Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

A Unified View of Some Numerical Methods for Fractional Diffusion

C. Hofreither

RICAM-Report 2019-12

A Unified View of Some Numerical Methods for Fractional Diffusion

Clemens Hofreither

Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenbergerstr. 69, 4040 Linz, Austria

Abstract

In recent years, a number of numerical methods for the solution of fractional Laplace and, more generally, fractional diffusion problems have been proposed. The approaches are quite diverse and include, among others, the use of best uniform rational approximations, quadrature for Dunford-Taylor-like integrals, finite element approaches for a localized elliptic extension into a space of increased dimensions, and time stepping methods for a parabolic reformulation of the fractional differential equation.

A systematic comparison, both theoretical and experimental, of these approaches has thus far been lacking. A main contribution of the present work is the observation that all approaches mentioned above can, in fact, be interpreted as realizing different rational approximations of a univariate function over the spectrum of the original (non-fractional) diffusion operator. While this is obvious for some of the methods, it is a new result in particular for extension-based and time stepping approaches.

This observation allows us to cast all described methods into a unified theoretical and computational framework, which has a number of benefits. Theoretically, it enables us to develop new convergence proofs for several of the studied methods, clarifies similarities and differences between the approaches, suggests how to design new and improved methods, and allows a direct comparison of the relative performance of the various methods. Practically, it provides a single, simple to implement, efficient and fully parallel algorithm for the realization of all studied methods; for instance, this does away with the need for constructing specific multilevel methods for the efficient realization of the extension methods and lets us parallelize the otherwise inherently sequential time stepping approach.

Based on the insights gained, we also propose a new method based on direct rational approximation which outperforms all investigated methods by a significant margin in our examples.

Finally, we present a detailed numerical study comparing all investigated methods for various fractional exponents and draw conclusions from the results.

Keywords: fractional diffusion, fractional Laplace, numerical methods, rational approximation

1. Introduction

Recently there has been significantly increased interest in fractional differential equations, for instance, as a tool for modeling nonlocal diffusion processes (see, e.g, [1, 2] and the references therein). Correspondingly, the research interest in efficient numerical methods for the approximate solution of such fractional differential equations has risen as well. The main difficulty in problems of this type is that fractional powers of diffusion operators are no longer local, and thus standard FEM discretizations result in dense matrices and therefore are not computationally feasible. An overview of some recently developed methods which attempt to overcome this problem follows below.

It has long been known that the square root of the negative Laplace operator in $\mathbb{R}^d$ can be obtained as the Dirichlet-to-Neumann map of a Laplace problem in the upper half-plane $\mathbb{R}^d \times (0, \infty)$. The characterization of

Email address: clemens.hofreither@ricam.oeaw.ac.at (Clemens Hofreither)

March 25, 2019

an arbitrary fractional power of the Laplacian as a Dirichlet-to-Neumann mapping of an extended problem in a semi-infinite cylinder in a space of one higher dimension was rigorously studied by Caffarelli and Silvestre [3] for problems posed on the full space $\mathbb{R}^d$, and in the later publications [4, 5, 6] for bounded domains $\Omega \subset \mathbb{R}^d$. Nochetto, Otárola and Salgado [7] comprehensively described and analyzed a numerical method for the solution of fractional Laplace problems which uses these ideas by transforming the original nonlocal fractional problem in a d-dimensional domain into a Neumann-to-Dirichlet map for a local, elliptic problem in a $d+1$-dimensional cylinder built on the original domain, and discretizing it with tensor product finite elements. An efficient multilevel solver for the resulting $d+1$-dimensional problem was described in [8]. A hybrid FEM-spectral method based on similar ideas was recently studied by Ainsworth and Glusa [9], and a similar approach based on *hp*-FEM by Banjai et al. [10].

A quite different approach has been proposed by Harizanov, Lazarov, Margenov, Marinov and Vutov [11] based on best uniform rational approximations (BURA). It is observed that, on the discrete level, the fractional diffusion problem may be viewed as taking a fractional matrix power of a standard stiffness matrix. In order to approximate the negative power $-s$ of a matrix, a rational function approximating the mapping $z \to z^{-s}$ is constructed via a modified Remez algorithm. This rational function is then decomposed into partial fractions, which allows the computation of the approximate matrix power based only on inversions of spectrally equivalent shifted versions of the original (sparse) stiffness matrix. A variant of this approach has been proposed in [12]. Rational approximations, although not best uniform ones, have also been used in the setting of fractional stochastic differential equations in [13].

Another class of methods was suggested by Bonito and Pasciak [14]. Here, integral representations of the mapping $z \mapsto z^{-s}$ are discretized by means of specially constructed quadrature formulae. These discretizations again take the form of rational functions and thus can be applied to the stiffness matrix as described above for the BURA-based method. This idea has been further used for fractional elliptic stochastic partial differential equations in [15].

Finally, a reformulation of the fractional diffusion problem as a parabolic problem with a local operator and an artificial time axis was given by Vabishchevich [16, 17]. He then applies an implicit time stepping method to this problem in order to obtain an approximation to the original fractional differential equation.

A main contribution of the present work is that all of the above methods can in fact be interpreted as methods based on rational approximation of a univariate function. While this is obvious for the BURA and integral quadrature-based approaches, it is a new result for the extension- and time stepping-based methods. This observation allows us to develop a unified error analysis for all mentioned methods which is only based on the quality of this rational approximation. In particular, this analysis fully decouples the spatial discretization from the rational approximation error, is dimension-independent and depends only on the spectrum of the discrete diffusion operator.

Importantly, the involved rational functions are not merely theoretical tools, but are in fact computable for all methods. This means that all investigated methods can be realized using the simple, efficient and fully parallel algorithm that applied to the BURA method. For instance, for the efficient solution of the extension method from [7], a sophisticated multilevel method with line smoothers was developed and analyzed in [8]. The new point of view based on rational approximation makes it clear that this is not necessary since the extension method can now be efficiently realized as a rational approximation method. This also immediately provides an efficient realization for extension methods with higher-order discretizations in the extended dimension, which we will exploit in order to give a variant based on cubic splines. Another example of the benefit of the new viewpoint is that the time stepping method of Vabishchevich [16] can be trivially parallelized using this approach, whereas the original method was by necessity sequential.

As another application of the new insights gained, we also develop a novel rational approximation method which is not based on best uniform rational approximations, but instead on the adaptive AAA algorithm [18] for rational approximation, increasing robustness and speed of this approximation step. The resulting method performs significantly better than the other approaches in our comparison.

Comparisons of some of these numerical methods have been performed in earlier works [19, 12]. The interpretation of extension and time stepping methods as rational approximation methods appears to be novel to the present work.

The remainder of this paper is structured as follows. We first introduce the fractional diffusion model

problem, its discretization and a simple, but inefficient solution method in Section 2. We then introduce rational approximation methods abstractly in Section 3, discuss their efficient implementation and establish a simple theoretical framework for their analysis. As a particular instance of this class of methods, we describe the BURA-based method in Section 4 and briefly discuss its analysis in the unified framework. Based on the discussion here, we propose a new rational approximation method in Section 5. We then discuss the quadrature-based method of Bonito and Pasciak as a rational approximation method in Section 6. We study extension methods in Section 7, where we develop the main part of our new theory, namely how to interpret and analyze such methods in the abstract rational approximation framework. The last method under investigation is the parabolic time stepping method of Vabishchevich in Section 8, which we also rewrite as a rational approximation method. The final part of this study is a detailed numerical investigation of all investigated methods in Section 9, where we also discuss the results.

2. The fractional diffusion problem and its discretization

Given some open and bounded domain $\Omega \subset \mathbb{R}^d$, $s \in (0,1)$, and a suitable right-hand side f defined on Ω, we seek a function u which solves the equation involving the fractional Laplace operator

$$\begin{aligned} (-\Delta)^s u &= f \qquad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \Gamma = \partial\Omega, \end{aligned}$$

or more generally, the fractional diffusion equation

$$\begin{aligned} \mathcal{L}^s u &= f \qquad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \Gamma = \partial\Omega, \end{aligned} \tag{1}$$

where $\mathcal{L}u = -\operatorname{div}(A\nabla u)$ is a self-adjoint, elliptic diffusion operator with $A(x) \in \mathbb{R}^{d\times d}$ symmetric and uniformly positive definite. The definition of the fractional operator power $\mathcal{L}^s$ is in itself not trivial, and we refer the reader to [20, 2] and the references therein for various approaches to the definition of fractional derivatives and fractional operator powers.

In the present paper, we will restrict ourselves to bounded domains with homogeneous Dirichlet boundary conditions, which allows a relatively simple spectral definition of the involved operators. Under the above assumptions, $\mathcal{L}$ admits a system of eigenfunctions u_j with corresponding eigenvalues $\lambda_j > 0$ such that

$$\mathcal{L}u_j = \lambda_j u_j \qquad \forall j = 1, 2, \dots$$

and $(u_i, u_j) = \delta_{ij}$, where $(\cdot,\cdot)$ denotes the L_2-inner product in Ω. A fractional power of $\mathcal{L}$ can then be defined as

$$\mathcal{L}^s u = \sum_{j=1}^{\infty} \lambda_j^s (u, u_j) u_j. \tag{2}$$

We are interested in discretizations of the above problem. To that end, let $V_h \subset H_0^1(\Omega)$ denote a finite-dimensional space which satisfies the homogeneous Dirichlet boundary conditions and which is spanned by a basis $(\varphi_j)_{j=1}^n$. By Galerkin discretization, we obtain a discrete system of eigenfunctions $u_j^h \in V_h$ such that

$$(A\nabla u_j^h, \nabla v^h) = \lambda_j^h (u_j^h, v^h) \qquad \forall v^h \in V_h \; \forall j \in \{1, \dots, n\}$$

with discrete eigenvalues $\lambda_j^h > 0$. Motivated by (2), we might approximate the action of a fractional power of $\mathcal{L}$ as

$$\mathcal{L}^s u \approx \sum_{j=1}^{n} (\lambda_j^h)^s (u, u_j^h) u_j^h.$$

In particular, by applying $\mathcal{L}^{-s}$ on both sides, we may obtain an approximation to the solution u of (1) by the formula

$$u \approx \sum_{j=1}^{n} \frac{(f, u_j^h)}{(\lambda_j^h)^s} u_j^h. \tag{3}$$

This method is called the *discrete eigenfunction method* (DEM) in [2].

In order to realize the DEM, we introduce the standard symmetric and positive definite stiffness and mass matrices $K, M \in \mathbb{R}^{n\times n}$ as

$$K_{ij} = (A\nabla\phi_j, \nabla\phi_i), \quad M_{ij} = (\phi_j, \phi_i) \qquad \forall i, j = 1, \ldots, n.$$

Then the matrix $U \in \mathbb{R}^{n\times n}$ which contains, columnwise, the coefficients of the discrete eigenfunctions with respect to the basis $(\varphi_j)_{j=1}^n$, that is, $u_j^h = \sum_{k=1}^n U_{kj}\varphi_k$, is given by the solution of the generalized eigenvalue problem

$$KU = MU\Lambda, \qquad U^T MU = I_n, \tag{4}$$

where $\Lambda = \mathrm{diag}(\lambda_j^h)_{j=1}^n$ is the diagonal matrix containing the discrete eigenvalues and I_n is the $n \times n$ identity matrix. Let now $\mathbf{f} \in \mathbb{R}^n$ denote the coefficient vector of the L_2-projection of f with respect to the basis of V_h, and $\mathbf{u} \in \mathbb{R}^n$ the coefficient vector of the discrete solution, i.e., the right-hand side of (3). Due to

$$(f, u_j^h) = (M\mathbf{f}, U\mathbf{e}_j) = (U^T M\mathbf{f}, \mathbf{e}_j) \qquad \forall j = 1, \ldots, n$$

with $\mathbf{e}_j \in \mathbb{R}^n$ being the unit vector pointing in the positive j-th coordinate direction, (3) can be written in matrix-vector notation as

$$\mathbf{u} = U\Lambda^{-s}U^T M\mathbf{f}. \tag{5}$$

From (4) we also have $U^T M = U^{-1}$ and $M^{-1}K = U\Lambda U^{-1}$. The latter is a diagonalization of $M^{-1}K$ and therefore allows us to take arbitrary matrix powers $s \in \mathbb{R}$ as $(M^{-1}K)^s = U\Lambda^s U^{-1}$. Using these relations in (5), we obtain

$$\mathbf{u} = U\Lambda^{-s}U^{-1}\mathbf{f} = (M^{-1}K)^{-s}\mathbf{f}. \tag{6}$$

Note that $M^{-1}K$ can be viewed as the discretization of the operator $\mathcal{J}\mathcal{L}$, where $\mathcal{J} : (H_0^1(\Omega))' \to H_0^1(\Omega)$ is the Riesz isomorphism.

The discrete eigenfunction method is a convenient way of computing solutions to fractional diffusion problems, and a quasi-optimal error estimate for it is given by [14, Theorem 4.3]. Unfortunately, it requires the solution of the discrete generalized eigenvalue problem (4) which has computational complexity $\mathcal{O}(n^3)$ and thus can only be used for problems of modest size.

3. Rational approximation methods

Referring to (6), we see that the solution

$$\mathbf{u} = (M^{-1}K)^{-s}\mathbf{f}$$

of the discrete eigenfunction method can be computed by taking a matrix power of $M^{-1}K$. Instead of computing this matrix power exactly, one class of methods tries to determine a rational function

$$r(z) = \frac{P_{k'}(z)}{Q_k(z)},$$

where $P_{k'} \in \mathcal{P}_{k'}$ and $Q_k \in \mathcal{P}_k$ are algebraic polynomials of the given maximum degree, which approximates the function z^{-s} for $z > 0$. A discrete solution is then obtained by

$$\mathbf{u}_r = r(M^{-1}K)\mathbf{f} = Ur(\Lambda)U^{-1}\mathbf{f}. \tag{7}$$

If, assuming $k' \le k$, the rational function is given in partial fraction decomposition form

$$r(z) = c_0 + \sum_{j=1}^{k} \frac{c_j}{z - d_j}, \tag{8}$$

the solution can be computed as

$$\mathbf{u}_r = c_0 + \sum_{j=1}^{k} c_j (M^{-1}K - d_j I_n)^{-1}\mathbf{f}.$$

Computing $\mathbf{u}_r$ using this formula involves computing solutions $\mathbf{w}_j = (M^{-1}K - d_j I_n)^{-1}\mathbf{f}$ to some shifted problems which are equivalently described by

$$(K - d_j M)\mathbf{w}_j = M\mathbf{f}.$$

This shifted problem amounts to solving the diffusion-reaction problem $-\operatorname{div}(A\nabla w_j) - d_j w_j = f$ with homogenous Dirichlet boundary conditions in Ω, which is spectrally equivalent to the original non-fractional diffusion problem assuming that the poles d_j are nonpositive. A good preconditioner for the diffusion problem can then serve as a preconditioner for the shifted problem. Thus, assuming we have an optimal complexity solver for the diffusion problem, we can realize the rational approximation method in $\mathcal{O}(kn)$ operations.

It is a simple matter to obtain an error estimate for the rational approximation method in terms of the approximation quality of $r(z)$ to the function z^{-s}.

Theorem 1. *The solution $u_r \in V_h$ obtained by the rational approximation method* (7) *and the solution $u \in V_h$ obtained by the discrete eigenfunction method* (5) *satisfy the relation*

$$\|u - u_r\|_{L_2(\Omega)} \leq \|(\cdot)^{-s} - r(\cdot)\|_{L_\infty(\lambda^h_{\min},\lambda^h_{\max})} \|f\|_{L_2(\Omega)},$$

where $\lambda^h_{\min}$ and $\lambda^h_{\max}$ are the smallest and largest discrete eigenvalues of the spatial discretization as described by (4).

Proof. From (5) and (7) we immediately obtain

$$\mathbf{u} - \mathbf{u}_r = U(\Lambda^{-s} - r(\Lambda))U^{-1}\mathbf{f}.$$

The functions $u, u_r \in V_h$ are represented by the coefficient vectors $\mathbf{u}, \mathbf{u}_r$, respectively. For any such function, (4) yields

$$\|u\|^2_{L_2(\Omega)} = \langle M\mathbf{u}, \mathbf{u}\rangle = \langle U^{-T}U^{-1}\mathbf{u}, \mathbf{u}\rangle = \|U^{-1}\mathbf{u}\|^2.$$

Thus we obtain

$$\|u - u_r\|_{L_2(\Omega)} = \|U^{-1}(\mathbf{u} - \mathbf{u}_r)\| = \|(\Lambda^{-s} - r(\Lambda))U^{-1}\mathbf{f}\| \leq \max_{j=1,\ldots,n} |(\lambda^h_j)^{-s} - r(\lambda^h_j)|\ \|f\|_{L_2(\Omega)}.$$

The maximum over all eigenvalues can be bounded from above by the maximum norm over the interval containing all eigenvalues. □

4. The BURA method

A rational approximation method for the solution of (1) was proposed in [11] based on best uniform rational approximations (BURA). The approach first computes the best rational approximation, with degree p for both the numerator and the denominator, of the function $z^{\beta-s}$, $z \in [0, 1]$, where a positive integer $\beta > s$, usually $\beta = 1$, is chosen in order to avoid the blowup at $z = 0$. Denote this approximation by $\sigma_{\beta,s,p}(z)$. A rational approximation to z^{-s} in $[0, 1]$ is obtained by taking $\sigma_{\beta,s,p}(z)/z^\beta$, which results in a rational function with degrees $(p, p + \beta)$ and thus allows a partial fraction decomposition of the form

$$\frac{\sigma_{\beta,s,p}(z)}{z^\beta} = \sum_{j=1}^{p+\beta} \frac{c_j}{z - d_j},$$

where the additive constant c_0 from (8) vanishes since the numerator has lower degree than the denominator, and the poles d_j are nonpositive. In [11], the matrix under consideration is then rescaled so as to have its spectrum contained in $[0,1]$. Equivalently, we can rescale the rational approximation to the interval $[0,\lambda^h_{\max}]$ by noting that

$$\begin{aligned} z^{-s} &= (\lambda^h_{\max})^{-s}\left(\frac{z}{\lambda^h_{\max}}\right)^{-s} \\ &\approx (\lambda^h_{\max})^{-s}\frac{\sigma_{\beta,s,p}(z/\lambda^h_{\max})}{(z/\lambda^h_{\max})^{\beta}} = (\lambda^h_{\max})^{-s}\sum_{j=1}^{p+\beta}\frac{c_j}{z/\lambda^h_{\max}-d_j} = \sum_{j=1}^{p+\beta}\frac{(\lambda^h_{\max})^{1-s}c_j}{z-\lambda^h_{\max}d_j} =: r_{\beta,s,p}(z), \end{aligned}$$

which yields the rescaled partial fraction decomposition which can finally be used in the rational approximation method as described in Section 3. Note that

$$r_{\beta,s,p}(z) = \frac{\sigma_{\beta,s,p}(z/\lambda^h_{\max})}{z^{\beta}(\lambda^h_{\max})^{s-\beta}}. \tag{9}$$

It should be noted that the original derivation of the BURA method in [11] does not explicitly mention the use of the mass matrix as we did in Section 3, but incorporating the mass matrix appears to be the theoretically sound approach especially for non-structured FEM discretizations.

The computation of the best approximation $\sigma_{\beta,s,p}$ in [11] is based on a modified Remez algorithm and involves subtle details related to stability and convergence. Even the sophisticated algorithm proposed in [11] requires the use of extended precision arithmetic already for relatively small values of p, and is not easy to implement. A high-quality implementation of an algorithm for best rational approximation is contained in the `chebfun` package [21], and could be used instead.

In [11], the error estimate

$$\max_{z\in[0,1]}|z^{\beta-s}-\sigma_{\beta,s,p}(z)| \approx 4^{1+\beta-s}|\sin\pi(\beta-s)|\exp(-2\pi\sqrt{(\beta-s)p}) := E_{\beta,s,p}$$

is given. Although this estimate stems from an asymptotic identity, it was demonstrated numerically in [11] that it is rather accurate already for low values of p. From (9) we obtain

$$\begin{aligned} |z^{-s}-r_{\beta,s,p}(z)| &= z^{-\beta}|z^{\beta-s}-z^{\beta}r_{\beta,s,p}(z)| = z^{-\beta}\left|z^{\beta-s}-\frac{\sigma_{\beta,s,p}(z/\lambda^h_{\max})}{(\lambda^h_{\max})^{s-\beta}}\right| \\ &= z^{-\beta}(\lambda^h_{\max})^{\beta-s}\left|\left(\frac{z}{\lambda^h_{\max}}\right)^{\beta-s}-\sigma_{\beta,s,p}\left(\frac{z}{\lambda^h_{\max}}\right)\right| \end{aligned}$$

and thus, using the above error estimate,

$$\max_{z\in[\lambda^h_{\min},\lambda^h_{\max}]}|z^{-s}-r_{\beta,s,p}(z)| \lesssim \frac{(\lambda^h_{\max})^{\beta-s}}{(\lambda^h_{\min})^{\beta}}E_{\beta,s,p}.$$

With Theorem 1, we obtain the error bound for $u_{\beta,s,p}\in V_h$ computed using the BURA method

$$\|u-u_{\beta,s,p}\|_{L_2(\Omega)} \lesssim \frac{(\lambda^h_{\max})^{\beta-s}}{(\lambda^h_{\min})^{\beta}}E_{\beta,s,p}\|f\|_{L_2(\Omega)} = (\lambda^h_{\min})^{-s}\kappa^{\beta-s}E_{\beta,s,p}\|f\|_{L_2(\Omega)}. \tag{10}$$

We can conclude that, while the convergence remains exponentially fast in p due to the term $E_{\beta,s,p}$, the estimate depends unfavorably on the condition number $\kappa=\lambda^h_{\max}/\lambda^h_{\min}$ of the matrix $M^{-1}K$. The reason for this lies in the fact that, while $\sigma_{\beta,s,p}(z)$ is the best rational approximation to $z^{\beta-s}$, there is no reason to believe that $\sigma_{\beta,s,p}(z)/z^{\beta}$ is the best or even a particularly good rational approximation to z^{-s}. This issue is illustrated in Table 1, where errors for these two approaches to approximation are compared. Indeed, whereas a best approximation exhibits an error that equioscillates over the entire interval, the error $|z^{-s}-r_{\beta,s,p}(z)|$ dominates for z close to $\lambda^h_{\min}$. We also note that the dependence on κ gets worse for smaller s.

p	$\|\sigma_{1,\frac{1}{2},p}(z) - z^{\frac{1}{2}}\|_{L^\infty(0,1)}$	$\|\sigma_{1,\frac{1}{2},p}(z)/z - z^{-\frac{1}{2}}\|_{L^\infty(10^{-6},1)}$	$\|q(z) - z^{-\frac{1}{2}}\|_{L^\infty(10^{-6},1)}$
1	$4.368902 \cdot 10^{-2}$	$4.2692 \cdot 10^{+4}$	$4.3379 \cdot 10^{+01}$
2	$8.501610 \cdot 10^{-3}$	$7.5175 \cdot 10^{+3}$	$8.2779 \cdot 10^{+00}$
3	$2.282106 \cdot 10^{-3}$	$1.3419 \cdot 10^{+3}$	$2.1071 \cdot 10^{+00}$
4	$7.365654 \cdot 10^{-4}$	$7.9756 \cdot 10^{+1}$	$6.0665 \cdot 10^{-01}$
5	$2.689574 \cdot 10^{-4}$	$2.5765 \cdot 10^{+2}$	$1.8228 \cdot 10^{-01}$
6	$1.074714 \cdot 10^{-4}$	$9.2119 \cdot 10^{+0}$	$5.5341 \cdot 10^{-02}$
7	$4.603655 \cdot 10^{-5}$	$3.8088 \cdot 10^{+1}$	$1.6832 \cdot 10^{-02}$
8	$2.085162 \cdot 10^{-5}$	$2.0737 \cdot 10^{+1}$	$5.1207 \cdot 10^{-03}$
9	$9.889319 \cdot 10^{-6}$	$8.2912 \cdot 10^{+0}$	$1.5578 \cdot 10^{-03}$
10	$4.875955 \cdot 10^{-6}$	$3.2105 \cdot 10^{+0}$	$4.7394 \cdot 10^{-04}$
11	$2.485587 \cdot 10^{-6}$	$1.3973 \cdot 10^{+0}$	$1.4419 \cdot 10^{-04}$
12	$1.304380 \cdot 10^{-6}$	$7.5802 \cdot 10^{-1}$	$4.3869 \cdot 10^{-05}$
13	$7.022329 \cdot 10^{-7}$	$4.8613 \cdot 10^{-1}$	$1.3349 \cdot 10^{-05}$
14	$3.867572 \cdot 10^{-7}$	$3.2962 \cdot 10^{-1}$	$4.0632 \cdot 10^{-06}$

Table 1: Errors for rational approximation of $z^{1/2}$ and $z^{-1/2}$ for varying degree p (equal degree for numerator and denominator). Here, $\sigma_{1,\frac{1}{2},p}(z)$ is the (p,p)-BURA to $z^{\frac{1}{2}}$ in $[0,1]$, and $q(z)$ is the (p,p)-BURA to $z^{-\frac{1}{2}}$ in $[10^{-6},1]$. Note that the latter approximation to $z^{-1/2}$ is not only better by several orders of magnitude than the one based on approximation of $z^{1/2}$, but also converges with a higher rate.

5. A novel method based on direct rational approximation

The discussion in the previous section suggests that better approximations may be obtained by approximating not $z^{\beta-s}$, but instead directly z^{-s} with a rational function. This necessitates knowledge of both the largest and the smallest eigenvalue $[\lambda^h_{\min}, \lambda^h_{\max}]$, such that the function z^{-s} can be approximated over this interval in order to avoid the blowup at $z = 0$. Good bounds for these eigenvalues can be computed, say, by power iteration or by the Implicitly Restarted Lanczos Method (IRLM) implemented in the ARPACK package[1].

One approach is then to compute again the best uniform rational approximation over this interval, which would yield, according to Theorem 1, a close to optimal method. An analysis of the dependence of the best approximation error on the condition number κ and on s is an interesting research direction. However, computing this approximation again suffers from numerical stability problems and may only be feasible for low degrees, at least using standard double precision arithmetic.

Another approach is to use a different method for rational approximation which does not necessarily yield best uniform approximations. A number of such methods is available in the literature, and we consider the so-called AAA ("adaptive Antoulas-Anderson") algorithm recently proposed by Nakatsukasa, Sète, and Trefethen [18]. The method is based on representation of the rational approximant in barycentric form and greedy selection of the support points. In addition to being simple to implement, it does not suffer from the same numerical instabilities as the Remez algorithm does and appears to provide good approximations. A Matlab implementation is contained in the `chebfun` package [21], and a Python implementation was made available by the author of this paper[2].

As the AAA algorithm requires a discrete set of function values, we sample z^{-s} over a uniform grid with M points in the interval $[\lambda^h_{\min}, \lambda^h_{\max}]$ and pass these values as input to the AAA algorithm, furthermore specifying the desired degree p of the approximation or the desired error tolerance. The used barycentric representation of the rational function allows a simple computation of the poles (d_j) and residues (c_j) for the representation (8) (cf. [18]), and the poles are nonpositive in our setting. The resulting rational

[1] `https://www.caam.rice.edu/software/ARPACK/`

[2] `https://github.com/c-f-h/aaa`

function with degrees (p, p) is then used for a rational approximation method as described in Section 3. The AAA algorithm has complexity $\mathcal{O}(Mp^3)$ and terminated in less than a second for the relevant examples in Section 9.

6. Rational approximations based on quadrature of integral representations

A different class of rational approximations was proposed by Bonito and Pasciak [14] based on the integral representation

$$z^{-s} = \frac{2\sin(\pi s)}{\pi} \int_{-\infty}^{\infty} \frac{e^{2sy}}{1 + e^{2y} z} dy.$$

In fact, three families of quadrature-based approximations are given in [14], but we focus on the third class based on the above integral as it is the only one to exhibit exponential convergence. The authors chose a parameter $q > 0$ and quadrature nodes $y_\ell = \ell q$ for $\ell \in \{-M, \dots, N\}$ with integers M, N. A quadrature rule is then constructed as

$$\frac{2q\sin(\pi s)}{\pi} \sum_{\ell=-M}^{N} \frac{e^{2sy_\ell}}{1 + e^{2y_\ell} z} =: r^{\mathrm{BP3}}(z),$$

which obviously can be brought into the form (8) with simple transformations and has negative poles, allowing a realization as in Section 3. For $M = N$ and $q = 1/\sqrt{N}$, the authors prove the error bound

$$|z^{-s} - r^{\mathrm{BP3}}(z)| \le \frac{2\sin(\pi s)}{\pi} \left[\frac{s^{-1} + ((1-s)\lambda_{\min}^h)^{-1}}{2\sinh(\pi^2/(4q))} e^{-\pi^2/(4q)} + \frac{1}{2s} e^{-2\beta/q} + \frac{1}{(2-2s)\lambda_{\min}^h} e^{-(2-2s)/q} \right].$$

In order to equilibrate the error terms, they suggest for $q > 0$ the choices

$$M = \left\lceil \frac{\pi^2}{4sq^2} \right\rceil, \qquad N = \left\lceil \frac{\pi^2}{4(1-s)q^2} \right\rceil \tag{11}$$

and show that then asymptotically

$$|z^{-s} - r^{\mathrm{BP3}}(z)| \lesssim \frac{2\sin(\pi s)}{\pi} \left(\frac{1}{s} + \frac{1}{(1-\beta)\lambda_{\min}^h} \right) e^{-\pi^2/(2q)}$$

as $q \to 0$. Since roughly $q \sim 1/\sqrt{N_{\mathrm{terms}}}$ with the number of terms in the rational approximation, the exponential term seems to exhibit a similar rate as that in the BURA method; numerical comparisons will be presented in Section 9.

7. The extension method

7.1. Formulation and discretization

We summarize here the approach of Nochetto, Otárola and Salgado [7]. We let $\alpha = 1 - 2s \in (-1, 1)$ and introduce an extended boundary value problem for a function $U(x, y)$, $x \in \Omega$, $y \in (0, \infty)$ given by

$$\begin{aligned} -\operatorname{div}(y^\alpha \tilde{A} \nabla U) &= 0 && \text{in } \Omega \times (0, \infty), \\ \lim_{y \to \infty} U(x, y) &= 0 && \forall x \in \Omega, \\ U(x, y) &= 0 && \forall x \in \partial\Omega, y \in (0, \infty), \\ -(\lim_{y \to 0} y^\alpha \partial_y U(x, y)) &= d_s f(x) && \forall x \in \Omega, \end{aligned}$$

where $\tilde{A}(x, y) = \operatorname{diag}(A(x), 1) \in \mathbb{R}^{(d+1)\times(d+1)}$ and

$$d_s = 2^{1-2s} \frac{\Gamma(1-s)}{\Gamma(s)}$$

is a dimension-independent constant involving the gamma function Γ. The solution u of problem (1) is then given by the Dirichlet trace

$$u(x) = U(x, 0) \qquad \forall x \in \Omega.$$

Since U decays exponentially with y, it is justified to truncate the above problem to a finite cylinder $\Omega \times (0, Y)$ with some $Y > 0$ and enforce homogeneous Dirichlet boundary conditions on $\Omega \times \{Y\}$ instead of the decay condition. With suitable test functions $V(x, y)$ which vanish at the Dirichlet boundaries $\partial\Omega \times (0, Y)$ and $\Omega \times \{Y\}$, the variational formulation reads

$$a(U, V) := \int_{\Omega\times(0,Y)} y^\alpha \tilde{A}\nabla U \cdot \nabla V \, d(x, y) = d_s \int_\Omega f(x)V(x, 0) \, dx. \tag{12}$$

We introduce a discrete space for the extended direction $W_h = \operatorname{span}\{\psi_j : j = 1, \dots, m\} \subset C(0, Y)$ with the Dirichlet boundary condition $\psi_j(Y) = 0$. The problem (12) is then discretized using tensor product basis functions

$$\varphi_i(x)\psi_j(y) \qquad \forall i = 1, \dots, n, j = 1, \dots, m, x \in \Omega, y \in (0, Y).$$

In order to compute the stiffness matrix, we let $U(x, y) = \varphi_{i_1}(x)\psi_{i_2}(y)$ and $V(x, y) = \varphi_{j_1}(x)\psi_{j_2}(y)$ and find that

$$\begin{aligned} a(U, V) &= \int_{\Omega\times(0,Y)} y^\alpha \begin{pmatrix} A\nabla_x\varphi_{i_1}(x)\psi_{i_2}(y) \\ (\varphi_{i_1}(x)\psi'_{i_2}(y)) \end{pmatrix} \cdot \begin{pmatrix} \nabla_x\varphi_{j_1}(x)\psi_{j_2}(y) \\ (\varphi_{j_1}(x)\psi'_{j_2}(y)) \end{pmatrix} d(x, y) \\ &= \left(\int_0^Y y^\alpha \psi_{i_2}\psi_{j_2} dy\right)\left(\int_\Omega A\nabla_x\varphi_{i_1} \cdot \nabla_x\varphi_{j_1} dx\right) + \left(\int_0^Y y^\alpha \psi'_{i_2}\psi'_{j_2} dy\right)\left(\int_\Omega \varphi_{i_1}\varphi_{j_1} dx\right). \end{aligned}$$

Thus the stiffness matrix for the extended problem is given as the sum of the two Kronecker products

$$\mathcal{A}^{(\alpha)} = M_y^{(\alpha)} \otimes K + K_y^{(\alpha)} \otimes M, \tag{13}$$

where $M_y^{(\alpha)} \in \mathbb{R}^{m\times m}$ and $K_y^{(\alpha)} \in \mathbb{R}^{m\times m}$ are the weighted mass and stiffness matrices

$$[M_y^{(\alpha)}]_{ij} = \int_0^Y y^\alpha \psi_j(y)\psi_i(y) \, dy, \qquad [K_y^{(\alpha)}]_{ij} = \int_0^Y y^\alpha \psi'_j(y)\psi'_i(y) \, dy,$$

and K and M are as introduced in Section 2. Formula (13) provides a simple way of computing the stiffness matrix for the extended problem.

7.2. Connection to the rational approximation method

In the following, we demonstrate that the extension method can be interpreted as a rational approximation method. As we did for the spatial problem in (4), we introduce the eigenvalue decomposition of the discretization in the extended dimension

$$K_y^{(\alpha)}V = M_y^{(\alpha)}V\Lambda_y, \qquad V^T M_y^{(\alpha)}V = I_m, \tag{14}$$

where $\Lambda_y = \operatorname{diag}(\mu_1^h, \dots, \mu_m^h)$ contains the discrete eigenvalues and $V \in \mathbb{R}^{m\times m}$ contains the coefficients of the discrete eigenfunctions. From these relations, we also obtain $V^T K_y^{(\alpha)} V = \Lambda_y$. With this, and using basic properties of the Kronecker product, (13) may be diagonalized as

$$\mathcal{A}^{(\alpha)} = (V^{-T} \otimes U^{-T})(\Lambda_y \otimes I_n + I_m \otimes \Lambda)(V^{-1} \otimes U^{-1}) = (V^{-T} \otimes U^{-T})D(V^{-1} \otimes U^{-1}) \tag{15}$$

with the diagonal matrix

$$D = \Lambda_y \otimes I_n + I_m \otimes \Lambda \in \mathbb{R}^{mn\times mn}.$$

The right-hand side of (12), represented as a vector in the tensor product basis, can be written as

$$d_s(\mathbf{c} \otimes M\mathbf{f}) \in \mathbb{R}^{mn},$$

where $\mathbf{c} \in \mathbb{R}^m$ is the collocation vector of the basis functions in the extended space basis evaluated at 0, that is, $c_k = \psi_k(0)$ for $k = 1, \dots, m$. (For a basis which is interpolatory at 0, we simply have $\mathbf{c} = \mathbf{e}_1$.) We want to solve the extended problem and then read off the Dirichlet data in order to obtain the coefficients $\mathbf{u}$ of the discrete solution. This can be expressed as

$$\mathbf{u} = d_s(\mathbf{c}^T \otimes I_n)(\mathcal{A}^{(\alpha)})^{-1}(\mathbf{c} \otimes M\mathbf{f}).$$

By direct inversion of (15), we obtain

$$\mathbf{u} = d_s(\mathbf{c}^T V \otimes U)D^{-1}(V^T\mathbf{c} \otimes U^T M\mathbf{f}),$$

and with $\mathbf{b} := V^T\mathbf{c}$, which is the vector of Dirichlet boundary values of the discrete eigenfunctions, we obtain the representation

$$\mathbf{u} = U\left[d_s(\mathbf{b}^T \otimes I_n)D^{-1}(\mathbf{b} \otimes I_n)\right]U^T M\mathbf{f}$$

for the solution of the fractional diffusion problem using the extension method. Comparing this to the analogous expression (5) for the discrete eigenfunction method, the question becomes how

$$E := d_s(\mathbf{b}^T \otimes I_n)D^{-1}(\mathbf{b} \otimes I_n)$$

relates to $\Lambda^{-1/2}$. Denoting the diagonal blocks of D by

$$D_k = \mu_k^h I_n + \Lambda \in \mathbb{R}^{n\times n}, \qquad k = 1, \dots, m,$$

and doing some linear algebra we find that

$$E = d_s \sum_{k=1}^m b_k^2 D_k^{-1} = d_s \operatorname{diag}\left(\sum_{k=1}^m \frac{b_k^2}{\mu_k^h + \lambda_i^h}\right)_{i=1}^m,$$

a diagonal matrix. Denoting the discrete eigenfunctions associated with problem (14) by $v_1^h, \dots, v_m^h \in W_h$, we may write

$$E = r(\Lambda), \qquad r(z) = d_s \sum_{k=1}^m \frac{v_k^h(0)^2}{\mu_k^h + z}. \tag{16}$$

and thus

$$\mathbf{u} = Ur(\Lambda)U^T M\mathbf{f} = Ur(\Lambda)U^{-1}\mathbf{f} = r(M^{-1}K)\mathbf{f},$$

the form of a rational approximation method as introduced in Section 3. We summarize this result as follows.

Theorem 2. *The extension method for solving* (1) *using basis functions* $(\psi_j)_{j=1}^m$ *in the extended direction with* $\psi_j(Y) = 0$ *is equivalent to a rational approximation method of the form* (7) *with the rational function*

$$r(z) = d_s \sum_{k=1}^m \frac{v_k^h(0)^2}{\mu_k^h + z},$$

where $v_k^h \in W_h$ *and* $\mu_k^h > 0$ *are the discrete eigenfunctions and eigenvalues of the one-dimensional problem*

$$(y^\alpha (v_k^h)',\ \psi_i')_{L_2(0,Y)} = \mu_k^h (y^\alpha v_k^h,\ \psi_i)_{L_2(0,Y)} \qquad \forall i = 1, \dots, m.$$

Note that the eigenvalue problem (14) tends to be small as usually $m \ll n$ and thus can be numerically solved without problems. This means that the rational function r is computable, and thus the extension method can be realized as a rational approximation method without needing to construct an efficient solver for the extended problem (12).

7.3. Error analysis as a rational approximation method

Since we have now established that the extension method is, in fact, a particular rational approximation method, an error estimate for it follows from Theorem 1 in terms of how well $r(z)$ approximates z^{-s} in the interval $[\lambda_{\min}^h, \lambda_{\max}^h]$. In the following, we will study the rescaled rational function

$$\rho_h(z) := \sum_{k=1}^{m} \frac{v_k^h(0)^2}{\mu_k^h + z}.$$

We can characterize $\rho_h(z)$ in terms of a Neumann-to-Dirichlet mapping for a one-dimensional problem.

Theorem 3. *Consider, for $z > 0$, the variational problem to find a function $v \in W_h$ such that*

$$\int_0^Y (y^\alpha v' w' + z y^\alpha v w) dy = w(0) \qquad \forall w \in W_h. \tag{17}$$

Then

$$\rho_h(z) = v(0).$$

Proof. The problem (17) has the matrix-vector formulation

$$(K_y^{(\alpha)} + z M_y^{(\alpha)})\mathbf{v} = \mathbf{c}.$$

Since

$$K_y^{(\alpha)} + z M_y^{(\alpha)} = V^{-T}(\Lambda_y + zI)V^{-1},$$

we obtain

$$\mathbf{v} = V(\Lambda_y + zI)^{-1} V^T \mathbf{c}.$$

By reading off the Dirichlet data at $y = 0$, we obtain

$$v(0) = \mathbf{c}^T \mathbf{v} = \mathbf{b}^T (\Lambda_y + zI)^{-1} \mathbf{b} = \sum_{k=1}^{m} \frac{b_k^2}{\mu_k^h + z} = \rho_h(z). \qquad \square$$

It is easy to see that (17) is the variational formulation of the ordinary differential equation

$$-(y^\alpha v'(y))' + z y^\alpha v(y) = 0 \quad \forall y \in (0, Y), \quad -\lim_{y \to 0^+} (y^\alpha v'(y)) = 1, \quad v(Y) = 0. \tag{18}$$

This shows that, indeed, $\rho_h(z)$ realizes the discrete Neumann-to-Dirichlet mapping at $y = 0$ for this problem when discretized in W_h. These considerations also make it clear that $\rho_h(z)$ converges to the exact Neumann-to-Dirichlet map as W_h becomes dense in $C(0, Y)$. Thus, it makes sense to define

$$\rho(z) := v_z(0),$$

where $v = v_z$ is the exact solution of (18).

Remark 1. Until now, we have not made clear why $\rho_h(z)$ should approximate z^{-s}. We can obtain a first hint in the case $s = 1/2$ ($\alpha = 0$), in which case we can use the known eigenfunctions and eigenvalues

$$v_j(y) = \sqrt{\frac{2}{Y}} \cos\left(\frac{(2j-1)\pi y}{2Y}\right), \qquad \mu_j = \frac{(2j-1)^2 \pi^2}{4Y^2}$$

of the negative second derivative operator with mixed Neumann-Dirichlet boundary conditions. These functions have the required normalization $\|v_j\|_{L_2(0,Y)} = 1$. Using some trigonometric identities, we find that

$$\sum_{j=1}^{\infty} \frac{v_j(0)^2}{\mu_j + z} = \frac{\tanh(Y\sqrt{z})}{\sqrt{z}}.$$

Indeed, this is the function $\rho(z)$ for the case $s = 1/2$. Since $\tanh x \to 1$ exponentially fast as $x \to \infty$, we have that pointwise $\rho(z) \to z^{-1/2}$ as $Y \to \infty$ for $z > 0$, and the convergence is uniform if we consider only $z \geq \lambda_{\min}^h > 0$.

In the general case, $\rho(z)$ has a more complicated form, which we now derive.

Theorem 4. *For $s \in (0,1)$ and $z > 0$, we have*

$$\rho(z) = \frac{2^{2s-1} z^{-s} \Gamma(s)^2 I_s(\sqrt{z}Y)}{2K_s(\sqrt{z}Y) + I_s(\sqrt{z}Y)\Gamma(1-s)\Gamma(s)},$$

where I_s and K_s are the modified Bessel functions of first and second kind, respectively, and Γ is the gamma function.

Proof. The problem (18) can be equivalently rewritten as

$$-v''(y) - \frac{\alpha}{y} v'(y) + zv(y) = 0 \quad \forall y \in (0, Y), \qquad \lim_{y \to 0^+} (y^\alpha v'(y)) = -1, \quad v(Y) = 0.$$

Two linearly independent solutions of the problem

$$v''(y) + \frac{\alpha}{y} v'(y) = zv(y)$$

are given by

$$v_1(y) = (\sqrt{z}y)^s I_s(\sqrt{z}y), \qquad v_2(y) = (\sqrt{z}y)^s K_s(\sqrt{z}y).$$

The relevant properties of the Bessel functions that we use in the following can be found in standard references such as [22] Using classical relations of the Bessel functions, we find that

$$y^{1-2s} v_1'(y) = y^{1-s} z^{\frac{s+1}{2}} I_{s-1}(\sqrt{z}y), \qquad y^{1-2s} v_2'(y) = -y^{1-s} z^{\frac{s+1}{2}} K_{1-s}(\sqrt{z}y),$$

and then using the limiting behavior of the Bessel functions at 0 that

$$\lim_{y \to 0^+} y^\alpha v_1'(y) = 2 \left(\frac{z}{2}\right)^s / \Gamma(s), \qquad \lim_{y \to 0^+} y^\alpha v_2'(y) = -\left(\frac{z}{2}\right)^s \Gamma(1-s).$$

Making the ansatz $v(y) = c_1 v_1(y) + c_2 v_2(y)$ and solving for the boundary conditions, we obtain the constants

$$c_1 = -\left(\frac{2}{z}\right)^s \frac{K_s(\sqrt{z}Y)\Gamma(s)}{2K_s(\sqrt{z}Y) + I_s(\sqrt{z}Y)\Gamma(1-s)\Gamma(s)}, \quad c_2 = \left(\frac{2}{z}\right)^s \frac{I_s(\sqrt{z}Y)\Gamma(s)}{2K_s(\sqrt{z}Y) + I_s(\sqrt{z}Y)\Gamma(1-s)\Gamma(s)}.$$

Since

$$\lim_{y \to 0^+} v_1(y) = 0, \qquad \lim_{y \to 0^+} v_2(y) = \Gamma(s) 2^{s-1},$$

the Dirichlet value at 0 is given by

$$v(0) = c_2 \Gamma(s) 2^{s-1} = \frac{2^{2s-1} z^{-s} \Gamma(s)^2 I_s(\sqrt{z}Y)}{2K_s(\sqrt{z}Y) + I_s(\sqrt{z}Y)\Gamma(1-s)\Gamma(s)}. \qquad \square$$

Using the asymptotics of I_s and K_s at infinity (in particular, that $K_s(z) \to 0$ exponentially fast as $z \to \infty$), we find that

$$\lim_{Y \to \infty} \rho(z) = \frac{2^{2s-1} z^{-s} \Gamma(s)^2}{\Gamma(1-s)\Gamma(s)} = \frac{z^{-s}}{d_s}.$$

Somewhat more precisely, using the bounds $K_s(x) \le C\sqrt{\pi/(2x)}\exp(-x)$ and $I_s(x) \ge c\exp(x)/\sqrt{2\pi x}$ for sufficiently large arguments from [22], we obtain the bound

$$\left| \frac{z^{-s}}{d_s} - \rho(z) \right| \le C_s \frac{\exp(-2\sqrt{z}Y)}{\sqrt{z}Y} \le C_s \frac{\exp(-2\sqrt{\lambda_{\min}^h}Y)}{\sqrt{\lambda_{\min}^h}Y}$$

for Y sufficiently large. Here we assumed that $z \geq \lambda_{\min}^h$, which is the relevant case for our estimates.

Let now v denote the exact solution of (18) and $v_h \in W_h$ the discrete solution from (17). From Theorem 3 and the definition of ρ, we have $\rho(z) - \rho_h(z) = v(0) - v_h(0)$. We are thus interested in the discretization error of the functional which evaluates v at 0. The standard approach to obtain error estimates for such functionals is via the dual problem, i.e., we solve a version of (17) where the right-hand side is the functional of interest. In our particular case, (17) already has this right-hand side, and therefore the solution of the dual problem is just v_h itself. Denoting the involved bilinear form by

$$b(v,w) := \int_0^Y (y^\alpha v'w' + zy^\alpha vw)dy,$$

we obtain, using Galerkin orthogonality, that

$$v(0) - v_h(0) = b(v,v) - b(v_h,v_h) = b(v - v_h, v) = b(v - v_h, v - v_h) = \|v - v_h\|_b^2,$$

where $\|v\|_b^2 = b(v,v)$ is the energy norm associated with this one-dimensional problem. Since v_h is the b-orthogonal projection of v into W_h, it follows further that

$$\rho(z) - \rho_h(z) = v(0) - v_h(0) = \inf_{w_h \in W_h} \|v_z - w_h\|_b^2, \tag{19}$$

where we wrote $v = v_z$ to remind ourselves that v depends on z. In other words, $\rho_h(z)$ converges to $\rho(z)$ with double the rate of best approximation to v_z in W_h (in the energy norm). This suggests that higher-order discretizations in the extended dimension provide a particularly significant improvement in convergence, and spectral methods, such as proposed in [9], are very attractive. For the case $s = 1/2$, we obtain by standard estimates that, if we use splines of degree p with maximum continuity C^{p-1} as the basis functions $(\psi_j)_{j=1}^m$, we have the rate $|\rho(z) - \rho_h(z)| = \mathcal{O}(m^{-2p})$.

Recalling that, for the extension method, $r(z) = d_s\rho_h(z)$, the relevant error term in Theorem 1 can now be bounded as

$$\begin{aligned}\|(\cdot)^{-s} - r(\cdot)\|_{L_\infty(\lambda_{\min}^h, \lambda_{\max}^h)} &\leq \|(\cdot)^{-s} - d_s\rho(\cdot)\|_{L_\infty(\lambda_{\min}^h, \lambda_{\max}^h)} + d_s\|\rho(\cdot) - \rho_h(\cdot)\|_{L_\infty(\lambda_{\min}^h, \lambda_{\max}^h)} \\ &\leq C_s \left(\frac{\exp(-2\sqrt{\lambda_{\min}^h}Y)}{\sqrt{\lambda_{\min}^h}Y} + \sup_{z \in [\lambda_{\min}^h, \lambda_{\max}^h]} \inf_{w_h \in W_h} \|v_z - w_h\|_b^2, \right) =: E_{\mathrm{EXM}}\end{aligned}$$

for Y sufficiently large. If we then denote by $u \in V_h$ the approximate solution to problem (1) obtained by the discrete eigenfunction method and by $u_{\mathrm{EXM}} \in V_h$ the one obtained by the extension method, Theorem 1 yields

$$\|u - u_{\mathrm{EXM}}\|_{L_2(\Omega)} \leq E_{\mathrm{EXM}}\|f\|_{L_2(\Omega)}.$$

The exponential error term depending on Y should be balanced with the term stemming from the approximation of v_z in W_h. We do not carry out these details here as they depend on the particular choice of the discretization space W_h. Results for W_h being a piecewise linear spline space can be found in [7], whereas a spectral method in the extended direction has been analyzed in [9].

Remark 2. (a) The error estimate we arrive at using the analysis as a rational approximation method is in the $L_2(\Omega)$-norm, whereas in both [7] and [9], the error is estimated in a fractional Sobolev space, namely, in a $H^s(\Omega)$-like norm.

(b) Our analysis fully decouples the error in the extended direction from the error in the spatial variable and thus can be used to obtain new error bounds for different spaces W_h without the need to analyze the properties of the full tensor product discretization space.

7.4. Numerical stabilization

Depending on the discretization space W_h and in particular for values of s close to 1, meaning α close to -1, the eigenvalue problem (14) may become extremely poorly conditioned, leading to a loss of accuracy in the numerical realization. In order to mitigate this, we may introduce a small stabilization parameter $\sigma > 0$ and instead solve the modified eigenvalue problem

$$K_y^{(\alpha)} V = (M_y^{(\alpha)} + \sigma K_y^{(\alpha)}) V \widetilde{\Lambda}_y,$$

which clearly has the same eigenvectors V as (14). Due to

$$K_y^{(\alpha)} V = (M_y^{(\alpha)} + \sigma K_y^{(\alpha)}) V \widetilde{\Lambda}_y = M V (\widetilde{\Lambda}_y + \sigma \Lambda_y \widetilde{\Lambda}_y),$$

we obtain $\Lambda_y = \widetilde{\Lambda}_y + \sigma \Lambda_y \widetilde{\Lambda}_y$ and finally

$$\Lambda_y = (1 - \sigma \widetilde{\Lambda}_y)^{-1} \widetilde{\Lambda}_y,$$

which allows us to recover the original eigenvalues from those of the stabilized problem. We note, however, that also the simple approximation

$$\Lambda_y \approx \widetilde{\Lambda}_y$$

seems to produce good results.

8. The time stepping method of Vabishchevich

Vabishchevich [16] has proposed a reformulation of the fractional diffusion problem (1) as a parabolic problem which is then solved by a time stepping scheme. He chooses a parameter $\delta > 0$ such that $\mathcal{L} \geq \delta I$ and observes that the solution of (1) is given by $w(1)$, where w solves the parabolic initial value problem

$$(t(\mathcal{L} - \delta I) + \delta I) \frac{dw}{dt} + s(\mathcal{L} - \delta I) w = 0 \quad \forall t \in (0,1), \qquad w(0) = \delta^{-s} f.$$

After semidiscretization in space and introducing the matrix $D = K - \delta M$, this problem reads

$$(tD + \delta M) \mathbf{w}' + sD\mathbf{w} = 0 \quad \forall t \in (0,1), \qquad \mathbf{w}(0) = \delta^{-s} \mathbf{f},$$

where $\mathbf{w} : [0,1] \to \mathbb{R}^n$. An implicit time stepping scheme over an equidistant time grid and with a parameter $\theta \in (0,1]$ is then introduced. We use, for $k \in \{0, \ldots, m\}$, the notations

$$t^k = \tau k, \qquad \tau = \frac{1}{m}, \qquad t^{\theta(k)} := \theta t^{k+1} + (1-\theta) t^k = t^k + \theta\tau, \qquad \mathbf{w}^{\theta(k)} := \theta \mathbf{w}^{k+1} + (1-\theta)\mathbf{w}^k$$

and introduce the implicit scheme

$$(t^{\theta(k)} D + \delta M) \frac{\mathbf{w}^{k+1} - \mathbf{w}^k}{\tau} + sD\mathbf{w}^{\theta(k)} = 0 \qquad \forall k = 0, \ldots, m-1.$$

By rearranging, we find the time stepping rule

$$\left[\left(\frac{t^{\theta(k)}}{\tau} + s\theta \right) D + \frac{\delta}{\tau} M \right] \mathbf{w}_{k+1} = \left[\left(\frac{t^{\theta(k)}}{\tau} - s(1-\theta) \right) D + \frac{\delta}{\tau} M \right] \mathbf{w}_k \qquad \forall k = 0, \ldots, m-1.$$

The approximate solution to (1) is then given by the coefficient vector $\mathbf{w}_m$.

We can now again show that the above scheme is equivalent to a rational approximation method. To this end, we again make use of the spectral decomposition from Section 2,

$$D = U^{-T}(\Lambda - \delta I_n) U^{-1}, \qquad M = U^{-T} U^{-1},$$

and obtain that, for all k,

$$U^{-T}\left[\left(\frac{t^{\theta(k)}}{\tau}+s\theta\right)(\Lambda-\delta I_n)+\frac{\delta}{\tau}I_n\right]U^{-1}\mathbf{w}_{k+1}=U^{-T}\left[\left(\frac{t^{\theta(k)}}{\tau}-s(1-\theta)\right)(\Lambda-\delta I_n)+\frac{\delta}{\tau}I_n\right]U^{-1}\mathbf{w}_k.$$

Introducing the rational functions

$$\omega_k(z):=\frac{\left(\frac{t^{\theta(k)}}{\tau}-s(1-\theta)\right)(z-\delta)+\frac{\delta}{\tau}}{\left(\frac{t^{\theta(k)}}{\tau}+s\theta\right)(z-\delta)+\frac{\delta}{\tau}},\qquad k=0,\dots,m-1$$

and writing $\mathbf{w}_k$ in the spectral basis, $\mathbf{q}_k:=U^{-1}\mathbf{w}_k$, we obtain

$$\mathbf{q}_{k+1}=\omega_k(\Lambda)\mathbf{q}_k.$$

Since $\mathbf{q}_0=\delta^{-s}U^{-1}\mathbf{f}$, the final solution is given by

$$\mathbf{u}=\mathbf{w}_m=U\mathbf{q}_m=Ur(\Lambda)U^{-1}\mathbf{f},$$

the form of a rational approximation method, with the rational function

$$r(z)=\delta^{-s}\prod_{k=0}^{m-1}\omega_k(z)$$

with degrees (m,m). The rational function $r(z)$ can again be brought into the form (8) by partial fraction decomposition. From the definition of ω_k, we can easily determine that the roots of $r(z)$ are given by

$$\delta\left(1-\frac{1}{\tau(k+\theta(1+s))}\right),\qquad k=0,\dots,m-1,$$

which is nonpositive provided that $\theta(1+s)\le 1$. Since $s<1$, the Crank-Nicolson method ($\theta=0.5$), which is the most accurate method in the proposed class of implicit schemes, guarantees that the roots of $r(z)$ are negative. Thus the method of Vabishchevich can equivalently be realized as a rational approximation method as in Section 3. Indeed, such a realization has the advantage of being fully parallel, whereas the original time stepping method is by necessity sequential.

We note that the above discussion could be extended to the higher-order time stepping method introduced in [17].

9. Numerical study

In this section, we perform a numerical comparison of the methods described in the previous sections. To this end, we discretize, for $0<s<1$, the problem

$$-\left(\frac{d^2}{dx^2}\right)^s u(x)=1\quad\forall x\in(-1,1),\qquad u(-1)=u(1)=0 \tag{20}$$

using continuous and piecewise linear finite element functions over a uniform grid with $2^{10}=1024=2h^{-1}$ intervals. After eliminating the Dirichlet degrees of freedom, this results in a problem with $\dim V_h=1023$ unknowns, and the eigenvalues of $M^{-1}K$ satisfy $\lambda^h_{\min}\approx 9.87$, $\lambda^h_{\max}\approx 1.26\cdot 10^7$. It has been shown in [14]) that the discretization error for this problem behaves like $\mathcal{O}(h^{\min\{2,2s+0.5\}})$.

Considering only a one-dimensional problem may seem too limited, however Theorem 1 shows that the performance of rational approximation methods depends essentially only on the spectrum of the discretized problem. Since the condition number of the discrete problem depends essentially only on the mesh size h and not the dimension (namely, $\kappa\sim h^{-2}$), the following results may be considered representative also for 2D and 3D problems with a similar mesh size $h\approx 10^{-3}$.

We then implement all methods described above as rational approximation methods. The parameters were chosen as follows.

- For the BURA method, we always choose $\beta = 1$ and rescale using the exact eigenvalue $\lambda^h_{\max}$. The rational approximations were computed using the `minimax` function from `chebfun` [21].
- For the extension method, we always choose the cutoff at $Y = 3$. We then introduce a graded mesh with a parameter $\gamma \geq 1$ as suggested in [7], namely,

$$y_\ell = Y \left(\frac{\ell}{m}\right)^\gamma, \qquad \ell = 0, \dots, m,$$

  and build a continuous spline space over this mesh. We try both the standard piecewise linear functions suggested in [7], as well as cubic splines with maximum continuity ($\psi_j \in C^2(0, Y)$) using a B-spline basis, which is a novel approach. We choose $\gamma = 12$ for $s = 0.25$ and $\gamma = 5$ for $s \geq 0.5$, except for $\gamma = 4.5$ for linear splines in the case $s = 0.75$. Finally, we always choose the stabilization parameter introduced in Section 7.4 as $\sigma = 10^{-12}$ with the inexact recovery formula.
- For the quadrature method of Bonito and Pasciak, we use the balanced parameter choice rule (11) suggested in [14]. This means that only $q > 0$ is a free parameter and directly determines the number of terms.
- For Vabishchevich's method, we choose the lower bound $\delta = \lambda^h_{\min}$ and always use the parameter $\theta = 0.5$ corresponding to the Crank-Nicolson scheme.
- For the novel AAA approximation method described in Section 5, we sample z^{-s} as described there with $M = 10^6$ points over the exact interval $[\lambda^h_{\min}, \lambda^h_{\max}]$.

The resulting errors for all methods are shown in Figure 1 for $s = 0.25$, Figure 2 for $s = 0.5$, and Figure 3 for $s = 0.75$. In all figures, the horizontal axis displays the number k of terms in the rational approximation of the form (8), or equivalently, the number of system solves for the stiffness matrix K to be performed when realizing the method. As all methods can be realized as rational approximation methods, this quantity allows a direct performance comparison between the methods. For each example, the upper plot displays the spectral error $\max_{z \in [\lambda^h_{\min}, \lambda^h_{\max}]} |z^{-s} - r(z)|$ with $r(z)$ being the rational function associated with the corresponding method, whereas the lower plot shows the actual error $\|\widetilde{u} - u_r\|_{L_2(-1,1)}$, where $\widetilde{u}$ is the true solution of (20) and u_r is the solution obtained by the respective method. The methods are labeled as **BP3** for the third quadrature method of Bonito and Pasciak [14], **BURA** for the BURA method [11], **EXM** for the extension method [7] with linear splines, **EXM(3)** for the variant of the extension method with cubic splines, **AAA** for the rational approximation method using the AAA algorithm (Section 5), and **Vab** for the original time stepping method of Vabishchevich [16].

For the spectral error, the dashed line labeled "best" shows the error of the best uniform rational approximation with degrees (k, k) of z^{-s} over $[\lambda^h_{\min}, \lambda^h_{\max}]$ and is thus the best achievable error for a rational approximation method in the sense of Theorem 1 and a lower bound for all other methods (compare also Table 1 and the discussion thereof). Indeed, if one obtains numerical bounds for the smallest and largest eigenvalues, one can in principle compute this rational approximation and use it for a numerical method. This approach performs significantly better than all other studied methods, as the plots show, but does suffer from the numerical difficulties of computing best rational approximations as described in Section 4 for the BURA method.

For the L_2-error, the dashed horizontal line labeled "DEM" shows the error achieved by the discrete eigenfunction method described in Section 2. Due to the known estimates (cf. [14]), this error is quasi-optimal with respect to the approximation error and indeed is not improved upon by any method.

The used software implementations in Python of all methods may be found on the author's homepage, a link to which is placed at the ORCID profile[3]. An interpretation and discussion of the numerical results follows below.

[3] `https://orcid.org/0000-0002-6616-5081`

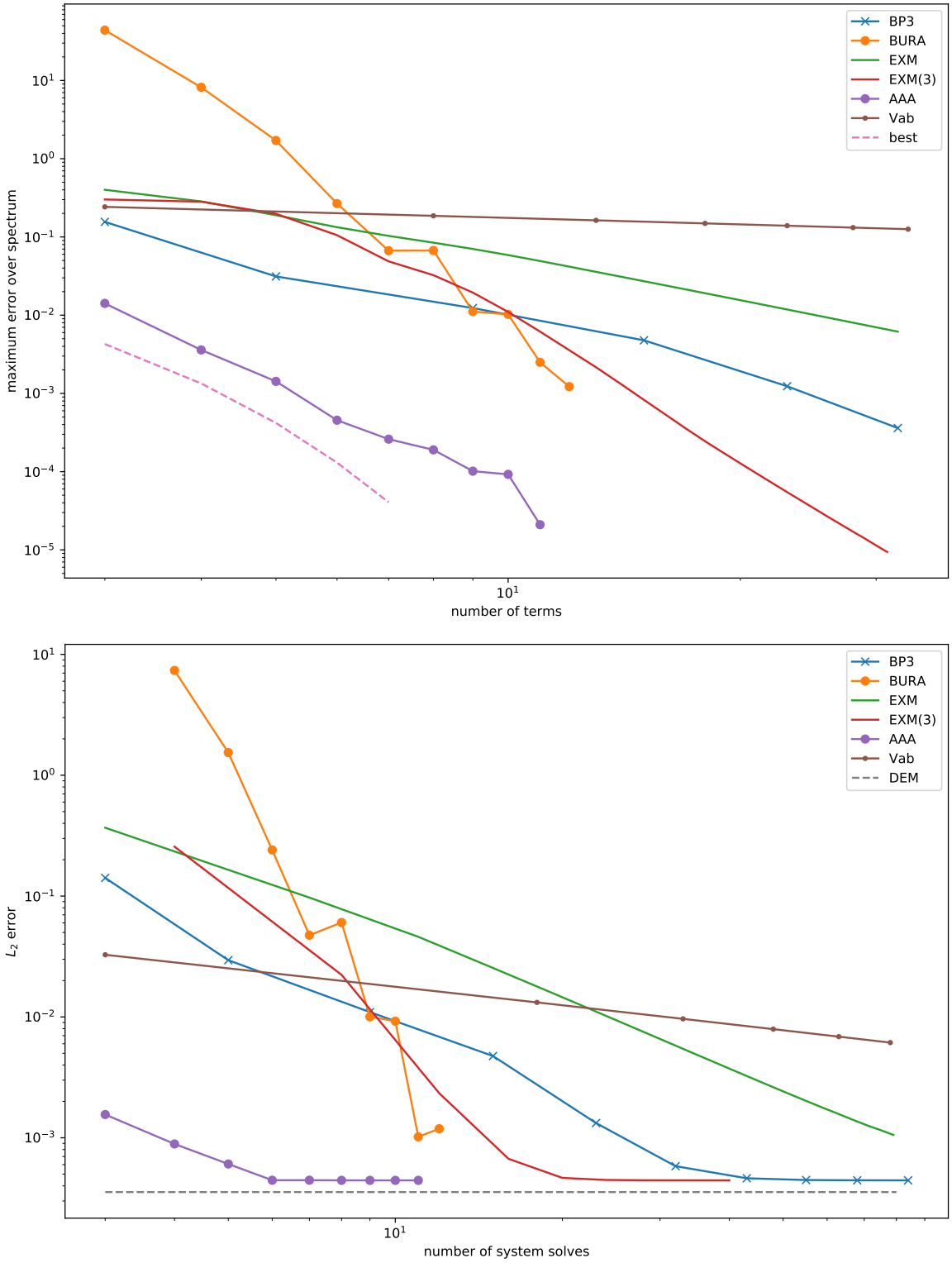

Figure 1: Errors in the case $s = 0.25$. *Top:* spectral error $\max_{z\in[\lambda^h_{\min},\lambda^h_{\max}]} |z^{-s} - r(z)|$. *Bottom:* L_2-errors $\|\widetilde{u} - u_r\|_{L_2(0,1)}$, where $\widetilde{u}$ is the true solution of the fractional Laplace problem and u_r is the solution obtained by the respective method.

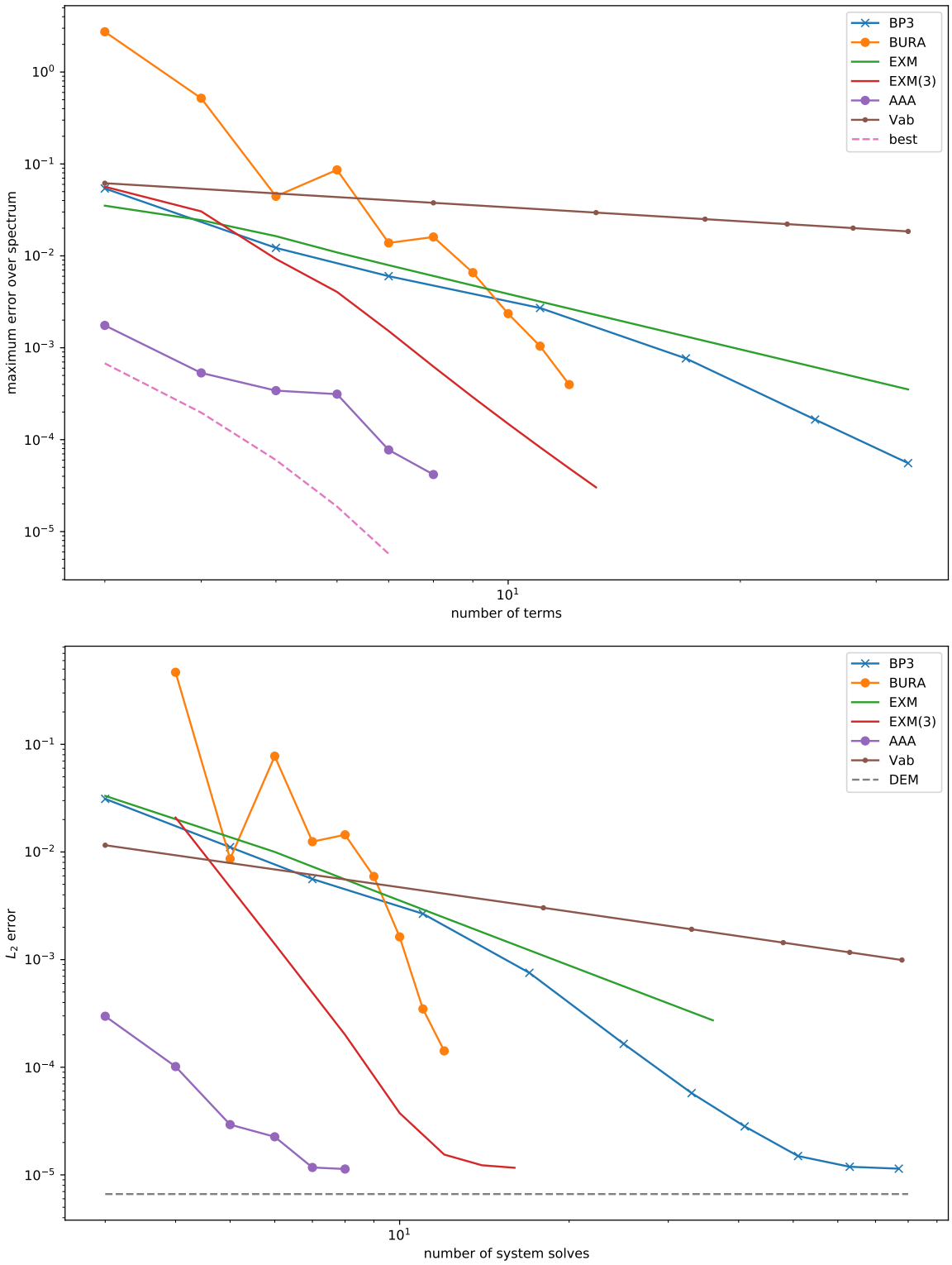

Figure 2: Errors in the case $s = 0.5$. *Top:* spectral error $\max_{z\in[\lambda^h_{\min},\lambda^h_{\max}]} |z^{-s} - r(z)|$. *Bottom:* L_2-errors $\|\widetilde{u} - u_r\|_{L_2(0,1)}$, where $\widetilde{u}$ is the true solution of the fractional Laplace problem and u_r is the solution obtained by the respective method.

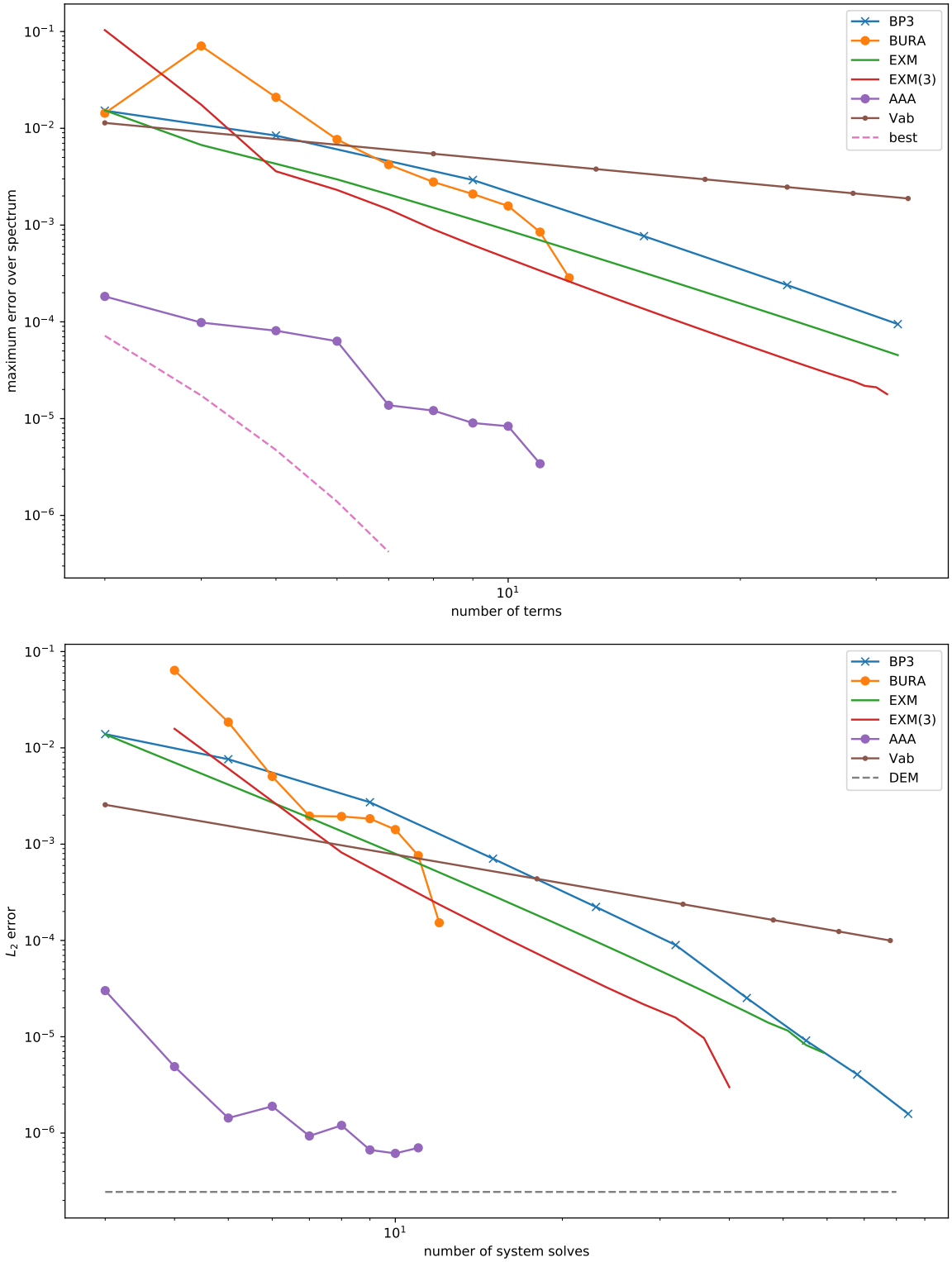

Figure 3: Errors in the case $s = 0.75$. *Top:* spectral error $\max_{z\in[\lambda_{\min}^h,\lambda_{\max}^h]} |z^{-s} - r(z)|$. *Bottom:* L_2-errors $\|\widetilde{u} - u_r\|_{L_2(0,1)}$, where $\widetilde{u}$ is the true solution of the fractional Laplace problem and u_r is the solution obtained by the respective method.

- In all cases, the spectral error $\max_z |z^{-s} - r(z)|$ (top plot) is a reasonable predictor for the actual performance of the method in terms of the L_2-error to the true solution (bottom plot), as predicted by Theorem 1. Nevertheless, there are some differences between both errors, as the plots show. One reason for this is that the spectral error weights the error uniformly over the entire interval containing the spectrum, whereas the true spectrum is a discrete set of points in this interval.

  Perhaps more importantly, the spectral coefficients of a function tend to decay for larger eigenvalues due to smoothness considerations. For this reason, the spectral error close to $\lambda_{\min}^h$ tends to be more important than that close to $\lambda_{\max}^h$. For instance, the spectral error for the BURA method is dominant close to the minimum eigenvalue and therefore the method does worse in practice, whereas the time stepping method of Vabishchevich concentrates its spectral error towards the largest eigenvalue and thus does better in practice relative to its spectral error estimate. One could analyze this effect by weighting the spectral error estimates, but this requires additional knowledge on the smoothness of f, i.e., the decay of $U^{-1}\mathbf{f}$ (cf. the proof of Theorem 1).

- Both the BURA method and the quadrature method of Bonito and Pasciak converge exponentially, but with quite different rates.

- As predicted by (10), the BURA method has larger initial error (i.e., for a small number of terms) for smaller values of s due to the stronger dependence on the condition number.

- Using standard double precision arithmetic, the best rational approximations of z^{1-s} for the BURA method could not be computed for high enough rational degree in order to resolve the spatial error, i.e., reach the error floor of the DEM. The computations were done using the `minimax` function of the `chebfun` package [21] using double precision. Still, the degrees for which `minimax` converged are higher than the reported limits for the modified Remez algorithm in [11, Table 5]. In fact, `minimax` using double precision appears to converge for degrees at least as high, and often higher, than the modified Remez algorithm from [11] using quadruple precision. It appears that `chebfun` has no provisions for computing with extended precision.

- For the extension method, some numerical estimates for the convergence rate in terms of the number $k = \dim W_h$ of system solves are given in the table below. We distinguish the case with linear splines studied in [7] as well as the novel variant with cubic, maximally smooth splines.

s	linear	cubic
0.25	k^{-2}	k^{-6}
0.5	k^{-2}	k^{-6}
0.75	$k^{-2.5}$	k^{-3}

  The rates k^{-2} and k^{-6} in the case $s = 0.5$ follow from (19) and standard estimates for spline approximation. A careful analysis of spline approximation in weighted Sobolev spaces and using graded meshes for (19) is required in order to understand the results for different choices of s.

  Overall, the advantage of cubic over linear splines is very significant in the cases $s = 0.25$ and $s = 0.5$, but rather mild for $s = 0.75$.

- The time stepping method of Vabishchevich [16] converges for $s = 0.25$ with the rate $\sim k^{-0.5}$, for $s = 0.5$ with the rate $\sim k^{-0.75}$, and for $s = 0.75$ with the rate $\sim k^{-1}$. These rates should be expected to improve when using the higher-order time stepping scheme proposed in [17].

- The rational approximation method based on the AAA algorithm has lower convergence rate than the BURA method, but avoids the strong dependence on the condition number. Overall, it seems to be the most efficient method of all the tested ones (except for the direct best approximation labeled "best", which may suffer from numerical instabilities for larger numbers of terms) by a significant margin for all choices of s.

Acknowledgments

This work was partially supported by the National Research Network "Geometry + Simulation" (NFN S117), funded by the Austrian Science Fund (FWF).

References

[1] C. Bucur, E. Valdinoci, Nonlocal Diffusion and Applications, Springer International Publishing, 2016 (2016). doi:10.1007/978-3-319-28739-3.

[2] A. Lischke, G. Pang, M. Gulian, F. Song, C. Glusa, X. Zheng, Z. Mao, W. Cai, M. M. Meerschaert, M. Ainsworth, G. E. Karniadakis, What is the fractional Laplacian?, arXiv e-prints (2018). arXiv:1801.09767.

[3] L. Caffarelli, L. Silvestre, An extension problem related to the fractional Laplacian, Communications in Partial Differential Equations 32 (8) (2007) 1245–1260 (2007). doi:10.1080/03605300600987306.

[4] C. Brändle, E. Colorado, A. de Pablo, U. Sánchez, A concave-convex elliptic problem involving the fractional Laplacian, Proceedings of the Royal Society of Edinburgh: Section A Mathematics 143 (01) (2013) 39–71 (2013). doi:10.1017/s0308210511000175.

[5] A. Capella, J. Dávila, L. Dupaigne, Y. Sire, Regularity of radial extremal solutions for some non-local semilinear equations, Communications in Partial Differential Equations 36 (8) (2011) 1353–1384 (2011). doi:10.1080/03605302.2011.562954.

[6] P. R. Stinga, J. L. Torrea, Extension problem and harnack's inequality for some fractional operators, Communications in Partial Differential Equations 35 (11) (2010) 2092–2122 (2010). doi:10.1080/03605301003735680.

[7] R. H. Nochetto, E. Otárola, A. J. Salgado, A PDE approach to fractional diffusion in general domains: A priori error analysis, Foundations of Computational Mathematics 15 (3) (2015) 733–791 (2015). doi:10.1007/s10208-014-9208-x.

[8] L. Chen, R. H. Nochetto, E. Otárola, A. J. Salgado, Multilevel methods for nonuniformly elliptic operators and fractional diffusion, Mathematics of Computation 85 (302) (2016) 2583–2607 (2016). doi:10.1090/mcom/3089.

[9] M. Ainsworth, C. Glusa, Hybrid finite element–spectral method for the fractional Laplacian: Approximation theory and efficient solver, SIAM Journal on Scientific Computing 40 (4) (2018) A2383–A2405 (2018). doi:10.1137/17m1144696.

[10] L. Banjai, J. M. Melenk, R. H. Nochetto, E. Otárola, A. J. Salgado, C. Schwab, Tensor FEM for spectral fractional diffusion, Foundations of Computational Mathematics (2018). doi:10.1007/s10208-018-9402-3.

[11] S. Harizanov, R. Lazarov, S. Margenov, P. Marinov, Y. Vutov, Optimal solvers for linear systems with fractional powers of sparse SPD matrices, Numerical Linear Algebra with Applications 25 (5) (2018) e2167 (2018). doi:10.1002/nla.2167.

[12] S. Harizanov, R. Lazarov, P. Marinov, S. Margenov, J. Pasciak, Comparison analysis on two numerical methods for fractional diffusion problems based on rational approximations of t^γ, $0 \leq t \leq 1$, arXiv e-prints (2018). arXiv:1805.00711.

[13] D. Bolin, K. Kirchner, The rational SPDE approach for Gaussian random fields with general smoothness (2017). arXiv:1711.04333.

[14] A. Bonito, J. E. Pasciak, Numerical approximation of fractional powers of elliptic operators, Mathematics of Computation 84 (295) (2015) 2083–2110 (2015). doi:10.1090/s0025-5718-2015-02937-8.

[15] D. Bolin, K. Kirchner, M. Kovács, Weak convergence of galerkin approximations for fractional elliptic stochastic PDEs with spatial white noise, BIT Numerical Mathematics.

[16] P. N. Vabishchevich, Numerically solving an equation for fractional powers of elliptic operators, Journal of Computational Physics 282 (2015) 289–302 (2015). doi:10.1016/j.jcp.2014.11.022.

[17] P. N. Vabishchevich, Numerical solving unsteady space-fractional problems with the square root of an elliptic operator, Mathematical Modelling and Analysis 21 (2) (2016) 220–238 (2016). doi:10.3846/13926292.2016.1147000.

[18] Y. Nakatsukasa, O. Sète, L. N. Trefethen, The AAA algorithm for rational approximation, SIAM Journal on Scientific Computing 40 (3) (2018) A1494–A1522 (2018). doi:10.1137/16m1106122.

[19] A. Bonito, J. P. Borthagaray, R. H. Nochetto, E. Otárola, A. J. Salgado, Numerical methods for fractional diffusion, Computing and Visualization in Science 19 (5-6) (2018) 19–46 (2018). doi:10.1007/s00791-018-0289-y.

[20] M. D. Ortigueira, J. A. T. Machado, What is a fractional derivative?, Journal of Computational Physics 293 (2015) 4–13 (2015). doi:10.1016/j.jcp.2014.07.019.

[21] T. A. Driscoll, N. Hale, L. N. Trefethen, Chebfun Guide, Pafnuty Publications, 2014 (2014).
URL http://www.chebfun.org/docs/guide/

[22] NIST digital library of mathematical functions, release 1.0.21 of 2018-12-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds.
URL http://dlmf.nist.gov/