ANALYSIS OF NUMERICAL METHODS FOR SPECTRAL FRACTIONAL ELLIPTIC EQUATIONS BASED ON THE BEST UNIFORM RATIONAL APPROXIMATION

STANISLAV HARIZANOV, RAYTCHO LAZAROV, PENCHO MARINOV, SVETOZAR MARGENOV, AND JOSEPH PASCIAK

Abstract. Here we study theoretically and compare experimentally with the methods developed in [19, 8] an efficient method for solving systems of algebraic equations \( \tilde{A}^\alpha \tilde{u}_h = \tilde{f}_h, 0 < \alpha < 1 \), where \( \tilde{A} \) is an \( N \times N \) matrix coming from the discretization of a fractional diffusion operator. More specifically, we focus on matrices obtained from finite difference or finite element approximation of second order elliptic problems in \( \mathbb{R}^d \), \( d = 1,2,3 \). The proposed methods are based on the best uniform rational approximation (BURA) \( r_{\alpha,k}(t) \) of \( t^\alpha \) on \([0,1]\). Here \( r_{\alpha,k} \) is a rational function of \( t \) involving numerator and denominator polynomials of degree at most \( k \).

The approximation of \( \tilde{u}_h = \tilde{A}^{-\alpha} \tilde{f}_h \) is then \( \tilde{w}_h = \lambda_1^{-\alpha} r_{\alpha,k}(\lambda_1 \tilde{A}^{-1}) \tilde{f}_h \), where \( \lambda_1 \) is the smallest eigenvalue of \( \tilde{A} \). We show that the proposed method is exponentially convergent with respect to \( k \) and has some attractive properties. First, it reduces the solution of the nonlocal system to solution of \( k \) systems with matrix \( (\tilde{A} + c_j \tilde{I}) \) and \( c_j > 0, j = 1,2,\ldots,k \). Thus, good computational complexity can be achieved if fast solvers are available for such systems. Second, the original problem and its rational approximation in the finite difference case are positivity preserving. In the finite element case, this valid for schemes obtained by mass lumping under certain mild conditions on the mesh. Further, we prove that the lumped mass schemes still have the expected rate of convergence, at times assuming additional regularity on the right hand side. Finally, we present comprehensive numerical experiments on a number of model problems for various \( \alpha \) in one and two spatial dimensions. These illustrate the computational behavior of the proposed method and compare its accuracy and efficiency with that of other methods developed by Harizanov et. al. [19] and Bonito and Pasciak [8].

Key words: fractional diffusion reaction, best uniform rational approximation, error analysis
AMS classification: 65F10, 65D15, 65M06, 65M60

1. Introduction

1.1. Spectral fractional powers of elliptic operators. In this paper we consider the following second order elliptic equation with homogeneous Dirichlet data:

\[
\begin{align*}
-\nabla \cdot (a(x) \nabla v(x)) &= f(x), & \text{for } x \in \Omega, \\
v(x) &= 0, & \text{for } x \in \partial \Omega.
\end{align*}
\]

(1)

Here \( \Omega \) is a bounded domain in \( \mathbb{R}^d \), \( d \geq 1 \), and we assume that \( 0 < a_0 \leq a(x) \) for \( x \in \Omega \).

The fractional powers of the elliptic operator associated with the problem (1) are defined in terms of the weak form of (1), namely, \( v(x) \) is the unique function in \( V = H^1_0(\Omega) \) satisfying

\[
a(v, \theta) = (f, \theta) \quad \text{for all } \theta \in V.
\]

\( a(v, \theta) = \int_\Omega a(x) \nabla v(x) \cdot \nabla \theta(x) \, dx \)

\( (f, \theta) = \int_\Omega f(x) \theta(x) \, dx \)
Here
\[ a(w, \theta) := \int_{\Omega} a(x) \nabla w(x) \cdot \nabla \theta(x) \, dx \quad \text{and} \quad (w, \theta) := \int_{\Omega} w(x) \theta(x) \, dx. \]
For \( f \in X := L^2(\Omega) \), (2) defines a solution operator \( Tf := v \). Following [26], we define an unbounded operator \( A \) on \( X \) as follows. The operator \( A \) with domain
\[ D(A) = \{ Tf : f \in X \} \]
is defined by \( Av = g \) for \( v \in D(A) \) where \( g \in X \) with \( Tg = v \). This is well defined as \( T \) is injective.

Thus, the focus of our work in this paper is numerical approximation and algorithm development for the equation:
\[ A^\alpha u = f \quad \text{with a solution} \quad u = A^{-\alpha} f. \]
Here \( A^{-\alpha} = T^\alpha \) for \( \alpha > 0 \) is defined by Dunford-Taylor integrals which can be transformed when \( \alpha \in (0,1) \), to the Balakrishnan integral, e.g. [4]: for \( f \in X \),
\[ u = A^{-\alpha} f = \frac{\sin(\pi \alpha)}{\pi} \int_{0}^{\infty} \mu^{-\alpha}(\mu I + A)^{-1} f \, d\mu. \]
This definition is sometimes referred to as the spectral definition of fractional powers. One can also use an equivalent definition through the expansion with respect to the eigenfunctions of \( A \), e.g. [28, 3]. We note that there are also problems on bounded domains involving fractional powers, for example, those related to Lévy diffusion [3, 33]. These problems involve the restriction of non-local operators defined on \( \mathbb{R}^d \) applied to bounded domain functions extended, e.g., by 0, outside of \( \Omega \). However, in this paper, we focus on the spectral definition (4) and the corresponding approximations by the finite element or finite difference methods.

An operator \( L \) is positivity preserving if \(Lf \geq 0 \) when \( f \geq 0 \). We note that by the maximum principle, \((\mu I + A)^{-1}\) is a positivity preserving operator for \( \mu \geq 0 \) and the formula (4) shows that \( A^{-\alpha} \) is also. In many applications, it is important that the discrete approximations share this property.

1.2. Some semi-discrete schemes. We study approximations to \( u = A^{-\alpha} f \) defined in terms of finite difference or finite element approximation of the operator \( T \). We shall use the following convention regarding the approximate solutions by these two methods. The finite element solution is a function in \( V_h \), an \( N \)-dimensional space of continuous piece-wise linear functions over a partition \( \mathcal{T}_h \) of the domain. Such functions will be denoted by \( u_h, v_h \), etc. Also we shall denote by \( \bar{A}, \bar{I}, \) etc operators acting on the elements \( u_h, \theta_h \), etc in the finite dimensional space of functions \( V_h \). When a nodal basis of the finite element space is introduced, then the vector coefficients in this basis are denoted \( \bar{u}_h, \bar{v}_h \), etc. Under this convention operator equations in \( V_h \) such as \( \bar{A}u_h = f_h \) will be written as a system of linear algebraic equations \( \bar{A}\bar{u}_h = \bar{f}_h \in \mathbb{R}^N \).

In the finite difference case, discrete solutions are vectors in \( \mathbb{R}^N \) and are also denoted \( \bar{u}_h, \bar{v}_h \), etc. Then the corresponding counterparts of operators action on these vectors are denoted by \( \bar{A}, \bar{I}, \) etc. The finite difference approximation. In this case the approximation \( \bar{u}_h \in \mathbb{R}^N \) of \( u \) is given by
\[ \bar{A}^{\alpha}\bar{u}_h = \bar{I}_h f := \bar{f}_h, \quad \text{or equivalently} \quad \bar{u}_h = \bar{A}^{-\alpha} \bar{f}_h, \]
where \( \bar{A} \) is an \( N \times N \) symmetric and positive definite matrix coming from a finite difference approximation to the differential operator appearing in (1), \( \bar{u}_h \) is the vector in \( \mathbb{R}^N \) of the approximate solution at the interior \( N \) grid points, and \( \bar{I}_h f := \bar{f}_h \in \mathbb{R}^N \) denotes the vector of the values of the data \( f \) at the grid points. Examples of such matrices are given in Subsection 3.1.
The finite element approximation. The approximation in the finite element case is defined in terms of a conforming finite dimensional space $V_h \subset V$ of piece-wise linear functions over a quasi-uniform partition $\mathcal{T}_h$ of $\Omega$ into $d$-simplices (intervals, triangles, and simplices in 1-D, 2-D, and 3-D, respectively). Note that the construction (4) of negative fractional powers carries over to the finite dimensional case, replacing $V$ and $X$ by $V_h$ with $a(\cdot, \cdot)$ and $(\cdot, \cdot)$ unchanged.

The discrete operator $\hat{A}$ is defined to be the inverse of $T_h : V_h \to V_h$ with $T_h g_h := v_h$ where $v_h \in V_h$ is the unique solution to
\begin{equation}
\tag{6}
a(v_h, \theta_h) = (g_h, \theta_h), \quad \text{for all } \theta_h \in V_h.
\end{equation}
The finite element approximation $u_h \in V_h$ of $u$ is then given by
\begin{equation}
\tag{7}
\hat{A}^{\alpha} u_h = \pi_h f, \quad \text{or equivalently} \quad u_h = \hat{A}^{-\alpha} \pi_h f := \hat{A}^{-\alpha} f_h,
\end{equation}
where $\pi_h$ denotes the $L^2(\Omega)$ projection into $V_h$. In this case, $N$ denotes the dimension of the space $V_h$ and equals the number of (interior) degrees of freedom. The operator $\hat{A}$ in the finite element case is a map of $V_h$ into $V_h$ so that $\hat{A} v_h := g_h$, where $g_h \in V_h$ is the unique solution to
\begin{equation}
\tag{8}
(g_h, \theta_h) = a(v_h, \theta_h), \quad \text{for all } \theta_h \in V_h.
\end{equation}
Let $\{\phi_j\}$ denote the standard “nodal” basis of $V_h$. In terms of this basis
\begin{equation}
\tag{9}
\hat{A} \text{ corresponds to the matrix } \hat{A} = \hat{M}^{-1/2} \hat{S}, \quad \text{where } \hat{S}_{i,j} = a(\phi_i, \phi_j), \quad \hat{M}_{i,j} = (\phi_i, \phi_j).
\end{equation}
In the terminology of the finite element method, $\hat{M}$ and $\hat{S}$ are the mass (consistent mass) and stiffness matrices, respectively.

Obviously, if $\theta = \hat{A} \eta$ and $\tilde{\theta}, \tilde{\eta} \in \mathbb{R}^N$ are the coefficient vectors corresponding to $\theta, \eta \in V_h$, then $\tilde{\theta} = \hat{A} \tilde{\eta}$. Now, for the coefficient vector $\tilde{f}_h$ corresponding to $f_h = \pi_h f$ we have $\tilde{f}_h = \hat{M}^{-1} \tilde{F}$, where $\tilde{F}$ is the vector with entries
\[ \tilde{F}_j = (f, \phi_j), \quad \text{for } j = 1, 2, \ldots, N. \]
Then using vector notation so that $\tilde{u}_h$ is the coefficient vector representing the solution $u_h$ through the nodal basis, we can write the finite element approximation of (1) in the form of system
\begin{equation}
\tag{10}
\tilde{A} \tilde{u}_h = \hat{M}^{-1} \tilde{F} \quad \text{which implies} \quad \hat{S} \tilde{u}_h = \tilde{F}.
\end{equation}
Consequently, the finite element approximation of the sub-diffusion problem (7) becomes
\begin{equation}
\tag{11}
\hat{M} \hat{A}^{\alpha} \tilde{u}_h = \tilde{F} \quad \text{or equivalently} \quad \tilde{u}_h = \hat{A}^{-\alpha} \hat{M}^{-1} \tilde{F}.
\end{equation}

The lumped mass finite element approximation. We shall also introduce the finite element method with “mass lumping” for two reasons. First, it leads to positivity preserving fully discrete methods (see, Section 2.4). Second, it is well known that lumped mass schemes for linear elements on uniform rectangular meshes are equivalent to the simplest finite difference approximations. In fact, as shown later, the matrix (41) of the finite difference approximation of 1-D problem is the same as the matrix of the lumped finite element method for linear elements. Therefore, the theoretical study of the lumped mass method answers the question about the convergence of the finite difference method for solving the problem (3), an outstanding issue in this area.

We introduce the lumped mass (discrete) inner product $(\cdot, \cdot)_h$ on $V_h$ in following way (see, e.g. [39, pp. 239–242]) for $d$-simplices in $\mathbb{R}^d$:
\begin{equation}
\tag{12}
(z, v)_h = \frac{1}{d+1} \sum_{\tau \in \mathcal{T}_h} \sum_{i=1}^{d+1} |\tau| z(P_i) v(P_i) \quad \text{and} \quad \hat{M}_h = \{(\phi_i, \phi_k)_h\}^{N}_{i,k}.
\end{equation}
Here \( P_1, \ldots, P_{d+1} \) are the vertexes of the \( d \)-simplex \( \tau \) and \( |\tau| \) is its \( d \)-dimensional measure. The matrix \( \widetilde{M}_h \) is called lumped mass matrix. Simply, the “lumped mass” inner product is defined by replacing the integrals determining the finite element mass matrix by local quadrature approximation, specifically, the quadrature defined by summing values at the vertices of a triangle weighted by the area of the triangle.

In this case, we define \( \Lambda \) by \( \Lambda v_h := g_h \) where \( g_h \in V_h \) is the unique solution to
\[
(g_h, \theta_h)_h = a(v_h, \theta_h), \quad \text{for all } \theta_h \in V_h
\]
so that
\[
\Lambda \text{ corresponds to the matrix } \widetilde{\Lambda} = \widetilde{M}_h^{-1} \widetilde{S}, \quad \text{where } \widetilde{M}_h = \{(\phi_i, \phi_k)_h\}_{i,k}^N.
\]
Here \( \widetilde{M}_h \) is the lumped mass semi-discrete approximation is given by
\[
u_h = \Lambda^{-\alpha} \mathcal{I}_h f := f_h \quad \text{or} \quad \tilde{u}_h = \tilde{\Lambda}^{-\alpha} \tilde{F}.
\]
Here \( \tilde{F} \) is the coefficient vector in the representation of the function \( \mathcal{I}_h f \) with respect to the nodal basis in \( V_h \). We shall call \( \tilde{u}_h \) in (5) and \( u_h \) in (7) and (15) semi-discrete approximations of \( u \).

We note that the matrix \( \Lambda \in \mathbb{R}^{N \times N}, N = O(h^{-d}) \), produced by the standard finite element or finite difference method is positive definite, large, sparse, with a condition number growing like \( h^{-2} \) as \( h \to 0 \). We shall assume in this paper, that the systems of the type \( (\Lambda + cI)u_h = f_h, \ c \geq 0 \) and \( u_h, f_h \in \mathbb{R}^N \) can be solved approximately in an optimal way, namely, by an algorithm that requires \( O(N) \) arithmetic operations. This could be achieved by using fast solution methods based on multi-grid, multi-level, domain decomposition, or other techniques. The aim of our paper is to construct a solution method for (15) with optimal computational complexity \( O(N) \).

We also note that fractional powers of a symmetric and positive definite matrix are well defined by matrix diagonalization so we can write
\[
\widetilde{\Lambda} = \Xi \Lambda \Xi
\]
with \( \Xi \) an orthogonal matrix and \( \Lambda \) a diagonal matrix with entries, \( \Lambda_{ii} = \lambda_i \) where \( 0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N \) are the eigenvalues of \( \Lambda \). In this case,
\[
\widehat{\Lambda}^{-\alpha} = \Xi^{-t} \Lambda^{-\alpha} \Xi.
\]
Of course, \( \Lambda^{-\alpha} \) is a diagonal matrix with diagonal entries \( \lambda_i^{-\alpha}, \ i = 1, 2, \ldots, N \).

The direct computation of \( \tilde{u}_h \) involves the computation of the eigenvalues and eigenvectors of the matrix \( \tilde{\Lambda} \). Such computation using this factorization is, generally, quite expensive, except for a very narrow class of equations with constant coefficients on rectangular domains. Similar techniques can be employed in both the standard mass and lumped mass finite element cases, but requires expansion in a basis of eigenvectors satisfying generalized eigenvector problems involving the matrices \( \overline{M}, \tilde{M}_h \) and \( \overline{S} \) and, again, direct computation is quite expensive.

Nevertheless, such approach could be made quite efficient for approximation of the corresponding elliptic operator by a spectral numerical method in simple domains, e.g. [36]. For such problems the spectral methods are known to be very accurate due to exponential convergence rate with respect to the number of the degrees of freedom. Such examples on square domains are presented in [36]. Although, the case of spectral approximation on a disk domain is discussed there, the application of their discretization to the fractional power problem would require computing the generalized eigenvectors for which fast methods are not available. In contrast, their discretization would be an ideal candidate for the method discussed here and only limited by the availability of fast solvers for the stationary problem. In our paper, the targeted area is a steady state problem.
in a complex domain with low regularity solution discretized by standard finite element or finite
difference method, naturally leading to large scale linear systems.

1.3. Fully discrete schemes based on the best uniform rational approximation. Here we
will introduce approximations of \( u_h = \mathcal{A}^{-\alpha} f_h \) by employing best rational approximations (BURA)
to \( t^\gamma \) on \([0, 1]\) with \( \gamma > 0 \). Specifically, we consider BURA along the diagonal of the Walsh table and
take \( \mathcal{R}_k \) to be the set of rational functions of the form \( P_k(t)/Q_k(t) \) with \( P_k(t) \) and \( Q_k(t) \) polynomials
of degree \( k \) and \( Q_k(0) = 1 \). The best rational approximation (BURA) of \( t^\gamma \) is the rational function
\( r_{\gamma,k} \in \mathcal{R}_k \) satisfying

\[
(17) \quad r_{\gamma,k}(t) := \arg\min_{s(t) \in \mathcal{R}_k} \| s(t) - t^\gamma \|_{L^\infty[0,1]}.
\]

Denoting the error by

\[
E_{\gamma,k} := \| r_{\gamma,k}(t) - t^\gamma \|_{L^\infty[0,1]},
\]

we apply Theorem 1 of [37] to claim that there is a constant \( C_\gamma > 0 \), independent of \( k \), such that

\[
(18) \quad E_{\gamma,k} \leq C_\gamma e^{-2\pi \sqrt{k}}.
\]

Thus, the BURA error converges exponentially to zero as \( k \) becomes large.

Rescaling and the semi-discrete approximation. We rescale the equations (5), (7) and (15):

\[
(19) \quad \tilde{u}_h = \lambda_1^{-\alpha}(\lambda_1 \tilde{\mathcal{A}}^{-1})^\alpha \tilde{f}_h
\]

where \( \lambda_1 \) denotes the smallest eigenvalue of \( \mathcal{A} \) in (5), (7) and (14), respectively. The scaling by \( \lambda_1 \)
maps the eigenvalues of \( \lambda_1 \tilde{\mathcal{A}}^{-1} \) to the interval \((0, 1] \).

We note that instead of scaling with \( \lambda_1 \), we can scale with any \( \delta \in (0, \lambda_1] \). In this case the eigenvalues of \( \delta \tilde{\mathcal{A}}^{-1} \) will be again in the interval \((0, 1] \) and the method will work in the same way.

This will allow to use any lower bound for the eigenvalue \( \lambda_1 \). Such a bound could be obtained using
the coercivity of the form \( a(\cdot, \cdot) \) in \( V \) and the Poincaré-Friedrichs inequality. For example, if \( \mathcal{A} \) is
obtained by finite element discretization of \(-\Delta \) in a bounded, convex, Lipschitz domain \( \Omega \subset \mathbb{R}^d \) with
a diameter \( d_\Omega = \text{diam}(\Omega) \) with Dirichlet boundary conditions, then the bound \( \lambda_1 \geq C_P = \frac{\pi^2}{d_\Omega^2} \)
follows from the Poincaré-Friedrichs inequality: \( C_P \| u \|_{L^2(\Omega)} \leq \| \nabla u \|_{L^2(\Omega)} \) for all \( u \in H^1_0(\Omega) \), [32, inequality (1.9)].

Another possibility is to find an estimate for \( \lambda_1 \) by running few iterations of the inverse power method.

Now we introduce the fully discrete approximations: \( w_h \in V_h \) of the finite element approximation
\( u_h \in \mathcal{V}_h \in \mathbb{R}^N \) of the finite difference approximation \( \tilde{u}_h \in \mathbb{R}^N \) by

\[
(20) \quad w_h = \lambda_1^{-\alpha} r_{\alpha,k}(\lambda_1 \tilde{\mathcal{A}}^{-1}) f_h \quad \text{and} \quad \tilde{w}_h = \lambda_1^{-\alpha} r_{\alpha,k}(\lambda_1 \tilde{\mathcal{A}}^{-1}) \tilde{f}_h.
\]

Here \( \lambda_1 \) and \( f_h \) are as in (7) or (15) and \( \tilde{\mathcal{A}} \) and \( \tilde{f}_h \) are as in (5).

In Section 2, we study the error of these fully discrete solutions. For the finite element case we obtain the error estimate

\[
(21) \quad \| u_h - w_h \| \leq \lambda_1^{-\alpha} E_{\alpha,k} \| f_h \|
\]

with \( \| \cdot \| \) denoting the norm in \( L^2(\Omega) \). In the finite difference case, we have

\[
(22) \quad \| \tilde{u}_h - \tilde{w}_h \|_{\ell_2} \leq \lambda_1^{-\alpha} E_{\alpha,k} \| \tilde{f}_h \|_{\ell_2}
\]

where the norm \( \| \cdot \|_{\ell_2} \) denotes the Euclidean norm in \( \mathbb{R}^N \).

We note that the schemes of [19] are closely related to our scheme. These were given for the finite difference case by writing

\[
\tilde{u}_h = \tilde{\mathcal{A}}^{-p}(\tilde{\mathcal{A}}^{p-\alpha}) \tilde{f}_h = \lambda_1^{\alpha-p} \tilde{\mathcal{A}}^{-p}(\lambda_1)^{p-\alpha} \tilde{f}_h, \quad p = 1, 2.
\]
Their approximation becomes

\[
\tilde{u}_h = \lambda_N^{p-\alpha} \tilde{A}^{-p} r_{p-\alpha,k}(\tilde{A} / \lambda_N) \tilde{f}_h
\]

that implies

\[
\| \tilde{u}_h - \tilde{w}_h \|_{\ell^2} \leq \lambda_N^{p-\alpha} E \| \tilde{f}_h \|_{\ell^2}.
\]

The main disadvantage of this method compared to ours is that \( \lambda_N \) grows on the order of \( h_{min}^{-2} \), with \( h_{min} \) denoting the minimal distance between mesh points so the factor of \( \lambda_N^{p-\alpha} \) deteriorates the convergence rate. This is especially harmful when local mesh refinement is used. In contrast, \( \lambda_1 \) is related to the constant in the Poincaré inequality and remains bounded away from zero independently of the mesh parameter so the appearance of \( \lambda_1^{p-\alpha} \) in our method is harmless.

**Existing solution methods for fractional powers of SPD matrices.** Due to the current interest of the computational mathematics and physics communities in modeling and simulations involving fractional powers of elliptic operators, a number of approaches and algorithms has been developed, studied, and tested on various problems, \([2, 30, 5, 8, 31, 25]\). However, the goal of this paper is to develop efficient methods for solving large systems (hundreds of thousands or even millions of unknowns) algebraic equations (5) that utilize efficient methods for solving the system \( \tilde{A} \tilde{u}_h = \tilde{f}_h \).

Below we make a concise survey of such methods.

1. In the finite difference case, \( \tilde{u}_h \) is expressed though a fractional power of a symmetric and positive definite matrix. We can look at this problem as a particular case of the well established methods of stable computations of the matrix square root or/and other functions of matrices, see, e.g. \([15, 24, 27]\). Often these are based on Newton iteration with suitable Padé stabilization. Application of such approach is limited to small-size matrices.

2. An extension of the problem from \( \Omega \subset \mathbb{R}^d \) to a problem in \( \Omega \times (0, \infty) \subset \mathbb{R}^{d+1} \), see, e.g. \([10]\). Nochetto and co-authors in \([30, 31]\) developed efficient computational method based on finite element discretization of the extended problem and subsequent use of multi-grid technique. The main deficiency of the method is that instead of problem in \( \mathbb{R}^d \), one needs to work in a domain in one dimension higher which adds to the complexity of the developed algorithms.

3. Reformulation of the problem as a pseudo-parabolic on the cylinder \( (0, 1) \times \Omega \) by adding a time variable \( t \in (0, 1) \). Such methods were proposed, developed, and tested by Vabishchevich in \([41, 42]\). As shown in the numerical experiments in \([25]\), the method is very slow when using uniform time stepping. However, the improvement proposed in \([16, 13]\) make this method quite competitive.

4. Approximation of the Dunford-Taylor integral representation of the solution of equations involving fractional powers of elliptic operators, proposed in the pioneering paper of Bonito and Pasciak \([8]\). Further the idea was extended and augmented in various directions in \([1, 9, 7, 6]\). These methods use exponentially convergent sinc quadratures.

5. Best uniform rational approximation of the function \( t^\alpha \) on \([0, 1]\), proposed in \([23, 19]\), further developed in \([22, 20, 21]\) and called BURA methods. In this paper we propose, analyze, and test a new method given by (20) that is based on rescaling the problem with the smallest eigenvalue of matrix \( \tilde{A} \) (or the operator \( A \)).

As shown recently in \([25]\), in appearance different, these methods are interrelated and all seem to involve some rational approximation of fractional powers of the underlying elliptic operator, see, e.g. \([1, 5, 8]\). In the mentioned above works the numerical algorithm results in a rational approximation of \( \mathcal{A}^{-\alpha} f \) where the elliptic operator \( \mathcal{A} \) is replaced by some approximation by finite elements. The algorithm we propose and study in this paper is based on the best uniform rational approximation and in principle should be at least as good as any of these methods. In fact, our comparisons show that in many cases the proposed method performs significantly better.

However, one should realize that BURA-based methods involve Remez method of finding the best uniform rational approximation by solving the highly non-linear min-max problem (17). It is
well known that Remez algorithm is very sensitive to the precision of the computer arithmetic, cf. [29, 43, 14]. One of the main reasons is that almost all extreme points of the error function tend to the origin as \( k \to \infty \), cf. [34, Theorem 4]. Various techniques for stabilization of the method have been used, mostly by using Tchebyshev orthogonal polynomials, cf. [14]. It is observed that to achieve high accuracy one needs to use high arithmetic precision. For example, in [43] the first 25 correct decimal digits of the BURA error of \( t^\alpha \) for six values of \( \alpha \in [0, 1] \) are reported for degree up to \( k = 30 \) by using computer arithmetic with 200 significant digits.

### Positivity preserving schemes.

In the finite difference case, if \( f(x) \geq 0 \) for \( x \in \Omega \) then the vector \( \tilde{u}_h \) defined by (5) has non-negative entries. The issue of positivity preservation of approximate solution of problems involving the spectral fractional Laplacian by finite difference method was first discussed and established for the BURA scheme (23) in [22]. We will show that the solution \( \tilde{w}_h \) of (20) has non-negative entries as well. However, in the finite element case, we will show that \( u_h(x) \) can have negative values even when \( f \) is non-negative, i.e. the consistent finite element approximation may loose non-negativity of the solution. Instead, we shall use schemes obtained by mass lumping discussed in Subsection 1.2. Note that in (7) we also replace \( \pi_h \) by the interpolant \( \mathcal{I}_h : C^0(\Omega) \to V_h \). This leads to a semi-discrete solution

\[
\begin{align}
\tilde{u}_h &= \lambda_1^{-\alpha} \mathcal{I}_hf = \lambda_1^{-\alpha}(\lambda_1 A^{-1})^\alpha \mathcal{I}_hf \\
\tilde{w}_h &= \lambda_1^{-\alpha} r_{\alpha,k}(\lambda_1 A^{-1}) \mathcal{I}_hf
\end{align}
\]

and its BURA approximation (fully discrete approximation)

\[
\begin{align}
\tilde{w}_h &= \lambda_1^{-\alpha} r_{\alpha,k}(\lambda_1 A^{-1}) \mathcal{I}_hf
\end{align}
\]

with \( A \) defined by (14). In this case, \( \tilde{w}_h \) is non-negative when \( f \) is non-negative and most of the approximation properties of \( u_h \) are still preserved (see, Theorem 4.2).

### 1.4. Organization of the paper and our contributions.

Section 2 examines the implementation of (20) for both, the finite element (7) and finite difference approximations (5), and also proves the estimates (21) and (22) for their error. Here we also consider the lumped mass method and discuss the non-negativity of the solution produced by non-negative data. In Section 3 we give several matrices obtained by finite difference method and perform some extensive computations on a number of test problems in one and two spatial dimensions. We compare the accuracy of the proposed in this paper new method, called P-BURA, with the BURA method (23) of [19] and with the method of Bonito and Pasciak, [8], on two 2-dimensional model problems with smooth (Table 3) and non-smooth right hand sides (Table 2). Further in Section 3 we study the efficiency of the method on some non-uniform meshes refined locally in order to capture the interior layers of the solution. The results are reported in Tables 4 and 5. Section 4 focuses on the finite element approximations. In Theorem 4.2 we provide error estimates for \( u - u_h \) both in the consistent mass finite element method (cf. [8]) and the case of mass lumping. As a consequence, in Corollary 4.4 we establish an error bound for the finite difference approximation of boundary value problem for the spectral fractional elliptic equation. To best of our knowledge, error bounds for the approximations of the spectral fractional Laplacian by finite differences is not available.

The main contributions of this paper are: (1) derivation an analysis of an efficient BURA method for solving systems of equations \( \tilde{A}^\alpha \tilde{u}_h = \tilde{f}_h, \quad 0 < \alpha < 1 \), in \( \mathbb{R}^N \), where \( \tilde{A} \) is a symmetric and positive sparse matrix obtained from finite difference or finite element approximations of elliptic operators; (2) analysis of lumped mass schemes that lead to positivity preserving methods; (3) estimates of approximation error of the spectral fractional Laplacian by finite differences.
2. Implementation and basic estimates of the error

2.1. Properties of the best uniform rational approximation. In this section, we discuss the implementation of (20) in the finite difference and finite element cases.

First, on Table 1 we present the computed error of BURA \( r_{\alpha,k}(t) \) of \( t^\alpha \) using the modified Remez algorithm, e.g. [19]. As expected, the approximation error for large \( \alpha \) is in the very reasonable range of \( 10^{-5} - 10^{-7} \) for \( k = 7 - 10 \). Moreover, for this range of \( k \) the Remez algorithm is relatively stable and the coefficients of BURA function \( r_{\alpha,k}(t) \) are determined with good accuracy.

| \( \alpha \) | \( E_{\alpha,5} \) | \( E_{\alpha,6} \) | \( E_{\alpha,7} \) | \( E_{\alpha,8} \) | \( E_{\alpha,9} \) | \( E_{\alpha,10} \) |
|---|---|---|---|---|---|---|
| 0.75 | 2.8676e-5 | 9.2522e-6 | 3.2566e-6 | 1.2288e-6 | 4.9096e-7 | 2.0584e-7 |
| 0.50 | 2.6896e-4 | 1.0747e-4 | 4.6037e-5 | 2.0852e-5 | 9.8893e-6 | 4.8760e-6 |
| 0.25 | 2.7348e-3 | 1.4312e-3 | 7.8650e-4 | 4.4950e-4 | 2.6536e-4 | 1.6100e-4 |

Next, we prove the estimates (21) and (22). It is known that the best rational approximation \( r_{\alpha,k}(x) = \frac{P(x)}{Q(x)} \) for \( \alpha \in (0,1) \) is non-degenerate, i.e., the polynomials \( P \) and \( Q \) are of full degree. Let the roots of \( P \) and \( Q \) be denoted by \( \zeta_1, \ldots, \zeta_k \) and \( d_1, \ldots, d_k \), respectively. It is shown in [34, 38] that the roots interlace and satisfy

\[
0 > \zeta_1 > d_1 > \zeta_2 > d_2 > \cdots > \zeta_k > d_k.
\]

We then have

\[
r_{\alpha,k}(t) = b \prod_{i=1}^{k} \frac{t - \zeta_i}{t - d_i}
\]

where, by (26) and the fact that \( r_{\alpha,k} \) is a best approximation to a non-negative function, \( b > 0 \) and \( P(x) > 0 \) and \( Q(x) > 0 \) for \( x \geq 0 \).

We consider \( \tilde{r}_{\alpha,k} \) defined by

\[
\tilde{r}_{\alpha,k}(\lambda) := r_{\alpha,k}(1/\lambda) = \frac{\tilde{P}(\lambda)}{\tilde{Q}(\lambda)}.
\]

Here \( \tilde{P}(\lambda) = \lambda^k P(\lambda^{-1}) \) and \( \tilde{Q}(\lambda) = \lambda^k Q(\lambda^{-1}) \) and hence their coefficients are defined by reversing the order of the coefficients in \( P, Q \) appearing in \( r_{\alpha,k} \). In addition, (26) implies

\[
0 > \tilde{d}_k > \tilde{\zeta}_k > \tilde{d}_{k-1} > \tilde{\zeta}_{k-1} > \cdots > \tilde{d}_1 > \tilde{\zeta}_1.
\]

Here \( \tilde{d}_i = 1/d_i \) and \( \tilde{\zeta}_i = 1/\zeta_i \) are the roots of \( \tilde{P} \) and \( \tilde{Q} \), respectively.

**Proposition 2.1.** For \( \alpha \in (0,1) \),

\[
\tilde{r}_{\alpha,k}(\lambda) = c_0 + \sum_{i=1}^{k} \frac{c_i}{\lambda - \tilde{d}_i}
\]

where \( c_i > 0 \) for \( i = 0, 1, \ldots, k \).

**Proof.** We note that

\[
c_0 = \lim_{\lambda \to \infty} \tilde{r}_{\alpha,k}(\lambda) = \lim_{x \to 0} r_{\alpha,k}(x) = b \prod_{i=1}^{k} \frac{\zeta_i}{d_i} > 0.
\]
The remaining coefficients in (29) are determined by the equation
\[ \tilde{P}(\lambda) = \tilde{Q}(\lambda) \left[ c_0 + \sum_{i=1}^{k} \frac{c_i}{\lambda - \tilde{d}_i} \right] \]
which when evaluated at \( \tilde{d}_i \) implies
\[ (30) \quad \tilde{P}(\tilde{d}_i) = c_i \prod_{j \neq i} (\tilde{d}_i - \tilde{d}_j). \]
It follows from (28) that both the signs of \( \tilde{P}(\tilde{d}_i) \) and those of the product on the right hand side of (30) oscillate with \( i \). In addition, (28) also implies that the product in (30) is positive for \( i = k \) and the sign of \( \tilde{P}(\tilde{d}_k) \) is the same as that of \( \tilde{P}(1) = P(1) > 0 \).

It follows that \( c_i > 0 \), for \( i = 1, 2, \ldots, k \).

Now consider the implementation of (20). By (11), the coefficients vector \( \tilde{w}_h \) of the representation of \( w_h \) with respect to the nodal basis in \( V_h \) is given by
\[ (31) \quad \tilde{w}_h = \lambda_1^{-\alpha} \tilde{r}_{\alpha,k}(\lambda_1^{-1}\tilde{A})\tilde{M}^{-1}\tilde{F} \]
with \( \tilde{A} \) denoting the matrix in (9). Applying Proposition 2.1 gives
\[ (32) \quad \tilde{w}_h = \lambda_1^{-\alpha} \left( c_0 \tilde{M}^{-1} \tilde{F} + \sum_{i=1}^{k} (\lambda_1 c_i)(\tilde{A} - \lambda_1 \tilde{d}_i \tilde{M})^{-1} \tilde{M}^{-1} \tilde{F} \right) \]
We note that even though \( \tilde{F} \) is a non-negative vector when \( f \) is non-negative, the matrix \( (\tilde{S} - \lambda_1 \tilde{d}_i \tilde{M})^{-1} \) is not positivity preserving when \( -\lambda_1 \tilde{d}_i \) is large. This is because even if \( \tilde{S} \) is an \( M \)-matrix, \( \tilde{S} + \gamma \tilde{M} \) for large \( \gamma \) becomes a matrix with the same sparsity pattern where every matrix entry in the pattern is positive. The inverse of such a matrix is NOT positivity preserving. As the matrix \( \tilde{M} \) has positive entries, its inverse appearing in the first term above is also not positive preserving.

In the finite difference case, \( \tilde{w}_h \) satisfies (31) with \( \tilde{M}f_h \) replaced by \( \tilde{f}_h \) and \( \tilde{A} \) denoting the finite difference matrix. Applying Proposition 2.1 gives
\[ (33) \quad \tilde{w}_h = \lambda_1^{-\alpha} \left( c_0 \tilde{f}_h + \sum_{i=1}^{k} (\lambda_1 c_i)(\tilde{A} - \lambda_1 \tilde{d}_i \tilde{M})^{-1} \tilde{f}_h \right) \]
The finite difference matrix \( \tilde{A} \) is generally an \( M \)-matrix and we have the following theorem (is a consequence of Proposition 2.1 and (33)).

**Proposition 2.2.** Assume that the finite difference matrix \( \tilde{A} \) is an \( M \)-matrix, i.e. all diagonal entries are positive and all non-diagonal entries are non-positive. If \( \tilde{f}_h \) has all its entries non-negative then the solution \( \tilde{w}_h \) represented by (32) has all its entries non-negative, i.e. the method is positivity preserving.

\[ \text{This, in turn, implies that the matrix } \tilde{A} \text{ in the finite element case CANNOT be an } M \text{-matrix.} \]
The above sum is trivially parallelizable as the result of each term is independent of all others. Alternatively, by (27),
\[
\tilde{r}_{\alpha,k}(\lambda) = b \prod_{i=1}^{k} \frac{1 - \lambda_i}{1 - \lambda d_i}
\]
and hence
\[
w_h = \lambda_1^{-\alpha} b \left[ \prod_{i=1}^{k} (\lambda_1 - \tilde{\zeta}_i A) (\lambda_1 - d_i A)^{-1} \right] f_h.
\]
This product needs to be computed sequentially, computing the \( j \)-term product by applying the \( j \)'th operator to the \( j - 1 \)'st term product.

We note that both, the additive (32) and the multiplicative (34) versions of the method led to stable computations. Indeed, due to the interlacing properties (26) the \( L^2 \)-norm of \((\lambda_1 \mathbb{I} - \tilde{\zeta}_i A)(\lambda_1 \mathbb{I} - d_i A)^{-1}\) is less than 1 and the product computation is stable. Similarly, since \( c_j > 0 \), \( j = 0, \ldots, k \) the summation in (32) is stable.

2.2. Finite difference scheme. We first consider the finite difference case. In this case, each term in the sum (32) requires a sparse matrix solve that involves a matrix which is a sum of \( A \) involving a sparse matrix multiply and a sparse matrix solve.

We next show that (22) holds. We note that
\[
\tilde{u}_h - \tilde{w}_h = \lambda_1^{-\alpha} \left[ (\lambda_1 \tilde{A}^{-1})^\alpha - r_{\alpha,k}(\lambda_1 \tilde{A}^{-1}) \right] \tilde{I}_h f := \lambda_1^{-\alpha} \mathcal{G} \tilde{I}_h f.
\]
The above matrix \( \mathcal{G} = (\lambda_1 \tilde{A}^{-1})^\alpha - r_{\alpha,k}(\lambda_1 \tilde{A}^{-1}) \) is symmetric and hence
\[
\|\tilde{u}_h - \tilde{w}_h\|_{\ell_2} \leq \lambda_1^{-\alpha} \rho(\mathcal{G}) \|\tilde{I}_h f\|_{\ell_2}
\]
with \( \rho(\mathcal{G}) \) denoting the spectral radius of \( \mathcal{G} \). The inequality (22) follows from noting that the eigenvalues of \( \mathcal{G} \) come from those of \( A \), i.e.,
\[
\rho(\mathcal{G}) = \max_{i=1}^{N} |(\lambda_1 / \lambda_i)^\alpha - r_{\alpha,k}(\lambda_1 / \lambda_i)| \leq \max_{1 \leq t \leq \infty} |t^\alpha - r_{\alpha,k}(t)| = E_{\alpha,k}.
\]

2.3. Consistent mass finite element method. The implementation of finite element problems is done in terms of stiffness and mass matrices denoted by \( \tilde{S} \) and \( \tilde{M} \), respectively, and vectors in \( \mathbb{R}^N \) where \( N \) is the dimension of \( V_h \) defined by (9). Then the coefficients \( \tilde{w}_h \) for the function \( w_h \) using (32) are given by
\[
\tilde{w}_h = \lambda_1^{-\alpha} \left( c_0 \tilde{M}^{-1} \tilde{F} + \sum_{i=1}^{k} (\lambda_1 c_i)(\tilde{S} - \lambda_1 \tilde{d}_i \tilde{M})^{-1} \tilde{F} \right).
\]
Similarly, for the product case we get
\[
\tilde{w}_h = \lambda_1^{-\alpha} b \left[ \prod_{i=1}^{k} (\lambda_1 \tilde{M} - d_i \tilde{S})^{-1} (\lambda_1 \tilde{M} - \tilde{\zeta}_i \tilde{S}) \right] \tilde{M}^{-1} \tilde{F}.
\]
The matrices in parenthesis appearing both in (35) and (36) are positive linear combinations of the symmetric and positive definite stiffness and mass matrices.

The validation of (21) in the finite element case is similar. Let \( \{\psi_1, \psi_2, \ldots, \psi_N\} \) be an \( L^2(\Omega) \) orthonormal basis of eigenfunctions in \( V_h \) with eigenvalues \( 0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N \) satisfying the generalized eigenvalue problem:
\[
a(\psi_i, \theta) = \lambda_i(\psi_i, \theta), \quad \text{for all } \theta \in V_h.
\]
Note that $A$ defined through $(A\psi_i, \theta) = a(\psi_i, \theta)$ has matrix representation $\tilde{M}^{-1}\tilde{S}$ so that this eigenvalue problem has equivalent algebraic form $\tilde{S}\tilde{\psi}_j = \lambda_j\tilde{M}\tilde{\psi}_j$. Expanding $\pi_h f$, $w_h$ and $u_h$ in this basis leads to

$$u_h - w_h = \lambda_1^{-\alpha} \sum_{i=1}^{n} [(\lambda_1/\lambda_i)^\alpha - r_{\alpha,k}(\lambda_1/\lambda_i)](f, \psi_i) \psi_i.$$  

The quantities in brackets above are bounded in absolute value by $E_{\alpha,k}$ and the inequality (21) follows from Parseval’s formula, i.e.,

$$\|u_h - w_h\|^2 \leq \lambda_1^{-\alpha} E_{\alpha,k}^2 \sum_{i=1}^{n} (f, \psi_i)^2 = \lambda_1^{-\alpha} E_{\alpha,k}^2 \|\pi_h f\|^2$$  

and (21) follows.

Though the consistent mass matrix $\tilde{M}$ has the same sparsity pattern as the stiffness matrix it is not an $M$-matrix. Then the formula (35) shows that even when $\tilde{S}$ is an $M$-matrix, the semidiscrete solution $\tilde{w}_h$ may not be non-negative for $f(x) \geq 0$ since $\tilde{S} + \mu\tilde{M}$ fails to be an $M$-matrix for large $\mu$. The issue of positivity preservation is discussed in more details and illustrated with numerical examples in Subsection 4.1.

2.4. Lumped mass finite element method. Since in this time $A$ has matrix representation $\tilde{M}_h^{-1}\tilde{S}$ then (35) becomes

$$\tilde{w}_h = \lambda_1^{-\alpha} \left( c_0\tilde{M}_h^{-1}\tilde{F} + \sum_{i=1}^{k} (\lambda_1 c_i)(\tilde{S} - \lambda_1\tilde{d}_i\tilde{M}_h)^{-1}\tilde{F} \right).$$  

The analysis of this scheme is the same as the analysis of the standard FEM scheme. The only difference is that now we need to use the eigenvalues and eigenfunctions of $a(\psi_i, \theta) = \lambda_i(\psi_i, \theta)_h$ for all $\theta \in V_h$, and the analysis follows easily.

The main purpose of introducing the lumped mass method is to ensure non-negativity of the fully discrete solution $\tilde{w}_h$ in case of non-negative data $f$. Due to (28) and Proposition 2.1 we have $\tilde{d}_i < 0$, and $c_i > 0$ for $i = 1, \ldots, k$. Matrix $\tilde{M}_h$ is diagonal with positive elements and representation (38) shows that if $\tilde{S}$ is an $M$-matrix, then $\tilde{S} - \lambda_1\tilde{d}_i\tilde{M}_h$, $i = 1, \ldots, k$ will be all $M$-matrices and the fully discrete solution will satisfy $\tilde{w}_h \geq 0$ if $f \geq 0$. Thus, to ensure non-negativity it is sufficient the stiffness matrix $\tilde{S}$ to be an $M$-matrix. This phenomenon is well understood in the case considered in this paper, namely, conforming linear finite elements on simplicial meshes. For $a(x) = 1$ and $d = 2$ in [12] this was shown to hold provided that the mesh triangles do not have any angles exceeding $\pi/2$. The most general result is established in [44, Lemma 2.1], namely, a sufficient and necessary condition $\tilde{S}$ to be an $M$-matrix for simplicial meshes in $\mathbb{R}^d$ for any $d \geq 2$. The condition is expressed through the angles between faces and areas of the simplex faces and improves the result from [12] for $d = 2$.

3. Finite difference approximation of the fractional diffusion problem

The linear operators we consider in this section are approximations of (1) by finite differences. We begin with some simple examples.

3.1. Example of Finite Difference Approximations. Now we give two particular examples of finite difference approximations of elliptic operators. These are used to illustrate the above theory and are also a basis of our numerical experiments.
Example 1. We first consider the one-dimensional equation (1) with variable coefficient, namely, we study the following boundary value problem \( -(a(x)u')' = f(x), \ u(0) = 0, \ u(1) = 0, \) for \( 0 < x < 1, \) where \( a(x) \) is uniformly positive function on \([0,1] \). On a uniform mesh \( x_i = ih, \ i = 0, \ldots, N+1, \) \( h = 1/(N+1), \) we consider the three-point approximation of the second derivative

\[
u''(x_i) \approx \frac{1}{h} \left( a_i \frac{u(x_{i+1}) - u(x_i)}{h} - a_i \frac{u(x_i) - u(x_{i-1})}{h} \right) \]

Here \( a_i = a(x_i - h/2) \) or \( a_i = \frac{1}{h} \int_{x_{i-1}}^{x_i} a(x) \, dx. \) Note that the former is the standard finite difference approximation obtained from the balanced method (see, e.g. [35, pp. 155–157]), while the latter is a result of finite element method with mass lumping, see Subsection 2.3.

Then the finite difference approximation in this case is the matrix equation (5) with (39)

\[
\tilde{A} = \frac{1}{h^2} \begin{bmatrix}
  a_1 + a_2 & -a_1 & & & \\
  -a_1 & a_2 + a_3 & -a_2 & & \\
  & \cdots & \cdots & \cdots & \\
  -a_1 & a_2 + a_3 & -a_2 & & \\
  & & & \cdots & \\
  & & & \cdots & \\
  -a_N & a_{N-1} + a_N & -a_N & & \\
  & & & & \cdots
\end{bmatrix}, \quad \tilde{I}_h f = \tilde{f}_h = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_N) \end{bmatrix}.
\]

The eigenvalues \( \lambda_i \) of the matrix \( \tilde{A} \) satisfy

\[
4\pi^2 \min_x a(x) \leq \lambda_i \leq 4 \max_x a(x)/h^2, \quad i = 1, \ldots, N.
\]

Example 2. The next example is for problem (1) on \( \Omega = (0,1) \times (0,1) \) on a \((n+1) \times (n+1) \) square mesh. The standard 5-point stencil finite difference approximation of the Laplace operator gives the matrix \( \tilde{A} \in \mathbb{R}^{N \times N}, \) \( N = n^2, \) that has the following block structure (here \( \tilde{A}_{i,i} \in \mathbb{R}^{n \times n}, \) \( i = 1, \cdots, n \) and \( \tilde{I}_n \) is the identity matrix in \( \mathbb{R}^{n} \))

\[
\tilde{A} = (n + 1)^2 \begin{bmatrix}
  \tilde{A}_{1,1} & -\tilde{I}_n & & & \\
  -\tilde{I}_n & \tilde{A}_{2,2} & -\tilde{I}_n & & \\
  & \cdots & \cdots & \cdots & \\
  -\tilde{I}_n & \tilde{A}_{i,i} & -\tilde{I}_n & & \\
  & & & \cdots & \\
  & & & \cdots & \\
  -\tilde{I}_n & \tilde{A}_{n,n} & -\tilde{I}_n & & \\
  & & & & \cdots
\end{bmatrix}, \quad \tilde{A}_{i,i} = \begin{bmatrix} 4 & -1 & & & \\
 -1 & 4 & -1 & & \\
 & \cdots & \cdots & \cdots & \\
 -1 & 4 & -1 & & \\
 & & & \cdots & \\
 & & & \cdots & \\
 -1 & 4 & -1 & & \\
 & & & & \cdots
\end{bmatrix}.
\]

This matrix could be obtained by the finite element method applied to triangular meshes that generated on triangulations obtained by splitting each rectangle into two triangles (by connecting the lower left vertex with the upper right one) and using the “lumped” mass inner product (12). Since the mesh is square, all diagonal elements of \( \tilde{M}_h^{-1} \) are equal to \( h^{-2} = (n + 1)^2. \) Then the operator \( \tilde{A} : V_h \to V_h \) is defined as \( (\tilde{A} u_h, v)_h = a(u_h, v) \) has a matrix representation \( \tilde{A} = \tilde{M}_h^{-1} \tilde{S}, \) see also [11, Chapter 4, p. 203–205].

Remark 3.1. We note that on an uniform mesh with step-size \( h = 1/(N+1) \) the matrix (40) has the following extreme eigenvalues:

\[
\lambda_1 = 8(n + 1)^2 \sin^2 \frac{\pi}{2(n + 1)} \approx 2\pi^2, \quad \lambda_{n^2} = 8(n + 1)^2 \sin^2 \frac{\pi n}{2(n + 1)} \approx 8(n + 1)^2 = 8h^{-2}.
\]
Now we consider smooth data techniques on a uniform mesh with mesh-size \( h \) of both the Laplace and the discrete Laplace operators, the exact discrete solution on the uniform mesh-size \( \tilde{h} \) is not continuous, \( \tilde{u} = \tilde{\Lambda}^{-\alpha} \tilde{I} \tilde{h} f_1 \), where \( \tilde{\Lambda} \) is as in (40), has been computed using FFT techniques on a uniform mesh with mesh-size \( h = 2^{-15} \).

**Example 1.** The right hand side \( f_1(x, y) \), used also in the numerical tests in \([8, 19]\), is piece-wise constant function (CheckerBoard), which has jump discontinuities along the lines \( x = 0.5 \) and \( y = 0.5 \)

\[
(42) \\
 f_1(x, y) = \begin{cases} 
 1, & \text{if } (x - 0.5)(y - 0.5) > 0, \\
 -1, & \text{if } (x - 0.5)(y - 0.5) < 0.
\end{cases}
\]

As \( f_1 \) is not continuous, \( \tilde{I} \tilde{h} f_1 \) is not well defined. Instead, we set \( \tilde{f}_h(x_j) = 0 \) at points of discontinuity and \( \tilde{f}_h(x_j) = f(x_j) \) otherwise. Here \( \{x_j\} \) are the interior nodes of the finite element mesh. A discrete reference solution of \( \tilde{u}_h = \tilde{\Lambda}^{-\alpha} \tilde{I} \tilde{h} f_1 \), where \( \tilde{\Lambda} \) is as in (40), has been computed using FFT techniques on a uniform mesh with mesh-size \( h = 2^{-15} \).

**Example 2.** Now we consider smooth data \( f_2(x, y) = \sin(2\pi x) \sin(2\pi y) \). Since \( f_2 \) is an eigenfunction of both the Laplace and the discrete Laplace operators, the exact discrete solution on the uniform mesh with mesh-size \( h \) is

\[
\tilde{u}_h = \tilde{\Lambda}^{-\alpha} \tilde{I} \tilde{h} f_2 = \left( \frac{8 \sin^2(\pi h)}{h^2} \right)^{-\alpha} \tilde{I} \tilde{h} f_2.
\]

Together with the two BURA-related solvers (BURA and P-BURA), we apply also the method, proposed by Bonito and Pasciak in \([8]\) that incorporates an exponentially convergent quadrature scheme for approximation of integral representation of the solution (4)

\[
t^{-\alpha} \approx Q_\alpha(t) := \frac{2k'}{\pi} \sum_{k=-M}^M e^{2(\alpha-1)k't} e^{-2k't}, \quad t \in (0, \infty),
\]
where \( m = \lfloor (1 - \alpha)k \rfloor, \) \( M = \lceil \alpha k \rceil, \) \( k' = \pi/(2\sqrt{\alpha(1-\alpha)k}). \) Note that \( Q_{\alpha} \) is in the class of rational functions \( R_{k+1} \) or \( R_{k+2}. \) In particular, for \( k = 7, \) \( Q_{\alpha} \in R_9 \) when \( \alpha = \{0.25, 0.5, 0.75\}. \) The approximate solution is of the form

\[
\tilde{u}_{h,i} := \frac{2k' \sin(\pi \alpha)}{\pi} \sum_{\ell = -m}^{M} e^{2(\alpha - 1)\ell k'} (\hat{h} + e^{-2\ell k'})^{-1} f_h.
\]

The parameter \( k' > 0 \) controls the accuracy of \( u_{h,i} \) and the number of linear systems to be solved. For example, \( k' = 1/3 \) gives rise to 120 systems for \( \alpha = \{0.25, 0.75\} \) and 91 systems for \( \alpha = 0.5 \) guaranteeing \( \|\tilde{u}_{h,i} - u_h\|_{\ell^2} \approx 10^{-7}\|f_h\|_{\ell^2}. \) We will refer to such a parameter choice as the \( k' \)-Q-method. On the other hand, taking \( k = 7 \) we need to solve nine linear systems in order to derive \( u_{h,i} \), and this will be called the Q-method. Although the theoretical foundation of the Q-method is quite different from the one of the BURA-related methods, both of the approaches are computationally very similar and this allows us to perform a meaningful comparison analysis.

### 3.3. Numerical tests for uniform mesh.

Now we analyze the computational results from the efficiency point of view. We fix the number \( k \) in such a way that the three methods, BURA, P-BURA and Q-method, require 9 systems of the type \((\bar{I} - d_i\hat{A})\tilde{w}_h = \tilde{v}_h \) to be solved. This means that all three methods need almost the same amount of computational work. For comparison, we also give the results of the \( k' \)-Q-method that has the best accuracy, but requires about 10 - 15 times more computational work.

Tables 2 and 3 present the computational results for the four solvers discussed above for three values of \( \alpha = 0.25, 0.5, 0.75 \) requiring a number of solves as discussed above. Together with the \( \ell^2 \)-norm of the error we also provide (just for comparison purposes) the error measured on the maximum norm \( \| \cdot \|_{\ell^\infty}. \)

The first general comment is that all methods work according to the developed theory. In terms of efficiency the P-BURA method seems to be the best. In agreement with the theory, the accuracy of the method does not depend on the mesh size \( h. \) Also, in agreement with the approximation error reported in Table 1 its error decreases when \( \alpha \) increases. But even in the worst approximation, the case when \( \alpha = 0.25, \) P-BURA produces a reasonable error in the range of \( 10^{-4} \) when using only 9 system solves. Moreover, for a fixed mesh of medium mesh-size (say, \( h = 10^{-8} - 10^{-9} \) and \( \alpha = 0.25 \) P-BURA is as accurate as BURA method and outperforming BURA and Q-method on all meshes for \( \alpha = 0.5 \) and \( \alpha = 0.75 \) on both Problem 1 and Problem 2. In contrast, the \( k' \)-Q-method has the same accuracy, but needs 120 system solves.

Second, we note that from the first row of Table 1 we see that the BURA approximation for \( t^{1-\alpha} = t^{0.75} \) has good accuracy for relatively small values of \( k. \) We see that for \( k = 8 \) the error ranges from \( 4.4 \times 10^{-4} \) for \( \alpha = 0.75 \) to \( 1.2 \times 10^{-6} \) for \( \alpha = 0.25. \) However, the computational results on Tables 2 and 3 show that the factor \( \kappa(\hat{A})^{1-\alpha} \) in the error bound for BURA method is polluting the approximate solution and reducing the accuracy. This pollution is especially visible in the computational results for \( \alpha = 0.25. \) In this case \( \kappa(\hat{A})^{1-\alpha} = \kappa(\hat{A})^{3/4} = O(h^{-3/2}) \) and every time one halves the mesh-size the error is increased by a factor of \( 2^{3/2} \approx 3.8. \) This pollution is less visible for \( \alpha = 0.75 \) since the factor is \( 2^{1/2} \approx 1.4. \) Regardless of this pollution, the BURA method with 8 system solves is, in general, more accurate than the Q-method for all three values of \( \alpha, \) when using the same number of system solves.

### 3.4. Numerical tests on locally refined mesh.

Since the P-BURA error estimate (22) is independent of the condition number of the discretization matrix \( \hat{A}, \) one can also apply local refinement techniques for efficiently capturing the solution behavior around possible singularities of the solution \( u. \) In this section we illustrate the advantages of such an approach, considering one-dimensional
Table 2. Relative errors of the approximate solution for $\alpha = 0.25, 0.5, 0.75$ obtained by four different methods on various uniform meshes. Each of the methods BURA, P-BURA, and Q-method uses 9 linear systems, while the $k'-Q$-method, $k' = 1/3$, uses 120 linear system solves for $\alpha = 0.25, 0.75$ and 91 solves for $\alpha = 0.5$. The reference solution is the discrete solution on a mesh with step-size $h = 2^{-15}$ computed via FFT.

| $\alpha$ | $h$ | Example 1, CheckerBoard right-hand-side |
|---------|-----|----------------------------------------|
|         |     | **BURA** | **P-BURA** | **Q-method** | **$k'$-Q-method** |
|         |     | $\ell_2$ | $\ell_\infty$ | $\ell_2$ | $\ell_\infty$ | $\ell_2$ | $\ell_\infty$ |
| 0.25    | $2^{-8}$ | 2.929e-4 | 2.612e-3 | 3.255e-4 | 2.550e-3 | 1.045e-2 | 1.288e-2 | 2.772e-4 | 2.612e-3 |
|         | $2^{-9}$ | 1.747e-4 | 1.847e-3 | 2.292e-4 | 1.875e-3 | 1.040e-2 | 1.207e-2 | 1.371e-4 | 1.847e-3 |
|         | $2^{-10}$ | 8.217e-4 | 1.829e-3 | 2.029e-4 | 1.339e-3 | 1.039e-2 | 1.152e-2 | 6.815e-5 | 1.305e-3 |
|         | $2^{-11}$ | 5.077e-3 | 1.094e-2 | 1.939e-4 | 8.219e-4 | 1.038e-2 | 1.097e-2 | 3.388e-5 | 9.196e-4 |
|         | $2^{-12}$ | 1.129e-2 | 2.610e-2 | 1.922e-4 | 7.451e-4 | 1.038e-2 | 1.152e-2 | 1.671e-5 | 6.413e-4 |
| 0.50    | $2^{-8}$ | 9.688e-5 | 1.900e-4 | 2.212e-5 | 1.849e-4 | 2.847e-3 | 2.910e-3 | 2.331e-5 | 1.821e-4 |
|         | $2^{-9}$ | 2.337e-4 | 4.995e-4 | 1.013e-5 | 8.787e-5 | 2.835e-3 | 2.904e-3 | 8.058e-6 | 9.110e-5 |
|         | $2^{-10}$ | 3.828e-4 | 8.616e-4 | 8.304e-6 | 4.742e-5 | 2.830e-3 | 2.902e-3 | 2.840e-6 | 4.559e-5 |
|         | $2^{-11}$ | 2.172e-4 | 6.274e-4 | 8.263e-6 | 2.433e-5 | 2.829e-3 | 2.902e-3 | 2.840e-6 | 4.559e-5 |
|         | $2^{-12}$ | 1.424e-3 | 2.814e-3 | 8.291e-6 | 1.909e-5 | 2.828e-3 | 2.902e-3 | 1.138e-7 | 2.280e-5 |
| 0.75    | $2^{-8}$ | 1.219e-4 | 2.741e-4 | 2.443e-6 | 9.168e-6 | 1.507e-3 | 1.825e-3 | 2.561e-6 | 9.103e-6 |
|         | $2^{-9}$ | 1.761e-4 | 3.976e-4 | 1.761e-5 | 8.219e-5 | 1.502e-3 | 1.824e-3 | 7.118e-7 | 3.263e-6 |
|         | $2^{-10}$ | 2.172e-4 | 4.958e-4 | 1.884e-7 | 3.110e-5 | 1.501e-3 | 1.823e-3 | 2.355e-7 | 1.198e-6 |
|         | $2^{-11}$ | 1.401e-4 | 3.478e-4 | 1.500e-7 | 6.592e-7 | 1.500e-3 | 1.823e-3 | 1.138e-7 | 4.677e-7 |
|         | $2^{-12}$ | 1.803e-4 | 3.264e-4 | 1.547e-7 | 4.574e-7 | 1.499e-3 | 1.823e-3 | 8.334e-8 | 2.079e-7 |

Note that with respect to the standard $L^2[0,1]$ inner product, we have $(\psi_i, \psi_j) = \delta_{ij}$ for all $i,j \in \mathbb{N}$. Therefore, for any right-hand side function $f$ on $(0,1)$ we can explicitly compute the solution of the continuous fractional diffusion problem with homogeneous boundary conditions

$$u(x) := A^{-\alpha} f = \sum_{i=1}^{\infty} \mu_i^{-\alpha} (f, \psi_i) \psi_i(x).$$

Furthermore, we have $\lambda_1 = \pi^2 > 1$ so for all meaningful grids on $[0,1]$ (e.g., coming from finite element or finite difference discretization) the first eigenvalue $\lambda_1$ of the corresponding discrete operator $\tilde{A}$ satisfies $\lambda_1 \geq \mu_1 > 1$. Thus the spectrum of $\tilde{A}$ is always in $[1,\infty)$ and we do not need to normalize the matrix in order to apply the P-BURA solver.

Now we take a smooth function, namely, $f(x) = 1$, but the solution of (3) will exhibit boundary layers near the end-points $x = 0$ and $x = 1$. Those layers are steeper as $\alpha \rightarrow 0$ and in order for the numerical solver to correctly capture them, we need very fine mesh near the boundary, especially for small $\alpha$. Thus, we consider the following class of locally refined meshes: Firstly, we start with a uniform mesh of size $h_0$. Then, at each refinement step we take the first and the last segments
Table 3. Relative errors of the approximate solution for $\alpha = 0.25, 0.5, 0.75$ obtained by four different methods on various uniform meshes. Each of the first three solvers incorporates 9 linear systems, while the last solver incorporates 120 for $\alpha = 0.25, 0.75$ and 91 system solves for $\alpha = 0.5$. On each level we have computed the exact Galerkin solution $\tilde{u}_h = \tilde{K}^{-\alpha}\mathcal{I}_h f_2 = \lambda_2^{-\alpha}\mathcal{I}_h \Psi_2$.

\[
\begin{array}{|c|c|c|c|c|c|c|c|c|c|}
\hline
\alpha & h & & & & & & & \\
\hline
& & BURA & & & & & & \\
& & $\ell_2$ & $\ell_\infty$ & $\ell_2$ & $\ell_\infty$ & $\ell_2$ & $\ell_\infty$ & $\ell_2$ & $\ell_\infty$ \\
\hline
0.25 & $2^{-8}$ & 4.615e-5 & 9.193e-5 & 1.086e-4 & 2.162e-4 & 5.223e-3 & 1.040e-2 & 1.955e-6 & 3.894e-6 \\
& $2^{-9}$ & 6.035e-5 & 1.205e-4 & 1.067e-4 & 2.131e-4 & 5.214e-3 & 1.041e-2 & 3.688e-7 & 7.362e-7 \\
& $2^{-10}$ & 4.993e-4 & 9.976e-4 & 1.062e-4 & 2.123e-4 & 5.209e-3 & 1.041e-2 & 2.663e-8 & 5.321e-8 \\
& $2^{-11}$ & 3.122e-3 & 6.240e-3 & 1.061e-4 & 2.121e-4 & 5.207e-3 & 1.041e-2 & 1.253e-7 & 2.506e-7 \\
& $2^{-12}$ & 6.904e-3 & 1.380e-2 & 1.060e-4 & 2.120e-4 & 5.206e-3 & 1.041e-2 & 1.500e-7 & 2.999e-7 \\
& $2^{-8}$ & 6.388e-5 & 1.273e-4 & 5.703e-6 & 1.136e-5 & 1.428e-3 & 2.845e-4 & 1.329e-6 & 2.648e-6 \\
& $2^{-9}$ & 1.426e-4 & 2.846e-4 & 4.630e-6 & 9.243e-6 & 1.426e-3 & 2.847e-3 & 2.650e-7 & 5.290e-7 \\
& $2^{-10}$ & 2.360e-4 & 4.715e-4 & 4.361e-6 & 8.713e-6 & 1.425e-3 & 2.847e-3 & 3.273e-10 & 6.539e-10 \\
& $2^{-11}$ & 1.469e-4 & 2.936e-4 & 4.292e-6 & 8.580e-6 & 1.424e-3 & 2.847e-3 & 6.656e-8 & 1.331e-7 \\
& $2^{-12}$ & 8.771e-4 & 1.754e-3 & 4.275e-6 & 8.547e-6 & 1.424e-3 & 2.848e-3 & 8.310e-8 & 1.662e-7 \\
0.50 & $2^{-8}$ & 7.299e-5 & 1.454e-4 & 7.564e-7 & 1.507e-6 & 8.331e-4 & 1.660e-3 & 6.733e-7 & 1.341e-6 \\
& $2^{-9}$ & 1.086e-4 & 2.168e-4 & 2.208e-7 & 4.408e-7 & 8.320e-4 & 1.661e-3 & 1.379e-7 & 2.753e-7 \\
& $2^{-10}$ & 1.332e-4 & 2.662e-4 & 8.724e-8 & 1.743e-7 & 8.314e-4 & 1.661e-3 & 4.375e-9 & 8.742e-9 \\
& $2^{-11}$ & 8.546e-5 & 1.708e-4 & 5.387e-8 & 1.077e-7 & 8.310e-4 & 1.661e-3 & 2.896e-8 & 5.789e-8 \\
& $2^{-12}$ & 1.110e-4 & 2.219e-4 & 4.553e-8 & 9.103e-8 & 8.308e-4 & 1.661e-3 & 3.728e-8 & 7.454e-8 \\
0.75 & $2^{-8}$ & 7.299e-5 & 1.454e-4 & 7.564e-7 & 1.507e-6 & 8.331e-4 & 1.660e-3 & 6.733e-7 & 1.341e-6 \\
& $2^{-9}$ & 1.086e-4 & 2.168e-4 & 2.208e-7 & 4.408e-7 & 8.320e-4 & 1.661e-3 & 1.379e-7 & 2.753e-7 \\
& $2^{-10}$ & 1.332e-4 & 2.662e-4 & 8.724e-8 & 1.743e-7 & 8.314e-4 & 1.661e-3 & 4.375e-9 & 8.742e-9 \\
& $2^{-11}$ & 8.546e-5 & 1.708e-4 & 5.387e-8 & 1.077e-7 & 8.310e-4 & 1.661e-3 & 2.896e-8 & 5.789e-8 \\
& $2^{-12}$ & 1.110e-4 & 2.219e-4 & 4.553e-8 & 9.103e-8 & 8.308e-4 & 1.661e-3 & 3.728e-8 & 7.454e-8 \\
\hline
\end{array}
\]

The numerical results are summarized on Table 4. We have considered only dyadic mesh refinement (i.e., $p = 2$). As expected, the smaller the $\alpha$ is, the steeper the boundary layers are, thus the bigger the benefit of the local refinement is. For example, for $\alpha = 0.25$ we observe that

\[
(f, \psi_i) = \int_0^1 \sqrt{2} \sin(\pi x) dx = \begin{cases} 0, & \text{if } i \text{ is even;} \\ \frac{2 \sqrt{2}}{\pi i}, & \text{if } i \text{ is odd.} \end{cases}
\]

and due to (45) we have the explicit representation of the exact solution

\[
u(x) = \frac{2 \sqrt{2}}{\pi i + 2\alpha} \sum_{i=0}^{\infty} \frac{\psi_{2i+1}(x)}{(2i + 1)^{1+2\alpha}}.
\]

As a reference solution, we consider the truncated series representation (46) by taking the first $10^4$ terms. Of course, this is an approximation to the exact solution. However, the error of such approximations for $\alpha = 0.25, 0.5, 0.75$ are all less than $1.270e-5, 4.136e-8, 1.429e-10$, respectively. Since these are all substantially below the approximation error of BURA (see, Table 1), we use them as substitute of the exact solution.

The accuracy of this reference solution is higher than the accuracy of the P-BURA method. We study the relative $L_2$-error $\|w_h - u_h\|/\|u\|$, where $w_h$ is a piece-wise linear function that interpolates the P-BURA solution on the locally refined mesh, while $u_h$ is a piece-wise linear function that samples $u$ on the uniform mesh with $h = 2^{-18}$.

The numerical results are summarized on Table 4. We have considered only dyadic mesh refinement (i.e., $p = 2$). As expected, the smaller the $\alpha$ is, the steeper the boundary layers are, thus the bigger the benefit of the local refinement is. For example, for $\alpha = 0.25$ we observe that
starting with a uniform mesh of size \( h_0 = 2^{-6} \) and performing 9 additional local refinement steps, we end up with a numerical solution that is as accurate as the numerical solution on a uniform mesh of size \( h_0 = 2^{-10} \). On the other hand, the first mesh consists of 81 nodes, while the second one - of 1023. Moreover, as \( h_0 \) increases the order of the error is the same as the order of the 9-BURA accuracy \( E_{0.25,9} = 2.654e-4 \) (see Table 1), meaning that the geometrical 6-step adaptive refinement with \( h_0 = 2^{-9} \) leads to almost optimal results at the numerical cost of solving nine tridiagonal linear systems of size 523. The latter is further illustrated on Fig. 1. There, using that the CheckerBoard function on \([0,1] \times [0,1]\) can be split into four squared sub-domains, such that on each of them we solve a tensor product of two 1D problems like Example 3, together with the linearity of the fractional Laplace operator, we numerically compute solution of Example 1 for \( \alpha = 0.25 \) on a \((2 \cdot 523 + 1) \times (2 \cdot 523 + 1) = 1047 \times 1047\) mesh, which is locally adapted along the boundary of the domain and the lines of discontinuity of the right-hand-side. On the left, we plot the computed numerical solution. In the middle we show the mesh refined around the central point (0.5,0.5). On the right we plot the point-wise error between the numerical solution and the true exact solution, sampled at a uniform grid of size \( h = 2^{-16} \) in the same region of interest. Note that the region captures the boundary layers along \( x = 0.5 \) and \( y = 0.5 \) and the point-wise error is less than 2e-4 overall.

| Table 4. Computing \( \|w_h - u_h\|/\|u\| \) for \( f(x) = 1 \) and \( \alpha = \{0.25, 0.5, 0.75\} \) on various uniform and locally refined meshes. 9-P-BURA is used as solver and all meshes have been iteratively refined until the smallest mesh segment is of size \( 2^{-15} \). Dyadic refinement \( (p = 2) \) has been applied. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Ref. level | \( h_0 = 2^{-6} \) | \( h_0 = 2^{-7} \) | \( h_0 = 2^{-8} \) | \( h_0 = 2^{-9} \) | \( h_0 = 2^{-10} \) |
| \( \alpha = 0.25 \) | 0 | 1.813e-2 | 9.071e-3 | 4.544e-3 | 2.287e-3 | 1.179e-3 |
| | last | 1.294e-3 | 6.931e-4 | 4.446e-4 | 3.351e-4 | 3.301e-4 |
| \( \alpha = 0.50 \) | 0 | 2.705e-3 | 9.547e-4 | 3.380e-4 | 1.233e-4 | 5.498e-5 |
| | last | 7.620e-4 | 2.648e-4 | 9.487e-5 | 4.563e-5 | 3.734e-5 |
| \( \alpha = 0.75 \) | 0 | 6.415e-4 | 1.713e-4 | 4.882e-5 | 2.158e-5 | 1.795e-5 |
| | last | 4.712e-4 | 1.321e-4 | 4.054e-5 | 2.056e-5 | 1.792e-5 |
| \# mesh | | 63 | 127 | 255 | 511 | 1023 |
| nodes | | 81 | 143 | 269 | 523 | 1033 |

**Figure 1.** Computing the CheckerBoard problem for \( \alpha = 0.25 \) on a locally adaptive 2-dimensional mesh. Left: Tensor-product numerical solution. Center: Geometrical refinement along discontinuities: the mesh within the domain \( \Omega' = [0.496, 0.504] \times [0.496, 0.504] \). Right: Plot of the error \( (u_P - u_h)(x) \) over \( \Omega' \) for \( h = 2^{-16} \).
Table 5. Computing $\|w_h - u_h\|/\|u\|$ for $f(x) = \delta_{1/2}(x)$ and $\alpha = \{0.5, 0.75\}$ on various uniform and locally refined meshes. P-BURA is used as solver and meshes have been iteratively refined until the smallest mesh segment is of size $2^{-13}$, for $\alpha = 0.5$, $\alpha = 0.75$. Dyadic refinement ($p = 2$) has been applied.

| $\alpha$ | Ref. level | $h_0 = 2^{-6}$ | $h_0 = 2^{-7}$ | $h_0 = 2^{-8}$ | $h_0 = 2^{-9}$ | $h_0 = 2^{-10}$ |
|----------|------------|----------------|----------------|----------------|----------------|----------------|
| 0.50     | 0          | 9.218e-2       | 6.510e-2       | 4.610e-2       | 3.243e-2       | 2.309e-2       |
|          | last       | 1.026e-2       | 7.975e-3       | 6.559e-3       | 5.690e-3       | 5.267e-3       |
| 0.75     | 0          | 5.776e-3       | 2.882e-3       | 1.452e-3       | 7.138e-4       | 3.610e-4       |
|          | last       | 1.456e-3       | 7.345e-4       | 3.826e-4       | 2.084e-4       | 1.423e-4       |
| # mesh   | 0          | 63             | 127            | 255            | 511            | 1023           |
| nodes    | last       | 83/77          | 145/139        | 269/263        | 525/519        | 1035/1029      |

Example 4. We consider the problem $-u'' = \delta_{1/2}(x)$ for $0 < x < 1$, $u(0) = u(1) = 0$, where the Dirac delta function $\delta_{1/2}(x)$ is centered at 0.5. The weak formulation of this problem is: find $u \in H^1_0(0,1)$ satisfying

$$
\int_0^1 u' \phi' \, dx = \int_0^1 \delta_{1/2} \phi \, dx = \phi(1/2), \quad \text{for all } \phi \in H^1_0(0,1).
$$

It follows that

$$
\delta_{1/2} = \sum_{i=1}^{\infty} \psi_i(1/2) \psi_i(x) = \sqrt{2} \sum_{i=0}^{\infty} (-1)^i \psi_{2i+1}(x).
$$

Hence,

$$
(47) \quad u(x) = \sqrt{2\pi}^{-2\alpha} \sum_{i=0}^{\infty} \frac{(-1)^i}{(2i+1)^{2\alpha}} \psi_{2i+1}(x) \quad \text{and} \quad \|u\| = \sqrt{2\pi}^{-2\alpha} \left( \sum_{i=0}^{\infty} (2i + 1)^{-4\alpha} \right)^{1/2}.
$$

Since for $\alpha \leq 0.25$ the series $\sum_{i=0}^{\infty} (2i + 1)^{-4\alpha}$ does not converge, $u(x) \in L^2(0,1)$ only for $\alpha > 0.25$. Therefore, here we considered the cases $\alpha = 0.5$ and $\alpha = 0.75$ only. Since the singularity is at 0.5, we start with a uniform mesh of size $h_0$, at each refinement step we take the two central segments (those that have a boundary point at $1/2$) and divide them in halves. Once the mesh is fixed, we take $\tilde{h}_0$ to be zero everywhere but in the middle, where the value is set to $h_{*1}$ – the size of the segments, attached to the midpoint. As before, we truncate the infinite series (47) to produce approximation to the solution $u$ which error is far below the error due to BURA. In Table 5 we summarize our study of the relative $L_2$-error: $\|w_h - u_h\|/\|u\|$.

4. Error estimates for the semi-discrete finite element approximations

In this section, we consider finite element approximations to the solution of the fractional problem $A^\alpha u = f$ (or equivalently $u = A^{-\alpha} f$). For simplicity, we only consider the case when the solution operator $T$ satisfies full elliptic regularity, i.e. for $f \in L^2(\Omega)$, the solution $v = Tf$ of (2) is in $H^2(\Omega) \cup H^1_0(\Omega)$ and satisfies

$$
(48) \quad \|v\|_{H^2(\Omega)} \leq c \|f\|.
$$

The assumption of full regularity greatly simplifies the semi-discrete error analysis. Further, to avoid proliferation of various constant related to the maximum and minimum values of $a(x)$ in this Section we shall use the norm generate by the bilinear form $a(\cdot, \cdot)$ which is equivalent to $H^1$:

$$
(49) \quad \|u\|_a = a(u, u)^{1/2} \quad \text{for all } u \in H^1_0(\Omega).
$$
4.1. Consistent mass finite element method. Now we consider the finite elements method (11) with consistent mass computation.

Approximation properties of the method. To study the approximation properties of $u_h$, we follow the technologies developed in [17, 40] and include the proof for completeness.

**Theorem 4.1.** (Fujita-Suzuki, [17, Theorem 5.2, p. 806]) Suppose that (48) holds. Then for $f \in L^2(\Omega),$ 
$$\|A^{-\alpha} f - \mathfrak{A}^{-\alpha} \pi_h f\| \leq C h^{2\alpha}\|f\|$$ 
with $C$ not depending on $h$. Here $\mathfrak{A}$ denotes the finite element operator without lumping, i.e., that appearing in (7).

**Proof.** Using the Balikrishnan formula (4) gives 
$$A^{-\alpha} f - \mathfrak{A}^{-\alpha} \pi_h f = \frac{\sin(\pi \alpha)}{\pi} \int_0^\infty \mu^{-\alpha}[w_\mu - w_{\mu, h}] d\mu. \tag{50}$$
where 
$$w_\mu := (\mu I + A)^{-1} f$$ and 
$$w_{\mu, h} = (\mu I + \mathfrak{A})^{-1} \pi_h f.$$ 
We clearly have that $w_\mu \in H^1_0(\Omega)$ is the unique solution of
$$\mu(w_\mu, \theta) + a(w_\mu, \theta) = (f, \theta) \quad \text{for all } \theta \in H^1_0(\Omega) \tag{51}$$
and $w_{\mu, h} \in \mathcal{V}_h$ is the unique solution of
$$\mu(w_{\mu, h}, \theta) + a(w_{\mu, h}, \theta) = (f, \theta) \quad \text{for all } \theta \in \mathcal{V}_h. \tag{52}$$
Now, taking $\theta = w_\mu$ in (51) and applying the Schwarz inequality to the right hand side implies that 
$$\|w_\mu\| \leq \mu^{-1}\|f\|$$
and subsequently
$$\|w_\mu\|_a \leq \mu^{-1/2}\|f\|. \tag{53}$$
Here $\| \cdot \|_a := a(\cdot, \cdot)^{1/2}$ denotes the $a$-norm on $H^1_0(\Omega)$. Let $e_\mu = w_\mu - w_{\mu, h}$. Using (53), Galerkin orthogonality and standard error estimates for finite element approximation gives, for all $\chi \in \mathcal{V}_h,$ 
$$\mu\|e_\mu\|^2 + a(e_\mu, e_\mu) = \mu(e_\mu, w_\mu - \chi) + a(e_\mu, w_\mu - \chi)$$
$$\leq ch\left[\mu^{1/2}\|e_\mu\| + \|e_\mu\|_a \right] \|f\|.$$ 
A simple application of the arithmetic-geometric mean inequality implies
$$\|e_\mu\| := (\mu\|e_\mu\|^2 + a(e_\mu, e_\mu))^{1/2} \leq ch\|f\|. \tag{54}$$
We apply finite element duality defining $z \in H^1_0(\Omega)$ to be the solution of
$$\mu(\theta, z) + a(\theta, z) = (\theta, e_\mu), \quad \text{for all } \theta \in H^1_0(\Omega),$$
so that Galerkin orthogonality implies
$$\|e_\mu\|^2 = \mu(e_\mu, z) + a(e_\mu, z) = \mu(e_\mu, z - \chi) + a(e_\mu, z - \chi), \quad \text{for all } \chi \in \mathcal{V}_h.$$ 
Now the Schwarz inequality, (54) and arguments leading to (54) applied to $e_\mu^z := z - \chi$ gives
$$\|e_\mu\|^2 \leq \|e_\mu\|\|e_\mu^z\| \leq ch^2\|f\|\|e_\mu\|$$
and so
$$\|e_\mu\| \leq ch^2\|f\|. \tag{55}$$
Now, taking $\theta = w_\mu$ in (51) and $\theta = w_{\mu,h}$ in (52) gives

$$\|w_\mu\| \leq \mu^{-1}\|f\| \text{ and } \|w_{\mu,h}\| \leq \mu^{-1}\|f\|$$

and so

(56) $$\|e_\mu\| \leq 2\mu^{-1}\|f\|.$$

Using the above estimates in (50) gives

$$\|A^{-\alpha}f - \tilde{A}^{-\alpha} \pi_h f\| \leq \frac{\sin(\pi \alpha)}{\pi} \left[ \int_0^{h^{-2}} \mu^{-\alpha}\|e_\mu\| \, d\mu + \int_{h^{-2}}^{\infty} \mu^{-\alpha}\|e_\mu\| \, d\mu \right]$$

$$\leq c\|f\| \left[ \int_0^{h^{-2}} \mu^{-\alpha} h^2 \, d\mu + \int_{h^{-2}}^{\infty} \mu^{-1-\alpha} \, d\mu \right] \leq ch^{2\alpha}\|f\|.$$

This completes the proof of the theorem. \(\square\)

**Positivity of the approximate solution.** Note that diffusion problem (1) is nonnegative if $f$ is nonnegative. This property is retained by the finite difference approximation of the problem. In the case of finite element method (10) we see that if $\tilde{S}$ is an $M$-matrix, then for $f \geq 0$ and consequently $\tilde{F} \geq 0$ and $\tilde{u}_h \geq 0$, i.e. the finite element method preserves positivity.

Next, we ask the question whether the finite element solution (11) of the sub-diffusion problem (3) retains this property. Obviously, the solution (11) can be expressed by (4)

$$\tilde{u}_h = \tilde{A}^{-\alpha} \tilde{M}^{-1} \tilde{F} = \frac{\sin(\pi \alpha)}{\pi} \int_0^{\infty} \mu^{-\alpha} (\mu \tilde{M} + \tilde{S})^{-1} \, d\mu \tilde{F}$$

Since $\tilde{M}$ is not an $M$-matrix, from this representation one can conjecture that even if $\tilde{S}$ is an $M$-matrix, $\tilde{M} \tilde{A}^{-\alpha}$ could fail to be an $M$-matrix and the finite element scheme may be not positivity preserving. For this, we have made some direct computations of the entries of the matrix $\tilde{M} \tilde{A}^{-\alpha}$ for the case of one-dimensional problems for various $\alpha$ and step-sizes $h = 1/(N + 1)$. In this case the $N \times N$ matrix $\tilde{S}$ is defined by (41) (with $h_i = h$) and $\tilde{M} = \frac{h}{N} \text{diag}(1, 4, 1)$ (a tridiagonal matrix).

In Table 6 we present the following information regarding the matrix $\tilde{M} \tilde{A}^{-\alpha}$ for various $\alpha$ and step-size $h$: in columns MrowS, we report the maxim row-sum and in columns MoffD, we report the maximum of all off-diagonal elements. We see that all row sums are positive. It is clear that if the maximal off-diagonal element is positive, then the matrix is NOT an $M$-matrix and therefore we cannot conclude positivity in this case. From this table, we also see that for $\alpha \geq 0.3$ the matrix $\tilde{M} \tilde{A}^{-\alpha}$ has all off-diagonal entries negative, thus it is an $M$-matrix and consequently the scheme will preserve positivity.

**Table 6.** The maximum row-sum (MrowS) and largest off-diagonal entries (MoffD) of matrix $\tilde{M} \tilde{A}^{-\alpha}$ for the one-dimensional problem.

| $\alpha$ | $h = 1/10$ | $h = 1/20$ | $h = 1/40$ | $h = 1/80$ |
|----------|------------|------------|------------|------------|
|          | MrowS  | MoffD | MrowS  | MoffD | MrowS  | MoffD | MrowS  | MoffD |
| 0.100    | 0.11950 | 0.018190 | 0.05958 | 0.010435 | 0.02977 | 0.005992 | 0.01488 | 0.003442 |
| 0.200    | 0.14097 | 0.014288 | 0.07006 | 0.009397 | 0.03498 | 0.006198 | 0.01748 | 0.004089 |
| 0.300    | 0.16354 | -0.000025 | 0.08102 | -0.000025 | 0.04042 | -0.000002 | 0.02501 | -0.000002 |
| 0.500    | 0.20417 | -0.0000852 | 0.10052 | -0.000049 | 0.05006 | -0.000003 | 0.02501 | -0.000002 |
| 0.700    | 0.21119 | -0.000177 | 0.10342 | -0.000068 | 0.05142 | -0.000004 | 0.02568 | -0.000003 |
| 0.800    | 0.18326 | -0.000138 | 0.08963 | -0.000065 | 0.04452 | -0.000004 | 0.02223 | -0.000002 |
| 0.900    | 0.11824 | -0.0000805 | 0.05786 | -0.000045 | 0.02872 | -0.000003 | 0.01433 | -0.000002 |
On Table 7 we report more computations of this kind using refined values around $\alpha = 0.3$. We see that in the one-dimensional case the matrix becomes an $M$-matrix for $\alpha \geq 0.287$.

Table 7. The maximum row-sum ($\text{MrowS}$) and largest off-diagonal entries ($\text{MoffD}$) of matrix $\tilde{M}A^\alpha$ for one-dimensional problem.

| $\alpha$ | $h = 1/10$ | $h = 1/20$ | $h = 1/40$ | $h = 1/80$ |
|----------|------------|------------|------------|------------|
|          | $\text{MrowS}$ | $\text{MoffD}$ | $\text{MrowS}$ | $\text{MoffD}$ | $\text{MrowS}$ | $\text{MoffD}$ | $\text{MrowS}$ | $\text{MoffD}$ |
| 0.284    | 0.15991    | 0.000733   | 0.07927    | 0.0004962  | 0.03955    | 0.0003949  | 0.01976    | 0.0002703  |
| 0.286    | 0.16036    | 0.000221   | 0.07949    | 0.0001166  | 0.03966    | 0.0000838  | 0.01982    | 0.0000621  |
| 0.288    | 0.16082    | -0.000302  | 0.07971    | -0.0000436 | 0.03977    | -0.0000015| 0.01987    | -0.0000001|
| 0.290    | 0.16127    | -0.000406  | 0.07993    | -0.0000239 | 0.03988    | -0.0000015| 0.01993    | -0.0000001|
| 0.292    | 0.16172    | -0.000413  | 0.08037    | -0.0000243 | 0.03999    | -0.0000015| 0.01998    | -0.0000001|

We also performed similar computations for the Poisson equation in an $L$-shaped domain, namely $\Omega = \{(0,1) \times (0,1)\} \setminus \{(0.5,1) \times (0.5,1)\}$. In this case we introduce an uniform mesh with step-size in both directions $h = 1/(n+1)$ so that the stiffness matrix is an $M$-matrix of size $N = 0.75n^2$. The results are reported in Table 8. We note that the matrix has many negative elements in a row. However, the existence of a positive off-diagonal entry in all cases suggests that $\tilde{M}A^\alpha$ is NOT an $M$-matrix when $\tilde{M}$ is consistent mass matrix and therefore, the method fails to preserve the positivity for all $\alpha \in (0,1)$. Similar are the results of a rectangular domain on a uniform square mesh. These results are a bit different from the one-dimensional computations shown on Table 6 and 7, where we see that for $\alpha \geq 0.3$ we have an $M$-matrix and the method will preserve positivity. We expect that in 3-D problems the consistent mass methods will not be positivity preserving.

Table 8. The maximum row sum ($\text{MrowS}$) and the largest off-diagonal entries ($\text{MoffD}$) of matrix $\tilde{M}A^\alpha$ for L-shaped domain.

| $\alpha$ | $h = 1/10$ | $h = 1/20$ | $h = 1/40$ |
|----------|------------|------------|------------|
|          | $\text{MrowS}$ | $\text{MoffD}$ | $\text{MrowS}$ | $\text{MoffD}$ | $\text{MrowS}$ | $\text{MoffD}$ |
| 0.300    | 0.02403    | 0.002695   | 0.00578    | 0.0001022  | 0.00143    | 0.000387  |
| 0.500    | 0.03820    | 0.004974   | 0.00905    | 0.002483   | 0.00222    | 0.001242  |
| 0.700    | 0.04891    | 0.007004   | 0.01170    | 0.004558   | 0.00285    | 0.003007  |
| 0.800    | 0.04628    | 0.006979   | 0.01134    | 0.005223   | 0.00275    | 0.003958  |
| 0.900    | 0.03176    | 0.004897   | 0.00818    | 0.004222   | 0.00198    | 0.003675  |

4.2. **Lumped mass finite element method.** As discuses in Subsection 2.4, when we employ mass lumping, both the semi-discrete and fully discrete approximations satisfy the positivity property, i.e., if $f$ is continuous and $f \geq 0$, then $u_h$ and $w_h$ given by (24) and (25) are both non-negative. This is a consequence of the fact that the lumped mass matrix is diagonal with positive diagonal entries, (4) and Proposition 2.1. Besides, since many finite difference schemes could be considered as obtained by lumped mass FEM, we have as a by-product of the result below, an error estimate for the finite difference approximations of spectral sub-diffusion problems. We are not aware of rigorous proof of such result.

We conclude this section with an error estimate in the case of two-dimensional problems with full regularity and data $f \in H^{1+\gamma}(\Omega)$:
Theorem 4.2. Let $\Omega \subset \mathbb{R}^2$ and suppose that (48) holds. Then for $f \in H^{1+\gamma}(\Omega)$ with $\gamma > 0$, $u_h$ given by (24) satisfies
\begin{equation}
\|A^{-\alpha}f - u_h\| \leq C(h^{2\alpha} + h^{1+\gamma})\|f\|_{H^{1+\gamma}(\Omega)}
\end{equation}
with $C$ not depending on $h$.

Proof. Let $\tilde{V}_h$ denote the set of continuous piecewise linear functions with respect to the mesh on $\Omega$ including non-vanishing functions on $\partial \Omega$. For the purposes of this proof, we consider $I_h$ as a map from $C^0(\tilde{\Omega})$ into $\tilde{V}_h$ even though the boundary values do not enter into (24) (or (25)). The resulting mass lumped matrix satisfies the following estimate:
\begin{equation}
\|I_h f\| + \|I_h f\|_a + h^{-1-\gamma}\|(f - I_h f)\| \leq C\|f\|_{H^{1+\gamma}(\Omega)}.
\end{equation}

By (59) and the stability of $A^{-\alpha}$,
\begin{equation}
\|A^{-\alpha} (f - I_h f)\| \leq C\|f - I_h f\| \leq Ch^{1+\gamma}\|f\|.
\end{equation}
Thus, we are left to bound
\begin{equation}
\|A^{-\alpha}I_h f - u_h\| \leq \|A^{-\alpha}I_h f - A^{-\alpha}I_h f\| + \|A^{-\alpha}I_h f - u_h\|.
\end{equation}
Here $A$ is the finite element operator appearing in (7) and Theorem 4.1. By Theorem 4.1 and (59),
\begin{equation}
\|A^{-\alpha}I_h f - A^{-\alpha}I_h f\| \leq Ch^{2\alpha}\|I_h f\| \leq Ch^{2\alpha}\|f\|_{H^{1+\gamma}(\Omega)}.
\end{equation}

For the second term in (61), we use the Balakrishnan formula and write
\begin{equation}
\|A^{-\alpha}I_h f - u_h\| \leq \sin(\pi\alpha) \left\| \int_0^\infty \mu^{-\alpha}(w_{\mu,h} - \tilde{w}_{\mu,h}) d\mu \right\|
\end{equation}
\begin{equation}
\leq \sin(\pi\alpha) \int_0^\infty \mu^{-\alpha}\|w_{\mu,h} - \tilde{w}_{\mu,h}\| d\mu
\end{equation}
where $w_{\mu,h}$ satisfies (52) with $f$ replaced by $I_h f$ and $\tilde{w}_{\mu,h} \in V_h$ satisfies
\begin{equation}
\mu(\tilde{w}_{\mu,h}, \theta)_h + a(\tilde{w}_{\mu,h}, \theta) = (I_h f, \theta)_h \quad \text{for all } \theta \in V_h.
\end{equation}

It follows that for $e_{\mu} = w_{\mu,h} - \tilde{w}_{\mu,h}$ and $\phi \in V_h$,
\begin{equation}
\mu(e_{\mu}, \phi) + a(e_{\mu}, \phi) = \mu[(\tilde{w}_{\mu,h}, \phi)_h - (\tilde{w}_{\mu,h}, \phi)] + [(I_h f, \phi) - (I_h f, \phi)_h].
\end{equation}
Taking $\phi = e_{\mu}$ and applying (58) gives
\begin{equation}
\mu\|e_{\mu}\|^2 + \|e_{\mu}\|^2_a \leq ch^2[\mu\|w_{\mu,h}\|_a\|e_{\mu}\|_a + \|I_h f\|_a\|e_{\mu}\|_a].
\end{equation}
The same argument that showed (53) leads to
\begin{equation}
\|\tilde{w}_{\mu,h}\|_a \leq \mu^{-1/2}\|I_h f\|_a.
\end{equation}
Thus, (64) implies
\begin{equation}
\mu\|e_{\mu}\|^2 + \|e_{\mu}\|^2_a \leq ch^2[\mu^{1/2} + 1]\|f\|_{H^{1+\gamma}(\Omega)}\|e_{\mu}\|_a.
\end{equation}
A straightforward application of the arithmetic-geometric mean inequality then gives
\begin{equation}
\mu\|e_{\mu}\|^2 + \|e_{\mu}\|^2_a \leq ch^4[\mu + 1]\|f\|^2_{H^{1+\gamma}(\Omega)}.
\end{equation}
We finally bound the integral in \((62)\) by breaking up the integration interval and bounding the resulting subinterval integrals. By \((65)\) and the Poincaré inequality,
\[
\int_0^1 \mu^{-\alpha} \|e_\mu\| \, d\mu \leq c \int_0^1 \mu^{-\alpha} \|e_\mu\|_\alpha \, d\mu \leq ch^2 \|f\|_{H^{1+\gamma}(\Omega)}^2
\]
and
\[
\int_1^{h^{-2}} \mu^{-\alpha} \|e_\mu\| \, d\mu \leq Ch^2 \|f\|_{H^{1+\gamma}(\Omega)} \int_1^{h^{-2}} \mu^{-\alpha} \, d\mu \leq Ch^{2\alpha} \|f\|_{H^{1+\gamma}(\Omega)}.
\]
As in \((56)\),
\[
\|e_\mu\| \leq C \mu^{-1} \|\mathcal{I}_h f\| \leq C \mu^{-1} \|f\|_{H^{1+\gamma}(\Omega)}
\]
so that
\[
\int_{h^{-2}}^{\infty} \mu^{-\alpha} \|e_\mu\| \, d\mu \leq C \|f\|_{H^{1+\gamma}(\Omega)} \int_{h^{-2}}^{\infty} \mu^{-1-\alpha} \, d\mu = Ch^{2\alpha} \|f\|_{H^{1+\gamma}(\Omega)}.
\]
Combining the above estimates completes the proof of the theorem. 

\[\square\]

**Remark 4.1.** The above theorem and its proof remains valid when \(\Omega \subset \mathbb{R}^d\) for \(d = 1\) and \(d = 3\) provided that \(1 + \gamma\) is replaced by \(d/2 + \gamma\). We believe that more refined analysis could lower the required smoothness of \(f\), but this will involve solution of number of technical issues that are beyond the scope of this paper.

**Corollary 4.3.** Let \(\Omega \subset \mathbb{R}^2\) and suppose that \((48)\) holds. Using BURA approximation error estimate \((21)\), the error bound of the lumped mass finite element method \((57)\), and the inequality \(\|f_h\| = \|\pi_h f\| \leq \|f\|\) we get
\[
\|A^{-\alpha} f - w_h\| \leq C(h^{2\alpha + h^{1+\gamma}}) \|f\|_{H^{1+\gamma}(\Omega)} + \lambda_1^{-\alpha} \ell_{\alpha,k} \|f\| \leq C(h^{2\alpha + h^{1+\gamma} + e^{-2\pi\sqrt{k\alpha}}} \|f\|_{H^{1+\gamma}(\Omega)}
\]
with \(C\) not depending on \(h\) and \(k\). Then the contributions of the finite element discretization and the BURA approximation to the total error can be balanced choosing properly the parameters \(h\) and \(k\). We see that the BURA error is fully controlled by \(\ell_{\alpha,k}\) and the \(L^2\)-norm of the data \(f\). This allows to choose \(k\) by using Table 1 once we fix the desired accuracy of the computations. To choose the mesh that guarantees the same accuracy we need to use either Richardson extrapolation or any other technique for error control of the finite element method by mesh refinement. Such discussion is beyond the scope of the paper.

**Corollary 4.4.** As a by-product of the error analysis of the lumped mass finite element method we also obtain an error bound for the finite difference method for the two-dimensional case through it equivalence to the lumped mass approximation on uniform meshes. To the best of our knowledge, this fact has not been know before.

5. Concluding remarks

In this paper we study algorithms of optimal complexity for solving the system of algebraic equations \(\tilde{A}^\alpha \tilde{u}_h = \tilde{f}_h\), \(0 < \alpha < 1\) for \(\tilde{u}_h \in \mathbb{R}^N\), where \(\tilde{A}\) is a symmetric and positive definite \(N \times N\) matrix with spectrum in \([\lambda_1, \lambda_N]\) which is obtained from discretization of a second order elliptic problem by a finite difference or finite elements method. Two methods, BURA and P-BURA, are analyzed and experimentally studied. They are based on the best uniform rational approximation \(r_{\gamma,k}(t) = P_k(t)/Q_k(t)\) of \(t^\gamma\). We note that these could be precomputed and later used in the computations. Such results for various values of \(\alpha\) and \(k\) could be found in the report \([18]\).

The presented estimates show that both methods have exponential convergent rate with respect to \(k\). They reduce the nonlocal fractional diffusion problem to solution of small number (determined by \(k\)) of systems in the form \((\tilde{A} + c\tilde{I})u_h = f_h\), \(c \geq 0\). The algorithm is optimal with respect to
$N$, assuming that solvers of optimal complexity (e.g. multigrid or multilevel) are used for the related sparse symmetric and positive definite (discrete elliptic) problems. More precisely, the computational complexity is $O(kN)$.

The presented numerical tests support the theoretical estimates. They prove the concept of the new P-BURA method and show its high efficiency. In contrast to BURA, the accuracy of P-BURA method does not depend on the condition number of $A$. This makes P-BURA robust with respect to the mesh parameter $h$, which also holds true in the case of approximations on locally refined meshes.

In general, the regularity of solution of the considered fractional diffusion problems decreases with decreasing of $\alpha$. As shown theoretically and numerically in [8], the convergence rate of $\|u - u_h\|_{L^2}$ is at best $O(h^{2\alpha})$. Here we studied also the lumped mass method that preserves the positivity property. In this context, the used here local mesh refinement for the CheckerBoard right hand side $f$ shows new promising opportunities for a substantial increase of the accuracy based on the robustness of P-BURA method. Even more impressive are the obtained results for Example 4 the case of Dirac delta-function right-hand-side.

Within the context of this paper, the question about the proper norms and algorithms for adaptive mesh refinement is very important, but not studied. This holds as well for the case when $f$ has lower than $L^2$-regularity. We feel that a study of these issues needs a separate rigorous technical analysis which remains out of the scope of this paper.

ACKNOWLEDGEMENT

This research has been partially supported by the Bulgarian National Science Fund under grant No. BNSF-DN12/1. The work of R. Lazarov was supported in parts by NSF-DMS #1620318 grant.

We acknowledge also the provided access to the e-infrastructure of the Centre for Advanced Computing and Data Processing, with the financial support by the Grant No. BG05M2OP001-1.001-0003, financed by the Science and Education for Smart Growth Operational Program (2014-2020) and co-financed by the European Union through the European structural and Investment funds.

REFERENCES

[1] L. Aceto and P. Novati. Rational approximation to the fractional Laplacian operator in reaction-diffusion problems. *SIAM J. Sci. Comput.*, 39(1):A214 – A228, 2017.
[2] L. Aceto and P. Novati. Efficient implementation of rational approximations to fractional differential operators. *Journal of Scientific Computing*, 76(1):651–671, 2018.
[3] G. Acosta and J. P. Borthagaray. A fractional Laplace equation: regularity of solutions and finite element approximations. *SIAM Journal on Numerical Analysis*, 55(2):472–495, 2017.
[4] A. Balakrishnan. Fractional powers of closed operators and the semigroups generated by them. *Pacific Journal of Mathematics*, 10(2):419–437, 1960.
[5] A. Bonito, W. Lei, and J. E. Pasciak. On sinc quadrature approximations of fractional powers of regularly accretive operators. *Journal of Numerical Mathematics*, 2018.
[6] A. Bonito, W. Lei, and J. E. Pasciak. Numerical approximation of the integral fractional laplacian. *Numerische Mathematik*, 142(2):235–278, 2019.
[7] A. Bonito, W. Lei, and J. E. Pasciak. On sinc quadrature approximations of fractional powers of regularly accretive operators. *Journal of Numerical Mathematics*, 27(2):57–68, 2019.
[8] A. Bonito and J. Pasciak. Numerical approximation of fractional powers of elliptic operators. *Mathematics of Computation*, 84(295):2083–2110, 2015.
[9] A. Bonito and J. Pasciak. Numerical approximation of fractional powers of regularly accretive operators. *IMA J Numer Anal*, 37(3):1245–1273, 2017.
[10] L. Caffarelli and L. Silvestre. An extension problem related to the fractional laplacian. *Communications in Partial Differential Equations*, 32(8):1245–1260, 2007.
[11] P. Ciarlet. *The Finite Element Method for Elliptic Problems*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 2002.

[12] P. Ciarlet and P.-A. Raviart. Maximum principle and uniform convergence for the finite element method. *Computer Methods in Appl. Mechanics and Engineering*, 2:17–31, 1973.

[13] R. Ciegis and P. N. Vabishchevich. Two-level schemes of cauchy problem method for solving fractional powers of elliptic operators. *Computers and Mathematics with Applications*, 2019.

[14] T. A. Driscoll, N. Hale, and L. N. Trefethen. *Chebfun Guide*. Pafnuty Publications, 2014.

[15] V. Druskin and L. Knizhnerman. Extended Krylov subspaces: approximation of the matrix square root and related functions. *SIAM Journal on Matrix Analysis and Applications*, 19(3):755–771, 1998.

[16] B. Duan, R. D. Lazarov, and J. E. Pasciak. Numerical approximation of fractional powers of elliptic operators. *IMA Journal of Numerical Analysis*, 03 2019.

[17] H. Fujita and T. Suzuki. Evolution problems. In J. L. P.G. Ciarlet, editor, *Handbook of Numerical Analysis*, volume II, pages 789–923. Elsevier, 1991.

[18] S. Harizanov, R. Lazarov, S. Margenov, and P. Marinov. The best uniform rational approximation (BURA) of $t^\alpha$, $t \in [0, 1]$, $\alpha \in (0, 1)$: applications to solving equations involving fractional powers of elliptic operators. *Report*, pages 1–66, 2019.

[19] S. Harizanov, R. Lazarov, S. Margenov, P. Marinov, and Y. Vutov. Optimal solvers for linear systems with fractional powers of sparse SPD matrices. *Numerical Linear Algebra with Applications*, 25(4):115–128, 2018.

[20] S. Harizanov, R. Lazarov, P. Marinov, S. Margenov, and J. Pasciak. Analysis of numerical methods for spectral fractional elliptic equations based on the best uniform rational approximation. *arXiv preprint arXiv:1905.08155*, 2019.

[21] S. Harizanov and S. Margenov. Positive approximations of the inverse of fractional powers of spd m-matrices. *Control Systems and Mathematical Methods in Economics: Essays in Honor of Vladimir M. Veliov*, 687:147–163, 2018.

[22] S. Harizanov, S. Margenov, P. Marinov, Y. Vutov. Volume constrained 2-phase segmentation method utilizing a linear system solver based on the best uniform polynomial approximation of $x^{-1/2}$. *J. Computational and Applied Mathematics*, 310:115–128, 2017.

[23] N. J. Higham. Stable iterations for the matrix square root. *Numerical Algorithms*, 15(2):227–242, 1997.

[24] C. Kenney and A. J. Laub. Rational iterative methods for the matrix sign function. *SIAM J. Matrix Anal. Appl.*, 12(2):273–291, 1991.

[25] A. Lisichke, G. Pang, M. Gulian, F. Song, C. Glusa, X. Zheng, Z. Mao, W. Cai, M. M. Meerschaert, M. Ainsworth, and G. Karniadakis. What is the fractional laplacian? *arXiv preprint arXiv:1801.09767*, 2018.

[26] P. G. Marinov and A. S. Andreev. A modified Remez algorithm for approximate determination of the rational function of the best approximation in Hausdorff metric. *Comptes rendus de l'Academie bulgare des Sciences*, 40(3):13–16, 1987.

[27] R. H. Nochetto, E. Otárola, and A. J. Salgado. A pde approach to fractional diffusion in general domains: a priori error analysis. *Foundations of Computational Mathematics*, 15(3):733–791, 2015.

[28] R. H. Nochetto, E. Otárola, and A. J. Salgado. A pde approach to space-time fractional parabolic problems. *SIAM Journal on Numerical Analysis*, 54(2):848–873, 2016.

[29] L. E. Payne and H. F. Weinberger. An optimal Poincaré inequality for convex domains. *Archive for Rational Mechanics and Analysis*, 5:286–292, Jan. 1960.

[30] X. Ros-Oton. Nonlocal elliptic equations in bounded domains: a survey. *arXiv preprint arXiv:1504.04099*, 2015.

[31] E. B. Saff and H. Stahl. Asymptotic distribution of poles and zeros of best rational approximants to $x^\alpha$ on $[0, 1]$. In "Topics in Complex Analysis", *Banach Center Publications*, volume 31. Institute of Mathematics, Polish Academy of Sciences, Warsaw, 1995.

[32] A. A. Samarskii. *The theory of difference schemes*, volume 240 of *Pure and Applied Mathematics*. Marcel Dekker, Inc., New York, 2001.

[33] F. Song, C. Xu, and G. E. Karniadakis. Computing fractional laplacians on complex-geometry domains: Algorithms and simulations. *SIAM J. Scientific Computing*, 39(4):A1320 – A1344, 2017.
[37] H. Stahl. Best uniform rational approximation of $x^\alpha$ on $[0, 1]$. *Bulletin of the American Mathematical Society*, 28(1):116–122, 1993.

[38] H. R. Stahl. Best uniform rational approximation of $x^\alpha$ on $[0, 1]$. *Acta Mathematica*, 190(2):241–306, 2003.

[39] V. Thomée. *Galerkin finite element methods for parabolic problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006.

[40] T. Usuijima. Error estimates for the lumped mass approximation of the heat equation. *Mem. Numer. Math.*, 6:65–82, 1979.

[41] P. N. Vabishchevich. Numerically solving an equation for fractional powers of elliptic operators. *Journal of Computational Physics*, 282:289–302, 2015.

[42] P. N. Vabishchevich. Numerical solution of time-dependent problems with fractional power elliptic operator. *Computational Methods in Applied Mathematics*, 18(1):111–128, 2018.

[43] R. S. Varga and A. J. Carpenter. Some numerical results on best uniform rational approximation of $x^\alpha$ on $[0, 1]$. *Numerical Algorithms*, 2(2):171–185, 1992.

[44] J. Xu and L. Zikatanov. A monotone finite element scheme for convection-diffusion equations. *Mathematics of Computation*, 68(228):1429–1446, 1999.

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev, bl. 25A, 1113 Sofia, Bulgaria (sharizanov@parallel.bas.bg)

Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, USA (lazarov@math.tamu.edu)

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev, bl. 8, 1113 Sofia, Bulgaria

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev, bl. 25A, 1113 Sofia, Bulgaria (pencho@parallel.bas.bg)

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev, bl. 25A, 1113 Sofia, Bulgaria (margenov@parallel.bas.bg)

Department of Mathematics, Texas A&M University, College Station, TX 77843, USA (pasciak@math.tamu.edu)