NUMERICAL APPROXIMATION OF PARTIAL DIFFERENTIAL EQUATIONS INVOLVING FRACTIONAL DIFFERENTIAL OPERATORS

A Dissertation

by

WENYU LEI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Joseph E. Pasciak
Co-Chair of Committee,	Andrea Bonito
Committee Members,	William Rundell
	Vivek Sarin
Head of Department,	Emil Straube

August 2018

Major Subject: Mathematics

ABSTRACT

The negative powers of an elliptic operator can be approximated via its Dunford-Taylor integral representation, i.e. we approximate the Dunford-Taylor integral with an exponential convergent sinc quadrature scheme and discretize the integrand (a diffusion-reaction problem) at each quadrature point using the finite element method. In this work, we apply this discretization strategy for a parabolic problem involving fractional powers of elliptic operators and a stationary problem involving the integral fractional Laplacian. The approximation of the parabolic problem is twofold: the homogenous problem and the non-homogeneous problem. We propose an approximation scheme for the homogeneous problem based on a complex-valued integral representation of the solution operator. An exponential convergent sinc quadrature scheme with a hyperbolic contour and a complex-valued finite element method are developed. The approximation of the non-homogeneous problem in space follows the same idea from the homogeneous problem but we need to additionally discretize the problem in the time domain. Here we consider two different approaches: a pseduo-midpoint quadrature scheme in time based on Duhamel's principle and the Crank-Nicolson time stepping method. Both methods guarantee second order convergence in time but require different sinc quadrature schemes to approximate the corresponding fractional operators. The time stepping method is stable provided that the sinc quadrature spacing is sufficiently small. In terms of the approximation of the stationary problem involving integral fractional Laplacian, we consider a Dunford-Taylor integral representation of the bilinear form in the weak formulation. After approximating the integral with a sinc quadrature scheme, we need to approximate the integrand at each quadrature point which contains a solution of a diffusion-reaction equation defined on the whole space. We approximate the integrand problem on a truncated domain together with the finite element method.

For both problems, we provide L^2 error estimates between solutions and their final approximations. Numerical implementation and results illustrating the behavior of the algorithms are also provided.

DEDICATION

This dissertation is dedicated to my parents for their endless love and support.

ACKNOWLEDGMENTS

I owe my deepest gratitude to my advisor Dr. Joseph Pasciak. His constant guidance, support and kindly consideration were extremely helpful to me to complete my research at Texas A&M University. I also express my warmest gratitude to my co-advisor Dr. Andrea Bonito. I have been very much impressed and influenced by his extraordinary insight, persistence and high standards in research.

I wish to thank Dr. Ricardo Nochetto, Dr. Tobias von Petersdorff and Dr. Abner Salgado. They provided me research motivations and showed me many invaluable insights. Appreciation is also expressed to Dr. Wolfgang Bangerth and Dr. Spencer Patty, for their consistent support of my numerical simulations using the deal.II finite element library.

I would like to thank Dr. William Rundell and Dr. Vivek Sarin, who served as members in my graduate research committee and have been a constant help.

Dr. Jean-Luc Germond taught me theory of finite element methods. Dr. Bojan Popov taught me theory and numerics of hyperbolic conservation laws. Dr. Alan Demlow taught me theory of adaptive finite element methods. I would like to thank them and all the professors who taught me at Texas A&M University. I wish to extend my thank to the numerical analysis research group, my friends, colleagues, the department staffs for making my time at Texas A&M University a great experience.

Finally, I would like to express my deepest gratitude to my parents for their constant love, encouragement and support.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Joseph Pasciak (co-chair), Andrea Bonito (chair) and William Rundell of the Department of Mathematics and Professor Vivek Sarin of the Department of Computer Science & Engineering.

Part of this work, especially Chapter V is a summary of joint work with Professor Joseph Pasciak and Andrea Bonito. All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a fellowship from Department of Mathematics, Texas A&M University and a research fellowship form National Science Foundation through Grants DMS-1216551 and DMS-1254618.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION

Various natural phenomena can be modeled by fractional diffusion or anomalous diffusion [53]. A typical example is the Lévy flights with the isotropic measure $|x|^{-d-2s}dx$, where d is the space dimension and $s \in (0,1)$. This is also called the symmetric s-stable Lévy process [4] and the corresponding infinitesimal generator is called fractional Laplacian and denoted by $(-\Delta)^s$ (see also the definition below). The fractional Laplacian has been applied to various areas such as finance [22], predator search patterns [65], peridynamics [64] and porous media flow [27]. Another motivation comes from some physical models in $\mathbb{R}^d$ involving the Dirichlet-to-Neumann (DTN) map in $\mathbb{R}^{d+1}$, e.g. electroconvection [26] and the surface quasigeostrophic models [39]. One can realizes the DTN map as the fractional Laplacian with the power $1/2$. In fact, it is shown in [20] that the fractional Laplacian with any power $s \in (0,1)$ can be treated as a DTN map via the so-called s-harmonic extension problem.

In this work, we consider two kinds of spatial fractional differential operators which are frequently used in above models: fractional powers of elliptic operators and the fractional Laplacian. The former is from spectral theory. Let X and Y be two Hilbert spaces such that Y is dense in X and Y can be compactly embedded into X. The unbounded L mapping from Y to X is defined based on a bilinear form $d(\cdot,\cdot)$ which is symmetric, coercive and bounded on $Y \times Y$, namely,

$$(Lw,\theta) = d(w,\theta), \qquad \text{for all } \theta \in Y,$$

where $(\cdot,\cdot)$ denotes the X inner product. We also refer to Section II.2 for a detailed definition. The fractional differential operator L^s for $s \in (0,1)$ is defined by the eigenfunction expansion

$$L^s v = \sum_{j=1}^{\infty} \lambda_j^s (v,\psi_j)\psi_j. \tag{I.1}$$

Here $\{\psi_j\}$ is an X-orthonormal basis of eigenfunctions of L with positive non-decreasing eigenvalues $\{\lambda_j\}$. In this work, we set Ω be a bounded Lipschitz domain, $X = L^2(\Omega)$, and $Y = H_0^1(\Omega)$. If Ω is convex, the Dirichlet form $d(u, v) = \int_\Omega \nabla u \cdot \nabla v \, dx$ for $u, v \in H_0^1(\Omega)$ defines the unbounded operator $L = -\Delta$ with the domain $D(L) = H_0^1(\Omega) \cap H^2(\Omega)$ (also called the Dirichlet Laplacian) and L^s is referred to as the *spectral* fractional Laplacian. Thanks to the Fredholm Alternative Theorem, the definition (I.1) makes sense with the domain of the operator $D(L^s) = \{v \in L^2(\Omega) : L^s v \in L^2(\Omega)\}$.

The definition of the second operator relies on the Fourier transform. Given v in the Schwartz class, the fractional Laplacian is defined as a pseudo-differential operator with symbol $|\zeta|^{2s}$, i.e.

$$\mathcal{F}((-\Delta)^s v)(\zeta) = |\zeta|^{2s} \mathcal{F}(v)(\zeta) \tag{I.2}$$

with $\mathcal{F}$ denoting the Fourier transform. Parseval's theorem implies that the domain of $(-\Delta)^s$ is a subspace of $H^{2s}(\mathbb{R}^d)$. An equivalent pointwise definition is given by (cf. [47])

$$(-\Delta)^s v(x) = c_{d,s} \text{P.V.} \int_{\mathbb{R}^d} K(x-y)(v(x) - v(y)) \, dy, \quad \text{with } K(y) = \frac{1}{|y|^{d+2s}}, \tag{I.3}$$

where $c_{d,s}$ is a normalization constant and P.V. stands for the principle value. To distinguish with the spectral fractional Laplacian, $(-\Delta)^s$ is also referred to as the *integral* fractional Laplacian.

A crucial tool to investigate the PDEs involving the above fractional differential operators is the equivalent DTN map mentioned above. With this approach, one can study the solution of a local problem in $\mathbb{R}^{d+1}$, which implies the properties of the solution to the original problem; see [21, 59, 67].

I.1 Model Problems and Existing Numerical Approaches

The numerical approximation to PDEs involving fractional differential operators is necessary. Unfortunately, unlike the case $s = 1$, a direct numerical approach (e.g. the finite element method using local basis functions), i.e. assembling and solving a linear system with a dense system matrix is unattractive from a computational point of view. It is also a demanding computational problem as the size of the system matrix becomes large. Our goal is to seek efficient “indirect approaches” to construct and solve such linear systems.

The approximation of stationary problems involving the fractional differential operators has become a popular topic in recent years; see e.g. [41, 54, 12, 13, 2, 31, 40] and the references therein.

A Stationary Problem involving Fractional Powers of Elliptic Operators

Given a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$ and a right hand side data $f \in L^2(\Omega)$, we want to find the solution $u \in D(L^s)$ satisfying

$$L^s u = f, \quad \text{in } \Omega, \tag{I.4}$$

where the fractional differential operator L^s is defined by (I.1). In [43], Kato proposed to represent the solution $u = L^{-s} f$ as a Dunford-Taylor integral with a complex contour encompassing the eigenvalues of L (Cauchy integral in the operator form). Deforming the contour of the integral appropriately, one derives the so-called Balakrishnan formula [6],

$$u = L^{-s} f = \frac{\sin(\pi s)}{\pi} \int_0^\infty \mu^{-s} (\mu I + L)^{-1} f \, d\mu. \tag{I.5}$$

We note that the formula above is a general definition of negative powers of regularly accretive operators; see [43]. The formula (I.5) also suggests the following discretization strategy: apply a numerical integration scheme to the integral with a quadrature spacing $k > 0$ together with a set of quadrature points $\{\mu_j\}$; use finite element methods to approximate $(\mu_j I + L)^{-1} f$ for each j. In [12], an exponentially convergent sinc quadrature scheme [48] is developed and a conforming finite element discretization is applied to approximate the integrand at each quadrature point.

Two alternative approaches are available. In [41], the numerical approximation of the stationary problem is based on the eigenfunction expansion of L_h^s, where L_h is an approximation of L in a finite dimensional space. The resulting approximation is essentially the same as its Dunford-Taylor integral representation (cf. (II.24)) but it requires the computation of the discrete eigenvectors and their eigenvalues (e.g. singular value decomposition). Another approach [54] is to treat the fractional power of L as a DTN map via the s-harmonic extension problem on the semi-infinite cylinder $\Omega \times (0, \infty)$. If $\mathsf{u}(x, y)$ for $(x, y) \in \Omega \times (0, \infty)$ is the solution to the local extension problem, then $u(x) = \mathsf{u}(x, 0)$. Numerically, the extension problem can be approximated using the finite element method in a truncated domain $\Omega \times (0, \mathcal{Y})$ for some $\mathcal{Y} > 1$. The truncation error in $\mathcal{Y}$ becomes exponentially small as $\mathcal{Y}$ increases. We refer to the review paper [8] for extensions based on this approach, such as time dependent problems, obstacle problems and adaptivity.

A Stationary Problem involving the Integral Fractional Laplacian

The stationary problem involving the integral fractional Laplacian reads: given $\Omega \subset \mathbb{R}^d$ and $f \in L^2(\Omega)$, we want to find $u \in L^2(\mathbb{R}^d)$ satisfying

$$u = 0, \quad \text{in } \Omega^c, \qquad (-\Delta)^s u = f, \quad \text{in } \Omega. \tag{I.6}$$

Here $(-\Delta)^s$ is defined by (I.2). We first provide a variational formulation of (I.6): seek $u \in \widetilde{H}^s(\Omega)$ so that

$$a(u, \phi) := \int_{\mathbb{R}^d} \big((-\Delta)^{s/2}\widetilde{u}\big)\big((-\Delta)^{s/2}\widetilde{\phi}\big)\, dx = \int_\Omega f\phi\, dx, \qquad \forall \phi \in \widetilde{H}^s(\Omega), \tag{I.7}$$

where $\widetilde{H}^s(\Omega)$ stands for the set of functions in Ω whose extension by zero belongs to $H^s(\mathbb{R}^d)$ and $\widetilde{u}$ and $\widetilde{\phi}$ denote extensions by zero. The bilinear form $a(\cdot, \cdot)$ on the left hand side of (I.7) is bounded and coercive in $\widetilde{H}^s(\Omega)$. Thus, the Lax-Milgram theory guarantees the existence and uniqueness of the solution u satisfying (I.7) (see Section V.1 for details).

For simplicity, we use continuous piecewise linear finite element spaces to approximate the solution of (I.7). The convergence analysis is classical once the regularity properties of solutions to problem (I.7) are understood (regularity results for (I.7) have been studied; see [2] and [59]).

However, the implementation of the resulting discretization suffers from the fact that, the entries of the stiffness matrix, namely, $a(\phi_i, \phi_j)$, with $\{\phi_k\}$ denoting the finite element basis, cannot be computed exactly except for the one dimensional case.

Similar to the operator form (I.3), an integral representation of the bilinear form is given by

$$a(u, \phi) = \frac{c_{d,s}}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (\widetilde{u}(x) - \widetilde{\phi}(y))(\widetilde{u}(x) - (\widetilde{\phi}(y))K(x-y)\,dy\,dx. \tag{I.8}$$

It is possible to apply the techniques developed for the approximation of boundary integral stiffness matrices [60] to deal with some of the issues associated with the approximation of the double integral above, namely, the application of special techniques for handling the singularity and quadratures. However, (I.8) requires additional truncation techniques as the non-locality of the kernel implies a non-vanishing integrand over $\mathbb{R}^d$. These techniques are used to approximate (I.8) in [31, 2]; see also [1] for a detailed implementation. In particular, [2] use their regularity theory to do *a priori* mesh refinement near the boundary to the rate of convergence under the assumption of exact evaluation of the stiffness matrix.

I.2 Contents of this Work

In this work, we follow the Dunford-Taylor integral approach for the stationary problem (I.4) to develop numerical algorithms for a time dependent problem involving fractional powers of elliptic operators as well as the stationary problem (I.6). Our presentation is organized as follows. We start from Chapter I.2 with basic notations and norm equivalence between different Sobolev spaces. Our instrumental methods of discretization, the finite element method and sinc approximation on the real line are also introduced in Chapter I.2. In Chapter III and Chapter IV, we develop approximation schemes for the time dependent problem. Chapter V discuss the numerical approximation of the bilinear form in (I.7) via its Dunford-Taylor integral representation and provide error estimates for the resulting discrete problem in both the energy and $L^2(\Omega)$ norms.

More detailed descriptions of the contents of this work are given in the following.

Approximation of the Homogeneous Parabolic Problem

The numerical approximation to the time depended problem

$$u_t + L^s u = f, \quad \text{in } \Omega, \qquad u(t=0) = v \tag{I.9}$$

is twofold: the approximation to the homogeneous problem ($f = 0$ and $v \neq 0$) and the approximation to the non-homogeneous problem ($f \neq 0$ and $v = 0$). The former is similar to the stationary problem (I.4). The only difference is that we apply a complex-valued integral representation of the solution operator $W(t) = e^{-tL^s}$ with a hyperbolic contour; see [36, 52, 62] for similar techniques. A sinc quadrature scheme together with a complex-valued finite element approximation are developed in Chapter III. Given a fixed time t, we provide the $L^2(\Omega)$ error estimate for the finite element approximation $W_h(t) = e^{-tL_h^s}$ in Theorem III.3. Our proof is based on the elliptic regularity for the unbounded operator L. We also note that when the fractional power $s = 1$, Theorem III.3 provides the same rate of convergence as in [69, Chapter 20]. Theorem III.5 and Remark III.4 show that the $L^2(\Omega)$ error between the finite element approximation and its sinc approximation with $2N+1$ quadrature points decays in the rate $O(e^{-cN/\ln N})$ for some positive constant c. The total error can be estimated by combining the finite element error and the sinc quadrature error (see Theorem III.6).

Approximation of the Non-homogeneous Parabolic Problem

The numerical approximation the non-homogeneous problem requires two steps: the finite element approximation to static problem and the numerical discretization in the time domain.

The error estimates for the finite element approximation to the static problem follows the homogeneous case; see Theorem IV.1. It turns out that the rate of convergence may degenerate if f is nonsmooth in space.

Starting from the finite element approximation to the static problem, we further consider two different approaches for the discretization in time. One can directly discretize the solution operator, which by Duhamel's principle, is a convolution in time between $W_h(t)$ and the $L^2(\Omega)$-projection of f onto the finite element space. Here we apply a pseudo midpoint quadrature scheme, i.e. given a partition $\{t_j\}$ on $[0,t]$,

$$\begin{aligned}\int_0^t W_h(\xi)\pi_h f(t-\xi)\,d\xi &= \sum_j \int_{t_{j-1}}^{t_j} W_h(\xi)\pi_h f(t-\xi)\,d\xi \\ &\approx \sum_j \Big(\int_{t_{j-1}}^{t_j} W_h(\xi)\,d\xi\,\pi_h f(t-t_{j-\frac{1}{2}})\Big).\end{aligned}$$

where π_h is the $L^2(\Omega)$ projection into the finite element space. Theorem IV.2 shows that the error for the above approximation with $\mathcal{N}$ uniform subintervals is bounded by $O(\mathcal{N}^{-2})$ provided that the right hand side data is in $H^2(0,t;L^2(\Omega))$. We then apply an exponentially convergent sinc quadrature scheme to approximate the Dunford-Taylor integral representation of the discrete operator $\int_{t_{j-1}}^{t_j} W_h(r)\,dr$ with $2N+1$ quadrature points. This leads to an additional error with the rate $O(e^{-c\sqrt{N}})$; see Theorem IV.4.

The second approach is to apply the implicit time stepping methods. Here we consider the Crank-Nicolson scheme. Theorem IV.7 guarantees the second order convergence rate in time but the right hand side data needs to be more regular in space compared with the previous approach. At each time step, one still needs to approximate the discrete fractional operator $(I+\tau L_h^s)^{-1}$, where $\tau > 0$ denotes the time step. This can be approximated by invoking a similar Balakrshnan formula (see (IV.27)). Theorem IV.9 provides the error estimate for the exponentially convergent sinc approximation at each time step. The error between the solution and its final approximation consists of the error from the discretization in both time and space, and errors from the sinc approximation at each time step (see Theorem IV.11).

Both time discretization schemes have their own advantages. As mentioned before, the numerical integration approach requires less regularity in space on the right hand side data. The time stepping method leads to a set of approximations at all time nodes in $[0,t]$, while the numerical integration approach only gives an approximation at the target value of t.

Approximation of the Integral Fractional Laplacian

In Chapter V, we propose a nonconforming finite element algorithm to discretize the problem (I.7) via the Dunford-Taylor integral approach which avoids computing the singular integral (I.8). We first derive an alternative integral representation of the bilinear form in (I.7). That is

$$a(\eta,\theta) = c_s \int_0^\infty t^{1-2s} \int_\Omega ((-\Delta)(I - t^2\Delta)^{-1}\widetilde{\eta})\theta \, dx \, dt \tag{I.10}$$

for $\eta, \theta \in \widetilde{H}^s(\Omega)$. There are two issues needed to be addressed in developing the numerical methods based on the above integral formulation:

- The discretization of the infinite integral with respect to t using a quadrature scheme;
- The approximation of the function $w(t) := w(t,x) = (-\Delta)(I - t^2\Delta)^{-1}\widetilde{\eta}$.

We address the former issue above by first making the change of variable $t^{-2} = e^y$ which results in an integral over $\mathbb{R}$. We then apply a sinc quadrature to obtain the approximation of the bilinear form

$$a^k(\eta,\theta) := \frac{c_s k}{2} \sum_{j=-N^-}^{N^+} e^{sy_j} \int_\Omega ((-\Delta)(e^{y_j} I - \Delta)^{-1}\widetilde{\eta})\theta \, dx, \qquad \text{for all } \theta, \eta \in L^2(\Omega), \tag{I.11}$$

where k is the quadrature spacing, $y_j = kj$, and N^- and N^+ are positive integers. For $\theta \in \widetilde{H}^s(\Omega)$ and $\eta \in \widetilde{H}^\delta(\Omega)$ with $\delta \in (s, 2-s]$. Theorem V.3 together with Remark V.3 shows that the error between $a(\eta,\theta)$ and $a^k(\eta,\theta)$ is bounded by $O(e^{-\pi^2/(2k)})$ provided that $N^+ + N^- = O(1/k^2)$.

Note that for each sinc quadrature point $t_j = e^{-y_j/2}$, $w(t_j)$ is defined in whole space $\mathbb{R}^d$ and decays exponentially as $|x|$ tends to infinity since u is supported on Ω; see [5, Lemma 2.1]. By taking the advantage of this decay property, we approximate $w(t_j)$ in the following two steps. We first approximate $w(t_j)$ in a truncated domain $B^M(t_j)$ which is defined by

$$B^M(t_j) := \begin{cases} \{(1 + t_j(1+M))x \;:\; x \in B\}, & t \geq 1 \\ \{(2+M)x \;:\; x \in B\}, & t < 1, \end{cases}$$

where B is a bounded convex set containing Ω. Then we define the approximation $w^M(t_j)$ solving the same equation but vanishing on the boundary of $B^M(t_j)$. Replacing $w(t_j)$ with $w^M(t_j)$ in (I.11) leads to the approximation of $a^k(\cdot,\cdot)$, denoting by $a^{k,M}(\cdot,\cdot)$. Theorem V.5 guarantees that for sufficiently large M and $\eta,\theta \in L^2(\Omega)$, the error between $a^k(\eta,\theta)$ and $a^{k,M}(\eta,\theta)$ is bounded by $O(e^{-cM})$. Next, we associate to a subdivision of $B^M(t_j)$ the finite element space $\mathbb{V}_h^M(t_j)$ and the restriction $a_h^{k,M}(\cdot,\cdot)$ of $a^{k,M}(\cdot,\cdot)$ to $\mathbb{V}_h^M(t_j) \times \mathbb{V}_h^M(t_j)$. For simplicity, the subdivisions of $B^M(t_j)$ are constructed to coincide on Ω. Denoting by $\mathbb{V}_h(\Omega)$ the set of finite element functions restricted to Ω and vanishing on $\partial\Omega$, our approximation to the solution of (I.7) is the function $u_h \in \mathbb{V}_h(\Omega)$ satisfying

$$a_h^{k,M}(u_h, v_h) = \int_\Omega f v_h \, dx, \qquad \text{for all } v_h \in \mathbb{V}_h(\Omega). \tag{I.12}$$

Lemma V.11 guarantees the $\mathbb{V}_h(\Omega)$-coercivity of the bilinear form $a_h^{k,M}(\cdot,\cdot)$ assuming the sinc quadrature spacing is sufficiently small. Consequently, u_h is well defined again from the Lax-Milgram theory. Moreover, for every t_j, given a sequence of quasi-uniform subdivisions of $B^M(t_j)$, we show (in Theorem V.9) that for η in $\widetilde{H}^\beta(D)$ with $\beta \in (s, 3/2)$ and for $\theta_h \in \mathbb{V}_h(D)$,

$$|a^{k,M}(\eta_h,\theta_h) - a_h^{k,M}(\eta_h,\theta_h)| \le C(1+\ln(h^{-1}))h^{\beta-s}\|\eta\|_{\widetilde{H}^\beta(\Omega)}\|\theta_h\|_{\widetilde{H}^s(\Omega)}.$$

Here C is a constant independent of M, k and h, and $\eta_h \in \mathbb{V}_h(\Omega)$ denotes the Scott-Zhang interpolation or the L^2 projection of v depending on whether $\beta \in (1, 3/2)$ or $\beta \in (s, 1]$.

The first Strang's Lemma implies that the error between u and u_h in the $\widetilde{H}^s(\Omega)$-norm is bounded by the error of the best approximation in $\widetilde{H}^s(\Omega)$ and the sum of the consistency errors from the above discretization steps (see Theorem V.13). If the domain is smooth, we can apply the regularity results from [38] together with standard duality argument to obtain the $L^2(\Omega)$ error estimate; see Theorem V.14.

CHAPTER II

PRELIMINARIES

In this chapter, we review some mathematical tools for both discretization strategies and numerical analysis for our model problems. The outline is as follows. We introduce Sobolev spaces and scales of interpolations spaces in Section II.1 and Section II.2, respectively. We shall use the intepolation spaces to define fractional powers of elliptic operators. This will be discussed in Section II.3. Our numerical techniques, the Galerkin finite element method and the sinc approximation, are reviewed in Section II.4 and Section II.5, respectively. In Section II.6 we review the numerical algorithm for the problem (I.4) as well as the error analysis.

We use the notation

$$A \le cB \qquad \text{and} \qquad A \le CB$$

where c and C are generic constants independent of A, B and discretization parameters. We may hide the above constants by using the notation

$$A \preceq B \qquad \text{and} \qquad A \succeq B.$$

Given two Hilbert spaces X and Y, $\|\cdot\|_{X\to Y}$ denotes the operator norm with the definition

$$\|F\|_{X\to Y} = \sup_{\theta\in X, \theta\neq 0} \frac{\|F(\theta)\|_Y}{\|\theta\|_X}.$$

II.1 Sobolev Spaces

In this work, Sobolev spaces are used to characterize the smoothness of functions. For more details about Sobolev spaces, we refer to [3, 51, 66].

The Weak Derivative.

Let $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_d)$ be a multi-index and define $|\alpha| = \sum_{j=1}^{d} \alpha_j$. For a smooth real-valued function v defined on Ω, the differential operator D^α is given by

$$D^\alpha v = \frac{\partial^{|\alpha|} v}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_d^{\alpha_d}}.$$

Let $C_0^\infty(\Omega)$ be the space of infinity differentiable functions which are compactly supported in Ω. A locally integrable function u has a weak derivative v, if v is locally integrable and satisfies

$$\int_\Omega v\phi \, dx = (-1)^{|\alpha|} \int_\Omega v D^\alpha \phi \, dx, \qquad \text{for all } \phi \in C_0^\infty(\Omega).$$

Throughout the disseration, D^α denotes the weak derivative.

Sobolev Spaces.

Given a Lebesgue measurable function v defined in Ω and $p \geq 1$, define the norm $\|\cdot\|_{L^p(\Omega)}$ by

$$\|v\|_{L^p(\Omega)} := \left(\int_\Omega |v(x)|^p \, dx \right)^{1/p}.$$

Denote $L^p(\Omega)$ the collection of all such v for which $\|v\|_{L^p(\Omega)}$ is finite. Then, given an integer $r \geq 0$, the Sobolev space $W^{r,p}(\Omega)$ is defined by

$$W^{r,p}(\Omega) := \{ v \in L^p(\Omega) \ : \ D^\alpha v \in L^p(\Omega) \text{ for all } |\alpha| \leq r \}.$$

The corresponding Sobolev norm and semi-norm of $W^{r,p}(\Omega)$ are given by

$$\|v\|_{W^{r,p}(\Omega)} := \left(\sum_{|\alpha| \leq r} \|D^\alpha v\|_{L^p(\Omega)}^p \right)^{1/p} \quad \text{and} \quad |v|_{W^{r,p}(\Omega)} := \left(\sum_{|\alpha| = r} \|D^\alpha v\|_{L^p(\Omega)}^p \right)^{1/p},$$

respectively.

For $\sigma \in (0,1)$, define the fractional Sobolev semi-norm $|\cdot|_{W^{\sigma,p}(\Omega)}$ by

$$|v|_{W^{\sigma,p}(\Omega)} := \left(\int_\Omega \int_\Omega \frac{|v(x)-v(y)|^p}{|x-y|^{d+\sigma p}}\, dx\, dy \right)^{1/p}$$

and set $W^{\sigma,p}(\Omega) := \{v \in L^p(\Omega) \ : \ |v|_{W^{\sigma,p}(\Omega)} < \infty\}$. For a positive and non-integer r, we write $r = m + \sigma$ where m is the largest integer less than r so that $\sigma \in (0,1)$. Then, the Sobolev space $W^{r,p}(\Omega)$ is defined by

$$W^{r,p}(\Omega) := \{v \in W^{m,p} \ : \ D^\alpha v \text{ exists and in } W^{\sigma,p}(\Omega) \text{ for all } |\alpha| = m\}$$

and the full norm is given by

$$\|u\|_{W^{r,p}(\Omega)} := \left(\|u\|^p_{W^{m,p}(\Omega)} + \sum_{|\alpha|=m} |D^\alpha u|^p_{W^{\sigma,p}(\Omega)} \right)^{1/p}.$$

For $r \geq 0$, $W^{r,p}(\Omega)$ is a Banach space (see [3]). In particular, when $p = 2$, the Sobolev space $H^r(\Omega) := W^{r,2}(\Omega)$ is a Hilbert space.

We note that in the space

$$H^1_0(\Omega) := \{v \in H^1(\Omega) \mid u = 0 \text{ on } \partial\Omega\},$$

the semi-norm $|\cdot|_{H^1(\Omega)}$ is equivalent with the full norm $\|\cdot\|_{H^1(\Omega)}$ thanks to the Poincaré inequality. Define the dual space $H^{-1}(\Omega)$ to be the set of bounded linear functionals acting on $H^1_0(\Omega)$ such that the norm

$$\|F\|_{H^{-1}(\Omega)} := \sup_{v \in H^1_0(\Omega)} \frac{\langle F, v\rangle}{|v|_{H^1(\Omega)}}$$

is finite. Here $\langle \cdot, \cdot \rangle$ denotes the duality pairing between $H^{-1}(\Omega)$ and $H^1_0(\Omega)$. Note that $L^2(\Omega) \subset H^{-1}(\Omega)$ by identifying $f \in L^2(\Omega)$ with the linear functional $(f, \cdot)_\Omega \in H^{-1}(\Omega)$, where $(\cdot,\cdot)_\Omega$ is the $L^2(\Omega)$ inner product $(v, w)_\Omega = \int_\Omega vw\, dx$.

II.2 Scales of Interpolations Spaces

The Unbounded Operator L and the Dotted Spaces.

Let us assume that $d_\Omega(\cdot,\cdot)$ is a symmetric, coercive and bounded bilinear form on $H_0^1(\Omega)\times H_0^1(\Omega)$. This means that there exists two positive constants c_0 and c_1 such that

$$c_0|v|^2_{H^1(\Omega)} \le d_\Omega(v,v); \qquad |d_\Omega(v,w)| \le c_1 |v|_{H^1(\Omega)} |w|_{H^1(\Omega)}, \qquad \text{for all } v,w \in H_0^1(\Omega).$$

For $f \in L^2(\Omega)$ define $Tf := w \in H_0^1(\Omega)$ to be the unique solution (guaranteed by the Lax-Milgram Theorem) of

$$d_\Omega(w,\theta) = (f,\theta)_\Omega, \qquad \text{for all } \theta \in H_0^1(\Omega). \tag{II.1}$$

Note that T is obviously one to one. So we denote L to be the inverse of T and define the domain of L to be the image of $L^2(\Omega)$ under T.

Now we define the dotted spaces for $r \ge 0$ with respect to the operator L. We note that since T is compact and symmetric on $L^2(\Omega)$, Fredholm theory guarantees that T has a $L^2(\Omega)$-orthonormal basis of eigenfunctions $\{\psi_j\}_{j=1}^\infty$ (also orthogonal in $H_0^1(\Omega)$) with non-increasing real eigenvalues $\mu_1 \ge \mu_2 \ge \mu_3 \ge ... > 0$. Clearly, ψ_j is also an eigenfunction of L with the eigenvalue $\lambda_j = 1/\mu_j$ for every positive integer j. For $r \ge 0$, the dotted space $\dot{H}^r(\Omega)$ is defined by

$$\dot{H}^r(\Omega) := \left\{ f \in L^2(\Omega) : \sum_{j=1}^\infty \lambda_j^r |(f,\psi_j)_\Omega|^2 < \infty \right\}.$$

These spaces form a Hilbert scale of interpolation spaces equipped with the norm

$$\|v\|_{\dot{H}^r(\Omega)} := \left(\sum_{j=1}^\infty \lambda_j^r |(v,\psi_j)_\Omega|^2 \right)^{1/2}.$$

We also denote by $\dot{H}^{-r}(\Omega)$ the dual space of $\dot{H}^r(\Omega)$ when $s \in [0,1]$. It is known that

$$\dot{H}^{-r}(\Omega) = \left\{ F \in H^{-1}(\Omega) : \|F\|_{\dot{H}^{-r}(\Omega)} := \left(\sum_{j=1}^{\infty} \lambda_j^{-r} |\langle F, \psi_j \rangle|^2 \right)^{1/2} < \infty \right\}.$$

The following lemma is instrumental to our numerical analysis.

Lemma II.1. *Let $a \in [0,2]$ and $b \in [0,1]$ satisfying $a+b \le 2$. Then for $\mu \in (0,\infty)$, there holds*

$$\|(\mu I + T)^{-1}\phi\|_{\dot{H}^{-b}(\Omega)} \le \mu^{(a+b)/2-1} \|\phi\|_{\dot{H}^a(\Omega)}, \qquad \textit{for all } \phi \in \dot{H}^a(\Omega).$$

Proof. Writing ϕ in the form of the eigenfunction expansion yields

$$\begin{aligned} \|(\mu I + T)^{-1}\phi\|^2_{\dot{H}^{-b}(\Omega)} &= \sum_{j=1}^{\infty} \frac{\lambda_i^{-b}}{(\mu + \lambda_j^{-1})^2} |(\phi, \psi_j)_\Omega|^2 \\ &= \sum_{j=1}^{\infty} \lambda_j^a \left(\frac{\lambda_j^{-(a+b)/2}}{\mu + \lambda_j^{-1}} \right)^2 |(\phi, \psi_j)_\Omega|^2 \\ &\le \mu^{a+b-2} \|\phi\|^2_{\dot{H}^a(\Omega)}. \end{aligned}$$

The above inequality follows from the Young's inequality

$$\frac{(\lambda_j^{-1})^{(a+b)/2} \mu^{1-(a+b)/2}}{\mu + \lambda_j^{-1}} \le 1.$$

□

Real Interpolation between two Hilbert Spaces.

Let X and Y be two Hilbert spaces. Y is continuously embedded and dense in X. For each $t > 0$ and $v \in X$, define the K-functional

$$K_{X,Y}(t,v) := \inf_{\theta \in Y} \|v - \theta\|_X^2 + t^2 \|\theta\|_Y^2. \tag{II.2}$$

For $\sigma \in (0,1)$, we define the norm (cf. [45, 18])

$$\|u\|_{[X,Y]_\sigma} := \left(\int_0^\infty t^{-2\sigma} K_{X,Y}(t,u)\, \frac{dt}{t} \right)^{1/2}.$$

The corresponding intermediate space $[X,Y]_\sigma$ is then given by

$$[X,Y]_\sigma = \{u \in X \;:\; \|u\|_{[X,Y]_\sigma} < \infty\}.$$

By convention, $[X,Y]_0 = X$ and $[X,Y]_1 = Y$. The space $[X,Y]_\sigma$ is a Hilbert scale between X and Y. If $[X_0,Y_0]_\sigma$ is another scale (as above) and $\mathcal{L}$ is a linear operator which is simultaneously bounded from X_0 to X and Y_0 to Y, i.e. $\|\mathcal{L}u\|_X \le C_1\|u\|_{X_0}$ and $\|\mathcal{L}u\|_Y \le C_2\|u\|_{Y_0}$, then,

$$\|\mathcal{L}u\|_{[X,Y]_\sigma} \le C_1^{1-\sigma} C_2^{\sigma} \|u\|_{[X_0,Y_0]_\sigma}. \tag{II.3}$$

The Intermediate Spaces between $H^{-1}(\Omega)$ and $H^2(\Omega) \cap H_0^1(\Omega)$.

Define the intermediate spaces $\mathbb{H}^s(\Omega)$ for $r \in [-1,2]$ by

$$\mathbb{H}^r(\Omega) := \begin{cases} H^r(\Omega) \cap H_0^1(\Omega), & 1 \le r \le 2, \\ [L^2(\Omega), H_0^1(\Omega)]_r, & 0 \le r \le 1, \\ [H^{-1}(\Omega), L^2(\Omega)]_{r+1}, & -1 \le r \le 0. \end{cases}$$

Here we note that for $r \in [0,1]$, we use the semi-norm $|\cdot|_{H^1(\Omega)}$ as the norm of $H_0^1(\Omega)$. Hence, the corresponding K-functional is

$$K_\Omega(t,u) := \inf_{\theta \in H_0^1(\Omega)} \|u-\theta\|_{L^2(\Omega)}^2 + t^2 |\theta|_{H^1(\Omega)}^2. \tag{II.4}$$

Norm Equivalency.

Since $\|\cdot\|_{\dot{H}^r(\Omega)} = \|\cdot\|_{\mathbb{H}^r(\Omega)}$ for $r = -1, 0, 1$, the spaces $\mathbb{H}^r(\Omega)$ and $\dot{H}^r(\Omega)$ coincide for $r \in [-1, 1]$ and their norms are equal. In order to extend the equivalency to $r \in [1, 2]$, we shall extend the definition of the solution operator T, i.e. we define $TF = w$ as the solution to (II.1) with the right hand side replaced by $\langle F, \theta\rangle$. We then assume that

Assumption II.1. *There exists $\alpha \in (0, 1]$ so that for every $r \in [0, \alpha]$,*

(a) T is a bounded operator mapping from $\mathbb{H}^{-1+r}(\Omega)$ into $\mathbb{H}^{1+r}(\Omega)$;

(b) The functional F defined by

$$\langle F, \theta\rangle := d_\Omega(u, \theta), \qquad \textit{for all } \theta \in H_0^1(\Omega)$$

is a bounded operator from $\mathbb{H}^{1+r}(\Omega)$ to $\mathbb{H}^{-1+r}(\Omega)$.

Under the above assumption, we are able to extend the equivalency property to $r \in [1, 1+\alpha]$:

Proposition II.2 ([12], Proposition 4.1)**.** *$\mathbb{H}^r(\Omega)$ and $\dot{H}^r(\Omega)$ coincide for $r \in [-1, 1+\alpha]$ and their norms are equivalent. In particular when $r \in [-1, 1]$, $\|v\|_{\dot{H}^r(\Omega)} = \|v\|_{\mathbb{H}^r(\Omega)}$ for $v \in \mathbb{H}^r(\Omega)$.*

Remark II.1. *For $r \in (1, 1+\alpha)$, the equivalency constants between $\|\cdot\|_{\dot{H}^r(\Omega)}$ and $\|\cdot\|_{\mathbb{H}^r(\Omega)}$ may depend on Ω due to the regularity estimates in Assumption II.1. In particular, when Ω is a dilation of a fixed domain, these constants are independent of the dilation parameter. This will be proved in Chapter V.*

II.3 Fractional Powers of Elliptic Operators

We shall give the definition of fractional powers of elliptic operators based on the eigenfunction expansion. When the fractional power is negative, we introduce an equivalent definition using the Dunford-Taylor integral. We also refer to [47, Chapter 4] for general definitions of fractional powers of operators, .

Fractional Powers of Elliptic Operators.

For $s \in (0,1)$, define the fractional power of L by eigenfunction expansion (I.1) under the setting in Section II.2. That is,

$$L^s v := \sum_{j=1}^{\infty} \lambda_j^s (v, \psi_j)_\Omega \psi_j, \tag{II.5}$$

with v in the domain of the operator is $D(L^s) := \{v \in L^2(\Omega) : L^s v \in L^2(\Omega)\} = \dot{H}^{2s}(\Omega)$. For $v, w \in \dot{H}^s(\Omega)$, the bilinear form

$$A(v,w) := (L^{s/2}v, L^{s/2}w)_\Omega = \sum_{j=1}^{\infty} \lambda_j^s (v,\psi_j)_\Omega (w,\psi_j)_\Omega. \tag{II.6}$$

satisfies $A(v,v) = \|v\|^2_{\dot{H}^s(\Omega)}$.

The Dunford-Taylor Integral Representation.

For $s \in (0,1)$ and $f \in L^2(\Omega)$, we define $L^{-s}f$ by replacing s with $-s$ in (II.5). L^{-s} is a bounded operator on $L^2(\Omega)$ and we have the following Dunford-Taylor integral representation

$$L^{-s}f = \frac{1}{2\pi i}\int_{\mathcal{C}(r_0,\theta_0)} z^{-s}(zI - L)^{-1} f \, dz \qquad \text{for } f \in L^2(\Omega). \tag{II.7}$$

Here $z^{-s} = e^{-s\ln z}$ with the branch cut for the logarithm along the negative real axis. The above integral contour should encompass the eigenvalues of L and avoid the negative real axis. Indeed, we may choose the contour $\mathcal{C}(r_0,\theta_0)$ for $r_0 > 0$ and $0 < \theta_0 < \pi$ consisting in the following three segments (see also Figure II.1) :

$$\begin{aligned} &\mathcal{C}_1 = \left\{ z(r) := re^{i\theta_0} \text{ with } r \text{ real going from } +\infty \text{ to } r_0 \right\} \text{ followed by} \\ &\mathcal{C}_2 = \left\{ z(\theta) := r_0 e^{i\theta} \text{ with } \theta \text{ going from } \theta_0 \text{ to } -\theta_0 \right\} \text{ followed by} \\ &\mathcal{C}_3 = \left\{ z(r) := re^{-i\theta_0} \text{ with } r \text{ real going from } r_0 \text{ to } +\infty \right\}. \end{aligned} \tag{II.8}$$

Letting $r_0 \to 0$ and $\theta_0 \to \pi$,we obtain the well known Balakrishman formula (cf. [47])

$$L^{-s} f = \frac{\sin(\pi s)}{\pi} \int_0^\infty \mu^{-s} (\mu I + L)^{-1} f \, d\mu. \tag{II.9}$$

Above two integrals converge as Bochner integrals and coincide with the definition of $L^{-s} f$ for all $f \in L^2(\Omega)$ (Theorem 2.1 of [12]).

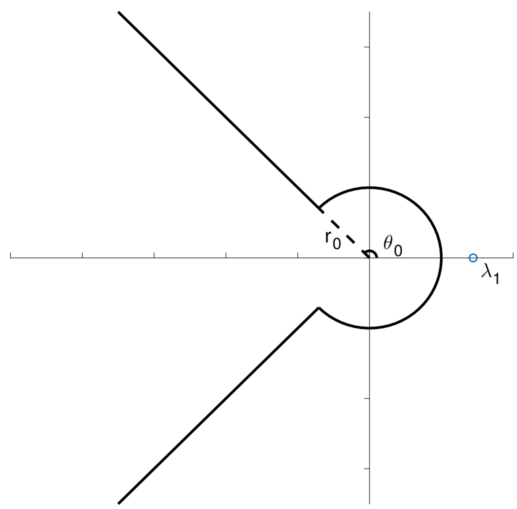

Figure II.1: The example contour $\mathcal{C}(r_0, \theta_0)$ in (II.7).

II.4 Galerkin Finite Element Approximation

Finite Element Spaces.

We additionally assume that Ω is polyhedral. Let $\{\mathcal{T}_j(\Omega)\}_{j=1}^\infty$ be a sequence of globally shape regular and quasi-uniform (cf. [34]) conforming subdivisions of Ω made of simplexes. This means that for any positive integer j and any simplex $\tau \in \mathcal{T}_j(\Omega)$, there are positive constants ρ and c independent of j such that if R_τ denotes the diameter of τ and r_τ denotes the radius of the largest ball which can be inscribed in τ, then,

$$\text{(shape regular)} \qquad R_\tau / r_\tau \leq c \text{ and} \tag{II.10}$$

$$\text{(quasi-uniform)} \qquad \max_{\tau \in \mathcal{T}_j(\Omega)} R_\tau \leq \rho \min_{\tau \in \mathcal{T}_j(\Omega)} R_\tau. \tag{II.11}$$

Fix the positive integer j and denote by $\mathbb{V}_j(\Omega) \subset H_0^1(\Omega)$ the space of continuous piecewise linear functions subodinate to $\mathcal{T}_j(\Omega)$ and by M_j the dimension of $\mathbb{V}_j(\Omega)$. Let $h_j = \max_{\tau \in \mathcal{T}_j(\Omega)} R_\tau$ and set $h_j \to 0$ as $j \to \infty$. Throughout this dissertation, we omit the subscript j and relate the finite element spaces to $h = h_j$, i.e. fix $h > 0$, denote by $\mathcal{T}_h(\Omega)$ the above subdivision, by $\mathbb{V}_h(\Omega)$ the finite element space, and by M_h the dimension of $\mathbb{V}_h(\Omega)$.

Discrete Operators and Discrete Norms.

For any $F \in H^{-1}(\Omega)$, we define the finite element approximation $T_h F \in \mathbb{V}_h(\Omega)$ of $TF \in H_0^1(\Omega)$ as the unique solution to (again, invoking the Lax-Milgram Theorem)

$$d_\Omega(T_h F, \phi_h) = \langle F, \phi_h \rangle, \qquad \text{for all } \phi_h \in \mathbb{V}_h(\Omega). \tag{II.12}$$

For $f \in L^2(\Omega)$, using the previously mentioned identification, $T_h f = T_h \pi_h f$, where π_h is the $L^2(\Omega)$ orthogonal projection onto $\mathbb{V}_h(\Omega)$, namely,

$$(\pi_h f, \phi_h)_\Omega = (f, \phi_h)_\Omega, \qquad \text{for all } \phi_h \in \mathbb{V}_h(\Omega).$$

The operator $L_h : \mathbb{V}_h(\Omega) \to \mathbb{V}_h(\Omega)$ defined by

$$(L_h v_h, \phi_h)_\Omega = d_\Omega(v_h, \phi_h), \qquad \text{for all } \phi_h \in \mathbb{V}_h(\Omega).$$

is the discrete counterpart of L. L_h is the inverse of T_h restricted to $\mathbb{V}_h(\Omega)$. Similar to T, $T_h|_{\mathbb{V}_h(\Omega)}$ has positive eigenvalues $\{\mu_{j,h}\}_{j=1}^{M_h}$ with corresponding $L^2(\Omega)$-orthonormal eigenfunctions $\{\psi_{j,h}\}_{j=1}^{M_h}$. The eigenvalues of L_h are denoted by $\lambda_{j,h} := \mu_{j,h}^{-1}$ for $j = 1, 2, \ldots, M_h$.

Define the discrete fractional operator $L_h^s : \mathbb{V}_h(\Omega) \to \mathbb{V}_h(\Omega)$ by

$$L_h^s v_h := \sum_{j=1}^{M_h} \lambda_{j,h}^s (v_h, \psi_{j,h})_\Omega \psi_{j,h}, \tag{II.13}$$

and the bilinear form $A_h(\cdot,\cdot)$ on $\mathbb{V}_h(\Omega) \times \mathbb{V}_h(\Omega)$ by

$$A_h(v_h, w_h) := \sum_{j=1}^{M_h} \lambda_{j,h}^s (v_h, \psi_{j,h})_\Omega (w_h, \psi_{j,h})_\Omega. \tag{II.14}$$

For $r \geq 0$, we also define the discrete dotted norms $\|\cdot\|_{\dot{H}_h^r(\Omega)}$ by

$$\|v_h\|_{\dot{H}_h^r(\Omega)} := \left(\sum_{j=1}^{M_h} \lambda_{j,h}^r |(v_h, \psi_{j,h})_\Omega|^2 \right)^{1/2}, \qquad \text{for } v_h \in \mathbb{V}_h(\Omega).$$

Due to the fact that $\max_j \lambda_{j,h} \leq Ch^{-2}$ (cf. [18, eq. (2.8)]), the above discrete dotted norms directly implies the following inverse estimate: for $r, \sigma \geq 0$, we have

$$\|v_h\|_{\dot{H}_h^{r+\sigma}(\Omega)} \leq Ch^{-\sigma} \|v_h\|_{\dot{H}_h^r(\Omega)}, \qquad \text{for } v_h \in \mathbb{V}_h(\Omega), \tag{II.15}$$

where the constant C only depends on c and ρ in (II.10) and (II.11), respectively. A discrete version of Lemma II.1 is the following.

Lemma II.3. *Let $a \in [0, 2]$ and $b \in [0, 1]$ with $a + b \leq 2$. Then for any $\mu \in (0, \infty)$,*

$$\|(\mu I + T_h)^{-1} \phi_h\|_{\dot{H}_h^{-b}(\Omega)} \leq \mu^{(a+b)/2-1} \|\phi_h\|_{\dot{H}_h^a(\Omega)}, \qquad \textit{for all } \phi_h \in \mathbb{V}_h(\Omega).$$

Approximation Results.

By Lemma 5.1 of [13], there exists a constant $c(r, \sigma)$ independent of h such that for $\sigma \in [0, 2]$ and $r \in [0, 1]$ satisfying $r + \sigma \leq 2$, there holds

$$\|(I - \pi_h) f\|_{\mathbb{H}^r(\Omega)} \leq c(r, \sigma) h^\sigma \|f\|_{\mathbb{H}^{r+\sigma}(\Omega)}. \tag{II.16}$$

Also we recall the following finite element error estimate from [13].

Proposition II.4 ([13], Lemma 6.1)**.** *Let part (a) of Assumption II.1 holds for some $\alpha \in (0, 1]$. Let $r \in [0, \frac{1}{2}]$ and set $\alpha_* = (\alpha + \min(1 - 2r, \alpha))/2$. Then there is a constant C independent of h such that for all $f \in \mathbb{H}^{\alpha-1}(\Omega)$,*

$$\|(T - T_h)f\|_{\mathbb{H}^{2r}(\Omega)} \leq Ch^{2\alpha_*}\|f\|_{\mathbb{H}^{\alpha-1}(\Omega)}. \tag{II.17}$$

Norm Equivalency between Continuous and Discrete Scales.

We shall need the norm equivalency between $\|v_h\|_{\dot{H}_h^r(\Omega)}$ and $\|v_h\|_{\dot{H}^r(\Omega)}$ for $v_h \in \mathbb{V}_h(\Omega)$. When $r \in [0, 1]$, there holds

$$c\|v_h\|_{\dot{H}_h^r(\Omega)} \leq \|v_h\|_{\dot{H}^r(\Omega)} \leq \|v_h\|_{\dot{H}_h^r(\Omega)}, \tag{II.18}$$

where the constant c only depending on shape-regularity and quasi-uniformity. The right inequality above follows from the interpolation and the fact that the norms in $L^2(\Omega)$ and $H^1(\Omega)$ coincide with those of $\dot{H}_h^0(\Omega)$ and $\dot{H}_h^1(\Omega)$ when restricted to $v_h \in \mathbb{V}_h(\Omega)$. In view of (II.16) with $\sigma = 0$, π_h is stable in $\mathbb{H}^r(\Omega)$ norms (see also [17]) for $r \in [0, 1]$, i.e.

$$\|\pi_h f\|_{\mathbb{H}^r(\Omega)} \leq c\|f\|_{\mathbb{H}^r(\Omega)}, \qquad \text{for all } f \in \mathbb{H}^r(\Omega). \tag{II.19}$$

Letting $r = 0$ and $r = 1$ respectively yields that π_h is simultaneously bounded from $L^2(\Omega)$ to $\dot{H}_h^0(\Omega)$ and $\mathbb{H}^1(\Omega)$ to $\dot{H}_h^1(\Omega)$. Invoking the interpolation estimates (II.3) and setting $f \in \mathbb{V}_h(\Omega)$ yield the left inequality of (II.18).

Remark II.2. *Let α be the regularity index given by Assumption II.1. For $r \in [1, 1 + \alpha] \cap [1, 3/2)$ and for all $v_h \in \mathbb{V}_h(\Omega)$, it follows that (see [70, Prop. 3.10] for a proof)*

$$c\|v_h\|_{\dot{H}_h^r(\Omega)} \leq \|v_h\|_{\dot{H}^r(\Omega)} \leq C\|v_h\|_{\dot{H}_h^r(\Omega)},$$

where the dependency of the constant c is the same as the constant in (II.18) *but the constant C*

additionally depends on constants in Assumption II.1.

When $r \in (1, 1+\alpha]$, we have

Lemma II.5. *For $v \in \dot{H}^r(\Omega)$ with $r \in [1, 1+\alpha]$, there exists a constant C independent of h so that*

$$\|\pi_h v\|_{\dot{H}_h^r(\Omega)} \le C\|v\|_{\dot{H}^r(\Omega)}. \tag{II.20}$$

Proof. Let $w \in H_0^1(\Omega)$, define the Riesz projection $r_h w := w_h \in \mathbb{V}_h(\Omega)$ which uniquely solves

$$d_\Omega(w_h, \phi_h) = d_\Omega(w, \phi_h), \qquad \text{for all } \phi_h \in \mathbb{V}_h(\Omega).$$

We claim that $\|r_h v\|_{\dot{H}_h^r(\Omega)} \le C\|v\|_{\dot{H}^r(\Omega)}$. Indeed,

$$\|r_h v\|_{\dot{H}_h^r(\Omega)} = \sup_{\phi \in \mathbb{V}_h(\Omega)} \frac{d_\Omega(r_h v, \phi)}{\|\phi\|_{\dot{H}_h^{2-r}(\Omega)}} \le C \sup_{\phi \in \dot{H}^{2-r}(\Omega)} \frac{d_\Omega(v, \phi)}{\|\phi\|_{\dot{H}^{2-r}(\Omega)}} = C\|v\|_{\dot{H}^r(\Omega)}$$

Here the inequality above holds due to (II.18) and $\mathbb{V}_h(\Omega) \subset \dot{H}^{2-r}(\Omega)$. Hence, for $v \in \dot{H}^r(\Omega)$, we apply the above result to obtain

$$\begin{aligned}
\|\pi_h v\|_{\dot{H}_h^r(\Omega)} &\le \|r_h v\|_{\dot{H}_h^r(\Omega)} + \|(\pi_h - r_h)v\|_{\dot{H}_h^r(\Omega)} \\
&\le \|r_h v\|_{\dot{H}_h^r(\Omega)} + Ch^{1-r}\left(\|(I-\pi_h)v\|_{\dot{H}^1(\Omega)} + \|(I-r_h)v\|_{\dot{H}^1(\Omega)}\right) \\
&\le C\|v\|_{\dot{H}^r(\Omega)}.
\end{aligned}$$

Note that in the second inequality above we use the inverse estimate (II.15) and in the third inequality we apply (II.16), Proposition II.2 and the approximation property of r_h (e.g. [69, eq. (2.23)])

$$\|(I - r_h)v\|_{H^1(\Omega)} \le Ch^{r-1}\|v\|_{\dot{H}^r(\Omega)}.$$

The proof is complete. □

II.5 Sinc Quadrature on the Real Line

The sinc method is our primary technique to discretize the integral representation such as (II.9). In this section, we introduce a class of functions whose integrals along the real line can be approximated with exponential accuracy with respect to the spacing using a sinc method.

The Sinc Method.

Given $f \in L^1(\mathbb{R})$, we want to approximate the integral $\int_{\mathbb{R}} f(x)\,dx$. We first subdivide the real line by equally spaced intervals with a quadrature spacing $k > 0$ and approximate

$$\int_{\mathbb{R}} f(x)\,dx \approx k \sum_{j=-\infty}^{\infty} f(jk). \tag{II.21}$$

Then we approximate the above infinite sum by a finite sum, i.e. given two positive integers N^+ and N^- depending on k, define the sinc approximation by

$$k \sum_{j=-N^-}^{N^+} f(jk). \tag{II.22}$$

Error Analysis.

Let us consider the following function space for the sinc approximation.

Definition II.1. *Given $\mathsf{d} > 0$, define the space $S(B_{\mathsf{d}})$ to be the set of functions f defined on $\mathbb{R}$ satisfying*

(a) f extends to an analytic function in the infinite strip

$$B_{\mathsf{d}} := \{z \in \mathbb{C} : \ \mathfrak{Im}(z) < \mathsf{d}\}$$

and is continuous on $\overline{B_{\mathsf{d}}}$.

(b) There exists a constant C independent of $y \in \mathbb{R}$ such that

$$k \int_{-\mathsf{d}}^{\mathsf{d}} |g(y + iw)| \, dw \leq C;$$

(c)

$$N(B_{\mathsf{d}}) := \int_{-\infty}^{\infty} (|g(y + i\mathsf{d})| + |g(y - i\mathsf{d})|) \, dy < \infty.$$

Note that the above definition is stronger than that used in [48]. By [48, Theorem 2.20], we have the error estimate for the quadrature scheme (II.21):

$$\left| \int_{-\infty}^{\infty} f(x) \, dx - k \sum_{j=-\infty}^{\infty} f(jk) \right| \leq \frac{N(B_{\mathsf{d}})}{e^{2\pi\mathsf{d}/k} - 1}. \tag{II.23}$$

To further obtain the error estimate between (II.21) and (II.22), we need the exponentially decay estimates of the function f. We will provide the full error estimate when f is explicitly defined.

II.6 Approximation of Negative Powers of Elliptic Operators

Let L be the unbounded operator defined in Section II.3. Given $f \in L^2(\Omega)$ and $s \in (0, 1)$, we consider the stationary problem (I.4). The solution is $u = L^{-s} f$. In this section, we briefly overview the finite element approximation of u from [12, 13]. We shall follow the ideas provided in this section and consider the approximation of a time dependent problem in the next chapter.

II.6.1 Space Discretization

We first apply the Balakrishman formula (II.9) to the negative power of L_h. That is

$$u_h := L_h^{-s} \pi_h f = \frac{\sin(\pi s)}{\pi} \int_0^{\infty} \mu^{-s} w_h(t) \, d\mu. \tag{II.24}$$

where $w_h(t) = (\mu I + L_h)^{-1} \pi_h f$ solves

$$\mu(w_h, \phi_h)_\Omega + d_\Omega(w_h, \phi_h) = (f, \phi_h)_\Omega, \qquad \text{for all } \phi_h \in \mathbb{V}_h(\Omega).$$

The theorem below provides the $L^2(\Omega)$ error estimate between u and u_h. The idea of the proof is to bound the error with the Balakrishman formula and utilize the the finite element error estimate (II.17) together with (II.16), Lemma II.1, Lemma II.3 and the norm equivalency results (Proposition II.2 and (II.18)).

Theorem II.6 (Theorem 4.3 of [12])**.** *Suppose that u is the solution of* (I.4) *and u_h is finite element approximation given by* (II.24)*. Let Assumption II.1 holds for some $\alpha \in (0,1]$. Set $\gamma = \alpha - s$ when $\alpha \geq s$ and $\gamma = 0$ when $\alpha < s$. For $f \in \mathbb{H}^{2\delta}(\Omega)$ with $\delta \geq \gamma$, there exists a constant C uniform in h such that*

$$\|u - u_h\|_{L^2(\Omega)} \leq C_h h^{2\alpha} \|f\|_{\mathbb{H}^{2\delta}(\Omega)}. \tag{II.25}$$

Here

$$C_h = \begin{cases} C\ln(1/h): & \text{if } \delta = \gamma \text{ and } \alpha \geq s, \\ C \quad : & \text{if } \delta > \gamma \text{ and } \alpha \geq s, \\ C \quad : & \text{if } s > \alpha. \end{cases}$$

II.6.2 Sinc Quadrature

We further approximate u_h by discretizing the Dunford-Taylor integral (II.24) with a sinc quadrature scheme. To this end, we use the change of variable $t = e^y$ and apply the formula (II.22), i.e. given the quadrature spacing $k > 0$ and two positive integers N^-, N^+ depending on k, the approximation of u_h is

$$u_h^k := \frac{k\sin(\pi s)}{\pi} \sum_{j=-N^-}^{N^+} e^{(1-s)y_j} w_h(t_j) \tag{II.26}$$

with $y_j = kj$ and $t_j = e^{y_j}$.

We follow Section II.5 to prove the exponential convergence of u_h^k. The idea of the proof will be applied in Section III.3, IV.2 and IV.3. So we sketch the proof and refer to [12, Theorem 3.5] for more details.

We expand u_h^k in the discrete eigenvector basis and get

$$\begin{aligned}\|u_h^k - u_h\|_{L^2(\Omega)}^2 &= \frac{\sin(\pi s)}{\pi}\sum_{j=1}^{M_h} |\mathcal{E}(\lambda_{j,h})|^2 |(\pi_h f, \psi_{j,h})_\Omega|^2 \\ &\le \frac{\sin(\pi s)}{\pi}\max_{j\in\{1,\ldots,M_h\}} |\mathcal{E}(\lambda_{j,h})|^2 \|f\|_{L^2(\Omega)}^2,\end{aligned} \tag{II.27}$$

where

$$\mathcal{E}(\lambda) := \int_{-\infty}^{\infty} g_\lambda(y)\, dy - k\sum_{j=-N^-}^{N^+} g_\lambda(jk)$$

with

$$g_\lambda(y) = e^{(1-s)y}(e^y+\lambda)^{-1}$$

Thus, we can focus on scaler case, namely, proving the exponential decay of $\mathcal{E}(\lambda)$ for all $\lambda \ge \lambda_1$ (recalling that λ_1 is the smallest eigenvalue of L). In fact, letting $\mathsf{d} = \pi/2$, we can show that g_λ satisfies the decay property

$$|g_\lambda(z)| \le \begin{cases} e^{-s\mathfrak{Re} z} : & \text{for } \mathfrak{Re} z > 0, \\ \dfrac{1}{\lambda_1} e^{(1-s)\mathfrak{Re} z} : & \text{for } \mathfrak{Re} z \le 0, \end{cases}$$

for all $z \in B_{\mathsf{d}}$ and $\lambda \ge \lambda_1$. Apply the decay property yields $g_\lambda \in S(B_{\mathsf{d}})$ and hence the error bound (II.23) holds when f is replaced by g_λ. Also, the decay estimate implies that

$$k\sum_{j=-N^--1}^{-\infty} |g_\lambda(jk)| \le \sum_{j=-N^--1}^{\infty} \frac{1}{\lambda_1} e^{(1-s)jk} \le \frac{1}{\lambda_1}\int_{-\infty}^{-kN^-} e^{(1-s)y}\, dy \le \frac{1}{(1-s)\lambda_1} e^{-(1-s)N^-k}.$$

Similarly,

$$k\sum_{j=N^++1}^{\infty} |g_\lambda(jk)| \le \frac{1}{s} e^{-sN^+k}.$$

Combing above two estimates together with (II.23) yields the full error estimate of the sinc approximation.

Theorem II.7 (Theorem 3.5 of [12]). *Let u_h and u_h^k be defined by* (II.24) *and* (II.26)*, respectively.*

Let $\mathsf{d} = \pi/2$ *and* $S(B_{\mathsf{d}})$ *be defined by Definition II.1. Then, there holds*

$$\|u_h - u_h^k\|_{L^2(\Omega)} \leq \frac{\sin(\pi s)}{\pi}\Bigg(\frac{N(B_{\mathsf{d}})}{e^{\pi^2/k}-1} + \frac{1}{s}e^{-sN^+k} + \frac{1}{(1-s)\lambda_1}e^{-(1-s)N^-k}\Bigg). \tag{II.28}$$

Here the constant $N(B_{\mathsf{d}})$ *is defined as in part (c) of Definition II.1.*

II.6.3 Choice of Parameters

In practice, we advocate to balance the three exponentials on the right hand side of (II.28), thereby imposing

$$\pi^2/k \approx sN^+k \approx (1-s)N^-k.$$

Thus, given $k > 0$, we choose

$$N^+ = \left\lceil \frac{\pi^2}{sk^2} \right\rceil, \text{ and } N^- = \left\lceil \frac{\pi^2}{(1-s)k^2} \right\rceil, \tag{II.29}$$

which leads to

$$\|u_h - u_h^k\|_{L^2(\Omega)} \preceq e^{-\pi^2/k}. \tag{II.30}$$

Combing the right hand side of (II.25) and (II.30) gives the $L^2(\Omega)$ estimate of the error between the solution u and its final approximation u_h^k. We can take the advantage of the exponential decay (II.30) by letting k as a function of h. Without the knowledge of the regularity of the right hand side data, we set

$$h^2 = e^{-\pi^2/k} \Rightarrow k = \frac{\pi^2}{\ln(h^{-1})}.$$

Now we let $U_h := u_h^k$ with k, N^+, N^- defined as above. According to (II.26), we shall compute U_h by solving $(e^{y_j}I + L_h)^{-1}\pi_h f$ independently. Totally there are $N^- + N^+ + 1 \sim O((\ln h)^2)$ equations need to be solved.

CHAPTER III

APPROXIMATION OF THE HOMOGENEOUS PARABOLIC PROBLEM *

Let us consider the numerical approximation of time dependent problems involving fractional powers of elliptic operators. In this chapter we focus on the homogeneous parabolic problem: find $u : \Omega \times [0, \mathsf{T}] \to \mathbb{R}$ satisfying

$$\begin{cases} u_t + L^s u = 0, & \text{in } \Omega \times (0, \mathsf{T}], \\ u = v, & \text{on } \Omega \times \{0\}, \end{cases} \tag{III.1}$$

where $v \in L^2(\Omega)$, $s \in (0, 1)$ and L^s is defined by (II.5).

Here is the online of this chapter. We first provide the weak formulation of the homogenous problem and a Dunford-Taylor integral representation of the weak solution in Section III.1. In Section III.2, we show a $L^2(\Omega)$ error estimate between the solution $u(t)$ and its finite element approximation $u_h(t)$ for a fixed time t. Section III.3 discusses an exponentially convergent sinc quadrature scheme approximating $u_h(t)$. Numerical results are provided in Section III.4.

III.1 Integral Representation of the Solution

Given a final time $\mathsf{T} > 0$, the weak formulation of (III.1) is: find $u \in L^2(0, \mathsf{T}; \dot{H}^s(\Omega))$ with $u_t \in L^2(0, \mathsf{T}; \dot{H}^{-s}(\Omega))$ such that

$$\begin{cases} (u_t, \phi)_\Omega + A(u, \phi) = 0 \text{ for all } \phi \in \dot{H}^s(\Omega) \text{ and for a.e. } t \in (0, \mathsf{T}], \\ u(0) = v, \end{cases} \tag{III.2}$$

where the bilinear form $A(\cdot, \cdot)$ is given by (II.6). The standard analysis for time dependent parabolic equations, see for example [35, Section 7.1.2], implies the existence and uniqueness of a solution

*Chapter III is reprinted from "The approximation of parabolic equations involving fractional powers of elliptic operators", 2017, Journal of Computational and Applied Mathematics, 315, 32–48, Copyright [2017] by Elsevier.

of (III.2) (it is the limit of the partial sums below) given by

$$u(t) = \sum_{j=1}^{\infty} e^{-t\lambda_j^s} (v, \psi_j)_\Omega \psi_j. \tag{III.3}$$

We shall use the contour $\mathcal{C} := \mathcal{C}(r_0, \theta)$ given by (II.8) with $r_0 \in (0, \lambda_1)$ and $\theta = \pi/4$ to define the Dunford-Taylor integral representation of the solution. The Cauchy's theorem applied to the partial sums of (III.3) and the Bochner integrability of the Dunford-Taylor integral implies that

$$u(t) = e^{-tL^s} v := \frac{1}{2\pi i} \int_{\mathcal{C}} e^{-tz^s} R_z(L) v \, dz. \tag{III.4}$$

Here $R_z(L) = (zI - L)^{-1}$ and $z^s := e^{s \ln z}$ with the logarithm defined with branch cut along the negative real axis. Note that e^{-tz^s} is also well defined since the range of z^s is smaller than its preimage and hence avoids the branch cut.

Remark III.1. *We have to point out that it is sufficient to set $\theta \in (0, \pi/2]$ in* (II.8) *so that the integral* (III.4) *is finite. This is because when $z = re^{\pm i\theta}$ with $r \geq r_0$, $\mathfrak{Re}(e^{-tz^s}) = e^{-tr^s \cos(s\theta)}$. So we require $\theta \leq \pi/(2s)$ for all $s \in (0, 1)$ so that $\mathfrak{Re}(e^{-tz^s})$ decays exponentially as $r \to \infty$.*

Notations.

According to the above remark, we cannot deform the contour (II.8) and rewrite (III.4) as a real integral like (II.9). So we will directly discretize (III.4) over a more appropriate path (not $\mathcal{C}$) in the complex plane. Given a complex number z, the definition of $R_z(L) = (zI - L)^{-1}$ requires the Sobolev spaces of complex valued functions. As usual, a complex Sobolev space is just two copies of the real valued space. For instance, the complex version of $H_0^1(\Omega)$ is the set of functions $v + iw$ where v, w are real valued functions in $H_0^1(\Omega)$ with the norm $\|v + iw\|_{H_0^1(\Omega)} := (\|v\|_{H_0^1(\Omega)}^2 + \|w\|_{H_0^1(\Omega)}^2)^{1/2}$. For convenience, in the complex case, we shall use the same notations for the Sobolev spaces. We have to point out that the negative spaces are defined using bounded antilinear functionals. Also the $L^2(\Omega)$ inner product should be $(v, w) = \int_\Omega v\overline{w} \, dx$ with $\overline{w}$ denoting the conjugate of w and definitions of bilinear forms like $A(\cdot, \cdot)$ follows a similar change. Note that the properties of Sobolev spaces and discrete operators introduced in Chapter I.2 still hold in the complex setting.

In the rest of this chapter, all spaces are defined on Ω. So we omit the domain part in the notation of function spaces and inner-products. For example, we write $\dot{H}^s = \dot{H}^s(\Omega)$, $\mathbb{V}_h = \mathbb{V}_h(\Omega)$ and $(\cdot,\cdot) = (\cdot,\cdot)_\Omega$.

III.2 Finite Element Approximation

III.2.1 The Approximation Scheme

The finite element approximation of (III.2) considered here reads: find $u_h \in H^1(0,\mathsf{T};\mathbb{V}_h)$ such that

$$\begin{cases} (u_{h,t}(t),\phi_h) + A_h(u_h(t),\phi_h) = 0, \quad \text{for } t \in (0,\mathsf{T}],\ \text{and } \phi_h \in \mathbb{V}_h, \text{ and} \\ u_h(0) = v_h := \pi_h v. \end{cases} \tag{III.5}$$

Here $u_{h,t}$ is the partial derivative of u_h with respect to t and $A_h(\cdot,\cdot)$ is given by (II.14). Similar to (III.4), the solution of (III.5) is given by

$$u_h(t) = e^{-tL_h^s} v_h := \sum_{j=1}^{M_h} e^{-t\lambda_{j,h}^s}(v_h,\psi_{j,h})\psi_{j,h} = \frac{1}{2\pi i}\int_{\mathcal{C}} e^{-tz^s} R_z(L_h) v_h\, dz. \tag{III.6}$$

III.2.2 Error estimates

We shall adopt the same procedure as we mentioned in Section II.6.1 to estimate the error between $u(t)$ and $u_h(t)$. To this end, we need to update Lemma II.1 and Lemma II.3 in the complex setting.

Lemma III.1. *Let T and T_h defined by* (II.1) *and* (II.12)*, respectively. Then there is a positive constant C not depending on $z \in \mathcal{C}$ and $f \in L^2$ such that*

$$|z|^{-s}\|T^{1-s}(z^{-1}I - T)^{-1} f\|_{L^2} \le C\|f\|_{L^2}. \tag{III.7}$$

The same inequality holds on $\mathbb{V}_h$ with T replaced by T_h.

Proof. $\boxed{1}$ We first show that

$$|\mu_j - z^{-1}|^{-1} \leq C(\mu_j + |z|^{-1})^{-1}, \quad \text{for } z \in \mathcal{C}, \tag{III.8}$$

with $\{\mu_j\}$ the set of non-increasing eigenvalues of T. According to the definition of $\mathcal{C}$ in (II.8), we shall bound the left hand side of (III.8) for $z \in \mathcal{C}_i$ for $i = 1, 2, 3$.

For $z \in \mathcal{C}_2$, the triangle inequality implies that

$$r_0^{-1} - \mu_1 \leq r_0^{-1} - \mu_j \leq |z^{-1} - \mu_j|.$$

Also,

$$|z|^{-1} + \mu_j \leq 2r_0^{-1}.$$

So

$$\frac{r_0^{-1} - \mu_1}{2r_0^{-1}}(|z|^{-1} + \mu_j) \leq |z^{-1} - \mu_j|. \tag{III.9}$$

If $z \in \mathcal{C}_3$, then z^{-1} is on the line segment connecting 0 to $r_0^{-1}e^{i\pi/4}$; see Figure III.1. It follows that $|z^{-1} - \mu_j| \geq \mathfrak{Im}(z^{-1}) = |z|^{-1}/\sqrt{2}$. Also, $|z^{-1} - \mu_j|$ is greater than or equal to the distance from μ_j to the line segment, i.e., $|z^{-1} - \mu_j| \geq \mu_j/\sqrt{2}$. The same inequalities hold for $\mathcal{C}_1$ and hence

$$|\mu_j - z^{-1}| \geq \frac{\sqrt{2}}{4}(\mu_j + |z|^{-1}), \qquad \text{for all } z \in \mathcal{C}_1 \cup \mathcal{C}_3. \tag{III.10}$$

Combining (III.9) and (III.10) completes the proof of (III.8).

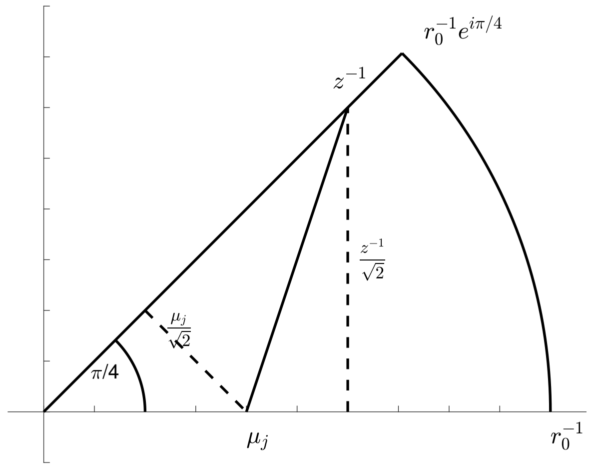

Figure III.1: Plot of $z^{-1} - \mu_j$ for $z \in \mathcal{C}_3$.

2 Now we are ready to show (III.7). As the proof of the continuous and discrete cases are essentially identical, we provide the proof for the former. Expanding the square of left hand side of (III.7) gives

$$\mathcal{W} := |z|^{-2s} \|T^{1-s}(z^{-1}I - T)^{-1}f\|_{L^2}^2 = \sum_{j=1}^{\infty} \left(\frac{|z|^{-s}\mu_j^{1-s}}{|z^{-1} - \mu_j|} \right)^2 |(f, \psi_j)|^2. \tag{III.11}$$

Thus, invoking (III.8) yields

$$\mathcal{W} \le C \sum_{j=1}^{\infty} \left(\frac{|z|^{-s}\mu_j^{1-s}}{|z|^{-1} + \mu_j} \right)^2 |(f, \psi_j)|^2.$$

The Young's inequality yields

$$\frac{|z|^{-s}\mu_j^{1-s}}{|z|^{-1} + \mu_j} \le \frac{s|z|^{-1} + (1-s)\mu_j}{|z|^{-1} + \mu_j} \le 1$$

so that combining the above two inequalities gives

$$\mathcal{W} \le C\|f\|_{L^2}^2$$

and completes the proof of the lemma. $\square$

Remark III.2. *We note that the constant C in* (III.7) *only depends on r_0 and λ_1 by* (III.9). *In fact, since we can choose any $r_0 \in (0, \lambda_1)$, we can avoid the dependencies by choosing $r_0 = \lambda_1/2$.*

The following lemma provides the estimate between $\pi_h R_z(L)v$ and $R_z(L_h)\pi_h v$ for $z \in \mathcal{C}$. This is the crucial step for the error analysis.

Lemma III.2 (Approximation of the resolvent)**.** *Let Assumption II.1 hold for some $\alpha \in (0,1]$. Then for $z \in \mathcal{C}$ and $v \in \dot{H}^{2\delta}$ with $\delta \in [0,(1+\alpha)/2]$, we have*

$$\|(\pi_h R_z(L) - R_z(L_h)\pi_h)v\|_{L^2} \le C|z|^{-1+\alpha-\delta}h^{2\alpha}\|v\|_{\dot{H}^{2\delta}}. \tag{III.12}$$

with C independent of z, v and h.

Proof. $\boxed{1}$ Noting that $R_z(L) = (zI-L)^{-1} = T(zT-I)^{-1}$ and $R_z(L_h)\pi_h = (zI-L_h)^{-1}\pi_h = (zT_h - I)^{-1}T_h\pi_h = (zT_h-I)^{-1}T_h$, we obtain

$$\begin{aligned}
\pi_h R_z(L) - R_z(L_h)\pi_h &= \pi_h(R_z(L) - R_z(L_h)\pi_h) \\
&= \pi_h[T(zT-I)^{-1} - (zT_h-I)^{-1}T_h] \\
&= \pi_h(zT_h-I)^{-1}[(zT_h-I)T - T_h(zT-I)](zT-I)^{-1} \\
&= \pi_h(zT_h-I)^{-1}(T_h-T)(zT-I)^{-1} \\
&= -(zT_h-I)^{-1}\pi_h(T-T_h)(zT-I)^{-1},
\end{aligned}$$

where for the last step we used the fact that $\pi_h(zT_h - I)^{-1} = (zT_h - I)^{-1}\pi_h$. Let

$$\begin{aligned} W(z) &:= |z|^{1-\alpha+\delta}(zT_h - I)^{-1}\pi_h(T - T_h)(zT - I)^{-1} \\ &= |z|^{-1-\alpha+\delta}(T_h - z^{-1}I)^{-1}\pi_h(T - T_h)(T - z^{-1}I)^{-1}. \end{aligned}$$

To complete the proof, it suffices to show that

$$\|W(z)\|_{\dot{H}^{2\delta}\to L^2} \le Ch^{2\alpha}. \tag{III.13}$$

Notice that

$$\begin{aligned} \|W(z)\| \le & \underbrace{\|z^{-(1+\alpha)/2}(T_h - z^{-1}I)^{-1}\pi_h\|_{\dot{H}^{1-\alpha}\to L^2}}_{=:\text{I}} \underbrace{\|(T - T_h)\|_{\dot{H}^{\alpha-1}\to \dot{H}^{1-\alpha}}}_{=:\text{II}} \\ & \underbrace{\|z^{-(1+\alpha)/2+\delta}(T - z^{-1}I)^{-1}\|_{\dot{H}^{2\delta}\to \dot{H}^{\alpha-1}}}_{=:\text{III}}. \end{aligned} \tag{III.14}$$

Now, we estimate the three terms on the right hand side above separately. For III, we have

$$\begin{aligned} \|(T - z^{-1})^{-1}\|_{\dot{H}^{2\delta}\to\dot{H}^{\alpha-1}} &= \sup_{w\in\dot{H}^{2\delta}} \frac{\|T^{(1-\alpha)/2}(T - z^{-1}I)^{-1}w\|_{L^2}}{\|L^{\delta}w\|_{L^2}} \\ &= \sup_{\theta\in L^2} \frac{\|T^{(1-\alpha)/2}(T - z^{-1}I)^{-1}T^{\delta}\theta\|_{L^2}}{\|\theta\|_{L^2}} \\ &= \|T^{1-[(1+\alpha)/2-\delta]}(T - z^{-1}I)^{-1}\|_{L^2}. \end{aligned}$$

Applying Lemma III.7 for $s := (1+\alpha)/2 - \delta \in [0,1]$ to obtain

$$\text{III} = \|z^{-(1+\alpha)/2+\delta}(T - z^{-1}I)^{-1}\|_{\dot{H}^{2\delta}\to\dot{H}^{\alpha-1}} \le C. \tag{III.15}$$

To estimate I, we invoke (II.19) and (II.18) to write

$$\begin{aligned}\|(T_h - z^{-1}I)^{-1}\pi_h\|_{\dot H^{1-\alpha}\to L^2} &\le C\|(T_h - z^{-1}I)^{-1}\pi_h\|_{\dot H_h^{1-\alpha}\to L^2}\\ &\le C\|(T_h - z^{-1}I)^{-1}\|_{\dot H_h^{1-\alpha}\to L^2}\|\pi_h\|_{\dot H^{1-\alpha}\to \dot H_h^{1-\alpha}}\\ &\le C\|(T_h - z^{-1}I)^{-1}\|_{\dot H_h^{1-\alpha}\to L^2}\\ &\le C\|T_h^{1-[(1+\alpha)/2]}(T_h - z^{-1}I)^{-1}\|_{L^2}.\end{aligned} \tag{III.16}$$

Since $(1+\alpha)/2 \in [0,1]$, applying Lemma III.7 again gives

$$\mathrm{I} = \|z^{-(1+\alpha)/2}(T_h - z^{-1})^{-1}\pi_h\|_{\dot H^{1-\alpha}\to L^2} \le C. \tag{III.17}$$

Combining (III.15), (III.17) and applying Proposition II.4 with $r = (1-\alpha)/2$ to estimate II yield (III.13). The proof is complete. □

Now we are in a position to show the main result of this section.

Theorem III.3 (Finite element approximation)**.** *Assume that Assumption II.1 holds for some $\alpha \in (0,1]$. Let $u(t)$ be the solution of the homogeneous problem* (III.2) *and let $u_h(t)$ be the finite element approximation given by* (III.5)*. Given the initial data $v \in \mathbb{H}^{2\delta}$ with $\delta \in [0,(1+\alpha)/2]$, there exists a constant $C(t)$ independent of h such that*

$$\|u(t) - u_h(t)\|_{L^2} \le C(t)h^{2\alpha}\|v\|_{\mathbb{H}^{2\delta}}, \tag{III.18}$$

where

$$C(t) = \begin{cases} C: & \text{if } \alpha < \delta,\\ C\max\{1,\ln(1/t)\}: & \text{if } \alpha = \delta,\\ Ct^{-(\alpha-\delta)/s}: & \text{if } \alpha > \delta.\end{cases} \tag{III.19}$$

Remark III.3. *If $\alpha < 1$, then the above theorem guarantees the rate of $h^{2\alpha}$ for all $\delta > \alpha$ without any degeneration as $t \to 0$. Note that the theorem only guarantees a rate of $\ln(1/t)h^2$ for small t,*

$\delta \geq 1$ *and* $\alpha = 1$. *In contrast, the classical analysis when* $s = \alpha = \delta = 1$ *[69] provides the rate* Ch^2 *(without the* $\ln(1/t)$ *for small* t*).*

Proof of Theorem III.3. By the norm equivalency mentioned in Remark II.1, we use the $\dot{H}^r$ norm instead of $\mathbb{H}^r$ norm for $r \in [-1, 1+\alpha]$ in the proof.

$\boxed{1}$ We shall analyze the error in two parts, i.e. by triangle inequality

$$\begin{aligned}\|u(t) - u_h(t)\|_{L^2} &= \|(e^{-tL^s} - e^{-tL_h^s}\pi_h)v\|_{L^2} \\ &\leq \|(I - \pi_h)e^{-tL^s}v\|_{L^2} + \|\pi_h(e^{-tL^s} - e^{-tL_h^s}\pi_h)v\|_{L^2}.\end{aligned}$$

In view of (II.16),

$$\|(I - \pi_h)e^{-tL^s}v\|_{L^2} \leq Ch^{2\alpha}\|e^{-tL^s}v\|_{\dot{H}^{2\alpha}}.$$

We bound the term on the right hand side by eigenfunction expansion setting $c_j = (v, \psi_j)$, the Fourier coefficient of v, and distinguish two cases. When $\delta \geq \alpha$, we use the representation (III.3) of $e^{-tL^s}v$ to write

$$\|e^{-tL^s}v\|_{\dot{H}^{2\alpha}}^2 = \sum_{j=1}^{\infty} \lambda_j^{2\alpha} \underbrace{e^{-2\lambda_j^s}}_{\leq 1} |c_j|^2 \leq \lambda_1^{2(\alpha-\delta)} \sum_{j=1}^{\infty} \lambda_j^{2\delta}|c_j|^2 = \lambda_1^{2(\alpha-\delta)}\|v\|_{\dot{H}^{2\delta}}^2.$$

Otherwise, when $\delta < \alpha$,

$$\|e^{-tL^s}v\|_{\dot{H}^{2\alpha}}^2 = t^{-2(\alpha-\delta)/s} \sum_{j=1}^{\infty} \lambda_j^{2\delta}|(t\lambda_j^s)^{(\alpha-\delta)/s}e^{-t\lambda_j^s}|^2|c_j|^2.$$

Using the fact that for $x \geq 0$ and $\eta = (\alpha - \delta)/s$, $x^\eta e^{-x} \leq C(\eta) = C$ and hence

$$\|e^{-tL^s}v\|_{\dot{H}^{2\alpha}}^2 \leq Ct^{-2(\alpha-\delta)/s}\|v\|_{\dot{H}^{2\delta}}^2.$$

$\boxed{2}$ Thus, we are left to bound

$$\|\pi_h(e^{-tL^s} - e^{-tL_h^s}\pi_h)v\|_{L^2}. \tag{III.20}$$

Since

$$\pi_h(e^{-tL^s} - e^{-tL_h^s}\pi_h)v = \frac{1}{2\pi i}\int_{\mathcal{C}} e^{-tz^s}\pi_h(R_z(L) - R_z(L_h)\pi_h)v\,dz,$$

we have

$$\|\pi_h(e^{-tL^s} - e^{-tL_h^s}\pi_h)v\|_{L^2} \leq \frac{1}{2\pi}\int_{\mathcal{C}} |e^{-tz^s}|\|\pi_h(R_z(L) - R_z(L_h)\pi_h)v\|_{L^2}\,d|z|.$$

Applying Lemma III.2 gives

$$\|\pi_h(e^{-tL^s} - e^{-tL_h^s}\pi_h)v\|_{L^2} \leq Ch^{2\alpha}\|v\|_{\dot{H}^{2\delta}}\int_{\mathcal{C}} |e^{-tz^s}||z|^{-1+\alpha-\delta}\,d|z|.$$

3 It remains to show that for a constant $C(t)$ independent of h,

$$\int_{\mathcal{C}} |e^{-tz^s}||z|^{-1+\alpha-\delta}\,d|z| \leq C(t). \tag{III.21}$$

Note that $|z| = r_0$ for $z \in \mathcal{C}_2$ and hence

$$\int_{\mathcal{C}_2} |e^{-tz^s}||z|^{-1+\alpha-\delta}\,d|z| \leq C.$$

For the remaining part of the contour, we have

$$\mathscr{I}_1 := \int_{\mathcal{C}_1\cup\mathcal{C}_3} |e^{-tz^s}||z|^{-1+\alpha-\delta}\,d|z| = 2\int_{r_0}^{\infty} e^{-\cos(s\pi/4)tr^s} r^{-1+\alpha-\delta} dr. \tag{III.22}$$

If $\delta > \alpha$,

$$\mathscr{I}_1 \leq C\int_{r_0}^{\infty} r^{-1+\alpha-\delta}\,dr \leq C.$$

If $\delta < \alpha$, we use the change of variable $y = \cos(s\pi/4)tr^s$ to get

$$\mathscr{I}_1 = Ct^{(\delta-\alpha)/s}\int_{\cos(\pi s/4)tr_0^s}^{\infty} e^{-y}y^{-1+(\alpha-\delta)/s}\,dy \leq \frac{C}{s}t^{(\delta-\alpha)/s}\Gamma\Big(\frac{\alpha-\delta}{s}\Big)$$

where $\Gamma(x)$ is the Gamma function. Finally, when $\delta = \alpha$, the same change of variables yields

$$\mathscr{I}_1 = C \int_{\cos(\pi s/4)tr_0^s}^{\infty} e^{-y} y^{-1} \, dy.$$

If t is large such that $\cos(\pi s/4)tr_0^s \geq 1$, then

$$\mathscr{I}_1 \leq C \int_1^{\infty} e^{-y} y^{-1} \, dy \leq C \int_1^{\infty} e^{-y} \, dy = C/e.$$

Otherwise, splitting the integral gives

$$\begin{aligned} \mathscr{I}_1 &\leq C \int_{\cos(\pi s/4)tr_0^s}^{1} e^{-y} y^{-1} \, dy + C/e \\ &\leq C \int_{\cos(\pi s/4)tr_0^s}^{1} y^{-1} \, dy + C/e \leq C \max\{1, \ln(1/t)\}. \end{aligned}$$

This completes the proof of (III.21) and hence the theorem. □

III.3 Sinc Approximation

In this section, we develop an exponentially convergent quadrature approximation to (III.6).

III.3.1 The Sinc Method

For a real constant $b \in (0, \lambda_1/\sqrt{2})$, define

$$\gamma(z) := b(\cosh z + i \sinh z), \quad z \in \mathbb{C} \tag{III.23}$$

and set the contour

$$\widehat{\mathcal{C}} \; := \; \{\gamma(y) \; : y \in \mathbb{R}\}.$$

We replace $\mathcal{C}$ with $\widehat{\mathcal{C}}$ in (III.6) to obtain,

$$\begin{aligned} u_h = e^{-tL_h^s} v_h &= \frac{1}{2\pi i} \int_{\widehat{\mathcal{C}}} e^{-tz^s} R_z(L_h) v_h \, dz \\ &= \frac{1}{2\pi i} \int_{\infty}^{-\infty} e^{-t\gamma(y)^s} \gamma'(y) [(\gamma(y) I - L_h)^{-1} v_h] \, dy. \end{aligned} \tag{III.24}$$

The sinc approximation $Q_h^{N,k}(t)v_h$ to (III.6) with $2N+1$ quadrature points and quadrature spacing $k>0$ is defined by

$$Q_h^{N,k}(t)v_h := -\frac{k}{2\pi i}\sum_{j=-N}^{N} e^{-t\gamma(y_j)^s}\gamma'(y_j)[(\gamma(y_j)I-L_h)^{-1}v_h] \tag{III.25}$$

with $y_j := jk$.

III.3.2 Error of the Sinc Approximation and the Total Error

Following the idea from Section II.6.2, it suffices to bound the error

$$\mathcal{E}(\lambda,t) := \int_{-\infty}^{\infty} g_\lambda(y,t)\,dy - k\sum_{j=-N}^{N} g_\lambda(jk,t), \tag{III.26}$$

where

$$g_\lambda(z,t) = e^{-t\gamma(z)^s}(\gamma(z)-\lambda)^{-1}\gamma'(z), \qquad \text{for } z\in\mathbb{C}, t>0. \tag{III.27}$$

The lemma below (see the proof in the next section) guarantees the exponential decay of $\mathcal{E}(\lambda,t)$ uniformly in $\lambda\ge\lambda_1$. It uses the following notations for $b\in(0,\lambda_1/\sqrt{2})$, N an integer, $k>0$ as above and $\mathsf{d}\in(0,\pi/4)$

$$\begin{aligned} \kappa &:= \cos\left[s(\pi/4+\mathsf{d})\right]\left[\sqrt{2}b\sin(\pi/4-\mathsf{d})\right]^s, \\ N(\mathsf{d},t) &:= \max_{\lambda\ge\lambda_1}\left\{\int_{-\infty}^{\infty}|g_\lambda(y+i\mathsf{d},t)|+|g_\lambda(y-i\mathsf{d},t)|\,dy\right\}, \quad \text{and} \\ M(t) &:= (1+\mathcal{L}(\kappa t)), \qquad \text{where } \mathcal{L}(x) := 1+|\ln(1-e^{-x})|. \end{aligned} \tag{III.28}$$

Lemma III.4. *Let $\mathcal{E}(\lambda,t)$ be given by* (III.26) *for an integer N and $k>0$ and κ be defined by* (III.28)*. Then, there is a constant C not depending on t, h, k and N satisfying*

$$|\mathcal{E}(\lambda,t)| \le C\Big(\frac{N(\mathsf{d},t)}{e^{2\pi\mathsf{d}/k}-1}+\frac{M(t)\cosh(kN)}{\sinh(kN)}e^{-\kappa 2^{-s}te^{kNs}}\Big). \tag{III.29}$$

The function $N(\mathsf{d},t)$ is uniformly bounded when t is bounded away from zero and bounded by

$CM(t)$ as $t \to 0$.

Theorem III.5 (Sinc approximation)**.** *Let $u_h(t)$ be the finite element approximation defined by* (III.5)*. $Q_h^{N,k}(t)v_h$ defined by* (III.25) *is the sinc approximation of $u_h(t)$. Then there exists a constant C independent of t, h, k and N, satisfying*

$$\|u_h(t) - Q_h^{N,k}(t)v_h\|_{L^2} \le C\Big(\frac{N(\mathsf{d},t)}{e^{2\pi\mathsf{d}/k}-1} + \frac{M(t)\cosh(kN)}{\sinh(kN)}e^{-\kappa 2^{-s}te^{kNs}}\Big)\|v\|_{L^2}. \tag{III.30}$$

Given a positive constant c, assume that $kN \ge c$. Since $M(t) \preceq \max(1, \ln(1/t))$, we have

$$\|u_h(t) - Q_h^{N,k}(t)v_h\|_{L^2} \le C\max(1,\ln(1/t))\Big(e^{-2\pi\mathsf{d}/k} + e^{-\kappa 2^{-s}te^{kNs}}\Big)\|v\|_{L^2}.$$

Proof. We expand v_h in the discrete eigenvector basis $\{\psi_{j,h}\}$ and get

$$\begin{aligned}\|(e^{-tL_h^s} - Q_h^{N,k}(t))v_h\|_{L^2}^2 &= (2\pi)^{-2}\sum_{j=1}^{M_h}|\mathcal{E}(\lambda_{j,h},t)|^2|(v_h,\psi_{j,h})|^2 \\ &\le (2\pi)^{-2}\max_{j=1,\ldots,M_h}|\mathcal{E}(\lambda_{j,h},t)|^2\|v_h\|_{L^2}^2,\end{aligned}$$

Applying Lemma III.4 into the right hand side above and using the fact $\|v_h\|_{L^2} \le \|v\|_{L^2}$ complete the proof of the theorem. □

Remark III.4 (Rate of convergence)**.** *As in [46], we set $k := \ln N/(sN)$ for some $N > 1$. The mononicity of $\cosh x/\sinh x$ for positive x implies that*

$$\frac{\cosh(kN)}{\sinh(kN)} \le \frac{\cosh(\ln 2/s)}{\sinh(\ln 2/s)} \tag{III.31}$$

and hence for a fixed time t,

$$\|u_h(t) - Q_h^{N,k}(t)v_h\|_{L^2} \le C\left(\frac{N(\mathsf{d},t)}{e^{4s\pi\mathsf{d}N/\ln N}-1} + \frac{M(t)}{e^{\kappa t 2^{-s}N}}\right)\|v\|_{L^2} = O\left(e^{-CN/\ln N}\right). \tag{III.32}$$

Remark III.5 (A balanced scheme)**.** *When t is small, we attempt to balance the error coming from the two terms in* (III.30) *by setting*

$$\frac{2\pi \mathsf{d}}{k} \approx \kappa t 2^{-s} e^{sNk}.$$

To this end, given an integer $N > 0$, we define k to be the unique positive solution of

$$k e^{sNk} = \frac{2^{1+s}\pi \mathsf{d}}{\kappa t}.$$

It follows that for $t \leq 1$ and $N > 1$,

$$Nk\, e^{sNk} = \frac{2^{2+s} N\pi \mathsf{d}}{\kappa t} \geq \frac{2^{1+s}\pi \mathsf{d}}{\kappa}.$$

It follows that $Nk \geq \zeta$ with ζ being the root of

$$\zeta e^{s\zeta} = \frac{2^{1+s}\pi \mathsf{d}}{\kappa}$$

so that

$$\frac{\cosh(kN)}{\sinh(kN)} \leq \frac{\cosh(\zeta)}{\sinh(\zeta)}.$$

Thus,

$$\|u_h(t) - Q_h^{N,k}(t) v_h\|_{L^2} \leq C \max(1, \ln(1/t)) e^{-2\pi \mathsf{d}/k}.$$

We combine Theorem III.3 and Theorem III.5 with the balanced scheme from Remark III.5 to obtain the estimate of the total error.

Theorem III.6 (Total error)**.** *Let $u(t)$ be the solution of the problem* (III.2)*. Set $u_h^k = Q_h^{N,k}(t) v_h$ with N determined by k according to Remark III.5. Let Assumption II.1 hold for some $\alpha \in (0, 1]$. Then. given the initial data $v \in \mathbb{H}^{2\delta}$, there holds*

$$\|u - u_h^k\|_{L^2} \preceq (C(t) h^{2\alpha} + \max(1, \ln(1/t)) e^{-2\pi \mathsf{d}/k}) \|v\|_{\mathbb{H}^{2\delta}},$$

where $C(t)$ is given by (III.19).

III.3.3 Proof of Lemma III.4

We first mention a fundamental ingredient provided in [46].

Lemma III.7 ([46], Lemma 1)**.** *For $r, \gamma > 0$,*

$$\int_0^\infty e^{-\gamma \cosh(x)}\, dx \le \mathcal{L}(\gamma) \tag{III.33}$$

and

$$\int_r^\infty e^{-\gamma \cosh(x)}\, dx \le (1 + \mathcal{L}(\gamma)) e^{-\gamma \cosh r}, \tag{III.34}$$

with $\mathcal{L}(\gamma)$ defined as in (III.28).

Two properties of $g_\lambda(z,t)$ are required to show Lemma III.4: $g_\lambda(z,t) \in B_{\mathsf{d}}$ for some $\mathsf{d} > 0$ and $g_\lambda(z,t)$ decays exponentially as $\mathfrak{Re} z \to \pm\infty$. The following lemma provides the ingredients for showing these properties.

Lemma III.8. *Let $0 < \mathsf{d} < \pi/4$ and $\lambda > \lambda_1$, $\gamma(z)$ be defined by* (III.23) *and $B_{\mathsf{d}} = \{z \in \mathbb{C} : \ \mathfrak{Im}(z) < \mathsf{d}\}$. The following assertions hold.*

(a) There exists a constant $C > 0$ only depending on λ_1, b and d such that

$$|\gamma(z) - \lambda| \ge C \qquad \textit{for all } z \in \overline{B}_{\mathsf{d}}; \tag{III.35}$$

(b) There exists a constant $C > 0$ only depending on λ_1, b and d such that

$$|\gamma'(z)(\gamma(z) - \lambda)^{-1}| \le C \qquad \textit{for all } z \in B_{\mathsf{d}};$$

(c)

$$\mathfrak{Re}(\gamma(z)^s) \ge \kappa \cosh(\mathfrak{Re}(z))^s \ge \kappa 2^{-s} e^{s|\mathfrak{Re} z|} \qquad \textit{for all } z \in B_{\mathsf{d}}.$$

Proof. For $z \in \mathbb{C}$, we rewrite $\gamma(z)$ by

$$\begin{aligned}\mathfrak{Re}(\gamma(z)) &= \sqrt{2}b\cosh(\mathfrak{Re}(z))\sin\left(\frac{\pi}{4} - \mathfrak{Im}(z)\right) \quad \text{and} \\ \mathfrak{Im}(\gamma(z)) &= \sqrt{2}b\sinh(\mathfrak{Re}(z))\sin\left(\frac{\pi}{4} + \mathfrak{Im}(z)\right).\end{aligned} \tag{III.36}$$

In order to show part (a), let $y_0 > 0$ be any number such that $C_0 := \lambda_1 - \sqrt{2}b\cosh(y_0) > 0$. Then for $z \in \overline{B}_{\mathsf{d}}$ with $|\mathfrak{Re}(z)| \leq y_0$,

$$\begin{aligned}|\gamma(z) - \lambda| \geq |\mathfrak{Re}(\gamma(z) - \lambda)| &= |\sqrt{2}b\cosh(\mathfrak{Re}(z))\sin\left(\frac{\pi}{4} - \mathfrak{Im}(z)\right) - \lambda| \\ &\geq \lambda_1 - \sqrt{2}b\cosh(y_0) = C_0.\end{aligned} \tag{III.37}$$

On the other hand, if $z \in \overline{B}_{\mathsf{d}}$ with $|\mathfrak{Re}(z)| > y_0$,

$$\begin{aligned}|\gamma(z) - \lambda| \geq |\mathfrak{Im}(\gamma(z) - \lambda)| &= \sqrt{2}b|\sinh(\mathfrak{Re}(z))|\sin\left(\mathfrak{Im}(z) + \frac{\pi}{4}\right) \\ &\geq \sqrt{2}b\sinh(y_0)\sin\left(\frac{\pi}{4} - \mathsf{d}\right).\end{aligned} \tag{III.38}$$

Combing (III.37) and (III.38) proves parts (a).

To show part (b), note that

$$|\gamma'(z)| \leq |\mathfrak{Re}(\gamma'(z))| + |\mathfrak{Im}(\gamma'(z))| \leq 2\sqrt{2}b\cosh(\mathfrak{Re}(z)). \tag{III.39}$$

For $z \in \overline{B}_{\mathsf{d}}$ with $|\mathfrak{Re}(z)| \leq y_0$, (III.37) yields

$$|\gamma'(z)(\gamma(z) - \lambda)^{-1}| \leq \frac{2\sqrt{2}b\cosh(y_0)}{C_0}. \tag{III.40}$$

Similarly, for $z \in \overline{B}_{\mathsf{d}}$ with $|\mathfrak{Re}(z)| > y_0$, there holds

$$|\gamma'(z)(\gamma(z) - \lambda)^{-1}| \leq \frac{2\cosh(\mathfrak{Re}(z))}{|\sinh(\mathfrak{Re}(z))|\,\sin(\pi/4 - \mathsf{d})} \leq C, \tag{III.41}$$

where to derive the last inequality we used the fact

$$\frac{\cosh(x)}{|\sinh(x)|} \leq \left|1 + \frac{2}{e^{2y_0} - 1}\right|, \qquad \text{for } x \in \mathbb{R} \text{ with } |x| \geq y_0.$$

Hence (III.40) and (III.41) imply part (b).

To show part (c), we note that by (III.36),

$$\left|\frac{\mathfrak{Im}(\gamma(z))}{\mathfrak{Re}(\gamma(z))}\right| = \frac{|\sinh(\mathfrak{Re}(z))| \sin(\pi/4 + \mathfrak{Im}(z))}{\cosh(\mathfrak{Re}(z)) \sin(\pi/4 - \mathfrak{Im}(z))} \leq \tan\left(\frac{\pi}{4} + \mathsf{d}\right), \qquad \text{for all } z \in \overline{B}_{\mathsf{d}}.$$

Thus,

$$|\arg(\gamma(z))| \leq \frac{\pi}{4} + \mathsf{d}, \qquad \text{for all } z \in \overline{B}_{\mathsf{d}},$$

so that together with the observation $|\gamma(z)| \geq |\mathfrak{Re}(\gamma(z))|$ and (III.36), we arrive at

$$\begin{aligned}
\mathfrak{Re}(\gamma(z)^s) &= |\gamma(z)|^s \cos(s \arg(\gamma(z))) \\
&\geq \cos(s(\pi/4 + \mathsf{d}))|\gamma(z)|^s \\
&\geq \cos(s(\pi/4 + \mathsf{d}))|\mathfrak{Re}(\gamma(z))|^s \\
&\geq \kappa \cosh(\mathfrak{Re}(z))^s \geq \kappa 2^{-s} e^{s|\mathfrak{Re} z|}, \qquad \text{for all } z \in \overline{B}_{\mathsf{d}}.
\end{aligned}$$

□

Lemma III.9. *Let g_λ be defined by* (III.27).

(a) (Exponential decay) *There holds*

$$|g_\lambda(z,t)| \leq C e^{-t\kappa(\cosh \mathfrak{Re} z)^s}, \qquad \textit{for all } z \in \overline{B}_{\mathsf{d}} \textit{ and } \lambda \geq \lambda_1. \tag{III.42}$$

(b) $g_\lambda(y,t) \in S(B_{\mathsf{d}})$ *for all* $\lambda \geq \lambda_1$.

Proof. (a) follows from part (b) and (c) in Lemma III.8. We now show part (b) by verifying Definition II.1 for g_λ. Part (a) of Lemma III.8 implies part (a) of Definition II.1. Applying (III.42) yields

$$\int_{-\mathsf{d}}^{\mathsf{d}} |g_\lambda(y+i\eta,t)|\, d\eta \le 2\mathsf{d}Ce^{-t\kappa(\cosh y)^s} \le C, \qquad \text{for all } y \in RR,$$

and

$$\begin{aligned} N(\mathsf{d},t) &= \max_{\lambda\ge\lambda_1} \int_{-\infty}^{\infty} (|g_\lambda(y-i\mathsf{d},t)| + |g_\lambda(y+i\mathsf{d},t)|)\, dy \\ &\le C\int_0^\infty e^{-\kappa t(\cosh y)^s}\, dy \qquad \text{(III.43)} \\ &\le C\int_0^1 e^{-\kappa t(\cosh y)^s}\, dy \, + C\int_1^\infty e^{-\kappa t(\cosh y)^s}\, dy. \end{aligned}$$

The first integral on the right hand side above is bounded by 1. For the second, making the change of integration variable, $(\cosh y)^s = \cosh\eta$, gives

$$I_2 := \int_1^\infty e^{-\kappa t(\cosh y)^s}\, dy = \frac{1}{s}\int_{\eta_0}^\infty e^{-\kappa t\cosh\eta}\frac{\sinh\eta\cosh y}{\cosh\eta\sinh y}\, d\eta \qquad \text{(III.44)}$$

where $\eta_0 = \cosh^{-1}[(\cosh(1))^s]$. As $\cosh(y)/\sinh(y)$ is decreasing for positive y and $\sinh(\eta)/\cosh(\eta) < 1$ for positive u, applying Lemma III.7 gives

$$I_2 \le \frac{\cosh(1)}{s\sinh(1)}\int_{\eta_0}^\infty e^{-\kappa t\cosh\eta}\, d\eta \le \frac{\cosh(1)}{s\sinh(1)}(1+\mathcal{L}(\kappa t))e^{-\kappa t\cosh(1)^s}. \qquad \text{(III.45)}$$

Combining this with the bound for the first integral of the right hand side of (III.43) proves (c) □

Proof of Lemma III.4. We split the error $\mathcal{E}(\lambda,t)$ in two parts, i.e.

$$|\mathcal{E}(\lambda,t)| \le \left|\int_{-\infty}^\infty g_\lambda(y,t)\, dy - k\sum_{j=-\infty}^{\infty} g_\lambda(jk,t)\right| + k\sum_{|j|>N}|g_\lambda(jk,t)|. \qquad \text{(III.46)}$$

In view of Lemma III.9, part (b), we invoke (II.23) to get

$$J_1 := \left|\int_{-\infty}^\infty g_\lambda(y,t)\, dy - k\sum_{j=-\infty}^{\infty} g_\lambda(jk,t)\right| \le \frac{N(\mathsf{d},t)}{e^{2\pi\mathsf{d}/k}-1}.$$

For the second term of (III.46), we take the advantage of (III.42) to derive that

$$J_2 := \left| k \sum_{|j|>N} g_\lambda(jk,t) \right| \leq Ck \sum_{|j|>N} e^{-t\kappa \cosh(jk)^s} \leq C \int_{kN}^{\infty} e^{-t\kappa \cosh(y)^s}\, dy. \tag{III.47}$$

Repeating the arguments in (III.44) and (III.45) (with $u_0 := \cosh^{-1}(\cosh(kN)^s)$) gives

$$\begin{aligned} J_2 &\leq \frac{C\cosh(kN)}{\sinh(kN)}(1+\mathcal{L}(\kappa t))e^{-\kappa t \cosh(kN)^s} \\ &\leq \frac{C\cosh(kN)}{\sinh(kN)}(1+\mathcal{L}(\kappa t))e^{-\kappa 2^{-s} t e^{kNs}}. \end{aligned}$$

This completes the estimate for the second term of (III.46) and proof. □

III.4 Numerical Illustration

In this section, we present some numerical experiments illustrating the error estimates provided in Section III.2 and Section III.3.

III.4.1 Error from the Finite Element Approximation

Consider the one-dimensional domain $\Omega := (0,1)$. Let L be the unbounded operator associated with the form $d(u,v) = \int_0^1 u'v'\, dx$ and set the initial data

$$v(x) := \begin{cases} 2x, & x < 0.5, \\ 2-2x, & x \geq 0.5. \end{cases} \tag{III.48}$$

To illustrate the error behavior predicted by Theorem III.3, we use a mesh of equally spaced points, i.e., $h = 1/(M+1)$ with M being the number of interior nodes. We set $\mathbb{V}_h$ to be the set of continuous piecewise linear functions surbodinate this mesh vanishing at 0 and 1. The resulting stiffness $\widetilde{A}_h$ and mass $\widetilde{M}_h$ matrices defined in terms of the standard hat-function finite element basis $\{\phi_i\}$, $i = 1, \ldots, M$ are tri-diagonal matrices with (tri-)diagonal entries $h^{-1}(-1,2,-1)$ and $h(1/6,4/6,1/6)$, respectively. The matrix form of L_h is given by $\widetilde{L}_h = \widetilde{M}_h^{-1}\widetilde{A}_h$.

These matrices can be diagonalized using the discrete sine transform, i.e., the $M \times M$ matrix with entries $S_{ij} := \sqrt{2h}\sin(ij\pi h)$. Thus, the matrix representation of L_h can be rewritten as $\widetilde{L}_h = S^{-1}\Lambda S$ with Λ denoting the diagonal matrix with diagonal $\Lambda_{ii} = 6h^{-2}(1-\cos(i\pi h))/(2+\cos(i\pi h))$, $i = 1\ldots, M$. The matrix $\widetilde{L}_h$ takes coefficients of a function $w_h \in \mathbb{V}_h$ to those of $L_h w_h$. Let $\widetilde{V}$ be the vector in $\mathbb{R}^M$ defined by

$$\widetilde{V}_j = (v, \phi_j), \quad j = 1, \ldots, M.$$

Then, the matrix representing $e^{-tL_h^s} w_h$ is thus given by $S^{-1}e^{-t\Lambda^s}S$ so the vector of coefficients representing $u_h(t) = e^{-tL_h^s}\pi_h v$ is given by

$$S^{-1}\mathcal{D}(t)S\widetilde{V}$$

where $\mathcal{D}(t)$ is the diagonal matrix with diagonal entries

$$\mathcal{D}(t)_{ii} = \frac{3e^{-t\Lambda_{ii}^s}}{h(2+\cos(i\pi h))}, \quad i = 1, \ldots, M.$$

The action of S on a vector can be efficiently computed using the Fast Fourier Transform in $O(M \ln M)$ operations and $S^{-1} = S$.

Note that v of (III.48) belongs to $\dot{H}^{3/2-2\epsilon}(0,1)$ for any $\epsilon > 0$. Theorem III.3 with $\alpha = 1$ guarantees

$$\|u(t) - u_h(t)\|_{L^2} \leq Ct^{-(1/4+\epsilon)/s}h^2.$$

To compute the solution $u(t)$ at the finite element nodes, the exact solution u is approximated using the first 50000 modes of its Fourier representation. The number of modes is chosen large enough such that it does not influence the space discretization error.

Table III.1 reports the L^2 error $e_i := \|u(t) - u_{h_i}(t)\|_{L^2}$ and its observed rate of convergence

$$L^2\text{-OROC}_i := \frac{\ln(e_i/e_{i+1})}{\ln(h_i/h_{i+1})}$$

for different s at time $t = 0.5$. In all cases, we observed $\|u(t) - u_h(t)\|_{L^2} \sim h^2$ as predicted by Theorem III.3, see also (III.18).

h	$s = 0.25$		$s = 0.5$		$s = 0.75$	
$1/8$	1.13×10^{-3}		1.21×10^{-3}		9.41×10^{-4}	
$1/16$	3.08×10^{-4}	1.87	3.03×10^{-4}	2.00	2.37×10^{-4}	1.99
$1/32$	7.97×10^{-5}	1.95	7.58×10^{-5}	2.00	5.94×10^{-5}	1.99
$1/64$	2.01×10^{-5}	1.98	1.89×10^{-5}	2.00	1.48×10^{-5}	2.00
$1/128$	5.05×10^{-6}	2.00	4.74×10^{-6}	2.00	3.71×10^{-6}	2.00

Table III.1: L^2 errors and observed rate of convergence (L^2-OROC) for different values of s. The observed error decay is in accordance with Theorem III.3.

III.4.2 Error from the Sinc Quadrature

We now focus on the quadrature error estimate given in Theorem III.5 and study the two different relations between k and N discussed in Remark III.4 and Remark III.5. Here we consider the scalar case to verify Lemma III.4. To do this we introduce an approximation to $\|\mathcal{E}(\cdot, t)\|_{L^\infty(10,\infty)}$ defined by the following procedure.

(I) We examine the value of $|\mathcal{E}(\lambda, t)|$ for $\lambda_j = 10\mu^j$ for $j = 0, 1, \cdots, \mathcal{N}$. Here $\mu > 1$ and $\mathcal{N}$ is chosen sufficiently large so that $|\mathcal{E}(\lambda, t)|$ is monotonically decreasing when $\lambda \geq \lambda_{\mathcal{N}}$ (for $t = .5$, we take $\mu = 3/2$ and $\mathcal{N} = 40$).

(II) We set $k := \operatorname{argmax}_{j=1,\ldots,\mathcal{N}}(|\mathcal{E}(\lambda_j, t)|)$ and approximate

$$\|\mathcal{E}(\cdot, t)\|_{L^\infty(10,\infty)} \approx \max_{l=1,\ldots,\mathcal{N}} |\mathcal{E}(\rho_l, 0.5)|, \qquad \text{where } \rho_l := \lambda_{k-1} + \frac{\lambda_{k+1} - \lambda_{k-1}}{\mathcal{N}} l.$$

By adjusting μ, $\mathcal{N}$ and l, we can obtain $\|\mathcal{E}(\cdot, t)\|_{L^\infty(10,\infty)}$ to any desired accuracy.

In Figures III.2 and III.3, we report values of $\|\mathcal{E}(\cdot,t)\|_{L^\infty(10,\infty)}$ as a function of N obtained by runing the above algorithm with $\mathcal{N}$ and μ adjusted so that the results are accurate to the number of digits reported. When considering Remark III.5, we choose $\mathsf{d} = \pi/8$.

For Figure III.2, we take $t = 0.5$. The blue lines give the results for Remark III.4 while the red lines give the results for Remark III.5. Except for the case of $s = .25$, the scheme in Remark III.5 is somewhat better.

For Figure III.3, we take $N = 32$ and report the errors as a function of t. In all cases, we see significant improvement from Remark III.4 to Remark III.5 when t is small. When considering Remark III.5, we choose $\mathsf{d} = \pi/8$ so that k can be computed as a function of N.

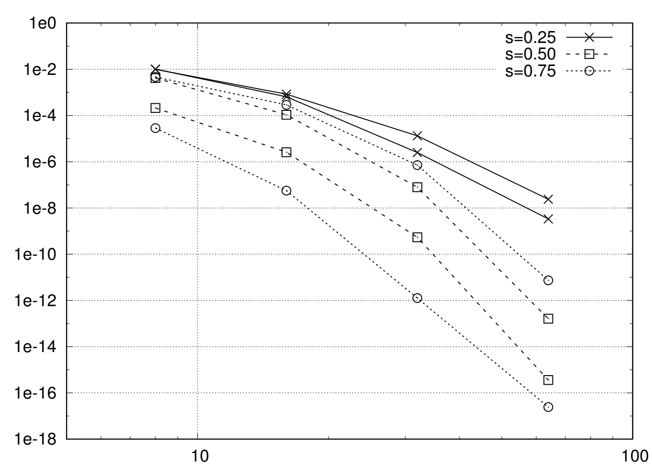

Figure III.2: $\|\mathcal{E}(\cdot,0.5)\|_{L^\infty(10,\infty)}$ as a function of N discussed in Remark III.4 and Remark III.5 reported in blue and red, respectively.

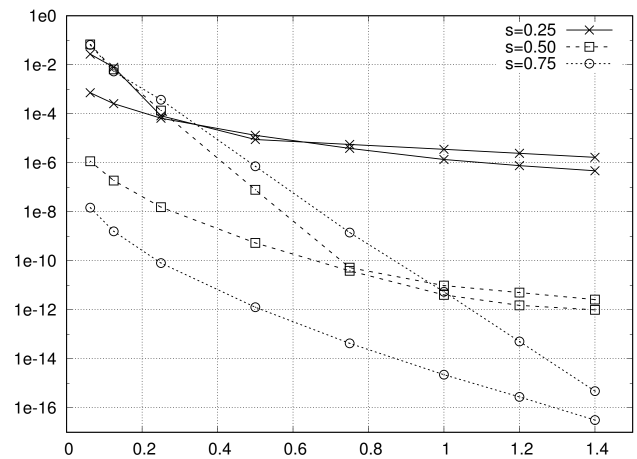

Figure III.3: $\|\mathcal{E}(\cdot,t)\|_{L^\infty(10,\infty)}$ with $N = 32$ as a function of t discussed in Remark III.4 and Remark III.5 reported in blue and red, respectively.

III.4.3 A Two Dimensional Problem

Set Ω to be the unit square $(0,1)^2$ and L is the unbounded operator associated with the form $d(u,v) = \int_\Omega \nabla u \cdot \nabla v \, dx$. The initial data is set to be the checkerboard function

$$v(x_1,x_2) = \begin{cases} 1 & \text{if } (x_1-0.5)(x_2-0.5) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{III.49}$$

Since we have $v \in \dot{H}^{1/2-2\epsilon}(\Omega)$ for all $\epsilon > 0$, Theorem III.3 guarantees an L^2 error decay

$$\|u(t) - u_h(t)\|_{L^2} \le Ct^{-(3/4+\epsilon)/s} h^2.$$

To approximate the solution, we use the scheme (III.25) with $N = 40$ and $k = \ln(N)/(sN)$ (the scheme in Remark III.4). Here we use *triangle* [63] to generate meshes such that each mesh is quasi-uniform and controlled by maximum area of cells. Approximation $Q_h^{N,k}(0.5)v_h$ for different values of s are provided in Figure III.4, thereby illustrating the effect of s on the diffusion strength. In addition, snapshots of $Q_h^{N,k}(t)v_h$ at different times t are provided in Figure III.5.

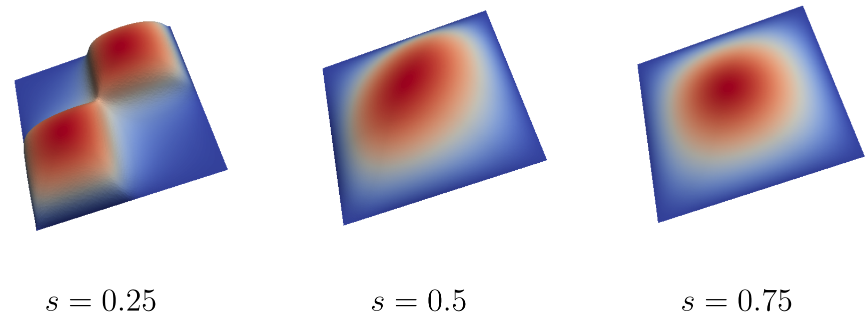

Figure III.4: Approximations $Q_h^{N,k}(0.5)v_h$ for initial data problem for different values of s. The diffusion process is faster when increasing s.

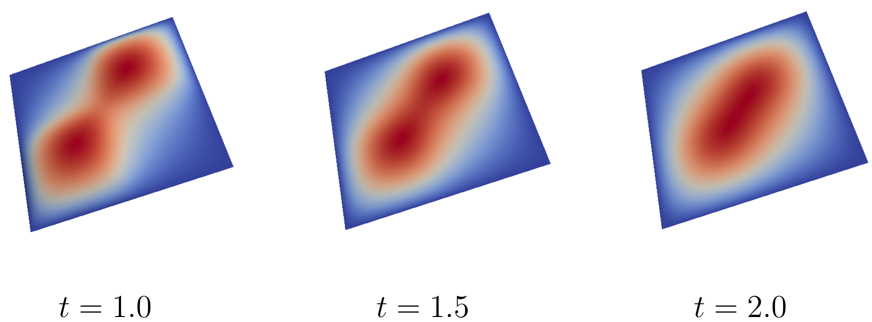

Figure III.5: Evaluation of the solution $Q_h^{N,k}(t)v_h$ at different time when $s = 0.25$.

Finally, the total approximation errors $\|Q_h^{N,k}(t)v_h - u(t)\|_{L^2}$ at $t = 0.5$ are reported in Table III.2 for different values of s. The optimal order 2 predicted by Theorem III.6 is obtained for large s, while the asymptotic regime for $s = 0.25$ was not reached in the computations.

h^2	$s = 0.25$		$s = 0.5$		$s = 0.75$	
0.02	2.38×10^{-2}		1.47×10^{-3}		6.11×10^{-4}	
0.005	6.20×10^{-3}	1.94	4.72×10^{-4}	1.64	1.66×10^{-4}	1.88
0.00125	1.59×10^{-3}	1.96	1.21×10^{-4}	1.96	4.32×10^{-5}	1.94
0.0003125	4.26×10^{-4}	1.90	3.17×10^{-5}	1.93	1.09×10^{-5}	1.99
0.000078125	1.13×10^{-4}	1.91	7.88×10^{-6}	2.01	2.73×10^{-6}	2.00

Table III.2: Total approximation error $\|Q_h^{N,k}(t) - u(t)\|_{L^2}$ at $t = 0.5$ and convergence rate for initial data (III.49) with different values of s. The optimal order 2 predicted by Theorem III.6 is obtained for large s, while the asymptotic regime for $s = 0.25$ was not reached in the computations.

CHAPTER IV

APPROXIMATION OF THE NON-HOMOGENEOUS PARABOLIC PROBLEM *

This chapter is continuation of the preceding chapter on the numerical approximation of the parabolic problem. So notations mentioned in Chapter III are still applied to this chapter. We now focus on the numerical approximation of the non-homogeneous problem: given a right hand data $f \in L^2(0,\mathsf{T};L^2(\Omega))$, we want to find $u : \Omega \times [0,\mathsf{T}] \to \mathbb{R}$ satisfying

$$\begin{cases} u_t + L^s u = f, & \text{in } \Omega \times (0,\mathsf{T}], \\ u = 0, & \text{on } \Omega \times \{0\}. \end{cases} \tag{IV.1}$$

The weak formulation reads: find $u \in L^2(0,\mathsf{T};\dot{H}^s(\Omega))$ with $u_t \in L^2(0,\mathsf{T};\dot{H}^{-s}(\Omega))$ such that

$$\begin{cases} (u_t,\phi)_\Omega + A(u,\phi) = (f(t),\phi) \text{ for all } \phi \in \dot{H}^s(\Omega) \text{ and for a.e. } t \in (0,\mathsf{T}], \\ u(0) = 0. \end{cases} \tag{IV.2}$$

By Duhamel's principle, the solution of the above problem is given by

$$u(t) = \int_0^t e^{-\xi L^s} f(t-\xi)\, d\xi. \tag{IV.3}$$

We again obtain the finite element approximation of (IV.1) by replacing L with L_h. The error analysis is provided in Section IV.1. The major task of this chapter, is the discretization in time. Here we consider two approaches. In Section IV.2, we discuss the approximation to the solution (IV.3) using a numerical quadrature scheme. In Section IV.3, we consider the time discretization of (IV.1) with the Crank-Nicolson time stepping scheme. Some numerical results are provided in Section IV.4 to support the error analysis for above discretization schemes.

---

*Section IV.1 and Section IV.4.1 are reprinted from "The approximation of parabolic equations involving fractional powers of elliptic operators", 2017, Journal of Computational and Applied Mathematics, 315, 32–48, Copyright [2017] by Elsevier.

IV.1 The Finite Element Approximation

The finite element approximation of (IV.2) reads: find $u_h \in H^1(0, \mathsf{T}; \mathbb{V}_h)$ such that

$$\begin{cases} (u_{h,t}, \phi_h) + A_h(u_h, \phi_h) = (f(t), \phi_h), & \forall \phi_h \in \mathbb{V}_h \text{ and a.e. } t \in (0, \mathsf{T}), \\ u_h(0) = 0, \end{cases} \tag{IV.4}$$

with the solution given by

$$u_h(t) = \int_0^t e^{-\xi L_h^s} \pi_h f(t - \xi)\, d\xi. \tag{IV.5}$$

If $f \in L^\infty(0, \mathsf{T}; \mathbb{H}^{2\delta})$, Theorem III.3 immediately implies the following L^2 error estimate

$$\begin{aligned} \|u(t) - u_h(t)\|_{L^2} &\le \int_0^t \|(e^{-\xi L^s} - e^{-\xi L_h^s} \pi_h) f(\xi)\|_{L^2}\, d\xi \\ &\le \int_0^t C(\xi) h^{2\alpha} \|f(t-\xi)\|_{\mathbb{H}^{2\delta}}\, d\xi \\ &\le C(h,t) \|f\|_{L^\infty(0,\mathsf{T};\mathbb{H}^{2\delta})}. \end{aligned}$$

Here $C(t)$ is given by (III.19) and

$$C(h,t) = h^{2\alpha} \int_0^t C(\xi)\, d\xi = \begin{cases} Cth^{2\alpha}: & \delta > \alpha, \\ Ct \max(1, \ln(1/t)) h^{2\alpha}: & \delta = \alpha, \\ Ch^{2\alpha} \int_0^t \xi^{-(\alpha-\delta)/s}\, d\xi: & \delta < \alpha. \end{cases} \tag{IV.6}$$

When $\delta < \alpha$, the above argument shows that the optimal convergence rate $2\widetilde{\alpha}$ is achieved provided that the integral $\int_0^t \xi^{-(\widetilde{\alpha}-\delta)/s}\, d\xi$ is finite, i.e. $\widetilde{\alpha} = s + \delta - \epsilon$ for every $\epsilon > 0$. We summarize the above discussion in the following theorem.

Theorem IV.1 (Finite element approximation)**.** *Suppose Assumption II.1 holds for some* $\alpha \in (0, 1]$. *Furthermore, we assume that* $f \in L^\infty(0, \mathsf{T}; \mathbb{H}^{2\delta})$ *with* $\delta \in [0, (1+\alpha)/2]$. *Let* u *be the solution of* (IV.2) *and* u_h *is the finite element approximation given by* (IV.4). *Let* $\widetilde{\alpha} = \min(\alpha, s + \delta - \epsilon)$ *with*

ϵ positive and sufficiently small. Then we have

$$\|u(t) - u_h(t)\|_{L^2} \le \widetilde{C}(t) h^{2\widetilde{\alpha}} \|f\|_{L^\infty(0,\mathsf{T};\mathbb{H}^{2\delta})}.$$

Here

$$\widetilde{C}(t) := \begin{cases} Ct: & \delta > \alpha, \\ Ct \max(1, \ln(1/t)): & \delta = \alpha, \\ Ct^{1-\frac{\widetilde{\alpha}-\delta}{s}}: & \delta < \alpha. \end{cases} \tag{IV.7}$$

IV.2 Time Discretization via Numerical Integration

IV.2.1 Discretization in Time

We now further approximate u_h given by (IV.5) by discretizing the integral with respect to ξ. To do this, we decompose the integral $[0, t]$ onto $\mathcal{N}$ subintervals

$$0 = t_0 < t_1 < \ldots < t_{\mathcal{N}-1} < t_{\mathcal{N}} = t,$$

where, for $\ell = 0, \ldots, \mathcal{N}$,

$$t_\ell := \ell\tau, \qquad \text{with } \tau = t/\mathcal{N}. \tag{IV.8}$$

On each subinterval we set $t_{\ell-\frac{1}{2}} = \frac{1}{2}(t_{\ell-1} + t_\ell)$ and propose the pseudo-midpoint approximation

$$\begin{aligned} \int_{t_{\ell-1}}^{t_\ell} e^{-\xi L_h^s} \pi_h f(t-\xi)\, d\xi &\approx \int_{t_{\ell-1}}^{t_\ell} e^{-\xi L_h^s}\, d\xi\, \pi_h f(t - t_{\ell-\frac{1}{2}}) \\ &= L_h^{-s}\big(e^{-t_{\ell-1}L_h^s} - e^{-t_\ell L_h^s}\big)\pi_h f(t - t_{\ell-\frac{1}{2}}) \\ &=: L_h^{-s}(W_h(t_{\ell-1}) - W_h(t_\ell)) g_h(t_{\ell-\frac{1}{2}}). \end{aligned}$$

Here we set $W_h(r) = e^{-rL_h^s}$ and $g_h(r) = \pi_h f(t-r)$. We further use the bar symbol to denote average quantities over the subinterval $[t_{\ell-1}, t_\ell]$, e.g.

$$\overline{W}_\ell : \mathbb{V}_h \to \mathbb{V}_h, \qquad \overline{W}_\ell := \frac{1}{\tau}\int_{t_{\ell-1}}^{t_\ell} W_h(\xi)\, d\xi.$$

and

$$\overline{g}_\ell := \frac{1}{\tau}\int_{t_{\ell-1}}^{t_\ell} g_h(\xi)\, d\xi.$$

Hence, by summing up the contributions from each subinterval, the time discretization of u_h becomes

$$u_h^{\mathcal{N}}(t) := \sum_{\ell=1}^{\mathcal{N}} L_h^{-s}(W_h(t_{\ell-1}) - W_h(t_\ell))g_h(t_{\ell-\frac{1}{2}}) = \tau \sum_{\ell=1}^{\mathcal{N}} \overline{W}_\ell g_h(t_{\ell-\frac{1}{2}}). \tag{IV.9}$$

Let us estimate the error between $u_h(t)$ and $u_h^{\mathcal{N}}(t)$ in the measure of L^2.

Theorem IV.2 (Time discretization)**.** *Assume that $g(r) = f(t-r)$ belongs to $H^2(0, \mathsf{T}; L^2)$. Then for $t \in [t_0, \mathsf{T}]$ with $t_0 > 0$, there exists a constant C independent of h and τ satisfying*

$$\|u_h(t) - u_h^{\mathcal{N}}(t)\|_{L^2} \le \tau^2((1+\sqrt{t})\|f_{tt}\|_{L^2(0,t;L^2)} + C\ln\mathcal{N}\|f_t\|_{L^\infty(0,t;L^2)}). \tag{IV.10}$$

where f_t and f_{tt} denote the first and second partial derivative in time of f.

Proof. $\boxed{1}$ We note that since $\|W_h(r)\|_{L^2\to L^2} \le 1$ for $r \in (0,t)$ and $\|\pi_h\|_{L^2\to L^2} \le 1$,

$$\begin{aligned}\left\| \int_0^{t_1} W_h(\xi)\pi_h(g(\xi) - g(t_{\frac{1}{2}}))\, d\xi \right\|_{L^2} &\le \int_0^{t_1} \|W_h(\xi)\|_{L^2\to L^2}\|(g(\xi) - g(t_{\frac{1}{2}}))\|_{L^2}\, d\xi \\ &\le \tau^2\|g_t\|_{L^\infty(0,T;L^2)}.\end{aligned}$$

Here we use the fact that

$$\|g(\xi) - g(t_{\frac{1}{2}})\|_{L^2} = \|\int_{t_{\frac{1}{2}}}^{\xi} g_t(\eta)\, d\eta\|_{L^2} \le |\xi - t_{\frac{1}{2}}|\|g_t\|_{L^\infty(0,t;L^2)} \le \tau\|g_t\|_{L^\infty(0,t;L^2)}.$$

For the rest of subintervals, namely $I_\ell = [t_{\ell-1}, t_\ell]$ with $\ell \geq 2$, we use the following decoposition

$$\begin{aligned}&\int_{t_{\ell-1}}^{t_\ell} W_h(\xi)(g_h(\xi) - g_h(t_{\ell-\frac{1}{2}}))\, d\xi \\ &\qquad = \underbrace{\tau \overline{W}_\ell \pi_h(\overline{g}_\ell - g(t_{\ell-\frac{1}{2}}))}_{=:E_1} + \underbrace{\int_{t_{\ell-1}}^{t_\ell} (W_h(\xi) - \overline{W}_\ell)\pi_h(g(\xi) - g(t_{\ell-\frac{1}{2}}))\, d\xi}_{=:E_2}\,.\end{aligned}$$

$\boxed{2}$ We estimate E_1 by

$$\|E_1\|_{L^2} \leq \tau \|\overline{W}_\ell \pi_h\|_{L^2 \to L^2} \|\overline{g}_\ell - g(t_{\ell-\frac{1}{2}})\|_{L^2}. \tag{IV.11}$$

We now bound $\|\overline{W}_\ell \pi_h\|_{L^2\to L^2}$ and $\|\overline{g}_\ell - g(t_{\ell-\frac{1}{2}})\|_{L^2}$ separately. For the latter, we expand $g(\eta)$ at $\eta = t_{\ell-\frac{1}{2}}$ to get

$$g(\eta) - g(t_{\ell-\frac{1}{2}}) = (\eta - t_{\ell-\frac{1}{2}}) g_t(t_{\ell-\frac{1}{2}}) + \int_{t_{\ell-\frac{1}{2}}}^{\eta} (r - t_{\ell-\frac{1}{2}}) g_{tt}(r)\, dr.$$

As a consequence, taking advantage of $t_{\ell-\frac{1}{2}}$ being the midpoint of the interval I_ℓ, we obtain

$$\begin{aligned}\overline{g}_\ell - g(t_{\ell-\frac{1}{2}}) &= \frac{1}{\tau}\int_{t_{\ell-1}}^{t_\ell} \big(g(\eta) - g(t_{\ell-\frac{1}{2}})\big)\, d\eta \\ &= \frac{1}{\tau}\int_{t_{\ell-1}}^{t_\ell}\int_{t_{\ell-\frac{1}{2}}}^{\eta} (r - t_{\ell-\frac{1}{2}}) g_{tt}(r)\, dr\, d\eta\end{aligned}$$

and hence using a Cauchy-Schwarz inequality

$$\|\overline{g}_\ell - g(t_{\ell-\frac{1}{2}})\|_{L^2} \leq \tau^{3/2} \|g_{tt}\|_{L^2(t_{\ell-1}, t_\ell; L^2)}. \tag{IV.12}$$

In order to bound $\|\overline{W}_\ell \pi_h\|_{L^2\to L^2}$, we note that form the definition of $W_h(r)$ and the stability of π_h, we derive that

$$\begin{aligned}\|\overline{W}_\ell \pi_h\|_{L^2\to L^2} &\le \frac{1}{\tau}\int_{t_{\ell-1}}^{t_\ell} \|e^{-\xi L_h^s}\pi_h\|_{L^2\to L^2}\,d\xi \\ &\le \frac{1}{\tau}\int_{t_{\ell-1}}^{t_\ell} d\xi = 1.\end{aligned} \tag{IV.13}$$

Estimates (IV.12) and (IV.13) into (IV.11) give the final bound for E_1:

$$\|E_1\|_{L^2} \le \tau^{5/2}\|g_{tt}\|_{L^2(t_{\ell-1},t_\ell;L^2)}. \tag{IV.14}$$

$\boxed{3}$ We estimate E_2

$$\|E_2\|_{L^2} \le \int_{t_{\ell-1}}^{t_\ell} \|(W_h(\xi)-\overline{W}_\ell)\pi_h\|_{L^2\to L^2}\|g(\xi)-g(t_{\ell-\frac{1}{2}})\|_{L^2}\,d\xi. \tag{IV.15}$$

In this case as well, we need to estimate two terms separately, namely $\|(W_h(\xi)-\overline{W}_\ell)\pi_h\|_{L^2\to L^2}$ and $\|g(\xi)-g(t_{\ell-\frac{1}{2}})\|_{L^2}$. For the latter, we write

$$\|g(\xi)-g(t_{\ell-\frac{1}{2}})\|_{L^2} = \|\int_{t_{\ell-\frac{1}{2}}}^{\xi} g_t(\eta)\,d\eta\|_{L^2} \le \tau\|g_t\|_{L^\infty(t_{\ell-1},t_\ell;L^2)}. \tag{IV.16}$$

Next, we bound $\|(W_h(\xi)-\overline{W}_\ell)\pi_h\|_{L^2\to L^2}$. As before, it suffices to estimate $\|W_h(\xi)-\overline{W}_\ell\|_{L^2\to L^2}$. To achieve this, we use set of the orthonormal eigenfunctions $\{\psi_{j,h}\}$ of L_h and derive that

$$\|W_h'(r)\psi_{j,h}\|_{L^2} = \lambda_{j,h}^s e^{-r\lambda_{j,h}^s} \le r^{-1} \le t_{\ell-1}^{-1}, \quad \text{for } r\in[t_{\ell-1},t_\ell].$$

Whence, $\|W_h'(r)\|_{L^2\to L^2} \le t_{\ell-1}^{-1}$ and

$$\|W_h(\xi)-\overline{W}_\ell\|_{L^2\to L^2} \le \tau \sup_{r\in I_\ell}\|W_h'(r)\|_{L^2\to L^2} \le \tau t_{\ell-1}^{-1}.$$

The above estimate and (IV.16) in (IV.15) imply the final bound on E_2

$$\|E_2\|_{L^2} \le \tau^3 t_{\ell-1}^{-1}\|g_t\|_{L^\infty(t_{\ell-1},t_\ell;L^2)}.$$

$\boxed{4}$ Summing up the estimates of E_1 and E_2 from each subinterval and using a Cauchy-Schwarz inequality, we yield that

$$\begin{aligned}\|u_h - u_h^{\mathcal{N}}\|_{L^2} &\leq \tau^2 \|g_t\|_{L^\infty(0,t;L^2)} + \tau^{5/2} \sum_{\ell=2}^{\mathcal{N}} \|g_{tt}\|_{L^2(t_{\ell-1},t_\ell;L^2)} + \tau^3 \|g_t\|_{L^\infty(0,t;L^2)} \sum_{\ell=2}^{\mathcal{N}} t_{\ell-1}^{-1} \\ &\leq \tau^2 \|g_{tt}\|_{L^\infty(0,t;L^2)} + \tau^{5/2}\mathcal{N}^{1/2} \|g_{tt}\|_{L^2(0,t;L^2)} + \tau^2 \sum_{\ell=1}^{\mathcal{N}-1} \frac{1}{\ell} \|g_t\|_{L^\infty(0,t;L^2)} \\ &= \tau^2((1+\sqrt{t})\|g_{tt}\|_{L^\infty(0,t;L^2)} + C\ln\mathcal{N}\|g_t\|_{L^\infty(0,t;L^2)}).\end{aligned} \tag{IV.17}$$

We note that for the last inequality above we used the fact that $\sum_{j=1}^{\mathcal{N}} j^{-1} \leq C\ln(\mathcal{N}-1)$. To conclude, we observe that

$$\|g_t\|_{L^\infty(0,t;L^2)} = \|f_t\|_{L^\infty(0,t;L^2)}, \qquad \|g_{tt}\|_{L^2(0,t;L^2)} = \|f_{tt}\|_{L^2(0,t;L^2)}$$

and that the embedding $H^1(0,t) \subset L^\infty(0,t)$ is continuous with norm independent of $t \geq t_0$. $\square$

Remark IV.1. *In* (IV.10)*, we can remove the term* $\ln\mathcal{N}$ *when* $f_t \in L^\infty(0,t;\mathbb{H}^\varepsilon)$ *with* $\varepsilon > 0$*. In fact, we can estimate* $\|E_2\|_{L^2}$ *in* (IV.15) *by*

$$\|E_2\|_{L^2} \leq \int_{t_{\ell-1}}^{t_\ell} \|(W_h(\xi) - \overline{W}_\ell)\|_{\dot{H}_h^\varepsilon \to L^2} \|\pi_h\|_{\dot{H}^\varepsilon \to \dot{H}_h^\varepsilon} \|g(\xi) - g(t_{\ell-\frac{1}{2}})\|_{\dot{H}^\varepsilon}\, d\xi.$$

In view of (II.18) *and Proposition II.2, we have* $\|\pi_h\|_{\dot{H}^\varepsilon \to \dot{H}_h^\varepsilon} \leq C$ *and also* $\|g(\xi) - g(t_{\ell-\frac{1}{2}})\|_{\mathbb{H}^\varepsilon} \leq \tau\|g_t\|_{L^\infty(t_{\ell-1},t_\ell;\mathbb{H}^\varepsilon)}$*. We also note that*

$$\begin{aligned}\|W_h'(r)\psi_{j,h}\|_{L^2} &= \lambda_{j,h}^s e^{-r\lambda_{j,h}^s} = r^{-1+\varepsilon/s}\lambda_{j,h}^\varepsilon (r\lambda_{j,h}^s)^{1-\varepsilon/s} e^{-r\lambda_{j,h}^s} \\ &\leq C\lambda_{j,h}^\varepsilon t_{\ell-1}^{-1+\varepsilon/s}, \quad \textit{for } r \in [t_{\ell-1}, t_\ell]\end{aligned}$$

and hence $\|W_h(\xi) - \overline{W}_\ell\|_{\dot{H}_h^\varepsilon \to L^2} \leq C\tau t_{\ell-1}^{-1+\varepsilon/s}$*. This leads to* $\|E_2\|_{L^2} \leq C\tau^3 t_{\ell-1}^{-1+\varepsilon/s}\|g_t\|_{L^2(t_{\ell-1},t_\ell;\mathbb{H}^\varepsilon)}$*. Since* $\tau\sum_{j=2}^{\mathcal{N}-1} t_{\ell-1}^{-1+\varepsilon/s} \leq \int_0^t \xi^{-1+\varepsilon/s}\, d\xi \leq C$*, we insert these results back into* (IV.17) *to obtain*

$$\|u_h - u_h^{\mathcal{N}}\|_{L^2} \leq \tau^2((1+\sqrt{t})\|f_{tt}\|_{L^\infty(0,t;L^2)} + C\|f_t\|_{L^\infty(0,t;\mathbb{H}^\varepsilon)}).$$

IV.2.2 Sinc Approximation of (IV.9)

In view of (IV.9), one remaining problem is to compute the quantity

$$\mathsf{D}_h(t,\tau)g_h := L_h^{-s}(e^{-tL_h^s} - e^{-(t+\tau)L_h^s})g_h.$$

for $t > 0$, $\tau > 0$ and $g_h \in \mathbb{V}_h$. We proceed as in the homogeneous case discussed in Section III.3.

We consider the Dunford-Taylor integral representation of $\mathsf{D}_h(t,\tau)$ with the contour $\widehat{\mathcal{C}}$ given by (III.23). Given a sinc quadrature spacing $k > 0$, we use a positive integer N to be defined below. For $t, \tau > 0$ and $g_h \in \mathbb{V}_h$, we propose the following sinc approximation to $\mathsf{D}_h(t,\tau)$:

$$Q_h^k(t,\tau)g_h := -\frac{k}{2\pi i}\sum_{j=-N}^{N}\big(e^{-t\gamma(y_j)^s} - e^{-(t+\tau)\gamma(y_j)^s}\big)\gamma(y_j)^{-s}\gamma'(y_j)[(\gamma(y_j)I - L_h)^{-1}g_h], \quad \text{(IV.18)}$$

with $y_j = jk$. Hence, a computable approximation of the solution to the non-homogeneous problem becomes

$$u_h^{\mathcal{N},k}(t) := \sum_{\ell=1}^{\mathcal{N}} Q_h^k(t_{\ell-1},\tau)\pi_h f(t - t_{\ell-\frac{1}{2}}). \quad \text{(IV.19)}$$

Remark IV.2 (Implementation)**.** *To minimize the number of system* $((\gamma(y_j)I - L_h)^{-1})$ *solves* g_h *in* (IV.19)*, we write*

$$\begin{aligned} u_h^{\mathcal{N},k}(t) &= -\frac{k}{2\pi i}\sum_{\ell=1}^{\mathcal{N}}\sum_{j=-N}^{N}\big(e^{-t_{\ell-1}\gamma(y_j)^s} - e^{-t_\ell\gamma(y_j)^s}\big)\gamma(y_j)^{-s}\gamma'(y_j)(\gamma(y_j)I - L_h)^{-1}\pi_h f(t - t_{\ell-\frac{1}{2}}). \\ &= -\frac{k}{2\pi i}\sum_{j=-N}^{N}\gamma(y_j)^{-s}\gamma'(y_j)(\gamma(y_j)I - L_h)^{-1}\mathsf{D}_j, \end{aligned}$$

where

$$\mathsf{D}_j := \sum_{\ell=1}^{\mathcal{N}}\big(e^{-t_{\ell-1}\gamma(y_j)^s} - e^{-t_\ell\gamma(y_j)^s}\big)\pi_h f(t - t_{\ell-\frac{1}{2}}). \quad \text{(IV.20)}$$

To implement the above, we proceed as follows:

1) Compute the inner product vectors, i.e., the integral of $f(t - t_{\ell-1/2})$ *against the finite element basis vectors, for all* ℓ.

2) *For each, j:*

   a) *compute the sums in* (IV.20) *but replacing $\pi_h f(t-t_{\ell-1/2})$ by the corresponding inner product vector, and*

   b) *compute $\gamma(y_j)^{-s}\gamma'(y_j)(\gamma(y_j)I - L_h)^{-1}\mathsf{D}_j$ by inversion of the corresponding stiffness matrix applied to the vector of Part a).*

3) *Sum up all contribution and multiply the result by $-\frac{k}{2\pi i}$.*

To analysis the error between $u_h^{\mathcal{N}}(t)$ and $u_h^{\mathcal{N},k}(t)$, we start with the approximation of $\mathsf{D}_h(t,\tau)$ by $Q_h^k(t,\tau)$ with $t > 0$.

Lemma IV.3. *Let $\tau > 0$ and $\mathsf{d} \in (0, \pi/4)$. Assume that $kN > c$ for some positive constant c. When $t > 0$, there exists a constant C independent of h, $\mathcal{N}$ and k such that for any $g_h \in \mathbb{V}_h$,*

$$\|(\mathsf{D}_h(t,\tau) - Q_h^k(t,\tau))g_h\|_{L^2} \le C\tau \max(1, \ln(1/t))\Big(e^{-2\pi\mathsf{d}/k} + e^{-\kappa 2^{-s} t e^{kNs}}\Big)\|g_h\|_{L^2}. \quad \text{(IV.21)}$$

When $t > 0$, we have

$$\|(\mathsf{D}_h(0,\tau) - Q_h^k(0,\tau))g_h\|_{L^2} \le C(e^{-2\pi\mathsf{d}/k} + e^{-sNk})\|g_h\|_{L^2}. \quad \text{(IV.22)}$$

Proof. For $z \in B_{\mathsf{d}}$, define

$$g_\lambda(z,t,\tau) = \big(e^{-t\gamma(z)^s} - e^{-(t+\tau)\gamma(z)^s}\big)\gamma(z)^{-s}\gamma'(z)(\gamma(z)I - \lambda)^{-1}$$

Note that in view of part (b) of Lemma III.8, $g_\lambda(z,t,\tau)$ with $t, \tau > 0$ has the exponential decay property

$$\begin{aligned} |g_\lambda(z,t,\tau)| &\le C|\gamma(z)^{-s}| \int_t^{t+\tau} |\gamma(z)^s e^{-r\gamma(z)^s}|\, dr \\ &= C\int_t^{t+\tau} e^{-r\mathfrak{Re}(\gamma(z)^s)}\, dr \le C\tau e^{-t\mathfrak{Re}(\gamma(z)^s)} \le C\tau e^{-t\kappa(\cosh\mathfrak{Re} z)^s} \end{aligned} \quad \text{(IV.23)}$$

with κ defined by (III.28). Hence we follow the proof of Lemma III.4 and Remark III.5 to obtain the estimate (IV.21).

When $t = 0$, we note that

$$\mathsf{D}_h(0,\tau)g_h = (L_h^{-s} - L_h^{-s}e^{-\tau L_h^s})g_h.$$

does not have the exponential decay property (IV.23) due to the term $L_h^{-s}g_h$. So we estimate the sinc approximation error for $L_h^{-s}g_h$ and $L_h^{-s}e^{-\tau L_h^s}g_h$ separately. In view of Lemma III.8.(c), we have

$$|g_\lambda(z)| := |z^{-s}\gamma'(z)(\gamma(z)I - \lambda)^{-1}| \le C\kappa 2^s e^{-s|\mathfrak{Re}z|}.$$

So we follow the argument before Theorem II.7 together with the above decay property to conclude that the L^2 error between $L_h^{-s}g_h$ and its sinc approximation in $Q_h^k(t,\tau)g_h$ should be bounded by $C(e^{-2\pi\mathsf{d}/k} + e^{-sNk})$. Here the constant C is independent of h, τ, N and k. On the other hand, we simply estimate the sinc approximation error for $L_h^{-s}e^{-\tau L_h^s}g_h$ by using the same exponential decay estimate as above since $|e^{-\tau\gamma(z)^s}| \le 1$ for $z \in B_{\mathsf{d}}$. Hence we have shown (IV.22) and the proof is complete. □

We are now ready to show the error estimate for the sinc approximation on the non-homogeneous problem.

Theorem IV.4. *Let $t > 0$ and assume that $f \in L^\infty(0,t;L^2)$. Assume that $Nk \ge c$ for some positive constant c. Let $u_h^{\mathcal{N}}$ and $u_h^{\mathcal{N},k}$ be defined by* (IV.9) *and* (IV.18)*, respectively. Then there exists a positive constant C independent of h, τ, k and N satisfying*

$$\|u_h^{\mathcal{N}} - u_h^{\mathcal{N},k}\|_{L^2} \le Ct\max(1,\ln(1/t))(e^{-2\pi\mathsf{d}/k} + e^{-kNs/2})\|f\|_{L^\infty(0,t;L^2)}. \qquad \text{(IV.24)}$$

Proof. We write

$$\|u_h^{\mathcal{N}} - u_h^{\mathcal{N},k}\|_{L^2} \le \|(\mathsf{D}_h(0,\tau) - Q_h^k(0,\tau))\pi_h f(t_{\frac{1}{2}})\|_{L^2} + \sum_{\ell=2}^{\mathcal{N}-1} \|(\mathsf{D}_h(t_{\ell-1},\tau) - Q_h^k(t_{\ell-1},\tau))\pi_h f(t - t_{\ell-\frac{1}{2}})\|_{L^2}.$$

The first term on the right hand side follows from (IV.22). In terms of the summation above, we apply (IV.21) to get

$$\begin{aligned}
&\sum_{\ell=2}^{\mathcal{N}-1} \|(\mathsf{D}_h(t_{\ell-1},\tau) - Q_h^k(t_{\ell-1},\tau))\pi_h f(t - t_{\ell-\frac{1}{2}})\|_{L^2} \\
&\quad\le C\|f\|_{L^\infty(0,t;L^2)}\tau \sum_{\ell=2}^{\mathcal{N}-1} \max(1, \ln(1/t_{\ell-1}))\Big(e^{-2\pi\mathsf{d}/k} + e^{-\kappa 2^{-s} t_{\ell-1} e^{kNs}}\Big) \\
&\quad\le C\|f\|_{L^\infty(0,t;L^2)}\Big(e^{-2\pi\mathsf{d}/k}\tau \sum_{\ell=2}^{\mathcal{N}-1} \max(1, \ln(1/t_{\ell-1})) \\
&\qquad + e^{-kNs/2}\tau \sum_{\ell=2}^{\mathcal{N}-1} t_{\ell-1}^{-1/2} \max(1, \ln(1/t_{\ell-1}))\Big) \\
&\quad\le C(t\max(1,\ln(1/t))e^{-2\pi\mathsf{d}/k} + \sqrt{t}\max(1,\ln(1/t))e^{-kNs/2})\|f\|_{L^\infty(0,t;L^2)}.
\end{aligned}$$

Combing above two estimate yields the desired estimate. □

Remark IV.3 (Relation between N and k)**.** *We balance the error in* (IV.24) *by letting* $2\pi\mathsf{d}/k = sNk/2$ *so that* $Nk = \sqrt{4\pi\mathsf{d}N/s} > \sqrt{4\pi\mathsf{d}/s}$ *when* $N > 1$*. Hence the error of the sinc approximation is bounded by* $C(t\max(1,\ln(1/t))e^{-2\pi\mathsf{d}/k}$.

IV.2.3 The Total Error

We end this section by combining the error estimates from Theorem IV.1, IV.2 and IV.4.

Theorem IV.5 (Total Error)**.** *Let* $u(t)$ *be the solution of the non-homogeneous problem* (IV.2)*.* $u_h^{\mathcal{N},k}(t)$ *is the approximation of* u *defined by* (IV.19) *together with the choice of* N *based on Remark IV.3. Given the regularity index* $\alpha \in (0,1]$ *from Assumption II.1, assume that* $f \in L^\infty(0,t;\mathbb{H}^{2\delta}) \cap H^2(0,t;L^2)$ *with* $\delta \in [0,(1+\delta)/2]$*. Then for* $t \ge t_0 > 0$*, there exists a posi-*

tive constant C satisfying

$$\begin{aligned}\|u(t) - u_h^{\mathcal{N},k}(t)\|_{L^2} \le \widetilde{C}(t) h^{2\widetilde{\alpha}} \|f\|_{L^\infty(0,\mathsf{T};\dot{H}^{2\delta})} \\ + \tau^2((1+\sqrt{t})\|f\|_{H^2(0,t;L^2)} + C \ln \frac{t}{\tau} \|f_t\|_{L^\infty(0,t;L^2)}) \\ + Ct \max(1, \ln(1/t)) e^{-2\pi \mathsf{d}/k} \|f\|_{L^\infty(0,t;L^2)}.\end{aligned}$$

Here $\widetilde{C}(t)$ is given by (IV.6).

IV.3 The Crank-Nicolson Time Stepping Method

In this section, we consider an alternative approximation in the time domain. Recall from (IV.4) that the finite element approximation $u_h(t)$ satisfies the equation

$$u_{h,t} + L_h^s u_h = \pi_h f(t), \quad \text{for } t > 0 \text{ with } u_h(0) = 0.$$

The Crank-Nicolson time stepping scheme reads: given a time step $\tau > 0$ and a nonnegative integer n, we approximate $\{u_h(t_n)\}_{n=0}^\infty$ with $t_n = n\tau$ by $\{u_h^n\}_{n=0}^\infty \subset \mathbb{V}_h$ satisfying

$$\frac{u_h^{n+1} - u_h^n}{\tau} + L_h^s \left(\frac{u_h^{n+1} + u_h^n}{2} \right) = f_h^{n+1/2}, \quad \text{for } n \ge 0 \quad \text{with } u_h^0 = \pi_h v, \tag{IV.25}$$

where $f_h^{n+1/2} = \pi_h f(t_{n+\frac{1}{2}})$. We simplify the above equation and find that

$$\begin{aligned} u_h^{n+1} &= (I + \frac{\tau}{2} L_h^s)^{-1} (I - \frac{\tau}{2} L_h^s) u_h^n + \tau (I + \frac{\tau}{2} L_h^s)^{-1} f_h^{n+1/2} \\ &= (I + \frac{\tau}{2} L_h^s)^{-1} (2u_h^n + \tau f_h^{n+1/2}) - u_h^n. \end{aligned} \tag{IV.26}$$

The remaining issue is to approximate $E_\tau g_h := (I + \tau L^s)^{-1} g_h$ for $g_h \in \mathbb{V}_h$. To do this, we invoke the following Balakrishman formula:

$$(I + L_h^s)^{-1} = \frac{\sin(\pi s)}{\pi} \int_0^\infty \frac{\mu^s}{1 + 2\mu^s \cos(\pi s) + \mu^{2s}} (\mu I + L_h)^{-1} \, d\mu. \tag{IV.27}$$

Hence

$$E_\tau g_h = \frac{\sin(\pi s)}{\pi} \int_0^\infty \frac{\mu^s}{1 + 2\mu^s \cos(\pi s) + \mu^{2s}} (\mu I + \tau^{1/s} L_h)^{-1} g_h \, d\mu.$$

Now we apply a sinc method to the above integral representation. We use the change of variable $\mu = e^y$ and set the quadrature spacing $k > 0$ together with a positive integer N. The approximation of $E_\tau g_h$ is given by

$$Q_\tau^k g_h := \frac{\sin(\pi s) k}{\pi} \sum_{j=-N}^{N} \frac{e^{y_j(s+1)}}{1 + 2e^{s y_j} \cos(\pi s) + e^{2 s y_j}} (e^{y_j} I + \tau^{1/s} L_h)^{-1} g_h. \tag{IV.28}$$

We replace $E_{\tau/2} = (I + \frac{\tau}{2} L_h^s)^{-1}$ with $Q_{\tau/2}^k$ to obtain our final approximation, i.e. we approximate $\{u_h(t_n)\}$ using $\{u_h^{n,k}\}$ which satisfies

$$u_h^{n+1,k} = Q_{\tau/2}^k (2u_h^{n,k} + \tau f_h^{n+1/2}) - u_h^{n,k} \quad \text{for } n > 0 \text{ with } u_h^{0,k} = 0. \tag{IV.29}$$

IV.3.1 Error Analysis for the Time Discretization

Let us consider the estimate of the L^2 error between the finite element approximation $u_h(t_n)$ and its time discretization u_h^n. Let $W_\tau := (I + \frac{\tau}{2} L_h^s)^{-1}(I - \frac{\tau}{2} L_h^s)$. Recalling in Section IV.2 that $W_h(t_n) = e^{-t_n L_h^s}$, W_τ^n should be an approximation of $W_h(t_n)$. The error between the two is given by the following lemma.

Lemma IV.6. *Assume that $\tau \max_{j=1,\ldots,M_h} \{\lambda_{j,h}^s\} \le \alpha_0$ for some $0 < \alpha_0 < \infty$. For $g_h \in \mathbb{V}_h$, there exists a positive constant C independent of h, τ and n satisfying*

$$\|(W_h(t_n) - W_\tau^n) g_h\|_{L^2} \le C t_n^{\delta/s-2} \tau^2 \|g_h\|_{\dot{H}_h^{2\delta}}, \tag{IV.30}$$

where $\delta \in [0, 2s]$.

Proof. For the cases $\delta = 0$ and $\delta = 2s$, (IV.30) follows from Theorem 7.2 and 7.1 in [69], respectively. The proof is complete by invoking the interpolation estimate (II.3) with these two cases. □

Remark IV.4. *We can remove the assumption $\tau \max_j\{\lambda_{j,h}^s\} \le \alpha_0$ in the above lemma by applying the Backward scheme at the first two time steps, i.e.*

$$u_h^{n+1} = E_\tau(u_h^n + \tau f_h^{n+1}), \qquad \textit{for } n = 0, 1.$$

Then we apply the Crank-Nicolson scheme for $n \ge 2$. The same error estimate has been proved in [69, Theorem 7.4]; see also [49] for more details.

Theorem IV.7 (Error estimate for the Crank-Nisolson scheme)**.** *Let $t_n = n\tau$ with $\tau > 0$ and the nonnegative integer n. Let $u_h(t_n)$ and u_h^n given by* (IV.4) *and* (IV.25)*, respectively. Under Assumption II.1 for some $\alpha \in (0,1)$, we assume that $s \in (0, (1+\alpha)/2]$. Assume that the right hand side data f satisfies $f \in L^\infty(0, t_n; \mathbb{H}^{2\delta})$, $f_t \in L^\infty(0, t_n; \mathbb{H}^{2\delta'})$ and $f_{tt} \in L^1(0, t_n; L^2)$, where $\delta \in (s, (1+\alpha)/2]$ and $\delta' \in (0, (1+\alpha)/2]$. We additionally assume that $\tau \max_j\{\lambda_{j,h}^s\} \le \alpha_0$. Then there exists a positive constant C independent of h and τ satisfying*

$$\begin{aligned}\|u_h(t_n) - u_h^n\|_{L^2} \le & C\tau^2\bigg(1 + (t_n^{\delta/s-1})\|f\|_{L^\infty(0,t_n;\mathbb{H}^{2\delta})} \\ & \qquad + (1 + t_n^{\delta'/s})\|f_t\|_{L^\infty(0,t_n;\mathbb{H}^{2\delta'})} + \|f_{tt}\|_{L^1(0,t_n;L^2)}\bigg).\end{aligned}$$

Proof. $\boxed{1}$ Recalling that $f_h^n = \pi_h f(t_n)$, we apply the first equation of (IV.26) recursively from $n-1$ to 0 and get

$$u_h^n = \tau \sum_{j=0}^{n-1} W_\tau^{n-j-1} E_{\tau/2} f_h^{j+1/2}. \tag{IV.31}$$

On the other hand, we let $f_h(t) = \pi_h f(t)$ and rewrite $u_h(t_n)$ as

$$\begin{aligned}u_h(t_n) &= \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} W_h(t_n - \xi) f_h(\xi)\, d\xi \\ &= \tau \sum_{j=0}^{n-1} W_h(t_{n-j-1}) \underbrace{\int_0^1 W_h(\tau - \tau\xi) f_h(t_j + \tau\xi)\, d\xi}_{=: I_\tau^j}.\end{aligned}$$

Thus,

$$
\begin{aligned}
e_h^n := u_h^n - u_h(t_n) &= \tau \sum_{j=0}^{n-1} \left(W_\tau^{n-j-1} E_{\tau/2} f_h^{j+1/2} - W_h(t_{n-j-1}) I_\tau^j \right) \\
&= \tau \sum_{j=0}^{n-2} \left(W_\tau^{n-j-1} - W_h(t_{n-j-1}) \right) E_{\tau/2} f_h^{j+1/2} \\
&+ \tau \sum_{j=0}^{n-2} W_h(t_{n-j-1}) \left(E_{\tau/2} f_h^{j+1/2} - I_\tau^j \right) \\
&+ \tau \left(E_{\tau/2} f_h^{n-1/2} - I_\tau^{n-1} \right) =: e_{h,1}^n + e_{h,2}^n + e_{h,3}^n.
\end{aligned}
$$

We shall bound $e_{h,i}^n$ for $i = 1, 2, 3$ separately.

$\boxed{2}$ We first bound $e_{h,1}^n$. Since $E_{\tau/2}$ and $W_h(t_{n-j-1})$ commute and $\|E_{\tau/2}\|_{L^2 \to L^2} \leq 1$, we have

$$
\|e_{h,1}^n\|_{L^2} \leq \tau \sum_{j=0}^{n-2} \left\| (W_\tau^{n-j-1} - W_h(t_{n-j-1})) f_h^{j+1/2} \right\|_{L^2}.
$$

Applying Lemma IV.6 yields

$$
\begin{aligned}
\|e_{h,1}^n\|_{L^2} &\leq C\tau^3 \sum_{j=0}^{n-2} t_{n-j-1}^{-(2-\frac{\delta}{s})} \|f_h^{j+1/2}\|_{\dot{H}_h^{2\delta}} \\
&\leq C\tau^2 \int_0^{t_{n-1}} \xi^{-2+\frac{\delta}{s}} \, d\xi \, \|f\|_{L^\infty(0,t_n;\dot{H}^{2\delta})} \leq C t_n^{-1+\delta/s} \tau^2 \|f\|_{L^\infty(0,t_n;\mathbb{H}^{2\delta})}.
\end{aligned}
$$

Note that for the second inequality, $\|f_h^{j+1/2}\|_{\dot{H}_h^{2\delta}} \leq C\|f(t_{j+1/2})\|_{\dot{H}^{2\delta}} \leq C\|f\|_{L^\infty(0,t_n;\dot{H}^{2\delta})}$ due to Lemma II.20.

$\boxed{3}$ Before we estimate $\|e_{h,2}^n\|_{L^2}$, let us consider $\|E_{\tau/2} f_h^{j+1/2} - I_\tau^j\|_{L^2}$ for $j = 0, \ldots, n-2$. According to the Taylor expansion of $f_h^{j+1/2}$ at t_j, we write

$$
E_{\tau/2} f_h^{j+1/2} = E_{\tau/2} \pi_h f(t_j) + \tau E_{\tau/2} \pi_h f_t(t_j) + E_{\tau/2} \int_{t_j}^{t_j+\tau/2} (t_j + \tau/2 - \xi) \pi_h f_{tt}(\xi) \, d\xi.
$$

Similarly,

$$I_\tau^j = \int_0^1 W_h(\tau - \xi\tau)\, d\xi\ \pi_h f(t_j) + \tau \int_0^1 W_h(\tau - \xi\tau)\xi\, d\xi\ \pi_h f_t(t_j)$$
$$+ \int_0^1 W_h(\tau - \xi\tau) \left(\int_{t_j}^{t_j+\xi\tau} (t_j + \xi\tau - \eta)\pi_h f_{tt}(\eta)\, d\eta \right) d\xi.$$

Notice that in above two equations, both the L^2 norm of last terms on the right hand side can be bounded by

$$C\tau \int_{t_j}^{t_{j+1}} \|f_{tt}\|_{L^2}\, d\xi.$$

It remains to bound $\|\tau^l E_{\tau/2}\pi_h f(t_j) - \tau^l \int_0^1 W_h(\tau - \xi\tau)\xi^l\, d\xi\ \pi_h f^{(l)}(t_j)\|_{L^2}$ for $l = 0, 1$, where $f^{(0)} = f$ and $f^{(1)} = f_t$. Note that for $\lambda > 0$, there exists a positive constant C satisfying

$$\left|(1 + \lambda/2)^{-1} - \int_0^1 \xi^l e^{-(1-\xi)\lambda}\, d\xi\right| \le C\lambda^{2-l} \qquad \text{for } 0 \le l \le 2.$$

So let us consider the case $l = 1$ and the case $l = 0$ should follow the same argument. Using the eigenfunction expansion with respect to $\{\psi_{\ell,h}\}$ and letting $c_{\ell,h}$ be the corresponding decomposition coefficients for $\ell = 1, 2, \ldots, M_h$, we obtain

$$\begin{aligned}
&\|\tau(E_{\tau/2} - \int_0^1 W_h(\tau - \xi\tau)\xi\, d\xi)W_h(t_{n-j-1})\pi_h f_t(t_j)\|_{L^2}^2 \\
&= \tau^2 \sum_{\ell=1}^{M_h} \left((1 + \tau\lambda_{\ell,h}^s/2)^{-1} - \int_0^1 \xi e^{-(1-\xi)\tau\lambda_{\ell,h}^s}\, d\xi \right)^2 e^{-2t_{n-j-1}\lambda_{\ell,h}^s} c_{\ell,h}^2 \\
&\le C\tau^2 \sum_{\ell=1}^{M_h} (\tau\lambda_{\ell,h}^s)^2 e^{-2t_{n-j-1}\lambda_{\ell,h}^s} c_{\ell,h}^2 \\
&\le C t_{n-j-1}^{-2(1-\delta'/s)} \tau^4 \sum_{\ell=1}^{M_h} \lambda_{\ell,h}^{2\delta'} \left(t_{n-j-1}\lambda_{\ell,h}^s\right)^{2-2\delta'/s} e^{-2t_{n-j-1}\lambda_{\ell,h}^s} c_{\ell,h}^2 \\
&\le C t_{n-j-1}^{-2(1-\delta'/s)} \tau^4 \sum_{\ell=1}^{M_h} \lambda_{\ell,h}^{2\delta'} c_{\ell,h}^2 = C t_{n-j-1}^{-2(1-\delta'/s)} \tau^4 \|\pi_h f_t(t_j)\|_{\dot H_h^{2\delta'}}^2.
\end{aligned} \tag{IV.32}$$

Similarly,

$$\|(E_{\tau/2} - \int_0^1 W_h(\tau - \xi\tau)\, d\xi)W_h(t_{n-j-1})\pi_h f(t_j)\|_{L^2} \le C t_{n-j-1}^{-(2-\delta/s)} \tau^2 \|\pi_h f(t_j)\|_{\dot H_h^{2\delta}}$$

Now we can apply the above two estimates to bound $e^n_{h,2}$. That is

$$\|e^n_{h,2}\| \le C\tau^3 \sum_{j=0}^{n-2} \left(t_{n-j-1}^{-(2-\delta/s)} \|f_h(t_j)\|_{\dot H_h^{2\delta}} + t_{n-j-1}^{-(1-\delta/s)} \|\pi_h f_t(t_j)\|_{\dot H_h^{2\delta'}} \right) + \tau^2 \sum_{j=0}^{n-2} \int_{t_j}^{t_{j+1}} \|f_{tt}(\xi)\| \, d\xi$$
$$\le C\tau^2 \left(t_n^{\delta'/s-1} \|f\|_{L^\infty(0,t_n;\mathbb{H}^{2\delta})} + t_n^{\delta'/s} \|f_t\|_{L^\infty(0,t_n;\mathbb{H}^{2\delta'})} + \|f_{tt}\|_{L^1(0,t_{n-1};L^2)} \right).$$

$\boxed{4}$ We follow the same argument in the previous step to bound $\|e^n_{3,h}\|_{L^2}$ with $j = n-1$. The only difference is in (IV.32). Note that

$$\|\tau(E_{\tau/2} - \int_0^1 W_h(\tau - \xi\tau)\xi \, d\xi)\pi_h f_t(t_j)\|^2_{L^2} \le C\tau^2 \sum_{\ell=1}^{M_h} (\tau\lambda^s_{\ell,h})^2 c^2_{\ell,h} \le C\tau^2 \|f_t(t_{n-1})\|^2_{L^2},$$

where we used the assumption $\tau\lambda^s_{\ell,h} \le \alpha_0$ to bound the second inequality. Hence

$$\|e^n_{3,h}\|_{L^2} \le C\tau^2 \Big(\|f(t_{n-1})\|_{L^2} + \|f_t(t_{n-1})\|_{L^2} + \int_{t_{n-1}}^{t_n} \|f_{tt}\|_{L^2} \, d\xi \Big).$$

$\boxed{5}$ The proof is complete by combing the estimates of $e^n_{i,h}$ for $i = 1, 2, 3$. $\square$

IV.3.2 Error Analysis for the Sinc Approximation

The key to the error analysis between Q^k_τ and E_τ is showing that the integrand function

$$g_\lambda(\tau; z) = \frac{e^{z(s+1)}}{1 + 2\cos(\pi s)e^{sz} + e^{2sz}} (e^z + \tau^{1/s}\lambda)^{-1},$$

defined in B_{d} for some $\mathsf{d} > 0$ has an exponential decay as $|\mathfrak{Re} z| \to \infty$. In fact, we can show that

Lemma IV.8 (Exponential decay)**.** *Given $\mathsf{d} \in (0, \min(\pi/2, \pi/s - \pi))$, there exists a positive constant C independent of τ satisfying*

$$|g_\lambda(\tau; z)| \le Ce^{-s|\mathfrak{Re} z|}, \qquad \textit{for } z \in B_{\mathsf{d}}.$$

Proof. We factorize $g_\lambda(\tau; z)$ as

$$g_\lambda(\tau; z) = \frac{e^{sz}}{(1+e^{sz+is\pi})(1+e^{sz-is\pi})} \frac{e^z}{e^z+\tau^{1/s}\lambda}. \tag{IV.33}$$

We first note that

$$\left|\frac{e^z}{e^z+\tau^{1/s}\lambda}\right| \le \frac{e^{\mathfrak{Re} z}}{e^{\mathfrak{Re} z}\cos(\mathfrak{Im} z)+\tau^{1/s}\lambda} \le \frac{1}{\cos \mathsf{d}}. \tag{IV.34}$$

For $\mathfrak{Re} z \le 0$,

$$\begin{aligned} \left|1+e^{sz+is\pi}\right| &\ge \left|\mathfrak{Re}(1+e^{sz+is\pi})\right| = 1+e^{s\mathfrak{Re} z}\cos s(\mathfrak{Im} z+\pi) \\ &\ge 1+e^{s\mathfrak{Re} z}\cos s(\mathsf{d}+\pi) \ge 1+\min(0,\cos s(\mathsf{d}+\pi)) =: c_0 > 0. \end{aligned}$$

Analogously, $|1+e^{sz-is\pi}| \ge c_0$. Thus, we combine above estimates and conclude that for $\mathfrak{Re} z \le 0$,

$$|g_\lambda(\tau; z)| \le \frac{e^{s\mathfrak{Re} z}}{c_0^2 \cos \mathsf{d}}.$$

For $\mathfrak{Re} z \ge 0$, we write $g_\lambda(\tau; z)$ as

$$g_\lambda(\tau; z) = \frac{e^{-sz}}{(1+e^{-sz+is\pi})(1+e^{-sz-is\pi})} \frac{e^z}{e^z+\lambda} \tag{IV.35}$$

and follow the same argument as the case $\mathfrak{Re} z \le 0$. Hence the proof is complete. □

We are able to provide the error estimate between Q_τ^k and E_τ.

Theorem IV.9 (SINC quadrature error for one-step time iteration)**.** *Let Q_τ^k defined by* (IV.28) *be the sinc approximation of $E_\tau = (I+\tau L_h^s)^{-1}$ Given $\mathsf{d} \in (0, \min(\pi/2, \pi/s-\pi))$, there exists a positive constant C independent of h, k and τ and N satisfying*

$$\|Q_\tau^k - E_\tau\|_{L^2\to L^2} \le \frac{2\sin(\pi s)}{\pi s c_0^2 \cos \mathsf{d}} \left(\frac{2}{e^{2\pi \mathsf{d}/k}-1} + e^{-sNk}\right). \tag{IV.36}$$

Proof. It suffices to show the scalar case, i.e.

$$\mathcal{E}(\tau;\lambda) := \left|k\int_{-\infty}^{\infty} g_\lambda(\tau;y)\,dy - k\sum_{j=-N}^{N} g_\lambda(\tau;kj)\right| \le \frac{2}{sc_0^2\cos \mathsf{d}}\left(\frac{2}{e^{2\pi\mathsf{d}/k}-1}+e^{-sNk}\right). \quad \text{(IV.37)}$$

Note that for $z \in B_{\mathsf{d}}$ and $\lambda > 0$. $|e^z + \tau^{1/s}\lambda| \ge \tau^{1/s}\lambda > 0$. Also based on the proof of Lemma IV.8, the denominators of $g_\lambda(\tau;z)$ in (IV.33) and (IV.35) are bounded away from zero. So $g_\lambda(\tau;z)$ is analytic in $\overline{B}_{\mathsf{d}}$. Also, invoking the decay estimate (IV.36),

$$\int_{-\mathsf{d}}^{\mathsf{d}} |g_\lambda(\tau;y+i\eta)|\,d\eta \le \frac{2\mathsf{d}}{c_0^2\cos\mathsf{d}} < \infty$$

and

$$N(B_{\mathsf{d}}) \le \frac{4}{c_0^2\cos\mathsf{d}}\int_0^\infty e^{-sy}\,dy = \frac{4}{sc_0^2\cos\mathsf{d}} < \infty.$$

Hence $g_\lambda(\tau;z) \in S(B_{\mathsf{d}})$ and (II.23) holds for $f = g_\lambda(\tau;\cdot)$. Also,

$$k\sum_{|j|\ge N+1} |g_\lambda(\tau;jk)| \le \frac{2}{c_0^2\cos\mathsf{d}}\int_{Nk}^\infty e^{-sy}\,dy \le \frac{2e^{-sNk}}{sc_0^2\cos\mathsf{d}}.$$

Combing the above estimate as well as (II.23) gives the desired estimate. □

Remark IV.5. *In spite of the usual choose $k = 1/\sqrt{N}$, we balance the two exponential terms on the right hand side of* (IV.36)*, i.e. given a fixed k, we choose $N = 2\pi\mathsf{d}/(sk^2)$ so the quadrature error estimate becomes $O(e^{-2\pi\mathsf{d}/k})$.*

Before we analysis the error $\|u_h^n - u_h^{n,k}\|_{L^2}$, let us discuss the stability of the numerical scheme (IV.29). Let $E_\tau(\lambda)$ and $Q_\tau^k(\lambda)$ be the scalar version of E_τ and Q_τ^k by replacing L_h with λ. Recalling the notation $\mathcal{E}(\tau;\lambda)$ in (IV.37), we assume that N is large enough so that

$$r(k) := \frac{\sin(\pi s)}{\pi}\frac{2}{sc_0^2\cos\mathsf{d}}\left(\frac{2}{e^{2\pi\mathsf{d}/k}-1}+e^{-sNk}\right) \le \min\left(\frac{\tau\lambda_{1,h}^s/2}{1+\tau\lambda_{1,h}^s/2}, \frac{1}{1+\tau\lambda_{M_h,h}^s/2}\right) \quad \text{(IV.38)}$$

so that for $j = 1, \ldots, M_h$,

$$\begin{aligned}|Q_\tau^k(\lambda_{j,h})| &\le |E_{\tau/2}(\lambda_{j,h})| + |Q_\tau^k(\lambda_{j,h}) - E_{\tau/2}(\lambda_{j,h})| \\ &= |E_{\tau/2}(\lambda_{j,h})| + \frac{\sin(\pi s)}{\pi}|\mathcal{E}(\lambda_{j,h})| \\ &\le |E_{\tau/2}(\lambda_{j,h})| + \frac{\tau\lambda_{j,h}^s}{2+\tau\lambda_{j,h}^s} = 1.\end{aligned}$$

Also for $j = 1, \ldots, M_h$,

$$\begin{aligned}Q_\tau^k(\lambda_{j,h}) &\ge E_{\tau/2}(\lambda_{j,h}) - |Q_\tau^k(\lambda_{j,h}) - E_{\tau/2}(\lambda_{j,h})| \\ &= E_{\tau/2}(\lambda_{j,h}) - \frac{\sin(\pi s)}{\pi}|\mathcal{E}(\tau;\lambda_{j,h})| \\ &\ge E_{\tau/2}(\lambda_{j,h}) - \frac{1}{1+\tau\lambda_{j,h}^s/2} = 0.\end{aligned}$$

So $0 \le Q_\tau^k(\lambda_{j,h}) \le 1$ for all $\tau > 0$. Applying this to the scheme (IV.29) yields that $\|2Q_\tau^k - I\|_{L^2\to L^2} \le 1$ and hence

$$\|u_h^{n+1,k}\|_{L^2} \le \|u_h^{n,k}\|_{L^2} + \tau\|f(t_{n+\frac{1}{2}})\|_{L^2}.$$

This implies the stability of the numerical scheme.

Theorem IV.10 (Error estimate on sinc approximation)**.** *Given a time step τ, the number of time steps n and a sinc quadrature spacing k, let $u^{n,k}$ defined by* (IV.29) *be the approximation of u_k^n given by* (IV.26)*. We also assume that k is small enough so that* (IV.38) *holds. If $f \in L^\infty(0, t_n; L^2)$, then there holds*

$$\|u_h^n - u_h^{n,k}\|_{L^2} \le \frac{t_n^2 r(k)}{2\tau}\|f\|_{L^\infty(0,t_n;L^2)}.$$

Proof. Note that $Q_{\tau/2}^k$ and $2Q_\tau^k - I$ are approximations of $E_{\tau/2}$ and W_τ, respectively. Similar to (IV.31), we write

$$u_h^{n,k} = \tau\sum_{j=0}^{n-1}(2Q_{\tau/2}^k - I)^{n-j-1}Q_{\tau/2}^k f_h^{j+1/2}.$$

Hence,

$$u_h^n - u_h^{n,k} = \tau \sum_{j=0}^{n-1} \big(W_\tau^{n-j-1} - (2Q_{\tau/2}^k - I)^{n-j-1}\big) Q_{\tau/2}^k f_h^{j+1/2}$$
$$+ \tau \sum_{j=0}^{n-1} W_\tau^{n-j-1}(E_{\tau/2} - Q_{\tau/2}^k) f_h^{j+1/2} =: \widetilde{e}_1 + \widetilde{e}_2.$$

We simply bound $\widetilde{e}_2$ by

$$\|\widetilde{e}_2\|_{L^2} \le \tau n r(k) \|f\|_{L^\infty(0,t_n;L^2)}. \tag{IV.39}$$

In terms of $\widetilde{e}_1$, noting that $\|(2Q_{\tau/2}^k - I)\|_{L^2\to L^2} \le 1$ by stability,

$$\|W_\tau^n - (2Q_{\tau/2}^k - I)^n\|_{L^2\to L^2}$$
$$\le \|E_{\tau/2} - Q_{\tau/2}^k\|_{L^2\to L^2} \sum_{j=0}^{n-1} \|W_\tau^j\|_{L^2\to L^2} \|(2Q_{\tau/2}^k - I)^{n-1-j}\|_{L^2\to L^2} \le n r(k).$$

So

$$\|\widetilde{e}_1\|_{L^2} \le \tau \sum_{j=0}^{n-1} (n-j-1) \|f\|_{L^\infty(0,t_n;L^2)} = \frac{n(n-1)}{2} \tau \|f\|_{L^\infty(0,t_n;L^2)}. \tag{IV.40}$$

Combing (IV.40) and (IV.39) together with the relation $t_n = n\tau$ gives the desired estimate. □

IV.3.3 The Total Error

Let us summarize the error estimates from Theorem IV.1, IV.7 and IV.10.

Theorem IV.11. *Given a time step τ, the number of time stepping n and a sinc quadrature spacing k, let $u_h^{n,k}$ defined by* (IV.26) *be the approximation of $u(t_n)$ defined by* (IV.2)*. Assume that Assumption* (II.1) *holds for some $\alpha \in (0,1]$. We also assume that $\tau\lambda_{M_h,h}^s \le \alpha_0$ for some $\alpha_0 > 0$ and* (IV.38) *holds so that the numerical scheme* (IV.26) *is stable. If the right hand side data f satisfies $f \in L^\infty(0,t_n;\mathbb{H}^{2\delta})$ for $\delta \in (s,(1+\alpha)/2]$, $f_t \in L^\infty(0,t_n;\mathbb{H}^{2\delta'})$ for $\delta' \in (0,(1+\alpha)/2]$*

and $f_{tt} \in L^1(0,t_n;L^2)$. *Then*

$$\|u(t_n) - u_h^{n,k}\|_{L^2} \le \widetilde{C}(t) h^{2\widetilde{\alpha}} \|f\|_{L^\infty(0,\mathsf{T};\mathbb{H}^{2\delta})}$$
$$+ C\tau^2\Big((1+t_n^{\delta/s-1})\|f\|_{L^\infty(0,t_n;\mathbb{H}^{2\delta})} + (1+t_n^{\delta/s})\|f_t\|_{L^\infty(0,t_n;\mathbb{H}^{2\delta'})} + \|f_{tt}\|_{L^1(0,t_n;L^2)}\Big)$$
$$+ \frac{t_n^2 r(k)}{2\tau}\|f\|_{L^\infty(0,t_n;L^2)}.$$

Here $\widetilde{C}(t)$ *is given by* (IV.6).

Let us discuss the relations between discretization parameters to end this section.

Remark IV.6 (Choice of parameters)**.** *Given a fixed mesh size* h*, we can first choose* τ *so that* $\tau \le \alpha_0 \lambda_{M_h,h}^{-s} \sim Ch^{2s}$*. Then we set* $r(k) \le C\tau^3$ *so that the numerical scheme is stable and the second order convergence in time is also guaranteed.*

IV.4 Numerical Illustration

In this section, we provide some numerical simulations to verity the error estimates we have shown in Section IV.1, Section IV.2 and Section IV.3.

IV.4.1 Error Behavior based on Finite Element Approximation

Consider the one dimensional non-homogeneous problem (IV.4) with the bilinear form $d(u,v) = \int_0^1 u'v'\,dx$ and the right hand side $f(t,x) = f(x) = v(x)$ in (III.48). We compute the finite element approximation $u_h(t)$ using the method mentioned in Section III.4.1. In Table IV.1, we report the asymptotic observed convergence rate for $t = 0.5$. This rate is defined by OROC:= $\ln(e_{h_9}/e_{h_{10}})/\ln 2$ for $s > 1/4$ and OROC:= $\ln(e_{h_{12}}/e_{h_{13}})/\ln 2)$ for $s < 1/4$ where the mesh size $h_i = 1/2^i$. The finer mesh sizes were used in the case of $s < 1/4$ to get closer to the asymptotic convergence order.

	$s < 0.25$		$s > 0.25$						
s	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
OROC	1.73	1.88	1.95	1.99	2.00	2.00	2.00	2.00	2.00
THM	1.7	1.9	2.0	2.0	2.0	2.0	2.0	2.0	2.0

Table IV.1: Observed rate of convergence (OROC) for different values of s together the rates predicted by Theorem IV.1 (THM).

IV.4.2 Error Behavior in Time via Numerical Integration

We consider the non-homogeneous problem (IV.4) in the square domain $\Omega = [0,1]^2$ and we set the bilinear form $d(u,v) = \int_\Omega \nabla u \cdot \nabla v \, dx$ for $u, v \in H_0^1(\Omega)$. We also let the right hand side data $f(t,x,y) = (3t^2 + t^2(2\pi^2)^s)\sin(\pi x)\sin(\pi y)$. Since $\sin(\pi x)\sin(\pi y)$ is a eigenfunction of L and $2\pi^2$ is the corresponding eigenvalue, the solution should be $u(t,x,y) = t^3 \sin(\pi x)\sin(\pi y)$.

We test the numerical scheme (IV.19) by fixing a uniform triangle mesh with the mesh size $h = 1/128$. We also choose $N = 160$ and the sinc quadrature spacing $k = \sqrt{\pi \mathsf{d}/(sN)}$ with $\mathsf{d} = \pi/8$. This guarantees that the error from the time discretization dominants the total error. Table IV.2 reports the L^2 error between $u(t)$ and $u_h^{\mathcal{N},k}$ for $t = 1$ and $\mathcal{N} = t/\tau$ against the time step τ. The convergence order τ^2 is observed.

τ	$s = 0.3$		$s = 0.5$		$s = 0.7$	
1	3.03×10^{-1}		3.55×10^{-1}		3.91×10^{-1}	
1/2	9.83×10^{-2}	1.62	1.40×10^{-1}	1.34	1.90×10^{-1}	1.04
1/4	2.65×10^{-2}	1.89	4.04×10^{-2}	1.79	6.29×10^{-2}	1.59
1/8	6.94×10^{-3}	1.93	1.05×10^{-2}	1.94	1.72×10^{-2}	1.87
1/16	1.95×10^{-3}	1.83	2.67×10^{-3}	1.98	4.39×10^{-3}	1.97

Table IV.2: L^2 errors and observed rate of convergence for different values of s. The observed error decay in time is in accordance with Theorem IV.2.

IV.4.3 Error Behavior in Time using the Crank-Nicolson Time Stepping Method

We consider solving the same problem as above but using the scheme (IV.29). We again set $h = 1/128$. In terms of the sinc approximation (IV.28), we set $N = 160$ and $k = 1/\sqrt{2\pi \mathsf{d}/(sN)}$ with $\mathsf{d} = \frac{1}{2}\min(\pi/2, \pi/s - \pi)$ so that the error from the quadrature scheme is sufficiently small. Table IV.3 reports the L^2 error between $u(t)$ and $u_h^{n,k}$ for $t = 1$ and $n = t/\tau$ against the time step τ. The convergence order τ^2 is observed.

τ	$s = 0.3$		$s = 0.5$		$s = 0.7$	
1	2.63×10^{-1}		2.97×10^{-1}		3.25×10^{-1}	
1/2	7.00×10^{-2}	1.91	7.97×10^{-2}	1.90	8.69×10^{-2}	1.90
1/4	1.75×10^{-2}	1.99	1.99×10^{-2}	2.00	2.14×10^{-2}	2.02
1/8	4.35×10^{-3}	2.01	4.96×10^{-3}	2.01	5.36×10^{-3}	2.00
1/16	1.06×10^{-3}	2.04	1.21×10^{-3}	2.03	1.33×10^{-3}	2.01

Table IV.3: L^2 errors and observed rate of convergence for different values of s. The observed error decay in time is in accordance with Theorem IV.7.

CHAPTER V

APPROXIMATION OF INTEGRAL FRACTIONAL LAPLACIAN

The goal of this chapter is to study the numerical approximation of solutions of partial differential equations on bounded domains involving the integral fractional Laplacian (I.2). Recall the problem (I.6): given the right hand side data f defined in Ω, we want to find u satisfying

$$(-\Delta)^s \widetilde{u} = f, \qquad \text{in } \Omega, \tag{V.1}$$

where $\widetilde{\cdot}$ denotes the extension by zero. A weak formulation of the above problem is introduced in Section V.1. We will discretize the weak formulation based the Dunford-Taylor integral representation of the bilinear form. The representation is provided in Section V.2. In Section V.3, Section V.4 and Section V.5, the approximation scheme is developed in three steps and consistency error estimates are also provided for each step. We use the final approximation of the bilinear form to generate our discrete problem and show an energy norm error estimate for the approximated solution in (V.6). The $L^2(\Omega)$ error estimate is also discussed assuming that the domain is smooth. Detailed implementation of the approximation scheme together with numerical experiments are provided in Section V.7.

V.1 Test Space and Variational Formulation

V.1.1 The Sobolev spaces $H^r(\mathbb{R}^d)$ and $\widetilde{H}^r(\Omega)$

For $v \in L^2(\mathbb{R}^d)$, the Fourier transform

$$(\mathcal{F}v)(\zeta) := \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-ix\cdot\zeta} v(x)\, dx, \qquad \text{for } \zeta \in \mathbb{R}^d$$

is an invertible mapping form $L^2(\mathbb{R}^d)$ onto $L^2(\mathbb{R}^d)$. It is known that for $r \geq 0$

$$\left(\int_{\mathbb{R}^d} (1 + |\zeta|^2)^r |\mathcal{F}v(\zeta)|^2 \, d\zeta \right)^{1/2}$$

is an equivalent norm for $H^r(\mathbb{R}^d)$ (see e.g. [68]).

For $r \in [0, 2]$, the set of functions in Ω whose extension by zero are in $H^r(\mathbb{R}^d)$ is denoted by $\widetilde{H}^r(\Omega)$. The norm of $\widetilde{H}^r(\Omega)$ is given by $\|\widetilde{\cdot}\|_{H^r(\mathbb{R}^d)}$. Note that for $r \in (0, 1)$ and v in the Schwartz space,

$$((-\Delta)^r v, v) = \int_{\mathbb{R}^d} |\xi|^{2r} |\mathcal{F}v(\xi)|^2 \, d\xi = |c_{d,r}| \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{(v(x) - v(y))^2}{|x - y|^{d+2r}} \, dx \, dy. \tag{V.2}$$

Thus, we prefer to use

$$|v|_{\widetilde{H}^r(\Omega)} = \left(\int_{\mathbb{R}^d} |\xi|^{2r} |\mathcal{F}\widetilde{v}(\xi)|^2 \, d\xi \right)^{1/2} = \left(|c_{d,r}| \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{(\widetilde{v}(x) - \widetilde{v}(y))^2}{|x - y|^{d+2r}} \, dx \, dy \right)^{1/2} \tag{V.3}$$

as the norm on $\widetilde{H}^r(\Omega)$ for $r \in (0, 1)$. This is justified upon invoking a variant of the Peetre-Tartar compactness argument. Suppose that the sequence $\{u_n\}_{n=1}^{\infty} \subset \widetilde{H}^r(\Omega)$ satifsies $\|u_n\|_{\widetilde{H}^r(\Omega)} = 1$ but $\|u_n\|_{L^2(\Omega)} > n|u_n|_{\widetilde{H}^r(\Omega)}$. Since $\widetilde{H}^r(\Omega) \subset H^r(\Omega)$ and the injection of $H^r(\Omega)$ into $L^2(\Omega)$ is compact (cf. [37, Theorem 1.4.5.2 and 1.4.3.2]), without loss of generality by passing to a subsequence, u_n convergences in $L^2(\Omega)$ as $n \to \infty$. This implies that $\{u_n\}$ is a Cauchy sequence in $\widetilde{H}^r(\Omega)$ and we set the limit to be v. Notice that

$$0 = |v|_{\widetilde{H}^r(\Omega)} = \left(\int_{\mathbb{R}^d} |\xi|^{2r} |\mathcal{F}\widetilde{v}(\xi)|^2 \, d\xi \right)^{1/2},$$

$\mathcal{F}\widetilde{v} = 0$ and hence $v = 0$, which contradicts to the assumption $\|u_n\|_{\widetilde{H}^r(\Omega)} = 1$. Therefore, $\|v\|_{L^2(\Omega)} \leq C|v|_{\widetilde{H}^r(\Omega)}$ for $u \in \widetilde{H}^r(\Omega)$ and the full norm of $\widetilde{H}^r(\Omega)$ is equivalent its semi norm.

V.1.2 The Variational Formulation

The variational formulation of (V.1) is: find $u \in \widetilde{H}^s(\Omega)$ satisfying

$$a(u, v) = (f, v)_\Omega, \qquad \text{for all } v \in \widetilde{H}^s(\Omega), \tag{V.4}$$

where

$$a(u, v) = \int_{\mathbb{R}^d} [(-\Delta)^{s/2}\widetilde{u}][(-\Delta)^{s/2}\widetilde{v}]\, dx. \tag{V.5}$$

We refer to Section V.7.1 for the description of model problems. According to the discussion in Section V.1.1, the bilinear form $a(\cdot, \cdot)$ is bounded on $\widetilde{H}^s(\Omega) \times \widetilde{H}^s(\Omega)$ and coercive on $\widetilde{H}^s(\Omega)$. Thus, the Lax-Milgram Theory guarantees existence and uniqueness.

V.1.3 More Notations

We define the Dirichlet form on $H^1(\Omega) \times H^1(\Omega)$ to be

$$d_\Omega(\eta, \phi) := \int_\Omega \nabla\eta \cdot \nabla\phi\, dx.$$

According to Section II.2, we denote T_Ω the solution operator associated with the form $(v, w)_\Omega + d_\Omega(v, w)$ for $v, w \in H^1_0(\Omega)$. This means that for $F \in H^{-1}(\Omega)$, $\theta = T_\Omega F \in H^1_0(\Omega)$ solves

$$(\theta, \phi)_\Omega + d_\Omega(\theta, \phi) = \langle F, \phi\rangle, \qquad \text{for all } \phi \in H^1_0(\Omega). \tag{V.6}$$

Set L_Ω to be the inverse of T_Ω and define the dotted spaces $\dot{H}^r(\Omega)$ with $r \in [-1, 2]$ is defined using the unbounded operator L_Ω.

Remark V.1 (Norm equivalence for Lipschitz domains)**.** *For $r \in (1, 3/2)$, it is known that $\widetilde{H}^r(\Omega) = H^r(\Omega) \cap H^1_0(\Omega) = \mathbb{H}^r(\Omega)$. On the other hand, we note that when $\partial\Omega$ is Lipschitz, $-\Delta$ is an isomorphism from $\mathbb{H}^r(\Omega)$ to $\mathbb{H}^{r-2}(\Omega)$, i.e. Assumption II.1 holds for $\alpha \in (0, 1/2)$; see [42, Theorem 0.5(b)]. We apply this regularity result into Proposition II.2 to obtain $\mathbb{H}^r(\Omega) = \dot{H}^r(\Omega)$. Since $\widetilde{H}^r(\Omega)$ coincides with $\mathbb{H}^r(\Omega)$ for $r \in [0, 1]$ (see e.g. [23, Lemma 4.11]), the norms of $\widetilde{H}^r(\Omega)$ and*

$\dot{H}^r(\Omega)$ *are equivalent for* $r \in [0, 3/2)$ *and the equivalence constant may depend on* Ω*. In this chapter, we use* $\widetilde{H}^r(\Omega)$ *to describe the smoothness of functions defined in* Ω*. When functions defined on a larger domain (see Section V.4 and V.5), we will use these interpolation spaces separately so that we can investigate the dependency of constants.*

V.2 An Alternative Integral Representation of the Bilinear Form

The goal of this section is to derive the Dunford-Taylor integral representation of the bilinear form $a(\cdot, \cdot)$ and some of its properties.

Theorem V.1 (Equivalent Representation)**.** *Let* $s \in (0, 1)$ *and* $0 \le r \le s$*. For* $\eta \in H^{s+r}(\mathbb{R}^d)$ *and* $\theta \in H^{s-r}(\mathbb{R}^d)$*,*

$$((-\Delta)^{(s+r)/2}\eta, (-\Delta)^{(s-r)/2}\theta)_{\mathbb{R}^d} = c_s \int_0^\infty t^{2-2s}(-\Delta(I - t^2\Delta)^{-1}\eta, \theta)\, \frac{dt}{t}, \qquad \text{(V.7)}$$

where

$$c_s := \left(\int_0^\infty \frac{y^{1-2s}}{1 + y^2}\, dy \right)^{-1} = \frac{2 \sin(\pi s)}{\pi}. \qquad \text{(V.8)}$$

Proof. Let $I(\eta, \theta)$ denotes the right hand side of (V.7). Parseval's theorem implies that

$$(-\Delta(I - t^2\Delta)^{-1}\eta, \theta) = \int_{\mathbb{R}^d} \frac{|\zeta|^2}{1 + t^2|\zeta|^2} \mathcal{F}(\eta)(\zeta)\overline{\mathcal{F}(\theta)(\zeta)}\, d\zeta. \qquad \text{(V.9)}$$

and so

$$I(\eta, \theta) = c_s \int_0^\infty t^{1-2s} \int_{\mathbb{R}^d} \frac{|\zeta|^2}{1 + t^2|\zeta|^2} \mathcal{F}(\eta)(\zeta)\overline{\mathcal{F}(\theta)(\zeta)}\, d\zeta\, dt. \qquad \text{(V.10)}$$

In order to invoke Fubini's theorem, we now show that

$$c_s \int_{\mathbb{R}^d} \int_0^\infty t^{1-2s} \frac{|\zeta|^2}{1 + t^2|\zeta|^2} |\mathcal{F}(\eta)(\zeta)|\, |\mathcal{F}(\theta)(\zeta)|\, d\zeta\, dt < \infty.$$

Indeed, the change of variable $y = t|\zeta|$ and the definition (V.8) of c_s implies that the above integral is equal to

$$c_s \int_{\mathbb{R}^d} |\mathcal{F}(\eta)(\zeta)||\mathcal{F}(\theta)(\zeta)| \int_0^\infty t^{1-2s} \frac{|\zeta|^2}{1+t^2|\zeta|^2}\, dt\, d\zeta = \int_{\mathbb{R}^d} |\zeta|^{2s} |\mathcal{F}(\eta)(\zeta)|\, |\mathcal{F}(\theta)(\zeta)|\; d\zeta,$$

which is finite for $\eta \in H^r(\mathbb{R}^d)$ and $\theta \in H^{s-r}(\mathbb{R}^d)$. We now apply Fubini's theorem and the same change of variable $y = t|\zeta|$ in (V.10) to arrive at

$$I(\eta, \theta) = \int_{\mathbb{R}^d} |\zeta|^{2s} \mathcal{F}(\eta)(\zeta) \overline{\mathcal{F}(\theta)(\zeta)}\, d\zeta = ((-\Delta)^{(s+r)/2}\eta, (-\Delta)^{(s-r)/2}\theta)_{\mathbb{R}^d}.$$

This completes the proof. □

Theorem V.1 above implies that for η, θ in $\widetilde{H}^s(\Omega)$,

$$a(\eta, \theta) = c_s \int_0^\infty t^{-2s} (w(\widetilde{\eta}, t), \theta)_\Omega \frac{dt}{t}, \tag{V.11}$$

where for $\psi \in L^2(\mathbb{R}^d)$

$$w(t) := w(\psi, t) := -t^2 \Delta (I - t^2 \Delta)^{-1} \psi.$$

Examining the Fourier transform of $w(\psi, t)$, we realize that $w(t) := w(\psi, t) := \psi + v(\psi, t)$ where $v(t) := v(\psi, t) \in H^1(\mathbb{R}^d)$ solves

$$(v(t), \phi)_{\mathbb{R}^d} + t^2 d_{\mathbb{R}^d}(v(t), \phi) = -(\psi, \phi)_{\mathbb{R}^d}, \qquad \text{for all } \phi \in H^1(\mathbb{R}^d). \tag{V.12}$$

The integral in (V.11) is the basis of our numerical method for (V.4). The following lemma, instrumental in our analysis, provides an alternative characterization for the inner product appearing on the right hand side of (V.11).

Lemma V.2. *Let η be in $L^2(\mathbb{R}^d)$. Then,*

$$(w(\eta, t), \eta)_{\mathbb{R}^d} = \inf_{\theta \in H^1(\mathbb{R}^d)} \{\|\eta - \theta\|^2 + t^2 d_{\mathbb{R}^d}(\theta, \theta)\} =: K(\eta, t). \tag{V.13}$$

Proof. Let η be in $L^2(\mathbb{R}^d)$. We start by observing that for any positive t and $\zeta \in \mathbb{R}^d$,

$$\hat{\phi}(\zeta) := \frac{\mathcal{F}(\eta)(\zeta)}{1 + t^2|\zeta|^2}$$

solves the minimization problem

$$\inf_{z\in\mathbb{C}} \{|\mathcal{F}(\eta)(\zeta) - z|^2 + t^2|\zeta|^2|z|^2\}$$

and so

$$\inf_{z\in\mathbb{C}} \{|\mathcal{F}(\eta)(\zeta) - z|^2 + t^2|\zeta|^2|z|^2\} = \frac{t^2|\zeta|^2}{1 + t^2|\zeta|^2}|\mathcal{F}(\eta)(\zeta)|^2. \tag{V.14}$$

We denote ϕ to be the inverse Fourier transform of $\hat{\phi}$. Note that ϕ is in $H^1(\mathbb{R}^d)$ (actually, ϕ is in $H^2(\mathbb{R}^d)$). Applying the Fourier transform, we find that

$$K(\eta, t) = \inf_{\theta\in H^1(\mathbb{R}^d)} \int_{\mathbb{R}^d} (|\mathcal{F}(\eta)(\zeta) - \mathcal{F}(\theta)(\zeta)|^2 + t^2|\zeta|^2|\mathcal{F}(\theta)(\zeta)|^2)\, d\zeta. \tag{V.15}$$

Now, ϕ is the pointwise minimizer of the integrand in (V.15) and since $\phi \in H^1(\mathbb{R}^d)$, it is also the minimizer of (V.13). In addition, (V.14), (V.15) and (V.9) imply that

$$K(\eta, t) = \int_{\mathbb{R}^d} \frac{t^2|\zeta|^2}{1 + t^2|\zeta|^2}|\mathcal{F}(\eta)(\zeta)|^2\, d\zeta = (w(\eta, t), \eta)_{\mathbb{R}^d}.$$

This completes the proof of the lemma. □

Remark V.2 (Relation with the vanishing Dirichlet boundary condition case)**.** *The above lemma implies that for* $\eta \in \widetilde{H}^s(\Omega)$*,*

$$a(\eta, \eta) = c_s \int_0^\infty t^{-2s} K(\widetilde{\eta}, t)\, \frac{dt}{t}.$$

On the other hand,

$$c_s \int_0^\infty t^{-2s} K_\Omega^0(\eta, t)\, \frac{dt}{t}$$

with

$$K_\Omega^0(\eta, t) := \inf_{\theta \in H_0^1(\Omega)} \{\|\eta - \theta\|_{L^2(\Omega)}^2 + t^2 d_\Omega(\theta, \theta)\} \tag{V.16}$$

is an equivalent norm for $\eta \in \dot{H}^s(\Omega)$*. Let* $\{\psi_i^0\} \subset H_0^1(\Omega)$ *denote the* $L^2(\Omega)$*-orthonormal basis of eigenfunctions satisfying*

$$d_\Omega(\psi_i^0, \theta) = \lambda_i(\psi_i^0, \theta)_\Omega, \qquad \textit{for all } \theta \in H_0^1(\Omega).$$

As the proof in Lemma V.2 but using the expansion in the above eigenfunctions, it is not hard to see that

$$(w_\Omega(\eta, t), \eta)_\Omega = K_\Omega^0(\eta, t) \tag{V.17}$$

with $w_\Omega(\eta, t) = \eta + v$ *and* $v \in H_0^1(\omega)$ *solving*

$$(v, \theta)_\Omega + t^2 d_\Omega(v, \theta) = -(u, \theta)_\Omega, \qquad \textit{for all } \theta \in H_0^1(\Omega).$$

This means that if $\eta \in L^2(\Omega)$*,* $K(\widetilde{\eta}, t) \le K_\Omega^0(\eta, t)$ *and hence*

$$(w(\widetilde{\eta}, t), \eta)_\Omega \le (w_\Omega(\eta, t), \eta)_\Omega.$$

V.3 Approximation of the Bilinear Form: Sinc Approximation

In this section, we analyze a sinc quadrature scheme applied to the integral (V.11).

V.3.1 The Sinc Quadrature Scheme

We first use the change of variable $t^{-2} = e^y$ so that (V.11) becomes

$$a(\eta, \theta) = \frac{c_s}{2} \int_{-\infty}^{\infty} e^{sy} (w(\widetilde{\eta}, t(y)), \theta)_\Omega \, dy.$$

Given a quadrature spacing $k > 0$ and two positive integers N^- and N^+, set $y_j := jk$ so that

$$t_j = e^{-y_j/2} = e^{-jk/2} \tag{V.18}$$

and define the approximation of $a(\eta, \theta)$ by

$$a^k(\eta, \theta) := \frac{c_s k}{2} \sum_{j=-N^-}^{N^+} e^{sy_j} (w(\widetilde{\eta}, t_j), \theta)_\Omega. \tag{V.19}$$

V.3.2 Consistency Bound

The convergence of the sinc quadrature depends on the properties of the integrand

$$g(y; \eta, \theta) := e^{sy} (w(\widetilde{\eta}, t(y)), \theta)_\Omega = e^{sy} \left(-\Delta (e^y I - \Delta)^{-1} \widetilde{\eta}, \widetilde{\theta} \right)_{\mathbb{R}^d}. \tag{V.20}$$

More precisely, we shall show that $g(y; \eta, \theta) \in S(B_{\mathsf{d}})$ with $\mathsf{d} = \pi/4$. In our context, this leads to the following estimates for the sinc quadrature error.

Theorem V.3 (Sinc quadrature consistency)**.** *Let* $\mathsf{d} = pi/4$. *Suppose* $\theta \in \widetilde{H}^s(\Omega)$ *and* $\eta \in \widetilde{H}^\delta(\Omega)$ *with* $\delta \in (s, 2-s]$. *Let* $a(\cdot, \cdot)$ *and* $a^k(\cdot, \cdot)$ *be defined by* (V.4) *and* (V.19)*, respectively. Then we have*

$$\begin{aligned} |a(\eta, \theta) - a^k(\eta, \theta)| \le {} & \frac{N(B_{\mathsf{d}})}{e^{\pi^2/(2k)} - 1} \\ & + \frac{2\sqrt{2}}{\delta - s} e^{(s-\delta)N^+ k/2} \|\eta\|_{\widetilde{H}^\delta(\Omega)} \|\theta\|_{\widetilde{H}^s(\Omega)} \\ & + \frac{\sqrt{2}}{s} e^{-sN^- k} \|\eta\|_{L^2(\Omega)} \|\theta\|_{L^2(\Omega)}. \end{aligned} \tag{V.21}$$

Proof. We start by showing that the conditions (a), (b) and (c) of Definition II.1 hold. For (a), we note that $g(\cdot; \eta, \theta)$ in analytic on B_{d} if and only if the operator mapping $z \mapsto (e^z I - \Delta)^{-1}$ is analytic on B_{d}. To see the latter, we fix $z_0 \in B$ and set $p_0 := e^{z_0}$. Clearly, $p_0 I - \Delta$ is invertible from $L^2(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$. Let $M_0 := \|(p_0 I - \Delta)^{-1}\|_{L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)}$. For $p \in \mathbb{C}$, we write

$$pI - \Delta = (p - p_0)I + (p_0 I - \Delta) = (p_0 I - \Delta) \left((p - p_0)(p_0 I - \Delta)^{-1} + I \right),$$

so that the Neumann series representation

$$(pI - \Delta)^{-1} = \left(\sum_{j=0}^{\infty} (-1)^j (p - p_0)^j (p_0 I - \Delta)^{-j} \right) (p_0 I - \Delta)^{-1}$$

is uniformly convergent provided $\|(p - p_0)(p_0 I - \Delta)^{-1}\|_{L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)} < 1$ or

$$|p - p_0| < 1/M_0.$$

Hence $(pI - \Delta)^{-1}$ is analytic in an open neighborhood of $p_0 = e^{z_0}$ for all $p_0 \in B$ and (a) follows.

To prove (b) and (c), we first bound $g(z; \eta, \theta)$ for z in the band B_{d}. Assume $\eta \in \widetilde{H}^\delta(\Omega)$ and $\theta \in \widetilde{H}^s(\Omega)$ with $\delta \in (s, 2-s]$. For $z \in B$, we use the Fourier transform and estimate $|g|$ as follows

$$\begin{aligned} |g(z; \eta, \theta)| &= \left| e^{sz} \int_{\mathbb{R}^d} \frac{|\zeta|^2}{e^z + |\zeta|^2} \mathcal{F}(\widetilde{\eta}) \overline{\mathcal{F}(\widetilde{\theta})} \, d\zeta \right| \\ &\leq \sqrt{2} e^{s \mathfrak{Re} z} \int_{\mathbb{R}^d} \frac{|\zeta|^2}{e^{\mathfrak{Re} z} + |\zeta|^2} |\mathcal{F}(\widetilde{\eta})| |\mathcal{F}(\widetilde{\theta})| \, d\zeta. \end{aligned}$$

Here we used the fact that $|e^z + |\zeta|^2| \geq \mathfrak{Re} e^z + |\zeta|^2 \geq (e^{\mathfrak{Re} z} + |\zeta|^2)/\sqrt{2}$. If $\mathfrak{Re} z < 0$, we deduce that

$$|g(z; \eta, \theta)| \leq \sqrt{2} e^{s \mathfrak{Re} z} \|\eta\|_{L^2(\Omega)} \|\theta\|_{L^2(\Omega)}. \tag{V.22}$$

Instead, when $\mathfrak{Re} z \geq 0$, we write

$$|g(z; \eta, \theta)| \leq \sqrt{2} e^{(s-\delta)\mathfrak{Re} z/2} \int_{\mathbb{R}^d} \frac{(|\zeta|^2)^{1-(\delta+s)/2} (e^{\mathfrak{Re} z})^{(\delta+s)/2}}{e^{\mathfrak{Re} z} + |\zeta|^2} |\zeta|^{\delta+s} |\mathcal{F}(\widetilde{\eta})| |\mathcal{F}(\widetilde{\theta})| \, d\zeta.$$

Whence, Young's inequality guarantees that

$$|g(z; \eta, \theta)| \leq \sqrt{2} e^{(s-\delta)\mathfrak{Re} z/2} \|\eta\|_{\widetilde{H}^\delta(\Omega)} \|\theta\|_{\widetilde{H}^s(\Omega)}. \tag{V.23}$$

Gathering the above two estimates (V.22) and (V.23) gives

$$\int_{-d}^{d} |g(y+iw;\eta,\theta)|\,dw \leq \begin{cases} \dfrac{\pi}{\sqrt{2}}\|\eta\|_{L^2(\Omega)}\|\theta\|_{L^2(\Omega)}, & y<0, \\ \dfrac{\pi}{\sqrt{2}}\|\eta\|_{\widetilde{H}^\delta(\Omega)}\|\theta\|_{\widetilde{H}^s(\Omega)}, & y \geq 0, \end{cases} \tag{V.24}$$

and

$$N(B_{\mathsf{d}}) \leq \frac{4\sqrt{2}}{\delta - s}\|\eta\|_{\widetilde{H}^\delta(\Omega)}\|\theta\|_{\widetilde{H}^s(\Omega)} + \frac{2\sqrt{2}}{s}\|\eta\|_{L^2(\Omega)}\|\theta\|_{L^2(\Omega)}. \tag{V.25}$$

Estimates (V.24) and (V.25) prove (b) and (c) respectively.

Having established (a), (b), and (c), we can use the sinc quadrature estimate (II.23). In addition, from (V.22) and (V.23) we also deduce that

$$\begin{aligned} k\sum_{j\leq -N^- -1}^{-\infty} |g(kj;\eta,\theta)| &\leq \frac{\sqrt{2}}{s}e^{-sN^-k}\|\eta\|_{L^2(\Omega)}\|\theta\|_{L^2(\Omega)} \qquad \text{and} \\ k\sum_{j\geq N^+ +1}^{\infty} |g(kj;\eta,\theta)| &\leq \frac{2\sqrt{2}}{\delta - s}e^{(s-\delta)N^+k/2}\|\eta\|_{\widetilde{H}^\delta(\Omega)}\|\theta\|_{\widetilde{H}^s(\Omega)}. \end{aligned} \tag{V.26}$$

Combining (II.23) and (V.26) shows (V.21) and completes the proof. □

Remark V.3 (Choice of N^- and N^+)**.** *Balancing the three exponentials in* (V.21) *leads to the following choice*

$$\pi^2/(2k) \approx (\delta - s)N^+k/2 \approx sN^-k.$$

Hence, for given the quadrature spacing $k>0$, we set

$$N^+ := \left\lceil \frac{\pi^2}{k^2(\delta - s)} \right\rceil \qquad \textit{and} \qquad N^- := \left\lceil \frac{\pi^2}{2sk^2} \right\rceil. \tag{V.27}$$

With this choice, (V.21) *becomes*

$$|a(\eta,\theta) - a^k(\eta,\theta)| \leq \gamma(k)\|\eta\|_{\widetilde{H}^\delta(\Omega)}\|\theta\|_{\widetilde{H}^s(\Omega)} \tag{V.28}$$

where

$$\gamma(k) := 3\left(\frac{2\sqrt{2}}{\delta - s} + \frac{\sqrt{2}}{s}\right) e^{-\pi^2/(2k)}. \tag{V.29}$$

V.4 Approximation of the Bilinear Form: Domain Truncation

Let B be a convex bounded domain containing Ω and the origin. Without loss of generality, we assume that the diameter of B is 1. This auxiliary domain is used to generate suitable truncation domains to approximate the solution of (V.12). We introduce a domain parameter $M > 0$ and define the dilated domains

$$B^M(t) := \begin{cases} \{y = (1 + t(1+M))x \;:\; x \in B\}, & t \geq 1, \\ \{y = (2+M)x \;:\; x \in B\}, & t < 1. \end{cases} \tag{V.30}$$

The approximation of $a^k(\cdot,\cdot)$ in (V.19) reads

$$a^{k,M}(\eta,\theta) := \frac{c_s k}{2} \sum_{j=-N^-}^{N^+} e^{\beta y_j} (w^M(\widetilde{\eta}, t_j), \theta)_\Omega, \tag{V.31}$$

with $t_j := t(y_j) = e^{-y_j/2}$, according to (V.18), and

$$w^M(t) := w^M(\widetilde{\eta}, t) = \widetilde{\eta}|_{B^M(t)} + v^M(\widetilde{\eta}, t), \tag{V.32}$$

where $v^M(t) := v^M(\widetilde{\eta}, t)$ solves

$$(v^M(t), \phi)_{B^M(t)} + t^2 d_{B^M(t)}(v^M(t), \phi) = -(\eta, \phi)_\Omega, \qquad \text{for all } \phi \in H_0^1(B^M(t)); \tag{V.33}$$

compare with (V.12). The domains $B^M(t_j)$ are constructed for the truncation error to be exponentially decreasing as a function of M. This is the subject of next section.

V.4.1 Consistency

The main result of this section provides an estimate for $a^k - a^{k,M}$. It relies on decay properties of $v(\widetilde{\eta}, t)$ satisfying (V.12). In fact, Lemma 2.1 of [5] guarantees the existence of universal constants c and C such that

$$t\|\nabla v(\widetilde{\eta}, t)\|_{L^2(A^M(t))} + \|v(\widetilde{\eta}, t)\|_{L^2(A^M(t))} \leq Ce^{-\max(1,t)cM/t}\|\eta\|_{L^2(\Omega)}, \tag{V.34}$$

provided $\eta \in L^2(\Omega)$ and $v(t) := v(\widetilde{\eta}, t)$ is given in (V.12). Here

$$A^M(t) := \{x \in A^M(t) : \text{dist}(x, \partial A^M(t)) < t\}$$

so that the minimal distance between points in $\Omega \subset B$ and $A^M(t)$ is greater than $M \max(1, t)$. An illustration of the different domains is provided in Figure V.1.

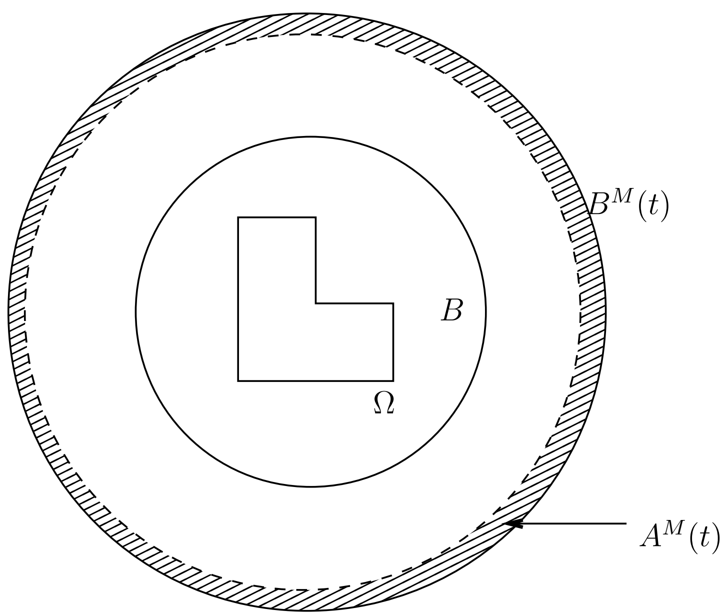

Figure V.1: Illustration of the different domains in $\mathbb{R}^2$. The domain of interest Ω is a L-shaped domain, $B \subset B^M(t)$ are interior of discs, and $A^M(t)$ is the filled portion of $B^M(t)$.

Lemma V.4 (Truncation error)**.** *Let* $\eta \in L^2(\Omega)$, $e(t) := v(\tilde{\eta}, t) - v^M(\tilde{\eta}, t)$ *and* c *be the constant appearing in* (V.34)*. There is a positive constant* C *not depending on* M *and* t *satisfying*

$$\|e(t)\|_{L^2(B^M(t))} \leq Ce^{-\max(1,t)cM/t}\|\eta\|_{L^2(\Omega)}. \tag{V.35}$$

Proof. In this proof, C denotes a generic constant only depending on Ω. Note that $e(t)$ satisfies the relations

$$\begin{aligned} (e(t), \phi)_{B^M(t)} + t^2 d_{B^M(t)}(e(t), \phi) = 0, \quad \forall \phi \in H_0^1(B^M(t)), \\ e(t) = v(t), \quad \text{on } \partial B^M(t). \end{aligned} \tag{V.36}$$

Let $\chi(t) \geq 0$ be a bounded cut off function satisfying $\chi(t) = 1$ on $\partial B^M(t)$ and $\chi(t) = 0$ on $B^M(t) \setminus A^M(t)$. Without loss of generality, we may assume that $\|\nabla \chi(t)\|_{L^\infty(\mathbb{R}^d)} \leq C/t$. This implies

$$\begin{aligned} \|\chi(t)v(t)\|_{L^2(B^M(t))} &+ t\|\nabla(\chi(t)v(t))\|_{L^2(B^M(t))} \\ &\leq C(\|v(t)\|_{L^2(A^M(t))} + t\|\nabla v(t)\|_{L^2(A^M(t))}) \\ &\leq Ce^{-\max(1,t)cM/t}\|\eta\|_{L^2(\Omega)}. \end{aligned}$$

Here we use the decay estimate (V.34) for last inequality above. Now, setting $e(t) := \chi(t)v(t) + \zeta(t)$, we find that $\zeta(t) \in H_0^1(B^M(t))$ satisfies

$$(\zeta(t), \phi)_{B^M(t)} + t^2 d_{B^M(t)}(\zeta(t), \phi) = -(\chi(t)v(t), \phi)_{B^M(t)} - t^2 d_{B^M(t)}(\chi(t)v(t), \phi)$$

for all $\phi \in H_0^1(B^M(t))$. Taking $\phi = \zeta(t)$, we deduce that

$$\begin{aligned} \|\zeta(t)\|^2_{L^2(B^M(t))} + t^2\|\nabla \zeta(t)\|^2_{L^2(B^M(t))} &\leq \|\chi(t)v(t)\|^2_{L^2(A^M(t))} + t^2\|\nabla(\chi(t)v(t))\|^2_{L^2(A^M(t))} \\ &\leq Ce^{-2\max(1,t)cM/t}\|\eta\|^2_{L^2(\Omega)}. \end{aligned}$$

Thus, combining the estimates for $\zeta(t)$ and $\chi(t)v(t)$ completes the proof. □

Lemma V.4 above is instrumental to derive exponentially decaying consistency error as $M \to$

∞. Indeed, we have the following theorem.

Theorem V.5 (Truncation error)**.** *Let c be the constant appearing in* (V.34) *and assume $M > 2(s+1)/c$. Then, there is a positive constant C not depending on M nor k satisfying*

$$|a^k(\eta,\theta) - a^{k,M}(\eta,\theta)| \le Ce^{-cM}\|\eta\|_{L^2(\Omega)}\|\theta\|_{L^2(\Omega)}, \qquad \textit{for all } \eta,\theta \in L^2(\Omega). \tag{V.37}$$

Proof. In this proof C denotes a generic constant only depending on Ω. Let η, θ be in $L^2(\Omega)$. It suffices to bound

$$\begin{aligned} E &:= \left| \frac{c_s k}{2} \sum_{j=-N^-}^{N^+} e^{sy_j}(w(t_j) - w^M(t_j), \theta)_\Omega \right| \\ &\le C\left(k \sum_{j=-N^-}^{-1} e^{sy_j} |(v(t_j) - v^M(t_j), \theta)_\Omega| + k \sum_{j=0}^{N^+} e^{sy_j} |(v(t_j) - v^M(t_j), \theta)_\Omega| \right) \\ &=: E_1 + E_2 \end{aligned}$$

with $v(t) = v(\widetilde{\eta}, t)$ defined by (V.12) and $v^M(t) = v^M(\widetilde{\eta}, t)$ defined by (V.33). We estimate E_1 and E_2 separately, starting with E_1.

From the definition $t_j = e^{-y_j/2}$, we deduce that when $j < 0$, $t_j > 1$ so that (V.35) gives

$$\begin{aligned} E_1 &\le Cke^{-cM} \sum_{j=-N^-}^{-1} e^{sy_j} \|\eta\|_{L^2(\Omega)}\|\theta\|_{L^2(\Omega)} \\ &\le Ce^{-cM} \frac{ke^{-sk}}{1-e^{-sk}} \|\eta\|_{L^2(\Omega)}\|\theta\|_{L^2(\Omega)} \le Ce^{-cM}\|\eta\|_{L^2(\Omega)}\|\theta\|_{L^2(\Omega)}. \end{aligned}$$

Similarly, for $j \geq 0$, i.e. $t_j < 1$, using (V.35) again, we have

$$
\begin{aligned}
E_2 &\leq Ck \sum_{j=0}^{N^+} e^{sy_j} e^{-cM/t_j} \|\eta\|_{L^2(\Omega)} \|\theta\|_{L^2(\Omega)} \\
&\leq Ck \sum_{j=0}^{N^+} e^{sy_j} e^{-cM(1+y_j/2)} \|\eta\|_{L^2(\Omega)} \|\theta\|_{L^2(\Omega)} \\
&= Cke^{-cM} \sum_{j=0}^{N^+} e^{(s-cM/2)y_j} \|\eta\|_{L^2(\Omega)} \|\theta\|_{L^2(\Omega)} \\
&\leq Ce^{-cM} \frac{k}{1-\exp(k(s-cM/2))} \|\eta\|_{L^2(\Omega)} \|\theta\|_{L^2(\Omega)} \\
&\leq \frac{Ce^{-cM}}{cM/2 - s} \|\eta\|_{L^2(\Omega)} \|\theta\|_{L^2(\Omega)} \leq Ce^{-cM} \|\eta\|_{L^2(\Omega)} \|\theta\|_{L^2(\Omega)},
\end{aligned}
$$

where we have also used the property $cM/2 - s > 1$ guaranteed by the assumption $M > 2(s+1)/c$. □

V.4.2 Uniform Norm Equivalence on Convex Domains

Since the domains $B^M(t)$ are convex, the regularity index α in Assumption II.1 is 1. Thus the norms in $\dot{H}^r(B^M(t))$ are equivalent to those in $H^r(B^M(t)) \cap H^1_0(B^M(t))$ for $r \in [1,2]$. However, as we mentioned in Remark V.1, the equivalence constants depend *a priori* on $B^M(t)$ and therefore on M and t. We show in this section that they can be bounded uniformly independently of both parameters.

To simplify the notation introduced in Section V.1 and Section II.2. We shall denote $T_{\Omega^M(t)}$ by T_t, $L_{B^M(t)}$ by L_t and $\dot{H}^s(B^M(t))$ by $\dot{H}^s$. We recall that $B^M(t)$ is a dilatation of the convex and bounded domain Ω containing the origin, see (V.30). We then have the following lemma.

Lemma V.6 (Ellipitic Regularity on Convex Domains)**.** *Let $f \in L^2(B^M(t))$. Then $\theta := T_t f$ is in $H^2(B^M(t)) \cap H^1_0(B^M(t))$ and satisfies*

$$\|\theta\|_{H^2(B^M(t))} \leq C\|f\|_{L^2(B^M(t))}, \tag{V.38}$$

where C is a constant independent of t and M.

Proof. It is well known that the convexity of B and hence that of $B^M(t)$ implies that $\theta \in H^2(B^M(t)) \cap H_0^1(B^M(t))$. Therefore, the crucial point is to show that the constant in (V.38) does not depend on M or t. To see this, the H^2 elliptic regularity on convex domains implies that for $\hat{\theta} \in H_0^1(B)$ with $\Delta\hat{\theta} \in L^2(B)$ then $\hat{\theta} \in H^2(B)$ and there is a constant C only depending on B such that

$$|\hat{\theta}|_{H^2(B)} \leq C\|\Delta\hat{\theta}\|_{L^2(B)}. \tag{V.39}$$

Let γ be such that $B^M(t) = \{\gamma x,\ x \in B\}$ (see (V.30)) and $\hat{\theta}(\hat{x}) = \theta(\gamma\hat{x})$ for $\hat{x} \in B$. Once scaled back to $B^M(t)$, estimate (V.39) gives

$$|\theta|_{H^2(B^M(t))} \leq C\|\Delta\theta\|_{L^2(B^M(t))} = C\|f - \theta\|_{L^2(B^M(t))}. \tag{V.40}$$

Now $\theta = T_t f$ immediately implies that $\|\theta\|_{H^1(B^M(t)))} \leq \|f\|_{L^2(B^M(t))}$ and (V.38) follows by the triangle inequality and obvious manipulations. □

Remark V.4 (Intermediate Spaces)**.** *Lemma V.6 implies that* $D(L_t) = \dot{H}^2 = H^2(B^M(t)) \cap H_0^1(B^M(t))$ *with norm equivalence constants independent of* M *and* t*. As* $D(L_t^{1/2}) = \dot{H}^1 = H_0^1(B^M(t))$*, for* $r \in [1,2]$

$$\dot{H}^r = [H^1(B^M(t)), H^2(B^M(t)) \cap H_0^1(B^M(t))]_{r-1} = H^r(B^M(t)) \cap H_0^1(B^M(t))$$

with norm equivalence constants independent of M *and* t*.*

Lemma V.7 (Norm Equivalence)**.** *For* $\beta \in [1, 3/2)$*, let* θ *be in* $\dot{H}^\beta$ *and* $\widetilde{\theta}$ *denote its extension by zero outside of* $B^M(t)$*. Then* $\widetilde{\theta}$ *is in* $H^\beta(\mathbb{R}^d)$ *and*

$$\|\theta\|_{\dot{H}^\beta} \leq C\|\widetilde{\theta}\|_{H^\beta(\mathbb{R}^d)}$$

with C *not depending on* t *or* M*.*

Proof. Given $\theta \in H^1(B^M(t))$, we denote $R\theta$ to be the elliptic projection of θ into $H_0^1(B^M(t))$, i.e., $R\theta \in H_0^1(B^M(t))$ is the solution of

$$(R\theta, \phi)_{B^M(t)} + d_{B^M(t)}(R\theta, \phi) \\ = (\theta, \phi)_{B^M(t)} + d_{B^M(t)}(\theta, \phi), \qquad \text{for all } \phi \in H_0^1(B^M(t)).$$

It immediately follows that

$$\|R\theta\|_{\dot{H}^1} = \|R\theta\|_{H^1(B^M(t))} \leq \|\theta\|_{H^1(B^M(t))}.$$

Also, if $\theta \in H^2(B^M(t))$, Lemma V.6 (see also Remark V.4) implies

$$\|R\theta\|_{\dot{H}^2} \leq C\|R\theta\|_{H^2(B^M(t))} \leq C\|\theta\|_{H^2(B^M(t))}$$

with C not depending on t or M. Hence, it follows by interpolation that

$$\|R\theta\|_{\dot{H}^\beta} \leq C_\beta\|\theta\|_{[H^1(B^M(t)),H^2(B^M(t))]_{\beta-1}}. \tag{V.41}$$

Now when $\theta \in \dot{H}^\beta \subset H^1(B^M(t))$, $R\theta = \theta$ so that in view of (V.41), it remains to show that

$$\|\theta\|_{[H^1(B^M(t)),H^2(B^M(t))]_{\beta-1}} \leq C\|\widetilde{\theta}\|_{H^\beta(\mathbb{R}^d)},$$

for a constant C independent of M and t. To see this, note that $\widetilde{\theta}$ is in $H^1(\mathbb{R}^d)$ and the extension of $\nabla\theta$ by zero is in $H^{\beta-1}(\mathbb{R}^d)$. We refer to Theorem 1.4.4.4 of [37] for a proof when $d = 1$ proof and the techniques used in Lemma 4.33 of [32] for the extension to the higher dimensional spaces. This implies that $\widetilde{\theta}$ belongs to $H^\beta(\mathbb{R}^d)$. Moreover, the restriction operator is simultaneously bounded from $H^j(\mathbb{R}^d)$ to $H^j(B^M(t))$ for $j = 1, 2$. Hence, by interpolation again, we have that

$$\|\theta\|_{[H^1(B^M(t)),H^2(B^M(t))]_{\beta-1}} \leq \|\widetilde{\theta}\|_{H^\beta(\mathbb{R}^d)}.$$

This completes the proof of the lemma. □

V.5 Approximation of the Bilinear Form: Finite Element Approximation

In this section, we turn our attention to the finite element approximation of each subproblems (V.33) in $a^{k,M}(\cdot,\cdot)$. Throughout this section, we omit when no confusion is possible the subscript j in t_j, i.e. we consider a generic t keeping in mind that the subsequent statements only hold for $t = t_j$ with $j = -N^-, ..., N^+$. We also make the additional unrestrictive assumption that B used to define $B^M(t)$ (see (V.30)) is polygonal. In turn, so are all the dilated domains $B^M(t)$.

V.5.1 Finite Element Approximation of $a^{k,M}(\cdot,\cdot)$

Based on the notations introduced in Section II.4, we set $\mathcal{T}_h^M(t) := \mathcal{T}_h(B^M(t))$ for $t = t_j$, $j = -N^-, ..., N^+$, given by (V.18). We assume that the conditions (II.10) and (II.11) hold for $\mathcal{T}_h^M(t_j)$ with constants c, ρ not depending on j. We also require that all the subdivisions match on Ω, i.e.

$$\mathcal{T}_h(\Omega) \subset \mathcal{T}_h^M(t_j) \tag{V.42}$$

for each j. We discuss in Section V.7 how to generate subdivisions meeting these requirements. Finally, in terms of the finite element space, we use the short notation $\mathbb{V}_h^M(t) := \mathbb{V}_h(B^M(t))$.

We are now in position to define the fully discrete/implementable problem. For η_h and θ_h in $\mathbb{V}_h(\Omega)$, the finite element approximation of $a^{k,M}(\cdot,\cdot)$ given by (V.31) is

$$a_h^{k,M}(\eta_h, \theta_h) := \frac{c_s k}{2} \sum_{j=-N^-}^{N^+} e^{sy_j} (w_h^M(\widetilde{\eta}_h, t_j), \theta_h)_\Omega \tag{V.43}$$

with

$$w_h^M(\widetilde{\eta}_h, t) := \widetilde{\eta}_h|_{B^M(t)} + v_h^M(t) \tag{V.44}$$

and where $v_h^M(t) \in \mathbb{V}_h^M(t)$ solves

$$(v_h^M(t), \phi_h)_{B^M(t)} + t^2 d_{B^M(t)}(v_h^M(t), \phi_h) = -(\widetilde{\eta}_h, \phi_h)_{B^M(t)}, \quad \forall \phi_h \in \mathbb{V}_h^M(t). \tag{V.45}$$

Remark V.5 (Assumption (V.42))**.** *Two critical properties follow from* (V.42)*. On the one hand,* *our analysis below relies on the fact that the extension by zero* $\widetilde{v}_h$ *of* $v_h \in \mathbb{V}_h(\Omega)$ *belongs to all* $\mathbb{V}_h^M(t)$*. This property greatly simplifies the computation of* $(w_h^M(\widetilde{\eta}_h, t_j), \theta_h)_D$ *in* (V.43).

V.5.2 Approximations on $B^M(t)$

Since $a_h^{k,M}(\cdot,\cdot)$ requires approximations by the finite element methods on domains $B^M(t)$. Standard finite element argumentations would lead to estimates with constants depending on $B^M(t)$ and therefore M and t. In this section, we exhibit results where this is not the case due to the particular definition (V.30) of $B^M(t)$.

We can use interpolation to develop approximation results for functions in the intermediate spaces with constants independent of M and t. The Scott-Zhang interpolation construction [61] gives rise to an approximation operator $\pi_h^{sz} : H_0^1(B^M(t)) \to \mathbb{V}_h^M(t)$ satisfying

$$\|\eta - \pi_h^{sz}\eta\|_{H^1(B^M(t))} \leq C\|\eta\|_{H^1(B^M(t))},$$

for all $\eta \in H_0^1(B^M(t)) = \dot{H}^1$ and

$$\|\eta - \pi_h^{sz}\eta\|_{H^1(B^M(t))} \leq Ch\|\eta\|_{H^2(B^M(t))},$$

for all $\eta \in H^2(B^M(t)) \cap H_0^1(B^M(t)) = \dot{H}^2$. The Scott-Zhang argument is local so the constants appearing above depend on the shape regularity of the triangulations but not on t or M. Interpolating the above inequalities shows that for all $r \in [0,1]$

$$\inf_{\chi \in \mathbb{V}_h^M(t)} \|\eta - \chi\|_{H^1(B^M(t))} \leq Ch^r\|\eta\|_{\dot{H}^{1+r}}, \qquad \text{for all } \eta \in \dot{H}^{1+r} \tag{V.46}$$

with C not depending on t or M.

Let $T_{t,h}$ denote the finite element approximation to T_t, i.e., for $F \in \dot{H}^{-1}$, $T_{t,h}F := w_h$ with $w_h \in {\mathbb{V}_h}^M(t)$ being the unique solution of

$$(w_h, \phi_h)_{B^M(t)} + d_{B^M(t)}(w_h, \phi_h) = \langle F, \phi_h \rangle, \qquad \text{for all } \phi_h \in \mathbb{V}_h^M(t).$$

The approximation result (V.46) and standard finite element analysis techniques implies that for any $r \in [0,1]$,

$$\|T_t F - T_{t,h} F\|_{L^2(B^M(t))} \le Ch^{1+r} \|T_t F\|_{\dot{H}^{1+r}} \le Ch^{1+r} \|F\|_{\dot{H}^{-1+r}}, \tag{V.47}$$

where the last inequality follows from interpolation since $\|T_t F\|_{H^1(B^M(t))} \le \|F\|_{H^{-1}(B^M(t))}$ and (V.38) hold.

For $f \in L^2(B^M(t))$, we define the solution operator associated with the Dirichlet form $d_{B^M(t)}(\cdot,\cdot)$. That is

$$S_t f := \eta \in H_0^1(B^M(t)) \tag{V.48}$$

satisfying,

$$d_{B^M(t)}(\eta, \phi) = (f, \phi)_{B^M(t)}, \qquad \text{for all } \phi \in H_0^1(B^M(t))$$

and let $S_{t,h} f \in \mathbb{V}_h^M(t)$ denote its finite element approximation; compare with T_t and $T_{h,t}$. Although the Poincaré constant depends on the diameter of $B^M(t)$, we still have the following lemma.

Lemma V.8. *There is a constant C independent of h, t, or M satisfying*

$$\|S_t f - S_{t,h} f\|_{L^2(B^M(t))} \le Ch^2 \|f\|_{L^2(B^M(t))}.$$

Proof. For $f \in L^2(B^M(t))$, set $e_h := (S_t - S_{t,h})f$. The elliptic regularity estimate (V.40) on convex domain and Cea's Lemma imply

$$\begin{aligned} |e_h|_{H^1(B^M(t))} &= \inf_{\chi_h \in \mathbb{V}_h^M(t)} |S_t f - \chi_h|_{H^1(B^M(t))} \le Ch |S_t f|_{H^2(B^M(t))} \\ &\le Ch \|\Delta S_t f\|_{L^2(B^M(t))} = Ch \|f\|_{L^2(B^M(t))}, \end{aligned}$$

where C is a constant independent of h, t and M. Galerkin orthogonality and the above estimate give

$$\|e_h\|^2_{L^2(B^M(t))} = d_{B^M(t)}(e_h, S_t e_h) = d_{B^M(t)}(e_h, (S_t - S_{t,h})e_h)$$
$$\leq |e_h|_{H^1(B^M(t))} |(S_t - S_{t,h})e_h|_{H^1(B^M(t))}$$
$$\leq Ch|e_h|_{H^1(B^M(t))} \|e_h\|_{L^2(B^M(t))}.$$

Combining the above two inequalities and obvious manipulations completes the proof of the lemma. □

We shall also need norm equivalency on discrete scales. Set $\dot{H}^r_h := \dot{H}^r_h(B^M(t))$ for r$\in [-1,2]$. Recall from (II.18) and Remark II.2 that for $r \in [0, 3/2)$, there exists a constant c and C independent of h satisfying

$$c\|v_h\|_{\dot{H}^r_h} \leq \|v_h\|_{\dot{H}^r} \leq \|v_h\|_{\dot{H}^r_h}, \qquad \text{for all } v_h \in \mathbb{V}^M_h(t). \tag{V.49}$$

Lemma V.6 guarantees that the constants above also independent of M and t. The spaces for negative r are defined by duality and the stability of the $L^2(B^M(t))$-projection π_h yields for $r \in [0, 1]$,

$$c\|v_h\|_{\dot{H}^{-r}} \leq \|v_h\|_{\dot{H}^{-r}_h} \leq \|v_h\|_{\dot{H}^{-r}}. \qquad \text{for all } v_h \in \mathbb{V}^M_h(t). \tag{V.50}$$

V.5.3 Consistency

We are now ready to estimate the consistency error between $a^{k,M}(\cdot,\cdot)$ and $a^{k,M}_h(\cdot,\cdot)$ on $\mathbb{V}_h(\Omega)$. Its decay depends on a parameter $\beta \in (s, 3/2)$, which will be related later to the regularity of the solution u to (V.4).

Theorem V.9 (Finite element consistency)**.** *Let* $\beta \in (s, 3/2)$ *and* $\alpha \in [s, \min(2s, s + 1/2)]$. *We assume that the quadrature parameters* N^- *and* N^+ *are chosen according to* (V.27). *There exists a constant* C *independent of* h*,* k *and* M *satisfying*

$$|a^{k,M}(\eta_h, \theta_h) - a^{k,M}_h(\eta_h, \theta_h)| \leq C(1 + \ln(h^{-1}))h^{\beta+\alpha-2s}\|\eta_h\|_{\widetilde{H}^\beta(\Omega)}\|\theta_h\|_{\widetilde{H}^\alpha(\Omega)} \tag{V.51}$$

for all $\eta_h, \theta_h \in \mathbb{V}_h(\Omega)$.

Proof. In this proof, C denotes a generic constant independent of h, M, k and t.

Fix $\eta_h \in \mathbb{V}_h(\Omega)$ and denote by $\tilde{\eta}_h$ its extension by zero outside Ω. We first observe that for $\theta_h \in \mathbb{V}_h(\Omega)$ and $\tilde{\theta}_h$ its extension by zero outside Ω, we have

$$(w^M(\widetilde{\eta}_h, t_j), \theta_h)_\Omega = (\widetilde{\pi}_h w^M(\widetilde{\eta}_h, t_j), \widetilde{\theta}_h)_{B^M(t)}, \tag{V.52}$$

where $\widetilde{\pi}_h$ denotes the L^2 projection onto $\mathbb{V}_h^M(t)$. Using the above identity and recalling that $t_j = e^{-y_j/2}$, we obtain

$$\begin{aligned} a^{k,M}(\eta_h, \theta_h) - a_h^{k,M}(\eta_h, \theta_h) &= \underbrace{\frac{c_s}{2} k \sum_{t_j \le \frac{1}{2}} e^{sy_j} (\widetilde{\pi}_h w^M(\widetilde{\eta}_h, t_j) - w_h^M(\widetilde{\eta}_h, t_j), \widetilde{\theta}_h)_{B^M(t)}}_{=:E_1} \\ &\quad + \underbrace{\frac{c_s}{2} k \sum_{t_j > \frac{1}{2}} e^{sy_j} (\widetilde{\pi}_h w^M(\widetilde{\eta}_h, t_j) - w_h^M(\widetilde{\eta}_h, t_j), \widetilde{\theta}_h)_{B^M(t)}}_{=:E_2} . \end{aligned}$$

We bound the two terms separately and start with the latter.

1 In view of the definitions (V.32) of $w^M(t)$ and (V.44) of $w_h^M(t)$, we have

$$\widetilde{\pi}_h w^M(\widetilde{\eta}_h, t) - w_h^M(\widetilde{\eta}_h, t) = \widetilde{\pi}_h v^M(\widetilde{\eta}_h, t) - v_h^M(\widetilde{\eta}_h, t). \tag{V.53}$$

We recall that $T_t = T_{B^M(t)}$ and S_t are defined by (V.6) with the domain $B^M(t)$ and (V.48) respectively. Using these operators and the relations satisfied by $v^M(t)$ and $v_h^M(t)$ (see (V.33) and (V.45)), we arrive at

$$\begin{aligned} \widetilde{\pi}_h w^M(\widetilde{\eta}_h, t) - w_h^M(\widetilde{\eta}_h, t) &= [S_{t,h}(S_{t,h} + t^2 I)^{-1} - \widetilde{\pi}_h S_t(S_t + t^2 I)^{-1}]\widetilde{\eta}_h \\ &= t^2 (S_{h,t} + t^{-2} I)^{-1} \widetilde{\pi}_h (S_{t,h} - S_t)(S_t + t^2 I)^{-1} \widetilde{\eta}_h. \end{aligned} \tag{V.54}$$

Thus,

$$
\begin{aligned}
&\|\widetilde{\pi}_h w^M(\widetilde{\eta}_h,t) - w_h^M(\widetilde{\eta}_h,t)\|_{L^2(\Omega^M(t))} \\
&\qquad \le t^2 \|(S_{t,h}+t^2I)^{-1}\widetilde{\pi}_h(S_{t,h}-S_t)(S_t+t^2I)^{-1}\| \, \|\widetilde{\eta}_h\|_{L^2(\Omega^M(t))} \\
&\qquad \le t^2 \|(S_{t,h}+t^2I)^{-1}\widetilde{\pi}_h\| \, \|S_{t,h}-S_t\| \, \|(S_t+t^2I)^{-1}\| \, \|\widetilde{\eta}_h\|_{L^2(\Omega^M(t))}.
\end{aligned}
$$

Here we have used $\|\cdot\|$ to denote the operator norm of operators from $L^2(B^M(t))$ to $L^2(B^M(t))$. Combining

$$
\|(S_{t,h}+t^2I)^{-1}\widetilde{\pi}_h\| \le t^{-2}, \quad \|(S_t+t^2I)^{-1}\| \le t^{-2}
$$

and Lemma V.8 gives

$$
\|\widetilde{\pi}_h w^M(\widetilde{\eta}_h,t) - w_h^M(\widetilde{\eta}_h,t)\| \le Ct^{-2}h^2\|\widetilde{\eta}_h\|_{L^2(B^M(t))}.
$$

Whence,

$$
\begin{aligned}
|E_2| &\le Ch^2 k \sum_{t_j > \frac{1}{2}} e^{(s+1)jk} \|\widetilde{\eta}_h\|_{L^2(\Omega^M(t))} \|\widetilde{\theta}_h\|_{L^2(\Omega^M(t))} \\
&\le Ch^2 \|\eta_h\|_{L^2(\Omega)} \|\theta_h\|_{L^2(\Omega)} \left(k \sum_{jk<2\ln 2} e^{(s+1)jk} \right) \le Ch^2 \|\eta_h\|_{L^2(\Omega)} \|\theta_h\|_{L^2(\Omega)}.
\end{aligned}
$$

$\boxed{2}$ We now focus on E_1 which requires a finer analysis using intermediate spaces. Also, we argue differently for $\beta \in (1, 3/2)$ and for $\beta \in (s, 1]$. In either case, we define

$$
\epsilon := \min(1+\alpha-2s, 1/\ln(1/h))
$$

and note that

$$
\epsilon^{-1} \le c(1+\ln(1/h)) \quad \text{and} \quad h^{-\epsilon} \le c \tag{V.55}
$$

with c depending on s but not h.

When $\beta \in (1, 3/2)$, we invoke (V.53) again to deduce

$$|E_1| \le k \sum_{t_j \le \frac{1}{2}} e^{sy_j} \|\widetilde{\pi}_h v^M(\widetilde{\eta}_h, t_j) - v_h^M(\widetilde{\eta}_h, t_j)\|_{\dot{H}_h^{-\alpha}} \|\widetilde{\theta}_h\|_{\dot{H}_h^{\alpha}}. \tag{V.56}$$

We set $\mu(t) := t^{-2} - 1$ and compute

$$\begin{aligned}\widetilde{\pi}_h v^M(\widetilde{\eta}_h, t) - v_h^M(\widetilde{\eta}_h, t) &= t^{-2}[(I + \mu(t)T_{t,h})^{-1}T_{t,h} - \widetilde{\pi}_h T_t(I + \mu(t)T_t)^{-1}]\widetilde{\eta}_h \\ &= (t\mu(t))^{-2}(\mu(t)^{-1}I + T_{t,h})^{-1}\widetilde{\pi}_h(T_{t,h} - T_t)(\mu(t)^{-1}I + T_t)^{-1}\widetilde{\eta}_h,\end{aligned} \tag{V.57}$$

which is now estimated in three parts. In view of Lemma II.1 with $T = T_t$, we have

$$\|(\mu(t)^{-1}I + T_t)^{-1}\|_{\dot{H}^{\beta} \to \dot{H}^{\beta-2}} \le 1,$$

For the second part, the error estimate (V.47) with $1 + r = \beta$ reads

$$\|T_{t,h} - T_t\|_{\dot{H}^{\beta-2} \to L^2(B^M(t))} \le Ch^{\beta}.$$

We estimate the last term of the product in the right hand side of (V.57) by

$$\begin{aligned}&\|(\mu(t)^{-1} + T_{t,h})^{-1}\widetilde{\pi}_h\|_{L^2(B^M(t)) \to \dot{H}_h^{-\alpha}} \\ &\qquad \le C\|(\mu(t)^{-1} + T_{t,h})^{-1}\|_{\dot{H}_h^{-\alpha+2s+\epsilon} \to \dot{H}_h^{-\alpha}} \|\widetilde{\pi}_h\|_{L^2(B^M(t)) \to \dot{H}_h^{-\alpha+2s+\epsilon}}.\end{aligned}$$

Thus, Lemma II.3, the inverse estimate (II.15) and (V.55) yield

$$\|(\mu(t)^{-1} + T_{t,h})^{-1}\widetilde{\pi}_h\|_{L^2(B^M(t)) \to \dot{H}_h^{-\alpha}} \le Ch^{\alpha-2s-\epsilon}t^{(2s+\epsilon-2)} \le Ch^{\alpha-2s}t^{(2s+\epsilon-2)}.$$

Note that for $t \in (0, 1/2]$, $0 < t^2 \le \mu(t)^{-1} \le \frac{4}{3}t^2 \le \frac{1}{3}$ so that

$$(t\mu(t))^{-2} \le \frac{16t^2}{9}.$$

Combining the above estimates with (V.57) gives

$$\|\widetilde{\pi}_h v^M(\widetilde{\eta}_h, t) - v_h^M(\widetilde{\eta}_h, t)\|_{\dot{H}_h^{-\alpha}} \le Ct^{2s+\epsilon} h^{\beta+\alpha-2s} \|\widetilde{\eta}_h\|_{\dot{H}^\beta}, \tag{V.58}$$

Since $t_j = e^{-y_j/2}$,

$$e^{sy_j} t_j^{2s+\epsilon} = e^{-\epsilon y_j/2}.$$

Estimates (V.56) and (V.58) then yield

$$\begin{aligned} |E_1| &\le Ch^{\beta+\alpha-2s} k \sum_{ky_j \ge 2\ln 2} e^{-\epsilon y_j/2} \|\widetilde{\eta}_h\|_{\dot{H}^\beta} \|\widetilde{\theta}_h\|_{\dot{H}^\alpha} \\ &\le Ch^{\beta+\alpha-2s} \epsilon^{-1} \|\widetilde{\eta}_h\|_{\dot{H}^\beta} \|\widetilde{\theta}_h\|_{\dot{H}_h^\alpha}. \end{aligned} \tag{V.59}$$

$\boxed{3}$ Now we bound the right hand side above in two cases. If $\alpha \le 1$ (i.e. $s \le 1/2$), in view of Remark V.1 and (V.49), there exists a constant C satisfying

$$\|\widetilde{\theta}_h\|_{\dot{H}_h^\alpha} \le C \|\widetilde{\theta}_h\|_{\dot{H}^\alpha} \le C \|\theta_h\|_{\dot{H}^\alpha(\Omega)} \le C \|\theta_h\|_{\widetilde{H}^\alpha(\Omega)} \tag{V.60}$$

We apply Lemma V.7 for η_h together with the above inequality to arrive at

$$\begin{aligned} |E_1| &\le Ch^{\beta+\alpha-2s-\epsilon} \|\widetilde{\eta}_h\|_{H^\beta(\mathbb{R}^d)} \|\theta_h\|_{\widetilde{H}^\alpha(\Omega)} \left(k \sum_{t(y_j) \le \frac{1}{2}} e^{-\epsilon jk/2} \right) \\ &\le C\epsilon^{-1} h^{\beta+\alpha-2s-\epsilon} \|\widetilde{\eta}_h\|_{H^\beta(\mathbb{R}^d)} \|\theta_h\|_{\widetilde{H}^\alpha(\Omega)} = Ch^{\beta+\alpha-2s}(1+\ln(h^{-1})) \|\eta_h\|_{\widetilde{H}^\beta(\Omega)} \|\theta_h\|_{\widetilde{H}^\alpha(\Omega)}. \end{aligned}$$

If $\alpha > 1$, we estimate $\|\widetilde{\theta}_h\|_{\dot{H}^\alpha}$ using Lemma V.7 to get the same bound as above.

$\boxed{4}$ When $\beta \in (s, 1]$, we bound (V.57) with different norms. In fact, there hold

$$\|(\mu(t)^{-1} I + T_t)^{-1}\|_{\dot{H}^\beta \to \dot{H}^{-1}} \le t^{\beta-1}, \quad \|T_{t,h} - T_t\|_{\dot{H}^{-1} \to L^2(B^M(t))} \le Ch,$$

and

$$\|(\mu(t)^{-1}+T_{t,h})^{-1}\widetilde{\pi}_h\|_{L^2(\Omega^M(t))\to\dot{H}_h^{-\alpha}}\le Ch^{-1+\beta+\alpha-2s-\epsilon}t^{(2s+\epsilon-\beta-1)}.$$

Combining these estimates leads to the same result as in (V.59). The rest of the proof are the same as in the previous step except that we use the fact $\|\widetilde{\eta}_h\|_{\dot{H}^\beta}\le C\|\eta_h\|_{\widetilde{H}^\beta(\Omega)}$ as in (V.60).

$\boxed{5}$ The proof of the theorem is complete upon combining the estimates for E_1 and E_2. □

V.6 The Discrete Problem and Error Estimates

The finite element approximation of the problem (V.1) is to find $u_h\in\mathbb{V}_h(\Omega)$ so that

$$a_h^{k,M}(u_h,\theta_h)=(f,\theta_h)_\Omega \qquad \text{for all } \theta_h\in\mathbb{V}_h(\Omega). \tag{V.61}$$

Analogous to Lemma V.2, we have the following representation using K-functional. The proof of the lemma is similar to that of Lemma V.2 and is omitted.

Lemma V.10 (K-functional formulation on the discrete space)**.** *For $\eta_h\in\mathbb{V}_h(\Omega)$, there holds*

$$(w_h^M(\widetilde{\eta}_h,t),\eta_h)_D=(w_h^M(\widetilde{\eta}_h,t),\widetilde{\eta}_h)_{B^M(t)}=:K_h(\widetilde{\eta}_h,t),$$

where

$$K_h(\widetilde{\eta}_h,t):=\min_{\varphi_h\in\mathbb{V}_h^M(t)}\Big(\|\widetilde{\eta}_h-\varphi_h\|^2_{L^2(B^M(t))}+t^2d_{B^M(t)}(\varphi_h,\varphi_h)\Big).$$

We emphasize that for $v_h\in\mathbb{V}_h^M(t)$, its extension by zero $\widetilde{\eta}_h$ belongs to $H^1(\mathbb{R}^d)$ and therefore

$$K_h(\tilde{v}_h,t)\ge K(\tilde{v}_h,t). \tag{V.62}$$

This property is critical in the proof of next theorem, which ensures the $\mathbb{V}_h(\Omega)$-ellipticity of the discrete bilinear for $a_h^{k,M}$. Before describing this next result, we recall that according to (V.28)

$$|a(\eta_h,\theta_h)-a^k(\eta_h,\theta_h)|\le\gamma(k)\|\eta_h\|_{\widetilde{H}^\delta(\Omega)}\|\theta_h\|_{\widetilde{H}^s(\Omega)}$$

with δ between s and $\min(2-s, 3/2)$ (since $\mathbb{V}_h(\Omega) \subset \widetilde{H}^{3/2-\epsilon}(\Omega)$ for any $\epsilon > 0$) and $\gamma(k) \sim Ce^{-\pi^2/(2k)}$. Also, we note that based on (II.15) and the norm equivalency (cf. Remark V.1 and (II.18)), there holds the inverse estimate

$$\|v_h\|_{\widetilde{H}^{r^+}(\Omega)} \leq c_I h^{r^- - r^+} \|v_h\|_{\widetilde{H}^{r^-}(\Omega)}, \qquad \forall v_h \in \mathbb{V}_h^M(t). \tag{V.63}$$

Theorem V.11 ($\mathbb{V}_h(\Omega)$-ellipticity)**.** *Let δ in Theorem V.3 between s and $\min(2-s, 3/2)$, k be the quadrature spacing and c_I be the inverse constant in* (V.63). *We assume that the quadrature parameters N^- and N^+ are chosen according to* (V.27). *Let $\gamma(k)$ be given by* (V.29) *and assume that k is chosen sufficiently small so that*

$$c_I \gamma(k) h^{s-\delta} < 1.$$

Then, there is a constant c independent of h, k and M such that

$$a_h^{k,M}(\eta_h, \eta_h) \geq c\|\eta_h\|^2_{\widetilde{H}^s(\Omega)}, \qquad \textit{for all } \eta_h \in \mathbb{V}_h(\Omega).$$

Proof. Let $\eta_h \in \mathbb{V}_h(\Omega)$ so that $\widetilde{\eta}_h \in H^1(\mathbb{R}^d)$. We use the equivalence relations provided by Lemmas V.2 and V.10 together with the monotonicity property (V.62) to write

$$a_h^{k,M}(\eta_h, \eta_h) = \frac{c_s k}{2} \sum_{j=-N^-}^{N^+} e^{sy_j} K_h(\widetilde{\eta}_h, t_j) \geq \frac{c_s k}{2} \sum_{j=-N^-}^{N^+} e^{sy_j} K(\widetilde{\eta}_h, t_j) = a^k(\eta_h, \eta_h).$$

The quadrature consistency bound (V.28) supplemented by an inverse inequality (V.63) yields

$$a_h^{k,M}(\eta_h, \eta_h) \geq a(\eta_h, \eta_h) - \gamma(k)\|\eta_h\|_{\widetilde{H}^\delta(\Omega)}\|\eta_h\|_{\widetilde{H}^s(\Omega)} \geq a(\eta_h, \eta_h) - c_I\gamma(k)h^{s-\delta}\|\eta_h\|^2_{\widetilde{H}^s(\Omega)}.$$

The desired result follows from assumption $c_I\gamma(k)h^{s-\delta} < 1$ and the coercivity of $a(\cdot,\cdot)$; see (V.2). □

Now that the consistency error between $a(\cdot,\cdot)$ and $a_h^{k,M}(\cdot,\cdot)$ is obtained, we can apply the first Strang's lemma to deduce the convergence of the approximation u_h towards u in the energy norm. To achieve this, we need a result regarding the stability and approximability of the Scott-Zhang interpolant π_h^{sz} [61] in the fractional spaces $\widetilde{H}^\beta(\Omega)$ with $\beta \in (1, 3/2)$.

This is the subject of the next lemma. We refer to [10, Appendix A] for a detailed proof. We also refer to [24] for the proof when $\beta \in (1/2, 1]$.

Lemma V.12 (Scott-Zhang Interpolant)**.** *Let $\beta \in (1, 3/2)$. Then, there is a constant C independent of h such that*

$$\|\pi_h^{sz} v\|_{\widetilde{H}^\beta(\Omega)} \le C\|v\|_{\widetilde{H}^\beta(\Omega)} \tag{V.64}$$

and for $s \in [0,1]$,

$$\|\pi_h^{sz} v - v\|_{\widetilde{H}^s(\Omega)} \le Ch^{\beta-s}\|v\|_{\widetilde{H}^\beta(\Omega)}, \tag{V.65}$$

for all $v \in \widetilde{H}^\beta(\Omega)$.

We note that the above lemma holds for $\beta \in (0,1)$ and $s \in (0,\beta)$ provided that π_h^{sz} is replaced by π_h, the L^2 projection onto $\mathbb{V}_h(\Omega)$; see (II.19) and (II.16). In order to consider both case simultaneously in the following proof, we set $\Pi_h = \pi_h$ when $\beta \in [0,1]$ and $\Pi_h = \pi_h^{sz}$ when $\beta \in (1, 3/2)$.

Theorem V.13. *Assume that the solution u of* (V.1) *belongs to $\widetilde{H}^\beta(\Omega)$ for $\beta \in (s, 3/2)$. Let $\delta := \min(2-s, \beta)$ be as in Theorem V.3, k be the quadrature spacing and c_I be the inverse constant in* (V.63). *We assume that the quadrature parameters N^- and N^+ are chosen according to* (V.27). *Let $\gamma(k)$ be given by* (V.29) *and assume that k is chosen sufficiently small so that*

$$c_I\gamma(k)h^{s-\delta} < 1.$$

Moreover, let $u_h \in \mathbb{V}_h(\Omega)$ be the solution of (V.61). *Then there is a constant C independent of h,*

M and k satisfying

$$\|u-u_h\|_{\widetilde{H}^s(\Omega)} \le C(\gamma(k)+e^{-cM}+(1+\ln(h^{-1}))h^{\beta-s})\|u\|_{\widetilde{H}^\beta(\Omega)}. \tag{V.66}$$

Proof. In our context, the first Strang's lemma (see e.g. Theorem 4.1.1 in [25]) reads

$$\|u-u_h\|_{\widetilde{H}^s(\Omega)} \le C \inf_{v_h\in\mathbb{V}_h(\Omega)} \left(\|u-v_h\|_{\widetilde{H}^s(\Omega)} + \sup_{w_h\in\mathbb{V}_h(\Omega)} \frac{|(a-a_h^{k,M})(v_h,w_h)|}{\|w_h\|_{\widetilde{H}^s(\Omega)}} \right),$$

where C is a constant independent of h, k and M. From the consistency estimates (V.28), (V.37) and (V.51) with $\alpha=s$, we deduce that

$$\begin{aligned}\|u-u_h\|_{\widetilde{H}^s(\Omega)} &\le C\|u-\Pi_h u\|_{\widetilde{H}^s(\Omega)} \\ &\quad + C(\gamma(k)+e^{-cM}+(1+\ln(h^{-1}))h^{\beta-s})\|\Pi_h u\|_{\widetilde{H}^\beta(\Omega)}\end{aligned}$$

The desired estimate follows from the approximability and stability of Π_h. $\square$

Next, we show a $L^2(\Omega)$ error estimates using a duality argument. We note that Theorem 7.1 together with Theorem 5.4 in [38] (see also Proposition 2.7 in [14]) guarantees that when $\partial\Omega$ is of C^∞ class and f is in $L^2(\Omega)$, the solution of the problem (V.1) is in $\widetilde{H}^\alpha(\Omega)$ for $\alpha=\min\{2s,1/2+s\}-\epsilon$ for every $\epsilon>0$. Now we apply this regularity result to a dual problem. That is, if $\mathsf{z}\in\widetilde{H}^s(\Omega)$ solves

$$a(\theta,\mathsf{z}) = (u-u_h,\theta)_\Omega, \qquad \text{for all } \theta\in\widetilde{H}^s(\Omega), \tag{V.67}$$

we have

$$\|\mathsf{z}\|_{\widetilde{H}^\alpha(\Omega)} \le C\|u-u_h\|_{L^2(\Omega)}. \tag{V.68}$$

Theorem V.14. *Under the assumptions in Theorem V.13, assume $\partial\Omega$ is of C^∞ class and let α =min$\{2s,s+1/2\}-\epsilon$. We have*

$$\|u-u_h\|_{L^2(\Omega)} \le C\ln(h^{-1})(\gamma(k)+e^{-cM}+\ln(h^{-1})h^{\beta+\alpha-2s})\|u\|_{\widetilde{H}^\beta(\Omega)}. \tag{V.69}$$

Proof. We let $e_h = u - u_h$ and $\theta = e_h$ in (V.67) to write

$$\|e_h\|^2_{L^2(\Omega)} = a(z, e_h).$$

By adding and subtracting $a(\Pi_h \mathsf{z}, e_h)$ and $a(\Pi_h \mathsf{z}, \Pi_h u)$ together with the relation

$$a(\Pi_h \mathsf{z}, u) = (f, \Pi_h \mathsf{z}) = a_h^{k,M}(\Pi_h \mathsf{z}, u_h),$$

we obtain that

$$\begin{aligned}
\|e_h\|^2_{L^2(\Omega)} &= a((I - \Pi_h)\mathsf{z}, e_h) + a(\Pi_h z, e_h) \\
&= a((I - \Pi_h)\mathsf{z}, e_h) + a(\Pi_h z, \Pi_h e_h) + a(\Pi_h z, u - \Pi_h u) \\
&= a((I - \Pi_h)\mathsf{z}, e_h) + a(\Pi_h z, \Pi_h e_h) + a_h^{k,M}(\Pi_h z, u_h) - a(\Pi_h z, \Pi_h u) \qquad \text{(V.70)} \\
&= a((I - \Pi_h)\mathsf{z}, e_h) + [a(\Pi_h \mathsf{z}, \Pi_h e_h) - a_h^{k,M}(\Pi_h \mathsf{z}, \Pi_h e_h)] \\
&+ [a_h^{k,M}(\Pi_h \mathsf{z}, \Pi_h u) - a(\Pi_h \mathsf{z}, \Pi_h u)] =: I + II + III.
\end{aligned}$$

We first deduce with the help of Lemma V.12 and (V.68) that

$$\begin{aligned}
|I| &\le \|\mathsf{z} - \Pi_h \mathsf{z}\|_{\widetilde{H}^s(\Omega)} \|e_h\|_{\widetilde{H}^s(\Omega)} \\
&\le C h^{\alpha - s} \|\mathsf{z}\|_{\widetilde{H}^\alpha(\Omega)} \|e_h\|_{\widetilde{H}^s(\Omega)} \le C h^{\alpha - s} \|e_h\|_{L^2(\Omega)} \|e_h\|_{\widetilde{H}^s(\Omega)}.
\end{aligned}$$

Next, we invoke the consistency error estimates in Theorem V.3, V.5 and V.9 with $\theta_h = \Pi_h e_h$ and $\eta_h = \Pi_h \mathsf{z}$ to obtain

$$\begin{aligned}
|II| &\le C(\gamma(k) + e^{-cM} + \ln(h^{-1}) h^{\alpha - s}) \|\mathsf{z}\|_{\widetilde{H}^\alpha(\Omega)} \|e_h\|_{\widetilde{H}^s(\Omega)} \\
&\le C(\gamma(k) + e^{-cM} + \ln(h^{-1}) h^{\alpha - s}) \|e_h\|_{L^2(\Omega)} \|e_h\|_{\widetilde{H}^s(\Omega)}.
\end{aligned}$$

Finally, we apply the consistency error estimates again with $\theta_h = \Pi_h \mathsf{z}$ and $\eta_h = \Pi_h u$ to get

$$\begin{aligned} |III| &\leq C(\gamma(k) + e^{-cM} + \ln(h^{-1}) h^{\beta+\alpha-2s}) \|\mathsf{z}\|_{\widetilde{H}^{\alpha}(\Omega)} \|u\|_{\widetilde{H}^{\beta}(\Omega)} \\ &\leq C(\gamma(k) + e^{-cM} + \ln(h^{-1}) h^{\beta+\alpha-2s}) \|e_h\|_{L^2(\Omega)} \|u\|_{\widetilde{H}^{\beta}(\Omega)}. \end{aligned}$$

Combining above three estimates into (V.70) and the energy norm estimate (V.66) for $\|e_h\|_{\widetilde{H}^s(\Omega)}$ gives the L^2 error bound (V.69). □

V.7 Numerical Implementation and Results

In this section, we present detailed numerical implementation to solve the following model problems.

V.7.1 Model Problems

One of the difficulties in developing numerical approximation to (V.4) is that there are relatively few examples where analytical solutions are available. One exception is the case when Ω is the unit ball in $\mathbb{R}^d$. In that case, the solution to the variational problem

$$a(u, \phi) = (1, \phi)_D, \qquad \text{for all } \phi \in \widetilde{H}^s(\Omega) \tag{V.71}$$

is radial and given by, (see [31])

$$u(x) = \frac{2^{-2s}\Gamma(d/2)}{\Gamma(d/2+s)\Gamma(1+s)}(1-|x|^2)^s. \tag{V.72}$$

It is also possible to compute the right hand side corresponding to the solution $u(x) = 1 - |x|^2$ in the unit ball. The corresponding right hand side can be derived by first computing the Fourier transform of $\widetilde{u}$, i.e.,

$$\mathcal{F}(\widetilde{u}) = 2J_2(|\zeta|)/|\zeta|^2,$$

where J_n is the Bessel function of the first kind. When $0 < s < 1$, we obtain

$$f(x) = \mathcal{F}^{-1}(2|\zeta|^{2s-2} J_2(|\zeta|)) = \frac{2^{2s}\Gamma(d/2+s)}{\Gamma(d/2)\Gamma(2-s)} {}_2F_1\left(d/2+s, s-1, d/2, |x|^2\right), \tag{V.73}$$

where ${}_2F_1$ is the Gaussian or ordinary hypergeometric function.

Remark V.6 (Smoothness)**.** *Even though the solution $u(x) = 1 - |x|^2$ is infinitely differentiable on the unit ball, the right hand side f has limited smoothness. Note that f is the restriction of $(-\Delta)^s \widetilde{u}$ to the unit ball. Now $\widetilde{u} \in H^{3/2-\epsilon}(\mathbb{R}^d)$ for $\epsilon > 0$ but is not in $H^{3/2}(\mathbb{R}^d)$. This means that $(-\Delta)^s \widetilde{u}$ is only in $H^{3/2-2s-\epsilon}(\mathbb{R}^d)$ and hence f is only in $H^{3/2-2s-\epsilon}(\Omega)$. This is in agreement with the singular behavior of ${}_2F_1(d/2+s, s-1, d/2, t)$ at $t = 1$ (see [58, Section 15.4]). In fact,*

$$\begin{aligned}
{}_2F_1(d/2+s, s-1, d/2, 1) &= \frac{\Gamma(d/2+s)\Gamma(1-2s)}{\Gamma(d/2+1-s)\Gamma(-s)} \quad \text{when } 0 < s < 1/2, \\
\lim_{t\to 1^-} \frac{{}_2F_1(d/2+s, s-1, d/2, t)}{-\log(1-t)} &= \frac{\Gamma(d/2)}{\Gamma(-1/2)\Gamma(1/2)} \quad \text{when } s = 1/2, \\
\lim_{t\to 1^-} \frac{{}_2F_1(d/2+s, s-1, d/2, t)}{(1-t)^{-2s+1}} &= \frac{\Gamma(d/2)}{\Gamma(-1/2)\Gamma(1/2)} \quad \text{when } 1/2 < s < 1.
\end{aligned}$$

This implies that for $s \geq 1/2$, the trace on $|x| = 1$ of $f(x)$ given by (V.73) *fails to exist (as for generic functions in $H^{3/2-2s}(\mathbb{R}^d)$). This singular behavior affects the convergence rate of the finite element method when the finite element data vector is approximated using standard numerical quadrature (e.g. Gaussian quadrature).*

V.7.2 Numerical Implementation

Based on the notations in Section V.4, we set $\Omega = B$ to be either the unit disk in $\mathbb{R}^2$ or $\Omega = (-1, 1)$ in $\mathbb{R}$. Let $B^M(t)$ be corresponding dilated domains. In one dimensional case, we consider $\mathcal{T}_h(\Omega)$ to be a uniform mesh and $\mathbb{V}_h(\Omega)$ to be the continuous piecewise linear finite element space. For the two dimensional case, $\mathcal{T}_h(\Omega)$ a regular (in the sense of page 247 in [25]) subdivision made of quadrilaterals. In this case, $\mathbb{V}_h(\Omega)$ is the set of continuous piecewise bilinear functions.

Non-uniform Meshes for $B^M(t)$.

We extend $\mathcal{T}_h(\Omega)$ to non-uniform meshes $\mathcal{T}_h^M(t)$, thereby violating the quasi-uniform assumption. For $t \leq 1$, we use a quasi-uniform mesh on $B^M(t) = B^M(1)$ with the same mesh size h. When $t > 1$ and $\Omega = (-1, 1)$, we use an exponentially graded mesh outside of Ω, i.e. the mesh points are $\pm e^{ih_0}$ for $i = 1, \ldots, \lceil M/h \rceil$ with $h_0 = h(\ln \gamma)/M$, where γ is the radius of $B^M(t)$ (see (V.30)). Therefore, we maintain the same number of mesh points for all $B^M(t)$. When Ω is a unit disk in $\mathbb{R}^2$, we start with a coarse subdivision of $B^M(t)$ as in the left of Figure V.2 (the coarse mesh of Ω in grey). Note that all vertices of a square have the same radial coordinates. We also point out that the position of the vertices along the radial direction and outside of Ω follow the same exponential distribution as in the one dimensional case. Then we refine each cell in Ω by connecting the midpoints between opposite edges. For the cells outside of Ω, we consider the same refinement in the polar coordinate system $(\ln r, \theta)$ with $r > 1$ and $\theta \in [0, 2\pi]$. This guarantees that mesh points on the same radial direction still follows the exponential distribution after global refinements and the number of mesh points in $\mathcal{T}_h^M(t)$ is unchanged for all $t > 0$. The figure on the right of Figure V.2 shows the exponentially graded mesh after three times global refinement.

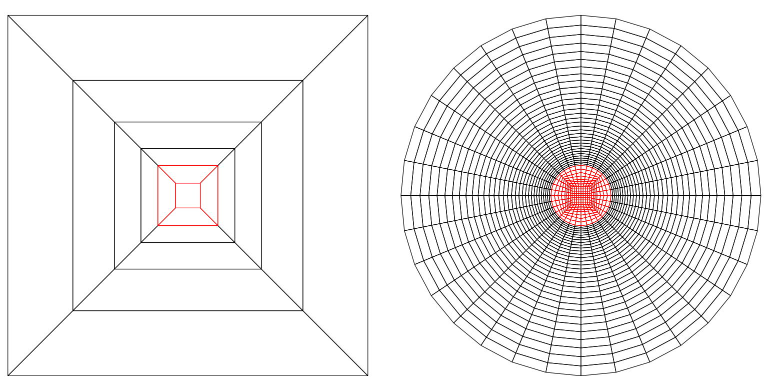

Figure V.2: Coarse gird (left) and three-times-refined non-uniform grid (right) of $B^M(t)$ with $M = 4$ and $t = 1$. Grids of D are in grey.

Matrix Aspects

To express the linear system to be solved, we denote by U to be the coefficient vector of u_h and F to be the coefficient vector of the L^2 projection of f onto $\mathbb{V}_h(\Omega)$. Let $M_h(t)$ and $A_h(t)$ be the mass and stiffness matrix in $\mathbb{V}_h^M(t)$. Denote $M_{\Omega,h}$ to be the mass matrix in $\mathbb{V}_h(\Omega)$. The linear system is given by

$$\frac{\sin(\pi\beta)k}{\pi}\sum_{i=-N^-}^{N^+} e^{sy_i} M_{\Omega,h}(e^{y_i}M_h(t_i)+A_h(t_i))^{-1}A_h(t_i)U = F \tag{V.74}$$

with $y_i = ik$ and $t_i = e^{-y_i/2}$. Here $M_{\Omega,h}$, U and F are all extended by zeros so that the dimension of the system is equal to the dimension of $\mathbb{V}_h^M(t)$.

Preconditioner

Since the linear system is symmetric, we apply the Conjugate Gradient method to solve the above linear system. Due to the norm equivalence between $\mathbb{H}^s(\Omega)$ and $\widetilde{H}^s(\Omega)$, the condition number of the system matrix is bounded by Ch^{-2s}. In order to reduce the number of iterations in one dimensional space, we use fractional powers of the discrete Laplacian $L_{\Omega,h}$ as a preconditioner, where $L_{\Omega,h} : H_0^1(\Omega) \to L^2(\Omega)$ is defined by

$$d_D(L_{\Omega,h}w, \phi_h) = d_\Omega(w, \phi_h), \qquad \text{for all } \phi_h \in \mathbb{V}_h(\Omega).$$

This can be computed by the discrete sine transform similar to the implementation discussed in Section III.4.

In two dimensional space, we use the multilevel preconditioner advocated in [16].

V.7.3 Numerical Illustration for the Non-smooth Solution

We first consider the numerical experiments for the model problem (V.71) and study the behavior of the $L^2(\Omega)$ error.

Influence from the Sinc Quadrature and Domain Truncation.

When $D = (-1, 1)$, we approximate the solution on the fixed uniform mesh with the mesh size $h = 1/8192$. The domain truncation parameter M is also fixed to be 20. Thus, h is small enough and M is large enough so that the $L^2(\Omega)$-error is dominant by the sinc quadrature spacing k. The left part of Figure V.3 shows that the $L^2(\Omega)$-error quickly converges to the error dominant by the Galerkin approximation when k approaches zero. Similar results are observed from the right part of Figure V.3 when the domain truncation parameter M increases. In this case, the mesh size $h = 1/8192$ and the quadrature step size $k = 0.2$.

Error Convergence from the Finite Element Approximation

We note that we implement the numerical algorithm for the two dimensional case using the deal.ii Library [7] and we invert matrices in (V.74) using the direct solver from UMFPACK [28]. Figure V.4 shows the approximated solutions for $s = 0.3$ and $s = 0.7$, respectively. Table V.1 reports errors $\|u - u_h\|_{L^2(\Omega)}$ and rates of convergence with $s = 0.3, 0.5$ and 0.7. Here the quadrature spacing ($k = 0.25$) and the domain truncation parameter ($M = 4$) are fixed so that the finite element discretization dominates the error. Since the solution u is in $H^{s+1/2-\epsilon}(\Omega)$ (see [2] for a proof), Table V.1 matches the expected rate of convergence $\min(1, s + 1/2)$ according to Theorem V.14.

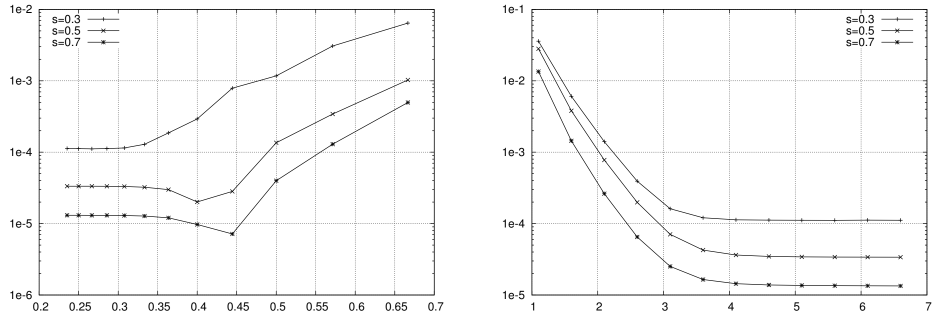

Figure V.3: The above figures report the $L^2(\Omega)$-error behavior when $D = (-1, 1)$. The left one shows the error as a function of the quadrature spacing k for a fixed mesh size ($h = 1/8192$) and domain truncation parameter ($M = 20$). The right plot reports the error as a function of the domain truncation parameter M with fixed mesh size ($h = 1/8192$) and quadrature spacing ($k = 0.2$). The spatial error dominates when k is small (left) and M is large (right).

#DOFS	$s = 0.3$		$s = 0.5$		$s = 0.7$	
345	2.69×10^{-1}	-	1.63×10^{-1}	-	1.03×10^{-1}	-
1361	1.59×10^{-1}	0.7575	9.07×10^{-2}	0.8426	5.55×10^{-2}	0.8918
5409	9.56×10^{-2}	0.7323	5.05×10^{-2}	0.8438	2.95×10^{-2}	0.9091
21569	5.71×10^{-2}	0.7447	2.78×10^{-2}	0.8633	1.54×10^{-2}	0.9366
86145	3.38×10^{-2}	0.7547	1.51×10^{-2}	0.8832	7.91×10^{-3}	0.9641
344321	1.99×10^{-2}	0.7644	8.07×10^{-3}	0.9004	3.97×10^{-3}	0.9936

Table V.1: $L^2(\Omega)$-errors for different values of s versus the number of degree of freedom used for the 2-D nonsmooth computations. #DOFS denotes the dimension of the finite element space $\mathbb{V}_h^M(t)$.

V.7.4 Numerical Illustration for the Smooth Solution

When the solution is smooth, the finite element error (assuming the exact computation of the stiffness entries, i.e. no consistency error) satisfies

$$\|u - u_h\|_{L^2(\Omega)} \leq ch^{2-2s+\min(s,1/2)}.$$

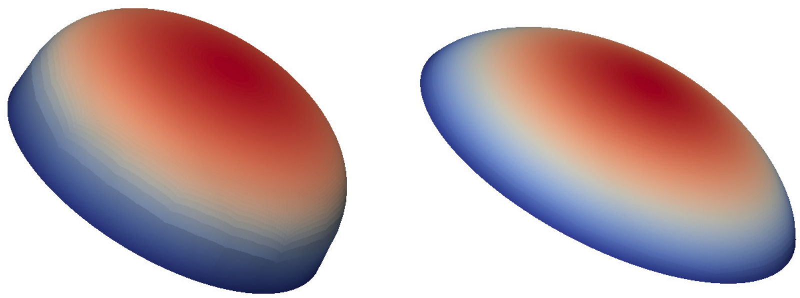

Figure V.4: Approximated solutions of (V.72) for $s = 0.3$ (left) and $s = 0.7$ (right) on the unit disk.

In contrast, because of the inherent consistency error, our method only guarantees (c.f., Theorem V.13)

$$\|u - u_h\|_{L^2(\Omega)} \leq ch^{3/2-2s+\min(s,1/2)}. \tag{V.75}$$

Table V.2 reports $L^2(\Omega)$-errors and rates for the problem (V.4) with the smooth solution $u(x) = 1 - |x|^2$ and the corresponding right hand side data (V.73) in the unit disk. To see the error decay, here we choose the quadrature step size $k = 0.2$ and the domain truncation parameter $M = 5$. The observed decay in the error does not match the expected rate (V.75). We think this loss of accuracy may be due either to the deterioration of the shape regularity constant in generating the subdivisions of $B^M(t)$ (see Section V.7.2) or to the imprecise numerical integration of the singular right hand side in (V.73).

To illustrate this, we consider the one dimensional problem. Instead of using (V.73) to compute the right hand side vector, we compute

$$(f, \phi_j) = a(u, \phi_j) = \frac{(\partial_L^{2s-1}\phi_j, u')_\Omega + (\partial_L^{2s-1}u, \phi_j')_\Omega}{2\,\cos(s\pi)} \tag{V.76}$$

with $\Omega = (-1, 1)$. We note that when $s < 1/2$, the fractional derivative with the negative power $2s - 1$ still makes sense for the local basis function ϕ_j. The right hand side of (V.76) can now be computed exactly.

#DOFS	$s = 0.3$		$s = 0.5$		$s = 0.7$	
409	6.24×10^{-2}	-	9.55×10^{-2}	-	1.35×10^{-1}	-
1617	2.90×10^{-2}	1.10	4.33×10^{-2}	1.14	6.27×10^{-2}	1.10
6433	1.44×10^{-2}	1.01	1.94×10^{-2}	1.15	2.81×10^{-2}	1.16
25665	7.21×10^{-3}	1.00	8.55×10^{-3}	1.19	1.20×10^{-2}	1.23
102529	3.56×10^{-3}	1.02	3.67×10^{-3}	1.22	4.78×10^{-3}	1.32
409857	1.74×10^{-3}	1.04	1.54×10^{-3}	1.25	1.73×10^{-3}	1.47

Table V.2: $L^2(\Omega)$-errors and rates for $s = 0.3$, 0.5 and 0.7 for the problem (V.4) with the right hand side (V.73). #DOFS denotes the number of degree of freedoms of $\Omega^M(t)$.

We illustrate the convergence rate for the one dimensional case in Table V.3 when the $L^2(\Omega)$-projection of right hand side is computed from (V.76). In this case, we compute at $s = 0.3, 0.4, 0.7$ as the expression in (V.76) is not valid for $s = 0.5$. We also fix $k = 0.2$ and $M = 6$. In all cases, we observe the predicted rate of convergence $\min(3/2, 2 - s)$, see (V.75).

h	$s = 0.3$		$s = 0.4$		$s = 0.7$	
1/16	4.51×10^{-4}		3.47×10^{-4}		9.27×10^{-4}	
1/32	1.42×10^{-4}	1.58	1.02×10^{-4}	1.77	4.16×10^{-4}	1.16
1/64	4.25×10^{-5}	1.63	3.31×10^{-5}	1.62	1.80×10^{-4}	1.21
1/128	1.34×10^{-5}	1.66	1.14×10^{-5}	1.54	7.66×10^{-5}	1.23
1/256	4.43×10^{-6}	1.59	4.06×10^{-6}	1.49	3.21×10^{-5}	1.25
1/512	1.50×10^{-6}	1.56	1.46×10^{-6}	1.48	1.33×10^{-5}	1.27

Table V.3: $L^2(\Omega)$-errors and rates for $s = 0.3$, 0.4 and 0.7 for the one dimensional problem when right hand side of the discrete problem is computed by (V.76).

CHAPTER VI

CONCLUSIONS, EXTENSIONS AND FUTURE RESEARCH

This dissertation has provided numerical schemes for parabolic problems involving fractional powers of elliptic operators and a stationary problem involving the integral fractional Laplacian. The approximations of both problems are based on the Dunford-Taylor integral representation of the corresponding solution operators. We note that since the integrand subproblems are diffusion-reaction problems, finite element software libraries can be applied.

Our studies in this dissertation have also contributed to the theoretical understanding our numerical methods. As a natural extension of the stationary problem (I.4), Chapter III shows that the $L^2(\Omega)$ error between the solution to the homogenous problem and its final approximation consists of two parts: the error from the space approximation and the exponentially convergent sinc approximation. The convergence rate in space only depends on the elliptic regularity index (Assumption II.1). In Chapter IV, the approximation schemes for the non-homogeneous problem have an extra time discretization error. The convergence rate in space for the non-homogeneous problem not only depends on the regularity index but also on the smoothness of the right hand side data. Both time discretization approaches for the non-homogenous problem, i.e. the pseduo-midpoint quadrature scheme in time and the Crank-Nicolson time stepping method, guarantee the second order convergence. In Chapter V, the error of the approximation to the stationary problem involving the integral fractional Laplacian consists of three terms: the error of the sinc approximation, domain truncation and the finite element approximation. The first two errors decay exponentially and the convergence rate for the finite element approximation depends on the regularity of the solution.

We next consider two extensions.

- *Approximation of Space-time Fractional Parabolic Equations.* Numerical methods discussed in Chapter III and Section IV.2 can be applied to the following space-time fractional parabolic problem: given $v \in L^2(\Omega)$ and $f \in L^\infty(0,\mathsf{T};L^2(\Omega))$, find u satisfying

$$\begin{cases} \partial_t^\gamma u + L^s u = 0, & \text{in } \Omega \times (0,\mathsf{T}], \\ u = v, & \text{on } \Omega \times \{0\}, \end{cases}$$

where ∂_t^γ denotes the left-sided Caputo fractional derivative with the order $\gamma \in (0,1)$. The above problem has a unique solution (see [55, Theorem 6]), which can be explicitly written as

$$u(t) := u(t,\cdot) = E(t)v + \int_0^t W(r)f(t-r)\,dr.$$

Here, for $w \in L^2(\Omega)$,

$$E(t)w := e_{\gamma,1}(-t^\gamma L^s)w = \sum_{j=1}^\infty e_{\gamma,1}(-t^\gamma \lambda_j^s)(w,\psi_j)_\Omega \psi_j$$

and

$$W(t)w := t^{\gamma-1}e_{\gamma,\gamma}(-t^\gamma L^s)w = \sum_{j=1}^\infty t^{\gamma-1}e_{\gamma,\gamma}(-t^\gamma \lambda_j^s)(w,\psi_j)_\Omega \psi_j,$$

with $e_{\gamma,\mu}(z)$ denoting the Mittag-Leffler function. Noting that the kernel $W(t)$ is singular at $t = 0$, we can apply the numerical scheme from Section IV.2 and overcome the singularity by utilizing a geometric refinement towards $t = 0$ so that we still obtain the second order convergence in time (up to a logarithmic factor). We refer to [9] for the details.

- *Error Estimates in $\mathbb{H}^s(\Omega)$ Norm.* In this dissertation, we only consider $L^2(\Omega)$ error estimate for the numerical approximation of the parabolic problem. We note that we can also show error estimates in the energy norm, i.e. $\mathbb{H}^s(\Omega)$ norm. The $\mathbb{H}^s(\Omega)$ norm error estimate for the space-time fractional parabolic equation is provided in [9] and the idea of the proof can be applied to the parabolic problem (I.9). Refined error estimates for sinc approximation are given in [11].

We end this chapter with two direction for future research.

- *Fractional Diffusion on Surfaces.* The Matérn field on a manifold can be constructed through a fractional stochastic partial differential equation, i.e. solving the problem (I.4) on a smooth surface with L replaced by the Laplace-Beltrami operator and f replaced by the Gaussian white noise; see [44]. We can also invoke the formula (I.5) to approximate this problem by replacing L with an finite element approximation on the surface. We note that the common finite element methods on surface like parametric finite element methods [33], trace finite element methods [57] and the narrow band methods [30, 29] are non-comforming. The error analysis using these numerical methods with a non-smooth right hand side data is an open question.

- *Obstacle Problems involving the Integral Fractional Laplacian.* The American option pricing problem is modeled by a parabolic obstacle problem involving a pseudodifferential operator [15]. In particular, the pseudodifferential operator is defined by a convolution with a kernel function $K(y) = e^{-\lambda|y|}/|y|^{1+2s}$. This diffusion process is called the tempered stable process and when $\lambda = 0$, we are back to the s-stable Lévy process (i.e. integral fractional Laplaican). A priori error analysis and a posteriori error analysis for the obstacle problems using the singular integral representation are discussed in [50, 56] as well as $1d$ numerical simulations. Here we consider the numerical approximation in two or higher dimensional space using the Dunford-Tarylor integral approach to approximate the American option pricing problems on multiple assets (cf. [19]).

REFERENCES

[1] Gabriel Acosta, Francisco M. Bersetche, and Juan Pablo Borthagaray. A short FE implementation for a 2d homogeneous Dirichlet problem of a fractional Laplacian. *Comput. Math. Appl.*, 74(4):784–816, 2017.

[2] Gabriel Acosta and Juan Pablo Borthagaray. A fractional Laplace equation: regularity of solutions and finite element approximations. *SIAM J. Numer. Anal.*, 55(2):472–495, 2017.

[3] Robert A. Adams and John J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.

[4] David Applebaum. *Lévy processes and stochastic calculus*, volume 93 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2004.

[5] Pascal Auscher, Steve Hofmann, Michael Lacey, Alan McIntosh, and Ph. Tchamitchian. The solution of the Kato square root problem for second order elliptic operators on $\mathbb{R}^n$. *Ann. of Math. (2)*, 156(2):633–654, 2002.

[6] A Balakrishnan. Fractional powers of closed operators and the semigroups generated by them. *Pacific Journal of Mathematics*, 10(2):419–437, 1960.

[7] W. Bangerth, R. Hartmann, and G. Kanschat. deal.II—a general-purpose object-oriented finite element library. *ACM Trans. Math. Software*, 33(4):Art. 24, 27, 2007.

[8] Andrea Bonito, Juan Pablo Borthagaray, Ricardo H Nochetto, Enrique Otarola, and Abner J Salgado. Numerical methods for fractional diffusion. *Comput. Vis. Sci., to appear. ArXiv preprint arXiv:1707.01566*, 2017.

[9] Andrea Bonito, Wenyu Lei, and Joseph E Pasciak. Numerical approximation of space-time fractional parabolic equations. *Comput. Methods Appl. Math.*, 17:679–705, 2017.

[10] Andrea Bonito, Wenyu Lei, and Joseph E Pasciak. Numerical approximation of the integral fractional laplacian. *arXiv preprint arXiv:1707.04290*, 2017.

[11] Andrea Bonito, Wenyu Lei, and Joseph E Pasciak. On sinc quadrature approximations of fractional powers of regularly accretive operators. *J. Numer. Math.*, 2018.

[12] Andrea Bonito and Joseph E. Pasciak. Numerical approximation of fractional powers of elliptic operators. *Math. Comp.*, 84(295):2083–2110, 2015.

[13] Andrea Bonito and Joseph E. Pasciak. Numerical approximation of fractional powers of regularly accretive operators. *IMA J. Numer. Anal.*, 37(3):1245–1273, 2017.

[14] Juan Pablo Borthagaray, Leandro M Del Pezzo, and Sandra Martínez. Finite element approximation for the fractional eigenvalue problem. *arXiv preprint arXiv:1603.00317*, 2016.

[15] Svetlana I Boyarchenko and Sergei Z Levendorskii. Perpetual american options under lévy processes. *SIAM Journal on Control and Optimization*, 40(6):1663–1696, 2002.

[16] James H. Bramble, Joseph E. Pasciak, and Panayot S. Vassilevski. Computational scales of Sobolev norms with application to preconditioning. *Math. Comp.*, 69(230):463–480, 2000.

[17] James H. Bramble and Jinchao Xu. Some estimates for a weighted L^2 projection. *Math. Comp.*, 56(194):463–476, 1991.

[18] James H. Bramble and Xuejun Zhang. The analysis of multigrid methods. In *Handbook of numerical analysis, Vol. VII*, Handb. Numer. Anal., VII, pages 173–415. North-Holland, Amsterdam, 2000.

[19] Mark Broadie and Jérôme Detemple. The valuation of american options on multiple assets. *Mathematical Finance*, 7(3):241–286, 1997.

[20] Luis Caffarelli and Luis Silvestre. An extension problem related to the fractional laplacian. *Communications in partial differential equations*, 32(8):1245–1260, 2007.

[21] Luis A Caffarelli, Sandro Salsa, and Luis Silvestre. Regularity estimates for the solution and the free boundary of the obstacle problem for the fractional laplacian. *Inventiones mathematicae*, 171(2):425–461, 2008.

[22] Peter Carr, Hélyette Geman, Dilip B Madan, and Marc Yor. The fine structure of asset returns: An empirical investigation. *The Journal of Business*, 75(2):305–332, 2002.

[23] S. N. Chandler-Wilde, D. P. Hewett, and A. Moiola. Interpolation of Hilbert and Sobolev spaces: quantitative estimates and counterexamples. *Mathematika*, 61(2):414–443, 2015.

[24] P. Ciarlet, Jr. Analysis of the Scott-Zhang interpolation in the fractional order Sobolev spaces. *J. Numer. Math.*, 21(3):173–180, 2013.

[25] Philippe G. Ciarlet. *The finite element method for elliptic problems*, volume 40 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002. Reprint of the 1978 original [North-Holland, Amsterdam; MR0520174 (58 #25001)].

[26] Peter Constantin, Tarek Elgindi, Mihaela Ignatova, and Vlad Vicol. On some electroconvection models. *Journal of Nonlinear Science*, 27(1):197–211, 2017.

[27] John H Cushman and TR Ginn. Nonlocal dispersion in media with continuously evolving scales of heterogeneity. *Transport in Porous Media*, 13(1):123–138, 1993.

[28] Timothy A Davis. Umfpack version 5.2. 0 user guide. *University of Florida*, 2007.

[29] Klaus Deckelnick, Gerhard Dziuk, Charles M Elliott, and Claus-Justus Heine. An h-narrow band finite-element method for elliptic equations on implicit surfaces. *IMA Journal of Numerical Analysis*, 30(2):351–376, 2009.

[30] Klaus Deckelnick, Charles M Elliott, and Thomas Ranner. Unfitted finite element methods using bulk meshes for surface partial differential equations. *SIAM Journal on Numerical Analysis*, 52(4):2137–2162, 2014.

[31] Marta D'Elia and Max Gunzburger. The fractional Laplacian operator on bounded domains as a special case of the nonlocal diffusion operator. *Comput. Math. Appl.*, 66(7):1245–1260, 2013.

[32] Françoise Demengel and Gilbert Demengel. *Functional spaces for the theory of elliptic partial differential equations*. Universitext. Springer, London; EDP Sciences, Les Ulis, 2012. Translated from the 2007 French original by Reinie Erné.

[33] Gerhard Dziuk. Finite elements for the beltrami operator on arbitrary surfaces. *Partial differential equations and calculus of variations*, 1357:142–155, 1988.

[34] Alexandre Ern and Jean-Luc Guermond. *Theory and practice of finite elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.

[35] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[36] Ivan P Gavrilyuk, Wolfgang Hackbusch, and Boris N Khoromskij. $\mathcal{H}$-matrix approximation for the operator exponential with applications. *Numerische Mathematik*, 92(1):83–111, 2002.

[37] Pierre Grisvard. *Elliptic problems in nonsmooth domains*, volume 69 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011. Reprint of the 1985 original [MR0775683], With a foreword by Susanne C. Brenner.

[38] Gerd Grubb. Fractional Laplacians on domains, a development of Hörmander's theory of μ-transmission pseudodifferential operators. *Adv. Math.*, 268:478–528, 2015.

[39] Isaac M Held, Raymond T Pierrehumbert, Stephen T Garner, and Kyle L Swanson. Surface quasi-geostrophic dynamics. *Journal of Fluid Mechanics*, 282:1–20, 1995.

[40] Yanghong Huang and Adam Oberman. Numerical methods for the fractional Laplacian: a finite difference–quadrature approach. *SIAM J. Numer. Anal.*, 52(6):3056–3084, 2014.

[41] M Ilic, F Liu, I Turner, and V Anh. Numerical approximation of a fractional-in-space diffusion equation, i. *Fractional Calculus and Applied Analysis*, 8(3):323–341, 2005.

[42] David Jerison and Carlos E. Kenig. The inhomogeneous Dirichlet problem in Lipschitz domains. *J. Funct. Anal.*, 130(1):161–219, 1995.

[43] Tosio Kato. Fractional powers of dissipative operators. *J. Math. Soc. Japan*, 13:246–274, 1961.

[44] Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

[45] J.-L. Lions and E. Magenes. *Non-homogeneous boundary value problems and applications. Vol. I*. Springer-Verlag, New York-Heidelberg, 1972. Translated from the French by P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 181.

[46] M. López-Fernández and C. Palencia. On the numerical inversion of the Laplace transform of certain holomorphic mappings. *Appl. Numer. Math.*, 51(2-3):289–303, 2004.

[47] Alessandra Lunardi. *Interpolation theory*. Appunti. Scuola Normale Superiore di Pisa (Nuova Serie). [Lecture Notes. Scuola Normale Superiore di Pisa (New Series)]. Edizioni della Normale, Pisa, second edition, 2009.

[48] John Lund and Kenneth L. Bowers. *Sinc methods for quadrature and differential equations*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[49] Mitchell Luskin and Rolf Rannacher. On the smoothing property of the Crank-Nicolson scheme. *Applicable Anal.*, 14(2):117–135, 1982/83.

[50] Ana-Maria Matache, Christoph Schwab, and Thomas P Wihler. Fast numerical solution of parabolic integrodifferential equations with applications in finance. *SIAM Journal on Scientific Computing*, 27(2):369–393, 2005.

[51] William McLean. *Strongly elliptic systems and boundary integral equations*. Cambridge University Press, Cambridge, 2000.

[52] William McLean and Vidar Thomee. Iterative solution of shifted positive-definite linear systems arising in a numerical method for the heat equation based on laplace transformation and quadrature. *The ANZIAM Journal*, 53(02):134–155, 2011.

[53] Ralf Metzler and Joseph Klafter. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.*, 339(1):77, 2000.

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

[55] Ricardo H Nochetto, Enrique Otarola, and Abner J Salgado. A pde approach to space-time fractional parabolic problems. *SIAM Journal on Numerical Analysis*, 54(2):848–873, 2016.

[56] Ricardo H Nochetto, Tobias von Petersdorff, and Chen-Song Zhang. A posteriori error analysis for a class of integral equations and variational inequalities. *Numerische Mathematik*, 116(3):519–552, 2010.

[57] Maxim A Olshanskii, Arnold Reusken, and Jörg Grande. A finite element method for elliptic equations on surfaces. *SIAM journal on numerical analysis*, 47(5):3339–3358, 2009.

[58] FWJ Olver, DW Lozier, RF Boisvert, and CW Clark. Nist digital library of mathematical functions. *Online companion to [65]: http://dlmf. nist. gov*, 2010.

[59] Xavier Ros-Oton and Joaquim Serra. The dirichlet problem for the fractional laplacian: regularity up to the boundary. *Journal de Mathématiques Pures et Appliquées*, 101(3):275–302, 2014.

[60] Stefan A Sauter and Christoph Schwab. Boundary element methods. In *Boundary Element Methods*, pages 183–287. Springer, 2010.

[61] L. Ridgway Scott and Shangyou Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54(190):483–493, 1990.

[62] Dongwoo Sheen, Ian Sloan, and Vidar Thomée. A parallel method for time-discretization of parabolic problems based on contour integral representation and quadrature. *Mathematics of Computation of the American Mathematical Society*, 69(229):177–195, 2000.

[63] Jonathan Richard Shewchuk. Delaunay refinement algorithms for triangular mesh generation. *Comput. Geom.*, 22(1-3):21–74, 2002. 16th ACM Symposium on Computational Geometry (Hong Kong, 2000).

[64] Stewart A Silling. Reformulation of elasticity theory for discontinuities and long-range forces. *Journal of the Mechanics and Physics of Solids*, 48(1):175–209, 2000.

[65] David W Sims, Emily J Southall, Nicolas E Humphries, Graeme C Hays, Corey JA Bradshaw, Jonathan W Pitchford, Alex James, Mohammed Z Ahmed, Andrew S Brierley, Mark A Hindell, et al. Scaling laws of marine predator search behaviour. *Nature*, 451(7182):1098, 2008.

[66] Olaf Steinbach. *Numerical approximation methods for elliptic boundary value problems*. Springer, New York, 2008. Finite and boundary elements, Translated from the 2003 German original.

[67] Pablo Raúl Stinga and José Luis Torrea. Extension problem and harnack's inequality for some fractional operators. *Communications in Partial Differential Equations*, 35(11):2092–2122, 2010.

[68] Luc Tartar. *An introduction to Sobolev spaces and interpolation spaces*, volume 3 of *Lecture Notes of the Unione Matematica Italiana*. Springer, Berlin; UMI, Bologna, 2007.

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

[70] Jinchao Xu. *Theory of multilevel methods*, volume 8924558. Cornell University Ithaca, NY, 1989.