IMA Journal of Numerical Analysis (2020) **40**, 1746–1771
doi:10.1093/imanum/drz013
Advance Access publication on 28 March 2019

Numerical approximation of fractional powers of elliptic operators

Beiping Duan
School of Mathematics and Statistics, Central South University, 410083 Changsha, P.R. China and Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, USA

Raytcho D. Lazarov*
Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, USA and Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Block 8, 1113 Sofia, Bulgaria
*Corresponding author: lazarov@math.tamu.edu

and

Joseph E. Pasciak
Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, USA

[Received on 09 March 2018; revised on 8 February 2019]

In this paper, we develop and study algorithms for approximately solving linear algebraic systems: $\mathcal{A}_h^\alpha u_h = f_h$, $0 < \alpha < 1$, for $u_h, f_h \in V_h$ with V_h a finite element approximation space. Such problems arise in finite element or finite difference approximations of the problem $\mathcal{A}^\alpha u = f$ with $\mathcal{A}$, for example, coming from a second-order elliptic operator with homogeneous boundary conditions. The algorithms are motivated by the method of Vabishchevich (2015, Numerically solving an equation for fractional powers of elliptic operators. *J. Comput. Phys.*, **282**, 289–302) that relates the algebraic problem to a solution of a time-dependent initial value problem on the interval [0, 1]. Here we develop and study two time-stepping schemes based on diagonal Padé approximation to $(1+x)^{-\alpha}$. The first one uses geometrically graded meshes in order to compensate for the singular behaviour of the solution for t close to 0. The second algorithm uses uniform time stepping, but requires smoothness of the data f_h in discrete norms. For both methods, we estimate the error in terms of the number of time steps, with the regularity of f_h playing a major role for the second method. Finally, we present numerical experiments for $\mathcal{A}_h$ coming from the finite element approximations of second-order elliptic boundary value problems in one and two spatial dimensions.

Keywords: fractional powers of elliptic operators; finite element approximation; Padé; approximation; solution methods for equations involving powers of SPD matrices.

1. Introduction

1.1 *Motivation and problem formulation*

Nonlocal operators arise in a wide variety of mathematical models such as modes of long-range interaction in elastic deformations (Silling, 2000), nonlocal electromagnetic fluid flows (McCay & Narasimhan, 1981), image processing (Gilboa & Osher, 2008; Pu *et al.*, 2010) and many more. A recent discussion about the properties of such models and their applications to chemistry, geosciences and engineering can be found in (Kilbas *et al.*, 2006; Metzler *et al.*, 2014).

The nonlocal operators considered in this paper involve fractional powers of operators $\mathcal{A}$ associated with second-order elliptic equations in bounded domains with homogeneous Dirichlet boundary conditions. The fractional power of $\mathcal{A}$ is defined through the Dunford–Taylor integral, (Kato, 1961; Lunardi, 2007), which is equivalent to the definition by the spectrum of $\mathcal{A}$. For a detailed discussion about this setting and other possible ways to define fractional powers of the Laplacian (and more general elliptic operators) we refer to (Kwaśnicki, 2017, Bonito *et al.*, 2018; Lischke *et al.*, 2018). We focus on the issue of solving the corresponding algebraic system that arises in approximating such operators by the finite element method, e.g., Bonito & Pasciak (2015), Bonito *et al.* (2018) and Lei (2018).

We begin with the definition of the fractional power of a second-order elliptic operator in a bounded domain $\Omega \subset \mathbb{R}^d$, $d = 1, 2, \ldots$ with a Lipschitz continuous boundary. On $V \times V$, with $V = H_0^1(\Omega)$, we consider the bilinear form

$$A(w, \phi) = \int_\Omega \Big(a(x) \nabla w \cdot \nabla \phi + q(x) w \phi \Big) \, \mathrm{d}x, \tag{1.1}$$

under the assumption $a(x), q(x) \in L^\infty(\Omega)$ and $a(x) \geqslant \bar{a}, q(x) \geqslant 0$ with $\bar{a}$ a positive constant, so that the bilinear form is coercive and continuous on V. Further, we define $\mathcal{T} : L^2(\Omega) \to V$, where $\mathcal{T} v = w \in V$ is the unique solution to

$$A(w, \phi) = (v, \phi), \qquad \phi \in V. \tag{1.2}$$

Here $(\cdot, \cdot)$ denotes the $L^2(\Omega)$-inner product.

Following Kato (1961), we define an unbounded operator $\mathcal{A}$ with domain of definition $D(\mathcal{A})$ being the image of $\mathcal{T}$ on $L^2(\Omega)$ and set $\mathcal{A}u = \mathcal{T}^{-1}u$ for $u \in D(\mathcal{A})$. This is well defined as $\mathcal{T}$ is injective on $L^2(\Omega)$. Negative fractional powers can be defined by Dunford–Taylor integrals, i.e., for $\alpha > 0$ and $v \in L^2(\Omega)$,

$$\mathcal{A}^{-\alpha} v = \frac{1}{2\pi i} \int_{\mathcal{C}} z^{-\alpha} R_z(\mathcal{A}) v \, \mathrm{d}z,$$

where $R_z(\mathcal{A}) = (\mathcal{A} - z\mathcal{I})^{-1}$ is the resolvent operator and $\mathcal{C}$ is an appropriate contour in the complex plane (see, e.g., Lunardi, 2007).

Equivalently, fractional powers for the above example can be defined by eigenvector expansions. As $\mathcal{T}$ is a compact, symmetric and positive definite operator on $L^2(\Omega)$, its eigenpairs $\{\psi_j, \mu_j\}$, for $j = 1, 2, \ldots, \infty$, with suitably normalized eigenvectors, provide an orthonormal basis for $L^2(\Omega)$. We also set $\lambda_j = \mu_j^{-1}$. For $\alpha \geqslant 0$ and $v \in L^2(\Omega)$,

$$\mathcal{A}^{-\alpha} v = \sum_{j=1}^{\infty} \mu_j^\alpha (v, \psi_j) \, \psi_j.$$

Positive fractional powers of $\mathcal{A}$ are also given by similar series. For $\alpha \geqslant 0$, define

$$D(\mathcal{A}^\alpha) := \left\{ v \in L^2(\Omega) \;\; \sum_{j=1}^{\infty} \lambda_j^{2\alpha} |(v, \psi_j)|^2 < \infty \right\}$$

and

$$\mathcal{A}^\alpha v := \sum_{j=1}^{\infty} \lambda_j^\alpha (v, \psi_j) \, \psi_j \;\text{ for } v \in D(\mathcal{A}^\alpha).$$

We consider the fractional order elliptic equation. Find $u \in D(\mathcal{A}^{\alpha})$ satisfying

$$\mathcal{A}^{\alpha} u = f \ \text{ for } f \in L^2(\Omega) \tag{1.3}$$

and note that its solution is given by

$$u = \mathcal{T}^{\alpha} f := \sum_{j=1}^{\infty} \mu_j^{\alpha} (f, \psi_j)\, \psi_j.$$

REMARK 1.1 Note that more realistic for some applications is the subdiffusion–reaction problem $\mathcal{A}^{\alpha} u + qu = f$, where the operator $\mathcal{A}$ is introduced through the form (1.1) with $q(x) \equiv 0$. However, the proposal in this paper method has no obviously extension to such problems.

Our goal is to approximate u by using finite element or finite differences. The finite element approximation on $V_h \subset V$ is based on the discrete solution operator $\mathcal{T}_h : V_h \to V_h$ defined by $\mathcal{T}_h v_h := w_h$, where w_h is the unique function in V_h satisfying

$$A(w_h, \phi_h) = (v_h, \phi_h), \qquad \text{for all } \phi_h \in V_h.$$

The inverse of $\mathcal{T}_h$ is denoted by $\mathcal{A}_h$ and satisfies

$$(\mathcal{A}_h w_h, \phi_h) = A(w_h, \phi_h), \qquad \text{for all } \phi_h \in V_h.$$

We obtain a 'semidiscrete' approximation to u of equation (1.3) by defining

$$u_h = \mathcal{T}_h^{\alpha} \pi_h f := \mathcal{A}_h^{-\alpha} \pi_h f, \tag{1.4}$$

where π_h is the $L^2(\Omega)$-orthogonal projection into V_h. Note that u_h can be expanded in the $L^2(\Omega)$-orthogonal eigenfunctions of $\mathcal{T}_h$, i.e., if $\{\psi_{h,j}, \mu_{h,j}\}$, for $j = 1, \ldots, M$, denotes the eigenpairs and M is the dimension of V_h then

$$\pi_h f = \sum_{j=1}^{M} (f, \psi_{h,j})\, \psi_{h,j}$$

and

$$u_h = \sum_{j=1}^{M} \mu_{h,j}^{\alpha} (f, \psi_{h,j})\, \psi_{h,j}. \tag{1.5}$$

In this paper, we shall study a technique for approximating the solution to (1.4) that avoids computing the eigenvectors and eigenvalues of $\mathcal{T}_h$.

We note that the technique to be developed can be applied to finite difference approximations as well. In this case, the discrete space is a finite-dimensional space of grid point values and $\pi_h f$ is replaced by the interpolant of f at the grid nodes. The matrix $\mathcal{T}_h$ comes from applying finite difference approximations to the derivatives in the strong form (see (4.1)) of problem (1.2) with proper scaling.

It has been shown in Bonito & Pasciak (2015, Theorem 4.3) that, if the operator $\mathcal{T}$ satisfies elliptic regularity pickup with index $s \in (0, 1]$ (see Assumption 4.2) then for appropriate f,

$$\|u - u_h\| = \left\|\mathcal{T}^{\alpha} f - \mathcal{T}_h^{\alpha} \pi_h f\right\| = O(h^{2s}).$$

See Section 4 for more details. Here $\|\cdot\|$ denotes the $L^2(\Omega)$-norm.

Obviously, the discrete operator $\mathcal{A}_h$ is symmetric and positive definite and the corresponding matrix is full. We note that problems involving finding approximations of $\mathcal{A}_h^{1/2}$, cf. Higham (1997), evaluating the sign function of $\mathcal{A}_h$, cf. Kenney & Laub (1991), and other related functions of matrices have a long history in numerical linear algebra.

1.2 *The idea of the method of Vabishchevich*

In this paper, we develop and study a method for approximating the solution (1.4). Our proposed method is related to an idea of Vabishchevich (2015) which exhibits u_h as a solution of a special time-dependent problem. Now we briefly explain the idea of his method.

We start with the observation that the unique solution $\hat{u}(t)$ to the ordinary differential equation initial value problem,

$$\hat{u}_t + \frac{\alpha(\lambda - \delta)}{\delta + t(\lambda - \delta)}\hat{u} = 0, \quad \hat{u}(0) = \delta^{-\alpha}\hat{v} \tag{1.6}$$

with $0 < \delta < \lambda$ and $\alpha \geqslant 0$, is given by

$$\hat{u}(t) = (\delta + t(\lambda - \delta))^{-\alpha}\hat{v} \tag{1.7}$$

and hence

$$\hat{u}(1) = \lambda^{-\alpha}\hat{v}. \tag{1.8}$$

Note also that for $k > 0$,

$$\hat{u}(t + k) = \left[1 + \frac{k(\lambda - \delta)}{\delta + t(\lambda - \delta)}\right]^{-\alpha} \hat{u}(t). \tag{1.9}$$

Now suppose that the spectrum of $\mathcal{A}_h$ is contained in the interval $[\lambda_1, \lambda_M]$ with $\lambda_1 > 0$. Such a bound could be obtained using the coercivity of the form $a(\cdot, \cdot)$ in V and the Poincaré–Friedrichs inequality. Also, we can get a sufficiently good estimate for λ_1 by running a few iterations of the inverse power method. In fact, we need to know only the growth rate of λ_M, e.g., λ_M grows like a constant times h^{-2} for the geometric refinement case.

Let $0 < \delta \leqslant \lambda_1$ and set $\mathcal{B} = \mathcal{A}_h - \delta\mathcal{I}$. We consider the vector-valued ODE. Find $U(t) : [0, 1] \to V_h$ satisfying

$$\begin{aligned} U_t(t) + \alpha\mathcal{B}(\delta\mathcal{I} + t\mathcal{B})^{-1}U(t) &= 0, \\ U(0) &= \delta^{-\alpha} f_h, \end{aligned} \tag{1.10}$$

where $f_h := \pi_h f$. Expanding the solution to (1.10) as

$$U(t) = \sum_{j=1}^{M} c_j(t)\psi_{h,j},$$

we find that $c_j(t)$ solves (1.6) with $\lambda = \lambda_{h,j}$. Moreover, it follows from (1.7) and (1.9) that

$$U(t) := (\delta\mathcal{I} + t\mathcal{B})^{-\alpha} f_h \tag{1.11}$$

and

$$U(t+k) = (\mathcal{I} + k\mathcal{B}(\delta\mathcal{I} + t\mathcal{B})^{-1})^{-\alpha} U(t). \tag{1.12}$$

As proposed by Vabishchevich (2015), it is then natural to consider numerical approximations to (1.10) based on a time-stepping method.

In Vabishchevich (2015), Vabishchevich proposed a time stepping scheme based on the backward Euler method and applied it to approximate fractional powers of a discrete approximation $\mathcal{A}_h$ of the Laplace operator with homogeneous Dirichlet boundary conditions. The results of numerical computations illustrating the accuracy, convergence and some theoretical aspects of the method were provided.

In this paper, we take a different but related approach. Instead of approximating the solution of (1.10), we simply approximate the function $U(t_i)$ given by (1.11) on an increasing sequence of nodes $0 = t_0 < t_1 < \cdots < t_K = 1$. We start from the recurrence (1.12). The 'time stepping' methods that we shall study are based on diagonal Padé approximation to $(1+x)^{-\alpha}$, i.e.,

$$(1+x)^{-\alpha} \approx r_m(x) := \frac{P_m(x)}{Q_m(x)} \tag{1.13}$$

with P_m and Q_m being polynomials of degree m and $Q_m(0) = 1$. The polynomials P_m and Q_m are uniquely defined by requiring the first $2m+1$ terms of the Maclaurin expansion of

$$(1+x)^{-\alpha} - r_m(x)$$

vanish. The method that we study and analyse is given by setting $U_0 := u(0) = \delta^{-\alpha} f_h$ and applying the recurrence

$$U_l = r_m\left(k_l\mathcal{B}(\delta\mathcal{I} + t_{l-1}\mathcal{B})^{-1}\right) U_{l-1}, \quad l = 1, 2, \ldots, K \tag{1.14}$$

with $k_l = t_l - t_{l-1}$. Here U_l is our approximation of $U(t_l)$ so that U_K approximates $U(1) = u_h$. We shall see that these methods are unconditionally stable for $m = 1, 2, \ldots$ and $\alpha \in (0, 1)$.

Even though we take a different point of view, we are still solving (1.10), as suggested by Vabishchevich (2015). It is important to note that, although problem (1.10) appears harmless, it behaves considerably different than, for example, the classical parabolic problem

$$w_t + \mathcal{A}_h w = 0. \tag{1.15}$$

For example, if $w(0) = \psi_{h,j}$ then $w(1) = e^{-\lambda_{h,j}}\psi_{h,j}$ while, if $f_h = \psi_{h,j}$, the solution of (1.10) is $u(1) = \mathcal{A}_h^{-\alpha}\psi_{h,j} = \lambda_{h,j}^{-\alpha}\psi_{h,j}$. This means that initial time-step errors in the high-frequency components for our problem have much stronger effect on the accuracy of the final solution. This is especially important for problems whose solutions have minimal regularity.

1.3 *Our contributions*

In this paper, we consider two time-stepping schemes, one involving mesh refinement near $t = 0$ and the other using a fixed time step. In both cases, we shall be using (1.14) to define our solution, but on different meshes in time.

The refinement scheme starts with an initial basic mesh with $t_0 = 0$, $t_i = 2^{i-1-L}$, for $i = 1, 2, \ldots L+1$ with L chosen so that $2^{-L} < (\lambda_M)^{-1}$. Subsequent finer meshes are defined by partitioning each of the above intervals into N equally spaced subdivisions. The refinement scheme leads to an error estimate

$$\left\| \mathcal{A}_h^{-\alpha} f_h - U_{N(L+1)} \right\| \leqslant C N^{-2m} \| f_h \|$$

for the Padé scheme based on $r_m(x)$.

The second scheme that we study is the simpler one using a fixed step size $k_N = 1/N$. In this case, we obtain the error estimate

$$\left\| \mathcal{A}_h^{-\alpha} f_h - U_N \right\| \leqslant C k_N^{\alpha+\gamma} \left\| A_h^{\gamma} f_h \right\|, \quad \text{for } 0 \leqslant \gamma \leqslant 2m - \alpha. \tag{1.16}$$

It is clear that L in the first method grows like the logarithm of λ_M so that more steps are required by the refinement scheme when the same N is used in both. However, in all of our numerical examples, if one adjusts the values of N in both schemes to obtain the same absolute convergence, the refinement scheme requires less steps overall.

The question of when the norm on the right-hand side of (1.16) can be controlled by natural norms on the data f has been only partially addressed in the literature. Although, such a result for $\gamma \leqslant 1$ was provided in (Lei, 2018), the result for larger γ is not known even in the finite element case. We discuss this in more detail in Section 4. In fact, our numerical results in Section 5 suggest that the result is not true in general.

2. Padé approximations

In this section, we develop diagonal Padé approximations to $(1 + x)^{-\alpha}$ for $\alpha \in (0, 1)$ based on the classical theory of Padé approximations given by Baker (1975).

Our approximations are of the form of (1.13) with $m = 1, 2, \ldots$ and we shall write down explicit formula for the polynomials $P_m(x)$ and $Q_m(x)$. The starting point is the formula (Baker, 1975, relation (5.2)) or (Kenney & Laub, 1989, formula (2.1))

$$(1 + x)^{-\alpha} = {}_2F_1(\alpha, 1; 1; -x).$$

Here ${}_2F_1(a, b; c; x)$ denotes the hyper-geometric function defined by

$${}_2F_1(a, b; c; x) = \sum_{j=0}^{\infty} \frac{(a)_j (b)_j}{j!\,(c)_j} x^j.$$

Here $(a)_0 = 1$ and $(a)_j = a(a + 1) \cdots (a + j - 1)$ for $j > 0$. This series converges for $|x| < 1$ provided that c is not in $\{0, -1, -2, \ldots\}$.

Baker (1975, Formula (5.12)), also gives an explicit expression for the denominator

$$\begin{aligned} Q_m(x) &= {}_2F_1(-m, -\alpha - m; -2m; -x) \\ &:= \sum_{j=0}^{m} \frac{(-m)_j\,(-\alpha - m)_j}{j!\ (-2m)_j}(-x)^j = 1 + \sum_{j=1}^{m} a_j b_j(\alpha) x^j. \end{aligned} \tag{2.1}$$

Here $b_0(\alpha) = a_0 = 1$,

$$b_j(\alpha) = (m+\alpha)((m-1)+\alpha)\cdots((m+1-j)+\alpha)$$

and

$$a_j = \frac{m(m-1)\cdots(m+1-j)}{j!\,(2m(2m-1)\cdots(2m+1-j))} \qquad \text{for } j = 1, 2, \ldots, m.$$

Now, Theorem 9.2 of Baker (1975) implies that $Q_m(x)/P_m(x)$ is the diagonal Padé approximation to $(1+x)^\alpha$ and again applying (5.12) of Baker (1975), we find that

$$\begin{aligned} P_m(x) &= {}_2F_1(-m, \alpha - m; -2m; -x) \\ &:= \sum_{j=0}^{m} \frac{(-m)_j(\alpha - m)_j}{j!\,(-2m)_j}(-x)^j = 1 + \sum_{j=1}^{m} a_j b_j(-\alpha) x^j. \end{aligned} \tag{2.2}$$

Using the above formulas, we find, for example,

$$r_1(x) = \frac{1 + [(1-\alpha)/2]x}{1 + [(1+\alpha)/2]x}$$

and

$$r_2(x) = \frac{1 + [(2-\alpha)/2]x + [(2-\alpha)(1-\alpha)/12]x^2}{1 + [(2+\alpha)/2]x + [(2+\alpha)(1+\alpha)/12]x^2}.$$

For our further considerations, we need to discuss a relation between the denominators appearing in Padé approximations and orthogonal polynomials with respect to an appropriate weight w. We first note the series expansion of

$$(1-x)^{-\alpha} = \sum_{i=0}^{\infty} \frac{(\alpha)_i\, x^i}{i!} := \sum_{i=0}^{\infty} c_i x^i.$$

The coefficients above satisfy

$$c_i = \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} \int_0^1 x^i \big((1-x)^{-\alpha} x^{\alpha-1}\big)\ \mathrm{d}x := \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} \int_0^1 x^i w(x)\ \mathrm{d}x, \tag{2.3}$$

where $w(x) = (1-x)^{-\alpha} x^{\alpha-1}$. By (7.7) of Baker (1975), utilizing the fact that $P_m(-x)/Q_m(-x)$ is the Padé approximation of $(1-x)^{-\alpha}$, we obtain that the denominator $Q_m(x)$ can be expressed by

$$Q_m(x) = (-x)^m q_m(-1/x),$$

where q_m is the monic polynomial of order m that is orthogonal to the set of polynomials of degree less than m with the weight $xw(x)$. As the roots of $q_m(x)$ are all distinct and in the interval $(0, 1)$, those of $Q_m(x)$ are distinct and in the interval $(-\infty, -1)$.

The expressions for the numerator and denominator in $r_m(x)$ imply the following proposition.

PROPOSITION 2.1 Let α be in $(0, 1)$ and m be a positive integer. Then, there are positive constants ρ_m satisfying

$$\rho_m \leqslant r_m(x) \leqslant 1, \qquad \text{for all } x \geqslant 0. \tag{2.4}$$

Proof. As α is in $(0, 1)$,

$$0 < b_j(-\alpha) < b_j(\alpha).$$

This means for $x \geqslant 0$, $a_j b_j(-\alpha)x^j \leqslant a_j b_j(\alpha)x^j$ with strict inequality when $x > 0$. The second inequality of (2.4) follows by summation.

For the first inequality in (2.4), we note that both $Q_m(x)$ and $P_m(x)$ are positive for $x \geqslant 0$ and

$$\lim_{x\to\infty} r_m(x) = b_m(-\alpha)/b_m(\alpha) > 0.$$

The first inequality in (2.4) immediately follows from the fact that $r_m(x)$ is continuous on $[-1, \infty)$ since all of the roots of $Q_m(x)$ are in $(-\infty, -1)$ and takes values in $(0, 1]$ for $x \geqslant 0$. $\square$

To clarify further the convergence of these approximations, we include the following proposition.

PROPOSITION 2.2 For $0 \leqslant s \leqslant 2m + 1$,

$$\left|(1 + x)^{-\alpha} - r_m(x)\right| \leqslant c_{m,s}x^s \quad \text{for } x \in [0, \infty), \tag{2.5}$$

where

$$c_{m,s} = \max\left\{Q_m(-1)2^{s-2m}, 2^{1+s}\right\}.$$

Proof. We use Theorem 5 of Kenney & Laub (1989) that gives that

$$\begin{aligned}(1 + x)^{-\alpha} - r_m(x) &= \frac{Q_m(-1)}{Q_m(x)} \sum_{n=2m+1}^{\infty} \frac{(\alpha)_n (n - 2m)_m}{n!\,(n + \alpha - m)_m}(-x)^n \\ &:= \frac{Q_m(-1)}{Q_m(x)}(-x)^{2m+1} \sum_{i=0}^{\infty} d_{2m+1+i}(-x)^i.\end{aligned}$$

A simple computation shows that $0 < d_n < 1$ for $n \geqslant 2m + 1$ and so for $|x| < \tau < 1$,

$$\left|\sum_{i=0}^{\infty} d_{2m+1+i}(-x)^i\right| \leqslant (1 - \tau)^{-1}.$$

Thus, for $x \in [0, \tau]$ and $s \in [0, 2m+1]$,

$$\begin{aligned} \left| r_m(x) - (1+x)^{-\alpha} \right| &\leqslant |Q_m(-1)|(1-\tau)^{-1} x^{2m+1} \\ &\leqslant |Q_m(-1)| \tau^{2m+1-s} (1-\tau)^{-1} x^s. \end{aligned}$$

Finally, for the case $x \in [\tau, \infty)$ by Proposition 2.1,

$$|r_m(x) - (1+x)^{-\alpha}| \leqslant 2 \leqslant 2x^s/\tau^s, \quad \text{for } x \in [\tau, \infty)$$

and (2.5) follows by taking $\tau = 1/2$. □

3. The time-stepping schemes and their analysis

In this section, we define and analyse both equally spaced time-stepping schemes as well as schemes employing refinement near the origin. We shall restrict ourselves to approximating the solution to finite-dimensional problem (1.4) described in the introduction, even though generalizations to Hermitian and nonHermitian bounded operators on infinite-dimensional spaces are possible. Recall that $\mathcal{B} = \mathcal{A}_h - \delta \mathcal{I}$ with $\delta \in (0, \lambda_1]$ and that $\|\cdot\|$ and $(\cdot, \cdot)$ denote, the norm and inner product in $L^2(\Omega)$, respectively.

3.1 *Time-stepping method on geometrically refined meshes*

We first consider the geometrically refined mesh (GRM) that is constructed in two steps.

First, we take

$$L = \lceil \log(\lambda_M) / \log(2) \rceil$$

and set $t_i = 2^{i-1-L}$ for $i = 1, \ldots, L+1$ and $t_0 = 0$. Note that $t_{L+1} = 1$ and $2^{-L} \leqslant (\lambda_M)^{-1}$. Next, we define

$$k_n = \begin{cases} t_n/N: & \text{for } n = 1, \ldots, L, \\ t_1/N: & \text{when } n = 0. \end{cases} \tag{3.1}$$

Note that $k_0 = k_1$ and

$$k_0 \lambda_M = \frac{2^{-L} \lambda_M}{N} \leqslant N^{-1}. \tag{3.2}$$

The grid that we use in our computations is obtained by partitioning each subinterval $I_n = [t_n, t_{n+1}]$, $n = 0, 1, \ldots, L$, into N subintervals with end points

$$t_{n,j} := t_n + j k_n \quad \text{for } j = 0, \ldots, N.$$

It can be seen that $t_{n,0} = t_n$ and $t_{n,N} = t_{n+1}$, $n = 0, 1, \ldots, L$, so that the mesh has totally $(L+1)N$ intervals.

Based on this partitioning, the refined time stepping method for approximating $\mathcal{A}_h^{-\alpha} v$ for $v \in V_h$ is given by the following.

Algorithm 3.1

(a) Set $U_0 = \delta^{-\alpha} v$.

(b) For $n = 0, 1, \ldots, L$:

(i) Set $U_{n,0} = U_n$.

(ii) For $j = 1, 2, \ldots, N$, set

$$U_{n,j} = r_m(k_n \mathcal{B}(\delta \mathcal{I} + t_{n,j-1}\mathcal{B})^{-1}) U_{n,j-1}.$$

(iii) Set $U_{n+1} = U_{n,N}$.

After executing the above algorithm, U_{L+1} is the approximation to $\mathcal{A}_h^{-\alpha} v$. Note that the notation differs slightly from that used in the introduction. The computation of U_{L+1} requires $K = (L+1)N$ time steps.

REMARK 3.1 Since all the roots of $Q_m(x)$ are real, distinct and in the interval $(-\infty, -1)$ we have $Q_m(x) = \prod_{i=1}^{m}(1 + \xi_i x)$, $\xi_i > 0$. Then r_m can be presented in terms of partial fractions

$$r_m(x) = \beta_0 + \sum_{i=1}^{m} \frac{\beta_i}{1 + \xi_i x}. \tag{3.3}$$

Then the evaluation of $r_m(\mathcal{C}_{n,j})U_{n,j-1}$, with $\mathcal{C}_{n,j} = k_n(\delta \mathcal{B}^{-1} + t_{n,j-1}\mathcal{I})^{-1}$ could be implemented by using (3.3) that will result in solving concurrently m systems

$$(\mathcal{I} + \xi_i \mathcal{C}_{n,j})V_i = U_{n,j-1}, \quad i = 1, \ldots, m \ \text{ so that } U_{n,j} = \beta_0 U_{n,j-1} + \sum_{i=1}^{m} \beta_i V_i.$$

The discrete eigenvalues and eigenvectors will play a major role in our analysis so, for notational simplicity, we denote them by $\{\psi_j, \lambda_j\}$ (instead of $\{\psi_{h,j}, \lambda_{h,j}\}$ as in the introduction). We then have

$$v = \sum_{j=1}^{M} (v, \psi_j)\psi_j.$$

Moreover,

$$\mathcal{A}_h^{-\alpha} v = \sum_{j=1}^{M} \lambda_j^{-\alpha}(v, \psi_j)\psi_j.$$

The expansion for U_{L+1} is given by

$$U_{L+1} = \sum_{j=1}^{M} \mu(\lambda_j)(v, \psi_j)\psi_j,$$

where the coefficient $\mu(\lambda_j)$ is given by the following algorithm.

Algorithm 3.2

(a) Set $\mu_0 = \delta^{-\alpha}$.

(b) For $n = 0, 1, \ldots, L$:

(i) Set $\mu_{n,0} = \mu_n$.

(ii) For $j = 1, 2, \ldots, N$, set

$$\mu_{n,j} = r_m(k_n(\lambda - \delta)/(\delta + t_{n,j-1}(\lambda - \delta)))\mu_{n,j-1}.$$

(iii) Set $\mu_{n+1} = \mu_{n,N}$.

(c) Set $\mu(\lambda) = \mu_{L+1}$.

THEOREM 3.2 Let N be a positive integer. Then

$$|\lambda^{-\alpha} - \mu(\lambda)| \leqslant \widetilde{c}\max\{\delta^{-2m-\alpha-1}, \delta^{-\alpha}\}N^{-2m}, \qquad \text{for all } \lambda \in [\lambda_1, \lambda_M] \tag{3.4}$$

and

$$\|\mathcal{A}_h^{-\alpha}v - U_{L+1}\| \leqslant \widetilde{c}\max\{\delta^{-2m-\alpha-1}, \delta^{-\alpha}\}N^{-2m}\|v\|, \qquad \text{for all } v \in V_h. \tag{3.5}$$

Here $\widetilde{c}$ is a constant depending only on m and α.

Proof. Fix λ in $[\lambda_1, \lambda_M]$ and let $\{\mu_n, \mu_{n,j}\}$ be as in Algorithm 3.2. Further, for $n = 0, \ldots, L$ and $j = 1, \ldots, N$ let

$$v_{n,j} = (\delta + t_{n,j}(\lambda - \delta))^{-\alpha}, \;\; e_{n,j} = v_{n,j} - \mu_{n,j}, \;\; \text{and} \;\; \theta_{n,j} = k_n(\lambda - \delta)/(\delta + t_{n,j-1}(\lambda - \delta)),$$

with $v_n := v_{n,0}, e_n := e_{n,0}$ and $e_{L+1} := v_{L,N} - \mu_{L,N}$. We note that as in (1.12),

$$v_{n,j} = (1 + \theta_{n,j})^{-\alpha}v_{n,j-1}.$$

Thus,

$$e_{n,j} = r_m(\theta_{n,j})e_{n,j-1} + \Big[(1 + \theta_{n,j})^{-\alpha} - r_m(\theta_{n,j})\Big]v_{n,j-1}$$

and by Proposition 2.1 and the triangle inequality,

$$|e_{n,j}| \leqslant |e_{n,j-1}| + \Big|(1 + \theta_{n,j})^{-\alpha} - r_m(\theta_{n,j})\Big|\;|v_{n,j-1}|, \tag{3.6}$$

for $n = 0, 1, \ldots, L$ and $j = 1, \ldots, N$.

Using $e_0 = 0$, (3.6), Proposition 2.2 and the definition of $v_{n,j-1}$, we have

$$
\begin{aligned}
|\lambda^{-\alpha} - \mu(\lambda)| = |e_{L+1}| &= \sum_{n=0}^{L} \sum_{j=1}^{N} \Big(|e_{n,j}| - |e_{n,j-1}| \Big) \\
&\leqslant c_{m,2m+1} \sum_{n=0}^{L} \sum_{j=1}^{N} \theta_{n,j}^{2m+1} |v_{n,j-1}| \\
&= c_{m,2m+1} \sum_{n=0}^{L} \sum_{j=1}^{N} \frac{(k_n(\lambda-\delta))^{2m+1}}{(\delta + t_{n,j-1}(\lambda-\delta))^{2m+\alpha+1}}. \qquad (3.7)
\end{aligned}
$$

We note that by (3.2),

$$
\sum_{j=1}^{N} \frac{(k_0(\lambda-\delta))^{2m+1}}{(\delta + t_{0,j-1}(\lambda-\delta))^{2m+\alpha+1}} \leqslant \delta^{-2m-\alpha-1} \sum_{j=1}^{N} N^{-2m-1} = \delta^{-2m-\alpha-1} N^{-2m}. \qquad (3.8)
$$

The remaining terms in (3.7) will be bounded by integration. Applying mean value theorem for integration it follows that for some $t_\theta \in [t_{n,j-1}, t_{n,j}]$,

$$
\begin{aligned}
&\int_{t_{n,j-1}}^{t_{n,j}} \frac{t^{2m}(\lambda-\delta)^{2m}}{(\delta + t(\lambda-\delta))^{2m+\alpha+1}} \,\mathrm{d}t = \frac{t_\theta^{2m}(\lambda-\delta)^{2m}}{(\delta + t_\theta(\lambda-\delta))^{2m+\alpha+1}} k_n \\
&= \left(\frac{t_\theta}{t_{n,j-1}} \right)^{2m} \left(\frac{\delta + t_{n,j-1}(\lambda-\delta)}{\delta + t_\theta(\lambda-\delta)} \right)^{2m+\alpha+1} \frac{t_{n,j-1}^{2m}(\lambda-\delta)^{2m}}{(\delta + t_{n,j-1}(\lambda-\delta))^{2m+\alpha+1}} k_n \\
&\geqslant \left(\frac{t_{n,j-1}}{t_\theta} \right)^{\alpha+1} \frac{t_{n,j-1}^{2m}(\lambda-\delta)^{2m}}{(\delta + t_{n,j-1}(\lambda-\delta))^{2m+\alpha+1}} k_n \\
&\geqslant \left(\frac{N}{N+1} \right)^{\alpha+1} \frac{t_{n,j-1}^{2m}(\lambda-\delta)^{2m}}{(\delta + t_{n,j-1}(\lambda-\delta))^{2m+\alpha+1}} k_n.
\end{aligned}
$$

Now $(N+1)/N \leqslant 2$ and $t_n \leqslant t_{n,j-1}$ for $j = 1, 2, \ldots, N$ so that

$$
\begin{aligned}
\frac{(k_n(\lambda-\delta))^{2m+1}}{(\delta + t_{n,j-1}(\lambda-\delta))^{2m+\alpha+1}} &= N^{-2m} \frac{t_n^{2m} k_n (\lambda-\delta)^{2m+1}}{(\delta + t_{n,j-1}(\lambda-\delta))^{2m+\alpha+1}} \\
&\leqslant 2^{\alpha+1} N^{-2m} (\lambda-\delta) \int_{t_{n,j-1}}^{t_{n,j}} \frac{t^{2m}(\lambda-\delta)^{2m}}{(\delta + t(\lambda-\delta))^{2m+\alpha+1}} \,\mathrm{d}t. \qquad (3.9)
\end{aligned}
$$

Applying (3.9) to the remaining terms in (3.7) with (3.8), we have

$$
\begin{aligned}
&\sum_{n=0}^{L}\sum_{j=1}^{N}\Big(|e_{n,j}|-|e_{n,j-1}|\Big)\\
&\leqslant \frac{c_{m,2m+1}}{\delta^{2m+\alpha+1}}N^{-2m}+c_{m,2m+1}2^{\alpha+1}N^{-2m}(\lambda-\delta)\int_0^1\frac{t^{2m}(\lambda-\delta)^{2m}}{(\delta+t(\lambda-\delta))^{2m+\alpha+1}}\,\mathrm{d}t\\
&\leqslant \frac{c_{m,2m+1}}{\delta^{2m+\alpha+1}}N^{-2m}+c_{m,2m+1}2^{\alpha+1}N^{-2m}\int_0^\infty\frac{z^{2m}}{(\delta+z)^{2m+\alpha+1}}\,\mathrm{d}z
\end{aligned}
\tag{3.10}
$$

with the last integral on the right satisfying

$$
\begin{aligned}
\int_0^\infty\frac{z^{2m}}{(\delta+z)^{2m+\alpha+1}}\,\mathrm{d}z&=\int_0^\delta\frac{z^{2m}}{(\delta+z)^{2m+\alpha+1}}\,\mathrm{d}z+\int_\delta^\infty\frac{z^{2m}}{(\delta+z)^{2m+\alpha+1}}\,\mathrm{d}z\\
&\leqslant\int_0^\delta\frac{\delta^{2m}}{\delta^{2m+\alpha+1}}\,\mathrm{d}z+\int_\delta^\infty\frac{z^{2m}}{z^{2m+\alpha+1}}\,\mathrm{d}z=\left(1+\frac{1}{\alpha}\right)\delta^{-\alpha}.
\end{aligned}
$$

Then (3.4) follows by combining (3.7) and (3.10).

The inequality (3.5) follows immediately from (3.4) and the Parseval's identities

$$
\begin{aligned}
\left\|\mathcal{A}_h^{-\alpha}v-U_{L+1}\right\|^2&=\sum_{j=1}^{M}\left|\lambda_j^{-\alpha}-\mu(\lambda_j)\right|^2|(v,\psi_j)|^2\\
&\le\widetilde{c}^2(\max\{\delta^{-2m-\alpha-1},\delta^{-\alpha}\})^2N^{-4m}\sum_{j=1}^{M}|(v,\psi_j)|^2\\
&=\widetilde{c}^2(\max\{\delta^{-2m-\alpha-1},\delta^{-\alpha}\})^2N^{-4m}\|v\|^2.
\end{aligned}
$$

□

REMARK 3.3 For any $s\in\mathbb{R}$ and $v\in V_h$, let

$$
\|v\|_{s,h}:=\left(\sum_{j=1}^{M}\lambda_j^{s}|(v,\psi_j)|^2\right)^{1/2}=\left\|\mathcal{A}_h^{s/2}v\right\|.
\tag{3.11}
$$

Then, it follows from the proof of the above theorem that for any $s\in\mathbb{R}$,

$$
\left\|\mathcal{A}_h^{-\alpha}v-U_{L+1}\right\|_{s,h}\leqslant\widetilde{c}N^{-2m}\|v\|_{s,h}.
$$

3.2 *Time-stepping method on uniform meshes*

We next consider uniform time stepping. In this case, given a positive integer N, we set $k_N = 1/N$ and $t_n = nk_N$. The approximation is obtained from the recurrence

$$\begin{aligned} U_0 &= \delta^{-\alpha} v \in V_h, \\ U_n &= r_m\left(k_N \mathcal{B}(\delta\mathcal{I} + t_{n-1}\mathcal{B})^{-1}\right) U_{n-1}, \quad \text{for } n = 1, 2, \ldots, N. \end{aligned} \tag{3.12}$$

In this case, U_N is our approximation to $\mathcal{A}_h^{-\alpha} v$. The analysis of the error requires the following proposition.

PROPOSITION 3.4 For $\lambda \geqslant \delta$, set

$$\theta_n = \frac{k_N(\lambda - \delta)}{\delta + t_{n-1}(\lambda - \delta)}. \tag{3.13}$$

Then for $q \geqslant p > 1$,

$$\prod_{n=p}^{q} r_m(\theta_n) \leqslant c \prod_{n=p}^{q} (1 + \theta_n)^{-\alpha}, \tag{3.14}$$

with c depending only on m and α.

Proof. We note that for $n > 1$, $\theta_n \leqslant (n-1)^{-1} \in [0, 1]$. By Proposition 2.2,

$$\begin{aligned} r_m(\theta_n) &\leqslant (1 + \theta_n)^{-\alpha} + c_{m,2m+1} |\theta_n|^{2m+1} \\ &\leqslant (1 + \theta_n)^{-\alpha} \left(1 + 2^\alpha c_{m,2m+1} (n-1)^{-2m-1}\right) \end{aligned}$$

and hence

$$\frac{r_m(\theta_n)}{(1 + \theta_n)^{-\alpha}} \leqslant \left(1 + 2^\alpha c_{m,2m+1} (n-1)^{-2m-1}\right).$$

Thus, for $q \geqslant p > 1$,

$$\prod_{n=p}^{q} \frac{r_m(\theta_n)}{(1 + \theta_n)^{-\alpha}} \leqslant \prod_{n=p}^{q} \left(1 + 2^\alpha c_{m,2m+1} (n-1)^{-2m-1}\right) \leqslant c$$

with

$$c = \prod_{j=1}^{\infty} \left(1 + 2^\alpha c_{m,2m+1} j^{-2m-1}\right).$$

□

Theorem 7.2 of Thomée (2006) provides error estimates for single-step approximations for the standard parabolic problem (1.15) with nonsmooth initial data. The next theorem has the same flavour, however, differs significantly as the solutions of our problem exhibit less regularity.

THEOREM 3.5 Let $N > 1$, $v \in V_h$ and U_N be defined by (3.12). Then, for $\gamma \geqslant 0$ and $\alpha + \gamma \leqslant 2m$,

$$\left\| \mathcal{A}_h^{-\alpha} v - U_N \right\| \leqslant \frac{c}{\delta^{\alpha+\gamma}} k_N^{\alpha+\gamma} \left\| \mathcal{A}_h^{\gamma} v \right\| \tag{3.15}$$

with c depending only on α, m and γ. As in Remark 3.3, the left-hand norm above can be replaced by $\| \cdot \|_{r,h}$ provided that the right is replaced by $\| \mathcal{A}_h^{\gamma} v \|_{r,h}$.

Proof. In this proof, c denotes a generic positive constant only depending on α, γ, m and δ. We fix $\lambda \geqslant \delta$ and define, for $j \geqslant l \geqslant 1$,

$$r_m^{j,l}(\lambda) := r_m^{j,l} = r_m(\theta_j)\, r_m(\theta_{j-1}) \cdots r_m(\theta_l)$$

and

$$w^{j,l}(\lambda) := w^{j,l} = (1+\theta_j)^{-\alpha} (1+\theta_{j-1})^{-\alpha} \cdots (1+\theta_l)^{-\alpha}.$$

Finally, we set

$$e_N(\lambda) := e_N = \left(w^{N,1} - r_m^{N,1} \right) \delta^{-\alpha}.$$

We note that it is a consequence of (1.8) and (1.9) that

$$\delta^{-\alpha} w^{N,1} = \lambda^{-\alpha}. \tag{3.16}$$

For any $j > 1$,

$$w^{j,1} - r_m^{j,1} = \left[(1+\theta_j)^{-\alpha} - r_m(\theta_j) \right] w^{j-1,1} + r_m(\theta_j) \left[w^{j-1,1} - r_m^{j-1,1} \right].$$

Repeated application of this identity leads to

$$\begin{aligned} e_N &= \delta^{-\alpha} \left[w^{N,1} - r_m^{N,1} \right] \\ &= \delta^{-\alpha} \left[(1+\theta_N)^{-\alpha} - r_m(\theta_N) \right] w^{N-1,1} \\ &\quad + \delta^{-\alpha} r_m(\theta_N) \left[(1+\theta_{N-1})^{-\alpha} - r_m(\theta_{N-1}) \right] w^{N-2,1} \\ &\quad + \cdots \\ &\quad + \delta^{-\alpha} r_m^{N,2} \left[(1+\theta_1)^{-\alpha} - r_m(\theta_1) \right]. \end{aligned} \tag{3.17}$$

Proposition 3.4 implies that for $j \geqslant 2$,

$$\left| r_m^{N,j} \right| \leq c w^{N,j}. \tag{3.18}$$

We first bound the last term of (3.17) by applying this and (3.16) to obtain

$$\begin{aligned} T_1 := \delta^{-\alpha} \left| r_m^{N,2} \left[(1+\theta_1)^{-\alpha} - r_m(\theta_1) \right] \right| &\leqslant c \delta^{-\alpha} w^{N,2} \left| (1+\theta_1)^{-\alpha} - r_m(\theta_1) \right| \\ &= c \lambda^{-\alpha} (1+\theta_1)^{\alpha} \left| (1+\theta_1)^{-\alpha} - r_m(\theta_1) \right|. \end{aligned} \tag{3.19}$$

When $\theta_1 \leqslant 1$, Proposition 2.2 with $s = \alpha + \gamma$ gives

$$T_1 \leqslant c\theta_1^{\alpha+\gamma}\lambda^{-\alpha}.$$

When $\theta_1 > 1$, since $\gamma \geqslant 0$,

$$T_1 \leqslant c(1+\theta_1)^{\alpha}\lambda^{-\alpha} \leqslant c\theta_1^{\alpha+\gamma}\lambda^{-\alpha}.$$

Thus, in either case, since $\theta_1 = k_N(\lambda-\delta)/\delta \leqslant k_N\lambda/\delta$,

$$T_1 \leqslant \frac{c}{\delta^{\alpha+\gamma}} k_N^{\alpha+\gamma}\lambda^{\gamma}. \tag{3.20}$$

The absolute value of the other terms in (3.17) are given by

$$T_j := \delta^{-\alpha}\left| r_m^{N,j+1}\left[(1+\theta_j)^{-\alpha} - r_m(\theta_j)\right] w^{j-1,1}\right|, \qquad \text{for } j = 2,3,\dots N,$$

where we have defined $r_m^{N,N+1} = 1$ for convenience of notation. Similar to (3.19), we have

$$T_j \leqslant c\lambda^{-\alpha}(1+\theta_j)^{\alpha}\left|(1+\theta_j)^{-\alpha} - r_m(\theta_j)\right|.$$

In this case, θ_j is in $[0,1]$ and we apply Proposition 2.2 with $s = 1+\alpha+\gamma$ to obtain

$$T_j \leqslant c\lambda^{-\alpha}\theta_j^{1+\alpha+\gamma} \leq c\frac{k_N(\lambda-\delta)k_N^{\alpha+\gamma}\lambda^{\gamma}}{(\delta + t_{j-1}(\lambda-\delta))^{1+\alpha+\gamma}}.$$

Thus,

$$\begin{aligned}\sum_{j=2}^{N} T_j &\leqslant c(\lambda-\delta)k_N^{\alpha+\gamma}\lambda^{\gamma}\int_0^1 (\delta + t(\lambda-\delta))^{-1-\alpha-\gamma}\,\mathrm{d}t \\ &\leqslant ck_N^{\alpha+\gamma}\lambda^{\gamma}\int_0^{\infty}(\delta+z)^{-1-\alpha-\gamma}\,\mathrm{d}z \\ &\leqslant \frac{c}{\delta^{\alpha+\gamma}}k_N^{\alpha+\gamma}\lambda^{\gamma}.\end{aligned} \tag{3.21}$$

The last inequality can be obtained by using the following reasoning:

$$\begin{aligned}\int_0^{\infty}(\delta+z)^{-1-\alpha-\gamma}\,\mathrm{d}z &= \int_0^{\delta}(\delta+z)^{-1-\alpha-\gamma}\,\mathrm{d}z + \int_{\delta}^{\infty}(\delta+z)^{-1-\alpha-\gamma}\,\mathrm{d}z \\ &\leqslant \int_0^{\delta}\delta^{-1-\alpha-\gamma}\,\mathrm{d}z + \int_{\delta}^{\infty} z^{-1-\alpha-\gamma}\,\mathrm{d}z = \frac{1+\alpha+\gamma}{\alpha+\gamma}\delta^{-\alpha-\gamma}.\end{aligned}$$

Combining (3.20) and (3.21) gives

$$|e_N(\lambda)| \leqslant \frac{c}{\delta^{\alpha+\gamma}}k_N^{\alpha+\gamma}\lambda^{\gamma}.$$

We note that by (3.16),

$$\mathcal{A}_h^{-\alpha} v = \delta^{-\alpha} \sum_{j=1}^{M} w^{N,1}(\lambda_j)(v, \psi_j)\psi_j \quad \text{and} \quad U_N = \delta^{-\alpha} \sum_{j=1}^{M} r_m^{N,1}(\lambda_j)(v, \psi_j)\psi_j.$$

Thus,

$$\begin{aligned}\left\|\mathcal{A}_h^{-\alpha} v - U_N\right\|^2 &= \sum_{j=1}^{M} |e_N(\lambda_j)|^2 |(v, \psi_j)|^2 \\ &\leqslant \frac{c}{\delta^{2\alpha+2\gamma}} k_N^{2\alpha+2\gamma} \sum_{j=1}^{M} \lambda_j^{2\gamma} |(v, \psi_j)|^2 = \frac{c}{\delta^{2\alpha+2\gamma}} k_N^{2\alpha+2\gamma} \left\|\mathcal{A}_h^{\gamma} v\right\|^2\end{aligned}$$

and this completes the proof. □

As seen in the above theorem, the discrete regularity of the solution determines the rate of convergence for the uniform step size time-stepping method. To some extent, the regularity of the discrete solution is related to the regularity properties of the continuous problem which is being approximated. This will be discussed in the next section.

4. Finite element approximation to fractional powers of second-order elliptic operators

We start with the second-order elliptic problem associated with the bilinear form (1.1) of the introduction, namely the boundary value problem

$$\begin{aligned} -\nabla \cdot (a(x)\nabla w) + q(x)w &= f, \quad \text{for } x \in \Omega, \\ w(x) &= 0, \quad \text{for } x \in \partial\Omega. \end{aligned} \tag{4.1}$$

Here $q(x)$, $a(x)$ and Ω are as in the introduction. The bilinear form (1.1) results from (4.1) in the usual way, i.e., integration against a test function and integration by parts.

REMARK 4.1 We note that all results below are also valid for the operator corresponding to equations (4.1) with Neumann or Robin type boundary conditions, provided that the corresponding bilinear form is coercive in $H^1(\Omega)$, and a proper error estimates theory for the finite-element approximation (cf. Theorem 4.4) is established.

We start by providing some results for the error between the semidiscrete approximation u_h given by (1.4) and the solution u of (1.3). These results depend on the regularity of the data expressed though the spaces $\widetilde{H}^t$. For for $-1 \leqslant t \leqslant 2$, $\widetilde{H}^t(\Omega)$ is defined as

$$\widetilde{H}^t(\Omega) := \begin{cases} H_0^1(\Omega) \cap H^t(\Omega), & 1 \leqslant t \leqslant 2 \\ [L^2(\Omega), H_0^1(\Omega)]_t, & 0 \leqslant t \leqslant 1 \\ [H^{-1}(\Omega), L^2(\Omega)]_{1+t}, & -1 \leqslant t \leqslant 0 \end{cases}$$

with $[\cdot, \cdot]_t$ denoting the real interpolation method and $H^{-1}(\Omega)$ denoting the dual of $H_0^1(\Omega)$.

ASSUMPTION 4.2 $\mathcal{T}$ satisfies elliptic regularity pickup with index $s \in (0, 1]$, that is,

(a) For $f \in \widetilde{H}^{-1+s}(\Omega)$, $\mathcal{T}f$ is in $\widetilde{H}^{1+s}(\Omega)$ and there is a constant c independent of f satisfying

$$\|\mathcal{T} f\|_{\widetilde{H}^{1+s}(\Omega)} \leqslant c\|f\|_{\widetilde{H}^{-1+s}(\Omega)}.$$

(b) The unbounded operator $\mathcal{A}$ extends to a bounded map of $\widetilde{H}^{1+s}(\Omega)$ into $\widetilde{H}^{-1+s}(\Omega)$, i.e., with constant independent of $v \in \widetilde{H}^{1+s}(\Omega)$, we have

$$\|\mathcal{A} v\|_{\widetilde{H}^{-1+s}(\Omega)} \leqslant c\|v\|_{\widetilde{H}^{1+s}(\Omega)}.$$

REMARK 4.3 In Bonito & Pasciak (2015), it has been shown that this implies

$$D(\mathcal{A}^{t/2}) = H_0^1(\Omega) \cap H^t(\Omega), \quad \text{for } t \in [1, 1+s]$$

with equivalent norms.

The above remark shows that $D(\mathcal{A}^{t/2})$ coincides with a Sobolev space of index t. Accordingly, we introduce the notation

$$\dot{H}^t = D(\mathcal{A}^{t/2}), \quad \text{for } t \geqslant 0.$$

A detailed estimation of the error $u - u_h = \mathcal{T}^{\alpha} f - \mathcal{T}_h^{\alpha} \pi_h f$ can be found in Bonito & Pasciak (2015, Theorem 4.3) and is summarized below (see also Fujita & Suzuki, 1991 for the case when $s = 1$).

THEOREM 4.4 (see Bonito & Pasciak, 2015, Theorem 4.3). Let Assumption 4.2 hold and assume that the mesh is globally quasi-uniform so that the inverse inequality holds, e.g., Thomée (2006) or Bonito & Pasciak (2015, Inequality (45)). Set $\beta = s - \alpha$ when $s > \alpha$ and $\beta = 0$ when $s \leqslant \alpha$. For $\gamma \geqslant \beta$, there is a constant C uniform in h such that

$$\left\|\mathcal{T}^{\alpha} f - \mathcal{T}_h^{\alpha} \pi_h f\right\| \leqslant C_{h,\gamma} h^{2s} \|f\|_{\dot{H}^{2\gamma}} \qquad \text{for all } f \in \dot{H}^{2\gamma}, \tag{4.2}$$

where

$$C_{h,\gamma} = \begin{cases} C\log(1/h), & \text{when } \gamma = \beta \text{ and } s \geqslant \alpha, \\ C, & \text{when } \gamma > \beta \text{ and } s \geqslant \alpha, \\ C, & \text{when } \alpha > s. \end{cases}$$

As we see from this theorem, the rate of convergence in the $L^2(\Omega)$-norm is the result of an interplay between the fractional order α, the regularity pickup s of the solution of problem (1.3) and the regularity of the right-hand side f. The bottom line is that one recovers optimal convergence rate $O(h^{2s})$ for $\alpha > s$ when $f \in L^2(\Omega)$. However, if $\alpha < s$, the solution is not in $H^{2s}(\Omega)$ without extra regularity from f so this additional smoothness is needed to get the same rate.

Now if we approximate problem (1.4) using the method of Vabishchevich on the GRM as described in Algorithm 3.1, we get the following bound for the total error (approximation by finite elements and time stepping).

Corollary 4.5 Assume the conditions of Theorem 4.4 hold. Then U_{L+1}, obtained by Algorithm 3.1 with $v = \pi_h f$, $\mathcal{B} = \mathcal{A}_h - \delta\mathcal{I}$, and $K = (L+1)N$ steps, satisfies

$$\left\| \mathcal{T}^{\alpha} f - U_{L+1} \right\| \leqslant C(h^{2s}\|f\|_{\dot{H}^{2\gamma}} + N^{-2m}\|f\|).$$

Corollary 4.6 Assume the conditions of Theorem 4.4 hold. Then U_N, obtained by applying (3.12) with $v = \pi_h f$ and $\mathcal{B} = \mathcal{A}_h - \delta\mathcal{I}$, satisfies

$$\left\| \mathcal{T}^{\alpha} f - U_N \right\| \leqslant C\left(h^{2s}\|f\|_{\dot{H}^{2\gamma}} + N^{-\alpha-\beta}\left\| \mathcal{A}_h^{\beta}\pi_h f \right\|\right), \quad \text{for } 0 \leqslant \beta \leqslant 2m - \alpha.$$

We next consider the question of bounding the norm $\|\mathcal{A}_h^{\beta}\pi_h f\|$ in terms of the regularity of f, i.e., when is there a constant c satisfying

$$\left\| \mathcal{A}_h^{\beta}\pi_h f \right\| \leqslant c\|\mathcal{A}^{\beta} f\|, \qquad \text{for all } f \in D(\mathcal{A}^{\beta}). \tag{4.3}$$

Remark 4.7 We note that the inequality (4.3) holds for any β when V_h is a spectral space, i.e., $V_h = span\{\psi_i\}_{i=1}^{M}$ with ψ_i the eigenfunction of $\mathcal{A}$, $M = O(h^{-2})$. In that case $\mathcal{A}_h^{\beta}\pi_h f = \pi_h \mathcal{A}^{\beta} f$ for any β and (4.3) is immediate.

For general finite element spaces, (4.3) is more subtle. For $\beta \in [0, 1/2]$, this follows when the $L^2(\Omega)$-projector into V_h is a bounded operator on $H_0^1(\Omega)$ with bound independent of h. For globally quasi-uniform meshes, the boundedness result is given in Bank & Dupont (1981) and Bramble & Xu (1991), while the case of certain refined meshes is given in Bank & Yserentant (2014). When the H^1 bound holds, by interpolation, there is a constant c depending only on $\beta \in [0, 1/2]$ satisfying (4.3).

We extend the above inequality to $\beta \in [1/2, (1+s)/2]$ in the next lemma whose proof is included for completeness as it was already observed in Lei (2018).

Lemma 4.8 Suppose that Assumption 4.2 holds and the mesh is globally quasi-uniform. Then there is a constant c depending on $\beta \in [1/2, (1+s)/2]$ such that (4.3) holds.

Let $P_h : H_0^1(\Omega) \to V_h$ denote the elliptic projector, i.e., $P_h w = w_h \in V_h$ is the unique solution of

$$A(w_h, \theta_h) = A(w, \theta_h), \qquad \text{for all } \theta_h \in V_h.$$

Without loss of generality, we can take the norm on $H_0^1(\Omega)$ to be

$$\|v\|_{H^1(\Omega)} = A(v, v)^{1/2} = \|\mathcal{A}^{1/2} v\|.$$

We then have

$$\left\| \mathcal{A}_h^{1/2} P_h w \right\| = \|P_h w\|_{H^1(\Omega)} \leqslant \|w\|_{H^1(\Omega)} = \|\mathcal{A}^{1/2} w\|, \tag{4.4}$$

for all $w \in H_0^1(\Omega) = D(\mathcal{A}^{1/2})$, while the identity $\mathcal{A}_h P_h v = \pi_h \mathcal{A} v$ for $v \in D(\mathcal{A})$ implies that

$$\|\mathcal{A}_h P_h v\| = \|\pi_h \mathcal{A} v\| \leqslant \|\mathcal{A} v\|, \qquad \text{for all } v \in D(\mathcal{A}).$$

It follows by interpolation that for $r \in [1/2, 1]$,

$$\left\| \mathcal{A}_h^r P_h v \right\| \leqslant \| \mathcal{A}^r v \|, \qquad \text{for all } v \in D(\mathcal{A}^r). \tag{4.5}$$

Now for $t \in [1, 1+s]$, Remark 4.3 implies that for $\dot{H}^t = D(\mathcal{A}^{t/2}) = H^t(\Omega) \cap H_0^1(\Omega)$. Thus, for $v \in \dot{H}^t$,

$$\|\pi_h v\|_{t,h} \leqslant \|(\pi_h - P_h)v\|_{t,h} + \|P_h v\|_{t,h} \leq C \left(h^{1-t} \|(\pi_h - P_h)v\|_{1,h} + \|P_h v\|_{t,h} \right), \tag{4.6}$$

where we used the inverse inequality. For the second term we apply $\|P_h v\|_{t,h} \leqslant C\|v\|_{\dot{H}^t}$ that follows from

$$\|P_h v\|_{1+s,h} = \sup_{\chi \in V_h} \frac{\left(\mathcal{A} v, \mathcal{A}_h^{(s-1)/2} \chi \right)}{\|\chi\|} \leqslant \sup_{\chi \in V_h} \frac{\|\mathcal{A} v\|_{\widetilde{H}^{-1+s}} \|\mathcal{A}_h^{(s-1)/2} \chi\|_{\widetilde{H}^{1-s}}}{\|\chi\|} \leqslant C\|v\|_{\widetilde{H}^{1+s}},$$

where we used the equivalence of the norms $\|\chi\|_{1-s,h}$ and $\|\chi\|_{\widetilde{H}^{1-s}}$ (established in Bank & Dupont, 1981; Bramble & Xu, 1991 for quasi-uniform meshes). To bound $\|\mathcal{A} v\|_{\widetilde{H}^{-1+s}}$ we invoke part (b) of Assumption 4.2. Then for $\beta = t/2 \in [1/2, (1+s)/2]$ the inequality (4.3) follows from (4.6), the triangle inequality and the well known error estimates

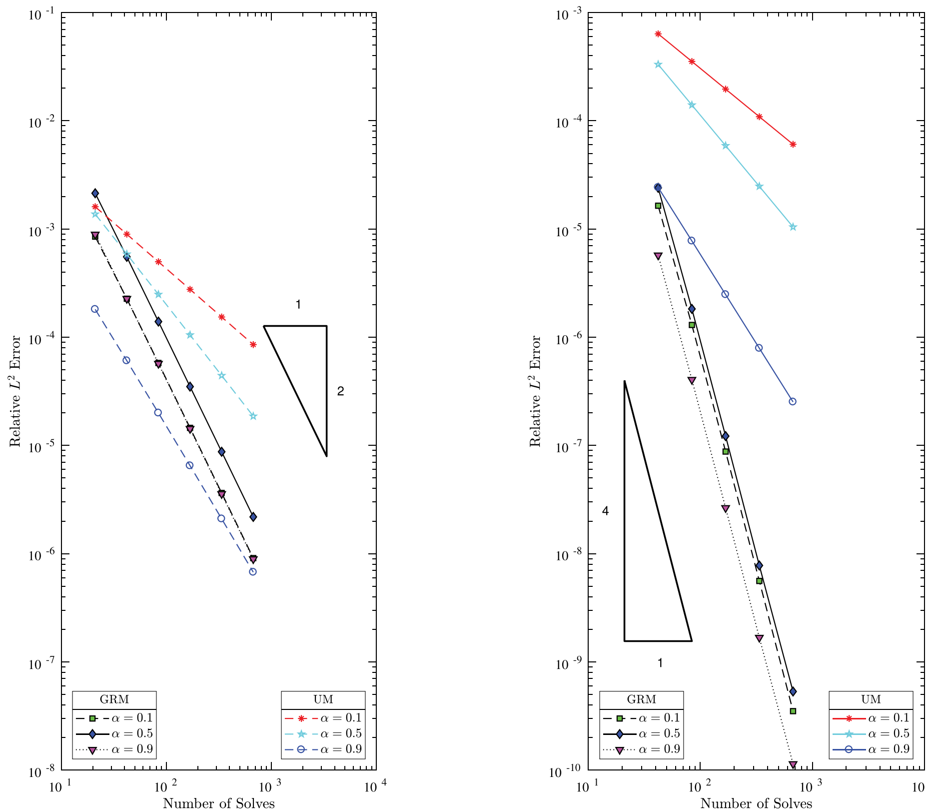
$$\|(I - \pi_h)v\|_{H^1(\Omega)} + \|(I - P_h)v\|_{H^1(\Omega)} \leqslant ch^{t-1} \|v\|_{\dot{H}^t(\Omega)}.$$

5. Numerical examples

In this section, we present numerical examples for the problem coming from (1.1) with $a(x) = 1$, $q(x) = 0$ and $\Omega \subset \mathbb{R}^d$, for $d = 1, 2$. Except for the tests in Section 5.1.1, we take $\delta = 4$ for all one-dimensional examples (in this case $\lambda_1 = \pi^2$) and $\delta = 10$ for two-dimensional examples (recall $\lambda_1 = 2\pi^2$).

TABLE 1 *The error $\|U_N - u_h\|_{L^2(\Omega)}/\|u_h\|_{L^2(\Omega)}$ for case (d) with $\alpha = 0.5$ using GRM scheme with different δ*

m	$\delta \backslash NS$	21	2×21	4×21	8×21	16×21	Conv. rate
$m = 1$	10^{-3}	1.20e−01	3.16e−02	8.10e−03	2.09e−03	5.88e−04	1.83(2)
	1	1.56e−02	4.13e−03	1.05e−03	2.64e−04	6.61e−05	2.00(2)
	4	3.99e−03	1.06e−03	2.69e−04	6.76e−05	1.69e−05	2.00(2)
	9	2.29e−03	6.10e−04	1.55e−04	3.90e−05	9.77e−06	2.00(2)
	9.86	2.16e−03	5.76e−04	1.47e−04	3.69e−05	9.22e−06	2.00(2)
$m = 2$	10^{-3}	3.30e−03	7.24e−04	3.86e−04	2.16e−04	1.18e−04	0.87(4)
	1	2.84e−04	2.19e−05	1.47e−06	9.41e−08	4.84e−09	4.28(4)
	4	7.75e−05	6.14e−06	4.15e−07	2.65e−08	2.10e−09	3.66(4)
	9	4.59e−05	3.60e−06	2.43e−07	1.55e−08	1.08e−09	3.85(4)
	9.86	4.30e−05	3.37e−06	2.27e−07	1.45e−08	9.99e−10	3.86(4)



(A) Case (c): $m = 1$ (B) Case (c): $m = 2$

FIG. 1. Case (c): relative L^2 error on GRM and UM for $m = 1$ (left) and $m = 2$ (right).

5.1 *One-dimensional examples:* $\Omega = (0, 1)$

We will consider approximating the solution of (1.3) with the following choices of f:

1. $f = \exp(-1/x - 1/(1-x) + 4)$ so that $f \in \dot{H}^s$ for any s.
2. $f = x(1-x)$ so that $f \in \dot{H}^s$ for $s < \frac{5}{2}$.
3. $f = \min(x, 1-x)$ so that $f \in \dot{H}^s$ for $s < \frac{3}{2}$.
4. $f = 1$, so that $f \in \dot{H}^s$ for $s < \frac{1}{2}$.

We note that $f = 1$ fails to be in $\dot{H}^{1/2}$ since functions in $\dot{H}^{1/2}$ vanish at $x = 0$ and $x = 1$. For all examples in the one-dimensional case, we use an equally spaced mesh with $h = 1/1000$ and $L = \lceil 2|\log h|/\log 2\rceil$. The total number of time steps for the GRM scheme is thus $(L+1)N$, with $N = 1, 2, 4, 8, \ldots$. To make the comparison of UM and GRM algorithms more meaningful, we report the error as a function of the number of system solves.

TABLE 2 *Approximate order vs. theoretical convergence rates for $m = 1$*

	GRM scheme			UM scheme		
α	0.1	0.5	0.9	0.1	0.5	0.9
f by (a)	2.00(2)	2.00(2)	2.00(2)	1.87(2)	1.95(2)	1.99(2)
f by (b)	2.00(2)	2.00(2)	2.00(2)	1.34(1.35)	1.71(1.75)	1.94(2)
f by (c)	2.00(2)	2.00(2)	2.00(2)	0.85(0.85)	1.25(1.25)	1.63(1.65)
f by (d)	2.00(2)	2.00(2)	2.00(2)	0.40(0.35)	0.77(0.75)	1.16(1.15)

TABLE 3 *Approximate order vs. theoretical convergence rates, $m = 2$*

	GRM scheme			UM scheme		
α	0.1	0.5	0.9	0.1	0.5	0.9
f by (a)	4.00(4)	3.98(4)	3.97(4)	2.63(4)	2.87(4)	2.96(4)
f by (b)	3.97(4)	3.61(4)	3.87(4)	1.35(1.35)	1.75(1.75)	2.01(2.15)
f by (c)	4.00(4)	3.87(4)	3.88(4)	0.85(0.85)	1.25(1.25)	1.65(1.65)
f by (d)	4.00(4)	3.98(4)	4.00(4)	0.42(0.35)	0.79(0.75)	1.17(1.15)

5.1.1 *Influence of the choice of δ.* We first test the robustness of the computations with respect to $\delta \in (0, \lambda_1)$. According to estimate (3.5), the constant in the error bound decreases when $\delta \to \lambda_1$. This is fully confirmed by our numerical experiments shown in Table 1. One can also notice that the accuracy of the method is fairly robust for $\frac{1}{2}\lambda_1 \leqslant \delta \leqslant \lambda_1$ and in full agreement with Theorem 3.2. The best are the results for $\delta \approx \lambda_1$. Here GRM has 21 subintervals and each subinterval is divided by $N = 2, 3, 8, 16$. This results in solving $m \times N \times 21$ systems of the type $(\mathcal{I} + d\mathcal{A}_h)U = V$, $d \geqslant 0$.

5.1.2 *Numerical study of the accuracy of the method for UM and GRM in time.* For the first plot, we use f given by (c) above and $\alpha = 0.1, 0.5, 0.9$. Figure 1 gives log–log plots of the relative L^2 error between $u_h = \mathcal{A}_h^{-\alpha}\pi_h f$ and the result obtained using the refined and uniform time stepping schemes with $m = 1$ (left plot) and $m = 2$ (right plot) as a function of the number of solves. Note that the GRM method leads to smaller error using the same number of solves. Plots for f given by (b) and (d) are similar and are omitted.

To further demonstrate that the numerical results reflect the theoretical results proved earlier, we report the approximate order of convergence going from $N = n$ to $N = 2n$,

$$\text{Approximate order} := \log(E_n/E_{2n})/\log(2). \tag{5.1}$$

We used $n = 8$ for both the UM runs and the GRM runs. The reason that we chose this n for this computation is that the errors were getting so small in the GRM algorithm for $N \geqslant 32$ that, we suspect, computer round off was affecting their significance.

We report the approximate order of convergence computed using (5.1) and compare it with the theoretical rate (in parenthesis) in Tables 2 and 3. Note that the approximate order of convergence was under the assumption that inequality (4.3) holds. Tables 2 and 3 give the rates when $m = 1$ and $m = 2$, respectively, for varying α and f given above. In most cases, the computed order is in good agreement with the theoretical rate for both the refinement and uniform time stepping schemes. In contrast, the

TABLE 4 *Error and the number of steps for the local refinement in space*

N	n_x	e_{semi}	$m=1$ E_{GRM}	$m=1$ NS	$m=1$ E_{UM}	$m=2$ E_{GRM}	$m=2$ NS	$m=2$ E_{UM}
4	72	6.85×10^{-4}	2.24×10^{-4}	92	2.21×10^{-5}	7.78×10^{-5}	23	5.47×10^{-6}
8	176	1.71×10^{-4}	5.62×10^{-5}	232	2.75×10^{-5}	7.75×10^{-5}	29	1.44×10^{-5}
16	416	4.29×10^{-5}	1.41×10^{-5}	560	2.77×10^{-5}	6.14×10^{-6}	70	1.47×10^{-5}
32	960	1.07×10^{-5}	3.52×10^{-6}	1312	2.77×10^{-5}	6.14×10^{-6}	82	1.47×10^{-5}
64	2176	2.68×10^{-6}	8.79×10^{-7}	3008	2.78×10^{-5}	4.15×10^{-7}	188	1.47×10^{-5}
128	4864	6.71×10^{-7}	2.19×10^{-7}	6784	2.77×10^{-5}	4.15×10^{-7}	212	1.47×10^{-5}

TABLE 5 *The error* $\|U_N - u_h\|_{L^2(\Omega)}/\|u_h\|_{L^2(\Omega)}$ *for the fourth-order UM scheme*

Ex.	$\alpha\backslash NS$	15	2×15	4×15	8×15	16×15	32×15	Conv. rate
(e)	0.1	3.10e−05	1.24e−05	4.98e−06	2.00e−06	7.87e−07	2.87e−07	1.31(1.35)
	0.3	3.62e−05	1.26e−05	4.39e−06	1.54e−06	5.35e−07	1.76e−07	1.52(1.55)
	0.5	2.19e−05	6.60e−06	2.00e−06	6.13e−07	1.87e−07	5.49e−08	1.72(1.75)
	0.7	9.39e−06	2.47e−06	6.51e−07	1.73e−07	4.62e−08	1.20e−08	1.92(1.95)
	0.9	2.06e−06	4.71e−07	1.08e−07	2.50e−08	5.83e−09	1.44e−09	2.12(2.15)
(f)	0.1	1.85e−02	1.23e−02	7.70e−03	4.48e−03	2.34e−03	1.05e−03	0.67(0.35)
	0.3	1.81e−02	1.08e−02	6.13e−03	3.25e−03	1.56e−03	6.54e−04	0.82(0.55)
	0.5	8.97e−03	4.78e−03	2.44e−03	1.17e−03	5.11e−04	1.97e−04	0.97(0.75)
	0.7	3.18e−03	1.50e−03	6.81e−04	2.93e−04	1.16e−04	4.07e−05	1.14(0.95)
	0.9	5.77e−04	2.41e−04	9.69e−05	3.71e−05	1.32e−05	4.21e−06	1.31(1.15)

theoretical rate of the smooth problem (for f given by (a)) would be $2m$ if (4.3) held. The results in Table 3 suggests that (4.3) does not hold for $\beta = 4 - \alpha$. In all of the above examples, the error for the refinement scheme is presented as a function of the number of solves and is smaller than the error of the uniform time stepping scheme.

Note that the convergence of the GRM schemes is more robust than that of the UM schemes. The GRM schemes always yield $2m$th order convergence while the convergence rate of UM schemes are related to the parameter α and the (discrete) regularity of the initial data v as suggested by the theory, cf. Corollary 4.6. The advantages of the refinement scheme are especially evident for problems with nonsmooth initial data.

5.2 *A spatial refinement example*

The last one-dimensional example is for $f = 1$, but uses a sequence of refined spatial grids. By (4.2), the semidiscrete error for an unrefined mesh is $O(h^{2s})$ for $s < 1/4 + \alpha$. As the singular behaviour is at the endpoints of the interval, it is natural to use refinement there to try to improve the error behaviour. We consider a mesh resulting from a geometric refinement near 0 and 1, similar to the geometric time stepping refinement at 0. Specifically, our meshes on $[0, 1/2]$ are constructed by restricting the mesh of Section 3.1 to [0,1/2] as a function of N, the number of points per interval. In this construction, we choose L so that $2^{-L} \leqslant h^{-2}$, where $h = 1/(4N)$ is the mesh size on [1/4,1/2]. The mesh on

TABLE 6 *The error* $\|U_{L+1} - u_h\|_{L^2(\Omega)}/\|u_h\|_{L^2(\Omega)}$ *for the fourth-order GRM scheme*

Ex.	$\alpha\backslash NS$	15	2×15	4×15	8×15	16×15	32×15	Conv. rate
(e)	0.1	3.67e−06	2.79e−07	1.86e−08	1.20e−09	1.01e−10	4.33e−11	3.91(4)
	0.3	7.23e−06	5.42e−07	3.59e−08	2.29e−09	1.47e−10	1.48e−11	3.91(4)
	0.5	7.29e−06	5.35e−07	3.52e−08	2.20e−09	1.24e−10	6.53e−11	3.92(4)
	0.7	5.25e−06	3.76e−07	2.46e−08	1.53e−09	8.13e−11	3.39e−11	3.94(4)
	0.9	1.96e−06	1.36e−07	8.74e−09	4.61e−10	1.33e−10	1.64e−10	3.96(4)
(f)	0.1	9.86e−05	7.99e−06	5.46e−07	3.50e−08	2.20e−09	1.45e−10	3.87(4)
	0.3	1.27e−04	1.01e−05	6.86e−07	4.39e−08	2.76e−09	1.74e−10	3.88(4)
	0.5	9.10e−05	7.16e−06	4.83e−07	3.08e−08	1.93e−09	1.34e−10	3.89(4)
	0.7	4.91e−05	3.80e−06	2.55e−07	1.63e−08	1.02e−09	7.05e−11	3.90(4)
	0.9	1.39e−05	1.06e−06	7.08e−08	4.49e−09	3.06e−10	1.63e−10	3.90(4)

[1/2,1] is obtained by reflecting the mesh on [0,1/2] about 1/2. The number of mesh points in space is $O(h^{-1}\log(1/h))$.

Table 4 reports errors using the GRM and UM time-stepping schemes applied to the case when $\mathcal{A}_h$ comes from a sequence of refined spatial meshes as discussed above. For brevity, we only report results for $\alpha = 0.5$. For each spatial mesh, we compute an accurate approximation $u_{h,ref}$ to the semidiscrete solution $u_h = \mathcal{A}_h^{-\alpha}\pi_h f$ by using a highly refined (in time) 4th order GRM time-stepping scheme. We then report the semidiscrete error norm $e_{semi} := \|I_h(u) - u_{h,ref}\|$, where I_h denotes the finite element interpolation operator on the refined spatial mesh. The solution u is computed at the nodes by using 800 000 terms in its Fourier series expansion. The error e_{semi} is important as it gives us an idea how small we need to make the time-stepping error so that the overall error is, for example, less than or equal to $2e_{semi}$. In Table 4, n_x is the number of intervals in the spatially refined grid, NS is the number of time steps used to reduce the GRM error $E_{GRM} := \|u_{GRM} - u_{h,ref}\|$ below e_{semi} and $E_{UM} := \|u_N - u_{h,ref}\|$ is the UM error for $N = 10^5$ time steps. It is clear that the uniform time-stepping method is inefficient for this problem. Indeed, in many cases, the uniform time stepping fails to reduce the error below e_{semi} even when using 10^5 time steps.

5.3 *Two-dimensional examples:* $\Omega = (0,1)^2$

In the two-dimensional case, we consider f given by

1. $f(x,y) = x(1-x)y(1-y)$ so that $f \in \dot{H}^s$ for $s < 5/2$;
2. $f(x,y) = \begin{cases} 1 & 0.25 \leqslant x,y \leqslant 0.75, \\ 0 & \text{otherwise,} \end{cases}$ so that $f \in \dot{H}^s$ for $s < 1/2$.

For brevity, we only report results for the case of $m = 2$. For all runs, we use a uniform mesh in space of size $h = 1/100$ and $L = 14$ in the GRM case. We report the relative error of the time stepping solution compared with $u_h = \mathcal{A}_h^{-\alpha}\pi_h f$. Table 5 gives the errors as a function of α and NS, the number of time steps, for the uniform stepping approximation, while those of Table 6 are for the geometric stepping approximation. Similar to one-dimensional case, the reported convergence rates are obtained by (5.1) using $n := NS = 8 \times 15$ with the theoretical rates in parenthesis.

6. Conclusions

We proposed two time-stepping methods based on Padé approximation for solving a special pseudo-parabolic equation introduced by Vabishchevich for solving equations involving powers of symmetric positive elliptic operators. We consider two schemes that use geometrically refined and uniform meshes in time. The scheme that uses geometrically refined meshes has a convergence rate that does not depend on the smoothness of the solution, while the scheme involving uniform time-mesh depends crucially on the discrete regularity of the solution. Both the theoretical estimates and the numerical tests show that the scheme on geometrically refined meshes is more efficient compared with the uniform time-stepping scheme, especially in the nonsmooth data case.

Acknowledgements

The authors express their thanks to the anonymous reviewers whose remarks and suggestions resulted in improvement of the presentation.

Funding

China Scholarship Council and the Fundamental Research Funds for the Central Universities of Central South University (2016zzts015 to B.D.) and partly by USA National Science Foundation, Division of Mathematical Science (#1620318 grant to R.L.).

References

Baker, G. A. (1975) *Essentials of Padé Approximants*. New York: Academic Press.

Bank, R. E. & Dupont, T. (1981) An optimal order process for solving finite element equations. *Math. Comp.*, **36**, 35–51.

Bank, R. E. & Yserentant, H. (2014) On the H^1-stability of the L^2-projection onto finite element spaces. *Numer. Math.*, **126**, 361–381.

Bonito, A., Borthagaray, J. P., Nochetto, R. H., Otárola, E. & Salgado, A. J. (2018) Numerical methods for fractional diffusion. *Comput. Visual. Sci*. **19**, 19–46.

Bonito, A. & Pasciak, J. (2015) Numerical approximation of fractional powers of elliptic operators. *Math. Comput.*, **84**, 2083–2110.

Bramble, J. H. & Xu, J. (1991) Some estimates for a weighted L^2 projection. *Math. Comp.*, **56**, 463–476.

Fujita, H. & Suzuki, T. (1991) Evolution problems. *Handbook of Numerical Analysis, vol. II.* North-Holland, Amsterdam: Handb. Numer. Anal. II, pp. 789–928.

Gilboa, G. & Osher, S. (2008) Nonlocal operators with applications to image processing. *Multiscale Model. Simul.*, **7**, 1005–1028.

Higham, N. J. (1997) Stable iterations for the matrix square root. *Numer. Algorithms*, **15**, 227–242.

Kato, T. (1961) Fractional powers of dissipative operators. *J. Math. Soc. Japan*, **13**, 246–274.

Kenney, C. & Laub, A. (1989) Padé error estimates for the logarithm of a matrix. *Int. J. Control*, **50**, 707–730.

Kenney, C. & Laub, A. J. (1991) Rational iterative methods for the matrix sign function. *SIAM J. Matrix Anal. Appl*, **12**, 273–291.

Kilbas, A. A., Srivastava, H. M. & Trujillo, J. J. (2006) *Theory and Applications of Fractional Differential Equations*. Amsterdam: Elsevier.

Kwaśnicki, M. (2017) Ten equivalent definitions of the fractional laplace operator. *Fract. Calc. Appl. Anal.*, **20**, 7–51.

Lei, W. (2018) *Numerical approximation of partial differential equations involving fractional differential operators*. *Ph.D. Thesis*, Texas A&M University, College Station, Texas, USA.

Lischke, A., Pang, G., Gulian, M., Song, F., Glusa, C., Zheng, X., Mao, Z., Cai, W., Meerschaert, M. M., Ainsworth, M., Karniadakis, G. E. (2018) What is the fractional laplacian? *Preprint arXiv:1801.09767*.

Lunardi, A. (2007) *Interpolation Theory*, 2nd edn. Pisa, Italy: Edizioni dela Normale.

McCay, B. M. & Narasimhan, M. N. L. (1981) Theory of nonlocal electromagnetic fluids. *Arch. Mech.*, **33**, 365–384.

Metzler, R., Jeon, J.-H., Cherstvy, A. G. & Barkai, E. (2014) Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Phys. Chem. Chem. Phys.*, **16**, 24128–24164.

Pu, Y.-F., Zhou, J.-L. & Yuan, X. (2010) Fractional differential mask: a fractional differential-based approach for multiscale texture enhancement. *IEEE Transactions Image Process.*, **19**, 491–511.

Silling, S. A. (2000) Reformulation of elasticity theory for discontinuities and long-range forces. *J. Mech. Phys. Solids*, **48**, 175–209.

Thomée, V. (2006) *Galerkin Finite Element Methods for Parabolic Problems*. Berlin: Springer.

Vabishchevich, P. N. (2015) Numerically solving an equation for fractional powers of elliptic operators. *J. Comput. Phys.*, **282**, 289–302.