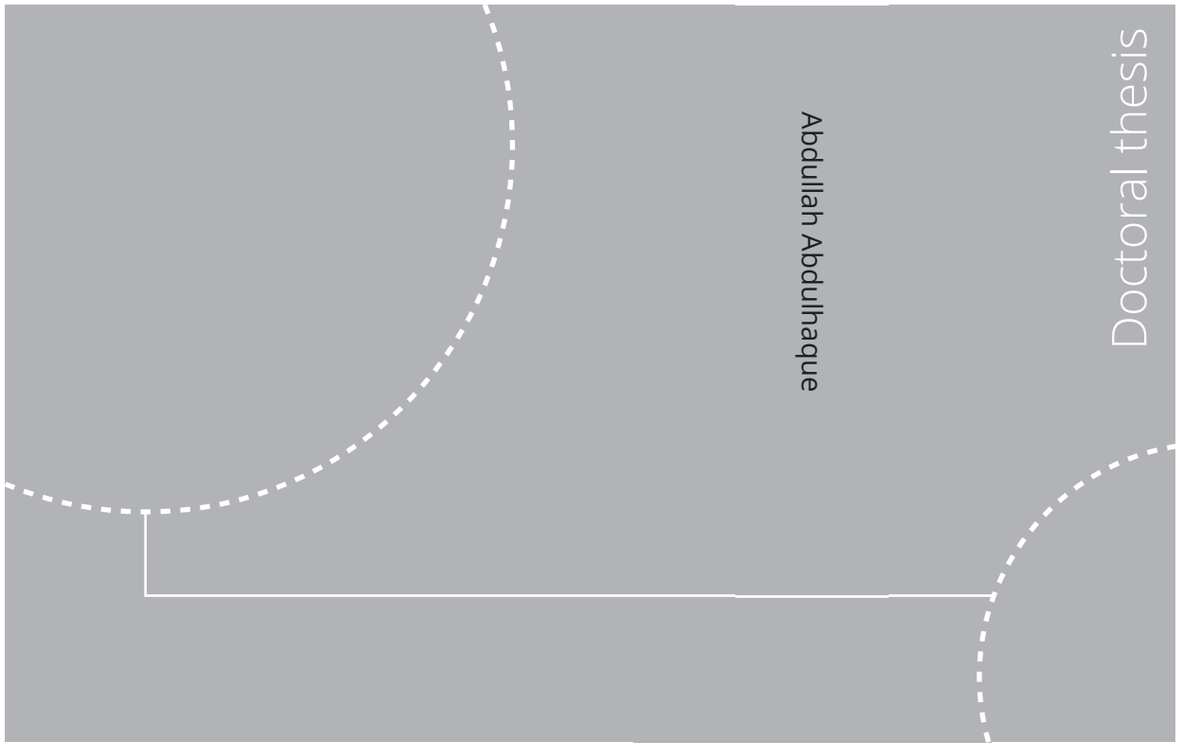
ISBN 978-82-326-5296-9 (printed ver.)
ISBN 978-82-326-6002-5 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (electronic ver.)



Doctoral theses at NTNU, 2022:261

Abdullah Abdulhaque

Adaptive Isogeometric Methods for Boussinesq Problems

NTNU
Norwegian University of Science and Technology

Doctoral theses at NTNU, 2022:261

NTNU

NTNU
Norwegian University of
Science and Technology
Thesis for the degree of
Philosophiae Doctor
Faculty of Information Technology
and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

Abdullah Abdulhaque

Adaptive Isogeometric Methods for Boussinesq Problems

Thesis for the degree of Philosophiae Doctor

Trondheim, September 2022

Norwegian University of Science and Technology
Faculty of Information Technology
and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology
and Electrical Engineering
Department of Mathematical Sciences

ISBN 978-82-326-5296-9 (printed ver.)
ISBN 978-82-326-6002-5 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (electronic ver.)

Doctoral theses at NTNU, 2022:261

Printed by Skipnes Kommunikasjon AS

Abstract

Computational Fluid Dynamics (CFD) is the numerical study of fluid flow, heat transfer, turbulence modelling and several conservation laws. The fluids can be liquids, gases and even plasmas. There is a vast number of differential equations describing the physical problems in specific situations, and many of them have a nonlinear structure. The main goals of CFD are constructing an efficient and stable numerical technique for discretizing the equations with respect to space and time, and then solving the (nonlinear) algebraic systems of equations arising from the discretization as quick and accurate as possible. This might not be straightforward because the procedure always depends on the specific situation. The simulations are often carried out over long time intervals, and that will make high accuracy in space and time desirable. Furthermore, nonlinearity and local instabilities can also slow down the computational speed. When we couple several equations together, the solution procedure becomes even more complex.

The most common numerical procedures utilized in CFD are the *Finite Difference Method* (FDM), the *Finite Volume Method* (FVM) and the *Finite Element Method* (FEM). The latter one is most the general and widespread because we can apply it on arbitrary complex domains that are sufficiently smooth, and it can perform local refinement in those parts of the domain where the unknown solution lacks sufficient regularity. There are a lot of similar FEM-approaches, and they differ most with respect to the choice of basis functions. One such method is called *Isogeometric Analysis* (IGA). It has superior approximation properties compared with classical FEM, and its signature ability is creating an exact mesh of the domain's geometry. The discretized equations can be solved quickly, and all these advantages make IGA well-suited for CFD applications.

The main focus of the thesis is solving the hydrodynamic Boussinesq equations for buoyancy-driven flow numerically. The PDE system consists of the Navier-Stokes equation and Advection-Diffusion equation coupled together. In particular, our research emphasizes adaptive error estimation

and local refinement using isogeometric discretization. Adaptive refinement originated in the late 1970s. It was designed for reducing approximation error by generating a new mesh repeatedly until it resembled the unknown solution's physical structure. In classical FEM, the theory of a posteriori error estimation is complete and has been applied widely to large classes of differential equations. This method is far better than a priori error estimation because it allows us to analyse local parts of the solution effectively and determine the corresponding local error. In CFD, there are many well-known situations where adaptive refinement and error estimation are desirable. We need a suitable method for reducing the error quickly without too much computational effort at the same time.

We consider qualitative analysis of efficient a posteriori error estimators for IGA. This topic is still in a development stage although the classical refinement theory is compatible with the isogeometric paradigm. Splines are in general not interpolatory like the shape functions from FEM. Since they have higher continuity and better approximation properties, there is a good reason to believe that isogeometric refinement yields very good results for smooth problems. We will investigate whether some of these classical error estimators can be adapted directly to IGA, and then test them on some major PDEs in CFD: the Stokes equation, the Navier-Stokes equations, the Advection-Diffusion equation, and the Boussinesq equations.

Preface

This thesis is submitted in partial fulfilment of the requirements for the degree of Philosophiae Doctor (PhD) in mathematics at the Norwegian University of Science and Technology (NTNU). The work was carried out at the Department of Mathematical Sciences (IMF), NTNU, and the Department of Applied Mathematics and Cybernetics, SINTEF Digital, both in Trondheim, from August 2016 to June 2022.

I was introduced to Isogeometric Analysis (IGA) in May 2015 by my supervisor, Professor Trond Kvamsdal. At that time, I had some knowledge of the Finite Element Method (FEM) and fluid mechanics (CFD), so we decided that the topic of my master's thesis should be isogeometric analysis of the Boussinesq equations. This system of partial differential equations has been solved with finite and spectral elements before, and it has a wide range of applications. But it had never been solved with IGA, which would be unique for my research.

In my master's thesis, I demonstrated that the Boussinesq equations can be solved with IGA, and the numerical results are far better than those ones obtained with classical FEM. Therefore, Kvamsdal proposed that our next step should be solving this equation system with adaptive refinement. Instead of the traditional approach with classical FEM, we chose IGA as our tool. To simplify the working process optimally, we decided to apply the same flexible and effective strategy from my master's thesis: analyse and solve the Navier-Stokes equation and Advection-Diffusion equation separately, and then combine their solution strategies together.

With this strategy in mind, I could decide each article's topic early, find relevant literature quickly, and choose benchmark problems for comparing numerical simulations. The simulation software was ready, but it had to be continuously refined and improved. It also took much time to figure out the best adaptive strategy, especially for the first article. After this clarification, the other simulations went fast. The first three articles constitute the basis for the final one, which covers the main topic of the thesis.

Acknowledgements

I would like to express my most profound gratitude to professor Trond Kvamsdal for his valuable guidance since he began supervising my master's thesis in 2015. Throughout these years, I was blessed with his frequently advising sessions where he showed genuine interest in my work. He guided me through every stage of my career as a research scientist and gave me a lot of academic experience. Already as a graduate, he gave me the opportunity to participate in the *3rd International Conference on Isogeometric Analysis* (IGA 2015), which provided a huge insight in the mathematical field that I was going to work with. During the PhD study, we participated together in seven conferences on computational mechanics, and there was a lot of exciting knowledge to get here.

I would like to thank my co-supervisor Dr. Arne Morten Kvarving[1] for his invaluable support during my PhD study. He taught me a lot about supercomputing and how to use the finite element method in computational fluid dynamics. Every time when I got stuck with the numerical simulations, he always helped me so I could detect the reason for the failure. Thus, I could proceed almost immediately. It saved a lot of time on several occasions during the PhD work.

I would also like to thank my second co-supervisor Dr. Mukesh Kumar[2] for his excellent feedbacks on my work. His discussions on methodology improvement and solution analysis in my PhD thesis have been of high importance. He had always a lot of remarkable proposals to emphasize how our work was differing from previous research on the same topics.

At SINTEF, I would like to thank Dr. Eivind Fonn[3] for his fantastic tutorials on isogeometric analysis, supercomputing, and post-processing. He played a major role during my master's thesis and the early stages of my PhD.

[1] Resarch scientist, SINTEF ICT
[2] Associate professor, Department of Mathematics College of Charleston, South Carolina
[3] Research scientist, SINTEF ICT

I would also like to thank Dr. Kjetil André Johannessen[4] at SINTEF for his contribution to the software development. It would not be possible to run the simulations without his adaptive refinement extensions in the numerical software.

[4]Research manager, SINTEF ICT

Abbreviations

AFEM	Adaptive Finite Element Method
AMR	Adaptive Mesh Refinement
BVP	Boundary Value Problem
CAD	Computer Assisted Design
CFD	Computational Fluid Dynamics
CGL2	Continuous Global L^2-projection
FDM	Finite Difference Method
FEA	Finite Element Analysis
FEM	Finite Element Method
FVM	Finite Volume Method
IFEM	Isogeometric Finite Element Method
IGA	Isogeometric Analysis
HR	Hierarchically refined
LBB	Ladyzhenskaya-Babuška-Brezzi
LR	Locally refined
MFEM	Mixed Finite Element Method
MsFEM	Multiscale Finite Element Method
Ndof	Number of Degrees of Freedom
NURBS	Non Uniform Rational B-splines
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
SEM	Spectral Element Method
SFEM	Smoothed Finite Element Method
SPR	Superconvergent Patch Recovery
SUPG	Streamline-Upwind/Petrov-Galerkin
TH	Truncated Hierarchical
UMR	Uniform Mesh Refinement
ZZ	Zienkiewicz-Zhu

Nomenclature

Latin Capital Letters

Br	Brinkman number
C^k	Space of continuous functions of order k
$C_{\mathcal{M}}$	Shape parameter of mesh $\mathcal{M}$
$\mathbf{C}$	Convection matrix
Cr	Courant number
D^α	Partial derivative of multi-index α
$\mathbf{D}$	Divergence matrix
$\mathbf{D}^T$	Gradient matrix
E	Energy space
$\mathcal{E}$	Set of edges
$\mathcal{E}_0$	Set of interior edges
$\mathcal{E}_D$	Set of Dirichlet edges
$\mathcal{E}_N$	Set of Neumann edges
$\mathcal{E}_R$	Set of Robin edges
G_h	Recovery operator
Gr	Grashof number
H^k	Sobolev space of indices $p = 2$ and $k \in \mathbb{N}$
$\mathcal{H}$	Knot vector in η-direction
$\mathbf{I}$	Identity matrix
$\mathbb{J}$	Jump in a function
$\mathcal{J}$	Jacobian
K	Element on the mesh
$\widehat{K}$	Reference element
$\mathbf{K}$	Stiffness matrix
L^p	Lebesgue space of index $p \in [1, \infty)$
$L_{l,r}$	B-spline basis function of index l and order r in ζ-direction
$\mathcal{L}$	Linear partial differential operator
$M_{j,q}$	B-spline basis function of index j and order q in η-direction

$\mathcal{M}$	Mesh on the domain
$\mathbf{M}$	Mass matrix
$\mathbb{N}$	Space of natural numbers (excluding 0)
$\mathcal{N}$	Set of degrees of freedom
N_ξ	Number of elements in ξ-direction
N_η	Number of elements in η-direction
N_ζ	Number of elements in ζ-direction
$N_{i,p}$	B-spline basis function of index i and order p in ξ-direction
Nu	Nusselt number
$\mathcal{P}$	Space of shape functions
$\mathbb{P}^p$	Polynomial space of degree p
Pe	Péclet number
Pr	Prandtl number
$\mathbb{Q}$	Space of rational numbers
$\mathbb{Q}^+$	Space of positive rational numbers (including 0)
$\mathbb{Q}^-$	Space of negative rational numbers
R_{ijl}	NURBS basis function of index (i, j, l)
$\mathbb{R}$	Space of real numbers
$\mathbb{R}^+$	Space of positive real numbers (including 0)
$\mathbb{R}^-$	Space of negative real numbers
$\mathbb{R}^d$	Euclidean space of dimension d
Ra	Rayleigh number
Re	Reynolds number
Ri	Richardson number
$\mathbb{RP}^d$	Real projective space of dimension d
$\mathcal{RT}$	Raviart-Thomas element
$\mathbb{S}_k^p$	Spline space of degree p and continuity k
$\mathcal{SG}$	Subgrid Taylor-Hood element
$\mathcal{TH}$	Taylor-Hood element
V	Trial space
W	Test space
$W^{k,p}$	Sobolev space of indices $p \in [1, \infty)$ and $k \in \mathbb{N}$
$\mathbb{Z}$	Space of integers
$\mathbb{Z}^+$	Space of positive integers (including 0)
$\mathbb{Z}^-$	Space of negative integers
$\mathcal{Z}$	Knot vector in ζ-direction

Latin Small Letters

d	Euclidean dimension
$\mathbf{e}_i$	Standard Cartesian basis vector
$\mathbf{f}$	Force vector
$\mathbf{g}$	Gravity vector
h_K	Diameter of element K
$\hbar_K$	Modified diameter of element K
k	Continuity
m	Maximal number of refinements
m_i	Multiplicity of knot i
n_1	Number of basis functions in η-direction
n_2	Number of basis functions in ξ-direction
n_3	Number of basis functions in ζ-direction
$\mathbf{n}$	Unit normal vector
p, r	Polynomial degree
u	Exact solution
u_h	Numerical solution
u^*	Recovered solution
$\mathbf{x}$	Cartesian vector
(x, y, z)	Cartesian coordinate

Greek Capital Letters

Δ_{x_i}	Knot partition in x_i-direction
Ξ	Knot vector in ξ-direction
Π	Projection operator
Ω	Physical domain
$\widehat{\Omega}$	Parametric domain
$\widetilde{\Omega}$	Parent domain

Greek Small Letters

δ_{ij}	Kronecker's delta
γ	Edge
η	Global error estimator
η_K	Local error estimator on element K
θ	Global effectivity index
θ_K	Local effectivity on element K
κ_K	Shape ratio on element K
μ	Dynamic viscosity
μ_d	Lebesgue measure on $\mathbb{R}^d$
(ξ, η, ζ)	Parametric coordinate
ρ	Mass density

ρ_K	Diameter of inscribed circle in element K
ψ	Shape function
ν	Kinematic viscosity
ϑ_K	SUPG tuning parameter on element K

Contents

Introduction

1 Historical background

We present some important historical facts on the development of finite element modelling that are relevant for the background of the thesis before starting with the formal introduction of the theoretical parts. The focus is the transition from the Finite Element Method to Isogeometric Analysis.

1.1 A brief history of the Finite Element Method

The *Finite Element Method* (FEM) is a major numerical procedure used for solving partial differential equations. It can handle any boundary conditions and discretize arbitrary domains with a complex and reasonably smooth geometrical structure. The main idea is converting a PDE from its original strong form into its equivalent weak form, and afterwards we approximate the weak solution as a finite linear combination of shape functions.

FEM originated as an ad hoc numerical technique in computational mechanics, particularly for structural engineering, thanks to the work of Galerkin. A groundbreaking development took place during the 1940s and 1950s, when Courant and Argyris formalized and generalized the finite element modelling concept by equipping it with a consistent and rigorous mathematical foundation. Consequently, it was demonstrated that FEM could be applied to any type of PDE arising in various physical sciences. This influential paradigm gave the impetus of constructing various families of shape functions with special characteristic properties.

A fundamental paradigm was launched in the 1960s, when Zienkewicz and his collaborators introduced the well-known *isoparametric concept*, one of the most important facilities of modern FEM technology. It requires that we use the same basis functions for approximating the PDE's unknown solution field and generating a suitable mesh on the physical domain. This new point of view became prominent due to many underlying factors like creating direct geometry-to-mesh mappings, developing flexible elements with curved edges, and avoiding inefficient conversions between different types of shape functions [43]. All these factors were quite time-consuming, and their influence needed to be reduced.

Another ground-breaking leap occurred in the 1970s with the invention of the *Mixed Finite Element Method* (MFEM) for high-order PDEs and systems of PDEs. The main focus was creating sophisticated approximation spaces for special variational formulations in order to provide both effective discretization and good preservation of physical structures. This paradigm had a profound influence on computational multi-physics problems. It also gave rise to many new finite element functions, thanks to leading experts

like Babuška, Brezzi, Nédélec, Fortin, Raviart and many others [16].

Since the 1980s, several new variants and pitchforks of finite element modelling have been invented and developed extensively, like the *Spectral Element Method* (SEM), *Boundary Element Method* (BEM), *Multiscale Finite Element Method* (MsFEM) and *Infinite Element Method* (IEM), just to mention a few. Today, there are many coexisting paradigms in the finite element hierarchy. The method of discretization depends on the specific problem, making the techniques more flexible.

Despite its strengths and advantages, FEM had many shortcomings and drawbacks. The isoparametric concept was only applicable for C^0-elements, and transferring it to high order continuity became very complicated and expensive with respect to implementation. Babuška demonstrated with his famous paradox that curved boundaries can never be represented exactly by straight-edged elements. They are just approximated. High accuracy was also difficult and inefficient to obtain with respect to computational effort, and the preclusion of spurious error propagation was not so straightforward either. These disadvantages sparked the motivation for creating more robust basis functions for the existing paradigms. A potential solution for this challenge was proposed in 2005 by Thomas Hughes, a well-known leading expert on finite element modelling [29].

1.2 A brief history of Isogeometric Analysis

In 2005, Hughes, Cottrell and Bazilevs introduced a new finite element method which they called *Isogeometric Analysis* (IGA) [62]. The main idea is using splines as basis functions. They can represent complex geometries exactly and have superior approximation properties. For example, splines are isoparametric for any level of continuity, and they provide direct and efficient geometry-to-mesh mappings. High continuity increases numerical accuracy very fast, ensures enhanced stability, and smooths out global error propagation quickly. Splines can even handle discontinuous data without many complications [12, 111, 31, 36, 64, 63, 74, 92]. This finite element paradigm has now received widespread recognition and experienced a rapid development since its launch.

The most characteristic feature of IGA is the exact representation of the domain's geometry, from which the term "isogeometric" arises. In classical FEM, we must interpolate the solution field with a certain type of basis functions, and they are used later for constructing a suitable mesh on the domain. In IGA, this process is totally reversed. We create an exact mesh on the domain with the help of appropriate basis functions first, and then

we apply them afterwards for approximating the solution field. As a result, the numerical strategy is geometry independent. The domain's geometry determines which type of spline function it is most convenient to use [29]. New research has also shown that IGA works well for large general classes of PDEs [11, 32, 78, 119].

Another remarkable advantage of IGA is the reduced computational running time. Splines have high continuity, so there is a great overlapping between the elements. Hence, the systems of equations arising from the isogeometric discretization are sparse, have lower spectral radius, and are significantly smaller than those from classical finite element discretization. Iterative solution algorithms will work faster for such discrete systems [50, 49, 96]. This increased computational speed is not possible to achieve with classical FEM. As a consequence of all these new benefits, IGA has given impetus for extensive research in numerical linear algebra [25, 4, 3, 46, 47, 60, 58, 59, 115].

Splines were originally invented as computational geometry tools, with focus on Computer Assisted Design (CAD). Finite Element Analysis (FEA) was designed as an equipment for solving PDEs. In a historical perspective, FEA and CAD evolved in separate communities with different purposes, but they are both essential for modern product development. The efficient interoperability between these two important technologies was disturbed because each community focused on how to improve disjoint stages in the product development instead of relating these stages to each other. But the invention of the isogeometric paradigm provides full interoperability, and therefore, IGA bridges the gap between FEA and CAD. The other current development trend in IGA is to generalize, combine and improve all the other well-established facilities of classical FEM.

The idea of discretizing PDEs by splines has existed for many years, and there has also been some sporadic research on this topic before 2005. But the further systematic development has been quite limited. Thanks to the emerge of efficient object-oriented programming languages and powerful computers, it has now been possible to carry out thorough analysis [88, 97]. IGA has also contributed to extensive research in several disciplines of computational mechanics like electromagnetism [23, 24, 95], fluid dynamics [13, 21, 37, 38, 39], structural engineering [28, 27, 89, 114], and biomechanics [118]. There have also been creative attempts to combine spline technology with other techniques like the Finite Volume Method (FVM) [55, 86] and Boundary Element Method (BEM) [41, 42, 82, 90, 108].

1.3 A brief history of local refinement

Local refinement, which is used in the *Adaptive Finite Element Method* (AFEM), is a special process where we solve a PDE on a coarse mesh, and then we loop over each element and estimate the numerical approximation error locally. If this error exceeds a predefined tolerance, the corresponding element is subdivided into smaller elements with almost the same shape as the original one. All the other elements with sufficiently low estimated error remain unchanged. Afterwards, we solve the PDE and repeat the subdivision procedure again until the global error is low enough. Just by splitting some selected elements where the estimated error is too high, we can increase the accuracy faster. At the end, the mesh will resemble the unknown solution's physical structure, from which the term *adaptive refinement* originates. In comparison with uniform refinement, where every single element is divided, the total computational effort is significantly reduced.

This procedure of local refinement and adaptive mesh generation is fully available in classical FEM and has been studied extensively by many leading experts like Zienkiewicz, Babuška, Demkowicz and Oden. Both B-splines and NURBS have better approximation properties than the classical finite element interpolants, but they only provide tensor refinement. This means that if we subdivide an arbitrary element, all the other adjacent elements must also be divided to preserve the conformal mesh, for any mesh line on the domain traverses the whole length. This approach increases the running time substantially.

A natural question is whether the local refinement technique can be transferred to IGA, and this is indeed possible. Although the concept of adaptive refinement in the isogeometric context is still in the development phase, there has been a lot of progress. Parts of the theoretical foundation for adaptive isogeometric refinement was established in other contexts.

Hierarchical B-splines (HB-splines), proposed by Forsey and Bartels [44], were the first attempt to permit local refinement for splines, and this has been studied further in [30, 22, 48, 66, 77, 85, 112]. Giannelli et al. [52, 51, 53] developed this new idea further by introducing *Truncated Hierarchical B-splines* (THB-splines) to ensure partition of unity. Finally, Dokken et al. proposed Locally Refined (LR) B-splines for making CAD and FEA interoperable with respect to local refinement [35]. A systematic description and numerical verification of local refinement in IGA has also been studied by Johannessen et al. in [69, 70, 68, 75, 76].

T-splines are the corresponding attempt for providing local refinement for NURBS, introduced by Sederberg et al. in 2003 [107]. Developed as a tool for computational geometry, it provides good representation of objects with complicated geometrical structure. Unfortunately, these basis functions

are linearly dependent, so M. Scott introduced *Analysis Suitable T-splines* (AS T-splines) for ensuring compatibility with FEA [106, 80]. A rational analogue of HB-splines, hierarchical NURBS, is studied in [102, 116].
In addition to LR- and T-splines, there has also been some efforts to make other classes of splines applicable to local isogeometric refinement, like *hierarchical box-splines* [71], *Multi-Patch B-splines* (MPB-splines) [20] and *Polynomial splines over Hierachical T-meshes* (PHT-splines) [109, 120]. Scott et al. [105] introduced *Isogeometric spline forests* as a new tool for adaptive refinement. In [56, 100, 101, 117], many types of splines permitting local refinement have been compared in a systematic way with respect to several features like approximation properties, computational effort, and underlying linear algebra structure.

1.4 A brief history of error estimation

Error estimation is a standard subroutine of local refinement in AFEM. First, we detect parts of the domain where the numerical solution might be too inaccurate, and then we subdivide the selected elements to reduce the error. This process is repeated several times, so we need to construct an efficient error estimator which can be calculated very fast and straightforwardly. Babuška and Rheinboldt were among the first ones to study this topic [5, 6]. Today, there are two general families:

- *Residual estimators*: The FEM-solution does not satisfy its governing PDE, and the corresponding residual error is estimated by solving local problems where load functions are given by local residuals.
- *Recovery estimators*: Projection is used to recover post-processed quantities from the solution, and the error is estimated by taking the difference between recovered solution and current FEM-solution.

The development of a posteriori estimators was initially unsystematic due to focus on special individual BVPs. Since the 1990s, the theory has been applied to more general and larger classes of PDEs. Demkowicz and Oden were among the leading figures in the creation of residual estimators. The Zienkiewicz-Zhu estimator from *Superconvergent Patch Recovery* (SPR), based on the improved gradient, is a very common error estimator in AFEM. It is effective, requires little implementation effort, and is most robust for smooth problems approximated by linear and quadratic shape functions [8, 9, 7, 121, 122]. This has been verified for B-splines and LR B-splines. Many recent studies have demonstrated that the classical theory of adaptive refinement can be incorporated and improved in IGA [2, 76, 73].

2 Isogeometric Analysis

This chapter highlights the most important facilities of the shape functions utilized in the isogeometric finite element discretization. Splines have many advantages not shared with their predecessors from classical FEM. We present their characteristic properties and illustrate how they are related to each other. Much of the relevant theory described here will be invoked later when the main research of the thesis really begins, in particular the derivation and reliability analysis of error estimators applied in the adaptive isogeometric mesh refinement.

2.1 Introduction to B-splines

The basics of B-splines

Let Δ_x be a uniform partition on the compact interval $[a,b] \subset \mathbb{R}$ such that

$$\Delta_x = \{x_i\}_{i=0}^{N} \quad , \quad a = x_0 < x_1 < x_2 < \cdots < x_{N-1} < x_N = b$$

Then, we split $[a,b]$ into N disjoint subintervals $\{I_i\}_{i=1}^{N}$ as follows:

$$[a,b] = \bigcup_{i=1}^{N} I_i \quad , \quad I_i = [x_{i-1}, x_i]$$
$$i \neq j \implies I_i \cap I_j = \emptyset$$

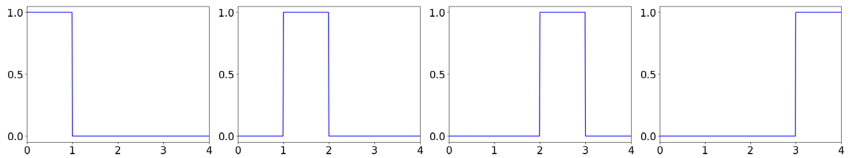
A *knot vector* is a sequence of nondecreasing knots $\Xi = \{\xi_i\}_{i=1}^{n+p+1}$ on the partition Δ_x such that the knots ξ_i equal the grid points x_i. It is used for creating a *spline function*, a piecewise defined and globally differentiable function which is expressed as a linear sum of n B-spline basis functions of polynomial degree p [17, 84]:

$$s(x) = \sum_{i=1}^{n} c_i N_{i,p}(\xi)$$

The B-spline basis functions are uniquely represented by the *Cox-de Boor formula*. This recursive relation is defined as follows:

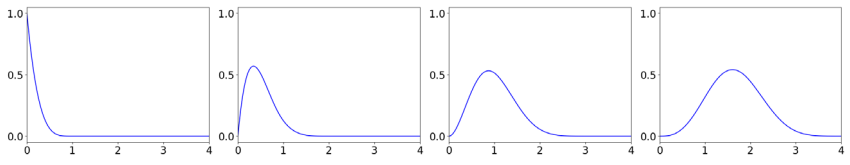
$$N_{i,p}(\xi) = \frac{\xi - \xi_i}{\xi_{i+p} - \xi_i} N_{i,p-1}(\xi) + \frac{\xi_{i+p+1} - \xi}{\xi_{i+p+1} - \xi_{i+1}} N_{i+1,p-1}(\xi) \tag{1a}$$

$$N_{i,0}(\xi) = \chi_{[\xi_i, \xi_{i+1})} = \begin{cases} 1 & \xi_i \leq \xi < \xi_{i+1} \\ 0 & \text{otherwise} \end{cases} \tag{1b}$$

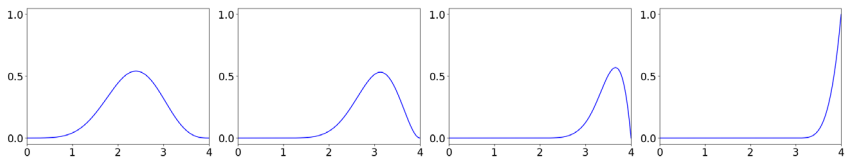


(a) $N_{1,0}$ **(b)** $N_{2,0}$ **(c)** $N_{3,0}$ **(d)** $N_{4,0}$

Figure 1. Plot of the four constant B-spline basis functions on the open knot vector $\{0, 1, 2, 3, 4\}$.

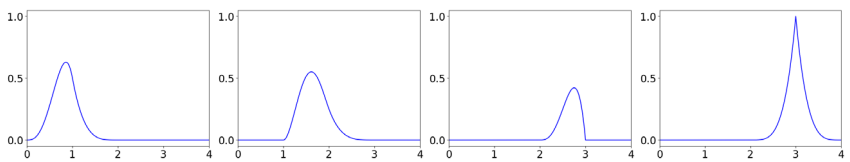


(a) $N_{1,4}$ **(b)** $N_{2,4}$ **(c)** $N_{3,4}$ **(d)** $N_{4,4}$



(e) $N_{5,4}$ **(f)** $N_{6,4}$ **(g)** $N_{7,4}$ **(h)** $N_{8,4}$

Figure 2. Plot of the eight quartic B-spline basis functions on the open knot vector $\{0, 0, 0, 0, 0, 1, 2, 3, 4, 4, 4, 4, 4\}$.



(a) $N_{4,4}$ **(b)** $N_{6,4}$ **(c)** $N_{9,4}$ **(d)** $N_{10,4}$

Figure 3. Plot of some selected quartic B-spline basis functions on the knot vector $\{0, 0, 0, 0, 0, 1, 1, 1, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4\}$.

The first-order derivatives are similarly given by

$$\frac{d}{d\xi}N_{i,p}(\xi) = \frac{p}{\xi_{i+p} - \xi_i}N_{i,p-1}(\xi) - \frac{p}{\xi_{i+p+1} - \xi_{i+1}}N_{i+1,p-1}(\xi) \tag{2}$$

The derivative formula (2) can be generalized to any order [91]:

$$\frac{d^\alpha}{d\xi^\alpha}N_{i,p}(\xi) = \frac{p!}{(p-\alpha)!}\sum_{j=0}^{\alpha} a_{\alpha,j}N_{i+j,p-\alpha}(\xi) \tag{3a}$$

$$a_{0,0} = 1 \tag{3b}$$

$$a_{\alpha,0} = \frac{a_{\alpha-1,0}}{\xi_{i+p-\alpha+1} - \xi_i} \tag{3c}$$

$$a_{\alpha,j} = \frac{a_{\alpha-1,j} - a_{\alpha-1,j-1}}{\xi_{i+p+j-\alpha+1} - \xi_{i+j}} \qquad 1 \le j \le \alpha - 1 \tag{3d}$$

$$a_{\alpha,\alpha} = -\frac{a_{\alpha-1,\alpha-1}}{\xi_{i+p+1} - \xi_{i+\alpha}} \tag{3e}$$

The recursive relation (1) is useful because it provides fast evaluation of spline functions at given points. This approach is far more efficient than complete symbolic derivation of exact expressions. The B-spline formula is recursive, but can be implemented with elementary dynamic programming techniques, and the original exponential running time will drop down to polynomial running time. This acceleration of the computational speed is stable [18]. It has been proven that Cox-de Boor recursion is preserved for some translation invariant operators, and that will speed up the transform of B-splines. A full characterization of such operators is found in [83].
The B-splines have many important properties:

1. **Uniqueness**: The B-spline $N_{i,p}$ depends only on the knots $\{\xi_j\}_{j=i}^{i+p+1}$.
2. **Local support**: $\mathrm{supp}(N_{i,p}) = (\xi_i, \xi_{i+p+1})$.
3. **Positivity**: $\xi \in (\xi_i, \xi_{i+p+1}) \implies N_{i,p} > 0$.
4. **Continuity**: B-splines are smooth polynomials of degree p between the knots. On a knot with multiplicity m, the continuity is C^{p-m}.
5. **Stability**: The B-spline basis is stable and linearly independent.
6. **Partition of unity**: $\sum_{i=1}^{n} N_{i,p}(\xi) = 1$.

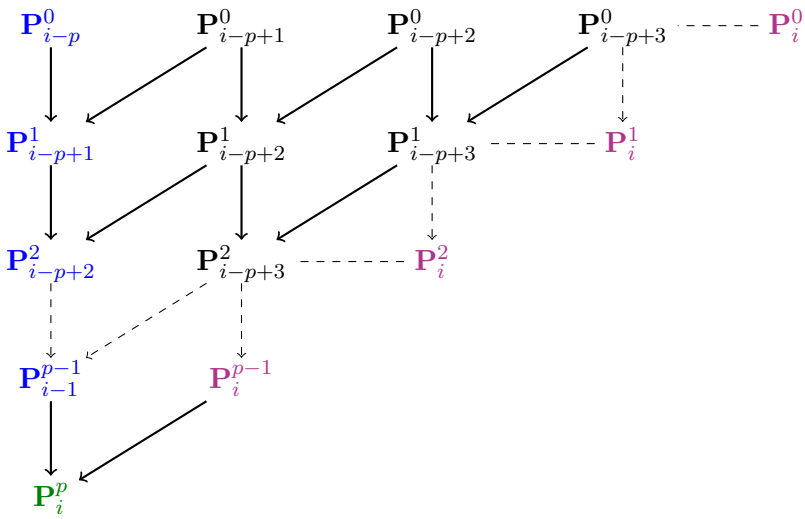


Figure 4. Visualization of the Cox-de Boor algorithm for B-spline evaluation.

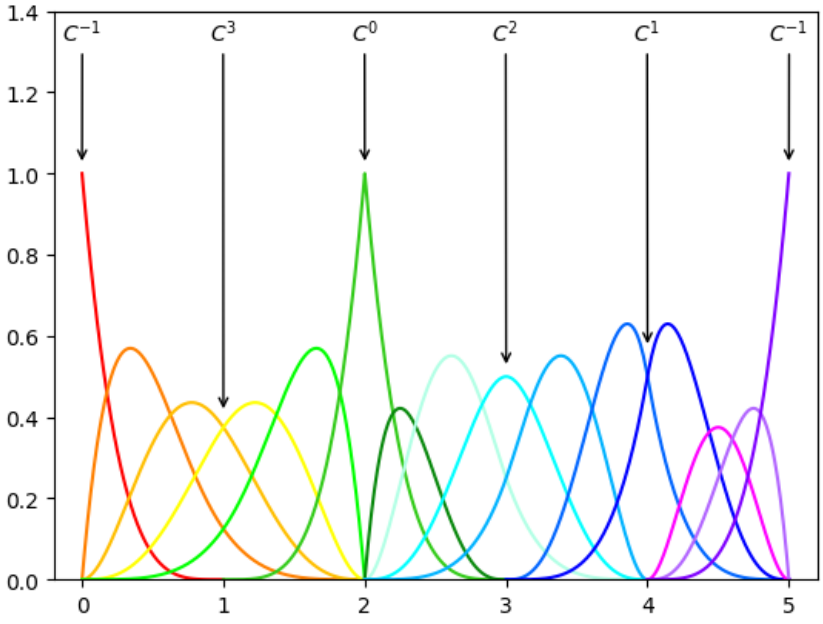
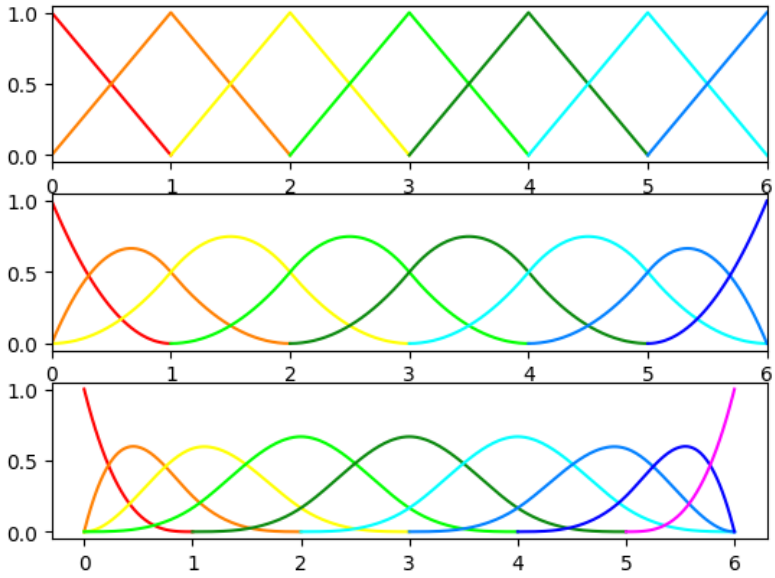
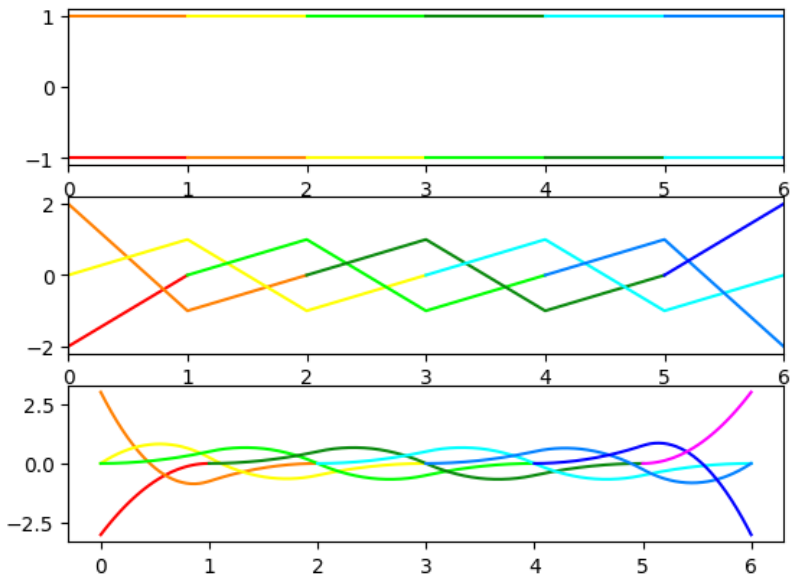


Figure 5. B-splines on $\{0, 0, 0, 0, 0, 1, 2, 2, 2, 2, 3, 3, 4, 4, 4, 5, 5, 5, 5, 5\}$.



(a) Linear, quadratic and cubic B-splines with full continuity.



(b) First-order derivatives of the previous B-splines.

Figure 6. B-splines with full continuity and their corresponding derivatives.

In higher dimensions, we need the extra knot vectors $\mathcal{H}$ and $\mathcal{Z}$. This allows us to define the *bivariate* and *trivariate tensor splines* as

$$s_2(\xi,\eta) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} c_{ij} N_{i,p}(\xi) M_{j,q}(\eta)$$
$$s_3(\xi,\eta,\zeta) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{l=1}^{n_3} c_{ijl} N_{i,p}(\xi) M_{j,q}(\eta) L_{l,r}(\zeta)$$

B-splines have equal polynomial degree p over the whole knot span, but it is possible to make a generalization with *multi-degree splines* (MD-splines). They have sections of different polynomial degree. Most of the standard B-spline theory can be directly extended for these new splines [14, 110].

Spaces of B-splines

The multivariate B-spline spaces in $\mathbb{R}^d$ are based on the idea that we partition the interval $[a_j, b_j]$ uniformly as follows, for all $1 \leq j \leq d$:

$$\Delta_{x_j} = \{x_i^{(j)}\}_{i=0}^{N_j} \quad , \quad a_j = x_0^{(j)} < x_1^{(j)} < \cdots < x_{N_j-1}^{(j)} < x_{N_j}^{(j)} = b_j$$

In 2D and 3D, the domains can be split into disjoint subdomains:

$$\Omega = \bigcup_{i=1}^{N_x}\bigcup_{j=1}^{N_y} I_{ij} \qquad , \quad I_{ij} = [x_{i-1}, x_i] \otimes [y_{j-1}, y_j]$$
$$\Omega = \bigcup_{i=1}^{N_x}\bigcup_{j=1}^{N_y}\bigcup_{l=1}^{N_z} I_{ijl} \quad , \quad I_{ijl} = [x_{i-1}, x_i] \otimes [y_{j-1}, y_j] \otimes [z_{l-1}, z_l]$$

If $I_d = \bigotimes_{i=1}^{d}[a_i, b_i]$ is a hypercube in $\mathbb{R}^d$, and p and k are the polynomial degree and continuity, respectively, then we have

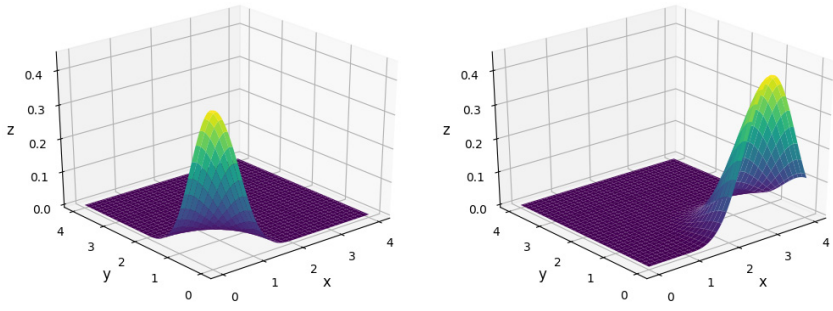
$$\mathbb{S}^{p_1,\ldots,p_d}_{k_1,\ldots,k_d}(\Delta_{x_1},\ldots,\Delta_{x_d}) = \left\{ s_d \,:\, \{c_{i_1,\ldots,i_d}\}_{i_1=1,\ldots,i_d=1}^{n_1,\ldots,n_d} \in \mathbb{R}^d \right\} \tag{5a}$$
$$s_d(\xi_1,\ldots,\xi_d) = \sum_{i_1=1}^{n_1}\cdots\sum_{i_d=1}^{n_d}\left[c_{i_1,\ldots,i_d} \prod_{j=1}^{d} N^{(j)}_{i_j,p_j}(\xi_j) \right] \tag{5b}$$

on the domain I_d. The dimension of the univariate B-spline space [103] is

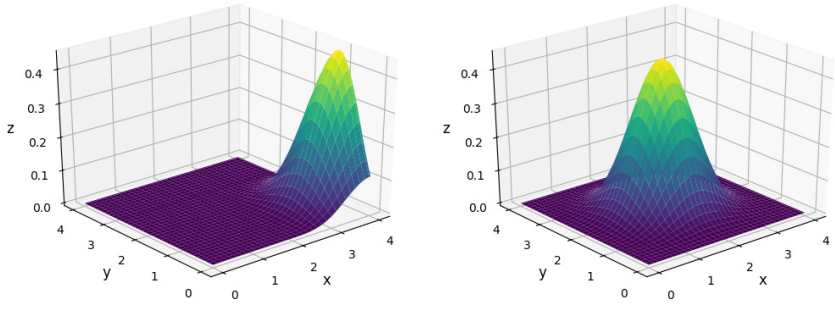
$$\dim\left(\mathbb{S}^p_k(\Delta)\right) = n(p-k) + p + 1 \tag{6}$$

The following multiplicative relation with respect to continuity holds:

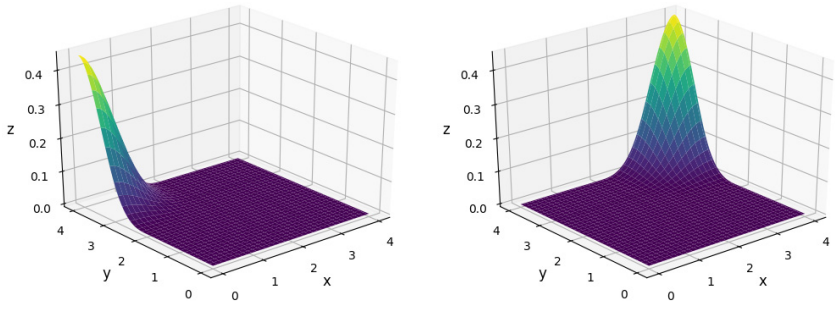
$$f_1 \in \mathbb{S}^{p_1}_{k_1}(\Delta), f_2 \in \mathbb{S}^{p_2}_{k_2}(\Delta) \implies f_1 f_2 \in \mathbb{S}^{p_1+p_2}_{\min(k_1,k_2)}(\Delta) \tag{7}$$



(a) Basis function $N_{2,2,3}$ **(b)** Basis function $N_{2,5,3}$



(c) Basis function $N_{3,6,3}$ **(d)** Basis function $N_{4,4,3}$



(e) Basis function $N_{6,2,3}$ **(f)** Basis function $N_{6,6,3}$

Figure 7. Bivariate B-splines on $\Xi, \mathcal{H} = \{0, 0, 0, 0, 1, 2, 3, 4, 4, 4, 4\}$.

Multivariate B-splines satisfy some general decomposition relations:

$$\mathbb{P}^{p_1,\dots,p_d}(\mathbb{R}^d) = \bigotimes_{i=1}^{d} \mathbb{P}^{p_i}(\mathbb{R}) \tag{8a}$$

$$\mathbb{S}^{p_1,\dots,p_d}_{k_1,\dots,k_d}(\Delta_{x_1},\dots,\Delta_{x_d}) = \bigotimes_{i=1}^{d} \mathbb{S}^{p_i}_{k_i}(\Delta_{x_i}) \tag{8b}$$

$$C^{k_1,\dots,k_d}\left(\bigotimes_{i=1}^{d}[a_i,b_i]\right) = \bigotimes_{i=1}^{d} C^{k_i}([a_i,b_i]) \tag{8c}$$

$$\dim\left(\mathbb{S}^{p_1,\dots,p_d}_{k_1,\dots,k_d}(\Delta_1,\dots,\Delta_d)\right) = \prod_{i=1}^{d} \dim\left(\mathbb{S}^{p_i}_{k_i}(\Delta_i)\right) \tag{8d}$$

Curves, surfaces and volumes

We can construct curves, surfaces and volumes from tensor B-splines by replacing the scalar weights with control points. This provides easier shape manipulation and more flexibility [84]. The tensor product formula is

$$\sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} \left[\prod_{j=1}^{d} N^{(j)}_{i_j,p_j}(\xi_j)\right] \mathbf{P}_{i_1,\dots,i_d} \tag{9}$$

where $\{N^{(j)}_{i_j,p_j}\}_{i_j=1}^{n_j}$ is the set of B-spline basis functions in the ξ_j-direction defined by the knot vector Ξ_j, where $1 \le j \le d$. The set of control points $\{\mathbf{P}_{i_1,\dots,i_d}\}$ form a *control polygon* (*control net* in 2D, *control lattice* in 3D):

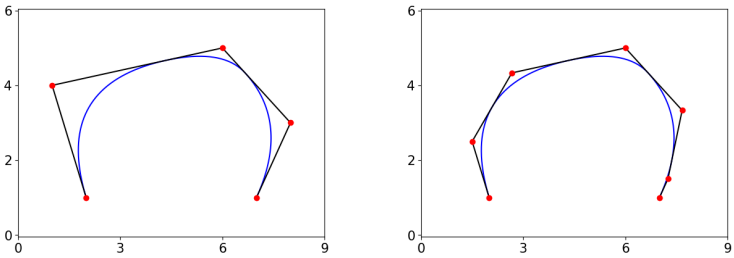
$$\mathbf{CP} = \bigoplus_{i_1=1}^{n_1} \cdots \bigoplus_{i_d=1}^{n_d} \mathbf{P}_{i_1,\dots,i_d} \tag{10}$$

Knot insertion

Knot insertion [91] is a common spline operation. It is directly related to h-refinement. We add the knot $\widehat{\xi} \in [t_s, t_{s+1})$ to a knot vector $\Xi = \{\xi_i\}_{i=1}^{n+p+1}$, obtain a new knot vector $\widehat{\Xi} = \{\xi_1,\dots,\xi_s,\widehat{\xi},\xi_{s+1},\dots,\xi_{n+p+1}\}$, and

$$\mathbf{C}(\xi) = \sum_{i=1}^{n} N_{i,p}\mathbf{P}_i = \sum_{i=1}^{n+1} \widehat{N}_{i,p}\widehat{\mathbf{P}}_i$$

This provides better shape control of the curve. The geometrical shape is preserved although the control polygon changes, as depicted in Figure 8.



(a) Original curve. **(b)** After inserting two knots.

Figure 8. Control polygon comparison for B-spline knot insertion.

According to *Böhm's theorem*, the new control points are defined as follows:

$$\widehat{\mathbf{P}}_i = \begin{cases} \mathbf{P}_i, & 1 \leq i \leq s-p \\ \alpha_i \mathbf{P}_i + (1-\alpha_i)\mathbf{P}_{i-1}, & s-p+1 \leq i \leq s \\ \mathbf{P}_{i-1}, & s+1 \leq i \leq n+1 \end{cases} \tag{11}$$

$$\alpha_i = \frac{\widehat{\xi} - \xi_i}{\xi_{i+p} - \xi_i} \tag{12}$$

The *Oslo algorithm* is a generalized version of this process, and it allows insertion of multiple knots simultaneously.

Degree elevation

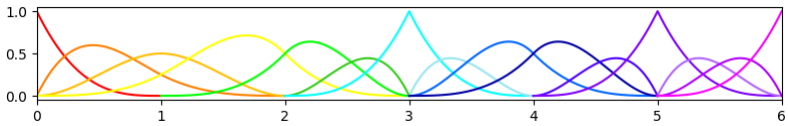
Degree elevation [91] is another spline operation. It involves increasing polynomial order and is directly related to p-refinement. The degree can also be reduced in some cases. Let Ξ be an open knot vector on $[a, b]$ such that we have $a = x_0 < x_1 < \cdots < x_s < x_{s+1} = b$, and $\{m_k\}_{k=1}^{s}$ are the internal knot multiplicities:

$$\Xi = \{\underbrace{a, \ldots, a}_{p+1}, \underbrace{x_1, \ldots, x_1}_{m_1}, \underbrace{x_2, \ldots, x_2}_{m_2}, \ldots, \underbrace{x_s, \ldots, x_s}_{m_s}, \underbrace{b, \ldots, b}_{p+1}\}$$

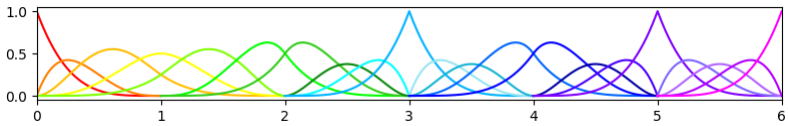
If we elevate or decrease the order, i.e. $p \to p+1$ or $p \to p-1$, then

$$\widehat{\Xi} = \{\underbrace{a, \ldots, a}_{p+2}, \underbrace{x_1, \ldots, x_1}_{m_1+1}, \underbrace{x_2, \ldots, x_2}_{m_2+1}, \ldots, \underbrace{x_s, \ldots, x_s}_{m_s+1}, \underbrace{b, \ldots, b}_{p+2}\}$$

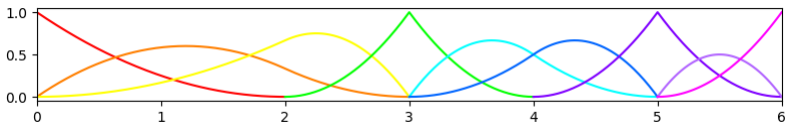
$$\widetilde{\Xi} = \{\underbrace{a, \ldots, a}_{p}, \underbrace{x_1, \ldots, x_1}_{m_1-1}, \underbrace{x_2, \ldots, x_2}_{m_2-1}, \ldots, \underbrace{x_s, \ldots, x_s}_{m_s-1}, \underbrace{b, \ldots, b}_{p}\}$$



(a) Cubic B-splines on $\{0, 0, 0, 0, 1, 2, 2, 3, 3, 3, 4, 4, 5, 5, 5, 6, 6, 6, 6\}$.



(b) Quartic B-splines on $\{0, 0, 0, 0, 0, 1, 1, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 6, 6, 6\}$.



(c) Quadratic B-splines on $\{0, 0, 0, 2, 3, 3, 4, 5, 5, 6, 6, 6\}$.

Figure 9. Degree elevation and reduction of cubic B-spline basis functions.

If we elevate a B-spline curve's degree, we get a new curve with the exactly same parametrization and geometry representation, but the control points change. The process of calculating them is based on the principle that since the two curves are identical, their derivatives of any order are equal despite different control points. This derivative argument acquires an open knot vector where the number of unique knots is $S + 1$. We denote $\{m_i\}_{i=1}^{S-1}$ as the multiplicities of the interior knots and define the auxiliary scalars

$$\beta_i = \sum_{j=1}^{i} m_j \qquad , \qquad 1 \leq i \leq S - 1$$

The control points of the curve's j-th derivative are defined recursively by

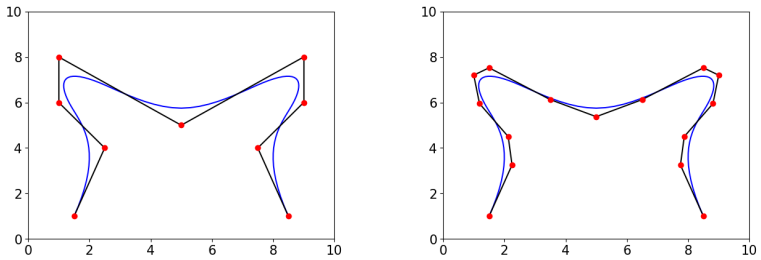
$$\mathbf{P}_i^j = \begin{cases} \frac{p+1-j}{\xi_{i+p+1}-\xi_{i+j}}(\mathbf{P}_{i+1}^{j-1} - \mathbf{P}_i^{j-1}) & , \xi_{i+p+1} > \xi_{i+j} \\ \mathbf{0} & , \text{else} \end{cases} \tag{13}$$

The new curve has degree $p+r$, and the knot vector changes from $\{\xi_i\}_{i=1}^{n+p+1}$ to $\{\xi_i'\}_{i=1}^{n'+p+r+1}$, where $n' = n + Sr$. The set of control points defined through backwards recursion:

$$\mathbf{Q}_{i+1}^{j-1} = \mathbf{Q}_i^{j-1} + \frac{\xi_{i+1+p+r}' - \xi_{i+j}'}{p + r + 1 - j}\mathbf{Q}_i^j \tag{14}$$

Algorithm 2.1 B-spline curve degree elevation

procedure DEGREE_ELEVATION($\{\mathbf{P}_i\}$,p,r,$\{\xi_i\}$,$\{\xi'_i\}$,S,$\{m_i\}$,$\{\beta_i\}$)
  for $1 \leq i \leq n$ **do**
    $\mathbf{P}_i^0 = \mathbf{P}_i$
  for $1 \leq j \leq p$ **do**
    Calculate $\mathbf{P}_1^j$ from (13)
  for $1 \leq i \leq S-1$ **do**
    for $p+1-z_i \leq j \leq p$ **do**
      Calculate $\mathbf{P}_{\beta_i+1}^j$ from (13)
  for $0 \leq j \leq p$ **do**
    $\mathbf{Q}_1^j = \mathbf{P}_1^j$
  for $1 \leq i \leq S-1$ **do**
    for $p+1-z_i \leq j \leq p$ **do**
      $\mathbf{Q}_{\beta_i+1+ir}^j = \mathbf{P}_{\beta_i+1}^j$
  for $1 \leq i \leq S-1$ **do**
    for $1 \leq k \leq r$ **do**
      $\mathbf{Q}_{\beta_i+1+ir+k}^j = \mathbf{Q}_{\beta_i+1+ir}^j$
  for $j = \{p, \ldots, 1\}$ **do**
    for $1 \leq i \leq n'-1$ **do**
      Calculate $\mathbf{Q}_{i+1}^{j-1}$ from (14)
  return $\{\mathbf{Q}_i^0\}$



(a) Original cubic curve. **(b)** Elevation into a quartic curve.

Figure 10. Control polygon comparison for B-spline degree elevation.

A complete description of this algorithm and its underlying mathematical structure can be found in [61].

2.2 Introduction to NURBS

The basics of NURBS

Non-Uniform Rational B-Splines (NURBS) can represent conic sections exactly and enables high-accuracy meshing of curved domains. The generic expression for this spline type is

$$R(\xi) = \sum_{i=1}^{n} R_{i,p}(\xi)\mathbf{P}_i \tag{15}$$

Exact representation is a central characteristic feature of IGA because an accurate mesh reduces the numerical approximation error quite much. This is suitable for defining curves, surfaces and volumes in the same way as tensor B-splines. The general representation becomes

$$\frac{\sum_{i_1=1}^{n_1}\cdots\sum_{i_d=1}^{n_d} w_{i_1,\ldots,i_d}\left[\prod_{j=1}^{d} N_{i_j,p_j}^{(j)}(\xi_j)\right]\mathbf{P}_{i_1,\ldots,i_d}}{\sum_{i_1=1}^{n_1}\cdots\sum_{i_d=1}^{n_d} w_{i_1,\ldots,i_d}\left[\prod_{j=1}^{d} N_{i_j,p_j}^{(j)}(\xi_j)\right]} \tag{16}$$

where w are the *weights*, and the denominator is the *weighting function*. A NURBS curve in $\mathbb{R}^d$ is the projection of a $(d+1)$-dimensional B-spline curve [40]. The evaluation requires a projection $\Pi : \mathbb{R}^d \longmapsto \mathbb{PR}^d$ on the control points, for all $i \in [1, n]$:

$$\mathbf{P}^i = (x_i, y_i, z_i) \longmapsto \mathbf{Q}^i = (w_i x_i, w_i y_i, w_i z_i, w_i) \tag{17}$$

The new B-spline curve expressed by the original knot vector Ξ and the projected points $\{\mathbf{Q}^i\}_{i=1}^{n}$ can be evaluated directly with the help of Cox-de Boor recursion. Thus, the inverse projection $\Pi^{-1} : \mathbb{RP}^d \longmapsto \mathbb{R}^d$ required for the final evaluation becomes, for all $i \in [1, n]$:

$$\mathbf{Q}_e^i = (\widetilde{x}_i, \widetilde{y}_i, \widetilde{z}_i, \widetilde{w}_i) \longmapsto \mathbf{P}_e^i = \frac{1}{\widetilde{w}_i}(\widetilde{x}_i, \widetilde{y}_i, \widetilde{z}_i) \tag{18}$$

The *real projective space*, $\mathbb{RP}^d$, is a d-dimensional manifold with quotient topology, consisting of every $x \in \mathbb{R}^{d+1}$ such that x and αx define the same point when $\alpha \neq 0$. It is also used for describing the properties of NURBS [40, 79]. The derivative of NURBS basis functions is given by

$$R'_{i,p}(\xi) = \frac{N'_{i,p}(\xi)W(\xi) - N_{i,p}(\xi)W'(\xi)}{(W(\xi))^2} \tag{19}$$

The process of evaluating and differentiating NURBS can be generalized to higher dimensions with matrix tensor products, and all the characteristic properties still hold. For further details, we refer to [91].

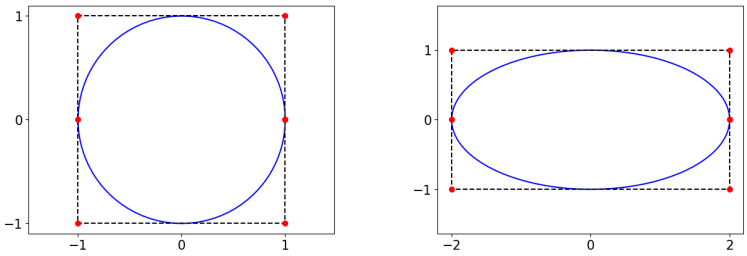


Figure 11. Two curves generated by NURBS.

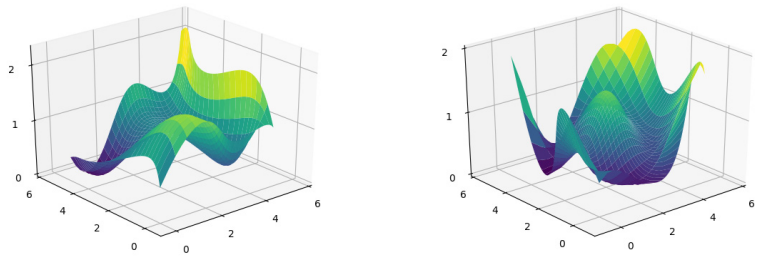
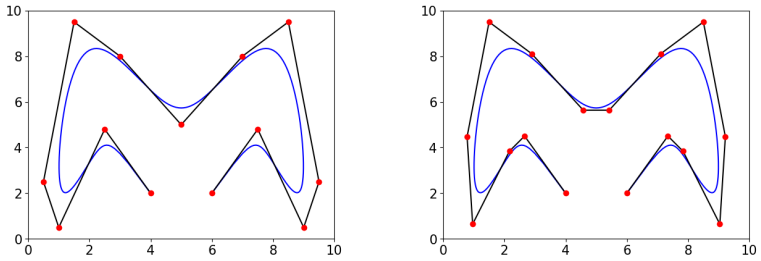


Figure 12. Two surfaces generated by NURBS.

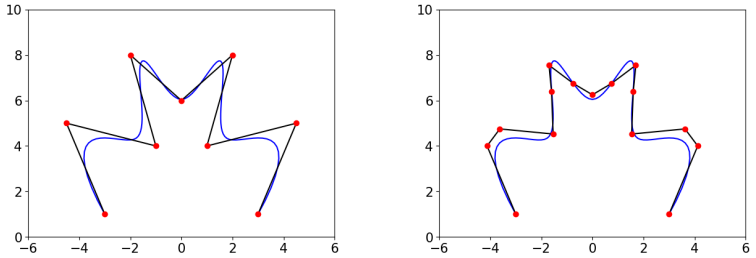
Like B-splines, NURBS have many special properties [91]:

1. **Uniqueness**: A NURBS $R_{i,p}$ depends only on the knots $\{\xi_j\}_{j=i}^{i+p+1}$.
2. **Local support**: $\text{supp}(R_{i,p}) = (\xi_i, \xi_{i+p+1})$.
3. **Positivity**: $\xi \in (\xi_i, \xi_{i+p+1}) \implies R_{i,p} > 0$.
4. **Continuity**: NURBS are smooth rational polynomials between the knots. On a knot with multiplicity m, the continuity is C^{p-m}.
5. **Unique maximum**: If $p > 0$, then $R_{i,p}$ has one unique maximum.
6. **Stability**: The NURBS basis is stable and linearly independent.
7. **Partition of unity**: $\sum_{i=1}^{n} R_{i,p}(\xi) = 1$.
8. **Nonsingularity**: All derivatives of $R_{i,p}$ exist in the knot span interior.
9. **Invariance**: NURBS are always invariant of scaling, rotation, shear, translation and projection.



(a) Original curve. **(b)** After inserting three knots.

Figure 13. Control polygon comparison for NURBS knot insertion.



(a) Original cubic curve. **(b)** New quartic curve.

Figure 14. Control polygon comparison for NURBS degree elevation.

Projection of geometric algorithms

Knot insertion and degree elevation can also be applied to NURBS in the same way as B-spline curves. The only difference is that we must invoke the projection $\Pi : \mathbb{R}^d \longmapsto \mathbb{PR}^d$ described earlier for evaluating NURBS curves and surfaces. After doing so, we carry out the desired geometric algorithm to the projected vector function and obtain new projected control points. The real physical control points are then found straightforwardly by the inverse projection $\Pi^{-1} : \mathbb{RP}^d \longmapsto \mathbb{R}^d$.

2.3 Introduction to LR B-splines

Types of meshes

A *box-mesh* (T-mesh) is a mesh where the domain is split into smaller rectangles (2D) or boxes (3D). A box in $\mathbb{R}^d$ is defined as $\beta = \bigtimes_{i=1}^{d} J_i$, where $J_i = [a_i, b_i]$. This can also be written as $\beta = [\mathbf{a}, \mathbf{b}]$. We call J_i trivial if $a_i = b_i$. The dimension of β is the number of nontrivial intervals, i.e. $\dim \beta = \#\{a_i < b_i\}$, and we have the following rules:

1. If $\dim \beta = l$, then β is called an (l, d)-box.
2. If $\dim \beta = d$, then β is called an element.
3. If $\dim \beta = d - 1$, then β is called a k-mesh rectangle, where k is the index of the trivial interval J_k.

The boxes has some important properties:

1. A d-box contains $2^{d-1}\binom{d}{l}$ l-boxes, where $0 \leq l \leq d$.
2. A mesh-rectangle $\lambda = [\mathbf{c}, \mathbf{e}]$ is the face of a d-box $[\mathbf{a}, \mathbf{b}]$ if
   (a) $c_k = a_k < b_k = e_k$ (nontrivial)
   (b) $c_k = e_k = a_k$ or $c_k = e_k = b_k$ (trivial)
3. The boundary of a d-box β is the union of its faces. The interior is β^0.

The last property can be compactly stated as follows:

$$\partial\beta = \bigcup_{\substack{1 \leq i \leq d \\ a_i < b_i}} J_1 \times \widehat{J}_i \times J_d \qquad , \qquad \widehat{J}_i = \{a_i, b_i\}$$

If Ω is a d-box in $\mathbb{R}^d$, and $\mathcal{E}$ is a box partition, then

1. For every $\beta_1, \beta_2 \in \mathcal{E}$, we have $\beta_1^0 \cap \beta_2^0 = \emptyset$.
2. $\beta^0 = \beta \backslash \partial\beta$, and $\bigcup_{\beta \in \mathcal{E}} = \Omega$.

The intersection of boxes in $\mathcal{E}$ containing $\mathbf{q} \in \mathbb{R}^d$ is denoted as

$$\beta_{\mathbf{q}}(\mathcal{E}) = \bigcap_{\beta \in \mathcal{E}, \mathbf{q} \in \beta} \beta$$

It is also common to use the auxiliary set Ω^+, defined as

$$\Omega^+ = \left\{ \bigtimes_{i=1}^{d} J_i : J_k \in \{[a_k - 1, b_k], [a_k, b_k], [a_k, b_k + 1]\}, \forall k \right\} \backslash \Omega$$

Then $\mathcal{E} \cap \Omega^+$ is a box partition of $\bigtimes_{i=1}^{d}[a_i - 1, b_i + 1]$. We have four sets:

$$\mathcal{F}(\mathcal{E}) = \bigcup_{\mathbf{q}\in\Omega} \{\beta_{\mathbf{q}}(\mathcal{E} \cap \Omega^+)\}$$
$$\mathcal{F}^0(\mathcal{E}) = \bigcup_{\mathbf{q}\in\Omega^0} \{\beta_{\mathbf{q}}(\mathcal{E})\}$$
$$\mathcal{F}_l(\mathcal{E}) = \{\beta \in \mathcal{F}(\mathcal{E}) : \dim \beta = l\}$$
$$\mathcal{F}_l^0(\mathcal{E}) = \{\beta \in \mathcal{F}^0(\mathcal{E}) : \dim \beta = l\}$$

For $k \in [1, d]$, $\mathcal{F}_{d-1,k}(\mathcal{E})$ is a set of k-mesh-rectangles in $\mathcal{F}_{d-1}(\mathcal{E})$. A *box mesh* on $[\mathbf{a}, \mathbf{b}]$, $\mathcal{M} = \mathcal{F}_{d-1}(\mathcal{E})$, is a minimal collection of $(d-1)$-boxes. A μ-extended box mesh $(\mathcal{M}, \mu)$ has an associated integer $\mu(\lambda)$ for all $\lambda \in \mathcal{M}$, where $\mu : \mathcal{M} \longmapsto \mathbb{N}$. Tensor meshes have no T-joints, so horizontal and vertical lines span the entire length in each direction. If $a_{k,1} < a_{k,2} < \cdots < a_{k,n_k}$, the tensor-mesh is given by

$$\mathcal{E} = \left\{ \bigtimes_{i=1}^{d} [a_{k,i_k}, a_{k,i_k+1}] : 1 \le i_k \le n_k - 1, 1 \le k \le d \right\}$$

A μ-extended tensor-mesh is a μ-extended box-mesh $(\mathcal{M}, \mu)$ where $\mathcal{M}$ is a box-mesh, and $\mu(\gamma) = \mu(\gamma')$ if γ and γ' are in the same hyperplane. The tensor-mesh expansion $\mathcal{M}^T$ of $\mathcal{M}$ is the smallest tensor-mesh containing $\mathcal{M}$, and the map $\mu^T : \mathcal{M}^T \longmapsto \mathbb{Z}^+$ is an extension of μ such that

$$\mu^T(\beta) = \begin{cases} \mu(\gamma), & \beta \subseteq \gamma \in \mathcal{M} \\ 0, & \beta \nsubseteq \gamma, \forall \gamma \in \mathcal{M} \end{cases}$$

We call $(\mathcal{M}^T, \mu^T)$ the μ-extended tensor-mesh expansion of $(\mathcal{M}, \mu)$.

Define a mesh-rectangle γ and a d-box β in $\mathbb{R}^d$. If $\beta \backslash \gamma$ is not connected, then γ splits β. The split is minimal if $\gamma \subseteq \beta$. $\beta \backslash \gamma$ has two connected components β_1 and β_2, and $X_{\beta,\gamma} = \{\overline{\beta}_2, \overline{\beta}_2\}$ is the closure. Assume that $\mathcal{E}$ is a box partition on a d-box Ω, and γ is a mesh-rectangle, both of them in $\mathbb{R}^d$. We say that γ splits $\mathcal{E}$ if it is a finite union of mesh-rectangles, where γ_i is either a split of a box in $\mathcal{E}$ or a mesh-rectangle in $\mathcal{M}(\mathcal{E})$. If $\mathcal{E}_1$ is the set of all boxes in $\mathcal{E}$ split by γ, $\mathcal{E}_2 = \mathcal{E} \backslash \mathcal{E}_1$, and $\mathcal{M} = \mathcal{F}_{d-1}(\mathcal{E})$, then

$$\mathcal{E} + \gamma = \mathcal{E}_2 \cup \left(\bigcup_{\beta\in\mathcal{E}} X_{\beta,\gamma} \right) \tag{21a}$$
$$\mathcal{M} + \gamma = \mathcal{F}_{d-1}(\mathcal{E} + \gamma) \tag{21b}$$

Thus, we can express the μ-extension of $\beta \in \mathcal{M} + \gamma$ as

$$\mu^T(\beta) = \begin{cases} 1, & \beta \not\subseteq \beta' \\ \mu(\beta') + 1, & \beta \subseteq \beta' \subseteq \gamma \\ \mu(\beta'), & \beta \subseteq \beta' \not\subseteq \gamma \end{cases}$$

for all $\beta' \in \mathcal{M}$. We call γ a constant split of $(\mathcal{M}, \mu)$ with multiplicity $\mu(\gamma)$ if $\mu(\gamma) = \mu_\gamma(\beta)$ for all $\beta \in \mathcal{M} + \gamma$ satisfying $\beta \subseteq \gamma$.

A μ-extended LR-mesh is a μ-extended box-mesh $(\mathcal{M}, \mu)$ where either one of the following criterions are satisfied:

1. $(\mathcal{M}, \mu)$ is a μ-extended tensor-mesh.
2. $(\mathcal{M}, \mu) = (\widetilde{\mathcal{M}} + \gamma, \widetilde{\mu}_\gamma)$, where $(\widetilde{\mathcal{M}}, \widetilde{\mu})$ is a μ-extended LR-mesh and γ is a constant split of it.

If $\{\epsilon_i\}_{i=1}^n$ is a collection of line insertions such that $\mathcal{M}_{i+1} = \mathcal{M}_i \cap \epsilon_i$, then the LR-meshes generate the sequence $\mathcal{M}_0 \subset \cdots \subset \mathcal{M}_n$. A meshline ϵ traverses the support of $B : \mathbb{R}^2 \longmapsto \mathbb{R}$ if one of the properties below hold:

1. If $\xi_0^* \leq \xi_0$, $\xi_{p_1+1} \leq \xi_1^*$ and $\eta_0 \leq \eta^* \leq \eta_{p_2+1}$, then $\epsilon = [\xi_0^*, \xi_1^*] \times \eta^*$.
2. If $\xi_0 \leq \xi^* \leq \xi_{p_1+1}$, $\eta_0^* \leq \eta_0$ and $\eta_{p_2+1} \leq \eta_1^*$, then $\epsilon = \xi^* \times [\eta_0^*, \eta_1^*]$.

LR B-spline

If $(\mathcal{M}, \mu)$ is a μ-extended box-mesh, $\mathbf{q} \in \mathbb{R}^d$, $X \subset \mathbb{R}^d$, and $1 \leq k \leq d$, then we have the following rules:

$$\mu_k(\mathbf{q}) = \max\left(\{0\} \cup \{\mu(\gamma) : \mathbf{q} \in \gamma \in \mathcal{F}_{d-1,k}(\mathcal{M})\}\right)$$
$$\nu(X) = \inf\left\{\mu_k(\{\mathbf{q}\}) : \mathbf{q} \in X\right\}$$

A tensor B-spline B has support on $(\mathcal{M}, \mu)$ if

$$m_{B_k(t)} \leq \nu\left(\operatorname{supp}(B) \cap \phi_{k,t}\right) \qquad \forall t \in \operatorname{supp}(B_k)$$

where $m_{B_k(t)}$ is the knot multiplicity of B, and $\phi_{k,t} = \mathbb{R}^{k-1} \times \{t\}\mathbb{R}^{d-k}$. If this holds with equality for $t \in \operatorname{supp}(B_k)^0$, we have minimal support.

Let $(\mathcal{M}, \mu)$ be a μ-extended LR-mesh. $B : \mathbb{R}^d \longmapsto \mathbb{R}$ is an LR B-spline if B is a tensor B-spline with minimal support on $(\mathcal{M}, \mu)$. For further details on this topic, we refer to [35, 19, 69].

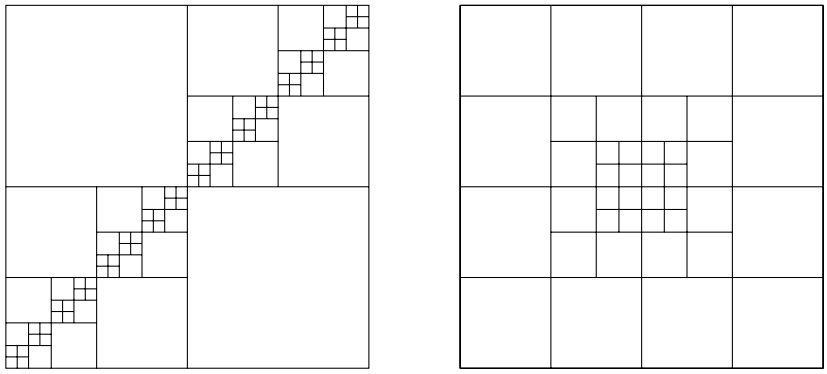
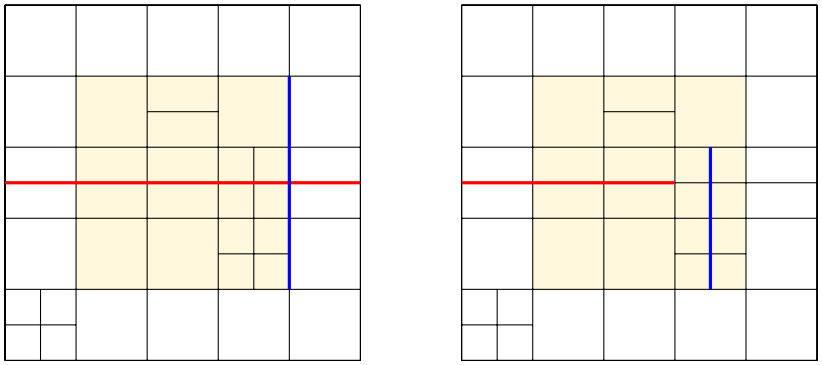


Figure 15. Two examples of LR-meshes in two dimensions.



(a) Meshlines traversing the support of a tensor B-spline

(b) Meshlines not traversing the support of a tensor B-spline

Figure 16. Illustration of meshlines on the support of a tensor B-spline.

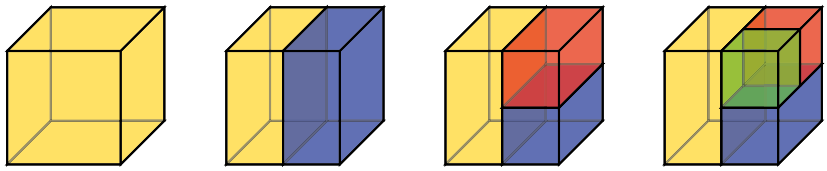
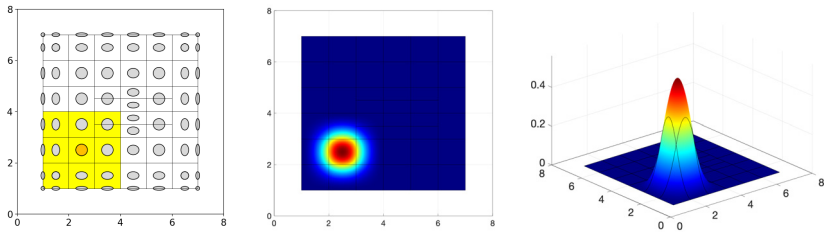
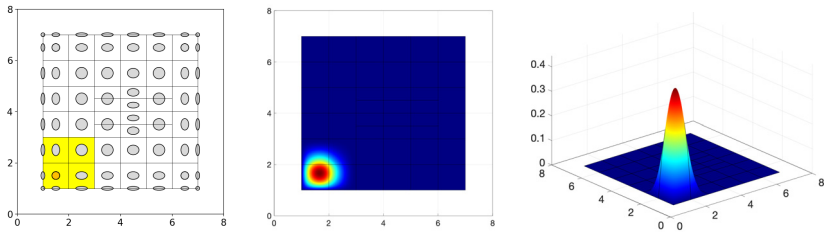


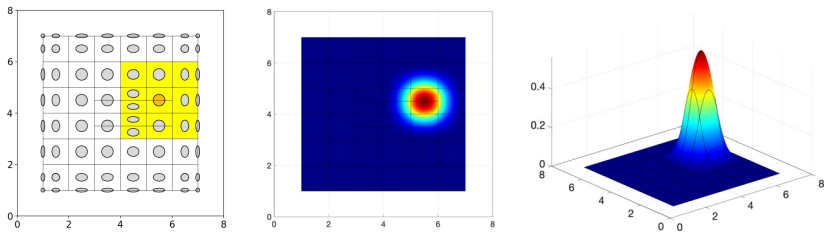
Figure 17. Construction of an LR-mesh in three dimensions.



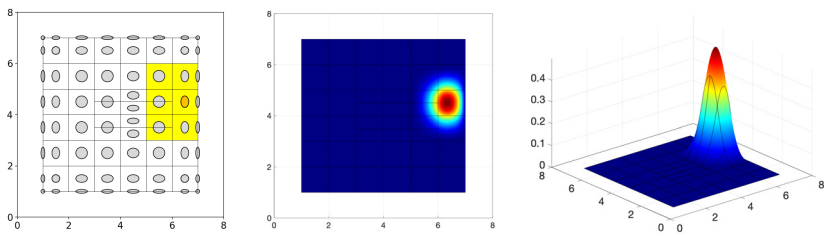
(a) LR B-spline B[1-2-3-4;1-2-3-4].



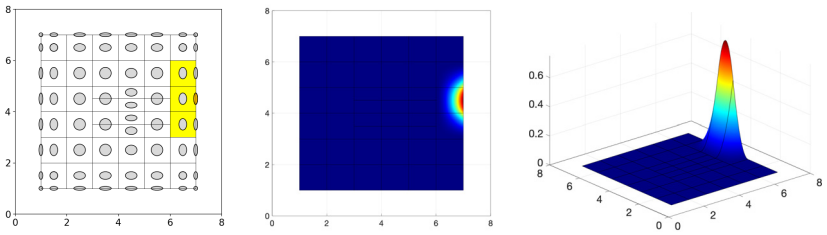
(b) LR B-spline B[1-1-2-3;1-1-2-3].



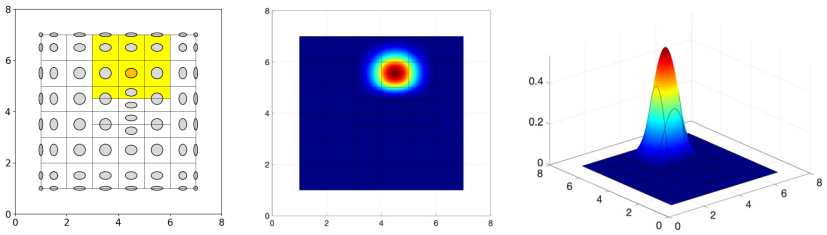
(c) LR B-spline B[4-5-6-7;3-4-5-6].



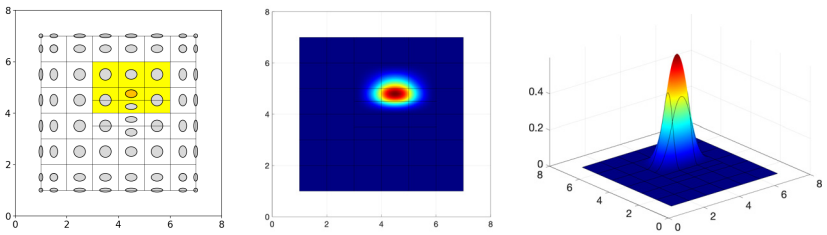
(d) LR B-spline B[5-6-7-7;3-4-5-6].



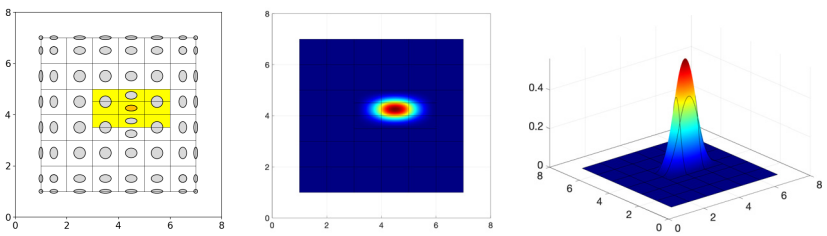
(e) LR B-spline B[6-7-7-7;3-4-5-6].



(f) LR B-spline B[4.5-5-6-7;3-4-5-6].

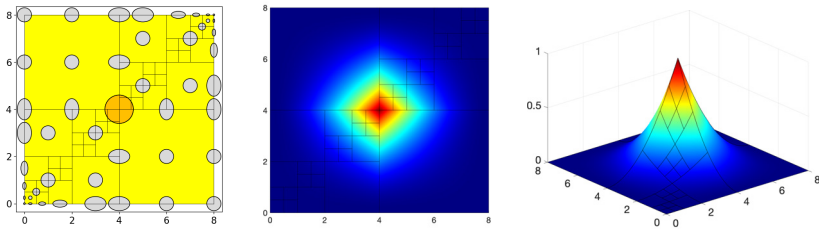


(g) LR B-spline B[4-4.5-5-6;3-4-5-6].

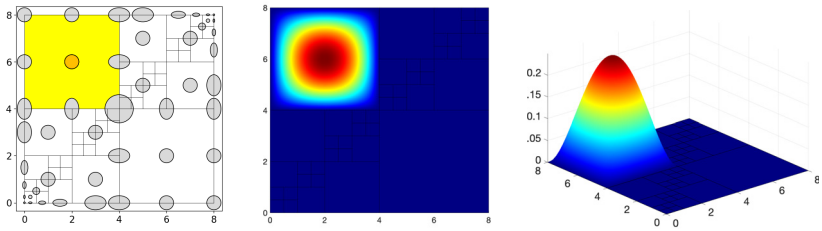


(h) LR B-spline B[3.5-4-4.5-5;3-4-5-6].

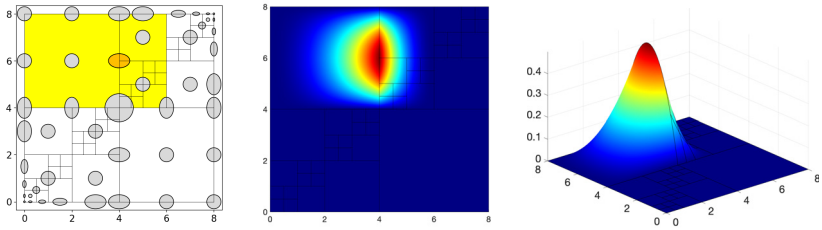
Figure 19. Support, top view and perspective view of some selected LR B-spline basis functions in the spline space $\mathbb{S}_1^2(\mathcal{M})$. The sizes of the ellipses centred at the Greville points correspond to the size of the chosen basis functions' support.



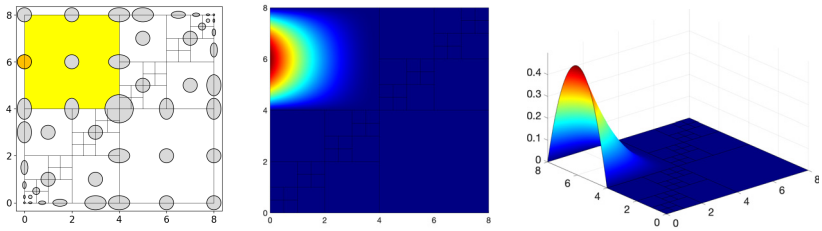
(a) LR B-spline B[0-4-4-8;0-4-4-8].



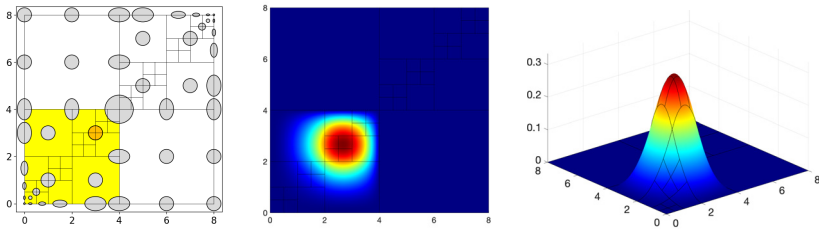
(b) LR B-spline B[0-0-4-4;4-4-8-8].



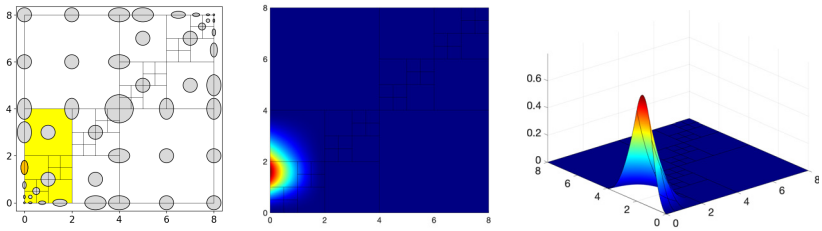
(c) LR B-spline B[0-4-4-6;4-4-8-8].



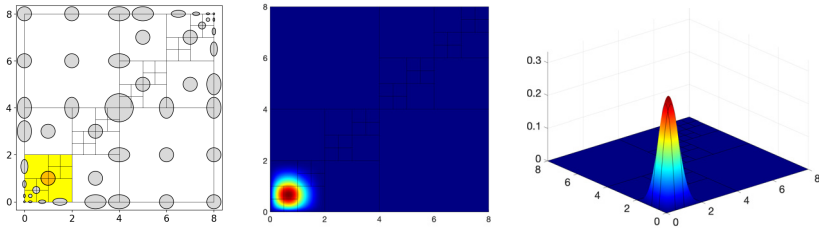
(d) LR B-spline B[0-0-0-4;4-4-8-8].



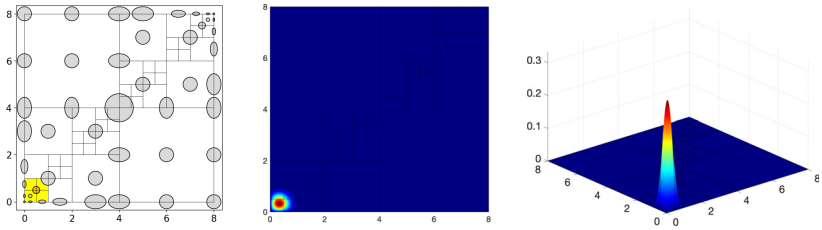
(e) LR B-spline B[0-0-4-4;0-0-4-4].



(f) LR B-spline B[0-0-2-2;0-2-2-4].



(g) LR B-spline B[0-0-2-2;0-0-2-2].



(h) LR B-spline B[0-0-1-1;0-0-1-1].

Figure 21. Support, top view and perspective view of some selected LR B-spline basis functions in the spline space $\mathbb{S}_0^2(\mathcal{M})$. The sizes of the ellipses centred at the Greville points correspond to the size of the chosen basis functions' support.

2.4 Topological and geometrical aspects

In a general setting, we assume that a spline is defined on an d-dimensional simplicial or polyhedral subdivision $\Delta \subseteq \mathbb{R}^d$. The dimension of $\mathbb{S}_k^p(\Delta)$ depends not solely on the continuity k and polynomial degree p, but also on geometrical, combinatorial and topological properties of Δ. This is a very complicated problem. In many cases, we can only make estimating bounds on the dimension formula of spline spaces, which requires advanced techniques from homological algebra, commutative algebra, and algebraic geometry [1, 81, 87]. Recently, several articles have confirmed how to apply the theory of spline modules and homologies to IGA [33, 34, 98, 99]. The celebrated index theorem from algebraic topology has also been related to splines [26, 93].

3 Finite Element Modelling

This section is emphasizing the main differences between classical finite element modelling and the new methods used in the isogeometric framework for doing the same tasks in the discretization process. Most of the implementation methodology is similar for both approaches, but IGA has some important and quite advantageous distinctions.

3.1 General theory of Ritz-Galerkin discretization

Assume that $\mathcal{L}$ is a linear partial differential operator of order s_1, f is an inhomogeneous term, B is a boundary condition operator of order $s_2 \in [0, s_1 - 1]$ (s_2 is the highest-order derivative in the boundary conditions), and g is a piecewise function for each boundary segment. Then the boundary value problem can be expressed as

$$\mathcal{L}(u) = f \qquad \mathbf{x} \in \Omega \tag{23a}$$
$$B(u) = g \qquad \mathbf{x} \in \partial\Omega \tag{23b}$$

When the PDE above is equipped with a well-posed BVP/IVP, we have the *strong formulation*. It has a *strong solution* satisfying $u \in C^{s_1}(\Omega) \cup C^{s_2}(\overline{\Omega})$. In any FEM-approach, we apply *Galerkin projection* where equation (23) is multiplied with a sufficiently smooth test function v, and then we integrate. This yields the *weak formulation*, which is solving an integral equation where the highest order derivative is reduced by a half:

$$\int_\Omega \mathcal{L}(u) v \, d\Omega = \int_\Omega f v \, d\Omega \tag{24}$$

The exact solution u belongs to a *trial space* V, and the numerical solution u_h belongs to the finite-dimensional space V_h. If $V_h \subset V$, our method is *conforming*. In the *Ritz-Galerkin discretization*, we express u_h as a linear combination of shape functions that are constructed from a linearly independent basis $\mathcal{P}$:

$$u_h(\mathbf{x}) = \sum_{m \in \mathcal{P}} u_m \psi_m(\mathbf{x}) = \mathbf{\Psi}^T \mathbf{u} \tag{25}$$

In this setting, v belongs to a *test space* W. The choice of the shape functions distinguishes the different FEM-approaches from each other. When $V \equiv W$, we have the standard *Bubnov-Galerkin method*, which works for most PDEs. This formulation is optimal if the operator $\mathcal{L}$ is self-adjoint.

If the PDE is advection-dominated, the *Petrov-Galerkin method* is better because it can stabilize boundary layers and preclude spurious oscillation. This is common for PDEs with odd-order derivatives. We usually add a perturbation function to the test function v such that W is decomposed in two parts. In general, the procedure of determining the perturbation depends on the PDE itself [29].

$$\text{Bubnov-Galerkin:} \qquad \int_\Omega \mathcal{L}(u) v \, d\Omega = F(v) \tag{26a}$$

$$\text{Petrov-Galerkin:} \qquad \int_\Omega \mathcal{L}(u) \left(v + \widetilde{v}\right) \, d\Omega = F(v + \widetilde{v}) \tag{26b}$$

Usually, the perturbation $\widetilde{v}$ is the derivative of v multiplied with a special tuning parameter, depending on our stabilization approach. This parameter is adjusted such that artificial diffusion is added in a controlled way.

3.2 Boundary value problems

Boundary conditions for second order PDEs

Most PDEs are subject to boundary conditions. For second order PDEs, there are three standard types of boundary conditions:

Dirichlet (essential):	$u = g_D$	$\mathbf{x} \in \partial\Omega_D$
Neumann (natural):	$\dfrac{\partial u}{\partial n} = g_N$	$\mathbf{x} \in \partial\Omega_N$
Robin (mixed):	$\alpha \dfrac{\partial u}{\partial n} + \beta u = g_R$	$\mathbf{x} \in \partial\Omega_R$

We assume that $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N \cup \partial\Omega_R$, and all these boundary segments are mutually disjoint. For higher order PDEs, there is a numerous amount of different BVP types, but they will not be considered here.

Dirichlet conditions describe displacement and are directly enforced in the solution, making them *essential*. Let us assume that the discretized equation system is $\mathbf{A}\mathbf{u} = \mathbf{f}$. We remove rows and columns corresponding to $\partial\Omega$, solve the modified system, and then we insert the boundary conditions into the solution vector at the entries representing $\partial\Omega$.

Neumann conditions describe flux and are automatically included in the weak formulation, making them *natural*. We loop over the element edges belonging to $\partial\Omega$, add their contributions into an extra vector $\mathbf{h}$, and define the right-hand side of the equation system as $\mathbf{f}+\mathbf{h}$. We do not need to modify the matrix and vectors after solving the extended system $\mathbf{A}\mathbf{u} = \mathbf{f} + \mathbf{h}$.

Robin conditions describe radiation on the boundary. They represent a combination of Dirichlet and Neumann conditions, hence *mixed*. In this case, the matrix $\mathbf{A}$ is modified by adding another matrix $\mathbf{H}$ which is zero everywhere except at those entries corresponding to $\partial\Omega$. Like the Neumann conditions, we are done after solving the system $(\mathbf{A}+\mathbf{H})\mathbf{u} = \mathbf{f}+\mathbf{h}$.

Trace lifting

After applying Galerkin projection on (23), which has order 2, the weak solution belongs to $H^1(\Omega)$. For homogenous Dirichlet problems ($g_D = 0$), we simply choose $u \in H_0^1(\Omega) \subset H^1(\Omega)$. For inhomogeneous conditions ($g_D \neq 0$), the resulting space is not a closed subset of $H^1(\Omega)$, so we need a *trace lifting* $\widetilde{w} \in H^1(\Omega)$ that equals $g_D = 0$ on $\partial\Omega$. This lifting is not unique in general, but it can be expressed in terms of a *trace operator*, which is a continuous linear mapping $\gamma_0 : H^1(\Omega) \longmapsto H^{1/2}(\partial\Omega)$ that satisfies $\gamma_0 u = g_D$ [104]. We obtain the decomposition $u = u_0 + w$ where $u_0 \in H_0^1(\Omega)$, and the modified equation system becomes $\mathbf{A}\mathbf{u}_0 = \mathbf{f} - \mathbf{A}\mathbf{w}$. The right-hand side must be calculated first before we remove the rows and columns corresponding to $\partial\Omega$.

If the boundary conditions are both mixed and inhomogeneous, then the weak formulation gets a new extended form:

$$\int_\Omega \mathcal{L}(u_0)v\,d\Omega = \int_\Omega fv\,d\Omega - \int_{\partial\Omega_N} g_N\gamma_0 v\,ds - \int_\Omega \mathcal{L}(w)v\,d\Omega \qquad (27)$$

The second integral on the right-hand side must always define a linear and continuous functional, so the set of admissible Neumann data g_N is the dual of the set of traces $\gamma_0 v$. Thus, the range of γ_0 on $\partial\Omega_N$ is $H_{00}^{1/2}(\Omega_N)$, a linear closed subspace of $H^{1/2}(\Omega_N)$, and the admissible Neumann data belongs to the dual space $(H^{1/2}(\Omega_N))'$ [104].

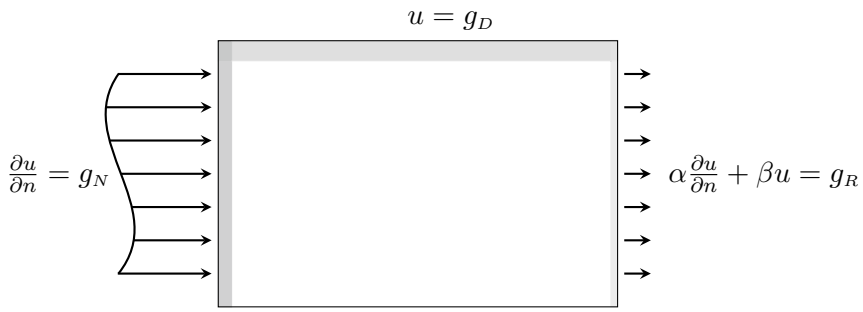


Figure 22. Illustration of different boundary conditions.

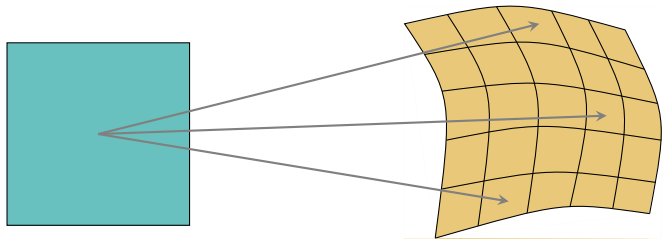
3.3 Assembly process

In the finite element assembly, we loop over every element and store the locally computed integrals into the matrices and vectors. This will require an efficient enumeration of the elements and an appropriate mapping which ensures correct insertion. Using the same notation as Hughes [65], we can express the assembly process as

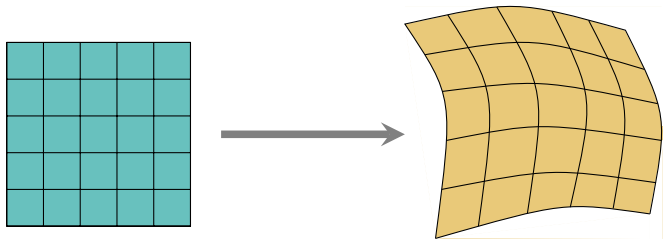
$$\mathbf{M} = \mathop{\mathsf{A}}_{e=1}^{n_e} (\mathbf{M}_e) \qquad , \qquad \mathbf{f} = \mathop{\mathsf{A}}_{e=1}^{n_e} (\mathbf{f}_e) \tag{28}$$

In classical FEM, we have isoparametric mapping on the individual elements, but in IGA, the mapping works on entire patches of the global domain. This is very time-saving because we do not need many different mappings, and it simplifies the implementation. We distinguish between three spaces:

Physical space:	Ω_e	(x, y, z)
Parameter space:	$\widehat{\Omega}_e$	(ξ, η, ζ)
Parent element:	$\widetilde{\Omega}_e$	$(\widetilde{\xi}, \widetilde{\eta}, \widetilde{\zeta})$



(a) Isoparametric mapping using classical FEM



(b) Isoparametric mapping using IGA

Figure 23. Comparison of isoparametric mappings in the finite element methods.

We have already assumed that $u_h = \mathbf{\Psi}^T \mathbf{u}$, where $\mathbf{\Psi}$ is the vector of all the shape functions. In IGA, we also need numerical quadrature in the assembly process for evaluating matrices and vectors in the discrete system. Exact analytical integration is too complicated and inefficient. We must also map the integrand on the local physical domain Ω_e bijectively to the parent element $\widetilde{\Omega}_e$. By using the chain rule, we can describe the coordinate transform of the mapping $g : \Omega_e \longmapsto \widetilde{\Omega}_e$ as

$$\begin{bmatrix} \frac{\partial \psi_i}{\partial \widetilde{\xi}} \\ \frac{\partial \psi_i}{\partial \widetilde{\eta}} \\ \frac{\partial \psi_i}{\partial \widetilde{\zeta}} \end{bmatrix} = \begin{bmatrix} \frac{\partial x}{\partial \widetilde{\xi}} & \frac{\partial y}{\partial \widetilde{\xi}} & \frac{\partial z}{\partial \widetilde{\xi}} \\ \frac{\partial x}{\partial \widetilde{\eta}} & \frac{\partial y}{\partial \widetilde{\eta}} & \frac{\partial z}{\partial \widetilde{\eta}} \\ \frac{\partial x}{\partial \widetilde{\zeta}} & \frac{\partial y}{\partial \widetilde{\zeta}} & \frac{\partial z}{\partial \widetilde{\zeta}} \end{bmatrix} \begin{bmatrix} \frac{\partial \psi_i}{\partial x} \\ \frac{\partial \psi_i}{\partial y} \\ \frac{\partial \psi_i}{\partial z} \end{bmatrix} = \mathbf{J} \nabla \psi_i \tag{29}$$

Here, $\mathbf{J}$ is the Jacobian. We use the chain rule again:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \xi}{\partial \widetilde{\xi}} & 0 & 0 \\ 0 & \frac{\partial \eta}{\partial \widetilde{\eta}} & 0 \\ 0 & 0 & \frac{\partial \zeta}{\partial \widetilde{\zeta}} \end{bmatrix} \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} & \frac{\partial z}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} & \frac{\partial z}{\partial \eta} \\ \frac{\partial x}{\partial \zeta} & \frac{\partial y}{\partial \zeta} & \frac{\partial z}{\partial \zeta} \end{bmatrix} = \mathbf{AB} \tag{30}$$

By using the compact notations $\{\xi, \eta, \zeta\}$ and $x_j = \{x, y, z\}$, we can express all the entries of the matrix $\mathbf{B}$ as follows:

$$B_{ij} = \frac{\partial x_j}{\partial \xi_i} = \sum_{k=1}^{N} \frac{\partial \psi_k}{\partial \xi_i} x_j^{(k)} \tag{31}$$

We express the determinant of $\mathbf{J}$ as

$$\begin{aligned} \det(\mathbf{J}) &= \det(\mathbf{A}) \det(\mathbf{B}) \\ &= \frac{\partial \xi}{\partial \widetilde{\xi}} \times \frac{\partial \eta}{\partial \widetilde{\eta}} \times \frac{\partial \zeta}{\partial \widetilde{\zeta}} \times \det(\mathbf{B}) \end{aligned}$$

The transformed integrand evaluated by Gaussian quadrature is defined as

$$\begin{aligned} A_e &= \int_{\Omega} G(x, y, z)\, d\Omega \\ &= \int_{-1}^{1} \int_{-1}^{1} \int_{-1}^{1} G(\widetilde{\xi}, \widetilde{\eta}, \widetilde{\zeta}) |\det(\mathbf{J})|\, d\widetilde{\xi}\, d\widetilde{\eta}\, d\widetilde{\zeta} \\ &\approx \sum_{i=1}^{m_x} \sum_{j=1}^{m_y} \sum_{k=1}^{m_z} w_i w_j w_k G(\widetilde{\xi}_i, \widetilde{\eta}_j, \widetilde{\zeta}_k) |\det(\mathbf{J})| \end{aligned}$$

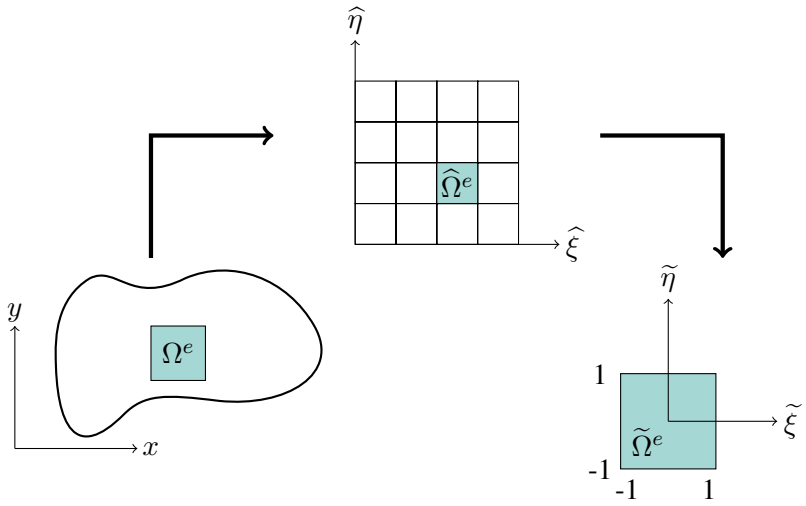


Figure 24. Mapping between different spaces in the quadrature process.

Splines form a smooth subspace of C^0, so Gaussian quadrature will work as long as we have enough quadrature points. But this scheme ignores high continuity between the elements, so it integrates more than necessary, and the computational increases a lot. A possible solution is using optimal quadrature algorithms which calculate nodes and weights for individual knot vectors. This approach, combined with homotopy continuation for ensuring that nothing fails, makes the running time of the assembly drop down from $\mathcal{O}(N^d)$ to $\mathcal{O}((N/2)^d)$. It works for tensor-splines [67]. The present point of view is that the quadrature should be based on the type of splines in the isogeometric discretization. Several articles have been discussing this relevant topic [57, 94, 113].

3.4 Comparison of FEM and IGA

In classical FEM, we have only a single mesh on the domain, but in IGA, there are two distinctive meshes because we are using splines [29]. The geometry of the domain is decomposed into local patches with their own individual knot spans to make the discretization flexible, and this is referred to as the *physical mesh*. The splines' control points are used for proper adjustment of the geometry, and they constitute the *control mesh*. These control nets and control lattices can be viewed respectively as quadrilaterals and hexahedrons combined.

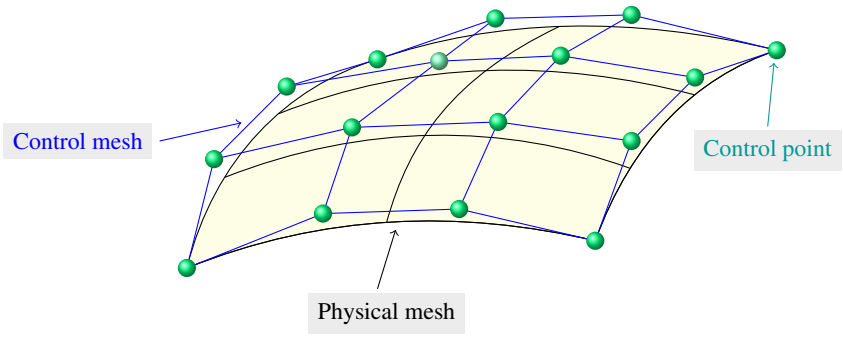


Figure 25. Illustration of physical mesh and control mesh.

As we see in Figure 25, we have a physical mesh and control mesh in IGA. The power of IGA is exact meshing. In classical FEM, we find basis functions for interpolating the numerical solution first, and then we use them to create a mesh, which is not exact in general. This is reversed in IGA, which provides a geometry-independent framework. We pick appropriate functions to represent the domain exactly, and then we apply them to approximate the unknown solution. Creating curved edges is not possible in FEM, for the elements are straight-edged, as shown in Figure 26. The exact mesh reduces the numerical error.

Isoparametric elements are of high importance because it is required that we use the same shape functions for modelling the solution and meshing the domain. In IGA, splines are isoparametric for any continuity, unlike FEM. The spline parameter space is local to a patch on the domain, not a single individual element. Since the total number of maps depends just on the patches, it simplifies the whole implementation and saves much time. The geometry-to-mesh mappings are also more efficient compared to FEM.

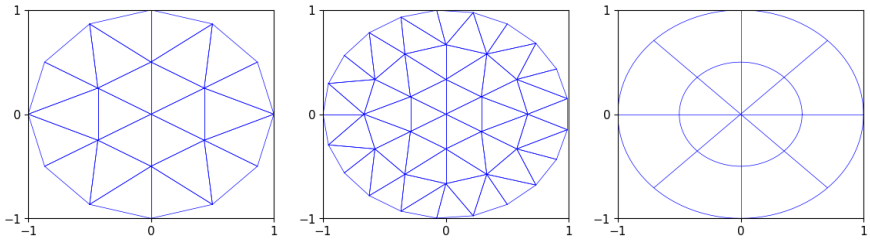
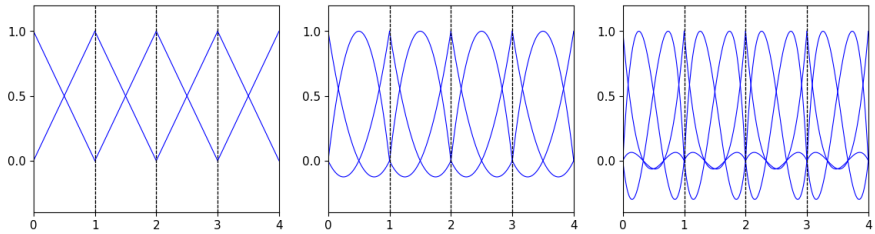
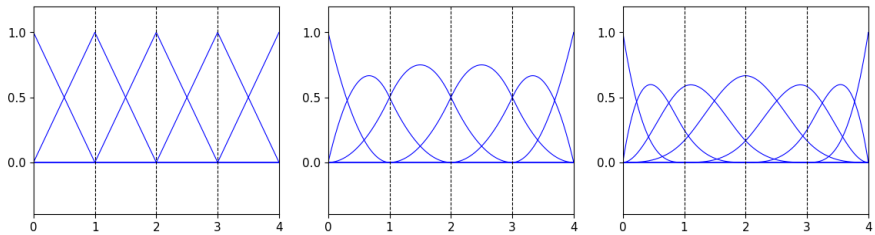


Figure 26. Comparison of meshing in IGA and FEM. Although the number of degrees of freedom increases, the two first meshes (FEM) cannot represent the circular plate exactly. Only the third mesh (IGA) with NURBS can do it.



(a) Linear, quadratic and cubic Lagrange basis functions.



(b) Linear, quadratic and cubic B-spline basis functions.

Figure 27. Comparison of first-, second- and third-degree Lagrange and B-spline basis functions on the interval $[0, 4]$, split into four elements of the same size. Only the B-splines have a uniform continuity pattern without jumps.

High continuity of IGA provides great accuracy because the solution's increased smoothness reduces the global error. As we see from Figure 27, the Lagrange interpolants have C^0-continuity at $\{1, 2, 3\}$, which is invariant of increasing the polynomial degree p. This might generate high error oscillations and spurious error propagation. But the B-splines can be adjusted to have continuity between 0 and $p-1$, and global C^{p-1}-continuity is optimal. We achieve greater accuracy per degree of freedom, and the high continuity creates more overlapping between the elements. Thus, the discretization matrices are not so large and have lower spectral radius, which increases the speed of iterative algorithms.

Splines provide computational stability because they are nonnegative and form a partition of unity. Since they also have the *variation diminishing approximation property*, IGA can also tackle discontinuous data better than the previous finite element approaches. This new enhancement yields better opportunities to reduce error oscillation and smooth out discontinuities. The effect is maximal for full continuity. Other basis functions from older finite element techniques do not share this advantageous property of splines.

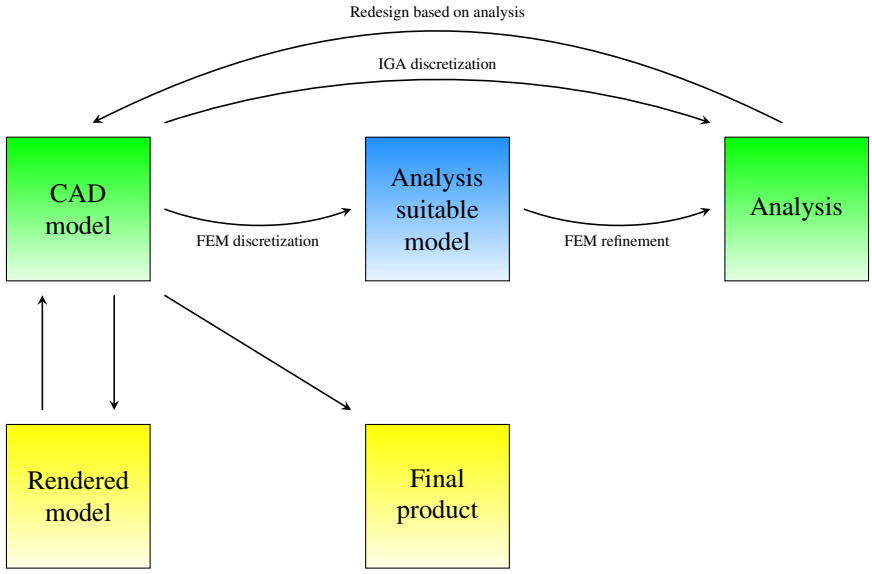


Figure 28. Comparison of FEM and IGA framework: IGA avoids the step in FEM that requires an analysis suitable model by using the discretization from CAD directly.

As we see from Figure 28, IGA enables us to create an exact CAD model of the domain, and then we can solve a PDE on it. This is possible because we use the same spline basis for the geometry and the unknown solution field, and the splines are isoparametric for any continuity, unlike FEM. It is unecessary to translate the CAD framework to an analysis suitable model before solving the PDE, as required in the old FEM paradigm.

According to Hughes [29], the CAD-translation from FEM is far from trivial in complex engineering designs and takes 80 % of the analysis time. As described earlier, FEM does not provide accurate geometry and mesh adaptivity when the continuity increases. This makes high convergence and precision cumbersome to obtain. But the choice of splines as basis functions creates direct and efficient interoperability between CAD and FEA, and we can jump directly to the analysis stage without translating the CAD model. This is one of the most important reasons why IGA could be more appropriate for future industrial applications.

Table 1. The most important differences between FEM and IGA.

Finite Element Method	**Isogeometric Analysis**
The domain's geometry is simply approximated	The domain's geometry is exactly represented
Single physical mesh	Physical mesh and control mesh
Nodal points on the mesh	Control points on the control mesh
Mesh consists of elements	Mesh consists of knot spans
Isoparametric mapping on single elements	Isoparametric mapping on entire patches
Basis functions interpolate the nodal points and variables	Basis functions do not interpolate the control points and variables
Basis functions can take any value	Basis functions are nonnegative
Degrees of freedom at the nodes	Degrees of freedom at the control points
Continuity between elements is C^0	Continuity between elements can be adjusted from C^0 to C^{p-1}
h- and p-refinement available	h-, p- and k-refinement available
Discontinuous data causes spurious oscillation	Discontinuous data is smoothed out due to variation diminishing property
Discretized system takes long time to solve iteratively	Discretized system takes shorter time to solve iteratively
The solution is defined by nodal variables	The solution is defined by control variables
Basis support is over a patch where elements share a common node	Basis support is over a rectangular array of knot spans whose sizes are depending on the basis continuity

4 Mesh generation

Constructing a suitable mesh on the physical domain is an important and challenging task in every finite element approach. A well-constructed mesh will accelerate the convergence of numerical accuracy. There is not a single universal algorithm for creating a suitable mesh because the procedure is always situation dependent. But there are many helpful criterions that we should try to satisfy as best as possible to create a mesh that does indeed work for the chosen PDE.

Since the meshes used in the simulations are implemented from scratch, we find it convenient to present the main ideas of the meshing. The technique is a flexible combination of other methods that work well for our problem, and they are applied in such a way that they can mesh different parts of the global domain based on the local geometric shape and subdivisions.

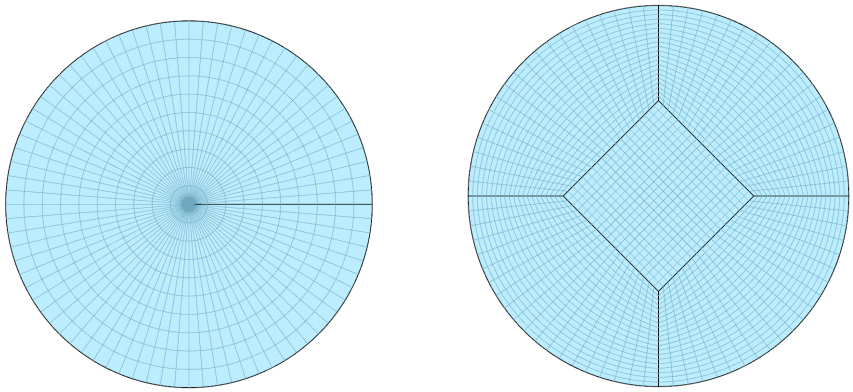
4.1 Multiple patching on a conformal mesh

Our first approach is utilizing *multiple patches* [29, 45]. A patch is a subdomain where we can define the mesh freely as we want. If Ω is the global domain, then $\{\Omega_i\}_{i=1}^{P}$ is a collection of mutually disjoint and quadrangular patches such that their disjoint union equals the whole domain Ω:

$$\Omega = \bigcup_{i=1}^{P} \Omega_i \quad , \quad i \neq j \implies \Omega_i \cap \Omega_j = \emptyset \tag{32}$$

The major advantage with multiple patching is the opportunity to create the mesh on each patch in such a way that it becomes consistent with the local geometric shape. Every patch has their own individual mesh. But the individual meshing on every patch Ω_i must always be done in such a way that the whole global mesh on the domain Ω becomes *conformal*. This means that the vertices of each element never touch the edges of the adjacent elements. Every element must coincide with each other just at the vertices to make the mesh regular.

Another advantage of multiple patching is avoiding *singularities* on the global mesh. In a singularity, the gradient of the numerical solution tends to infinity, and this generates high error in the finite element approximation. A good strategy to avoid this defect is analysing the entire domain first, localizing the singularities on forehand, and then use multiple patching to eliminate them. This makes the global mesh sufficiently regular, and we prevent the numerical accuracy from being deteriorated.



(a) Single patch with singularity **(b)** Multiple patch without singularity.

Figure 29. A circle with two different types of mesh.

We illustrate an introductory example of multiple patching by creating meshes on the unit disc in two different ways. The first mesh is commonly used in the *Finite Difference Method*, the Finite Element Method's ancestor. For some PDEs, this mesh works fine for finite differences, but not with finite elements. This is because of the singularity at the origin. With the help of multiple patching, we can put a square at the middle, rotate it $45°$, and divide the resting area in four equal parts. Then we can refine all the five patches and obtain a conformal mesh without any singularity. This improves the finite element approach.

If we are going to perform adaptive refinement on a complex domain which must to be partitioned into several patches, it is best to let the initial mesh be conformal and very coarse. Then, we can carry out the refinement such that the continuity along the common facets shared by the patches is preserved. Thus, we maintain the same degree of continuity and improve the global approximation.

4.2 The importance of geometric continuity in IGA

In multiple patching, *geometric continuity* is a central concept. It provides a formal definition of isogeometric elements and is used to match geometric invariants. Following the approach as in [54], we start by introducing *jets*. Given $m, d, k \in \mathbb{N}$ and $s \in \mathbb{R}^m$, we define the set of pairs

$$\mathcal{F}_{s,d} = \{(f, \mathcal{N}) \,:\, \mathcal{N} \text{ is an open neighbourhood of } s,\ f \in C^k(\mathcal{N})\} \quad (33)$$

Let $D^{\boldsymbol{\alpha}}$ be the partial differential operator of multi-index $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$, and define the equivalence relation $\sim_s^k$ on $\mathcal{F}_{s,d}$ by

$$(f_1, \mathcal{N}_1) \sim_s^k (f_2, \mathcal{N}_2) \quad , \quad \text{if } D^{\boldsymbol{\alpha}} f_1(s) = D^{\boldsymbol{\alpha}} f_2(s)\,,\ |\boldsymbol{\alpha}| \leq k \tag{34}$$

The equivalence class $\mathbf{j}_s^k f$ of f under $\sim_s^k$ is a k-jet of f at s. For $i = \{1, 2\}$, we define $\square_i \subset \mathbb{R}^d$ as a d-dimensional polytope, and E_i is an $(d-1)$-dimensional facet of $\square_i$. Let $\mathcal{N}_i \subset \mathbb{R}^d$ be an open set such that $\mathcal{N}_i \subset \text{int}(E_i)$, and define a C^k-diffeomorphism $\rho : \mathcal{N}_1 \mapsto \mathcal{N}_2$ by

$$\rho(\text{int}(E_1)) = \text{int}(E_2) \tag{35a}$$
$$\rho(\mathcal{N}_1 \cup \text{int}(\square_1)) = \mathcal{N}_2 \backslash \square_1 \tag{35b}$$
$$\rho(\mathcal{N}_1 \cup \square_1) = \mathcal{N}_2 \backslash \text{int}(\square_1) \tag{35c}$$

Lastly, we let $\mathbf{x}_1 : \square_i \mapsto \mathbb{R}^d$ be C^k-maps such that

$$\mathbf{x}_2(\rho(s)) = \mathbf{x}_1(s) \quad , \quad \forall s \in \text{int}(E_1) \tag{36}$$

We say that $\mathbf{x}_1$ joins $\mathbf{x}_2$ with continuity G^k and reparametrization ρ along the common interface $E = \mathbf{x}_1(E_1) = \mathbf{x}_2(E_2)$ if

$$\mathbf{j}_s^k \mathbf{x}_1 = \mathbf{j}_s^k (\mathbf{x}_2 \circ \rho) \quad , \quad \forall s \in \text{int}(E_1) \tag{37}$$

Let $\{\mathbf{G}^{(i)} : [0,1]^3 \mapsto \mathbb{R}^3,\ 1 \leq i \leq P\}$ be a finite set of bijective regular geometry mappings defined by $(\xi^{(i)}, \eta^{(i)}, \zeta^{(i)}) \mapsto (G_1^{(i)}, G_2^{(i)}, G_3^{(i)})$. Thus, we can rewrite equation (32) as follows:

$$\Omega = \bigcup_{i=1}^{P} \mathbf{G}^{(i)}([0,1]^d) \tag{38}$$

In the isoparametric approach, we construct a local function $u^{(i)} : \Omega_i \mapsto \mathbb{R}^d$ such that $u^{(i)} \circ (\mathbf{G}^{(i)})^{-1}$ solves the PDE on Ω_i. When the scalar components of $u^{(i)}$ and $\mathbf{x}_i$ are from the same function space, then $u^{(i)} \circ (\mathbf{G}^{(i)})^{-1}$ is an isogeometric element [72]. Furthermore, when every patch of the domain are matched together with such G^k-mappings, the isogeometric elements will be C^k. For a more detailed description, we start by defining the space

$$V^{(k)} = \{u \in C^k(\Omega) : u\big|_{\Omega_i} \in \mathbb{S} \circ (\mathbf{G}^{(i)})^{-1}\,,\ 1 \leq i \leq P\}$$

where $\mathbb{S}$ is the parametric spline space $\mathbb{S} \circ (\mathbf{G}^{(i)})^{-1}$ is the corresponding physical space on patch Ω_i. Thus, the solution u of a PDE on Ω satisfies

$$u\big|_{\Omega_i}(\mathbf{x}) = u^{(i)}(\mathbf{x}) = (U^{(i)} \circ (\mathbf{G}^{(i)})^{-1})(\mathbf{x}) \quad , \quad \mathbf{x} \in \Omega^{(i)}$$

Here, $U^{(i)} \in \mathbb{S}$ is a parametric spline. Then, for any disjoint patches Ω_1 and

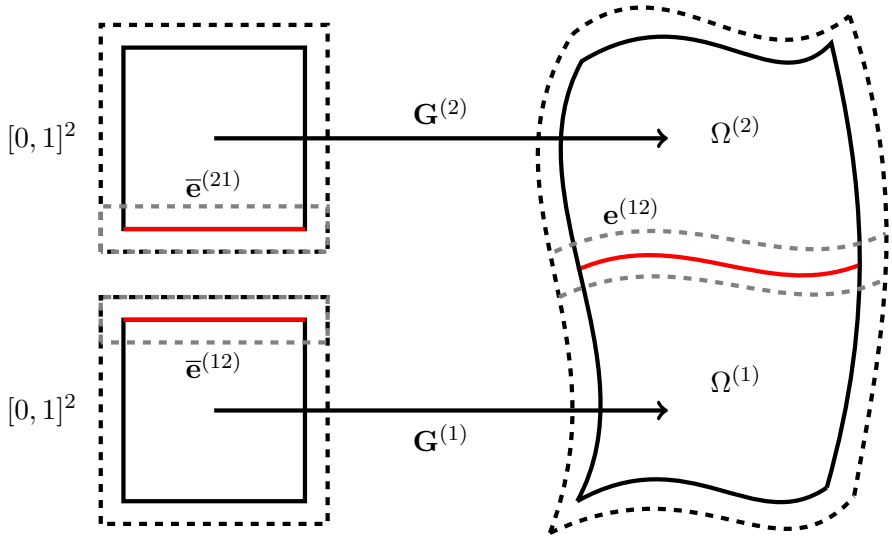


Figure 30. A G^k-mapping joining patches in $\mathbb{R}^2$ into one domain.

Ω_2 sharing a common edge $\mathbf{e}^{(12)}$, we require that

$$D^{(\alpha)}u^{(1)}(\mathbf{x}) = D^{(\alpha)}u^{(2)}(\mathbf{x}) \quad , \quad \mathbf{x} \in \mathbf{e}^{(12)} \, , \, |\boldsymbol{\alpha}| \leq k \tag{39}$$

We denote the sets of vertices and edges respectively as $\mathcal{N}$ and $\mathcal{E}$. The mesh $\mathcal{M}$ is a *topological surface* if we have

- A collection $\{K_i\}$ of polygons (elements) that are pairwise disjoint.
- A collection $\{\phi_{ij}\}$ of homeomorphisms between disjoint polygons K_1 and K_2 sharing a common face.

If ϕ_{ij} and its inverse ϕ_{ji} are C^1-diffeomorphisms, then they are *transition maps*. Together with the common shared face between the elements, they constitute a *gluing structure*. The relationship between isogeometric domain decomposition and commutative algebra has been studied extensively by Blidia et al. in [15]. It has also been shown that the space $V^{(k)}$ can be decomposed in a special way with respect to vertices and edges [72].

4.3 Solid Modelling representation

The multiple patching technique can be directly transferred from 2D to 3D. In addition, we will use *solid modelling representation* [10] to represent the domain Ω efficiently. The first method, *constructive solid geometry* (CSG), is constructing Ω block by block from physical primitives. Typical blocks for our purposes are boxes, both with straight and curvilinear faces.

Boxes with one skew face follows another procedure named *boundary representation* (BRep). We define the top and bottom as parallel rectangles at different heights, and then we fill the gap between them by a *regularized union*. This means that lower-dimensional features of the solid are discarded, and we restrict the volume by constraining the rectangles. For boxes with circular top and bottom, we must first parametrize the two curvilinear areas before lofting the volume between them to obtain a solid 3D-block.

In any case, surfaces and volumes are defined explicitly. We also apply *spatial subdivision*. The complex domain is split in several nonoverlapping and disjoint parts on forehand. These minor parts are redefined in terms of straight or curvilinear solids. Then, we combine them together and refine their individual meshes, which yields a regular and conformal mesh.

The figures below are examples on how to utilize the BRep-technique for lofting the volume between two surfaces and then generate a solid. The strength of this approach is that the top and bottom do not have to be congruent geometric objects, only similar in the shape.

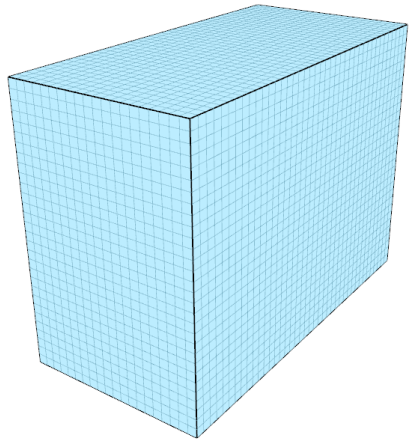
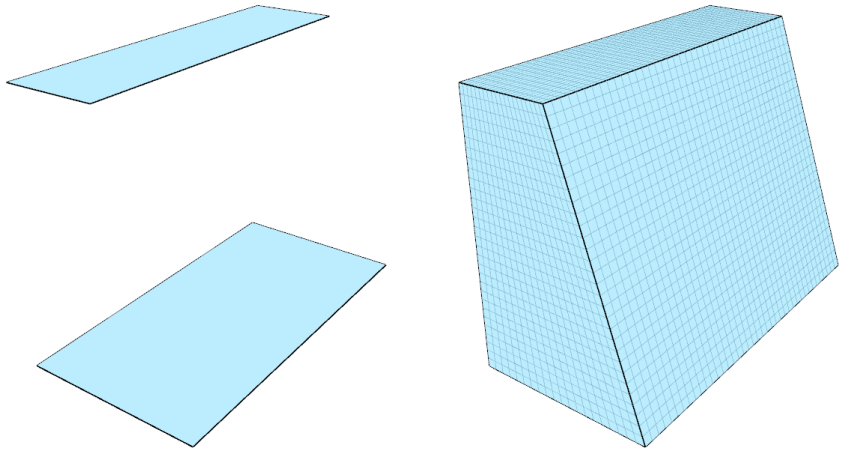
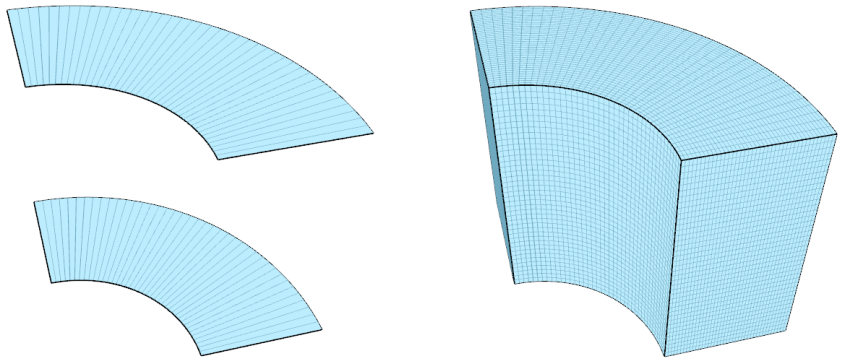


Figure 31. Box with uniform mesh



(a) Top and bottom of box **(b)** Box with a skew edge

Figure 32. Creating a box with skew edges by lofting between the top and bottom



(a) Top and bottom **(b)** Box with circular form

Figure 33. Creating a box with circular top and bottom by lofting between them

5 Summary of papers

Report: Error Estimation in Isogeometric Analysis

Author: Abdullah Abdulhaque

This paper is a summary of the self-study course *Error Estimation and Adaptive Finite Element Methods*, which I took during the autumn of 2016 under the guidance of professor Trond Kvamsdal. The main textbooks of this course were the monographs *A Posteriori Error Estimation in Finite Element Analysis* by M. Ainsworth and J. T. Oden, and *A Posteriori Error Estimation Techniques for Finite Element Methods* by Rüdiger Verfürth.

After reading these books and related material, the main task was showing how to adapt as many as possible of the classical error estimation techniques to IGA. There are also some small numerical simulations for verifying the quality and reliability of the a posteriori error estimators.

Article 1: A Posteriori Error Estimates for Isogeometric Analysis of the Stokes Equation

Authors: Abdullah Abdulhaque, Trond Kvamsdal, Kjetil André Johannessen, Mukesh Kumar, Arne Morten Kvarving

In this paper, we solve the Stokes equation with adaptive isogeometric refinement, using residual and recovery estimators. Both are derived from scratch and analysed thoroughly. In this process, we demonstrate how these a posteriori estimators, originally developed in the context of classical FEM, are fully compatible with IGA and can be adapted for higher continuity than C^0. They remain the same for Taylor-Hood, Sub-Grid Taylor-Hood and Raviart-Thomas elements.

We perform numerical simulations at the end to examine how well the estimators work. In these experiments, we test some benchmark problems with analytical solutions, making it possible to compare the exact numerical error with the estimated error. The estimated errors' behaviour is then used to determine the residual and recovery estimators' quality. All the three proposed discretization elements are tested for each estimator.

Article 2: A Posteriori Error Estimation for Isogeometric Analysis of the Navier-Stokes equation

Authors: Abdullah Abdulhaque, Trond Kvamsdal, Mukesh Kumar, Arne Morten Kvarving

This paper is a direct continuation of the previous one. We examine how to solve the incompressible Navier-Stokes equation with adaptive isogeometric refinement. The focus is illustrating how to extend the same methodology of the first paper to handle the nonlinearity occurring in the new equation. Hence, the new residual estimator is just a slight extension of its linear counterpart, but the recovery estimator is totally unchanged.

At the end, we examine some benchmark problems with analytical solutions to determine which estimator is the best. The new simulations will take some longer time because of the Newton-iteration required for solving the discrete equation systems, but the post-processing remains the same. We will test Taylor-Hood, Sub-Grid Taylor-Hood and Raviart-Thomas elements like we did previously for the Stokes equation.

Article 3: Error Estimation for Isogeometric Analysis of the Advection-Diffusion-Reaction equation

Authors: Abdullah Abdulhaque, Trond Kvamsdal, Mukesh Kumar, Arne Morten Kvarving

The third paper concerns how to solve the Advection-Diffusion-Reaction equation with adaptive isogeometric refinement. This time, the solution process is more straightforward compared to the previous articles because we have a single linear PDE, not a system. The derivation and analysis of the residual and recovery estimators for this PDE are not so cumbersome either. We will only use splines with full continuity since they provide the best approximation.

In the numerical simulations, we examine smooth benchmark problems with boundary and interior layers. We vary a couple of parameters and illustrate how singular perturbation affects the behaviour of the estimators, and then we demonstrate how SUPG-discretization combined with IGA works for these problems.

Article 4: Adaptive Isogeometric Analysis of the Boussinesq Equations for Buoyancy-Driven Flow

Authors: Abdullah Abdulhaque, Trond Kvamsdal, Mukesh Kumar, Arne Morten Kvarving

This paper covers the main topic of the thesis. We combine the previous strategies for solving the Navier-Stokes and Advection-Diffusion equations together to solve the stationary incompressible Boussinesq equation for buoyancy-driven flow. Although this is a large PDE system with several couplings between the subequations, it is possible to estimate the errors of each individual equation and then add them together. There is no need for proving robustness since we are combining estimators that have been proven to be robust, which simplifies the methodology.

For the numerical simulations, we will test two problems with analytical solutions. Lastly, we examine a special benchmark problem without that feature. The numerical results will be compared with other articles instead, and much of the post-processing is implemented entirely from scratch.

6 Software development

Various programming languages, software and computers have been used for the numerical simulations and software development in this thesis. We present them briefly and explain how the simulations work.

6.1 Computer facilities

Programming languages

- *Python*: Used for creating convergence plots, figures in the thesis, tailored post-processing algorithms for each PDE, and for writing g2-files in the mesh generation (secondary IFEM-input).
- *MATLAB*: Used for plotting the graph of LR B-splines basis functions.
- *XML*: Used for writing xinp-files for defining BVPs, solution methods and post-processing in the simulations (primary IFEM-input).
- *Shellscript*: Used to write bash-files for running multiple simulations and post-processing tasks simultaneously. Runs on UNIX systems.

Software

- *IFEM*: Open-source software developed by SINTEF for solving a variety of PDEs with IGA, written in C++ and FORTRAN. https://github.com/OPM/IFEM
- *SpliPy*: Open-source software library created by SINTEF for defining spline-based meshes on physical domains, written in Python. https://github.com/sintefmath/Splipy
- *lrbsplines*: Open-source software library created by SINTEF similar to SpliPy, using LR B-splines to permit local refinement. https://github.com/TheBB/lrsplines-python/blob/master/lrsplines.pyx
- *HDFView*: Open-source software from The HDF Group, used to analyse hdf-files visually.
- *GLview Inova*: Proprietary software from Ceetron ASA for visualizing the numerical solution of PDEs, written in OpenGL.

Hardware

- *Personal computer* - All the simulation and visualization codes are written, tested and updated on this machine before being transferred to larger computers for simulation. It has 2 cores (Intel Core i7).
- *Markov* - Mainframe at Department of Mathematical Sciences, NTNU, used for simulations. It has one unit with 28 cores (Intel(R) Xeon(R) CPU E5-2690 v4).
- *Syvert* - Mainframe at Department of Mathematical Sciences, NTNU, used for simulations. It has three units:
  - *Syvert 0* - 24 cores (Intel(R) Xeon(R) CPU X7542).
  - *Syvert 1* - 8 cores (Intel(R) Xeon(R) CPU E5-2637 v3).
  - *Syvert 2* - 32 cores (Intel(R) Xeon(R) CPU E5-4650 0).
- *Afem* - Workstation at Department of Applied Mathematics and Cybernetics, SINTEF, used for building and updating simulation software. It has 4 cores (Intel(R) Xeon(R) CPU X5450).
- *Flop* - Mainframe at Department of Applied Mathematics and Cybernetics, SINTEF, used for simulations. It has three units:
  - *Flop 1* - 32 cores (Intel(R) Xeon(R) CPU E5-2650 0).
  - *Flop 2* - 32 cores (Intel(R) Xeon(R) CPU E5-2650 0).
  - *Flop 3* - 48 cores (Intel(R) Xeon(R) CPU E5-2670 v3).

6.2 Simulation facilities

The IFEM-software is non-graphical and is run in a computer terminal, preferably UNIX (Linux) terminals which always have Shellscript as a built-in language for communicating directly with the computer. This allows us to run many simulations at once and automatize much of the post-processing. Although IFEM is run in the terminal and has no GUI, the output can be interpreted by graphical software displaying extensively how the numerical results behave.

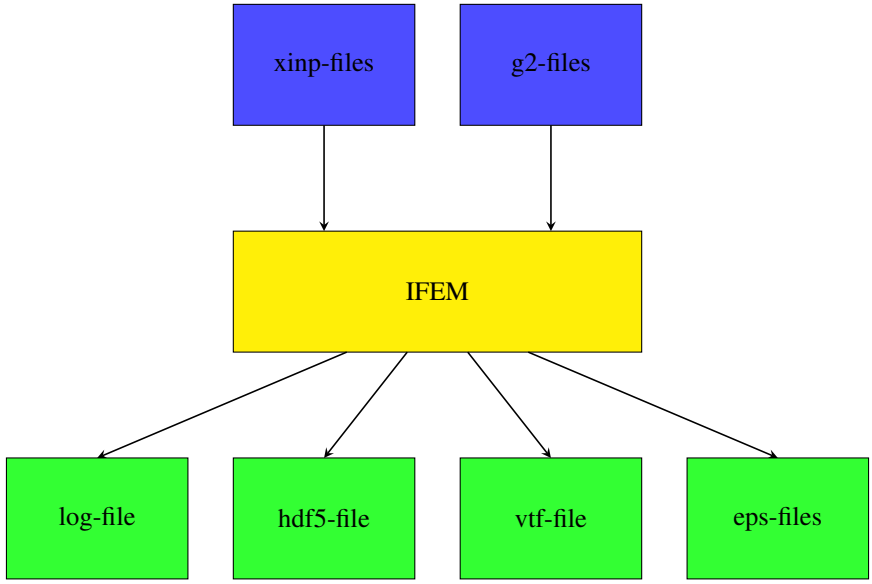


Figure 34. Schematic representation of the input/output-processing in IFEM.

There are six types of files used in the simulations:

- *xinp-files*: The primary IFEM input. This XML-file contains all information about the PDE, boundary conditions, discretization type, solution method and post-processing. Multiple files can also be passed if necessary.
- *g2-files*: The secondary IFEM input. This file is created from Python-scripts using SpliPy and lrsplines. It contains all necessary information about the physical domain that is required for a suitable discretization.
- *log-file*: This file contains all information about the numerical solution, like description of the discretized system, values of the PDE-specific norms, and discretization errors.
- *hdf5-file*: This file contains the numerical solution data and is accessed through Python in order to perform customary post-processing.
- *vtf-file*: This is the input file of GLview. It displays the behaviour of the numerical solution, its derivatives, and the different norms.
- *eps-files*: These files describe how the mesh evolves.

Bibliography

[1] P. Alfeld and L. L. Schumaker. “Bounds on the dimensions of trivariate spline spaces”. In: *Adv. Comput. Math.* 29 (2008), pp. 315–335.

[2] C. Anitescu, M. N. Hossain, and T. Rabczuk. “Recovery-based error estimation and adaptivity using high-order splines over hierarchical T-meshes”. In: *Computer Methods in Applied Mechanics and Engineering* 328 (2018), pp. 638–662.

[3] M. D. et at. “Robust and optimal multi-iterative techniques for IgA collocation linear systems”. In: *Computer Methods in Applied Mechanics and Engineering* 284 (2015), pp. 1120–1146.

[4] M. D. et at. “Robust and optimal multi-iterative techniques for IgA Galerkin linear systems”. In: *Computer Methods in Applied Mechanics and Engineering* 284 (2015), pp. 230–264.

[5] I. Babuška and W. C. Rheinboldt. “A posteriori error estimates for the finite element method”. In: *International Journal for Numerical Methods in Engineering* 12 (1978), pp. 1597–1615.

[6] I. Babuška and W. C. Rheinboldt. “Error estimates for adaptive finite element computations”. In: *SIAM Journal of Numerical Analysis* 15.4 (1978), pp. 736–754.

[7] I. Babuška, T. Strouboulis, and C. S. Upadhyay. “A model study of the quality of a posteriori error estimators for finite element solutions of linear elliptic problems, with particular reference to the behavior near the boundary”. In: *International Journal for Numerical Methods in Engineering* 40 (1997), pp. 2521–2577.

[8] I. Babuška, T. Strouboulis, and C. S. Upadhyay. “A model study of the quality of a posteriori error estimators for linear elliptic problems. Error estimation in the interior of patchwise uniform grids of triangles”. In: *Computer Methods in Applied Mechanics and Engineering* 114 (1994), pp. 307–378.

[9] I. Babuška, C. Upadhyay, S. Gangaraj, and K. Copps. “Validation of a posteriori error estimators by numerical approach”. In: *International Journal for Numerical Methods in Engineering* 37.7 (1994), pp. 1073–1123.

[10] I. Babuška, J. E. Flaherty, W. D. Henshaw, J. E. Hopcroft, J. Oliger, and T. Tezduyar. *Modeling, Mesh Generation, and Adaptive Numerical Methods for Partial Differential Equations*. New York: Springer-Verlag, 1995.

[11] A. Bartezzaghi, L. Dedè, and A. Quarteroni. “Isogeometric Analysis for high order Partial Differential Equations on surfaces”. In: *Computer Methods in Applied Mechanics and Engineering* 295 (2015), pp. 446–469.

[12] Y. Bazilevs, L. B. da Veiga, J. A. Cottrell, T. J. R. Hughes, and G. Sangalli. “Isogeometric analysis: Approximation, stability and error estimates for h-refined meshes”. In: *Mathematical Models and Methods in Applied Sciences* 16 (2006), pp. 1031–1090.

[13] Y. Bazilevs, V. M. Calo, T. J. R. Hughes, and Y. Zhang. “Isogeometric fluid-structure interaction: theory, algorithms, and computations”. In: *Comput. Mech.* 43 (2008), pp. 3–37.

[14] C. Beccari, G. Casciola, and S. Morigi. “On multi-degree splines”. In: *Computer Aided Geometric Design* 58 (2017), pp. 8–23.

[15] A. Blidia, B. Mourrain, and N. Villamizar. “G^1-smooth splines on quad meshes with 4-split macro-patch elements”. In: *Computer Aided Geometric Design* 52-53 (2017), pp. 106–125.

[16] D. Boffi, F. Brezzi, and M. Fortin. *Mixed Finite Element Methods and Applications*. Berlin: Springer-Verlag, 2013.

[17] B. D. Bojanov, H. A. Hakopian, and A. A. Sahakian. *Spline Functions and Multivariate Interpolations*. Dordrecht, Netherlands: Springer-Verlag, 1993.

[18] C. de Boor. *A Practical Guide to Splines*. New York, USA: Springer-Verlag, 2013.

[19] A. Bressan. “Some properties of LR-splines”. In: *Computer Aided Geometric Design* 30 (2013), pp. 778–794.

[20] F. Buchegger, B. Jüttler, and A. Mantzaflaris. “Adaptively refined multi-patch B-splines with enhanced smoothness”. In: *Applied Mathematics and Computation* 272 (2016), pp. 159–172.

[21] A. Buffa, C. de Falco, and G. Sangalli. “IsoGeometric Analysis: Stable elements for the 2D Stokes equation”. In: *International Journal for Numerical Methods in Fluids* 65 (2011), pp. 1407–1422.

[22] A. Buffa and E. M. Garau. “Refinable spaces and local approximation estimates for hierarchical splines”. In: *IMA Journal of Numerical Analysis* 37 (2017), pp. 1125–1149.

[23] A. Buffa, G. Sangalli, and R. Vázquez. “Isogeometric analysis in electromagnetics: B-splines approximation”. In: *Computer Methods in Applied Mechanics and Engineering* 199 (2010), pp. 1143–1152.

[24] A. Buffa, G. Sangalli, and R. Vázquez. “Isogeometric methods for computational electromagnetics: B-spline and T-spline discretizations”. In: *Journal of Computational Physics* 257 (2014), pp. 1291–1320.

[25] M. Charina, M. Donatelli, L. Romani, and V. Turati. “Multigrid methods: Grid transfer operators and subdivision schemes”. In: *Linear Algebra and its Applications* 520 (2017), pp. 151–190.

[26] C. de Concini, C. Procesi, and M. Vergne. “Box splines and the equivariant index theorem”. In: *J. Inst. Math. Jussieu* 12.3 (2012), pp. 503–544.

[27] J. A. Cottrell, T. J. R. Hughes, and A. Reali. “Studies of refinement and continuity in isogeometric structural analysis”. In: *Computer Methods in Applied Mechanics and Engineering* 196 (2007), pp. 4160–4183.

[28] J. A. Cottrell, A. Reali, Y. Bazilevs, and T. J. R. Hughes. “Isogeometric analysis of structural vibrations”. In: *Computer Methods in Applied Mechanics and Engineering* 195 (2006), pp. 5257–5296.

[29] J. A. Cottrell, T. J. R. Hughes, and Y. Bazilevs. *Isogeometric Analysis: Toward Integration of CAD and FEA*. UK: John Wiley & Sons, Ltd, 2009.

[30] D. D’Angella, S. Kollmannsberger, E. Rank, and A. Reali. “Multilevel Bézier extraction for hierarchical local refinement of Isogeometric Analysis”. In: *Computer Methods in Applied Mechanics and Engineering* 328 (2018), pp. 147–174.

[31] L. Dedè, C. Jäggli, and A. Quarteroni. “Isogeometric numerical dispersion analysis for two-dimensional elastic wave propagation”. In: *Computer Methods in Applied Mechanics and Engineering* 284 (2015), pp. 320–348.

[32] L. Dedè and A. Quarteroni. “Isogeometric Analysis for second order Partial Differential Equations on surfaces”. In: *Computer Methods in Applied Mechanics and Engineering* 284 (2015), pp. 807–834.

[33] M. DiPasquale. “Associated primes of spline complexes”. In: *Journal of Symbolic Computation* 76 (2016), pp. 158–199.

[34] M. DiPasquale, F. Sottile, and L. Sun. “Semialgebraic splines”. In: *Computer Aided Geometric Design* 55 (2017), pp. 29–47.

[35] T. Dokken, T. Lyche, and K. Pettersen. “Polynomial splines over locally refined box-partitions”. In: *Computer Aided Geometric Design* 30 (2013), pp. 331–356.

[36] B. Dortdivanlioglu, A. Javili, and C. Linder. “Computational aspects of morphological instabilities using isogeometric analysis”. In: *Computer Methods in Applied Mechanics and Engineering* 316 (2017), pp. 261–279.

[37] J. A. Evans and T. J. R. Hughes. “Isogeometric divergence-conforming B-splines for the Darcy-Stokes-Brinkman equations”. In: *Mathematical Models and Methods in Applied Sciences* 23 (2013), pp. 671–741.

[38] J. A. Evans and T. J. R. Hughes. “Isogeometric divergence-conforming B-splines for the steady Navier-Stokes equations”. In: *Mathematical Models and Methods in Applied Sciences* 23 (2013), pp. 1421–1478.

[39] J. A. Evans and T. J. R. Hughes. “Isogeometric divergence-conforming B-splines for the unsteady Navier-Stokes equations”. In: *Journal of Computational Physics* 241 (2013), pp. 141–167.

[40] G. Farin. *NURBS from Projective Geometry to Practical Use*. Massachusetts: A K Peters Ltd, 1999.

[41] M. Feischl, G. Gantner, and D. Praetorius. “Reliable and efficient a posteriori error estimation for adaptive IGA boundary element methods for weakly-singular integral equations”. In: *Computer Methods in Applied Mechanics and Engineering* 290 (2015), pp. 362–386.

[42] M. Feischl, G. Gantner, A. Haberl, and D. Praetorius. “Optimal convergence for adaptive IGA boundary element methods for weakly-singular integral equations”. In: *Numerische Mathematik* 136 (2017), pp. 147–182.

[43] J. Fish and T. Belytschko. *A First Course in Finite Elements*. Berlin: John Wiley & Sons, Ltd, 2007.

[44] D. Forsey and R. Bartels. "Hierarchical B-Spline Refinement". In: *Computer Graphics* 22.4 (1988), pp. 205–212.

[45] P. J. Frey and P.-L. George. *Mesh Generation*. London: John Wiley & Sons, Inc., 2008.

[46] K. P. S. Gahalaut, J. K. Kraus, and S. K. Tomar. "Multigrid methods for isogeometric discretization". In: *Computer Methods in Applied Mechanics and Engineering* 253 (2013), pp. 413–425.

[47] K. P. S. Gahalaut, S. K. Tomar, and J. K. Kraus. "Algebraic multilevel preconditioning in isogeometric analysis. Construction and numerical studies". In: *Computer Methods in Applied Mechanics and Engineering* 266 (2013), pp. 40–56.

[48] E. M. Garau and R. Vázquez. "Algorithms for the implementation of adaptive isogeometric methods using hierarchical B-splines.pdf". In: *Applied Numerical Mathematics* 123 (2018), pp. 58–87.

[49] D. Garcia, D. Pardo, L. Dalcin, M. Paszýnski, N. Collier, and V. M. Calo. "The value of continuity: Refined isogeometric analysis and fast direct solvers". In: *Computer Methods in Applied Mechanics and Engineering* 316 (2017), pp. 586–605.

[50] C. Garoni, C. Manni, F. Pelosi, S. Serra-Capizzano, and H. Speleers. "On the spectrum of stiffness matrices arising from isogeometric analysis". In: *Numerische Mathematik* 127 (2014), pp. 751–799.

[51] C. Giannelli, B. Jüttler, and H. Speleers. "Strongly stable bases for adaptively refined multilevel spline spaces". In: *Adv. Comput. Math.* 40 (2014), pp. 459–490.

[52] C. Giannelli, B. Jüttler, and H. Speleers. "THB-splines: The truncated basis for hierarchical splines". In: *Computer Aided Geometric Design* 29 (2012), pp. 485–498.

[53] C. Giannelli, B. Jüttler, S. K. Kleiss, A. Mantzaflaris, B. Simeon, and J. Špeh. "THB-splines: An effective mathematical technology for adaptive refinement in geometric design and isogeometric analysis". In: *Computer Methods in Applied Mechanics and Engineering* 299 (2016), pp. 337–365.

[54] D. Groisser and J. Peters. "Matched G^k-constructions always yield C^k-continuous isogeometric elements". In: *Computer Aided Geometric Design* 34 (2015), pp. 67–72.

[55] C. Heinrich, B. Simeon, and S. Boschert. “A finite volume method on NURBS geometries and its application in isogeometric fluid-structure interaction”. In: *Mathematics and Computers in Simulation* 82 (2012), pp. 1645–1666.

[56] P. Hennig, M. Kästner, P. Morgenstern, and D. Peterseim. “Adaptive mesh refinement strategies in isogeometric analysis - A computational comparison”. In: *Computer Methods in Applied Mechanics and Engineering* 316 (2017), pp. 424–448.

[57] R. R. Hiemstra, F. Calabrò, D. Schillinger, and T. J. Hughes. “Optimal and reduced quadrature rules for tensor product and hierarchically refined splines in isogeometric analysis”. In: *Computer Methods in Applied Mechanics and Engineering* 316 (2017), pp. 966–1004.

[58] C. Hofreither, S. Takacs, and W. Zulehner. “A robust multigrid method for Isogeometric Analysis in two dimensions using boundary correction”. In: *Computer Methods in Applied Mechanics and Engineering* 316 (2017), pp. 22–42.

[59] C. Hofreither and S. Takacs. “Robust Multigrid for Isogeometric Analysis Based on Stable Splittings of Spline Spaces”. In: *SIAM Journal of Numerical Analysis* 55.4 (2017), pp. 2004–2024.

[60] C. Hofreither, B. Jüttler, G. Kiss, and W. Zulehner. “Multigrid methods for isogeometric analysis with THB-splines”. In: *Computer Methods in Applied Mechanics and Engineering* 308 (2016), pp. 96–112.

[61] Q.-X. Huang, S.-M. Hu, and R. R. Martin. “Efficient Degree Elevation and Knot Insertion for B-spline Curves using Derivatives”. In: *Computer-Aided Design and Applications* 1.1–4 (2004), pp. 719–725.

[62] T. J. R. Hughes, J. A. Cottrell, and Y. Bazilevs. “Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement”. In: *Computer Methods in Applied Mechanics and Engineering* 194 (2005), pp. 4135–4195.

[63] T. J. R. Hughes, J. A. Evans, and A. Reali. “Finite element and NURBS approximations of eigenvalue, boundary-value, and initial-value problems”. In: *Computer Methods in Applied Mechanics and Engineering* 272 (2014), pp. 290–320.

[64] T. J. R. Hughes, A. Reali, and G. Sangalli. "Duality and unified analysis of discrete approximations in structural dynamics and wave propagation: Comparison of p-method finite elements with k-method NURBS". In: *Computer Methods in Applied Mechanics and Engineering* 197 (2008), pp. 4104–4124.

[65] T. J. R. Hughes. *The finite element method*. USA: Prentice-Hall, 1987.

[66] W. Jiang and J. E. Dolbow. "Adaptive refinement of hierarchical B-spline finite elements with an efficient data transfer algorithm". In: *International Journal for Numerical Methods in Engineering* 102 (2015), pp. 233–256.

[67] K. A. Johannessen. "Optimal quadrature for univariate and tensor product splines". In: *Computer Methods in Applied Mechanics and Engineering* 316 (2017), pp. 84–99.

[68] K. A. Johannessen, M. Kumar, and T. Kvamsdal. "Divergence-conforming discretization for Stokes problem on locally refined meshes using LR B-splines". In: *Computer Methods in Applied Mechanics and Engineering* 293 (2015), pp. 38–70.

[69] K. A. Johannessen, T. Kvamsdal, and T. Dokken. "Isogeometric analysis using LR B-splines". In: *Computer Methods in Applied Mechanics and Engineering* 269 (2014), pp. 471–514.

[70] K. A. Johannessen, F. Remonato, and T. Kvamsdal. "On the similarities and differences between Classical Hierarchical, Truncated Hierarchical and LR B-splines". In: *Computer Methods in Applied Mechanics and Engineering* 291 (2015), pp. 64–101.

[71] T. Kanduč, C. Giannelli, F. Pelosi, and H. Speleers. "Adaptive isogeometric analysis with hierarchical box splines". In: *Computer Methods in Applied Mechanics and Engineering* 316 (2017), pp. 817–838.

[72] M. Kapl, V. Vitrih, B. Jüttler, and K. Birner. "Isogeometric analysis with geometrically continuous functions on two-patch geometries". In: *Computers and Mathematics with Applications* 70 (2015), pp. 1518–1538.

[73] S. K. Kleiss and S. K. Tomar. "Guaranteed and sharp a posteriori error estimates in isogeometric analysis". In: *Computers and Mathematics with Applications* 70 (2015), pp. 167–190.

[74] R. Kolman, J. Plešek, and M. Okrouhlík. “Complex wavenumber Fourier analysis of the B-spline based finite element method”. In: *Wave Motion* 15 (2014), pp. 348–359.

[75] M. Kumar, T. Kvamsdal, and K. A. Johannessen. “Simple a posteriori error estimators in adaptive isogeometric analysis”. In: *Computers and Mathematics with Applications* 70 (2015), pp. 1555–1582.

[76] M. Kumar, T. Kvamsdal, and K. A. Johannessen. “Superconvergent patch recovery and a posteriori error estimation technique in adaptive isogeometric analysis”. In: *Computer Methods in Applied Mechanics and Engineering* 293 (2017), pp. 1086–1156.

[77] G. Kuru, C. Verhoosel, K. van der Zee, and E. van Brummelen. “Goal-adaptive Isogeometric Analysis with hierarchical splines”. In: *Computer Methods in Applied Mechanics and Engineering* 270 (2014), pp. 270–292.

[78] U. Langer, S. E. Moore, and M. Neumüller. “Space-time isogeometric analysis of parabolic evolution problems”. In: *Computer Methods in Applied Mechanics and Engineering* 306 (2016), pp. 342–363.

[79] J. M. Lee. *Introduction to Smooth Manifolds*. New York: Springer-Verlag, 2013.

[80] X. Li and M. A. Scott. “Analysis-suitable T-splines: Characterization, refineability, and approximation”. In: *Mathematical Models and Methods in Applied Sciences* 24 (2014), pp. 1141–1164.

[81] X. Li, J. Deng, and F. Chen. “Dimensions of Spline Spaces over 3D Hierarchical T-Meshes”. In: *Journal of Information and Computational Science* 3.3 (2006), pp. 487–501.

[82] B. Marussig, J. Zechner, G. Beer, and T.-P. Fries. “Fast isogeometric boundary element method based on independent field approximation”. In: *Computer Methods in Applied Mechanics and Engineering* 284 (2015), pp. 458–488.

[83] C. A. Micchelli, Y. Xu, and H. Zhang. “On translation invariant operators which preserve the B-spline recurrence”. In: *Adv. Comput. Math.* 28 (2008), pp. 157–169.

[84] G. Micula and S. Micula. *Handbook of Splines*. Netherlands: Springer-Verlag, 1999.

[85] D. Mokriš, B. Jüttler, and C. Giannelli. “On the completeness of hierarchical tensor-product B-splines”. In: *Journal of Computational and Applied Mathematics* 271 (2014), pp. 53–70.

[86] M. R. Moosavi and A. Khelil. “Isogeometric meshless finite volume method in nonlinear elasticity”. In: *Acta Mechanica* 226 (2015), pp. 123–135.

[87] B. Mourrain, R. Vidunas, and N. Villamizar. “On the dimension of spline spaces on planar T-meshes”. In: *Computer Aided Geometric Design* 45 (2016), pp. 108–133.

[88] V. P. Nguyen, C. Anitescu, S. P. Bordas, and T. Rabczuk. “Isogeometric analysis: An overview and computer implementation aspects”. In: *Mathematics and Computers in Simulation* 117 (2015), pp. 89–116.

[89] J. Niiranen, S. Khakalo, V. Balobanov, and A. Niemi. “Variational formulation and isogeometric analysis for fourth-order boundary value problems of gradient-elastic bar and plane strain/stress problems”. In: *Computer Methods in Applied Mechanics and Engineering* 308 (2016), pp. 182–211.

[90] M. J. Peake, J. Trevelyan, and G. Coates. “Extended isogeometric boundary element method (XIBEM) for three-dimensional medium-wave acoustic scattering problems”. In: *Computer Methods in Applied Mechanics and Engineering* 284 (2015), pp. 762–780.

[91] L. Piegl and W. Tiller. *The NURBS Book*. Berlin: Springer-Verlag, 1997.

[92] E. Pilgerstorfer and B. Jüttler. “Bounding the influence of domain parameterization and knot spacing on numerical stability in Isogeometric Analysis”. In: *Computer Methods in Applied Mechanics and Engineering* 268 (2014), pp. 589–613.

[93] C. Procesi. “Splines and index theorem”. In: *Bull. Math. Sci.* 2 (2012), pp. 57–123.

[94] V. Puzyrev, Q. Deng, and V. Calo. “Dispersion-optimized quadrature rules for isogeometric analysis: Modified inner products, their dispersion properties, and optimally blended schemes”. In: *Computer Methods in Applied Mechanics and Engineering* 320 (2017), pp. 421–443.

[95] A. Ratnani and E. Sonnendrücker. “An Arbitrary High-Order Spline Finite Element Solver for the Time Domain Maxwell Equations”. In: *J Sci. Comput.* 51 (2012), pp. 87–106.

[96] F. Roman, C. Manni, and H. Speleers. “Spectral analysis of matrices in Galerkin methods based on generalized B-splines with high smoothness”. In: *Numerische Mathematik* 135 (2017), pp. 169–216.

[97] D. Rypl and B. Patzák. “From the finite element analysis to the isogeometric analysis in an object oriented computing environment”. In: *Advances in Engineering Software* 44 (2012), pp. 116–125.

[98] G. Sangalli, T. Takacs, and R. Vázquez. “Unstructured spline spaces for isogeometric analysis based on spline manifolds”. In: *Computer Aided Geometric Design* 47 (2016), pp. 61–82.

[99] H. Schenck. “Algebraic methods in approximation theory”. In: *Computer Aided Geometric Design* 45 (2016), pp. 14–31.

[100] D. Schillinger and E. Rank. “An isogeometric design-through-analysis methodology based on adaptive hierarchical refinement of NURBS, immersed boundary methods, and T-spline CAD surfaces”. In: *Computer Methods in Applied Mechanics and Engineering* 200 (2011), pp. 3358–3380.

[101] D. Schillinger, L. Dedè, M. A. Scott, J. A. Evans, M. J. Borden, E. Rank, and T. J. Hughes. “An isogeometric design-through-analysis methodology based on adaptive hierarchical refinement of NURBS, immersed boundary methods, and T-spline CAD surfaces”. In: *Computer Methods in Applied Mechanics and Engineering* 249-252 (2012), pp. 116–150.

[102] D. Schillinger, J. A. Evans, A. Reali, M. A. Scott, and T. J. Hughes. “Isogeometric collocation: Cost comparison with Galerkin methods and extension to adaptive hierarchical NURBS discretizations”. In: *Computer Methods in Applied Mechanics and Engineering* 267 (2013), pp. 170–232.

[103] L. Schumaker. *Spline Functions, Computational Methods*. Philadelphia: Society for Industrial and Applied Mathematics, 2015.

[104] C. Schwab. *p- and hp- Finite Element Methods*. UK: Oxford University Press, 1998.

[105] M. A. Scott, D. C. Thomas, and E. J. Evans. “Isogeometric spline forests”. In: *Computer Methods in Applied Mechanics and Engineering* 269 (2014), pp. 222–264.

[106] M. A. Scott, X. Li, T. W. Sederberg, and T. J. R. Hughes. “Local refinement of analysis-suitable T-splines”. In: *Computer Methods in Applied Mechanics and Engineering* 213-216 (2012), pp. 206–222.

[107] T. W. Sederberg, J. Zheng, A. Bakenov, and A. Nasri. “T-splines and T-NURCCs”. In: *ACM Trans. Graph.* 22 (2003), pp. 477–484.

[108] R. N. Simpson and Z. Liu. “Acceleration of isogeometric boundary element analysis through a black-box fast multipole method”. In: *Engineering Analysis with Boundary Elements* 66 (2016), pp. 168–182.

[109] L. Tian, F. Chen, and Q. Du. “Adaptive finite element methods for elliptic equations over hierarchical T-meshes”. In: *Journal of Computational and Applied Mathematics* 236 (2011), pp. 878–891.

[110] D. Toshniwala, H. Speleers, R. R. Hiemstraa, and T. J. Hughes. “Multi-degree smooth polar splines: A framework for geometric modeling and isogeometric analysis”. In: *Computer Aided Geometric Design* 316 (2017), pp. 1005–1061.

[111] L. B. da Veiga, A. Buffa, G. Sangalli, and R. Vázquez. “Mathematical analysis of variational isogeometric methods”. In: *Acta Numerica* 23 (2014), pp. 157–287.

[112] A.-V. Vuong, C. Giannelli, B. Jüttler, and B. Simeon. “A hierarchical approach to adaptive local refinement in isogeometric analysis”. In: *Computer Methods in Applied Mechanics and Engineering* 200 (2011), pp. 3554–3567.

[113] D. Wang, Q. Liang, and J. Wu. “A quadrature-based superconvergent isogeometric frequency analysis with macro-integration cells and quadratic splines”. In: *Computer Methods in Applied Mechanics and Engineering* 320 (2017), pp. 272–744.

[114] D. Wang, W. Liu, and H. Zhang. “Superconvergent isogeometric free vibration analysis of Euler-Bernoulli beams and Kirchhoff plates with new higher order mass matrices”. In: *Computer Methods in Applied Mechanics and Engineering* 286 (2015), pp. 230–267.

[115] M. Woźniak, K. Kuźnik, M. Paszyński, V. Calo, and D. Pardo. “Computational cost estimates for parallel shared memory isogeometric multi-frontal solvers”. In: *Computers and Mathematics with Applications* 67 (2014), pp. 1864–1883.

[116] Z. jun Wu, Z. dong Huang, Q. hua Liu, and B. quan Zuo. “A local solution approach for adaptive hierarchical refinement in isogeometric analysis”. In: *Computer Methods in Applied Mechanics and Engineering* 283 (2015), pp. 1467–1492.

[117] L. Xin, C. F. Lai, K. H. Mei, and D. J. Song. “A survey on the local refinable splines”. In: *Science China* 59 (2016), pp. 617–644.

[118] Y. Zhang, Y. Bazilevs, S. Goswami, C. L. Bajaj, and T. J. Hughes. "Patient-specific vascular NURBS modeling for isogeometric analysis of blood flow". In: *Computer Methods in Applied Mechanics and Engineering* 196 (2007), pp. 2943–2959.

[119] S. Zhu, L. Dedè, and A. Quarteroni. "Isogeometric analysis and proper orthogonal decomposition for parabolic problems". In: *Numerische Mathematik* 135 (2016), pp. 333–370.

[120] Y. Zhu and F. Chen. "Modified Bases of PHT-Splines". In: *Commun. Math. Stat.* 5 (2017), pp. 381–397.

[121] O. Zienkiewicz and J. Z. Zhu. "The superconvergent patch recovery and a posteriori error estimates. I. The recovery technique". In: *International Journal for Numerical Methods in Engineering* 33 (1992), pp. 1331–1364.

[122] O. Zienkiewicz and J. Z. Zhu. "The superconvergent patch recovery and a posteriori error estimates. II. Error estimates and adaptivity". In: *International Journal for Numerical Methods in Engineering* 33 (1992), pp. 1365–1382.

Report: Error Estimation in Isogeometric Analysis

Abdullah Abdulhaque

Technical report accepted at NTNU

Error Estimation in Isogeometric Analysis

Abdullah Abdulhaque

Department of Mathematical Sciences
Norwegian University of Science and Technology, Trondheim, Norway
e-mail: Abdullah.Abdulhaque@ntnu.no

Abstract

This paper is a detailed summary of the most common techniques used for local refinement and error estimation in the Finite Element Method. We will study whether it is possible to adapt these classical techniques to Isogeometric Analysis and examine their qualitative properties to determine which error estimator is best for adaptive refinement. There is also some discussion on the new methods only available when using B-splines for approximating the unknown solution.

1 Adaptive finite element modelling

1.1 Importance of local refinement

Local refinement is an important topic in the *Finite Element Method* (FEM). In some cases, the numerical approximation of a partial differential equation (PDE) is deteriorated by local singularities or very sharp gradients. They can be generated by several factors like shock fronts, interior layers, boundary layers, rarefaction waves, discontinuity propagation and re-entrant corners on the domain. In all the corresponding regions where these deteriorating factors are located, the solution of the PDE is lacking sufficient regularity. Therefore, it can be convenient to refine the elements of these regions locally to increase the accuracy faster.

The main advantage of local refinement is that we only refine those elements where the estimated error of the solution exceeds a predefined tolerance. By doing so, the discrete system of equations will not grow too large, and the computational effort required for solving it is limited. This is more effective than standard tensor-refinement, where all the adjacent elements must also be refined to ensure that the mesh is uniform. After refining the solution a few times and reducing the approximation error, the final mesh resembles the physical structure of the unknown solution. This is where the term "*adaptive refinement*" occurs, and we denote our numerical approach as *Adaptive Finite Element Method* (AFEM).

1.2 Historical background

In finite element modelling, a differential equation is transformed into an integral equation. Since the integrand is measurable, we can split the integral into a sum of integrals over disjoint subdomains whose union constitutes the original domain. Thus, we can analyse the problem globally and locally. In any field of computational mechanics, this characteristic construction allows us to discretize arbitrary geometries by using unstructured meshes.

In the process of solving a differential equation, we must transform the continuum model of a physical problem into a suitable discrete model which can be handled properly by a computer. Because of the discretization in space and time, parts of the original model's information are lost, so there will always be some error in the approximation. The new quest becomes how to measure, control, and minimize the error as best as possible. If we succeed, then the quality of the numerical solution is optimized.

A priori error estimates have been known for a long time, and they work well for most differential equations. The method of a posteriori estimation, anyway, is relatively new in a historical perspective and became popular thanks to the rapid emerge of powerful computers allowing better analysis and quality control of the numerical results. Babuška and Rheinboldt were among the first ones to study both a posteriori estimation and adaptive finite element modelling in the late 1970s [9, 12, 11, 10, 8]. The research resulted in several different error estimators. Their pioneering work gave an impetus to intensive research on these topics.

Many important facilities have been developed in connection with the theory a posteriori error estimation. Ladevèze and Leguillon [68] discovered the *equilibrated boundary data* concept, where some small complementary problems are solved elementwise and then combined together. Demkowicz and his collaborators [43, 42, 44] introduced the *element residual method*, where we use the residual error of the approximation as an error indicator, mostly applied in connection with p- and hp-FEM. Both Ainsworth and Oden have shown that this method works well for saddle-point problems, variational inequalities and elliptic BVPs [3, 73].

Despite these various efforts, most research work in the 1980s was ad hoc and designed for specific individual PDEs. The complete theoretical framework for a posteriori error estimation reached sufficient maturity in the 1990s. At that time, the focus was changed on how to create universal error estimators for large general classes of PDEs (elliptic, parabolic, hyperbolic), and this new point of view accelerated the research further.

The superconvergent patch recovery, invented by Zienkiewicz and Zhu [93, 94], is one of the greatest achievements in the context of a posteriori error estimation. We smooth the gradient in a certain region of the domain

and compare it with other gradients of the original solution. According to Babuška et al. [14, 21, 13], this estimator is straightforward to implement, and it is most effective and robust for smooth problems approximated by linear and quadratic shape functions.

In 2005, Hughes and his collaborators introduced *Isogeometric Analysis* (IGA), a new finite element technique [62]. It has superior approximation properties when compared with classical FEM and provides very accurate solutions with minimal computational effort. This is because we use splines (B-splines, NURBS, etc.) as basis functions, and that creates enhanced interoperability between Finite Element Analysis (FEA) and Computer Assisted Design (CAD) [28, 37, 86, 85].

Today, it seems like the long-standing quest for creating shape functions with optimal approximation properties has been fulfilled. Since B-splines and NURBS just allow tensor-refinement, the next goal was incorporating local refinement from classical AFEM with IGA. This would lead to the development of new basis functions like locally refined (LR) B-splines [45, 63] and analysis suitable (AS) T-splines [81, 69]. Despite several efforts, the complete theory of adaptive IGA is still relatively new. A natural question arising is whether the classical theory of a posteriori error estimation is compatible with the isogeometric framework, where splines provide high continuity, lower degrees of freedom, and curved boundaries for the essential reduction of global approximation error. Recently, various articles about adaptive isogeometric refinement have been published, with particular focus on the chosen basis functions [41, 52, 54, 55, 60, 64, 79, 91].

1.3 Aim and outline of the paper

We start with a formal description of technical concepts related to finite element analysis and regular mesh generation. Much of the theory here will be applied in the later sections where error estimators are derived. Some special properties of reference elements and subdivision of meshes will also be introduced.

Then, we move on to the general theory of a priori estimates. This part is relevant because it forms some of the basis for the underlying theory behind a posteriori estimators. We start with some elementary estimates in one spatial dimension and prove convergence results for classical interpolants. After that, we will show how these convergence properties can be adapted to spline functions. The derivation of similar estimates in several dimensions will be as general as possible, assuming that the domain has a curved boundary, and the continuity is variable.

After that, we will give a brief introduction of the theory for a posteriori estimates. This includes the recent optimal control interpretation of adaptive refinement, and how to reduce pollution error in critical parts of the solution's domain, which has high importance.

The next step is deriving a posteriori estimators and determining which of them can be adapted to IGA. We will both derive and prove the robustness of explicit residual estimators and recovery estimators. A comprehensive discussion on the concepts like gradient recovery, superconvergence and computation of residuals will be given. The main goal here is to demonstrate that IGA is fully compatible with the old adaptive refinement framework.

Uniform refinement is a relevant topic to be discussed because IGA offers new ways of refining elements, and there is an elegant mathematical theory behind the comparison between these techniques. The Serendipity pairing between spaces that are obtained from k-refinement will also be presented, for this does not exist in classical finite element modelling.

Lastly, we will discuss the universal axioms behind adaptive mesh refinement and present different strategies for subdividing elements. This is the main part where we use the error estimators presented earlier.

2 Finite element nomenclature

This section considers the general theory of finite elements, conformal meshes, shape regularity, and mapping between the reference and physical domains. We will also derive some important inequalities related to the geometry mappings.

2.1 Properties of finite elements and partitions

We start first with the formal definition of finite elements and the partitions they generate on the domain of a differential equation.

Definition 1 (Finite element [39])**.** *We call the triple* $(\Omega, \mathcal{P}, \mathcal{N})$ *a finite element if it satisfies the following criterions:*

1. *The element domain* $\Omega \subseteq \mathbb{R}^d$ *is a compact set with nonempty interior, and the boundary* $\partial\Omega$ *is piecewise smooth.*
2. *The space of shape functions on* Ω*,* $\mathcal{P}$*, has finite dimension.*
3. *The set of degrees of freedom,* $\mathcal{N} = \{N_i\}_{i=1}^m$*, is a basis for* $\mathcal{P}'$*.*

It should be noted that the dimension of $\mathcal{P}$, and hence $\mathcal{P}'$, is usually referred to as the *number of degrees of freedom*. Since we have a finite-dimensional space, it is complete with respect to the norm [65]. For classical Lagrangian finite elements with the basis $\{\psi_i\}_{i=1}^m$ for $\mathcal{P}$, we have $\psi_i(N_j) = \delta_{ij}$, which is called a *nodal basis*. However, this basis is not a requirement for being a proper finite element. The important factor is that the degrees of freedom determines the chosen basis [36], i.e. it is a proper space for the dual space $\mathcal{P}'$. In IGA, we do not have a nodal basis. The degrees of freedom are the *control point values*, and it is possible to show that these values determine the spline function space.

If the PDE depends on one variable, the creation of a finite element partition $\mathcal{M}$ on an arbitrary open interval $I = (a, b)$ is straightforward. We just need to ensure that the partition is regular and includes the endpoints:

$$\mathcal{M} \,:\, a = x_0 < x_1 < x_2 < \cdots < x_{N-1} < x_N = b \tag{1}$$

Figure 1. A simple one-dimensional and non-uniform finite element partition.

The boundary conditions are just specified at the endpoints $x = \{a, b\}$. Generally, we do not need a uniform partition on the interval, and the mesh size h_K can vary. In such situations, it is appropriate to use the maximal width h in the process of deriving error estimates in one spatial dimension. The size of the N elements can be chosen freely, so we introduce

$$h_i = x_i - x_{i-1} \quad , \quad h = \max_{1 \leq i \leq N} h_i \tag{2}$$

In two or three spatial dimensions, the new requirements for a proper finite element partition become more complicated because several conditions must be satisfied. The following criterions are found in [39]:

Definition 2 (Finite element partition)**.** *Let Ω be a closed domain with Lipschitz boundary $\partial\Omega$. A proper finite element partition $\mathcal{M}$ on Ω satisfies*

1. ***Nonemptiness**: Every element $K \in \mathcal{M}$ is nonempty, and $\overline{K}$ is closed.*
2. ***Closure**: The finite element partition $\mathcal{M}$ ensures that $\overline{\Omega} = \bigcup_{K \in \mathcal{M}} \overline{K}$.*
3. ***Admissibility**: If $K_1, K_2 \in \mathcal{M}$, then either $K_1 \cap K_2 = \emptyset$, or $\partial K_1 \cap \partial K_2$ is a complete lower-dimensional face.*
4. ***Continuous boundary**: For all $K \in \mathcal{M}$, ∂K is Lipschitz continuous.*
5. ***Shape regularity**: If $K \in \mathcal{M}$, then its shape ratio κ_K is bounded away from zero and independent of K.*
6. ***Affine equivalence**: $\mathcal{M}$ will consist of triangles or tetrahedrons, or convex parallelograms or parallelepipeds in 2D and 3D, respectively.*

We denote $\mathcal{N}$ and $\mathcal{E}$, as the sets of vertices and edges, respectively. Furthermore, we define some special sets [88]:

$$\begin{aligned}
\omega_K &= \bigcup_{\mathcal{E}_K \cap \mathcal{E}_{K^*} \neq \emptyset} K^* & \widetilde{\omega}_K &= \bigcup_{\mathcal{N}_K \cap \mathcal{N}_{K^*} \neq \emptyset} K^* \\
\omega_\gamma &= \bigcup_{\gamma \in \mathcal{E}_{K^*}} K^* & \widetilde{\omega}_\gamma &= \bigcup_{\mathcal{N}_\gamma \cap \mathcal{N}_{K^*} \neq \emptyset} K^* \\
\omega_z &= \bigcup_{z \in \mathcal{N}_{K^*}} K^* & \sigma_z &= \bigcup_{z \in \mathcal{N}_\gamma} \gamma \\
\Sigma &= \bigcup_{K \in \mathcal{M}} \mathcal{E}_K
\end{aligned} \tag{3}$$

All these special sets above are visualized in Figure 3 and 4.

Figure 2. Illustration of a face of an element in two and three dimensions.

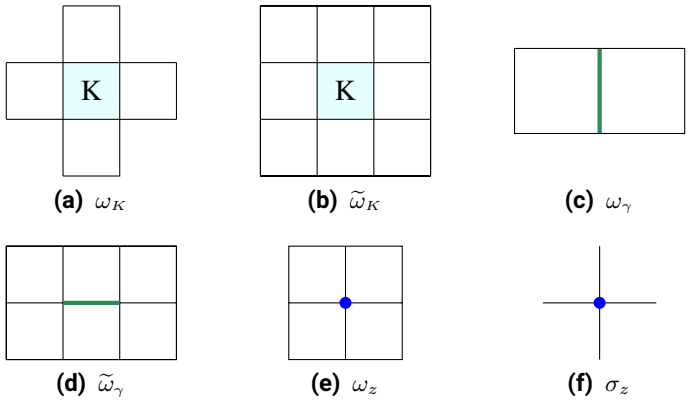


Figure 3. Illustration of the sets ω_K, $\widetilde{\omega}_K$, ω_γ, $\widetilde{\omega}_\gamma$, ω_z and σ_z for quadrilaterals.

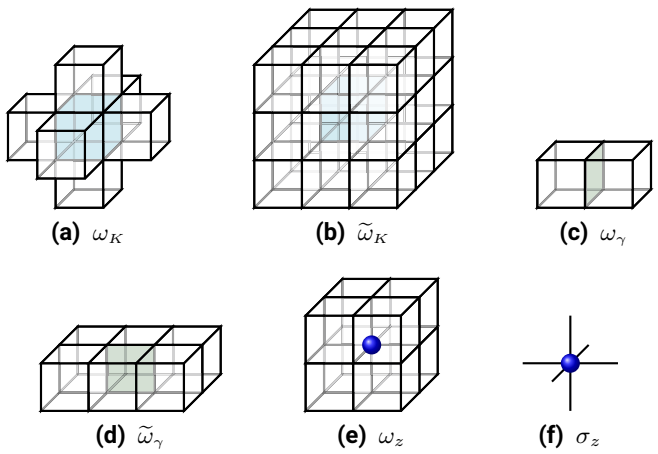


Figure 4. Illustration of the sets ω_K, $\widetilde{\omega}_K$, ω_γ, $\widetilde{\omega}_\gamma$, ω_z and σ_z for boxes.

Definition 3 (Domain [48]). *In several dimensions ($d \geq 2$), a domain Ω is an open, bounded and connected subset of $\mathbb{R}^d$. The boundary $\partial\Omega$ has some special properties. If $\alpha, \beta > 0$ are constants, $\{x^r = (x^{r*}, x_d^r) \in (\mathbb{R}^{d-1}, \mathbb{R}) : 1 \leq r \leq R\}$ is a finite set of local coordinate systems, and $\{\phi^r\}_{r=1}^R$ is a set of local maps which are Lipschitz continuous on their domain $\{x^{r*} \in \mathbb{R}^{d-1} : |x^{r*}| < \alpha\}$, then*

$$\partial\Omega = \bigcup_{r=1}^{R} \{(x^{r*}, x_d^r) : \phi^r(x^{r*}), |x^{r*}| < \alpha\}$$
$$\{(x^{r*}, x_d^r) : \phi^r(x^{r*}) < x_d^r < \phi^r(x^{r*}) + \beta, |x^{r*}| < \alpha\} \subset \Omega$$
$$\{(x^{r*}, x_d^r) : \phi^r(x^{r*}) - \beta < x_d^r < \phi^r(x^{r*}), |x^{r*}| < \alpha\} \subset \mathbb{R}^d \backslash \overline{\Omega}$$

where $|x^{r*}| \leq \alpha$ is a shorthand notation for $\{|x_i^{r*}| \leq \alpha : 1 \leq i \leq d-1\}$. If all the local maps belong to C^m, then Ω is said to be of class C^m.

In the assembly and post-processing, we need efficient computation of the local contributions for each element on the global mesh. Thus, we define the reference element as the standard triple $(\widehat{\Omega}, \widehat{\mathcal{P}}, \widehat{\mathcal{N}})$ and construct a geometric transformation $\mathcal{F} : \widehat{\Omega} \longmapsto \Omega$. Since $\mathcal{F}$ is a C^1-diffeomorphism, the physical nodes' numbering should be compatible with the numbering of the reference nodes. When splines are used as basis functions, it is sufficient with only one single reference mapping on each patch of the whole domain, not one for every individual element as in classical finite element analysis.

Definition 4 (Geometrically conformal mesh [48]). *Define the domain $\Omega \subset \mathbb{R}^d$, and $\mathcal{M}$ is a mesh on Ω. We call $\mathcal{M}$ geometrically conformal if it satisfies a special matching condition: If $K_m, K_n \in \mathcal{M}$ satisfies $K_m \cap K_n = F$, where F is a non-empty $(d-1)$-dimensional face, then there is a face $\widehat{F} \in \widehat{K}$ and a renumbering of the geometric nodes corresponding to K_m and K_n such that the following identities hold:*

$$F = \mathcal{F}_m(\widehat{F}) = \mathcal{F}_n(\widehat{F})$$
$$\mathcal{F}_{m|\widehat{F}} = \mathcal{F}_{n|\widehat{F}}$$

Most splines provide a tensor mesh on the domain. This is an elementary example of a geometrically conformal mesh, as seen in Figure 7. Furthermore, this mesh can be smoothly transformed into a new mesh with curved boundaries, making it possible to depict the domain exactly. This motivates the necessity of measuring the regularity of the elements on the mesh.

Definition 5 (Shape regularity)**.** *If h_K is the diameter of a triangle (the diameter of the smallest circle containing it), and ρ_K is the diameter of the largest inscribed circle inside it, we define the triangle's shape regularity κ_K as the finite ratio*

$$\kappa_K = \frac{h_K}{\rho_K} \tag{6}$$

For quadrilaterals, let $\{a_l\}_{l=1}^4$ be the vertices enumerated anti-clockwise, and T_l is a triangle with vertices $\{a_l, a_{l+1}, a_{l+2}\}$ (indexes are counted modulo 4) [2]. Then, we introduce the new diameters

$$h_K = \max_l h_{T_l} \qquad , \qquad \rho_K = \min_l \rho_l \tag{7}$$

Thus, for a convex quadrilateral, h_K is equal to the longest edge. In any case, we call the finite element partition $\mathcal{M}$ regular if there is a κ such that

$$\kappa_K \leq \kappa \qquad , \qquad K \in \mathcal{M} \tag{8}$$

For any finite element partition $\mathcal{M}$, we can define its *shape parameter* $C_\mathcal{M}$ as the *maximal shape ratio* of all elements $K \in \mathcal{M}$:

$$C_\mathcal{M} = \max_{K \in \mathcal{M}} \frac{h_K}{\rho_K} \tag{9}$$

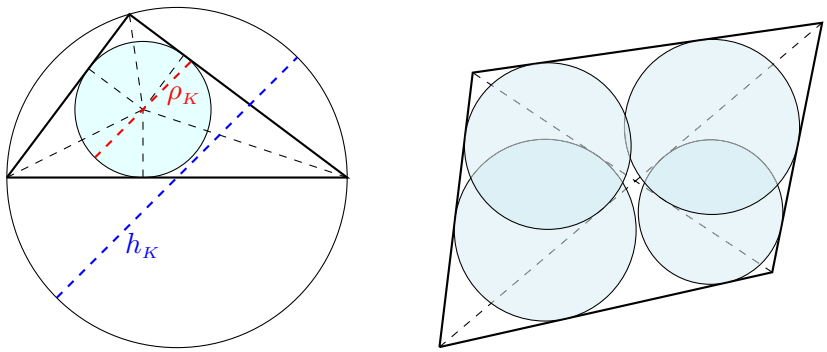


Figure 5. To the left, the circumscribed circle (radius h_K) and the inscribed circle (radius ρ_K) are shown for the triangle K. To the right, the four inscribed circles for a convex quadrilateral are displayed.

2.2 Properties of the reference element

We denote the *reference element* for convex quadrilaterals and boxes as

$$\widehat{K} = \{\widehat{\mathbf{x}} \in \mathbb{R}^d \,:\, \widehat{x}_i \in [0,1]\,,\, 1 \leq i \leq d\} \tag{10}$$

Each element K is the image of $\widehat{K}$ under $\mathcal{F}_K \,:\, \widehat{K} \longmapsto K$, which is a surjective and orientation-preserving diffeomorphism. In general, $\mathcal{F}_K$ might not be affine, especially for quadrilateral elements. This is more obvious in IGA, where the edges can be curved. The function A_K associated with $\mathcal{F}_K$ is therefore a vector function of the coordinates, so it must satisfy the following conditions [2]:

$$\|\mathcal{J}A_K\|_{L^\infty(\widehat{K})} \leq Ch_K \tag{11a}$$

$$\|\mathcal{J}A_K^{-1}\|_{L^\infty(\widehat{K})} \leq C\frac{\kappa_K}{\rho_K} \tag{11b}$$

$$C\rho_K^2 \leq \|\det(\mathcal{J}A_K)\|_{L^\infty(\widehat{K})} \leq Ch_K^2 \tag{11c}$$

where $\mathcal{J}A_K$ is the Jacobian matrix of A_K. From [88], the reference element $\widehat{K}$ allows us to define a *reference cube*. If $\alpha = (\alpha_1, \ldots, \alpha_d)$ is a multi-index such that $|\alpha| = \alpha_1 + \cdots + \alpha_d$, and $x^\alpha = x_1^{\alpha_1} \cdot \cdots \cdot x_d^{\alpha_d}$, then the cube is

$$R_p(\widehat{K}_d) = \text{span}\{x^\alpha : |\alpha|_\infty \leq p\} \tag{12a}$$

$$R_p(K) = \{\psi \circ \mathcal{F}_K^{-1} : \psi \in R_p(\widehat{K}_d)\} \tag{12b}$$

where p is the polynomial degree. If k is the continuity, we define the spaces

$$\mathcal{S}^{p,-1}(\mathcal{M}) = \{\psi : \Omega \longmapsto \mathbb{R}\,,\, \psi|_K \in R_p(K)\,,\, \forall K \in \mathcal{M}\} \tag{13a}$$

$$\mathcal{S}^{p,k}(\mathcal{M}) = \mathcal{S}^{p,-1}(\mathcal{M}) \cap C^k(\overline{\Omega}) \tag{13b}$$

$$\mathcal{S}_D^{p,k}(\mathcal{M}) = \mathcal{S}^{p,k}(\mathcal{M}) \cap H_D^1(\Omega) \tag{13c}$$

$$\mathcal{S}_0^{p,k}(\mathcal{M}) = \mathcal{S}^{p,k}(\mathcal{M}) \cap H_0^1(\Omega) \tag{13d}$$

We denote $\widetilde{K}$ as the subdomain consisting of K and the other elements sharing at least one common vertex with K [2]:

$$\widetilde{K} = \text{int}\left\{\bigcup K^* : K^* \in \mathcal{M}, \overline{K}^* \cap \overline{K} \neq \emptyset\right\} \tag{14}$$

This special patch will be used extensively in the derivation of central a priori and a posteriori estimates. For this patch, we denote the diameters as

$$h_{\widetilde{K}} = \max_{K^* \subseteq \widetilde{K}} h_{K^*} \qquad , \qquad \rho_{\widetilde{K}} = \min_{K^* \subseteq \widetilde{K}} \rho_{K^*} \tag{15}$$

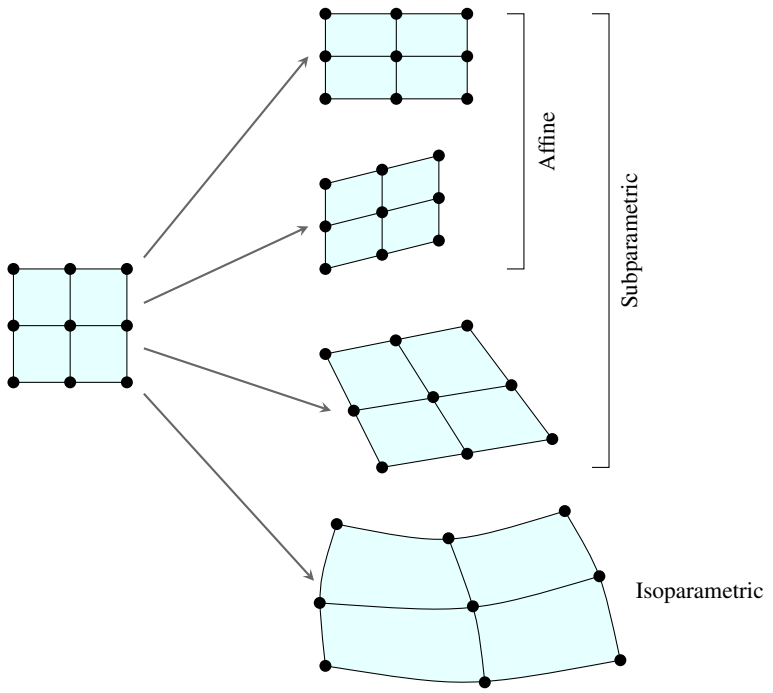
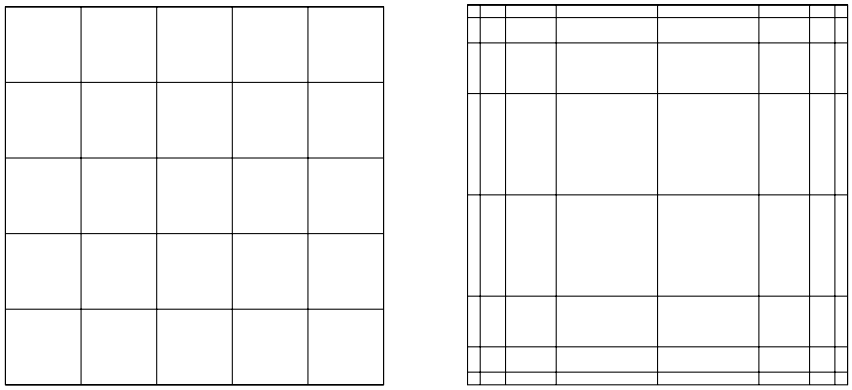


Figure 6. Visualization of different mappings from the reference quadrilateral element.

As we see from Figure 6, the standard affine mapping in classical FEM belongs to the category of *sub-parametric* mappings. This means that the interpolation order of the $\mathbf{x}$-coordinates is lower than that for the shape function. The *isoparametric* mapping will enable us to make the edges of the quadrilateral element curved, and this works well for NURBS. Adjusting the control points is relatively easy. In the case of affine mappings, the Jacobian of the transform is diagonal because a square is transformed to a parallelogram (the rectangle is actually a parallelogram where all the angles are $90°$). In the general sub-parametric case, it is an invertible and constant matrix. The isoparametric map describes the coordinates as functions of the other ones, and this enables us to generate curved edges on the quadrilaterals in a simpler way when compared with the classical finite element framework.

Definition 6 (Quasi-uniform mesh [2])**.** *If the mesh generated by the finite element partition $\mathcal{M}$ is quasi-uniform, then there is a constant C such that*

$$\left.\begin{aligned} h_{\widetilde{K}} &\leq C h_{K^*} \\ \kappa_{\widetilde{K}} &\leq C \kappa_{K^*} \end{aligned}\right\} \qquad \forall K^* \subset \widetilde{K} \tag{16}$$



(a) Uniform tensor mesh **(b)** Quasi-uniform mesh

Figure 7. Uniform and quasi-uniform meshes in two dimensions.

This definition is essential because it provides much simplification of several assumptions in the derivation of theorems about approximation, convergence and boundedness. Now, we will prove and generalize a proposition from [48]. First, we need some notational facilities.

For the reference element, denoted as $(\widehat{K}, \widehat{\mathcal{P}}, \widehat{\mathcal{N}})$, we define $\{\widehat{\sigma}_i\}_{i=1}^m$ and $\{\widehat{\psi}_i\}_{i=1}^m$ respectively as the local sets for degrees of freedom and shape functions, and $V(\widehat{K})$ is the closed domain of the finite element projection $\Pi_{\widehat{K}} : V(\widehat{K}) \longmapsto \widehat{\mathcal{P}}$ associated with $\widehat{K}$, which is given by

$$\Pi_{\widehat{K}} : \widehat{v} \longmapsto \sum_{i=1}^{m} \widehat{\sigma}_i(\widehat{v})\widehat{\psi}_i$$

For all $K \in \mathcal{M}$, we denote $V(K)$ as a Banach space of $\mathbb{R}^d$-valued functions, and there exists a linear bijective map given by

$$\phi_K : V(K) \longmapsto V(\widehat{K})$$

We assume that ϕ_K is defined for a whole patch as in IGA, not for a single element as in classic FEM. Thus, we can establish the following result:

Proposition 1 (Commutativity of the projection operator)**.** *An arbitrary element $K \in \mathcal{M}$, denoted as the triple $(K, \mathcal{P}_K, \mathcal{N}_K)$, satisfies*

$$
\begin{aligned}
K &= \mathcal{F}_K(\widehat{K}) \\
\mathcal{P}_K &= \left\{ \phi_K^{-1}(\widehat{p}),\, \widehat{p} \in \widehat{\mathcal{P}} \right\} \\
\mathcal{N}_K &= \{ \{\sigma_{K,i}\}_{i=1}^m : \sigma_{K,i} = \widehat{\sigma}_i(\phi_K(p)),\, \forall p \in \mathcal{P}_K \}
\end{aligned}
$$

The local shape functions and projection operator $\Pi_K : V(K) \longmapsto \mathcal{P}$ are

$$
\begin{aligned}
\psi_{K,i} &= \phi_K^{-1}(\widehat{\psi}_i) \\
\Pi_K : v &\longmapsto \sum_{i=1}^m \sigma_{K,i}(v)\psi_{K,i}
\end{aligned}
$$

Then the following diagram below commutes:

$$
\begin{array}{ccc}
V(K) & \xrightarrow{\phi_K} & V(\widehat{K}) \\
{\scriptstyle \Pi_K}\downarrow & & \downarrow{\scriptstyle \Pi_{\widehat{K}}} \\
\mathcal{P}_K & \xrightarrow{\phi_K} & \widehat{\mathcal{P}}
\end{array}
\tag{19}
$$

Proof. The geometry transform $\mathcal{F}_K$ is continuous, so the image of a compact set under $\mathcal{F}_K$ is also compact, and K will have the same set properties like $\widehat{K}$ [65]. Since ϕ_K is bijective, the reference basis $\widehat{\mathcal{P}}$ (inverse image of $\mathcal{P}$) will also serve as a basis for $\widehat{K}$. The same holds for the degrees of freedom, so K is a proper finite element. The commutative diagram (19) is valid because ϕ_K is linear:

$$
\Pi_{\widehat{K}}(\phi_K(v)) = \sum_{i=1}^m \widehat{\sigma}_i(\phi_K(v))\widehat{\psi}_i = \sum_{i=1}^m \sigma_{K,i}(v)\phi_K(\psi_{K,i}) = \phi_K(\Pi_K(v))
$$

Hence, the commutativity of the projection operator is established. □

Before starting the theoretical analysis, we will prove some propositions related to the reference element $\widehat{K}$. They will be crucial for deriving a priori estimates. In section 3.4 in [78], these propositions were proved by assuming that $\mathcal{F}_K$ is affine, i.e. on the form $A_K\widehat{x} + b_K$. From a technical point of view, we see that the matrix A_K corresponds to the Jacobian matrix of $\mathcal{F}_K$. We will generalize the procedure such that the propositions hold for a diffeomorphic vector function A_K, and its Jacobian is nonsingular on K. Furthermore, it should be noted that in IGA, the mapping $\mathcal{F}_K$ is local to a whole patch instead of a single element as in classical FEM, but these propositions will still be true. So if a set of elements belong to the same patch, then $\mathcal{F}_K$ for each element K will actually be the same map.

Proposition 2. *Let $\mathcal{M}$ be a finite element partition on a domain, $K \in \mathcal{M}$ is an arbitrary element, and $\widehat{K}$ is the reference element. For each element, we have a diffeomorphic mapping $\mathcal{F}_K = A_K(\widehat{x})$ for all $\widehat{x} \in \widehat{K}$. If $v \in W^{k,p}(K)$ for $k \in \mathbb{Z}^+$ and $\widehat{v} = v \circ \mathcal{F}_K$, then $\widehat{v} \in W^{k,p}(\widehat{K})$, and*

$$|\widehat{v}|_{W^{k,p}(\widehat{K})} \leq C_1 \|\mathcal{J}A_K\|^k \| \det(\mathcal{J}A_K)\|^{-1/p}_{L^\infty(K)} |v|_{W^{k,p}(K)} \tag{20a}$$

$$|\widehat{v}|_{W^{k,\infty}(\widehat{K})} \leq C_2 \|\mathcal{J}A_K\|^k |v|_{W^{k,\infty}(K)} \tag{20b}$$

$$|v|_{W^{k,p}(K)} \leq C_3 \|\mathcal{J}A_K^{-1}\|^k \| \det(\mathcal{J}A_K)\|^{1/p}_{L^\infty(K)} |\widehat{v}|_{W^{k,p}(\widehat{K})} \tag{20c}$$

$$|v|_{W^{k,\infty}(K)} \leq C_4 \|\mathcal{J}A_K^{-1}\|^k |\widehat{v}|_{W^{k,\infty}(\widehat{K})} \tag{20d}$$

where we have used the natural matrix norm $\|A\| = \max_{|w|=1} |Aw|$, and the arbitrary real constants $\{C_i\}_{i=1}^4$ depend on the Jacobian.

Proof. We consider first the case where $p \in [1, \infty)$. By using the chain rule, the definition of the $W^{k,p}$-seminorm, the standard change-of-variable formula for multidimensional integrals, and Hölder's L^p-inequality, we get

$$\begin{aligned}
|\widehat{v}|^p_{W^{k,p}(\widehat{K})} &= \sum_{|\alpha|=k} \int_{\widehat{K}} |D^\alpha \widehat{v}|^p \, d\widehat{\Omega} \\
&= \sum_{|\alpha|=k} \int_{\widehat{K}} |D^\alpha (v \circ \mathcal{F}_K)|^p \, d\widehat{\Omega} \\
&= \sum_{|\alpha|=k} \int_{\widehat{K}} |(D^\alpha v) \circ \mathcal{F}_K|^p |D^\alpha \mathcal{F}_K|^p \, d\widehat{\Omega} \\
&\leq \sum_{|\alpha|=k} \int_{\widehat{K}} \Big(C_1 \|\mathcal{J}A_K\|^k \Big)^p |(D^\alpha v) \circ \mathcal{F}_K|^p \, d\widehat{\Omega} \\
&= \Big(C_1 \|\mathcal{J}A_K\|^k \Big)^p \sum_{|\alpha|=k} \int_K |D^\alpha v|^p |\det(\mathcal{J}A_K^{-1})| \, d\Omega \\
&\leq \Big(C_1 \|\mathcal{J}A_K\|^k \Big)^p \sum_{|\alpha|=k} \| \det(\mathcal{J}A_K)\|^{-1}_{L^\infty(K)} \|(D^\alpha v)^p\|_{L^1(K)} \\
&= \Big(C_1 \|\mathcal{J}A_K\|^k \Big)^p \| \det(\mathcal{J}A_K)\|^{-1}_{L^\infty(K)} \sum_{|\alpha|=k} \|D^\alpha v\|^p_{L^p(K)} \\
&= \Big(C_1 \|\mathcal{J}A_K\|^k \Big)^p \| \det(\mathcal{J}A_K)\|^{-1}_{L^\infty(K)} |v|^p_{W^{k,p}(K)}
\end{aligned}$$

Taking p-root of both sides yields (20a). If $p = \infty$, the seminorm does not contain any integral, so changing variables is inapplicable. The domain and the solution are bounded, so we assume that $|(D^\alpha v) \circ \mathcal{F}_K| \leq c_2 |(D^\alpha v)|$. This yields the following derivation:

$$\begin{aligned}
|\widehat{v}|_{W^{k,\infty}(\widehat{K})} &= \max_{|\alpha|=k} \|D^\alpha \widehat{v}\|_{L^\infty(\widehat{K})} \\
&= \max_{|\alpha|=k} \|D^\alpha (v \circ \mathcal{F}_K)\|_{L^\infty(\widehat{K})} \\
&= \max_{|\alpha|=k} \left(\sup_{\mathbf{x}\in\Omega} |(D^\alpha v) \circ \mathcal{F}_K| \cdot |D^\alpha \mathcal{F}_K| \right) \\
&\leq c_1 \|\mathcal{J}A_K\|^k \left(\max_{|\alpha|=k} \sup_{\mathbf{x}\in\Omega} |(D^\alpha v) \circ \mathcal{F}_K| \right) \\
&\leq c_1 c_2 \|\mathcal{J}A_K\|^k \left(\max_{|\alpha|=k} \sup_{\mathbf{x}\in\Omega} |(D^\alpha v)| \right) \\
&= C_2 \|\mathcal{J}A_K\|^k |v|_{W^{k,\infty}(K)}
\end{aligned}$$

where $C_2 = c_1 c_2$, and estimate (20b) is proved. The two last estimates are derived in a similar way as the two first ones. Hence, we have also shown that the following limits below are true:

$$\lim_{p\to\infty} \|\det(\mathcal{J}A_K)\|_{L^\infty(K)}^{-1/p} = 1 \quad , \quad \lim_{p\to\infty} \|\det(\mathcal{J}A_K)\|_{L^\infty(K)}^{1/p} = 1 \qquad \square$$

Proposition 3. *Using the same assumptions as in the previous proposition, we have the following estimates:*

$$\|\mathcal{J}A_K\| \leq \frac{h_K}{\widehat{\rho}} \qquad , \qquad \|\mathcal{J}A_K^{-1}\| \leq \frac{\widehat{h}}{\rho_K} \tag{21}$$

Proof. We redefine the natural matrix norm as follows:

$$\|B\| = \max_{|w|=1} |Bw| = \frac{1}{\widehat{\rho}} \sup\{|B\xi| : |\xi| = \widehat{\rho}\}$$

If $\widehat{x}, \widehat{y} \in \widehat{K}$, we can introduce $\xi = \widehat{x} - \widehat{y}$, and $\|\xi\| \leq h_K$. Hence, we obtain

$$\|\mathcal{J}A_K(\xi)\| \leq \|\mathcal{J}A_K\| \cdot \|\xi\| \leq \frac{h_K}{\widehat{\rho}}$$

A similar procedure holds for the second estimate too. $\square$

3 A priori error estimation

In this section, we review some of the most relevant theory related to a priori estimation. There is a concise of B-spline approximation, interpolation, and quasi-interpolation, including the most important estimates related to these topics.

3.1 Main characteristics

In numerical computations, the chosen method satisfies two criterions:

- *Reliability*: Computational error is controlled within a tolerance level.
- *Efficiency*: Computational effort required to find a solution is minimal.

This is also important in finite element error estimation because the error can be generated from many different sources, and the discrete system of equations to be solved is often quite large, so we need an appropriate way for ensuring that nothing goes out of control. We do not consider the modelling error herein. The common sources of error in finite element modelling are

- *Spatial discretization error* from discrete Galerkin projection.
- *Temporal discretization error* from the chosen time-integrator.
- *Quadrature error* from the assembly and the post-processing.
- *Iterative method error* applied to the discrete system of equations.

A priori error estimation works well for problems with sufficient regularity. This method involves a thorough analysis of the PDE's weak formulation such that we can derive a suitable inequality describing the solution's global asymptotic behaviour. It works in many cases, but not for local refinement. This is because a priori estimation requires the following criterions [47]:

- *Accuracy*: Regarding the quantity of interest.
- *Stability*: Global measure of the discretization error's degree of interaction and accumulation in the total error.
- *Regularity*: The exact solution must possess certain differentiability.

We formulate general paradigm for a priori error estimation as follows:

$$\text{Small discretization error} + \text{Stability of discrete problem} \implies \text{Small error}$$

A priori analysis provides some insight of what can be classified as an optimal convergence rate, and the underlying prerequisites to be satisfied. The solution's regularity is not so restricting because we can overcome it by applying adaptive refinement. A priori analysis enables us to compare different methods and decide the kind of problems they are best suited for. Many standard a priori estimates can be used for analysing the quality of a posteriori estimators.

3.2 Underlying assumptions

Standard weak formulation of the BVP

In a priori error estimation, we assume first that $\mathcal{L}$ is a linear, second-order and uniformly elliptic partial differential operator, such that u is a unique solution of the strong problem $\mathcal{L}u = f$, where f is the continuous and inhomogeneous source term. Furthermore, we define V as a function space where u belongs to.

To be as general as possible, we assume that the BVP for our linear PDE is defined as follows:

$$\mathcal{L}u = f \quad , \quad \mathbf{x} \in \Omega \tag{22a}$$

$$u = 0 \quad , \quad \mathbf{x} \in \partial\Omega_D \tag{22b}$$

$$\mathbf{n}^T \mathbf{A} \nabla u = g_N \quad , \quad \mathbf{x} \in \partial\Omega_N \tag{22c}$$

where $\Omega \subset \mathbb{R}^d$ an open domain in d dimensions, and the segments $\partial\Omega_D$ and $\partial\Omega_N$ denote respectively the Dirichlet and Neumann boundaries such that we have $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$. The general expression for $\mathcal{L}$ is given by

$$\mathcal{L}u = -\nabla^T \mathbf{A} \nabla u + (\mathbf{b} \cdot \nabla)u + cu \tag{23}$$

where $\mathbf{A} \in W^{1,\infty}(\Omega)^{d \times d}$, $\mathbf{b} \in L^\infty(\Omega)^d$ and $c \in L^\infty(\Omega)$. We assume that $\mathbf{A}$ is positive definite, i.e. the eigenvalues are positive for any $\mathbf{x} \in \mathbb{R}^d$, and $\mathbf{n}$ is the normal unit vector on the boundary $\partial\Omega$.

For inhomogeneous Dirichlet conditions, we decompose u as $\overline{u} + \widetilde{u}$, where $\overline{u}$ solves the homogeneous problem, and $\widetilde{u}$ is an explicit prolongation which equals the inhomogeneous Dirichlet conditions at $\partial\Omega_D$. Otherwise, it may be chosen freely as long as it is smooth enough to belong to V [22]. This procedure yields an extra source term, so we assume hereafter homogeneous Dirichlet conditions for simplicity. After using Galerkin projection, the *weak formulation* becomes

$$u \in V \quad : \quad a(u, v) = l(v) \qquad \forall v \in V \tag{24}$$

where $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ and $l(\cdot) : V \to \mathbb{R}$ are respectively the bilinear and linear forms. After integration by parts, they are given by

$$a(u,v) = \int_\Omega \nabla u^T \mathbf{A} \nabla v + (\mathbf{b} \cdot \nabla u)v + cuv \, d\Omega \tag{25a}$$

$$l(v) = \int_\Omega fv \, d\Omega + \int_{\partial\Omega_N} g_N v \, ds \tag{25b}$$

Galerkin orthogonality and its consequences

Many of the a priori estimates relies on the fact that $a(\cdot,\cdot)$ is coercive:

$$\exists\, \alpha > 0 \quad : \quad a(u,u) \geq \alpha \|u\|_{H^1}^2 \tag{26}$$

This is possible to achieve if there are no first-order derivatives in (23), for the operator $\mathcal{L}$ will be self-adjoint. If $\mathcal{L}$ is not self-adjoint, $a(\cdot,\cdot)$ might not be coercive. Therefore, we can apply *Gårding's inequality* [36], which states that it is always possible to construct a real positive constant K for our problem such that

$$a(u,u) + K\|u\|_{L^2}^2 \geq \frac{\alpha}{2}\|u\|_{H^1}^2 \tag{27}$$

In any case, $a(\cdot,\cdot)$ induces the specific *energy norm* of equation (23):

$$|||u||| = \sup_{v \in V \setminus \{0\}} \frac{|a(u,v)|}{\|v\|} = \sqrt{a(u,u)} \tag{28}$$

This is an important norm used in the derivation and analysis of estimators. A main property provided by the energy norm is *Galerkin orthogonality* [23]. It states the following relation between the exact and numerical solutions:

$$a(u - u_h, v_h) = 0 \qquad v_h \in V \tag{29}$$

Theorem 1. *Let u be the exact solution of (23), and $u_{V^{(n)}}$ is a solution of (24), where $V^{(n)}$ is a discrete subspace of the trial space V. Then*

$$|||e_{V^{(n)}}||| = \min_{\chi \in V^{(n)}} |||u - \chi||| = \sqrt{|||u|||^2 - |||u_{V^{(n)}}|||^2} \tag{30}$$

Proof. We decompose $u - \chi$ to obtain the expansion below:

$$\begin{aligned}
|||u - \chi|||^2 &= |||(u - u_{V^{(n)}}) + (u_{V^{(n)}} - \chi)|||^2 \\
&= |||e_{V^{(n)}} + v_{V^{(n)}}|||^2 \\
&= |||e_{V^{(n)}}|||^2 + |||v_{V^{(n)}}|||^2 + 2a(e_{V^{(n)}}, v_{V^{(n)}}) \\
&= |||e_{V^{(n)}}|||^2 + |||v_{V^{(n)}}|||^2
\end{aligned}$$

The third term $2a(e_{V^{(n)}}, v_{V^{(n)}})$ vanishes because of Galerkin orthogonality. Let $\mathcal{B}u \in V^{(n)}$ be the best finite element approximation of u in the energy norm $|||\cdot|||$ over $V^{(n)}$. Using the previous derivation, we get

$$\begin{aligned} |||u - \mathcal{B}u|||^2 &= \min_{\chi \in V^{(n)}} |||u - \chi|||^2 \\ &= |||e_{V^{(n)}}|||^2 + \min_{\chi \in V^{(n)}} |||v_{V^{(n)}}|||^2 \\ &= |||e_{V^{(n)}}|||^2 \end{aligned}$$

The final identity (30) becomes true due to the following implication:

$$\chi \equiv 0 \implies |||u|||^2 = |||e_{V^{(n)}}|||^2 + |||v_{V^{(n)}}|||^2 \qquad \square$$

Theorem 2. *Let $\mathcal{J} : V \longmapsto \mathbb{R}$ be a functional given by*

$$\mathcal{J}(v) = \frac{1}{2}a(v,v) - f(v) \quad , \quad v \in V$$

If $V^{(n)} \subset V$ is a discrete subspace, then the following identity holds:

$$\mathcal{J}(u_{V^{(n)}}) = \min_{v \in V^{(n)}} \mathcal{J}(v) = -\frac{1}{2}|||u_{V^{(n)}}|||^2 \tag{31}$$

Proof. If $\chi \in V^{(n)}$, then we can state that

$$\begin{aligned} \mathcal{J}(u_{V^{(n)}} + \chi) &= \frac{1}{2}a(u_{V^{(n)}} + \chi, u_{V^{(n)}} + \chi) - f(u_{V^{(n)}} + \chi) \\ &= \frac{1}{2}\left[a(u_{V^{(n)}}, u_{V^{(n)}}) + 2a(u_{V^{(n)}}, \chi) + a(\chi, \chi)\right] \\ &\quad - f(u_{V^{(n)}}) - f(\chi) \\ &= \underbrace{\frac{1}{2}|||u_{V^{(n)}}|||^2 - f(u_{V^{(n)}})}_{\mathcal{J}(u_{V^{(n)}})} + \frac{1}{2}|||\chi|||^2 + \underbrace{a(u_{V^{(n)}}, \chi) - f(\chi)}_{0} \end{aligned}$$

The term $u_{V^{(n)}}$ minimizes $\mathcal{J}$ over $V^{(n)}$ because

$$\mathcal{J}(u_{V^{(n)}}) \leq \mathcal{J}(u_{V^{(n)}} + \chi)$$

The final result follows directly as shown below:

$$\begin{aligned} \mathcal{J}(u_{V^{(n)}}) &= \frac{1}{2}|||u_{V^{(n)}}|||^2 - f(u_{V^{(n)}}) \\ &= \frac{1}{2}|||u_{V^{(n)}}|||^2 - |||u_{V^{(n)}}|||^2 \\ &= -\frac{1}{2}|||u_{V^{(n)}}|||^2 \end{aligned} \qquad \square$$

Corollary 1. *Let* $\{V^{(i)}\}_{i=1}^{n}$ *be a sequence of finite-dimensional spaces, and*

$$V^{(n)} \subset V^{(n-1)} \subset \cdots \subset V^{(2)} \subset V^{(1)} \subseteq V$$

If $\{u_{V^{(i)}}\}_{i=1}^{n}$ *is the corresponding Galerkin approximation sequence, then*

$$|||u_{V^{(n)}}||| \leq \cdots \leq |||u_{V^{(1)}}||| \tag{32}$$

Proof. First, we utilize an important implication:

$$V^{(2)} \subseteq V^{(1)} \implies \min_{v \in V^{(2)}} \mathcal{J}(v) \leq \min_{v \in V^{(1)}} \mathcal{J}(v)$$

By invoking identity (31), we obtain

$$-\frac{1}{2}|||v_{V^{(2)}}|||^2 = \min_{v \in V^{(2)}} \mathcal{J}(v) \leq \min_{v \in V^{(1)}} \mathcal{J}(v) = -\frac{1}{2}|||v_{V^{(1)}}|||^2$$
$$|||v_{V^{(1)}}|||^2 \leq |||v_{V^{(2)}}|||^2$$

Thus, we get a new inequality:

$$|||e_{V^{(2)}}|||^2 = |||u|||^2 - |||u_{V^{(2)}}|||^2 \leq |||u|||^2 - |||u_{V^{(1)}}|||^2 = |||e_{V^{(1)}}|||^2$$

This telescoping property of $\{V^{(i)}\}_{i=1}^{n}$ means that (32) holds. □

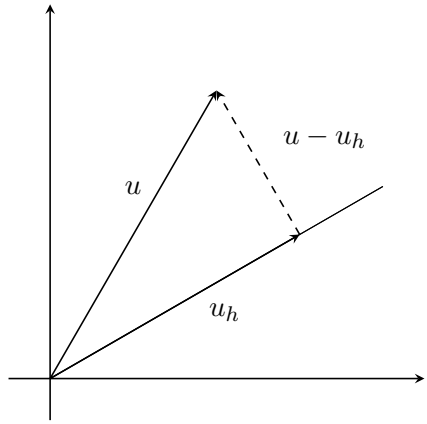


Figure 8. Geometric representation of Galerkin orthogonality

Equivalent adjoint formulation

In the development of new error estimators, it can be actual to utilize the *adjoint formulation* of the weak formulation, which simplifies the derivation and might yield interesting results. First, we need some general formality. If X and Y are Banach spaces, then $\mathcal{L}(X,Y)$ is the space of continuous linear mappings $L : X \longmapsto Y$ with the finite norm

$$\|L\|_{\mathcal{L}(X,Y)} = \sup_{\varphi \in X \setminus \{0\}} \frac{\|L\varphi\|_Y}{\|\varphi\|_X}$$

Thus, the dual space of Y, $Y^* = \mathcal{L}(Y,\mathbb{R})$, consists of continuous linear functionals on Y. The dual mapping $L^* : Y^* \longmapsto X^*$ satisfies

$$\begin{aligned} \langle L^* y, x \rangle &= \langle y, Lx \rangle \\ \|L^*\|_{\mathcal{L}(Y^*,X^*)} &= \|L\|_{\mathcal{L}(X,Y)} \end{aligned}$$

We denote $\mathcal{L}^2(X,Y,\mathbb{R})$ as the space of bilinear mappings $a : X \times Y \longmapsto \mathbb{R}$ with the finite norm given by

$$\|a\|_{\mathcal{L}^2(X,Y,\mathbb{R})} = \sup_{\varphi \in X \setminus \{0\}} \sup_{\psi \in Y \setminus \{0\}} \frac{|a(\varphi,\psi)|}{\|\varphi\|_X \|\psi\|_Y}$$

As a consequence of *Riesz's representation theorem*, we can express our bilinear form uniquely by the inner product with a canonical isometry [88]:

$$a(\varphi,\psi) = \langle A\varphi, \psi \rangle \tag{34}$$

We assume that the standard formulation (24) is known. Since $a(\cdot,\cdot)$ is on the form $V \times V \longmapsto \mathbb{R}$, it implies that $A : V \longmapsto V^*$ is a linear and bounded elliptic operator.
Canonical isomorphisms can only be achieved for Hilbert spaces, for in this case, the transpose of any operator is adjoint, and we define it as

$$L^T : Y \longmapsto X \qquad (L^T y, x)_X = (y, Lx)_Y$$

If $\Lambda : X \longmapsto Y$ is an isomorphism and $X \subseteq Y$, it is canonical, and

$$\Lambda_X L^T = L^* \Lambda_Y \qquad \begin{array}{ccc} Y & \xrightarrow{L^T} & X \\ {\scriptstyle \Lambda_Y}\downarrow & & \downarrow{\scriptstyle \Lambda_X} \\ Y^* & \xrightarrow[L^*]{} & X^* \end{array} \tag{35}$$

3.3 Some classical lemmas

Lemma 1 (Generalized Céa's lemma)**.** *If $a(\cdot,\cdot)$ is continuous, the numerical solution of the variational problem (24) satisfies a special estimate:*

$$\|u-u_h\|_{H^1} \leq \frac{2K}{\alpha}\|u-u_h\|_{H^1} + \frac{2M}{\alpha}\inf_{v_h\in V_h}\|u-v_h\|_{H^1} \tag{36}$$

Proof. Assume that $a(\cdot,\cdot)$ is not coercive. We combine Gårding's inequality with the Galerkin orthogonality and continuity of $a(\cdot,\cdot)$ to obtain:

$$\begin{aligned}\frac{\alpha}{2}\|u-u_h\|_{H^1}^2 &\leq K\|u-u_h\|_{L^2}^2 + a(u-u_h,u-u_h)\\ &\leq K\|u-u_h\|_{H^1}^2 + a(u-u_h,u-v_h)\\ &\leq K\|u-u_h\|_{H^1}^2 + M\|u-u_h\|_{H^1}\|u-v_h\|_{H^1}\end{aligned}$$

We divide everything with $(\alpha/2)\|u-u_h\|_{H^1}$ and take infimum with respect to v_h on both sides. This yields the final result. □

It should be noted that in the case where $a(\cdot,\cdot)$ is coercive, the first term on the right-hand side of (36) vanishes, and $2M/\alpha$ becomes just M/α.

Lemma 2 (Deny-Lions lemma)**.** *The seminorm of $W^{k+1,p}$ is equivalent to the quotient norm of $W^{k+1,p}\backslash\mathbb{P}^k$, i.e. there is a constant $C>0$ such that*

$$\forall v\in W^{k+1,p}(\Omega) \quad , \quad \inf_{\widehat{p}\in\mathbb{P}^k(\Omega)}\|v+\widehat{p}\|_{W^{k+1,p}(\Omega)} \leq C|v|_{W^{k+1,p}(\Omega)} \tag{37}$$

where $\mathbb{P}^p(\Omega)$ is the space of polynomials of degree at most p in each variable. A general proof of this lemma can be found in [39], which is based on the *Hahn-Banach extension theorem* and *Sobolev's embedding theorem*.

Lemma 3 (Bramble-Hilbert lemma)**.** *Let $\Omega\subset\mathbb{R}^d$ be an open domain with a connected Lipschitz boundary, and p and q are conjugate exponents. Let g be a continuous and linear functional on $W^{k+1,p}(\Omega)$ with the property*

$$g(p)=0 \quad , \quad \forall p\in\mathbb{P}^k(\Omega)$$

Then there is a constant C such that

$$|g(v)| \leq C\|g\|_{W^{k+1,q}(\Omega)}|v|_{W^{k+1,p}(\Omega)} \tag{38}$$

Proof. The properties of g imply that we have:

$$\begin{aligned} |g(v)| = |g(v+p)| &\leq \|g\|_{W^{k+1,q}(\Omega)} \|v+p\|_{W^{k+1,p}(\Omega)} \\ \implies |g(v)| &\leq \|g\|_{W^{k+1,q}(\Omega)} \inf_{p \in \mathbb{P}^k(\Omega)} \|v+p\|_{W^{k+1,p}(\Omega)} \end{aligned}$$

The conclusion follows directly from the Deny-Lions lemma (37). □

Lemma 4 (Aubin-Nitsche lemma)**.** *Let H be a Hilbert space with $\overline{V} = H$ and $V \hookrightarrow H$, such that V is a subspace of H equipped with a continuous injection $l : V \longmapsto H$. Define the dual variational problem as*

$$\phi_f \in V \quad : \quad a(v, \phi_f) = (v, f) \quad , \quad \forall v \in V$$

where $f \in H$ and $v \in V$. Then the following estimate holds:

$$|u - u_h| \leq M \|u - u_h\| \left(\sup_{f \in H \setminus \{0\}} \left\{ \frac{1}{|f|} \inf_{\phi_h \in V_h} \|\phi_f - \phi_h\| \right\} \right) \tag{39}$$

Proof. From the Galerkin orthogonality, $a(u - u_h, \phi_h) = 0$, so we get

$$\begin{aligned} a(u - u_h, \phi_f) &= a(u - u_h, \phi_f - \phi_h) \\ &= (f, u - u_h) \end{aligned}$$

Since $a(\cdot, \cdot)$ is continuous, we obtain

$$|(f, u - u_h)| \leq M \|u - u_h\| \inf_{\phi_h \in V_h} \|\phi_f - \phi_h\|$$

By the definition of norms, we know that

$$|u - u_h| = \sup_{f \in H \setminus \{0\}} \frac{|(f, u - u_h)|}{|f|}$$

Hence, the final conclusion holds. □

All these classical lemmas play a vital role in FEA because they can be used for proving that error estimators are robust. It should also be noted that they are independent of the choice of basis functions. Therefore, they can be applied directly to IGA.

3.4 Polynomial interpolation theory

Classical interpolation estimates in 1D

First, we prove a priori estimates in one spatial dimension, and then adapt them to several dimensions. They are used for analysing the approximation properties of splines. We start with a classical interpolation result:

Theorem 3 (General interpolation estimate)**.** *Assume that* $f \in C^{n+1}([a,b])$ *with* $|f^{(n+1)}(x)| \leq M$*, and* $\Pi \in \mathbb{P}([a,b])$ *interpolates* f *at* $n+1$ *equally spaced and distinct nodes* $\{x_i\}_{i=0}^n$ *(including endpoints) on* $[a,b]$*. Then*

$$|f(x) - \Pi(x)| \leq \frac{M}{4(n+1)} h^{n+1} \qquad , \qquad h = \frac{b-a}{n} \tag{40}$$

Proof. The proof consists of two stages. In the first stage, we define

$$\begin{aligned} P(x) &= \prod_{i=0}^{n} (x - x_i) && \text{(Global node polynomial)} \\ c(t) &= \frac{f(t) - \Pi(t)}{P(t)} && \text{(Constant expression)} \\ d(x) &= f(x) - \Pi(x) - c(t)P(x) && \text{(Auxiliary function)} \end{aligned}$$

From *Hôpital's rule*, $d(x_i) = 0$. Elsewhere, the function is continuous. From *Rolle's theorem*, we know that between any two zeros x_i and x_{i+1} of d, d' has a root, so d' has at least $n+1$ nodes in total. Repeating the same argument n more times, $d^{(n+1)}$ has at least one root ζ, so

$$f^{(n+1)}(\zeta) - \Pi^{(n+1)}(\zeta) - c(t)P^{(n+1)}(\zeta) = 0$$

We have the following implications:

$$\begin{aligned} \Pi \in \mathbb{P}([a,b]) \quad &\Longrightarrow \quad \Pi^{(n+1)}(x) \equiv 0 \\ \deg(P) = n+1 \quad &\Longrightarrow \quad P^{n+1}(x) \equiv (n+1)! \end{aligned}$$

Using the definition of c and rearranging the terms, we get

$$f(x) - \Pi(x) = \frac{f^{(n+1)}(t)}{(n+1)!} \prod_{i=0}^{n} (x - x_i) \tag{42}$$

For each $x \in [a,b]$, there is a $t \in (a,b)$ such that the equality above holds. The next stage is finding an upper estimate for P. Assume that $x_j \leq x \leq x_{j+1}$ for some $j \in [0, n-1]$. If $g(x) = (x - x_j)(x - x_{j+1})$, then $g'(x) = 2x - x_j - x_{j+1}$, and $\hat{x} = (x_j + x_{j+1})/2$ is the coordinate of

the extremal value $-h^2/4$. We know that $h = x_{j+1} - x_j$ for any j since we have a uniform partition on $[a, b]$.

Now, we can find a maximal upper bound for the absolute value of the global node polynomial P. The main idea of this estimate is splitting the big product in three parts and finding upper bounds for each of them, which are later combined into a final upper bound. We split as follows:

$$\begin{aligned}
\prod_{i=0}^{n} |x - x_i| &= \prod_{i=0}^{j-1} (x - x_i)|x - x_j||x - x_{j+1}| \prod_{i=j+2}^{n} (x_i - x) \\
&\leq \frac{h^2}{4} \prod_{i=0}^{j-1} (x - x_i) \prod_{i=j+2}^{n} (x_i - x) \\
&\leq \frac{h^2}{4} \prod_{i=0}^{j-1} (x_{j+1} - x_i) \prod_{i=j+2}^{n} (x_i - x_j) \\
&= \frac{h^2}{4} h^j h^{n-j-1} \prod_{i=0}^{j-1} (j - i + 1) \prod_{i=j+2}^{n} (i - j) \\
&\leq \frac{h^{n+1}}{4} (j+1)!(n-j)! \\
&\leq \frac{h^{n+1}}{4} n!
\end{aligned}$$

It can be shown by induction that $(j+1)!(n-j)! \leq n!$ holds for $j \in [0, n-1]$. By taking absolute value of (42) and using the result above, we obtain

$$|f(x) - \Pi(x)| \leq \frac{M}{4(n+1)} h^{n+1}$$

Hence, the desired conclusion has been proved. □

This general interpolation estimate is still valid although the partition on $[a, b]$ is non-uniform. The only adjustment required is denoting h as the maximal mesh size of the partition. Then the inequality holds, and the error is automatically bounded.

In general approximation theory, it is well-known that equidistant nodes on the domain might cause unexpected trouble like wild oscillations. This behaviour depends often on the function itself, but we know that adding more interpolation nodes increases the interpolation polynomial's degree quite much. Thus, it is convenient to use non-equidistant nodes to minimize the upper bound of the global node polynomial. A celebrated interpolation technique used for smoothing out wild oscillations is using the zeros of

Chebyshev polynomials. We refer to [53, 82] for a thorough description of their main properties. The most interesting feature is orthogonality with respect to the Chebyshev measure $(1-x^2)^{-1/2}\,dx$, which implies that we have an analytical formula for the zeros:

$$x_j = \cos\left(\frac{2j+1}{2n+2}\pi\right) \quad , \quad 0 \leq j \leq n \tag{43}$$

By using these nodes, we can establish an important approximation estimate:

Theorem 4 (Chebyshev interpolation estimate)**.** *Define $f \in C^{n+1}([a,b])$ with $|f^{(n+1)}(x)| \leq M$, and $\Pi \in \mathbb{P}^n([a,b])$ is interpolating f at the shifted zeros of the Chebyshev polynomial T_{n+1}. Then we have the bound*

$$|f(x) - \Pi(x)| \leq \frac{(b-a)^{n+1}}{2^{2n+1}(n+1)!}M \tag{44}$$

Proof. If $\{t_i\}_{i=0}^n$ is the set of zeros of T_{n+1}, we can utilize the identity

$$\prod_{i=0}^{n}(t-t_i) = 2^{-n}T_{n+1}(t)$$

We define the linear transformation $\mathcal{T} : [-1,1] \longmapsto [a,b]$ as

$$x = \frac{b-a}{2}t + \frac{b+a}{2}$$

Thus, we get the following equality:

$$\prod_{i=0}^{n}|t-t_i| = \left(\frac{b-a}{2}\right)^{n+1}\prod_{i=0}^{n}|x-\mathcal{T}(t_i)| = \left(\frac{b-a}{2}\right)^{n+1}2^{-n}|T_{n+1}(x)|$$

Since $|T_{n+1}(x)| \leq 1$, the final estimate follows directly. □

Now, we use the interpolation inequality (42) to establish a classical result which holds for Lagrange interpolants from classical FEM.

Theorem 5 (A priori interpolation estimate in H^r-seminorm)**.** *Let $u \in H^{r+1}(I)$, $\Pi^r \in \mathbb{P}^r \cap C^0$ is the classical FEM-interpolant, and $I = (a,b)$ is an open interval with regular partition. Then we have the a priori estimate*

$$|u - \Pi^r u|_{H^k(I)} \leq Ch^{r+1-k}|u|_{H^{r+1}(I)} \quad , \quad 0 \leq k \leq r+1 \tag{45}$$

Proof. We start with the initial case $k = 0$, where $H^0 \equiv L^2$ by definition. The interval is split up in n elements. We define $I_j = [x_j - x_{j-1}]$ as the j-th subinterval of I with the local mesh width $h_j = x_j - x_{j-1}$. By using inequality (40) and denoting the local generic constant as C_j, we get

$$\begin{aligned}
\|u - \Pi^r u\|^2_{L^2(I)} &= \sum_{j=1}^{n} \|u - \Pi^r u\|^2_{L^2(I_j)} \\
&\leq \sum_{j=1}^{n} \left(C_j h_j^{r+1}\right)^2 |u|^2_{H^{r+1}(I_j)} \\
&\leq \left(\max_{1\leq j\leq n} C_j\right)^2 \left(\max_{1\leq j\leq n} h_j^{r+1}\right)^2 \sum_{j=1}^{n} |u|^2_{H^{r+1}(I_j)} \\
&= \left(Ch^{r+1}\right)^2 |u|^2_{H^{r+1}(I)}
\end{aligned}$$

Taking the square root of the inequality yields the desired result. If $k = 1$, then the exponent of h drops from $r + 1$ to r. This is because u was interpolated by a polynomial of degree r, and we had $r + 1$ degrees of freedom. But the derivative is interpolated by another polynomial of degree $r - 1$, so we lose one degree of freedom, and this makes the exponent drop down to r. Repeating the same argument k times, where $0 \leq k \leq r + 1$, the final result is established. □

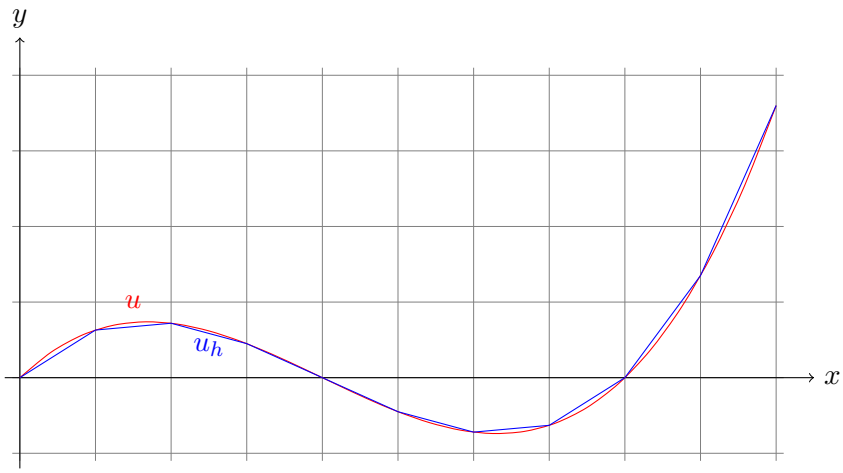
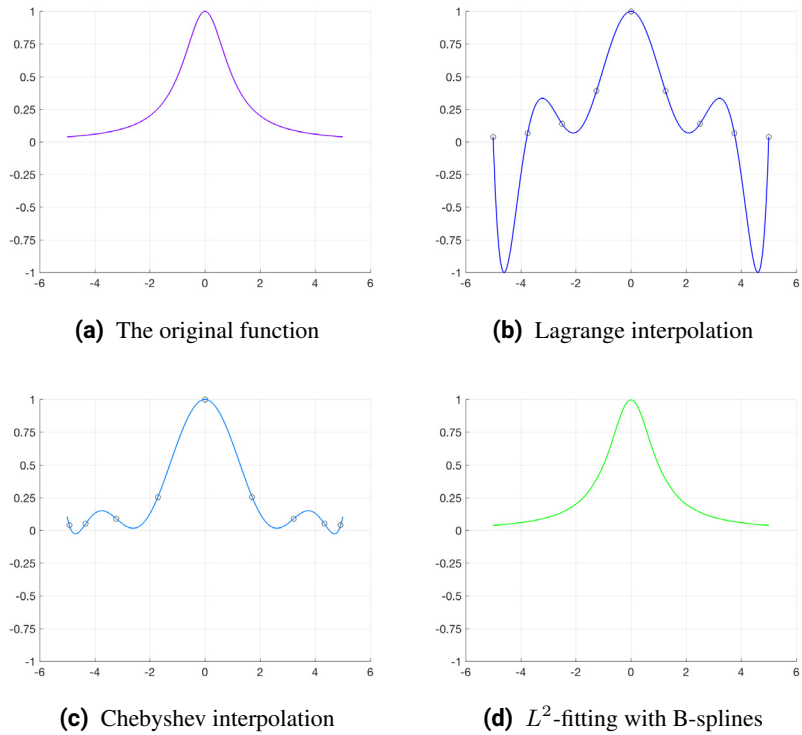


Figure 9. Linear piecewise interpolation of a continuous function.



(a) The original function

(b) Lagrange interpolation

(c) Chebyshev interpolation

(d) L^2-fitting with B-splines

Figure 10. The Runge phenomenon handled in three different ways.

As an illustration of the theory above, we look at a famous interpolation result known as *Runge's phenomenon*. The function under consideration is

$$f(x) = \frac{1}{x^2 + 1}$$

In the first interpolation with nine equidistant nodes, the global Lagrange polynomial oscillates wildly and deteriorates the approximation quality. But if we use nine Chebyshev nodes instead, the oscillation is significantly reduced. Lastly, the oscillation is smoothed out by using 8th degree B-splines on 20 subintervals on the domain. Thus, we see that the last approach yields an approximation that almost coincides with the original function.

B-spline estimates in 1D

We turn our attention to the estimates where a function is approximated by B-splines instead of Lagrange interpolants. The new procedure is almost the same as before. First, we need some new notation. According to [77], we can write a single B-spline basis function on the following form:

$$N_i(x) = \sum_{j=i}^{i+p+1} \prod_{\substack{k=i \\ k\neq j}}^{i+p+1} \frac{1}{\xi_K - \xi_j}(x-\xi_j)_+^p \tag{46}$$

$$(x-\xi_j)_+ = \max\{0, x-\xi_j\}$$

We recognize $\Xi = \{\xi_i\}_{i=1}^{n+p+1}$ as the standard knot vector on $[a,b]$, which generates the familiar partition $\mathcal{M}$, given by

$$\underbrace{a = \xi_1 = \cdots = \xi_{p+1}}_{\text{End knots}} < \underbrace{\xi_{p+2} < \cdots < \xi_N}_{\text{Interior knots}} < \underbrace{\xi_{n+1} = \cdots = \xi_{n+p+1} = b}_{\text{End knots}}$$

It is important to recall that n is the total number of B-spline basis functions, while $N = n - p$ is the number of subintervals on $[a,b]$, assuming full continuity and no repeated interior knots for the sake of simplicity.

Theorem 6 (A priori B-spline estimate in H^r-seminorm)**.** *Let $u \in H^{r+1}(I)$, $\Pi_K^r u \in \mathbb{P}^r(\Omega) \cap C^k(\Omega)$ is the B-spline approximation, and the domain $I = (a,b)$ has a regular partition. Then we have the a priori estimate*

$$|u - \Pi^r u|_{H^k(I)} \leq Ch^{r+1-k}|u|_{H^{r+1}(I)} \qquad , \qquad 0 \leq k \leq r+1 \tag{47}$$

Proof. The proof follows the same argument as the end of the proof for Theorem 5. The difference is that splines are used for the approximation. □

Multidimensional estimates

We express the numerical solution by the finite element operator Π_K^p, such that $u_h = \Pi_K^p u$. This operator belongs to V_h, a finite-dimensional subspace of V, and approximates u in $\mathbb{P}^p(\overline{\Omega}) \cap C^k(\overline{\Omega})$. Since the shape functions are splines, the continuity restriction is $0 \leq k \leq p-1$.

Theorem 7 (Multidimensional a priori estimate in H^m-seminorm)**.** *Let $u \in C^k(\Omega)$, and denote $u_h = \Pi_m^{p+1} u$ as a finite element approximation of u. If $0 \leq m < k$, then the general elliptic estimate in $H^m(\Omega)$ is*

$$|u - u_h|_{H^m(\Omega)} \leq Ch^{p+1-m}|u|_{H^m(\Omega)} \tag{48}$$

Proof. From equation (20c) and (21), we can deduce:

$$\begin{aligned}|u-u_h|_{H^m(K)} &\leq C\|\mathcal{J}A_K^{-1}\|^m\|\det(\mathcal{J}A_K)\|^{1/2}_{L^\infty(K)}|\widehat{u}-\widehat{u}_h|_{H^m(\widehat{K})}\\ &\leq C\left(\frac{\widehat{h}}{\rho_K}\right)^m\|\det(\mathcal{J}A_K)\|^{1/2}_{L^\infty(K)}|(I-\Pi_m^{p+1})\widehat{u}|_{H^m(\widehat{K})}\end{aligned}$$

We know that $(I-\Pi_m^{p+1})(q)=0$ because q is invariant on $\mathbb{P}^p(\Omega)$, so if we use the Deny-Lions lemma (37), we obtain:

$$\begin{aligned}|u-u_h|_{H^m(K)} &\leq C\left(\frac{\widehat{h}}{\rho_K}\right)^m\|\det(\mathcal{J}A_K)\|^{1/2}_{L^\infty(K)}\inf_{\widehat{q}\in\mathbb{P}^p}\|\widehat{u}+\widehat{q}\|_{H^{p+1}(\widehat{K})}\\ &\leq C\left(\frac{\widehat{h}}{\rho_K}\right)^m\|\det(\mathcal{J}A_K)\|^{1/2}_{L^\infty(K)}|\widehat{u}|_{H^{p+1}(\widehat{K})}\end{aligned}$$

Applying equation (20a) and the Bramble-Hilbert lemma (38), we get:

$$\begin{aligned}|u-u_h|_{H^m(K)} &\leq C'\left(\frac{\widehat{h}}{\rho_K}\right)^m\left(\frac{h_K}{\widehat{\rho}}\right)^{p+1}|u|_{H^{p+1}(K)}\\ &\leq C'D(\widehat{h}\kappa_K)^m(\widehat{\rho})^{-p-1}h_K^{p+1-m}|u|_{H^m(K)}\end{aligned}$$

We know that $\widehat{h}$ and $\widehat{\rho}$ are constant on the reference element, and the maximal value of κ_K is the shape parameter $C_{\mathcal{M}}$. Summing over all the elements and combining everything into a generic constant C yields inequality (48). □

It should be noted that the approximation error drops down to machine precision immediately if $m \geq k$, for it is possible to represent any function of finite continuity exactly as a finite linear combination of (piecewise) polynomials of sufficiently high degree. If the function is non-polynomial, this drop does not occur. The error decreases when the polynomial degree increases, which holds for any smooth function.

Furthermore, we remark that the FEM-approximation estimate holds for IGA too although splines are non-interpolatory in general. The difference is that the splines have much higher continuity, which in turn makes the generic constant C in the inequality significantly lower, a crucial ingredient for optimal reduction of global approximation error [28, 86, 85, 49].

Now, we establish another useful estimate in the energy norm [63]:

Theorem 8 (General a priori estimate in energy norm)**.** *Given a function u solving equation (22), let p be the polynomial degree of the finite element approximation, and the parameter $\lambda \in \mathbb{R}$ characterizes the strength of the solution's singularities. Thus, we define $\beta = \min\{p, \lambda\}$. The following estimates hold in the energy norm induced by the bilinear form $a(\cdot, \cdot)$:*

$$\textit{Smooth:} \qquad |||u - u_h||| \leq Ch^p \|u\|_{H^{p+1}} \tag{49a}$$

$$\textit{Non-smooth:} \qquad |||u - u_h||| \leq Ch^\beta \|u\|_{H^{\beta+1}} \tag{49b}$$

Proof. Since $a(\cdot, \cdot)$ is continuous, there is a real constant $M > 0$ such that

$$|a(u - u_h, u - v_h)| \leq M \|u - u_h\|_V \|u - v_h\|_V$$

To be as general as possible, we assume that $a(\cdot, \cdot)$ is not coercive and invoke Gårding's inequality (3.3). Combining this with the continuity, we obtain

$$\|u - u_h\|_V \leq \frac{2(K + M)}{\alpha} \|u - v_h\|_V$$

Next, we assume that there is a constant D such that $\|u\|_V \leq D|u|_{H^{p+1}}$. By invoking the previous approximation result and applying the universal fact that $|u|_{H^{p+1}} \leq \|u\|_{H^{p+1}}$ for any u, we obtain the smooth estimate (49a). A similar procedure would hold for the other. □

This theorem is very useful for insufficiently smooth functions. If there are no singularities, then $\beta = p$, whereas for the case $\beta < p$ the convergence is governed by the unknown solution's regularity, not the polynomial order of the shape functions [83]. However, proper adaptive mesh refinement may circumvent this loss of convergence rate. The global asymptotic convergence rate of the error in the H^k-norm is given by

$$\mathcal{O}\left(N_{\text{dof}}^{-\frac{p+1-k}{2^{d-1}}}\right) \tag{50}$$

If the PDE is time-dependent, then the two sources of inaccuracy are spatial and temporal error. Hence, the general a priori estimate becomes

$$|||u - \Pi_h u||| \leq \{C_1 h^p + C_2 (\Delta t)^s\} \|u\|_{H^{\beta+1}} \tag{51}$$

We denote s as the order of the chosen time-integrator, and Δt is the uniform time step on $[0, T]$. Although we have assumed that $\mathcal{L}$ is a second-order operator, the same estimates hold for higher order elliptic operators too.

Comparison with NURBS

The presented estimates for B-splines are the same for NURBS, but the proofs for these estimates are complicated. This is because we are using rational B-splines, and the mapping $\mathcal{F} : \widehat{\Omega} \to \Omega$ might not be regular among mesh lines. Even if the scalar function f belongs to $H^k(\Omega)$, it does not guarantee that $\widehat{f} = f \circ \mathcal{F}$ is in $H^k(\widehat{\Omega})$ too. These complexities related to NURBS implies that the proofs of their approximation properties require more technical facilities. Some of them include *bent Sobolev spaces* and special projection operators, whose properties have been studied extensively and verified in [28, 84, 85].

3.5 Least-squares approximation

A common procedure for discrete and continuous curve fitting is *least-squares approximation*, which minimizes the error in the L^2-norm [35]. The same procedure can be adapted to B-splines, which is more optimal because the shape functions are piecewise instead of global. This yields higher local control of the approximated curve, and the linear system of equations to solve is not so ill-conditioned, even if we increase the polynomial degree to a high level. We derive the minimization problem as follows:

$$\min_{\mathbf{u}\in\mathbb{R}^N} \int_I \left| f(x) - \sum_{i=1}^{N} c_i \psi_i(x) \right|^2 dx = \min_{\mathbf{u}\in\mathbb{R}^N} \|f(x) - \mathbf{\Psi}(x)^T \mathbf{c}\|_{L^2}^2$$

Here, $\mathbf{c}$ is the coefficient vector, and $\mathbf{\Psi}$ is the vector with shape functions on the partition of the interval I. Taking the gradient with respect to $\mathbf{c}$ yields

$$\int_I 2\mathbf{\Psi}(x)\mathbf{\Psi}(x)^T \mathbf{c} - 2\mathbf{\Psi}(x) f(x)\, dx = 0$$

$$\int_I \psi_i(x)\psi_j(x) c_j \, dx = \int_I f(x)\psi_i(x)\, dx \quad , \quad 1 \leq i, j \leq n$$

$$\mathbf{M}\mathbf{c} = \mathbf{f}$$

where $\mathbf{M}$ and $\mathbf{f}$ are the mass matrix and load vector, respectively. If f is an r-degree polynomial, and the B-splines have degree $p \geq r$, then the least-squares fitting yields an exact representation without any error. Since we use piecewise polynomials instead of a single polynomial for the entire approximation, the global error becomes minimized.

3.6 Quasi-interpolation

General construction

Another common approximation method is *quasi-interpolation* [85]. Splines are not interpolatory in general, so we define a projection in form of a dual basis and dual functionals. This yields a spline-preserving quasi-interpolant.

Definition 7. *Let Ξ be an open knot vector of degree p. Then we can define a spline quasi-interpolation operator $\Pi_{p,\Xi} : C^\infty([0,1]) \to \mathbb{S}^p(\Xi)$ as*

$$\Pi_{p,\Xi}(f) = \sum_{j=1}^{n} \lambda_{j,p}(f) N_{j,p} \tag{52a}$$

$$\lambda_{j,p}(N_{k,p}) = \delta_{jk} \tag{52b}$$

where we have used the following formulas:

$$\lambda_{j,p}(f) = \int_{\xi_j}^{\xi_{j+p+1}} f(s) \frac{d^{p+1}}{ds^{p+1}} \phi_j(s)\, ds \tag{53a}$$

$$\phi(\xi) = \left[\frac{1}{p!} \prod_{j=1}^{p} (\xi - \xi_{j+i}) \right] g\left(\frac{2\xi - \xi_j - \xi_{j+p+1}}{\xi_{j+p+1} - \xi_j} \right) \tag{53b}$$

Here, g is a *transition function* expressed in terms of *perfect splines* [33, 70]. If the polynomial degree is p, the knots are $\{\xi\}_{i=1}^{m}$, and we choose some real constants γ and $\{\alpha\}_{i=1}^{p}$, then the perfect spline is given by

$$P(t) = \sum_{i=1}^{p} \alpha_i t^{i-1} + \gamma \left(t^p + 2 \sum_{k=1}^{m} (-1)^k (t - \xi_K)_+^p \right). \tag{54}$$

The procedure of constructing perfect splines originates from *Favard's interpolation problem*, an extremal problem in $W^{p,\infty}([-1,1])$. We create a function f in this space satisfying the following interpolation conditions:

$$f(-1) = 0\ ,\ f(1) = 1$$
$$f^{(j)}(-1) = f^{(j)}(1) = 0\ ,\ 1 \le j \le p-1$$

It can be shown that the only function minimizing $\|f^{(p)}\|_\infty$ optimally is

$$\psi(x) = \frac{(-1)^{p-1}}{2} \int_{-1}^{1} (x-t)_+^{p-1} \operatorname{sgn}\left[\frac{U_p(t)}{\sqrt{1-t^2}} \right] dt$$

where $U_p(x) = \sin(p\cos^{-1}(x))$ is the second kind Chebyshev polynomial. We can prove from this construction that the knots $\{\xi\}_{i=1}^m$ become extremal points of $T_p(x) = \cos(p\cos^{-1}(x))$, the first kind Chebyshev polynomial:

$$\xi_i = \cos\left[\left(\frac{m-i}{m}\right)\pi\right] \quad , \quad 0 \leq i \leq p$$

As an example, the first and second order perfect B-splines are

$$B_1^*(x) = (x+1)\chi_{[-1,0]} + (1-x)\chi_{[0,1]}$$
$$B_2^*(x) = 2(x+1)^2\chi_{[-1,-1/2]} + (1-2x^2)\chi_{[-1/2,1/2]} + 2(1-x)^2\chi_{[1/2,1]}$$

Finally, the transition function g becomes [80]

$$g(x) = \begin{cases} 0, & x < -1 \\ \int_{-1}^x B_m^*(t)\,dt, & -1 \leq x < 1 \\ 1, & x \geq 1 \end{cases}.$$

To create commuting projectors, we define a new spline preserving quasi-interpolant on $\Xi' = \{\xi_2, \ldots, \xi_{n+p}\}$ such that

$$\Pi_{p-1,\Xi'}^c g = \frac{d}{d\xi}\Pi_{p,\Xi}\int_0^\xi g(s)\,ds = \sum_{j=1}^{n-1}\lambda_{j,p-1}^c(g)\widehat{N}_{j,p-1}$$
$$\lambda_{j,p-1}^c(g) = \lambda_{j+1,p}\left(\int_{\xi_j}^\xi g(s)\,ds\right) - \lambda_{j,p}\left(\int_{\xi_j}^\xi g(s)\,ds\right)$$
$$\Pi_{p-1,\Xi'}^c\frac{df}{d\xi} = \frac{d}{d\xi}\Pi_{p,\Xi}f$$

The new projector generates a commutative diagram [85]:

$$\begin{array}{ccccccc} \mathbb{R} & \longrightarrow & H^1(0,1) & \xrightarrow{\frac{d}{d\xi}} & L^2(0,1) & \longrightarrow & 0 \\ & & \Pi_{p,\Xi}\downarrow & & \Pi_{p-1,\Xi'}^c\downarrow & & \\ \mathbb{R} & \longrightarrow & \mathbb{S}^p(\Xi) & \xrightarrow{\frac{d}{d\xi}} & \mathbb{S}^p(\Xi') & \longrightarrow & 0 \end{array} \tag{55}$$

Special estimates

Now, we will prove some special multi-dimensional inequalities for spline quasi-interpolation. These inequalities were presented in [88] using weighted averages and interpolation at the nodes. They can be transferred to IGA, but the procedure for proving these results under the new framework becomes slightly different.

Theorem 9. *Let $K \in \mathcal{M}$ be an arbitrary element, $\gamma \in \mathcal{E}_K$ is an arbitrary face, and Π is a spline quasi-interpolation operator of degree r on the corresponding local tensor mesh. Then, for all Lebesgue exponents p and functions $u \in W^{1,p}$, we have the following local L^p-estimates:*

$$\|u - \Pi u\|_{L^p(K)} \leq C_1 \|v\|_{L^p(\widetilde{\omega}_K)} \tag{56a}$$

$$\|u - \Pi u\|_{L^p(K)} \leq C_2 h_K \|\nabla v\|_{L^p(\widetilde{\omega}_K)} \tag{56b}$$

$$\|\nabla(u - \Pi u)\|_{L^p(K)} \leq C_3 \|\nabla v\|_{L^p(\widetilde{\omega}_K)} \tag{56c}$$

$$\|u - \Pi u\|_{L^p(\gamma)} \leq C_4 h_\gamma^{1-1/p} \|\nabla v\|_{L^p(\widetilde{\omega}_\gamma)} \tag{56d}$$

Proof. The proof has four stages:

Stage 1+2. Let $v \in L^q(K)\backslash\{0\}$, where q is the conjugate exponent of p. Since B-splines form a partition of unity, we get

$$\begin{aligned}\int_K (u - \Pi u) v \, d\Omega &= \int_K \left[\left(\sum_{j \in \mathcal{N}_K} N_{j,r} \right) u - \sum_{j \in \mathcal{N}_K} \lambda_{j,r} N_{j,r} \right] v \, d\Omega \\ &= \sum_{j \in \mathcal{N}_K} \int_K N_{j,r} (u - \lambda_{j,r}) v \, d\Omega\end{aligned}$$

where $\lambda_{j,r}$ is the coefficient expressed in Definition 7. By combining Hölder's inequalities for integrals and sums simultaneously, we get

$$\begin{aligned}&\sum_{j \in \mathcal{N}_K} \int_K N_{j,r} (u - \lambda_{j,r}) v \, d\Omega \\ &= \sum_{j \in \mathcal{N}_K} \int_K N_{j,r}^{1/p} (u - \lambda_{j,r}) N_{j,r}^{1/q} v \, d\Omega \\ &\leq \sum_{j \in \mathcal{N}_K} \|N_{j,r}^{1/p} (u - \lambda_{j,r})\|_{L^p(K)} \|N_{j,r}^{1/q} v\|_{L^q(K)} \\ &\leq \left[\sum_{j \in \mathcal{N}_K} \|N_{j,r}^{1/p} (u - \lambda_{j,r})\|_{L^p(K)}^p \right]^{\frac{1}{p}} \left[\sum_{j \in \mathcal{N}_K} \|N_{j,r}^{1/q} v\|_{L^q(K)}^q \right]^{\frac{1}{q}}\end{aligned}$$

Applying positivity and partition of unity for B-splines, we get

$$\sum_{j\in\mathcal{N}_K} \|N_{j,r}^{1/q} v\|_{L^q(K)}^q = \int_K \underbrace{\left[\sum_{j\in\mathcal{N}_K} N_{j,r}\right]}_{\leq 1} |v|^q \, d\Omega \leq \|v\|_{L^q(K)}^q$$

Since $v \in L^q(K)\backslash\{0\}$ is arbitrary, we can use the general norm definition:

$$\begin{aligned}\|u - \Pi u\|_{L^p(K)} &= \sup_v \frac{\int_K (u - \Pi u) v \, d\Omega}{\|v\|_{L^q(K)}} \\ &\leq \left[\sum_{j\in\mathcal{N}_K} \|N_{j,r}^{1/p}(u - \lambda_{j,r})\|_{L^p(K)}^p\right]^{\frac{1}{p}}\end{aligned}$$

By the help of best polynomial approximation in the L^p-norm, we can denote the coefficient $\lambda_{j,r}$ as the best spline quasi-interpolation of u by constants. Since ω_j has greater measure that K, as shown in Figure 3 and 4, we get

$$\begin{aligned}\|N_{j,r}^{1/p}(u - \lambda_{j,r})\|_{L^p(K)} &= \inf_{c\in\mathbb{R}} \|N_{j,r}^{1/p}(u - c)\|_{L^p(K)} \\ &\leq C_K \inf_{c\in\mathbb{R}} \|N_{j,r}^{1/p}(u - c)\|_{L^p(\omega_j)} \\ &\leq C_K \|N_{j,r}^{1/p} u\|_{L^p(\omega_j)}\end{aligned}$$

where $C_K > 0$ is depending on K. The generalized version of Poincaré's inequality in L^p [50] implies that

$$\begin{aligned}\|N_{j,r}^{1/p} u\|_{L^p(\omega_j)} &\leq C_P(\omega_j) h_j \|N_{j,r}^{1/p} \nabla u\|_{L^p(\omega_j)} \\ &= \left(C_P(\omega_j)\frac{h_j}{h_K}\right) h_K \|N_{j,r}^{1/p} \nabla u\|_{L^p(\omega_j)}\end{aligned}$$

The partition of unity implies that

$$\left[\sum_{j\in\mathcal{N}_K} \|N_{j,r}^{1/p} u\|_{L^p(\omega_j)}^p\right]^{\frac{1}{p}} \leq \|u\|_{L^p(\widetilde{\omega}_j)} \quad , \quad \left[\sum_{j\in\mathcal{N}_K} \|N_{j,r}^{1/p} \nabla u\|_{L^p(\omega_j)}^p\right]^{\frac{1}{p}} \leq \|\nabla u\|_{L^p(\widetilde{\omega}_j)}$$

Thus, we have proven inequality (18a) and (18b). It should be noted that

$$C_2 = C_1 \max_{j\in\mathcal{N}_K} C_P(\omega_j)\frac{h_j}{h_K}$$

Stage 3. Let $\mathbf{v} \in L^q(K)\backslash\{0\}^d$ be a vector field. The product rule yields

$$\begin{aligned}\int_K \nabla(u-\Pi u)\cdot\mathbf{v}\,d\Omega &= \sum_{j\in\mathcal{N}_K}\int_K [\nabla N_{j,r}(u-\lambda_{j,r})]\cdot\mathbf{v}\,d\Omega \\ &= \sum_{j\in\mathcal{N}_K}\left[\int_K N_{j,r}\nabla u\cdot\mathbf{v}\,d\Omega + \int_K (u-\lambda_{j,r})\nabla N_{j,r}\cdot\mathbf{v}\,d\Omega\right]\end{aligned}$$

By combining Hölder's inequalities for integrals and sums with the partition of unity for B-splines, we get an inequality for the first sum:

$$\begin{aligned}\sum_{j\in\mathcal{N}_K}\int_K N_{j,r}\nabla u\cdot\mathbf{v} &\le \left[\sum_{j\in\mathcal{N}_K}\|N_{j,r}^{1/p}\nabla u\|_{L^p(K)}^p\right]^{\frac{1}{p}}\left[\sum_{j\in\mathcal{N}_K}\|N_{j,r}^{1/q}\mathbf{v}\|_{L^q(K)}^q\right]^{\frac{1}{q}} \\ &\le \|\nabla u\|_{L^p(K)}\|\mathbf{v}\|_{L^q(K)}\end{aligned}$$

Similarly, the other sum satisfies

$$\sum_{j\in\mathcal{N}_K}\int_K (u-\lambda_{j,r})\nabla N_{j,r}\cdot\mathbf{v} \le \sum_{j\in\mathcal{N}_K}\|u-\lambda_{j,r}\|_{L^p(K)}\|\nabla N_{j,r}\|_{L^\infty(K)}\|\mathbf{v}\|_{L^q(K)}$$

Since $\mathbf{v}$ is arbitrary, the general definition of norms implies that

$$\|\nabla(u-\Pi u)\|_{L^p(K)} \le \|\nabla u\|_{L^p(K)} + \sum_{j\in\mathcal{N}_K}\|u-\lambda_{j,r}\|_{L^p(K)}\|\nabla N_{j,r}\|_{L^\infty(K)}$$

Poincaré's inequality yields $\|u-\lambda_{j,r}\|_{L^p(K)} \le C_P(\omega_j)\|\nabla u\|_{L^p(\widetilde{\omega}_j)}$, and since the term $\|\nabla N_{j,r}\|_{L^\infty(K)}$ depends just on K, we obtain (18c) with

$$C_3 = 1 + \sum_{j\in\mathcal{N}_K}\|\nabla N_{j,r}\|_{L^\infty(K)}$$

Stage 4. In the same way as before, we have the inequality

$$\|u-\Pi u\|_{L^p(\gamma)} \le \left[\sum_{j\in\mathcal{N}_\gamma}\|N_{j,r}^{1/p}(u-\lambda_{j,r})\|_{L^p(\gamma)}^p\right]^{\frac{1}{p}}$$

The difference is that we changed the domain from an arbitrary element K to one of its corresponding edges γ.

From [88], there is a trace inequality in the L^p which states that

$$\|N_{j,r}^{\frac{1}{p}}(u-\lambda_{j,r})\|_{L^p(\gamma)}^p \leq c_1\|N_{j,r}^{\frac{1}{p}}(u-\lambda_{j,r})\|_{L^p(\omega\gamma)}^p + c_2\|N_{j,r}^{\frac{1}{p}}(u-\lambda_{j,r})\|_{L^p(\omega\gamma)}^{p-1}\|N_{j,r}^{\frac{1}{p}}u\|_{L^p(\omega\gamma)}$$

The final approach is just applying the other estimates we derived previously in Stage 1 and 2. On the two terms, we get

$$\|N_{j,r}^{\frac{1}{p}}(u-\lambda_{j,r})\|_{L^p(\omega_\gamma)}^p \leq \left(\frac{C_\gamma^p h_z^p}{h_\gamma^{p-1}}\right) h_\gamma^{p-1}\|\nabla u\|_{L^p(\widetilde{\omega}_\gamma)}^p$$

$$\|N_{j,r}^{\frac{1}{p}}(u-\lambda_{j,r})\|_{L^p(\omega_\gamma)}^{p-1}\|N_{j,r}^{\frac{1}{p}}u\|_{L^p(\omega_\gamma)} \leq \left(\frac{C_\gamma h_z}{h_\gamma}\right)^{p-1} h_\gamma^{p-1}\|\nabla u\|_{L^p(\widetilde{\omega}_\gamma)}^p$$

We can finally define the generic constant C_4 as

$$C_4 = \max\left\{c_1\left(\frac{C_\gamma^p h_z^p}{h_\gamma^{p-1}}\right), c_2\left(\frac{C_\gamma h_z}{h_\gamma}\right)^{p-1}\right\}^{\frac{1}{p}}$$

Thus, we have proved inequality (18d). □

4 A posteriori error estimation

In this section, we present the main theory of a posteriori estimation, which is relevant for the derivation of our error estimators discussed later. This includes the formal properties of estimators, optimal control interpretation of adaptive refinement, and the important concept of pollution error.

4.1 Main characteristics

The common drawbacks of a priori estimation demonstrate that *a posteriori error estimation* can be more advantageous. This approach requires solving the PDE on a coarse mesh and refine the elements where the estimated error is too high, and we repeat the process until the global estimated error is low enough. Local refinement is easy because we post-process the final computed solution instead of assuming its properties on forehand. There are many different a posteriori estimators available, and they share some common properties [88]. First, the computation of the local error should be done as fast as possible. Second, we require two bounds:

1. *Upper bound* constraining the global error within a given tolerance.
2. *Lower bound* ensuring that all local parts of the domain are refined correctly.

This time, stability is measured in terms of multiplicative factors whose size reflects the computational effort. Small size means less sensitivity to local perturbations, and we call the stability *strong* because the norm involves derivatives. Thus, we can formulate the general paradigm for a posteriori error estimation as follows:

$$\text{Small residual} + \text{Stability of continuous problem} \implies \text{Small error}$$

In classical FEM, we can construct a conformal mesh on an arbitrary domain by subdividing it into triangles or quadrilaterals. Sometimes, we can create hybrid meshes consisting of both types. When we refine the elements, it is common to subdivide it into four smaller elements. In this way, the new refined elements will almost resemble the original element with respect to the shape, as shown in Figure 11. Triangulation for any continuity is not fully available in IGA because the spline triangles are not compatible enough with the discretization process. It is possible to discretize a PDE with isogeometric C^0-triangles, but this is not of significant interest, for the numerical results become almost similar to those ones obtained with classical FEM. Therefore, we will always restrict ourselves to non-hybrid meshes with convex quadrilaterals.

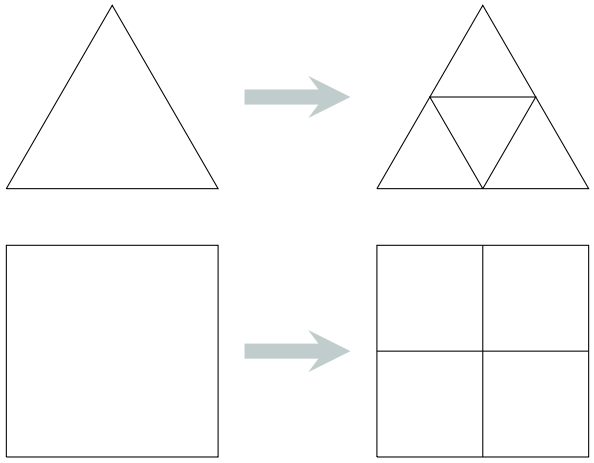


Figure 11. Refinement of a triangle and a quadrilateral.

The global norm error does not provide useful bounds for the errors in the target quantities of the real physical interest, and their sensitivity to local error sources cannot be represented appropriately enough by global stability constants. In a priori error estimation, it is possible to overcome such typical deficiencies by employing several duality techniques. We replace global stability constants by computationally obtained local sensitivity factors, and then we combine them with Galerkin orthogonality to derive appropriate a posteriori error estimators. In a posteriori estimates, local residuals of the computed solution are multiplied with certain weights measuring their error dependence, which can be controlled by local adaptive coarsening and refining. These weights are obtained as the approximate solution of an adjoint linear problem related to the original physical model. The resulting adaptive meshes are economic with respect to computational effort.

The approximation quality depends on the mesh and characteristic properties of the shape functions used for discretization. Controlling error requires correct determination of existing influence factors affecting the local error indicator on the target quantity we want to model properly. This type of sensitivity analysis of local perturbations in the error motivates the usage of adjoint operators. For any a posterior estimator, we must detect interplay of many effects caused by error propagation to achieve suitable error control and solution-adapted meshing. The use of duality arguments and adjoint operators in the derivation of a posteriori estimators was originally proposed by Babuška and Miller [5, 6, 7]. This has been studied for more general situations by Eriksson et al. [47], and by Giles and Süli [56]. We start with the formal definitions of efficiency, reliability, and asymptotic exactness of error estimators, which will be used throughout this paper.

Definition 8 (Efficiency of error estimator [2]). *If $\mathcal{M}$ is a finite element partition on a domain Ω, then the global error estimator η is the l^2-norm of the local errors η_K on each individual element K:*

$$\eta = \sqrt{\sum_{K \in \mathcal{M}} \eta_K^2} \tag{57}$$

From this definition, we define the local and global effectivity indices as

$$\theta_K = \frac{\eta_K}{|||e|||_{(K)}} \quad , \quad \theta = \frac{\eta}{|||e|||} \tag{58}$$

We call the error estimator asymptotically exact if the following limit holds:

$$\lim_{h \to 0} \theta = 1 \tag{59}$$

If the error estimator is robust, then there are constants C and D such that

$$C\eta \leq |||e||| \leq D\eta \tag{60}$$

The lower bound provides the efficiency, and the upper bound ensures that the estimator is reliable and indicates the elements to be refined correctly.

4.2 Optimal control interpretation

Most a posteriori estimates are on the standard form

$$|||u - u_h||| \leq C|||\rho(u_h)|||_* \tag{61}$$

where $\rho(u_h) = f - Au_h$ is a computable residual, and E^* is the energy space's dual. Although the energy norm is generic and applicable for any PDE, it does not always provide useful bounds on the target quantities' errors. To be more versatile, we can analyse the error measures by duality principles. Let u be the PDE's solution, and $J(u)$ is a physical quantity derived from it. We want to control the functional error $J(u) - J(u_h)$ in terms of local computable residuals $\rho_K(u_h)$ on each element K. Assuming that the PDE is linear, the error equation becomes $A(u - u_h) = \rho(u_h)$. The effect of the cell residual ρ_K on the error $e_{K'}$ of another cell K', a complex interaction, cannot be determined analytically in general, only detected by computation. This gives a "weighted" a posteriori estimate:

$$|J(u) - J(u_h)| \approx \langle \rho(u_h), \omega_h(z) \rangle \tag{62}$$

The sensitivity factor ω_h describes the effect of local variations in $\rho(u_h)$. It is governed by solving the adjoint problem $A^*z = j$ approximately, where j is a density function associated with J. This approach is known as the *dual-weighted residual method* (DWR). The goal is to minimize $J(u) - J(u_h)$, and this is indeed a constrained optimization problem:

$$\min_{u \in V} J(u) \quad , \quad \text{subject to} \quad \begin{cases} A(u, \varphi) = F(\varphi) & \varphi \in V \\ A(\varphi, z) = J(\varphi) & z \in V \end{cases} \tag{63}$$

where z is an adjoint variable. The corresponding Lagrangian is

$$L(u, z) = J(u) + F(z) - A(u, z)$$

The identity $J(u) = F(z) = A(u, z)$ makes u and z mutually disjoint and dual. We define the error and its dual respectively as $e = x - x_h$ and $e^* = z - z_h$. Then

$$J(e) = A(e, z) = A(e, e^*) = A(u, e^*) = F(e^*)$$

The corresponding residuals are given by

$$\rho(u_h, \cdot) = F(\cdot) - A(u_h, \cdot)$$
$$\rho^*(z_h, \cdot) = J(\cdot) - A(\cdot, z_h)$$

From Galerkin orthogonality, we see that

$$\rho(u_h, z - \varphi_h) = A(e, e^*) = \rho^*(z_h, u - \varphi_h)$$

Thus, we obtain the final identity

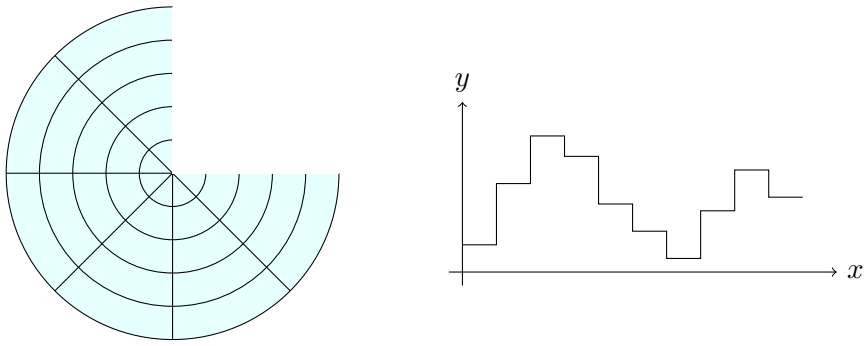
$$\begin{aligned} J(e) &= \min_{\varphi_h \in V_h} \rho(u_h, z - \varphi_h) \\ &= \min_{\varphi_h \in V_h} \rho^*(z_h, u - \varphi_h) \\ &= F(e^*) \end{aligned}$$

This connection between a posteriori error estimation and optimal control was demonstrated by Becker and Rannacher [29], and it provides many general results which can be adapted to many PDEs. Furthermore, the approach holds both for Bubnov-Galerkin and Petrov-Galerkin formulations, even if the PDE is semilinear.

4.3 The effect of pollution error

The a priori estimates discussed until now require the unknown solution to be sufficiently smooth, e.g. it belongs to $H^m(\Omega)$. Sometimes, the solution of a PDE has a very sharp gradient in certain regions of the domain Ω, but the solution itself is sufficiently smooth and does not contain discontinuities. When we solve this PDE with adaptive FEM, the mesh is coarse at the beginning, but in the regions with sharp gradient, the error here will not be so influent on the rest of the solution. After some refinements, the global error is low enough.

In many practical situations, the solution is not smooth enough, and there are even situations where the error can have a very deteriorating effect. There exist certain cases where the gradient of the solution is singular, like domains with re-entrant corners or jump discontinuities in the boundary conditions. The approximation error here is classified as *pollution error* because it affects the rest of the whole global solution very much when it is too high, and the results become quite poor. Simplification of the physical model's data is a common reason for this defect, for we do not know on forehand how the real solution will behave. Thus, the asymptotic behaviour of the solution must be described in terms of *intensity factors* describing the strength of the singularities. We use also a local averaging scheme for extracting a post-processed gradient value, which is more accurate than the *raw-gradient* originating from the finite element solution [20].



(a) A circle with re-entrant corner

(b) Discontinuous jumps in the solution

Figure 12. Examples on sources causing pollution error.

According to Babuška [13], pollution error is characterized as follows:

- The error e_K of an element K can be split up in two parts: *local error* (caused in element K and the neighbouring elements, i.e. the patch $\widetilde{\omega}_K$) and *pollution error* (originating in the rest of the global mesh, especially in neighbourhoods of singularities).
- Error indicators based only on local computation neglect pollution error.
- The pollution error is the most influent factor on the global error and can only be controlled properly by global adaptive refinement.
- We can the estimate pollution error in small patches with a global extraction consisting of two parts: the finite element approximation of an auxiliary function constructed appropriately, and the standard error indicators. To make the process effective, we use a direct solver with resolution capability on the linear equation system arising from the discretization.
- On a uniform mesh, the global energy norm equals the energy norm over the patches of elements with vertices at the singularity, so the global effectivity index reflects just the accuracy at these elements. The pollution error here is negligible because the error indicators almost equal the exact error's norm.
- Pollution error can be significant almost everywhere.

These characteristics demonstrate that if we want optimal error reduction on elements, the error estimator must be composed of distinct components estimating the local error and pollution error separately from each other. Babuška has also proved in [14] that when we suppress the pollution error correctly, then the element effectivity indexes depend on the local $(p+1)$-Taylor expansion of the exact solution. A good choice is analysing element error indicators through the value of effectivity indexes corresponding to worst Taylor expansions. From [14, 21, 18], the following conclusions hold:

1. Since the error of element patches has a local and global component, and local estimators neglect the global part, we cannot describe the effectivity index on an element without verifying that the pollution error is very small compared with the local error here.
2. Local error coincides with the error of finite element approximation of a local $(p+1)$-Taylor expansion of the exact solution.
3. Local geometry of the mesh determines effectivity indices of elements.

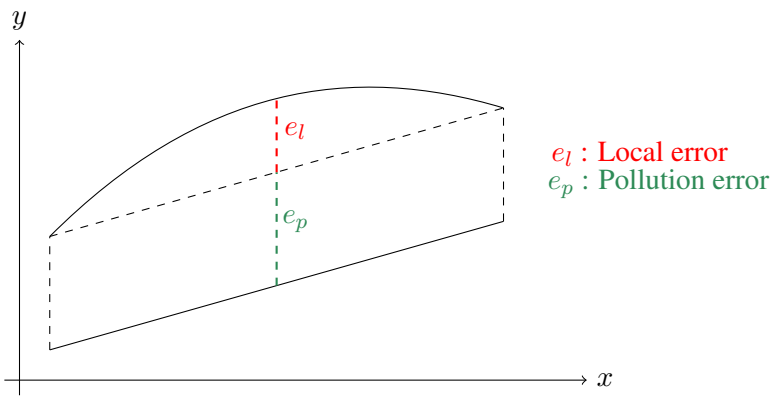


Figure 13. Decomposing the total approximation error in two parts.

With the same notation from [57], we define the local and pollution error as

$$e_h = \underbrace{(u - \Pi_h u)}_{\substack{e_{H,\,\text{loc}} \\ \text{Locally produced} \\ \text{truncation error}}} + \underbrace{(\Pi_h u - u_h)}_{\substack{e_{H,\,\text{glob}} \\ \text{Globally transported} \\ \text{pollution error}}} \tag{64}$$

where Π_h is the operator for the finite element projection. If there are no sources of pollution error at all, then the last term is negligible. According to Oden [74], we can assume that the subdomain Ω_0 has a singularity. The local error will depend on the local properties of Π_h and the regularity of the exact solution near Ω_0, while the pollution error is determined by the solution's regularity outside Ω_0, the boundary operator, and the boundary's regularity. It is also more flexible to analyse the pollution error on each individual element.

Furthermore, if we denote the original local error estimator as η_K^h, we can extend it with the estimator $\eta_K^{h,pol}$ for the pollution error. To make this new extended estimator work properly, we need a bound on the form

$$\begin{aligned} &C|||\eta_K^h|||^2_{\Omega_K,\,\text{ext}} \leq |||u_h - u_H|||^2_{\Omega_K} \leq D|||\eta_K^h|||^2_{\Omega_K,\,\text{ext}} \\ &|||\eta_K^h|||^2_{\Omega_K,\,\text{ext}} = |||\eta_K^h|||^2_{\Omega_K} + |||\eta_K^{h,\,\text{pol}}|||^2_{\Omega_K} \end{aligned} \tag{65}$$

where u_h and u_H represents the finite element approximation on a fine and a coarse mesh, respectively. The *Element Residual Method* (ERM), described by Oden et al. in [72], is a common a posteriori error estimator. It has been shown in [74] that this method can be extended to the *Equivalent Pollution Residual Method*, enabling us to estimate the pollution error in each element.

According to Babuška [16], there are two types of pollution error. The first one is *A-pollution* (Approximation pollution) and is usually caused by non-smoothness of input data, like the load function or initial conditions. In this case, the non-smoothness of the solution in one specific area affects the accuracy in another different area. The second type is *B-pollution* (Boundary pollution) and is common in several dimensions. It occurs if the domain or its boundary is not smooth, or if the boundary conditions change in a sudden non-smooth way.

Until now, we have assumed that our approach is *conforming*, which means that $V_h \subset V$. A problem with numerical integration is conformity violation triggered by the lack of enough quadrature points in the assembly process. This is a common variational crime and main source of both A- and B-pollution. In regions where the solution lacks smoothness, the effect of pollution becomes stronger when the quadrature scheme is not precise enough. This has been verified with the Bramble-Hilbert lemma [90].

4.4 Methodology for comparing quality

Although local estimators work better than a priori estimators, they cannot take pollution error into account, which is most significant in most of the elements. In small patches, we can estimate the pollution error by *global extraction*, which is using a finite element approximation of an appropriately constructed function and a standard element indicator. If a direct solver with resolution capability is employed, the running time is bounded. We can also create separate estimates for each error component. It is meaningless to report effectivity without confirming that pollution is negligible, as Babuška et al. mentioned in [13].

Let $S(\overline{x}, H)$ be a subdomain of size H centred at $\overline{x}$. We can impose some criterions on the analytical solution and the computational mesh:

- Locally uniform mesh for analysis, $C_1 h^\gamma \leq H \leq C_2 h^\gamma$, $\gamma \in (0, 1)$.
- Convergence in L^2, $\|e_h\|_{L^2(S(\overline{x},H))} \leq C h^{p+1-\epsilon} H$
- Sufficient smoothness of the exact solution u

$$\max_{\substack{0\leq i,j\leq p+2 \\ i+j=p+2}} \left\| \frac{\partial^{p+2} u}{\partial x_1^i \partial x_2^j} \right\|_{L^\infty(S(\overline{x},H))} \leq K < \infty$$

$$\sum_{\substack{0\leq i,j\leq p+2 \\ i+j=p+2}} \left| \frac{\partial^{p+2} u}{\partial x_1^i \partial x_2^j} \right| (\overline{x}) \geq C_0 > 0$$

5 Residual-based estimators

Before we derive some of the common residual-based estimators used in adaptive refinement, which rely on complementary variational methods, we will first give a concise overview on the relevant theory of this topic. Much of this theory was originally developed in the context of classical finite element modelling, but it will also work for splines since they are piecewise polynomials with high continuity.

5.1 Preliminaries

To derive the explicit a posteriori estimators, we need an important theorem:

Theorem 10 (Bernardi and Girault [30])**.** *Let $p \in [1, \infty)$, $k \in \mathbb{Z}^+$, $s \in [0, 1]$, and $t \in [s, k]$. Let Ω be a polygon domain with a regular finite element partition $\mathcal{M}$ (triangles or quadrilaterals) such that V_h is a finite element subspace. Then there exists a bounded linear operator $\Pi_h : W^{t,p} \longmapsto V_h$ and constants C_1 and C_2 such that for all $u \in W^{t,p}$ and $K \in \mathcal{M}$, we have*

$$|(I - \Pi_h)u|_{W^{s,p}(K)} \leq C_1 h_K^{t-s} |u|_{W^{t,p}(\widetilde{K})} \tag{67}$$

$$|(I - \Pi_h)u|_{W^{s,p}(\gamma)} \leq C_2 h_\gamma^{t-s-1/p} \|u\|_{W^{t,p}(\widetilde{K})} \tag{68}$$

where K is an arbitrary element and γ is any edge of it. The constants C_1 and C_2 depend on the regularity constant κ, not the element diameter h_K.

We also need an inequality based on the general Hölder inequality for sums:

Lemma 5. *Let $\{a_i\}_{i=1}^{\infty}$, $\{b_i\}_{i=1}^{\infty}$, $\{c_j\}_{j=1}^{\infty}$ and $\{d_j\}_{j=1}^{\infty}$ are l^1-summable sequences, and p and q are conjugate exponents. Then we have*

$$\sum_{i=1}^{m} a_i b_i + \sum_{j=1}^{n} c_j d_j \leq 2 \left[\sum_{i=1}^{m} a_i^p + \sum_{j=1}^{n} c_j^p \right]^{\frac{1}{p}} \left[\sum_{i=1}^{m} b_i^q + \sum_{j=1}^{n} d_j^q \right]^{\frac{1}{q}} \tag{69}$$

We introduce cut-off functions [58], which play a vital role in the proofs for efficiency. In general, a cut-off function $\psi \in C_0^{\infty}$ on an open set A satisfies

$$\psi(\mathbf{x}) = \begin{cases} 1 & , \ \mathbf{x} \in A' \\ \in [0, 1] & , \ \mathbf{x} \in A \backslash A' \\ 0 & , \ \mathbf{x} \in \mathbb{R}^n \backslash A \end{cases}$$

where $A' \subsetneq A$ is nonempty. Then, we use an important proposition from [88], which generalizes corresponding lemmas and theorems from [2]:

Proposition 4. *For all elements K and their faces γ, we have the following bounds for any $v \in R_s(K)$ and $w \in R_s(\gamma)$:*

$$\|v\|_{L^p(K)} \leq C_1 \|\psi_K^{1/p} v\|_{L^p(K)} \tag{70a}$$
$$\|\nabla(\psi_K v)\|_{L^p(K)} \leq C_2 h_K^{-1} \|v\|_{L^p(K)} \tag{70b}$$
$$\|w\|_{L^p(\gamma)} \leq C_3 \|\psi_\gamma^{1/p} w\|_{L^p(\gamma)} \tag{70c}$$
$$\|\nabla(\psi_\gamma w)\|_{L^p(\omega_\gamma)} \leq C_4 h_\gamma^{1/p-1} \|w\|_{L^p(\gamma)} \tag{70d}$$
$$\|\psi_\gamma w\|_{L^p(\omega_\gamma)} \leq C_5 h_\gamma^{1/p} \|w\|_{L^p(\gamma)} \tag{70e}$$

Proof. These estimates are derived quite similarly. First, we transform K and γ respectively to $\widehat{K}$ and $\widehat{\gamma}$ by using $\mathcal{F} : \Omega \longmapsto \widehat{\Omega}$. Next, we use the general theorem stating that all norms are equivalent on a finite-dimensional Banach space [65]. This is valid because the numerical solution is a finite linear combination of splines. Since $\widehat{\psi} > 0$ on $\text{int}(\widehat{K})$, we can introduce two complete norms:

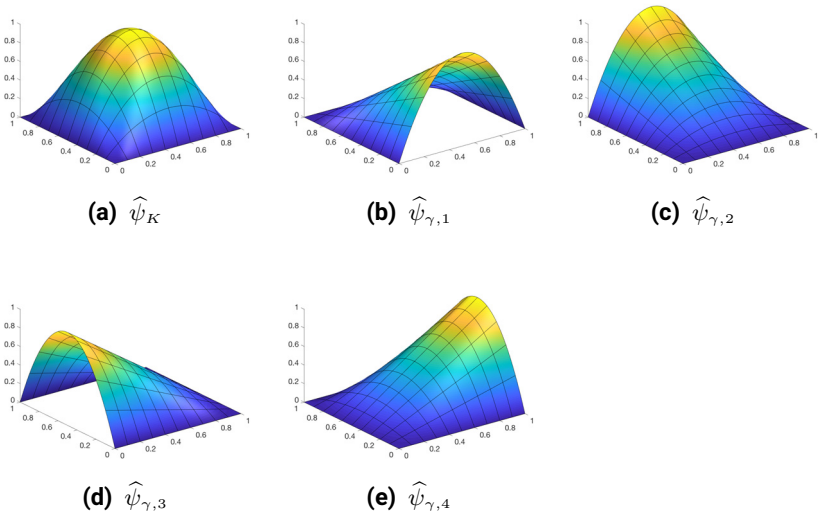
$$\widehat{v} \longmapsto \|\widehat{\psi}^{1/p}\widehat{v}\|_{L^p(\widehat{K})} \qquad , \qquad \widehat{v} \longmapsto \|\widehat{\psi}\widehat{v}\|_{W^{1,p}(\widehat{K})}$$

We transform $\widehat{K}$ and $\widehat{\gamma}$ respectively to K and γ by $\mathcal{F}^{-1} : \widehat{\Omega} \longmapsto \Omega$. □

We notice that R_s is the reference square (12), and we have two cut-off functions ψ_K and ψ_γ respectively on the element interior and element edge. The optimal constants $\{C_i\}_{i=1}^5$ depend only on s, p, h_K and ψ. This means in practice that the estimates are determined by the discretization, not the model problem. A detailed proof for the optimal constants is described in [88]. In our case, we expect that they will be small because we are using splines with high continuity.

The later efficiency proofs require another technical facility. Verfürth [87, 88] has constructed five bubble functions (cut-off functions) on the reference quadrilateral $\widehat{K} = [-1, 1]^2$ for localizing the residual. We will generalize this approach by focusing on the reference elements $[0, 1]^2$ and $[0, 1]^3$, since splines are used for discretization. On $[0, 1]^2$, we get

$$\widehat{\psi}_{\gamma,1} = 4x(1-x)(1-y) \qquad \widehat{\psi}_{\gamma,2} = 4x(1-x)y$$
$$\widehat{\psi}_{\gamma,3} = 4(1-x)y(1-y) \qquad \widehat{\psi}_{\gamma,4} = 4xy(1-y)$$
$$\widehat{\psi}_K = 16x(1-x)y(1-y)$$



(a) $\widehat{\psi}_K$ **(b)** $\widehat{\psi}_{\gamma,1}$ **(c)** $\widehat{\psi}_{\gamma,2}$

(d) $\widehat{\psi}_{\gamma,3}$ **(e)** $\widehat{\psi}_{\gamma,4}$

Figure 14. The cut-off functions $\widehat{\psi}_K$, $\widehat{\psi}_{\gamma,1}$, $\widehat{\psi}_{\gamma,2}$, $\widehat{\psi}_{\gamma,3}$ and $\widehat{\psi}_{\gamma,4}$ on $\widehat{K} = [0,1]^2$.

On $[0,1]^3$, we get the similar analogous functions

$$\begin{aligned}
\widehat{\psi}_{\gamma,1} &= 16x(1-x)(1-y)z(1-z) & \widehat{\psi}_{\gamma,2} &= 16x(1-x)yz(1-z)\\
\widehat{\psi}_{\gamma,3} &= 16(1-x)y(1-y)z(1-z) & \widehat{\psi}_{\gamma,4} &= 16xy(1-y)z(1-z)\\
\widehat{\psi}_{\gamma,5} &= 16x(1-x)y(1-y)(1-z) & \widehat{\psi}_{\gamma,6} &= 16x(1-x)y(1-y)z\\
\widehat{\psi}_K &= 64x(1-x)y(1-y)z(1-z)
\end{aligned}$$

The cut-off function ψ_K can be expressed in terms of tensor B-splines and belong to $W^{1,\infty}(K)$. To make them even more smooth, we define $\psi_{K,m} = (\psi_K)^{m+1}$, which belongs to $C^m(K)$ and hence $W^{m+1,\infty}(K)$.

Now, we will finally present an important auxiliary lemma used later:

Lemma 6. *If $\{x_i\}_{i=1}^m$ are strictly positive real numbers, and $p \geq 1$, then*

$$\left(\sum_{i=1}^m x_i\right)^p \leq 2^{p-1}\left(\sum_{i=1}^m x_i^p\right) \tag{73}$$

5.2 Standard explicit estimator

We start with the standard explicit residual estimator for Poisson's equation, and generalize the procedures described in [2, 88].

Derivation of the estimator

We consider the BVP from (22), restrict ourselves to the Poisson problem, and define $\mathcal{M}$ as the partition on Ω. Since the error $e = u - u_h$ satisfies Galerkin orthogonality (29), integration by parts allows us to express the sum of integrals as follows:

$$\begin{aligned}
a(e,v) &= a(u,v) - a(u_h,v) \\
&= \sum_{K\in\mathcal{M}} \left(a_K(u,v) - a_K(u_h,v)\right) \\
&= \sum_{K\in\mathcal{M}} \left\{ \int_K fv \, d\Omega + \int_{\partial K_N} g_N v \, ds - \int_K \nabla u_h \cdot \nabla v \, d\Omega \right\} \\
&= \sum_{K\in\mathcal{M}} \left\{ \int_K fv \, d\Omega + \int_{\partial K_N} g_N v \, ds + \int_K v\nabla^2 u_h \, d\Omega - \int_{\partial K} \frac{\partial u_h}{\partial n_K} v \, ds \right\} \\
&= \sum_{K\in\mathcal{M}} \int_K rv \, d\Omega + \sum_{\gamma\in\Sigma} \int_\gamma jv \, ds \\
&= \sum_{K\in\mathcal{M}} \int_K r(v - v_h) \, d\Omega + \sum_{\gamma\in\Sigma} \int_\gamma j(v - v_h) \, ds
\end{aligned}$$

where $\partial K_N = \partial K \cap \partial\Omega_N$. This result holds because $a(e, v_h) = 0$. In this setting, we denote the *interior* and *boundary residuals* respectively as $r : \Omega \longmapsto \mathbb{R}$ and $j : \Sigma \longmapsto \mathbb{R}$ [88]. The last term refers to the whole skeleton Σ of $\mathcal{M}$, and is not restricted just to the boundary $\partial\Omega$. If $K \in \mathcal{M}$, these two residuals are given by

$$r|_K = f + \nabla^2 u_h \qquad \mathbf{x} \in \text{int}(K) \tag{74a}$$

$$j|_K = \begin{cases} -\mathbb{J}_\gamma(\mathbf{n}_\gamma \cdot \nabla u_h), & \mathbf{x} \in \mathcal{E}_0 \\ g_N - \mathbf{n}_\gamma \cdot \nabla u_h, & \mathbf{x} \in \mathcal{E}_N \\ 0, & \mathbf{x} \in \mathcal{E}_D \end{cases} \tag{74b}$$

The set of all edges on $\mathcal{M}$ is partitioned as $\mathcal{E} = \mathcal{E}_0 \cup \mathcal{E}_D \cup \mathcal{E}_N$, such that $\mathcal{E}_0$, $\mathcal{E}_D$ and $\mathcal{E}_N$ represent the interior edges, Dirichlet edges and Neumann edges. The jump of u in the direction of $\mathbf{n}_\gamma$ measures the boundary discontinuity between K and another adjacent element K^*. It is given by

$$\mathbb{J}_\gamma(\mathbf{n}_\gamma \cdot \nabla u_h) = \lim_{t\to 0^+} \{u(\mathbf{x} - t\mathbf{n}_\gamma) - u(\mathbf{x} + t\mathbf{n}_\gamma)\} \tag{75}$$

By using inequality (67) and (68), we choose $t = 1$, $s = 0$ and $p = 2$. Then, we apply inequality (69) with $p = q = 2$ and continue the derivation:

$$\begin{aligned}
a(e,v) &\leq \sum_{K\in\mathcal{M}} \|r\|_{L^2(K)}\|(I-\Pi_h)v\|_{L^2(K)} + \sum_{\gamma\in\Sigma}\|j\|_{L^2(\gamma)}\|(I-\Pi_h)v\|_{L^2(\gamma)} \\
&\leq \sum_{K\in\mathcal{M}} C_1 h_K \|v\|_{H^1(K)}\|r\|_{L^2(K)} + \sum_{\gamma\in\Sigma} C_2 h_\gamma^{1/2}\|v\|_{H^1(\gamma)}\|j\|_{L^2(\gamma)} \\
&\leq C_1\left[\sum_{K\in\mathcal{M}}\|v\|^2_{H^1(K)}\right]^{1/2}\left[\sum_{K\in\mathcal{M}} h_K^2\|r\|^2_{L^2(K)}\right]^{1/2} \\
&\quad + C_2\left[\sum_{\gamma\in\Sigma}\|v\|^2_{H^1(\gamma)}\right]^{1/2}\left[\sum_{\gamma\in\Sigma} h_\gamma\|j\|^2_{L^2(\gamma)}\right]^{1/2} \\
&\leq 2C_3\|v\|_{H^1(\Omega)}\underbrace{\left\{\sum_{K\in\mathcal{M}} h_K^2\|r\|^2_{L^2(K)} + \sum_{\gamma\in\Sigma} h_\gamma\|j\|^2_{L^2(\gamma)}\right\}^{1/2}}_{W}
\end{aligned}$$

where $C_3 = \max\{C_1, C_2\}$. From Poincaré's inequality, we see that

$$\begin{aligned}
\|e\|^2_{H^1} &= \|e\|^2_{L^2} + |e|^2_{H^1} \\
&\leq C_4^2|e|^2_{H^1} + |e|^2_{H^1} \\
&= (C_4^2+1)|||e|||^2
\end{aligned}$$

Since $a(\cdot,\cdot)$ is coercive, we can replace v by e and obtain

$$\begin{aligned}
\alpha|||e|||^2 &\leq a(e,e) \\
&\leq 2C_3\|e\|_{H^1(\Omega)}W \\
&\leq 2C_3(C_4^2+1)^{1/2}|||e|||W \\
|||e||| &\leq \left(\frac{2C_3}{\alpha}(C_4^2+1)^{1/2}\right)W \\
|||e|||^2 &\leq C^2_{\text{RES}}\sum_{K\in\mathcal{M}}\left\{h_K^2\|r\|^2_{L^2(K)} + h_\gamma\|j\|^2_{L^2(\partial K)}\right\}
\end{aligned}$$

where $C^2_{\text{RES}} = (C_4^2+1)(2C_3/\alpha)^2$. Hence, the local estimator becomes

$$\boxed{\eta^2_{\text{RES},K} = h_K^2\|r\|^2_{L^2(K)} + h_\gamma\|j\|^2_{L^2(\partial K)}} \tag{76}$$

Robustness of the estimator

If we sum the error in the energy norm over every element on the domain, we get an inequality on the following form:

$$\begin{aligned}
|||e|||^2 &= \sum_{K\in\mathcal{M}} |||e|||_K^2 \\
&\leq \sum_{K\in\mathcal{M}} C_{\text{RES},K}^2 \eta_{\text{RES},K}^2 \\
&\leq \left(\max_{K\in\mathcal{M}} C_{\text{RES},K}^2\right) \eta_{\text{RES}}^2
\end{aligned}$$

This bound shows that reliability is satisfied. If ψ_K is the cut-off function on K, let $w_K = (f_K + \nabla^2 u_h)\psi_K$. Our first derivation is combining inequality (70a) and (70b) with the L^2-representation (74a), Hölder's L^p-inequality and $\psi_K \in [0,1]$. This yields

$$\begin{aligned}
C_1^{-2}\|f_K + \nabla^2 u_h\|_{L^2(K)}^2 &\leq \int_K (f_K + \nabla^2 u_h)^2 \psi_K \, d\Omega \\
&= \int_K (f + \nabla^2 u_h) w_K \, d\Omega \\
&= \int_K \nabla(u - u_h) \cdot \nabla w_K \, d\Omega \\
&\leq \|\nabla(u - u_h)\|_{L^2(K)} \|\nabla w_K\|_{L^2(K)} \\
&\leq \|\nabla(u - u_h)\|_{L^2(K)} \cdot C_2 h_K^{-1} \|f_K + \nabla^2 u_h\|_{L^2(K)}
\end{aligned}$$

Then, we divide all terms by $\|f_K + \nabla^2 u_h\|_{L^2(K)}$ and multiply with $C_1^2 h_K$:

$$h_K \|f_K + \nabla^2 u_h\|_{L^2(K)} \leq C_1^2 C_2 \|\nabla(u - u_h)\|_{L^2(K)} \tag{77}$$

This estimate holds for the interior of K, and we will use a similar procedure for $\mathcal{E}_0$ and $\mathcal{E}_N$. First, we consider $\mathcal{E}_0$ and define $w_\gamma = -\mathbb{J}_\gamma(\mathbf{n}_\gamma \cdot \nabla u_h)\psi_\gamma$. This time, we use inequality (70c):

$$\begin{aligned}
&C_3^{-2}\|\mathbb{J}_\gamma(\mathbf{n}_\gamma \cdot \nabla u_h)\|_{L^2(\gamma)}^2 \\
&\leq \int_{\omega_\gamma} \mathbb{J}_\gamma(\mathbf{n}_\gamma \cdot \nabla u_h)^2 \psi_\gamma \, ds \\
&= \int_{\omega_\gamma} \nabla(u - u_h) \cdot \nabla w_\gamma \, ds - \int_{\omega_\gamma} r w_\gamma \, ds \\
&= \int_{\omega_\gamma} \nabla(u - u_h) \cdot \nabla w_\gamma \, ds - \sum_{K\subset\omega_\gamma} \int_K (f_K + \nabla^2 u_h) w_\gamma \, d\Omega
\end{aligned}$$

Then, we combine (70d) and (70e) with Hölder's L^p-inequality:

$$\int_{\omega_\gamma} \nabla(u-u_h)\cdot\nabla w_\gamma \,d\Omega \leq \|\nabla(u-u_h)\|_{L^2(\omega_\gamma)} \cdot C_4 h_\gamma^{-1/2} \|\mathbb{J}_\gamma(\mathbf{n}_\gamma\cdot\nabla u_h)\|_{L^2(\gamma)}$$
$$\sum_{K\subset\omega_\gamma}\int_K (f_K+\nabla^2 u_h) w_\gamma \,d\Omega \leq \sum_{K\subset\omega_\gamma} \|f_K+\nabla^2 u_h\|_{L^2(K)} \cdot C_5 h_\gamma^{1/2} \|\mathbb{J}_\gamma(\mathbf{n}_\gamma\cdot\nabla u_h)\|_{L^2(\gamma)}$$

Collecting everything, dividing by the common factor $\|\mathbb{J}_\gamma(\mathbf{n}_\gamma\cdot\nabla u_h)\|_{L^2(\gamma)}$ and then multiplying all terms by $C_3^2 h_\gamma^{1/2}$ yields

$$\begin{aligned} &h_\gamma^{1/2}\|\mathbb{J}_\gamma(\mathbf{n}_\gamma\cdot\nabla u_h)\|_{L^2(\gamma)} \\ &\leq C_3^2 C_4 \|\nabla(u-u_h)\|_{L^2(\omega_\gamma)} + C_3^2 C_5 \sum_{K\subset\omega_\gamma} h_\gamma \|f_K+\nabla^2 u_h\|_{L^2(K)} \end{aligned} \tag{78}$$

Lastly, we define $w_\gamma = (g_N - \mathbf{n}_\gamma\cdot\nabla u_h)\psi_\gamma$ on $\mathcal{E}_N$ and use the procedure:

$$\begin{aligned} C_3^{-2}\|g_N-\mathbf{n}_\gamma\cdot\nabla u_h\|_{L^2(\gamma)}^2 &\leq \int_\gamma (g_N-\mathbf{n}_\gamma\cdot\nabla u_h)^2\psi_\gamma \,ds \\ &= \int_K \nabla(u-u_h)\cdot\nabla w_\gamma \,d\Omega - \int_K (f_K+\nabla^2 u_h) w_\gamma \,d\Omega \end{aligned}$$

By combining (70d) and (70e) with Hölder's L^p-inequality, we obtain

$$\int_{\omega_\gamma} \nabla(u-u_h)\cdot\nabla w_\gamma \,d\Omega \leq \|\nabla(u-u_h)\|_{L^2(\omega_\gamma)} \cdot C_4 h_\gamma^{-1/2} \|g_N-\mathbf{n}_\gamma\cdot\nabla u_h\|_{L^2(\gamma)}$$
$$\int_K (f_K+\nabla^2 u_h) w_\gamma \,d\Omega \leq \|f_K+\nabla^2 u_h\|_{L^2(K)} \cdot C_5 h_\gamma^{1/2} \|g_N-\mathbf{n}_\gamma\cdot\nabla u_h\|_{L^2(\gamma)}$$

Collecting everything, dividing by the common factor $\|g_N-\mathbf{n}_\gamma\cdot\nabla u_h\|_{L^2(\gamma)}$ and then multiplying all terms by $C_3^2 h_\gamma^{1/2}$ again yields

$$\begin{aligned} &h_\gamma^{1/2}\|g_N-\mathbf{n}_\gamma\cdot\nabla u_h\|_{L^2(\gamma)} \\ &\leq C_3^2 C_4 \|\nabla(u-u_h)\|_{L^2(\omega_\gamma)} + C_3^2 C_5 h_\gamma \|f_K+\nabla^2 u_h\|_{L^2(K)} \end{aligned} \tag{79}$$

The final step is applying Lemma 6 with $p=2$ on inequality (77), (78) and (79), summing the new bounds and then taking maximum over every generic constant. In this way, we can summarize the derivations in this subsection:

Theorem 11. *The explicit residual estimator* (13a) *is reliable and efficient.*

5.3 General explicit L^p-estimator

This residual estimator is a generalization of the previous one, and differs mostly in its derivation. But its underlying structure is quite similar.

Derivation of the estimator

The explicit L^p-estimator is applicable when $p \in (1, \infty)$ and the domain Ω is convex. To derive it, we apply the Aubin-Nitsche method [4] where we consider the dual (adjoint) version of the original weak formulation (22).

$$\Phi_F \in V \quad : \quad a(v, \Phi_F) = (v, F) \qquad \forall v \in V \tag{80}$$

where $\|\Phi_F\|_{W^{2,q}} \leq C_F \|F\|_{L^q}$ due to convexity. By using inequalities (67) and (68), we let $t = 2$, $s = 0$, and replace p by q. The conjugate identity, $\frac{1}{p} + \frac{1}{q} = 1$, is applied from now on. We use the adjoint formulation:

$$\begin{aligned}
(e, F) &= a(e, \Phi_F) \\
&= \sum_{K \in \mathcal{M}} \int_K r[(I - \Pi_h)\Phi_F] \, d\Omega + \sum_{\gamma \in \Sigma} \int_\gamma j[(I - \Pi_h)\Phi_F] \, ds \\
&\leq \sum_{K \in \mathcal{M}} \|r\|_{L^p(K)} \|(I - \Pi_h)\Phi_F\|_{L^q(K)} + \sum_{\gamma \in \Sigma} \|j\|_{L^p(\gamma)} \|(I - \Pi_h)\Phi_F\|_{L^q(\gamma)} \\
&\leq \sum_{K \in \mathcal{M}} C_1 h_K^2 \|\Phi_F\|_{W^{2,q}(\tilde{K})} \|r\|_{L^p(K)} + \sum_{\gamma \in \Sigma} C_2 h_\gamma^{1+1/p} \|\Phi_F\|_{W^{2,q}(\tilde{K})} \|j\|_{L^p(\gamma)} \\
&\leq C_1 \left[\sum_{K \in \mathcal{M}} \|\Phi_F\|^q_{W^{2,q}(\tilde{K})} \right]^{1/q} \left[\sum_{K \in \mathcal{M}} h_K^{2p} \|r\|^p_{L^p(K)} \right]^{1/p} \\
&\quad + C_2 \left[\sum_{\gamma \in \Sigma} \|\Phi_F\|^q_{W^{2,q}(\tilde{K})} \right]^{1/q} \left[\sum_{\gamma \in \Sigma} h_\gamma^{p+1} \|j\|^p_{L^p(\gamma)} \right]^{1/p} \\
&\leq 2C_3 C_F \|F\|_{L^q(\Omega)} \left\{ \sum_{K \in \mathcal{M}} h_K^{2p} \|r\|^p_{L^p(K)} + \sum_{\gamma \in \Sigma} h_\gamma^{p+1} \|j\|^p_{L^p(\gamma)} \right\}^{1/p}
\end{aligned}$$

where $C_3 = \max\{C_1, C_2\}$. Here, we have used inequality (69) in a general way with conjugate exponents. The L^p-norm in this case will be

$$\|e\|_{L^p(\Omega)} = \sup_{F \in L^q(\Omega)} \frac{(e, F)}{\|F\|_{L^q(\Omega)}}$$

By setting $C_{L^p} = 2C_3 C_F$ in addition, we obtain the final inequality:

$$\|e\|^p_{L^p(\Omega)} C^p_{L^p} \left[\sum_{K \in \mathcal{M}} h_K^{2p} \|r\|^p_{L^p(K)} + \sum_{\gamma \in \Sigma} h_\gamma^{p+1} \|j\|^p_{L^p(\gamma)} \right]$$

The error estimator is defined as

$$\boxed{\eta^p_{L^p(K)} = h_K^{2p}\|r\|^p_{L^p(K)} + h_\gamma^{p+1}\|j\|^p_{L^p(\partial K)}} \tag{81}$$

Robustness of the estimator

In the same way as the explicit estimator for the energy norm (13a), we just sum the error in the L^p-norm over every element on the domain. This yields a lower bound ensuring that the estimator is reliable:

$$\begin{aligned}\|e\|^p_{L^p} &= \sum_{K\in\mathcal{M}} \|e\|^p_{L^p(K)} \\ &\leq \sum_{K\in\mathcal{M}} C^p_{L^p(K)}\eta^p_{L^p(K)} \\ &\leq \left(\max_{K\in\mathcal{M}} C^p_{L^p(K)}\right)\eta^p_{L^p}\end{aligned}$$

Since Ω has finite measure, we can invoke the L^p-embedding [31]

$$1 \leq p_1 < p_2 \leq \infty \implies L^{p_2} \hookrightarrow L^{p_1}$$

In addition, the approximated solution is constructed from a finite-dimensional subspace in $L^p(\Omega)$, so we can apply norm equivalence to state that

$$C_{p_2,p_1}\|u\|_{L^{p_2}(\Omega)} \leq \|u\|_{L^{p_1}(\Omega)} \leq C_{p_1,p_2}\|u\|_{L^{p_2}(\Omega)} \tag{82}$$

The right-hand side holds when Ω has finite measure, and the left-hand side is a consequence of norm equivalence for finite-dimensional spaces. To prove the L^p-estimator's efficiency, we adapt much of the proof in Subsection 5.2 and generalize it by invoking inequality (82).

Our first step is defining $w_K = (f_K + \nabla^2 u_h)\psi_K$, applying inequality (82) from L^q to L^2, and then invoking Hölder's inequality such that

$$\begin{aligned}\|f_K + \nabla^2 u_h\|^2_{L^q(K)} &\leq C^2_{q,2}\|f_K + \nabla^2 u_h\|^2_{L^2(K)} \\ &\leq C^2_{q,2}C_1^2\int_K (f_K + \nabla^2 u_h)^2\psi_K \, d\Omega \\ &= C^2_{q,2}C_1^2\int_K \nabla(u - u_h)\cdot\nabla w_K \\ &\leq C^2_{q,2}C_1^2\|\nabla(u-u_h)\|_{L^p(K)}\cdot C_2 h_K^{-1}\|f_K + \nabla^2 u_h\|_{L^q(K)}\end{aligned}$$

We divide both sides with $h_K^{-1}\|f_K + \nabla^2 u_h\|_{L^q(K)}$, apply inequality (82) from L^p to L^q, and set $\widetilde{C}_1 = (C_{q,2}C_1)^2/C_{p,q}$ to obtain the first bound:

$$h_K\|f_K + \nabla^2 u_h\|_{L^p(K)} \leq C_2\widetilde{C}_1\|\nabla(u - u_h)\|_{L^p(K)} \tag{83}$$

The next step is setting $w_\gamma = -\mathbb{J}_\gamma(\mathbf{n}_\gamma \cdot \nabla u_h)\psi_\gamma$. As we did above, we get

$$\begin{aligned}
&\|\mathbb{J}_\gamma(\mathbf{n}_\gamma \cdot \nabla u_h)\|^2_{L^q(\gamma)} \\
&\leq C_{q,2}^2\|\mathbb{J}_\gamma(\mathbf{n}_\gamma \cdot \nabla u_h)\|^2_{L^2(\gamma)} \\
&\leq C_{q,2}^2 C_3^2 \int_\gamma (\mathbb{J}_\gamma(\mathbf{n}_\gamma \cdot \nabla u_h))^2\psi_K \, ds \\
&= C_{q,2}^2 C_3^2 \left[\sum_{K\subset\omega_\gamma}\int_K -(f_K + \nabla^2 u_h)w_\gamma \, d\Omega + \int_{\omega_\gamma} \nabla(u - u_h)\cdot\nabla w_\gamma \, ds\right] \\
&\leq C_{q,2}^2 C_3^2 \Bigg[C_4 h_\gamma^{1/p-1}\|\nabla(u - u_h)\|_{L^p(\omega_\gamma)}\|\mathbb{J}_\gamma(\mathbf{n}_\gamma \cdot \nabla u_h)\|_{L^q(\gamma)} \\
&\qquad + \sum_{K\subset\omega_\gamma} C_5 h_\gamma^{1/p}\|f_K + \nabla^2 u_h\|_{L^p(K)}\|\mathbb{J}_\gamma(\mathbf{n}_\gamma \cdot \nabla u_h)\|_{L^q(\gamma)}\Bigg]
\end{aligned}$$

We divide both sides with $h_K^{-1/q}\|\mathbb{J}_\gamma(\mathbf{n}_\gamma \cdot \nabla u_h)\|_{L^q(\gamma)}$, apply inequality (82) from L^p to L^q, and set $\widetilde{C}_3 = (C_{q,2}C_3)^2/C_{p,q}$ to obtain the next bound:

$$\begin{aligned}
&h_\gamma^{1/q}\|\mathbb{J}_\gamma(\mathbf{n}_\gamma \cdot \nabla u_h)\|_{L^p(\gamma)} \\
&\leq \widetilde{C}_3\left[C_4\|\nabla(u - u_h)\|_{L^p(\omega_\gamma)} + \sum_{K\subset\omega_\gamma} C_5 h_\gamma\|f_K + \nabla^2 u_h\|_{L^p(K)}\right]
\end{aligned} \tag{84}$$

Lastly, we define $w_\gamma = (g_N - \mathbf{n}_\gamma \cdot \nabla u_h)\psi_\gamma$ and derive as follows:

$$\begin{aligned}
&\|g_N - \mathbf{n}_\gamma \cdot \nabla u_h\|^2_{L^q(\gamma)} \\
&\leq C_{q,2}\|g_N - \mathbf{n}_\gamma \cdot \nabla u_h\|^2_{L^2(\gamma)} \\
&\leq C_{q,2}^2 C_3^2 \int_\gamma (g_N - \mathbf{n}_\gamma \cdot \nabla u_h)^2\psi_\gamma \, ds \\
&= C_{q,2}^2 C_3^2 \left[\int_K \nabla(u - u_h)\cdot\nabla w_\gamma \, d\Omega - \int_K (f_K + \nabla^2 u_h)w_\gamma \, d\Omega\right] \\
&\leq C_{q,2}^2 C_3^2 \left[C_4 h_\gamma^{1/p-1}\|\nabla(u - u_h)\|_{L^p(K)} + C_5 h_\gamma^{1/p}\|f_K + \nabla^2 u_h\|_{L^p(K)}\right] \\
&\qquad \times \|g_N - \mathbf{n}_\gamma \cdot \nabla u_h\|_{L^q(\gamma)}
\end{aligned}$$

We divide both sides with $h_K^{-1/q}\|g_N - \mathbf{n}_\gamma \cdot \nabla u_h\|_{L^q(\gamma)}$, invoke inequality (82) from L^p and L^q, and insert (83) such that we obtain the final bound

$$\begin{aligned} & h_\gamma^{1/q}\|g_N - \mathbf{n}_\gamma \cdot \nabla u_h\|_{L^p(\gamma)} \\ & \leq \widetilde{C}_3 \left[C_4 \|\nabla(u - u_h)\|_{L^p(\omega_\gamma)} + \sum_{K \subset \omega_\gamma} C_5 h_\gamma \|f_K + \nabla^2 u_h\|_{L^p(K)} \right] \end{aligned} \tag{85}$$

Using the same argumentation as the previous estimator, we can summarize everything as follows:

Theorem 12. *The general explicit L^p-estimator (81) is reliable and efficient.*

5.4 Adaptation to IGA

All the error estimators presented in this chapter can be applied directly in adaptive finite element modelling using B-splines to discretize the problem. Doing so has some useful advantages which are not available in the previous approaches. The B-splines have higher continuity than the classical FEM interpolants, so computation of the error estimator goes significantly faster, in addition to lack of jumps caused by discontinuities in the derivative of the shape functions. If we increase the continuity in addition to the polynomial degree, which is straightforward and requires less computational effort than classical FEM, then the global approximation error is also reduced, and the refinement speeds up.

However, there are some cases where introducing C^0-lines on the mesh is inevitable. This can be caused by the domain's structure, like the need for multiple patches or some special handling of discontinuities in the analytical solution. Another reason might be variable material parameters occurring in the PDE itself, which changes abruptly on different parts of the domain. In such cases, it is best to construct the mesh such that it has full continuity everywhere except on the C^0-lines. The jump $-\mathbb{J}_\gamma(\mathbf{n}_\gamma \cdot \nabla u_h)$ occurring in (74b) can be restricted just to the C^0-lines such that the computation of the error estimator still goes fast.

6 Enhancement-based estimators

This section focuses on error estimators used in connection with h-, p- and k-refinement in IGA. The two first ones were originally developed in the context of classical FEM. It has been demonstrated that they are fully compatible with IGA. The last one, a kind of hybridization, does not exist in classical FEM.

Consider the original model problem (22) such that the linear operator $\mathcal{L}$ is self-adjoint, and $V_h \subset V$ has finite dimension such that $u \in V$ and $u_h \in V_h$. We define a new subspace $V_h^* \subset V$ obtained from global h-, p- and k-refinement. If $u_h^* \in V_h^*$ is another FE-approximation of (22), we can derive the following inequality:

$$\begin{aligned} |||e||| &= |||u - u_h||| \\ &= |||u - u_h^* + u_h^* - u_h||| \\ &\leq \underbrace{|||u - u_h^*|||}_{\text{Non-Computable}} + \underbrace{|||u_h^* - u_h|||}_{\text{Computable}} \end{aligned}$$

This splitting is useful in the derivation of error estimators and analysis of their properties, which has been investigated in [2, 24, 25, 51]. If u_h^* is a sufficiently accurate approximation of u, we may approximate the error by the computable part of the inequality above, which reduces to

$$|||e||| \approx |||u_h^* - u_h||| = \eta_h^*$$

The error estimator above is clearly computable because it is independent of u. By the statement "if u_h^* is sufficiently accurate", we mean that there exists a constant $C_\theta \in [0, 1)$ satisfying $C_\theta = \mathcal{O}(h)$ such that

$$|||u - u_h^*||| \leq C_\theta |||u - u_h||| \tag{86}$$

This is called the *saturation assumption* [88]. For the model problem, we know from the previous sections that the error in the energy norm is

$$|||e||| = a(u - u_h, u - u_h)$$

6.1 *h*-refinement

In h-refinement, we keep the polynomial degree p of the shape functions constant, and the mesh size h is reduced by refining the elements. It is common to measure the convergence with respect to number of degrees of freedom (total number of unknown variables in the discrete equation system) because the elements can be locally refined, and the mesh size is variable.

There are three general ways of performing h-refinement [92]:

1. *Uniform Mesh Refinement* (UMR): All elements are divided into smaller elements repeatedly, and the global mesh is preserved [51].
2. *Element Subdivision*: We just select some individual elements and refine them locally. The conformal mesh can be preserved by proper handling of so-called "hanging nodes" [24].
3. *Remeshing*: The initial global mesh is completely discarded and replaced by a better new mesh [25].

In the context of IGA, h-refinement corresponds to *knot insertion* [85, 75]. We add an extra knot $\widehat{\xi}$ to a knot vector Ξ. This yields more basis functions and better shape control of the spline. It does not require subdivision of spline or changing geometric shape. The method relies on *Böhm's* theorem.

Theorem 13 (Böhm's theorem [34])**.** *If $\widehat{\xi} \in [\xi_s, \xi_{s+1})$, the knot insertion process can be described as follows:*

$$\sum_{i=0}^{n} N_{i,p}\mathbf{P}_i \longmapsto \sum_{i=0}^{n} \widehat{N}_{i,p}\widehat{\mathbf{P}}_i \tag{87a}$$

$$\Xi = \{\xi_0, \ldots, \xi_m\} \longmapsto \widehat{\Xi} = \{\xi_0, \ldots, \xi_s, \widehat{\xi}, \xi_{s+1}, \ldots, \xi_m\} \tag{87b}$$

$$\widehat{\mathbf{P}}_i = \begin{cases} \mathbf{P}_i, & 0 \leq i \leq s-p \\ (1-\alpha_i)\mathbf{P}_{i-1} + \alpha_i \mathbf{P}_i, & s-p+1 \leq i \leq s \\ \mathbf{P}_{i-1}, & s+1 \leq i \leq n+1 \end{cases} \tag{87c}$$

$$\alpha_i = \frac{\widehat{\xi} - \xi_i}{\xi_{i+p} - \xi_i} = \frac{\widehat{\xi} - \widehat{\xi}_i}{\widehat{\xi}_{i+p+1} - \widehat{\xi}_i} \tag{87d}$$

This process can be generalized by direct simultaneous insertion of multiple knots, and the *Oslo algorithm* [27, 59] is well-suited for this application. The algorithm is also compatible with NURBS, with some additional features.

As shown in Figures 15 and 16, the initial knot vectors in both directions are $\Xi, \mathcal{H} = \{0, 0, 0, 1, 1, 1\}$. After inserting the knots $1/4$, $1/2$ and $3/4$, they just become the new vectors $\widehat{\Xi}, \widehat{\mathcal{H}} = \{0, 0, 0, 1/4, 1/2, 3/4, 1, 1, 1\}$. Since we have only one single element to begin with, the control points and Greville points will coincide, but not after the addition of more knots.

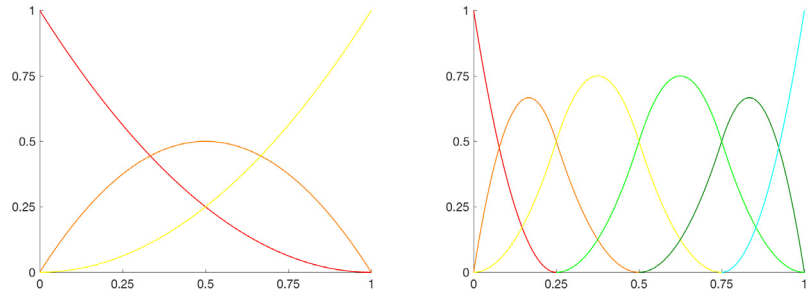


Figure 15. h-refinement: Original and refined basis functions

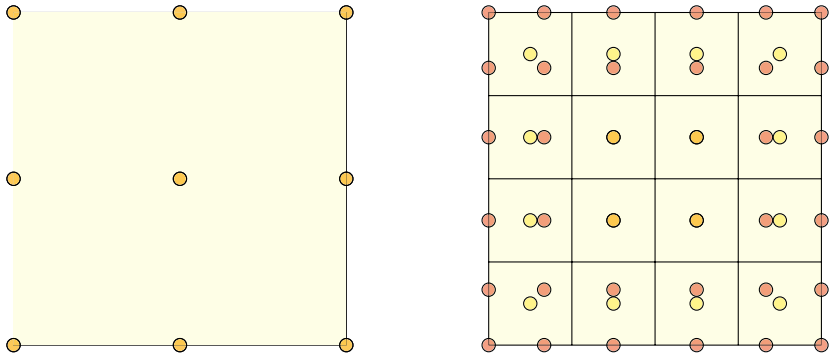


Figure 16. h-refinement: Physical mesh, control points (•) and Greville points (•).

6.2 *p*-refinement

In p-refinement, the global mesh remains completely unchanged while we increase the polynomial degree p. This can be done uniformly such that the degrees of all the shape functions increases, or locally by elevating some selected elements. The convergence is therefore naturally measured in the polynomial degree, and not the number of degrees of freedom.

In the context of IGA, p-refinement corresponds to *degree elevation* [85, 75]. This is very useful for splines because the continuity can be increased, and compatibility with the geometric shape becomes better.

If the original knot vector is Ξ, we can increase or reduce the order just by adding ($\widehat{\Xi}$) or removing ($\widetilde{\Xi}$) the end knots:

$$\Xi = \{\underbrace{\xi_0, \ldots, \xi_0}_{p+1}, \xi_1, \xi_2, \ldots, \xi_{s-1}, \underbrace{\xi_s, \ldots, \xi_s}_{p+1}\}$$
$$\widehat{\Xi} = \{\underbrace{\xi_0, \ldots, \xi_0}_{p+2}, \xi_1, \xi_2, \ldots, \xi_{s-1}, \underbrace{\xi_s, \ldots, \xi_s}_{p+2}\}$$
$$\widetilde{\Xi} = \{\underbrace{\xi_0, \ldots, \xi_0}_{p}, \xi_1, \xi_2, \ldots, \xi_{s-1}, \underbrace{\xi_s, \ldots, \xi_s}_{p}\}$$

As we see from Figures 17 and 18, the original knot vectors in both directions are $\Xi, \mathcal{H} = \{0, 0, 0, 1, 1, 1\}$. We increase the multiplicity of every knot, and the new vectors are $\widehat{\Xi}, \widehat{\mathcal{H}} = \{0, 0, 0, 0, 1, 1, 1, 1\}$.

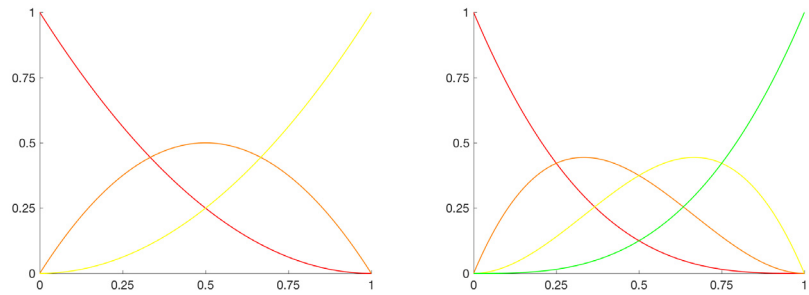


Figure 17. p-refinement: Original and refined basis functions

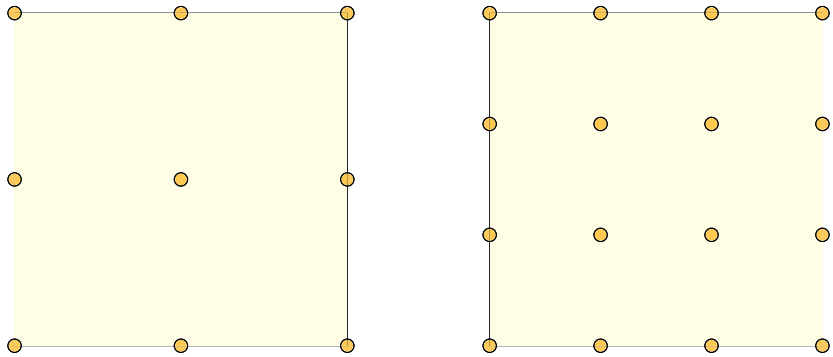
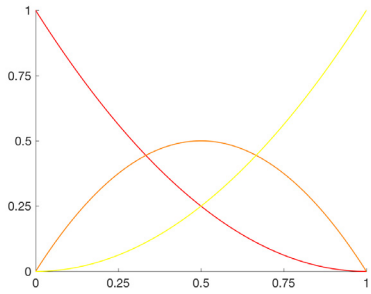
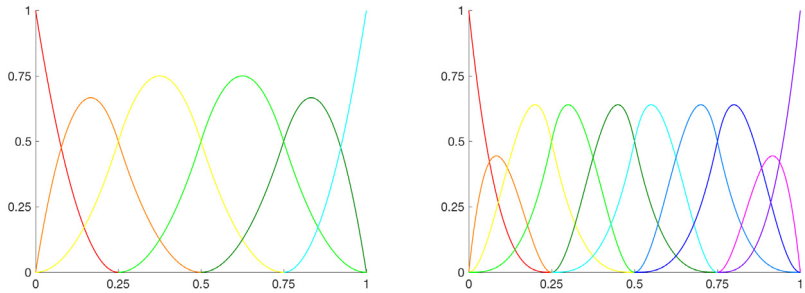


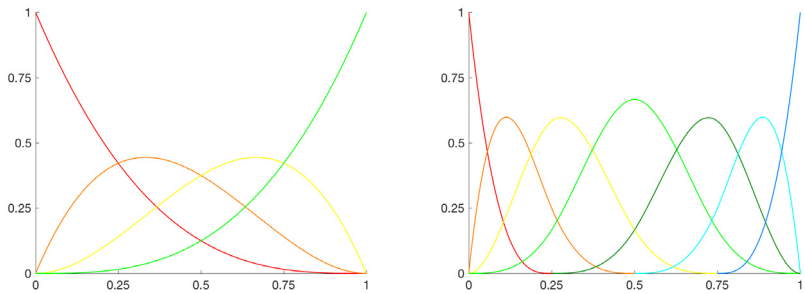
Figure 18. p-refinement: Physical mesh, control points (•) and Greville points (•).



(a) Initial mesh

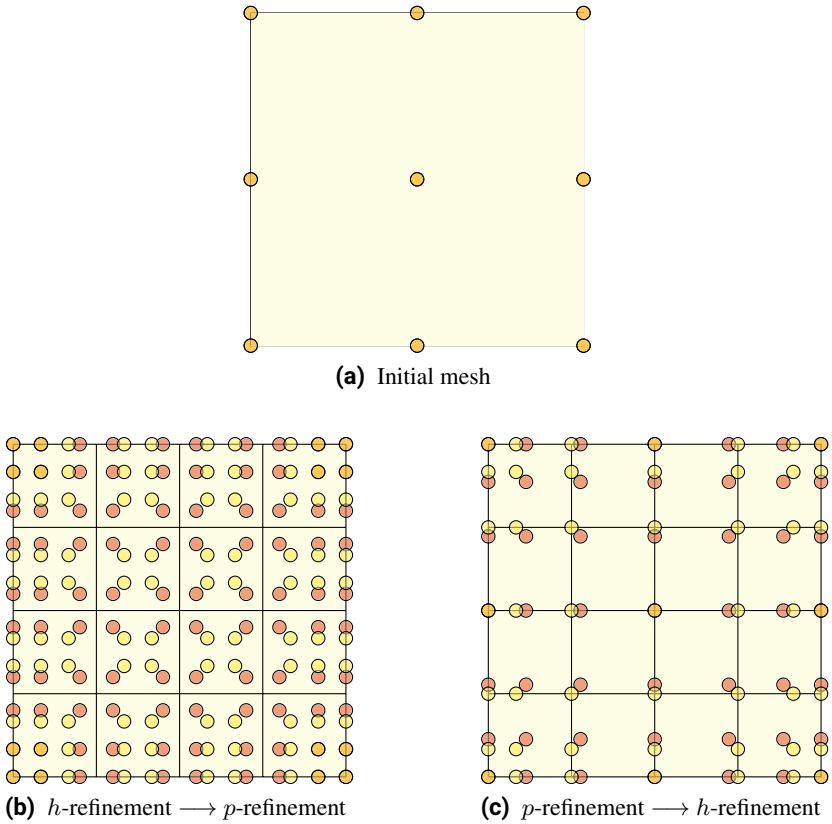


(b) h-refinement $\longrightarrow$ p-refinement



(c) p-refinement $\longrightarrow$ h-refinement

Figure 19. Combining h- and p-refinement: Original and refined basis functions. The operations knot insertion and degree elevation are non-commutative.



(a) Initial mesh

(b) h-refinement $\longrightarrow$ p-refinement

(c) p-refinement $\longrightarrow$ h-refinement

Figure 20. Combination of h- and p-refinement: Physical mesh, control points (•) and Greville points (•).

Figures 19 and 20 demonstrate that degree elevation and knot insertion do not commute. In the first approach, knot insertion on $\Xi, \mathcal{H} = \{0,0,0,1,1,1\}$ yields $\widetilde{\Xi}', \widetilde{\mathcal{H}}' = \{0,0,0,1/4,1/2,3/4,1,1,1\}$. Degree elevation results in $\widetilde{\Xi}, \widetilde{\mathcal{H}} = \{0,0,0,0,1/4,1/4,1/2,1/2,3/4,3/4,1,1,1,1\}$. The new basis functions are only C^1-continuous on the knots although they are cubic. In the next approach, degree elevation yields $\widehat{\Xi}', \widehat{\mathcal{H}}' = \{0,0,0,0,1,1,1,1\}$ first, and knot insertion results in $\widehat{\Xi}, \widehat{\mathcal{H}} = \{0,0,0,0,1/4,1/2,3/4,1,1,1,1\}$. This time, the number of basis functions is smaller, and the continuity is C^2. We see clearly from this illustration that the second approach yields the best approximation due to higher continuity, and the number of degrees of freedom is reduced.

6.3 *k*-refinement

k-refinement is a brand-new method which is only available in IGA. It allows us to control the continuity and growth of control variables. We can apply it because the patches on the domain have a homogeneous structure. If we want to increase the polynomial degree from p to $p+1$, we increase the continuity similarly from q to $q+1$. To do so, we just increase the multiplicity of the end knots, not the interior knots in addition as in p-refinement. In this way, the dimension and number of degrees of freedom for the new spline space will not grow too large. Thus, we have elevated the degree and continuity simultaneously, which is not possible in classical FEM. The mesh is still the same, but higher continuity yields better approximation since the error decreases more [85, 40].

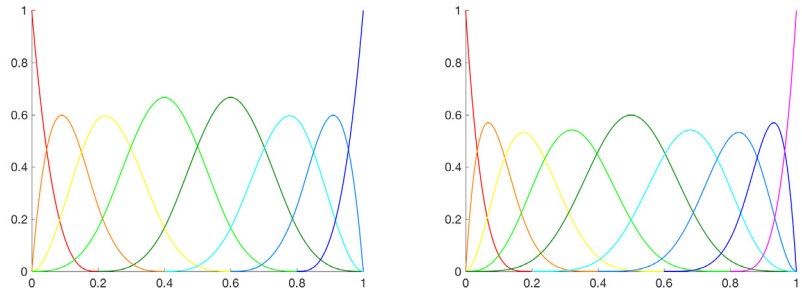


Figure 21. k-refinement: Original and refined basis functions

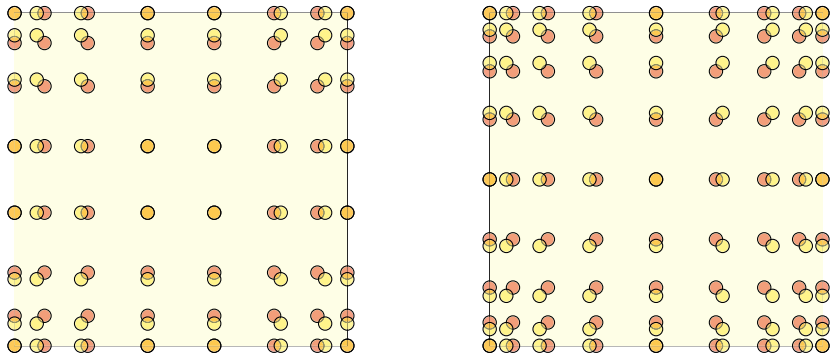
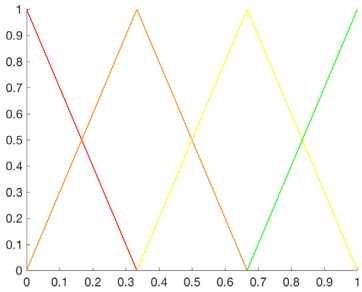
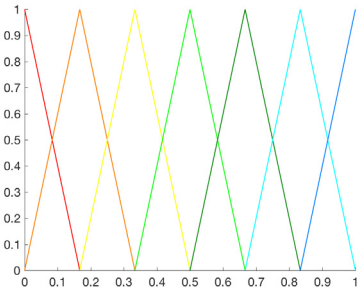


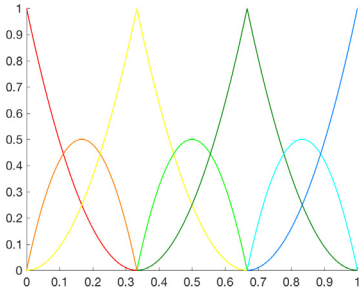
Figure 22. k-refinement: Physical mesh, control points (•) and Greville points (•).



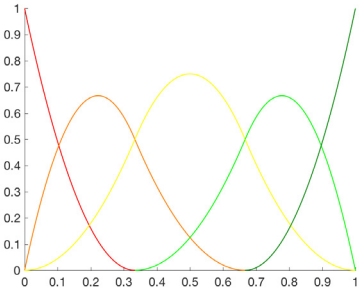
(a) $\Xi = \{0, 0, \frac{1}{3}, \frac{2}{3}, 1, 1\}, p = 1$



(b) $\Xi = \{0, 0, \frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}, 1, 1\}, p = 1$



(c) $\Xi = \{0, 0, 0, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1, 1, 1\}, p = 2$



(d) $\Xi = \{0, 0, 0, \frac{1}{3}, \frac{2}{3}, 1, 1, 1\}, p = 2$

Figure 23. Collective comparison of h-, p- and k-refinement.

As we see from Figure 23, we start with $\Xi = \{0, 0, 1/3, 2/3, 1, 1\}$ and refine it in three ways using the methods described above. The dimension of the spline space $\mathbb{S}_h^{p+1,k+1}(\mathcal{M})$ obtained by k-refinement is just one unit more than $\mathbb{S}_h^{p,k}(\mathcal{M})$, the original one. This is because we only increased the multiplicity of the end knots. But for the spline spaces $\mathbb{S}_{h/2}^{p,k}(\mathcal{M})$ and $\mathbb{S}_h^{p+1,k}(\mathcal{M})$, obtained respectively from h- and p-refinement, the dimensions are significantly larger than for the k-refinement due to addition of more knots. Furthermore, we observe that p-refinement does not yield enough differentiable splines since it ignored the continuity at each knot.

6.4 Analysis and quality comparison of the refinements

Comparison of dimension, embedding and asymptotics

Let $\mathcal{M}$ be the original tensor-mesh, and $\overline{\mathcal{M}}$ is the new mesh obtained by halving every element of $\mathcal{M}$ in each direction, such that $\mathbb{S}_h^{p,k}(\mathcal{M})$ is an isogeometric finite element space. Then, we have the following mappings which are induced by the refinements discussed previously:

$$\mathbb{S}_h^{p,k}(\mathcal{M}) \xrightarrow{h\text{-refinement}} \mathbb{S}_{h/2}^{p,k}(\overline{\mathcal{M}}) \tag{88a}$$

$$\mathbb{S}_h^{p,k}(\mathcal{M}) \xrightarrow{p\text{-refinement}} \mathbb{S}_h^{p+1,k}(\mathcal{M}) \tag{88b}$$

$$\mathbb{S}_h^{p,k}(\mathcal{M}) \xrightarrow{k\text{-refinement}} \mathbb{S}_h^{p+1,k+1}(\mathcal{M}) \tag{88c}$$

According to [66], it can be shown that all of these subspaces above satisfy some very important inclusions:

$$\mathbb{S}_h^{p,k}(\mathcal{M}) \subseteq \mathbb{S}_{h/2}^{p,k}(\overline{\mathcal{M}}) \qquad \mathbb{S}_h^{p,k}(\mathcal{M}) \nsubseteq \mathbb{S}_h^{p+1,k+1}(\mathcal{M}) \tag{89a}$$

$$\mathbb{S}_h^{p,k}(\mathcal{M}) \subseteq \mathbb{S}_h^{p+1,k}(\mathcal{M}) \qquad \mathbb{S}_h^{p,k}(\mathcal{M}) \nsupseteq \mathbb{S}_h^{p+1,k+1}(\mathcal{M}) \tag{89b}$$

Let the parametric domain be $\widehat{\Omega} = [0,1]^d$ such that $\dim \mathbb{S}_h^{p,k}(\mathcal{M}) = N^d$ and $\mathcal{M}$ is a uniform tensor mesh with mesh size h. Then the spaces obtained from uniform refinement have the following dimensions:

$$\dim \mathbb{S}_{h/2}^{p,k}(\overline{\mathcal{M}}) = (2N - k - 1)^d \tag{90a}$$

$$\dim \mathbb{S}_h^{p+1,k}(\mathcal{M}) = \left(N + \frac{1}{h}\right)^d \tag{90b}$$

$$\dim \mathbb{S}_h^{p+1,k+1}(\mathcal{M}) = (N+1)^d \tag{90c}$$

When the linear system of equations grows large, it can be shown that

$$\frac{\dim \mathbb{S}_{h/2}^{p,k}(\overline{\mathcal{M}})}{\dim \mathbb{S}_h^{p,k}(\mathcal{M})}, \frac{\dim \mathbb{S}_h^{p+1,k}(\mathcal{M})}{\dim \mathbb{S}_h^{p,k}(\mathcal{M})} \longrightarrow 2^d$$

$$\frac{\dim \mathbb{S}_h^{p+1,k+1}(\mathcal{M})}{\dim \mathbb{S}_h^{p,k}(\mathcal{M})} \longrightarrow 1$$

These limits demonstrate that k-refinement provides the lowest number of degrees of freedom, implying minimal computational cost for assembling and solving the equation system arising from the isogeometric discretization.

Comparison by Kolmogorov *n*-widths

The advantages of k-refinement over h- and p-refinement can be explored further through an established approximation estimate [28]. First, we denote $\widehat{\Omega}$ and Ω as the parametric and physical domains, respectively. If $\mathcal{F} : \widehat{\Omega} \to \Omega$, $u : \Omega \to \mathbb{R}^d$, $u \in H^l$ and $0 \leq k \leq l \leq p+1$, then

$$\sum_{e=1}^{N} |u - \Pi_k u|^2_{H^k(\Omega^e)} \leq C \sum_{e=1}^{N} \sum_{i=0}^{l} h_e^{2(l-k)} \|\nabla \mathcal{F}\|_{L^\infty(\mathcal{F}^{-1}(\Omega_e))} |u|^2_{H^i(\Omega_e)} \tag{92}$$

where Ω^e is an arbitrary physical element and N is the total number of elements. The order of convergence in the refinement process depends just on the polynomial degree p, and IGA with p-th order NURBS yields the same convergence rate as FEM with p-th order interpolants. But the size of the constant C decreases when the continuity increases, which is not explicitly given in (92). The reason for this improved efficiency can be explained by Kolmogorov n-widths [76].

Let X be a Sobolev space such that $A, X_n \subset X$ and $\dim X_n = n$. We approximate and arbitrary element $x \in A$ in terms of another element $x_n \in X_n$. The distance between the point x and the space X_n is given by

$$E(x, X_n; X) = \inf_{x_n \in X_n} \|x - x_n\|_X \tag{93}$$

We call x_n^* the best approximation of x when $\|x - x_n^*\|_X = E(x, X_n; X)$. If we want to approximate all elements $x \in A$, we need a deviation given by

$$E(A, X_n; X) = \sup_{x \in A} \inf_{x_n \in X_n} \|x - x_n\|_X \tag{94}$$

The best n-dimensional subspace for approximating the subset A is given by the Kolmogorov n-width, which is expressed as

$$d_n(A, X) = \inf_{\substack{X_n \subset X_n \\ \dim X_n = n}} \sup_{x \in A} \inf_{x_n \in X_n} \|x - x_n\|_X \tag{95}$$

Let us assume that there are several n-dimensional subspaces that can be used for approximating the subset A, for example Y_n. In this case, we have to introduce a *comparison ratio* which is given by

$$\kappa(A, X_n, Y_n; X) = \frac{E(A, X_n; X)}{E(A, Y_n; X)} \tag{96}$$

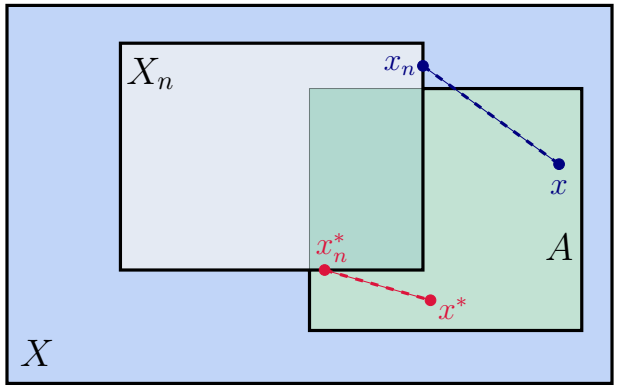


Figure 24. Distance between the subspaces A and X_n, measured in the worst scenario when the deviation is too high.

The value of κ reveals which space is appropriate for the approximation:

$\kappa \approx 1$	Both spaces work
$\kappa \ll 1$	Choose X_n
$\kappa \gg 1$	Choose Y_n

If $E(A, \widetilde{X}_n; X) = d_n(A, X)$, we call $\widetilde{X}_n$ an optimal n-dimensional subspace. In this case, we can define the *optimality ratio* is

$$\Lambda(A, \widetilde{X}_n; X) = \frac{E(A, \widetilde{X}_n; X)}{d_n(A, X)} \tag{97}$$

We consider the parametric domain $(0, 1)$, and $X = B^m(0, 1)$ is the unit ball on $H^m(0, 1)$:

$$B^m(0, 1) = \{x \in H^m(0, 1) \, : \, \|x\|_{H^m(0,1)} \leq 1\} \tag{98}$$

Lastly, we assume that p and k are fixed, such that $k \in [0, p-1]$. If the number of degrees of freedom increases, then Λ converges to a unique limit $L \geq 1$. By increasing k gradually, this limit L will be smaller, and if $k = p - 1$, which is the optimal continuity, then $L = 1$. Furthermore, there is no other value of k than $p - 1$ which makes the limit equal 1.

$$k = p - 1 \quad \Longrightarrow \quad \inf_{0 \leq k \leq p-1} \lim_{N_{\text{dof}} \to \infty} \Lambda = 1$$

This convergence has been verified numerically in [49]. Using Kolmogorov n-widths for Sobolev spaces demonstrates that the k-refinement works best, since we can control and preserve the continuity.

6.5 Serendipity pairing of $\mathbb{S}_h^{p,k}(\mathcal{M}) - \mathbb{S}_h^{p+1,k+1}(\mathcal{M})$

We will now discuss how to create a *Serendipity pairing* between $\mathbb{S}_h^{p,k}(\mathcal{M})$ and $\mathbb{S}_h^{p+1,k+1}(\mathcal{M})$, a useful technique in adaptive IGA which uses the local refinement methodology as developed in [63, 66]. Let $\mathcal{M}_0$ be the initial tensor mesh on the domain Ω. As mentioned earlier, the error in the energy norm can be expressed as

$$|||e||| \leq \underbrace{|||u - u_h^*|||}_{\text{Non-Computable}} + \underbrace{|||u_h^* - u_h|||}_{\text{Computable}} \tag{99}$$

We use LR B-splines to construct a discrete pair of k-refined approximation spaces $\mathbb{S}_h^{p,k}(\mathcal{M})$ and $\mathbb{S}_h^{p+1,k+1}(\mathcal{M})$. For adaptive LR-meshes, the k-refined space's dimension almost equals the dimension of the original one, and it will not grow too large. At each adaptive refinement level, the integration LR-meshes are the same for both spaces. We obtain two error estimators η_h^* and η_h^{RES}, which correspond respectively to the computable and non-computable parts of (99).

Since k-refinement does not share the same embedding property as h- and p-refinement as shown in (89), using it in adaptive refinement is easier, for there is no "embedding property to fulfil" here according to Kumar et al. [66]. Although k-refinement speeds up the whole process because of better approximation and computational efficiency, a little problem is remaining. The spaces $\mathbb{S}_h^{p,k}(\mathcal{M})$ and $\mathbb{S}_h^{p+1,k+1}(\mathcal{M})$ have some common elements, but they are not subspaces of each other. The natural question becomes how we should choose the elements to be refined correctly. Since $\mathbb{S}_h^{p+1,k+1}(\mathcal{M})$ provides better approximation than $\mathbb{S}_h^{p,k}(\mathcal{M})$, an appropriate solution for this problem is using their set difference $\mathbb{S}_h^{p+1,k+1}(\mathcal{M}) \backslash \mathbb{S}_h^{p,k}(\mathcal{M})$ to construct a new mesh $\mathcal{M}_{l+1}$ which is indeed a proper subspace of $\mathcal{M}_l$. Furthermore, the integration LR-mesh at each refinement level is be same for both the spaces. At each refinement level, we have the following inclusions:

$$\mathbb{S}_h^{p,k}(\mathcal{M}_l) \subset \mathbb{S}_h^{p,k}(\mathcal{M}_{l+1}) \quad , \quad \mathbb{S}_h^{p+1,k+1}(\mathcal{M}_l) \subset \mathbb{S}_h^{p+1,k+1}(\mathcal{M}_{l+1})$$

Another problem arising in the Serendipity pairing is the choice of the estimators η_h^* and η_h^{RES}. A natural choice is the standard explicit residual estimator (13a) for the non-computable part.

In some cases, the explicit residual estimator might be very conservative, causing too many elements to be refined due to over-estimation. Since the convergence in the computable quantity $|||u - u_h^*|||$ increases by one degree per refinement step, it might reduce the conservative effect a bit.

For smooth problems with quasi-uniform grids and even distribution of error, the k-refinement and Serendipity pairing work well. If the initial spline space is $\mathbb{S}_h^{p,k}$, it is possible to create a coarser space $\mathbb{S}_{mh}^{p,k}$ of a factor m to reduce the global error quickly. But this does not work for problems with pollution error and layers on the boundary or interior of the computational domain. In such cases, we must increase the grid in the critical regions first to restrain the deteriorating effect on the approximation quality.

Serendipity pairing

1: **procedure** SERENDIPITY_PAIRING($\mathcal{M}_0$, $\mathbb{S}_h^{p,k}(\mathcal{M}_0)$, $\mathbb{S}_h^{p+1,k+1}(\mathcal{M}_0)$)
2: **for** each refinement level l **do**
3: Use an error estimator to choose ϵ % of $B_i \in \mathbb{S}_h^{p+1,k+1}(\mathcal{M}_l)$.
4: Refine the chosen B-spline functions to obtain $\mathbb{S}_h^{p+1,k+1}(\mathcal{M}_{l+1})$.
5: Store information about the mesh-line of length $p+2$ in $\mathcal{E}_l$.
6: Refine $B_i \in \mathbb{S}_h^{p,k}(\mathcal{M}_l)$ with mesh-line $\mathcal{E}_l$ to obtain $\mathbb{S}_h^{p,k}(\mathcal{M}_{l+1})$.

7 Recovery-based estimation

In cases where the exact solution is smooth or the pollution error is controlled by adaptive refinement, it can be shown that measuring the error in the energy norm is suitable for determining which elements should be refined. The natural approach will therefore be to calculate the difference between the direct and post-processed gradients, and then use it as an error indicator for the refinement. If we have a smooth error distribution for the discretization, this error indicator may be both reliable and efficient [13, 67].
The key ingredient in this new approach is the *recovery operator* G_h. It computes ∇u_h^* in equation (86) by using the gradient of the computed FE-solution. The superconvergent recovery estimator reads:

$$\eta_{SPR}^2 = \int_\Omega |G_h[u_h] - \nabla u_h|^2 \, d\mathbf{x} \tag{100}$$

7.1 Characteristic properties

Following Ainsworth and Craig [1], we present some important conditions which are required to make recovery operators work properly:

1. *Consistency*, correct reproduction of the true gradient:

$$u \in \mathbb{P}^{p+1}(\widetilde{K}) \Longrightarrow G_h[\Pi_p u] \equiv \Pi_p(\nabla u) \tag{R1}$$

2. *Localization*, minimal computation of G_h:

$$x_0 \in K \Longrightarrow \nabla u\text{-values sampled on } \widetilde{K} \text{ define } G_h[u](x_0) \tag{R2}$$

3. *Boundedness and linearity*, reliable recovery of u_h:

$$|G_h[v]|_{L^\infty(K)} \leq D|v|_{W^{1,\infty}(\widetilde{K})} \quad , \quad K \in \mathcal{M},\ v \in V_h \tag{R3}$$

(R1) does not determine G_h uniquely but is together with (R3) sufficient to make the error estimator asymptotically exact. (R3) ensures that the recovered gradient belongs to the proper space for the problem at hand. (R2) is of practical nature because it constrains the computational effort involved in the recovery step. We recall that $\widetilde{K}$ is a patch containing an element K, and its size can vary.

Theorem 14 (Robustness of recovery estimators)**.** *Recovery operators that satisfy conditions* (R1)-(R3) *are reliable and efficient, and thus robust.*

Proof. We assume that $V_h, V_h^* \subset H^1$ such that $u \in H^1$, $\nabla u_h \in V_h$ and $G_h[u_h] \in V_h^*$. We derive an upper bound for the reliability by combining the saturation assumption (86) with Minkowski's L^p-inequality:

$$\begin{aligned}
&\|G_h[u_h] - \nabla u_h\|_{L^2(\Omega)} \\
&= \|G_h[u_h] - \nabla u + \nabla u - \nabla u_h\|_{L^2(\Omega)} \\
&\leq \|G_h[u_h] - \nabla u\|_{L^2(\Omega)} + \|\nabla u - \nabla u_h\|_{L^2(\Omega)} \\
&\leq C_\theta \|\nabla(u - u_h)\|_{L^2(\Omega)} + \|\nabla(u - u_h)\|_{L^2(\Omega)}
\end{aligned}$$

Similarly, we derive a lower bound for the efficiency:

$$\begin{aligned}
&\|\nabla(u - u_h)\|_{L^2(\Omega)} \\
&= \|\nabla u - G_h[u_h] + G_h[u_h] - \nabla u_h\|_{L^2(\Omega)} \\
&\leq \|\nabla u - G_h[u_h]\|_{L^2(\Omega)} + \|G_h[u_h] - \nabla u_h\|_{L^2(\Omega)} \\
&\leq C_\theta \|\nabla(u - u_h)\|_{L^2(\Omega)} + \|G_h[u_h] - \nabla u_h\|_{L^2(\Omega)}
\end{aligned}$$

We define the constants $C_\pm = (1 \pm C_\theta)^{-1}$ and combine everything together to obtain a two-sided bound:

$$C_+ \|G_h[u_h] - \nabla u_h\|_{L^2(\Omega)} \leq \|\nabla(u - u_h)\|_{L^2(\Omega)} \leq C_- \|G_h[u_h] - \nabla u_h\|_{L^2(\Omega)} \tag{101}$$

Hence, the recovery estimator is robust. □

Theorem 15 (Convergence of the recovery operator)**.** *If G_h is a recovery operator that satisfies the required criterions, and $u \in H^{r+2}(\Omega)$, then*

$$\|\nabla u - G_h(\Pi_h u)\|_{L^2(\Omega)} \leq C_{SPR} h^{r+1} |u|_{H^{r+2}(\Omega)} \tag{102}$$

Proof. We start locally by assuming that $u \in H^{r+2}(\widetilde{K})$ and $w \in \mathbb{P}^{r+1}(\widetilde{K})$. The error of the recovered gradient is decomposed as follows:

$$\begin{aligned}
E[u] &= \nabla u - G_h(\Pi_h u) \\
&= \nabla(u - w) + (\nabla w - \Pi_h \nabla w) + (\Pi_h \nabla w - G_h(\Pi_h u)) \\
&= \nabla(u - w) + \nabla(w - \Pi_h w) + G_h(\Pi_h(w - u))
\end{aligned}$$

From Minkowski's inequality, we obtain

$$\|E[u]\|_{L^2} \leq \|\nabla(u-w)\|_{L^2} + \|\nabla(w-\Pi_h w)\|_{L^2} + \|G_h(\Pi_h(u-w))\|_{L^2}$$

Now, we can analyse the three terms individually. By using the general a priori estimate in the H^k-norm (48), we get

$$\begin{aligned}\|\nabla(u-w)\|_{L^2(K)} &= |u-w|_{H^1(K)} \\ &\leq |u-w|_{H^1(\widetilde{K})} \\ &\leq C_1 h_K^{r+1} |u|_{H^{r+2}(\widetilde{K})}\end{aligned}$$

Minkowski's inequality and the general approximation property of Π_h yield

$$\begin{aligned}\|\nabla(w-\Pi_h w)\|_{L^2(K)} &\leq C_2 h_K^{r+1} |\nabla w|_{H^{r+1}(K)} \\ &= C_2 h_K^{r+1} |w|_{H^{r+2}(K)} \\ &\leq C_2 h_K^{r+1} |u-(u-w)|_{H^{r+2}(\widetilde{K})} \\ &\leq C_2 h_K^{r+1} \left[|u|_{H^{r+2}(\widetilde{K})} + |u-w|_{H^{r+2}(\widetilde{K})}\right]\end{aligned}$$

From the L^p-inclusion for domains of finite measure, the equivalence of L^p-norms, Hölder and Minkowski's inequalities, and the boundedness of Π_h, we obtain the following derivation

$$\begin{aligned}&\|G_h(\Pi_h(w-u))\|_{L^2(K)} \\ &\leq D_1 \|G_h(\Pi_h(w-u))\|_{L^1(K)} \\ &\leq D_1 \|1\|_{L^1(K)} \|G_h(\Pi_h(w-u))\|_{L^\infty(K)} \\ &= D_1 h_K \|G_h(\Pi_h(w-u))\|_{L^\infty(K)} \\ &\leq D_1 D_2 h_K \|\Pi_h(w-u)\|_{W^{1,\infty}(\widetilde{K})} \\ &\leq D_1 D_2 h_K D_3 h_K^{-1} \|\Pi_h(w-u)\|_{L^\infty(\widetilde{K})} \\ &\leq D_1 D_2 D_3 D_4 h_K^{-1} \|\Pi_h(w-u)\|_{L^2(\widetilde{K})} \\ &= D_5 h_K^{-1} \|(w-u) + \Pi_h(w-u) - (w-u)\|_{L^2(\widetilde{K})} \\ &\leq D_5 h_K^{-1} \left[\|w-u\|_{L^2(\widetilde{K})} + \|(1-\Pi_h)(w-u)\|_{L^2(\widetilde{K})}\right] \\ &\leq D_5 h_K^{-1} \left[D_6 h_K^{r+2} |w-u|_{H^{r+2}(\widetilde{K})} + D_7 h_K^{r+2} |w-u|_{H^{r+2}(\widetilde{K})}\right] \\ &= C_3 h_K^{r+1} |w-u|_{H^{r+2}(\widetilde{K})}\end{aligned}$$

where $D_5 = D_1 D_2 D_3 D_4$ and $C_3 = D_5(D_6 + D_7)$ are generic constants.

Since w is a polynomial approximation of u, we can take the infimum over all polynomials to exclude w from every inequality:

$$\begin{aligned}\|E[u]\|_{L^2(K)} &\leq \inf_{w\in\mathbb{P}^{r+1}} \|E[u]\|_{L^2} \\ &= h_K^{r+1}(C_1 + 2C_2 + C_3)|u|_{H^{r+2}(K)}\end{aligned}$$

By L^2-summation over the elements and invoking $h = \max h_K$, we obtain

$$\|E[u]\|_{L^2(\Omega)} = h^{r+1} C_{\text{SPR}} |u|_{H^{r+2}(\Omega)}$$

Thus, the global estimate for recovery operators has been established. □

In general, if C is a constant depending on u, $\tau \in (0, 1]$, and u is sufficiently regular, then superconvergence is obtained if the following estimate is valid:

$$\|u_h - \Pi_h u\|_{H^1(\Omega)} \leq C h^{p+\tau} \tag{103}$$

Corollary 2. *Suppose that* (102) *holds. Then we have the inequality*

$$\|\nabla u - G_h(u_h)\|_{L^2(\Omega)} \leq C h^{p+\tau} \tag{104}$$

Proof. Minkowski's inequality and the superconvergence criterion (103) yield the following derivation:

$$\begin{aligned}&\|\nabla u - G_h(u_h)\|_{L^2(K)} \\ &= \|\nabla u - G_h(\Pi_h u) + G_h(\Pi_h u) - G_h(u_h)\|_{L^2(K)} \\ &\leq \|\nabla u - G_h(\Pi_h u)\|_{L^2(K)} + \|G_h(\Pi_h u) - G_h(u_h)\|_{L^2(K)} \\ &\leq C_1 h_K^{p+1} |u|_{H^{p+2}(\widetilde{K})} + C_2 \|u_h - \Pi_h u\|_{H^1(\widetilde{K})} \\ &\leq C_1 h_K^{p+1} |u|_{H^{p+2}(\widetilde{K})} + C_2 C_3(u) h_K^{p+\tau}\end{aligned}$$

Summing over all the elements in the L^2-norm yields the desired estimate. The constant C is depending on u, but not on h. □

Definition 9 (Superconvergent points [89])**.** *Let I be a partition of the domain with maximal mesh width h, and $\xi = \xi(h)$ is a family of points. We call them superconvergent for function values of order $\sigma > 0$ if*

$$|e(\xi)| \leq C h^{r+\sigma} \tag{105a}$$

$$\left|\frac{d^k}{dx^k} e(\xi)\right| \leq C h^{r+\sigma-k} \tag{105b}$$

If the FE-solution u_h of a PDE is a (piecewise) polynomial of degree $r+1$, and the exact solution u is in $W^{r+2,k}$, then the error $e = u - u_h$ satisfies

$$\|e\|_{L^p(I)} + h\|e\|_{W^{1,p}(I)} \leq Ch^{r+2}\|u\|_{W^{r+2,p}(I)} \tag{106a}$$

$$\|e\|_{W^{-s,\infty}(I)} \leq Ch^{r+2+s}\|u\|_{W^{r+2,\infty}(I)} \tag{106b}$$

where $p \in [1, \infty]$ and $s \leq r-1$. If there are no mesh restrictions, we have

$$|e(x_i)| \leq Ch^{2r}\|u\|_{H^{r+2}(I)} \tag{107}$$

Theorem 16 (Shifted Legendre points [89])**.** *If the mesh is continuous with $I_i = (x_i, x_{i+1})$, we can translate the zeros of the k-th order Legendre polynomial $P_k(x)$ to I_k by the standard linear mapping*

$$\xi_i(x) = \frac{2x - (x_i + x_{i+1})}{x_{i+1} - x_i}$$

This yields the superconvergent estimate

$$|e'(\eta_i)| \leq Ch^r \left(\|u\|_{W^{r,\infty}(I)} + \|u\|_{W^{r+1,\infty}(I_i)}\right) \tag{108}$$

where η_i is a zero of $P_k(x)$, the Gauss-Legendre (GL) points, mapped to I_i. If $r \geq 3$, then h^r changes to h^{r+1}.

Theorem 17 (Shifted Lobatto points [89])**.** *If r is even, $k = 1$, and the nodal errors $|e'(x_i)|$, $|e'(x_{i+1/2})|$ and $|e'(x_{i+1})|$ are of order $\mathcal{O}(h^r)$ on I_i, then the error satisfies the order estimate*

$$|e(\xi)| = \mathcal{O}(h^{r+1})$$

where ξ is any of the $r-2$ roots of Q in $\overline{I}_i$, a polynomial given by

$$Q(x) = C\frac{d^{r-2}}{dx^{r-2}}\left[(x^2-1)^{r-2}\left(x^2 - \frac{r+2}{r-2}\right)\right]$$

These new points are the zeros of $(x^2-1)P'_{k-1}(x)$, the Gauss-Legendre-Lobatto (GLL) points. Both the GL- and GLL-points can be evaluated efficiently for any order by Newton iteration, which is most accurate [82]. A comprehensive table over superconvergence results depending on the polynomial degree and continuity is provided by Wahlbin in [89].

Definition 10 (Distance between sets)**.** *Assume that some well-defined sets satisfy $\Omega_0 \subseteq \Omega_1 \subseteq \mathcal{D}_h$. Then we define the distance between them as*

$$\partial_<(\Omega_0, \Omega_1) = dist(\partial\Omega_0 \backslash \partial\mathcal{D}_h, \partial\Omega_1 \backslash \partial\mathcal{D}_h)$$

From this, we introduce the following sets following sets:

$$C^\infty_<(\overline{\Omega}) = \{v \in C^\infty(\overline{\Omega}) : \partial_<(\mathrm{supp}(v), \Omega) > 0\}$$
$$\mathcal{S}^h_<(\overline{\Omega}) = \{v \in \mathcal{S}^h(\overline{\Omega}) : \partial_<(\mathrm{supp}(v), \Omega) > 0\}$$

Theorem 18 (Superapproximation [89])**.** *We have two constants c and C, and a number L. Let $\Omega_0 \subseteq \Omega_1 \subseteq \mathcal{D}_h$ and $\omega \in C^\infty_<(\overline{\Omega})$ such that*

$$d = \partial_<(\Omega_0, \Omega_1) > ch$$
$$\|\omega\|_{W^{l,\infty}(\Omega_0)} \leq \Lambda d^{-l} \quad , \quad 0 \leq l \leq L$$

Then, for all $\chi \in \mathbb{S}_h$, there is a $\psi \in \mathcal{S}^h_<(\Omega_1)$ such that

$$\|\omega\chi - \psi\|_{L^2(\Omega_1)} \leq C\Lambda \left(\frac{h}{d}\right) \|\chi\|_{L^2(\Omega_1)} \tag{109}$$

For any $1 \leq p, q \leq \infty$ and element K_i, we have

$$\|\chi\|_{L^p(K_i)} \leq Ch^{-n\left(\frac{1}{q}-\frac{1}{p}\right)} \|\chi\|_{L^q(K_i)} \quad , \quad \chi \in \mathbb{S}_h \tag{110}$$

Corollary 3. *Let the superapproximation property (110) hold with constants c_0 and C, and assume that the following orthogonality relation is true:*

$$(v_h, \chi) = 0 \quad , \quad \forall \chi \in \mathbb{S}^h_<(\Omega_1)$$

Then there is a constant c_1 such that the following estimate holds:

$$\|v_h\|_{L^2(\Omega_0)} \leq Ce^{c_1 d/h} \|v_h\|_{L^2(\Omega_1)} \tag{111}$$

Superconvergence can be obtained in many different ways. In most cases, we examine convergence of derivatives computed from the approximate solution at special mesh points. These are the superconvergence points, and the corresponding values are superconvergent, as we have seen now. In general, there are two types of superconvergence used for error estimation:

- *Direct superconvergence*: We obtain the superconvergent values from direct evaluation of the approximate solution at all the chosen superconvergent points.
- *Superconvergence via averaging*: We obtain superconvergent values by local averaging the approximate solution.

Because we need an exact conformal mesh on an arbitrary domain with potential complex geometry, it affects the location of superconvergent points strongly. They are very sensitive to the mesh geometry.

Definition 11 ($\eta\%$-superconvergence [19, 15])**.** *Let u be the exact solution, and $\{u_h\}$ is a sequence of FEM-solutions computed on $M = \{\mathcal{M}_h\}$. For the linear functional $F(u)$ and each $K \in \mathcal{M}_h$, a geometry dependent point $\overline{x}$ is given. Define the function*

$$\Psi_K(u - u_h) = \max_{x \in K} |F(u - u_h)(x)| \tag{112}$$

We define the relative error in $F(u)$ at a chosen mesh point $\overline{x}$ as

$$\Theta(\overline{x}; F; u, u_h; h, K) = \begin{cases} \frac{|F(u-u_h)(\overline{x})|}{\Psi_K(u-u_h)} & , \quad \text{if } \Psi_K(u - u_h) \neq 0 \\ 0 & , \quad \text{if } \Psi_K(u - u_h) = 0 \end{cases} \tag{113}$$

We call $\overline{x}$ a u-$\eta\%$-superconvergence point relative to u and M if

$$\lim_{h \to 0} \Theta(\overline{x}; F; u, u_h; h, K) \leq 0.01\eta \tag{114}$$

Let $\mathcal{U}$ be a class of exact solutions. The definition above allows us to introduce several other quantities of interest:

$\eta\%$-contour of $F(u)$ in $K \in \mathcal{K}$ for u:

$$\mathscr{C}^{\eta\%}_{F(u)}(u; K, \mathcal{M}_h) = \{x \in K : \Theta(x; F; u, u_h; h, K) = 0.01\eta\}$$

$\eta\%$-band of $F(u)$ in $K \in \mathcal{K}$ for u:

$$\mathscr{B}^{\eta\%}_{F(u)}(u; K, \mathcal{M}_h) = \{x \in K : \Theta(x; F; u, u_h; h, K) < 0.01\eta\}$$

Superconvergence points of $F(u)$ in $K \in \mathcal{K}$ for $\mathcal{U}$:

$$\mathscr{X}^{\text{sup}}_{F(u)}(\mathcal{U}; K, \mathcal{M}_h) = \bigcap_{u \in \mathcal{U}} \mathscr{C}^{\eta\%}_{F(u)}(u; K, \mathcal{M}_h)$$

$\eta\%$-superconvergence regions of $F(u)$ in $K \in \mathcal{K}$ for $\mathcal{U}$:

$$\mathscr{R}^{\eta\%}_{F(u)}(\mathcal{U}; K, \mathcal{M}_h) = \bigcap_{u \in \mathcal{U}} \mathscr{B}^{\eta\%}_{F(u)}(u; K, \mathcal{M}_h)$$

Common $\eta\%$-regions of $F(u)$ in $K \in \mathcal{K}$ for $\mathcal{U}$:

$$\overline{\mathscr{R}}^{\eta\%}_{F(u)}(\mathcal{U}, M; K, \mathcal{M}_h) = \bigcap_{\mathcal{M}_h \in M} \bigcap_{K \in \mathcal{M}} \mathscr{R}^{\eta\%}_{F(u)}(\mathcal{U}; K, \mathcal{M}_h)$$

7.2 Global recovery estimators

We present two global recovery estimators based on least-squares fitting, the so-called *projection* and *variational recovery operators*, originally proposed in [71] and [61]. From now on, we will denote the errors as

$$e = u - u_h \qquad , \qquad e_\sigma = \nabla(u - u_h)$$

We define $\boldsymbol{\sigma}^* = \mathbf{R}\widehat{\mathbf{c}}_\sigma$, where $\mathbf{R}$ is a matrix that corresponds to the functions representing the displacement field, and $\widehat{\mathbf{c}}_\sigma$ is the unknown global vector field of required new control variables.

Continuous L^2-projection (CL2P)

In this approach, we minimize a functional $\mathcal{J}_{L^2}$ with respect to $\widehat{\mathbf{c}}_\sigma$, where

$$\mathcal{J}_{L^2}(\widehat{\mathbf{c}}_\sigma) = \int_\Omega (\boldsymbol{\sigma}^* - \boldsymbol{\sigma}_h)^T (\boldsymbol{\sigma}^* - \boldsymbol{\sigma}_h)\, d\Omega \tag{115}$$

Taking the gradient with respect to $\widehat{\mathbf{c}}_\sigma$ yields a linear system:

$$\left[\int_\Omega \mathbf{R}^T\mathbf{R}\, d\Omega\right] \widehat{\mathbf{c}}_\sigma = \int_\Omega \mathbf{R}^T \boldsymbol{\sigma}_h\, d\Omega$$
$$\mathbf{A}\widehat{\mathbf{c}}_\sigma = \mathbf{b}_\sigma$$

This is global because the field $\boldsymbol{\sigma}^*$ is obtained by projecting the computed gradient components $\boldsymbol{\sigma}_h$ onto the same function space as u_h.

Discrete least-squares fitting (DLSF)

We choose a set of optimal sampling points on each patch, $\{x_k\}_{i=1}^{N_{\text{samp}}}$, such that the functional to be minimized becomes

$$\mathcal{H}_{L^2}(\widehat{\mathbf{c}}_\sigma) = \sum_{i=1}^{N_{\text{samp}}} (\boldsymbol{\sigma}^*(x_k) - \boldsymbol{\sigma}_h(x_k))^2 \tag{116}$$

Since $\boldsymbol{\sigma}^* = \mathbf{R}\widehat{\mathbf{c}}_\sigma$, we take the gradient of $\mathcal{H}_{L^2}$ with respect to $\widehat{\mathbf{c}}_\sigma$ and get

$$\left[\sum_{i=1}^{N_{\text{samp}}} \mathbf{R}(x_k)^T\mathbf{R}(x_k)\right] \widehat{\mathbf{c}}_\sigma = \sum_{i=1}^{N_{\text{samp}}} \mathbf{R}(x_k)^T \boldsymbol{\sigma}_h(x_k)$$
$$\mathbf{A}\widehat{\mathbf{c}}_\sigma = \mathbf{b}_\sigma$$

7.3 The Zienkiewicz-Zhu estimator

A famous recovery method applied in a posteriori error estimation is the *Zienkiewicz-Zhu Superconvergent Patch Recovery* (ZZ-SPR) scheme [95]. For every element $K \in \mathcal{M}$, we construct a recovered gradient σ_K^{ZZ} by following a two-step procedure [20]:

1. Use least-squares fitting on the gradient for vertex-patches of elements
The set $\widetilde{\omega}_K$ from (3) is a patch of vertices. For each patch, recover an averaged gradient $\sigma_{\mathcal{N}}^*$ by solving a constrained minimization problem:

$$\text{Compute} \quad \sigma_{\mathcal{N}}^* \in \mathbb{P}^2(\widetilde{\omega}_K) = \left\{ P : P(x,y) = \sum_{i,j=0}^{2} a_{ij} x^i y^j \right\} \tag{117a}$$

$$\text{Subject to} \quad \sum_{L=1}^{s} |\sigma_{\mathcal{N}}^* - \nabla u_h|^2(\mathbf{x}_l) = \min_{\sigma \in \mathbb{P}^p(\widetilde{\omega}_K)} \sum_{L=1}^{s} |\sigma - \nabla u_h|^2(\mathbf{x}_l) \tag{117b}$$

where $\{\mathbf{x}_l\}_{L=1}^s$ is the set of sampling-points, usually the mapped GL-points in the elements belonging to $\widetilde{\omega}_K$. For rectangular meshes, these sampling-points become the gradient's superconvergence points.

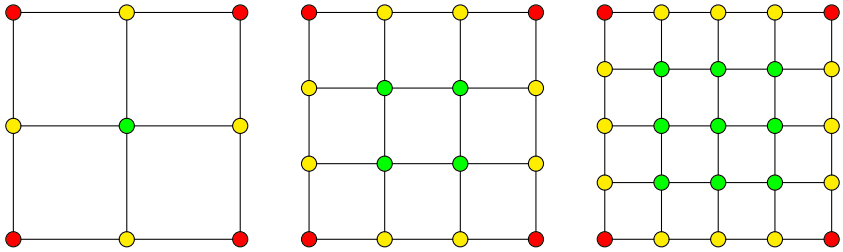
2. Construct the recovered gradient over the element K
On the interior element K, let $\{\mathcal{N}_i^K\}_{i=1}^4$ be the set of its vertices, $\{\widetilde{\omega}_{\mathcal{N}_i^K}\}_{i=1}^4$ is the set of element patches connected to the vertices, and $\{\omega_{\mathcal{N}_i^K}^*\}_{i=1}^4$ are the averaged gradients obtained from the previous stage. We define a recovered C^0-gradient over the entire mesh as the linear finite sum

$$\sigma_K^{ZZ} = \sum_{i}^{N} \alpha_i^K \varphi_i^K \tag{118}$$

where φ_i^K are the shape functions on K, and the coefficients α_i^K are degrees of freedom for the recovered gradient.

According to Babuška et al. [20], there are three types of degrees of freedom for the recovered gradient. We generalize these definitions to any polynomial degree $p \geq 2$ such that we can adapt them to IGA. To do so, we partition the set of indices for the nodes, $\mathcal{I}$, into three mutually disjoint sets:

$\mathcal{I}_1 = \{\text{Nodes at the element vertices.}\}$
$\mathcal{I}_2 = \{\text{Nodes on the element boundary, excluding the vertices.}\}$
$\mathcal{I}_3 = \{\text{Nodes on the element's interior part.}\}$



(a) Polynomial degree 2. **(b)** Polynomial degree 3. **(c)** Polynomial degree 4.

Figure 25. The recovered gradient's degrees of freedom in 2D: Vertex degree of freedom (•), edge degree of freedom (•), and internal degree of freedom (•).

Now, we can classify the degrees of freedom on the recovered gradient:

a) *Vertex degree of freedom*. These coefficients are found directly from

$$\alpha_i^K = \sigma^*_{\mathcal{N}_i^K}(\mathcal{N}_i^K) \quad , \quad i \in \mathcal{I}_1 \tag{119}$$

b) *Edge degree of freedom*. These coefficients are found by solving

$$\frac{\partial I_i}{\partial \alpha}(\alpha_i^K) = 0 \quad , \quad i \in \mathcal{I}_2 \tag{120a}$$

$$I_i(\alpha) = \left\| \alpha \varphi_i^K + \sum_{j \in \mathcal{I}_2^{(i)}} \left(\alpha_j^K \varphi_j^K - \frac{1}{2} \sigma^*_{X_j^K} \right) \right\|_{L^2(V_i)} \tag{120b}$$

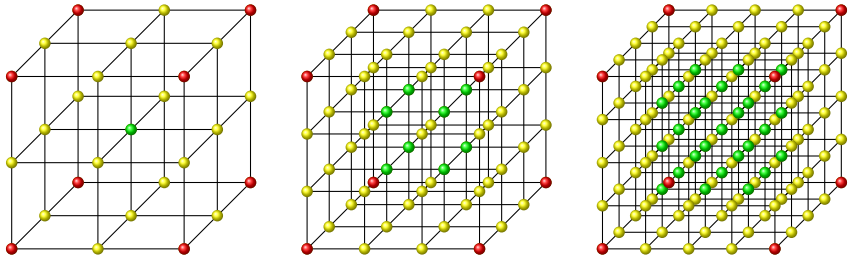
c) *Internal degree of freedom*. These last quantities are determined from

$$\frac{\partial I_i}{\partial \alpha}(\alpha_i^K) = 0 \quad , \quad i \in \mathcal{I}_3 \tag{121a}$$

$$I_i(\alpha) = \left\| \alpha \varphi_i^K + \sum_{j \in \mathcal{I}_3^{(i)}} \left(\alpha_j^K \varphi_j^K - \frac{1}{4} \sigma^*_{X_j^K} \right) \right\|_{L^2(W_i)} \tag{121b}$$

In this setting, V_i and W_i are patches enclosed by the adjacent nodes of node i, and V_i consists of boundary nodes. We have used two other auxiliary sets:

$$\mathcal{I}_2^{(i)} = \{\text{Nodes on the element boundary adjacent to node } i.\}$$
$$\mathcal{I}_3^{(i)} = \{\text{Nodes diagonally adjacent to node } i.\}$$



(a) Polynomial degree 2. **(b)** Polynomial degree 3. **(c)** Polynomial degree 4.

Figure 26. The recovered gradient's degrees of freedom in 3D: Vertex degree of freedom (•), face degree of freedom (•), and internal degree of freedom (•).

The procedure of calculating the Zienkiewicz-Zhu estimator (118) by the three-stage algorithm described above can be extended from two to three dimensions. A slight difference in this case is that the sets $\mathcal{I}_2$ and $\mathcal{I}_2^{(i)}$ are defined for nodes on a face, not just an edge. Furthermore, the functional I_i must also be modified because we are taking more nodes into account. For the face and internal degrees of freedom, we get the following formulas:

$$I_i(\alpha) = \left\| \alpha\varphi_i^K + \sum_{j\in\mathcal{I}_2^{(i)}} \left(\alpha_j^K \varphi_j^K - \frac{1}{\delta}\sigma^*_{X_j^K} \right) \right\|_{L^2(V_i)} \qquad i \in \mathcal{I}_2 \tag{122a}$$

$$\delta = \begin{cases} 5 & \text{if } i \text{ is on the edge of a face} \\ 8 & \text{if } i \text{ is not on the edge of a face} \end{cases}$$

$$I_i(\alpha) = \left\| \alpha\varphi_i^K + \sum_{j\in\mathcal{I}_3^{(i)}} \left(\alpha_j^K \varphi_j^K - \frac{1}{8}\sigma^*_{X_j^K} \right) \right\|_{L^2(W_i)} \qquad i \in \mathcal{I}_3 \tag{122b}$$

7.4 General SPR procedure

As we have seen until now, the original idea behind SPR is improving the gradient of the FE-solution at nodal points, so we must define an *element patch* consisting of all the elements connected to a given node. This can be done separately for each gradient component. We construct a global polynomial from the basis function's monomials, and then calculate its coefficients in such a way that it will match the gradient component optimally at the patch's reduced integration points. The improved gradient follows by direct evaluation of this new polynomial.

In [67], Kumar et al. provide a computer-based proof for the existence of superconvergent points in context of IGA, and a three-step SPR-procedure:

1) *Patch recovery*. Let $\sigma_d^* = \mathbf{P}(\mathbf{x})\mathbf{a}_d$, such that $d \in \{x, y, z\}$, $\mathbf{P}$ is a matrix of monomials, and $\mathbf{a}_d$ is found from least-squares fitting of σ_d^* to the values of σ_d^h at the sampling points $\{\mathbf{x}_i\}$. Thus, we minimize

$$\mathcal{F}(\mathbf{a}_d) = \sum_{i=1}^{n_{sp}^{el}} (\sigma_{d,i}^* - \sigma_{d,i}^h)^T (\sigma_{d,i}^* - \sigma_{d,i}^h)$$

By invoking the stationary criterion, we must solve $\mathbf{D}\mathbf{a}_d = \mathbf{G}$, where

$$\mathbf{D} = \sum_{i=1}^{n_{sp}^{el}} \mathbf{P}_i(\mathbf{x}_i)^T \mathbf{P}_i(\mathbf{x}_i) \quad , \quad \mathbf{G} = \sum_{i=1}^{n_{sp}^{el}} \mathbf{P}_i(\mathbf{x}_i)^T \sigma_{d,i}^h$$

2) *Patch configuration*. The interior patch consists of all elements that belong to the given basis function's support. For the boundary patch, lacking enough elements for the discrete least-squares fit, there are two alternatives:

   (a) Extending the domain of element patches.

   (b) Using recovery on the regular patch for that basis function.

3) *Global recovery*. We conjoin the polynomial expansions $\sigma^* = \mathbf{P}\mathbf{a}$ for every patch containing the actual element using the basis as weighting function. If R_A is the solution's basis function and σ_A^* is the local recovered gradient, we can apply partition of unity to obtain

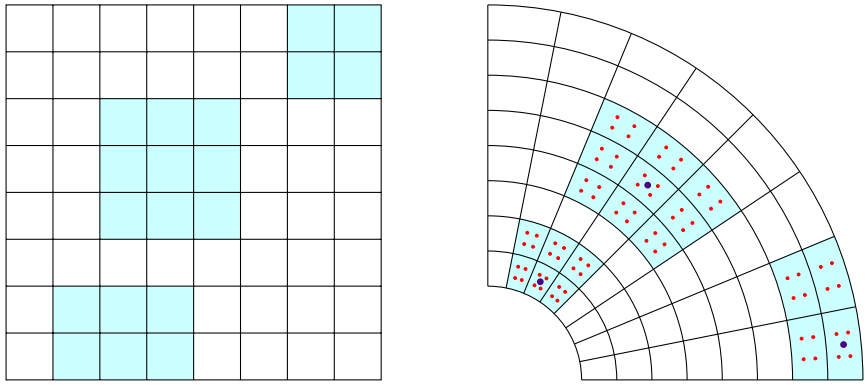
$$\sigma^*(\mathbf{x}) = \sum_{\forall A} \sigma_A^* R_A(\mathbf{x})$$

This SPR-approach for IGA is a generalization of the methodology described in [32, 93]. It has been verified numerically that IGA-SPR works for linear problems and yields excellent results, i.e. good effectivity indices [67]. The sampling points $\{\mathbf{x}_i\}$ must satisfy the consistency condition (R1), so the natural choice is *Barlow points* [26]. According to Zienkiewicz and Zhu [95], the algorithm above is easy to implement, but we should take some things into account to make it work properly:

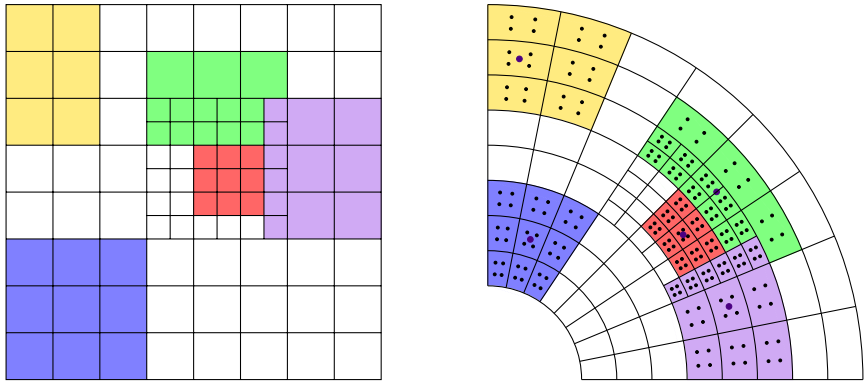
i) The linear equation system $\mathbf{D}\mathbf{a}_d = \mathbf{G}$ is solved component-wise.

ii) Local normalized coordinates ensure that $\mathbf{D}$ becomes well-conditioned.

iii) If $p \geq 2$, the internal nodes should be chosen as the average of other internal nodes from several patches.

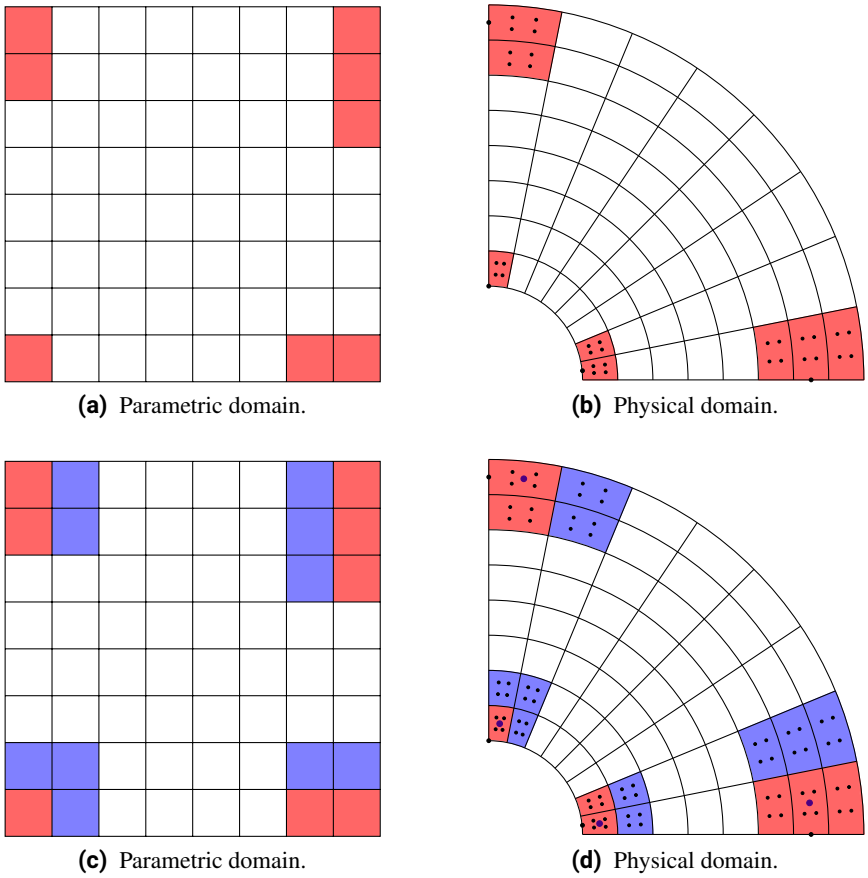


(a) Parametric domain. **(b)** Physical domain.



(c) Parametric domain. **(d)** Physical domain.

Figure 27. Regular element patch configuration: In the first row, we have element patches for the support of quadratic B-splines and NURBS, which have tensor product structure. In the second row, we have a general LR mesh with support of quadratic LR B-splines.



(a) Parametric domain. **(b)** Physical domain.

(c) Parametric domain. **(d)** Physical domain.

Figure 28. Boundary element patch configuration: The first row represents element patches for the support of quadratic B-splines and NURBS at the boundary. In the second row, we have a general LR mesh with support of quadratic LR B-splines, and the boundary patch is extended.

8 General theory of adaptive refinement

In this final section, we present the adaptive refinement theory in a more abstract way and present a general subdivision procedure.

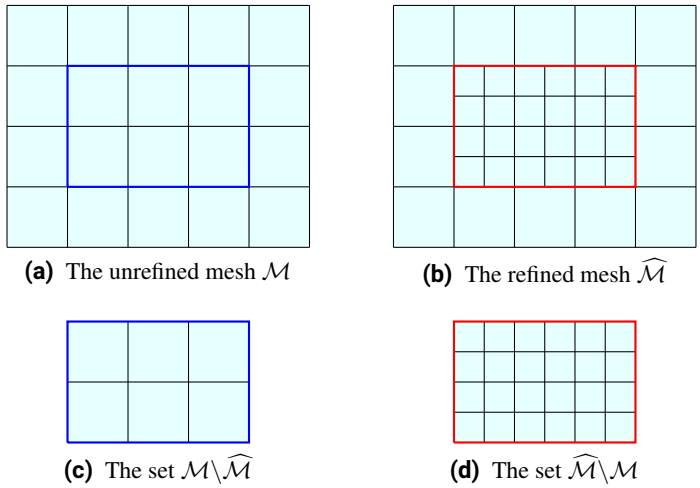
8.1 Theoretical background

As before, we assume that the trial and test spaces coincide as V (Bubnov-Galerkin discretization), $\mathcal{M}$ is a shape-regular mesh on the domain Ω, and $V(\mathcal{M})$ is the discrete finite-dimensional space on $\mathcal{M}$. The space $V \cup V(\mathcal{M})$ has a quasi-metric d satisfying three properties, for any element u, v and w:

$$\begin{aligned}
&\textit{Non-negativity}: && d[\widehat{\mathcal{M}}; u, v] \geq 0 \\
&\textit{Quasi-symmetry}: && d[\widehat{\mathcal{M}}; u, v] \geq C_\Delta d[\widehat{\mathcal{M}}; v, u] \\
&\textit{Quasi triangle inequality}: && C_\Delta^{-1} d[\widehat{\mathcal{M}}; u, v] \leq d[\widehat{\mathcal{M}}; u, w] + d[\widehat{\mathcal{M}}; w, v]
\end{aligned}$$

If V is a Banach space, then d is reduced to a regular norm. The local error contributions can be characterized as follows:

$$\begin{aligned}
&\eta_K(\mathcal{M}; \cdot) : V(\mathcal{M}) \longmapsto \mathbb{R}^+ && K \in \mathcal{M} \\
&\eta(\mathcal{M}; u_h)^2 = \sum_{K \in \mathcal{M}} \eta_K(\mathcal{M}; u_h)^2 && u_h \in V(\mathcal{M})
\end{aligned}$$



(a) The unrefined mesh $\mathcal{M}$

(b) The refined mesh $\widehat{\mathcal{M}}$

(c) The set $\mathcal{M} \backslash \widehat{\mathcal{M}}$

(d) The set $\widehat{\mathcal{M}} \backslash \mathcal{M}$

Figure 29. Visualization of the sets $\mathcal{M} \backslash \widehat{\mathcal{M}}$ (—) and $\widehat{\mathcal{M}} \backslash \mathcal{M}$ (—).

To ensure proper adaptive refinement, it is necessary to invoke a special *compatibility condition* [38], which holds in many applications.

Definition 12 (Adaptive refinement compatibility)**.** *For any refinement* $\widehat{\mathcal{M}}$ *of* $\mathcal{M}$*, we assume that* $d[\widehat{\mathcal{M}};\cdot,\cdot]$ *is well-defined on* $V \cup V(\mathcal{M}) \cup V(\widehat{\mathcal{M}})$ *with* $d[\widehat{\mathcal{M}};u,u_h] = d[\mathcal{M};u,u_h]$*. Then, we have*

$$d[\widehat{\mathcal{M}};u,u_h] \leq \epsilon \qquad , \qquad \forall \epsilon > 0$$

We denote the *set of admissible meshes*, and its subset with at most N elements more than the initial mesh, as

$$\mathbb{M} = \{\mathcal{M} : \mathcal{M} \text{ is an admissible refinement of } \mathcal{M}_0\} \tag{123a}$$

$$\mathbb{M}(N) = \{\mathcal{M} \in \mathbb{M} : |\mathcal{M}| - |\mathcal{M}_0| \leq N\} \tag{123b}$$

The adaptive mesh-refinement works if we satisfy the estimates below:

$$|\widehat{\mathcal{M}} \backslash \mathcal{M}| \leq |\widehat{\mathcal{M}}| - |\mathcal{M}| \tag{124a}$$

$$|\mathcal{M}_{l+1}| \leq C_s |\mathcal{M}_l| \tag{124b}$$

$$|\mathcal{M}_l| - |\mathcal{M}_0| \leq C_m \sum_{k=0}^{l-1} |\mathcal{M}_k| \tag{124c}$$

$$|\mathcal{M} \oplus \mathcal{M}'| \leq |\mathcal{M}| + |\mathcal{M}'| - |\mathcal{M}_0| \tag{124d}$$

Here, $|\cdot|$ is a counting-measure for the mesh cardinality (number of mesh elements). The two first estimates follow from splitting each element into C_s elements ($C_s \geq 2$). The third estimate represents closure and holds for $C_m > 0$, depending on $\mathbb{M}$. Lastly, for any two meshes $\mathcal{M}, \mathcal{M}' \in \mathbb{M}$, there is a common coarsest mesh $\mathcal{M} \oplus \mathcal{M}' \in \mathbb{M}$ satisfying (124d).

The adaptivity axioms [38] provide an abstract framework independent of the specific PDE to be solved and the chosen type of basis functions used for the approximation. If they are satisfied, then the general algorithm for adaptive mesh-refinement will converge quasi-optimally. For these axioms, we assume that $C_{\text{stab}}, C_{\text{red}}, C_{\text{osc}}, C_{\text{drel}}, C_{\text{ref}}, C_{\text{qo}}(\epsilon_{\text{stab}}) \geq 1$ and $\rho_{\text{red}} \in (0,1)$ are auxiliary constants just depending on the set $\mathbb{M}$.

1. *Stability on non-refined elements*. If $\mathcal{M}$ is the initial mesh, $\widehat{\mathcal{M}}$ is a refined mesh, $u \in V(\mathcal{M})$ and $\widehat{u}_h \in V(\widehat{\mathcal{M}})$ are discrete solutions, and $S \subseteq \mathcal{M} \cap \widehat{\mathcal{M}}$ consists of unrefined elements, we have the bound

$$\left| \sqrt{\sum_{K \in S} \eta_K(\widehat{\mathcal{M}};\widehat{u}_h)^2} - \sqrt{\sum_{K \in S} \eta_K(\mathcal{M};u_h)^2} \right| \leq C_{\text{stab}} d[\widehat{\mathcal{M}};\widehat{u}_h,u_h] \tag{A1}$$

2. *Reduction property on refined elements*. For the parameter $\rho_{\text{red}} \in (0,1)$ and the set $\mathcal{W} = \widehat{\mathcal{M}} \backslash \mathcal{M}$, we have the bound

$$\sum_{K \in \mathcal{W}} \eta_K(\widehat{\mathcal{M}}; \widehat{u}_h)^2 \leq \sum_{K \in \mathcal{W}} \rho_{\text{red}} \eta_K(\mathcal{M}; u_h)^2 + C_{\text{red}} d[\widehat{\mathcal{M}}; \widehat{u}_h, u_h]^2 \tag{A2}$$

3. *General quasi-orthogonality*. If $\epsilon_{\text{qo}} \in [0, \epsilon^*_{\text{qo}})$ and $N \geq l$, we can generalize Pythagoras's theorem as follows:

$$\sum_{k=l}^{N} (d[\mathcal{M}_{k+1}; u_h^{(k+1)}, u_h^{(k)}]^2 - \epsilon_{\text{qo}} d[\mathcal{M}_k; u, u_h^{(k)}]^2) \leq C_{\text{qo}}(\epsilon_{\text{qo}}) \eta(\mathcal{M}_l; u_h(\mathcal{M}_l))^2 \tag{A3}$$

4. *Discrete reliability*. Let $\mathcal{R}$ be a set containing up to a multiplicative constant the same number of elements as $\mathcal{M} \backslash \widehat{\mathcal{M}}$, e.g. $\mathcal{R}$ contains $\mathcal{M} \backslash \widehat{\mathcal{M}}$ and an additional element layer. The quasi-metric satisfies

$$d[\widehat{\mathcal{M}}; \widehat{u}_h, u_h]^2 \leq C_{\text{drel}} \left(\sum_{K \in \mathcal{R}(\mathcal{M}, \widehat{\mathcal{M}})} \eta_K(\mathcal{M}; u_h)^2 \right) \tag{A4}$$

The first axiom (A1), providing algorithm convergence, can be verified with the triangle inequality and inverse estimates. The second (A2) originates from observations that the error estimators' contributions are weighted by local mesh-size and decrease uniformly on refined elements.

If V is a Hilbert space, then (A3) reduces to standard orthogonality because d becomes a normal metric. Using quasi-orthogonality might be beneficial for nonsymmetric operators, nonconforming discretization, or mixed problems. It can be shown that discrete reliability (A4) implies continuous reliability. This is a very important property of error estimators:

$$d[\mathcal{M}; u, u_h] \leq C_{\text{rel}} \eta(\mathcal{M}; u_h) \tag{125}$$

If $\theta < (1 + C_{\text{stab}}^2 C_{\text{rel}}^2)^{-1}$, then the optimal value of ϵ^*_{qo} satisfies

$$\epsilon^*_{\text{qo}} \geq \frac{\theta^2 (1 - \rho_{\text{red}})^2 C_{\text{stab}}^2}{2 C_{\text{rel}}^2 (C_{\text{red}} + C_{\text{stab}}^2)^2} \tag{126}$$

Theorem 19 (Adaptive refinement convergence [38]). *Axioms* (A1)-(A4) *imply linear convergence of adaptive refinement. If* $C_{conv} > 0$ *and* $\rho_{conv} \in (0, 1)$*, then the following inequality holds:*

$$\eta(\mathcal{M}_{k+l}; u_h^{(k+l)})^2 \leq C_{conv}\rho_{conv}^k \eta(\mathcal{M}_l; u_h^{(l)})^2 \quad , \quad \forall k, l \in \mathbb{Z}^+ \tag{127}$$

We also have a best possible algebraic convergence order for $s > 0$*:*

$$\begin{aligned} \|(\eta(\cdot), u_h(\cdot))\|_{\mathbb{B}_s} &= \sup_{N\in\mathbb{Z}^+} \inf_{|\mathcal{M}|-|\mathcal{M}_0|\leq N} \eta(\mathcal{M}; u_h)(N+1)^s \\ &\simeq \sup_{l\in\mathbb{Z}^+} \eta(\mathcal{M}_l; u_h^{(l)})(|\mathcal{M}_l| - |\mathcal{M}_0| + 1)^s \end{aligned} \tag{128}$$

Reliability means that the true error approaches zero when the computable error, estimated by residuals, also approaches zero by adaptive refinement. This is crucial for sufficient convergence rate. Efficiency of the estimator prevents overestimation of error up to some oscillation. We have

$$d[\mathcal{M}; u, u_h] \leq C_{\text{rel}}\eta(\mathcal{M}, u_h) \tag{129a}$$

$$C_{\text{eff}}^{-1}\eta(\mathcal{M}, u_h) \leq d[\mathcal{M}; u, u_h] + \text{osc}(\mathcal{M}; u_h) \tag{129b}$$

This expression relates quasi-optimal estimator convergence with the true error's convergence rate, and it includes oscillation. When the estimator is both reliable and efficient, then the estimated error decays asymptotically in the same way as the true error. It has been demonstrated that for conforming discretizations, the approximation sequence will always converge [17].

Theorem 20 (Quasi-monotonicity [38]). *Axioms* (A1), (A2) *and* (A4) *imply that the error estimator* η *is quasi-monotone, i.e. there is a* C_{mon} *such that*

$$\eta(\widehat{\mathcal{M}}; U(\widehat{\mathcal{M}}) \leq C_{mon}\eta(\mathcal{M}; U(\mathcal{M}) \tag{130}$$

Theorem 21 (Convergence of error estimator [38]). *Assume that the four axioms* (A1)-(A4) *are satisfied. If the error is quasi-monotone, and*

$$\|\operatorname{osc}(\cdot)\|_{\mathbb{O}_s} = \sup_{N\in\mathbb{N}} \inf_{|\mathcal{M}|-|\mathcal{M}_0|\leq N} \text{osc}(\mathcal{M}; u_h)(N+1)^s < \infty$$

then the approximation ability $\|\cdot\|_{\mathbb{B}_s}$ *can be characterized by*

$$\|(u, u_h(\cdot))\|_{\mathbb{A}_s} = \sup_{N\in\mathbb{N}} \inf_{|\mathcal{M}|-|\mathcal{M}_0|\leq N} d[\mathcal{M}; u, u_h](N+1)^s \tag{131a}$$

$$\|(u, u_h(\cdot))\|_{\mathbb{A}_s} \simeq \sup_{l\in\mathbb{Z}^+} \frac{d[\mathcal{M}_l; u, u_h^l]}{(|\mathcal{M}_l| - |\mathcal{M}_0| + 1)^{-s}} \tag{131b}$$

8.2 Marking techniques

Any type of adaptive refinement requires some essential features:

1. A suitable discretization method.
2. Efficient solver for the discrete problems.
3. An appropriate a posteriori error estimator.
4. An effective refinement strategy for detecting elements.

Maximum and equilibration strategies

A central part of the refinement routine is determining elements to be refined. For each step k, we take a partition $\mathcal{M}_k$ of the domain Ω containing the current elements used for approximating the unknown solution. Then we determine a subset $\widetilde{\mathcal{M}}_k \subseteq \mathcal{M}_k$ with those elements to be refined. This yields a new partition $\mathcal{M}_{k+1}$ which can be refined again if necessary. Originally developed for classical FEM, this procedure is also applicable for IGA. In our case, it is appropriate to split the basis functions instead of the elements to improve the refinement.

First, we will focus on two well-known procedures: the *Maximum strategy* and the *Equilibration strategy* (Dörfler) [46, 88]. Both of them require the same input: a partition $\mathcal{M}_k$, a set of error indicators $\{\eta_K\}_{K\in\mathcal{M}_k}$, and a bulk parameter $\beta \in (0,1)$. The Maximum strategy is cheaper than the equilibration strategy. If β is high, few elements are marked, and if β is low, many elements are marked. This effect is reversed when using equilibration. To obtain a certain equilibrium, we can choose $\beta \approx 0.5$ such that the number of marked elements becomes balanced.

Algorithm 8.1 Equilibration Strategy (β-EQU)

procedure EQUILIBRATION_STRATEGY($\mathcal{M}_k$, $\{\eta_K\}_{K\in\mathcal{M}_k}$, β)
  Set $\widetilde{\mathcal{M}}_k = \emptyset$ and $\Sigma_{\mathcal{M}_k} = 0$
  Compute $\Theta_{\mathcal{M}_k} = \sum_{K\in\mathcal{M}_k} \eta_K^2$
  while $\Sigma_{\mathcal{M}_k} < \beta\Theta_{\mathcal{M}_k}$ **do**
    Compute $\widetilde{\eta}_{\max} = \max\{\eta_K \,:\, K \in \mathcal{M}_k \backslash \widetilde{\mathcal{M}}_k\}$
    for $K \in \mathcal{M}_k \backslash \widetilde{\mathcal{M}}_k$ **do**
      if $\eta_K = \widetilde{\eta}_{\max}$ **then**
        Store K in $\widetilde{\mathcal{M}}_k$
        Add η_K^2 to $\Sigma_{\mathcal{M}_k}$
  return $\widetilde{\mathcal{M}}_k$

Algorithm 8.2 Maximum Strategy (β-MAX)

procedure MAXIMUM_STRATEGY($\mathcal{M}_k$, $\{\eta_K\}_{K\in\mathcal{M}_k}$, β)
  Set $\widetilde{\mathcal{M}}_k = \emptyset$
  Compute $\eta_{\max} = \max\{\eta_K \,:\, K \in \mathcal{M}_k\}$
  for $\eta_K \in \mathcal{M}_k$ **do**
    if $\eta_K \geq \beta\eta_{\max}$ **then**
      Mark K for refinement and store it in $\widetilde{\mathcal{M}}_k$
  return $\widetilde{\mathcal{M}}_k$

Both the maximum and equilibration strategies minimize the cardinality of $\widetilde{\mathcal{M}}_k$, but for the equilibration strategy, we also have the estimate

$$\sum_{K\in\widetilde{\mathcal{M}}_k} \eta_K^2 \geq \beta \sum_{K\in\mathcal{M}_k} \eta_K^2 \tag{132}$$

Maximal error method

The *Maximal error method*, developed for classical FEM, is looping over every element, computing their individual error estimates, sorting them, and then choose $\beta\%$ of those elements which possess the highest estimated error. To adapt this method to IGA, it is preferable to split basis functions, not elements. Thus, we define $\operatorname{supp}(N_i) = \mathcal{M}(N_i)$, where N_i is a B-spline, and $\|e\|_{E(K)}$ is energy error on K. The *B-spline error* [63] is

$$\|e\|^2_{\mathcal{M}(N_i)} = \sum_{K\in\mathcal{M}_k(N_i)} \|e\|^2_{E(K)} \tag{133}$$

Algorithm 8.3 Maximal Error Method (β-N_{dof})

procedure MAXIMAL_ERROR_METHOD($\mathcal{M}_k$, β)
  Set $\widetilde{\mathcal{M}}_k = \emptyset$
  Create an array T of size $2 \times N_{\text{dof}}$
  for $N_i \in N_{\text{dof}}$ **do**
    Compute $\|e\|_{\mathcal{M}_k(N_i)}$
    Store $\|e\|_{\mathcal{M}_k(N_i)}$ and N_i in T
  Sort the errors in decreasing order, with their corresponding B-splines
  Refine $\beta\%$ of the first B-splines in T
  Update N_{dof} and $\widetilde{\mathcal{M}}_k$
  return $\widetilde{\mathcal{M}}_k$

Algorithm 8.4 Symmetric Maximal Error Method (adjusted β-N_{dof})

procedure SYMMETRIC_MAXIMAL_ERROR_METHOD($\mathcal{M}_k$, β)
  Set $\widetilde{\mathcal{M}}_k = \emptyset$
  Create an array T of size $2 \times N_{\text{dof}}$
  for $N_i \in N_{\text{dof}}$ **do**
    Compute $\|e\|_{\mathcal{M}_k(N_i)} = \sqrt{\sum_{K \in \mathcal{M}_k(N_i)} \|e\|^2_{E(K)}}$
    Store $\|e\|_{\mathcal{M}_k(N_i)}$ and N_i in T
  Sort the errors in decreasing order, with their corresponding B-splines
  Choose $\beta\%$ of the first B-splines in T.
  Define η_{crit} as the error of the last element to be refined
  Refine every element whose error is less than or equal to η_{crit}
  Update N_{dof} and $\widetilde{\mathcal{M}}_k$
  return $\widetilde{\mathcal{M}}_k$

Preservation of symmetry

If a PDE is expressed by a symmetric differential operator, it might happen that solving it numerically with adaptive refinement will produce a non-symmetric mesh. To illustrate this, we focus on the β-N_{dof} method. There is an improved version of this technique called the adjusted β-N_{dof} method, which ensures that every refined mesh remains symmetric. In this way, it preserves the underlying property of the differential operator.

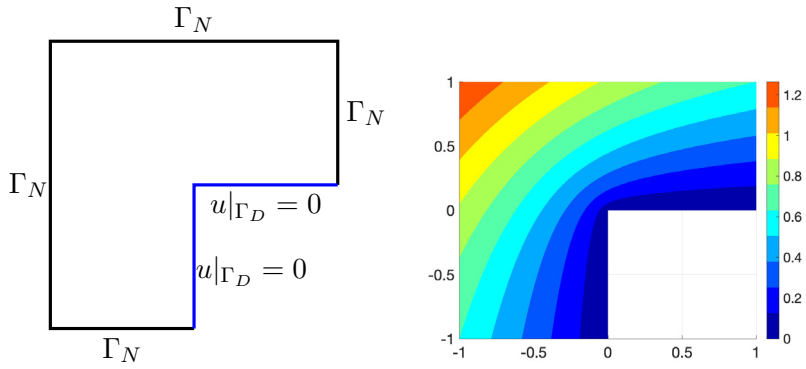
Although we have marked the elements correctly by the β-N_{dof} method, there might be a certain risk that the mesh contains some elements with the same error as the last element to be refined. When these elements are not refined, the new mesh loses its symmetry although the underlying differential operator is symmetric. But if these elements are refined in addition to the original ones that we subdivided previously, the symmetry of the mesh is preserved for every new refinement.

To illustrate the advantages of the adjusted β-N_{dof} method, we consider a well-known BVP for Poisson's equation called the L-shape problem. The domain is defined as $\Omega = (-1,1)^2 \backslash (0,1) \times (-1,0)$, and the BVP is

$$\boxed{\begin{aligned} -\nabla^2 u &= f, \quad u \in \Omega \\ u &= u|_{\partial\Omega}, \quad u \in \partial\Omega \end{aligned}} \tag{134a, 134b}$$

In this setting, there is no source ($f = 0$), and the analytical solution is

$$u(r,\theta) = r^{2/3} \sin\left(\frac{2\theta}{3}\right) \tag{135}$$



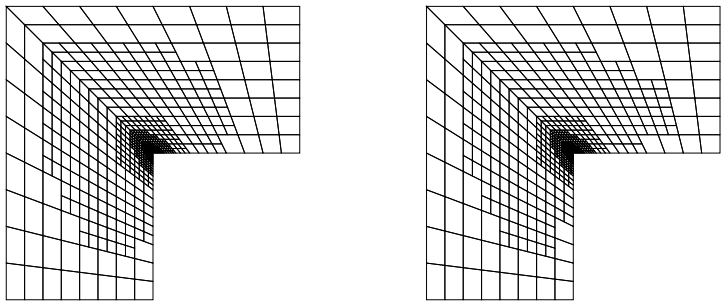
(a) L-shaped problem description. **(b)** Exact solution u.

Figure 30. Illustration of the L-shape problem for Poisson's equation (134).

Now, we perform two simulations, both with and without symmetrization. We use the recovery estimator and choose $\beta = 10\%$. There are three stopping criterions:

- The simulation stops if the estimated error is below 10^{-8}.
- The number of degrees of freedom is not more than 10000.
- The number of refinements do not exceed 30.

In this simulation with polynomial degree 2, we see that the meshes obtained after 5 steps with and without symmetrization are not the same. This is demonstrating that the adjusted β-N_{dof} method preserves the symmetry of the differential operator in Poisson's equation (134).



(a) Unsymmetric mesh. **(b)** Symmetric mesh.

Figure 31. Different meshes for the L-shape problem.

Algorithm 8.5 Averaging Strategy (β-AVG)

procedure AVERAGING_STRATEGY($\mathcal{M}$, η, β, $k_{\max}$)
  Initialize a tensor-mesh $\overline{\mathcal{M}}$, and let $k = 0$
  while $k \leq k_{\max}$ **do**
    Compute $\|e\|_{\mathcal{M}}$
    for $K \in \mathcal{M}$ **do**
      if $\|e\|_{E(K)} > \beta \|e\|_{\mathcal{M}}$ **then**
        Refine K
    Update N_{dof} and $\overline{\mathcal{M}}$
    $k = k + 1$
  return $\overline{\mathcal{M}}$

Averaging strategy

The *Averaging strategy* is a suitable method for PDEs where the symmetry might influence the outcome of the adaptive refinement. After solving the PDE on a given mesh, we compute the estimated average error. Then we loop over the mesh again and refine all the elements whose estimated error exceeds β % of this estimated average error. This will continue within a predefined number of iteration steps.

Special strategies

If a PDE is convection-dominated, then Algorithms 8.1 and 8.2 may not work properly. In this case, the elements on $\mathcal{M}$ are categorized as follows:

- Very few elements with extremely large estimated error.
- A big majority with extremely small estimated error.
- A medium group with reasonably estimated error.

Only the elements belonging to the first very small group will be refined, and the adaptive refinement deteriorates. We can overcome this defect by doing a small modification that works for both marking strategies. In addition to the threshold θ, we define a *very* small percentage β (≤ 10), mark the first $\beta\%$ elements with largest estimated error, and then apply Algorithm 8.1 and 8.2. For unsteady PDEs, where the solution changes over time, we require some additional features:

1. Adaptive mesh refinement at every time-step.
2. Coupling of time-step control and spatial refinement.
3. Partial coarsening if necessary.
4. Re-meshing if necessary.

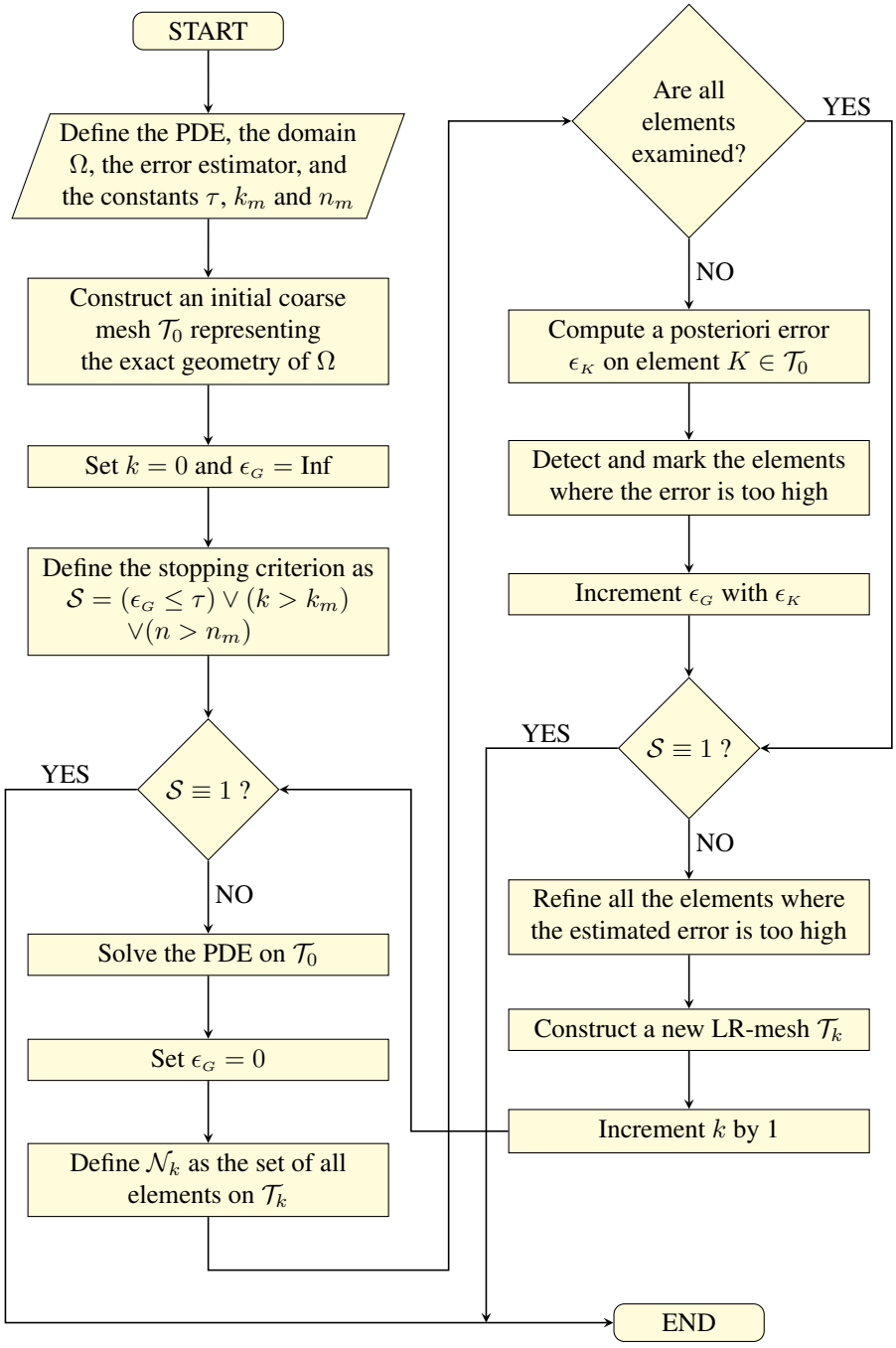


Figure 32. A rough flowchart for adaptive finite element refinement.

9 Conclusion

We have seen that the established theory of adaptive refinement in classical finite element modelling can be transferred directly to isogeometric analysis, equipped with some extra advantages that were not available in the previous paradigm, like high continuity, exact meshing on the domain, and reduced computational effort. Such benefits improve the refinement and makes the numerical solution converge faster to the analytical solution compared with the old finite element method.

The residual estimator works better for isogeometric discretization because there are no jumps in the gradient of the numerical solution when the continuity is at least C^1. This advantage speeds up the computation of the estimator on each element. The combination of tensor B-splines with hierarchical refinement (structured mesh refinement) and LR B-splines gives good recovery estimators.

Most of the error estimators still satisfy the same inequality bounds as before, but the main difference is that the generic constants occurring in these inequalities are significantly lower compared with classical finite element modelling. This new observation indicates that we do not need to refine so many elements before the estimated global error reaches the desired tolerance.

Lastly, we see that IGA offers more refinement techniques that were not available in the past, like Serendipity pairing and hierarchical refinement. We can therefore conclude that IGA has a good potential for adaptive refinement in modern finite element technology.

Bibliography

[1] M. Ainsworth and A. Craig. “A posteriori error estimators in the finite element method”. In: *Numerische Mathematik* 60 (1992), pp. 429–463.

[2] M. Ainsworth and J. T. Oden. *A Posteriori Error Estimation in Finite Element Analysis*. New York: John Wiley & Sons, Ltd, 2000.

[3] M. Ainsworth and J. T. Oden. “A unified approach to a posteriori error estimation based on element residual methods”. In: *Numerische Mathematik* 65 (1993), pp. 23–50.

[4] J. P. Aubin. “Behavior of the error of the approximate solutions of boundary value problems for linear elliptic operators by Galerkin’s and finite difference methods”. In: *Ann. Scuola Norm. Sup. Pisa* 21 (1967), pp. 599–637.

[5] I. Babuška and A. Miller. “The post-processing approach in the finite element method - Part 1: Calculation of displacements, stresses and other higher derivatives of the displacements”. In: *International Journal for Numerical Methods in Engineering* 20 (1984), pp. 1085–1109.

[6] I. Babuška and A. Miller. “The post-processing approach in the finite element method - Part 2: The calculation of stress intensity factors”. In: *International Journal for Numerical Methods in Engineering* 20 (1984), pp. 1111–1129.

[7] I. Babuška and A. Miller. “The post-processing approach in the finite element method - Part 3: A posteriori error estimates and adaptive mesh selection”. In: *International Journal for Numerical Methods in Engineering* 20 (1984), pp. 2311–2324.

[8] I. Babuška and W. C. Rheinboldt. “A posteriori error analysis of finite element solutions for one dimensional problems”. In: *SIAM Journal of Numerical Analysis* 18 (1981), pp. 565–589.

[9] I. Babuška and W. C. Rheinboldt. “A posteriori error estimates for the finite element method”. In: *International Journal for Numerical Methods in Engineering* 12 (1978), pp. 1597–1615.

[10] I. Babuška and W. C. Rheinboldt. “Adaptive approaches and reliability estimations in finite element analysis”. In: *Computer Methods in Applied Mechanics and Engineering* 17 (1979), pp. 519–540.

[11] I. Babuška and W. C. Rheinboldt. “Analysis of optimal finite element meshes in $\mathbb{R}^1$”. In: *Mathematics of Computation* 33 (1979), pp. 435–463.

[12] I. Babuška and W. C. Rheinboldt. “Error estimates for adaptive finite element computations”. In: *SIAM Journal of Numerical Analysis* 15.4 (1978), pp. 736–754.

[13] I. Babuška, T. Strouboulis, and C. S. Upadhyay. “A model study of the quality of a posteriori error estimators for finite element solutions of linear elliptic problems, with particular reference to the behavior near the boundary”. In: *International Journal for Numerical Methods in Engineering* 40 (1997), pp. 2521–2577.

[14] I. Babuška, T. Strouboulis, and C. S. Upadhyay. “A model study of the quality of a posteriori error estimators for linear elliptic problems. Error estimation in the interior of patchwise uniform grids of triangles”. In: *Computer Methods in Applied Mechanics and Engineering* 114 (1994), pp. 307–378.

[15] I. Babuška, T. Strouboulis, and C. S. Upadhyay. “$\eta\%$-superconvergence of finite element approximations in the interior of general meshes of triangles”. In: *Computer Methods in Applied Mechanics and Engineering* 122 (1995), pp. 273–305.

[16] I. Babuška and M. Suri. “The p and hp versions of the finite element method, basic principles and properties”. In: *SIAM review* 36.4 (1994), pp. 578–632.

[17] I. Babuška and M. Vogelius. “Feedback and adaptive finite element solution of one-dimensional boundary value problems”. In: *Numerische Mathematik* 44.1 (1984), pp. 75–102.

[18] I. Babuška, T. Strouboulis, C. S. Upadhyay, and S. K. Gangaraj. “A model study of element residual estimators for linear elliptic problems: The quality of the estimators in the interior of meshes of triangles and quadrilaterals”. In: *Computers and Structures* 57.6 (1995), pp. 1009–1028.

[19] I. Babuška, T. Strouboulis, S. K. Gangaraj, and C. S. Upadhyay. "$\eta\%$-super-convergence in the interior of locally refined meshes of quadrilaterals: superconvergence of the gradient in finite element solutions of Laplace's and Poisson's equations". In: *Applied Numerical Mathematics* 16 (1994), pp. 3–49.

[20] I. Babuška, T. Strouboulis, S. K. Gangaraj, and C. S. Upadhyay. "Pollution error in the h-version of the finite element method and the local quality of the recovered derivatives". In: *Computer Methods in Applied Mechanics and Engineering* 140 (1997), pp. 1–37.

[21] I. Babuška, C. Upadhyay, S. Gangaraj, and K. Copps. "Validation of a posteriori error estimators by numerical approach". In: *International Journal for Numerical Methods in Engineering* 37.7 (1994), pp. 1073–1123.

[22] I. Babuška and T. Strouboulis. *The Finite Element Method and its Reliability*. UK: Oxford University Press, 2001.

[23] I. Babuška, J. Whiteman, and T. Strouboulis. *Finite Elements. An Introduction to the Method and Error Estimation*. UK: Oxford University Press, 2011.

[24] R. E. Bank. "Hierarchical bases and the finite element method". In: *Acta Numerica* 5 (1996), pp. 1–43.

[25] R. E. Bank and A. Weiser. "Some a posteriori error estimators for elliptic partial differential equations". In: *Mathematics of Computation* 44.170 (1985), pp. 283–301.

[26] J. Barlow. "Optimal stress locations in finite element models". In: *International Journal for Numerical Methods in Engineering* 10 (1976), pp. 243–251.

[27] R. Bartels, J. Beatty, and B. Barsky. *An Introduction to Splines for use in Computer Graphics and Geometric Modeling*. Los Altos, USA: Morgan Kaufmann Publishers, Inc, 1987.

[28] Y. Bazilevs, L. B. da Veiga, J. A. Cottrell, T. J. R. Hughes, and G. Sangalli. "Isogeometric analysis: Approximation, stability and error estimates for h-refined meshes". In: *Mathematical Models and Methods in Applied Sciences* 16 (2006), pp. 1031–1090.

[29] R. Becker and R. Rannacher. "An optimal control approach to a posteriori error estimation in finite element methods". In: *Acta Numerica* 10 (2001), pp. 1–102.

[30] C. Bernardi and V. Girault. "A local regularization operator for triangular and quadrilateral finite elements". In: *SIAM Journal of Numerical Analysis* 35.5 (1998), pp. 1893–1916.

[31] P. K. Bhattacharyya. *Distributions: generalized functions with applications in Sobolev spaces*. Göttingen: Walter de Gruyter GmbH, 2012.

[32] T. Blacker and T. Belytschko. "Superconvergent patch recovery with equilibrium and conjoint interpolant enhancements". In: *International Journal for Numerical Methods in Engineering* 37 (1994), pp. 517–536.

[33] B. D. Bojanov, H. A. Hakopian, and A. A. Sahakian. *Spline Functions and Multivariate Interpolations*. Dordrecht, Netherlands: Springer-Verlag, 1993.

[34] C. de Boor. *A Practical Guide to Splines*. New York, USA: Springer-Verlag, 2013.

[35] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge: Cambridge University Press, 2004.

[36] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. New York: Springer-Verlag, 2008.

[37] A. Buffa, J. Rivas, G. Sangalli, and R. Vázquez. "Isogeometric Discrete Differential Forms in Three Dimensions". In: *SIAM Journal of Numerical Analysis* 49.2 (2011), pp. 818–844.

[38] C. Carstensen, M. Feischl, M. Page, and D. Praetorius. "Axioms of adaptivity". In: *Computers and Mathematics with Applications* 67 (2014), pp. 1195–1253.

[39] P. Ciarlet. *The Finite Element Method for Elliptic Problems*. Philadelphia: Society for Industrial and Applied Mathematics, 2002.

[40] J. A. Cottrell, T. J. R. Hughes, and Y. Bazilevs. *Isogeometric Analysis: Toward Integration of CAD and FEA*. UK: John Wiley & Sons, Ltd, 2009.

[41] D. D'Angella, S. Kollmannsberger, E. Rank, and A. Reali. "Multilevel Bézier extraction for hierarchical local refinement of Isogeometric Analysis". In: *Computer Methods in Applied Mechanics and Engineering* 328 (2018), pp. 147–174.

[42] L. Demkowicz, P. Devloo, and J. T. Oden. "On an h-type mesh refinement strategy based on minimization of interpolation errors". In: *Computer Methods in Applied Mechanics and Engineering* 53 (1985), pp. 67–89.

[43] L. Demkowicz, J. T. Oden, and T. Strouboulis. “Adaptive finite elements for flow problems with moving boundaries. Part 1: Variational principles and a posteriori error estimates”. In: *Computer Methods in Applied Mechanics and Engineering* 46 (1984), pp. 217–251.

[44] L. Demkowicz, J. T. Oden, W. Rachowicz, and O. Hardy. “Toward a universal hp-adaptive finite element strategy. Part 1: Constrained approximation and data structure”. In: *Computer Methods in Applied Mechanics and Engineering* 77 (1989), pp. 79–112.

[45] T. Dokken, T. Lyche, and K. Pettersen. “Polynomial splines over locally refined box-partitions”. In: *Computer Aided Geometric Design* 30 (2013), pp. 331–356.

[46] W. Dörfler. “A convergent adaptive algorithm for Poisson’s equation”. In: *SIAM Journal of Numerical Analysis* 33.3 (1996), pp. 1106–1124.

[47] K. Eriksson, D. Estep, P. Hanson, and C. Johnson. “Introduction to adaptive methods for differential equations”. In: *Acta Numerica* 4 (1995), pp. 105–158.

[48] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*. New York: Springer-Verlag, 2004.

[49] J. A. Evans, Y. Bazilevs, I. Babuška, and T. J. R. Hughes. “n-widths, sup-infs, and optimality ratios for the k-version of the isogeometric finite element method”. In: *Computer Methods in Applied Mechanics and Engineering* 198 (2009), pp. 1726–1741.

[50] L. Evans. *Partial differential equations*. Providence, USA: American Mathematical Society, 2010.

[51] S. Ferraz-Leite, C. Ortner, and D. Praetorius. “Convergence of simple adaptive Galerkin schemes based on h-h/2 error estimators”. In: *Numerische Mathematik* 116.2 (2010), pp. 291–316.

[52] E. M. Garau and R. Vázquez. “Algorithms for the implementation of adaptive isogeometric methods using hierarchical B-splines.pdf”. In: *Applied Numerical Mathematics* 123 (2018), pp. 58–87.

[53] W. Gautschi. *Orthogonal Polynomials, Computation and Approximation*. Oxford: Oxford University Press, 2004.

[54] C. Giannelli, B. Jüttler, and H. Speleers. “Strongly stable bases for adaptively refined multilevel spline spaces”. In: *Adv. Comput. Math.* 40 (2014), pp. 459–490.

[55] C. Giannelli, B. Jüttler, S. K. Kleiss, A. Mantzaflaris, B. Simeon, and J. Špeh. "THB-splines: An effective mathematical technology for adaptive refinement in geometric design and isogeometric analysis". In: *Computer Methods in Applied Mechanics and Engineering* 299 (2016), pp. 337–365.

[56] M. B. Giles and E. Süli. "Adjoint methods for PDEs: a posteriori error analysis and postprocessing by duality". In: *Acta Numerica* 11 (2002), pp. 145–236.

[57] T. Grätcsh and K. J. Bathe. "A posteriori error estimation techniques in practical finite element analysis". In: *Computers and Structures* 83 (2005), pp. 235–265.

[58] G. Grubb. *Distributions and Operators*. New York: Springer-Verlag, 2009.

[59] W. B. H. Prautzsch and M. Paluszny. *Bézier and B-Spline Techniques*. Berlin: Springer-Verlag, 2002.

[60] P. Hennig, M. Kästner, P. Morgenstern, and D. Peterseim. "Adaptive mesh refinement strategies in isogeometric analysis - A computational comparison". In: *Computer Methods in Applied Mechanics and Engineering* 316 (2017), pp. 424–448.

[61] E. Hinton and J. S. Campbell. "Local and global smoothing of discontinuous finite element functions using a least squares method". In: *International Journal for Numerical Methods in Engineering* 8 (1974), pp. 461–480.

[62] T. J. R. Hughes, J. A. Cottrell, and Y. Bazilevs. "Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement". In: *Computer Methods in Applied Mechanics and Engineering* 194 (2005), pp. 4135–4195.

[63] K. A. Johannessen, T. Kvamsdal, and T. Dokken. "Isogeometric analysis using LR B-splines". In: *Computer Methods in Applied Mechanics and Engineering* 269 (2014), pp. 471–514.

[64] T. Kanduč, C. Giannelli, F. Pelosi, and H. Speleers. "Adaptive isogeometric analysis with hierarchical box splines". In: *Computer Methods in Applied Mechanics and Engineering* 316 (2017), pp. 817–838.

[65] E. Kreyszig. *Introductory functional analysis with applications*. USA: John Wiley & Sons, Inc., 1978.

[66] M. Kumar, T. Kvamsdal, and K. A. Johannessen. “Simple a posteriori error estimators in adaptive isogeometric analysis”. In: *Computers and Mathematics with Applications* 70 (2015), pp. 1555–1582.

[67] M. Kumar, T. Kvamsdal, and K. A. Johannessen. “Superconvergent patch recovery and a posteriori error estimation technique in adaptive isogeometric analysis”. In: *Computer Methods in Applied Mechanics and Engineering* 293 (2017), pp. 1086–1156.

[68] P. Ladevèze and D. Leguillon. “Error estimate procedure in the finite element method and applications”. In: *SIAM Journal of Numerical Analysis* 20 (1983), pp. 485–509.

[69] X. Li and M. A. Scott. “Analysis-suitable T-splines: Characterization, refineability, and approximation”. In: *Mathematical Models and Methods in Applied Sciences* 24 (2014), pp. 1141–1164.

[70] G. Micula and S. Micula. *Handbook of Splines*. Netherlands: Springer-Verlag, 1999.

[71] J. T. Oden and H. J. Brauchli. “On the calculation of consistent stress distributions in finite element approximations”. In: *International Journal for Numerical Methods in Engineering* 3 (1971), pp. 317–325.

[72] J. T. Oden, L. Demkowicz, T. Strouboulis, and P. Devloo. “Adaptive methods for problems in solid and fluid mechanics”. In: *Accuracy Estimates and Adaptive Refinements in Finite Element Computations*. Ed. by I. Babuška, O. Zienkiewicz, J. Gago, and de A. E. R. Oliveira. New York: Wiley, pp. 249–280.

[73] J. T. Oden, L. Demkowicz, W. Rachowicz, and T. A. Westermann. “Toward a universal hp-adaptive finite element strategy. Part 2: A posteriori error estimation”. In: *Computer Methods in Applied Mechanics and Engineering* 77 (1989), pp. 113–180.

[74] J. T. Oden and Y. Feng. “Local and pollution error estimation for finite element approximations of elliptic boundary value problems”. In: *Journal of Computational and Applied Mathematics* 74 (1996), pp. 245–293.

[75] L. Piegl and W. Tiller. *The NURBS Book*. Berlin: Springer-Verlag, 1997.

[76] A. Pinkus. *n-Widths in Approximation Theory*. Berlin: Springer-Verlag, 1985.

[77] M. J. D. Powell. *Approximation theory and practice*. New York, USA: Cambridge University Press, 1981.

[78] A. Quarteroni and A. Valli. *Numerical approximation of partial differential equations*. Berlin: Springer-Verlag, 2008.

[79] D. Schillinger, J. A. Evans, A. Reali, M. A. Scott, and T. J. Hughes. "Isogeometric collocation: Cost comparison with Galerkin methods and extension to adaptive hierarchical NURBS discretizations". In: *Computer Methods in Applied Mechanics and Engineering* 267 (2013), pp. 170–232.

[80] L. Schumaker. *Spline Functions: Basic Theory*. New York: Cambridge University Press, 2007.

[81] M. A. Scott, X. Li, T. W. Sederberg, and T. J. R. Hughes. "Local refinement of analysis-suitable T-splines". In: *Computer Methods in Applied Mechanics and Engineering* 213-216 (2012), pp. 206–222.

[82] J. Shen, T. Tang, and L.-L. Wang. *Spectral Methods*. Berlin: Springer-Verlag, 2011.

[83] B. Szabó and I. Babuška. *Finite Element Analysis*. New York: John Wiley & Sons, Ltd, 1991.

[84] L. B. da Veiga, D. Cho, and G. Sangalli. "Anisotropic NURBS approximation in isogeometric analysis". In: *Computer Methods in Applied Mechanics and Engineering* 209-212 (2012), pp. 1–11.

[85] L. B. da Veiga, A. Buffa, G. Sangalli, and R. Vázquez. "Mathematical analysis of variational isogeometric methods". In: *Acta Numerica* 23 (2014), pp. 157–287.

[86] L. B. da Veiga, A. Buffa, J. Rivas, and G. Sangalli. "Some estimates for h-p-k-refinement in Isogeometric Analysis". In: *Numerische Mathematik* 118 (2011), pp. 271–305.

[87] R. Verfürth. "A posteriori error estimation and adaptive mesh refinement techniques". In: *J. Comput. Appl. Math.* 50 (1994), pp. 67–83.

[88] R. Verfürth. *A Posteriori Error Estimation Techniques for Finite Element Methods*. UK: Oxford University Press, 2013.

[89] L. Wahlbin. *Superconvergence in Galerkin finite element methods*. Berlin: Springer-Verlag, 1995.

[90] L. B. Wahlbin. "Local Behaviour in Finite Element Methods". In: *Handbook of Numerical Analysis, Vol. II*. Ed. by P. Ciarlet and J. L. Lions. Amsterdam: Elsevier, pp. 353–522.

[91] Z. jun Wu, Z. dong Huang, Q. hua Liu, and B. quan Zuo. “A local solution approach for adaptive hierarchical refinement in isogeometric analysis”. In: *Computer Methods in Applied Mechanics and Engineering* 283 (2015), pp. 1467–1492.

[92] O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu. *The Finite Element Method: Its basis and fundamentals*. Massachusetts: Elsevier, 2013.

[93] O. Zienkiewicz and J. Z. Zhu. “The superconvergent patch recovery and a posteriori error estimates. I. The recovery technique”. In: *International Journal for Numerical Methods in Engineering* 33 (1992), pp. 1331–1364.

[94] O. Zienkiewicz and J. Z. Zhu. “The superconvergent patch recovery and a posteriori error estimates. II. Error estimates and adaptivity”. In: *International Journal for Numerical Methods in Engineering* 33 (1992), pp. 1365–1382.

[95] O. Zienkiewicz and J. Z. Zhu. “The superconvergent patch recovery (SPR) and adaptive finite element refinement”. In: *Computer Methods in Applied Mechanics and Engineering* 101 (1992), pp. 207–224.

Paper 1: A Posteriori Error Estimates for Isogeometric Analysis of the Stokes Equation

Abdullah Abdulhaque, Trond Kvamsdal, Kjetil André Johannessen, Mukesh Kumar and Arne Morten Kvarving

Sent to *Computer Methods in Applied Mechanics and Engineering*

Paper 2: A Posteriori Error Estimation for Isogeometric Analysis of the Navier-Stokes equation

Abdullah Abdulhaque, Trond Kvamsdal, Mukesh Kumar and Arne Morten Kvarving

Sent to *International Journal for Numerical Methods in Engineering*

Paper 3: Error Estimation for Isogeometric Analysis of Advection-Diffusion-Reaction Problems

Abdullah Abdulhaque, Trond Kvamsdal, Mukesh Kumar and Arne Morten Kvarving

To be submitted

Paper 4: Adaptive Isogeometric Analysis of the Boussinesq Equations for Buoyancy-Driven Flow

Abdullah Abdulhaque, Trond Kvamsdal, Mukesh Kumar and Arne Morten Kvarving

To be submitted

Appendix

Chapter A

Multivariate Calculus

Definition 1. *Let* $f : \mathbb{R}^3 \to \mathbb{R}$, $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^3$ *and* $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ *be respectively a scalar field, a vector field and a tensor field. If* $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ *and* $\hat{\mathbf{k}}$ *are the respective unit vectors in the* x*-,* y*- and* z*-directions, then the differential vector operators in Cartesian coordinates are defined as follows:*

$$\nabla f = \frac{\partial f}{\partial x}\hat{\mathbf{i}} + \frac{\partial f}{\partial y}\hat{\mathbf{j}} + \frac{\partial f}{\partial z}\hat{\mathbf{k}} \qquad \nabla \mathbf{f} = \begin{bmatrix} \frac{\partial g^1}{\partial x} & \frac{\partial g^1}{\partial y} & \frac{\partial g^1}{\partial z} \\ \frac{\partial g^2}{\partial x} & \frac{\partial g^2}{\partial y} & \frac{\partial g^2}{\partial z} \\ \frac{\partial g^3}{\partial x} & \frac{\partial g^3}{\partial y} & \frac{\partial g^3}{\partial z} \end{bmatrix}$$

$$\nabla \cdot \mathbf{f} = \frac{\partial g^1}{\partial x} + \frac{\partial g^2}{\partial y} + \frac{\partial g^3}{\partial z} \qquad \nabla \cdot \mathbf{F} = \begin{bmatrix} \frac{\partial g^{11}}{\partial x} + \frac{\partial g^{12}}{\partial y} + \frac{\partial g^{13}}{\partial z} \\ \frac{\partial g^{21}}{\partial x} + \frac{\partial g^{22}}{\partial y} + \frac{\partial g^{23}}{\partial z} \\ \frac{\partial g^{31}}{\partial x} + \frac{\partial g^{32}}{\partial y} + \frac{\partial g^{33}}{\partial z} \end{bmatrix}$$

$$\nabla \times \mathbf{f} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ g^1 & g^2 & g^3 \end{vmatrix}$$

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \qquad \nabla^2 \mathbf{f} = \begin{bmatrix} \frac{\partial^2 g^1}{\partial x^2} + \frac{\partial^2 g^1}{\partial y^2} + \frac{\partial^2 g^1}{\partial z^2} \\ \frac{\partial^2 g^2}{\partial x^2} + \frac{\partial^2 g^2}{\partial y^2} + \frac{\partial^2 g^2}{\partial z^2} \\ \frac{\partial^2 g^3}{\partial x^2} + \frac{\partial^2 g^3}{\partial y^2} + \frac{\partial^2 g^3}{\partial z^2} \end{bmatrix}$$

Theorem 1 (The theorems of Green, Stokes and Gauss [8])**.** *Let A be a closed area with the piecewise smooth boundary C, and V is a closed volume with the piecewise smooth surface S. Let $\mathbf{n}$ be the unit normal vector on the surface. Then we have the following integral theorems:*

$$\textit{Green's theorem 1:} \qquad \oint_C \mathbf{F} \cdot d\mathbf{r} = \iint_R (\nabla \times \mathbf{F}) \cdot \hat{\mathbf{k}}\, dA \tag{A.1a}$$

$$\textit{Green's theorem 2:} \qquad \oint_C \mathbf{F} \cdot \mathbf{n}\, ds = \iint_R \nabla \cdot \mathbf{F}\, dA \tag{A.1b}$$

$$\textit{Stokes' theorem:} \qquad \oint_C \mathbf{F} \cdot d\mathbf{r} = \iint_S (\nabla \times \mathbf{F}) \cdot \mathbf{n}\, dS \tag{A.1c}$$

$$\textit{Gauss' theorem:} \qquad \oiint_S \mathbf{F} \cdot \mathbf{n}\, dS = \iiint_V \nabla \cdot \mathbf{F}\, dV \tag{A.1d}$$

Corollary 1 (Green's identities [8])**.** *From Gauss' theorem, we have Green's identities:*

$$\iiint_V g\nabla^2 f\, dV = \oiint_S g\frac{\partial f}{\partial n}\, dS - \iiint_V \nabla f \cdot \nabla g\, dV \tag{A.2a}$$

$$\iiint_V g\nabla^2 f - f\nabla^2 g\, dV = \oiint_S g\frac{\partial f}{\partial n} - f\frac{\partial g}{\partial n}\, dS \tag{A.2b}$$

Corollary 2 (Special integral identities)**.** *From Stokes' and Gauss' theorems, we have the following integral identities:*

$$\iiint_V \nabla f\, dV = \oiint_S f\mathbf{n}\, dS \tag{A.3a}$$

$$\iiint_V \nabla^2 f\, dV = \oiint_S \frac{\partial f}{\partial n}\, dS \tag{A.3b}$$

$$\iiint_V f(\nabla \cdot \mathbf{F})\, dV = \oiint_S f(\mathbf{F} \cdot \mathbf{n})\, dS - \iiint_V \nabla f \cdot \mathbf{F}\, dV \tag{A.3c}$$

$$\iiint_V \nabla \times \mathbf{F}\, dV = -\oiint_S \mathbf{F} \times \mathbf{n}\, dS \tag{A.3d}$$

$$\iiint_V \mathbf{F} \cdot (\nabla \times \mathbf{G})\, dV = \iiint_V \mathbf{G} \cdot (\nabla \times \mathbf{F})\, dV - \oiint_S (\mathbf{F} \times \mathbf{G}) \cdot \mathbf{n}\, dS \tag{A.3e}$$

CHAPTER B

Function Space Theory

1 The space of differentiable functions

Definition 2. *The support of a function f is given by*

$$\operatorname{supp}(u) = \{\mathbf{x} \in \Omega \,:\, u(\mathbf{x}) \neq 0\} \tag{B.1}$$

Definition 3. *Define $u : \mathbb{R}^d \to \mathbb{R}$ as function of d variables. Then the multi-index derivative of order α is denoted as*

$$\partial^\alpha u = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_d^{\alpha_d}} \qquad |\alpha| = \sum_{i=1}^{d} \alpha_i \tag{B.2}$$

Definition 4. *Let f be a function on $I = [a, b]$. It is absolutely continuous if*

$$f(x) = f(a) + \int_a^x f'(t)\, dt \tag{B.3}$$

If f is Lipschitz continuous, then there is an $M > 0$ such that

$$|f(x) - f(y)| \leq M|x - y| \qquad , \qquad \forall x, y \in [a, b] \tag{B.4}$$

If f is continuously differentiable, then f' is continuous.

Definition 5. [2, 7] *Let $\Omega \subset \mathbb{R}^d$ be an open set. If $k \in \mathbb{Z}^+$ and $\alpha \in \mathbb{R}^d$, we define the spaces of continuous and differentiable functions as*

$$C(\Omega) = \{u : u \text{ is continuous on } \Omega\} \tag{B.5a}$$
$$C_c(\Omega) = \{u : u \in C(\Omega), \text{ supp} \subset\subset \Omega\} \tag{B.5b}$$
$$C_0(\Omega) = \{u : u \in C(\Omega),\ u(\partial\Omega) = 0\} \tag{B.5c}$$
$$C^k(\Omega) = \{u : \partial^{|\alpha|}u \in C(\Omega),\ \forall |\alpha| \leq k\} \tag{B.5d}$$
$$C_c^k(\Omega) = C^k(\Omega) \cap C_c(\Omega) \tag{B.5e}$$
$$C_0^k(\Omega) = C^k(\Omega) \cap C_0(\Omega) \tag{B.5f}$$
$$C^\infty(\Omega) = \mathcal{E}(\Omega) = \bigcap_{k=0}^{\infty} C^k(\Omega) \tag{B.5g}$$
$$C_c^\infty(\Omega) = \mathcal{D}(\Omega) = C^\infty(\Omega) \cap C_c(\Omega) \tag{B.5h}$$

If $p \in [1, \infty)$, these spaces can be equipped with the norms

$$\|u\|_{C^k(\Omega),p} = \left[\sum_{|\alpha|=0}^{k} \left(\sup_{x\in\Omega} |\partial^\alpha u|\right)^p\right]^{\frac{1}{p}} = \left[\sum_{i=0}^{k} |u|^p_{C^i(\Omega),p}\right]^{\frac{1}{p}} \tag{B.6a}$$
$$\|u\|_{C^k(\Omega),\infty} = \max_{0\leq|\alpha|\leq k} \sup_{x\in\Omega} |\partial^\alpha u| = \max_{0\leq i\leq k} |u|_{C^i(\Omega),\infty} \tag{B.6b}$$

They induce the similar seminorms

$$|u|_{C^k(\Omega),p} = \left[\sum_{|\alpha|=k} \left(\sup_{x\in\Omega} |\partial^\alpha u|\right)^p\right]^{\frac{1}{p}} \tag{B.7a}$$
$$|u|_{C^k(\Omega),\infty} = \max_{|\alpha|=k} \sup_{x\in\Omega} |\partial^\alpha u| \tag{B.7b}$$

Definition 6. [2] *The space of distributions is $\mathcal{D}'(\mathbb{R}^d)$, while $\mathcal{E}'(\mathbb{R}^d)$ consists of distributions with compact support. They satisfy the following inclusions:*

$$\mathcal{D}(\mathbb{R}^d) \subset \mathcal{E}(\mathbb{R}^d) \tag{B.8a}$$
$$\mathcal{E}'(\mathbb{R}^d) \subset \mathcal{D}'(\mathbb{R}^d) \tag{B.8b}$$

Definition 7. *The α-th weak partial derivative w of an arbitrary function u on a domain $\Omega \subset \mathbb{R}^d$ is defined as*

$$\int_\Omega u\partial^\alpha\varphi \, d\Omega = (-1)^\alpha \int_\Omega w\varphi \, d\Omega \qquad \forall\varphi \in C_0^\infty \tag{B.9}$$

Theorem 2 (Convergence in $\mathcal{D}(\Omega)$ [2]). *If $\{\phi_n\}_{n=1}^{\infty}$ is a function sequence converging in $\mathcal{D}(\Omega)$, then*

$$\begin{aligned} &\exists K \subset\subset \Omega : \operatorname{supp}(\phi_n - \phi) \subset K, && \forall n \in \mathbb{N} \\ &\lim_{n\to\infty} \partial^\alpha \phi_n \to \partial^\alpha \phi \text{ uniformly}, && \forall \alpha : |\alpha| \in \mathbb{Z}^+ \end{aligned}$$

Theorem 3 (Strong and weak derivative coincidence [2]). *If f is absolutely continuous, then the strong and weak derivatives coincide.*

Theorem 4 (Completeness and inclusion of $C^k(\Omega)$). *$C^k(\Omega)$ is a Banach space with respect to the supremum norm, but not the integration norm. The inclusion of differentiable functions is given by the following lattice:*

$$\begin{array}{ccccccc} C^\infty(\Omega) & \subset & \cdots & \subset & C^2(\Omega) & \subset & C^1(\Omega) \subset C^0(\Omega) \\ \cup & & & & \cup & & \cup \quad \cup \\ C_c^\infty(\Omega) & \subset & \cdots & \subset & C_c^2(\Omega) & \subset & C_c^1(\Omega) \subset C_c^0(\Omega) \end{array}$$

If $k_1, k_2 \in \mathbb{Z}^+$ are finite, and $k_1 < k_2$ then we have a continuous embedding:

$$C^{k_2}(\Omega) \hookrightarrow C^{k_1}(\Omega) \tag{B.11}$$

Definition 8 (C^k boundary [10]). *Let $\Omega \subset \mathbb{R}^d$ be an open domain. The boundary $\partial\Omega$ is C^k if there is a constant $r > 0$, a function $g : \mathbb{R}^{d-1} \longmapsto \mathbb{R}$ in C^k, and a coordinate system $(e_1, \ldots, e_d)$ such that*

$$\Omega \cap B(x^0, r) = \{x \in B(x^0, r) \,:\, x_d > g(\widetilde{x})\}$$

The same holds for $C^{k,\alpha}$ with $\alpha \in (0, 1]$. A C^1 boundary is Lipschitz, but not in the opposite direction.

2 The Lebesgue space

Definition 9. [1] *Let $\Omega \subset \mathbb{R}^d$ be a domain of nonzero measure, where $p \in [1, \infty)$ is the Lebesgue index. The Lebesgue space $L^p(\Omega)$ consists of Lebesgue measurable functions on Ω with norms given by*

$$\|u\|_{L^p(\Omega)} = \left[\int_\Omega |u(x)|^p \, dx\right]^{\frac{1}{p}} \qquad p \in [1, \infty) \tag{B.12a}$$

$$\|u\|_{L^\infty(\Omega)} = \sup_{x \in \Omega} |u(x)| \qquad p = \infty \tag{B.12b}$$

If $p = 2$, we have a Hilbert space $L^2(\Omega)$, and the inner product is

$$\langle u, v \rangle_{L^2(\Omega)} = \int_\Omega u(x) v(x) \, dx \tag{B.13}$$

Definition 10 (Local integrability)**.** *The space of locally integrable functions on a domain Ω is defined as follows:*

$$L^p_{loc}(\Omega) = \{f \in L^p(K) \, : \, K \subset\subset int(\Omega)\} \tag{B.14}$$

Definition 11 (Quotient Lebesgue space)**.** *The quotient Lebesgue space is a Banach space where the elements have zero mean average:*

$$L^p_0(\Omega) = \left\{ f \in L^p(\Omega) : \int_\Omega f \, dx = 0 \right\} \tag{B.15}$$

$$\|u\|_{L^p_0} = \inf_{\alpha \in \mathbb{R}} \|u + \alpha\|_{L^p} \tag{B.16}$$

Theorem 5 (Completeness of $L^p(\Omega)$ [1])**.** *$L^p(\Omega)$ is a Banach space with respect to the integration norm and the completion of $C^0(\Omega)$:*

$$L^p \equiv \overline{C^0}^{\|\cdot\|_{L^p}}$$

Theorem 6 (Reflexivity and separability of $L^p(\Omega)$ [1, 2])**.** *Let $p, q \in (1, \infty)$ be conjugate exponents such that $p^{-1} + q^{-1} = 0$. Then the following holds:*

1. *$p \in (1, \infty)$: L^p is uniformly convex and reflexive.*
2. *$p \in [1, \infty)$: L^p is separable.*
3. *$(L^p)^* = L^q$.*

Theorem 7 (Interpolation in L^p-spaces [1]). *Let $f \in L^p(\Omega) \cap L^q(\Omega)$ such that $1 \leq p < q \leq \infty$, $\alpha \in [0,1]$, $r \in (p,q)$ and $\frac{1}{r} = \frac{\alpha}{p} + \frac{1-\alpha}{q}$. Then $f \in L^r(\Omega)$, and the following inequality holds:*

$$\|f\|_{L^r(\Omega)} \leq \|f\|^{\alpha}_{L^p(\Omega)} \|f\|^{1-\alpha}_{L^q(\Omega)} \tag{B.17}$$

Theorem 8 (General L^p-embeddings [1, 2]). *Let $\Omega \subset \mathbb{R}^d$ be open and bounded. If the Lebesque indices p_1 and p_2 satisfy $p_1, p_2 \geq 1$ and $p_1 < p_2$, we have a continuous chain of embeddings:*

$$L^{\infty}(\Omega) \hookrightarrow L^{p_2}(\Omega) \hookrightarrow L^{p_1}(\Omega) \hookrightarrow L^1(\Omega) \tag{B.18}$$

Theorem 9 (L^p-inequalities [1]). *Define $\{p_i\}_{i=1}^r$ as a set of conjugate exponents, such that $\sum_{i=1}^r \frac{1}{p_i} = 1$ for all $p_i \in [1,\infty)$, and let $\{f_i\}_{i=1}^r$ be a set of functions. Then, Hölder's integral inequality is defined as*

$$\left\| \prod_{i=1}^{r} f_i \right\|_{L^1(\Omega)} \leq \prod_{i=1}^{r} \|f_i\|_{L^{p_i}(\Omega)} \tag{B.19}$$

Minkowski's integral inequality is defined as

$$\left\| \sum_{i=1}^{r} f_i \right\|_{L^p(\Omega)} \leq \sum_{i=1}^{r} \|f_i\|_{L^p(\Omega)} \tag{B.20}$$

If p and q be conjugate exponents, and we have the numbers $a, b, \epsilon > 0$, then Young's inequalities are given by:

$$\textit{Young's inequality 1:} \qquad ab \leq \frac{a^p}{p} + \frac{b^q}{q} \tag{B.21a}$$

$$\textit{Young's inequality 2:} \qquad ab \leq \epsilon a^2 + \frac{b^2}{4\epsilon} \tag{B.21b}$$

For any $a, b \in \mathbb{R}$, the arithmetic-geometric mean inequality is defined as

$$ab \leq \frac{1}{2}(a^2 + b^2) \tag{B.22}$$

3 The Hölder space

Definition 12. [2] *Let $\Omega \subset \mathbb{R}^d$ and $\lambda \in (0,1)$. We call the function f Hölder continuous of order λ if there is an $M > 0$ such that*

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq M\|\mathbf{x} - \mathbf{y}\|^\lambda \quad , \quad \forall \mathbf{x}, \mathbf{y} \in \Omega \tag{B.23}$$

The space $C^{0,\lambda}(\overline{\Omega})$ consists of functions that are Hölder continuous of order λ. It has a norm and semi-norm:

$$\|f\|_{C^{0,\lambda}(\overline{\Omega})} = \|f\|_{C^0(\overline{\Omega})} + |f|_{C^{0,\lambda}(\overline{\Omega})} \tag{B.24a}$$

$$|f|_{C^{0,\lambda}(\overline{\Omega})} = \sup_{\substack{\mathbf{x},\mathbf{y}\in\Omega \\ \mathbf{x}\neq\mathbf{y}}} \frac{|f(\mathbf{x}) - f(\mathbf{y})|}{\|\mathbf{x} - \mathbf{y}\|^\lambda} \tag{B.24b}$$

Definition 13. [2] *Let $\Omega \subset \mathbb{R}^d$, $\lambda \in (0,1]$ and $k \in \mathbb{N}$. The space $C^{k,\lambda}(\overline{\Omega})$ consists of functions $f \in C^k(\overline{\Omega})$ such that $\partial^\alpha f \in C^{0,\lambda}(\overline{\Omega})$ for $|\alpha| \leq k$. The norm and semi-norm are defined as follows:*

$$\|f\|_{C^{k,\lambda}(\overline{\Omega})} = \sum_{|\alpha|\leq k} \|\partial^\alpha f\|_{C^{0,\lambda}(\overline{\Omega})} \tag{B.25a}$$

$$|f|_{C^{k,\lambda}(\overline{\Omega})} = \sum_{|\alpha|=k} \sup_{\substack{\mathbf{x},\mathbf{y}\in\Omega \\ \mathbf{x}\neq\mathbf{y}}} \frac{|\partial^\alpha[f(\mathbf{x}) - f(\mathbf{y})]|}{\|\mathbf{x} - \mathbf{y}\|^\lambda} \tag{B.25b}$$

If f is Lipschitz continuous, it is Hölder continuous of order 1. Any Hölder continuous function is uniformly continuous, and hence continuous.

Theorem 10 ($C^{k,\lambda}$-embeddings [2])**.** *If $k \in \mathbb{Z}^+$ and $0 < \mu < \lambda \leq 0$, then the following embeddings er continuous for any domain $\Omega \subset \mathbb{R}^d$:*

$$C^{k,\lambda}(\overline{\Omega}) \hookrightarrow C^k(\overline{\Omega}) \tag{B.26a}$$

$$C^{k,\lambda}(\overline{\Omega}) \hookrightarrow C^{k,\mu}(\overline{\Omega}) \tag{B.26b}$$

If Ω is bounded, then the same embeddings are compact.

Definition 14 (Lipschitz boundary [10])**.** *Let $\Omega \subset \mathbb{R}^d$ be an open and connected domain. For $d \leq 2$, we call $\partial\Omega$ a Lipschitz boundary if there is a finite open cover $\{U^i\}_{i=1}^m$ of $\partial\Omega$ such that for $1 \leq j \leq m$, the following criterions hold:*

1. *$\partial\Omega \cap U^j$ is the graph of a Lipschitz function g^j.*
2. *$\Omega \cap U^j$ is on one side of this graph.*

For $1 \leq j \leq m$ there is an Euclidean coordinate system $\{e_i^j\}_{i=1}^d$, real numbers $r^j, h^j > 0$, and a Lipschitz function $g^j : \mathbb{R}^{d-1} \longmapsto \mathbb{R}$, such that

$$x = \sum_{i=1}^{d} x_i^j e_i^j \in U^j, \ x = (\widetilde{x}^j, x_n^j), \ |\widetilde{x}^j| < r^j \tag{B.27}$$

These coordinates satisfy some special rules:

$$x_d^j = g^j(\widetilde{x}^j) \implies x \in \partial\Omega \tag{B.28a}$$

$$0 < x_d^j - g^j(\widetilde{x}^j) < h^j \implies x \in \Omega \tag{B.28b}$$

$$0 > x_d^j - g^j(\widetilde{x}^j) > -h^j \implies x \notin \Omega \tag{B.28c}$$

Thus, the open covers can be defined as follows:

$$U^j = \{x \in \mathbb{R}^d : |\widetilde{x}^j| < r^j, \ |x_d^j - g^j(\widetilde{x}^j)| < h^j\} \tag{B.29}$$

4 The Sobolev space

Definition 15. [1] *Let $\Omega \subset \mathbb{R}^d$ be a domain of nonzero measure, where $k \in \mathbb{Z}^+$ is the derivative order and $p \in [1, \infty)$ is the Lebesgue index. The Sobolev space $W^{k,p}(\Omega)$ consists of all functions whose weak partial derivatives up to order k belong to $L^p_{loc}(\Omega)$. It is equipped with the norms*

$$\|u\|_{W^{k,p}(\Omega)} = \left[\sum_{|\alpha|\leq k} \|\partial^\alpha u\|^p_{L^p(\Omega)}\right]^{\frac{1}{p}} = \left[\sum_{i=0}^{k} |u|^p_{W^{i,p}(\Omega)}\right]^{\frac{1}{p}} \qquad p \in [1, \infty) \tag{B.30a}$$

$$\|u\|_{W^{k,\infty}(\Omega)} = \max_{|\alpha|\leq k} \|\partial^\alpha u\|_{L^\infty(\Omega)} = \max_{0\leq i\leq k} |u|_{W^{i,\infty}(\Omega)} \qquad p = \infty \tag{B.30b}$$

They induce the similar seminorms

$$|u|_{W^{k,p}(\Omega)} = \left[\sum_{|\alpha|=k} \|\partial^\alpha u\|^p_{L^p(\Omega)}\right]^{\frac{1}{p}} \tag{B.31a}$$

$$|u|_{W^{k,\infty}(\Omega)} = \max_{|\alpha|=k} \|\partial^\alpha u\|_{L^\infty(\Omega)} \tag{B.31b}$$

If $p = 2$, we have a Hilbert space $H^k(\Omega) \equiv W^{k,2}(\Omega)$ with the inner product

$$\langle u, v\rangle_{H^k(\Omega)} = \sum_{|\alpha|\leq k} \langle \partial^\alpha u, \partial^\alpha v\rangle_{L^2(\Omega)} \tag{B.32}$$

Definition 16. [1, 2] *The space $W^{k,p}_0(\Omega)$ consists of functions $u \in W^{k,p}(\Omega)$ such that the weak partial derivatives up to order k vanish on $\partial\Omega$:*

$$W^{k,p}_0(\Omega) = \left\{u \in W^{k,p}(\Omega) \,:\, \partial^\alpha u\big|_{\partial\Omega} = 0,\ |\alpha| \leq k\right\} \tag{B.33}$$

If u is an arbitrary function, its null extension is

$$\widetilde{u} = \begin{cases} u & x \in \Omega \\ 0 & x \in \mathbb{R}^d \backslash \Omega \end{cases}$$

The Sobolev space $W^{k,p}_{00}(\Omega)$ is, for $p \in (1, \infty)$, defined as

$$W^{k,p}_{00}(\Omega) = \{u \in W^{k,p}(\Omega) \,:\, \widetilde{u} \in W^{k,p}(\mathbb{R}^d)\} \tag{B.34}$$

If $k \in \mathbb{N}$, then $W^{k,p}_0(\Omega)$ and $W^{k,p}_{00}(\Omega)$ are equipped with the standard $W^{k,p}$-norm. Furthermore, $W^{k,p}_{00}(\Omega) = W^{k,p}_0(\Omega)$ when $k \in \mathbb{N}$ and $p \in (1, \infty)$.

Definition 17. [2] *Let $p \in [1,\infty)$ and $s = k + \sigma$ such that $l \in \mathbb{Z}^+$ and $\sigma \in (0,1)$. Then $W^{s,p}(\Omega) \subset W^{k,p}(\Omega)$, equipped with the Slobodetskii norm and seminorm:*

$$\|u\|_{W^{s,p}(\Omega)} = \left[\|u\|^p_{W^{s,p}(\Omega)} + \sum_{|\alpha|=m} \iint_{\substack{\Omega\times\Omega \\ \mathbf{x}\neq\mathbf{y}}} \frac{|\partial^\alpha[u(\mathbf{x}) - u(\mathbf{y})]|^p}{\|\mathbf{x}-\mathbf{y}\|^{d+\sigma p}}\, d\mathbf{x}\, d\mathbf{y} \right]^{\frac{1}{p}}$$

$$= \left[\|u\|^p_{W^{s,p}(\Omega)} + |u|^p_{W^{s,p}(\Omega)} \right]^{\frac{1}{p}} \tag{B.35a}$$

$$|u|_{W^{s,p}(\Omega)} = \left[\sum_{|\alpha|=m} \iint_{\substack{\Omega\times\Omega \\ \mathbf{x}\neq\mathbf{y}}} \frac{|\partial^\alpha[u(\mathbf{x}) - u(\mathbf{y})]|^p}{\|\mathbf{x}-\mathbf{y}\|^{d+\sigma p}}\, d\mathbf{x}\, d\mathbf{y} \right]^{\frac{1}{p}} \tag{B.35b}$$

Theorem 11 (Completeness of $W^{k,p}$ [1])**.** *$W^{k,p}$ is a Banach space with respect to the integration norm and the completion of C^k. The same holds for $W_0^{k,p}$ and $\mathcal{D}$ in this norm. If $s = k + \sigma$, $W^{s,p}$ is the completion of $C^{s,p}$:*

$$W^{k,p} \equiv \overline{C^k}^{\|\cdot\|_{W^{k,p}}} \tag{B.36a}$$

$$W_0^{k,p} \equiv \overline{\mathcal{D}}^{\|\cdot\|_{W^{k,p}}} \tag{B.36b}$$

$$W^{s,p} \equiv \overline{C^{s,p}}^{\|\cdot\|_{W^{s,p}}} \tag{B.36c}$$

Theorem 12 (Duality of $W^{k,p}$ [1, 2])**.** *If $p, q \in [1,\infty)$ are two conjugate exponents, then the dual space of $W_0^{k,p}(\Omega)$ is $W^{-k,q}(\Omega)$. The norm is*

$$\|L\|_{W^{-k,q}(\Omega)} = \sup_{u\in W^{k,p}(\Omega)\setminus\{0\}} \frac{|L(u)|}{\|u\|_{W^{k,p}(\Omega)}} \tag{B.37}$$

Riesz' representation theorem implies the existence of a unique function set $\{v_\alpha\}_{|\alpha|\leq k} \in L^q(\Omega)$ such that L can be expressed as

$$L(u) = \sum_{|\alpha|\leq k} \langle \partial^\alpha u, v_\alpha \rangle \tag{B.38}$$

This representation is unique for $p \in (1,\infty)$. We have also a special rule:

$$\Omega \neq \mathbb{R}^d \implies W_0^{k,q}(\Omega) = W^{k,p}$$
$$\Omega = \mathbb{R}^d \implies W_0^{k,q}(\Omega) \neq W^{k,p}$$

Thus, $(W_0^{k,p}(\Omega))' = W^{-k,q}(\Omega)$ when $\Omega = \mathbb{R}^d$.

Theorem 13 (Reflexivity and separability of $W^{k,p}$ [1, 2])**.** *Let p and q be conjugate exponents, and $k \in \mathbb{N}$ is arbitrary. Then the following holds:*

1. *$p \in (1, \infty)$: $W^{k,p}$ is uniformly convex and reflexive.*
2. *$p \in [1, \infty)$: $W^{k,p}$ is separable.*

The same rules hold for $W_0^{k,p}$, $W_{00}^{k,p}$ and $W^{s,p}$ ($s > 0$).

Theorem 14 (Density of $W^{k,p}$ [2])**.** *For $W_0^{k,p}$, We have some density rules:*

1. *$W_0^{k_2,p}$ is dense in $W_0^{k_1,p}$ for $k_1 < k_2$ and fixed $p \geq 1$.*
2. *$\mathcal{D}(\mathbb{R}^d)$ is dense in $W_0^{k,p}(\mathbb{R}^d)$ for $p \in [1, \infty)$ and $k \in \mathbb{Z}^+$.*
3. *$\mathcal{D}(\Omega)$ is dense in $W_{00}^{k,p}(\Omega)$ if $p \in (1, \infty)$ and $\partial\Omega$ is continuous.*

For $s > 0$, $p \in (1, \infty)$ and $\Omega \subset \mathbb{R}^d$, we have similar density rules for $W^{s,p}$:

1. *$\mathcal{D}(\Omega)$ is dense in $W_0^{s,p}(\Omega)$.*
2. *$\mathcal{D}(\mathbb{R}^d)$ is dense in $W_0^{k,p}(\mathbb{R}^d)$ and $W_0^{k,p}(\mathbb{R}^d) \equiv W^{k,p}(\mathbb{R}^d)$.*
3. *$\mathcal{D}(\Omega)$ is dense in $W_0^{s,p}(\Omega)$ and $W_0^{k,p}(\Omega) \equiv W^{k,p}(\Omega)$ if $s \in (0, 1/p]$ and $\partial\Omega$ is Lipschitz continuous.*

Theorem 15 (Compactness of $W^{s,p}$ [2])**.** *Let $\Omega \subset \mathbb{R}^d$ such that $\partial\Omega$ is Lipschitz continuous and $p \in (0, \infty)$. Let $s_1, s_2 \in \mathbb{R}^+$ and $m_1, m_2 \in \mathbb{Z}^+$ such that $s_1 < s_2$ and $k_1 < k_2$. If $s_1 < k_2 + \frac{1}{p}$ and $s_2 > k_1 + \frac{1}{p}$, then*

$$W^{s_2,p}(\Omega) \hookrightarrow\hookrightarrow W^{s_1,p}(\Omega) \tag{B.39a}$$

$$W_0^{s_2,p}(\Omega) \hookrightarrow\hookrightarrow W_0^{s_1,p}(\Omega) \tag{B.39b}$$

$$W_{00}^{k_2+\frac{1}{p},p}(\Omega) \hookrightarrow\hookrightarrow W_{00}^{k_1+\frac{1}{p},p}(\Omega) \tag{B.39c}$$

$$W_{00}^{k_2+\frac{1}{p},p}(\Omega) \hookrightarrow\hookrightarrow W_0^{s_1,p}(\Omega) \tag{B.39d}$$

$$W_0^{s_2,p}(\Omega) \hookrightarrow\hookrightarrow W_{00}^{k_1+\frac{1}{p},p}(\Omega) \tag{B.39e}$$

Theorem 16 (Sobolev's embedding theorem [2])**.** *Define the constants $r \in [0, s]$, $m \in \mathbb{N}$, $\sigma \in (0, 1)$ and $1 < p \leq q < \infty$. Let $\Omega \subset \mathbb{R}^d$ be an open and bounded domain with Lipschitz continuous boundary $\partial\Omega$ and the s-extension property, i.e. there is a continuous and linear extension operator $P_s : W^{s,p}(\Omega) \longmapsto W^{s,p}(\mathbb{R}^d)$ defined by $(P_s u) \downarrow_\Omega = u$ that satisfies*

$$\|P_s u\|_{W^{s,p}(\mathbb{R}^d)} \leq C \|u\|_{W^{s,p}(\Omega)}$$

Then we have the following continuous embeddings in $W^{k,p}(\Omega)$:

$$r - \frac{d}{q} \leq s - \frac{d}{p} \qquad \Longrightarrow \qquad W^{s,p}(\Omega) \hookrightarrow W^{r,q}(\Omega) \tag{B.40a}$$

$$m + \sigma \leq s - \frac{d}{p} \qquad \Longrightarrow \qquad W^{s,p}(\Omega) \hookrightarrow C^{m,\sigma}(\overline{\Omega}) \tag{B.40b}$$

$$p = 1,\ q = \infty,\ r \leq s - d \qquad \Longrightarrow \qquad W^{s,1}(\Omega) \hookrightarrow C^{r,\infty}(\overline{\Omega}) \tag{B.40c}$$

If $1 \leq p \leq q \leq \infty$, we have the following compact embeddings in $W^{k,p}(\Omega)$:

$$r - \frac{d}{q} < s - \frac{d}{p} \qquad \Longrightarrow \qquad W^{s,p}(\Omega) \hookrightarrow\hookrightarrow W^{r,q}(\Omega) \tag{B.41a}$$

$$m + \sigma < s - \frac{d}{p} \qquad \Longrightarrow \qquad W^{s,p}(\Omega) \hookrightarrow\hookrightarrow C^{m,\sigma}(\overline{\Omega}) \tag{B.41b}$$

Theorem 17 (Special $W^{k,p}$-embeddings [1, 2])**.** *Let $\Omega \subseteq \mathbb{R}^d$ be a domain. If $p \geq 1$ and $k \in \mathbb{Z}^+$ are fixed, then*

$$\mathcal{D}(\Omega) \hookrightarrow W_0^{k,p}(\Omega) \hookrightarrow W^{-k,p}(\Omega) \hookrightarrow \mathcal{D}'(\Omega) \tag{B.42}$$

For the space $W_{00}^{k,p}(\Omega)$ with $p \in (1, \infty)$, we have

$$W_{00}^{k,p}(\Omega) \hookrightarrow W^{k,p}(\overline{\Omega}) \hookrightarrow W^{k,p}(\Omega) \tag{B.43}$$

Theorem 18 (Poincaré's inequality [1, 6])**.** *If $\Omega \subset \mathbb{R}^d$ is a domain with finite measure, and $u \in W_0^{1,p}(\Omega)$, then Poincaré's inequality is defined as*

$$\|u\|_{L^p(\Omega)} \leq C_\Omega |u|_{W^{1,p}(\Omega)} \tag{B.44}$$

Theorem 19 (Gagliardo-Nirenberg-Sobolev inequality [6])**.** *If $p \in [1, d)$, $q = np/(n-p)$, $u \in C_c^1(\mathbb{R}^d)$ and $\operatorname{supp}(u) \subset\subset \Omega$, then the Gagliardo-Nirenberg-Sobolev inequality is defined as*

$$\|u\|_{L^q(\mathbb{R}^d)} \leq C \|\nabla u\|_{L^p(\mathbb{R}^d)} \tag{B.45}$$

If $u \in W^{1,p}(\Omega)$, and $\Omega \subset \mathbb{R}^d$ is bounded and open with Lipschitz boundary, then the following inequality holds:

$$\|u\|_{L^q(\Omega)} \leq C \|\nabla u\|_{W^{1,p}(\Omega)} \tag{B.46}$$

CHAPTER C

Finite Element Analysis

1 Differential operator theory

Definition 18 (Linear partial differential operators [3])**.** *Let $\mathcal{L}$ be a linear partial differential operator of order $2m$. Then the strong form of $\mathcal{L}$ and its equivalent divergence form are defined as follows:*

$$S_{\mathcal{L}}(u) = \sum_{|\eta|\leq 2m} a_\eta \partial^\eta u = \sum_{|\alpha|,|\beta|\leq m} (-1)^{|\beta|} \partial^\beta (a_{\alpha,\beta} \partial^\alpha u) \tag{C.1}$$

where $a_\eta = (-1)^m a_{\eta,\beta}$, $\alpha + \beta = \eta$, $|\alpha| = |\beta| = m$ and $a_{\eta,\beta} \in C^\infty(\overline{\Omega})$.

Definition 19 (Uniform ellipticity [6])**.** *Let $\{a_{ij}\}_{i=1,j=1}^{d,d} \in L^\infty$ such that they satisfy $a_{ij} = a_{ji}$. Define a linear partial differential operator as*

$$\mathcal{L}u = -\sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial}{\partial x_j} \left(a_{ij}(\mathbf{x}) \frac{\partial u}{\partial x_i} \right)$$

We call $\mathcal{L}$ uniformly elliptic if there is a constant $\alpha > 0$ such that

$$\sum_{i=1}^{d} \sum_{j=1}^{d} a_{ij}(\mathbf{x}) \frac{\partial u}{\partial x_i} \frac{\partial u}{\partial x_j} \geq \alpha |u|_{H^1}^2 \tag{C.2}$$

Every elliptic operator has even order.

Definition 20 (Strong ellipticity [5]). *Assume that a second-order linear partial differential operator of is defined as*

$$\mathcal{L}u = -\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial}{\partial x_j}\left(a_{ij}(\mathbf{x})\frac{\partial u}{\partial x_i}\right) + \sum_{i=1}^{d} b_i(\mathbf{x})\frac{\partial u}{\partial x_i} + b_0(\mathbf{x})u \qquad \text{(C.3)}$$

such that $\mathbf{A} = \{a_{ij}\}_{i=1,j=1}^{d,d}$. *If* $Re(\mathbf{A})$ *and* $Im(\mathbf{A})$ *commute, and the eigenvalues of* $\mathbf{A}$ *are positive and have the same argument,* $\mathcal{L}$ *is strongly elliptic.*

Definition 21 (Hypo-ellipticity [5]). *The characteristic polynomial of a second-order linear partial differential operator* $\mathcal{L}$ *is given by:*

$$p(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$$

Let $\mathbf{A}$ *be a constant matrix, and* $C : \mathbb{R}^d \longmapsto \mathbb{R}$ *satisfies* $C(\mathbf{x}) \to 0$ *when* $|\mathbf{x}| \to \infty$ *such that*

$$|\partial^{\alpha}(\mathbf{x})| \leq C(\mathbf{x})|p(\mathbf{x})| \qquad , \qquad \forall \mathbf{x} \in \mathbb{R}^d\,, \forall \alpha \in \mathbb{N}^d$$

We call $\mathcal{L}$ *hypo-elliptic if it satisfies the listed criterions.*

Definition 22 (Semi-ellipticity [5]). *Let* $\mathcal{L} = \sum a_{\alpha}D^{\alpha}$ *be a constant linear partial differential operator, and* $\{m_i\}_{i=1}^{d}$ *are partial orders with respect to* $\{\partial/\partial x_i\}_{i=1}^{d}$. *Denote* $|\alpha : m|$ *as* $\alpha_1/m_1 + \cdots + \alpha_d/m_d$. *We call* $\mathcal{L}$ *semi-elliptic if it satisfies*

$$\begin{aligned} &a_{\alpha} = 0 && \forall \alpha \in \mathbb{N}^d\,, |\alpha : m| > 1 \\ &p(\mathbf{x}) \neq 0 && \forall \mathbf{x} \in \mathbb{R}^d \backslash \{\mathbf{0}\} \end{aligned}$$

Any semi-elliptic operator is hypo-elliptic, but not conversely.

Proposition 1. [5] *Let* $\mathcal{L}$ *be a constant linear partial differential operator, and* $\mathcal{L}_p$ *is the corresponding principal part (highest-order derivatives). Then* $\mathcal{L}$ *is elliptic iff it is hypo-elliptic and satisfies*

$$\sum_{i=1}^{d}\left|\frac{\partial \mathcal{L}_p}{\partial x_i}(\mathbf{x})\right| \neq 0 \qquad , \qquad \forall \mathbf{x} \in \mathbb{R}^d \backslash \{\mathbf{0}\}$$

where $\mathbf{x} = (x_1, \ldots, x_d)$ *are the standard coordinates in* $\mathbb{R}^d$.

Theorem 20 (Gårding's inequality [4]). *Assume that $\mathcal{L}$ as a non-symmetric linear partial differential operator of second order, as described by (C.3). If $\{a_{ij}\}_{i=1,j=1}^{d,d}$, $\{b_i\}_{i=0}^{d} \in L^\infty$ and (C.2) holds, there is a finite positive constant K such that*

$$a(u,u) + K\|u\|_{L^2}^2 \geq \frac{\alpha}{2}\|u\|_{H^1}^2 \tag{C.4}$$

Definition 23 (Boundary operators [3]). *If a PDE has order $2m$, then the Dirichlet, Neumann and General operators are defined as*

$$B_D u = \frac{\partial^i u}{\partial x_i} \qquad 0 \leq i \leq m-1 \tag{C.5}$$

$$B_N u = \frac{\partial^i u}{\partial x_i} \qquad m \leq i \leq 2m-1 \tag{C.6}$$

$$B_j u = \sum_{|\alpha| \leq m_j} b_{j,\alpha} \partial^\alpha u \qquad m_j < 2m,\ 1 \leq j \leq m \tag{C.7}$$

2 Weak formulation of PDEs

Definition 24. [9] *Let V be a function space, and $F : V \longmapsto \mathbb{R}$ is a functional associating a real number to each element in V. For all $u, v \in V$ and $a, b \in \mathbb{R}$, let $C > 0$ be an arbitrary constant. We call F*

Linear:	$F(au + bv) = aF(u) + bF(v)$
Bounded:	$\lvert F(u)\rvert \leq C\lVert u\rVert_V$

Let $A : V \times V \longmapsto \mathbb{R}$ be a form on V. For all $u, v, w \in V$ and $a, b \in \mathbb{R}$, let $M > 0$ and $\alpha > 0$ be arbitrary constants. We call A

Bilinear:	$A(au + bv, w) = aA(u, w) + bA(v, w)$
	$A(u, av + bw) = aA(u, v) + bA(u, w)$
Symmetric:	$A(u, v) = A(v, u)$
Continuous:	$\lvert A(u, v)\rvert \leq M\lVert u\rVert_V \lVert v\rVert_V$
Positive:	$A(u, u) > 0$
Coercive:	$A(u, u) > \alpha\lVert u\rVert_V^2$

The operator norms of A and F are defined as

$$\lVert F\rVert = \sup_{u \in V\setminus\{0\}} \frac{\lVert F(u)\rVert}{\lVert u\rVert}$$

$$\lVert A\rVert = \sup_{u \in V\setminus\{0\}} \sup_{v \in V\setminus\{0\}} \frac{\lVert A(u, v)\rVert}{\lVert u\rVert\lVert v\rVert}$$

Theorem 21 (Lax-Milgram theorem [4])**.** *Let V be a Hilbert space, such that $a \in \mathcal{L}(V \times V, \mathbb{R})$, $f \in V^*$, and $a(\cdot, \cdot)$ is coercive. The variational problem is defined as*

$$u \in V \quad : \quad a(u, v) = f(v), \qquad \forall v \in V$$

This problem is well-posed, and the solution satisfies the a priori estimate

$$\lVert u\rVert_V = \frac{1}{\alpha}\lVert f\rVert_{V^*}, \qquad \forall f \in V^*$$

Theorem 22 (Banach-Nečas-Babuška theorem [10]). *Let V and W be Banach spaces, such that $a \in \mathcal{L}(W \times V, \mathbb{R})$, $f \in V^*$, and $V^{**} = V$. Assume that the variational problem is defined as*

$$u \in W \quad : \quad a(u,v) = f(v), \qquad \forall v \in V$$

This problem is well-posed if and only if

$$\exists \alpha > 0, \qquad \inf_{w \in W} \sup_{v \in V} \frac{a(w,v)}{\|w\|_W \|v\|_V} \geq \alpha$$
$$\forall v \in V, w \in W, \qquad a(w,v) = 0 \implies v = 0$$

The solution satisfies the a priori estimate

$$\|u\|_V = \frac{1}{\alpha} \|f\|_{V^*}, \qquad \forall f \in V^*$$

Theorem 23. [3] *Let $\Omega \subset \mathbb{R}^d$ be open and bounded with Lipschitz boundary $\partial\Omega$ such that $d \geq 2$ and $p \in [1, \infty)$. Define $\gamma_0 : C_0^\infty \longmapsto C^r(\partial\Omega)$ as a restriction to $\partial\Omega$ such that $\gamma_0 u = u\big|_{\partial\gamma}$. The extension γ of γ_0 has the property $\gamma_p = \gamma \in \mathcal{L}(W^{s,p}(\Omega), W^{t,p}(\partial\Omega))$ for certain s and t. Then*

1. *$\gamma_p : W^{1,p}(\Omega) \longmapsto W^{1-1/p,p}(\partial\Omega)$ is linear, surjective, continuous, and it satisfies the bound*

$$\|\gamma_p u\|_{W^{1-1/p,p}(\partial\Omega)} \leq C \|u\|_{W^{1,p}(\Omega)}$$

2. *$u \in W_0^{1,p}(\Omega)$ is equivalent to $u \in W^{1,p}(\Omega)$ with $\gamma u = 0$ on $\partial\Omega$).*
3. *If $\Omega \in C^{m-1,1}$, then $u \in W_0^{m,p}(\Omega)$ is equivalent to $u \in W^{m,p}(\Omega)$ and $\partial^\alpha u \in W_0^{1,p}(\Omega)$, i.e. $\gamma \partial^\alpha u = 0$ for $|\alpha| \leq m-1$.*
4. *The extension operator $E_p : W^{1-1/p,p}(\partial\Omega) \longmapsto W^{1,p}(\Omega)$ is linear, continuous, and corresponds to the inverse of γ_0.*

Theorem 24 (Trace theorem [3]). *Let $\Omega \subset \mathbb{R}^d$ have a Lipschitz boundary, $k \in \mathbb{Z}^+$ and $p \in [1, \infty]$. Then there is a positive constant C such that*

$$\|u\|_{W^{k,p}(\partial\Omega)} \leq C \|u\|_{W^{k,p}(\Omega)}^{1-1/p} \|u\|_{W^{k+1,p}(\Omega)}^{1/p} \quad , \quad \forall u \in W^{k+1,p}(\Omega) \quad \text{(C.8)}$$

Bibliography

[1] R. A. Adams and J. J. F. Fournier. *Sobolev Spaces*. Amsterdam: Elsevier, 2003.

[2] P. K. Bhattacharyya. *Distributions: generalized functions with applications in Sobolev spaces*. Göttingen: Walter de Gruyter GmbH, 2012.

[3] K. Böhmer. *Numerical Methods for Nonlinear Elliptic Differential Equations*. New York: Oxford University Press, 2003.

[4] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. New York: Springer-Verlag, 2008.

[5] R. Dautray and J.-L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology, Volume 2*. Paris: Springer-Verlag, 2000.

[6] L. Evans. *Partial differential equations*. Providence, USA: American Mathematical Society, 2010.

[7] C. Gasquet and P. Witomski. *Fourier Analysis and Applications*. New York: Springer-Verlag, 1999.

[8] J. Marsden and A. Tromba. *Vector calculus*. New York: W. H. Freeman and Company, 2012.

[9] A. Quarteroni and A. Valli. *Numerical approximation of partial differential equations*. Berlin: Springer-Verlag, 2008.

[10] C. Schwab. *p- and hp- Finite Element Methods*. UK: Oxford University Press, 1998.