Optimal preconditioners for Fractional Differential Equations

P. Vassalos

Department of Informatics
Athens University of Business and Economics
Athens, Greece

Numerical Analysis and Scientific Computation with Applications

A joint work with: N. Barakitis

Introduction

Why Fractional Differential Equations

- Fractional -in space- partial order diffusion equations are a generalization of classical partial differential equations.
- They are used to describe diffusion phenomena, that cannot be modeled by the second order diffusion equations. Replacing the Laplacian operator by fractional Laplacians is motivated by the need to represent processes involving anomalous diffusion.
- In probabilistic terms, it features long- distance interactions instead of the next-neighbour interaction of random walks and the short-distance interactions of their limit, the Brownian motion.
- The FDEs are of numerical interest, since there exist only few cases in which the analytic solution is known!

Preliminaries

We focus on the initial-boundary value problem of the form

$$\begin{cases} \frac{\partial u(x,t)}{\partial t} = d_+(x,t)\frac{\partial^\alpha u(x,t)}{\partial_+ x^\alpha} + d_-(x,t)\frac{\partial^\alpha u(x,t)}{\partial_- x^\alpha} + f(x,t) & (x,t) \in (L,R)\times(0,T], \\ u(L,t) = u(R,t) = 0, & t \in [0,T], \\ u(x,0) = u_0(x), & x \in [L,R], \end{cases}$$

where $\alpha \in (0,2)$ is the **fractional** derivative order, and the nonnegative functions $d_\pm(x,t)$ are the **diffusion coefficients**. The right-handed (-) and the left-handed (+) fractional derivatives in the above equation are defined in Riemann-Liouville form as follows

Definition of fractional derivative

$$\frac{\partial^\alpha u(x,t)}{\partial_+ x^\alpha} = \frac{1}{\Gamma(n-\alpha)} \frac{\partial^n}{\partial x^n} \int_L^x \frac{u(\xi,t)}{(x-\xi)^{\alpha+1-n}} d\xi,$$

$$\frac{\partial^\alpha u(x,t)}{\partial_- x^\alpha} = \frac{(-1)^n}{\Gamma(n-\alpha)} \frac{\partial^n}{\partial x^n} \int_x^R \frac{u(\xi,t)}{(x-\xi)^{\alpha+1-n}} d\xi$$

where n is an integer such that $n-1 < \alpha \leq n$ and $\Gamma()$ is the gamma function.

Remark

If $\alpha \in \mathbb{N}$, the fractional derivatives reduce to the standard derivatives.

Finite differences approximation

Applying the shifted Grünwald formulas we can approximate the left and right-handed fractional derivatives with

$$\frac{\partial^\alpha u(x,t)}{\partial_+ x^\alpha} = \frac{1}{\Delta x^\alpha} \sum_{k=0}^{\lfloor (x-L)/\Delta x \rfloor} g_k^{(\alpha)} u(x-(k-1)\Delta x, t) + O(\Delta x),$$

and

$$\frac{\partial^\alpha u(x,t)}{\partial_- x^\alpha} = \frac{1}{\Delta x^\alpha} \sum_{k=0}^{\lfloor (R-x)/\Delta x \rfloor} g_k^{(\alpha)} u(x+(k-1)\Delta x, t) + O(\Delta x),$$

Thus, using backward finite differences for the discretization in time (implicit Euler) and the above formulas for the discretization in space, we lead to the following numerical scheme

Numerical Scheme

$$\frac{u_i^{(m)} - u_i^{(m-1)}}{\Delta t} = \frac{d_{+,i}^{(m)}}{\Delta x^\alpha} \sum_{k=0}^{i+1} g_k^{(\alpha)} u_{i-k+1}^{(m)} + \frac{d_{-,i}^{(m)}}{\Delta x^\alpha} \sum_{k=0}^{N-i+2} g_k^{(\alpha)} u_{i+k-1}^{(m)} + f_i^{(m)},$$

where $u_i^{(m)}$ denotes the numerical approximation of $u(x_i, t_m)$.
Equivalently, in matrix form notation the above scheme is written as

$$\left(\nu_{M,N} I + D_+^{(m)} T_{\alpha,N} + D_-^{(m)} T_{\alpha,N}^T\right) u^{(m)} = \nu_{M,N} u^{(m-1)} + \Delta x^\alpha f^{(m)}$$

where

1. $\nu_{M,N} = \frac{\Delta x^\alpha}{\Delta t}$, I the idetinty matrix of order N,
2. $u^{(m)} = [u_1^{(m)}, \ldots, u_N^{(m)}]^T, \quad f^{(m)} = [f_1^{(m)}, \ldots, f_N^{(m)}]^T,$
3. $D_\pm^{(m)} = \mathrm{diag}(d_{\pm,1}^{(m)}, \ldots, d_{\pm,N}^{(m)}),$

and

The coefficient matrix

$$T_{\alpha,N} = -\begin{bmatrix} g_1^{(\alpha)} & g_0^{(\alpha)} & 0 & \cdots & 0 & 0 \\ g_2^{(\alpha)} & g_1^{(\alpha)} & g_0^{(\alpha)} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ g_{N-1}^{(\alpha)} & \ddots & \ddots & \ddots & g_1^{(\alpha)} & g_0^{(\alpha)} \\ g_N^{(\alpha)} & g_{N-1}^{(\alpha)} & \ddots & \ddots & g_2^{(\alpha)} & g_1^{(\alpha)} \end{bmatrix}_{N\times N}$$

is a lower Hessenberg Toeplitz matrix. We denote the coefficient matrix of the linear system that in each step we have to solve by $\mathcal{M}_{\alpha,N}^{(m)}$, that is

$$\mathcal{M}_{\alpha,N}^{(m)} = \nu_{M,N} I + D_+^{(m)} T_{\alpha,N} + D_-^{(m)} T_{\alpha,N}^T$$

Spectral analysis of $\mathcal{M}_{\alpha,N}^{(m)}$

- From the definition of $g_i^{(\alpha)}$ i.e.

$$\begin{cases} g_0^{(\alpha)} = 1, & g_1^{(\alpha)} = -\alpha & g_0^{(\alpha)} > g_2^{(\alpha)} > g_3^{(\alpha)} \ldots > 0, \\ \sum_{k=0}^{\infty} g_k^{(\alpha)} = 0, & \sum_{k=0}^{n} g_k^{(\alpha)} < 0, & n \geq 1. \end{cases}$$

  we can conclude that $\mathcal{M}_{\alpha,N}^{(m)}$ is strictly diagonal dominant.

- When the coefficients $d_{\pm}(x,t)$ are constants then $\mathcal{M}_{\alpha,N}^{(m)}$ is Toeplitz.
- In general, assuming that $d_{\pm}(x,t)$ are nonconstant, Riemman integrable positive functions, then the generated matrix sequences belong to the so called Generalized Locally Toeplitz (GLT) class .

Spectral analysis of $\mathcal{M}_{\alpha,N}^{(m)}$. The Constant case

Theorem (Donatelli et al)

Let us assume that $d_{\pm}(x,t) = d > 0$ *and that* $\nu_{M,N} = o(1)$. *Then for the matrix-sequence* $\{\mathcal{M}_{\alpha,N}^{(m)}\}_N$ *holds that:*

$$\{\mathcal{M}_{\alpha,N}^{(m)}\}_N \sim_\lambda (d \cdot p_\alpha(\theta), [-\pi, \pi])$$

i.e. for all $F \in C_0(\mathbb{C})$:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} F(\lambda_j(\mathcal{M}_{\alpha,N}^{(m)}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(p_\alpha(\theta))\, dt$$

where

$$p_\alpha(\theta) = -e^{-\mathrm{i}\theta} \left(1 - e^{\mathrm{i}\theta}\right)^{\alpha} - e^{\mathrm{i}\theta} \left(1 - e^{-\mathrm{i}\theta}\right)^{\alpha}$$

The order of the symbol

Definition

Let $f : [a, b] \to \mathbb{R}$ be a continuous nonnegative function. We say that f has a zero of order $\rho > 0$ at $\theta_0 \in [a, b]$ if there exist two real constants $C_1, C_2 > 0$ such that

$$\liminf_{\theta \to \theta_o} \frac{f(\theta)}{|\theta - \theta_0|^\rho} = C_1, \quad \limsup_{\theta \to \theta_o} \frac{f(\theta)}{|\theta - \theta_0|^\rho} = C_2.$$

Theorem

Let $\alpha \in (1, 2), \theta \in [-\pi, \pi]$. Then, the function $p_\alpha(\theta)$ has a zero at $\theta_0 = 0$ of order α.

Remark

If $\alpha = 1$ then $p_1(\theta) = (2 - 2\cos\theta)$ while if $\alpha = 2$ then $p_2(\theta) = (4 - 4\cos\theta)$. Thus, in both cases the zero at 0, is of order 2.

Spectral analysis of $\mathcal{M}_{\alpha,N}^{(m)}$. The non-constant case

The following theorem describes the singular values of $\mathcal{M}_{\alpha,N}^{(m)}$

Theorem

Let $\nu_{M,N} = o(1)$ and that $d_+(x), d_-(x)$ are both Riemann integrable over $[L, R]$, $h_\alpha(x,\theta) = d_+(x)f_\alpha(\theta) + d_-(x)f_\alpha(-\theta)$, where

$$f_\alpha(\theta) = -e^{-\mathrm{i}\theta}\left(1 + e^{\mathrm{i}(\theta+\pi)}\right)^\alpha$$

then

$$\{\mathcal{M}_{\alpha,N}^{(m)}\}_N \sim_\sigma (h_\alpha(x,\theta), [L,R]\times[-\pi,\pi]),$$

and whenever $h_\alpha(x,\theta)$ is real-valued, i.e iff $d_+(x) = d_-(x)$ we have

$$\{\mathcal{M}_{\alpha,N}^{(m)}\}_N \sim_\lambda (h_\alpha(x,\theta), [L,R]\times[-\pi,\pi]),$$

Numerical solution of $\mathcal{M}_{\alpha,N}^{(m)} x = b$

Remark

From the previous spectral theorems, both in constant and non constant case, we conclude that the condition number of $\mathcal{M}_{\alpha,N}^{(m)}$ will tend to infinity. Thus, for a fixed tolerance, the iterations of any Krylov type method applying to the system $\mathcal{M}_{\alpha,N}^{(m)} x = b$ will increase as $N \to \infty$. Hence, the need for preconditioning is obvious.

In the literature there are two main proposals:

- Circulant preconditioners
- Band preconditioners

Even thought are both quite efficient, for different reason each of them, fail in general to ensure number of iterations independent of the dimension N even in the constant coefficient case i.e. $d_{\pm}(x) = d > 0$.

Disadvantages. Circulant case

- Circulant preconditioners cannot be optimal in the multidimensional problems i.e. two and three spacial dimensional spaces. The reason is due to a negative result (Noutsos,Serra Capizzano and V., TCS 2004) that ensure us that whenever the symbol of a Toeplitz matrix has root of order greater than one, then the condition number of $\kappa_2(C_N^{-1}T_N)$, for any C_N belonging to this matrix algebra, must tend to infinity.
- The other negative result about circulant preconditioners concerns the non constant case. Using GLT technology we can easily see that the preconditioned sequence cannot be clustered at one, since the function $\frac{h_\alpha(x,\theta)}{f(\theta)}$ is a nontrivial function depending on the variable x, whenever the diffusion coefficients are non-constant functions. Therefore the superlinear behavior of any preconditioned Krylov method is lost, in contrast with what happens in the constant coefficient case

Band preconditioners case

The band preconditioners proposed by Donatelli, Serra Capizzano, and Massa (J. Comput. Phys, 2016) catch also the influence of the non constant character of the diffusion coefficients. Specifically, are defined by the symbols

$$p_1(x,\theta) = d_+(x)(1-e^{-\mathrm{i}\theta}) + d_-(x)(1-e^{\mathrm{i}\theta})$$

and

$$p_2(x,\theta) = (d_+(x)+d_-(x))(2-2\cos\theta)$$

in matrix notation the above statements lead to

$$P_{1,N} = \nu_{M,N} I + D_+ B_N + D_- B_N^T$$
$$P_{2,N} = \nu_{M,N} I + D_+ L_N + D_- L_N^T$$

where B_N, L_N are the band Toeplitz matrices

$$B_N = \mathrm{trid}[0\ 1\ -1], \qquad L_N = \mathrm{trid}[-1\ 2\ -1]$$

Band preconditioners case. Disadvandages

- Since
$$\lim_{\theta \to 0} \frac{h_\alpha(x,\theta)}{p_k(x,\theta)} = |\theta|^{\alpha-k}, \qquad k \in \{1,2\}$$
we deduce that the condition number of the preconditioned matrices will behave asymptotical to $N^{\alpha-k}$
- In multidimensional case, this kind of preconditioners become, at least in constant coefficient case, block band Toeplitz with Toeplitz blocks. In each iteration of the preconditioned Krylov methods, we need to solve fast such kind of systems. Unfortunately, there is no similar to Thomas algorithm direct method in multidimensional setting! Thus, the time for the "inversion" of the preconditioner will overcome the $O(N^d \log N^d)$ of the GMRES.

Our proposal when $d_{\pm}(x,t) = d > 0$

Our proposal tries to eliminate the aforementioned drawbacks. In the case of $d_{\pm}(x,t) = d > 0$ we propose as preconditioner the τ preconditioner of the form

$$P_\tau = S_N D_N S_N$$

where S_N is the sine matrix

$$S_N(i,j) = \sqrt{\frac{2}{n+1}} \left(\sin\left(\frac{\pi i j}{N+1} \right) \right)_{i,j=1}^{n}$$

and

$$D_N = \operatorname{diag}\left(p\left(\frac{\pi j}{N+1} \right) \right)$$

where

$$p_\alpha(\theta) = -e^{-\mathrm{i}\theta} \left(1 - e^{\mathrm{i}\theta}\right)^{\alpha} - e^{\mathrm{i}\theta} \left(1 - e^{-\mathrm{i}\theta}\right)^{\alpha}$$

Our proposal when $d_{\pm}(x,t) = d > 0$

- Contrary to circulant case, the negative result for the τ class of matrix algebra holds when the order of root is greater than 3.
- Contrary to the band approach, our preconditioner is constructed directly from the symbol $p_\alpha(\theta)$ and not from its approximation $p_1(\theta)$ or $p_2(\theta)$. As a consequence we can ensure strong clustering around unity and superlinear convergence for the Krylov methods.
- Contrary to band approach, in multidimension case, the solution of every system of the form $P_\tau x = b$ can be very efficient and costs $O(N^d \log N^d)$.
- Using Krylov methods and matrix algebra preconditioner in such kind of systems, the total procedure becomes highly parallelizable.

Our proposal, $d_{\pm}(x,t)$ non constant case

In order to take into account also the influence of the nonconstant coefficients $d_+(x,t), d_-(x,t)$ we add into the previous preconditioner one more term, i.e.

$$P_\tau^* = KP_\tau \qquad d_+(x,t) \neq d_-(x,t)$$

or

$$P_\tau^* = K^{\frac{1}{2}} P_\tau K^{\frac{1}{2}} \qquad d_+(x,t) = d_-(x,t)$$

where K is the diagonal matrix having

$$K_{ii} = \frac{d_-(x_i,t_m) + d_+(x_i,t_m)}{2}$$

The underlying idea of the above proposal is to approximate the symbol

$$h_\alpha(x,\theta) = d_-(x) f_\alpha(\theta) + d_+(x) f_\alpha(-\theta)$$

where

$$f_\alpha(\theta) = -e^{-\mathrm{i}\theta}\left(1 + e^{\mathrm{i}(\theta+\pi)}\right)^\alpha$$

Numerical experiments

Table: 1D case: $d_-(x,t) = \Gamma(3-\alpha)|x-1/2|^\alpha = d_+(x,t)$, $\alpha = 1.5$

N	P_τ	C	B_2
2^7	4	7	8
2^8	4	8	8
2^9	4	8	8
2^{10}	5	8	9

Numerical experiments

Table: 2D case: $c_-(x,y,t) = c_+(x,y,t) = d_-(x,y,t) = d_+(x,y,t) = 0.005$, $\alpha = \beta = 1.5$

$N_1, N_2 = n$	P_τ	C	B_2
2^4	4	6	7
2^5	5	9	10
2^6	5	13	12
2^7	6	18	14

Thank you very much for your attention!