

Full Length Article



Fractional differential problems with numerical anti-reflective boundary conditions from a numerical linear algebra perspective: A computational study with an extensive numerical validation

Ercília Sousa^a, Cristina Tablino-Possio^{b,*}, Rolf Krause^c, Stefano Serra-Capizzano^{d,e}

^a CMUC, Department of Mathematics, University of Coimbra, Coimbra, Portugal

^b Department of Mathematics and Applications, University of Milano - Bicocca, Milano, Italy

^c Center for Computational Medicine in Cardiology, University of Italian Switzerland, Lugano, Switzerland

^d Department of Science and High Technology, University of Insubria, Como, Italy

^e Department of Information Technology, Uppsala University, Uppsala, Sweden

ARTICLE INFO

2000 MSC:
35R11
15B05
65F08

Keywords:

Physical boundary conditions (BCs)
Numerical BCs
Anti-reflective BCs
Fractional differential equations
Fast numerical solvers

ABSTRACT

In the current work, we propose numerical anti-reflective boundary conditions (BCs) in the context of nonlocal problems of fractional differential type: the numerical linear algebra goal is a $O(N \log N)$ complexity of the resulting direct and iterative algorithms, accompanied by a qualitative better approximation, with the mitigation of boundary artifacts. In fact, for showing the quality of the numerical anti-reflective BCs, we compare various types of numerical BCs, including the anti-symmetric ones considered in the case of fractional differential problems for modeling reasons. More in detail, given important similarities between anti-symmetric and anti-reflective BCs, we compare them from the perspective of computational efficiency, by considering non-truncated and truncated versions, and also other standard numerical BCs such as periodic BCs or reflective/Neumann BCs. A short theoretical analysis and several numerical tests, tables, and visualizations are provided and critically discussed. The conclusion is that the truncated numerical anti-reflective BCs perform better, both in terms of low computational cost and accuracy.

1. Introduction

We consider a specific class of fractional differential equations with the idea of testing the effectiveness of numerical boundary conditions (BCs), in terms of computational cost and quality of the approximation. We start to define the left Riemann-Liouville fractional derivative of order α , with $1 < \alpha < 2$ and $x \in \mathbb{R}$, as

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

and the right Riemann-Liouville fractional derivative of order α , with $1 < \alpha < 2$ and $x \in \mathbb{R}$, expressed as

$${}_x^{RL} D_\infty^\alpha u(x, t) = \frac{1}{\Gamma(2-\alpha)} \frac{\partial^2}{\partial x^2} \int_x^\infty u(\xi, t)(\xi-x)^{1-\alpha} d\xi.$$

* Corresponding author.

E-mail addresses: ecs@mat.uc.pt (E. Sousa), cristina.tablinopossio@unimib.it (C. Tablino-Possio), rolf.krause@usi.ch (R. Krause), s.serracapizzano@uninsubria.it, serra@it.uu.se (S. Serra-Capizzano).

<https://doi.org/10.1016/j.amc.2025.129751>

Received 24 February 2025; Received in revised form 31 July 2025; Accepted 24 September 2025

Available online 27 October 2025

0096-3003/© 2025 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Setting $\kappa_\alpha = (\cos(\pi\alpha/2))^{-1}$, the following class of general weighted operators is considered

$$\Delta_\beta^{\alpha/2} u(x, t) = \frac{1 + \beta}{2} {}^{RL}D_{-\infty}^\alpha u(x, t) + \frac{1 - \beta}{2} {}^{RL}D_x^\alpha u(x, t),$$

with $\alpha \in (1, 2)$, $\beta \in (-1, 1)$, so that $\Delta_\beta^{\alpha/2}$ becomes a linear convex combination of the given left and right derivatives, with $\Delta_0^{\alpha/2} \equiv \Delta^{\alpha/2}$ being their arithmetic mean for $\beta = 0$ and

$$(-\Delta)^{\alpha/2} u(x, t) = \frac{1}{2 \cos(\pi\alpha/2)} [{}^{RL}D_x^\alpha u(x, t) + {}^{RL}D_\infty^\alpha u(x, t)], \tag{1}$$

according to [1].

In the current work we consider the one-dimensional time-dependent problem, given by

$$\frac{\partial u}{\partial t}(x, t) = \kappa_\alpha \Delta_\beta^{\alpha/2} u(x, t), \quad x \in (a, b), \quad t > 0, \tag{2}$$

$\alpha \in (1, 2)$, $\beta \in (-1, 1)$, with initial condition in time and with (homogeneous) physical Dirichlet BCs in space. However, when we discretize (2) using standard formulae, the nonlocality of the operator implies that we need evaluations of the numerical solution well outside of the physical window $[a, b]$. For coping with this unavoidable difficulty, inspired by a similar nonlocal effect in signal processing and imaging [2], we use numerical BCs, that is, we impose relations among the values inside $[a, b]$ and those outside $[a, b]$, with the idea of forcing simultaneously uniqueness in the solution of the discretized problem, reduction of boundary artifacts, and low computational cost. More precisely, at $x = a$, $x = b$, we use the physical BCs and for $x < a$ and $x > b$ we make use of numerical zero Dirichlet BCs, numerical reflection, anti-symmetry and anti-reflection. We stress that this work is the first in using such sophisticated BCs for fractional differential equations (FDEs), for diminishing simultaneously the computational cost and the ringing effects.

Below we report the organization of the work, with a focus on the novel contribution of the present work with respect to the relevant literature. The present work is organized as follows. In Section 2 we describe the problem in an open domain, while in Sections 3 and 4 we enforce the two types of numerical BCs, connected with the presence of walls, and we discuss the implication of the infinite number of (Fourier) coefficients coming from the nonlocal nature of the underlying operators. Section 5 deals with several experiments and visualizations, which look very interesting from a numerical viewpoint due to a very low complexity of $O(N \log N)$ real arithmetic operations, with N being the number of space grid points. Section 6 is devoted to the study of ad hoc truncations, which impose that the resulting matrices lie in the anti-reflective matrix algebra [3,4] associated with the anti-reflective transform [5]. In this way we diminish further the computational cost of $O(N \log N)$ real arithmetic operations, where the hidden constant is substantially lower with respect to the previous nontruncated case, as formally discussed in Section 7. Several numerical results are reported and discussed for checking the computational advantages of our proposals. Furthermore, in Section 6.1 we provide numerical evidences of the reduction of the boundary artifacts, using numerical anti-symmetric and numerical anti-reflective BCs, while Section 6.2 contains a concise discussion on numerical experiments in a two-dimensional setting. The theoretical analysis of the $O(N \log N)$ computational cost and a comparison with the literature in various fields are reported in Section 7: as partly expected from the theoretical study in [2] in the context of signal processing and imaging, the quality of the approximation is better, when considering the anti-reflective choice with respect to all the other considered numerical BCs, even in our considered fractional setting. Finally, Section 8 contains final remarks and a mention to a list of open problems.

2. Open domain

Inspired by [6–8], we start by considering a problem with anti-symmetric physical BCs. More precisely, we suppose that we have the time dependent fractional diffusion equation in the open space domain i.e.

$$\frac{\partial u}{\partial t}(x, t) = \kappa_\alpha \Delta_\beta^{\alpha/2} u(x, t), \quad x \in \mathbb{R}, \quad t > 0,$$

with anti-symmetric physical BCs $u(-x, t) = -u(x, t)$ given in [6] and

$$\Delta_\beta^{\alpha/2} u(x, t) = \frac{1 + \beta}{2} {}^{RL}D_{-\infty}^\alpha u(x, t) + \frac{1 - \beta}{2} {}^{RL}D_x^\alpha u(x, t),$$

with $\alpha \in (1, 2)$, $\beta \in (-1, 1)$, so that $\Delta_\beta^{\alpha/2}$ becomes a linear convex combination of the given left and right derivatives, with $\Delta_0^{\alpha/2} \equiv \Delta^{\alpha/2}$ being their arithmetic mean for $\beta = 0$.

Hereafter, we will consider the Grünwald-Letnikov approximations of the left and right Riemann-Liouville fractional derivatives at (x_j, t_n) , that is

$${}^{RL}D_{-\infty}^\alpha u(x_j, t_n) \approx \frac{1}{(\Delta x)^\alpha} \sum_{k=0}^{\infty} g_k^\alpha u(x_{j+1-k}, t_n), \tag{3}$$

$${}^{RL}D_x^\alpha u(x_j, t_n) \approx \frac{1}{(\Delta x)^\alpha} \sum_{k=0}^{\infty} g_k^\alpha u(x_{j-1+k}, t_n), \tag{4}$$

with $x_j = j\Delta x$ and $t_n = n\Delta t$, $j \in \mathbb{Z}$, $n \geq 0$, where the Grünwald-Letnikov coefficients are defined through the subsequent recurrence formula for all $\alpha > 0$ as

$$g_0^\alpha = 1, \quad g_{k+1}^\alpha = -\frac{\alpha - k}{k + 1} g_k^\alpha, \quad k \geq 0.$$

Let U_j^n represent the approximate solution $u(x_j, t_n)$ in the discrete domain and let $\mathbf{U}^n = [U_{-N}^n, \dots, U_N^n]^T$, where the truncation is performed safely taking into consideration that the function goes to zero as we go to infinity. Then the θ -method for the time integration in matrix form gives

$$\frac{\mathbf{U}^{n+1} - \mathbf{U}^n}{\Delta t} = \frac{\kappa_\alpha}{(\Delta x)^\alpha} A_\beta (\theta \mathbf{U}^{n+1} + (1 - \theta) \mathbf{U}^n),$$

or, equivalently,

$$(I - \mu_\alpha \theta A_\beta) \mathbf{U}^{n+1} = (I + \mu_\alpha (1 - \theta) A_\beta) \mathbf{U}^n,$$

where the matrix $A_\beta / (\Delta x)^\alpha$ represents the chosen approximation of $\Delta_\beta^{\alpha/2}(\cdot)$, I represents the identity matrix, and $\mu_\alpha = \kappa_\alpha \Delta t / (\Delta x)^\alpha$.

As well known, the Explicit Euler method is obtained for $\theta = 0$, the Implicit Euler method for $\theta = 1$, and the Crank-Nicolson method for $\theta = 1/2$. Clearly, the application of the Explicit Euler method simply requires matrix-vector multiplications with Toeplitz structures. In Section 5 we will consider the more interesting case of application of both Implicit Euler and Crank-Nicolson methods where we have to solve the structured linear systems above. These methods with $\theta = 1, 1/2$ have much better stability and approximation properties.

Thus, to form the matrix A_β , we simply need to consider (3)-(4) and the assumption $\mathbf{U}^n = [U_{-N}^n, \dots, U_N^n]^T$, so that

$$A_\beta = \frac{1 + \beta}{2} A_L + \frac{1 - \beta}{2} A_R,$$

$$A_L = \begin{bmatrix} g_1^\alpha & g_0^\alpha & 0 & \dots & 0 & 0 \\ g_2^\alpha & g_1^\alpha & g_0^\alpha & \dots & 0 & 0 \\ g_3^\alpha & g_2^\alpha & g_1^\alpha & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ g_{2N+1}^\alpha & g_{2N}^\alpha & g_{2N-1}^\alpha & \dots & g_2^\alpha & g_1^\alpha \end{bmatrix},$$

$A_R = A_L^T$. It is evident both A_L and A_R and hence A_β share a Toeplitz structure. Furthermore, the choice of $\beta = 0$ leads to the Riesz operator (fractional Laplacian in 1D) and

$$A_0 = \frac{1}{2} A_L + \frac{1}{2} A_R.$$

Therefore, in accordance with the continuous operator, we find a symmetric Toeplitz matrix of the form

$$A_0 = \begin{bmatrix} 2g_1^\alpha & g_0^\alpha + g_2^\alpha & g_3^\alpha & \dots & g_{2N}^\alpha & g_{2N+1}^\alpha \\ g_2^\alpha + g_0^\alpha & 2g_1^\alpha & g_0^\alpha + g_2^\alpha & \dots & g_{2N-1}^\alpha & g_{2N}^\alpha \\ g_3^\alpha & g_2^\alpha + g_0^\alpha & 2g_1^\alpha & \dots & g_{2N-2}^\alpha & g_{2N-1}^\alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ g_{2N+1}^\alpha & g_{2N}^\alpha & g_{2N-1}^\alpha & \dots & g_2^\alpha + g_0^\alpha & 2g_1^\alpha \end{bmatrix}.$$

Notice that we could even consider other types of approximations that would end up having different values for the coefficients. However, when using uniform griddings, the resulting matrix structure would remain unchanged, since it is essentially decided by the nonlocal nature of the underlying continuous operator. Hence, independently of the approximation scheme as long as the grid points are equispaced, the resulting matrix structure is the same and essentially of dense Toeplitz type (i.e. shift invariant in the terminology used in signal processing and imaging [9]). We point out that the latter represents the main target of our present study, since the computational cost of the related algorithms strongly depends on matrix structural features. Regarding structured matrices, in detail a square matrix A of size m is called Toeplitz if $a_{j,k} = \alpha_{j-k}$, $\alpha_s \in \mathbb{C}$, $s = 1 - m, \dots, m - 1$. A square matrix B of size m is called Hankel if $a_{j,k} = \beta_{j+k-1}$, $\beta_s \in \mathbb{C}$, $s = 1, \dots, 2m - 1$. We point out that any Hankel matrix B can be written as JA and $\tilde{A}J$ where A, \tilde{A} are Toeplitz and J is the antidiagonal or flip matrix with $J_{j,k} = 1$ if $j + k - 1 = m$ and zero otherwise. Hence J is a special Hankel matrix. For more results on Toeplitz and Hankel matrices refer to [10].

3. Anti-symmetric numerical BCs

We move now to the problem with an initial condition $u(x, 0) = u_0(x)$, $x \in [a, b]$, and Dirichlet physical BCs

$$\frac{\partial u}{\partial t}(x, t) = \kappa_\alpha \Delta_\beta^{\alpha/2} u(x, t), \quad x \in (a, b). \tag{5}$$

According to the formulae in (3) and (4), setting $x_j = a + j\Delta x$, $t_n = n\Delta t$, $j \in \mathbb{Z}$, $n \geq 0$, since the values U_j^n approximate the solution $u(x_j, t_n)$ and since the index j ranges between $-\infty$ and ∞ , the discrete scheme requires the evaluations well outside of the physical domain $[a, b]$. The conditions outside the physical domain are referred as numerical BCs, since their occurrence is imposed by the numerical scheme.

More specifically, for the use of values outside the interval $[a, b]$, in analogy with what is done in imaging/signal processing outside the field of values [9], we start by using numerical anti-symmetric BCs. Indeed, given $u(x, t)$ defined for $t \geq 0$ and $x \in [a, b]$, we extend it in a larger interval as

$$u(-\hat{x} + a, t) = -u(\hat{x} + a, t), \quad \text{for all } 0 < \hat{x} < b - a, \tag{6}$$

$$u(\hat{x} + b, t) = -u(-\hat{x} + b, t), \quad \text{for all } 0 < \hat{x} < b - a. \tag{7}$$

We consider a equispaced discrete domain $x_j = a + j\Delta x, j = 0, 1, \dots, N, x_0 = a, x_N = b$, and $\Delta x = \frac{b-a}{N}$. As a consequence we find

$$u(-j\Delta x + a, t) = -u(j\Delta x + a, t), \quad \text{for all } 0 < j\Delta x < b - a,$$

$$u(j\Delta x + b, t) = -u(-j\Delta x + b, t), \quad \text{for all } 0 < j\Delta x < b - a.$$

When a numerical anti-symmetric boundary condition is imposed at $x = a$, from (6) we deduce

$$U_{j+1-k}^n = -U_{-j-1+k}^n.$$

Consequently, the approximation of the left fractional Riemann-Liouville derivative becomes

$$\begin{aligned} {}_{-x}^{RL} D_x^{\alpha,ref} u(x_j, t_n) &\approx \frac{1}{(\Delta x)^\alpha} \left(\sum_{k=0}^{j+1} g_k^\alpha U_{j+1-k}^n + \sum_{k=j+2}^{\infty} g_k^\alpha U_{j+1-k}^n \right) \\ &= \frac{1}{(\Delta x)^\alpha} \left(\sum_{k=0}^{j+1} g_k^\alpha U_{j+1-k}^n - \sum_{k=j+2}^{\infty} g_k^\alpha U_{k-j-1}^n \right). \end{aligned}$$

However, the fact that the second sum goes until infinity looks problematic, since the anti-symmetric condition would send points to the right boundary wall and not inside the interior domain (a, b) . As already observed, the interior domain (a, b) is called field of values in imaging and signal processing terminology [9]. In fact, we decide to stop the second sum as described in what follows, that is,

$${}_{-x}^{RL} D_x^{\alpha,ref} u(x_j, t_n) \approx \frac{1}{(\Delta x)^\alpha} \left(\sum_{k=0}^{j+1} g_k^\alpha U_{j+1-k}^n - \sum_{k=j+2}^{N+j+1} g_k^\alpha U_{k-j-1}^n \right) \tag{8}$$

and the latter implies that we consider a bounded wall, for the numerical anti-symmetric boundary, of the same size as the interior domain. We observe that the truncation gives an approximation that makes sense if we take into consideration the fact that the sequence g_k goes to zero as k tends to infinity.

For the numerical anti-symmetric boundary condition at $x = b$, by following a similar approach, we infer

$${}_x^{RL} D_\infty^{\alpha,ref} u(x_j, t_n) \approx \frac{1}{(\Delta x)^\alpha} \left(\sum_{k=0}^{N-j+1} g_k^\alpha U_{j-1+k}^n + \sum_{k=N-j+2}^{\infty} g_k^\alpha U_{j-1+k}^n \right).$$

We also need to stop at a finite point as in the previous case, that is

$${}_x^{RL} D_\infty^{\alpha,ref} u(x_j, t_n) \approx \frac{1}{(\Delta x)^\alpha} \left(\sum_{k=0}^{N-j+1} g_k^\alpha U_{j-1+k}^n + \sum_{k=N-j+2}^{2N-j+1} g_k^\alpha U_{j-1+k}^n \right).$$

Owing to $u(x + b) = -u(-x + b)$, for $k \geq N - j + 2$, we have

$$U_{j-1+k} = U_{N-N+j-1+k} = -U_{N+N-j+1-k} = -U_{2N-j+1-k}.$$

Therefore

$${}_x^{RL} D_\infty^{\alpha,ref} u(x_j, t_n) \approx \frac{1}{(\Delta x)^\alpha} \left(\sum_{k=0}^{N-j+1} g_k^\alpha U_{j-1+k}^n - \sum_{k=N-j+2}^{2N-j+1} g_k^\alpha U_{2N-j+1-k}^n \right). \tag{9}$$

Thus, by applying again the θ -method, the matrix form of the problem is

$$(I - \mu_\alpha \theta A_\beta^{\text{anti}}) \mathbf{U}^{n+1} = (I + \mu_\alpha (1 - \theta) A_\beta^{\text{anti}}) \mathbf{U}^n,$$

with $\mathbf{U}^n = [U_0^n, \dots, U_N^n]^T$, I being the identity matrix, and $\mu_\alpha = \kappa_\alpha \Delta t / (\Delta x)^\alpha$. By combining (8) and (9), the matrix A_β^{anti} is expressed as the $(N + 1) \times (N + 1)$ matrix

$$A_\beta^{\text{anti}} = \frac{1 + \beta}{2} A_L^{\text{anti}} + \frac{1 - \beta}{2} A_R^{\text{anti}} \tag{10}$$

with

$$A_L^{\text{anti}} = \begin{bmatrix} g_1^\alpha & g_0^\alpha - g_2^\alpha & -g_3^\alpha & \dots & -g_N^\alpha & -g_{N+1}^\alpha \\ g_2^\alpha & g_1^\alpha - g_3^\alpha & g_0^\alpha - g_4^\alpha & \dots & -g_{N+1}^\alpha & -g_{N+2}^\alpha \\ g_3^\alpha & g_2^\alpha - g_4^\alpha & g_1^\alpha - g_5^\alpha & \dots & -g_{N+2}^\alpha & -g_{N+3}^\alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ g_{N+1}^\alpha & g_N^\alpha - g_{N+2}^\alpha & g_{N-1}^\alpha - g_{N+3}^\alpha & \dots & g_2^\alpha - g_{2N}^\alpha - g_0^\alpha & g_1^\alpha - g_{2N+1}^\alpha \end{bmatrix}$$

and

$$A_R^{\text{anti}} = \begin{bmatrix} g_1^\alpha - g_{2N+1}^\alpha & g_2^\alpha - g_{2N}^\alpha - g_0^\alpha & g_3^\alpha - g_{2N-1}^\alpha & \dots & g_N^\alpha - g_{N+2}^\alpha & g_{N+1}^\alpha \\ g_0^\alpha - g_{2N}^\alpha & g_1^\alpha - g_{2N-1}^\alpha & g_2^\alpha - g_{2N-2}^\alpha & \dots & g_{N-1}^\alpha - g_{N+1}^\alpha & g_N^\alpha \\ 0 - g_{2N-1}^\alpha & g_0^\alpha - g_{2N-2}^\alpha & g_1^\alpha - g_{2N-3}^\alpha & \dots & & g_{N-1}^\alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 - g_{N+1}^\alpha & 0 - g_N^\alpha & 0 - g_{N-1}^\alpha & \dots & g_0^\alpha - g_2^\alpha & g_1^\alpha \end{bmatrix}.$$

We stress that the terms appearing because of the presence of the boundary are highlighted in a different color. We also notice as the presence of the correction term g_0^α in the first and last equation directly originates from the fact that the approximation of (3) and (4) starts from $k = 0$, that is the approximation across the considered point becomes an approximation across the boundary, so requiring a proper reflection inside. For example, in the last equation, the left fractional Riemann-Liouville derivative in x_N is approximated as

$$\begin{aligned} {}^{RL}D_x^{\alpha,ref} u(x_N, t_n) &\approx \frac{1}{(\Delta x)^\alpha} \left(\sum_{k=0}^{N+1} g_k^\alpha U_{N+1-k}^n - \sum_{k=N+2}^{\infty} g_k^\alpha U_{k-N-1}^n \right) \\ &= \frac{1}{(\Delta x)^\alpha} \left(-g_0^\alpha U_{N-1}^n + \sum_{k=1}^{N+1} g_k^\alpha U_{N+1-k}^n - \sum_{k=N+2}^{\infty} g_k^\alpha U_{k-N-1}^n \right), \end{aligned}$$

being $U_{N+1}^n = -U_{N-1}^n$.
When $\beta = 0$ we find

$$A_0^{anti} = \frac{1}{2} A_L^{anti} + \frac{1}{2} A_R^{anti}$$

so that $2A_0^{anti}$ equals the matrix

$$\begin{bmatrix} 2g_1^\alpha - g_{2N+1}^\alpha & g_0^\alpha + g_2^\alpha - g_3^\alpha - g_{2N}^\alpha - g_0^\alpha & g_3^\alpha - g_3^\alpha - g_{2N-1}^\alpha & \dots & \dots & g_{N-1}^\alpha - g_{N-1}^\alpha - g_{N+3}^\alpha & g_{N-1}^\alpha - g_{N-1}^\alpha - g_{N+2}^\alpha & g_{N+1}^\alpha - g_{N+1}^\alpha \\ g_0^\alpha + g_2^\alpha - g_{2N}^\alpha & 2g_1^\alpha - g_3^\alpha - g_{2N-1}^\alpha & g_0^\alpha + g_2^\alpha - g_4^\alpha - g_{2N-2}^\alpha & g_3^\alpha - g_5^\alpha - g_{2N-3}^\alpha & \dots & \dots & g_{N-1}^\alpha - g_{N+1}^\alpha - g_{N+1}^\alpha & g_{N-1}^\alpha - g_{N+2}^\alpha \\ g_3^\alpha - g_{2N-1}^\alpha & g_0^\alpha + g_2^\alpha - g_4^\alpha - g_{2N-2}^\alpha & \vdots & \vdots & \vdots & \vdots & g_{N-2}^\alpha - g_{N+2}^\alpha - g_{N-2}^\alpha & g_{N-1}^\alpha - g_{N+3}^\alpha \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{N-1}^\alpha - g_{N+3}^\alpha & g_{N-2}^\alpha - g_{N-1}^\alpha - g_{N+2}^\alpha & \vdots & \vdots & \vdots & \vdots & \vdots & g_3^\alpha - g_{2N-1}^\alpha \\ g_{N-1}^\alpha - g_{N+2}^\alpha & g_{N-1}^\alpha - g_{N+1}^\alpha - g_{N+1}^\alpha & g_{N-2}^\alpha - g_{N+2}^\alpha - g_{N-2}^\alpha & \dots & \dots & \dots & 2g_1^\alpha - g_{2N-1}^\alpha - g_3^\alpha & g_0^\alpha + g_3^\alpha - g_{2N}^\alpha \\ g_{N+1}^\alpha - g_{N+1}^\alpha & g_{N-1}^\alpha - g_{N+2}^\alpha - g_{N-1}^\alpha & g_{N-1}^\alpha - g_{N+3}^\alpha - g_{N-1}^\alpha & \dots & \dots & g_3^\alpha - g_{2N-1}^\alpha - g_3^\alpha & g_0^\alpha + g_2^\alpha - g_{2N}^\alpha - g_{2N}^\alpha - g_0^\alpha & 2g_1^\alpha - g_{2N+1}^\alpha \end{bmatrix}$$

We can rewrite A_0^{anti} in several ways according to the type of structural properties we are looking for. For instance,

$$A_0^{anti} = S + B,$$

where S is a symmetric matrix and B only has non-zero elements on the first and last rows and columns in the following manner

$$B = \frac{1}{2} \begin{bmatrix} 0 & -g_0^\alpha & 0 & \dots & 0 & 0 \\ g_2^\alpha & 0 & 0 & \dots & 0 & g_N^\alpha \\ g_3^\alpha & 0 & 0 & \dots & g_{N-1}^\alpha & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{N-1}^\alpha & 0 & 0 & 0 & g_3^\alpha & \\ g_N^\alpha & 0 & 0 & 0 & g_2^\alpha & \\ 0 & 0 & 0 & \dots & -g_0^\alpha & 0 \end{bmatrix}$$

Alternatively, a different way of looking at the same matrix structure is put in evidence in the following equations. We have

$$\begin{aligned} A_L^{anti} &= T_L - \tilde{H}_L - \tilde{R}_L^{anti}, \\ A_R^{anti} &= T_R - \tilde{H}_R - \tilde{R}_R^{anti}, \end{aligned}$$

where

$$T_L = \begin{bmatrix} g_1^\alpha & g_0^\alpha & 0 & \dots & 0 & 0 \\ g_2^\alpha & g_1^\alpha & g_0^\alpha & \dots & 0 & 0 \\ g_3^\alpha & g_2^\alpha & g_1^\alpha & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{N+1}^\alpha & g_N^\alpha & g_{N-1}^\alpha & \dots & g_2^\alpha & g_1^\alpha \end{bmatrix}, \quad T_R = T_L^T \tag{11}$$

are Toeplitz matrices in lower and upper Hessenberg form, respectively,

$$\tilde{H}_L = \begin{bmatrix} 0 & g_2^\alpha & g_3^\alpha & \dots & g_N^\alpha & g_{N+1}^\alpha \\ 0 & g_3^\alpha & g_4^\alpha & \dots & g_{N+1}^\alpha & g_{N+2}^\alpha \\ 0 & g_4^\alpha & g_5^\alpha & \dots & g_{N+2}^\alpha & g_{N+3}^\alpha \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & g_{N+2}^\alpha & g_{N+3}^\alpha & \dots & g_{2N}^\alpha & g_{2N+1}^\alpha \end{bmatrix}, \quad \tilde{H}_R = J \tilde{H}_L J,$$

J being the antidiagonal or flip matrix, are Hankel matrices apart the first and last zero columns, respectively, and

$$\tilde{R}_L^{anti} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & g_0^\alpha & 0 \end{bmatrix}, \quad \tilde{R}_R^{anti} = J \tilde{R}_L^{anti} J = \begin{bmatrix} 0 & g_0^\alpha & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

are just rank one correction matrices. Thus, the matrix A_0^{anti} can be written as

$$A_0^{\text{anti}} = \frac{1}{2}(T_L + T_R - H_L - H_R) + R_0^{\text{anti}}$$

where H_L and H_R denote the full Hankel matrices linked to \tilde{H}_L and \tilde{H}_R , respectively, and where

$$R_0^{\text{anti}} = \frac{1}{2} \begin{bmatrix} 0 & -g_0^\alpha & 0 & \dots & 0 & g_{N+1}^\alpha \\ g_2^\alpha & 0 & 0 & \dots & 0 & g_N^\alpha \\ g_3^\alpha & 0 & 0 & \dots & 0 & g_{N-1}^\alpha \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_N^\alpha & 0 & 0 & \dots & 0 & g_2^\alpha \\ g_{N+1}^\alpha & 0 & 0 & \dots & -g_0^\alpha & 0 \end{bmatrix}.$$

Clearly, $S_0 = \frac{1}{2}(T_L + T_R - H_L - H_R)$ is a symmetric matrix and, in particular, the matrix

$$T_0 = \frac{1}{2}(T_L + T_R)$$

is the real symmetric Toeplitz matrix with generating function $g_\alpha(\theta)$ (see [11] for the specific case and [12, Section 6.1] for the general notion of generating function of Toeplitz matrices and Toeplitz matrix-sequences), whose Fourier coefficients are defined as

$$t_0 = g_1^\alpha, \quad t_1 = (g_0^\alpha + g_2^\alpha)/2, \quad t_i = g_{i+1}^\alpha/2, \quad i = 2, \dots, N. \tag{12}$$

From [11], we know that the generating function $g_\alpha(\theta)$ is nonnegative and not identically zero so that T_0 is positive definite for any matrix-size. Furthermore $g_\alpha(\theta)$ has a unique zero at $\theta = 0$ and according to the order of the fractional derivative this zero has order α so that, by varying the matrix-sizes, the sequence of the minimal eigenvalues tends to zero as $m^{-\alpha}$ where $m = N + 1$ is the matrix-size, in accordance with (11) and taking into account [13–15].

4. Anti-reflective numerical BCs

A possible proposal to restore the continuity of the function and not only of its derivative is to consider numerical anti-reflective BCs, as first introduced in the context of signal/image deblurring and restoration [2]: see also [16,17] for applications and further results in presence of noise and blurring. More precisely, we set

$$u(-\hat{x} + a, t) - u(a, t) = u(a, t) - u(\hat{x} + a, t), \quad \text{for all } 0 < \hat{x} < b - a, \tag{13}$$

$$u(\hat{x} + b, t) - u(b, t) = u(b, t) - u(-\hat{x} + b, t), \quad \text{for all } 0 < \hat{x} < b - a, \tag{14}$$

Therefore, by considering the same arguments as before with respect to the boundaries, the approximation of the left fractional Riemann-Liouville derivative becomes

$$\begin{aligned} {}_{-\infty}^{RL} D_x^{\alpha,ref} u(x_j, t_n) &\approx \frac{1}{(\Delta x)^\alpha} \left(\sum_{k=0}^{j+1} g_k^\alpha U_{j+1-k}^n + \sum_{k=j+2}^{\infty} g_k^\alpha U_{j+1-k}^n \right) \\ &= \frac{1}{(\Delta x)^\alpha} \left(\sum_{k=0}^{j+1} g_k^\alpha U_{j+1-k}^n + \sum_{k=j+2}^{\infty} g_k^\alpha (2U_0^n - U_{k-(j+1)}^n) \right) \\ &\approx \frac{1}{(\Delta x)^\alpha} \left(\sum_{k=0}^{j+1} g_k^\alpha U_{j+1-k}^n + \sum_{k=j+2}^{N+j+1} g_k^\alpha (2U_0^n - U_{k-(j+1)}^n) \right) \end{aligned} \tag{15}$$

and the approximation of the right fractional Riemann-Liouville derivative becomes

$$\begin{aligned} {}_x^{RL} D_\infty^{\alpha,ref} u(x_j, t_n) &\approx \frac{1}{(\Delta x)^\alpha} \left(\sum_{k=0}^{N-j+1} g_k^\alpha U_{j-1+k}^n + \sum_{k=N-j+2}^{\infty} g_k^\alpha U_{j-1+k}^n \right) \\ &= \frac{1}{(\Delta x)^\alpha} \left(\sum_{k=0}^{N-j+1} g_k^\alpha U_{j-1+k}^n + \sum_{k=N-j+2}^{\infty} g_k^\alpha (2U_N^n - U_{2N-j+1-k}^n) \right) \\ &\approx \frac{1}{(\Delta x)^\alpha} \left(\sum_{k=0}^{N-j+1} g_k^\alpha U_{j-1+k}^n + \sum_{k=N-j+2}^{2N-j+1} g_k^\alpha (2U_N^n - U_{2N-j+1-k}^n) \right). \end{aligned} \tag{16}$$

Thus, by applying again the θ -method, the matrix form of the problem is

$$(I - \mu_\alpha \theta A_\beta^{\text{antiR}}) \mathbf{U}^{n+1} = (I + \mu_\alpha (1 - \theta) A_\beta^{\text{antiR}}) \mathbf{U}^n,$$

with $\mathbf{U}^n = [U_0^n, \dots, U_N^n]^T$, I being the identity matrix, and $\mu_\alpha = \kappa_\alpha \Delta t / (\Delta x)^\alpha$. By combining (15) and (16), the matrix A_β^{anti} is expressed as the $(N + 1) \times (N + 1)$ matrix

$$A_\beta^{\text{antiR}} = \frac{1 + \beta}{2} A_L^{\text{antiR}} + \frac{1 - \beta}{2} A_R^{\text{antiR}} \tag{17}$$

and for $\beta = 0$ the matrix $2A_0^{\text{antiR}}$ equals the matrix

$$\begin{bmatrix} 2g_1^\alpha - g_{2N+1}^\alpha & g_0^\alpha + g_2^\alpha - g_2^\alpha - g_{2N}^\alpha - g_0^\alpha & g_3^\alpha - g_3^\alpha - g_{2N-1}^\alpha & \dots & \dots & g_{N-1}^\alpha - g_{N-1}^\alpha - g_{N+3}^\alpha & g_N^\alpha - g_N^\alpha - g_{N+2}^\alpha & g_{N+1}^\alpha - g_{N+1}^\alpha \\ g_0^\alpha + g_2^\alpha - g_{2N}^\alpha & 2g_1^\alpha - g_3^\alpha - g_{2N-1}^\alpha & g_0^\alpha + g_2^\alpha - g_4^\alpha - g_{2N-2}^\alpha & g_3^\alpha - g_5^\alpha - g_{2N-3}^\alpha & \dots & \dots & g_{N-1}^\alpha - g_{N+1}^\alpha - g_{N+1}^\alpha & g_N^\alpha - g_{N+2}^\alpha \\ g_3^\alpha - g_{2N-1}^\alpha & g_0^\alpha + g_2^\alpha - g_4^\alpha - g_{2N-2}^\alpha & \vdots & \vdots & \vdots & \vdots & g_{N-2}^\alpha - g_{N+2}^\alpha - g_N^\alpha & g_{N-1}^\alpha - g_{N+3}^\alpha \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{N-1}^\alpha - g_{N+3}^\alpha & g_{N-2}^\alpha - g_N^\alpha - g_{N+2}^\alpha & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_N^\alpha - g_{N+2}^\alpha & g_{N-1}^\alpha - g_{N+1}^\alpha - g_{N+1}^\alpha & g_{N-2}^\alpha - g_{N+2}^\alpha - g_N^\alpha & \dots & \dots & \dots & 2g_1^\alpha - g_{2N-1}^\alpha - g_3^\alpha & g_0^\alpha + g_2^\alpha - g_{2N}^\alpha \\ g_{N+1}^\alpha - g_{N+1}^\alpha & g_N^\alpha - g_{N+2}^\alpha - g_N^\alpha & g_{N-1}^\alpha - g_{N+3}^\alpha - g_{N-1}^\alpha & \dots & \dots & g_3^\alpha - g_{2N-1}^\alpha - g_3^\alpha & g_0^\alpha + g_2^\alpha - g_{2N}^\alpha - g_2^\alpha - g_0^\alpha & 2g_1^\alpha - g_{2N+1}^\alpha \end{bmatrix}$$

with

$$z_r^\alpha = 2 \sum_{k=r+1}^{N+r} g_k^\alpha, \quad r = 1, \dots, N + 1.$$

More in detail, the obtained structure takes the more explicit form

$$\begin{aligned} A_L^{\text{antiR}} &= T_L - \tilde{H}_L - \hat{R}_L^{\text{anti}} + \hat{Z}_L^{\text{antiR}}, \\ A_R^{\text{antiR}} &= T_R - \tilde{H}_R - \hat{R}_R^{\text{anti}} + \hat{Z}_L^{\text{antiR}}, \end{aligned}$$

where

$$\hat{R}_L^{\text{antiR}} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & g_0^\alpha & -2g_0^\alpha \end{bmatrix}, \quad \hat{R}_R^{\text{antiR}} = J \hat{R}_L^{\text{antiR}} J$$

and

$$\hat{Z}_L^{\text{antiR}} = [2\tilde{H}_L \mathbf{e} \mid O^{(N+1) \times N}], \quad \hat{Z}_R^{\text{antiR}} = [O^{(N+1) \times N} \mid 2\tilde{H}_R \mathbf{e}],$$

with $\mathbf{e} = [1, \dots, 1]^T$ and $O^{(N+1) \times N}$ zero matrix of dimension $(N + 1) \times N$.

Thus, as in the case of numerical anti-symmetric BCs, for $\beta = 0$ the matrix A_0^{antiR} can be written as

$$A_0^{\text{antiR}} = \frac{1}{2} (T_L + T_R - H_L - H_R) + R_0^{\text{antiR}} = S_0 + R_0^{\text{antiR}}$$

where again H_L and H_R denote the full Hankel matrices linked to \tilde{H}_L and \tilde{H}_R respectively and where

$$R_0^{\text{antiR}} = \frac{1}{2} \left[\begin{array}{cccc|cccc} 2g_0^\alpha + z_1^\alpha & & -g_0^\alpha & 0 & \dots & 0 & g_{N+1}^\alpha + z_{N+1}^\alpha & \\ g_2^\alpha + z_2^\alpha & & 0 & 0 & \dots & 0 & g_N^\alpha + z_N^\alpha & \\ g_3^\alpha + z_3^\alpha & & 0 & 0 & \dots & 0 & g_{N-1}^\alpha & \\ \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \\ g_N^\alpha + z_N^\alpha & & 0 & 0 & \dots & 0 & g_2^\alpha + z_2^\alpha & \\ g_{N+1}^\alpha + z_{N+1}^\alpha & & 0 & 0 & \dots & -g_0^\alpha & 2g_0^\alpha + z_1^\alpha & \end{array} \right].$$

5. Numerical experiments

In the following, in view of better approximation/stability properties, we consider numerical experiments related the more relevant case of application of Implicit Euler scheme or Crank-Nicolson scheme to the solution of (5)–(7). Indeed, in such case the matrix form of the problem is

$$(I - \mu_\alpha A_\beta) \mathbf{U}^{n+1} = \mathbf{U}^n, \tag{18}$$

and

$$\left(I - \frac{\mu_\alpha}{2} A_\beta \right) \mathbf{U}^{n+1} = \left(I + \frac{\mu_\alpha}{2} A_\beta \right) \mathbf{U}^n, \tag{19}$$

respectively, where A_β is as in (10) or as in (17).

Since the implicit schemes involve the solution of a linear system, in the following we will focus on the application of Krylov methods and related effective preconditioned techniques in the case $\beta = 0$. On that point of view, we start by giving numerical evidence of the spectral analysis of the involved structured matrices in the case $\beta = 0$, namely T_0 , A_0^{anti} , and A_0^{antiR} . First of all, we highlight as the minimal eigenvalues goes to zero asymptotically as $m^{-\alpha}$, m being the matrix dimension, according to the order of zero of the Toeplitz generating function (see [11] and references therein)

$$f_{\alpha, T_0}(\theta) = -(f_{\alpha, T_L}(\theta) + f_{\alpha, T_R}(\theta))/2$$

Table 1
Spectral analysis of $X \in \{T_0, A_0^{\text{anti}}, A_0^{\text{antiR}}\}$ - Stationary discrete problem: minimal eigenvalue and γ as in (20).

m	$\lambda_{\min}(T_0)$	$\gamma(T_0)$	$\lambda_{\min}(A_0^{\text{anti}})$	$\gamma(A_0^{\text{anti}})$	$\lambda_{\min}(A_0^{\text{antiR}})$	$\gamma(A_0^{\text{antiR}})$
$\alpha = 1.2$						
1000	2.33167e-04	-	3.15528e-04	-	3.11001e-05	-
2000	1.01115e-04	1.20536	1.37002e-04	1.20358	1.35354e-05	1.20019
4000	4.39242e-05	1.20292	5.95594e-05	1.20179	5.89123e-06	1.20009
8000	1.90983e-05	1.20158	2.59087e-05	1.20090	2.56422e-06	1.20005
$\alpha = 1.5$						
1000	1.01144e-04	-	1.26438e-04	-	6.05846e-06	-
2000	3.57435e-05	1.50066	4.46521e-05	1.50164	2.14144e-06	1.50037
4000	1.26338e-05	1.50039	1.57779e-05	1.50082	7.57015e-07	1.50019
8000	4.46604e-06	1.50023	5.57676e-06	1.50041	2.67628e-07	1.50009
$\alpha = 1.8$						
1000	2.69766e-05	-	2.99102e-05	-	4.47444e-07	-
2000	7.75208e-06	1.79905	8.58130e-06	1.80137	1.28440e-07	1.80061
4000	2.22692e-06	1.79954	2.46316e-06	1.80069	3.68771e-08	1.80030
8000	6.39615e-07	1.79977	7.07189e-07	1.80034	1.05890e-08	1.80015

where

$$f_{\alpha, T_L}(\theta) = \sum_{k=-1}^{\infty} g_{k+1}^{\alpha} e^{ik\theta} = e^{-i\theta} (1 + e^{i(\theta+\pi)})^{\alpha}$$

$$f_{\alpha, T_R}(\theta) = \overline{f_{\alpha, T_L}(\theta)}.$$

Indeed in Table 1 we report the minimal eigenvalue of the matrix $X_m \in \mathbb{R}^{m \times m}$, with $X \in T_0, A_0^{\text{anti}}, A_0^{\text{antiR}}$ for increasing dimension together with the corresponding quantity

$$\gamma(X_m) = \log_2 \left(\frac{\lambda_{\min}(X_m)}{\lambda_{\min}(X_{2m})} \right), \tag{20}$$

whose limit for the dimension m tending to infinity is exactly α . In fact, the latter is expected since the related Toeplitz matrix admits a generating function in the Wiener class which is nonnegative and with a unique zero of order α at $\theta = 0$. Hence, in the light of the results in [13–15], we know that the minimal eigenvalue of $T_m(f_{\alpha, T_0})$ is asymptotic to $m^{-\alpha}$. We notice that the result in [13] is far more general, where the result concerns the minimal modulus of the eigenvalues and where the assumption of a nonnegative generating function is replaced by a complex-valued generating function with weak sectorial character.

In addition, Fig. 1.a highlights how the whole eigenvalue distribution of the matrix A_0^{anti} mimics in quite good measure the quoted generating function even in the case of a moderate matrix dimension as 16000, while for the sake of completeness in Fig. 1.b the absolute error with respect the generating function is plotted for different values of the parameter α . Regarding the absolute error reported in Fig. 1.b, it is interesting to observe the smooth shape of the error which would suggest the use of the extrapolation techniques for the fast eigenvalue computation as in [18,19]: we also notice the exception of a unique double eigenvalue with much smaller error close to the abscissa value of 0.5.

Then, we consider the spectral analysis of the whole matrix $A_0^{\text{anti}} = I - \mu_{\alpha} A_0^{\text{anti}}$ and of the Toeplitz counterpart $T_0 = I - \mu_{\alpha} T_0$. In Table 2 the minimal and maximal eigenvalues are reported, together with the spectral condition number K_2 for increasing dimensions in the case $k = 1$ and $\Delta t = \Delta x$ and Implicit Euler scheme for different values of the parameter α . The same analysis is considered in Table 3 with respect to the Crank-Nicolson scheme.

It is evident the worsening of the condition number for increasing dimension as the parameter α increases, approaching the standard second order differential problem case, due to a faster decrease of the minimal eigenvalue to zero, while the maximal one tends to a constant (the maximum of the generating function which is known to be continuous and 2π -periodic [11]).

Then, we consider the GMRES method alone or with suitable preconditioners taken in the algebra of the circulant matrices and in the algebra of sine transforms (also called τ matrices [10]) for the solution of a linear system with matrix A_0^{anti} . More precisely, we compare the case of no preconditioning, Strang Circulant preconditioning C_0 , Frobenius Optimal Circulant preconditioning C_0^* , natural τ preconditioning \mathbb{J}_0 , Frobenius Optimal τ preconditioning \mathbb{J}_0^* (see [20–23] and references therein), all built by referring to the Toeplitz part T_0 .

In Table 4 (first and second panel) the number of iterations required to reach convergence within a tolerance of 10^{-6} is reported in the case $k = 1$ and $\Delta t = \Delta x$ and both Implicit Euler and Crank-Nicolson schemes for different values of the parameter α . The constant number of iterations independent of the dimension testify the effectiveness and robustness of the proposed preconditioning strategies. An essentially negligible dependence on the chosen scheme is also observed.

In addition, we test the robustness also when increasing the value k . In Table 4 ((third and fourth panel) the very same analysis is reported, giving evidence of the strong robustness of Strang Circulant and τ preconditioners. In Tables 5 and 6 we collect the

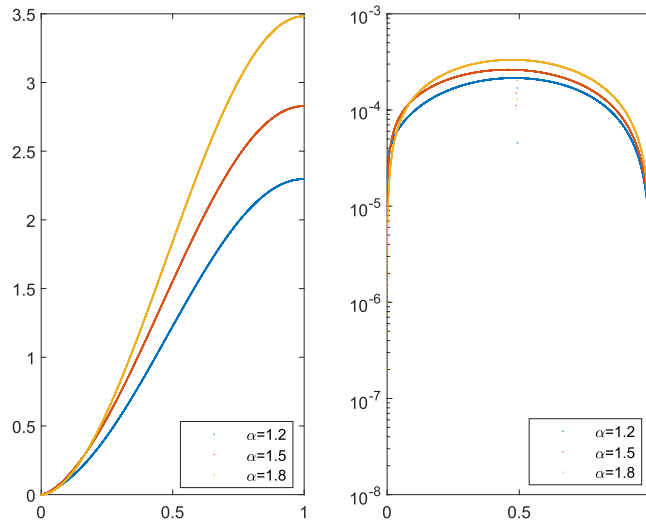


Fig. 1. Eigenvalue distribution of A_0^{anti} with size 16000 for $\alpha = 1.2, 1.5, 1.8$ and absolute error with respect to the generating function $f_{a,T_0}(\theta)$.

Table 2

Spectral analysis of $A_0^{\text{anti}} = I - \mu_\alpha A_0^{\text{anti}}$ and $T_0 = I - \mu_\alpha T_0$: minimal and maximal eigenvalue, spectral condition number - Implicit Euler method case - case $k = 1$.

N	$\lambda_{\min}(A_0^{\text{anti}})$	$\lambda_{\max}(A_0^{\text{anti}})$	$K_2(A_0^{\text{anti}})$	$\lambda_{\min}(T_0)$	$\lambda_{\max}(T_0)$	$K_2(T_0)$
$\alpha = 1.2$						
1000	1.00125	1.01443e+01	1.01315e+01	1.00093	1.01443e+01	1.01349e+01
2000	1.00063	1.15051e+01	1.14979e+01	1.00046	1.15051e+01	1.14997e+01
4000	1.00031	1.30677e+01	1.30637e+01	1.00023	1.30677e+01	1.30647e+01
8000	1.00016	1.48625e+01	1.48602e+01	1.00012	1.48625e+01	1.48608e+01
$\alpha = 1.5$						
1000	1.00399	9.03978e+01	9.00385e+01	1.00319	9.03978e+01	9.01102e+01
2000	1.00199	1.27459e+02	1.27206e+02	1.00160	1.27459e+02	1.27256e+02
4000	1.00100	1.79863e+02	1.79684e+02	1.00080	1.79863e+02	1.79720e+02
8000	1.00050	2.53966e+02	2.53840e+02	1.00040	2.53966e+02	2.53865e+02
$\alpha = 1.8$						
1000	1.00749	8.74988e+02	8.68480e+02	1.00676	8.74988e+02	8.69114e+02
2000	1.00375	1.52331e+03	1.51762e+03	1.00339	1.52331e+03	1.51817e+03
4000	1.00187	2.65203e+03	2.64707e+03	1.00169	2.65203e+03	2.64755e+03
8000	1.00094	4.61718e+03	4.61285e+03	1.00085	4.61718e+03	4.61327e+03

same tests in the case of numerical anti-reflective BCs and the comments are very similar. Finally, Table 7 account for the use of a random solution for numerical anti-symmetric and anti-reflective BCs, respectively: the randomness is nonphysical and is introduced for checking the numerical robustness of the considered methods and the conclusion is that no changes are observed in the related convergence history.

6. Truncated approximations, the anti-reflective transform, and numerical experiments

In order to increase the computational efficiency we may consider a differently truncated version of the previous approximation of the left and right fractional Riemann-Liouville derivatives, suitable tailored so that the arising matrix belongs to the anti-reflective matrix algebra [3] and the solution of the linear systems can be achieved by a direct solver within $O(N \log N)$ real arithmetic operations via few fast discrete sine transform of type I (for other fast trigonometric and Fourier-like transforms and their use, see [21,24] and references there reported). As emphasized in [25] the cost of one fast discrete transform of type I is around one half of the cost of the celebrated fast Fourier transform. Furthermore the related solver is of direct type and we do not need any preconditioned Krylov iterative solver so that the overall cost is much lower, when compared with the techniques proposed for the nontruncated versions.

More precisely, we will consider the approximations as follows

$$\begin{aligned}
 {}_{-\infty}^{RL}D_x^{\alpha,ref} u(x_j, t_n) &\approx \frac{1}{(\Delta x)^\alpha} \sum_{k=0}^{j+1} g_k^\alpha U_{j+1-k}^n + \frac{1}{(\Delta x)^\alpha} \sum_{k=j+2}^N g_k^\alpha U_{j+1-k}^n, \\
 {}_x^{RL}D_\infty^{\alpha,ref} u(x_j, t_n) &\approx \frac{1}{(\Delta x)^\alpha} \sum_{k=0}^{N-j+1} g_k^\alpha U_{j-1+k}^n + \frac{1}{(\Delta x)^\alpha} \sum_{k=N-j+2}^N g_k^\alpha U_{j-1+k}^n,
 \end{aligned}$$

Table 7

Number of preconditioned GMRES iterations to solve the linear system with matrix $A_0^{\text{anti}} = \nu I - kA_0^{\text{anti}}$ (Implicit Euler method), $A_0^{\text{anti}} = \nu I - \frac{k}{2}A_0^{\text{anti}}$ (Crank-Nicolson method) $A_0^{\text{antiR}} = \nu I - kA_0^{\text{antiR}}$ (Implicit Euler method), $A_0^{\text{antiR}} = \nu I - \frac{k}{2}A_0^{\text{antiR}}$ (Crank-Nicolson method) for increasing dimension n till $tol = 1.e - 6$ - case $k = 1$ - random exact solution.

N	A_0^{anti}					A_0^{antiR}														
	Implicit Euler		Crank-Nicolson			Implicit Euler		Crank-Nicolson												
	-	C_0	C_0^*	\mathbb{I}_0	\mathbb{I}_0^*	-	C_0	C_0^*	\mathbb{I}_0	\mathbb{I}_0^*	-	C_0	C_0^*	\mathbb{I}_0	\mathbb{I}_0^*					
$\alpha = 1.2$																				
1000	21	5	5	3	3	16	5	5	3	3	21	6	6	4	4	15	5	5	4	4
2000	22	5	5	3	3	16	5	5	3	3	22	6	6	4	4	16	5	5	4	4
4000	23	5	5	3	3	17	5	5	3	3	23	6	6	4	4	17	5	5	4	4
8000	24	5	5	3	3	18	5	5	3	3	24	6	6	4	4	18	5	5	4	4
$\alpha = 1.5$																				
1000	54	5	5	4	4	41	5	5	4	4	53	7	8	5	5	39	7	7	5	5
2000	62	5	5	4	4	46	5	5	4	4	59	7	8	5	5	45	7	7	5	5
4000	70	5	5	4	4	53	5	5	4	4	69	7	8	5	5	52	7	7	5	5
8000	80	5	5	4	4	61	5	5	4	4	77	8	8	5	5	59	7	7	5	5
$\alpha = 1.8$																				
1000	136	4	7	4	4	104	5	6	4	4	128	7	13	5	5	98	7	10	5	5
2000	166	5	7	4	4	126	5	6	4	4	154	7	13	5	5	120	7	10	5	5
4000	200	5	6	3	4	152	5	6	3	4	192	8	12	5	5	149	7	10	5	5
8000	242	5	6	4	4	187	5	6	4	4	217	6	12	5	5	172	7	10	5	5

Consequently, by imposing the anti-symmetric BCs, we square the system and the matrix becomes

$$\begin{bmatrix}
 2g_1^\alpha + 2g_0^\alpha + z_1^\alpha & g_0^\alpha + g_2^\alpha - g_2^\alpha - g_2^\alpha - g_0^\alpha & g_3^\alpha - g_3^\alpha - g_2^\alpha - g_1^\alpha & \dots & \dots & g_{N-1}^\alpha - g_{N-1}^\alpha - g_{N+3}^\alpha & g_N^\alpha - g_N^\alpha - g_{N+2}^\alpha & g_{N+1}^\alpha - g_{N+1}^\alpha + z_{N+1}^\alpha \\
 g_0^\alpha + g_2^\alpha - g_2^\alpha - g_2^\alpha + z_2^\alpha & 2g_1^\alpha - g_3^\alpha - g_2^\alpha - g_1^\alpha & g_0^\alpha + g_2^\alpha - g_4^\alpha - g_2^\alpha - g_2^\alpha & g_3^\alpha - g_3^\alpha - g_2^\alpha - g_3^\alpha & \dots & \dots & g_{N-1}^\alpha - g_{N+1}^\alpha - g_{N+1}^\alpha & g_N^\alpha - g_N^\alpha + z_N^\alpha \\
 g_3^\alpha - g_2^\alpha - g_1^\alpha + z_3^\alpha & g_0^\alpha + g_2^\alpha - g_4^\alpha - g_2^\alpha - g_2^\alpha & \dots & \dots & \dots & \dots & g_{N-2}^\alpha - g_{N+2}^\alpha - g_N^\alpha & g_{N-1}^\alpha - g_{N+3}^\alpha + z_{N-1}^\alpha \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 g_{N-1}^\alpha - g_{N+3}^\alpha + z_{N-1}^\alpha & g_{N-2}^\alpha - g_{N-1}^\alpha - g_{N+2}^\alpha & \dots & \dots & \dots & \dots & \dots & g_3^\alpha - g_2^\alpha - g_1^\alpha + z_3^\alpha \\
 g_N^\alpha - g_{N+2}^\alpha + z_N^\alpha & g_{N-1}^\alpha - g_{N+1}^\alpha - g_{N+1}^\alpha & g_{N-2}^\alpha - g_{N+2}^\alpha - g_N^\alpha & \dots & \dots & \dots & 2g_1^\alpha - g_2^\alpha - g_1^\alpha - g_3^\alpha & g_0^\alpha + g_2^\alpha - g_2^\alpha + z_2^\alpha \\
 g_{N+1}^\alpha - g_{N+1}^\alpha + z_{N+1}^\alpha & g_N^\alpha - g_{N+2}^\alpha - g_N^\alpha & g_{N-1}^\alpha - g_{N+3}^\alpha - g_{N-1}^\alpha & \dots & \dots & g_3^\alpha - g_2^\alpha - g_1^\alpha - g_3^\alpha & g_0^\alpha + g_2^\alpha - g_2^\alpha - g_0^\alpha & 2g_1^\alpha - g_2^\alpha - g_1^\alpha + 2g_0^\alpha + z_1^\alpha
 \end{bmatrix},$$

that is the anti-symmetric matrix

$$\frac{1}{2} \begin{bmatrix}
 2g_1^\alpha & 0 & 0 & \dots & \dots & 0 & 0 & 0 \\
 g_0^\alpha + g_2^\alpha & & & & & & & g_N^\alpha \\
 g_3^\alpha & & & & & & & g_{N-1}^\alpha \\
 \vdots & \tau(t_0, t_1, t_2, \dots, t_N) & & & & & & \vdots \\
 g_{N-1}^\alpha & & & & & & & g_3^\alpha \\
 g_N^\alpha & & & & & & & g_0^\alpha + g_2^\alpha \\
 0 & 0 & 0 & \dots & \dots & 0 & 0 & 2g_1^\alpha
 \end{bmatrix}, \tag{24}$$

where $\tau(t_0, t_1, t_2, \dots, t_N)$ is the matrix belonging to the τ algebra (associated to the discrete sine transform of type I) with coefficients $\{t_i\}_{i=0, \dots, N}$ given by the Toeplitz coefficients defined in (12). In the same way the matrix obtained by imposing numerical anti-reflective BCs shows the expression

$$\frac{1}{2} \begin{bmatrix}
 2g_1^\alpha + 2g_0^\alpha + \bar{z}_1 & 0 & 0 & \dots & \dots & 0 & 0 & 0 \\
 g_0^\alpha + g_2^\alpha + \bar{z}_2 & & & & & & & g_N^\alpha \\
 g_3^\alpha + \bar{z}_3 & & & & & & & g_{N-1}^\alpha + \bar{z}_{N-1} \\
 \vdots & \tau(t_0, t_1, t_2, \dots, t_N) & & & & & & \vdots \\
 g_{N-1}^\alpha + \bar{z}_{N-1} & & & & & & & g_3^\alpha + \bar{z}_3 \\
 g_N^\alpha & & & & & & & g_0^\alpha + g_2^\alpha + \bar{z}_2 \\
 0 & 0 & 0 & \dots & \dots & 0 & 0 & 2g_1^\alpha + 2g_0^\alpha + \bar{z}_1
 \end{bmatrix}, \tag{25}$$

where

$$\bar{z}_r = 2 \sum_{k=r+1}^N g_k^\alpha, \quad r = 1, \dots, N - 1$$

are just the truncated version of the previously defined quantities z_r . Therefore, the choice between the two different types of conditions pertains to the quality of the approximation and does not influence the computational efficiency.

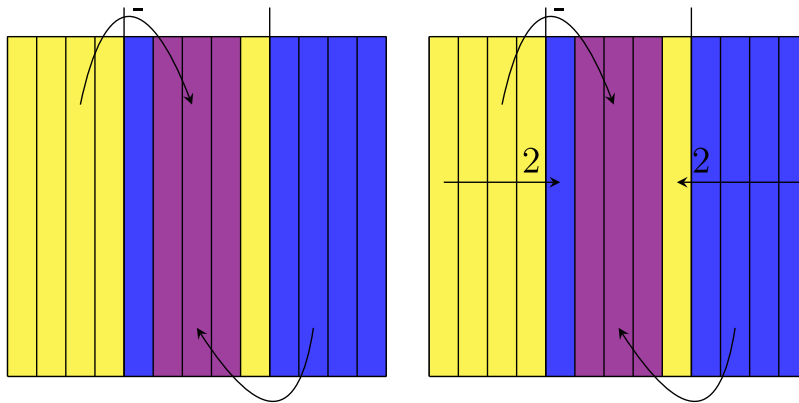


Fig. 2. Effect of the numerical anti-symmetric (left) and anti-reflective (right) BCs on the matrix structure A_0^{full} .

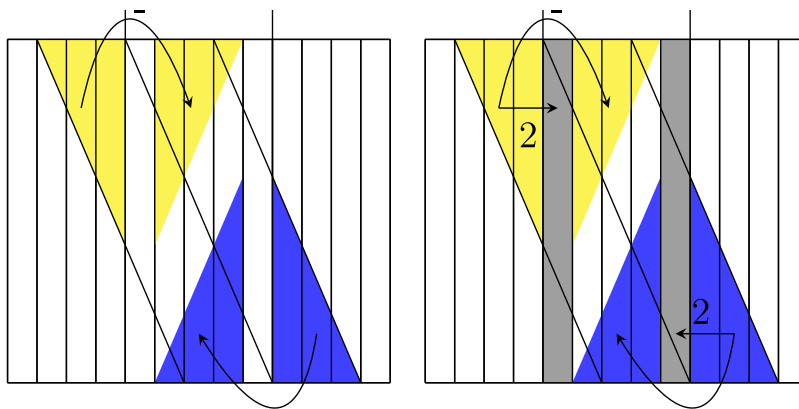


Fig. 3. Truncated anti-symmetric (left) and anti-reflective (right) boundaries effect on the matrix structure A_0^{full} .

In Figs. 2 and 3 the differences in applying the considered numerical BCs in full and its truncated version on A_0^{full} is highlighted. Notice as the external yellow and blue full wings in Fig. 2 are transferred inside the matrix A_0^{anti} with a purple overlap in the central part in the first case, while no overlap is originated in the truncated one.

Furthermore, when using physical Dirichlet BCs, the first and the last rows and columns in (24), (25) just disappear and the resulting linear systems become even simpler, since the coefficient matrix reduces to

$$\tau(t_0, t_1, t_2, \dots, t_N).$$

Therefore the whole computational cost is that of three sine I transforms owing to the τ structure [10] of $\tau(t_0, t_1, t_2, \dots, t_N)$; see the discussion in Section 7.1.

6.1. Numerical precision

Finally, we want to check the numerical precision of numerical anti-symmetric and anti-reflective BCs, by considering both nontruncated and truncated versions. Comparison is made also with respect to more classical numerical BCs of Dirichlet and reflective (Neumann) type. Notice that the numerical reflective BCs also leads to an algebra of matrices related to fast transform and hence the computational cost is essentially the same as that arising with numerical anti-reflective BCs: however the numerical anti-reflective BCs are more accurate as discussed in [2,26] and references therein.

We consider a test problem for $\beta = 0$ and homogeneous physical BCs

$$\begin{aligned} \frac{\partial u}{\partial t}(x, t) &= \Delta_\beta^{\alpha/2} u(x, t) + S(x, t), \quad x \in (0, 2), \quad t > 0, \\ u(x, 0) &= x^4(2 - x)^4, \quad x \in (0, 2), \\ u(0, t) &= 0, \quad t \geq 0, \\ u(2, t) &= 0, \quad t \geq 0, \end{aligned} \tag{26}$$

Table 8
Maximal absolute error at t in the case of numerical Dirichlet (D), reflective (R), and anti-reflective (AR) approximations - Problem with physical homogeneous Dirichlet BCs.

Implicit Euler			Crank-Nicolson			
$\alpha = 1.2$						
t	D	R	AR	D	R	AR
2.e-03	8.93e-04	9.33e-04	8.54e-04	8.93e-04	9.32e-04	8.53e-04
1	1.93e-01	2.07e-01	1.81e-01	1.93e-01	2.06e-01	1.81e-01
2	1.78e-01	1.98e-01	1.62e-01	1.78e-01	1.98e-01	1.61e-01
$\alpha = 1.4$						
t	D	R	AR	D	R	AR
2.e-03	2.16e-03	2.22e-03	2.10e-03	2.16e-03	2.23e-03	2.10e-03
1	2.98e-01	3.21e-01	2.78e-01	2.98e-01	3.20e-01	2.78e-01
2	2.17e-01	2.50e-01	1.92e-01	2.17e-01	2.50e-01	1.91e-01
$\alpha = 1.6$						
t	D	R	AR	D	R	AR
2.e-03	3.80e-03	3.87e-03	3.74e-03	3.82e-03	3.88e-03	3.75e-03
1	3.47e-01	3.70e-01	3.27e-01	3.47e-01	3.70e-01	3.27e-01
2	2.33e-01	2.67e-01	2.06e-01	2.33e-01	2.67e-01	2.06e-01
$\alpha = 1.8$						
t	D	R	AR	D	R	AR
2.e-03	5.76e-03	5.80e-03	5.72e-03	5.80e-03	5.84e-03	5.76e-03
1	3.80e-01	3.94e-01	3.67e-01	3.80e-01	3.95e-01	3.67e-01
2	2.37e-01	2.58e-01	2.20e-01	2.37e-01	2.58e-01	2.20e-01

with

$$S(x, t) = e^{-t} \left(-x^4(2-x)^4 - \frac{1}{2} \sum_{p=0}^4 (-1)^p 2^{4-p} \binom{4}{p} \frac{\Gamma(p+5)}{\Gamma(p+5-\alpha)} (x^{p+4-\alpha} + (2-x)^{p+4-\alpha}) \right),$$

and exact solution

$$u(x, t) = e^{-t} x^4(2-x)^4.$$

As it can be observed in Table 8 the numerical anti-reflective BCs guarantees a substantially higher precision for small values of t , also in the truncated version. For larger t the difference and advantage are less evident when compared with numerical Dirichlet and reflective BCs. Furthermore, also the parameter α plays a role and for the quality of the reconstruction is better for $\alpha = 1.2, 1.4$ while for $\alpha = 1.6, 1.8$ the differences are negligible at least for $t = 2$. Furthermore, it is worth noticing that in the homogeneous setting numerical anti-symmetric and anti-reflective BCs coincide. It remains to study the numerical anti-reflective BCs in the non-homogeneous setting: as emphasized in the conclusions, this will be the subject of future research. Finally, Figs. 4 and 5 reinforce the same conclusions on the higher precision of the numerical anti-symmetric/anti-reflective BCs.

6.2. Two-dimensional framework

In the two-dimensional case, for $\alpha \in (1, 2)$, we consider the fractional differential equation

$$\frac{\partial u}{\partial t}(x, y, t) = \frac{\kappa_\alpha}{2} \left({}_{-\infty}^{RL} D_x^\alpha u(x, y, t) + {}_x^{RL} D_\infty^\alpha u(x, y, t) + {}_{-\infty}^{RL} D_y^\alpha u(x, y, t) + {}_y^{RL} D_\infty^\alpha u(x, y, t) \right), \quad (x, y) \in \Omega, t > 0,$$

with

$${}_{-\infty}^{RL} D_x^\alpha u(x_i, y_j, t_n) \approx \frac{1}{(\Delta x)^\alpha} \sum_{k=0}^{\infty} g_k^\alpha u(x_{i+1-k}, y_j, t_n),$$

$${}_x^{RL} D_\infty^\alpha u(x_i, y_j, t_n) \approx \frac{1}{(\Delta x)^\alpha} \sum_{k=0}^{\infty} g_k^\alpha u(x_{i-1+k}, y_j, t_n),$$

and likewise

$${}_{-\infty}^{RL} D_y^\alpha u(x_i, y_j, t_n) \approx \frac{1}{(\Delta x)^\alpha} \sum_{k=0}^{\infty} g_k^\alpha u(x_i, y_{j+1-k}, t_n),$$

$${}_y^{RL} D_\infty^\alpha u(x_i, y_j, t_n) \approx \frac{1}{(\Delta x)^\alpha} \sum_{k=0}^{\infty} g_k^\alpha u(x_i, y_{j-1+k}, t_n).$$

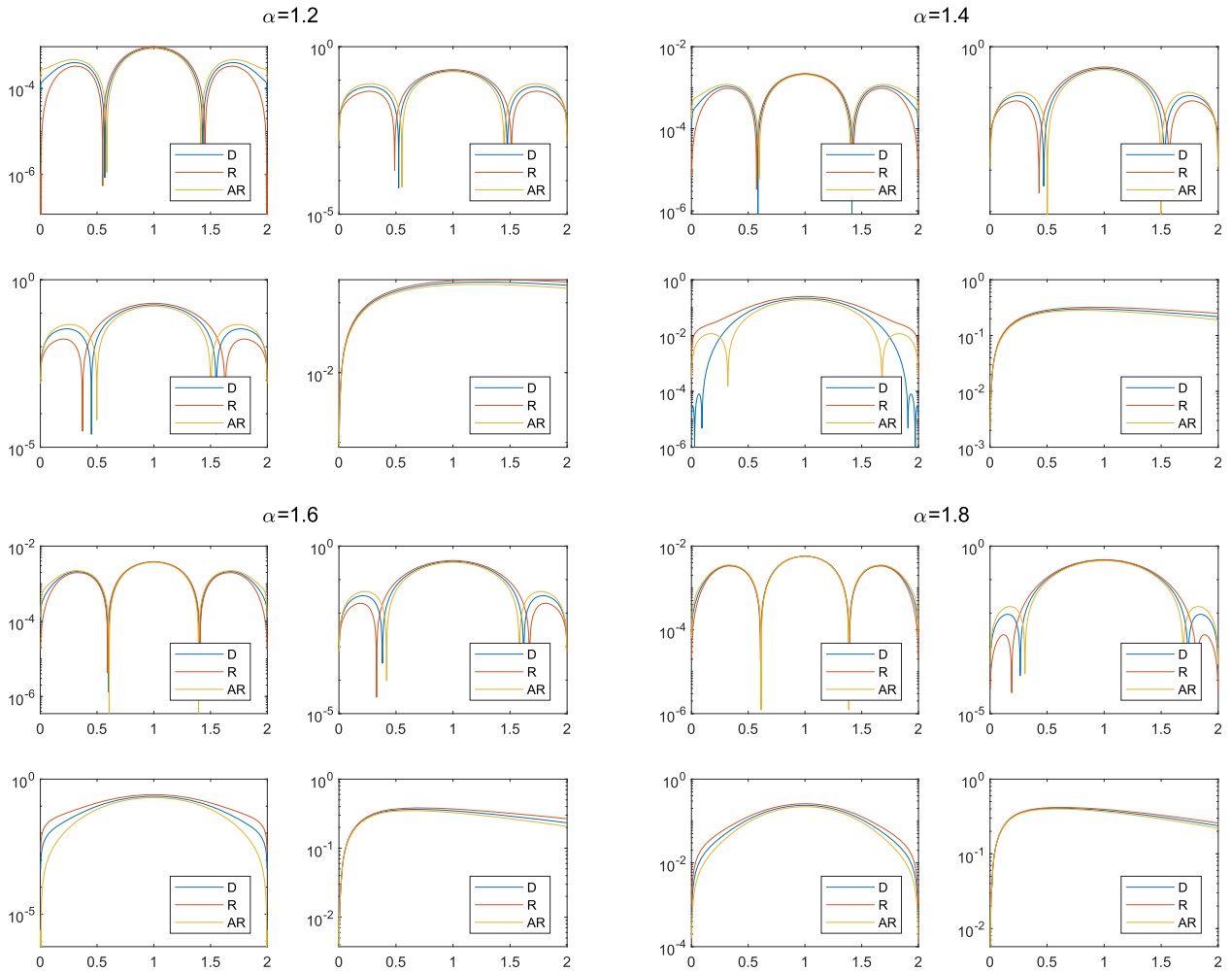


Fig. 4. Absolute error vs x at $t \in [2.e - 3, 1, 2]$ and Maximal absolute error vs t - Implicit Euler, $\alpha = 1.2, 1.4, 1.6, 1.8$ - Homogeneous Dirichlet BCs.

The separability of the operator gives rise to a tensorial matrix structure with respect to the one-dimensional matrices as follows

$$A_{2D}^{\square} = A_{1D}^{\square} \otimes I + I \otimes A_{1D}^{\square}$$

where \square denote any numerical boundary conditions type which is imposed. In this way the resulting matrix-size is N^2 ; see also [2,26] for more details on similar structures in the context of imaging. Likewise the preconditioner P in the two-dimensional case is built as

$$P_{2D}^{\triangle} = P_{1D}^{\triangle} \otimes I + I \otimes P_{1D}^{\triangle}$$

where \triangle denote any previously considered one-dimensional preconditioner, that is Strang Circulant preconditioning C_0 , Frobenius Optimal Circulant preconditioning C_0^* , natural τ preconditioning \mathbb{J}_0 , Frobenius Optimal τ preconditioning \mathbb{J}_0^* .

As reported in Table Table 9, the two-level τ preconditioning is the most effective, especially in connection with numerical anti-symmetric BCs. With respect to the one-dimensional setting, we observe a deterioration in the case of larger α for numerical anti-reflective BCs. This can be attributed to the simultaneous presence of two facts: the ill-conditioning which grows with α and the rank distance of the anti-reflective structures from the τ algebra which grows as N (i.e. the square root of the matrix-size), in accordance with the analysis in [3]. For the computation cost refer to Remark 3 in the subsequent section.

7. Theoretical analysis and comparison with the literature

In this section first we formally derive the computational costs of the proposed techniques. Then we discuss the connections of the current study with the literature.

7.1. Computational costs

We discuss the cost in terms of arithmetic operations of the algorithms introduced and used in the previous sections.

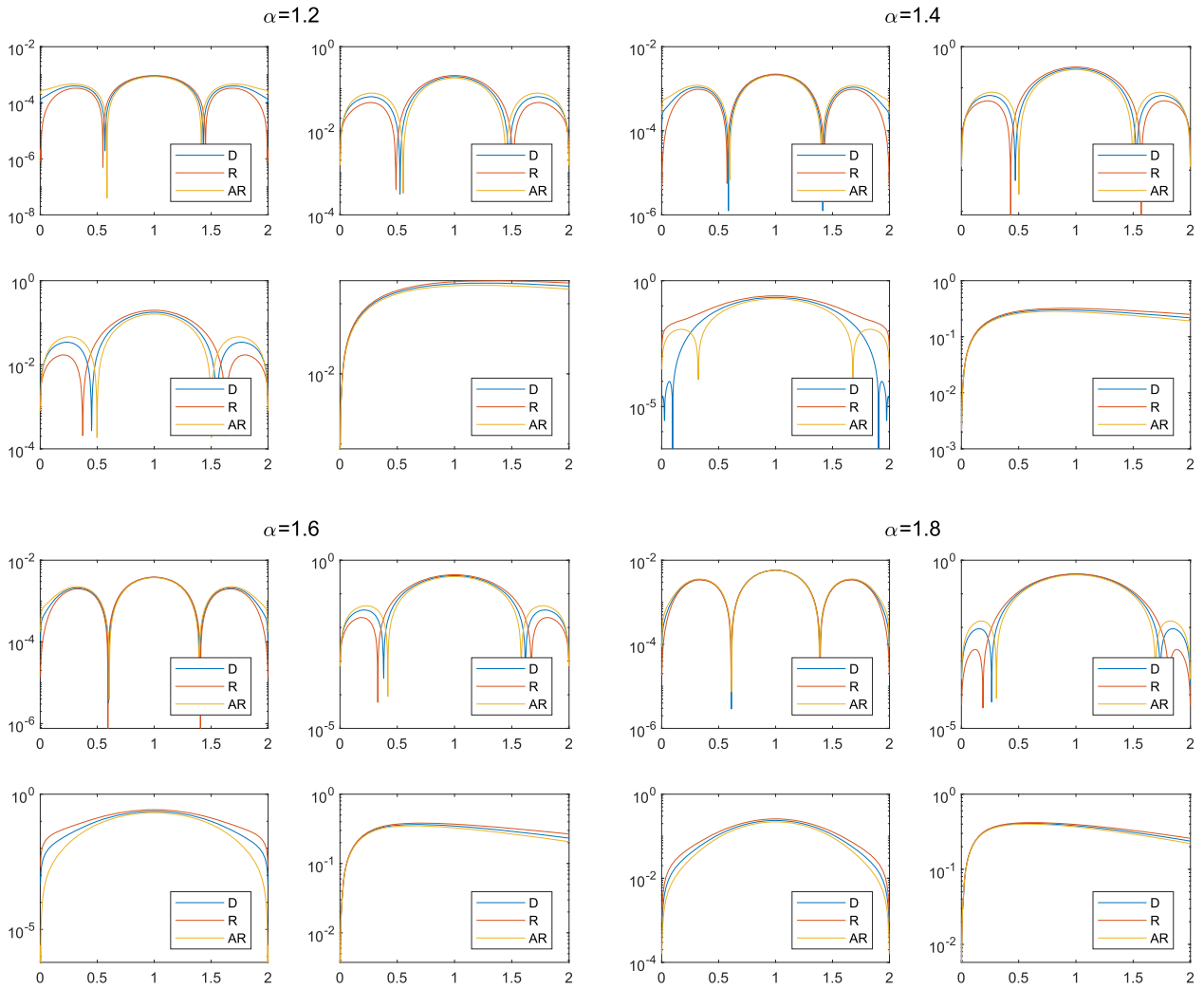


Fig. 5. Absolute error vs x at $t \in \{0.002, 1, 2\}$ and Maximal absolute error vs t - Crank-Nicolson, $\alpha = 1.2, 1.4, 1.6, 1.8$ - Homogeneous Dirichlet BCs.

Theorem 1. *The matrix-vector product with zero Dirichlet numerical BCs, periodic numerical BCs, anti-symmetric numerical BCs, and anti-reflective numerical BCs, can be performed within $O(N \log(N))$ arithmetic operations.*

Proof Looking at the representations in (22) and (23) we observe two facts:

- a. the boundary conditions are decided by imposing constraints on the vector: for instance, looking at (21) for numerical zero Dirichlet BCs in the vector $U^{n,\text{full}}$, it is enough to set $U_{-N}^n = \dots = U_{-1}^n = U_{N+1}^n = \dots = U_{2N}^n = 0$;
- b. because of a.) the sum of the two matrices in (22) and (23) is the same for all the considered numerical BCs.

Hence a fast matrix-vector product $A_0^{\text{full}} \mathbf{v}$ can be implemented by multiplying the rectangular $(N + 1) \times (3N - 1)$ Toeplitz matrix A_0^{full} by the corresponding vector \mathbf{v} . For implementing such a multiplication we do a double embedding. First we complete the Toeplitz matrix to become a square $M \times M$ Toeplitz matrix, with $M = 3N - 1$ and the same coefficients as A_0^{full} . Second we embed the resulting square $M \times M$ Toeplitz matrix into a $(2M) \times (2M)$ circulant matrix C_{2M} , following the standard trick in [20, Eq. (1.6), p. 430].

Consequently the computation

$$A_0^{\text{full}} \mathbf{v}$$

can be obtained by taking the first $N + 1$ components of $C_{2M} \mathbf{v}_{\text{ext}}$, where \mathbf{v}_{ext} is the vector of size $2M$ obtained by stacking \mathbf{v} and a proper number of zeros. In this way, independently of the chosen numerical BCs, the total cost is given by the cost of 3 fast Fourier transforms (FFTs) of size $2M$ plus linear cost and hence $O(N \log(N))$ arithmetic operations are needed since $M = 3N - 1$. •

Remark 1. Notice that the algorithm suggested in the previous theorem is not optimal and in fact it can be reduced not in order, but at least in terms of multiplicative constants. For instance, using numerical zero Dirichlet BCs, the matrix A_0^{full} reduces directly

Table 9

Number of preconditioned GMRES iterations in the 2D setting to solve the linear system with matrix $\mathcal{A}_0^{\text{anti},2D} = \nu I - kA_0^{\text{anti},2D}$ (Implicit Euler method), $\mathcal{A}_0^{\text{anti},2D} = \nu I - \frac{k}{2}A_0^{\text{anti},2D}$ (Crank-Nicolson method) $\mathcal{A}_0^{\text{antiR},2D} = \nu I - kA_0^{\text{antiR},2D}$ (Implicit Euler method), $\mathcal{A}_0^{\text{antiR},2D} = \nu I - \frac{k}{2}A_0^{\text{antiR},2D}$ (Crank-Nicolson method) for increasing dimension n till $\text{tol} = 1.e - 6$ - case $k = 1$ - random exact solution.

N^2	$\mathcal{A}_0^{\text{anti},2D}$					$\mathcal{A}_0^{\text{antiR},2D}$														
	Implicit Euler		Crank-Nicolson			Implicit Euler		Crank-Nicolson												
	-	C_0	C_0^*	\mathbb{J}_0	\mathbb{J}_0^*	-	C_0	C_0^*	\mathbb{J}_0	\mathbb{J}_0^*	-	C_0	C_0^*	\mathbb{J}_0	\mathbb{J}_0^*	-	C_0	C_0^*	\mathbb{J}_0	\mathbb{J}_0^*
$\alpha = 1.2$																				
100	16	8	8	5	5	13	7	7	4	4	15	11	11	9	9	11	8	9	7	7
400	19	8	8	5	5	14	7	7	4	4	17	11	11	9	9	12	8	9	7	7
1600	22	9	8	5	5	16	7	7	4	4	18	11	12	9	9	12	9	9	7	7
6400	24	9	9	5	5	17	7	7	4	4	19	11	11	9	9	13	9	9	7	7
$\alpha = 1.5$																				
100	18	8	9	5	6	16	8	8	5	5	24	14	15	12	12	17	11	12	10	10
400	27	10	11	5	5	22	9	9	5	5	29	16	18	13	13	21	13	13	10	10
1600	36	11	12	5	5	28	10	10	5	5	37	19	20	14	14	26	14	15	11	11
6400	45	12	13	6	5	35	11	11	5	5	44	21	22	17	17	30	16	17	13	13
$\alpha = 1.8$																				
100	21	9	10	6	6	20	8	10	6	6	31	17	19	15	15	25	15	16	12	12
400	35	10	13	6	6	31	10	12	6	6	48	24	26	19	19	36	19	21	14	14
1600	55	13	17	6	6	47	12	14	6	6	70	32	35	23	23	51	24	26	18	18
6400	84	15	20	6	6	65	15	17	6	6	97	39	44	28	28	69	29	32	22	22

to a $(N + 1) \times (N + 1)$ Toeplitz matrix and hence the double embedding described in Theorem 1 is not necessary (only the second is required as described in [20, Eq. (1.6), p. 430]). A further example is represented by the case of numerical periodic BCs, in which A_0^{full} is a $(N + 1) \times (N + 1)$ circulant matrix: therefore no embedding is required at all.

Theorem 2. *The anti-symmetric linear systems with coefficient matrix as in (24) and the anti-reflective linear systems with coefficient matrix as in (25) can be solved within $O(N \log(N))$ arithmetic real operations. Furthermore, the related matrix-vector product has exactly the same cost.*

Proof Let us consider $S^{(1)}$ the linear space of structured matrices of the following form

$$M = \begin{bmatrix} \alpha & & \\ \mathbf{v} & \hat{M} & \mathbf{w} \\ & & \beta \end{bmatrix}, \tag{27}$$

with $\alpha, \beta \in \mathbb{R}$, $\mathbf{v}, \mathbf{w} \in \mathbb{R}^{n-2}$, and $\hat{M} \in \tau_{n-2}^{(1)}$. As proven in [4], not only $S^{(1)}$ is a vector space of dimension $3n - 4$, but it is also closed under multiplication and inversion (whenever M is invertible). Hence this space is a matrix-algebra as the circulants are and, instead of being diagonalized by the discrete Fourier transform, any matrix M as (27) is diagonalized by the anti-reflective fast transform [5].

Now it is clear that the matrices in (24) and (25) approximating our fractional problem with anti-symmetric and anti-reflective BCs belong to the algebra $S^{(1)}$ with $n = N + 2$ and hence they inherits all the favourable computational features, as in [3, Theorem 2.1].

Hence, in the light of the Algorithm in [3, Theorem 2.1] for $d = 1$, both the matrix-vector product and the solution of a related linear system have the same $O(N \log(N))$ computational cost, where $N + 2$ multiplications are replaced by $N + 2$ divisions when passing from the matrix-vector product to the solution of a linear system. •

Remark 2. When comparing the FFT and the fast anti-reflective transform [5], it should be observed that the cost is essentially given by that of the fast sine I transform (see [5] and references therein). In this direction, for n power of 2, the Van Loan book [25] indicates a floating point cost of the FFT as $4n \log 2(n)$ real operations, while with the same setting the fast sine I transform has an operation count amounting to $\frac{5}{2}n \log 2(n)$ real operations and the same type of advantage in favor of the fast sine I transform holds for generic sizes. Therefore, with a careful implementation of the fast sine I transform, the CPU timings are substantially in favor of the anti-reflective approach.

Remark 3. We observe that Theorem 2.1 in [3] holds also in dD for $d > 1$ i.e. for higher dimensional problems, as considered in Section 6.2 for $d = 2$. As a consequence, the good computational features described in Theorem 2 carry over for every $d > 1$. We remind that the study of the algebra $S^{(1)}$ can be found in [4], while a detailed analysis of the anti-reflective transform is reported in [5].

7.2. Relations with the literature

First of all it is worth noticing that the anti-reflective BCs were introduced [2] more than twenty years ago in a context of signal processing and imaging, for increasing the quality of the reconstruction of a blurred signal/image contaminated by noise and for reducing the overall complexity to that of few fast sine transforms i.e. to $O(N \log N)$ real arithmetic operations, where N is the number of pixels: see Section 7.1 and also [3,5,16,17,27] for a series of results regarding the anti-reflective algebra of matrices, the low complexity transform, and deblurring techniques associated with anti-reflective BCs. In other words, the anti-reflective BCs represent a numerical trick for obtaining higher precision at low computational cost. As in the present case in the context of FDEs, for quality of reconstruction we mean a better accuracy and the elimination of disturbing boundary artifacts, which are called ringing effects [28–30] in the imaging community.

We observe that boundary artifacts can be observed also in connection with fractional operators in imaging [31] or when solving numerically FDEs [32].

Here we proposed the anti-reflective BCs in the context of nonlocal problems modeled by FDEs of general type. In this study its further potential in imagining when the blur is of fractional nature is not considered here and it will be the subject of future investigations. We observe indeed that

- the link between fractional models and imaging problems relies on the nonlocality of the underlying operator;
- rough artifacts stemming from the boundaries have been observed also in a fractional differential setting [32,33].

Regarding the first item, in the imaging case an integral operator describes the model, while the fractional operators are also described in terms of a global integral. This is the reason for which they share a nonlocal nature.

The difference between the two settings is that fractional equations are endowed with physical BCs, while this is not the case in general in signal and image processing. Hence in this work we have been careful in making a distinction between physical and numerical BCs. Interestingly enough, since the observation in [33], it is now generally accepted that also in the numerical approximation of fractional differential equations boundary artifacts appear and they become worse and worse, when increasing the matrix-size N (see also [32]). Hence the goal of setting additional numerical BCs for diminishing the boundary artifacts and for having an optimal $O(N \log N)$ complexity is of concrete interest in real-world applications modeled by fractional operators (see e.g. [34,35] and references therein). We also remark that fractional operators are now used also in imaging and hence making a connection between the two areas is crucial for a mutual benefit: in other words the idea of numerical BCs is general and can be applied in a much wider context of nonlocal operators.

Regarding further connections with the open domain problem in Section 2, we recall the following class of nonlinear FDEs appearing in several works [6–8], usually in d -dimensional domains and without the time variable. The precise formulation can be described by the following two equations

$$(-\Delta)^{\alpha/2} u(x) = f(u), \quad x \in \mathbb{R}^d, \tag{28}$$

$$u(x', -x_d) = -u(x', x_d), \quad x = (x', x_d) \in \mathbb{R}^d, \tag{29}$$

with $x' = (x_1, \dots, x_{d-1})$, $(-\Delta)^{\alpha/2}$ being the fractional Laplacian with fractional order $\alpha \in (1, 2)$, and $f(u) = -c(x)u^p$, $c, p \geq 0$.

When considering $d = 1$, condition (29) reduces to $u(-x) = -u(x)$ with $x_d \equiv x$ and x' not present, while the operator in (28) in \mathbb{R} is defined as in [1, p.11], which reduces to (1), if we omit the time variable. We observe that the present physical anti-symmetry is somehow a special case of the numerical anti-reflection used in imaging: the latter statement is crystal clear by comparing Eqs. (24) and (25) with [2][Eq. (3.3)], since the matrix structures emerging in the present work belong to the anti-reflective algebra introduced in [2]. We stress again and it is worth noticing that the conditions (29) are called anti-symmetric BCs and characterize the continuous equation for modeling reasons, while numerical anti-reflective BCs are not part of a continuous model as the anti-symmetric BCs in (29).

Therefore, inspired by the previous considerations on the physical anti-reflections and by the connections emphasized in the previous lines, we have considered the numerical version of these BCs and more precisely anti-symmetry and anti-reflection. Conversely, our analysis of the computations costs in Section 7.1 could be used for designing fast algorithms for solving problems as in (28)-(29), given the Toeplitz/Hankel structures emphasized in Section 2.

8. Conclusions

In the present work we have combined the idea of fractional differential equations and numerical anti-symmetric and anti-reflective BCs, where the latter were introduced in a context of signal processing and imaging for increasing the quality of the reconstruction of a blurred signal/image contaminated by noise and for reducing the overall complexity to that of few fast sine transforms i.e. to $O(N \log N)$ real arithmetic operations, where N is the number of pixels. The idea of ending up with matrix structures belonging to the anti-reflective algebra or well approximated by sine transform matrices seems quite effective also in the considered setting of nonlocal fractional problems. In fact, we should emphasize that from a matrix viewpoint this is not surprising since both operators, the fractional one and those related to the blurring in imaging are all of nonlocal type. The only relevant differences rely on the presence of physical BCs in the FDEs setting and on the subspace of ill-conditioning, which corresponds to low frequencies in the fractional differential case and to the high frequencies in the case of blurring (see [36]).

Several numerical tests, tables, and visualizations have been provided and critically discussed, also in connection with other classical numerical BCs and the truncation of the two types of numerical BCs mainly considered in the current work (anti-reflective and anti-symmetric).

More research remains to be performed especially in connection with the following items:

- multidimensional domains and more involved nonlocal operators which could be treated since the τ algebra and the related sine I and anti-reflective transforms admit multilevel versions via tensorizations [2,5];
- of course, if the considered matrices do not belong to an algebra and preconditioned Krylov solvers have to be employed, then it is necessary to take into account the computational barriers in [37,38];
- the impact of the present study in imaging deblurring problems when the blur operator is of fractional nature is an interesting subject for future researches;
- non-equispaced grids or variable coefficients which could be treated, taking inspiration from [39] and by referring to the theory of generalized locally Toeplitz matrix-sequences [40] also in a multidimensional setting [41,42];
- coming back to the inspiration given by the physical anti-symmetric BCs, many issues can be considered. For instance, if the boundary is only at one side the related jumping problem between boundaries would not exist. With boundaries at both sides, we can suppose that we have this reflecting boundary as an intermediate boundary, where at the final boundary wall the function could be set to zero. Furthermore, our numerical proposals and the related analysis of the computations costs could be used for designing fast algorithms for solving problems [6–8] as in (28)-(29), given the Toeplitz/Hankel structures emphasized in Section 2; furthermore, our numerical BCs could be used in connection with the approximation technique in [39], when the domain is truncated;
- a numerical and theoretical comparison in terms of precision between the truncated and nontruncated versions of the considered numerical BCs, even if few numerics suggest that the differences are not relevant;
- a delicate issue concerns the case of non-homogeneous physical Dirichlet BCs. When using the Riemann-Liouville operator, the presence of non-homogeneous BCs lead to a further term in the operator and to a corresponding further term in its numerical approximation [43,44], so that the relations (3), (4) are no longer shift invariant. In other words, the rectangular Toeplitz structures are lost and the associated matrices have an additional term, which is still Toeplitz-like but not purely Toeplitz. As a consequence the way of adding the most precise numerical BCs is nontrivial and it will be the subject of future investigations. Of course this delicate issue is not present when using Caputo fractional derivatives.

The last two items are of particular interest since they have the potential to give a formal answer to the problem of unphysical oscillations (rough artifacts similar to those in imaging [2,26]) appearing in the numerical solution, close to the boundaries and for small stepsizes, as observed in [32,33] for stationary fractional problems.

Data availability

Data will be made available on request.

Acknowledgements

Ercília Sousa was partially supported by the Centre for Mathematics of the University of Coimbra - UIDB/00324/ 2020, funded by the Portuguese Government through FCT/MCTES. The work of Stefano Serra-Capizzano and Cristina Tablino-Possio is supported by GNCS-INdAM. The work of Rolf Krause and Stefano Serra-Capizzano is funded from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955701. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Belgium, France, Germany, Switzerland. Furthermore, Stefano Serra-Capizzano is grateful for the support of the Laboratory of Theory, Economics and Systems - Department of Computer Science at Athens University of Economics and Business.

Finally, we express deep gratitude to the Reviewers, whose suggestions have been crucial for improving the presentation of our work and for the quality of the presented results.

References

- [1] A. Lischke, G. Pang, M. Gulian, F. Song, C. Glusa, X. Zheng, Z. Mao, W. Cai, M.M. Meerschaert, M. Ainsworth, G.E. Karniadakis, What is the fractional Laplacian? A comparative review with new results, *J. Comput. Phys.* 404 (2020) 109009.
- [2] S. Serra-Capizzano, A note on antireflective boundary conditions and fast deblurring models, *SIAM J. Sci. Comput.* 25 (4) (2004) 1307–1325.
- [3] A. Aricò, M. Donatelli, S. Serra-Capizzano, The anti-reflective algebra: structural and computational analysis with application to image deblurring and denoising, *Calcolo* 45 (3) (2008) 149–175.
- [4] A. Aricò, M. Donatelli, S. Serra-Capizzano, Spectral analysis of the anti-reflective algebra, *Linear Algebra Appl.* 428 (2–3) (2008) 657–675.
- [5] A. Aricò, M. Donatelli, J. Nagy, S. Serra-Capizzano, The Anti-Reflective Transform and Regularization by Filtering. *Numerical Linear Algebra in Signals, Systems and Control*, 1–21, 80, Springer, Dordrecht, Dordrecht, 2011.
- [6] S. Dipierro, J. Thompson, E. Valdinoci, On the Harnack inequality for antisymmetric s -harmonic functions, *J. Funct. Anal.* 285 (2023) 109917.
- [7] S. Dipierro, G. Poggesi, J. Thompson, E. Valdinoci, The role of antisymmetric functions in nonlocal equations, *Trans. Amer. Math. Soc.* 377 (3) (2024) 1671–1692.
- [8] R. Zhuo, C. Li, Classification of anti-symmetric solutions to nonlinear fractional Laplace equations, *Calc. Var.* 61 (2022) 17.
- [9] P.C. Hansen, J.G. Nagy, D.P. O'leary, Deblurring images. matrices, spectra, and filtering. fundamentals of algorithms, *Soc. Ind. Appl. Math. (SIAM)* 3 (2006).
- [10] D. Bini, M. Capovani, Spectral and computational properties of band symmetric Toeplitz matrices, *Linear Algebra Appl.* 52 (53) (1983) 99–126.

- [11] M. Donatelli, M. Mazza, S. Serra-Capizzano, Spectral analysis and structure preserving preconditioners for fractional diffusion equations, *J. Comput. Phys.* 307 (2016) 262–279.
- [12] C. Garoni, S. Serra-Capizzano, Generalized Locally Toeplitz Sequences: Theory and Applications, I of *Cham*, Springer, 2017.
- [13] A. Böttcher, S. Grudsky, On the condition numbers of large semi-definite Toeplitz matrices, *Linear Algebra Appl.* 279 (1–3) (1998) 285–301.
- [14] S. Serra-Capizzano, On the extreme spectral properties of Toeplitz matrices generated by L^1 functions with several minima/maxima, *BIT* 36 (1) (1996) 135–142.
- [15] S. Serra-Capizzano, On the extreme eigenvalues of Hermitian (block) Toeplitz matrices, *Linear Algebra Appl.* 270 (1998) 109–129.
- [16] M. Donatelli, C. Estatico, A. Martinelli, S. Serra-Capizzano, Improved image deblurring with anti-reflective boundary conditions and re-blurring, *Inverse Probl.* 21 (1) (2005) 169.
- [17] M. Donatelli, S. Serra-Capizzano, Anti-Reflective Boundary Conditions and Re-Blurring, *Inverse Problems*, 21, 2005.
- [18] M. Bogoya, S.M. Grudsky, S. Serra-Capizzano, Fast non-Hermitian Toeplitz eigenvalue computations, joining matrixless algorithms and FDE approximation matrices, *SIAM J. Matrix Anal. Appl.* 45 (1) (2024) 284–305.
- [19] S.-E. Ekström, C. Garoni, S. Serra-Capizzano, Are the eigenvalues of banded symmetric Toeplitz matrices known in almost closed form?, *Exp. Math.* 27 (4) (2018) 478–487.
- [20] R. Chan, M. Ng, Conjugate gradient methods for Toeplitz systems, *SIAM Rev.* 38 (3) (1996) 427–482.
- [21] F.D. Benedetto, S. Serra-Capizzano, Optimal multilevel matrix algebra operators, *Linear Multilinear Algebra* 48 (1) (2000) 35–66.
- [22] S. Serra-Capizzano, Preconditioning strategies for Hermitian Toeplitz systems with nondefinite generating functions, *SIAM J. Matrix Anal. Appl.* 17 (4) (1996) 1007–1019.
- [23] S. Serra-Capizzano, Toeplitz preconditioners constructed from linear approximation processes, *SIAM J. Matrix Anal. Appl.* 20 (2) (1999) 446–465.
- [24] T. Kailath, V. Olshevsky, Displacement structure approach to discrete-trigonometric-transform based preconditioners of G. Strang type and of T. Chan type, *SIAM J. Matrix Anal. Appl.* 26 (3) (2005) 706–734.
- [25] C.V. Loan, Computational frameworks for the fast Fourier transform, *Soc. Ind. App. Math. (SIAM)* 10 (1992). *Frontiers in Applied Mathematics*.
- [26] J. Nagy, M. Ng, L. Perrone, Kronecker product approximations for image restoration with reflexive boundary conditions, *SIAM J. Matrix Anal. Appl.* 25 (3) (2003) 829–841.
- [27] L. Perrone, Kronecker product approximations for image restoration with anti-reflective boundary conditions, *Numer. Linear Algebra Appl.* 13 (1) (2006) 1–22.
- [28] Y. Cai, M. Donatelli, D. Bianchi, T.-Z. Huang, Regularization preconditioners for frame-based image deblurring with reduced boundary artifacts, *SIAM J. Sci. Comput.* 38 (1) (2016) 164–B189.
- [29] N.-Y. Lee, Suppression of defective data artifacts for deblurring images corrupted by random valued noise, *J. Comput. Math.* 33 (3) (2015) 263–282.
- [30] N.-Y. Lee, B.J. Lucier, Preconditioned conjugate gradient method for boundary artifact-free image deblurring, *Inverse Probl. Imaging* 10 (1) (2016) 195–225.
- [31] F. Dong, Q. Ma, Single image blind deblurring based on the fractional-order differential, *Comput. Math. Appl.* 78 (6) (2019) 1960–1977.
- [32] Z.-H. She, X. Zhang, X.-M. Gu, S. Serra-Capizzano, On τ preconditioners for a Quasi-compact difference scheme to Riesz fractional diffusion equations with variable coefficients, *arXiv* 2024. <https://arxiv.org/abs/2404.10221>.
- [33] W. Bu, Y. Tang, J. Yang, Galerkin finite element method for two-dimensional Riesz space fractional diffusion equations, *J. Comput. Phys.* 276 (2014) 26–38.
- [34] G. Espinosa-Paredes, M.A. Polo-Labarrios, J. Alvarez-Ramirez, Anomalous diffusion processes in nuclear reactors, *Ann. Nuclear Energy* 54 (2013) 227–232.
- [35] I. Podlubny, Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of their Solution and Some of their Applications, 198, Academic Press, New York, New York, 1998.
- [36] M. Donatelli, S. Serra-Capizzano, On the regularizing power of multigrid-type algorithms, *SIAM J. Sci. Comput.* 27 (6) (2006) 2053–2076.
- [37] S. Serra-Capizzano, Matrix algebra preconditioners for multilevel Toeplitz matrices are not superlinear. Special issue on structured and infinite systems of linear equations, *Linear Algebra Appl.* 343 (2002) 303–319.
- [38] S. Serra-Capizzano, E. Tyrtshnikov, How to prove that a preconditioner cannot be superlinear, *Math. Comp.* 72 (243) (2003) 1305–1316.
- [39] A. Simmons, Y. Qianqian, T. Moroney, A finite volume method for two-sided fractional diffusion equations on non-uniform meshes, *J. Comput. Phys.* 335 (2017) 747–759.
- [40] C. Garoni, H. Speleers, S.-E. Ekström, S. Serra-Capizzano, T.J.R. Hughes, Symbol-based analysis of finite element and isogeometric B-Spline discretizations of eigenvalue problems: exposition and review, *Arch. Comput. Methods Eng.* 26 (5) (2019) 1639–1690.
- [41] A. Dorostkar, M. Neytcheva, S. Serra-Capizzano, Spectral analysis of coupled PDEs and of their Schur complements via generalized locally Toeplitz sequences in 2D, *Comput. Methods Appl. Mech. Eng.* 309 (2016) 74–105.
- [42] C. Garoni, S. Serra-Capizzano, Generalized Locally Toeplitz Sequences: Theory and Applications, II of *Cham*, Springer, Cham, 2018.
- [43] K. Diethelm, The analysis of fractional differential equations. An application-oriented exposition using differential operators of Caputo type, in: *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, 2004.
- [44] M. Mazza, B-Spline collocation discretizations of Caputo and Riemann-Liouville derivatives: a matrix comparison, *Fract. Calc. Appl. Anal.* 24 (6) (2021) 1670–1698.