



An Efficient Second-Order Convergent Scheme for One-Side Space Fractional Diffusion Equations with Variable Coefficients

Xue-lei Lin¹ · Pin Lyu² · Michael K. Ng³ · Hai-Wei Sun⁴ · Seakweng Vong⁴

Received: 2 June 2019 / Revised: 11 October 2019 / Accepted: 17 October 2019 /
Published online: 17 January 2020
© Shanghai University 2020

Abstract

In this paper, a second-order finite-difference scheme is investigated for time-dependent space fractional diffusion equations with variable coefficients. In the presented scheme, the Crank–Nicolson temporal discretization and a second-order weighted-and-shifted Grünwald–Letnikov spatial discretization are employed. Theoretically, the unconditional stability and the second-order convergence in time and space of the proposed scheme are established under some conditions on the variable coefficients. Moreover, a Toeplitz preconditioner is proposed for linear systems arising from the proposed scheme. The condition number of the preconditioned matrix is proven to be bounded by a constant independent of the discretization step-sizes, so that the Krylov subspace solver for the preconditioned linear systems converges linearly. Numerical results are reported to show the convergence rate and the efficiency of the proposed scheme.

Keywords One-side space fractional diffusion equation · Variable diffusion coefficients · Stability and convergence · High-order finite-difference scheme · Preconditioner

Mathematics Subject Classification 26A33 · 35R11 · 65M06 · 65M12

1 Introduction

In the paper, we study an efficient numerical method for solving the one-side space fractional diffusion equation (OSFDE) with variable coefficients. To begin with, we first present the one-spatial-dimensional (1-D) OSFDE (the two-dimensional case will be discussed in Sect. 3) [20, 37, 38, 40]:

This research was supported by research Grants, 12306616, 12200317, 12300519, 12300218 from HKRGC GRF, 11801479 from NSFC, MYRG2018-00015-FST from University of Macau, and 0118/2018/A3 from FDCT of Macao, Macao Science and Technology Development Fund 0005/2019/A, 050/2017/A, and the Grant MYRG2017-00098-FST and MYRG2018-00047-FST from University of Macau.

✉ Pin Lyu
plyu@swufe.edu.cn

Extended author information available on the last page of the article

$$\begin{cases} \frac{\partial u(x,t)}{\partial t} = d(x) {}_{x_L} D_x^\alpha u(x,t) + f(x,t), & x \in (x_L, x_R), t \in (0, T], \\ u(x_L, t) = 0, \quad u(x_R, t) = \psi(t), & t \in [0, T], \\ u(x, 0) = \varphi(x), & x \in [x_L, x_R], \end{cases} \quad (1)$$

where $d(x)$, which satisfies $0 < d_- \leq d(x) \leq d_+ < \infty$, is a strictly positive known function, ψ , φ , and f are all known functions, u is unknown to be solved, and ${}_{x_L} D_x^\alpha u(x, t)$ is the Riemann–Liouville (RL) fractional derivative of order $\alpha \in (1, 2)$ defined as [29, 36]

$${}_{x_L} D_x^\alpha u(x, t) = \frac{1}{\Gamma(2-\alpha)} \frac{\partial^2}{\partial x^2} \int_{x_L}^x \frac{u(\xi, t)}{(x-\xi)^{\alpha-1}} d\xi \quad (2)$$

with $\Gamma(\cdot)$ denoting the gamma function.

Due to the nonlocal dependence, fractional derivatives model many challenging phenomena more accurately than integer-order derivatives do, which has, therefore, attracted lots of interests in recent years. For example, in [1], a two-dimensional version of the OSFDE is applied to image denoising, where the noisy image is the initial value and the solution of the diffusion equation at final time is the denoised image. Due to the nonlocal property of the fractional derivative, the fractional diffusion-based image denoising model [1] has a good capability of texture preserving. For more applications of the fractional diffusion equation, one may refer to viscoelasticity [12], fractal dynamics [34], signal processing [3, 35, 42], image processing [30–32], and the references therein.

Nevertheless, it is well known that closed-form analytic solutions of fractional diffusion equations are usually not available, especially in the existence of variable coefficients. Because of this, many numerical discretization schemes have been developed for fractional diffusion equations; see, e.g., [5, 10, 17–19, 22, 25, 28, 45–47].

For those schemes applicable to or solely developed for OSFDEs, one may refer to [2, 6, 9, 15, 24, 26, 33, 37–41, 44]. In [9, 15, 44], numerical schemes with spatial fourth-order convergence for a space fractional diffusion equation are developed by applying the technique of compact operators, which is, however, only available for the constant-coefficient case. Another spatially fourth-order accurate scheme is studied in [2] by implementing weighted-and-shifted Lubich difference operators whose convergent property is established only for constant diffusion coefficients. Some second-order numerical schemes are proposed in [37–40] for solving OSFDEs with variable coefficients, which, however, does not provide convergence proof. In [33], the stability and convergence of the second-order numerical scheme for variable coefficient equations are established for the 1-D OSFDE, which cannot be extended to the case of higher spatial dimension. In [21], a series of numerical schemes for the Riesz space fractional diffusion equation have been proven to be convergent and stable. Nevertheless, the proof technique used in [21] heavily depends on the symmetry of discretization matrix of the Riesz fractional derivative, which is not applicable to the OSFDE that involves the non-symmetric one-sided fractional derivatives weighted by variable coefficients. The shift-Grünwald spatial scheme together with the backward difference temporal scheme proposed in [23] is applicable to the OSFDE with variable coefficients, whose stability and convergence can be established under infinity norm. Nevertheless, its convergence rate is only of first order in time and space. Hence, there is no second-order scheme for the two-spatial-dimension (2-D) OSFDE with variable coefficients.

In this paper, we propose a second-order scheme for 1-D and 2-D OSFDEs with variable coefficients. The Crank–Nicolson method and a second-order weighted-and-shifted

Grünwald–Letnikov difference (WSGD, see [41]) operator are employed to discretize temporal and spatial derivatives, respectively. The key point of stability and convergence proof is to find an inner product under which the spatial discretization matrix is negative semi-definite. For the 1-D case, we choose the inner product associated with the diagonal matrix arising from discretization of d^{-1} . As a result, the unconditional stability and convergence of the proposed scheme are established without additional assumption on $d(x)$ for the 1-D case. For the 2-D case, to construct a desired inner product, some assumptions are made on the variable coefficients (see the assumptions in Corollaries 3.1–3.2), with which the unconditional stability and convergence of the proposed scheme are established.

Moreover, because of the nonlocality of the fractional derivative and the existence of the variable coefficients, the discretization of the OSFDE tends to generate dense matrix with high displacement rank [27], for which the discrete linear systems related to the variable coefficients OSFDE are time-consuming to directly solve. Fortunately, the discretization matrix has Toeplitz-like structure due to which its matrix–vector multiplication can be fast computed via fast Fourier transforms (FFTs). Because of the fast matrix–vector multiplication, fast iterative solvers for the linear systems can be possibly developed. However, the discretization matrix of the OSFDE is ill-conditioned when τ/h^α is large, where τ and h represent the temporal and spatial step-sizes, respectively. Thus, a Toeplitz preconditioner is proposed to reduce the condition number of 1-D and 2-D discretization matrices. Theoretically, we show that the condition number of the preconditioned matrix is uniformly bounded by a constant independent of τ and h under certain conditions on the diffusion coefficients (see the assumptions in Theorems 2.4, 3.3), so that the Krylov subspace method for the preconditioned linear systems converges linearly no matter the unpreconditioned matrix is ill-conditioned or not.

To summarize, the contribution of this paper is of twofold: (i) the proposed scheme is the first scheme of second-order accuracy for the 2-D OSFDE with variable coefficients; (ii) the proposed preconditioning technique is the first fast solver with the convergence rate independent of discretization step-sizes for linear systems from second-order discretization of the 2-D OSFDE with variable coefficients.

This paper is organized as follows. In Sect. 2, we propose a second-order scheme and its corresponding Toeplitz preconditioner for the one-dimensional OSFDE, analyze the unconditional stability and convergence of the proposed scheme, and estimate the condition number of the preconditioned matrix. In Sect. 3, we extend the scheme and the preconditioner to the two-dimensional case. In Sect. 4, numerical results are reported to show the efficiency and accuracy of the proposed scheme.

2 Stability and Convergence of Discrete One-Dimensional OSFDE and Its Preconditioning

We need some notations to describe the discretization for (1). Let $h = (x_R - x_L)/(M + 1)$ and $\tau = T/N$ be the space and time step-sizes, respectively, where M and N are given positive integers. And denote $x_i = x_L + ih$ for $i = 0, 1, \dots, M + 1$, $t_n = n\tau$ for $n = 0, 1, \dots, N$. Throughout this paper, the discretization on the RL fractional derivative is based on the following second-order WSGD formula [41]:

$${}_{x_L}D_x^\alpha u(x_i) = \frac{1}{h^\alpha} \sum_{k=0}^i w_k^{(\alpha)} u(x_{i-k+1}) + \mathcal{O}(h^2), \tag{3}$$

which is under the smooth assumptions $u, {}_{-\infty}D_x^{\alpha+2}u$ and Fourier transform of ${}_{-\infty}D_x^{\alpha+2}u$ belong to $L^1(\mathbb{R})$ (see, e.g., [4]). The coefficients $w_k^{(\alpha)}$ were defined by [41]

$$w_0^{(\alpha)} = \frac{\alpha}{2}g_0^{(\alpha)}, w_k^{(\alpha)} = \frac{\alpha}{2}g_k^{(\alpha)} + \frac{2-\alpha}{2}g_{k-1}^{(\alpha)} \text{ for } k \geq 1, \tag{4}$$

where $g_k^{(\alpha)}$ are the coefficients of the power series of $(1-z)^\alpha$, and they can be obtained recursively as

$$g_0^{(\alpha)} = 1, \quad g_k^{(\alpha)} = \left(1 - \frac{\alpha+1}{k}\right)g_{k-1}^{(\alpha)} \text{ for } k = 1, 2, \dots.$$

Next, we introduce the finite-difference scheme for solving (1). Let u_i^n be the numerical approximation of $u(x_i, t_n)$. Denote $d_i = d(x_i)$, $\varphi_i = \varphi(x_i)$, $f_i^{n-\frac{1}{2}} = f(x_i, t_{n-\frac{1}{2}})$ for $1 \leq i < M$ and $f_M^{n-\frac{1}{2}} = f(x_M, t_{n-\frac{1}{2}}) + w_0^{(\alpha)}\psi(t_{n-\frac{1}{2}})/h^\alpha$, where $t_{n-\frac{1}{2}} = (t_{n-1} + t_n)/2$, $n = 1, 2, \dots, N$. Then, applying the Crank–Nicolson technique and approximation (3) to the time derivative and the space fractional derivatives of (1), respectively, we get

$$\frac{u_i^n - u_i^{n-1}}{\tau} = \frac{1}{2h^\alpha}d_i \sum_{k=0}^i w_k^{(\alpha)}(u_{i-k+1}^{n-1} + u_{i-k+1}^n) + f_i^{n-\frac{1}{2}} + R_i^{n-\frac{1}{2}}, \quad 1 \leq i \leq M, \tag{5}$$

where $|R_i^{n-\frac{1}{2}}| \leq c_1(\tau^2 + h^2)$ for a positive constant c_1 ; see, e.g., [41].

Denote $u^n = [u_1^n, u_2^n, \dots, u_M^n]^T$, $f^{n-\frac{1}{2}} = [f_1^{n-\frac{1}{2}}, f_2^{n-\frac{1}{2}}, \dots, f_M^{n-\frac{1}{2}}]^T$, and

$$D = \text{diag}(d_1, d_2, \dots, d_M), \quad G_\alpha = \begin{bmatrix} w_1^{(\alpha)} & w_0^{(\alpha)} & 0 & \dots & 0 \\ w_2^{(\alpha)} & w_1^{(\alpha)} & w_0^{(\alpha)} & \ddots & \vdots \\ \vdots & w_2^{(\alpha)} & w_1^{(\alpha)} & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & w_0^{(\alpha)} \\ w_M^{(\alpha)} & \dots & \dots & w_2^{(\alpha)} & w_1^{(\alpha)} \end{bmatrix}, \tag{6}$$

where $\{w_k^{(\alpha)}\}_{k=0}^M$ are the coefficients given in (4).

Omitting the small term $R_i^{n-\frac{1}{2}}$ in (5), Eq. (1) can be solved numerically by the following finite-difference scheme in the matrix form:

$$\frac{1}{\tau}(u^n - u^{n-1}) = \frac{1}{2h^\alpha}DG_\alpha(u^{n-1} + u^n) + f^{n-\frac{1}{2}}, \quad n = 1, 2, \dots, N. \tag{7}$$

2.1 Stability and Convergence

Some general notations:

- $\mathbb{C}^{m \times n}$ ($\mathbb{R}^{m \times n}$, respectively) denotes the set of all $m \times n$ complex (real, respectively) matrices;
- $\mathcal{H}(X)$ denotes the symmetric part of a square matrix X .

Lemma 2.1 [41] *The matrix $G_\alpha + G_\alpha^T$ is negative definite.*

Lemma 2.2 [11, 13] *Let the symmetric matrix $H \in \mathbb{R}^{n \times n}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Then, for all $w \in \mathbb{R}^{n \times 1}$,*

$$\lambda_n w^T w \leq w^T H w \leq \lambda_1 w^T w.$$

Now, we show the stability and convergence of the scheme (7) by the energy method.

Theorem 2.1 *The finite-difference scheme (7) is unconditionally stable and its solution satisfies the following estimate:*

$$\|u^n\|_{D^{-1}}^2 \leq \exp(2T)\|\varphi\|_{D^{-1}}^2 + [\exp(2T) - 1] \max_{1 \leq k \leq n} \|f^{k-\frac{1}{2}}\|_{D^{-1}}^2, \quad n = 1, 2, \dots, N,$$

where $\|\cdot\|_{D^{-1}}$ is the norm induced by the inner product $\langle v_1, v_2 \rangle_{D^{-1}} := hv_1^T D^{-1} v_2$.

Proof Some steps of this proof are similar to those of Theorem 3.8 in [43]. Multiplying $h(u^{n-1} + u^n)^T D^{-1}$ on the both sides of (7), we get

$$\begin{aligned} \frac{1}{\tau} h(u^{n-1} + u^n)^T D^{-1} (u^n - u^{n-1}) &= \frac{1}{2h^\alpha} h(u^{n-1} + u^n)^T G_\alpha (u^{n-1} + u^n) \\ &\quad + h(u^{n-1} + u^n)^T D^{-1} f^{n-\frac{1}{2}}. \end{aligned} \tag{8}$$

Notice that $w^T G_\alpha w = w^T \mathcal{H}(G_\alpha) w$ for any real vector w . Therefore, by Lemma 2.1, the first term on the right-hand side of (8) can be estimated as

$$\begin{aligned} &\frac{1}{2h^\alpha} h(u^{n-1} + u^n)^T G_\alpha (u^{n-1} + u^n) \\ &= \frac{1}{2h^\alpha} h(u^{n-1} + u^n)^T \mathcal{H}(G_\alpha) (u^{n-1} + u^n) \leq 0. \end{aligned}$$

As a result

$$\begin{aligned} &h(u^n)^T D^{-1} u^n - h(u^{n-1})^T D^{-1} u^{n-1} \\ &\leq \tau h(u^n)^T D^{-1} f^{n-\frac{1}{2}} + \tau h(u^{n-1})^T D^{-1} f^{n-\frac{1}{2}}. \end{aligned} \tag{9}$$

Applying the Cauchy–Schwarz inequality on the right-hand side of (9), we get

$$\|u^n\|_{D^{-1}}^2 \leq \|u^{n-1}\|_{D^{-1}}^2 + \frac{\tau}{2} \|u^n\|_{D^{-1}}^2 + \frac{\tau}{2} \|u^{n-1}\|_{D^{-1}}^2 + \tau \|f^{n-\frac{1}{2}}\|_{D^{-1}}^2,$$

which is equivalent to

$$\|u^n\|_{D^{-1}}^2 \leq \frac{2 + \tau}{2 - \tau} \|u^{n-1}\|_{D^{-1}}^2 + \frac{2\tau}{2 - \tau} \|f^{n-\frac{1}{2}}\|_{D^{-1}}^2. \tag{10}$$

Iterating (10) for n times, we obtain

$$\begin{aligned} \|u^n\|_{D^{-1}}^2 &\leq \left(\frac{2 + \tau}{2 - \tau}\right)^n \|u^0\|_{D^{-1}}^2 \\ &\quad + \frac{2\tau}{2 - \tau} \left[1 + \frac{2 + \tau}{2 - \tau} + \left(\frac{2 + \tau}{2 - \tau}\right)^2 + \dots + \left(\frac{2 + \tau}{2 - \tau}\right)^{n-1}\right] \max_{1 \leq k \leq n} \|f^{k-\frac{1}{2}}\|_{D^{-1}}^2. \end{aligned} \tag{11}$$

For the small τ ($\tau \leq 1$), we have

$$\left(\frac{2+\tau}{2-\tau}\right)^n = \left(1 + \frac{2\tau}{2-\tau}\right)^n \leq (1+2\tau)^n \leq \lim_{N \rightarrow +\infty} \left(1 + \frac{2T}{N}\right)^N = \exp(2T), \tag{12}$$

and

$$\frac{2\tau}{2-\tau} \sum_{k=1}^n \left(\frac{2+\tau}{2-\tau}\right)^{k-1} = \left(\frac{2+\tau}{2-\tau}\right)^n - 1 \leq \exp(2T) - 1. \tag{13}$$

The result follows from (11)–(13).

Theorem 2.2 *Let $u(x_i, t_n)$ be the exact solution of (1) and u_i^n be the solution of the finite-difference scheme (7). Denote $e_i^n = u(x_i, t_n) - u_i^n$, $0 \leq i \leq M + 1$, $0 \leq n \leq N$. Then, there exists a positive constant c_2 such that*

$$\|e^n\| \leq c_2(\tau^2 + h^2),$$

where $e^n = [e_1^n, e_2^n, \dots, e_M^n]^T$ and $\|\cdot\|$ denotes the discrete L^2 norm, i.e., $\|v\| = \sqrt{hv^T v}$.

Proof Denote $R^{n-\frac{1}{2}} = [R_1^{n-\frac{1}{2}}, R_2^{n-\frac{1}{2}}, \dots, R_M^{n-\frac{1}{2}}]^T$. We can easily show that e^n and e_i^n satisfy the following error equations:

$$\begin{aligned} \frac{1}{\tau}(e^n - e^{n-1}) &= \frac{1}{2h^\alpha} DG_\alpha(e^{n-1} + e^n) + R^{n-\frac{1}{2}}, \quad 1 \leq n \leq N, \\ e_0^n &= e_{M+1}^n = 0, \quad 1 \leq n \leq N, \quad e_i^0 = 0, \quad 0 \leq i \leq M + 1. \end{aligned}$$

By Theorem 2.1, we have

$$\|e^n\|_{D^{-1}}^2 \leq [\exp(2T) - 1] \max_{1 \leq k \leq n} \|R^{k-\frac{1}{2}}\|_{D^{-1}}^2, \quad n = 1, 2, \dots, N.$$

As D^{-1} is a positive diagonal matrix, utilizing Lemma 2.2, we get

$$\|e^n\|^2 \leq [c_2(\tau^2 + h^2)]^2, \quad n = 1, 2, \dots, N.$$

2.2 An Estimate on the Field of Values of $DG_\alpha + G_\alpha^T D$

In this subsection, we focus on estimating the field of values of $DG_\alpha + G_\alpha^T D$, the results of which will be further applied to the analysis of one-dimensional preconditioning and the extension to the two-dimensional OSFDE. First, we denote $g(\alpha, x)$ as the generating function [27] of the Toeplitz matrix G_α . The next two lemmas describe some properties concerning $g(\alpha, x)$, which will be useful in obtaining the desired estimation.

Lemma 2.3 [27] *Let $\mathbf{u} = [u_1, u_2, \dots, u_M]^T$, $\mathbf{v} = [v_1, v_2, \dots, v_M]^T \in \mathbb{R}^{M \times 1}$. Then, we have*

$$\mathbf{u}^T G_\alpha \mathbf{v} = \frac{1}{2\pi} \int_{-\pi}^\pi \sum_{k=1}^M u_k e^{-ikx} \sum_{k=1}^M v_k e^{ikx} g(\alpha, x) dx.$$

Lemma 2.4 [43] *It holds that*

$$\varsigma_\alpha \triangleq \min_x \frac{\Re[-g(\alpha, x)]}{|g(\alpha, x)|} = \left| \cos\left(\frac{\alpha}{2}\pi\right) \right|,$$

where $\Re[g(\alpha, x)]$ denotes the real part of $g(\alpha, x)$.

The following lemma provides a novel bound to the field of values of $\tilde{D}G\tilde{D}$, where $G = -G_\alpha - G_\alpha^T$ and \tilde{D} is a diagonal matrix satisfying some properties.

Lemma 2.5 [43] *Denote $G = -G_\alpha - G_\alpha^T$. Suppose that $\tilde{D} = \text{diag}(\tilde{d}(x_1), \tilde{d}(x_2), \dots, \tilde{d}(x_M))$ for some function $\tilde{d}(x)$ defined on (x_L, x_R) . For any real vector $\mathbf{u} = [u_1, u_2, \dots, u_M]^T$, we have*

$$\mathbf{u}^T \tilde{D}G\tilde{D}\mathbf{u} \leq 2 \max_i \{|\tilde{d}_i|^2\} \mathbf{u}^T G\mathbf{u}, \tag{14}$$

if $\tilde{d}(x)$ is convex and $\tilde{d}(x) \geq 0$, or $\tilde{d}(x)$ is concave and $\tilde{d}(x) \leq 0$.

Assuming that $0 \leq \kappa_{\min} \leq d(x) \leq \kappa_{\max} < \infty$. The following theorem reveals some inclusion relations between numerical ranges of G and $-DG_\alpha - G_\alpha^T D$, which acts an important role in the analysis of the proposed preconditioner.

Theorem 2.3 *For any $\mathbf{u} = [u_1, u_2, \dots, u_M]^T$, we have*

$$\begin{aligned} & \left(\kappa - \frac{\sqrt{2}(\kappa_{\max} - \kappa_{\min})}{\varsigma_\alpha} \right) \mathbf{u}^T G\mathbf{u} \\ & \leq \mathbf{u}^T (-DG_\alpha - G_\alpha^T D)\mathbf{u} \leq \left(\kappa + \frac{\sqrt{2}(\kappa_{\max} - \kappa_{\min})}{\varsigma_\alpha} \right) \mathbf{u}^T G\mathbf{u}, \end{aligned} \tag{15}$$

where $\kappa = \kappa_{\max}$ when $d(x)$ is concave, and $\kappa = \kappa_{\min}$ when $d(x)$ is convex.

Proof Denote $\tilde{D} = D - \kappa I$, then $DG_\alpha + G_\alpha^T D = \kappa(G_\alpha + G_\alpha^T) + \tilde{D}G_\alpha + G_\alpha^T \tilde{D}$. And, for any $\mathbf{u} = [u_1, u_2, \dots, u_M]^T$, we have

$$\mathbf{u}^T (-DG_\alpha - G_\alpha^T D)\mathbf{u} = \kappa \mathbf{u}^T G\mathbf{u} + \mathbf{u}^T (-\tilde{D}G_\alpha - G_\alpha^T \tilde{D})\mathbf{u}.$$

Denote $u(x) = \sum_{k=1}^M u_k e^{ikx}$ and $v(x) = \sum_{k=1}^M (\tilde{D}u)_k e^{ikx}$, it follows by Lemmas 2.3 and 2.5 that

$$\mathbf{u}^T G\mathbf{u} = \mathbf{u}^T (-G_\alpha - G_\alpha^T)\mathbf{u} = \frac{1}{\pi} \int_{-\pi}^{\pi} \Re[-g(\alpha, x)] |u(x)|^2 dx, \tag{16}$$

$$\begin{aligned} \mathbf{u}^T \tilde{D}G\tilde{D}\mathbf{u} &= \frac{1}{\pi} \int_{-\pi}^{\pi} \Re[-g(\alpha, x)] |v(x)|^2 dx \\ &\leq \frac{2(\kappa_{\max} - \kappa_{\min})^2}{\pi} \int_{-\pi}^{\pi} \Re[-g(\alpha, x)] |u(x)|^2 dx. \end{aligned} \tag{17}$$

Using Lemma 2.3 again and applying the Cauchy–Schwarz inequality, Lemma 2.4, (16) and (17), we get

$$\begin{aligned}
 & \left| \mathbf{u}^T (-\tilde{D}G_\alpha - G_\alpha^T \tilde{D}) \mathbf{u} \right| \\
 &= \frac{1}{2\pi} \left| \int_{-\pi}^\pi (-v^* g u - u^* g^* v) dx \right| \\
 &\leq \frac{1}{\pi} \int_{-\pi}^\pi |g(\alpha, x)| |v(x)| |u(x)| dx \\
 &\leq \frac{1}{\pi \zeta_\alpha} \int_{-\pi}^\pi \Re[-g(\alpha, x)] |v(x)| |u(x)| dx \\
 &\leq \frac{1}{\pi \zeta_\alpha} \sqrt{\int_{-\pi}^\pi \Re[-g(\alpha, x)] |v(x)|^2 dx} \sqrt{\int_{-\pi}^\pi \Re[-g(\alpha, x)] |u(x)|^2 dx} \\
 &\leq \frac{\sqrt{2}(\kappa_{\max} - \kappa_{\min})}{\pi \zeta_\alpha} \int_{-\pi}^\pi \Re[-g(\alpha, x)] |u(x)|^2 dx \\
 &= \frac{\sqrt{2}(\kappa_{\max} - \kappa_{\min})}{\zeta_\alpha} \mathbf{u}^T G \mathbf{u}.
 \end{aligned}$$

Thus, the desired result can be obtained just by utilizing the following inequality:

$$\begin{aligned}
 \kappa \mathbf{u}^T G \mathbf{u} - \left| \mathbf{u}^T (-\tilde{D}G_\alpha - G_\alpha^T \tilde{D}) \mathbf{u} \right| &\leq \mathbf{u}^T (-DG_\alpha - G_\alpha^T D) \mathbf{u} \\
 &\leq \kappa \mathbf{u}^T G \mathbf{u} + \left| \mathbf{u}^T (-\tilde{D}G_\alpha - G_\alpha^T \tilde{D}) \mathbf{u} \right|.
 \end{aligned}$$

2.3 Toeplitz Preconditioner for the Discrete One-Dimensional Fractional Diffusion Equation

To solve (7) is equivalent to recursively solve the following linear systems:

$$\mathbf{A} \mathbf{u}^n = b^n, \quad n = 1, 2, \dots, N, \tag{18}$$

where $\mathbf{A} = I_M - \eta DG_\alpha$, $\eta = \tau / (2h^\alpha)$, I_k denotes the $k \times k$ identity matrix, $b^n = (I_M + \eta DG_\alpha) \cdot u^{n-1} + \tau \mathbf{f}^{n-\frac{1}{2}}$. As explained in the introduction section, a good preconditioner is required for the linear systems in (18).

For any $m \times m$ diagonal matrix, $\mathbf{C} = \text{diag}(c_1, c_2, \dots, c_m)$, denote $\text{mean}(\mathbf{C}) = \frac{1}{m} \sum_{i=1}^m c_i$. In this subsection, we propose a Toeplitz preconditioner for the linear systems in (18), such that

$$\mathbf{P} = I_M - \eta \bar{d} G_\alpha, \tag{19}$$

where $\bar{d} = \text{mean}(D)$. In the following, we discuss a computationally effective representation of \mathbf{P}^{-1} , which allows fast matrix–vector multiplication of \mathbf{P}^{-1} .

Let $\mathbf{v} = (v_1, v_2, \dots, v_M)^T$ and $\tilde{\mathbf{v}} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_M)^T$ be solutions of following linear systems:

$$\mathbf{P} \mathbf{v} = \mathbf{e}_1 \equiv (1, 0, 0, \dots, 0)^T, \quad \mathbf{P} \tilde{\mathbf{v}} = \mathbf{e}_M \equiv (0, 0, \dots, 0, 1)^T. \tag{20}$$

According to the Gohberg–Semencul-type formula [8], \mathbf{P}^{-1} can be expressed as follows:

$$\mathbf{P}^{-1} = \frac{1}{2v_1}(\mathbf{S}_1\mathbf{C}_1 - \mathbf{S}_2\mathbf{C}_2), \tag{21}$$

where $\mathbf{S}_1, \mathbf{S}_2$ are skew-circulant matrices with $\mathbf{v}, \bar{\mathbf{v}} = (-\tilde{v}_M, \tilde{v}_1, \dots, \tilde{v}_{M-1})^T$ as their first columns, respectively; $\mathbf{C}_1, \mathbf{C}_2$ are circulant matrices with $\hat{\mathbf{v}} = (\tilde{v}_M, \tilde{v}_1, \dots, \tilde{v}_{M-1})^T, \mathbf{v}$ as their first columns, respectively. From (20), we see that v_1 is the first diagonal entry of \mathbf{P}^{-1} . From Lemma 2.1, we see that $\mathbf{P} + \mathbf{P}^T$ is positive definite. Thus,

$$v_1 = \mathbf{e}_1^T \mathbf{P}^{-1} \mathbf{e}_1 = \frac{1}{2} \mathbf{e}_1^T (\mathbf{P}^{-1} + \mathbf{P}^{-T}) \mathbf{e}_1 = \frac{1}{2} \mathbf{e}_1^T \mathbf{P}^{-1} (\mathbf{P} + \mathbf{P}^T) \mathbf{P}^{-T} \mathbf{e}_1 > 0,$$

which means that (21) is applicable. Moreover, the Toeplitz linear systems in (20) can be efficiently solved by the super fast direct solver proposed in [7].

For $\mathbf{C} \in \mathbb{C}^{m \times n}$, denote by $\Sigma(\mathbf{C})$, the set of singular values of \mathbf{C} . Also denote $\Sigma^2(\mathbf{C}) = \{\lambda^2 | \lambda \in \Sigma(\mathbf{C})\}$. For any matrix $\mathbf{C} \in \mathbb{C}^{m \times m}$, denote by $\sigma(\mathbf{C})$, the spectrum of \mathbf{C} . For a number λ , denote by $\Re(\lambda)$, the real part of λ .

For any invertible matrix $\mathbf{C} \in \mathbb{C}^{m \times m}$, define its condition number as

$$\text{cond}(\mathbf{C}) \triangleq \|\mathbf{C}\|_2 \|\mathbf{C}^{-1}\|_2.$$

Lemma 2.6 (see [41, Lemma 2.7]) *For any $\mathbf{C} \in \mathbb{C}^{m \times m}$, it holds*

$$\{\Re(\lambda) | \lambda \in \sigma(\mathbf{C})\} \subset \left[\min_{z \in \sigma((\mathbf{C} + \mathbf{C}^*)/2)} z, \max_{z \in \sigma((\mathbf{C} + \mathbf{C}^*)/2)} z \right].$$

As a preconditioner, the invertibility is essential.

Proposition 2.1 *\mathbf{P} is invertible for any $\alpha \in (1, 2)$.*

Proof By Lemma 2.1, it is easy to see that $\mathbf{P} + \mathbf{P}^T$ is positive definite and thus has positive eigenvalues. From Lemma 2.6, we see that $\{\Re(\lambda) | \lambda \in \sigma(\mathbf{P})\} \subset (0, +\infty)$. Therefore, \mathbf{P} is invertible.

For any Hermitian matrices $\mathbf{H}_1, \mathbf{H}_2 \in \mathbb{C}^{m \times m}$, denote $\mathbf{H}_1 < \mathbf{H}_2$ or $\mathbf{H}_2 > \mathbf{H}_1$ if $\mathbf{H}_2 - \mathbf{H}_1$ is Hermitian positive definite. Especially, we denote $\mathbf{O} < \mathbf{H}_1$ or $\mathbf{H}_1 > \mathbf{O}$, when \mathbf{H}_1 itself is Hermitian positive definite. Also, we use $\mathbf{H}_1 \leq \mathbf{H}_2$ or $\mathbf{H}_2 \geq \mathbf{H}_1$ to denote a Hermitian positive semi-definite $\mathbf{H}_2 - \mathbf{H}_1$ and use $\mathbf{O} \leq \mathbf{H}_1$ or $\mathbf{H}_1 \geq \mathbf{O}$ to denote a Hermitian positive semi-definite \mathbf{H}_1 .

Next, we are to estimate the condition number of the preconditioned matrix $\mathbf{A}\mathbf{P}^{-1}$.

Proposition 2.2 *For positive numbers $\xi_i, \zeta_i (1 \leq i \leq m)$, it obviously holds that*

$$\min_{1 \leq i \leq m} \frac{\xi_i}{\zeta_i} \leq \left(\sum_{i=1}^m \zeta_i \right)^{-1} \left(\sum_{i=1}^m \xi_i \right) \leq \max_{1 \leq i \leq m} \frac{\xi_i}{\zeta_i}.$$

Theorem 2.4 *Assume*

- (i) for any $x \in (x_L, x_R)$, $d(x) \in [\kappa_{\min}, \kappa_{\max}]$ for positive constants κ_{\min} and κ_{\max} ,
- (ii) $\kappa_{\max} - \nu_\alpha > 0$, with $\nu_\alpha = \sqrt{2}(\kappa_{\max} - \kappa_{\min})/\zeta_\alpha$,
- (iii) $d(x)$ is concave.

Then, for any $N \geq 1$, any $M \geq 1$, $\Sigma^2(\mathbf{A}\mathbf{P}^{-1}) \subset [\check{s}, \hat{s}]$ and thus

$$\sup_{N, M \geq 1} \text{cond}(\mathbf{A}\mathbf{P}^{-1}) \leq \sqrt{\hat{s}/\check{s}},$$

where \check{s} and \hat{s} are positive constants independent of τ, h , and given by

$$\check{s} = \min \left\{ \frac{\kappa_{\max} - \nu_\alpha}{\kappa_{\max}}, \frac{\kappa_{\min}^2}{\kappa_{\max}^2} \right\}, \quad \hat{s} = \max \left\{ \frac{\kappa + \nu_\alpha}{\kappa_{\min}}, \frac{\kappa_{\max}^2}{\kappa_{\min}^2} \right\}.$$

Proof By straightforward calculation:

$$\begin{aligned} \mathbf{A}^T \mathbf{A} &= I_M - \eta(G_\alpha^T D + D G_\alpha) + \eta^2 G_\alpha^T D^2 G_\alpha, \\ \mathbf{P}^T \mathbf{P} &= I_M + \eta \bar{d} G + \eta^2 \bar{d}^2 G_\alpha^T G_\alpha. \end{aligned}$$

By Theorem 2.3, we see that

$$\begin{aligned} \mathbf{0} &< I_M + (\kappa_{\max} - \nu_\alpha)\eta G + \kappa_{\min}^2 \eta^2 G_\alpha^T G_\alpha \\ &\leq \mathbf{A}^T \mathbf{A} \leq I_M + (\kappa_{\max} + \nu_\alpha)\eta G + \kappa_{\max}^2 \eta^2 G_\alpha^T G_\alpha. \end{aligned} \tag{22}$$

For any non-zero vector $\mathbf{y} \in \mathbb{R}^{M \times 1}$, denote $\mathbf{z} = \mathbf{P}^{-1}\mathbf{y}$. Then, it holds

$$\frac{\mathbf{y}^T (\mathbf{A}\mathbf{P}^{-1})^T (\mathbf{A}\mathbf{P}^{-1}) \mathbf{y}}{\mathbf{y}^T \mathbf{y}} = \frac{\mathbf{z}^T \mathbf{A}^T \mathbf{A} \mathbf{z}}{\mathbf{z}^T \mathbf{P}^T \mathbf{P} \mathbf{z}}.$$

By (22),

$$\begin{aligned} &\frac{\mathbf{z}^T [I_M + (\kappa_{\max} - \nu_\alpha)\eta G + \kappa_{\min}^2 \eta^2 G_\alpha^T G_\alpha] \mathbf{z}}{\mathbf{z}^T (I_M + \eta \bar{d} G + \eta^2 \bar{d}^2 G_\alpha^T G_\alpha) \mathbf{z}} \\ &\leq \frac{\mathbf{z}^T \mathbf{A}^T \mathbf{A} \mathbf{z}}{\mathbf{z}^T \mathbf{P}^T \mathbf{P} \mathbf{z}} \leq \frac{\mathbf{z}^T [I_M + (\kappa_{\max} + \nu_\alpha)\eta G + \kappa_{\max}^2 \eta^2 G_\alpha^T G_\alpha] \mathbf{z}}{\mathbf{z}^T (I_M + \eta \bar{d} G + \eta^2 \bar{d}^2 G_\alpha^T G_\alpha) \mathbf{z}}. \end{aligned} \tag{23}$$

By Proposition 2.2 and (23),

$$\check{s} = \min \left\{ 1, \frac{\kappa_{\max} - \nu_\alpha}{\kappa_{\max}}, \frac{\kappa_{\min}^2}{\kappa_{\max}^2} \right\} \leq \frac{\mathbf{z}^T \mathbf{A}^T \mathbf{A} \mathbf{z}}{\mathbf{z}^T \mathbf{P}^T \mathbf{P} \mathbf{z}} \leq \max \left\{ 1, \frac{\kappa + \nu_\alpha}{\kappa_{\min}}, \frac{\kappa_{\max}^2}{\kappa_{\min}^2} \right\} = \hat{s}.$$

During the proof above, there is no constraint on M and N . Thus, for any $N \geq 1$, any $M \geq 1$, $\Sigma^2(\mathbf{A}\mathbf{P}^{-1}) \subset [\check{s}, \hat{s}]$ and $\sup_{N, M \geq 1} \text{cond}(\mathbf{A}\mathbf{P}^{-1}) \leq \sqrt{\hat{s}/\check{s}}$.

Similar to proof of Theorem 2.4, one can prove the following theorem.

Theorem 2.5 *Assume*

- (i) for any $x \in (x_L, x_R)$, $d(x) \in [\kappa_{\min}, \kappa_{\max}]$ for positive constants κ_{\min} and κ_{\max} ,
- (ii) $\kappa_{\min} - \nu_\alpha > 0$, with $\nu_\alpha = \sqrt{2(\kappa_{\max} - \kappa_{\min})}/\zeta_\alpha$
- (iii) $d(x)$ is convex.

Then, for any $N \geq 1$, any $M \geq 1$, $\Sigma^2(\mathbf{AP}^{-1}) \subset [\check{s}, \hat{s}]$ and thus

$$\sup_{N, M \geq 1} \text{cond}(\mathbf{AP}^{-1}) \leq \sqrt{\hat{s}/\check{s}},$$

where \check{s} and \hat{s} are positive constants independent of τ, h and given by

$$\check{s} = \min \left\{ \frac{\kappa_{\min} - \nu_\alpha}{\kappa_{\max}}, \frac{\kappa_{\min}^2}{\kappa_{\max}^2} \right\}, \quad \hat{s} = \max \left\{ \frac{\kappa_{\min} + \nu_\alpha}{\kappa_{\min}}, \frac{\kappa_{\max}^2}{\kappa_{\min}^2} \right\}.$$

Remark 2.1 Theorems 2.4–2.5 show that $\text{cond}(\mathbf{AP}^{-1})$ has an upper bound independent of τ and h under certain assumptions on the coefficient function d . Thus, the Krylov subspace method for such preconditioned linear systems converges linearly and independently on the discretization step-sizes.

3 Extension to 2-D OSFDE

In this section, we study the following 2-D OSFDE [39]. For ease of statement, we set $u|_{\partial\Omega} \equiv 0$ in (24), although $u(x, y, t)$ could be non-zero for $t \in (0, T]$ and $(x, y) \in (\{x_R\} \times (y_L, y_R)) \cup ((x_L, x_R) \times \{y_R\})$,

$$\begin{cases} \frac{\partial u(x, y, t)}{\partial t} = d(x, y) {}_{x_L}D_x^\alpha u(x, y, t) + e(x, y) {}_{y_L}D_y^\beta u(x, y, t) + f(x, y, t), \\ \quad (x, y) \in \Omega, t \in (0, T], \\ u(x, y, t) = 0, \quad (x, y) \in \partial\Omega \\ u(x, y, 0) = \varphi(x, y), \quad (x, y) \in \bar{\Omega}, \end{cases} \tag{24}$$

where $\alpha, \beta \in (1, 2)$, $\Omega = (x_L, x_R) \times (y_L, y_R)$, $\partial\Omega$ denotes the boundary of Ω and $d(x, y)$, $e(x, y)$ are nonnegative functions, and ${}_{x_L}D_x^\alpha u(x, y, t)$ denotes the α -order RL derivative with respect to the x direction defined as

$${}_{x_L}D_x^\alpha u(x, y, t) = \frac{1}{\Gamma(2 - \alpha)} \frac{\partial^2}{\partial x^2} \int_{x_L}^x \frac{u(\xi, y, t)}{(x - \xi)^{\alpha-1}} d\xi,$$

${}_{y_L}D_y^\beta u(x, y, t)$ can be defined in a similar way.

To state a finite-difference scheme for (24), we need more notations. Let $\tau = T/N$, $h_1 = \frac{x_R - x_L}{M_1 + 1}$, $h_2 = \frac{y_R - y_L}{M_2 + 1}$, where M_1, M_2 , and N are some positive integers. For $i = 0, 1, \dots, M_1 + 1$, $j = 0, 1, \dots, M_2 + 1$ and $n = 0, 1, \dots, N$, denote $x_i = ih_1$, $y_j = jh_2$, and $t_n = n\tau$. Denote $t_{n-\frac{1}{2}} = \frac{t_n + t_{n-1}}{2}$ for $n = 1, 2, \dots, N$. Let $\bar{\Omega}_h = \{(x_i, y_j) | 0 \leq i \leq M_1 + 1, 0 \leq j \leq M_2 + 1\}$, $\Omega_h = \bar{\Omega}_h \cap \Omega$, $\partial\Omega_h = \bar{\Omega}_h \cap \partial\Omega$. Furthermore, denote $d_{i,j} = d(x_i, y_j)$, $e_{i,j} = e(x_i, y_j)$, $f_{i,j}^{n-\frac{1}{2}} = f(x_i, y_j, t_{n-\frac{1}{2}})$, and $\varphi_{i,j} = \varphi(x_i, y_j)$, and let $u_{i,j}^n$ be the numerical approxi-

mation of $u(x_i, y_j, t_n)$. Then, in a similar way with the one-dimensional case, we can derive the Crank–Nicolson scheme for the 2-D problem (24) as follows:

$$\begin{aligned} \frac{u_{i,j}^n - u_{i,j}^{n-1}}{\tau} &= \frac{1}{2h^\alpha} d_{i,j} \sum_{k=0}^i w_k^{(\alpha)} \left(u_{i-k+1,j}^{n-1} + u_{i-k+1,j}^n \right) \\ &+ \frac{1}{2h^\beta} e_{i,j} \sum_{k=0}^j w_k^{(\beta)} \left(u_{i,j-k+1}^{n-1} + u_{i,j-k+1}^n \right) + f_{i,j}^{n-\frac{1}{2}} + \hat{R}_{i,j}^{n-\frac{1}{2}}, \end{aligned} \tag{25}$$

$$1 \leq i \leq M_1, 1 \leq j \leq M_2, 1 \leq n \leq N,$$

where $\hat{R}_{i,j}^{n-\frac{1}{2}} \leq c_3(\tau^2 + h_1^2 + h_2^2)$ for a positive constant c_3 .

Take

$$\begin{aligned} u^n &= [u_{1,1}^n, u_{2,1}^n, \dots, u_{M_1,1}^n, u_{1,2}^n, \dots, u_{M_1,2}^n, \dots, u_{1,M_2}^n, \dots, u_{M_1,M_2}^n]^T, \\ f^{n-\frac{1}{2}} &= [f_{1,1}^{n-\frac{1}{2}}, f_{2,1}^{n-\frac{1}{2}}, \dots, f_{M_1,1}^{n-\frac{1}{2}}, f_{1,2}^{n-\frac{1}{2}}, \dots, f_{M_1,2}^{n-\frac{1}{2}}, \dots, f_{1,M_2}^{n-\frac{1}{2}}, \dots, f_{M_1,M_2}^{n-\frac{1}{2}}]^T, \\ D &= \text{diag}(d_{1,1}, d_{2,1}, \dots, d_{M_1,1}, d_{1,2}, \dots, d_{M_1,2}, \dots, d_{M_1,M_2}), \\ E &= \text{diag}(e_{1,1}, e_{2,1}, \dots, e_{M_1,1}, e_{1,2}, \dots, e_{M_1,2}, \dots, e_{M_1,M_2}). \end{aligned}$$

Omitting the small term $\hat{R}_{i,j}^{n-\frac{1}{2}}$ in (25), the finite-difference scheme in the matrix form for (24) can be given as

$$\begin{aligned} \frac{1}{\tau} (u^n - u^{n-1}) &= \left(\frac{1}{2h_1^\alpha} D(I \otimes G_\alpha) + \frac{1}{2h_2^\beta} E(G_\beta \otimes I) \right) (u^{n-1} + u^n) + f^{n-\frac{1}{2}}, \\ 1 \leq n \leq N, \end{aligned} \tag{26}$$

where I is the identity matrix, the symbol “ \otimes ” denotes the Kronecker product, and G_β has the similar definition to G_α .

3.1 Stability and Convergence of the 2-D Problem

To discuss the stability and convergence of the scheme (26), we denote

$$A = \frac{1}{2h_1^\alpha} D(I \otimes G_\alpha) + \frac{1}{2h_2^\beta} E(G_\beta \otimes I),$$

and introduce a set

$$\mathfrak{D} = \{X | X > \mathbf{O}, -\mathcal{H}(XA) \geq \mathbf{O}, \text{cond}(X) \leq c \text{ for } c \text{ independent of } \tau, h_1 \text{ and } h_2\}.$$

Now, we present the stability of the scheme (26).

Theorem 3.1 *For any $Q \in \mathfrak{D}$, the finite-difference scheme (26) is unconditionally stable and its solution satisfies the following estimate:*

$$\|u^n\|_Q^2 \leq \exp(2T) \|\varphi\|_Q^2 + [\exp(2T) - 1] \max_{1 \leq k \leq n} \|f^{k-\frac{1}{2}}\|_Q^2, \quad n = 1, 2, \dots, N,$$

where $\|\cdot\|_Q$ is defined as $\|v\|_Q^2 := hv^T Qv$.

Proof Multiplying $h(u^{n-1} + u^n)^T Q$ on the both sides of (26), we get

$$\begin{aligned} & \frac{1}{\tau} h(u^{n-1} + u^n)^T Q(u^n - u^{n-1}) \\ & = h(u^{n-1} + u^n)^T QA(u^{n-1} + u^n) + h(u^{n-1} + u^n)^T Qf^{n-\frac{1}{2}}. \end{aligned}$$

Since $\mathcal{H}(QA)$ is negative semi-definite, we have

$$h(u^{n-1} + u^n)^T QA(u^{n-1} + u^n) = h(u^{n-1} + u^n)^T \mathcal{H}(QA)(u^{n-1} + u^n) \leq 0.$$

Then, it follows

$$h(u^n)^T Qu^n - h(u^{n-1})^T Qu^{n-1} \leq \tau h(u^n)^T Qf^{n-\frac{1}{2}} + \tau h(u^{n-1})^T Qf^{n-\frac{1}{2}}.$$

The rest of the proof is similar to that in Theorem 2.1.

With Theorem 3.1, the convergence of the scheme (26) can be directly obtained:

Theorem 3.2 *Let $u(x_i, y_j, t_n)$ be the exact solution of (24) and smooth enough, $u^n_{i,j}$ be the solution of the finite-difference scheme (26). Denote $e^n_{i,j} = u(x_i, y_j, t_n) - u^n_{i,j}$, $0 \leq i \leq M_1 + 1$, $0 \leq j \leq M_2 + 1$, $0 \leq n \leq N$. For any $Q \in \mathfrak{D}$, there exists a positive constant c_4 , such that*

$$\|e^n\| \leq c_4(\tau^2 + h_1^2 + h_2^2).$$

The remaining and important thing is to give the feature of the set \mathfrak{D} . However, it seems difficult to depict all the elements of \mathfrak{D} . In the following Corollaries 3.1 and 3.2, we show that there are some matrices belong to \mathfrak{D} when the variable coefficients $d(x, y)$, $e(x, y)$ satisfy some certain conditions, and this ensures that \mathfrak{D} is not an empty set which is necessary for the stability and convergence. We discuss the existence of those matrices in two cases.

- **Case 1** When $d(x, y)$, $e(x, y)$ are separable respect to x and y .

In this case, we denote $d(x, y) = \tilde{d}(x)\hat{d}(y)$ and $e(x, y) = \tilde{e}(x)\hat{e}(y)$, and take

$$\begin{aligned} \tilde{D} &= \text{diag}(\tilde{d}_1, \tilde{d}_2, \dots, \tilde{d}_{M_1}), \quad \hat{D} = (\hat{d}_1, \hat{d}_2, \dots, \hat{d}_{M_2}), \\ \tilde{E} &= \text{diag}(\tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_{M_1}), \quad \hat{E} = (\hat{e}_1, \hat{e}_2, \dots, \hat{e}_{M_2}). \end{aligned}$$

Then, $D = \hat{D} \otimes \tilde{D}$, $E = \hat{E} \otimes \tilde{E}$.

Corollary 3.1 *If $\tilde{d}_- \leq \tilde{d}(x) \leq \tilde{d}_+$ and $\hat{e}_- \leq \hat{e}(y) \leq \hat{e}_+$ for some positive constants $\tilde{d}_-, \tilde{d}_+, \hat{e}_-$ and \hat{e}_+ , then $\hat{E}^{-1} \otimes \tilde{D}^{-1} \in \mathfrak{D}$.*

Proof We have $A = \frac{1}{2h_1^\alpha}(\hat{D} \otimes \tilde{D}G_\alpha) + \frac{1}{2h_2^\beta}(\hat{E}G_\beta \otimes \tilde{E})$, then

$$\mathcal{H}((\hat{E}^{-1} \otimes \tilde{D}^{-1})A) = \frac{1}{4h_1^\alpha}(\hat{E}^{-1}\hat{D} \otimes (G_\alpha + G_\alpha^T)) + \frac{1}{4h_2^\beta}((G_\beta + G_\beta^T) \otimes \tilde{D}^{-1}\tilde{E}),$$

which is negative semi-definite. Thus $\hat{E}^{-1} \otimes \tilde{D}^{-1} \in \mathfrak{D}$.

- **Case 2** When $d(x, y)$ and $e(x, y)$ are non-separable.

As in Lemma 2.4, we denote

$$\varsigma_\beta \triangleq \min_x \frac{\Re[-g(\beta, x)]}{|g(\beta, x)|} = \left| \cos \left(\frac{\beta}{2} \pi \right) \right|,$$

where $g(\beta, x)$ is the generating function of the matrix G_β .

Corollary 3.2

- (i) Assume that

$$\begin{aligned} 0 \leq \kappa_{\min}^d(y) \leq d(x, y) \leq \kappa_{\max}^d(y) < \infty \text{ for every } (x, y), \\ 0 \leq \kappa_{\min}^e(x) \leq e(x, y) \leq \kappa_{\max}^e(x) < \infty \text{ for every } (x, y). \end{aligned}$$

Then, $I \in \mathfrak{D}$ if the following conditions are fulfilled:

$$\kappa^d(y) - \frac{\sqrt{2}(\kappa_{\max}^d(y) - \kappa_{\min}^d(y))}{\varsigma_\alpha} \geq 0 \text{ for every } y, \tag{27}$$

$$\kappa^e(x) - \frac{\sqrt{2}(\kappa_{\max}^e(x) - \kappa_{\min}^e(x))}{\varsigma_\beta} \geq 0 \text{ for every } x, \tag{28}$$

where $\kappa^d(y) = \kappa_{\max}^d(y)$ when $d(x, y)$ is a concave function of x , $\kappa^d(y) = \kappa_{\min}^d(y)$ when $d(x, y)$ is a convex function of x , $\kappa^e(x) = \kappa_{\max}^e(x)$ when $e(x, y)$ is a concave function of y , $\kappa^e(x) = \kappa_{\min}^e(x)$ when $e(x, y)$ is a convex function of y .

- (ii) Assume that

$$0 \leq \kappa_{\min}^{e'}(x) \leq \frac{e(x, y)}{d(x, y)} \leq \kappa_{\max}^{e'}(x) < \infty \text{ with } 0 < d(x, y) < \infty.$$

Then, $D^{-1} \in \mathfrak{D}$ if the following condition is fulfilled:

$$\kappa^{e'}(x) - \frac{\sqrt{2}(\kappa_{\max}^{e'}(x) - \kappa_{\min}^{e'}(x))}{\varsigma_\beta} \geq 0,$$

where $\kappa^{e'}(x) = \kappa_{\max}^{e'}(x)$ when $\frac{e(x, y)}{d(x, y)}$ is a concave function of y , and $\kappa^{e'}(x) = \kappa_{\min}^{e'}(x)$ when $\frac{e(x, y)}{d(x, y)}$ is a convex function of y .

- (iii) Assume that

$$0 \leq \kappa_{\min}^{d'}(y) \leq \frac{d(x, y)}{e(x, y)} \leq \kappa_{\max}^{d'}(y) < \infty \text{ with } 0 < e(x, y) < \infty.$$

Then, $E^{-1} \in \mathfrak{D}$ if the following condition is fulfilled:

$$\kappa^{d'}(y) - \frac{\sqrt{2}(\kappa_{\max}^{d'}(y) - \kappa_{\min}^{d'}(y))}{\varsigma_\alpha} \geq 0,$$

where $\kappa^{d'}(y) = \kappa_{\max}^{d'}(y)$ when $\frac{d(x,y)}{e(x,y)}$ is a concave function of x , and $\kappa^{d'}(y) = \kappa_{\min}^{d'}(y)$ when $\frac{d(x,y)}{e(x,y)}$ is a convex function of x .

Proof We first prove (i). Denote

$$K^d = \text{diag}(k^d(y_1), k^d(y_2), \dots, k^d(y_{M_2})), \quad K^e = \text{diag}(k^e(x_1), k^e(x_2), \dots, k^e(x_{M_1})).$$

Take $\tilde{D} = D - K^d \otimes I$ and $\tilde{E} = E - I \otimes K^e$. Then,

$$A = \frac{1}{2h^\alpha} [(K^d \otimes G_\alpha) + \tilde{D}(I \otimes G_\alpha)] + \frac{1}{2h^\beta} [(G_\beta \otimes K^e) + \tilde{E}(G_\beta \otimes I)].$$

For any $\mathbf{u} = [u_{1,1}, u_{2,1}, \dots, u_{M_1,1}, u_{1,2}, \dots, u_{M_1,2}, \dots, u_{1,M_2}, \dots, u_{M_1,M_2}]^T$, we have

$$\begin{aligned} 2\mathbf{u}^T \mathcal{H}(A)\mathbf{u} &= \frac{1}{2h^\alpha} [\mathbf{u}^T (K^d \otimes (G_\alpha + G_\alpha^T))\mathbf{u} + \mathbf{u}^T (\tilde{D}(I \otimes G_\alpha) + (I \otimes G_\alpha^T)\tilde{D})\mathbf{u}] \\ &\quad + \frac{1}{2h^\beta} [\mathbf{u}^T ((G_\beta + G_\beta^T) \otimes K^e)\mathbf{u} + \mathbf{u}^T (\tilde{E}(G_\beta \otimes I) + (G_\beta^T \otimes I)\tilde{E})\mathbf{u}]. \end{aligned}$$

Referring to the proof of Theorem 2.3, it is easy to obtain

$$\begin{aligned} \left| \mathbf{u}^T (\tilde{D}(I \otimes G_\alpha) + (I \otimes G_\alpha^T)\tilde{D})\mathbf{u} \right| &\leq \frac{-\sqrt{2}}{\varsigma_\alpha} \mathbf{u}^T (K_\alpha \otimes (G_\alpha + G_\alpha^T))\mathbf{u}, \\ \left| \mathbf{u}^T (\tilde{E}(G_\beta \otimes I) + (G_\beta^T \otimes I)\tilde{E})\mathbf{u} \right| &\leq \frac{-\sqrt{2}}{\varsigma_\beta} \mathbf{u}^T ((G_\beta + G_\beta^T) \otimes K_\beta)\mathbf{u}, \end{aligned}$$

where

$$\begin{aligned} K_\alpha &= \text{diag}(\kappa_{\max}^d(y_1) - \kappa_{\min}^d(y_1), \kappa_{\max}^d(y_2) - \kappa_{\min}^d(y_2), \dots, \kappa_{\max}^d(y_{M_2}) - \kappa_{\min}^d(y_{M_2})), \\ K_\beta &= \text{diag}(\kappa_{\max}^e(x_1) - \kappa_{\min}^e(x_1), \kappa_{\max}^e(x_2) - \kappa_{\min}^e(x_2), \dots, \kappa_{\max}^e(x_{M_1}) - \kappa_{\min}^e(x_{M_1})). \end{aligned}$$

Therefore,

$$\begin{aligned} -2\mathbf{u}^T \mathcal{H}(A)\mathbf{u} &\geq \frac{1}{2h^\alpha} \mathbf{u}^T \left(\left(K^d - \frac{\sqrt{2}}{\varsigma_\alpha} K_\alpha \right) \otimes (-G_\alpha - G_\alpha^T) \right) \mathbf{u} \\ &\quad + \frac{1}{2h^\beta} \mathbf{u}^T \left((-G_\beta - G_\beta^T) \otimes \left(K^e - \frac{\sqrt{2}}{\varsigma_\beta} K_\beta \right) \right) \mathbf{u}, \end{aligned}$$

which implies that $\mathcal{H}(A)$ is negative semi-definite if the conditions in (27)–(28) hold. Hence, $I \in \mathfrak{D}$.

Similarly, one can show (ii) and (iii).

3.2 The Two-Dimensional Toeplitz Preconditioner

In this subsection, we extend the Toeplitz preconditioner to the 2-D case. To solve (26), it is equivalent to solve the following N linear systems:

$$\mathbf{A}u^n = b^n, \quad n = 1, 2, \dots, N, \tag{29}$$

where I_k denotes the $k \times k$ identity, $\mathbf{A} = I_{\hat{M}} + DB_x + EB_y$, $B_x = -\eta_x(I_{M_2} \otimes G_\alpha)$, $B_y = -\eta_y(G_\beta \otimes I_{M_1})$, $\hat{M} = M_1M_2$, $\eta_x = \tau/(2h_1^\alpha)$, $\eta_y = \tau/(2h_2^\beta)$, $\mathbf{b}^n = (I_{\hat{M}} - DB_x - EB_y) \cdot u^{n-1} + \tau \mathbf{f}^{n-\frac{1}{2}}$. Our two-level Toeplitz preconditioner for preconditioning (29) is defined as follows:

$$\mathbf{P} = I_{\hat{M}} + \bar{d}B_x + \bar{e}B_y, \tag{30}$$

where $\bar{d} = \text{mean}(D)$, $\bar{e} = \text{mean}(E)$. The preconditioned Krylov subspace method with the preconditioner \mathbf{P} is employed to solve the linear systems in (29). Hence, in each iteration, it requires to compute some matrix–vector multiplications like $\mathbf{P}^{-1}\mathbf{z}$ for some randomly given \mathbf{z} , i.e., it requires to solve the linear system of the form:

$$\mathbf{P}\mathbf{x} = \mathbf{z}. \tag{31}$$

Next, we introduce a multigrid method to solve (31).

For the choices of coarse-grid matrices, interpolation, and restriction, we refer to the geometric grid coarsening, piecewise linear interpolation, and its transpose. For the choice of pre-smoothing iteration, we refer to the block Jacobi iteration, that is

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{T}_x^{-1}(\mathbf{z} - \mathbf{P}\mathbf{x}^k), \tag{32}$$

where $\mathbf{T}_x = I_{\hat{M}} + \bar{d}B_x$ is the block diagonal part of \mathbf{P} , and \mathbf{x}^k is an initial guess of \mathbf{x} in (31). Since \mathbf{T}_x is a block diagonal matrix with identical Toeplitz blocks, its inversion, \mathbf{T}_x^{-1} can be computed efficiently with the help of Gohberg–Semencul-type formula as discussed in Sect. 2. For the choice of post-smoother, we refer to the block Jacobi iteration for the permuted linear system, that is

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{T}_y^{-1}(\mathbf{z} - \mathbf{P}\mathbf{x}^k), \tag{33}$$

where $\mathbf{T}_y = I_{\hat{M}} + \bar{e}B_y$, \mathbf{x}^k is an initial guess of \mathbf{x} in (31). One can easily find a x - y ordering permutation matrix \mathbf{Q} , such that

$$\mathbf{T}_y = \mathbf{Q}^T(I_{\hat{M}} - \bar{e}\eta_y I_{M_1} \otimes G_\beta)\mathbf{Q}. \tag{34}$$

Thus, $\mathbf{T}_y^{-1} = \mathbf{Q}^T(I_{\hat{M}} - \bar{e}\eta_y I_{M_1} \otimes G_\beta)^{-1}\mathbf{Q}$, which means that the implementation of (33) still requires to compute an inversion of a block diagonal matrix with identical Toeplitz blocks. Therefore, (33) can still be fast implemented using the Gohberg–Semencul-type formula. Similar to proof of Proposition 2.1, one can prove the following proposition.

Proposition 3.1 \mathbf{P} defined in (30) is invertible for any $\alpha \in (1, 2)$.

Theorem 3.3 Let $d(x, y) \equiv v_1 a(x, y)$ and $e(x, y) \equiv v_2 a(x, y)$ for any $(x, y) \in \Omega$ with nonnegative constants v_1 and v_2 . Assume

- (i) $a(x, y) \in [\check{a}, \hat{a}]$ with $\check{a} > 0$ for any $(x, y) \in \Omega$;
- (ii) for any $x \in (x_L, x_R)$, $a(x, \cdot)$ is convex or concave on $y \in (y_L, y_R)$; and for any $y \in (y_L, y_R)$, $a(\cdot, y)$ is convex or concave on $x \in (x_L, x_R)$;
- (iii) $\check{c}_1 = \inf_{y \in (y_L, y_R)} [\tilde{\mathcal{M}}_1(y) - \sqrt{2}(\hat{\mathcal{M}}_1(y) - \check{\mathcal{M}}_1(y))/\varsigma_\alpha] > 0$ with $\hat{\mathcal{M}}_1(y) =: \sup_{x \in (x_L, x_R)} a(x, y)$ and $\check{\mathcal{M}}_1(y) =: \inf_{x \in (x_L, x_R)} a(x, y)$,

$$\tilde{\mathcal{M}}_1(y) = \begin{cases} \hat{\mathcal{M}}_1(y) & \text{if } a(\cdot, y) \text{ is convex,} \\ \check{\mathcal{M}}_1(y) & \text{if } a(\cdot, y) \text{ is concave,} \end{cases}$$

$$\check{c}_2 = \inf_{x \in (x_L, x_R)} [\tilde{\mathcal{M}}_2(x) - \sqrt{2}(\hat{\mathcal{M}}_2(x) - \check{\mathcal{M}}_2(x))/\varsigma_\beta] > 0 \text{ with } \hat{\mathcal{M}}_2(x) =: \sup_{y \in (y_L, y_R)} a(x, y) \text{ and } \check{\mathcal{M}}_2(x) =: \inf_{y \in (y_L, y_R)} a(x, y),$$

$$\tilde{\mathcal{M}}_2(x) = \begin{cases} \hat{\mathcal{M}}_2(x) & \text{if } a(x, \cdot) \text{ is convex,} \\ \check{\mathcal{M}}_2(x) & \text{if } a(x, \cdot) \text{ is concave.} \end{cases}$$

Then, for any positive integers, N, M_1 and M_2 , it holds $\Sigma^2(\mathbf{AP}^{-1}) \subset [\check{s}, \hat{s}]$ and thus

$$\sup_{M_1, M_2, N \geq 1} \text{cond}(\mathbf{AP}^{-1}) \leq \sqrt{\hat{s}/\check{s}},$$

where \check{s}, \hat{s} are positive constants independent of τ, h_1 , and h_2 ,

$$\check{s} = \min \left\{ \frac{\check{c}_1}{\hat{a}}, \frac{\check{c}_2}{\hat{a}}, \frac{\check{a}^2}{\hat{a}^2} \right\}, \hat{s} = \max \left\{ \frac{\hat{c}_1}{\check{a}}, \frac{\hat{c}_2}{\check{a}}, \frac{\hat{a}^2}{\check{a}^2} \right\},$$

$$\hat{c}_1 = \sup_{y \in (y_L, y_R)} \left[\tilde{\mathcal{M}}_1(y) + \frac{\sqrt{2}}{\varsigma_\alpha} (\hat{\mathcal{M}}_1(y) - \check{\mathcal{M}}_1(y)) \right],$$

$$\hat{c}_2 = \sup_{x \in (x_L, x_R)} \left[\tilde{\mathcal{M}}_2(x) + \frac{\sqrt{2}}{\varsigma_\beta} (\hat{\mathcal{M}}_2(x) - \check{\mathcal{M}}_2(x)) \right].$$

Proof Denote

$$D_a = \text{diag}(a_{1,1}, a_{2,1}, \dots, a_{M_1,1}, a_{1,2}, a_{2,2}, \dots, a_{M_1,2}, \dots, a_{1,M_2}, a_{2,M_2}, \dots, a_{M_1,M_2})$$

with $a_{i,j} = a(x_i, y_j)$. Also, denote $\bar{a} = \text{mean}(D_a)$. By straightforward calculation,

$$\mathbf{A}^T \mathbf{A} = I_{\check{M}} + \nu_1 (B_x^T D_a + D_a B_x) + \nu_2 (B_y^T D_a + D_a B_y) + W^T D_a^2 W, \tag{35}$$

$$\mathbf{P}^T \mathbf{P} = I_{\check{M}} + \nu_1 \bar{a} (B_x^T + B_x) + \nu_2 \bar{a} (B_y^T + B_y) + \bar{a}^2 W^T W, \tag{36}$$

where $W = \nu_1 B_x + \nu_2 B_y$. Rewrite $D_a = \text{diag}(D_{a,1}, D_{a,2}, \dots, D_{a,M_2})$ with $D_{a,i} = \text{diag}(a_{1,i}, a_{2,i}, \dots, a_{M_1,i})$. Then, it is easy to see that $B_x^T D_a + D_a B_x = \text{diag}(H_1, H_2, \dots, H_{M_2})$ with $H_i = -\eta_x (D_{a,i} G_\alpha + G_\alpha^T D_{a,i})$. Denote $l_1(y) = \tilde{\mathcal{M}}_1(y) - \sqrt{2}(\hat{\mathcal{M}}_1(y) - \check{\mathcal{M}}_1(y))/\varsigma_\alpha$ and $s_1(y) = \tilde{\mathcal{M}}_1(y) + \sqrt{2}(\hat{\mathcal{M}}_1(y) - \check{\mathcal{M}}_1(y))/\varsigma_\alpha$. Then, applying Theorem 2.3 to (i)–(iii), we have

$$-\check{c}_1 \eta_x (G_\alpha + G_\alpha^T) \leq l_1(y_i) \eta_x G \leq H_i \leq s_1(y_i) \eta_x G \leq -\hat{c}_1 \eta_x (G_\alpha + G_\alpha^T).$$

Therefore,

$$\mathbf{O} < \check{c}_1(B_x^T + B_x) \leq B_x^T D_a + D_a B_x \leq \hat{c}_1(B_x^T + B_x), \tag{37}$$

where the first “<” is obvious. Recall the permutation matrix defined in (34). Denote $\tilde{B}_y := \mathbf{Q}B_y\mathbf{Q}^T = -\eta_y I_{M_1} \otimes G_\beta$, $\tilde{D}_a = \text{diag}(\tilde{D}_{a,1}, \tilde{D}_{a,2}, \dots, \tilde{D}_{a,M_1})$ with $\tilde{D}_{a,i} = \text{diag}(a_{i,1}, a_{i,2}, \dots, a_{i,M_2})$. Then, it is easy to check that

$$B_y^T D_a + D_a B_y = \mathbf{Q}^T(\tilde{B}_y^T \tilde{D}_a + \tilde{D}_a \tilde{B}_y)\mathbf{Q}.$$

Similarly to proof of (37), applying Theorem 2.3 to (i), (ii), and (iii) yields

$$\begin{aligned} \mathbf{O} < \check{c}_2(B_y^T + B_y) &= \check{c}_2 \mathbf{Q}^T(\tilde{B}_y^T + \tilde{B}_y)\mathbf{Q} \leq B_y^T D_a + D_a B_y \leq \hat{c}_2 \mathbf{Q}^T(\tilde{B}_y^T + \tilde{B}_y)\mathbf{Q} \\ &= \hat{c}_2(B_y^T + B_y). \end{aligned} \tag{38}$$

Moreover, it is easy to see that

$$\check{a}^2 W^T W \leq W^T D_a^2 W \leq \hat{a}^2 W^T W. \tag{39}$$

By (37)–(39),

$$\begin{aligned} \mathbf{O} < I_{\hat{M}} + v_1 \check{c}_1(B_x^T + B_x) + v_2 \check{c}_2(B_y^T + B_y) + \check{a}^2 W^T W \\ &\leq \mathbf{A}^T \mathbf{A} \\ &\leq I_{\hat{M}} + v_1 \hat{c}_1(B_x^T + B_x) + v_2 \hat{c}_2(B_y^T + B_y) + \hat{a}^2 W^T W. \end{aligned} \tag{40}$$

For any non-zero vector $y \in \mathbb{R}^{M \times 1}$, denote $z = \mathbf{P}^{-1}y$. Then, it holds

$$\frac{y^T (\mathbf{A}\mathbf{P}^{-1})^T (\mathbf{A}\mathbf{P}^{-1}) y}{y^T y} = \frac{z^T \mathbf{A}^T \mathbf{A} z}{z^T \mathbf{P}^T \mathbf{P} z}.$$

By (40),

$$\begin{aligned} 0 < \frac{z^T [I_{\hat{M}} + v_1 \check{c}_1(B_x^T + B_x) + v_2 \check{c}_2(B_y^T + B_y) + \check{a}^2 W^T W] z}{z^T [I_{\hat{M}} + v_1 \bar{a}(B_x^T + B_x) + v_2 \bar{a}(B_y^T + B_y) + \bar{a}^2 W^T W] z} \\ &\leq \frac{z^T \mathbf{A}^T \mathbf{A} z}{z^T \mathbf{P}^T \mathbf{P} z} \\ &\leq \frac{z^T [I_{\hat{M}} + v_1 \hat{c}_1(B_x^T + B_x) + v_2 \hat{c}_2(B_y^T + B_y) + \hat{a}^2 W^T W] z}{z^T [I_{\hat{M}} + v_1 \bar{a}(B_x^T + B_x) + v_2 \bar{a}(B_y^T + B_y) + \bar{a}^2 W^T W] z}. \end{aligned} \tag{41}$$

By Proposition 2.2 and (41),

$$\check{s} \leq \min \left\{ \frac{\check{c}_1}{\bar{a}}, \frac{\check{c}_2}{\bar{a}}, \frac{\check{a}^2}{\bar{a}^2} \right\} \leq \frac{z^T \mathbf{A}^T \mathbf{A} z}{z^T \mathbf{P}^T \mathbf{P} z} \leq \max \left\{ \frac{\hat{c}_1}{\bar{a}}, \frac{\hat{c}_2}{\bar{a}}, \frac{\hat{a}^2}{\bar{a}^2} \right\} \leq \hat{s}.$$

During the proof above, there is no constraint on M and N . Thus, for any $N \geq 1$, any $M \geq 1$, $\Sigma^2(\mathbf{A}\mathbf{P}^{-1}) \subset [\check{s}, \hat{s}]$ and $\sup_{N, M \geq 1} \text{cond}(\mathbf{A}\mathbf{P}^{-1}) \leq \sqrt{\hat{s}/\check{s}}$.

4 Numerical Experiments

In this section, we test several examples to support theoretical results of Theorems 2.2, 3.2, and to show the efficiency of the Toeplitz preconditioner. We compare the proposed Toeplitz preconditioner with circulant preconditioners (two-level circulant preconditioner in the 2-D case) proposed in [16] and [14] and the Laplacian preconditioners proposed in [6, 26]. For fairness, the 2-D Laplacian preconditioner is implemented by the same multi-grid method as the one used for the implementation of the proposed 2-D Toeplitz preconditioner. We use **C** and **L** to denote the circulant preconditioner and Laplacian preconditioner, respectively. The preconditioned generalized minimal residual (PGMRES) method is employed to solve various preconditioned systems of (7) and (26). We denote the PGMRES method with the Toeplitz preconditioner, the Circulant preconditioner and the Laplacian preconditioner by PGMRES-T, PGMRES-C, and PGMRES-L, respectively. We note that these preconditioners are all used as right preconditioners in the implementation. The GMRES method employed in this paper is an un-restarted version with a maximal iteration number, 200.

The stopping criterion for PGMRES is set as $\frac{\|r_k\|_2}{\|r_0\|_2} \leq 1E-7$, where r_k denotes the residual vector at the k th iteration. All numerical experiments are performed via MATLAB R2018a on a PC with system information: Ubuntu 18.04.1 LTS 64-bit and configuration: Intel(R) Core(TM) i5-7 500 T CPU 2.70 GHz×4 15.6 GB RAM.

Recall that h is the spatial step-size for the one-dimensional discretization. We also set $h_1 = h_2 = h$ in the 2-D discretization for the related experiments in this section. Define the error as

$$E(h, \tau) = \max_{0 \leq n \leq N} \|e^n\|.$$

Denote by CPU, the running time by unit seconds. Denote by “iter”, the average of iteration numbers of the PGMRES method for the N linear systems in (7) or (26).

Example 4.1 Consider a one-dimensional OSFDE with $[x_L, x_R] = [0, 1]$, $T = 1$, and

$$d(x) = \cos(\pi x/2) + 0.1,$$

$$f(x, t) = 192x^3(1-x)^3t^2 - 2^6t^3d(x) \sum_{k=3}^6 \frac{\binom{3}{k-3} k! x^{k-\alpha}}{(-1)^{k-1} \Gamma(k+1-\alpha)}.$$

The explicit expression of the exact solution for the example is $u(x, t) = 2^6x^3(1-x)^3t^3$.

To show the convergence order of the proposed scheme on Example 4.1, we plot the values of $\ln(E(h, \tau))$ with different h and fixed τ in Fig. 1. As illustrated in Fig. 1, the values of $\ln(E(h, \tau))$ are distributed like a straight line with slope “2”, which demonstrates the second-order accuracy in space of the proposed scheme.

We test the three preconditioners on Example 4.1, the results of which are listed in Table 1. Overall, the choices of N , $N = 1$ typically maximize the condition number of the unpreconditioned matrix, which is difficult to fast solve. Hence, reporting the results of the three preconditioners in the case of $N = 1$ is convincing and representative for testing their performance, which is why we fix $N = 1$ in Table 1. Since $E(h, \tau)$ of the three solvers are almost the same, the results of $E(h, \tau)$ are skipped in Table 1. From Table 1, we see that the iteration numbers of PGMRES-T and PGMRES-C are stable with respect to M while

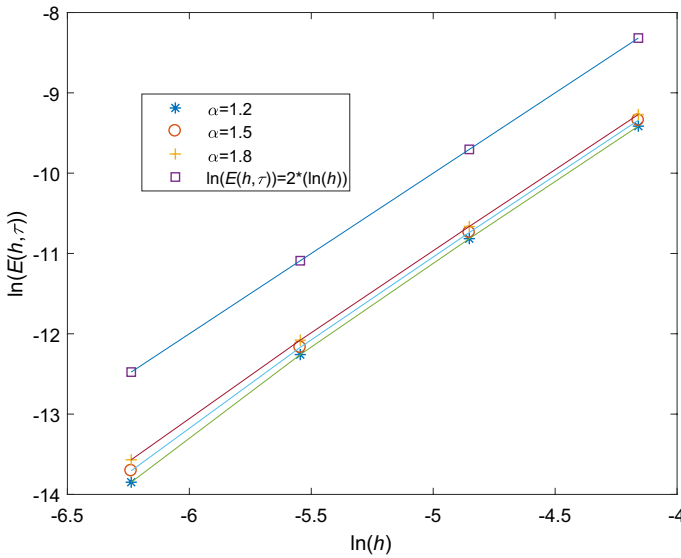


Fig. 1 $\ln(E(h, \tau))$ when $\tau = 2^{-11}$

Table 1 Performance of three preconditioners on Example 4.1 with $N = 1$

α	M	PGMRES-T		PGMRES-L		PGMRES-C	
		iter	CPU/s	iter	CPU/s	iter	CPU/s
1.5	2^9	23	0.03	72	0.03	25	0.01
	2^{10}	23	0.05	87	0.06	25	0.02
	2^{11}	23	0.09	105	0.09	25	0.04
1.6	2^9	23	0.02	49	0.02	25	0.01
	2^{10}	23	0.05	57	0.04	25	0.02
	2^{11}	23	0.09	67	0.06	25	0.04
1.7	2^9	23	0.02	33	0.01	25	0.01
	2^{10}	23	0.04	37	0.02	25	0.02
	2^{11}	23	0.09	42	0.04	25	0.04
1.8	2^9	23	0.02	21	0.01	25	0.01
	2^{10}	23	0.04	23	0.01	25	0.02
	2^{11}	23	0.09	25	0.02	25	0.04
1.9	2^9	23	0.02	12	0.01	24	0.01
	2^{10}	23	0.04	13	0.01	25	0.02
	2^{11}	23	0.09	14	0.01	25	0.04

the iteration numbers of PGMRES-L keep increasing as M increases. However, since the problem in Example 4.1 has only one-spatial dimension and the matrix size is not large, the three solvers are all efficient in terms of computational time in regardless of the difference in iteration numbers.

Example 4.2 Consider a 2-D OSFDE with $[x_L, x_R] = [y_L, y_R] = [0, 2]$, $T = 1$, and

$$d(x, y) = x^2 + y^2 + 20, \quad e(x, y) = \sin \left[\frac{\pi}{24}(x + 4) \right] + \sin \left[\frac{\pi}{24}(y + 4) \right],$$

$$f(x, y, t) = 3x^4(2 - x)^4y^4(2 - y)^4t^2 - t^3y^4(2 - y)^4d(x, y) \sum_{k=4}^8 \frac{\binom{4}{k-4} 2^{8-k} k! x^{k-\alpha}}{(-1)^k \Gamma(k+1-\alpha)}$$

$$- t^3x^4(2 - x)^4e(x, y) \sum_{k=4}^8 \frac{\binom{4}{k-4} 2^{8-k} k! y^{k-\beta}}{(-1)^k \Gamma(k+1-\beta)}.$$

The explicit expression of the exact solution for the example is $u(x, y, t) = x^4(2 - x)^4y^4(2 - y)^4t^3$.

To show the convergence order of the proposed scheme on Example 4.2, we plot the values of $\ln(E(h, \tau))$ with different h and fixed τ in Fig. 2. As illustrated in Fig. 2, the values of $\ln(E(h, \tau))$ are distributed like a straight line with slope “2”, which demonstrates the second-order accuracy in space of the proposed scheme in the 2-D case.

We test the three preconditioners on Example 4.2, the results of which are listed in Table 2. Again, we fix $N = 1$ in Table 2. Since $E(h, \tau)$ of the three solvers are almost the same, the results of $E(h, \tau)$ are skipped in Table 2. Table 2 shows that the iteration numbers of PGMRES-T are more stable with respect to variation of (α, β) and M than that of PGMRES-L and PGMRES-C. Moreover, since PGMRES-T has smaller iteration numbers on Example 4.2 than those of PGMRES-L and PGMRES-C, PGMRES-T solver requires less computation time. Hence, Table 2 demonstrates that the proposed Toeplitz preconditioner outperforms the other two preconditioners for Example 4.2.

The convexity/concavity assumption presented in (ii) of Theorem 3.3 implies that the diffusion coefficients d and e are continuous functions. However, this is only for

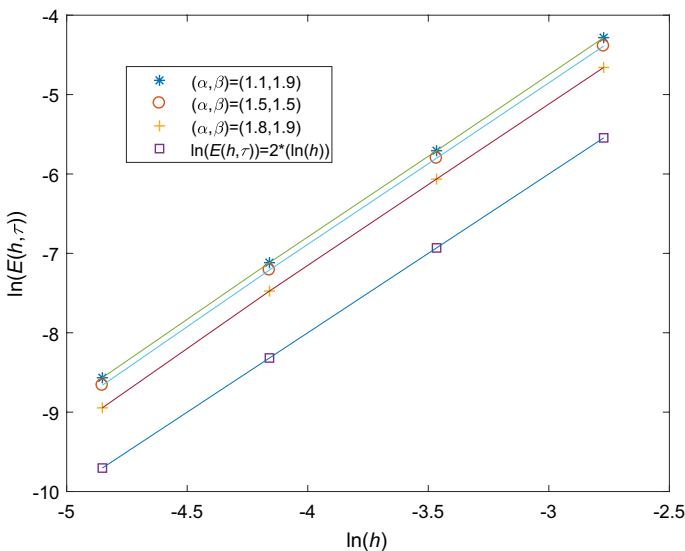


Fig. 2 $\ln(E(h, \tau))$ when $\tau = 2^{-7}$

Table 2 Performance of three preconditioners on Example 4.2 with $N = 1$

(α, β)	M	PGMRES-T		PGMRES-L		PGMRES-C	
		iter	CPU/s	iter	CPU/s	iter	CPU/s
(1.6,1.9)	2^9	12	2.01	35	5.00	47	2.73
	2^{10}	12	8.47	37	24.77	57	31.25
	2^{11}	13	47.02	38	115.55	71	171.19
(1.8,1.9)	2^9	14	2.32	26	3.51	43	4.10
	2^{10}	15	10.71	28	17.74	52	21.16
	2^{11}	16	58.26	30	87.07	64	186.62
(1.9,1.9)	2^9	15	2.46	28	3.82	40	3.15
	2^{10}	16	11.37	30	19.13	46	24.66
	2^{11}	17	61.95	32	92.89	54	121.69
(1.8,1.8)	2^9	15	2.47	28	3.72	39	4.67
	2^{10}	16	11.37	30	19.11	44	17.06
	2^{11}	16	58.17	32	92.76	51	87.10
(1.6,1.6)	2^9	16	2.61	41	5.88	36	2.01
	2^{10}	16	11.35	43	29.96	40	12.20
	2^{11}	17	61.91	46	148.58	44	70.12

theoretical consideration. Actually, the proposed Toeplitz preconditioner also works for the OSFDE with discontinuous coefficients. To demonstrate this, we test the three solvers PGMRES-L and PGMRES-C, PGMRES-T on Example 4.3.

Example 4.3 Consider a 2-D OSFDE with $[x_L, x_R] = [y_L, y_R] = [0, 2], T = 1$, and

$$\begin{aligned}
 d(x, y) &= \begin{cases} 1.1, & x \geq 1, \\ 1, & \text{otherwise,} \end{cases} & e(x, y) &= \begin{cases} 1.1, & y \leq 1, \\ 1, & \text{otherwise,} \end{cases} \\
 f(x, y, t) &= 3x^4(2-x)^4y^4(2-y)^4t^2 - t^3y^4(2-y)^4d(x, y) \sum_{k=4}^8 \frac{\binom{4}{k-4} 2^{8-k} k! x^{k-\alpha}}{(-1)^k \Gamma(k+1-\alpha)} \\
 &\quad - t^3x^4(2-x)^4e(x, y) \sum_{k=4}^8 \frac{\binom{4}{k-4} 2^{8-k} k! y^{k-\beta}}{(-1)^k \Gamma(k+1-\beta)}.
 \end{aligned}$$

The explicit expression of the exact solution for the example is $u(x, y, t) = x^4(2-x)^4y^4 \cdot (2-y)^4t^3$.

We solve the three preconditioners on Example 4.3, the results of which are listed in Table 3. Since $E(h, \tau)$ of the three solvers are almost the same, the results of $E(h, \tau)$ are skipped in Table 2. Note that the coefficients d and e are both discontinuous. From Table 3, we see that the iteration numbers of PGMRES-T are stable with respect to changes of M , which demonstrates the robustness of the proposed Toeplitz preconditioner even for the case of discontinuous coefficients. Table 3 also shows that PGMRES-T is the most efficient one among the three solvers in terms of CPU measure for sufficiently large M .

Table 3 Performance of three preconditioners on Example 4.3 with $N = 1$

(α, β)	M	PGMRES-T		PGMRES-L		PGMRES-C	
		iter	CPU/s	iter	CPU/s	iter	CPU/s
(1.6,1.9)	2^9	16	2.70	41	6.56	51	2.44
	2^{10}	16	10.57	45	30.52	66	15.35
	2^{11}	17	49.00	48	136.86	87	105.97
(1.6,1.8)	2^9	14	2.28	36	5.35	40	1.79
	2^{10}	15	9.77	39	25.40	50	10.30
	2^{11}	15	42.79	42	114.21	62	64.95
(1.6,1.7)	2^9	12	1.95	31	4.48	31	1.36
	2^{10}	12	7.83	33	20.78	36	6.76
	2^{11}	13	36.93	35	91.17	42	38.37
(1.7,1.8)	2^9	11	1.80	29	4.12	33	1.46
	2^{10}	11	7.15	31	19.20	39	7.47
	2^{11}	12	34.19	33	84.70	46	43.77
(1.8,1.9)	2^9	11	1.78	29	4.15	34	1.63
	2^{10}	11	7.13	31	19.23	41	8.26
	2^{11}	12	34.20	33	84.91	49	48.47

5 Concluding Remarks

We study the second-order schemes for time-dependent 1-D and 2-D OSFDEs with variable diffusion coefficients, in which the implicit Crank–Nicolson scheme and WSGD formula are employed to discretize the temporal and the spatial derivatives, respectively. Theoretically, we have established the unconditional stability and second-order convergence for the one-dimensional scheme without additional assumption, and for the two-dimensional scheme with certain assumptions on diffusion coefficients presented in Corollaries 3.1–3.2. To accelerate the solution process, Toeplitz preconditioners have been proposed for both one- and two-dimensional schemes. The condition numbers of the preconditioned matrices have been proven to be bounded by a constant independent of discretization step-sizes under certain assumptions on the diffusion coefficients presented in Theorems 2.4, 2.5, and 3.3. Numerical results reported have shown the second-order convergence rate of the proposed schemes and the efficiency of the proposed preconditioners.

Acknowledgements The authors would like to thank the editor and referees for valuable comments and suggestions, which helped to improve the quality of the manuscript.


References

1. Abirami, A., Prakash, P., Thangavel, K.: Fractional diffusion equation-based image denoising model using CN-GL scheme. *Int. J. Comput. Math.* **95**(6/7), 1222–1239 (2018)
2. Chen, M.H., Deng, W.H.: Fourth order accurate scheme for the space fractional diffusion equations. *SIAM J. Numer. Anal.* **52**, 1418–1438 (2014)
3. Chen, Y., Vinagre, B.M.: A new IIR-type digital fractional order differentiator. *Signal Process.* **83**(11), 2359–2365 (2003)

4. Ciarlet, P.G.: *Linear and Nonlinear Functional Analysis with Applications*, vol. 130. SIAM, Philadelphia (2013)
5. Ding, H., Li, C.: A high-order algorithm for time-Caputo-tempered partial differential equation with Riesz derivatives in two spatial dimensions. *J. Sci. Comput.* **80**(1), 81–109 (2019)
6. Donatelli, M., Mazza, M., Serra-Capizzano, S.: Spectral analysis and structure preserving preconditioners for fractional diffusion equations. *J. Comput. Phys.* **307**, 262–279 (2016)
7. De Hoog, F.: A new algorithm for solving Toeplitz systems of equations. *Linear Algebra Appl.* **88**, 123–138 (1987)
8. Gohberg, I., Olshevsky, V.: Circulants, displacements and decompositions of matrices. *Integr. Equ. Oper. Theory* **15**, 730–743 (1992)
9. Hao, Z.P., Sun, Z.Z., Cao, W.R.: A fourth-order approximation of fractional derivatives with its applications. *J. Comput. Phys.* **281**, 787–805 (2015)
10. Ilic, M., Liu, F., Turner, I., Anh, V.: Numerical approximation of a fractional-in-space diffusion equation (II)-with nonhomogeneous boundary conditions. *Fract. Calc. Appl. Anal.* **9**(4), 333–349 (2006)
11. Jin, X.Q., Vong, S.W.: *An Introduction to Applied Matrix Analysis*. Higher Education Press, Beijing (2016)
12. Koeller, R.: Applications of fractional calculus to the theory of viscoelasticity. *J. Appl. Mech.* **51**(2), 299–307 (1984)
13. Laub, A.J.: *Matrix Analysis for Scientists and Engineers*, vol. 91. SIAM, Philadelphia (2005)
14. Lei, S.L., Chen, X., Zhang, X.H.: Multilevel circulant preconditioner for high-dimensional fractional diffusion equations. *East Asian J. Appl. Math.* **6**, 109–130 (2016)
15. Lei, S.L., Huang, Y.C.: Fast algorithms for high-order numerical methods for space-fractional diffusion equations. *Int. J. Comput. Math.* **94**(5), 1062–1078 (2017)
16. Lei, S.L., Sun, H.W.: A circulant preconditioner for fractional diffusion equations. *J. Comput. Phys.* **242**, 715–725 (2013)
17. Li, C., Deng, W., Zhao, L.: Well-posedness and numerical algorithm for the tempered fractional differential equations. *Discrete Contin. Dyn. Syst.* **24**(4), 1989–2015 (2019)
18. Li, M., Gu, X.M., Huang, C., Fei, M., Zhang, G.: A fast linearized conservative finite element method for the strongly coupled nonlinear fractional Schrödinger equations. *J. Comput. Phys.* **358**, 256–282 (2018)
19. Lin, X.L., Ng, M.K.: A fast solver for multidimensional time-space fractional diffusion equation with variable coefficients. *Comput. Math. Appl.* **78**(5), 1477–1489 (2019)
20. Lin, X.L., Ng, M.K., Sun, H.W.: Efficient preconditioner of one-sided space fractional diffusion equation. *BIT* **58**, 729–748 (2018). <https://doi.org/10.1007/s10543-018-0699-8>
21. Lin, X.L., Ng, M.K., Sun, H.W.: Stability and convergence analysis of finite difference schemes for time-dependent space-fractional diffusion equations with variable diffusion coefficients. *J. Sci. Comput.* **75**, 1102–1127 (2018)
22. Lin, X.L., Ng, M.K., Sun, H.W.: Crank–Nicolson alternative direction implicit method for space-fractional diffusion equations with nonseparable coefficients. *SIAM J. Numer. Anal.* **57**(3), 997–1019 (2019)
23. Meerschaert, M.M., Scheffler, H.P., Tadjeran, C.: Finite difference methods for two-dimensional fractional dispersion equation. *J. Comput. Phys.* **211**, 249–261 (2006)
24. Meerschaert, M.M., Tadjeran, C.: Finite difference approximations for fractional advection–dispersion flow equations. *J. Comput. Appl. Math.* **172**(1), 65–77 (2004)
25. Moghaddam, B., Machado, J.T., Morgado, M.: Numerical approach for a class of distributed order time fractional partial differential equations. *Appl. Numer. Math.* **136**, 152–162 (2019)
26. Moghaderi, H., Dehghan, M., Donatelli, M., Mazza, M.: Spectral analysis and multigrid preconditioners for two-dimensional space-fractional diffusion equations. *J. Comput. Phys.* **350**, 992–1011 (2017)
27. Ng, M.K.: *Iterative Methods for Toeplitz Systems*. Oxford University Press, USA (2004)
28. Osman, S., Langlands, T.: An implicit Keller Box numerical scheme for the solution of fractional sub-diffusion equations. *Appl. Math. Comput.* **348**, 609–626 (2019)
29. Podlubny, I.: *Fractional Differential Equations*. Academic Press, New York (1999)
30. Pu, Y.F., Siarry, P., Zhou, J.L., Zhang, N.: A fractional partial differential equation based multiscale denoising model for texture image. *Math. Methods Appl. Sci.* **37**(12), 1784–1806 (2014)
31. Pu, Y.F., Zhou, J.L., Yuan, X.: Fractional differential mask: a fractional differential-based approach for multiscale texture enhancement. *IEEE Trans. Image Process.* **19**(2), 491–511 (2009)
32. Pu, Y.F., Zhou, J.L., Zhang, Y., Huang, G., Siarry, P.: Fractional extreme value adaptive training method: fractional steepest descent approach. *IEEE Trans. Neural Networks Learn. Syst.* **26**(4), 653–662 (2013)

33. Qu, W., Lei, S.L., Vong, S.: A note on the stability of a second order finite difference scheme for space fractional diffusion equations. *Numer. Algebra Control Optim.* **4**, 317–325 (2014)
34. Rossikhin, Y.A., Shitikova, M.V.: Applications of fractional calculus to dynamic problems of linear and nonlinear hereditary mechanics of solids. *Appl. Mech. Rev.* **50**(1), 15–67 (1997)
35. Roy, S.: On the realization of a constant-argument immittance or fractional operator. *IEEE Trans. Circ. Theory* **14**(3), 264–274 (1967)
36. Samko, S.G., Kilbas, A.A., Marichev, O.I.: *Fractional Integrals and Derivatives: Theory and Applications*. Gordon and Breach Science Publishers, Switzerland (1993)
37. Sousa, E.: Numerical approximations for fractional diffusion equations via splines. *Comput. Math. Appl.* **62**(3), 938–944 (2011)
38. Sousa, E., Li, C.: A weighted finite difference method for the fractional diffusion equation based on the Riemann–Liouville derivative. *Appl. Numer. Math.* **90**, 22–37 (2015)
39. Tadjeran, C., Meerschaert, M.M.: A second-order accurate numerical method for the two-dimensional fractional diffusion equation. *J. Comput. Phys.* **220**, 813–823 (2007)
40. Tadjeran, C., Meerschaert, M.M., Scheffler, H.P.: A second-order accurate numerical approximation for the fractional diffusion equation. *J. Comput. Phys.* **213**, 205–213 (2006)
41. Tian, W.Y., Zhou, H., Deng, W.H.: A class of second order difference approximations for solving space fractional diffusion equations. *Math. Comp.* **84**, 1703–1727 (2015)
42. Tseng, C.C.: Design of fractional order digital FIR differentiators. *IEEE Signal Process Lett.* **8**(3), 77–79 (2001)
43. Vong, S., Lyu, P.: On a second order scheme for space fractional diffusion equations with variable coefficients. *Appl. Numer. Math.* **137**, 34–48 (2019)
44. Vong, S., Lyu, P., Chen, X., Lei, S.L.: High order finite difference method for time–space fractional differential equations with Caputo and Riemann–Liouville derivatives. *Numer. Algorithms* **72**, 195–210 (2016)
45. Yuttanan, B., Razzaghi, M.: Legendre wavelets approach for numerical solutions of distributed order fractional differential equations. *Appl. Math. Modell.* **70**, 350–364 (2019)
46. Zeng, F., Liu, F., Li, C., Burrage, K., Turner, I., Anh, V.: A Crank–Nicolson ADI spectral method for a two-dimensional Riesz space fractional nonlinear reaction–diffusion equation. *SIAM J. Numer. Anal.* **52**, 2599–2622 (2014)
47. Zhuang, P., Liu, F., Anh, V., Turner, I.: Numerical methods for the variable-order fractional advection–diffusion equation with a nonlinear source term. *SIAM J. Numer. Anal.* **47**(3), 1760–1781 (2009)

Affiliations

Xue-lei Lin¹ · Pin Lyu²  · Michael K. Ng³ · Hai-Wei Sun⁴ · Seakweng Vong⁴

Xue-lei Lin
hxuellin@gmail.com

Michael K. Ng
mng@maths.hku.hk

Hai-Wei Sun
hsun@um.edu.mo

Seakweng Vong
swvong@um.edu.mo

¹ Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

² School of Economic Mathematics, Southwestern University of Finance and Economics, Chengdu 611130, China

³ Department of Mathematics, The University of Hong Kong, Hong Kong, China

⁴ Department of Mathematics, University of Macau, Macao, China