Journal of Scientific Computing (2022) 91:55
https://doi.org/10.1007/s10915-022-01837-4

Exponentially Convergent Trapezoidal Rules to Approximate Fractional Powers of Operators

Lidia Aceto[1] · Paolo Novati[2]

Received: 9 December 2021 / Revised: 15 March 2022 / Accepted: 18 March 2022

Abstract
In this paper we are interested in the approximation of fractional powers of self-adjoint positive operators. Starting from the integral representation of the operators, we apply the trapezoidal rule combined with a double-exponential transform of the integrand function. In this work we show how to improve the existing error estimates for the scalar case and also extend the analysis to operators. We report some numerical experiments to show the reliability of the estimates obtained.

Keywords Matrix functions · Double-exponential transform · Trapezoidal rule · Fractional Laplacian

Mathematics Subject Classification 47A58 · 65F60 · 65D32

1 Introduction

In this work we are interested in the numerical approximation of $\mathcal{L}^{-\alpha}$, $\alpha \in (0, 1)$. Here $\mathcal{L}$ is a self-adjoint positive operator acting in an Hilbert space $\mathcal{H}$ in which the eigenfunctions of $\mathcal{L}$ form an orthonormal basis of $\mathcal{H}$, so that $\mathcal{L}^{-\alpha}$ can be written through the spectral decomposition of $\mathcal{L}$. In other words, for a given $g \in \mathcal{H}$, we have

$$\mathcal{L}^{-\alpha} g = \sum_{j=1}^{+\infty} \mu_j^{-\alpha} \langle g, \phi_j \rangle \phi_j \tag{1}$$

✉ Lidia Aceto
lidia.aceto@uniupo.it

Paolo Novati
novati@units.it

1 Dipartimento di Scienze e Innovazione Tecnologica, Università del Piemonte Orientale, viale T. Michel, 11, 15121 Alessandria, Italy

2 Dipartimento di Matematica e Geoscienze, Università di Trieste, Via Valerio 12/1, 34127 Trieste, Italy

Published online: 04 April 2022

where μ_j and ϕ_j are the eigenvalues and the eigenfunctions of $\mathcal{L}$, respectively, and $\langle\cdot,\cdot\rangle$ denotes the $\mathcal{H}$-inner product. Throughout the paper we also assume $\sigma(\mathcal{L}) \subseteq [1, +\infty)$, where $\sigma(\mathcal{L})$ denotes the spectrum of $\mathcal{L}$.

Applications of (1) include the numerical solution of fractional equations involving the anomalous diffusion, in which $\mathcal{L}$ is related to the Laplacian operator, and this is the main reason for which in recent years a lot of attention has been placed on the efficient approximation of fractional powers. Among the approaches recently introduced we quote here the methods based on the best uniform rational approximations of functions closely related to $\lambda^{-\alpha}$ that have been studied in [8–11]. Another class of methods relies on quadrature rules arising from the Dunford-Taylor integral representation of $\lambda^{-\alpha}$ [1–4, 6, 7, 19, 20]. Very recently, time stepping methods for a parabolic reformulation of fractional diffusion equations, proposed in [21], have been interpreted by Hofreither in [12] as rational approximations of $\lambda^{-\alpha}$.

In this work, starting from the integral representation (see [5])

$$\mathcal{L}^{-\alpha} = \frac{2\sin(\alpha\pi)}{\pi}\int_0^{+\infty} t^{2\alpha-1}(\mathcal{I}+t^2\mathcal{L})^{-1}dt, \qquad \alpha \in (0,1), \tag{2}$$

where $\mathcal{I}$ is the identity operator in $\mathcal{H}$, we consider the trapezoidal rule applied to the double-exponential transform of the integrand function. We recall here that the method based on the single-exponential (SE) transform has been extensively studied in [6, 7], where the authors also provide reliable error estimates. The rate of convergence has been shown to be of type

$$\exp(-c\sqrt{n}), \tag{3}$$

where n is closely related to the number of nodes. The double-exponential transform has been widely investigated in [14–18] for general scalar functions. In this work we show how to improve the existing error estimates for the function $\lambda^{-\alpha}$. We also extend the analysis to operators, showing that it is possible to reach a convergence rate of type

$$\exp\left(-c\sqrt{\frac{n}{\ln n}}\right).$$

While theoretically disadvantageous with respect to the single-exponential approach, we show that the double-exponential approach is actually faster at least for $\alpha \in (1/2, 1)$.

The paper is organized as follows. In Sect. 2 we make a short background concerning the trapezoidal rule with particular attention to functions that decay exponentially at infinity. In Sect. 3 we consider the trapezoidal rule combined with a double-exponential transform. Here the convergence analysis is derived for the approximation of the scalar function $\lambda^{-\alpha}$ and is then extended in Sects. 4 and 5 to the case of the operator $\mathcal{L}^{-\alpha}$. Some concluding remarks are finally reported in Sect. 6.

2 A General Convergence Result for the Trapezoidal Rule

Given a generic continuous function $f:\mathbb{R}\to\mathbb{R}$, in this section we make a short background concerning the trapezoidal approximation

$$I(f) = \int_{-\infty}^{+\infty} f(x)dx \approx h\sum_{\ell=-\infty}^{+\infty} f(\ell h),$$

where h is a suitable positive value. Given M and N positive integers, we denote the truncated trapezoidal rule by

$$T_{M,N,h}(f) = h \sum_{\ell=-M}^{N} f(\ell h).$$

For the error we have

$$\mathcal{E}_{M,N,h}(f) := \left| I(f) - T_{M,N,h}(f) \right| \leq \mathcal{E}_D + \mathcal{E}_{T_L} + \mathcal{E}_{T_R},$$

where

$$\mathcal{E}_D = \left| \int_{-\infty}^{+\infty} f(x)dx - h \sum_{\ell=-\infty}^{+\infty} f(\ell h) \right|,$$

$$\mathcal{E}_{T_L} = h \sum_{\ell=-\infty}^{-M-1} |f(\ell h)|, \qquad \mathcal{E}_{T_R} = h \sum_{\ell=N+1}^{+\infty} |f(\ell h)|.$$

The quantities $\mathcal{E}_D$ and $\mathcal{E}_T := \mathcal{E}_{T_L} + \mathcal{E}_{T_R}$ are referred to as the discretization error and the truncation error, respectively.

Definition 1 [13,Definition 2.12] For $d > 0$, let $\mathcal{D}_d$ be the infinite strip domain of width $2d$ given by

$$\mathcal{D}_d = \{\zeta \in \mathbb{C} : |\text{Im}(\zeta)| < d\}.$$

Let $B(\mathcal{D}_d)$ be the set of functions analytic in $\mathcal{D}_d$ that satisfy

$$\int_{-d}^{d} |f(x + i\eta)|d\eta = \mathcal{O}(|x|^a), \quad x \to \pm\infty, \quad 0 \leq a < 1,$$

and

$$\mathcal{N}(f, d) = \lim_{\eta \to d^-} \left\{ \int_{-\infty}^{+\infty} |f(x + i\eta)|dx + \int_{-\infty}^{+\infty} |f(x - i\eta)|dx \right\} < +\infty.$$

The next theorem gives an estimate for the discretization error of the trapezoidal rule when applied to functions in $B(\mathcal{D}_d)$.

Theorem 1 [13,Theorem 2.20] *Assume* $f \in B(\mathcal{D}_d)$. *Then*

$$\mathcal{E}_D \leq \frac{\mathcal{N}(f, d)}{2 \sinh(\pi d/h)} e^{-\pi d/h}. \tag{4}$$

Theorem 2 *Assume* $f \in B(\mathcal{D}_d)$ *and that there are positive constants* β, γ *and* C *such that*

$$|f(x)| \leq C \begin{cases} \exp(\beta x), & x < 0, \\ \exp(-\gamma x), & x \geq 0. \end{cases} \tag{5}$$

Then,

$$\mathcal{E}_{M,N,h}(f) \leq \frac{N(f, d)}{2 \sinh(\pi d/h)} e^{-\pi d/h} + \frac{C}{\beta} e^{-\beta M h} + \frac{C}{\gamma} e^{-\gamma N h}. \tag{6}$$

Proof By (5) we immediately have

$$\mathcal{E}_{T_L} \leq \frac{C}{\beta} e^{-\beta M h}, \quad \mathcal{E}_{T_R} \leq \frac{C}{\gamma} e^{-\gamma N h}.$$

Using Theorem 1 we obtain (6). □

The above result states that for functions that decay exponentially for $x \to \pm\infty$ it may be possible to have exponential convergence after a proper selection of h. When working with the more general situation

$$I(g) := \int_a^b g(t)dt, \tag{7}$$

one can consider a conformal map

$$\psi : (-\infty, +\infty) \to (a, b),$$

and, through the change of variable $t = \psi(x)$, transform (7) to

$$I(g_\psi) := \int_{-\infty}^{+\infty} g_\psi(x)dx, \qquad g_\psi(x) = g(\psi(x))\psi'(x).$$

A suitable choice of the mapping ψ may allow to work with a function g_ψ that fulfills the hypothesis of Theorem 2 so that $I(g)$ can be evaluated with an error that decays exponentially.

Since the aim of the paper is the computation of $\mathcal{L}^{-\alpha}$ with $\sigma(\mathcal{L}) \subseteq [1, +\infty)$, for $\lambda \geq 1$ we consider now the integral representation (2)

$$\lambda^{-\alpha} = \frac{2\sin(\alpha\pi)}{\pi} \int_0^{+\infty} t^{2\alpha-1}(1 + t^2\lambda)^{-1}dt, \qquad \alpha \in (0, 1). \tag{8}$$

Defining

$$g_\lambda(t) := t^{2\alpha-1}(1 + t^2\lambda)^{-1}, \tag{9}$$

and a change of variable $t = \psi(x)$, $\psi : (-\infty, +\infty) \to (0, +\infty)$, let

$$g_{\lambda,\psi}(x) = g_\lambda(\psi(x))\psi'(x). \tag{10}$$

Let moreover

$$\mathcal{Q}^\alpha_{M,N,h}(g_{\lambda,\psi}) = \frac{2\sin(\alpha\pi)}{\pi} h \sum_{\ell=-M}^{N} g_{\lambda,\psi}(\ell h)$$

be the truncated trapezoidal rule for the computation of $\lambda^{-\alpha}$, that is, for the computation of

$$\frac{2\sin(\alpha\pi)}{\pi} \int_0^{+\infty} g_\lambda(t)dt = \frac{2\sin(\alpha\pi)}{\pi} \int_{-\infty}^{+\infty} g_{\lambda,\psi}(x)dx.$$

We denote the error by

$$\begin{aligned} E_{M,N,h}(\lambda) &= \left|\lambda^{-\alpha} - \mathcal{Q}^\alpha_{M,N,h}(g_{\lambda,\psi})\right| \\ &= \frac{2\sin(\alpha\pi)}{\pi} \mathcal{E}_{M,N,h}(g_{\lambda,\psi}), \end{aligned} \tag{11}$$

and for operator argument

$$E_{M,N,h}(\mathcal{L}) = \left\|\mathcal{L}^{-\alpha} - \mathcal{Q}^\alpha_{M,N,h}(g_{\mathcal{L},\psi})\right\|_{\mathcal{H}\to\mathcal{H}}. \tag{12}$$

3 Double-Exponential Transformation

The DE transform we use here is given by

$$\psi_{DE}(x) = \tau^{-1/2} \exp\left(\frac{\pi}{2}\sinh(x)\right), \qquad \tau > 0. \tag{13}$$

We consider in (8) the change of variable

$$\tau t^2 = \tau\,(\psi_{DE}(x))^2 = \exp(\pi\,\sinh(x)), \qquad \tau > 0.$$

The function involved in this case is

$$\begin{aligned} g_{\lambda,\psi_{DE}}(x) &= \frac{\pi}{2}\tau^{1-\alpha}\frac{\exp(\alpha\pi\,\sinh(x))}{\tau+\lambda\exp(\pi\,\sinh(x))}\cosh(x) \\ &= \frac{\pi}{2}\lambda^{-\alpha}\frac{(\lambda/\tau\exp(\pi\,\sinh(x)))^{\alpha}}{1+\lambda/\tau\exp(\pi\,\sinh(x))}\cosh(x), \end{aligned} \tag{14}$$

and we employ the trapezoidal rule to compute

$$\lambda^{-\alpha} = \frac{2\sin(\alpha\pi)}{\pi}\int_{-\infty}^{+\infty} g_{\lambda,\psi_{DE}}(x)dx.$$

The parameter τ needs to be selected in some way and the analysis is provided in Sect. 5.4. Its introduction is motivated by the fact that, when moving from λ to $\mathcal{L}$, the method (the choice of M, N and h) and the error estimates have to be derived by working uniformly in the interval $[1, +\infty)$ containing $\sigma(\mathcal{L})$. As in the SE case applied to (8), the function $g_{\lambda,\psi_{DE}}(x)$ exhibits a fast decay for $x \to \pm\infty$ (see [7]), but the definition of the strip of analyticity is now much more difficult to handle since everything now depends on λ and τ.

3.1 Asymptotic Behavior of the Integrand Function

In order to apply Theorem 2 we need to study $|g_{\lambda,\psi_{DE}}(x+i\eta)|$. From (14) we have

$$|g_{\lambda,\psi_{DE}}(x+i\eta)| = \frac{\pi}{2}\lambda^{-\alpha}\left|\frac{(\lambda/\tau\exp(\pi\,\sinh(x+i\eta)))^{\alpha}}{1+\lambda/\tau\exp(\pi\,\sinh(x+i\eta))}\right||\cosh(x+i\eta)|.$$

After simple manipulations based on standard relations we find

$$|\cosh(x+i\eta)| = \sqrt{\cosh^2 x - \sin^2\eta},$$

and therefore

$$|\cosh(x+i\eta)| \le \cosh x.$$

Moreover

$$\left|(\lambda/\tau\exp(\pi\,\sinh(x+i\eta)))^{\alpha}\right| = \left(\frac{\lambda}{\tau}\right)^{\alpha}|\exp(\alpha\pi\,\sinh x\cos\eta))|.$$

In addition, we can bound the denominator using the results given in [15, p. 388], that is,

$$\left|\frac{1}{1+\lambda/\tau\exp(\pi\,\sinh(x+i\eta))}\right| \le \frac{1}{1+\lambda/\tau\exp(\pi\,\sinh x\cos\eta)\cos(\pi/2\sin\eta)}.$$

From the above relations we find

$$\left|g_{\lambda,\psi_{DE}}(x+i\eta)\right| \le \frac{\pi}{2}\lambda^{-\alpha}\frac{\cosh x}{\cos(\pi/2\sin\eta)}G_{\alpha}(x,\eta),$$

where

$$G_{\alpha}(x,\eta) = \frac{(\lambda/\tau\exp(\pi\,\sinh x\cos\eta))^{\alpha}}{1+\lambda/\tau\exp(\pi\,\sinh x\cos\eta)}.$$

Let x^* be such that

$$\pi \sinh x^* \cos\eta = \ln(\tau/\lambda);$$

we have

$$G_\alpha(x,\eta) \le \begin{cases} (\lambda/\tau)^\alpha \exp(\alpha\pi \cos\eta \sinh x), & x \le x^*, \\ (\lambda/\tau)^{\alpha-1} \exp(-(1-\alpha)\pi \cos\eta \sinh x), & x > x^*. \end{cases} \tag{15}$$

3.2 Error Estimate for the Scalar Case

The bound (15) implies that

$$\begin{aligned}
\mathcal{N}\left(g_{\lambda,\psi_{DE}}, d\right) &= \lim_{\eta\to d^-} \left\{ \int_{-\infty}^{+\infty} \left|g_{\lambda,\psi_{DE}}(x+i\eta)\right| dx + \int_{-\infty}^{+\infty} \left|g_{\lambda,\psi_{DE}}(x-i\eta)\right| dx \right\} \\
&\le \lim_{\eta\to d^-} \pi\lambda^{-\alpha} \left\{ \frac{1}{\cos(\pi/2\sin\eta)} \int_{-\infty}^{+\infty} G_\alpha(x,\eta)\cosh x dx \right\} \\
&\le \lim_{\eta\to d^-} \frac{\pi\lambda^{-\alpha}}{\cos(\pi/2\sin\eta)} \left\{ (\lambda/\tau)^\alpha \int_{-\infty}^{x^*} \exp(\alpha\pi\cos\eta\sinh x)\cosh x\, dx \right. \\
&\quad \left. + (\lambda/\tau)^{\alpha-1} \int_{x^*}^{+\infty} \exp(-(1-\alpha)\pi\cos\eta\sinh x)\cosh x\, dx \right\} \\
&\le \frac{1}{\alpha(1-\alpha)} \frac{2}{\cos d\cos(\pi/2\sin d)} \lambda^{-\alpha}.
\end{aligned}$$

In addition, assuming $d = d(\lambda,\tau) < \pi/2$, it can be observed that

$$\begin{aligned}
\int_{-d(\lambda,\tau)}^{d(\lambda,\tau)} |g_{\lambda,\psi_{DE}}(x+i\eta)| d\eta &\le \frac{\pi}{2}\lambda^{-\alpha} \int_{-d(\lambda,\tau)}^{d(\lambda,\tau)} \frac{G_\alpha(x,\eta)\cosh x}{\cos(\pi/2\sin\eta)} d\eta \\
&= \mathcal{O}(1) \quad \text{for } x \to \pm\infty.
\end{aligned}$$

Using Theorem 1, for the discretization error we have

$$\left| \int_{-\infty}^{+\infty} g_{\lambda,\psi_{DE}}(x)dx - h\sum_{\ell=-\infty}^{+\infty} g_{\lambda,\psi_{DE}}(\ell h) \right| \le \xi(d) \frac{1}{\alpha(1-\alpha)} \lambda^{-\alpha} \frac{e^{-\pi d/h}}{2\sinh(\pi d/h)},$$

where

$$\xi(d) = \frac{2}{\cos d\cos(\pi/2\sin d)}. \tag{17}$$

The remaining task is to estimate the truncation error. Using (15) we obtain

$$\begin{aligned}
h\sum_{\ell=-\infty}^{-M-1} \left|g_{\lambda,\psi_{DE}}(\ell h)\right| &\le \frac{\pi}{2}\tau^{-\alpha} h \sum_{\ell=-\infty}^{-M-1} \exp(\alpha\pi\sinh(\ell h))\cosh(\ell h) \\
&\le \frac{\pi}{2}\tau^{-\alpha} \int_{-\infty}^{-Mh} \exp(\alpha\pi\sinh x)\cosh(x)dx \\
&\le \frac{\tau^{-\alpha}}{2\alpha} \exp\left(-\alpha\pi\sinh(Mh)\right) \\
&\le \frac{\tau^{-\alpha}}{2\alpha} \exp\left(\frac{\alpha\pi}{2}\right) \exp\left(-\frac{\alpha\pi}{2}\exp(Mh)\right).
\end{aligned}$$

Similarly,

$$h\sum_{\ell=N+1}^{+\infty}\left|g_{\lambda,\psi_{DE}}(\ell h)\right|\le \frac{\pi}{2}\lambda^{-1}\tau^{1-\alpha}h\sum_{\ell=N+1}^{+\infty}\exp(-(1-\alpha)\pi\sinh(\ell h))\cosh(\ell h)$$
$$\le \frac{\pi}{2}\lambda^{-1}\tau^{1-\alpha}h\int_{Nh}^{+\infty}\exp(-(1-\alpha)\pi\sinh x)\cosh(x)dx$$
$$\le \frac{\lambda^{-1}\tau^{1-\alpha}}{2(1-\alpha)}\exp\left(\frac{(1-\alpha)\pi}{2}\right)\exp\left(-\frac{(1-\alpha)\pi}{2}\exp(Nh)\right).$$

The above results are summarized as follows.

Proposition 3 *Using the double-exponential transform, for the quadrature error it holds*

$$\mathcal{E}_{M,N,h}(g_{\lambda,\psi_{DE}})\le\frac{1}{\alpha(1-\alpha)}\xi(d)\lambda^{-\alpha}\frac{e^{-\pi d/h}}{2\sinh(\pi d/h)}+ \tag{18}$$
$$\frac{\tau^{-\alpha}}{2\alpha}\exp\left(\frac{\alpha\pi}{2}\right)\exp\left(-\frac{\alpha\pi}{2}\exp(Mh)\right)+ \tag{19}$$
$$\frac{\lambda^{-1}\tau^{1-\alpha}}{2(1-\alpha)}\exp\left(\frac{(1-\alpha)\pi}{2}\right)\exp\left(-\frac{(1-\alpha)\pi}{2}\exp(Nh)\right), \tag{20}$$

where $\xi(d)$ is defined by (17).

Defining

$$h=\ln\left(\frac{4dn}{\mu}\right)\frac{1}{n},\quad \text{for}\quad n\ge\frac{\mu e}{4d},\quad \mu=\min(\alpha,1-\alpha) \tag{21}$$

as in [15,Theorem 2.14], we first observe that (see (18))

$$\frac{\exp\left(-\frac{\pi d}{h}\right)}{2\sinh\left(\frac{\pi d}{h}\right)}\le\frac{1}{1-e^{-\frac{\pi}{2}\mu e}}\exp\left(\frac{-2\pi dn}{\ln\left(\frac{4dn}{\mu}\right)}\right). \tag{22}$$

Setting $M=N=n$, the choice of h as in (21) leads to a truncation error that decays faster than the discretization one, because for an arbitrary constant c (see (19)-(20))

$$\exp\left(-c\exp(nh)\right)=\exp\left(-\frac{4cdn}{\mu}\right).$$

As consequence the idea is to assume the discretization error as estimator for the global quadrature error, that is, using (18) and (22),

$$\mathcal{E}_{n,n,h}(g_{\lambda,\psi_{DE}})\approx K_\alpha\xi(d)\lambda^{-\alpha}\exp\left(\frac{-2\pi dn}{\ln\left(\frac{4dn}{\mu}\right)}\right), \tag{23}$$

where

$$K_\alpha=\frac{1}{\alpha(1-\alpha)}\frac{1}{1-e^{-\frac{\pi}{2}\mu e}}. \tag{24}$$

Formula (23) is very similar to the one given in [15,Theorem 2.14], that reads

$$\hat{\mathcal{E}}_{M,N,h}(g_{\lambda,\psi_{DE}})\approx\frac{\tau^{-\alpha}}{\mu}\alpha(1-\alpha)\left(K_\alpha\xi(d)+e^{\frac{\pi}{2}\nu}\right)\exp\left(\frac{-2\pi dn}{\ln\left(\frac{4dn}{\mu}\right)}\right) \tag{25}$$

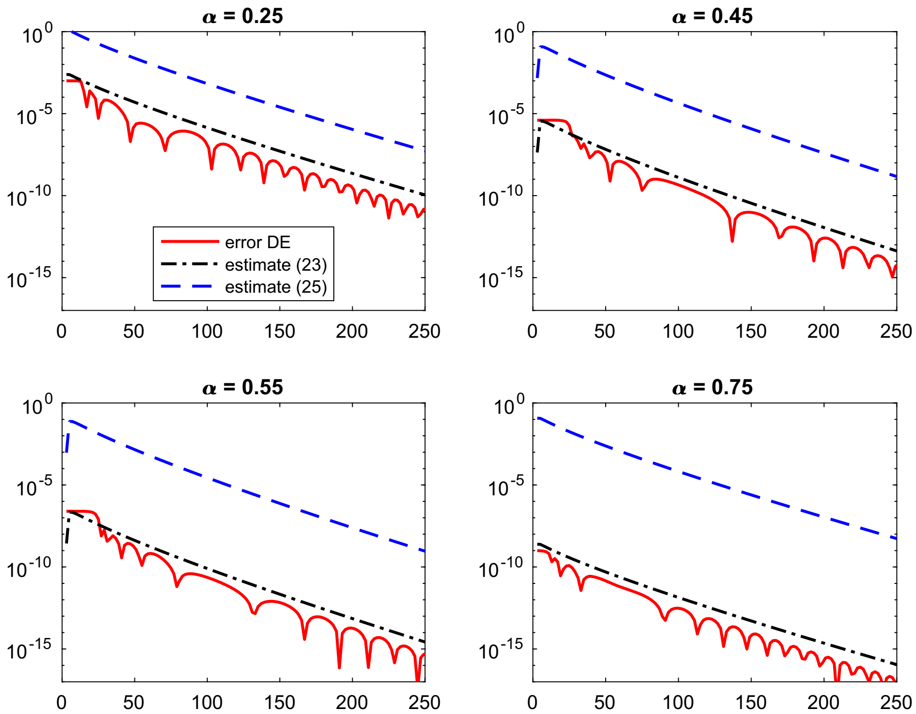

Fig. 1 Error for the trapezoidal rule applied with the double-exponential transform (error DE), estimates (23) and (25) versus the number of inversions, for the computation of $\lambda^{-\alpha}$ with $\lambda = 10^{12}$ and $\tau = 100$

where $\nu = \max(\alpha, 1-\alpha)$ and $M = n$, $N = n - \chi$ (or viceversa depending on α), where $\chi > 0$ is defined in order to equalize the contribute of the truncation errors [15,Theorem 2.11]. The important difference is given by the factor $\lambda^{-\alpha}$ that replaces $\tau^{-\alpha}$, and this is crucial to correctly handle the case of $\lambda \to +\infty$. In this situation the error of the trapezoidal rule goes 0 because $g_{\lambda,\psi_{DE}}(x) \to 0$ as $\lambda \to +\infty$ (see (14)). Anyway, as we shall see, $d \to 0$ as $\lambda \to +\infty$, so that the exponential term itself is not able to reproduce this situation. An example is given in Fig. 1 in which we consider $\lambda = 10^{12}$ and $\tau = 100$.

4 The Poles of the Integrand Function

All the analysis presented so far is based on the assumption that the integrand function

$$g_{\lambda,\psi_{DE}}(x) = \frac{\pi}{2}\tau^{1-\alpha}\frac{\exp(\alpha\pi\sinh x)}{\tau + \lambda\exp(\pi\sinh x)}\cosh x$$

is analytic in the strip $\mathcal{D}_d$, for a certain $d = d(\lambda, \tau)$. Therefore we have to study the poles of this function, that is, we have to study the equation

$$\tau + \lambda\exp(\pi\sinh x) = 0.$$

We have

$$\exp(\pi \sinh x) = \frac{\tau}{\lambda} e^{i\pi},$$
$$\sinh x = \frac{1}{\pi} \ln \frac{\tau}{\lambda} + i(2k+1), \quad k \in \mathbb{Z}.$$

By solving the above equation for each k, we obtain the complete set of poles. Assuming to work with the principal value of the logarithm and taking $k = 0$, we obtain the poles closest to the real axis x_0 and its conjugate $\overline{x_0}$, where

$$x_0 = \sinh^{-1}\left(\frac{1}{\pi} \ln \frac{\tau}{\lambda} + i\right)$$
$$= \ln\left(\frac{1}{\pi} \ln \frac{\tau}{\lambda} + i + \sqrt{\left(\frac{1}{\pi} \ln \frac{\tau}{\lambda}\right)^2 + 2i \frac{1}{\pi} \ln \frac{\tau}{\lambda}}\right). \tag{26}$$

In order to apply the bound on the strip we have to define

$$d = d(\lambda, \tau) = r \operatorname{Im} x_0, \quad 0 < r < 1. \tag{27}$$

The introduction of the factor r is necessary to avoid $\xi(d) \to +\infty$ as $\operatorname{Im} x_0 \to \pi/2$, which verifies for $\lambda \to \tau$ (see (17)).

4.1 Asymptotic Behaviors

Setting

$$s = \frac{1}{\pi} \ln \frac{\lambda}{\tau},$$

we have

$$\frac{1}{\pi} \ln \frac{\tau}{\lambda} = -s,$$

and therefore we can write (26) as

$$x_0 = \ln\left(s\left(-1 + \frac{i}{s} + \sqrt{1 - \frac{2i}{s}}\right)\right).$$

Assuming $\lambda \gg \tau$, that is, $s \gg 1$, and using

$$\sqrt{1-x} \approx 1 - \frac{1}{2}x - \frac{1}{8}x^2 - \frac{1}{16}x^3, \quad x \approx 0, \tag{28}$$

we obtain

$$\sqrt{1 - \frac{2i}{s}} \approx 1 - \frac{i}{s} + \frac{1}{2s^2} + \frac{i}{2s^3}.$$

Using also $\ln(1+x) \approx x$,

$$x_0 \approx \ln\left(s\left(-1 + \frac{i}{s} + 1 - \frac{i}{s} + \frac{1}{2s^2} + \frac{i}{2s^3}\right)\right)$$
$$= \ln\left(s\left(\frac{1}{2s^2} + \frac{i}{2s^3}\right)\right)$$

$$= \ln\left(\frac{1}{2s}\right) + \ln\left(1 + \frac{i}{s}\right)$$
$$\approx \ln\left(\frac{1}{2s}\right) + \frac{i}{s}.$$

Therefore, for $\lambda \gg \tau$,

$$\mathrm{Im}\, x_0 \approx \frac{1}{s} = \frac{\pi}{\ln\frac{\lambda}{\tau}}. \tag{29}$$

Assume now $\lambda = 1$ and $\tau \gg 1$. By (26) we have

$$x_0 = \ln\left(\frac{1}{\pi}\ln\tau + i + \sqrt{\left(\frac{1}{\pi}\ln\tau\right)^2 + 2i\frac{1}{\pi}\ln\tau}\right).$$

Setting

$$s = \frac{1}{\pi}\ln\tau,$$

we have

$$x_0 = \ln\left(s\left(1 + \frac{i}{s} + \sqrt{1 + \frac{2i}{s}}\right)\right)$$
$$\approx \ln\left(s\left(1 + \frac{i}{s} + 1 + \frac{i}{s}\right)\right)$$
$$= \ln\left(2s\left(1 + \frac{i}{s}\right)\right)$$
$$\approx \ln(2s) + \frac{i}{s},$$

that finally leads to

$$\mathrm{Im}\, x_0 \approx \frac{1}{s} = \frac{\pi}{\ln\tau}. \tag{30}$$

5 The Minimax Problem

Let us define the function

$$\varphi(\lambda, \tau) = \xi(d)\lambda^{-\alpha}\exp\left(\frac{-2\pi dn}{\ln\left(\frac{4dn}{\mu}\right)}\right), \quad d = d(\lambda, \tau),$$

representing the (λ, τ)-dependent factor of the error estimate given by (23), that is,

$$\mathcal{E}_{n,n,h}(g_{\lambda,\psi_{DE}}) \approx K_\alpha \varphi(\lambda, \tau),$$

where K_α is defined by (24). Since our aim is to work with a self-adjoint operator with spectrum contained in $[1, +\infty)$ the problem consists in defining properly the parameter τ. This can be done by solving

$$\min_{\tau\geq 1}\max_{\lambda\geq 1}\varphi(\lambda, \tau). \tag{31}$$

As for the true error, experimentally one observes that τ must be taken much greater than 1, independently of α. Therefore, from now on the analysis will be based on the assumption $\tau \gg$

1. Regarding the function $\varphi(\lambda, \tau)$, by taking $d = d(\lambda, \tau)$ as in (27) and n sufficiently large, again, one experimentally observes that with respect to λ the function initially decreases, reaches a local minimum (for $\lambda = \tau$ in which $d = r\pi/2$), then a local maximum (much greater than τ), and finally goes to 0 for $\lambda \to +\infty$ (see Fig. 2). In this view, denoting by $\overline{\lambda}$ the local maximum, for n sufficiently large the problem (31) reduces to the solution of

$$\varphi(1, \tau) = \varphi(\overline{\lambda}, \tau). \tag{32}$$

5.1 Evaluating the Local Maximum

Since $0 < d \le r\pi/2$, $0 < r < 1$, we have

$$0 < C \le \cos d \cos\left(\frac{\pi}{2}\sin d\right) < 1,$$

where C is a constant depending on r. Therefore by (17),

$$2 < \xi(d) \le \frac{2}{C},$$

so that we neglect the contribution of this function in what follows.

Since the maximum is seen to be much larger than τ, we consider the approximation (29). Therefore we have to solve

$$\frac{d}{d\lambda}\lambda^{-\alpha}\exp\left(-\frac{2\pi r\frac{\pi}{\ln\frac{\lambda}{\tau}}n}{\ln\left(\frac{4}{\mu}nr\frac{\pi}{\ln\frac{\lambda}{\tau}}\right)}\right) = 0,$$

that, after some manipulation leads to

$$\frac{d}{d\lambda}\lambda^{-\alpha}\exp\left(-\frac{c_1 n}{\ln\frac{\lambda}{\tau}\, q(\lambda)}\right) = 0,$$

where

$$c_1 = 2\pi^2 r, \quad q(\lambda) = \ln(c_2 n) - \ln\left(\ln\frac{\lambda}{\tau}\right), \quad c_2 = \frac{4}{\mu}\pi r. \tag{33}$$

We find the equation

$$-\alpha\lambda^{-1} - \frac{d}{d\lambda}\left(\frac{c_1 n}{\ln\frac{\lambda}{\tau}\, q(\lambda)}\right) = 0,$$

and since

$$\frac{d}{d\lambda}\left(\frac{c_1 n}{\ln\frac{\lambda}{\tau}\, q(\lambda)}\right) = \frac{c_1 n}{\lambda}\frac{1 - q(\lambda)}{\left(\ln\frac{\lambda}{\tau}\right)^2 q(\lambda)^2},$$

we finally have to solve

$$\alpha + c_1 n\frac{1 - q(\lambda)}{\left(\ln\frac{\lambda}{\tau}\right)^2 q(\lambda)^2} = 0. \tag{34}$$

For large n we have

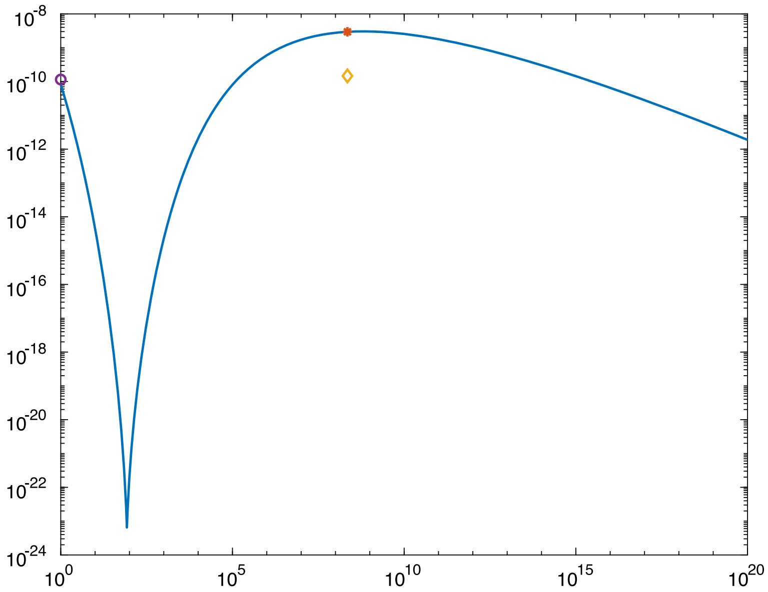

Fig. 2 Plot of the function $\varphi(\lambda, \tau^*)$ for $n = 40$ and $\alpha = 1/2$. The asterisk represents the approximation of the local maximum given by (35), that is, the point $\left(\lambda^*, \varphi(\lambda^*, \tau^*)\right)$. The diamond represents the approximation of $\varphi(\lambda^*, \tau^*)$ stated in (38). Finally the circle is the approximation of $\varphi(1, \tau^*)$ given in (39)

$$q(\lambda) \approx \ln(c_2 n),$$
$$\frac{q(\lambda) - 1}{q(\lambda)^2} \approx \frac{1}{q(\lambda)} \approx \frac{1}{\ln(c_2 n)},$$

so that the solution of (34) can be approximated by

$$\lambda^* = \tau \exp\left(\sqrt{\frac{c_1 n}{\alpha \ln(c_2 n)}}\right). \tag{35}$$

For any given $\tau \geq 1$, it can be observed experimentally that λ^* is a very good approximation of the local maximum (see Fig. 2).

We also remark that the assumption on n stated in (21), that leads to the error estimate (23), is automatically fulfilled for $\lambda = \lambda^*$, at least for α not too small. Indeed, using (27) and (29) we first observe that

$$d(\lambda^*, \tau) \approx \frac{r\pi}{\ln \frac{\lambda^*}{\tau}} = r\pi \sqrt{\frac{\alpha \ln(c_2 n)}{c_1 n}}. \tag{36}$$

Then by (33), using $\mu \leq 1/2$ and assuming for instance $0.9 < r < 1$, after some simple computation we find

$$\frac{\mu e}{4d(\lambda^*, \tau)} \approx \frac{\mu e}{4\pi r} \sqrt{\frac{c_1 n}{\alpha \ln(c_2 n)}}$$
$$\leq \frac{1}{3}\sqrt{\frac{n}{\alpha}}.$$

Therefore the condition (21) holds true for $n \geq 1/(9\alpha)$.

5.2 The Error at the Local Maximum

By (36) clearly $d(\lambda^*, \tau) \to 0$ for $n \to +\infty$, and therefore from (17) we deduce that $\xi(d(\lambda^*, \tau)) \to 2$ for $n \to +\infty$. As consequence

$$\varphi(\lambda^*, \tau) \approx 2\left(\lambda^*\right)^{-\alpha} \exp\left(\frac{-2\pi d(\lambda^*, \tau)n}{\ln\left(\frac{4d(\lambda^*,\tau)n}{\mu}\right)}\right).$$

By defining

$$s_n = \sqrt{\frac{c_1 n}{\alpha \ln(c_2 n)}}, \tag{37}$$

from (35) and (36) we have

$$\begin{aligned} \lambda^* &= \tau \exp(s_n), \\ d(\lambda^*, \tau) &\approx \frac{r\pi}{s_n}, \end{aligned}$$

and hence, after some computation

$$\begin{aligned} \left(\lambda^*\right)^{-\alpha} \exp\left(\frac{-2\pi d(\lambda^*, \tau)n}{\ln\left(\frac{4d(\lambda^*,\tau)n}{\mu}\right)}\right) &\approx \tau^{-\alpha} \exp(-\alpha s_n) \exp\left(\frac{-2\pi \frac{r\pi}{s_n} n}{\ln\left(\frac{4\frac{r\pi}{s_n}n}{\mu}\right)}\right) \\ &= \tau^{-\alpha} \exp\left(-\alpha\left(s_n + \frac{c_1 n}{\alpha s_n \ln\left(\frac{c_2 n}{s_n}\right)}\right)\right). \end{aligned}$$

By (37) we have

$$\begin{aligned} s_n + \frac{c_1 n}{\alpha s_n \ln\left(\frac{c_2 n}{s_n}\right)} &= \sqrt{\frac{c_1 n}{\alpha \ln(c_2 n)}} + \frac{c_1 n}{\sqrt{\frac{\alpha c_1 n}{\ln(c_2 n)}} \ln\left(\frac{c_2 n}{s_n}\right)} \\ &= \sqrt{\frac{c_1 n}{\alpha \ln(c_2 n)}}\left(1 + \frac{\ln(c_2 n)}{\ln\left(\frac{c_2 n}{s_n}\right)}\right) \\ &\approx 3\sqrt{\frac{c_1 n}{\alpha \ln(c_2 n)}}, \end{aligned}$$

because

$$\frac{\ln(c_2 n)}{\ln\left(\frac{c_2 n}{s_n}\right)} \to 2 \quad \text{for} \quad n \to +\infty.$$

Joining the above approximations we finally obtain

$$\begin{aligned} \varphi(\lambda^*, \tau) &\approx 2\tau^{-\alpha} \exp\left(-3\sqrt{\alpha}\sqrt{\frac{c_1 n}{\ln(c_2 n)}}\right) \\ &= 2\tau^{-\alpha} \exp(-3\alpha s_n). \end{aligned} \tag{38}$$

5.3 Error at $\lambda = 1$

By (30), that is,

$$d(1,\tau) \approx r\frac{\pi}{\ln\tau}, \quad \tau \gg 1,$$

we have again $\xi(d(1,\tau)) \approx 2$ and therefore

$$\varphi(1,\tau) \approx 2\exp\left(\frac{-2\pi d(1,\tau)n}{\ln\left(\frac{4d(1,\tau)n}{\mu}\right)}\right).$$

Using (33) we find

$$\begin{aligned}\varphi(1,\tau) &\approx 2\exp\left(-\frac{2\pi r\frac{\pi}{\ln\tau}n}{\ln\left(\frac{4}{\mu}nr\frac{\pi}{\ln\tau}\right)}\right)\\ &= 2\exp\left(-\frac{c_1 n}{\ln\tau\,(\ln(c_2 n) - \ln(\ln\tau))}\right)\\ &\approx 2\exp\left(-\frac{c_1 n}{\ln\tau\,\ln(c_2 n)}\right)\\ &= 2\exp\left(-\frac{\alpha s_n^2}{\ln\tau}\right). \qquad (39)\end{aligned}$$

5.4 Approximating the Optimal Value for τ

We need to solve (32) for τ. Using the approximations (39) and (38) we impose

$$\begin{aligned}\exp\left(-\frac{\alpha s_n^2}{\ln\tau}\right) &= \tau^{-\alpha}\exp(-3\alpha s_n)\\ &= \exp(-3\alpha s_n - \alpha\ln\tau),\end{aligned}$$

that is,

$$-\frac{\alpha s_n^2}{\ln\tau} = -3\alpha s_n - \alpha\ln\tau.$$

Solving the above equation we find

$$\begin{aligned}\ln\tau &= \frac{\left(-3+\sqrt{13}\right)s_n}{2}\\ &\approx 0.3 s_n,\end{aligned}$$

so that

$$\tau^* = \exp(0.3 s_n) \qquad (40)$$

represents an approximate solution of (32).

In Fig. 2 we plot the function $\varphi(\lambda,\tau^*)$ for $\lambda \in [1, 10^{20}]$, in an example in which $n = 40$, $\alpha = 1/2$, and $\tau^* \cong 84.4$ defined by (40). Moreover we show the results of the approximations (35), (38) and (39), for $\tau = \tau^*$. Clearly the ideal situation would be to have τ^* such that $\varphi(1,\tau^*) = \varphi(\overline{\lambda},\tau^*)$, but notwithstanding all the approximations used, the results are fairly good and allow to have a simple expression for τ^*.

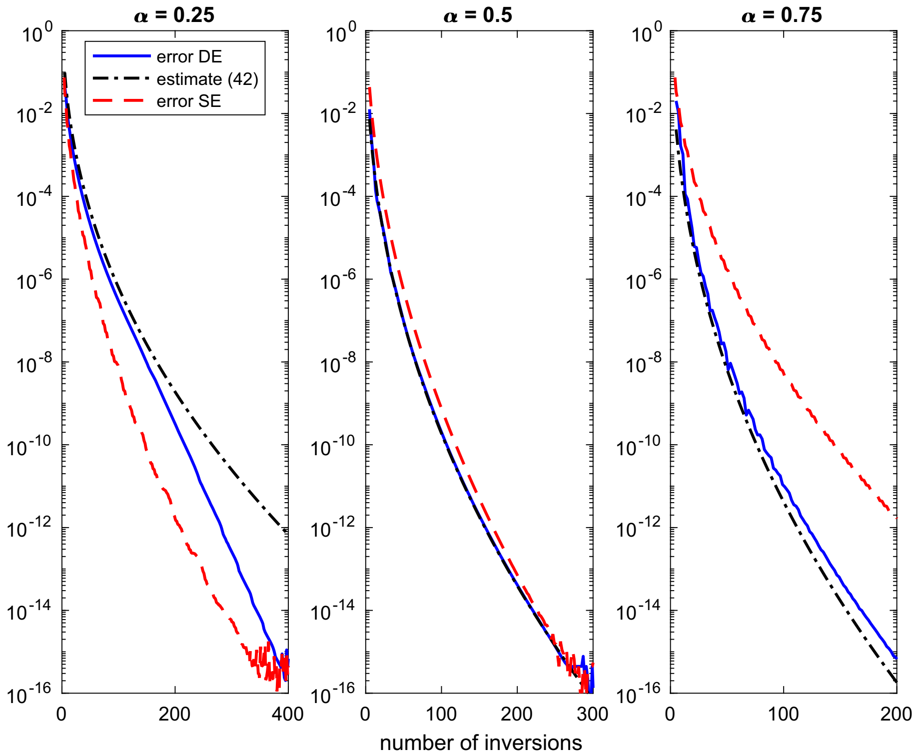

Fig. 3 Error for the trapezoidal rule applied with the double-exponential transform (error DE), with the single-exponential transform (error SE) and error estimate given by (42)

By using (40) in (38) we obtain

$$\varphi(\lambda^*, \tau^*) \approx 2 \exp\left(-3.3\sqrt{\alpha}\sqrt{\frac{c_1 n}{\ln(c_2 n)}}\right). \tag{41}$$

Remembering that by (11)

$$E_{n,n,h}(\mathcal{L}) \leq 2\frac{\sin(\alpha\pi)}{\pi} \max_{\lambda \geq 1} \mathcal{E}_{n,n,h}(g_{\geq,\psi_{DE}}),$$

using (41) we finally obtain the error estimate

$$E_{n,n,h}(\mathcal{L}) \approx \overline{K}_\alpha \exp\left(-3.3\sqrt{\alpha}\sqrt{\frac{c_1 n}{\ln(c_2 n)}}\right), \tag{42}$$

where

$$\overline{K}_\alpha = 4\frac{\sin(\alpha\pi)}{\pi} K_\alpha = 4\frac{\sin(\alpha\pi)}{\pi}\frac{1}{\alpha(1-\alpha)}\frac{1}{1-e^{-\frac{\pi}{2}\mu e}}.$$

In Fig. 3 we show the behavior of the method for the computation of $\mathcal{L}^{-\alpha}$, where $\mathcal{L}$ is the artificial operator

$$\mathcal{L} = [\mathrm{diag}(1, 2, \ldots, 100)]^8, \quad \sigma(\mathcal{L}) \subseteq [1, 10^{16}],$$

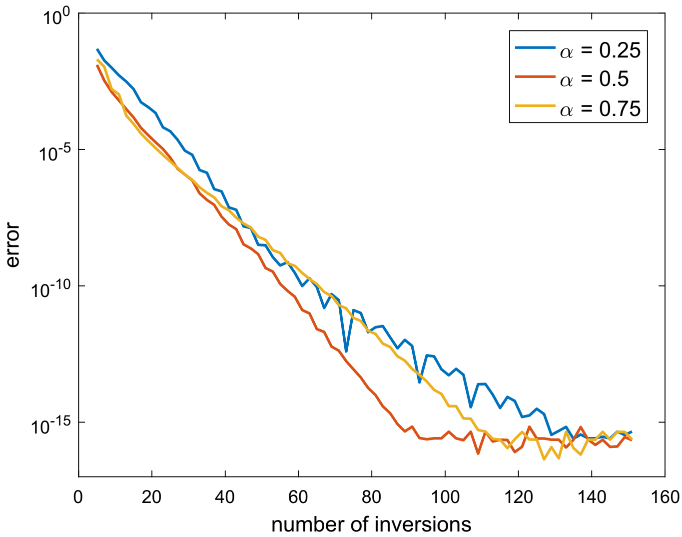

Fig. 4 Spectral norm of the error of the method applied to the computation of $\mathcal{L}^{-\alpha}$, where $\mathcal{L}$ is the discrete one dimensional Laplacian

together with the estimate (42). For comparison, in the same pictures we also plot the error of the SE approach. As mentioned in the introduction, the DE approach appears to be faster for $1/2 < \alpha < 1$.

Finally, we also consider the case of $\mathcal{L}$ equal to the operator obtained by applying the three point finite difference discretization to the one dimensional Laplacian on $(0, \pi)$, with Dirichlet boundary conditions. In Fig. 4 we show the results obtained by considering a uniform mesh with $N = 200$ internal points.

6 Conclusions

In this work we have analyzed the behavior of the trapezoidal rule for the computation of $\mathcal{L}^{-\alpha}$, in connection with the double-exponential transformations. All the analysis has been based on the assumption of $\mathcal{L}$ unbounded, so that the results can be applied even to discrete operators, with spectrum arbitrarily large, without the need to know its amplitude, that is, the largest eigenvalue. In particular we have introduced new error estimates for the scalar and the operator case for the double-exponential transform. The sharp estimate obtained for the scalar case has been fundamental for the proper selection of the parameter τ that is necessary to obtain good results also for the operator case.

Author Contributions The authors contributed equally to this work.

Funding This work was partially supported by GNCS-INdAM and FRA-University of Trieste. The authors are members of the INdAM research group GNCS.

Data Availability Enquiries about data availability should be directed to the authors.

Declarations

Conflict of interest The authors have not disclosed any competing interests.

References

1. Aceto, L., Bertaccini, D., Durastante, F., Novati, P.: Rational Krylov methods for functions of matrices with applications to fractional partial differential equations. J. Comput. Phys. **396**, 470–482 (2019)
2. Aceto, L., Novati, P.: Fast and accurate approximations to fractional powers of operators. IMA J. Numer. Anal. (2021). https://doi.org/10.1093/imanum/drab002
3. Aceto, L., Novati, P.: Padé-type approximations to the resolvent of fractional powers of operators. J. Sci. Comput. **83**(1), 1–17 (2020)
4. Aceto, L., Novati, P.: Rational approximations to fractional powers of self-adjoint positive operators. Numer. Math. **143**, 1–16 (2019)
5. Balakrishnan, A.V.: Fractional powers of closed operators and the semigroups generated by them. Pac. J. Math. **10**, 419–437 (1960)
6. Bonito, A., Pasciak, J.E.: Numerical approximation of fractional powers of elliptic operators. Math. Comput. **84**, 2083–2110 (2015)
7. Bonito, A., Lei, W., Pasciak, J.E.: On sinc quadrature approximations of fractional powers of regularly accretive operators. J. Numer. Math. **27**(2), 57–68 (2019)
8. Harizanov, S., Lazarov, R., Margenov, S., Marinov, P., Vutov, Y.: Optimal solvers for linear systems with fractional powers of sparse SPD matrices. Numer. Linear Algebra Appl. **25**(5), e2167 (2018)
9. Harizanov, S., Lazarov, R., Marinov, P., Margenov, S., Pasciak, J.E.: Analysis of numerical methods for spectral fractional elliptic equations based on the best uniform rational approximation. J. Comput. Phys. **408**, 109285 (2020)
10. Harizanov, S., Lazarov, R., Marinov, P., Margenov, S., Pasciak, J.E.: Comparison analysis of two numerical methods for fractional diffusion problems based on the best rational approximations of t^{γ} on [0, 1], In: Chemnitz Fine Element Symposium, pp. 165–185. Springer, Cham (2017)
11. Harizanov, S., Margenov, S.: Positive approximations of the inverse of fractional powers of SPD M-Matrices, In: Control Systems and Mathematical Methods in Economics, pp. 147–163. Springer, Cham (2018)
12. Hofreither, C.: A unified view of some numerical methods for fractional diffusion. Comput. Math. Appl. **80**(2), 351–366 (2020)
13. Lund, J., Bowers, K.L.: Sinc Methods for Quadrature and Differential Equations. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1992)
14. Okayama, T., Tanaka, K., Matsuo, T., Sugihara, M.: DE-Sinc methods have almost the same convergence property as SE-Sinc methods even for a family of functions fitting the SE-Sinc methods. Part I: definite integration and function approximation. Numer. Math. **125**, 511–543 (2013)
15. Okayama, T., Matsuo, T., Sugihara, M.: Error estimates with explicit constants for Sinc approximation, Sinc quadrature and Sinc indefinite integration. Numer. Math. **124**, 361–394 (2013)
16. Mori, M.: Discovery of the double exponential transformation and its developments. Publ. RIMS Kyoto Univ. **41**, 897–935 (2005)
17. Tanaka, K., Sugihara, M., Murota, K., Mori, M.: Function classes for double exponential integration formulas. Numer. Math. **111**, 631–655 (2009)
18. Trefethen, L.N., Weideman, J.A.C.: The exponentially convergent trapezoidal rule. SIAM Rev. **56**(3), 385–458 (2014)
19. Vabishchevich, P.N.: Approximation of a fractional power of an elliptic operator. Numer. Linear Algebra Appl. **27**(3), e2287 (2020)

20. Vabishchevich, P.N.: Numerical solution of time-dependent problems with fractional power elliptic operator. Comput. Methods Appl. Math. **18**(1), 111–128 (2018)
21. Vabishchevich, P.N.: Numerically solving an equation for fractional powers of elliptic operators. J. Comput. Phys. **282**, 289–302 (2015)