Sparse Space-time Boundary Element Methods for the Heat Equation

Anne Reinarz

July 2015

This thesis is submitted to the Department of Mathematics and Statistics in partial fulfilment of the requirement for the degree of Doctorate of Philosophy

Abstract

The goal of this work is the efficient solution of the heat equation with Dirichlet or Neumann boundary conditions using the Boundary Elements Method (BEM). Efficiently solving the heat equation is useful, as it is a simple model problem for other types of parabolic problems. In complicated spatial domains as often found in engineering, BEM can be beneficial since only the boundary of the domain has to be discretised. This makes BEM easier than domain methods such as finite elements and finite differences, conventionally combined with time-stepping schemes to solve this problem.

The contribution of this work is to further decrease the complexity of solving the heat equation, leading both to speed gains (in CPU time) as well as requiring smaller amounts of memory to solve the same problem. To do this we will combine the complexity gains of boundary reduction by integral equation formulations with a discretisation using wavelet bases. This reduces the total work to $O(h_x^{-(d-1)})$, when the solution of the linear system is performed with linear complexity.

We show that the discretisation with a wavelet basis leads to a numerically sparse matrix. Further, we show that this matrix can be compressed without losing accuracy of the underlying Galerkin scheme. This matrix compression reduces the number of non-zero matrix entries from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$. Thus, we can indeed solve the linear system in linear time.

It has been shown theoretically that using sparse grid methods leads to considerably higher convergence rates in the energy norm of the problem. In this work we will show that the convergence can be further improved for some choices of polynomial degrees by using more general sparse grid spaces. We also give numerical results to verify the theoretical bounds from [Chernov, Schwab, 2013].

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Signed.....
Anne Reinarz

Acknowledgments

I would like to thank my supervisors Alexey Chernov and Steve Langdon. For the past four years, Alexey has continuously provided me with support and with plenty of good ideas. He always helped me when I had questions. He continued support even after moving to Oldenburg, providing many excellent corrections to this thesis. Steve has always been very encouraging, especially in helping me to attend conferences to show my work and involving me in interesting discussions.

I would like to thank the University of Reading and the Department of Mathematics and Statistics for financial support, as well as SIAM for enabling me to go to CSE15 in Salt Lake City.

I would also like to thank my fellow PhD students for some very enjoyable tea breaks. In particular, I thank Kasia and Noeleene for their help with proof-reading this thesis. I am also thankful to Peta for her help with organising travel and her constant encouragement.

I would like to thank my cousin Maike for proof-reading my thesis and giving excellent feedback, my husband Matthias for providing punctuation and moral support, my sister Lisa for always believing in me, and my parents and aunt for fostering a love of science in me and for their continued support over the years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The goal of this work is the efficient solution of the heat equation with Dirichlet or Neumann boundary conditions. The heat equation is a simple model problem for other types of parabolic problems. The numerical solution of non-stationary parabolic problems is needed in numerous fields, which we describe below.

We solve the heat equation

$$(\partial_t - \Delta)u = f \tag{1.1}$$

for some right-hand side f, posed in a spatial domain $\Omega \subset \mathbb{R}^d$ and on the time interval $(0, T)$. Throughout we will use zero initial conditions

$$u = 0 \qquad \text{at } \{t = 0\} \times \Omega$$

and either Dirichlet boundary conditions, which means that the value of the solution on the boundary is given

$$u|_{\partial\Omega} = g$$

or Neumann boundary conditions, which means that the normal derivative of the solution on the boundary is given

$$\partial_n u|_{\partial\Omega} = g.$$

Solving the heat equation has many applications in physics and engineering [51]. The primary application in three dimensions is modelling heat flow in an isotropic medium. Other applications include pressure diffusion in porous media or diffusion of a chemical substance from a region of higher to one of lower concentration. For the latter problem the diffusion coefficients may depend on the concentration, leading to a non-linear equation, which is not covered in this work.

The heat equation can also arise in problems in image analysis and machine learning, such as shape recognition problems [53]. Further, one can use an equation of this form for image processing problems such as linear denoising [43].

The heat equation also appears in financial modelling [52]. In particular, it is used for the valuation of financial derivatives. Further, the differential equation derived from the Black-Scholes option pricing model can easily be transformed into the heat equation. Since these forms of the problem typically do not have analytical solutions, efficient numerical methods for solving them are important.

1.1 Background

Conventional methods for solving parabolic boundary value problems include Finite Element Methods (FEM), numerical schemes which approximate the solution using a variational formulation on a simple subdivision of the domain Ω. This is combined with a low-order time stepping scheme, such as implicit Euler or Crank-Nicholson [50].

Another alternative is to use convolution quadrature (see [37] and [38]) for the time discretisation. Convolution quadrature provides a stable time-stepping scheme by using a Laplace transform of the kernel function. It can be applied to a variety of problems, see e.g. [4].

In complicated spatial domains as often found in engineering, the Boundary Element Method (BEM) can be very useful since only the boundary of the domain has to be discretised, making it easier than domain methods such as finite elements and finite differences. In several applications, the needed data is not the solution of the problem itself, instead it is given by the boundary values of the solution or by its derivatives. Another advantage is that this data can be obtained directly from the boundary integral formulation.

Further, BEM can be used for problems with unbounded domains since a volume mesh of the unbounded domain does not need to be generated. An example of the discrete solution to an exterior problem and an interior problem on a smooth domain are shown in Figure 1.1. BEM are introduced in detail in the books [36], [49] and [44], which cover only elliptic problems. However, most of the ideas are easily transferable to the case of parabolic problems.

We will use a Galerkin discretisation for the boundary integral formulation of the

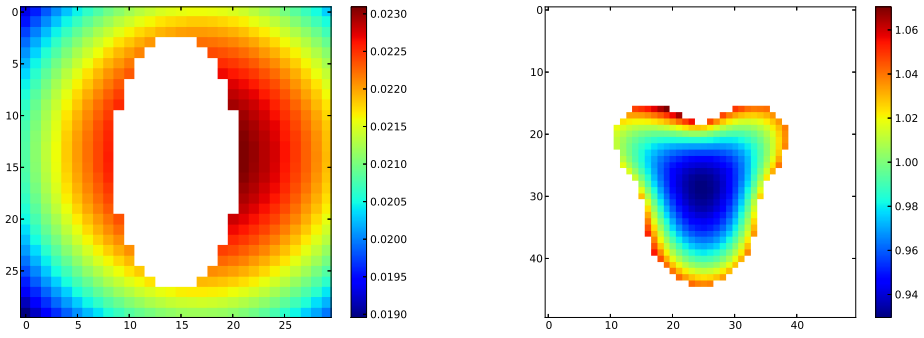

Figure 1.1: Solutions calculated with BEM for the outside of an ellipse and for a star-shaped domain .

heat equation. This has the advantages of being stable for any combination of mesh widths h_t, h_x and of allowing for a straight-forward error analysis. An alternative to Galerkin methods is offered by collocation methods (see [16] and [2]). In collocation methods a suitable set of points is chosen and the equation is required to be satisfied at those points.

The boundary element method (BEM) relies on finding a formulation of the problem (1.1) which is posed on the mantle of the space-time cylinder $\Omega \times (0, T)$. For this we require the fundamental solution of the heat equation, which is

$$G(t,x) = \begin{cases} (4\pi t)^{-d/2} e^{-|x|^2/4t} & t \geq 0 \\ 0 & t < 0. \end{cases} \tag{1.2}$$

Then we can apply Green's second theorem to the problem with either Dirichlet or Neumann boundary conditions. Thus we get the following representation for the solution of the heat equation

$$\begin{aligned} u(x,t) = \int_0^T \int_{\partial\Omega} \left[G(x-y, t-s) \frac{\partial}{\partial n_y} u(y,s) - \frac{\partial}{\partial n} G(x-y, t-s) u(y,s) \right] dyds \\ + \int_0^T \int_{\Omega} G(x-y, t-s) f(y,t) dydt, \end{aligned} \tag{1.3}$$

where n_y is outward unit normal to $\partial\Omega$. The boundary element method then consists of finding either $\frac{\partial}{\partial n} u|_{\partial\Omega}$ for the Dirichlet problem or $u|_{\partial\Omega}$ for the Neumann problem. This means we only need to solve a problem on the boundary of the domain, lowering

the dimension of the problem.

The BEM formulation of the heat equation becomes coercive after the boundary reduction. This means that the method is stable for all choices of mesh size versus time steps, and allows for more flexibility. In particular, for a problem with an inhomogeneous source term which does not vary significantly in time, a small number of time steps may be sufficient and allows for much faster solving.

To compare use of FEM and BEM for solving the heat equation we compare their relative complexity. Complexity is a measure of the number of single operations (FLOPs) needed to complete a computation. The complexity of these methods depends strongly on the complexity of the solution of the resulting linear system. Linear complexity for the solution of the linear system is attainable for FEM since that formulation results in sparse matrices. However, the BEM formulation generally results in densely populated matrices. We will resolve this issue by using a wavelet basis. This leads to a numerically sparse matrix and the corresponding linear system can be solved with linear complexity as required.

Typically, FEM combined with a low-order time-stepping scheme give a complexity of

$$O(h_t^{-1} h_x^{-d}),$$

where the spatial dimension is given by d, h_x is the mesh width in space and h_t is the time step size. According to [46] if one allows increasing the polynomial degree in time along with a mesh refinement in the temporal dimension, i.e. with hp-FEM the complexity can be reduced to

$$O(h_x^{-d} |\log h_x|^2).$$

In [47] space-time compressive, adaptive Galerkin methods are used to further reduce the complexity to

$$O(h_x^{-d}).$$

The contribution of this work is to further decrease the complexity of these methods. This leads both to speed gains (in CPU time) as well as requiring smaller amounts of memory to solve the same problem. To do this we will combine the complexity gains of boundary reduction by integral equation formulations with a sparse tensor space-time discretisation. This reduces the total work to

$$O(h_x^{-(d-1)})$$

when the solution of the linear system is performed with linear complexity.

1.2 Motivation

The boundary integral operators of the heat equation have very similar properties to the operators in the elliptic case. More precisely, it has been shown in [15] that these operators are coercive and continuous in the appropriate anisotropic Sobolev spaces. This means that unlike in the case of the domain heat operator we can assure stability for any conforming Galerkin discretisation using the classical Lemma of Lax-Milgram and Lemma of Céa.

The first step to achieving the required complexity gains is finding a way to solve the linear system in linear complexity. This is in general not possible for densely populated matrices such as those given by the boundary integral operators since they are non-local. However, we can obtain numerically sparse matrices by using a wavelet basis. Wavelet bases (see e.g. [19], [13], and [45]) were initially used for signal analysis (sound, images). There are also numerous other applications in numerical analysis.

Most research into using wavelet bases for BEM has been done for the elliptic case (see e.g. [34]). There has also been some work on using wavelets for the heat equation in two dimensions in [8]. As in elliptic problems it will be possible to compress the resulting matrix by setting small entries to zero. One of the main results of this work is proving that a matrix compression results in no loss of accuracy for this problem. We will also discuss some alternative types of wavelet basis.

When trying to get sparse matrices one alternative to wavelets is to use panel clustering methods. For example, in [40] fast multipole methods are used in space and time. In the near-field they use numerical quadrature to calculate the time-integrals which leaves them with a smooth kernel in space.

Another alternative is adaptive cross approximation in which one uses rough approximations for the far-field and precise calculations only in the near-field (see e.g. [5], [6] and [7]).

The second step is improving the approximation properties of the method itself, i.e. improving the convergence rates. In order to improve the expected convergence results, sparse grid techniques (see e.g. [32], [28]) can be used. It has been shown theoretically in [12] and [11] that this approach does indeed improve convergence.

In this work we will show that the convergence can be further improved for some choices of polynomial degrees by using more general sparse grid spaces. We use the combination technique (see e.g. [33],[22]) to implement this and verify the theoretical bounds.

1.3 Chapter Overview

In Chapter 2 we introduce the non-stationary heat equation and outline the boundary reduction. This chapter also contains some well-known theoretical results on the heat equation, they will be used throughout.

In Chapter 3 we introduce several concepts related to wavelet basis functions. We discuss multiresolution analysis and the construction of biorthogonal wavelets. We also give many explicit examples of wavelet bases as they are used in this work.

Chapter 4 shows the discretisation of the integral equations and discusses some implementational issues, such as matrix structure and quadrature rules. It also contains a comparison between FEM and BEM.

In Chapter 5 we summarise several known results on the convergence rates of full tensor product BEM for the heat equation. Then we show new estimates in the energy norm which lead to improved convergence rates. Finally we show numerical results to verify these estimates.

In Chapter 6 we introduce sparse grid spaces with several choices of index set. We show a known proof for the convergence rates of standard sparse grids and verify these rates numerically. Further, we introduce an optimised sparse grid index set and prove new results for the convergence rates of these sets.

In Chapter 7 we prove that when using a wavelet basis a matrix compression reduces the number of non-zero matrix entries to $O(N)$ and does not lead to a loss of accuracy in the scheme. Numerical results are also given.

In Chapter 8 we conclude with a summary and a discussion of future work.

Chapter 2

The Heat Equation

In this chapter we introduce the boundary integral formulation of the heat equation. The results of this chapter are well known and can also be found in [15] and [42].

We start out by giving a problem formulation on domains $\Omega \subset \mathbb{R}^d$. The appropriate function spaces for this formulation are not the well known Sobolev spaces H^r, but rather the anisotropic Sobolev spaces $H^{r,s}$. We introduce these function spaces in Section 2.1.3.

Then we summarise the reduction of the problem to the boundary. Then we show some properties of the boundary integral operators. Notably, even though the heat equation is a parabolic differential equation, the associated boundary integral operators have similar properties to those of elliptic operators. Finally, we give some regularity results for the solution of the problem.

2.1 Problem Formulation

Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain with boundary $\Gamma := \partial\Omega$. For simplicity we will later restrict ourselves to smooth domains. However, all theoretical results in this chapter hold for general Lipschitz domains.

Further, let n be the outer normal vector field of Γ. We assume that it exists almost everywhere on the boundary Γ.

With $T > 0$ we denote a finite time horizon and with $\mathcal{I} := (0, T)$ the time interval of interest.

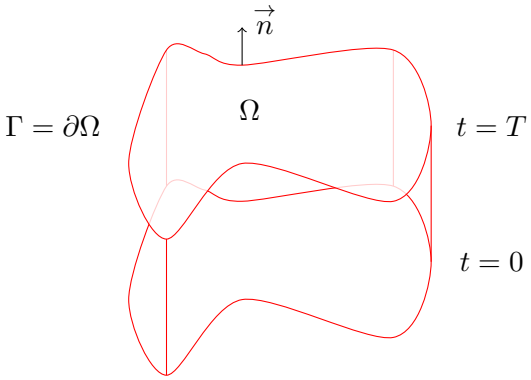

Figure 2.1: The domain Q for $\Omega \subset \mathbb{R}^2$.

Then we set $Q := \mathcal{I} \times \Omega$ the space-time cylinder. The domain heat equation is defined on Q. However, after the boundary reduction we will mainly work with the mantle of the space-time cylinder $\Sigma = \mathcal{I} \times \Gamma$.

In Q we consider a linear nonstationary heat equation with Dirichlet or Neumann boundary conditions.

2.1.1 Trace Operators

To formulate the heat equation we first introduce two types of trace operators for sufficiently smooth functions w.

Definition 2.1.1. *We denote the trace operator by γ_0, so*

$$\gamma_0 w = w|_\Sigma, \tag{2.1}$$

is the function w restricted to the mantle of the space-time cylinder.

Definition 2.1.2. *We denote by γ_1 the conormal derivative of a function, so*

$$\gamma_1 w = \partial_n w = (\nabla w|_\Sigma) \cdot n, \tag{2.2}$$

is the normal derivative of a function w restricted to the mantle of the space-time cylinder.

After the relevant function spaces have been introduced we will show continuity results for these trace operators.

2.1.2 Formulation of the Domain Heat Equation

The heat equation describes heat diffusion through a given region over time. In order to give a full description of a heat diffusion problem we need to supplement the heat equation

$$(\partial_t - \Delta)u = f, \qquad \text{in } Q$$

with a combination of initial and boundary values. For simplicity we always assume that the initial conditions are zero. This means that we prescribe

$$u = 0, \qquad \text{at } \{t = 0\} \times \Omega$$

in the entire domain. Further we need to prescribe values on the boundary (Dirichlet problem) or the boundary heat flux (Neumann problem).

Thus, the Dirichlet and Neumann problems are formulated as follows.

Definition 2.1.3 (Dirichlet Problem)**.** *Given $g : \Sigma \to \mathbb{R}$ and $f : Q \to \mathbb{R}$, find $u : Q \to \mathbb{R}$ satisfying:*

$$\begin{aligned} (\partial_t - \Delta)u &= f, && \text{in } Q \\ u &= 0, && \text{at } \{t = 0\} \times \Omega \\ \gamma_0 u &= g, && \text{in } \Sigma. \end{aligned} \tag{2.3}$$

Definition 2.1.4 (Neumann Problem)**.** *Given $h : \Sigma \to \mathbb{R}$ and $f : Q \to \mathbb{R}$, find $u : Q \to \mathbb{R}$ satisfying:*

$$\begin{aligned} (\partial_t - \Delta)u &= f, && \text{in } Q \\ u &= 0, && \text{at } \{t = 0\} \times \Omega \\ \gamma_1 u &= h, && \text{in } \Sigma. \end{aligned} \tag{2.4}$$

Remark 2.1.5. *It is possible to pose the heat equation with other types of boundary conditions, such as Robin-type boundary conditions*

$$a\gamma_0 u + b\gamma_1 u = c.$$

Newton's law of cooling states that the boundary heat flux is proportional to the temperature difference between the domain Ω and the surrounding environment $\mathbb{R}^d \backslash \Omega$. This makes Robin boundary conditions the natural formulation to model this.

2.1.3 Function spaces

A variety of function spaces are needed in the course of this work. For example, in order to give the solvability and uniqueness results for the Neumann and Dirichlet problems above we will require certain anisotropic Sobolev spaces.

Thus, we start this section by introducing L^2 spaces and the standard Sobolev spaces H^r. Then we define the anisotropic spaces $H^{r,s}$ and $H^{r,s}_{\text{mix}}$. The mix-spaces will be useful in the error analysis of the sparse grid spaces in Chapter 6.

The Sobolev spaces needed for this work are constructed using the function spaces $L^2(\Sigma)$.

Definition 2.1.6. *We denote the $L^2(\Sigma)$ inner product as follows*

$$\langle u, v\rangle := \int_\Gamma \int_0^T u(x,t)v(x,t)dtdx.$$

Thus, we have a norm defined as $\|u\|_{L^2(\Sigma)} = \sqrt{\langle u, u\rangle}$ and we can define the space of square integrable functions

$$L^2(\Sigma) = \{u \ : \ \|u\|_{L^2(\Sigma)} < \infty\}$$

For simplicity we will start by defining isotropic Sobolev spaces. We will then introduce two types of anisotropic Sobolev space.

Note that we denote multi-indices (i.e. sequences of natural numbers) by $\mathbf{k} = (k_1, ..., k_d) \in \mathbb{N}^d$. Further, we write the 1-norm of these vectors as $|\mathbf{k}| := \sum_{i=1}^d k_i$.

Definition 2.1.7 (weak derivative)**.** *Let $U \subset \mathbb{R}^d$ be an open set. We say v is the $\boldsymbol{k}$-th weak derivative of the function u if*

$$\int_U uD^{\boldsymbol{k}}\varphi = (-1)^{|\boldsymbol{k}|}\int_U v\varphi, \quad \forall\varphi \in C_0^\infty(U), \quad \text{where } D^{\boldsymbol{k}}\varphi = \frac{\partial^{|\boldsymbol{k}|}}{\partial^{k_1}\dots\partial^{k_n}}\varphi$$

where $C_0^\infty(U)$ is the space of infinitely differentiable functions with compact support in U. We denote the weak derivative v by $D^{\boldsymbol{k}}u$.

Whole-numbered Sobolev spaces can be understood as spaces of L^2-functions with weak derivatives.

Definition 2.1.8. *Let $s \in \mathbb{N}$ and $U \subset \mathbb{R}^d$ an open set, then*

$$H^s(U) = \{u \in L^2(U) \ : \ \sum_{0\le|\boldsymbol{k}|\le s} \|D^{\boldsymbol{k}}u\|^2_{L^2(U)} < \infty\},$$

where $D^{\boldsymbol{k}}u$ is the weak derivative of u.

There are a variety of ways to define Sobolev spaces with real-valued regularity exponents. We define them directly using Sobolev-Slobodeckij semi-norms. Alternatively, they can be understood as interpolation spaces of the whole-numbered Sobolev spaces or they can be defined via Fourier transforms.

Definition 2.1.9. *For an open subset $U \subset \mathbb{R}^d$, for $\theta \in (0,1)$ and for $f \in L^2(U)$, the Slobodeckij semi-norm is defined by*

$$|f|^2_{H^\theta(U)} := \int_U \int_U \frac{|f(x) - f(y)|^2}{|x-y|^{2\theta+d}}\, dx\, dy.$$

Let $s > 0$ be a non-integer and set $\theta = s - \lfloor s \rfloor \in (0,1)$. Then

$$H^s(U) := \left\{ f \in H^{\lfloor s \rfloor}(\Omega) \ : \ \sup_{|\boldsymbol{k}|=\lfloor s \rfloor} |D^{\boldsymbol{k}} f|_{H^\theta(U)} < \infty \right\}.$$

Next we introduce the $H^{r,s}(\Sigma)$ and $H^t_{\text{mix}}(\Sigma)$ spaces, more general spaces than the standard isotropic Sobolev spaces defined above.

Definition 2.1.10. *Let $r, s > 0$. Then the anisotropic Sobolev spaces $H^{r,s}(\Sigma)$ and $\tilde{H}^{r,s}(\Sigma)$ are given by*

$$H^{r,s}(\Sigma) := L^2(\mathcal{I}, H^r(\Gamma)) \cap H^s(\mathcal{I}, L^2(\Gamma)) \tag{2.5}$$

We can restrict ourselves to spaces which have zero initial conditions,

$$\tilde{H}^{r,s}(\Sigma) := \{ u \in H^{r,s}((-\infty, T) \times \Gamma) \ : \ u(t,x) = 0, \ t < 0 \}. \tag{2.6}$$

Both types of anisotropic spaces can be equipped with a simple graph norm

$$\|u\|_{H^{r,s}(\Sigma)} = \|u\|_{L^2(\mathcal{I}, H^r(\Gamma))} + \|u\|_{H^s(\mathcal{I}, L^2(\Gamma))}.$$

Using the dual space we can define $H^{-r,-s} = (H^{r,s})'$.

Next we introduce the so called mix-spaces. Let $\Omega_i \subset \mathbb{R}^{d_i}$ for $1 \le i \le n$. We define

$$H^{\mathbf{k}}_{\text{mix}}(\Omega_1 \times ... \times \Omega_n) := H^{k_1}(\Omega_1) \otimes \cdots \otimes H^{k_n}(\Omega_n).$$

Further, for ease of notation we will denote

$$H^{t,l}_{\text{mix}}(\Omega_1 \times \Omega_2) := H^t(\Omega_1) \otimes H^l(\Omega_2).$$

For $t, l < 0$, $H^{t,l}_{\text{mix}}$ is defined as the dual of $H^{-t,-l}_{\text{mix}}$, i.e. we set

$$H^{t,l}_{\text{mix}} := (H^{-t,-l}_{\text{mix}})'.$$

These are spaces of dominating mixed derivative.

The following relation holds between the isotropic Sobolev spaces and these mix-spaces:

$$H^{t,l}_{\text{mix}}(\Omega_1 \times \Omega_2) \subset H^{t,l}(\Omega_1 \times \Omega_2).$$

Further, the following embeddings hold

Lemma 2.1.11. *Let $\Omega_1 \subset \mathbb{R}^{d_1}$, $\Omega_2 \subset \mathbb{R}^{d_2}$. Further, let $a, b, k \geq 0$ and $k \geq a + 2b$, then there holds*

$$H^{k,\frac{k}{2}}(\Omega_1 \times \Omega_2) \subset H^{a,b}_{mix}(\Omega_1 \times \Omega_2).$$

Proof. See Lemma 5.2, [12]. □

2.1.4 Uniqueness and Solvability

The existence and uniqueness of solutions to the domain heat equation depends on properties of the trace operators. These properties are also needed to show the regularity results at the end of this chapter.

Lemma 2.1.12. *The trace operator γ_0 is continuous and surjective as a mapping*

$$\tilde{H}^{1,\frac{1}{2}}(Q) \to H^{\frac{1}{2},\frac{1}{4}}(\Sigma).$$

Proof. See Lemma 2.4 in [15]. □

Lemma 2.1.13. *For $s \in (-\frac{1}{2}, \frac{1}{2})$ the conormal derivative is continuous as a mapping*

$$\{v \in \tilde{H}^{1+s,\frac{1+s}{2}}(Q) \ : \ (\partial_t - \Delta)v \in L^2(Q)\} \to H^{-\frac{1}{2}+s,(-\frac{1}{2}+s)/2}(\Sigma).$$

Proof. See Corollary 4.14 in [15]. □

The well-posedness and solvability of the Neumann and Dirichlet problems (2.3) and (2.4) are well known.

Lemma 2.1.14. *For every $f \in \tilde{H}^{-1,-1/2}(Q)$ and $g \in H^{1/2,1/4}(\Sigma)$ there exists a unique $u \in \tilde{H}^{1,1/2}(Q)$ satisfying (2.3).*

Proof. See Theorem 2.9, [15]. □

Lemma 2.1.15. *For every $f \in L^2(Q)$ and $h \in L^2(I, H^{-1/2}(\Gamma))$ there exists a unique $u \in \tilde{H}^{1,1/2}(Q)$ satisfying (2.4).*

Proof. See Lemma 2.21, [15]. □

2.2 Boundary Reduction

We now want to transform the boundary value problems (2.3) and (2.4) into integral equations on the boundary Γ. For this we require a version of Green's Theorem.

Theorem 2.2.1. *Let $u \in \tilde{H}^{1,\frac{1}{2}}(Q)$ with $(\partial_t - \Delta)u \in L^2(Q)$ and $v \in \tilde{H}^{1,\frac{1}{2}}(Q)$. Then for any $t_0 \in \mathbb{R}$ there holds Green's first formula:*

$$\int_Q \nabla u(x,t)\cdot\nabla v(x,t_0-t)dxdt + \int_Q \partial_t u(x,t)v(x,t_0-t)dxdt \\ = \int_\Sigma \gamma_1 u(t,x)\gamma_0 v(x,t_0-t)dxdt + \int_Q (\partial_t - \Delta)u(x,t)v(x,t-t_0)dxdt$$

If additionally $(\partial_t - \Delta)v \in L^2(Q)$, then there holds Green's second formula

$$\int_Q (\partial_t - \Delta)u(x,t)v(x,t_0-t) - u(x,t_0-t)(\partial_t - \Delta)v(x,t)dxdt \\ = \int_\Sigma \gamma_0 u(x,t)\cdot\gamma_1 v(x,t_0-t)dxdt - \int_\Sigma \gamma_1 u(x,t)\cdot\gamma_0 v(x,t_0-t)dxdt$$

Proof. See Proposition 2.19 in [15]. □

The fundamental solution of the heat equation is

$$G(x,t) = \begin{cases} (4\pi t)^{-d/2} e^{-|x|^2/4t} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{2.7}$$

for any dimension $d > 1$.

Applying the second Green's theorem to the Dirichlet problem (2.3) or the Neumann problem (2.4) with the choice $v(x,t) = G(x,t)$ yields that the solution $u \in \tilde{H}^{1,\frac{1}{2}}(Q)$ admits the representation:

$$u(x,t) = \int_\Sigma \left[G(x-y,t-s)\frac{\partial u}{\partial n_y}(y,s) - \frac{\partial}{\partial n_y}G(x-y,t-s)u(y,s)\right] dyds \\ + \int_Q G(x-y,t)f(y,s)dyds, \tag{2.8}$$

However, for simplicity we will assume $f = 0$ in the following.

With this simple representation of the solution it becomes natural to define the following two operators.

Definition 2.2.2. *The single layer heat potential is defined as*

$$\mathcal{K}_0(\varphi)(x,t) := \int_\Sigma \varphi(y,s)G(x-y,t-s)dyds \qquad (x,t) \in Q$$

Definition 2.2.3. *The double layer heat potential is defined as*

$$\mathcal{K}_1(\psi)(x,t) := \int_\Sigma \psi(y,s)\frac{\partial}{\partial n_y}G(x-y,t-s)dyds \qquad (x,t) \in Q.$$

This means that using the trace operators defined in Section 2.1.1 we can rewrite the equation (2.8) as:

$$u = \mathcal{K}_0(\gamma_1 u) - \mathcal{K}_1(\gamma_0 u), \qquad \text{in Q}, \tag{2.9}$$

we call this the representation formula.

Next we restrict the heat potential operators to the mantle of the space-time cylinder using a trace operator. This simplifies the notation. We refer to the resulting operators as boundary integral operators.

Let $\varphi \in H^{\frac{1}{2},\frac{1}{4}}(\Sigma)$ and $\psi \in H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)$.

Definition 2.2.4. *The single layer operator V is defined as*

$$V\psi := \gamma_0 \mathcal{K}_0 \psi. \tag{2.10}$$

Further, the hypersingular operator W is defined as

$$W\varphi := -\gamma_1 \mathcal{K}_1 \varphi \tag{2.11}$$

and the double layer operator K is defined as

$$K\varphi := \gamma_0 \left(\mathcal{K}_1 \varphi\right)|_Q + \frac{1}{2}\varphi. \tag{2.12}$$

Lastly, the related operator N is defined as

$$N\psi := \gamma_1(\mathcal{K}_0\psi)|_Q - \frac{1}{2}\psi. \tag{2.13}$$

In order to make use of these operators, we need to know more about them. These operators have been well studied in [15], from which we take this theory.

As in the elliptic case, there hold certain jump relations on Σ. These relations ensure that the above operators V, W, K and N are indeed well defined.

Definition 2.2.5. *Let B_R be a unit ball in $\mathbb{R}^d$ large enough to contain Ω. Further, let $Q^c = \mathcal{I} \times B_R \backslash \Omega$ and let $u \in \tilde{H}^{1,\frac{1}{2}}(\mathcal{I} \times B_R)$. Then the jumps across the boundary Γ are defined as*

$$[\gamma_0 u] = \gamma_0(u|_{Q^c}) - \gamma_0(u|_Q)$$

and

$$[\gamma_1 u] = \gamma_1(u|_{Q^c}) - \gamma_1(u|_Q).$$

These definitions are independent of the choice of R.

Theorem 2.2.6. *For all $\psi \in H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)$ and all $\varphi \in H^{\frac{1}{2},\frac{1}{4}}(\Sigma)$ there hold the jump relations:*

$$[\gamma_0 \mathcal{K}_0 \psi] = 0, \quad [\gamma_1 \mathcal{K}_0 \psi] = -\psi, \quad [\gamma_0 \mathcal{K}_1 \varphi] = \varphi, \quad [\gamma_1 \mathcal{K}_1 \varphi] = 0.$$

Proof. See Theorem 4.3 in [15]. □

Further, if Γ is sufficiently smooth all the integral operators used in the methods above are one-to-one mappings.

Theorem 2.2.7. *Assume that $\Gamma \in C^\infty(\Sigma)$. Then for any $s \geq 0$ the mappings*

$$\begin{aligned} V &: \tilde{H}^{s-\frac{1}{2},(s-\frac{1}{2})/2}(\Sigma) \to \tilde{H}^{s+\frac{1}{2},(s+\frac{1}{2})/2}(\Sigma) \\ \left(\frac{1}{2}I + K\right), \left(\frac{1}{2}I - N\right) &: \tilde{H}^{s+\frac{1}{2},(s+\frac{1}{2})/2}(\Sigma) \to \tilde{H}^{s+\frac{1}{2},(s+\frac{1}{2})/2}(\Sigma) \\ W &: \tilde{H}^{s+\frac{1}{2},(s+\frac{1}{2})/2}(\Sigma) \to \tilde{H}^{s-\frac{1}{2},(s-\frac{1}{2})/2}(\Sigma) \end{aligned}$$

are isomorphisms.

Proof. See Theorem 4.3 in [15]. □

This provides the basis for the analysis of Galerkin methods for these operators. Further, we can show that V and W are positive and define isomorphisms in anisotropic trace spaces.

Theorem 2.2.8. *The single layer operator* $V : H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma) \to H^{\frac{1}{2},\frac{1}{4}}(\Sigma)$ *is an isomorphism and*

$$\exists c > 0 \ : \ \langle \psi, V\psi \rangle \geq c \|\psi\|^2_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \qquad \forall \psi \in H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma).$$

The hypersingular operator $W : H^{\frac{1}{2},\frac{1}{4}}(\Sigma) \to H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)$ *is an isomorphism and*

$$\exists c > 0 \ : \ \langle \phi, W\phi \rangle \geq c \|\phi\|^2_{H^{\frac{1}{2},\frac{1}{4}}(\Sigma)} \qquad \forall \phi \in H^{\frac{1}{2},\frac{1}{4}}(\Sigma).$$

Proof. See Corollary 3.13 in [15]. □

Taken together with the continuity results this theorem implies invertibility of the operators V and W. Due to the invertibility and coercivity of the operators we can ensure that any discrete scheme will be stable and have a unique solution. In Chapter 4 we will use these properties to show best approximation properties of the discrete approximation with the Lemma of Céa.

From these properties we can formulate two methods for solving the Dirichlet problem (2.3) and the Neumann problem (2.4).

2.2.1 Direct Method

Using the direct method the boundary integral formulation of the Dirichlet Problem is

1. Find $\psi \in H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)$ such that:

$$V\psi = \left(\frac{1}{2}I + K\right) g. \tag{2.14}$$

2. Then the unique solution $u \in \tilde{H}^{1,\frac{1}{2}}(Q)$ of the Dirichlet problem with $f = 0$ can be represented by

$$u = \mathcal{K}_0 \psi - \mathcal{K}_1 g. \tag{2.15}$$

Using the direct method the boundary integral formulation of the Neumann Problem is

1. Find $\varphi \in H^{\frac{1}{2},\frac{1}{4}}(\Sigma)$ such that:

$$W\varphi = \left(\frac{1}{2}I - N\right) h. \tag{2.16}$$

2. Then the unique solution $u \in \tilde{H}^{1,\frac{1}{2}}(Q)$ of the Neumann problem with $f = 0$ can be represented by

$$u = \mathcal{K}_0 h - \mathcal{K}_1 \varphi. \tag{2.17}$$

In this method $\psi = \gamma_1 u$ is exactly the boundary flux on Σ, so this method is useful if the boundary fluxes are also of interest.

2.2.2 Indirect Method

Using the indirect method the boundary integral formulation of the Dirichlet Problem is

1. Find $\psi \in H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)$ such that:

$$V\psi = g. \tag{2.18}$$

2. Then the unique solution $u \in \tilde{H}^{1,\frac{1}{2}}(Q)$ of the Dirichlet problem with $f = 0$ can be represented by

$$u = \mathcal{K}_0 \psi. \tag{2.19}$$

Using the indirect method the boundary integral formulation of the Neumann Problem is

1. Find $\varphi \in H^{\frac{1}{2},\frac{1}{4}}(\Sigma)$ such that:

$$W\varphi = -h. \tag{2.20}$$

2. Then the unique solution $u \in \tilde{H}^{1,\frac{1}{2}}(Q)$ of the Neumann problem with $f = 0$ can be represented by

$$u = \mathcal{K}_1 \varphi. \tag{2.21}$$

Remark 2.2.9. *This method is simpler to implement than the direct method since the matrix of the double layer operator K and the matrix of the operator N do not need to be evaluated for the Dirichlet and Neumann problem respectively.*

2.3 Regularity

In this section we will summarise some of the regularity results for the Dirichlet and Neumann problems.

Theorem 2.3.1. *The single layer operator V is a continuous map from*

$$V \,:\, H^{-\frac{1}{2}+s,-\frac{1}{4}+\frac{s}{2}}(\Sigma) \to H^{\frac{1}{2}+s,\frac{1}{4}+\frac{s}{2}}(\Sigma),$$

for any $s \in (-\frac{1}{2}, \frac{1}{2})$.

Proof. See Theorem 4.8 in [15]. □

Theorem 2.3.2. *For any $s \in (-\frac{1}{2}, \frac{1}{2})$ the operators*

$$\begin{aligned} W \,:\, H^{\frac{1}{2}+s,\frac{1}{4}+\frac{s}{2}}(\Sigma) &\to H^{-\frac{1}{2}+s,-\frac{1}{4}+\frac{s}{2}}(\Sigma) \\ \frac{1}{2}I + K,\ \frac{1}{2}I - K \,:\, H^{\frac{1}{2}+s,\frac{1}{4}+\frac{s}{2}}(\Sigma) &\to H^{\frac{1}{2}+s,\frac{1}{4}+\frac{s}{2}}(\Sigma) \\ \frac{1}{2}I + N,\ \frac{1}{2}I - N \,:\, H^{-\frac{1}{2}+s,-\frac{1}{4}+\frac{s}{2}}(\Sigma) &\to H^{-\frac{1}{2}+s,-\frac{1}{4}+\frac{s}{2}}(\Sigma) \end{aligned}$$

are continuous.

Proof. See Theorem 4.16 in [15] □

Combining these results we get the following regularity results.

Theorem 2.3.3. *The inverse operators*

$$\begin{aligned} V^{-1} \,:\, \tilde{H}^{1,\frac{1}{2}}(\Sigma) &\to L^2(\Sigma) \\ \left(\frac{1}{2}I + K\right)^{-1},\ \left(\frac{1}{2}I - K\right)^{-1} \,:\, \tilde{H}^{1,\frac{1}{2}}(\Sigma) &\to H^{1,\frac{1}{2}}(\Sigma) \\ \left(\frac{1}{2}I + N\right)^{-1},\ \left(\frac{1}{2}I - N\right)^{-1} \,:\, L^2(\Sigma) &\to L^2(\Sigma) \\ W^{-1} \,:\, L^2(\Sigma) &\to \tilde{H}^{1,\frac{1}{2}}(\Sigma) \end{aligned}$$

are continuous.

Proof. See Theorem 4.18 in [15]. □

Chapter 3

Wavelets

In this chapter we introduce wavelets, in particular biorthogonal wavelets. Wavelets are useful in a many different applications. They are used in pure mathematics for the analysis of harmonic operators. They are also widely used in signal analysis. A general introduction to wavelets can be found in [13] or [19].

In this chapter we start by introducing multiresolution analysis and biorthogonal wavelets. Then we give examples of wavelet bases. These bases will be important throughout this work, mainly in Chapter 7 which introduces a matrix compression based on properties of certain types of biorthogonal wavelets. Further, the norm equivalences that hold for wavelets are used for the proofs in Chapters 5 and 6.

3.1 Notation

In this chapter we assume that the domain Ω is simply connected and that its boundary Γ is smooth. In two dimensions this means that it can be parameterised by a single function

$$\gamma : [0,1] \to \Gamma.$$

Further, we assume that the parameterisation γ is 1-periodic and that the derivative

$$\alpha(t) := \|\gamma'(t)\| > 0 \text{ for all } t \in [0,1].$$

Remark 3.1.1. *In [35] the more general case of a piecewise smooth boundary is discussed.*

Definition 3.1.2. *We denote the* ***characteristic function*** *of an interval by χ, i.e.*

$$\chi_{[a,b]}(x) := \begin{cases} 1, & x \in [a,b] \\ 0, & else. \end{cases}$$

Definition 3.1.3. *A family of functions $\{\varphi_k\}_{k\in\mathbb{Z}}$ is a* ***Riesz basis*** *of the Hilbert space H if it is dense in H and there exist $0 < C_1 \leq C_2$ such that for all finitely supported sequences (x_k), we have*

$$C_1 \sum_k |x_k|^2 \leq \| \sum_k x_k \varphi_k \|_H^2 \leq C_2 \sum_k |x_k|^2.$$

3.2 Multiresolution Analysis

Multiresolution analysis was first formulated in 1986 by Mallat and Meyer (see [39] and [41]). It provides a framework to construct wavelets.

A multiresolution analysis consists of a sequence of nested approximation spaces

$$... \subset V_{-1} \subset V_0 \subset V_1 \subset ... \subset V_j \subset ... \subset L^2(\mathbb{R}). \tag{3.1}$$

Further, the union of these spaces should be dense in $L^2(\mathbb{R})$

$$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}) \tag{3.2}$$

and their intersection should be the null function

$$\bigcap_{j\in\mathbb{Z}} V_j = \{0\}. \tag{3.3}$$

The spaces are related to each other with a dyadic scaling:

$$f(\cdot) \in V_j \Leftrightarrow f(2^j \cdot) \in V_0. \tag{3.4}$$

Finally, we require that there exists a function $\phi \in V_0$ such that

$$\{\phi(\cdot - k) \ : \ k \in \mathbb{Z}\} \text{ forms a Riesz basis of } V_0. \tag{3.5}$$

This means that all spaces are scaled versions of the initial space V_0, so we call this a multiresolution analysis.

Due to (3.5) together with (3.4) we have that

$$\phi_{j,k} = 2^{j/2}\phi(2^j \cdot -k), \quad k \in \mathbb{Z}. \tag{3.6}$$

forms a Riesz basis of V_j.

The function ϕ is referred to as a scaling function since every V_j is generated by scaled versions of ϕ.

Since $V_0 \subset V_1$, we can expand $\phi \in V_0$ in terms of the basis of V_1

$$\phi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k \phi(2x - k).$$

This type of equation is called a refinement equation and the coefficients h_k are called refinement coefficients.

Next we construct a system of pairwise orthogonal subspaces W_j. These spaces are orthogonal with respect to the L^2 inner product. These give a multilevel decomposition of the spaces V_j, i.e. there exist spaces W_j such that

$$V_{j+1} = V_j \oplus W_j, \; W_j \perp W_{\tilde{j}} \;\; \forall j \neq \tilde{j}.$$

Due to (3.2) and (3.3) this implies

$$L^2(\mathbb{R}) = \overline{\bigoplus_{j\in\mathbb{Z}} W_j}.$$

The spaces W_j inherit scaling property (3.4) from the spaces V_j.

Together this means that if we have an orthonormal basis $\{\psi(\cdot - k), \; k \in \mathbb{Z}\}$ of W_0, then $\{\psi_{j,k} = 2^{-j/2}\psi(2^{-j} \cdot -k), \; j,k \in \mathbb{Z}\}$ is a basis of W_j. This means that in order to find an orthonormal wavelet basis of L^2 we only need to find a ψ so that its translations form an orthonormal basis of W_0. We refer to such a ψ as a **mother wavelet** since the entire wavelet basis can be derived from it.

Theorem 3.2.1 (Theorem 5.1.1, [19]). *If a sequence of nested approximation spaces satisfies (3.1) – (3.5), i.e. when we have a multiresolution analysis, there exists an associated wavelet basis* $\{\psi_{j,k}, \; j,k \in \mathbb{Z}\}$, *such that*

$$\Pi_{V_{j+1}} = \Pi_{V_j} + \sum_k \langle \cdot, \psi_{j,k}\rangle \psi_{j,k},$$

where Π_{V_j} is the L^2-orthogonal projection onto V_j. The wavelet ψ can be constructed using the refinement equation

$$\psi = \sum_k (-1)^{k-1} h_{-k-1} \phi_{-1,k}. \tag{3.7}$$

Remark 3.2.2. *This series representation of ψ is not unique.*

Remark 3.2.3. *We can define the basis of the space W_j as $\psi_{jk} = \psi(2^j x - k)$ using this representation.*

Next we will give an example of using the refinement coefficients to construct a mother wavelet for the case of piecewise constant basis functions.

3.2.1 Example: Haar Wavelet

The simplest example of such a multiresolution analysis uses piecewise constant functions and is called the Haar multiresolution analysis. It is associated with the Haar wavelet. Since we will later use wavelets only on finite intervals we give the Haar multiresolution analysis on the interval $[0,1]$ instead of $\mathbb{R}$.

For $j \in \mathbb{N}$ and $k \in \{1, ..., 2^j\}$ consider the decomposition $\tau_k^j = [(k-1)2^{-j}, k2^{-j}]$ of the interval $[0,1]$. This decomposition has an associated space of piecewise constant basis functions

$$V_j = \{f \in L^2 \ : \ f \text{ is constant on } \tau_k^j, \ k \in \{1, ..., 2^j\}\}.$$

The scaling function of these spaces is the box function

$$\phi(x) = \chi_{[0,1]}(x).$$

Thus, these spaces are spanned by scaled versions of ϕ, i.e.

$$V_j = \text{span}\{\phi(2^j \cdot -k)\}_{k=1}^{2^j}.$$

By construction the the inclusions

$$V_0 \subset ... \subset V_L \subset ... \subset L^2([0,1])$$

hold. We now construct the subspaces W_j as the $L^2([0,1])$-orthogonal complements of V_{j-1} in V_j.

$$W_0 = V_0, \ W_j \oplus V_{j-1} = V_j,$$

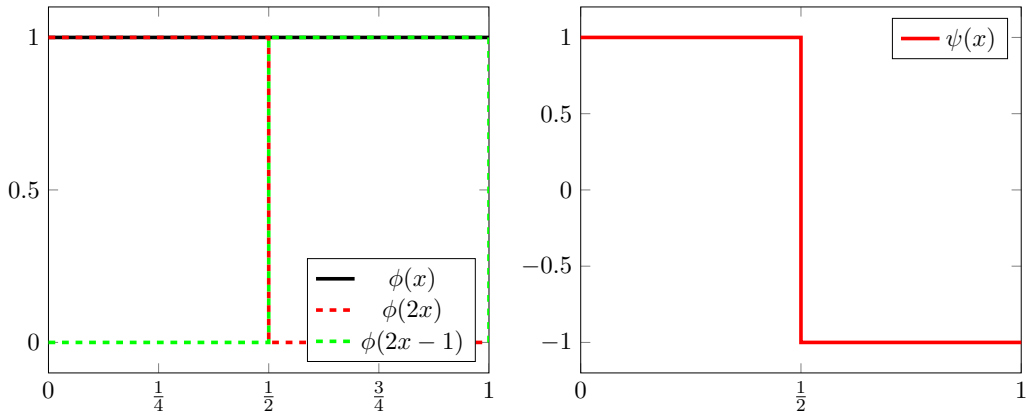

Figure 3.1: The box function $\phi(x)$ and the components of the refinement equation (left) and the Haar wavelet (right).

where $\dim W_0 = 1,\ \dim W_j = 2^{j-1},\ j > 0$.

Clearly the box function ϕ satisfies the following refinement equation

$$\phi(x) = \phi(2x) + \phi(2x-1).$$

Using Theorem 3.2.1 we get the following representation of the Haar mother wavelet

$$\psi = 2^{-1/2}(\phi_{-1,0} - \phi_{-1,1}) = \begin{cases} 1 & \text{in } [0, \frac{1}{2}) \\ -1 & \text{in } [\frac{1}{2}, 1) \\ 0 & \text{else.} \end{cases}$$

Figure 3.1 depicts the refinement equation and the resulting Haar wavelet. This means that the basis of the space W_j is

$$\{\psi_{j,k} := 2^{(j-1)/2}\psi(2^j \cdot -k),\ k = 1, ..., 2^j\}.$$

Since the length of the support of $\psi_{j,k}$ is 2^{1-j}, the above basis functions are normalised, i.e. $\|\psi\|_{L^2([0,1])} = 1$. Further, the orthogonality relations hold by definition.

Using these orthogonality conditions we can derive one moment condition for these wavelets,

$$\int_0^1 \psi_{j,k}(x)dx = \int_0^1 \psi_{j,k}(x)\psi_{0,k}(x)dx = 0.$$

3.3 Biorthogonal Multiresolution Analysis

Instead of using the multiresolution analysis from the previous section we can define a so called biorthogonal multiresolution analysis.

For the biorthogonal multiresolution analysis we require two scaling functions $\phi, \tilde{\phi}$. In turn they generate two different multiresolution analyses, and two different wavelet functions $\psi, \tilde{\psi}$, the wavelet and the dual wavelet. We call these sequences dual multiresolution sequences.

Definition 3.3.1 (dual pair)**.** *We say that two refinable functions θ, $\tilde{\theta}$ form a dual pair if*

$$\langle \theta, \tilde{\theta}(\cdot - k) \rangle = \delta_{0,k}, \quad k \in \mathbb{Z}.$$

Using biorthogonal wavelets gives the necessary freedom to construct basis functions which are symmetric around 0 or $\frac{1}{2}$ and to choose the number of vanishing moments and the degree of polynomial exactness seperately. This is necessary to ensure that the matrices of the integral operators can be compressed to sparse matrices [14].

We start with two hierarchies of approximation spaces

$$\begin{aligned} &... \subset V_0 \subset ... \subset V_j \subset ... \subset L^2(\mathbb{R}) \\ &... \subset \tilde{V}_0 \subset ... \subset \tilde{V}_j \subset ... \subset L^2(\mathbb{R}). \end{aligned}$$

Now we define the complement spaces W_j to V_j in V_{j+1}. The new construction is chosen so that we have orthogonality between W_j and $\tilde{W}_{\tilde{j}}$ for $j \neq \tilde{j}$, instead of between W_j and $W_{\tilde{j}}$ for $j \neq \tilde{j}$, as in the previous construction. This means that it is no longer clear that the basis functions of W_j form a Riesz-basis.

We need to use the dual hierarchy to ensure this, so we also find complement spaces $\tilde{W}_j$ to $\tilde{V}_j$ in $\tilde{V}_{j+1}$. The construction is so that

$$\tilde{W}_j \perp V_j \text{ and } W_j \perp \tilde{V}_j,$$

and thus,

$$W_j \perp \tilde{W}_{\tilde{j}}, \qquad \text{for } j \neq \tilde{j}.$$

This allows us to prove that the bases are indeed Riesz bases. To give this result we first define Fourier transforms and frames.

Definition 3.3.2 (Fourier Transform). *We denote by $\hat{\psi}$ the Fourier transform of ψ:*

$$\hat{\psi}(\xi) := (2\pi)^{-\frac{1}{2}} \int_{\mathbb{R}} e^{-i\xi x} \psi(x) dx.$$

Definition 3.3.3 (Frame). *Let $f \in L^2(\mathbb{R})$. Then we call $(u_n)_n$ a frame if there exist $c_1 > 0$ and $c_2 < \infty$ so that*

$$c_1 \|f\|^2_{L^2(\mathbb{R})} \leq \sum_n |\langle f, u_n \rangle|^2 \leq c_2 \|f\|^2_{L^2(\mathbb{R})}.$$

Given a frame $(u_n)_n$, we call a second frame $(v_n)_n$ a dual frame if

$$\langle u_n, v_{n-k} \rangle = \delta_{0,k}, \qquad \forall n, k.$$

Theorem 3.3.4 (Theorem 3.8, [14]). *Let h_n and $\tilde{h}_n$ be two real sequences with*

$$\sum_{n \in \mathbb{Z}} h_n \tilde{h}_{n+2k} = \delta_{k,0}.$$

Define the single scale functions ϕ and $\tilde{\phi}$ using h_n and $\tilde{h}_n$ as refinement sequences as follows

$$m_0(\xi) = 2^{-\frac{1}{2}} \sum_n h_n e^{-in\xi}, \qquad \tilde{m}_0(\xi) = 2^{-\frac{1}{2}} \sum_n \tilde{h}_n e^{-in\xi}$$
$$\hat{\phi}(\xi) = (2\pi)^{-\frac{1}{2}} \prod_{j=1}^{\infty} m_0(2^{-j}\xi), \ \hat{\tilde{\phi}}(\xi) = (2\pi)^{-\frac{1}{2}} \prod_{j=1}^{\infty} \tilde{m}_0(2^{-j}\xi).$$

Further, assume that their Fourier transforms decay sufficiently rapidly, more precisely, for some $c, \epsilon > 0$

$$|\hat{\phi}(\xi)| \leq c(1 + |\xi|)^{-\frac{1}{2}-\epsilon}, \ |\hat{\tilde{\phi}}(\xi)| \leq c(1 + |\xi|)^{-\frac{1}{2}-\epsilon}.$$

Then we define ψ and $\tilde{\psi}$ as

$$\psi = 2^{j/2} \sum_n (-1)^n \tilde{h}_{-n+1} \phi(2 \cdot + n)$$
$$\tilde{\psi} = 2^{j/2} \sum_n (-1)^n h_{-n+1} \tilde{\phi}(2 \cdot + n).$$

Then $\psi_{j,k} = 2^{j/2} \psi(2^{-j} \cdot - k)$ constitute a frame in $L^2(\mathbb{R})$. Further, $\tilde{\psi}_{j,k} = 2^{j/2} \tilde{\psi}(2^{-j} \cdot$

$-k$) *constitute a dual frame and there holds*

$$f = \sum_{j,k} \langle f, \tilde{\psi}_{jk} \rangle \psi_{jk} = \sum_{j,k} \langle f, \psi_{jk} \rangle \tilde{\psi}_{jk}, \quad \forall f \in L^2(\mathbb{R}).$$

where the series converges strongly in $L^2(\mathbb{R})$.

Further, if ϕ *and* $\tilde{\phi}$ *fulfill*

$$\langle \phi, \tilde{\phi}(\cdot - k) \rangle = \delta_{k,0},$$

the wavelets $\psi_{j,k}$ *and* $\tilde{\psi}_{j,k}$ *are dual Riesz bases with*

$$\langle \psi_{j,k}, \tilde{\psi}_{j',k'} \rangle = \delta_{j,j'} \delta_{k,k'},$$

i.e. they are biorthogonal.

When we use wavelets we will often need the following norm equivalences. For this type of wavelet basis the Jackson and Bernstein inequalites hold [19]. That means we can use an estimate of the form

$$\inf_{u_j \in V_j} \|u - u_j\|_{L^2} \leq c 2^{-jm} \|u\|_{H^m}, \quad \forall u \in H^m$$

for some $m \in \mathbb{N}$. Further, there holds an inverse estimate of the form

$$\|u_j\|_{H^r} \leq c 2^{jq} \|u_j\|_{L^2}, \quad \forall u_j \in V_j,$$

for $q < r$ with $r \in (0, m]$. When we have these two estimates the following norm equivalences hold.

Theorem 3.3.5 (Theorem 3.3, [31]). *Let* $u \in H^t$, $u = \sum_{\boldsymbol{j}=(j_1,\ldots,j_k)} w_{\boldsymbol{j}}$ *for* $w_{\boldsymbol{j}} \in W_{j_1} \otimes \ldots \otimes W_{j_k}$. *Then*

$$\|u\|_{H^t}^2 \sim \sum_{\boldsymbol{j}} 2^{2t \max\{j_1,\ldots,j_k\}} \|w_{\boldsymbol{j}}\|_{L^2}^2, \tag{3.8}$$

for $t \in (-\tilde{r}, r)$ *where* r *and* $\tilde{r}$ *is the number of vanishing moments of the wavelets and the dual wavelets respectively.*

Remark 3.3.6. *These norm equivalences can easily be extended to anisotropic spaces. Let* $\Omega_1 \subset \mathbb{R}^{d_1}$, $\Omega_1 \subset \mathbb{R}^{d_1}$ *and let* $u \in H^{r,s}(\Omega_1 \times \Omega_2)$ *with* $u = \sum_{(i,j)\geq 0} w_{i,j}$ *for* $w_{i,j} \in W_i \otimes W_j$, *then*

$$\|u\|_{H^{r,s}(\Omega_1 \times \Omega_2)}^2 \sim \sum_{(i,j)\geq 0} 2^{2 \max\{ri,sj\}} \|w_{i,j}\|_{L^2(\Omega_1 \times \Omega_2)}^2.$$

Further, if $u \in H^{r,s}_{mix}(\Omega_1 \times \Omega_2)$, *then*

$$\|u\|^2_{H^{r,s}_{mix}(\Omega_1\times\Omega_2)} \sim \sum_{(i,j)\geq 0} 2^{2(ri+sj)} \|w_{i,j}\|^2_{L^2(\Omega_1\times\Omega_2)}.$$

Remark 3.3.7. *Analogous equivalences hold for the dual wavelet.*

In the following two sections we give two examples of the construction of biorthogonal wavelet basis functions.

3.3.1 Example: Wavelet with 3 Vanishing Moments

As we did for the construction of the Haar wavelet we start with a basis of box functions. Let $\phi = \chi_{[0,1]}$. Then the scaled and translated versions of ϕ are

$$\phi_{jk} = 2^{j/2}\chi_{[t^{(j)}_k, t^{(j)}_{k+1}]}, \quad \text{with } t^{(j)}_k = k2^{-j}, \quad k = 0, 1, ..., 2^j - 1, \quad j \in \mathbb{N}_0.$$

Remark 3.3.8. *This corresponds to the piecewise constant basis used in Chapter 4.*

Now we define the space spanned by these basis functions

$$V_j = \text{span}\,\{\phi_{j,k} \ : \ k = 0, 1, ..., 2^j - 1\}.$$

Since this space fulfills the requirements of Theorem 3.3.4 we know that there exists a biorthogonal basis generated by $\tilde{\phi}$ such that

$$\langle \phi, \tilde{\phi}(\cdot - k)\rangle = \delta_{0,k}, \qquad k \in I_j.$$

Let W_j be the complement space to V_j in V_{j-1}. Then Theorem 3.3.4 further gives the existence of wavelets ψ, $\tilde{\psi}$ such that $\psi_{j,k}$ and $\tilde{\psi}_{j,k}$ are Riesz bases of W_j and $\tilde{W}_j$ respectively.

Writing the biorthogonal wavelet $\tilde{\psi}$ explicitely is not necessary since we only require its existence for the theory in Chapter 7.

Using an appropriate refinement sequence we can construct the mother wavelet ex-

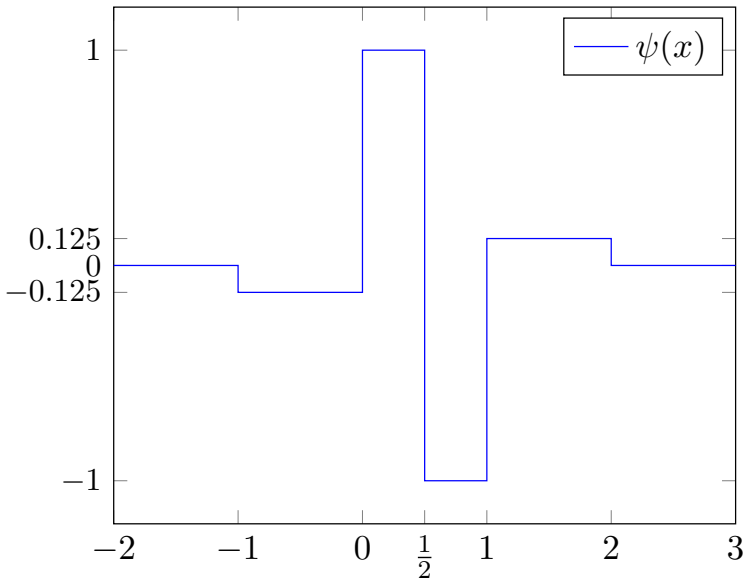

Figure 3.2: A piecewise constant wavelet with three vanishing moments.

plicitely:

$$\psi(x) := \begin{cases} -\frac{1}{8} & x \in [-1, 0) \\ 1 & x \in [0, \frac{1}{2}) \\ -1 & x \in [\frac{1}{2}, 1) \\ \frac{1}{8} & x \in [1, 2] \\ 0 & \text{else.} \end{cases} \tag{3.9}$$

This wavelet is shown in Figure 3.2.

These wavelets have two important properties. Firstly, they have three vanishing moments, i.e.:

$$\langle (\cdot)^\alpha, \psi_{j,k} \rangle = 0, \quad \forall 0 \leq \alpha < 3.$$

and secondly they have a compact support, i.e.:

$$|\text{supp}\psi_{jk}| = 3 \cdot 2^{-j}.$$

3.3.2 Example: B-Spline Wavelets

As before we start with a dual pair of refinable functions $(\theta, \tilde{\theta})$:

$$\theta(x) = \sum_k a_k \theta(2x - k), \qquad \tilde{\theta}(x) = \sum_k \tilde{a}_k \tilde{\theta}(2x - k)$$

and

$$\int_{\mathbb{R}} \theta(x)\tilde{\theta}(x-k)dx = \delta_{k,0}, \quad \forall k \in \mathbb{Z}.$$

For the function θ we choose standard B-spline functions ${}_m\theta$, which are polynomials of degree $m-1$. To define the B-splines we first define the divided differences.

Definition 3.3.9. *We define recursively the m-th order divided difference of $f \in C^m(\mathbb{R})$ at the points $t_i, ..., t_{i+m}$*

$$[t_i, ..., t_{i+m}]f = \frac{[t_{i+1}, ..., t_{i+m}]f - [t_i, ..., t_{i+m-1}]f}{t_{i+m} - t_i},$$

where $[t_i]f = f(t_i)$.

Definition 3.3.10. *Let $x_+^m = (\max\{0, x\})^m$. Then, the m-th order centered cardinal B-spline is given by*

$${}_m\theta(x) = m[0, 1, ..., m]\left(\cdot - x - \left\lfloor \frac{m}{2} \right\rfloor\right)_+^{m-1}$$

Remark 3.3.11. *Using the above formula we easily get the first order cardinal B-spline:*

$${}_2\theta(x) = \begin{cases} x, & 0 \le x \le 1 \\ 2-x, & 1 \le x < 2, \\ 0 & else. \end{cases}$$

The higher-order splines follow analogously.

Before we define the corresponding multiresolution analysis we will discuss properties of these B-splines. In Figure 3.3 we plot the first four cardinal B-splines.

The centered B-splines have compact support

$$\operatorname{supp}{}_m\theta = \left[-\left\lfloor \frac{m}{2} \right\rfloor, \left\lceil \frac{m}{2} \right\rceil\right] =: [l_1, l_2]$$

The centered B-splines are refinable and the refinement sequence $\{a_k\}$ is finite. The refinement sequence is known and is given by

$${}_m\theta(x) = \sum_{k=l_1}^{l_2} a_{km}\theta(2x-k), \tag{3.10}$$

with $a_k = 2^{2-m}\binom{m}{k + \lfloor \frac{m}{2} \rfloor}$.

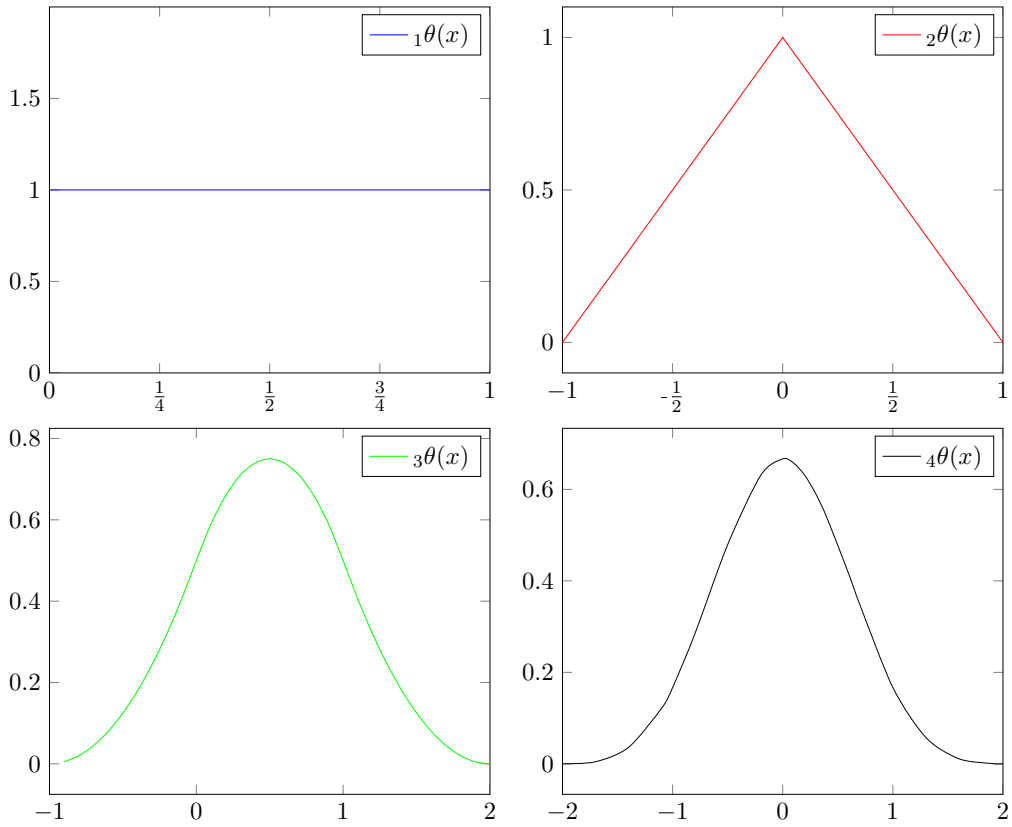

Figure 3.3: The first four cardinal B-spline functions.

Remark 3.3.12. *We can use formula (3.10) to find the refinement sequence for the first order centered B-spline* ${}_2\theta$. *We know that* $\operatorname{supp} {}_2\theta = [-1, 1]$. *This means there are three refinement coefficients to be calculated. Clearly they are*

$$a_{-1} = \frac{1}{2},\ a_0 = 1,\ a_1 = \frac{1}{2}.$$

We show this refinement sequence in Figure 3.4.

We know from [14] that for each m and for any $\tilde{m} \geq m$ with $m + \tilde{m}$ there exists a function ${}_{m,\tilde{m}}\tilde{\theta}$ such that $({}_m\theta, {}_{m,\tilde{m}}\tilde{\theta})$ form a dual pair, i.e.

$$\langle {}_m\theta, {}_{m,\tilde{m}}\tilde{\theta}(\cdot - k)\rangle = \delta_{0,k}.$$

The function ${}_{m,\tilde{m}}\tilde{\theta}$ has a compact support

$$\operatorname{supp} {}_{m,\tilde{m}}\tilde{\theta} = [l_1 - \tilde{m} + 1, l_2 + \tilde{m} - 1] =: [\tilde{l}_1, \tilde{l}_2]$$

and the same symmetry properties as ${}_m\theta$ [17].

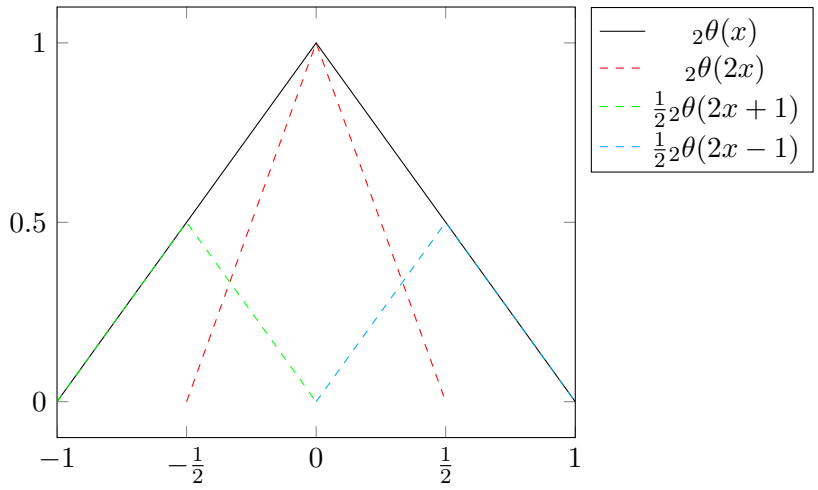

Figure 3.4: The first-order centered cardinal B-spline function and its refinement sequence.

The function ${}_{m,\tilde{m}}\tilde{\theta}$ is also refinable, with refinement sequence

$$ {}_{m,\tilde{m}}\tilde{\theta}(x) = \sum_k \tilde{a}_k \; {}_{m,\tilde{m}}\tilde{\theta}(2x-k). $$

```
def cascade(h, phi_0, j_max):
    start = min(phi_0.keys())
    end   = max(phi_0.keys())
    il    = (end-start)/2
    phi_j = phi_0
    h     = defaultdict(int, h)
    for j in range(1,j_max+1):
        # Previously calculated values:
        phi_jm1 = dict(phi_j)
        # Current values:
        phi_j   = {}
        ind_1 = il * 2**j
        ind_2 = il * 2**(j-1) - 1
        x = 2**(-j)*m
        phi_j[x] = 0
        # Use previously calculated values in refinement eq.
        for l in range(-ind_2,ind_2+1):
            phi_j[x] += h[m-2*l] * phi_jm1[2**(-(j-1))*l]
    return phi_j
```

Figure 3.5: The cascade algorithm (in Python).

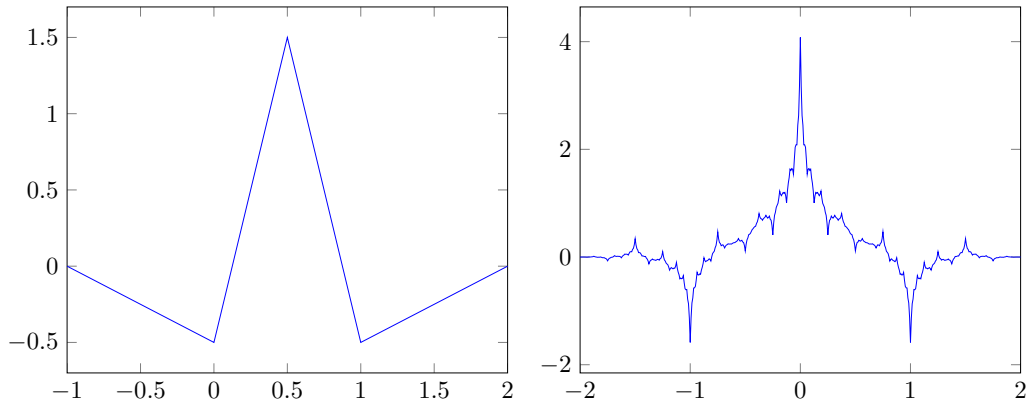

Figure 3.6: The functions ψ (left) and $\tilde{\theta} = {}_{2,2}\theta$ (right) for $m = \tilde{m} = 2$.

Notation: For any function f we denote $f_{j,k} = 2^{j/2} f(2^j x - k)$.

Then the spaces V_j and $\tilde{V}_j$ can be defined as the spans of ${}_m\theta_{j,k}$ and ${}_{m,\tilde{m}}\tilde{\theta}_{j,k}$. Using the single scale bases we can define the wavelets ψ and $\tilde{\psi}$ using the refinement sequences as described in Theorem 3.3.4. The complement spaces W_j and $\tilde{W}_j$ are spanned by the corresponding wavelets $\psi_{j,k}$ and $\tilde{\psi}_{j,k}$.

We do not have an analytic representation of the dual scaling function $\tilde{\theta}$. Instead we can evaluate $\tilde{\theta}$ at point values using the cascade algorithm [19]. The algorithm is shown in Figure 3.5. The wavelet and the dual scaling function as approximated by the cascade algorithm are shown in Figure 3.6.

3.4 Wavelets on Intervals

Wavelets defined on non-periodic domains such as intervals need to be chosen carefully. Wavelets chosen to fulfill certain boundary conditions have been introduced in [18]. They build on the work from [17]. We give here a brief summary of the construction. Essentially these wavelets are constructed in such a way that when the wavelet vanishes on one side of the interval the dual wavelet is unconstrained and vice versa. This allows us to require bounds at the edges of the interval without losing properties such as norm equivalences.

We use the set Z to specify the location of Dirichlet bounds. $Z = \{\}$ corresponds to no Dirichlet bounds, $Z = \{0\}$ corresponds to Dirichlet bounds on the left side of the interval, $Z = \{1\}$ corresponds to bounds on the right and $Z = \{0, 1\}$ corresponds to bounds on both ends of the interval.

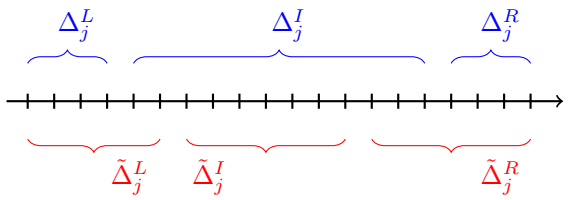

Figure 3.7: The interior, left and right index sets.

In this section we fix the order m of the B-spline wavelet and of the order $\tilde{m}$ of its dual. Let $(\theta, \tilde{\theta}) = ({}_m\theta, {}_{m,\tilde{m}}\tilde{\theta})$ be the chosen dual pair of refinable single scale functions as defined in the previous section. Now we divide the generators of V_j and $\tilde{V}_j$ into left, right and interior basis functions as follows:

$$\Theta'_j = \Theta_j^L \cup \Theta_j^I \cup \Theta_j^R, \text{ and } \tilde{\Theta}'_j = \tilde{\Theta}_j^L \cup \tilde{\Theta}_j^I \cup \tilde{\Theta}_j^R.$$

The interior basis functions are left unchanged, i.e.

$$\Theta_j^I = \{\theta_{j,k} \ : \ k \in \Delta_j^I\}, \quad \tilde{\Theta}_j^I = \{\tilde{\theta}_{j,k} \ : \ k \in \tilde{\Delta}_j^I\},$$

where $\Delta_j^I = \{l, ..., 2^j - l - (m \mod 2)\}$ and $\tilde{\Delta}_j^I = \{\tilde{l}, ..., 2^j - \tilde{l} - (m \mod 2)\}$ with $l = \tilde{l} - (m - \tilde{m})$ and $\tilde{l} \geq \tilde{l}_2$. The index sets are plotted in Figure 3.7.

The left and right generator functions need to be modified in order to ensure the boundary conditions are met. We define

$$\alpha_{nr} = \int x^r \theta(x - n) dx, \quad \tilde{\alpha}_{nr} = \int x^r \tilde{\theta}(x - n) dx.$$

Using these coefficients we redefine the left boundary generating functions

$$\theta_{j,l-m+r}^L = \sum_{n=-l_2+1}^{l-1} \tilde{\alpha}_{nr} \theta_{jn}|_{[0,1]}, \quad r = 0, ..., m-1 \text{ and}$$
$$\tilde{\theta}_{j,\tilde{l}-\tilde{m}+r}^L = \sum_{n=-\tilde{l}_2+1}^{\tilde{l}-1} \alpha_{nr} \tilde{\theta}_{jn}|_{[0,1]}, \quad r = 0, ..., \tilde{m}-1.$$

Then we redefine the right boundary functions symmetrically

$$\theta_{j,2^j-l+m-m \bmod 2-r}^R(1-x) = \theta_{j,l-m+r}^L(x), \quad r = 0, ..., m-1 \text{ and}$$
$$\tilde{\theta}_{j,2^j-\tilde{l}+\tilde{m}-m \mod 2-r}^R(1-x) = \tilde{\theta}_{j,\tilde{l}-\tilde{m}+r}^L(x), \quad r = 0, ..., \tilde{m}-1.$$

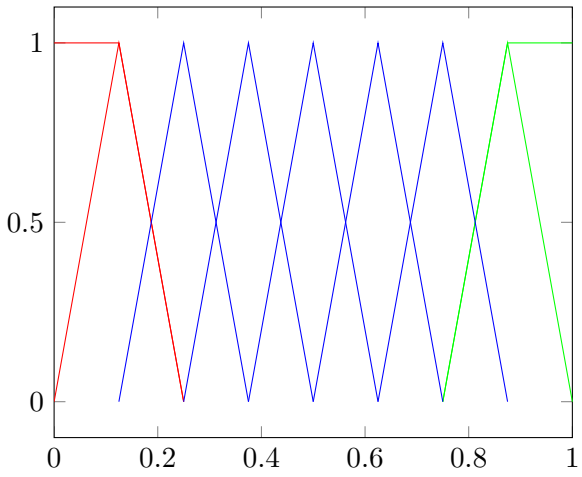

Figure 3.8: The modified generator functions for $d = 2$, the left boundary functions are red, the interior functions are blue and the right boundary functions are green.

We have to assume that $j \geq j_0$ to ensure that the boundary wavelets do not interfere. Then the nested spaces $V_j = \text{span } \Theta'_j$ and $\tilde{V}_j = \text{span } \tilde{\Theta}'_j$ are exact of order m and $\tilde{m}$ respectively. The modified scaling functions are shown for the case of linear B-splines in Figure 3.8.

We denote by Θ_j^Z the functions Θ'_j with the boundary conditions corresponding to Z. Now we need to ensure biorthogonality while keeping the boundary conditions Z valid. More precisely, this means we want

$$\langle \theta_{j,k}^Z, \tilde{\theta}_{j,k'}^{\tilde{Z}} \rangle = \delta_{k,k'}, \qquad \forall k \in \Delta_j, k' \in \tilde{\Delta}_j$$

with the boundary conditions

$$\theta_{j,k}^Z(x) = 0, \quad \forall k \in \Delta_j \text{ and } \tilde{\theta}_{j,k}^{\tilde{Z}}(x) = 0, \quad x \in Z, \ \forall k \in \tilde{\Delta}_j.$$

Remark 3.4.1. *In [18] boundary conditions on higher derivatives are also considered. However, to ensure initial conditions are met we will only require these.*

In the following we ensure that the boundary conditions are met and then biorthogonalise the resulting system. As before the interior functions do not require any modification.

Let $a', \tilde{a}' \in \mathbb{N}$ with $a' \geq m - 1$, $\tilde{a}' \geq \tilde{m} + 1$, then we supplement the generating

functions with higher and lower order B-splines as follows

$$\Theta^{(+0)} = \{\theta^L_{j,a'-(m-1)+r},\ r = 0, ..., m-2\} \cup \{{}_{m-1}\theta_{[j,a']}, ..., {}_{m-1}\theta_{[j,a'+b']}\}$$
$$\tilde{\Theta}^{(-0)} = \{\tilde{\theta}^L_{j,\tilde{a}'-(\tilde{m}+1)+r},\ r = 0, ..., \tilde{m}\} \cup \{{}_{m-1,\tilde{m}+1}\tilde{\theta}_{[j,\tilde{a}']}, ..., {}_{m-1,\tilde{m}+1}\tilde{\theta}_{[j,\tilde{a}'+\tilde{b}']}\},$$

where $a' - m + 1 = \tilde{a}' - \tilde{m} - 1$ and $a' + b' = \tilde{a}' + \tilde{b}'$.

Integrating the primal system on the left and differentiating the dual system yields

$$\Theta^{(+1)} = \{\theta^L_{j,l-1-m+r},\ r = 0, ..., m-1\} \cup \{{}_m\theta_{[j,l-1]}, ..., {}_m\theta_{[j,l-1+b']}\}$$
$$\tilde{\Theta}^{(-1)} = \{\tilde{\theta}^L_{j,\tilde{l}-\tilde{m}-+r},\ r = 0, ..., \tilde{m}-1\} \cup \{{}_{m,\tilde{m}}\tilde{\theta}_{[j,\tilde{l}]}, ..., {}_{m,\tilde{m}}\tilde{\theta}_{[j,\tilde{l}+\tilde{b}'+1]}\}.$$

Theorem 3.4.2 (Theorem 4.2, [18]). *For every Z, $m > 0$, $\tilde{m} > 0$ and $j \geq j_0$ the dual pair Θ^Z_j, $\tilde{\Theta}^Z_j$ defined above satisfies the complementary boundary conditions and can be biorthogonalised.*

For $Z = \{\}$ we can define the wavelet basis functions using the refinement sequences for the generating functions.

$$\Theta^{(+0)}_j = \sum_k (-1)^{k-1} a_{-k-1} \theta^{(+0)}_{-1,k},\ \tilde{\Theta}^{(-0)}_j = \sum_k (-1)^{k-1} \tilde{a}_{-k-1} \tilde{\theta}^{(+0)}_{-1,k}.$$

These wavelets are compactly supported, biorthogonal and the functions have $\tilde{m}+1$ vanishing moments. Now we can introduce

$$\psi^{(+1)}_j = 2^j \left(\int \psi^{(+0)}_j \right) \subset V^{(+1)}_{j+1} \text{ and } \tilde{\psi}^{(-1)}_j = (-1)2^{-j} \frac{d}{dx} \tilde{\psi}^{(-0)}_j \subset \tilde{V}^{(-1)}_{j+1}.$$

Theorem 3.4.3 (Proposition 3.8, [8]). *The collections $\psi^{(+1)}$, $\tilde{\psi}^{(-1)}$ are biorthogonal bases with*

$$V^{(+1)}_{j+1} = V^{(+1)}_j \bigoplus \text{span } \psi^{(+1)}_j, \quad \tilde{V}^{(-1)}_{j+1} = \tilde{V}^{(-1)}_j \bigoplus \text{span } \tilde{\psi}^{(-1)}_j.$$

Now we can define the wavelets for symmetric boundary conditions:

$$\psi^{\{0,1\}}_j := \psi^{(+1)}_j,\ \tilde{\psi}^{\{\}}_j := \psi^{(-1)}_j \text{ and}$$
$$\psi^{\{\}}_j := \psi^{(-1)}_j,\ \tilde{\psi}^{\{0,1\}}_j := \psi^{(+1)}_j.$$

Now we can use these two definitions to find the corresponding results for asymmetric

boundary conditions. More precisely, for $Z = \{0\}$ and if we let $p = l - m + m \mod 2$:

$$\begin{aligned}
\psi_j^{L,\{0\}} &:= \{\psi_{jk}^{\{0,1\}} \ : \ k = 1, ..., p-1\}, \\
\psi_j^{I,\{0\}} &:= \{\psi_{jk}^{\{\}} \ : \ k = p, p+1, ..., 2^j - p\}, \\
\psi_j^{R,\{0\}} &:= \{\psi_{jk}^{\{\}} \ : \ k = 2^j - p + 1, ..., 2^j\}
\end{aligned}$$

and for the dual system

$$\begin{aligned}
\tilde{\psi}_j^{L,\{1\}} &:= \{\psi_{jk}^{\{0,1\}} \ : \ k = 1, ..., p-1\}, \\
\tilde{\psi}_j^{I,\{1\}} &:= \{\psi_{jk}^{\{\}} \ : \ k = p, p+1, ..., 2^j - p\}, \\
\tilde{\psi}_j^{R,\{1\}} &:= \{\psi_{jk}^{\{\}} \ : \ k = 2^j - p + 1, ..., 2^j\}.
\end{aligned}$$

Chapter 4

Galerkin Boundary Element Methods

In this chapter we discuss the discretisation of the direct and indirect formulations of the heat equation given in Section 2.2. Essentially, we need to approximate the anisotropic Sobolev space $H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)$ that these equations are formulated in. To approximate this space we choose a finite dimensional subspace of the anisotropic Sobolev space.

In this chapter we discuss such discretisations by full tensor product spaces of piecewise polynomials. However, due to the coercivity of the single-layer operator many other kinds of discretisation are possible. Wavelet bases have been introduced in Chapter 3. Discretisations using wavelet bases will be discussed in Chapter 7 and discretisations using sparse grid spaces will be discussed in Chapter 6.

First we discuss in general terms the discretisation in space and time. Then we introduce the discretisation in time by piecewise constant basis functions and in space by piecewise polynomial basis functions. Next, we look in detail at the discretisation of the single and double layer operators. This includes finding the analytical solutions for the time integrals of both operators. Further, we discuss implementational issues, such as the solution of the resulting linear system and quadrature rules. Finally, we give numerical results showing a comparison between the boundary element method described in this chapter and a finite element discretisation.

4.1 Space-Time Discretisation

In this section we discuss the discretisation in space and time. We initially discuss the discretisations without giving the discrete space since the following results hold for all choices of discrete subspace. In the following sections we will give details of the construction of a full tensor product discretisation with piecewise polynomials.

Let $\mathcal{X}_L$ be a nested sequence of discrete spaces, i.e.

$$\mathcal{X}_0 \subset \mathcal{X}_1 \subset ... \subset \mathcal{X}_L \subset ... \subset H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma).$$

Further, let $\psi_L \in \mathcal{X}_L$ be the solution to either the direct or indirect formulations of the heat equation with Dirichlet data, i.e.

$$\begin{aligned} &\text{Find } \psi_L \in \mathcal{X}_L \text{ such that} \\ &\qquad \langle V\psi_L, \eta\rangle = \langle g, \eta\rangle, \qquad \text{for all } \eta \in \mathcal{X}_L \quad \text{(Direct method)} \\ &\text{or} \quad \langle V\psi_L, \eta\rangle = \langle (\frac{1}{2} + K)g, \eta\rangle, \quad \text{for all } \eta \in \mathcal{X}_L \quad \text{(Indirect method)} \end{aligned} \tag{4.1}$$

Lemma 4.1.1. *The solution $\psi_L \in \mathcal{X}_L$ of both problems is unique and quasi-optimal:*

$$\|\psi - \psi_L\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq \frac{\|V\|}{c_v} \inf_{\eta_L \in \mathcal{X}_L} \|\psi - \eta_L\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)}. \tag{4.2}$$

Proof. This follows directly from the Lemma of Lax-Milgram and the Lemma of Céa respectively, using the coercivity and continuity of V in the appropriate spaces. See [15] or [42] for more details. □

Remark 4.1.2. *Analogous results hold for the Neumann problem.*

4.1.1 Time Discretisation

Now we give an explicit construction for a discrete space in time. We will refer to this space as $\mathcal{X}_{l_t}$.

For a given level $l_t \in \mathbb{N}$, choose $N_t = 2^{l_t}$ and the index set $\Delta_{l_t} = \{0, 1, ..., N_t - 1\}$.

We subdivide the time interval of interest $\mathcal{I} = (0, T)$ by $t_k^{l_t} = Tk/N_t$ with $k \in \Delta_{l_t}$. This gives us an equidistant partition of the time interval. The time step size h_t is given by $h_t = T/N_t$.

For the discretisation we employ piecewise constant functions

$$\chi_k(t) = \begin{cases} 1, & \text{if } t_k^{l_t} < t < t_{k+1}^{l_t} \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

Functions of higher polynomial degree degree can also be used (see e.g. [42]). Another option is to use wavelets in time, for this see Chapter 3.

Then the discrete space in time is given as the span of these functions

$$\mathcal{X}_{l_t} = \operatorname{span}\{\chi_k\}_{k=1}^{N_t}.$$

Once we have defined the space discretisation we can tensorise the two spaces to form the discrete space $\mathcal{X}_L$.

4.1.2 Space discretisation

Let Γ denote the boundary of the domain Ω. In the following Γ is assumed to be smooth, however, more general boundaries are possible. For example, polygonal domains or other piecewise smooth domains are easily handled.

In two dimensions, the smooth boundary Γ of a simply connected domain can be parameterised by a single 1-periodic function:

$$\gamma \ : \ [0,1] \to \Gamma.$$

In the following we assume that the function γ is analytic [34].

Remark 4.1.3. *In higher dimension [42], $d > 2$, the domain needs to be cut up into smaller non-overlapping patches Γ_i, each with its own parameterisation*

$$\gamma_i \ : \ [0,1]^{d-1} \to \Gamma_i.$$

Each patch is meshed individually.

We create a mesh $\mathcal{T}_h$ on $[0,1]^{d-1}$, for example, by division into intervals, cubes or simplices. We denote the elements of this mesh by $\tau \in \mathcal{T}_h$. For $d = 2$ this is shown in Figure 4.1.

Then we define the discrete space $\mathcal{X}_{l_x}^{p_x}$ as the image of the space of piecewise polynomials of degree p_x. Here l_x gives the number of elements in the mesh. More precisely,

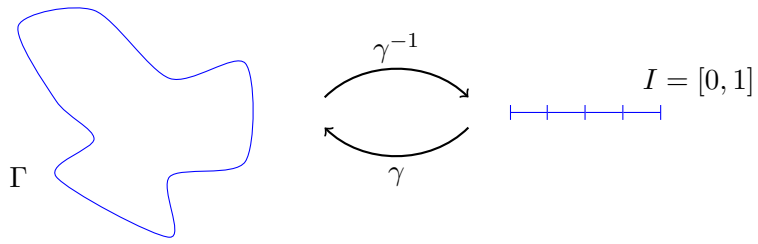

Figure 4.1: The mapping γ and its inverse mapping by γ^{-1} for $d = 2$.

there are 2^{l_x} elements $\tau \in \mathcal{T}_h$ and

$$\mathcal{X}_{l_x}^{p_x} = \{v \in L^2(\Gamma) \ : \ v|_\tau \circ \gamma \in \mathcal{P}_{p_x} \ \ \forall \tau \in \mathcal{T}_h\},$$

where $\mathcal{P}_{p_x}$ is an appropriate space of polynomials of degree p_x.

The basis functions on Γ can also be given using the parametrisation γ. This gives a basis defined on each element τ of the triangulation:

$$b_j = \hat{b}_j \circ \gamma^{-1}, \quad j = 1, ..., (p_x + 1)^{d-1},$$

where $\hat{b}_j$ are the basis functions on the interval $I = [0, 1]$.

Remark 4.1.4. *The number of basis functions on each elements is given under the assumption that tensor product polynomials of degree p_x in each direction are used.*

The collection of these functions for all $\tau \in \mathcal{T}_h$ forms a basis for $\mathcal{X}_{l_x}^{p_x}$. Thus, if there are N_x elements in $\mathcal{T}_h$, then there are $(p_x + 1)^{d-1} N_x$ basis functions. It is convenient to denote them by $\{b_\alpha(x)\}_\alpha$. Then $\{b_\alpha(x)\chi_n(t)\}_{\alpha,n}$ forms a basis of $\mathcal{X}_L := \mathcal{X}_{l_t} \otimes \mathcal{X}_{l_x}^{p_x}$. This is the well known full tensor product space. Alternatively it is possible to combine space and time discretisations using a sparse grid space. This will be discussed in Chapter 6.

The Galerkin solution ψ_L belongs to $\mathcal{X}_L$, so we can write it as

$$\psi_L(x, t) = \sum_{n=0}^{N_t-1} \sum_{\beta=0}^{N_x-1} \psi_{\beta n} b_\beta(x) \chi_n(t),$$

where N_x is the number of basis functions in space and N_t is the number of basis functions in time.

This gives us the discretised form of the equation to be solved for the Indirect Method

(2.2.2):

$$\sum_{n=0}^{N_t-1}\sum_{\beta=0}^{N_x-1}\langle b_\alpha\chi_m, Vb_\beta\chi_n\rangle\psi_{\beta n} = \langle b_\alpha\chi_m, g\rangle, \quad \alpha = 0,...,N_x-1, \quad m = 0,...,N_t-1. \tag{4.4}$$

The Direct Method (2.2.1) can be discretised completely analogously and gives a similar linear system to solve.

Next we look at some examples of parameterisations for different smooth boundaries Γ that will be used in the numerical tests in Section 5.4.

Example: The unit circle in two dimensions

The simplest example of a smooth domain in $d = 2$ is the circle $\Gamma = \partial B_R(0)$. This domain is shown in Figure 4.2. It is easy to see that it can be mapped bijectively and smoothly onto the interval $[0, 1]$.

We denote the mapping from the unit interval $[0, 1]$ to the boundary by γ , it is given by

$$\begin{aligned}\gamma \;:\; [0,1] &\to \Gamma = \partial B_R(0)\\ \varphi \quad &\mapsto R\begin{pmatrix}\cos(\pi(2\varphi-1))\\ \sin(\pi(2\varphi-1))\end{pmatrix}\end{aligned}$$

The inverse mapping is denoted by γ^{-1} and is given by:

$$\begin{aligned}\gamma^{-1} \;:\; \Gamma = \partial B_R(0) &\to [0,1].\\ \begin{pmatrix}x\\ y\end{pmatrix} \quad &\mapsto \frac{1}{2\pi}\operatorname{atan2}(y,x)+\frac{1}{2},\end{aligned}$$

where atan2 is the function given by:

$$\operatorname{atan2}(y,x) = 2\arctan\left(\frac{y}{\sqrt{x^2+y^2}+x}\right)$$

In this case the outer normal at the point $\gamma(\varphi)$ is easy to calculate. It is given by:

$$\begin{pmatrix}n_1\\ n_2\end{pmatrix} = \begin{pmatrix}\cos(2\pi\varphi-\pi)\\ \sin(2\pi\varphi-\pi)\end{pmatrix}.$$

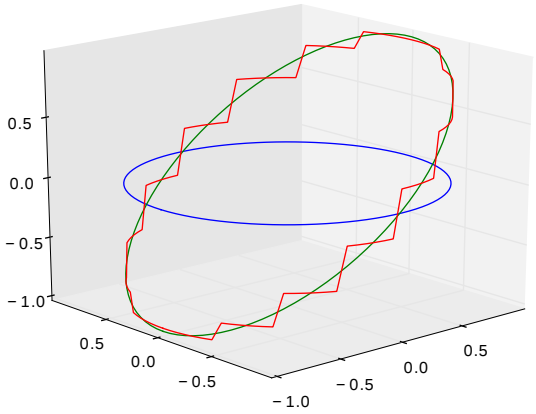

Figure 4.2: A circular domain $\Omega = B_1(0)$ in blue and the exact boundary flux in green, as well as the approximated boundary flux in red.

For the integration it is also necessary to calculate the derivatives of the mapping γ. For a circle these derivatives have a very simple form:

$$\|\gamma'(x)\| = 2\pi R. \tag{4.5}$$

Remark 4.1.5. *The derivative γ does not depend on x, so it is possible to speed up the numerical quadrature needed to evaluate the boundary integral operators by moving it out of the integrals.*

Example: Ellipse

Another easily parameterised smooth domain is the ellipse. In our tests we choose an ellipse where the major axis coincides with the x-axis. The major axis of an ellipse is its longest diameter. These ellipses are described by two parameters a and b which give the eccentricity of the ellipse. The values a, b and the major axis of an ellipse are shown in Figure 4.3.

As before we denote the smooth, 1-periodic mapping from the interval $[0, 1]$ onto the boundary of the ellipse by γ.

$$\begin{aligned} \gamma \;:\; [0,1] &\to \Gamma \\ \varphi \;&\mapsto \begin{pmatrix} a\cos(2\pi\varphi) \\ b\sin(2\pi\varphi) \end{pmatrix} \end{aligned}$$

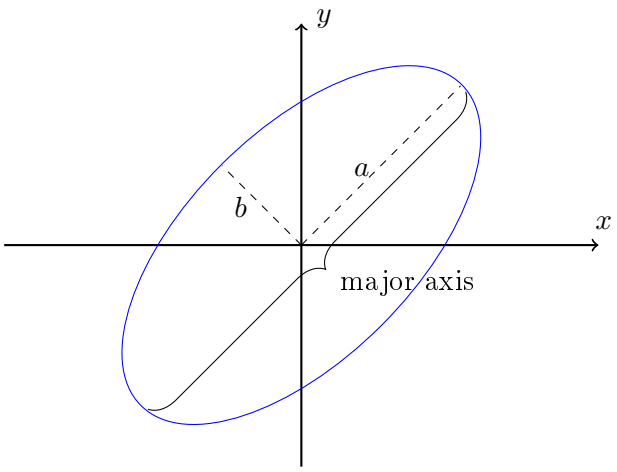

Figure 4.3: An ellipse, the major axis and the values of a and b.

It is simple to calculate that

$$\|\gamma'(\varphi)\| = 2\pi\sqrt{(a\sin(2\pi\varphi))^2 + (b\cos(2\pi\varphi))^2}.$$

Further, the inverse of γ is given by:

$$\gamma^{-1}(x) = \frac{1}{2\pi}\operatorname{atan2}(ax_2, bx_1).$$

The outer normal for the ellipse is given by:

$$\begin{pmatrix}\tilde{n}_1\\ \tilde{n}_2\end{pmatrix} = \begin{pmatrix}1/a\cos(2\pi\varphi)\\ 1/b\sin(2\pi\varphi)\end{pmatrix}.$$

And normalising gives the unit outer normal $n = \tilde{n}/\|\tilde{n}\|$ as required.

Example: A star-shaped domain

A more complicated domain that is still easy to parameterise, is the star-shaped domain shown in Figure 4.4. This domain was chosen to be smooth and less symmetric than the previous tests.

In this case the smooth mapping γ is given by:

$$\begin{aligned}\gamma \;:\; [0,1] &\to \Gamma\\ s \;&\mapsto \frac{1}{20}\begin{pmatrix}\cos(2\pi s)(4+\cos(6\pi s)+\cos(2\pi s))\\ \sin(2\pi s)(4+\cos(6\pi s)+\cos(2\pi s))\end{pmatrix}\end{aligned} \tag{4.6}$$

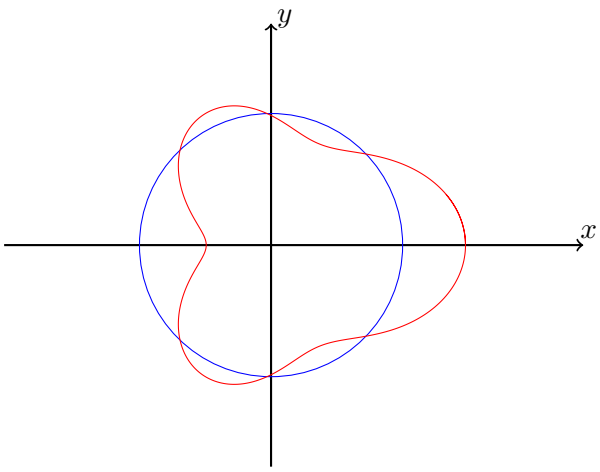

Figure 4.4: The star-shaped domain used for tests in red and a circle of radius 1 as a reference in blue.

It is also necessary to calculate the derivative of the mapping γ. In this case it is given by

$$\|\gamma'(s)\| = \frac{\pi}{10}\sqrt{(4+\cos(6\pi s)+\cos(2\pi s))^2+(3\sin(6\pi s)+\sin(2\pi s))^2} \tag{4.7}$$

4.2 The Single-layer Operator

Discretisation of the single-layer operator V leads to a square matrix $G_{\alpha+nN_x,\beta+mN_x}$. When we discretise with piecewise constant basis functions in time the matrix has a block Toeplitz structure. We examine each of the N_t blocks corresponding to a pair of time steps m, n. The blocks each have size $N_x \times N_x$. To keep the notation compact we will also refer to the matrix blocks as (G_{mn}), their entries are

$$\begin{aligned}(G_{mn})_{\alpha\beta} &:= \langle b_\alpha \chi_m, V b_\beta \chi_n\rangle \\ &= \int_\Sigma \int_\Sigma b_\alpha(x) b_\beta(y) \chi_m(t) \chi_n(s) G(x-y, t-s) dy\, ds\, dx\, dt.\end{aligned}$$

Assume in the following that constant basis functions are used in time. We change the order of integration and define a time-integrated kernel:

$$g_{mn}(x) := \int_{mh_t}^{(m+1)h_t} \int_{nh_t}^{(n+1)h_t} G(x, t-s) ds dt.$$

Remark 4.2.1. *Since the fundamental solution $G(x,t)$ is zero when $t < 0$, the time-integrated kernel $g_{mn}(x)$ will also be zero when $m < n$.*

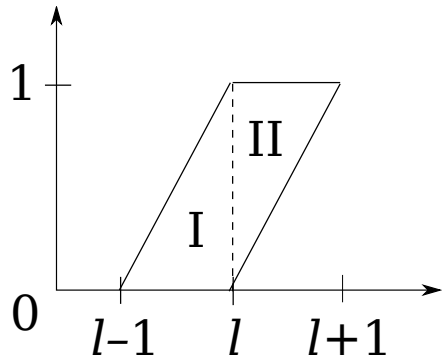

Figure 4.5: The transformed subdomains I and II.

In this section easily simplify the integration in time by reducing this double-integral to a one-dimensional integral (see [15], Section 6). After the reduction one can either apply a numerical quadrature rule or evaluate the one-dimensional integral analytically.

First, let $l = m - n$ and scale the integration variables in $g_{mn}(x)$ to get

$$\begin{aligned} g_{mn}(x) &= h_t^2 \int_0^1 \int_0^1 G(x, h_t(t + m - (s + n)))dsdt \\ &= h_t^2 \int_0^1 \int_{t+l-1}^{t+l} G(x, h_t\tilde{s})d\tilde{s}dt, \end{aligned}$$

where $\tilde{s} = t - s + l$.

By dividing the domain into two triangular domains (shown in Figure 4.5) and changing the order of integration we get

$$g_{mn}(x) = h_t^2 \underbrace{\int_{l-1}^{l} \int_0^{\tilde{s}-l+1} G(x, \tilde{s}h_t)dtd\tilde{s}}_{I} + h_t^2 \underbrace{\int_{l}^{l+1} \int_{\tilde{s}-l}^{1} G(x, \tilde{s}h_t)dtd\tilde{s}}_{II}. \tag{4.8}$$

The first integrand, which corresponds to domain I above is 0 in the case $l = 0$.

Now the integration over t can easily be performed and we get

$$g_{mn}(x) = h_t^2 \int_{l-1}^{l} G(x, \tilde{s}h_t)(\tilde{s} - l + 1)d\tilde{s} + h_t^2 \int_{l}^{l+1} G(x, \tilde{s}h_t)(l + 1 - \tilde{s})d\tilde{s}. \tag{4.9}$$

The integration over $\tilde{s}$ can be done analytically or using a quadrature rule. Since the integral of $g_{mn}(x)$ has an algebraic singularity at $\tilde{s} = 0$ one would use a composite

Gauss-Legendre rule for $l = 0, 1$ and a Gauss-Legendre rule for $l > 1$.

Integrating analytically involves evaluating the exponential integral function. In languages such as C/C++ or Python this can be done efficiently. In other languages it may be preferable to use an integration rule instead.

Next we will show how to derive the integral value analytically for any $d > 1$. We will later use the same approach to calculate the time integrals for the double layer operator.

Definition 4.2.2. *We define the exponential integral functions as*

$$\mathrm{Ei}(x) := \int_{-\infty}^{x} e^t t^{-1} dt.$$

Further, for ease of notation we define as in [15]:

$$\mathrm{E}_1(x) := -\mathrm{Ei}(-x).$$

We will use the following simple integration rules:

$$\int_a^b e^{-r/s} ds = \left[r\,\mathrm{Ei}(-r/s) + s e^{-r/s} \right]_a^b \tag{4.10}$$

and

$$\int_a^b e^{-r/s} s^{-1} ds = \left[-\mathrm{Ei}(-r/s) \right]_a^b . \tag{4.11}$$

When the lower integration limit a is zero and the upper integration limit $b > 0$, we have

$$\int_0^b e^{-r/s} ds = r\,\mathrm{Ei}(-r/b) + s e^{-r/b} \tag{4.12}$$

and

$$\int_0^b e^{-r/s} s^{-1} ds = -\mathrm{Ei}(-r/b). \tag{4.13}$$

We start with the simplest case $l = 0$, i.e. the elements on the diagonal. In this case the integral (4.8) has the form:

$$\begin{aligned} g_{mm}(x) &= h_t^2 \left[\int_0^1 G(x, \tilde{s} h_t) d\tilde{s} - \int_0^1 G(x, \tilde{s} h_t) \tilde{s} d\tilde{s} \right] \\ &= h_t^2 (4\pi)^{-d/2} \left[\int_0^1 (\tilde{s} h_t)^{-1} e^{-|x|^2/(4\tilde{s} h_t)} d\tilde{s} - \int_0^1 (\tilde{s} h_t)^{-1} e^{-|x|^2/(4\tilde{s} h_t)} \tilde{s} d\tilde{s} \right] \end{aligned}$$

We look at the two integrals separately. Using the integration rule (4.11) the first

integral gives

$$\begin{aligned}\int_0^1 (\tilde{s}h_t)^{-1} e^{-|x|^2/(4\tilde{s}h_t)} d\tilde{s} &= \frac{1}{h_t}\int_0^{h_t} \sigma^{-1} e^{-|x|^2/(4\sigma)} d\sigma \\ &= -\frac{1}{h_t}\,\mathrm{Ei}(-|x|^2/(4h_t)).\end{aligned}$$

Further, the second integral is

$$\begin{aligned}\int_0^1 (\tilde{s}h_t)^{-1} e^{-|x|^2/(4\tilde{s}h_t)} \tilde{s} d\tilde{s} &= \frac{1}{h_t}\int_0^{h_t} \sigma^{-1} e^{-|x|^2/(4\sigma)} \sigma d\sigma \\ &= \frac{1}{h_t}\int_0^{h_t} e^{-|x|^2/(4\sigma)} d\sigma.\end{aligned}$$

By using equation (4.10) we find that the second integral gives

$$\begin{aligned}\frac{1}{h_t}\int_0^{h_t} e^{-|x|^2/(4\sigma)} d\sigma &= \left[\frac{|x|^2}{4}\,\mathrm{Ei}(-|x|^2/(4\sigma)) + \sigma e^{-|x|^2/(4\sigma)}\right]_0^{h_t} \\ &= \frac{1}{h_t}\left(\frac{|x|^2}{4}\,\mathrm{Ei}\left(-|x|^2/(4h_t)\right) + e^{-|x|^2/(4h_t)}\right).\end{aligned}$$

In order to simplify notation we set

$$a_k(x) = \frac{\|x\|^2}{(4kh_t)}.$$

Here we only need a_1, however in the other cases other values of k will be used. Then we sum up the two integrals and get the following solution for the time integral on the diagonal:

$$g_{mm}(x) = h_t(4\pi)^{-d/2}(\mathrm{E}_1(a_1)(1+a_1) - e^{-a_1}).$$

Next we look at the case $l = 1$. This case can be handled in much the same way as the calculation above, giving

$$\begin{aligned}g_{m,m-1}(x) &= h_t^2\left(\int_0^1 G(x,\tilde{s}h_t)d\tilde{s} + \int_1^2 G(x,\tilde{s}h_t)(2-\tilde{s})d\tilde{s}\right) \\ &= (4\pi)^{-d/2}\left(h_t\int_0^{h_t} e^{-|x|^2/(4\sigma)}d\sigma - \int_{h_t}^{2h_t} e^{-|x|^2/(4\sigma)}d\sigma\right. \\ &\qquad \left. + 2h_t\int_{h_t}^{2h_t} e^{-|x|^2/(4\sigma)}\sigma^{-1}d\sigma\right).\end{aligned}$$

Again we look at each of the summands individually. It is easy to see that using

equations (4.10) and (4.11) respectively, we get

$$h_t \int_0^{h_t} e^{-|x|^2/(4\sigma)} d\sigma = \left[\frac{|x|^2}{4} \operatorname{Ei}(-|x|^2/(4\sigma)) + \sigma e^{-|x|^2/(4\sigma)} \right]_0^{h_t},$$

$$\int_{h_t}^{2h_t} e^{-|x|^2/(4\sigma)} d\sigma = \left[\frac{|x|^2}{4} \operatorname{Ei}(-|x|^2/(4\sigma)) + \sigma e^{-|x|^2/(4\sigma)} \right]_{h_t}^{2h_t},$$

and

$$2h_t \int_{h_t}^{2h_t} e^{-|x|^2/(4\sigma)} \sigma^{-1} d\sigma = 2h_t \left[-\operatorname{Ei}(-|x|^2/(4\sigma)) \right]_{h_t}^{2h_t}.$$

In total this gives us

$$(4\pi)^{-d/2} \left(-2h_t E_1(a_1)(a_1+1) + 2h_t e^{-a_1} - 2h_t e^{-a_2} + h_t(2+a_2) E_1(a_2) \right).$$

For $l < 0$ it is clear that $g_{m,m-l}(x) = 0$ due to the form of the fundamental solution. Thus, the remaining case is $l > 1$. Here the integral has the form

$$\begin{aligned} g_{m,m-l}(x) &= h_t^2 \left(\int_{l-1}^{l} G(x, \tilde{s}h_t)(\tilde{s} - l + 1) d\tilde{s} + \int_{l}^{l+1} G(x, \tilde{s}h_t)(l + 1 - \tilde{s}) d\tilde{s} \right) \\ &= h_t^2 (4\pi)^{-d/2} \Bigg(\frac{1}{h_t} \int_{l-1}^{l} e^{-|x|^2/(4\tilde{s}h_t)} d\tilde{s} \\ &\qquad - (l-1) \int_{l-1}^{l} e^{-|x|^2/(4\tilde{s}h_t)} (\tilde{s}h_t)^{-1} d\tilde{s} \\ &\qquad - \frac{1}{h_t} \int_{l}^{l+1} e^{-|x|^2/(4\tilde{s}h_t)} d\tilde{s} \\ &\qquad + (l+1) \int_{l}^{l+1} e^{-|x|^2/(4\tilde{s}h_t)} (\tilde{s}h_t)^{-1} d\tilde{s} \Bigg). \end{aligned}$$

Then, we sum up all the integrands and get

$$\begin{aligned} g_{m,m-l}(x) = (4\pi)^{-d/2} \Bigg(& \int_{(l-1)h_t}^{lh_t} e^{-|x|^2/(4\sigma)} d\sigma - h_t(l-1) \int_{(l-1)h_t}^{lh_t} e^{-|x|^2/(4\sigma)} \sigma^{-1} d\sigma \\ & - \int_{lh_t}^{(l+1)h_t} e^{-|x|^2/(4\sigma)} d\sigma \\ & + (l+1)h_t \int_{lh_t}^{(l+1)h_t} e^{-|x|^2/(4\sigma)} \sigma^{-1} d\sigma \Bigg). \end{aligned}$$

Next we integrate using the simple integration rules (4.10) and (4.11), giving

$$\begin{aligned} g_{m,m-l}(x) = (4\pi)^{-d/2}\Big(&h_t(a_1+l-1)E_1(a_{l-1}) + h_t(l+1+a_1)E_1(a_{l+1}) \\ &- 2h_t(a_1+l)E_1(a_l) - h_t(l-1)e^{-a_{l-1}} \\ &- (l+1)h_t e^{-a_{l+1}} + 2lh_t e^{-a_l}\Big). \end{aligned}$$

Finally, summarising the results we have

$$\begin{aligned} g_{m,m}(x) &= h_t(4\pi)^{-d/2} f_1(x), \\ g_{m,m-1}(x) &= h_t(4\pi)^{-d/2}(-2f_1(x) + f_2(x)), \\ g_{m,m-l}(x) &= h_t(4\pi)^{-d/2}(f_{l-1}(x) + f_{l+1}(x) - 2f_l(x)), \quad l > 1. \end{aligned} \tag{4.14}$$

Where

$$f_l(x) = E_1(a_l)(l + a_l) - le^{-a_l}. \tag{4.15}$$

The function $g_{mn}(x)$ has a logarithmic singularity for x tending to zero. This is easy to see using the Taylor expansion of $\mathrm{Ei}(x)$:

$$\mathrm{Ei}(x) = \gamma + \ln|x| + \sum_{k=1}^{\infty} \frac{x^k}{k\,k!}.$$

This series representation holds for all $x > 0$ (see [1]). However, for large x it converges slowly and should not be used in calculations.

This means that for the integration we need a quadrature rule suitable for dealing with functions with logarithmic singularities. See Section 4.6 for details on the construction of suitable quadrature rules.

4.2.1 Structure of the Matrix

The structure of the matrix of the single-layer operator depends on the choice of basis functions in time and space. We use piecewise constant basis functions in time, leading to a block Toeplitz structure for the matrix.

As before we refer to the matrix block corresponding to the time intervals m and n as G_{mn}. Several of these block matrices are zero, more precisely

$$\langle \chi_m, V\chi_n \rangle = 0, \qquad \text{if } m < n,$$

since $G(x, t-s) = 0$ if $s > t$.

Lemma 4.2.3. *The diagonal matrix blocks G_{nn} for $n = 1...N_t$ are symmetric positive definite.*

Proof. To show that the matrices on the diagonal are indeed symmetric we look at the matrix entries

$$(G_{nn})_{\alpha\beta} = \int_\Gamma \int_\Gamma b_\alpha(x) b_\beta(y) g_{nn}(x-y) dx dy.$$

The time integrated kernel $g_{nn}(x)$ depends only on x^2, so switching x and y above does not change the value of the integral. Thus, $(G_{nn})_{\alpha\beta} = (G_{nn})_{\beta\alpha}$.

The single layer operator has been shown to be coercive in Theorem 2.2.8, as such the matrix G must be positive definite. Thus, the diagonal blocks must also be positive definite.

Together this gives the assertion that the diagonal blocks are symmetric positive definite matrices. □

For piecewise constant basis functions in time steps a further simplification is possible.

Lemma 4.2.4. *For a piecewise constant polynomial basis with constant time steps there holds*

$$G_{n_1 k_1} = G_{n_2 k_2} \qquad \text{if } n_1 - k_1 = n_2 - k_2.$$

Proof. Since

$$(G_{nk})_{\alpha\beta} = \int_\Gamma \int_\Gamma b_\alpha(x) b_\beta(y) g_{nk}(x-y) ds_x ds_y,$$

we need only show that $g_{n_1,k_1} = g_{n_2,k_2}$. Let $l = n_1 - k_1 = n_2 - k_2$. According to equation (4.9) we can rewrite g_{n_1,k_1} for piecewise constant basis functions as

$$g_{n_1,k_1}(x) = h_t^2 \int_{l-1}^{l} G(x, \tilde{s} h_t)(\tilde{s} - l + 1) d\tilde{s} + h_t^2 \int_l^{l+1} G(x, \tilde{s} h_t)(l + 1 - \tilde{s}) d\tilde{s}.$$

Since this equation depends only on l the assertion is clear. □

Remark 4.2.5. *Lemma 4.2.4 does not hold for higher order polynomials in time. The time-integrated kernel is given by*

$$g_{mn}(x) = \int_{\mathcal{I}} \int_{\mathcal{I}} \chi_m(t) \chi_n(s) G(x, t-s) ds\, dt.$$

If χ_m and χ_n are not piecewise constant the roles of t and s cannot simply be exchanged.

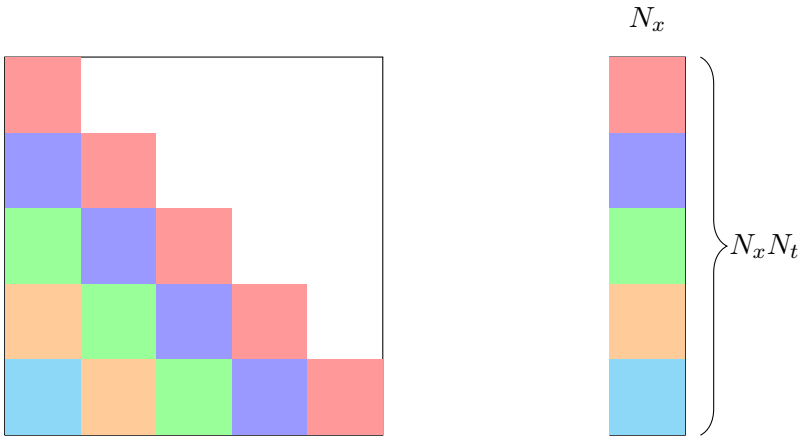

Figure 4.6: Structure of the matrix of the single layer operator and the matrix as it is stored for implementational purposes.

This means that when using piecewise constant polynomial basis functions in time the matrix of the single-layer operator G has the form

$$G = \begin{pmatrix} G_{00} & 0 & 0 & \dots & 0 \\ G_{01} & G_{00} & 0 & \dots & 0 \\ G_{02} & G_{01} & G_{00} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \end{pmatrix}$$

This means we can save storage space by only saving one block matrix for each $n-k$ and we store only $N_t N_x^2$ matrix entries instead of $(N_t N_x)^2$. The structure of the matrix and the structure of the stored matrix are shown in Figure 4.6.

4.3 The Double-layer Operator

Now we look at the discretisation of the double-layer operator. We require this in order to assemble the right-hand side for the indirect method. It is very similar to the discretisation of the single-layer operator given in Section 4.2. However, it requires the evaluation of the normal derivative of the fundamental solution.

The normal derivative has a relatively simple form:

$$\begin{aligned} \frac{\partial}{\partial n_y} G(x-y,t) &= \begin{cases} (4\pi t)^{-d/2}(2t)^{-1} e^{-\|x-y\|^2/(4t)} \langle n_y, x-y\rangle & t \geq 0 \\ 0 & t < 0, \end{cases} \\ &= G(x-y,t)/(2t)\langle n_y, x-y\rangle. \end{aligned} \tag{4.16}$$

One can either evaluate $\langle Kg, b_\alpha(x)\xi_m(t)\rangle$ directly or approximate the function g by a polynomial g_h. The use of linear (or higher order) polynomials is necessary in that case to ensure $g_h \in H^{1/2,1/4}(\Sigma)$.

When approximating g by polynomials the advantage is that the matrix of the double-layer operator can be set up with analytically evaluated time integrals as in the case of the single-layer operator. This can save computational effort. However, choosing an approximation for the function g leads to an additional error term.

The matrix of the double layer operator is given by:

$$\begin{aligned}(K_{mn})_{\alpha\beta} &= \langle b_\alpha\chi_m, Kb_\beta\chi_n\rangle \\ &= \int_\Sigma\int_\Sigma \frac{\partial}{\partial n_y}G(x-y,t-s)b_\alpha(x)b_\beta(y)\chi_n(t)\chi_k(s)dy\,ds\,dx\,dt \\ &= \int_\Gamma\int_\Gamma\int_{mh_t}^{(m+1)h_t}\int_{nh_t}^{(n+1)h_t} \frac{\partial}{\partial n_y}G(x-y,t-s)b_\alpha(x)b_\beta(y)ds\,dt\,dy\,dx \\ &= \int_\Gamma\int_\Gamma k_{mn}(x-y)b_\alpha(x)b_\beta(y)dy\,dx,\end{aligned}$$

where $k_{mn}(x-y)$ is the time-integrated kernel of the double layer operator.

Using the same method as for the time-integration of the single-layer operator, we split the integration into two domains:

$$\begin{aligned}k_{mn}(x-y) = h_t^2\Big(&\int_{l-1}^{l}\frac{\partial}{\partial n_y}G(x-y,sh_t)(s-l+1)ds \\ &+\int_l^{l+1}\frac{\partial}{\partial n_y}G(x-y,sh_t)(l+1-s)ds\Big)\end{aligned}$$

To evaluate this expression, we need the integrals used previously, as well as the integral

$$\int_a^b x^{-2}e^{-r/4x}dx = \Big[4r^{-1}e^{-r/4x}\Big]_a^b. \tag{4.17}$$

Since the calculations have been done in detail for the single-layer potential we will only summarise the results of the corresponding calculation for the double layer potential. As before

$$a_k(x) := \frac{\|x\|^2}{4kh_t}.$$

Then the results of the calculation are

$$\begin{aligned} k_{m,m}(x) &= (8\pi)^{-d/2}\langle n_y, x\rangle \tilde{f}_1(x), \\ k_{m,m-1}(x) &= (8\pi)^{-d/2}\langle n_y, x\rangle \left(\tilde{f}_2(x) - 2\tilde{f}_1(x)\right), \\ k_{m,m-l}(x) &= (8\pi)^{-d/2}\langle n_y, x\rangle \left(\tilde{f}_{l-1}(x) - 2\tilde{f}_l(x) + \tilde{f}_{l+1}(x)\right), \quad l > 1. \end{aligned} \tag{4.18}$$

Where

$$\tilde{f}_l(x) = \frac{e^{-a_l}}{a_l} - E_1(a_l).$$

As in the case of the single-layer operator the analytically evaluated time integrals have a logarithmic singularity. This makes finding suitable quadrature rules simpler, as the same rule can be applied to both operators. The choice of quadrature rules is discussed in detail in Section 4.6.

4.4 Assembling the Right Hand Side

The direct and indirect methods for solving the Dirichlet problem were given in equation (4.1). The right hand side for these problems was given by g or $\frac{1}{2}g + Kg$, for the indirect and direct methods respectively. Thus, to solve the resulting linear systems we need to compute

$$\begin{aligned} F(x,t) &= \frac{1}{2}g(x,t) + K_1(g)(x,t) \\ &= \frac{1}{2}g(x,t) + \int_\Sigma \frac{\partial}{\partial n_y} G(x-y, t-s)g(y,s)dy\,ds \qquad (x,t) \in \Sigma, \end{aligned}$$

for the indirect method and

$$F(x,t) = g(x,t) \qquad (x,t) \in \Sigma,$$

for the direct method.

So, to assemble the right hand side of the linear system, we need to calculate:

$$\begin{aligned} (b_m)_\alpha &= \int_\Sigma F(x,t)b_\alpha(x)\chi_m(t)dx\,dt \\ &= \int_\Gamma \int_{mh_t}^{(m+1)h_t} F(x,t)b_\alpha(x)dt\,dx. \end{aligned}$$

4.5 Solving the Linear System

The next step is to solve the resulting linear system. Due to the block lower triangular form of the matrix of the single layer operator we can find an efficient solver for the resultant systems. This simple forwards substitution was suggested in [42]. Thus, for every $n \leq N_t$ we solve

$$G_{nn}q_n = F_n - \sum_{k=1}^{n-1} G_{nk}q_k. \tag{4.19}$$

Since the symmetric positive definite matrix G_{nn} is the same for every step n, it can be inverted once and then reused. For large problems evaluating the inverse is costly, in this case we calculate the LU decomposition of the matrix once and then use it to solve efficiently in each step.

We obtain a very simple method for solving the linear system both for the direct and the indirect method. This algorithm only works for constant time steps. A similar algorithm can be used for variable time step size.

```
def solveMem(A,B,Nx,Nt):
    B = B.reshape(-1)
    x = zeros(B.shape)
    for i in range(Nt):
        sumAx = zeros([Nx])
        for k in range(i):
            sumAx += dot(A[(i-k)*Nx:(i-k+1)*Nx,:],
                         x[k*Nx:(k+1)*Nx])
        x[i*Nx:(i+1)*Nx] = solve(A[0:Nx,:],
                                 B[i*Nx:(i+1)*Nx]-sumAx)
    return x
```

Figure 4.7: The algorithm used to solve the linear system (in Python).

4.6 Quadrature Rules in Space

In Sections 4.2 and 4.3 we saw that evaluating the time integrals for the single- and double-layer operators results in double integrals of the form

$$\int_\Gamma \int_\Gamma F(x,y)dx\,dy,$$

where the integrand is given either by

$$g_{mn}(x-y)b_\alpha(x)b_\beta(y) \text{ or } k_{mn}(x-y)b_\alpha(x)b_\beta(y). \tag{4.20}$$

Using the parameterisation γ of the boundary we can easily rewrite this as an integral over the unit square:

$$\begin{aligned}\int_\Gamma \int_\Gamma F(x,y)dx\,dy &= \int_0^1 \int_0^1 F(\gamma^{-1}(\tilde{x}),\gamma^{-1}(\tilde{y}))|\gamma'(\tilde{x})|\cdot|\gamma'(\tilde{y})|d\tilde{x}\,d\tilde{y} \\ &= \int_0^1 \int_0^1 \hat{F}(\tilde{x},\tilde{y})d\tilde{x}\,d\tilde{y},\end{aligned} \tag{4.21}$$

where $\hat{F}(\tilde{x},\tilde{y}) = F(\gamma^{-1}(\tilde{x}),\gamma^{-1}(\tilde{y}))|\gamma'(\tilde{x})|\cdot|\gamma'(\tilde{y})|$.

Since the kernel functions g_{mn} and k_{mn} contain exponential integral functions (see (4.14) and (4.18)) with logarithmic singularities, we need to find an efficient quadrature rule for logarithmic singularities.

There are several ways to evaluate these integrals efficiently. In higher dimensional cases I is a double integral over $d-1$-dimensional parallelotopes. An algorithm for calculating those integrals was given in [10].

4.6.1 One-dimensional Rules

First we will discuss some of the one-dimensional quadrature rules that can be used for the types of integrals that need to be evaluated. In particular, we examine generalised Gauss-Jacobi, Gauss-Laguerre and composite Gauss-Legendre rules for the singular coordinates and a Gauss-Legendre quadrature for the regular coordinates.

Generalised Gauss-Jacobi

Gauss-Jacobi rules are used to integrate functions with singularities at the endpoints. The generalised Gauss-Jacobi rules proposed in [24] generalise these rules so that they integrate functions with logarithmic singularities. In particular, these rules can integrate polynomials of degree up to $2n-1$ multiplied by a logarithmic singularity exactly.

The rule is given by

$$\begin{aligned}\int_0^1 g(x)(1-x)^\alpha x^\beta \log(x)dx &= -\int_0^1 g(x)(1-x)^\alpha x^\beta \log(1/x)dx \\ &= -\sum_{\nu=1}^{n} w_\nu^{(\alpha,\beta)} g(x_\nu^{(\alpha,\beta)}), \qquad \alpha,\beta > -1,\ g \in \mathbb{P}_{2n-1},\end{aligned}$$

where $\mathbb{P}_{2n-1}$ denotes the set of polynomials of degree $\leq 2n-1$.

Further details on the construction of these quadrature rules, as well as code to generate them is given in [23].

Gauss-Laguerre

An alternative to the generalised Gauss-Jacobi quadrature is Gauss-Laguerre quadrature. These rules are defined as follows

$$\int_0^\infty x^\alpha e^{-x} f(x)dx = \sum_{i=1}^{n} w_i f(x_i).$$

We can easily transform our integrand into the form required in order to use these rules:

$$\begin{aligned}\int_0^1 g(x)\log(x)dx &= \int_\infty^0 g(e^{-y})e^{-y}ydy \\ &= -\int_0^\infty y^1 g(e^{-y})e^{-y}dy.\end{aligned}$$

Gauss-Legendre

Gauss-Legendre quadrature is not suitable for singular integrands. We will use this rule for the regular integrands that occur. Let $\{t_i, w_i\}_{i=1}^N$ be N quadrature points and weights respectively. The quadrature points t_i for the quadrature order N are given by the roots of the Legendre polynomials $P_N(x)$. The weights w_i are given by

$$w_i = \frac{2}{NP'_{N-1}(t_i)P'_N(t_i)}. \tag{4.22}$$

Composite Gauss-Legendre

Next we look at a rule which can be used for more general types of singularities. We break up the interval of integration and use the Gauss-Legendre rule described above on each of the intervals.

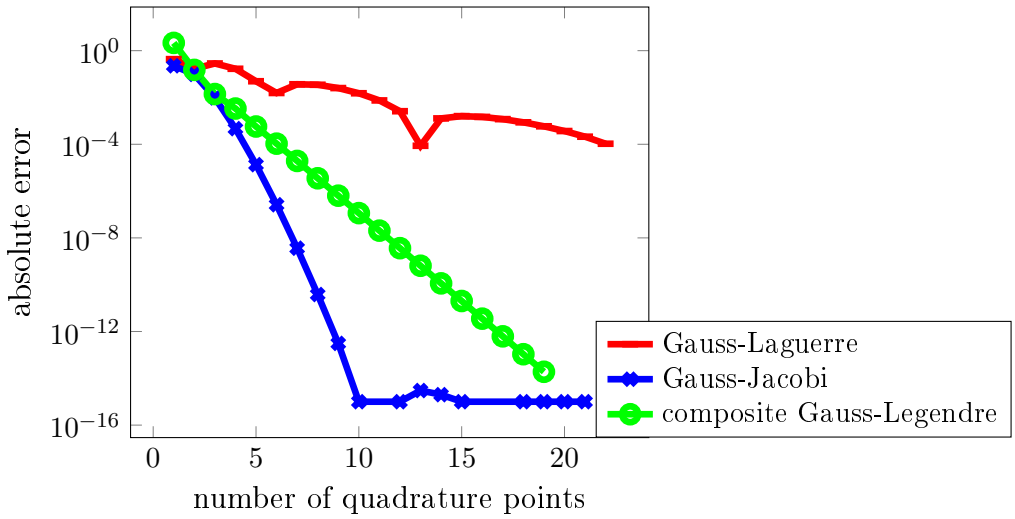

Figure 4.8: A comparison of the convergence of the three one-dimensional quadrature rules for the test function $f(x) = \log(x)(4 + \cos(2\pi x))$.

Let $m \in \mathbb{N}$ and $\sigma \in (0, 1)$. We define the geometric subdivision $[0, 1] = I_1 \cup ... \cup I_m$ with

$$I_j = [\sigma^j, \sigma^{j-1}] \text{ for } j = 1, ..., m-1, \quad \text{and } I_m = [0, \sigma^{m-1}].$$

We then define composite Gauss-Legendre [48] quadrature rules on this subdivision. For $m, n \in \mathbb{N}$ and $\sigma \in (0, 1)$ let I_j be given as above. Let

$$n_j = \left\lceil \frac{n(m+1-j)^\delta}{m^\delta} \right\rceil \text{ for } j = 1...m. \tag{4.23}$$

We define the composite Gauss-Legendre quadrature rule for f as

$$Q_{n,m,\sigma,\delta} f := \sum_{j=1}^{m} Q_{n_j}^{I_j} f, \tag{4.24}$$

where $Q_n^{I_j} f$ is the Gauss-Legendre quadrature rule on the interval I_j.

The composite Gauss-Legendre rule uses n_1 Gauss-Legendre points in the rightmost interval I_1, and a decreasing number of Gauss-Legendre points towards 0. The total number of quadrature points is $\sum_{j=1}^{m} n_j \approx nm/(\delta + 1)$.

Comparing the One-dimensional quadrature rules

Figure 4.6.1 shows a comparison of these three quadrature rules. We see that the generalised Gauss-Jacobi quadrature converges much more quickly for logarithmic

singularities than the other two rules for integrands similar to those that appear in the discretisation of the single and double layer potentials. In all numerical tests generalised Gauss-Jacobi rules were used in the singular coordinates.

4.6.2 Higher-dimensional Rules

In order to create quadrature rules in higher dimensions, we use Duffy transforms. A Duffy transform transforms a triangle to square [20]. We use them to move the logarithmic singularity so that it is only in one coordinate direction. Then we can subtract the singularity leading to a non-singular integral. These transformations are possible in arbitrary dimensions, here we use them for the two-dimensional case.

Starting from the integral (4.21) we first need to seperate the regular summand F_{reg} in the integrand from the summand F_{sing}, which has a logarithmic singularity:

$$\int_0^1 \int_0^1 \hat{F}(\tilde{x},\tilde{y})d\tilde{x}\,d\tilde{y} = \int_0^1 \int_0^1 \tilde{F}_{\text{reg}}(\tilde{x},\tilde{y})d\tilde{x}\,d\tilde{y} + \int_0^1 \int_0^1 \tilde{F}_{\text{sing}}(\tilde{x},\tilde{y})d\tilde{x}\,d\tilde{y}.$$

The first integral can be computed using a Gauss-Legendre rule. In the following sections we will discuss the computation of the second integral depending on the location of the supports of the two basis functions b_α and b_β (see (4.20)).

Identical Elements

In this case the two basis functions b_α and b_β have identical supports. This means that the integrand F_{sing} can be written as $F_{\text{sing}}(x,y) = f(x,y)\log|x-y|$. To isolate the singularity which is currently located on the diagonal of the square $[0,1]^2$, we first divide the domain into to triangles along the diagonal:

$$\begin{aligned} I &= \int_0^1 \int_0^1 f(x,y)\log|x-y|dxdy \\ &= \int_0^1 \int_0^y f(x,y)\log|x-y|dxdy + \int_0^1 \int_y^1 f(x,y)\log|x-y|dxdy =: I_1 + I_2. \end{aligned}$$

Then, using the Duffy transform $x = (1-t)y$ for the first summand and the transform

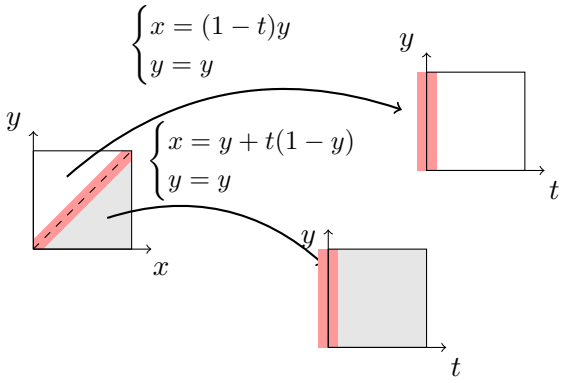

Figure 4.9: Division of the square into two triangles and the Duffy-transformation of each triangle to a square.

$x = y + t(1-y)$ for the second summand as in Figure 4.9, we get:

$$\int_0^1 \int_y^1 f(x,y) \log|x-y| dxdy = \int_0^1 \int_0^1 \log|t| f(y+(1-y)t, y)(1-y) dtdy + \int_0^1 \int_0^1 \log|y| f(y+1+yt, y+1) y dtdy.$$

Adjacent elements

In this case the two basis functions b_α and b_β have supports which coincide in one point. Depending on the location of this point the integral needs to be handled differently. First there are singularities in the left upper corner of the square given by the tensor product of the two intervals. In this case the integrand can be written as $F_{\text{sing}}(x,y) = f(x,y)\log(1+x-y)$. This gives integrals of the form:

$$\begin{aligned} I &= \int_0^1 \int_0^1 f(x,y) \log|1+x-y| dxdy \\ &= \int_0^1 \int_0^{1-y} f(x,y) \log|1+x-y| dxdy + \int_0^1 \int_{1-y}^1 f(x,y) \log|1+x-y| dxdy \\ &=: I_1 + I_2. \end{aligned}$$

Then using the transformation $x = (1-y)t$ on the first summand we get:

$$\begin{aligned} I_1 &= \int_0^1 \int_0^{1-y} f(x,y) \log|1+x-y| dxdy \\ &= \int_0^1 \int_0^1 \log|y| f(yt, 1-y) y dt dy \\ &\quad + \int_0^1 \int_0^1 \log|t+1| f((1-y)t, y)(1-y) dt dy. \end{aligned}$$

Further, using the transformation $y = sx + 1 - x$ on the second summand we get:

$$\begin{aligned} I_2 &= \int_0^1 \int_{1-y}^1 f(x,y) \log|1+x-y| dxdy \\ &= \int_0^1 \int_{1-y}^1 \log|x| f(x, sx+1-x) x dx ds \\ &\quad + \int_0^1 \int_{1-y}^1 \log|2-s| f(x, sx+1-x) x dx ds. \end{aligned}$$

Singularities in the right upper corner correspond to the second case of adjacent elements in which the first element is to right of the second element. To isolate the singularity in this case, the form of the integrand needs to be $F_{\text{sing}}(x,y) = \log(-1+x-y) f(x,y)$. This gives integrals of the form:

$$\begin{aligned} I &= \int_0^1 \int_0^{1-y} f(x,y) \log|x-y-1| dxdy \\ &= \int_0^1 \int_0^{1-y} f(x,y) \log|x-y-1| dxdy + \int_0^1 \int_{1-y}^1 f(x,y) \log|x-y-1| dxdy \\ &=: I_1 + I_2. \end{aligned}$$

Then using the transformation $y = s(1-x)$ on the first summand we get:

$$\begin{aligned} I_1 &= \int_0^1 \int_{1-y}^1 f(x,y) \log|-1+x-y| dxdy \\ &= \int_0^1 \int_{1-y}^1 \log|x| f(1-x, sx) x dx ds \\ &\quad + \int_0^1 \int_{1-y}^1 \log|s+1| f(x, s(1-x))(1-x) dx ds. \end{aligned}$$

Further, using the transformation $x = yt + 1 - y$ on the second summand we get:

$$\begin{aligned} I_2 &= \int_0^1 \int_0^{1-y} f(x,y) \log|-1+x-y| dxdy \\ &= \int_0^1 \int_0^1 \log|y| f(yt+1-y,y) y dt dy \\ &\quad + \int_0^1 \int_0^1 \log|2-t| f(yt+1-y,y) y dt dy. \end{aligned}$$

4.7 Numerical Experiments

In this section we compare the convergence of a boundary element discretisation with that of a finite element discretisation of the same problem. For this comparison we choose a homogeneous problem with Dirichlet boundary conditions (2.3). It is formulated as follows

$$\begin{aligned} (\partial_t - \Delta)u &= 0, && \text{in } \mathcal{I} \times \Omega \\ u &= 0, && \text{at } \{t=0\} \times \Omega \\ \gamma_0 u &= g, && \text{in } \Sigma. \end{aligned}$$

Definition 4.7.1. *The circle of radius R, centered around x is denoted by*

$$B_R(x) := \{(y_1, y_2) \ : \ (y_1 - x_1)^2 + (y_2 - x_2)^2 \leq R^2\}.$$

As a domain we choose a circle of radius 1, i.e. $\Omega = B_1(0)$. Since the exact solution is known for this particular problem, we choose as a right hand side $g(r, \varphi, t) = t^2 \cos(\varphi)$. According to [42] the exact solution is:

$$u(r,\varphi,t) = \left(rt^2 - 4 \sum_{k=1}^{\infty} \frac{J_1(\beta_k r)}{\beta_k^3 J_2(\beta_k)} (t - \frac{1}{\beta_k^2}(1 - e^{-\beta_k^2 t})) \right) \cos(\varphi), \tag{4.25}$$

By taking the normal derivative we easily see that the boundary flux is

$$q(r,\varphi,t) = \left(t^2 - \frac{1}{4}t + 4 \sum_{k=0}^{\infty} \frac{1 - e^{-\beta_k^2 t}}{\beta_k^4} \right) \cos(\varphi). \tag{4.26}$$

4.7.1 Finite Element Implementation

The finite element discretisation requires a time-stepping scheme. We choose a Crank-Nicolson scheme. Crank-Nicolson is a second-order method, that is implicit in time. Discretising only in time gives a semi-discrete scheme. The semi-discrete scheme is given as follows

$$\frac{u_{n+1} - u_n}{h_t} = \frac{1}{2}\left[\Delta u_{n+1} + \Delta u_n\right] + \frac{1}{2}\left[f_{n+1} + f_n\right],$$

where as before $u_n = u(t_n, \cdot)$ and $t_n = nh_t$.

For the volume mesh in space we use piecewise linear basis functions on a mesh of triangles. Since our domain is a circle we approximate its boundary by a polygon and then discretise with triangles. A sample mesh is shown in Figure 4.10.

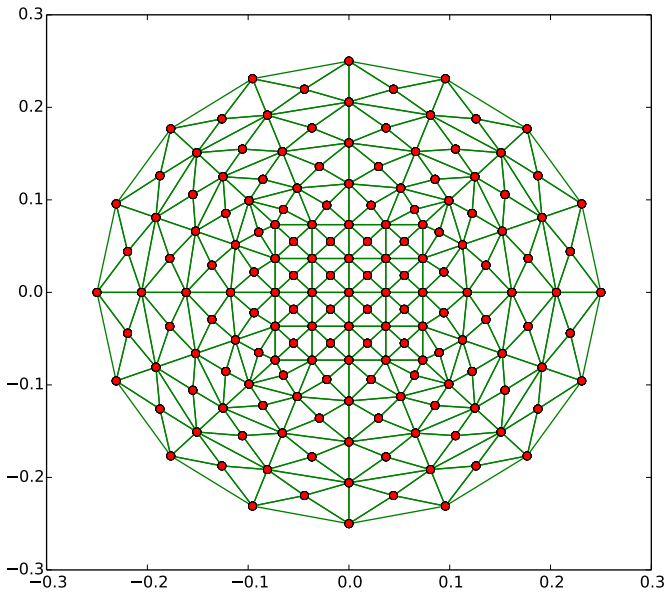

Figure 4.10: The FE mesh used on the domain $B_1(0)$. The nodes on the boundary are marked in green, while the inner nodes are marked in red.

Discretising in space as well as in time leads to the following fully discrete system:

$$\left(M + \frac{1}{2}h_t A\right) u_{n+1} = h_t B + M u_n - \frac{1}{2}h_t A u_n.$$

We denote the piecewise linear FE basis functions by $b_j : \Omega \to \mathbb{R}$. Then, M is the

mass matrix, given by

$$M_{jk} = \langle b_j, b_k \rangle,$$

A is the stiffness matrix, given by

$$A_{jk} = \langle \nabla b_j, \nabla b_k \rangle,$$

and B is the vector of the right hand side, given by

$$B_j = \frac{1}{2} \left(\langle f_n, b_j \rangle + \langle f_{n+1}, b_j \rangle \right).$$

Then to solve using FEM we need to find a smooth extension of g from $\partial\Omega$ to Ω. This extensions is not uniquely defined. We denote this extension by $\tilde{g}$.

Next we rewrite equation (2.3) such that it fulfills zero Dirichlet boundary conditions. Set $\tilde{u} = u - \tilde{g}$ in (2.3) and solve

$$\begin{aligned} \partial_t \tilde{u} - \Delta \tilde{u} = -(\partial_t - \Delta)\tilde{g} =: f \qquad & \text{in } \mathcal{I} \times \Omega \\ \tilde{u} = 0 \qquad & \text{at } \{t = 0\} \times \Omega \\ \gamma_0 \tilde{u} = 0 \qquad & \text{in } \Sigma \end{aligned} \tag{4.27}$$

Here we give two alternatives for the choice of extension $\tilde{g}$ to $g(r, \varphi, t) = t^2 \cos(\varphi)$.

Alternative 1: Use the extension

$$g(r, \varphi, t) = r^2 t^2 \cos(\varphi).$$

It follows that the right hand side is given by

$$f(r, \varphi, t) = -(2t - 3t^2) \cos(\varphi)$$

Alternative 2: Use the extension

$$g(x, y, t) = t^2 x.$$

It follows that the right hand side is given by

$$f(x, y, t) = -2tx.$$

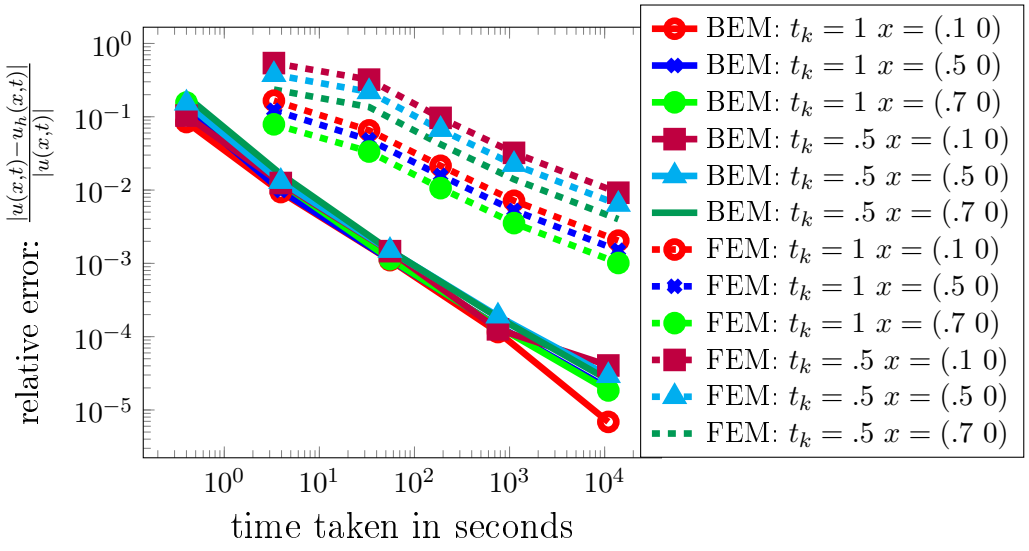

Figure 4.11: The pointwise error plotted against time taken in seconds for a BEM versus a FEM implementation.

4.7.2 Comparison between FEM and BEM

In this section we compare the error of the FE discretisation described in Section 4.7.1 with $h_t \sim h_x$ to the error of the BE discretisation of the same problem. We compare the pointwise error at several time and space coordinates in the domain and we compare the convergence of the boundary flux in the $L^2(\Gamma)$-norm at different points in time.

The BE discretisation used for this test uses piecewise constant polynomial basis functions in time and space with $h_t \sim h_x$. The discretisation does not use wavelets or a sparse grid discretisation.

Figure 4.11 shows the absolute pointwise error at several different points in time and space plotted against the time taken. We see that the BE method converges to the exact solution more quickly than the FE discretisation. However, if one were to compare the computation of the solution in the entire domain an finite element implementation would be faster, since evaluating the representation formula requires the numerical solution of a double integral.

Next we compare the $L^2(\Gamma)$-error of the boundary flux at certain points in time. Since we used piecewise linear basis functions in space for the FE discretisation it is easy to calculate an approximation to the boundary flux. We use a four-point forward finite difference stencil for the approximation. For the BE implementation

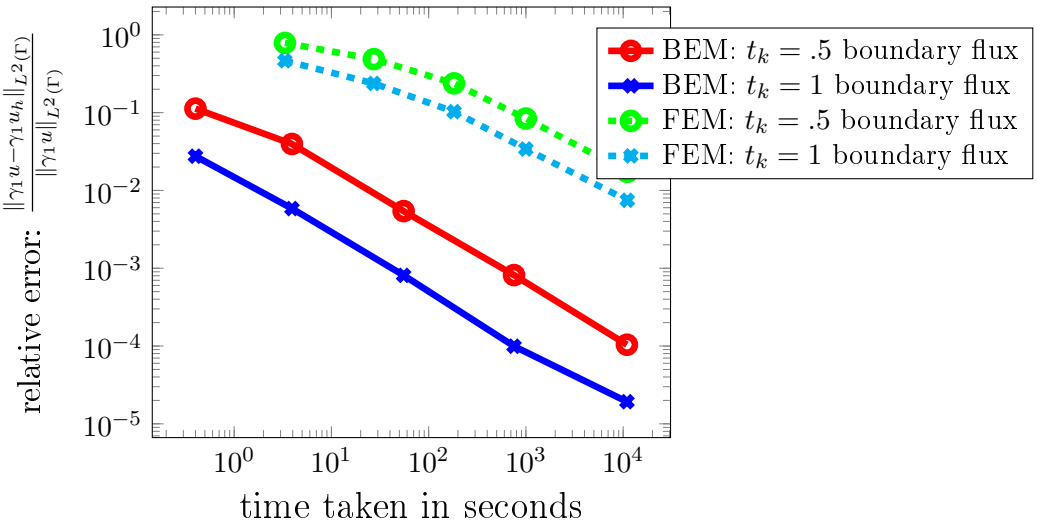

Figure 4.12: The L^2-error of the boundary flux plotted against time taken in seconds for a BEM versus a FEM implementation.

the boundary flux is calculated directly and no post-processing is necessary.

Figure 4.12 shows the $L^2(\Gamma)$-error of the boundary flux at two different points in time. Again the BE implementation is faster and shows a higher rate of convergence.

All in all, we conclude that using a boundary element discretisation is particularly beneficial when the boundary flux is the quantity of interest. Boundary elements are also useful when the solution needs to be evaluated at only a few points. However, if the solution is needed in the interior of the entire domain a FE implementation may be the better choice. For an outside domain, which is not bounded, BEM offers an easy alternative to FEM.

Chapter 5

Error Analysis for Full Tensor Product Approximation Spaces

In this chapter we give some basic results of the error analysis for the boundary integral formulation of the heat equation. First we summarise the classical results from [15] and [42] for different choices of polynomial degrees. Then we give new results obtained for the case of identical polynomial degrees in time and space.

The results of this chapter are for full tensor product discretisations with piecewise polynomial basis functions. Results on the error analysis for sparse grid spaces can be found in Chapter 6.

5.1 L^2- orthogonal Projections

Throughout this and the following chapters we will require the properties of L^2-orthogonal operators.

Let $\mathcal{X}$ be a closed subspace of $L^2(\Sigma)$. Then there exists a uniquely defined projection operator

$$\Pi_{\mathcal{X}} \, : \, L^2(\Sigma) \to \mathcal{X},$$

such that

$$\langle f, g - \Pi_{\mathcal{X}} g \rangle = 0 \qquad \forall f \in \mathcal{X}, \; g \in L^2(\Sigma).$$

Definition 5.1.1. *We refer to the projection*

$$\Pi_{\mathcal{X}} \, : \, L^2(\Sigma) \to \mathcal{X}$$

defined above as the L^2-orthogonal projection.

5.2 Classical Error Estimates

In the following we give some classical results on the approximation properties of piecewise polynomial full tensor product spaces $\mathcal{X}_L = \mathcal{X}_{l_x} \otimes \mathcal{X}_{l_t} \subset H^{p,q}(\Sigma)$.

The following well-known theorem on the convergence in the energy norm is taken from [42].

Theorem 5.2.1. *Let $\psi_L \in \mathcal{X}_L$ be the Galerkin approximation to the Dirichlet problem and let $\psi \in H^{p_x+1,p_t+1}(\Sigma)$ be the solution. Here p_x and p_t are the polynomial degrees of the spaces $\mathcal{X}_{l_x}$ and $\mathcal{X}_{l_t}$ respectively. Then*

$$\|\psi - \psi_L\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq c(h_x^{\frac{1}{2}} + h_t^{\frac{1}{4}})(h_x^{p_x+1} + h_t^{p_t+1})\|\psi\|_{H^{p_x+1,p_t+1}(\Sigma)}.$$

We give the detailed proof to illustrate how the improvements of the next section can be attained. In particular, note that the Aubin-Nitsche argument used in the following proof is not sharp.

The proof of this theorem requires knowledge of the approximation properties of the L^2-projection operators to the spaces $\mathcal{X}_{l_x}$ and $\mathcal{X}_{l_t}$. These are denoted by $\Pi_{\mathcal{X}_{l_x}}$ and $\Pi_{\mathcal{X}_{l_t}}$ respectively. The polynomial degrees of the spaces $\mathcal{X}_{l_x}$ and $\mathcal{X}_{l_t}$ are p_x and p_t respectively and the mesh widths in the spaces are h_x and h_t.

Lemma 5.2.2 (Section 5, [15]). *Let β_1, β_2 satisfy*

$$-(p_t+1) \leq \beta_1 < \beta_2 \leq p_t + 1,\ \beta_2 > -1/2 \text{ and } \beta_1 < 1/2.$$

Then,

$$\|u - \Pi_{\mathcal{X}_{l_t}} u\|_{H^{\beta_1}(I)} \leq c h_t^{\beta_2-\beta_1} \|u\|_{H^{\beta_2}(I)}, \quad u \in H^{\beta_2}(I).$$

Further, let α_1, α_2 satisfy

$$-(p_x+1) \leq \alpha_1 < \alpha_2 \leq p_x + 1,\ \alpha_2 > -1/2 \text{ and } \alpha_1 < 1/2.$$

Then,

$$\|u - \Pi_{\mathcal{X}_{l_x}} u\|_{H^{\alpha_1}(\Gamma)} \leq c h_x^{\alpha_2-\alpha_1} \|u\|_{H^{\alpha_2}(\Gamma)}, \quad u \in H^{\alpha_2}(\Gamma).$$

For ease of notation we denote by $\Pi_{\mathcal{X}_{l_x}}$ also the projection:

$$(\Pi_{l_x} u)(x,t) = (\Pi_{\mathcal{X}_{l_x}} u(x,\cdot))(t), \quad \text{for } x \in \Gamma.$$

Analogously,

$$(\Pi_{l_t} u)(x,t) = (\Pi_{\mathcal{X}_{l_t}} u(\cdot,t))(x), \quad \text{for } t \in \mathcal{I}.$$

Then $\Pi_{\mathcal{X}_{l_x}}\Pi_{\mathcal{X}_{l_t}} = \Pi_{\mathcal{X}_{l_t}}\Pi_{\mathcal{X}_{l_x}}$ is the $L^2(\Sigma)$-orthogonal projection onto $\mathcal{X}_L = \mathcal{X}_{l_x} \otimes \mathcal{X}_{l_t}$.

Combining the two estimates from Lemma 5.2.2:

Lemma 5.2.3 (Proposition 5.3, [15]). *Let λ, μ, r, s denote values satisfying*

$$\begin{aligned} -p_x \leq \lambda \leq 0 \leq r \leq p_x + 1 \text{ and} \\ -p_t \leq \mu \leq 0 \leq s \leq p_t + 1. \end{aligned}$$

Then, for all $u \in H^{r,s}(\Sigma)$, there exists $c \geq 0$ which depends on λ, μ, r, s such that

$$\|u - \Pi_{\mathcal{X}_{l_x}}\Pi_{\mathcal{X}_{l_t}} u\|_{H^{\lambda,\mu}(\Sigma)} \leq c(h_x^{-\lambda} + h_t^{-\mu})(h_x^r + h_t^s)\|u\|_{H^{r,s}(\Sigma)},$$

where $\Pi_{\mathcal{X}_{l_x}}, \Pi_{\mathcal{X}_{l_t}}$ are the L^2 projections on to $\mathcal{X}_{l_x}$ and $\mathcal{X}_{l_t}$ respectively.

Proof. For this proof λ, μ, r, s are fixed. Remember that $\lambda, \mu \leq 0$.

Adding zero gives $u - \Pi_{\mathcal{X}_{l_x}}\Pi_{\mathcal{X}_{l_t}} u = (u - \Pi_{\mathcal{X}_{l_x}} u) + \Pi_{\mathcal{X}_{l_x}}(u - \Pi_{\mathcal{X}_{l_t}} u)$. Using the triangle inequality and Lemma 5.2.2 we get

$$\begin{aligned} \|u - \Pi_{\mathcal{X}_{l_x}}\Pi_{\mathcal{X}_{l_t}} u\|_{L^2(\Sigma)} &\leq \|u - \Pi_{\mathcal{X}_{l_x}} u\|_{L^2(\Sigma)} + \|\Pi_{\mathcal{X}_{l_x}}(u - \Pi_{\mathcal{X}_{l_t}} u)\|_{L^2(\Sigma)} \\ &\leq c h_x^r \|u\|_{L^2(I, H^r(\Gamma))} + h_t^s \|u\|_{H^s(I, L^2(\Gamma))}. \end{aligned}$$

It follows,

$$\|u - \Pi_{\mathcal{X}_{l_x}}\Pi_{\mathcal{X}_{l_t}} u\|_{L^2(\Sigma)} \leq c(h_x^r + h_t^s)\|u\|_{H^{r,s}(\Sigma)}. \tag{5.1}$$

Then we use an Aubin-Nitsche argument to get

$$\begin{aligned} \|u - \Pi_{\mathcal{X}_{l_x}}\Pi_{\mathcal{X}_{l_t}} u\|_{H^{\lambda,\mu}(\Sigma)} &= \sup_{v \in \tilde{H}^{-\lambda,-\mu}(\Sigma)} \frac{|\langle u - \Pi_{\mathcal{X}_{l_x}}\Pi_{\mathcal{X}_{l_t}} u, v\rangle|}{\|v\|_{H^{-\lambda,-\mu}(\Sigma)}} \\ &= \sup_{v \in \tilde{H}^{-\lambda,-\mu}(\Sigma)} \frac{|\langle u, v - \Pi_{\mathcal{X}_{l_x}}\Pi_{\mathcal{X}_{l_t}} v\rangle|}{\|v\|_{H^{-\lambda,-\mu}(\Sigma)}} \\ &\leq \|u\|_{L^2(\Sigma)} \sup_{v \in \tilde{H}^{-\lambda,-\mu}(\Sigma)} \frac{\|v - \Pi_{\mathcal{X}_{l_x}}\Pi_{\mathcal{X}_{l_t}} v\|_{L^2(\Sigma)}}{\|v\|_{H^{-\lambda,-\mu}(\Sigma)}} \\ &\leq c(h_x^{-\lambda} + h_t^{-\mu})\|u\|_{L^2(\Sigma)}. \end{aligned}$$

We note that $(Id - \Pi_{\mathcal{X}_{l_x}}\Pi_{\mathcal{X}_{l_t}}) = (Id - \Pi_{\mathcal{X}_{l_x}}\Pi_{\mathcal{X}_{l_t}})^2$ and get

$$\begin{aligned} \|u - \Pi_{\mathcal{X}_{l_x}}\Pi_{\mathcal{X}_{l_t}} u\|_{H^{\lambda,\mu}(\Sigma)} &= \|(Id - \Pi_{\mathcal{X}_{l_x}}\Pi_{\mathcal{X}_{l_t}})^2 u\|_{H^{\lambda,\mu}(\Sigma)} \\ &\overset{(5.1)}{\leq} c(h_x^{-\lambda} + h_t^{-\mu})\|u - \Pi_{\mathcal{X}_{l_x}}\Pi_{\mathcal{X}_{l_t}} u\|_{L^2(\Sigma)} \\ &\leq c(h_x^{-\lambda} + h_t^{-\mu})(h_x^r + h_t^s)\|u\|_{H^{r,s}(\Sigma)} \end{aligned}$$

as asserted. □

In Chapter 2 we showed the coercivity of the single layer operator. So, using the classical Lemma of Céa and Galerkin orthogonality Theorem 5.2.1 follows directly from this Lemma.

This theorem can be applied to different choices of polynomial degrees. The term $(h_x^{-\lambda} + h_t^{-\mu})$ in the estimate is determined by the $H^{\mu,\lambda}(\Sigma)$-norm in the left-hand side of the estimate. For all further estimates we will choose $\lambda = -\frac{1}{2}$ and $\mu = -\frac{1}{4}$, leading to estimates in the energy norm of our problem.

Then we need to balance the term $(h_x^{-\lambda} + h_t^{-\mu})$ with the term $(h_x^r + h_t^s)$. If our right hand side is assumed to be arbitrarily smooth, the only restrictions on r and s come from the choice of polynomial degree. Due to Theorem 5.2.1 we have the restrictions $r \leq p_t + 1$ and $s \leq p_x + 1$. If we choose $p_x = 2p_t + 1$, then s can be at most $p_t + 1$ and r at most $p_x + 1 = 2p_t + 2 = 2s$. This leaves us with two terms of the same form and Theorem 5.2.1 gives

$$\|\psi - \psi_L\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq c(h_x^2 + h_t)^{s+\frac{1}{4}} \|\psi\|_{H^{2s,s}(\Sigma)}$$

for a scaling of $h_x^2 \sim h_t$. For fixed polynomial degrees p_x and p_t the total number of degrees of freedom N is proportional to $h_x^{-(d-1)} h_x^{-2} = h_x^{-(d+1)}$. Rewriting the convergence estimate with respect to degrees of freedom gives

$$\|\psi - \psi_L\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq cN^{-2(s+\frac{1}{4})/(d+1)} \|\psi\|_{H^{2s,s}(\Sigma)},$$

with a constant $c > 0$ depending on the polynomial degrees p_x and p_t.

With the restriction $p_x = 2p_t + 1$ the basis functions in time and space can not be chosen independently. In particular, at least piecewise linear basis functions must be chosen in space. However, from an implementational standpoint it is easiest to work with low polynomial degrees both in time and space.

We are mainly interested in the case of $p_x = p_t = 0$, i.e. piecewise constant basis functions in time and space. These are easiest to implement and they result in a block Toeplitz structure of the matrix, leading to an easily solvable linear system. Further, piecewise constant basis functions allow analytic evaluation of the time integrals. This was detailed in Chapter 4.

When we no longer have the restriction $p_x = 2p_t + 1$ the optimal scaling between

h_x and h_t is not clear. In the following we find the optimal scaling for the case $p_x = p_t$ and then apply it to the case of piecewise constant basis functions. In the next section we will improve further upon these results.

Let $s = p_x + 1 = p_t + 1$. Then we have

$$\|\psi - \psi_L\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq c(h_x^{\frac{1}{2}} + h_t^{\frac{1}{4}})(h_x^s + h_t^s)\|\psi\|_{H^{s,s}(\Sigma)}.$$

We let σ parameterise the scaling between h_x and h_t by $h_t \sim h_x^\sigma$. Now, our goal is to find the value of σ for which the upper bound on the error of the energy norm in the above estimate is smallest. This means we need to minimise the expression with regard to σ.

Clearly,

$$\begin{aligned}(h_x^{\frac{1}{2}} + h_t^{\frac{1}{4}})(h_x^s + h_t^s) &= (h_x^{\frac{1}{2}} + h_x^{\frac{\sigma}{4}})(h_x^s + h_x^{s\sigma}) \\ &= h_x^{\frac{1+2s}{2}} + h_x^{\frac{1+2s\sigma}{2}} + h_x^{\frac{4s+\sigma}{4}} + h_x^{\frac{4s+1}{4}\sigma}.\end{aligned}$$

This means we need to find

$$m := \min\left\{1 + 2s, 1 + 2s\sigma, \frac{4s+\sigma}{2}, \frac{4s+1}{2}\sigma\right\}.$$

Lemma 5.2.4. *For any $d > 2$ the minimum m is given by*

$$\min\left\{1 + 2s, 1 + 2s\sigma, \frac{4s+\sigma}{2}, \frac{4s+1}{2}\sigma\right\} = \begin{cases} \frac{4s+1}{2}\sigma, & \sigma \leq 1 \\ \frac{4s+\sigma}{2}, & 1 < \sigma \leq 2 \\ 1 + 2s, & \textit{else.} \end{cases}$$

Proof. First we note that since $\frac{4s+1}{2}\sigma \leq 1 + 2s\sigma$ for $\sigma \leq 2$ and $1 + 2s \leq 1 + 2s\sigma$ for $\sigma \geq 1$, we can simplify the minimum by removing $1 + 2s\sigma$. So we find

$$m = \min\left\{1 + 2s, \frac{4s+\sigma}{2}, \frac{4s+1}{2}\sigma\right\}.$$

We easily see that

$$\frac{4s+1}{2}\sigma \leq 1 + 2s \Leftrightarrow \sigma \leq \frac{2+4s}{1+4s}.$$

Further,

$$\frac{4s+1}{2}\sigma \leq \frac{4s+\sigma}{2} \Leftrightarrow \sigma \leq 1.$$

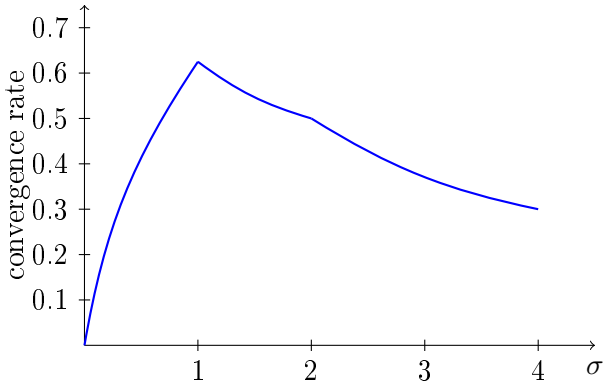

Figure 5.1: The convergence rate in the energy norm plotted against the value of σ for $d = 2$ and $s = 1$. The maximum is attained at $\sigma = 1$.

And finally,

$$\frac{4s+\sigma}{2} \leq 1 + 2s \Leftrightarrow \sigma \leq 2.$$

This concludes the proof □

We start by examining the case $s = 1$, i.e. piecewise constant basis functions, since this is the case we are most interested in. In this case the convergence rate with respect to the number of degrees of freedom in the energy norm is given as $\frac{m}{2(d-1+\sigma)}$ as shown in Figure 5.1 for $d = 2$.

As we can see in Figure 5.1 the choice leading to the highest convergence rate for $s = 1$ is $\sigma = 1$, giving:

$$\|\psi - \psi_L\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq cN_L^{-\frac{5}{4d}} \|\psi\|_{H^{1,1}(\Sigma)}. \tag{5.2}$$

Remark 5.2.5. *This demonstrates that for $s = 1$ the optimal scaling in time and space suggested by Theorem 5.2.1 is $h_t \sim h_x$.*

Now we look at the remaining cases, where $s \in \mathbb{N}$, $s > 1$. The results in these cases are very similar to those when $s = 1$.

Again we examine the convergence rate in the energy norm $\frac{m}{2(d-1+\sigma)}$, which is shown in Figure 5.2 for a few values of s. We note that for these values of s the scaling σ which leads to the highest convergence rate in the energy norm is 1, i.e. we choose $h_t \sim h_x$.

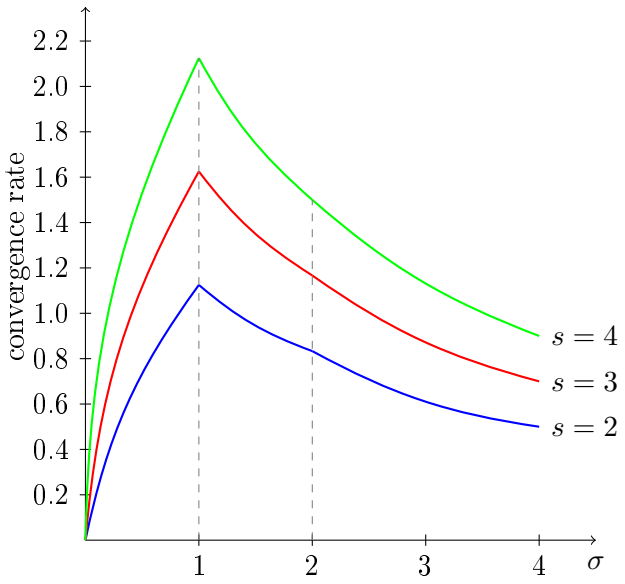

Figure 5.2: The convergence rate in the energy norm plotted against the value of σ for $d = 2$ and $s = 2, 3, 4$.

In the next section we will improve upon these convergence rates, this will also lead to a different choice of optimal scaling.

Remark 5.2.6. *Let $s \geq 1$ and let the dimension d be 2 or 3. In the interval $\sigma \in [0, 1]$ the convergence rate is given by $\frac{4s+1}{4}\frac{\sigma}{d-1+\sigma}$, which is monotonically increasing. Since further both $\frac{4s+\sigma}{4(d-1+\sigma)}$ and $\frac{1+2s}{2(d-1+\sigma)}$ are monotonically decreasing for $\sigma > 1$ the maximum must indeed be indeed achieved at $\sigma = 1$.*

According to Lemma 5.2.4 the estimate for $\sigma = 1$ in the energy norm is

$$\|\psi - \psi_L\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq c N_L^{-\frac{4s+1}{4d}} \|\psi\|_{H^{s,s}(\Sigma)},$$

with the constant $c = c(p_x, p_t) > 0$.

In Table 5.1 we give a summary of these convergence rates for different polynomial degrees in time and space for 2 and 3 dimensions. We also summarise the optimal scalings for these cases.

Full tensor product, $d = 2$			Full tensor product, $d = 3$		
(p_x, p_t)	conv. rate γ	scaling σ	(p_x, p_t)	conv. rate γ	scaling σ
$(0,0)$	$\frac{5}{8}$	1	$(0,0)$	$\frac{5}{12}$	1
$(1,0)$	$\frac{5}{6}$	2	$(1,0)$	$\frac{5}{8}$	2
$(1,1)$	$\frac{9}{8}$	1	$(1,1)$	$\frac{3}{4}$	1
$(2,2)$	$\frac{13}{8}$	1	$(2,2)$	$\frac{13}{12}$	1
$(3,1)$	$\frac{3}{2}$	2	$(3,1)$	$\frac{9}{8}$	2
$(3,3)$	$\frac{17}{8}$	1	$(3,3)$	$\frac{17}{12}$	1

Table 5.1: Convergence rates and optimal scaling σ for full tensor product discretisation in 2 and 3 dimensions.

5.3 Error Bounds for Equal Polynomial Degrees in Time and Space

In this section we find error bounds for the convergence rate of full tensor product Galerkin BEM, where $p_x = p_t$, that are sharper than those obtained with the classical results in the previous section. These results are new to this work.

The main ingredient used for the new proof are norm equivalences which can be shown using wavelet bases. The theory behind these is summarised in Chapter 3.

In particular, Theorem 3.3.5 gives that for $u \in H^{r,s}(\Sigma)$ with $u = \sum_{(l_x,l_t)\geq 0} w_{l_x,l_t}$ and $w_{l_x,l_t} \in W_{l_x} \otimes W_{l_t}$, we have

$$\|u\|^2_{H^{r,s}(\Sigma)} \sim \sum_{l_x,l_t} 2^{2\max\{rl_x,sl_t\}} \|w_{l_x,l_t}\|^2_{L^2(\Sigma)}. \tag{5.3}$$

The norm equivalences given above deliver upper and lower bounds for our estimates. This means that our estimates are sharper than the estimates derived using an Aubin-Nitsche argument. Now we use the norm equivalences to calculate new error bounds.

We define the full tensor product index set as follows

$$I_L^\sigma = \{(l_x, l_t) \ : \ l_x \leq L, \ l_t \leq \sigma L\},$$

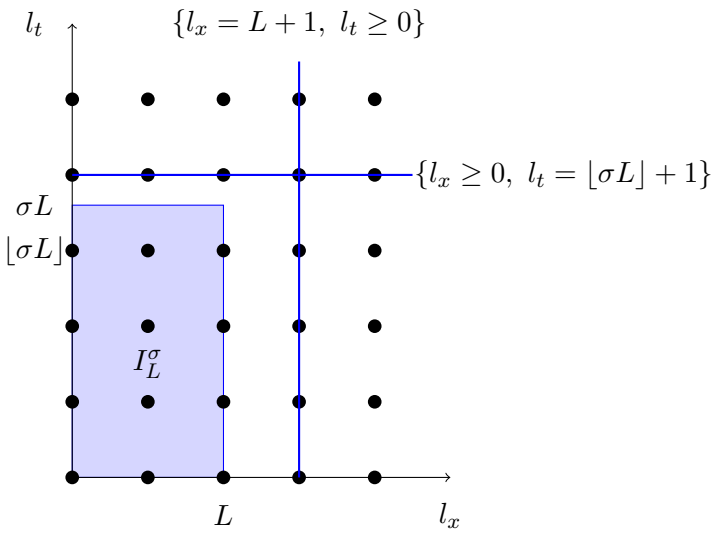

Figure 5.3: The full tensor product index set I_L^σ.

where σ is a real parameter that can be chosen. As before it denotes the scaling in time and space. This means

$$(l_x, l_t) \notin I_L^\sigma \Leftrightarrow (l_x, l_t) \in \{l_x \geq L+1\} \cup \{l_t \geq \lfloor \sigma L \rfloor + 1\}.$$

The index set is shown in Figure 5.3. Now we write $u = \sum_{(l_x,l_t)\geq 0} w_{(l_x,l_t)}$. Then by using an extension to Theorem 3.3.5 and letting $v = \sum_{(l_x,l_t)\in I_L^\sigma} w_{(l_x,l_t)} \in \mathcal{X}_L$ we get

$$\begin{aligned}\inf_{v\in\mathcal{X}_L} \|u-v\|^2_{H^{r,s}(\Sigma)} &\sim \sum_{(l_x,l_t)\notin I_L^\sigma} 2^{2\max\{rl_x,sl_t\}} \|w_l\|^2_{L^2(\Sigma)} \\ &= \sum_{(l_x,l_t)\notin I_L^\sigma} 2^{2\max\{rl_x,sl_t\}-2\max\{\mu l_x,\lambda l_t\}} 2^{2\max\{\mu l_x,\lambda l_t\}} \|w_l\|^2_{L^2(\Sigma)}.\end{aligned}$$

This can be estimated as follows

$$\inf_{v\in\mathcal{X}_L} \|u-v\|^2_{H^{r,s}(\Sigma)} \leq \left(\max_{(l_x,l_t)\notin I_L^\sigma} 2^{2\max\{rl_x,sl_t\}-2\max\{\mu l_x,\lambda l_t\}}\right) \cdot \underbrace{\sum_{(l_x,l_t)\notin I_L^\sigma} 2^{2\max\{\mu l_x,\lambda l_t\}} \|w_l\|^2_{L^2(\Sigma)}}_{\sim \|u\|^2_{H^{\mu,\lambda}(\Sigma)}}.$$

This means we need to estimate the maximum in order to get the convergence rates.

Since we want to find estimates for the energy norm we set $r = -\frac{1}{2}$, $s = -\frac{1}{4}$.

$$\max_{(l_x,l_t)\notin I_L^\sigma} 2^{2\max\{rl_x,sl_t\}-2\max\{\mu l_x,\lambda l_t\}} = \max_{(l_x,l_t)\notin I_L^\sigma} 2^{-\max\{l_x,\frac{l_t}{2}\}-2\max\{\mu l_x,\lambda l_t\}}$$

The term $2^{-(\max\{l_x,\frac{l_t}{2}\}+2\max\{\mu l_x,\lambda l_t\})}$ reaches its maximum when the negative exponent is as small as possible. We define

$$G(l_x,l_t) := \max\left\{l_x,\frac{l_t}{2}\right\} + 2\max\{\mu l_x,\lambda l_t\}. \tag{5.4}$$

Then we need to find

$$n := \min_{(l_x,l_t)\notin I_L^\sigma} G(l_x,l_t).$$

To find this minimum we use some properties of monotonically increasing functions.

Definition 5.3.1. *The function $F(l_x,l_t)$ is a monotonically increasing function if*

$$\begin{aligned} F(l_x+k,l_t) &\geq F(l_x,l_t), \qquad \forall k \geq 0 \\ F(l_x,l_t+k) &\geq F(l_x,l_t), \qquad \forall k \geq 0. \end{aligned}$$

Lemma 5.3.2. *Let F be a monotonically increasing function. Then its minimum outside the set I_L^σ is*

$$\min_{(l_x,l_t)\notin I_L^\sigma} F(l_x,l_t) = \min\{F(L+1,0), F(0,\lfloor\sigma L\rfloor+1\}.$$

Proof. Let $l_x \geq L+1$ Then there holds

$$F(l_x,l_t) \geq F(L+1,l_t)$$

by definition of monotonically increasing. Analogously if we let $l_t \geq \lfloor\sigma L\rfloor+1$, there holds

$$F(l_x,l_t) \geq F(l_x,\lfloor\sigma L\rfloor+1)$$

Together this tells us that the minimum must lie in the subset

$$\{(l_x,l_t) \ : \ l_x = L+1 \text{ or } l_t = \lfloor\sigma L\rfloor+1\} \subset \{(l_x,l_t) \notin I_L\sigma\}.$$

In Figure 5.3 this subset is depicted by the blue lines.

Now let $l_x = L + 1$ and $l_t \geq 0$, then there holds

$$F(l_x, l_t) \geq F(L + 1, 0).$$

Analogously, for $l_x \geq 0$ and $l_t = \lfloor \sigma L \rfloor + 1$ we have

$$F(l_x, l_t) \geq F(0, \lfloor \sigma L \rfloor + 1).$$

This shows that the minimum can only be attained at $(L + 1, 0)$ or $(0, \lfloor \sigma L \rfloor + 1)$ as desired. □

To estimate the convergence rates we require the minimum n. Clearly, the function of the exponent $G(l_x, l_t)$ is a monotonically increasing function. Using Lemma 5.3.2 this means that n is given by

$$\begin{aligned} n &= \min_{(l_x,l_t)\notin I_L^\sigma} G(l_x, l_t) = \min\{G(L + 1, 0), G(0, \lfloor \sigma L \rfloor + 1)\} \\ &= \min\left\{L + 1 + 2\mu(L + 1), \frac{\lfloor \sigma L \rfloor + 1}{2} + 2\lambda(\lfloor \sigma L \rfloor + 1))\right\} \\ &= \min\left\{(L + 1)(2\mu + 1), (\lfloor \sigma L \rfloor + 1))\left(\frac{4\lambda + 1}{2}\right)\right\} \\ &\sim \min\left\{2\mu + 1, \sigma\frac{4\lambda + 1}{2}\right\}(L + 1). \end{aligned}$$

Thus, the minimum is

$$n \sim (L + 1)\begin{cases} \sigma\frac{4\lambda+1}{2}, & \sigma \leq \frac{4\mu+2}{4\lambda+1} \\ 2\mu + 1, & \text{else.} \end{cases}$$

In Figure 5.4 we examine the case $\mu = \lambda$ more closely. The polynomial degrees restrict the choice of μ and λ, since we require $\lambda \leq p_x + 1$ and $\mu \leq p_t + 1$ to ensure that the approximation space is embedded in the appropriate Sobolev space. The figure shows the exponent n for different values of μ.

We know that the number of degrees of freedom for the full tensor product spaces is given by:

$$N_L = \dim \mathcal{X}_L \sim 2^{L(d-1+\sigma)}.$$

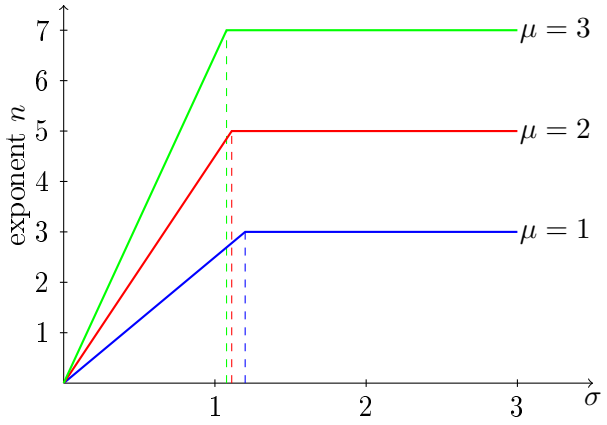

Figure 5.4: The exponent n against σ for several choices of $\mu = \lambda$.

We have now proven the following theorem.

Theorem 5.3.3. *Let $d > 1$ and let μ, λ fulfill $\lambda \leq p_x + 1$ and $\mu \leq p_t + 1$ and let $c > 0$ be a constant depending only on the polynomial degrees p_x and p_t. Then the convergence in the energy norm is*

$$\|u - u_h\|^2_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq cN_L^{-\frac{2\mu+1}{d-1+\sigma}} \|u\|^2_{H^{\mu,\lambda}(\Sigma)}, \text{ for } \sigma \leq \frac{4\mu+2}{4\lambda+1},$$

and

$$\|u - u_h\|^2_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq cN_L^{-\frac{(4\lambda+1)\sigma}{2(d-1+\sigma)}} \|u\|^2_{H^{\mu,\lambda}(\Sigma)}, \text{ for } \sigma > \frac{4\mu+2}{4\lambda+1},$$

where the scaling in space and time is given by $h_t \sim h_x^\sigma$.

In Figure 5.5 we see a plot of the convergence rates for the case $p_x = p_t = 0$ and in Table 5.2 we give the convergence rates and optimal choices of σ for some other values of $\mu = \lambda$.

In two dimensions and for $p_x = p_t = 0$ the convergence rate at $\sigma = \frac{6}{5}$ is $2\gamma = \frac{15}{11} = 1.36$ for the squares of the error. At $\sigma = 1$ the rate is expected to be $5/4 = 1.25$ and at $\sigma = 2$ we expect a rate of exactly 1. These rates coincide with those of the classical error estimates for $\sigma \leq 1$ and for $\sigma \geq 2$. However the maximum is now attained at $\frac{6}{5}$ and it is greater than the convergence rate at $\sigma = 1$, suggesting that this scaling should be used instead.

As μ and λ increase the improvement becomes smaller. These results give the largest improvement for the case $\mu = \lambda = 1$. This happens since for large $\mu = \lambda$ the term $\frac{4\mu+2}{4\mu+1}$ approaches 1 and our results approach the results given in the previous section.

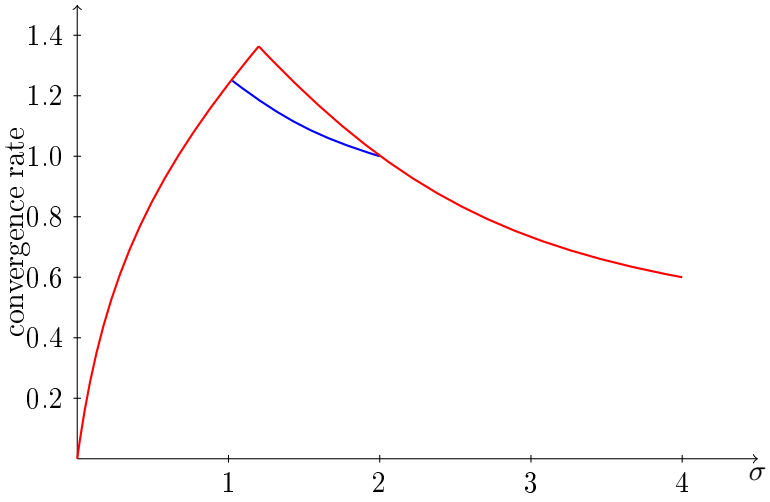

Figure 5.5: Convergence rate of the energy norm squared plotted against σ for $\mu = 1$. The blue line shows the results of the classical error analysis again, while the red line shows our improvements.

Full tensor product, $d = 2$			Full tensor product, $d = 3$		
(p_x, p_t)	conv. rate γ	scaling σ	(p_x, p_t)	conv. rate γ	scaling σ
$(0,0)$	$\frac{15}{22}$	$\frac{6}{5}$	$(0,0)$	$\frac{15}{32}$	$\frac{6}{5}$
$(1,1)$	$\frac{45}{38}$	$\frac{10}{9}$	$(1,1)$	$\frac{45}{56}$	$\frac{10}{9}$
$(2,2)$	$\frac{91}{54}$	$\frac{14}{13}$	$(2,2)$	$\frac{91}{80}$	$\frac{14}{13}$
$(3,3)$	$\frac{153}{70}$	$\frac{18}{17}$	$(3,3)$	$\frac{153}{104}$	$\frac{18}{17}$

Table 5.2: Improved convergence rates and optimal values of σ for full product discretisations in 2 and 3 dimensions.

5.4 Numerical Experiments

In this section we give some tests to confirm the convergence rates in the energy norm that were derived in this chapter.

First we give some brief definitions for Bessel functions, since they are needed to give the exact solutions for some of the tests. Then we move on to giving numerical experiments. First we show tests on a circle. These have the advantage that the exact solution can be calculated easily. One method for calculating solutions of the heat equation on a circle is given in Appendix A. These tests were also used in [42].

Then we give results calculated on ellipses of varying eccentricity and on a star-shaped domain. For these tests the exact solutions are not known, however, they offer a more challenging test for these methods.

5.4.1 Bessel Functions

Bessel functions, are the solutions to the Bessel differential equations:

$$z^2\frac{\partial^2 f(z)}{\partial z^2} + z\frac{\partial f(z)}{\partial z} + (z^2 - \alpha^2)f(z) = 0, \tag{5.5}$$

for an arbitrary complex number α.

Definition 5.4.1. *We denote by J_k k-th -Bessel function of the first kind. More precisely, a solution to (5.5) for $\alpha = k$, which is finite at the origin $x = 0$.*

5.4.2 Experiments on Circles

We solve the Dirichlet problem on a circle of radius $R = 1$, i.e. on the domain $\Omega = B_R(0)$. With $T > 0$ we denote a finite time horizon and with $\mathcal{I} := (0, T)$ the time intervall. We set $Q := \mathcal{I} \times \Omega$ the space-time cylinder with mantle $\Sigma = \mathcal{I} \times \Gamma$.

Then we want to find $u : Q \to \mathbb{R}$ satisfying:

$$\begin{aligned} (\partial_t - \Delta)u &= 0, && \text{in } Q \\ u &= 0, && \text{at } \{t = 0\} \times \Omega \\ \gamma_0 u &= g, && \text{in } \Sigma, \end{aligned} \tag{5.6}$$

where γ_0 is the trace operator.

The tests in this section show numerical results for three different choices of the right hand side g. In all three cases the exact boundary flux ψ is known. Using the coercivity and continuity of V in $H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)$ and Galerkin orthogonality, we have

$$\begin{aligned} \|\psi - \psi_L\|^2_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} &\sim \langle V(\psi - \psi_L), \psi - \psi_L\rangle \\ &\overset{\text{Galerkin orth.}}{=} \langle V(\psi - \psi_L), \psi\rangle \end{aligned}$$

We use this equation to calculate the error for all experiments in this section. For simplicity we plot the error in the energy norm squared.

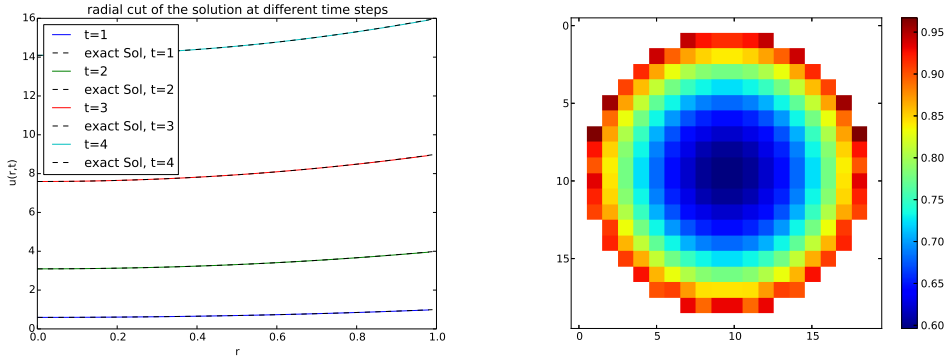

Figure 5.6: A radial cut of the solution at four different time steps, where the exact solution is shown in black and the discrete approximations in colour (right) and the solution $u(r,t)$ at the time step $t = 1$ (right). Both plots are calculated with 16 elements in space and 256 in time with constant basis functions.

Tests for Space-independent Right Hand Side

The first example we choose has a right hand side which is constant in space, in particular we choose $g(x,t) = t^2$.

In this case the exact solution due to [42] (note the sign error in that work) in polar coordinates is

$$u(r,\varphi,t) = t^2 + 4\sum_{k=1}^{\infty} \frac{J_0(\alpha_k r)}{\alpha_k^3 J_1(\alpha_k)} \left(t - \frac{1}{\alpha_k^2}(1 - e^{-\alpha_k^2 t}) \right),$$

where α_k are the roots of the 0-th Bessel function J_0 with $\alpha_1 < \alpha_2 < \ldots$. This solution is radially symmetric.

In Figure 5.6 we give plots of this exact solution and the approximated solution. One can see that the approximation is good, particularly in the center of the domain, but cannot be calculated near the boundary of the domain. Since the representation formula (2.9) used to calculate these values has a singularity at the boundary of the domain, this is not surprising.

The exact boundary flux is given by

$$q(\varphi,t) = t + 4\sum_{k=0}^{\infty} \frac{1 - e^{-\alpha_k^2 t}}{\alpha_k^4}.$$

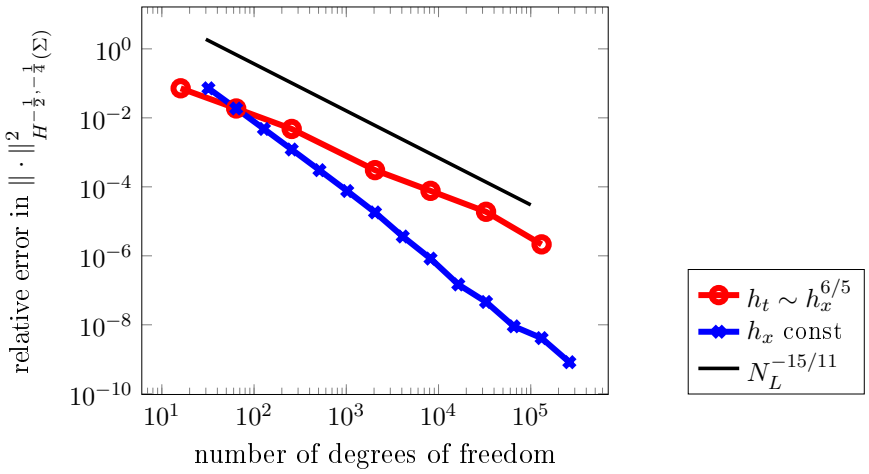

Figure 5.7: Convergence of the boundary flux in the energy norm for the right hand side $g(x,t) = t^2$.

To test the convergence rates we now calculate the convergence of the solution to the exact boundary flux given above. Let q_L be the approximated boundary flux in the discrete space $\mathcal{X}_L^\sigma$. Then the expected convergence rate in the energy norm is

$$\|q - q_L\|^2_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq c N_L^{-\frac{15}{11}} \|u\|^2_{H^{1,1}(\Sigma)}, \text{ for } \sigma = \frac{6}{5}, \tag{5.7}$$

according to Theorem 5.3.3.

Figure 5.7 shows the convergence rates in the energy norm for this right hand side. The red plot shows the convergence when $h_t \sim h_x^{6/5}$. Note that we have plotted the squares of the energy norm, and as such our expected convergence rate is $\frac{15}{11}$. As we can see the convergence rate coincides with the expected values.

Since for this particular solution the boundary flux is only time-dependent, we do not have to refine in space to improve convergence. In order to show convergence to a higher accuracy, we also show a test in which only 4 elements in space are used and only h_t is refined. This is also shown in Figure 5.7.

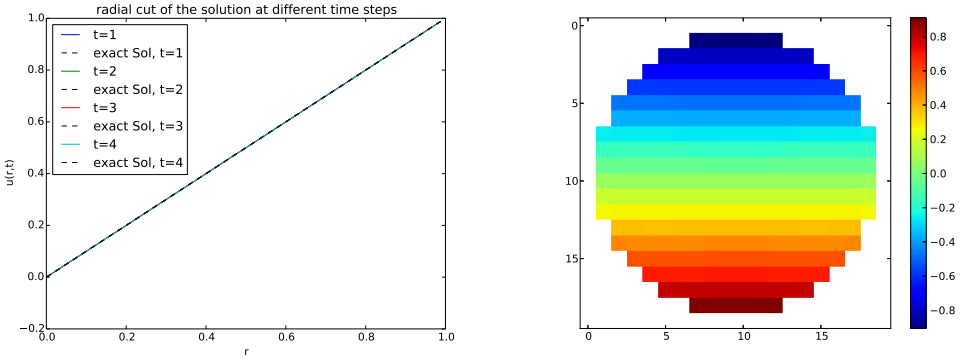

Figure 5.8: A radial cut of the solution at four different time steps, where the exact solution is shown in black and the discrete approximations in colour (right) and the solution $u(r, \varphi)$ at the time step $t = 1$ (left). Both plots were calculated with constant basis functions in time and space, with 8 elements used in each.

Tests for a Stationary Right Hand Side

Next we look at a solution which is stationary. The right hand side we choose for this test is $g(r, \varphi, t) = R\cos(\varphi)$.

In this case the exact solution is easy to calculate, it is

$$u(r, \varphi, t) = r\cos(\varphi).$$

The solution and its boundary flux are constant in time as can be seen in Figure 5.8. This figure shows the solution at $t = 1$ and a radial cut of the solution at different time steps. Even though only a few degrees of freedom are used in space, the discrete solution nevertheless provides a good approximation. The exact boundary flux in this case is

$$q(\varphi, t) = \cos(\varphi).$$

Figure 5.9 shows the convergence rates of the squares of the energy norm for this right hand side. The red plot shows the convergence when $h_t \sim h_x^{6/5}$ and our expected convergence rate is $\frac{15}{11}$ according to Theorem 5.3.3. As we can see the convergence rate is close to the predicted values.

For this solution the boundary flux is constant in time, so we do not have to refine in time to improve convergence. In order to show convergence to a higher accuracy, we also show a test in which only 1 element is used in time and only the mesh width h_x is refined. This is also shown in Figure 5.9

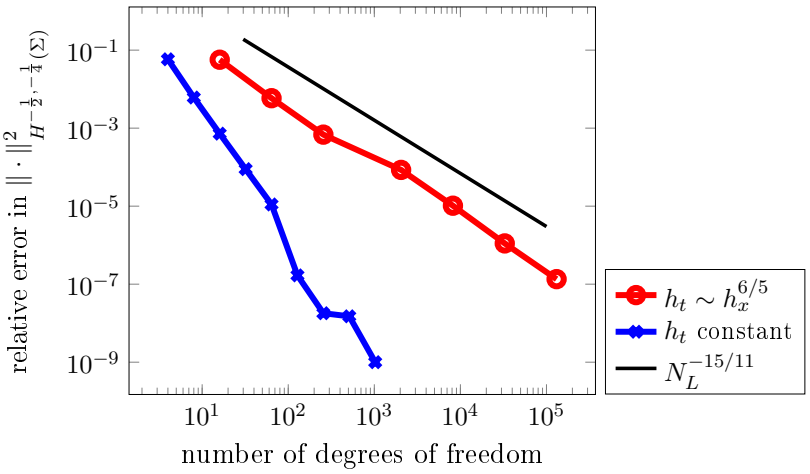

Figure 5.9: Convergence of the boundary flux in the energy norm for the right hand side $g(r, \varphi, t) = R\cos(\varphi)$.

Tests with a Time- and Space-dependent Right Hand Side

The last test calculated on the circle combines the two previous tests, using the right hand side $g(r, \varphi, t) = t^2 \cos(\varphi)$. This right hand side leads to a solution that is not constant in time or space. The exact solution for this problem is

$$u(r, \varphi, t) = \left(rt^2 - 4 \sum_{k=1}^{\infty} \frac{J_1(\beta_k r)}{\beta_k^3 J_2(\beta_k)} \left(t - \frac{1}{\beta_k^2} \left(1 - e^{-\beta_k^2 t} \right) \right) \right) \cos(\varphi), \tag{5.8}$$

where β_k are the roots of the first Bessel function J_1 with $\beta_1 < \beta_2 < \ldots$. In Figure 5.10 we show the calculated solution u at the time steps $t = .25, .5, .75$ and 1. We see that the differences in the extrema of solution increasing as time passes.
Taking the normal derivative of the exact solution, it is easy to see that the exact boundary flux is

$$q(r, \varphi, t) = \left(t^2 - \frac{1}{4} t + 4 \sum_{k=0}^{\infty} \frac{1 - e^{-\beta_k^2 t}}{\beta_k^4} \right) \cos(\varphi). \tag{5.9}$$

To check the convergence rates we again calculate the convergence of the solution to the exact boundary flux given above. The expected convergence rate of the squares of the energy norm is $\frac{15}{11}$, where the scaling $h_t \sim h_x^{6/5}$ is chosen.

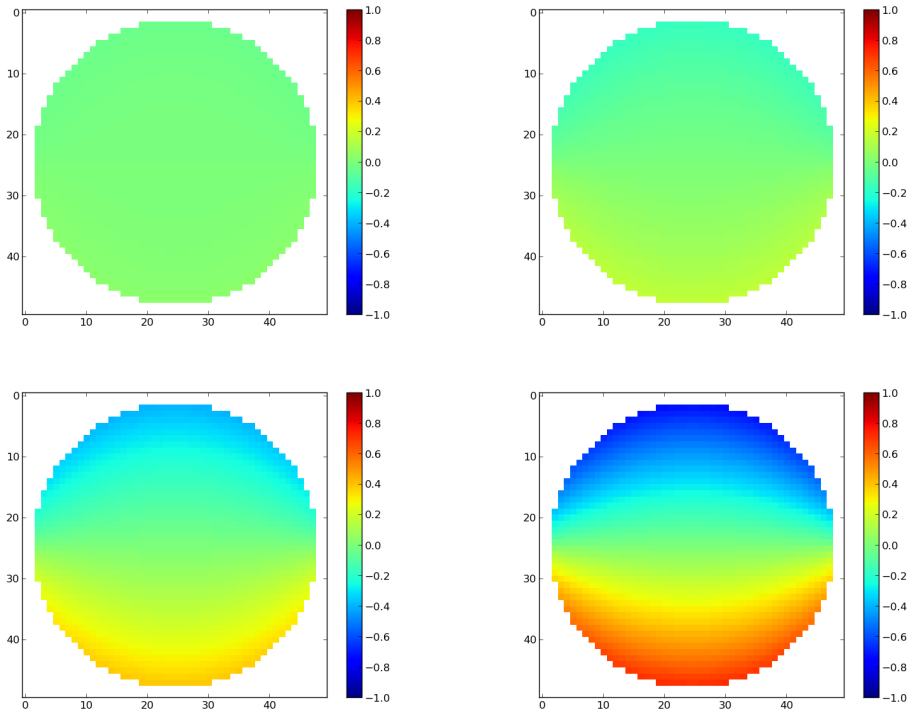

Figure 5.10: The approximated solution using the indirect method for the right hand side $g(r, \varphi, t) = t^2 \cos(\varphi)$ at four different time steps, $t = 1, 2, 3, 4$. Piecewise constant basis functions were used in time and space, with 16 elements used in each.

Figure 5.11 shows the convergence rates in the energy norm for this right hand side. The red plot shows the convergence when $h_t \sim h_x^{6/5}$. Again we have plotted the squares of the energy norm, and our expected convergence rate is $\frac{15}{11}$. As we can see the convergence rate is close to the predicated rate.

We also run tests with two other values of σ. When $\sigma = 1$ we have as expected a slightly larger error. The convergence rate in this case is expected to be $\frac{5}{4}$. Lastly, when $\sigma = 2$ we expect a slower convergence rate of 1. The numerical tests in Figure 5.11 confirm these rates.

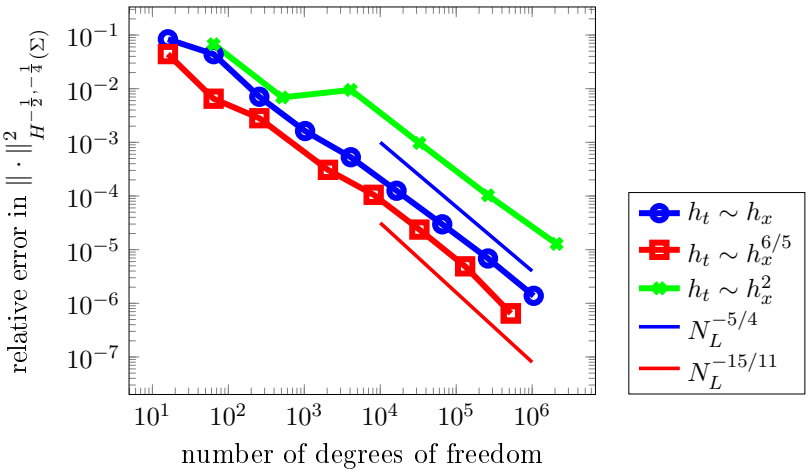

Figure 5.11: Convergence of the boundary flux in the energy norm for the right hand side $g(r, \varphi, t) = Rt^2 \cos(\varphi)$.

5.4.3 Experiments on Ellipses

In this section we give some more challenging tests on ellipses. For these tests it is simpler to use the indirect method, as the exact solution is not known. We use a value calculated with as many degrees of freedom as possible, as an approximation of the exact solution to calculate the error.

In Figure 5.12 we show the approximated solutions for two ellipses with different right hand sides. These tests are described in the following sections.

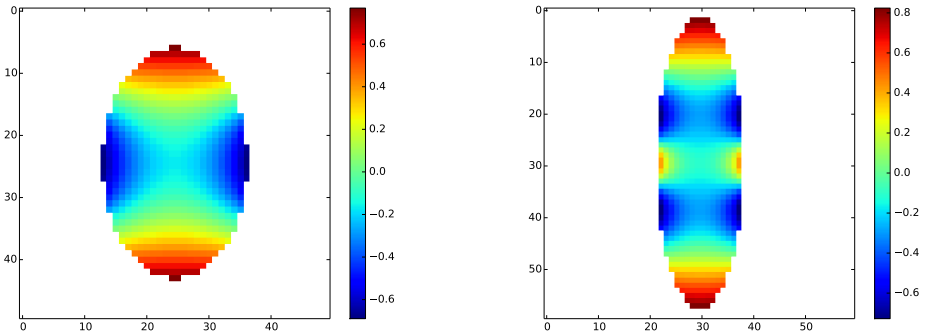

Figure 5.12: The approximated solution on an ellipse for the right hand side $g(\varphi, t) = t^2 \cos(2\varphi)$ at the time-step $t = 1$ (left), calculated with 16 elements in time and space. The approximated solution for the right hand side $g(\varphi, t) = t^2 \cos(4\varphi)$ at the time-step $t = 1$ (right), calculated with 64 elements in time and space.

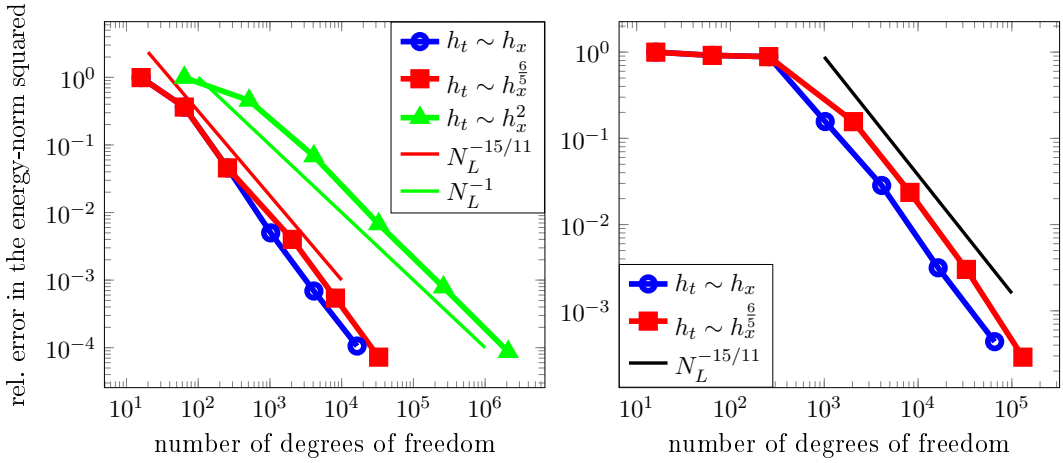

Figure 5.13: Convergence of the boundary flux in squares of the energy norm for the right hand side $g(\varphi, t) = t^2 \cos(2\varphi)$ on an ellipse with eccentricities $a = 0.8, b = 0.5$ (left) and for $g(\varphi, t) = t^2 \cos(4\varphi)$ on an ellipse with eccentricities $a = 1, b = 0.3$ (right).

Tests for Time- and Space-dependent Right Hand Side

The first test is on an ellipse with semi-axes: $a = 0.8, b = 0.5$. The right hand side that was chosen, is $g(\varphi, t) = t^2 \cos(2\varphi)$. The solution is shown in Figure 5.12.

In Figure 5.13 one can see that the correspondence to the expected rates is good for $\sigma = 2$, where we expect a rate of exactly 1. At $\sigma = 1$ the rate should be $5/4 = 1.25$, and is in fact somewhat higher than that. In particular, the error for $\sigma = 1$ is smaller than the error for $\sigma = \frac{6}{5}$. The reason for this discrepancy is not clear. The expected convergence rate for $\sigma = \frac{6}{5}$ is $\frac{15}{11}$, and Figure 5.12 shows a good correspondence to this rate.
The next test features a thinner ellipse with $a = 1, b = 0.3$. The right hand side was chosen to be more oscillatory than in the previous case, with $g(\varphi, t) = t^2 \cos(4\varphi)$. This solution is shown in Figure 5.12. This plot was generated using 64 elements in time and space, more elements were necessary to resolve the oscillations.

One can see in Figure 5.13 that due to the larger number of oscillations, the pre-asymptotic range has increased. Three uniform refinements are necessary, before the convergence curves reach their asymptotic rates. At $\sigma = 1$ the rate should be $5/4 = 1.25$ and is again somewhat higher. At $\sigma = \frac{6}{5}$ the expected convergence rate for the squares of the energy norm is $\frac{15}{11}$ and we see a good correspondence to this rate. The error for $\sigma = 1$ is smaller than the error for $\sigma = \frac{6}{5}$ as in the previous test using ellipses, and unlike the tests on the circle. It is unclear why $\sigma = 1$ leads to

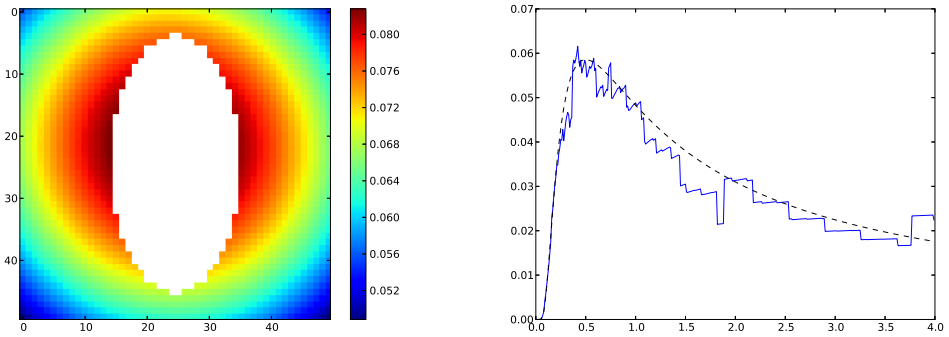

Figure 5.14: The approximated solution on the exterior of an ellipse for the right hand side $g(\varphi, t) = t^2 \cos(\varphi)$, at the time-step $t = 1$ (left), and the time evolution of the solution at the point $x = (.9, .9)$, with the exact solution shown in black (right).

higher convergence rates for ellipses.

Tests for an Exterior Problem on an Ellipse

In this section we give a numerical experiment for an exterior problem. More exactly, we solve the heat equation on the exterior of an ellipse. Using the boundary integral formulation of the heat equation, this problem can be handled with the same method as an interior problem. The only change to the tests, given previously for the interior problems on ellipses, is that the outer normal now points into the ellipse.

For this test we used an ellipse with eccentricities $a = 0.8,\ b = 0.5$. We used the fundamental solution itself as a right hand side

$$g(x, t) = G(x, t).$$

This means, that we have the exact solution and its boundary flux in the entire domain. The solution at time-step $t = 1$, and the time evolution of the solution are shown in Figure 5.14.

We show tests for the exterior problem only for the optimal scaling $\sigma = \frac{6}{5}$. In Figure 5.15, we plot the convergence in the energy-norm squared. As in the previous tests for ellipses there is a pre-asymptotic range where there is no convergence. However, after three steps we see a good correspondence with the theoretically expected convergence rate of $\frac{15}{11}$.

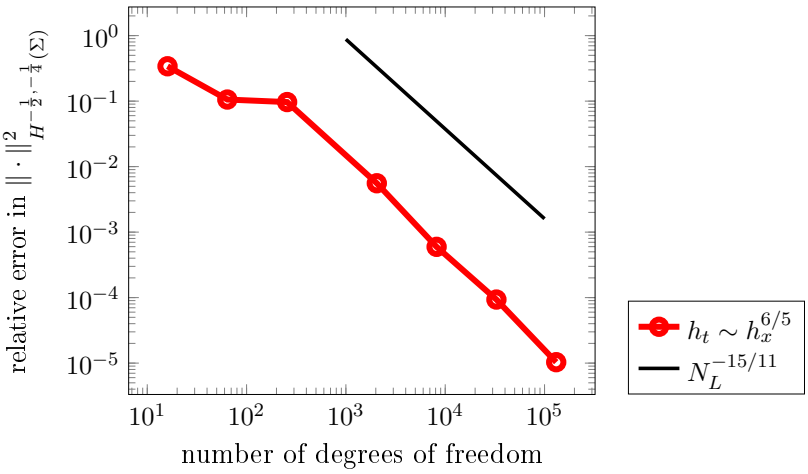

Figure 5.15: Convergence of the boundary flux in the energy norm squared for the right hand side $g(x,t) = G(x,t)$, for the exterior of an ellipse.

5.4.4 Experiments on Star-shaped Domains

In this section we show one experiment on the star-shaped domain parametrised by (4.6). This domain was chosen to show the convergence of the method on a smooth domain, that is less symmetric than the circle and ellipse.

The right hand side that was chosen for this test is $g(\varphi,t) = t^2$. In Figure 5.16 we show the approximated solution to this problem at the time step $t = 1$. This solution was calculated with 16 elements in time and space. For this problem the exact solution and boundary flux are not known. To calculate the convergence, we use the last calculated value as an approximation to the exact solution.

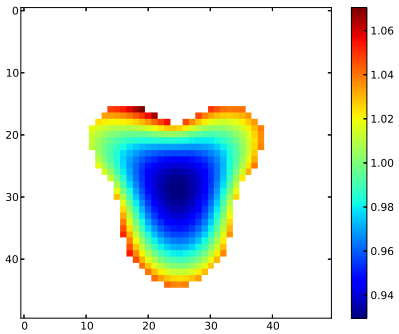

Figure 5.16: The approximated solution on a star-shaped domain at the time-step $t = 1$

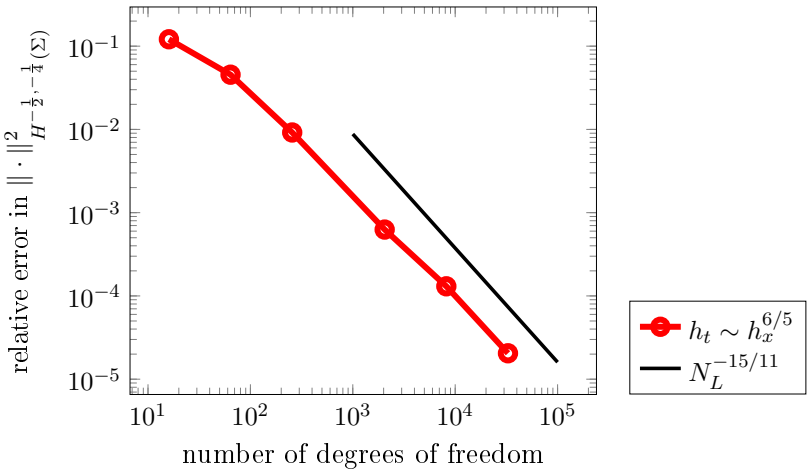

Figure 5.17: Convergence of the boundary flux in the energy norm squared for the right hand side of $g(x,t) = t^2$, on a star-shaped domain.

We used the optimal scaling $\sigma = \frac{6}{5}$ for this test. For that scaling, the expected convergence rate for the squares of the energy norm is $\frac{15}{11}$. In Figure 5.17 we plot the convergence of the squares of the energy norm for the problem. We see that the calculated convergence rate is close to the predicted rate for this test.

Chapter 6

Sparse Grids

This chapter introduces sparse grids. We define their structure and summarise their approximation properties. Two types of sparse grid index sets will be studied, the standard sparse grid index set and an optimised sparse grid index set.

Sparse grids (see e.g. [54], [3], [9]) have been applied successfully to a variety of different problems, such as quantum mechanics [22], high- dimensional quadrature [25] or elliptic partial differential equations [27].

The approximation properties of standard (Smolyak) sparse grids for the BEM formulation of the heat equation will be summarised in Section 6.2.1. Further, we show new results obtained for the approximation of the optimised sparse grids applied to the heat equation in Section 6.2.2. These results are useful as they allow more general choices of polynomial degree for the basis functions.

We also explain the combination technique, which gives a faster algorithm for sparse grid methods (see [33], [22], or [26]) in Section 6.3. Finally, we give numerical results for these methods in Section 6.4.

6.1 Construction of Sparse Grid Spaces

Essentially the idea behind sparse grid methods is truncating a tensor-product expansion of a one-dimensional multilevel basis . The main advantage to using sparse Galerkin discretisations is that they yield a mild dependence on the dimension. More precisely, sparse grid methods scale in dimension with $\mathcal{O}(N(\log N)^{d-1})$, while the full tensor product scales with $\mathcal{O}(N^d)$, where N is the number of degrees of freedom.

We use sparse grids to improve the cardinality of the tensor product in space-time.

These have a natural tensor product structure between space and time which makes such a discretisation easy. In general the spatial dimensions do not have a tensor product structure, so applying sparse grids there as well is more difficult.

The first step in defining sparse grid structures is the definition of one-dimensional multilevel decompositions. They can then be combined to form sparse grid spaces.

Let $\mathcal{X}_L^x$ be the discrete space in the spatial dimensions and let $\mathcal{X}_L^t$ be the discrete space in time. Assume there exists a multilevel decomposition of these spaces

$$\begin{aligned}\mathcal{X}_L^x &= \mathcal{W}_0^x \oplus \cdots \oplus \mathcal{W}_L^x,\\ \mathcal{X}_L^t &= \mathcal{W}_0^t \oplus \cdots \oplus \mathcal{W}_L^t.\end{aligned}$$

A variety of different bases can be used to obtain the required multilevel decomposition. We will mainly use the wavelet bases described in Chapter 3. Another commonly used basis is the piecewise linear spline basis [27].

Figure 6.1 shows the one-dimensional multilevel decomposition given by the Haar wavelet basis. One can easily see the hierarchial structure of the subspaces, this structure is also present for other multilevel decompositions.
The full tensor product space from Chapter 4 can easily be rewritten using the above multilevel decompositions.

$$\begin{aligned}\mathcal{X}_L^x \otimes \mathcal{X}_L^t &= \left(\oplus_{i=0}^L \mathcal{W}_i^x\right) \otimes \left(\oplus_{j=0}^L \mathcal{W}_j^t\right)\\ &= \sum_{\max\{i,j\}\leq L} \mathcal{W}_i^x \otimes \mathcal{W}_j^t.\end{aligned}$$

The sparse grid method relies on cutting off the above sum in a way that balances the accuracy of the approximation space and the cardinality of each complement space. When approximating a smooth function the spaces with a large number of degrees of freedom in both time and space are not the most important ones. Instead spaces which are refined heavily in only one of the dimensions are needed. In particular, computationally expensive spaces such as $\mathcal{W}_L^x \otimes \mathcal{W}_L^t$ with a dimension of 2^{2L} are not necessary to decrease the error.

The general form of a sparse grid space in two dimensions is

$$\hat{\mathcal{X}}_L := \bigotimes_{(\ell_x,\ell_t)\in I_L} \mathcal{W}_{\ell_x}^x \otimes \mathcal{W}_{\ell_t}^t \subset \mathcal{X} = H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)$$

where I_L is an index set. In the following sections we discuss two different choices

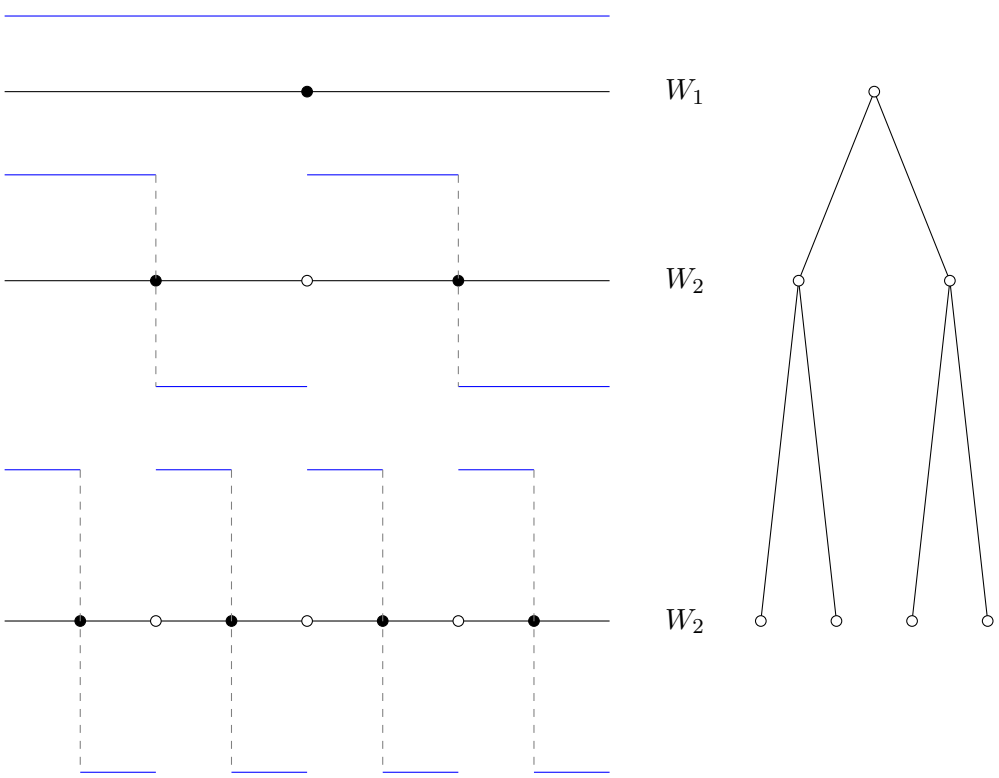

Figure 6.1: The multilevel one-dimensional Haar wavelet basis on 3 levels. The dots are at the center of the basis function they represent. The black dots represent basis functions on that level. The unfilled dots represent the location of basis functions on previous levels. On the right the tree structure of this multilevel decomposition is shown.

for these index sets. In Figure 6.2 we show how the space is set up for a standard sparse grid index set (see e.g. [28]).

Definition 6.1.1. *The standard anisotropic sparse grid index set is defined as follows*

$$\hat{I}_L^\sigma = \{(\ell_x, \ell_t) \ : \ \ell_t/\sigma + \ell_x\sigma \leq L\}\,,$$

where σ is a free variable. In this case we write $\mathcal{X}_L = \mathcal{X}_L^\sigma$.

In Figure 6.3 we plot this index set for the choices $\sigma = 1$ and $\sigma = \sqrt{2}$.

In the following section on the error anlysis the optimal choice for the free variable σ will be clarified. The choice depends on the spatial dimension d.

We remember that the L^2-orthogonal projection from a given approximation space

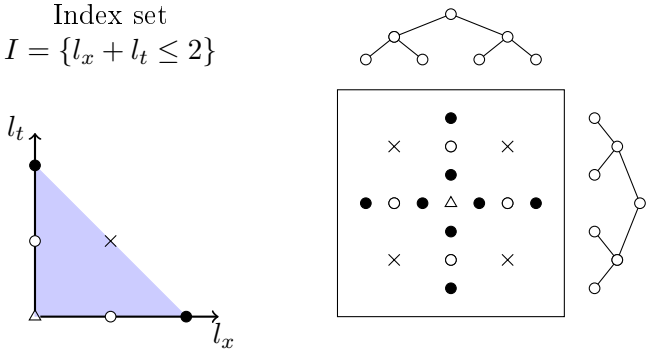

Figure 6.2: The standard sparse grid index set on the left and to the right the corresponding basis functions. The basis functions are represented by markers at the center of their support.

$\mathcal{X}$ (Definition 5.1.1) is given by

$$\Pi_{\mathcal{X}} : L^2(\Sigma) \to \mathcal{X}.$$

Independentally of the choice of approximation space we can make the following an Aubin-Nitsche argument. Let us assume the solution $\psi \in L^2(\Sigma)$.

$$\begin{aligned}\|\psi - \Pi_{\mathcal{X}}\psi\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} &= \sup_{\xi \in H^{\frac{1}{2},\frac{1}{4}}(\Sigma)} \frac{\langle \psi - \Pi_{\mathcal{X}}\psi, \xi \rangle}{\|\xi\|_{H^{\frac{1}{2},\frac{1}{4}}(\Sigma)}} \\ &= \sup_{\xi \in H^{\frac{1}{2},\frac{1}{4}}(\Sigma)} \frac{\langle \psi - \Pi_{\mathcal{X}}\psi, \xi - \Pi_{\mathcal{X}}\xi \rangle}{\|\xi\|_{H^{\frac{1}{2},\frac{1}{4}}(\Sigma)}}\end{aligned}$$

Then we can estimate

$$\begin{aligned}\|\psi - \Pi_{\mathcal{X}}\psi\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} &\leq \|\psi - \Pi_{\mathcal{X}}\psi\|_{L^2(\Sigma)} \sup_{\xi \in H^{\frac{1}{2},\frac{1}{4}}(\Sigma)} \frac{\|\xi - \Pi_{\mathcal{X}}\xi\|_{L^2(\Sigma)}}{\|\xi\|_{H^{\frac{1}{2},\frac{1}{4}}(\Sigma)}} \\ &\leq \|\psi\|_{H^{s_x,s_t}_{\text{mix}}(\Sigma)} \underbrace{\frac{\|\psi - \Pi_{\mathcal{X}}\psi\|_{L^2(\Sigma)}}{\|\psi\|_{H^{s_x,s_t}_{\text{mix}}(\Sigma)}}}_{\substack{\text{small for} \\ \text{standard} \\ \text{sparse grids}}} \underbrace{\sup_{\xi \in \|\psi\|_{H^{\frac{1}{2},\frac{1}{4}}(\Sigma)}} \frac{\|\xi - \Pi_{\mathcal{X}}\xi\|_{L^2(\Sigma)}}{\|\xi\|_{H^{\frac{1}{2},\frac{1}{4}}(\Sigma)}}}_{\substack{\text{small for full tensor} \\ \text{product grids}}}\end{aligned}$$

for any approximation space $\mathcal{X}$.

This argument leads to the idea of finding a compromise between the full tensor

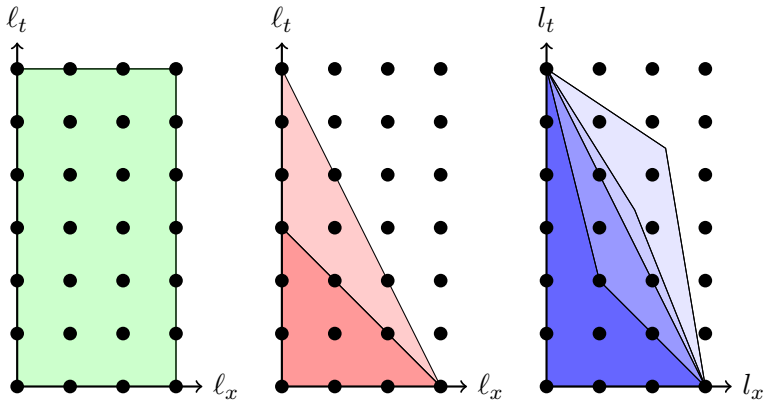

Figure 6.3: Index sets for the full tensor product discretisation and for the sparse tensor products with $\sigma = 1$ and $\sigma = \sqrt{2}$ respectively, as well as the optimised sparse grid index set for $\mathcal{T} = \frac{1}{2}, 0, -\frac{1}{4}, -2$.

product discretisation and the sparse grid discretisation. The optimised sparse grid space is such a space for certain parameters.

The optimised sparse grid index sets were first introduced in [31]. They give optimal results for the sparse grid convergence in Sobolev norms of the spaces $H^s(\Omega)$, $s \in \mathbb{R}$. We use these index sets to discretise the anisotropic Sobolev spaces $H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)$.

Due to the anisotropy in the Sobolev space it is beneficial to also introduce an anisotropy in the index set. Care has to be taken when comparing to [31], where this anisotropy is not present in the definition.

Definition 6.1.2. *The optimised sparse grid index set is defined as follows*

$$J_L^{\mathcal{T}} = \{(l_x, l_t) \ : \ l_x + l_t/2 - \mathcal{T}\max\{l_x, l_t/2\} \leq (1-\mathcal{T})L\}\,.$$

where $\mathcal{T} \in [-\infty, 1)$ is a free variable. In this case we write $\mathcal{X}_L = \mathcal{X}_L^{\mathcal{T}}$.

These index sets allow more flexibility through the parameter $\mathcal{T}$. The index set for $\mathcal{T} = 0$ corresponds to the standard sparse grids and $\mathcal{T} = -\infty$ corresponds to full tensor product spaces.

The index set is plotted in Figure 6.3 for different values of $\mathcal{T}$. One can see that as the $\mathcal{T}$ gets smaller the index set gets larger, eventually approaching the full tensor product space.

6.2 Error Analysis

In this section we give an error analysis for the discretisation of the heat equation using sparse grid spaces. First we give an error analysis for the standard sparse grid spaces following the proofs in [12]. Then we give some new results for the error anlysis for optimised sparse grid spaces.

To apply these methods we first reiterate the discrete formulation of the heat equation.

Given $\mathcal{X}_L \subset \mathcal{X} := H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)$. Find $\psi_L \in \mathcal{X}_L$ such that

$$\begin{aligned} \text{Indirect Method:} \quad & \langle \eta_L, V\psi_L \rangle = \langle \eta_L, (\frac{1}{2}I + K)g \rangle \quad \forall \eta_L \in \mathcal{X}_L, \\ \text{Direct Method:} \quad & \langle \eta_L, V\psi_L \rangle = \langle \eta_L, g \rangle \qquad\qquad\quad \forall \eta_L \in \mathcal{X}_L. \end{aligned} \tag{6.1}$$

We showed in Chapter 2.2 that the single layer operator V is coercive. This means that we immediately get a best approximation property for the discrete spaces from the classical Lemma of Céa. We will use this property in both sections.

6.2.1 Error Analysis for Standard Sparse Grids

In this section we find and prove error estimates for the standard sparse grid spaces. This section follows [12] closely. This proofs are given for completeness.

The error estimate relies mainly on an Aubin-Nitsche argument for the L^2-orthogonal projection.

Definition 6.2.1. *We denote by*

$$\Pi_{\mathcal{X}_L^\sigma} \, : \, L^2(\Sigma) \to \mathcal{X}_L^\sigma$$

the L^2-orthogonal projection onto the discrete space $\mathcal{X}_L^\sigma$.

The main result of this section is given below.

Theorem 6.2.2. *Suppose $\psi \in \tilde{H}_{mix}^{(d-1)\mu,\mu}(\Sigma)$ for μ, p_x, p_t satisfying*

$$\mu = \frac{p_x + 1}{d-1}, \qquad \text{and} \qquad p_t + 1 \geq \mu. \tag{6.2}$$

Then the error of the sparse tensor Galerkin approximation $\psi_L \in \mathcal{X}_L^\sigma$, with $\sigma = \sqrt{d-1}$ is

$$\|\psi - \psi_L\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq c N_L^{-\lambda} \left(\log(N_L)\right)^{\mu+\frac{1}{2}} \|\psi\|_{H^{(d-1)\mu,\mu}_{mix}(\Sigma)}, \tag{6.3}$$

where N_L is the number of degrees of freedom and $\lambda = \mu + \frac{1}{2(d+1)}$.

Proof. Due to coercivity of the single-layer operator we can estimate the Galerkin error by

$$\|psi - \psi_L\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq c\|\psi - \Pi_{\mathcal{X}_L^\sigma}\psi\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)}.$$

Further, let us assume that $\psi \in L^2(\Sigma)$. Then, we can use an Aubin-Nitsche argument to get the following estimate

$$\begin{aligned}
\|\psi - \Pi_{\mathcal{X}_L^\sigma}\psi\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} &= \sup_{\xi \in H^{\frac{1}{2},\frac{1}{4}}(\Sigma)} \frac{\langle \psi - \Pi_{\mathcal{X}_L^\sigma}\psi, \xi\rangle}{\|\xi\|_{H^{\frac{1}{2},\frac{1}{4}}(\Sigma)}} \\
&= \sup_{\xi \in H^{\frac{1}{2},\frac{1}{4}}(\Sigma)} \frac{\langle \psi - \Pi_{\mathcal{X}_L^\sigma}\psi, \xi - \Pi_{\mathcal{X}_L^\sigma}\xi\rangle}{\|\xi\|_{H^{\frac{1}{2},\frac{1}{4}}(\Sigma)}} \\
&\leq \|\psi - \Pi_{\mathcal{X}_L^\sigma}\psi\|_{L^2(\Sigma)} \sup_{\xi \in H^{\frac{1}{2},\frac{1}{4}}(\Sigma)} \frac{\|\xi - \Pi_{\mathcal{X}_L^\sigma}\xi\|_{L^2(\Sigma)}}{\|\xi\|_{H^{\frac{1}{2},\frac{1}{4}}(\Sigma)}}
\end{aligned} \tag{6.4}$$

The above result holds for all discrete spaces $\mathcal{X}_L$. In order to show the desired convergence results we use some well-known properties of sparse grid spaces. More precisely, we use Corollary 4.5 from [28]. It states that for $0 < r < p_x + 1$ and $0 < s < p_t + 1$

$$\|\xi - \Pi_{\mathcal{X}_L^\sigma}\xi\|_{L^2(\Sigma)} = \inf_{\xi_L \in \mathcal{X}_L^\sigma} \|\xi - \xi_L\|_{L^2(\Sigma)} \leq c N_L^{-\alpha} (\log N_L)^\beta \|\xi\|_{H^{r,s}_{\text{mix}}(\Sigma)}, \tag{6.5}$$

with a convergence rate of $\alpha = \frac{\min\{r, s\sigma^2\}}{\max\{d-1, \sigma^2\}}$ and with some $\beta \geq 0$, that will be specified later.

Our goal is to choose the free variable σ such that the convergence rate α is as large as possible. In our setting we have $s = \mu$ and $r = (d-1)\mu$. This means that the convergence rate is

$$\alpha = \frac{\min\{\mu(d-1), \mu\sigma^2\}}{\max\{d-1, \sigma^2\}},$$

which attains its maximum of $\alpha_{\max} = \mu$ at $\sigma^2 = d - 1$.

We remember that $H^{k,k/2}(\Sigma) \subset H^{(d-1)\mu,\mu}_{\text{mix}}(\Sigma)$ for $k \geq (d+1)\mu$. This gives

$$\|\xi - \Pi_{\mathcal{X}_L^\sigma}\xi\|_{L^2(\Sigma)} \leq cN_L^{-\alpha}(\log N_L)^\beta \|\xi\|_{H^{\mu,\mu(d-1)}_{\text{mix}}(\Sigma)} \leq cN_L^{-\alpha}(\log N_L)^\beta \|\xi\|_{H^{k,k/2}(\Sigma)}.$$

Setting $\alpha = \frac{1}{2(d+1)}$ and $k = \frac{1}{2}$ by [28] we get

$$\sup_{\xi \in H^{\frac{1}{2},\frac{1}{4}}(\Sigma)} \frac{\|\xi - \Pi_{\mathcal{X}_L^\sigma}\xi\|_{L^2(\Sigma)}}{\|\xi\|_{H^{\frac{1}{2},\frac{1}{4}}(\Sigma)}} \leq cN_L^{-\frac{1}{2(d+1)}} (\log N_L)^{\frac{1}{2(d+1)}+\frac{1}{2}}.$$

Further, using (6.5) we can estimate

$$\|\psi - \Pi_{\mathcal{X}_L^\sigma}\psi\|_{L^2(\Sigma)} \leq cN_L^{-\mu} (\log N_L)^{\mu+\frac{1}{2}} \|\psi\|_{H^{(d-1\mu,\mu)}_{\text{mix}}(\Sigma)}.$$

Combining these two results using (6.4) we get as desired

$$\|\psi - \Pi_{\mathcal{X}_L^\sigma}\psi\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq cN_L^{-\mu-\frac{1}{2(d+1)}} (\log N_L)^{\frac{1}{2(d+1)}+\mu+1} \|\psi\|_{H^{(d-1)\mu,\mu}_{\text{mix}}(\Sigma)}.$$

□

Remark 6.2.3. *In the case $d = 2$ and by choosing $\sigma = 1$ we get*

$$\|\psi - \psi_L\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq cN_L^{-p_x-\frac{7}{6}} (\log(N_L))^{p_x+\frac{3}{2}} \|\psi\|_{H^{p_x+1,p_x+1}_{mix}(\Sigma)}, \tag{6.6}$$

where $p_x \leq p_t$.

Remark 6.2.4. *In the case $d = 3$ and by choosing $\sigma = \sqrt{2}$ we get*

$$\|\psi - \psi_L\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq cN_L^{-\frac{p_x}{2}-\frac{5}{8}} (\log(N_L))^{\frac{p_x}{2}+1} \|\psi\|_{H^{2(p_x+1),p_x+1}_{mix}(\Sigma)}, \tag{6.7}$$

where $p_x \leq 2p_t + 1$.

Corollary 6.2.5. *Suppose $g \in \tilde{H}^{k,\frac{k}{2}}(\Sigma)$ and μ, λ, p_x, p_t satisfy the requirements of Theorem 6.2.2 and*

$$k = \frac{d+1}{d-1}(p_x + 1) + 1.$$

Then the error of the sparse tensor Galerkin solution $\psi \in \mathcal{X}_L^{\sqrt{d-1}}$ has the bounds

$$\|\psi - \psi_L\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq cN_L^{-\lambda} (\log N_L)^{\lambda+1} \|g\|_{H^{k,\frac{k}{2}}(\Sigma)}.$$

Proof. See Corollary 4.8 in [12]. The proof uses the embedding $H^{(d+1)\mu,\frac{(d+1)\mu}{2}}_{\text{mix}}(\Sigma) \subset H^{(d-1)\mu,\mu}(\Sigma)$ and the fact that the single layer operator V is an isomorphism in the

appropriate anisotropic Sobolev spaces. □

The convergence rates for different dimensions and choices of polynomial degrees are summarised in Table 6.1 for $d = 2$ and in Table 6.2 for $d = 3$. The convergence rates given for full tensor products are improved from [12] using the results from Section 5. The regularity refers to the required regularity on the right hand side for these methods, i.e. in $H^{k,\frac{k}{2}}(\Sigma)$.

The table shows that in discretisations with low polynomial degree the sparse grids yield higher rates than the full tensor products. However, they require slightly higher regularity assumptions on the data. They also have restrictions on the choice of polnomial degrees in time and space due to (6.2) in Theorem 6.2.2.

In the case we are most interested in: piecewise constant basis functions, i.e. $p_x = p_t = 0$ and $d = 2$, the convergence rate in the energy norm using these sparse grids is almost twice as high as that of full tensor products. This large improvement can be seen in the numerical tests of these methods given in Section 6.4.

For $d = 3$ the improvements to the convergence rates γ in the energy norm using sparse grids are not quite as large as in $d = 2$. For example, when $p_x = p_t = 0$ the improvement is from $\frac{15}{32} \sim 0.47$ to $\frac{5}{8} = 0.625$.

Tests with $p_x = 2p_t + 1$ are also given since these give optimal results for the full tensor products. In $d = 3$ we see that even in this case the sparse grids outperform full tensor product grids.

Full tensor product, $d = 2$				Standard sparse grids, $d = 2$		
(p_x, p_t)	conv. rate γ	regularity k	σ	(p_x, p_t)	conv. γ	regularity k
$(0,0)$	$\frac{15}{22}$	3	$\frac{6}{5}$	$(0,0)$	$\frac{7}{6}$	4
$(1,0)$	$\frac{5}{6}$	3	2	$(1,0)$	-	-
$(1,1)$	$\frac{45}{38}$	5	$\frac{10}{9}$	$(1,1)$	$\frac{13}{6}$	7
$(3,1)$	$\frac{3}{2}$	5	2	$(3,1)$	-	-

Table 6.1: Convergence rates and required regularity assumptions on the right hand side for full and sparse tensor product discretisation in 2 dimensions.

Full tensor product, $d = 3$				Standard sparse grids, $d = 3$		
(p_x, p_t)	conv. rate γ	regularity k	σ	(p_x, p_t)	conv. γ	regularity k
$(0,0)$	$\frac{15}{32}$	3	$\frac{6}{5}$	$(0,0)$	$\frac{5}{8}$	3
$(1,0)$	$\frac{5}{8}$	3	2	$(1,0)$	$\frac{9}{8}$	5
$(1,1)$	$\frac{45}{56}$	5	$\frac{10}{9}$	$(1,1)$	$\frac{9}{8}$	5
$(3,1)$	$\frac{9}{8}$	5	2	$(3,1)$	$\frac{17}{8}$	9

Table 6.2: Convergence rates and required regularity assumptions on the right hand side for full and sparse tensor product discretisation in 3 dimensions.

6.2.2 Error Analysis for Optimised Sparse Grids

The Aubin-Nitsche argument given earlier led to the idea of finding a compromise between standard sparse grid spaces and full tensor product spaces. In this section we give an error analysis for some such spaces, those based on optimised sparse grid index sets.

We remember that the index set for the optimised sparse grids are given by

$$J_L^{\mathcal{T}} = \left\{ (l_x, l_t) \ : \ l_x + \frac{l_t}{2} - \mathcal{T} \max \left[l_x, \frac{l_t}{2} \right] \leq (1 - \mathcal{T}) L \right\}.$$

We will refer to the sparse grid space resulting from this choice of index set as follows.

$$\mathcal{X}_L^{\mathcal{T}} := \bigotimes_{(\ell_x, \ell_t) \in J_L^{\mathcal{T}}} \mathcal{W}_{\ell_x}^x \otimes \mathcal{W}_{\ell_t}^t. \tag{6.8}$$

Our goal in this section is to find convergence estimates in these spaces.

As in Chapter 5 the main ingredient used for the convergence proof are norm equivalences, which can be shown using wavelet bases.

Let $\psi_{j,k}$ be a biorthogonal wavelet basis.

Then we recall from Chapter 3 the following norm equivalences. Let

$$\psi = \sum_{(l_x, l_t) \geq 0} w_l \text{ with } w_{l_x, l_t} \in W_{l_x}^x \otimes W_{l_t}^t.$$

Then,

$$\|\psi\|^2_{H^{s_x,s_t}(\Sigma)} \sim \sum_{(l_x,l_t)\geq 0} 2^{2\max\{s_x l_x, s_t l_t\}} \|w_{l_x,l_t}\|^2_{L^2(\Sigma)} \text{ and}$$
$$\|\psi\|^2_{H^{s_x,s_t}_{\text{mix}}(\Sigma)} \sim \sum_{(l_x,l_t)\geq 0} 2^{2s_x l_x + 2s_t l_t} \|w_{l_x,l_t}\|^2_{L^2(\Sigma)}.$$

More specifically for the energy norm and the mix-spaces that we require in the following estimates, we have

$$\|\psi\|^2_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \sim \sum_{(l_x,l_t)\geq 0} 2^{-\max\{l_x,l_t/2\}} \|w_{l_x,l_t}\|^2_{L^2(\Sigma)} \text{ and}$$
$$\|\psi\|^2_{H^{s,\frac{s}{2}}_{\text{mix}}(\Sigma)} \sim \sum_{(l_x,l_t)\geq 0} 2^{2sl_x + sl_t} \|w_{l_x,l_t}\|^2_{L^2(\Sigma)}.$$

Next we combine these two estimates to get an approximation in the energy norm. We choose the discrete approximation $\psi_L = \sum_{(l_x,l_t)\in J_l^{\mathcal{T}}} w_{l_x,l_t}$ and use the best approximation property to get

$$\begin{aligned}
\inf_{\psi_L \in \mathcal{X}_L^{\mathcal{T}}} \|\psi - \psi_L\|^2_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} &\leq \sum_{(l_x,l_t)\notin J_L^{\mathcal{T}}} 2^{-\max\{l_x,l_t/2\}} \|w_l\|^2_{L^2(\Sigma)} \\
&\leq \max_{(l_x,l_t)\notin J_L^{\mathcal{T}}} 2^{-\max\{l_x,l_t/2\} - (2sl_x + sl_t)} \sum_{(l_x,l_t)\notin J_L^{\mathcal{T}}} 2^{(2sl_x+sl_t)} \|w_l\|^2_{L^2(\Sigma)} \qquad (6.9) \\
&\leq \max_{(l_x,l_t)\notin J_L^{\mathcal{T}}} 2^{-\max\{l_x,l_t/2\} - (2sl_x + sl_t)} \|\psi\|^2_{H^{s,s/2}_{\text{mix}}(\Sigma)}.
\end{aligned}$$

for any $-\infty \leq \mathcal{T} \leq 1$.

In order to estimate the convergence we find the maximum for $(l_x, l_t) \notin J_L^{\mathcal{T}}$. The maximum is attained when the negative exponent $\tilde{m} := \max\{l_x, l_t/2\} + 2sl_x + sl_t$ attains its minimum. We have not yet chosen $\mathcal{T}$ and in the following will choose $\mathcal{T}$ to maximise the convergence rate.

$$\begin{aligned}
\tilde{m} &= \min_{(l_x,l_t)\notin J_L^{\mathcal{T}}} (\max\{l_x, l_t/2\} + 2s(l_x + l_t/2)) \\
&= \min_{(l_x,l_t)\notin J_L^{\mathcal{T}}} \Big(2s(l_x + l_t/2) + \max\{l_x, l_t/2\} \Big) \\
&= \min_{(l_x,l_t)\notin J_L^{\mathcal{T}}} \Big(2s(l_x + l_t/2 - \mathcal{T}\max\{l_x, l_t/2\}) + (1 + 2s\mathcal{T})\max\{l_x, l_t/2\} \Big) \\
&= 2s(\lfloor (1-\mathcal{T})L \rfloor + 1) + (1 + 2s\mathcal{T}) \min_{(l_x,l_t)\notin J_L^{\mathcal{T}}} \max\{l_x, l_t/2\}
\end{aligned}$$

The last equation holds since $(l_x, l_t) \notin J_L^{\mathcal{T}}$, if $(l_x + l_t/2) - \mathcal{T}\max\{l_x, l_t/2\}$ were smaller than $(1-\mathcal{T})L + 1$ it would by definition of the index set lie in $J_L^{\mathcal{T}}$.

The function $G(l_x, l_t) := \max\{l_x, \frac{l_t}{2}\}$ is monotonically increasing. Thus, by analogous arguments to those of Lemma 5.3.2 we get

$$\begin{aligned}
\min_{(l_x,l_t)\notin J_L^{\mathcal{T}}} G(l_x, l_t/2) &= \min\left\{G(L+1,0), G(0,2L+1), G\left(\left\lfloor 2\frac{1-\mathcal{T}}{2-\mathcal{T}}L\right\rfloor + 1\right)\right\} \\
&= \min\left\{L+1, \left\lfloor 2\frac{1-\mathcal{T}}{2-\mathcal{T}}L\right\rfloor + 1\right\} \\
&= \begin{cases} L+1, & \mathcal{T} \leq 0 \\ \left\lfloor 2\frac{1-\mathcal{T}}{2-\mathcal{T}}L\right\rfloor + 1, & 0 \leq \mathcal{T} \leq 1. \end{cases}
\end{aligned}$$

We will handle the two different cases $\mathcal{T} < 0$ and $\mathcal{T} \geq 0$ seperately. Firstly, if $\mathcal{T} < 0$:

$$\begin{aligned}
\tilde{m} &= 2s\left[\lfloor(1-\mathcal{T})L\rfloor + 1\right] + (1+2s\mathcal{T})(L+1) \\
&\leq 2s((1-\mathcal{T})L + 1) + (1+2s\mathcal{T})(L+1) \\
&= (1+2s)(L+1) + 2s\mathcal{T}
\end{aligned}$$

On the other hand if $\mathcal{T} \geq 0$:

$$\begin{aligned}
\tilde{m} &= 2s\left[\lfloor(1-\mathcal{T})L\rfloor + 1\right] + (1+2s\mathcal{T})\left(\left\lfloor 2\frac{1-\mathcal{T}}{2-\mathcal{T}}L\right\rfloor + 1\right) \\
&\leq 2s(1-\mathcal{T})L + 2s + (1+2s\mathcal{T})\left(2L\frac{1-\mathcal{T}}{2-\mathcal{T}} + 1\right) \\
&= 2(1-\mathcal{T})\left[s + \frac{1}{2-\mathcal{T}} + \frac{2s\mathcal{T}}{2-\mathcal{T}}\right]L + 2s + 1 + 2s\mathcal{T} \\
&\leq (2s+1+2s\mathcal{T})\left(2L\frac{1-\mathcal{T}}{2-\mathcal{T}} + 1\right)
\end{aligned}$$

Remark 6.2.6. *If we choose $\mathcal{T} = -\frac{1}{2s}$ and combine this estimate with the best approximation estimate from equation (6.9), we get*

$$\begin{aligned}
\inf_{\psi_L \in \mathcal{X}_L^{\mathcal{T}}} \|\psi - \psi_L\|^2_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} &\leq c2^{-(1+2s)L} \sum_{l=(l_x,l_t)\notin J_L^T} 2^{(2sl_x+sl_t)}\|w_l\|^2_{L^2(\Sigma)} \\
&\leq c2^{-(1+2s)L}\|\psi\|^2_{H^{s,\frac{s}{2}}_{mix}(\Sigma)}.
\end{aligned} \tag{6.10}$$

In order to find the convergence rates for this choice of index sets we now need to calculate the cardinality of the optimised sparse grid space $\mathcal{X}_L^{\mathcal{T}}$ in dependence of spatial dimension d and the choice of $\mathcal{T}$.

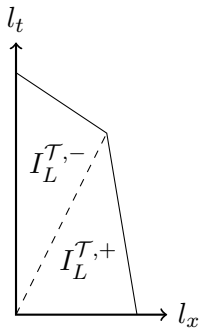

Figure 6.4: The two index sets $I_L^{\mathcal{T},+}$ and $I_L^{\mathcal{T},-}$ for $\mathcal{T} = -2$.

Lemma 6.2.7. *The dimension of the approximation spaces $\mathcal{X}_L^{\mathcal{T}}$ is*

$$\dim \mathcal{X}_L^{\mathcal{T}} \leq c \begin{cases} 2^{d\frac{1-\mathcal{T}}{2-\mathcal{T}}L} & \text{if } 2 < d\frac{1-\mathcal{T}}{2-\mathcal{T}} \\ 2^{2L} & \text{if } d\frac{1-\mathcal{T}}{2-\mathcal{T}} < 2 < (d+1)\frac{1-\mathcal{T}}{2-\mathcal{T}} \\ 2^{(d+1)L\frac{1-\mathcal{T}}{2-\mathcal{T}}} & \text{else.} \end{cases}$$

Proof. The index set corresponding to $\mathcal{X}_L^{\mathcal{T}}$ is

$$J_L^{\mathcal{T}} = \left\{ l_x + \frac{l_t}{2} - \mathcal{T} \max\left\{ l_x, \frac{l_t}{2} \right\} \leq (1-\mathcal{T})L \right\}.$$

In order to calculate the dimension of $\mathcal{X}_L^{\mathcal{T}}$ more easily we will divide this index set into two parts.

$$I_L^{\mathcal{T},+} := \left\{ \frac{l_t}{2} + (1-\mathcal{T})l_x \leq (1-\mathcal{T})L \right\} \cap \left\{ l_x \geq \frac{l_t}{2} \right\}$$

and

$$I_L^{\mathcal{T},-} := \left\{ l_x + \frac{l_t}{2}(1-\mathcal{T}) \leq (1-\mathcal{T})L \right\} \cap \left\{ l_x \leq \frac{l_t}{2} \right\}.$$

These two index sets are shown in Figure 6.4. Now, the dimension is given as

$$\dim \mathcal{X}_L^{\mathcal{T}} = \sum_{l \in I_L^{\mathcal{T}}} 2^{(d-1)l_x + l_t} = \sum_{l \in I_L^{\mathcal{T},+}} 2^{(d-1)l_x + l_t} + \sum_{l \in I_L^{\mathcal{T},-}} 2^{(d-1)l_x + l_t}.$$

We look at the two summands individually.

The index set $I_L^{\mathcal{T},+}$ is simply a triangle, to simplify the sum we will use a transfor-

mation. The index set $I_L^{\mathcal{T},+}$ has the vertices

$$\left((0,0),(L,0),\left(\frac{1-\mathcal{T}}{2-\mathcal{T}}L,\frac{1-\mathcal{T}}{2-\mathcal{T}}2L\right)\right).$$

Using a standard affine transformation onto these vertices we can reparameterise the index set $I_L^{\mathcal{T},+}$ as

$$(l_x,l_t)=\left(\tilde{l}_t+\left\lfloor\tilde{l}_x\frac{1-\mathcal{T}}{2-\mathcal{T}}\right\rfloor,2\left\lfloor\tilde{l}_x\frac{1-\mathcal{T}}{2-\mathcal{T}}\right\rfloor\right).$$

where $\tilde{l}_x=0,...,L$, $\tilde{l}_t=0,...,L-\tilde{l}_x$.

Then we get

$$\begin{aligned}\sum_{l\in I_L^{\mathcal{T},+}}2^{(d-1)l_x+l_t}&=\sum_{\tilde{l}_x=0}^{L}\sum_{\tilde{l}_t=0}^{L-\tilde{l}_x}2^{(d-1)\tilde{l}_t+(d+1)\left\lfloor\tilde{l}_x\frac{1-\mathcal{T}}{2-\mathcal{T}}\right\rfloor}\\&=\sum_{\tilde{l}_x=0}^{L}2^{(d+1)\left\lfloor\tilde{l}_x\frac{1-\mathcal{T}}{2-\mathcal{T}}\right\rfloor}\underbrace{\sum_{l_t=0}^{L-\tilde{l}_x}2^{(d-1)\tilde{l}_t}}_{\leq 2\cdot 2^{(d-1)(L-\tilde{l}_x)}}\\&\leq 2\cdot 2^{(d-1)L}\sum_{l_x=0}^{L}2^{(d+1)\left\lfloor\tilde{l}_x\frac{1-\mathcal{T}}{2-\mathcal{T}}\right\rfloor-(d-1)\tilde{l}_x}\\&\leq 2\cdot 2^{(d-1)L}\sum_{l_x=0}^{L}2^{\left[(d+1)\frac{1-\mathcal{T}}{2-\mathcal{T}}-(d-1)\right]\tilde{l}_x}\end{aligned}$$

since $(d+1)\frac{1-\mathcal{T}}{2-\mathcal{T}}-(d-1)$ is always positive, we get in total

$$\sum_{l\in I_L^{\mathcal{T},+}}2^{(d-1)l_x+l_t}\leq c2^{(d-1)L}2^{(d+1)L\frac{1-\mathcal{T}}{2-\mathcal{T}}-(d-1)L}=c2^{(d+1)L\frac{1-\mathcal{T}}{2-\mathcal{T}}}.$$

Next we use a similar affine transformation on the second index set $I_L^{\mathcal{T},-}$, which has the vertices

$$\left((0,0),(0,2L),\left(\frac{1-\mathcal{T}}{2-\mathcal{T}}L,\frac{1-\mathcal{T}}{2-\mathcal{T}}2L\right)\right)$$

This gives the following reparameterisation.

$$(l_x,l_t)=\left(\left\lfloor\hat{l}_x\frac{1-\mathcal{T}}{2-\mathcal{T}}\right\rfloor,2\hat{l}_t+2\left\lfloor\hat{l}_x\frac{1-\mathcal{T}}{2-\mathcal{T}}\right\rfloor\right).$$

where $\hat{l}_x=0,...,L$, $\hat{l}_t=0,...,L-\hat{l}_x$.

For this index set we get:

$$\begin{aligned}\sum_{l\in I_L^{\mathcal{T},-}} 2^{(d-1)l_x+l_t} &= \sum_{\hat{l}_x=0}^{L}\sum_{\hat{l}_t=0}^{L-\hat{l}_x} 2^{2\hat{l}_t+(d+1)\lfloor \hat{l}_x\frac{1-\mathcal{T}}{2-\mathcal{T}}\rfloor} \\ &= \sum_{\hat{l}_x=0}^{L} 2^{(d+1)\lfloor \hat{l}_x\frac{1-\mathcal{T}}{2-\mathcal{T}}\rfloor} \underbrace{\sum_{\hat{l}_t=0}^{L-\hat{l}_x} 2^{2\hat{l}_t}}_{\leq c2^{2(L-\hat{l}_x)}} \\ &\leq c2^{2L}\sum_{\hat{l}_x=0}^{L} 2^{(d+1)\lfloor \hat{l}_x\frac{1-\mathcal{T}}{2-\mathcal{T}}\rfloor-2\hat{l}_x} \leq c2^{2L}\sum_{\hat{l}_x=0}^{L} 2^{\lceil (d+1)\frac{1-\mathcal{T}}{2-\mathcal{T}}-2\rceil\hat{l}_x}.\end{aligned}$$

We now split into two cases. In those where $(d+1)\frac{1-\mathcal{T}}{2-\mathcal{T}}-2$ is negative the remaining sum is bounded from above by 1 and we get in total

$$\sum_{l\in I_L^{\mathcal{T},-}} 2^{(d-1)l_x+l_t} \leq c2^{2L}.$$

On the other hand if $(d+1)\frac{1-\mathcal{T}}{2-\mathcal{T}}-2$ is positive, we estimate

$$\sum_{l\in I_L^{\mathcal{T},-}} 2^{(d-1)l_x+l_t} \leq c2^{(d+1)\frac{1-\mathcal{T}}{2-\mathcal{T}}L}.$$

Now we can add up the two summands to get the estimate for the dimension of the entire approximation space

$$\begin{aligned}\dim \mathcal{X}_L^{\mathcal{T}} &= \sum_{l\in I_L^{\mathcal{T},+}} 2^{(d-1)l_x+l_t} + \sum_{l\in I_L^{\mathcal{T},-}} 2^{(d-1)l_x+l_t} \\ &\leq c\begin{cases} 2^{d\frac{1-\mathcal{T}}{2-\mathcal{T}}L} & \text{if } 2 < d\frac{1-\mathcal{T}}{2-\mathcal{T}} \\ 2^{2L} & \text{if } d\frac{1-\mathcal{T}}{2-\mathcal{T}} < 2 < (d+1)\frac{1-\mathcal{T}}{2-\mathcal{T}} \\ 2^{(d+1)L\frac{1-\mathcal{T}}{2-\mathcal{T}}} & \text{else.}\end{cases}\end{aligned}$$

□

Remark 6.2.8. *We can rewrite this result for $d = 2$ in a simpler form:*

$$\dim \mathcal{X}_L^{\mathcal{T}} \leq c\begin{cases} 2^{2L} & \textit{if } \mathcal{T} > -1 \\ 2^{3L\frac{1-\mathcal{T}}{2-\mathcal{T}}} & \textit{else.}\end{cases}$$

We now examine the convergence rates in the energy norm for $d = 2$ to find the

optimal choice of $\mathcal{T}$ for that dimension. The same methodology can be applied to higher dimensions.

We now give the main result of this section.

Theorem 6.2.9. *Let $d = 2$ and suppose that $\psi \in H^{s,s/2}_{mix}(\Sigma)$ with $1 \leq s \leq \min\{p_x + 1, p_t + 1\}$. Then for all $\mathcal{T}$ in the interval $[-1, 0)$ the error of the optimised sparse tensor Galerkin approximation $\psi_L \in \mathcal{X}_L^{\mathcal{T}}$ is*

$$\inf_{v \in \mathcal{X}_L^{\mathcal{T}}} \|u - v\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq c N_L^{-\frac{(1+2s)}{4}} \|u\|_{H^{s,\frac{s}{2}}_{mix}(\Sigma)}.$$

This gives the highest convergence rate attained under the constraint $\mathcal{T} < 0$.

Further, if $\mathcal{T} \geq 0$ then the highest convergence rate is reached at $\mathcal{T} = 2 - \sqrt{3 + \frac{1}{2s}}$ and the error is

$$\inf_{v \in \mathcal{X}_L^{\mathcal{T}}} \|u - v\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq c N_L^{-(s(1+\mathcal{T})+\frac{1}{2})\frac{1-\mathcal{T}}{2-\mathcal{T}}} \|u\|_{H^{s,\frac{s}{2}}_{mix}(\Sigma)}.$$

Proof. Above we have almost finished showing this result. We combine the calculation from Lemma 6.2.7 with the best approximation results.

First let $0 > \mathcal{T} > -1$. Then we get

$$\begin{aligned}\inf_{v \in \mathcal{X}_L^{\mathcal{T}}} \|u - v\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} &\leq c N_L^{-\frac{(1+2s)(L+1)+2s\mathcal{T}}{2\cdot 2L}} \|u\|_{H^{s,\frac{s}{2}}_{\text{mix}}(\Sigma)} \\ &\leq c N_L^{-\frac{1+2s}{4}} \|u\|_{H^{s,\frac{s}{2}}_{\text{mix}}(\Sigma)}.\end{aligned}$$

Further, if $\mathcal{T} < -1$,

$$\begin{aligned}\inf_{v \in \mathcal{X}_L^{\mathcal{T}}} \|u - v\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} &\leq c N_L^{-\frac{(1+2s)(L+1)+2s\mathcal{T}}{2\cdot 3\frac{1-\mathcal{T}}{2-\mathcal{T}} L}} \|u\|_{H^{s,\frac{s}{2}}_{\text{mix}}(\Sigma)} \\ &\leq c N_L^{-\frac{1+2s}{6\frac{1-\mathcal{T}}{2-\mathcal{T}}}} \|u\|_{H^{s,\frac{s}{2}}_{\text{mix}}(\Sigma)}.\end{aligned}$$

In this case the convergence rate is highest when $\frac{1-\mathcal{T}}{2-\mathcal{T}}$ is smallest, i.e. when $\mathcal{T} = -1$. This choice gives

$$\inf_{v \in \mathcal{X}_L^{\mathcal{T}}} \|u - v\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq c N_L^{-\frac{1+2s}{4}} \|u\|_{H^{s,\frac{s}{2}}_{\text{mix}}(\Sigma)},$$

as desired.

If $\mathcal{T} > 0$, we have

$$\begin{aligned}\inf_{v\in\mathcal{X}_L^{\mathcal{T}}} \|u-v\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} &\leq cN_L^{-\frac{(1+2s+2s\mathcal{T})(2L\frac{1-\mathcal{T}}{2-\mathcal{T}}+1)}{2\cdot 2L}} \|u\|_{H^{s,\frac{s}{2}}_{\text{mix}}(\Sigma)} \\ &\leq cN_L^{-\frac{1}{2}(1+2s+2s\mathcal{T})\frac{1-\mathcal{T}}{2-\mathcal{T}}} \|u\|_{H^{s,\frac{s}{2}}_{\text{mix}}(\Sigma)}.\end{aligned}$$

We now find the value of $\mathcal{T}$ that maximises $(1+2s+2s\mathcal{T})\frac{1-\mathcal{T}}{2-\mathcal{T}}$. To find the maximum we derive the expression,

$$\frac{d}{d\mathcal{T}}\left[(1+2s+2s\mathcal{T})\frac{1-\mathcal{T}}{2-\mathcal{T}}\right] = 2s\frac{1-\mathcal{T}}{2-\mathcal{T}} - (1+2s+2s\mathcal{T})\frac{1}{(2-\mathcal{T})^2}.$$

Setting this expression to zero gives us the extrema

$$2s\frac{1-\mathcal{T}}{2-\mathcal{T}} - (1+2s+2s\mathcal{T})\frac{1}{(2-\mathcal{T})^2} = 0 \Leftrightarrow \mathcal{T}^2 - 4\mathcal{T} + (1-\frac{1}{2s}) = 0.$$

This gives us $\mathcal{T} = 2-\sqrt{3+\frac{1}{2s}}$ as the value which maximises the expression, as required. □

Remark 6.2.10. *If we choose constant polynomial degrees $p_x = 0$ and $p_t = 0$ and $\mathcal{T} \in (0,1]$, then the regularity is $s = 2$ and the convergence estimate is:*

$$\inf_{\psi_L\in X_L^{\mathcal{T}}} \|\psi-\psi_L\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq N_L^{-3/4}\|\psi\|_{H^{2,1}_{mix}(\Sigma)}$$

Corollary 6.2.11. *Suppose that $g \in \tilde{H}^{s,\frac{s}{2}}(\Sigma)$ with $s \geq \min\{p_x+1, p_t+1\}$. Then the error of the optimised sparse tensor Galerkin solution $\psi_L \in \mathcal{X}_L^{\mathcal{T}}$ to (6.1) has the bounds*

$$\|\psi-\psi_L\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} \leq cN_L^{-\gamma}\|g\|_{H^{2s,s}(\Sigma)},$$

with the convergence rate

$$\gamma = \begin{cases} -(s(1+\mathcal{T})+\frac{1}{2})\frac{1-\mathcal{T}}{2-\mathcal{T}} & \text{if } \mathcal{T} = 2-\sqrt{3+\frac{1}{2s}}, \\ \frac{1+2s}{4} & \text{if } \mathcal{T} > -1. \end{cases}$$

Proof. According to Lemma 2.1.11 we have the following embedding

$$H^{r,\frac{r}{2}}(\Sigma) \subset H^{a,b}(\Sigma) \text{ for } r \geq a+2b.$$

This implies,

$$\|\psi\|_{H^{s,\frac{s}{2}}(\Sigma)} \leq \|\psi\|_{H^{2s,s}(\Sigma)}.$$

Further, Theorem 2.2.7 yields that the mapping

$$V : \tilde{H}^{s,\frac{s}{2}}(\Sigma) \to \tilde{H}^{s+1,\frac{s+1}{2}}(\Sigma)$$

is an isomorphism. This together with Theorem 6.2.9 gives the assertion. □

In the Tables 6.3 and 6.4 we compare the convergence rates obtained with the optimised sparse grids to the convergence rates we proved in the previous section for the standard sparse grids with $\sigma = 1$ and $\sigma = \sqrt{2}$ respectively.

We see that the rates for the optimised sparse grids are lower than those for the standard sparse grids especially for high polynomial degrees. However, they require lower regularity assumptions on the right hand side and have no restrictions on the choice of polynomial degrees.

Further, one can see that in higher dimensions, such as $d = 3$, the optimised sparse grids start yielding higher convergence rates for some configurations of polynomial degree.

Note that the choice $\mathcal{T} = 0$ does not lead to the standard sparse tensor product we are comparing with in $d = 2$ since σ was chosen to be 1. Allowing the same flexibility of scaling in time and space for the index set $J_L^{\mathcal{T}}$ might improve the convergence results.

Standard sparse grids, $d = 2$			Optimised sparse grids, $d = 2$		
(p_x, p_t)	conv. rate γ	reg. r	$\mathcal{T}$	conv. rate γ	reg. r
$(0,0)$	$\frac{7}{6} \sim 1.17$	4	$2 - \sqrt{\frac{7}{2}}$	$\frac{9}{2} - \sqrt{14} \sim 0.76$	2
$(1,0)$	-	-	$2 - \sqrt{\frac{7}{2}}$	$\frac{9}{2} - \sqrt{14}$	2
$(1,1)$	$\frac{13}{6} \sim 2.17$	7	$2 - \frac{\sqrt{13}}{2}$	$\frac{17}{2} - 2\sqrt{13} \sim 1.28$	4
$(3,1)$	-	-	$2 - \frac{\sqrt{13}}{2}$	$\frac{17}{2} - 2\sqrt{13}$	4

Table 6.3: Convergence rates and required regularity assumptions on the right hand side for standard and optimised sparse and for full tensor product discretisations in 2 dimensions.

Standard sparse grids, $d = 3$			Optimised sparse grids, $d = 3$		
(p_x, p_t)	conv. rate γ	reg. r	$\mathcal{T}$	conv. rate γ	reg. r
$(0,0)$	$\frac{5}{8} = 0.625$	3	$2 - \sqrt{\frac{7}{2}}$	$\frac{9}{2} - \sqrt{14} \sim 0.76$	2
$(1,0)$	$\frac{9}{8} = 1.125$	5	$2 - \sqrt{\frac{7}{2}}$	$\frac{9}{2} - \sqrt{14}$	2
$(1,1)$	$\frac{9}{8} = 1.125$	5	$2 - \frac{\sqrt{13}}{2}$	$\frac{17}{2} - 2\sqrt{13} \sim 1.28$	4
$(3,1)$	$\frac{17}{8} = 2.125$	9	$2 - \frac{\sqrt{13}}{2}$	$\frac{17}{2} - 2\sqrt{13}$	4

Table 6.4: Convergence rates and required regularity assumptions on the right hand side for standard and optimised sparse and for full tensor product discretisations in 3 dimensions.

6.3 The Sparse Grid Combination Technique

The combination technique for the solution of sparse grid problems was first introduced in [33]. The basic idea behind the technique is to find a sparse grid approximation using a linear combination of smaller full grid solutions. The advantage of this is that the necessary full grids are much smaller than the full sparse grid and can be computed more quickly, while still giving the same accuracy. It also gives an easier implementation since the need for the solution in a sparse grid space is replaced with the solution of several full grids. Further, the solution of the systems corresponding to these full grids can be performed in parallel, see e.g. [26] and [30].

No general proof of convergence for the combination technique exists. However, it has been shown in [29] that it produces the same order of convergence with the same complexity as the Galerkin approach in the standard sparse tensor product case for certain elliptic operators.

The combination technique can also be used for the discretisation with more general sparse grid spaces, however, there the rates of convergence are not clear.

First we revisit the setting of our specific problem. We are working on a tensor product domain $\Sigma = \mathcal{I} \times \Gamma$. As before, our discrete spaces in time and space are

$$\mathcal{X}_0^x \subset \mathcal{X}_1^x \subset ... \subset \mathcal{X}_{l_x}^x \subset ... \subset L^2(\Gamma) \subset H^{-\frac{1}{2}}(\Gamma) \text{ and}$$
$$\mathcal{X}_0^t \subset \mathcal{X}_1^t \subset ... \subset \mathcal{X}_{l_t}^t \subset ... \subset L^2(\mathcal{I}) \subset H^{-\frac{1}{4}}(\mathcal{I}).$$

We are solving one of the following problems:

Find $\varphi \in H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)$ such that

$$\langle V\varphi, \eta\rangle = \langle g, \eta\rangle, \qquad \text{for all } \eta \in H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma) \quad \text{(Direct method)}$$

$$\text{or} \quad \langle V\varphi, \eta\rangle = \langle (\frac{1}{2} + K)g, \eta\rangle, \quad \text{for all } \eta \in H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma) \quad \text{(Indirect method)}$$

Before giving the combination technique we first define the following projection.

Definition 6.3.1. *Let $\hat{\Pi}_{l_x,l_t}$ be a mapping*

$$\hat{\Pi}_{l_x,l_t} \,:\, H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma) \to \mathcal{X}^x_{l_x} \otimes \mathcal{X}^t_{l_t},$$

which satisfies Galerkin-orthogonality

$$\langle V(\varphi - \hat{\Pi}_{l_x,l_t}\varphi), v\rangle = 0, \qquad \forall v \in \mathcal{X}^x_{l_x} \otimes \mathcal{X}^t_{l_t}.$$

We refer to this projection as the Galerkin projection.

The Galerkin projection is well-defined due to the coercivity of the single-layer operator V.

Then we define the combination technique sparse grid solution φ_L using the Galerkin projection:

$$\varphi_L = \left(\sum_{l=0}^{L} \hat{\Pi}_{l,L-l}\varphi - \sum_{l=0}^{L} \hat{\Pi}_{l-1,L-l}\varphi \right) \in \mathcal{X}^\sigma_L, \ \sigma = 1. \tag{6.11}$$

This combination of spaces is shown in Figure 6.5. Essentially one adds the spaces denoted by $+$ and then subtracts the spaces denoted by $-$ on the figure.

Note that another of the advantages of the combination technique is that we solve only systems of full tensor products and do not require a multilevel decomposition. This gives us greater flexibility in the choice of basis functions.

Now we consider all such spaces $\mathcal{X}^x_{l_x} \otimes \mathcal{X}^t_{l_t}$ such that

$$l_x + l_t = L - l, \ l = 0, 1, \ l_x, l_t \geq 0.$$

We look at the solution in one of the full tensor product spaces $\mathcal{X}^x_{l_x} \otimes \mathcal{X}^t_{l_t}$. The related

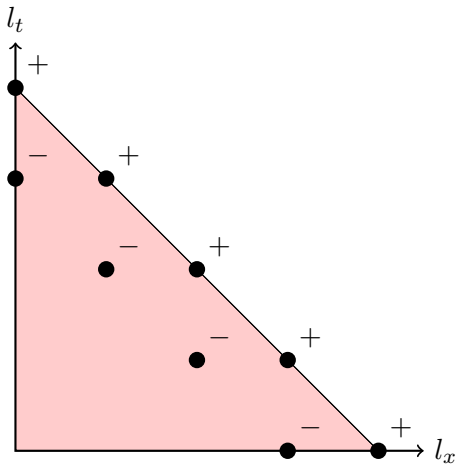

Figure 6.5: The sign contributions of the subspaces used for the combination technique for standard sparse grids with $\sigma = 1$.

Galerkin solution φ_{l_x,l_t} is the solution of

$$\begin{aligned}
&\text{Find } \varphi_{l_x,l_t} \in \mathcal{X}^x_{l_x} \otimes \mathcal{X}^t_{l_t} \text{ such that} \\
&\qquad \langle V\varphi_{l_x,l_t}, \eta\rangle = \langle g, \eta\rangle, \qquad \text{for all } \eta \in \mathcal{X}^x_{l_x} \otimes \mathcal{X}^t_{l_t} \quad \text{(Direct method)} \\
&\text{or} \quad \langle V\varphi_{l_x,l_t}, \eta\rangle = \langle (\frac{1}{2} + K)g, \eta\rangle, \quad \text{for all } \eta \in \mathcal{X}^x_{l_x} \otimes \mathcal{X}^t_{l_t} \quad \text{(Indirect method).}
\end{aligned}$$

Further, let a_{l_x,l_t} be the vector corresponding to the solution φ_{l_x,l_t}, i.e.

$$\varphi_{l_x,l_t} = \sum_{k_1,k_2} a_{k_1,k_2} b_{k_1,k_2}(x),$$

where b_{k_1,k_2} are the basis functions of $\mathcal{X}^x_{l_x} \otimes \mathcal{X}^t_{l_t}$.

Then the summation of two vectors of different sizes is calculated as follows. Let the vectors u and v have the coefficients a_{j_x,j_t} and c_{j_x,j_t} respectively. Then,

$$u_{l_x,l_t} + v_{k_x,k_t} = \sum_{(j_x,j_t)} (a_{j_x,j_t} + c_{j_x,j_t}) b_{j_x,j_t}(x),$$

where unknown coefficients are assumed to be 0.

Now we combine the vector solutions u_{l_x,l_t} to the problems in $\mathcal{X}^x_{l_x} \otimes \mathcal{X}^t_{l_t}$ according to equation (6.11), giving us a sparse grid approximation.

Remark 6.3.2. *If we want to use the combination technique for an anisotropic sparse grid index set, i.e. for a set of the form*

$$\hat{I}_L^\sigma = \{(l_x, l_t) \ : \ \sigma l_x + l_t/\sigma \leq L\}.$$

Then the formula is changed as follows (see [29])

$$\lceil \sigma^2 l_x \rceil + l_t = \lceil \sigma L \rceil - l, \ l = 0, 1.$$

6.4 Numerical Experiments

In this section we verify the given convergence rates with numerical experiments. We start with experiments for the standard sparse grid rule. We use the tests described in more detail in 5.4.

We solve the Dirichlet problem on a circle with radius 1, i.e. we want to find $u : Q \to \mathbb{R}$ satisfying:

$$\begin{aligned} (\partial_t - \Delta)u &= 0, && \text{in } \mathcal{I} \times B_1(0) \\ u &= 0, && \text{at } \{t = 0\} \times B_1(0) \\ \gamma_0 u &= g, && \text{in } \mathcal{I} \times \partial B_1(0), \end{aligned} \tag{6.12}$$

where we choose the right hand side $g(\varphi, t) = t^2 \cos(\varphi)$. The exact boundary flux for this problem is (5.9).

In Figure 6.6 we compare the convergence rates of the square of the energy norm of the full tensor product discretisation and the standard sparse grid discretisations with $\sigma = 1$ in both. The convergence rate for the full tensor product discretisation, namely $\frac{15}{11}$, is as expected from Chapter 5. The convergence rates are given in Table 5.2. The expected convergence rate for the square of the energy norm for the standard sparse grid method is $\frac{14}{6}$. These rates are summarised in Tables 6.1 and 6.2. The tests show a correspondence to the expected rates.

Tests for the optimised sparse grid index sets are not given here since the index set only starts diverging from the standard sparse grid index set with $\sigma = \sqrt{2}$ at $L = 8$, which makes it difficult to confirm the expected rates.

Lastly, we give some numerical results for the combination technique. The left plot in Figure 6.7 shows convergence of the energy norm against the total number of degrees

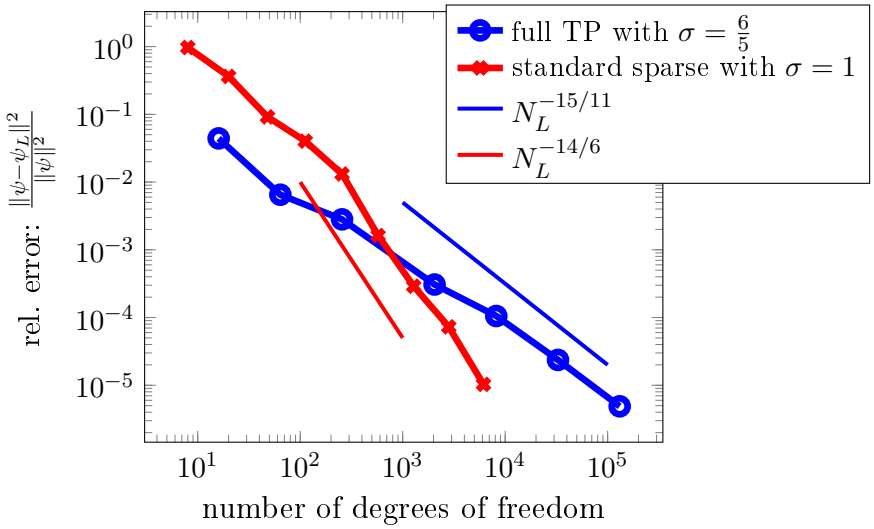

Figure 6.6: Convergence of the squares of the energy norm for the right hand side $g(\varphi, t) = t^2 \cos(\varphi)$ on a circle of radius 1.

of freedom. As expected, the convergence rates are identical to those obtained by implementing the sparse grid method using a multilevel decomposition. However, as the right plot in Figure 6.7 shows the combination technique provides a large improvement in the time taken for the calculation.

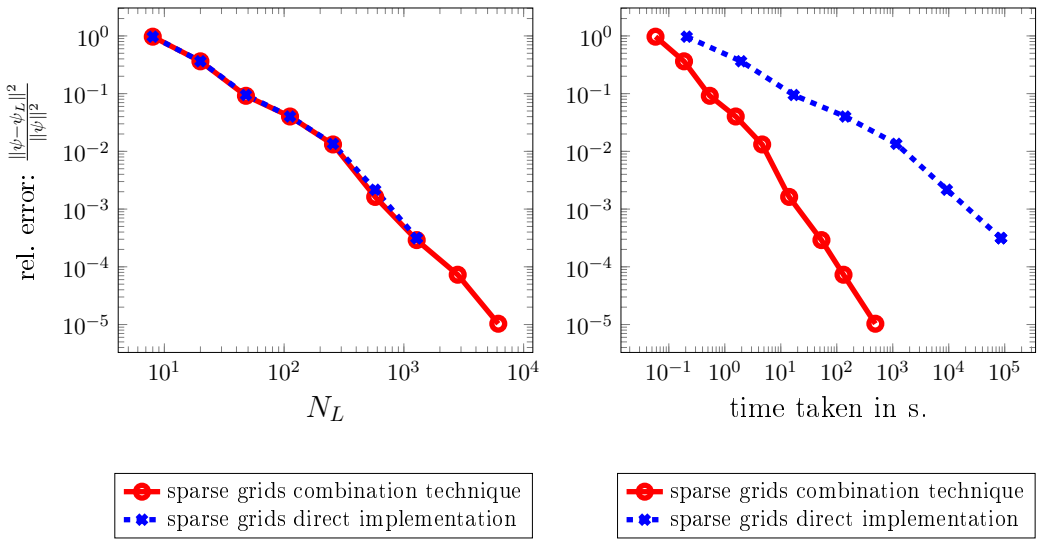

Figure 6.7: Convergence of the relative error in the energy norm squared versus number of degrees of freedom (left) and time taken in seconds (right) for the standard sparse grid space with the combination technique and without.

Chapter 7

Matrix Compression

In this chapter we discuss the compression of the matrix of the single-layer heat operator. In general, the discretisation of boundary integral equations leads to densely populated matrices. The resulting linear systems cannot be solved in linear time. One way to regain sparse matrices is to use a wavelet basis, and then remove small entries with a matrix compression.

First, we give some new results derived for the matrix compression in space using a piecewise constant wavelet basis. This allows us to reduce the number of non-zero matrix entries in each time block from $O(N_x^2)$ to $O(N_x)$.

Results on the matrix compression have already been derived in [8] for B-spline wavelets in two dimensions. They have shown that the use of B-spline wavelet basis functions in time and space allows the compression of the matrix of the single-layer heat operator. This reduces the number of non-zero matrix entries from $O(N_x^2 N_t^2)$ to $O(N_x N_t)$. These results are summarised in Section 7.4.

These results can not be applied to piecewise constant wavelets due to the low number of vanishing moments of the Haar wavelet. In the first part of this section we show a matrix compression using a piecewise constant wavelet with three vanishing moments in time. We use piecewise constant polynomial basis functions in time instead of a wavelet basis defined on intervals. In Chapter 4 we showed that we only need to store $O(N_t)$ time blocks when we are using piecewise constant basis functions and we only need to invert a symmetric positive definite sparse matrix with $O(n_x)$ entries. This means it is not necessary to use wavelet basis functions in time as well.

This chapter concludes with some remarks on implementational issues such as reusing calculated integrals and calculating the distances between supports of the basis func-

tions. Finally we give numerical results for the piecewise constant wavelet basis.

7.1 Background and Notation

We discretise the integral operator V by a wavelet basis ψ_{jk} in space and piecewise constant functions in time. The construction of several suitable wavelet bases was described in Chapter 3.

Let G be the matrix of the single layer heat potential. For ease of notation we denote the block matrix G_{mn} corresponding to the m-th and n-th time interval by v, more precisely

$$v_{\alpha,\beta} := (G_{mn})_{\alpha,\beta} = \int_\Gamma \int_\Gamma g_{mn}(x-y) b_\alpha(x) b_\beta(y) ds_y \, ds_x,$$

where b_α and b_β denote the basis functions in space. Further, we remember that we denoted the time integrated kernel by g_{mn}.

The discretisation by biorthogonal wavelet bases leads to numerically sparse matrices v. In the first compression step, the matrix entries, for which the distances of the supports of the corresponding ansatz and test functions are bigger than a certain cut-off parameter, are set to zero. Since the resulting matrix is still not sparse, the second compression step sets some of the matrix entries with overlapping supports to zero as well.

This has been covered extensively in the case of elliptic equations, see e.g. [34] and [45]. Here we apply similar arguments to the case of the heat equation.

In the following we prove that the matrix compression does not result in a loss of accuracy, for a piecewise constant wavelet with three vanishing moments (see Section 3.3.1). In this case the mother wavelet is:

$$\psi(x) := \begin{cases} -\frac{1}{8} & x \in [-1, 0], \\ 1 & x \in [0, \frac{1}{2}], \\ -1 & x \in [\frac{1}{2}, 1], \\ \frac{1}{8} & x \in [1, 2], \\ 0 & \text{otherwise.} \end{cases}$$

We denoted the parameterisation of the boundary Γ by γ. Thus, we can define the wavelet basis functions as

$$\psi_{jk} = (\tilde{\psi}_{jk} \circ \gamma^{-1})(x), \quad x \in \Gamma,$$

where $\tilde{\psi}_{jk} = 2^{j/2}\psi(2^j x - k)$.

Further, we denote the supports of the wavelet basis functions by

$$\Omega_{j,k} := \text{convhull}\{x \in \Gamma \ : \ \psi_{jk}(\gamma^{-1}(x)) \neq 0\}.$$

7.1.1 Differentiation Rules

To prove that the matrix compression does not result in a loss of accuracy, we will need the following two well-known differentiation rules.

Lemma 7.1.1 (Formula of Fáa di Bruno, [21]). *Let g be defined in a neighborhood of x and have derivatives of order up to n at x. Further, let f be defined in a neighborhood of $g(x)$ and have derivatives of order up to n at $g(x)$. Then*

$$\partial_x^n f(g(x)) = \sum_{I_n} \frac{n!}{\prod_{j=1}^n k_j!} \partial_x^a \left(f(g(x))\right) \prod_{j=1}^n \left(\frac{\partial_x^j g(x)}{j!}\right)^{k_j},$$

with $a = \sum_{j=1}^n k_j$ and $I_n := \{k_1 + 2k_2 + ... + nk_n = n\}$.

Corollary 7.1.2. *Applying this rule in the case $\partial_x^k g(x) = 0$, for $k > 2$ gives*

$$\partial_x^n f(g(x)) = \sum_{k_1+2k_2=n} \frac{n!}{k_1!k_2!} \partial_x^{k_1+k_2} \left(f(g(x))\right) \left(\partial_x g(x)\right)^{k_1} \left(\frac{\partial_x^2 g(x)}{2}\right)^{k_2}.$$

Lemma 7.1.3 (Leibniz Rule, [1]). *Let f and g have derivatives of order up to n. Then their product $f \cdot g$ also has derivatives of order up to n and its n-th derivative is given by*

$$\partial_x^n (f(x)g(x)) = \sum_{k=0}^n \binom{n}{k} \partial_x^k f(x) \partial_x^{n-k} g(x).$$

Corollary 7.1.4. *Applying this rule in the case $\partial_x^k g(x) = 0$, for $k > 1$ gives*

$$\begin{aligned} \partial_x^n (f(x)g(x)) &= \sum_{k=0}^n \binom{n}{k} \partial_x^k f(x) \underbrace{\partial_x^{n-k} g(x)}_{=0, \ \text{for } k<n-1} \\ &= \partial_x^n (f(x))g(x) + n\partial_x^{n-1}(f(x))\partial_x(g(x)). \end{aligned}$$

We will also need an expression for derivatives of the exponential integral function $\mathrm{E}_1(x)$.

Lemma 7.1.5. *For any $n \geq 1$, $x > 0$ there holds*

$$\partial_x^n(\mathrm{E}_1(x)) = e^{-x}x^{-n}n!(-1)^n \sum_{k=0}^{n-1} \frac{x^k}{k!}. \tag{7.1}$$

Proof of Lemma 7.1.5. By definition of the exponential integral function (see Definition 4.2.2), the first derivative is clearly

$$\partial_x(\mathrm{E}_1(x)) = \partial_x \left(\int_x^\infty e^{-t}t^{-1}dt \right) = -e^{-x}x^{-1}.$$

We can now use the Leibniz rule (see Lemma 7.1.3) to find the higher derivatives

$$\begin{aligned}\partial_x^{n+1}(\mathrm{E}_1(x)) = \partial_x^n(-e^{-x}x^{-1}) &= -\sum_{k=0}^{n} \binom{n}{k} (\partial_x^k e^{-x})(\partial_x^{n-k}x^{-1}) \\ &= -\sum_{k=0}^{n} \frac{n!}{k!}(-1)^k e^{-x}(-1)^{n-k}x^{-1-(n-k)}.\end{aligned}$$

Rearranging the terms of the sum gives the required result. □

7.2 First compression step

In the first matrix compression we set to zero those matrix entries that correspond to wavelet basis functions with supports that are far apart. We denote this compressed matrix by v^ϵ, its entries are given by

$$(v_{(j,k),(j',k')})^\epsilon = \begin{cases} 0, & \text{if } \operatorname{dist}(\Omega_{j,k}, \Omega_{j',k'}) > B_{j,j'}, \\ v_{(j,k),(j',k')}, & \text{else.} \end{cases} \tag{7.2}$$

where the cut-off parameter is

$$B_{j,j'} \geq a \max \left\{ 2^{-j}, 2^{-j'}, 2^{\frac{J(2\delta+1)-j(4+\delta)-j'(4+\delta)}{5}} \right\}, \tag{7.3}$$

with $a, \delta \in \mathbb{R}$, $a > 1$, $2 < \delta < 3$ and J denoting the highest level.

Next, we show that by setting these entries to zero we did not lose accuracy or the stability of the underlying Galerkin scheme. To show this, we need estimates for the derivatives of the time-integrated kernel. We start by showing the following lemma.

Lemma 7.2.1. *As in equation (4.15) let*

$$f_l(x-y) = l\,\mathrm{E}_1(a_l) + a_l\,\mathrm{E}_1(a_l) - le^{-a_l},$$

where $a_l = \frac{\|x-y\|^2}{lh_t}$. *Then, there exists a constant* $c_{\alpha,\beta,l} > 0$ *independent of* x, y *such that the derivatives of this term fulfill the following estimates*

$$\left|\partial_x^\alpha \partial_y^\beta f_l(a_l)\right| \leq c_{\alpha,\beta,l}\|x-y\|^{-(\alpha+\beta)},$$

where $l \geq 1$, $\alpha+\beta > 0$, $x, y \in \Gamma \subset \mathbb{R}^d$ *and* $x \neq y$.

Proof of Lemma 7.2.1. Due to the symmetry of f_l it suffices to find the derivatives with respect to x. The derivatives with respect to y have the same form.

For ease of notation we set $z := x - y$ in the following.

Combining the formula of Fáa di Bruno (see Lemma 7.1.1) and Lemma 7.1.5, which gives the derivatives of E_1 we get:

$$\begin{aligned}\partial_x^n(\mathrm{E}_1(a_l)) &= \sum_{k_1+2k_2=n} \frac{n!}{k_1!k_2!}(-1)^{k_1+k_2}F_{k_1+k_2}(a_l)(2\|z\|)^{k_1}\left(\frac{1}{lh_t}\right)^{k_1+k_2} \\ &= \sum_{k_1+2k_2=n} b_{k_1,k_2}F_{k_1+k_2}(a_l)a_l^{k_1/2},\end{aligned} \tag{7.4}$$

where we denote the n-th derivative of E_1 by F_n. Thus,

$$F_0(x) := \mathrm{E}_1(x),$$

and

$$F_n(x) := e^{-x}x^{-n}(-1)^n n!\sum_{k=0}^{n-1}\frac{x^k}{k!}, \quad n \geq 1. \tag{7.5}$$

The coefficients of the sum are given by

$$b_{k_1,k_2} = \frac{n!}{k_1!k_2!}2^{k_1}\left(\frac{1}{lh_t}\right)^{k_1+k_2-k_1/2}.$$

Next, we find derivatives for terms of the form $\mathrm{E}_1(a_l)a_l$. We use the Leibniz rule (see

Lemma 7.1.3) and the definition of F_n given in (7.5):

$$\begin{aligned}\partial_x^n \mathrm{E}_1(x)x &= n\partial_x^{n-1}\mathrm{E}_1(x) + (\partial_x^n \mathrm{E}_1(x))x \\ &= nF_{n-1}(x) + xF_n(x).\end{aligned}$$

For $n \geq 2$ this expression can be simplified as follows.

$$\begin{aligned}&xF_n(x) + nF_{n-1}(x) \\ &\quad= e^{-x}x^{-n}(-1)^n n! \sum_{k=0}^{n-1}\left(\frac{x^{k+1}}{k!}\right) + e^{-x}x^{-n+1}(-1)^{n-1}(n-1)!\sum_{k=0}^{n-2}\left(\frac{x^k}{k!}\right) \\ &\quad= e^{-x}x^{-n}(-1)^n n! \underbrace{\left(\sum_{k=0}^{n-1}\frac{x^{k+1}}{k!} - \sum_{k=0}^{n-2}\frac{x^k}{k!}\right)}_{=\dfrac{x^n}{(n-1)!}} \\ &\quad= e^{-x}(-1)^n n.\end{aligned}$$

Taken together with the formula of Fáa di Bruno this gives (for $n \geq 2$)

$$\partial_x^n(\mathrm{E}_1(a_l)a_l) = \sum_{k_1+2k_2=n} b_{k_1,k_2}\Big(e^{-a_l}(-1)^{k_1+k_2}(k_1+k_2)\Big)a_l^{k_1/2}. \tag{7.6}$$

Lastly, we have terms of the form e^{-a_l}. These terms are easy to derive using the formula of Fáa di Bruno:

$$\begin{aligned}\partial_x^n e^{-a_l} &= \sum_{k_1+2k_2=n} \frac{n!}{k_1!k_2!}(2\|z\|)^{k_1}\left(\frac{1}{lh_t}\right)^{k_1+k_2}(-1)^{k_1+k_2}e^{-a_l} \\ &= \sum_{k_1+2k_2=n} b_{k_1,k_2}a_l^{k_1/2}(-1)^{k_1+k_2}e^{-a_l}.\end{aligned} \tag{7.7}$$

Now we can add up the three terms (7.4), (7.6) and (7.7) to find an expression for the derivatives of $f_l(a_l)$.

$$\partial_x^n f_l(a_l) = \sum_{k_1+2k_2=n} b_{k_1,k_2}\Big(lF_{|k|}(a_l) + (-1)^{|k|}e^{-a_l}(|k|-l)\Big)a_l^{k_1/2},\ n \geq 2,$$

where $|k| = k_1 + k_2$.

Reordering these terms gives

$$\partial_x^n f_l(a_l) = \sum_{k_1+2k_2=n} \tilde{b}_{k_1,k_2} a_l^{-n/2} \Big(|k|! e^{-a_l} l \sum_{i=0}^{|k|-1} \frac{a_l^i}{i!} + e^{-a_l} a_l^{|k|} (|k| - l) \Big), \tag{7.8}$$

where $\tilde{b}_{k_1,k_2} = (-1)^{k_1+k_2} b_{k_1,k_2}$.

We note two inequalities that hold for all $x > 0$ and for all $n \geq 1$

$$e^{-x} \sum_{k=0}^{n-1} \frac{x^k}{k!} \leq 1,$$

and

$$e^{-x} x^n \leq \left(\frac{n}{e}\right)^n.$$

Taking the two inequalities above together and inserting them into (7.8) gives us the following estimate for $n \geq 2$.

$$\partial_x^n f_l(a_l) \leq \sum_{k_1+2k_2=n} \tilde{b}_{k_1,k_2} a_l^{-n/2} \Big(l(k_1+k_2)! + \left(\frac{k_1+k_2}{e}\right)^{k_1+k_2} (k_1+k_2-l) \Big).$$

Now we estimate the absolute value of this sum for $n \geq 2$:

$$|\partial_x^n f_l(a_l)| \leq c_{n,0,l} |a_l|^{-n/2},$$

where

$$c_{n,0,l} = \left| \sum_{k_1+2k_2=n} \tilde{b}_{k_1,k_2} \Big(l(k_1+k_2)! + \left(\frac{k_1+k_2}{e}\right)^{k_1+k_2} (k_1+k_2-l) \Big) \right|.$$

The remaining case to be covered is $n = 1$. Deriving f_l once gives

$$\begin{aligned} \partial_x f_l(a_l) &= (\partial_x a_l)(-e^{-a_l} a_l^{-1} + E_1(a_l) - e^{-a_l} + e^{-a_l}) \\ &= \frac{2}{\sqrt{l h_t}} a_l^{1/2} (E_1(a_l) - e^{-a_l} a_l^{-1}). \end{aligned}$$

We can estimate the absolute value of this derivative as follows

$$\begin{aligned} |\partial_x f_l(a_l)| &\leq |a_l|^{-1/2} b_1 \underbrace{\left| E_1(a_l) a_l - e^{-a_l} \right|}_{\leq 2} \\ &\leq c_{0,0,l} \|z\|^{-1}, \end{aligned}$$

where the constants are given by $c_{0,0,l} = 4\sqrt{lh_t}$. □

Corollary 7.2.2. *The time integrated kernel $g_{m,m-l}(x-y)$ fulfills the following estimate*

$$\left|\partial_x^\alpha \partial_y^\beta g_{m,m-l}(x-y)\right| \leq c_{\alpha,\beta,l}\|x-y\|^{-(\alpha+\beta)},$$

where $\alpha+\beta > 0,\ x, y \in \Gamma \subset \mathbb{R}^d$ and $x \neq y$.

Proof of Corollary 7.2.2. According to equation (4.14) the time integrated kernel can be written as follows

$$\begin{aligned} g_{m,m}(x) &= h_t(4\pi)^{-d/2} f_1(x), \\ g_{m,m-1}(x) &= h_t(4\pi)^{-d/2}(-2f_1(x) + f_2(x)), \\ g_{m,m-l}(x) &= h_t(4\pi)^{-d/2}(f_{l-1}(x) + f_{l+1}(x) - 2f_l(x)), \quad l > 1. \end{aligned}$$

Then, combining the Lemma 7.2.1 and the triangle inequality, the required estimate follows immediately. □

We now use the above lemma to show that the matrix entries, that are set to zero during the first matrix compression, are sufficiently small.

Lemma 7.2.3. *Let* $\operatorname{dist}(\Omega_{j,k}, \Omega_{j',k'}) > 0$. *Then, there exists a constant $c_l > 0$ depending only on l, such that the coefficients of v fulfill*

$$\left|v_{(j,k),(j',k')}\right| = \left|\langle V\psi_{jk}\chi_m, \psi_{j'k'}\chi_n\rangle\right| \leq c_l 2^{-\frac{7}{2}(j+j')} \operatorname{dist}(\Omega_{j,k}, \Omega_{j',k'})^{-6},$$

where χ_m were the piecewise constant basis functions used in time.

Proof. The entries of the matrix v are given by

$$v_{(j,k),(j',k')} = \int_{\Omega_{j,k}} \int_{\Omega_{j',k'}} g_{mn}(x-y)\psi_{jk}(x)\psi_{j'k'}(y)dydx, \tag{7.9}$$

where ψ_{jk} are the wavelet basis functions and $\Omega_{j,k}$ gives their support.

Let $x \in \Omega_{j,k} = \operatorname{supp}(\psi_{jk})$ be fixed. Then, we can approximate the function

$$y \mapsto g_{mn}(x-y),$$

by a Taylor series (see e.g. Chapter 25, [1]). We cut off this Taylor series after the second term, giving

$$g_{mn}(x-y) = \sum_{\alpha \leq 2} c_\alpha(y_0, x)(y-y_0)^\alpha + R(y, y_0, x), \tag{7.10}$$

where R is the remainder. We will write this remainder in its integral form for convenience

$$R(y, y_0, x) = c(x)(y - y_0)^3 \int_0^1 (1 - m)^2 \partial_y^3 g_{mn}(x - \tilde{y}_m) dm,$$

where we write $\tilde{y}_m = y_0 + m(y - y_0)$.

The basis functions ψ_{jk} have three vanishing moments. This means that

$$\int_{\Omega_{j',k'}} \psi_{j'k'}(y)(p \circ \gamma^{-1})(y) = 0,$$

for all polynomials p of degree less than three. As such, if we insert (7.10) into equation (7.9), all polynomials of degree less than three vanish. More precisely, the terms

$$\sum_{\alpha \leq 2} c_\alpha(y_0, x)(y - y_0)^\alpha$$

vanish and we are left with the integral of the remainder R.

We still have a remaining dependence on x. To remove it we form a Taylor series with regard to x around the point $x_0 \in \Omega_{j,k}$. Let $y \in \text{supp}(\psi_{j'k'})$ be fixed. Then the Taylor-series of order two of the function

$$x \mapsto R(y, y_0, x)$$

around x_0 is given by

$$R(y, y_0, x) = \sum_{\beta \leq 2} c_\beta(y, x_0)(x - x_0)^\beta + R_1(y, y_0, x, x_0). \tag{7.11}$$

Here R_1 denotes the remainder of the second Taylor series.

We insert (7.11) into equation (7.9) and since $\psi_{jk}(x)$ have three vanishing moments, the terms

$$\sum_{\beta \leq 2} c_\beta(y, x_0)(x - x_0)^\beta$$

vanish from the integral and we are left with the remainder

$$R_1(y, y_0, x, x_0) = c(x - x_0)^3 \int_0^1 (1 - m)^2 \partial_x^3 R(y, y_0, \tilde{x}_m, x_0) dm,$$

where $\tilde{x}_m = x_0 + m(x - x_0)$.

Next, we estimate the absolute value of R_1, the remainder of the second Taylor polynomial. This gives

$$\begin{aligned}|R_1(y,y_0,x,x_0)| &\leq c\|x-x_0\|^3\left|\int_0^1 (1-m)^2\partial_x^3 R(y,y_0,\tilde{x}_m,x_0)dm\right| \\ &\leq c\|x-x_0\|^3\|y-y_0\|^3\Big|\int_0^1\int_0^1 (1-m_1)^2(1-m_2)^2\partial_x^3\partial_y^3 g_{mn}(\tilde{x}_{m_1}-\tilde{y}_{m_2})dm_1dm_2\Big|.\end{aligned}$$

Since we assumed $\operatorname{dist}(\Omega_{j,k},\Omega_{j',k'})>0$, we have $\tilde{x}_{m_1}\neq\tilde{y}_{m_2}$. Consequently by Corollary 7.2.2, $g_{mn}(\tilde{x}_{m_1}-\tilde{y}_{m_2})$ is bounded. This means that we can estimate the integral as follows,

$$\begin{aligned}&\Big|\int_0^1\int_0^1 (1-m_1)^2(1-m_2)^2\partial_x^3\partial_y^3 g_{mn}(\tilde{x}_{m_1}-\tilde{y}_{m_2})dm_1dm_2\Big| \\ &\quad\leq \sup_{\substack{x\in\Omega_{j,k}\\ y\in\Omega_{j',k'}}}\left|\partial_x^3\partial_y^3 g_{mn}(x-y)\right|\cdot\underbrace{\left|\int_0^1\int_0^1 (1-m_1)^2(1-m_2)^2dm_1dm_2\right|}_{=\frac{1}{9}}.\end{aligned}$$

Next, we use Corollary 7.2.2 to estimate the derivatives of the time-integrated kernel g_{mn}. Clearly we have that

$$\begin{aligned}\sup_{\substack{x\in\Omega_{j,k}\\ y\in\Omega_{j',k'}}}\left|\partial_x^3\partial_y^3 g_{mn}(x-y)\right| &\leq c\sup_{\substack{x\in\Omega_{j,k}\\ y\in\Omega_{j',k'}}}\|x-y\|^{-6} = c\sup_{\substack{x\in\Omega_{j,k}\\ y\in\Omega_{j',k'}}}\operatorname{dist}(x,y)^{-6} \\ &\leq c\operatorname{dist}(\Omega_{j,k},\Omega_{j',k'})^{-6}.\end{aligned}$$

Now we are ready to estimate the absolute value of the matrix entries:

$$\begin{aligned}\left|v_{(j,k),(j',k')}\right| &= \left|\int_{\Omega_{j,k}}\int_{\Omega_{j',k'}} g_{mn}(x-y)\psi_{jk}(x)\psi_{j'k'}(y)dydx\right| \\ &\leq c\operatorname{dist}(\Omega_{j,k},\Omega_{j',k'})^{-6}\left|\int_{\Omega_{j,k}}\int_{\Omega_{j',k'}}\|x-x_0\|^3\|y-y_0\|^3\psi_{jk}(x)\psi_{j'k'}(y)dydx\right|\end{aligned}$$

We showed in Section 3.3.1 that the supports of the wavelet basis functions have length $3\cdot 2^{-j}$. It folllows, that the distance between x and x_0 can be at most $3\cdot 2^{-j}$. Analogously, the distance between y and y_0 is at most $3\cdot 2^{-j'}$. Thus,

$$\|x-x_0\|^3\|y-y_0\|^3\leq c2^{-3(j+j')}.$$

It remains to estimate the integrals over the wavelet basis functions

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k).$$

Using their properties given in 3.3.1 we get

$$\left|\int_{\Omega_{j,k}} \psi_{jk}(x)dx\right| \leq \int_{\Omega_{j,k}} \underbrace{|\psi_{jk}(x)|}_{\leq 2^{j/2}} dx \leq 3 \cdot 2^{j/2}2^{-j}.$$

Taken together we have

$$\left|v_{(j,k),(j',k')}\right| \leq c2^{-3(j+j')}2^{-j/2}2^{-j'/2}\operatorname{dist}(\Omega_{j,k}, \Omega_{j',k'})^{-6}$$

as required.

□

Remark 7.2.4. *This proof only used the fact that the wavelets have three vanishing moments, which means that any wavelet with three vanishing moments can be used. If a wavelet with higher vanishing moments is used, a higher proportion of the matrix entries are small.*

Let R be the matrix containing the error made by the matrix compression. It is given by

$$R = (r_{(j,k),(j',k')}) = (v_{(j,k),(j',k')}) - (v^{\epsilon}_{(j,k),(j',k')}). \tag{7.12}$$

We remember that the entry $v^{\epsilon}_{(j,k),(j',k')}$ in the compressed matrix was zero if the distance between the supports of the corresponding wavelet basis functions was smaller than a cut-off parameter $B_{j,j'}$. Next we need to show that the entries of the error matrix R are sufficiently small. To do so we use the previous lemma, which showed that the entries set to zero are small.

Let $\tilde{I}_j$ be the index set of indices corresponding to the level j. It contains 2^j elements.

Lemma 7.2.5. *If for the cut-off parameter we have $B_{j,j'} \geq a\max\{2^{-j}, 2^{-j'}\}$ with $a > 1$, the following estimate holds:*

$$\sum_{k\in\tilde{I}_j} |r_{(j,k),(j',k')}| \leq c2^{-\frac{7}{2}(j+j')}2^j B^{-5}_{j,j'}. \tag{7.13}$$

Proof. The sum can be written as

$$\begin{aligned}\sum_{k\in\tilde{I}_j}|r_{(j,k),(j',k')}| &= \sum_{\{k\in\tilde{I}_j\,:\,\mathrm{dist}(\Omega_{(j,k)},\Omega_{(j',k')})>B_{j,j'}\}}|v_{(j,k),(j',k')}| \\ &\overset{\text{Lemma 7.2.3}}{\leq} c2^{-\frac{7}{2}(j+j')}\sum_{\{k\in\tilde{I}_j\,:\,\mathrm{dist}(\Omega_{(j,k)},\Omega_{(j',k')})>B_{j,j'}\}}\mathrm{dist}(\Omega_{(j,k)},\Omega_{(j',k')})^{-6}\end{aligned}$$

The index set $\{k\in\tilde{I}_j\ :\ \mathrm{dist}(\Omega_{(j,k)},\Omega_{(j',k')})>B_{j,j'}\}$ contains at most $N_j=2^j$ elements. It follows that

$$\begin{aligned}\sum_{\{k\in\tilde{I}_j\,:\,\mathrm{dist}(\Omega_{(j,k)},\Omega_{(j',k')})>B_{j,j'}\}}\left[\mathrm{dist}(\Omega_{(j,k)},\Omega_{(j',k')})\right]^{-6} &\leq c2^j\int_{|x|\geq B_{j,j'}}|x|^{-6}dx \\ &\leq c2^jB_{j,j'}^{-5}.\end{aligned}$$

In total this gives the assertion. □

Lemma 7.2.6. *If the cut-off parameter $B_{j,j'}$ is sufficiently large, or more precisely, if*

$$B_{j,j'}\geq a\max\left\{2^{-j},2^{-j'},2^{\frac{J(2\delta+1)-j(4+\delta)-j'(4+\delta)}{5}}\right\}$$

with $a,\delta\in\mathbb{R}$, $a>1$ and $2<\delta<3$, we get the following bound on the entries of the error matrix r.

$$\sum_{k\in\tilde{I}_j}2^{-j/2}2^{-(j+j')}|r_{(j,k),(j',k')}|\leq c2^{-j'/2}a^{-7}2^{j(\delta-1)}2^{j'(\delta-1)}2^{-J(2\delta+1)}.$$

Proof. Applying Lemma 7.2.5 gives

$$\begin{aligned}\sum_{k\in\tilde{I}_j}2^{-j/2}2^{-(j+j')}|r_{(j,k),(j',k')}| &\overset{\text{Lemma 7.2.5}}{\leq} c2^{-j/2}2^{-(j+j')}2^{-\frac{7}{2}(j+j')}2^jB_{j,j'}^{-5} \\ &\leq c2^{-4(j+j')}2^{-j'/2}B_{j,j'}^{-5}.\end{aligned}$$

We assume without loss of generality that $j\geq j'$. If this is not the case, the roles can be reversed. For ease of notation we define

$$\eta:=\max\left\{2^{-j},2^{-j'},2^{\frac{J(2\delta+1)-j(4+\delta)-j'(4+\delta)}{5}}\right\}$$

We now look at the different values the maximum can attain separately. Under the assumption $j\geq j'$, there are only two cases.

Case 1: If $\eta = 2^{\frac{J(2\delta+1)-j(4+\delta)-j'(4+\delta)}{5}}$, we know that $B_{j,j'} \geq a2^{\frac{J(2\delta+1)-j(4+\delta)-j'(4+\delta)}{5}}$. Then, using the above we get

$$\sum_{k\in\tilde{I}_j} 2^{-j/2}2^{-(j+j')}|r_{(j,k),(j',k')}| \leq c2^{-j'/2}a^{-5}2^{j(\delta-1)}2^{j'(\delta-1)}2^{-J(2\delta+1)},$$

which gives the assertion.

Case 2: If $\eta = 2^{-j'}$, we know that $B_{j,j'} \geq a2^{-j'}$. This gives us

$$\begin{aligned}\sum_{k\in\tilde{I}_j} 2^{-j/2}2^{-(j+j')}|r_{(j,k),(j',k')}| &\leq 2^{-4(j+j')}2^{-j'/2}B_{j,j'}^{-5} \\ &\leq ca^{-5}2^{-4j}2^{j'/2}\end{aligned}$$

It remains to show that

$$2^{-4j}2^{j'/2} \leq 2^{j(\delta-1)}2^{j'(\delta-1)}2^{-J(2\delta+1)},$$

when $\frac{J(2\delta+1)-j(3+\delta)-j'(3+\delta)}{5} \geq -j'$ and $1 < \delta < 2$.
Since

$$\begin{aligned}&J(2\delta+1) - j(3+\delta) - j'(3+\delta) \geq 5j' \\ \Leftrightarrow\ &-J(2\delta+1) + j(\delta-1) + j'(\delta-1) \leq j' - 4j\end{aligned}$$

we obtain the assertion.

□

Definition 7.2.7. *We define the spectral norm of a matrix as follows:*

$$\|A\| := \max_{\|x\|_2=1} \|Ax\|_2. \tag{7.14}$$

Lemma 7.2.8 (Schur's Lemma, Lemma 6.2.3 [45])**.** *Let $(A_{ij})_{ij\in I}$ be a matrix and let I be a countable index set. Then, for every vector $u = (u_i)_{i\in I}$ and for an arbitrary $s \in \mathbb{R}$ we have*

$$\|Au\| \leq c\left(\sup_{i\in I}\sum_{j\in I} 2^{s(i-j)}|A_{ij}|\right)^{\frac{1}{2}}\left(\sup_{j\in I}\sum_{i\in I} 2^{s(j-i)}|A_{ij}|\right)^{\frac{1}{2}}\|u\|.$$

Define the matrices $\mathcal{R}_{(j,j')}$ by

$$(\mathcal{R}_{(j,j')})_{k,k'} := 2^{-(j+j')}|r_{(j,k),(j',k')}|. \tag{7.15}$$

Lemma 7.2.9. *The spectral norm of $\mathcal{R}_{(j,j')}$ is bounded by*

$$\|\mathcal{R}_{(j,j')}\| \leq ca^{-5}2^{-J(2\delta+1)}2^{j(\delta-1)}2^{-j'(\delta-1)}. \tag{7.16}$$

Proof. Applying Schur's Lemma with $s = \frac{1}{2}$ gives

$$\|\mathcal{R}_{(j,j')}\| \leq \left(\sup_{k'\in\tilde{I}_{j'}} \sum_{k\in\tilde{I}_j} 2^{-(j+j')}2^{-(j'-j)/2}|r_{(j,k),(j',k')}|\right)^{1/2} \cdot \left(\sup_{k\in\tilde{I}_j} \sum_{k'\in\tilde{I}_{j'}} 2^{-(j+j')}2^{-(j-j')/2}|r_{(j,k),(j',k')}|\right)^{1/2}.$$

Since $(a+b)^2 \geq 0$ we have $a^{\frac{1}{2}}b^{\frac{1}{2}} \leq \frac{a+b}{2}$. Using this estimate we get

$$\|\mathcal{R}_{(j,j')}\| \leq c\Big(\sup_{k\in\tilde{I}_j} \sum_{k'\in\tilde{I}_{j'}} 2^{-(j+j')}2^{-(j-j')/2}|r_{(j,k),(j',k')}| + \sup_{k'\in\tilde{I}_{j'}} \sum_{k\in\tilde{I}_j} 2^{-(j+j')}2^{-(j'-j)/2}|r_{(j,k),(j',k')}|\Big)$$

Finally, applying Lemma 7.2.6 gives the assertion:

$$\|\mathcal{R}_{(j,j')}\| \leq ca^{-5}2^{-J(2\delta+1)}2^{j(\delta-1)}2^{-j'(\delta-1)}.$$

□

It remains to check that v^ϵ is sufficiently sparse to allow the solution of the linear system in linear complexity. The number of non-zero matrix entries is estimated in what follows.

Theorem 7.2.10 (Theorem 8.2.11, [45])**.** *Here the number of degrees of freedom is given by $N_J = 2^J$. The compressed matrix v^ϵ contains*

$$\mathcal{O}\left((\log N_J)^b N_J\right)$$

non-zero entries. The constant $b > 2$ depends on the spatial dimension d and on the number of vanishing moments of the wavelet and dual wavelet.

Proof. Since the proof depends only on the structure of the wavelet basis, the proof from [45] can be applied without change. □

Remark 7.2.11. *The number of non-zero matrix entries still contains a logarithmic term, to attain linear complexity $\mathcal{O}(N_J)$ we need to remove further matrix entries.*

7.3 Second compression step

The first compression step does not sparsify the matrix enough to achieve linear complexity in solving the resulting linear system. In the second step we set to zero some entries for which the supports of the ansatz and test functions overlap as well. Then, we show that the number of nonzero entries in the matrix reduces to $O(N_j)$, where N_j is the number of degrees of freedom, without a loss of accuracy or stability.

We recall, that the scaling functions associated with the piecewise constant wavelet basis are

$$\phi_{jk} = 2^{j/2}\phi(2^j\gamma^{-1}(x) - k),$$

with $\phi = \chi_{[0,1]}$. Further, we recall that the piecewise constant basis functions used for the time discretisation are denoted by $\chi_n(t)$ and that the elements

$$\Omega_{jk} = \text{conv hull}\,\{x \in \Gamma \ : \ \psi_{jk}(\gamma^{-1}(x)) \neq 0\}.$$

Lemma 7.3.1. *There exists a constant $c > 0$, such that the following estimate holds*

$$|\langle V\psi_{jk'}\chi_m, \phi_{jk}\chi_n\rangle| \leq c2^{-4j}\left[\text{dist}(\Omega_{j,k'}, \gamma(\text{supp}\,\phi_{jk}))\right]^{-3},$$

with $j_0 < j < J$.

Proof. To show this result we use a similar technique to that used in Lemma 7.2.3. Let $x \in \Omega_{j,k'}$ and let $y \in \text{supp}\,\phi_{jk}$. Then, we use a Taylor-series of degree two around the point $x \in \Omega_{j,k'}$ to represent the function $x \to g_{mn}(x-y)$. This gives

$$g_{mn}(x-y) = \sum_{\alpha<3} c_\alpha(y, x_0)(x-x_0)^\alpha + R(x, x_0, y).$$

When we insert this representation into the integrals, giving

$$|\langle V\psi_{jk'}\chi_m, \phi_{jk}\chi_n\rangle| = \left|\int_{\Omega_{j,k'}}\int_{\text{supp}\,\phi_{jk}} g_{mn}(x-y)\psi_{jk'}(x)\phi_{jk}(y)dydx\right|,$$

the terms $(x-x_0)^\alpha$, for $\alpha < 3$ vanish due to the three vanishing moments of $\psi_{jk'}$. For ease of notation let $\tilde{x}_m = x_0 + m(x-x_0)$. This leaves integrals over the remainder

R:

$$|\langle V\psi_{jk'}\chi_m, \phi_{jk}\chi_n\rangle| = \left| \int_{\Omega_{j,k'}} \int_{\operatorname{supp}\phi_{jk}} g_{mn}(x-y)\psi_{jk'}(x)\phi_{jk}(y)dydx \right|.$$

This can be estimated as follows

$$|\langle V\psi_{jk'}\chi_m, \phi_{jk}\chi_n\rangle| \leq c \int_{\Omega_{j,k'}} \int_{\operatorname{supp}\phi_{jk}} |x-x_0|^3 H(x,y)|\psi_{jk'}(x)\phi_{jk}(y)|dydx,$$

where

$$H(x,y) = \int_0^1 (1-m)^2 \left|\partial_x^3 g_{mn}(\tilde{x}_m - y)\right| dm,$$

which can be estimated as follows

$$H(x,y) \leq c \sup_{\substack{x\in\gamma^{-1}(\Omega_{j,k'}) \\ y\in\operatorname{supp}(\phi_{jk})}} |\partial_x^3 g_{mn}(x-y)|.$$

Since we have $\operatorname{dist}(\Omega_{j,k'}, \gamma(\operatorname{supp}(\phi_{jk}))) > 0$, we can apply Lemma 7.2.1 to the time-integrated kernel:

$$\begin{aligned} \sup_{\substack{x\in\gamma^{-1}(\Omega_{j,k'}\cap\Gamma) \\ y\in\operatorname{supp}(\phi_{jk})}} |\partial_x^3 g_{mn}(x-y)| &\leq c|x-y|^{-3} \\ &\leq c \operatorname{dist}(\Omega_{j,k'}, \operatorname{supp}\phi_{jk})^{-3}. \end{aligned}$$

Analogously to the estimates for integrals over the wavelet basis functions in Lemma 7.2.3 we estimate as follows.

$$\begin{aligned} |\langle V\psi_{jk'}\chi_m, \phi_{jk}\chi_n\rangle| &\leq c2^{-3j}2^{-j/2}2^{-j/2} \operatorname{dist}(\Omega_{j,k'}, \operatorname{supp}\phi_{jk})^{-3} \\ &\leq c2^{-4j} \operatorname{dist}(\Omega_{j,k'}, \operatorname{supp}\phi_{jk})^{-3}. \end{aligned}$$

□

Corollary 7.3.2. *There holds*

$$\left| \int_{\Omega_{j,k'}} g_{mn}(x-y)\psi_{jk'}(y)dy \right| \leq c2^{-\frac{7}{2}j} \operatorname{dist}(x, \Omega_{j,k'})^{-3},$$

for all x in Γ.

Proof. Using the vanishing moments of $\psi_{jk'}$ and Lemma 7.2.1, this lemma can be shown analogously to Lemma 7.3.1. □

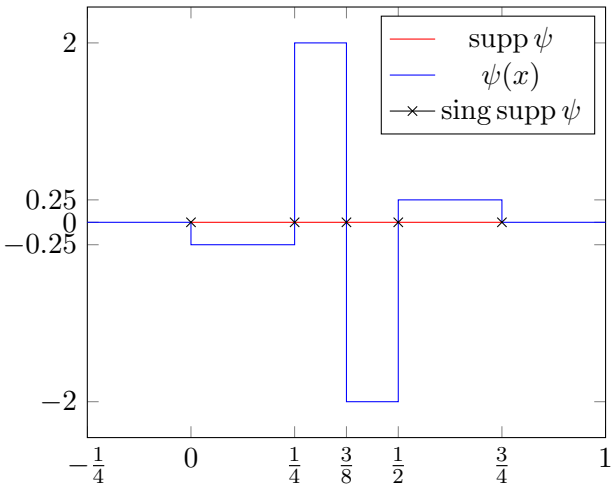

Figure 7.1: A wavelet ψ and its support and singular support.

Next, we define the so-called singular support of a function.

Definition 7.3.3. *The singular support of ψ_{jk} is defined as the set of points at which ψ_{jk} is not a smooth function. We denote the singular support of ψ_{jk} by*

$$\Omega_{j,k}^S := \operatorname{sing\,supp} \psi_{jk}.$$

Remark 7.3.4. *The singular support of the mother wavelet ψ is shown in Figure 7.1. We see, that the singular support of the wavelet basis functions consists only of 5 distinct points.*

In the following proofs we always assume that $j \geq j'$.

Lemma 7.3.5. *We can estimate the absolute value of the matrix entries as follows*

$$\left|v_{(j,k),(j',k')}\right| \leq c 2^{-3j} 2^{-|j-j'|/2} \operatorname{dist}(\Omega_{j,k}, \Omega_{j',k'}^S)^{-2},$$

with $j_0 \leq j \leq J$.

Proof. The matrix entries can be written as

$$|v_{(j,k),(j',k')}| = \left| \int_{\Omega_{j,k}} \int_{\Omega_{j',k'}} g_{mn}(x-y) \psi_{jk}(x) \psi_{j'k'}(y) dy dx \right|$$

Now we can apply Corollary 7.3.2 to estimate

$$|v_{(j,k),(j',k')}| \leq \int_{\Omega_{j,k}} \underbrace{\left| \int_{\Omega_{j',k'}} g_{mn}(x-y)\psi_{j'k'}(y)dy \right|}_{\leq c2^{-\frac{7}{2}j'} \operatorname{dist}(x,\Omega_{j',k'})^{-3}} |\psi_{jk}(x)|dx$$

Further, using the estimate $|\psi_{jk}(x)| \leq 2^{j/2}$ we obtain

$$|v_{(j,k),(j',k')}| \leq c2^{-\frac{7}{2}j'}2^{j/2} \int_{\Omega_{j,k}} \operatorname{dist}(x,\Omega_{j',k'})^{-3}dx.$$

Let $x_0 \in \operatorname{sing\,supp} \psi_{j'k'}$, and let $x \in \Omega_{j,k}$. Then we can estimate,

$$\begin{aligned} \int_{\Omega_{j,k}} \operatorname{dist}(x,\Omega_{j',k'})^{-3}dx &\leq \int_{\Omega_{j,k}} \|x-x_0\|^{-3}dx \\ &\leq c \operatorname{dist}(\Omega_{j,k},\Omega^S_{j',k'})^{-2}, \end{aligned}$$

which gives the assertion. □

Having defined the singular support it is now possible to give the form of the second compression matrix $(v^\epsilon)'$ as follows

$$(v^\epsilon)'_{(jk),(j'k')} = \begin{cases} v^\epsilon_{(jk),(j'k')}, & \text{if } j' \leq j \text{ and } \operatorname{dist}(\Omega_{(j,k)},\Omega^S_{(j',k')}) \leq B^S_{j,j'}, \\ v^\epsilon_{(jk),(j'k')}, & \text{if } j' > j \text{ and } \operatorname{dist}(\Omega^S_{(j,k)},\Omega_{(j',k')}) \leq B^S_{j,j'}, \\ 0, & \text{else.} \end{cases}$$

where the second cut-off parameter

$$B^S_{j,j'} = a' \max\left\{2^{-j}, 2^{-j'}, 2^{\frac{J(2\delta'+1)-3\max\{j,j'\}-(j+j')(\delta'+1)}{2}}\right\}, \tag{7.17}$$

with $a', \delta' \in \mathbb{R}$ and $2 < \delta' < 3$, $a' > 1$.

Remark 7.3.6. *Lemma 7.3.5 tells us that*

$$\left|v_{(j,k),(j',k')}\right| \leq c2^{-3j}2^{-|j-j'|/2} \operatorname{dist}(\Omega_{j,k},\Omega^S_{j',k'})^{-2}.$$

Consequently,

$$\left|v_{(j,k),(j',k')} - (v^\epsilon_{(j,k),(j',k')})'\right| \leq c2^{-3j}2^{-|j-j'|/2}(B^S_{j,j'})^{-2}.$$

Let the error matrix corresponding to the second compression be given as

$$r'_{(j,k),(j',k')} = v_{(jk),(j'k')} - (v^{\epsilon})'_{(jk),(j'k')}.$$

Further, let the block matrix corresponding to the levels j, j' be denoted by

$$(\mathcal{R}'_{(j,j')})_{k,k'} = 2^{-(j+j')} r'_{(j,k),(j',k')}.$$

It remains to show, that the second compression does not reduce the convergence order.

Lemma 7.3.7. *There holds*

$$\|R'_{(jj')}\| \leq c \cdot (a')^{-2} 2^{-J(2\delta'+1)} 2^{(j+j')\delta'} 2^{j-j'},$$

with a constant $c > 0$ independent of a'.

Proof. This proof follows the same lines as the proof of Lemma 7.2.9. To show the result we apply Schur's Lemma and use $a^{1/2}b^{1/2} \leq \frac{a+b}{2}$, giving

$$\begin{aligned}\|\mathcal{R}'_{(jj')}\| \leq c\Big(&\sup_{k' \in I_{j'}} \sum_{k \in I_j} 2^{-(j'-j)/2} 2^{-(j+j')} |r'_{(jk),(j'k')}| \\ &+ \sup_{k \in I_j} \sum_{k' \in I_{j'}} 2^{-(j-j')/2} 2^{-(j+j')} |r'_{(jk),(j'k')}| \Big)\end{aligned}$$

Applying Lemma 7.3.5 to this estimate gives

$$|r'_{(jk),(j'k')}| \leq c 2^{-3j} 2^{-(j-j')/2} (B^S_{j,j'})^{-2}.$$

Now we insert the definition of $B^S_{j,j'}$ into the equation. Since we have restricted ourselves without loss of generality to the case $j \geq j'$ there are two cases to account for.

Case 1: If $B^S_{j,j'} = 2^{\frac{J(2\delta'+1) - 3\max\{j,j'\} - (j+j')(\delta'+1)}{2}}$

In this case we can estimate

$$|r'_{(jk),(j'k')}| \leq c(a')^{-2} 2^{-(j-j')/2} 2^{-J(2\delta'+1)} 2^{(j+j')(\delta'+1)}.$$

Case 2: If $B^S_{j,j'} = 2^{-j'}$

In this case we obtain

$$|r'_{(jk),(j'k')}| \leq c(a')^{-2}2^{-3j}2^{-(j-j')/2}2^{2j'}.$$

Since

$$2^{-3j}2^{-(j-j')/2}2^{2j'} \leq 2^{-(j-j')/2}2^{-J(2\delta'+1)}2^{(j+j')(\delta'+1)},$$

when $2j' \geq J(2\delta'+1) - 3j - (j+j')(\delta'+1)$ this can be estimated by the same term as in the first case.

Further, we can remove all summands corresponding to zero entries in r'. We call the index sets with the removed indices $I^j \subset I_j$ and $I^{j'} \subset I_{j'}$, respectively. Thus,

$$\begin{aligned}\|\mathcal{R}'_{(jj')}\| \leq c(a')^{-2}\Big(&\sup_{k'\in I_{j'}}\sum_{k\in I^j} 2^{(j+j')\delta'}2^{-J(2\delta'+1)} \\ &+ \sup_{k\in I^j}\sum_{k'\in I_{j'}} 2^{-(j-j')}2^{(j+j')\delta'}2^{-J(2\delta'+1)}\Big).\end{aligned}$$

Using the definition of the first compression we find that I^j contains at most $\mathcal{O}(2^{j-j'})$ non-zero entries and $I^{j'}$ contains at most $\mathcal{O}(2^{j'-j})$ non-zero entries.

This gives

$$\begin{aligned}\|\mathcal{R}'_{(jj')}\| &\leq c(a')^{-2}\left(2^{(j+j')\delta'}2^{-J(2\delta'+1)}2^{j-j'} + 2^{j'-j}2^{-(j-j')}2^{(j+j')\delta'}2^{-J(2\delta'+1)}\right) \\ &\leq c(a')^{-2}2^{(j+j')\delta'}2^{-J(2\delta'+1)}2^{j-j'}.\end{aligned}$$

as asserted. □

We would like to show that the convergence rates of the original Galerkin scheme are preserved for the compressed scheme. This is easily seen using the following version of Strang's Lemma.

Theorem 7.3.8 (Theorem 6.1, [8])**.** *Let H be a separable Hilbert space with norm $\|\cdot\|$, and let H_n be a sequence of finite-dimensional subspaces of H. Let P_n denote the orthogonal projection onto H_n.*

Further, let A be a bijective, continuous operator on H and A_n an injective sequence of operators on H_n. Then the error between the exact solution y of the problem

$$Ay = f$$

and the approximated solution y_n of

$$A_n y_n = P_n f$$

can be estimated as follows

$$\|y - y_n\| \leq c\|y - P_n y\| + \|P_n A y_n - A_n y_n\|.$$

Remark 7.3.9. *After applying this theorem to the error of the compressed scheme we can estimate the second summand using the estimates derived for $\mathcal{R}'$.*

Finally, we have to show that after the second matrix compression we are left with only $\mathcal{O}(N_J)$ matrix entries. This is covered by the following theorem.

Theorem 7.3.10. *The matrix $(v^\epsilon)'$, defined by the two matrix compressions contains*

$$\mathcal{O}(N_J), \quad N_J = 2^J$$

non-zero entries.

Proof. Since the proof depends only on the structure of the wavelet basis, the proof of Theorem 8.2.10 from [45] can be applied without change. □

7.4 Wavelets in Time

The wavelet basis suggested in [8] uses the B-spline wavelet forms a basis in space, denoted by ψ^X_{jk} and the wavelets on the interval as a basis in time, denoted by ψ^T_{jk}. These wavelets have been described in Section 3.3.2 and 3.4 respectively. We give here the results for the matrix compression using these wavelets.

We denote the matrix of the single layer operator with this basis by w. As before we denote the matrix sub-block corresponding to the levels j and j' by $w_{j,j'}$.

To define the compressed matrix when wavelets are used in time and space we need to define the distance between elements. Let $\lambda = (j,k)$ and $\lambda' = (j',k')$, then

$$\operatorname{dist}_{(\lambda,\lambda')} = \operatorname{dist}\{\operatorname{supp}\, \psi^X_{j,k}, \operatorname{supp}\, \psi^X_{j',k'}\}^2 + \operatorname{dist}\{\operatorname{supp}\, \psi^T_{j,k}, \operatorname{supp}\, \psi^T_{j',k'}\}.$$

Then for j, j' the compressed blocks of the matrix are

$$w^\epsilon_{j,j'} = \left(w^\epsilon_{\lambda=(j,k),\lambda'=(j',k')}\right)_{j,k,j',k'},$$

where

$$w^{\epsilon}_{\lambda,\lambda'} = \begin{cases} 0, & \text{if dist } _{\lambda,\lambda'} \geq \delta_{j,j'} \\ w_{\lambda,\lambda'}, & \text{else.} \end{cases} \tag{7.18}$$

Theorem 7.4.1 (Propostion 5.5, [8]). *Let the compression parameter $\delta_{j,j'} > 0$, then*

$$\begin{aligned} \#\ \textit{non-zero entries}\ w^{\epsilon}_{j,j'} \leq c2^{3(j+j')} \min\{&2^{-3j} + 2^{-3j'} + 2^{-j-2j'} + 2^{-j'-2j} \\ &+ \sqrt{\delta_{j,j'}}(2^{-2j} + 2^{-2j'}) + \delta_{j,j'}(2^{-j}, 2^{-j'}), 1\} \end{aligned}$$

Theorem 7.4.2 (Proposition 5.6, [8]). *Let the compression parameter $\delta_{j,j'} > 0$, then*

$$\|w_{j,j'} - w^{\epsilon}_{j,j'}\| \leq c2^{-bj}2^{-(b-3)j'}\delta^{-b}_{j,j'} \max\{\delta^{3/2}_{j,j'}, 2^{-3j}, 2^{-3j'}\},$$

with $b = \tilde{m}^X + 2\tilde{m}^T + \frac{3}{2}$, where $\tilde{m}^X$ and $\tilde{m}^T$ are the number of vanishing moments of the dual system in space and time respectively.

7.5 Implementation

In this section we discuss some of the issues related to the implementation of wavelet bases and in particular of the matrix compressions given for piecewise constant wavelets.

Firstly we ensure that we do not reevaluate the same integrals several times. Then we discuss a method for calculating the distances between the elements in space as this is needed for the matrix compression.

7.5.1 Reevaluating Integrals

To compute the matrix of the single layer heat potential one has to compute the coefficients

$$\langle V\psi_{jk}\chi_m, \psi_{j'k'}\chi_n\rangle. \tag{7.19}$$

We define coefficients of the matrix corresponding to the single scale basis ϕ_{jk} are given by

$$\alpha_{(j,k,m),(j',k',n)} = \langle V\phi_{jk}\chi_m, \phi_{j'k'}\chi_n\rangle.$$

Our goal is to write the wavelet basis functions ψ_{jk} as linear combinations of the single scale functions ϕ_{jk} . We use this representation to write the matrix entries unsing only the coefficients $\alpha_{(j,k,m),(j',k',m')}$ as follows:

$$\langle V\psi_{jk}\chi_m, \psi_{j'k'}\chi_n\rangle = \sum_l \sum_{l'} b_l b_{l'} \alpha_{(j+1,2k+l,m),(j'+1,2k'+l',n)}.$$

with coefficients b_l as given in refinement relation (3.6) in [34].

The formulas require some of the values for $\alpha_{(j,k,m),(j',k',n)}$ to be calculated several times.

To avoid reevaluation of the expensive integrals $\alpha_{(j,k,m),(j',k',n)}$ we use a technique called memoization. A memoize function speeds up a computation by storing the results of a function call and returning the result when the input occurs again. In Figure 7.2 we give an implementation of the memoize function, it can be applied to speed up any function that gets called multiple times with the same input.

```
def memoize(f):
    # The memoized version of f looks up its
    # function arguments in this dictionary:
    class memodict(dict):
        # Only when the value for this argument
        # was not found, the following function
        # is called:
        def __missing__(self, key):
            # We calculate the value f(key),
            # store it, and return it.
            ret = self[key] = f(key)
            return ret
    return memodict().__getitem__
```

Figure 7.2: A memoize decorator function (in Python).

7.5.2 Calculating distances between elements

To calculate the compressed matrix we need to calculate the distance between supports and singular supports of the wavelet basis functions. For the first compression it is sufficient to calculate the distances between supports. However, for the second compression we also need to evaluate the distance between the singular supports of wavelet functions.

In the case of the circle the distances between the elements can be calculated directly. More precisely, if the mesh is sufficiently refined the distance between the support of ψ_{jk} and $\psi_{j'k'}$ is proportional to the distance between the supports of the basis after projection to the (periodic) unit interval. Distances on the unit interval are easy to calculate.

In the more general case of a smooth closed curve, there is no simple formula that

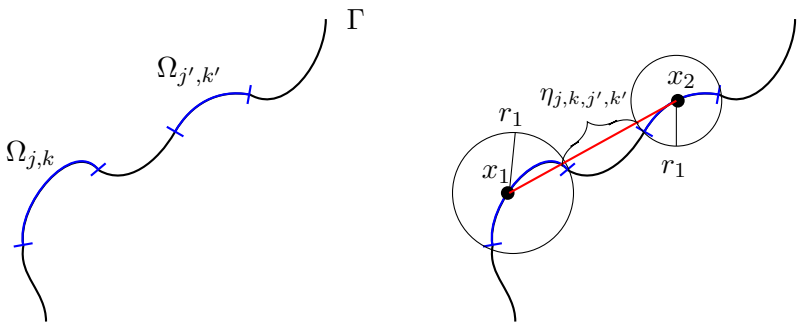

Figure 7.3: Calculating the distance between the supports of two basis functions ψ_{jk} and $\psi_{j'k'}$.

can be efficiently evaluated. We use the following method to estimate the distances instead.

We denote the estimate for the distance between two elements $\Omega_{j,k} = \text{supp } \psi_{jk}$ and $\Omega_{j',k'} = \text{supp } \psi_{j'k'}$ by $\eta_{j,k,j',k'}$. The calculation of the distance is shown in Figure 7.3.

We find the smallest circles $B_{r_1}(x_1)$ and $B_{r_2}(x_2)$ such that $x_1, x_2 \in \Gamma$ and such that they contain the elements Ω_{jk} and $\Omega_{j'k'}$ respectively. Then we take the distance between the circles $\eta_{j,k,j',k'}$. This method will underestimate the distance between the two elements. Particularly, when there are few elements in the spatial mesh, this means that the matrix will not be sparsified as strongly as it should be.

7.6 Numerical Experiments

In this section we give some numerical experiments using wavelet basis functions in space and piecewise constant basis functions in time. First we discuss the structure of the matrix of the compressed and uncompressed wavelet schemes. Then we give numerical results on the speed-up attained by using a matrix compression and verify that the compressed scheme does not lead to a loss of accuracy. Finally, we explore the impact the choice of the cut-off parameters a, δ and a', δ'.

7.6.1 Structure of the Matrix

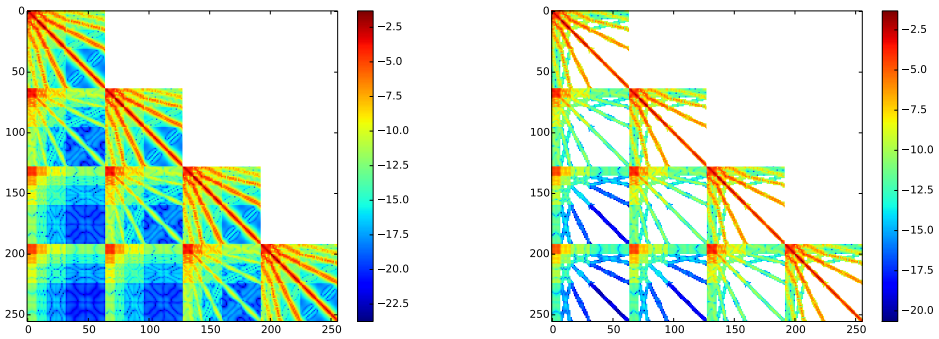

Figure 7.4: The natural logarithm of the matrix coefficients of the single layer heat operator with four time blocks (left), and the non-zero matrix entries after the matrix compression (right).

The structure of the matrix is shown in in Figure 7.4. The left plot shows the natural logarithm of the matrix entries. We can see that essentially the matrix blocks corresponding to one time step have a finger structure, and all entries not in the finger structure are small. The right plot shows the structure of the matrix after the small entries have been set to zero.

In Figure 7.5 we plot the time-integrated heat kernel for $z = x - y$. When we have identical time intervals, i.e. $l = 0$ the heat kernel becomes more peaked and approaches the δ function. For time intervals which are further apart, i.e. $l > 0$, the heat kernel is smaller.

Due to this behavior the block matrices in Figure 7.4 corresponding to larger l values have smaller matrix entries than the block matrices on the diagonal corresponding to identical time intervals.

This type of matrix structure implies that different compression rates for different time steps may be effective. In the following, the same compression is used for all time steps.

7.6.2 Speed Comparisons

In this section we compare the time needed to set up and solve the linear system, for the compressed and uncompressed case.

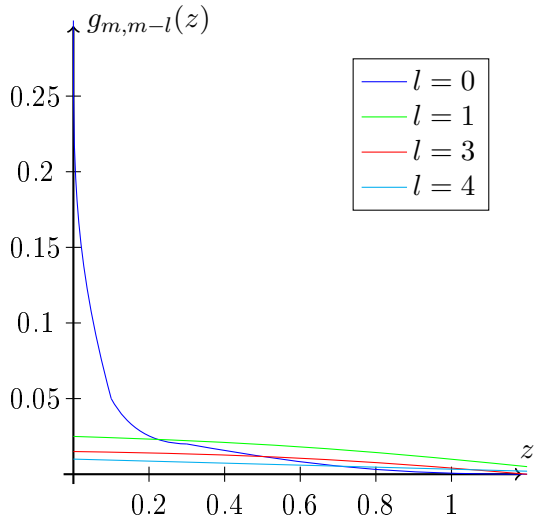

Figure 7.5: Plot of the analytically evaluated time integrals $g_{m,m-l}$ for $z \in [0,1]$ for different values of l.

Time taken to solve the linear system		
j_{max}	compressed matrix	no compression
3	0.00037	0.00039
4	0.00131	0.00135
5	0.00643	0.00665
6	0.05680	0.05698
7	0.72420	0.74965
8	10.5130	11.3976

Table 7.1: The time taken in seconds to solve the linear system for the compressed and uncompressed matrix.

In Table 7.1 we show the time taken in seconds to solve the linear system with matrix compression and without. The compressed system can be solved faster than the uncompressed system in all cases. However, the time needed to solve the uncompressed system is also low. This is probably due to the efficient solver. When the number of degrees of freedom is increased we expect the compressed system to be considerably faster to solve.

Next we look at the time it takes to assemble the matrix, this is shown in Table 7.2. Here we can compare the time taken to assemble the matrix with and without compression, and also the time it takes to assemble the compressed matrix with the time saving measures discussed in Section 7.5.1. We see that the memoization is necessary to ensure the speed-up.

We see that the matrix compression gives a large improvement to the time taken to assemble. The memoize function yields a further improvement in time taken. In total we can quickly assemble much larger systems when using the matrix compression.

j_{max}	compressed matrix with memoize	compressed matrix without memoize	no compression with memoize
3	0.37	0.7460	0.3713
4	2.99	4.05	6.6358
5	13.70	15.84	28.3224
6	37.32	55.49	76.0679
7	82.77	187.28	209.0140
8	206.9	886.7	-
9	417.4	3617.9	-

Table 7.2: The time taken in seconds to assemble the matrix for the compressed and uncompressed matrix.

7.6.3 Complexity and Accuracy

In this section we verify the complexity and accuracy results from the previous sections.

In Table 7.3 we compare the number of non-zero matrix entries. For ease of comparison we plot this data in Figure 7.6. We expect that the after the matrix compression the number of non-zero matrix entries decreases from $\mathcal{O}(n^2)$ to $\mathcal{O}(n)$. We see that the numerical experiments verify this.

Number of non-zero matrix entries			
j_{max}	N	no compression	compressed matrix
3	8	64	24
4	16	576	304
5	32	3136	1264
6	64	14400	3984
7	128	61504	11200

Table 7.3: The number of non-zero matrix entries for the compressed and uncompressed wavelet basis.

Lastly, we verify that the matrix compression does not lead to a loss of accuracy. We use the same problem (5.6) as was used in Chapter 5. As a domain we use a circle and as a right hand side we use $g(\varphi, t) = \cos(\varphi)t^2$. We use the scaling $\sigma = 2$, i.e. $h_t \sim h_x^2$.

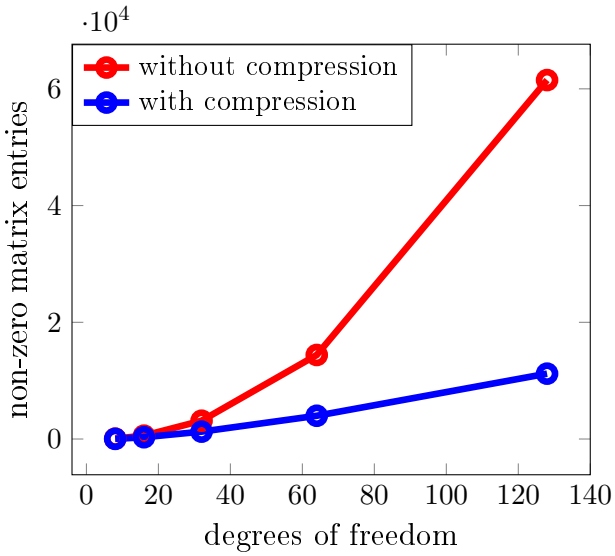

Figure 7.6: The number of non-zero matrix entries for the compressed and uncompressed matrix.

Figure 7.7 shows the convergence in the energy norm. As expected, the convergence rates are exactly those of Chapter 5. The piecewise constant wavelet basis spans the same discrete space as the piecewise constant polynomial basis functions used in that chapter.

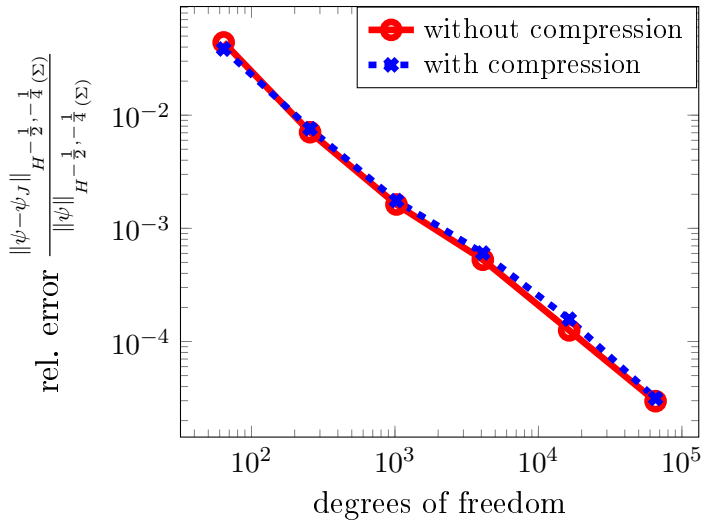

Figure 7.7: Plot of the convergence with the right hand side $g(\varphi, t) = \cos(\varphi)t^2$. Constant basis functions are used in time and piecewise constant wavelets are used in space.

7.6.4 Sensitivity to Compression Parameters

In this section we discuss how the constants a, δ and a', δ' in the definition of the cut-off parameters $B_{j,j'}$ and $B^S_{j,j'}$ given in equations (7.3) and (7.17) affect the accuracy of the scheme and the number of non-zero matrix entries. A similar comparison was shown in [34] for the Laplace equation.

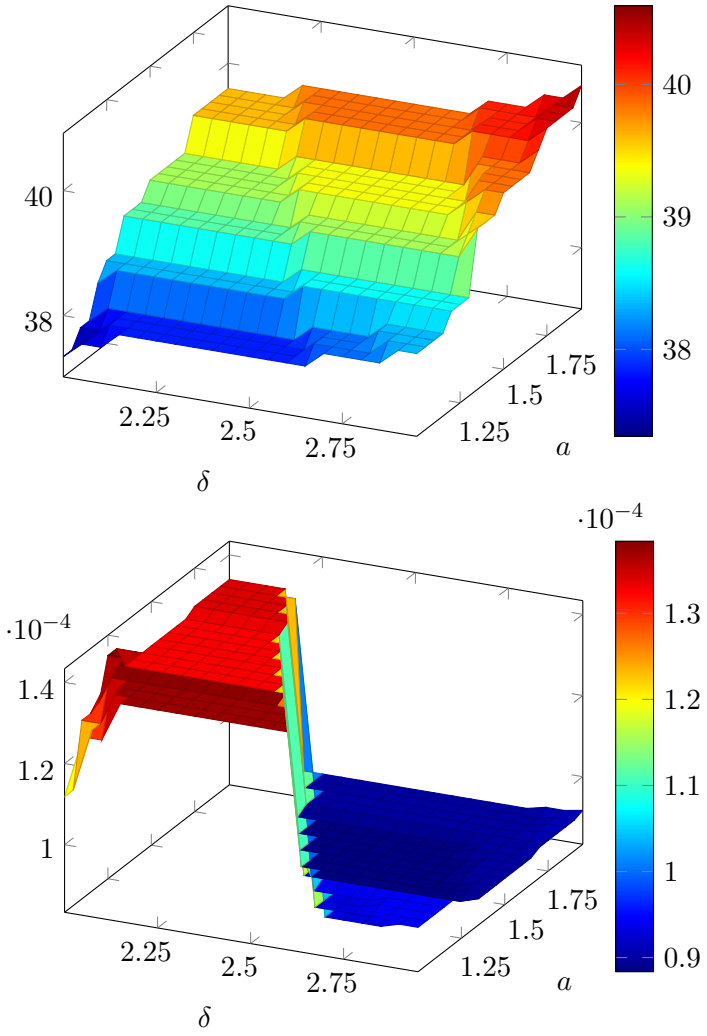

Figure 7.8: The effect of varying the parameters a, a' and δ, δ' of the compression on the proportion of non-zero matrix entries (in percentage) to the total number of matrix entries (top) and on the error of the energy norm (bottom).

In the bottom of Figure 7.8 we plot the error of the energy norm

$$\frac{\|\psi\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)} - \|\psi_J\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)}}{\|\psi\|_{H^{-\frac{1}{2},-\frac{1}{4}}(\Sigma)}}$$

and the number of non-zero matrix entries in dependance of the parameters of the two matrix compressions. We used $a = a'$ and $\delta = \delta'$ in the plots.

We see that number of non-zero matrix entries is lowest when we choose the parameters as small as possible, and the error is smallest when the parameters are chosen as large as possible. These results are similar to those attained in [34] for the elliptic case. In total the effects of varying the compression parameters is small and a choice in the middle of the admissable ranges can be made.

Chapter 8

Conclusions

In this chapter we briefly summarise the main results of this thesis. Then, we offer an outlook on possible directions for future research.

8.1 Summary

We started this thesis with an introduction of wavelets, in particular of biorthogonal wavelets, and an introduction of the boundary reduction of the non-stationary heat equation. The first two chapters reiterated elementary results on both topics, that were used in the subsequent chapters.

In Chapter 4 we discussed the Galerkin discretisation of the boundary integral formulation of the heat equation. This chapter contained a comparison between FEM and BEM. The take-away from this comparison was that BEM are faster in terms of CPU time when individual point evaluations of the solution are needed in the domain, or when the boundary flux itself is required.

In this chapter we also gave analytical formulas for the time integrals, both for the single- and double-layer heat operators. This meant, that when setting up the matrices corresponding to these operators, we were left with integrals in space, over integrands with logarithmic singularities. To evaluate these integrals, we gave efficient quadrature rules for dealing with integrands with logarithmic singularities. Taken together, this gave us an efficient method for numerically evaluating all needed integrals.

Chapter 5 gave an error analysis of the full-tensor product approximation spaces for the boundary reduced heat equation. In particular, we examined the choice of scaling between mesh width, in space h_x, and in time h_t. We found, that when using

piecewise constant polynomial basis functions in time and space, the scaling $h_t \sim h_x^{\frac{6}{5}}$ leads to higher convergence rates in the energy norm. These results are supported by numerical experiments.

In Chapter 6 we introduced sparse grid discretisations. In [12] theoretical results for the convergence rates in the energy norm for a standard sparse grid method were proven. We verified these rates with numerical experiments. Next, we found bounds for the error in the energy norm for an optimised sparse grid space. These results show an improvement over the standard sparse grid spaces in three dimensions. However, in two dimensions it is preferable to use the standard sparse grid index set.

Finally, in Chapter 7 we discussed matrix compression. These can be applied without loss of accuracy when a wavelet basis with a sufficiently high number of vanishing moments is used. We use wavelet basis functions only in time, and show that each matrix block has only $\mathcal{O}(N_x)$ non-zero entries, since we already showed in Chapter 4 that we only need to store $\mathcal{O}(N_t)$ matrix blocks. We compare this with the results of [8], in which wavelet basis functions are used in time and space. Both methods leave in total $\mathcal{O}(N_x N_t)$ non-zero matrix entries. However, our method is easier to implement and allows for piecewise constant wavelet bases in space.

In total, we have achieved both of our main goals. We have reduced the complexity to

$$\mathcal{O}(h_x^{-(d-1)})$$

using boundary reduction. We have used wavelet matrix compressions to reduce the matrix to a sparse matrix, and to solve the linear system in linear complexity. Further, we have also shown methods for increasing the convergence rates in the energy norm, i.e. sparse grid discretisations and a different scaling for full tensor product discretisations.

8.2 Future Work

There are some possible extensions to the theory for the optimised sparse grids. Currently, they do not out-perform standard sparse grids even though they should be more flexible. This is due to the scaling of the optimised sparse grids being $\sigma = 2$. Allowing more flexibility in the scaling between time and space may lead to higher

convergence rates. Changing the index set to

$$J_L^{T,\sigma} = \left\{ (l_x, l_t) \ : \ l_x + \frac{l_t}{\sigma} - T \max\{l_x, l_t/2\} \leq (1-T)L \right\},$$

should lead to an improvement over the standard sparse grid discretisation.

On the implementational side, there are several numerical experiments that could produce interesting results. For example, it would be interesting to allow higher order polynomials as basis functions, and to allow higher spatial dimensions, in order to verify the theoretical results. Further, one might allow more general domains, such as piecewise smooth domains, i.e. polygons.

Using the improvements to CPU speed, gained from the boundary element implementation, it may be possible to solve high dimensional versions of the problem to allow uncertainity in the domain or data. Another possibility would be to modify the method to allow some forms of non-linearity.

Bibliography

[1] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, 1972.

[2] D.N. Arnold and W.L. Wendland. On the asymptotic convergence of collocation methods. *Math. Comp.*, 41:349–381, 1983.

[3] K. Babenko. Approximation by trigonometric polynomials in a certain class of periodic functions of several variables. *Soviet Math. Dokl.*, 1:672–675, 1960.

[4] L. Banjai and S. Sauter. Rapid solution of the wave equation in unbounded domains. *SIAM Journal on Numerical Analysis*, 47(1):227–249, 2009.

[5] M. Bebendorf. Approximation of boundary element matrices. *Numerische Mathematik*, 2000.

[6] M. Bebendorf. Adaptive cross approximation of multivariate functions. *Constructive Approximation*, 34, 2011.

[7] M. Bebendorf and S. Hardesty. Adaptive cross approximation of tensors arising in the discretization of boundary integral operator shape derivatives. *Engineering Analysis with Boundary Elements*, 37(1):60–67, 2013.

[8] C. Bourgeois and R. Schneider. Biorthogonal wavelets for the direct integral formulation of the heat equation. 2000.

[9] H.J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:1–123, 2004.

[10] A. Chernov and A. Reinarz. Numerical quadrature for high-dimensional singular integrals over parallelotopes. *Comp, & Math. with Applications*, 66(7):1213–1231, 2013.

[11] A. Chernov and C. Schwab. First order k-th moment Finite Element analysis of nonlinear operator equations with stochastic data. *Mathematics of Computation*, 82:1859–1888, 2013.

[12] A. Chernov and C. Schwab. Sparse space-time Galerkin BEM for the nonstationary heat equation. *ZAMM Z. Angew. Math. Mech.*, 93:403–413, 2013.

[13] A. Cohen. *Numerical Analysis of Wavelet Methods.* Elsevier, 2003.

[14] A. Cohen, I. Daubechies, and J.C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Pure Appl. Math.*, (5):485–560, 1992.

[15] M. Costabel. Boundary integral operators for the heat equation. *Integral Equations Operator Theory*, 13(4):498–552, 1990.

[16] M. Costabel and E. Stephan. On the convergence of collocation methods for boundary integral equations on polygons. *Math. Comp.*, 49:461–478, 1987.

[17] W. Dahmen, A. Kunoth, and K. Urban. Biorthogonal spline wavelets on the interval–stability and moment conditions. *Applied and Computational Harmonic Analysis*, 6:132–196, 1999.

[18] W. Dahmen and R. Schneider. Wavelets with complementary boundary conditions–function spaces on the cube. *Results Math.*, 34(3-4):255–293, 1998.

[19] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, 1992.

[20] M.G. Duffy. Quadrature over a pyramid or cube of integrands with a singularity at a vertex. *SIAM J. Numer. Anal.*, 19(6):1260–1262, 1982.

[21] G. M. Constantine and T. H. Savits. A multivariate Faa di Bruno formula with applications. *Transactions of the American Mathematical Society*, 348(2):503–520, February 1996.

[22] J. Garcke and M. Griebel. On the computation of the eigenproblems of hydrogen and helium in strong magnetic and electric fields with the sparse grid combination technique. *Journal of Computational Physics*, 165(2):694 – 716, 2000.

[23] W. Gautschi. Orthogonal polynomials (in Matlab). *Journal of Computational and Applied Mathematics*, 178:215–234, 2005.

[24] W. Gautschi. Numerical integration over the square in the presence of algebraic/logarithmic singularities with an application to aerodynamics. *Numerical Algorithms*, 61(2):275–290, 2012.

[25] T. Gerstner and M. Griebel. Numerical integration using sparse grids. *Numer. Algorithms*, 18:209–232, 1998.

[26] M. Griebel. The combination technique for the sparse grid solution of PDEs on multiprocessor machines. *Parallel Processing Letters*, 2(1):61–70, 1992.

[27] M. Griebel. Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences. *Computing*, 61(2):151–179, 1998.

[28] M. Griebel and H. Harbrecht. On the construction of sparse tensor product spaces. *Mathematics of Computations*, 82(282):975–994, April 2013. Also available as INS Preprint No. 1104, 2011.

[29] M. Griebel and H. Harbrecht. On the convergence of the combination technique. In *Sparse grids and Applications*, volume 97 of *Lecture Notes in Computational Science and Engineering*, pages 55–74. Springer, 2014. Also available as INS Preprint No. 1304.

[30] M. Griebel, W. Huber, T. Störtkuhl, and C. Zenger. On the parallel solution of 3D PDEs on a network of workstations and on vector computers. In *Lecture Notes in Computer Science 732, Parallel Computer Architectures: Theory, Hardware, Software, Applications*, pages 276–291. Springer Verlag, 1993.

[31] M. Griebel and S. Knapek. Optimized general sparse grid approximation spaces for operator equations. *Math. Comp.*, 78(268):2223–2257, 2009.

[32] M. Griebel, D. Oeltz, and P. Vassilevski. Space-time approximation with sparse grids. *SIAM Journal on Scientific Computing*, 2005.

[33] M. Griebel, M. Schneider, and C. Zenger. A combination technique for the solution of sparse grid problems. In P. de Groen and R. Beauwens, editors, *Iterative Methods in Linear Algebra*, pages 263–281. IMACS, Elsevier, North Holland, 1992.

[34] H. Harbrecht and R. Schneider. Wavelet Galerkin schemes for 2D-BEM. In *Problems and Methods in Mathematical Physics*, volume 121 of *Operator Theory: Advances and Applications*, pages 221–260. Birkhäuser Basel, 2001.

[35] H. Harbrecht and R. Schneider. Wavelet galerkin schemes for boundary integral equations—implementation and quadrature. *SIAM Journal on Scientific Computing*, 27(4):1347–1370, 2006.

[36] G. Hsiao and W.L. Wendland. *Boundary integral equations*, volume 164 of *Applied Mathematical Sciences*. Springer-Verlag, Berlin, 2008.

[37] C. Lubich. Convolution quadrature and discretized operational calculus. I. *Numer. Math.*, 52(2):129–145, 1988.

[38] C. Lubich. Convolution quadrature and discretized operational calculus. II. *Numer. Math.*, 52(4):413–425, 1988.

[39] S. Mallat. Multiresolution approximations and wavelet orthonormal bases of L2(R). *Transactions of the American Mathematical Society*, 315(1):69–87, September 1989.

[40] M. Messner, M. Schanz, and J. Tausch. Fast Galerkin method for space-time boundary integral equations. *Journal of Computational Physics*, 258(0):15 – 30, 2014.

[41] Y. Meyer. Ondelettes, fonctions splines et analyses graduées. Lectures given at the University of Torino, Italy, 1986.

[42] P.J. Noon. The single layer heat potential and Galerkin boundary element methods for the heat equation. 1988. Phd thesis.

[43] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629–639, Jul 1990.

[44] S. Sauter and C. Schwab. *Boundary Element Methods.* Springer, 2004.

[45] R. Schneider. *Multiskalen- und Wavelet Matrixkompressionen.* Teubner, 1998.

[46] Dominik Schötzau and Christoph Schwab. Time discretization of parabolic problems by the hp-version of the discontinuous Galerkin Finite Element Method. *SIAM Journal on Numerical Analysis*, 38(3):837–875, 2000.

[47] C. Schwab and R. Stevenson. Space-time adaptive wavelet methods for parabolic problems. *Math. Comp*, 78(267):1293–1318, 2009.

[48] C. Schwab and W.L. Wendland. On numerical cubatures of singular surface integrals in boundary element methods. *Numerische Mathematik*, 62(1):343–369, 1992.

[49] O. Steinbach. *Numerical approximation methods for elliptic boundary value problems.* Springer, New York, 2008. Finite and boundary elements, Translated from the 2003 German original.

[50] V. Thomee. *Galerkin finite element methods for parabolic problems.* Springer, 1984.

[51] H. Wayland. *Differential equations applied in science and engineering.* D. Van Nostrand Co., Inc., Princeton, N. J.-Toronto-New York-London, 1957.

[52] P. Wilmott, S. Howison, and J. Dewynne. *The Mathematics of Financial Derivatives.* Cambridge University Press, 1995. Cambridge Books Online.

[53] A. Witkin. Scale-space filtering: A new approach to multi-scale description. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.*, volume 9, pages 150–153, Mar 1984.

[54] C. Zenger. Sparse grids. *Notes Numer. Fluid Mech.*, 31:241–251, 1991.

Appendix A

Solutions to the Heat Equation on the Circle

In this appendix we describe one method of deriving analytic solutions to the heat equation on the circle. We use these solutions to verify our numerical results.

We describe the points of the unit circle $B_1(0)$ by scaled polar coordinates (r, φ) where the angle $\varphi \in [0, 1]$ and the radius $r \in [0, 1]$. For simplicity we assume zero initial conditions, i.e. that $u(\cdot, t = 0) = 0$.

We transform to a Dirichlet problem where the inhomogeneous condition appears as a forcing function.

Let $\tilde{g}(r, \varphi, t)$ denote the harmonic extension of the boundary conditions g to the entire unit disk, i.e.:

$$\begin{aligned} \Delta\tilde{g}(r, \varphi, t) &= 0 \qquad \text{in } \Omega,\ t > 0 \\ \tilde{g}(r, \varphi, t)|_\Gamma &= g(r, \varphi, t) \qquad t > 0 \end{aligned}$$

Then we set $U = u - \tilde{g}$. U satisfies:

$$\begin{aligned} \partial_t U(r, \varphi, t) - \Delta U(r, \varphi, t) &= -\partial_t \tilde{g}(r, \varphi, t) \qquad \text{in } Q \\ U(r, \varphi, 0) &= 0 \qquad \text{in } \Omega \\ U(r, \varphi, t) &= 0 \qquad \text{in } \Sigma \end{aligned}$$

Now we can apply Duhamel's principle, which states that the solution to this problem is

$$U(r, \varphi, t) = \int_0^t v(r, \varphi, t, s)\, ds,$$

where v is the solution to

$$\begin{aligned} \partial_t v - \Delta v &= 0 && \text{in } Q,\ t > s \\ v|_\Gamma &= 0 && t > s \\ v(r,\varphi,s,s) &= -\partial_t \tilde{g}(r,\varphi,s) && \text{in } Q \end{aligned} \tag{A.1}$$

The variable s is viewed as a parameter.

Duhamel's principle can be applied to general problems, however, in the case of the unit disk it is particularly useful since the equation for v can be easily solved using separation of variables.

The solution $v(r,\varphi,t,s)$ is by separation of variables:

$$v(r,\varphi,t,s) = R(r,s)H(\varphi,s)T(t,s).$$

First we solve for t.

$$\partial_t T(t,s) = T(t,s) \underbrace{\frac{R(r,s)H(\varphi,s)}{\Delta R(r,s)H(\varphi,s)}}_{\text{independent of } t}$$

So we have a solution of the form

$$T(t,s) = e^{-\lambda t},$$

where λ can still be chosen freely. Once we insert this, the remaining problem has the form of the Sturm-Liouville problem:

$$\lambda R(r,s)H(\varphi,s) + \Delta R(r,s)H(\varphi,s) = 0.$$

Separating the variables and multiplying by $r^2/R(r,s)H(\varphi,s)$ gives

$$r\partial_r(r\partial_r R(r,s))\frac{1}{R(r,s)} + \lambda r^2 = -\partial_\varphi^2 H(\varphi,s)\frac{1}{H(\varphi,s)} = \mu \neq \mu(r).$$

In order to keep the required boundary conditions we need:

$$H(\pi,s) = H(-\pi,s) \text{ and } \partial_\varphi H(\pi,s) = \partial_\varphi H(-\pi,s).$$

So the problem that needs to be solved is

$$\partial_\varphi^2 H + \mu H = 0.$$

This problem only has non-trivial solutions for $\mu = m^2$ with $m \in \mathbb{N}_0$.

In this case the solutions to the problem are

$$H(\varphi, s) = a(s)e^{im\varphi} \text{ and } H(\varphi, s) = a(s)e^{-im\varphi}$$

and any linear combination of these solutions.

Next we look at the solution for $R(r, s)$. The equation to be solved is

$$\begin{aligned} &r^2\partial_r^2 R(r,s) + r\partial_r R(r,s) + (\lambda r^2 - m^2)R(r,s) = 0 \\ &R(1,s) = 0 \text{ and } |R(r,s)| < \infty. \end{aligned}$$

The boundary conditions for R come from the boundary conditions for v.

We set $p = \sqrt{\lambda}r$ and substitute $R(r, s) = \overline{R}(p, s)$ giving

$$\begin{aligned} &p^2\partial_p^2 \overline{R}(p,s) + p\partial_p \overline{R}(p,s) + (p^2 - m^2)\overline{R}(r,s) = 0 \\ &\overline{R}(\sqrt{\lambda},s) = 0 \text{ and } |\overline{R}(r,s)| < \infty. \end{aligned}$$

The equation above is Bessels equation. It has two linearly independent solutions $J_m(s)$ and $Y_m(s)$, the Fourier-Bessel functions of first and second type respectively.

$J_m(s)$ is bounded at 0, while $Y_m(s)$ is not, so it is clear that we use $J_m(s)$ as solutions.

So we have

$$\overline{R}(p) = c_m(s)J_m(p)$$

and any linear combination of these as solutions.

To satisfy the Dirichlet boundary conditions of the problem $p = \sqrt{\lambda}r$ must be a zero of J_m at $r = 1$. It follows that $\lambda = \alpha_{k,m}^2$, where $\alpha_{k,m}$ is the k-th zero of the m-th Fourier-Bessel function.

In total this gives

$$v(r, \varphi, t, s) = \sum_{m=-\infty}^{\infty} \sum_{k=1}^{\infty} A_{k,m}(s)J_m(\alpha_{k,m}r)e^{-\alpha_{k,m}^2 t}e^{im\varphi}. \tag{A.2}$$

The boundary conditions of the problem for v give us the necessary information to determine the coefficients $A_{k,m}$.

We need

$$\sum_{m=-\infty}^{\infty}\sum_{k=1}^{\infty} A_{k,m}(s)J_m(\alpha_{k,m}r)e^{-\alpha_{k,m}^2 s}e^{im\varphi} = -\partial_s\tilde{g}(r,\varphi,s).$$

This gives for the solution of the original problems

$$\begin{aligned} u &= \tilde{g}(r,\varphi,t) + \int_0^t v(r,\varphi,t,s)ds \\ &= \tilde{g}(r,\varphi,t) + \sum_{m=-\infty}^{\infty}\sum_{k=1}^{\infty}\int_0^t A_{k,m}(s)ds\, J_m(\alpha_{k,m}r)e^{im\varphi}e^{-\alpha_{k,m}^2 t}. \end{aligned} \tag{A.3}$$

Application to $g(t) = t^2$

Since g does not depend on r and φ harmonic extension of g is is g itself.

Since $-\partial_t\tilde{g}$ does not depend on φ the dependence on φ can be dropped. So all coefficients with $m \neq 0$ are zero.

What remains is

$$v(r,\varphi,s,s) = \sum_{k=1}^{\infty} A_{k,0}(s)e^{-\alpha_{k,m}^2 s}J_0(\alpha_{k,0}r) = -s^2.$$

Now it remains to find the coefficients $A_{k,0}$. We know that $A_{k,0}(s)e^{-\alpha_{k,m}^2 s}$ should be the Fourier-Bessel coefficients for the function $-\partial\tilde{g}(s) = -2s$. This gives

$$A_{k,0}(s)e^{-\alpha_{k,m}^2 s} = \frac{-2s}{\alpha_k \frac{1}{2}J_1(\alpha_{k0})}.$$

It follows that u has the form

$$\begin{aligned} u &= \tilde{g}(r,\varphi,t) + \sum_{k=1}^{\infty}\int_0^t A_{k,0}(s)ds\, J_m(\alpha_{k,0}r)e^{-\alpha_{k,0}^2 t} \\ &= t^2 + \sum_{k=1}^{\infty}\int_0^t \frac{-2se^{\alpha_{k,m}^2 s}}{\alpha_k \frac{1}{2}J_1(\alpha_{k0})}ds\, J_m(\alpha_{k,0}r)e^{-\alpha_{k,0}^2 t} \\ &= t^2 + \sum_{k=1}^{\infty}\frac{-4}{\alpha_k J_1(\alpha_{k0})}\int_0^t se^{\alpha_{k,m}^2 s}ds\, J_m(\alpha_{k,0}r)e^{-\alpha_{k,0}^2 t} \\ &= t^2 + 4\sum_{k=1}^{\infty}\frac{J_0(\alpha_{k0}r)}{\alpha_{k0}^3 J_1(\alpha_k)}(t - \frac{1}{\alpha_{k0}^2}(1 - e^{-\alpha_{k0}^2 t}). \end{aligned}$$