Simplex Splines on the Powell–Sabin 12-Split

Components of the finite element method

Ivar Haugaløkken Stangeby

Master’s Thesis, Spring 2018
This master’s thesis is submitted under the master’s programme Mathematics, with programme option Mathematics, at the Department of Mathematics, University of Oslo. The scope of the thesis is 60 credits.

The front page depicts a section of the root system of the exceptional Lie group $E_8$, projected into the plane. Lie groups were invented by the Norwegian mathematician Sophus Lie (1842–1899) to express symmetries in differential equations and today they play a central role in various parts of mathematics.
Abstract

In this thesis we implement and employ a simplex spline basis developed by Cohen, Lyche, and Riesenfeld for the space of $C^1$ quadratic splines on the Powell–Sabin 12-split of a triangle. A matrix recursion is used for evaluation and differentiation, where the need for standard Bernstein–Bézier techniques are avoided. A brief account of the construction of multivariate splines and the theoretical foundation underlying the finite element method is given. Subsequently, using an explicit conversion to the Hermite nodal basis known from the finite element method, we solve the biharmonic equation using conforming methods. Interpolation estimates are derived, and numerical results are seen to comply.
First and foremost, I would like to express my gratitude to my advisors Andrea Bressan, Tom Lyche and Georg Muntingh, for always taking their time to answer my questions. A big thank you to my fellow students, friends and “partners in crime” — The last five years have been a blast! Aksel, Martin and Ola: Thank you for the countless hours of fun! Last but not least, I wish to thank my parents for piquing my interest in the natural sciences and always believing in me!
Contents

Abstract i

Acknowledgements iii

Contents v

List of Figures vii

List of Tables ix

1 Introduction 1
  1.1 Aim and structure of the thesis 1
  1.2 Motivation 3
  1.3 Suggestions for future work 4

2 An Introduction to Simplex Splines 7
  2.1 Definitions 8
  2.2 Knot insertion and recursion 10
  2.3 The Powell–Sabin 12-split 14

3 Sobolev Spaces and the Biharmonic Equation 23
  3.1 Preliminary definitions 24
  3.2 Sobolev spaces 27
  3.3 Gauss–Green formulas 31
  3.4 The biharmonic equation 33
# Contents

3.5 Existence and uniqueness of weak solutions . . . . . . . . 35  

4 The Finite Element Method 39  
4.1 The finite element formulation . . . . . . . . . . . . . . 39  
4.2 Cea’s lemma . . . . . . . . . . . . . . . . . . . . . . . . . 41  
4.3 Constructing a finite element space . . . . . . . . . . . . 42  
4.4 Powell–Sabin 12-splits and a $C^1$ finite element space . . 50  
4.5 Interpolation error for $S^1_2(T_{PS12})$ . . . . . . . . . . . 54  

5 Numerical Results 67  
5.1 Deriving a linear system . . . . . . . . . . . . . . . . . . . 67  
5.2 The finite element assembly process . . . . . . . . . . . . 69  
5.3 Model problems for the biharmonic equation . . . . . . . . 73  

Appendices 81  

A Implementation 83  
A.1 SSplines . . . . . . . . . . . . . . . . . . . . . . . . . . . 83  
A.2 Finite element implementation . . . . . . . . . . . . . . . 86  

Bibliography 89
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Four common triangular $C^1$ finite elements</td>
<td>4</td>
</tr>
<tr>
<td>2.1</td>
<td>The area of a 2-simplex in terms of the length of a 1-simplex</td>
<td>12</td>
</tr>
<tr>
<td>2.2</td>
<td>Powell–Sabin 12-split</td>
<td>14</td>
</tr>
<tr>
<td>2.3</td>
<td>Quadratic spline function</td>
<td>17</td>
</tr>
<tr>
<td>2.4</td>
<td>Quadratic S-spline basis</td>
<td>18</td>
</tr>
<tr>
<td>2.5</td>
<td>Hermite nodal basis</td>
<td>21</td>
</tr>
<tr>
<td>3.1</td>
<td>A bounded domain $\Omega$</td>
<td>25</td>
</tr>
<tr>
<td>4.1</td>
<td>Finite element diagrams</td>
<td>44</td>
</tr>
<tr>
<td>4.2</td>
<td>Triangulations</td>
<td>49</td>
</tr>
<tr>
<td>4.3</td>
<td>Local basis indentification</td>
<td>53</td>
</tr>
<tr>
<td>4.4</td>
<td>Global continuity</td>
<td>54</td>
</tr>
<tr>
<td>4.5</td>
<td>Global basis functions for $S^1_2(\mathcal{T}_{PS12})$</td>
<td>55</td>
</tr>
<tr>
<td>4.6</td>
<td>Alternative degrees of freedom</td>
<td>60</td>
</tr>
<tr>
<td>5.1</td>
<td>Local versus global numbering</td>
<td>72</td>
</tr>
<tr>
<td>5.2</td>
<td>Stiffness matrix sparsity pattern</td>
<td>73</td>
</tr>
<tr>
<td>5.3</td>
<td>Triangle subdivision</td>
<td>74</td>
</tr>
<tr>
<td>5.4</td>
<td>Triangulation: uniform</td>
<td>75</td>
</tr>
<tr>
<td>5.5</td>
<td>Model problem 1: analytical</td>
<td>76</td>
</tr>
<tr>
<td>5.6</td>
<td>Model problem 1: approximate</td>
<td>77</td>
</tr>
<tr>
<td>5.7</td>
<td>Model problem 2: analytical</td>
<td>79</td>
</tr>
</tbody>
</table>
List of Figures

5.8  Model problem 2: approximate  . . . . . . . . . . . . . . . . . . . . . . . . . 80
List of Tables

2.1 Quadratic S-spline knot multisets ......................... 15
2.2 Spline basis conversion matrix ............................... 22
5.1 Model problem 1: convergence rates ....................... 77
5.2 Model problem 2: convergence rates ....................... 78
A.1 SplineFunction properties ................................. 86
CHAPTER 1

Introduction

The Finite Element Method (FEM) is a technique for solving partial differential equations (PDEs) numerically, which is used in a wide variety of engineering disciplines. By multiplying the PDE by certain test functions and integrating, one arrives at the so-called weak form of the PDE. By discretizing and restricting the test functions to a finite dimensional space of relatively simple functions, typically piecewise linear functions, or so-called “hat functions” on triangles, one obtain an approximate solution by solving a linear system numerically. Recently, the idea to replace these hat functions by functions with higher smoothness, like splines, has gained momentum, most notably in the new research field of isogeometric analysis (IGA) [HCB05].

1.1 Aim and structure of the thesis

In this thesis we solve a fourth-order partial differential equation by the finite element method using the Hermite nodal basis for the $C^1$ continuous quadratic splines on the Powell–Sabin 12-split of a triangle. A set of simplex splines described in [CLR13] is used for evaluation. We provide an overarching introduction to the various components of the finite element method, and formulate the method in such a fashion that the transition from the mathematical statement to an implementation on a computer is as small as possible.
1. Introduction

Thesis outline

Chapter 2 gives a general introduction to multivariate splines, and summarizes some of their properties. We then turn to the specifics of a space of $C^1$ quadratic splines on the Powell–Sabin 12-split of a triangle, and introduce two sets of bases for this space: the S-basis which will be used for the evaluation and differentiation of splines; and the Hermite nodal basis for which continuity conditions are satisfied automatically.

Chapter 3 introduces the theoretical framework underlying the finite element method, namely that of Sobolev spaces, consisting of functions admitting weak derivatives. A few integral identities known as Gauss–Green formulas are derived. Subsequently, the biharmonic equation is introduced, and existence and uniqueness of weak solutions are shown.

Chapter 4 focuses on the finite element method. Based on the weak formulations from the preceding chapter, a finite element formulation is derived and the notion of a finite element is rigorously defined. A finite element based on the $C^1$ quadratic splines on the Powell–Sabin 12-split is introduced, and interpolation estimates are derived by introducing a similar element with affine properties, enabling us to utilize standard interpolation results.

Chapter 5 demonstrates the transition from a mathematical statement, to the derivation of a linear system suitable for solving on a computer. Two model problems are constructed, and error convergence rate estimates are seen to agree with the theoretical results.

Appendix A gives a small showcase of the SSplines library implemented during the writing of this thesis, and some useful Python packages for the implementation of finite elements are mentioned.
1.2 Motivation

Solving a fourth order partial differential equation using a direct finite element method requires $C^1$ continuity between adjacent finite elements. By introducing degrees of freedom corresponding to derivatives on the element boundary, such a continuity may be enforced. There are many finite elements of $C^1$ continuity. The *Argyris triangle* consisting of a space of pure quintic polynomials being one commonly used element. However, if one desires to use a polynomial space of lower degree, then there are not necessarily enough degrees of freedom to enforce the inter-element continuity requirements. However, by subdividing each triangle into a set of sub-triangles and instead considering piecewise polynomials, one may add enough internal degrees of freedom to preserve the internal $C^1$ continuity, while still having enough degrees of freedom left to enforce the inter element continuity conditions. Such elements are often referred to as *macro-elements* or *composite elements*. Notable examples are the *Clough–Tocher 3-split* [CT65] consisting of piecewise cubic polynomials; the *Powell–Sabin 6-split* (used for the solution of PDEs in [SDV06; Spe08]) and the *Powell–Sabin 12-split* [PS77] (used for example in [Osw92]) consisting of piecewise quadratic polynomials, the latter being the element of study in this thesis. All of these are displayed diagrammatically in Figure 1.1.

The classical approach to working with triangular macro elements has been to employ Bernstein–Bézier techniques for the evaluation of each polynomial piece over the sub-triangles of the macro element and then enforcing the internal continuity requirements manually. In recent studies [CLR13] an explicit simplex spline basis for the $C^1$ quadratic spline space over the Powell–Sabin 12-split has been developed, which removes the need for manually enforcing internal continuity conditions. Furthermore, continuity conditions are given for smooth joins of adjacent elements, however these can be a bit tricky to implement in a finite element setting. In this thesis, we employ this explicit basis for easy evaluation of splines over each macro triangle. In addition to this, the Hermite nodal basis known from the finite element literature is used for easy inter-element continuity
1. Introduction

**Figure 1.1:** Four common triangular $C^1$ finite elements. Note that the Argyris triangle is the only element using pure polynomials of degree five, and no subdivision is needed. For the cubic Clough–Tocher a subdivision into three triangles is used. For the quadratic Powell–Sabin 6- and 12-split, a subdivision into six and twelve triangles respectively are needed.

conditions, ensuring global $C^1$ continuity across a full triangulation. The Hermite nodal basis is expressed in terms of the simplex spline basis for ease of evaluation. Consequently, we utilize the strength of both bases when used in the finite element setting.

1.3 Suggestions for future work

It is of interest to examine whether the method described and implemented in this thesis can be extended to spline spaces of higher smoothness, for instance, the space of internally $C^3$ and globally $C^2$ quintic splines over the Powell–Sabin 12-split of a triangulation, introduced in [LM15]. Furthermore, the method of adaptive refinement often employed in the finite element method could be studied in this spline setting. Firstly, $h$-refinement, where the triangulation is refined by splitting each macro-triangle into a set of sub-triangles; $p$-refinement, or degree elevation, where the polynomial degrees are increased on each element; and finally, $k$-refinement (introduced in the context of IGA), where both mesh refinement and degree elevation are performed: $p$-refinement, followed by $h$-refinement.

The finite element method relies on the evaluation of basis functions and their derivatives over each element in the mesh. The most common
finite element implementation paradigm relies on the use of reference elements, where the values and derivatives of the basis functions are described on a single reference element and subsequently transformed via an affine map to each element in the mesh. This methodology however is not trivial when some of the degrees of freedom are directional derivatives, as is the case for \( C^1 \) elements, and this is the reason that high end robust and commercial finite element implementations seldom support \( C^1 \) elements. Recently, a general method for the transformation of elements was proposed [Kir17] and it would be interesting to see whether this method could be applied to the finite element discussed in this thesis.

There is ongoing work [LM18] on developing similar simplex spline bases for the set of \( C^1 \) cubic splines over the Clough–Tocher split. The transition from the implementation from this thesis, to one employing this newly constructed basis for the Clough–Tocher split is minimal, due to the similar construction. Furthermore, a simplex spline basis for \( C^2 \) cubic splines on the Powell–Sabin 12-split, analogous to the \( C^1 \) quadratic basis discussed in this thesis, is currently being developed [LM].

A macro element similar to the PS12-element studied here is presented in [LM14] for the space of quintic splines with global \( C^2 \) and local \( C^3 \) smoothness. For fast evaluation a subdivision scheme is proposed. The implementation done related to this thesis could be extended to this setting.
In this chapter, we give a general definition of a simplex spline. These splines are non-negative piecewise polynomials with local support, and form a multivariate generalization of univariate B-splines. Recall that univariate B-splines are piecewise polynomials defined over certain sets of knots, where the restriction to any such knot interval is a polynomial of some prescribed degree. The following multivariate construction is geometric in nature with the univariate splines as a special case. The material follows closely that of [PBP02], which gives a thorough summary of spline techniques, and provides a solid reference for the interested reader.

We start by giving some preliminary definitions, and define the main object of study, namely the normalized simplex spline. We then discuss recurrence relations which form the basis for much of the computational methods involving splines. We then summarize some results pertaining to the simplex splines, which consider continuity properties of simplex splines and how these relate to the simplex spline knots. In order to compute with simplex splines, a suitable basis is needed. By fixing the geometry it is possible to construct explicit spline bases. We will discuss two particular bases for the $C^1$ quadratic simplex splines on the fixed geometry of the Powell–Sabin 12-split, which will be employed throughout this thesis.
2. An Introduction to Simplex Splines

2.1 Definitions

We start by giving some preliminary definitions, starting with the notion of a simplex—a geometric object which generalizes the concept of points, lines and triangles.

Preliminaries

We say that a set of $k+1$ points $p_1, \ldots, p_{k+1}$ are affinely independent, if the $k$ vectors $p_2 - p_1, \ldots, p_{k+1} - p_1$ are linearly independent.

**Definition 2.1.1** ($k$-Simplex). Given a set of $k+1$ affinely independent points $v_1, \ldots, v_{k+1}$ in $\mathbb{R}^k$, then the $k$-simplex $\sigma$ defined by them is given as

$$\sigma := \left\{ \sum_{i=1}^{k+1} c_i v_i : \sum_{i=1}^{k+1} c_i = 1 \text{ and } c_i \geq 0 \text{ for all } i \right\}.$$ 

In other words, $\sigma$ consists of all convex combinations of the vectors $v_i$.

Some standard examples of simplices are the 1-simplex, a line segment; the 2-simplex, a triangle; and the 3-simplex; a tetrahedron. If we are given a $k$-simplex, it is natural to talk about its $k$-dimensional volume, being a generalization of length in one dimension, and area in two dimensions. The $k$-dimensional volume of a $k$-simplex $\sigma$ with vertices $v_1, \ldots, v_{k+1}$ is given by

$$\text{Vol}_k(\sigma) := \frac{1}{k!} \left| \det(v_2 - v_1, \ldots, v_{k+1} - v_1) \right|. \quad (2.1)$$

**Remark 2.1.2.** Note that for $k = 2$, $\text{Vol}_2(\sigma)$ coincides with the standard definition of the area, and for $k = 1$ it coincides with the length. We will in this specific cases use the notation

$$\text{area}(\sigma) = \text{Vol}_2(\sigma)$$

$$\text{length}(\sigma) = \text{Vol}_1(\sigma)$$

Given $k$ and $s$ two positive integers with $k \geq s$, we denote by $\pi$ the canonical projection from $\mathbb{R}^k$ to $\mathbb{R}^s$. That is, if $x = (x_1, \ldots, x_k)$ is a point in
2.1. Definitions

$\mathbb{R}^k$, then $\pi(x) = (x_1, \ldots, x_s)$. The fiber of a point $y$ in $\mathbb{R}^s$ is defined as the set of points in $\mathbb{R}^k$ that project down to $y$:

$$\pi^{-1}(y) := \left\{ x \in \mathbb{R}^k : \pi(x) = y \right\}.$$

**Simplex spline**

With this, we have all the components we need to define a simplex spline.

**Definition 2.1.3** (Simplex Spline). Let $\sigma$ be a $k$-simplex in $\mathbb{R}^k$ with $\operatorname{Vol}_k(\sigma) > 0$. On $\mathbb{R}^s$, we define the $s$-variate simplex spline $M_{k,\sigma}: \mathbb{R}^s \to \mathbb{R}$ from $\sigma$ by

$$M_{k,\sigma}(x) := \operatorname{Vol}_{k-s}(\sigma \cap \pi^{-1}(x)).$$

In general, depending on the integers $k$ and $s$, there are infinitely many simplices in $\mathbb{R}^k$ whose vertices coincide when projected down to $\mathbb{R}^s$. Ultimately, we would like a representation of the simplex spline that is independent of which specific $k$-simplex we choose. We introduce the notion of knots, associated to a given simplex spline and projection.

**Definition 2.1.4** (Simplex Spline Knots). Given a simplex spline $M_{k,\sigma}$, where $\sigma$ is a $k$-simplex with vertices $v_1, \ldots, v_{k+1}$, we define the knots $a_1, \ldots, a_{k+1}$ of $M_{k,\sigma}$ to be the projection of the vertices:

$$a_i := \pi(v_i)$$

for $i = 1, \ldots, k+1$.

By normalizing the simplex spline with the volume of the defining simplex $\sigma$, we obtain a spline that is independent of the specific simplex used [PBP02, Chapter 18]. We introduce the normalized simplex spline as follows:
2. An Introduction to Simplex Splines

**Definition 2.1.5** (Normalized simplex spline). Let $a_1, \ldots, a_{k+1}$ be a set of $k + 1$ points in $\mathbb{R}^s$. Choose representatives $v_i$ in the set of fibers of $a_i$ for each $i$ so that $v_1, \ldots, v_{k+1}$ constitutes an affinely independent set spanning a $k$-simplex $\sigma$. We then define the normalized simplex spline $M$ in terms of the knots $a_i$ as:

$$M(x \mid a_1, \ldots, a_{k+1}) := \frac{M_{k,\sigma}(x)}{\text{Vol}_{k}(\sigma)}.$$ 

This normalized simplex spline $M$ has unit integral:

$$\int_{\mathbb{R}^s} M(x \mid a_1, \ldots, a_{k+1}) \, dx.$$

**Remark 2.1.6.** The notation $M(x \mid a_1, \ldots, a_{k+1})$ can be a bit cumbersome. To ease the notation, we often denote the knot dependency using a multiset $K = \{k_1^{m_1}, \ldots, k_{\ell}^{m_{\ell}}\}$, where the non-negative integers $m_i$ represent the knot multiplicities and we write

$$M(x \mid K) := M(x \mid k_1^{m_1}, \ldots, k_{\ell}^{m_{\ell}}).$$

Furthermore, we denote by $[K]$ the convex hull of the knot set, and $|K|$ denotes the total number of knots, $m_1 + \ldots + m_{\ell}$ counted with multiplicity. In this notation, a triangle with vertices $a_1, a_2, a_3$ can conveniently be expressed using the convex hull notation:

$$K = [a_1, a_2, a_3]. \quad (2.2)$$

By sticking to the normalized simplex splines, we can work in a “bottom up” fashion, by specifying the domain of the simplex spline in terms of the knots, and not worry about what simplices we should choose. We will later see how the knots of a simplex spline influences the continuity and degree of the resulting spline.

### 2.2 Knot insertion and recursion

One of the nice properties of univariate $B$-splines, is that they satisfy a recursion formula. We would like our simplex splines to satisfy a similar
2.2. Knot insertion and recursion

kind of relation. We first extend the concept of knot insertion. To this end, assume that we are working with a $k$-simplex $\sigma$ with vertices $v_1, \ldots, v_{k+1}$ and corresponding knots $a_i = \pi(v_i)$.

**Knot insertion**

Given an arbitrary point $a_0 \in \mathbb{R}^s$, we can represent this new knot $a_0$ as an affine combination of the old knots, say $a_0 = \sum_{i=1}^{k+1} \xi_i a_i$, where the weights $\xi_i$ sum to one. We view this new point $a_0$ as the projection of a point

$$v_0 = \sum_{i=1}^{k+1} \xi_i v_i.$$  

Since the original vertices $v_1, \ldots, v_{k+1}$ are affinely independent, we can create $k + 1$ new simplices $\sigma_i$ by replacing vertex $v_i$ in $\sigma$ by $v_0$. It turns out that these new simplices partition $\sigma$ in such a way that the corresponding simplex splines relate to each other. This is known as *Micchelli’s knot insertion formula* [Mic80].

**Theorem 2.2.1** (Micchelli’s Knot Insertion Formula). Let $\sigma$ be a $k$-simplex with vertices $v_1, \ldots, v_{k+1}$ and corresponding knots $K = \{a_1, \ldots, a_{k+1}\}$. Introduce a new knot $a_0$ as an affine combination of the old knots with weights $\xi_i$ that sum to one. Then the corresponding normalized simplex spline satisfies the following relation:

$$M(x \mid K) = \sum_{i=1}^{k+1} \xi_i M(x \mid (K \setminus \{a_i\}) \cup \{a_0\})$$

where $K \setminus \{a_i\} = \{a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_{k+1}\}$ denotes the knot multiset obtained by removing $a_i$.

**Simplex spline recursion**

If we assume that the knot $a_0$ in Micchelli’s knot insertion formula coincides with the point of evaluation $x$ we obtain a recurrence relation where the spline $M$ can be represented as a combination of splines of one degree lower. This is known as *Micchelli’s recurrence relation*. 

11
2. An Introduction to Simplex Splines

Theorem 2.2.2 (Micchelli’s Recurrence Relation). Let $\sigma$ be a $k$-simplex with vertices $v_1, \ldots, v_{k+1}$ and corresponding knots $K = \{a_1, \ldots, a_{k+1}\}$. Assume $k > s$ and that $x$ is an affine combination of the knots with weights $\xi_i$ that sum to one. Then the corresponding normalized simplex spline satisfies the following relation:

$$M(x \mid K) = \frac{k}{k-s} \sum_{i=1}^{k+1} \xi_i M(x \mid K \setminus \{a_i\}).$$

Consequently, a spline defined by $k + 1$ knots can be represented as a weighted sum of splines defined by $k$ knots.

![Figure 2.1](image)

**Figure 2.1:** Here, the base simplex $\rho$ in red is a line segment, and the full simplex $\sigma$ in blue, a triangle. In this case, $k = 2$, and we have that $\text{area}(\sigma) = \frac{1}{2} \text{length}(\rho)$, which we recognize as the standard formula for the area of a triangle, base times height over two.

**Proof.** Let $\sigma$ be the $k+1$-simplex with vertices $p, v_1, \ldots, v_{k+1}$, where $\pi(p) = x$. Furthermore, let $\rho$ be the $k$-simplex spline with vertices $v_1, \ldots, v_{k+1}$.

As in [PBP02] we consider only a special case, where the simplex $\rho$ lies in a hyperplane orthogonal to the fibers of the canonical projection, see Figure 2.1 for an example. The proof of the general case can be found in [Mic80].

For simplicity, denote by

$$M_i(x) := M(x \mid (K \setminus \{a_i\}) \cup \{x\}).$$
These are the splines occurring on the right hand side in Micchelli’s knot insertion formula where the new knot is $x$. If $h$ denotes the euclidean distance from the point $p$ to the simplex $\rho$, then we have that

$$
\begin{align*}
\text{Vol}_k(\sigma) &= \frac{h}{k} \text{Vol}_{k-1}(\rho), \\
\text{Vol}_{k-s}(\sigma \cap \pi^{-1}(x)) &= \frac{h}{k-s} \text{Vol}_{k-s-1}(\rho \cap \pi^{-1}(x))
\end{align*}
$$

In view of Definition 2.1.5 on page 10, the terms to the left are exactly the terms in the definition of $M(x | K)$, and similarly, the ones to the right contain the terms occurring in the definition of $M_i(x)$, so we have that

$$
M_i(x) = \frac{k}{k-s} M(x | K \setminus \{a_i\})
$$

Substituting this into the relation in Theorem 2.2.1 yields the result. □

**Properties of simplex splines**

We wrap up this section by summarizing some general properties of these normalized simplex splines that follow directly from the construction, or from the recurrence relation. These properties are used in subsequent sections.

**Lemma 2.2.3** (Simplex Spline Properties). The $s$-variate normalized simplex spline $M(x | K)$ with knot set $K = \{a_1, \ldots, a_{k+1}\}$ has the following properties:

1. $M$ is a piecewise polynomial of degree at most $k - s$;

2. $M$ is $k - m - 1$ continuously differentiable over any knot line containing at most $m$ collinear knots.

3. $M$ is an $s$-variate Bernstein polynomial whenever the knot set $K = [a_1^{m_1}, \ldots, a_{s+1}^{m_{s+1}}]$ contains exactly $s + 1$ distinct knots.

4. $M$ is supported in the convex hull of the defining knot set,

$$
\text{Supp}(M) = [K].
$$
2. An Introduction to Simplex Splines

2.3 The Powell–Sabin 12-split

In this section we focus our attention to bivariate simplex splines defined over triangular regions. We will spend time looking at the Powell–Sabin 12-split (PS12) [PS77] of a triangle, which is a subdivision that enables us to define simplex splines that have continuous derivatives internal to the triangle. Splines on this subdivision will be the central topic in this thesis. We start by fixing some notation, and subsequently we introduce the S-basis for piecewise quadratic $C^1$ splines on the PS12-split.

Notation

Let $K = [a_1, a_2, a_3]$ denote a non-degenerate triangle in $\mathbb{R}^2$. Given such a triangle, we denote by $K_{PS12}$ the Powell–Sabin 12-split of the triangle. This is a triangulation consisting of ten vertices $a_1, \ldots, a_{10}$ and twelve faces $K_1, \ldots, K_{12}$ delineated by the complete graph formed by connecting the mid-points of the triangle edges. We employ the numbering of sub-triangles and vertices shown in Figure 2.2.

![Figure 2.2: The Powell–Sabin 12-split of a triangle, with its associated face and vertex numbering.](image)

Given a triangulation $\mathcal{T}$ of some domain $\Omega$ in $\mathbb{R}^2$ and integers $-1 \leq r < d$ we are interested in spline spaces of the form

$$\mathcal{S}_r^d(\mathcal{T}) = \{ f \in C^r(\Omega) : f|_K \in \Pi_d \text{ for all } K \in \mathcal{T} \}$$
2.3. The Powell–Sabin 12-split

consisting of splines of degree $d$ with $C^r$ continuity over $\Omega$. In particular, we are interested in the spaces $S^1_2(K_{PS12})$ and $S^1_2(T_{PS12})$. Here, $T_{PS12}$ denotes the triangulation obtained by splitting each triangle in $T$ according to the Powell–Sabin 12-split.

**A basis for $C^1$ quadratic splines on the PS12-split**

Having a fixed geometry to work over, as is the case for the PS12, we may consider basis functions for spline spaces over this geometry. In this section we look at the basis for the space $S^1_2(K_{PS12})$ developed in [CLR13]. One of the desirable properties of this specific basis is the matrix recurrence relation down to hat functions for both evaluation and differentiation of splines on the PS12. These basis functions are specified in terms of scaled versions of the normalized simplex splines from Definition 2.1.5.

Let $\nu_d := \dim(\Pi_d) = \binom{d+2}{2}$ denote the dimension of the set of bivariate polynomials of degree at most $d$. Following [CLR13, Definition 2.1], we define the knot multisets $K_i$ as in Table 2.1, and the quadratic S-spline basis as follows:

**Table 2.1:** The knot multisets corresponding to the quadratic S-spline basis. Here $a_1, \ldots, a_6$ are the first six vertices of the Powell–Sabin 12-split, see Figure 2.2 on page 14.

<table>
<thead>
<tr>
<th>$K_1$</th>
<th>$K_2$</th>
<th>$K_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${a_1^2, a_4, a_6}$</td>
<td>${a_2^2, a_4, a_2, a_6}$</td>
<td>${a_1, a_4, a_2, a_5, a_6}$</td>
</tr>
<tr>
<td>$K_4$</td>
<td>$K_5$</td>
<td>$K_6$</td>
</tr>
<tr>
<td>${a_2^2, a_5, a_1, a_4}$</td>
<td>${a_2^2, a_5, a_4}$</td>
<td>${a_2^2, a_5, a_4, a_1}$</td>
</tr>
<tr>
<td>$K_7$</td>
<td>$K_8$</td>
<td>$K_9$</td>
</tr>
<tr>
<td>${a_2^2, a_5, a_3, a_6, a_4}$</td>
<td>${a_2^2, a_6, a_2, a_5}$</td>
<td>${a_2^2, a_6, a_5}$</td>
</tr>
<tr>
<td>$K_{10}$</td>
<td>$K_{11}$</td>
<td>$K_{12}$</td>
</tr>
<tr>
<td>${a_2^2, a_6, a_4, a_3}$</td>
<td>${a_4, a_6, a_4, a_5}$</td>
<td>${a_2^2, a_4, a_3, a_6}$</td>
</tr>
</tbody>
</table>

**Definition 2.3.1** (Quadratic S-Spline Basis). Given the knot multiset defined in Table 2.1, the **quadratic S-spline basis functions** are defined as

$$S_i(x) := \frac{\text{area}(\left[K_i\right])}{\nu_2} M(x | K_i)$$

for $i = 1, \ldots, 12$ where $M(x | K_i)$ is a normalized bivariate simplex spline (see Definition 2.1.5), and $\nu_2 = 6$. 

15
Remark 2.3.2. As a sanity check, we verify that these splines satisfy the properties listed in Lemma 2.2.3.

1. Note first that $|K_i| = k + 1 = 5$. With $s = 2$ and $k = 4$, we see that these splines are piecewise polynomials of degree $k - s = 2$.

2. Furthermore, since all knot lines crossing the interior of the triangle $K$ contains $m = 2$ knots, we can deduce that $S_i$ is $k - m - 1 = 1$ times continuously differentiable in the interior of $K$. Therefore, $S_i \in C^1(K)$.

3. The knot multisets $K_1, K_5$ and $K_9$ contain exactly $s + 1 = 3$ distinct knots, and consequently, the basis functions $S_1, S_5$ and $S_9$ are bivariate Bernstein polynomials.

4. The splines are supported in the convex hull $[K_i]$ of the defining knot multiset. Consequently, by counting, we see there are exactly $\nu_2 = 6$ non-zero S-splines on each sub-triangle $K_i$.

The twelve S-splines on an equilateral triangle can be seen in Figure 2.4. Since there are six non-zero S-splines on each sub-triangle, the scaling factors $\text{area}([K_i])/\nu_2$ ensure that the S-splines form a partition of unity, that is

$$\sum_{i=1}^{12} S_i = 1.$$ 

Given a set of twelve coefficients $c_1, \ldots, c_{12}$, we refer to a linear combination $s = c_1S_1 + c_2S_2 + \ldots + c_{12}S_{12}$ as a quadratic spline function on $K_{PS12}$. These span the entire spline space $S_2^2(K_{PS12})$. An example with alternating coefficients $c_i = (-1)^i$ for $i = 1, \ldots, 12$ is shown in Figure 2.3.

**Evaluation and differentiation**

The S-spline basis satisfies a matrix recurrence relation for the evaluation and differentiation of quadratic simplex splines on the PS12. The algorithm is described in [CLR13, Algorithm 4.2], where the S-spline matrices only
2.3. The Powell–Sabin 12-split

\[ s = -S_1 + S_2 - \ldots - S_{11} + S_{12}. \]

depend on the barycentric coordinates of the point of evaluation. This algorithm has the benefit of computational efficiency.

Recurrence relation and polynomial pieces

Since the simplex splines satisfy Theorem 2.2.2, it is possible to explicitly find the twelve polynomial pieces of a simplex spline over the PS12. Polynomial expressions of splines can be useful in for instance symbolic computation or for debugging numerical evaluation routines. Over the Powell–Sabin 12-split, with \( K \) composed of vertices of \( K_{ps12} \), the normalized simplex spline \( M(x \mid K) \) can be equivalently defined recursively as

\[
M(x \mid K) = Q(x \mid K) \frac{\text{area}(K)}{\left| K \right| - 1} \]

where

\[
Q(x \mid K) = \begin{cases} 
0, & \text{if } \text{area}(K) = 0; \\
\frac{1}{\text{area}(K)} I(K)(x), & \text{if } \text{area}(K) \neq 0 \text{ and } \left| K \right| = 3; \\
\sum_{i=1}^{6} \beta_i Q(x \mid K \setminus \{a_i\}), & \text{if } \text{area}(K) \text{ and } \left| K \right| > 3.
\end{cases}
\]

Here, \( x = \beta_1 a_1 + \ldots + \beta_6 a_6 \), with the \( \beta_i \) summing to one and \( \beta_i = 0 \) whenever \( a_i \notin K \). This means that we can construct an algorithm.
2. An Introduction to Simplex Splines

\( S_i \) on an equilateral triangle. Note that for the cases \( i = 1, 5, 9 \), the S-splines \( S_i \) are scaled Bernstein polynomials.

**Figure 2.4:** The twelve quadratic S-spline basis functions \( S_i \) on an equilateral triangle. Note that for the cases \( i = 1, 5, 9 \), the S-splines \( S_i \) are scaled Bernstein polynomials.
for computing pieces of the S-spline basis by recursing down to scaled Bernstein polynomials.

**The Hermite nodal basis for \( C^1 \) quadratic splines on the PS12-split**

An alternative to the S-basis for piecewise quadratic splines on the PS12 is the so-called **Hermite nodal basis**, which interpolates values and gradients at the vertices as well as outward facing unit normal derivatives at the midpoint of each edge. This basis is very useful in the finite element method, as it provides easy conditions for smooth joins of function spaces on adjacent triangles.

A conversion between the S-basis and the Hermite nodal basis is explicitly given in [CLR13, Section 8.2] in terms of an invertible matrix \( H \), as seen in Table 2.2. Due to the simple construction of smooth spline spaces on triangulations using the Hermite nodal basis, and the numerical efficiency of the matrix recursion for the S-spline basis, we will make extensive use of the conversion between these two in the implementation. The twelve Hermite basis functions can be seen in Figure 2.5 on page 21.

**Definition 2.3.3.** The Hermite nodal basis are the functions \( \mathcal{H}_i \) for \( i = 1, \ldots, 12 \) satisfying the relations

\[
\Psi_j(\mathcal{H}_i) = \delta_{ij}
\]

where \( \mathcal{N}_{PS12} = \{ \Psi_j \}_{j=1}^{12} \) are the linear functionals defined by

\[
\Psi(f) = [\Psi_1(f), \ldots, \Psi_{12}(f)]
\]

\[
= [f(a_1), f_x(a_1), f_y(a_1), f_n_3(a_4)), \ldots, f_n_2(a_6)]]
\]

Here, \( n_i \) denotes the **outward facing unit normal** to the edge opposite \( a_i \). Note that this is in contrast to [CLR13], where inward normal vectors are used instead.
2. An Introduction to Simplex Splines

It was shown in [PS77] that the values taken by the functionals in $N_{PS12}$ uniquely determine a spline in $S^1_2(K_{PS12})$, and we here briefly summarize the proof. See for instance [LS07] for an alternative proof using so called minimal determining sets.

**Lemma 2.3.4.** The set $N_{PS12}$ determine the space $S^1_2(K_{PS12})$. If $s \in S^1_2(K_{PS12})$ satisfies $\Psi_i(s) = 0$ for all $\Psi_i \in N_{PS12}$, then $s = 0$.

**Proof.** Recall the numbering scheme used for the PS12-split shown in Figure 2.2.

1. Using that $f$ is a piecewise quadratic $C^1$ spline whose values and gradients vanishes at the vertices of the triangle, deduce that $f$ and that the gradient component parallel to the perimeter is zero along the perimeter.

2. Using that the component of the gradient that is perpendicular to the perimeter vanishes at the midpoints $a_4, a_5, a_6$, deduce that $f$ and the component of the gradient parallel to the perimeter of the interior triangle $K^* = [a_4, a_5, a_6]$ is zero along this perimeter.

3. Using that the gradient component perpendicular to the perimeter of the interior triangle $K^*$ varies linearly along the perimeter, deduce that this is zero due to the data prescribed at the vertices of $K^*$.

4. Using the continuity of the gradient, deduce that $f$ is zero at the midpoint $a_{10}$, which implies that $f$ is zero over all of $K^*$, which again implies that $f$ is zero over the entirety of $K$. ■
2.3. The Powell–Sabin 12-split

**Figure 2.5:** The twelve Hermite nodal basis functions $H_i$ on an equilateral triangle. Make note of the difference in scale. Also, note that some of the basis functions take on negative values, in contrast with the S-spline basis functions which are all positive.
Table 2.2: With $h = [H_1, \ldots, H_{12}]$ and $s = [S_1, \ldots, S_{12}]$ being two vectors containing the Hermite nodal basis functions and the S-basis functions respectively, we have that $h = sH$. That is, column $i$ of $H$ corresponds to the spline coefficients of $H_i$ in terms of the S-basis. In the matrix, $a_{ij} = a_i - a_j$ where $a_i = (x_i, y_i)$ are the vertices of the triangle,

$$
\ell_{ijk} = a_{ij}^T a_{jk} / (a_{ij}^T a_{ij})
$$

and $\delta$ is twice the signed area of the triangle. Note that the number of zeros are high so this matrix is easy to compute and incurs no significant overhead.

$$
H = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & \frac{1}{3}x_{21} & \frac{1}{3}y_{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-\frac{2}{3}\ell_{126} & \frac{1}{2}x_{12}\ell_{126} & \frac{1}{2}y_{12}\ell_{126} & \delta/(6\|a_{12}\|_2) & -\frac{2}{3}\ell_{125} & \frac{1}{6}x_{21}\ell_{125} & \frac{1}{6}y_{21}\ell_{125} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & \frac{1}{3}x_{12} & \frac{1}{3}y_{12} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & \frac{1}{3}x_{32} & \frac{1}{3}y_{32} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -\frac{2}{3}\ell_{234} & \frac{1}{6}x_{23}\ell_{234} & \frac{1}{6}y_{23}\ell_{234} & \delta/(6\|a_{23}\|_2) & -\frac{2}{3}\ell_{246} & \frac{1}{6}x_{23}\ell_{246} & \frac{1}{6}y_{23}\ell_{246} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \frac{1}{3}x_{23} & \frac{1}{3}y_{23} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \frac{1}{3}x_{13} & \frac{1}{3}y_{13} & 0 \\
-\frac{2}{3}\ell_{134} & \frac{1}{2}x_{13}\ell_{134} & \frac{1}{2}y_{13}\ell_{134} & 0 & 0 & 0 & 0 & 0 & -\frac{2}{3}\ell_{315} & \frac{1}{6}x_{31}\ell_{315} & \frac{1}{6}y_{31}\ell_{315} & \delta/(6\|a_{31}\|_2) \\
1 & \frac{1}{2}x_{31} & \frac{1}{2}y_{31} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$
The Finite Element Method, which we will discuss in Chapter 4, is based on the variational formulation of the partial differential equation in question. The problem to be solved is expressed in terms of abstract operators $a$ and $L$ between function spaces of so-called admissible functions:

$$a: U \times V \to \mathbb{R},$$
$$L: V \to \mathbb{R}$$

The task at hand is then to solve the variational problem: Find $u \in U$ such that $a(u, v) = L(v)$ for all $v \in V$. As we shall see, the intricate part of the variational formulation of a partial differential equation is not necessarily to determine the appropriate abstract operators $a$ and $L$, but rather determine which function spaces $U$ and $V$ one should employ.

In this chapter, we will give an introduction to Sobolev spaces, which provide the correct framework for finding proper choices of $U$ and $V$. The concept of Sobolev spaces is based upon the notion of a weak derivative, which extends the notion of differentiability to functions not necessarily differentiable in the classical pointwise sense.

We start by giving some preliminary definitions, and a motivating example of a weak derivative. We then proceed by defining Sobolev spaces, and look at some of their properties. We then focus our attention on
specifics related to the biharmonic equation, which will be the partial
differential equation of interest in the rest of the thesis.

In the following, we denote by Ω an open bounded subset of \( \mathbb{R}^n \) with a
sufficiently well-behaved boundary \( \Gamma \) and outward facing normal vector \( \nu \), as seen in Figure 3.1. We will employ the multi-index notation for mixed partial derivatives, where \( \alpha = (\alpha_1, \ldots, \alpha_n) \) is a **multi-index** of order
\( |\alpha| = \alpha_1 + \ldots + \alpha_n \). We then define the mixed partial derivative operator \( D^\alpha \) as
\[
D^\alpha := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}.
\]
We reserve the notation \( \partial_i \) for the operator \( D^\alpha \) for which
\[
\alpha = (\delta_{i,1}, \ldots, \delta_{i,n}).
\]
I.e., the first order partial derivative with respect to component \( x_i \). We
denote by \( \partial_{ij} = \partial_i \partial_j \) the second order mixed partial derivative with respect
to \( x_i \) and \( x_j \), and similarly for higher order derivatives. Furthermore, for a
direction vector \( \nu = (\nu_1, \ldots, \nu_n) \), we denote by \( \partial_\nu \) the directional derivative
in the direction \( \nu \):
\[
\partial_\nu = \sum_{i=1}^{n} \partial_i \nu_i.
\]
The directional derivative can also be expressed as \( \nabla u \cdot \nu \) where \( \nabla \) denotes the gradient
\[
\nabla = (\partial_1, \ldots, \partial_n). \tag{3.1}
\]

## 3.1 Preliminary definitions

In order to define the concept of **weak derivative**, we first need to define
two specific classes of functions.

**Definition 3.1.1** (Test functions). Given an open subset \( \Omega \) of \( \mathbb{R}^n \), we
denote by \( \mathcal{D}(\Omega) \) the set of infinitely differentiable real valued functions
\( \varphi : \Omega \to \mathbb{R} \) with compact support in \( \Omega \). We refer to a function \( \varphi \in \mathcal{D}(\Omega) \) as a **test function**.
3.1. Preliminary definitions

Remark 3.1.2. Note that due to test functions $\varphi \in \mathcal{D}(\Omega)$ being compactly supported, they vanish sufficiently near the boundary $\Gamma$, hence any boundary integral involving products with $\varphi$ is zero. This fact will be used extensively in the following.

Definition 3.1.3 (Locally integrable function). Given an open subset $\Omega$ of $\mathbb{R}^n$, we say that a function $v : \Omega \rightarrow \mathbb{R}$ is locally integrable (also often referred to as locally summable in the literature) if the integral

$$\int_{\Omega} |v\varphi| \, dx < \infty$$

for all test functions $\varphi \in \mathcal{D}(\Omega)$. We denote the set of all locally integrable functions on $\Omega$ by $L^1_{\text{loc}}(\Omega)$.

Before we give the definition of a weak derivative, we start with a motivating example following closely that of [Eva10].

Example 3.1.4. Consider a function $u \in C^k(\Omega)$ being $k$ times continuously differentiable, and let $\alpha$ be a multi-index of order $|\alpha| = k$. For any test-function $\varphi \in \mathcal{D}(\Omega)$, integration by parts gives us

$$\int_{\Omega} uD^\alpha \varphi \, dx = (-1)^{|\alpha|} \int_{\Omega} (D^\alpha u) \varphi \, dx. \quad (3.2)$$
3. Sobolev Spaces and the Biharmonic Equation

Here we have used the fact that boundary integrals vanish, due to \( \varphi \) being compactly supported (see Remark 3.1.2 on page 25). Entertain now the thought that \( u \) was not \( k \) times continuously differentiable. The left hand side of Equation (3.2) would only make sense if \( u \) was a locally integrable function, so the left hand side can be remedied by making a restriction on \( u \). The right hand side however has no clear cut meaning, due to \( u \) not being in \( \mathcal{C}^k(\Omega) \). The solution to this problem is to look for a locally integrable function \( v \) such that

\[
\int_{\Omega} u D^\alpha \varphi \, dx = (-1)^{|\alpha|} \int_{\Omega} v \varphi \, dx,
\]

for all \( \varphi \in \mathcal{D}(\Omega) \). It is this function \( v \) that we will refer to as a weak derivative of \( u \). We summarize this in a definition.

**Definition 3.1.5** (Weak derivative). Given two locally integrable functions \( u, v \in L^1_{\text{loc}}(\Omega) \), and a multi-index \( \alpha = (\alpha_1, \ldots, \alpha_n) \) of order \( |\alpha| = k \), we say that \( v \) is a \( \alpha^{th} \) weak partial derivative of \( u \) if the following relation is satisfied:

\[
\int_{\Omega} u D^\alpha \varphi \, dx = (-1)^{|\alpha|} \int_{\Omega} v \varphi \, dx,
\]

for all \( \varphi \in \mathcal{D}(\Omega) \). We denote this as \( v = D^\alpha u \), and often specify that this derivative is in the weak sense.

Before we proceed by defining Sobolev spaces, it is worthwhile to give an example of a function that admits a weak derivative, but no traditional derivative.

**Example 3.1.6.** We consider a function that is defined piecewise and is a member of \( \mathcal{C}^0(\Omega) \). As we shall see, its weak derivative turns out to be what we “expect” it to be, in the sense that the derivative can be taken piecewise. To this end, let \( \Omega = (0, 2) \) be an open interval in \( \mathbb{R} \). Define the function \( u: \Omega \to \mathbb{R} \) by

\[
u(x) := \begin{cases} 
  x, & 0 < x \leq 1, \\
  2 - x, & 1 < x < 2.
\end{cases}
\]
We wish to find the weak first derivative of $u$. Given an arbitrary test function $\varphi \in \mathcal{D}(\Omega)$, we have that
\[
\int_0^2 u D^1 \varphi \, dx = \int_0^1 x D^1 \varphi \, dx + \int_1^2 (2 - x) D^1 \varphi \, dx.
\]
Applying integration by parts, and using the fact that $\varphi$ is compactly supported so that boundary integrals vanish, we have that
\[
\int_0^2 u D^1 \varphi \, dx = \varphi(1) - \varphi(1) - \int_\Omega v \varphi \, dx.
\]
Here, $v$ is defined as the equivalence class of the function
\[
x \mapsto \begin{cases} 1, & 0 < x \leq 1, \\ -1, & 1 < x < 2. \end{cases}
\]
Consequently, $v$ satisfies the requirement of being a first weak derivative of $u$, and we denote this $v = Du$.

### 3.2 Sobolev spaces

Having defined what it means for a function $v$ to be a weak partial derivative of a function $u$, we are ready to look at the main object of study in this chapter. We start with the key definition:

**Definition 3.2.1** (Sobolev Space). For an integer $m \geq 0$ we define the **Sobolev space of order $m$ over $\Omega$** as the following set of functions:

\[
H^m(\Omega) := \{ u: \Omega \to \mathbb{R} : D^\alpha u \in L^2(\Omega) \text{ for all } \alpha \text{ with } |\alpha| \leq m \}. \tag{3.3}
\]

That is, $H^m(\Omega)$ consists of all functions whose weak partial derivatives of order at most $m$ exist and are $L^2(\Omega)$-integrable.

**Remark 3.2.2.** There are more general definitions of Sobolev spaces, where the weak derivatives are allowed to be in more general $L^p(\Omega)$ spaces, and these are typically denoted $W^{m,p}(\Omega)$. Setting $p = 2$ coincides with our definition of $H^m(\Omega)$, and this class of functions will suffice for our setting.
3. Sobolev Spaces and the Biharmonic Equation

The spaces $H^m(\Omega)$ are Hilbert spaces when endowed with a certain norm, which we state without proof:

**Lemma 3.2.3.** For $m \geq 0$, the space $H^m(\Omega)$ are Hilbert spaces when endowed with the inner product norm defined by

$$\|u\|_{m,\Omega} := \left( \sum_{|\alpha| \leq m} \int_{\Omega} |D^\alpha u|^2 \, dx \right)^{1/2} \quad (3.4)$$

which measures the $L^2$-norms of the weak partial derivatives.

We also often make use of the associated seminorm defined by

$$|u|_{m,\Omega} := \left( \sum_{|\alpha| = m} \int_{\Omega} |D^\alpha u|^2 \, dx \right)^{1/2}, \quad (3.5)$$

which picks out only the $m^{th}$ order weak partial derivatives.

**Remark 3.2.4.** Note that for $m = 0$, the Sobolev space $H^0(\Omega)$ coincides with the standard $L^2(\Omega)$ space of square integrable functions. Consequently, we will often use the notation $H^0(\Omega)$ with associated norm for $L^2(\Omega)$ and its norm.

By completing the space $D(\Omega)$ with respect to the Sobolev norms $\|\cdot\|_{m,\Omega}$, we obtain the following spaces

$$H^m_0(\Omega) := \overline{D(\Omega)}, \quad (3.6)$$

which can be considered as subspaces of $H^m(\Omega)$ consisting of functions whose $\alpha$th derivatives all vanish at the boundary for $|\alpha| \leq m - 1$.

Later on, when deriving weak formulations of partial differential equations, we need to specify functions that vanish on the boundary $\Gamma$ of the domain $\Omega$. Such a specification of function value along the boundary is in the most general setting far from trivial, and sets requirements on the smoothness of the boundary $\Gamma$. It is sufficient that the boundary is so-called *Lipschitz-continuous*, and in this case it is possible to define the *trace operator*. We will however in the following always assume that the domain
3.2. Sobolev spaces

Ω admits a suitably smooth boundary Γ so that we may characterize the following two spaces, defined above, as:

\[ H^{1,0}_0(\Omega) := \{ u \in H^1(\Omega) : u = 0 \text{ on } \Gamma \}, \]
\[ H^{2,0}_0(\Omega) := \{ u \in H^2(\Omega) : u = \partial_\nu u = 0 \text{ on } \Gamma \}. \] (3.7)

These spaces will used for encoding homogeneous boundary conditions. See [Cia02, §1.2] for a more thorough discussion on the theory of trace operators, and the difficulties of specifying function values along the boundary.

The Sobolev embedding theorem gives conditions for when one Sobolev space is continuously embedded in another. We will need these results later when deriving interpolation estimates, so we include the embedding theorem here without proof, specialized to our setting. A general formulation can be found in for instance [OCS86].

**Theorem 3.2.5** (The Sobolev Embedding Theorem). Let \( \Omega \in \mathbb{R}^n \) be a sufficiently smooth domain. Let \( j \) and \( m \) be non-negative integers. Then the following continuous embedding exist: If \( 2m \leq n \), then

\[ H^{m+j}(\Omega) \hookrightarrow H^j(\Omega). \] (3.8)

Moreover, if \( 2m > n \), we also have

\[ H^{m+j}(\Omega) \hookrightarrow C^j(\Omega). \] (3.9)

Note that this embedding theorem tells us under what conditions we can expect functions in Sobolev spaces to be continuous functions.

There is an important inequality relating the \( L^2 \) norm of a function to the \( L^2 \) norm of its first derivatives. This result is key when deriving equivalent norms on Sobolev spaces, and while we will not use this result directly, we include it, and an application, for completeness.
3. Sobolev Spaces and the Biharmonic Equation

**Proposition 3.2.6** (Poincaré–Friedrich inequality). When the domain $\Omega$ is bounded, there exists a constant $C(\Omega)$ which only depends on $\Omega$ such that

$$\|u\|_{0,\Omega} \leq C(\Omega)|u|_{1,\Omega}$$

(3.10)

for all $u \in H^1_0(\Omega)$.

We can immediately deduce that on a bounded domain, the Sobolev seminorm $|\cdot|_{1,\Omega}$ is equivalent to the full norm $\|\cdot\|_{1,\Omega}$ on $H^1_0(\Omega)$:

**Corollary 3.2.7** (Equivalence of norm and seminorm on $H^1_0(\Omega)$). If the domain $\Omega$ is bounded, then there exists a constant $M$ such that

$$|u|_{1,\Omega} \leq \|u\|_{1,\Omega} \leq M|u|_{1,\Omega}$$

(3.11)

for all $u \in H^1_0(\Omega)$.

**Proof.** The first inequality follows immediately as the seminorm only includes the terms containing derivatives of highest order, and is therefore always smaller than or equal to the full norm. For the second inequality, note that

$$\|u\|^2_{1,\Omega} = \|u\|^2_{0,\Omega} + |u|^2_{1,\Omega}.$$

Applying the Poincaré–Friedrichs inequality yields

$$\|u\|^2_{1,\Omega} \leq (C(\Omega)^2 + 1)|u|^2_{1,\Omega},$$

and taking square roots proves the result with $M = \sqrt{C(\Omega)^2 + 1}$.

**Remark 3.2.8.** By chaining the Poincaré–Friedrich inequality, it is possible to show that on $H^2_0(\Omega)$ over bounded domains, the seminorm $|\cdot|_{2,\Omega}$ is in fact a norm, and it is equivalent to the full norm $\|\cdot\|_{2,\Omega}$. Consequently, there exists a $\kappa > 0$ such that

$$|u|_{2,\Omega} \leq \|u\|_{2,\Omega} \leq \kappa|u|_{2,\Omega}.$$

See for instance [Cia02, §1.2].

30
3.3 Gauss–Green formulas

We now deduce some integration formulas that will be used when deriving variational formulations of partial differential equations. Note that by construction, a fundamental Gauss–Green's lemma is applicable to functions in the space $H^1(\Omega)$ and this fact will be of much use to us. This formula gives a relationship between the $n$-fold integral over an $n$-dimensional domain $\Omega$ and the $(n-1)$-fold integral over a sufficiently smooth $(n-1)$-dimensional boundary $\Gamma$.

**Proposition 3.3.1** (Gauss–Green’s lemma). For any two functions $u, v \in H^1(\Omega)$, we have that

$$
\int_\Omega u \partial_i v \, dx = - \int_\Omega (\partial_i u)v \, dx + \int_\Gamma u \nu_i \, d\gamma \tag{3.12}
$$

for $i = 1, \ldots, n$. Here, $\nu_i$ denotes the $i$th component of the outward pointing unit normal $\nu = (\nu_1, \ldots, \nu_n)$ to $\Omega$.

We can repeatedly apply Proposition 3.3.1 to deduce other similar formulas for functions of higher smoothness, that is, functions residing in higher order Sobolev spaces. For functions in $H^2(\Omega)$, we have the following result.

**Corollary 3.3.2.** If $u \in H^2(\Omega)$ and $v \in H^1(\Omega)$, then

$$
\int_\Omega (\Delta u)v \, dx = - \int_\Omega \nabla u \cdot \nabla v \, dx + \int_\Gamma (\partial_\nu u)v \, d\gamma. \tag{3.13}
$$

**Proof.** Since $u \in H^2(\Omega)$, we have that $\partial_i u \in H^1(\Omega)$ for all $i = 1, \ldots, n$. Applying Proposition 3.3.1 for each $i$ and summing over $i = 1, \ldots, n$ yields

$$
\int_\Omega \sum_{i=1}^n \partial_i u \partial_i v \, dx = - \int_\Omega \sum_{i=1}^n (\partial_i u)v \, dx + \int_\Gamma \sum_{i=1}^n (\partial_\nu u)\nu_i \, d\gamma.
$$

Using the identities

$$
\partial_\nu u = \sum_{i=1}^n \partial_i u \nu_i, \quad \Delta u = \sum_{i=1}^n \partial_i u, \quad \nabla u \cdot \nabla v = \sum_{i=1}^n \partial_i u \partial_i v
$$

31
and rearranging, we get
\[ \int_{\Omega} (\Delta u)v \, dx = -\int_{\Omega} \nabla u \cdot \nabla v \, dx + \int_{\Gamma} (\partial_{\nu} u)v \, d\gamma, \]
as we wanted to show. ■

A similar result for functions in \( H^4(\Omega) \) can be deduced, which will be used in the derivation of the variational formulation of fourth order partial differential equations.

**Corollary 3.3.3.** If \( u \in H^4(\Omega) \) and \( v \in H^2(\Omega) \), then
\[ \int_{\Omega} (\Delta^2 u)v \, dx = \int_{\Omega} \Delta u \Delta v \, dx - \int_{\Gamma} \Delta u \partial_{\nu} v \, d\gamma + \int_{\Gamma} (\partial_{\nu} \Delta u)v \, d\gamma. \]  
(3.14)

**Proof.** We start by applying Corollary 3.3.2 to \( u \) and \( v \) separately, and subtract these equations. This yields
\[ \int_{\Omega} (v \Delta u - u \Delta v) \, dx = \int_{\Gamma} (\partial_{\nu} u)v - u \partial_{\nu} v \, d\gamma \]
for all \( u, v \in H^2(\Omega) \), where the gradient terms have cancelled due to symmetry. Since \( u \in H^4(\Omega) \), it follows that \( \Delta u \in H^2(\Omega) \), so we may apply the above identity with \( u \) replaced by \( \Delta u \). Rearranging terms gives us
\[ \int_{\Omega} v \Delta^2 u \, dx = \int_{\Omega} \Delta u \Delta v \, dx - \int_{\Gamma} \Delta u \partial_{\nu} v \, d\gamma + \int_{\Gamma} (\partial_{\nu} \Delta u)v \, d\gamma \]
as stated. ■

Finally, we show that over the space \( H^2_0(\Omega) \) the \( L^2 \)-norm of the laplacian of a function is equal to the semi-norm.

**Corollary 3.3.4.** For all functions \( u \in H^2_0(\Omega) \) the following equality holds:
\[ \| \Delta u \|_0,\Omega = |u|_{2,\Omega}. \]  
(3.15)

**Proof.** This can be shown using a density argument. The set \( D(\Omega) \) is dense in \( H^2_0(\Omega) \) by definition. Furthermore, the maps \( u \mapsto |u|_{2,\Omega} \) and \( u \mapsto \| \Delta u \|_0,\Omega \) can be shown to be continuous with respect to \( \| \cdot \|_{2,\Omega} \). In view of [RY08,
3.4. The biharmonic equation

Corollary 1.29], it suffices to show that the equality holds on the dense subset $D(\Omega)$ of $H^2_0(\Omega)$.

Let $u \in D(\Omega)$. By expanding, we see that

$$\|\Delta u\|^2_{0,\Omega} = \int_\Omega \sum_i (\partial_i u)^2 \, dx + \int_\Omega \sum_{i \neq j} \partial_i u \partial_{jj} u \, dx,$$

$$|u|^2_{2,\Omega} = \int_\Omega \sum_i (\partial_i u)^2 \, dx + \int_\Omega \sum_{i \neq j} \partial_i u \partial_{jj} u \, dx,$$

where the diagonal terms have been extracted. By applying Proposition 3.3.1 twice, we have that

$$\int_\Omega \sum_{i \neq j} \partial_i \partial_{jj} u \, dx = -\int_\Omega \sum_{i \neq j} \partial_i u \partial_{jj} u \, dx = \int_\Omega \sum_{i \neq j} \partial_i u \partial_{jj} u \, dx,$$

where all the boundary integrals vanish due to $u$ being in $H^2_0(\Omega)$. Consequently, the two expressions $\|\Delta u\|^2_{0,\Omega}$ and $|u|^2_{2,\Omega}$ are equal over $D(\Omega)$, and the result follows by taking square roots and extending to $H^2_0(\Omega)$. ■

3.4 The biharmonic equation

The biharmonic equation is a fourth order partial differential equation arising in the study of continuum mechanics, for instance in the study of plate bending. In this section we study the variational formulation of the biharmonic equation and how this ties in to the theory of Sobolev spaces introduced above. In the following, $\Omega$ is an open bounded domain in $\mathbb{R}^n$ with a sufficiently smooth boundary $\Gamma$ and an outward unit normal $\nu = (\nu_1, \ldots, \nu_n)$. In its classical form, the biharmonic equation reads:

$$\begin{cases}
\Delta^2 u = f \text{ in } \Omega, \\
u = \partial_{\nu} u = 0 \text{ on } \Gamma.
\end{cases}$$

(3.16)

Here, $f$ is some given source function $f: \Omega \to \mathbb{R}$ and $u: \Omega \to \mathbb{R}$ is the unknown. We will refer to Equation (3.16) as the strong form of the biharmonic equation.
3. Sobolev Spaces and the Biharmonic Equation

The weak form

The weak form of a partial differential equation is a reformulation of the strong form in terms of test functions. We multiply the equation by a test function \( v \) in some function space of suitable functions \( V \)—which will be determined later—and integrate over the domain. This yields:

\[
\int_{\Omega} (\Delta^2 u) v \, dx = \int_{\Omega} f v \, dx,
\]  

(3.17)

and we are now interested in finding the \( u \) that satisfies Equation (3.17) for all \( v \in V \). In its current form, we require \( u \) to be four times differentiable, due to the \( \Delta^2 u \) term. We can greatly weaken this requirement by applying Corollary 3.3.3 and moving some of the derivatives onto \( v \), giving us the following:

\[
\int_{\Omega} \Delta u \Delta v \, dx - \int_{\Gamma} \Delta u \partial_{\nu} v \, d\gamma + \int_{\Gamma} (\partial_{\nu} \Delta u) v \, d\gamma = \int_{\Omega} f v \, dx.
\]

In order to encode the clamped boundary conditions for the biharmonic equation in this weak formulation of the problem, we need to be clever in the choice of function space \( V \). By letting \( V = H^2_0(\Omega) \) we are guaranteed that the boundary integrals vanish due to the construction of \( H^2_0(\Omega) \), as \( v = \partial_{\nu} v = 0 \) on the boundary \( \Gamma \). Furthermore, if a solution \( u \) exists in this space, it automatically satisfies the boundary conditions. We may now restate the biharmonic equation with boundary conditions in its weak form: Find \( u \in H^2_0(\Omega) \) such that

\[
\int_{\Omega} \Delta u \Delta v \, dx = \int_{\Omega} f v \, dx
\]

(3.18)

for all \( v \in H^2_0(\Omega) \). Note that this weak equation can be studied independently of the strong form of the governing partial differential equation.

Remark 3.4.1. In general, a solution \( u \) to the weak formulation in Equation (3.18) does not necessarily solve the strong form in Equation (3.16). If however, the weak solution \( u \) is sufficiently smooth to lie in the space \( H^4(\Omega) \cap H^2_0(\Omega) \), then by applying Corollary 3.3.3 backwards, we can deduce that it also solves the strong equation.
3.5 Existence and uniqueness of weak solutions

Having derived a suitable weak formulation of the biharmonic equation, the question now is: Under what circumstances can we expect the weak formulation to have a solution, and is this solution unique — if this is the case, we call the weak formulation well posed. In this section, we aim to provide an answer to this question. Before we begin, we recast the weak formulation of the equation into a more abstract setting, more suitable for the following analysis. Let $a: H^2_0(\Omega) \times H^2_0(\Omega) \to \mathbb{R}$ and $L: H^2_0(\Omega) \to \mathbb{R}$ be the bilinear and linear forms respectively defined by

$$a(u, v) := \int_{\Omega} \Delta u \Delta v \, dx,$$
$$L(v) := \int_{\Omega} fv \, dx.$$  

(3.19)

The abstract weak formulation of the biharmonic equation with clamped boundary conditions then reads: Find $u \in H^2_0(\Omega)$ such that

$$a(u, v) = L(v)$$

for all $v \in H^2_0(\Omega)$.

The Lax–Milgram theorem gives sufficient conditions for such a weak formulation to have a solution. In order to state it, we need the definition of ellipticity of an abstract operator.

**Definition 3.5.1 (V-ellipticity).** If $V$ is a normed vector space with associated norm $\| \cdot \|_V$, we say that a bilinear form $a: V \times V \to \mathbb{R}$ is $V$-elliptic (often referred to as $V$-coercive in the literature) if there exists a scalar $\kappa > 0$ such that

$$a(v, v) \geq \kappa \| v \|_V^2$$

(3.20)

for all $v \in V$.

With this definition in mind, we state the following important key result. A proof can be found in [Eva10, §6.2.1].

35
3. Sobolev Spaces and the Biharmonic Equation

**Theorem 3.5.2** (Lax–Milgram). If $V$ is a Hilbert space, and $a: V \times V \to \mathbb{R}$ is a bounded and $V$-elliptic bilinear form with ellipticity constant $\kappa$, then for any bounded functional $L: V \to \mathbb{R}$ there exists a unique $u \in V$ such that

$$a(u, v) = L(v)$$

for all $v \in V$. In addition, this unique solution satisfies the bound

$$\|u\|_V \leq \frac{1}{\kappa} \|L\|_{V^*} \tag{3.21}$$

where $V^*$ is the dual space of $V$ with associated norm.

We now have everything we need to show that the weak formulation of the biharmonic equation is well posed. We formulate this as a theorem.

**Theorem 3.5.3** (Well posedness of the weak biharmonic equation). Given a function $f \in L^2(\Omega)$ there exists a unique solution to the weak formulation of the biharmonic equation: Find $u \in H^2_0(\Omega)$ such that

$$\int_\Omega \Delta u \Delta v \, dx = \int_\Omega f v \, dx$$

for all $v \in H^2_0(\Omega)$.

**Proof.** We need to show that the bilinear form

$$a(u, v) = \int_\Omega \Delta u \Delta v \, dx \tag{3.22}$$

is both bounded and $H^2_0(\Omega)$-elliptic, and that

$$L(v) = \int_\Omega f v \, dx \tag{3.23}$$

is bounded, so that the abstract problem satisfies the hypothesis of the Lax–Milgram theorem.

**Boundedness of $a$:** By the Cauchy–Schwartz inequality, we have that

$$|a(u, v)| \leq \|\Delta u\|_{0,\Omega} \|\Delta v\|_{0,\Omega} = \|u\|_{2,\Omega} \|v\|_{2,\Omega} \leq \|u\|_{2,\Omega} \|v\|_{2,\Omega}, \tag{3.24}$$
3.5. Existence and uniqueness of weak solutions

where the first equality follows from Corollary 3.3.4, and the last inequality follows from the fact that the semi-norm is always smaller than or equal to the full norm. Consequently, we see that the bilinear form $a$ is bounded.

**Boundedness of $L$:** Since both $f$ and $v$ lie in $L^2(\Omega)$, an application of the Cauchy–Schwartz inequality yields

$$|L(v)| \leq \|f\|_{0,\Omega} \|v\|_{0,\Omega} \leq \|f\|_{2,\Omega} \|v\|_{2,\Omega}.$$  \hspace{1cm} (3.25)

We therefore see that $L$ is bounded.

**Ellipticity of $a$:** We have by definition,

$$a(v, v) = \int_{\Omega} \Delta v \Delta v \, dx = \|\Delta v\|_{0,\Omega}^2$$

which by Corollary 3.3.4 is equal to $|u|_{2,\Omega}^2$. In view of Remark 3.2.8, there exists a $\kappa$ such that $|u|_{2,\Omega}^2 \geq \kappa \|v\|_{2,\Omega}^2$, so it follows that

$$a(v, v) \geq \kappa \|v\|_{2,\Omega}^2$$

and consequently $a$ is $H^2_0(\Omega)$-elliptic.

Combining the three results above, we see that the weak biharmonic equation is well posed, and admits a unique solution in $H^2_0(\Omega)$. ■

Having concluded that the weak biharmonic equation admits a solution, we turn in the next chapter to the task of finding approximate solutions.
CHAPTER 4

The Finite Element Method

Having derived the weak formulation of the biharmonic equation in Chapter 3, we are ready to put the Finite Element Method to use. In this chapter we derive a finite element formulation for an abstract variational problem of the form: Find \( u \in V \) such that \( a(u, v) = L(v) \) for all \( v \in V \). We then derive a fundamental convergence result in finite element theory, known as Cea’s lemma, which tells us exactly when the finite element method converges. Furthermore, we discuss the systematic construction of a discrete function space, and under what conditions one can expect this space to be a proper subset of the set \( V \) and some of the finite element language is introduced.

We then focus our attention to the specifics of the biharmonic equation. We construct a suitable discrete function space based on \( C^1 \) continuous quadratic simplex splines over the PS12 of a triangle. By combining Cea’s lemma with interpolation results pertaining to the specific finite element space used, one can derive error bounds for the finite element solution in the \( H^2(\Omega) \)-norm. We will also employ the so-called Aubin-Nitsche duality argument to derive error estimates in the \( L^2(\Omega) \)-norm.

4.1 The finite element formulation

The function space \( V \) in the weak formulation is infinite dimensional. We wish to compute an approximation to the analytical solution by instead
4. The Finite Element Method

considering a finite dimensional subspace of $V$. We will in the following denote this subspace by $V_h \subset V$. Given such a subspace, we may restrict the bilinear and linear forms $a$ and $L$ to this subspace, and we therefore define the associated forms $a_h : V_h \times V_h \to \mathbb{R}$ and $L : V_h \to \mathbb{R}$ as the restrictions: $a_h := a|_{V_h \times V_h}$, and $L_h := L|_{V_h}$. The finite element formulation of the abstract problem is then: Find $u_h \in V_h$ such that

$$a_h(u_h, v_h) = L_h(v_h) \quad (4.1)$$

for all $v_h \in V_h$.

Well posedness of the finite element formulation

Having transitioned from the full space $V$ to the finite dimensional subspace $V_h$, the question is whether the finite element formulation is well posed in the sense of the Lax–Milgram lemma. Fortunately, this follows directly.

**Corollary 4.1.1.** Let $a : V \times V \to \mathbb{R}$ and $L : V \to \mathbb{R}$ be a bilinear and linear operator satisfying the hypothesis of Lax–Milgram (Theorem 3.5.2). Let $V_h$ be a finite dimensional subspace of $V$, and $a_h$, $L_h$ the corresponding bilinear and linear forms. Then, there exists exactly one $u_h \in V_h$ such that

$$a_h(u_h, v_h) = L_h(v_h) \quad (4.2)$$

for all $v_h \in V_h$.

**Proof.** Every finite dimensional subspace of a Hilbert space $V$ is closed, hence $V_h$ is closed. Any closed subspace of a Hilbert space is itself a Hilbert space with respect to the inherited norm, thus $V_h$ is a Hilbert space. The restriction $a_h$ is both bounded and $V_h$-elliptic, and $L_h$ is bounded. Consequently, $V_h$, $a_h$ and $L_h$ satisfy the hypothesis of the Lax–Milgram lemma in their own right. We can therefore conclude that the finite element formulation admits a unique solution. ■

40
4.2 Cea’s lemma

The following inequality is of fundamental importance in the Finite Element Method. The so called Cea’s lemma tells us that the solution to the finite element formulation is in some sense the best achievable. We start by proving an auxiliary result known as Galerkin orthogonality and then formulate Cea’s result as a theorem.

**Lemma 4.2.1** (Galerkin Orthogonality). Let $u$ be the solution to the weak formulation and $u_h$ the solution to the corresponding finite element formulation. Then the error $e := u - u_h$ is orthogonal to the finite dimensional subspace $V_h$ of $V$. That is,

$$ a(e, v_h) = 0 $$

for all $v_h \in V_h$.

**Proof.** By assumption, both $u$ and $u_h$ satisfy $a(u, v_h) = L_h(v_h) = a(u_h, v_h)$ for all $v_h \in V_h$. By subtracting the two equations, we have

$$ a(e, v_h) = a(u - u_h, v_h) = 0 $$

for all $v_h \in V_h$.

**Theorem 4.2.2** (Cea’s Lemma). Let $a$ be a bounded and $V$-elliptic bilinear form with bound $M$ and ellipticity constant $\kappa$. Furthermore, let $u$ be the solution to the abstract variational problem: Find $u \in V$ such that

$$ a(u, v) = L(v) $$

for all $v \in V$. Let $u_h \in V_h$ be the solution to the corresponding finite element formulation. Then, $u_h$ satisfies

$$ \|u - u_h\|_V \leq \frac{M}{\kappa} \inf_{v_h \in V_h} \|u - v_h\|_V $$

**Proof.** Using the $V$-ellipticity of $a$, we have that

$$ \kappa \|u - u_h\|_V^2 \leq a(u - u_h, u - u_h). $$
We add \( 0 = a(u - u_h, v_h - v_h) \) to the right hand side, and by using the bilinearity of \( a \), rearrange the right hand side as
\[
a(u - u_h, u - u_h) = a(u - u_h, u - u_h) + a(u - u_h, v_h - v_h) = a(u - u_h, u - v_h) + a(u - u_h, v_h - u_h).
\]

Since \( v_h - u_h \in V_h \), we have by Galerkin orthogonality, Lemma 4.2.1, that \( a(u - u_h, v_h - u_h) = 0 \). By boundedness of \( a \), it follows that
\[
\kappa \| u - u_h \|_{V_h}^2 \leq a(u - u_h, u - v_h) \leq M \| u - u_h \|_V \| u - v_h \|_V
\]
for all \( v_h \in V_h \). Since this is independent on which \( v_h \in V_h \) we choose, dividing both sides by \( \| u - u_h \|_V \), rearranging, and taking the infimum over \( V_h \) yields
\[
\| u - u_h \|_V \leq \frac{M}{\kappa} \inf_{v_h \in V_h} \| u - v_h \|_V
\]
as we wanted to show.

**Remark 4.2.3.** This result tells us that the approximate solution obtained by solving the finite element formulation, is up to a constant, the best possible. Note that the constant only depends on the bilinear operator \( a \).

## 4.3 Constructing a finite element space

The Finite Element Method encompasses a systematic way of constructing finite element spaces. We start by giving a general treatment, then we focus our attention on the specifics of the biharmonic equation, and the quadratic simplex splines on the Powell–Sabin 12-split. The construction of a finite element space is a two step process involving the description of local degrees of freedom for each element, and subsequently the construction of global degrees of freedom for the entire domain by identifying the local degrees of freedom.

In order to derive further properties of the finite element method, we need to be very precise in exactly what we mean by a finite element. Informally, we would like to associate to a geometric region, a set of
functions. These functions should be uniquely specifiable by the values of a certain set of linear functionals. The geometric region of interest is typically triangles and quadrangles when the spatial dimension is two, however the theory generalizes nicely to arbitrary spatial dimension.

We introduce the finite element as in [Gia02].

**Definition 4.3.1** (Finite Element). A finite element is a triple \((K, \mathcal{P}, \mathcal{N})\) where

1. \(K\) is a closed subset of \(\mathbb{R}^n\) with a Lipschitz-continuous boundary, called the element domain;
2. \(\mathcal{P}\) is a space of real valued functions defined over \(K\), called the set of shape functions; and
3. \(\mathcal{N}\) is a finite set of linearly independent functionals \(\Psi_i : \mathcal{P} \to \mathbb{R}\) with \(i = 1, \ldots, N\), that form a basis for the dual space \(\mathcal{P}^*\) of \(\mathcal{P}\), referred to as the degrees of freedom for \(\mathcal{P}\).

The requirement that \(\mathcal{N}\) is a basis for the dual space of \(\mathcal{P}\) gives rise to an explicit basis for \(\mathcal{P}\) which we will make much use of:

**Definition 4.3.2** (Nodal Basis). Let \((K, \mathcal{P}, \mathcal{N})\) be a finite element. The set of \(N\) functions \(B = \{\varphi_1, \ldots, \varphi_N\}\) satisfying

\[
\Psi_i(\varphi_j) = \delta_{ij}
\]

for \(i, j = 1, \ldots, N\) constitutes a basis for \(\mathcal{P}\), and is referred to as the nodal basis for \(\mathcal{P}\).

This means that we may specify a finite element in terms of the geometric region, a function space \(\mathcal{P}\) and a set of degrees of freedom, for which a set of nodal basis functions is implicitly defined through the degrees of freedom. When \(\mathcal{N}\) form a basis for the dual space of \(\mathcal{P}\), we say that \(\mathcal{N}\) is \(\mathcal{P}\)-unisolvent. The following lemma gives a sufficient condition for unisolvence.
4. The Finite Element Method

**Figure 4.1:** The diagram on the left describes an element where the shape functions are linear polynomials, and the degrees of freedom are the evaluation at the vertices. The diagram on the right corresponds to the degrees of freedom for the PS12 element, described in Definition 2.3.3.

**Lemma 4.3.3** ($\mathcal{P}$-unisolvence). Let $\mathcal{P}$ be a function space of dimension $N$, and denote by $\mathcal{P}^*$ its dual. Let $\mathcal{N} = \{\Psi_1, \ldots, \Psi_N\} \subseteq \mathcal{P}^*$. Then the following are equivalent:

1. $\mathcal{N}$ constitutes a basis for $\mathcal{P}^*$;
2. Given $f \in \mathcal{P}$ such that $\Psi(f) = 0$ for all $\Psi \in \mathcal{N}$, then $f = 0$.

**Remark 4.3.4** (Diagrammatic Representation of Finite Elements). The prescription of degrees of freedom for a finite element can be a tedious task. Fortunately, a diagrammatic convention for the description of finite elements have been developed, whereby the degrees of freedom easily can be read off.

A dot represents the evaluation at the corresponding point, a circle represents evaluation of the first partial derivatives at the point, a double circle the second partial derivatives, etc. Furthermore, an arrow denotes the evaluation of the directional derivative at the root of the arrow in the direction of the arrow. This is best shown by example, and some can be seen in Figure 4.1.
The local interpolation operator

We now turn to finite element interpolation. We associate to a finite element a local finite element interpolation operator. These will be used in the construction of $C^1$ quadratic finite element spaces, by piecing together finite elements and constructing a corresponding global finite element interpolation operator.

In our case, the set of degrees of freedom $\mathcal{N}$ involve directional derivatives up to first order, evaluated at certain points in the domain $K$. That is, the functionals $\Psi_i : \mathcal{P} \rightarrow \mathbb{R}$ are for example of the form

$$\Psi_i(f) = f(a_i^0), \text{ or } \Psi_i(f) = \nabla f(a_i^1) \cdot \xi_{ij} \quad (4.5)$$

where the points $a_i^r$ are referred to as the carrier of the corresponding functionals, and the vectors $\xi_{ij}$ are direction vectors. In any case, we let $m$ denote the highest order of derivatives occurring in the set of degrees of freedom (in the example above $m = 1$). Consequently, the degrees of freedom of a finite element can be applied to any function in $C^m(K)$, as they exhibit sufficient smoothness. In view of this, we define the following interpolation operator.

**Definition 4.3.5 (Local Interpolation Operator).** Let $(K, \mathcal{P}, \mathcal{N})$ be a finite element with $m$ being the highest degree of derivative occurring in the degrees of freedom. We define the $\mathcal{P}$-interpolation operator $\mathcal{I}_K : C^m(K) \rightarrow \mathcal{P}$ as

$$\mathcal{I}_K v := \sum_{i=1}^{N} \Psi_i(v) \varphi_i \quad (4.6)$$

where $\varphi_i$ are the nodal basis functions corresponding to the degrees of freedom.

The local interpolation operators will be of key importance when deriving error estimates for the finite element method. One important property of these local interpolation operators is that they preserve the nodal basis functions.
Lemma 4.3.6. Let \((K, P, N)\) be a finite element with associated interpolation operator \(I_K\). Then
\[
I_K \varphi_i = \varphi_i \tag{4.7}
\]
for all \(i = 1, \ldots, N\).

Proof. Since \(\Psi_j(\varphi_i) = \delta_{ij}\), it follows that
\[
I_K \varphi_i = \sum_{j=1}^{N} \Psi_j(\varphi_i) \varphi_j = \varphi_i. \tag{4.8}
\]

Corollary 4.3.7. The interpolation operator \(I_K\) preserves the shape functions, i.e.,
\[
I_K v = v \tag{4.9}
\]
for all \(v \in \mathcal{P}\).

Proof. Let \(v = \sum_{i=1}^{N} c_i \varphi_i\). Then
\[
I_K v = \sum_{j=1}^{N} \Psi_j(v) \varphi_j = \sum_{j=1}^{N} \sum_{i=1}^{N} c_i \Psi_j(\varphi_i) \varphi_j
= \sum_{i=1}^{N} c_i I_K \varphi_i = \sum_{i=1}^{N} c_i \varphi_i = v.
\]

Affine families of finite elements

We briefly digress by considering one specific type of finite elements. One common approach to the finite element method, is to construct families of so-called affine-equivalent finite elements by means of a fixed reference element, and a set of affine maps. Here, we discuss properties of such affine families, as we will be needing them later in finite element error estimation. Let \((\hat{K}, \hat{P}, \hat{N})\) be a finite element with fixed shape functions and degrees of freedom. Assume that the degrees of freedom are of the form given in Equation (4.5) where the points \(\hat{\alpha}_i\) are the carriers and \(\hat{\xi}_{ij}\)
are directional vectors. Let $K$ be some arbitrary triangle, related to $\hat{K}$ by an invertible affine map $F: \hat{K} \rightarrow K$ given by

$$F(\hat{x}) = B\hat{x} + c = x.$$ 

This affine map induces a correspondence between the functions and functionals over respective triangles known as the pull-back and the push-forward operations.

1. The pullback operator $F^*$ takes elements of $C^1(\hat{K})$ to elements of $C^1(K)$ by

$$F^*(\hat{v}) = \hat{v} \circ F^{-1}.$$ 

2. The pushforward operator $F_*$ takes elements of the dual space $(C^1(\hat{K}))^*$ to elements of the dual space $(C^1(K))^*$ by

$$F_*(\hat{\Psi}) = \hat{\Psi} \circ (F^*)^{-1} = \hat{\Psi}(v \circ F). \quad (4.10)$$

We are now in position to prescribe a set of shape functions and degrees of freedom for the triangle $K$ as follows:

1. Define the set $\mathcal{P}$ in terms of $F$ and $\hat{P}$ as

$$\mathcal{P} = \left\{ v := F^*(\hat{v}) = (\hat{v} \circ F^{-1}) \text{ for all } \hat{v} \in \hat{P} \right\}.$$ 

That is, $v(x) = \hat{v}(F^{-1}(x)) = \hat{v}(\hat{x})$.

2. Define the degrees of freedom $\mathcal{N}$ as

$$\mathcal{N} = \left\{ \Psi := F_*(\hat{\Psi}) \text{ for all } \hat{\Psi} \in \hat{\mathcal{N}} \right\}. \quad (4.11)$$

That is, the points $a^r_i = F(\hat{a}^r_i)$ are the carriers of $\mathcal{N}$, and the vectors $\xi_{ij} = B\hat{\xi}_{ij}$ are the corresponding directional vectors.

The triple $(K, \mathcal{P}, \mathcal{N})$ then constitutes a finite element, and due to its construction we say that $(K, \mathcal{P}, \mathcal{N})$ is affine equivalent to $(\hat{K}, \hat{P}, \hat{\mathcal{N}})$. 47
4. The Finite Element Method

Remark 4.3.8. If $\hat{\varphi}_i$ are the nodal basis functions corresponding to the reference element, then the functions $\varphi_i = \hat{\varphi}_i \circ F^{-1}$ are the nodal basis functions corresponding to the affine equivalent element. Furthermore, by a direct computation, it can be seen that

$$\hat{I}v = \hat{I}\hat{v},$$

that is - it does not matter whether the function is transformed then interpolated, or whether the function is interpolated then transformed.

Finite element spaces

By representing our domain of interest as a disjoint union of element domains, and on each element domain form a finite element, with its associated shape functions and degrees of freedom, we may piece together the local interpolants in such a way that a global function space is formed. We specialize our attention to a subdivision of the domain $\Omega$ in the case where $\Omega$ is a polygonal domain in $\mathbb{R}^2$ and the element domains are triangles.

Remark 4.3.9. By a \textit{polygonal domain}, we mean a domain $\Omega$ whose boundary $\Gamma$ is comprised of line segments. That is, the domain is piecewise linear. This ensures that the domain can be represented exactly in a triangulation as a union of disjoint triangles.

We make precise what we mean by such a subdivision:
4.3. Constructing a finite element space

**Figure 4.2:** The subdivision on the left forms a triangulation, while the one on the right does not, as it violates the third requirement of Definition 4.3.10

**Definition 4.3.10** (Triangulation). Let $\Omega$ be a polygonal domain. A *triangulation* $T = \{K_i\}_{i=1}^M$ of $\Omega$ is a finite collection of triangles (element domains) such that:

1. $\text{int}(K_i) \cap \text{int}(K_j) = \emptyset$ when $i \neq j$;
2. $\overline{\Omega} = \bigcup_{i=1}^N K_i$; and
3. no vertex of any triangle lies in the interior of an edge or any other triangle.

To a triangulation we associate the sets $\mathcal{V}$, $\mathcal{E}$ and $\mathcal{F}$, which denote the collections of vertices, edges and faces in the triangulation respectively. We let $|\mathcal{V}|$, $|\mathcal{E}|$ and $|\mathcal{F}|$ denote the cardinality of respective sets.

In view of this definition, Figure 4.2 shows one example of a subdivision that is a triangulation, and one that is not.

We also define the following properties of a triangulation, which will be needed later.

**Definition 4.3.11.** For a single triangle $K$, $h_K$ denotes the largest side length, and $\rho_K$ denotes the diameter of inscribed circle of $K$. We say that a triangulation $T$ is *regular* if there exists a constant $\sigma$ such that

$$\frac{h_K}{\rho_K} \leq \sigma \quad (4.12)$$

for all triangles $K \in T$. 

49
Assume we are given a triangulation $\mathcal{T}$ of a polygonal domain $\Omega$. Furthermore, assume that to each element domain, we have associated a finite element $(K, \mathcal{P}_K, \mathcal{N}_K)$. Let $m$, as in the preceding section, denote the highest order of derivative occurring in the degrees of freedom for all the elements and recall that on each element we have a local interpolation operator $\mathcal{I}_K: C^m(K) \to \mathcal{P}_K$. By piecing together these local interpolants, we may define a global interpolant over the whole domain:

**Definition 4.3.12** (Global Interpolation Operator). We define for $f \in C^m(\Omega)$ the **global interpolation operator** $\mathcal{I}$ as

$$\mathcal{I}f|_K := \mathcal{I}_K f.$$  (4.13)

Equipped with this global interpolation operator, we have everything we need to define a finite element space.

**Definition 4.3.13** (Finite Element Space). Let $\mathcal{T}$ be a triangulation of the polygonal domain $\Omega$. Let a global interpolation operator $\mathcal{I}$ be defined for all $f \in C^m(\Omega)$. We define the corresponding **finite element space** $V_\mathcal{T}$ as

$$V_\mathcal{T} := \{ \mathcal{I}f : f \in C^m(\Omega) \},$$  (4.14)

and we say that the interpolant $\mathcal{I}$ has **continuity order** $r$ if $\mathcal{I}f \in C^r(\Omega)$ for all $f \in C^m(\Omega)$. If this is the case, the inclusion $V_\mathcal{T} \subseteq C^r(\Omega)$ holds, and we call $V_\mathcal{T}$ a $C^r$ finite element space.

### 4.4 Powell–Sabin 12-splits and a $C^1$ finite element space

In this section we use the spline spaces discussed in Chapter 2 in the construction of a $C^1$ finite element space. As always, let $\mathcal{T}$ be a triangulation of a polygonal domain $\Omega$. Recall that for a triangle $K$, the spline space $S^1_{12}(K_{PS12})$ consists of piecewise quadratic polynomials with $C^1$ smoothness. We will on each triangle $K$ in the triangulation, construct a finite element where the space of shape functions is $S^1_{2}(K_{PS12})$, and the degrees of freedom
4.4. Powell–Sabin 12-splits and a $C^1$ finite element space

are the functionals specified in Definition 2.3.3 on page 19 and shown in Figure 4.1. We now formally define the element we will be employing for the approximate solution of the biharmonic equation, and look at some of its properties.

**Definition 4.4.1** (The Powell–Sabin Element). Let $K$ be a triangle, and let $S^1_2(\overline{K}_{PS12})$ be the set of piecewise $C^1$ quadratics on $K_{PS12}$. Let $N_{PS12}$ be the set of twelve Hermite degrees of freedom from Definition 2.3.3. Then the triple

\[(K, S^1_2(\overline{K}_{PS12}), N_{PS12})\]

constitutes a finite element, which we refer to as a **PS12 finite element**. The nodal basis corresponding to the degrees of freedom are the twelve Hermite nodal basis functions $H_i$ defined in Definition 2.3.3 and shown in Figure 2.5.

This element is what we will use to construct a $C^1$ finite element space. We define its local interpolant as in Definition 4.3.5. Note that the highest degree of derivative occurring in the degrees of freedom is $m = 1$.

**Definition 4.4.2** (The Powell–Sabin Local Interpolant). Let

\[(K, S^1_2(\overline{K}_{PS12}), N_{PS12})\]

be a PS12 finite element, we denote by $I^K_{PS12} : C^1(K) \rightarrow S^1_2(\overline{K}_{PS12})$ the local interpolant defined by

\[I^K_{PS12} f := \sum_{i=1}^{12} \Psi_i(f) H_i\]

(4.16)

for all $f \in C^1(K)$, where $\Psi_i \in N_{PS12}$.

**A global basis for $S^1_2(\mathcal{T}_{PS12})$**

In this section, we discuss how we may construct a set of basis functions for the whole space $S^1_2(\mathcal{T}_{PS12})$ by identifying local basis functions on adjacent elements. Since we require a global $C^1$ continuity across the whole triangulation, some care has to be exercised when identifying the local basis

51
functions related to the normal derivatives at edge midpoints, otherwise no such continuity can be assumed to hold. We give an example for which the triangulation consists of two triangles. The general approach can be inferred from this.

**Example 4.4.3** (Global basis functions over two triangles). It suffices to consider the case where the triangulation $\mathcal{T} = \{K_1, K_2\}$ consists of two triangles $K_1 = [a_1, a_2, a_3]$ and $K_2 = [a_1, a_2, \tilde{a}_3]$ with a common edge $e = [a_1, a_2]$ and outward facing unit normals $\mathbf{n}_3$ and $\mathbf{\hat{n}}_3$ to $e$ respectively.

We want to construct a set of basis functions $\phi_i : K_1 \cup K_2 \to \mathbb{R}$ for the space $S^1_2(\mathcal{T}_{PS12})$.

Let $\mathcal{H}_1^1, \ldots, \mathcal{H}_1^{12}$ be the local basis functions over $K_1$ and $\mathcal{H}_2^1, \ldots, \mathcal{H}_2^{12}$ the local basis functions over $K_2$ numbered as in Definition 2.3.3. Figure 4.3 shows the numbering of the basis functions corresponding to the common edge. We start by considering a global basis function corresponding to the evaluation at the vertex $a_1$. Define $\phi_1 : K_1 \cup K_2 \to \mathbb{R}$ by

$$
\phi_1(x) = \begin{cases} 
\mathcal{H}_1^1(x), & \text{if } x \in K_1, \\
\mathcal{H}_2^1(x), & \text{if } x \in K_2.
\end{cases}
$$

We can easily verify that this basis function is $C^1$ continuous, as by construction, the derivatives of $\phi_1$ vanish at the point $a_1$ (recall how the nodal basis functions were defined, satisfying $\Psi_i(\mathcal{H}_j) = \delta_{ij}$). Similarly, we define

$$
\phi_i(x) = \begin{cases} 
\mathcal{H}_1^i(x), & \text{if } x \in K_1, \\
\mathcal{H}_2^i(x), & \text{if } x \in K_2,
\end{cases}
$$

for $i = 2, 3, 5, 6, 7$, and these are all seen to be continuous. This takes care of the basis functions corresponding to shared vertices, whose support is the union of $K_1$ and $K_2$. Consider now the function $\phi_4$ corresponding to the shared edge, defined by

$$
\phi_4(x) = \begin{cases} 
\mathcal{H}_1^4(x), & \text{if } x \in K_1, \\
-\mathcal{H}_2^4(x), & \text{if } x \in K_2.
\end{cases}
$$
and make specific note of the negative coefficient. This ensures that the cross boundary derivative of $\phi_4$ is continuous across the common edge. This is illustrated in Figure 4.4. Furthermore, define

\[
\phi_i(x) = \begin{cases} 
\mathcal{H}_1^i(x), & \text{if } x \in K_1, \\
0, & \text{otherwise}
\end{cases} \quad \phi_{i+5}(x) = \begin{cases} 
\mathcal{H}_2^i(x), & \text{if } x \in K_2, \\
0, & \text{otherwise}
\end{cases}
\]

for $i = 8, 9, 10, 11, 12$. These are the global basis functions corresponding to degrees of freedom whose carriers are not shared between the two triangles. This gives us a total of 17 global basis functions, and consequently $\dim(S_{12}^1(T_{PS12})) = 17$.

\[\text{Figure 4.3: The numbering used for the local basis functions in Example 4.4.3.}\]

Since we have three global nodal basis functions per vertex, and one global nodal basis function per edge, we can deduce that in general, the dimension of the spline space is

\[\dim(S_{12}^1(T_{PS12})) = 3|V| + |E|\]

where $V$ and $E$ denotes the set of vertices and edges in the triangulation $T$. We denote the global basis functions by $\phi_1, \ldots, \phi_N$. An example of some
global basis functions can be seen in Figure 4.5 along with the associated mesh.

By [Cia02, Theorem 2.1.2], it follows directly that the inclusions

\[ V_h := S^1_2(T_{PS12}) \subset H^2(\Omega), \]
\[ V_{ho} := \{ s \in S^1_2(T_{PS12}) : s = \partial_\nu s = 0 \text{ on } \Gamma \} \subset H^2_0(\Omega) \]

hold, where \( \nu \) is the outward unit normal to the boundary \( \Gamma \). Consequently, in view of earlier discussion, we will be employing the space \( V_{ho} \) in the solution of the biharmonic equation with homogeneous boundary conditions.

\[ \text{FIGURE 4.4: The figure on the left demonstrates a global nodal function that is not in } C^1(\Omega) \text{ as the local nodal basis functions have been stitched together wrongly on two adjacent triangles. The figure on the right shows the situation where the signs of the two basis functions are opposite, yielding a } C^1(\Omega) \text{ global nodal basis function.} \]

4.5 Interpolation error for \( S^1_2(T_{PS12}) \)

Let \( V_{ho} \subset H^2_0(\Omega) \) used for solving the biharmonic equation be the finite element space in question. We are interested in finding a bound for the approximation error \( e = u - u_h \) in the finite element method. By considering interpolation onto the finite element space, we may derive
suitable error estimates. In view of Cea’s lemma (Theorem 4.2.2), we know that the error is bounded as

\[ \|u - u_h\|_{2,\Omega} \leq \inf_{v_h \in V_h} \|u - v_h\|_{2,\Omega} \]  

(4.17)

By interpolating \( u \) onto \( V_{ho} \), we have the following inequality:

\[ \inf_{v_h \in V_{ho}} \|u - v_h\|_{2,\Omega} \leq \|u - \mathcal{I}_{PS12} u\|_{2,\Omega}. \]  

(4.18)

The question at hand is then whether it is possible to find an estimate for the interpolation error \( \|u - \mathcal{I}_{PS12} u\|_{2,\Omega} \).

**An interpolation estimate for affine families of elements**

Central to interpolation estimates for the finite element method is the notion of affine equivalent elements as introduced in Section 4.3, where
the nodal basis and degrees of freedom of one element is related to the nodal basis and degrees of freedom over a reference element through an affine map $F$. For such affine families of elements, it is possible to derive interpolation estimates on arbitrary triangles where the bound is dependent on the affine map and the reference element. While our element $(K, S^1_2(K_{PS12}), N_{PS12})$ turns out to not be affine equivalent, we will still find use of the results pertaining to such affine families in the derivation of error bounds for our specific element.

We start by proving a purely geometrical result that relates maps $F$ and $F^{-1}$ to the geometrical properties of the respective triangles.

**Lemma 4.5.1.** Let $\hat{K}$ and $K$ be related through the invertible affine map $F: \hat{K} \rightarrow K$ defined by

$$F(\hat{x}) = B\hat{x} + c. \quad (4.19)$$

Then the following bounds hold:

$$\|B\| \leq \frac{h}{\hat{h}} \quad \|B^{-1}\| \leq \frac{\hat{h}}{\rho}, \quad (4.20)$$

where $h$ and $\hat{h}$ are the largest side lengths; and $\rho$ and $\hat{\rho}$ are the diameters of the inscribed circles of triangles $K$ and $\hat{K}$ respectively.

**Proof.** We show the bound for $\|B^{-1}\|$ only and note that the other bound is completely analogous. By definition, the operator norm is given by

$$\|B^{-1}\| = \sup_{\|x\|=1} \|B^{-1}x\| = \frac{1}{\rho} \sup_{\|x\|=1} \|B^{-1}\rho x\| = \frac{1}{\rho} \sup_{\|z\|=\rho} \|B^{-1}z\|. \quad (4.21)$$

obtained by setting $z = \rho x$. Now, by definition of $\rho$, for any vector $z$ with $\|z\| = \rho$ there exists two points $x$ and $y$ in $K$, such that $z = x - y$. Note that

$$\|B^{-1}z\| = \|B^{-1}(x - c) - B^{-1}(y - c)\| = \|\hat{x} - \hat{y}\| \leq \hat{h} \quad (4.22)$$

obtained by adding $0 = B^{-1}c - B^{-1}c$ and observing that since both $\hat{x}$ and $\hat{y}$ are elements of $\hat{K}$, their distance cannot exceed that of $\hat{h}$. Consequently

$$\|B^{-1}\| \leq \frac{\hat{h}}{\rho}. \quad (4.23)$$
4.5. Interpolation error for $S_2^1(T_{PS12})$

At the heart of the error estimates for affine families of finite elements, is the relation between Sobolev seminorms of functions on one element in terms of the reference element. By bounding the derivatives occurring in the definition of the seminorms, and applying the transformation rule for integrals:

$$\int_K v \, dx = \int_{\hat{K}} (\hat{v} \circ F^{-1}) \| \det(B^{-1}) \| \, d\hat{x}. \tag{4.24}$$

it is possible to derive the next result which is of key importance [Cia02, Theorem 3.1.2].

**Theorem 4.5.2.** Let $\hat{K}$ and $K$ be two triangles related through the invertible affine map $F: \hat{K} \rightarrow K$ defined by

$$F(\hat{x}) = B\hat{x} + c. \tag{4.25}$$

Then, there exists a constant $C$ only dependent on $m$, such that for all $v$ in $H^m(K)$:

$$|\hat{v}|_{m,\hat{K}} \leq C\|B\|^m |\det(B)|^{-1/2} |v|_{m,K} \tag{4.26}$$

and for all $\hat{v}$ in $H^m(\hat{K})$:

$$|v|_{m,K} \leq C\|B^{-1}\|^m |\det(B)|^{1/2} |\hat{v}|_{m,\hat{K}} \tag{4.27}$$

Furthermore, there is a result pertaining to polynomial preserving interpolation operators on Sobolev spaces.

**Lemma 4.5.3.** Assume that $H^{k+1}(K) \hookrightarrow H^m(K)$ is continuously embedded, and let $I$ be a continuous interpolation operator that preserves polynomials of degree $\leq k$:

$$Iv = v \tag{4.28}$$

for all $v \in \Pi_k(K)$. Then, there exists a constant $C$ depending only on $K$ and $I$ such that

$$|v - Iv|_{m,K} \leq C|v|_{k+1,K} \tag{4.29}$$

for all $v \in H^{k+1}(K)$. 

57
4. The Finite Element Method

With these results in mind, it is now possible to derive an interpolation result for an finite element affinely equivalent to some given reference element. Indeed, the following estimate holds [Cia02, Theorem 3.1.5]:

**Theorem 4.5.4 (Affine Interpolation Estimate).** Let $(\hat{K}, \hat{P}, \hat{N})$ be a reference finite element with $s$ being the highest order of derivative occurring in the degrees of freedom. If the following inclusions hold for integers $m, k > 0$:

\[
H^{k+1}(\hat{K}) \hookrightarrow C^s(\hat{K}),
\]

\[
H^{k+1}(\hat{K}) \hookrightarrow H^m(\hat{K}),
\]

\[
\Pi_k(\hat{K}) \subset \hat{P} \subset H^m(\hat{K}),
\]

then there exists a constant $C$ depending only on the reference element, such that for all affine equivalent elements $(K, P, N)$ and functions $v \in H^{k+1}(K)$:

\[
|v - \mathcal{I}_K v|_{m,K} \leq C \frac{h_{k+1}^{k+1} \rho_K^m}{\rho_K} |v|_{k+1,K}. \tag{4.30}
\]

**Proof.** We start by noting that the inclusion $\Pi_k(\hat{K}) \subset \hat{P}$ ensures that the corresponding local interpolation operator $\hat{I}$ replicates polynomials of degree $\leq k$. Pairing this with the inclusion $H^{k+1}(\hat{K}) \hookrightarrow H^m(\hat{K})$ we may apply Lemma 4.5.3 to deduce that there exists a constant $C_1$ such that

\[
|\hat{v} - \hat{I} \hat{v}|_{m,\hat{K}} \leq C_1 |\hat{v}|_{k+1,\hat{K}}.
\]

From Theorem 4.5.2 we know that the factor $|\hat{v}|_{k+1,\hat{K}}$ can be bounded

\[
|\hat{v}|_{k+1,\hat{K}} \leq C_2 \|B\|^{k+1} |\det(B)|^{-1/2} |v|_{k+1,K}
\]

and that

\[
|v - \mathcal{I} v|_{m,K} \leq C_3 \|B^{-1}\|^{m} |\det(B)|^{1/2} |\hat{v} - \hat{I} \hat{v}|_{m,\hat{K}}.
\]

By chaining the three above inequalities, we obtain

\[
|v - \mathcal{I} v|_{m,K} \leq C \|B^{-1}\|^{m} \|B\|^{k+1} |v|_{m,K}
\]
4.5. Interpolation error for $S_2^1(T_{PS12})$

By applying Lemma 4.5.1, we have that

$$|v - I v|_{m,K} \leq C h^{k+1} |v|_{k+1,K},$$  \hspace{1cm}  (4.31)

where the quantities $\hat{\rho}$ and $\hat{h}$ were absorbed into the constant $C$. \hfill ■

A direct consequence of extra assumed regularity on the triangles is the following, where the parameter $\sigma$ is contained in the constant $C$.

**Corollary 4.5.5.** If the family of elements are regular (Definition 4.3.11), then under the same hypothesis as in Theorem 4.5.4 there exists a constant $C$ such that for all $v \in H^{k+1}(K)$:

$$|v - I_K v|_{m,K} \leq C h^{k+1-m} |v|_{k+1,K}.\hspace{1cm} (4.32)$$

**An interpolation estimate for a non-affine element**

We now turn to deriving a similar error estimate for the non-affine finite element $(K, S_2^1(K_{PS12}), N_{PS12})$. Denote for brevity the interpolation operator $I_{PS12}$ by $I$.

We start by constructing a set of degrees of freedom for an element that is in affine correspondence with a reference element. We will use this set of degrees of freedom as a stepping stone for obtaining our interpolation estimate. Let $\Xi = \{\lambda_1, \ldots, \lambda_{12}\}$ be a set of degrees of freedom defined as follows:

1. $\lambda(v) = v(a_i)$: evaluation at the vertices $a_i$ of $K$, for $i = 1, 2, 3$;

2. $\lambda(v) = \nabla v(a_i) \cdot \xi_{ij}$: directional derivatives at the vertices in the direction $\xi_{ij} = a_j - a_i$ for $1 \leq i, j \leq 3$ and $i \neq j$.

3. $\lambda(v) = \nabla v(b_i) \cdot r_i$: directional derivatives at the edge midpoints in the direction $r_i = b_i - a_i$, where $b_i$ is the midpoint of the edge not containing $a_i$ (see Figure 4.6).

The directional vectors in $\Xi$ corresponding to the point $a_1$ is shown in Figure 4.6, the rest is constructed similarly.
4. The Finite Element Method

Figure 4.6: Some of the degrees of freedom in $\Xi$, constructed from the geometry of the triangle. Note that finite elements with this type of degrees of freedom do not assemble to form a $C^1$ space.

Lemma 4.5.6. The element $(K, S_1^1(K_{PS12}), \Xi)$ is affine equivalent to a reference element $(\hat{K}, S_1^1(\hat{K}_{PS12}), \hat{\Xi})$ under the affine map $F: \hat{K} \rightarrow K$.

Proof. In view of Section 4.3 we show that the degrees of freedom $\Xi$ are equal to the pushforward of the degrees of freedom $\hat{\Xi}$. For the evaluation at the vertices, we note that

$$v(a_i) = \hat{v}(\hat{a}_i)$$

for $i = 1, 2, 3$. For the directional derivatives, we need to examine how the gradient is transformed under $F$. We have that

$$\nabla v(a_i) \cdot (a_j - a_i) = \nabla v(F(\hat{a}_i)) \cdot B(\hat{a}_j - \hat{a}_i)$$

Using the fact that for an affine map $B = \nabla F(\hat{a}_i)$ it follows that

$$= \nabla v(F(\hat{a}_i)) \nabla F(\hat{a}_i) \cdot (\hat{a}_j - \hat{a}_i)$$

$$= \nabla v(F(\hat{a}_i)) \nabla F(\hat{a}_i) \cdot (\hat{a}_j - \hat{a}_i)$$

$$= \nabla \hat{v}(\hat{a}_i) \cdot (\hat{a}_j - \hat{a}_i).$$
4.5. Interpolation error for $S_2^1(T_{PS12})$

Consequently, all the degrees of freedom of $\Xi$ are equivalently defined in terms of $\hat{\Xi}$, and $(K, S_2^1(K_{PS12}), \Xi)$ is therefore affine equivalent to $(\hat{K}, S_2^1(\hat{K}_{PS12}), \hat{\Xi})$.

Let $(K, S_2^1(K_{PS12}), \Xi)$ be the finite element from the lemma above. As proven, this element is affinely equivalent to some reference element. Note that the same element domain and shape functions are used for both this auxiliary element, and our PS12-element. Let its associated interpolation operator be denoted $\Lambda$.

The goal is to obtain a bound for the estimate $|v - \mathcal{I}v|_{m,K}$ and the derivation relies on the following observation. Using the triangle inequality we have that

$$|v - \mathcal{I}v|_{m,K} \leq |v - \Lambda v|_{m,K} + |\mathcal{I}v - \Lambda v|_{m,K}. \quad (4.33)$$

Since $(K, S_2^1(K_{PS12}), \Xi)$ is affine-equivalent to some reference element, we already have a bound for this first term under the conditions of Theorem 4.5.4. It therefore suffices to find a bound for the second term.

By arguing as in Lemma 2.3.4, it is seen that $\Xi$ is $S_2^1(K_{PS12})$-unisolvent. We start by showing that the interpolation difference $\mathcal{I}v - \Lambda v$ can be bounded in the $|\cdot|_{2,K}$-norm. We will use the following special case of [Cia02, Theorem 3.1.5] to obtain one of the bounds:

**Lemma 4.5.7.** For an affine-equivalent finite element $(K, S_2^1(K_{PS12}), \Xi)$ there exists a constant $C$ only dependent on the reference element, such that for all functions $v \in H^3(K)$,

$$\max_{i=1,2} \left\{ \text{ess sup}_{x \in K} \left| \partial_i(v - \Lambda v)(x) \right| \right\} \leq C \text{area}(K)^{-1/2} h_K^3 \rho_K |v|_{3,\Omega}. \quad (4.34)$$

**Lemma 4.5.8.** Under the assumptions of Theorem 4.5.4, with $m = 2$, $k = 2$ and $s = 1$ there exists regular finite element a constant $C$ only depending on the reference element, such that

$$|\mathcal{I}v - \Lambda v|_{2,K} \leq Ch_K |v|_{3,K} \quad (4.35)$$

for all $v \in H^3(K)$. 61
4. The Finite Element Method

Proof. As before, we let $\Psi_i$ and $\mathcal{H}_i$ denote the degrees of freedom and nodal basis functions for the PS12-element. Let $L_i$ be the nodal basis functions corresponding to the degrees of freedom $\lambda_i$ in $\Xi$. We wish to express the interpolation difference $\Theta = I v - \Lambda v$ as a linear combination of these $L_i$. Since $\Theta$ is an element of $S^1_1(K_{PS12})$, from Corollary 4.3.7 we can write

$$\Theta = \Lambda \Theta = \sum_{i=1}^{12} \lambda_i(\Theta) L_i.$$  \hfill (4.36)

It therefore suffices to determine the coefficients $\lambda_i(\Theta)$. We examine the three types of degrees of freedom in question in turn.

**Values at the vertices**  Note that $\Psi_1 = \lambda_1$, and consequently

$$\lambda_1(\Theta) = \lambda_1(I v - \Lambda v) = \lambda_1(v) - \lambda_1(v) = 0.$$  \hfill (4.37)

The same argument shows that $\lambda_5(\Theta) = \lambda_9(\Theta) = 0$.

**Directional derivatives at the vertices**  Assume that the degrees of freedom corresponding to derivatives at the first vertex are numbered as follows:

$$\Psi_2(v) = \partial_x v(a_1) \quad \Psi_3(v) = \partial_y v(a_1) \quad \lambda_2(v) = \nabla v(a_1) : \xi_{12} \quad \lambda_3(v) = \nabla v(a_1) : \xi_{13}$$

Let $\Psi = (\Psi_2, \Psi_3)^T$ and $\lambda = (\lambda_2, \lambda_3)^T$. By definition, these are related through the relation $\lambda = A \Psi$ where $A = (\xi_{12}, \xi_{13})^T$ is invertible as long as the vectors are not parallel, which is true so long as $K$ is non-degenerate. Similarly, $\Psi = A^{-1} \lambda$. We then have

$$\lambda(\Theta) = A \Psi(I v - \Lambda v) = A \Psi(v - \Lambda v) = AA^{-1} \lambda(v - \Lambda v) = \lambda(v - v) = 0.$$  

A similar argument shows that $\lambda_6(\Theta) = \lambda_7(\Theta) = \lambda_{10}(\Theta) = \lambda_{11}(\Theta) = 0$.

**Cross boundary derivative**  Combining the two steps from above, we see that both the component of the gradient along the edges and the value
of $\Theta$ is zero at the vertices. Since $\Theta$ is a spline in $S^1_2(K_{PS12})$, we can deduce that $\Theta(b_1) = 0$ and that $\nabla \Theta(b_1) \cdot \xi_{12}$ is zero. Note that $\xi_{12}$ is perpendicular to the outward unit normal $n_3$. We can therefore write the directional vector $r_3$ in terms of the normal vector $n_3$ and the vector $\xi_{12}$:

$$r_3 = (r_3 \cdot n_3)n_3 + (r_3 \cdot \xi_{12})\xi_{12}/\|\xi_{12}\|.$$ 

Recall that $\lambda_4(\Theta) = \nabla \Theta(b_3) \cdot r_3$. It then follows that

$$\lambda_4(\Theta) = (r_3 \cdot n_3)\nabla \Theta(b_3) \cdot n_3 + (r_3 \cdot \xi_{12})\nabla \Theta(b_3) \cdot \xi_{12}/\|\xi_{12}\|$$

$$= (r_3 \cdot n_3)\nabla \Theta(b_3) \cdot n_3$$

$$= (r_3 \cdot n_3)\Psi_4(v - \Lambda v).$$

Similar arguments hold for the other two degrees of freedom corresponding to cross boundary derivatives, $\lambda_8$ and $\lambda_{12}$.

By the analysis done above, we can now write

$$\Theta = \sum_{i=4,8,12} (r_i \cdot n_i)\Psi_i(v - \Lambda v)L_i$$

(4.38)

where the vectors $r_i$ and $n_i$ are the vectors corresponding to the functionals in question.

**Bound in the Sobolev seminorm** We are interested in the seminorm of $\Theta$, and since $(r_i \cdot n_i)$ and $\Psi_i(v - \Lambda v)$ are scalars, we can write

$$|\Theta|_{2,K} = \sum_{i=4,8,12} |r_i \cdot n_i||\Psi_i(v - \Lambda v)||L_i|_{2,K}.$$ 

(4.39)

We can bound these terms separately.

1. Firstly, note that

$$r_i \cdot n_i = \|r_i\||n_i||\cos(\theta_i),$$

where $\theta_i$ is the angle between the two vectors. Since $n_i$ is of unit length, and $r_i$ cannot, by construction, have magnitude larger than the longest edge of $K$, it follows that $|r_i \cdot n_i| \leq h_K$. 

63
4. The Finite Element Method

2. Secondly, we have that

$$|Ψ_i(v − Λv)| = |∇(v − Λv)(b_i) · n_i|. $$

By writing the dot product in terms of the angle, as above, we have that

$$|∇(v − Λv)(b_i) · n_i| \leq ∇(v − Λv)(b_i)(b_i) · n_i| \leq \sqrt{2} \max_{i=1,2} \left\{ \text{ess sup}_{x \in K} |∇_i(v − Λv)(x)| \right\}. $$

Invoking Lemma 4.5.7, we obtain the bound

$$|∇(v − Λv)(b_i) · n_i| \leq C_1 \left( \frac{\text{area}(K)}{\rho_K} \right)^{1/2} h_{3,K}^3 |v|_{3,K}. \quad (4.40)$$

3. Since \((K, S^1(K_{PS12}), Ω)\) is affine equivalent to the reference element, we may apply Theorem 4.5.2 to obtain the bound

$$|L_i|_{2,K} \leq C_2 \left\| B^{-1} \right\|^2 \text{det}(B)|L_i|_{2,K}. \quad (4.41)$$

Note that \(|\text{det}(B)| = \text{area}(K)/\text{area}(\hat{K})\), and that by Lemma 4.5.1 \(\|B^{-1}\| \leq \hat{h}/\rho_K\). This yields

$$|L_i|_{2,K} \leq C_2 \frac{\hat{h}^2}{\rho_K^2} \frac{\text{area}(K)^{1/2}}{\text{area}(\hat{K})^{1/2}} |L_i|_{2,K}. \quad (4.42)$$

By moving all quantities related to the reference element (anything with a hat) into the constant \(C_2\), we obtain

$$|L_i|_{2,K} \leq C_2 \frac{\text{area}(K)^{1/2}}{\rho_K^2} |L_i|_{2,K}. \quad (4.43)$$

Taking \(C_3 = \max_i |L_i|_{2,K}\) and plugging these three bounds into Equation (4.39), we obtain

$$|Θ|_{2,K} \leq C_1 C_2 C_3 \frac{h_{3,K}^4}{\rho_K^3} |v|_{3,K} = C \frac{h_{3,K}^4}{\rho_K^3} |v|_{3,K}. \quad (4.44)$$

By using the regularity properties, we have that \(1/\rho_K \leq σ/h_K\), and consequently

$$|Θ|_{2,K} \leq C h_K |v|_{3,K}, \quad (4.45)$$

which is what we wanted to show. \(\blacksquare\)
4.5. Interpolation error for $S_2^1(T_{PS12})$

We can now state the following:

**Lemma 4.5.9** (Interpolation Estimate for a regular PS12 Element). For a regular element $(K, S_2^1(K_{PS12}), N_{PS12})$, then for all $v \in H^3(K)$, the local interpolation error is bounded as

\[ |v - \mathcal{I}_v|_{2,K} \leq Ch_K |v|_{3,K}. \]  \hspace{1cm} (4.46)

**Proof.** By the triangle inequality, we have that

\[ |v - \mathcal{I}_v|_{2,K} \leq |v - \Lambda v|_{2,K} + |\mathcal{I} v - \Lambda v|_{2,K}. \]

By applying Corollary 4.5.5 and Lemma 4.5.8 with $m = 2, k = 2$ to respective terms, we obtain

\[ |v - \mathcal{I}_v|_{2,K} \leq Ch_K |v|_{3,K} \]

as desired. \(\square\)

**Error estimates in the $L^2$-norm.**

So far, the only estimates we have are for the $H^2(\Omega)$-norm. It is often of interest to also have some estimates for lower order norms, like the $L^2$-norm, even for a problem stated and solved in $H^2_0(\Omega)$. By employing the so-called *Aubin-Nitsche duality argument* it is possible to derive such estimates. A proof of the following result is given in [OCS86, Section 3.3]:

**Lemma 4.5.10** ($L^2$ Error Estimate). If $u$ is the analytical solution to the weak biharmonic equation, and $u_h$ the corresponding finite element approximation solved with piecewise polynomials of degree $k$, then

\[ \|u - u_h\|_{0,\Omega} \leq Ch_{\min\{2(k-1),k+1,s+4\}} \|f\|_{s,\Omega} \]  \hspace{1cm} (4.47)

where $f \in H^s(\Omega)$ is the corresponding source term and $u$ satisfies the regularity $u \in H^2_0(\Omega) \cap H^r(\Omega)$ for $r = 4 + s$.

By approximating $u$ with piecewise quadratic polynomials ($k = 2$), we directly obtain the following estimate in the $L^2$-norm with a convergence rate independent of the smoothness of $f$:
4. The Finite Element Method

**Theorem 4.5.11.** If \( u_h \) solves the finite element formulation of the biharmonic equation using a regular family of piecewise quadratic \( C^1 \) splines, then the error is bounded in the \( L^2 \)-norm as

\[
\| u - u_h \|_{0, \Omega} \leq C h^2 \max \| f \|_{s, \Omega}
\]

under the assumptions in Lemma 4.5.10.

**Proof.** Setting \( k = 2 \) in Lemma 4.5.10 yields

\[
\min \{2(k - 1), k + 1, s + 4 \} = \min \{2, 3, s + 4 \} = 2
\]

We are now equipped with error estimates for the solution to the biharmonic equation using the PS12-element. We will later see in the numerical results that these estimates are sharp.

Consequently, given a regular triangulation of a domain \( \Omega \), we may obtain interpolation error estimates for the global PS12 interpolation operator over the whole domain by summing over each element domain.

**Theorem 4.5.12** (Global interpolation estimates). For a regular family of PS12-elements with corresponding global interpolation operator \( I \), the following global estimates hold. For all \( v \in H^3(\Omega) \):

\[
\| v - I v \|_{2, \Omega} \leq C h_{\text{max}} \| v \|_{3, \Omega}
\]

**Remark 4.5.13.** Recall that over the space \( H^2_0(\Omega) \), the norms \( \| \cdot \|_{2, \Omega}, \| \cdot \|_{3, \Omega} \) and \( \| \Delta \cdot \|_{0, \Omega} \) are all equivalent, hence similar estimates hold for all \( v \in H^2_0(\Omega) \cap H^3(\Omega) \) in all three norms.
CHAPTER 5

Numerical Results

In this chapter, we will verify the theoretical estimates for the approximation error using the spline space $S^1_2(T_{PS12})$. We start by discussing how such an approximation may be computed on a computer.

5.1 Deriving a linear system

Recall that the space $S^1_2(T_{PS12})$ has dimension $N = 3|V| + |E|$, and that we have associated with this space a set of global basis functions $\phi_1, \ldots, \phi_N$. Any function $u_h$ in $S^1_2(T_{PS12})$ can then be written as a linear combination

$$u_h = \sum_{i=1}^{N} c_i \phi_i.$$  \hspace{1cm} (5.1)

If $u_h$ is the solution to the finite element formulation of the biharmonic equation (Equation (4.1)), then in particular, it satisfies $a_h(u_h, \phi_j) = L_h(\phi_j)$ for all the basis functions $\phi_j$. By expanding $u_h$ as a linear combination and using the linearity of the bilinear form $a_h$, we have that

$$\sum_{i=1}^{N} c_i a_h(\phi_i, \phi_j) = L_h(\phi_j)$$ \hspace{1cm} (5.2)
5. Numerical Results

for all $j = 1, \ldots, N$. This we recognize as a linear system $Ac = b$ where the matrix $A \in \mathbb{R}^{N \times N}$ and vector $b \in \mathbb{R}^N$ have entries given by

$$A_{ij} = a_h(\phi_i, \phi_j) = \int_\Omega \Delta \phi_i \Delta \phi_j \, dx;$$

$$b_j = \int_\Omega f \phi_j \, dx.$$

The method then reduces to the assembly and solution of a linear system.

**Remark 5.1.1.** It is customary to refer to the matrix $A$ as the **stiffness matrix** and the vector $b$ as the **load vector**, with these terms being borrowed from solid mechanics, even if the underlying differential equation is not related to mechanics at all.

The matrix $A$ has properties that depend both on the bilinear operator $a_h$, as well as the specific basis $\phi_1, \ldots, \phi_N$ chosen for the finite element space.

**Lemma 5.1.2 (Properties of $A$).** The matrix $A$ corresponding to the finite element formulation of the Biharmonic equation using the Hermite nodal basis for the space $S_1^2(T_{PS12})$ has the following properties:

1. It is symmetric,
2. positive definite, and
3. sparse.

**Proof.** The fact that the matrix is symmetric follows readily from the definition of the bilinear form $a_h$:

$$A_{ij} = \int_\Omega \Delta \phi_i \Delta \phi_j \, dx = \int_\Omega \Delta \phi_j \Delta \phi_i \, dx = A_{ji}. \quad (5.4)$$

To show positive definiteness, recall that $A$ is positive definite if $c^T Ac > 0$ whenever $c \neq 0$. Let $c \neq 0$ be the coefficient vector corresponding to some element $u_h$ in $V_h$. Expanding the vector matrix product we see that

$$c^T Ac = \sum_{i=1}^N \sum_{j=1}^N c_i \int_\Omega \Delta \phi_i \Delta \phi_j \, dx \ c_j$$

$$= \int_\Omega \Delta u_h \Delta u_h \, dx = a_h(u_h, u_h) \geq C \| u_h \|_2^2$$
5.2. The finite element assembly process

where the last inequality follows by the $H^2_0$-ellipticity of the linear form $a_h$. Furthermore, since $c \neq 0$, we have that $u_h \neq 0$, so $\|u_h\|^2 > 0$. The sparsity of the matrix follows directly from the local support of the global basis functions. If $\phi_i$ is a basis function which corresponds to a vertex carrier $v$, then the support $\text{Supp}(\phi_i)$ is the union of the triangles containing the vertex $v$. Similarly, if $\phi_i$ is a basis function corresponding to an edge $e$, then the support $\text{Supp}(\phi_i)$ consists of the union of the triangles sharing the edge $e$. ■

5.2 The finite element assembly process

Since the domain $\Omega$ is partitioned into a set of macrotriangles $\mathcal{T}$, the integrals occurring in the matrix $A$ and the vector $b$ may be computed by iterating over the macrotriangles. On each macrotriangle, we have a local Hermite nodal basis (recall Definition 4.3.2 on page 43), which we will denote by

$$\mathcal{H}^k = [\mathcal{H}^k_1, \ldots, \mathcal{H}^k_{12}]$$

(5.5)

where $k$ denotes the macrotriangle number. That is, any global nodal basis function $\phi_i$ is represented locally on macrotriangle $K_k$ as a linear combination of the local basis functions $\mathcal{H}^k$. Due to the nature of the global basis functions, with their local support, there are only twelve non-zero global basis functions on each macrotriangle.

A connectivity map

It is important in the assembly of the matrix $A$ and vector $b$ to have an explicit map which given a triangle number $k$ and local basis number $i$, tells us which global basis function $j$ is the corresponding one. That is, we are interested in a map $\sigma$ such that $\sigma(k, i) = j$ gives us the number of the corresponding global basis function. Such a map is often called a connectivity map, and these can be constructed explicitly.

An algorithm for computing the map $\sigma$ in the case where each triangle is endowed with the Powell–Sabin nodal basis is given in Algorithm 5.2.1.
5. Numerical Results

on the facing page. For a triangulation consisting of two triangles, the local and global numbering of the degrees of freedom are depicted in Figure 5.1 on page 72. The sparsity structure of the resulting matrix $A$ depends on how you choose to number the degrees of freedom. From a computational point of view, choosing a good connectivity map is important. The sparsity pattern resulting from the map $\sigma$ generated by Algorithm 5.2.1 can be seen in Figure 5.2 on page 73 along with the corresponding mesh.

**Assembling $A$ and $b$**

Given such a connectivity map $\sigma$, the assembly of the stiffness matrix $A$ and the load vector $b$ is fairly straightforward, and can be accomplished by looping over each macrotriangle, and adding the local contributions. This routine is described in Algorithm 5.2.2 on page 74. The integrals are evaluated using an appropriate numerical quadrature rule for the triangle.

**Remark 5.2.1.** Since the basis functions $H^k_i$ are piecewise polynomials of degree two, the numerical integration in Algorithm 5.2.2 should be performed over each subtriangle in the Powell–Sabin 12-split. Furthermore, the integrand

$$\Delta H^k_i \Delta H^k_j$$

is a piecewise constant function, hence a one-point scheme should suffice on each subtriangle of the 12-split. The integrand

$$f H^k_i$$

on the other hand, depends on the source term $f$. Here, a second-order quadrature rule was used.
Algorithm 5.2.1 Constructing a connectivity map for the $C^1$ quadratic global Hermite basis. The local degrees of freedom are numbered triangle by triangle in a counterclockwise order, in such a way that the degrees of freedom lying close geometrically are given numbers close to each other. The maps $\alpha$ and $\beta$ are auxiliary maps used for storing the numbers of previously accessed global degrees of freedom.

```
procedure DOF($T$)
    $N := 1$
    for triangle $K_k = [v_1, v_2, v_3]$ in $T$ do
        $n := 1$
        for edge $e_j = [v_j, v_{j+1}]$ in $K_k$ do
            if vertex $v_j$ not visited then
                $\sigma(k, n) := N$
                $\sigma(k, n + 1) := N + 1$
                $\sigma(k, n + 2) := N + 2$
                $\alpha(v_j) := (N, N + 1, N + 2)$
                $N = N + 3; n = n + 3$
                else
                    $\sigma(k, n) = \alpha(v_j)$
                    $n = n + 3$
                end if
            if edge $e_j$ not visited then
                $\sigma(k, n + 3) := N$
                $\beta(e_j) := N$
                $N = N + 1; n = n + 1$
                else
                    $\sigma(k, n) := \beta(e_j)$
                    $n = n + 1$
                end if
        end for
    end for
end procedure
```
5. Numerical Results

(a) Global Numbering

(b) Local Numbering

Figure 5.1: Local and global numbering of the respective degrees of freedom, as induced by the map $\sigma$. As seen, the local degrees of freedom are numbered in a counterclockwise fashion.
5.3. Model problems for the biharmonic equation

Recall from Lemma 4.5.9 on page 65 that for any function $u$ in $H^3(\Omega)$, we can expect the finite element approximation error to be of the form

$$\|u - u_h\|_{2,\Omega} \leq \mathcal{O}(h)$$

in the $H^2$-norm, while in the $L^2$-norm (Lemma 4.5.10), we can expect an error of the form

$$\|u - u_h\|_{0,\Omega} \leq \mathcal{O}(h^2).$$
5. Numerical Results

Algorithm 5.2.2 Assembling the finite element matrix $A$ and the vector $b$. Note that the symmetric properties of $A$ are exploited in the inner for-loop.

```
procedure ASSEMBLE
  Allocate memory for the sparse matrix $A$ and for the $N \times 1$ vector $b$.
  for triangle $K_k$ in $T$ do
    for $j = 1, \ldots, 12$ do
      for $i = 1, \ldots, j$ do
        $A(\sigma(k, i), \sigma(k, j)) += \int_{K_k} \Delta H^k_i \Delta H^k_j \, dx$.
        if $i \neq j$ then $\triangleright$ Symmetry
          $A(\sigma(k, j), \sigma(k, i)) += A(\sigma(k, i), \sigma(k, j))$.
        end if
      end for
      $b(\sigma(k, j)) = \int_{K_k} f H^k_j \, dx$.
    end for
  end for
  return $A$, $b$
end procedure
```

By the method of manufactured solutions, we construct two model problems for the biharmonic equation, which are subsequently solved on a sequence of uniformly refined regular meshes as seen in Figure 5.4.

![Initial triangle](a) Initial triangle  ![Subdivided triangle](b) Subdivided triangle

Figure 5.3: In order to keep the meshes nested and regular, we subdivide each triangle by connecting the edge midpoints, partitioning each triangle into four sub-triangles. This ensures that all knotlines of the coarse triangulation are present in the fine triangulation, and consequently the corresponding spline spaces are nested.
5.3. Model problems for the biharmonic equation

Figure 5.4: A sequence of nested triangulations \( T^{(i)} \subseteq T^{(i+1)} \) of the unit square, obtained by subdividing each triangle by connecting the edge mid points. Such a subdivision ensures that the corresponding spline spaces are nested. Note that one subdivision yields a triangulation with four times as many triangles.

The first model problem

We start by constructing a model problem where the analytical solution is

\[
    u(x, y) = 4\pi \sin^2(\pi x) \sin^2(\pi y) \sin(2\pi(y - x)),
\]

and the corresponding source term \( f \) can be computed as the bilaplacian of \( u \). This yields

\[
    f(x, y) = 8\pi^4 \left[ \sin(2\pi x) - 8\sin(4\pi x) + 25\sin(2\pi(x - 2y)) \\
    - 32\sin(4\pi(x - y)) - \sin(2\pi y) + 8\left( \sin(4\pi y) \\
    + \sin(2\pi(-x + y)) \right) + 25\sin(4\pi x - 2\pi y) \right].
\]

This function \( u \), along with its outward normal derivative, vanishes at the boundary.

Results

On a nested sequence of uniform meshes with \( 2n^2 \) triangles, with \( n = 2^i \) for \( i = 1, \ldots, 5 \), we solve the corresponding finite element problem to the first model problem. The results are summarized in Table 5.1, and are seen to agree with the expected theoretical estimates. Two of the resulting finite element solutions are shown in Figure 5.6. These should be compared to the analytical solution in Figure 5.5.
5. Numerical Results

FIGURE 5.5: The analytical solution and the source term for the homogeneous model problem on the unit square.
5.3. Model problems for the biharmonic equation

Table 5.1: On a sequence of meshes of $2n^2$ triangles we compute the finite element solution to the first model problem. The convergence rate $\alpha$ in the relative $H^2$-norm, and the rate $\beta$ in the relative $L^2$-norm is computed. The results are seen to agree with the theoretical estimates from Lemmas 4.5.9 and 4.5.10, of $O(h)$ and $O(h^2)$ respectively.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$h_{\text{max}}$</th>
<th>$|\Delta e|_{0,\Omega}$</th>
<th>$|\Delta e|<em>{0,\Omega}/|\Delta u|</em>{0,\Omega}$</th>
<th>$\alpha$</th>
<th>$|e|_{0,\Omega}$</th>
<th>$|e|<em>{0,\Omega}/|u|</em>{0,\Omega}$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.414</td>
<td>84.144</td>
<td>0.628</td>
<td>N/A</td>
<td>0.507</td>
<td>0.485</td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td>0.707</td>
<td>56.660</td>
<td>0.423</td>
<td>0.570</td>
<td>0.228</td>
<td>0.218</td>
<td>1.152</td>
</tr>
<tr>
<td>4</td>
<td>0.353</td>
<td>43.970</td>
<td>0.328</td>
<td>0.365</td>
<td>0.087</td>
<td>0.083</td>
<td>1.392</td>
</tr>
<tr>
<td>8</td>
<td>0.176</td>
<td>21.265</td>
<td>0.158</td>
<td>1.048</td>
<td>0.019</td>
<td>0.018</td>
<td>2.186</td>
</tr>
<tr>
<td>16</td>
<td>0.088</td>
<td>10.479</td>
<td>0.078</td>
<td>1.020</td>
<td>0.005</td>
<td>0.004</td>
<td>2.055</td>
</tr>
<tr>
<td>32</td>
<td>0.044</td>
<td>5.213</td>
<td>0.038</td>
<td>1.007</td>
<td>0.001</td>
<td>0.001</td>
<td>2.018</td>
</tr>
</tbody>
</table>

Figure 5.6: The finite element solution to the first model problem with 8 and 512 triangles respectively. The corresponding spline spaces have dimensions 43 and 1667.

The second model problem

Our second model problem is similar to the first one in that it is described using homogeneous boundary conditions. Letting

$$u(x, y) := x^2(1 - x)^2y^2(1 - y)^2$$

(5.8)
5. Numerical Results

we obtain the corresponding source term

\[
f(x, y) = 8 \left[ 3x^4 - 6x^3 + 9x^2(1 - 2y)^2 \\ - 6x(6y^2 - 6y + 1) + 3y^4 \\ - 6y^3 + 9y^2 - 6y + 1 \right]. \tag{5.9}
\]

The source term and the analytical solution \( u \) can be seen in Figure 5.7 on the next page. Again, some tedious calculations show that this \( u \) satisfies the boundary conditions.

Results

The approximate solution is found in a sequence of finite element spaces over the unit square, as in the previous model problem. The results are summarized in Table 5.2 and two of the approximations are shown in Figure 5.8.

**Table 5.2:** The approximations to the second model problem were computed over a sequence of meshes of \( 2^n \) triangles. The convergence rate \( \alpha \) in the relative \( H^2 \)-norm, and the rate \( \beta \) in the relative \( L^2 \)-norm is computed. The results are seen to agree with the theoretical estimates from Lemmas 4.5.9 and 4.5.10 of \( O(h) \) and \( O(h^2) \) respectively.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( h_{\text{max}} )</th>
<th>( | \Delta e |_{\Omega} )</th>
<th>( | \Delta e |<em>{\Omega}/| \Delta u |</em>{\Omega} )</th>
<th>( \alpha )</th>
<th>( | e |_{\Omega} )</th>
<th>( | e |<em>{\Omega}/| u |</em>{\Omega} )</th>
<th>( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.414</td>
<td>0.057</td>
<td>1.000</td>
<td>N/A</td>
<td>0.0015870</td>
<td>1.0000</td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td>0.707</td>
<td>0.045</td>
<td>0.780</td>
<td>0.358</td>
<td>0.0009920</td>
<td>0.6250</td>
<td>0.6781</td>
</tr>
<tr>
<td>4</td>
<td>0.354</td>
<td>0.020</td>
<td>0.356</td>
<td>1.134</td>
<td>0.0002070</td>
<td>0.1307</td>
<td>2.2579</td>
</tr>
<tr>
<td>8</td>
<td>0.177</td>
<td>0.009</td>
<td>0.164</td>
<td>1.117</td>
<td>0.0000440</td>
<td>0.0279</td>
<td>2.2298</td>
</tr>
<tr>
<td>16</td>
<td>0.088</td>
<td>0.004</td>
<td>0.078</td>
<td>1.077</td>
<td>0.0000100</td>
<td>0.0063</td>
<td>2.1536</td>
</tr>
<tr>
<td>32</td>
<td>0.044</td>
<td>0.002</td>
<td>0.038</td>
<td>1.039</td>
<td>0.0000020</td>
<td>0.0015</td>
<td>2.0804</td>
</tr>
</tbody>
</table>
5.3. Model problems for the biharmonic equation

(a) The analytical solution $u$.

(b) The analytical source term $f = \Delta^2 u$.

**Figure 5.7:** The analytical solution and corresponding source term to the second model problem on the unit square.
5. Numerical Results

(a) Approximate solution with 8 triangles.  (b) Approximate solution with 512 triangles.

Figure 5.8: The finite element solution to the second model problem on a mesh with 8 and 512 triangles respectively. The corresponding spline spaces have dimensions 43 and 1667 respectively.
Appendices
APPENDIX A

Implementation

The implementation for this thesis was written in the programming language PYTHON, due to its easy prototyping, with comparatively high speed. In this chapter, we discuss the implementation of the SSplines-package. A python package for the evaluation of constant, linear or quadratic simplex splines over the Powell–Sabin 12-split of a single triangle. The package relies heavily on the use of the numpy-library for C-optimized numerical routines. The full source code is openly available at GitHub:

github.com/qTipTip

A.1 SSplines

The SSplines package provides two main objects. The SplineFunction object which represents a callable spline function over a given triangle, and the SplineSpace object which facilitates the instantiation of several SplineFunction objects.

Basic usage

In order to showcase basic functionality, we will demonstrate how to generate the pictures in Figures 2.3 to 2.5. The visualization was done using the matplotlib library. We start by instantiating a SplineSpace-object.
A. Implementation

### Setup

```python
# import necessary packages
import SSplines
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D

# instantiate a spline space over the PS12-split of a triangle
degree = 2
triangle = np.array([[0, 0],
                     [1, 0],
                     [0.5, np.sqrt(3)/2]])
S = SSplines.SplineSpace(triangle, degree)
```

The object `S` is now a `SplineSpace` object, which when given a set of coefficients, returns the corresponding `SplineFunction` object. Any such `SplineFunction` can be evaluated. We create a function for the evaluation and visualization of a `SplineFunction`.

### Evaluation and Visualization

```python
def visualize(f):
    # visualizes a ‘SplineFunction’ object
    points = SSplines.sample_triangle(f.triangle, 20)
    values = f(points)

    fig = plt.figure()
    axs = Axes3D(fig)
```
The spline function in Figure 2.3 is a specific linear combination of the twelve basis functions of $S$. This can be represented and visualized using the code from above as follows:

```python
# represent a specific spline
coefficients = np.array([-(-1)**(i+1) for i in range(S.dimension)])
spline_funct = S.function(coefficients)
visualize(spline_funct)
```

The SplineSpace object contains getters for the S-basis and the Hermite nodal basis. So, Figures 2.4 and 2.5 can be evaluated as follows:

```python
for b in S.basis():
    visualize(b)
for H in S.hermite_basis():
    visualize(H)
```

The SplineFunction object supports basic arithmetical operations like addition and scalar multiplication. Furthermore, the evaluation of directional derivatives are supported. In Table A.1, the methods of the SplineFunction object is presented.
A. Implementation

<table>
<thead>
<tr>
<th>Python-command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>f(x)</td>
<td>Evaluate at the point $x$.</td>
</tr>
<tr>
<td>f.D(x,u,r)</td>
<td>Evaluates the $r$th directional derivative in the direction $u$ at the point $x$.</td>
</tr>
<tr>
<td>f.dx(x) or f.dy(x)</td>
<td>Evaluates the partial derivative at the point $x$.</td>
</tr>
<tr>
<td>f.ddx(x) or f.ddy(x)</td>
<td>Evaluates the second order partial derivative at the point $x$.</td>
</tr>
<tr>
<td>f.grad(x)</td>
<td>Evaluates the gradient at the point $x$.</td>
</tr>
<tr>
<td>f.div(x)</td>
<td>Evaluates the divergence at the point $x$.</td>
</tr>
<tr>
<td>f.lapl(x)</td>
<td>Evaluates the laplacian at the point $x$.</td>
</tr>
</tbody>
</table>

**Validation of implementation**

The numerical routines implemented in the SSplines package was thoroughly tested against symbolic expressions using scipy. By using the recurrence relation, a symbolic polynomial representation of each basis function on each of the twelve subtriangles was computed. The numerical evaluation and differentiation of the S-splines were verified to agree with the evaluation and differentiation of these polynomial pieces.

**A.2 Finite element implementation**

In this section we gloss over some useful packages in the implementation of the finite element method.

**Numerical integration**

Initially, the idea was to use the dblquad function from scipy for the numerical integration of choice. However, this integration method failed to converge when integrating over the PS12-split, even when the integrations were performed over each sub-triangle. Instead we employed quadpy [Sch18], a well-documented and thoroughly tested package for
A.2. Finite element implementation

numerical integration of arbitrary order over a wide range of geometrical regions, even higher order simplices. For the evaluation of the bilinear and linear forms, the second order triangular quadrature rule quadpy.triangle.XiaoGimbutas(2) was used, while for the computation of the approximation errors, a fourth order quadrature rule was used: quadpy.triangle.XiaoGimbutas(4). All integration was performed over the subtriangles of the PS12-split using the auxiliary method SSplines.ps12_sub_triangles from the SSplines-package.

Mesh generation

Storing meshes on a computer is purely a matter of choosing the right data structure. We chose to go for the structure involving two matrices $V$ and $T$ called the *point-matrix* and *connectivity-matrix*, respectively, where $V$ is an $(N \times 2)$-matrix containing the coordinates of each of the $N$ vertices in the mesh, and $T$ is an $(M \times 3)$ matrix where each row contains the indices of the vertices making up one of the $M$ triangles. In this thesis, all triangulations have been of the unit rectangle, where such matrices are easily generated. However, when the domain gets complicated, mesh generation becomes a topic in and of itself.

There are several heavy duty python packages (pygmsh, meshr and meshpy to name a few) dealing with meshing and mesh-generation, most of them using the Triangle software [She96] under the hood. However, for someone new to the finite element method, a simple mesh is all that is needed to get started. The python package meshzoo (by the author behind quadpy) gives access of helper functions for the generation of meshes of triangles, rectangles and other simple geometric shapes. As an example, the point-matrix and the connectivity-matrix of the first mesh in Figure 5.2 can be generated using the following code.

```python
Meshzoo
```
A. Implementation

```python
import meshzoo
vertices, triangles = meshzoo.rectangle(
    xmin=0, xmax=1, ymin=0, ymax=1,
    nx = 3, ny = 3, zizag=False
)
```

**Sparse linear algebra**

The sparsity structure of the stiffness matrix resulting from the finite element assembly is one of the major computational benefits of the finite element method and should be exploited. The scipy-library offers a set of sparse matrix formats. We used the lil-format (linked list sparse format) for efficient array slicing as this tied well into the SSplines-implementation. However, in hindsight, the coo-format (coordinate format) is probably more suited for finite element assembly.

For the solution of the sparse linear system, a direct method implemented as `scipy.sparse.linalg.spsolve` was used due to the relatively small matrices occurring in the model problem. Since the stiffness matrix is symmetric positive definite, the conjugate gradient method ( `scipy.sparse.linalg.cg`) is a good method for an iterative solution.


A. Implementation


[Sch18] Schlömer, N. nschloe/quadpy v0.11.3. 2018.
A.2. Finite element implementation


