Optimal Numerical Basis Functions in the Partition of Unity Method

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Denis Alexander Düsseldorf
aus
Köln

Bonn, Dezember 2023

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Marc Alexander Schweitzer
2. Gutachterin: Prof. Dr. Ira Neitzel

Tag der Promotion: 20.02.2024
Erscheinungsjahr: 2024

Acknowledgements

Für meine Eltern, Andrea und Heinz, die mich in jeder Situation unterstützen und mir Halt geben.

Für meine Verlobte, Flor, die mich motiviert und mir immer den Rücken frei hält.

Für meine Urgroßeltern, Luise und Karl, die gespannt auf diesen Tag warteten, ihn aber leider nicht mehr erleben durften.

Für all meine Freunde, die mich mit schlechter Laune ertragen mussten, weil wieder etwas nicht geklappt hat.

I would also like to thank my supervisor, Marc Alexander Schweitzer, for making this work possible and supporting me throughout these years. I'm grateful for the countless discussions and help from the whole PUMA group of Fraunhofer SCAI and the people from the Institute of Numerical Simulation, including the former colleagues Albert Ziegenhagel and Christian Rieger. All of you have made this work possible.

Contents

List of Figures

List of Tables

1 Introduction and motivation

Partial differential equations (PDE) arise from the modeling of a wide range of physical problems and knowing how to solve them is of large interest in many industrial undertakings. Starting from the need to understand wave propagation and heat diffusion, Leonhard Euler was among the first people to develop easy discretization schemes to solve such problems in the 18th century, resulting in the first Finite Difference Methods. In the 19th century Spectral Methods came to light, introduced by Jean-Baptiste Joseph Fourier as a result of his idea of representing any function in terms of a trigonometric series. With the invention of electronic computers, Finite Difference Methods were developed further and applied to a broader range of PDE problems such as fluid dynamics and structural analysis. While all of the existing methods had been based on mostly regular grids to discretize a physical body of interest, the Finite Element Method (FEMs), which was developed in the midst of the 20th century, proved itself to be a gamechanger for its ability to handle complicated domains using non-uniform grids. The early Finite Element Methods were developed by several researches in the 1940s, 1950s and 1960s, among them Richard Courant, John Tinsley Oden, Olgierd Zienkiewicz and Ray William Clough ([Cou43, TCMT56, ZC67]). For fluid dynamics and transport phenomena which can hardly be discretized using fixed grids, the Finite Volume Methods emerged shortly after the introduction of the Finite Element Method. Its earliest contributors were Arthur Allan Harlow, John E. Welch ([HW65]).

To this day, the mentioned methods have been improved and adapted to handle ever more exotic domains and partial differential equations. On one hand, this is due to the development of high-performance hardware such as multi-core processors and advances in semiconductor technology. On the other hand, smart changes and adaptions of the generic algorithms describing the numerical schemes to certain applications of interest have had a tremendous impact on the performance of the numerical schemes, such as the generalization of the FEM into the Partition of Unity Finite Element Method (PUFEM) due to Ivo Babuška and Jens M. Melenk ([MB96]). While the classical FEM is based on a global mesh, the PUFEM allowed to use locally defined, independent approximation spaces, which were then combined in the definition of a global approximation space. Unfortunately, even the limits of modern-day hardware can be reached with ease, since the mere quantity and size of the respective systems of linear equations to be solved are simply too large. As strong simplifications of the models under study simultaneously decrease their reliability and thus their interestingness, advanced numerical methods have to be developed in order to make these problems computable. Yet, to provide sufficient accuracy and reliability requires substantial

research into these methods and the particular design problem.

Among the reasons various industries are interested in solving partial differential equations is the engineering and design of materials, since the development and construction of complex mechanical structures relies on the use of materials that have use-case adapted properties. Oftentimes, such materials are not known a priori and must themselves be developed. However, real-life testing involves a huge financial risk since many physical material coupons and prototypes must be manufactured and exposed to a series of physical tests designed to measure the materials response in situations of interest. There is an infinite number of possible materials to be tested, and the material design cycle may hence be arbitrarily expensive. A key factor to cut costs is numerical simulation, which could be used to reduce the amount of physical testing by identifying promising material designs. Consequently, physical testing has to be performed only for these identified designs. A scheme of the material design cycle together with a possible bypass due to an increased use of numerical simulation is shown in Figure 1.1. The employed numerical methods must, however,

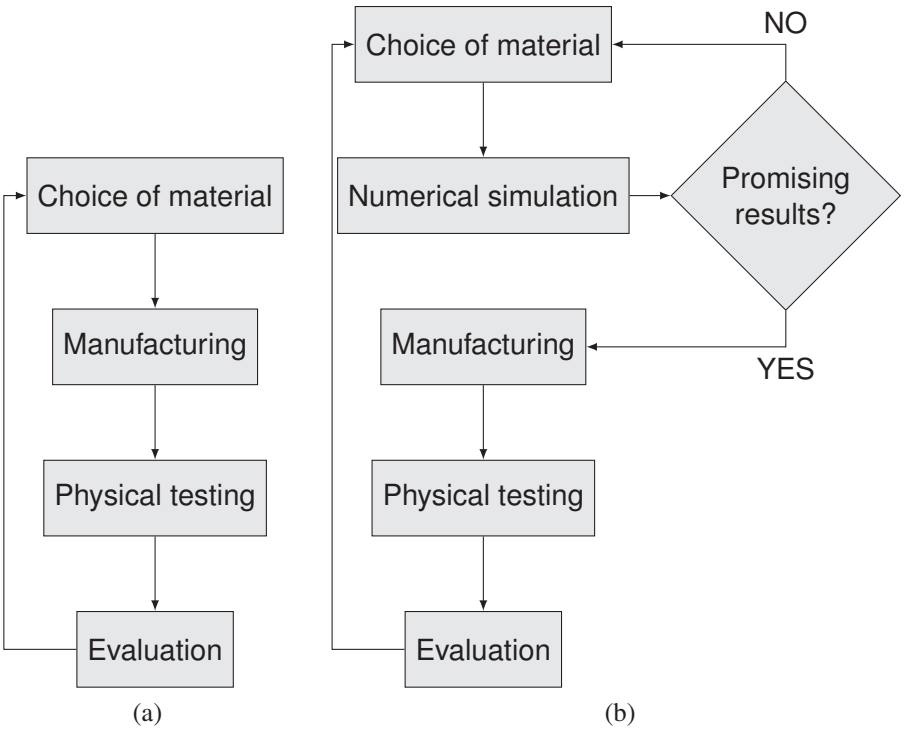


Figure 1.1: (a) Simplified traditional design cycle: Manufacturing and testing are the most expensive steps. (b) Pre-identification of promising designs may cut costs by bypassing manufacturing and testing.

be sufficiently accurate and fast, since the numerical simulation step includes all loading scenarios that would also be considered in physical testing. In total, a large number of partial differential equations for a range of materials has to be solved in order to obtain the most promising design.

This thesis considers the general case of elliptic, even-order partial differential equations, which are introduced in Chapter 2. In Chapter 3, the impact and difficulty arising from heterogeneous coefficients is briefly described, before introducing the partial differential equations of interest in this

thesis and showing that they are indeed elliptic partial differential equations with unique solutions. The Partition of Unity Method (PUM) due to Marc Alexander Schweitzer, a further abstraction of the PUFEM is introduced in Chapter 4. Traditional mesh-based methods typically use piecewise polynomial approximation spaces and rely on spatial refinement to improve the approximation power of the global space. In contrast, the PUM allows for a smart, operator-dependent choice of local approximation spaces, which are then linked in a global approximation space. Their quality can be enhanced either by using spatial refinement or by adding additional functions to the local approximation spaces. Local approximation quality has a direct influence on global approximation quality, and using local spaces with strong approximation properties is hence favorable. Chapter 5 presents a way of constructing local spaces that are optimal in a certain sense. The so-constructed local spaces are independent of the load and imposed boundary conditions, and can be reused. They will be pre-computed in a so-called offline phase, before any global simulation that uses these spaces is started. The computation of optimal local approximation spaces, including the impact of parameters used in their construction, is investigated in detail in Chapter 6. Using four benchmark problems with a yet computable reference solution, a general approach for their computation and use in more complex situations is identified. Many problems of (industrial) interest require discretizations using tens of millions of degrees of freedom, and while these problems may be computationally infeasible using standard methods, optimal local approximation spaces may be used to replace heavy local refinement with a smarter choice of basis functions. As a result, the number of degrees of freedom is reduced drastically and the previously computationally infeasible problem becomes feasible, all while maintaining the quality of high-resolution solutions. Two such complicated problems are investigated in Chapter 7, the first being posed on a detailed two-dimensional domain describing the rib of an airplane wing (Section 7.1). The second problem considered in Section 7.2 describes the dynamic propagation of a wave in a three-dimensional heterogeneous material with multiple spherical inclusions.

This thesis proposes details of a constructive method to compute optimal local approximation spaces. The original framework, which was introduced for the case of second-order elliptic PDE ([BL11]), was generalized to the case of even-order elliptic PDE. Moreover, the algebraic conditions ensuring reusability of the optimal basis functions, whose computation is numerically expensive, were developed during the writing of this thesis. The effect of various parameters that are to be chosen during their computation is investigated in a series of experiments. All of the experiments were conducted using `PUMA` (**P**artition of **U**nity **M**ethod and **A**pplications), an efficient implementation of the Partition of Unity Method developed by Fraunhofer SCAI ([SCA]). The back end of `PUMA` is implemented in `C++` while its functionality can be controlled using a `Python 3` front end. In order to simplify the computation of optimal basis functions and their use in `PUMA`, an additional `Python` package, `optbasefun`, was developed. Many of the `Python` classes implemented in this package, handle instances of `PUMA` classes together with additional meta information. This way, the entire computational process related to optimal basis functions is modularized into mostly interchangeable blocks of code. Having the computational setup in place, the last goal of this thesis is to identify trade-offs between the computational effort involved in the construction of optimal basis functions and their performance.

2 Elliptic partial differential operators

In this chapter, the class of elliptic partial differential operators, which is considered throughout this thesis is introduced. Elliptic partial differential equations generalize the Laplace equation and arise from the description of a variety of static problems. The theory presented in this chapter is based on [Hac17], especially Chapters 5 and 7. The definitions and theoretical results will for the scope of this thesis be extended to hold for vector-valued differential operators. In order to consider the scalar and vector-valued cases conjointly, an additional parameter n is used in the respective formulas, which takes either the value 1 (scalar PDE) or d (vector-valued PDE), with $d \in \mathbb{N}$ denoting the spatial dimension of the domain under study.

Section 2.1 describes the general notation and introduces basic definitions for working with partial differential equations of arbitrary order $\mathrm{k} \in \mathbb{N}$. The case of even order $2\,\mathrm{k} \in \mathbb{N}$, together with its implications, is then presented in detail in Section 2.2. This includes the concept of variational formulations, as well as conditions for existence and uniqueness of solutions.

2.1 Partial differential operators

Throughout the whole thesis, $\Omega \subset \mathbb{R}^d$ denotes a domain of interest and $d \in \mathbb{N}$ is the spatial dimension. By $\mathcal{X}(\Omega)$, a Banach space of functions with certain regularity to be specified later is denoted, and this space is referred to as the *space of coefficient functions*. For $0 < \mathrm{k} \in \mathbb{N}$ and $\mathrm{n} \in \{1, d\}$, let $\mathcal{L} : [\mathcal{C}^{\mathrm{k}}(\Omega)]^{\mathrm{n}} \to [\mathcal{C}^0(\Omega)]^{\mathrm{n}}$ be a linear partial differential operator of the form

$$\mathcal{L}\,u \coloneqq \left[\mathcal{L}_{i,j}\right]_{i,j=1}^{n} \left[u_i\right]_{i=1}^{\mathrm{n}} = \begin{bmatrix} \mathcal{L}_{1,1} & \dots & \mathcal{L}_{1,\mathrm{n}} \\ \vdots & \ddots & \vdots \\ \mathcal{L}_{n,1} & \dots & \mathcal{L}_{\mathrm{n},\mathrm{n}} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_{\mathrm{n}} \end{bmatrix}, \tag{2.1}$$

for all $u \in [\mathcal{C}^{\mathrm{k}}(\Omega)]^{\mathrm{n}}$, with $\mathcal{L}_{i,j} : \mathcal{C}^{\mathrm{k}}(\Omega) \to \mathcal{C}^0(\Omega)$ and

$$\mathcal{L}_{i,j}\,v \coloneqq \sum_{\substack{\alpha^{i,j} \in \mathbb{N}_0^d \\ |\alpha^{i,j}| \leqslant \mathrm{k}}} a_{\alpha^{i,j}} \partial^{\alpha^{i,j}} v \quad \forall i,j = 1, \dots, \mathrm{n},\ \forall v \in \mathcal{C}^{\mathrm{k}}(\Omega). \tag{2.2}$$

The operator $\mathcal{L}$ is of order k, meaning that at most k-th order partial derivatives appear in its definition. In (2.2), $\alpha^{i,j} = [\alpha_p^{i,j}]_{p=1}^d \in \mathbb{N}_0^d$ denote multi-indices with $|\alpha^{i,j}| \coloneqq \sum_{p=1}^d \alpha_p^{i,j}$, partial

derivatives are denoted

$$\partial^{\alpha^{i,j}} v(x) = \frac{\partial^{\alpha_1^{i,j}}}{\partial x_1^{\alpha_1^{i,j}}} \dots \frac{\partial^{\alpha_d^{i,j}}}{\partial x_d^{\alpha_d^{i,j}}} v(x), \tag{2.3}$$

and $a_{\alpha^{i,j}} \in \mathcal{X}(\Omega)$ for all $\alpha^{i,j} \in \mathbb{N}_0^d$ with $|\alpha^{i,j}| \leqslant \mathrm{k}$ and all $i, j = 1, \dots, \mathrm{n}$ are the coefficients.

Note: For now, the space of coefficients $\mathcal{X}(\Omega)$ is arbitrary. Usual choices are $\mathcal{C}^0(\Omega)$, $\mathcal{C}^1(\Omega)$, or $\mathrm{L}^\infty(\Omega)$, but the assumed degree of regularity always depends on the PDE at hand. In the upcoming definitions and theory, many conditions will be formulated 'for all $x \in \Omega$', implicitly meaning 'for almost every $x \in \Omega$' whenever this seems appropriate (e.g. when $\mathcal{X}(\Omega) = \mathrm{L}^\infty(\Omega)$).

Remark 2.1. *Note that any operator $\mathcal{L}$ as in (2.1) can be written as the sum*

$$\mathcal{L} = \mathcal{L}^0 + \dots + \mathcal{L}^{\mathrm{k}}, \tag{2.4}$$

where $\mathcal{L}^m : [\mathcal{C}^m(\Omega)]^{\mathrm{n}} \to [\mathcal{C}^0(\Omega)]^{\mathrm{n}}$ are linear differential operators of order $m = 0, \dots, \mathrm{k}$ taking the form

$$\mathcal{L}^m u = [\mathcal{L}_{i,j}^m]_{i,j=1}^{\mathrm{n}} [u_i]_{i=1}^{\mathrm{n}}, \quad \forall u \in [\mathcal{C}^m(\Omega)]^{\mathrm{n}} \tag{2.5}$$

with $\mathcal{L}_{i,j}^m : \mathcal{C}^m(\Omega) \to \mathcal{C}^0(\Omega)$ and

$$\mathcal{L}_{i,j}^m v = \sum_{\substack{\alpha^{i,j} \in \mathbb{R}^d \\ |\alpha^{i,j}| = m}} a_{\alpha^{i,j}} \partial^{\alpha^{i,j}} v, \quad \forall i, j = 1, \dots, \mathrm{n}, \quad \forall v \in \mathcal{C}^m(\Omega) \tag{2.6}$$

and all $m = 0, \dots, \mathrm{k}$.

*

Definition 1 (Main component)**.** *If a linear differential operator $\mathcal{L}$ of order $\mathrm{k} \in \mathbb{N}$ such as in (2.1) is written in the form (2.4), then the operator $\mathcal{L}^{\mathrm{k}}$ is called* main-component *of $\mathcal{L}$.* ○

Remark 2.2. *In the case of $\mathrm{n} = 1$, the operator $\mathcal{L} = \mathcal{L}_{1,1}$ is linear and scalar-valued. After dropping unused indices it reads*

$$\mathcal{L} u = \sum_{\substack{\alpha \in \mathbb{N}_0^d \\ |\alpha| \leqslant \mathrm{k}}} a_\alpha \partial^\alpha u, \quad \forall u \in \mathcal{C}^{\mathrm{k}}(\Omega). \tag{2.7}$$

If moreover $\mathcal{L}$ is of even order $2\,\mathrm{k}$ it can be written as

$$\mathcal{L} u = \sum_{\substack{\alpha \in \mathbb{N}_0^d \\ |\alpha| \leqslant \mathrm{k}}} \sum_{\substack{\beta \in \mathbb{N}_0^d \\ |\beta| \leqslant \mathrm{k}}} (-1)^{|\beta|} \partial^\beta \left(a_{\alpha,\beta} \partial^\alpha v \right), \quad \forall v \in \mathcal{C}^{\mathrm{k}}(\Omega). \tag{2.8}$$

*

Remark 2.3. *In the case of* $\mathrm{n} = \mathrm{k} = 1$ *and* $\mathcal{X}(\Omega)$ *containing functions with at least one continuous derivative, the operator* $\mathcal{L} = \mathcal{L}_{1,1}$ *can be written in standard divergence form,*

$$\mathcal{L}\,u = -\operatorname{div}(A\nabla u) + b \cdot \nabla u + cu, \tag{2.9}$$

with the coefficients $c \in \mathcal{X}(\Omega)$*,*

$$A = \begin{bmatrix} a_{1,1} & \dots & a_{1,d} \\ \vdots & \ddots & \vdots \\ a_{d,1} & \dots & a_{d,d} \end{bmatrix} \in \left[\mathcal{X}(\Omega)\right]^{d\times d}, \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_d \end{bmatrix} \in \left[\mathcal{X}(\Omega)\right]^{d}. \tag{2.10}$$

*

The operator $\mathcal{L}$ may satisfy useful conditions guaranteeing solvability of partial differential equations including $\mathcal{L}$. Among them is *ellipticity*.

Definition 2 (Ellipticity of scalar-valued linear differential operators)**.** *A scalar-valued linear differential operator of order* $\mathrm{k} \in \mathbb{N}$ *of the form (2.7) is called* elliptic*, if for all* $x \in \Omega$ *and all* $\xi \in \mathbb{R}^d\backslash\{0\}$ *it holds that*

$$\sum_{\substack{\alpha\in\mathbb{N}_0^d \\ |\alpha|=\mathrm{k}}} a_\alpha(x)\xi^\alpha \neq 0, \tag{2.11}$$

where $\xi^\alpha = \xi_1^{\alpha_1}\dots\xi_d^{\alpha_d}$ *for all appearing multi-indices* α*.* ○

Remark 2.4. *If the assumptions from Remark 2.3 hold with a constant and symmetric* $A \in \mathbb{R}^{d\times d}_{sym}$*, then* $\mathcal{L}$ *is elliptic if and only if all eigenvalues of* A *are bigger than zero.*

*

The definition of ellipticity can be extended easily to vector-valued linear differential operators.

Definition 3 (Ellipticity of vector-valued linear differential operators)**.** *A vector-valued linear differential operator of order* $\mathrm{k} \in \mathbb{N}$ *of the form (2.1) with* $\mathrm{n} = d$ *is called* elliptic*, if for all* $x \in \Omega$ *and all* $\xi \in \mathbb{R}^d\backslash\{0\}$ *it holds that*

$$\sum_{\substack{\alpha^{i,1}\in\mathbb{N}_0^d \\ |\alpha^{i,1}|=\mathrm{k}}} a_{\alpha^{i,1}}(x)\xi^{\alpha^{i,1}} + \ldots + \sum_{\substack{\alpha^{i,d}\in\mathbb{N}_0^d \\ |\alpha^{i,d}|=\mathrm{k}}} a_{\alpha^{i,d}}(x)\xi^{\alpha^{i,d}} \neq 0, \quad \forall i = 1,\dots,d. \tag{2.12}$$

○

The general form of a partial differential equation reads as follows.

Problem 1: General Partial Differential Equation.

Let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ be open and bounded, let $\mathcal{L}$ be a linear differential operator of order $\mathrm{k} \in \mathbb{N}$ as in (2.1) and let $\mathrm{n} \in \{1, d\}$. Find a function $u \in [\mathcal{C}^{\mathrm{k}}(\Omega)]^{\mathrm{n}}$ satisfying

$$\begin{aligned} \mathcal{L}\, u &= f, \quad \text{in } \Omega \\ \mathcal{B}^0\, u &= g^0, \quad \text{on } \Gamma^0 \subset \partial\,\Omega \\ &\vdots \\ \mathcal{B}^{\mathrm{k}-1}\, u &= g^{\mathrm{k}-1}, \quad \text{on } \Gamma^{\mathrm{k}-1} \subset \partial\,\Omega \end{aligned} \tag{2.13}$$

where $\mathcal{B}^m$ are (piecewise) linear differential operators of order $m = 0, \ldots, \mathrm{k}-1$ defined on corresponding parts of the boundary and $f : \Omega \to \mathbb{R}^{\mathrm{n}}$, $g^m : \Gamma^m \to \mathbb{R}^{\mathrm{n}}, m = 0, \ldots, \mathrm{k}-1$ are sufficiently smooth functions.

Remark 2.5. *The operator $\mathcal{B}^0$ is of order 0 and hence does not contain any derivatives. The coefficients of the operators $\mathcal{B}^0, \ldots, \mathcal{B}^{\mathrm{k}-1}$ may depend on the domain, e.g. on its outer normal. Essential, as well as natural boundary conditions may be expressed conjointly in this form.*

*

Remark 2.6. *The functions $f, g^0, \ldots, g^{\mathrm{k}-1}$ are arbitrary and they may also be defined piecewise. Furthermore, the coefficients of the boundary operators $\mathcal{B}^0, \ldots, \mathcal{B}^{\mathrm{k}-1}$ may be defined piecewise as well, meaning that several boundary conditions of the same order may be expressed conjointly.*

*

The boundary operators $\mathcal{B}^0, \ldots, \mathcal{B}^{\mathrm{k}-1}$ may not be chosen arbitrarily, but they need to form a *normal system* on all parts of the boundary and *cover* the operator $\mathcal{L}$, in order for Problem 1 to satisfy a certain form of ellipticity and hence be solvable. The definitions and very technical results regarding these concepts can for example be found in [LMM68]. They are rather abstract and can hardly be applied directly to show the solvability of a given partial differential equation. Throughout the next section, more practical and easier-to-check conditions are established for the case of even-order partial differential operators.

2.2 Partial differential operators of even order

In this section, the special case of differential operators of even order $2\,\mathrm{k} \in \mathbb{N}$ are investigated. Whenever the coefficients of $\mathcal{L}_{i,j}$ for $i, j = 1, \ldots, \mathrm{n}$ appearing in the definition of the even-order operator $\mathcal{L}$ are sufficiently smooth, the scalar operators $\mathcal{L}_{i,j}$ from (2.2) can be written in the form

$$\mathcal{L}_{i,j} v = \sum_{\substack{\alpha^{i,j} \in \mathbb{N}_0^d \\ |\alpha^{i,j}| \leqslant \mathrm{k}}} \sum_{\substack{\beta^{i,j} \in \mathbb{N}_0^d \\ |\beta^{i,j}| \leqslant \mathrm{k}}} (-1)^{|\beta^{i,j}|} \partial^{\beta^{i,j}} \left(a_{\alpha^{i,j},\beta^{i,j}} \partial^{\alpha^{i,j}} v \right), \quad \forall v \in \mathcal{C}^{\mathrm{k}}(\Omega). \tag{2.14}$$

For a differential operator of even order $2\,\mathrm{k}$, Problem 1 can be written in the following form.

Problem 2: General Partial Differential Equation of even order.

Let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ be open and bounded, let $\mathcal{L}$ be a linear differential operator of even order $2\,\mathrm{k} \in \mathbb{N}$ as in (2.1) and let $\mathrm{n} \in \{1, d\}$. Find a function $u \in [\mathcal{C}^{2\,\mathrm{k}}(\Omega)]^{\mathrm{n}}$ satisfying

$$\begin{aligned} \mathcal{L}\, u &= f, \quad \text{in } \Omega \\ \mathcal{B}^0\, u &= g^0, \quad \text{on } \Gamma^0 \subset \partial\,\Omega \\ &\vdots \\ \mathcal{B}^{2\,\mathrm{k}-1} &= g^{2\,\mathrm{k}-1}, \quad \text{on } \Gamma^{2\,\mathrm{k}-1} \subset \partial\,\Omega \end{aligned} \tag{2.15}$$

where $\mathcal{B}^m$ are (piecewise) linear differential operators of order $m = 0, \ldots, 2\,\mathrm{k}-1$ defined on corresponding parts of the boundary and $f : \Omega \to \mathbb{R}^{\mathrm{n}}$, $g^m : \Gamma^m \to \mathbb{R}^{\mathrm{n}}, m = 0, \ldots, 2\,\mathrm{k}-1$ are sufficiently smooth functions.

Note: Problem 2 is stated in the most general form. It is not necessary to define boundary operators of all orders for the problem and its variational formulation (cf. Section 2.2.1) to be well-posed, so possibly $\Gamma^j = \varnothing$ for some $j \in \{0, \ldots, 2\,\mathrm{k}-1\}$. In general, for even order $2\,\mathrm{k}$ only k boundary conditions have to be imposed.

For the upcoming theory, it is common to consider the case that only the first k boundary operators prescribe essential values of normal derivatives. These constraints can be incorporated into the trial and test space. The presence of other boundary conditions mainly results in additional constraints and terms appearing in the variational formulation that will be developed below. For the outer normal $\vec{n} = \left[\vec{n}_1, \ldots, \vec{n}_d\right]^T$, the first order normal derivative of a function u is defined as

$$\partial_{\vec{n}} u := \vec{n} \cdot \nabla u = \vec{n}_1 \partial_{x_1} u + \ldots + \vec{n}_d \partial_{x_d} u. \tag{2.16}$$

The same concept can be applied recursively to obtain higher order normal derivatives of u,

$$\partial_{\vec{n}}^i = \partial_{\vec{n}} \left(\partial_{\vec{n}}^{i-1} u \right). \tag{2.17}$$

From now on, the boundary operators are supposed to be of the form

$$\mathcal{B}^0 = \mathbb{I}, \quad \mathcal{B}^j = \partial_{\vec{n}}^j, \quad \forall j = 1, \ldots, \mathrm{k}-1, \tag{2.18}$$

and $\Gamma^0 = \ldots = \Gamma^{\mathrm{k}-1} = \partial\,\Omega$, as well as $\Gamma^{\mathrm{k}} = \ldots = \Gamma^{2\,\mathrm{k}-1} = \varnothing$, meaning that they explicitly fix the value of terms appearing due to a repeated application of integration by parts, and in the

simplest case these values are zero. Supposing that the values of the i-th normal derivatives for $i = 0, \ldots, \mathrm{k}-1$ are prescribed globally, the resulting problem reads as follows.

Problem 3: General PDE of even order with homogeneous Dirichlet b.c.

Let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ open and bounded with a sufficiently smooth boundary. Let $\mathcal{L}$ a differential operator of even order $2\,\mathrm{k} \in \mathbb{N}$ and let $\mathrm{n} \in \{1, d\}$. Assume that $f : \Omega \to \mathbb{R}^{\mathrm{n}}$ is sufficiently smooth. Find a function $u \in [\mathcal{C}^{2\,\mathrm{k}}(\Omega)]^{\mathrm{n}}$ satisfying

$$\begin{aligned} \mathcal{L}[u_j]_{j=1}^{\mathrm{n}} &= f, \quad \text{in } \Omega \\ u_j &= 0, \quad \text{on } \partial\,\Omega \\ \partial_{\vec{n}}^i u_j &= 0, \quad \text{on } \partial\,\Omega \end{aligned} \tag{2.19}$$

for all $i = 0, \ldots, \mathrm{k}-1$ and $j = 1, \ldots, \mathrm{n}$.

For operators of even order, a property that is stronger than ellipticity and that will be important in the upcoming part of the theory is the so-called *uniform ellipticity*. Again, uniform ellipticity is first described for scalar linear differential operators and extended to the vector-valued case afterwards.

Definition 4 (Uniform ellipticity of scalar linear differential operators)**.** *A scalar linear differential operator of even order $2k$ of the form (2.7) is called* uniformly elliptic*, if there exists a constant $\mathrm{C}_e^{\mathcal{L}} \in \mathbb{R}_+$ such that for all $x \in \Omega$ and all $\xi \in \mathbb{R}^d \backslash \{0\}$ it holds that*

$$\sum_{\substack{\alpha \in \mathbb{N}_0^d \\ |\alpha| = 2k}} a_\alpha(x) \xi^\alpha \geqslant \mathrm{C}_e^{\mathcal{L}} \, |\xi|^{2k} \tag{2.20}$$

○

Definition 5 (Uniform ellipticity for vector-valued linear differential operators)**.** *A vector-valued linear differential operator of even order $2k$ of the form (2.7) with $\mathrm{n} = d$ is called* uniformly elliptic*, if there exists a constant $\mathrm{C}_e^{\mathcal{L}} \in \mathbb{R}_+$ such that for all $x \in \Omega$ and all $\xi \in \mathbb{R}^d \backslash \{0\}$ and all $i = 1, \ldots, d$ it holds that*

$$\sum_{\substack{\alpha^{i,1} \in \mathbb{R}^d \\ |\alpha^{i,1}| = 2k}} a_{\alpha^{i,1}} \xi^{\alpha^{i,1}} + \ldots + \sum_{\substack{\alpha^{i,d} \in \mathbb{R}^d \\ |\alpha^{i,d}| = 2k}} a_{\alpha^{i,d}} \xi^{\alpha^{i,d}} \geqslant \mathrm{C}_e^{\mathcal{L}} \, |\xi|^{2k}. \tag{2.21}$$

○

Remark 2.7. *In the situation described in Remark 2.3 for a symmetric coefficient $A = [A_{i,j}]_{i,j=1}^d$, as well as $b = \begin{bmatrix} 0 & \ldots & 0 \end{bmatrix}$ and $c = 0$, uniform ellipticity of the PDE operator*

$$\mathcal{L}(u) = -\operatorname{div}(A \nabla u) \tag{2.22}$$

is assured whenever $A(x)$ is symmetric and positive definite for all $x \in \Omega$, i.e.

$$\xi^T A(x) \xi \geqslant |\xi|^2, \quad \forall \xi \in \mathbb{R}^d, \, \forall x \in \Omega\,. \tag{2.23}$$

In this case, the map $\langle \cdot, \cdot \rangle_{A(x)} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ *defined by* $\langle \xi_1, \xi_2 \rangle := \xi_1^T A(x) \xi_2$ *defines a scalar product, and since all scalar products on* $\mathbb{R}^d$ *are equivalent, it holds that*

$$\begin{aligned} \langle \xi, \xi \rangle_{A(x)} &= \xi^T A(x) \xi \\ &= \begin{bmatrix} \xi_1 & \dots & \xi_d \end{bmatrix} \begin{bmatrix} \sum_{j=1}^d A_{1,j}(x) \xi_1 \\ \vdots \\ \sum_{j=1}^d A_{d,j}(x) \xi_d \end{bmatrix} \\ &= \sum_{i,j=1}^d A_{i,j} \xi_i \xi_j \geqslant c(x) |\xi|^2 \end{aligned} \tag{2.24}$$

for all $x \in \Omega$*, constants* $0 < c(x) \in \mathbb{R}$ *and all* $\xi \in \mathbb{R}^d \backslash \{0\}$*. By taking* $c = \min_{x \in \Omega} c(x)$*, the result matches the definition of uniform ellipticity.*

*

Remark 2.8. *Consider a vector-valued linear differential operator* $\mathcal{L}$ *of even order* $2\,\mathrm{k}$ *for* $\mathrm{n} = d$ *in the form (2.1). If the scalar linear differential operators* $\mathcal{L}_{i,j}$ *for* $i, j = 1, \dots, d$ *are uniformly elliptic with constant* $\mathrm{C}_e^{\mathcal{L}_{i,j}} \in \mathbb{R}_+$*, then also* $\mathcal{L}$ *is elliptic with uniform ellipticity constant*

$$\mathrm{C}_e^{\mathcal{L}} := \min_{i=1}^d \sum_{j=1}^d \mathrm{C}_e^{\mathcal{L}_{i,j}} \tag{2.25}$$

*

So far, the solution u of Problem 3 needs to be $2\,\mathrm{k}$ times continuously differentiable. In order to relax this very strong regularity condition, the problem is transformed into a variational (weak) formulation, as will be seen in the next section.

2.2.1 Variational formulation

Problem 3, i.e. the problem of finding a solution to the $2\,\mathrm{k}$-th order linear partial differential equation with homogeneous Dirichlet boundary conditions, is stated in its so-called *strong form*, meaning that a solution must satisfy *strong* regularity conditions such as being $2\,\mathrm{k}$ times continuously differentiable. In this section, the PDE is transformed into its corresponding *variational formulation* (or *weak form*). This variational formulation relaxes the conditions imposed on solutions and can be solved using numerical schemes such as the Partition of Unity Method (PUM), which will be presented in Chapter 4. The Dirichlet boundary conditions in Problem 3 assure that the solution u and all components of all derivatives of u of order up to $\mathrm{k}-1$ vanish on the whole boundary. The boundary conditions are incorporated into the approximation space, leading to the condition $u \in [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}}$, and the problem can be written in a more compact form.

Problem 4: Reformulation of Problem 3.

Let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ open and bounded with a piecewise smooth boundary. Let $\mathcal{L}$ be a differential operator of even order $2\,\mathrm{k} \in \mathbb{N}$ and let $\mathrm{n} \in \{1, d\}$. Assume that $f : \Omega \to \mathbb{R}^{\mathrm{n}}$ is sufficiently smooth. Find a function $u \in [\mathcal{C}^{2\,\mathrm{k}}(\Omega)]^{\mathrm{n}} \cap [\mathrm{H}^{\mathrm{k}}_0(\Omega)]^{\mathrm{n}}$ satisfying

$$\mathcal{L}\,u = f, \quad \text{in } \Omega\,. \tag{2.26}$$

The variational (weak) formulation of Problem 4 can now be developed. For this purpose, assume that the operator $\mathcal{L}$ appearing in the formulation is of even order $2k$, and that its coefficients are sufficiently smooth. Starting from (2.26), multiply the left-hand side $\mathcal{L}\,u$ by a test function $v \in [\mathcal{C}^{\infty}_0(\Omega)]^{\mathrm{n}}$ and integrate over the domain Ω. Afterwards, repeatedly integrate by parts to move k partial derivatives to the test function v. Since v has compact support, the boundary integrals over $\partial\,\Omega$ appearing due to the integration by parts vanish. The calculation reads

$$\begin{aligned}
\big\langle \mathcal{L}u, v \big\rangle_{\mathrm{L}^2(\Omega)} &= \int_\Omega \mathcal{L}\,u \cdot v \, dx \\
&= \int_\Omega \begin{bmatrix} \mathcal{L}_{1,1} & \dots & \mathcal{L}_{1,\mathrm{n}} \\ \vdots & \ddots & \vdots \\ \mathcal{L}_{\mathrm{n},1} & \dots & \mathcal{L}_{\mathrm{n},\mathrm{n}} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_{\mathrm{n}} \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ \vdots \\ v_{\mathrm{n}} \end{bmatrix} dx \\
&= \int_\Omega \sum_{i,j=1}^{\mathrm{n}} \mathcal{L}_{i,j}\, u_j v_i \, dx \\
&= \int_\Omega \sum_{i,j=1}^{\mathrm{n}} \sum_{\substack{\alpha^{i,j} \in \mathbb{N}_0^d \\ |\alpha^{i,j}| \leqslant \mathrm{k}}} \sum_{\substack{\beta^{i,j} \in \mathbb{N}_0^d \\ |\beta^{i,j}| \leqslant \mathrm{k}}} v_i a_{\alpha^{i,j},\beta^{i,j}} \partial^{\alpha^{i,j}+\beta^{i,j}} u_j \, dx \\
&= \sum_{i,j=1}^{\mathrm{n}} \sum_{\substack{\alpha^{i,j} \in \mathbb{N}_0^d \\ |\alpha^{i,j}| \leqslant \mathrm{k}}} \sum_{\substack{\beta^{i,j} \in \mathbb{N}_0^d \\ |\beta^{i,j}| \leqslant \mathrm{k}}} (-1)^{|\beta^{i,j}|} \int_\Omega \partial^{\beta^{i,j}} (a_{\alpha^{i,j},\beta^{i,j}} v_i) \partial^{\alpha^{i,j}} u_j \, dx
\end{aligned} \tag{2.27}$$

Analogously, the right-hand side of the PDE is multiplied by the test function and integrated over Ω. Note that the right-hand side of eq. (2.27) does not require u to be $2\,\mathrm{k}$-times continuously differentiable anymore, but it only requires u to provide k weak derivatives. This allows to state variational (weak) formulation of Problem 4.

Problem 5: Weak formulation of uniformly elliptic PDE of even order with homogeneous Dirichlet b.c.

Let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ be open and bounded with a piecewise smooth boundary. Let $\mathcal{L}$ a uniform elliptic differential operator of even order $2\,\mathrm{k} \in \mathbb{N}$ and let $\mathrm{n} \in \{1, d\}$. Assume $f \in [\mathrm{L}^2(\Omega)]^{\mathrm{n}}$. Find a function $u \in [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}}$ satisfying

$$\mathbf{a}[u, v] = \ell(v), \quad \forall v \in [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}} \tag{2.28}$$

with bilinear form $\mathbf{a} : [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}} \times [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}} \to \mathbb{R}$ and linear functional $\ell : [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}} \to \mathbb{R}$ given by

$$\begin{aligned} \mathbf{a}[u, v] &:= \sum_{i,j=1}^{\mathrm{n}} \sum_{\substack{\alpha^{i,j} \in \mathbb{N}_0^d \\ |\alpha^{i,j}| \leqslant \mathrm{k}}} \sum_{\substack{\beta^{i,j} \in \mathbb{N}_0^d \\ |\beta^{i,j}| \leqslant \mathrm{k}}} (-1)^{|\beta^{i,j}|} \int_\Omega \partial^{\beta^{i,j}} (a_{\alpha^{i,j},\beta^{i,j}} v_i) \partial^{\alpha^{i,j}} u_j \, dx \\ \ell(v) &:= \int_\Omega f \cdot v \, dx \end{aligned} \tag{2.29}$$

Remark 2.9. *Even if the regularity assumptions on the solution were weakened in the variational formulation, any strong solution $u \in [\mathcal{C}^{2\,\mathrm{k}}(\Omega)]^{\mathrm{n}}$ of Problem 4, with a differential operator $\mathcal{L}$ of order $2\,\mathrm{k}$ having sufficiently regular coefficients, also solves Problem 5.*

*

Remark 2.10. *In case that the coefficients are constant and can be written in a product form,*

$$a_{\alpha^{i,j},\beta^{i,j}} = a_{\alpha^{i,j}} a_{\beta^{i,j}}, \quad \forall \alpha^{i,j}, \beta^{i,j} \in \mathbb{N}_0^d, \, |\alpha^{i,j}|, |\beta^{i,j}| \leqslant \mathrm{k} \tag{2.30}$$

and for all $i, j = 1, \ldots, n$, then eq. (2.27) *can be rewritten as*

$$\langle \mathcal{L}\, u, v \rangle_{\mathrm{L}^2(\Omega)} = \sum_{i,j=1}^{\mathrm{n}} \langle \sum_{\substack{\alpha^{i,j} \in \mathbb{N}_0^d \\ |\alpha^{i,j}| \leqslant \mathrm{k}}} a_{\alpha^{i,j}} \partial^{\alpha^{i,j}} u_j, \sum_{\substack{\beta^{i,j} \in \mathbb{N}_0^d \\ |\beta^{i,j}| \leqslant \mathrm{k}}} (-1)^{|\beta^{i,j}|} a_{\beta^{i,j}} \partial^{\beta^{i,j}} v_i \rangle_{\mathrm{L}^2(\Omega)}$$

that is

$$\langle \mathcal{L}\, u, v \rangle_{\mathrm{L}^2(\Omega)} = \sum_{i,j=1}^{n} \langle \mathcal{L}_{i,j}^{\mathit{trial}} u_j, \mathcal{L}_{i,j}^{\mathit{test}} v_i \rangle_{\mathrm{L}^2(\Omega)} \tag{2.31}$$

for all $v \in [\mathcal{C}_0^\infty(\Omega)]^{\mathrm{n}}$, being the sum of inner products of linear differential operators that are applied only to components of the trial and test functions. These operators read

$$\begin{aligned} \mathcal{L}_{i,j}^{\mathit{trial}} &= \sum_{\substack{\alpha^{i,j} \in \mathbb{N}_0^d \\ |\alpha^{i,j}| \leqslant \mathrm{k}}} a_{\alpha^{i,j}} \partial^{\alpha^{i,j}} \\ \mathcal{L}_{i,j}^{\mathit{test}} &= \sum_{\substack{\beta^{i,j} \in \mathbb{N}_0^d \\ |\beta^{i,j}| \leqslant \mathrm{k}}} (-1)^{|\beta^{i,j}|} a_{\beta^{i,j}} \partial^{\beta^{i,j}}, \end{aligned} \tag{2.32}$$

for all $i, j = 1, \ldots, \mathrm{n}$. *Several classical partial differential equations, among them the Laplace and Poisson equations, can be written in this way.*

*

Remark 2.11. *In the scalar case* $\mathrm{n} = 1$*, after dropping unused indices, Remark 2.10 simplifies to*

$$\langle \mathcal{L}\, u, v\rangle_{\mathrm{L}^2(\Omega)} = \langle \mathcal{L}^{\mathit{trial}}\, u, \mathcal{L}^{\mathit{test}}\, v\rangle_{\mathrm{L}^2(\Omega)}, \quad \forall v \in \mathcal{C}_0^\infty(\Omega). \tag{2.33}$$

$$\mathcal{L}^{\mathit{trial}} = \sum_{\substack{\alpha\in\mathbb{N}_0^d \\ |\alpha|\leqslant \mathrm{k}}} a_\alpha \partial^\alpha, \quad \mathcal{L}^{\mathit{test}} = \sum_{\substack{\beta\in\mathbb{N}_0^d \\ |\beta|\leqslant \mathrm{k}}} (-1)^{|\beta|} a_\beta \partial^\beta. \tag{2.34}$$

*

The variational formulation has been set up. In the next section, existence and uniqueness of solutions will be investigated.

2.2.2 Existence and uniqueness of weak solutions

This section presents definitions and theoretical results related to the existence and uniqueness of solutions of the variational Problem 5 introduced in the previous section. In particular, continuity and ellipticity of the bilinear form and continuity of the linear functional are sufficient to guarantee a unique solution due to the theorem of Lax and Milgram presented in further below. Ellipticity is a very strong condition which under some circumstances can be weakened to coercivity. This may still be enough to guarantee existence and uniqueness of a weak solution.

Continuity of linear and bilinear forms

Continuity is equivalent to boundedness for linear functionals as well as bilinear forms.

Definition 6 (Boundedness of linear functionals)**.** *Let* $(W, \|\cdot\|_W)$ *be a normed space. Then, the linear functional* $\ell : W \to \mathbb{R}$ *is* bounded*, if there exists a constant* $\mathrm{C}^\ell_{\mathit{cont}} \in \mathbb{R}_+$ *such that*

$$\ell(v) \leqslant \mathrm{C}^\ell_{\mathit{cont}} \|v\|_W, \quad \forall v \in W. \tag{2.35}$$

○

Definition 7 (Boundedness of bilinear forms)**.** *Let* $(W, \|\cdot\|_W)$ *be a normed space. Then, the bilinear form* $\mathbf{a} : W \times W \to \mathbb{R}$ *is* bounded *if there exists a constant* $\mathrm{C}_{\mathit{cont}} \in \mathbb{R}_+$ *satisfying*

$$\mathbf{a}[u, v] \leqslant \mathrm{C}_{\mathit{cont}} \|u\|_W \|v\|_W, \quad \forall u, v \in W. \tag{2.36}$$

○

Note that the linear functional ℓ appearing in Problem 5 is bounded due to

$$\ell(v) = \int_\Omega f \cdot v \, dx \leqslant \|f\|_{[\mathrm{L}^2(\Omega)]^\mathrm{n}} \|v\|_{[\mathrm{L}^2(\Omega)]^\mathrm{n}} \leqslant \|f\|_{[L^2(\Omega)]^\mathrm{n}} \|v\|_{[\mathrm{H}^\mathrm{k}(\Omega)]^\mathrm{n}}, \tag{2.37}$$

for all test functions $v \in [\mathrm{H}^\mathrm{k}(\Omega)]^\mathrm{n}$, that is $\mathrm{C}^\ell_{\mathrm{cont}} = \|f\|_{[\mathrm{L}^2(\Omega)]^\mathrm{n}}$. The bilinear form appearing in Problem 5 is continuous under mild assumptions.

Lemma 2.1. *Let the assumptions from Problem 5 hold and let* $\mathbf{a} : [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}} \times [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}} \to \mathbb{R}$ *be the bilinear form appearing in the variational formulation 2.28. Suppose that the coefficient space is* $\mathcal{X}(\Omega) = \mathrm{L}^\infty(\Omega)$*. Then,* $\mathbf{a}$ *is continuous in the sense that*

$$\mathbf{a}[u, v] \leqslant \mathrm{C}_{cont} \, \|u\|_{[\mathrm{H}^{\mathrm{k}}(\Omega)]^n} \|v\|_{[\mathrm{H}^{\mathrm{k}}(\Omega)]^n} \tag{2.38}$$

for all $u, v \in [\mathrm{H}^{\mathrm{k}}(\Omega)]^{\mathrm{n}}$*, with*

$$\mathrm{C}_{cont} \coloneqq \sum_{i,j=1}^{\mathrm{n}} \sum_{\substack{\alpha^{i,j} \in \mathbb{N}_0^d \\ |\alpha^{i,j}| \leqslant \mathrm{k}}} \sum_{\substack{\beta^{i,j} \in \mathbb{N}_0^d \\ |\beta^{i,j}| \leqslant \mathrm{k}}} \|a_{\alpha^{i,j},\beta^{i,j}}\|_{\mathrm{L}^\infty(\Omega)}. \tag{2.39}$$

•

Proof. Since the coefficients are in $\mathcal{X}(\Omega) = \mathrm{L}^\infty(\Omega)$ and all components of u and v are in $\mathrm{L}^2(\Omega)$, the Hölder inequality is applied to see that for all $i, j = 1, \ldots, \mathrm{n}$ and all $\alpha^{i,j}, \beta^{i,j} \in \mathbb{N}_0^d$ it holds that

$$\begin{aligned} &\int_\Omega (-1)^{|\beta^{i,j}|} a_{\alpha^{i,j},\beta^{i,j}} \partial^{\alpha^{i,j}} u_j \partial^{\beta^{i,j}} v_i \, dx \\ &\qquad \leqslant \|a_{\alpha^{i,j},\beta^{i,j}}\|_{\mathrm{L}^\infty(\Omega)} \|\partial^{\alpha^{i,j}} u_j\|_{\mathrm{L}^2(\Omega)} \|\partial^{\beta^{i,j}} v_i\|_{\mathrm{L}^2(\Omega)} \\ &\qquad \leqslant \|a_{\alpha^{i,j},\beta^{i,j}}\|_{\mathrm{L}^\infty(\Omega)} \|u\|_{[\mathrm{H}^{\mathrm{k}}(\Omega)]^{\mathrm{n}}} \|v\|_{[\mathrm{H}^{\mathrm{k}}(\Omega)]^{\mathrm{n}}}. \end{aligned} \tag{2.40}$$

This estimate is used for all terms appearing in $\mathbf{a}[u, v]$ to see that the claim holds with

$$\mathrm{C}_{\mathrm{cont}} = \sum_{i,j=1}^{\mathrm{n}} \sum_{\substack{\alpha^{i,j} \in \mathbb{N}_0^d \\ |\alpha^{i,j}| \leqslant \mathrm{k}}} \sum_{\substack{\beta^{i,j} \in \mathbb{N}_0^d \\ |\beta^{i,j}| \leqslant \mathrm{k}}} \|a_{\alpha^{i,j},\beta^{i,j}}\|_{\mathrm{L}^\infty(\Omega)}. \tag{2.41}$$

□

Ellipticity & coercivity of bilinear forms

Showing ellipticity of bilinear forms is usually more involved. In case that the bilinear form $\mathbf{a}$ results from bringing a PDE with differential operator $\mathcal{L}$ into its variational formulation, it strongly depends on the coefficients of $\mathcal{L}$ whether $\mathbf{a}$ is elliptic.

Definition 8 (Ellipticity of bilinear forms)**.** *Let* $(V, \|\cdot\|_V)$ *be a Hilbert space. The bilinear form* $\mathbf{a} : V \times V \to \mathbb{R}$ *is called* elliptic*, if there exists a constant* $\mathrm{C}_e \in \mathbb{R}_+$ *such that*

$$\mathbf{a}[u, u] \geqslant \mathrm{C}_e \, \|u\|_V^2, \quad \forall u \in V. \tag{2.42}$$

○

Ellipticity of a bilinear form is a very strong condition and under some circumstances it is enough that the bilinear form is coercive. In the theory of linear operators on arbitrary Hilbert spaces $(V, \|\cdot\|_V)$, coercivity and ellipticity are defined identically and used interchangeably. If the bilinear form is defined on a Sobolev space, coercivity is defined as follows.

Definition 9 (Coercivity of bilinear forms)**.** *Consider a bilinear form* $\mathbf{a} : [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}} \times [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}} \to \mathbb{R}$*. Then,* $\mathbf{a}$ *is called* coercive*, if there exist constants* $\mathrm{C}_1^c, \mathrm{C}_2^c \in \mathbb{R}$ *with* $\mathrm{C}_1^c > 0$*, such that*

$$\mathbf{a}[u,u] \geqslant \mathrm{C}_1^c \, \|u\|^2_{[\mathrm{H}^{\mathrm{k}}(\Omega)]^{\mathrm{n}}} - \mathrm{C}_2^c \, \|u\|^2_{[\mathrm{L}^2(\Omega)]^{\mathrm{n}}}, \quad \forall u \in [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}}. \tag{2.43}$$

○

Unique solvability for elliptic bilinear forms

The Lax-Milgram Theorem connects continuity and ellipticity to guarantee the existence and uniqueness of solutions. The proof relies on the Riesz representation theorem.

Theorem 2.2 (Riesz)**.** *Let* $(V, \|\cdot\|_V)$ *a real Hilbert space. Then, the* Riesz operator

$$\mathcal{R} : V^\star \to V, \quad \textit{with } \ell\, v = \langle \mathcal{R}\,\ell, v\rangle_V, \quad \forall\, \ell \in V^\star,\, \forall v \in V \tag{2.44}$$

exists and satisfies $\|\,\ell\,\|_{V^\star} = \|\mathcal{R}\,\ell\,\|_V$*.*

◇

Proof. Define $\mathcal{R}$ piecewise. If $\ell = 0$ in $V^\star$, define $\mathcal{R}\,\ell = 0 \in V$. For all other $\ell \in V^\star$ denote its kernel by K_ℓ. Since ℓ is bounded, it is continuous. Note that $\{0\} \subset \mathbb{R}$ is a closed subspace, and continuous operators map closed subspaces to closed subspaces, implying that also K is closed. The space V can hence be decomposed as a direct sum,

$$V = K_\ell \oplus K_\ell^\perp. \tag{2.45}$$

Since ℓ is not the zero functional, $K_\ell^\perp \neq \{0\}$ must hold. Pick an element $k \in K_\ell^\perp$ with $\|k\|_V = 1$. Using the linearity of ℓ, for all $v \in V$ it holds that

$$\ell(k\,\ell\,v - v\,\ell\,k) = \ell\,k\,\ell\,v - \ell\,v\,\ell\,k = 0, \tag{2.46}$$

so

$$k\,\ell\,v - v\,\ell\,k \in \ker(\ell) = K_\ell. \tag{2.47}$$

The linearity of the scalar product, a null-addition of $v\,\ell\,k - v\,\ell\,k$, conjugate symmetry and the fact that $\ell : V \to \mathbb{R}$ show that

$$\begin{aligned}
\ell\,v &= \ell\,v\langle k,k\rangle_V = \langle k\,\ell\,v, k\rangle_V \\
&= \langle k\,\ell\,v - \underbrace{v\,\ell\,k + v\,\ell\,k}_{=0}, k\rangle_V \\
&= \underbrace{\langle k\,\ell\,v - v\,\ell\,k, k\rangle_V}_{\overset{(2.47)}{=}0} + \langle v\,\ell\,k, k\rangle_V \\
&= \langle v(\ell\,k), k\rangle_V \\
&= \langle v, \underbrace{(\ell\,k)k}_{=:\mathcal{R}\,\ell}\rangle_V.
\end{aligned} \tag{2.48}$$

Concluding, the Riesz operator $\mathcal{R}$ is defined as

$$\mathcal{R} : V^\star \to V, \quad \ell \mapsto \mathcal{R}\,\ell = \begin{cases} 0, & \text{if } \ell = 0 \in V^\star \\ (\ell\,k)k, & \text{else} \end{cases} \tag{2.49}$$

and this operator satisfies $\ell\, v = \langle \mathcal{R}\,\ell, v\rangle_V$ for all $\ell \in V^\star$ and all $v \in V$. The norm equality follows since for all $\ell \in V^\star$ it holds that

$$\|\mathcal{R}\,\ell\,\|_V = \|(\ell\, k)k\|_V = |\,\ell\, k|\,\underbrace{\|k\|_V}_{=1} = |\,\ell\, k| \leqslant \sup_{\substack{v\in V\\ \|v\|_V=1}} |\,\ell\, v| = \|\,\ell\,\|_{V^\star} \tag{2.50}$$

as well as

$$\begin{aligned} \|\,\ell\,\|_{V^\star} &= \sup_{\substack{v\in V\\ \|v\|_V=1}} |\,\ell\, v| \\ &= \sup_{\substack{v\in V\\ \|v\|_V=1}} |\langle \mathcal{R}\,\ell, v\rangle_V| \\ &\leqslant \sup_{\substack{v\in V\\ \|v\|_V=1}} \|v\|_V \|\mathcal{R}\,\ell\,\|_V \\ &= \|\mathcal{R}\,\ell\,\|_V \end{aligned} \tag{2.51}$$

so in total $\|\,\ell\,\|_{V^\star} = \|\mathcal{R}\,\ell\,\|_V$. □

Another component used in the proof of the Theorem of Lax and Milgram is the Banach fixed point Theorem.

Definition 10. *Let (X, d) a metric space. A mapping $K : X \to X$ is called a* contraction *on X, if there exists a constant $1 > \varrho \in \mathbb{R}_+$ such that*

$$d[K(x), K(y)] \leqslant \varrho\, d[x, y], \quad \forall x, y \in X. \tag{2.52}$$

○

This means that geometrically the images of two arbitrary points $x, y \in X$ are closer together than the points originally were.

Theorem 2.3 (Banach fixed point)**.** *Let (X, d) be a complete metric space and let $K : X \to X$ be a contraction on X with constant $1 > \varrho \in \mathbb{R}_+$. Then, K has a unique fixed point, i.e. there exists a unique $x \in X$ such that*

$$K(x) = x. \tag{2.53}$$

◇

Proof. Consider the sequence $\{x_i\}_{i=1}^{\infty}$ with $x_{i+1} = K(x_i)$ for an arbitrary $x_0 \in X$. First, it is shown that the sequence is a Cauchy sequence, then that it has a limit, and lastly that the limit is unique.
Step 1: Since K is a contraction,

$$\begin{aligned} d[x_{i+1],x_i] &= d[K(x_i), K(x_{i-1})] \\ &\leqslant \varrho\, d[x_i, x_{i-1}] \\ &= \varrho\, d[K(x_{i-1}), K(x_{i-2})] \\ &\leqslant \varrho^2\, d[x_{i-1}, x_{i-2}] \\ &\;\;\vdots \\ &\leqslant \varrho^i\, d[x_1, x_0] \end{aligned} \tag{2.54}$$

and using the triangle inequality for $j \geqslant i$ and the summation formula for a geometric sum, it holds that

$$\begin{aligned} d[x_i, x_j] &\leqslant d[x_i, x_{i+1}] + d[x_{i=1}, x_{i+2}] + \ldots + d[x_{j-1}, x_j] \\ &\leqslant (\varrho^i + \varrho^{i+1} + \ldots + \varrho^{j-1}) d[x_1, x_0] \\ &= \varrho^i \frac{1-\varrho^{j-i}}{1-\varrho} d[x_1, x_0]. \end{aligned} \tag{2.55}$$

Since $0 < \varrho < 1$ it holds that $1 - \varrho^{j-i} < 1$, implying that

$$d[x_i, x_j] \leqslant \frac{\varrho^i}{1-\varrho} d[x_1, x_0]. \tag{2.56}$$

From (2.56) it is seen that the right hand side can be made arbitrarily small by choosing i sufficiently large and $j > i$. This proves, that $\{x_i\}_{i=1}^{\infty}$ is a Cauchy sequence. Since the space (X, d) is complete, the limit of the sequence, $x_i \overset{i\to\infty}{\to} x$ exists in X.
Step 2: The limit of the sequence is a fixed point. This follows from a use of the triangle inequality

$$\begin{aligned} d[x, K(x)] &\leqslant d[x, x_i] + d[x_i, K(x)] = d[x, x_i] + d[K(x_{i-1}), K(x)] \\ &\leqslant d[x, x_i] + \varrho\, d[x_{i-1}, x] \end{aligned} \tag{2.57}$$

and since $x_i \to x$ the right-hand side can be made arbitrarily small by choosing i large enough. It follows that

$$d[x, K(x)] = 0 \quad \Rightarrow \quad K(x) = x, \tag{2.58}$$

so $x \in X$ is a fixed point of K.
Step 3: Suppose there exists another fixed point $\tilde{x}$. Since K is a contraction it follows that

$$d[x, \tilde{x}] = d[K(x), K(\tilde{x})] \leqslant \varrho\, d[x, \tilde{x}] \tag{2.59}$$

implying that $d[x, \tilde{x}] = 0$ since $\varrho < 1$. The fixed points must hence coincide. □

The previous Theorems 2.2 and 2.3 finally allow to state and prove the Theorem of Lax and Milgram.

Theorem 2.4 (Lax-Milgram)**.** *Let $(V, \|\cdot\|_V)$ be a Hilbert space and $\mathbf{a} : V \times V \to \mathbb{R}$ a bilinear form. If a is continuous with constant $\mathrm{C}_{cont} \in \mathbb{R}_+$, then there exists a unique continuous linear operator $A : V \to V$ satisfying*

$$\mathbf{a}[u, v] = \langle Au, v \rangle_V, \quad \forall u, v \in V, \tag{2.60}$$

and

$$\|A\| := \sup_{\substack{u \in V \\ \|u\|_V = 1}} \|A(u)\|_V \leqslant \mathrm{C}_{cont}\,. \tag{2.61}$$

If moreover a is elliptic with constant $\mathrm{C}_e \in \mathbb{R}_+$, then the linear operator A is invertible with

$$\|A^{-1}\| \leqslant {\mathrm{C}_e}^{-1}, \tag{2.62}$$

implying that for any bounded linear operator $\ell : V \to V$ the equation

$$\mathbf{a}[u, v] = \ell(v), \quad \forall v \in V \tag{2.63}$$

is uniquely solvable.

◇

Proof. Let $A : V \to V^\star$ be the map defined via

$$A : V \to V^\star, \quad u \mapsto Au := \mathbf{a}[u, \cdot]. \tag{2.64}$$

This map is continuous, since a is continuous. Let $\mathcal{R} : V^\star \to V$ be the Riesz operator, satisfying

$$Auv = \langle v, \mathcal{R}A\rangle_V, \quad \forall v \in V. \tag{2.65}$$

Using

$$\langle \mathcal{R}Au, v\rangle_V = Au(v) = \mathbf{a}[u, v], \quad \forall u, v \in V \tag{2.66}$$

it holds that

$$\begin{aligned} \mathrm{C_e}\,\|v\|_V^2 \leqslant \mathbf{a}[v, v] \quad &= \quad |\langle \mathcal{R}A, v\rangle_V| \leqslant \|\mathcal{R}A\|_V \|v\|_V \\ \overset{\|\mathcal{R}\|_{V^\star}=1}{=} \quad & \|Av\|_{V^\star}\|v\|_V \leqslant \|A\|_{V^\star}\|v\|_V^2, \end{aligned} \tag{2.67}$$

implying

$$\mathrm{C_e} \leqslant \|A\|_{V^\star} \quad \Rightarrow \quad 0 < \frac{\mathrm{C_e}}{\|A\|_{V^\star}} \leqslant 1. \tag{2.68}$$

For $\varrho \in \mathbb{R}$ define the map

$$K : V \to V, \quad u \mapsto K(u) := u - \varrho(\mathcal{R}Au - \mathcal{R}\,\ell). \tag{2.69}$$

Note that $u^\star \in V$ is a fix point of K if and only if

$$\mathcal{R}Au^\star - \mathcal{R}\,\ell = 0, \tag{2.70}$$

meaning that

$$\underbrace{\langle \mathcal{R}Au^\star, v\rangle_V}_{=\mathbf{a}[u^\star, v]} = \underbrace{\langle \mathcal{R}\,\ell, v\rangle_V}_{=\ell(v)}, \quad \forall v \in V, \tag{2.71}$$

so

$$\langle \mathcal{R}Au^\star - \mathcal{R}\,\ell, v\rangle_V = 0, \quad \forall v \in V. \tag{2.72}$$

Choose $v = \mathcal{R}Au^\star - \mathcal{R}\,\ell \in V$ to see that

$$\|\mathcal{R}Au^\star - \mathcal{R}\,\ell\,\|_V^2 = 0 \quad \Rightarrow \quad \mathcal{R}Au^\star = \mathcal{R}\,\ell, \tag{2.73}$$

showing that $u^\star$ is a solution to Problem 5. It remains to show, that such a fix point $u^\star$ of K exists. For $v_1, v_2 \in V$ and $v = v_1 - v_2$ it holds that

$$\begin{aligned} \|K(v_1) - K(v_2)\|_V^2 &= \langle v - \varrho\mathcal{R}Av, v - \varrho\mathcal{R}Av\rangle_V \\ &= \|v\|_V^2 - 2\varrho\langle \mathcal{R}Av, v\rangle_V + \varrho^2\|\mathcal{R}Av\|_V^2 \\ &= \|v\|_V^2 - 2\varrho\,\mathbf{a}[v, v] + \varrho^2\,\mathbf{a}[v, v] \\ &\leqslant \|v\|_V^2 - 2\varrho\,\mathrm{C_e}\,\|v\|_V^2 + \varrho^2\|A\|_{V^\star}\|v\|_V\|\mathcal{R}Av\|_V \\ &\leqslant \|v\|_V^2 - 2\varrho\,\mathrm{C_e}\,\|v\|_V^2 + \varrho^2\|A\|_{V^\star}\|v\|_V^2 \\ &= \|v\|_V^2(1 - 2\varrho\,\mathrm{C_e} + \varrho^2\|A\|_{V^\star}). \end{aligned} \tag{2.74}$$

Finally, for $\varrho = \frac{\mathrm{C_e}}{\|A\|^2_{V^\star}}$ one obtains

$$\begin{aligned} 1 - 2\varrho\,\mathrm{C_e} + \varrho^2\|A\|_{V^\star} &= 1 - 2\frac{\mathrm{C_e}}{\|A\|^2_{V^\star}}\,\mathrm{C_e} + \left(\frac{\mathrm{C_e}}{\|A\|^2_{V^\star}}\right)^2 \|A\|^2_{V^\star} \\ &= 1 - \frac{{\mathrm{C_e}}^2}{\|A\|^2_{V^\star}} \\ &< 1, \end{aligned} \tag{2.75}$$

where the last inequality holds since (2.68) implies $0 < \frac{{\mathrm{C_e}}^2}{\|A\|^2_{V^\star}} \leqslant 1$. This shows that K, for ϱ defined as above, is a contraction. Due to Theorem 2.3, it has a unique fixed point. Concluding, eq. (2.63) is uniquely solvable. □

The Theorem of Lax and Milgram can be applied directly to the operators defined in Problem 5. Whenever the bilinear form is continuous and elliptic, a unique solution $u_f \in [\mathrm{H^k}(\Omega)]^\mathrm{n}$ exists. An a priori error estimate of the approximation error using a finite dimensional subspace $V_h \subset V$ is given in Céas Lemma 2.5.

Lemma 2.5 (Céa)**.** *Let $(V, \|\cdot\|_V)$ be a real Hilbert space, $V_h \subset V$ a finite-dimensional subspace and let $\mathbf{a} : V \times V \to \mathbb{R}$ be an elliptic and continuous bilinear form with constants $0 < \mathrm{C}_e \leqslant \mathrm{C}_{cont}$,*

$$\begin{aligned} \mathbf{a}[u, v] &\leqslant \mathrm{C}_{cont}\,\|u\|_V\|v\|_V, \quad \forall u, v \in V \\ \mathbf{a}[u, u] &\geqslant \mathrm{C}_e\,\|u\|^2_V, \quad \forall u \in V. \end{aligned} \tag{2.76}$$

Let $u \in V$ and $u_h \in V_h$ be the solutions of

$$\mathbf{a}[u, v] = \ell(v), \quad \forall v \in V \tag{2.77}$$

respectively

$$\mathbf{a}[u_h, v_h] = \ell(v_h), \quad \forall v_h \in V_h, \tag{2.78}$$

which exist due to the Lax-Milgram Theorem 2.4. Then it holds that

$$\|u - u_h\|_V \frac{\mathrm{C}_{cont}}{\mathrm{C}_e} \inf_{v_h \in V_h} \|u - v_h\|_V. \tag{2.79}$$

•

Proof. Since $V_h \subset V$ it holds that

$$\mathbf{a}[u_h, v_h] - \mathbf{a}[u, v_h] = \ell(v_h) - \ell(v_h) = 0, \quad \forall v_h \in V_h. \tag{2.80}$$

Hence, also

$$\begin{aligned} \mathbf{a}[u - u_h, u - u_h] &= \mathbf{a}[u - u_h, u - v_h + v_h - u_h] \\ &= \mathbf{a}[u - u_h, u - v_h] + \underbrace{\mathbf{a}[u - u_h, v_h - u_h]}_{=0 \text{ by } (2.80)} \\ &= \mathbf{a}[u - u_h, u - v_h]. \end{aligned} \tag{2.81}$$

This yields

$$\begin{aligned} \mathrm{C_e}\,\|u-u_h\|_V^2 &= \mathbf{a}[u-u_h,u-u_h] \\ &\overset{(2.81)}{=} \mathbf{a}[u-u_h,u-v_h] \\ &\leqslant \mathrm{C_{cont}}\,\|u-u_h\|_V\|u-v_h\|_V, \end{aligned} \tag{2.82}$$

which in turn implies

$$\|u-u_h\|_V \leqslant \frac{\mathrm{C_{cont}}}{\mathrm{C_e}}\|u-v_h\|_V, \quad \forall v_h \in V_h, \tag{2.83}$$

and since $v_h \in V_h$ is arbitrary, this concludes the proof. □

Ellipticity of the bilinear form is oftentimes shown by hand. This is done by invoking a suitable type of Poincaré Friedrichs inequality, which helps to provide one part of the norm equivalence between the standard Sobolev norm, and the energy norm defined by the bilinear form on a certain subspace. In case that the bilinear form $\mathbf{a}[\cdot,\cdot]$, corresponding to the partial differential operator $\mathcal{L}$ of order $2\,\mathrm{k}$ under study, has a nontrivial kernel $\mathcal{K}(\mathcal{L})$, the bilinear form is no scalar product on the full Sobolev space $[\mathrm{H}^{\mathrm{k}}(\Omega)]^{\mathrm{n}}$, but only on the quotient space $[\mathrm{H}^{\mathrm{k}}(\Omega)]^{\mathrm{n}}_{/\mathcal{K}(\mathcal{L})}$, and the variant of Poincaré Friedrichs inequality that needs to be invoked in this case should show that

$$\mathbf{a}[u,u] \overset{!}{\geqslant} \mathrm{C_e}\,\|u\|^2_{[\mathrm{H}^{\mathrm{k}}(\Omega)]^{\mathrm{n}}}, \quad \forall u \in [\mathrm{H}^{\mathrm{k}}(\Omega)]^{\mathrm{n}}_{\mathcal{K}(\mathcal{L})}. \tag{2.84}$$

For the scalar, pure Dirichlet Problem 5, the corresponding Poincaré Friedrichs inequality is presented in Theorem 2.6.

Theorem 2.6 (Poincaré Friedrichs inequality)**.** *For $d \in \mathbb{N}$, let $\Omega \subset \mathbb{R}^d$ be bounded. Then, there exists a constant $\mathrm{C}_{PF} > 0$ such that*

$$\mathrm{C}_{PF}\left(\sum_{\substack{\alpha\in\mathbb{R}^d \\ |\alpha|=\mathrm{k}}} \|\partial^\alpha u\|^2_{\mathrm{L}^2(\Omega)}\right)^{\frac{1}{2}} \geqslant \|u\|_{\mathrm{L}^2(\Omega)}, \quad \forall u \in \mathrm{H}^{\mathrm{k}}_0(\Omega). \tag{2.85}$$

◇

Remark 2.12. *The term appearing in the brackets on the left-hand side of (2.85) is the Sobolev seminorm. In the case of second-order scalar partial differential equations, (2.85) can be written as*

$$\mathrm{C}_{PF}\,\|\nabla u\|_{[\mathrm{L}(\Omega)]^d} \geqslant \|u\|_{\mathrm{L}^2(\Omega)}, \quad \forall u \in \mathrm{H}^1(\Omega). \tag{2.86}$$

*

Unique solvability for coercive bilinear forms

As stated before, ellipticity is a very strong condition that can under some circumstances be weakened to coercivity. Conditions for the unique solvability of Problem 5 with a coercive bilinear form are presented in Theorem 2.7.

Theorem 2.7 (Coercivity and unique solvability)**.** *Let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ be open and bounded. Also, let $\mathbf{a} : [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^n \times [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}} \to \mathbb{R}$ be coercive with constants $\mathrm{C}_1^c, \mathrm{C}_2^c \in \mathbb{R}$ with $\mathrm{C}_1^c > 0$, and let $\ell : [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}} \to \mathbb{R}$ be a linear form. Then, the problem*

$$\mathbf{a}[u, v] = \ell[v], \quad \forall v \in [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}} \tag{2.87}$$

has a unique solution $u \in [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}}$, if and only if the kernels of both linear functionals $u \mapsto \mathbf{a}[u, \cdot]$ and $v \mapsto \mathbf{a}[\cdot, v]$ only contain the zero function, i.e.

$$\begin{aligned} \dim\{w \in [\mathrm{H}^{\mathrm{k}}(\Omega)]^{\mathrm{n}} : \quad \mathbf{a}[w, v] = 0, \quad \forall v \in [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}}\} &= 0 \\ \dim\{w \in [\mathrm{H}^{\mathrm{k}}(\Omega)]^{\mathrm{n}} : \quad \mathbf{a}[u, w] = 0, \quad \forall u \in [\mathrm{H}_0^{\mathrm{k}}(\Omega)]^{\mathrm{n}}\} &= 0. \end{aligned} \tag{2.88}$$

◇

Proof. The proof is very technical and can be found in [Hac17, Theorem 6.108 & Theorem 7.14]. □

2.2.3 Conditions for ellipticity of the bilinear form

In the previous section, the abstract concepts of ellipticity and coercivity have been introduced, which due to Theorem 2.4 and Theorem 2.7 are directly linked to the solvability of Problem 5 and the uniqueness of its solutions. In this section, assumptions on the coefficients of the differential operator are reviewed, which ensure ellipticity of the corresponding bilinear form $\mathbf{a}[\cdot, \cdot]$. Again, the theory is based on [Hac17] with some additional information and modifications for vector-valued problems.

Theorem 2.8 (Ellipticity of bilinear form I)**.** *Let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ open and bounded with a piecewise smooth boundary. Consider a scalar second-order linear differential operator $\mathcal{L}$ that is uniformly elliptic with constant $\mathrm{C}_e^{\mathcal{L}}$. Also suppose that the bilinear form $\mathbf{a} : \mathrm{H}_0^1(\Omega) \times \mathrm{H}_0^1(\Omega) \to \mathbb{R}$ appearing in the pure Dirichlet Problem 5 takes the form*

$$\mathbf{a}[u, v] = \sum_{\substack{\alpha \in \mathbb{N}_0^d \\ |\alpha|=1}} \sum_{\substack{\beta \in \mathbb{N}_0^d \\ |\beta|=1}} \int_\Omega a_{\alpha,\beta} \partial^\alpha u \partial^\beta v \, dx, \quad \forall u, v \in \mathrm{H}_0^1(\Omega) \tag{2.89}$$

with coefficients $\alpha^{\alpha,\beta} \in \mathcal{X}(\Omega) = \mathrm{L}^\infty(\Omega)$ for all $\alpha, \beta \in \mathbb{N}_0^d$ with $|\alpha| = |\beta| = 1$. Then, $\mathbf{a}$ is elliptic, i.e.

$$\mathbf{a}[u, u] \geqslant \mathrm{C}_e \, \|u\|^2_{\mathrm{H}^1(\Omega)}, \quad \forall u \in \mathrm{H}_0^1(\Omega), \tag{2.90}$$

with the constant $\mathrm{C}_e = \frac{\mathrm{C}_e^{\mathcal{L}}}{2} \min\{1, \mathrm{C}_{PF}^{-2} \quad\} \in \mathbb{R}_+$, where C_{PF} is the Poincaré-Friedrichs constant.

◇

Proof. Since $\mathcal{L}$ contains only second-order derivatives, all terms $\partial^\alpha u$ and $\partial^\beta v$ appearing in the bilinear form refer to first-order weak derivatives. For fixed $x \in \Omega$, the definition of uniform

ellipticity with $\xi = \nabla u$ yields

$$\begin{aligned} \mathbf{a}[u,u] &= \sum_{\substack{\alpha\in\mathbb{N}_0^d \\ |\alpha|=1}} \sum_{\substack{\beta\in\mathbb{N}_0^d \\ |\beta|=1}} \int_\Omega a_{\alpha,\beta}\partial^\alpha u \partial^\beta u \, dx \\ &= \int_\Omega \sum_{\substack{\alpha\in\mathbb{N}_0^d \\ |\alpha|=1}} \sum_{\substack{\beta\in\mathbb{N}_0^d \\ |\beta|=1}} a_{\alpha,\beta}(\nabla u)^{\alpha+\beta} \, dx \\ &\geqslant \mathrm{C}_{\mathrm{e}}^{\mathcal{L}} \int_\Omega |\nabla u|^2 \, dx \\ &= \mathrm{C}_{\mathrm{e}}^{\mathcal{L}} \|\nabla u\|^2_{\mathrm{L}^2(\Omega)}. \end{aligned} \tag{2.91}$$

An application of the Poincaré-Friedrichs inequality yields the claim,

$$\begin{aligned} \mathbf{a}[u,v] &\geqslant \frac{\mathrm{C}_{\mathrm{e}}^{\mathcal{L}}}{2} \left(\|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^d} + \|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^d} \right) \\ &\geqslant \frac{\mathrm{C}_{\mathrm{e}}^{\mathcal{L}}}{2} \left(\frac{1}{\mathrm{C}_{\mathrm{PF}}^2} \|u\|^2_{\mathrm{L}^2(\Omega)} + \|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^d} \right) \\ &\geqslant \underbrace{\min\left\{ \frac{\mathrm{C}_{\mathrm{e}}^{\mathcal{L}}}{2}, \frac{\mathrm{C}_{\mathrm{e}}^{\mathcal{L}}}{2\,\mathrm{C}_{\mathrm{PF}}^2} \right\}}_{=:\mathrm{C_e}} \|u\|^2_{\mathrm{H}^1(\Omega)}. \end{aligned} \tag{2.92}$$

□

Theorem 2.9 (Ellipticity of bilinear form II)**.** *Let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ open and bounded with a piecewise smooth boundary. Consider a vector-valued linear differential operator $\mathcal{L}$ of even order 2, that has a diagonal form, i.e.*

$$\mathcal{L} = \begin{bmatrix} \mathcal{L}_{1,1} & 0 & \dots & 0 \\ 0 & \mathcal{L}_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mathcal{L}_{d,d} \end{bmatrix} \tag{2.93}$$

and suppose further that all second-order linear and scalar differential operators $\mathcal{L}_{i,i}$ appearing in $\mathcal{L}$ are uniformly elliptic with constant $\mathrm{C}_e^{\mathcal{L}_{i,i}}$ for all $i = 1, \dots, d$. Additionally, assume that the corresponding bilinear form $\mathbf{a} : [\mathrm{H}_0^1(\Omega)]^d \times [\mathrm{H}_0^1(\Omega)]^d \to \mathbb{R}$ of the pure Dirichlet Problem 5 takes the form

$$\mathbf{a}[u,v] = \sum_{i=1}^d \sum_{\substack{\alpha^{i,i}\in\mathbb{N}_0^d \\ |\alpha^{i,i}|=1}} \sum_{\substack{\beta^{i,i}\in\mathbb{N}_0^d \\ |\beta^{i,i}|=1}} a_{\alpha^{i,i},\beta^{i,i}} \partial^{\alpha^{i,i}} u_i \partial^{\beta^{i,i}} v_i, \quad \forall u, v \in [\mathrm{H}_0^1(\Omega)]^d \tag{2.94}$$

with coefficients $a_{\alpha^{i,i},\beta^{i,i}} \in \mathcal{X}(\Omega) = \mathrm{L}^\infty(\Omega)$ for all $\alpha^{i,i}, \beta^{i,i} \in \mathbb{N}_0^d$ with $|\alpha^{i,i}| = |\beta^{i,i}| = 1$ for all $i = 1, \dots, d$. Then, $\mathbf{a}$ is elliptic.

◇

Proof. The bilinear form reads

$$\mathbf{a}[u,v] = \sum_{i=1}^{d} \mathbf{a}_{i,i}[u_i, v_i] \tag{2.95}$$

with

$$\begin{aligned} \mathbf{a}_{i,i} &: \mathrm{H}_0^1(\Omega) \times \mathrm{H}_0^1(\Omega) \to \mathbb{R}, \\ \mathbf{a}_{i,i}[u_i, v_i] &= \sum_{\substack{\alpha^{i,i} \in \mathbb{N}_0^d \\ |\alpha^{i,i}|=1|}} \sum_{\substack{\beta^{i,i} \in \mathbb{N}_0^d \\ |\beta^{i,i}|=1}} a_{\alpha^{i,i},\beta^{i,i}} \partial^{\alpha^{i,i}} u_i \partial^{\beta^{i,i}} v_i, \end{aligned} \tag{2.96}$$

for all $u_i, v_i \in \mathrm{H}_0^1(\Omega)$ being the bilinear form corresponding to the operator $\mathcal{L}_{i,i}$, for $i = 1, \ldots, d$. Since all $\mathcal{L}_{i,i}$ satisfy the assumptions from Theorem 2.8, the bilinear forms $\mathbf{a}_{i,i}$ are elliptic with constants $\mathrm{C_e}^{\mathbf{a}_{i,i}}$. Hence, for all $u \in [\mathrm{H}_0^1(\Omega)]^d$ it holds that

$$\begin{aligned} \mathbf{a}[u,u] &= \sum_{i=1}^{d} \mathbf{a}_{i,i}[u_i, u_i] \\ &\geqslant \sum_{i=1}^{d} \mathrm{C_e}^{\mathbf{a}_{i,i}} \|u_i\|^2_{\mathrm{H}^1(\Omega)} \\ &\geqslant \underbrace{\min_{i=1}^{d} \mathrm{C_e}^{\mathbf{a}_{i,i}}}_{=:\mathrm{C_e}} \underbrace{\sum_{i=1}^{d} \|u_i\|^2_{\mathrm{H}^1(\Omega)}}_{=\|u\|^2_{[\mathrm{H}^1(\Omega)]^n}} \\ &= \mathrm{C_e} \|u\|^2_{[\mathrm{H}^1(\Omega)]^d}. \end{aligned} \tag{2.97}$$

□

Ellipticity of the bilinear form can also be shown for higher order operators.

Theorem 2.10 (Ellipticity of bilinear form III)**.** *Let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ be open and bounded and let $\mathcal{L}$ be a scalar linear differential operator of even order $2\,\mathrm{k}$, that is uniformly elliptic with constant $\mathrm{C}_e^{\mathcal{L}}$. Furthermore, suppose that the corresponding bilinear form $\mathbf{a} : \mathrm{H}_0^{\mathrm{k}}(\Omega) \times \mathrm{H}_0^{\mathrm{k}}(\Omega)$ appearing in the pure Dirichlet Problem 5 reads*

$$\mathbf{a}[u,v] = \sum_{\substack{\alpha \in \mathbb{N}_0^d \\ |\alpha| \leqslant \mathrm{k}}} \sum_{\substack{\beta \in \mathbb{N}_0^d \\ |\beta| \leqslant \mathrm{k}}} \int_\Omega a_{\alpha,\beta} \partial^\alpha u \partial^\beta v \, dx, \quad \forall u, v \in \mathrm{H}_0^{\mathrm{k}}(\Omega), \tag{2.98}$$

with coefficients of the form

$$\begin{aligned} a_{\alpha,\beta} &= \text{const} && \forall \alpha, \beta \in \mathbb{N}_0^d,\ |\alpha| + |\beta| = 2\,\mathrm{k} \\ a_{\alpha,\beta} &= 0 && \forall \alpha, \beta \in \mathbb{N}_0^d,\ 0 < |\alpha| + |\beta| \leqslant 2\,\mathrm{k} - 1 \\ a_{0,0}(x) &\geqslant 0, && \forall x \in \Omega\,. \end{aligned} \tag{2.99}$$

Then, the bilinear form $\mathbf{a}$ is elliptic.

◇

Proof. The proof is performed similar to the one from Theorem 2.8. Using the uniform ellipticity of $\mathcal{L}$, $\mathbf{a}[u,u,]$ is estimated downwards to the Sobolev seminorm $\|u\|^2_{\mathrm{H}^{2\,\mathrm{k}}(\Omega)}$. Afterwards, parts of this seminorm are estimated using variants of the Poincaré Friedrichs inequality, leading to seminorms of u on all spaces $\mathrm{H}^m(\Omega)$ for $m = 0,\dots,2\,\mathrm{k}-1$. Taking the minimum over all appearing coefficients proves the claim. □

Theorem 2.11 (Ellipticity of bilinear form IV)**.** *Let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ open and bounded. Consider a vector-valued linear differential operator $\mathcal{L}$ of even order $2\,\mathrm{k}$, that has a diagonal form, i.e.*

$$\mathcal{L} = \begin{bmatrix} \mathcal{L}_{1,1} & 0 & \dots & 0 \\ 0 & \mathcal{L}_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mathcal{L}_{d,d} \end{bmatrix}. \tag{2.100}$$

Suppose further, that all linear scalar-valued differential operators $\mathcal{L}_{i,i}$ appearing in $\mathcal{L}$ are uniformly elliptic with constants $\mathrm{C}_e^{\mathcal{L}_{i,i}}$ for all $i = 1,\dots,d$. Additionally, suppose that the corresponding bilinear form $\mathbf{a} : [\mathrm{H}^{\mathrm{k}}_0(\Omega)]^d \times [\mathrm{H}^{\mathrm{k}}_0(\Omega)]^d \to \mathbb{R}$ appearing in the pure Dirichlet Problem 5 reads

$$\mathbf{a}[u,v] = \sum_{i=1}^{d} \sum_{\substack{\alpha^{i,i}\in\mathbb{N}_0^d \\ |\alpha^{i,i}|\leqslant k}} \sum_{\substack{\beta^{i,i}\in\mathbb{N}_0^d \\ |\beta^{i,i}|\leqslant k}} a_{\alpha^{i,i},\beta^{i,i}} \partial^{\alpha^{i,i}} u_i \partial^{\beta^{i,i}} v_i, \quad \forall u,v \in [\mathrm{H}^{\mathrm{k}}_0(\Omega)]^d, \tag{2.101}$$

with coefficients of $\mathcal{L}^{i,i}$ for $i = 1,\dots,d$ satisfying

$$\begin{aligned} a_{\alpha^{i,i},\beta^{i,i}} &= \mathit{const} && \forall \alpha^{i,i},\beta^{i,i} \in \mathbb{N}_0^d,\ |\alpha^{i,i}| + |\beta^{i,i}| = 2\,\mathrm{k} \\ a_{\alpha^{i,i},\beta^{i,i}} &= 0, && \forall \alpha^{i,i},\beta^{i,i} \in \mathbb{N}_0^d, 0 < |\alpha^{i,i}| + |\beta^{i,i}| \leqslant 2\,\mathrm{k}-1 \\ a_{\alpha^{i,i}=0,\beta^{i,i}=0} &\geqslant 0 \end{aligned} \tag{2.102}$$

Then, the bilinear form $\mathbf{a}$ is elliptic.

◇

Proof. The bilinear form $\mathbf{a}$ reads

$$\mathbf{a}[u,v] = \sum_{i=1}^{d} \mathbf{a}_{i,i}[u_i,v_i] \tag{2.103}$$

with

$$\mathbf{a}_{i,i} : \mathrm{H}^{\mathrm{k}}_0(\Omega) \times \mathrm{H}^{\mathrm{k}}_0(\Omega) \to \mathbb{R}, \quad (u_i,v_i) \mapsto \sum_{\substack{\alpha^{i,i}\in\mathbb{N}_0^d \\ |\alpha^{i,i}|\leqslant k}} \sum_{\substack{\beta^{i,i}\in\mathbb{N}_0^d \\ |\beta^{i,i}|\leqslant k}} \int_\Omega a_{\alpha^{i,i},\beta^{i,i}} \partial^{\alpha^{i,i}} u_i \partial^{\beta^{i,i}} v_i \, dx \tag{2.104}$$

being the bilinear form corresponding to the operator $\mathcal{L}_{i,i}$ for $i = 1,\dots,d$. The bilinear forms $\mathbf{a}_{i,i}$ satisfy the assumptions from Theorem 2.10 and hence are elliptic with constants $C_{\mathrm{e}}^{\mathbf{a}_{i,i}} \in \mathbb{R}_+$ for all $i = 1,\dots,d$, i.e.

$$\mathbf{a}_{i,i}[u_i,u_i] \geqslant C_{\mathrm{e}}^{\mathbf{a}_{i,i}} \|u_i\|^2_{\mathrm{H}^{\mathrm{k}}(\Omega)}, \quad \forall u_i \in \mathrm{H}^{\mathrm{k}}(\Omega). \tag{2.105}$$

Hence, also a is elliptic with

$$\begin{aligned} \mathbf{a}[u,u] &= \sum_{i=1}^{d} \mathbf{a}_{i,i}[u_i,u_i] \\ &\geqslant \sum_{i=1}^{d} C_{\mathrm{e},\mathbf{a}_{i,i}} \|u_i\|^2_{H^1(\Omega)} \\ &\geqslant \left(\min_{i=1}^{d} C_{\mathrm{e}}^{\mathbf{a}_{i,i}}\right) \underbrace{\sum_{i=1}^{d} \|u_i\|^2_{\mathrm{H}^{\mathrm{k}}(\Omega)}}_{=\|u\|^2_{[\mathrm{H}^{\mathrm{k}}(\Omega)]^d}} \\ &= \min_{i=1}^{d} C_{\mathrm{e}}^{\mathbf{a}_{i,i}} \|u\|^2_{[\mathrm{H}^{\mathrm{k}}(\Omega)]^d}. \end{aligned} \tag{2.106}$$

□

Remark 2.13. *If Ω is not bounded, the result from Theorems 2.10 and 2.11 still holds if the coefficients of the terms not involving any weak derivatives, $a_{0,0}$ resp. $a_{\alpha^{i,i}=0,\beta^{i,i}=0}$ for $i = 1,\dots,d$ are bounded below by a constant $c > 0$ resp. $c_i > 0$ for $i = 1,\dots,d$. This was proven in [Hac17, Theorem 7.7].*

*

3

PDE with heterogeneous coefficients

The problems of interest in this thesis are described using strongly heterogeneous coefficients, leading to numerical solutions which may vary on a fine scale. Problems of this type are in general hard to handle using standard numerical methods, since they require a very fine discretization of the underlying computational domain. If the discretization is chosen too coarse, the desired fine-scale features of the true solution cannot be resolved and numerical solutions are inaccurate. On the other hand, high levels of detail lead to large numbers of degrees of freedom needed to describe the discrete numerical solutions. The identification of a good trade-off between the accuracy of solutions and the complexity of numerical models cannot be generic and must be based on the properties of the considered partial differential operators, the domain, as well as the data of the problem.

In general, complicating features come in two categories. First, there are geometric features that possibly cause fine-scale behavior of numerical solutions. The well-known Poisson equation on an L-shaped domain shows, that this may even happen for apparently simple physical domains, partial differential operators, loads and boundary conditions. A straight-forward application of standard numerical methods, such as the Finite Element Method or more general the Partition of Unity Method (cf. Chapter 4) using polynomial bases, will lead to poor results if the corresponding regions around the complicating features of the domain are not resolved finely enough, for example using an adaptive refinement strategy. In Figure 3.1 two of these complicating geometric features are sketched. Figure 3.1 (a) shows a domain with a reentrant corner, which possibly causes singularities. In order to improve the accuracy of numerical solutions, adaptive refinement towards the reentrant corner is required for standard numerical methods in such cases. Figure 3.1 (b) shows a domain with a circular hole in the center, which in most applications won't cause singularities but still require heavy (adaptive) spatial refinement for standard elements / patches / volumes / cells having straight edges. Note that a hole in the domain may also be interpreted as a sudden jump in the operators' coefficients to zero. In general, the coefficients have a strong impact on the behavior of numerical solutions, and their explicit meaning depends on the physical problem described by the partial differential operator. They may for example express thermal conductivity in the study of the heat equation, or the elastic modulus of a material when investigating linear elasticity. In the simplest case, the coefficients are constant throughout the whole body under study, but this does not necessarily hold, leading to the second category of complicating features. The body may contain inclusions made of other materials, with very different physical properties than those of

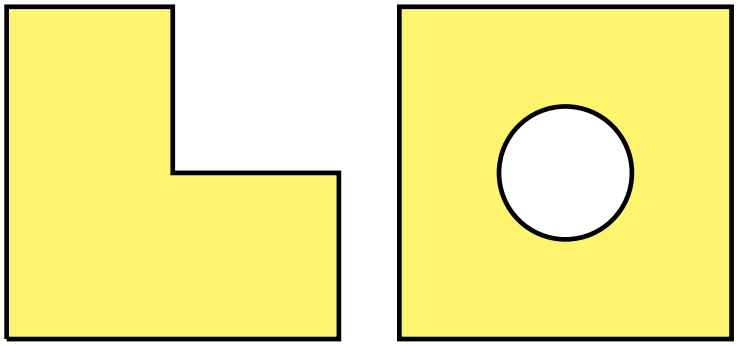


Figure 3.1: Examples of complicating geometric features are reentrant corners (a) and holes (b)

the matrix material, or it may even be made entirely from various material plies. Throughout this thesis, the different subdomains are assumed to be perfectly connected to each other. Figure 3.2 (a) sketches the cross-section of a domain into which a rectangular rod has been inserted. In (b), a domain with multiple inclusions of different sizes and materials is shown. Figure 3.2 (c) shows the cross-section of a three-ply laminate.

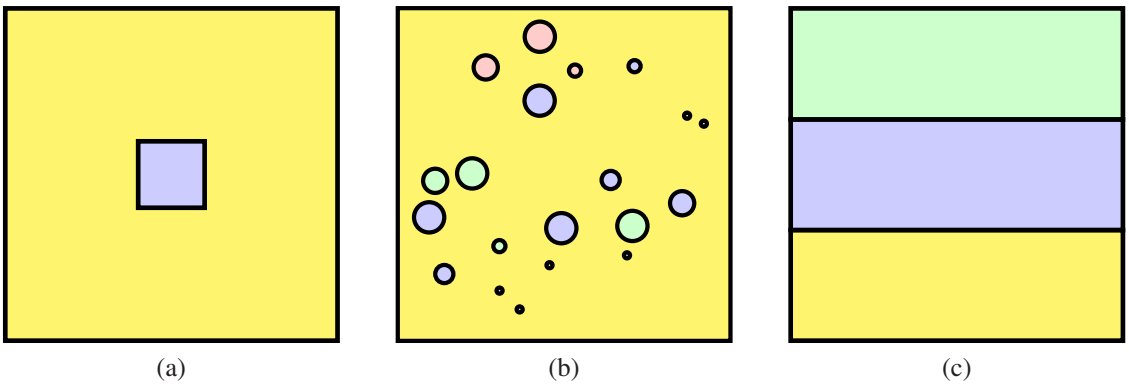


Figure 3.2: (a) Domain with an inclusion, (b) domain with many inclusions of different sizes, (c) 3-ply laminate. The various colors denote different materials.

Many problems of interest in industrial applications combine various of the complicating features presented so far: For a high strength-to-weight ratio, a physical body is made from a laminated material. The geometry of the body may include reentrant corners and each ply is possibly fiber-reinforced, i.e. contains inclusions of a very small size. In order to further reduce weight and save material costs, there may be holes in the body. Finally, bolts and screws, modeled by inclusions throughout the thickness, are inserted in order to connect the body to other structural parts.

Since it is a priori not clear which material to choose for a given purpose, many laminates have to be tested in order to identify good candidates. Furthermore, material coupons in a structural design cycle are exposed to a wide range of conditions and tested for failure. Replacement of physical testing by numerical testing may drastically cut overall costs, since only coupons for the most promising materials will have to be manufactured and tested.

Traditional numerical methods, however, fail to produce accurate results for complex simulations in a feasible amount of time. The Partition of Unity Method, which is introduced in Chapter 4, allows for the use of additional local basis functions, so-called enrichment functions, which may encode a-priori knowledge on the true solutions. For some of the complicating features and differential operators, such enrichment functions can be derived analytically, for example using the stress recovery method ([Mel05]). Unfortunately, analytical enrichments are not available for general operators and domains. Chapter 5 presents a method to compute numerical enrichment functions, which will then be applied to a number of model problems in Chapter 6. The remainder of this chapter introduces these model problems and shows that they are well-posed.

3.1 Second-order scalar problems in divergence form

Second-order elliptic partial differential equations in divergence form were considered before in Chapter 2, and their general form is presented again in Problem 6.

Problem 6: Second-order elliptic PDE in divergence form.

Let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ be open, bounded and sufficiently regular, and let $\mathcal{L} : \mathcal{C}^2(\Omega) \to \mathcal{C}^0(\Omega)$ be a scalar partial differential operator of order $\mathrm{k} = 2$ in the form

$$u \mapsto \mathcal{L}\,u := -\operatorname{div}(A\nabla u) + b \cdot \nabla u + cu, \tag{3.1}$$

for data $A \in [\mathcal{C}^1(\Omega)]^{d\times d}$, $b \in [\mathcal{C}^0(\Omega)]^d$ and $c \in \mathcal{C}^0(\Omega)$. Furthermore, let $\Gamma_D \subset \partial\,\Omega, \Gamma_N \subset \partial\,\Omega \backslash \Gamma_D$ and $f : \Omega \to \mathbb{R}$, $g : \Gamma_D \to \mathbb{R}$ as well as $h : \Gamma_N \to \mathbb{R}$ sufficiently smooth functions. Find a function $u \in \mathcal{C}^2(\Omega)$ satisfying

$$\begin{aligned} \mathcal{L}\,u &= f, && \text{in } \Omega \\ u &= g, && \text{on } \Gamma_D \\ A\nabla u \cdot \vec{n} &= h, && \text{on } \Gamma_N. \end{aligned} \tag{3.2}$$

Remark 3.1. *Problem 6 can also be formulated for vector-valued operators.*

*

Partial differential equations of this form can be used to model a wide variety of stationary phenomena. The theory and conditions for ellipticity of the operator $\mathcal{L}$ have already been investigated in Chapter 2.

Weak formulation

On all parts of the boundary where no boundary condition is imposed in eq. (3.2), a zero Neumann boundary condition is implicitly assumed, i.e. $A\nabla u \cdot \vec{n} = 0$ on $\partial\,\Omega \backslash (\Gamma_D \cup \Gamma_N)$, allowing the solution to behave freely. Hence, multiply the left-hand side of the partial differential equation with a test function v that vanishes on the Dirichlet boundary Γ_D, integrate over Ω and apply

integration by parts to see that

$$\begin{aligned}\int_\Omega \mathcal{L}\,uv\,dx &= \int_\Omega (-\operatorname{div}(A\nabla u) + b\cdot\nabla u + cu)v\,dx \\ &= \int_\Omega -\operatorname{div}(A\nabla u)v + b\cdot\nabla uv + cuv\,dx \\ &= \int_\Omega \nabla u\cdot A\nabla v + b\cdot\nabla uv + cuv\,dx \\ &\quad - \int_{\partial\Omega} v(A\nabla u)\cdot\vec{n}\,ds \\ &= \int_\Omega \nabla u\cdot A\nabla v + b\cdot\nabla uv + cuv\,dx \\ &\quad - \int_{\Gamma_N} hv\,ds,\end{aligned} \tag{3.3}$$

where the last inequality holds due to the boundary conditions. The weak form of Problem 6 can now be formulated for the trial and test spaces

$$\begin{aligned}\mathrm{V}^{\text{trial}}(\Omega) &:= \{u\in \mathrm{H}^1(\Omega) :\ \operatorname{tr} u = g,\ \text{on } \Gamma_D\} \\ \mathrm{V}^{\text{test}}(\Omega) &:= \{v\in \mathrm{H}^1(\Omega) :\ \operatorname{tr} v = 0,\ \text{on } \Gamma_D\}.\end{aligned} \tag{3.4}$$

Problem 7: Weak second-order PDE in divergence form.

Consider a domain of interest, $\Omega\subset\mathbb{R}^d$ for $d\in\mathbb{N}$, which is open, bounded and sufficiently regular. Consider the differential operator $\mathcal{L}$ as in Problem 6, with data $A\in[\mathrm{L}^\infty(\Omega)]^{d\times d}, b\in[\mathrm{L}^\infty(\Omega)]^d, c\in\mathrm{L}^\infty(\Omega)$, For given functions $f\in\mathrm{L}^2(\Omega), g\in\mathrm{L}^2(\Gamma_D), h\in\mathrm{L}^2(\Gamma_N)$ find a function $u\in\mathrm{V}^{\text{trial}}(\Omega)$ satisfying

$$\mathbf{a}[u,v] = \ell(v),\quad \forall v\in\mathrm{V}^{\text{test}}(\Omega), \tag{3.5}$$

with bilinear form $\mathbf{a}:\mathrm{V}^{\text{trial}}(\Omega)\times\mathrm{V}^{\text{test}}(\Omega)\to\mathbb{R}$ and linear functional $\ell:\mathrm{V}^{\text{trial}}(\Omega)\to\mathbb{R}$ defined by

$$\begin{aligned}\mathbf{a}[u,v] &:= \int_\Omega \nabla u\cdot A\nabla v\,dx \\ \ell(v) &:= \int_\Omega fv\,dx + \int_{\Gamma_N} hv\,ds.\end{aligned} \tag{3.6}$$

Ellipticity of the bilinear form $\mathbf{a}[\cdot,\cdot]$ has already been discussed in Chapter 2. The proof of ellipticity requires an application of a Poincaré Friedrichs inequality, that shows the equivalence of the Sobolev seminorm (resp. the energy norm) and the full Sobolev norm on H^1. For the pure Dirichlet problem, i.e. for $\Gamma_N=\varnothing$ and $\Gamma_D=\partial\Omega$, the corresponding inequality is given in Theorem 2.6. This in turn allowed to formulate easy-to-check conditions for ellipticity of $\mathbf{a}$, see

Theorems 2.8 and 2.10. For other types of boundary conditions, other variants of the Poincaré Friedrichs inequality have to be employed.

3.2 Linear Elasticity

An application of forces (loads) to a physical body results in movement of the materials particles, and solid materials can only resist so much tension before breaking. Linear elasticity is a simplification of small (in magnitude) responses of the material to the applied load. There is a large industrial interest in (efficient) numerical methods to solve such problems, as they arise in nearly every sector of manufacturing.
Section 3.2.1 introduces the concept of the stiffness tensor of material properties to describe the constitutive equations of a general orthotropic (anisotropic) material. In general, orthotropic materials may respond differently to outer forces depending on the direction of these forces. They are a generalization of isotropic materials, which share the same material properties in all coordinate directions. The description of a general orthotropic material is hence based on the choice of a coordinate system, the so-called principal axes of the material. This section, which is based on [BC09, GF18, Gur73], also develops the strong formulation of the partial differential equations describing linear elasticity and shows that it is elliptic. In Section 3.2.2, the corresponding weak formulation is derived and it is shown that this is an elliptic variational problem according to the definitions presented in Chapter 2, implying existence and uniqueness of solutions.

3.2.1 Constitutive equations

Outer forces (loads) that are applied to a material lead to movements of its particles. The relation between the applied force and the resulting deformation of the material points is described in terms of the materials' stress tensor σ and strain tensor ε, which are both symmetric by definition. Materials are called emphelastic, if the deformation is reversed whenever the load is removed. Elastic materials are able to deform without suffering damages to their internal structure. If the relation between an elastic materials' stress and strain is linear, the material is referred to as a linear elastic material. In this case Hooke's law holds, stating that the entries of the stress tensor σ in spatial dimension $d = 2$ and $d = 3$ can be written as

$$\sigma_{ij} = \sum_{k,l=1}^{d} \underline{\underline{\mathbf{C}}}_{ijkl} \varepsilon_{kl}. \tag{3.7}$$

This is the tensor contraction $\sigma = \underline{\underline{\mathbf{C}}} : \varepsilon$ of the fourth-order tensor $\underline{\underline{\mathbf{C}}} \in \mathbb{R}^{d\times d\times d\times d}$, called the *Hooke's tensor* or the *stiffness tensor of material properties*, and the infinitesimal second-order strain tensor $\varepsilon \in \mathbb{R}^{d\times d}$. For a vector-valued displacement function $u \in [\mathcal{C}^1(\Omega)]^d$, the latter is defined as

$$\varepsilon(u) := \frac{1}{2}\left(\nabla u + (\nabla u)^T\right), \tag{3.8}$$

with $\nabla u \in [\mathcal{C}^0(\Omega)]^{d\times d}$ being the Jacobian of u. The material tensor has $2^4 = 16$ components in 2d and $3^4 = 81$ components in 3d, and various simplifications are derived in the following. Since the stress is symmetric, i.e. $\sigma_{ij} = \sigma_{ji}$ and ε is arbitrary, it holds that

$$\sigma_{ij} = \sum_{k,l=1}^{d} \underline{\underline{\mathbf{C}}}_{ijkl} \varepsilon_{kl} = \sum_{k,l=1}^{d} \underline{\underline{\mathbf{C}}}_{jikl} \varepsilon_{kl} = \sigma_{ji}, \tag{3.9}$$

so $\underline{\underline{\mathbf{C}}}_{ijkl} = \underline{\underline{\mathbf{C}}}_{jikl}$ for all i, j. This reduces the number of components from $2^4 = 16$ to $2^2 \cdot 3 = 12$ in 2d and from $3^4 = 81$ to $3^2 \cdot 6 = 54$ in 3d. Similarly, the symmetry of the strain tensor gives

$$\sigma_{ij} = \sum_{k,l=1}^{d} \underline{\underline{\mathbf{C}}}_{ijkl}\varepsilon_{kl} = \sum_{k,l=1}^{d} \underline{\underline{\mathbf{C}}}_{ijkl}\varepsilon_{lk} \overset{k\leftrightarrow l}{=} \sum_{k,l=1}^{d} \underline{\underline{\mathbf{C}}}_{ijlk}\varepsilon_{kl}, \tag{3.10}$$

that is

$$\sum_{k,l=1}^{d} (\underline{\underline{\mathbf{C}}}_{ijkl} - \underline{\underline{\mathbf{C}}}_{ijlk})\varepsilon_{kl} = 0. \tag{3.11}$$

Since this holds for all ε, the stiffness tensor satisfies $\underline{\underline{\mathbf{C}}}_{ijkl} = \underline{\underline{\mathbf{C}}}_{ijlk}$, which reduces the number of coefficients to $3 \cdot 3 = 9$ in 2d and to $6 \cdot 6 = 36$ in 3d.

Hooke's tensor is derived from the quadratic strain energy density ψ (for details see [BC09]),

$$\underline{\underline{\mathbf{C}}}_{ijkl} := \frac{\partial^2 \psi}{\partial \varepsilon_{kl} \partial \varepsilon_{ij}}. \tag{3.12}$$

If ψ is a $\mathcal{C}^2$ function, the order of derivation can be changed and hence

$$\underline{\underline{\mathbf{C}}}_{ijkl} \overset{\text{Def.}}{=} \frac{\partial^2 \psi}{\partial \varepsilon_{kl} \partial \varepsilon_{ij}} = \frac{\partial^2 \psi}{\partial \varepsilon_{ij} \varepsilon_{kl}} \overset{\text{Def.}}{=} \underline{\underline{\mathbf{C}}}_{klij}, \tag{3.13}$$

further reducing the number of coefficients to 6 in 2d and 21 in 3d. In the following, the three-dimensional case is considered, but for $d = 2$ the corresponding relations are derived analogously. Let $\{e_1, e_2, e_3\}$ be an orthonormal basis of $\mathbb{R}^3$. Representing the tensor in this basis reads

$$\mathbf{C} = \underline{\underline{\mathbf{C}}}_{ijkl} e_i \otimes e_j \otimes e_k \otimes e_l. \tag{3.14}$$

Using the above-described symmetries, the stress-strain relation from Hooke's law can also be written in Voigt notation as a relation of a matrix (second-order tensor) and stress and strain vectors,

$$\sigma = \mathbf{C}\,\varepsilon, \tag{3.15}$$

with

$$\sigma := \begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{13} \\ \sigma_{12} \end{bmatrix}, \quad \varepsilon := \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ \varepsilon_{23} \\ \varepsilon_{13} \\ \varepsilon_{12} \end{bmatrix}, \tag{3.16}$$

as well as

$$\mathbf{C} := \begin{bmatrix} \underline{\underline{\mathbf{C}}}_{1111} & \underline{\underline{\mathbf{C}}}_{1122} & \underline{\underline{\mathbf{C}}}_{1133} & \underline{\underline{\mathbf{C}}}_{1123} & \underline{\underline{\mathbf{C}}}_{1113} & \underline{\underline{\mathbf{C}}}_{1112} \\ \underline{\underline{\mathbf{C}}}_{1122} & \underline{\underline{\mathbf{C}}}_{2222} & \underline{\underline{\mathbf{C}}}_{2233} & \underline{\underline{\mathbf{C}}}_{2223} & \underline{\underline{\mathbf{C}}}_{2213} & \underline{\underline{\mathbf{C}}}_{2212} \\ \underline{\underline{\mathbf{C}}}_{1133} & \underline{\underline{\mathbf{C}}}_{2233} & \underline{\underline{\mathbf{C}}}_{3333} & \underline{\underline{\mathbf{C}}}_{3323} & \underline{\underline{\mathbf{C}}}_{3313} & \underline{\underline{\mathbf{C}}}_{3312} \\ \underline{\underline{\mathbf{C}}}_{1123} & \underline{\underline{\mathbf{C}}}_{2223} & \underline{\underline{\mathbf{C}}}_{3323} & \underline{\underline{\mathbf{C}}}_{2323} & \underline{\underline{\mathbf{C}}}_{2313} & \underline{\underline{\mathbf{C}}}_{2312} \\ \underline{\underline{\mathbf{C}}}_{1113} & \underline{\underline{\mathbf{C}}}_{2213} & \underline{\underline{\mathbf{C}}}_{3313} & \underline{\underline{\mathbf{C}}}_{2313} & \underline{\underline{\mathbf{C}}}_{1313} & \underline{\underline{\mathbf{C}}}_{1312} \\ \underline{\underline{\mathbf{C}}}_{1112} & \underline{\underline{\mathbf{C}}}_{2212} & \underline{\underline{\mathbf{C}}}_{3312} & \underline{\underline{\mathbf{C}}}_{2312} & \underline{\underline{\mathbf{C}}}_{1312} & \underline{\underline{\mathbf{C}}}_{1212} \end{bmatrix}. \tag{3.17}$$

It can be shown that the matrix form $\mathbf{C}$ of the elasticity tensor $\underline{\underline{\mathbf{C}}}$ is invertible, so

$$\begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ \varepsilon_{23} \\ \varepsilon_{13} \\ \varepsilon_{12} \end{bmatrix} = \mathbf{C}^{-1} \begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{13} \\ \sigma_{12} \end{bmatrix}. \tag{3.18}$$

Using another orthonormal basis $\{\tilde{e}_1, \tilde{e}_2, \tilde{e}_3\}$, and representing the tensor also in this basis, it holds that

$$\mathbf{C} = \tilde{\mathbf{C}} \quad \Longleftrightarrow \quad \underline{\underline{\mathbf{C}}}_{ijkl} e_i \otimes e_j \otimes e_k \otimes e_l = \underline{\underline{\tilde{\mathbf{C}}}}_{pqrs} \tilde{e}_p \otimes \tilde{e}_q \otimes \tilde{e}_r \otimes \tilde{e}_s. \tag{3.19}$$

Using the relation

$$(e_m \otimes e_n) \cdot \tilde{e}_k = (e_n \cdot \tilde{e}_k) e_m \tag{3.20}$$

and associativity of the Kronecker product, it holds that

$$\begin{aligned} \underline{\underline{\mathbf{C}}}_{ijkl}(e_i \otimes e_j) \otimes (e_k \otimes e_l) \cdot e_l &= \underline{\underline{\mathbf{C}}}_{ijkl}(e_i \otimes e_j) \otimes (e_l \cdot e_l) e_k \\ &= \underline{\underline{\mathbf{C}}}(e_l \cdot e_l) e_i \otimes e_j \otimes e_k \\ &= \underline{\underline{\mathbf{C}}} e_i \otimes e_j \otimes e_k. \end{aligned} \tag{3.21}$$

Hence, from (3.19) one obtains

$$\underline{\underline{\mathbf{C}}}_{ijkl} e_i \otimes e_j \otimes e_k = \underline{\underline{\tilde{\mathbf{C}}}}_{pqrs} (\tilde{e}_s \cdot e_l) \tilde{e}_p \otimes \tilde{e}_q \otimes \tilde{e}_r \tag{3.22}$$

A similar step is performed with e_k, e_j and e_i yielding

$$\underline{\underline{\mathbf{C}}}_{ijkl} = \underline{\underline{\tilde{\mathbf{C}}}}_{pqrs} (\tilde{e}_s \cdot e_l)(\tilde{e}_r \cdot e_k)(\tilde{e}_q \cdot e_j)(\tilde{e}_p \cdot e_i), \tag{3.23}$$

which provides a way to compute changes of bases for the fourth-order material stiffness tensor, possibly leading to further simplifications for materials with special properties. Note that any symmetry transformation can be described by an orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{3\times 3}$, with

$$\mathbf{Q} = Q_{ij} e_i \otimes e_j. \tag{3.24}$$

From the transformation formula (3.23) it is seen, that a material symmetry with respect to $\mathbf{Q}$ gives rise to the conditions

$$\underline{\underline{\mathbf{C}}}_{ijkl} = Q_{ip} Q_{jq} Q_{kr} Q_{ls} \underline{\underline{\mathbf{C}}}_{pqrs}. \tag{3.25}$$

A material that has three orthogonal planes of reflection symmetry is called orthotropic. These materials will be considered in the remainder of this work intensively. The orthogonal matrices related to the reflection symmetries are

$$\mathbf{Q}^x = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{Q}^y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{Q}^z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \tag{3.26}$$

Considering $\mathbf{Q}^x$, the material invariance (3.25) gives rise to the conditions

$$
\begin{aligned}
\underline{\underline{\mathbf{C}}}_{1111} &= Q^x_{1p}Q^x_{1q}Q^x_{1r}Q^x_{1s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^4\underline{\underline{\mathbf{C}}}_{1111} = \underline{\underline{\mathbf{C}}}_{1111}\\
\underline{\underline{\mathbf{C}}}_{1122} &= Q^x_{1p}Q^x_{1q}Q^x_{2r}Q^x_{2s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^2 1^2\underline{\underline{\mathbf{C}}}_{1122} = \underline{\underline{\mathbf{C}}}_{1122}\\
\underline{\underline{\mathbf{C}}}_{1133} &= Q^x_{1p}Q^x_{1q}Q^x_{3r}Q^x_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^2 1^2\underline{\underline{\mathbf{C}}}_{1133} = \underline{\underline{\mathbf{C}}}_{1133}\\
\underline{\underline{\mathbf{C}}}_{1123} &= Q^x_{1p}Q^x_{1q}Q^x_{2r}Q^x_{3s}\underline{\underline{\mathbf{C}}}_{prqs} = (-1)^2 1^2\underline{\underline{\mathbf{C}}}_{1123} = \underline{\underline{\mathbf{C}}}_{1123}\\
\underline{\underline{\mathbf{C}}}_{1113} &= Q^x_{1p}Q^x_{1q}Q^x_{1r}Q^x_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^3 1\underline{\underline{\mathbf{C}}}_{1113} = -\underline{\underline{\mathbf{C}}}_{1113}\\
\underline{\underline{\mathbf{C}}}_{1112} &= Q^x_{1p}Q^x_{1q}Q^x_{1r}Q^x_{2s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^3 1\underline{\underline{\mathbf{C}}}_{1112} = -\underline{\underline{\mathbf{C}}}_{1112}\\
\underline{\underline{\mathbf{C}}}_{2222} &= Q^x_{2p}Q^x_{2q}Q^x_{2r}Q^x_{2s}\underline{\underline{\mathbf{C}}}_{pqrs} = 1^4\underline{\underline{\mathbf{C}}}_{2222} = \underline{\underline{\mathbf{C}}}_{2222}\\
\underline{\underline{\mathbf{C}}}_{2233} &= Q^x_{2p}Q^x_{2q}Q^x_{3r}Q^x_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = 1^4\underline{\underline{\mathbf{C}}}_{2233} = \underline{\underline{\mathbf{C}}}_{2233}\\
\underline{\underline{\mathbf{C}}}_{2223} &= Q^x_{2p}Q^x_{2q}Q^x_{2r}Q^x_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = 1^4\underline{\underline{\mathbf{C}}}_{2223} = \underline{\underline{\mathbf{C}}}_{2223}\\
\underline{\underline{\mathbf{C}}}_{2213} &= Q^x_{2p}Q^x_{2q}Q^x_{1r}Q^x_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)1^3\underline{\underline{\mathbf{C}}}_{2213} = -\underline{\underline{\mathbf{C}}}_{2213}\\
\underline{\underline{\mathbf{C}}}_{2212} &= Q^x_{2p}Q^x_{2q}Q^x_{1r}Q^x_{2s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)1^3\underline{\underline{\mathbf{C}}}_{2212} = -\underline{\underline{\mathbf{C}}}_{2212}\\
\underline{\underline{\mathbf{C}}}_{3333} &= Q^x_{3p}Q^x_{3q}Q^x_{3r}Q^x_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = 1^4\underline{\underline{\mathbf{C}}}_{3333} = \underline{\underline{\mathbf{C}}}_{3333}\\
\underline{\underline{\mathbf{C}}}_{3323} &= Q^x_{3p}Q^x_{3q}Q^x_{2r}Q^x_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = 1^4\underline{\underline{\mathbf{C}}}_{3323} = \underline{\underline{\mathbf{C}}}_{3323}\\
\underline{\underline{\mathbf{C}}}_{3313} &= Q^x_{3p}Q^x_{3q}Q^x_{1r}Q^x_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)1^3\underline{\underline{\mathbf{C}}}_{3313} = -\underline{\underline{\mathbf{C}}}_{3313}\\
\underline{\underline{\mathbf{C}}}_{3312} &= Q^x_{3p}Q^x_{3q}Q^x_{1r}Q^x_{2s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)1^3\underline{\underline{\mathbf{C}}}_{3312} = -\underline{\underline{\mathbf{C}}}_{3312}\\
\underline{\underline{\mathbf{C}}}_{2323} &= Q^x_{2p}Q^x_{3q}Q^x_{2r}Q^x_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = 1^4\underline{\underline{\mathbf{C}}}_{2323} = \underline{\underline{\mathbf{C}}}_{2323}\\
\underline{\underline{\mathbf{C}}}_{2313} &= Q^x_{2p}Q^x_{3q}Q^x_{1r}Q^x_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)1^3\underline{\underline{\mathbf{C}}}_{2313} = -\underline{\underline{\mathbf{C}}}_{2313}\\
\underline{\underline{\mathbf{C}}}_{2312} &= Q^x_{2p}Q^x_{3q}Q^x_{1r}Q^x_{2s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)1^3\underline{\underline{\mathbf{C}}}_{2312} = -\underline{\underline{\mathbf{C}}}_{2312}\\
\underline{\underline{\mathbf{C}}}_{1313} &= Q^x_{1p}Q^x_{3q}Q^x_{1r}Q^x_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^2 1^2\underline{\underline{\mathbf{C}}}_{1313} = \underline{\underline{\mathbf{C}}}_{1313}\\
\underline{\underline{\mathbf{C}}}_{1312} &= Q^x_{1p}Q^x_{3q}Q^x_{1r}Q^x_{2s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^2 1^2\underline{\underline{\mathbf{C}}}_{1312} = \underline{\underline{\mathbf{C}}}_{1312}\\
\underline{\underline{\mathbf{C}}}_{1212} &= Q^x_{1p}Q^x_{2q}Q^x_{1r}Q^x_{2s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^2 1^2\underline{\underline{\mathbf{C}}}_{1212} = \underline{\underline{\mathbf{C}}}_{1212}.
\end{aligned}
\tag{3.27}
$$

showing that

$$
\begin{aligned}
0 &= \underline{\underline{\mathbf{C}}}_{1113} = \underline{\underline{\mathbf{C}}}_{1112} = \underline{\underline{\mathbf{C}}}_{2213} = \underline{\underline{\mathbf{C}}}_{2212} = \underline{\underline{\mathbf{C}}}_{3313}\\
&= \underline{\underline{\mathbf{C}}}_{3312} = \underline{\underline{\mathbf{C}}}_{2313} = \underline{\underline{\mathbf{C}}}_{2312}.
\end{aligned}
\tag{3.28}
$$

Considering $\mathbf{Q}^y$ implies that the coefficients which are nonzero yet satisfy

$$\begin{aligned}
\underline{\underline{\mathbf{C}}}_{1111} &= Q^y_{1p}Q^y_{1q}Q^y_{1r}Q^y_{1s}\underline{\underline{\mathbf{C}}}_{pqrs} = 1^4\underline{\underline{\mathbf{C}}}_{1111} = \underline{\underline{\mathbf{C}}}_{1111}\\
\underline{\underline{\mathbf{C}}}_{1122} &= Q^y_{1p}Q^y_{1q}Q^y_{2r}Q^y_{2s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^2 1^2\underline{\underline{\mathbf{C}}}_{1122} = \underline{\underline{\mathbf{C}}}_{1122}\\
\underline{\underline{\mathbf{C}}}_{1133} &= Q^y_{1p}Q^y_{1q}Q^y_{3r}Q^y_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = 1^4\underline{\underline{\mathbf{C}}}_{1133} = \underline{\underline{\mathbf{C}}}_{1133}\\
\underline{\underline{\mathbf{C}}}_{1123} &= Q^y_{1p}Q^y_{1q}Q^y_{2r}Q^y_{3s}\underline{\underline{\mathbf{C}}}_{prqs} = (-1)1^3\underline{\underline{\mathbf{C}}}_{1123} = -\underline{\underline{\mathbf{C}}}_{1123}\\
\underline{\underline{\mathbf{C}}}_{2222} &= Q^y_{2p}Q^y_{2q}Q^y_{2r}Q^y_{2s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^4\underline{\underline{\mathbf{C}}}_{2222} = \underline{\underline{\mathbf{C}}}_{2222}\\
\underline{\underline{\mathbf{C}}}_{2233} &= Q^y_{2p}Q^y_{2q}Q^y_{3r}Q^y_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^2 1^2\underline{\underline{\mathbf{C}}}_{2233} = \underline{\underline{\mathbf{C}}}_{2233}\\
\underline{\underline{\mathbf{C}}}_{2223} &= Q^y_{2p}Q^y_{2q}Q^y_{2r}Q^y_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^3 1\underline{\underline{\mathbf{C}}}_{2223} = -\underline{\underline{\mathbf{C}}}_{2223}\\
\underline{\underline{\mathbf{C}}}_{3333} &= Q^y_{3p}Q^y_{3q}Q^y_{3r}Q^y_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = 1^4\underline{\underline{\mathbf{C}}}_{3333} = \underline{\underline{\mathbf{C}}}_{3333}\\
\underline{\underline{\mathbf{C}}}_{3323} &= Q^y_{3p}Q^y_{3q}Q^y_{2r}Q^y_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)1^3\underline{\underline{\mathbf{C}}}_{3323} = -\underline{\underline{\mathbf{C}}}_{3323}\\
\underline{\underline{\mathbf{C}}}_{2323} &= Q^y_{2p}Q^y_{3q}Q^y_{2r}Q^y_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^2 1^2\underline{\underline{\mathbf{C}}}_{2323} = \underline{\underline{\mathbf{C}}}_{2323}\\
\underline{\underline{\mathbf{C}}}_{1313} &= Q^y_{1p}Q^y_{3q}Q^y_{1r}Q^y_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = 1^4\underline{\underline{\mathbf{C}}}_{1313} = \underline{\underline{\mathbf{C}}}_{1313}\\
\underline{\underline{\mathbf{C}}}_{1312} &= Q^y_{1p}Q^y_{3q}Q^y_{1r}Q^y_{2s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)1^3\underline{\underline{\mathbf{C}}}_{1312} = -\underline{\underline{\mathbf{C}}}_{1312}\\
\underline{\underline{\mathbf{C}}}_{1212} &= Q^y_{1p}Q^y_{2q}Q^y_{1r}Q^y_{2s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^2 1^2\underline{\underline{\mathbf{C}}}_{1212} = \underline{\underline{\mathbf{C}}}_{1212}.
\end{aligned} \tag{3.29}$$

showing that

$$\underline{\underline{\mathbf{C}}}_{1123} = \underline{\underline{\mathbf{C}}}_{2223} = \underline{\underline{\mathbf{C}}}_{3323} = \underline{\underline{\mathbf{C}}}_{1312} = 0. \tag{3.30}$$

Lastly, considering $\mathbf{Q}^z$ yields the following relations for the nonzero coefficients

$$\begin{aligned}
\underline{\underline{\mathbf{C}}}_{1111} &= Q^z_{1p}Q^z_{1q}Q^z_{1r}Q^z_{1s}\underline{\underline{\mathbf{C}}}_{pqrs} = 1^4\underline{\underline{\mathbf{C}}}_{1111} = \underline{\underline{\mathbf{C}}}_{1111}\\
\underline{\underline{\mathbf{C}}}_{1122} &= Q^z_{1p}Q^z_{1q}Q^z_{2r}Q^z_{2s}\underline{\underline{\mathbf{C}}}_{pqrs} = 1^4\underline{\underline{\mathbf{C}}}_{1122} = \underline{\underline{\mathbf{C}}}_{1122}\\
\underline{\underline{\mathbf{C}}}_{1133} &= Q^z_{1p}Q^z_{1q}Q^z_{3r}Q^z_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^2 1^2\underline{\underline{\mathbf{C}}}_{1133} = \underline{\underline{\mathbf{C}}}_{1133}\\
\underline{\underline{\mathbf{C}}}_{2222} &= Q^z_{2p}Q^z_{2q}Q^z_{2r}Q^z_{2s}\underline{\underline{\mathbf{C}}}_{pqrs} = 1^4\underline{\underline{\mathbf{C}}}_{2222} = \underline{\underline{\mathbf{C}}}_{2222}\\
\underline{\underline{\mathbf{C}}}_{2233} &= Q^z_{2p}Q^z_{2q}Q^z_{3r}Q^z_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^2 1^2\underline{\underline{\mathbf{C}}}_{2233} = \underline{\underline{\mathbf{C}}}_{2233}\\
\underline{\underline{\mathbf{C}}}_{3333} &= Q^z_{3p}Q^z_{3q}Q^z_{3r}Q^z_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^4\underline{\underline{\mathbf{C}}}_{3333} = \underline{\underline{\mathbf{C}}}_{3333}\\
\underline{\underline{\mathbf{C}}}_{2323} &= Q^z_{2p}Q^z_{3q}Q^z_{2r}Q^z_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^2 1^2\underline{\underline{\mathbf{C}}}_{2323} = \underline{\underline{\mathbf{C}}}_{2323}\\
\underline{\underline{\mathbf{C}}}_{1313} &= Q^z_{1p}Q^z_{3q}Q^z_{1r}Q^z_{3s}\underline{\underline{\mathbf{C}}}_{pqrs} = (-1)^2 1^2\underline{\underline{\mathbf{C}}}_{1313} = \underline{\underline{\mathbf{C}}}_{1313}\\
\underline{\underline{\mathbf{C}}}_{1212} &= Q^z_{1p}Q^z_{2q}Q^z_{1r}Q^z_{2s}\underline{\underline{\mathbf{C}}}_{pqrs} = 1^4\underline{\underline{\mathbf{C}}}_{1212} = \underline{\underline{\mathbf{C}}}_{1212}.
\end{aligned} \tag{3.31}$$

Hence, two symmetry planes are enough to reduce the matrix form of the stiffness tensor of the material to the final, simplified form

$$\mathbf{C} = \begin{bmatrix}
\underline{\underline{\mathbf{C}}}_{1111} & \underline{\underline{\mathbf{C}}}_{1122} & \underline{\underline{\mathbf{C}}}_{1133} & 0 & 0 & 0\\
\underline{\underline{\mathbf{C}}}_{1122} & \underline{\underline{\mathbf{C}}}_{2222} & \underline{\underline{\mathbf{C}}}_{2233} & 0 & 0 & 0\\
\underline{\underline{\mathbf{C}}}_{1133} & \underline{\underline{\mathbf{C}}}_{2233} & \underline{\underline{\mathbf{C}}}_{3333} & 0 & 0 & 0\\
0 & 0 & 0 & \underline{\underline{\mathbf{C}}}_{2323} & 0 & 0\\
0 & 0 & 0 & 0 & \underline{\underline{\mathbf{C}}}_{1313} & 0\\
0 & 0 & 0 & 0 & 0 & \underline{\underline{\mathbf{C}}}_{1212}
\end{bmatrix} \tag{3.32}$$

An orthotropic material can be fully described by the Young's moduli, the shear moduli and the Poisson ratios. They are defined using the three main axes (the axes of the coordinate system). Young's moduli E_x, E_y, E_z describe the stiffness of the material along the axes. The shear moduli G_{yz}, G_{zx}, G_{xy} describes the ratio of shear stress to shear strain. The first index denotes the direction of stress and strain (along one of the main axes) and the second index denotes the direction of the normal of the plane (also one of the main axes), whose movement is considered. The Poisson ratios $\nu_{yz}, \nu_{zx}, \nu_{xy}$ describe the contraction in direction of the main axes corresponding to the second index when applying a force in direction of the main axes corresponding to the first index.

Using these material constants, the stiffness matrix $\mathbf{C}^{-1}$ in the stress strain relations (3.18) reads

$$\mathbf{C}^{-1} = \begin{bmatrix} \frac{1}{E_x} & -\frac{\nu_{yx}}{E_y} & -\frac{\nu_{yx}}{E_y} & 0 & 0 & 0 \\ -\frac{\nu_{xy}}{E_x} & \frac{1}{E_y} & -\frac{\nu_{yz}}{E_y} & 0 & 0 & 0 \\ -\frac{\nu_{xy}}{E_x} & -\frac{\nu_{yz}}{E_y} & \frac{1}{E_y} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2G_{yz}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2G_{zx}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2G_{xy}} \end{bmatrix} \tag{3.33}$$

and inverting it yields

$$\mathbf{C} = \begin{bmatrix} \frac{1-\nu_{yz}\nu_{zy}}{E_yE_z\Delta} & \frac{\nu_{yx}+\nu_{zx}\nu_{yz}}{E_yE_z\Delta} & \frac{\nu_{zx}+\nu_{yx}\nu_{zy}}{E_yE_z\Delta} & 0 & 0 & 0 \\ \frac{\nu_{xy}+\nu_{xz}\nu_{zy}}{E_zE_x\Delta} & \frac{1-\nu_{zx}\nu_{xz}}{E_zE_x\Delta} & \frac{\nu_{zy}+\nu_{zx}\nu_{xy}}{E_zE_x\Delta} & 0 & 0 & 0 \\ \frac{\nu_{xz}+\nu_{xy}\nu_{yz}}{E_xE_y\Delta} & \frac{\nu_{yz}+\nu_{xz}\nu_{yx}}{E_xE_y\Delta} & \frac{1-\nu_{xy}\nu_{yx}}{E_xE_y\Delta} & 0 & 0 & 0 \\ 0 & 0 & 0 & 2G_{yz} & 0 & 0 \\ 0 & 0 & 0 & 0 & 2G_{zx} & 0 \\ 0 & 0 & 0 & 0 & 0 & 2G_{xy} \end{bmatrix} \tag{3.34}$$

with

$$\Delta = \frac{1 - \nu_{xy}\nu_{yx} - \nu_{yz}\nu_{zy} - \nu_{zx}\nu_{xz} - 2\nu_{xy}\nu_{yz}\nu_{zx}}{E_xE_yE_z}. \tag{3.35}$$

Note that one can compute ν_{zy} from ν_{yz}, ν_{xz} from ν_{zx} and ν_{yx} from ν_{xy} via the relations

$$\begin{aligned} \frac{\nu_{yz}}{E_y} &= \frac{\nu_{zy}}{E_z} \\ \frac{\nu_{zx}}{E_z} &= \frac{\nu_{xz}}{E_x} \\ \frac{\nu_{xy}}{E_x} &= \frac{\nu_{yx}}{E_y}. \end{aligned} \tag{3.36}$$

Remark 3.2. *For isotropic linear elastic materials, it holds that*

$$\begin{aligned} E_x &= E_y = E_z =: E \\ \nu_{yz} &= \nu_{zx} = \nu_{xy} =: \nu \\ G_{yz} &= G_{zx} = G_{xy} =: G. \end{aligned} \tag{3.37}$$

In this case, the shear modulus can be expressed using the Poisson ratio and the modulus of elasticity via

$$G = \frac{E}{2(1+\nu)}. \tag{3.38}$$

The stiffness tensor reads

$$\mathbf{C}^{-1} = \begin{bmatrix} \frac{1}{E} & -\frac{\nu}{E} & -\frac{\nu}{E} & 0 & 0 & 0 \\ -\frac{\nu}{E} & \frac{1}{E} & -\frac{\nu}{E} & 0 & 0 & 0 \\ -\frac{\nu}{E} & -\frac{\nu}{E} & \frac{1}{E} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1+\nu}{E} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1+\nu}{E} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1+\nu}{E} \end{bmatrix} \tag{3.39}$$

and its inverse can be written using the Lamé constants

$$\lambda := \frac{\nu E}{(1-2\nu)(1+\nu)}, \quad \mu := \frac{E}{2(1+\nu)} \tag{3.40}$$

as

$$\mathbf{C}^{-1} = \begin{bmatrix} 2\mu+\lambda & \lambda & \lambda & 0 & 0 & 0 \\ \lambda & 2\mu+\lambda & \lambda & 0 & 0 & 0 \\ \lambda & \lambda & 2\mu+\lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu \end{bmatrix}. \tag{3.41}$$

*

Next, the strong form of the partial differential equations describing linear elasticity is stated. In order to do so, the divergence operator is extended to matrices: The divergence of the matrix $A \in [\mathcal{C}^1(\Omega)]^{d\times d}$ is defined as the vector containing the divergence of the columns of A, i.e.

$$\operatorname{div}(A) := \begin{bmatrix} \sum_{i=1}^{d} \frac{\partial A_{i,1}}{\partial x_i} \\ \vdots \\ \sum_{i=1}^{d} \frac{\partial A_{i,d}}{\partial x_i} \end{bmatrix}. \tag{3.42}$$

The general form of the partial differential equations of linear elasticity, which prescribe the value of the divergence of the strain tensor $\sigma(u)$ subject to boundary conditions, are presented in Problem 8.

Problem 8: Linear elasticity.

Let $\Omega \subset \mathbb{R}^d$ for $d \in \{2, 3\}$ be open and bounded with piecewise smooth boundary and let a stiffness tensor $\underline{\underline{\mathbf{C}}}$ be given. Let

$$\mathcal{L} : [\mathcal{C}^2(\Omega)]^d \to [\mathcal{C}^0(\Omega)]^d, \tag{3.43}$$

with

$$\mathcal{L}\, u(x) := -\operatorname{div} \sigma(u(x)) \tag{3.44}$$

and strain tensor $\sigma(u) = \underline{\underline{\mathbf{C}}} : \varepsilon(u)$. Furthermore, let $\Gamma_D \subset \partial\,\Omega$, $\Gamma_N \subset \partial\,\Omega \backslash \Gamma_D$ and $f : \Omega \to \mathbb{R}^d$, $g : \Gamma_D \to \mathbb{R}^d$ as well as $h : \Gamma_N \to \mathbb{R}^d$ sufficiently smooth functions. Find a function $u \in [\mathcal{C}^2(\Omega)]^d$ satisfying

$$\begin{aligned} \mathcal{L}\, u &= f, \quad \text{in } \Omega \\ u &= g, \quad \text{on } \Gamma_D \\ \sigma(u) \cdot \vec{n} &= h, \quad \text{on } \Gamma_N. \end{aligned} \tag{3.45}$$

Remark 3.3. *The form of the differential operator from Problem 8 and the definition of the matrix divergence from eq.* (3.42) *show, that Problem 8 is a coupled system of* d *second-order partial differential equations in divergence form.*

$*$

The operator $\mathcal{L}$ from Problem 8 is continuous. Whenever Assumption 3.1 holds, it can be shown that $\mathcal{L}$ is elliptic.

Assumption 3.1 (Ellipticity of $\mathcal{L}$)**.** *The material under study is not perfectly incompressible, i.e. at least one of the Young moduli is less than* 0.5*, and the shear moduli are positive.*

$\diamond$

Assumption 3.1 ensures $\Delta > 0$ in (3.35) implying that all nontrivial entries from the stiffness tensor $\underline{\underline{\mathbf{C}}}$ are positive. It can now be shown that $\mathcal{L}$ is an elliptic operator. Expanding the action of the differential operator $\mathcal{L}$ applied to $u \in [\mathcal{C}^2(\Omega)]^d$ reads

$$\begin{aligned} \mathcal{L}\, u = \;& \operatorname{div} \sigma(u) = \operatorname{div}(\underline{\underline{\mathbf{C}}} : \varepsilon(u)) \\ =\; & \operatorname{div} \left[\textstyle\sum_{k,l=1}^{3} \frac{1}{2} \underline{\underline{\mathbf{C}}}_{ijkl} (\partial_{x_k} u_l + \partial_{x_l} u_k) \right]_{j=1}^{3} \\ =\; & \frac{1}{2} \left[\textstyle\sum_{i=1}^{3} \partial_{x_i} \frac{1}{2} \sum_{k,l=1}^{3} \underline{\underline{\mathbf{C}}}_{ijkl} (\partial_{x_k} u_l + \partial_{x_l} u_k) \right]_{j=1}^{3} \\ =\; & \frac{1}{2} \left[\textstyle\sum_{i,k,l=1}^{3} \underline{\underline{\mathbf{C}}}_{ijkl} \partial_{x_i} \partial_{x_k} u_l + \sum_{i,k,l=1}^{3} \underline{\underline{\mathbf{C}}}_{ijkl} \partial_{x_i} \partial_{x_l} u_k \right]_{j=1}^{3}. \end{aligned} \tag{3.46}$$

The main component $\mathcal{L}^2$ of $\mathcal{L}$ is obtained by taking $i = k$ in the first sum, and $i = l$ in the second sum. It reads

$$\mathcal{L}^2 u = \frac{1}{2}\left[\sum_{k,l=1}^{3} \underline{\underline{\mathbf{C}}}_{kjkl}\partial^2_{x_k} u_l + \sum_{k,l=1}^{3} \underline{\underline{\mathbf{C}}}_{ljkj}\partial^2_{x_j} u_k\right]_{j=1}^{3}. \tag{3.47}$$

Note that for any fixed j, only the choice $l = j$ leads to nontrivial entries $\underline{\underline{\mathbf{C}}}_{kjkl}$ of the stiffness tensor in the first sum. Analogously for the second sum, $\underline{\underline{\mathbf{C}}}_{ljkj}$ is nontrivial only for $k = l$. Hence,

$$\mathcal{L}^2 u = \frac{1}{2}\left[\sum_{k=1}^{3} \underline{\underline{\mathbf{C}}}_{kjkj}\partial^2_{x_k} u_j + \sum_{k=1}^{3} \underline{\underline{\mathbf{C}}}_{kjkj}\partial^2_{x_k} u_k\right]_{j=1}^{3}, \tag{3.48}$$

which according to eq. (2.1) can be written in the form

$$\mathcal{L}^2 u = \begin{bmatrix} \mathcal{L}^2_{1,1} & \mathcal{L}^2_{1,2} & \mathcal{L}^2_{1,3} \\ \mathcal{L}^2_{2,1} & \mathcal{L}^2_{2,2} & \mathcal{L}^2_{2,3} \\ \mathcal{L}^2_{3,1} & \mathcal{L}^2_{3,2} & \mathcal{L}^2_{3,3} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}, \tag{3.49}$$

with scalar differential operators

$$\mathcal{L}^2_{i,j} = \frac{1}{2}\begin{cases} 2\underline{\underline{\mathbf{C}}}_{jjjj}\partial^2_{x_j} + \sum_{\substack{p=1 \\ p\neq j}}^{3} \underline{\underline{\mathbf{C}}}_{pjpj}\partial^2_{x_p}, & \text{if } i = j \\ \sum_{\substack{p=1 \\ p\neq j}}^{3} \underline{\underline{\mathbf{C}}}_{jpjp}\partial^2_{x_j} u_p, & \text{else.} \end{cases} \tag{3.50}$$

Note that in the above equation, symmetry of the stiffness coefficients is implicitly used, such that e.g. $\underline{\underline{\mathbf{C}}}_{2121}$ refers to $\underline{\underline{\mathbf{C}}}_{1212}$ etc. The explicit representations of the scalar differential operators appearing in the constitutive equations of linear elasticity can now be used to show uniform ellipticity of the equations. For any $\xi \in \mathbb{R}^3$ and all $j = 1, \ldots, d$ it holds that

$$\left(\underline{\underline{\mathbf{C}}}_{jjjj} + \sum_{\substack{p=1 \\ p\neq j}}^{3} \underline{\underline{\mathbf{C}}}_{pjpj}\right)\xi_j^2 + \sum_{\substack{p=1 \\ p\neq j}}^{3} \underline{\underline{\mathbf{C}}}_{jpjp}\xi_p^2 \geqslant C_{\text{e,j}}|\xi|^2, \tag{3.51}$$

with

$$C_{\text{e,j}} = \min\left\{\underline{\underline{\mathbf{C}}}_{jjjj} + \sum_{\substack{p=1 \\ p\neq j}}^{3} \underline{\underline{\mathbf{C}}}_{jpjp}, \min_{\substack{p=1,2,3 \\ p\neq j}}\left\{\underline{\underline{\mathbf{C}}}_{jpjp}\right\}\right\}, \tag{3.52}$$

which is precisely the definition of uniform ellipticity.

3.2.2 Weak formulation

In this section, the weak form of the PDE of Linear Elasticity for $d \in \{2, 3\}$ is established, and it is shown that the corresponding bilinear form is elliptic. This section is based on the works of J. L. Lions & G. Duvaut [LD72], and of L. Tartar [Tar82]. The proofs of all relevant theoretical results are presented in full length in Chapter A.

The variational formulation is obtained as in the case of any other partial differential equation, by multiplying all components $i = 1, \ldots, d$ of (3.45) by a test function v_i and integrating them over the domain Ω. By denoting the j-th column of $\sigma(u)$ as $\sigma_{\cdot,j}(u)$, this reads

$$\int_\Omega -\operatorname{div}(\sigma_{\cdot,j}(u))v_j\,dx = \int_\Omega f_j v_j\,dx, \quad \forall j = 1, \ldots, d. \tag{3.53}$$

Using Greens identity (integration by parts) individually on the left-hand side of all equations results in

$$\int_\Omega \sigma_{\cdot,j}(u) \cdot \nabla v_j \, dx - \int_{\partial\,\Omega} v_j \sigma_{\cdot,j}(u) \cdot \vec{n} \, ds = \int_\Omega f_j v_j \, dx, \quad j = 1, \ldots, d. \tag{3.54}$$

Note that this system of variational formulations is coupled via the strain tensor $\sigma(u)$, and the sum of the equations from (3.54) is solved for the vector-valued function $v = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}^T$. Formally, the sum of all integrands from the left-hand side volume integrals reads

$$\sum_{j=1}^{d} \sigma_{\cdot,j}(u) \cdot \nabla v_j = \sum_{i,j=1}^{d} \sigma_{i,j}(u) \frac{\partial v_j}{\partial x_i} = \sigma(u) : \nabla v. \tag{3.55}$$

The sum of all integrands from the boundary integrals is

$$\sum_{j=1}^{d} v_j \sigma_{\cdot,j}(u) \cdot \vec{n} = v \cdot \sigma(u) \vec{n} \tag{3.56}$$

and the sum of all integrands from the right-hand sides is

$$\sum_{j=1}^{d} v_j \sigma_{\cdot,j}(u) \cdot \vec{n} = \sum_{i,j=1}^{d} v_j \sigma_{ij} \vec{n}_i = v \cdot \sigma(u) \vec{n}. \tag{3.57}$$

In total, one obtains the equation

$$\int_\Omega \sigma(u) : \nabla v \, dx - \int_{\partial\,\Omega} v \cdot \sigma(u) \vec{n} \, ds = \int_\Omega f \cdot v \, dx. \tag{3.58}$$

Note that the boundary integral does not appear on the left-hand side in practical computations: It moves to the right-hand side of the equation for all parts of the boundary on which a boundary condition is imposed. If no explicit boundary condition is set, u is assumed to satisfy a zero Neumann boundary condition, allowing it to behave freely and leading to the boundary integral being zero on these parts of the boundary. Therefore, inserting the boundary conditions into (3.58) reads

$$\int_\Omega \sigma(u) : \nabla v \, dx = \int_\Omega f \cdot v \, dx + \int_{\Gamma_N} v \cdot h \, ds. \tag{3.59}$$

The symmetry of $\underline{\underline{\mathbf{C}}}$ formally implies that

$$\begin{aligned} \sigma(u) : \nabla v = & \sum_{i,j=1}^{d} \sigma_{ij}(u) \frac{\partial v_i}{\partial x_j} = \sum_{i,j=1}^{d} \sigma_{ji}(u) \frac{\partial v_i}{\partial x_j} \\ \overset{i \leftrightarrow j}{=} & \sum_{i,j=1}^{d} \sigma_{ij}(u) \frac{\partial v_j}{\partial x_i} = \sigma(u) : \nabla^T v, \end{aligned} \tag{3.60}$$

so the left-hand side of (3.59) is symmetric and can be written as

$$\int_\Omega \sigma(u) : \nabla v \, dx = \int_\Omega \sigma(u) : \varepsilon(v) \, dx = \int_\Omega (\underline{\underline{\mathbf{C}}} : \varepsilon(u)) : \varepsilon(u) \, dx. \tag{3.61}$$

Using the trial and test spaces

$$\begin{aligned} \mathrm{V}^{\text{trial}}(\Omega) &:= \{u \in [H^1(\Omega)]^d : \operatorname{tr} u = g \text{ on } \Gamma_D\} \\ \mathrm{V}^{\text{test}}(\Omega) &:= \{v \in [H^1(\Omega)]^d : \operatorname{tr} v = 0 \text{ on } \Gamma_D\}, \end{aligned} \tag{3.62}$$

the weak formulation of Problem 8 can be formulated.

Problem 9: Weak linear elasticity.

Let $\Omega \subset \mathbb{R}^d$ for $d \in \{2, 3\}$ be open, bounded and sufficiently regular. Let $\Gamma_D \subset \partial\Omega$, $\Gamma_N \subset \Omega \backslash \Gamma_D$ and suppose that $f \in [\mathrm{L}^2(\Omega)]^d$, $g \in [\mathrm{L}^2(\Gamma_D)]^d$ as well as $h \in [\mathrm{L}^2(\Gamma_N)]^d$. Find a function $u \in \mathrm{V}^{\text{trial}}(\Omega)$ satisfying

$$\mathbf{a}[u, v] = \ell(v), \quad \forall v \in \mathrm{V}^{\text{test}}(\Omega), \tag{3.63}$$

with bilinear form $\mathbf{a} : \mathrm{V}^{\text{trial}}(\Omega) \times \mathrm{V}^{\text{test}}(\Omega) \to \mathbb{R}$ and linear functional $\ell : \mathrm{V}^{\text{test}}(\Omega) \to \mathbb{R}$ defined by

$$\begin{aligned} \mathbf{a}[u, v] &:= \int_\Omega \sigma(u) : \varepsilon(v)\, dx \\ \ell(v) &= \int_\Omega f \cdot v\, dx + \int_{\Gamma_N} h \cdot v\, ds. \end{aligned} \tag{3.64}$$

In the following, it will be shown that $\mathbf{a}$ is an elliptic bilinear form. In order to do so, the equivalence of the strain energy norm

$$\begin{aligned} \|v\|_{\mathcal{E}(\Omega)} &:= \left[\int_\Omega \varepsilon(v)(x) : \varepsilon(v)(x)\, dx + \int_\Omega v(x) \cdot v(x)\, dx \right]^{\frac{1}{2}} \\ &= \left[\|\varepsilon(v)\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} + \|v\|^2_{[\mathrm{L}^2(\Omega)]^d} \right]^{\frac{1}{2}}, \end{aligned} \tag{3.65}$$

and the standard Sobolev norm on $[\mathrm{H}^1(\Omega)]^d$ is needed. The Korn inequality shows one part of this estimate.

Theorem 3.1 (Korn inequality)**.** *Let $\Omega \subset \mathbb{R}^d$ an open, bounded and regular domain. Then, there exists a constant $C_{\text{korn}}(\Omega) > 0$ such that*

$$\|v\|_{\mathcal{E}(\Omega)} \geqslant C_{\text{korn}}(\Omega) \|v\|_{[\mathrm{H}^1(\Omega)]^d}, \quad \forall v \in [\mathrm{H}^1(\Omega)]^d. \tag{3.66}$$

◇

Proof. The proof can be found in [LD72, Tar82], as well as in Chapter A. □

The other part of the norm equivalence follows by a straight-forward calculation: Let $v \in [\mathrm{H}^1(\Omega)]^d$, then

$$\|v\|^2_{[\mathrm{H}^1(\Omega)]^d} = \|\varepsilon(v)\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} + \|v|^2_{[\mathrm{L}^2(\Omega)]^d}. \tag{3.67}$$

The matrix norm is rewritten as

$$\begin{aligned}
\|\varepsilon(v)\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} &= \int_\Omega \varepsilon(v) : \varepsilon(v)\, dx \\
&= \frac{1}{4}\sum_{i,j=1}^d \int_\Omega \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)^2 dx \\
&= \frac{1}{4}\sum_{i,j=1}^d \int_\Omega \left(\frac{\partial u_i}{\partial x_j}\right)^2 + \left(\frac{\partial u_j}{\partial x_i}\right)^2 + 2\frac{\partial u_i}{\partial x_j}\frac{\partial u_j}{\partial x_i}\, dx \\
&= \frac{1}{4}\|\nabla v\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} + \frac{1}{4}\|\nabla v\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} \\
&\qquad + \frac{1}{2}\sum_{i,j=1}^d \int_\Omega \frac{\partial u_i}{\partial x_j}\frac{\partial u_j}{\partial x_i}\, dx \\
&= \frac{1}{2}\|\nabla v\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} + \frac{1}{2}\sum_{i,j=1}^d \langle \frac{\partial u_i}{\partial x_j}, \frac{\partial u_j}{\partial x_i}\rangle_{L^2(\Omega)}
\end{aligned} \tag{3.68}$$

and the scalar product terms are estimated using the Cauchy Schwarz inequality, followed by an application of the inequality of the arithmetic and geometric mean,

$$\begin{aligned}
\sum_{i,j=1}^d \langle \frac{\partial u_i}{\partial x_j}, \frac{\partial u_j}{\partial x_i}\rangle_{\mathrm{L}^2(\Omega)} &\leqslant \sum_{i,j=1}^d \left\|\frac{\partial u_i}{\partial x_j}\right\|_{\mathrm{L}^2(\Omega)} \left\|\frac{\partial u_j}{\partial x_i}\right\|_{\mathrm{L}^2(\Omega)} \\
&= \sum_{i,j=1}^d \sqrt{\left\|\frac{\partial u_i}{\partial x_j}\right\|^2_{\mathrm{L}^2(\Omega)} \left\|\frac{\partial u_j}{\partial x_i}\right\|^2_{\mathrm{L}^2(\Omega)}} \\
&\leqslant \sum_{i,j=1}^d \frac{1}{2}\left(\left\|\frac{\partial u_i}{\partial x_j}\right\|^2_{\mathrm{L}^2(\Omega)} + \left\|\frac{\partial u_j}{\partial x_i}\right\|^2_{\mathrm{L}^2(\Omega)}\right) \\
&= \frac{1}{2}\|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} + \frac{1}{2}\|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} \\
&= \|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}}.
\end{aligned} \tag{3.69}$$

Using (3.69) in (3.68) yields the other side of the norm equivalence,

$$\|\varepsilon(v)\|_{[\mathrm{L}^2(\Omega)]^d} \leqslant \|\nabla v\|_{[\mathrm{L}^2(\Omega)]^d}. \tag{3.70}$$

Remark 3.4. *Due to the norm equivalence, it holds that*

$$\begin{aligned} C_{\mathrm{korn}}(\Omega)\|v\|_{[H^1(\Omega)]^d} &\leqslant \|v\|_{\mathcal{E}(\Omega)} \\ &= \left(\|\varepsilon(v)\|^2_{[L^2(\Omega)]^{d\times d}} + \|v\|^2_{[L^2(\Omega)]^d}\right)^{\frac{1}{2}} \\ &\leqslant \left(\|\nabla v\|^2_{[L^2(\Omega)]^{d\times d}} + \|v\|^2_{[L^2(\Omega)]^d}\right)^{\frac{1}{2}} \\ &= \|v\|_{[H^1(\Omega)]^d}, \end{aligned} \tag{3.71}$$

that is

$$C_{\mathrm{korn}}(\Omega)\|v\|_{[H^1(\Omega)]^d} \leqslant \|v\|_{[H^1(\Omega)]^d}, \quad \forall v \in [H^1(\Omega)]^d, \tag{3.72}$$

implying $C_{\mathrm{korn}}(\Omega) \leqslant 1$.

*

In the following, the material under study is assumed to be admissible according to definition 11.

Definition 11 (Admissibility of fourth-order tensors)**.** *A tensor* $\underline{\underline{\mathbf{T}}} \in \mathbb{R}^{d\times d\times d\times d}$ *is called admissible for* $\beta \geqslant \gamma > 0$*, if* $\underline{\underline{\mathbf{T}}}$ *satisfies*

$$\begin{aligned} \underline{\underline{\mathbf{T}}}_{ijkl} &= \underline{\underline{\mathbf{T}}}_{jikl} = \underline{\underline{\mathbf{T}}}_{klij} \\ |\underline{\underline{\mathbf{T}}}_{ijkl}| &\leqslant \beta, \quad \forall i,j,k,l = 1,\dots,d \\ (\underline{\underline{\mathbf{T}}} : A(x)) : A(x) &\geqslant \gamma A(x) : A(x), \quad \forall A : \Omega \to \mathbb{R}^{d\times d}_{sym}, \text{ a. e. } x \in \Omega\,. \end{aligned} \tag{3.73}$$

○

For admissible materials, ellipticity of the bilinear form in the pure Dirichlet case now follows from Theorem 3.2

Theorem 3.2 (Ellipticity / boundedness)**.** *Suppose that the assumptions from Problem 9 hold with* $\Gamma_D = \partial\,\Omega$ *and* $\Gamma_N = \varnothing$*. Let* $\underline{\underline{\mathbf{C}}}$ *be an admissible stiffness tensor with admissibility constants* (γ, β)*. Then, the bilinear form* $\mathbf{a}[\cdot,\cdot]$ *appearing in Problem 9 is elliptic and continuous, i.e.*

$$\begin{aligned} \mathbf{a}[u,u] &\geqslant \mathrm{C}_e\,\|u\|^2_{[\mathrm{H}^1(\Omega)]^d}, \\ \mathbf{a}[u,v] &\leqslant \mathrm{C}_{cont}\,\|u\|_{[\mathrm{H}^1(\Omega)]^d}\|v\|_{[\mathrm{H}^1(\Omega)]^d}, \end{aligned} \tag{3.74}$$

for all $u \in \mathrm{V}^{trial}(\Omega)$ *and* $v \in \mathrm{V}^{test}(\Omega)$ *with constants*

$$\begin{aligned} \mathrm{C}_e &= \min\left\{\gamma C^2_{\mathrm{korn}}(\Omega), \frac{\gamma C^2_{\mathrm{korn}}(\Omega)}{c^2_{\mathrm{PF}}(\Omega)}\right\}, \\ \mathrm{C}_{cont} &= \beta \geqslant, \end{aligned} \tag{3.75}$$

satisfying $\mathrm{C}_{cont} \geqslant \mathrm{C}_e > 0$.

◇

Proof. The stiffness tensor $\underline{\underline{\mathbf{C}}}$ is admissible, meaning that

$$(\underline{\underline{\mathbf{C}}} : \varepsilon(u)(x)) : \varepsilon(u)(x) \geqslant \gamma\varepsilon(u)(x) : \varepsilon(u)(x), \tag{3.76}$$

for all $u \in \mathrm{V}^{\text{trial}}$ and almost every $x \in \Omega$. Hence,

$$\begin{aligned} \mathbf{a}[u,u] &= \int_\Omega (\mathbf{C} : \varepsilon(u)) : \varepsilon(u)\, dx \\ &\geqslant \gamma \int_\Omega \varepsilon(u) : \varepsilon(u)\, dx \\ &= \gamma \|\varepsilon(u)\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}}. \end{aligned} \tag{3.77}$$

Inserting the relation

$$\|\varepsilon(u)\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} = \|u\|^2_{\mathcal{E}(\Omega)} - \|u\|^2_{[\mathrm{L}^2(\Omega)]^d} \tag{3.78}$$

into (3.77) yields

$$\begin{aligned} \mathbf{a}[u,v] \geqslant\ & \gamma \|\varepsilon(u)\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} \\ =\ & \gamma \left(\|u\|^2_{\mathcal{E}(\Omega)} - \|u\|^2_{[\mathrm{L}^2(\Omega)]^d} \right) \\ \overset{\text{Korn}}{\geqslant}\ & \gamma \left(C^2_{\text{korn}}(\Omega) \|u\|^2_{[\mathrm{H}^1(\Omega)]^d} - \|u\|^2_{[\mathrm{L}^2(\Omega)]^d} \right) \\ \geqslant\ & \min\{\gamma, \gamma C^2_{\text{korn}}(\Omega)\} \left(\|u\|^2_{[\mathrm{H}^1(\Omega)]^d} - \|u\|^2_{[\mathrm{L}^2(\Omega)]^d} \right) \\ =\ & \min\{\gamma, \gamma C^2_{\text{korn}}(\Omega)\} \|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} \\ =\ & \gamma C^2_{\text{korn}}(\Omega) \|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} \end{aligned} \tag{3.79}$$

where the last inequality holds since $C_{\text{korn}}(\Omega) \leqslant 1$, as stated in Remark 3.4. Since the domain is bounded, a vectir-valued Poincaré Friedrichs inequality holds, stating that there exists a constant $c_{\text{PF}}(\Omega) > 0$ such that

$$\|\nabla u\|_{[\mathrm{L}^2(\Omega)]^{d\times d}} \geqslant \frac{1}{c_{\text{PF}}(\Omega)} \|u\|_{[\mathrm{L}^2(\Omega)]^d}, \quad \forall u \in [\mathrm{H}^1_0(\Omega)]^d. \tag{3.80}$$

Thus, eq. (3.79) yields

$$\begin{aligned} \mathbf{a}[u,u] &\geqslant \gamma C^2_{\text{korn}}(\Omega) \|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} \\ &= \gamma C^2_{\text{korn}}(\Omega) \left(\frac{1}{2} \|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} + \frac{1}{2} \|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} \right) \\ &\geqslant \gamma C^2_{\text{korn}}(\Omega) \left(\|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} + \frac{1}{c_{\text{PF}^2}} \|u\|^2_{[\mathrm{L}^2(\Omega)]^d} \right) \\ &\geqslant \min \left\{ \gamma C^2_{\text{korn}}(\Omega), \frac{\gamma C^2_{\text{korn}}(\Omega)}{c^2_{\text{PF}}(\Omega)} \right\} \|u\|^2_{[\mathrm{H}^1(\Omega)]^d} \end{aligned} \tag{3.81}$$

showing that the bilinear form $\mathbf{a}[\cdot,\cdot]$ is elliptic with ellipticity constant

$$\mathrm{C_e} := \min \left\{ \gamma C^2_{\text{korn}}(\Omega), \frac{\gamma C^2_{\text{korn}}(\Omega)}{c^2_{\text{PF}}(\Omega)} \right\} > 0. \tag{3.82}$$

Boundedness of the bilinear form is a direct consequence of the admissibility of the material. For any $u \in \mathrm{V}^{\text{trial}}(\Omega)$ and $v \in \mathrm{V}^{\text{test}}(\Omega)$ it holds that

$$\begin{aligned} \mathbf{a}[u,v] &= \int_\Omega (\underline{\underline{\mathbf{C}}} : \varepsilon(u)) : \varepsilon(v)\, dx \\ &\leqslant \beta \int_\Omega \varepsilon(u) : \varepsilon(v)\, dx \end{aligned} \tag{3.83}$$

and using the Cauchy Schwarz inequality,

$$\begin{aligned}
\mathbf{a}[u,v] &\leqslant \beta\|\varepsilon(u)\|_{[\mathrm{L}^2(\Omega)]^{d\times d}}\|\varepsilon(v)\|_{[\mathrm{L}^2(\Omega)]^{d\times d}} \\
&\leqslant \beta\left(\|\varepsilon(u)\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} + \|u\|^2_{[\mathrm{L}^2(\Omega)]^d}\right)^{\frac{1}{2}} \\
&\qquad \cdot \left(\|\varepsilon(v)\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} + \|v\|^2_{[\mathrm{L}^2(\Omega)]^d}\right)^{\frac{1}{2}} \\
&= \beta\|u\|_{\mathcal{E}(\Omega)}\|v\|_{\mathcal{E}(\Omega)} \\
&\leqslant \beta\|u\|_{[\mathrm{H}^1(\Omega)]^d}\|v\|_{[\mathrm{H}^1(\Omega)]^d}.
\end{aligned} \tag{3.84}$$

This proves the result. □

Remark 3.5. *Theorem 3.2 proved the ellipticity of the bilinear form in the pure Dirichlet problem. In other well-posed cases, different versions of the Poincaré Friedrichs inequality must be invoked.*

*

4 Partition of Unity Method

In this chapter, a brief overview of the Partition of Unity Method (PUM) is given, as introduced in [Sch03] as a further abstraction to the Generalized Finite Element Method (GFEM). The GFEM, which was presented in [MB96, MB97], is itself a broad generalization of the standard Finite Element Method (FEM). In the following, the general process of constructing a PUM for the solution of a partial differential equation is presented, starting with a description of the domain discretization in Section 4.1. In Section 4.2, the construction of local and global approximation spaces is described. Since the partial differential operators considered in this thesis are of even order $2\,\mathrm{k}$, not necessarily for $\mathrm{k} = 1$, the error estimate of PUM-approximations is formulated accordingly in Section 4.3. Lastly, Section 4.4 shortly presents practical details of the cover construction.

4.1 Spatial discretization

In the following, let $\Omega \subset \mathbb{R}^d$ be a $d \in \mathbb{N}$ dimensional open and bounded with a sufficiently regular boundary. Let $\{\omega_i\}_{i=1}^{\mathrm{m}}$ be an open cover of Ω, that is a collection of open sets, the so-called *patches*, with

$$\omega_i \subset \mathbb{R}^d, \quad \bigcup_{i=1}^{\mathrm{m}} \omega_i \supset \overline{\Omega}. \tag{4.1}$$

Note that there are no specifications on the form or the alignment of these patches. However, there are some conditions that can be imposed on the patches to make computations easier. In practical applications, numerical integration needs to be performed on the patches. Hence, the patches are usually chosen to be rectangles or circles in $d = 2$, resp. cubes and spheres in $d = 3$, since efficient quadrature rules are available for these types of geometry. In Assumptions 4.1 and 4.2, conditions for efficiency and stability of the global method are presented. These conditions assure that the patches from the cover are chosen in a sensible way, for example preventing redundant patches. The pointwise overlap condition from Assumption 4.1 ensures, that at most a certain number of patches overlap at all points of the domain.

Assumption 4.1 (Pointwise overlap condition)**.** *There is a generic bound on the number of patches any point $x \in \Omega$ belongs to, i.e. there exists a number $M \in \mathbb{N}$ such that for all points $x \in \Omega$ it holds that*

$$\mathrm{card}\{i |\, x \in \omega_i\} \leqslant M. \tag{4.2}$$

◇

Assumption 4.2 guarantees, that each patch exclusively covers a part of the domain. This property will be essential to prevent linear dependencies in the global approximation space, as will be seen in Section 4.2.

Assumption 4.2 (Flat top property)**.** *Assume that there exists a constant* $\mathrm{C}_{FT} > 0$ *such that for all patches* ω_i*, the* flat top region

$$\omega_i^{\mathrm{FT}} := \{x \in \omega_i \mid \varphi_i(x) = 1\} \tag{4.3}$$

satisfies the condition

$$\mathrm{meas}(\omega_i) \leqslant \mathrm{C}_{FT}\,\mathrm{meas}(\omega_i^{\mathrm{FT}}), \quad \forall i = 1, \ldots, \mathrm{m}\,. \tag{4.4}$$

◇

Figure 4.1 shows an exemplary discretization of a bounded domain using rectangular patches. The sketched discretization satisfies the pointwise overlap condition, as well as the flat top property.

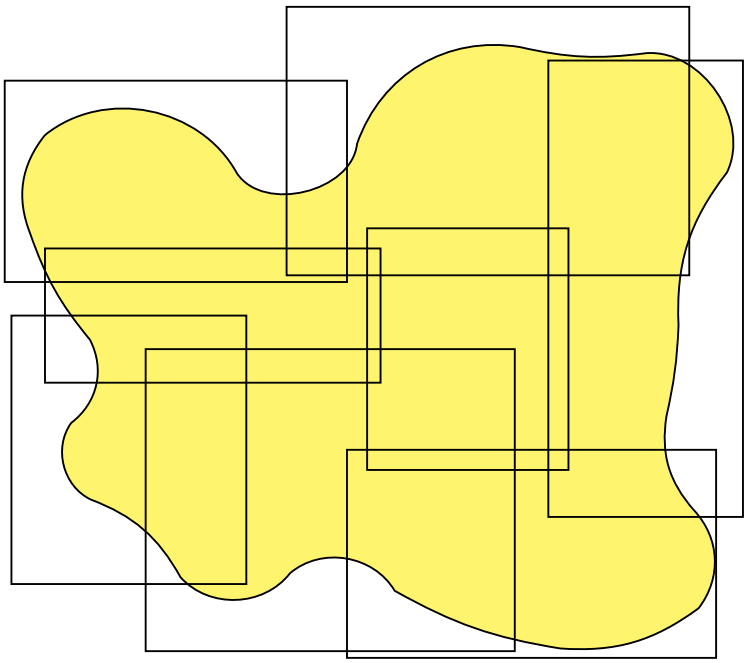


Figure 4.1: Domain and discretization using rectangular patches. The patches fully cover the closure of the domain and satisfy Assumptions 4.1 and 4.2.

4.2 Construction of the approximation space

In the following, the notion of a Partition of Unity (PU) is introduced. A PU is a set of functions with certain properties, which prove useful when constructing a global approximation space from local approximation spaces.

Definition 12 (Partition of Unity)**.** *Let* $\{\varphi_i\}_{i=1}^{\mathrm{m}}$ *be a set of Lipschitz functions, such that for all* $i = 1, \ldots, \mathrm{m}$ *and all* $x \in \Omega$ *it holds that*

$$\operatorname{supp} \varphi_i = \overline{\omega_i}, \tag{4.5}$$

$$\varphi_i(x) \in [0, 1], \tag{4.6}$$

$$\sum_{i=1}^{\mathrm{m}} \varphi_i(x) = 1. \tag{4.7}$$

In this case, $\{\varphi_i\}_{i=1}^{\mathrm{m}}$ *is called a Partition of Unity (PU) subordinate to the cover* $\{\omega_i\}_{i=1}^{\mathrm{m}}$*. If* $\{\varphi_i\}_{i=1}^{\mathrm{m}} \subset \mathcal{C}^r(\Omega)$*, the Partition of Unity is said to be of order* r*.* ○

In order to introduce the Partition of Unity Method and to simplify the notation, multi-indices are used. For $\alpha \in \mathbb{N}_0^d$ and any sufficiently smooth function $u : \Omega \to \mathbb{R}$,

$$\partial^\alpha u \coloneqq \partial_{x_1}^{\alpha_1} \partial_{x_2}^{\alpha_2} \ldots \partial_{x_d}^{\alpha_d} u. \tag{4.8}$$

Furthermore, multi-indices satisfy

$$|\alpha| \coloneqq \sum_{i=1}^{d} \alpha_i, \quad \text{and} \quad \alpha \pm \beta = (\alpha_1 \pm \beta_1, \ldots, \alpha_d \pm \beta_d), \quad \forall \alpha, \beta \in \mathbb{N}_0^d. \tag{4.9}$$

Additionally, the notation $\beta \leqslant \alpha$ is used whenever $\beta_i \leqslant \alpha_i$ for all $i = 1, \ldots, d$. In the following, it is assumed that the Partition of Unity satisfies Assumption 4.3.

Assumption 4.3. *Let* $\{\varphi_i\}_{i=1}^{\mathrm{m}}$ *be a Partition of Unity subordinate to the cover* $\{\omega_i\}_{i=1}^{\mathrm{m}}$ *and suppose that the constructed PUM should be used to solve PDE of even order* $2\,\mathrm{k}$*. Furthermore, suppose that for all* $\beta \in \mathbb{N}_0^d$ *with* $|\beta| \leqslant \mathrm{k}$ *there exists a constant* $C_\beta > 0$ *such that*

$$\| \partial^\beta \varphi_i \|_{L^\infty(\mathbb{R}^d)} \leqslant \frac{C_\beta}{\operatorname{diam}(\omega_i)^{|\beta|}}, \quad \forall i = 1, \ldots, \mathrm{m}\,. \tag{4.10}$$

◇

Next, the construction of a global approximation space is presented, and its approximation quality investigated. To this end, a local approximation space $\mathcal{V}(\omega_i)$ is introduced on each patch ω_i, $i = 1, \ldots, \mathrm{m}$. These local spaces are of the form

$$\mathcal{V}(\omega_i) \coloneqq \mathcal{P}^{p_i}(\omega_i) + \mathcal{E}(\omega_i), \quad \forall i = 1, \ldots, \mathrm{m}, \tag{4.11}$$

where $\mathcal{P}^{p_i}(\omega_i)$ is the space of polynomials of degree $p_i \in \mathbb{N}$ on ω_i. The space $\mathcal{E}(\omega_i)$ is called *space of enrichments*, and it is of the form

$$\mathcal{E}(\omega_i) = \operatorname{span}\{\psi_{i,1}, \ldots, \psi_{i,\mathrm{n}}\}. \tag{4.12}$$

The *enrichment functions* $\psi_{i,j}$ encode a priori knowledge on the behavior of the overall solution, such as fine-scale oscillations due to heterogeneous coefficients, or singularities due to reentrant corners. In Chapter 5, a numerical scheme to pre-compute local functions encoding this behavior of the solution will be presented, which can then be used as enrichment functions on the corresponding patch(es). The set of functions $\{\eta_{i,j}\}_{j=1}^{d_i}$ denotes a basis of $\mathcal{V}(\omega_i)$ for all $i = 1, \ldots, \mathrm{m}$,

and d_i is the corresponding dimension of the local space. The global approximation space is the result of a combination of the Partition of Unity with the local spaces,

$$\mathcal{V}^{\mathrm{PU}}(\Omega) := \sum_{i=1}^{\mathrm{m}} \varphi_i \, \mathcal{V}(\omega_i) = \{\varphi_i \, v_i | \, v_i \in \mathcal{V}(\omega_i)\}. \tag{4.13}$$

The global space is spanned by the local bases in the sense that

$$\mathcal{V}^{\mathrm{PU}}(\Omega) = \operatorname{span}\{\varphi_i \, \eta_{i,j} \, | \, i = 1, \dots, \mathrm{m}, \; j = 1, \dots, d_i\}. \tag{4.14}$$

For a covering of the domain that satisfies the flat top property from assumption 4.2, any local basis function $\eta_{i,j}$ corresponds to a global basis function $\varphi_i \, \eta_{i,j}$ and in this case

$$\dim \mathcal{V}^{\mathrm{PU}}(\Omega) = \sum_{i=1}^{\mathrm{m}} d_i. \tag{4.15}$$

The flat top property can hence be used as a stability criterion. Moreover, note that for local approximation spaces only consisting of polynomials and a Partition of Unity of order r, the global basis functions are all r-times continuously differentiable as well. This shows, that higher global regularity is easy to achieve with the PUM, and that the PUM can (in contrast to FEM) be used without modifications to solve higher order problems.

The described construction works for local spaces consisting of scalar or vector-valued functions. For the experimental part of this thesis, it is assumed that a vector-valued local approximation space is the tensor product of scalar local spaces. Adding a vector-valued enrichment function means enriching the scalar spaces with the components of the enrichment.

4.3 Approximation properties

In the following theorem, the error estimates from [Sch03, Theorem 2.1] are formulated for the case of higher order approximations.

Theorem 4.1 (Approximation properties)**.** *Let $\Omega \subset \mathbb{R}^d$ be sufficiently regular and let $\{\omega_i\}_{i=1}^{\mathrm{m}}$ be a covering of Ω satisfying Assumptions 4.1 and 4.2. Moreover, let $\{\varphi_i\}_{i=1}^{\mathrm{m}}$ be a Partition of Unity subordinate to the cover that satisfies Assumption 4.3. Furthermore, let $\{\mathcal{V}(\omega_i)\}_{i=1}^{\mathrm{m}}$ be local approximation spaces, and $\mathcal{V}^{\mathrm{PU}}(\Omega)$ as in (4.13). Let $u \in \mathrm{H}^{\mathrm{k}}(\Omega)$ be the function to be approximated and suppose that there exist local functions $v_i \in \mathcal{V}(\omega_i)$ satisfying*

$$\|\partial^\alpha (u - v_i)\|_{L^2(\Omega \cap \omega_i)} \leqslant \varepsilon_{\alpha,i} \tag{4.16}$$

with constants $\varepsilon_{\alpha,i}$, for any $\alpha \in \mathbb{N}_0^d$, $|\alpha| \leqslant \mathrm{k}$ and all $i = 1, \dots, \mathrm{m}$. Then, the function

$$u^{\mathrm{PU}} := \sum_{i=1}^{\mathrm{m}} \varphi_i \, v_i \in \mathcal{V}^{\mathrm{PU}}(\Omega) \subset \mathrm{H}^{\mathrm{k}}(\Omega) \tag{4.17}$$

satisfies

$$\|\partial^\alpha (u - u^{\mathrm{PU}})\|_{L^2(\Omega)} \leqslant \sqrt{2^{|\alpha|} M} \left[\sum_{\substack{\beta \in \mathbb{N}_0^d \\ |\beta| \leqslant |\alpha|}} \sum_{i=1}^{\mathrm{m}} \left(\frac{C_\beta \varepsilon_{\alpha-\beta,i}}{(\operatorname{diam} \omega_i)^{|\beta|}} \right)^2 \right]^{\frac{1}{2}} \tag{4.18}$$

for all $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leqslant \mathrm{k}$.

◇

Proof. Let $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leqslant \mathrm{k}$ be arbitrary. Using the fact that $u = \sum_{i=1}^{\mathrm{m}} \varphi_i \, u$, it holds that

$$\begin{aligned} \partial^\alpha (u - u^{\mathrm{PU}}) &= \partial^\alpha \sum_{i=1}^{\mathrm{m}} \varphi_i (u - v_i) \\ &= \sum_{i=1}^{\mathrm{m}} \partial^\alpha [\varphi_i (u - v_i)] \\ &= \sum_{i=1}^{\mathrm{m}} \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leqslant \alpha}} \partial^\beta \varphi_i \, \partial^{\alpha-\beta} (u - v_i). \end{aligned} \tag{4.19}$$

Note that there are $2^{|\alpha|}$ multi-indices β in the inner sum, since the product rule is applied $|\alpha|$ times. From the inequality of arithmetic and geometric means, one can see for any fixed α that

$$\left[\sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leqslant \alpha}} \sum_{i=1}^{\mathrm{m}} \partial^\beta \varphi_i \, \partial^{\alpha-\beta} (u - v_i) \right]^2 \leqslant 2^{|\alpha|} \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leqslant \alpha}} \left[\sum_{i=1}^{\mathrm{m}} \partial^\beta \varphi_i \, \partial^{\alpha-\beta} (u - v_i) \right]^2 . \tag{4.20}$$

The pointwise overlap condition from Assumption 4.1 reads

$$\operatorname{card}\{i \,|\, x \in \omega_i\} \leqslant M, \quad \forall x \in \Omega, \tag{4.21}$$

implying that for any α and β from (4.20) it holds that

$$\left[\sum_{i=1}^{\mathrm{m}} \partial^\beta \varphi_i \, \partial^{\alpha-\beta} (u - v_i) \right]^2 \leqslant M \sum_{i=1}^{\mathrm{m}} \left[\partial^\beta \varphi_i \, \partial^{\alpha-\beta} (u - v_i) \right]^2 . \tag{4.22}$$

Combining everything,

$$\| \partial^\alpha (u - u^{\mathrm{PU}}) \|_{L^2(\Omega)}^2 \leqslant 2^{|\alpha|} M \sum_{i=1}^{\mathrm{m}} \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leqslant \alpha}} \| \partial^\beta \varphi_i \, \partial^{\alpha-\beta} (u - v_i) \|_{L^2(\Omega)}^2, \tag{4.23}$$

and the summands appearing on the right-hand side can be estimated as

$$\| \partial^\beta \varphi_i \, \partial^{\alpha-\beta} (u - v_i) \|_{L^2(\Omega)}^2 \leqslant \left(\frac{C_\beta}{(\operatorname{diam} \omega_i)^{|\beta|}} \right)^2 \varepsilon_{\alpha-\beta,i}^2. \tag{4.24}$$

Inserting (4.24) into (4.23) proves the claim,

$$\| \partial^\alpha (u - u^{\mathrm{PU}}) \|_{L^2(\Omega)} \leqslant \sqrt{2^{|\alpha|} M} \left[\sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leqslant \alpha}} \sum_{i=1}^{\mathrm{m}} \left(\frac{C_\beta \varepsilon_{\alpha-\beta,i}}{(\operatorname{diam} \omega_i)^{|\beta|}} \right)^2 \right]^{\frac{1}{2}} . \tag{4.25}$$

□

Remark 4.1. *Under the assumptions from Theorem 4.1, the estimates of the partial derivatives (4.18) imply estimates of the full* H^{k}*-norm,*

$$\begin{aligned} \|u - u^{\mathrm{PU}}\|_{\mathrm{H}^{\mathrm{k}}(\Omega)} &\overset{\text{Def.}}{=} \left[\sum_{\substack{\alpha \in \mathbb{N}_0^d \\ |\alpha| \leqslant \mathrm{k}}} \|\partial^\alpha (u - u^{\mathrm{PU}})\|^2_{L^2(\Omega)} \right]^{\frac{1}{2}} \\ &\overset{(4.18)}{\leqslant} \left[\sum_{\substack{\alpha \in \mathbb{N}_0^d \\ |\alpha| \leqslant \mathrm{k}}} 2^{|\alpha|} M \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leqslant \alpha}} \sum_{i=1}^{\mathrm{m}} \left(\frac{c_\beta \varepsilon_{\alpha-\beta,i}}{(\operatorname{diam} \omega_i)^{|\beta|}} \right)^2 \right]^{\frac{1}{2}}. \end{aligned} \tag{4.26}$$

*

The estimates from Theorem 4.1 show, that the global error is bounded by the sum of the local errors. Hence, it is important to choose local approximation spaces with good approximation properties. In the following, the Partition of Unity Method will be formulated as an h- and p-version. In order to do so, recall the Lemma of Bramble and Hilbert.

Lemma 4.2 (Bramble & Hilbert)**.** *Let* $\omega \subset \mathbb{R}^d$ *be a bounded Lipschitz domain. Let* $r \in \mathbb{N}$ *arbitrary and* $u \in \mathrm{H}^r(\omega)$*. Then, there exists a polynomial* $v \in \mathcal{P}^{r-1}(\omega)$ *such that for all* $t \in \{0, \ldots, r\}$

$$|u - v|_{\mathrm{H}^t(\omega)} \leqslant C(r, \omega) \operatorname{diam}(\omega)^{r-t} |u|_{\mathrm{H}^t(\omega)}, \tag{4.27}$$

with the Sobolev seminorms $|u|_{\mathrm{H}^t(\omega)}$ *on* $\mathrm{H}^t(\omega)$*.*

•

Proof. The proof can be found in [Bra07]. □

Since the true solution u of the problem under study is in $\mathrm{H}^{\mathrm{k}}(\Omega)$, it is also locally in $\mathrm{H}^{\mathrm{k}}(\omega_i)$ for all $i = 1, \ldots, \mathrm{m}$. Therefore, the Bramble Hilbert Lemma 4.2 can be applied locally on all patches, whenever Assumption 4.4 holds.

Assumption 4.4 (Patch size & polynomial degree)**.** *All patches* ω_i *for* $i = 1, \ldots, \mathrm{m}$ *have a similar size* h*, i.e.* $\operatorname{diam} \omega_i \approx h$ *for all* $i = 1, \ldots, \mathrm{m}$*. Also assume, that the local approximation spaces consist of polynomials of degree at least* $\mathrm{k} - 1$*, i.e.*

$$\mathcal{V}(\omega_i) = \mathcal{P}^{p_i}(\omega_i) + \mathcal{E}(\omega_i), \tag{4.28}$$

with $p_i \geqslant \mathrm{k} - 1$*.*

◇

Consequently, the local approximation quality can be expressed in terms of h and p_i as follows.

Corollary 4.3 (Local approximation as h-, p- & hp-version)**.** *Let $u \in \mathrm{H}^{\mathrm{k}}(\Omega)$ and let Assumptions 4.1 to 4.4 hold with $p_i = \mathrm{k}-1$ for all $i = 1, \ldots, \mathrm{m}$. Then, there exist local polynomials $v_i \in \mathcal{V}(\omega_i)$ satisfying*

$$|u - v_i|_{\mathrm{H}^r(\omega_i)} \leqslant C(\mathrm{k}, \omega_i) h^{p_i+1-r} |u^{\mathrm{PU}}|_{\mathrm{H}^r(\omega_i)}, \quad \forall i = 1, \ldots, \mathrm{m} \tag{4.29}$$

for all $r = 0, \ldots, \mathrm{k}$.

•

Proof. Since $u \in \mathrm{H}^{\mathrm{k}}(\omega_i)$ for all ω_i, $i = 1, \ldots, \mathrm{m}$, the Bramble Hilbert Lemma 4.2 can be applied on all patches, ensuring the existence of a polynomial $v_i \in \mathcal{P}^{\mathrm{k}-1}(\omega_i) \subset \mathcal{V}(\omega_i)$ satisfying

$$\begin{aligned} |u - v_i|_{\mathrm{H}^r(\omega_i)} &\leqslant C(\mathrm{k}, \omega_i) h^{\mathrm{k}-r} |u^{\mathrm{PU}}|_{\mathrm{H}^r(\omega_i)} \\ &= C(\mathrm{k}, \omega_i) h^{p_i+1-r} |u^{\mathrm{PU}}|_{\mathrm{H}^r(\omega_i)} \end{aligned} \tag{4.30}$$

for all $r = 0, \ldots, \mathrm{k}$. □

Corollary 4.3 shows, that the Partition of Unity method can be used as h-, p- or hp-version.

Remark 4.2. *From Corollary 4.3 it especially follows, that the local approximation $v_i \in \mathcal{V}(\omega_i)$ satisfies*

$$\|u - v_i\|_{L^2(\omega_i)} \leqslant C(\mathrm{k}, \omega_i) h^{p_i+1} \|u^{\mathrm{PU}}\|_{L^2(\omega_i)} \tag{4.31}$$

and

$$\begin{aligned} \|\nabla(u - v_i)\|^2_{L^2(\omega_i)} &= \sum_{i=1}^{d} \|\partial_{x_i}(u - v_i)\|^2_{L^2(\omega_i)} \\ &= \sum_{\substack{\alpha \in \mathbb{N}_0^d \\ |\alpha|=1}} \|\partial^\alpha (u - v_i)\|^2_{L^2(\omega_i)} \\ &= |u - v_i|^2_{\mathrm{H}^1(\omega_i)} \\ &\leqslant C(\mathrm{k}, \omega_i) h^{p_i} |u^{\mathrm{PU}}|_{\mathrm{H}^1(\omega_i)} \end{aligned} \tag{4.32}$$

and similar estimates hold for derivatives of higher order.

*

Remark 4.3. *The estimates from Corollary 4.3 can directly be used as the bounds $\varepsilon_{\alpha,i}$ in (4.16) from Theorem 4.1, yielding global estimates of the Partition of Unity Method in terms of the local polynomial degree and the diameter of the patches.*

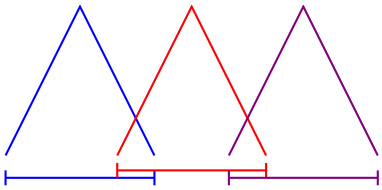
*

4.4 Practical details of the cover construction

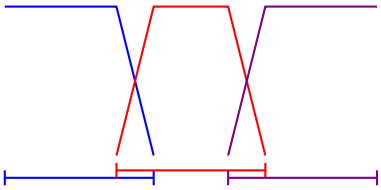
In this section, the practical process of constructing the cover and a corresponding Partition of Unity is briefly described. After defining the spatial domain Ω, a bounding box of Ω is determined, which on its own forms an open cover. This cover is said to be of level 0. Subsequent levels $i+1$ are constructed hierarchically by splitting all patches from the cover on level i at hand in half in all coordinate directions. Newly created patches that do not have an intersection with Ω are discarded. Hence, the number of patches increases roughly by a factor of 2^d for each new level and spatial dimension d. Since the Partition of Unity Method is intended to work on discretizations consisting of overlapping patches, the patches are stretched in all coordinate directions by a factor $\kappa > 1$. By choosing this *stretch factor* appropriately, Assumptions 4.1 and 4.2 are satisfied automatically from the cover construction. After the cover was generated, a Partition of Unity subordinate to this cover may be constructed. A straight-forward approach to do this is to first define weight functions $\{W_i\}_{i=1}^{\mathrm{m}}$ for the patches, and taking their weighted sums,

$$\varphi_i := \frac{W_i}{\sum_{j=1}^{\mathrm{m}} W_j}, \quad \forall i = 1, \ldots, \mathrm{m}, \tag{4.33}$$

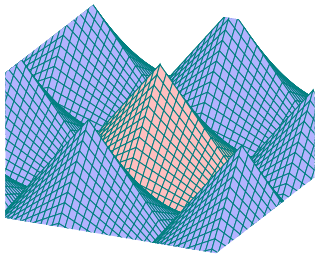
The functions $\{\varphi\}_{i=1}^{\mathrm{m}}$ form a Partition of Unity, and the method of construction is referred to as *Shepards approach*. Figure 4.2 shows possible weight functions and the corresponding Shepard PU on an exemplary discretization consisting of overlapping patches in one and two spatial dimensions. The two-dimensional weight functions are tensor products of one-dimensional weight functions, and the figure also shows overlapping patches.



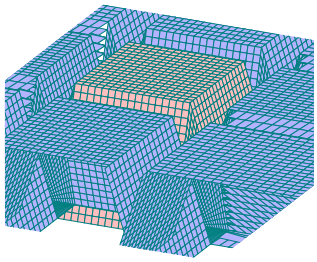
(a) Discretization of a one dimensional domain into overlapping integrals, and weight functions based on $|x|$



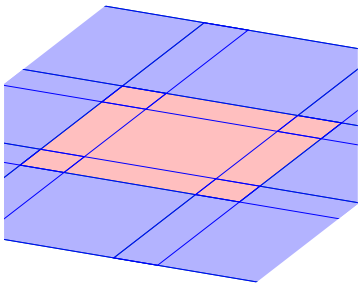
(b) Shepard PU resulting from the weight functions from (a)



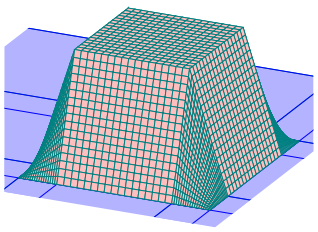
(c) Weight functions in 2d that are tensor products of the 1d weight functions from (a)



(d) Shepard PU resulting from the weight functions from (c)



(e) Cover made from overlapping rectangles used in (c) and (d)



(f) PU function corresponding to the center patch from (d)

Figure 4.2: Weight functions based on the absolute value and the corresponding Shepard PU functions. In two spatial dimensions, the weight functions are tensor products of the one-dimensional weight functions. In the flat top region of each patch, the corresponding PU function is equal to 1.

5

Optimal basis functions

Finding the best approximation to a given element in a certain space is a classical problem in numerical analysis. In the case of functions which are to be best-approximated in a Sobolev space, the first quantitative, yet abstract estimates were presented by Andrey Kolmogorov in 1936 ([Kol36]). In the works of Tichomirov, Babadjanov and Garkavi ([TB67, Gar62]), the original definitions were modified and various explicit formulas for worst-case best approximation errors, together with estimates for certain other cases of function spaces, were computed. Conditions for the existence of subspaces with best-possible approximation quality were developed in [Bro64] (Theorem 6) and extended in the works of Singer and Pinkus ([Sin13, Pin85]). In the latter, Theorem 2.2 in chapter 4 also gives an explicit representation of the optimal subspaces in terms of the solutions of a generalized eigenvalue problem. In the works of Babuška and Lipton ([BL11]), this representation was exploited in order to compute optimal local approximation spaces for second-order elliptic partial differential equations. The performance of the so-constructed spaces is superior to that of standard polynomial approximation spaces, and hence allows for an improvement of the quality of numerical solutions to second-order partial differential equations, without the need to use heavy spatial refinement. As mentioned before, it is possible to obtain analytical enrichments for certain combinations of differential operators, domains and data, for example using the stress recovery method ([Mel05]). Whenever such analytical enrichments are available, they should be used, but this is only the case for very specific problems. In Section 5.1, the framework presented in [BL11], which allows to compute operator-dependent numerical enrichments with superior approximation qualities, is recapitulated. While the original publication considered second-order elliptic problems, the framework is extended to the more general case of elliptic partial differential equations of even order. Since the framework is formulated for the case of homogeneous partial differential equations, this section also introduces the concept of lifting of solutions, allowing to handle non-homogeneous data via particular solutions. Section 5.2 presents practical aspects for the computation of locally optimal approximation spaces, which will be applied in the experimental section of this thesis. The computation of optimal local approximation spaces can be done in an offline phase, prior to any global computation using these spaces. As a result, a coarse global computation using optimal basis functions will provide results that are comparable to computations on much finer discretizations, while requiring only a fraction of the corresponding degrees of freedom. Nevertheless, the offline phase requires a substantial amount of numerical work. Section 5.3 therefore investigates and develops conditions which allow the optimal shape functions to

be reused in changed settings.

5.1 Theoretical construction of optimal bases

In the previous Chapter 4, the Partition of Unit Method was introduced. For a patch $\omega \subset \Omega$, it was seen that the corresponding local approximation space used in the Partition of Unity Method, $\mathcal{V}(\omega)$, takes the general form

$$\mathcal{V}(\omega) = \mathcal{P}(\omega)^p + \mathcal{E}(\omega), \tag{5.1}$$

where $\mathcal{P}(\omega)^p$ denotes the space of polynomials up to a certain order p and $\mathcal{E}(\omega)$ is referred to as the space of enrichments. A fundamental property of the Partition of Unity Method is, that all local approximation spaces are independent of each other, and the space of enrichments $\mathcal{E}(\omega)$ should hence be chosen in a way that maximizes the approximation power of the whole local space $\mathcal{V}(\omega)$. This section is concerned with revisiting and extending the framework from [BL11], allowing for a smart choice of the space of enrichments $\mathcal{E}(\omega)$. The theory is formulated for homogeneous problems, and hence the method of lifting of solutions will be presented beforehand. Lifting of solutions allows to split the solution of the original problem into a sum of particular solutions and function satisfying the homogeneous problem. The latter encodes fine-scale information which is inherent to the differential operator appearing in the problem under study.

5.1.1 Lifting of solutions and homogeneous sampling problem

This section describes a decomposition of the solution of a partial differential equation which is referred to as *lifting of solutions*. Consider a general PDE in the form of Problem 1, which is specified by the tuple

$$(\mathrm{k}, f, g^0, \ldots, g^{\mathrm{k}-1}, \Gamma_0, \ldots, \Gamma^{\mathrm{k}-1}) \tag{5.2}$$

describing the order of the partial differential equation, the load, boundary values and boundary parts. For clarity, recall the definition of Problem 1.

Problem 1: General Partial Differential Equation.

Let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ be open and bounded, let $\mathcal{L}$ be a linear differential operator of order $\mathrm{k} \in \mathbb{N}$ as in (2.1) and let $\mathrm{n} \in \{1, d\}$. Find a function $u \in [\mathcal{C}^{\mathrm{k}}(\Omega)]^{\mathrm{n}}$ satisfying

$$\begin{aligned} \mathcal{L}\, u &= f, \quad \text{in } \Omega \\ \mathcal{B}^0\, u &= g^0, \quad \text{on } \Gamma^0 \subset \partial\,\Omega \\ &\vdots \\ \mathcal{B}^{\mathrm{k}-1}\, u &= g^{\mathrm{k}-1}, \quad \text{on } \Gamma^{\mathrm{k}-1} \subset \partial\,\Omega \end{aligned}$$

where $\mathcal{B}^m$ are (piecewise) linear differential operators of order $m = 0, \ldots, \mathrm{k}-1$ defined on corresponding parts of the boundary and $f : \Omega \to \mathbb{R}^{\mathrm{n}}$, $g^m : \Gamma^m \to \mathbb{R}^{\mathrm{n}}, m = 0, \ldots, \mathrm{k}-1$ are sufficiently smooth functions.

Since the partial differential operator $\mathcal{L}$ and all partial differential boundary operators $\mathcal{B}^i$ for $i = 0, \ldots, \mathrm{k}-1$ are linear, the superposition principle can be used to express the solution $u \in [\mathcal{C}^{\mathrm{k}}(\Omega)]^{\mathrm{n}}$ in the form $u = u_f + u_0 + u_1 + \ldots + u_{\mathrm{k}-1} + u_{\mathrm{hom}}$, where

$$\begin{aligned} u_f \text{ solves Problem 1 with } & (\mathrm{k}, f, 0, \ldots, 0, \Gamma^0, \ldots, \Gamma^{\mathrm{k}-1}) \\ u_0 \text{ solves Problem 1 with } & (\mathrm{k}, 0, g^0, 0, \ldots, 0, \Gamma^0, \ldots, \Gamma^{\mathrm{k}-1}) \\ \vdots & \\ u_{\mathrm{k}-1} \text{ solves Problem 1 with } & (\mathrm{k}, 0, \ldots, 0, g^{\mathrm{k}-1}, \Gamma^0, \ldots, \Gamma^{\mathrm{k}-1}) \\ u_{\mathrm{hom}} \text{ solves Problem 1 with } & (\mathrm{k}, 0, \ldots, 0, \Gamma^0, \ldots, \Gamma^{\mathrm{k}-1}) \end{aligned} \tag{5.3}$$

In the above equation (5.3), 0 denotes the zero vector in the case of a vector-valued partial differential equation. The functions $u_f, u_0, \ldots, u_{\mathrm{k}-1}$ are called *particular solutions* for the given data $f, g^0, \ldots, g^{\mathrm{k}-1}$. The function u_{hom} satisfies the homogeneous problem, in which all data has been replaced by zero. It encodes fine-scale information which is inherent to the involved differential operators $\mathcal{L}, \mathcal{B}^0, \ldots, \mathcal{B}^{\mathrm{k}-1}$ and totally independent of the data of the original problem.

In general, the homogeneous problem has an infinite number of solutions, among them the zero function. The computation of the unique function u_{hom} from the lifting is no easy task, and a method to approximate it is investigated in detail in Section 5.1.2.

Remark 5.1. *Consider a second-order partial differential equation in divergence form that is subject to classical Dirichlet and Neumann boundary conditions, i.e. find a function u solving*

$$\begin{aligned} -\operatorname{div}(A\nabla u) + b \cdot \nabla u + cu &= f, && \text{in } \Omega \\ u &= g, && \text{on } \Gamma_D \subset \partial\Omega \\ A\nabla u \cdot \vec{n} &= h, && \text{on } \Gamma_N \coloneqq \partial\Omega \backslash \Gamma_D \end{aligned} \tag{5.4}$$

for sufficiently regular data A, b, c, g, h. Using particular solutions, i.e. a function u_f solving

$$\begin{aligned} -\operatorname{div}(A\nabla u_f) + b \cdot \nabla u_f + cu_f &= f, && \text{in } \Omega \\ u_f &= 0, && \text{on } \Gamma_D \subset \partial\Omega \\ A\nabla u_f \cdot \vec{n} &= 0, && \text{on } \Gamma_N \coloneqq \partial\Omega \backslash \Gamma_D \end{aligned} \tag{5.5}$$

another function u_g solving

$$\begin{aligned} -\operatorname{div}(A\nabla u_g) + b \cdot \nabla u_g + cu_g &= 0, && \text{in } \Omega \\ u_g &= g, && \text{on } \Gamma_D \subset \partial\Omega \\ A\nabla u_g \cdot \vec{n} &= 0, && \text{on } \Gamma_N \coloneqq \partial\Omega \backslash \Gamma_D \end{aligned} \tag{5.6}$$

and a function u_h solving

$$\begin{aligned} -\operatorname{div}(A\nabla u_h) + b \cdot \nabla u_h + cu_h &= 0, && \text{in } \Omega \\ u_h &= 0, && \text{on } \Gamma_D \subset \partial\Omega \\ A\nabla u_h \cdot \vec{n} &= h, && \text{on } \Gamma_N \coloneqq \partial\Omega \backslash \Gamma_D \end{aligned} \tag{5.7}$$

as well as a function u_{hom} satisfying the homogeneous problem

$$\begin{aligned} -\operatorname{div}(A\nabla u_{hom}) + b\cdot\nabla u_{hom} + cu_{hom} &= 0, \quad \textit{in } \Omega \\ u_{hom} &= 0, \quad \textit{on } \Gamma_D \subset \partial\Omega \\ A\nabla u_{hom}\cdot\vec{n} &= 0, \quad \textit{on } \Gamma_N := \partial\Omega\backslash\Gamma_D \end{aligned} \tag{5.8}$$

a solution of the original problem is given by $u = u_f + u_g + u_h + u_{hom}$.

*

Lifting of local solutions

The superposition principle described before will be extensively used throughout this thesis and the remainder of this chapter specifically. Optimal shape functions, whose construction will be presented in Section 5.1.2, are however computed and employed locally. In fact, not even the particular solutions for the data should be computed on a global scale, and hence the previously described lifting of solutions is localized in the following. The locally computed particular solutions, as well as the correct function satisfying the local homogeneous problem, are then used as enrichments on the corresponding patches in a global Partition of Unity Method. The restriction of Problem 1 to a subdomain $\omega \subset \Omega$ is presented in Problem 10.

Problem 10: Local General Partial Differential Equation.

Let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ be open and bounded, $\omega \subset \Omega$, let $\mathcal{L}$ be a linear differential operator of order $\mathrm{k} \in \mathbb{N}$ as in (2.1) and let $\mathrm{n} \in \{1, d\}$. Find a function $u_\omega \in [\mathcal{C}^{\mathrm{k}}(\omega)]^{\mathrm{n}}$ satisfying

$$\begin{aligned} \mathcal{L}\,u &= f, \quad \text{in } \omega \\ \mathcal{B}^0\,u &= g^0, \quad \text{on } \Gamma^0_\omega := \Gamma^0 \cap \partial\,\omega \\ &\vdots \\ \mathcal{B}^{\mathrm{k}-1}\,u &= g^{\mathrm{k}-1}, \quad \text{on } \Gamma^{\mathrm{k}-1}_\omega := \Gamma^{\mathrm{k}-1} \cap \partial\,\omega \end{aligned} \tag{5.9}$$

where $\mathcal{B}^m$ are (piecewise) linear differential operators of order $m = 0, \ldots, \mathrm{k}-1$ defined on corresponding parts of the boundary and $f : \Omega \to \mathbb{R}^{\mathrm{n}}$, $g^m : \Gamma^m \to \mathbb{R}^{\mathrm{n}}$, $m = 0, \ldots, \mathrm{k}-1$ are sufficiently smooth functions from the corresponding global problem.

It is worth noticing that the local Problem 10 does not necessarily have to be well-posed, even if the global problem was. The reason is, that Γ^i_ω may be empty for some $i \in \{0, \ldots, \mathrm{k}-1\}$. Consider for example the case that there is no boundary condition of order 0, i.e. $\Gamma^0_\omega = \varnothing$, making solutions at most unique up to a constant. However, one has total freedom in adding additional boundary data on the faces of ω which are not fixed by the globally imposed boundary conditions in order to restore the well-posedness of the problem. The superposition principle presented before

implies that the solution u_ω can be written as the sum $u_\omega = u_{\omega,f} + u_{\omega,0} + \ldots + u_{\omega,\mathrm{k}-1} + u_{\omega,\mathrm{hom}}$, where $u_{\omega,f}, u_{\omega,0}, \ldots, u_{\omega,\mathrm{k}-1}$ are particular solutions corresponding to the data, and $u_{\omega,\mathrm{hom}}$ is a local function satisfying the homogeneous problem, i.e.

$$\begin{aligned}
u_f \text{ solves Prb. 1 with} \quad & (\mathrm{k}, f, 0, \ldots, 0, \Gamma_\omega^0, \ldots, \Gamma_\omega^{\mathrm{k}-1}) \\
u_0 \text{ solves Prb. 1 with} \quad & (\mathrm{k}, 0, g^0, 0, \ldots, 0, \Gamma_\omega^0, \ldots, \Gamma_\omega^{\mathrm{k}-1}) \\
\vdots & \\
u_{\mathrm{k}-1} \text{ solves Prb. 1 with} \quad & (\mathrm{k}, 0, \ldots, 0, g^{\mathrm{k}-1}, \Gamma_\omega^0, \ldots, \Gamma_\omega^{\mathrm{k}-1}) \\
u_{\mathrm{hom}} \text{ solves Prb. 1 with} \quad & (\mathrm{k}, 0, \ldots, 0, \Gamma_\omega^0, \ldots, \Gamma_\omega^{\mathrm{k}-1}).
\end{aligned} \tag{5.10}$$

Remark 5.2. *It is not always necessary to compute all parts of the decomposed solution: Consider for example an interior patch $\omega \subset \Omega$ with $\omega \cap \partial\,\Omega = \varnothing$. In this case, a particular solutions to any boundary condition function g^i for $i \in \{0, \ldots, \mathrm{k}-1\}$ is only defined by artificially prescribed boundary values, which need to be added to the problem in order to make it uniquely solvable. In fact, such particular solutions are again solutions of the homogeneous problem.*

*

Remark 5.3. *It may also happen, that a particular solution is a constant or a lower order polynomial, which is contained in the local approximation space anyway. Adding this function as enrichment will not improve the approximation quality of the local approximation space.*

*

Solving the homogeneous local problem requires the prescription of artificial boundary conditions. Otherwise, the zero function is a trivial solution for any PDE of order $\mathrm{k} > 0$. In general, there is an infinite-dimensional function space spanned by the solutions of the homogeneous problem, and samples from this space can be computed by solving the local homogeneous sampling Problem 11 for a given function b, which is imposed as boundary condition on the *free boundary* $\partial\,\omega \setminus \partial\,\Omega$ of the local patch ω.

Problem 11: Local homogeneous sampling problem.

Let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ be open and bounded, $\omega \subset \Omega$, let $\mathcal{L}$ be a linear differential operator of order $\mathrm{k} \in \mathbb{N}$ as in (2.1) and let $\mathrm{n} \in \{1, d\}$. Furthermore, let $\Gamma_\omega^{\mathrm{F}} := \partial\omega \setminus \partial\Omega$ and $b : \Gamma_\omega^{\mathrm{F}} \to \mathbb{R}^{\mathrm{n}}$ sufficiently regular. Find a function $u_\omega \in [\mathcal{C}^{\mathrm{k}}(\omega)]^{\mathrm{n}}$ satisfying

$$\begin{aligned} \mathcal{L}\, u^b_{\omega,\mathrm{hom}} &= 0, \quad \text{in } \omega \\ \mathcal{B}^0\, u^b_{\omega,\mathrm{hom}} &= 0, \quad \text{on } \Gamma^0_\omega \\ &\vdots \\ \mathcal{B}^{\mathrm{k}-1}\, u^b_{\omega,\mathrm{hom}} &= 0, \quad \text{on } \Gamma^{\mathrm{k}-1}_\omega \\ u^b_{\omega,\mathrm{hom}} &= b, \quad \text{on } \Gamma^{\mathrm{F}}_\omega . \end{aligned} \tag{5.11}$$

where $\mathcal{B}^m$ are the (piecewise) linear differential operators of order $m = 0, \ldots, \mathrm{k}-1$ defined on corresponding parts of the boundary from the global problem, and 0 denoting the zero vector in the case of $\mathrm{n} = d$.

Remark 5.4. *The sampling problem can also be defined differently: Instead of prescribing the value of the solution on the free boundary, other conditions such as a divergence may be prescribed.*

Depending on the location of ω in Ω and the order of the PDE it may be necessary to prescribe additional conditions on the free boundary to ensure well-posedness of the sampling problem. For the sake of simplicity, it will be assumed throughout the remainder of this thesis that Problem 11 is well-posed.*

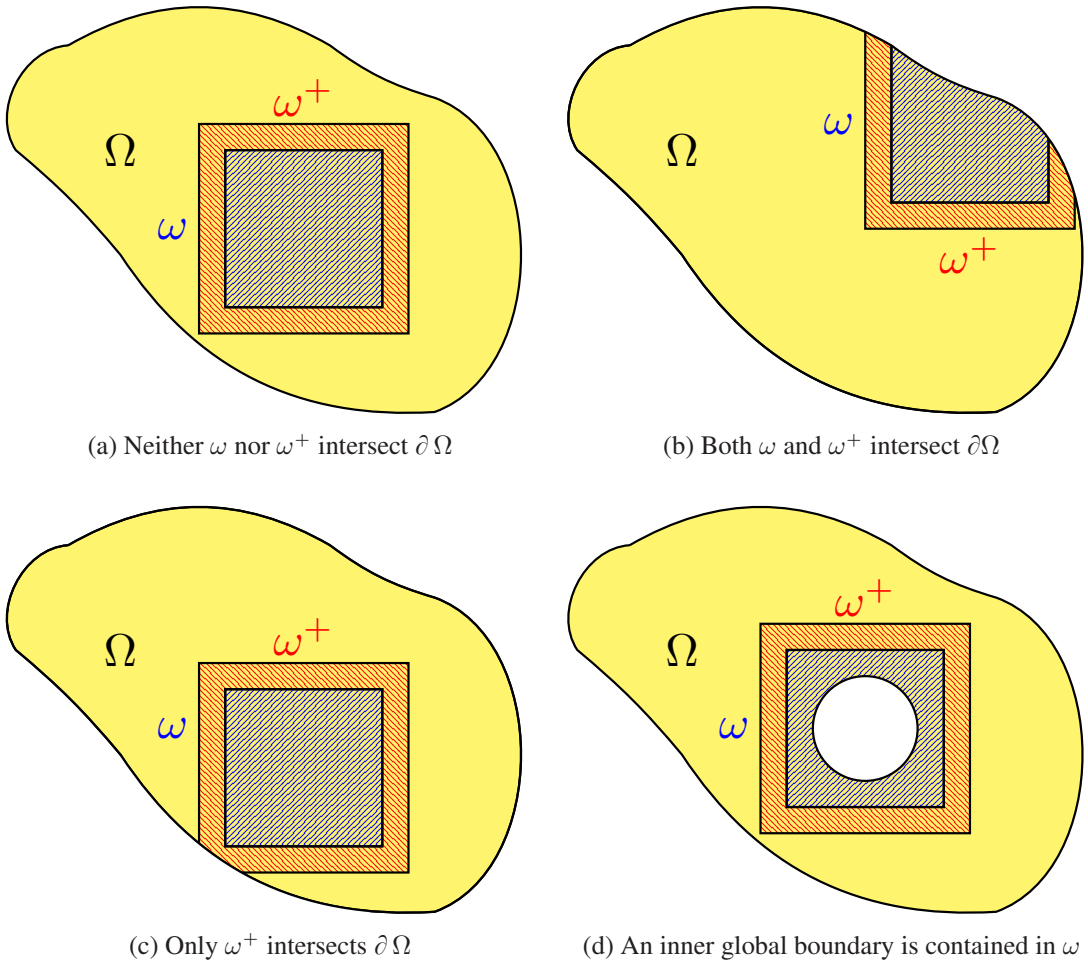
As mentioned before, the solutions of the homogeneous problem and hence the form of an explicit basis of the space of local weakly harmonic functions are only known for very few partial differential operators. In all other cases, the space of local weakly harmonic functions needs to be sampled, for example using the sampling Problem 11. The process of identifying harmonic functions encoding valuable fine-scale information is described in the following Section 5.1.2.

5.1.2 Construction of optimal bases

As seen in the previous Section 5.1.1, the solution of a partial differential equation can be expressed locally as a sum of particular solutions corresponding to the data of the problem, and a function that satisfies the homogeneous problem. The space of homogeneous solutions is in general infinite-dimensional, and the true homogeneous solution appearing in the lifting can only be computed from the sampling Problem 11 if the correct boundary boundary value b is known. This, however, requires knowledge of the true solution, which is not available. Instead, a low-dimensional space of harmonic functions can be constructed, such that the best-approximation error of $u_{\omega,\mathrm{hom}}$ in this space is arbitrarily small. A framework for the construction of such best-possible approximation

spaces is presented in the following.

The original publication [BL11] introduced the construction in the case of second-order elliptic partial differential equations, and in this section the framework will be extended to the more general case of elliptic partial differential equations of even order. Consider the general Problem 1 and suppose that it is of even order $2\,\mathrm{k}$ and scalar, i.e. $\mathrm{n} = 1$, in order to simplify the notation. Similar results hold for the vector-valued case. The construction is done locally on a subdomain $\omega \subset \Omega$, where ω is a patch, resp. a collection of several patches in the context of the Partition of Unity Method (cf. Chapter 4). Improving the local approximation quality on ω will also improve the global error bound from Theorem 4.1. In order to construct the local approximation space on ω, an oversampled version ω^+ of ω must be invoked, which can be obtained by stretching ω with factors $\tau_1, \ldots, \tau_d > 1$ in all coordinate direction. Possible geometric relations between ω and ω^+ are sketched in Figure 5.1.



(a) Neither ω nor ω^+ intersect $\partial\,\Omega$

(b) Both ω and ω^+ intersect $\partial\Omega$

(c) Only ω^+ intersects $\partial\,\Omega$

(d) An inner global boundary is contained in ω

Figure 5.1: Possible geometrical relations between ω, ω^+ and Ω. The

The problem under study is transformed into its variational formulation, and the corresponding bilinear form is denoted by $\mathbf{a}_\Omega[\cdot,\cdot]$, where the subscript is added in order to indicate the integration domain. $\mathcal{R}$ denotes the space of rigid body modes, that is the space of functions whose contributions to the solution are fixed by imposing suitable boundary conditions. Hence, the homogeneous

problem at hand is well-posed for

$$\mathbf{a}_\Omega : \mathrm{H}^{\mathrm{k}}(\Omega)_{/\mathcal{R}} \times \mathrm{H}^{\mathrm{k}}(\Omega)_{/\mathcal{R}} \to \mathbb{R}, \tag{5.12}$$

and the corresponding linear functional defined on this quotient space.

Remark 5.5. *The bilinear form corresponding to the Laplace operator,*

$$\mathbf{a}_\Omega[u,v] \coloneqq \int_\Omega \nabla u \cdot \nabla v \, dx, \tag{5.13}$$

is a scalar product on $\mathrm{H}^1(\Omega)_{/\mathcal{R}}$ *with* $\mathcal{R} = \operatorname{span}\{1\}$.

The rigid body modes corresponding to three-dimensional linear elasticity,

$$\mathbf{a}_\Omega[u,v] \coloneqq \int_\Omega \sigma(u) : \varepsilon(v) \, dx, \tag{5.14}$$

are translations and rotations, i.e.

$$\mathcal{R} \coloneqq \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} x_2 \\ -x_1 \\ 0 \end{bmatrix}, \begin{bmatrix} x_3 \\ 0 \\ -x_1 \end{bmatrix}, \begin{bmatrix} 0 \\ x_3 \\ -x_2 \end{bmatrix} \right\}. \tag{5.15}$$

*

Similarly, the local problems on ω and ω^+ are well-posed for

$$\begin{aligned} \mathbf{a}_\omega &: \mathrm{H}^{\mathrm{k}}(\omega)_{/\mathcal{R}} \times \mathrm{H}^{\mathrm{k}}(\omega)_{/\mathcal{R}} \to \mathbb{R}, \\ \mathbf{a}_{\omega^+} &: \mathrm{H}^{\mathrm{k}}(\omega^+)_{/\mathcal{R}} \times \mathrm{H}^{\mathrm{k}}(\omega^+)_{/\mathcal{R}} \to \mathbb{R}, \end{aligned} \tag{5.16}$$

with corresponding linear functionals, and the bilinear forms define local energy inner products

$$\langle u,v\rangle_{\mathcal{E}(\omega)} \coloneqq \mathbf{a}_\omega[u,v], \quad \langle u,v\rangle_{\mathcal{E}(\omega^+)} \coloneqq \mathbf{a}_{\omega^+}[u,v]. \tag{5.17}$$

Local spaces of weakly harmonic functions are defined as

$$\begin{aligned} \mathcal{H}(\omega) &\coloneqq \{u \in \mathrm{H}^{\mathrm{k}}(\omega)_{/\mathcal{R}} |\ \mathbf{a}_\omega[u,\varphi] = 0,\ \forall \varphi \in \mathcal{C}_0^\infty(\omega)\} \\ \mathcal{H}(\omega^+) &\coloneqq \{u \in \mathrm{H}^{\mathrm{k}}(\omega^+)_{/\mathcal{R}} |\ \mathbf{a}_{\omega^+}[u,\varphi] = 0,\ \forall \varphi \in \mathcal{C}_0^\infty(\omega^+)\} \end{aligned} \tag{5.18}$$

and by definition it holds that the weak local homogeneous solutions satisfy $u_{\omega,\mathrm{hom}} \in \mathcal{H}(\omega)$, resp. $u_{\omega^+,\mathrm{hom}} \in \mathcal{H}(\omega^+)$. A 'good' local approximation space for $u_{\omega^+,\mathrm{hom}}$ must hence be a subspace of $\mathcal{H}(\omega)$ and its construction will involve the harmonic space on the oversampled patch $\mathcal{H}(\omega^+)$. Furthermore, the construction relies on a Cacciopoly-type inequality presented in Theorem 5.1, which allows to estimate the energy of a function on ω in terms of the $\mathrm{H}^{\mathrm{k}-1}$ norm on ω^+.

In the original publication [BL11], the difference between the construction on inner patches and patches touching the boundary was highlighted, and Theorems 5.1 and 5.2 were shown for both cases and $\mathrm{k} = 2$. In the latter case, the construction of optimal local approximation spaces is technically more involved, but the results coincide. For the following generalization, the geometric relation $\omega \subset \omega^+ \subset \Omega$ with $\omega^+ \cap \partial\,\Omega = \varnothing$ will be assumed.

Theorem 5.1 (Caccioppoli-type inequality for local energy norm)**.** *Consider a uniformly elliptic differential operator $\mathcal{L}$ of order $2\,\mathrm{k}$, and assume the corresponding elliptic bilinear form $\mathbf{a}_\Omega : \mathrm{H}^{\mathrm{k}}(\Omega)_{/\mathcal{R}} \times \mathrm{H}^{\mathrm{k}}(\Omega)_{/\mathcal{R}} \to \mathbb{R}$ takes the form*

$$\mathbf{a}_\Omega[u,v] \coloneqq \sum_{\substack{\alpha,\beta\in\mathbb{N}_0^d \\ |\alpha|,|\beta|\leqslant \mathrm{k}}} \int_\Omega a_{\alpha,\beta}\partial^\alpha u \partial^\beta v \, dx \tag{5.19}$$

with coefficients $a_{\alpha,\beta} \in \mathrm{L}^\infty(\Omega)$ and $a_{\alpha,\beta} \in \mathbb{R}$ constant for any terms involving derivatives of the highest order k, i.e. $\max\{|\alpha|,|\beta|\} = \mathrm{k}$. Then, for all $u \in \mathcal{H}(\omega^+)$ the following inequality holds,

$$\|u\|_{\mathcal{E}(\omega)} \leqslant C\|u\|_{\mathrm{H}^{\mathrm{k}-1}(\omega^+)}. \tag{5.20}$$

◇

Proof. Let $u \in \mathcal{H}(\omega^+)$ be a harmonic function, i.e. $a_{\omega^+}[u,\eta] = 0$ for all $\mathcal{C}_0^\infty(\omega^+)$. Since $\omega \subset \omega^+$ with $\operatorname{dist}(\partial\,\omega, \partial\,\omega^+) > 0$, there exists a cutoff (bump) function $\eta \in \mathcal{C}_0^\infty(\omega^+)$ with $\eta = 1$ on ω. The funtion η, as well as all of its partial derivatives have zero trace on $\partial\,\omega^+$, which proves especially useful during integration by parts. Using the chain rule,

$$\begin{aligned} 0 \quad &= \quad \mathbf{a}_{\omega^+}[u,\eta u] = \sum_{\substack{\alpha,\beta\in\mathbb{N}_0^d \\ |\alpha|,|\beta|\leqslant \mathrm{k}}} \int_{\omega^+} a_{\alpha,\beta}\partial^\alpha u \partial^\beta(\eta u)\, dx \\ &= \sum_{\substack{\alpha,\beta\in\mathbb{N}_0^d \\ |\alpha|,|\beta|\leqslant \mathrm{k}}} \sum_{\substack{\gamma\in\mathbb{N}_0^d \\ \gamma\leqslant\beta}} \int_{\omega^+} a_{\alpha,\beta}\partial^\alpha u \partial^\gamma \eta \partial^{\beta-\gamma} u \, dx, \end{aligned} \tag{5.21}$$

and by bringing all terms for $|\gamma| > 0$ to the other side,

$$\begin{aligned} &\sum_{\substack{\alpha,\beta\in\mathbb{N}_0^d \\ |\alpha|,|\beta|\leqslant \mathrm{k}}} \int_{\omega^+} a_{\alpha,\beta}\partial^\alpha u \partial^\beta u \eta \, dx \\ &\qquad = - \sum_{\substack{\alpha,\beta\in\mathbb{N}_0^d \\ |\alpha|,|\beta|\leqslant \mathrm{k}}} \sum_{\substack{\gamma\in\mathbb{N}_0^d \\ 0<|\gamma| \\ \gamma\leqslant\beta}} \int_{\omega^+} a_{\alpha,\beta}\partial^\alpha u \partial^\gamma \eta \partial^{\beta-\gamma} u \, dx. \end{aligned} \tag{5.22}$$

Next, the terms appearing on the right-hand side are estimated individually. Since $|\gamma| > 0$, also $|\beta - \gamma| \leqslant \mathrm{k}-1$. For any α with $|\alpha| < \mathrm{k}$, the Hölder and Cauchy Schwarz inequalities are applied to see that

$$\int_{\omega^+} a_{\alpha,\beta}\partial^\alpha u \partial^\gamma \eta \partial^{\beta-\gamma} u \, dx \leqslant \|a_{\alpha,\beta}\|_{L^\infty(\omega^+)} \|\partial^\gamma \eta\|_{C^\infty(\omega^+)} \|u\|^2_{\mathrm{H}^{\mathrm{k}-1}(\omega^+)}. \tag{5.23}$$

In the case of $|\alpha| = \mathrm{k}$ and $|\beta - \gamma| < \mathrm{k}-1$, apply integration by parts, having in mind that $a_{\alpha,\beta}$ is constant by assumption. Let $i \in \{1,\dots,d\}$ such that $\alpha_i > 0$, then $\alpha = \alpha - e_i + e_i$ with i-th unit

vector e_i, and

$$\begin{aligned}\int_{\omega^+} \partial^\alpha u \partial^\gamma \eta \partial^{\beta-\gamma} u \, dx &= \int_{\omega^+} \partial^{\alpha-e_i+e_i} u \partial^\gamma \eta \partial^{\beta-\gamma} u \, dx \\ &= -\int_{\omega^+} \partial^{\alpha-e_i} u \partial^{e_i} \left(\partial^\gamma \eta \partial^{\beta-\gamma} u \right) \, dx \\ &= -\int_{\omega^+} \partial^{\alpha-e_i} u \partial^{\gamma+e_i} \eta \partial^{\beta-\gamma} u + \partial^{\alpha-e_i} u \partial^\gamma \eta \partial^{\beta-\gamma+e_i} u \, dx\end{aligned} \tag{5.24}$$

Due to the assumptions on the multiindices, it holds that

$$|\alpha - e_i| = \mathrm{k} - 1, \quad |\beta - \gamma| < |\beta - \gamma + e_i| \leqslant \mathrm{k} - 1, \tag{5.25}$$

and similar estimates to (5.23) can be performed for both integrands with $|a_{\alpha,\beta}|$ instead of $\|a_{\alpha,\beta}\|_{\mathrm{L}^\infty(\omega^+)}$. In the last case, consider $|\alpha| = \mathrm{k}$ and $|\beta - \gamma| = \mathrm{k} - 1$, again noting that $a_{\alpha,\beta}$ is constant and can hence be regarded as a multiplicative factor of the integral. Apply integration by parts once to remove one partial derivative from $\alpha = \alpha - e_i + e_i$, and the calculation is identical to the one in (5.24). However, note that now $|\beta - \gamma + e_i| = k$, and the estimate from (5.23) can only be applied for one of the two integrands. Integration by parts is now applied again, this time removing a partial derivative from $\beta - \gamma + e_i = \beta - \gamma + e_i - e_j + e_j$. Again, one of the terms can be estimated as in (5.23), while for the other the process is iterated. All in all, integration by parts is applied over and over again to move all partial derivatives from $\beta - \gamma$ and α to the corresponding other part of the integrand. Each time, an additional term appears involving a partial derivative of η of maximum order $|\gamma| + 2\,\mathrm{k} - 1$, and the corresponding terms can be estimated as before. Combining everything, and noting that integration by parts is performed $2\,\mathrm{k} - 1$ times, an odd number, it holds that

$$\begin{aligned}\int_{\omega^+} \partial^\alpha u \partial^\gamma \eta \partial^{\beta-\gamma} u \, dx \leqslant \quad & -\int_{\omega^+} \partial^{\beta-\gamma} u \partial^\gamma \eta \partial^\alpha u \, dx \\ & + \|u\|^2_{\mathrm{H}^{k-1}(\omega^+)} \sum_{\substack{\delta_1, \delta_2 \in \mathbb{N}_0^d \\ \delta_1 \leqslant \alpha \\ \delta_2 \leqslant \beta}} \|\partial^\delta \eta\|_{C_0^\infty(\omega^+)},\end{aligned} \tag{5.26}$$

or

$$\int_{\omega^+} \partial^\alpha u \partial^\gamma \eta \partial^{\beta-\gamma} u \, dx \leqslant \frac{1}{2} \|u\|^2_{\mathrm{H}^{k-1}(\omega^+)} \sum_{\substack{\delta \in \mathbb{N}_0^d \\ |\delta| \leqslant |\gamma| + 2\,\mathrm{k} - 1}} \|\partial^\delta \eta\|_{C_0^\infty(\omega^+)} \tag{5.27}$$

The proof is finished by noting that uniform ellipticity of $\mathcal{L}$ implies

$$\sum_{\substack{\alpha, \beta \in \mathbb{N}_0^d \\ |\alpha| \leqslant \mathrm{k} \\ |\beta| \leqslant \mathrm{k}}} a_{\alpha,\beta} \partial^\alpha u \partial^\beta u \geqslant 0 \quad \text{almost everywhere in } \Omega, \tag{5.28}$$

so

$$\begin{aligned}
\|u\|^2_{\mathcal{E}(\omega)} &= \int_\omega \sum_{\substack{\alpha,\beta\in\mathbb{N}_0^d \\ |\alpha|\leqslant \mathrm{k} \\ |\beta|\leqslant \mathrm{k}}} a_{\alpha,\beta}\partial^\alpha u \partial^\beta u \, dx \\
&= \int_\omega \sum_{\substack{\alpha,\beta\in\mathbb{N}_0^d \\ |\alpha|\leqslant \mathrm{k} \\ |\beta|\leqslant \mathrm{k}}} a_{\alpha,\beta}\partial^\alpha u \partial^\beta u\eta \, dx \\
&\leqslant \int_{\omega^+} \sum_{\substack{\alpha,\beta\in\mathbb{N}_0^d \\ |\alpha|\leqslant \mathrm{k} \\ |\beta|\leqslant \mathrm{k}}} a_{\alpha,\beta}\partial^\alpha u \partial^\beta u\eta \, dx \qquad (5.29) \\
&\leqslant \left(C(\eta) \sum_{\substack{\alpha,\beta\in\mathbb{N}_0^d \\ |\alpha|\leqslant \mathrm{k} \\ |\beta|\leqslant \mathrm{k}}} \|a_{\alpha,\beta}\|_{L^\infty(\omega^+)} \right) \|u\|^2_{\mathrm{H}^{\mathrm{k}-1}(\omega^+)},
\end{aligned}$$

where in the last inequality all previously described estimates were used, and $C(\eta)$ contains norms of partial derivatives of η. □

It will turn out, that the functions spanning the desired optimal approximation space on ω are in fact restrictions of weakly harmonic functions on ω^+. Theorem 5.2 shows a crucial property of the corresponding restriction operator.

Theorem 5.2 (Compactness of the restriction operator)**.** *Consider the restriction operator* $\mathrm{P} : \mathcal{H}(\omega^+) \hookrightarrow \mathcal{H}(\omega)$*, defined by* $\mathrm{P}\,u(x) = u(x)$ *for all* $x \in \omega$*. Under the assumptions from Theorem 5.1,* P *is compact.*

◇

Proof. The compactness of P is shown using sequences. If $\{u_i\}_{i=1}^\infty \subset \mathcal{H}(\omega^+)$ is a bounded sequence, it needs to be shown that $\{\mathrm{P}\,u_i\}_{i=1}^\infty \subset \mathcal{H}(\omega)$ has a convergent subsequence. It follows from the Rellich-Kondrachov theorem, that the embedding $\iota^{\mathrm{k}}_{\mathrm{k}-1} : \mathrm{H}^{\mathrm{k}}(\omega^+) \hookrightarrow \mathrm{H}^{\mathrm{k}-1}(\omega^+)$, $u \mapsto u$ is compact. Hence, the sequence $\{\iota^{\mathrm{k}}_{\mathrm{k}-1}\,u_i\}_{i=1}^\infty = \{u_i\}_{i=1}^\infty$ has a convergent subsequence in $\mathrm{H}^{\mathrm{k}-1}(\omega^+)$, named $\{u_{i_j}\}_{j=1}^\infty$. The Caccioppoli-type inequality from Theorem 5.1 holds, and can be applied to $\mathrm{P}\,u_{i_j} - \mathrm{P}\,u_{i_k} \in \mathcal{H}(\omega)$, reading

$$\|\,\mathrm{P}\,u_{i_j} - \mathrm{P}\,u_{i_k}\|_{\mathcal{E}(\omega)} \leqslant C\|u_{i_j} - u_{i_k}\|_{\mathrm{H}^{\mathrm{k}-1}(\omega^+)}, \quad \forall j,k \in \mathbb{N}. \tag{5.30}$$

The subsequence $\{u_{i_j}\}_{j=1}^\infty$ at hand converges in $\mathrm{H}^{\mathrm{k}-1}(\omega^+)$, so it is a Cauchy sequence, meaning that for any $\varepsilon > 0$ there exists $r \in \mathbb{N}$ such that

$$\|u_{i_j} - u_{i_k}\|_{\mathrm{H}^{\mathrm{k}-1}(\omega^+)} < \varepsilon, \quad \forall j,k \geqslant r. \tag{5.31}$$

From (5.30) and (5.31) it follows that

$$\|\,\mathrm{P}\,u_{i_j} - \mathrm{P}\,u_{i_k}\|_{\mathcal{E}(\omega)} \leqslant C\|u_{i_j} - u_{i_k}\|_{\mathrm{H}^{\mathrm{k}-1}(\omega^+)} < C\varepsilon, \quad \forall j,k,\geqslant r, \tag{5.32}$$

showing that also $\{\mathrm{P}\, u_{i_j}\}_{j=1}^{\infty}$ is a Cauchy sequence in $\mathcal{H}(\omega)$, thus convergent. □

Using the restriction operator P, n-dimensional approximation spaces $\mathcal{V}_{\mathrm{n}}(\omega) \subset \mathcal{H}(\omega)$, for $n \in \mathbb{N}$, will be constructed in the following. Their approximation quality is quantified in terms of the worst-case best approximation error,

$$\sup_{u \in \mathcal{H}(\omega^+)} \inf_{v \in \mathcal{V}_{\mathrm{n}}(\omega)} \frac{\|\mathrm{P}\, u - v\|_{\mathcal{E}(\omega)}}{\|u\|_{\mathcal{E}(\omega^+)}}. \tag{5.33}$$

Consequently, the best n dimensional approximation space, $\mathcal{V}_{\mathrm{n}}^{\mathrm{opt}}(\omega)$, is defined by

$$\mathcal{V}_{\mathrm{n}}^{\mathrm{opt}}(\omega) := \operatorname*{arginf}_{\substack{\mathcal{V}_{\mathrm{n}}(\omega) \subset \mathcal{H}(\omega) \\ \dim \mathcal{V}_{\mathrm{n}}(\omega) = n}} \sup_{u \in \mathcal{H}(\omega^+)} \inf_{v \in \mathcal{V}_{\mathrm{n}}(\omega)} \frac{\|\mathrm{P}\, u - v\|_{\mathcal{E}(\omega)}}{\|u\|_{\mathcal{E}(\omega^+)}} \tag{5.34}$$

and the corresponding worst-case best approximation error is denoted

$$\begin{aligned} d_n(\omega, \omega^+) := & \inf_{\substack{\mathcal{V}_{\mathrm{n}}(\omega) \subset \mathcal{H}(\omega) \\ \dim \mathcal{V}_{\mathrm{n}}(\omega) = n}} \sup_{u \in \mathcal{H}(\omega^+)} \inf_{v \in \mathcal{V}_{\mathrm{n}}(\omega)} \frac{\|\mathrm{P}\, u - v\|_{\mathcal{E}(\omega)}}{\|u\|_{\mathcal{E}(\omega^+)}} \\ = & \sup_{u \in \mathcal{H}(\omega^+)} \inf_{v \in \mathcal{V}_{\mathrm{n}}^{\mathrm{opt}}(\omega)} \frac{\|\mathrm{P}\, u - v\|_{\mathcal{E}(\omega)}}{\|u\|_{\mathcal{E}(\omega^+)}}. \end{aligned} \tag{5.35}$$

The quantity d_n is called *Kolmogorov n-width* of the compact operator P.

The space $\mathcal{V}_{\mathrm{n}}^{\mathrm{opt}}(\omega)$ can now be constructed explicitly. Consider the adjoint of the restriction operator, $\mathrm{P}^\star : \mathcal{H}(\omega) \to \mathcal{H}(\omega^+)$. The operator $\mathrm{P}^\star$ is linear and continuous. Under the assumptions from Theorem 5.1, P is compact, so for any bounded sequence $\{u_i\}_{i=1}^{\infty} \subset \mathcal{H}(\omega^+)$ there exists a convergent subsequence $\{\mathrm{P}\, u_{i_j}\}_{j=1}^{\infty} \subset \mathcal{H}(\omega)$. From the continuity of $\mathrm{P}^\star$, it follows that

$$\|\mathrm{P}^\star(\mathrm{P}\, u_{i_j} - \mathrm{P}\, u_{i_k})\|_{\mathcal{H}(\omega^+)} \leqslant \|\mathrm{P}^\star\|_{\mathrm{op}} \|P u_{i_j} - P u_{i_k}\|_{\mathcal{H}(\omega^+)}, \quad \forall j, k \in \mathbb{N}. \tag{5.36}$$

with the operator norm

$$\|\mathrm{P}^\star\|_{\mathrm{op}} := \sup_{u \in \mathcal{H}(\omega)} \frac{\|\mathrm{P}^\star u\|_{\mathrm{H}^{\mathrm{k}}(\omega^+)}}{\|u\|_{\mathrm{H}^{\mathrm{k}}(\omega)}}, \tag{5.37}$$

showing that $\{\mathrm{P}^\star\, \mathrm{P}\, u_{i_j}\}_{j=1}^{\infty} \subset \mathcal{H}(\omega^+)$ converges in $\mathcal{H}(\omega^+)$. This means, that also the operator $\mathrm{P}^\star\, \mathrm{P}$ is compact. Moreover, it is self-adjoint and non-negative. In the books of [Gar62, Sin13, Pin85], it was shown that the Kolmogorov n-width is directly related to (and computable using) the eigenvalues of the eigenvalue problem $\mathrm{P}^\star\, \mathrm{P}\, u = \lambda u$. This is presented in Theorem 5.3.

Theorem 5.3 (n-width and eigenvalue problem)**.** *The n-widths of the compact, self-adjoint, non-negative operator $\mathrm{P}^\star\, \mathrm{P}$ and the corresponding optimal n-dimensional approximation spaces $\mathcal{V}_{\mathrm{n}}^{opt}(\omega)$ satisfy*

$$\begin{aligned} d_n(\omega, \omega^+) = & \sqrt{\lambda_{n+1}} \\ \mathcal{V}_{\mathrm{n}}^{opt}(\omega) = & \operatorname{span}\{\psi_1, \ldots, \ldots, \psi_n\} \end{aligned} \tag{5.38}$$

for all $n \in \mathbb{N}$, where $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant 0$ are the ordered eigenvalues of $\mathrm{P}^\star\, \mathrm{P}$ and ψ_i is the eigenfunction corresponding to λ_i for all i.

◇

Proof. See [Sin13, Chapter IV, Theorem 2.2]. □

Using the definition of the adjoint operator and the local energy inner products, the eigenvalue problem can be rewritten. While the operator $\mathrm{P}^\star\,\mathrm{P}$ is rather abstract, this reformulation can be used to explicitly compute the optimal approximation space $\mathcal{V}_{\mathrm{n}}^{\mathrm{opt}}(\omega)$.

Theorem 5.4 (n-width and gEVP)**.** *The n-widths of the compact, self-adjoint, non-negative operator $\mathrm{P}^\star\,\mathrm{P}$ and the corresponding optimal n-dimensional approximation spaces $\mathcal{V}_{\mathrm{n}}^{opt}(\omega)$ satisfy*

$$\begin{aligned} d_n(\omega,\omega^+) &= \sqrt{\lambda_{n+1}} \\ \mathcal{V}_{\mathrm{n}}^{opt}(\omega) &= \operatorname{span}\{\mathrm{P}\,\psi_1,\ldots,\ldots,\mathrm{P}\,\psi_n\}, \end{aligned} \tag{5.39}$$

where $1 \geqslant \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant 0$ and $\mathcal{H}(\omega^+) \ni \psi_1, \psi_2, \ldots$ are the eigenvalues and eigenfunctions of the generalized eigenvalue problem

$$\langle \psi, v\rangle_{\mathcal{E}(\omega)} = \lambda \langle \psi, v\rangle_{\mathcal{E}(\omega^+)}, \quad \forall v \in \mathcal{H}(\omega^+). \tag{5.40}$$

◇

Proof. From Theorem 5.3 it is known that the eigenpairs of the eigenvalue problem are the solutions of

$$\mathrm{P}^\star\,\mathrm{P}\,\psi = \lambda\,\psi\,. \tag{5.41}$$

Given a solution (λ, ψ) of (5.41) it also holds that

$$\langle \mathrm{P}^\star\,\mathrm{P}\,\psi, v\rangle_{\mathcal{E}(\omega^+)} = \lambda \langle \psi, v\rangle_{\mathcal{E}(\omega)}, \quad \forall v \in \mathcal{H}(\omega^+). \tag{5.42}$$

Using the definition of the adjoint operator, and recalling that P is the restriction, it follows

$$\langle \mathrm{P}\,\psi, \mathrm{P}\,v\rangle_{\mathcal{E}(\omega^+)} = \langle \psi, v\rangle_{\mathcal{E}(\omega)}, \quad \forall v \in \mathcal{H}(\omega^+), \tag{5.43}$$

and the generalized eigenvalue problem reads

$$\langle \psi, v\rangle_{\mathcal{E}(\omega)} = \lambda \langle \psi, v\rangle_{\mathcal{E}(\omega)}, \quad \forall v \in \mathcal{H}(\omega^+). \tag{5.44}$$

Due to Theorem 5.3, the eigenvalues encode information on the n-widths, whereas the restrictions of the eigenfunctions span the space $\mathcal{V}_{\mathrm{n}}^{\mathrm{opt}}(\omega)$. □

5.2 Practical construction

The framework for the computation of optimal local approximation spaces, which was presented in the previous Section 5.1.2, will be extensively used in the experimental part of this thesis. This section describes practical details of the computational process using the Partition of Unity Method. Consider Problem 1 and suppose the differential operator is uniformly elliptic, scalar and of even

order $2\,\mathrm{k}$. Suppose further that $\omega \subset \Omega$ is a subdomain which will be used as a patch in the Partition of Unity Method and on which the local approximation quality of the global space is to be enhanced, for example due to the presence of a complicating feature. According to the lifting of solutions described in Section 5.1.1, the solution u_ω of the corresponding local Problem 10 can be decomposed as

$$u_\omega = u_{\omega,f} + u_{\omega,g^0} + \ldots + u_{\omega,g^{2\,\mathrm{k}-1}} + u_{\omega,\mathrm{hom}}, \tag{5.45}$$

with the function $u_{\omega,\mathrm{hom}}$ being a unique function satisfying the local homogeneous problem and the particular solutions $u_{\omega,f}, u_{\omega,g^0}, \ldots, u_{\omega,g^{2\,\mathrm{k}-1}}$, which can be computed in a straight-forward fashion by solving the corresponding local problems. In the following, all functions appearing in (5.45) refer to variational solutions. As described before, it may be necessary to impose artificial boundary conditions in order to fix rigid modes and ensure well-posedness of the local problems.

The computation of $u_{\omega,\mathrm{hom}}$ on a local level is a very difficult task for many reasons, so the framework from Section 5.1.2 is employed, providing an optimal local approximation space $\mathcal{V}_\mathrm{n}^\mathrm{opt}(\omega)$ to approximate $u_{\omega,\mathrm{hom}}$, with dimension $n \in \mathbb{N}$ to be decided upon.

Let now $\mathcal{V}^\mathrm{PU}(\Omega)$ be the global approximation space of a Partition of Unity Method, which uses ω as one of the patches in the underlying cover. The local approximation space on ω appearing in $\mathcal{V}^\mathrm{PU}(\Omega)$ has the general form

$$\mathcal{V}(\omega) = \mathcal{P}^p(\omega) + \mathcal{E}(\omega) \tag{5.46}$$

for some polynomial degree p. Using the splitting (5.45) and $\mathcal{V}_\mathrm{n}^\mathrm{opt}(\omega)$, a space of enrichments on ω is given by

$$\mathcal{E}(\omega) = \mathrm{span}\{u_{\omega,f}, u_{\omega,g^0}, \ldots, u_{\omega,g^{2\,\mathrm{k}-1}}\} \cup \mathcal{V}_\mathrm{n}^\mathrm{opt}(\omega). \tag{5.47}$$

In order to construct $\mathcal{V}_\mathrm{n}^\mathrm{opt}(\omega)$ explicitly, an oversampled patch ω^+ containing ω is chosen and the generalized eigenvalue problem from Theorem 5.4, that is

$$\langle \psi, v \rangle_{\mathcal{E}(\omega)} = \lambda \langle \psi, v \rangle_{\mathcal{E}(\omega^+)}, \quad \forall v \in \mathcal{H}(\omega^+) \tag{5.48}$$

is solved. As mentioned before, an explicit basis of the infinite-dimensional space of harmonic functions $\mathcal{H}(\omega^+)$ is only known for very few differential operators and must in general be sampled. This section describes practical details of the sampling process and the construction of $\mathcal{V}_\mathrm{n}^\mathrm{opt}(\omega)$.

5.2.1 Choice of the oversampling factor

The construction of optimal shape functions which are to be used on a patch ω are constructed on an oversampled version ω^+ of this patch and afterwards restricted to ω. A priori, the shapes of ω and ω^+ are arbitrary, and the only necessary geometrical condition is $\omega \subset \omega^+$. For practical applications, ω is chosen to be a hypercube and ω^+ is obtained by stretching all sides of ω. Without loss of generality, this oversampling factor $\tau > 1$ is chosen identical in all coordinate directions.

Remark 5.6. *Depending on the problem at hand, it can make sense to use differently shaped subdomains such as spheres. The shape of the patch can also have a strong influence on reusability of the optimal basis functions, as will be seen in Section 5.3.*

*

Suppose now, that $(\lambda, \psi) \in \mathbb{R} \times \mathcal{H}(\omega^+)$ is an eigenpair of the infinite-dimensional eigenvalue problem (5.48). Since $\langle \cdot, \cdot \rangle_{\mathcal{E}(\omega)}$ and $\langle \cdot, \cdot \rangle_{\mathcal{E}(\omega^+)}$ are scalar products on $\mathcal{H}(\omega)$, resp $\mathcal{H}(\omega^+)$, the choice $v = \psi$ leads to the equation

$$\underbrace{\langle \psi, \psi \rangle_{\mathcal{E}(\omega)}}_{\| \psi \|^2_{\mathcal{E}(\omega)}} = \lambda \underbrace{\langle \psi, \psi \rangle_{\mathcal{E}(\omega^+)}}_{\| \psi \|^2_{\mathcal{E}(\omega^+)}}, \quad \forall v \in \mathcal{H}(\omega^+), \tag{5.49}$$

and hence $\lambda \in (0, 1)$ for an oversampling factor $\tau > 1$. Moreover, supposing that ω is an interior patch with side lengths $h_1, \dots, h_d$, i.e. $\operatorname{meas} \omega = \prod_{i=1}^d h_i$, it holds that $\operatorname{meas} \omega^+ \leqslant \prod_{i=1}^d \tau\, h_i$, and equality holds for $\omega^+ \subset \Omega$. As a result,

$$\operatorname{meas}(\omega^+ \setminus \omega) \leqslant \prod_{i=1}^d \tau\, h_i - \prod_{i=1}^d h_i = (\tau^d - 1) \prod_{i=1}^d h_i, \tag{5.50}$$

and for $\tau \approx 1$ also $\omega^+ \approx \omega$, implying that the energy norm of a function on ω will be arbitrarily close to the energy norm of the function on ω^+. Hence, (5.49) implies $\lambda \approx 1$. Following similar lines, choosing a very large oversampling factor tends to lead to very small eigenvalues, since the fraction of a functions energy on ω eventually gets negligibly small compared to its energy on ω^+, leading to eigenvalues $\lambda \approx 0$.

Solving the sampling problem means propagating the boundary data prescribed on the free boundary according to the PDE coefficients towards the center of the oversampled patch ω^+. Oversampling factor and boundary data hence influence each other and must be balanced. In the upcoming numerical experiments (Chapters 6 and 7), the oversampling factor $\tau = 2.0$ worked well, and this is also the value chosen in the original publication [BL11]. It was slightly shrunk if this led to increases of $\operatorname{meas}(\Gamma^{\mathrm{F}}_{\omega^+})$ or to avoid interference from other subregions of interest.

5.2.2 Sampling of harmonic functions

As introduced in Section 5.1.1, denote by $\Gamma^{\mathrm{F}}_{\omega^+}$ the free part of the boundary of ω^+ on which no global boundary condition holds, and recall that weakly local harmonic functions can be sampled by solving the variational formulation of Problem 11. In the following, it is assumed that the sampling problem is well-posed for any provided nonzero boundary value on the free boundary. Let now $\{b^i\}_{i=1}^m \subset \mathrm{L}^2(\Gamma^{\mathrm{F}}_{\omega^+})$, and denote by $\mathcal{D}(\Gamma^{\mathrm{F}}_{\omega^+}, \{b^i\}_{i=1}^m)$ the space spanned by these boundary data functions,

$$\mathcal{D}(\Gamma^{\mathrm{F}}_{\omega^+}, \{b^i\}_{i=1}^m) \coloneqq \operatorname{span}\{b^1, \dots, b^m\} \subset \mathrm{L}^2(\Gamma^{\mathrm{F}}_{\omega^+}). \tag{5.51}$$

Furthermore, denote by u_b the solution of the sampling Problem 11 with additional boundary data b prescribed on $\Gamma^{\mathrm{F}}_{\omega^+}$. For any $b_1, b_2 \in \mathcal{D}(\Gamma^{\mathrm{F}}_{\omega^+}, \{b^i\}_{i=1}^m)$, it follows that also $b_1 + b_2 \in \mathcal{D}(\Gamma^{\mathrm{F}}_{\omega^+}, \{b^i\}_{i=1}^{\mathrm{m}})$. Moreover, for all $\varphi \in \mathcal{C}_0^\infty(\omega^+)$ it holds that

$$\begin{aligned} \mathbf{a}_{\omega^+}[u_{b_1} + u_{b_2}, v] = \;& \mathbf{a}_{\omega^+}[u_{b_1}, v] + \mathbf{a}_{\omega^+}[u_{b_2}, v] = 0, \\ (u_{b_1} + u_{b_2})_{|\,\Gamma^{\mathrm{F}}} = \;& u_{b_1\,|\,\Gamma^{\mathrm{F}}} + u_{b_2\,|\,\Gamma^{\mathrm{F}}} = b_1 + b_2. \end{aligned} \tag{5.52}$$

Hence,

$$u_{b_1+b_2} = u_{b_1} + u_{b_2}, \quad \forall b_1, b_2 \in \mathcal{D}(\omega^+), \tag{5.53}$$

showing that the space $\mathcal{H}_{\mathrm{m}}(\omega^+, \{b^i\}_{i=1}^m) := \operatorname{span}\{u_{b^1}, \ldots, u_{b^m}\}$ spanned by the solutions satisfies

$$\dim \mathcal{H}_{\mathrm{m}}(\omega^+, \{b^i\}_{i=1}^m) = \dim \mathcal{D}(\Gamma^{\mathrm{F}}_{\omega^+}, \{b^i\}_{i=1}^m). \tag{5.54}$$

The choice of boundary data functions for the sampling is essential to the quality of the finite dimensional harmonic space $\mathcal{H}_{\mathrm{m}}$, and various ways of sampling will be described in Section 5.2.5.

Remark 5.7. *The sampling of a finite-dimensional subspace of the infinite-dimensional, abstract harmonic space is the first source of errors in any practical computation of the optimal local basis functions.*

*

Note that the sampling of harmonic functions is the most expensive part of the overall computation. The reason for this is that a partial differential equation has to be solved and that the solution of this PDE should be able to capture the fine-scale behavior of the homogeneous solution, $u_{0_{|\omega^+}}$, meaning that the sampling problem must be resolved finely enough to capture it. In practice, this means using a large number of very fine patches in the Partition of Unity Method.

5.2.3 Setup and solution of the generalized eigenvalue problem

After a set of boundary data functions $\{b^i\}_{i=1}^m \subset \mathrm{L}^2(\Gamma^{\mathrm{F}})$ has been chosen and a finite-dimensional subspace of the infinite-dimensional space of weakly local harmonic functions has been sampled,

$$\mathcal{H}_{\mathrm{m}}(\omega^+, \{b^i\}_{i=1}^m) \subset \mathcal{H}(\omega^+), \tag{5.55}$$

the discrete generalized eigenvalue problem

$$\langle \psi, v \rangle_{\mathcal{E}(\omega)} = \lambda \langle \psi, v \rangle_{\mathcal{E}(\omega^+)}, \quad \forall v \in \mathcal{H}_{\mathrm{m}}(\omega^+, \{b^i\}_{i=1}^m) \tag{5.56}$$

can be set up. Since $\{u_{b^i}\}_{i=1}^m$ is a generating set of $\mathcal{H}_{\mathrm{m}}(\omega^+, \{b^i\}_{i=1}^m)$, there are m eigenpairs (λ_i, ψ_i), $i = 1, \ldots, m$, with

$$\psi_i = \sum_{j=1}^m (\boldsymbol{\psi}_i)_j u_{b^j}, \tag{5.57}$$

and $\boldsymbol{\psi}_i$ is the coefficient vector solving the generalized matrix eigenvalue problem

$$M_\omega \, \boldsymbol{\psi}_i = \lambda_i M_{\omega^+} \, \boldsymbol{\psi}_i \tag{5.58}$$

with matrices

$$\begin{aligned} M_\omega &= \begin{bmatrix} \langle u_{b^1}, u_{b^1} \rangle_{\mathcal{E}(\omega)} & \ldots & \langle u_{b^1}, u_{b^m} \rangle_{\mathcal{E}(\omega)} \\ \vdots & \ddots & \vdots \\ \langle u_{b^m}, u_{b^1} \rangle_{\mathcal{E}(\omega)} & \ldots & \langle u_{b^m}, u_{b^m} \rangle_{\mathcal{E}(\omega)} \end{bmatrix} \\ M_{\omega^+} &= \begin{bmatrix} \langle u_{b^1}, u_{b^1} \rangle_{\mathcal{E}(\omega^+)} & \ldots & \langle u_{b^1}, u_{b^m} \rangle_{\mathcal{E}(\omega^+)} \\ \vdots & \ddots & \vdots \\ \langle u_{b^m}, u_{b^1} \rangle_{\mathcal{E}(\omega^+)} & \ldots & \langle u_{b^m}, u_{b^m} \rangle_{\mathcal{E}(\omega^+)} \end{bmatrix}. \end{aligned} \tag{5.59}$$

Remark 5.8. *The solution of the generalized eigenvalue problem is a post-processing step, and can be regarded as a filtering to identify functions from the finite-dimensional harmonic space that provide the most valuable information. If the number of samples m is small, the generating functions $\{u_{b^i}\}_{i=1}^m$ could also directly be used as local enrichment functions in a global computation. Choosing the 'right' type of boundary data is however not an easy task, and the simulations presented in Chapter 6 will reveal that the required number of boundary data functions is usually quite large.*

*

5.2.4 Adding additional boundary data

Suppose now that an additional boundary data function $b^{m+1} \in \mathrm{L}^2(\Gamma^{\mathrm{F}}_{\omega^+})$ is added, and that all boundary data functions $\{b^i\}_{i=1}^{m+1}$ are linearly independent, leading to an $m+1$ dimensional discrete space of weakly harmonic functions, $\mathcal{H}_{m+1}(\omega^+, \{b^i\}_{i=1}^{m+1})$. It is not clear a priori, how the eigenpairs obtained from the m and the $m+1$ dimensional approximation spaces are related, even if the spaces themselves are nested. In order to see this, note that the extended $(m+1)\times(m+1)$ matrices appearing in the $(m+1)$-dimensional generalized eigenvalue problem contain the $m\times m$ matrices appearing in the m-dimensional generalized eigenvalue problem,

$$\begin{aligned} M_{\omega}^{m+1} &= \left[\begin{array}{c|c} M_{\omega}^{m} & d_{\omega} \\ \hline d_{\omega}^{T} & c_{\omega} \end{array}\right] \\ M_{\omega^+}^{m+1} &= \left[\begin{array}{c|c} M_{\omega^+}^{m} & d_{\omega^+} \\ \hline d_{\omega^+}^{T} & c_{\omega^+} \end{array}\right] \end{aligned} \tag{5.60}$$

with the following vectors from $\mathbb{R}^m$,

$$d_{\omega} = \begin{bmatrix} \langle u_{b^1}, u_{b^{m+1}}\rangle_{\mathcal{E}(\omega)} \\ \vdots \\ \langle u_{b^m}, u_{b^{m+1}}\rangle_{\mathcal{E}(\omega)} \end{bmatrix}, \quad d_{\omega^+} = \begin{bmatrix} \langle u_{b^1}, u_{b^{m+1}}\rangle_{\mathcal{E}(\omega^+)} \\ \vdots \\ \langle u_{b^m}, u_{b^{m+1}}\rangle_{\mathcal{E}(\omega^+)} \end{bmatrix} \tag{5.61}$$

and scalars

$$c_{\omega} = \langle u_{b^{m+1}}, u_{b^{m+1}}\rangle_{\mathcal{E}(\omega)}, \quad c_{\omega^+} = \langle u_{b^{m+1}}, u_{b^{m+1}}\rangle_{\mathcal{E}(\omega^+)}. \tag{5.62}$$

Suppose that $(\lambda, \boldsymbol{\psi}^m)$ is an eigenpair of the m dimensional generalized eigenvalue problem. Let now $T : \mathbb{R}^m \to \mathbb{R}^m$ be a linear transformation, $e \in \mathbb{R}$ and consider the vector

$$\boldsymbol{\psi}^{m+1} := \begin{bmatrix} T\,\boldsymbol{\psi}^m \\ e \end{bmatrix} \in \mathbb{R}^{m+1}. \tag{5.63}$$

The vector $\boldsymbol{\psi}^{m+1}$ is an eigenvector of the $m+1$ dimensional gEVP corresponding to an eigenvalue μ, whenever

$$M_{\omega}^{m+1}\,\boldsymbol{\psi}^{m+1} = \mu M_{\omega^+}^{m+1}\,\boldsymbol{\psi}^{m+1}, \tag{5.64}$$

i.e.

$$\left[\begin{array}{c|c} M_{\omega}^{m} & d_{\omega} \\ \hline d_{\omega}^{T} & c_{\omega} \end{array}\right] \begin{bmatrix} T\,\boldsymbol{\psi}^m \\ e \end{bmatrix} = \mu \left[\begin{array}{c|c} M_{\omega^+}^{m} & d_{\omega^+} \\ \hline d_{\omega^+}^{T} & c_{\omega^+} \end{array}\right] \begin{bmatrix} T\,\boldsymbol{\psi}^m \\ e \end{bmatrix}. \tag{5.65}$$

For orthonormal $\{u_{b^i}\}_{i=1}^{m+1}$ in the energy scalar product over ω^+, it holds that

$$M_{\omega^+}^m = \mathbb{I}_m \in \mathbb{R}^{m\times m}, \quad M_{\omega^+}^m = \mathbb{I}_{m+1} \in \mathbb{R}^{(m+1)\times(m+1)}, \tag{5.66}$$

meaning that d_{ω^+} is the zero vector and $c_{\omega^+} = 1$. Consequently, (5.65) simplifies to

$$\begin{aligned} \begin{bmatrix} M_\omega^m T\,\boldsymbol{\psi}^m + e d_\omega \\ d_\omega T\,\boldsymbol{\psi}^m + e c_\omega \end{bmatrix} &= \mu \begin{bmatrix} T\,\boldsymbol{\psi}^m \\ e \end{bmatrix} \\ &\Leftrightarrow \\ e d_\omega &= (\mu-\lambda) T\,\boldsymbol{\psi}^m \\ d_\omega^T T\,\boldsymbol{\psi}^m &= (\mu - c_\omega) e. \end{aligned} \tag{5.67}$$

Case: Optimal shape function remains unchanged
Suppose that the optimal shape function remains unchanged, i.e. consider the canonical embedding of $\boldsymbol{\psi}^m$ into $\mathbb{R}^{m+1}$. Then, $T = \mathbb{I}_m$, $e = 0$ and the conditions from (5.65 resp. (5.67)) read

$$\begin{aligned} (\mu-\lambda)\,\boldsymbol{\psi}^m &= \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \\ d_\omega^T\,\boldsymbol{\psi}^m &= 0. \end{aligned} \tag{5.68}$$

Equation (5.68) shows that the canonical embedding of $\boldsymbol{\psi}^m$ can only be the eigenvector corresponding to the same eigenvalue as before, $\mu = \lambda$. Furthermore, the orthogonality condition $d_\omega \perp \boldsymbol{\psi}^m$ must hold.

Case: Eigenvalue remains unchanged
Consider now the case that $\mu = \lambda$. The conditions from (5.67) read

$$\begin{aligned} e d_\omega &= \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \\ d_\omega^T T\,\boldsymbol{\psi}^m &= (\lambda - c_\omega) e. \end{aligned} \tag{5.69}$$

The first condition from (5.69) holds if either $e = 0$, or d_ω is the zero vector. In the first case, the second equation from (5.69) implies the orthogonality condition

$$d_\omega \perp T\,\boldsymbol{\psi}^m\,. \tag{5.70}$$

This condition can be used to explicitly compute feasible transformations T.

In the second case, the second equation from (5.69) reads

$$(\lambda - c_\omega) e = 0, \tag{5.71}$$

which is satisfied if either $c_\omega = \lambda$, or $e = 0$ holds.

Conclusion
The foregoing discussion shows that it may under some circumstances be possible for optimal shape functions to remain invariant when increasing the dimension of the discrete space of weakly harmonic functions. In such cases, the corresponding eigenvalue will not change either. On the other hand, if an eigenvalue of the m dimensional generalized eigenvalue problem is also an eigenvalue of the $(m + 1)$ dimensional generalized eigenvalue problem, then the corresponding eigenvectors may be a transformation of each other. In general, it is not possible to estimate whether the largest eigenvalue of the m dimensional problem will still be the dominant eigenvalue of the $m+1$ dimensional problem. In the experiments presented in Chapter 6, it will be observed that the dominant eigenvalues tend to stabilize at certain values for sufficiently large numbers of samples, even when changing the entire set of boundary data functions.

5.2.5 Generation of structured boundary data

Different ways of structurally generating boundary data functions in two spatial dimensions, which can then be used in the sampling Problem 11, are presented in the following. In the original publication [BL11], as well as the follow-up paper [BHL14], the generalized Finite Element Method was employed to numerically compute the optimal local basis functions. The authors chose the nodal finite element basis functions as boundary data in the sampling problem, since every boundary data function that is representable in the FEM approximation space is a linear combination of these nodal basis functions. On the other hand, the number of basis functions in this case is maximal for a fixed grid. In the more recent article [LSS22], oscillating boundary data functions having support on the full free boundary were investigated.

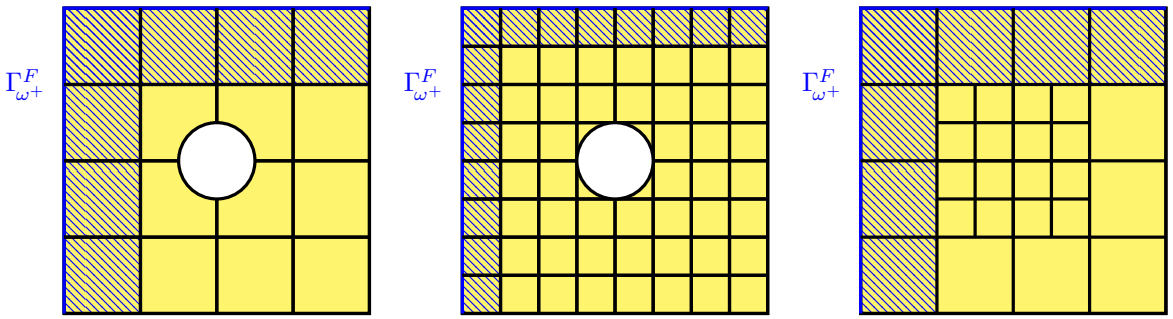
In this section, three variants to generate structured boundary data in the Partition of Unity Method are introduced. The first one generates functions that are defined patchwise and can be local polynomials of any desired order. This approach to construct boundary data with possibly very small support is the PUM equivalent of the approach presented in [BHL14]. Since the sampling problem is expensive to solve, one is interested in solving it for 'relevant' boundary data only. However, since the optimal basis functions are a linear combination of the solutions of the sampling problem, they can by definition only have support on those parts of the free boundary where at least one of the boundary data functions is supported. This motivates the use of boundary data functions having larger support, possibly on the whole free boundary of the domain. Two ways of generating structured boundary data of this type are presented further below, the first one using edgewise B-Splines, and the second one using oscillating trigonometric functions.

Ultimately, the local homogeneous solution $u_{\omega,\text{hom}}$ is to be approximated, and this homogeneous solution lies in the space of harmonic functions, which is infinite-dimensional and will hence be approximated by a finite dimensional subspace. The functions spanning this finite dimensional subspace are solutions to the sampling Problem 11. Different choices of boundary data used in the sampling problem correspond to expanding $u_{\omega,\text{hom}}$ in different bases. Depending on the properties of $u_{\omega,\text{hom}}$, an expansion of the function on the boundary in one of the three presented bases may be favorable.

Note: ω, resp. ω^+, are subdomains of the global domain Ω, on which the original problem of interest was formulated. The sampling problem on the other hand is local, meaning that ω^+ now acts as the full domain for the sampling problem. In order to solve the sampling problem, ω^+ has to be discretized itself. In the enriched global method, all fine patches appearing in the discretization of the local problem must then be resolved for an accurate numerical quadrature.

Boundary hats

In the publication [BHL14], nodal basis functions in a P1-FEM were considered. These are the boundary data functions with smallest possible support, and for the interpolant P1-FEM they are defined by setting the coefficient corresponding to one single boundary node to one, while all other coefficients are set to zero. In contrast to this nodal construction, the Partition of Unity Method uses patches, and a local set of basis functions that are supported throughout the corresponding patch. It is hence straight-forward to define the boundary data functions to be used in the sampling problem patchwise. In the following, suppose that the oversampled patch ω^+ is discretized in such a way that the patches on the boundary stem from a uniform discretization of level $\ell \in \mathbb{N}$, in the following referred to as *boundary level*. Figure 5.2 shows sketches of several exemplary situations. It should be highlighted, that the discretization is not necessarily assumed to be uniform throughout the whole oversampled patch, as can be seen in Figure 5.2 (c): Only the patches touching any part of the free boundary $\Gamma^{\mathrm{F}}_{\omega^+}$ must be constructed on the given boundary level in the initial discretization, and other parts of the domain may for example be the result of adaptive refinement.



(a) Uniform discretization of level 2

(b) Uniform discretization of level 3

(c) Adaptive discretization, level 2 to 3

Figure 5.2: Several possible ways of discretizing an exemplary oversampled patch ω^+. The free boundary consists of the left and top face of ω^+ (blue line), and the boundary patches are hatched in blue. In contrast to the theory of the PUM and for simplicity of the visualization, patches are shown without overlap.

Remark 5.9. *The assumption on the uniform level of boundary patches is only imposed for the formalization of the practical approach presented in this thesis and not due to theoretical restrictions.*

*

The set of boundary patches from a uniform discretization of level ℓ is denoted $\mathbb{P}_\ell(\omega^+)$. On each of the boundary patches, the Partition of Unity method defines a local, polynomial approximation space of order $p \in \mathbb{N}$. The set of scalar, patchwise boundary data functions of polynomial degree q in x_j, for $0 \leqslant q \leqslant p$ and $j = 1, \ldots, d$ defined on level ℓ is denoted $\mathfrak{H}^{q,1}_{\ell,j}$. Combining these sets of functions for all variables j leads to

$$\mathfrak{H}^{q,1}_\ell := \bigcup_{j=1}^{d} \mathfrak{H}^{q,1}_{\ell,j}. \tag{5.72}$$

If vector-valued boundary data in dimension d is desired, define

$$\mathfrak{H}^{q,d}_\ell := \bigcup_{i=1}^{d} \left\{ h e_i \middle| \, h \in \mathfrak{H}^{q,1}_\ell \right\} \tag{5.73}$$

where e_i denotes the i-th unit vector in d dimensions. In general, the problem at hand clearly defines whether scalar or vector-valued boundary data is needed. In order to cover both cases conjointly, $\mathfrak{H}^q_\ell$ is introduced as

$$\mathfrak{H}^q_\ell := \begin{cases} \mathfrak{H}^{q,1}_\ell, & \text{in the scalar case} \\ \mathfrak{H}^{q,d}_\ell, & \text{in the vector-valued case.} \end{cases} \tag{5.74}$$

Regardless of the polynomial degree q or the dimension of the data, the functions from $\mathfrak{H}^q_\ell$ will be referred to as *boundary hats*, or *boundary hats of degree* q. It makes sense to investigate the impact of the polynomial degree q, as well as of the boundary level ℓ on the quality of the discrete space of harmonic functions.

Note that the set of boundary patches on a coarser level, $\mathbb{P}_{\tilde{\ell}}(\omega^+)$, for $0 \leqslant \tilde{\ell} \leqslant \ell$, can in general only be approximately reconstructed by grouping various boundary patches of level ℓ. This is due to the fact that the size of the overlap region results from multiplication of the patch size by a stretch factor, and its volume hence becomes smaller for higher levels. In Figure 5.3 it can be seen that a reconstruction of lower-level patches is only possible for a stretch factor of 1.0, i.e. non-overlapping patches. Figure 5.4 sketches the problems arising for a stretch factor bigger than 1.0. In such cases, a uniform refinement of a boundary patch on level ℓ leads to patches on level $\ell+1$ whose union does not cover the full stretched domain of the level ℓ patch. Moreover, the overlap between the smaller patches is covered multiple times. In general, patches should overlap in the Partition of Unity Method, and the constant boundary hats on level ℓ are hence linearly independent of the boundary hats on level $\ell+1$.

Remark 5.10. *Depending on the problem at hand, it may not be necessary to consider polynomials in all coordinate directions. In the case of extruded two-dimensional domains, polynomials acting in the extruded coordinate direction may oftentimes be omitted in practical applications.*

*

Basis-Splines

The next type of boundary data is supported on larger parts of the free boundary, not only on single patches as in the case of the boundary hats. This is motivated by the fact that the solution of the

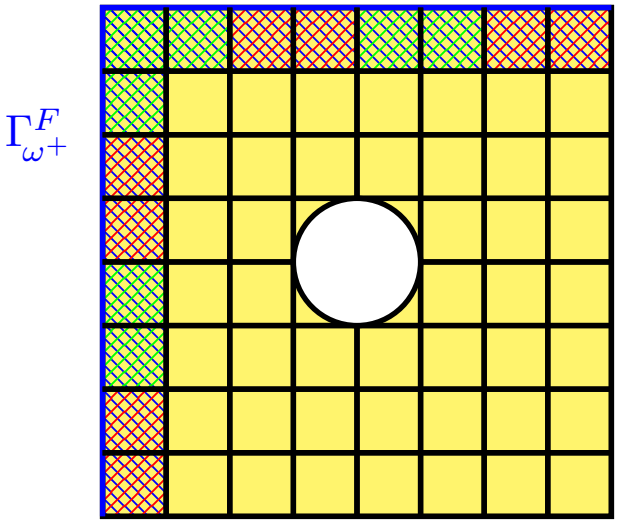
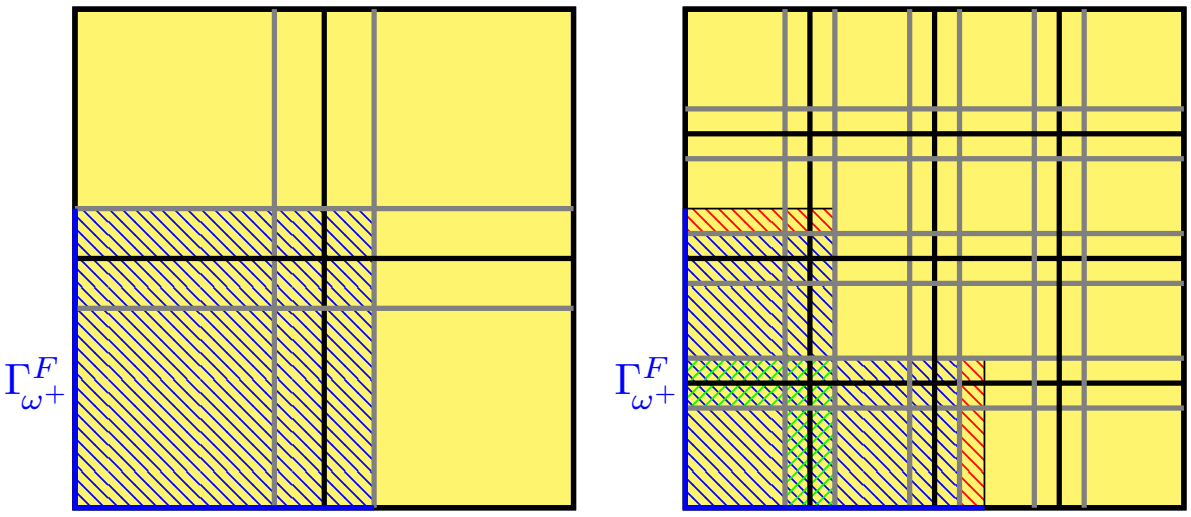


Figure 5.3: Exemplary two-dimensional discretization of ω^+ without overlap. The boundary patches on level $\ell = 3$ are hatched in blue. The reconstructed parts of boundary patches on level $\tilde{\ell} = 2$ are hatched alternating in red and green for better visibility. In this case, reconstruction of lower-level patches is possible.



(a) boundary level $\ell = 1$ (b) boundary level $\ell = 2$

Figure 5.4: Exemplary two-dimensional discretization of ω^+ with stretch factor 1.2. The patches from level $\ell = 1$ (a) are once refined uniformly (b). The coarse boundary patch (blue) is refined to four smaller patches, but only three of them are boundary patches on level 2. The red dashed domain in (b) was part of the level 1 stretched coarse patch, but is not covered by the stretched refined patches on level 2. The green dashed region is covered by two adjacent patches on level 2.

sampling problem is numerically expensive and the number of boundary data functions should therefore be as small as possible. In the following, Basis-Splines (B-Splines) are the functions of choice (cf. [PBP02]). The construction can be performed either for the full free boundary in case it is connected, or only a connected part of it. In order to formalize the ideas, suppose that the

domain is two-dimensional, $d = 2$. Consider a splitting of the free boundary,

$$\Gamma^{\mathrm{F}}_{\omega^+} = \bigcup_{s=1}^{r} \Gamma_s, \tag{5.75}$$

into connected parts $\Gamma_s \subset \Gamma^{\mathrm{F}}_{\omega^+}$, $s = 1, \ldots, r$, which can also overlap, and assume bijections

$$T_i : [0, \operatorname{meas} \Gamma_s] \to \Gamma_s, \quad j = 1, \ldots, r \tag{5.76}$$

are given. However, note that the boundary data is always imposed on the whole free boundary in the sampling problem, and the boundary data functions on Γ are extended by zero for this purpose to the full free boundary. The following construction is repeated for all boundary parts Γ from the splitting and corresponding bijection T, leading to a set of B-Splines in x_j for $j = 1, \ldots, d$, which are of order $k \in \mathbb{N}$ and defined on Γ. An example of a free boundary and a corresponding boundary map is sketched in Figure 5.5. Let $n \in \mathbb{N}$ and consider the *knot sequence*

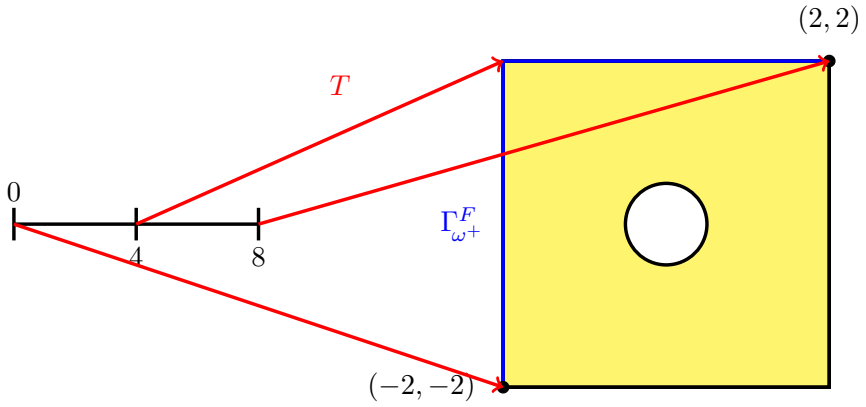


Figure 5.5: Sketch of the boundary map T corresponding to the free boundary marked in blue, which is considered entirely, $\Gamma = \Gamma^{\mathrm{F}}_{\omega^+}$. The length of the free boundary is $\operatorname{meas} \Gamma^{\mathrm{F}}_{\omega^+} = 8$, and T satisfies the following conditions: $T(0) = (-2, -2)^T$, $T(4) = (-2, 2)^T$, $T(8) = (2, 2)^T$.

$$0 \leqslant b_0 \leqslant b_1 \leqslant \ldots \leqslant b_n \leqslant \operatorname{meas} \Gamma. \tag{5.77}$$

The B-splines of order 1 on $[0, \operatorname{meas} \Gamma]$ are defined as

$$B^i_{1,n} : [0, \operatorname{meas} \Gamma] \to \mathbb{R}, \quad B^i_{1,n}(x) := \begin{cases} 1, & \text{if } b_i \leqslant x < b_{i+1} \\ 0, & \text{else} \end{cases} \tag{5.78}$$

for all $i = 0, \ldots, n-1$. B-splines of higher order $p > 1$ are computed using the three term recursion

$$B^i_{p,n}(x) := \frac{x - b_i}{b_{i+p} - b_i} B^i_{p-1,n}(x) + \frac{b_{i+p+1} - x}{b_{i+p+1} - b_{i+1}} B^{i+1}_{p-1,n}(x), \tag{5.79}$$

for $i = 0, \ldots, n-p-1$. The B-Splines of oder p in x_j on Γ are now defined as the concatenation of the B-Splines of order p on $[0, \operatorname{meas} \Gamma]$ with the j-th component of the inverse transformation, $(T^{-1})_j$,

$$\mathfrak{B}^1_{p,n,j}(\Gamma) := \left\{ B^i_{p,n} \circ (T^{-1})_j \,|\, i = 0, \ldots, n-p-1 \right\}. \tag{5.80}$$

Repeating this construction for all variables x_j, $j = 1, \ldots, d$ and combining the sets of B-Splines yields

$$\mathfrak{B}^1_{p,n}(\Gamma) \coloneqq \bigcup_{j=1}^{d} \mathfrak{B}^1_{p,n,j}(\Gamma). \tag{5.81}$$

In case vector-valued B-Splines (d-dimensional) are desired,

$$\mathfrak{B}^d_{p,n}(\Gamma) \coloneqq \left\{ be_i \middle| \, b \in \mathfrak{B}^1_{p,n}(\Gamma),\, i = 1, \ldots, d \right\}, \tag{5.82}$$

with e_i being the i-th unit vector in d dimensions. Scalar boundary data is needed for scalar partial differential equations, and vector-valued boundary data is needed for vector-valued problems. Since no confusion is possible, both cases will be denoted conjointly as

$$\mathfrak{B}_{p,n}(\Gamma) \coloneqq \begin{cases} \mathfrak{B}^1_{p,n}(\Gamma), & \text{in the scalar case} \\ \mathfrak{B}^d_{p,n}(\Gamma), & \text{else.} \end{cases} \tag{5.83}$$

The process is repeated for all connected parts of the boundary appearing in the splitting (5.75), leading to the full set of boundary B-Splines on the free boundary $\Gamma^{\mathrm{F}}_{\omega^+}$,

$$\mathfrak{B}^{\Gamma^{\mathrm{F}}_{\omega^+}}_{p,n} \coloneqq \bigcup_{s=1}^{r} \mathfrak{B}_{p,n}(\Gamma_s). \tag{5.84}$$

The previous construction can be extended to the case of higher-dimensional domains by mapping multivariate B-Splines from a reference domain to a hyperplane Γ being part of the free boundary.

Remark 5.11. *In the case of three-dimensional domains, variation of the boundary data in the third coordinate direction x_3 may not be necessary. In such cases, the B-Splines from the $2d$ approach above can be assumed to be constant in x_3.*

*

Oscillating trigonometric functions

The mapping process described before is not inherent to B-Splines and can be applied to other sets of functions as well. In the following, oscillating trigonometric functions are considered, which appear in Fourier series expansions. These functions oscillate on arbitrarily fine scales in order to compensate for their large support. The functions are constructed in exactly the same way as described above, with the univariate B-Splines on $[0, \operatorname{meas}\Gamma]$ replaced by the sets of functions

$$\left\{ \cos\left(\frac{2\pi i\, \varsigma(\cdot)}{\operatorname{meas}\Gamma} \right) \right\}_{i=1}^{n} \tag{5.85}$$

or

$$\left\{ \sin\left(\frac{2\pi i\, \varsigma(\cdot)}{\operatorname{meas}\Gamma} \right) \right\}_{i=1}^{n}. \tag{5.86}$$

The number of oscillations, i.e. the number of maxima and minima of the functions increase with their index and are specified by $1 \leqslant \varsigma \in \mathbb{N}$. For general $i = 1, \ldots, n$ the number of oscillations is $(i+1)\,\varsigma$. The sets of functions of oscillating boundary data on $\Gamma^{\mathrm{F}}_{\omega^+}$ are denoted $\mathfrak{C}^{\varsigma}_n$ for the cosine functions, resp. $\mathfrak{S}^{\varsigma}_n$ for the sine functions.

General remarks regarding the construction of boundary data

The B-Splines in the experiments from Chapter 6, which consider rectangular and cubic patches, will not be defined on the entire free boundary, but edge-wise. This doubles the number of boundary data functions defined in each coordinate direction, but adds more variability. Moreover, additional knots located slightly outside of the bounds of the edges may be inserted into the knot sequence, leading to additional B-Splines that are referred to as *corner splines* in the experiments. In order to support the boundary data also along vertices, the B-Spline boundary data functions in the experiments are defined as tensor products of B-Splines of an arbitrary order on one edge with the B-Spline of order 1 defined on an adjacent edge of the free boundary, which takes value 1 at the vertex and value 0 towards the interior of the adjacent edge.

5.3 Reusability of optimal basis functions

As seen before, the numerical computation of optimal basis functions is based on the repeated numerical solution of the sampling Problem 11. The restriction of the homogeneous solution of the original partial differential equation under study is contained in the infinite-dimensional space of local weakly harmonic functions, and the solutions of the sampling problem span a discrete, finite-dimensional approximation of this infinite-dimensional space. Since the optimal basis functions are linear combinations of the solutions of the sampling problem, these solutions should be computed as accurate as possible, in turn causing a significant amount of numerical work.

Fortunately, the construction of optimal basis functions is independent of the load and the explicit value of boundary conditions that apply in the problem. Handling changes in the data only requires the recomputation of local particular solutions.

On the other hand, the question arises whether local optimal basis functions computed on one patch may be employed on another patch. This geometric reusability is not clear by construction, and will be investigated in the remainder of this section. The presented theoretical results regarding geometric reusability were developed in cooperation with Prof. Dr. Christian Rieger from the University of Marburg. In Section 5.3.1, general considerations regarding reusability of optimal shape functions are presented. Since it is impossible to state relations for partial differential equations of any given order and structure, more specific and easy-to-check conditions will be developed for the special case of scalar second-order partial differential equations in divergence form in Section 5.3.2. Similar results can also be developed for the vector-valued case.

Note that this section is focussed on geometric reusability, which can be analytically investigated in order to establish algebraic conditions linking the transformations and the coefficients of the partial differential equations. These conditions ensure, that optimal basis functions computed on one patch are still optimal when transformed to another patch. In the case that the underlying differential operator changes, it is in general not possible to find an algebraic relation between optimal shape functions computed for the old and the new operator (except for example if the operator is just scaled by a fixed constant). Since the differential operators under study are assumed to be elliptic, small changes of the operators also lead to small changes of the solutions of the sampling problem, which consequently also lead to relatively small changes of the spectrum of the generalized eigenvalue problems. Straight-forward estimates of the errors are, however, too greedy to be of any practical use, and the similarity between sets of optimal shape functions corresponding to different differential operators will not be investigated in this thesis.

5.3.1 General conditions for geometric reusability

This section presents general considerations on the reusability of optimal shape functions under geometric transformations. In the following, let $\Omega \subset \mathbb{R}^d$ be a d-dimensional domain, and let $\mathcal{L} : [\mathcal{C}^{2\,\mathrm{k}}(\Omega)]^{\mathrm{n}} \to \mathbb{R}^{\mathrm{n}}$ be an elliptic partial differential operator of even order $2\,\mathrm{k}$, for which the optimal shape functions are to be computed. For scalar PDE it holds that $\mathrm{n} = 1$, otherwise $\mathrm{n} = d$. Let $\omega \subset \Omega$ be a patch from the domain, on which optimal shape functions are to be used, and let $\omega^+ \supset \omega$ be an oversampled version of the patch, that is used in their construction. The generalized eigenvalue problem consists of finding eigenpairs $(\lambda, \psi) \in \mathbb{R} \times \mathcal{H}(\omega^+)$ satisfying

$$\langle \psi, v \rangle_{\mathcal{E}(\omega)} = \lambda \langle \psi, v \rangle_{\mathcal{E}(\omega^+)}, \quad \forall v \in \mathcal{H}(\omega^+). \tag{5.87}$$

In order to investigate geometric reusability of the shape functions, consider another patch $\tilde{\omega}$ together with its oversampled version $\tilde{\omega}^+$. For the sake of simplicity, suppose that $\omega^+ \cap \partial\,\Omega = \tilde{\omega}^+ \cap \partial\,\Omega = \varnothing$. The original patches and the new ones are linked via a bijective map

$$T : \mathbb{R}^d \to \mathbb{R}^d, \quad x \mapsto \tilde{x} := Tx = \begin{bmatrix} T_1(x) \\ \vdots \\ T_d(x) \end{bmatrix}, \quad \text{with } T(\omega^+) = \tilde{\omega}^+ . \tag{5.88}$$

A sketch of this situation is shown in Figure 5.6. Similar to (5.87), the eigenpairs $(\tilde{\lambda}, \tilde{\psi})$ of the

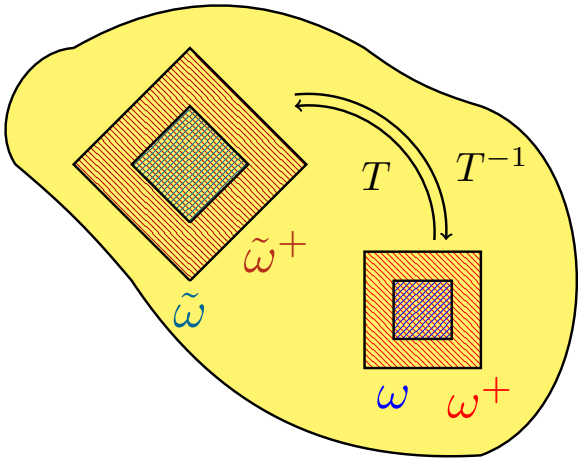


Figure 5.6: Sketch of the geometric relation between ω and $\tilde{\omega}$, resp. ω^+ and $\tilde{\omega}^+$. The patches, which are linked through the transformation T, are supposed to be in the interior of Ω.

generalized eigenvalue problem posed on the new patches $\tilde{\omega}$ and $\tilde{\omega}^+$ satisfy

$$\langle \tilde{\psi}, \tilde{v} \rangle_{\mathcal{E}(\tilde{\omega})} = \tilde{\lambda} \langle \tilde{\psi}, \tilde{v} \rangle_{\mathcal{E}(\tilde{\omega}^+)}, \quad \forall \tilde{v} \in \mathcal{H}(\tilde{\omega}^+). \tag{5.89}$$

In the following, conditions to be imposed on the operator $\mathcal{L}$ and the map T will be identified, such that an eigenpair (λ, ψ) can be mapped to an eigenpair $(\tilde{\lambda}, \tilde{\psi})$. Although one is in general interested in mapping the space spanned by various eigenfunctions on ω^+ to a space spanned by eigenfunctions on $\tilde{\omega}^+$, this is very hard to investigate analytically. Hence, a mapping of the individual basis functions will be analyzed in the following.

By the multidimensional chain rule, one obtains with the new variables $\tilde{x} = T(x)$ for a function $h \in \mathcal{C}^1(\mathbb{R}^d)$ that

$$\frac{\partial h(T(x))}{\partial x_j} = \sum_{i=1}^{d} \frac{\partial h(\tilde{x})}{\partial \tilde{x}_k} \frac{\partial T_k(x)}{\partial_{x_j}}. \tag{5.90}$$

For $T \in [\mathcal{C}^1(\mathbb{R}^d)]^d$ with components $T_1, \dots, T_d$ in the x variables, the Jacobian is defined as

$$J_{T,x}(x) := \begin{bmatrix} \frac{\partial T_1(x)}{\partial x_1} & \cdots & \frac{\partial T_1(x)}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial T_d(x)}{\partial x_1} & \cdots & \frac{\partial T_d(x)}{\partial x_d} \end{bmatrix}. \tag{5.91}$$

In the scope of Linear Elasticity from Section 3.2, the Jacobian of a vector-valued function was denoted similar to the case of scalar functions with the gradient symbol ∇. Since in this section Jacobians and regular gradients appear in the same equations, the notion from (5.91) will be used for better readability.

Using this notation, the chain rule (5.90) reads

$$\nabla_x h(T(x)) = J_{T,x}^T \nabla_{\tilde{x}} h(T(x)), \tag{5.92}$$

or equivalently

$$J_{T,x}^{-T}(x) \nabla_x h(\tilde{x}) = \nabla_{\tilde{x}} h(\tilde{x}). \tag{5.93}$$

A basic property that is needed in order to show reusability of the optimal shape functions is presented in the following.

Lemma 5.5 (Transformation of smooth local functions)**.** *Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a $\mathcal{C}^\infty$ diffeomorphism with $T(\omega^+) = \tilde{\omega}^+$. Then, the map $\mathcal{M}$ with*

$$\tilde{\psi} \mapsto \mathcal{M}(\tilde{\psi}) := \tilde{\psi} \circ T, \quad \forall \tilde{\psi} \in \mathcal{C}_0^\infty(\tilde{\omega}^+) \tag{5.94}$$

has the image $\mathcal{C}_0^\infty(\omega^+)$ and is bijective.

•

Proof. For any $\tilde{\psi} \in \mathcal{C}_0^\infty(\tilde{\omega}^+)$, the function

$$\mathcal{M}(\tilde{\psi}) := \tilde{\psi} \circ T : \omega^+ \to \mathbb{R} \tag{5.95}$$

is in $\mathcal{C}^\infty(\omega^+)$. Moreover, it holds that

$$\operatorname{supp} \psi = T^{-1}(\operatorname{supp} \tilde{\psi}). \tag{5.96}$$

Since $\operatorname{supp} \tilde{\psi} \subset \tilde{\omega}^+$ is compact and T^{-1} is continuous, also $\operatorname{supp} \psi \subset \omega^+$ is compact, so $\mathcal{M}(\mathcal{C}_0^\infty(\tilde{\omega}^+)) \subset \mathcal{C}^\infty(\omega^+)$.

Injectivity: Let $\mathcal{M}(\tilde{\psi}_1) = \psi = \mathcal{M}(\tilde{\psi}_2)$ for some $\psi_1, \psi_2 \in \mathcal{C}_0^\infty(\tilde{\omega}^+)$. This means

$$\tilde{\psi}_1(T(x)) = \tilde{\psi}_2(T(x)), \quad \forall x \in \omega^+ \tag{5.97}$$

and since $T(\omega^+) = \tilde{\omega}^+$, this means that

$$\tilde{\psi}_1(\tilde{x}) = \tilde{\psi}_2(\tilde{x}), \quad \forall \tilde{x} \in \tilde{\omega}^+, \tag{5.98}$$

so $\tilde{\psi}_1 = \tilde{\psi}_2$ on $\tilde{\omega}^+$.

Surjectivity: Let $\psi \in \mathcal{C}_0^\infty(\omega^+)$. By repeating the same argument as before, the function $\tilde{\psi} := \psi \circ T^{-1}$ is in $\mathcal{C}_0^\infty(\tilde{\omega}^+)$.
Therefore, the map $\mathcal{M} : \mathcal{C}_0^\infty(\tilde{\omega}) \to \mathcal{C}_0^\infty(\omega^+)$ is bijective. □

Remark 5.12. *Using Lemma 5.5, a function $\tilde{\psi} \in \mathcal{C}_0^\infty(\tilde{\omega}^+)$ can be identified with the function $(\psi \circ T) \in \mathcal{C}_0^\infty(\tilde{\omega}^+)$, for a unique $\psi \in \mathcal{C}_0^\infty(\omega^+)$. Analogously, any function $\psi \in \mathcal{C}_0^\infty(\omega^+)$ can be identified with the function $(\tilde{\psi} \circ T^{-1}) \in \mathcal{C}_0^\infty(\tilde{\omega}^+)$, for a unique $\tilde{\psi} \in \mathcal{C}_0^\infty(\tilde{\omega}^+)$.*

*

A similar transformation can be derived for H^1 functions.

Lemma 5.6 (Transformation of weakly differentiable local functions)**.** *Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a $\mathcal{C}^\infty$ diffeomorphism with $T(\omega^+) = \tilde{\omega}^+$ and $J_{T,x} = \mathrm{const}$. Then, the map $\mathcal{M}$ with*

$$\tilde{v} \mapsto \mathcal{M}(\tilde{v}) := \tilde{v} \circ T, \quad \forall \tilde{v} \in \mathrm{H}^1(\tilde{\omega}^+) \tag{5.99}$$

has the image $\mathrm{H}^1(\omega^+)$ and is bijective.

•

Proof. Transformation of L^2 functions: Let $\tilde{v} \in L^2(\tilde{\omega}^+)$. Using the transformation formula for integrals,

$$\begin{aligned} \int_{\tilde{\omega}^+} \tilde{v}(\tilde{x})^2 \, d\tilde{x} &= \int_{T(\omega^+)} \tilde{v}(\tilde{x})^2 \, d\tilde{x} \\ &= \int_{\omega^+} \tilde{v}(T(x))^2 \underbrace{|\det J_{T,x}|}_{=\mathrm{const}} \, dx. \end{aligned} \tag{5.100}$$

Hence, $v = \mathcal{M}(\tilde{v}) = (\tilde{v} \circ T) \in L^2(\omega^+)$.

Injectivity of $\mathcal{M}$: Let $\mathcal{M}(\tilde{v}_1) = v = \mathcal{M}(\tilde{v}_2)$ for some $\tilde{v}_1, \tilde{v}_2 \in H^1(\tilde{\omega}^+)$. In the L^2-sense this means $\mathcal{M}(\tilde{v}_1) = \mathcal{M}(\tilde{v}_2)$ almost everywhere, that is

$$\mathcal{M}(\tilde{v}_1)(x) = \mathcal{M}(\tilde{v}_2)(x), \text{ a.e.} \quad \Longleftrightarrow \quad \tilde{v}_1(T(x)) = \tilde{v}_2(T(x)), \text{ a.e.} \tag{5.101}$$

and since T is a diffeomorphism, this means $\tilde{v}_1(\tilde{x}) = \tilde{v}_2(\tilde{x})$ for almost every $\tilde{x} \in \tilde{\omega}^+$.

Surjectivity of $\mathcal{M}$: Let $v \in L^2(\omega^+)$. Then, for any $\eta \in \mathcal{C}_0^\infty(\omega^+)$ it holds that

$$\begin{aligned} \infty \quad &> \quad \int_{\omega^+} v(x)\eta(x) \, dx = \int_{T^{-1}(\omega^+)} v(x)\eta(x) \, dx \\ &= \quad \int_{\tilde{\omega}^+} |\det J_{T^{-1},\tilde{x}}| v(T^{-1}(\tilde{x}))\eta(T^{-1}(\tilde{x})) \, d\tilde{x}. \end{aligned} \tag{5.102}$$

From Lemma 5.5 it follows that $\eta \circ T^{-1}$ can be identified with a unique function $\tilde{\eta} \in \mathcal{C}_0^\infty(\tilde{\omega}^+)$. Since $|\det J_{T^{-1},\tilde{x}}|$ is constant and nonzero, one can choose $\tilde{v} = v \circ T^{-1}$ and obtain that

$$\int_{\tilde{\omega}^+} \tilde{v}(\tilde{x})\tilde{\eta}(\tilde{x})\, d\tilde{x} < \infty, \quad \forall \tilde{\eta} \in \mathcal{C}_0^\infty(\tilde{\omega}^+) \tag{5.103}$$

implying that $\tilde{v} \in L^2(\tilde{\omega}^+)$. Concluding, the map $\mathcal{M}$ is bijective.

Transformation of weak derivatives: Let $\tilde{v} \in H^1(\tilde{\omega}^+)$. Let $\tilde{\eta} \in \mathcal{C}_0^\infty(\tilde{\omega}^+)$ and let v be the unique function $v = \tilde{v} \circ T$. It holds that $\tilde{\eta}' \in \mathcal{C}_0^\infty(\tilde{\omega}^+)$ and one can also identify $\eta' = \tilde{\eta}' \circ T$ (this is the transformation from Lemma 5.5). Next, it is shown that $w := \tilde{v}' \circ T$ is the weak derivative of v and for this, the definition of the weak derivative for $\tilde{v}$ will be used.

$$\begin{aligned}
\int_{\omega^+} w(x)\eta(x)\, dx &= \int_{T^{-1}(\tilde{\omega}^+)} w(x)\eta(x)\, dx \\
&= \int_{\tilde{\omega}^+} w(T^{-1}(\tilde{x}))\eta(T^{-1}(\tilde{x}))|\det J_{T^{-1},\tilde{x}}|\, d\tilde{x} \\
&= \int_{\tilde{\omega}^+} \tilde{v}'(T(T^{-1}(\tilde{x})))\tilde{\eta}(\tilde{x})|\det J_{T^{-1},\tilde{x}}|\, d\tilde{x} \\
&= \int_{\tilde{\omega}^+} \tilde{v}'(\tilde{x})\tilde{\eta}(\tilde{x})|\det J_{T^{-1},\tilde{x}}|\, d\tilde{x} \\
&= -\int_{\tilde{\omega}^+} \tilde{v}(\tilde{x})\tilde{\eta}'(\tilde{x})|\det J_{T^{-1},\tilde{x}}|\, d\tilde{x} \\
&= -\int_{T(\omega^+)} \tilde{v}(\tilde{x})\tilde{\eta}'(\tilde{x})|\det J_{T^{-1},\tilde{x}}|\, d\tilde{x} \\
&= -\int_{\omega^+} \tilde{v}(T(x))\tilde{\eta}'(T(x)) \underbrace{|\det J_{T^{-1},\tilde{x}}|\,|\det J_{T,x}|}_{=1}\, dx \\
&= -\int_{\omega^+} v(x)\eta'(x)\, dx.
\end{aligned} \tag{5.104}$$

This shows that $v' := w = \tilde{v}' \circ T$ is the weak derivative of v. □

Remark 5.13. *The preceding Lemma 5.6 can be generalized easily to the case of vector-valued functions. Also note, that the statements from Lemma 5.5 and Lemma 5.6 can be generalized to hold in more abstract settings. These results are well-known and can be found in the literature.*

*

In the proof of Lemma 5.6 it was shown that weak derivatives can be transformed using diffeomorphisms with constant Jacobians. Especially this can be used to transform all terms including derivatives that occur in the weak form of the PDE from one patch to another. Consequently, any harmonic function $\psi \in \mathcal{H}(\omega^+)$ on ω^+ can be identified with a unique harmonic function $\tilde{\psi} \in \mathcal{H}(\tilde{\omega}^+)$ on $\tilde{\omega}^+$, with

$$\psi = \tilde{\psi} \circ T. \tag{5.105}$$

In particular, this also holds for the eigenfunctions of the generalized eigenvalue problem.

Using Lemma 5.5 and Lemma 5.6, the generalized eigenvalue problem (5.89) posed on the patches $\tilde{\omega}$ and $\tilde{\omega}^+$ can be transformed to the patches ω and ω^+, and explicit conditions on T and the coefficients appearing in the partial differential equation can be derived. In the next section, this will be done for the special case of second-order partial differential equations in divergence form.

5.3.2 Geometric reusability for second-order elliptic PDE in divergence form

In this section, geometric reusability of optimal shape functions for the case of scalar second-order differential equations in divergence form is investigated. The goal of this section is to present general conditions that both the data of the partial differential equation and the map T need to satisfy in order for an eigenfunction of (5.87) to be transformable via T to an eigenfunction of (5.89). The partial differential operator $\mathcal{L} : \mathcal{C}^2(\Omega) \to \mathcal{C}^0(\Omega)$ in this section reads

$$\mathcal{C}^2(\Omega) \ni u \mapsto \mathcal{L}[u] := -\operatorname{div}(A\nabla u) + b \cdot \nabla u + cu, \tag{5.106}$$

with coefficients $A \in [\mathcal{C}^1(\Omega)]^{d\times d}$, $b \in [\mathcal{C}^0(\Omega)]^d$ and $c \in \mathcal{C}^0(\Omega)$. Next, the integrals appearing in the gEVP eq. (5.89) will be rewritten. In order to do this, let $\tilde{u}, \tilde{v} \in \mathcal{H}(\tilde{\omega}^+)$ and use the transformation formula for integrals to see that the integral appearing on the left-hand side of (5.89) reads

$$\begin{aligned}
\mathbf{a}_{\tilde{\omega}}[\tilde{u},\tilde{v}] &= \int_{\tilde{\omega}} \nabla_{\tilde{x}}\tilde{u}(\tilde{x}) \cdot A(\tilde{x})\nabla_{\tilde{x}}\tilde{v}(\tilde{x})\, d\tilde{x} + \int_{\tilde{\omega}} b(\tilde{x}) \cdot \nabla_{\tilde{x}}\tilde{u}(\tilde{x})\, d\tilde{x} \\
&\quad + \int_{\tilde{\omega}} c(\tilde{x})\tilde{u}(\tilde{x})\tilde{v}(\tilde{x})\, d\tilde{x} \\
&\overset{\text{Trafo.}}{=} \int_{\omega} \nabla_{\tilde{x}}\tilde{u}(T(x)) \cdot A(T(x))\nabla_{\tilde{x}}\tilde{v}(T(x))|\det J_{T,x}(x)|\, dx \\
&\quad + \int_{\omega} b(T(x)) \cdot \nabla_{\tilde{x}}\tilde{u}(T(x))\tilde{v}(T(x))|\det J_{T,x}(x)|\, dx \\
&\quad + \int_{\omega} c(T(x))\tilde{u}(T(x))\tilde{v}(T(x))|\det J_{T,x}(x)|\, dx \\
&= \int_{\omega} J_{T,x}^{-T}(x)\nabla_x\tilde{u}(T(x)) \cdot A(T(x))J_{T,x}^{-T}(x)\nabla_x\tilde{v}(T(x))|\det J_{T,x}(x)|\, dx \\
&\quad + \int_{\omega} b(T(x)) \cdot J_{T,x}^{-T}(x)\nabla_x\tilde{u}(T(x))\tilde{v}(T(x))|\det J_{T,x}(x)|\, dx \\
&\quad + \int_{\omega} c(T(x))\tilde{u}(T(x))\tilde{v}(T(x))|\det J_{T,x}(x)|\, dx.
\end{aligned}$$

For practical applications, the transformations of interest are translations, scalings, rotations and sometimes shearings. All of these transformations are linear mappings and have constant Jacobians. This will be assumed in the following.

Assumption 5.1. *The map T is linear, i.e. the Jacobian $J_{T,x}$ is constant.*

◇

Lemma 5.6 can be used to identify $\tilde{u}$ and $\tilde{v}$ with unique functions $u, v \in \mathcal{H}(\tilde{\omega}^+)$, implying

that

$$\begin{aligned}\mathbf{a}_{\tilde{\omega}}[\tilde{u},\tilde{v}] \quad &= \quad \int_{\omega} \nabla_x u(x) \cdot J_{T,x}^{-1} A(T(x)) J_{T,x}^{-T} \nabla_x v(x) |\det J_{T,x}|\, dx \\ &\quad + \int_{\omega} b(T(x)) \cdot J_{T,x}^{-T}(x) \nabla_x u(x) v(x) |\det J_{T,x}|\, dx \\ &\quad + \int_{\omega} c(T(x)) u(x) v(x) |\det J_{T,x}|\, dx.\end{aligned} \tag{5.107}$$

In the same way, the integral over $\tilde{\omega}^+$ appearing on the right-hand side of (5.89) can be transformed to an integral over ω^+,

$$\begin{aligned}\mathbf{a}_{\tilde{\omega}^+}[\tilde{u},\tilde{v}] \quad &= \quad \int_{\omega^+} \nabla_x u(x) \cdot J_{T,x}^{-1} A(T(x)) J_{T,x}^{-T} \nabla_x v(x) |\det J_{T,x}|\, dx \\ &\quad + \int_{\omega^+} b(T(x)) \cdot J_{T,x}^{-T}(x) \nabla_x u(x) v(x) |\det J_{T,x}|\, dx \\ &\quad + \int_{\omega^+} c(T(x)) u(x) v(x) |\det J_{T,x}|\, dx.\end{aligned} \tag{5.108}$$

This allows to state compatibility conditions for different types of transformation T and the data A, b, c of the partial differential equation. Obviously, the eigenvalues of the generalized eigenvalue problems posed on $(\tilde{\omega}, \tilde{\omega}^+)$ and (ω, ω^+) coincide, if the integrals appearing on the left-hand side and on the right-hand side of (5.87) and (5.89) are identical up to a multiplicative factor. Using eqs. (5.107) and (5.108), this means

$$\begin{aligned}&\alpha \int_{\omega^+} \nabla_x u(x) \cdot A(x) \nabla_x v(x)\, dx + \alpha \int_{\omega^+} b(x) \cdot \nabla_x u(x) v(x)\, dx \\ &\qquad + \alpha \int_{\omega^+} c(x) u(x) v(x)\, dx \\ &= \int_{\omega^+} \nabla_x u(x) \cdot J_{T,x}^{-1} A(T(x)) J_{T,x}^{-T} \nabla_x v(x) |\det J_{T,x}| \\ &\qquad + \int_{\omega^+} b(T(x)) \cdot J_{T,x}^{-T} \nabla_x u(x) v(x) |\det J_{T,x}|\, dx \\ &\qquad + \int_{\omega^+} c(T(x)) u(x) v(x) |\det J_{T,x}|\, dx\end{aligned} \tag{5.109}$$

for $0 \neq \alpha \in \mathbb{R}$ and a similar condition with the same factor α should hold for the integration domain ω. One possibility for (5.109) to hold is that the following three conditions are satisfied

$$\begin{aligned}&\alpha \int_{\omega^+} \nabla_x u(x) \cdot A(x) \nabla_x v(x)\, dx \quad = \\ &\qquad \int_{\omega^+} \nabla_x u(x) \cdot J_{T,x}^{-1} A(T(x)) J_{T,x}^{-T} \nabla_x v(x) |\det J_{T,x}| \\ &\alpha \int_{\omega^+} b(x) \cdot \nabla_x u(x) v(x)\, dx \quad = \\ &\qquad \int_{\omega^+} b(T(x)) \cdot J_{T,x}^{-T} \nabla_x u(x) v(x) |\det J_{T,x}|\, dx \\ &\alpha \int_{\omega^+} c(x) u(x) v(x)\, dx \quad = \int_{\omega^+} c(T(x)) u(x) v(x) |\det J_{T,x}|\, dx\end{aligned} \tag{5.110}$$

and the same conditions should hold for all integrals replaced by integrals over ω. A sufficient condition for (5.110) to hold is that the continuous data A, b, c and the diffeomorphism T satisfy

$$\begin{aligned} \alpha A(x) &= J_{T,x}^{-1} A(T(x)) J_{T,x}^{-T} |\det J_{T,x}|, && \forall x \in \omega^+ \\ \alpha b(x) &= b(T(x)) \cdot J_{T,x}^{-T} |\det J_{T,x}|, && \forall x \in \omega^+ \\ \alpha c(x) &= c(T(x)) |\det J_{T,x}|, && \forall x \in \omega^+ . \end{aligned} \tag{5.111}$$

These relations are formulated for any diffusion coefficient A, any sink / source terms b and any constant changes c. They can be simplified substantially in the case of simpler PDEs. Consider e.g. the diffusion equation with a changing scalar diffusion coefficient, that is $A = \gamma \mathbb{I}, b^T = \begin{bmatrix} 0 & \dots & 0 \end{bmatrix}, c = 0$ for $\gamma \in \mathbb{R}$ in the reference configuration, and $A = \tilde{\gamma} \mathbb{I}, b^T = \begin{bmatrix} 0 & \dots & 0 \end{bmatrix}, c = 0$ in the new configuration. Suppose further that $T = \mathbb{I}$. In this case, (5.111) reads

$$\alpha \gamma = \tilde{\gamma} \tag{5.112}$$

which always can be satisfied by choosing $\alpha = \frac{\tilde{\gamma}}{\gamma}$. The optimal shape functions for any diffusion coefficient will hence coincide. Note that in the case of non-constant diffusion, $\gamma = \gamma(x)$, the relation can only be satisfied whenever $\tilde{\gamma}(x)$ is a multiple of $\gamma(x)$.

In the following, the investigation will be performed the other way around: First, the transformation is fixed, and then necessary conditions to be imposed on the data are developed.

Translations

Consider a translation of the form,

$$T : \mathbb{R}^d \to \mathbb{R}^d, \quad T : x \mapsto T(x) := Bx + y \tag{5.113}$$

for $B = \mathbb{I}$ and $y \in \mathbb{R}^d$. In this case we have $J_{T,x} = \mathbb{I}$ and $|\det J_{T,x}| = 1$. A sketch is shown in Figure 5.7. The conditions from eq. (5.111) read

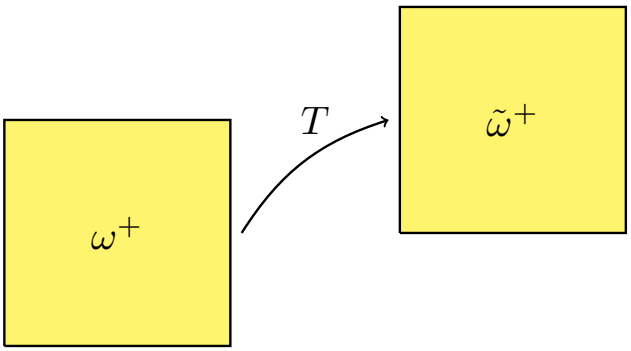


Figure 5.7: Sketch of a translation $x \mapsto x + y$

$$\begin{aligned} \alpha A(x) &= A(x + y), && \forall x \in \omega^+ \\ \alpha b(x) &= b(x + y), && \forall x \in \omega^+ \\ \alpha c(x) &= c(x + y), && \forall x \in \omega^+ . \end{aligned} \tag{5.114}$$

A sufficient condition for (5.114) to hold is that the data is constant. In this case $\alpha = 1$.

Isotropic scalings

Consider an isotropic scaling of the patch, that is scaling by the same factor $\beta > 0$ in all coordinate directions,

$$T : \mathbb{R}^d \to \mathbb{R}^d, \quad x \mapsto T(x) := \beta x = \beta \, \mathbb{I} \, x, \tag{5.115}$$

as sketched in Figure 5.8. Here it holds that

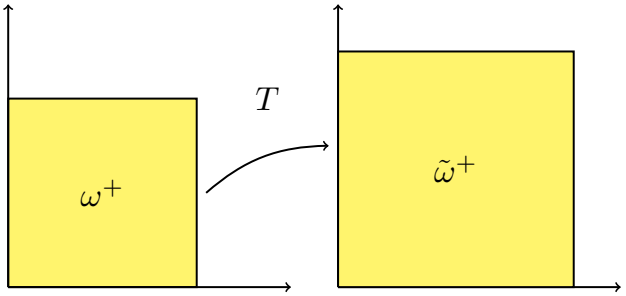


Figure 5.8: Sketch of an isotropic scaling $x \mapsto \beta x$

$$\begin{aligned} J_{T,x} &= \beta \, \mathbb{I}, \qquad \forall x \in \omega^+ \\ |\det J_{T,x}| &= \beta^d, \qquad \forall x \in \omega^+ \, . \end{aligned} \tag{5.116}$$

The conditions (5.111) read

$$\begin{aligned} \alpha A(x) &= \beta^{d-2} A(\beta x), \qquad \forall x \in \omega^+ \\ \alpha b(x) &= \beta^{d-1} b(\beta x), \qquad \forall x \in \omega^+ \\ \alpha c(x) &= \beta^d c(\beta x), \qquad \forall x \in \omega^+ \, . \end{aligned} \tag{5.117}$$

In the following, β is interpreted to establish conditions that the data needs to satisfy. Note that even for the case of constant coefficients, that is $A \in \mathbb{R}^{d\times d}, b \in \mathbb{R}^d, c \in \mathbb{R}$, the transport term $b \cdot \nabla u$ and the source term cu from (5.106) cause problems, since (5.117) implies that

$$b = \frac{\alpha}{\beta^{d-1}} b, \quad c = \frac{\alpha}{\beta^d} c, \tag{5.118}$$

must hold, i.e. $\alpha = \beta = 1$ is the only feasible choice for general b, c. Consequently, T must be the identity. In the case that b and c vanish, the conditions (5.117) for constant $A \in \mathbb{R}^{d\times d}$ read

$$A = \frac{\alpha}{\beta^{d-2}} A. \tag{5.119}$$

This means that any isotropic scaling $T(x) := \beta x$ can be chosen in the case of $b^T = \begin{bmatrix} 0 & \dots & 0 \end{bmatrix}$, $c = 0$ and $A \in \mathbb{R}^{d\times d}$, $\alpha = \beta^{d-2}$ is feasible.

Remark 5.14. *In the more general case of $b^T = \begin{bmatrix} 0 & \dots & 0 \end{bmatrix}, c = 0$ and a non-constant $A : \mathbb{R}^d \to \mathbb{R}$, the conditions (5.117) are satisfied whenever A coincides on ω^+ and $\tilde{\omega}^+$.*

*

Orthogonal transformations

In this section, general orthogonal transformations are considered, that is

$$T : \mathbb{R}^d \to \mathbb{R}^d, \quad x \mapsto Qx \tag{5.120}$$

for an orthogonal matrix $Q \in \mathbb{R}^{d\times d}$. It holds that

$$\begin{aligned} Q^{-1} &= Q^T \\ J_{T,x} &= Q \\ |\det J_{T,x}| &= 1. \end{aligned} \tag{5.121}$$

The conditions (5.117) read

$$\begin{aligned} \alpha A(x) &= Q^T A(Qx)Q, && \forall x \in \omega^+ \\ \alpha b(x) &= b(Qx)\cdot Q, && \forall x \in \omega^+ \\ \alpha c(x) &= c(Qx), && \forall x \in \omega^+ . \end{aligned} \tag{5.122}$$

The factor of the source term, c, does not cause problems in this case, as long as c coincides on ω^+ and $\tilde{\omega}^+$. Neither does the transport term b, whenever on $\tilde{\omega}^+$ it is the transformed transport term from ω^+. In the case of $b^T = \begin{bmatrix}0 & \dots & 0\end{bmatrix}, c = 0$ and constant A, it must only be assumed that $AQ^T = \alpha Q^T A$ for some $\alpha \in \mathbb{R}$. A special case of orthogonal transformations are rotations, exemplarily shown in Figure 5.9.

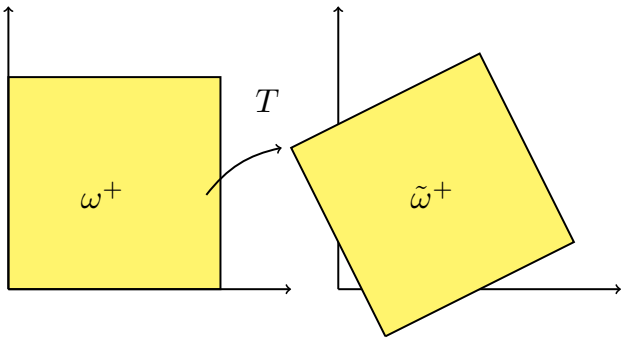


Figure 5.9: Sketch of a rotation. Rotations are a special case of orthogonal transformations.

Shearings

In this section, shearings of the domain are considered. Without loss of generality, consider a shearing in x_1-direction, that is

$$T : \mathbb{R}^d \to \mathbb{R}^d, \quad x \mapsto T(x) := Sx, \quad \text{with } S = \begin{bmatrix} 1 & \varepsilon & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ for } \varepsilon \in \mathbb{R}. \tag{5.123}$$

A sketch of this type of transformation is shown in Figure 5.10 In this case, it holds that

$$J_{T,x} = S, \quad |\det J_{T,x}| = 1, \quad S^{-1} = \begin{bmatrix} 1 & -\varepsilon & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{5.124}$$

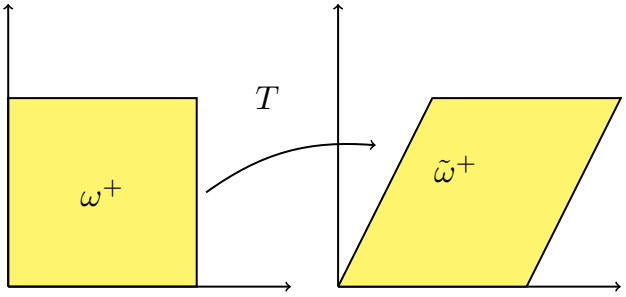


Figure 5.10: Sketch of a shearing in x_1-direction.

The conditions (5.117) read

$$\begin{aligned} \alpha A(x) &= S^{-1}A(Sx)S^{-T}, && \forall x \in \omega^+ \\ \alpha b(x) &= b(Sx) \cdot S^{-T}, && \forall x \in \omega^+ \\ \alpha c(x) &= c(Sx), && \forall x \in \omega^+ . \end{aligned} \tag{5.125}$$

For the case of $b^T = \begin{bmatrix} 0 & \dots & 0 \end{bmatrix}, c = 0$, the conditions are satisfied whenever for a fixed nonzero α it holds that

$$\alpha S A(x) = A(Sx)S^{-T}, \quad \forall x \in \omega^+ . \tag{5.126}$$

Combination of transformations

The transformations that were considered previously can be combined in order to obtain compatibility conditions for more complex transformations. Let two linear transformations T_1, T_2 be given as

$$T_i : \mathbb{R}^d \to \mathbb{R}^d, \quad x \mapsto T_i(x) := B_i x + y_i, \quad i = 1, 2. \tag{5.127}$$

The concatenation of the movements is given by

$$\begin{aligned} (T_2 \circ T_1) : \mathbb{R}^d \to & \ \mathbb{R}^d, \\ x \mapsto & \ (T_2 \circ T_1)(x) := B_2(B_1 x + y_1) + y_2 \\ & = B_2 B_1 x + B_2 y_1 + y_2 \end{aligned} \tag{5.128}$$

and therefore

$$J_{T_2 \circ T_1, x} = B_2 B_1, \quad |\det J_{T_2 \circ T_1, x}| = |\det B_2 B_1|. \tag{5.129}$$

The conditions (5.117) for an application of $T_2 \circ T_1$, that must be fulfilled for a fixed nonzero $\alpha \in \mathbb{R}$, read

$$\begin{aligned} \alpha A(x) &= (B_2 B_1)^{-1} A(T_2(T_1(x)))(B_2 B_1)^{-T} |\det B_2 B 1|, \\ \alpha b(x) &= b(T_2(T_1(x))) \cdot B_2 B_1 |\det B_2 B_1|, \\ \alpha c(x) &= c(T_2(T_1(x)))|\det B_2 B_1| \end{aligned} \tag{5.130}$$

for all $x \in \omega^+$.

6 Numerical computation of Optimal Bases

This chapter investigates numerically computed optimal basis functions, as well as global approximation properties of the enriched Partition of Unity Method. The differential equations considered in this section all present some kind of fine-scale behavior that is hard to capture using standard polynomial bases on coarse patches. The missing fine-scale information of coarse solutions should then be recovered using optimal local basis functions.

Four benchmark problems are investigated in Sections 6.1 to 6.4, and the structure of these sections is as follows: First, the corresponding partial differential equation is presented in its strong form. Trial and test spaces are introduced, together with the variational formulation of the problem, whose solvability is briefly discussed. Next, the influence of the various types of structured boundary data and their parameters, presented in Section 5.2.5, on the solutions of the generalized eigenvalue problem is investigated. Claims regarding the expected performance in global simulations can be made based on the number and magnitude of the largest eigenvalues that were computed in each case, since these eigenvalues encode information on the worst case best-approximation error, cf. Section 5.1. Finally, the claims are validated by identifying a numerical reference solution and discussing the error of global computations using increasing numbers of optimal shape functions as enrichments. This allows to draw conclusions from the benchmark problems and better understand the overall computational process.

Consider an elliptic partial differential equation of even order $2\,\mathrm{k}$, and let ω be a patch from the cover employed in a Partition of Unity Method. The approximation power of the local approximation space $\mathcal{V}(\omega)$ should be enhanced, and as discussed in Chapters 4 and 5, the local enrichment space takes the form

$$\mathcal{E}(\omega) = \operatorname{span}\{u_{\omega,f}, u_{\omega,g^0}, \ldots, u_{\omega,g^{2\,\mathrm{k}-1}}\} + \mathcal{V}_{\mathrm{n}}^{\mathrm{opt}}(\omega), \tag{6.1}$$

with particular solutions $u_{\omega,f}, u_{\omega,g^0}, \ldots, u_{\omega,g^{2\,\mathrm{k}-1}}$ for the data and an optimal approximation space $\mathcal{V}_{\mathrm{n}}^{\mathrm{opt}}(\omega)$ for the homogeneous part of the solution, $u_{\omega,\mathrm{hom}}$. A suitable choice of the dimension n of this optimal local approximation space is not generic and depends on the problem at hand. In the following benchmark problems, the fine-scale behavior of the solutions is expected to be localized in only few patches, and the corresponding enrichment spaces take the form (6.1). On the remaining patches, no fine-scale behavior is anticipated and only particular solutions will be used to span the corresponding enrichment spaces.

The first benchmark problem from Section 6.1 is a Poisson equation with a coefficient that jumps from value 1.0 to a value of 100.0 in a small square region located in the center of the domain. If not resolved correctly, the jump of the coefficient has a strong influence on the true solution, since its behavior in both parts of the domain will differ significantly. Section 6.2 considers a stationary convection diffusion equation with oscillating coefficients on a fine scale around the center of the domain, taking values between 1 and 121. The solution is expected to oscillate on a fine-scale, and resolving the coefficients oscillations requires heavy spatial refinement in traditional methods. The third benchmark, presented in Section 6.3, investigates the equations of linear elasticity for an isotropic material in two spatial dimensions on a plate with a circular hole located in its center. Resolving the periphery of the hole, as well as approximating sufficiently accurate values of the strains near this periphery, usually requires heavy spatial refinement. Section 6.4 then considers linear elasticity in $d = 3$, again for the case of a plate with a circular hole located in its center. Additionally, the domain is made from two isotropic material plies modeling steel (bottom ply) and aluminum (top ply). The first three benchmarks are pure Dirichlet problems, having boundary conditions prescribed globally. The fourth benchmark only prescribes the values of the solution on the left and right face of the three dimensional plate, which will lead to additional difficulties.

In the following, all optimal shape functions are computed on a discretization of level 8 using polynomials of degree two. B-Splines are quadratic unless stated otherwise and trigonometric functions are treated as patchwise polynomials of order 4. The oversampling factor is chosen as $\tau = 2.0$. All simulations were performed using the `PUMA` software toolkit developed by Fraunhofer SCAI ([SCA]) and the `optbasefun` module for `Python 3` developed in the scope of this thesis for the computation and use of optimal basis functions.

6.1 Poisson equation with jumping coefficient

This section investigates a Poisson problem with a jumping coefficient. The coefficient appearing in the problem is 100 times larger in a small region located in the center of the domain than outside, leading to a steep gradient of the weak solutions. This fine-scale behavior is difficult to resolve using coarse discretizations and standard polynomial bases.

In Section 6.1.1, the strong and weak formulation of the Poisson equation with a jumping coefficient is presented. The set of boundary conditions, as well as the trial and test spaces are introduced. In Section 6.1.2, the effects of different choices of structured boundary data on the set of optimal shape functions are investigated. Finally, Section 6.1.3 presents the results of the global, enriched computation. In all of the tables presented in the remainder of this section, dim is the dimension of the discrete space of harmonic functions, which coincides with the number of employed boundary functions. Furthermore, $\lambda_0 \geqslant \lambda_1 \geqslant \dots$ denote the obtained eigenvalues in descending order, and n_{-i} is the number of eigenvalues that are larger than 10^{-i}, i.e.

$$n_{-i} = \mathrm{card}\{\lambda : \lambda > 10^{-i}\}, \quad i = 1, 2, 4, 8. \tag{6.2}$$

6.1.1 Problem formulation

Problem 12 presents the strong form of the PDE under study.

Problem 12: Poisson problem with jumping coefficient.

Let $\Omega := [-6,6]^2 \subset \mathbb{R}^2$, let $A_{1,1} : \mathbb{R}^2 \to \mathbb{R}$ and $A : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$\begin{aligned} A_{1,1}(x) &:= \begin{cases} 100, & x \in [-1,1]^2 \\ 1, & \text{else} \end{cases}, \\ A(x) &:= A_{1,1}(x) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{6.3}$$

and consider the partial differential operator

$$\mathcal{L} : \mathcal{C}^2(\Omega) \to \mathcal{C}^0(\Omega), \quad v \mapsto \mathcal{L}\, v := -\operatorname{div}(A \nabla u). \tag{6.4}$$

Moreover, let

$$\begin{aligned} f : \mathbb{R}^2 \to \mathbb{R}, \quad x = (x_1, x_2) \mapsto f(x) := x_1 \\ g : \mathbb{R}^2 \to \mathbb{R}, \quad x = (x_1, x_2) \mapsto g(x) := x_2 \end{aligned} \tag{6.5}$$

Find a function $u \in \mathcal{C}^2(\Omega)$ satisfying

$$\begin{aligned} -\mathcal{L}\, u(x) &= f(x), \quad \text{in } \Omega \\ u(x) &= g(x), \quad \text{on } \partial\Omega. \end{aligned} \tag{6.6}$$

Problem 12 is a special case of the stationary convection diffusion equation for $b = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$, $c = 0$. According to Section 3.1, the problem under study is uniformly elliptic.

Define the trial and test spaces

$$\begin{aligned} \mathrm{V}^{\text{trial}}(\Omega) &:= \{u \in \mathrm{H}^1(\Omega) : \operatorname{tr}(u) = g \text{ on } \partial\Omega\} \\ \mathrm{V}^{\text{test}}(\Omega) &:= \{v \in \mathrm{H}^1(\Omega) : \operatorname{tr}(v) = 0 \text{ on } \partial\Omega\}, \end{aligned} \tag{6.7}$$

multiply the differential equation with a test function $v \in \mathrm{V}^{\text{test}}(\Omega)$, integrate over Ω and use integration by parts. The obtained weak formulation corresponding to Problem 12 is stated in Problem 13.

Problem 13: Weak Poisson problem with jumping coefficient.

Let $\Omega := [-6,6]^2 \subset \mathbb{R}^2$, let $A_{1,1} : \mathbb{R}^2 \to \mathbb{R}$ and $A : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$\begin{aligned} A_{1,1}(x) &:= \begin{cases} 100, & x \in [-1,1]^2 \\ 1, & \text{else} \end{cases}, \\ A(x) &:= A_{1,1}(x) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{6.8}$$

as well as

$$\begin{aligned} f : \mathbb{R}^2 \to \mathbb{R}, \quad x = (x_1, x_2) \mapsto f(x) := x_1 \\ g : \mathbb{R}^2 \to \mathbb{R}, \quad x = (x_1, x_2) \mapsto g(x) := x_2. \end{aligned} \tag{6.9}$$

Define the bilinear form $\mathbf{a} : \mathrm{V}^{\text{trial}}(\Omega) \times \mathrm{V}^{\text{test}}(\Omega) \to \mathbb{R}$ and linear functional $\ell : \mathrm{V}^{\text{test}}(\Omega) \to \mathbb{R}$,

$$\begin{aligned} \mathbf{a}[u,v] &:= \int_\Omega A_{1,1} \nabla u \cdot \nabla v \, dx, \\ \ell(v) &:= \int_\Omega f v \, dx. \end{aligned} \tag{6.10}$$

Find a function $u \in \mathrm{V}^{\text{trial}}(\Omega)$ satisfying

$$\mathbf{a}[u,v] = \ell(v), \quad \forall v \in \mathrm{V}^{\text{test}}(\Omega). \tag{6.11}$$

The bilinear form and linear functional are both continuous since the coefficients are sufficiently regular. Moreover, the Poincaré Friedrichs inequality (Theorem 2.6) can be applied, showing that $\mathbf{a}$ is elliptic. Hence, the assumptions of the Lax-Milgram Theorem 2.4 are satisfied, showing the existence of a unique solution.

6.1.2 Influence of the boundary data

Now, the impact of different choices of boundary data on the eigenvalues is investigated, and as described in Section 5.2, the boundary data consisted of boundary hats, B-Splines and Fourier-type oscillating functions.

Boundary hats

The following Tables 6.1 - 6.5 present key numbers obtained from the computation of optimal shape functions using boundary hats as boundary data in the sampling problem. In all of these tables, ℓ refers to the boundary level on which the boundary hats are piecewise defined with support on only one of the patches on this level at a time. Moreover, pd stands for the polynomial degree(s) used in the computation. In Figure 6.1, the two largest eigenvalues of the computations are plotted against the corresponding boundary level. It is visible, that the largest eigenvalues of the computations for polynomial degrees $\{0\}$, $\{1\}$, $\{2\}$, $\{0,1\}$ and $\{0,1,2\}$ approach approximately

ℓ	pd	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	0	12	$1.355 \cdot 10^{-1}$	$1.355 \cdot 10^{-1}$	2	4	11	11
3	0	28	$1.89 \cdot 10^{-1}$	$1.89 \cdot 10^{-1}$	2	6	12	23
4	0	60	$2.21 \cdot 10^{-1}$	$2.21 \cdot 10^{-1}$	2	6	12	26
5	0	124	$2.385 \cdot 10^{-1}$	$2.385 \cdot 10^{-1}$	2	6	14	28
6	0	252	$2.472 \cdot 10^{-1}$	$2.472 \cdot 10^{-1}$	2	6	14	28
7	0	508	$2.514 \cdot 10^{-1}$	$2.514 \cdot 10^{-1}$	2	6	14	28
8	0	1020	$2.537 \cdot 10^{-1}$	$2.537 \cdot 10^{-1}$	2	6	14	28

Table 6.1: Key numbers obtained from the computation of the optimal shape functions using constant boundary hats on various boundary levels.

ℓ	pd	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	1	24	$2.395 \cdot 10^{-1}$	$6.062 \cdot 10^{-2}$	1	4	11	12
3	1	56	$2.424 \cdot 10^{-1}$	$7.835 \cdot 10^{-2}$	1	5	12	24
4	1	120	$2.463 \cdot 10^{-1}$	$1.099 \cdot 10^{-1}$	2	5	12	26
5	1	248	$2.497 \cdot 10^{-1}$	$1.365 \cdot 10^{-1}$	2	6	13	28
6	1	504	$2.519 \cdot 10^{-1}$	$1.574 \cdot 10^{-1}$	2	6	14	28
7	1	1016	$2.531 \cdot 10^{-1}$	$1.747 \cdot 10^{-1}$	2	6	14	28
8	1	2040	$2.538 \cdot 10^{-1}$	$1.947 \cdot 10^{-1}$	2	6	14	28

Table 6.2: Key numbers obtained from the computation of the optimal shape functions using linear boundary hats in x_1 and x_2 direction on various boundary levels.

ℓ	pd	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	2	24	$1.977 \cdot 10^{-1}$	$5.708 \cdot 10^{-2}$	1	4	11	12
3	2	56	$2.055 \cdot 10^{-1}$	$6.03 \cdot 10^{-2}$	1	4	12	24
4	2	120	$2.175 \cdot 10^{-1}$	$6.834 \cdot 10^{-2}$	1	5	12	26
5	2	248	$2.307 \cdot 10^{-1}$	$9.32 \cdot 10^{-2}$	1	5	13	28
6	2	504	$2.412 \cdot 10^{-1}$	$1.179 \cdot 10^{-1}$	2	6	14	28
7	2	1016	$2.484 \cdot 10^{-1}$	$1.411 \cdot 10^{-1}$	2	6	14	28
8	2	2040	$2.534 \cdot 10^{-1}$	$1.662 \cdot 10^{-1}$	2	6	14	28

Table 6.3: Key numbers obtained from the computation of the optimal shape functions using quadratic boundary hats in x_1 and x_2 direction on various boundary levels.

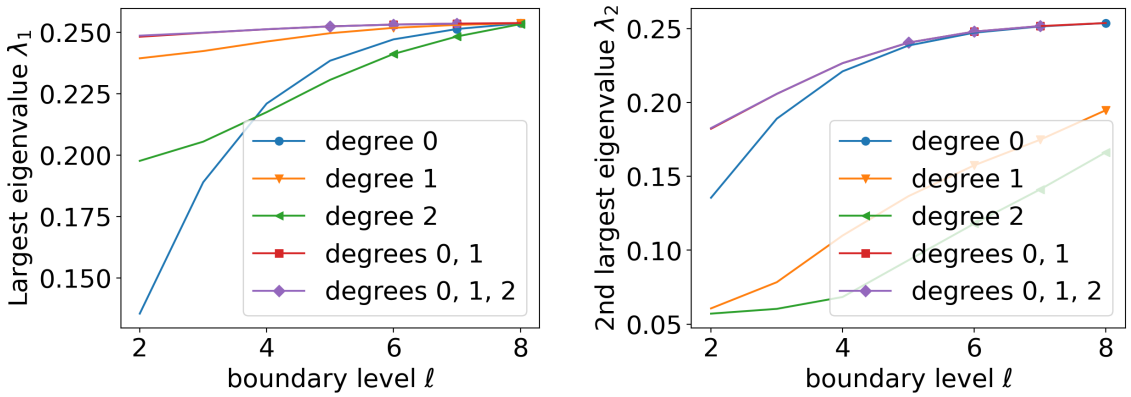
the same value when increasing the boundary level. The second eigenvalue tends to stabilize in the same region as the first eigenvalue whenever constants are included. Higher order boundary hats do not yield a large second eigenvalue on their own.

ℓ	pd	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	0, 1	36	$2.482 \cdot 10^{-1}$	$1.822 \cdot 10^{-1}$	2	6	12	19
3	0, 1	84	$2.498 \cdot 10^{-1}$	$2.058 \cdot 10^{-1}$	2	6	12	26
4	0, 1	180	$2.513 \cdot 10^{-1}$	$2.266 \cdot 10^{-1}$	2	6	14	28
5	0, 1	372	$2.525 \cdot 10^{-1}$	$2.405 \cdot 10^{-1}$	2	6	14	28
6	0, 1	756	$2.532 \cdot 10^{-1}$	$2.479 \cdot 10^{-1}$	2	6	14	28
7	0, 1	1524	$2.536 \cdot 10^{-1}$	$2.517 \cdot 10^{-1}$	2	6	14	28
8	0, 1	3060	$2.538 \cdot 10^{-1}$	$2.537 \cdot 10^{-1}$	2	6	14	29

Table 6.4: Key numbers obtained from the computation of the optimal shape functions using constant and linear boundary hats in x_1 and x_2 direction on various boundary levels.

ℓ	pd	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	0, 1, 2	60	$2.487 \cdot 10^{-1}$	$1.826 \cdot 10^{-1}$	2	6	12	22
3	0, 1, 2	140	$2.499 \cdot 10^{-1}$	$2.058 \cdot 10^{-1}$	2	6	12	28
4	0, 1, 2	300	$2.513 \cdot 10^{-1}$	$2.266 \cdot 10^{-1}$	2	6	14	29
5	0, 1, 2	620	$2.525 \cdot 10^{-1}$	$2.405 \cdot 10^{-1}$	2	6	14	29
6	0, 1, 2	1260	$2.532 \cdot 10^{-1}$	$2.479 \cdot 10^{-1}$	2	6	14	28
7	0, 1, 2	2540	$2.536 \cdot 10^{-1}$	$2.517 \cdot 10^{-1}$	2	6	14	28

Table 6.5: Key numbers obtained from the computation of the optimal shape functions using constant, linear and quadratic boundary hats in x_1 and x_2 direction on various boundary levels.



(a) (b)

Figure 6.1: Largest eigenvalue λ_1 (a) and second largest eigenvalue λ_2 (b) for increasing boundary levels ℓ obtained from the computations for boundary hats of various polynomial degrees. Markers are plotted every 500 degrees of freedom.

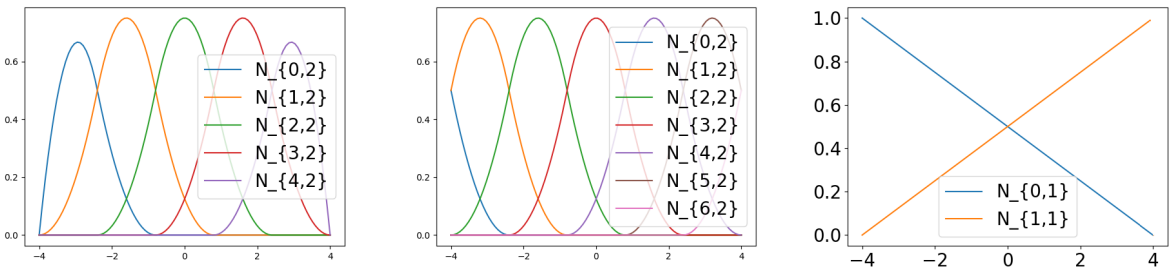
Increasing the boundary level shows, for all polynomial degrees, that the PDE under study seems to yield two very large and six more large eigenvalues. The largest eigenvalue is moreover seem to stabilize at around 0.253 - 0.254 for increasing boundary levels, regardless of the polynomial degree(s). On the other hand, the second eigenvalue variates more for different types of boundary

hats, even for high-dimensional approximation spaces. The largest values are obtained whenever the constant boundary hats are included in the boundary data. It seems that constant boundary hats have the strongest influence on the eigenvalues of the benchmark problem under study.

Note that increasing the boundary level or using multiple polynomial degrees increases the number of boundary hats, and since their number coincides with the number of sample problems to be solved, a suitable trade-off has to be identified. For the benchmark problem under study, it can be concluded from the previous discussion and the numbers of degrees of freedom, that using constant boundary hats on boundary level 8 yields the most promising results for a moderate 1020 degrees of freedom.

B-Splines

In this subsection, B-Splines are used as boundary data in the sampling problem. The B-Splines under study are quadratic and in some cases cubic. Inner knots are repeated in order to assure a maximum number of continuous B-Splines. Note that four ways of generating the boundary data were tested: The splines can either be generated in one coordinate direction only, or in both coordinate directions. When adding artificial knots outside of the coordinate range, the original vertices become inner knots and this in turn leads to additional splines having support on the endpoints of the domain edges. These splines are referred to as corner splines in the following. Figure 6.2 exemplarily shows the sets of quadratic univariate B-Splines for 4 different inner knots (without repetitions) with and without corner splines. As described in Section 5.2.5, bivariate B-Splines to be used as boundary data in the sampling problem are obtained by multiplying the univariate splines by linear splines in the other coordinate direction. Figure 6.2 (c) shows linear B-Splines without inner knots. The so-constructed bivariate splines are also referred to as quadratic B-Splines in the following.



(a) (b) (c)

Figure 6.2: Quadratic B-Splines with four different inner knots (without repetitions) without corner splines (a) and with corner splines (b). Linear B-Splines without inner knots are shown in (c).

Tables 6.6 - 6.9 present key numbers obtained from the four computations using quadratic B-Splines, with n denoting the number of inner knots without repetitions. Table 6.10 shows the results for cubic B-Splines in x_1 and x_2 direction with included corner splines. From Table 6.6 it can be seen, that using quadratic B-Splines only in x_1 direction and without considering corner splines results in one very large and four additional large eigenvalues. The size of the largest eigenvalue is 0.1403 in all computations shown. When adding corner splines, and / or considering

n	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
3	16	$1.403 \cdot 10^{-1}$	$5.721 \cdot 10^{-2}$	1	4	8	8
4	20	$1.403 \cdot 10^{-1}$	$5.74 \cdot 10^{-2}$	1	4	8	10
5	24	$1.403 \cdot 10^{-1}$	$5.743 \cdot 10^{-2}$	1	4	8	12
6	28	$1.403 \cdot 10^{-1}$	$5.744 \cdot 10^{-2}$	1	4	8	14
7	32	$1.403 \cdot 10^{-1}$	$5.744 \cdot 10^{-2}$	1	4	8	16
8	36	$1.403 \cdot 10^{-1}$	$5.744 \cdot 10^{-2}$	1	4	8	18
9	40	$1.403 \cdot 10^{-1}$	$5.744 \cdot 10^{-2}$	1	4	8	18

Table 6.6: Key numbers obtained from the computation of the optimal shape functions using quadratic B-Splines in x_1 direction, not including corner splines.

n	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
3	24	$2.47 \cdot 10^{-1}$	$2.443 \cdot 10^{-1}$	2	6	11	11
4	28	$2.47 \cdot 10^{-1}$	$2.446 \cdot 10^{-1}$	2	6	10	13
5	32	$2.471 \cdot 10^{-1}$	$2.448 \cdot 10^{-1}$	2	6	11	15
6	36	$2.471 \cdot 10^{-1}$	$2.449 \cdot 10^{-1}$	2	6	11	17
7	40	$2.471 \cdot 10^{-1}$	$2.449 \cdot 10^{-1}$	2	6	11	19
8	44	$2.471 \cdot 10^{-1}$	$2.449 \cdot 10^{-1}$	2	6	11	21
9	48	$2.471 \cdot 10^{-1}$	$2.45 \cdot 10^{-1}$	2	6	11	20

Table 6.7: Key numbers obtained from the computation of the optimal shape functions using quadratic B-Splines in x_1 direction, including corner splines.

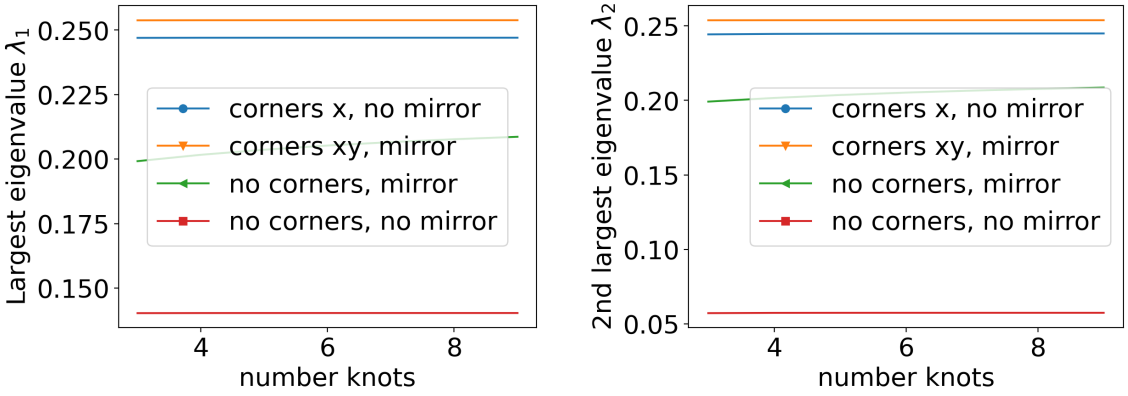
n	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
3	32	$1.992 \cdot 10^{-1}$	$1.992 \cdot 10^{-1}$	2	6	12	16
4	40	$2.017 \cdot 10^{-1}$	$2.017 \cdot 10^{-1}$	2	6	14	20
5	48	$2.036 \cdot 10^{-1}$	$2.036 \cdot 10^{-1}$	2	6	14	23
6	56	$2.053 \cdot 10^{-1}$	$2.053 \cdot 10^{-1}$	2	6	14	24
7	64	$2.066 \cdot 10^{-1}$	$2.066 \cdot 10^{-1}$	2	6	14	27
8	72	$2.077 \cdot 10^{-1}$	$2.077 \cdot 10^{-1}$	2	6	14	27
9	80	$2.087 \cdot 10^{-1}$	$2.087 \cdot 10^{-1}$	2	6	14	28

Table 6.8: Key numbers obtained from the computation of the optimal shape functions using quadratic B-Splines in x_1 and x_2 direction, not including corner splines.

quadratic B-Splines also in x_2 direction, there are two very large and six additional large eigenvalues. In the case of B-Splines in x_1 and x_2 direction together with corner splines, the two dominant eigenvalues are of identical size 0.2538 for all considered numbers of inner knots, and only the number of very small eigenvalues keeps increasing. The development of the two dominant eigenvalues for increasing numbers of oscillations is shown in Figure 6.3. The following Table 6.10 shows the results for cubic B-Splines. Also for this type of boundary data, there are two very large and six large eigenvalues, and both dominant eigenvalues are of similar size as in the case of quadratic B-Splines.

n	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
3	96	$2.538 \cdot 10^{-1}$	$2.538 \cdot 10^{-1}$	2	6	12	19
4	112	$2.538 \cdot 10^{-1}$	$2.538 \cdot 10^{-1}$	2	6	14	23
5	128	$2.538 \cdot 10^{-1}$	$2.538 \cdot 10^{-1}$	2	6	14	23
6	144	$2.538 \cdot 10^{-1}$	$2.538 \cdot 10^{-1}$	2	6	14	24
7	160	$2.538 \cdot 10^{-1}$	$2.538 \cdot 10^{-1}$	2	6	14	27
8	176	$2.538 \cdot 10^{-1}$	$2.538 \cdot 10^{-1}$	2	6	14	27
9	192	$2.538 \cdot 10^{-1}$	$2.538 \cdot 10^{-1}$	2	6	14	28

Table 6.9: Key numbers obtained from the computation of the optimal shape functions using quadratic B-Splines in x_1 and x_2 direction, including corner splines.



(a) (b)

Figure 6.3: Largest eigenvalue λ_1 (a) and second largest eigenvalue λ_2 (b) obtained from the computations using quadratic B-Splines with an increasing number of inner knots (without repetitions). All computations use less than 200 degrees of freedom.

n	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
3	80	$2.532 \cdot 10^{-1}$	$2.532 \cdot 10^{-1}$	2	6	12	20
4	96	$2.532 \cdot 10^{-1}$	$2.532 \cdot 10^{-1}$	2	6	14	24
5	112	$2.533 \cdot 10^{-1}$	$2.533 \cdot 10^{-1}$	2	6	14	24
6	128	$2.533 \cdot 10^{-1}$	$2.533 \cdot 10^{-1}$	2	6	14	27
7	144	$2.533 \cdot 10^{-1}$	$2.533 \cdot 10^{-1}$	2	6	14	28
8	160	$2.533 \cdot 10^{-1}$	$2.533 \cdot 10^{-1}$	2	6	14	28
9	176	$2.533 \cdot 10^{-1}$	$2.533 \cdot 10^{-1}$	2	6	14	29

Table 6.10: Key numbers obtained from the computation of the optimal shape functions using cubic B-Splines in x_1 and x_2 direction, including corner splines.

Concluding, it can be said that the corner splines seem to have a strong influence on the size of the largest eigenvalues. The experimental approach of only considering splines in one coordinate direction seems to perform quite well for the benchmark problem under study. However, this ap-

proach will probably only work whenever the fine-scale properties of the true solution do not vary much in the other coordinate directions. This behavior will probably be hardly visible by diffused boundary conditions on x_1. In regard of the relatively low quantity of B-Splines per edge, it is therefore recommended to consider them in both coordinate directions. Moreover, corner splines should always be included. Using higher order B-Splines does not seem to be advantageous.

Oscillating trigonometric functions

This subsection considers boundary data defined by the fourier-type basis functions $\mathfrak{C}(\Gamma^{\mathrm{F}})$ from Section 5.2 and the maximum number of oscillations, n_{osc}, is increased subsequently. Table 6.11 shows key numbers obtained from the corresponding computations. For Fourier-type basis func-

n_{osc}	dim	λ_1	$\lambda_2 = \lambda_3$	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	26	$7.203 \cdot 10^{-2}$	$2.083 \cdot 10^{-2}$	0	3	7	8
5	56	$7.371 \cdot 10^{-2}$	$2.087 \cdot 10^{-2}$	0	3	9	17
10	106	$7.392 \cdot 10^{-2}$	$2.091 \cdot 10^{-2}$	0	3	9	19
15	156	$7.399 \cdot 10^{-2}$	$2.093 \cdot 10^{-2}$	0	3	9	19
20	206	$7.4 \cdot 10^{-2}$	$2.094 \cdot 10^{-2}$	0	3	9	19
25	256	$7.401 \cdot 10^{-2}$	$2.094 \cdot 10^{-2}$	0	3	9	19
30	306	$7.402 \cdot 10^{-2}$	$2.094 \cdot 10^{-2}$	0	3	9	19
35	356	$7.402 \cdot 10^{-2}$	$2.095 \cdot 10^{-2}$	0	3	9	19
40	406	$7.402 \cdot 10^{-2}$	$2.095 \cdot 10^{-2}$	0	3	9	19
45	456	$7.403 \cdot 10^{-2}$	$2.095 \cdot 10^{-2}$	0	3	9	19
50	506	$7.403 \cdot 10^{-2}$	$2.095 \cdot 10^{-2}$	0	3	9	19
55	556	$7.403 \cdot 10^{-2}$	$2.095 \cdot 10^{-2}$	0	3	9	19
60	606	$7.403 \cdot 10^{-2}$	$2.095 \cdot 10^{-2}$	0	3	9	19
65	656	$7.403 \cdot 10^{-2}$	$2.095 \cdot 10^{-2}$	0	3	9	19
70	706	$7.403 \cdot 10^{-2}$	$2.095 \cdot 10^{-2}$	0	3	9	19
75	756	$7.403 \cdot 10^{-2}$	$2.095 \cdot 10^{-2}$	0	3	9	19
80	806	$7.403 \cdot 10^{-2}$	$2.095 \cdot 10^{-2}$	0	3	9	19
85	856	$7.403 \cdot 10^{-2}$	$2.095 \cdot 10^{-2}$	0	3	9	19
90	906	$7.403 \cdot 10^{-2}$	$2.095 \cdot 10^{-2}$	0	3	9	19
95	956	$7.403 \cdot 10^{-2}$	$2.095 \cdot 10^{-2}$	0	3	9	19

Table 6.11: Key numbers obtained from the computation of the optimal shape functions using Fourier-type basis functions in x_1 and x_2 direction for increasing numbers of oscillations.

tions, the results only show three moderately large eigenvalues, with the largest being approximately 0.074. The largest eigenvalues do not change in size for increasing numbers of oscillations. Concluding, it seems that Fourier-type boundary data performs poor for the benchmark problem under study, even when using high oscillatory functions and consequently about one thousand solutions to the sampling problem. In a global computation, the error will probably decrease very slowly when increasing the number of optimal shape functions.

Conclusive remarks

In the previous discussion it was established, that constant boundary hats on the highest feasible boundary level 8 seem to produce the most promising results: Two very large and six large eigenvalues are captured, and the sampling problem has to be solved a moderate 1020 times. The constant hats are essential to increase the first two dominant eigenvalues, but combining them with boundary hats of higher degrees rapidly increases the number of sampling problems to be solved, being a computationally expensive step in the overall computation. It was also seen that B-Splines should be defined in both coordinate directions, not only in x_1, and corner splines should be included. Moreover, higher order B-Splines do not seem to further improve the results. For quadratic B-Splines defined using only three different inner knots, two very large and six large eigenvalues are captured. These eigenvalues also remain approximately unchanged when increasing the numbers of inner knots. Furthermore, the dominant eigenvalues are of similar size to the dominant eigenvalues from the boundary hats computations. Since less than 200 sampling problems need to be solved, the results from this subsection look very promising. In order to visualize the difference in computational effort, Figure 6.4 shows the largest achievable eigenvalue in terms of the degrees of freedom, i.e. the number of sampling problems needed to compute it. Furthermore, it turned

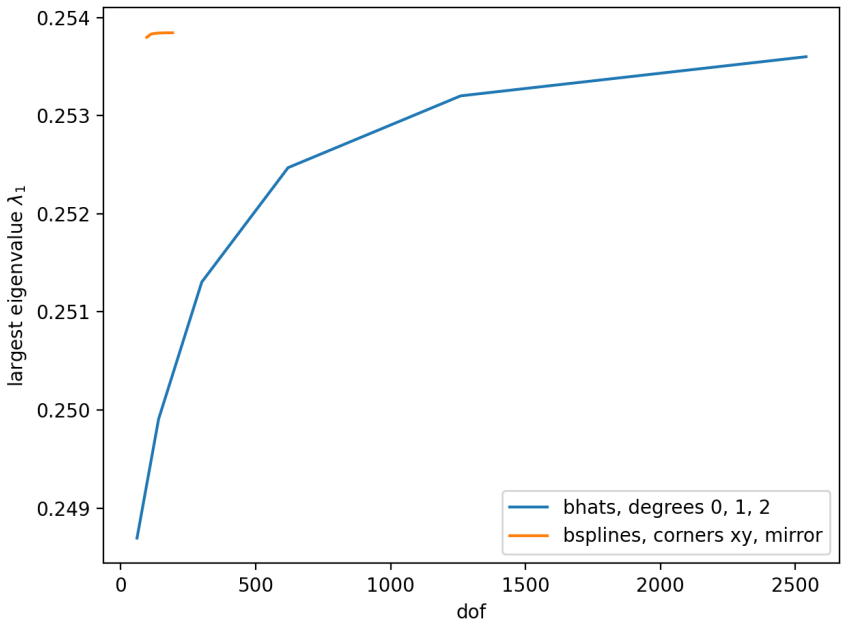


Figure 6.4: Comparison of the size of the largest eigenvalues attainable using boundary hats, and using B-Splines, in relation to the number of degrees of freedom (i.e. the dimension of the approximation to the space of harmonic functions).

out that only three moderate eigenvalues with the largest being of size ≈ 0.007 could be identified when using Fourier-type boundary data, even when using large numbers of highly oscillatory functions on the boundary. It is expected, that the optimal shape functions from these computations will perform worse than the other two approaches in a global experiment. However - and to be fair - this may be due the design of the model problem, whose solution is not expected to oscillate in any way.

6.1.3 Discusison of global errors

In the following, a reference solution for the benchmark problem is identified. Afterwards, the various sets of optimal shape functions investigated in the previous section are employed in a global computation and their approximation power is investigated. This will help to (in)validate the claims made in the previous section.

In order to identify a reasonable reference solution, numerical solutions of the PDE described in Problem 13 are computed for various discretization levels. The domain is given by $\Omega = [-6, 6]^2$ and the bounding box is chosen to be $[-6, 10]^2$, hence a discretization on any level ι consists of $\mathrm{m} = 3 \cdot 2^{\iota - 2}$ patches of size $h = \frac{12}{\mathrm{m}}$. The cover is chosen this way in order to have four patches of size 2 on level 3 covering ω, as will be seen further below. The energy of the solutions on different levels is computed, leading to a discrete approximation of the function

$$E : \mathbb{R}_+ \to \mathbb{R}, \quad h \mapsto E(h) := a_\Omega[u_h, u_h]^{\frac{1}{2}}, \tag{6.12}$$

where u_h is the solution computed on the corresponding discretization using patch size h. The obtained values of E, as well as the extrapolated limit for $h = 0$ are shown in Figure 6.5. It can be seen, that the extrapolated value at $h = 0$ approximately coincides with the value at $h = 0.015625$, corresponding to a discretization on level 10. The level 10 solution $u_{0.015625}$ will hence be chosen as reference solution in the upcoming analysis.

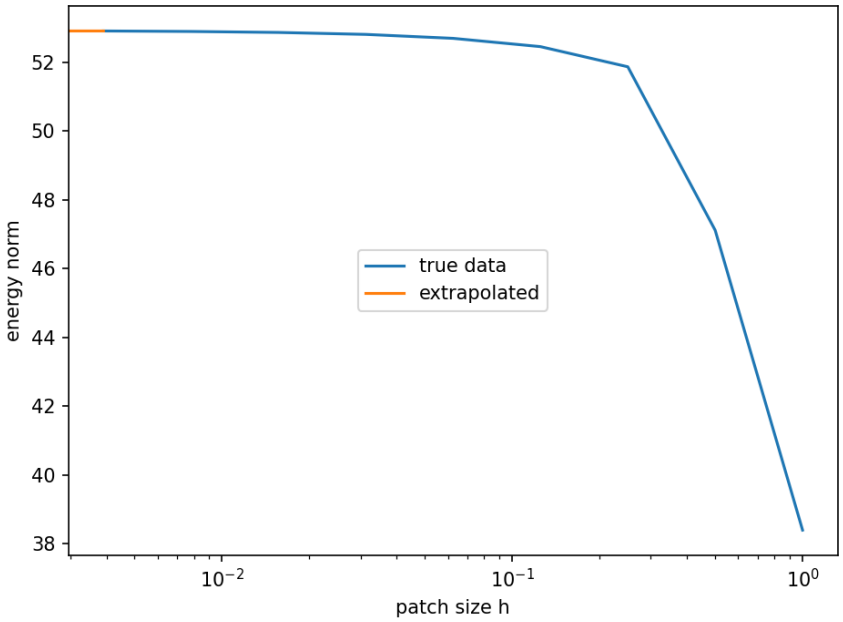
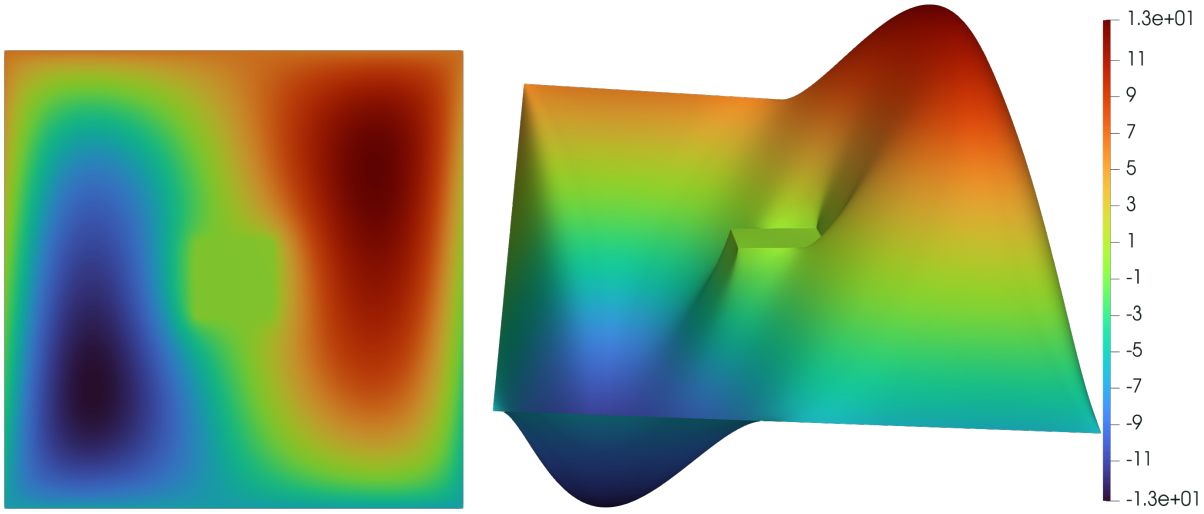


Figure 6.5: The energy of the solutions u_h for various values of h. The extrapolated limit value at $h = 0$ is also shown.

In the following, the enriched solutions obtained from the use of boundary hats, B-Splines and Fourier-type basis functions as boundary data in the sampling problem are compared to the reference solution. For an increasing number of enrichments, the enriched solutions are intuitively expected to be ever closer to the reference solution. However, note that additional enrichments used

in one part of the domain may improve the local error on this subdomain, but due to global regularity of the PUM solution also have an adverse effect on the error in the rest of the domain. Since the PUM minimizes the global (not the local) energy error, enrichments may hence be discarded. In such cases, the solutions remain invariant even when adding more and more enrichments. This was investigated in detail in [Sch11]. The reference solution is shown in Figure 6.6. In the fol-



(a) (b)

Figure 6.6: (a) Reference solution $u_{0.015625}$ for load $f(x) = x_1$ and global essential boundary condition $g(x) = x_2$. In (b), the warped state is shown.

lowing subsections, the domain is coarsely discretized as shown in Figure 6.7. Particular solutions for the load $f(x) = x_1$ and global essential boundary conditions $g(x) = x_2$ have been computed for all coarse patches, resp. all coarse boundary patches individually and will be used as local enrichments on the corresponding patches. Optimal shape functions are used as enrichments in the marked center of the domain, since fine-scale behavior is only expected in this region. From Figure 6.7, it is clearly visible that the jump interface of the coefficient is not resolved by the patches, and hence a solution without any further enrichments used on the red patches is expected to perform poor. This is confirmed by a relative global energy error of $\approx 25\%$.

In all upcoming computations, the optimal basis functions were computed on the oversampled patch $\omega^+ = [-4, 4]^2$ and used as enrichment in the four center patches, which are marked in red in Figure 6.7 and cover $\omega = [-2, 2]^2$.

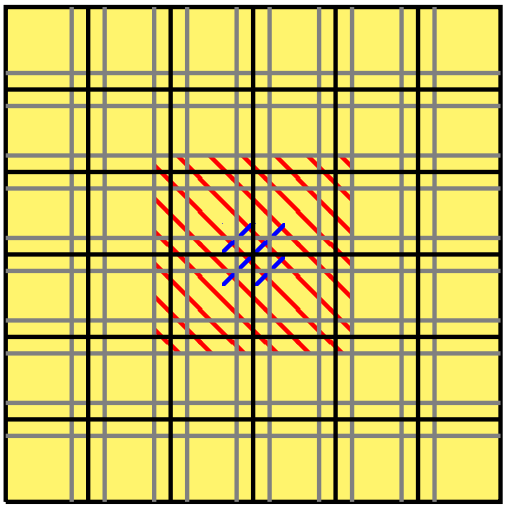


Figure 6.7: Coarse global discretization. In the blue marked region, the coefficient A takes the value 100 and outside it is equal to 1. Local particular solutions for load and essential boundary conditions are used on the corresponding patches, and the red patches will be further enriched using optimal shape functions. The red patches cover the patch ω.

Boundary hats

The relative energy error of the various series of enriched solutions on the full domain is shown for enrichments based on constant, linear and quadratic boundary hats, as well as combinations thereof, in tables 6.12 - 6.16. It can be seen, that all sets of optimal shape functions reduce the error from 25.22% to 4.69% when using at least 10 enrichments. The enrichments corresponding to the first two dominant eigenvalues have the largest impact, and are in all cases able to reduce the error to less than 6% on their own.

It stands out, that the error obtained using only one enrichment constructed from boundary hats of $pd = \{0, 1\}$ or $pd = \{0, 1, 2\}$ is significantly worse than in the case of a single polynomial degree.

ℓ	pd	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
2	0	25.22%	10.04%	5.07%	4.97%	4.93%	4.82%	4.69%	4.69%
3	0	25.22%	6.81%	5.08%	4.95%	4.92%	4.79%	4.69%	4.69%
4	0	25.22%	7.65%	5.07%	4.97%	4.92%	4.79%	4.69%	4.69%
5	0	25.22%	6.94%	5.07%	5.01%	4.92%	4.79%	4.69%	4.69%
6	0	25.22%	6.54%	5.07%	5.00%	4.92%	4.78%	4.69%	4.69%
7	0	25.22%	7.77%	5.07%	4.99%	4.92%	4.79%	4.69%	4.69%
8	0	25.22%	7.11%	5.07%	4.99%	4.92%	4.79%	4.69%	4.69%

Table 6.12: Development of the relative energy error for increasing numbers of enrichments obtained from constant boundary hats.

Concluding from the foregoing tables, it results that all sets of optimal shape functions can be used to drastically reduce the initial error of 25.22% to 4.69%. The boundary level moreover

ℓ	pd	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
2	1	25.22%	7.68%	5.43%	4.96%	4.93%	4.76%	4.69%	4.69%
3	1	25.22%	7.66%	5.28%	4.97%	4.93%	4.77%	4.69%	4.69%
4	1	25.22%	7.64%	5.21%	4.98%	4.93%	4.78%	4.69%	4.69%
5	1	25.22%	7.63%	5.16%	5.02%	4.93%	4.79%	4.69%	4.69%
6	1	25.22%	7.63%	5.12%	5.00%	4.93%	4.80%	4.69%	4.69%
7	1	25.22%	7.63%	5.10%	5.00%	4.93%	4.80%	4.69%	4.69%
8	1	25.22%	7.63%	5.08%	4.99%	4.93%	4.81%	4.69%	4.69%

Table 6.13: Development of the relative energy error for increasing numbers of enrichments obtained from linear boundary hats in x_1 and x_2.

ℓ	pd	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
2	2	25.22%	8.24%	5.55%	5.29%	5.00%	4.72%	4.69%	4.69%
3	2	25.22%	8.11%	5.45%	5.22%	4.99%	4.73%	4.69%	4.69%
4	2	25.22%	7.94%	5.63%	5.12%	4.96%	4.74%	4.69%	4.69%
5	2	25.22%	7.79%	5.40%	5.03%	4.95%	4.76%	4.69%	4.69%
6	2	25.22%	7.69%	5.24%	4.99%	4.94%	4.78%	4.69%	4.69%
7	2	25.22%	7.65%	5.16%	5.04%	4.93%	4.79%	4.69%	4.69%
8	2	25.22%	7.63%	5.11%	5.01%	4.93%	4.81%	4.69%	4.69%

Table 6.14: Development of the relative energy error for increasing numbers of enrichments obtained from quadratic boundary hats in x_1 and x_2.

ℓ	pd	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
2	0, 1	25.22%	19.00%	5.76%	5.18%	4.94%	4.85%	4.69%	4.69%
3	0, 1	25.22%	19.02%	5.30%	4.95%	4.93%	4.78%	4.69%	4.69%
4	0, 1	25.22%	19.07%	5.10%	4.93%	4.93%	4.79%	4.69%	4.69%
5	0, 1	25.22%	19.11%	5.06%	5.04%	4.92%	4.80%	4.69%	4.69%
6	0, 1	25.22%	19.13%	5.06%	5.01%	4.92%	4.81%	4.69%	4.69%
7	0, 1	25.22%	19.14%	5.07%	5.00%	4.92%	4.81%	4.69%	4.69%
8	0, 1	25.22%	19.15%	5.07%	4.99%	4.92%	4.81%	4.69%	4.69%

Table 6.15: Development of the relative energy error for increasing numbers of enrichments obtained from constant and linear boundary hats in x_1 and x_2.

does not need to be chosen very high. As seen before, there are two dominant eigenvalues in the computations for sufficiently large boundary levels, and hence both corresponding optimal shape functions should be used as enrichments. Using more enrichments only slightly improves the error, and using more than 10 enrichments does not change the result anymore. Plots of the difference between reference and enriched solutions give visual evidence for this, and are presented in Figure 6.8. In Figure 6.9 the decay of the energy error is visualized.

ℓ	pd	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
2	0, 1, 2	25.22%	18.98%	5.78%	5.18%	4.94%	4.86%	4.69%	4.69%
3	0, 1, 2	25.22%	19.02%	5.30%	4.95%	4.93%	4.78%	4.69%	4.69%
4	0, 1, 2	25.22%	19.07%	5.10%	4.93%	4.93%	4.79%	4.69%	4.69%
5	0, 1, 2	25.22%	19.11%	5.06%	5.04%	4.92%	4.80%	4.69%	4.69%
6	0, 1, 2	25.22%	19.13%	5.06%	5.01%	4.92%	4.81%	4.69%	4.69%
7	0, 1, 2	25.22%	19.14%	5.07%	5.00%	4.92%	4.81%	4.69%	4.69%

Table 6.16: Development of the relative energy error for increasing numbers of enrichments obtained from constant, linear and quadratic boundary hats in x_1 and x_2.

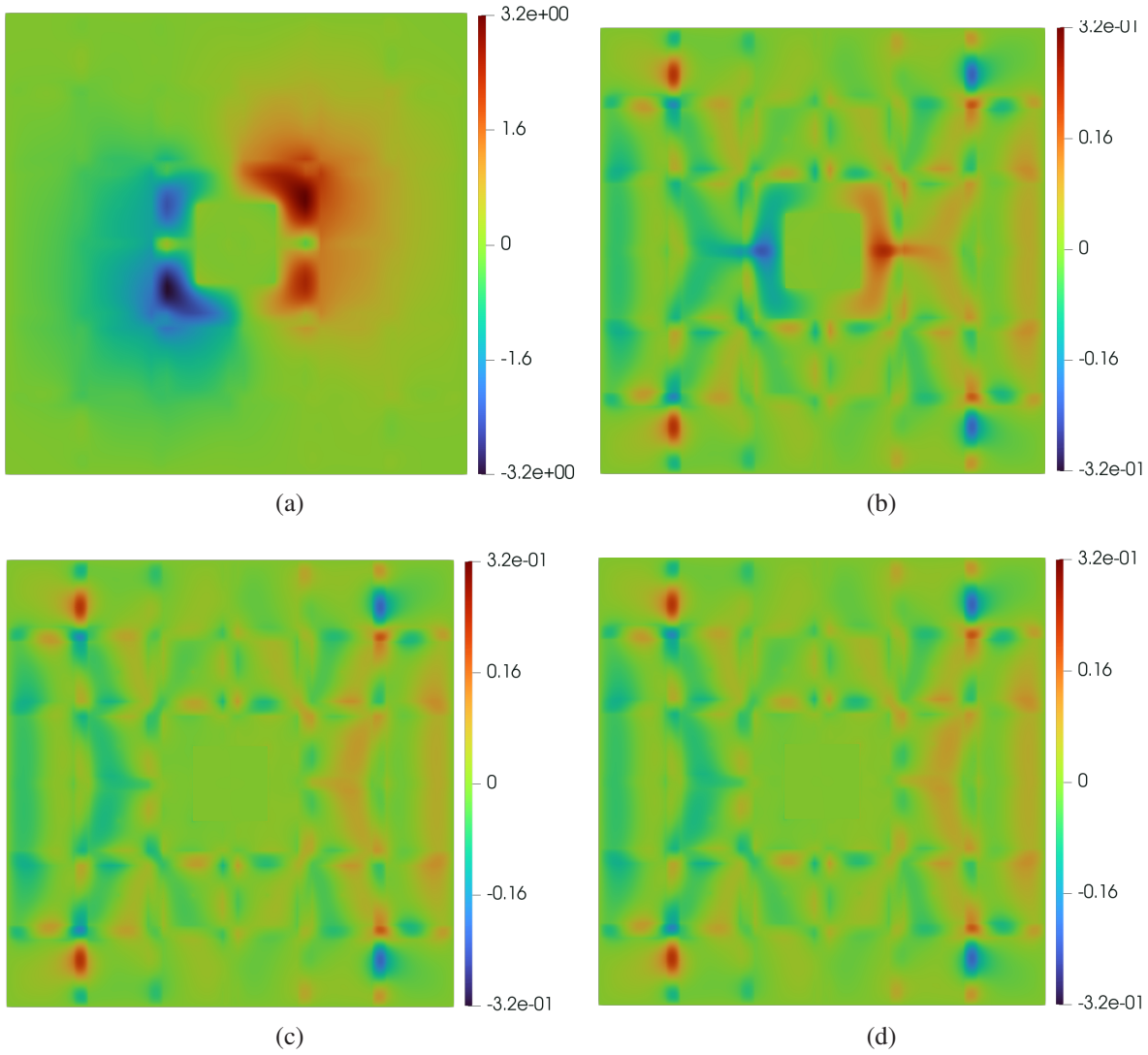


Figure 6.8: Difference between reference and enriched solutions, for (a) only particular solutions, (b) particular solutions and two enrichments, (c) particular solutions and ten enrichments, (d) particular solutions and twenty enrichments. Enrichments were computed using boundary level 8 and polynomial degrees $pd = \{0, 1\}$. Note that the minimum and maximum shown on the scale of (a) is ten times bigger than in the scales of (b), (c) and (d).

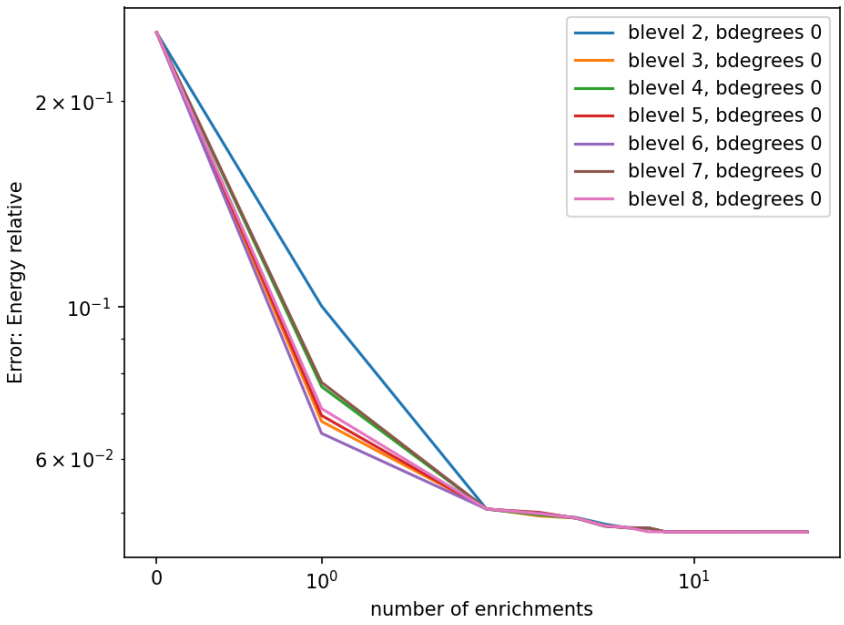


Figure 6.9: Relative energy error for increasing numbers of enrichments, constructed from boundary hats of different degrees and defined on various boundary levels.

B-Splines

The enrichments used in this section are constructed from B-Spline boundary data, and tables 6.17 - 6.20 present the relative energy errors of the enriched computations based on the four variants of quadratic B-Splines presented in Section 6.1.2. All sets of optimal shape functions reduce the error from 25.22% to $\approx 4.7\%$ when using at least 10 enrichments. Two enrichments from the optimal shape functions constructed from B-Splines only in x_1 direction and without considering corner splines lead to errors of 8.01%. Optimal shape functions constructed from all other types of B-Splines lead to errors of at most 5.3% whenever two enrichments are used, and the enrichments constructed from B-Splines in x_1 and x_2 direction including corner splines perform best.

n	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
4	25.22%	20.78%	8.01%	7.18%	5.00%	4.83%	4.70%	4.70%
5	25.22%	20.79%	8.01%	7.18%	5.00%	4.84%	4.70%	4.69%
6	25.22%	20.79%	8.01%	7.18%	5.00%	4.84%	4.70%	4.69%
7	25.22%	20.79%	8.01%	7.18%	5.00%	4.84%	4.70%	4.69%
8	25.22%	20.79%	8.01%	7.18%	5.00%	4.84%	4.70%	4.69%
9	25.22%	20.79%	8.01%	7.18%	5.00%	4.84%	4.70%	4.69%

Table 6.17: Development of the relative energy error for increasing numbers of enrichments obtained from quadratic B-Splines in x_1 direction without corner splines.

n	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
3	25.22%	7.71%	5.30%	4.91%	4.89%	4.77%	4.70%	4.70%
4	25.22%	7.71%	5.30%	4.91%	4.89%	4.77%	4.70%	4.70%
5	25.22%	7.71%	5.30%	4.91%	4.89%	4.77%	4.70%	4.69%
6	25.22%	7.71%	5.30%	4.91%	4.89%	4.77%	4.70%	4.69%
7	25.22%	7.71%	5.30%	4.91%	4.89%	4.77%	4.70%	4.69%
8	25.22%	7.71%	5.30%	4.91%	4.89%	4.77%	4.70%	4.69%
9	25.22%	7.71%	5.30%	4.91%	4.89%	4.77%	4.70%	4.69%

Table 6.18: Development of the relative energy error for increasing numbers of enrichments obtained from quadratic B-Splines in x_1 direction with corner splines.

The results for boundary data consisting of cubic B-Splines in x_1 and x_2 direction, including corner splines, are presented in Table 6.21. It is seen that the error decays very similar to that in the case of quadratic B-Splines presented in table 6.20. Again, the first enrichment is able to reduce a very large fraction of the error, leading to errors of 5.07% for only two enrichments.

Concluding, it results that all sets of optimal shape functions constructed from B-Spline boundary data can be used to drastically reduce the initial error of 25.22% to approximately 4.7%. It even seems that weakly chosen boundary data will only decrease the speed of decay, but ultimately, i.e. for sufficiently large numbers of enrichments, lead to errors of comparable size. For reasonable boundary data however, two enrichments corresponding to the two obtained dominant eigenvalues (cf. Section 6.1.2) were obtained, and hence the corresponding optimal shape functions should definitely be used as enrichments. Using more than ten enrichments, in case there are more, does

n	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
3	25.22%	8.73%	5.20%	5.05%	5.02%	4.89%	4.69%	4.69%
4	25.22%	8.56%	5.20%	5.05%	5.02%	4.87%	4.69%	4.69%
5	25.22%	7.78%	5.19%	5.05%	5.02%	4.84%	4.69%	4.69%
6	25.22%	9.37%	5.18%	5.05%	5.02%	4.87%	4.69%	4.69%
7	25.22%	8.77%	5.17%	5.04%	5.02%	4.83%	4.69%	4.69%
8	25.22%	8.19%	5.16%	5.03%	5.01%	4.85%	4.69%	4.69%
9	25.22%	8.10%	5.16%	5.03%	5.01%	4.82%	4.69%	4.69%

Table 6.19: Development of the relative energy error for increasing numbers of enrichments obtained from quadratic B-Splines in x_1 and x_2 direction without corner splines.

n	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
3	25.22%	7.77%	5.07%	4.99%	4.92%	4.78%	4.69%	4.69%
4	25.22%	7.95%	5.07%	4.99%	4.92%	4.83%	4.69%	4.69%
5	25.22%	8.16%	5.07%	4.99%	4.92%	4.78%	4.69%	4.69%
6	25.22%	8.18%	5.07%	4.99%	4.92%	4.81%	4.69%	4.69%
7	25.22%	7.23%	5.07%	4.99%	4.92%	4.81%	4.69%	4.69%
8	25.22%	8.33%	5.07%	4.99%	4.92%	4.78%	4.69%	4.69%
9	25.22%	7.28%	5.07%	4.99%	4.92%	4.81%	4.69%	4.69%

Table 6.20: Development of the relative energy error for increasing numbers of enrichments obtained from quadratic B-Splines in x_1 and x_2 direction with corner splines.

n	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
3	25.22%	7.12%	5.07%	4.99%	4.92%	4.79%	4.69%	4.69%
4	25.22%	7.89%	5.07%	4.99%	4.92%	4.83%	4.69%	4.69%
5	25.22%	8.16%	5.07%	4.99%	4.92%	4.78%	4.69%	4.69%
6	25.22%	8.23%	5.07%	4.99%	4.92%	4.81%	4.69%	4.69%
7	25.22%	7.49%	5.07%	4.99%	4.92%	4.78%	4.69%	4.69%
8	25.22%	7.97%	5.07%	4.99%	4.92%	4.82%	4.69%	4.69%
9	25.22%	7.37%	5.07%	4.99%	4.92%	4.81%	4.69%	4.69%

Table 6.21: Development of the relative energy error for increasing numbers of enrichments obtained from cubic B-Splines in x_1 and x_2 direction with corner splines.

not further reduce the error. Plots of the difference between reference and enriched solutions, shown in Figure 6.10, presents visual evidence for this. In Figure 6.11, the decay of the energy errors from table 6.20 is visualized.

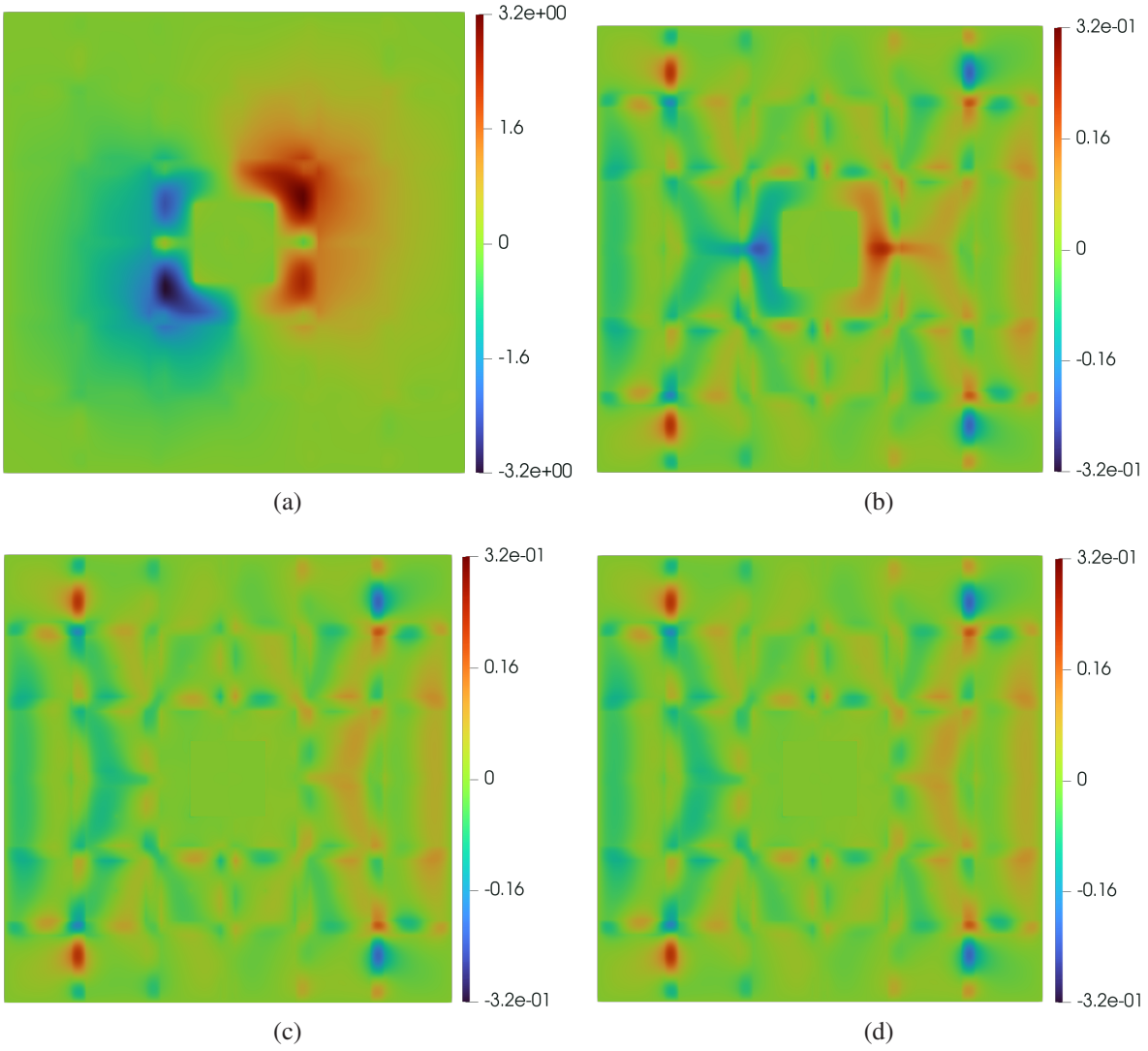


Figure 6.10: Difference between reference and enriched solutions, for (a) only particular solutions, (b) particular solutions and two enrichments, (c) particular solutions and ten enrichments, (d) particular solutions and twenty enrichments. Enrichments were computed from quadratic B-Splines in x_1 and x_2 direction, including the corner splines, for 9 different inner knots. Note that the scale of the errors in (a) is ten times larger than in (b), (c) and (d).

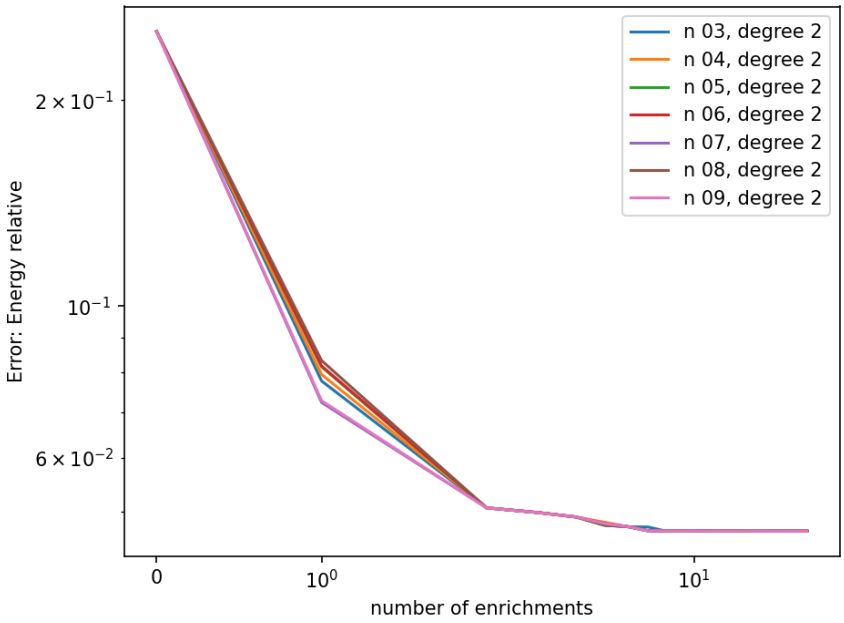


Figure 6.11: Relative energy error for increasing numbers of enrichments, constructed from quadratic B-Spline boundary data in x_1 and x_2, including corner splines, and defined for various numbers of inner knots (without repetitions).

Oscillating trigonometric functions

The following table 6.22 presents the development of the relative energy error for increasing numbers of optimal shape functions based on oscillatory Fourier-type basis functions used as enrichments. Note that the maximum number of optimal shape functions computed with this approach is 18. It is clearly visible, that the error decays very slow, and the first three enrichments for all numbers of oscillations are barely capable of reducing it. Additionally using the fourth enrichment results in a strong decay from $\approx 22\%$ to less than 8.3%. Additional enrichments then lead to slight improvements of the error to 4.7% for ten enrichments. The remaining nine enrichments furthermore only lead to a negligible improvement, and in the end all sets of optimal shape functions reduce the initial error of 25.22% to 4.69%. It can be concluded, that all sets of optimal shape

n_{osc}	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 19
2	25.22%	24.70%	24.00%	22.55%	8.81%	7.81%	-%	-%
5	25.22%	24.70%	23.87%	22.48%	8.28%	5.84%	4.70%	-%
10	25.22%	24.70%	23.78%	22.42%	7.98%	5.64%	4.70%	4.69%
15	25.22%	24.70%	23.74%	22.40%	7.87%	5.84%	4.70%	4.69%
20	25.22%	24.70%	23.72%	22.39%	7.82%	5.68%	4.70%	4.69%
25	25.22%	24.70%	23.71%	22.39%	7.79%	5.74%	4.70%	4.69%
30	25.22%	24.70%	23.71%	22.38%	7.77%	5.62%	4.70%	4.69%
35	25.22%	24.70%	23.70%	22.38%	7.76%	5.65%	4.70%	4.69%
40	25.22%	24.70%	23.70%	22.38%	7.75%	5.10%	4.70%	4.69%
45	25.22%	24.70%	23.70%	22.38%	7.75%	5.79%	4.70%	4.69%
50	25.22%	24.70%	23.70%	22.37%	7.74%	5.66%	4.70%	4.69%
55	25.22%	24.70%	23.70%	22.37%	7.74%	5.46%	4.70%	4.69%
60	25.22%	24.70%	23.69%	22.37%	7.74%	5.59%	4.70%	4.69%
65	25.22%	24.70%	23.69%	22.37%	7.73%	5.72%	4.70%	4.69%
70	25.22%	24.70%	23.69%	22.37%	7.73%	5.80%	4.70%	4.69%
75	25.22%	24.70%	23.69%	22.37%	7.73%	5.61%	4.70%	4.69%
80	25.22%	24.70%	23.69%	22.37%	7.73%	5.08%	4.70%	4.69%
85	25.22%	24.70%	23.69%	22.37%	7.73%	5.59%	4.70%	4.69%
90	25.22%	24.70%	23.69%	22.37%	7.73%	5.61%	4.70%	4.69%
95	25.22%	24.70%	23.69%	22.37%	6.78%	5.44%	4.70%	4.69%

Table 6.22: Development of the relative energy error for increasing numbers of enrichments obtained from Fourier-type basis functions in x_1 and x_2 direction with increasing number of maximum oscillations.

functions are able to substantially reduce the error. However, the error for increasing numbers of enrichments decays very slowly, and this coincides with the observation from Section 6.1.2, which revealed that no dominant eigenvalues were captured for Fourier-type boundary data. As stated before, it is possible that the Fourier-type data has a poor performance for the benchmark problem at hand since the homogeneous solution does not oscillate rapidly on a fine scale. For completeness, Figure 6.12 shows the development of the difference between various enriched solutions and the reference. In Figure 6.13, the slow decay of the the relative energy error is visualized.

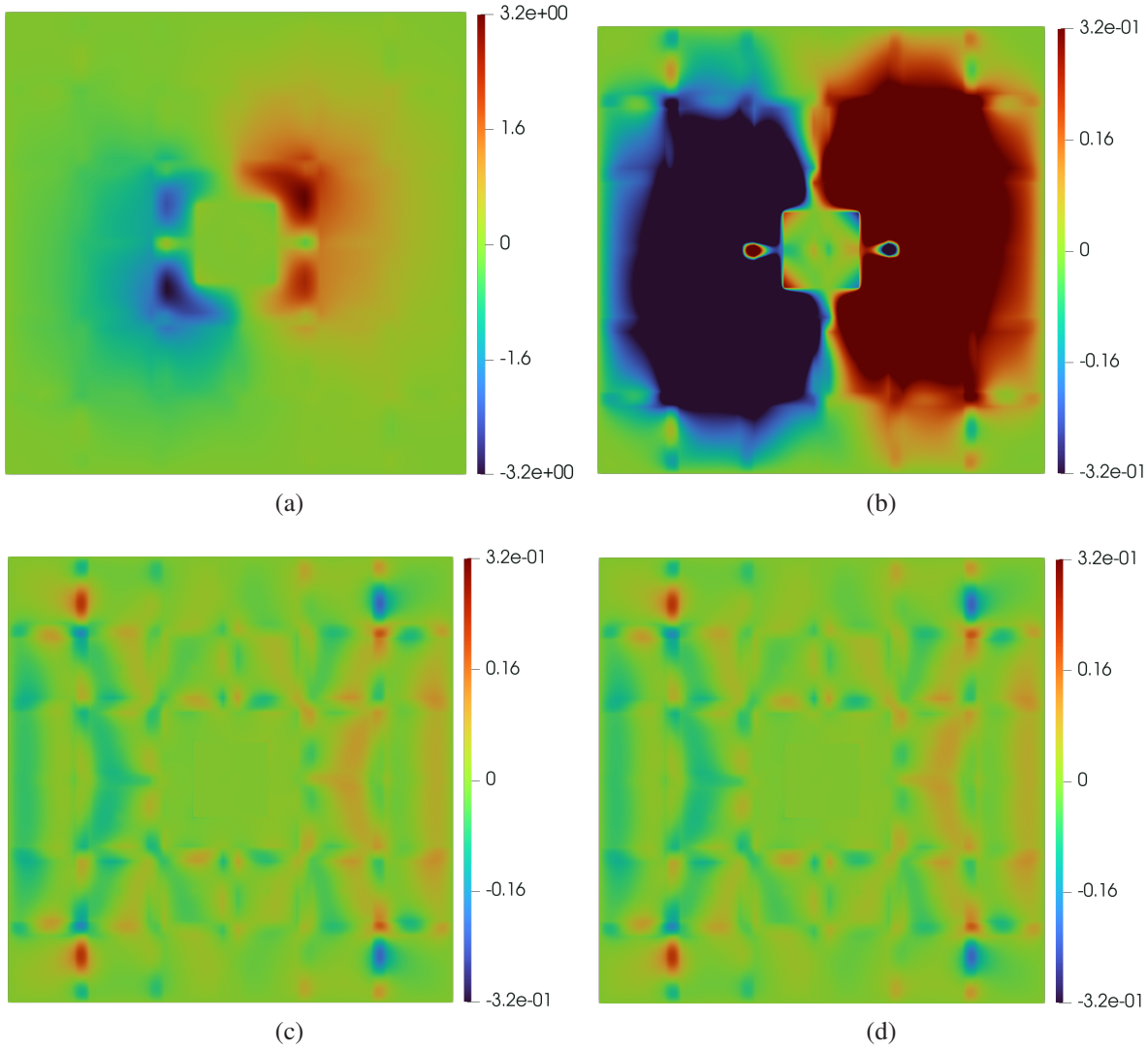


Figure 6.12: Difference between reference and enriched solutions, for (a) only particular solutions, (b) particular solutions and two enrichments, (c) particular solutions and ten enrichments, (d) particular solutions and nineteen enrichments. Enrichments were computed from Fourier-type boundary data in x_1 and x_2 direction, for a maximum of 95 oscillations. Note that the scale of the error in (a) is ten times larger than the scale of (b), (c) and (d).

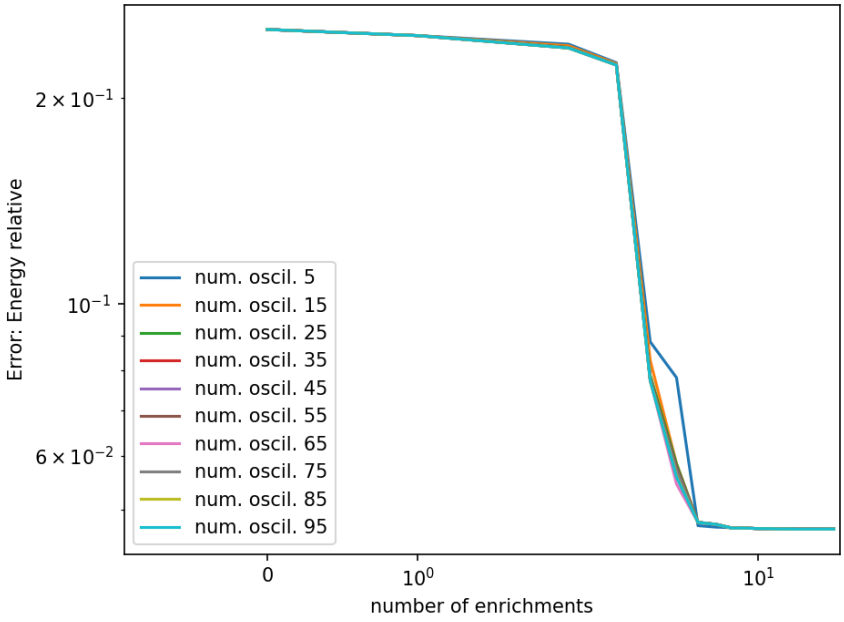


Figure 6.13: Relative energy error for increasing numbers of enrichments, constructed from Fourier-type boundary data in x_1 and x_2 direction for various maximum numbers of oscillations.

Conclusions from the benchmark problem

The foregoing detailed discussion revealed, that the homogeneous solution of the benchmark problem under study can be approximated well using optimal shape functions constructed from boundary hats, B-Splines or Fourier-type boundary data. All approaches were capable of reducing the initial error of 25.22% to $\approx 4.7\%$. However, the workload needed to compute the enrichments from the different approaches, as well as the rate of decay of the error, differ significantly. The boundary hats approach is the most versatile, but in general also the most expensive, while the B-Spline approach is the cheapest, since relatively few sampling problems have to be solved. Boundary hats, as well as B-Splines, led to two dominant eigenvalues. The Fourier approach on the other hand was not able to capture any dominant eigenvalues at all. It was also observed, that the optimal shape functions corresponding to dominant eigenvalues are able to reduce the error much faster. In total, the B-Spline approach, considering quadratic splines in x_1 and x_2 direction defined on 9 different inner nodes, together with the corresponding corner splines, showed the best overall performance. This is due to the fact that the sampling problem has to be solved relatively few times, and the resulting shape functions are able to reduce the energy error very quickly.

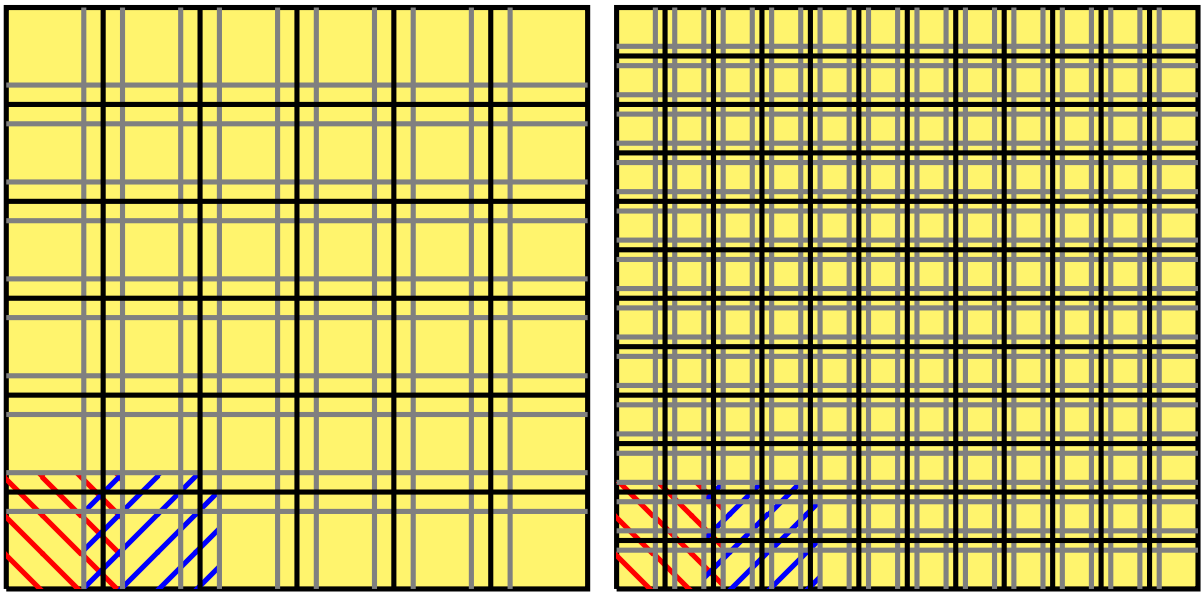
Residual error

In the previous discussion, particular solutions for the load were used on coarse patches and local function spaces on boundary patches were furthermore enriched with particular solutions for the essential boundary conditions. There are four patches in the center, whose corresponding local function spaces were enriched with an ever-increasing number of optimal shape functions computed using different types of boundary data in the sampling problem. It was seen in the previous Section 6.1.3, that all shape functions were capable of substantially reducing the initial error, but even for a large number of enrichments there was still a residual error which could not be improved anymore. The magnitude of the residual error was, furthermore, of the same size $\approx 4.7\%$ for all sets of optimal shape functions that were considered. This section aims to identify the source of this residual error.

The residual error of 4.7% may be due to three possible reasons: Either, the resolution of the enrichments was not high enough, or the error is due to the coarse discretization that is employed globally, especially as a result of the relatively large overlap between patches. The lifting of solutions furthermore implies that optimal basis functions for the approximation of the homogeneous part of local solution need to be computed and applied on all patches, not only in the center of the domain, and this may also have a negative influence on the results. Since optimal local basis functions basically replace the need for heavy refinement, the last two sources of error can be treated conjointly. When identifying the reference solution at the beginning of Section 6.1.3, it was observed that a global level 10 solution with patches of size $h = 0.015625$ captures all fine-scale features of the homogeneous solution. The enrichments have been computed on a global discretization on level 8, but since the patch ω^+ on which the enrichments are computed is much smaller, the patch sizes coincide. The enrichments are hence defined on the same sufficiently small length-scale as the reference solution. The previous argumentation shows, that the resolution of the enrichments cannot be responsible for the residual error.

In order to check the effect of the coarse discretization on the residual error, the particular solutions

corresponding to the coarse level 3 were used on a global discretization of the levels 3, 4 and 5. Figure 6.14 shows a discretization on two subsequent levels and two marked patches on the coarse level. The particular solutions are computed on these coarse patches, and they are also used on all patches on the fine level that result from a refinement of the corresponding coarse patch.



(a) (b)

Figure 6.14: (a) shows a level 3 discretization. Particular solutions are computed for all patches, including the ones marked in red and blue. (b) shows a level 4 discretization. The blue patches use the particular solutions computed on the blue patch from (a), and the red patches use the particular solutions computed on the red patch from (a).

This construction minimizes the influence of factors other than the coarse discretization level on the enriched computations. The following table 6.23 presents the results of the enriched computation, and Figure 6.15 shows plots of the differences of the enriched solutions to the reference solution $u_{0.015625}$.

coarse level	particular level	nn	relative energy error
3	3	20	4.70
4	3	20	2.80
5	3	20	2.50

Table 6.23: Relative errors for coarse discretization on levels 3, 4 and 5, when using particular solutions from level 3 and 20 optimal shape functions as enrichments in the center region.

The results show, that the coarse level has a strong influence on the residual error. This is due to the construction of the Partition of Unity Method, in which local function spaces (and local approximations) are chosen independently of each other and glued together with functions forming a partition of unity. Note that the use of optimal local basis functions on all patches, as described

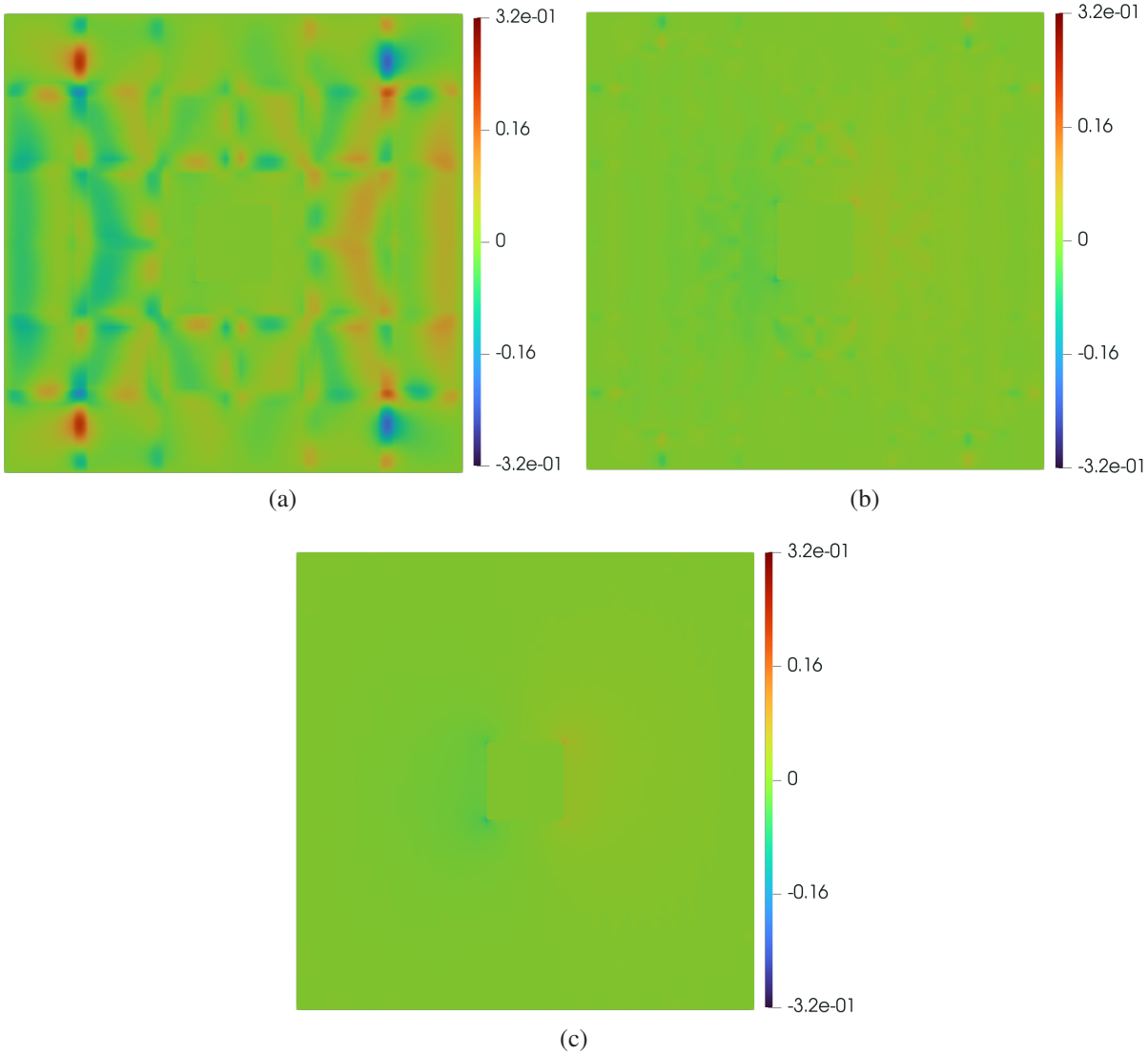


Figure 6.15: Difference between enriched solutions using particular solutions from level 3 and 20 optimal shape functions in the center region, for coarse discretization of level 3 (a), 4 (b) and 5 (c).

in the lifting (cf. Section 5.1.1), replaces the need for spatial refinement and a similar reduction of errors is expected in this case.

For subsequent benchmark problems, which show a residual error of similar structure, the previous discussion can be repeated analogously and will hence be omitted.

6.2 Stationary convection diffusion equation with leading $\mathcal{C}^1$ coefficient

This section presents a two-dimensional stationary convection diffusion equation with coefficients that oscillates in both coordinate directions within a small region located in the center of the domain. The maximum magnitude of the leading coefficient is $11^2 = 121$, whereas it is constant

and equal to 1 outside of the center region. The fine-scale behavior is hard to grasp using standard polynomial basis functions, especially since the coefficients imply steep gradients of the solution. Optimal basis functions will hence be locally employed to improve the approximation quality in enriched global simulations.

In Section 6.2.1, the strong and weak formulation of the stationary convection diffusion equation is presented. The set of boundary conditions, as well as the trial and test spaces are introduced as well. In Section 6.2.2, the effect of parameter variation on the set of optimal shape functions is investigated. Finally, Section 6.2.3 presents the results of the global, enriched computation.

6.2.1 Problem formulation

Problem 14 presents the strong form of the PDE under study.

Problem 14: Stationary convection diffusion with oscillating coefficients.

Let $\Omega := [-6, 6]^2 \subset \mathbb{R}^2$, let $A_{1,1} : \mathbb{R}^2 \to \mathbb{R}$ and $A : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$A_{1,1}(x) = (5\cos(3\pi x_1) + 6)(5\cos(3\pi x_2) + 6), \tag{6.13}$$

for $x \in [-1, 1]^2$, and $A_{1,1}(x) = 1$ for $x \notin [-1, 1]^2$. Consider the coefficient $A : \mathbb{R}^2 \to \mathbb{R}$ with

$$A(x) := A_{1,1}(x) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{6.14}$$

and the partial differential operator

$$\mathcal{L} : \mathcal{C}^2(\Omega) \to \mathcal{C}^0(\Omega), \quad v \mapsto \mathcal{L}\, v := -\operatorname{div}(A\nabla u) + \nabla A_{1,1} \cdot \nabla u. \tag{6.15}$$

Moreover, let

$$\begin{aligned} f : \mathbb{R}^2 \to \mathbb{R}, \quad x = (x_1, x_2) \mapsto f(x) := x_1 \\ g : \mathbb{R}^2 \to \mathbb{R}, \quad x = (x_1, x_2) \mapsto g(x) := x_2 \end{aligned} \tag{6.16}$$

Find a function $u \in \mathcal{C}^2(\Omega)$ satisfying

$$\begin{aligned} -\mathcal{L}\, u(x) &= f(x), \quad \text{in } \Omega \\ u(x) &= g(x), \quad \text{on } \partial\Omega. \end{aligned} \tag{6.17}$$

The coefficient $A_{1,1}$ is shown in Figure 6.16. Note that the gradient of the coefficient $A_{1,1}$ reads

$$\nabla A_{1,1}(x) = -15\pi \begin{bmatrix} \sin(3\pi x_1)(5\cos(3\pi x_2) + 6) \\ (5\cos(3\pi x_1) + 6)\sin(3\pi x_2) \end{bmatrix}. \tag{6.18}$$

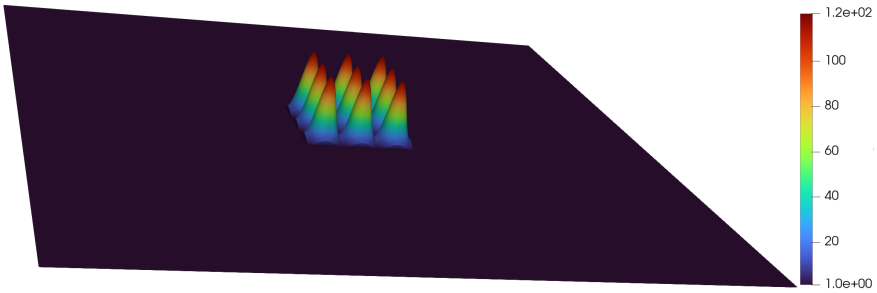


Figure 6.16: The leading coefficient $A_{1,1}$ of the PDE presented in Problem 14, which oscillates in the center of the domain.

The trial and test spaces are defined as

$$\begin{aligned} \mathrm{V}^{\text{trial}}(\Omega) &:= \{u \in \mathrm{H}^1(\Omega) : \operatorname{tr}(u) = g \text{ on } \partial\Omega\} \\ \mathrm{V}^{\text{test}}(\Omega) &:= \{v \in \mathrm{H}^1(\Omega) : \operatorname{tr}(v) = 0 \text{ on } \partial\Omega\}, \end{aligned} \tag{6.19}$$

and hence the weak formulation reads as follows.

Problem 15: Weak stationary convection diffusion with oscillating coffi-cients.

Let $\Omega := [-6, 6]^2 \subset \mathbb{R}^2$, let $A_{1,1} : \mathbb{R}^2 \to \mathbb{R}$ and $A : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$A_{1,1}(x) = (5\cos(3\pi x_1) + 6)(5\cos(3\pi x_2) + 6), \tag{6.20}$$

for $x \in [-1, 1]^2$, and $A_{1,1}(x) = 1$ for $x \notin [-1, 1]^2$. Consider the coefficient $A : \mathbb{R}^2 \to \mathbb{R}$ with

$$A(x) := A_{1,1}(x) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{6.21}$$

as well as

$$\begin{aligned} f : \mathbb{R}^2 \to \mathbb{R}, \quad x = (x_1, x_2) \mapsto f(x) := x_1 \\ g : \mathbb{R}^2 \to \mathbb{R}, \quad x = (x_1, x_2) \mapsto g(x) := x_2 \end{aligned} \tag{6.22}$$

Define the bilinear form $\mathbf{a} : \mathrm{V}^{\text{trial}}(\Omega) \times \mathrm{V}^{\text{test}}(\Omega) \to \mathbb{R}$ and linear functional $\ell : \mathrm{V}^{\text{test}}(\Omega) \to \mathbb{R}$,

$$\begin{aligned} \mathbf{a}[u, v] &:= \int_\Omega A_{1,1} \nabla u \cdot \nabla v + v \nabla u \cdot \nabla A_{1,1} \, dx, \\ \ell(v) &:= \int_\Omega f v \, dx. \end{aligned} \tag{6.23}$$

Find a function $u \in \mathrm{V}^{\text{trial}}(\Omega)$ satisfying

$$\mathbf{a}[u, v] = \ell(v), \quad \forall v \in \mathrm{V}^{\text{trial}}(\Omega). \tag{6.24}$$

From (6.18) it follows that

$$\nabla A_{1,1}(x) = -15\pi \begin{bmatrix} \sin(3\pi x_1)(5\cos(3\pi x_2)+6) \\ (5\cos(3\pi x_1)+6)\sin(3\pi x_2) \end{bmatrix} \geqslant \begin{bmatrix} -165\pi \\ -165\pi \end{bmatrix}, \tag{6.25}$$

and hence

$$\begin{aligned} \mathbf{a}[u,u] &= \int_\Omega \underbrace{A_{1,1}}_{\geqslant 1} \nabla u \cdot \nabla u + u \nabla u \cdot \nabla A_{1,1}\, dx \\ &\geqslant \int_\Omega \nabla u \cdot \nabla u\, dx - 165\pi \int_\Omega u\partial_{x_1} u + u\partial_{x_2} u\, dx. \end{aligned} \tag{6.26}$$

An application of the Poincaré-Friedrichs inequality (Theorem 2.6) on the first term yields

$$\begin{aligned} \int_\Omega \nabla u \cdot \nabla u\, dx &= \|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^2} \\ &= \frac{1}{2}\left[\|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^2} + \|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^2}\right] \\ &= \frac{1}{2}\left[\|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^2} + \frac{1}{{\mathrm{C_{PF}}}^2}\|u\|^2_{\mathrm{L}^2(\Omega)}\right] \\ &\geqslant \min\left\{\frac{1}{2}, \frac{1}{2\,{\mathrm{C_{PF}}}^2}\right\}\left[\|\nabla u\|^2_{[\mathrm{L}^2(\Omega)]^2} + \|u\|^2_{\mathrm{L}^2(\Omega)}\right] \\ &\geqslant \min\left\{\frac{1}{2}, \frac{1}{2\,{\mathrm{C_{PF}}}^2}\right\}\|u\|^2_{\mathrm{H}^1(\Omega)}. \end{aligned} \tag{6.27}$$

For the other terms appearing on the right-hand side of (6.26), integration by parts reads

$$\int_\Omega u\partial_{x_j} u\, dx = \int_{\partial\,\Omega} \vec{n}_j u^2\, ds - \int_\Omega \partial_{x_j} uu\, dx, \quad j = 1,2, \tag{6.28}$$

that is

$$\int_\Omega u\partial_{x_j} u\, dx = \frac{1}{2}\int_{\partial\,\Omega} \vec{n}_j u^2\, ds \overset{u=g \text{ on } \partial\,\Omega}{=} \frac{1}{2}\int_{\partial\,\Omega} \vec{n}_j g^2\, ds, \quad j = 1,2. \tag{6.29}$$

Using the boundary data $g(x) = x_2$, (6.29) for $j = 1$ reads

$$\begin{aligned} \int_\Omega u\partial_{x_1} u\, dx &= \frac{1}{2}\int_{\partial\,\Omega} \vec{n}_1 x_2^2\, dS \\ &= \underbrace{\frac{1}{2}\int_{-6}^{6} -1x_2^2\, dx_2 + \frac{1}{2}\int_{-6}^{6} 1x_2^2\, dx_2}_{=0} \\ &\qquad + \underbrace{\frac{1}{2}\int_{-6}^{6} 0\cdot(-6)^2\, dx_2}_{=0} + \underbrace{\frac{1}{2}\int_{-6}^{6} 0\cdot 6^2\, dx_2}_{=0} \\ &= 0, \end{aligned} \tag{6.30}$$

where the boundary integral was split into integrals over the left, right, bottom and top side of Ω.

Similarly, for $j = 2$ it holds that

$$\begin{aligned}\int_\Omega u\partial_{x_2} u\,dx &= \frac{1}{2}\int_{\partial\Omega} \vec{n}_2 x_2^2\,dS \\ &= \underbrace{\frac{1}{2}\int_{-6}^{6} 0x_2^2\,dx_2}_{=0} + \underbrace{\frac{1}{2}\int_{-6}^{6} 0x_2^2\,dx_2}_{=0} \\ &\quad + \underbrace{\frac{1}{2}\int_{-6}^{6} -1\cdot(-6)^2\,dx_2 + \frac{1}{2}\int_{-6}^{6} 1\cdot 6^2\,dx_2}_{=0} \\ &= 0.\end{aligned} \tag{6.31}$$

Inserting (6.27), (6.30) and (6.31) into (6.26) yields

$$\mathbf{a}[u,u] \quad \geqslant \quad \min\left\{\frac{1}{2}, \frac{1}{2\,{C_{\text{PF}}}^2}\right\} \|u\|^2_{\text{H}^1(\Omega)}, \tag{6.32}$$

proving the ellipticity the bilinear form $\mathbf{a}[\cdot,\cdot]$. Due to Theorem 2.4 (Lax and Milgram), Problem 15 has a unique weak solution.

6.2.2 Influence of the boundary data

In the following, the impact of different choices of boundary data on the eigenvalues is investigated. In the following, the results obtained from using the boundary hats approach, the B-Spline approach, and the Fourier approach are presented. Afterwards, conclusions drawn from the full investigation of the benchmark problem at hand are presented.

Boundary hats

The following tables 6.24 - 6.28 present key numbers obtained from the computation of optimal shape functions using boundary hats as boundary data in the sampling problem. The boundary level ℓ is varied, and pd denotes the polynomial degree(s) of the boundary hats that are used.

Figure 6.17 presents a visualization of the development of the two dominant eigenvalues in

ℓ	pd	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	0	12	$1.303\cdot10^{-1}$	$1.303\cdot10^{-1}$	2	4	11	11
3	0	28	$1.845\cdot10^{-1}$	$1.845\cdot10^{-1}$	2	6	12	23
4	0	60	$2.175\cdot10^{-1}$	$2.175\cdot10^{-1}$	2	6	12	35
5	0	123	$2.362\cdot10^{-1}$	$2.355\cdot10^{-1}$	2	6	14	55
6	0	252	$2.448\cdot10^{-1}$	$2.448\cdot10^{-1}$	2	6	14	64
7	0	508	$2.492\cdot10^{-1}$	$2.492\cdot10^{-1}$	2	6	14	65
8	0	1020	$2.516\cdot10^{-1}$	$2.516\cdot10^{-1}$	2	6	14	72

Table 6.24: Key numbers obtained from the computation of the optimal shape functions using constant boundary hats on various boundary levels.

ℓ	pd	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	1	24	$2.366 \cdot 10^{-1}$	$6.016 \cdot 10^{-2}$	1	4	11	11
3	1	56	$2.397 \cdot 10^{-1}$	$7.337 \cdot 10^{-2}$	1	5	12	23
4	1	120	$2.438 \cdot 10^{-1}$	$1.039 \cdot 10^{-1}$	2	5	12	36
5	1	248	$2.474 \cdot 10^{-1}$	$1.301 \cdot 10^{-1}$	2	6	13	55
6	1	504	$2.497 \cdot 10^{-1}$	$1.511 \cdot 10^{-1}$	2	6	14	64
7	1	1016	$2.509 \cdot 10^{-1}$	$1.687 \cdot 10^{-1}$	2	6	14	65
8	1	2040	$2.517 \cdot 10^{-1}$	$1.893 \cdot 10^{-1}$	2	6	14	72

Table 6.25: Key numbers obtained from the computation of the optimal shape functions using linear boundary hats in x_1 and x_2 direction on various boundary levels.

ℓ	pd	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	2	24	$1.94 \cdot 10^{-1}$	$5.695 \cdot 10^{-2}$	1	4	11	12
3	2	56	$2.02 \cdot 10^{-1}$	$6.027 \cdot 10^{-2}$	1	4	12	23
4	2	120	$2.143 \cdot 10^{-1}$	$6.473 \cdot 10^{-2}$	1	5	12	36
5	2	248	$2.278 \cdot 10^{-1}$	$8.746 \cdot 10^{-2}$	1	5	13	55
6	2	504	$2.387 \cdot 10^{-1}$	$1.116 \cdot 10^{-1}$	2	6	14	64
7	2	1016	$2.461 \cdot 10^{-1}$	$1.346 \cdot 10^{-1}$	2	6	14	64
8	2	2040	$2.513 \cdot 10^{-1}$	$1.6 \cdot 10^{-1}$	2	6	14	71

Table 6.26: Key numbers obtained from the computation of the optimal shape functions using quadratic boundary hats in x_1 and x_2 direction on various boundary levels.

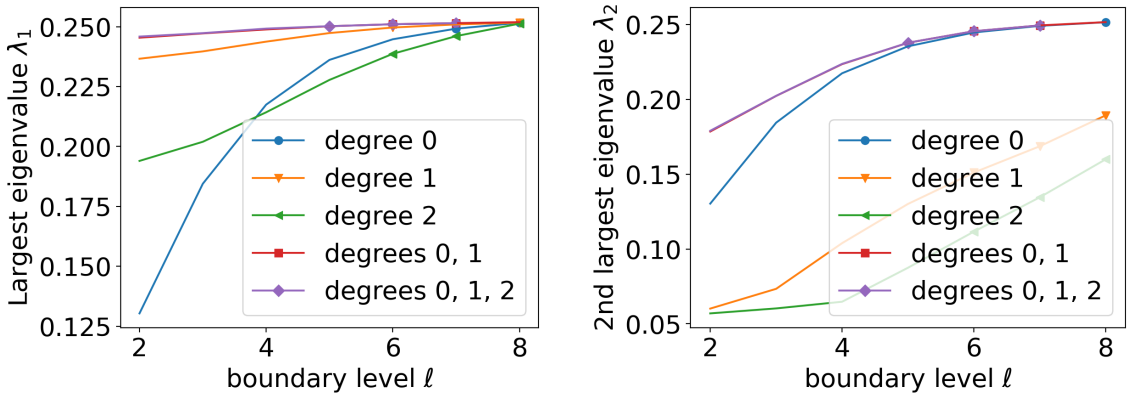
ℓ	pd	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	0, 1	36	$2.454 \cdot 10^{-1}$	$1.785 \cdot 10^{-1}$	2	6	12	15
3	0, 1	84	$2.472 \cdot 10^{-1}$	$2.023 \cdot 10^{-1}$	2	6	12	23
4	0, 1	180	$2.488 \cdot 10^{-1}$	$2.236 \cdot 10^{-1}$	2	6	14	36
5	0, 1	372	$2.502 \cdot 10^{-1}$	$2.379 \cdot 10^{-1}$	2	6	14	55
6	0, 1	756	$2.51 \cdot 10^{-1}$	$2.456 \cdot 10^{-1}$	2	6	14	66
7	0, 1	1524	$2.515 \cdot 10^{-1}$	$2.495 \cdot 10^{-1}$	2	6	14	67
8	0, 1	3060	$2.519 \cdot 10^{-1}$	$2.517 \cdot 10^{-1}$	2	7	14	72

Table 6.27: Key numbers obtained from the computation of the optimal shape functions using constant and linear boundary hats in x_1 and x_2 direction on various boundary levels.

terms of the boundary level. It is visible, that the largest eigenvalue λ_1 of the computations for polynomial degrees $\{0\}$, $\{1\}$, $\{2\}$, $\{0,1\}$ and $\{0,1,2\}$ approaches approximately the same value when increasing the boundary level. For a fixed boundary level, the optimal shape functions from $pd = \{0,1\}$ and $pd = \{0,1,2\}$ yield the largest values. However, in terms of degrees of freedom, also purely linear and purely constant boundary hats lead to large values of λ_1. Approximately 500 degrees of freedom are sufficient in all of these cases. Large values of the second eigenvalue λ_2 cannot be obtained in all cases. Only the cases $pd = \{0\}$, $pd = \{0,1\}$ and $pd = \{0,1,2\}$ give large values for 500 to 1000 degrees of freedom.

ℓ	pd	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	0, 1, 2	60	$2.459 \cdot 10^{-1}$	$1.79 \cdot 10^{-1}$	2	6	12	15
3	0, 1, 2	140	$2.473 \cdot 10^{-1}$	$2.024 \cdot 10^{-1}$	2	7	13	23
4	0, 1, 2	300	$2.492 \cdot 10^{-1}$	$2.238 \cdot 10^{-1}$	2	7	14	36
5	0, 1, 2	620	$2.502 \cdot 10^{-1}$	$2.379 \cdot 10^{-1}$	2	7	14	55
6	0, 1, 2	1260	$2.51 \cdot 10^{-1}$	$2.456 \cdot 10^{-1}$	2	6	14	66
7	0, 1, 2	2540	$2.515 \cdot 10^{-1}$	$2.495 \cdot 10^{-1}$	2	6	14	67

Table 6.28: Key numbers obtained from the computation of the optimal shape functions using constant, linear and quadratic boundary hats in x_1 and x_2 direction on various boundary levels.



(a) (b)

Figure 6.17: Largest eigenvalue λ_1 (a) and second largest eigenvalue λ_2 (b) for increasing boundary levels ℓ obtained from the computations for boundary hats of various polynomial degrees. Markers are plotted every 500 degrees of freedom.

Increasing the boundary level shows for all polynomial degrees, that the PDE under study seems to yield two very large and six or seven additional large eigenvalues. The largest eigenvalue stabilizes at around 0.25, and whenever constant boundary hats are included in the computation, the second largest eigenvalues are of around the same size. The second eigenvalue variates much more for the different polynomial degree(s) considered, and it seems that constant boundary hats have the strongest influence on the dominant eigenvalues.

In order to keep the degrees of freedom, i.e. the number of sampling problems to be solved, as small possible, a computation of the optimal shape functions using constant boundary hats on boundary level 8 seems to produce the most promising results. This way, 1020 sampling problems have to be solved.

B-Splines

In this subsection, B-Splines are used as boundary data in the sampling problem. The B-Splines under study are quadratic. Cubic B-Splines have been investigated in Section 6.1, but did not show advantages over quadratic B-Splines and will hence not be considered anymore. Inner knots

are repeated in order to ensure a maximum number of continuous B-Splines. The four ways of generating B-Splines, which were already used in Section 6.1, were also used for this benchmark problem: B-Splines can be constructed only in x_1, or in both coordinate directions, and corner splines can eventually be included. Key numbers from the computations using quadratic B-Splines are presented in tables 6.29 to 6.32, where n denotes the number of inner knots without repetitions.

The results of the construction considering B-Splines only in x_1 direction and without corner

n	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
3	15	$1.179 \cdot 10^{-1}$	$4.38 \cdot 10^{-2}$	1	4	7	7
4	20	$1.373 \cdot 10^{-1}$	$5.787 \cdot 10^{-2}$	1	4	8	10
5	24	$1.373 \cdot 10^{-1}$	$5.79 \cdot 10^{-2}$	1	4	8	11
6	28	$1.373 \cdot 10^{-1}$	$5.791 \cdot 10^{-2}$	1	4	8	13
7	32	$1.373 \cdot 10^{-1}$	$5.791 \cdot 10^{-2}$	1	4	8	14
8	36	$1.373 \cdot 10^{-1}$	$5.791 \cdot 10^{-2}$	1	4	8	15
9	40	$1.373 \cdot 10^{-1}$	$5.791 \cdot 10^{-2}$	1	4	8	16

Table 6.29: Key numbers obtained from the computation of the optimal shape functions using B-Splines in x_1 direction, not including corner splines.

n	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
3	24	$2.439 \cdot 10^{-1}$	$2.421 \cdot 10^{-1}$	2	6	11	11
4	28	$2.44 \cdot 10^{-1}$	$2.424 \cdot 10^{-1}$	2	6	10	12
5	32	$2.44 \cdot 10^{-1}$	$2.425 \cdot 10^{-1}$	2	6	11	14
6	36	$2.44 \cdot 10^{-1}$	$2.426 \cdot 10^{-1}$	2	6	12	15
7	40	$2.44 \cdot 10^{-1}$	$2.427 \cdot 10^{-1}$	2	6	11	16
8	44	$2.44 \cdot 10^{-1}$	$2.427 \cdot 10^{-1}$	2	6	11	17
9	48	$2.44 \cdot 10^{-1}$	$2.427 \cdot 10^{-1}$	2	6	11	19

Table 6.30: Key numbers obtained from the computation of the optimal shape functions using B-Splines in x_1 direction, including corner splines.

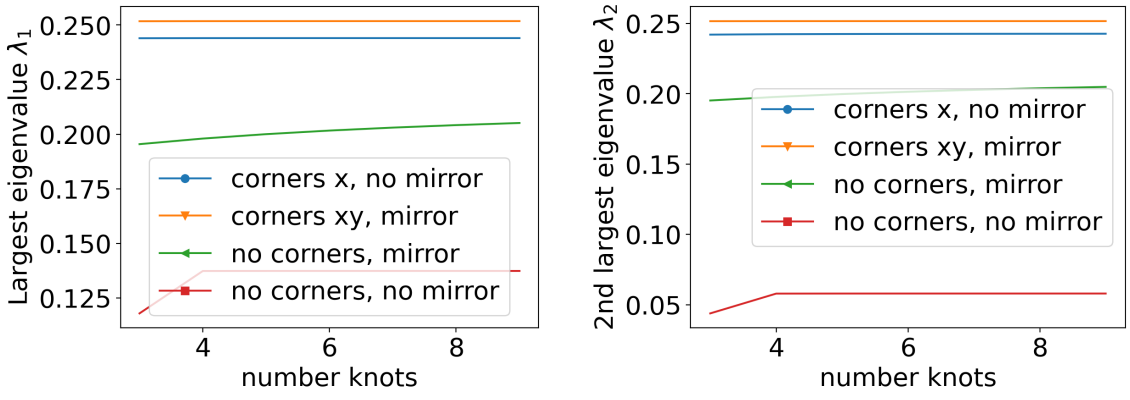
n	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
3	32	$1.954 \cdot 10^{-1}$	$1.952 \cdot 10^{-1}$	2	6	12	15
4	40	$1.98 \cdot 10^{-1}$	$1.978 \cdot 10^{-1}$	2	6	14	18
5	48	$2 \cdot 10^{-1}$	$1.998 \cdot 10^{-1}$	2	6	14	19
6	56	$2.016 \cdot 10^{-1}$	$2.015 \cdot 10^{-1}$	2	6	14	23
7	64	$2.03 \cdot 10^{-1}$	$2.028 \cdot 10^{-1}$	2	6	14	25
8	72	$2.041 \cdot 10^{-1}$	$2.039 \cdot 10^{-1}$	2	6	14	27
9	80	$2.051 \cdot 10^{-1}$	$2.049 \cdot 10^{-1}$	2	6	14	29

Table 6.31: Key numbers obtained from the computation of the optimal shape functions using B-Splines in x_1 and x_2 direction, not including corner splines.

n	dim	λ_1	λ_2	n_{-1}	n_{-2}	n_{-4}	n_{-8}
3	96	$2.517 \cdot 10^{-1}$	$2.517 \cdot 10^{-1}$	2	6	12	15
4	112	$2.518 \cdot 10^{-1}$	$2.517 \cdot 10^{-1}$	2	6	14	18
5	128	$2.518 \cdot 10^{-1}$	$2.517 \cdot 10^{-1}$	2	6	14	19
6	144	$2.518 \cdot 10^{-1}$	$2.517 \cdot 10^{-1}$	2	6	14	22
7	160	$2.518 \cdot 10^{-1}$	$2.517 \cdot 10^{-1}$	2	6	14	23
8	176	$2.518 \cdot 10^{-1}$	$2.517 \cdot 10^{-1}$	2	6	14	26
9	192	$2.518 \cdot 10^{-1}$	$2.517 \cdot 10^{-1}$	2	6	14	29

Table 6.32: Key numbers obtained from the computation of the optimal shape functions using B-Splines in x_1 and x_2 direction, including corner splines.

splines, which is presented in Table 6.29, show that there tends to be only one dominant eigenvalue with a size of about 0.137. The second-largest eigenvalue is very small in this case. The other variants of construction presented in Tables 6.30 to 6.32 lead to two very large and six more large eigenvalues. Both dominant eigenvalues are of approximately the same size and the construction based on B-Splines in both coordinate directions including corner splines - as also seen before in the previous experiment - produces dominant eigenvalues of largest values. Since the dimension of the discrete harmonic spaces, i.e. the number of sampling problems to be solved, is relatively small, the construction using B-Splines in both coordinate directions and including corner splines is favorable. Figure 6.18 presents the development of the dominant eigenvalues for increasing numbers of inner knots without repetitions for all four variants of construction.



(a) (b)

Figure 6.18: Largest eigenvalue λ_1 (a) and second largest eigenvalue λ_2 (b) obtained from the computations using quadratic B-Splines for increasing numbers of inner knots (without repetitions). All computations use less than 200 degrees of freedom.

Oscillating trigonometric functions

This subsection considers boundary data defined by the oscillatory Fourier-type basis functions $\mathfrak{C}$ from Section 5.2 and the maximum number of oscillations, n_{osc}, is increased subsequently.

Table 6.33 shows key numbers obtained from the corresponding computations. No very large

n_{osc}	dim	λ_1	$\lambda_2 = \lambda_3$	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	26	$7.217 \cdot 10^{-2}$	$2.097 \cdot 10^{-2}$	0	3	7	8
5	56	$7.385 \cdot 10^{-2}$	$2.101 \cdot 10^{-2}$	0	3	9	18
10	106	$7.405 \cdot 10^{-2}$	$2.105 \cdot 10^{-2}$	0	3	9	18
15	156	$7.412 \cdot 10^{-2}$	$2.107 \cdot 10^{-2}$	0	3	9	18
20	206	$7.414 \cdot 10^{-2}$	$2.107 \cdot 10^{-2}$	0	3	9	18
25	256	$7.415 \cdot 10^{-2}$	$2.108 \cdot 10^{-2}$	0	3	9	18
30	306	$7.415 \cdot 10^{-2}$	$2.108 \cdot 10^{-2}$	0	3	10	18
35	356	$7.416 \cdot 10^{-2}$	$2.108 \cdot 10^{-2}$	0	3	10	18
40	406	$7.416 \cdot 10^{-2}$	$2.109 \cdot 10^{-2}$	0	3	10	18
45	456	$7.416 \cdot 10^{-2}$	$2.109 \cdot 10^{-2}$	0	3	10	18
50	506	$7.416 \cdot 10^{-2}$	$2.109 \cdot 10^{-2}$	0	3	10	18
55	556	$7.416 \cdot 10^{-2}$	$2.109 \cdot 10^{-2}$	0	3	10	18
60	606	$7.416 \cdot 10^{-2}$	$2.109 \cdot 10^{-2}$	0	3	10	18
65	656	$7.416 \cdot 10^{-2}$	$2.109 \cdot 10^{-2}$	0	3	10	18
70	706	$7.416 \cdot 10^{-2}$	$2.109 \cdot 10^{-2}$	0	3	10	18
75	756	$7.416 \cdot 10^{-2}$	$2.109 \cdot 10^{-2}$	0	3	10	18
80	806	$7.416 \cdot 10^{-2}$	$2.109 \cdot 10^{-2}$	0	3	10	18
85	856	$7.416 \cdot 10^{-2}$	$2.109 \cdot 10^{-2}$	0	3	10	18

Table 6.33: Key numbers obtained from the computation of the optimal shape functions using Fourier-type basis functions in x_1 and x_2 direction for increasing numbers of oscillations.

eigenvalues are captured, and the largest eigenvalue is of size ≈ 0.07, even when using a large number of boundary data functions. As already seen in the previous benchmark problem, it is expected that the optimal shape functions based on Fourier-type boundary data will perform poor, meaning that many enrichments will be needed to decrease the error in a global computation to a reasonable level. The results from this subsection are surprising in the sense that the coefficient indeed oscillates, which in turn may indicate that also the homogeneous solution oscillates.

Conclusive remarks

The conclusion of this section are similar to the ones of the first benchmark problem. Using constant boundary hats on the highest feasible boundary level 8 promise the best results, with two very large, dominant eigenvalues. Furthermore, six to seven large eigenvalues were captured. The corresponding optimal shape functions are linear combinations of 1020 solutions to the sampling problem. When using B-Spline boundary data, the importance of considering corner splines was shown. Also, B-splines in both coordinate directions should be employed, even though quite promising values of the two dominant eigenvalues could also be computed from B-splines only in x_1 direction and including the corner splines. Boundary data consisting of oscillatory, Fourier-type functions was not able to capture any dominant eigenvalues, and the performance of the corresponding optimal shape functions in global computations is expected to be relatively poor.

6.2.3 Discussion of global errors

The procedure of the first benchmark problem is repeated. First, a numerical reference solution is identified. Afterwards, the various sets of optimal shape functions from Section 6.2.2 are used as enrichments and the enriched solutions are compared against the reference solution, which will help to (in)validate the claims previously made.

The bounding box is chosen to be $[-6, 10]^2$ and the reference solution is identified as in the first benchmark problem. Unenriched solutions u_h are computed for various discretization levels and corresponding patch sizes h, leading to discrete values of the function

$$E : \mathbb{R}_+ \to \mathbb{R}, \quad h \mapsto E(h) := a_\Omega[u_h, u_h]^{\frac{1}{2}}, \tag{6.33}$$

The discrete approximation of the function, as well as the extrapolated limit for $h = 0$ are shown in Figure 6.19. The value at $h = 0.015625$, corresponding to a discretization on level 10, approxi-

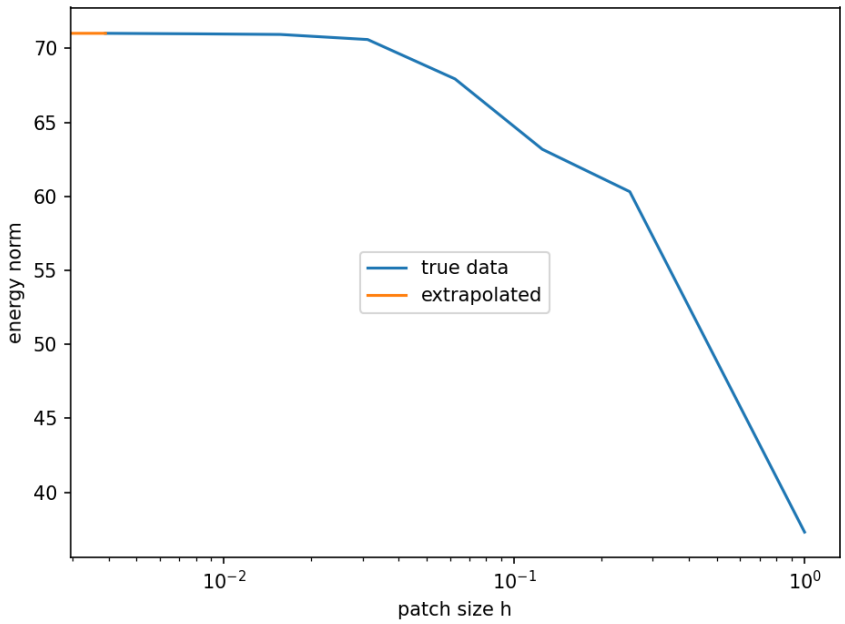
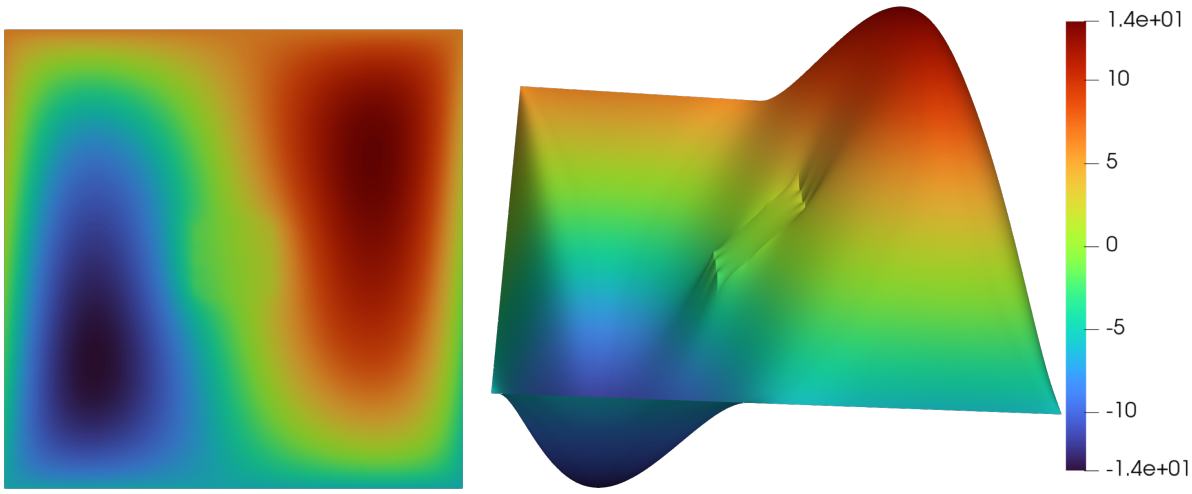


Figure 6.19: The energy of the solutions u_h for various values of h. The extrapolated limit value at $h = 0$ is also shown.

mately coincides with the extrapolated limit, and hence the corresponding level 10 approximation $u_{0.015625}$ will be used as a reference solution in the following.

In the following, the enriched solutions obtained from boundary hats, B-Splines and Fourier-type basis functions as boundary data in the sampling problem are compared to the reference solution. As already mentioned in the analysis of the first benchmark problem, additional local enrichments may lead to improvements of the local energy, but have adverse effects on other parts of the domain. Since the PUM minimizes the global (not the local) energy, enrichments may hence be discarded (cf. [Sch11]). The reference solution is shown in Figure 6.20. The domain is coarsely discretized on level $\iota = 3$ as shown in Figure 6.21. Particular solutions for the load $f(x) = x_1$ and global essential boundary conditions $g(x) = x_2$ have been computed for all coarse patches, resp. all coarse boundary patches individually and will be used as local enrichments on the corresponding



(a) (b)

Figure 6.20: (a) Reference solution $u_{0.015625}$ for load $f(x) = x_1$ and global essential boundary condition $g(x) = x_2$. In (b), the warped state is shown.

patches. Optimal shape functions are used as enrichments in the marked center of the domain. From Figures 6.16 and 6.21 it is clearly visible, that the behavior of the coefficient is not resolved

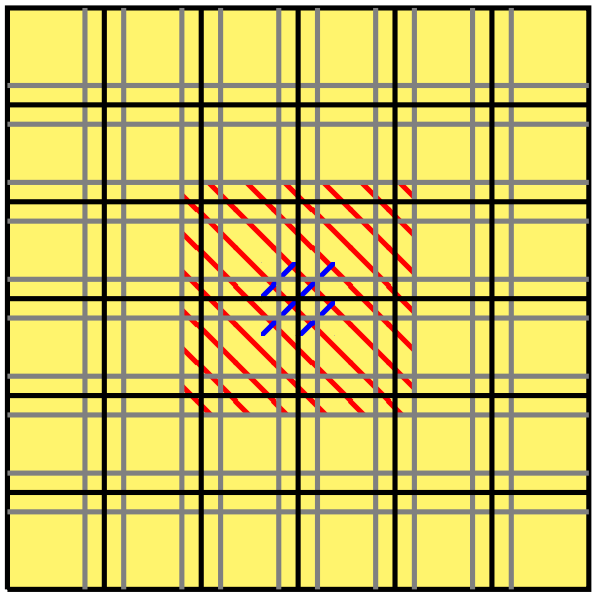


Figure 6.21: Coarse global discretization. In the blue marked region, the coefficient A oscillates, and outside it is equal to 1. Local particular solutions for load and essential boundary conditions are used on the corresponding patches, and the red patches will be further enriched using optimal shape functions. The red patches cover the patch ω.

by the coarse patches, leading to an expected poor performance of the unenriched solution. The global energy error of the unenriched solution is approximately 33%.

In all upcoming computations, the optimal basis functions were computed on the oversampled patch $\omega^+ = [-4, 4]^2$ and used as enrichment in the four center patches, which are marked in red

in Figure 6.21 and cover $\omega = [-2, 2]^2$.

Boundary hats

The following tables 6.34 to 6.38 show the relative energy error for enrichments based on constant, linear and quadratic boundary hats, as well as combinations thereof.

ℓ	pd	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
2	0	33.14%	24.41%	6.69%	4.87%	4.73%	4.51%	3.91%	3.91%
3	0	33.14%	24.36%	6.85%	4.86%	4.77%	4.73%	3.93%	3.90%
4	0	33.14%	24.36%	6.76%	4.88%	4.74%	4.74%	3.93%	3.90%
5	0	33.14%	8.07%	6.63%	4.91%	4.73%	4.67%	3.92%	3.90%
6	0	33.14%	24.44%	6.55%	4.92%	4.68%	4.73%	3.93%	3.90%
7	0	33.14%	24.46%	6.51%	6.15%	4.67%	4.73%	3.93%	3.90%
8	0	33.14%	24.47%	6.48%	6.11%	4.67%	4.73%	3.93%	3.90%

Table 6.34: Development of the relative energy error for increasing numbers of enrichments obtained from constant boundary hats.

ℓ	pd	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
2	1	33.14%	12.25%	7.88%	5.78%	5.74%	4.38%	3.90%	3.91%
3	1	33.14%	12.15%	6.83%	5.66%	5.63%	4.41%	3.93%	3.90%
4	1	33.14%	12.03%	6.67%	5.45%	5.44%	4.40%	3.93%	3.90%
5	1	33.14%	11.94%	6.54%	5.25%	5.25%	4.38%	3.93%	3.90%
6	1	33.14%	11.88%	6.48%	6.22%	5.09%	4.36%	3.93%	3.90%
7	1	33.14%	11.86%	6.44%	6.14%	4.96%	4.34%	3.93%	3.90%
8	1	33.14%	11.84%	6.44%	6.12%	4.84%	4.33%	3.93%	3.90%

Table 6.35: Development of the relative energy error for increasing numbers of enrichments obtained from linear boundary hats in x_1 and x_2.

ℓ	pd	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
2	2	33.14%	12.52%	8.35%	6.87%	6.63%	4.24%	3.91%	3.91%
3	2	33.14%	12.38%	8.11%	6.73%	6.38%	4.30%	3.93%	3.90%
4	2	33.14%	12.23%	8.16%	6.37%	6.03%	4.32%	3.93%	3.90%
5	2	33.14%	12.08%	7.26%	5.90%	5.68%	4.34%	3.93%	3.90%
6	2	33.14%	11.97%	6.85%	5.49%	5.39%	4.35%	3.93%	3.90%
7	2	33.14%	11.90%	6.59%	5.21%	5.18%	4.35%	3.93%	3.90%
8	2	33.14%	11.84%	6.42%	6.15%	5.01%	4.34%	3.93%	3.90%

Table 6.36: Development of the relative energy error for increasing numbers of enrichments obtained from quadratic boundary hats in x_1 and x_2.

ℓ	pd	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
2	0, 1	33.14%	27.32%	7.08%	4.14%	4.56%	4.67%	3.93%	3.91%
3	0, 1	33.14%	27.35%	6.88%	4.28%	4.48%	4.65%	3.93%	3.90%
4	0, 1	33.14%	27.42%	6.70%	4.55%	4.54%	4.25%	3.93%	3.90%
5	0, 1	33.14%	27.48%	6.58%	4.75%	4.61%	4.28%	3.93%	3.90%
6	0, 1	33.14%	27.52%	6.53%	6.26%	4.65%	4.33%	3.93%	3.90%
7	0, 1	33.14%	27.53%	6.50%	6.16%	4.68%	4.70%	3.93%	3.90%
8	0, 1	33.14%	25.11%	6.48%	6.11%	4.69%	4.45%	3.93%	3.90%

Table 6.37: Development of the relative energy error for increasing numbers of enrichments obtained from constant and linear boundary hats in x_1 and x_2.

ℓ	pd	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
2	0, 1, 2	33.14%	27.28%	7.09%	4.12%	4.52%	4.64%	3.94%	3.91%
3	0, 1, 2	33.14%	27.34%	6.87%	4.28%	4.47%	4.40%	3.93%	3.90%
4	0, 1, 2	33.14%	27.37%	6.70%	4.59%	4.43%	4.34%	3.93%	3.90%
5	0, 1, 2	33.14%	27.47%	6.58%	4.75%	4.60%	4.56%	3.93%	3.90%
6	0, 1, 2	33.14%	27.52%	6.53%	6.26%	4.65%	4.30%	3.93%	3.90%
7	0, 1, 2	33.14%	27.54%	6.50%	6.16%	4.68%	4.33%	3.93%	3.90%

Table 6.38: Development of the relative energy error for increasing numbers of enrichments obtained from constant, linear and quadratic boundary hats in x_1 and x_2.

All sets of enrichments can be used to reduce the initial error of 33.14% to about 3.9%. It can be seen that the first two enrichments, corresponding to the two dominant eigenvalues, have the largest impact on the error and using those two enrichments leads to relative errors of less than 8.4% in all cases and for all boundary levels. The performance of the first enrichment alone however varies significantly for the different types of boundary hats used in their construction. Using ten enrichments reduces the relative error to less than 3.95% in all cases, and using further enrichments only marginally changes the error. Concluding, the first two enrichments corresponding to the dominant eigenvalues have the strongest potential to reduce the errors, and they should both be employed. The boundary level does not need to be chosen very high in order to produce promising results. In Figure 6.22, the distribution of the energy error for various numbers of enrichments is shown visually. It can clearly be observed, that the error in ω is reduced. Furthermore, the remaining part of the error for 20 enrichments is mostly located outside of ω and with highest absolute values in the overlap region of coarse patches. The decay of the errors is shown in Figure 6.23

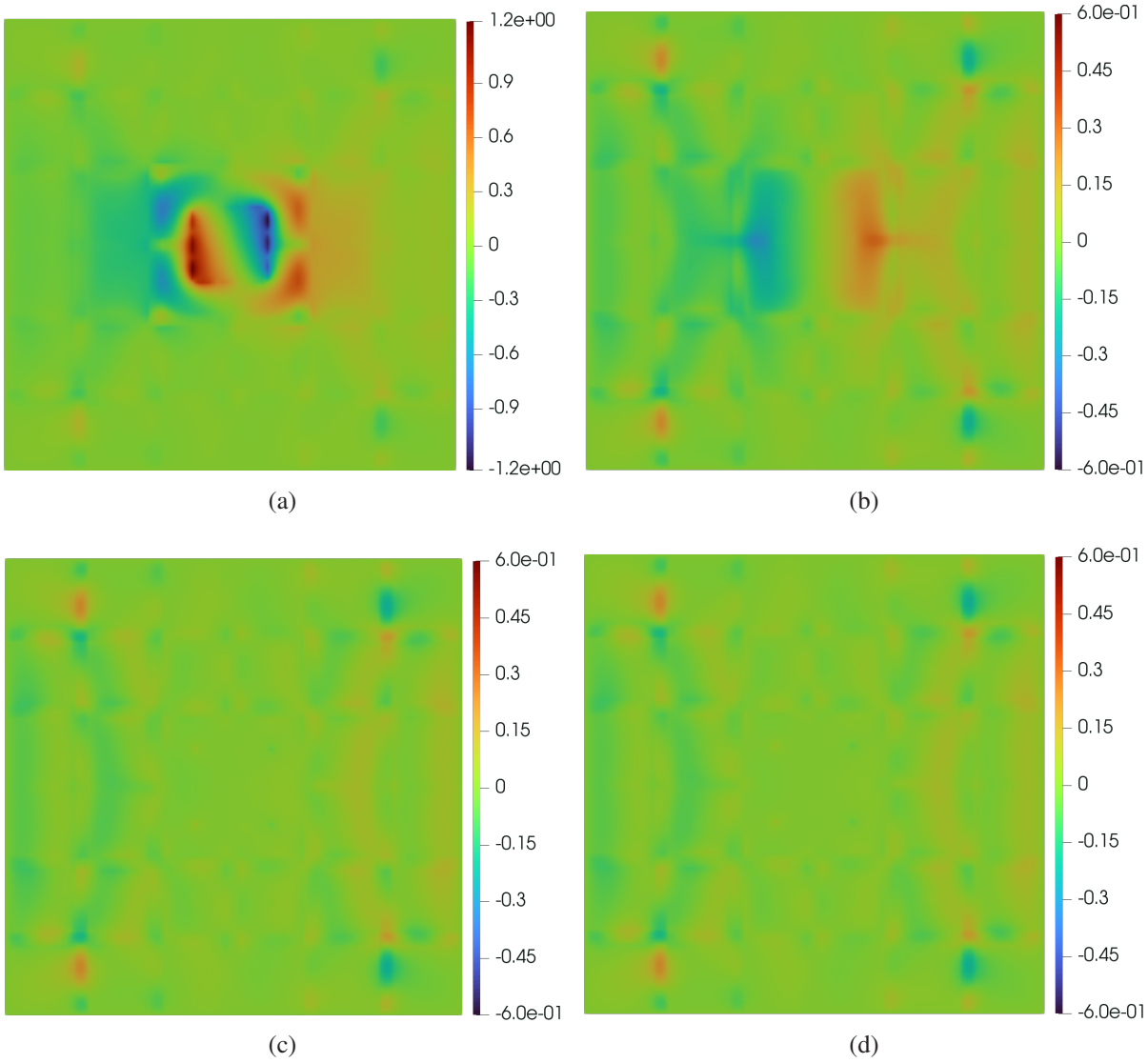


Figure 6.22: Difference between reference and enriched solutions, for (a) only particular solutions, (b) particular solutions and two enrichments, (c) particular solutions and ten enrichments, (d) particular solutions and eighteen enrichments. Enrichments were computed using boundary level 8 and polynomial degrees $pd = \{0, 1\}$. Note that the minimum and maximum shown on the scale of (a) is twice as large as in the scales of (b), (c) and (d).

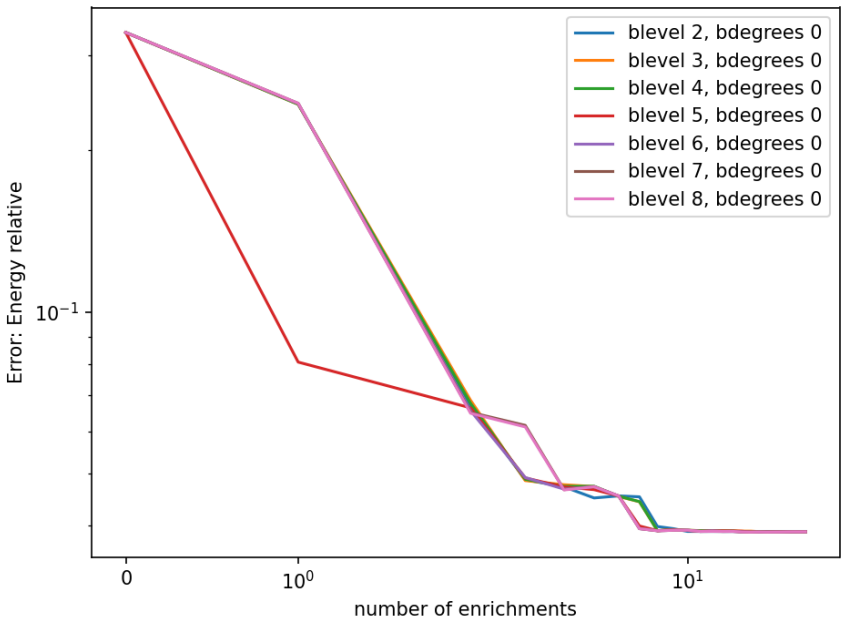


Figure 6.23: Relative energy error for increasing numbers of enrichments, constructed from boundary hats of different degrees and defined on various boundary levels.

B-Splines

In this section, the accuracy of the enriched solutions using enrichments constructed from B-Spline boundary data in the sampling problem is investigated. The results are shown in tables 6.39 - 6.42.

n	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
3	33.14%	29.77%	11.00%	9.03%	8.07%	5.23%	4.02%	4.02%
4	33.14%	30.20%	11.27%	9.88%	8.75%	5.51%	3.91%	3.91%
5	33.14%	30.20%	11.28%	9.87%	8.77%	5.51%	3.91%	3.91%
6	33.14%	30.20%	11.28%	9.87%	8.77%	5.51%	3.91%	3.89%
7	33.14%	30.20%	11.28%	9.87%	8.77%	5.52%	3.91%	3.90%
8	33.14%	30.20%	11.28%	9.87%	8.77%	5.52%	3.91%	3.90%
9	33.14%	30.20%	11.28%	9.87%	8.77%	5.52%	3.90%	3.90%

Table 6.39: Development of the relative energy error for increasing numbers of enrichments obtained from quadratic B-Splines in x_1 direction without corner splines.

n	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
3	33.14%	12.11%	6.76%	6.04%	5.35%	5.46%	3.91%	3.91%
4	33.14%	12.11%	6.76%	6.04%	5.38%	5.48%	3.92%	3.90%
5	33.14%	12.11%	6.76%	6.04%	5.38%	5.48%	3.92%	3.92%
6	33.14%	12.11%	6.76%	6.03%	5.38%	5.48%	3.92%	3.92%
7	33.14%	12.11%	6.76%	6.03%	5.38%	5.48%	3.92%	3.91%
8	33.14%	12.11%	6.76%	6.03%	5.38%	5.48%	3.92%	3.91%
9	33.14%	12.11%	6.76%	6.03%	5.38%	5.48%	3.92%	3.91%

Table 6.40: Development of the relative energy error for increasing numbers of enrichments obtained from quadratic B-Splines in x_1 direction with corner splines.

n	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
3	33.14%	28.06%	6.29%	4.61%	4.93%	4.26%	3.92%	3.90%
4	33.14%	28.05%	6.19%	4.58%	4.92%	4.25%	3.92%	3.89%
5	33.14%	28.01%	6.15%	4.58%	4.91%	4.25%	3.93%	3.90%
6	33.14%	27.99%	6.17%	4.60%	4.90%	4.25%	3.92%	3.90%
7	33.14%	27.97%	6.18%	4.61%	4.90%	4.25%	3.92%	3.90%
8	33.14%	27.96%	6.19%	4.62%	4.89%	4.26%	3.92%	3.90%
9	33.14%	27.95%	6.20%	4.62%	4.88%	4.26%	3.92%	3.90%

Table 6.41: Development of the relative energy error for increasing numbers of enrichments obtained from quadratic B-Splines in x_1 and x_2 direction without corner splines.

All sets of enrichments constructed from B-Spline boundary data are capable of reducing the initial error of 33.14% to $\approx 3.9\%$ in almost all cases. The only exceptions are enrichments constructed

n	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 20
3	33.14%	28.52%	6.49%	6.12%	4.59%	4.68%	3.94%	3.90%
4	33.14%	28.75%	6.48%	6.11%	4.58%	4.75%	3.92%	3.90%
5	33.14%	28.85%	6.48%	6.11%	4.58%	4.75%	3.94%	3.90%
6	33.14%	28.90%	6.48%	6.11%	4.57%	4.74%	3.93%	3.90%
7	33.14%	28.94%	6.48%	6.11%	4.57%	4.74%	3.93%	3.90%
8	33.14%	28.96%	6.48%	6.11%	4.57%	4.74%	3.93%	3.90%
9	33.14%	28.98%	6.48%	6.11%	4.57%	4.74%	3.93%	3.90%

Table 6.42: Development of the relative energy error for increasing numbers of enrichments obtained from quadratic B-Splines in x_1 and x_2 direction with corner splines.

from B-Splines only in x_1 direction, without corner splines, and for only 3 different inner knots, which perform slightly worse. In tables 6.40 - 6.42, the two enrichments corresponding to the dominant eigenvalues have the largest influence on the error and reduce the error to less than 7% in all cases. In the case of B-Splines only in x_1 direction and without corner splines, presented in table 6.39, the performance is slightly worse. Note that in this case only one dominant eigenvalue had been computed. In all cases, the relative error varies substantially when only using one enrichment corresponding to the largest eigenvalue. For 10 enrichments, the error is basically as small as it will get. Concluding, it can be said that any set of enrichments constructed from B-Spline boundary data performs well, and at least all enrichments corresponding to the dominant eigenvalues should be used. Note that weakly chosen boundary data in the sampling problem, such as only using B-Splines in one coordinate direction without corner splines, only decreases the speed of decay but will ultimately lead to promising results. The decay of the error for various numbers of enrichments is visualized in Figure 6.24, showing that the error drastically reduces inside of ω. The remaining error for 10 or 20 enrichments is mainly located outside of ω and the largest magnitudes are attained in the overlap regions of the coarse patches. Finally, Figure 6.25 graphically shows the decay of the relative errors.

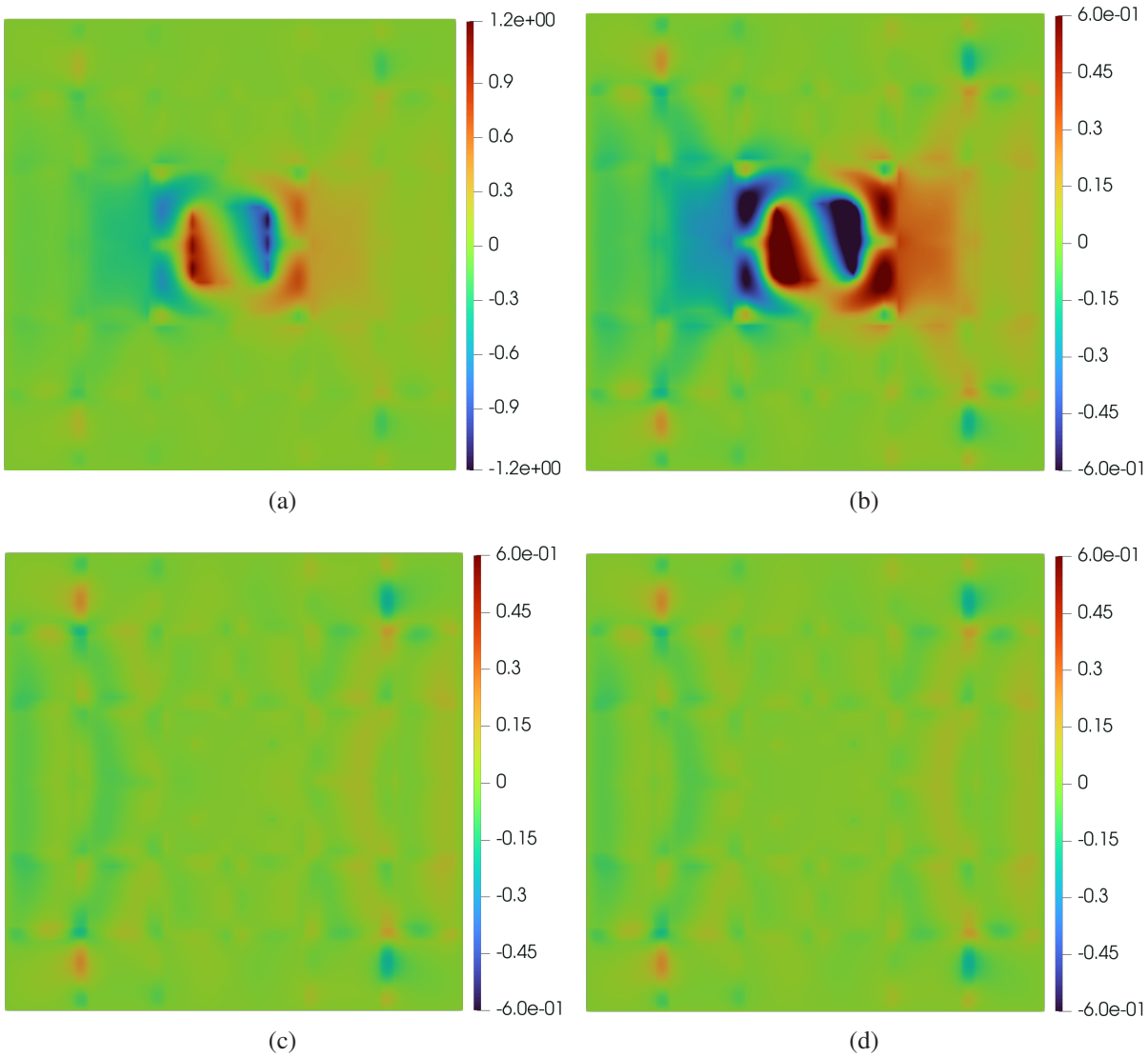


Figure 6.24: Difference between reference and enriched solutions, for (a) only particular solutions, (b) particular solutions and two enrichments, (c) particular solutions and ten enrichments, (d) particular solutions and eighteen enrichments. Enrichments were computed from quadratic B-Splines in x_1 and x_2 direction, including the corner splines, for 9 different inner knots. Note that the scale of the errors in (a) is ten times larger than in (b), (c) and (d).

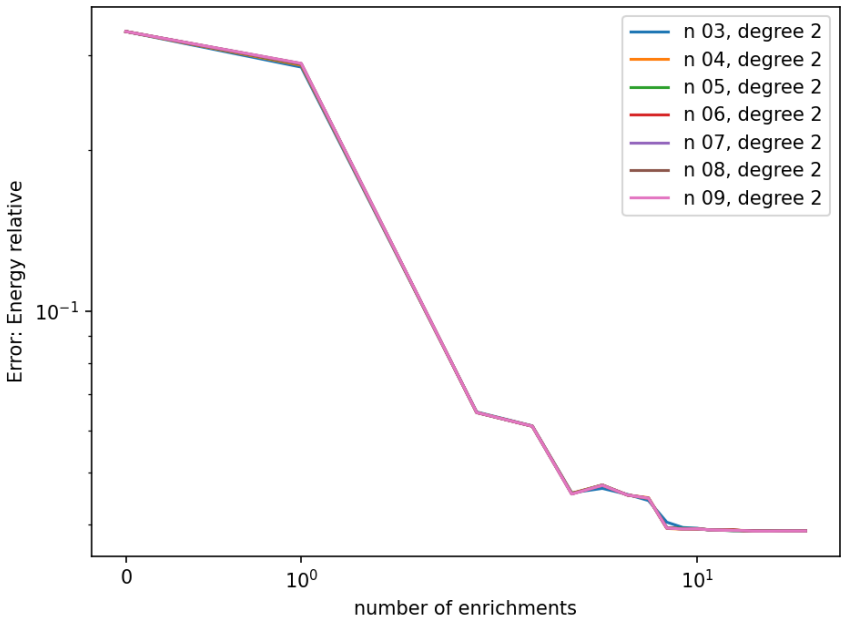


Figure 6.25: Relative energy error for increasing numbers of enrichments, constructed from quadratic B-Spline boundary data in x_1 and x_2, including corner splines, and defined for various numbers of inner knots (without repetitions).

Oscillating trigonometric functions

In table 6.43, the development of the relative energy error when using enrichments constructed from Fourier-type boundary data in the sampling problem is shown. As expected from the discussion in Section 6.2.2, which showed that no dominant eigenvalue and corresponding eigenfunctions could be captured, the decay of the error is slower as in the case of boundary hats or B-Splines. For 10 enrichments however, the error is less than 4% and the remaining eight enrichments hardly decrease the error any further. For 18 enrichments, the magnitude of the relative error is very similar to the previously investigated cases of boundary hats, resp. B-Spline boundary data. The foregoing discussion holds for all considered maximum numbers of oscillations. All sets of enrichments are

n_{osc}	nn 0	nn 1	nn 2	nn 3	nn 4	nn 5	nn 10	nn 18
2	33.14%	27.46%	25.65%	21.89%	14.38%	12.06%	-%	-%
5	33.14%	27.53%	25.53%	21.63%	13.62%	11.45%	3.95%	3.92%
10	33.14%	27.53%	25.43%	21.48%	13.08%	11.04%	3.95%	3.92%
15	33.14%	27.53%	25.39%	21.42%	12.89%	10.90%	3.95%	3.92%
20	33.14%	27.53%	25.37%	21.40%	12.81%	10.85%	3.95%	3.92%
25	33.14%	27.53%	25.36%	21.38%	12.77%	10.82%	3.95%	3.92%
30	33.14%	27.53%	25.35%	21.37%	12.75%	10.81%	3.95%	3.92%
35	33.14%	27.53%	25.35%	21.36%	12.73%	10.80%	3.95%	3.92%
40	33.14%	27.53%	25.35%	21.36%	12.72%	10.79%	3.95%	3.92%
45	33.14%	27.53%	25.34%	21.35%	12.72%	10.79%	3.95%	3.92%
50	33.14%	27.53%	25.34%	21.35%	12.71%	10.79%	3.95%	3.92%
55	33.14%	27.53%	25.34%	21.35%	12.71%	10.78%	3.95%	3.92%
60	33.14%	27.53%	25.34%	21.35%	12.70%	10.78%	3.95%	3.92%
65	33.14%	27.53%	25.34%	21.35%	12.70%	10.78%	3.95%	3.92%
70	33.14%	27.53%	25.34%	21.34%	12.70%	10.78%	3.95%	3.92%
75	33.14%	27.53%	25.34%	21.34%	12.70%	10.78%	3.95%	3.92%
80	33.14%	27.53%	25.34%	21.34%	12.70%	10.78%	3.95%	3.92%
85	33.14%	27.53%	25.34%	21.34%	12.70%	10.78%	3.95%	3.92%

Table 6.43: Development of the relative energy error for increasing numbers of enrichments obtained from Fourier-type basis functions in x_1 and x_2 direction with increasing number of maximum oscillations.

able to reduce the error to less than 4%, but the speed of decay is quite slow. This coincides with the observation from Section 6.2.2, showing that no dominant eigenvalues could be identified for Fourier-type boundary data. For completeness, the difference between various enriched solutions and the reference solution is presented in Figure 6.26. For 10 and 18 enrichments, the largest magnitudes of error are mainly located in the overlap region of the coarse patches outside of ω. In Figure 6.27, the decay of the relative errors from table 6.43 is visualized.

Conclusions from the benchmark problem

The previous discussion showed that the homogeneous part of the solution of the benchmark problem under study could be approximated well using enrichments constructed from boundary hats, B-Splines or oscillatory Fourier-type functions used as boundary data in the sampling problem. All

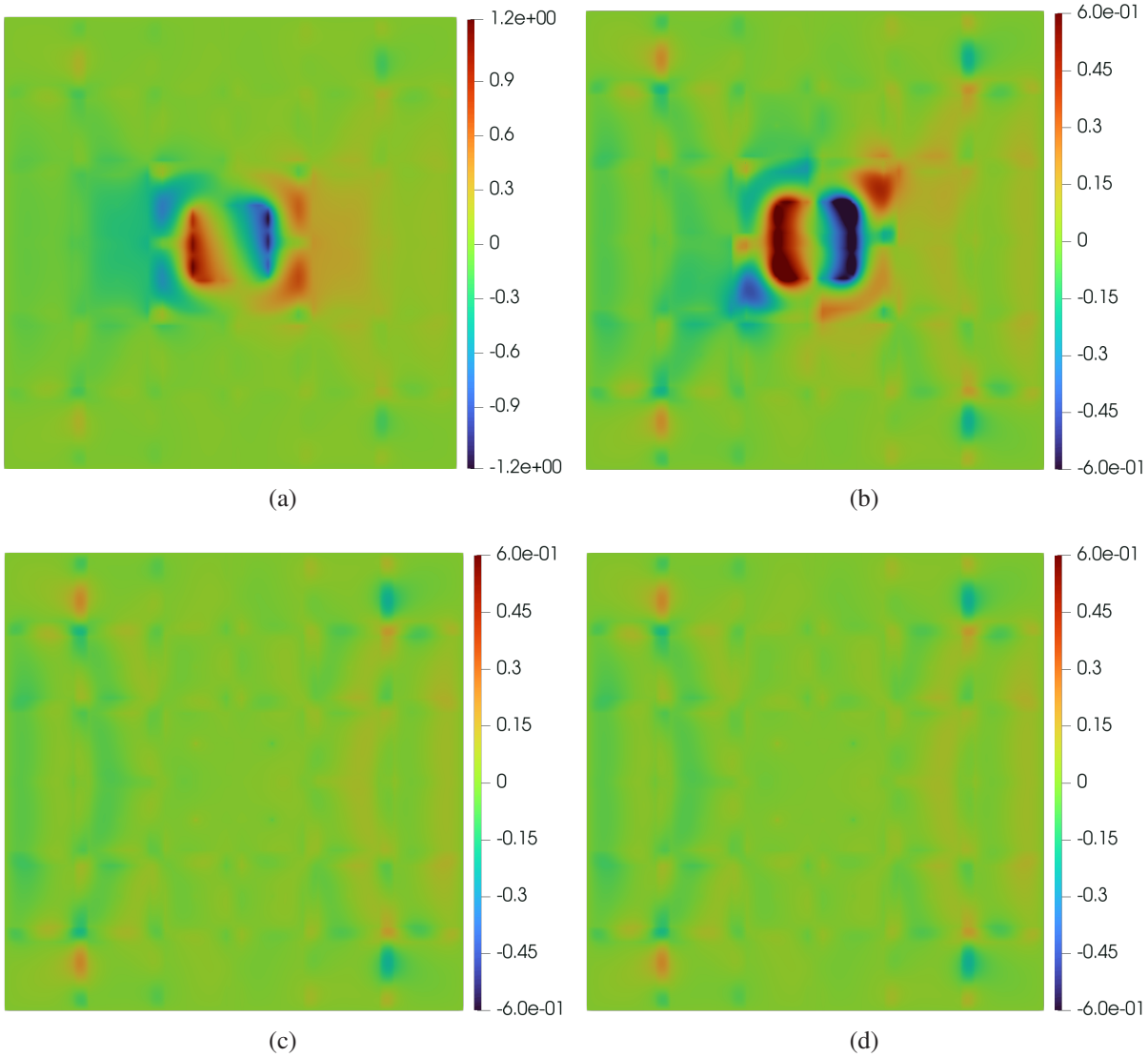


Figure 6.26: Difference between reference and enriched solutions, for (a) only particular solutions, (b) particular solutions and two enrichments, (c) particular solutions and ten enrichments, (d) particular solutions and eighteen enrichments. Enrichments were computed from Fourier-type boundary data in x_1 and x_2 direction, for a maximum of 85 oscillations. Note that the scale of the error in (a) is twice as large as the scale used in (b), (c) and (d).

approaches were capable of reducing the initial error of 33.14% to $\approx 3.9\%$. The workload needed to construct the various sets of enrichments, the number of dominant eigenvalues, and the speed of decay of the relative energy error in global computations vary significantly. As seen in the first benchmark problem, the boundary hats approach is the most versatile, but in general also the most expensive, since the number of boundary data functions grows very fast for increasing boundary levels. The B-Spline approach is the cheapest, since even for 9 pairwise different inner knots there are only a small number of sampling problems to be solved. B-Splines in both coordinate directions including the corner splines, as well as boundary hats defined on a sufficiently high boundary level, led to 2 dominant eigenvalues. The use of Fourier-type boundary data did not result in any dominant eigenvalue and yielded the slowest decay of the error in the global computations. The

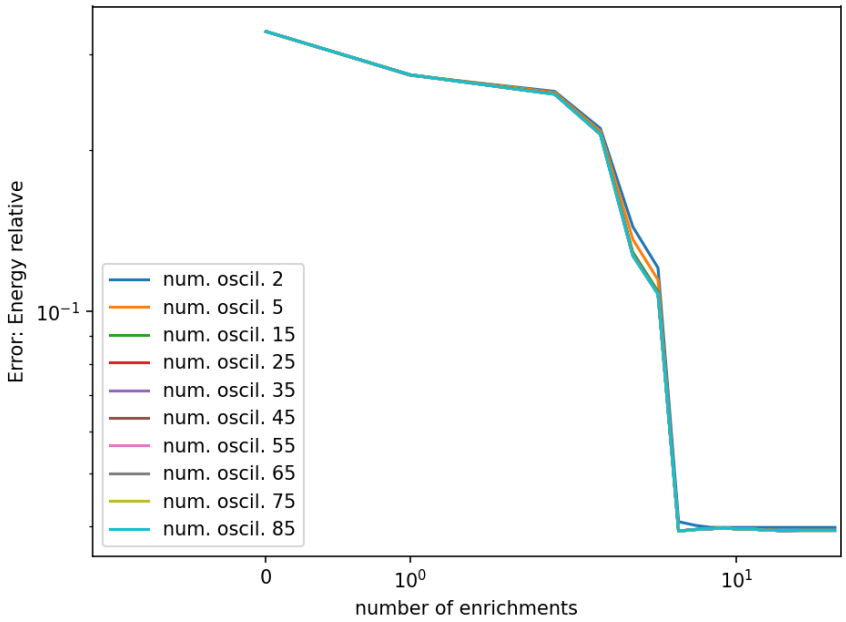


Figure 6.27: Relative energy error for increasing numbers of enrichments, constructed from Fourier-type boundary data in x_1 and x_2 direction for various maximum numbers of oscillations.

size of eigenvalues and the speed of decay of the (relative) energy error are clearly correlated. The best overall performance was achieved by using B-Splines in x_1 and x_2 direction, together with corner splines and defined on 9 pairwise different inner knots.

6.3 Isotropic linear elasticity in 2d

This section presents results regarding a problem of linear elasticity on a rectangular two-dimensional domain with a circular hole in the center. The material is assumed to be isotropic with a modulus of elasticity of $E = 100.00\,\mathrm{GPa}$ and a Poisson ratio of $\nu = 0.30$. No boundary condition is prescribed on the interior boundary of the hole. It is expected, that the strain of the solution along the whole surface presents fine-scale behavior which is hard to grasp using standard basis functions on coarse patches.

The remainder of this section is structured as in the previous two benchmark problems, i.e. Section 6.3.1 explicitly presents the problem under study and in Section 6.3.2 the effect of the different choices of boundary data on the eigenvalues obtained from the computation of the optimal shape functions are investigated. Finally, Section 6.3.3 investigates the performance of the enriched computations.

6.3.1 Problem formulation

Problem 16 presents the benchmark problem under study in this section.

Problem 16: Isotropic linear elasticity in 2d.

Let $\Omega = [-6,6] \backslash B_{0.5}(0,0) \subset \mathbb{R}^2$ be a two-dimensional plate with a circular hole in the center. Also, let $\underline{\underline{\mathbf{C}}}$ be the the stiffness tensor for the isotropic material defined by modulus of elasticity $E = 100.00$ and Poisson ratio $\nu = 0.30$. Finally, let $f(x) = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$ and

$$\begin{aligned} g_l(x) &= \begin{bmatrix} 0 & 0 \end{bmatrix}^T \\ g_b(x) &= \begin{bmatrix} 0 & 0 \end{bmatrix}^T \\ g_t(x) &= \frac{1}{12} \begin{bmatrix} 5(x_1+6) & 2(x_1+6) \end{bmatrix}^T \\ g_r(x) &= \frac{1}{12} \begin{bmatrix} 5(x_2+6) & 2(x_2+6) \end{bmatrix}^T, \end{aligned} \tag{6.34}$$

as well as $g : \partial\,\Omega \backslash \partial B_{0.5}(0,0) \to \mathbb{R}^2$ with

$$x \mapsto g(x) := \begin{cases} g_l, & \text{if } x_1 = -6 \\ g_r, & \text{if } x_1 = 6 \\ g_b, & \text{if } x_2 = -6 \\ g_b, & \text{if } x_2 = 6. \end{cases} \tag{6.35}$$

Consider the differential operator

$$\mathcal{L} : [\mathcal{C}^2(\Omega)]^2 \to [\mathcal{C}^0(\Omega)]^2, \quad u \mapsto \mathcal{L}\,u := -\operatorname{div}(\sigma(u)). \tag{6.36}$$

Find a function $u \in [\mathcal{C}^2(\Omega)]^2$ satisfying

$$\begin{aligned} -\mathcal{L}\,u(x) &= f(x), \quad \text{in } \Omega \\ u(x) &= g(x), \quad \text{on } \partial\,\Omega, \end{aligned} \tag{6.37}$$

with the stress tensor $\sigma(u) = \underline{\underline{\mathbf{C}}}(u)\varepsilon(u)$ and $\varepsilon(u)$ is the strain tensor $\varepsilon(u) = \frac{1}{2}\left(\nabla u + (\nabla u)^T\right)$.

The boundary conditions visually mean, that the bottom and left face of the plate are clamped, whereas the top right vertex is pulled to a fixed displacement of $\begin{bmatrix} 5 & 2 \end{bmatrix}$. This behavior is described by the linear functions g_t and g_r. The equations of isotropic linear elasticity are a special case of the equations of orthotropic linear elasticity, and in Section 3.2 it has been shown that $\mathcal{L}$ is an elliptic operator. Using the trial and test spaces

$$\begin{aligned} \mathrm{V}^{\text{trial}} &:= \{u \in [H^1(\Omega)]^2 : \operatorname{tr}(u) = g, \text{ on } \partial\,\Omega \backslash \partial B_{0.5}(0,0)\} \\ \mathrm{V}^{\text{test}} &:= \{u \in [H^1(\Omega)]^2 : \operatorname{tr}(u) = \begin{bmatrix} 0 & 0 \end{bmatrix}^T, \text{ on } \partial\,\Omega \backslash \partial B_{0.5}(0,0)\} \end{aligned} \tag{6.38}$$

the weak formulation of Problem 16 is derived.

Problem 17: Weak isotropic linear elasticity in 2d.

Let $\Omega = [-6,6] \backslash B_{0.5}(0,0) \subset \mathbb{R}^2$ be a two-dimensional plate with a circular hole in the center. Also, let $\underline{\underline{\mathbf{C}}}$ be the stiffness tensor for the isotropic material defined by modulus of elasticity $E = 100.00$ and Poisson ratio $\nu = 0.30$. Finally, let $f(x) = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$ and

$$\begin{aligned} g_l(x) &= \begin{bmatrix} 0 & 0 \end{bmatrix}^T \\ g_b(x) &= \begin{bmatrix} 0 & 0 \end{bmatrix}^T \\ g_t(x) &= \frac{1}{12} \begin{bmatrix} 5(x_1+6) & 2(x_1+6) \end{bmatrix}^T \\ g_r(x) &= \frac{1}{12} \begin{bmatrix} 5(x_2+6) & 2(x_2+6) \end{bmatrix}^T, \end{aligned} \tag{6.39}$$

as well as $g : \partial\Omega \backslash \partial B_{0.5}(0,0) \to \mathbb{R}^2$ with

$$x \mapsto g(x) := \begin{cases} g_l, & \text{if } x_1 = -6 \\ g_r, & \text{if } x_1 = 6 \\ g_b, & \text{if } x_2 = -6 \\ g_b, & \text{if } x_2 = 6. \end{cases} \tag{6.40}$$

Define the bilinear form $\mathbf{a} : \mathrm{V}^{\text{trial}}(\Omega) \times \mathrm{V}^{\text{test}}(\Omega) \to \mathbb{R}^2$ and linear functional $\ell : \mathrm{V}^{\text{test}} \to \mathbb{R}$,

$$\begin{aligned} \mathbf{a}[u,v] &:= \int_\Omega \sigma(u) : \varepsilon(v)\, dx \\ \ell(v) &:= \int_\Omega f \cdot v\, dx. \end{aligned} \tag{6.41}$$

Find a function $u \in \mathrm{V}^{\text{trial}}(\Omega)$ satisfying

$$\mathbf{a}[u,v] = \ell(v), \quad \forall v \in \mathrm{V}^{\text{test}}(\Omega). \tag{6.42}$$

As presented in Section 3.2, the bilinear form $\mathbf{a}$ for the isotropic material under study is continuous and elliptic. Since furthermore ℓ is a continuous linear functional, the theorem of Lax and Milgram (Theorem 2.4) guarantees solvability of Problem 17.

6.3.2 Influence of the boundary data

This section studies the influence of the various type of boundary data on the eigenvalues obtained from the computation of the optimal shape functions. In the following, the implications of boundary data resulting defined by the boundary hats approach, the B-Spline approach, and the Fourier-type approach are investigated. As in the previous benchmark problems, conclusions are drawn afterwards.

Boundary hats

The following Tables 6.44 to 6.47 present the results obtained from the use of constant, linear and quadratic boundary hats, as well as the combination of constant and linear boundary hats used as boundary data in the sampling problem. The tables show, that for boundary hats of all degrees

ℓ	pd	dim	λ_1	λ_2	λ_8	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	0	24	$1.954 \cdot 10^{-1}$	$1.929 \cdot 10^{-1}$	$1.757 \cdot 10^{-2}$	3	9	18	22
3	0	56	$3.023 \cdot 10^{-1}$	$2.777 \cdot 10^{-1}$	$3.375 \cdot 10^{-2}$	3	10	24	49
4	0	120	$3.972 \cdot 10^{-1}$	$3.552 \cdot 10^{-1}$	$5.492 \cdot 10^{-2}$	5	13	27	53
5	0	248	$4.692 \cdot 10^{-1}$	$4.108 \cdot 10^{-1}$	$7.664 \cdot 10^{-2}$	5	13	27	57
6	0	504	$5.182 \cdot 10^{-1}$	$4.433 \cdot 10^{-1}$	$8.941 \cdot 10^{-2}$	5	13	29	57
7	0	1016	$5.479 \cdot 10^{-1}$	$4.607 \cdot 10^{-1}$	$9.69 \cdot 10^{-2}$	6	13	29	57
8	0	2040	$5.666 \cdot 10^{-1}$	$4.708 \cdot 10^{-1}$	$1.018 \cdot 10^{-1}$	8	15	29	59

Table 6.44: Key numbers obtained from the computation of the optimal shape functions using constant boundary hats on various boundary levels.

ℓ	pd	dim	λ_1	λ_2	λ_8	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	1	48	$4.921 \cdot 10^{-1}$	$4.53 \cdot 10^{-1}$	$7.888 \cdot 10^{-2}$	5	13	25	47
3	1	112	$5.555 \cdot 10^{-1}$	$4.653 \cdot 10^{-1}$	$1.006 \cdot 10^{-1}$	8	13	29	55
4	1	240	$5.652 \cdot 10^{-1}$	$4.702 \cdot 10^{-1}$	$1.02 \cdot 10^{-1}$	8	15	29	57
5	1	496	$5.672 \cdot 10^{-1}$	$4.711 \cdot 10^{-1}$	$1.021 \cdot 10^{-1}$	8	15	29	59
6	1	1008	$5.676 \cdot 10^{-1}$	$4.713 \cdot 10^{-1}$	$1.021 \cdot 10^{-1}$	8	15	29	60
7	1	2032	$5.678 \cdot 10^{-1}$	$4.714 \cdot 10^{-1}$	$1.021 \cdot 10^{-1}$	8	15	29	61

Table 6.45: Key numbers obtained from the computation of the optimal shape functions using linear boundary hats in x_1 and x_2 direction on various boundary levels.

ℓ	pd	dim	λ_1	λ_2	λ_8	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	2	48	$4.878 \cdot 10^{-1}$	$3.89 \cdot 10^{-1}$	$7.133 \cdot 10^{-2}$	6	13	25	48
3	2	112	$5.549 \cdot 10^{-1}$	$4.439 \cdot 10^{-1}$	$9.527 \cdot 10^{-2}$	6	13	29	56
4	2	240	$5.653 \cdot 10^{-1}$	$4.648 \cdot 10^{-1}$	$1.006 \cdot 10^{-1}$	8	15	29	57
5	2	496	$5.672 \cdot 10^{-1}$	$4.698 \cdot 10^{-1}$	$1.018 \cdot 10^{-1}$	8	15	29	59
6	2	1008	$5.677 \cdot 10^{-1}$	$4.71 \cdot 10^{-1}$	$1.021 \cdot 10^{-1}$	8	15	29	59
7	2	2032	$5.677 \cdot 10^{-1}$	$4.711 \cdot 10^{-1}$	$1.02 \cdot 10^{-1}$	8	15	29	60
8	2	4080	$5.677 \cdot 10^{-1}$	$4.712 \cdot 10^{-1}$	$1.021 \cdot 10^{-1}$	8	15	29	59

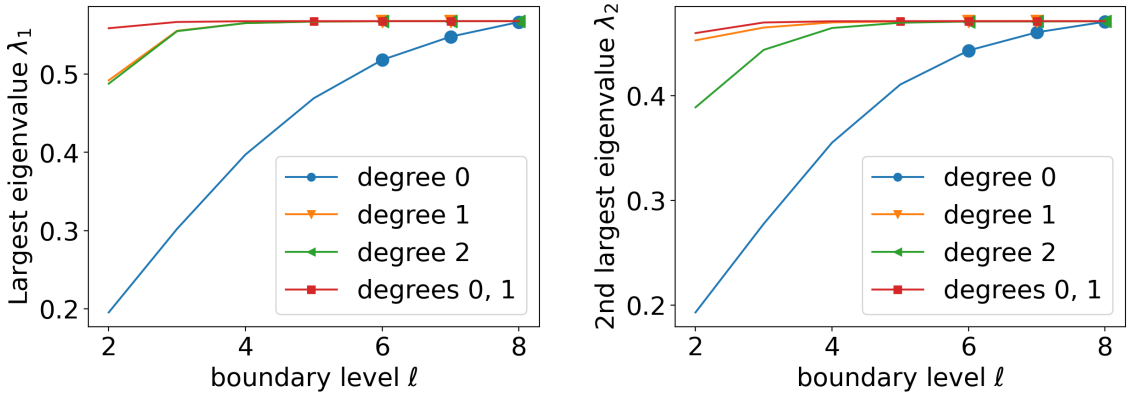
Table 6.46: Key numbers obtained from the computation of the optimal shape functions using quadratic boundary hats in x_1 and x_2 direction on various boundary levels.

and sufficiently high boundary level there are 8 dominant, i.e. very large eigenvalues and 15 large eigenvalues. For increasing boundary level, the largest eigenvalues stabilize at similar values, the

ℓ	pd	dim	λ_1	λ_2	λ_8	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	0, 1	72	$5.587 \cdot 10^{-1}$	$4.6 \cdot 10^{-1}$	$9.193 \cdot 10^{-2}$	5	13	26	49
3	0, 1	168	$5.667 \cdot 10^{-1}$	$4.701 \cdot 10^{-1}$	$1.008 \cdot 10^{-1}$	8	15	29	55
4	0, 1	360	$5.677 \cdot 10^{-1}$	$4.712 \cdot 10^{-1}$	$1.02 \cdot 10^{-1}$	8	15	29	57
5	0, 1	744	$5.678 \cdot 10^{-1}$	$4.714 \cdot 10^{-1}$	$1.021 \cdot 10^{-1}$	8	15	29	59
6	0, 1	1512	$5.678 \cdot 10^{-1}$	$4.714 \cdot 10^{-1}$	$1.021 \cdot 10^{-1}$	8	15	29	59
7	0, 1	3048	$5.678 \cdot 10^{-1}$	$4.714 \cdot 10^{-1}$	$1.021 \cdot 10^{-1}$	8	15	29	59
8	0, 1	6120	$5.678 \cdot 10^{-1}$	$4.714 \cdot 10^{-1}$	$1.021 \cdot 10^{-1}$	8	15	29	59

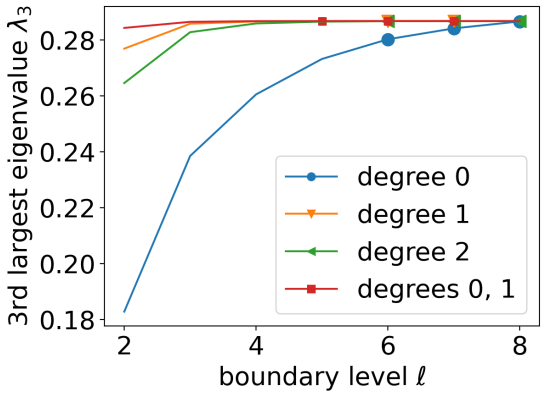
Table 6.47: Key numbers obtained from the computation of the optimal shape functions using constant and linear boundary hats in x_1 and x_2 direction on various boundary levels.

first four being approximately 0.57, 0.47, 0.29 and 0.24. The development of the four largest eigenvalues is visualized in Figure 6.28. The plots show, that approximately 500 degrees of freedom are sufficient to obtain large values of the four largest eigenvalues whenever linear or quadratic boundary hats are at least included as boundary data in the sampling problem. When using only constant boundary hats, a higher boundary level must be chosen to obtain results of the same magnitude. In total, boundary hats of all degrees can be employed to capture dominant eigenvalues of large magnitudes. Compared to the previously studied benchmark problems, however, it seems that constant boundary hats are not that important in the problem currently under study. In order to keep the number of degrees of freedom, i.e. the number of sampling problems to be solved as small as possible, enrichments based on linear or quadratic boundary hats on boundary level $\ell = 4$ or $\ell = 5$ seem most promising.

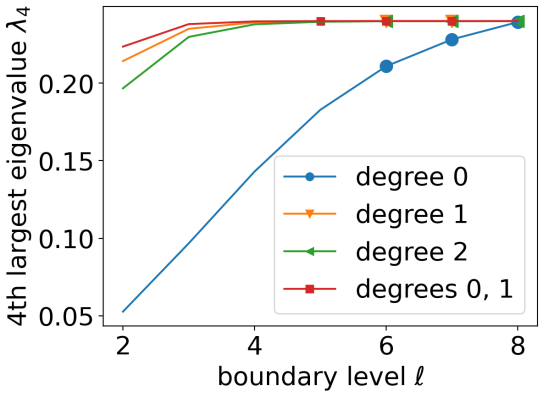


(a) λ_1

(b) λ_2



(c) λ_3



(d) λ_4

Figure 6.28: Development of the four largest eigenvalues for increasing boundary levels and all considered boundary hats degrees. Markers are plotted every 500 degrees of freedom.

B-Splines

This subsection considers optimal shape functions constructed from B-Splines used as boundary data in the sampling problem. The results for the four sets of B-Splines used already in the previous benchmark problems, i.e. only in x_1 or in x_1 and x_2 direction as well as with / without corner splines, are presented in Tables 6.48 to 6.51. The results show that the four approaches differ significantly in the number of dominant eigenvalues they capture, as well as their magnitude. The largest eigenvalue attainable using boundary data consisting of B-Splines in x_1 direction without corner splines is ≈ 0.343. Using B-Splines also in x_2 direction, as well as considering corner splines, the magnitude of the largest achievable eigenvalue is increased to ≈ 0.568. The implications of the various sets of B-Splines on the other eigenvalues are even stronger, as can be seen in Figure 6.29. The results show, that a reasonable choice of boundary data in the sampling problem is very important. It is expected, that the enrichments constructed from B-Splines in both coordinate direction together with corner splines will have superior performance in the global computations presented below in Section 6.3.3. Since the B-Spline approach usually relies on relatively few solutions of the sampling problem, this approach is recommended.

n	dim	λ_1	λ_2	λ_8	n_{-1}	n_{-2}	n_{-4}	n_{-8}
3	32	$3.411 \cdot 10^{-1}$	$1.647 \cdot 10^{-1}$	$1.274 \cdot 10^{-2}$	4	8	15	16
4	40	$3.423 \cdot 10^{-1}$	$1.706 \cdot 10^{-1}$	$1.698 \cdot 10^{-2}$	4	8	16	20
5	48	$3.427 \cdot 10^{-1}$	$1.719 \cdot 10^{-1}$	$1.785 \cdot 10^{-2}$	4	9	17	24
6	56	$3.428 \cdot 10^{-1}$	$1.723 \cdot 10^{-1}$	$1.863 \cdot 10^{-2}$	4	9	19	28
7	64	$3.429 \cdot 10^{-1}$	$1.725 \cdot 10^{-1}$	$1.892 \cdot 10^{-2}$	4	9	19	32
8	72	$3.429 \cdot 10^{-1}$	$1.725 \cdot 10^{-1}$	$1.904 \cdot 10^{-2}$	4	9	19	36
9	80	$3.429 \cdot 10^{-1}$	$1.725 \cdot 10^{-1}$	$1.909 \cdot 10^{-2}$	4	9	20	36

Table 6.48: Key numbers obtained from the computation of the optimal shape functions using B-Splines in x_1 direction, not including corner splines.

n	dim	λ_1	λ_2	λ_8	n_{-1}	n_{-2}	n_{-4}	n_{-8}
3	48	$4.309 \cdot 10^{-1}$	$3.795 \cdot 10^{-1}$	$4.471 \cdot 10^{-2}$	5	11	18	21
4	56	$4.32 \cdot 10^{-1}$	$3.809 \cdot 10^{-1}$	$5.294 \cdot 10^{-2}$	5	11	21	25
5	64	$4.323 \cdot 10^{-1}$	$3.814 \cdot 10^{-1}$	$5.457 \cdot 10^{-2}$	5	11	21	29
6	72	$4.324 \cdot 10^{-1}$	$3.815 \cdot 10^{-1}$	$5.532 \cdot 10^{-2}$	5	11	21	33
7	80	$4.324 \cdot 10^{-1}$	$3.816 \cdot 10^{-1}$	$5.559 \cdot 10^{-2}$	5	11	21	37
8	88	$4.324 \cdot 10^{-1}$	$3.817 \cdot 10^{-1}$	$5.568 \cdot 10^{-2}$	5	11	22	40
9	96	$4.324 \cdot 10^{-1}$	$3.817 \cdot 10^{-1}$	$5.573 \cdot 10^{-2}$	5	11	22	40

Table 6.49: Key numbers obtained from the computation of the optimal shape functions using B-Splines in x_1 direction, including corner splines.

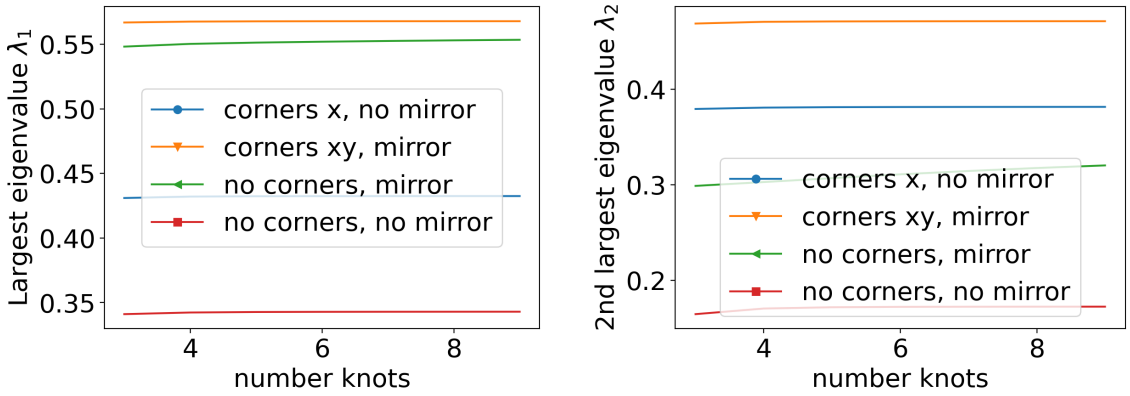
n	dim	λ_1	λ_2	λ_8	n_{-1}	n_{-2}	n_{-4}	n_{-8}
3	64	$5.481 \cdot 10^{-1}$	$2.99 \cdot 10^{-1}$	$8.915 \cdot 10^{-2}$	6	13	24	32
4	80	$5.502 \cdot 10^{-1}$	$3.03 \cdot 10^{-1}$	$9.278 \cdot 10^{-2}$	6	13	25	40
5	96	$5.512 \cdot 10^{-1}$	$3.071 \cdot 10^{-1}$	$9.508 \cdot 10^{-2}$	6	13	27	45
6	112	$5.519 \cdot 10^{-1}$	$3.111 \cdot 10^{-1}$	$9.616 \cdot 10^{-2}$	6	15	27	49
7	128	$5.525 \cdot 10^{-1}$	$3.146 \cdot 10^{-1}$	$9.668 \cdot 10^{-2}$	6	15	29	50
8	144	$5.53 \cdot 10^{-1}$	$3.177 \cdot 10^{-1}$	$9.697 \cdot 10^{-2}$	6	15	29	53
9	74	$5.534 \cdot 10^{-1}$	$3.204 \cdot 10^{-1}$	$9.715 \cdot 10^{-2}$	6	15	29	54

Table 6.50: Key numbers obtained from the computation of the optimal shape functions using B-Splines in x_1 and x_2 direction, not including corner splines.

Exemplarily, the first three enrichments obtained from boundary hats of polynomial degree $pd = \{0, 1\}$ and boundary level $\ell = 8$ are shown in Figure 6.30.

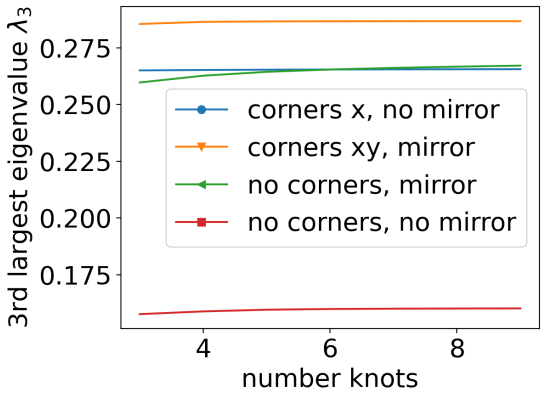
n	dim	λ_1	λ_2	λ_8	n_{-1}	n_{-2}	n_{-4}	n_{-8}
3	192	$5.668 \cdot 10^{-1}$	$4.689 \cdot 10^{-1}$	$9.544 \cdot 10^{-2}$	6	13	25	37
4	224	$5.675 \cdot 10^{-1}$	$4.707 \cdot 10^{-1}$	$9.95 \cdot 10^{-2}$	6	13	26	42
5	256	$5.677 \cdot 10^{-1}$	$4.711 \cdot 10^{-1}$	$1.013 \cdot 10^{-1}$	8	13	27	45
6	288	$5.677 \cdot 10^{-1}$	$4.713 \cdot 10^{-1}$	$1.018 \cdot 10^{-1}$	8	15	27	49
7	320	$5.678 \cdot 10^{-1}$	$4.713 \cdot 10^{-1}$	$1.02 \cdot 10^{-1}$	8	15	29	53
8	352	$5.678 \cdot 10^{-1}$	$4.713 \cdot 10^{-1}$	$1.021 \cdot 10^{-1}$	8	15	29	53
9	384	$5.678 \cdot 10^{-1}$	$4.714 \cdot 10^{-1}$	$1.021 \cdot 10^{-1}$	8	15	29	54

Table 6.51: Key numbers obtained from the computation of the optimal shape functions using B-Splines in x_1 and x_2 direction, including corner splines.

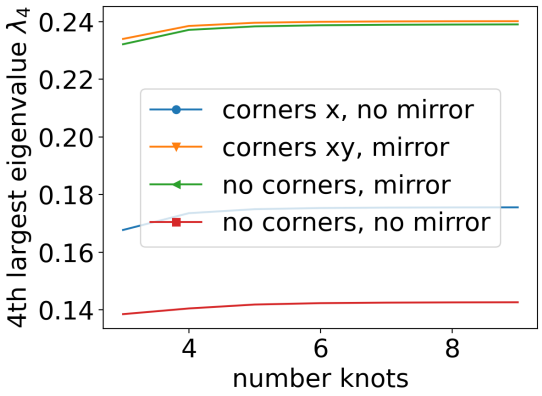


(a) λ_1

(b) λ_2

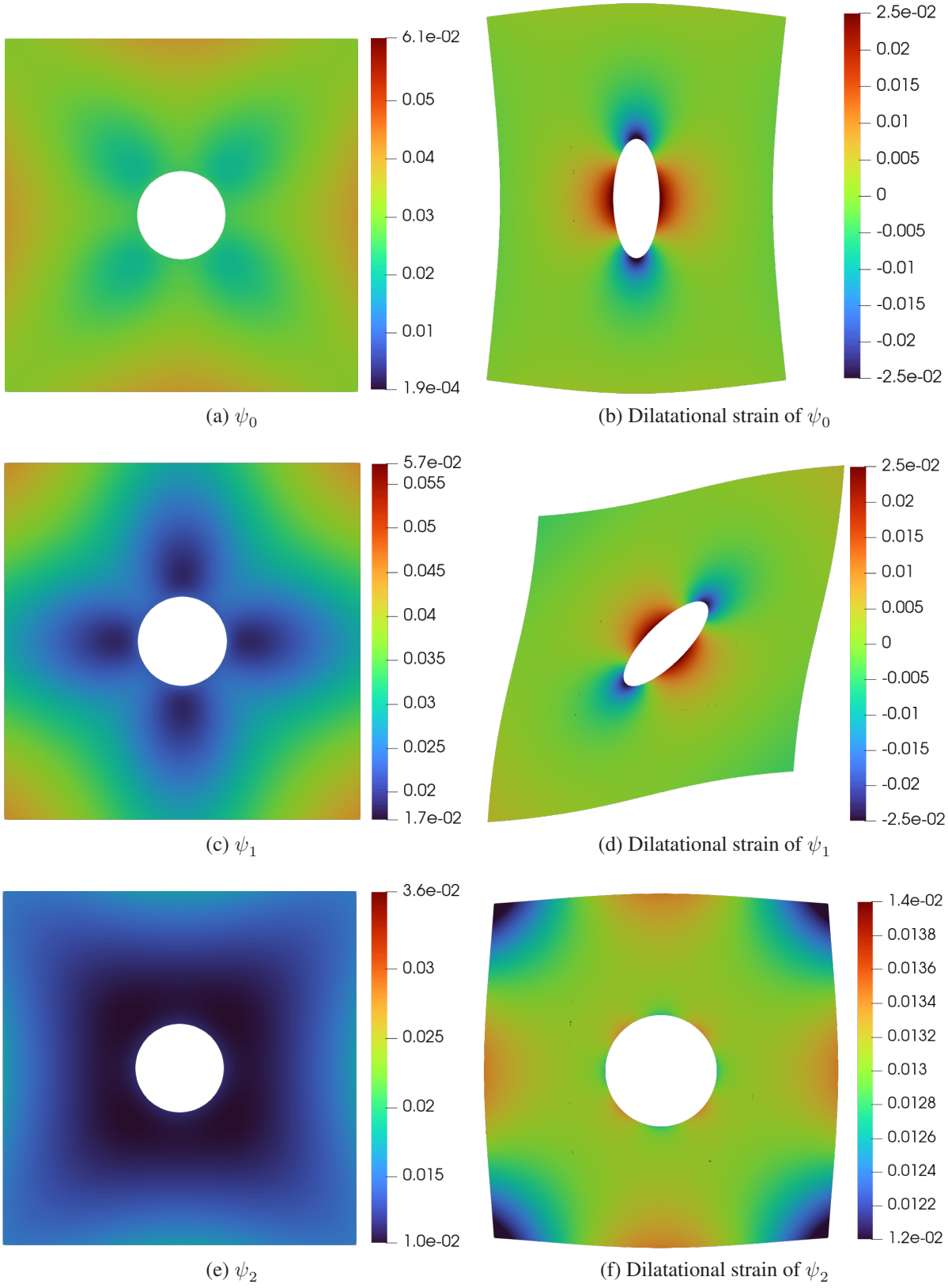


(c) λ_3



(d) λ_4

Figure 6.29: Development of the four largest eigenvalues for increasing boundary levels and the four different sets of B-Splines. All computations use less than 400 degrees of freedom.



(a) ψ_0 (b) Dilatational strain of ψ_0

(c) ψ_1 (d) Dilatational strain of ψ_1

(e) ψ_2 (f) Dilatational strain of ψ_2

Figure 6.30: The first three optimal shape functions computed using $pd = \{0, 1\}$ and $\ell = 8$. Left column: Restriction of the functions to ω, right column: Dilatational strain of the functions on the warped domain.

Oscillating trigonometric functions

Table 6.52 presents key numbers for the optimal shape functions computed from Fourier-type boundary data. For all maximum numbers of oscillations, 4 very large and 11 large eigenvalues were captured. It is also to be noted that $\lambda_1 \approx 0.28$ is far larger than $2 \approx 0.125$, $\lambda_2 \approx \lambda_3$, and also that the magnitudes of the four dominant eigenvalues change only slightly when using more than 10 oscillations. Due to the relatively small magnitude of the dominant eigenvalues obtained from

n_{osc}	dim	λ_1	λ_2	λ_8	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	46	$2.449 \cdot 10^{-1}$	$1.224 \cdot 10^{-1}$	$1.691 \cdot 10^{-2}$	4	11	16	16
5	106	$2.7 \cdot 10^{-1}$	$1.247 \cdot 10^{-1}$	$1.81 \cdot 10^{-2}$	4	11	23	40
10	206	$2.808 \cdot 10^{-1}$	$1.249 \cdot 10^{-1}$	$1.839 \cdot 10^{-2}$	4	11	23	44
15	306	$2.839 \cdot 10^{-1}$	$1.25 \cdot 10^{-1}$	$1.844 \cdot 10^{-2}$	4	11	23	44
20	406	$2.853 \cdot 10^{-1}$	$1.25 \cdot 10^{-1}$	$1.847 \cdot 10^{-2}$	4	11	23	44
25	506	$2.86 \cdot 10^{-1}$	$1.25 \cdot 10^{-1}$	$1.848 \cdot 10^{-2}$	4	11	23	44
30	606	$2.864 \cdot 10^{-1}$	$1.25 \cdot 10^{-1}$	$1.848 \cdot 10^{-2}$	4	11	23	44
35	706	$2.867 \cdot 10^{-1}$	$1.251 \cdot 10^{-1}$	$1.848 \cdot 10^{-2}$	4	11	23	44
40	806	$2.869 \cdot 10^{-1}$	$1.251 \cdot 10^{-1}$	$1.849 \cdot 10^{-2}$	4	11	23	44
45	906	$2.87 \cdot 10^{-1}$	$1.251 \cdot 10^{-1}$	$1.849 \cdot 10^{-2}$	4	11	23	44
50	1006	$2.871 \cdot 10^{-1}$	$1.251 \cdot 10^{-1}$	$1.849 \cdot 10^{-2}$	4	11	23	44
55	1106	$2.872 \cdot 10^{-1}$	$1.251 \cdot 10^{-1}$	$1.849 \cdot 10^{-2}$	4	11	23	44
60	1206	$2.872 \cdot 10^{-1}$	$1.251 \cdot 10^{-1}$	$1.849 \cdot 10^{-2}$	4	11	23	44
65	1306	$2.873 \cdot 10^{-1}$	$1.251 \cdot 10^{-1}$	$1.849 \cdot 10^{-2}$	4	11	23	44
70	1406	$2.873 \cdot 10^{-1}$	$1.251 \cdot 10^{-1}$	$1.849 \cdot 10^{-2}$	4	11	23	44
75	1506	$2.873 \cdot 10^{-1}$	$1.251 \cdot 10^{-1}$	$1.849 \cdot 10^{-2}$	4	11	23	44
80	1606	$2.873 \cdot 10^{-1}$	$1.251 \cdot 10^{-1}$	$1.849 \cdot 10^{-2}$	4	11	23	44
85	1706	$2.874 \cdot 10^{-1}$	$1.251 \cdot 10^{-1}$	$1.849 \cdot 10^{-2}$	4	11	23	44

Table 6.52: Key numbers obtained from the computation of the optimal shape functions using Fourier-type basis functions in x_1 and x_2 direction for increasing numbers of oscillations.

the Fourier-type approach, as well as the conclusions from the previous benchmark problems, it is expected that the enrichments constructed from Fourier-type boundary data will perform worse than the boundary hats resp. the B-Spline approach.

Conclusive remarks

The results presented for the benchmark problem at hand implicate that promising sets of optimal shape functions can be computed from the boundary hats approach, as well as the B-Spline approach. Similar to the conclusions drawn from the previous benchmark problems, enrichments constructed from Fourier-type boundary data is expected to perform poor, meaning that the number of enrichments needed to significantly reduce the error in a global computation is large. For the benchmark problem at hand, and in contrast to the previous ones, constant boundary hats do not promise the best results. Instead, it seems that linear / quadratic boundary hats on the moderate boundary level 4 or 5 should be used, in order to maintain the number of boundary data functions relatively small. Also, using B-Splines in both coordinate directions and including corner splines produced promising magnitudes of dominant eigenvalues.

6.3.3 Discussion of global errors

This section repeats the procedure of the previous benchmark problems. After a numerical reference solution is identified, the various sets of enrichments constructed in Section 6.3.2 are used and the performance of the enriched solutions is analyzed. This allows to validate the claims made in Section 6.3.2.

The bounding box is chosen to be $[-6, 10]^2$, and unenriched solutions u_h are computed for various cover levels ι and corresponding patch size

$$h = \frac{4}{2^{\iota-2}}. \tag{6.43}$$

This leads to discrete values of the function,

$$E : \mathbb{R}_+ \to \mathbb{R}, \quad h \mapsto E(h) := a_\Omega[u_h, u_h]^{\frac{1}{2}}, \tag{6.44}$$

whose value at the limit $h = 0$ is extrapolated. The data points are visualized in Figure 6.31. The

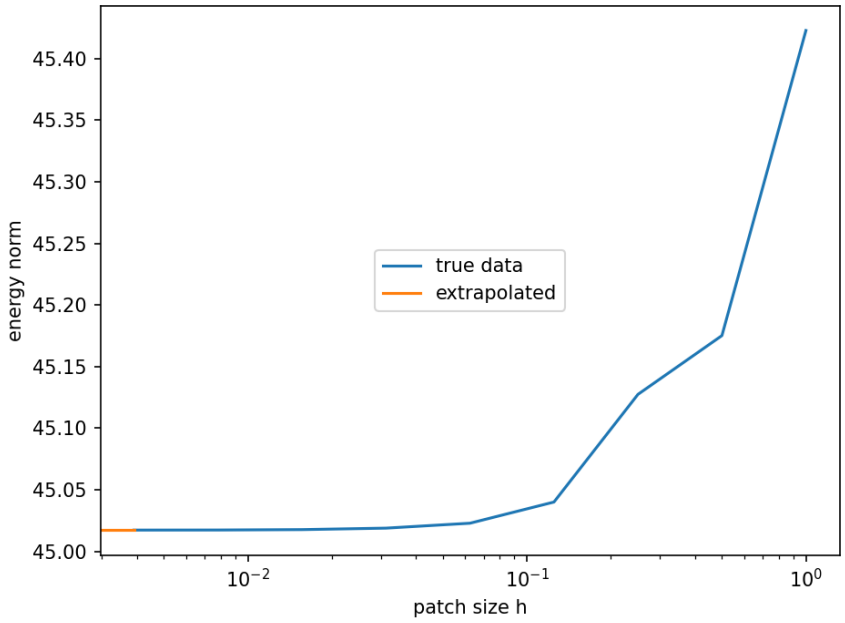


Figure 6.31: The energy of the solutions u_h for various various values of h. The extrapolated limit value at $h = 0$ is also shown.

energy norm of the level 10 solution, i.e. $h = 0.015625$, is approximately equal to the extrapolated value at $h = 0$, and will hence be used as reference solution in the following. The reference solution is shown in Figure 6.32.

As in the previous benchmark problems, local particular solutions on all patches of the coarse discretization for $\iota = 3$ are computed for the load and eventually for the boundary conditions, if the patch intersects the global boundary. The four center patches covering $[-1, 1]^2$ are furthermore enriched using the optimal shape functions constructed in the previous section 6.3.2. This situation is sketched in Figure 6.33. The displacement of the solution along the surface of the hole is

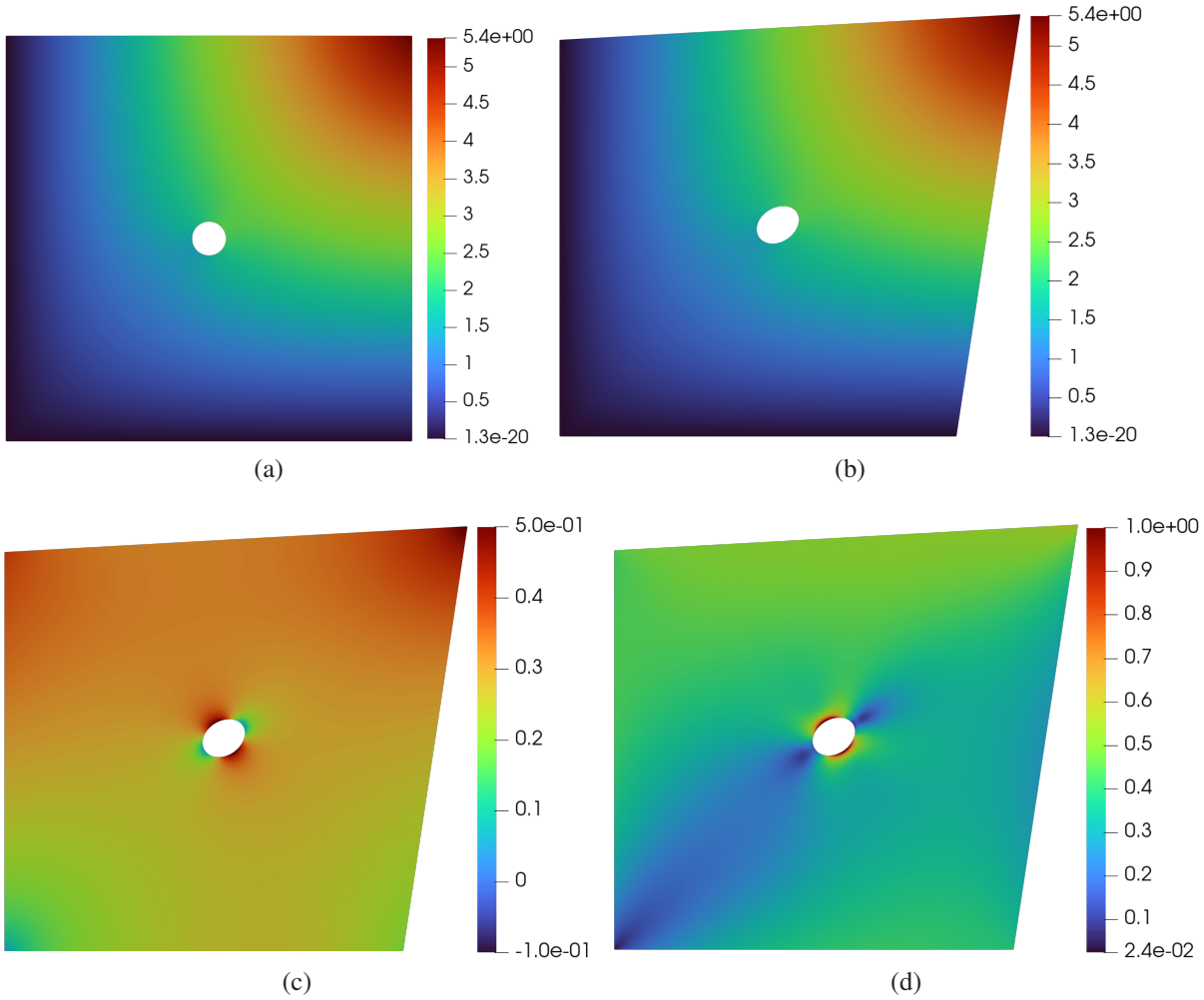


Figure 6.32: (a) Reference solution $u_{0.015625}$, (b) reference solution on warped domain, (c) dilatational strain on warped domain, (d) distortional strain on warped domain

expected to change on a scale far smaller than what can be resolved using low-degree polynomial basis functions, leading to an initial relative energy error of about 8%.

The optimal basis functions were computed on the oversampled patch $\omega^+ = [-4,4]^2 \backslash B_{0.5}(0,0)$ and will be used as enrichments in the four center patches marked in red in Figure 6.33, which cover $\omega = [-2,2]^2 \backslash B_{0.5}(0,0)$.

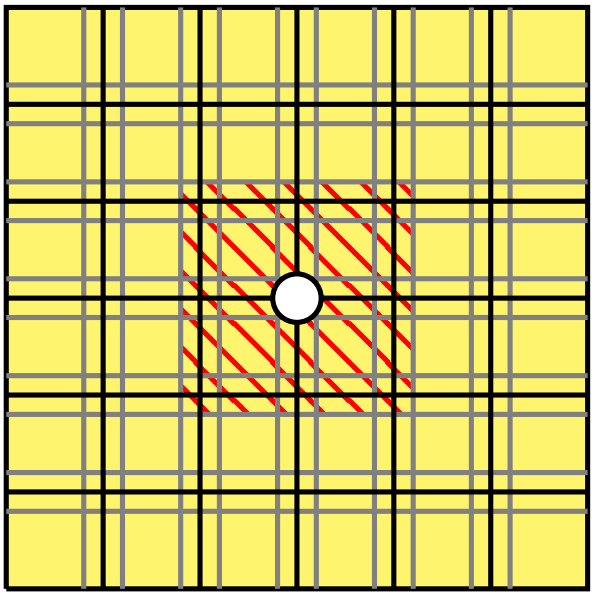


Figure 6.33: Coarse global discretization. Local particular solutions for load and essential boundary conditions are used on the corresponding patches, and the red patches will be further enriched using optimal shape functions. The red patches cover the patch ω.

Boundary hats

In Tables 6.53 to 6.56, the relative energy error of the enriched solutions based on constant, linear and quadratic boundary hats, as well as constant and the combination of constant and linear boundary hats are presented.

ℓ	pd	nn 0	nn 2	nn 4	nn 6	nn 8	nn 9	nn 10	nn 20
2	0	8.00%	7.15%	4.87%	4.60%	4.10%	3.70%	3.33%	2.64%
3	0	8.00%	7.15%	2.91%	2.78%	2.72%	2.72%	2.72%	2.62%
4	0	8.00%	7.15%	2.92%	2.79%	2.76%	2.73%	2.72%	2.62%
5	0	8.00%	7.15%	2.93%	2.79%	2.77%	2.73%	2.71%	2.62%
6	0	8.00%	7.16%	2.94%	2.87%	2.77%	2.73%	2.72%	2.62%
7	0	8.00%	7.16%	2.96%	2.88%	2.78%	2.73%	2.72%	2.62%
8	0	8.00%	7.16%	2.96%	2.89%	2.78%	2.73%	2.73%	2.62%

Table 6.53: Development of the relative energy error for increasing numbers of enrichments obtained from constant boundary hats.

From the tables it can be seen, that the first optimal shape function does not reduce the error at all. This is due to the previously described fact, that the Partition of Unity Method minimizes the global energy error, not the local ones. Hence, choosing a local enrichment as being part of the solution may improve the corresponding local error, but increase the error on adjacent parts of the domain, which could ultimately result in worse performance. This is exactly what happens in the benchmark problem at hand. However, from the tables it also becomes clear that all sets of enrichments based on boundary hats can be used to reduce the relative initial error of 8% to 2.62%. For all boundary data functions considered, the first four enrichments corresponding to the

ℓ	pd	nn 0	nn 2	nn 4	nn 6	nn 8	nn 9	nn 10	nn 20
2	1	8.00%	7.16%	2.93%	2.83%	2.75%	2.75%	2.74%	2.62%
3	1	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.73%	2.62%
4	1	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.73%	2.62%
5	1	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.73%	2.62%
6	1	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.73%	2.62%
7	1	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.73%	2.62%

Table 6.54: Development of the relative energy error for increasing numbers of enrichments obtained from linear boundary hats in x_1 and x_2.

ℓ	pd	nn 0	nn 2	nn 4	nn 6	nn 8	nn 9	nn 10	nn 20
2	2	8.00%	7.16%	2.92%	2.82%	2.77%	2.76%	2.75%	2.62%
3	2	8.00%	7.16%	2.96%	2.88%	2.78%	2.74%	2.73%	2.62%
4	2	8.00%	7.16%	2.96%	2.89%	2.78%	2.73%	2.73%	2.62%
5	2	8.00%	7.16%	2.96%	2.89%	2.78%	2.73%	2.73%	2.62%
6	2	8.00%	7.16%	2.96%	2.89%	2.78%	2.73%	2.73%	2.62%
7	2	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.73%	2.62%
8	2	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.73%	2.62%

Table 6.55: Development of the relative energy error for increasing numbers of enrichments obtained from quadratic boundary hats in x_1 and x_2.

ℓ	pd	nn 0	nn 2	nn 4	nn 6	nn 8	nn 9	nn 10	nn 20
2	0, 1	8.00%	7.16%	2.94%	2.86%	2.78%	2.74%	2.74%	2.62%
3	0, 1	8.00%	7.16%	2.96%	2.88%	2.78%	2.73%	2.73%	2.62%
4	0, 1	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.73%	2.62%
5	0, 1	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.73%	2.62%
6	0, 1	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.72%	2.62%
7	0, 1	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.72%	2.62%
8	0, 1	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.72%	2.62%

Table 6.56: Development of the relative energy error for increasing numbers of enrichments obtained from constant and linear boundary hats in x_1 and x_2.

four largest eigenvalues reduce the relative error to 2.97%, which is a 62.9% reduction. Additional enrichments only slightly improve the results. Note that in Section 6.3.2, 8 dominant eigenvalues were identified. However, the threshold of 0.1 was chosen generically, i.e. from empirical observations of many PDE, and did not consider individual properties of the underlying partial differential operator. In Figure 6.34, the distribution of the energy error for various numbers of enrichments is shown visually. The decay of the errors is shown in Figure 6.35

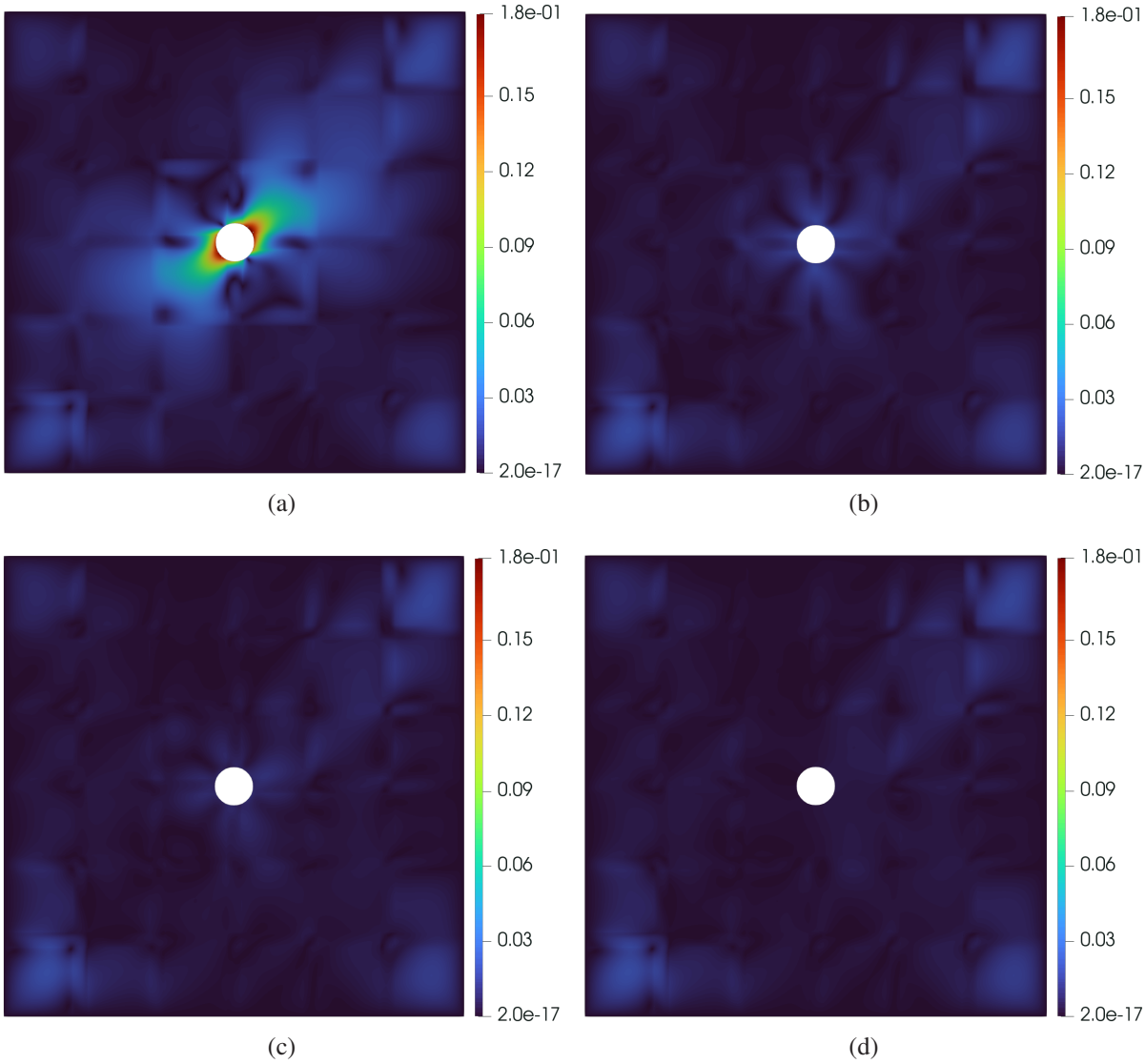


Figure 6.34: Magnitude of the difference between reference and enriched solutions, for (a) only particular solutions, (b) particular solutions and four enrichments, (c) particular solutions and ten enrichments, (d) particular solutions and twenty enrichments. Enrichments were constructed from boundary hats on level 8 with $pd = \{0, 1\}$

B-Splines

The accuracy of enriched solutions using enrichments constructed from B-Splines boundary data in the sampling problem is analyzed. Tables 6.57 to 6.60 present the obtained relative energy errors.

Again, all sets of enrichments could significantly reduce the initial error of 8% to 2.62%. However, there were significant differences in the speed of decay of the error in terms of the number of enrichments. The approach using B-Splines in x_1 and x_2 together with corner splines has the best performance and is able to reduce the error to 2.97% with only four enrichments. The same number of enrichments leads to errors of 3.08% when corner splines are omitted. When B-Splines only in x_1 direction are considered, the corresponding errors are $\approx 4.2\%$ with corner splines, and

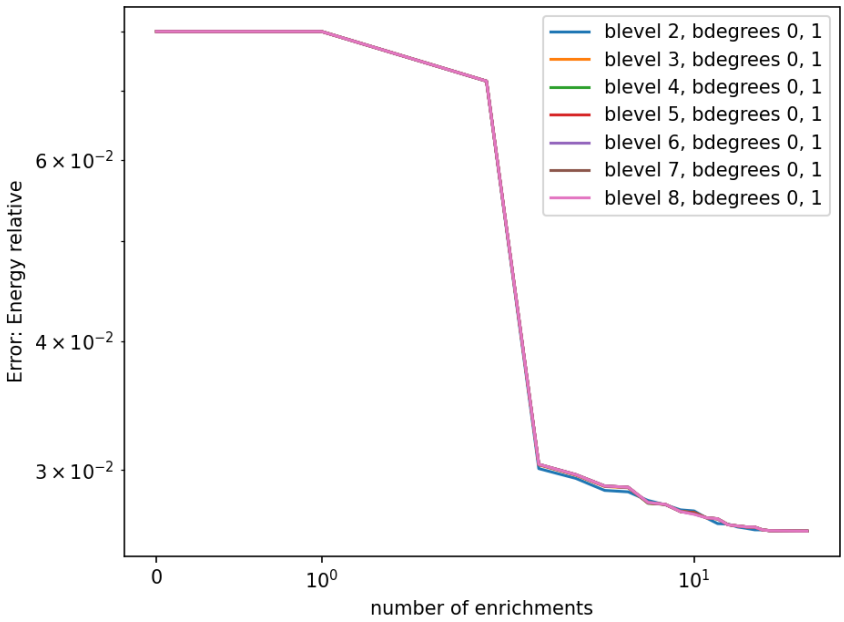


Figure 6.35: Relative energy error for increasing numbers of enrichments, constructed from constant and linear boundary hats and defined on various boundary levels.

n	nn 0	nn 2	nn 4	nn 6	nn 8	nn 9	nn 10	nn 20
3	8.00%	7.25%	5.72%	4.92%	3.72%	3.39%	3.12%	2.62%
4	8.00%	7.34%	5.87%	5.13%	3.73%	3.39%	3.16%	2.62%
5	8.00%	7.35%	5.89%	5.05%	3.71%	3.29%	3.13%	2.62%
6	8.00%	7.36%	5.90%	5.06%	3.71%	3.29%	3.13%	2.62%
7	8.00%	7.36%	5.90%	5.05%	3.71%	3.27%	3.14%	2.62%
8	8.00%	7.36%	5.90%	5.04%	3.71%	3.27%	3.14%	2.62%
9	8.00%	7.36%	5.90%	5.04%	3.71%	3.27%	3.14%	2.62%

Table 6.57: Development of the relative energy error for increasing numbers of enrichments obtained from quadratic B-Splines in x_1 direction without corner splines.

n	nn 0	nn 2	nn 4	nn 6	nn 8	nn 9	nn 10	nn 20
3	8.00%	6.38%	4.17%	3.77%	3.09%	3.03%	2.87%	2.62%
4	8.00%	6.37%	4.20%	3.70%	3.09%	3.02%	3.01%	2.62%
5	8.00%	6.37%	4.20%	3.68%	3.10%	3.02%	3.01%	2.62%
6	8.00%	6.37%	4.20%	3.63%	3.10%	3.03%	3.02%	2.62%
7	8.00%	6.37%	4.20%	3.61%	3.10%	3.03%	3.02%	2.62%
8	8.00%	6.37%	4.20%	3.60%	3.10%	3.03%	3.02%	2.62%
9	8.00%	6.37%	4.20%	3.60%	3.10%	3.03%	3.02%	2.62%

Table 6.58: Development of the relative energy error for increasing numbers of enrichments obtained from quadratic B-Splines in x_1 direction with corner splines.

n	nn 0	nn 2	nn 4	nn 6	nn 8	nn 9	nn 10	nn 20
3	8.00%	7.17%	3.06%	2.97%	2.77%	2.73%	2.72%	2.62%
4	8.00%	7.17%	3.07%	2.97%	2.78%	2.73%	2.70%	2.62%
5	8.00%	7.17%	3.08%	2.97%	2.78%	2.73%	2.72%	2.62%
6	8.00%	7.17%	3.08%	2.97%	2.78%	2.73%	2.72%	2.62%
7	8.00%	7.17%	3.08%	2.97%	2.78%	2.73%	2.72%	2.62%
8	8.00%	7.17%	3.08%	2.97%	2.78%	2.73%	2.72%	2.62%
9	8.00%	7.17%	3.08%	2.97%	2.78%	2.73%	2.72%	2.62%

Table 6.59: Development of the relative energy error for increasing numbers of enrichments obtained from quadratic B-Splines in x_1 and x_2 direction without corner splines.

n	nn 0	nn 2	nn 4	nn 6	nn 8	nn 9	nn 10	nn 20
3	8.00%	7.16%	2.94%	2.87%	2.77%	2.74%	2.73%	2.62%
4	8.00%	7.16%	2.96%	2.88%	2.78%	2.74%	2.73%	2.62%
5	8.00%	7.16%	2.96%	2.88%	2.78%	2.73%	2.73%	2.62%
6	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.73%	2.62%
7	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.73%	2.62%
8	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.73%	2.62%
9	8.00%	7.16%	2.97%	2.89%	2.78%	2.73%	2.73%	2.62%

Table 6.60: Development of the relative energy error for increasing numbers of enrichments obtained from quadratic B-Splines in x_1 and x_2 direction with corner splines.

$\approx 5.9\%$ without corner splines. For all sets of enrichments, there is hardly any more change for more than 10 enrichments. Also note, that the number of pairwise different inner knots did not have a large impact on the decay (and speed of decay) of the error. Similar to the boundary hats case, not all enrichments corresponding to dominant eigenvalues need to be employed to obtain accurate results, and again this may be due to the value 0.1 of the threshold.

Figure 6.36 presents a visualization of the difference between enriched and reference solutions for various numbers of enrichments, and it is clearly visible that the magnitude in ω, i.e. near the periphery of the hole, is reduced. The enrichments that were used had been constructed from B-Splines in x_1 and x_2 direction with corner splines and 9 pairwise different inner knots. The difference for 20 enrichments is mainly located outside of ω and is due to the coarse global discretization. Figure 6.37 shows a graphical representation of the relative energy error.

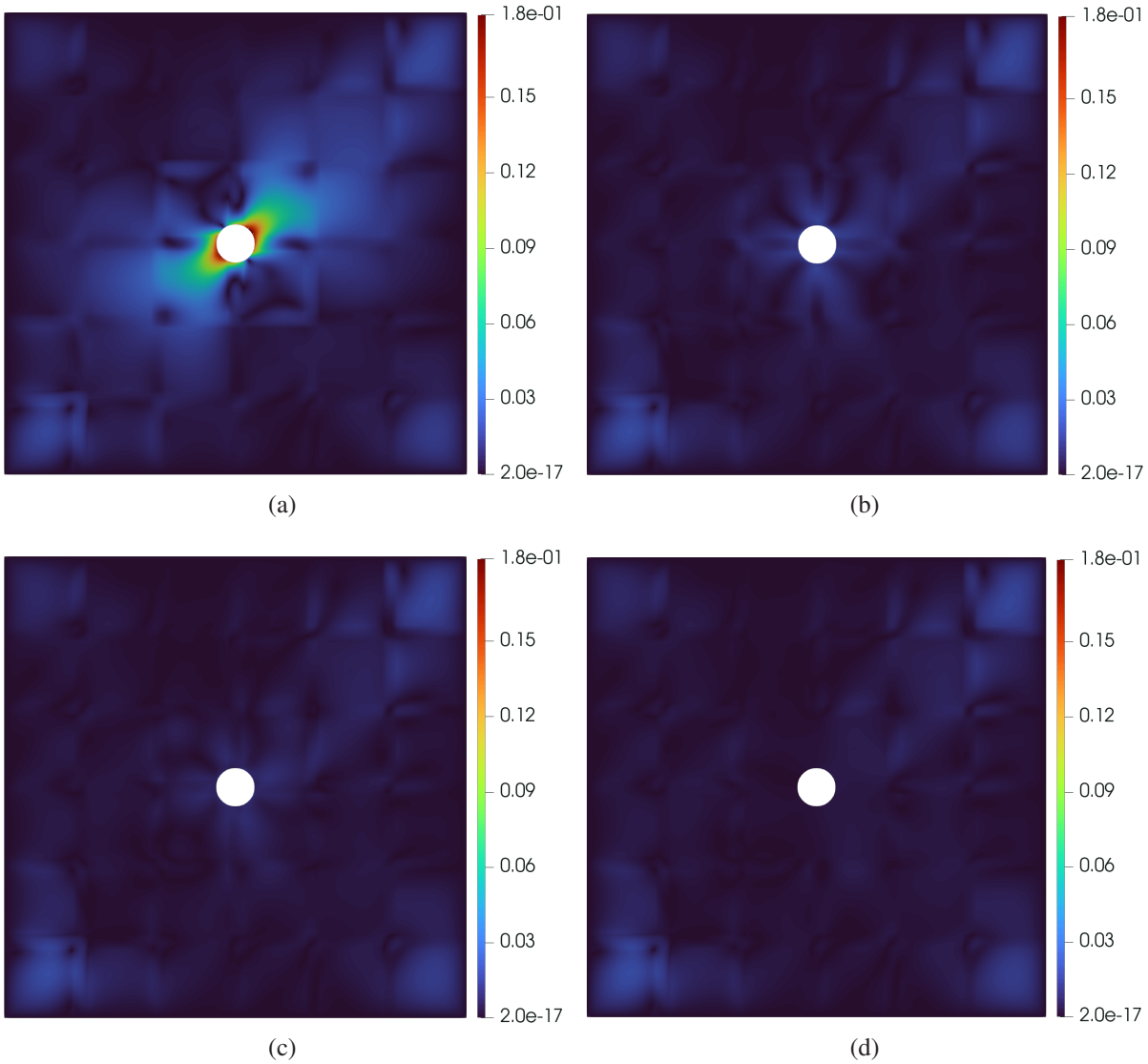


Figure 6.36: Magnitude of the difference between reference and enriched solutions, for (a) only particular solutions, (b) particular solutions and four enrichments, (c) particular solutions and ten enrichments, (d) particular solutions and twenty enrichments. Enrichments were computed from quadratic B-Splines in x_1 and x_2 direction, including the corner splines, for 9 different inner knots.

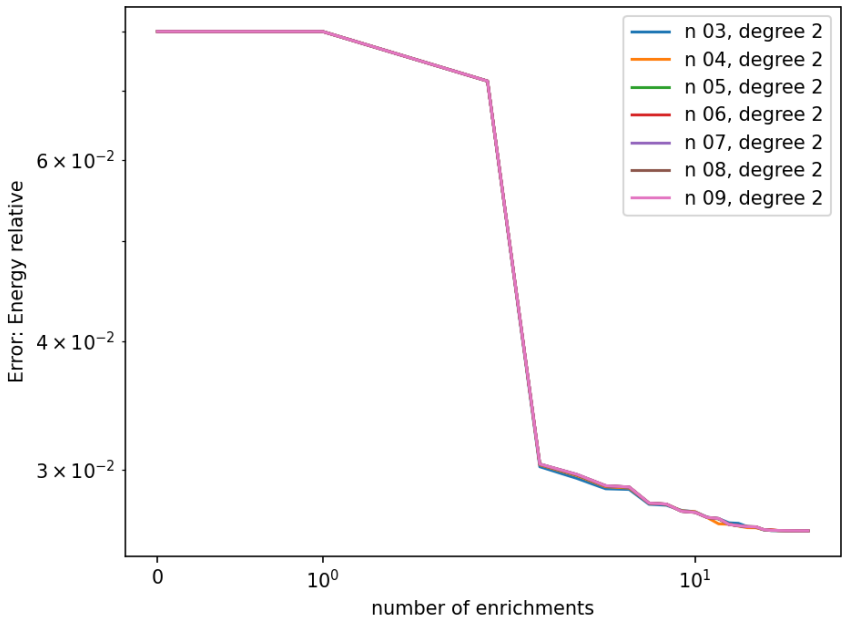


Figure 6.37: Relative energy error for increasing numbers of enrichments, constructed from quadratic B-Spline boundary data in x_1 and x_2, including corner splines, and defined for various numbers of inner knots (without repetitions).

Oscillating trigonometric functions

In Table 6.61, the development of the relative energy error when using enrichments constructed from Fourier-type boundary data in the sampling problem is shown. For a maximum number of oscillations of at least 5, all sets of enrichments are capable of reducing the error from 8% to 2.62%. From the discussion in Section 6.3.2 it was seen, that the magnitude of the largest eigenvalues is far smaller than the magnitude of the largest eigenvalues from the boundary hats or B-Spline approach, and this predicts the slower rate of decay of the relative energy error in the global computations. Increasing the number of maximum oscillations does - if it is at least 5 - not have any influence on the results. For completeness, the difference between reference and enriched solutions is shown

n_{osc}	nn 0	nn 2	nn 4	nn 6	nn 8	nn 9	nn 10	nn 20
2	8.00%	7.39%	6.34%	3.69%	3.53%	3.30%	2.93%	2.72%
5	8.00%	7.38%	6.31%	3.57%	3.43%	3.23%	3.07%	2.62%
10	8.00%	7.38%	6.31%	3.56%	3.42%	3.23%	3.07%	2.62%
15	8.00%	7.38%	6.31%	3.57%	3.42%	3.22%	3.06%	2.62%
20	8.00%	7.38%	6.31%	3.56%	3.42%	3.22%	3.06%	2.62%
25	8.00%	7.38%	6.31%	3.56%	3.42%	3.22%	3.06%	2.62%
30	8.00%	7.38%	6.31%	3.56%	3.42%	3.22%	3.06%	2.62%
35	8.00%	7.38%	6.31%	3.56%	3.42%	3.22%	3.06%	2.62%
40	8.00%	7.38%	6.31%	3.56%	3.42%	3.22%	3.06%	2.62%
45	8.00%	7.38%	6.31%	3.56%	3.42%	3.22%	3.06%	2.62%
50	8.00%	7.38%	6.31%	3.56%	3.42%	3.22%	3.06%	2.62%
55	8.00%	7.38%	6.31%	3.56%	3.42%	3.22%	3.06%	2.62%
60	8.00%	7.38%	6.31%	3.56%	3.42%	3.22%	3.06%	2.62%
65	8.00%	7.38%	6.31%	3.56%	3.42%	3.22%	3.06%	2.62%
70	8.00%	7.38%	6.31%	3.56%	3.42%	3.22%	3.06%	2.62%
75	8.00%	7.38%	6.31%	3.56%	3.42%	3.22%	3.06%	2.62%
80	8.00%	7.38%	6.31%	3.56%	3.42%	3.22%	3.06%	2.62%
85	8.00%	7.38%	6.31%	3.56%	3.42%	3.22%	3.06%	2.62%

Table 6.61: Development of the relative energy error for increasing numbers of enrichments obtained from Fourier-type basis functions in x_1 and x_2 direction with increasing number of maximum oscillations.

in Figure 6.38 for various numbers of enrichments. For 20 enrichments, the difference is mainly located outside of ω and due to the coarse global discretization. Finally, Figure 6.39 visualizes the decay of the relative energy error for increasing numbers of enrichments.

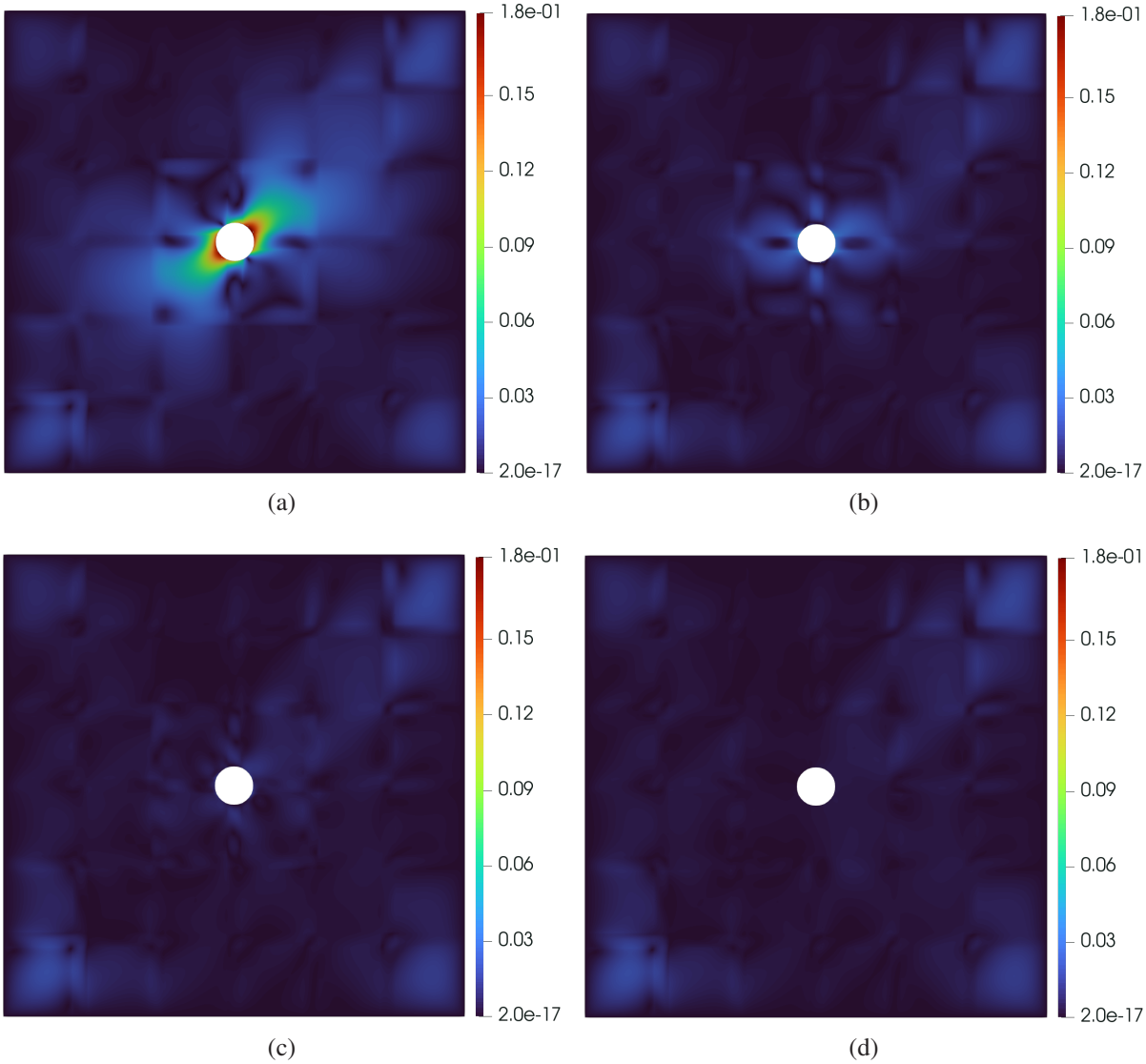


Figure 6.38: Magnitude of the difference between reference and enriched solutions, for (a) only particular solutions, (b) particular solutions and four enrichments, (c) particular solutions and ten enrichments, (d) particular solutions and twenty enrichments. Enrichments were computed from Fourier-type boundary data in x_1 and x_2 direction, for a maximum of 85 oscillations.

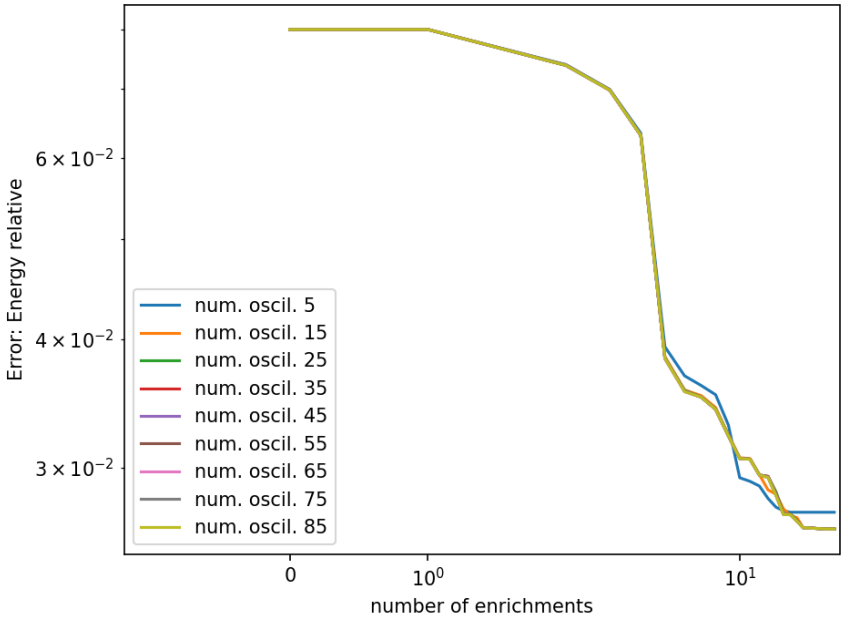


Figure 6.39: Relative energy error for increasing numbers of enrichments, constructed from Fourier-type boundary data in x_1 and x_2 direction for various maximum numbers of oscillations.

Conclusions from the benchmark problem

The conclusions of this benchmark problem mainly coincide with the conclusions of the previous benchmark problems. Ultimately, all sets of enrichments could be used to reduce the initial relative energy error of 8% to $\approx 2.62\%$. Using boundary hats in the sampling problem is the most expensive approach, but also the most versatile, since they allow for very fine scale variations of the enrichments. The B-Spline approach is the cheapest, since it leads to relatively small numbers of boundary data functions. Using B-Splines in both coordinate directions, and including corners splines, proved to be essential for obtaining large numbers and magnitudes of the dominant eigenvalues, and hence good performance in global computations. For sufficiently high boundary level, resp. number of inner knots, 8 dominant eigenvalues were captured. The global, enriched computations revealed, that not necessarily all enrichments corresponding to these 8 dominant eigenvalues are essential for small energy errors. This may be due to the empirically chosen threshold of 0.1 for the magnitude of eigenvalues in order to be called 'dominant'. Also note, that enrichments based on constant boundary hats did not perform as well as enrichments based on linear or quadratic boundary hats in this benchmark problem: Using constant boundary hats, the magnitude of the dominant eigenvalues was far smaller, even for boundary level 8. Using Fourier-type boundary data, 4 dominant eigenvalues were captured, with the largest being significantly smaller than in the other approaches. As expected, the size of the eigenvalues is a good predictor for the achievable speed of decay when using the corresponding optimal shape functions as enrichments. B-Splines in x_1 and x_2 direction, which additionally contained corner splines and were defined on 9 pairwise different inner knots, showed - again - the best overall performance, since they require relatively few solutions of the sampling problem, yield dominant eigenvalues of large magnitude, and ultimately show very good approximation qualities in global computations.

6.4 Isotropic linear elasticity in 3d

This section investigates the problem of Linear Elasticity posed on a on a rectangular three-dimensional domain having a circular hole in the center. The material is assumed to be isotropic and consists of two plies, the lower ply made of steel and the upper ply made of aluminum. The corresponding moduli of elasticity are $E = 210\,\mathrm{GPa}$ resp. $E = 70\,\mathrm{GPa}$, and the Poisson ratios are $\nu = 0.3$ resp. $\nu = 0.28$. As in the previous benchmark problem, no boundary condition is prescribed on the interior boundary. While all previous benchmarks considered essential boundary conditions prescribed on the outer whole boundary, the displacement in this case is only prescribed on two faces of the domain. On all other faces of the domain, including the inner boundary, the solution satisfies zero Neumann boundary conditions, meaning that it is allowed to behave freely on these faces.

It is expected, that the displacement presents fine-scale behavior near the periphery of the hole, which is hard to grasp using standard basis functions on coarse patches. However, on patches that touch faces with boundary conditions of both types, additional problems will occur in the upcoming analysis. Compared to the previous three benchmark problems, the discussion of global errors will hence be performed differently for this benchmark problem.

In Section 6.4.1, the problem under study is introduced. In Section 6.4.2, the influence of the previously used boundary data types to generate optimal shape functions is analyzed. The Section 6.4.3 studies their performance in global simulations.

6.4.1 Problem formulation

The benchmark problem under study is presented in Problem 18.

Problem 18: Isotropic linear elasticity in 3d.

Let

$$\Omega = \left([-6,6]^2 \backslash B_{0.5}(0,0)\right) \times [0,0.2] \subset \mathbb{R}^3 \tag{6.45}$$

be a three-dimensional plate with a circular hole in the center. Also, let $\underline{\underline{\mathbf{C}}}$ be the piecewise stiffness tensor for the isotropic material defined by two plies of thickness 0.1 each. The moduli of elasticity and the Poisson ratii are given by $E = [210\,\mathrm{GPa}, 70\,\mathrm{GPa}]$ and $\nu = [0.30, 0.28]$. Finally, let $f(x) = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T$ and

$$\begin{aligned} g_l(x) &= \begin{bmatrix} \frac{-1}{36} x_2 & 0 & 0 \end{bmatrix} \\ g_r(x) &= \begin{bmatrix} \frac{-1}{36} x_2 & 0 & 0 \end{bmatrix}, \end{aligned} \tag{6.46}$$

as well as $g : \partial\,\Omega \backslash \partial(B_{0.5}(0,0) \times [0,0.2]) \to \mathbb{R}^3$ with

$$x \mapsto g(x) := \begin{cases} g_l, & \text{if } x_1 = -6 \\ g_r, & \text{if } x_1 = 6. \end{cases} \tag{6.47}$$

Consider the differential operator

$$\mathcal{L} : [\mathcal{C}^2(\Omega)]^3 \to [\mathcal{C}^0(\Omega)]^3, \quad u \mapsto \mathcal{L}\,u := -\operatorname{div}(\sigma(u)). \tag{6.48}$$

Find a function $u \in [\mathcal{C}^2(\Omega)]^3$ satisfying

$$\begin{aligned} -\mathcal{L}\,u(x) &= f(x), \quad \text{in } \Omega \\ u(x) &= g(x), \quad \text{on } \Gamma_{\text{left}} \cup \Gamma_{\text{right}} := \{x \in \Omega : \ |x_1| = 6\}, \end{aligned} \tag{6.49}$$

with the stress tensor $\sigma(u) = \underline{\underline{\mathbf{C}}} : \varepsilon(u)$ and $\varepsilon(u)$ is the infinitesimal strain tensor $\varepsilon(u) = \frac{1}{2}\left(\nabla u + (\nabla u)^T\right)$.

This benchmark problem is an extruded version of the previous benchmark problem to three-dimensional space. Due to Section 3.2, the differential operator is elliptic. The weak formulation is posed on the trial and test spaces

$$\begin{aligned} \mathrm{V}^{\text{trial}} &:= \{u \in [H^1(\Omega)]^3 : \ \operatorname{tr}(u) = g, \ \text{on } \Gamma_{\text{left}} \cup \Gamma_{\text{right}}\} \\ \mathrm{V}^{\text{test}} &:= \{u \in [H^1(\Omega)]^3 : \ \operatorname{tr}(u) = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T, \ \text{on } \Gamma_{\text{left}} \cup \Gamma_{\text{right}}\} \end{aligned} \tag{6.50}$$

and reads

Problem 19: Weak isotropic linear elasticity in 3d.

Let

$$\Omega = \left([-6,6]^2 \backslash B_{0.5}(0,0)\right) \times [0,0.2] \subset \mathbb{R}^3 \tag{6.51}$$

be a three-dimensional plate with a circular hole in the center. Also, let $\underline{\underline{\mathbf{C}}}$ be the piecewise stiffness tensor for the isotropic material defined by two plies of thickness 0.1 each. The moduli of elasticity and the Poisson ratii are given by $E = [210\,\mathrm{GPa}, 70\,\mathrm{GPa}]$ and $\nu = [0.30, 0.28]$. Finally, let $f(x) = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T$ and

$$\begin{aligned} g_l(x) &= \begin{bmatrix} \frac{-1}{36} x_2 & 0 & 0 \end{bmatrix} \\ g_r(x) &= \begin{bmatrix} \frac{-1}{36} x_2 & 0 & 0 \end{bmatrix}, \end{aligned} \tag{6.52}$$

as well as $g : \partial\,\Omega \backslash \partial(B_{0.5}(0,0) \times [0,0.2]) \to \mathbb{R}^3$ with

$$x \mapsto g(x) := \begin{cases} g_l, & \text{if } x_1 = -6 \\ g_r, & \text{if } x_1 = 6. \end{cases} \tag{6.53}$$

Define the bilinear form $\mathbf{a} : \mathrm{V}^{\text{trial}}(\Omega) \times \mathrm{V}^{\text{test}}(\Omega) \to \mathbb{R}^2$ and linear functional $\ell : \mathrm{V}^{\text{test}} \to \mathbb{R}$,

$$\begin{aligned} \mathbf{a}[u,v] &:= \int_\Omega \sigma(u) : \varepsilon(v)\, dx \\ \ell(v) &:= \int_\Omega f \cdot v\, dx. \end{aligned} \tag{6.54}$$

Find a function $u \in \mathrm{V}^{\text{trial}}(\Omega)$ satisfying

$$\mathbf{a}[u,v] = \ell(v), \quad \forall v \in \mathrm{V}^{\text{test}}(\Omega). \tag{6.55}$$

Theorem 3.2 from Section 3.2.2 showed, that the bilinear form $\mathbf{a}$ for the isotropic two-ply material under study is continuous and elliptic in the pure Dirichlet problem. In the problem at hand, the function u is allowed to behave freely on the lower and upper face of the domain. This is modeled by a zero Neumann condition of the form

$$\sigma(u) \cdot \vec{n} = 0, \tag{6.56}$$

which is implicitly assumed in Problem 18. Ellipticity and boundedness are shown similar to Theorem 3.2 using another variant of Poincaré inequality. The linear functional ℓ is obviously continuous and the Lax-Milgram Theorem 2.4 hence ensures existence and uniqueness of solutions.

6.4.2 Influence of the boundary data

This section investigates the influence of the choice of boundary data used in the sampling problem on the number of dominant eigenvalues and their magnitude. The implications of using boundary

hats and B-Splines as boundary data on the eigenvalues corresponding to the optimal shape functions are investigated. The previous benchmark problems also considered oscillatory Fourier-type boundary data, but the resulting performance was relatively poor. This approach will therefore not be investigated in the scope of the benchmark problem at hand. For similar reasons, only the approach considering B-Splines in x_1 and x_2 direction together with corner splines will be considered, since it previously outperformed the other approaches.

Boundary hats

The following Tables 6.62 - 6.65 present the results for constant, linear and quadratic boundary hats. For a sufficiently large boundary level, six dominant eigenvalues and 9 additional large

ℓ	pd	dim	λ_1	λ_2	λ_8	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	0	24	$2.2 \cdot 10^{-1}$	$2.16 \cdot 10^{-1}$	$1.962 \cdot 10^{-2}$	5	11	20	24
3	0	56	$3.319 \cdot 10^{-1}$	$2.986 \cdot 10^{-1}$	$3.731 \cdot 10^{-2}$	7	14	28	52
4	0	120	$4.168 \cdot 10^{-1}$	$3.675 \cdot 10^{-1}$	$5.97 \cdot 10^{-2}$	7	15	29	58
5	0	248	$4.791 \cdot 10^{-1}$	$4.112 \cdot 10^{-1}$	$7.991 \cdot 10^{-2}$	7	15	29	60
6	0	504	$5.175 \cdot 10^{-1}$	$4.336 \cdot 10^{-1}$	$9.234 \cdot 10^{-2}$	7	15	31	60
7	0	1016	$5.386 \cdot 10^{-1}$	$4.443 \cdot 10^{-1}$	$9.692 \cdot 10^{-2}$	8	15	31	60

Table 6.62: Key numbers obtained from the computation of the optimal shape functions using constant boundary hats on various boundary levels.

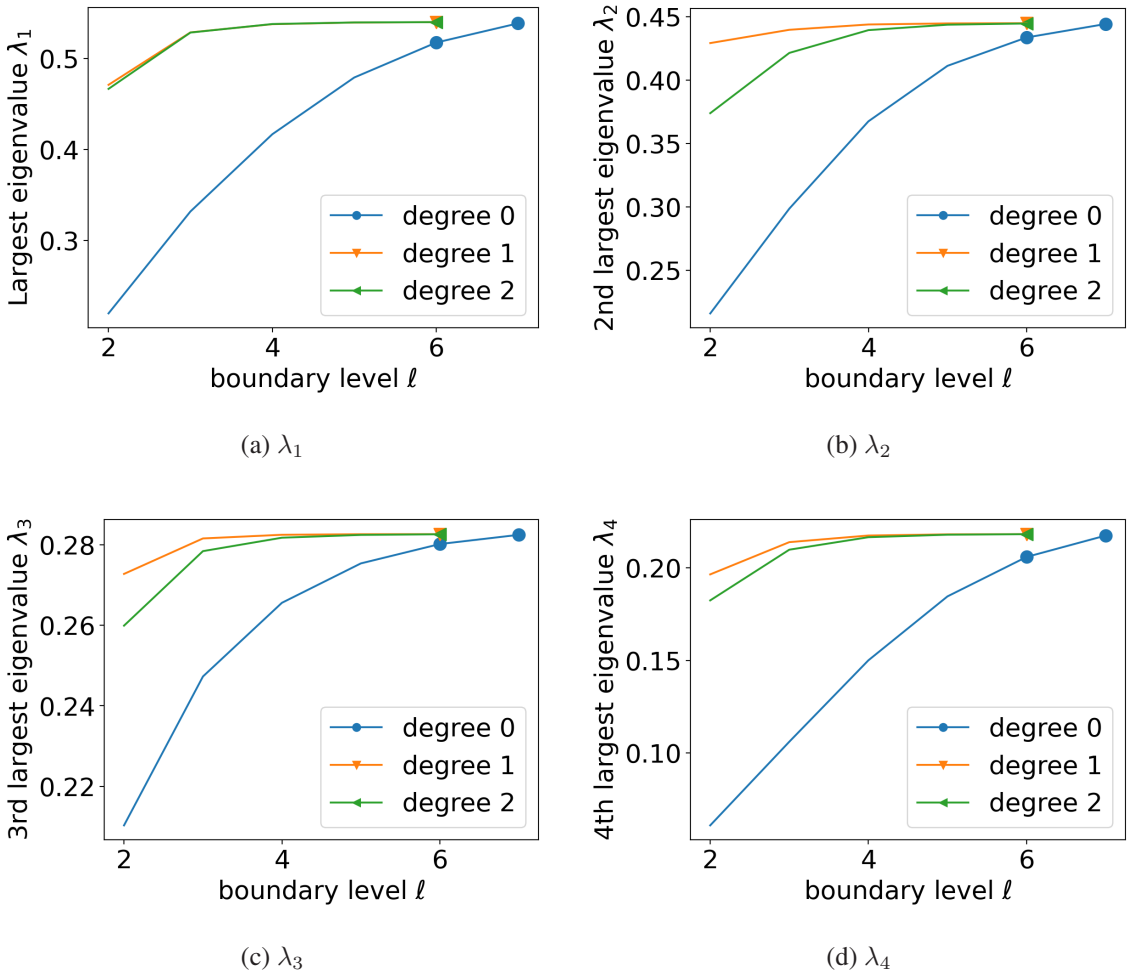
ℓ	pd	dim	λ_1	λ_2	λ_8	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	1	48	$4.711 \cdot 10^{-1}$	$4.292 \cdot 10^{-1}$	$7.214 \cdot 10^{-2}$	5	13	27	47
3	1	112	$5.287 \cdot 10^{-1}$	$4.397 \cdot 10^{-1}$	$9.604 \cdot 10^{-2}$	5	13	31	60
4	1	240	$5.377 \cdot 10^{-1}$	$4.439 \cdot 10^{-1}$	$9.71 \cdot 10^{-2}$	6	13	31	60
5	1	496	$5.396 \cdot 10^{-1}$	$4.447 \cdot 10^{-1}$	$9.723 \cdot 10^{-2}$	6	15	31	60
6	1	1008	$5.4 \cdot 10^{-1}$	$4.449 \cdot 10^{-1}$	$9.725 \cdot 10^{-2}$	6	15	31	61

Table 6.63: Key numbers obtained from the computation of the optimal shape functions using linear boundary hats in x_1 and x_2 direction on various boundary levels.

ℓ	pd	dim	λ_1	λ_2	λ_8	n_{-1}	n_{-2}	n_{-4}	n_{-8}
2	2	48	$4.667 \cdot 10^{-1}$	$3.739 \cdot 10^{-1}$	$7.146 \cdot 10^{-2}$	5	13	25	48
3	2	112	$5.285 \cdot 10^{-1}$	$4.214 \cdot 10^{-1}$	$9.158 \cdot 10^{-2}$	5	13	31	60
4	2	240	$5.379 \cdot 10^{-1}$	$4.394 \cdot 10^{-1}$	$9.606 \cdot 10^{-2}$	6	13	31	60
5	2	496	$5.396 \cdot 10^{-1}$	$4.437 \cdot 10^{-1}$	$9.701 \cdot 10^{-2}$	6	13	31	60
6	2	1008	$5.4 \cdot 10^{-1}$	$4.447 \cdot 10^{-1}$	$9.721 \cdot 10^{-2}$	6	15	32	61

Table 6.64: Key numbers obtained from the computation of the optimal shape functions using quadratic boundary hats in x_1 and x_2 direction on various boundary levels.

eigenvalues are captured. For all boundary hat degrees considered, the dominant eigenvalues stabilize at similar values of around 0.54, 0.44, 0.28, 0.22, 0.22 and 0.10 (the last one is slightly larger than the threshold). Figure 6.40 visualizes the development of the first four eigenvalues. Ultimately, constant, linear, as well as quadratic boundary hats can be used to capture dominant



(a) λ_1 (b) λ_2

(c) λ_3 (d) λ_4

Figure 6.40: Development of the four largest eigenvalues for increasing boundary levels and all considered boundary hats degrees. Markers are plotted every 500 degrees of freedom.

eigenvalues with large magnitudes. Since the number of degrees of freedom, i.e. the number of sampling problems to be solved should be as small as possible, the optimal shape functions constructed from linear and quadratic boundary hats seem to outperform the approach using constant boundary hats. In both cases, a boundary level of 4 or 5 seems to be sufficient, leading to less than 500 degrees of freedom.

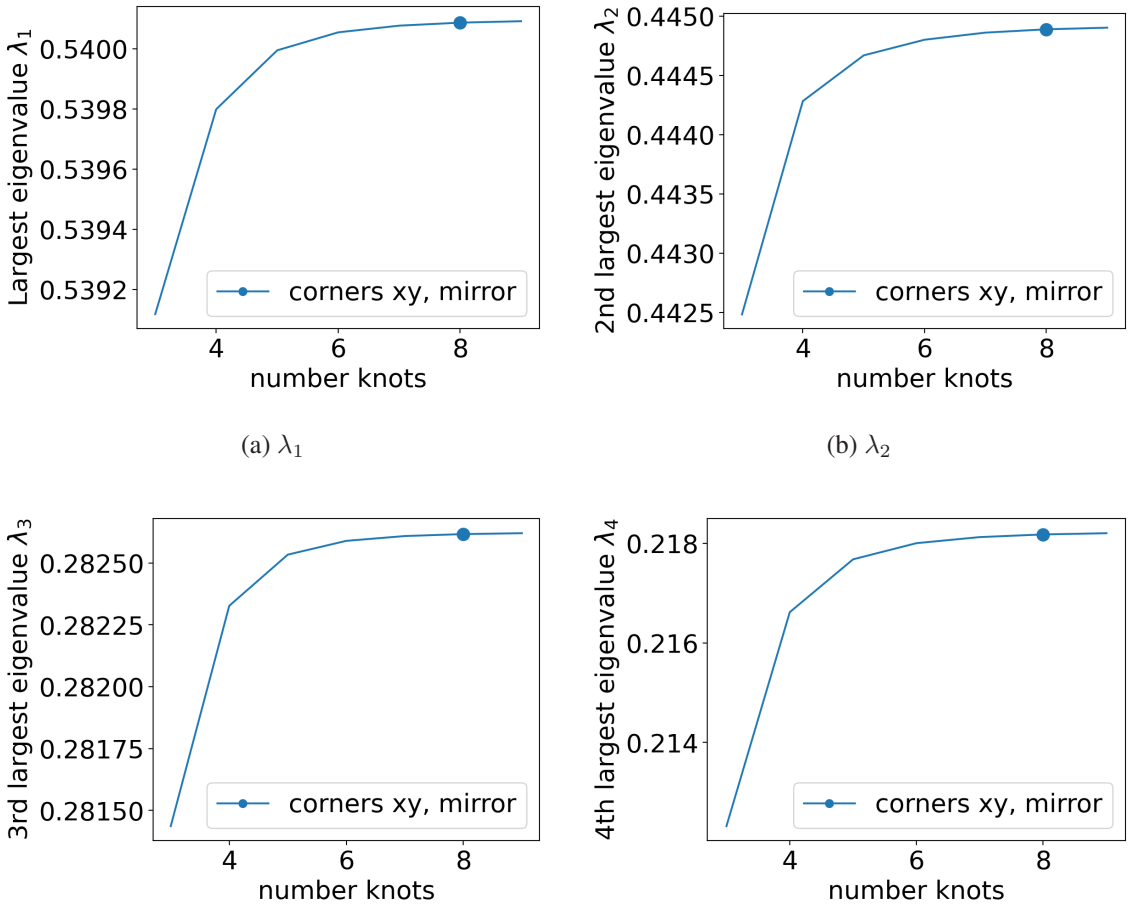
B-Splines

Table 6.65 presents key numbers obtained from the computation of the optimal shape functions when using B-Splines in x_1 and x_2 direction, and including corner splines, as boundary data in the sampling problem. The development of the first four eigenvalues for increasing numbers of

n	dim	λ_1	λ_2	λ_8	n_{-1}	n_{-2}	n_{-4}	n_{-8}
3	288	$1{\cdot}10^{0}$	$1{\cdot}10^{0}$	$2.123 \cdot 10^{-1}$	8	23	44	59
4	336	$1{\cdot}10^{0}$	$1{\cdot}10^{0}$	$2.166 \cdot 10^{-1}$	8	24	44	68
5	384	$1{\cdot}10^{0}$	$1{\cdot}10^{0}$	$2.177 \cdot 10^{-1}$	9	24	47	76
6	432	$1{\cdot}10^{0}$	$1{\cdot}10^{0}$	$2.18 \cdot 10^{-1}$	9	24	50	83
7	480	$1{\cdot}10^{0}$	$1{\cdot}10^{0}$	$2.181 \cdot 10^{-1}$	9	24	50	89
8	528	$1{\cdot}10^{0}$	$1{\cdot}10^{0}$	$2.182 \cdot 10^{-1}$	9	24	52	91
9	576	$1{\cdot}10^{0}$	$1{\cdot}10^{0}$	$2.182 \cdot 10^{-1}$	9	24	52	92

Table 6.65: Key numbers obtained from the computation of the optimal shape functions using B-Splines in x_1 and x_2 direction with corner splines.

distinct inner knots is shown in Figure 6.41. They stabilize at approximately the same values as the eigenvalues computed using boundary hats as boundary data.



(a) λ_1 (b) λ_2

(c) λ_3 (d) λ_4

Figure 6.41: Development of the four largest eigenvalues for increasing numbers of distinct inner knots. Markers are plotted every 500 degrees of freedom.

6.4.3 Discussion of global errors

In order to identify a numerical reference solution, the problem was solved for various levels of discretization. Since the domain under study is relatively thin in x_3 direction, it is discretized only in x_1 and x_2. The two material layers in x_3 direction are resolved implicitly using a second-order B-Spline basis. Consequently, the patch size h in the following refers to the patch size in x_1 resp. x_2 direction.

The energy norms of unenriched solutions on various discretization levels, as well as the extrapolated value for patch size $h = 0$ are shown in Figure 6.42. Since the energy norm for a discretiza-

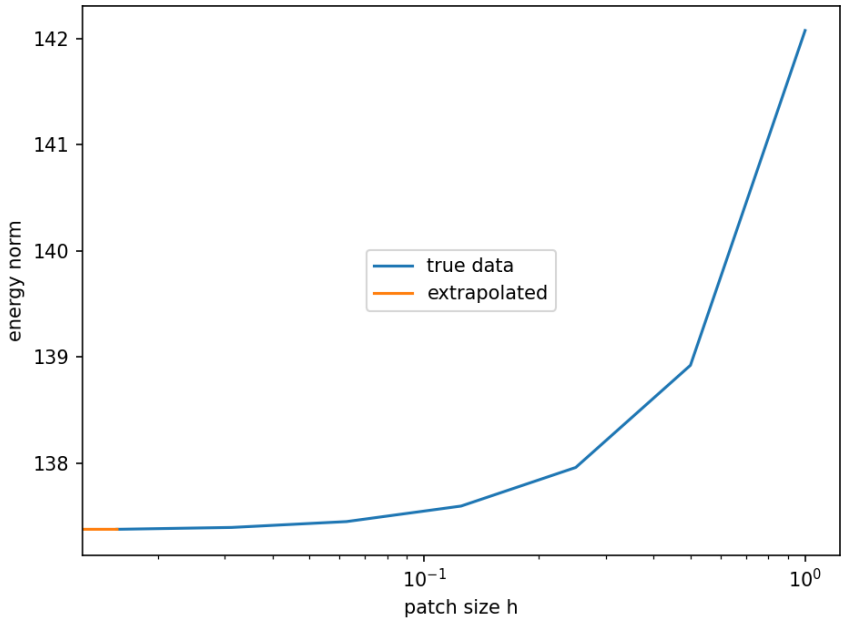


Figure 6.42: The energy of the solutions u_h for various various values of h. The extrapolated limit value at $h = 0$ is also shown.

tion on level $\iota = 8$, corresponding to a patch size of $h = 0.0625$ and the extrapolated value at $h = 0$ approximately coincide, the solution $u_{0.0625}$ will be used as a numerical reference solution in the following. The problem under study is subject to Dirichlet boundary conditions imposed on the left and right face, as well as zero Neumann boundary conditions on the other four outer faces as well as the interior boundary. As mentioned before, this allows the solution to behave freely among these parts of the boundary.

Similar to the previous benchmark problems, it is expected that the behavior of the solution at the interior boundary is hard to grasp, motivating the use of optimal basis functions in the center region. It will, however, turn out that the choice of boundary conditions leads to additional difficulties.

An unenriched solution using 6 patches in x_1 and x_2 direction was computed, and the top view of this coarse discretization is sketched in Figure 6.43. The original enrichment strategy is as follows: The local approximation spaces of all patches touching the Dirichlet boundary will be enriched with particular solutions for the boundary data, and the four center patches around the hole will further be enriched using optimal local basis functions.

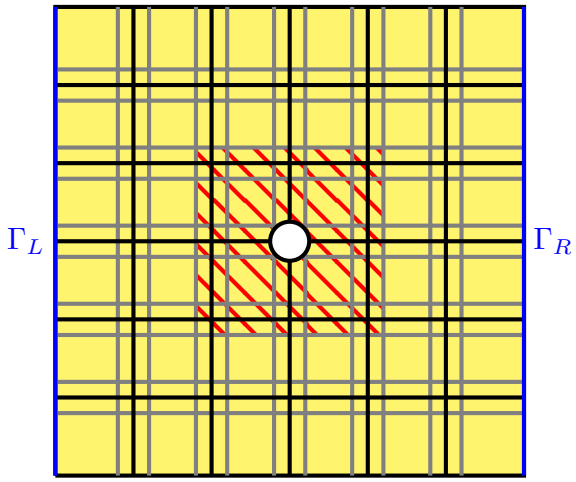


Figure 6.43: Top view of the coarse global discretization. Original enrichment strategy: Local particular solutions for essential boundary conditions (blue faces) are used on the corresponding patches, and the red patches will be further enriched using optimal shape functions. Only one patch is used through the thickness. The problem is 3 dimensional, and all patches touching the Dirichlet boundary also touch at least two zero Neumann boundaries.

The difference between the reference solution and the unenriched solution is shown in Figure 6.44. The largest magnitude of the error is observed in the x_1x_2 corner patches of the domain, not around the periphery of the circular hole as was expected. Hence, the enrichment strategy has to

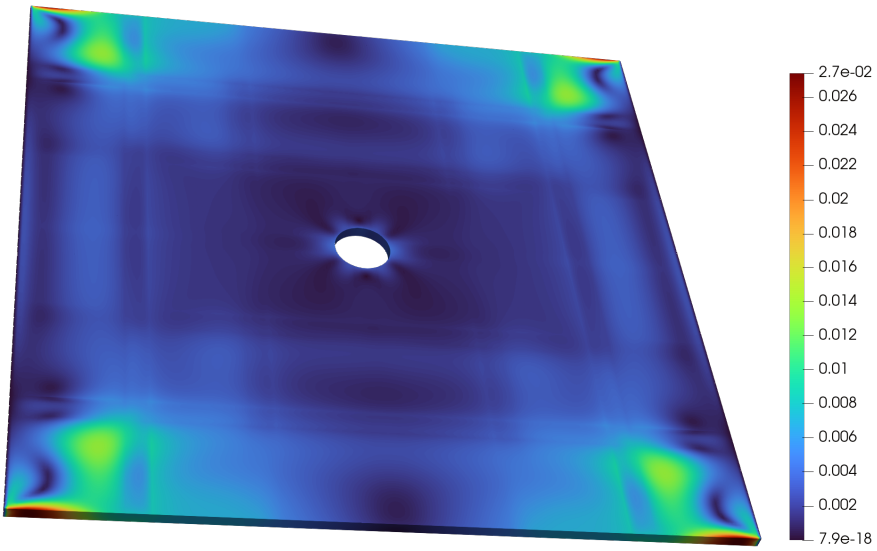


Figure 6.44: Magnitude of the difference between the reference solution and the unenriched solution. The difference is largest near the x_1x_2-vertices of the domain.

be modified. Note that due to the lifting of solutions, the homogeneous part of the solution must be included on all patches. For the previous benchmark problems, however, these homogeneous solutions were expected to behave nicely and hence be negligible. For the problem at hand, this is not the case, and before enriching the center patches near the periphery of the hole, the error at the x_1x_2-vertices of the domain must be reduced. As described, this is done by not only enriching the local spaces of the patches touching the Dirichlet boundary with a particular solution for g,

but also with corresponding sets of optimal basis functions. The free boundary of the patchwise local sampling problems, on which data needs to be prescribed, is formed by all interior edges of the patch which result from punching it out of the full domain Ω. A sketch of the free boundary for the front left patch is shown in Figure 6.45. For a sufficiently large amount of boundary data

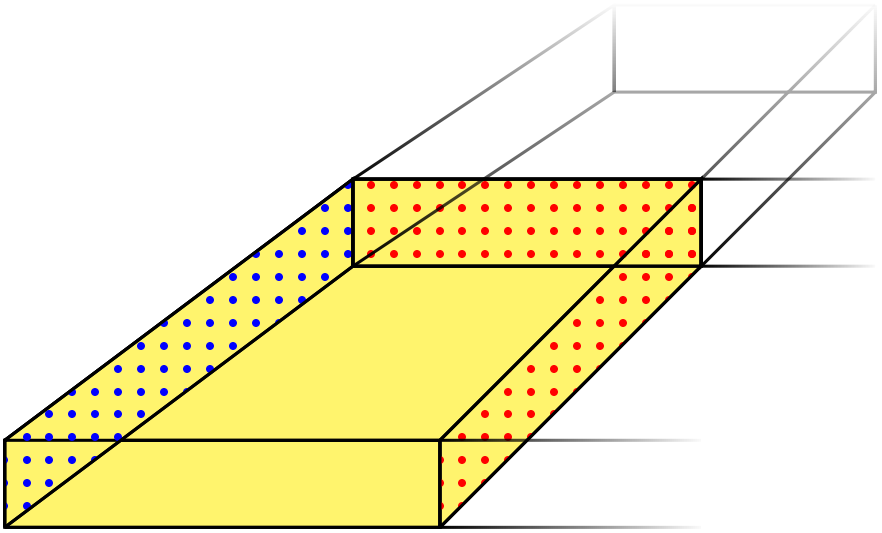
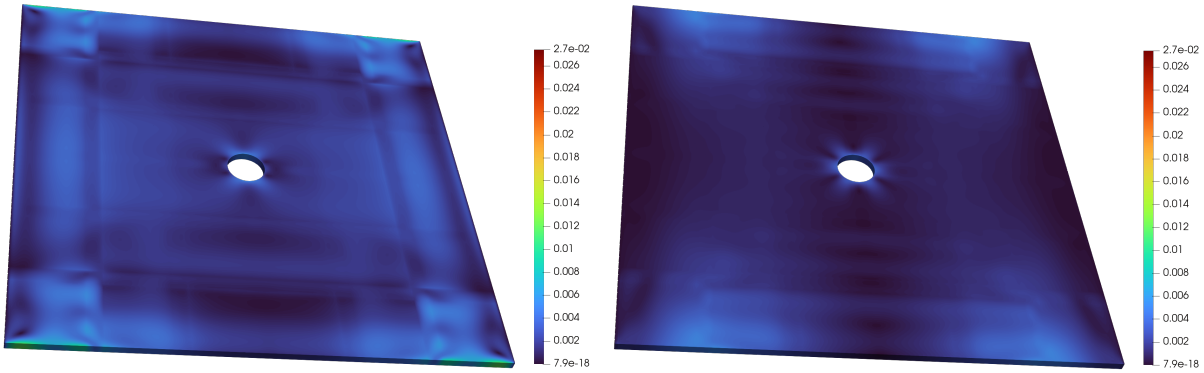


Figure 6.45: Sketch of the front left patch of the domain. The patch extends through the whole thickness. The face with global Dirichlet boundary conditions is marked with blue dots, zero Neumann faces are not marked (front, bottom and top faces). Red dots are used to mark the free boundary, i.e. artificial boundary obtained by punching ω out of Ω.

functions (cf. discussions from Sections 6.1 to 6.3), there are four dominant eigenvalues on the four corner patches, and six dominant eigenvalues on all other patches touching the Dirichlet boundary. Note that the front left corner patch can be shifted and rotated by 180° onto the rear right corner patch, and the rear left corner patch can be transformed similarly onto the front right corner patch. These transformations moreover correctly map the Dirichlet faces onto each other. Furthermore, all other (non-corner) patches touching the Dirichlet boundary are either a translation of each other, or a translation together with a rotation by 180°. In total, only three sets of optimal shape functions need to be computed due to the reusability results from Section 5.3. In Figure 6.46, the differences between the reference and solutions that are unenriched in the center and only use a particular solution for g, as well as one resp. six optimal basis functions on all patches touching the Dirichlet boundary, i.e. the left and right face of the plate, are shown. Using only one optimal basis function on each patch touching the Dirichlet boundary, the difference between the reference and enriched solutions near the x_1x_2 vertices is still dominant. For six optimal basis functions, the difference near the vertices and around the periphery of the hole is of the same magnitude. While the patches touching the Dirichlet boundary were enriched with optimal basis functions and a particular solution for g, their neighbors on the front resp. rear face were not enriched at all, leading to a significant jump in resolution. While the magnitude of the difference between both solutions is reduced significantly, it does not vanish entirely and the largest magnitude is now observed near the jump in resolution.



(a) (b)

Figure 6.46: Magnitude of the difference between the reference solution and solutions that do not use enrichments around the hole in the center, but particular solutions for g as well as (a) one and (b) six optimal shape functions on all patches touching the Dirichlet boundary.

Additionally to the particular solutions and six optimal basis functions for patches touching the Dirichlet boundary, 10 optimal basis functions are used in the center, and the results are shown in Figure 6.47. These functions are capable of reducing the local magnitude of the error around the periphery of the hole as expected.

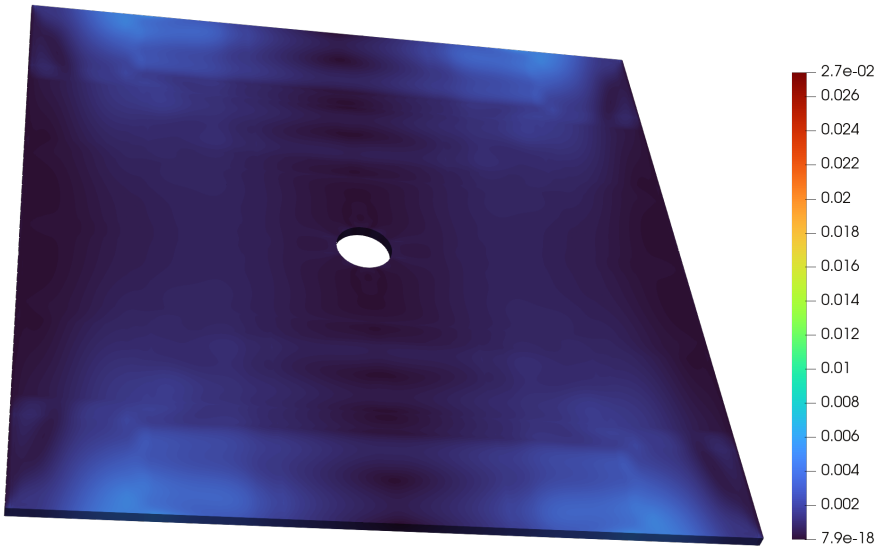


Figure 6.47: Difference between the reference solution and the solution using 10 optimal shape functions around the hole in the center, particular solutions for g and six optimal shape functions on all patches touching the Dirichlet boundary.

Similar to the previous benchmark problems, it was at first assumed that the solution behaves nicely away from a pre-identified region, on which fine-scale behavior was anticipated. Consequently, the homogeneous part of the solution appearing in the lifting was omitted on patches away from this region of interest. As long as the problem under study is subject to Dirichlet boundary conditions on the whole outer boundary, this tactic worked like a charm. In the current case of varying boundary conditions, however, the homogeneous part of the solution cannot be omitted in the lifting and must itself be approximated using optimal shape functions on the respective patches.

As shown in [Sch11], the global error of the enriched Partition of Unity Method is only expected to diminish when the approximation power is improved globally. The foregoing discussion showed, that no part of the lifted solution should be neglected unless the problem at hand is very well understood.

7

Numerical study of model problems

A full investigation of the various ways of constructing optimal shape functions, as well as their performance in global computations, was conducted in the previous Chapter 6, helping to gain a better understanding of the overall method. As a result, ways of construction, that probably lead to sets of optimal shape functions with a good performance in global enriched computations, were identified. This chapter considers problems which are too complex for a full investigation in the style of the benchmark problems. Furthermore, it may not even be possible to compute a reference solution to compare against. As described in Chapter 3, the method of constructing optimal shape functions investigated in this thesis is promising especially in the case of heterogeneous coefficients appearing in the considered partial differential equations. The coefficients may oscillate or even show discontinuities expressed in the form of jumps. Holes in the computational domain can be described by vanishing coefficients.

The main goal of this chapter is to show that problems, which are hard to solve and require millions of degrees of freedom in generic, traditional methods, become feasible when employing local enrichments in the form of optimal basis functions and particular solutions. These well-suited, operator-dependent enrichment functions replace the need for heavy spatial refinement, and can be used to significantly simplify the solution process. This chapter provides proof-of-concept examples, and all simulations were performed using the `PUMA` software toolkit developed by Fraunhofer SCAI ([SCA]) and the `optbasefun` module for `Python 3` developed in the scope of this thesis.

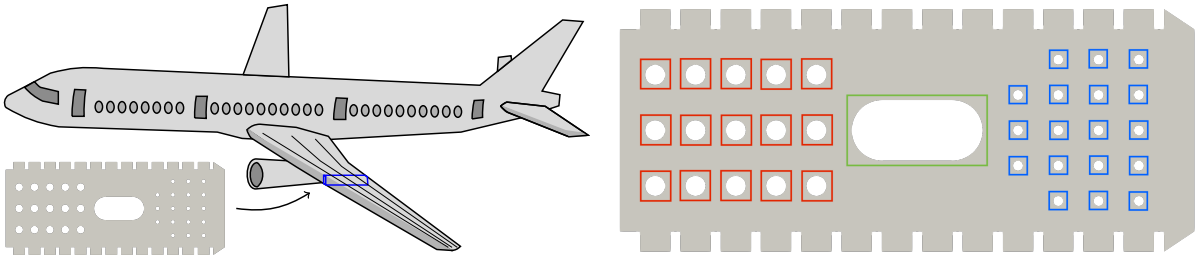
The problem in Section 7.1 investigates linear elasticity on the two-dimensional model of an airplane rib, a domain with many complicating features, among them 34 round holes and a large number of reentrant corners. In the experiment, analytical, as well as numerical enrichments in the form of optimal basis functions are employed. Since most of the circular holes have a similar structure, the results from Section 5.3.1 regarding geometric reusability of optimal basis functions are employed and thus validated numerically.

In Section 7.2, the propagation of waves due to an impact on a three-dimensional cube with 50 spherical inclusions is investigated by solving the equations of elastodynamics. The inclusions have different material properties from the matrix material, and this will lead to a reduction in the speed of propagation or even to local reflections of the propagating wave.

7.1 Orthotropic linear elasticity on an airplane rib

As mentioned and investigated before, it is numerically expensive to compute optimal basis functions. These basis functions are independent of the load and values of boundary conditions, and it was shown in Section 5.3.1 that they can also be transformed to other geometries as long as the coefficients satisfy some easy-to-check conditions. Moreover, it was mentioned that analytical enrichments should be used whenever available. All of these concepts are highlighted in the following use-case simulation.

An airplane has two wings, and each wing consists of multiple ribs which stabilize its shape. A sketch of the structure of a rib and its location in an airplane wing is sketched in Figure 7.1. All red patches can be translated onto each other. The blue patches in the right half of the domain are isotropically scaled and translated versions of the red patches. The green patch in the center has a porthole-shaped hole in its center. In order to further complicate things, there is a large



(a) (b)

Figure 7.1: (a) Sketch of an airplane and one of the ribs from its wings. (b) Detailed sketch of a rib. Subdomains marked in the same color have the same size and are translations of each other. Blue marked subdomains in the right half of the rib are a shrunken and translated version of the red marked subdomains from the left half of the rib.

number of reentrant corners on the outer periphery of the rib structure. All of the previously described complicating features need to be resolved by the employed discretization, and using adaptive spatial refinement towards all of the complicating features basically refines the whole domain.

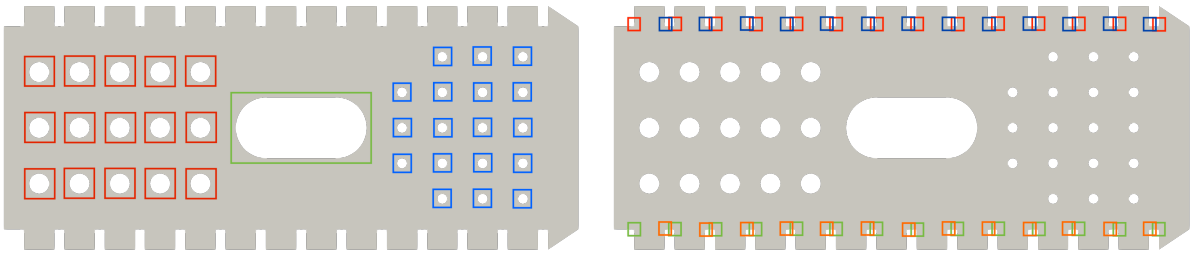
In the following, linear elasticity is investigated on the two-dimensional rib structure, consisting of 33 circular holes and the porthole-shaped hole in the center. The domain Ω used in the following experiment is of size $29\,\mathrm{cm} \times 12\,\mathrm{cm}$, and the radii of the circular inclusions are $1.0\,\mathrm{cm}$ (red patches), resp. $0.5\,\mathrm{cm}$ (blue patches). The modulus of elasticity and Poisson ratio of the material are $E = 100.0\,\mathrm{GPa}, \nu = 0.3$.

The left boundary of the rib is clamped, and the right boundary is pulled with a constant displacement of $1.0\,\mathrm{cm}$ in x_1 direction. A proper resolution of the geometry of the two-dimensional rib needs more than 5 million degrees of freedom. Instead of heavy spatial refinement, a coarse discretization on level 5 with polynomials of order 1 is chosen. The discretization is chosen this way, since all level 5 patches will at most contain (a part of) one interior hole. Note however, that the patches are too coarse to resolve details of the circular shape of the inclusions. In order to improve the approximation quality of the patchwise local approximation spaces, the following additional enrichment functions will be employed: For all patches touching the Dirichlet boundary

(left and right boundary), a particular solution is computed and used as local enrichment. Optimal basis functions are computed for the green patch one of the red patches. Following Section 5.3.1, optimal basis functions on all other red as well as all blue patches are obtained by concatenating the computed optimal basis functions on the red patch with corresponding transformations. All patches containing (a part of) an interior hole are enriched with 9 corresponding optimal basis functions. Following the stress recovery method from [Mel05, LPD18], the components of the stress field for a two-dimensional isotropic material can be expressed as a family of functions of the form

$$c_\lambda(r,\theta) = \lambda r^{\lambda-1}\tilde{c}_\lambda(r,\theta) \tag{7.1}$$

in the vicinity of discontinuities, where $\lambda \in \mathbb{R}$, (r,θ) are polar coordinates and $\tilde{c}_\lambda$ is a linear function in θ. The functions from (7.1) for corresponding textbook values of λ and the coefficients appearing in $\tilde{c}_\lambda$ are used as enrichment functions on all patches near reentrant corners. These functions are defined for L-shaped domains and may have to be rotated and / or reflected. The enrichment strategy is depicted and described again in Figure 7.2. In total, the global approximation space



(a) Subdomains of the same color have the same proportions. All patches touching the marked subdomains are enriched with correpsonding optimal basis functions.

(b) Reentrant corners are located in the center of the marked subdomains. Blue patches use the L-domain enrichments (7.1). Enrichments for the other cases are obtained from rotations and reflections of these functions.

Figure 7.2: Sketch of a rib used in the construction of airplane wings. Individual captions of (a) and (b) describe the enrichment strategy for the marked subdomains. Additionally, a particular solution for the boundary values is used as enrichment for all patches touching the left resp. right boundary of the domain.

only consisted of 375520 degrees of freedom, and the dilatational strain of the enriched solution on the deformed domain is shown in Figure 7.3. It can clearly be seen that the strain patterns around the holes differ and are not generic. The enriched Partition of Unity Method is capable of producing results with detailed fine-scale behavior while only needing a very small number of degrees of freedom. The experiment moreover validated the theoretical reusability results from Section 5.3.1 numerically.

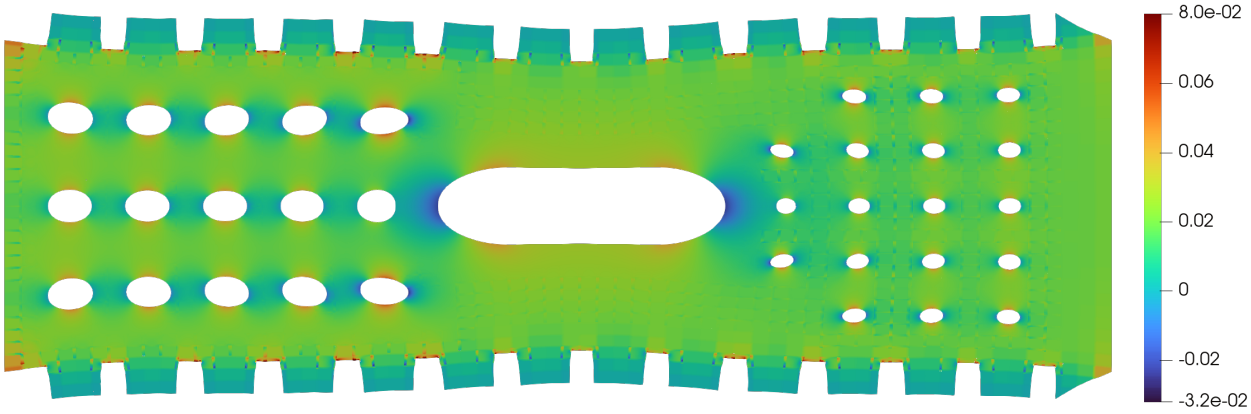


Figure 7.3: Plot of the dilatational strain of the enriched solution using 9 optimal basis functions on all patches near the inner holes on the deformed domain, one analytical enrichment on all patches near a reentrant corner, and a particular solution for the essential boundary values prescribed on the left and right boundary.

7.2 Wave propagation in heterogeneous media

In order to describe wave propagation in a linear elastic material, the *equations of elastodynamics* are solved, which links applied outer forces to vibrational responses of the material. The equations are based on Newtons second law, stating that the application of outer forces results in an acceleration. This acceleration in turn creates a velocity which moves a physical system. During this movement, resistance due to the internal structure of the material is encountered. The system of vibrational action and material resistance creates a steady-state vibration after some time, and the corresponding vibration coincides with the frequency of the applied outer force. Elastodynamics appears naturally in everyday situations and must for example be considered in civil engineering, in order to prevent structures from vibrating in their eigenfrequency due to outer influences, since this resonance effect could lead to the collapse of the entire structure. In the following, let $\Omega = [-6, 6]$ be a cube made from a linear elastic matrix material with modulus of elasticity and Poisson ratio specified by $E_M = 1.0, \nu_M = 0.3$. There are 50 spherical inclusions of radius $r = 0.5$ centered at random locations in $[-4, 4]^3$, which are made of another linear elastic material with $E_I = 100.0, \nu_I = 0.3$.

For $t_{\max} \in \mathbb{R}_+$ let $(0, t_{\max})$ be a considered time horizon. The equations of elastodynamics for linear elastic materials and a sufficiently smooth displacement field

$$u : \Omega \times (0, t_{\max}) \to \mathbb{R}^3 \tag{7.2}$$

reads

$$\rho \ddot{u}(x,t) - \operatorname{div} \sigma(u(x,t)) = f(x,t), \quad \forall (x,t) \in \Omega \times (0, t_{\max}), \tag{7.3}$$

with the material density $\rho \in \mathbb{R}$, a load function $f : \Omega \times (0, t_{\max}) \to \mathbb{R}^3$, a material stress tensor $\sigma(u)$ and a known initial displacement field $u(x, 0) : \Omega \to \mathbb{R}^3$. The functions $\dot{u}$ and $\ddot{u}$ are the first and second order derivatives in time.

In the following, the density is assumed to be $\rho = 1$ and the load function f is assumed to be zero.

The resulting hyperbolic problem reads

$$\ddot{u}(x,t) - \operatorname{div}\sigma(u(x,t)) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \forall (x,t) \in \Omega \times (0, t_{\max}). \tag{7.4}$$

The equation is discretized in space using the Partition of Unity Method, leading to the discrete variational formulation

$$M\ddot{\mathfrak{u}}(t) + K\mathfrak{u}(t) = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \forall t \in (0, t_{\max}), \tag{7.5}$$

with assembled mass matrix M and stiffness matrix K. In (7.5), $\mathfrak{u}(t) \in \mathbb{R}^{\dim \mathcal{V}^{\mathrm{PU}}(\Omega)}$ denotes the coefficients of the expansion of the solution u at timestep t in $\mathcal{V}^{\mathrm{PU}}(\Omega)$. Note that from (7.5), it directly follows that

$$\ddot{\mathfrak{u}}(t) = -M^{-1}K\mathfrak{u}(t). \tag{7.6}$$

In order to obtain a fully discretized iterative scheme, introduce the discrete time steps $t_i = i\Delta t$ and the notion $\mathfrak{u}_i := \mathfrak{u}(t_i)$, $\dot{\mathfrak{u}}_i := \dot{\mathfrak{u}}(t_i)$ and $\ddot{\mathfrak{u}}_i := \ddot{\mathfrak{u}}(t_i)$. Furthermore, a central finite difference scheme is applied. Instead of solving (7.6) in every time step, u_i on the right-hand side is replaced by u_{i-1}, allowing to calculate a new acceleration coefficient from the displacement in the previous time step. This new acceleration is then propagated into a new velocity and consequently a new displacement field. The entire iterative scheme reads

$$\begin{aligned} \ddot{\mathfrak{u}}_i &= -M^{-1}K\mathfrak{u}_{i-1} \\ \dot{\mathfrak{u}}_i &= \dot{\mathfrak{u}}_{i-1} + (\Delta t)\ddot{\mathfrak{u}}_i \\ \mathfrak{u}_i &= \mathfrak{u}_{i-1} + (\Delta t)\dot{\mathfrak{u}}_i. \end{aligned} \tag{7.7}$$

Note that in the Partition of Unity Method, mass lumping can be used for any employed local bases. Hence, M in the following refers to the lumped mass matrix, and M^{-1} is hence easy to compute ([Sch13]).

The initial displacement u_0 and acceleration $\ddot{u}_0$ are zero. The initial velocity is

$$\dot{u} = \begin{bmatrix} -\frac{1}{100} e^{\frac{-\|x-c\|_2}{0.6^2}} \\ 0 \\ 0 \end{bmatrix} \tag{7.8}$$

with $c = \begin{bmatrix} 6 & 0 & 0 \end{bmatrix}^T$ being the midpoint of the right face of Ω. The initial velocity describes an impact at c, and its magnitude is shown in Figure 7.4. For better visibility, a part of the cube has been cut out.

Using the iterative scheme (7.7) and the time step size $\Delta t = 0.0132$, values of the coefficients $\mathfrak{u}_i, \dot{\mathfrak{u}}_i, \ddot{\mathfrak{u}}_i$ are calculated for $i = 1, \ldots, 1500$.

When using the classical FEM or PUM without enrichments, the size of the elements / patches near the inclusions must be chosen small enough for their spherical shape to be resolved. For a large number of inclusions, adaptive refinement towards all of them basically refines the whole domain,

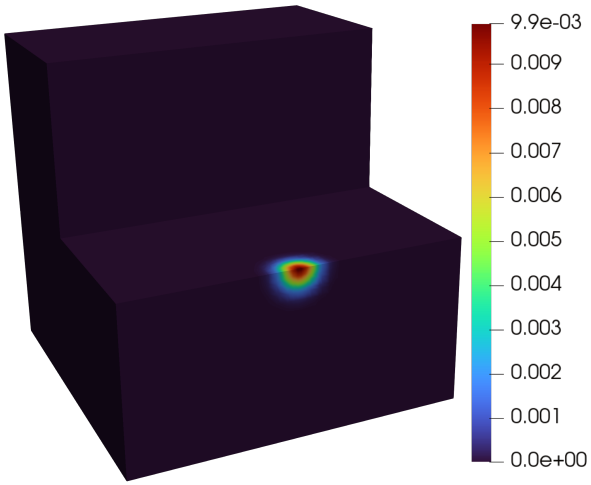
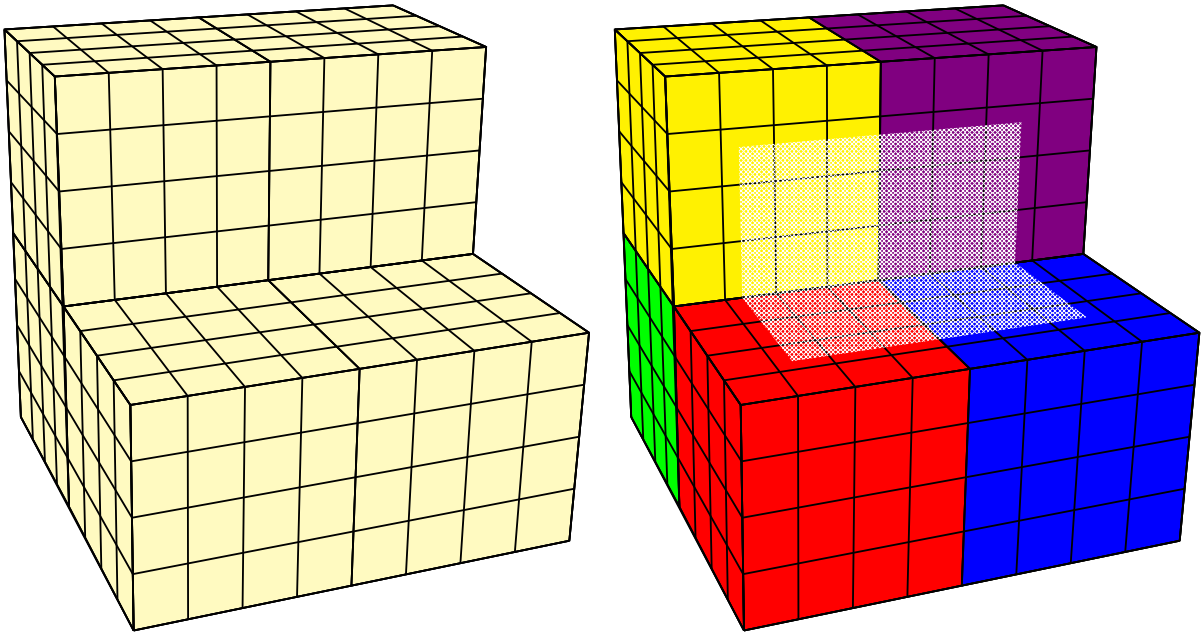


Figure 7.4: Magnitude of the initial velocity $\dot{u}$, describing an impact on the midpoint of the right face of Ω, i.e. the face with $x_1 = 6$.

leading to algebraic systems with millions of degrees of freedom that become infeasible without utilizing large parallel computers. A global discretization on level 3 with polynomials of degree 2 is chosen, that uses $8^3 = 512$ patches, and the discretization is sketched in Figure 7.5 (a). Again, a part of the domain has been cut out and the overlap region is not shown for better visibility. The sketched patches have an extent of 1.25 and the spherical inclusions, which have a radius of 0.5, can hence not be resolved by the chosen coarse discretization. In the following, the enrichment strategy used on the 512 coarse patches is described.

The benchmark problem from Section 6.4 showed, that the solution cannot be expected to only present interesting features in the subdomain containing all the inclusions, $[-4, 4]^3$, but also near the transition between Dirichlet and zero Neumann boundary conditions. Furthermore, computing and using enrichments only on patches touching boundaries of both types pushed (parts of) the error further into other regions of the domain instead of eliminating it entirely. Unfortunately, using individual enrichments for all 512 patches makes the global assembly step of the system of linear equations extremely expensive. This is due to the fact, that optimal basis functions for one coarse patch are defined on a local discretization, and all these local patches hence have to be resolved to perform a (sufficiently) exact global quadrature.

As a trade-off between complexity and accuracy, the domain Ω is cut into 8 cubes of identical size, see Figure 7.5 (b). On all of these cubes, optimal basis functions are computed. Furthermore, a particular solution for g is computed on all cubes touching the Dirichlet face. All local computations are performed using polynomials of degree 2 on a uniform discretization of level 5. Afterwards, the local approximation spaces of all 512 global patches are enriched with 25 optimal basis functions defined on the corresponding cube that the patch belongs to. All global patches forming the four cubes touching the Dirichlet face are further enriched with the particular solution for g computed on their corresponding cube. In order to further improve the approximation quality of the global space, optimal basis functions on level 6 using polynomials of degree 2 are computed on the subdomain $[-4, 4]^3$ (white shaded in Figure 7.5), and all global patches located inside of this subdomain are additionally enriched with the first 25 resulting optimal basis functions. In the end, the global approximation space $\mathcal{V}^{\mathrm{PU}}(\Omega)$ consists of 54552 degrees of freedom, thereof $512 \cdot 6 = 3072$ corresponding to local polynomials on the patches and 51480 due to enrichments.

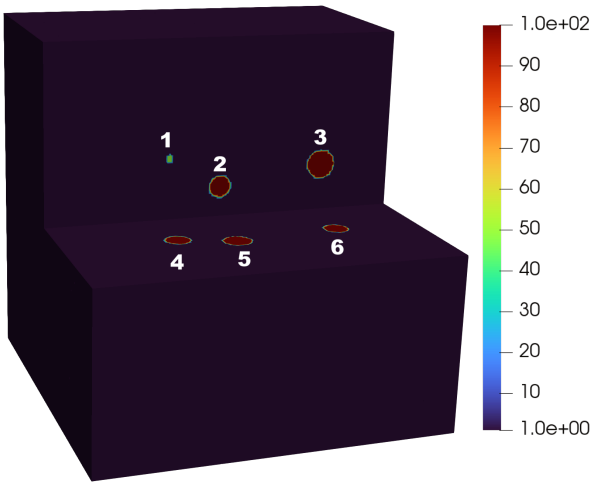


(a) Global discretization of Ω on level 3, using $8^3 = 512$ patches

(b) Splitting of Ω into 8 subdomains (colored cubes), and subdomain containing all inclusion (white shaded)

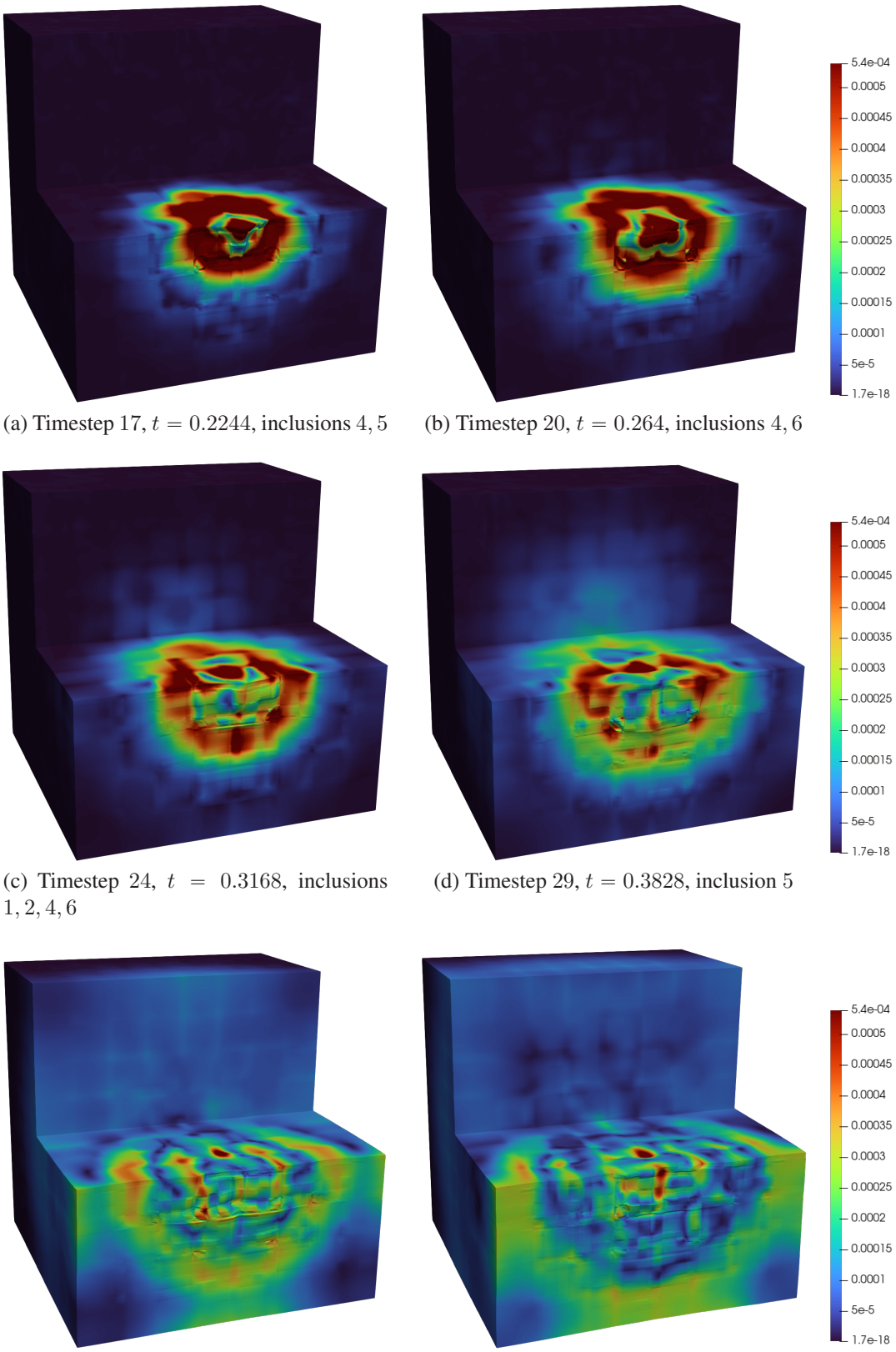
Figure 7.5: For better visibility, a part of the cubic domain $\Omega = [-6, 6]^3$ has been cut out.

Section 7.2 shows the modulus of elasticity of the multi-material under study. From the shown angle, six inclusions are visible and should be detected by the enriched method. In order to reference them, they have been numbered. The displacement field of the enriched PUM is shown for several



time steps in Figure 7.6. It is clearly visible, that the enriched Partition of Unity Method is able to detect the spherical shape of the inclusions. Since the inclusions have a much larger modulus of elasticity, the wave cannot propagate through the inclusions at the same rate as it passes trough the matrix material. This experiment shows, that the PUM can be used together with optimal basis functions to solve PDE which are infeasible unless large parallel computers are used. The initial assembly of $M^{-1}K$ is relatively expensive for the enriched PUM, since all local discretizations used in the computation of optimal basis functions and particular solutions have to be resolved

during quadrature. However, since the number of degrees of freedom used in the global approximation space is very small compared to the number of degrees of freedom that would be needed in an unenriched method, the matrix vector multiplications performed to compute the coefficients of new time steps are very cheap. For a sufficiently large number of time steps, the enriched PUM will hence also outperform any unenriched method that uses more degrees of freedom in terms of runtime.



(a) Timestep 17, $t = 0.2244$, inclusions $4, 5$

(b) Timestep 20, $t = 0.264$, inclusions $4, 6$

(c) Timestep 24, $t = 0.3168$, inclusions $1, 2, 4, 6$

(d) Timestep 29, $t = 0.3828$, inclusion 5

(e) Timestep 42, $t = 0.5544$, inclusion 4

(f) Timestep 51, $t = 0.6732$, inclusions $3, 4, 5$

Figure 7.6: The displacement field at several time steps, together with the indices of the detectable inclusions according to the labeling in Section 7.2.

8 Conclusive remarks

This thesis investigated a constructive way of computing optimal local approximation spaces. The framework was originally presented in [BL11] for the solution of second-order elliptic partial differential equations by the means of Finite Element Methods. During the writing of this thesis, the framework was extended to the case of even-order elliptic PDE and modified for the use in the Partition of Unity Method.

The PUM, is a broad generalization of the FEM and allows the use of independent local approximation spaces, which are then linked into a global approximation space with a Partition of Unity. The global approximation error of the PUM is a direct consequence of the local approximation errors, and a best-possible choice of local approximation spaces is hence favorable. The presented framework allows to pre-compute local approximation spaces, which are optimal in the sense that they minimize the Kolmogorov n-width. They are operator-dependent and independent of the explicit values of load and boundary data appearing in the PDE. Using the obtained space of optimal local basis functions instead of heavy spatial refinement in areas of interest has the potential to substantially reduce the number of degrees of freedom needed for an adequate discretization of a given PDE.

Since the construction of the optimal approximation spaces is expensive, algebraic conditions for their geometric reusability were derived. In the case of second order PDE in divergence form, this led to explicit, easy-to-check conditions that need to be satisfied by the coefficients of the PDE and any feasible geometric transformation. Whenever these conditions are satisfied, a set of optimal basis functions can be transformed to another patch and still be optimal. This has the potential to significantly reduce the computational effort for problems posed on domains with various regions of interest and moreover increase the possibility to reuse results from previous simulations.

Various numerical experiments were conducted for this thesis. The first set of experiments consisted of four benchmark problems, that underwent a full numerical investigation. The entire process of constructing optimal local approximation spaces and using them in enriched global simulations was investigated and quantified. The goal of these full investigations was to understand the impact of the choice of boundary data used in the sampling problem. It turned out that both boundary hats and quadratic B-Splines can be used as boundary data in the sampling problem to obtain a small number of relevant eigenpairs in the generalized eigenvalue problem. The two approaches lead to local approximation spaces with similar performance in global enriched sim-

ulations. Even tough both approaches perform similarly, the amount of B-Spline boundary data functions is smaller, favoring this approach.

The second type of conducted experiments were concerned with problems that are very hard to solve using traditional numerical methods. The enriched PUM was capable of producing results for these proof-of-concept problems with details on a very fine scale while only using a very small number of degrees of freedom, validating the presented approach.

All simulations are based on the `PUMA` software toolchain developed by Fraunhofer SCAI ([SCA]). In order to compute and use optimal local basis functions, a `Python 3` module, `optbasefun`, was developed in the scope of this thesis. This module allows to control the `Python` front end of `PUMA` for the computation of optimal local basis functions. Many of the classes defined in `optbasefun` handle instances of `PUMA` classes together with additional meta information needed for the computation or use of optimal basis functions, which leads to a modularization of the code into independent and interchangeable blocks. Additionally, it provides easy-to-use data handlers, for example to set up the different types of boundary condition functions which are used in the sampling of harmonic functions, or to simplify the setup and solution of the generalized eigenvalue problem.

Outlook

In this thesis, the framework for the computation of optimal local basis functions was investigated in terms of its feasibility. When shifting the focus towards performance, there are possible technical improvements to be made, for example regarding the structure of the output. Performing numerical quadrature is typically the most expensive part of any global method using enrichments defined on a very fine discretization. Instead of storing the coefficient vectors of the optimal basis functions and integrating their products with polynomial basis functions in every enriched simulation, the assembled entries of the mass / stiffness matrix may be stored and reused directly. This can lead to a significant speedup of the enriched method, but requires the definition of binding rules for the geometric alignment between the local domain and global patches.

Additionally, the results regarding reusability may be extended. This thesis presented algebraic conditions for the equality of optimal basis functions, and equality of several functions implies equality of the spaces spanned by these functions. However, a different set of spanning functions may still lead to the same (or a very similar) spanned space. Since the "approximation quality" of a set of optimal local basis functions refers to their spanned space instead of their individual contributions, similarity of optimal local approximation spaces should be investigated as well. The main reasoning is, that slight changes of an elliptic PDE lead to slight changes of the harmonic functions, which in turn lead to slight changes of the discrete matrices in the generalized eigenvalue problem and consequently the discrete eigenfunctions. The optimal local approximation spaces for two very similar PDE are hence expected to be sufficiently close to each other, and this can possibly be quantified by the computation of an angle between these spaces ([GNB05, Hit13]). Ultimately, this may lead to the development of a decision mechanism for the a priori selection of a set of pre-computed optimal basis functions according to the coefficients of a problem at hand. Using machine learning algorithms, additional sets of optimal basis functions may even be generated without the need to explicitly solve PDE.

While the problems considered in the experimental part of this thesis were of second order, the

framework has been shown to be feasible also for higher order PDE. Therefore, future applications should also involve higher order PDE, such as the biharmonic equation or the Cahn Hilliard equation. It is worth noticing that it is technically very involved to construct and use higher order Finite Elements, whereas the (enriched) PUM can easily be applied to such problems.

While this thesis investigated one way of constructing operator-dependent basis functions in detail, a multitude of methods for the construction of improved shape functions has been developed ([EHMP19, SW17, CELL19]). Moreover, there are other well established approaches to reduce the complexity of global approximation spaces. This ranges from the use of shell elements in the modeling of thin structures to the Global Local approach typically used in fracture mechanics, as presented by Birner in his dissertation ([Bir23]). It will be interesting for future works to investigate the combination of such methods with the presented framework for the computation of optimal basis functions.

Appendix A: Korn inequality

In the following, the works of [LD72] and [Tar82] regarding the Korn inequality (Theorem 3.1) are recapitulated. The proof of the Korn inequality on a sufficiently regular domain $\Omega \subset \mathbb{R}^{2,3}$ relies on an alternative representation of the space in the form of $\mathrm{L}^2(\Omega)$. Introduce

$$\mathcal{X}(\Omega) \coloneqq \{v \in \mathrm{H}^{-1}(\Omega) |\, \partial_{x_j} v \in \mathrm{H}^{-1}(\Omega)\, j = 1, \ldots, d\}. \tag{A.1}$$

The equality will be proven first for the full space $\mathbb{R}^d$, then for a half-space $\mathbb{R}^d_+$ and afterwards for a regular, bounded and open set $\Omega \subset \mathbb{R}^d$, corresponding to Lemma 1 to 5 in [Tar82, Functional spaces related to the Navier-Stokes equation, pp. 26 – 29].

Lemma A.1. *For $\Omega = \mathbb{R}^d$, it holds that $\mathcal{X}(\mathbb{R}^d) = \mathrm{L}^2(\mathbb{R}^d)$.*

•

Proof. Plancharel's Theorem states, that the Fourier transform $\mathcal{F} : \mathrm{L}^2(\mathbb{R}^d) \to \mathrm{L}^2(\mathbb{R}^d)$ is an isometry. It is given by

$$\begin{aligned} \mathcal{F}(v)(y) = \quad \hat{v}(y) &= \int_{\mathbb{R}^d} e^{-2ix\cdot y} v(x)\, dx \\ \mathcal{F}^{-1}(v)(x) = \quad v(x) &= \int_{\mathbb{R}^d} e^{2ix\cdot y} \hat{v}(y)\, dy. \end{aligned} \tag{A.2}$$

The Fourier transform allows to identify

$$v \in \mathrm{H}^1(\mathbb{R}^d) \quad \Longleftrightarrow \quad (1 + |\cdot|^2)^{\frac{1}{2}} \hat{v}(\cdot) \in \mathrm{L}^2(\mathbb{R}^d). \tag{A.3}$$

It also allows an identification of the partial derivatives via

$$\partial_{x_j} \in \mathrm{H}^1(\mathbb{R}^d) \quad \Longleftrightarrow \quad (1 + |\cdot|^2)^{\frac{1}{2}} \cdot_i \hat{v}(\cdot) \in \mathrm{L}^2(\mathbb{R}^d). \tag{A.4}$$

Since the Fourier transform is an isometry, i.e.

$$\|v\|_{\mathrm{L}^2(\mathbb{R}^d)} = \|\mathcal{F}(v)\|_{\mathrm{L}^2(\mathbb{R}^d)} = \|\hat{v}\|_{\mathrm{L}^2(\mathbb{R}^d)}, \tag{A.5}$$

the conditions (A.3) and (A.4) mount to

$$v \in \mathrm{H}^1(\mathbb{R}^d) \iff \frac{\hat{v}(\cdot)}{(1+|\cdot|^2)^{\frac{1}{2}}} \in \mathrm{L}^2(\mathbb{R}^d) \tag{A.6}$$

and

$$\partial_{x_j} v \in \mathrm{H}^1(\mathbb{R}^d) \iff \frac{|\hat{v}(\cdot)\,\cdot_i|}{(1+|\cdot|^2)^{\frac{1}{2}}} \in \mathrm{L}^2(\mathbb{R}^d) \tag{A.7}$$

for $j = 1, \dots, d$. The norm of the sum of the right-hand side from (A.6) and all right-hand sides from (A.7) reads

$$\begin{aligned}
\left\| \frac{\hat{u}(\cdot)}{(1+|\cdot|^2)^{\frac{1}{2}}} + \sum_{j=1}^{d} \frac{|\hat{u}(\cdot)\,\cdot_i|}{(1+|\cdot|^2)^{\frac{1}{2}}} \right\|_{\mathrm{L}^2(\mathbb{R}^d)} &\leqslant \left\| \frac{\hat{u}(\cdot)}{(1+|\cdot|^2)^{\frac{1}{2}}} \right\|_{\mathrm{L}^2(\mathbb{R}^d)} + \sum_{j=1}^{d} \left\| \frac{|\hat{u}(\cdot)\,\cdot_i|}{(1+|\cdot|^2)^{\frac{1}{2}}} \right\|_{\mathrm{L}^2(\mathbb{R}^d)} \\
&= \int_{\mathbb{R}^d} \frac{\hat{u}^2(y)}{1+|y|^2}\,dy + \sum_{j=1}^{d} \int_{\mathbb{R}^d} \frac{|\hat{u}(y)y_i|^2}{1+|y|^2}\,dy \\
&= \int_{\mathbb{R}^d} \frac{\hat{u}^2(y)}{1+|y|^2}\,dy + \sum_{j=1}^{d} \int_{\mathbb{R}^d} \frac{\hat{u}(y)^2 y_i^2}{1+|y|^2}\,dy \\
&= \int_{\mathbb{R}^d} \frac{\hat{u}^2(y)}{1+|y|^2}\,dy + \int_{\mathbb{R}^d} \frac{\hat{u}(y)^2|y|^2}{1+|y|^2}\,dy \\
&= \int_{\mathbb{R}^d} \frac{(1+|y|^2)\hat{u}(y)^2}{1+|y|^2}\,dy \\
&= \left\| \frac{(1+|\cdot|)\hat{u}(\cdot)}{(1+|\cdot|^2)^{\frac{1}{2}}} \right\|_{\mathrm{L}^2(\mathbb{R}^d)}
\end{aligned}$$

and the term on the right-hand side is finite, since all occurring integrals are finite. Therefore,

$$(A.6), (A.7) \quad \iff \quad \frac{(1+|\cdot|)\hat{u}(\cdot)}{(1+|\cdot|^2)^{\frac{1}{2}}} \in \mathrm{L}^2(\mathbb{R}^d). \tag{A.8}$$

This is true, whenever $\hat{u} \in \mathrm{L}^2(\mathbb{R}^d)$, which in turn holds whenever $u \in \mathrm{L}^2(\mathbb{R}^d)$. This proves the claim. □

In order to prove the statement for a half-space, the following statement is needed.

Lemma A.2. *$\overline{\mathcal{C}_0^\infty(\mathbb{R}^d_+)}$ is dense in $\mathcal{X}(\mathbb{R}^d_+)$.*

•

Proof. Let $u \in \mathcal{X}$ and let $\{T_{u_h}\}_h \subset \mathcal{X}(\mathbb{R}^d_+)$ defined by the relation

$$T_{u_h} := \langle u_h, v \rangle_{\mathrm{H}^1(\Omega)} = \langle u, u_{-h} \rangle_{\mathrm{H}^1(\Omega)}, \quad \forall v \in \mathrm{H}^1_0(\mathbb{R}^d_+). \tag{A.9}$$

It holds that $u_h \overset{h\to 0}{\to} u$ in $\mathcal{X}(\mathbb{R}^d_+)$. Let now $\varphi \in \mathcal{C}^\infty(\mathbb{R}^d)$ defined by

$$\varphi(x) := \begin{cases} C \exp\left(\frac{1}{|x|^2-1|}\right), & |x| < 1 \\ 0, & |x| \geqslant 1 \end{cases} \tag{A.10}$$

with $C > 0$ selected such that $\int_{\mathbb{R}^d} \varphi(x)\,dx = 1$. The sequence of standard mollifiers is defined by the functions

$$\varphi_\varepsilon(x) \coloneqq \frac{1}{\varepsilon^d}\varphi\left(\frac{x}{\varepsilon}\right) \in \mathcal{C}_0^\infty(\mathbb{R}^d). \tag{A.11}$$

For any h, the mollification of u_h satisfies $\rho_\epsilon^h \coloneqq (u_h \star \varphi_\varepsilon)_{|\mathbb{R}^d_+} \overset{\varepsilon\to 0}{\to} u_h$, and since $\overline{\mathcal{C}_0^\infty(\mathbb{R}^d_+)}$ is closed, it holds for the limit that $u_h \in \overline{\mathcal{C}_0^\infty(\mathbb{R}^d_+)}$. Hence, any element from $\mathcal{X}(\mathbb{R}^d_+)$ can be approximated arbitrarily well using a sequence from $\overline{\mathcal{C}_0^\infty(\mathbb{R}^d_+)}$, meaning that $\overline{\mathcal{C}_0^\infty(\mathbb{R}^d_+)}$ is dense in $\mathcal{X}(\mathbb{R}^d_+)$. □

Lemma A.3. *There exists a continuous extension* $\Theta : \mathcal{X}(\mathbb{R}^d_+) \to \mathcal{X}(\mathbb{R}^d)$.

•

Proof. Define the operator $Q : \mathrm{H}^1(\mathbb{R}^d) \to \mathrm{H}^1_0(\mathbb{R}^d_+)$ by

$$Q(u)(x_1,\dots,x_d) \coloneqq \begin{cases} 0, & x_d < 0 \\ u(x_1,\dots,x_d) + \sum_{j=1}^2 a_j u(x_1,\dots,x_{d-1},-jx_d), & x_d \geqslant 0 \end{cases}$$

where $a_1, a_2 \in \mathbb{R}$ are chosen such that $1 + a_1 + a_2 = 0$. By inserting $x_d = 0$ into the definition,

$$Q(u)(x_1,\dots,x_{d-1},0) = 0, \tag{A.12}$$

one sees that the image of Q are indeed functions from H^1_0. Define a second operator

$$R : \mathrm{H}^1(\mathbb{R}^d) \to \mathrm{H}^1_0(\mathbb{R}^d_+) \tag{A.13}$$

by

$$R(u)(x_1,\dots,x_d) = \begin{cases} 0, & x_d < 0 \\ u(x_1,\dots,x_d) + \sum_{j=1}^2 \frac{a_j}{-j} u(x_1,\dots,x_{d-1},-jx_d), & x_d < 0 \end{cases}$$

and since the image of R should be functions from H^1_0, $1 + \frac{a_1}{-1} + \frac{a_2}{-2} = 0$ is required, ensuring $R(u)(x_1,\dots,x_{d-1},0) = 0$. Using the linearity of Q yields

$$Q\left(\frac{\partial u}{\partial x_i}\right) = \begin{cases} \frac{\partial}{\partial x_i} Q(u), & i \neq d \\ \frac{\partial}{\partial x_d} R(u), & i = d. \end{cases} \tag{A.14}$$

Since $Q(u) = 0$ on $x_d = 0$, the operator Q can be extended to $\mathrm{H}^1(\mathbb{R}^d)$ by continuity, so

$$Q_{\mathrm{ext}} : \mathrm{H}^1(\mathbb{R}^d) \to \mathrm{H}^1(\mathbb{R}^d). \tag{A.15}$$

Let $\Theta = Q^T$ the formal adjoint of Q,

$$\Theta : \mathrm{H}^1_0(\mathbb{R}^d_+) \to \mathrm{H}^1(\mathbb{R}^d), \tag{A.16}$$

that is

$$\Theta : \mathrm{H}^{-1}(\mathbb{R}^d_+) \to \mathrm{H}^{-1}(\mathbb{R}^d). \tag{A.17}$$

For any $v \in \mathrm{H}^1(\mathbb{R}^d)$ and any $i \neq d$ it holds that

$$\begin{aligned}\langle \frac{\partial}{\partial x_i}\Theta(u), v\rangle = &\quad -\langle \Theta(u), \frac{\partial v}{\partial x_i}\rangle = -\langle u, Q\left(\frac{\partial v}{\partial x_i}\right)\rangle \\ = &\quad -\langle u, \frac{\partial}{\partial x_i}Q(v)\rangle = \langle \frac{\partial u}{\partial x_i}, Q(v)\rangle \\ = &\quad \langle \Theta\left(\frac{\partial u}{\partial x_i}\right), v\rangle,\end{aligned} \tag{A.18}$$

that is

$$\frac{\partial}{\partial x_i}\Theta = \Theta\frac{\partial}{\partial x_i}, \quad i \neq d. \tag{A.19}$$

For $i = d$ it similarly holds that

$$\frac{\partial}{\partial x_d}\Theta = R^T\frac{\partial}{\partial x_d}, \tag{A.20}$$

where $R^T : \mathrm{H}^{-1}(\mathbb{R}^d_+) \to \mathrm{H}^{-1}(\mathbb{R}^d)$ is the form adjoint of R. Let now $\frac{\partial u}{\partial x_d} \in \mathrm{H}^{-1}(\mathbb{R}^d_+)$. Then

$$\frac{\partial}{\partial x_d}\Theta(u) \in \mathrm{H}^{-1}(\mathbb{R}^d). \tag{A.21}$$

The last thing to prove is that the operator Θ is an extension. For this, use a restriction operator

$$\Pi : \mathrm{H}^{-1}(\mathbb{R}^d) \to \mathrm{H}^{-1}(\mathbb{R}^d_+), \quad \Pi(u) \coloneqq u_{|\mathbb{R}^d_+}, \quad \forall u \in \mathrm{H}^{-1}(\mathbb{R}^d) \tag{A.22}$$

and an extension by zero

$$\Psi : \mathrm{H}^1_0(\mathbb{R}^d_+) \to \mathrm{H}^1(\mathbb{R}^d), \quad \Psi(u) \coloneqq \begin{cases} u(x), & x \in \mathbb{R}^d_+ \\ 0, & x \notin \mathbb{R}^d_+, \end{cases} \tag{A.23}$$

which is the formal transpose of the restriction. The following relations are obtained

$$\begin{array}{ccccc} \mathrm{H}^{-1}(\mathbb{R}^d_+) & \overset{P}{\to} & \mathrm{H}^{-1}(\mathbb{R}^d) & \overset{\Pi}{\to} & \mathrm{H}^{-1}(\mathbb{R}^d_+) \\ \mathrm{H}^1_0(\mathbb{R}^d_+) & \overset{Q}{\leftarrow} & \mathrm{H}^1(\mathbb{R}^d) & \overset{\Psi}{\leftarrow} & \mathrm{H}^1_0(\mathbb{R}^d_+) \end{array} \tag{A.24}$$

Note that $Q\Psi = \mathrm{id}$ holds, which is equivalent to $\Pi\Theta = \mathrm{id}$, showing that Θ is an extension operator. □

Lemma A.4. *It holds that* $\mathcal{X}(\mathbb{R}^d_+) = \mathrm{L}^2(\mathbb{R}^d_+)$.

•

Proof. Using the continuous extension Θ from lemma A.3, $u \in \mathcal{X}(\mathbb{R}^d_+)$ implies

$$\Theta(u) \in \mathcal{X}(\mathbb{R}^d) \tag{A.25}$$

and hence $\Theta(u) \in \mathrm{L}^2(\mathbb{R}^d)$ by lemma A.1. □

Lemma A.5. *Let* $\Omega \subset \mathbb{R}^d$ *be regular, bounded and open. Then* $\mathcal{X}(\Omega) = \mathrm{L}^2(\Omega)$.

•

Proof. Let $\{\Omega_i\}_i$ be an open cover of Ω. Consider a smooth partition of unity subordinate to the cover, $\{\varphi_i\}_i$, that is functions

$$\varphi_i \in \mathcal{C}^\infty(\mathbb{R}^d; [0,1]), \quad \forall i \tag{A.26}$$

with

$$\sum_i \varphi_i(x) = 1, \quad \forall x \in \Omega\,. \tag{A.27}$$

One can write any $u \in \mathcal{X}(\Omega)$ as

$$u = \sum_i \varphi_i u. \tag{A.28}$$

If i is such that $\overline{\Omega_i}$ is entirely contained in Ω, then $\varphi_i u \in \mathcal{X}(\mathbb{R}^d)$, so $\varphi_i u \in \mathrm{L}^2(\mathbb{R}^d)$ using lemma A.1, and since $\sup \varphi_i u = \Omega_i$ also $\varphi_i u \in \mathrm{L}^2(\Omega)$ using an extension by 0 outside of the support. If j is such that φ_j intersects the boundary of Ω, take a local representation of the boundary, that is a function ρ_j that has two bounded derivatives and an inverse ρ_j^{-1} which also has two bounded derivatives. The function ρ_j maps the part of the boundary $\Omega_j \cap \partial\,\Omega$ onto the axis $x_d = 0$ (see Figure A.1 for a sketch).

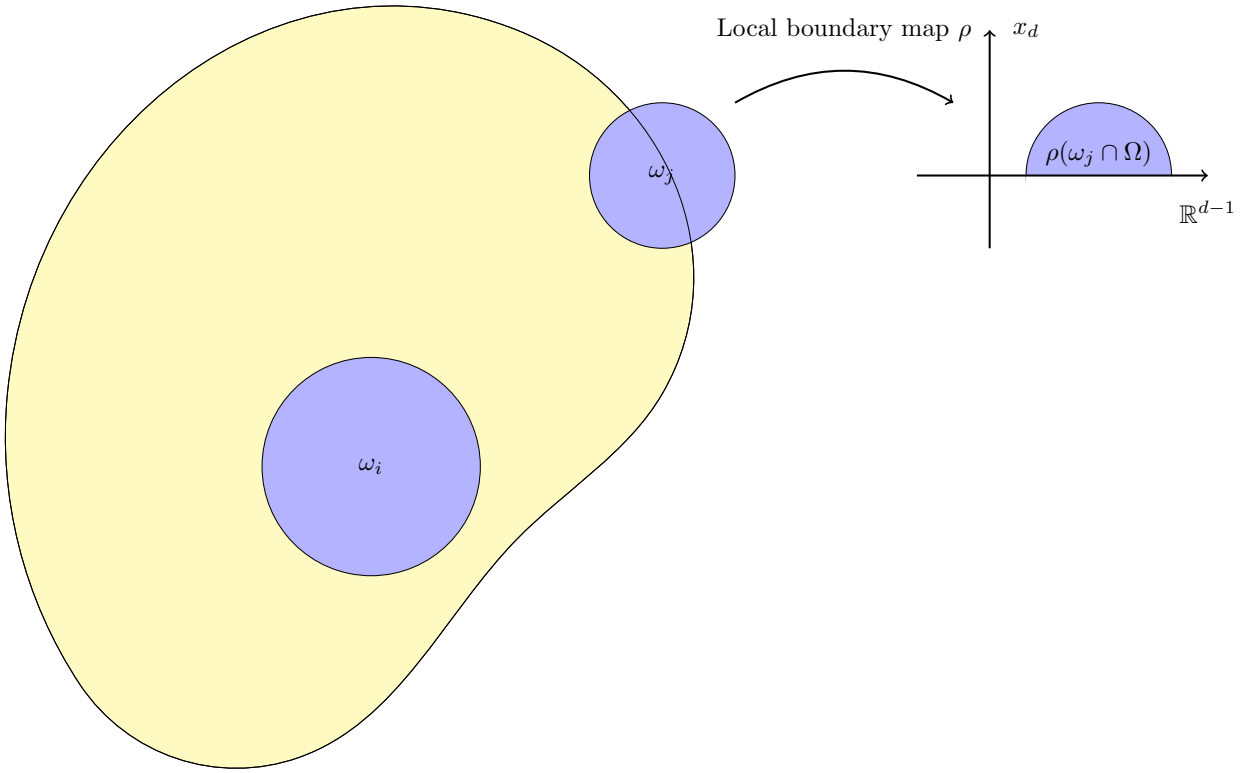


Figure A.1: Sketch of the action of the local boundary map ρ_j acting on patch Ω_j.

Therefore

$$\varphi_i u \in \mathcal{X}(\Omega) \tag{A.29}$$

implies that

$$(\varphi_i u) \circ \rho^{-1} \in \mathcal{X}(\mathbb{R}^d_+), \tag{A.30}$$

and

$$(\varphi_i u) \circ \rho^{-1} \in \mathrm{L}^2(\mathbb{R}^d_+). \tag{A.31}$$

by lemma Lemma A.4. Since L^2 measurability is preserved by ρ_i, it can be used to map back to Ω_i ending up with

$$\varphi_i u \in \mathrm{L}^2(\Omega). \tag{A.32}$$

By iterating over all elements of the cover, one obtains

$$u = \sum_i \underbrace{\varphi_i u}_{\in \mathrm{L}^2(\Omega)} \in \mathrm{L}^2(\Omega). \tag{A.33}$$

□

Using lemma A.5, the Korn inequality (Theorem 3.1) can finally be proven. This inequality shows one part of the equivalence of the norm $\|\cdot\|_{\mathrm{H}^1(\Omega)}$ to the strain energy norm

$$\begin{aligned} \|v\|_{\mathcal{E}(\Omega)} &:= \left[\int_\Omega \varepsilon(v)(x) : \varepsilon(v)(x)\, dx + \int_\Omega v(x) \cdot v(x)\, dx\right]^{\frac{1}{2}} \\ &= \left[\|\varepsilon(v)\|^2_{[\mathrm{L}^2(\Omega)]^{d\times d}} + \|v\|^2_{[\mathrm{L}^2(\Omega)]^d}\right]^{\frac{1}{2}}, \end{aligned}$$

Proof of Theorem 3.1. Let $E(\Omega)$ be the space of $v \in [\mathrm{L}^2(\Omega)]^d$ such that

$$\varepsilon_{ij}(v) \in \mathrm{L}^2(\Omega), \quad \forall i, j = 1, \ldots, d \tag{A.34}$$

It is obvious, that the strain energy norm is indeed a norm on $E(\Omega)$, and it is easily verified that $E(\Omega)$ is a Hilbert space. For all components $i = 1, \ldots, d$ it holds that

$$\frac{\partial\, \mathrm{L}^2\, v_i}{\partial x_j \partial x_k} = \frac{\partial \varepsilon_{ik}(v)}{\partial x_j} + \frac{\partial \varepsilon_{ij}(v)}{\partial x_k} - \frac{\partial \varepsilon_{jk}(v)}{\partial x_i}. \tag{A.35}$$

Since $v \in E(\Omega)$, it holds that $\varepsilon_{ij}(v) \in \mathrm{L}^2(\Omega)$, so

$$\frac{\partial \varepsilon_{ij}(v)}{\partial x_k} \in \mathrm{H}^{-1}(\Omega). \tag{A.36}$$

Therefore, eq. (A.35) yields

$$\frac{\partial\, \mathrm{L}^2\, v_i}{\partial x_j \partial x_k} \in \mathrm{H}^{-1}(\Omega), \quad \forall i, j, k = 1, \ldots, d. \tag{A.37}$$

It was observed that $\frac{\partial v_i}{\partial x_k}$ satisfies all assumptions from lemma A.5, and therefore

$$\frac{\partial v_i}{\partial x_k} \in \mathrm{L}^2(\Omega), \quad \forall i, k = 1, \ldots, d. \tag{A.38}$$

Therefore $v \in [\mathrm{H}^1(\Omega)]^d$ holds, and since the other inclusion is trivial $E(\Omega) = [\mathrm{H}^1(\Omega)]^d$ follows. Note that the inclusion

$$i : [\mathrm{H}^1(\Omega)]^d \hookrightarrow E(\Omega) \tag{A.39}$$

is continuous, and as seen before, surjective. Therefore (cf. closed graph theorem) it is an isomorphism satisfying

$$\|v\|_{E(\Omega)} \geqslant C_{\mathrm{korn}}(\Omega)\|v\|_{\mathcal{V}^d(\Omega)}, \quad \forall v \in [\mathrm{H}^1(\Omega)]^d, \tag{A.40}$$

where $C_{\mathrm{korn}}(\Omega)$ is the continuity constant of the inclusion from the closed graph theorem. □

Bibliography

[BC09] O.A. Bauchau and J.I. Craig. *Structural Analysis: With Applications to Aerospace Structures*. Solid Mechanics and Its Applications. Springer Netherlands, 2009.

[BHL14] Ivo Babuška, Xu Huang, and Robert Lipton. Machine computation using the exponentially convergent multiscale spectral generalized finite element method. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 48 no. 2:493–515, 2014.

[Bir23] Matthias Birner. Global-local techniques for fracture. unpublished, 2023.

[BL11] Ivo Babuška and Robert Lipton. Optimal local approximation spaces for generalized finite element methods with application to multiscale problems. *Multiscale Modeling & Simulation*, 9(1):373–406, 2011.

[Bra07] Dietrich Braess. *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, 3 edition, 2007.

[Bro64] A. L. Brown. Best n-Dimensional Approximation to Sets of Functions. *Proceedings of the London Mathematical Society*, s3-14(4):577–594, 10 1964.

[CELL19] Eric Chung, Yalchin Efendiev, Yanbo Li, and Qin Li. Generalized multiscale finite element method for the steady state linear boltzmann equation, 2019.

[Cou43] Richard Courant. Variational methods for the solution of problems of equilibrium and vibrations. *Bulletin of the American Mathematical Society*, 49:1–23, 1943.

[EHMP19] Christian Engwer, Patrick Henning, Axel Målqvist, and Daniel Peterseim. Efficient implementation of the localized orthogonal decomposition method, 2019.

[Gar62] A. L. Garkavi. On the optimal net and best cross-section of a set in a normed space. *Izv. Akad. Nauk SSSR Ser. Mat.; Translated: Amer. Math. Soc. Transl. (2) 39 (1964), 111-132. MR 25 #429.*, 26:87–106, 1962.

[GF18] Phillip L. Gould and Yuan Feng. *Introduction to linear elasticity*. Springer, Cham, Switzerland, fourth edition. edition, 2018.

[GNB05] Hendra Gunawan, Oki Neswan, and Wono Setya Budhi. A formula for angles between subspaces of inner product spaces. *Beiträge zur Algebra und Geometrie*, 46, 01 2005.

[Gur73] Morton E. Gurtin. The linear theory of elasticity. In C. Truesdell, editor, *Linear Theories of Elasticity and Thermoelasticity: Linear and Nonlinear Theories of Rods, Plates, and Shells*, pages 1–295, Berlin, Heidelberg, 1973. Springer Berlin Heidelberg.

[Hac17] Wolfgang Hackbusch. *Theorie und Numerik elliptischer Differentialgleichungen*. Springer, Wiesbaden, 4. auflage edition, 2017.

[Hit13] Eckhard Hitzer. Angles between subspaces, 2013.

[HW65] Francis H. Harlow and J. Eddie Welch. Numerical Calculation of Time-Dependent Viscous Incompressible Flow of Fluid with Free Surface. *The Physics of Fluids*, 8(12):2182–2189, 12 1965.

[Kol36] A. Kolmogoroff. Über die beste annaherung von funktionen einer gegebenen funktionenklasse. *Annals of Mathematics*, 37(1):107–110, 1936.

[LD72] Jacques-Louis Lions and Georges Duvaut. *Les inéquations en mécanique et en physique*, volume 21 of *Travaux Recherches Mathématiques*. Dunod, 1972.

[LMM68] Jacques-Louis Lions, Enrico Magenes, and Enrico Magenes. *Problemes aux limites non homogenes et applications*, volume 1 (2). Dunod Paris, 1968.

[LPD18] Rafael Lins, Sergio Persival Proenca, and Carlos Armando Duarte. Efficient and accurate stress recovery procedure and a-posteriori error estimator for the stable generalized/extended finite element method. *International Journal for Numerical Methods in Engineering*, 119, 10 2018.

[LSS22] Robert Lipton, Paul Sinz, and Michael Stuebner. Angles between subspaces and nearly optimal approximation in gfem. *Computer Methods in Applied Mechanics and Engineering*, 402:115628, 2022. A Special Issue in Honor of the Lifetime Achievements of J. Tinsley Oden.

[MB96] Jens Markus Melenk and Ivo Babuška. The partition of unity finite element method: Basic theory and applications. Report, Eidgenössische Technische Hochschule (ETH) Zürich, Zurich, 1996.

[MB97] Jens Markus Melenk and Ivo Babuška. The partition of unity method. *International Journal for Numerical Methods in Engineering*, 40(4):727–758, 1997.

[Mel05] Jens Markus Melenk. *On Approximation in Meshless Methods*, pages 65–141. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[PBP02] Hartmut Prautzsch, Wolfgang Boehm, and Marco Paluszny. *Bézier and B-Spline Techniques*. Springer Berlin, Heidelberg, 01 2002.

[Pin85] Allan Pinkus. n-widths in approximation theory. In *Ergebnisse der Mathematik und ihrer Grenzgebiete*, 1985.

[SCA] Fraunhofer SCAI. Puma – rapid enriched simulation application development. www.scai.fraunhofer.de/de/geschaeftsfelder/meshfree-multiscale-methods/produkte/puma.html. Accessed: 13.12.2023.

[Sch03] Marc Alexander Schweitzer. *Partition of Unity Method*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[Sch11] Marc Alexander Schweitzer. Stable enrichment and local preconditioning in the particle-partition of unity method. *Numerische Mathematik*, 118:137–170, 05 2011.

[Sch13] Marc Alexander Schweitzer. Variational mass lumping in the partition of unity method. *SIAM Journal on Scientific Computing*, 35(2):A1073–A1097, 2013.

[Sin13] Ivan Singer. *Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2013.

[SW17] Marc Alexander Schweitzer and Sa Wu. Evaluation of local multiscale approximation spaces for partition of unity methods. In M. Griebel and M. A. Schweitzer, editors, *Meshfree Methods for Partial Differential Equations VIII*, volume 115 of *Lecture Notes in Science and Engineering*, pages 163–194. Springer International Publishing, 2017.

[Tar82] Luc Tartar. *Topics in nonlinear analysis*. Publications Mathematiques D'orsay, 1982.

[TB67] V. M. Tichomirov and S. B. Babadjanov. On the width ofa functional class in the space $l_p(p \geqslant 1)$. *Izv. Akad. Nauk SSSR Ser. Mat.*, pages 24–30, 1967.

[TCMT56] M.J. Turner, R.W. Clough, H.C. Martin, and L.J. Topp. Stiffness and deflection analysis of complex structures. *Journal of Aeronautical Sciences*, 23(9):805–823, 1956.

[ZC67] O.C. Zienkiewicz and Y.K. Cheung. *The Finite Element Method in Structural and Continuum Mechanics: Numerical Solution of Problems in Structural and Continuum Mechanics*. Number v. 1 in European civil engineering series. McGraw-Hill, 1967.