

$$(\mathcal{D}_\tau^\alpha v)^n := \sum_{k=1}^{n} a_{n-k}^{(n)} \nabla_\tau v^k \quad \text{with} \quad a_{n-k}^{(n)} := \frac{1}{\tau_k} \int_{t_{k-1}}^{t_k} \omega_{1-\alpha}(t_n - s)\, ds.$$

This sum has the desired form (2.2) where (using the integral mean value theorem),

$$(2.3) \qquad A_{n-k}^{(n)} := a_{n-k}^{(n)} = \omega_{1-\alpha}(t_n - s_{nk}) \quad \text{for some } s_{nk} \in [t_{k-1}, t_k].$$

It follows that assumption A1 is satisfied, and A2 holds with $\pi_A = 1$.

Example 2 (fast L1 formula). In the two-level fast L1 approximation [17], the sum-of-exponentials technique is applied to approximate the weakly singular kernel $\omega_{1-\alpha}(t-s)$. That is, for a user-given absolute tolerance error $\epsilon \ll 1$ and a cutoff time $\Delta t > 0$, one determines a positive integer N_q, positive quadrature nodes θ^ℓ, and positive weights ϖ^ℓ ($1 \le \ell \le N_q$) such that

$$\left| \omega_{1-\alpha}(t_k - s) - \sum_{\ell=1}^{N_q} \varpi^\ell e^{-\theta^\ell(t_k - s)} \right| \le \epsilon \quad \forall t_k \in [s + \Delta t, T].$$

Then we use $\theta = 0$ and $v'(s) \approx \nabla_\tau v^k / \tau_k$ (linear interpolation) to obtain

$$(D_f^\alpha u)^n := a_0^{(n)} \nabla_\tau u^n + \sum_{\ell=1}^{N_q} \varpi^\ell e^{-\theta^\ell \tau_n} H^\ell(t_{n-1}), \quad n \geq 1,$$

where $H^\ell(t_k)$ satisfies $H^\ell(t_0) = 0$ and the recurrence relationship

$$H^\ell(t_k) = e^{-\theta^\ell \tau_k} H^\ell(t_{k-1}) + \frac{1}{\tau_k} \int_{t_{k-1}}^{t_k} e^{-\theta^\ell(t_k - s)} \nabla_\tau u^k \, ds, \quad k \geq 1, \ 1 \leq \ell \leq N_q.$$

This approximation also has the form (2.2) with $\theta = 0$,

$$A_0^{(n)} := a_0^{(n)} \quad \text{and} \quad A_{n-k}^{(n)} := \frac{1}{\tau_k} \int_{t_{k-1}}^{t_k} \sum_{\ell=1}^{N_q} \varpi^\ell e^{-\theta^\ell(t_n - s)} \, ds \quad \text{for } 1 \leq k \leq n-1.$$

If the tolerance error ϵ is small enough such that $\epsilon \leq \min\left\{ \frac{1}{3}\omega_{1-\alpha}(T), \alpha\, \omega_{2-\alpha}(1) \right\}$, then [17, Lemma 2.5] ensures that A1–A2 hold true with $\pi_A = 3/2$.

Example 3 (nonuniform Alikhanov formula). Let $\Pi_{1,k}v$ be the linear interpolant of a function v with respect to the nodes t_{k-1} and t_k, and let $\Pi_{2,k}v$ denote the quadratic interpolant with respect to t_{k-1}, t_k and t_{k+1}. Taking a special choice $\theta = \alpha/2$, and applying the linear and quadratic polynomial interpolations, we have the nonuniform Alikhanov formula [13, 16]

$$\begin{aligned}
(\mathcal{D}_\tau^\alpha v)^{n-\theta} := &\sum_{k=1}^{n-1} \int_{t_{k-1}}^{t_k} \omega_{1-\alpha}(t_{n-\theta} - s) \left(\Pi_{2,k}v\right)'(s) \, ds \\
&+ \int_{t_{n-1}}^{t_{n-\theta}} \omega_{1-\alpha}(t_{n-\theta} - s) \left(\Pi_{1,n}v\right)'(s) \, ds \quad \text{for } n \geq 1.
\end{aligned}$$

This formula can be written as the form (2.2) with $A_0^{(1)} := \hat{a}_0^{(1)}$ for $n = 1$ and, for $n \geq 2$,

$$A_{n-k}^{(n)} := \begin{cases} \hat{a}_0^{(n)} + \rho_{n-1}\hat{b}_1^{(n)} & \text{for } k = n, \\ \hat{a}_{n-k}^{(n)} + \rho_{k-1}\hat{b}_{n-k+1}^{(n)} - \hat{b}_{n-k}^{(n)} & \text{for } 2 \leq k \leq n-1, \\ \hat{a}_{n-1}^{(n)} - \hat{b}_{n-1}^{(n)} & \text{for } k = 1, \end{cases}$$

where the discrete coefficients $\hat{a}_{n-k}^{(n)}$ and $\hat{b}_{n-k}^{(n)}$ are defined by

$$\hat{a}_0^{(n)} := \frac{1}{\tau_n} \int_{t_{n-1}}^{t_{n-\theta}} \omega_{1-\alpha}(t_{n-\theta} - s) \, ds,$$

$$\hat{a}_{n-k}^{(n)} := \frac{1}{\tau_k} \int_{t_{k-1}}^{t_k} \omega_{1-\alpha}(t_{n-\theta} - s) \, ds \quad \text{for } 1 \leq k \leq n-1,$$

$$\hat{b}_{n-k}^{(n)} := \frac{2}{\tau_k(\tau_k + \tau_{k+1})} \int_{t_{k-1}}^{t_k} (s - t_{k-\frac{1}{2}})\omega_{1-\alpha}(t_{n-\theta} - s) \, ds \quad \text{for } 1 \leq k \leq n-1.$$

The theoretical properties in [16, Theorem 2.2] assure A1–A2 with $\pi_A = 11/4$ provided the local mesh assumption A3 holds with the maximum step size ratio $\rho = 7/4$.

We now continue to introduce an important tool: the complementary discrete convolution kernels. The semigroup property of the fractional integral, $\mathcal{I}^\alpha \mathcal{I}^\beta = \mathcal{I}^{\alpha+\beta}$, holds because the integral kernels satisfy $\omega_\alpha * \omega_\beta = \omega_{\alpha+\beta}$. It follows that

$$(2.4) \qquad \int_s^t \omega_\alpha(t-\mu)\omega_{1-\alpha}(\mu-s)\,\mathrm{d}\mu = \omega_1(t-s) = 1 \quad \text{for all } 0 < s < t < \infty,$$

and it motivates us to seek a family of complementary discrete convolution kernels $P_{n-j}^{(n)}$ having the identical property

$$(2.5) \qquad \sum_{j=m}^n P_{n-j}^{(n)} A_{j-m}^{(j)} \equiv 1 \quad \text{for } 1 \le m \le n \le N.$$

In fact, by taking $m = k$ and $m = k+1$,

$$P_{n-k}^{(n)} A_0^{(k)} + \sum_{j=k+1}^n P_{n-j}^{(n)} A_{j-k}^{(j)} = 1 = \sum_{j=k+1}^n P_{n-j}^{(n)} A_{j-(k+1)}^{(j)}, \quad 1 \le k \le n-1,$$

we see that

$$P_{n-k}^{(n)} = \frac{1}{A_0^{(k)}} \sum_{j=k+1}^n P_{n-j}^{(n)} \left(A_{j-k-1}^{(j)} - A_{j-k}^{(j)} \right), \quad 1 \le k \le n-1,$$

and the complementary discrete kernels may be defined via the recursion [15]

$$(2.6) \quad P_0^{(n)} := \frac{1}{A_0^{(n)}}, \quad P_j^{(n)} := \frac{1}{A_0^{(n-j)}} \sum_{k=0}^{j-1} \left(A_{j-k-1}^{(n-k)} - A_{j-k}^{(n-k)} \right) P_k^{(n)} \quad \text{for } 1 \le j \le n-1.$$

Example 4 (pictures of $A_j^{(n)}$ and $P_j^{(n)}$ of L1 formula). Consider the widespread L1 approximation in Example 1. Figure 1 plots the L1 discrete kernels $A_j^{(n)}$ and the complementary discrete kernels $P_j^{(n)}$ when $T = 1$ and $n = 30$ for three graded meshes of the form (1.3).

As a consequence of the identity (2.4), we find that

$$\int_0^t \omega_\alpha(t-s)(\mathcal{D}_t^\alpha v)(s)\,\mathrm{d}s = \int_0^t v'(s)\,\mathrm{d}s,$$

which provides the inspiration for the second part of the next lemma.

LEMMA 2.1. *Let the assumptions* A1 *and* A2 *hold.*

1. *The discrete kernels $P_j^{(n)}$ in* (2.6) *having the property* (2.5) *satisfy*

$$0 \le P_{n-j}^{(n)} \le \pi_A \Gamma(2-\alpha)\tau_n^\alpha \quad \text{for } 1 \le j \le n \le N$$

    and

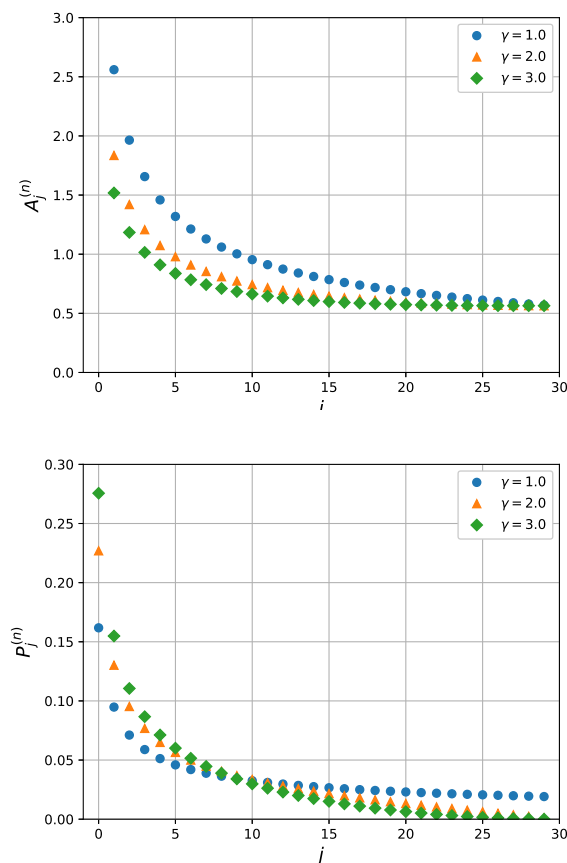$$(2.7) \qquad \sum_{j=1}^n P_{n-j}^{(n)} \omega_{1-\alpha}(t_j) \le \pi_A \quad \text{for } 1 \le n \le N.$$

FIG. 1. *Top: the* L1 *discrete kernel* (2.3) *for three different meshes of the form* (1.3) *in the case* $T = 1$ *and* $n = 30$. *Bottom: the complementary discrete kernels* $P_j^{(n)}$.

2. *If* $v : [0, T] \to \mathbb{R}$ *is any continuous, piecewise-C^1 function such that v' is nonnegative and monotone decreasing, then*

$$(2.8) \qquad \sum_{j=1}^{n} P_{n-j}^{(n)} (\mathcal{D}_t^\alpha v)(t_j) \leq \pi_A \int_0^{t_n} v'(s) \, \mathrm{d}s \quad \text{for } 1 \leq n \leq N.$$

Proof. It follows at once from the monotonicity assumption A1 that $A_0^{(n)} > 0$ and $A_{j-k-1}^{(n-k)} - A_{j-k}^{(n-k)} \geq 0$ for $0 \leq k \leq j-1$. The lower bound $P_j^{(n)} \geq 0$ is then clear from the recursion (2.6). Since all the discrete kernels are nonnegative, we have

$$P_{n-k}^{(n)} A_0^{(k)} \leq \sum_{j=k}^{n} P_{n-j}^{(n)} A_{j-k}^{(j)} = 1$$

and taking $n = k$ in the assumption A2 gives

$$A_0^{(k)} \geq \frac{1}{\pi_A \tau_k} \int_{t_{k-1}}^{t_k} \omega_{1-\alpha}(t_k - s) \, \mathrm{d}s = \frac{\omega_{2-\alpha}(\tau_k)}{\pi_A \tau_k} = \frac{1}{\Gamma(2-\alpha)\pi_A \tau_k^\alpha},$$

so the complementary discrete convolution kernels $P_{n-k}^{(n)}$ are well-defined and satisfy the upper bound $P_{n-k}^{(n)} \leq 1/A_0^{(k)} \leq \Gamma(2-\alpha)\pi_A \tau_k^\alpha$. Furthermore, the assumption A2 and the identity (2.5) imply that $\omega_{1-\alpha}(t_j) \leq \pi_A A_{j-1}^{(j)}$ and

$$\sum_{j=1}^n P_{n-j}^{(n)} \omega_{1-\alpha}(t_j) \leq \pi_A \sum_{j=1}^n P_{n-j}^{(n)} A_{j-1}^{(j)} = \pi_A \quad \text{for } n \geq 1,$$

which completes the proof of part 1.

Recall Chebyshev's sorting inequality [9, p. 168, item 236]: if f is monotone increasing and g is monotone decreasing on the interval $[a,b]$, and if both functions are integrable, then

$$(b-a)\int_a^b f(s)g(s)\,\mathrm{d}s \leq \int_a^b f(t)\,\mathrm{d}t \int_a^b g(s)\,\mathrm{d}s.$$

Taking $[a,b] = [t_{k-1}, t_k]$, $f(s) = \omega_{1-\alpha}(t_j - s)$, and $g(s) = v'(s) \geq 0$, and using A2, we see that

$$(\mathcal{D}_t^\alpha v)(t_j) = \sum_{k=1}^j \int_{t_{k-1}}^{t_k} \omega_{1-\alpha}(t_j - s)v'(s)\,\mathrm{d}s$$

$$\leq \sum_{k=1}^j \frac{1}{\tau_k}\int_{t_{k-1}}^{t_k} \omega_{1-\alpha}(t_j - t)\,\mathrm{d}t \int_{t_{k-1}}^{t_k} v'(s)\,\mathrm{d}s \leq \pi_A \sum_{k=1}^j A_{j-k}^{(j)} \int_{t_{k-1}}^{t_k} v'(s)\,\mathrm{d}s.$$

Thus, from the identical property (2.5) of the discrete kernels $P_{n-j}^{(n)}$, we conclude that

$$\sum_{j=1}^n P_{n-j}^{(n)}(\mathcal{D}_t^\alpha v)(t_j) \leq \sum_{j=1}^n P_{n-j}^{(n)} \pi_A \sum_{k=1}^j A_{j-k}^{(j)} \int_{t_{k-1}}^{t_k} v'(s)\,\mathrm{d}s$$

$$= \pi_A \sum_{k=1}^n \int_{t_{k-1}}^{t_k} v'(s)\,\mathrm{d}s \sum_{j=k}^n P_{n-j}^{(n)} A_{j-k}^{(j)} = \pi_A \sum_{k=1}^n \int_{t_{k-1}}^{t_k} v'(s)\,\mathrm{d}s,$$

and part 2 follows.   □

When A3 also holds, we have a variant of the second part of Lemma 2.1.

LEMMA 2.2. *Let the assumptions* A1–A3 *hold. If* $v : [0,T] \to \mathbb{R}$ *is any continuous, piecewise-C^1 function such that* v' *is nonnegative and monotone, then*

$$\sum_{j=1}^{n-1} P_{n-j}^{(n)}(\mathcal{D}_t^\alpha v)(t_j) \leq \max(1,\rho)\pi_A \int_0^{t_n} v'(s)\,\mathrm{d}s \quad \text{for } 1 \leq n \leq N.$$

Proof. If v' is nonnegative and monotone *decreasing*, then $\mathcal{D}_t^\alpha v(t_j) \geq 0$ and the results of Lemma 2.1 imply that

$$\sum_{j=1}^{n-1} P_{n-j}^{(n)}(\mathcal{D}_t^\alpha v)(t_j) \leq \sum_{j=1}^n P_{n-j}^{(n)}(\mathcal{D}_t^\alpha v)(t_j) \leq \pi_A \int_0^{t_n} v'(s)\,\mathrm{d}s\,.$$

Otherwise, if v' is monotonely *increasing*, then

$$
\begin{aligned}
\sum_{j=1}^{n-1} P_{n-j}^{(n)} (\mathcal{D}_t^\alpha v)(t_j) &= \sum_{j=1}^{n-1} P_{n-j}^{(n)} \sum_{k=1}^{j} \int_{t_{k-1}}^{t_k} \omega_{1-\alpha}(t_j - s) v'(s) \, \mathrm{d}s \\
&\leq \sum_{j=1}^{n-1} P_{n-j}^{(n)} \sum_{k=1}^{j} v'(t_k) \int_{t_{k-1}}^{t_k} \omega_{1-\alpha}(t_j - s) \, \mathrm{d}s \\
&\leq \pi_A \sum_{j=1}^{n-1} P_{n-j}^{(n)} \sum_{k=1}^{j} v'(t_k) \tau_k A_{j-k}^{(j)} \\
&= \pi_A \sum_{k=1}^{n-1} v'(t_k) \tau_k \sum_{j=k}^{n-1} P_{n-j}^{(n)} A_{j-k}^{(j)} \leq \pi_A \sum_{k=1}^{n-1} v'(t_k) \tau_k \\
&\leq \pi_A \sum_{k=1}^{n-1} v'(t_k) \rho_k \tau_{k+1} \leq \rho \pi_A \sum_{k=1}^{n-1} \int_{t_k}^{t_{k+1}} v'(s) \, \mathrm{d}s,
\end{aligned}
$$

and the desired estimate again holds. $\qquad\square$

We can use Lemma 2.2 to prove the following property of the Mittag–Leffler function (1.2).

LEMMA 2.3. *Let the assumptions* A1–A3 *hold. For any real* $\mu > 0$,

$$
\sum_{j=1}^{n-1} P_{n-j}^{(n)} E_\alpha(\mu t_j^\alpha) \leq \pi_A \max(1, \rho) \frac{E_\alpha(\mu t_n^\alpha) - 1}{\mu} \quad \text{for } 1 \leq n \leq N.
$$

Proof. The series definition (1.2) shows that

$$
E_\alpha(\mu t^\alpha) = 1 + \sum_{k=1}^{\infty} \frac{\mu^k t^{k\alpha}}{\Gamma(1 + k\alpha)} = 1 + \sum_{k=1}^{\infty} \mu^k v_k(t),
$$

where $v_k(t) = \omega_{1+k\alpha}(t)$ and we have $v_k'(t) = \omega_{k\alpha}(t) > 0$ for all $k \geq 1$. If $1 \leq k \leq 1/\alpha$, then $-1 \leq k\alpha - 1 \leq 0$ and $v_k''(t) = \omega_{k\alpha-1}(t) \leq 0$ for all $t > 0$. Otherwise, if $k > 1/\alpha$, then $k\alpha - 1 > 0$ and $v_k''(t) > 0$ for all $t > 0$. Thus, v_k' is always nonnegative and monotone, so we may apply Lemma 2.2, and deduce that

$$
\sum_{j=1}^{n-1} P_{n-j}^{(n)} (\mathcal{D}_t^\alpha v_k)(t_j) \leq \max(1, \rho) \pi_A \int_0^{t_n} v_k'(s) \, \mathrm{d}s = \max(1, \rho) \pi_A v_k(t_n) \quad \text{for } k \geq 1.
$$

Multiplying both sides of this inequality by μ^k, summing over the index k, and using the fact that

$$
\mathcal{D}_t^\alpha v_k(t) = \int_0^t \omega_{1-\alpha}(t - s) \omega_{k\alpha}(s) \, \mathrm{d}s = \omega_{1+(k-1)\alpha}(t) = v_{k-1}(t) \quad \text{for all } k \geq 1,
$$

we have

$$
\sum_{k=1}^{m} \mu^k \sum_{j=1}^{n-1} P_{n-j}^{(n)} v_{k-1}(t) \leq \max(1, \rho) \pi_A \sum_{k=1}^{m} \mu^k v_k(t_n).
$$

Because the series $\sum_{k=1}^{\infty} \mu^k v_k(t)$ is absolutely convergent and $\omega_1(t) = 1$, the desired inequality follows after interchanging the sums on the left-hand side and then sending $m \to \infty$. The proof is completed. $\qquad\square$

3. Discrete fractional Grönwall inequality. Our main result is stated in the next theorem. The proof is similar to that of [15, Lemma 2.2], but we include it here to incorporate the nonuniform mesh parameter ρ in A3, which does not appear in discrete Grönwall inequalities for classical parabolic equations.

THEOREM 3.1. *Let the assumptions* A1–A3 *hold, let* $0 \le \theta < 1$, *and let* $(g^n)_{n=1}^N$ *and* $(\lambda_l)_{l=0}^{N-1}$ *be given nonnegative sequences. Assume further that there exists a constant* Λ *(independent of the step sizes) such that* $\Lambda \ge \sum_{l=0}^{N-1} \lambda_l$, *and that the maximum step size satisfies*

$$\max_{1 \le n \le N} \tau_n \le \frac{1}{\sqrt[\alpha]{2\pi_A \Gamma(2-\alpha)\Lambda}} .$$

Then, for any nonnegative sequence $(v^k)_{k=0}^N$ *such that*

$$(3.1) \qquad \sum_{k=1}^n A_{n-k}^{(n)} \nabla_\tau (v^k)^2 \le \sum_{k=1}^n \lambda_{n-k} (v^{k-\theta})^2 + v^{n-\theta} g^n \quad for\ 1 \le n \le N,$$

it holds that

$$(3.2) \qquad v^n \le 2 E_\alpha \big(2\max(1,\rho)\pi_A \Lambda t_n^\alpha\big)\bigg(v^0 + \max_{1 \le k \le n} \sum_{j=1}^k P_{k-j}^{(k)} g^j\bigg) \quad for\ 1 \le n \le N.$$

Proof. We replace the index n with j in (3.1), then multiply by $P_{n-j}^{(n)}$ and sum over j to obtain

$$(3.3) \qquad \sum_{j=1}^n P_{n-j}^{(n)} \sum_{k=1}^j A_{j-k}^{(j)} \nabla_\tau (v^k)^2 \le \sum_{j=1}^n P_{n-j}^{(n)} \sum_{k=1}^j \lambda_{j-k} (v^{k-\theta})^2 + \sum_{j=1}^n P_{n-j}^{(n)} v^{j-\theta} g^j.$$

On the left-hand side, we exchange the order of summation and use the identity (2.5) to get

$$(3.4) \qquad \sum_{j=1}^n P_{n-j}^{(n)} \sum_{k=1}^j A_{j-k}^{(j)} \nabla_\tau (v^k)^2 = \sum_{k=1}^n \nabla_\tau (v^k)^2 \sum_{j=k}^n P_{n-j}^{(n)} A_{j-k}^{(j)}$$
$$= \sum_{k=1}^n \nabla_\tau (v^k)^2 = (v^n)^2 - (v^0)^2.$$

Thus, it follows from (3.3) that

$$(3.5) \qquad (v^n)^2 \le (v^0)^2 + \sum_{j=1}^n P_{n-j}^{(n)} \sum_{k=1}^j \lambda_{j-k} (v^{k-\theta})^2 + \sum_{j=1}^n P_{n-j}^{(n)} v^{j-\theta} g^j,$$

For brevity, let us write the claimed estimate (3.2) as $v^n \le F_n G_n$, where

$$F_n := 2 E_\alpha\big(2\max(1,\rho)\pi_A \Lambda t_n^\alpha\big) \quad and \quad G_n := v^0 + \max_{1 \le k \le n} \sum_{j=1}^k P_{k-j}^{(k)} g^j.$$

We will use complete induction, noting that the Mittag–Leffler function (1.2) satisfies $E_\alpha(0) = 1$ and $E_\alpha'(z) > 0$ for all real $z > 0$, so $F_n \ge F_{n-1} \ge 2$ for $n \ge 2$.

If $v^1 \leq v^0$, then $v^1 \leq G_1 \leq F_1 G_1$, as required. Otherwise, if $v^1 > v^0$, then $v^{1-\theta} \leq v^1$. One deduces from (3.5) that

$$\begin{aligned}
(v^1)^2 &\leq (v^0)^2 + P_0^{(1)} v^{1-\theta} g^1 + P_0^{(1)} \lambda_0 (v^{1-\theta})^2 \\
&\leq v^1 (v^0 + P_0^{(1)} g^1) + P_0^{(1)} \lambda_0 (v^1)^2 = v^1 G_1 + P_0^{(1)} \lambda_0 (v^1)^2.
\end{aligned}$$

Part 1 of Lemma 2.1 and the given restriction on the maximum time-step imply that

$$(3.6) \qquad P_0^{(1)} \lambda_0 \leq \pi_A \Gamma(2-\alpha) \tau_1^\alpha \Lambda \leq 1/2.$$

Thus, $(v^1)^2 \leq 2v^1 G_1$ and so $v^1 \leq 2G_1 \leq F_1 G_1$, which implies that the desired estimate holds for $n = 1$.

For the inductive step, let $2 \leq n \leq N$ and assume that

$$(3.7) \qquad v^k \leq F_k G_k \quad \text{for } 1 \leq k \leq n-1.$$

Choose some $k(n)$ such that $v^{k(n)} = \max_{0 \leq j \leq n-1} v^j$. If $v^n \leq v^{k(n)}$, then, since F_k and G_k are monotone increasing in k,

$$v^n \leq v^{k(n)} \leq F_{k(n)} G_{k(n)} \leq F_n G_n,$$

as required. Otherwise, if $v^n > v^{k(n)}$, then $v^{j-\theta} \leq \max(v^{j-1}, v^j) \leq v^n$ for $1 \leq j \leq n$. We deduce from (3.5) that

$$(3.8) \quad (v^n)^2 \leq v^n v^0 + v^n \sum_{j=1}^n P_{n-j}^{(n)} g^j + v^n \sum_{j=1}^{n-1} P_{n-j}^{(n)} \sum_{k=1}^j \lambda_{j-k} v^{k-\theta} + (v^n)^2 P_0^{(n)} \sum_{k=1}^n \lambda_{n-k}.$$

Using part 1 of Lemma 2.1,

$$(3.9) \qquad P_0^{(n)} \sum_{k=1}^n \lambda_{n-k} \leq \pi_A \Gamma(2-\alpha) \Lambda \tau_n^\alpha,$$

so the limitation on the maximum step size implies that

$$(3.10) \qquad (v^n)^2 \leq v^n \left(G_n + \sum_{j=1}^{n-1} P_{n-j}^{(n)} \sum_{k=1}^j \lambda_{j-k} v^{k-\theta} \right) + \frac{1}{2} (v^n)^2.$$

Thus, applying the induction hypothesis (3.7), we deduce from (3.10) that

$$\begin{aligned}
v^n &\leq 2G_n + 2 \sum_{j=1}^{n-1} P_{n-j}^{(n)} \sum_{k=1}^j \lambda_{j-k} \left[\theta v^{k-1} + (1-\theta) v^k \right] \\
&\leq 2G_n + 2 \sum_{j=1}^{n-1} P_{n-j}^{(n)} \sum_{k=1}^j \lambda_{j-k} \left[\theta F_{k-1} G_{k-1} + (1-\theta) F_k G_k \right] \\
&\leq 2G_n + 2 \sum_{j=1}^{n-1} P_{n-j}^{(n)} \sum_{k=1}^j \lambda_{j-k} F_k G_k \leq 2G_n + 2 \sum_{j=1}^{n-1} P_{n-j}^{(n)} F_j G_j \sum_{k=1}^j \lambda_{j-k} \\
&\leq 2G_n + 4\Lambda G_{n-1} \sum_{j=1}^{n-1} P_{n-j}^{(n)} E_\alpha \left(2 \max(1, \rho) \pi_A \Lambda t_j^\alpha \right).
\end{aligned}$$

Finally, by Lemma 2.3 with $\mu = 2\max(1, \rho)\pi_A\Lambda$,

$$v^n \leq 2G_n + 2\max(1, \rho)\pi_A\Lambda G_n \frac{E_\alpha\big(2\max(1, \rho)\pi_A\Lambda t_n^\alpha\big) - 1}{\max(1, \rho)\pi_A\Lambda} = F_n G_n,$$

which completes the inductive step and the proof. $\qquad\square$

Remark 1. One may use the inequality (2.7) in part 1 of Lemma 2.1 to bound the convolutional summation $\sum_{j=1}^{k} P_{k-j}^{(k)} g^j$, that is,

$$\sum_{j=1}^{k} P_{k-j}^{(k)} g^j \leq \sum_{j=1}^{k} P_{k-j}^{(k)} \omega_{1-\alpha}(t_j) \max_{1 \leq j \leq k} \frac{g^j}{\omega_{1-\alpha}(t_j)} \leq \pi_A \max_{1 \leq j \leq k} \frac{g^j}{\omega_{1-\alpha}(t_j)}.$$

So the discrete solution of (3.1) can also be bounded by

$$v^n \leq 2E_\alpha\big(2\max(1, \rho)\pi_A\Lambda t_n^\alpha\big) \left(v^0 + \pi_A\Gamma(1 - \alpha) \max_{1 \leq j \leq n} \{t_j^\alpha g^j\}\right) \quad \text{for } 1 \leq n \leq N.$$

On the other hand, if the given sequence $(\lambda_l)_{l=0}^{N-1}$ is nonpositive and the constant $\Lambda \leq 0$, a similar argument will show that the discrete inequality (3.2) holds in a simpler form, requiring only the assumptions A1–A2 but no restrictions on time steps,

$$(3.11) \quad v^n \leq v^0 + \max_{1 \leq k \leq n} \sum_{j=1}^{k} P_{k-j}^{(k)} g^j \leq v^0 + \pi_A\Gamma(1 - \alpha) \max_{1 \leq j \leq n} \{t_j^\alpha g^j\} \quad \text{for } 1 \leq n \leq N.$$

Remark 2. By including the nonnegative sequence $(\lambda_l)_{l=0}^{N-1}$ in (3.1), we are able to treat various numerical approaches to solving linear and nonlinear subdiffusion problems. Typically, the sequence takes only a few nonzero values. Recent examples include $\lambda_l = 0$ for $l \geq 1$ in the time-weighted method from section 4, and $\lambda_l = 0$ for $l \geq 2$ in the one-step linearized scheme [17] for a semilinear subdiffusion equation. Thus, the constant Λ is always not very large and the maximum time-step restriction $\max_{1 \leq n \leq N} \tau_n \leq 1/\sqrt[\alpha]{2\pi_A\Gamma(2 - \alpha)\Lambda}$ is also not stringent in practical applications.

Remark 3. The Mittag–Leffler function E_α also arises naturally in other discrete and continuous Grönwall inequalities for fractional diffusion and wave equations [1, Lemma 2], and for weakly singular Volterra equations [5, Theorems 1.3 and 1.6]. The presence of the nonuniform mesh parameter ρ in the argument of E_α indicates that sudden, drastic reductions of the time step should be avoided. Nevertheless, our discrete Grönwall inequality is valid for a general nonuniform mesh provided A1–A3 are satisfied.

We also have an alternative version of the above theorem.

THEOREM 3.2. *Theorem* 3.1 *remains valid if the condition* (3.1) *is replaced by*

$$(3.12) \qquad \sum_{k=1}^{n} A_{n-k}^{(n)} \nabla_\tau v^k \leq \sum_{k=1}^{n} \lambda_{n-k} v^{k-\theta} + g^n \quad \text{for } 1 \leq n \leq N.$$

Moreover, if the given sequence $(\lambda_l)_{l=0}^{N-1}$ *is nonpositive and the constant* $\Lambda \leq 0$,

$$(3.13) \qquad v^n \leq v^0 + \sum_{j=1}^{n} P_{n-j}^{(n)} g^j \leq v^0 + \pi_A\Gamma(1 - \alpha) \max_{1 \leq j \leq n} \{t_j^\alpha g^j\} \quad \text{for } 1 \leq n \leq N.$$

Proof. The structure of the proof is as before. However, instead of (3.3) and (3.4), we now have

$$\sum_{j=1}^{n} P_{n-j}^{(n)} \sum_{k=1}^{j} A_{j-k}^{(j)} \nabla_{\tau} v^k \leq \sum_{j=1}^{n} P_{n-j}^{(n)} \sum_{k=1}^{j} \lambda_{j-k} v^{k-\theta} + \sum_{j=1}^{n} P_{n-j}^{(n)} g^j$$

and

$$\sum_{j=1}^{n} P_{n-j}^{(n)} \sum_{k=1}^{j} A_{j-k}^{(j)} \nabla_{\tau} v^k = \sum_{k=1}^{n} \nabla_{\tau} v^k \sum_{j=k}^{n} P_{n-j}^{(n)} A_{j-k}^{(j)} = \sum_{k=1}^{n} \nabla_{\tau} v^k = v^n - v^0,$$

respectively, so that instead of (3.5) we obtain

$$v^n \leq v^0 + \sum_{j=1}^{n} P_{n-j}^{(n)} \sum_{k=1}^{j} \lambda_{j-k} v^{k-\theta} + \sum_{j=1}^{n} P_{n-j}^{(n)} g^j.$$

As before, if $v^1 \leq v^0$, then $v^1 \leq G_1$. For the alternative case $v^1 > v^0$, we again have $v^{1-\theta} \leq v^1$ which now yields

$$v^1 \leq v^0 + P_0^{(1)} g^1 + P_0^{(1)} \lambda_0 v^{1-\theta} = G_1 + P_0^{(1)} \lambda_0 v^{1-\theta} \leq G_1 + \tfrac{1}{2} v^1,$$

where the final step again relies on the step size assumption to ensure (3.6). Thus, once again, $v^1 \leq 2G_1$. In the inductive step, (3.8) is replaced by

$$v^n \leq v^0 + \sum_{j=1}^{n} P_{n-j}^{(n)} g^j + \sum_{j=1}^{n-1} P_{n-j}^{(n)} \sum_{k=1}^{j} \lambda_{j-k} v^{k-\theta} + v^n P_0^{(n)} \sum_{k=1}^{n} \lambda_{n-k},$$

and by again using (3.9) together with the limitation on the maximum step size, we see that

$$v^n \leq \left(G_n + \sum_{j=1}^{n-1} P_{n-j}^{(n)} \sum_{k=1}^{j} \lambda_{j-k} v^{k-\theta} \right) + \frac{v^n}{2},$$

which is equivalent to (3.10) so the remainder of the proof is unchanged. $\square$

Remark 4. The discrete fractional Grönwall inequalities in Theorems 3.1 and 3.2 are valid on very general nonuniform time meshes and differ substantially from the discrete fractional Grönwall inequality of Jin, Li, and Zhou [11, Theorem 2.8], which is built on the uniform mesh for both the L1 scheme and the convolution quadratures generated by backward difference formulas.

Remark 5 (multiterm and distributed-order Caputo derivatives). Note that our theory starts only from the discrete convolution form (2.2) and the three assumptions A1–A3, but not the continuous counterpart (2.1). Correspondingly, the complementary discrete kernels $P_{n-j}^{(n)}$ defined in (2.5) are also independent of (2.1). In other words, the fractional order α of Caputo's derivative $\mathcal{D}_t^{\alpha} v$ in Lemmas 2.1 and 2.2, and the fractional exponent α in the Mittag–Leffler function E_α in Lemma 2.3 and Theorems 3.1 and 3.2, are determined only by the integrand function $\omega_{1-\alpha}(t_n - s)$ of the lower bound in A2, but are independent of the continuous counterpart of (2.2).

To explain this point more clearly, suppose that the discrete convolution form (2.2) arises from some numerical formula for a multiterm Caputo derivative $\sum_{i=1}^{m} w_i \mathcal{D}_t^{\alpha_i} v$

with $0 < \alpha_i < 1$ and the weights $w_i > 0$; see [22]. Then all of the fractional exponents α_i or the maximum order $\max_{1 \le i \le m} \alpha_i$ can determine a single fractional exponent α for A2 and the Mittag–Leffler function E_α in Theorems 3.1 and 3.2. Hence, the presented results would also be useful for studying numerical approximations of multiterm Caputo derivatives and distributed-order Caputo derivatives, since the latter can be approximated by certain multiterm derivatives via a proper quadrature rule [14].

Remark 6 (Caputo BDF2-like formula and an open problem). There are other practically important formulas, such as the Caputo BDF2-like approach [7, 14, 19]. To start the time-stepping process, one computes the first-level solution by the L1 approach in Example 1, $(\mathcal{D}_\tau^\alpha v)^1 := a_0^{(1)} \nabla_\tau v^1$, or the Alikhanov formula in Example 3, $(\mathcal{D}_\tau^\alpha v)^1 := \hat{a}_0^{(1)} \nabla_\tau v^1$. For any time-level t_n with $n \ge 2$, taking $\theta = 0$ and applying the quadratic polynomial interpolation $\Pi_{2,k} v$, we have a Caputo backward differentiation formula [19]

$$(\mathcal{D}_\tau^\alpha v)^n := \sum_{k=1}^{n-1} \int_{t_{k-1}}^{t_k} \omega_{1-\alpha}(t_n - s) \left(\Pi_{2,k} v \right)'(s) \, \mathrm{d}s$$
$$+ \int_{t_{n-1}}^{t_n} \omega_{1-\alpha}(t_n - s) \left(\Pi_{2,n-1} v \right)'(s) \, \mathrm{d}s \quad \text{for } n \ge 2.$$

This formula tends to the classical BDF2 approximation in the limit as $\alpha \to 1$. One can obtain the compact form (2.2) with the discrete kernels $A_{n-k}^{(n)}$,

$$A_{n-k}^{(n)} := \begin{cases} a_0^{(n)} + \rho_{n-1} \big(b_1^{(n)} + b_0^{(n)} \big) & \text{for } k = n, \\ a_1^{(n)} + \rho_{n-2} b_2^{(n)} - \big(b_1^{(n)} + b_0^{(n)} \big) & \text{for } k = n-1, \\ a_{n-k}^{(n)} + \rho_{k-1} b_{n-k+1}^{(n)} - b_{n-k}^{(n)} & \text{for } 2 \le k \le n-2 \\ a_{n-1}^{(n)} - b_{n-1}^{(n)} & \text{for } k = 1, \end{cases}$$

where the coefficients $a_{n-k}^{(n)}$ are defined in Example 1, and the $b_{n-k}^{(n)}$ are defined by

$$b_0^{(n)} := \frac{2}{\tau_{n-1}(\tau_{n-1} + \tau_n)} \int_{t_{n-1}}^{t_n} (s - t_{n-\frac{1}{2}}) \omega_{1-\alpha}(t_n - s) \, \mathrm{d}s,$$
$$b_{n-k}^{(n)} := \frac{2}{\tau_k(\tau_k + \tau_{k+1})} \int_{t_{k-1}}^{t_k} (s - t_{k-\frac{1}{2}}) \omega_{1-\alpha}(t_n - s) \, \mathrm{d}s \quad \text{for } 1 \le k \le n-1.$$

Notice that if the fractional order $\alpha \to 1$, then $\omega_{3-\alpha}(t) \to t$, $\omega_{2-\alpha}(t) \to 1$, and $\omega_{1-\alpha}(t) \to 0$, uniformly for $t > 0$. Thus, we have $a_0^{(n)} = \omega_{2-\alpha}(\tau_n)/\tau_n \to 1/\tau_n$ and

$$b_0^{(n)} = \frac{2}{\tau_{n-1}(\tau_{n-1} + \tau_n)} \left[\omega_{3-\alpha}(\tau_n) - \frac{\tau_n}{2} \omega_{2-\alpha}(\tau_n) \right] \to \frac{\tau_n}{\tau_{n-1}(\tau_{n-1} + \tau_n)},$$

whereas $a_{n-k}^{(n)} \to 0$ and $b_{n-k}^{(n)} \to 0$ for $k \ge 1$. So, when the fractional order $\alpha \to 1$,

$$(\mathcal{D}_\tau^\alpha v)^n \to D_2 v^n := \left(\frac{1}{\tau_n} + \frac{1}{\tau_{n-1} + \tau_n} \right) \nabla_\tau v^n - \frac{\tau_n}{\tau_{n-1}(\tau_{n-1} + \tau_n)} \nabla_\tau v^{n-1}$$

which is the second-order BDF2 scheme for the classical diffusion equations. We see that the second kernel $A_1^{(n)}$ can be negative, at least, when α is close to 1 (whereas

the Caupto BDF2 scheme is shown in [14] to preserve the discrete maximum principle and nonnegativity property when α is close to 0).

The Caputo BDF2 formula may not meet our a priori assumptions A1–A2, which results in that our Grönwall inequality would be not applicable directly. It is not surprising because, for a classical parabolic equation, the standard discrete Grönwall inequality can also not be applied directly to the second-order BDF2 scheme. However, a weighted recombination technique works well; see the detailed analysis by Thomée [27, Theorem 1.7] for a uniform time mesh, and a similar technique for nonuniform meshes [3, 6]. For the Caputo BDF2 formula, Theorems 3.1 and 3.2 would be also useful for the stability and convergence analysis if it can be rearranged to meet the positive and monotone assumptions A1–A2. On the uniform mesh with $\tau_n = \tau$, Lv and Xu [19] developed a new technique of variable-weights recombination and achieved a new form of $(\mathcal{D}_\tau^\alpha v)^n$ with a new variable $\bar{v}^k := v^k - \eta v^{k-1}$ and $\bar{v}^0 := v^0$; in our notations,

$$(3.14) \qquad (\mathcal{D}_\tau^\alpha v)^n = \sum_{k=1}^{n} \bar{A}_{n-k}^{(n)} \nabla_\tau \bar{v}^k + v^0 \sum_{j=1}^{n} A_{n-j}^{(n)} \eta^j,$$

where the combination parameter $\eta := \frac{1}{2}(1 - A_1^{(n)}/A_0^{(n)})$. From the substitution formulas

$$v^k = \sum_{\ell=0}^{k} \eta^{k-\ell} \bar{v}^\ell \quad \text{and} \quad \nabla_\tau v^k = \sum_{\ell=1}^{k} \eta^{k-\ell} \nabla_\tau \bar{v}^\ell + \eta^k v^0,$$

one has a new series of discrete convolution weights

$$\bar{A}_{n-k}^{(n)} := \sum_{j=k}^{n} A_{n-j}^{(n)} \eta^{j-k} \quad \text{for } 1 \le k \le n.$$

The results of [19, Lemma 3.2] imply that $0 < \eta < 2/3$ and the new convolution kernels $\bar{A}_{n-k}^{(n)}$ are positive and monotone,

$$\bar{A}_0^{(n)} > \bar{A}_1^{(n)} > \cdots > \bar{A}_{n-1}^{(n)} > 0 \quad \text{for } 1 \le k \le n.$$

Thus, our discrete Grönwall inequalities (and the complementary discrete convolution kernels $P_{n-j}^{(n)}$ as well) could be applied for this new form (3.14) directly once a proper constant π_A in A2 is determined by a more careful examination.

Nonetheless, we do not know whether the variable-weights recombination technique [19] works on nonuniform time grids. More precisely, it has yet to be determined what constraints must be imposed on a nonuniform mesh so that the new discrete form (3.14) satisfies the a priori assumptions A1–A3 required by Theorems 3.1 and 3.2. This problem could be very challenging, at least technically, and remains open to us.

4. Stability and consistency. We will now outline how the results of section 3 can be applied to study a numerical solution of problem (1.1). For simplicity, we restrict our attention to the case of a linear reaction term $f(x,t,u) := \kappa u + \psi(x,t)$ with a constant $\kappa \ge 0$. By applying the first Green identity, the fractional PDE (1.1) is written in a weak form as

$$(4.1) \quad \langle \mathcal{D}_t^\alpha u, v \rangle + \mathcal{B}(u,v) = \kappa \langle u, v \rangle + \langle \psi(t), v \rangle \quad \text{for all } v \in H_0^1(\Omega) \text{ and for } 0 < t \le T,$$

where $\langle u, v \rangle$ denotes the inner product in $L_2(\Omega)$, and $\mathcal{B}(u,v) = \langle \mathcal{L}u, v \rangle$ is the bilinear form induced by the elliptic operator $\mathcal{L}$. Since the latter is strongly elliptic, by in-

creasing κ if necessary, we may assume that the bilinear form is coercive: there is a constant $c > 0$ such that

$$(4.2) \qquad \mathcal{B}(v, v) \geq c\|v\|_{H_0^1(\Omega)}^2 \quad \text{for all } v \in H_0^1(\Omega).$$

Let X_h be a finite-dimensional subspace of $H_0^1(\Omega)$; for example, a (conforming) finite element space based on a triangulation of Ω with the mesh size h. Galerkin's method yields a spatially discrete approximate solution $u_h : [0, T] \to X_h$ satisfying

$$(4.3) \quad \langle \mathcal{D}_t^\alpha u_h, \chi \rangle + \mathcal{B}(u_h, \chi) = \kappa\langle u_h, \chi \rangle + \langle \psi(t), \chi \rangle \quad \text{for all } \chi \in X_h \text{ and } 0 < t \leq T$$

with $u_h(0) = u_{h0} \approx u_0$ for a suitable $u_{h0} \in X_h$. To compute a fully discrete numerical solution $u_h^n \in X_h$, where $u(t_n) \approx u_h^n$ for $1 \leq n \leq N$, we apply the approximation (2.2) to the fractional derivative term in (4.3) so that

$$(4.4) \qquad \langle (\mathcal{D}_\tau^\alpha u_h)^{n-\theta}, \chi \rangle + \mathcal{B}\big(u_h^{n-\theta}, \chi\big) = \kappa\langle u_h^{n-\theta}, \chi \rangle + \langle \psi(t_{n-\theta}), \chi \rangle$$

for all $\chi \in X_h$ and for $1 \leq n \leq N$.

The next lemma is a discrete analogue of the inequality [1, Lemma 1]

$$\big(\mathcal{D}_t^\alpha \|v\|^2\big)(t) \leq 2\big\langle (\mathcal{D}_t^\alpha v)(t), v(t) \big\rangle \quad \text{for } 0 \leq t \leq T \text{ and } 0 < \alpha < 1,$$

and helps set the stage for applying our discrete fractional Grönwall inequality.

LEMMA 4.1. *Let the assumption* A1 *hold and fix the parameter* $\theta \in [0, 1)$. *Then every sequence* $(v^n)_{n=0}^N$ *in* $L_2(\Omega)$ *satisfies*

$$\sum_{k=1}^n A_{n-k}^{(n)} \nabla_\tau \big(\|v^k\|^2\big) \leq 2\big\langle (\mathcal{D}_\tau^\alpha v)^{n-\theta}, v^{n-\theta} \big\rangle - d_n\big(\theta^{(n)} - \theta\big)\big\|(\mathcal{D}_\tau^\alpha v)^{n-\theta}\big\|^2$$

for $1 \leq n \leq N$, *where* $0 < d_n < 1/A_0^{(n)}$ *and* $0 < \theta^{(n)} < 1/2$ *are given by*

$$d_n := \frac{2A_0^{(n)} - A_1^{(n)}}{A_0^{(n)}(A_0^{(n)} - A_1^{(n)})} > 0 \quad and \quad \theta^{(n)} := \frac{A_0^{(n)} - A_1^{(n)}}{2A_0^{(n)} - A_1^{(n)}} < \frac{1}{2}.$$

Proof. By Lemma A.1 (see Appendix A),

$$2\big\langle (\mathcal{D}_\tau^\alpha v)^{n-\theta}, v^{n-\theta} \big\rangle = 2\theta\big\langle \mathcal{D}_\tau^\alpha v)^{n-\theta}, v^{n-1} \big\rangle + 2(1-\theta)\big\langle \mathcal{D}_\tau^\alpha v)^{n-\theta}, v^n \big\rangle$$

$$\geq \sum_{k=1}^n A_{n-k}^{(n)} \left(\big\|v^k\big\|^2 - \big\|v^{k-1}\big\|^2 \right)$$

$$+ \left(\frac{1-\theta}{A_0^{(n)}} - \frac{\theta}{A_0^{(n)} - A_1^{(n)}} \right)\big\|(\mathcal{D}_\tau^\alpha v)^{n-\theta}\big\|^2,$$

and the second term on the right side equals $d_n(\theta^{(n)} - \theta)\|(\mathcal{D}_\tau^\alpha v)^{n-\theta}\|^2$. $\qquad\square$

THEOREM 4.2. *Let the assumption* A1 *hold and* $0 \leq \theta \leq \theta^{(n)}$ *for* $1 \leq n \leq N$. *Then the fully discrete solution* $u_h^n \in X_h$, *defined by* (4.4), *satisfies*

$$\sum_{k=1}^n A_{n-k}^{(n)} \nabla_\tau \big(\big\|u_h^k\big\|^2\big) \leq 2\kappa\big\|u_h^{n-\theta}\big\|^2 + 2\big\|u_h^{n-\theta}\big\|\big\|\psi(t_{n-\theta})\big\| \quad for \ 1 \leq n \leq N.$$

Proof. Put $\chi = 2u_h^{n-\theta}$ in the Galerkin discrete equation (4.4), apply Lemma 4.1 with $v^n = u_h^n$, and use positive definiteness (4.2) of the bilinear form. □

Applying the discrete fractional Grönwall inequality from Theorem 3.1 with

$$v^n := \|u_h^n\|, \quad g^n := 2\|\psi(t_{n-\theta})\|, \quad \lambda_0 := 2\kappa, \quad \text{and} \quad \lambda_j := 0 \text{ for } 1 \leq j \leq N-1,$$

we see from Theorem 4.2 that the scheme (4.4) is stable in $L_2(\Omega)$,

$$\|u_h^n\| \leq 2E_\alpha\big(4\max(1,\rho)\pi_A\kappa t_n^\alpha\big)\left(\|u_{0h}\| + 2\max_{1\leq k\leq n}\sum_{j=1}^k P_{k-j}^{(k)}\|\psi(t_{j-\theta})\|\right),$$

provided $\theta \leq \theta^{(n)}$ for $1 \leq n \leq N$. The inequality from Remark 1 yields a weaker but simpler stability estimate,

$$\|u_h^n\| \leq 2E_\alpha\big(4\max(1,\rho)\pi_A\kappa t_n^\alpha\big)\left(\|u_{0h}\| + 2\pi_A\Gamma(1-\alpha)\max_{1\leq k\leq n} t_k^\alpha\|\psi(t_{k-\theta})\|\right).$$

To bound the error in u_h^n, we introduce the Ritz projector $R_h : H_0^1(\Omega) \to X_h$, which is well-defined by

(4.5) $$\mathcal{B}(R_h v, \chi) = \mathcal{B}(v, \chi) \quad \text{for all } v \in H_0^1(\Omega) \text{ and } \chi \in X_h,$$

because the bilinear form satisfies (4.2). Put $e_h^n = u_h^n - R_h u^n \in X_h$, where $u^n = u(t_n)$, so that

$$\|u_h^n - u^n\| \leq \|u^n - R_h u^n\| + \|e_h^n\|.$$

The error in the Ritz projection $R_h u^n$ is estimated in the usual way from a study of the elliptic problem, so it suffices to deal with $\|e_h^n\|$. Using the weak form (4.1) at $t = t_{n-\theta}$ with $v = \chi$, we see that

(4.6) $$\langle(\mathcal{D}_t^\alpha u)(t_{n-\theta}), \chi\rangle + \mathcal{B}(u(t_{n-\theta}), \chi) = \kappa\langle u(t_{n-\theta}), \chi\rangle + \langle\psi(t_{n-\theta}), \chi\rangle.$$

It follows from (4.4) that

$$\begin{aligned}\langle(\mathcal{D}_\tau^\alpha e_h)^{n-\theta}, \chi\rangle + \mathcal{B}(e_h^{n-\theta}, \chi) = {}&\kappa\langle u_h^{n-\theta}, \chi\rangle + \langle\psi(t_{n-\theta}), \chi\rangle\\&- \langle(\mathcal{D}_\tau^\alpha R_h u)^{n-\theta}, \chi\rangle - \mathcal{B}(R_h u^{n-\theta}, \chi).\end{aligned}$$

Therefore, since (4.5) and (4.6) imply

$$\begin{aligned}\mathcal{B}(R_h u^{n-\theta}, \chi) = {}&\mathcal{B}(u^{n-\theta} - u(t_{n-\theta}), \chi) + \mathcal{B}(u(t_{n-\theta}), \chi)\\= {}&-\langle\triangle(u^{n-\theta} - u(t_{n-\theta})), \chi\rangle + \kappa\langle u(t_{n-\theta}), \chi\rangle\\&+ \langle\psi(t_{n-\theta}), \chi\rangle - \langle(\mathcal{D}_t^\alpha u)(t_{n-\theta}), \chi\rangle,\end{aligned}$$

we have

$$\langle(\mathcal{D}_\tau^\alpha e_h)^{n-\theta}, \chi\rangle + \mathcal{B}(e_h^{n-\theta}, \chi) = \kappa\langle e_h^{n-\theta}, \chi\rangle + \langle\mathcal{R}^n, \chi\rangle \quad \text{for all } \chi \in X_h,$$

where

$$\mathcal{R}^n = (\mathcal{D}_t^\alpha u)(t_{n-\theta}) - (\mathcal{D}_\tau^\alpha R_h u)^{n-\theta} - \kappa\big(u(t_{n-\theta}) - R_h u^{n-\theta}\big) + \triangle\big(u^{n-\theta} - u(t_{n-\theta})\big).$$

Choosing $\chi = 2e_h^{n-\theta}$ and arguing as before, but now with $v^n := \|e_h^n\|$ and $g^n := 2\|\mathcal{R}^n\|$, we see that (for appropriate θ)

$$\|e_h^n\| \leq 2E_\alpha\big(4\max(1,\rho)\pi_A\kappa t_n^\alpha\big)\left(\|u_{0h} - u_0\| + 2\max_{1\leq k\leq n}\sum_{j=1}^{k} P_{k-j}^{(k)}\|\mathcal{R}^j\|\right)$$

for $1 \leq n \leq N$. A complete error analysis would typically proceed by applying the triangle inequality to obtain

$$\|\mathcal{R}^j\| \leq \big\|(\mathcal{D}_t^\alpha u)(t_{j-\theta}) - (\mathcal{D}_\tau^\alpha u)^{j-\theta}\big\| + \big\|(\mathcal{D}_\tau^\alpha(u - R_h u))^{j-\theta}\big\|$$
$$+ \kappa\big\|(u - R_h u)^{j-\theta}\big\| + \big\|(\kappa + \triangle)\big(u^{j-\theta} - u(t_{j-\theta})\big)\big\|,$$

and estimating separately the resulting convolutional sums over j, refer to a new technique of global consistency error analysis developed in recent works [15, 16, 17]. The details would depend on the choice of the discrete kernels $A_{n-j}^{(n)}$ and of the space X_h, and would rely on some a priori estimates for the partial derivatives of u.

A similar approach works if finite differences are used for the space discretization [15], by introducing an appropriate discrete ℓ_2 inner product in place of the inner product $\langle u, v\rangle$.

Appendix A. Two technical inequalities.   The proof of Lemma 4.1 relies on the following result, essentially due to Alikhanov [2, Lemma 1].

LEMMA A.1. *If the assumption* A1 *holds, then every sequence* $(v^n)_{n=0}^N$ *in* $L_2(\Omega)$ *satisfies*

$$2\langle(\mathcal{D}_\tau^\alpha v)^{n-\theta}, v^n\rangle \geq \sum_{k=1}^n A_{n-k}^{(n)}\Big(\|v^k\|^2 - \|v^{k-1}\|^2\Big) + \frac{\|(\mathcal{D}_\tau^\alpha v)^{n-\theta}\|^2}{A_0^{(n)}}$$

and

$$2\langle(\mathcal{D}_\tau^\alpha v)^{n-\theta}, v^{n-1}\rangle \geq \sum_{k=1}^n A_{n-k}^{(n)}\Big(\|v^k\|^2 - \|v^{k-1}\|^2\Big) - \frac{\|(\mathcal{D}_\tau^\alpha v)^{n-\theta}\|^2}{A_0^{(n)} - A_1^{(n)}}$$

for $1 \leq n \leq N$, *provided we set* $A_1^{(1)} = 0$ *in the case* $n = 1$.

Proof. Fix n and consider the difference

$$J_n := 2\langle(\mathcal{D}_\tau^\alpha v)^{n-\theta}, v^n\rangle - \sum_{k=1}^n A_{n-k}^{(n)}\Big(\|v^k\|^2 - \|v^{k-1}\|^2\Big).$$

We have

$$J_n = \sum_{k=1}^n A_{n-k}^{(n)}\big(2\langle v^k - v^{k-1}, v^n\rangle - \langle v^k - v^{k-1}, v^k + v^{k-1}\rangle\big)$$

$$= \sum_{k=1}^n A_{n-k}^{(n)}\langle v^k - v^{k-1}, 2v^n - (v^k + v^{k-1})\rangle$$

and, using the identity $2v^n - (v^k + v^{k-1}) = v^k - v^{k-1} + 2\sum_{j=k+1}^n(v^j - v^{j-1})$,

$$J_n = \sum_{k=1}^n A_{n-k}^{(n)}\big\|v^k - v^{k-1}\big\|^2 + 2\sum_{k=1}^n A_{n-k}^{(n)}\sum_{j=k+1}^n\langle v^k - v^{k-1}, v^j - v^{j-1}\rangle$$

$$= \sum_{k=1}^n A_{n-k}^{(n)}\big\|v^k - v^{k-1}\big\|^2 + 2\sum_{j=2}^n\sum_{k=1}^{j-1} A_{n-k}^{(n)}\langle v^k - v^{k-1}, v^j - v^{j-1}\rangle.$$

To continue the proof, it is convenient to introduce

$$w^j := \sum_{k=1}^{j} A_{n-k}^{(n)}(v^k - v^{k-1}) \quad \text{and} \quad Q_j := \frac{1}{A_{n-j}^{(n)}} \quad \text{for } 1 \le j \le n.$$

Notice that $v^j - v^{j-1} = Q_j(w^j - w^{j-1})$ for $2 \le j \le n$, and that the assumption A1 implies $Q_1 \ge Q_2 \ge \cdots \ge Q_n$. Thus, one deduces that

$$\begin{aligned}
J_n &= Q_1\|w^1\|^2 + \sum_{j=2}^{n} Q_j\|w^j - w^{j-1}\|^2 + 2\sum_{j=2}^{n} Q_j\langle w^{j-1}, w^j - w^{j-1}\rangle \\
&= Q_1\|w^1\|^2 + \sum_{j=2}^{n} Q_j\left(\|w^j\|^2 - \|w^{j-1}\|^2\right) \\
&= Q_n\|w^n\|^2 + \sum_{j=1}^{n-1}(Q_j - Q_{j+1})\|w^j\|^2 \ge Q_n\|w^n\|^2.
\end{aligned}$$

The first inequality now follows by noting that $w^n = (\mathcal{D}_\tau^\alpha v)^{n-\theta}$ and $Q_n = 1/A_0^{(n)}$. Furthermore, by using the identity $v^{n-1} = v^n - (v^n - v^{n-1}) = v^n - Q_n(w^n - w^{n-1})$, we have

$$\begin{aligned}
&2\langle(\mathcal{D}_\tau^\alpha v)^{n-\theta}, v^{n-1}\rangle - \sum_{k=1}^{n} A_{n-k}^{(n)}\left(\|v^k\|^2 - \|v^{k-1}\|^2\right) \\
&= J_n - 2Q_n\langle w^n, w^n - w^{n-1}\rangle \\
&\ge Q_n\|w^n\|^2 + (Q_{n-1} - Q_n)\|w^{n-1}\|^2 - 2Q_n\langle w^n, w^n - w^{n-1}\rangle \\
&= -Q_n\|w^n\|^2 + 2Q_n\langle w^n, w^{n-1}\rangle + (Q_{n-1} - Q_n)\|w^{n-1}\|^2 \\
&= \frac{1}{Q_{n-1} - Q_n}\left(\|Q_n w^n + (Q_{n-1} - Q_n)w^{n-1}\|^2 - Q_n Q_{n-1}\|w^n\|^2\right) \\
&\ge -\frac{Q_n Q_{n-1}}{Q_{n-1} - Q_n}\|w^n\|^2 = \frac{-\|w^n\|^2}{A_0^{(n)} - A_1^{(n)}}.
\end{aligned}$$

Therefore the claimed second inequality follows and the proof is complete. $\qquad\square$

REFERENCES

[1] A. A. ALIKHANOV, *A priori estimates for solutions of boundary value problems for fractional-order equations*, Differ. Equ., 46 (2010), pp. 660–666.

[2] A. A. ALIKHANOV, *A new difference scheme for the time fractional diffusion equation*, J. Comput. Phys., 280 (2015), pp. 424–438.

[3] J. BECKER, *A second order backward difference method with variable steps for a parabolic problem*, BIT, 38 (1998), pp. 644–662.

[4] H. Brunner, *The numerical solution of weakly singular Volterra integral equations by collocation on graded meshes*, Math. Comp., 45 (1985), pp. 417–437.

[5] J. Dixon and S. McKee, *Weakly singular discrete Gronwall inequalities*, ZAMM Z. Angew. Math. Mech., 66 (1986), pp. 535–544.

[6] E. Emmrich, *Convergence of the variable two-step BDF time discretisation of nonlinear evolution problems governed by a monotone potential operator*, BIT, 49 (2009), pp. 297–323.

[7] G. Gao, Z. Sun, and H. Zhang, *A new fractional numerical differentiation formula to approximate the Caputo fractional derivative and its applications*, J. Comput. Phys., 259 (2014), pp. 33–50.

[8] I. G. Graham, *Galerkin methods for second kind integral equations with singularities*, Math. Comp., 39 (1982), pp. 519–533.

[9] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge University Press, Cambridge, 1934.

[10] B. Jin, R. Lazarov, and Z. Zhou, *An analysis of the L1 scheme for the subdiffusion equation with nonsmooth data*, IMA J. Numer. Anal., 36 (2016), pp. 197–221.

[11] B. Jin, B. Li, and Z. Zhou, *Numerical analysis of nonlinear subdiffusion equations*, SIAM J. Numer. Anal., 56 (2018), pp. 1–23.

[12] D. Li, H. Liao, W. Sun, J. Wang, and J. Zhang, *Analysis of L1-Galerkin FEMs for time-fractional nonlinear parabolic problems*, Commun. Comput. Phys., 24 (2018), pp. 86–103.

[13] H.-L. Liao, Y. Zhao, and X. Teng, *A weighted ADI scheme for subdiffusion equations*, J. Sci. Comput., 69 (2016), pp. 1144–1164.

[14] H.-L. Liao, P. Lyu, S. W. Vong and Y. Zhao, *Stability of fully discrete schemes with interpolation-type fractional formulas for distributed-order subdiffusion equations*, Numer. Algorithms, 75 (2017), pp. 845–878.

[15] H.-L. Liao, D. Li, and J. Zhang, *Sharp error estimate of the nonuniform L1 formula for linear reaction-subdiffusion equations*, SIAM J. Numer. Anal., 56 (2018), pp. 1112–1133.

[16] H.-L. Liao, W. McLean, and J. Zhang, *A second-order scheme with nonuniform time steps for a linear reaction-subdiffusion equation*, Found. Comput. Math., submitted.

[17] H.-L. Liao, Y. Yan, and J. Zhang, *Unconditional convergence of a two-level linearized fast algorithm for nonlinear subdiffusion equations*, J. Sci. Comput., submitted.

[18] Y. Lin and C. Xu, *Finite difference/spectral approximations for the time-fractional diffusion equation*, J. Comput. Phys., 225 (2007), pp. 1533–1552.

[19] C. Lv and C. Xu, *Error analysis of a high order method for time-fractional diffusion equations*, SIAM J. Sci. Comput., 38 (2016), A2699–A2724.

[20] W. McLean and K. Mustapha, *A second-order accurate numerical method for a fractional wave equation*, Numer. Math., 105 (2007), pp. 481–510.

[21] K. B. Oldham and J. Spanier, *The Fractional Calculus*, Academic Press, New York, 1974.

[22] I. Podlubny, *Fractional Differential Equations*, Academic Press, New York, 1999.

[23] K. Sakamoto and M. Yamamoto, *Initial value/boundary value problems for fractional diffusion-wave equations and applications to some inverse problems*, J. Math. Anal. Appl., 382 (2011), pp. 426–447.

[24] M. Stynes, *Too much regularity may force too much uniqueness*, Fract. Calc. Appl. Anal., 19 (2016), pp. 1554–1562.

[25] 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 (2017), pp. 1057–1079.

[26] Z. Sun and X. Wu, *A fully discrete difference scheme for a diffusion-wave system*, Appl. Numer. Math., 56 (2006), pp. 193–209.

[27] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, 2nd ed., Springer, Berlin, 2006.

[28] Y. Yan, M. Khan, and N. J. Ford, *An analysis of the modified L1 scheme for time fractional partial differential equations with nonsmooth data*, SIAM J. Numer. Anal., 56 (2018), pp. 210–227.