A Novel Analytic Diagonalization Technique Finite Element Method for the Spectral Fractional Laplacian

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Shane E. Sawyer

August 2024

To my wife Paula for her unending love and support through all of our journeys together. To the memory of my parents who provided support all along the way. And to Elaina, to whom I hope to be the best father I can be.

Acknowledgments

I whole heatedly wish to thank my advisor, Dr. Abner Salgado, whose unbounded patience was instrumental in the completion of my research and degree. I also wish to thank Dr. Steven Wise for his guidance on my understanding of numerical analysis. To Dr. Tuoc Phan, I deeply thank for is work on my committee and for the many classes I had with him. To Dr. Matthias Maier, I thank for his help in reading through my dissertation and providing insight on how to best utilize the deal.ii software library.

There have been many classmates that have been instrumental in providing assistance and commiseration. To name but a few, I wish to offer thanks to Joshua Siktar, Liet Vo, Daniel McBride, Mitchell Sutton, Anna Sisk, Wencel Valega-Mackenzie, Joseph Daws, and Michael Wise.

A special thanks to my "blood brother" David Turpin whose gift allowed me to finish this work, and the rest of my life.

Abstract

The goal of this dissertation is to present a new technique for approximating solutions to problems involving the spectral fractional Laplacian. Previous works have used the Dirichlet-to-Neumann extension technique of Caffarelli and Silvestre, together with a diagonalization method to reduce computational complexity. Building on this method, a novel scheme is proposed where the analytic solution to the associated eigenvalue problem in the extended dimension is used, thus avoiding the numerical issues of ill conditioning in computing the eigenpairs. It is then shown how this new method is related to a quadrature scheme to approximate the spectral fractional Laplacian via a Balakrishnan integral formula. The quadrature scheme used in the algorithm demonstrates exponential convergence to the analytic integral. Numerical examples illustrate the theoretical convergence rates. The parallel performance of the algorithm is studied using both strong and weak scaling.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Nonlocal and fractional order operators are of increasing interest in a variety of fields of study. Application areas include image processing and denoising [29], electroconvection [16], quasi-geostrophic flow [13], cardiac electrophysiology [21, 22], and anomalous diffusion [38], to name but a very few of them.

In this work, problems involving fractional powers of the Dirichlet Laplacian $(-\Delta)^s$, where $s \in (0,1)$, are studied. To be specific, the domain Ω will be a convex, open, and bounded subset of $\mathbb{R}^d$, $d \geq 1$, with Lipschitz boundary $\partial\Omega$, and the problem of interest is stated as: given $s \in (0,1)$ and a sufficiently smooth f, find u such that

$$(-\Delta)^s u = f \qquad \text{in } \Omega\,. \tag{1.1}$$

The meaning of this operator will be specified later in Chapter 2. At this stage it suffices to say that, stated this way, the fractional Laplacian is a nonlocal operator, making it difficult to solve by traditional numerical methods. It has been shown by Caffarelli and Silvestre [19] that the solution of (1.1) in $\mathbb{R}^d$ can be found by extending the problem to the upper half space $\mathbb{R}^{d+1}_+$ via a Dirichlet-to-Neumann map. This extension method of Caffarelli and Silvestre has been adapted to bounded domains, $\Omega \subset \mathbb{R}^d$, in [18, 43]. This poses the extension problem on the semi-infinite cylinder $\mathcal{C} = \Omega \times (0,\infty)$. The problem now reads as follows: find $\mathcal{U}$ such that

$$\begin{cases} \text{div } (y^\alpha \nabla \mathcal{U}) = 0\,, & \text{in } \mathcal{C}\,, \\ \mathcal{U} = 0\,, & \text{on } \partial_L \mathcal{C}\,, \\ \dfrac{\partial \mathcal{U}}{\partial \nu^\alpha} = d_s f\,, & \text{on } \Omega \times \{0\}\,, \end{cases} \tag{1.2}$$

where $\partial_L \mathcal{C} = \partial\Omega \times [0, \infty)$ is the lateral boundary of the cylinder, and by

$$\frac{\partial \mathcal{U}}{\partial \nu^\alpha} = -\lim_{y \downarrow 0} y^\alpha \mathcal{U}_y$$

the conormal derivative of $\mathcal{U}$ is denoted. Here, ν represents the unit outer normal vector to $\partial\mathcal{C}$ along $\Omega \times \{0\}$. The parameter α is related to s by

$$\alpha = 1 - 2s \in (-1, 1)\,. \tag{1.3}$$

Lastly, d_s is a positive normalization constant that depends only on the power s via

$$d_s = \frac{2^{1-2s}\Gamma(1-s)}{\Gamma(s)}\,. \tag{1.4}$$

As seen in [19, 18, 43], the solution to the fractional Laplacian in (1.1) is related to the Dirichlet-to-Neumann operator in (1.2) by

$$d_s(-\Delta)^s u = \frac{\partial \mathcal{U}}{\partial \nu^\alpha} \qquad \text{in } \Omega\,.$$

The approach taken in this dissertation is not the first attempt at solving the fractional Laplacian. Numerical schemes have been introduced to approximate solutions of (1.1) by computing numerical solutions of (1.2). A piecewise linear finite element scheme was introduced and analyzed in [39]. This method was improved in many ways in the subsequent work [8]. The improvements include a tensor product formulation with *hp*-FEM in the extended variable y coupled to P_1-FEM in Ω. A diagonalization technique was also introduced that allows the Caffarelli-Silvestre extension to decouple into independent, singularly perturbed second-order reaction-diffusion problems in Ω. Lastly, the use of *hp*-FEM in Ω, along with certain regularity assumptions, provided for exponential convergence to u in Ω.

The present work builds on the improvements presented in [8] by replacing the *hp*-FEM scheme in the extended variable with the analytical solution to the one dimensional second-order eigenvalue problem induced by the diagonalization technique. The resulting scheme is called an *exact diagonalization technique* to distinguish it from earlier approaches. By using the analytic solution, the instability inherent in the numerical approximation to a large eigenvalue problem [47] can be mitigated. For the analysis of this method, it is shown that this technique is related to similar work in [14] which uses a discretization of the so-called Balakrishnan formula for fractional powers of positive operators [7, 11].

Before moving along, it is instructive to mention some related approaches to the numerical approximation of solutions to (1.1). For a more thorough overview of the literature, there are several references available [12, 23, 17, 35]. Reference [46] studies the numerical solution to problems with elliptic operators with a fractional power. In order to apply the method proposed in [46] to problem (1.1) it should be first observed that, by setting $A = -\Delta$, it is seen that $A = A^* \geq \delta I$, where $\delta > 0$. The problem at hand is then transformed into a pseudo-parabolic problem

$$\big(t(A - \delta I) + \delta I\big)\frac{dw}{dt} + s(A - \delta I)w = 0\,.$$

The accompanying initial condition is $w(0) = \delta^{-s} f$. The solution of this pseudo-parabolic problem is related to the fractional problem by $u = w(1)$. A detailed analysis of this method, with further insight, is given in [26].

Another class of approaches uses the idea of rational approximation. In [2] this is applied to a fractional-in-space reaction-diffusion equation

$$\frac{\partial u(x,t)}{\partial t} = -\kappa_{2s}(-\Delta)^s u(x,t) + f(x,t,u)\,, \quad x \in \Omega \quad \text{and} \quad t \in (0,T)\,,$$

where κ_{2s} is the diffusion coefficient. The equation is subject to the boundary condition $u(x,t)|_{x\in\partial\Omega} = 0$ and has initial condition prescribed by $u(x,0) = u_0(x)$. Note that in this work s is restricted to $(1/2, 1]$. The first step in the proposed method is to discretize the Laplacian by means of a finite difference method over a uniform mesh resulting in the matrix approximation $h^{-2}L$. Then, the scheme applies the fractional power s to the matrix L obtaining

$$-(-\Delta)^s \approx -\frac{1}{h^{2s}} L^s \approx -\frac{1}{h^{2s}} M^{-1} K \,,$$

where M and K are banded matrices that are found by a rational approximation of z^{s-1} using the Gauss-Jacobi rule applied to the Dunford-Taylor integral representation of L^s [2]. A similar approach, but one using a best uniform rational approximation (BURA) is proposed in [31, 30]. These works start with a finite element discretization of an elliptic operator, $\hat{\mathcal{A}}$, and consider solving the algebraic problem

$$\mathcal{A}^s \mathbf{u} = \mathbf{f} \,,$$

where $\mathcal{A}$ is a rescaling of $\hat{\mathcal{A}}$ so that the spectrum lies in $(0, 1]$. The authors of [31] then determine the BURA to z^{1-s} on the interval $(0, 1]$ using a modified Remez algorithm.

A scheme involving a quadrature formula for the Dunford-Taylor integral representation of the fractional Laplacian is developed in [14]. In particular, it is observed that the Dunford-Taylor integral is equivalent to the the Balakrishnan integral

$$(-\Delta)^{-s} = \frac{\sin(s\pi)}{\pi} \int_0^\infty t^{-s} (tI - \Delta)^{-1} dt \,.$$

Then, to approximate the solution u, $(t_i I - \Delta)^{-1} f$ is computed via a finite element method at the quadrature nodes t_i. Using an appropriate change of variables, [14] shows that the quadrature scheme converges exponentially fast with respect to the number of quadrature nodes k. Subsequent work extended the approach to more general operators in [15]. The following error estimate was proved for convex domains Ω [14, 12]:

$$\|u - u_h^k\|_{L^2(\Omega)} \lesssim h^2 \|f\|_{\mathbb{H}^{2-2s}(\Omega)} \,,$$

where u is the solution to (1.1) and u_h^k represents its numerical approximation.

Other approaches have been proposed that utilize the Caffarelli-Silvestre extension but with different tactics than those in [39, 8] or in this work. A hybrid FE-spectral method to solve the extended problem based on the Caffarelli-Silvestre extension is proposed in [4]. Their approach uses the first M eigenvalues of the Laplacian. The lower part of the spectrum is approximated by a FEM on Ω and the upper part of the spectrum is approximated using Weyl's asymptotic [33]. For convenience let us denote by h the mesh size of the finite element space $\mathbb{V}_h$, defined in Ω, k the polynomial order of the FEM in $\mathbb{V}_h$, $r > -s$ the regularity of f, and $\mathcal{U}_h^M$ the discrete solution. Under the assumption that M is large enough so that $\lambda_M^{-(r+s)/2} \sim h^{\min\{k,r+s\}}$, the following error estimate is provided [4, Theorem 2]:

$$\|\nabla(\mathcal{U} - \mathcal{U}_h^M)\|_{L^2(y^\alpha,\mathcal{C}_\mathcal{Y})} \lesssim h^{\min\{k,r+s\}} \sqrt{|\log h|}\ \|f\|_{\mathbb{H}^r(\Omega)} .$$

If $M \sim \mathcal{O}(\log^p \dim\mathbb{V}_h)$ for some $p \geq 0$, the error estimate may be written as [4, Equation (28)]

$$\|\nabla(\mathcal{U} - \mathcal{U}_h^M)\|_{L^2(y^\alpha,\mathcal{C}_\mathcal{Y})} \lesssim |\log(M \cdot \dim\mathbb{V}_h)|^q \big(M \cdot \dim\mathbb{V}_h\big)^{-\min\{k,r+s\}/d} \|f\|_{\mathbb{H}^r(\Omega)} ,$$

for some $q \geq 0$.

A numerical method that uses generalized Laguerre functions as a basis for the extended dimension was proposed in [20]. Since the weight y^α appears in the problem, the use of Laguerre functions is natural. They call their scheme an enriched Laguerre spectral method, indicating that the basis in the extended dimension has been supplemented with additional functions to compensate for the singular/degenerate weight. Denoting by N the number of generalized Laguerre basis functions used in the extended dimension and by k the number of additional functions used to enrich the Laguerre basis, the following error estimate is proved in [20]:

$$\begin{aligned}\|\nabla(\mathcal{U} - \mathcal{U}_h)\|_{L^2(y^\alpha,\mathcal{C}_\mathcal{Y})} \lesssim \quad & N^{-m/2}\|f\|_{\mathbb{H}^{m/2-s}(\Omega)} + N^{1-m/2}\|f\|_{\mathbb{H}^{(m-1)/2-s}(\Omega)} \\ & + h^r\|f\|_{\mathbb{H}^{r+1/2-s}(\Omega)} + h^{r-1}\|f\|_{\mathbb{H}^{r-1-s}(\Omega)} ,\end{aligned}$$

where $r > -s$, it is assumed that $f \in \mathbb{H}^{l-s}(\Omega)$ with $l = \max\{\frac{m}{2}, r + \frac{1}{2}\}$, $\alpha = 1 - 2s$, and $m = 2k + 1 + [2s]$.

Finally, it should also be noted that [32] showed that all the schemes in [46, 2, 31, 14, 39, 8] can be understood as methods based on rational approximation of a univariate function.

It is now instructive to summarize the convergence results from the previous work in [39, 8] that is the foundation of the novel method described in this dissertation. As before, the discrete solution is denoted by $\mathcal{U}_h$. Recall, $\mathbb{V}_h$ denotes the finite element space. By $\mathcal{S}_\mathcal{Y}$ the finite element space in the extended direction is denoted. For the following estimates, $\mathbb{V}_h$ will consist of piecewise linear functions on a conforming and quasi-uniform mesh of quadrilaterals on Ω. If $\mathcal{S}_\mathcal{Y}$ is taken to be piecewise linear functions on quasi-uniform intervals, the following error estimate holds [39, Theorem 5.1]:

$$\|\nabla(\mathcal{U} - \mathcal{U}_h)\|_{L^2(y^\alpha, \mathcal{C}_\mathcal{Y})} \lesssim h^{s-\epsilon}\|f\|_{\mathbb{H}^{1-s}(\Omega)} \sim \left(\dim\mathbb{V}_h \cdot \dim\mathcal{S}_\mathcal{Y}\right)^{-\frac{s-\epsilon}{d+1}}\|f\|_{\mathbb{H}^{1-s}(\Omega)}\,.$$

If instead $\mathcal{S}_\mathcal{Y}$ consists of piecewise linear functions on quasi-uniform intervals that are graded towards $y = 0$, the error estimate can be improved as shown in [39, Equation (5.11)] to

$$\|\nabla(\mathcal{U} - \mathcal{U}_h)\|_{L^2(y^\alpha, \mathcal{C}_\mathcal{Y})} \lesssim |\log(\dim\mathbb{V}_h \cdot \dim\mathcal{S}_\mathcal{Y})|^s\left(\dim\mathbb{V}_h \cdot \dim\mathcal{S}_\mathcal{Y}\right)^{-1/(d+1)}\|f\|_{\mathbb{H}^{1-s}(\Omega)}\,.$$

Lastly, in [8], $\mathcal{S}_\mathcal{Y}$ uses *hp*-FEM on intervals with a geometric grading towards $y = 0$ in the extended dimension. The error estimate is given in [8, Equation (5.57)] as

$$\|\nabla(\mathcal{U} - \mathcal{U}_h)\|_{L^2(y^\alpha, \mathcal{C}_\mathcal{Y})} \lesssim |\log(\dim\mathbb{V}_h)|^q\left(\dim\mathbb{V}_h\right)^{-1/d}\|f\|_{\mathbb{H}^{1-s}(\Omega)}\,,$$

where $q > 0$.

Having reviewed the existing methods and their analysis the dissertation now proceeds with the development of our semi-analytic technique. The presentation is organized as follows. In Chapter 2, the notation and preliminary results needed to study the problem are presented. In Chapter 3, the diagonalization technique and the analytical solution to the eigenvalue problem is discussed. Chapter 4 contains the analysis of the new method. A connection between the exact diagonalization scheme and quadrature of the Balakrishnan formula for the inverse of the fractional Laplacian operator along with error estimates is drawn. Lastly, in Chapter 5, numerical results are presented to demonstrate the convergence of the scheme as well as the parallel performance of the algorithm.

Chapter 2

Preliminaries and Notation

Throughout our work, it assumed that Ω is a convex, open, bounded, and connected subset of $\mathbb{R}^d$, $d \geq 1$, such that $\partial\Omega$ is polyhedral. The semi-infinite cylinder is defined as $\mathcal{C} = \Omega \times (0, \infty)$, along with its lateral boundary $\partial_L \mathcal{C} = \partial\Omega \times [0, \infty)$. Similarly, for any $\mathcal{Y} > 0$, the truncated cylinder is denoted by $\mathcal{C}_{\mathcal{Y}} = \Omega \times (0, \mathcal{Y})$, along with the corresponding lateral boundary $\partial_L \mathcal{C}_{\mathcal{Y}} = \partial\Omega \times [0, \mathcal{Y}]$. It will also be prudent to distinguish between points in $\mathbb{R}^d$ and those in $\mathbb{R}^{d+1}_+$. Set $\mathbf{x} = (x, y)$, where $x \in \mathbb{R}^d$ and $y \in (0, \infty)$. Another common notation is given by $a \lesssim b$ to indicate that $a \leq Cb$, where the constant C is independent of either a or b and does not depend on the discretization scheme. By $a \sim b$, it is meant that $a \lesssim b \lesssim a$, where the constants on either end of the inequalities are generally different from one another. Standard Sobolev spaces such as $H^1(\Omega)$, $H^1_0(\Omega)$, and $H^\sigma(\Omega)$, with $\sigma \in \mathbb{R}$ [3, 24], are used throughout the dissertation.

2.1 The Spectral Fractional Laplacian

In this work, the spectral interpretation of $(-\Delta)^s$ is employed, which is briefly summarized below. Define the operator $(-\Delta) : L^2(\Omega) \to L^2(\Omega)$ with domain

$$\mathrm{Dom}\ (-\Delta) = \{w \in H^1_0(\Omega) : \Delta w \in L^2(\Omega)\}.$$

This operator is positive, unbounded, closed, and possesses a compact inverse. Hence, the spectrum of $(-\Delta)$ is discrete, positive, and accumulates at infinity.

In other words, there exists a collection of eigenpairs, $\{(\lambda_k, \phi_k)\}_{k\geq 1} \in \mathbb{R}_+ \times H_0^1(\Omega) \setminus \{0\}$, where, for each $k \geq 1$,

$$\begin{cases} -\Delta\phi_k = \lambda_k\phi_k\,, & \text{in } \Omega\,,\\ \phi_k = 0\,, & \text{on } \partial\Omega\,. \end{cases}$$

In addition, $\lambda_1 = \min\{\lambda_k\}_{k\geq 1}$ is simple and $\lambda_k \to \infty$ as $k \to \infty$. Moreover, $\{\phi_k\}_{k\geq 1}$ forms an orthonormal basis of $L^2(\Omega)$ [28, Section 6.5, Theorem 1]. Finally, $\{\phi_k\}_{k\geq 1}$ is also an orthogonal basis of $H_0^1(\Omega)$ with $\|\nabla\phi_k\|^2_{L^2(\Omega)} = \lambda_k$.

Fractional powers of the Dirichlet Laplacian can now be defined using this spectral decomposition. For any $w \in C_0^\infty(\Omega)$, the spectral fractional Laplacian can be written as

$$(-\Delta)^s w = \sum_{k=1}^{\infty} w_k \lambda_k^s \phi_k\,, \tag{2.1}$$

where $w_k = \int_\Omega w\phi_k\,dx$.

Suppose now that it is desired to solve (1.1) and $f = \sum_{k\geq 1} f_k\phi_k$. Owing to the orthonormality of $\{\phi_k\}_{k\geq 1}$, and defining $u_k = \lambda_k^{-s} f_k$ for all $k \in \mathbb{N}$, the function $u = \sum_{k\geq 1} u_k\phi_k$ solves (1.1), provided all these series converge in a suitable sense. This is made rigorous by introducing, for $r \geq 0$, the following Hilbert space:

$$\mathbb{H}^r(\Omega) = \left\{ w = \sum_{k=1}^{\infty} w_k\phi_k \in L^2(\Omega) : \|w\|^2_{\mathbb{H}^r(\Omega)} := \sum_{k=1}^{\infty} \lambda_k^r |w_k|^2 < \infty \right\}.$$

For $r < 0$ the space $\mathbb{H}^r(\Omega)$ is the dual of $\mathbb{H}^{-r}(\Omega)$. Further characterizations of these spaces can be seen in [39, Equation (2.13)].

After all this notation and these constructions, a rigorous meaning can be endowed to (1.1). Namely, for $s \in (0,1)$, the operator $(-\Delta)^s : \mathbb{H}^s(\Omega) \to \mathbb{H}^{-s}(\Omega)$ is an isometry. For this reason, if $f = \sum_{k\geq 1} f_k\phi_k \in \mathbb{H}^{-s}(\Omega)$, the solution to (1.1) is given by

$$u = \sum_{k\geq 1} \lambda_k^{-s} f_k\phi_k \in \mathbb{H}^s(\Omega), \tag{2.2}$$

with $\|u\|_{\mathbb{H}^s(\Omega)} = \|f\|_{\mathbb{H}^{-s}(\Omega)}$.

2.2 The Caffarelli-Silvestre Extension

The Caffarelli-Silvestre extension [19] necessitates the handling of both degenerate and singular elliptic equations on the space $\mathbb{R}^{d+1}_+ = \mathbb{R}^d \times \mathbb{R}_+$, or in the case of a bounded domain, the extension [18, 43] requires handling such equations on $\mathcal{C}$. To study such equations weighted Sobolev spaces [34] are used with the weight y^α, where α is defined in (1.3).

Starting with some domain $\mathcal{D} \subset \mathbb{R}^{d+1}_+$, define $L^2(y^\alpha, \mathcal{D})$ as the space containing all measurable functions w defined on $\mathcal{D}$ such that

$$\|w\|^2_{L^2(y^\alpha,\mathcal{D})} \coloneqq \int_{\mathcal{D}} |y|^\alpha w(\mathbf{x})^2 \, d\mathbf{x} < \infty \, .$$

Using this definition, a similar definition holds for the weighted Sobolev space

$$H^1(y^\alpha, \mathcal{D}) = \left\{ w \in L^2(y^\alpha, \mathcal{D}) : |\nabla w| \in L^2(y^\alpha, \mathcal{D}) \right\} \, ,$$

where $|\nabla w|$ should be interpreted as a distributional derivative. The space is equipped with the following norm:

$$\|w\|^2_{H^1(y^\alpha,\mathcal{D})} = \|w\|^2_{L^2(y^\alpha,\mathcal{D})} + \|\nabla w\|^2_{L^2(y^\alpha,\mathcal{D})}.$$

Of particular interest will be the weighted Sobolev space defined on $\mathcal{C}$ of functions that vanish at the lateral boundary, given by

$$\mathring{H}^1_L(y^\alpha, \mathcal{C}) = \left\{ w \in H^1(y^\alpha, \mathcal{C}) : w|_{\partial_L \mathcal{C}} = 0 \right\} \, .$$

It should be noted that this space has a weighted Poincaré inequality [8, (2.21)] and as a result the $\mathring{H}^1_L(y^\alpha, \mathcal{C})$ semi-norm is an equivalent norm on this space. Finally, it should be commented that functions in $\mathring{H}^1_L(y^\alpha, \mathcal{C})$ have well defined traces in $\Omega \times \{0\}$. By this it is meant that there is a continuous and surjective mapping $\operatorname{tr} : \mathring{H}^1_L(y^\alpha, \mathcal{C}) \to \mathbb{H}^s(\Omega)$ such that, for every smooth $w \in \mathring{H}^1_L(y^\alpha, \mathcal{C})$, it holds that

$$\text{tr } w(x) = w(x, 0), \qquad \forall x \in \Omega\,.$$

Details of this observation may be found in [39, Proposition 2.5].

A bilinear form can be defined as $a_{\mathcal{C}} : \mathring{H}^1_L(y^\alpha, \mathcal{C}) \times \mathring{H}^1_L(y^\alpha, \mathcal{C}) \to \mathbb{R}$ by

$$a_{\mathcal{C}}(v, w) = \int_{\mathcal{C}} y^\alpha \nabla v \cdot \nabla w \, d\mathbf{x}\,,$$

which is bounded and coercive. With these preliminaries at hand, a weak formulation of (1.2) can be constructed: find $\mathcal{U} \in \mathring{H}^1_L(y^\alpha, \mathcal{C})$ such that

$$a_{\mathcal{C}}(\mathcal{U}, v) = d_s \langle f, \text{tr } v \rangle\,, \quad \forall\, v \in \mathring{H}^1_L(y^\alpha, \mathcal{C})\,, \tag{2.3}$$

where $\langle \cdot, \cdot \rangle$ is the duality pairing between $\mathbb{H}^s(\Omega)$ and $\mathbb{H}^{-s}(\Omega)$.

The main result of Caffarelli and Silvestre can now be given.

Theorem 2.1 (extension)**.** *Let $f \in \mathbb{H}^{-s}(\Omega)$ and assume that $\mathcal{U} \in \mathring{H}^1_L(y^\alpha, \mathcal{C})$ solves (2.3). Then $u = tr\, \mathcal{U} \in \mathbb{H}^s(\Omega)$ solves (1.1) in the sense that it satisfies (2.2).*

Proof. See [19] for the case $\Omega = \mathbb{R}^d$, and [18, 43] for bounded domains. □

2.3 Truncation

As a first step towards the numerical approximation of the solution to (1.1) via (2.3), the infinite cylinder is truncated to a finite height. For $\mathcal{Y} > 0$, introduce

$$\mathring{H}^1_L(y^\alpha, \mathcal{C}_{\mathcal{Y}}) = \left\{ w \in H^1(y^\alpha, \mathcal{C}_{\mathcal{Y}}) : w|_{\partial_L \mathcal{C}_{\mathcal{Y}} \cup (\Omega \times \{\mathcal{Y}\})} = 0 \right\}\,.$$

The following problem is now introduced: find $\mathcal{U}_{\mathcal{Y}} \in \mathring{H}^1_L(y^\alpha, \mathcal{C}_{\mathcal{Y}})$ such that

$$a_{\mathcal{C}_{\mathcal{Y}}}(\mathcal{U}_{\mathcal{Y}}, v) = d_s \langle f, \text{tr } v \rangle\,, \quad \forall\, v \in \mathring{H}^1_L(y^\alpha, \mathcal{C}_{\mathcal{Y}})\,, \tag{2.4}$$

where the bilinear form $a_{\mathcal{C}_\mathcal{Y}}$ is defined in an analogous manner to $a_\mathcal{C}$.

The truncation error is quantified below.

Proposition 2.2 (truncation). *Let $\mathcal{Y} \geq 1$. If $\mathcal{U} \in \mathring{H}^1_L(y^\alpha, \mathcal{C})$ solves (2.3), then we have*

$$\int_\mathcal{Y}^\infty \int_\Omega y^\alpha |\nabla \mathcal{U}|^2 \, d\mathbf{x} \lesssim e^{-\sqrt{\lambda_1}\mathcal{Y}} \|f\|^2_{\mathbb{H}^{-s}(\Omega)}.$$

Consequently, if $\mathcal{U}_\mathcal{Y} \in \mathring{H}^1_L(y^\alpha, \mathcal{C}_\mathcal{Y})$ solves (2.4), we have

$$\|u - tr\, \mathcal{U}_\mathcal{Y}\|_{\mathbb{H}^s(\Omega)} = \|\nabla(\mathcal{U} - \mathcal{U}_\mathcal{Y})\|_{L^2(y^\alpha, \mathcal{C})} \lesssim e^{-\sqrt{\lambda_1}\mathcal{Y}/4} \|f\|_{\mathbb{H}^{-s}(\Omega)} \,.$$

Proof. See [39, Theorem 3.5]. □

2.4 Tensorial Discretization

Before discussing the discretization further, a general result regarding the tensor product of Hilbert spaces is given.

Theorem 2.3 (tensor products). *Let (M_i, μ_i), $i = 1, 2$, be measure spaces. Then we have the isomorphism*

$$L^2(M_1, \mu_1) \otimes L^2(M_2, \mu_2) \approx L^2(M_1 \times M_2, \mu_1 \otimes \mu_2) \,,$$

provided that all spaces are separable. In addition, if $\{\phi_k^{(i)}\}_{k\geq 1}$ is an orthonormal basis of $L^2(M_i, \mu_i)$, then $\{\phi_k^{(1)}\phi_m^{(2)}\}_{k,m\geq 1}$ is an orthonormal basis of their tensor product.

Proof. See [42, Theorem II.10] for the isomorphism, and [42, Proposition II.4.2] for the statement about bases. □

The following corollary motivates, as in [39, 8], the use of a tensorial discretization.

Corollary 2.4 (tensor products). *Let*

$$H^1_\mathcal{Y}(y^\alpha, (0, \mathcal{Y})) = \{w \in L^2(y^\alpha, (0, \mathcal{Y})) : w' \in L^2(y^\alpha, (0, \mathcal{Y}))\,,\, w(\mathcal{Y}) = 0\}\,.$$

Then we have

$$\mathring{H}^1_L(y^\alpha, \mathcal{C}_\mathcal{Y}) = H^1_0(\Omega) \otimes H^1_\mathcal{Y}(y^\alpha, (0, \mathcal{Y}))\,. \tag{2.5}$$

Proof. Let $\{\phi_j\}_{j\geq 1} \subset H^1_0(\Omega)$ and $\{\psi_k\}_{k\geq 1} \subset H^1_\mathcal{Y}(y^\alpha, (0, \mathcal{Y}))$ be orthonormal bases which, in addition, are orthogonal in $L^2(\Omega)$ and $L^2(y^\alpha, (0, \mathcal{Y}))$, respectively. Choose $\{\phi_j\}_{j\geq 1}$ to be the eigenfunctions of the Dirichlet Laplacian, as described in Section 2.1. An example of the family $\{\psi_k\}_{k\geq 1}$ will be given below in Theorem 3.1.

Clearly, $\phi_j\psi_k \in \mathring{H}^1_L(y^\alpha, \mathcal{C}_\mathcal{Y})$. Observe, in addition, that

$$\begin{aligned}
(\phi_j\psi_k, \phi_s\psi_t)_{\mathring{H}^1_L(y^\alpha, \mathcal{C}_\mathcal{Y})} &= \int_0^\mathcal{Y} \int_\Omega \big([\nabla\phi_j(x) \cdot \nabla\phi_s(x)]\psi_k(y)\psi_t(y) \\
&\qquad\qquad + \phi_j(x)\phi_s(x)\psi_k'(y)\psi_t'(y)\big)\, dx\, dy \\
&= \delta_{js} \int_0^\mathcal{Y} y^\alpha \psi_k(y)\psi_t(y)\, dy + \delta_{kt} \int_\Omega \phi_j(x)\phi_s(x)\, dx \\
&= C_{jkst}\delta_{js}\delta_{kt}\,.
\end{aligned}$$

Thus, the family $\{\phi_j\psi_k\}_{j,k\geq 1}$ is linearly independent. Using Gram-Schmidt, this linearly independent set contains an orthonormal subset which, to simplify the notation, we do not relabel.

Next, it will be shown that $\{\phi_j\psi_k\}_{j,k\geq 1}$ is complete in $\mathring{H}^1_L(y^\alpha, \mathcal{C}_\mathcal{Y})$. To see this, assume that $W \in \mathring{H}^1_L(y^\alpha, \mathcal{C}_\mathcal{Y})$ is such that

$$\int_0^\mathcal{Y} y^\alpha \int_\Omega \left([\nabla_x W(x,y) \cdot \nabla\phi_j(x)]\psi_k(y) + \partial_y W(x,y)\phi_j(x)\psi_k'(y)\right)\, dx\, dy = 0\,,$$

for all $j, k \in \mathbb{N}$. This implies that

$$\begin{gathered}
0 = \int_0^\mathcal{Y} y^\alpha \psi_k(y) \left(\int_\Omega \nabla_x W(x,y) \cdot \nabla_x \phi_j(x)\, dx\right) dy + \\
\int_\Omega \phi_j(x) \left(\int_0^\mathcal{Y} y^\alpha \partial_y W(x,y)\psi_k'(y)\, dy\right) dx\,.
\end{gathered}$$

But, because ϕ_j and ψ_k are linearly independent, this is only possible if

$$\int_\Omega \nabla_x W(x,y) \cdot \nabla_x \phi_j(x)\, dx = 0\,,$$

for all $j \in \mathbb{N}$ and a.e. $y \in (0, \mathcal{Y})$; and

$$\int_0^{\mathcal{Y}} y^\alpha \partial_y W(x,y) \psi_k'(y)\, dy = 0\,,$$

for all $k \in \mathbb{N}$ and a.e. $x \in \Omega$. Let now, for each $j \in \mathbb{N}$, $S_j \subset (0, \mathcal{Y})$ be the set where the first condition is false. Since S_j is null, i.e.,

$$\int_{S_j} y^\alpha\, dy = 0\,,$$

it holds that $S = \bigcup_{j=1}^\infty S_j$ is also a null set, as it is a countable union of null sets.

Similarly, for $k \in \mathbb{N}$, let $T_k \subset \Omega$ denote where the second condition fails. It again holds that

$$\int_{T_k} dx = 0\,,$$

and that $T = \bigcup_{k=1}^\infty T_k$ is another null set.

The previous observations show that

$$A(x,y) = |\nabla_x W(x,y)| + |\partial_y W(x,y)| = 0, \quad \forall (x,y) \in \mathcal{C}_{\mathcal{Y}} \setminus (T \times S).$$

However, $T \times S$ is a null set and therefore, $A(x,y) = 0$ almost everywhere in $\mathcal{C}_{\mathcal{Y}}$. As a consequence, $W = 0$ a.e. in $\mathcal{C}_{\mathcal{Y}}$.

In conclusion, it was shown that $\{\phi_j \psi_k\}_{j,k \geq 1}$ is a complete orthonormal basis in $\mathring{H}_L^1(y^\alpha, \mathcal{C}_{\mathcal{Y}})$.

This shows that the mapping defined by

$$\begin{aligned} \mathcal{W} : H_0^1(\Omega) \otimes H_{\mathcal{Y}}^1(y^\alpha, (0, \mathcal{Y})) &\to \mathring{H}_L^1(y^\alpha, \mathcal{C}_{\mathcal{Y}}) \\ \phi_j \otimes \psi_k &\mapsto \phi_j \psi_k\,, \end{aligned}$$

and extended by linearity, is the requisite isomorphism. $\square$

Having shown (2.5), an approximate solution of (2.4) is found via a Galerkin technique.

The space $H_0^1(\Omega)$ is approximated by a finite element space consisting of piecewise linear and continuous functions that vanish on the boundary [27], denoted $\mathbb{V}_h$. Usually this will be constructed on the basis of a quasi-uniform triangulation of Ω with size $h > 0$ so that

$$\dim \mathbb{V}_h \sim h^{-d} \, .$$

For the extended dimension, at this stage, a finite dimensional subspace $\mathcal{S}_\mathcal{Y}$ of $H_L^1(y^\alpha, (0, \mathcal{Y}))$ is introduced, but not further specified. Assume, for $\mathcal{K} \in \mathbb{N}$, that

$$\dim \mathcal{S}_\mathcal{Y} \sim \mathcal{K} \, .$$

Clearly, $\dim \mathbb{V}_h \otimes \mathcal{S}_\mathcal{Y} \sim \mathcal{K} h^{-d}$ and, as shown in (2.5),

$$\mathbb{V}_h \otimes \mathcal{S}_\mathcal{Y} \subset H_0^1(\Omega) \otimes H_\mathcal{Y}^1(y^\alpha, (0, \mathcal{Y})) = \mathring{H}_L^1(y^\alpha, \mathcal{C}_\mathcal{Y}) \, .$$

The Galerkin approximation of the solution to (2.4) is then denoted by $\mathcal{U}_{h,\mathcal{Y}}^\mathcal{K} \in \mathbb{V}_h \otimes \mathcal{S}_\mathcal{Y}$. From its definition, the best approximation result immediately follows:

$$\begin{aligned} \|\mathrm{tr}\ (\mathcal{U}_\mathcal{Y} - \mathcal{U}_{h,\mathcal{Y}}^\mathcal{K})\|_{\mathbb{H}^s(\Omega)} &= \|\nabla(\mathcal{U}_\mathcal{Y} - \mathcal{U}_{h,\mathcal{Y}}^\mathcal{K})\|_{L^2(y^\alpha, \mathcal{C}_\mathcal{Y})} \\ &= \inf_{W \in \mathbb{V}_h \otimes \mathcal{S}_\mathcal{Y}} \|\nabla(\mathcal{U}_\mathcal{Y} - W)\|_{L^2(y^\alpha, \mathcal{C}_\mathcal{Y})} \, . \end{aligned}$$

Some choices of $\mathcal{S}_\mathcal{Y}$ that have been used in the literature, and the corresponding error estimates, are detailed in Chapter 1.

Chapter 3

Diagonalization

3.1 Background

The diagonalization approach originally suggested in [8] is recounted here, as it forms the basis of the new approach described in this dissertation.

The tensorial discretization leads to a representation of the Galerkin solution to (2.4): For $(x, y) \in \mathcal{C}_{\mathcal{Y}}$,

$$\mathcal{U}_{h,\mathcal{Y}}^{\mathcal{K}}(x,y) = \sum_{k=1}^{\mathcal{K}} U_k(x) v_k(y) \ , \qquad u_{h,\mathcal{Y}}^{\mathcal{K}}(x) = \mathcal{U}_{h,\mathcal{Y}}^{\mathcal{K}}(x,0) = \sum_{k=1}^{\mathcal{K}} U_k(x) v_k(0) \, , \tag{3.1}$$

where $U_k \in \mathbb{V}_h$ and $v_k \in \mathcal{S}_{\mathcal{Y}}$. As a means to reduce computational complexity, a diagonalization technique is applied in [8], where $\mathcal{S}_{\mathcal{Y}}$ is an hp-FE space over a geometrically graded mesh. Then, the elements v_k in (3.1) are chosen as the solutions to the following discrete eigenvalue problem: find $(v, \mu) \in \mathcal{S}_{\mathcal{Y}} \setminus \{0\} \times \mathbb{R}$ such that

$$\int_0^{\mathcal{Y}} y^\alpha v'(y) w'(y) \, dy = \mu \int_0^{\mathcal{Y}} y^\alpha v(y) w(y) \, dy \ , \ \forall \, w \in \mathcal{S}_{\mathcal{Y}} \, . \tag{3.2}$$

A main drawback to this approach is that computing eigenvalues is inherently numerically unstable; see [47]. Some attempts at addressing this, by suitably choosing the parameters in the hp-FEM are discussed in the Section on Numerical Experiments and the Conclusions of [9]. This, however, is an essential difficulty. The eigenvalue problem (3.2) is very ill-conditioned.

Ignoring conditioning issues, let us explain what is the advantage of this diagonalization. Begin by observing that, as shown in [8], the solutions to (3.2) form an orthonormal basis of the FE space $\mathcal{S}_{\mathcal{Y}}$ which, in addition, satisfy

$$\int_0^{\mathcal{Y}} y^\alpha v_k'(y) v_i'(y) dy = \mu_k \delta_{ki} \,.$$

The solution representation in (3.1) may be written using as v_k the solutions to (3.2). Choosing, for the test function, $\zeta = V v_i$, where $V \in \mathbb{V}_h$, it is seen that

$$\begin{aligned}
&\int_\Omega \int_0^{\mathcal{Y}} y^\alpha \nabla \mathcal{U}_{h,\mathcal{Y}}^{\mathcal{K}}(x,y) \cdot \nabla \big(V(x) v_i(y)\big) \, dx \, dy \\
&= \int_\Omega \int_0^{\mathcal{Y}} y^\alpha \nabla \left(\sum_{k=1}^{\mathcal{K}} U_k(x) v_k(y) \right) \cdot \nabla \big(V(x) v_i(y)\big) \, dx \, dy \\
&= \int_\Omega \int_0^{\mathcal{Y}} y^\alpha \left(\sum_{k=1}^{\mathcal{K}} \left[\nabla U_k(x) v_k(y), \; U_k(x) v_k'(y) \right] \right) \cdot \left[\nabla V(x) v_i(y), \; V(x) v_i'(y) \right] dx \, dy \\
&= \int_\Omega \int_0^{\mathcal{Y}} y^\alpha \sum_{k=1}^{\mathcal{K}} \left[\nabla U_k(x) \cdot \nabla V(x) v_k(y) v_i(y) + U_k(x) V(x) v_k'(y) v_i'(y) \right] dx \, dy \\
&= \sum_{k=1}^{\mathcal{K}} \Bigg[\int_\Omega \nabla U_k(x) \cdot \nabla V(x) \int_0^{\mathcal{Y}} y^\alpha v_k(y) v_i(y) \, dy \, dx \\
&\qquad\qquad + U_k(x) V(x) \int_0^{\mathcal{Y}} y^\alpha v_k'(y) v_i'(y) \, dy \, dx \Bigg] \\
&= \int_\Omega \nabla U_k(x) \cdot \nabla V(x) \, dx + \mu_k \int_\Omega U_k(x) V(x) \, dx.
\end{aligned}$$

This, for $k = 1, \ldots, \mathcal{K}$, implies the functions $U_k \in \mathbb{V}_h$ solve

$$\int_\Omega \nabla U_k(x) \cdot \nabla V(x) \, dx + \mu_k \int_\Omega U_k(x) V(x) \, dx = d_s v_k(0) \langle f, V \rangle \,. \tag{3.3}$$

The efficiency of the diagonalization technique follows from this decoupling since each of the $\mathcal{K}$ problems now only involve solving over Ω and each problem is independent of each other, meaning they may be solved in parallel.

3.2 A Semi-Analytic Approach

To overcome the challenge in numerically solving the eigenvalue problem (3.2), an exact diagonalization technique is proposed. In other words, the eigenvalue problem in (3.2) will be solved exactly. Start by writing the following problem: find pairs (μ, ψ), with $\psi \neq 0$, such that:

$$\begin{cases} -(y^\alpha \psi'(y))' = \mu y^\alpha \psi(y), & \text{in } (0, \mathcal{Y}), \\ -y^\alpha \psi'(y) \to 0, & y \downarrow 0, \\ \psi(\mathcal{Y}) = 0. \end{cases} \tag{3.4}$$

Theorem 3.1 (eigenpairs)**.** *There is a countable number of eigenpairs* $\{(\mu_k, \psi_k)\}_{k=1}^{\infty} \subset \mathbb{R}_+ \times H^1_{\mathcal{Y}}(y^\alpha, (0, \mathcal{Y}))$ *that are solutions to* (3.4)*. The eigenvalues may be numbered so that*

$$0 < \mu_1 < \mu_2 \leq \cdots \leq \mu_n \leq \cdots, \qquad \mu_n \to \infty, \ n \to \infty.$$

Moreover, the normalized eigenfunctions $\{\psi_k\}_{k>\infty}$ *form an orthonormal basis of the space* $L^2(y^\alpha, (0, \mathcal{Y}))$*. As a consequence,*

$$\int_0^{\mathcal{Y}} y^\alpha \psi_j'(y) \psi_k'(y)\, dy = \mu_j \delta_{jk}. \tag{3.5}$$

Finally, $\{\psi_k\}_{k\geq 1}$ *is an orthogonal basis of* $H^1_{\mathcal{Y}}(y^\alpha, (0, \mathcal{Y}))$*.*

Proof. Notice that this ODE with the given boundary conditions is a singular Sturm-Liouville problem. Positivity of the first eigenvalue μ_1 is a result of minimizing the Rayleigh quotient over trial functions subject to the boundary conditions in (3.4). A proof of this can be found in [44, Theorem 1, Section 11.1] and requires only accounting for the boundary conditions. The remaining eigenvalues can be found in an analogous manner to [44, Theorem 2, Section 11.1], which establishes the orthogonality of the eigenfunctions with respect to y^α as a constraint in minimizing the Rayleigh quotient. The eigenfunctions can be seen to be complete in the weighted L^2 sense as in [44, Theorem 2, Section 11.3]. Equation (3.5) can be seen by integrating the ODE against another eigenfunction

$$\int_0^{\mathcal{Y}} y^\alpha \psi_j'(y)\psi_k'(y)\,dy = \int_0^{\mathcal{Y}} -(y^\alpha \psi_j'(y))'\psi_k(y)\,dy = \mu_j \int_0^{\mathcal{Y}} y^\alpha \psi_j(y)\psi_k(y)\,dy = \mu_j \delta_{jk}\,,$$

showing that $\{\psi_k\}_{k\geq 1}$ is an orthogonal basis of $H^1_{\mathcal{Y}}(y^\alpha,(0,\mathcal{Y}))$. $\square$

The goal now is to find the exact solutions to this problem. Expanding the ordinary differential equation, it is seen that

$$\psi''(y) + \frac{\alpha}{y}\psi'(y) + \mu\psi(y) = 0\ ,\ y \in (0,\mathcal{Y})\,. \tag{3.6}$$

The solutions to this second order linear ordinary differential equation behave differently depending on the value of s.

Lemma 3.2 (eigenpairs for $s = \frac{1}{2}$)**.** *Let $s = \frac{1}{2}$. The eigenpairs $\{(\mu_k,\psi_k)\}_{k=1}^\infty$ that solve (3.4) are given by*

$$\psi_k(y) = \sqrt{\frac{2}{\mathcal{Y}}}\cos(\mu_k y)\ ,\ k \in \mathbb{N}\,,$$

and

$$\mu_k = \left(\frac{(k-\frac{1}{2})\pi}{\mathcal{Y}}\right)^2\,. \tag{3.7}$$

Proof. Since $\alpha = 0$, equation (3.6) becomes a linear second order ODE with constant coefficients. Its solution is carried out by elementary means. $\square$

On the other hand, the case of $s \neq \frac{1}{2}$ requires some work.

Lemma 3.3 (eigenpairs for $s \neq \frac{1}{2}$)**.** *Let $s \neq \frac{1}{2}$. The eigenpairs $\{(\mu_k,\psi_k)\}_{k=1}^\infty$ that solve (3.4) are given by*

$$\psi_k(y) = \left(\frac{\sqrt{2}}{\mu_k^{s/2}\mathcal{Y}J_{1-s}(\eta_k)}\right)(y\sqrt{\mu_k})^s J_{-s}(y\sqrt{\mu_k})\ ,\ k \in \mathbb{N}\,,$$

where η_k is the k^{th} positive root of J_{-s}. The eigenvalues are given by

$$\mu_k = \left(\frac{\eta_k}{\mathcal{Y}}\right)^2\,. \tag{3.8}$$

Proof. Begin by proposing the solution ansatz $\psi(y) = z^s w(z)$, where $z = \sqrt{\mu}y$. Using this the ODE becomes

$$z^s w''(z) + zw'(z) + (z^2 - s^2)w(z) = 0\,.$$

This is a Bessel equation [1, Section 9.1]. According to [DLMF, Section 10.2(ii)], its general solution is given by

$$w(z) = AJ_s(z) + BY_s(z)\,,$$

where A and B are arbitrary constants; J_s and Y_s are, respectively, the Bessel functions of the first and second kind.

The objective now is to satisfy the boundary conditions. Substituting back that $\sqrt{\mu}y = z$, the Neumann boundary condition at $y = 0$ becomes

$$\begin{aligned}
-\lim_{y\downarrow 0} y^\alpha \psi'(y) &= -\lim_{y\downarrow 0} y^\alpha \frac{d}{dy}\left[(\sqrt{\mu}y)^s\big(AJ_s(\sqrt{\mu}y) + BY_s(\sqrt{\mu}y)\big)\right] \\
&= \frac{2^{1-s}\mu^s\big(A\pi + B\cos(\pi s)\Gamma(1-s)\Gamma(s)\big)}{\pi\Gamma(s)} \\
&= 0\,,
\end{aligned}$$

i.e.,

$$A\pi + B\cos(\pi s)\Gamma(1-s)\Gamma(s) = 0\,.$$

On the other hand, the Dirichlet boundary condition implies that

$$AJ_s(\sqrt{\mu}\mathcal{Y}) + BY_s(\sqrt{\mu}\mathcal{Y}) = 0\,.$$

Expressing B in terms of A, it can be shown that

$$\psi(y) = \frac{2^s A}{\Gamma(1-s)\cos(\pi s)}\,{}_0F_1(0, 1-s; -\mu y^2/4)\,,$$

where ${}_0F_1$ is the hypergeometric function. This exotic function can be expressed in terms of the Bessel function of the first kind by the relation [1, Equation (9.1.69)]

$$J_{-s}(\sqrt{\mu}y) = \frac{\left(\frac{\sqrt{\mu}y}{2}\right)^{-s}}{\Gamma(1-s)} {}_0F_1(0, 1-s; -\mu y^2/4)\,.$$

So, it may be written that

$$\psi(y) = \frac{A}{\cos(\pi s)} \left(\sqrt{\mu}y\right)^s J_{-s}(\sqrt{\mu}y)\,.$$

The coefficient A is determined so that $\|\psi\|^2_{L^2(y^\alpha,(0,\mathcal{Y}))} = 1$. This shows the result. □

According to (3.1), the eigenfunctions only need to be evaluated at $y = 0$.

Corollary 3.4 (value at $y = 0$). *If* $\{(\mu_k, \psi_k)\}_{k=1}^{\infty}$ *solve* (3.4)*, then*

$$\psi_k(0) = \sqrt{\frac{2}{\mathcal{Y}}}\,, \qquad\qquad s = \frac{1}{2}\,, \tag{3.9}$$

and

$$\psi_k(0) = \frac{2^{s+1/2}}{\mu_k^{s/2}\mathcal{Y}J_{1-s}(\eta_k)\Gamma(1-s)}\,, \qquad\qquad s \neq \frac{1}{2}\,. \tag{3.10}$$

Proof. The case $s = \frac{1}{2}$ is immediate.

On the other hand, if $s \neq \frac{1}{2}$, the asymptotic form for Bessel functions of the first kind [1, Equation (9.1.7)] is used. Namely,

$$J_{-s}(z) \sim \frac{1}{\Gamma(1-s)} \left(\frac{2}{z}\right)^s\,.$$

Using this and taking the limit $y \downarrow 0$ yields the desired result. □

The approach of the method can be summarized as follows. Given $f \in \mathbb{H}^{-s}(\Omega)$, a finite element space $\mathbb{V}_h$ is chosen. Then solve, for $k = 1, \ldots, \mathcal{K}$, problem (3.3) where the constant d_s is defined in (1.4), $v_k(0) = \psi_k(0)$, and the pair $(\mu_k, \psi_k(0))$ is given by either (3.7) and

(3.9) or by (3.8) and (3.10). Finally, the approximate solution is

$$u_{h,\mathcal{Y}}^{\mathcal{K}} = \sum_{k=1}^{\mathcal{K}} \psi_k(0) U_k \in \mathbb{V}_h \, .$$

The main computational appeal of the diagonalization developed in [8] remains. Namely, problems (3.3) can be solved in parallel. In addition, the values of μ_k and $\psi_k(0)$, which is all is needed to solve these problems and form the solution, are given explicitly. No unstable eigenvalue problem needs to be solved.

Chapter 4

Error Analysis

Our goal now shall be to derive an error estimate for this new semi-analytic diagonalization scheme. In order to achieve this, in what follows we shall assume that $f \in L^2(\Omega)$.

4.1 The discrete Laplacian

We begin by introducing the so-called discrete Laplacian $\Delta_h : \mathbb{V}_h \to \mathbb{V}_h$, which is defined as

$$\int_\Omega \Delta_h W V dx = -\int_\Omega \nabla W \cdot \nabla V dx, \quad \forall V, W \in \mathbb{V}_h .$$

This is clearly an invertible, symmetric, and negative definite operator. Let $M = \dim \mathbb{V}_h$. We have:

1. There exist a finite collection of eigenpairs, $\{(\lambda_{h,m}, \Phi_{h,m})\}_{m=1}^M \subset \mathbb{R}_+ \times \mathbb{V}_h$, such that,

$$\int_\Omega (-\Delta_h)\Phi_{h,m} V dx = \int_\Omega \nabla\Phi_{h,m} \cdot \nabla V dx = \lambda_{h,m} \int_\Omega \Phi_{h,m} V dx, \quad \forall V \in \mathbb{V}_h .$$

2. The eigenvectors are a basis of $\mathbb{V}_h$, and can be chosen to be orthonormal in $L^2(\Omega)$.

3. The eigenvalues satisfy

$$\lambda_{h,1} < \lambda_{h,2} \le \cdots \le \lambda_{h,M} .$$

4. More importantly, the eigenvalues satisfy the bounds

$$\lambda_{h,1} \geq \frac{1}{C_P}\,, \qquad\qquad \lambda_{h,M} \lesssim h^{-2}\,, \tag{4.1}$$

where C_P is the best constant in the Poincaré inequality in Ω. The first inequality in (4.1) can be shown directly from an application of the Rayleigh quotient. For the second inequality, the result may be found in [45, Estimate (3.28)] with $\beta = 2$ and setting $\chi = \Phi_{h,M}$, where $\Phi_{h,M}$ is the eigenfunction corresponding to $\lambda_{h,M}$.

5. Fractional powers of the discrete Laplacian, denoted $(-\Delta_h)^r$, for $r \in \mathbb{R}$, are given by standard spectral theory.

From the previous points two crucial estimates follow. First, if P_h is the $L^2(\Omega)$–projection onto $\mathbb{V}_h$,

$$\|(-\Delta_h)P_h\|_{L^2(\Omega)\to L^2(\Omega)} \lesssim \frac{1}{h^2}\,. \tag{4.2}$$

Next, since from the onset we have assumed that our domain is a convex polytope, we have that if $w \in H^1_0(\Omega)$ is such that $-\Delta w = g \in L^2(\Omega)$, then $w \in H^2(\Omega) \cap H^1_0(\Omega)$. Therefore, a standard application of Aubin-Nitsche duality [27, Lemma 2.31] says that,

$$\|w - w_h\|_{L^2(\Omega)} \lesssim h^2 |w|_{H^2(\Omega)} \lesssim h^2 \|g\|_{L^2(\Omega)}\,,$$

where $w_h = (-\Delta_h)^{-1} P_h g$ is the Galerkin approximation to w. In operator terms this reads

$$\|(-\Delta)^{-1} - (-\Delta_h)^{-1} P_h\|_{L^2(\Omega)\to L^2(\Omega)} \lesssim h^2. \tag{4.3}$$

4.2 Error decomposition

Having introduced the discrete Laplacian, we observe that the solution to (3.3) can be written as

$$U_k = d_s \psi_k(0)(\mu_k \mathrm{I} - \Delta_h)^{-1} P_h f\,,$$

where I is the identity operator and we recall that P_h denotes the $L^2(\Omega)$–projection onto $\mathbb{V}_h$. We may then rewrite our approximate solution as

$$u_{h,\mathcal{Y}}^{\mathcal{K}} = d_s \sum_{k=1}^{\mathcal{K}} |\psi_k(0)|^2 (\mu_k \mathrm{I} - \Delta_h)^{-1} P_h f \,. \tag{4.4}$$

With this at hand we begin the error analysis by writing

$$\|u - u_{h,\mathcal{Y}}^{\mathcal{K}}\|_{L^2(\Omega)} \leq \|u - u_h\|_{L^2(\Omega)} + \|u_h - u_{h,\mathcal{Y}}^{\infty}\|_{L^2(\Omega)} + \|u_{h,\mathcal{Y}}^{\infty} - u_{h,\mathcal{Y}}^{\mathcal{K}}\|_{L^2(\Omega)} \,. \tag{4.5}$$

Here $u_h = (-\Delta_h)^{-s} P_h f$ and, as the notation suggests,

$$u_{h,\mathcal{Y}}^{\infty} = d_s \sum_{k=1}^{\infty} |\psi_k(0)|^2 (\mu_k \mathrm{I} - \Delta_h)^{-1} P_h f \,.$$

For reasons that shall become clear during the course of our discussion, we shall call the first term on the right hand side of (4.5) the discretization error, the second one the quadrature error, whereas the last one we name the truncation error. Each of the three contributions of the error are studied below.

Discretization error

To control the discretization error, we observe that (4.2) and (4.3) are precisely the conditions of [37, Theorem 1], and so we conclude that the discretization error is controlled by

$$\|u - u_h\|_{L^2(\Omega)} \lesssim h^{2s} \|f\|_{L^2(\Omega)} \,. \tag{4.6}$$

Quadrature error: The Balakrishnan formula

To study the quadrature error, we recall the Balakrishnan formula for powers of positive operators; see [36, Chapter 4] for the general theory, and [14] and [12] for its application to the fractional Laplacian. As stated above, the operator $-\Delta_h$ is positive definite and symmetric. Therefore,

$$(-\Delta_h)^{-s} = \frac{2\sin(\pi s)}{\pi} \int_0^\infty t^\alpha (t^2 \mathrm{I} - \Delta_h)^{-1} \, dt \,. \tag{4.7}$$

We interpret (4.4) as a quadrature formula for the operator-valued weighted integral

$$\int_0^\infty t^\alpha F_h(t) \, dt \,, \qquad F_h(t) = \frac{2\sin(\pi s)}{\pi} (t^2 \mathrm{I} - \Delta_h)^{-1} \,.$$

The nodes $\{t_k^{(s)}\}_{k=1}^{\mathcal{K}}$ and weights $\{\omega_k^{(s)}\}_{k=1}^{\mathcal{K}}$ of the quadrature scheme are dependent on the value of s. The specific values for the two cases, $s = \frac{1}{2}$ and $s \neq \frac{1}{2}$, will be given in Sections 4.2.1 and 4.2.2, respectively.

It is now clear why we name the terms on the right hand side as such. The difference $u_h - u_{h,\mathcal{Y}}^\infty$ encodes the error in approximating the integral in (4.7) with a quadrature formula with an infinite number of nodes and weights. Finally, $u_{h,\mathcal{Y}}^\infty - u_{h,\mathcal{Y}}^{\mathcal{K}}$ measures the effect of truncating the quadrature rule and only accounting for $\mathcal{K}$ nodes.

In order to concisely state our results let us introduce some notation. Given $\lambda > 0$, we define the function

$$F_\lambda(t) = \frac{2\sin(\pi s)}{\pi} \frac{1}{t^2 + \lambda} \,.$$

The integral and quadrature formulas of interest are then denoted by

$$I_s(F_\lambda) = \int_0^\infty t^{1-2s} F_\lambda(t) \, dt \,,$$

and

$$Q_s^{\mathcal{K}}(F_\lambda) = \sum_{k=1}^{\mathcal{K}} (t_k^{(s)})^{1-2s} \omega_k^{(s)} F_\lambda(t_k^{(s)}) \,, \tag{4.8}$$

respectively. The symbol $Q_s^\infty(F_\lambda)$ has then the expected meaning.

The main error estimate of our work is the content of the next result.

Theorem 4.1 (error estimate)**.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded, convex polytope and $f \in L^2(\Omega)$. Let, for $s \in (0,1)$, $u \in \mathbb{H}^s(\Omega)$ be the solution to (1.1) and $u_{h,\mathcal{Y}}^{\mathcal{K}} \in \mathbb{V}_h$ be as defined in (4.4).*

Then, we have

$$\|u - u_{h,\mathcal{Y}}^{\mathcal{K}}\|_{L^2(\Omega)} \lesssim \Bigg(h^{2s} + \sup_{\lambda \geq \frac{1}{C_P}} |I_s(F_\lambda) - Q_s^\infty(F_\lambda)| \\ + \sup_{\lambda \geq \frac{1}{C_P}} \left| \sum_{k > \mathcal{K}} (t_k^{(s)})^{1-2s} \omega_k^{(s)} F_\lambda(t_k^{(s)}) \right| \Bigg) \|f\|_{L^2(\Omega)} ,$$

where the implicit constant is independent of $\mathcal{Y}$, $\mathcal{K}$, h, and f.

Proof. Owing to (4.6) it suffices to study the quadrature and truncation errors.

Let, for the sake of definiteness, $M = \dim \mathbb{V}_h$ and $\{\Phi_{h,m}\}_{m=1}^M \subset \mathbb{V}_h$ be as in Section 4.1. Then,

$$P_h f = \sum_{m=1}^{M} f_m \Phi_{h,m} , \qquad f_m = \int_\Omega f \Phi_{h,m} \, dx ,$$

and

$$u_h = (-\Delta_h)^s P_h f = \int_0^\infty t^\alpha F_h(t) P_h f \, dt = \sum_{m=1}^{M} f_m I_s(F_{\lambda_{h,m}}) \Phi_{h,m} .$$

Similarly, (4.4) can be rewritten as

$$u_{h,\mathcal{Y}}^{\mathcal{K}} = \sum_{m=1}^{M} f_m Q_s^{\mathcal{K}}(F_{\lambda_{h,m}}) \Phi_{h,m} .$$

Since $\{\Phi_{h,m}\}_{m=1}^M$ is orthonormal in $L^2(\Omega)$,

$$\|u_h - u_{h,\mathcal{Y}}^{\mathcal{K}}\|_{L^2(\Omega)}^2 = \sum_{m=1}^{M} |f_m|^2 \left| I_s(F_{\lambda_{h,m}}) - Q_s^{\mathcal{K}}(F_{\lambda_{h,m}}) \right|^2 \\ \leq \sup_{\lambda \geq \frac{1}{C_P}} \left| I_s(F_\lambda) - Q_s^{\mathcal{K}}(F_\lambda) \right|^2 \|f\|_{L^2(\Omega)}^2 ,$$

where we used (4.1). In summary we have obtained

$$\|u_h - u_{h,\mathcal{Y}}^{\mathcal{K}}\|_{L^2(\Omega)} \leq \sup_{\lambda \geq \frac{1}{C_P}} \left| I_s(F_\lambda) - Q_s^{\mathcal{K}}(F_\lambda) \right| \|f\|_{L^2(\Omega)} .$$

Since, clearly,

$$\left|I_s(F_\lambda) - Q_s^{\mathcal{K}}(F_\lambda)\right| \leq |I_s(F_\lambda) - Q_s^\infty(F_\lambda)| + \left|\sum_{k>\mathcal{K}} (t^{(s)})^{1-2s}\omega_k^{(s)} F_\lambda(t_k^{(s)})\right|,$$

the result follows. □

The usefulness of Theorem 4.1 is clear. To obtain an error estimate it suffices to, depending on the value of s, estimate the two suprema. We now embark in this task.

4.2.1 Error Analysis for $s = \frac{1}{2}$

In the case $s = \frac{1}{2}$, we have that $d_s = 1$. Substitute this, (3.7), and (3.9) into (4.4) to conclude that in this case the quadrature formula $Q_{\frac{1}{2}}^{\mathcal{K}}$, defined in (4.8), has the following nodes and weights:

$$t_k^{(\frac{1}{2})} = \frac{(k-\frac{1}{2})\pi}{\mathcal{Y}}, \qquad \omega_k^{(\frac{1}{2})} = \frac{\pi}{\mathcal{Y}}. \tag{4.9}$$

In addition

$$I_{\frac{1}{2}}(F_\lambda) = \int_0^\infty F_\lambda(t)\,dt = \frac{2}{\pi}\int_0^\infty \frac{1}{t^2+\lambda}\,dt = \frac{1}{\sqrt{\lambda}}, \tag{4.10}$$

as expected.

To aid in the error estimate, two lemmas are now presented.

Lemma 4.2 (quadrature error)**.** *Let $s = \frac{1}{2}$ and $\lambda > 0$. If the weights and nodes of (4.8) are specified by (4.9), then we have*

$$\left|I_{\frac{1}{2}}(F_\lambda) - Q_{\frac{1}{2}}^\infty(F_\lambda)\right| \lesssim \frac{\exp(-\sqrt{\lambda}\mathcal{Y})}{\sqrt{\lambda}}. \tag{4.11}$$

Proof. Set $a = \sqrt{\lambda}\mathcal{Y}/\pi$. Using the partial fraction expansion [41, Equation (30:6:5)], write

$$\begin{aligned} Q_{\frac{1}{2}}^\infty(F_\lambda) &= \frac{\pi}{\mathcal{Y}}\sum_{k=1}^\infty \frac{2}{\pi}\frac{1}{\frac{(k-\frac{1}{2})^2\pi^2}{\mathcal{Y}^2}+\lambda} = \frac{2}{\mathcal{Y}}\frac{\mathcal{Y}^2}{\pi^2}\sum_{k=1}^\infty \frac{1}{(k-\frac{1}{2})^2+\lambda\frac{\mathcal{Y}^2}{\pi^2}} \\ &= \frac{2\mathcal{Y}}{\pi^2}\frac{\sqrt{\lambda}}{\sqrt{\lambda}}\sum_{k=1}^\infty \frac{1}{(k-\frac{1}{2})^2+a^2} = \frac{2a}{\pi\sqrt{\lambda}}\sum_{k=1}^\infty \frac{1}{(k-\frac{1}{2})^2+a^2} \end{aligned}$$

$$= \frac{8a}{\pi\sqrt{\lambda}} \sum_{k=1}^{\infty} \frac{1}{(2k-1)^2 + (2a)^2} = \frac{8a}{\pi\sqrt{\lambda}} \frac{\pi \tanh(a\pi)}{8a} = \frac{1}{\sqrt{\lambda}} \tanh\left(\sqrt{\lambda}\mathcal{Y}\right).$$

Combine this with (4.10) to obtain

$$\begin{aligned}
\left| I_{\frac{1}{2}}(F_\lambda) - Q_{\frac{1}{2}}^{\infty}(F_\lambda) \right| &= \frac{1}{\sqrt{\lambda}}(1 - \tanh(\sqrt{\lambda}\mathcal{Y})) \\
&= \frac{1}{\sqrt{\lambda}} \left(1 - \frac{1 - \exp(-2\sqrt{\lambda}\mathcal{Y})}{1 + \exp(-2\sqrt{\lambda}\mathcal{Y})} \right) \\
&= \frac{1}{\sqrt{\lambda}} \left(\frac{1 + \exp(-2\sqrt{\lambda}\mathcal{Y}) - 1 + \exp(-2\sqrt{\lambda}\mathcal{Y})}{1 + \exp(-2\sqrt{\lambda}\mathcal{Y})} \right) \\
&= \frac{1}{\sqrt{\lambda}} \left(\frac{2\exp(-2\sqrt{\lambda}\mathcal{Y})}{1 + \exp(-2\sqrt{\lambda}\mathcal{Y})} \right) \lesssim \frac{\exp(-\sqrt{\lambda}\mathcal{Y})}{\sqrt{\lambda}},
\end{aligned}$$

as claimed. □

Lemma 4.3 (truncation error)**.** *Let $s = \frac{1}{2}$, $\lambda > 0$, and $\mathcal{K} \in \mathbb{N}$. Assume that the weights and nodes of $Q_{\frac{1}{2}}^{\mathcal{K}}$ are given in (4.9). If $\mathcal{K}$ is sufficiently large, then we have*

$$\left| Q_{\frac{1}{2}}^{\infty}(F_\lambda) - Q_{\frac{1}{2}}^{\mathcal{K}}(F_\lambda) \right| \lesssim \frac{\mathcal{Y}}{\mathcal{K}}.$$

Proof. Recall that from [10, Section 1.9, formula (10)]

$$\psi'(\mathcal{K}) = \sum_{k \geq \mathcal{K}} \frac{1}{k^2},$$

where here ψ is the digamma function. Using the asymptotic bound for ψ for large arguments [1, Formula 6.4.12], the first derivative of the digamma function behaves as

$$\psi'(x) \sim \frac{1}{x} + \frac{1}{2x^2} + \frac{1}{6x^3} - \frac{1}{30x^5} + \frac{1}{42x^7} - \cdots \lesssim \frac{1}{x}.$$

Combining these two observations, it holds that

$$\left| Q_{\frac{1}{2}}^{\infty}(F_\lambda) - Q_{\frac{1}{2}}^{\mathcal{K}}(F_\lambda) \right| = \frac{2}{\mathcal{Y}} \sum_{k > \mathcal{K}} \frac{1}{\frac{(k-\frac{1}{2})^2\pi^2}{\mathcal{Y}^2} + \lambda} = \frac{2\mathcal{Y}}{\pi^2} \sum_{k > \mathcal{K}} \frac{1}{(k-\frac{1}{2})^2 + \lambda\frac{\mathcal{Y}^2}{\pi^2}}$$

$$\leq \frac{2\mathcal{Y}}{\pi^2}\sum_{k\geq\mathcal{K}}\frac{1}{(k+\frac{1}{2})^2} \leq \frac{2\mathcal{Y}}{\pi^2}\sum_{k\geq\mathcal{K}}\frac{1}{k^2} = \frac{2\mathcal{Y}}{\pi^2}\psi'(\mathcal{K}) \lesssim \frac{\mathcal{Y}}{\mathcal{K}}\,,$$

which is the bound that was needed to be obtained. □

Using these lemmas, a corollary to Theorem 4.1 shows the error estimate for the case of $s = \frac{1}{2}$.

Corollary 4.4 (error estimate for $s = \frac{1}{2}$)**.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded, convex polytope and $f \in L^2(\Omega)$. Let $u \in \mathbb{H}^{1/2}(\Omega)$ be the solution to (1.1) for $s = 1/2$. Let $u^{\mathcal{K}}_{h,\mathcal{Y}}$ be given by (4.4), where $\mathbb{V}_h$ is constructed using a quasi-uniform mesh of size h. Then, we have that*

$$\|u - u^{\mathcal{K}}_{h,\mathcal{Y}}\|_{L^2(\Omega)} \lesssim \left(h + \sqrt{C_P}\exp\left(-\frac{\mathcal{Y}}{\sqrt{C_P}}\right) + \frac{\mathcal{Y}}{\mathcal{K}}\right)\|f\|_{L^2(\Omega)}\,,$$

where C_P is given in (4.1). In particular, if the truncation parameter is chosen as $\mathcal{Y} \sim |\log h|$, and the number of quadrature points is chosen as $\mathcal{K} \sim \frac{|\log h|}{h}$, we may conclude that

$$\|u - u^{\mathcal{K}}_{h,\mathcal{Y}}\|_{L^2(\Omega)} \lesssim h\|f\|_{L^2(\Omega)}\,.$$

Proof. The proof of the error estimate essentially entails using Lemma 4.2 and 4.3 to bound the estimate in Theorem 4.1 and using (4.1).

The practical error estimate, that is the one in terms of h, is given by using the prescribed relations between all discretization parameters. □

4.2.2 Error Analysis for $s \neq \frac{1}{2}$

First use (4.4) to realize that this solution representation is indeed a quadrature formula for the Balakrishnan formula (4.7).

Lemma 4.5 (quadrature)**.** *Let $s \in (0,1)\setminus\{\frac{1}{2}\}$. Then, the solution representation (4.4) can be equivalently written as*

$$u^{\mathcal{K}}_{h,\mathcal{Y}} = \sum_{m=1}^{M} f_m Q^{\mathcal{K}}_s(F_{\lambda_{h,m}})\Phi_{h,m}\,,$$

where the quadrature formula $Q_s^{\mathcal{K}}$ is defined in (4.8). Its nodes and weights are given by

$$t_k^{(s)} = \sqrt{\mu_k} = \frac{\eta_k}{\mathcal{Y}} \, , \qquad \omega_k^{(s)} = \frac{2}{\mathcal{Y}\eta_k J_{1-s}^2(\eta_k)} \, . \tag{4.12}$$

Proof. Start by substituting (1.4) and (3.10) into (4.4). Using the famous Euler's reflection formula [1, Formula 6.1.17] given by

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin(\pi s)} \, ,$$

it holds that

$$\begin{aligned} u_{h,\mathcal{Y}}^{\mathcal{K}} &= d_s \sum_{k=1}^{\mathcal{K}} |\psi_k(0)|^2 (\mu_k \mathrm{I} - \Delta_h)^{-1} P_h f \\ &= \frac{2^{1-2s}\Gamma(1-s)}{\Gamma(s)} \sum_{k=1}^{\mathcal{K}} \frac{2^{1+2s}}{\mu_k^s \mathcal{Y}^2 \Gamma^2(1-s) J_{1-s}^2(\eta_k)} (\mu_k \mathrm{I} - \Delta_h)^{-1} P_h f \\ &= \frac{2\sin(\pi s)}{\pi} \sum_{k=1}^{\mathcal{K}} \frac{2}{\mu_k^s \mathcal{Y}^2 J_{1-s}^2(\eta_k)} (\mu_k \mathrm{I} - \Delta_h)^{-1} P_h f \, . \end{aligned}$$

Substitute (3.8) into this result to obtain that

$$\begin{aligned} u_{h,\mathcal{Y}}^{\mathcal{K}} &= \sum_{k=1}^{\mathcal{K}} \left(\frac{\eta_k}{\mathcal{Y}}\right)^{1-2s} \frac{2}{\eta_k \mathcal{Y} J_{1-s}^2(\eta_k)} \left[\frac{2\sin(\pi s)}{\pi} \left(\left(\frac{\eta_k}{\mathcal{Y}}\right)^2 \mathrm{I} - \Delta_h \right)^{-1} \right] P_h f \\ &= \sum_{k=1}^{\mathcal{K}} \left(\frac{\eta_k}{\mathcal{Y}}\right)^{1-2s} \frac{2}{\eta_k \mathcal{Y} J_{1-s}^2(\eta_k)} \left[\frac{2\sin(\pi s)}{\pi} \sum_{m=1}^{M} f_m \frac{1}{\left(\frac{\eta_k}{\mathcal{Y}}\right)^2 + \lambda_{h,m}} \right] \Phi_{h,m} \\ &= \sum_{m=1}^{M} f_m \sum_{k=1}^{\mathcal{K}} \left(\frac{\eta_k}{\mathcal{Y}}\right)^{1-2s} \frac{2}{\eta_k \mathcal{Y} J_{1-s}^2(\eta_k)} F_{\lambda_{h,m}} \left(\frac{\eta_k}{\mathcal{Y}}\right) \Phi_{h,m} \\ &= \sum_{m=1}^{M} f_m \sum_{k=1}^{\mathcal{K}} \left(t_k^{(s)}\right)^{1-2s} \omega_k^{(s)} F_{\lambda_{h,m}}(t_k^{(s)}) \Phi_{h,m} = \sum_{m=1}^{M} f_m Q_s^{\mathcal{K}}(F_{\lambda_{h,m}}) \Phi_{h,m} \, , \end{aligned}$$

where, as claimed, the nodes and weights are given by (4.12). □

Having identified the nodes and weights of the quadrature rule (4.8), the formula is now related to the one studied in [40, Equations 1.1 and 1.2]. In this work, for $\nu > -1$, the following quadrature formula is studied

$$\int_{-\infty}^{\infty} |t|^{1+2\nu} G(t)\, dt \approx \tau \sum_{k\in\mathbb{Z}\setminus\{0\}} w_{\nu,k} |\tau\xi_{\nu,k}|^{1+2\nu} G(\tau\xi_{\nu,k})\,,$$

where $\tau > 0$ is a given parameter, the weights are given by

$$w_{\nu,k} = \frac{2}{\pi^2 \xi_{\nu,|k|} J_{\nu+1}^2(\pi\xi_{\nu,|k|})}\,, \quad k \in \mathbb{Z}\setminus\{0\}\,,$$

and, for $k \in \mathbb{N}$, $\xi_{\nu,k}$ are the zeros of $J_\nu(\pi x)$ ordered in such a way that

$$0 < \xi_{\nu,1} < \xi_{\nu,2} < \cdots\,.$$

Finally, $\xi_{\nu,-k} = -\xi_{\nu,k}$, $k \in \mathbb{N}$.

Let now G be an even function, then it is immediate to see that, with the same nodes and weights,

$$\int_0^{\infty} t^{1+2\nu} G(t) dt \approx \tau \sum_{k=1}^{\infty} w_{\nu,k} |\tau\xi_{\nu,k}|^{1+2\nu} G(\tau\xi_{\nu,k})\,. \tag{4.13}$$

It is now shown how this quadrature form is related to $Q_s^\infty(F_\lambda)$.

Lemma 4.6 (quadrature equivalence)**.** *Let $s \in (0,1)\setminus\{\frac{1}{2}\}$. Let the nodes and weights of the quadrature scheme $Q_s^\infty(F_\lambda)$ be given by (4.12). Then, this quadrature coincides with the one given in (4.13) with $\nu = -s > -1$ and $\tau = \frac{\pi}{\mathcal{Y}}$.*

Proof. First recall that, for any $\lambda > 0$, the function F_λ is even. Next, for $k \in \mathbb{N}$, note that $J_{-s}(\eta_k) = 0$. Use this observation to conclude that $\pi\xi_{-s,k} = \eta_k$. In addition, by setting $\tau = \frac{\pi}{\mathcal{Y}}$, it is seen that $\frac{\eta_k}{\mathcal{Y}} = \tau\xi_{-s,k}$. Substituting this into (4.13), it holds that

$$\begin{aligned}
\tau \sum_{k=1}^{\mathcal{K}} & \frac{2}{\pi^2 \xi_{-s,k} J_{1-s}^2(\pi\xi_{-s,k})} |\tau\xi_{-s,k}|^{1-2s} F_\lambda(\tau\xi_{-s,k}) \\
&= \sum_{k=1}^{\mathcal{K}} \frac{\tau}{\pi} \frac{2}{\pi\xi_{-s,k} J_{1-s}^2(\pi\xi_{-s,k})} |\tau\xi_{-s,k}|^{1-2s} F_\lambda(\tau\xi_{-s,k}) \\
&= \sum_{k=1}^{\mathcal{K}} \frac{2}{\mathcal{Y}\eta_k J_{1-s}^2(\eta_k)} \left(\frac{\eta_k}{\mathcal{Y}}\right)^{1-2s} F_\lambda\left(\frac{\eta_k}{\mathcal{Y}}\right) = Q_s^\infty(F_\lambda)\,.
\end{aligned}$$

□

Reference [40] provides an analysis for the quadrature (4.13). To describe it, following [40, Definition 2.1] the following class of functions is introduced.

Definition 4.7 (class $\mathcal{B}_{s,\ell}$)**.** *Let $\ell > 0$ and denote*

$$D_\ell = \{z \in \mathbb{C} : |\Im z| < \ell\}\,, \qquad \Gamma_\ell = \partial D_\ell\,.$$

By $\mathcal{B}_{s,\ell}$ we denote the collection of functions $G : \bar{D}_\ell \to \mathbb{C}$ that satisfy:

1. G is analytic in the strip D_ℓ.

2. For all $0 < c < \ell$, the integral

$$\mathcal{N}_{s,c}(G) = \int_{-\infty}^{\infty} \left[|x + \imath c|^{1-2s}|G(x + \imath c)| + |x - \imath c|^{1-2s}|G(x - \imath c)|\right]\, dx$$

   exists. In addition, we require that $\mathcal{N}_{s,\ell-0}(G) = \lim_{c\uparrow\ell} \mathcal{N}_{s,c}(G)$ exists and remains finite.

3. For all $0 < c < \ell$, we require that

$$\lim_{x\to\pm\infty} \int_{-c}^{c} |x + \imath y|^{1-2s}|G(x + \imath y)|\, dy = 0\,. \tag{4.14}$$

The usefulness of this class with regards to quadrature is the following result.

Theorem 4.8 (quadrature error estimate)**.** *Let $\ell > 0$, $s \in (0, 1)$, and $G \in \mathcal{B}_{s,\ell}$. Then*

$$|I_s(G) - Q_s^\infty(G)| \lesssim \mathcal{N}_{s,\ell-0}(G) \exp\left(-2\ell\mathcal{Y}\right)\,, \tag{4.15}$$

where the implicit constant depends only on s and ℓ.

Proof. See [40, Theorem 2.1], and use the equivalence of Lemma 4.6. □

The functions of interest in this dissertation are uniformly in a class $\mathcal{B}_{s,\ell}$.

Theorem 4.9 ($F_\lambda \in \mathcal{B}_{s,\ell}$)**.** *Let $s \in (0, 1)$, $\lambda_0 > 0$, and set $\ell = \frac{1}{2}\sqrt{\lambda_0}$. For every $\lambda \geq \lambda_0$ we have $F_\lambda \in \mathcal{B}_{s,\ell}$. Moreover, $\mathcal{N}_{s,\ell-0}(F_\lambda)$ depends only on s and ℓ.*

Proof. For this proof set $a_s = \frac{2\sin(\pi s)}{\pi}$. The conditions given in Definition 4.7 are now verified below.

1. First observe that the only poles of F_λ are $\pm\imath\sqrt{\lambda}$. Therefore, by setting $\ell = \frac{1}{2}\sqrt{\lambda_0}$ it is seen that, for every $\lambda \geq \lambda_0$, F_λ is analytic in D_ℓ.

2. Now, let $c \in (0, \ell)$. Observe that

$$
\begin{aligned}
|F_\lambda(x+\imath c)|^2 = \left|\frac{a_s}{x^2 - c^2 + \lambda + 2\imath cx}\right|^2 &= \frac{a_s^2}{|x^2 - c^2 + \lambda + 2\imath cx|^2} \\
&= \frac{a_s^2}{(x^2 - c^2 + \lambda)^2 + (2cx)^2} \\
&= \frac{a_s^2}{x^4 + 2c^2x^2 + 2x^2\lambda - 2c^2\lambda + c^4 + \lambda}.
\end{aligned}
$$

Then, for $x > c$,

$$
\begin{aligned}
|x+\imath c|^{1-2s}|F_\lambda(x+\imath c)| &= \frac{a_s(x^2+c^2)^{\frac{1-2s}{2}}}{\left((x^2 - c^2 + \lambda)^2 + 4c^2x^2\right)^{\frac{1}{2}}} \\
&= \frac{a_s x^{1-2s}\left(1 + \frac{c^2}{x^2}\right)^{\frac{1-2s}{2}}}{(x^4 + 2c^2x^2 + 2x^2\lambda - 2c^2\lambda + c^4 + \lambda^2)^{\frac{1}{2}}} \\
&= \frac{a_s x^{1-2s}\left(1 + \frac{c^2}{x^2}\right)^{\frac{1-2s}{2}}}{x^2\left(1 + \frac{2c^2}{x^2} + \frac{2\lambda}{x^2} - \frac{2c^2\lambda}{x^4} + \frac{c^4}{x^4} + \frac{\lambda^2}{x^4}\right)^{\frac{1}{2}}} \\
&\leq a_s x^{-1-2s}\left(1 + \frac{c^2}{x^2}\right)^{\frac{1-2s}{2}},
\end{aligned}
$$

upon noticing that $c < \frac{\sqrt{\lambda_0}}{2} \leq \frac{\sqrt{\lambda}}{2}$. Then,

$$
\frac{1}{a_s}\int_{-\infty}^{\infty} |x+\imath c|^{1-2s}|F_\lambda(x+\imath c)|\,dx \leq \int_{-\infty}^{-c^2} |x|^{-1-2s}\left(1+\frac{c^2}{x^2}\right)^{\frac{1-2s}{2}}\,dx + C + \int_{c^2}^{\infty} x^{-1-2s}\left(1+\frac{c^2}{x^2}\right)^{\frac{1-2s}{2}}\,dx\,.
$$

To obtain that

$$
\int_{c^2}^{\infty} x^{-1-2s}\,dx < \infty\,,
$$

it must be that $-1-2s<-1$, or $s>0$. A similar analysis holds for $|x-\imath c|^{1-2s}|F_\lambda(x-\imath c)|$. Hence, $\mathcal{N}_{s,c}(F_\lambda)$ exists.

Notice, finally, that $c \leq \ell = \frac{\sqrt{\lambda}}{2} < \sqrt{\lambda}$. In other words, the integral's domain is well separated from the poles of F_λ. Consequently,

$$\mathcal{N}_{s,\ell-0}(F_\lambda) = \lim_{c\uparrow\ell} \mathcal{N}_{s,c}(F_\lambda)$$

exists, is finite, and only depends on s and ℓ.

3. Now proceed to check the last condition given by (4.14). Suppose that $x>2c$. Then,

$$\begin{aligned}\int_{-c}^{c} |x+\imath y|^{1-2s}\left|\frac{1}{(x+\imath y)^2+\lambda}\right| dy &= 2\int_0^c \frac{(x^2+y^2)^{\frac{1-2s}{2}}}{\sqrt{|x^2+2\imath xy-y^2+\lambda|}}\, dy \\ &= 2\int_0^c \sqrt{\frac{(x^2+y^2)^{1-2s}}{(x^2+\lambda-y^2)^2+4x^2y^2}}\, dy \\ &\leq 2\int_0^c \frac{\sqrt{(x^2+y^2)^{1-2s}}}{x^2+\lambda-y^2}\, dy \\ &= \frac{2x^{1-2s}}{x}\int_0^c \frac{\left(1+\frac{y^2}{x^2}\right)^{\frac{1-2s}{2}}}{1+\frac{\lambda}{x^2}-\frac{y^2}{x^2}}\, \frac{dy}{x} \\ &\leq 2x^{-2s}\int_0^{c/x} \frac{(1+t^2)^{\frac{1-2s}{2}}}{1-t^2}\, dt\,,\end{aligned}$$

where the change of variables $t=\frac{y}{x}$ has been applied. Since $0<2y<2c<x$, it can be concluded that $t=\frac{y}{x}\in(0,\frac{1}{2})$, which implies $1-t^2>\frac{3}{4}$. Therefore,

$$\int_{-c}^{c} |x+\imath y|^{1-2s}\left|\frac{1}{(x+\imath y)^2+\lambda}\right| dy \leq \frac{8}{3}x^{-2s}\int_0^{c/x} (1+t^2)^{\frac{1-2s}{2}}\, dt\,.$$

Next, the proof proceeds depending on the sign of $1-2s$. If $1-2s>0$, then

$$\int_{-c}^{c} |x+\imath y|^{1-2s}\left|\frac{1}{(x+\imath y)^2+\lambda}\right| dy \leq \frac{8}{3}x^{-2s}\left(\frac{5}{4}\right)^{\frac{1-2s}{2}} \frac{c}{x}\,,$$

whereas if $1 - 2s < 0$,

$$\int_{-c}^{c} |x + \imath y|^{1-2s} \left| \frac{1}{(x + \imath y)^2 + \lambda} \right| dy \leq \frac{8}{3} x^{-2s} \frac{c}{x} .$$

In any event, these estimates are sufficient to conclude that

$$\lim_{x \to \pm\infty} \int_{-c}^{c} |x + \imath y|^{1-2s} |F_\lambda(x + \imath y)| \, dy = 0 .$$

All requirements have been verified and the proof is complete. □

An immediate consequence is an estimate for the quadrature error.

Corollary 4.10 (quadrature error)**.** *Let $s \in (0,1) \setminus \{\frac{1}{2}\}$. Let the weights and nodes of Q_s^∞ be given by (4.12). Then*

$$\sup_{\lambda \geq \frac{1}{C_P}} |I_s(F_\lambda) - Q_s^\infty(F_\lambda)| \lesssim \exp\left(-\frac{\mathcal{Y}}{\sqrt{C_P}} \right) ,$$

where C_P is the best constant in the Poincaré inequality; see (4.1).

Proof. It suffices to set $\lambda_0 = \frac{1}{C_P}$ in Theorem 4.9. □

The remaining step is to bound the truncation error, which is done by the following lemma.

Lemma 4.11 (truncation error)**.** *Let $s \in (0,1) \setminus \{\frac{1}{2}\}$ and $\mathcal{K} \in \mathbb{N}$. Let the nodes and weights of $Q_s^{\mathcal{K}}$ be given by (4.12). If $\mathcal{K}$ is sufficiently large, then the truncation error may be controlled via*

$$\sup_{\lambda \geq \frac{1}{C_P}} |Q_{(s)}^\infty(F_\lambda) - Q_{(s)}^{\mathcal{K}}(F_\lambda)| \lesssim \left(\frac{\mathcal{Y}}{\mathcal{K}} \right)^{2s} .$$

Proof. To begin the proof, start by using the asymptotic behavior of the Bessel function of the first kind for large arguments [1, Formula 9.2.1], i.e.,

$$J_{-s}(z) \sim \sqrt{\frac{2}{\pi z}} .$$

With this at hand, it holds that

$$
\begin{aligned}
|Q_s^\infty(F_\lambda) - Q_s^{\mathcal{K}}(F_\lambda)| &\leq \frac{\pi}{\mathcal{Y}} \sum_{k>\mathcal{K}} \frac{2}{\pi \eta_k J_{1-s}(\eta_k)^2} \left(\frac{\eta_k}{\mathcal{Y}}\right)^{1-2s} F_\lambda\left(\frac{\eta_k}{\mathcal{Y}}\right) \\
&= \frac{4\sin(\pi s)}{\mathcal{Y}} \sum_{k>\mathcal{K}} \frac{1}{\pi \eta_k J_{1-s}(\eta_k)^2} \left(\frac{\eta_k}{\mathcal{Y}}\right)^{1-2s} \frac{1}{\frac{\eta_k^2}{\mathcal{Y}^2} + \lambda} \\
&\lesssim \frac{1}{\mathcal{Y}} \sum_{k>\mathcal{K}} \left(\frac{\eta_k}{\mathcal{Y}}\right)^{1-2s} \frac{1}{\frac{\eta_k^2}{\mathcal{Y}^2} + \lambda} \\
&= \mathcal{Y}^{2s} \sum_{k>\mathcal{K}} \frac{\eta_k^{1-2s}}{\eta_k^2 + \mathcal{Y}^2 \lambda} \,.
\end{aligned}
$$

Next, McMahon's asymptotic expansion for large zeros [DLMF, Section 10.21 (vi)] is applied:

$$
\eta_k \sim k\pi \,,
$$

to observe that

$$
\sum_{k>\mathcal{K}} \frac{\eta_k^{1-2s}}{\eta_k^2 + \mathcal{Y}^2\lambda} \lesssim \sum_{k>\mathcal{K}} \frac{1}{k^{1+2s}} = \zeta(1+2s, \mathcal{K}+1) \,,
$$

where ζ denotes the Hurwitz Zeta function [DLMF, Section 25.11]. Using that, for large $\mathcal{K}$, the asymptotic [DLMF, Equation 25.11.43] shows that

$$
\zeta(1+2s, \mathcal{K}+1) \sim (\mathcal{K}+1)^{-2s} < \mathcal{K}^{-2s} \,.
$$

The claimed result follows. □

With Corollary 4.10 and Lemma 4.11 at hand, a final corollary to Theorem 4.1 is given to provide an error estimate for $s \neq \frac{1}{2}$.

Corollary 4.12 (error estimate for $s \neq \frac{1}{2}$). *Let $\Omega \subset \mathbb{R}^d$ be a bounded, convex polytope and $f \in L^2(\Omega)$. Let $u \in \mathbb{H}^s(\Omega)$ be the solution to (1.1) for $s \in (0,1) \setminus \{\frac{1}{2}\}$. Let $u_{h,\mathcal{Y}}^{\mathcal{K}}$ be given by (4.4), where $\mathbb{V}_h$ is constructed using a quasi-uniform mesh of size h. Then, we have that*

$$
\|u - u_{h,\mathcal{Y}}^{\mathcal{K}}\|_{L^2(\Omega)} \lesssim \left(h^{2s} + \exp\left(-\frac{\mathcal{Y}}{\sqrt{C_P}}\right) + \left(\frac{\mathcal{Y}}{\mathcal{K}}\right)^{2s}\right) \|f\|_{L^2(\Omega)} \,,
$$

where C_P is given by (4.1). In particular, if the truncation parameter is chosen as $\mathcal{Y} \sim 2s|\log h|$ and the number of quadrature points is chosen as $\mathcal{K} \sim \frac{\mathcal{Y}}{h}$, we may conclude that

$$\|u - u_{h,\mathcal{Y}}^{\mathcal{K}}\|_{L^2(\Omega)} \lesssim h^{2s}\|f\|_{L^2(\Omega)} \,.$$

Proof. Apply Corollary 4.10 and Lemma 4.11 to each one of the terms in the conclusion of Theorem 4.1.

The practical error estimate, that is the one given solely in terms of h, is given by using the prescribed relations between all discretization parameters. □

Chapter 5

Numerical Results

Previous work demonstrated the convergence of the tensor FEM of the Caffarelli-Silvestre extension problem without diagonalization [39] and with diagonalization [8]. In order to validate the convergence estimates for the new exact diagonalization technique, similar numerical experiments are conducted. The exact diagonalization method is implemented using the `deal.ii` finite element library [5, 6] which makes use of a wide array of scientific libraries including the Message Passing Interface (MPI) for parallel computation. Using the MPI interface in `deal.ii`, the parallel performance of the method is measured. Both strong scaling and weak scaling performance is measured. We note that all meshes in Ω use uniform, conforming $\mathbb{Q}_1$ elements. Using the error estimates from Chapter 4, the truncation height $\mathcal{Y}$ and number of eigenpair contributions can be determined from a given mesh size h in order to maintain the discretization error of h^{2s}. Recall, the recipe for these two parameters is

$$\mathcal{Y} = 2s|\log(h)| \ , \ \mathcal{K} = \frac{2s|\log(h)|}{h} \, . \tag{5.1}$$

5.1 Convergence Studies

To demonstrate the method's ability to achieve the theoretical convergence rate, the numerical solution is computed to approximate the solution to two smooth solutions. For the first problem, an L-shaped domain, $\Omega_L \subset \mathbb{R}^2$, is considered that is determined by the vertices

$$\mathbf{c} \in \{(0,0)\,,(1,0)\,,(-1,1)\,,(-1,-1)\,,(0,-1)\,\}\,.$$

For the right hand side, the following function is prescribed:

$$f(x_1,x_2) = (2\pi^2)^s \sin(\pi x_1)\sin(\pi x_2) + \\ \big((3^2+2^2)\pi^2\big)^s \sin(3\pi x_1)\sin(2\pi x_1) + \big((5^2+4^2)\pi^2\big)^s \sin(5\pi x_1)\sin(4\pi x_2)\,.$$

In this case, the exact solution is

$$u(x_1,x_2) = \sin(\pi x_1)\sin(\pi x_2) + \sin(3\pi x_1)\sin(2\pi x_2) + \sin(5\pi x_1)\sin(4\pi x_2)\,.$$

Notice that, since the domain is not convex, the theoretical developments do not apply as written. However, it is easy to realize that the only place where convexity is invoked is in the derivation of (4.3). One can easily develop a more general error analysis that takes into account a smaller regularity shift.

A $\mathbb{Q}_1$-FEM on uniformly refined meshes in Ω is used. For the parameters in the extended dimension, the relations in (5.1) are applied. The results in Table 5.1 provide details regarding the number of degrees of freedom for the mesh in Ω_L, truncation height, number of eigenpairs used in the approximation, $L^2(\Omega_L)$ error, and the observed convergence rate for $s = 0.25$ and $s = 0.75$. In both cases, the observed convergence rate matches the theoretical rate of $2s$. The error is also shown in Figure 5.1 for both $s = 0.25$ and $s = 0.75$ along with the corresponding theoretical convergence lines.

Table 5.1: Number of elements for the Q_1 FEM on uniform meshes in Ω_L, height $\mathcal{Y}$ of the truncation, $\mathcal{K}$ number of eigenpairs used, the resulting error in the $L^2(\Omega)$ norm, and the convergence rate for $s = 0.25$ and $s = 0.75$.

(a) $s = 0.25$

# $\mathcal{T}_{\Omega_L}$	$\mathcal{Y}$	$\mathcal{K}$	Error	Rate
65	0.69315	2	1.18473e+00	–
225	1.03972	8	9.22409e-01	0.36
833	1.38629	22	6.59965e-01	0.48
3201	1.73287	55	4.68283e-01	0.50
12545	2.07944	133	3.29807e-01	0.51
49665	2.42602	310	2.33219e-01	0.50
197633	2.77259	709	1.64819e-01	0.50
788481	3.11916	1597	1.16468e-01	0.50

(b) $s = 0.75$

# $\mathcal{T}_{\Omega_L}$	$\mathcal{Y}$	$\mathcal{K}$	Error	Rate
65	2.07944	8	8.62909e-01	–
225	3.11916	24	3.54492e-01	1.28
833	4.15888	66	1.15435e-01	1.62
3201	5.19860	166	3.69013e-02	1.65
12545	6.23832	399	1.19066e-02	1.63
49665	7.27805	931	3.91190e-03	1.61
197633	8.31777	2129	1.30633e-03	1.58
788481	9.35749	4791	4.42652e-04	1.56

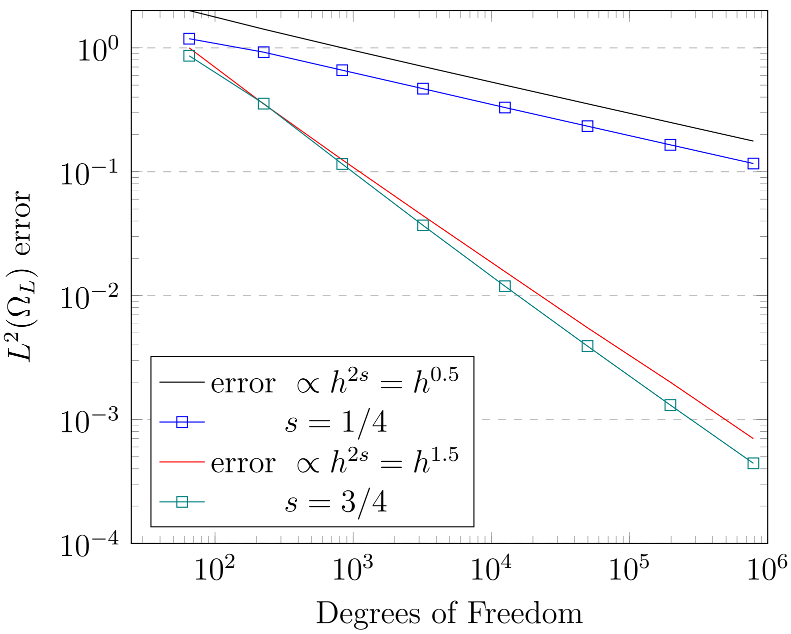

Figure 5.1: Error in the $L^2(\Omega_L)$ norm versus the number of degrees of freedom using Q_1 finite elements for $s = 1/4$ and $s = 3/4$ on uniformly refined meshes of Ω_L.

For the second problem, a circular domain is considered specified by

$$\Omega_C = \{x \in \mathbb{R}^2 : |x| < 1\}\,.$$

Using polar coordinates, it can be shown that

$$\phi_{m,n}(r,\theta) = J_m(j_{m,n}r)\big(A_{m,n}\cos(m\theta) + B_{m,n}\sin(m\theta)\big)\,,$$

where J_m denotes the Bessel function of the first kind with parameter m, $j_{m,n}$ is the n^{th} zero of $J_m(z)$, and $A_{m,n}$, $B_{m,n}$ are real normalization constants that ensure $\|\phi_{m,n}\|_{L^2(\Omega)} = 1$ for all $m,n \in \mathbb{N}$. It may also be shown that $\lambda_{m,n} = (j_{m,n})^2$. Now, take $f = (\lambda_{1,1})^s\phi_{1,1}$. Then by Equation (2.1),

$$u(r,\theta) = \phi_{1,1}(r,\theta)\,.$$

As in the case of the $L-$shaped domain, a $\mathbb{Q}_1$-FEM on uniformly refined meshes in Ω_C is used. Similarly, the formulas given in (5.1) are used. Convergence results are presented in Table 5.2 along with details regarding the number of degrees of freedom for the mesh in Ω_C, truncation height, number of eigenpairs used in the approximation, $L^2(\Omega_C)$ error, and the observed convergence rate for $s = 0.25$ and $s = 0.75$. For both values of s the observed convergence rates agree with the theoretical convergence rate of $2s$. Figure 5.2 shows the error for both $s = 0.25$ and $s = 0.75$ next to the corresponding theoretical error convergence lines.

Table 5.2: Number of elements for the Q_1 FEM on uniform meshes in Ω_C, height $\mathcal{Y}$ of the truncation, $\mathcal{K}$ number of eigenpairs used, the resulting error in the $L^2(\Omega)$ norm, and the convergence rate for $s = 0.25$ and $s = 0.75$.

(a) $s = 0.25$

# $\mathcal{T}_{\Omega_C}$	$\mathcal{Y}$	$\mathcal{K}$	Error	Rate
89	0.69315	2	3.95613e-01	–
337	1.03972	8	2.49627e-01	0.66
1313	1.38629	22	1.74086e-01	0.52
5185	1.73287	55	1.23124e-01	0.50
20609	2.07944	133	8.67400e-02	0.51
82177	2.42602	310	6.13695e-02	0.50
328193	2.77259	709	4.33828e-02	0.50
1311745	3.11916	1597	3.06600e-02	0.50

(b) $s = 0.75$

# $\mathcal{T}_{\Omega_C}$	$\mathcal{Y}$	$\mathcal{K}$	Error	Rate
89	2.07944	8	6.73748e-02	–
337	3.11916	24	2.09143e-02	1.69
1313	4.15888	66	6.38904e-03	1.71
5185	5.19860	166	2.04193e-03	1.65
20609	6.23832	399	6.70252e-04	1.61
82177	7.27805	931	2.24480e-04	1.58
328193	8.31777	2129	7.62235e-05	1.56
1311745	9.35749	4791	2.61759e-05	1.54

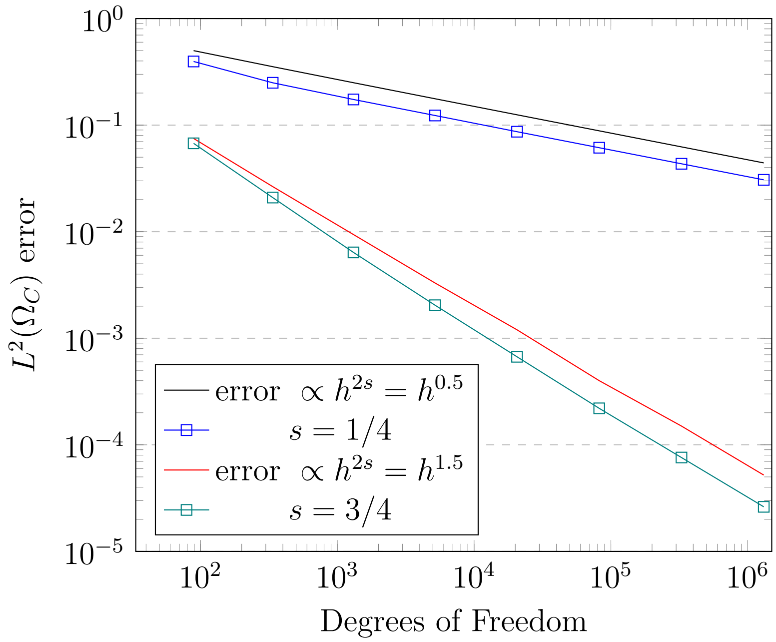

Figure 5.2: Error in the $L^2(\Omega_C)$ norm versus the number of degrees of freedom using Q_1 finite elements for $s = 1/4$ and $s = 3/4$ on uniformly refined meshes of Ω_C.

5.2 Parallel Performance

In Chapter 3, it is stated that the diagonalization technique exposes an inherent parallelization opportunity for the numerical algorithm. The parallel performance is measured in two scaling tests. For both tests, set $\Omega_\square = (0,1)^2$. In this case, it is known that

$$\phi_{m,n}(x_1,x_2) = \sin(m\pi x_1)\sin(n\pi x_2)\ ,\ \ \lambda_{m,n} = \pi^2(m^2+n^2)\ ,\ \ m,n \in \mathbb{N}\,.$$

Prescribing $f(x_1,x_2) = (2\pi^2)^s \sin(\pi x_1)\sin(\pi x_2)$, from (2.1) it holds that

$$u(x_1,x_2) = \sin(\pi x_1)\sin(\pi x_2)\,.$$

A uniform mesh of $\mathbb{Q}_1$ elements is used on $\Omega_\square$ consisting of 1089 degrees of freedom. For convenience, $\mathcal{Y}$ is set to 10 and $s = 0.25$ is used in all cases.

For the first test, the strong scaling performance is measured. For this test, the number of eigenpairs from the exact diagonalization is fixed at $\mathcal{K} = 1,024,000$. The number of MPI processes used to solve the problem is then increased from 1 up to 32. Figure 5.3 illustrates the speedup of the algorithm compared with the ideal linear speedup curve. At 32 MPI processes, the algorithm maintains a speedup of nearly 28.

The second parallel performance test measures the weak scaling properties of the algorithm. The design of this experiment is to scale the amount of computational work linearly with the number of MPI processes and examine the efficiency of the algorithm, i.e. does the total wall time remain constant for each parallel run? The number of eigenpairs from the exact diagonalization is calculated by 256,000 times the number of MPI processes launched. The parallel efficiency can then be calculated as a ratio of the runtime for N MPI processes to the runtime for a single MPI process. Figure 5.4 shows the parallel efficiency curve for the numerical algorithm. A parallel efficiency of nearly 85% is maintained with 32 MPI processes.

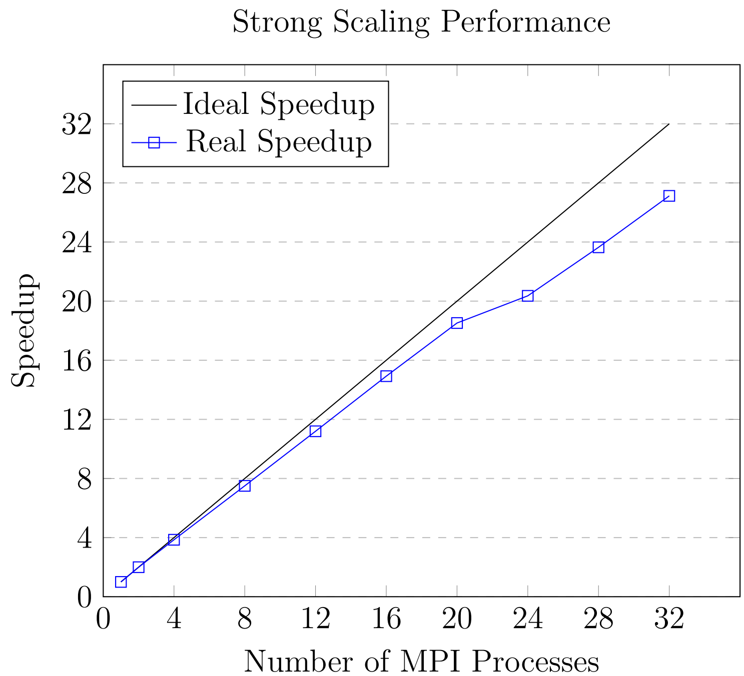

Figure 5.3: Strong scaling performance in a simple test case.

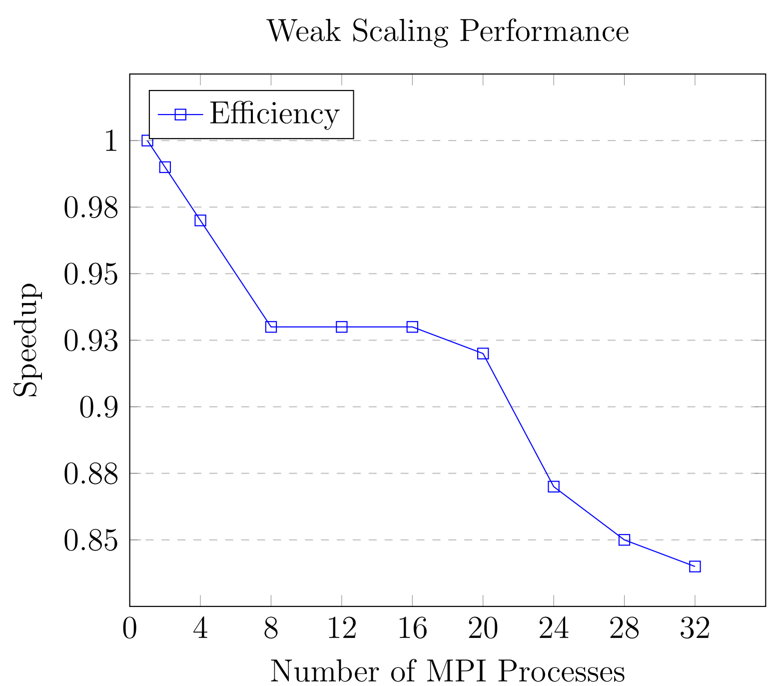

Figure 5.4: Weak scaling performance in a simple test case.

Chapter 6

Conclusion

This work introduced and analyzed a novel exact diagonalization technique for approximating the solution to the spectral fractional Laplacian using a localized FEM based on the Caffarelli-Silvestre extension technique [19]. The new technique extends previous work [8] which used hp-FEM to approximate the eigenpairs of the eigenvalue problem in the extended dimension. Error estimates measured in the $L^2(\Omega)$ norm were derived. Numerical experiments were conducted to demonstrate that the new technique was able to reproduce the theoretical convergence rate. Parallel performance of the new technique was measured by two scaling benchmarks. These benchmarks demonstrate that the algorithm is able to take advantage of the parallelizability inherent in the numerical scheme.

Bibliography

[1] Abramowitz, M., Stegun, I. A., and Romer, R. H. (1988). Handbook of mathematical functions with formulas, graphs, and mathematical tables. 20, 21, 29, 31, 36

[2] Aceto, L. and Novati, P. (2017). Rational approximation to the fractional laplacian operator in reaction-diffusion problems. *SIAM Journal on Scientific Computing*, 39(1):A214–A228. 3, 4, 6

[3] Adams, R. A. and Fournier, J. J. (2003). *Sobolev spaces*. Elsevier. 8

[4] Ainsworth, M. and Glusa, C. (2018). Hybrid finite element–spectral method for the fractional Laplacian: Approximation theory and efficient solver. *SIAM Journal on Scientific Computing*, 40(4):A2383–A2405. 5

[5] Arndt, D., Bangerth, W., Bergbauer, M., Feder, M., Fehling, M., Heinz, J., Heister, T., Heltai, L., Kronbichler, M., Maier, M., Munch, P., Pelteret, J.-P., Turcksin, B., Wells, D., and Zampini, S. (2023, submitted). The `deal.II` library, version 9.5. *Journal of Numerical Mathematics*. 39

[6] Arndt, D., Bangerth, W., Davydov, D., Heister, T., Heltai, L., Kronbichler, M., Maier, M., Pelteret, J.-P., Turcksin, B., and Wells, D. (2021). The deal.II finite element library: Design, features, and insights. *Computers & Mathematics with Applications*, 81:407–422. 39

[7] Balakrishnan, A. V. (1960). Fractional powers of closed operators and the semigroups generated by them. *Pacific J. Math.*, 10:419–437. 3

[8] Banjai, L., Melenk, J. M., Nochetto, R. H., Otárola, E., Salgado, A. J., and Schwab, C. (2019). Tensor FEM for spectral fractional diffusion. *Foundations of Computational Mathematics*, 19(4):901–962. 2, 3, 5, 6, 10, 12, 16, 17, 22, 39, 48

[9] Banjai, L., Melenk, J. M., and Schwab, C. (2023). Exponential convergence of *hp* FEM for spectral fractional diffusion in polygons. *Numer. Math.*, 153(1):1–47. 17

[10] Bateman, H. (1953). *Higher transcendental functions [volumes i-iii]*, volume 3. McGRAW-HILL book company. 29

[11] Birman, M. S. and Solomjak, M. Z. (1987). *Spectral theory of selfadjoint operators in Hilbert space*. Mathematics and its Applications (Soviet Series). D. Reidel Publishing Co., Dordrecht. Translated from the 1980 Russian original by S. Khrushchëv and V. Peller. 3

[12] Bonito, A., Borthagaray, J. P., Nochetto, R. H., Otárola, E., and Salgado, A. J. (2018). Numerical methods for fractional diffusion. *Computing and Visualization in Science*, 19(5):19–46. 3, 4, 25

[13] Bonito, A. and Nazarov, M. (2021). Numerical simulations of surface quasi-geostrophic flows on periodic domains. *SIAM Journal on Scientific Computing*, 43(2):B405–B430. 1

[14] Bonito, A. and Pasciak, J. (2015). Numerical approximation of fractional powers of elliptic operators. *Mathematics of Computation*, 84(295):2083–2110. 3, 4, 6, 25

[15] Bonito, A. and Pasciak, J. E. (2017). Numerical approximation of fractional powers of regularly accretive operators. *IMA Journal of Numerical Analysis*, 37(3):1245–1273. 4

[16] Bonito, A. and Wei, P. (2020). Electroconvection of thin liquid crystals: Model reduction and numerical simulations. *Journal of Computational Physics*, 405:109140. 1

[17] Borthagaray, J. P., Li, W., and Nochetto, R. H. ([2020] ©2020). Linear and nonlinear fractional elliptic problems. In *75 years of mathematics of computation*, volume 754 of *Contemp. Math.*, pages 69–92. Amer. Math. Soc., [Providence], RI. 3

[18] Cabré, X. and Tan, J. (2010). Positive solutions of nonlinear problems involving the square root of the Laplacian. *Advances in Mathematics*, 224(5):2052–2093. 1, 2, 10, 11

[19] Caffarelli, L. and Silvestre, L. (2007). An extension problem related to the fractional Laplacian. *Communications in partial differential equations*, 32(8):1245–1260. 1, 2, 10, 11, 48

[20] Chen, S. and Shen, J. (2020). An efficient and accurate numerical method for the spectral fractional laplacian equation. *Journal of Scientific Computing*, 82(1):17. 5

[21] Cusimano, N. and Gerardo-Giorda, L. (2018). A space-fractional monodomain model for cardiac electrophysiology combining anisotropy and heterogeneity on realistic geometries. *Journal of Computational Physics*, 362:409–424. 1

[22] Cusimano, N., Gerardo-Giorda, L., and Gizzi, A. (2021). A space-fractional bidomain framework for cardiac electrophysiology: 1d alternans dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(7). 1

[23] D'Elia, M., Du, Q., Glusa, C., Gunzburger, M., Tian, X., and Zhou, Z. (2020). Numerical methods for nonlocal and fractional models. *Acta Numer.*, 29:1–124. 3

[24] Di Nezza, E., Palatucci, G., and Valdinoci, E. (2012). Hitchhiker's guide to the fractional sobolev spaces. *Bulletin des sciences mathématiques*, 136(5):521–573. 8

[DLMF] DLMF. *NIST Digital Library of Mathematical Functions*. `https://dlmf.nist.gov/`, Release 1.1.11 of 2023-09-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds. 20, 37

[26] Duan, B., Lazarov, R. D., and Pasciak, J. E. (2020). Numerical approximation of fractional powers of elliptic operators. *IMA J. Numer. Anal.*, 40(3):1746–1771. 3

[27] Ern, A. and Guermond, J.-L. (2004). *Theory and practice of finite elements*, volume 159. Springer. 15, 24

[28] Evans, L. C. (2022). *Partial differential equations*, volume 19. American Mathematical Society. 9

[29] Gatto, P. and Hesthaven, J. S. (2015). Numerical approximation of the fractional Laplacian via *hp* hp-finite elements, with an application to image denoising. *Journal of Scientific Computing*, 65(1):249–270. 1

[30] Harizanov, S., Lazarov, R., Margenov, S., Marinov, P., and Pasciak, J. (2020). Analysis of numerical methods for spectral fractional elliptic equations based on the best uniform rational approximation. *Journal of Computational Physics*, 408:109285. 4

[31] Harizanov, S., Lazarov, R., Margenov, S., Marinov, P., and Vutov, Y. (2018). Optimal solvers for linear systems with fractional powers of sparse spd matrices. *Numerical Linear Algebra with Applications*, 25(5):e2167. 4, 6

[32] Hofreither, C. (2020). A unified view of some numerical methods for fractional diffusion. *Computers & Mathematics with Applications*, 80(2):332–350. 6

[33] Ivrii, V. (2016). 100 years of Weyl's law. *Bulletin of Mathematical Sciences*, 6(3):379–452. 5

[34] Kufner, A. and Opic, B. (1984). How to define reasonably weighted Sobolev spaces. *Commentationes Mathematicae Universitatis Carolinae*, 25(3):537–554. 10

[35] Lischke, A., Pang, G., Gulian, M., and et al. (2020). What is the fractional Laplacian? A comparative review with new results. *J. Comput. Phys.*, 404:109009, 62. 3

[36] Lunardi, A. et al. (2009). *Interpolation theory*, volume 9. Edizioni della normale Pisa. 25

[37] Matsuki, M. and Ushijima, T. (1993). A note on the fractional powers of operators approximating a positive definite selfadjoint operator. *Journal of the Faculty of Science, the University of Tokyo. Sect. 1 A, Mathematics*, 40(2):517–528. 25

[38] Metzler, R. and Klafter, J. (2000). The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Physics Reports*, 339(1):1–77. 1

[39] Nochetto, R. H., Otárola, E., and Salgado, A. J. (2015). A PDE approach to fractional diffusion in general domains: a priori error analysis. *Foundations of Computational Mathematics*, 15(3):733–791. 2, 5, 6, 9, 11, 12, 39

[40] Ogata, H. (2005). A numerical integration formula based on the bessel functions. *Publications of the Research Institute for Mathematical Sciences*, 41(4):949–970. 31, 33

[41] Oldham, K. B., Myland, J., and Spanier, J. (2009). *An atlas of functions: with equator, the atlas function calculator*. Springer. 28

[42] Reed, M. and Simon, B. (1981). *I: Functional analysis*, volume 1. Academic press. 12

[43] Stinga, P. R. and Torrea, J. L. (2010). Extension problem and Harnack's inequality for some fractional operators. *Communications in Partial Differential Equations*, 35(11):2092–2122. 1, 2, 10, 11

[44] Strauss, W. A. (2007). *Partial differential equations: An introduction*. John Wiley & Sons. 18

[45] Thomée, V. (2007). *Galerkin finite element methods for parabolic problems*, volume 25. Springer Science & Business Media. 24

[46] Vabishchevich, P. N. (2015). Numerically solving an equation for fractional powers of elliptic operators. *Journal of Computational Physics*, 282:289–302. 3, 6

[47] Zhang, Z. (2015). How many numerical eigenvalues can we trust? *Journal of Scientific Computing*, 65(2):455–466. 3, 16

Appendix

A Computation Details and Reproducibility

All numerical experiments were conducted using a single node containing one Intel Xeon Gold 6246R processor running at 3.40GHz with a total of 628 GB of RAM. The node was running Fedora release 36 configured with Linux kernel 6.2.13. All libraries and source code were compiled using GCC version 12.2.1. The source code for deal.ii and its dependencies were built using the *candi* system, available at https://github.com/dealii/candi. Versions for all libraries used are given in Table 1. The source code for the numerical algorithm is available at https://github.com/shedsaw/Exact-Diagonalization-Tensor-FEM.

Table 1: Library versions.

OpenMPI	4.1.5	NetCDF	4.7.4
METIS	5.1.0	Numdiff	5.9.0
deal.ii	9.4.2-r3	Oce-OCE	0.18.3
adloc	2.7.3	OpenBLAS	0.3.17
ARPACK-NG	3.8.0	P4est	2.3.2
assimp	4.1.0	ParMETIS	4.0.3
astyle	2.04	PETSc	3.16.4
boost	1.63.0	Scalapack	2.1.0
bzip2	1.0.6	SLEPc	3.16.2
Cmake	3.20.5	Sundials	5.7.0
Ginkgo	1.4.0	SuperLU Dist	5.1.2
Gmsh	4.8.4	Symengine	0.8.1
GSL	2.6	Trilinos Release	13.2.0
HDF5	1.10.8	Zlib	1.2.8
Mumps	5.4.0.5		

Vita

Shane Sawyer was born in Chattanooga, Tennessee. He completed his undergraduate degree in Applied Mathematics at the University of Tennessee at Chattanooga. He subsequently earned a Master's degree in Engineering with a concentration in Computational Engineering from the University of Tennessee at Chattanooga. His interests include computer programming, especially scientific and parallel varieties, physics, fluid dynamics, and reading.