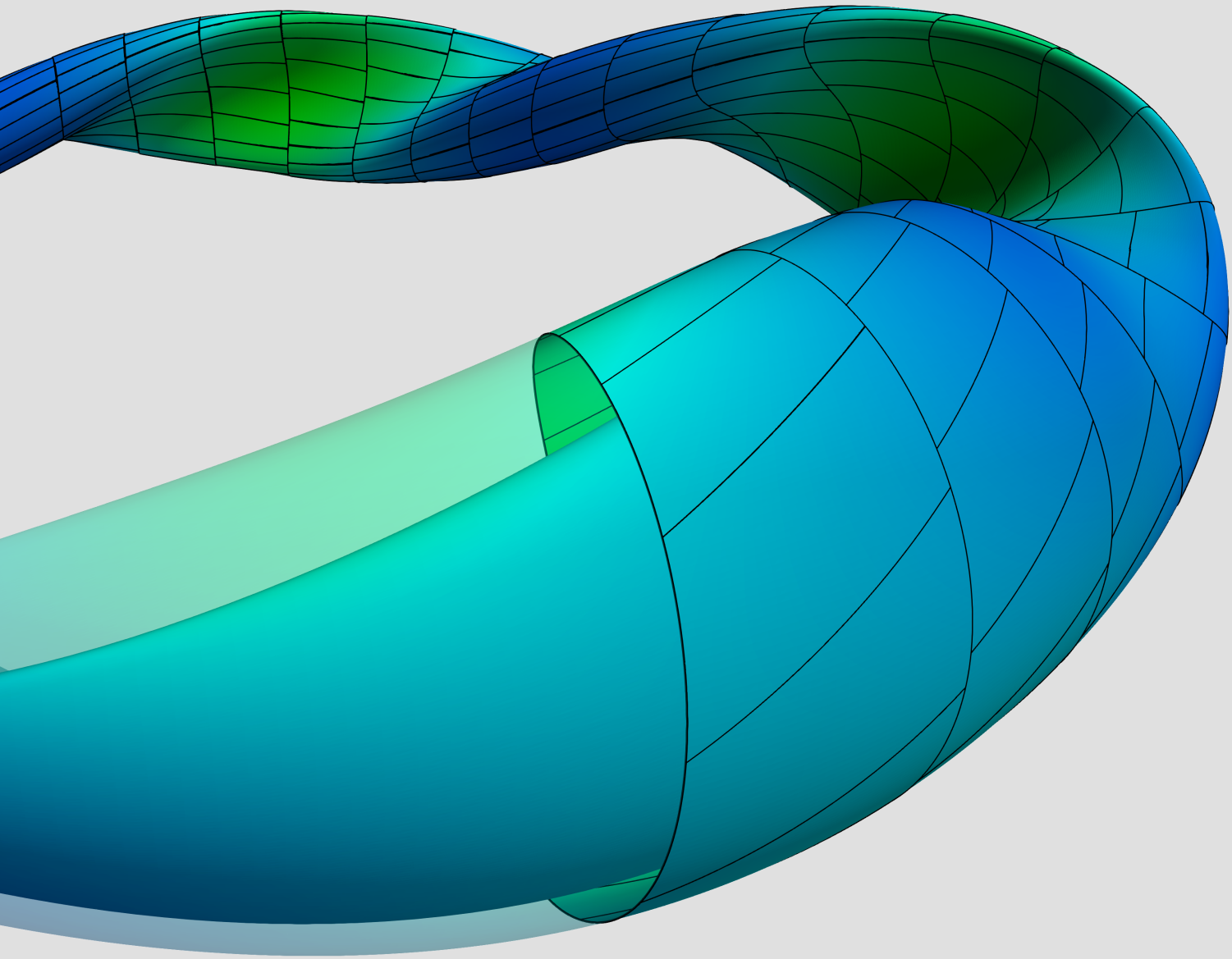
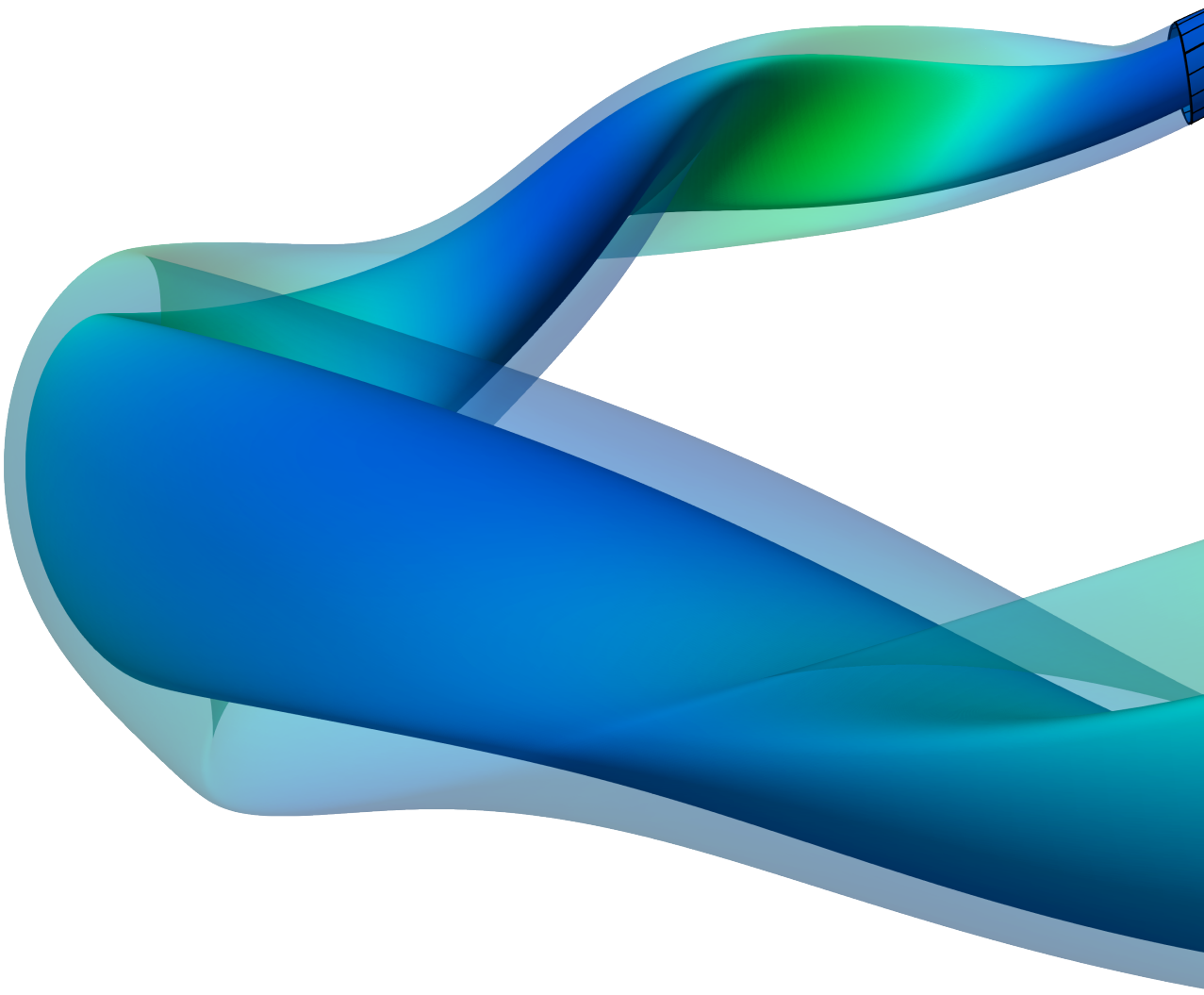
A LOCALLY FIELD-ALIGNED DISCONTINUOUS GALERKIN METHOD

Benedict Dingfelder



`Figure (front pages)`: Nested flux surfaces of a "Wendelstein 7-X"-like magnetohydrodynamic equilibrium. The outermost flux surface is discretized by a locally field-aligned mesh with 40×16 cells of which half is shown. The other half of the flux surface is plotted transparent. Colors depict the norm of the magnetic field with green being large and blue small.



Technische Universität München
Fakultät für Mathematik
Lehrstuhl für Numerische Methoden der Plasmaphysik

Max-Planck-Institut für Plasmaphysik

A locally field-aligned discontinuous Galerkin method

Benedict Dingfelder

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Prof. Dr. Daniel Matthes
Prüfende der Dissertation:	1.	Prof. Dr. Eric Sonnendrücker
	2.	Prof. Dr. Massimo Fornasier
	3.	Prof. Dr. Philippe Helluy

Die Dissertation wurde am 29.03.2018 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 25.06.2018 angenommen.

ABSTRACT

The question of building reliable code is not only a question of computational resources but also extends to the prospect of treating yet untreatable problems. Hence, mathematical structure of the underlying equations needs to be analyzed and incorporated in numerical methods.
With this thesis being motivated by problems in plasma physics, we start with the theory of ideal magnetohydrodynamics and deduce several eigenvalue model problems in two- and three-dimensional geometries. These models include anisotropic wave equations and differential operators up to the fourth order. Their structural analysis yields benefits when separating resolution parallel and perpendicular to the magnetic field. We therefore construct a discontinuous Galerkin method using a non-conforming locally field-aligned mesh in combination with a locally field-aligned basis.
We provide insight on the implementational challenges of the numerical method and evaluate the results of the code developed alongside this thesis. The approach of locally aligning mesh and basis allows to resolve highly oscillatory functions while providing the possibility of a coarse discretization of close-to-constant parts. Furthermore, the size of eigenvalue errors no longer exclusively depends on the frequency of the associated eigenfunction but also on the size of the eigenvalue which supports the accurate calculation of spectra of small eigenvalues. We gain up to 6.5 orders of magnitude in accuracy compared to a non-aligned discontinuous Galerkin method with the same number of degrees of freedom. In particular, results improve for eigenfunctions with high frequencies. The study of problems on flux surfaces of magnetohydrodynamic equilibria yields the same physical behaviour as results of several existing codes which provides the foundation for extending the method to three-dimensional applications.

ZUSAMMENFASSUNG

Die Frage der Erstellung zuverlässigen Codes stellt sich nicht nur hinsichtlich limitierter Rechnerkapazitäten, sondern erweitert sich auf die numerische Lösbarkeit bisher unlösbarer Probleme. Daher ist die Analyse und Einarbeitung der mathematischen Struktur zugrunde liegender Gleichungen unerlässlich.
Diese Arbeit ist motiviert durch plasmaphysikalische Fragestellungen. Wir beginnen mit der Einführung der Theorie der idealen Magnetohydrodynamik und leiten daraus mehrere Eigenwert-Modellprobleme in zwei- und dreidimensionalen Geometrien her. Diese Modelle beinhalten anisotrope Diffusionsgleichungen und Differentialoperatoren bis zu vierter Ordnung. Ihre strukturelle Analyse zeigt Vorteile auf, wenn die Auflösung parallel und senkrecht zum Magnetfeld getrennt festgelegt werden kann. Wir konstruieren daher ein unstetiges Galerkin-Verfahren, das nicht-konforme lokal feldausgerichtete Gitter in Kombination mit lokal feldausgerichteter Basis verwendet.
Wir geben einen Einblick in die Herausforderungen der Implementierung des numerischen Verfahrens und evaluieren die Ergebnisse des zu dieser Arbeit erstellten Codes. Der Ansatz, lokal ausgerichtete Gitter und Basen zu verwenden, ermöglicht die Auflösung hochoszillativer Funktionen bei gleichzeitiger Möglichkeit grober Approximation nahezu konstanter Anteile. Überdies hängt die Größe der Eigenwertfehler nicht mehr nur von der Frequenz der zugehörigen Eigenfunktion ab, sondern auch von der Größe des Eigenwerts selbst. Dies ermöglicht die genaue Berechnung von Spektren kleiner Eigenwerte. Wir erzielen bis zu 6.5 Größenordnungen mehr an Genauigkeit im Vergleich zu nicht-ausgerichteten unstetigen Galerkin-Verfahren mit derselben Zahl an Freiheitsgraden. Insbesondere verbessern sich die Resultate für hochfrequente Eigenfunktionen. Die Untersuchung von Gleichungen auf Flussflächen magnetohydrodynamischer Gleichgewichte reproduziert das physikalische Verhalten der Ergebnisse bereits existenter Codes. Dies bereitet die Grundlage für die Erweiterbarkeit des Verfahrens hin zu dreidimensionalen Anwendungsfällen.

CONTENTS

SYMBOLS AND ABBREVIATIONS

$\boldsymbol{b}$	Direction of the magnetic field on a flux surface
$\boldsymbol{b}_\perp$	Direction perpendicular to $\boldsymbol{b}$
$\boldsymbol{b}_{\text{mesh}}$	Direction of the local alignment of the mesh
DoF	Degrees of freedom of the discretization
$\text{DoF}_\parallel$	Degrees of freedom for discretizing the parallel direction
$\text{DoF}_\perp$	Degrees of freedom for discretizing the perpendicular direction
ι	Rotational transform, in practice first component of $\boldsymbol{b}$ if $\boldsymbol{b} \equiv (\iota, 1)^\top$
m_{max}	Maximal frequency/mode number of an eigenmode in the first dimension (poloidally)
n_{max}	Maximal frequency/mode number of an eigenmode in the second dimension (toroidally)
$\boldsymbol{n}$	Outer unit normal
nnzA	Number of non-zeroes of the left hand side system matrix
N_Σ	Total number of cells in mesh
N_x	Number of cells in x-direction, parallel mesh resolution
N_y	Number of cells in y-direction, perpendicular mesh resolution
p_ξ	Degree of the basis in ξ-direction, parallel degree
p_η	Degree of the basis in η-direction, perpendicular degree
s	Normalized flux surface coordinate, $s \in [0, 1]$
Ω	Fully periodic computational domain $[0, 2\pi)^2$
ω^2	Eigenvalue
ω^2_{max}	Upper bound for eigenvalues of interest
$\omega^2_{m,n}$	Eigenvalue associated to mode (m, n)

ADG	Locally field-aligned discontinuous Galerkin (method)
BR2	Bassi-Rebay 2
DG	Discontinuous Galerkin (method)
LDG	Local discontinuous Galerkin
LHD	Large Helical Device
MHD	Magnetohydrodynamic(s)
W7-X	Wendelstein 7-X

Chapter 1

MOTIVATION AND OUTLINE
Laying the foundation

One of the biggest challenges of the upcoming decades is the provision of clean and reliable energy to limit the impact of the climate change while meeting the ever growing global power demand. [1, 2] predict an increase of roughly 30% in the global energy needs until 2040 in comparison to 2017 with the demand for electricity increasing by roughly 40%.
As of March 2018, 195 states have signed the Paris Agreement which aims at limiting "the increase of the global average temperature to well below 2°C above pre-industrial levels" [3]. The German government aspires to a reduction of greenhouse gas emissions by $80 - 95\%$ until 2050 in comparison to 1990 [4] and a share of 65% of renewable energies in the power generation until 2030 [5]. To produce electricity while emitting close to no greenhouse gases, the inclusion of power plants relying on fossil fuels has to be cut to a minimum. A possibility is to increase the share of renewable energies to 100% in the long term.
Considering Germany, the expansion of wind and solar power plants should be limited to keep the public approval of the energy revolution [6, 7]. However, expanding other renewable energies exhibit limited potential [6, p.8]. Furthermore, the cost of the German energy revolution amounts to a total of 1 to 2×10^{12} Euro until 2050 [6, p.79].
By analyzing the electricity data of Germany, France and Italy, [8, 7, 9] have found that the amount of energy storage capacities needed to sustain a 100% share of renewable energies is enormous and systems held for backup operate at a capacity of less than 17%. Considering Germany, a backup-system supplementing about 89% of the peak load would have to be established [7]. Creating a sufficiently interconnected power grid across the European Union, allows to cut backup capacities by roughly 30% in comparison [10]. Relying on the construction of pumped-storage plants in Norway within the eStorage project [11], Germany's share of wind and solar energy can be increased loss-free to a maximum of 60% [12] which is still far away from a share of 100% and not accounting for storage necessary for the remaining countries within the European Union.

Hence, the provision of sufficient storage capacities for the whole European Union cannot be met by present technology and massive investments in research on storage techniques is required.
For cutting the amount of storage systems needed, strategies of combining storage with baseload plants are proposed [9]. Having fossil fuel based baseload plants in reserve in power systems with high shares of wind energy pushes these plants into part-time operation and ramping which results in increased outages and plant depreciation [13]. Besides, these baseload plants don't operate free of greenhouse gases. The only alternative to date is the production of energy via nuclear fission. Germany is shutting down its remaining nuclear power plants until the end of 2022 [14] and a change of the political policy is unlikely. Furthermore, the question of depositing highly radioactive waste remains unsolved. Hence, the necessity for research on another type of baseload plant emitting no greenhouse gases is evident.
A mere 100 years ago, Aston deduced in 1920 that the fusion of hydrogen atoms to helium yields energy as pointed out in [15] and ibidem Eddington suspected that exactly this is the reaction powering the stars which was later proven in 1939 by Bethe [16]. After research on nuclear fusion was mainly pursued for military purposes within the Manhattan Project [17], concepts for harnessing fusion energy were developed in the 1950s. Spitzer designed the setup known as a Stellarator [18]. Inspired by the ideas of Lavrentiev, Tamm and Sakharov designed the setup known as a Tokamak [19]. Both systems rely on the magnetic confinement of hot plasma in a toroidal geometry for achieving thermonuclear fusion [20]. Other approaches include inertial confinement [21], inertial electrostatic confinement [22] or muon-catalyzed fusion [23].
Considering the further development of machines relying on magnetically confined plasma, the findings of Tokamak experiment T3 [24] resulted in a concentration of research effort on the Tokamak design [25]. The later emerging problems of the Tokamak approach [26] and the progression of advanced computer aided design reanimated research on Stellarator configurations [27].
Currently operated Tokamak experiments include the Joint European Torus (JET) being "the only operational fusion experiment capable of producing fusion energy" [28] and ASDEX Upgrade, preparing the scientific foundation for a fusion power plant of the Tokamak type [29]. Both prepare the groundwork for the International Thermonuclear Experimental Reactor (ITER), a currently constructed Tokamak being the first experiment expected to yield a tenfold return on energy [30].
Currently operated Stellarator experiments include the Large Helical Device (LHD), evaluating plasma confinement properties in helical geometries [31], and Wendelstein 7-X (W7-X), evaluating the suitability of the Stellarator concept for a fusion power plant [32].
Fusion energy seems to be a perfect candidate for baseload plants. First of all, the fusion reaction yields Helium as a product and consequently no greenhouse gases. In fusion reactors, nuclear meltdowns are scientifically impossible. In the process of fusion, surrounding materials become lightly radioactively activated due to free neutrons created in the fusion reaction. However, the activity of the components quickly declines and exhibits only a ten-thousandth of its initial activity after 100 years [33]. The reactors are fueled by a mixture of hydrogen isotopes, namely deuterium and tritium, of which one gram yields the same heating power as 11.5 tons of coal or 7.5 to 8 tons

of oil, which can be deduced by the energy gain of the fusion reaction [34] and its conversion to the respective units, see [35, Appendix 1.B] and [36, Appendix B.]. Deuterium can be separated from water using the Girdler sulfide process and electrolysis [37] whereas tritium can be produced via a Lithium breeding process [38].
Politically, European institutes for fusion research are currently organized within the EUROfusion consortium, aiming to develop fusion as a viable energy source by 2050 [39].

This thesis is located within fusion research, namely within the division for developing new computational methods for plasma physics. As computing times and disk space are always limited, proper numerical algorithms have to use the available resources responsibly and focus their computational effort on whats necessary. The identification of these necessities is however an intricate question. Often, mathematical models in plasma physics are extremely complex due to the abundance of details which can be taken into consideration such as the simulation of a bandwidth of different physical processes within the plasma, the geometry of the domain of consideration and the inclusion of boundary conditions.
However, even complicated models can exhibit a certain structure when lifted to a high level of mathematical abstraction. Building numerical methods which preserve or exploit these identified structures yields numerical solutions in which certain physical conservation laws are already fulfilled by default or yields code applicable to problems which, if tackled by standard techniques, could only if at all be solved extremely slowly. Hence, the question of building reliable code is not only a question of computational resources but also extends to the prospect of treating yet untreatable problems.
This thesis focuses on the identifications of the just mentioned mathematical structure of the underlying physical equations and its incorporation in numerical methods as well as implementation and assessment of these methods. Hence, explanations of the physical background and its equations are cut to the level which is needed for introducing the mathematical theory and interpretation of the results.

We start with the introduction of ideal magnetohydrodynamic (MHD) equations in Chapter 2 and reduce them to a set of structurally equivalent model problems. These model problems are meant to act as representatives for the full set of physical equations, meaning that a method capable of treating the model problems should be able to treat the full set of equations as well. Amongst others, we deduce reduced MHD equations, the reduced MHD shear Alfvén wave equation and anisotropic wave problems. Furthermore, we dwell onto the physical motivations for considering these problems.
Chapter 3 constructs a discontinuous Galerkin (DG) method for treating anisotropic wave equations in two- and three-dimensional geometries. In comparison to regular heterogeneous diffusion equations, the anisotropic wave equation with constant coefficients is not purely elliptic due to the semidefiniteness of the tensor yielding an ill-posed problem. Different kinds of meshes for the

domain discretization are analyzed predominantly with respect to the capability of addressing resolution in different preset directions separately.
DG methods can be viewed as a combination of finite element and finite volume methods. The domain is subdivided into distinct mesh elements on which the solution is approximated typically by high order polynomials, a feature of finite element methods. Across the domain interfaces, solutions are allowed to be discontinuous and a communication mechanism known from finite volumes via so-called numerical fluxes is established. DG methods are highly scalable and well suited for large-scale computations as their discretization stencils are compact [40, p.V], and allow for an intuitive treatment of non-conforming meshes. Furthermore, the flexibility offered by the subdivision of the domain, the choice of basis and numerical fluxes can and should be exploited by the numerical code.
The first DG method was introduced in 1973 for hyperbolic equations [41]. An independent approach using discontinuous finite elements originated in the early 1970s in [42, 43, 44] later to be adapted and then known as interior penalty methods [45, 46, 47]. In the late 1990s, DG methods underwent a significant further development. We remark on the extension of DG methods to purely elliptical problems such as to the compressible Navier-Stokes equations [48], diffusion problems [49] and convection-diffusion problems [50] and [51] with the latter being known as the local discontinuous Galerkin (LDG) method. Summaries providing a common framework for DG methods are given by [52, 40].
Chapter 4 approaches a more complicated model problem involving up to fourth order differentials. A DG method based on a mixed variational form involving differentials of at most first order is constructed.
Chapter 5 focuses on the challenges of the development of the `FORTRAN`-code implemented alongside this thesis for the method developed in the preceding chapters. The choice for `FORTRAN` is justified by the considerations of performing large-scale computations in particular in view of future extensions to equations mirroring more physical behaviour than the ones focused on herein.
Chapter 6 numerically analyzes the described and implemented method. The results are evaluated from different points of view, compared with a non-aligned DG method and summarized. For three-dimensional geometries, we compare with multiple existing codes.
Chapter 7 closes with a summary of the findings for the locally field-aligned discontinuous Galerkin method constructed in this thesis and provides an outlook on future perspectives in development.

Chapter 2

PHYSICAL DERIVATION
Building mathematical models

This chapter provides a short introduction to magnetohydrodynamics (MHD) which "is a fluid model that describes the macroscopic equilibrium and stability properties of a plasma" [53, p.1]. We perform a gradual simplification of its equations for building model problems. As we focus on the analysis of the mathematical structure as outlined in Chapter 1, we keep the classification and explanation of physical phenomena and implications associated to the equations to a minimum. We consider the "most basic version" of MHD, called ideal MHD equations, which "assumes that the plasma can be represented by a single fluid with infinite electrical conductivity and zero ion gyro radius" [53, p.1].
The outline of this chapter is as follows: Starting with the equations of ideal MHD in Section 2.1, we introduce the notion of MHD equilibria and highlight several properties and their importance for fusion research in Section 2.2. After linearizing the ideal MHD equations around a static MHD equilibrium state in Section 2.3, we deduce reduced MHD equations in Section 2.4 given several physical assumptions. From this set of equations, we build a first model problem in Section 2.5 involving a fourth order spatial operator by introducing an ansatz for a wave solution in space and time. The transformation of the coordinate system to straight field line coordinates and further simplifications in Section 2.6 yield an anisotropic wave equation with a second order spatial operator. Section 2.7 closes with a motivation of the physically interesting properties of the solution of the constructed model problems and provides an outlook on important considerations for features of numerical methods when returning from model problems to their origin.

2.1 Ideal MHD

As a starting point, we introduce the time-dependent ideal MHD model in three dimensions given by [53, p.9, (2.1)]

$$\text{Mass conservation:} \quad \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0 \tag{2.1}$$

$$\text{Momentum:} \quad \rho \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = \boldsymbol{J} \times \boldsymbol{B} - \nabla p \tag{2.2}$$

$$\text{Energy:} \quad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{p}{\rho^{\gamma}}\right) = 0 \tag{2.3}$$

$$\text{Ohm's law:} \quad \boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B} = 0 \tag{2.4}$$

$$\text{Maxwell–Faraday equation:} \quad \nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{2.5}$$

$$\text{Ampère's circuital law:} \quad \nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{J} \tag{2.6}$$

$$\text{Gauss's law for magnetism:} \quad \nabla \cdot \boldsymbol{B} = 0 \tag{2.7}$$

where ρ is the mass density, v the fluid velocity, $\boldsymbol{J}$ is the current density, $\boldsymbol{B}$ the magnetic field, p the plasma pressure, $\gamma = 5/3$ the ratio of specific heats, $\boldsymbol{E}$ the electric field and μ_0 the permeability of free space. $\nabla\cdot$ depicts the divergence operator, $\nabla\times$ is the curl and $\times$ the cross product. Furthermore,

$$\frac{\mathrm{d}}{\mathrm{d}t} := \frac{\partial}{\partial t} + \boldsymbol{v} \cdot \nabla \tag{2.8}$$

is the convective derivative.
Combination of Maxwell-Faraday (2.5) and Ohm's law (2.4) yields

$$\frac{\partial \boldsymbol{B}}{\partial t} = \nabla \times (\boldsymbol{v} \times \boldsymbol{B}) \quad \text{(induction equation)} \ . \tag{2.9}$$

2.2 MHD equilibria

The goal of ideal MHD equilibrium theory is the discovery of magnetic geometries which are of interest for fusion reactors, i.e., they stably confine hot plasmas at a sufficiently high ratio of plasma pressure to magnetic pressure [53, p.80]. An overview of MHD equilibrium theory can be found in [53, Chapter 4]. We provide a short introduction to this theory focusing on the information needed to understand the results of Section 6.4.
A static MHD equilibrium is a steady state solution of the ideal MHD equations (2.1) – (2.7) for vanishing fluid velocity

$$\boldsymbol{v} = 0 \ . \tag{2.10}$$

Insertion in (2.1) – (2.7) yields the time-independent equations for static equilibria [53, p.58, (4.1)]

$$\text{Force balance equation:} \qquad \boldsymbol{J}_0 \times \boldsymbol{B}_0 = \nabla p_0 \tag{2.11}$$

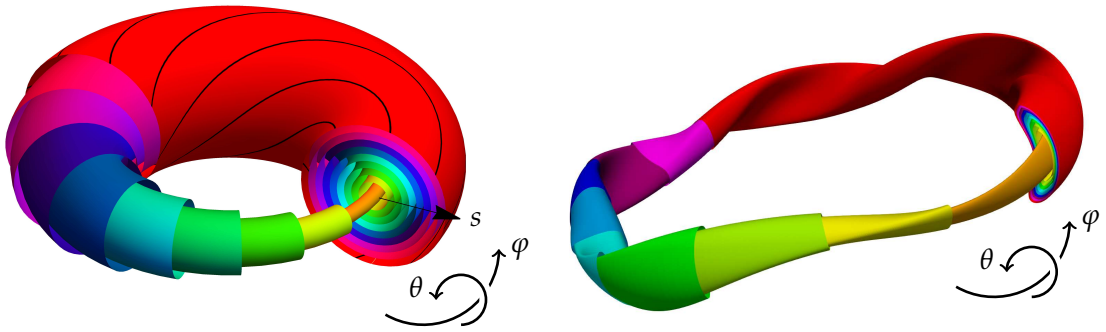
$$\text{Ampère's circuital law:} \qquad \nabla \times \boldsymbol{B}_0 = \mu_0 \boldsymbol{J}_0 \tag{2.12}$$

$$\text{Gauss's law for magnetism:} \qquad \nabla \cdot \boldsymbol{B}_0 = 0\,. \tag{2.13}$$

All equilibrium quantities are marked with subscript 0. For static equilibria, it holds $v_0 = 0$.
The virial theorem [53, Section 4.3] states that the existence of an MHD equilibrium requires an external magnetic field. Hence, no self-confining plasma configuration can be established and fusion reactors need the setup of an external system of coils providing the said magnetic field. A common feature of the design of Tokamaks and Stellarators is the toroidal geometry which is motivated by "the thermal conduction loss rate parallel to the magnetic field [being] enormous compared to that perpendicular to the field" [53, p.81].
Forming the dot product of (2.11) with $\boldsymbol{B}_0$ we obtain [53, p.63, (4.9)]

$$\boldsymbol{B}_0 \cdot \nabla p_0 = 0 \tag{2.14}$$

which shows that the three-dimensional equilibrium can be subdivided into a set of nested so-called flux surfaces on which magnetic lines reside, i.e., lines sharing the same direction on this surface, and $p_0 \equiv \text{const}$ [53, Section 4.5].



(a) Plain torus (b) W7-X-like geometry

FIGURE 2.1: *Sketch of nested flux surfaces for different geometries. The left picture shows magnetic field lines on the outermost surface.*

Figure 2.1 shows sketches of nested flux surfaces for a plain torus and a W7-X-like geometry. The innermost flux surface which degenerates to a single line is called the magnetic axis [53, p.63]. Being the surface of a (transformed) torus, a flux surface can be parameterized by two angles $(\theta, \varphi) \in [0, 2\pi)^2$ with θ being the angle of poloidal rotation and φ being the angle of toroidal rotation. Flux surfaces can be characterized by the rotational transform ι which is the average

change of the poloidal angle on a poloidal cross section after following a magnetic field line for one toroidal transit [53, Section 4.6.4]. If so-called straight field line coordinates are used, the two flux surface angles are chosen such that the direction of the magnetic field on a flux surface is described by

$$\begin{pmatrix} b^\theta \\ b^\varphi \end{pmatrix} = \begin{pmatrix} \iota(s) \\ 1 \end{pmatrix} \tag{2.15}$$

with $s \in [0,1]$ being a variable identifying a normalized distance from the center of the torus. $s = 0$ represents the magnetic axis and $s = 1$ the outermost flux surface. Examples for straight field line coordinates are Boozer [54] and PEST coordinates [55]. On so-called rational flux surfaces, $\iota(s)$ is a rational number which therefore yields field lines connecting on themselves.

2.3 Linearized ideal MHD

The linearized ideal MHD model "is a single-fluid model that describes the effects of magnetic geometry on the macroscopic equilibrium and stability properties of fusion plasmas" [53, p.35]. For investigating small deviations from a given time-independent MHD equilibrium state which we denote by subscript zero, we linearize all quantities in time around this state such that [53, p.329, (8.1)]

$$\boldsymbol{Q}(\boldsymbol{x},t) = \boldsymbol{Q}_0(\boldsymbol{x}) + \tilde{\boldsymbol{Q}}_1(\boldsymbol{x},t) \qquad , \qquad \left|\frac{\tilde{\boldsymbol{Q}}_1}{\boldsymbol{Q}_0}\right| = \varepsilon \ll 1 \tag{2.16}$$

where $\boldsymbol{Q}$ acts as a possibly vector-valued representative variable and $\boldsymbol{x} = (x_1, x_2, x_3)^\top$. In particular, we define the linear term of the fluid velocity $\boldsymbol{v}$ as [53, p.335, (8.13)]

$$\frac{\partial \tilde{\boldsymbol{\xi}}}{\partial t} := \tilde{\boldsymbol{v}}_1 \tag{2.17}$$

such that $\tilde{\boldsymbol{\xi}}$ depicts the displacement of the plasma from the equilibrium. As we consider static equilibria, the velocity $\boldsymbol{v}_0$ vanishes. As equations are linearized, we summarize terms including the product of two or more time-dependent variables which then are of order $\mathcal{O}\left(\varepsilon^2\right), \varepsilon \to 0$ and drop them. The linearized version of the mass conservation equation (2.1) is obtained by first writing

$$\begin{aligned}
\frac{\partial}{\partial t}\left(\rho_0 + \tilde{\rho}_1\right) + \nabla \cdot \left(\left(\rho_0 + \tilde{\rho}_1\right)\left(\boldsymbol{v}_0 + \tilde{\boldsymbol{v}}_1\right)\right) &= 0 \\
\frac{\partial \tilde{\rho}_1}{\partial t} + \nabla \cdot \left(\left(\rho_0 + \tilde{\rho}_1\right)\tilde{\boldsymbol{v}}_1\right) &= 0 \\
\frac{\partial \tilde{\rho}_1}{\partial t} + \nabla \cdot \left(\rho_0 \tilde{\boldsymbol{v}}_1\right) + \mathcal{O}\left(\varepsilon^2\right) &= 0 \\
\frac{\partial \tilde{\rho}_1}{\partial t} + \tilde{\boldsymbol{v}}_1 \cdot \nabla \rho_0 + \rho_0 \nabla \cdot \tilde{\boldsymbol{v}}_1 + \mathcal{O}\left(\varepsilon^2\right) &= 0
\end{aligned} \tag{2.18}$$

which gives

$$\frac{\partial \tilde{\rho}_1}{\partial t} + \tilde{\boldsymbol{v}}_1 \cdot \nabla \rho_0 + \rho_0 \nabla \cdot \tilde{\boldsymbol{v}}_1 = 0 \quad \text{(linearized mass conservation)} \ . \tag{2.19}$$

For linearizing the momentum equation (2.2), we analyze the left hand side using (2.8)

$$\begin{aligned} \rho \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} &= (\rho_0 + \rho_1) \left(\frac{\partial}{\partial t} (\boldsymbol{v}_0 + \tilde{\boldsymbol{v}}_1) + (\boldsymbol{v}_0 + \tilde{\boldsymbol{v}}_1) \cdot \nabla (\boldsymbol{v}_0 + \tilde{\boldsymbol{v}}_1) \right) \\ &= (\rho_0 + \rho_1) \left(\frac{\partial \tilde{\boldsymbol{v}}_1}{\partial t} + \tilde{\boldsymbol{v}}_1 \cdot \nabla \tilde{\boldsymbol{v}}_1 \right) = \rho_0 \frac{\partial \tilde{\boldsymbol{v}}_1}{\partial t} + \mathcal{O}\left(\varepsilon^2\right) \end{aligned} \tag{2.20}$$

and the right hand side using the force balance equation (2.11) to obtain

$$\begin{aligned} \boldsymbol{J} \times \boldsymbol{B} - \nabla p &= \left(\boldsymbol{J}_0 + \tilde{\boldsymbol{J}}_1\right) \times \left(\boldsymbol{B}_0 + \tilde{\boldsymbol{B}}_1\right) - \nabla \left(p_0 + \tilde{p}_1\right) \\ &= \boldsymbol{J}_0 \times \boldsymbol{B}_0 - \nabla p_0 + \tilde{\boldsymbol{J}}_1 \times \boldsymbol{B}_0 + \boldsymbol{J}_0 \times \tilde{\boldsymbol{B}}_1 + \tilde{\boldsymbol{J}}_1 \times \tilde{\boldsymbol{B}}_1 - \nabla \tilde{p}_1 \\ &= \tilde{\boldsymbol{J}}_1 \times \boldsymbol{B}_0 + \boldsymbol{J}_0 \times \tilde{\boldsymbol{B}}_1 - \nabla \tilde{p}_1 + \mathcal{O}\left(\varepsilon^2\right) \end{aligned} \tag{2.21}$$

which we combine using (2.17) to

$$\rho_0 \frac{\partial^2 \tilde{\boldsymbol{\xi}}}{\partial t^2} = \tilde{\boldsymbol{J}}_1 \times \boldsymbol{B}_0 + \boldsymbol{J}_0 \times \tilde{\boldsymbol{B}}_1 - \nabla \tilde{p}_1 \quad \text{(linearized momentum)} \ . \tag{2.22}$$

For linearizing the energy equation (2.3), we first consider

$$\begin{aligned} 0 = \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{p}{\rho^\gamma} \right) &= \left(\frac{\partial}{\partial t} + \boldsymbol{v} \cdot \nabla \right) \left(\frac{p}{\rho^\gamma} \right) \\ &= \rho^{-\gamma} \frac{\partial p}{\partial t} - \gamma p \rho^{-\gamma-1} \frac{\partial \rho}{\partial t} + \rho^{-\gamma} \boldsymbol{v} \cdot \nabla p - \gamma \rho^{-\gamma-1} p \boldsymbol{v} \cdot \nabla \rho \ . \end{aligned} \tag{2.23}$$

Multiplication with ρ^γ and insertion of the mass conservation (2.1) yields

$$\begin{aligned} 0 &= \frac{\partial p}{\partial t} + \gamma \frac{p}{\rho} \nabla \cdot (\rho \boldsymbol{v}) + \boldsymbol{v} \cdot \nabla p - \gamma \frac{p}{\rho} \boldsymbol{v} \cdot \nabla \rho \\ &= \frac{\partial p}{\partial t} + \gamma \frac{p}{\rho} \rho \nabla \cdot \boldsymbol{v} + \gamma \frac{p}{\rho} \boldsymbol{v} \cdot \nabla \rho + \boldsymbol{v} \cdot \nabla p - \gamma \frac{p}{\rho} \boldsymbol{v} \cdot \nabla \rho \\ &= \frac{\partial p}{\partial t} + \gamma p \nabla \cdot \boldsymbol{v} + \boldsymbol{v} \cdot \nabla p \end{aligned} \tag{2.24}$$

which we linearize to

$$\begin{aligned} 0 &= \frac{\partial}{\partial t} \left(p_0 + \tilde{p}_1\right) + \gamma \left(p_0 + \tilde{p}_1\right) \nabla \cdot \left(\boldsymbol{v}_0 + \tilde{\boldsymbol{v}}_1\right) + \left(\boldsymbol{v}_0 + \tilde{\boldsymbol{v}}_1\right) \cdot \nabla \left(p_0 + \tilde{p}_1\right) \\ &= \frac{\partial \tilde{p}_1}{\partial t} + \gamma \left(p_0 + \tilde{p}_1\right) \nabla \cdot \tilde{\boldsymbol{v}}_1 + \tilde{\boldsymbol{v}}_1 \cdot \nabla \left(p_0 + \tilde{p}_1\right) \\ &= \frac{\partial \tilde{p}_1}{\partial t} + \gamma p_0 \nabla \cdot \tilde{\boldsymbol{v}}_1 + \tilde{\boldsymbol{v}}_1 \cdot \nabla p_0 + \mathcal{O}\left(\varepsilon^2\right) \end{aligned} \tag{2.25}$$

such that the summarized linear version of the energy equation (2.3) writes

$$\frac{\partial \tilde{p}_1}{\partial t} + \gamma p_0 \nabla \cdot \tilde{v}_1 + \tilde{v}_1 \cdot \nabla p_0 = 0 \quad \text{(linearized pressure)} \tag{2.26}$$

which we title as the linearized pressure equation since only pressure and velocity terms are involved due to the substitution of the density via the mass conservation.
Furthermore, we linearize Ohm's law (2.4)

$$\begin{aligned} \boldsymbol{E}_0 + \tilde{\boldsymbol{E}}_1 &= -\left(\boldsymbol{v}_0 + \tilde{\boldsymbol{v}}_1\right) \times \left(\boldsymbol{B}_0 + \tilde{\boldsymbol{B}}_1\right) \\ &= -\tilde{\boldsymbol{v}}_1 \times \boldsymbol{B}_0 + \mathcal{O}\left(\varepsilon^2\right) \end{aligned} \tag{2.27}$$

which yields

$$\tilde{\boldsymbol{E}}_1 = -\tilde{\boldsymbol{v}}_1 \times \boldsymbol{B}_0 \quad \text{(linearized Ohm's law)} \tag{2.28}$$

and the induction equation (2.9) as a combination of Ohm's law (2.4) and the Maxwell-Faraday equation (2.5) by

$$\begin{aligned} \frac{\partial}{\partial t}\left(\boldsymbol{B}_0 + \tilde{\boldsymbol{B}}_1\right) &= \nabla \times \left(\left(\boldsymbol{v}_0 + \tilde{\boldsymbol{v}}_1\right) \times \left(\boldsymbol{B}_0 + \tilde{\boldsymbol{B}}_1\right)\right) \\ \frac{\partial}{\partial t}\tilde{\boldsymbol{B}}_1 &= \nabla \times \left(\tilde{\boldsymbol{v}}_1 \times \left(\boldsymbol{B}_0 + \tilde{\boldsymbol{B}}_1\right)\right) \\ \frac{\partial}{\partial t}\tilde{\boldsymbol{B}}_1 &= \nabla \times \left(\tilde{\boldsymbol{v}}_1 \times \boldsymbol{B}_0\right) + \nabla \times \left(\tilde{\boldsymbol{v}}_1 \times \tilde{\boldsymbol{B}}_1\right) \\ \frac{\partial}{\partial t}\tilde{\boldsymbol{B}}_1 &= \nabla \times \left(\tilde{\boldsymbol{v}}_1 \times \boldsymbol{B}_0\right) + \mathcal{O}\left(\varepsilon^2\right) \end{aligned} \tag{2.29}$$

which we integrate in time using (2.17) to obtain

$$\tilde{\boldsymbol{B}}_1 = \nabla \times \left(\tilde{\boldsymbol{\xi}} \times \boldsymbol{B}_0\right) \quad \text{(linearized induction equation)} \ . \tag{2.30}$$

For the linearization of Ampère's circuital law (2.6), we obtain

$$\mu_0 \boldsymbol{J}_0 = \nabla \times \boldsymbol{B}_0 \qquad , \qquad \mu_0 \tilde{\boldsymbol{J}}_1 = \nabla \times \tilde{\boldsymbol{B}}_1 \ . \tag{2.31}$$

Substituting these equations and the linearized induction equation (2.30) into the linearized momentum (2.22), we obtain the normal mode formulation of the linearized MHD stability problem for general three-dimensional equilibria [53, p.336]

$$\rho_0 \mu_0 \frac{\partial^2 \tilde{\boldsymbol{\xi}}}{\partial t^2} = \left(\nabla \times \left(\nabla \times \left(\tilde{\boldsymbol{\xi}} \times \boldsymbol{B}_0\right)\right)\right) \times \boldsymbol{B}_0 + \left(\nabla \times \boldsymbol{B}_0\right) \times \left(\nabla \times \left(\tilde{\boldsymbol{\xi}} \times \boldsymbol{B}_0\right)\right) - \mu_0 \nabla \tilde{p}_1 \ . \tag{2.32}$$

Lastly, the linearization of Gauss's law for magnetism (2.7) becomes

$$\nabla \cdot \boldsymbol{B}_0 = 0 \qquad , \qquad \nabla \cdot \tilde{\boldsymbol{B}}_1 = 0 \ . \tag{2.33}$$

2.4 Reduction of ideal MHD

For building model problems, we execute a gradual reduction process. We start with the linearized ideal MHD equations of Section 2.3.
First, we introduce the magnetic vector potential $\boldsymbol{A}$ and the electric potential ϕ defined by [56, p.15–15]

$$\boldsymbol{B} = \nabla \times \boldsymbol{A} \tag{2.34}$$

$$\boldsymbol{E} = -\nabla\phi - \frac{\partial \boldsymbol{A}}{\partial t} \,. \tag{2.35}$$

Furthermore, we define the parallel and perpendicular component of a vector-valued variable $\boldsymbol{Q}$ as

$$Q_{\parallel} \coloneqq \hat{\boldsymbol{b}} \cdot \boldsymbol{Q} \qquad , \qquad \boldsymbol{Q}_{\perp} \coloneqq -\hat{\boldsymbol{b}} \times \hat{\boldsymbol{b}} \times \boldsymbol{Q} \qquad , \qquad \hat{\boldsymbol{b}} \coloneqq \frac{\boldsymbol{B}_0}{\|\boldsymbol{B}_0\|_2} \,. \tag{2.36}$$

Note that the parallel component is defined as a scalar and not a vector. We denote this by removing the bold font of the vector-valued variable $\boldsymbol{Q}$. For reducing the equations, we split the electric field $\tilde{\boldsymbol{E}}_1$ and the current density $\tilde{\boldsymbol{J}}_1$ into parallel and perpendicular components.
First, linearization of (2.35) yields

$$\boldsymbol{E}_0 + \tilde{\boldsymbol{E}}_1 = -\nabla\left(\phi_0 + \tilde{\phi}_1\right) - \frac{\partial}{\partial t}\left(\boldsymbol{A}_0 + \tilde{\boldsymbol{A}}_1\right) \tag{2.37}$$

$$\tilde{\boldsymbol{E}}_1 = -\nabla\tilde{\phi}_1 - \frac{\partial \tilde{\boldsymbol{A}}_1}{\partial t} \tag{2.38}$$

from which we extract the parallel component of the electric field variation using the linearized Ohm's law (2.28) which yields

$$\left(\tilde{E}_1\right)_{\parallel} = \hat{\boldsymbol{b}} \cdot \left(-\tilde{\boldsymbol{v}}_1 \times \boldsymbol{B}_0\right) = 0 \tag{2.39}$$

since $\tilde{\boldsymbol{v}}_1 \times \boldsymbol{B}_0$ is perpendicular to $\boldsymbol{B}_0$. We obtain

$$0 = -\hat{\boldsymbol{b}} \cdot \nabla\tilde{\phi}_1 - \hat{\boldsymbol{b}} \cdot \frac{\partial \tilde{\boldsymbol{A}}_1}{\partial t} = -\hat{\boldsymbol{b}} \cdot \nabla\tilde{\phi}_1 - \frac{\partial}{\partial t}\left(\hat{\boldsymbol{b}} \cdot \tilde{\boldsymbol{A}}_1\right) \tag{2.40}$$

where the last equality holds as $\hat{\boldsymbol{b}}$ is time-independent. We summarize these findings in

$$\hat{\boldsymbol{b}} \cdot \nabla\tilde{\phi}_1 = -\frac{\partial\left(\tilde{A}_1\right)_{\parallel}}{\partial t} \,. \tag{2.41}$$

Second, the parallel component of the linearization of Ampère's circuital law (2.31) with insertion of the magnetic vector potential (2.34) yields

$$\mu_0\left(\tilde{J}_1\right)_{\parallel} = \hat{\boldsymbol{b}} \cdot \left(\nabla \times \left(\nabla \times \tilde{\boldsymbol{A}}_1\right)\right) \,. \tag{2.42}$$

When neglecting compressional Alfvén waves, the perpendicular component of the variation of the magnetic vector potential $\tilde{\boldsymbol{A}}_1$ is comparably small and therefore [57, p.159, (B.13)]

$$\tilde{\boldsymbol{A}}_1 = \hat{\boldsymbol{b}} \left(\tilde{A}_1\right)_{\parallel} + \left(\tilde{\boldsymbol{A}}_1\right)_{\perp} \approx \hat{\boldsymbol{b}} \left(\tilde{A}_1\right)_{\parallel} \ . \tag{2.43}$$

Then, (2.42) can be simplified via drift approximation to [58, p. 965, (13)]

$$\mu_0 \left(\tilde{J}_1\right)_{\parallel} \approx -\nabla_{\perp}^2 \left(\tilde{A}_1\right)_{\parallel} \tag{2.44}$$

where we introduce the notation for perpendicular derivatives

$$\nabla_{\perp}^2 \coloneqq \nabla \cdot \nabla_{\perp} \qquad , \qquad \nabla_{\perp} \coloneqq \nabla - \hat{\boldsymbol{b}}\hat{\boldsymbol{b}} \cdot \nabla = -\hat{\boldsymbol{b}} \times \left(\hat{\boldsymbol{b}} \times \nabla\right) \ . \tag{2.45}$$

Lastly, we apply the divergence operator to Ampère's circuital law (2.6) to obtain

$$\nabla \cdot (\nabla \times \boldsymbol{B}) = 0 = \mu_0 \nabla \cdot \boldsymbol{J} \tag{2.46}$$

whose linearized form is then decomposed into

$$0 = \mu_0 \nabla \cdot \left(\left(\tilde{\boldsymbol{J}}_1\right)_{\perp} + \hat{\boldsymbol{b}} \left(\tilde{J}_1\right)_{\parallel}\right) \ . \tag{2.47}$$

For considering the perpendicular component of $\tilde{\boldsymbol{E}}_1$, we take the cross product of the linearized Ohm's law (2.28) with $\boldsymbol{B}_0$ and obtain

$$\boldsymbol{B}_0 \times \tilde{\boldsymbol{E}}_1 = -\boldsymbol{B}_0 \times \tilde{\boldsymbol{v}}_1 \times \boldsymbol{B}_0 \ . \tag{2.48}$$

Furthermore, we build the cross product with the linearized electric potential equation (2.38) and obtain

$$\begin{aligned} \boldsymbol{B}_0 \times \tilde{\boldsymbol{E}}_1 &= -\boldsymbol{B}_0 \times \nabla\tilde{\phi}_1 - \frac{\partial}{\partial t}\left(\boldsymbol{B}_0 \times \tilde{\boldsymbol{A}}_1\right) \\ &\approx -\boldsymbol{B}_0 \times \nabla\tilde{\phi}_1 - \frac{\partial}{\partial t}\left(\boldsymbol{B}_0 \times \left(\hat{\boldsymbol{b}} \left(\tilde{A}_1\right)_{\parallel}\right)\right) \\ &= -\boldsymbol{B}_0 \times \nabla\tilde{\phi}_1 \end{aligned} \tag{2.49}$$

where we neglect $\left(\tilde{\boldsymbol{A}}_1\right)_{\perp}$ with (2.43). We combine (2.48) and (2.49) to

$$\boldsymbol{B}_0 \times \tilde{\boldsymbol{v}}_1 \times \boldsymbol{B}_0 = \boldsymbol{B}_0 \times \nabla\tilde{\phi}_1 \tag{2.50}$$

$$\hat{\boldsymbol{b}} \times \tilde{\boldsymbol{v}}_1 \times \hat{\boldsymbol{b}} = \frac{\hat{\boldsymbol{b}} \times \nabla\tilde{\phi}_1}{\|\boldsymbol{B}_0\|_2} \tag{2.51}$$

$$\hat{\boldsymbol{b}} \times \hat{\boldsymbol{b}} \times \tilde{\boldsymbol{v}}_1 \times \hat{\boldsymbol{b}} = \frac{\hat{\boldsymbol{b}} \times \hat{\boldsymbol{b}} \times \nabla\tilde{\phi}_1}{\|\boldsymbol{B}_0\|_2} \tag{2.52}$$

$$-\tilde{\boldsymbol{v}}_1 \times \hat{\boldsymbol{b}} = \frac{\hat{\boldsymbol{b}} \times \hat{\boldsymbol{b}} \times \nabla\tilde{\phi}_1}{\|\boldsymbol{B}_0\|_2} \tag{2.53}$$

where the last equation holds since $\hat{\boldsymbol{b}}$ is a unit vector. Using (2.17) and the linearized momentum equation (2.22) and forming the cross product with $\boldsymbol{B}_0$, we obtain

$$\rho_0 \frac{\partial}{\partial t} \left(\boldsymbol{B}_0 \times \tilde{\boldsymbol{v}}_1 \right) = \boldsymbol{B}_0 \times \tilde{\boldsymbol{J}}_1 \times \boldsymbol{B}_0 + \boldsymbol{B}_0 \times \boldsymbol{J}_0 \times \tilde{\boldsymbol{B}}_1 - \boldsymbol{B}_0 \times \nabla \tilde{p}_1 \tag{2.54}$$

which is rewritten using the force balance equation (2.11) to

$$\rho_0 \frac{\partial}{\partial t} \left(\boldsymbol{B}_0 \times \tilde{\boldsymbol{v}}_1 \right) = \boldsymbol{B}_0 \times \tilde{\boldsymbol{J}}_1 \times \boldsymbol{B}_0 - \nabla p_0 \times \tilde{\boldsymbol{B}}_1 - \boldsymbol{B}_0 \times \nabla \tilde{p}_1 \ . \tag{2.55}$$

Neglecting all pressure-related terms and division by $\| \boldsymbol{B}_0 \|_2^2$ yields

$$\frac{\rho_0}{\| \boldsymbol{B}_0 \|_2} \frac{\partial}{\partial t} \left(\hat{\boldsymbol{b}} \times \tilde{\boldsymbol{v}}_1 \right) = - \hat{\boldsymbol{b}} \times \hat{\boldsymbol{b}} \times \tilde{\boldsymbol{J}}_1 \ . \tag{2.56}$$

Insertion of (2.53) gives

$$\frac{\rho_0}{\| \boldsymbol{B}_0 \|_2^2} \frac{\partial}{\partial t} \left(\hat{\boldsymbol{b}} \times \hat{\boldsymbol{b}} \times \nabla \tilde{\phi}_1 \right) = - \hat{\boldsymbol{b}} \times \hat{\boldsymbol{b}} \times \tilde{\boldsymbol{J}}_1 \tag{2.57}$$

which is simplified using the definitions of the perpendicular component (2.36) and the perpendicular gradient (2.45) to

$$- \frac{\rho_0}{\| \boldsymbol{B}_0 \|_2^2} \frac{\partial}{\partial t} \left(\nabla_\perp \tilde{\phi}_1 \right) = \left(\tilde{\boldsymbol{J}}_1 \right)_\perp \ . \tag{2.58}$$

Using time-independence of the mass density ρ_0 and $\| \boldsymbol{B}_0 \|_2$ and insertion in (2.47) yields

$$\frac{\partial}{\partial t} \nabla \cdot \left(\frac{\rho_0 \mu_0}{\| \boldsymbol{B}_0 \|_2^2} \nabla_\perp \tilde{\phi}_1 \right) = \mu_0 \nabla \cdot \left(\hat{\boldsymbol{b}} \left(\tilde{J}_1 \right)_\parallel \right) \ . \tag{2.59}$$

Defining the Alfvén velocity [59, p.25]

$$v_\mathrm{A} := \frac{\| \boldsymbol{B}_0 \|_2}{\sqrt{\mu_0 \rho_0}} \ , \tag{2.60}$$

we summarize (2.41), (2.44) and (2.59) to the system of reduced MHD equations

$$\hat{\boldsymbol{b}} \cdot \nabla \tilde{\phi}_1 = - \frac{\partial \left(\tilde{A}_1 \right)_\parallel}{\partial t} \tag{2.61}$$

$$\mu_0 \left(\tilde{J}_1 \right)_\parallel = - \nabla_\perp^2 \left(\tilde{A}_1 \right)_\parallel \tag{2.62}$$

$$\frac{\partial}{\partial t} \nabla \cdot \left(\frac{1}{v_\mathrm{A}^2} \nabla_\perp \tilde{\phi}_1 \right) = \mu_0 \nabla \cdot \left(\hat{\boldsymbol{b}} \left(\tilde{J}_1 \right)_\parallel \right) \ . \tag{2.63}$$

for whose derivation we used drift approximation and neglected pressure-related terms as well as compressional Alfvén waves.

2.5 Derivation of a 4^{th}-order equation

This section transforms the reduced MHD equations (2.61) – (2.63) into an eigenvalue problem and reduces this system to a single equation. Therefrom, we build a first coarse model problem. For considering the stability of the MHD equilibrium, we introduce a time-periodic solution for the pertubations of the MHD equilibrium and rewrite time derivatives of (2.61) and (2.63) using [53, p.329, (8.2)]

$$\tilde{Q}_1(\boldsymbol{x}, t) = \boldsymbol{Q}_1(\boldsymbol{x}) \exp(-\mathrm{i}\omega t) \tag{2.64}$$

for a not necessarily vector-valued representative variable $\boldsymbol{Q}_1$. This limits the eigenfunction space to wave-type solutions. (2.61) – (2.63) then write

$$\hat{\boldsymbol{b}} \cdot \nabla \phi_1 = \mathrm{i}\omega (A_1)_{\parallel} \tag{2.65}$$

$$\mu_0 (J_1)_{\parallel} = -\nabla_{\perp}^2 (A_1)_{\parallel} \tag{2.66}$$

$$-\mathrm{i}\omega \nabla \cdot \left(\frac{1}{v_{\mathrm{A}}^2} \nabla_{\perp} \phi_1 \right) = \mu_0 \nabla \cdot \left(\hat{\boldsymbol{b}} (J_1)_{\parallel} \right) \tag{2.67}$$

as exponential contributions cancel. Note that the variables are now without a tilde as they are functions of $\boldsymbol{x}$ and not of $(\boldsymbol{x}, t)$. Successive insertion of (2.66) and (2.65) into (2.67) yields

$$\begin{aligned} -\mathrm{i}\omega \nabla \cdot \left(\frac{1}{v_{\mathrm{A}}^2} \nabla_{\perp} \phi_1 \right) &= -\nabla \cdot \left(\hat{\boldsymbol{b}} \nabla_{\perp}^2 (A_1)_{\parallel} \right) \\ &= -\nabla \cdot \left(\hat{\boldsymbol{b}} \nabla_{\perp}^2 \left(\frac{1}{\mathrm{i}\omega} \hat{\boldsymbol{b}} \cdot \nabla \phi_1 \right) \right) \end{aligned} \tag{2.68}$$

which we rearrange to

$$\omega^2 \nabla \cdot \left(\frac{1}{v_{\mathrm{A}}^2} \nabla_{\perp} \phi \right) = -\nabla \cdot \left(\hat{\boldsymbol{b}} \nabla_{\perp}^2 \left(\hat{\boldsymbol{b}} \cdot \nabla \phi \right) \right) \tag{2.69}$$

which is known as the eigenvalue problem associated to the reduced MHD shear Alfvén wave equation [57, p.43] and can be found in [60, p.3698, (27)]. We dropped the subscript 1. We deduce all model problems from (2.69).

We build a very coarse first model problem by considering (2.69) as a two-dimensional equation in a fully periodic domain with $v_{\mathrm{A}} \equiv 1$, $\hat{\boldsymbol{b}} = \boldsymbol{b} \equiv (b_1, b_2)^{\top}$. The reduction of dimensionality is motivated by the choice of $\hat{\boldsymbol{b}} = (b_1, b_2, 0)^{\top}$. For this problem, we rewrite the perpendicular gradient defined in (2.45) as

$$\nabla_{\perp} \coloneqq \boldsymbol{b}_{\perp} \boldsymbol{b}_{\perp} \cdot \nabla \tag{2.70}$$

with $\boldsymbol{b}_{\perp} \equiv (-b_2, b_1)^{\top}$. This redefinition removes the requirement of $\boldsymbol{b}$ being normed. Hence, we obtain a two-dimensional equation involving a 4^{th}-order differential operator given by

$$-\nabla \cdot (\boldsymbol{b} \nabla \cdot (\boldsymbol{b}_{\perp} \boldsymbol{b}_{\perp} \cdot \nabla (\boldsymbol{b} \cdot \nabla \phi))) = \omega^2 \nabla \cdot (\boldsymbol{b}_{\perp} \boldsymbol{b}_{\perp} \cdot \nabla \phi) \ , \qquad \text{in } \Omega \tag{2.71}$$

within the fully periodic domain $\Omega = [0, 2\pi)^2$. The mathematical treatment of this model problem is investigated in Chapter 4. Throughout this thesis, (2.71) is called the 4^{th}-order equation.

2.6 Derivation of the anisotropic wave equations

In this section, we derive another set of model problems. We start with the eigenvalue problem associated to the reduced MHD shear Alfvén equation (2.69) which is simplified further to a two-dimensional anisotropic wave equation. With all quantities being related to MHD equilibria, we aim at a change of the underlying coordinate system which represents the flux surfaces of the MHD equilibrium state introduced in Section 2.2 in an intuitive way. For this, we use straight field line coordinates, namely Boozer coordinates [54], and change the coordinate system from $\boldsymbol{x}$ to $(s, \theta, \varphi)^\top$ with the poloidal, toroidal angles θ, φ and the radial direction represented by the normalized flux surface coordinate s being perpendicular to a flux surface which is therefore labeled as the radial direction such that $\nabla s \cdot \hat{\boldsymbol{b}} = 0$. As resonance phenomena in plasma are related to the coefficient of the second radial derivative of the electric potential ϕ [61], we rewrite the reduced MHD shear Alfvén equation (2.69) and select the respective quantities. Important features of the spectrum are thereby unaffected.
First, the perpendicular gradient (2.45) can be decomposed as

$$\nabla_\perp \phi = \left(\frac{\nabla s}{\|\nabla s\|_2} \cdot \nabla \phi \right) \frac{\nabla s}{\|\nabla s\|_2} + \left(\frac{\hat{\boldsymbol{b}} \times \nabla s}{\|\nabla s\|_2} \cdot \nabla \phi \right) \frac{\hat{\boldsymbol{b}} \times \nabla s}{\|\nabla s\|_2} \tag{2.72}$$

as both ∇s and $\hat{\boldsymbol{b}} \times \nabla s$ are perpendicular to $\hat{\boldsymbol{b}}$ by definition of the coordinate system. Let Q be a scalar-valued function in the following. It holds

$$\nabla s \cdot \nabla Q = \|\nabla s\|_2^2 \frac{\partial Q}{\partial s} + \nabla s \cdot \nabla \theta \frac{\partial Q}{\partial \theta} + \nabla s \cdot \nabla \varphi \frac{\partial Q}{\partial \varphi} \tag{2.73}$$

which can be further simplified to

$$\nabla s \cdot \nabla Q = \|\nabla s\|_2^2 \frac{\partial Q}{\partial s} \tag{2.74}$$

as ∇s is defined to be perpendicular to flux surfaces and therefore to the directions of the flux surface coordinates θ and φ.
We then rewrite (2.72) as

$$\nabla_\perp \phi = \frac{\partial \phi}{\partial s} \nabla s + \boldsymbol{\phi}_{\neg \partial s} \tag{2.75}$$

where we collect all terms that do not involve an s-derivative of ϕ in

$$\boldsymbol{\phi}_{\neg \partial s} := \left(\nabla \phi \cdot \frac{\hat{\boldsymbol{b}} \times \nabla s}{\|\nabla s\|_2} \right) \frac{\hat{\boldsymbol{b}} \times \nabla s}{\|\nabla s\|_2} \tag{2.76}$$

We then obtain

$$\begin{aligned} \nabla \cdot (\nabla_\perp Q) &= \nabla \cdot \left(\frac{\partial Q}{\partial s} \nabla s + \boldsymbol{Q}_{\neg \partial s} \right) \\ &= (\nabla \cdot \nabla s) \frac{\partial Q}{\partial s} + \nabla \left(\frac{\partial Q}{\partial s} \right) \cdot \nabla s + \nabla \cdot \boldsymbol{Q}_{\neg \partial s} \\ &= \nabla^2 s \frac{\partial Q}{\partial s} + \frac{\partial^2 Q}{\partial s^2} \|\nabla s\|_2^2 + \nabla \cdot \boldsymbol{Q}_{\neg \partial s} \end{aligned} \tag{2.77}$$

where the third equation holds with (2.74). Applying this to the reduced MHD shear Alfvén equation (2.69), we obtain for the left hand side

$$\begin{aligned}\nabla \cdot \left(\frac{1}{v_{\mathrm{A}}^2} \nabla_\perp \phi \right) &= \left(\nabla \frac{1}{v_{\mathrm{A}}^2} \right) \cdot \nabla_\perp \phi + \frac{1}{v_{\mathrm{A}}^2} \nabla \cdot (\nabla_\perp \phi) \\ &= \left(\nabla \frac{1}{v_{\mathrm{A}}^2} \right) \cdot \nabla_\perp \phi + \frac{1}{v_{\mathrm{A}}^2} \left((\nabla^2 s) \frac{\partial \phi}{\partial s} + \frac{\partial^2 \phi}{\partial s^2} \left\| \nabla s \right\|_2^2 + \nabla \cdot \boldsymbol{\phi}_{\neg \partial s} \right)\end{aligned} \tag{2.78}$$

by momentarily omitting ω^2. The product rule of differentiation yields

$$\frac{\partial^2}{\partial s^2} \left(\hat{\boldsymbol{b}} \cdot \nabla \phi \right) = \hat{\boldsymbol{b}} \cdot \nabla \left(\frac{\partial^2 \phi}{\partial s^2} \right) + 2 \frac{\partial \hat{\boldsymbol{b}}}{\partial s} \cdot \nabla \left(\frac{\partial \phi}{\partial s} \right) + \frac{\partial^2 \hat{\boldsymbol{b}}}{\partial s^2} \cdot \nabla \phi \tag{2.79}$$

such that the right hand side of (2.69) writes

$$\begin{aligned}&- \nabla \cdot \left(\hat{\boldsymbol{b}} \nabla \cdot \left(\nabla_\perp \left(\hat{\boldsymbol{b}} \cdot \nabla \phi \right) \right) \right) = \\ &= -\nabla \cdot \left(\hat{\boldsymbol{b}} \left((\nabla^2 s) \frac{\partial}{\partial s} \left(\hat{\boldsymbol{b}} \cdot \nabla \phi \right) + \frac{\partial^2}{\partial s^2} \left(\hat{\boldsymbol{b}} \cdot \nabla \phi \right) \left\| \nabla s \right\|_2^2 + \nabla \cdot \left(\hat{\boldsymbol{b}} \cdot \nabla \boldsymbol{\phi} \right)_{\neg \partial s} \right) \right) \\ &= -\nabla \cdot \left(\hat{\boldsymbol{b}} \left((\nabla^2 s) \frac{\partial}{\partial s} \left(\hat{\boldsymbol{b}} \cdot \nabla \phi \right) + \nabla \cdot \left(\hat{\boldsymbol{b}} \cdot \nabla \boldsymbol{\phi} \right)_{\neg \partial s} \right) \right) \\ &\quad - \nabla \cdot \left(\hat{\boldsymbol{b}} \left(\hat{\boldsymbol{b}} \cdot \nabla \left(\frac{\partial^2 \phi}{\partial s^2} \right) + 2 \frac{\partial \hat{\boldsymbol{b}}}{\partial s} \cdot \nabla \left(\frac{\partial \phi}{\partial s} \right) + \frac{\partial^2 \hat{\boldsymbol{b}}}{\partial s^2} \cdot \nabla \phi \right) \left\| \nabla s \right\|_2^2 \right)\end{aligned} \tag{2.80}$$

using (2.77) and (2.79). We now gather the second radial derivatives of the electric potential ϕ in (2.78) and (2.80). Resolving the outermost divergence in (2.80) yields no further radial derivatives in the scalar-valued inner terms Q since

$$\nabla \cdot \left(\hat{\boldsymbol{b}} Q \right) = (\nabla \cdot \boldsymbol{B}_0) \frac{Q}{\left\| \boldsymbol{B}_0 \right\|_2} + \boldsymbol{B}_0 \cdot \nabla \left(\frac{Q}{\left\| \boldsymbol{B}_0 \right\|_2} \right) \tag{2.81}$$

where $\nabla \cdot \boldsymbol{B}_0 = 0$ due to Gauss's law of magnetism (2.7) and the second addend a parallel gradient which yields no radial derivatives. In summary, we reduce (2.69) to

$$\omega^2 \frac{\left\| \nabla s \right\|_2^2}{v_{\mathrm{A}}^2} \frac{\partial^2 \phi}{\partial s^2} = -\nabla \cdot \left(\hat{\boldsymbol{b}} \left\| \nabla s \right\|_2^2 \hat{\boldsymbol{b}} \cdot \nabla \left(\frac{\partial^2 \phi}{\partial s^2} \right) \right) \tag{2.82}$$

by just considering second radial derivatives of ϕ. We now substitute

$$\frac{\partial^2 \phi}{\partial s^2} \to \psi \tag{2.83}$$

and transform the differentials and vectors from the coordinate system $\boldsymbol{x}$ to $(s, \theta, \varphi)^\top$. An important property of the parallel gradient of a scalar-valued function Q is given by [62, p.138, (B.3.7)]

$$\boldsymbol{B}_0 \cdot \nabla_{\boldsymbol{x}} Q = -\frac{F_T'(s)}{\sqrt{g}} \left(\iota(s) \partial_\theta Q + \partial_\varphi Q \right) \tag{2.84}$$

for functions $F_T'(s) = \partial F_T(s)/\partial s$ and $\iota(s)$ which exclusively depend on the variable s. $\iota(s)$ is the rotational transform introduced in Section 2.2 and $F_T(s)$ is the toroidal flux. $\sqrt{g}$ is the Jacobian of the coordinate transform from $\boldsymbol{x}$ to $(s,\theta,\varphi)^\top$ given by [62, B.1.2. and p. 129]

$$\frac{1}{\sqrt{g}} = \nabla s \cdot (\nabla\theta \times \nabla\varphi) \ . \tag{2.85}$$

Note, that the radial derivative vanishes.
For the remaining part of this section, we identify the variables of differentiation on the respective operators. Using (2.36) for rewriting $\hat{\boldsymbol{b}}$ and Gauss's law for magnetism (2.33) on the right hand side of (2.82), we obtain

$$\begin{aligned} &-\nabla_{\boldsymbol{x}} \cdot \left(\hat{\boldsymbol{b}} \left\| \nabla s \right\|_2^2 \hat{\boldsymbol{b}} \cdot \nabla_{\boldsymbol{x}} \psi \right) = -\boldsymbol{B}_0 \cdot \nabla_{\boldsymbol{x}} \left(\frac{\left\| \nabla s \right\|_2^2}{\left\| \boldsymbol{B}_0 \right\|_2^2} \boldsymbol{B}_0 \cdot \nabla_{\boldsymbol{x}} \psi \right) \\ &= \frac{F_T'(s)}{\sqrt{g}} \begin{pmatrix} 0 \\ \iota(s) \\ 1 \end{pmatrix} \cdot \nabla_{s,\theta,\varphi} \left(-\frac{\left\| \nabla s \right\|_2^2 F_T'(s)}{\left\| \boldsymbol{B}_0 \right\|_2^2 \sqrt{g}} \begin{pmatrix} 0 \\ \iota(s) \\ 1 \end{pmatrix} \cdot \nabla_{s,\theta,\varphi} \psi \right) \end{aligned} \tag{2.86}$$

where we applied (2.84) twice for the second equation. $(0,\iota(s),1)^\top$ is represented in the $(s,\theta,\varphi)^\top$-coordinate system. We rearrange (2.86) by observing

$$\frac{1}{\sqrt{g}} \nabla_{s,\theta,\varphi} \cdot \left(F_T'(s) \begin{pmatrix} 0 \\ \iota(s) \\ 1 \end{pmatrix} \right) = 0 \tag{2.87}$$

as $F_T'(s)$ is constant in θ, φ. Then, we can rewrite (2.86) as

$$\begin{aligned} &\frac{F_T'(s)}{\sqrt{g}} \begin{pmatrix} 0 \\ \iota(s) \\ 1 \end{pmatrix} \cdot \nabla_{s,\theta,\varphi} \left(-\frac{\left\| \nabla s \right\|_2^2 F_T'(s)}{\left\| \boldsymbol{B}_0 \right\|_2^2 \sqrt{g}} \begin{pmatrix} 0 \\ \iota(s) \\ 1 \end{pmatrix} \cdot \nabla_{s,\theta,\varphi} \psi \right) \\ &= -\frac{1}{\sqrt{g}} \nabla_{s,\theta,\varphi} \cdot \left(\left(\begin{pmatrix} 0 \\ \iota(s) \\ 1 \end{pmatrix} \frac{\left\| \nabla s \right\|_2^2 \left(F_T'(s) \right)^2}{\left\| \boldsymbol{B}_0 \right\|_2^2 \sqrt{g}} \begin{pmatrix} 0 \\ \iota(s) \\ 1 \end{pmatrix} \cdot \nabla_{s,\theta,\varphi} \psi \right) \ . \end{aligned} \tag{2.88}$$

Noticing that s-derivatives vanish and using the definition of v_{A} in (2.60), we summarize the transformed version of (2.82) as

$$-\nabla_{\theta,\varphi} \cdot \left(\begin{pmatrix} \iota(s) \\ 1 \end{pmatrix} \frac{\left\| \nabla s \right\|_2^2 \left(F_T'(s) \right)^2}{\left\| \boldsymbol{B}_0 \right\|_2^2 \sqrt{g}} \begin{pmatrix} \iota(s) \\ 1 \end{pmatrix} \cdot \nabla_{\theta,\varphi} \psi \right) = \omega^2 \mu_0 \rho_0 \frac{\left\| \nabla s \right\|_2^2 \sqrt{g}}{\left\| \boldsymbol{B}_0 \right\|_2^2} \psi \ . \tag{2.89}$$

Assuming constant equilibrium density on each flux surface, we normalize the equation via $\mu_0\rho_0 \equiv 1$ and obtain

$$-\nabla_{\theta,\varphi} \cdot \left(\begin{pmatrix} \iota(s) \\ 1 \end{pmatrix} \frac{\left\| \nabla s \right\|_2^2 \left(F_T'(s) \right)^2}{\left\| \boldsymbol{B}_0 \right\|_2^2 \sqrt{g}} \begin{pmatrix} \iota(s) \\ 1 \end{pmatrix} \cdot \nabla_{\theta,\varphi} \psi \right) = \omega^2 \frac{\left\| \nabla s \right\|_2^2 \sqrt{g}}{\left\| \boldsymbol{B}_0 \right\|_2^2} \psi \ , \qquad \text{in } \Omega \tag{2.90}$$

with the fully periodic domain $\Omega = [0, 2\pi)^2$ which we call the anisotropic wave equation with metric terms. This is the second model problem. Neglecting the terms associated to the metric transformation and defining $\boldsymbol{b} \coloneqq (\iota(s), 1)^\top$ yields

$$-\nabla_{\theta,\varphi} \cdot \left(\boldsymbol{b}\boldsymbol{b} \cdot \nabla_{\theta,\varphi}\psi\right) = \omega^2 \psi \,, \qquad \text{in } \Omega \tag{2.91}$$

with the fully periodic domain $\Omega = [0, 2\pi)^2$ which we call the constant coefficient anisotropic wave equation. This is the third model problem. As the tensor $\boldsymbol{b}\boldsymbol{b}^\top$ is semidefinite, we tag these equations as anisotropic. Both anisotropic wave equations are analyzed in Chapter 3 with adaptions necessary for including all metric terms of (2.90) being discussed in Section 3.7.

2.7 Spectral properties and prospects

The spectrum of the reduced MHD shear Alfvén equation (2.69) and its related model problems are worthwhile to investigate. In general, the continuous spectrum "describes inherent plasma properties independent of external boundary conditions" [63, p.3207]. Discussions of these properties can be found in [64]. In axisymmetric toroidal geometries, plasma heating by resonance absorption with frequencies described by the spectrum can be constructed [60]. Furthermore, the spectrum is of importance for the current drive [65] and for the plasma stability in the presence of fast particles [66]. Stellarators also exhibit low frequency instabilities visible in the spectrum [67]. Therefore, we focus on the resolution of the spectrum in a neighbourhood of zero, i.e., we aim to resolve small eigenvalues of the model problems (2.71), (2.90) and (2.91).
We put emphasis on the extensibility of the method constructed in this thesis to more complex equations from which we deduced the model problems. For this, we particularly remark that all model problems (2.71), (2.90) and (2.91) are two-dimensional equations. For a more accurate reproduction of physical characteristics, e.g., consideration of the reduced MHD shear Alfvén equation (2.69) or the normal mode formulation of the linearized MHD stability problem for general three-dimensional equilibria (2.32), the extension to a three-dimensional discretization including the radial variable s is necessary. However, the radial discretization should be incorporated via a tensor-product approach using the two-dimensional discretization of the flux surface.

Chapter 3

ANISOTROPIC WAVE EQUATIONS Building methods (Part I)

In this chapter, we construct different numerical schemes for discretizing the anisotropic wave equations derived in Section 2.6. We therefore summarize (2.90) and (2.91) by momentarily omitting the right hand side metric terms of (2.90). The equation we consider throughout this chapter then writes

$$-\nabla \cdot (\boldsymbol{b}\boldsymbol{b} \cdot \nabla \phi) = \omega^2 \phi, \qquad \text{in } \Omega \tag{3.1}$$

with semidefinite tensor $\boldsymbol{b}\boldsymbol{b}^\top$ in the fully periodic domain $\Omega = [0, 2\pi)^2$. As a matter of notation, we use $\boldsymbol{x} = (x, y)^\top$-coordinates instead of $(\theta, \varphi)^\top$-coordinates.
Despite (3.1) being derived for straight field line coordinates, i.e.,

$$\boldsymbol{b}(\boldsymbol{x}) = \alpha(\boldsymbol{x}) \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \qquad b_1, b_2 \in \mathbb{R} \tag{3.2}$$

for a scalar-valued, non-zero, fully periodic function $\alpha(\boldsymbol{x})$ with the constant coefficient wave equation corresponding to $\alpha(\boldsymbol{x}) \equiv 1$, we execute important parts of the analysis of this chapter for general $\boldsymbol{b}(\boldsymbol{x})$ to explore the limitations of the herein constructed method. The basic principles of the construction of the method are however based on the consideration of the constant coefficient anisotropic wave equation, i.e., $\boldsymbol{b} = (b_1, b_2)^\top \equiv \text{const}$, and the straight field line approach. Whenever we assume constant $\boldsymbol{b}$ throughout a section, we put

$$\textit{Setting: } \boldsymbol{b} = (b_1, b_2)^\top \equiv \text{const}$$

at its beginning. If constant $\boldsymbol{b}$ is just necessary for intermediate results, we insert a remark.
The outline of this chapter is as follows: Section 3.1 discusses the exact solution of (3.1) for the case of constant $\boldsymbol{b}$. The choice of a suitable method is motivated in Section 3.2. Section 3.3 discusses the domain discretization for a discontinuous Galerkin method. General methodology of discontinuous Galerkin methods is introduced in Section 3.4. The system matrices of a primal variational form of

(3.1) are constructed in Section 3.5. In Section 3.6, we build different mixed variational forms and associated system matrices. Necessary adaptions for considering the anisotropic wave equation (2.90) on the flux surface of an MHD equilibrium are discussed in Section 3.7. As the assembly of system matrices can be drastically simplified for constant $\boldsymbol{b}$, Section 3.8 introduces Kronecker matrix products and illustrates the setup for the example of the primal variational form. This chapter closes with the presentation of asymptotic expansions of the discrete eigenvalues of the primal variational form for constant $\boldsymbol{b}$ and various underlying meshes in Section 3.9.

3.1 Analytical solution

$$\textit{Setting: } \boldsymbol{b} = (b_1, b_2)^\top \equiv \text{const}$$

To investigate properties of solutions of (3.1), we consider analytical eigenfunctions and their associated eigenvalues. Defining the norm

$$\|v\|^2_{L_2(\Omega)} \coloneqq \int_\Omega v\bar{v}\, \mathrm{d}\boldsymbol{x}\,, \tag{3.3}$$

we state the following theorem.

Theorem 3.1. *The functions*

$$\phi_{m,n}(x,y) = \exp\left(\mathrm{i}\left(mx + ny\right)\right), \qquad m,n \in \mathbb{Z},\ x,y \in \Omega \tag{3.4}$$

are analytic eigenfunctions of (3.1) *with corresponding eigenvalues* $\omega^2_{m,n} = (b_1 m + b_2 n)^2$*. The parallel gradient of each eigenfunction fulfills*

$$\|\boldsymbol{b}\cdot\nabla\phi_{m,n}(x,y)\|^2_{L^2_2(\Omega)} \leqslant 4\pi^2\omega^2_{m,n}\,. \tag{3.5}$$

Proof:
The gradient of $\phi_{m,n}$ is given by

$$\nabla\phi_{m,n}(x,y) = \begin{pmatrix} \mathrm{i}m\exp\left(\mathrm{i}\left(mx+ny\right)\right) \\ \mathrm{i}n\exp\left(\mathrm{i}\left(mx+ny\right)\right)\end{pmatrix} = \mathrm{i}\phi_{m,n}(x,y)\begin{pmatrix} m \\ n\end{pmatrix}\,. \tag{3.6}$$

Inserting the gradient in the left hand side of (3.1) yields

$$-\nabla\cdot(\boldsymbol{b}\boldsymbol{b}\cdot\nabla\phi_{m,n}(x,y)) = -\mathrm{i}^2(b_1m+b_2n)^2\phi_{m,n} = (b_1m+b_2n)^2\phi_{m,n}\,. \tag{3.7}$$

Using the representation of the gradient (3.6), we obtain

$$\begin{aligned}&\|\boldsymbol{b}\cdot\nabla\phi_{m,n}(x,y)\|^2_{L_2(\Omega)} = \|\phi_{m,n}(x,y)\,\mathrm{i}(b_1m+b_2n)\|^2_{L_2(\Omega)} = \\ &= \|\phi_{m,n}(x,y)\|^2_{L_2(\Omega)}\,|b_1m+b_2n|^2 \leqslant 4\pi^2\,|b_1m+b_2n|^2 = 4\pi^2\omega^2_{\mathrm{m,n}}\,.\end{aligned} \tag{3.8}$$

■

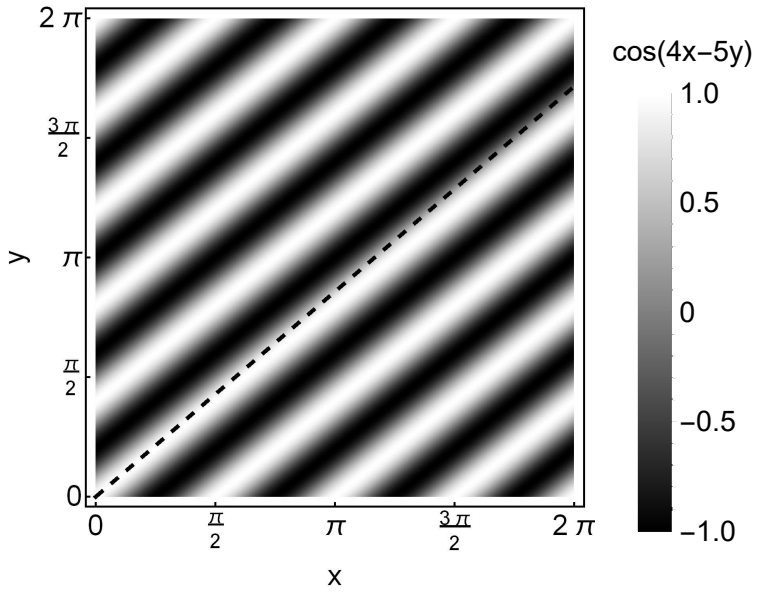


FIGURE 3.1: *Density plot of* $\cos(4x - 5y)$ *in* $\Omega = [0, 2\pi)^2$ *with associated eigenvalue* 0.11305798 *for* $\boldsymbol{b} = (1.1659398, 1)^\top$. *The black dashed line indicates the* $\boldsymbol{b}$*-direction.*

Theorem 3.1 shows that the norm of the parallel gradient depends on the size of the eigenvalue. For example, Figure 3.1 shows the density plot of the real part of $\phi_{4,-5}$. We observe that the function is almost constant in the given $\boldsymbol{b}$-direction which is plotted as the black dashed line. For this choice of $\boldsymbol{b}$, $\phi_{4,-5}$ is an eigenfunction with eigenvalue of order 10^{-1}.
Hence, when resolving eigenfunctions associated to small eigenvalues, less resolution in $\boldsymbol{b}$-direction than in $\boldsymbol{b}_\perp$-direction is needed as their variation along $\boldsymbol{b}$ is small. A method for solving (3.1) should be able to address the resolution of functions in $\boldsymbol{b}$- and $\boldsymbol{b}_\perp$-direction separately. This enables to resolve the highly oscillative behaviour along $\boldsymbol{b}_\perp$ accurately while providing a coarse discretization of a close to constant function along $\boldsymbol{b}$.
We call $\phi_{m,n}$ Fourier modes with frequencies or mode numbers (m, n) and remark that modes with frequency ratio m/n close to $-b_2/b_1$ are those producing small eigenvalues, as

$$\omega_{m,n}^2 = (b_1 m + b_2 n)^2 = n^2 (b_1 \frac{m}{n} + b_2)^2 \,. \tag{3.9}$$

When plotting eigenvalues over the domain of mode numbers, we observe that frequency pairs associated to small eigenvalues gather along $\boldsymbol{b}_\perp$ which is shown in Figure 3.2.

3.2 Choice of method

The discussion of separating degrees of freedom in a numerical method is an intricate question. Amongst others, [68, 69, 70, 71] study finite difference approaches for plasma turbulence. How-

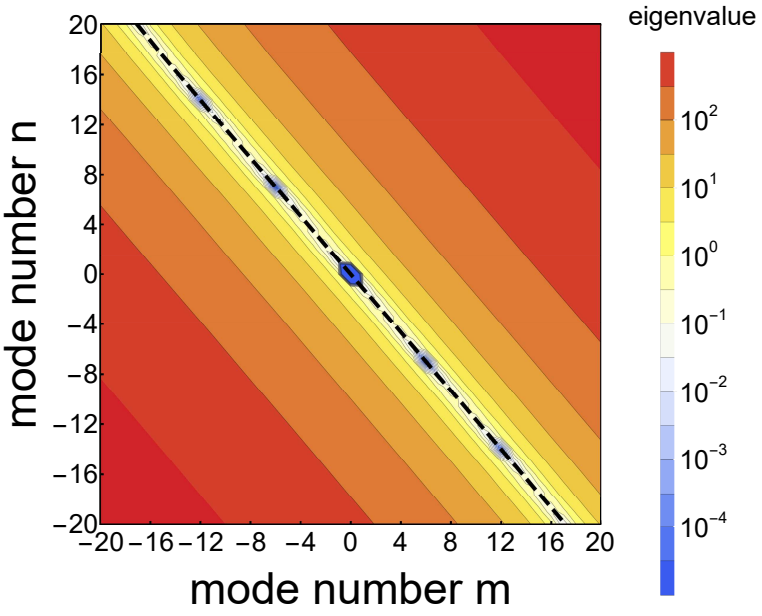


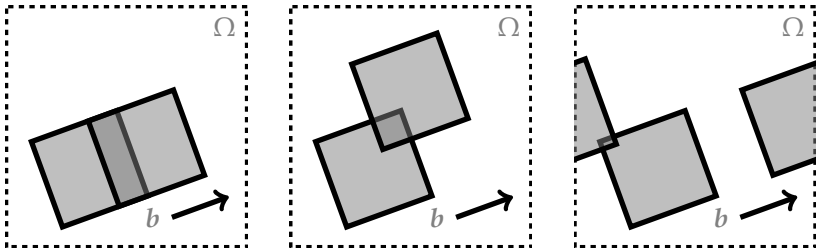
FIGURE 3.2: *Contour plot of the size of the exact eigenvalues for* $\boldsymbol{b} = (1.1659398, 1)^\top$ *with associated mode numbers up to* 20*. The black dashed line indicates the direction perpendicular to* $\boldsymbol{b}$*.*

ever, finite difference methods allow no variational formulations and solutions only can only be evaluated pointwise.

Considering the analysis of Section 3.1, Fourier methods seem to be the most appealing. For the given two-dimensional equation, Fourier methods indeed perform superior to all other kinds of methods. Considering the inclusion of metric terms as in (2.90) and the future applicability of the method to more sophisticated three-dimensional partial differential equations as outlined in Section 2.7, the global supports of the Fourier basis emerge to pose numerical difficulties as these lead to the treatment of huge dense matrices.

Finite elements can be chosen as tensor products of 1D finite elements aligned in $\boldsymbol{b}$- and $\boldsymbol{b}_\perp$-direction respectively. This however introduces difficulties when treating the periodic boundary as, dependent on $\boldsymbol{b}$, the supports of these finite elements may intersect differently than throughout the domain. When assembling a discretization matrix, finite elements with support affected by the periodic boundary generally have to be treated individually. This prevents the precalculation of all matrix entries using a single finite element as reference. Figure 3.3 shows a graphic representation of this argument which is discussed in detail in Section 3.3.2. Section 3.3 demonstrates that a suitable mesh has non-conforming interfaces. Even though mortar methods for finite elements exist [72], we choose to construct a discontinuous Galerkin method which naturally incorporates the treatment of non-conforming interfaces [73].

Furthermore, discontinuous Galerkin methods offer great freedom in the subdivision of the domain as well as the choice of the basis for each cell. The following sections therefore deal with the



(a) in domain (1) (b) in domain (2) (c) periodic boundary

FIGURE 3.3: *Intersection of supports of two finite elements in a periodic domain.*

construction of a discontinuous Galerkin method which is able to address the degrees of freedom for resolving $\boldsymbol{b}$- and $\boldsymbol{b}_\perp$-direction individually.

3.3 Choice of mesh

$$\textit{Setting: } \boldsymbol{b} = (b_1, b_2)^\top \equiv \text{const}$$

We state desirable properties of the subdivision of the fully periodic domain Ω.

1. Uniformity: The treatment of a single cell and its neighbours is exemplary for all other cells, i.e., the number and distribution of neighbours as well as the size of a cell and its interfaces is the same for all cells.

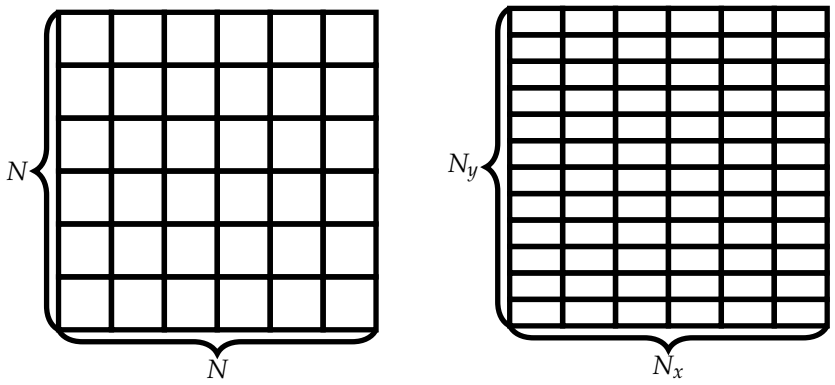
2. DoF-separation: The discretization assists in addressing the resolution along $\boldsymbol{b}$ and $\boldsymbol{b}_\perp$ individually, i.e., degrees of freedom for $\boldsymbol{b}$- and $\boldsymbol{b}_\perp$-direction can be assigned.

In the following, we evaluate different domain discretizations. We refer to interfaces as left, right, upper and bottom (instead of lower) to provide a framework for unique abbreviations (L,R,U,B) for objects in the subsequent parts. The number of cells in x,y-direction is defined by N_x,N_y. Then, the total number of cells is

$$N_\Sigma := N_x N_y \,. \tag{3.10}$$

3.3.1 Cartesian mesh

A straight forward discretization of the domain is given by a cartesian mesh as shown in Figure 3.4(a). It is conformal and fulfills the property of uniformity. In addition, no specific information about the differential equation (3.1) is used. Incorporating DoF-separation is not possible as increasing the resolution in x- or y-direction as shown in Figure 3.4(b) generally refines the resolution simultaneously in $\boldsymbol{b}$- and $\boldsymbol{b}_\perp$-direction.



(a) $N \times N$ square cells (b) $N_x \times N_y$ rectangular cells

FIGURE 3.4: *Cartesian meshes*

3.3.2 Fully aligned mesh

Considering the property of DoF-separation, the alignment of all cell interfaces with $\boldsymbol{b}$ and $\boldsymbol{b}_\perp$ allows to address the resolution along the $\boldsymbol{b}$- and $\boldsymbol{b}_\perp$-direction directly. Furthermore, aligned meshes are a desirable property in turbulence simulation of fusion plasmas [74]. Full alignment however encounters the same problems as a fully aligned finite element method as mentioned in Section 3.2. First, one set of upper and bottom or left and right interfaces has to be chosen conformal when using cells of equal size. A general sketch of such a fully aligned mesh is shown in Figure 3.5. Another restriction is given by the following theorem.

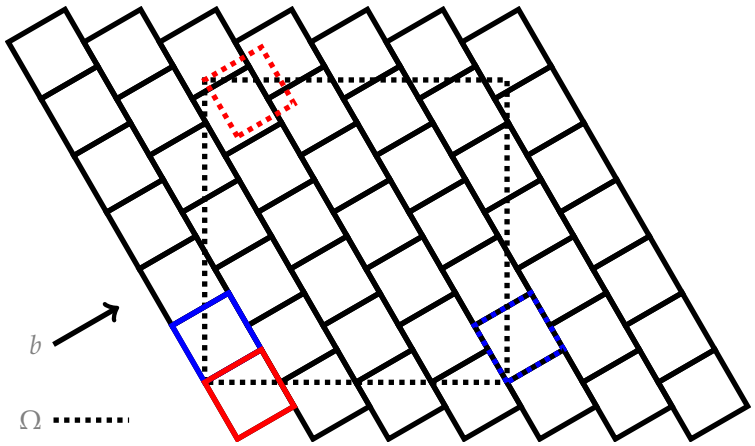


FIGURE 3.5: *Sketch of a fully aligned mesh. An example of a periodically shifted cell is shown with blue borders. Periodicity problems are highlighted in red.*

Theorem 3.2. *Let* $\boldsymbol{b} = (b_1, b_2)^\top$ *with* $b_1, b_2 > 0$*. If an aligned mesh with conforming* $\boldsymbol{b}$*-aligned and non-conforming* $\boldsymbol{b}_\perp$*-aligned interfaces can be constructed, then*

$$\exists k \in \mathbb{N}, \quad k = N_x \frac{b_2}{b_1} \,. \tag{3.11}$$

Proof:
Without loss of generality, we consider a representation of Ω such that the bottom left corner of a cell coincides with the bottom left corner of Ω. This is the blue bordered cell in Figure 3.6. Its bottom neighbour is depicted by a red boundary and their periodic continuations have dotted boundaries.

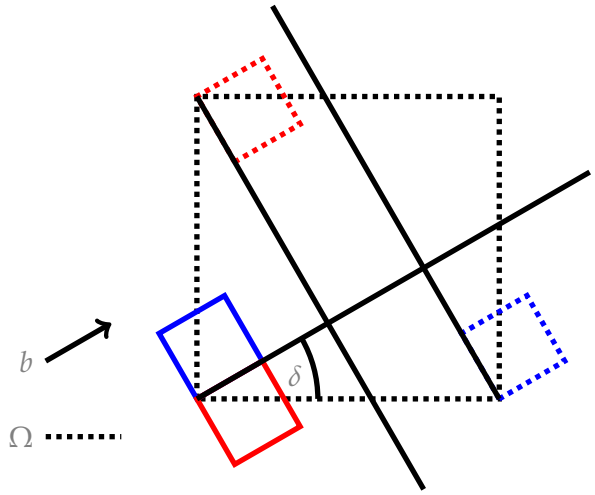


FIGURE 3.6: *Sketch for the proof of periodicity restrictions when aligning both interfaces of the cells.*

Considering the extended boundaries of the blue cells, we observe that the cell width is given by $\frac{2\pi}{N_x}\cos(\delta)$ with $\delta = \arctan(b_2/b_1)$ for some $N_x \in \mathbb{N}$. Considering the extended boundaries of the red cells, we observe that

$$\exists k \in \mathbb{N}, \quad 2\pi\cos\left(\frac{\pi}{2} - \delta\right) = 2\pi\sin(\delta) = k\frac{2\pi}{N_x}\cos(\delta) \tag{3.12}$$

which we simplify to

$$\exists k \in \mathbb{N}, \quad k = N_x \tan(\delta) = N_x \frac{b_2}{b_1} \tag{3.13}$$

using the definition of δ.

■

Figure 3.5 illustrates the problem when choosing the cell width such that the left and right periodicity constraints are fulfilled and (3.11) is not fulfilled. Hence, the upper and lower periodicity constraints cannot be met as shown by the red and red-dotted cell.

When choosing conforming $\boldsymbol{b}$-aligned interfaces, the statement and proof of Theorem 3.2 can be readily adapted by performing the same arguments using Figure 3.6 rotated by $\pi/2$. Condition (3.11) is generally not met as it may only apply for rational b_2/b_1 which is discussed in more detail later this section.
Hence, we consider the alignment of either the upper and bottom or left and right interfaces with $\boldsymbol{b}$ as shown in Figure 3.7. Here, N_x addresses the resolution in $\boldsymbol{b}$-direction exclusively whereas N_y addresses both $\boldsymbol{b}$- and $\boldsymbol{b}_\perp$-direction. To avoid strongly sheared cells, aligning left and right interfaces is preferable whenever $\left|b_2 N_y / \left(b_1 N_x\right)\right| > 1$ whereas the alignment of upper and bottom interfaces should be done for $\left|b_2 N_y / \left(b_1 N_x\right)\right| \leqslant 1$. A more detailed discussion of this can be found in Section 3.3.3.

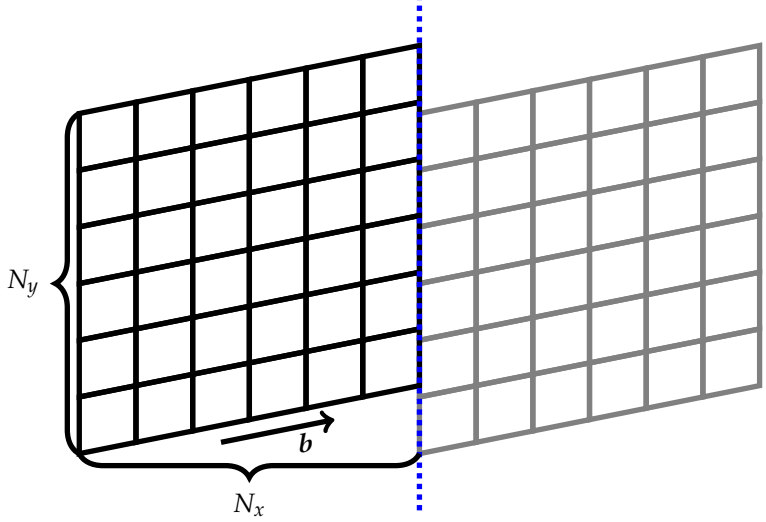


FIGURE 3.7: *Fully aligned mesh with aligned upper and bottom interfaces (black) and its periodic continuation (gray). The blue dashed line depicts the right hand side periodic boundary.*

Considering the property of uniformity, we observe that the interfaces of cells within the domain are conforming whereas there are non-conforming interfaces at the periodic boundary as shown in Figure 3.7 whenever

$$\frac{b_2 N_y}{b_1} \notin \mathbb{Z} \tag{3.14}$$

which holds for all irrational $b_2/b_1 \in \mathbb{R} \setminus \mathbb{Q}$. In the case of rational b_2/b_1 we want to avoid restricting the mesh resolution to this quantity as we would be forced to use a fixed cell number ratio. Hence, uniformity in general doesn't hold for the fully aligned mesh. However, this mesh provides a uniform treatment of all cells with no interfaces at the periodic boundary.

3.3.3 Locally aligned mesh

Section 3.3.1 shows a mesh fulfilling uniformity whereas Section 3.3.2 shows a mesh providing the property of DoF-separation. Combining the ideas of these sections, we construct a mesh with locally aligned interfaces as shown in Figure 3.8. The periodic domain is subdivided into $N_x \times N_y$ rectangular regions as in Figure 3.4(b) and the left corner points of each cell are kept fixed. The bottom and upper interfaces are then aligned with $\boldsymbol{b}$. As the direction of $\boldsymbol{b}$ is constant throughout Ω, this construction is the same for every cell. Hence, the resulting mesh fulfills the desired property of uniformity as each cell has the same neighbouring structure which is depicted in Figure 3.9(a). The left and right interfaces are non-conforming whenever

$$\frac{b_2 N_y}{b_1 N_x} \notin \mathbb{Z} \,. \tag{3.15}$$

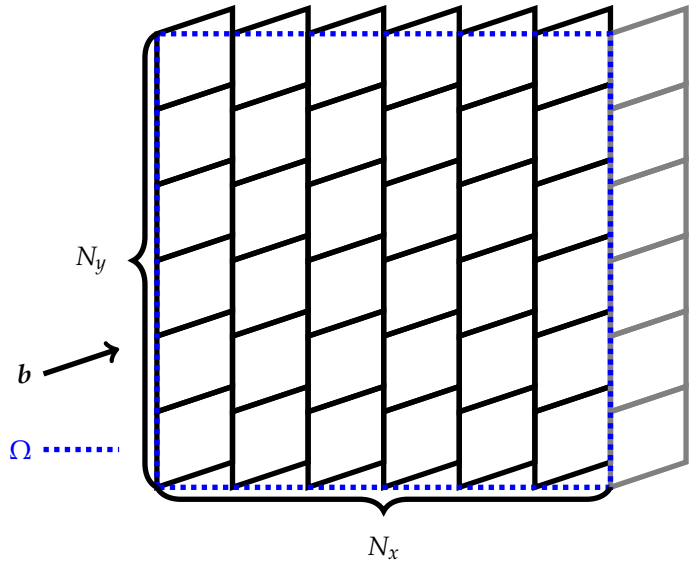
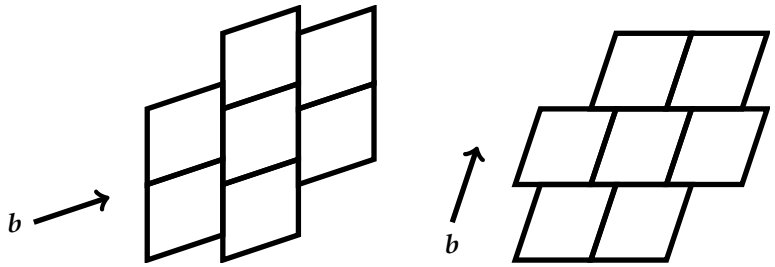


FIGURE 3.8: *Locally aligned $N_x \times N_y$ mesh in the periodic domain Ω (blue dashed) with right hand side periodic continuation (gray).*

In comparison to the fully aligned mesh of Figure 3.7 we obtain the locally aligned mesh by displacement of columns of cells in y-direction. For investigating the property of DoF-separation we consider an arbitrary point in Ω. Translating this point in $\boldsymbol{b}$-direction across the periodic boundaries, we exactly traverse N_x cells before reaching a point with the same x-coordinate. When the point is translated onto an aligned interface we still just count this as one cell instead of two. Hence, N_x addresses the resolution in $\boldsymbol{b}$-direction. As in the fully aligned case, N_y addresses the resolution in both $\boldsymbol{b}$- and $\boldsymbol{b}_\perp$-direction.
For $\left|b_1 N_x / \left(b_2 N_y\right)\right| \ll 1$ cells are strongly sheared as depicted in Figure 3.10 which can result in



(a) Aligned upper and bottom interfaces (b) Aligned left and right interfaces

FIGURE 3.9: *Neighbourhood of a single cell in a locally aligned mesh.*

numerically difficulties as the Jacobians of the element transformation tend to zero in the limit. To avoid this, left and right interfaces can be aligned instead of the upper and bottom interfaces. Starting with a rectangular mesh with $N_x \times N_y$ cells, we then fix the bottom corner points and align the left and right interface with $\boldsymbol{b}$. The upper and bottom interfaces are then non-conforming whenever

$$\frac{b_1 N_x}{b_2 N_y} \notin \mathbb{Z} \,. \tag{3.16}$$

The resulting neighbouring structure is depicted by Figure 3.9(b).

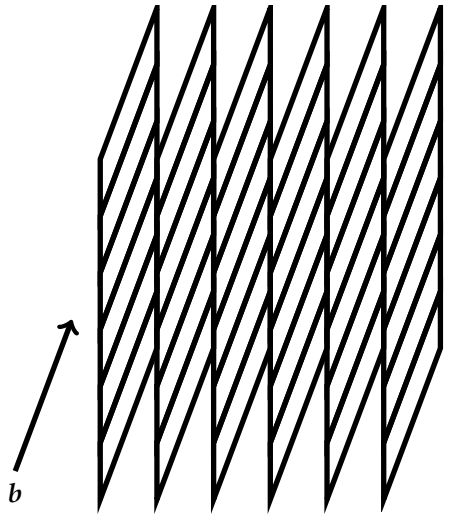


FIGURE 3.10: *Locally aligned mesh with strongly sheared cells.*

In summary, the proposed locally aligned mesh is generally non-conforming and provides both desired properties of uniformity and DoF-separation.
We remark that the construction process works in the same manner for $\boldsymbol{b}(\boldsymbol{x})$ with constant direction

and varying length, i.e.,

$$\boldsymbol{b}(\boldsymbol{x}) = \alpha(\boldsymbol{x}) \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \tag{3.17}$$

for a scalar-valued, non-zero, fully periodic function $\alpha(\boldsymbol{x})$. The mesh can therefore be used for treating straight field line coordinates. The properties of uniformity and DoF-separation are still fulfilled.
For general $\boldsymbol{b}(\boldsymbol{x})$ the local alignment could be generalized as shown in Figure 3.11. Depending on the vector field, either left and right or upper and bottom alignment has to be chosen, for example for minimizing the maximal shear. Uniformity cannot be achieved for general vector fields especially as the amount of neighbours per cell might vary.

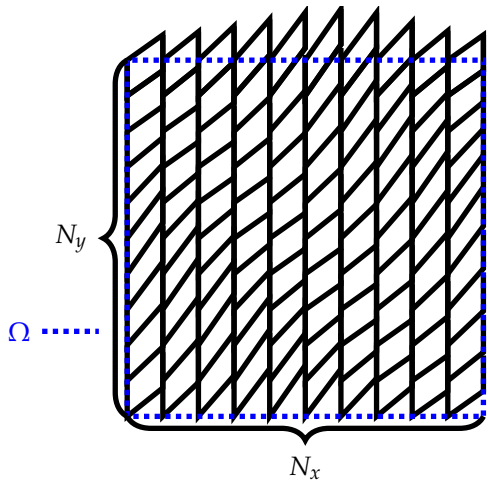


FIGURE 3.11: *Locally aligned $N_x \times N_y$ mesh in the periodic domain Ω (blue dashed) for the vector field $\boldsymbol{b}(x,y) = (1+\cos(x+y)/3, 1+\sin(x+y)/3)^\top$.*

3.4 General remarks on discontinuous Galerkin

The analysis of this section is strongly tied to [40, Section 4.5] where a discontinuous Galerkin method for a model problem with heterogeneous diffusion is constructed. The therein considered equation is given by

$$-\nabla \cdot (\kappa \nabla u) = f\,, \quad \text{in } \Omega\,, \quad u = 0\,, \quad \text{on } \partial\Omega\,. \tag{3.18}$$

The constant coefficient anisotropic wave equation is structurally equivalent to (3.18). In comparison, we consider (3.18) as an eigenvalue problem with periodic boundary conditions. Furthermore, we allow for positive semidefinite tensors $\kappa = \boldsymbol{b}\boldsymbol{b}^\top$. [40, Section 4.5] only discusses positive definite tensors.
We rewrite the anisotropic wave equation (3.1) in variational form using a space of functions V

such that the eigenvalue problem reads: Find pairs $(\phi, \omega^2) \in V \times \mathbb{R}$ such that

$$\int_\Omega -\nabla \cdot (\boldsymbol{b}\boldsymbol{b} \cdot \nabla\phi)\, \psi \, \mathrm{d}\boldsymbol{x} = \omega^2 \int_\Omega \phi\psi \, \mathrm{d}\boldsymbol{x} \qquad \forall \psi \in V \tag{3.19}$$

which can be written more general as

$$a(\phi, \psi) = \omega^2 \langle \phi_h, \psi_h \rangle \qquad \forall \psi \in V \tag{3.20}$$

for a bilinear form a and

$$\langle \phi_h, \psi_h \rangle = \int_\Omega \phi\psi \, \mathrm{d}\boldsymbol{x} \,. \tag{3.21}$$

Discretization of this weak form, i.e., imposing a finite dimensional function space V_h, allows the reformulation of the problem as: Find pairs $(\phi_h, \omega^2) \in V_h \times \mathbb{R}$ such that

$$a_h(\phi_h, \psi_h) = \omega^2 \langle \phi_h, \psi_h \rangle \qquad \forall \psi_h \in V_h \tag{3.22}$$

for a discrete bilinear form a_h defined on $V_h \times V_h$. Following [40, Sections 1.3, 4.2.1] this bilinear form should be consistent, continuous, coercive and symmetric.

Definition 3.3. *A discrete bilinear form $a_h : V_h \times V_h \to \mathbb{R}$ is called*

- ***consistent*** *if for an exact solution (ϕ, ω^2) of (3.1) it holds $a_h(\phi, \psi_h) = \omega^2 \langle \phi, \psi_h \rangle$, $\forall \psi_h \in V_h$. This strong form of consistency implies that a_h can be extended to $V_* \times V_h$ where $V_h \subset V_* \subset V$ as remarked in [40, Section 1.3.3].*
- ***continuous*** *if $\forall \phi \in V_*, \psi_h \in V_h$ it holds $a_h(\phi, \psi_h) \leqslant C \|\phi\|_* \|\psi_h\|_h$ for some $C > 0$ independent of h and norms $\|.\|_*$ and $\|.\|_h$ on the spaces V_* and V_h respectively fulfilling $\|\psi_h\|_h \leqslant \|\psi_h\|_*$, $\forall \psi_h \in V_h$, see [40, Section 1.3.4].*
- ***coercive*** *if $\forall \psi_h \in V_h$ it holds $a_h(\psi_h, \psi_h) \geqslant \varepsilon \|\psi_h\|_h^2$ for some $\varepsilon > 0$ independent of h [40, Definition 1.3].*
- ***symmetric*** *if $\forall \phi_h, \psi_h \in V_h$ it holds $a_h(\phi_h, \psi_h) = a_h(\psi_h, \phi_h)$.*

These properties ensure the well-posedness and consequently the uniqueness of the discrete solution of (3.22) using the Lax-Milgram theorem and allow the construction of a suitable error analysis.
However, the eigenvalue problem (3.1) is not well-posed.

Lemma 3.4. *The eigenvalue problem (3.1) is ill-posed.*

Proof:
The function $\phi \equiv c$ for $c \in \mathbb{R} \setminus \{0\}$ is an eigenfunction with eigenvalue 0. Let (ψ, ω^2) be an eigenpair with eigenvalue $\omega^2 \neq 0$. Then $\psi + \alpha\phi$ is an eigenfunction with eigenvalue ω^2 for all $\alpha \in \mathbb{R}$ since the differential operator of (3.1) is linear.

■

Corollary 3.5. *Let $\boldsymbol{b} = (b_1, b_2)^\top$ be constant with $b_1/b_2 \in \mathbb{Q}$ and (ψ, ω^2) an eigenpair with $\omega^2 \neq 0$. Then there exists an infinite-dimensional space V_0 of eigenfunctions of (3.1) with eigenvalue 0 and $\psi + \phi$ is an eigenfunction with eigenvalue ω^2 for all $\phi \in V_0$.*

Proof:
Theorem 3.1 shows that $\phi_{m,n}(x, y) = \exp(\mathrm{i}(mx + ny))$ are eigenfunctions of (3.1) with eigenvalue $(b_1 m + b_2 n)^2$. Then

$$(b_1 m + b_2 n)^2 = b_2^2 \left(\frac{b_1}{b_2} m + n \right)^2 . \tag{3.23}$$

As we assumed that $b_1/b_2 \in \mathbb{Q}$, there exists $\tilde{m} \in \mathbb{N} \setminus \{0\}$ such that $-\tilde{n} := b_1/b_2 \tilde{m} \in \mathbb{Z}$. Then the function $\phi_{\tilde{m},\tilde{n}}$ is an eigenfunction with eigenvalue 0 which is non-constant as $\tilde{m} \neq 0$. Furthermore the function $\phi_{k\tilde{m},k\tilde{n}}$ has eigenvalue 0 for all $k \in \mathbb{N}$ as

$$b_2^2 \left(k \frac{b_1}{b_2} \tilde{m} + k\tilde{n} \right)^2 = b_2^2 k^2 \left(\frac{b_1}{b_2} \tilde{m} + \tilde{n} \right)^2 = 0 \tag{3.24}$$

such that V_0 has infinite dimensions. As the differential operator of (3.1) is linear, $\psi + \phi$ has the eigenvalue $\omega^2 + 0 = \omega^2$ for all $\phi \in V_0$.

■

Lemma 3.6. *Any consistent bilinear form a_h discretizing (3.1) with $c \in V_h, \forall c \in \mathbb{R}$ cannot be coercive.*

Proof:
Let $c \in \mathbb{R} \setminus \{0\}$ and $\phi \equiv c \in V_h$. ϕ is an eigenfunction of (3.1) with eigenvalue 0. As a_h is consistent, $a_h(\phi, \phi) = 0$. Then we cannot find an $\varepsilon > 0$ such that $0 \geqslant \varepsilon \|\phi\|_h^2 > 0$ since $\|\phi\|_h^2 > 0$ with $\phi \neq 0$.

■

Fixing functions up to a constant is possible, for example by setting the integral mean of solutions over Ω to a fixed constant. However, the eigenspace for the eigenvalue 0 can be infinite-dimensional as shown by Corollary 3.5 and adding some kind of special treatment for its handling seems inevitable. Though these cases can generally be neglected using the following argument. For constant $\boldsymbol{b} = (b_1, b_2)^\top$ with $b_1/b_2 \in \mathbb{R} \setminus \mathbb{Q}$ the eigenvalues for non-constant Fourier modes are arbitrary close to 0, but the eigenspace for the eigenvalue 0 just consists of the constant mode. Additionally, due to the Nyquist-Shannon sampling theorem [75], imposing DoF_x and DoF_y degrees of freedom for discretization of the x- and y-direction, Fourier modes with mode numbers $m_{\max} = 2\,\mathrm{DoF}_x$ and $n_{\max} = 2\,\mathrm{DoF}_y$ can be resolved. If $b_1/b_2 \in \mathbb{Q}$ is an irreducible fraction and $|b_1| < 2\,\mathrm{DoF}_y$, $|b_2| < 2\,\mathrm{DoF}_x$, the eigenspace for 0 provided by a suitable method should consist of more functions than just the constant mode. We denote the set of these $\boldsymbol{b}$ as $B_{\mathrm{NS}} \subset \mathbb{Q}^2$. Given a randomly selected $\boldsymbol{b} \in \mathbb{R}^2$, we remark that $\mu_{\mathbb{R}^2}(\mathbb{R}^2 \setminus \mathbb{Q}^2) = 1$. Furthermore $\mu_{\mathbb{Q}^2}(\mathbb{Q}^2 \setminus B_{\mathrm{NS}}) = 1$. Hence, the cases where a method deals with an eigenspace for 0 of a size larger than 1 are generally negligible.

Considering Lemma 3.6, we focus on the construction of weak forms fulfilling consistency, continuity and symmetry.
In the following sections, we drop the h-indices when we address discrete bilinear forms to improve readability.
For describing discontinuous Galerkin methods, we often need to refer to objects defined on a cell and its neighbours.

Definition 3.7. *Let the following objects be defined as*

- *$\mathcal{F}$ is the set of all interfaces and $F \in \mathcal{F}$ an interface which is shared by two distinct neighbouring cells.*
- *$\mathcal{K}$ is the set of cells and $K \in \mathcal{K}$ a cell. $N_F(K) \in \mathcal{K}$ denotes the neighbour of cell K sharing the interface F. ∂K the set of all interfaces of K.*
- *$\boldsymbol{n}^K$, $\boldsymbol{n}^{N_F(K)}$ is the unit outer normal of an interface $F \in \mathcal{F}$ viewed from the main cell K, neighbouring cell $N_F(K)$. It holds $\boldsymbol{n}^K = -\boldsymbol{n}^{N_F(K)}$.*
- *Average $\{\!\{.\}\!\}$ and jump $[\![.]\!]$ on an interface F are given by*

$$\{\!\{f\}\!\} = \frac{1}{2}\left(f^K + f^{N_F(K)}\right) \qquad , \qquad [\![f]\!] = f^K \boldsymbol{n}^K + f^{N_F(K)} \boldsymbol{n}^{N_F(K)} \tag{3.25}$$

where f^K, $f^{N_F(K)}$ are defined as the limit on F from K, $N_F(K)$ respectively. For all $x \in F$ it holds

$$f^K(x) \coloneqq \lim_{\substack{x_k \to x \\ x_k \in K}} f(x_k) \qquad , \qquad f^{N_F(K)}(x) \coloneqq \lim_{\substack{x_k \to x \\ x_k \in N_F(K)}} f(x_k)\,. \tag{3.26}$$

We note that average and jump are well-defined as they are identical when viewed from each of the neighbouring cells of the interface.

3.5 Construction of a primal form matrix system

For the construction of a suited discontinuous Galerkin method, we consider the variational form of (3.1) when integrating on cells K of a mesh. This section deals with the discretization of a single integral equation. The problem reads: Find pairs $(\phi, \omega^2) \in V_\Phi \times \mathbb{R}$ such that

$$\sum_{K \in \mathcal{K}} \left(\int_K -\nabla \cdot (\boldsymbol{b}\boldsymbol{b} \cdot \nabla \phi)\, \psi \, \mathrm{d}\boldsymbol{x} \right) = \omega^2 \sum_{K \in \mathcal{K}} \left(\int_K \phi \psi \, \mathrm{d}\boldsymbol{x} \right) \quad \forall \psi \in V_\Phi\,. \tag{3.27}$$

Functions ϕ are called trial functions and ψ are called test functions. In general, the spaces for test and trial functions can be chosen independently. Throughout this thesis, we focus on the case where these spaces coincide but mention whenever this assumption is essential for a result. The

local space of test and trial functions $V_{K,\Phi}$ consists of functions with support exclusively in K. The space of test and trial functions for Ω is given by

$$V_\Phi = \bigcup_{K\in\mathcal{K}} V_{K,\Phi} \,. \tag{3.28}$$

Having defined this space, we omit its specification in reformulations of the variational form (3.27) to improve readability. Furthermore, we consider (3.27) for a single cell K first, which gives $\forall \psi^K \in V_{K,\Phi}, \forall K \in \mathcal{K}$

$$\int_K -\nabla\cdot\left(\boldsymbol{b}\boldsymbol{b}\cdot\nabla\phi^K\right)\psi^K \,\mathrm{d}\boldsymbol{x} = \omega^2\int_K \phi^K\psi^K \,\mathrm{d}\boldsymbol{x} \tag{3.29}$$

where we later omit the superscript of functions $\phi^K, \psi^K \in V_{K,\Phi}$ whenever there is no need for differentiating the cells of definition. (3.29) is obtained by choosing $\psi^K \in V_{K,\Phi}$ as test function in (3.27).

Section 3.5.1 constructs the primal variational form using symmetric interior penalty fluxes. Section 3.5.2 sets up the corresponding system matrices.

3.5.1 Construction of a primal form

Integration by parts allows to rewrite (3.27) as

$$\int_K \boldsymbol{b}\cdot\nabla\phi\boldsymbol{b}\cdot\nabla\psi \,\mathrm{d}\boldsymbol{x} - \sum_{F\in\partial K}\left(\int_F \widehat{\boldsymbol{b}\cdot\nabla\phi}\psi\boldsymbol{b}\cdot\boldsymbol{n}\,\mathrm{d}S(\boldsymbol{x})\right) = \omega^2\int_K \phi\psi \,\mathrm{d}\boldsymbol{x} \tag{3.30}$$

where $\boldsymbol{n} = \boldsymbol{n}^K$. As the solution is double-valued at the cell interfaces, we introduce a yet to be defined numerical flux $\widehat{\boldsymbol{b}\cdot\nabla\phi}$. Another integration by parts of (3.30) allows the reformulation to

$$\begin{aligned} &-\int_K \phi\nabla\cdot\left(\boldsymbol{b}\boldsymbol{b}\cdot\nabla\psi\right)\,\mathrm{d}\boldsymbol{x} - \sum_{F\in\partial K}\left(\int_F \widehat{\boldsymbol{b}\cdot\nabla\phi}\psi\boldsymbol{b}\cdot\boldsymbol{n}\,\mathrm{d}S(\boldsymbol{x})\right) \\ &\quad + \sum_{F\in\partial K}\left(\int_F \widehat{\phi}\boldsymbol{b}\cdot\nabla\psi\boldsymbol{b}\cdot\boldsymbol{n}\,\mathrm{d}S(\boldsymbol{x})\right) = \omega^2\int_K \phi\psi \,\mathrm{d}\boldsymbol{x} \end{aligned} \tag{3.31}$$

with a numerical flux $\widehat{\phi}$ that still needs to be defined. To resolve the second order differential, we again integrate by parts by evaluating the inner cell function ϕ on the boundary such that

$$\begin{aligned} &\int_K \boldsymbol{b}\cdot\nabla\phi\boldsymbol{b}\cdot\nabla\psi \,\mathrm{d}\boldsymbol{x} - \sum_{F\in\partial K}\left(\int_F \widehat{\boldsymbol{b}\cdot\nabla\phi}\psi\boldsymbol{b}\cdot\boldsymbol{n}\,\mathrm{d}S(\boldsymbol{x})\right) \\ &\quad + \sum_{F\in\partial K}\left(\int_F \left(\widehat{\phi}-\phi\right)\boldsymbol{b}\cdot\nabla\psi\boldsymbol{b}\cdot\boldsymbol{n}\,\mathrm{d}S(\boldsymbol{x})\right) = \omega^2\int_K \phi\psi \,\mathrm{d}\boldsymbol{x}\,. \end{aligned} \tag{3.32}$$

Following [40, Section 4.5], we define the fluxes as

$$\widehat{\boldsymbol{b}\cdot\nabla\phi} = \{\!\{\boldsymbol{b}\cdot\nabla\phi\}\!\} - \frac{\eta}{h_F}\boldsymbol{b}\cdot[\![\phi]\!]\,, \qquad \widehat{\phi} = \{\!\{\phi\}\!\} \tag{3.33}$$

to obtain the symmetric interior penalty (SIP) form with jumps and averages defined in (3.25). Here, η is a stabilization parameter larger than zero and h_F is a parameter dependent on the length of the cell edge with interface F. The larger the stabilization parameter the stronger discontinuities of the solution are penalized. We rewrite (3.32) using (3.33) $\forall K \in \mathcal{K}, \psi_K \in V_{K,\Phi}$

$$\begin{aligned}
&\int_K \boldsymbol{b}\cdot\nabla\phi^K\boldsymbol{b}\cdot\nabla\psi^K \,\mathrm{d}\boldsymbol{x} - \sum_{F\in\partial K}\left(\int_F \frac{1}{2}\left(\boldsymbol{b}\cdot\nabla\phi^K + \boldsymbol{b}\cdot\nabla\phi^{N_F(K)}\right)\psi^K\boldsymbol{b}\cdot\boldsymbol{n}^K \,\mathrm{d}S(\boldsymbol{x})\right)\\
&- \sum_{F\in\partial K}\left(\int_F \frac{1}{2}\left(\phi^K - \phi^{N_F(K)}\right)\boldsymbol{b}\cdot\nabla\psi^K\boldsymbol{b}\cdot\boldsymbol{n}^K \,\mathrm{d}S(\boldsymbol{x})\right)\\
&+ \sum_{F\in\partial K}\left(\frac{\eta}{h_F}\int_F \boldsymbol{b}\cdot\left(\phi^K\boldsymbol{n}^K + \phi^{N_F(K)}\boldsymbol{n}^{N_F(K)}\right)\psi^K\boldsymbol{b}\cdot\boldsymbol{n}^K \,\mathrm{d}S(\boldsymbol{x})\right) = \omega^2\int_K \phi^K\psi^K \,\mathrm{d}\boldsymbol{x}\,.
\end{aligned} \tag{3.34}$$

Building the sum over all elements, we impose the following lemmata for simplifying the surface terms.

Lemma 3.8. *Let $\widehat{\phi}$ be an arbitrary flux, $\psi^K \in V_{K,\Phi}$ locally defined in K and $\boldsymbol{v} \in \mathbb{C}^2$ a not necessarily constant vector. Then*

$$\sum_{K\in\mathcal{K}}\sum_{F\in\partial K}\left(\int_F \widehat{\phi}\psi^K\boldsymbol{v}\cdot\boldsymbol{n}^K \,\mathrm{d}S(\boldsymbol{x})\right) = \sum_{F\in\mathcal{F}}\left(\int_F \widehat{\phi}\boldsymbol{v}\cdot[\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x})\right)\,. \tag{3.35}$$

Proof:
When summing over all cells, each interface F is considered twice in total, once for a cell K and once for its unique neighbour $N_F(K)$. Hence, we obtain

$$\begin{aligned}
&\sum_{K\in\mathcal{K}}\sum_{F\in\partial K}\left(\int_F \widehat{\phi}\psi^K\boldsymbol{v}\cdot\boldsymbol{n}^K \,\mathrm{d}S(\boldsymbol{x})\right)\\
&= \sum_{F\in\mathcal{F}}\left(\int_F \widehat{\phi}\psi^K\boldsymbol{v}\cdot\boldsymbol{n}^K \,\mathrm{d}S(\boldsymbol{x}) + \int_F \widehat{\phi}\psi^{N_F(K)}\boldsymbol{v}\cdot\boldsymbol{n}^{N_F(K)} \,\mathrm{d}S(\boldsymbol{x})\right)\\
&= \sum_{F\in\mathcal{F}}\left(\int_F \widehat{\phi}\left(\psi^K\boldsymbol{v}\cdot\boldsymbol{n}^K + \psi^{N_F(K)}\boldsymbol{v}\cdot\boldsymbol{n}^{N_F(K)}\right)\,\mathrm{d}S(\boldsymbol{x})\right)\\
&= \sum_{F\in\mathcal{F}}\left(\int_F \widehat{\phi}\boldsymbol{v}\cdot[\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x})\right)\,.
\end{aligned} \tag{3.36}$$

■

Lemma 3.9. *Let $\widehat{\phi} = \{\!\{\phi\}\!\}$, $\psi^K \in V_{K,\Phi}$ locally defined in K and $\boldsymbol{v} \in \mathbb{C}^2$ a not necessarily constant vector. Then*

$$-\sum_{K\in\mathcal{K}}\sum_{F\in\partial K}\left(\int_F \left(\widehat{\phi} - \phi^K\right)\psi^K\boldsymbol{v}\cdot\boldsymbol{n}^K \,\mathrm{d}S(\boldsymbol{x})\right) = \sum_{F\in\mathcal{F}}\left(\int_F \boldsymbol{v}\cdot[\![\phi]\!]\{\!\{\psi\}\!\} \,\mathrm{d}S(\boldsymbol{x})\right)\,. \tag{3.37}$$

Proof:
When summing over all cells, each interface F is considered twice in total, once for a cell K and once for its unique neighbour $N_F(K)$. Hence, we obtain

$$\begin{aligned}
&-\sum_{K\in\mathcal{K}}\sum_{F\in\partial K}\left(\int_F\left(\widehat{\phi}-\phi^K\right)\psi^K\boldsymbol{v}\cdot\boldsymbol{n}^K\,\mathrm{d}S(\boldsymbol{x})\right)\\
&=-\sum_{K\in\mathcal{K}}\sum_{F\in\partial K}\left(\int_F\frac{1}{2}\boldsymbol{v}\cdot\boldsymbol{n}^K\left(-\phi^K+\phi^{N_F(K)}\right)\psi^K\,\mathrm{d}S(\boldsymbol{x})\right)\\
&=\sum_{K\in\mathcal{K}}\sum_{F\in\partial K}\left(\int_F\boldsymbol{v}\cdot\left(\phi^K\boldsymbol{n}^K-\phi^{N_F(K)}\boldsymbol{n}^K\right)\frac{1}{2}\psi^K\,\mathrm{d}S(\boldsymbol{x})\right)\\
&=\sum_{K\in\mathcal{K}}\sum_{F\in\partial K}\left(\int_F\boldsymbol{v}\cdot\left(\phi^K\boldsymbol{n}^K+\phi^{N_F(K)}\boldsymbol{n}^{N_F(K)}\right)\frac{1}{2}\psi^K\,\mathrm{d}S(\boldsymbol{x})\right)\\
&=\sum_{K\in\mathcal{K}}\sum_{F\in\partial K}\left(\int_F\boldsymbol{v}\cdot[\![\phi]\!]\frac{1}{2}\psi^K\,\mathrm{d}S(\boldsymbol{x})\right)\\
&=\sum_{F\in\mathcal{F}}\left(\int_F\boldsymbol{v}\cdot[\![\phi]\!]\frac{1}{2}\psi^K\,\mathrm{d}S(\boldsymbol{x})+\int_F\boldsymbol{v}\cdot[\![\phi]\!]\frac{1}{2}\psi^{N_F(K)}\,\mathrm{d}S(\boldsymbol{x})\right)\\
&=\sum_{F\in\mathcal{F}}\left(\int_F\boldsymbol{v}\cdot[\![\phi]\!]\frac{1}{2}\left(\psi^K+\psi^{N_F(K)}\right)\,\mathrm{d}S(\boldsymbol{x})\right)=\sum_{F\in\mathcal{F}}\left(\int_F\boldsymbol{v}\cdot[\![\phi]\!]\{\!\{\psi\}\!\}\,\mathrm{d}S(\boldsymbol{x})\right)\,.
\end{aligned}\tag{3.38}$$

■

Applying these lemmata to (3.34), we obtain $\forall\psi^K\in V_\Phi$

$$\begin{aligned}
&\sum_{K\in\mathcal{K}}\left(\int_K\boldsymbol{b}\cdot\nabla\phi^K\boldsymbol{b}\cdot\nabla\psi^K\,\mathrm{d}\boldsymbol{x}\right)-\sum_{F\in\mathcal{F}}\left(\int_F\{\!\{\boldsymbol{b}\cdot\nabla\phi\}\!\}\boldsymbol{b}\cdot[\![\psi]\!]\,\mathrm{d}S(\boldsymbol{x})\right)\\
&\quad-\sum_{F\in\mathcal{F}}\left(\int_F\boldsymbol{b}\cdot[\![\phi]\!]\{\!\{\boldsymbol{b}\cdot\nabla\psi\}\!\}\,\mathrm{d}S(\boldsymbol{x})\right)\\
&\quad+\sum_{F\in\mathcal{F}}\left(\frac{\eta}{h_F}\int_F\boldsymbol{b}\cdot[\![\phi]\!]\boldsymbol{b}\cdot[\![\psi]\!]\,\mathrm{d}S(\boldsymbol{x})\right)=\omega^2\sum_{K\in\mathcal{K}}\left(\int_K\phi^K\psi^K\,\mathrm{d}\boldsymbol{x}\right)\,.
\end{aligned}\tag{3.39}$$

Considering (3.39), the property of symmetry of the bilinear form is obvious as ϕ and ψ are interchangeable.
The method is consistent as for a continuous solution $\left(\phi_0,\omega_0^2\right)$ it holds $\{\!\{\phi_0\}\!\}=\phi_0$, $\{\!\{\boldsymbol{b}\cdot\nabla\phi_0\}\!\}=\boldsymbol{b}\cdot\nabla\phi_0$ and $[\![\phi_0]\!]=0$. Insertion in (3.39) and one integration by parts allows us to retrieve (3.29).

3.5.2 Construction of the system matrices

This section shows, that the matrices discretizing (3.1) based on the variational form (3.39) are symmetric if the spaces for test and trial functions are chosen identically. Generalized and standard eigenvalue problem systems are set up.

Given a basis $\left(\phi_k^K\right)_k$ of $V_{K,\Phi}$ for all K we can write the discrete solution ϕ as

$$\phi(\boldsymbol{x}) = \sum_{K\in\mathcal{K}} \sum_k \Phi_k^K \phi_k^K(\boldsymbol{x}) \qquad , \qquad \Phi_k^K \in \mathbb{R} \tag{3.40}$$

for coefficients $\boldsymbol{\Phi} = \left(\Phi_k^K\right)$. The variational form (3.39) can be translated to matrix form

$$A\boldsymbol{\Phi} = \omega^2 B\boldsymbol{\Phi} \tag{3.41}$$

where we call A the left hand side system matrix and B the right hand side system matrix. For assembling these matrices, we are left with the choice of suitable spaces for test and trial functions ψ, ϕ and an enumeration of the cells of the mesh. A detailed discussion of suitable bases for the implementation can be found in Section 5.2.
We observe that the system matrices are symmetric.

Theorem 3.10. *A and B are symmetric when choosing the same basis for discretizing test and trial functions.*

Proof:
Transposition of the system matrices is equivalent to exchanging test and trial functions in the formulae for building these matrices. As ϕ and ψ are interchangeable in (3.39) and the same basis for test and trial functions is chosen, the resulting system matrices are symmetric.

■

Hence, discretizing ϕ, ψ by using the same function space V_Φ ensures the symmetry of the system matrices and consequently real eigenvalues. The property of symmetry can simplify the calculation of eigenvalues by using particularly suited eigenvalue solvers. More details on this discussion are given in Section 5.5.1.
We can either solve the generalized eigenvalue problem (3.41) or reduce it to a standard eigenvalue problem. Multiplication of the inverse right hand side system matrix B produces a system matrix which is non-symmetric in general. This can be circumvented by using the matrix root of B which exists since B is symmetric positive definite as a block diagonal mass matrix [76, p.448, 46.(a)] and is defined as

$$B^{\frac{1}{2}} B^{\frac{1}{2}} = B \, . \tag{3.42}$$

We then obtain the standard eigenvalue problem

$$\left(B^{-\frac{1}{2}} A B^{-\frac{1}{2}}\right)\left(B^{\frac{1}{2}}\boldsymbol{\Phi}\right) = \omega^2 B^{\frac{1}{2}}\boldsymbol{\Phi} \tag{3.43}$$

with a symmetric system matrix as

$$\left(B^{-\frac{1}{2}} A B^{-\frac{1}{2}}\right)^\top = \left(B^{-\frac{1}{2}}\right)^\top A^\top \left(B^{-\frac{1}{2}}\right)^\top = B^{-\frac{1}{2}} A B^{-\frac{1}{2}} \, . \tag{3.44}$$

If eigenvectors are of interest, close attention has to be paid when using the reduced system (3.43) as the eigenvectors are scaled by $B^{\frac{1}{2}}$.
For constant $\boldsymbol{b}$, the system matrices can be constructed using the Kronecker matrix formalism outlined in Section 3.8.

3.6 Construction of mixed form matrix systems

In comparison to Section 3.5, we split the anisotropic wave equation (3.1) to obtain two integral equations in variational form and set up the resulting matrices for the discretization of this system of integral equations. The construction focuses on the properties for bilinear forms outlined in Section 3.4. In general, the spaces for test and trial functions can be chosen independently. Throughout this thesis, we focus on the case where these spaces coincide but mention whenever this assumption is essential for a result.
In Section 3.6.1, we set up the general variational mixed form. In Sections 3.6.2 and 3.6.3, we use local discontinuous Galerkin fluxes for the mixed variational form and constructs the respective system matrices. This is repeated in Sections 3.6.4 and 3.6.5 for Bassi-Rebay 2 fluxes.

3.6.1 Construction of a mixed form

For preserving symmetry of the system of integral equations, we propose to split (3.1) into two equations using the substitution

$$\begin{cases} \boldsymbol{b} \cdot \nabla \phi & = u \\ -\nabla \cdot (\boldsymbol{b} u) & = \omega^2 \phi \, . \end{cases} \tag{3.45}$$

Note that the symmetric splitting of the tensor leads to a scalar-valued equation for the parallel gradient represented by u. The associated weak form for test and trial functions $v, u \in V_U, \psi, \phi \in V_\Phi$ and a mesh with cells $K \in \mathcal{K}$ writes

$$\begin{cases} \sum\limits_{K \in \mathcal{K}} \left(\int_K \boldsymbol{b} \cdot \nabla \phi v \, \mathrm{d}\boldsymbol{x} \right) = \sum\limits_{K \in \mathcal{K}} \left(\int_K u v \, \mathrm{d}\boldsymbol{x} \right) & \forall v \in V_U \\ \sum\limits_{K \in \mathcal{K}} \left(\int_K -\nabla \cdot (\boldsymbol{b} u) \, \psi \, \mathrm{d}\boldsymbol{x} \right) = \omega^2 \sum\limits_{K \in \mathcal{K}} \left(\int_K \phi \psi \, \mathrm{d}\boldsymbol{x} \right) & \forall \psi \in V_\Phi \end{cases} \tag{3.46}$$

Using locally defined test and trial functions $v^K, u^K \in V_{K,U}$, $\psi^K, \phi^K \in V_{K,\Phi}$ in the locally defined function spaces $V_{K,U}, V_{K,\Phi}$ for a cell K, (3.46) writes

$$\begin{cases} \int_K \boldsymbol{b} \cdot \nabla \phi^K v^K \, \mathrm{d}\boldsymbol{x} = \int_K u^K v^K \, \mathrm{d}\boldsymbol{x} & \forall v^K \in V_{K,U} \\ \int_K -\nabla \cdot \left(\boldsymbol{b} u^K \right) \psi^K \, \mathrm{d}\boldsymbol{x} = \omega^2 \int_K \phi^K \psi^K \, \mathrm{d}\boldsymbol{x} & \forall \psi^K \in V_{K,\Phi} \end{cases} \tag{3.47}$$

The global spaces for test functions defined in Ω are

$$V_\Phi = \bigcup_{K\in\mathcal{K}} V_{K,\Phi} \quad , \quad V_U = \bigcup_{K\in\mathcal{K}} V_{K,U} \, . \tag{3.48}$$

In the following, we omit the spaces for test functions and the superscripts K whenever there is no need for differentiating the cells of definition.
Integration by parts of (3.47) leads to

$$\begin{cases} -\int_K \phi \nabla\cdot(\boldsymbol{b}v)\ \mathrm{d}\boldsymbol{x} + \sum_{F\in\partial K}\left(\int_F \widehat{\phi} v \boldsymbol{b}\cdot\boldsymbol{n}\ \mathrm{d}S(\boldsymbol{x})\right) &= \int_K uv\ \mathrm{d}\boldsymbol{x} \\ \int_K u\boldsymbol{b}\cdot\nabla\psi\ \mathrm{d}\boldsymbol{x} - \sum_{F\in\partial K}\left(\int_F \widehat{u}\psi\boldsymbol{b}\cdot\boldsymbol{n}\ \mathrm{d}S(\boldsymbol{x})\right) &= \omega^2\int_K \phi\psi\ \mathrm{d}\boldsymbol{x} \end{cases} \tag{3.49}$$

for which numerical fluxes $\widehat{u}$ and $\widehat{\phi}$ have to be defined and $\boldsymbol{n} = \boldsymbol{n}^K$. As the differential operators in the first integrals of (3.49) act on different function spaces, namely once on $V_{K,U}$ and once on $V_{K,\Phi}$, we integrate the first equation by parts once more. Using the inner cell function ϕ on the boundary, we obtain

$$\begin{cases} \int_K \boldsymbol{b}\cdot\nabla\phi v\ \mathrm{d}\boldsymbol{x} + \sum_{F\in\partial K}\left(\int_F (\widehat{\phi}-\phi)\, v\boldsymbol{b}\cdot\boldsymbol{n}\ \mathrm{d}S(\boldsymbol{x})\right) &= \int_K uv\ \mathrm{d}\boldsymbol{x} \\ \int_K u\boldsymbol{b}\cdot\nabla\psi\ \mathrm{d}\boldsymbol{x} - \sum_{F\in\partial K}\left(\int_F \widehat{u}\psi\boldsymbol{b}\cdot\boldsymbol{n}\ \mathrm{d}S(\boldsymbol{x})\right) &= \omega^2\int_K \phi\psi\ \mathrm{d}\boldsymbol{x} \end{cases} \tag{3.50}$$

where we reformulate the first equation to obtain

$$\begin{cases} -\int_K uv\ \mathrm{d}\boldsymbol{x} + \int_K \boldsymbol{b}\cdot\nabla\phi v\ \mathrm{d}\boldsymbol{x} + \sum_{F\in\partial K}\left(\int_F (\widehat{\phi}-\phi)\, v\boldsymbol{b}\cdot\boldsymbol{n}\ \mathrm{d}S(\boldsymbol{x})\right) = 0 \\ \int_K u\boldsymbol{b}\cdot\nabla\psi\ \mathrm{d}\boldsymbol{x} - \sum_{F\in\partial K}\left(\int_F \widehat{u}\psi\boldsymbol{b}\cdot\boldsymbol{n}\ \mathrm{d}S(\boldsymbol{x})\right) = \omega^2\int_K \phi\psi\ \mathrm{d}\boldsymbol{x} \end{cases} \tag{3.51}$$

3.6.2 Local discontinuous Galerkin fluxes

As a first example for the fluxes in (3.51), we use slightly modified local discontinuous Galerkin (LDG) fluxes proposed in [51] and summarized in [52, Table 3.1] with the choice of $\beta \equiv 0$ and adapted α_j such that

$$\widehat{u} = \{\!\{u\}\!\} - \frac{\eta}{h_F}\boldsymbol{b}\cdot[\![\phi]\!] \quad , \quad \widehat{\phi} = \{\!\{\phi\}\!\} \tag{3.52}$$

where jumps and averages are defined in (3.25). Similar to (3.33), η is a stabilization parameter larger than zero and h_F is a parameter dependent on the length of the cell edge with interface F. We note that there are more possible choices for the fluxes. Setting $\eta = 0$ leads to the Bassi-Rebay 1 scheme [48]. Another example, namely the Bassi-Rebay 2 scheme, is discussed in Sections 3.6.4

and 3.6.5. The impact of the choice of fluxes onto the numerical results is discussed in Section 6.2.5. Insertion of LDG fluxes (3.52) into (3.51) yields $\forall K \in \mathcal{K}$

$$\begin{cases} -\int_K u^K v^K \,\mathrm{d}\boldsymbol{x} + \int_K \boldsymbol{b}\cdot\nabla\phi^K v^K \,\mathrm{d}\boldsymbol{x} \\ \quad -\sum_{F\in\partial K}\left(\int_F \frac{1}{2}\boldsymbol{b}\cdot\left(\phi^K\boldsymbol{n}^K + \phi^{N_F(K)}\boldsymbol{n}^{N_F(K)}\right)v^K \,\mathrm{d}S(\boldsymbol{x})\right) = 0 \qquad \forall v_K \in V_{K,U} \\ \\ -\sum_{F\in\partial K}\left(\int_F \frac{1}{2}\left(u^K + u^{N_F(K)}\right)\psi^K\boldsymbol{b}\cdot\boldsymbol{n}^K \,\mathrm{d}S(\boldsymbol{x})\right) \\ \quad +\sum_{F\in\partial K}\left(\int_F \frac{\eta}{h_F}\boldsymbol{b}\cdot\left(\phi^K\boldsymbol{n}^K + \phi^{N_F(K)}\boldsymbol{n}^{N_F(K)}\right)\psi^K\boldsymbol{b}\cdot\boldsymbol{n}^K \,\mathrm{d}S(\boldsymbol{x})\right) \\ \quad +\int_K u^K\boldsymbol{b}\cdot\nabla\psi^K \,\mathrm{d}\boldsymbol{x} = \omega^2\int_K \phi^K\psi^K \,\mathrm{d}\boldsymbol{x} \qquad \forall\psi^K \in V_{K,\Phi} \end{cases} \tag{3.53}$$

To retrieve the global variational formulation, we sum (3.53) over all over all elements K. Using Lemma 3.9 on the surface term of the first equation of (3.53), we obtain

$$\sum_{K\in\mathcal{K}}\left(-\int_K u^K v^K \,\mathrm{d}\boldsymbol{x} + \int_K \boldsymbol{b}\cdot\nabla\phi^K v^K \,\mathrm{d}\boldsymbol{x}\right) - \sum_{F\in\mathcal{F}}\int_F \boldsymbol{b}\cdot[\![\phi]\!]\{\!\{v\}\!\} \,\mathrm{d}S(\boldsymbol{x}) = 0 \tag{3.54}$$

whereas the second equation of (3.53) using Lemma 3.8 yields

$$\begin{aligned} &-\sum_{F\in\mathcal{F}}\left(\int_F \{\!\{u\}\!\}\boldsymbol{b}\cdot[\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x})\right) + \sum_{F\in\mathcal{F}}\left(\int_F \frac{\eta}{h_F}\boldsymbol{b}\cdot[\![\phi]\!]\boldsymbol{b}\cdot[\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x})\right) \\ &\quad +\sum_{K\in\mathcal{K}}\left(\int_K u^K\boldsymbol{b}\cdot\nabla\psi^K \,\mathrm{d}\boldsymbol{x}\right) = \omega^2\sum_{K\in\mathcal{K}}\int_K \phi^K\psi^K \,\mathrm{d}\boldsymbol{x}\,. \end{aligned} \tag{3.55}$$

The method is consistent as for a continuous solution $(\phi_0, u_0, \omega_0^2)$ it holds $\{\!\{\phi_0\}\!\} = \phi_0, \{\!\{u_0\}\!\} = u_0$ and $[\![\phi_0]\!] = 0$. Insertion in (3.54) and (3.55) and one integration by parts in (3.55) allows us to retrieve (3.46).

3.6.3 Local discontinuous Galerkin system matrices

In comparison to Section 3.5.2, we are now dealing with a matrix system with two coefficient vectors $\boldsymbol{U}$ and $\boldsymbol{\Phi}$ given by the discretization of u and ϕ using the bases $\left(u_k^K\right)_k$ of $V_{K,U}$ and $\left(\phi_k^K\right)_k$ of $V_{K,\Phi}$ for all K such that

$$u(\boldsymbol{x}) = \sum_{K\in\mathcal{K}}\sum_k U_k^K u_k^K(\boldsymbol{x}) \qquad , \qquad U_k^K \in \mathbb{R} \tag{3.56}$$

$$\phi(\boldsymbol{x}) = \sum_{K\in\mathcal{K}}\sum_k \Phi_k^K \phi_k^K(\boldsymbol{x}) \qquad , \qquad \Phi_k^K \in \mathbb{R} \tag{3.57}$$

We obtain a system with coefficient vectors $\boldsymbol{U}$ and $\boldsymbol{\Phi}$ as the equivalent to (3.41)

$$\begin{pmatrix} M_1 & M_2 \\ M_3 & M_4 \end{pmatrix} \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{\Phi} \end{pmatrix} = \omega^2 \begin{pmatrix} 0 & 0 \\ 0 & M_5 \end{pmatrix} \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{\Phi} \end{pmatrix} \tag{3.58}$$

which can either be solved as a generalized eigenvalue problem or by reducing (3.58) using $\boldsymbol{U} = -M_1^{-1} M_2 \boldsymbol{\Phi}$ to

$$\left(-M_3 M_1^{-1} M_2 + M_4 \right) \boldsymbol{\Phi} = \omega^2 M_5 \boldsymbol{\Phi} \, . \tag{3.59}$$

However, this is only feasible if the inversion of M_1 is possible and numerically simple which is the case for numerical fluxes which don't couple neighbouring cells for u. Then M_1 is a sparse matrix of block diagonal structure. This is fulfilled for the choice of LDG fluxes as $\widehat{\phi}$ is independent of u. For constant $\boldsymbol{b}$, the individual matrices M_k, $k = 1, \ldots, 5$ can be built as Kronecker matrix products outlined in Section 3.8 using the respective spaces for test and trial functions and local cell contributions given in (3.53).

For building the blocks of the left and right hand side system matrices, we associate the matrix components using equations (3.54), (3.55) as follows

$$M_{UV} \leftrightarrow \sum_{K \in \mathcal{K}} \int_K u^K v^K \, \mathrm{d}\boldsymbol{x} \tag{3.60}$$

$$A_{\Phi V} \leftrightarrow \sum_{K \in \mathcal{K}} \int_K \boldsymbol{b} \cdot \nabla \phi^K v^K \, \mathrm{d}\boldsymbol{x} \tag{3.61}$$

$$B_{\Phi V} \leftrightarrow \sum_{F \in \mathcal{F}} \int_F \boldsymbol{b} \cdot [\![\phi]\!] \{\!\{ v \}\!\} \, \mathrm{d}S(\boldsymbol{x}) \tag{3.62}$$

$$B_{U\Psi} \leftrightarrow \sum_{F \in \mathcal{F}} \int_F \{\!\{ u \}\!\} \boldsymbol{b} \cdot [\![\psi]\!] \, \mathrm{d}S(\boldsymbol{x}) \tag{3.63}$$

$$B_{\Phi\Psi} \leftrightarrow \sum_{F \in \mathcal{F}} \int_F \frac{\eta}{h_F} \boldsymbol{b} \cdot [\![\phi]\!] \boldsymbol{b} \cdot [\![\psi]\!] \, \mathrm{d}S(\boldsymbol{x}) \tag{3.64}$$

$$A_{U\Psi} \leftrightarrow \sum_{K \in \mathcal{K}} \int_K u^K \boldsymbol{b} \cdot \nabla \psi^K \, \mathrm{d}\boldsymbol{x} \tag{3.65}$$

$$M_{\Phi\Psi} \leftrightarrow \sum_{K \in \mathcal{K}} \int_K \phi^K \psi^K \, \mathrm{d}\boldsymbol{x} \tag{3.66}$$

Equation (3.58) then writes

$$\begin{pmatrix} -M_{UV} & A_{\Phi V} - B_{\Phi V} \\ A_{U\Psi} - B_{U\Psi} & B_{\Phi\Psi} \end{pmatrix} \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{\Phi} \end{pmatrix} = \omega^2 \begin{pmatrix} 0 & 0 \\ 0 & M_{\Phi\Psi} \end{pmatrix} \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{\Phi} \end{pmatrix} \, . \tag{3.67}$$

Theorem 3.11. *The system matrices*

$$\begin{pmatrix} -M_{UV} & A_{\Phi V} - B_{\Phi V} \\ A_{U\Psi} - B_{U\Psi} & B_{\Phi\Psi} \end{pmatrix} \quad , \quad \begin{pmatrix} 0 & 0 \\ 0 & M_{\Phi\Psi} \end{pmatrix} \tag{3.68}$$

are symmetric if the spaces for test and trial functions for v, u coincide and the spaces for test and trial functions ψ, ϕ coincide.

Proof:

Choosing the same basis for the test and trial functions of each variable, it is obvious that the mass

matrices M_{UV} and $M_{\Phi\Psi}$ and the penalization matrix $B_{\Phi\Psi}$ are symmetric. Furthermore it holds $A_{\Phi V} = A_{U\Psi}^\top$ as well as $B_{\Phi V} = B_{U\Psi}^\top$ since ϕ, ψ and u, v are interchangeable between (3.61) and (3.65) as well as (3.62) and (3.63).

■

We either consider the generalized eigenvalue problem in full form (3.67) or reduced form (3.59) as M_{UV} is a block diagonal mass matrix and hence easily invertible. Using the same construction as at the end of Section 3.5.2, we obtain a standard eigenvalue problem using the matrix root $M_{\Phi\Psi}^{\frac{1}{2}}$

$$A\left(M_{\Phi\Psi}^{\frac{1}{2}}\mathbf{\Phi}\right) = \omega^2 M_{\Phi\Psi}^{\frac{1}{2}}\mathbf{\Phi} \tag{3.69}$$

with the symmetric system matrix

$$A := M_{\Phi\Psi}^{-\frac{1}{2}}\left(\left(A_{U\Psi} - B_{U\Psi}\right)M_{UV}^{-1}\left(A_{\Phi V} - B_{\Phi V}\right) + B_{\Phi\Psi}\right)M_{\Phi\Psi}^{-\frac{1}{2}}\,. \tag{3.70}$$

Corollary 3.12. $A = A^\top$.

Proof:
Using the symmetry of the matrices in Theorem 3.11 we conclude

$$\begin{aligned}
&\left(M_{\Phi\Psi}^{-\frac{1}{2}}\left(\left(A_{U\Psi} - B_{U\Psi}\right)M_{UV}^{-1}\left(A_{\Phi V} - B_{\Phi V}\right) + B_{\Phi\Psi}\right)M_{\Phi\Psi}^{-\frac{1}{2}}\right)^\top \\
&= \left(M_{\Phi\Psi}^{-\frac{1}{2}}\right)^\top\left(\left(\left(A_{U\Psi} - B_{U\Psi}\right)M_{UV}^{-1}\left(A_{\Phi V} - B_{\Phi V}\right)\right)^\top + B_{\Phi\Psi}^\top\right)\left(M_{\Phi\Psi}^{-\frac{1}{2}}\right)^\top \\
&= \left(M_{\Phi\Psi}^\top\right)^{-\frac{1}{2}}\left(\left(A_{\Phi V} - B_{\Phi V}\right)^\top\left(M_{UV}^{-1}\right)^\top\left(A_{U\Psi} - B_{U\Psi}\right)^\top + B_{\Phi\Psi}\right)\left(M_{\Phi\Psi}^\top\right)^{-\frac{1}{2}} \\
&= M_{\Phi\Psi}^{-\frac{1}{2}}\left(\left(A_{U\Psi} - B_{U\Psi}\right)M_{UV}^{-1}\left(A_{\Phi V} - B_{\Phi V}\right) + B_{\Phi\Psi}\right)M_{\Phi\Psi}^{-\frac{1}{2}}\,.
\end{aligned} \tag{3.71}$$

■

Again, if eigenvectors are of interest, close attention has to be paid when using the reduced system (3.69) as the eigenvectors are scaled by $M_{\Phi\Psi}^{\frac{1}{2}}$.

3.6.4 Bassi-Rebay 2 fluxes

Local discontinuous Galerkin fluxes are discussed in Sections 3.6.2 and 3.6.3. As a second example for fluxes in (3.51), we describe the Bassi-Rebay 2 (BR2) discontinuous Galerkin fluxes proposed in [77] and summarized in [52, Table 3.1, Bassi et al.]. These are defined as

$$\widehat{\phi} = \{\!\{\phi\}\!\} \qquad , \qquad \widehat{u} = \{\!\{u^*\}\!\} \tag{3.72}$$

for a yet undefined so-called lifted surface gradient u^* and the average defined in (3.25). The lifted volume gradient $u^K \in V_{K,U}$ is split into the local gradient $q^K \in V_{K,Q}$, $\forall K \in \mathcal{K}$ on the cell K and

interface contributions $r^{K,F} \in V_{K,F,R}$, $\forall K \in \mathcal{K}, F \in \partial K$, so-called lifting terms, associated with an interface F of the cell K such that

$$u^K = q^K + \sum_{F \in \partial K} r^{K,F} \qquad \text{on} \quad K \,. \tag{3.73}$$

The global function spaces are given by

$$V_Q = \bigcup_{K \in \mathcal{K}} V_{K,Q} \qquad , \qquad V_R = \bigcup_{K \in \mathcal{K}} \bigcup_{F \in \partial K} V_{K,F,R} \,. \tag{3.74}$$

Useful bases for BR2 are discussed in Section 5.2.
We define the flux containing the lifted gradient for a single interface F as

$$\widehat{u}\,\big|_F = \{\!\{q\}\!\}\big|_F + \eta_{\mathrm{BR2}} \{\!\{r^F\}\!\} \tag{3.75}$$

with the stabilization parameter η_{BR2} which should be chosen larger than the number of neighbours per cell. We note that $\{\!\{r^F\}\!\}$ is well-defined. The BR2 flux only penalizes the lifting terms belonging to interface F. Inserting (3.73) and (3.75) into (3.51), we obtain the following system of integral equations for q^K, $r^{K,F}$, ϕ^K, $\forall K \in \mathcal{K}$

$$\begin{cases} \displaystyle\int_K q^K v_1^K \,\mathrm{d}\boldsymbol{x} - \int_K \boldsymbol{b} \cdot \nabla \phi^K v_1^K \,\mathrm{d}\boldsymbol{x} = 0 & \forall v_1^K \in V_{K,Q} \\ \displaystyle\int_K r^{K,F} v_2^{K,F} \,\mathrm{d}\boldsymbol{x} - \int_F \frac{1}{2} \boldsymbol{b} \cdot \left(\phi^K \boldsymbol{n}^K + \phi^{N_F(K)} \boldsymbol{n}^{N_F(K)} \right) v_2^{K,F} \,\mathrm{d}S(\boldsymbol{x}) = 0 & \forall F \in \partial K \,, v_2^{K,F} \in V_{K,F,R} \\ \displaystyle\int_K q^K \boldsymbol{b} \cdot \nabla \psi^K \,\mathrm{d}\boldsymbol{x} - \sum_{F \in \partial K} \left(\int_F \frac{1}{2} \left(q^K + q^{N_F(K)} \right) \psi^K \boldsymbol{b} \cdot \boldsymbol{n}^K \,\mathrm{d}S(\boldsymbol{x}) \right) & \\ \displaystyle - \sum_{F \in \partial K} \left(\eta_{\mathrm{BR2}} \int_F \frac{1}{2} \left(r^{K,F} + r^{N_F(K),F} \right) \psi^K \boldsymbol{b} \cdot \boldsymbol{n}^K \,\mathrm{d}S(\boldsymbol{x}) \right) & \\ \displaystyle + \sum_{F \in \partial K} \left(\int_K r^{K,F} \boldsymbol{b} \cdot \nabla \psi^K \,\mathrm{d}\boldsymbol{x} \right) = \omega^2 \int_K \phi^K \psi^K \,\mathrm{d}\boldsymbol{x} & \forall \psi^K \in V_{K,\Phi} \end{cases} \tag{3.76}$$

where we note that the first equation was split into distinct volume terms q and lifting terms r_F by definition. Summing the element contributions (3.76) over all elements K, we obtain for the first equation

$$\sum_{K \in \mathcal{K}} \left(\int_K q^K v_1^K \,\mathrm{d}\boldsymbol{x} - \int_K \boldsymbol{b} \cdot \nabla \phi^K v_1^K \,\mathrm{d}\boldsymbol{x} \right) = 0 \,. \tag{3.77}$$

For the second equation, we obtain for each interface F

$$- \sum_{K \in \mathcal{K}} \left(\int_F \frac{1}{2} \boldsymbol{b} \cdot [\![\phi]\!] v_2^{K,F} \,\mathrm{d}S(\boldsymbol{x}) \right) + \sum_{K \in \mathcal{K}} \left(\int_K r^{K,F} v_2^{K,F} \,\mathrm{d}\boldsymbol{x} \right) = 0 \,. \tag{3.78}$$

The volume integrals of the third equation write

$$\sum_{K\in\mathcal{K}} \int_K q^K \boldsymbol{b}\cdot\nabla\psi^K \,\mathrm{d}\boldsymbol{x}\,, \tag{3.79}$$

$$\sum_{K\in\mathcal{K}}\sum_{F\in\partial K} \int_K r^{K,F} \boldsymbol{b}\cdot\nabla\psi^K \,\mathrm{d}\boldsymbol{x}\,, \tag{3.80}$$

$$\sum_{K\in\mathcal{K}} \int_K \phi^K\psi^K \,\mathrm{d}\boldsymbol{x}\,. \tag{3.81}$$

Using Lemma 3.8 on the surface terms of the third equation of (3.76) yields

$$\begin{aligned}\sum_{K\in\mathcal{K}}\sum_{F\in\partial K}\left(\int_F \{\!\{q\}\!\}\psi^K\boldsymbol{b}\cdot\boldsymbol{n}^K \,\mathrm{d}S(\boldsymbol{x})\right) &= \sum_{F\in\mathcal{F}}\int_F \{\!\{q\}\!\}\boldsymbol{b}\cdot[\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x})\\ &= \sum_{K\in\mathcal{K}}\sum_{F\in\partial K}\left(\int_F \frac{1}{2}q^K\boldsymbol{b}\cdot[\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x})\right)\,,\end{aligned} \tag{3.82}$$

$$\begin{aligned}\sum_{K\in\mathcal{K}}\sum_{F\in\partial K}\left(\eta_{\mathrm{BR2}}\int_F \{\!\{r^F\}\!\}\psi^K\boldsymbol{b}\cdot\boldsymbol{n}^K \,\mathrm{d}S(\boldsymbol{x})\right) &= \sum_{F\in\mathcal{F}}\eta_{\mathrm{BR2}}\int_F \{\!\{r^F\}\!\}\boldsymbol{b}\cdot[\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x})\\ &= \sum_{K\in\mathcal{K}}\sum_{F\in\partial K}\eta_{\mathrm{BR2}}\int_F \boldsymbol{b}\cdot[\![\psi]\!]\frac{1}{2}r^{K,F} \,\mathrm{d}S(\boldsymbol{x})\,.\end{aligned} \tag{3.83}$$

The method is consistent as for a continuous solution $(\phi_0, q_0, r_0, \omega_0^2)$ it holds $\{\!\{\phi_0\}\!\} = \phi_0$, $\{\!\{q_0\}\!\} = q_0$ and $[\![\phi_0]\!] = 0$. Insertion in (3.78) yields $r_0^{K,F} = 0$ for all K, F. Then (3.83) vanishes and we obtain

$$\begin{cases}\displaystyle\sum_{K\in\mathcal{K}}\left(\int_K q_0^K v_1^K \,\mathrm{d}\boldsymbol{x} - \int_K \boldsymbol{b}\cdot\nabla\phi_0^K v_1^K \,\mathrm{d}\boldsymbol{x}\right) &= 0\\ \displaystyle\sum_{K\in\mathcal{K}}\int_K q_0^K\boldsymbol{b}\cdot\nabla\psi^K \,\mathrm{d}\boldsymbol{x} - \sum_{F\in\mathcal{F}}\int_F q_0\boldsymbol{b}\cdot[\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x}) &= \omega_0^2\displaystyle\sum_{K\in\mathcal{K}}\int_K \phi_0^K\psi^K \,\mathrm{d}\boldsymbol{x}\end{cases} \tag{3.84}$$

of which one integration by parts of the second part of the equation allows us to retrieve (3.46) with $q_0 = u$.

3.6.5 Bassi-Rebay 2 system matrices

This section first deals with basics of the setup of the system matrices and their properties for the Bassi-Rebay 2 fluxes. At the end of this section, we give details on the assembly of the system matrices using the locally aligned mesh of Section 3.3.3.

In comparison to Sections 3.5.2 and 3.6.3, we are now dealing with a system with multiple coefficient vectors. Defining bases on all K, $\left(q_k^K\right)_k$ of $V_{K,Q}$, $\left(r_k^{K,F}\right)_k$ of $V_{K,F,R}$ defined individually $\forall F\in\partial K$ and $\left(\phi_k^K\right)_k$ of $V_{K,\Phi}$, we can retrieve the parallel gradient u and the solution ϕ as

$$u(\boldsymbol{x}) = \sum_{K\in\mathcal{K}}\sum_k\left(Q_k^K q_k^K(\boldsymbol{x})\right) + \sum_{K\in\mathcal{K}}\sum_{F\in\partial K}\sum_k\left(R_k^{K,F} r_k^{K,F}(\boldsymbol{x})\right) \quad , \quad Q_k^K, R_k^{K,F}\in\mathbb{R} \tag{3.85}$$

$$\phi(\boldsymbol{x}) = \sum_{K\in\mathcal{K}}\sum_k \Phi_k^K\phi_k^K(\boldsymbol{x}) \quad , \quad \Phi_k^K\in\mathbb{R} \tag{3.86}$$

yielding coefficient vectors $\boldsymbol{Q}$, $\boldsymbol{R}$ and $\boldsymbol{\Phi}$. The ordering of coefficients in $\boldsymbol{Q}$ and $\boldsymbol{\Phi}$ is done in the same manner as in Sections 3.5.2 and 3.6.3. On top of the ordering in the number of cell K, $\boldsymbol{R}$ is additionally indexed for the interfaces $F \in \partial K$ for each K.
Following Section 3.6.3, the system matrices are set up as block matrices. We associate matrix components using equations (3.77) – (3.83) as follows

$$M_{QV_1} \leftrightarrow \left(\int_K q^K v_1^K \, \mathrm{d}\boldsymbol{x} \right)_{K \in \mathcal{K}} \tag{3.87}$$

$$A_{\Phi V_1} \leftrightarrow \left(\int_K \boldsymbol{b} \cdot \nabla \phi^K v_1^K \, \mathrm{d}\boldsymbol{x} \right)_{K \in \mathcal{K}} \tag{3.88}$$

$$M_{RV_2} \leftrightarrow \left(\int_K r^{K,F} v_2^{K,F} \, \mathrm{d}\boldsymbol{x} \right)_{K \in \mathcal{K}, F \in \partial K} \tag{3.89}$$

$$B_{\Phi V_2} \leftrightarrow \left(\int_F \boldsymbol{b} \cdot [\![\phi]\!] \frac{1}{2} v_2^{K,F} \, \mathrm{d}S(\boldsymbol{x}) \right)_{K \in \mathcal{K}, F \in \partial K} \tag{3.90}$$

$$A_{Q\Psi} \leftrightarrow \left(\int_K q^K \boldsymbol{b} \cdot \nabla \psi^K \, \mathrm{d}\boldsymbol{x} \right)_{K \in \mathcal{K}} \tag{3.91}$$

$$A_{R\Psi} \leftrightarrow \left(\int_K r^{K,F} \boldsymbol{b} \cdot \nabla \psi^K \, \mathrm{d}\boldsymbol{x} \right)_{K \in \mathcal{K}, F \in \partial K} \tag{3.92}$$

$$M_{\Phi\Psi} \leftrightarrow \left(\int_K \phi^K \psi^K \, \mathrm{d}\boldsymbol{x} \right)_{K \in \mathcal{K}} \tag{3.93}$$

$$B_{Q\Psi} \leftrightarrow \left(\sum_{F \in \partial K} \int_F \frac{1}{2} q^K \boldsymbol{b} \cdot [\![\psi]\!] \, \mathrm{d}S(\boldsymbol{x}) \right)_{K \in \mathcal{K}} \tag{3.94}$$

$$B_{R\Psi} \leftrightarrow \left(\eta_{\mathrm{BR2}} \int_F \boldsymbol{b} \cdot [\![\psi]\!] \frac{1}{2} r^{K,F} \, \mathrm{d}S(\boldsymbol{x}) \right)_{K \in \mathcal{K}, F \in \partial K} \tag{3.95}$$

where the blocks of the matrices are either indexed by $K \in \mathcal{K}$ or by a common index accounting for $K \in \mathcal{K}$ and $F \in \partial K$ which is mentioned in the subscript. The system matrices then write

$$\begin{pmatrix} M_{QV_1} & 0 & -A_{\Phi V_1} \\ 0 & M_{RV_2} & B_{\Phi V_2} \\ A_{Q\Psi} - B_{Q\Psi} & A_{R\Psi} - B_{R\Psi} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{Q} \\ \boldsymbol{R} \\ \boldsymbol{\Phi} \end{pmatrix} = \omega^2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & M_{\Phi\Psi} \end{pmatrix} \begin{pmatrix} \boldsymbol{Q} \\ \boldsymbol{R} \\ \boldsymbol{\Phi} \end{pmatrix} . \tag{3.96}$$

The left hand side system matrix is generally not symmetric as for example $A_{\Phi V_1}$ has no boundary contributions in comparison to $A_{Q\Psi} - B_{Q\Psi}$. For obtaining a symmetric system matrix, we reduce (3.96). Being block diagonal mass matrices, M_{QV_1} and M_{RV_2} are efficiently invertible, yielding

$$\boldsymbol{Q} = M_{QV_1}^{-1} A_{\Phi V_1} \boldsymbol{\Phi} \tag{3.97}$$

$$\boldsymbol{R} = -M_{RV_2}^{-1} B_{\Phi V_2} \boldsymbol{\Phi} \tag{3.98}$$

Hence, we obtain

$$A\boldsymbol{\Phi} = \omega^2 M_{\Phi\Psi} \boldsymbol{\Phi} \tag{3.99}$$

with system matrix

$$A := (A_{Q\Psi} - B_{Q\Psi})\, M_{QV_1}^{-1} A_{\Phi V_1} - (A_{R\Psi} - B_{R\Psi})\, M_{RV_2}^{-1} B_{\Phi V_2}\ . \tag{3.100}$$

Theorem 3.13. *If $V_{K,Q} = V_{K,F,R} \quad \forall K \in \mathcal{K}, F \in \mathcal{F}$ and the respective spaces for test and trial functions are identical, then $A = A^\top$.*

Proof:
In parts, this proof is closely related to those of Theorems 3.10 and 3.11. We split A into

$$A_{Q\Psi} M_{QV_1}^{-1} A_{\Phi V_1} - B_{Q\Psi} M_{QV_1}^{-1} A_{\Phi V_1} - A_{R\Psi} M_{RV_2}^{-1} B_{\Phi V_2} + B_{R\Psi} M_{RV_2}^{-1} B_{\Phi V_2} \tag{3.101}$$

to investigate the symmetry of its components. First

$$\left(A_{Q\Psi} M_{QV_1}^{-1} A_{\Phi V_1}\right)^\top = A_{\Phi V_1}^\top \left(M_{QV_1}^{-1}\right)^\top A_{Q\Psi}^\top = A_{\Phi V_1}^\top M_{QV_1}^{-1} A_{Q\Psi}^\top = A_{Q\Psi} M_{QV_1}^{-1} A_{\Phi V_1} \tag{3.102}$$

where the 3rd identity is fulfilled as ϕ, ψ and q, v_1 are interchangeable in (3.88) and (3.91).
Secondly, it holds

$$\left(B_{R\Psi} M_{RV_2}^{-1} B_{\Phi V_2}\right)^\top = B_{\Phi V_2}^\top \left(M_{RV_2}^{-1}\right)^\top B_{R\Psi}^\top = B_{\Phi V_2}^\top M_{RV_2}^{-1} B_{R\Psi}^\top = B_{R\Psi} M_{RV_2}^{-1} B_{\Phi V_2} \tag{3.103}$$

as ϕ, ψ and $r^{K,F}, v_2^{K,F}$ are interchangeable in (3.90) and (3.95). We note that the stabilization factor η_{BR2} is a global scalar constant which can be translated multiplicatively.
To finish the proof, we need to show

$$\left(A_{R\Psi} M_{RV_2}^{-1} B_{\Phi V_2} + B_{Q\Psi} M_{QV_1}^{-1} A_{\Phi V_1}\right)^\top = A_{R\Psi} M_{RV_2}^{-1} B_{\Phi V_2} + B_{Q\Psi} M_{QV_1}^{-1} A_{\Phi V_1}\ . \tag{3.104}$$

As $V_{K,Q} = V_{K,F,R} \quad \forall K \in \mathcal{K}, F \in \mathcal{F}$, we can exchange $r^{K,F}, v_2^{K,F}, q^K, v_1^K$ in the integral expressions and obtain the same matrices. We consider the extended matrix

$$B_{Q\Psi}^{\mathcal{F}} \leftrightarrow \left(\int_F \frac{1}{2} q^K \boldsymbol{b} \cdot [\![\psi]\!]\, \mathrm{d}S(\boldsymbol{x})\right)_{K\in\mathcal{K}, F\in\partial K} \tag{3.105}$$

and observe $\left(B_{Q\Psi}^{\mathcal{F}}\right)^\top = B_{\Phi V_2}$ as ϕ and ψ as well as q^K and $v_2^{K,F}$ are interchangeable in (3.90) and (3.105).
Considering matrices M_{QV_1} and M_{RV_2}, the former is built of mass blocks for each $K \in \mathcal{K}$ and the latter is built of mass blocks for each $K \in \mathcal{K}$ and for each $F \in \partial K$. For both matrices, these mass blocks are identical due to the choice of $V_{K,Q} = V_{K,F,R}$. Considering matrices $A_{\Phi V_1}^\top$ and $A_{R\Psi}$, the former is built of blocks for each $K \in \mathcal{K}$ and the latter of blocks for each $K \in \mathcal{K}$ and for each $F \in \partial K$. As ϕ and ψ as well as v_1^K and $r^{K,F}$ are interchangeable in the integral expressions (3.88) and (3.92) these blocks are the same. Hence, we obtain

$$\begin{aligned}\left(B_{Q\Psi} M_{QV_1}^{-1} A_{\Phi V_1}\right)^\top &= A_{\Phi V_1}^\top \left(M_{QV_1}^{-1}\right)^\top B_{Q\Psi}^\top \\ &= A_{R\Psi} \left(M_{RV_2}^{-1}\right)^\top \left(B_{Q\Psi}^{\mathcal{F}}\right)^\top = A_{R\Psi} M_{RV_2}^{-1} B_{\Phi V_2}\end{aligned} \tag{3.106}$$

which concludes the proof.

■

System (3.99) can be transformed to a standard eigenvalue problem using the inverse square root of $M_{\Phi\Psi}$ as presented at the end of Section 3.5.2 in (3.43) as $M_{\Phi\Psi}$ is symmetric positive definite as a block diagonal mass matrix.
We remark that the reduced system matrix has a higher sparsity compared to, e.g., the reduced system matrix (3.70) for LDG fluxes, as the decomposition of u into local and lifted gradients for each edge reduces the coupling of neighbouring cells [77].
We now discuss the setup of system matrices when using the locally aligned mesh presented in Section 3.3.3 for constant $\boldsymbol{b}$. Then, each cell has the same neighbouring structure and the coefficient vector $\boldsymbol{R}$ can be organized as

$$\boldsymbol{R} = \left(\boldsymbol{R}^{F_1}, \dots, \boldsymbol{R}^{F_6}\right)^\top \tag{3.107}$$

for a given enumeration of the 6 interfaces where each $\boldsymbol{R}^{F_i} = \left(R_k^{K,F_i}\right)_{k,K}$ is organized in the same way as $\boldsymbol{Q}$ and $\boldsymbol{\Phi}$. We obtain the following structure for matrices (3.89), (3.90), (3.92) and (3.95)

$$M_{RV_2} = \begin{pmatrix} M_{R^{F_1}V_2^{F_1}} & & \\ & \ddots & \\ & & M_{R^{F_6}V_2^{F_6}} \end{pmatrix} \quad , \quad B_{\Phi V_2} = \begin{pmatrix} B_{\Phi V_2^{F_1}} \\ \vdots \\ B_{\Phi V_2^{F_6}} \end{pmatrix} , \tag{3.108}$$

$$A_{R\Psi} = \begin{pmatrix} A_{R^{F_1}\Psi} & \dots & A_{R^{F_6}\Psi} \end{pmatrix} \quad , \quad B_{R\Psi} = \begin{pmatrix} B_{R^{F_1}\Psi} & \dots & B_{R^{F_6}\Psi} \end{pmatrix} . \tag{3.109}$$

In case of a conforming locally aligned mesh, two of the six interfaces collapse to a point and the respective boundary integrals evaluate to zero.
Each block of the system matrices (3.108), (3.109) and the remaining matrices of (3.87) – (3.94) can now be built using the Kronecker matrix formalism presented in Section 3.8.

3.7 Mixed variational form for MHD equilibria

So far, the anisotropic wave equation (3.1) is solved within the logical, fully periodic domain Ω. For physical applications, we solve this equation on a flux surface of an MHD equilibrium as outlined in Section 2.2. This transformation yields additional metric factors which have to be taken into account as outlined in Section 2.6 and summarized in (2.90). We cite the anisotropic wave equation in the frequency domain here once again in the form

$$-\nabla \cdot \left(\bar{\boldsymbol{b}} \frac{\|\nabla s\|_2^2 \left(F_T'(s)\right)^2}{\|\boldsymbol{B}_0\|_2^2 \sqrt{g}} \bar{\boldsymbol{b}} \cdot \nabla \phi \right) = \omega^2 \frac{\|\nabla s\|_2^2 \sqrt{g}}{\|\boldsymbol{B}_0\|_2^2} \phi \tag{3.110}$$

where $\bar{\boldsymbol{b}} = (\iota(s), 1)^\top$ with the normalized flux surface coordinate s. Introducing the $\boldsymbol{x}$-dependent definitions for the metric terms

$$\mathcal{M}_1 \coloneqq \frac{\|\nabla s\|_2 \, F_T'(s)}{\|\boldsymbol{B}_0\|_2 \left|\sqrt{g}\right|^{\frac{1}{2}}} \quad , \quad \mathcal{M}_2 \coloneqq \frac{\|\nabla s\|_2^2 \left|\sqrt{g}\right|}{\|\boldsymbol{B}_0\|_2^2} \tag{3.111}$$

where we account for the sign of $\sqrt{g}$, we can construct a mixed form of (3.110) as

$$\begin{cases} \mathcal{M}_1 \bar{\boldsymbol{b}} \cdot \nabla \phi & = u \\ -\nabla \cdot (\mathcal{M}_1 \bar{\boldsymbol{b}} u) & = \omega^2 \mathcal{M}_2 \phi \end{cases} \tag{3.112}$$

to obtain a symmetric system of integral equation.
The addition of metric terms is readily incorporated in the mixed variational forms of Section 3.6 as they hold for general $\boldsymbol{b}(\boldsymbol{x})$. We only have to include $\mathcal{M}_2$ in the mass integrals of the right hand side which yields

$$M_{\Phi\Psi} \leftrightarrow \left(\int_K \mathcal{M}_2 \phi^K \psi^K \, \mathrm{d}\boldsymbol{x} \right)_{K \in \mathcal{K}} \tag{3.113}$$

for equations (3.66) and (3.93). The symmetry of the system matrices remains unaffected by this operation since $\mathcal{M}_2$ is only introduced in the right hand side mass matrix.
We remark that $\mathcal{M}_1 \bar{\boldsymbol{b}}$ fulfills the setting of $\boldsymbol{b}(\boldsymbol{x}) = \alpha(\boldsymbol{x}) \, (b_1, b_2)^\top$ with constant $(b_1, b_2)^\top = (\iota(s), 1)^\top$ on a single flux surface and scalar variation $\alpha(\boldsymbol{x}) = \mathcal{M}_1$. Therefore, a locally aligned mesh can be constructed as outlined in Section 3.3.3.

3.8 Kronecker system matrices

$$\textit{Setting: } \boldsymbol{b} = (b_1, b_2)^\top \equiv \text{const}$$

For constant $\boldsymbol{b}$, the process of assembling the system matrices of Sections 3.5.2, 3.6.3 and 3.6.5 can be simplified using Kronecker matrix products [76, Section 4.2]. This section provides a brief overview of the definition and a selection of properties and exemplarily constructs the system matrices of the primal variational form discussed in Section 3.5 using the locally aligned mesh of Section 3.3.3.

Definition 3.14. *See [76, p.243, 4.2.1]. Let $A = \left(a_{ij}\right)_{ij} \in \mathbb{C}^{k \times l}$, $B \in \mathbb{C}^{m \times n}$, then the Kronecker product of A and B is defined as $A \otimes B \in \mathbb{C}^{km \times ln}$ with*

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1l}B \\ a_{21}B & a_{22}B & \dots & a_{2l}B \\ \vdots & & \ddots & \vdots \\ a_{k1}B & a_{k2}B & \dots & a_{kl}B \end{pmatrix} . \tag{3.114}$$

Theorem 3.15. *Let $A, C \in \mathbb{C}^{k\times l}$, $B, D \in \mathbb{C}^{m\times n}$ then*

$$(A \otimes B)^\top = A^\top \otimes B^\top \tag{3.115}$$
$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1} \tag{3.116}$$
$$(A \otimes B)(C \otimes D) = AC \otimes BD \tag{3.117}$$

Proof:
See [76, p.243, 4.2.4], [76, p.244, 4.2.11] and [76, p.251, 14.(a,b)] respectively.

■

Theorem 3.16. *Let $A \in \mathbb{C}^{m\times m}$ have eigenvalues $(\lambda_i)_i$ and $B \in \mathbb{C}^{n\times n}$ have eigenvalues $(\mu_j)_j$. Then the eigenvalues of $A \otimes B$ are given by $(\lambda_i \mu_j)_{ij}$.*

Proof:
See [76, p.245, 4.2.12].

■

Theorem 3.17. *Let $A, C \in \mathbb{C}^{m\times m}$ and $B, D \in \mathbb{C}^{n\times n}$. Assume A and C diagonalize in the same basis. Let $(\lambda_i)_i, (\mu_i)_i$ be the eigenvalues of A, C and the spectrum of a matrix A be denoted by $\sigma(A)$. Then*

$$\sigma(A \otimes B + C \otimes D) = \bigcup_i \sigma(\lambda_i B + \mu_i D) \ . \tag{3.118}$$

Proof:
As A and C diagonalize in the same basis, there exists a transform matrix T such that

$$T^{-1}AT = \operatorname{diag}(\lambda_i)_i \quad , \quad T^{-1}CT = \operatorname{diag}(\mu_i)_i \tag{3.119}$$

where $\operatorname{diag}(v)$ is the diagonal matrix with main diagonal v. Then

$$(T \otimes \mathrm{Id}_n)^{-1}(A \otimes B + C \otimes D)(T \otimes \mathrm{Id}_n) = \operatorname{diag}(\lambda_i)_i \otimes B + \operatorname{diag}(\mu_i)_i \otimes D \ . \tag{3.120}$$

The basis transform using $T \otimes \mathrm{Id}_n$ leaves the spectrum of $A \otimes B + C \otimes D$ invariant where Id_n is the $n \times n$ identity matrix. $\operatorname{diag}(\lambda_i)_i \otimes B + \operatorname{diag}(\mu_i)_i \otimes D$ is a block diagonal matrix with blocks $\lambda_i \otimes B + \mu_i \otimes D$. This finishes the proof as the spectrum of a block diagonal matrix is the union of spectra of each block.

■

For constructing the system matrices of the primal variational form introduced in Section 3.5, we first note that the basis functions can be chosen identical up to translation. We then introduce a unique cell index for each cell $K_1, \ldots, K_{N_\Sigma}$ and assemble the system matrices as shown in

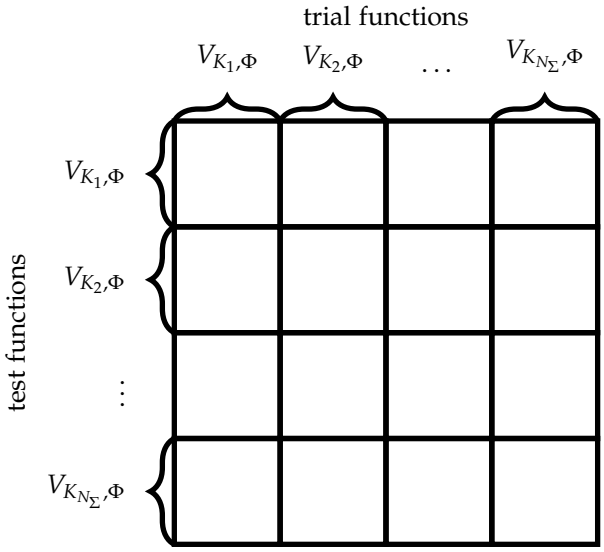


FIGURE 3.12: *Occupation block structure of a system matrix.*

Figure 3.12. This block matrix is sparse since the block (i, j) is 0 whenever K_i and K_j share no interface as the intersection of the supports of the respective test and trial functions is empty in this case.

We assemble the matrices by reducing the global variational form (3.39) to the local variational form (3.34) by using a test function $\psi^K \in V_{K,\Phi}$. As the same basis for each cell up to translation is chosen and the neighbouring structure of each cell is the same which is ensured by the usage of the locally aligned mesh, the local variational form (3.34) is representative for building each row of blocks in Figure 3.12. (3.34) is split in main and neighbouring cell contributions by separating integrals with ϕ^K from those with $\phi^{N_F(K)}$. The main cell contributions containing the volume integral and all boundary integrals with ϕ^K build a block diagonal substructure in the system matrix. Abbreviating the block containing these parts by A_M, the main cell contribution matrix is given by

$$\mathrm{Id}_{N_y} \otimes \mathrm{Id}_{N_x} \otimes A_M \tag{3.121}$$

using the Kronecker matrix product where Id_N is the $N \times N$ identity matrix.

We are left with the arrangement of blocks belonging to the neighbours of each cell. This assembly is summed up by introducing an appropriate enumeration of cells. Starting with the bottom left cell, we assign an index for the column and the row of the cell as shown for example in Figure 3.13.

Picking a cell with index (k, l), the indices of the neighbouring cells can now be calculated. Figure 3.14 shows the change of indices for the general case of possibly sheared meshes. We define the vertical shift constant c and the horizontal shift constant c_H as

$$c := \left\lfloor \frac{b_2}{b_1} \frac{N_y}{N_x} \right\rfloor \quad , \quad c_H := \left\lfloor \frac{b_1}{b_2} \frac{N_x}{N_y} \right\rfloor \tag{3.122}$$

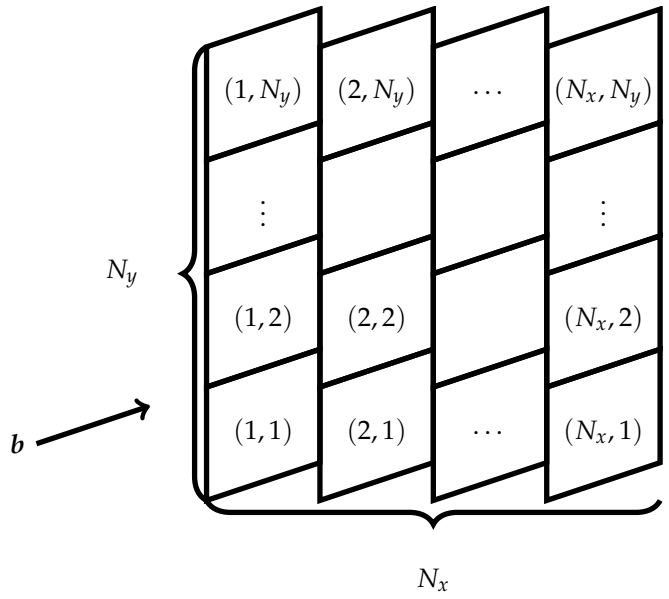
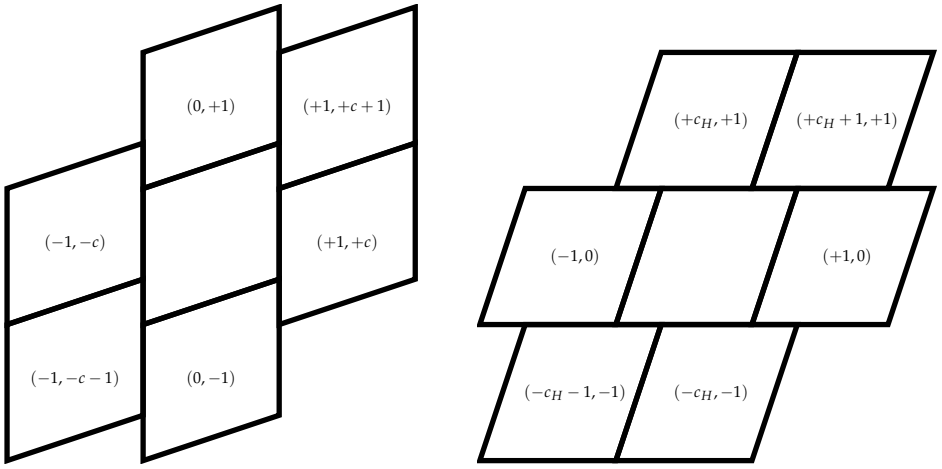


FIGURE 3.13: *Locally aligned $N_x \times N_y$ mesh with enumerated cells.*



(a) Aligned upper and bottom interfaces (b) Aligned left and right interfaces

FIGURE 3.14: *Change of indices of neighbouring cells respective to the index of a reference cell up to periodicity treatment.*

where $\lfloor . \rfloor$ are lower Gauss brackets. However, whenever the index (k,l) of a neighbouring cell fulfills one of the properties $k,l < 1$, $k > N_x$, $l > N_y$, we have to adapt this index due to the periodic boundary. Tables 3.1 and 3.2 summarize how to account for the periodicity.

upper, bottom interface aligned		
adaption	k (neighbour)	l (neighbour)
$(-1,-c-1)$	$\mathrm{mod}\,(k-2,\ N_x)+1$	$\mathrm{mod}\,\big(l-c-2,\ N_y\big)+1$
$(0,-1)$	k	$\mathrm{mod}\,\big(l-2,\ N_y\big)+1$
$(-1,-c)$	$\mathrm{mod}\,(k-2,\ N_x)+1$	$\mathrm{mod}\,\big(l-c-1,\ N_y\big)+1$
$(+1,+c)$	$\mathrm{mod}\,(k,\ N_x)+1$	$\mathrm{mod}\,\big(l+c-1,\ N_y\big)+1$
$(0,+1)$	k	$\mathrm{mod}\,\big(l,\ N_y\big)+1$
$(+1,+c+1)$	$\mathrm{mod}\,(k,\ N_x)+1$	$\mathrm{mod}\,\big(l+c,\ N_y\big)+1$

TABLE 3.1: *Indices of the neighbours of cell (k,l) for aligned upper and bottom interfaces.*

left, right interface aligned		
adaption	k (neighbour)	l (neighbour)
$(-c_H-1,-1)$	$\mathrm{mod}\,(k-c_H-2,\ N_x)+1$	$\mathrm{mod}\,\big(l-2,\ N_y\big)+1$
$(-c_H,-1)$	$\mathrm{mod}\,(k-c_H-1,\ N_x)+1$	$\mathrm{mod}\,\big(l-2,\ N_y\big)+1$
$(-1,0)$	$\mathrm{mod}\,(k-2,\ N_x)+1$	l
$(+1,0)$	$\mathrm{mod}\,(k,\ N_x)+1$	l
$(+c_H,+1)$	$\mathrm{mod}\,(k+c_H-1,\ N_x)+1$	$\mathrm{mod}\,\big(l,\ N_y\big)+1$
$(+c_H+1,+1)$	$\mathrm{mod}\,(k+c_H,\ N_x)+1$	$\mathrm{mod}\,\big(l,\ N_y\big)+1$

TABLE 3.2: *Indices of the neighbours of cell (k,l) for aligned left and right interfaces.*

Using this information about the neighbouring structure, we can now construct the left hand side system matrix A of (3.41). We use the definition and properties of circulant matrices [78]. Using the locally aligned mesh with aligned upper and bottom interfaces, defining the circulant matrix

$$P_N \coloneqq \begin{pmatrix} 0 & 1 & 0 & \dots & & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & 1 & 0 \\ 0 & 0 & & 0 & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{N\times N}\,, \tag{3.123}$$

and A_B, A_U, A_{LB},A_{LU}, A_{RB},A_{RU} as the matrices containing the bottom, upper, left bottom, left upper, right bottom, right upper interface contributions of the boundary integrals in (3.34), A

writes

$$\begin{aligned} A = &\mathrm{Id}_{N_y} \otimes \mathrm{Id}_{N_x} \otimes A_M + P_{N_y}^{-1} \otimes \mathrm{Id}_{N_x} \otimes A_B + P_{N_y} \otimes \mathrm{Id}_{N_x} \otimes A_U \\ &+ P_{N_y}^{-c-1} \otimes P_{N_x}^{-1} \otimes A_{LB} + P_{N_y}^{-c} \otimes P_{N_x}^{-1} \otimes A_{LU} + P_{N_y}^{c} \otimes P_{N_x} \otimes A_{RB} + P_{N_y}^{c+1} \otimes P_{N_x} \otimes A_{RU} \, . \end{aligned} \tag{3.124}$$

For aligned left and right interfaces, we define A_{BL}, A_{BR}, A_L, A_R, A_{UL}, A_{UR} as the matrices containing the bottom left, bottom right, left, right, upper left, upper right interface contributions of the boundary integrals in (3.34). A is given by

$$\begin{aligned} A = &\mathrm{Id}_{N_y} \otimes \mathrm{Id}_{N_x} \otimes A_M + P_{N_y}^{-1} \otimes P_{N_x}^{-c_H-1} \otimes A_{BL} + P_{N_y}^{-1} \otimes P_{N_x}^{-c_H} \otimes A_{BR} \\ &+ \mathrm{Id}_{N_y} \otimes P_{N_x}^{-1} \otimes A_L + \mathrm{Id}_{N_y} \otimes P_{N_x} \otimes A_R + P_{N_y} \otimes P_{N_x}^{c_H} \otimes A_{UL} + P_{N_y} \otimes P_{N_x}^{c_H+1} \otimes A_{UR} \, . \end{aligned} \tag{3.125}$$

We note that $P_N^0 = \mathrm{Id}_N$. The right hand side system matrix B of (3.41) uses the mass matrix M given by the mass integral on the right hand side of (3.34). It can be written as

$$B = \mathrm{Id}_{N_y} \otimes \mathrm{Id}_{N_x} \otimes M \, . \tag{3.126}$$

3.9 Asymptotic formulae

$$\textit{Setting: } \boldsymbol{b} = (b_1, b_2)^\top \equiv \text{const}$$

For constant $\boldsymbol{b}$, Section 3.8 shows that the system matrices for the primal variational form introduced in Section 3.5 using the locally aligned mesh can be represented by Kronecker products (3.124), (3.125) involving circulant matrices. In this case, this section shows that the effort for calculating eigenvalues can be reduced further to systems of the size of the underlying basis of $V_{K,\Phi}$. Furthermore, we present asymptotic expansions for the eigenvalues of the discrete system in the number of cells for the choices of different meshes and a two-element basis.
We first investigate the spectrum of circulant matrices and observe that they diagonalize in the discrete Fourier basis [78, Section 3.1].

Definition 3.18. *ω_N is defined as the N^{th} root of unity given by*

$$\omega_N := \exp\left(\frac{2\pi \mathrm{i}}{N}\right) \, . \tag{3.127}$$

F_N is defined as the $N \times N$ discrete Fourier matrix given by

$$F_N := \frac{1}{\sqrt{N}} \begin{pmatrix} \omega_N^{0\cdot 0} & \omega_N^{0\cdot 1} & \dots & \omega_N^{0\cdot(N-1)} \\ \omega_N^{1\cdot 0} & \omega_N^{1\cdot 1} & \dots & \omega_N^{1\cdot(N-1)} \\ \vdots & & \ddots & \vdots \\ \omega_N^{(N-1)\cdot 0} & \omega_N^{(N-1)\cdot 1} & \dots & \omega_N^{(N-1)\cdot(N-1)} \end{pmatrix} \, . \tag{3.128}$$

Theorem 3.19. *Let $k \in \mathbb{Z}$ and $N \in \mathbb{N}$. The eigenpairs of P_N^k with P_N defined by* (3.123) *are given by $\left(\omega_N^{k\cdot(l-1)}, (F_N)_l\right)$ where $(F_N)_l$ is the l^{th} column of F_N.*

Proof:
See [78, Section 3.1].

■

Corollary 3.20. $F_N^{-1} P_N^k F_N = \operatorname{diag}\left(1, \omega_N^k, \ldots, \omega_N^{(N-1)k}\right)$ *where* $\operatorname{diag}(\boldsymbol{v})$ *is the diagonal matrix with main diagonal* $\boldsymbol{v}$.

Proof:
Using Theorem 3.19 we have

$$P_N^k F_N = F_N \operatorname{diag}\left(1, \omega_N^k, \ldots, \omega_N^{(N-1)k}\right) . \tag{3.129}$$

Multiplication with F_N^{-1} concludes the proof.

■

In the following sections, we apply this information to obtain the desired asymptotic expansions for different discretizations. Section 3.9.1 explores locally aligned meshes, Section 3.9.2 explores cartesian meshes and Section 3.9.3 explores fully aligned meshes for special choices of $\boldsymbol{b}$.

3.9.1 Locally aligned mesh

Dealing with the locally aligned mesh, we transform the system matrix of (3.43) to obtain

Corollary 3.21. *Let $\sigma_{m,n}$ be the spectrum of*

$$\begin{aligned} M^{-\frac{1}{2}} \Big(& A_M + \omega_{N_y}^{-n} A_B + \omega_{N_y}^{n} A_U + \omega_{N_y}^{-(c+1)n} \omega_{N_x}^{-m} A_{LB} \\ & + \omega_{N_y}^{-cn} \omega_{N_x}^{-m} A_{LU} + \omega_{N_y}^{cn} \omega_{N_x}^{m} A_{RB} + \omega_{N_y}^{(c+1)n} \omega_{N_x}^{m} A_{RU} \Big) M^{-\frac{1}{2}} \end{aligned} \tag{3.130}$$

for aligned bottom and upper interfaces and

$$\begin{aligned} M^{-\frac{1}{2}} \Big(& A_M + \omega_{N_y}^{-n} \omega_{N_x}^{-(c_H+1)m} A_{BL} + \omega_{N_y}^{-n} \omega_{N_x}^{-c_H m} A_{BR} + \omega_{N_x}^{-m} A_L \\ & + \omega_{N_x}^{m} A_R + \omega_{N_y}^{n} \omega_{N_x}^{c_H m} A_{UL} + \omega_{N_y}^{n} \omega_{N_x}^{(c_H+1)m} A_{UR} \Big) M^{-\frac{1}{2}} \end{aligned} \tag{3.131}$$

for aligned left and right interfaces. The spectrum of $B^{-\frac{1}{2}} A B^{-\frac{1}{2}}$, namely the system matrix of (3.43) *using a locally aligned mesh, with A respectively defined by* (3.124) *or* (3.125) *and B defined by* (3.126)*, is given by*

$$\bigcup_{m=0}^{N_y-1} \bigcup_{n=0}^{N_x-1} \sigma_{m,n} . \tag{3.132}$$

Proof:
We transform $B^{-\frac{1}{2}}AB^{-\frac{1}{2}}$ using the transform matrix $F_{N_y} \otimes F_{N_x} \otimes \mathrm{Id}$ and obtain

$$\left(F_{N_y} \otimes F_{N_x} \otimes \mathrm{Id}\right)^{-1} \left(B^{-\frac{1}{2}}AB^{-\frac{1}{2}}\right) \left(F_{N_y} \otimes F_{N_x} \otimes \mathrm{Id}\right) . \tag{3.133}$$

Corollary 3.20 shows that P_N^k diagonalizes in the same basis for all powers k. Furthermore $P_{N_y}^k \otimes P_{N_x}^l$ diagonalize in the same basis, namely using the transform matrix $F_{N_y} \otimes F_{N_x}$, for all powers k, l. The respective eigenvalues are given by combination of Theorem 3.16 with Corollary 3.20 as $\left(\omega_{N_y}^{mk}\omega_{N_x}^{nl}\right)_{mn}$. Subsequent application of Theorem 3.17 then provides the desired result.

■

Corollary 3.21 shows, that the eigenvalues of (3.43) can be retrieved using those of (3.130) or (3.131) respectively. This is a reduction of the size of the matrices by a factor of the number of cells N_Σ. We present asymptotic expansions for these eigenvalues for $N_x, N_y \to \infty$ when choosing a basis with degree 0 in x-direction and 1 in y-direction for aligned upper and bottom interfaces and degree 1 in x-direction and 0 in y-direction for aligned left and right interfaces. Then, matrices (3.130) and (3.131) are in $\mathbb{R}^{2\times 2}$. The exact eigenvalues are given by Theorem 3.1. For (3.130), this writes

$$(b_1 m + b_2 n)^2 \, \eta \frac{N_y}{N_x} + \mathcal{O}\left(N_x^k N_y^l\right) \tag{3.134}$$

and for (3.131) we obtain

$$(b_1 m + b_2 n)^2 \, \eta \frac{N_x}{N_y} + \mathcal{O}\left(N_x^k N_y^l\right) \tag{3.135}$$

for $k, l \in \mathbb{Z}, k + l \leqslant -2$. Choosing $\eta = \frac{N_x}{N_y}$ for aligned upper and bottom interfaces and $\eta = \frac{N_y}{N_x}$ for aligned left and right interfaces, we obtain for (3.130) the asymptotic expansion

$$\begin{aligned}&(b_1 m + b_2 n)^2 - \frac{\pi^2}{3 b_1 N_x^2} m \left(b_1^3 m^3 + 4 b_1^2 b_2 m^2 n + 6 b_1 b_2^2 m n^2 + 4 b_2^3 n^3\right) \\ &- \frac{2\pi^2}{3 b_1 N_x N_y} b_2^3 \left(1 + 2c\right) n^4 + \frac{\pi^2}{3 N_y^2} b_2^2 \left(1 + 6c + 6c^2\right) n^4 + \mathcal{O}\left(N_x^k N_y^l\right)\end{aligned} \tag{3.136}$$

and for (3.131) this yields

$$\begin{aligned}&(b_1 m + b_2 n)^2 - \frac{\pi^2}{3 b_2 N_y^2} n \left(4 b_1^3 m^3 + 6 b_1^2 b_2 m^2 n + 4 b_1 b_2^2 m n^2 + b_2^3 n^3\right) \\ &- \frac{2\pi^2}{3 b_2 N_x N_y} b_1^3 \left(1 + 2c_H\right) m^4 + \frac{\pi^2}{3 N_x^2} b_1^2 \left(1 + 6c_H + 6c_H^2\right) m^4 + \mathcal{O}\left(N_x^k N_y^l\right)\end{aligned} \tag{3.137}$$

with c and c_H defined in (3.122) and $k, l \in \mathbb{Z}, k + l \leqslant -2, k < -2 \vee l < -2$. We remark that these particular choices for η are necessary whenever choosing a basis which resolves just one direction.

3.9.2 Cartesian mesh

Using a cartesian mesh presented in Section 3.3.1, the neighbouring structure simplifies in comparison to Figure 3.14 which is shown in Figure 3.15. The respective indices are summarized in Table 3.3.

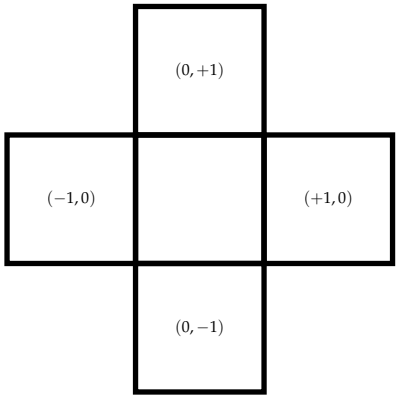


FIGURE 3.15: *Change of indices of neighbouring cells respective to the index of a reference cell up to periodicity treatment for the cartesian mesh in Figure 3.4(b).*

cartesian mesh		
adaption	k (neighbour)	l (neighbour)
$(-1, 0)$	$\mathrm{mod}\,(k-2,\ N_x)+1$	l
$(+1, 0)$	$\mathrm{mod}\,(k,\ N_x)+1$	l
$(0, -1)$	k	$\mathrm{mod}\,(l-2,\ N_y)+1$
$(0, +1)$	k	$\mathrm{mod}\,(l,\ N_y)+1$

TABLE 3.3: *Indices of the neighbours of cell (k, l) for the cartesian mesh in Figure 3.4(b).*

The left hand side system matrix in (3.41) then writes

$$\begin{aligned} A = &\mathrm{Id}_{N_y} \otimes \mathrm{Id}_{N_x} \otimes A_M + \mathrm{Id}_{N_y} \otimes P_{N_x}^{-1} \otimes A_L \\ &+ \mathrm{Id}_{N_y} \otimes P_{N_x} \otimes A_R + P_{N_y}^{-1} \otimes \mathrm{Id}_{N_x} \otimes A_B + P_{N_y} \otimes \mathrm{Id}_{N_x} \otimes A_U \end{aligned} \tag{3.138}$$

with the respective left, right, bottom, upper interface contribution matrices A_L, A_R, A_B, A_U. The right hand side system matrix stays the same as there are no neighbouring cell contributions. Transforming this system to a standard eigenvalue problem, we can structurally reduce the system matrix as in Section 3.9.1.

Corollary 3.22. *Let $\sigma_{m,n}$ be the spectrum of*

$$M^{-\frac{1}{2}}\left(A_M + \omega_{N_x}^{-m} A_L + \omega_{N_x}^{m} A_R + \omega_{N_y}^{-n} A_B + \omega_{N_y}^{n} A_U\right) M^{-\frac{1}{2}} \,. \tag{3.139}$$

The spectrum of $B^{-\frac{1}{2}} A B^{-\frac{1}{2}}$, namely the system matrix of (3.43) *using a cartesian mesh, with A defined by* (3.138) *and B defined by* (3.126)*, is given by*

$$\bigcup_{m=0}^{N_y-1} \bigcup_{n=0}^{N_x-1} \sigma_{m,n} \,. \tag{3.140}$$

Proof:
The result can be obtained using the same construction as in Corollary 3.21.

■

Using the same basis as at the end of Section 3.9.1 for aligned upper and bottom interfaces, i.e., a basis with degree 0 in x-direction and 1 in y-direction, the asymptotic expansion of (3.139) for $N_x, N_y \to \infty$ yields

$$2b_1 b_2 mn + b_2^2 n^2 + b_1^2 m^2 \eta \frac{N_y}{N_x} - b_1^2 m^2 \left(\frac{N_y}{\eta N_x} + \frac{N_y^2}{\eta^2 N_x^2} + \frac{N_y^3}{\eta^3 N_x^3} + \dots\right) + \mathcal{O}\left(N_x^k N_y^l\right) \tag{3.141}$$

with $k, l \in \mathbb{Z}, k + l \leqslant -2$. We obtain

$$\eta \frac{N_y}{N_x} - \sum_{i=1}^{\infty} \left(\frac{N_y}{\eta N_x}\right)^i \overset{!}{=} 1 \tag{3.142}$$

which holds for $\eta = \frac{N_x^2 + N_y^2}{N_x N_y}$ where $\frac{N_y}{\eta N_x} < 1$ holds in particular. Using this η, we obtain the asymptotic expansion for the eigenvalues

$$(b_1 m + b_2 n)^2 - \frac{\pi^2}{3N_x^2} b_1 m \left(4b_2 n \left(m^2 + n^2\right) + b_1 \left(m^3 - 2mn^2\right)\right) + \frac{\pi^2}{3N_y^2} b_2^2 n^4 + \mathcal{O}\left(N_x^k N_y^l\right) \tag{3.143}$$

with $k, l \in \mathbb{Z}, k + l \leqslant -2, k < -2 \vee l < -2$.
Comparing this expansion with (3.136) and (3.137), no structural advantage of the locally aligned mesh is evident in this representation. For a full comparison, all parts of the asymptotic expansions with exponents $k + l = -2$ of $N_x^k N_y^l$ have to be considered. Furthermore, bases of higher degrees should be implemented. However, the calculation of asymptotic expansions of analytical eigenvalues of a 3×3 matrix is a complex undertaking. Therefore, we resort this discussion to the comparison of numerical results in Section 6.2.1 for bases with higher degrees.

3.9.3 Fully aligned mesh

This section considers the fully aligned mesh presented in Figure 3.7 for the case of conforming boundary interfaces. This is the case whenever

$$c_{\mathrm{UB}} = \frac{b_2}{b_1} N_y \in \mathbb{N} \qquad , \qquad c_{\mathrm{LR}} = \frac{b_1}{b_2} N_x \in \mathbb{N} \tag{3.144}$$

for aligned upper and bottom or left and right interfaces respectively as previously mentioned in (3.14). In these cases, the neighbouring structure of each cell is the same as for the cartesian mesh presented in Figure 3.15 except for cells at the periodic boundary. Hence, the indices presented in Table 3.3 have to be adapted for the left- and rightmost cells in the case of aligned upper and bottom interfaces and the upper- and bottommost cells in the case of aligned left and right interfaces. These results are summarized in Table 3.4 and Table 3.5.

fully aligned mesh (upper, bottom interfaces aligned)		
adaption	k (neighbour)	l (neighbour)
$(-1,0)\,, k \neq 1, N_x$	$\mathrm{mod}\,(k-2,\ N_x)+1$	l
$(+1,0)\,, k \neq 1, N_x$	$\mathrm{mod}\,(k,\ N_x)+1$	l
$(-1,0)\,, k = 1$	N_x	$\mathrm{mod}\,\big(l-c_{\mathrm{UB}}-1,\ N_y\big)+1$
$(+1,0)\,, k = N_x$	1	$\mathrm{mod}\,\big(l+c_{\mathrm{UB}}-1,\ N_y\big)+1$
$(0,-1)$	k	$\mathrm{mod}\,\big(l-2,\ N_y\big)+1$
$(0,+1)$	k	$\mathrm{mod}\,\big(l,\ N_y\big)+1$

TABLE 3.4: *Indices of the neighbours of cell (k,l) for the fully aligned mesh in Figure 3.7 with aligned upper and bottom interfaces.*

fully aligned mesh (left, right interfaces aligned)		
adaption	k (neighbour)	l (neighbour)
$(-1,0)$	$\mathrm{mod}\,(k-2,\ N_x)+1$	l
$(+1,0)$	$\mathrm{mod}\,(k,\ N_x)+1$	l
$(0,-1)\,, l \neq 1, N_y$	k	$\mathrm{mod}\,\big(l-2,\ N_y\big)+1$
$(0,+1)\,, l \neq 1, N_y$	k	$\mathrm{mod}\,\big(l,\ N_y\big)+1$
$(0,-1)\,, l = 1$	$\mathrm{mod}\,(k-c_{\mathrm{LR}}-1,\ N_x)+1$	N_y
$(0,+1)\,, l = N_y$	$\mathrm{mod}\,(k+c_{\mathrm{LR}}-1,\ N_x)+1$	1

TABLE 3.5: *Indices of the neighbours of cell (k,l) for the fully aligned mesh in Figure 3.7 with aligned left and right interfaces.*

Defining the matrices

$$D_N = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \in \mathbb{R}^{N\times N} \quad , \quad S_N = \begin{pmatrix} 0 & \dots & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & & & 0 \\ 1 & 0 & \dots & 0 \end{pmatrix} \in \mathbb{R}^{N\times N}\,, \tag{3.145}$$

the Kronecker matrix structure presented in Section 3.8 can then be applied to build the system matrices of (3.41) which write

$$\begin{aligned} A = &\operatorname{Id}_{N_y} \otimes \operatorname{Id}_{N_x} \otimes A_M + \left(\operatorname{Id}_{N_y} \otimes D_{N_x}^\top + P_{N_y}^{-c_{\mathrm{UB}}} \otimes S_{N_x}^\top\right) \otimes A_L \\ &+ \left(\operatorname{Id}_{N_y} \otimes D_{N_x} + P_{N_y}^{c_{\mathrm{UB}}} \otimes S_{N_x}\right) \otimes A_R + P_{N_y}^{-1} \otimes \operatorname{Id}_{N_x} \otimes A_B + P_{N_y} \otimes \operatorname{Id}_{N_x} \otimes A_U \end{aligned} \tag{3.146}$$

for aligned upper and bottom interfaces and

$$\begin{aligned} A = &\operatorname{Id}_{N_y} \otimes \operatorname{Id}_{N_x} \otimes A_M + \operatorname{Id}_{N_y} \otimes P_{N_x}^{-1} \otimes A_L + \operatorname{Id}_{N_y} \otimes P_{N_x} \otimes A_R \\ &+ \left(D_{N_y}^\top \otimes \operatorname{Id}_{N_x} + S_{N_x}^\top \otimes P_{N_x}^{-c_{\mathrm{LR}}} \otimes\right) \otimes A_B + \left(D_{N_y} \otimes \operatorname{Id}_{N_x} + S_{N_x} \otimes P_{N_x}^{c_{\mathrm{LR}}} \otimes\right) \otimes A_U \end{aligned} \tag{3.147}$$

for aligned left and right interfaces where A_L, A_R, A_B, A_U denote the left, right, bottom, upper interface contribution matrices respectively. The right hand side system matrix stays the same as there are no neighbouring contributions. Reduction to a standard eigenvalue problem using the matrix root of the right hand side system matrix yields

Theorem 3.23. *Let $\sigma_{m,n}$ be the spectrum of*

$$M^{-\frac{1}{2}} \left(A_M + \omega_{N_y}^{-n\frac{c_{UB}}{N_x}} \omega_{N_x}^{-m} A_L + \omega_{N_y}^{n\frac{c_{UB}}{N_x}} \omega_{N_x}^{m} A_R + \omega_{N_y}^{n} A_B + \omega_{N_y}^{-n} A_U \right) M^{-\frac{1}{2}} \tag{3.148}$$

for aligned bottom and upper interfaces and

$$M^{-\frac{1}{2}} \left(A_M + \omega_{N_x}^{m} A_L + \omega_{N_x}^{-m} A_R + \omega_{N_y}^{-n} \omega_{N_x}^{-m\frac{c_{LR}}{N_y}} A_B + \omega_{N_y}^{n} \omega_{N_x}^{m\frac{c_{LR}}{N_y}} A_U \right) M^{-\frac{1}{2}} \tag{3.149}$$

for aligned left and right interfaces. The spectrum of $B^{-\frac{1}{2}} A B^{-\frac{1}{2}}$, namely the system matrix of (3.43) *using a fully aligned mesh, with A respectively defined by* (3.146) *or* (3.147) *and B defined by* (3.126)*, is given by*

$$\bigcup_{m=0}^{N_y-1} \bigcup_{n=0}^{N_x-1} \sigma_{m,n} \ . \tag{3.150}$$

Proof:
This proof is closely related to the proof of Corollary 3.21. We focus on the case of aligned upper and bottom interfaces as the the same methodology can be applied for the case of aligned left and right interfaces.
First, transforming $B^{-\frac{1}{2}} A B^{-\frac{1}{2}}$ using the matrix $F_{N_y} \otimes \operatorname{Id}_{N_x} \otimes \operatorname{Id}$ and Corollary 3.20 we obtain a block diagonal matrix with blocks

$$\begin{aligned} &\operatorname{Id}_{N_x} \otimes A_M + \left(D_{N_x}^\top + \omega_{N_y}^{nc_{\mathrm{UB}}} S_{N_x}^\top\right) \otimes A_L + \left(D_{N_x} + \omega_{N_y}^{-nc_{\mathrm{UB}}} \otimes S_{N_x}\right) \otimes A_R \\ &+ \omega_{N_y}^{n} \operatorname{Id}_{N_x} \otimes A_B + \omega_{N_y}^{-n} \operatorname{Id}_{N_x} \otimes A_U \end{aligned} \tag{3.151}$$

for $n = 1, \dots, N_y$. Note that the exponents of ω_{N_y} were reordered for standardizing the notation. We observe

$$\left(D_{N_x}^{\top} + \omega_{N_y}^{nc_{\mathrm{UB}}} S_{N_x}^{\top}\right)^{-1} = \left(D_{N_x} + \omega_{N_y}^{-nc_{\mathrm{UB}}} S_{N_x}\right) \tag{3.152}$$

such that both diagonalize in the same basis. The eigenpairs of $\left(D_{N_x} + \omega_{N_y}^{-nc_{\mathrm{UB}}} S_{N_x}\right)$ are given by

$$\left(\omega_{N_x}^{-m}\omega_{N_y}^{-n\frac{c_{\mathrm{UB}}}{N_x}}, \begin{pmatrix} \left(\omega_{N_x}^{m}\omega_{N_y}^{n\frac{c_{\mathrm{UB}}}{N_x}}\right)^{N_x-1} \\ \vdots \\ \left(\omega_{N_x}^{m}\omega_{N_y}^{n\frac{c_{\mathrm{UB}}}{N_x}}\right) \\ 1 \end{pmatrix}\right) \tag{3.153}$$

reproducible by straightforward calculation. The eigenvalues of the inverse matrix are the inverse eigenvalues. Denoting the matrix of eigenvectors by $F_{N_x}^{A}$ and using that all first matrices in the Kronecker products of (3.151) diagonalize in this basis, we can now use the transform matrix $F_{N_y} \otimes F_{N_x}^{A} \otimes \mathrm{Id}$ on $B^{-\frac{1}{2}} A B^{-\frac{1}{2}}$ and successively apply Theorem 3.17 to obtain the desired result.

■

We present asymptotic expansions for the eigenvalues of (3.148) and (3.149) for $N_x, N_y \to \infty$, using the same bases presented at the end of Section 3.9.1. Again, we choose $\eta = \frac{N_x}{N_y}$ for aligned upper and bottom interfaces and $\eta = \frac{N_y}{N_x}$ for aligned left and right interfaces. For (3.148) we obtain

$$(b_1 m + b_2 n)^2 + \frac{\pi^2}{3 b_1^2 N_x^2} (b_1 m + b_2 n)^4 + \mathcal{O}\left(N_x^k N_y^l\right) \tag{3.154}$$

and for (3.149) this yields

$$(b_1 m + b_2 n)^2 + \frac{\pi^2}{3 b_2^2 N_y^2} (b_1 m + b_2 n)^4 + \mathcal{O}\left(N_x^k N_y^l\right) \tag{3.155}$$

for $k, l \in \mathbb{Z}, k + l \leqslant -4$.

In comparison to (3.136) and (3.137), we see that the lowest order error terms of (3.154) and (3.155) scale with the size of the eigenvalue. Hence, the smaller the eigenvalue, the better the approximation using the fully aligned mesh. However, this is only possible whenever (3.144) is fulfilled.

Chapter 4

A 4^{TH}-ORDER EQUATION Building methods (Part II)

In this chapter, we construct a numerical scheme for discretizing the 4^{th}-order equation

$$-\nabla\cdot(\boldsymbol{b}\nabla\cdot(\boldsymbol{b}_\perp\boldsymbol{b}_\perp\cdot\nabla(\boldsymbol{b}\cdot\nabla\phi))) = \omega^2\nabla\cdot(\boldsymbol{b}_\perp\boldsymbol{b}_\perp\cdot\nabla\phi)\ , \qquad \text{in } \Omega \tag{4.1}$$

derived in Section 2.5 in (2.71) in the fully periodic domain $\Omega = [0,2\pi)^2$. The left hand side operator is of order 4 and the right hand side operator of order 2. The analysis of this chapter is inspired by and tied to the findings of Chapter 3. Despite (4.1) being derived for constant $\boldsymbol{b}$, we introduce a mixed variational form of this equation for general $\boldsymbol{b}(\boldsymbol{x})$. We remark that the generalized perpendicular direction $\boldsymbol{b}_\perp(\boldsymbol{x})$ is defined up to a sign in two dimensions which poses no difficulties as it cancels in (4.1).

The outline of this chapter is as follows: Section 4.1 explores the analytical solution of (4.1) for constant $\boldsymbol{b}$ and establishes similarities to the exact solution of the constant coefficient anisotropic wave equation discussed in Section 3.1. Section 4.2 constructs a discontinuous Galerkin method discretizing (4.1).

4.1 Analytical solution

$$\textit{Setting: } \boldsymbol{b} = (b_1, b_2)^\top \equiv \text{const}$$

A variant of Theorem 3.1 presented in Section 3.1 holds for (4.1).

Theorem 4.1. *The functions*

$$\phi_{m,n}(x,y) = \exp(\mathrm{i}(mx+ny)), \qquad m,n\in\mathbb{Z},\ x,y\in\Omega \tag{4.2}$$

are analytic eigenfunctions of (4.1) *with corresponding eigenvalues* $\omega_{m,n}^2 = (b_1m+b_2n)^2$ *whenever* $\boldsymbol{b}_\perp\cdot(m,n)^\top \neq 0$. *The parallel gradient of each eigenfunction fulfills*

$$\|\boldsymbol{b}\cdot\nabla\phi_{m,n}(x,y)\|_{L^2_2(\Omega)}^2 \leqslant 4\pi^2\omega_{m,n}^2\ . \tag{4.3}$$

Proof:
For $\boldsymbol{v} = (v_1, v_2)^\top \in \mathbb{C}^2$ and $c \in \mathbb{C}$ constant, it holds

$$\boldsymbol{v} \cdot \nabla (c\phi_{m,n}) = \mathrm{i}c\,(v_1 m + v_2 n)\,\phi_{m,n} = \nabla \cdot (\boldsymbol{v} c \phi_{m,n}) \ . \tag{4.4}$$

Insertion in (4.1) yields

$$\begin{aligned} &- \nabla \cdot (\boldsymbol{b} \nabla \cdot (\boldsymbol{b}_\perp \boldsymbol{b}_\perp \cdot \nabla (\boldsymbol{b} \cdot \nabla \phi_{m,n}))) \\ &= -\nabla \cdot (\boldsymbol{b} \nabla \cdot (\boldsymbol{b}_\perp \boldsymbol{b}_\perp \cdot \nabla (\mathrm{i}(b_1 m + b_2 n)\phi_{m,n}))) \\ &= -\nabla \cdot \left(\boldsymbol{b} \nabla \cdot \left(\boldsymbol{b}_\perp \mathrm{i}^2 (-b_2 m + b_1 n)(b_1 m + b_2 n)\phi_{m,n}\right)\right) \\ &= -\nabla \cdot \left(\boldsymbol{b} \mathrm{i}^3 (-b_2 m + b_1 n)^2 (b_1 m + b_2 n)\phi_{m,n}\right) \\ &= -\mathrm{i}^4 (-b_2 m + b_1 n)^2 (b_1 m + b_2 n)^2 \phi_{m,n} \end{aligned} \tag{4.5}$$

for the left hand side and

$$\nabla \cdot (\boldsymbol{b}_\perp \boldsymbol{b}_\perp \cdot \nabla \phi_{m,n}) = \nabla \cdot (\boldsymbol{b}_\perp \mathrm{i}(-b_2 m + b_1 n)\phi_{m,n}) = \mathrm{i}^2 (-b_2 m + b_1 n)^2 \phi_{m,n} \tag{4.6}$$

for the right hand side operator. Together this yields

$$\begin{aligned} -\mathrm{i}^4 (-b_2 m + b_1 n)^2 (b_1 m + b_2 n)^2 \phi_{m,n} &= \omega^2 \mathrm{i}^2 (-b_2 m + b_1 n)^2 \phi_{m,n} \\ (b_1 m + b_2 n)^2 \phi_{m,n} &= \omega^2 \phi_{m,n} \end{aligned} \tag{4.7}$$

for $(-b_2 m + b_1 n) = \boldsymbol{b}_\perp \cdot (m, n)^\top \neq 0$. We remark that for $(-b_2 m + b_1 n) = 0$ no eigenvalue ω can be associated to the problem.
The inequality holds as shown in Theorem 3.1.

■

As of Theorem 4.1, we expect difficulties for the approximation of eigenfunctions $\phi_{m,n}$ with mode numbers m, n such that $\boldsymbol{b}_\perp \cdot (m, n)^\top = 0$ which is particularly the case for the constant mode $\phi_{0,0}$. However, besides the constant mode, the results of Theorem 3.1 can be recovered whenever we consider eigenvalues $\omega_{m,n}^2 < 1$ as

$$\left\| \begin{pmatrix} m \\ n \end{pmatrix} \right\|_2 = \left\| \left(\boldsymbol{b} \cdot \begin{pmatrix} m \\ n \end{pmatrix} + \boldsymbol{b}_\perp \cdot \begin{pmatrix} m \\ n \end{pmatrix} \right) \begin{pmatrix} m \\ n \end{pmatrix} \right\| \leqslant \left(\left| \boldsymbol{b} \cdot \begin{pmatrix} m \\ n \end{pmatrix} \right| + \left| \boldsymbol{b}_\perp \cdot \begin{pmatrix} m \\ n \end{pmatrix} \right| \right) \left\| \begin{pmatrix} m \\ n \end{pmatrix} \right\|_2 \tag{4.8}$$

yields

$$1 - \left| \boldsymbol{b} \cdot \begin{pmatrix} m \\ n \end{pmatrix} \right| \leqslant \left| \boldsymbol{b}_\perp \cdot \begin{pmatrix} m \\ n \end{pmatrix} \right| \tag{4.9}$$

for $(m, n) \neq (0, 0)$. Hence, $\omega_{m,n}^2 < 1$ yields $\boldsymbol{b}_\perp \cdot (m, n)^\top \neq 0$. In this case, the theory developed for solving the anisotropic wave equation (3.1) can be extended for (4.1). We point out that the eigenvalue problem (4.1) is ill-posed with the same proof as Lemma 3.4.

4.2 Construction of a mixed form matrix system

Referring to Sections 3.4, 3.5 and 3.6, the goal of this section is the construction of a symmetric and consistent mixed variational form of (4.1) involving differentials of at most first order.
The primal variational form reads: Find pairs $\left(\phi, \omega^2\right) \in V_\Phi \times \mathbb{R}$ such that

$$\begin{aligned} &\sum_{K\in\mathcal{K}} \left(\int_K -\nabla\cdot\left(\boldsymbol{b}\nabla\cdot\left(\boldsymbol{b}_\perp\boldsymbol{b}_\perp\cdot\nabla\left(\boldsymbol{b}\cdot\nabla\phi\right)\right)\right)\psi \,\mathrm{d}\boldsymbol{x}\right) \\ &= \omega^2 \sum_{K\in\mathcal{K}} \left(\int_K \nabla\cdot\left(\boldsymbol{b}_\perp\boldsymbol{b}_\perp\cdot\nabla\phi\right)\psi \,\mathrm{d}\boldsymbol{x}\right) \quad \forall\psi\in V_\Phi\,. \end{aligned} \tag{4.10}$$

The element-local form writes

$$\begin{aligned} &\int_K -\nabla\cdot\left(\boldsymbol{b}\nabla\cdot\left(\boldsymbol{b}_\perp\boldsymbol{b}_\perp\cdot\nabla\left(\boldsymbol{b}\cdot\nabla\phi^K\right)\right)\right)\psi^K \,\mathrm{d}\boldsymbol{x} \\ &= \omega^2 \int_K \nabla\cdot\left(\boldsymbol{b}_\perp\boldsymbol{b}_\perp\cdot\nabla\phi^K\right)\psi^K \,\mathrm{d}\boldsymbol{x} \quad \forall\psi^K\in V_{K,\Phi}, K\in\mathcal{K} \end{aligned} \tag{4.11}$$

for the locally defined function spaces $V_{K,\Phi}$. The global and local function spaces are connected by $V_\Phi = \bigcup_{K\in\mathcal{K}} V_{K,\Phi}$. Again, we simplify the notation by omitting the superscript of functions $\phi^K, \psi^K \in V_{K,\Phi}$ whenever there is no differentiation between multiple cells.
Sections 4.2.1 and 4.2.3 derive a mixed variational form of the 4^{th}-order left hand side and the 2^{nd}-order right hand side operator respectively whereas Sections 4.2.2 and 4.2.4 construct the associated system matrices. The final system is built in Section 4.2.5 and the overall construction process is summarized.

4.2.1 4^{th}-order operator mixed form

To improve readability, we omit the definition of test spaces for the variational formulations and until the summary of the final system.
Using the differential degree as a lower index, we propose the mixed form

$$\begin{cases} u_1 & = \boldsymbol{b}\cdot\nabla\phi \\ u_2 & = \boldsymbol{b}_\perp\cdot\nabla u_1 \\ u_3 & = \nabla\cdot\left(\boldsymbol{b}_\perp u_2\right) \\ -\nabla\cdot\left(\boldsymbol{b}u_3\right) & = \mathrm{RHS} \end{cases} \tag{4.12}$$

where the right hand side equations of Section 4.2.3 are abbreviated by RHS, with corresponding variational forms

$$\begin{cases} -\int_K u_1 v_1 \,\mathrm{d}\boldsymbol{x} & +\int_K \boldsymbol{b}\cdot\nabla\phi v_1 \,\mathrm{d}\boldsymbol{x} & = 0 \\ \int_K u_2 v_2 \,\mathrm{d}\boldsymbol{x} & -\int_K \boldsymbol{b}_\perp\cdot\nabla u_1 v_2 \,\mathrm{d}\boldsymbol{x} & = 0 \\ -\int_K u_3 v_3 \,\mathrm{d}\boldsymbol{x} & +\int_K \nabla\cdot(\boldsymbol{b}_\perp u_2)\, v_3 \,\mathrm{d}\boldsymbol{x} & = 0 \\ -\int_K \nabla\cdot(\boldsymbol{b} u_3)\,\psi \,\mathrm{d}\boldsymbol{x} & & = \mathrm{RHS} \end{cases} \tag{4.13}$$

We now need to introduce boundary integrals for enabling the communication with neighbouring cells. Therefore, we integrate the first and second equation of (4.13) by parts twice. The first integration by parts introduces a new numerical flux for each equation which we define as $\widehat{\phi}$ and $\widehat{u}_1$. The second integration by parts is carried out using the respective inner cell function. Furthermore, we integrate the third and fourth equation of (4.13) by parts once, introducing a new numerical flux for each equation which we define as $\widehat{u}_2$ and $\widehat{u}_3$. Defining the local test spaces $V_{K,U_1}, V_{K,U_2}, V_{K,U_3}$, the full system is then given $\forall K \in \mathcal{K}$ by

$$\begin{cases} -\int_K u_1 v_1 \,\mathrm{d}\boldsymbol{x} + \int_K \boldsymbol{b}\cdot\nabla\phi v_1 \,\mathrm{d}\boldsymbol{x} \\ \qquad + \sum_{F\in\partial K}\left(\int_F (\widehat{\phi}-\phi)\, v_1 \boldsymbol{b}\cdot\boldsymbol{n}\,\mathrm{d}S(\boldsymbol{x})\right) & = 0 & \forall v_1 \in V_{K,U_1} \\ \int_K u_2 v_2 \,\mathrm{d}\boldsymbol{x} - \int_K \boldsymbol{b}_\perp\cdot\nabla u_1 v_2 \,\mathrm{d}\boldsymbol{x} \\ \qquad - \sum_{F\in\partial K}\left(\int_F (\widehat{u}_1-u_1)\, v_2 \boldsymbol{b}_\perp\cdot\boldsymbol{n}\,\mathrm{d}S(\boldsymbol{x})\right) & = 0 & \forall v_2 \in V_{K,U_2} \\ -\int_K u_3 v_3 \,\mathrm{d}\boldsymbol{x} + \int_K u_2 \boldsymbol{b}_\perp\cdot\nabla v_3 \,\mathrm{d}\boldsymbol{x} \\ \qquad + \sum_{F\in\partial K}\left(\int_F \widehat{u}_2 v_3 \boldsymbol{b}_\perp\cdot\boldsymbol{n}\,\mathrm{d}S(\boldsymbol{x})\right) & = 0 & \forall v_3 \in V_{K,U_3} \\ \int_K u_3\boldsymbol{b}\cdot\nabla\psi \,\mathrm{d}\boldsymbol{x} - \sum_{F\in\partial K}\left(\int_F \widehat{u}_3\psi\boldsymbol{b}\cdot\boldsymbol{n}\,\mathrm{d}S(\boldsymbol{x})\right) & = \mathrm{RHS} & \forall \psi \in V_{K,\Phi} \end{cases} \tag{4.14}$$

In summary, we are now dealing with four fluxes $\widehat{u}_1, \widehat{u}_2, \widehat{u}_3$ and $\widehat{\phi}$. As in Section 3.6.2, we choose fluxes inspired by local discontinuous Galerkin fluxes such that

$$\widehat{u}_1 = \{\!\{u_1\}\!\} - \frac{\eta_2}{h_F}\boldsymbol{b}_\perp\cdot[\![u_2]\!] \quad , \quad \widehat{u}_2 = \{\!\{u_2\}\!\} \tag{4.15}$$

$$\widehat{u}_3 = \{\!\{u_3\}\!\} - \frac{\eta_\phi}{h_F}\boldsymbol{b}\cdot[\![\phi]\!] \quad , \quad \widehat{\phi} = \{\!\{\phi\}\!\} \tag{4.16}$$

where η_2, η_ϕ are constant stabilization parameters and h_F is a parameter dependent on the length of the cell edge with interface F. Flux jumps and averages are defined in (3.25). Note, that omitting

the stabilization in u_2 by just using average fluxes for u_1 is beneficial for reducing the size of the system matrices as outlined in Section 4.2.5.
We return to the global variational forms by summing local contributions in (4.14) over all elements. Using the global test spaces $V_{U_j} = \bigcup_{K\in\mathcal{K}} V_{K,U_j}$, $j = 1,2,3$, Lemmata 3.8 and 3.9, the system then writes $\forall \psi \in V_\Phi, v_j \in V_{U_j}, j = 1,2,3$

$$
\begin{cases}
-\sum_{K\in\mathcal{K}} \left(\int_K u_1 v_1 \,\mathrm{d}\boldsymbol{x} \right) & + \sum_{K\in\mathcal{K}} \left(\int_K \boldsymbol{b}\cdot\nabla\phi v_1 \,\mathrm{d}\boldsymbol{x} \right) \\
\quad - \sum_{F\in\mathcal{F}} \left(\int_F \boldsymbol{b}\cdot[\![\phi]\!]\{\!\{v_1\}\!\} \,\mathrm{d}S(\boldsymbol{x}) \right) & = 0 \\
\quad \sum_{K\in\mathcal{K}} \left(\int_K u_2 v_2 \,\mathrm{d}\boldsymbol{x} \right) & - \sum_{K\in\mathcal{K}} \left(\int_K \boldsymbol{b}_\perp\cdot\nabla u_1 v_2 \,\mathrm{d}\boldsymbol{x} \right) \\
\quad + \sum_{F\in\partial K} \left(\int_F \boldsymbol{b}_\perp\cdot[\![u_1]\!]\{\!\{v_2\}\!\} \,\mathrm{d}S(\boldsymbol{x}) \right) & \\
\quad + \sum_{F\in\partial K} \left(\int_F \frac{\eta_2}{h_F}\boldsymbol{b}_\perp\cdot[\![u_2]\!]\boldsymbol{b}_\perp\cdot[\![v_2]\!] \,\mathrm{d}S(\boldsymbol{x}) \right) & = 0 \\
-\sum_{K\in\mathcal{K}} \left(\int_K u_3 v_3 \,\mathrm{d}\boldsymbol{x} \right) & - \sum_{K\in\mathcal{K}} \left(\int_K u_2\boldsymbol{b}_\perp\cdot\nabla v_3 \,\mathrm{d}\boldsymbol{x} \right) \\
\quad + \sum_{F\in\partial K} \left(\int_F \{\!\{u_2\}\!\}\boldsymbol{b}_\perp\cdot[\![v_3]\!] \,\mathrm{d}S(\boldsymbol{x}) \right) & = 0 \\
\quad \sum_{K\in\mathcal{K}} \left(\int_K u_3\boldsymbol{b}\cdot\nabla\psi \,\mathrm{d}\boldsymbol{x} \right) & - \sum_{F\in\mathcal{F}} \left(\int_F \{\!\{u_3\}\!\}\boldsymbol{b}\cdot[\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x}) \right) \\
\quad + \sum_{F\in\mathcal{F}} \left(\int_F \frac{\eta_\phi}{h_F}\boldsymbol{b}\cdot[\![\phi]\!]\boldsymbol{b}\cdot[\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x}) \right) & = \mathrm{RHS}
\end{cases}
\tag{4.17}
$$

The method is consistent as for a continuous solution $(\phi_0, u_{1,0}, u_{2,0}, u_{3,0}, \omega_0^2)$ it holds $\{\!\{\varphi_0\}\!\} = \varphi_0$ and $[\![\varphi_0]\!] = 0$ for all $\varphi_0 \in \{\phi_0, u_{1,0}, u_{2,0}, u_{3,0}\}$. Insertion in (4.17) and one integration by parts of the second and fourth equation directly yields the variational form of (4.12).

4.2.2 4^{th}-order operator system matrix

The setup of system matrices associated to the variational form (4.17) is closely related to Section 3.6.3. We are now dealing with a system matrix with four coefficient vectors $\boldsymbol{U}_j$, $j = 1,2,3$ and $\boldsymbol{\Phi}$ given by the discretization of u_j, $j = 1,2,3$ and ϕ using the bases $\left(u_{j,k}^K\right)_k$ of V_{K,U_j} and $\left(\phi_k^K\right)_k$ of $V_{K,\Phi}$ for all K such that

$$
u_j(\boldsymbol{x}) = \sum_{K\in\mathcal{K}} \sum_k U_{j,k}^K u_{j,k}^K(\boldsymbol{x}) \qquad , \qquad U_{j,k}^K \in \mathbb{R} \qquad j = 1,2,3 \tag{4.18}
$$

$$
\phi(\boldsymbol{x}) = \sum_{K\in\mathcal{K}} \sum_k \Phi_k^K \phi_k^K(\boldsymbol{x}) \qquad , \qquad \Phi_k^K \in \mathbb{R} \tag{4.19}
$$

For $\boldsymbol{b} \equiv \text{const}$, the blocks of this system matrix can be built as Kronecker matrix products as outlined in Section 3.8 using the respective spaces for test and trial functions and local cell contributions summarized in (4.14).
From a global standpoint, we associate the integrals of system (4.17) to block matrices as follows. The volume terms are given by

$$M_{U_1V_1} \leftrightarrow \sum_{K\in\mathcal{K}} \int_K u_1 v_1 \,\mathrm{d}\boldsymbol{x} \tag{4.20}$$

$$A_{\Phi V_1} \leftrightarrow \sum_{K\in\mathcal{K}} \int_K \boldsymbol{b}\cdot\nabla\phi v_1 \,\mathrm{d}\boldsymbol{x} \tag{4.21}$$

$$M_{U_2V_2} \leftrightarrow \sum_{K\in\mathcal{K}} \int_K u_2 v_2 \,\mathrm{d}\boldsymbol{x} \tag{4.22}$$

$$A_{U_1V_2} \leftrightarrow \sum_{K\in\mathcal{K}} \int_K \boldsymbol{b}_\perp\cdot\nabla u_1 v_2 \,\mathrm{d}\boldsymbol{x} \tag{4.23}$$

$$M_{U_3V_3} \leftrightarrow \sum_{K\in\mathcal{K}} \int_K u_3 v_3 \,\mathrm{d}\boldsymbol{x} \tag{4.24}$$

$$A_{U_2V_3} \leftrightarrow \sum_{K\in\mathcal{K}} \int_K u_2\boldsymbol{b}_\perp\cdot\nabla v_3 \,\mathrm{d}\boldsymbol{x} \tag{4.25}$$

$$A_{U_3\Psi} \leftrightarrow \int_K u_3\boldsymbol{b}\cdot\nabla\psi \,\mathrm{d}\boldsymbol{x} \tag{4.26}$$

whereas the boundary terms are

$$B_{\Phi V_1} \leftrightarrow \sum_{F\in\mathcal{F}} \int_F \boldsymbol{b}\cdot[\![\phi]\!]\{\!\{v_1\}\!\} \,\mathrm{d}S(\boldsymbol{x}) \tag{4.27}$$

$$B_{U_1V_2} \leftrightarrow \sum_{F\in\mathcal{F}} \int_F \boldsymbol{b}_\perp\cdot[\![u_1]\!]\{\!\{v_2\}\!\} \,\mathrm{d}S(\boldsymbol{x}) \tag{4.28}$$

$$B_{U_3\Psi} \leftrightarrow \sum_{F\in\mathcal{F}} \int_F \{\!\{u_3\}\!\}\boldsymbol{b}\cdot[\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x}) \tag{4.29}$$

$$B_{U_2V_3} \leftrightarrow \sum_{F\in\mathcal{F}} \int_F \{\!\{u_2\}\!\}\boldsymbol{b}_\perp\cdot[\![v_3]\!] \,\mathrm{d}S(\boldsymbol{x}) \tag{4.30}$$

$$B_{U_2V_2} \leftrightarrow \sum_{F\in\mathcal{F}} \int_F \frac{\eta_2}{h_F}\boldsymbol{b}_\perp\cdot[\![u_2]\!]\boldsymbol{b}_\perp\cdot[\![v_2]\!] \,\mathrm{d}S(\boldsymbol{x}) \tag{4.31}$$

$$B_{\Phi\Psi} \leftrightarrow \sum_{F\in\mathcal{F}} \int_F \frac{\eta_\phi}{h_F}\boldsymbol{b}\cdot[\![\phi]\!]\boldsymbol{b}\cdot[\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x}) \tag{4.32}$$

The left hand side system matrix then writes with the coefficient vectors

$$\begin{pmatrix} -M_{U_1V_1} & 0 & 0 & A_{\Phi V_1}-B_{\Phi V_1} \\ -A_{U_1V_2}+B_{U_1V_2} & M_{U_2V_2}+B_{U_2V_2} & 0 & 0 \\ 0 & -A_{U_2V_3}+B_{U_2V_3} & -M_{U_3V_3} & 0 \\ 0 & 0 & A_{U_3\Psi}-B_{U_3\Psi} & B_{\Phi\Psi} \end{pmatrix} \begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \\ \boldsymbol{U}_3 \\ \boldsymbol{\Phi} \end{pmatrix} . \tag{4.33}$$

We exchange the ordering of test functions, namely change v_1 and v_3, to obtain

$$\begin{pmatrix} 0 & -A_{U_2V_3}+B_{U_2V_3} & -M_{U_3V_3} & 0 \\ -A_{U_1V_2}+B_{U_1V_2} & M_{U_2V_2}+B_{U_2V_2} & 0 & 0 \\ -M_{U_1V_1} & 0 & 0 & A_{\Phi V_1}-B_{\Phi V_1} \\ 0 & 0 & A_{U_3\Psi}-B_{U_3\Psi} & B_{\Phi\Psi} \end{pmatrix} \begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \\ \boldsymbol{U}_3 \\ \boldsymbol{\Phi} \end{pmatrix} \tag{4.34}$$

which allows us to prove the symmetry of the system matrix.

Theorem 4.2. *If the respective spaces for test and trial functions coincide and besides $V_{U_1} = V_{U_3}$, then the system matrix* (4.34) *is symmetric.*

Proof:
When choosing the same basis for test and trial functions for each variable, it is obvious that the mass matrices $M_{U_jV_j}, j = 1,2,3$ and the penalization matrices $B_{U_2V_2}, B_{\Phi\Psi}$ are symmetric. Furthermore it holds $A^\top_{U_1V_2} = A_{U_2V_3}$ as well as $B^\top_{U_1V_2} = B_{U_2V_3}$ as u_2, v_2 and u_1, v_3 are interchangeable between (4.23) and (4.25) as well as (4.28) and (4.30) and $V_{U_1} = V_{U_3}$. It holds $M^\top_{U_1V_1} = M_{U_3V_3}$ as $V_{U_1} = V_{U_3}$. Further, it holds $A^\top_{U_3\Psi} = A_{\Phi V_1}$ as well as $B^\top_{U_3\Psi} = B_{\Phi V_1}$ as ϕ, ψ and u_3, v_1 are interchangeable between (4.26) and (4.21) as well as (4.29) and (4.27) and $V_{U_1} = V_{U_3}$.

■

4.2.3 2nd-order operator mixed form

The right hand side operator of (4.1) is structurally equivalent to the anisotropic wave problem (3.1) up to a sign with tensor $\boldsymbol{b}_\perp \boldsymbol{b}_\perp{}^\top$. We discretize it using the mixed variational form of Section 3.6.1 with local discontinuous Galerkin fluxes presented in Section 3.6.2 and summarize the results here.
Using the substitution variable u_4 where the lower index is no longer related to the degree of differentiation but a continuation of the previous indexing, the equation system for the right hand side is defined as

$$\begin{cases} \omega^2 u_4 & = \omega^2 \boldsymbol{b}_\perp \cdot \nabla\phi \\ \text{LHS} & = \omega^2 \nabla \cdot (\boldsymbol{b}_\perp u_4) \end{cases} \tag{4.35}$$

with $u_4 \in V_{U_4} = \bigcup_{K\in\mathcal{K}} V_{K,U_4}$, left hand side equations of Section 4.2.1 abbreviated by LHS and associated local variational form $\forall v_4 \in V_{K,U_4}, \psi \in V_{K,\Phi}, K \in \mathcal{K}$

$$\begin{cases} 0 = \omega^2 \left(\displaystyle\int_K u_4 v_4 \,\mathrm{d}\boldsymbol{x} - \int_K \boldsymbol{b}\cdot\nabla\phi v_4 \,\mathrm{d}\boldsymbol{x} - \sum_{F\in\partial K} \left(\int_F \left(\widehat{\phi} - \phi\right) v_4 \boldsymbol{b}_\perp \cdot \boldsymbol{n} \,\mathrm{d}S(\boldsymbol{x}) \right) \right) \\ \text{LHS} = \omega^2 \left(- \displaystyle\int_K u_4 \boldsymbol{b}_\perp \cdot \nabla\psi \,\mathrm{d}\boldsymbol{x} + \sum_{F\in\partial K} \left(\int_F \widehat{u}_4 \psi \boldsymbol{b}_\perp \cdot \boldsymbol{n} \,\mathrm{d}S(\boldsymbol{x}) \right) \right) \end{cases} \tag{4.36}$$

where the fluxes are given by

$$\widehat{u}_4 = \{\!\{u_4\}\!\} - \frac{\eta_\mathrm{E}}{h_F} \boldsymbol{b}_\perp \cdot [\![\phi]\!] \qquad , \qquad \widehat{\phi} = \{\!\{\phi\}\!\} \tag{4.37}$$

with stabilization constant η_E with subscript E for clarifying that it belongs to the eigenvalue part of (4.1). Flux jumps and averages are defined in (3.25). Using Lemma 3.9, the global variational forms of the first equation of (4.36) is given by

$$0 = \omega^2 \left(\sum_{K\in\mathcal{K}} \left(\int_K u_4^K v_4^K \,\mathrm{d}\boldsymbol{x} - \int_K \boldsymbol{b}_\perp \cdot \nabla\phi^K v_4^K \,\mathrm{d}\boldsymbol{x} \right) + \sum_{F\in\mathcal{F}} \left(\int_F \boldsymbol{b}_\perp \cdot [\![\phi]\!] \{\!\{v_4\}\!\} \,\mathrm{d}S(\boldsymbol{x}) \right) \right) \tag{4.38}$$

whereas the second equation of (4.36) using Lemma 3.8 yields

$$\begin{aligned}\mathrm{LHS} = \omega^2 \Bigg(&- \sum_{K\in\mathcal{K}} \left(\int_K u_4^K \boldsymbol{b}_\perp \cdot \nabla \psi^K \,\mathrm{d}\boldsymbol{x} \right) + \sum_{F\in\mathcal{F}} \left(\int_F \{\!\{ u_4 \}\!\} \boldsymbol{b}_\perp \cdot [\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x}) \right) \\ &- \sum_{F\in\mathcal{F}} \left(\int_F \frac{\eta_\mathrm{E}}{h_F} \boldsymbol{b}_\perp \cdot [\![\phi]\!] \boldsymbol{b}_\perp \cdot [\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x}) \right) \Bigg) .\end{aligned} \tag{4.39}$$

4.2.4 2nd-order operator system matrix

The setup of the left hand side system matrix associated to (4.38) and (4.39) is done in the same way as outlined in Section 3.6.3 using the coefficient vectors $\boldsymbol{U}_4$ and $\boldsymbol{\Phi}$ given by the discretization of u_4 and ϕ with the bases $\left(u_{4,k}^K\right)_k$ of V_{K,U_4} and $\left(\phi_k^K\right)_k$ of $V_{K,\Phi}$ for all K such that

$$u_4(\boldsymbol{x}) = \sum_{K\in\mathcal{K}} \sum_k U_{4,k}^K u_{4,k}^K(\boldsymbol{x}) \qquad , \qquad U_{4,k}^K \in \mathbb{R} \tag{4.40}$$

$$\phi(\boldsymbol{x}) = \sum_{K\in\mathcal{K}} \sum_k \Phi_k^K \phi_k^K(\boldsymbol{x}) \qquad , \qquad \Phi_k^K \in \mathbb{R} \tag{4.41}$$

The matrix block associations are given by formulas (4.38) and (4.39) such that

$$M_{U_4V_4}^\mathrm{E} \leftrightarrow \sum_{K\in\mathcal{K}} \int_K u_4^K v_4^K \,\mathrm{d}\boldsymbol{x} \tag{4.42}$$

$$A_{\Phi V_4}^\mathrm{E} \leftrightarrow \sum_{K\in\mathcal{K}} \int_K \boldsymbol{b}_\perp \cdot \nabla \phi^K v_4^K \,\mathrm{d}\boldsymbol{x} \tag{4.43}$$

$$B_{\Phi V_4}^\mathrm{E} \leftrightarrow \sum_{F\in\mathcal{F}} \int_F \boldsymbol{b}_\perp \cdot [\![\phi]\!] \{\!\{ v_4 \}\!\} \,\mathrm{d}S(\boldsymbol{x}) \tag{4.44}$$

$$A_{U_4\Psi}^\mathrm{E} \leftrightarrow \sum_{K\in\mathcal{K}} \int_K u_4^K \boldsymbol{b}_\perp \cdot \nabla \psi^K \,\mathrm{d}\boldsymbol{x} \tag{4.45}$$

$$B_{U_4\Psi}^\mathrm{E} \leftrightarrow \sum_{F\in\mathcal{F}} \int_F \{\!\{ u_4 \}\!\} \boldsymbol{b}_\perp \cdot [\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x}) \tag{4.46}$$

$$B_{\Phi\Psi}^\mathrm{E} \leftrightarrow \sum_{F\in\mathcal{F}} \int_F \frac{\eta_\mathrm{E}}{h_F} \boldsymbol{b}_\perp \cdot [\![\phi]\!] \boldsymbol{b}_\perp \cdot [\![\psi]\!] \,\mathrm{d}S(\boldsymbol{x}) \tag{4.47}$$

with superscript E for avoiding double definitions with the left hand side system from which we obtain the right hand side system matrix

$$\begin{pmatrix} 0 \\ \mathrm{LHS} \end{pmatrix} = \omega^2 \begin{pmatrix} M_{U_4V_4}^\mathrm{E} & -A_{\Phi V_4}^\mathrm{E} + B_{\Phi V_4}^\mathrm{E} \\ -A_{U_4\Psi}^\mathrm{E} + B_{U_4\Psi}^\mathrm{E} & -B_{\Phi\Psi}^\mathrm{E} \end{pmatrix} \begin{pmatrix} \boldsymbol{U}_4 \\ \boldsymbol{\Phi} \end{pmatrix} \tag{4.48}$$

which is symmetric if the respective spaces for test and trial functions coincide due to Theorem 3.11.

4.2.5 System matrices, reduction and summary

We summarize the results of Sections 4.2.2 and 4.2.4 to construct the generalized eigenvalue problem and discuss conditions for reduction to a generalized eigenvalue problem of smaller size. Using the system matrices given in (4.34) and (4.48), the generalized eigenvalue problem is given

by

$$\begin{aligned}
&\begin{pmatrix}
0 & -A_{U_2V_3}+B_{U_2V_3} & -M_{U_3V_3} & 0 & 0\\
-A_{U_1V_2}+B_{U_1V_2} & M_{U_2V_2}+B_{U_2V_2} & 0 & 0 & 0\\
-M_{U_1V_1} & 0 & 0 & 0 & A_{\Phi V_1}-B_{\Phi V_1}\\
0 & 0 & 0 & 0 & 0\\
0 & 0 & A_{U_3\Psi}-B_{U_3\Psi} & 0 & B_{\Phi\Psi}
\end{pmatrix}
\begin{pmatrix}
\boldsymbol{U}_1\\ \boldsymbol{U}_2\\ \boldsymbol{U}_3\\ \boldsymbol{U}_4\\ \boldsymbol{\Phi}
\end{pmatrix}\\
&= \omega^2
\begin{pmatrix}
0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & M^{\mathrm{E}}_{U_4V_4} & -A^{\mathrm{E}}_{\Phi V_4}+B^{\mathrm{E}}_{\Phi V_4}\\
0 & 0 & 0 & -A^{\mathrm{E}}_{U_4\Psi}+B^{\mathrm{E}}_{U_4\Psi} & -B^{\mathrm{E}}_{\Phi\Psi}
\end{pmatrix}
\begin{pmatrix}
\boldsymbol{U}_1\\ \boldsymbol{U}_2\\ \boldsymbol{U}_3\\ \boldsymbol{U}_4\\ \boldsymbol{\Phi}
\end{pmatrix}
\end{aligned} \tag{4.49}$$

with symmetric left and right hand side system matrices. The right hand side system is reduced analogously to the end of Section 3.6.3 while neglecting the inverse root of the mass matrix. Relating to (3.70), we define

$$A^{\mathrm{E}} := \left(A^{\mathrm{E}}_{U_4\Psi}-B^{\mathrm{E}}_{U_4\Psi}\right)\left(M^{\mathrm{E}}_{U_4V_4}\right)^{-1}\left(A^{\mathrm{E}}_{\Phi V_4}-B^{\mathrm{E}}_{\Phi V_4}\right)+B^{\mathrm{E}}_{\Phi\Psi} \tag{4.50}$$

such that the system writes

$$\begin{aligned}
&\begin{pmatrix}
0 & -A_{U_2V_3}+B_{U_2V_3} & -M_{U_3V_3} & 0\\
-A_{U_1V_2}+B_{U_1V_2} & M_{U_2V_2}+B_{U_2V_2} & 0 & 0\\
-M_{U_1V_1} & 0 & 0 & A_{\Phi V_1}-B_{\Phi V_1}\\
0 & 0 & A_{U_3\Psi}-B_{U_3\Psi} & B_{\Phi\Psi}
\end{pmatrix}
\begin{pmatrix}
\boldsymbol{U}_1\\ \boldsymbol{U}_2\\ \boldsymbol{U}_3\\ \boldsymbol{\Phi}
\end{pmatrix}\\
&= \omega^2
\begin{pmatrix}
0 & 0 & 0 & 0\\
0 & 0 & 0 & 0\\
0 & 0 & 0 & 0\\
0 & 0 & 0 & -A^{\mathrm{E}}
\end{pmatrix}
\begin{pmatrix}
\boldsymbol{U}_1\\ \boldsymbol{U}_2\\ \boldsymbol{U}_3\\ \boldsymbol{\Phi}
\end{pmatrix}.
\end{aligned} \tag{4.51}$$

The left hand side system is reduced by solving for the coefficient vectors $\boldsymbol{U}_j$, $j=1,2,3$. We retrieve the equations

$$\boldsymbol{U}_1 = M^{-1}_{U_1V_1}\left(A_{\Phi V_1}-B_{\Phi V_1}\right)\boldsymbol{\Phi} \tag{4.52}$$

$$\boldsymbol{U}_2 = \left(M_{U_2V_2}+B_{U_2V_2}\right)^{-1}\left(A_{U_1V_2}-B_{U_1V_2}\right)\boldsymbol{U}_1 \tag{4.53}$$

$$\boldsymbol{U}_3 = -M^{-1}_{U_3V_3}\left(A_{U_2V_3}-B_{U_2V_3}\right)\boldsymbol{U}_2 \tag{4.54}$$

from which we obtain the reduced generalized eigenvalue problem

$$A\boldsymbol{\Phi} = \omega^2 A^{\mathrm{E}}\boldsymbol{\Phi} \tag{4.55}$$

with the reduced left hand side system matrix

$$\begin{aligned}
A := &\left(A_{U_3\Psi}-B_{U_3\Psi}\right)M^{-1}_{U_3V_3}\left(A_{U_2V_3}-B_{U_2V_3}\right)\left(M_{U_2V_2}+B_{U_2V_2}\right)^{-1}\\
&\left(A_{U_1V_2}-B_{U_1V_2}\right)M^{-1}_{U_1V_1}\left(A_{\Phi V_1}-B_{\Phi V_1}\right)-B_{\Phi\Psi}\,.
\end{aligned} \tag{4.56}$$

Corollary 4.3. *If the respective spaces for test and trial functions coincide and besides $V_{U_1} = V_{U_3}$, then $A = A^\top$.*

Proof:
This follows directly from the symmetry of (4.34) given by Theorem 4.2 as

$$\begin{aligned} A^\top =& \Big(\left(A_{U_3\Psi} - B_{U_3\Psi}\right) M_{U_3V_3}^{-1} \left(A_{U_2V_3} - B_{U_2V_3}\right) \left(M_{U_2V_2} + B_{U_2V_2}\right)^{-1} \\ & \left(A_{U_1V_2} - B_{U_1V_2}\right) M_{U_1V_1}^{-1} \left(A_{\Phi V_1} - B_{\Phi V_1}\right) \Big)^\top - B_{\Phi\Psi}^\top \\ =& \left(A_{\Phi V_1} - B_{\Phi V_1}\right)^\top \left(M_{U_1V_1}^{-1}\right)^\top \left(A_{U_1V_2} - B_{U_1V_2}\right)^\top \left(\left(M_{U_2V_2} + B_{U_2V_2}\right)^{-1}\right)^\top \\ & \left(A_{U_2V_3} - B_{U_2V_3}\right)^\top \left(M_{U_3V_3}^{-1}\right)^\top \left(A_{U_3\Psi} - B_{U_3\Psi}\right) - B_{\Phi\Psi}^\top \\ =& \left(A_{U_3\Psi} - B_{U_3\Psi}\right) M_{U_3V_3}^{-1} \left(A_{U_2V_3} - B_{U_2V_3}\right) \left(M_{U_2V_2} + B_{U_2V_2}\right)^{-1} \\ & \left(A_{U_1V_2} - B_{U_1V_2}\right) M_{U_1V_1}^{-1} \left(A_{\Phi V_1} - B_{\Phi V_1}\right) - B_{\Phi\Psi} = A \ . \end{aligned} \tag{4.57}$$

■

However, the inversion of $M_{U_2V_2} + B_{U_2V_2}$ is costly as it is not a block diagonal matrix. In this case, the multiplication with the inverse of $M_{U_2V_2} + B_{U_2V_2}$ further couples neighbouring cells resulting in a system matrix of lower sparsity. This can be circumvented by setting $\eta_2 = 0$, i.e., impose average fluxes for u_1 and u_2, such that the fluxes in Section 4.2.1 write

$$\widehat{u}_1 = \{\!\{u_1\}\!\} \qquad , \qquad \widehat{u}_1 = \{\!\{u_2\}\!\} \ . \tag{4.58}$$

Then, $B_{U_2V_2} = 0$ and the inversions in (4.56) are carried out for block diagonal mass matrices $M_{U_jV_j}, j = 1,2,3$ which is numerically efficient.
We further remark that the huge nullspace of the right hand side system matrix in (4.51) poses difficulties for several solvers of generalized eigenvalue problems and prevents further simplifications to a standard eigenvalue problem. Reducing the system to (4.55) leaves us with a nullspace of dimension 1 for the right hand side system matrix A^{E} for most choices of $\boldsymbol{b}$ as shown in Section 3.4. We summarize the mixed form equations

$$\begin{cases} u_1 & = \boldsymbol{b} \cdot \nabla\phi \\ u_2 & = \boldsymbol{b}_\perp \cdot \nabla u_1 \\ u_3 & = \nabla \cdot (\boldsymbol{b}_\perp u_2) \\ -\nabla \cdot (\boldsymbol{b} u_3) & = \omega^2 \nabla \cdot (\boldsymbol{b}_\perp u_4) \\ \omega^2 u_4 & = \omega^2 \boldsymbol{b}_\perp \cdot \nabla\phi \end{cases} \tag{4.59}$$

and propose to use average fluxes for u_1, u_2 given in (4.58), local discontinuous Galerkin fluxes for u_3, ϕ given in (4.16) and for u_4, ϕ given in (4.37) for the setup of the reduced generalized eigenvalue problem (4.55).

Chapter 5

IMPLEMENTATION
Building code

This chapter focuses on the implementation of the numerical method presented in Chapters 3 and 4. Predominantly, we discuss these issues independent of the programming language. The code for this thesis is implemented in `FORTRAN`.
We consider the generation of a locally aligned mesh introduced in Section 3.3.3 with aligned upper and bottom interfaces as aligned left and right interfaces can be readily obtained by a change of variables. For this, the direction of $\boldsymbol{b}$ has to be constant. We allow the local alignment of the mesh to deviate from $\boldsymbol{b}$ and therefore introduce

$$\boldsymbol{b}_{\text{mesh}} = \begin{pmatrix} b_1^{\mathfrak{M}} \\ b_2^{\mathfrak{M}} \end{pmatrix} \in \mathbb{R}^2 \, . \tag{5.1}$$

This provides the possibility to investigate the impact of the mesh-alignment as well as the choice of $\boldsymbol{b}_{\text{mesh}} \approx \boldsymbol{b}$ such that the conformity condition (3.14) for fully aligned meshes is fulfilled.
The outline of this chapter is as follows: Section 5.1 is a technical derivation of the analysis necessary for transforming integrals of the variational forms to a reference element. Section 5.2 then presents quadratures for evaluating these transformed integrals and introduces the bases for the discontinuous Galerkin method. Section 5.3 deals with the construction of data structure for generating the underlying mesh. As a tool for postprocessing solutions, Section 5.4 outlines the framework for associating eigenvectors to Fourier modes. We follow with an overview of the libraries included in the `FORTRAN`-code in Section 5.5. This chapter closes with an overview of the main input parameters of the `FORTRAN`-code in Section 5.6.

5.1 Integral transform to reference element

For constructing the system matrices of Sections 3.5.2, 3.6.3, 3.6.5, 3.8, 4.2.2 and 4.2.4, various integrals need to be evaluated on cells and parts of their boundary. The goal of this section is

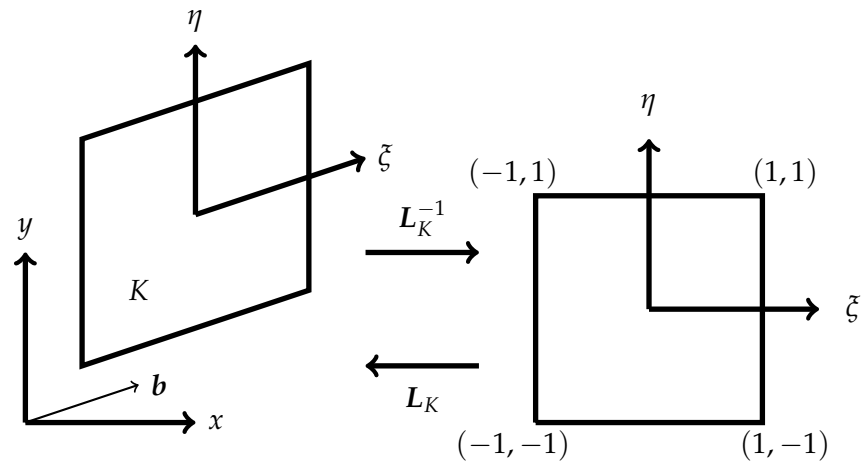


FIGURE 5.1: *Map of the reference element to a cell K and its inverse.*

to provide the groundwork for the numerical quadrature of volume and surface integrals by transforming the domain of integration to a reference element which we choose as $[-1,1]^2$. We define the coordinate system of the reference element by $(\xi,\eta)^\top$.
For fixed $\boldsymbol{b}_{\text{mesh}}$, explicit formulae for the transform maps and their derivatives are presented in Section 5.1.1. Transformed volume and boundary integrals are derived in Section 5.1.2.

5.1.1 Domain transform

This section provides explicit formulae for transform functions and their derivatives for the locally aligned mesh with aligned upper and bottom interfaces presented in Section 3.3.3. The results for aligned left and right interfaces can be obtained by performing the calculations for the coordinate system with exchanged coordinates.
We define

$$h_x := \frac{2\pi}{N_x} \quad , \qquad h_y := \frac{2\pi}{N_y} \, . \tag{5.2}$$

Using the cell indexation defined in Section 3.8, we define the map of the reference element to a cell K with index (k,l) as $\boldsymbol{L}_K$ and its inverse as

$$\begin{pmatrix} x \\ y \end{pmatrix} = \boldsymbol{L}_K(\xi,\eta) = \begin{pmatrix} \frac{h_x}{2} & 0 \\ \frac{b_2^{\mathfrak{M}} h_x}{b_1^{\mathfrak{M}} 2} & \frac{h_y}{2} \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} + \begin{pmatrix} \left(k-\frac{1}{2}\right) h_x \\ \frac{b_2^{\mathfrak{M}} h_x}{b_1^{\mathfrak{M}} 2} + \left(l-\frac{1}{2}\right) h_y \end{pmatrix} , \tag{5.3}$$

$$\begin{pmatrix} \xi \\ \eta \end{pmatrix} = \boldsymbol{L}_K^{-1}(x,y) = \begin{pmatrix} \frac{2}{h_x} & 0 \\ -\frac{b_2^{\mathfrak{M}} 2}{b_1^{\mathfrak{M}} h_y} & \frac{2}{h_y} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} 2k-1 \\ 2l-1-2(k-1)\frac{b_2^{\mathfrak{M}} h_x}{b_1^{\mathfrak{M}} h_y} \end{pmatrix} . \tag{5.4}$$

The corresponding sketch of this map and its inverse is depicted in Figure 5.1. The Jacobian

matrices are given by

$$D_{\xi,\eta}\boldsymbol{L}_K(\xi,\eta) = \begin{pmatrix} \frac{h_x}{2} & 0 \\ \frac{b_2^{\mathfrak{M}} h_x}{b_1^{\mathfrak{M}} 2} & \frac{h_y}{2} \end{pmatrix} \quad , \quad D_{x,y}\boldsymbol{L}_K^{-1}(x,y) = \begin{pmatrix} \frac{2}{h_x} & 0 \\ -\frac{b_2^{\mathfrak{M}} 2}{b_1^{\mathfrak{M}} h_y} & \frac{2}{h_y} \end{pmatrix} \tag{5.5}$$

which are constant in their respective variables with determinants given by

$$\mathcal{J} := \left|\det\left(D_{\xi,\eta}\boldsymbol{L}_K(\xi,\eta)\right)\right| = \frac{h_x h_y}{4} \quad , \quad \det\left(D_{x,y}\boldsymbol{L}_K^{-1}(x,y)\right) = \frac{4}{h_x h_y} \,. \tag{5.6}$$

The maps on boundaries of the reference element are given by

$$\begin{pmatrix} x \\ y \end{pmatrix} = \boldsymbol{L}_K(\pm 1,\eta) = \begin{pmatrix} 0 \\ \frac{h_y}{2} \end{pmatrix} \eta + \begin{pmatrix} \left(k - \frac{1}{2} \pm \frac{1}{2}\right) h_x \\ \frac{b_2^{\mathfrak{M}} h_x}{b_1^{\mathfrak{M}} 2} \pm \frac{b_2^{\mathfrak{M}} h_x}{b_1^{\mathfrak{M}} 2} + \left(l - \frac{1}{2}\right) h_y \end{pmatrix} \tag{5.7}$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \boldsymbol{L}_K(\xi,\pm 1) = \begin{pmatrix} \frac{h_x}{2} \\ \frac{b_2^{\mathfrak{M}} h_x}{b_1^{\mathfrak{M}} 2} \end{pmatrix} \xi + \begin{pmatrix} \left(k - \frac{1}{2}\right) h_x \\ \frac{b_2^{\mathfrak{M}} h_x}{b_1^{\mathfrak{M}} 2} + \left(l - \frac{1}{2} \pm \frac{1}{2}\right) h_y \end{pmatrix} \tag{5.8}$$

with the norms of their derivatives being

$$\mathcal{J}_{F,K} := \begin{cases} \|D_\eta \boldsymbol{L}_K(\pm 1,\eta)\|_2 = \frac{h_y}{2} & , \quad F \text{ y-aligned} \\ \|D_\xi \boldsymbol{L}_K(\xi,\pm 1)\|_2 = \frac{h_x}{2}\sqrt{1 + \left(\frac{b_2^{\mathfrak{M}}}{b_1^{\mathfrak{M}}}\right)^2} & , \quad F\ \boldsymbol{b}_{\text{mesh}}\text{-aligned} \end{cases} \tag{5.9}$$

5.1.2 Integral transforms

This section provides formulae for the transform of the volume and boundary integrals presented at the end of Sections 3.5.1, 3.6.2, 3.6.4, 4.2.1 and 4.2.3. The integration is carried out on the reference element $[-1,1]^2$. For simplifying the notation, sub- and superscripts indicating specific cells and associated maps are dropped whenever there is only one cell and map considered.
When performing integral transforms, we have to consider derivatives in the respective transformed coordinate system and hence state the following Lemma.

Lemma 5.1. *Let $f : K \to \mathbb{C}$ be a scalar-valued differentiable function and $\boldsymbol{L} : [-1,1]^2 \to K$ a vector-valued differentiable function, $(x,y)^\top$ the coordinate system on K and $(\xi,\eta)^\top$ the coordinate system on $[-1,1]^2$. Then, it holds*

$$\left(\nabla_{x,y}(f \circ \boldsymbol{L})\right)(\xi,\eta) = \left(D_{x,y}\boldsymbol{L}^{-1}(x,y)\right)^\top \cdot \left(\nabla_{\xi,\eta}(f \circ \boldsymbol{L})\right)(\xi,\eta) \,. \tag{5.10}$$

Proof:
Using the transform function $\boldsymbol{L}$ and its inverse $\boldsymbol{L}^{-1}$ given in the left equation of (5.3) and (5.4), we obtain

$$\begin{aligned}\left(\nabla_{x,y}\left(f\circ\boldsymbol{L}\right)\right)(\xi,\eta) &= \left(\nabla_{x,y}\left(f\circ\boldsymbol{L}\circ\boldsymbol{L}^{-1}\right)\right)(x,y)\\ &= \begin{pmatrix}\left(\nabla_{\xi,\eta}\left(f\circ\boldsymbol{L}\right)\right)(\xi,\eta)\cdot\left(\partial_x\boldsymbol{L}^{-1}\right)(x,y)\\ \left(\nabla_{\xi,\eta}\left(f\circ\boldsymbol{L}\right)\right)(\xi,\eta)\cdot\left(\partial_y\boldsymbol{L}^{-1}\right)(x,y)\end{pmatrix}\\ &= \left(D_{x,y}\boldsymbol{L}^{-1}(x,y)\right)^{\top}\cdot\left(\nabla_{\xi,\eta}\left(f\circ\boldsymbol{L}\right)\right)(\xi,\eta)\ .\end{aligned} \tag{5.11}$$

■

Using Lemma 5.1, we establish the transformed volume integrals using the rules for integration by substitution. Defining the scalar-valued functions $f,g:K\to\mathbb{C}$, these write for general $\boldsymbol{b}(x,y)$

$$\int_K fg\,\mathrm{d}(x,y)=\int_{[-1,1]^2}\left(f\circ\boldsymbol{L}\right)\left(g\circ\boldsymbol{L}\right)\mathcal{J}\,\mathrm{d}(\xi,\eta) \tag{5.12}$$

$$\begin{aligned}\int_K f\boldsymbol{b}\cdot\nabla_{x,y}g\,\mathrm{d}(x,y) &= \int_{[-1,1]^2}\left(f\circ\boldsymbol{L}\right)(\xi,\eta)\left(\boldsymbol{b}\circ\boldsymbol{L}\right)(\xi,\eta)\cdot\left(\nabla_{x,y}\left(g\circ\boldsymbol{L}\right)\right)(\xi,\eta)\,\mathcal{J}\,\mathrm{d}(\xi,\eta)\\ &= \int_{[-1,1]^2}\left(f\circ\boldsymbol{L}\right)(\xi,\eta)\left(\boldsymbol{b}\circ\boldsymbol{L}\right)(\xi,\eta)\cdot\left(D_{x,y}\boldsymbol{L}^{-1}(x,y)\right)^{\top}\cdot\left(\nabla_{\xi,\eta}\left(g\circ\boldsymbol{L}\right)\right)(\xi,\eta)\,\mathcal{J}\,\mathrm{d}(\xi,\eta)\end{aligned} \tag{5.13}$$

where $\mathcal{J}$ denotes the Jacobian given in (5.6). For fixed $\boldsymbol{b}_{\text{mesh}}$, the Jacobian matrix $D_{x,y}L^{-1}(x,y)$ is constant in (x,y). For extensions to general, (x,y)-dependent $\boldsymbol{b}_{\text{mesh}}$ this term has to be treated further to translate (x,y)-dependence into (ξ,η)-dependence.
For the transformation of surface integrals, we establish the building blocks for treating all presented combinations of flux-pairs. Constants are not affected by the transformation process.
Let $f,g:\Omega\to\mathbb{C}$ be scalar-valued functions. Considering a single interface F with neighbouring cells K and $N_F(K)$, the boundary integrals for flux jumps and averages defined in (3.25) are transformed as follows

$$\begin{aligned}\int_F\boldsymbol{b}\cdot[\![f]\!]\,\mathrm{d}S(x,y) &= \int_{\boldsymbol{L}_K^{-1}(F)}\left(\boldsymbol{b}\circ\boldsymbol{L}_K\right)\cdot\boldsymbol{n}_K\left(f^K\circ\boldsymbol{L}_K\right)\mathcal{J}_{F,K}\,\mathrm{d}S(\xi,\eta)+\\ &+\int_{\boldsymbol{L}_{N_F(K)}^{-1}(F)}\left(\boldsymbol{b}\circ\boldsymbol{L}_{N_F(K)}\right)\cdot\boldsymbol{n}^{N_F(K)}\left(f^{N_F(K)}\circ\boldsymbol{L}_{N_F(K)}\right)\mathcal{J}_{F,N_F(K)}\,\mathrm{d}S(\xi,\eta)\end{aligned} \tag{5.14}$$

$$\begin{aligned}\int_F\{\!\{f\}\!\}\,\mathrm{d}S(x,y) &= \int_{\boldsymbol{L}_K^{-1}(F)}\frac{1}{2}\left(f^K\circ\boldsymbol{L}_K\right)\mathcal{J}_{F,K}\,\mathrm{d}S(\xi,\eta)\\ &\quad+\int_{\boldsymbol{L}_{N_F(K)}^{-1}(F)}\frac{1}{2}\left(f^{N_F(K)}\circ\boldsymbol{L}_{N_F(K)}\right)\mathcal{J}_{F,N_F(K)}\,\mathrm{d}S(\xi,\eta)\end{aligned} \tag{5.15}$$

where $J_{F,K}$ are the norms of the path parameterization induced by the map L_K given in (5.9) as the transformation can be carried out over the full edge on which the interface F is located. Establishing the treatment of products of functions f^K and $g^{N_F(K)}$ with support on neighbouring cells as

$$\int_F f^K g^{N_F(K)}\,\mathrm{d}(x,y)=\int_{\boldsymbol{L}_K^{-1}(F)}\left(f^K\circ\boldsymbol{L}_K\right)\left(\left(g^{N_F(K)}\circ L_{N_F(K)}\right)\circ L_{N_F(K)}^{-1}\circ L_K\right)\mathcal{J}_{F,K}\,\mathrm{d}S(\xi,\eta)\ , \tag{5.16}$$

the framework necessary for dealing with all presented combinations of flux-pairs on interfaces is complete.

5.2 Bases and integral evaluation

This section presents quadrature rules for integral evaluation and deals with the construction of a particularly suited basis for the transformed integrals presented in Section 5.1.2.
For the evaluation of these integrals, we use Legendre-Gauss [79, p.887, 25.4.29] or Legendre-Gauss-Lobatto [79, p.888, 25.4.32] quadrature. For weights and nodes in the (ξ, η)-coordinate system, we use the notation

$$\omega_k^\xi, \xi_k\,, \quad k = 0, \ldots, p_\xi \qquad \omega_k^\eta, \eta_k\,, \quad k = 0, \ldots, p_\eta \tag{5.17}$$

for $p_\xi, p_\eta \in \mathbb{N}$. In the code, we globally assign one number of nodes each for all quadratures in ξ- and η-direction.
The quadrature formula for a function $f : [-1,1]^2 \to \mathbb{C}$ writes

$$\int_{[-1,1]^2} f(\xi,\eta)\ \mathrm{d}(\xi,\eta) \approx \sum_{i=0}^{p_\xi}\sum_{j=0}^{p_\eta} \omega_i^\xi \omega_j^\eta f(\xi_i, \eta_j) \tag{5.18}$$

which is exact for integrating polynomials with degrees in ξ, η less or equal than $2p_\xi + 1, 2p_\eta + 1$ for Legendre-Gauss and less or equal than $2p_\xi - 1, 2p_\eta - 1$ for Legendre-Gauss-Lobatto quadrature. Defining the Legendre polynomials $P_n(\xi)$ using the three term recurrence [79, p. 775, 22.3.8. and p. 782, 22.7.10]

$$\begin{aligned} &P_0(\xi) = 1\,, \quad P_1(\xi) = \xi \\ &(n+1)\,P_{n+1}(\xi) = (2n+1)\,\xi P_n(\xi) - nP_{n-1}(\xi) \end{aligned} \tag{5.19}$$

we can retrieve the Legendre-Gauss weights and nodes for $k = 0, \ldots, p_\xi$ as

$$\xi_k \text{ is the } (k+1)^{\text{th}} \text{ zero of } P_{p_\xi+1}(\xi) \tag{5.20}$$

$$\omega_k^\xi = \frac{2}{\left(1-\xi_k^2\right)\left(P'_{p_\xi+1}(\xi_k)\right)^2} \tag{5.21}$$

whereas for $k = 1, \ldots, p_\xi - 1$, the Legendre-Gauss-Lobatto weights and nodes are

$$\xi_k \text{ is the } k^{\text{th}} \text{ zero of } P'_{p_\xi}(\xi) \qquad \xi_0 = -1, \xi_{p_\xi} = 1 \tag{5.22}$$

$$\omega_k^\xi = \frac{2}{p_\xi\left(p_\xi+1\right)\left(P_{p_\xi}(\xi_k)\right)^2} \qquad \omega_0 = \omega_{p_\xi} = \frac{2}{p_\xi\left(p_\xi+1\right)} \tag{5.23}$$

The same holds respectively for ω_k^η, η_k.
Using a set of interpolation points $\{\xi_k\}_k^{p_\xi}$, we can build Lagrange basis polynomials [79, p.878,

25.2.2] defined as

$$l_p^\xi(\xi) = \prod_{\substack{k=0\\k\neq p}}^{p_\xi} \frac{\xi - \xi_k}{\xi_p - \xi_k}, \quad p = 0, \dots, p_\xi \qquad , \qquad l_q^\eta(\eta) = \prod_{\substack{k=0\\k\neq q}}^{p_\eta} \frac{\eta - \eta_k}{\eta_q - \eta_k}, \quad q = 0, \dots, p_\eta \,. \tag{5.24}$$

We now choose the spaces of basis functions $V_{K,\Phi}$, $V_{K,U}$, $V_{K,Q}$, $V_{K,F,R}$, $V_{K,U_j}, j = 1,2,3,4$ of Sections 3.5, 3.6, 4.2 for all $K \in \mathcal{K}$ structurally as

$$V_K := \operatorname{span}\left\{ \left(l_p^\xi l_q^\eta \right) \circ \boldsymbol{L}_K^{-1}, \quad p = 0, \dots, p_\xi, \quad q = 0, \dots, p_\eta \right\} \tag{5.25}$$

where the degrees of basis functions p_ξ, p_η are the same for all $K \in \mathcal{K}$ but may differ among the different approximated functions. Note that the degrees might have to coincide in some cases to ensure the symmetry of the resulting system matrices as outlined in the previously mentioned sections.
Exploring some properties of the chosen basis, we first observe

$$l_p^\xi(\xi_k) = \delta_{pk} \qquad , \qquad l_q^\eta(\eta_k) = \delta_{qk} \tag{5.26}$$

by definition where δ is the Kronecker delta. For $f(\xi, \eta) = l_p^\xi(\xi) l_q^\eta(\eta)$ in (5.18) this yields

$$\int_{[-1,1]^2} l_p^\xi(\xi) l_q^\eta(\eta) \, \mathrm{d}(\xi, \eta) = \sum_{i=0}^{p_\xi} \sum_{j=0}^{p_\eta} \omega_i^\xi \omega_j^\eta l_p^\xi(\xi_i) l_q^\eta(\eta_j) = \sum_{i=0}^{p_\xi} \sum_{j=0}^{p_\eta} \omega_i^\xi \omega_j^\eta \delta_{pi} \delta_{qj} = \omega_p^\xi \omega_q^\eta \tag{5.27}$$

which allows us to hugely simplify the evaluation of mass integrals. By insertion of the basis functions (5.25) into (5.12) for f and g, we obtain

$$\begin{aligned} &\int_K \left(l_{p_1}^\xi l_{q_1}^\eta \right) \circ \boldsymbol{L}_K^{-1} \left(l_{p_2}^\xi l_{q_2}^\eta \right) \circ \boldsymbol{L}_K^{-1} \, \mathrm{d}(x, y) = \int_{[-1,1]^2} l_{p_1}^\xi l_{q_1}^\eta l_{p_2}^\xi l_{q_2}^\eta \mathcal{J} \, \mathrm{d}(\xi, \eta) \\ &= \mathcal{J} \sum_{i=0}^{p_\xi} \sum_{j=0}^{p_\eta} \omega_i^\xi \omega_j^\eta l_{p_1}^\xi(\xi_i) l_{q_1}^\eta(\eta_j) l_{p_2}^\xi(\xi_i) l_{q_2}^\eta(\eta_j) \\ &= \mathcal{J} \sum_{i=0}^{p_\xi} \sum_{j=0}^{p_\eta} \omega_i^\xi \omega_j^\eta \delta_{p_1 i} \delta_{q_1 j} \delta_{p_2 i} \delta_{q_2 j} = \mathcal{J} \omega_{p_1}^\xi \omega_{q_1}^\eta \delta_{p_1 p_2} \delta_{q_1 q_2} \,. \end{aligned} \tag{5.28}$$

Hence, the mass matrices are diagonal as the integral is non-zero if and only if test and trial function coincide.
As derivatives appear in the remaining volume integrals, we need the derivative of the basis of Legendre polynomials

$$\frac{\partial l_p^\xi}{\partial \xi}(\xi) = \prod_{\substack{k=0\\k\neq p}}^{p_\xi} \frac{1}{\xi_p - \xi_k} \sum_{\substack{j=0\\j\neq p}}^{p_\xi} \prod_{\substack{k=0\\k\neq p,j}}^{p_\xi} (\xi - \xi_k) \,, \quad p = 0, \dots, p_\xi \tag{5.29}$$

which can be stated analogously for $\partial l_q^\eta / \partial \eta$. Defining

$$\Pi_p^\xi := \prod_{\substack{k=0 \\ k \neq p}}^{p_\xi} \frac{1}{\xi_p - \xi_k} \qquad , \qquad \Pi_q^\eta := \prod_{\substack{k=0 \\ k \neq q}}^{p_\eta} \frac{1}{\eta_q - \eta_k} \tag{5.30}$$

the derivatives evaluated at the quadrature nodes yield

$$\frac{\partial l_p^\xi}{\partial \xi}(\xi_l) = \Pi_p^\xi \sum_{\substack{j=0 \\ j \neq p}}^{p_\xi} \prod_{\substack{k=0 \\ k \neq p,j}}^{p_\xi} (\xi_l - \xi_k) = \begin{cases} \sum_{\substack{j=0 \\ j \neq p}}^{p_\xi} \frac{1}{\xi_p - \xi_j} & l = p \\ \Pi_p^\xi \prod_{\substack{k=0 \\ k \neq p,l}}^{p_\xi} (\xi_l - \xi_k) & l \neq p \end{cases} \tag{5.31}$$

The volume and boundary integrals are evaluated using the presented Legendre-Gauss or Legendre-Gauss-Lobatto quadrature.
When inserting test and trial functions in the volume and boundary integrals of Section 5.1.2, we note, that the evaluation of these functions is reduced to evaluating Lagrange polynomials and their derivatives at the quadrature nodes by definition of the bases (5.25). Information about the map to the reference element is only used in the Jacobians and transformation of $\boldsymbol{b}$.
We further note that the Lagrange polynomials can be built with other nodes than the ones used for the quadrature. This allows to increase the number of quadrature points while keeping the degrees of the bases fixed which is beneficial for variable $\boldsymbol{b}$, the Fourier postprocessing of Section 5.4 and if integrals include metric terms. However, in this case formulae (5.26), (5.27), (5.28) and (5.31) don't hold and the mass matrix becomes a block diagonal matrix with each block being a dense element mass matrix.
Considering the evaluation of boundary integrals on non-conforming interfaces, the integration is carried out over parts of the whole edge. The quadrature nodes and weights for boundary integrals are transformed to these intervals.
For solving the anisotropic wave equation of Chapter 3 using the mixed variational form with LDG fluxes presented in Sections 3.6.2 and 3.6.3, we choose the same degrees for all bases of $V_{K,\Phi}$ and $V_{K,U}$. For Bassi-Rebay 2 fluxes presented in Sections 3.6.4 and 3.6.5, we choose the same degrees for all bases of $V_{K,\Phi}$, $V_{K,Q}$ and $V_{K,F,R}$.
Considering the symmetry condition in Corollary 4.3 for the 4^{th}-order problem of Chapter 4, we opt for the same degrees for all bases of $V_{K,\Phi}$, $V_{K,U_j}, j = 1,2,3,4$. For all equations and bases, p_ξ and p_η may differ.

5.3 Mesh generation

This section outlines the implementation of data structures for the locally aligned mesh presented in Section 3.3.3 with aligned upper and bottom interfaces.

The mesh is generated using the parameters

$$\boldsymbol{b}_{\text{mesh}}, N_x, N_y \tag{5.32}$$

as input.
We first assign a global index to all cells. Using the (k,l)-index introduced in Section 3.8, we define the global index

$$i_e\,(k,l) \coloneqq k + (l-1)N_x \qquad , \qquad k = 1,\ldots,N_x \qquad , \qquad l = 1,\ldots,N_y\,. \tag{5.33}$$

The (k,l)-index can be retrieved via

$$(k,l) = \left(\text{mod}\,(i_e - 1,\ N_x) + 1,\ \left\lceil \frac{i_e}{N_x} \right\rceil\right) \tag{5.34}$$

where $\lceil . \rceil$ are upper Gauss brackets. For visualizing the mesh and calculating the transformed integrals presented in Section 5.1.2, we identify the corners and points of non-conformity of each cell as well as the interfaces as shown in Figure 5.2. Note that an index from $1,\ldots,6$ can be assigned to each point and interface. This index association is set globally throughout the code. Further note that the interfaces have a fixed orientation.

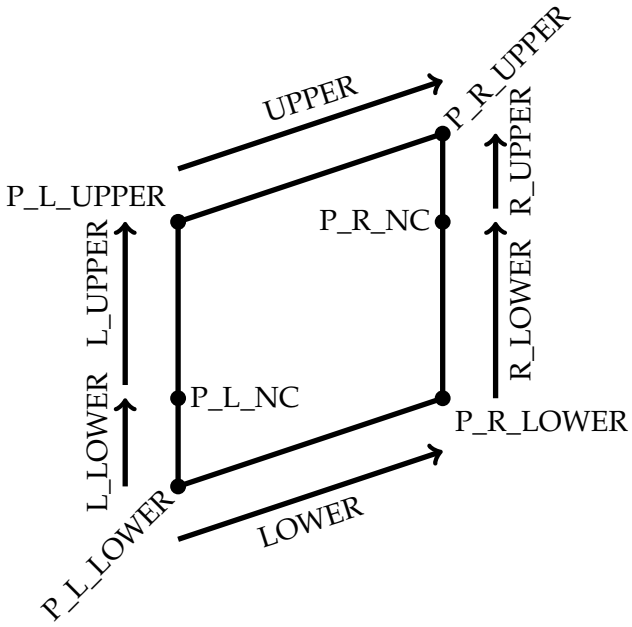


FIGURE 5.2: *Nomenclature of cell points and interfaces.*

Using the vertical shift constant $c = \left\lfloor \frac{b_2^{\mathfrak{M}}}{b_1^{\mathfrak{M}}} \frac{N_y}{N_x} \right\rfloor$ defined in (3.122) and map $\boldsymbol{L}_K^{-1}$ onto the reference element defined in (5.4), we summarize the coordinates of these points in Table 5.1. The start and end points of each interface can then be established by combining Figure 5.2 and Table 5.1.

point	x-abscissa	y-ordinate	ξ-abscissa	η-ordinate
P_L_LOWER	$(k-1)h_x$	$(l-1)h_y$	-1	-1
P_L_UPPER	$(k-1)h_x$	lh_y	-1	1
P_R_LOWER	kh_x	$\frac{b_2^{\mathfrak{M}}h_x}{b_1^{\mathfrak{M}}} + (l-1)h_y$	1	-1
P_R_UPPER	kh_x	$\frac{b_2^{\mathfrak{M}}h_x}{b_1^{\mathfrak{M}}} + lh_y$	1	1
P_L_NC	$(k-1)h_x$	$\frac{b_2^{\mathfrak{M}}h_x}{b_1^{\mathfrak{M}}} + (l-c-1)h_y$	-1	$-1+2\left(\frac{b_2^{\mathfrak{M}}h_x}{b_1^{\mathfrak{M}}h_y} - c\right)$
P_R_NC	kh_x	$(l+c)h_y$	1	$1-2\left(\frac{b_2^{\mathfrak{M}}h_x}{b_1^{\mathfrak{M}}h_y} - c\right)$

TABLE 5.1: *Coordinates of cell points of K with index (k,l).*

Note that the points of non-conformity coincide with a corner point for a conforming mesh. The respective integral evaluations then collapse to integrals over a single point which evaluate to 0. Boundary integrals are evaluated for each interface F. Hence, we establish a global interface index by successive enumeration. More details are provided later this section. Furthermore, knowledge of the indices of the neighbouring cells as well as the local position of the interface within these cells is necessary. Using the uniformity of the neighbourhood when using the locally aligned mesh, the local position of interfaces can be established as shown in Table 5.2.

local interface index	local interface index of neighbour
L_LOWER	R_UPPER
L_UPPER	R_LOWER
R_LOWER	L_UPPER
R_UPPER	L_LOWER
LOWER	UPPER
UPPER	LOWER

TABLE 5.2: *Local interface indices within a cell and its respective neighbour.*

The connection of the index (k,l) of a cell and its neighbour is summarized in Table 3.1. We update this table using the local position of interfaces in Table 5.3.
Lastly the unit outer normals for each interface are summarized in Table 5.4.
Using the number of all cells $N_\Sigma = N_x N_y$ previously defined in (3.10), we summarize this information in the following data structure with the given dimensionality.

- `KL_To_Elem`: $N_x \times N_y$, see (5.33)
  - 1st dimension: k-index
  - 2nd dimension: l-index

local interface index	k (neighbour)	l (neighbour)
L_LOWER	$\mathrm{mod}\,(k-2,\ N_x)+1$	$\mathrm{mod}\,(l-c-2,\ N_y)+1$
L_UPPER	$\mathrm{mod}\,(k-2,\ N_x)+1$	$\mathrm{mod}\,(l-c-1,\ N_y)+1$
R_LOWER	$\mathrm{mod}\,(k,\ N_x)+1$	$\mathrm{mod}\,(l+c-1,\ N_y)+1$
R_UPPER	$\mathrm{mod}\,(k,\ N_x)+1$	$\mathrm{mod}\,(l+c,\ N_y)+1$
LOWER	k	$\mathrm{mod}\,(l-2,\ N_y)+1$
UPPER	k	$\mathrm{mod}\,(l,\ N_y)+1$

TABLE 5.3: *Indices of the neighbours of cell (k,l) sharing a certain local interface.*

local interface index	unit outer normal
L_LOWER	$(-1,0)^\top$
L_UPPER	$(-1,0)^\top$
R_LOWER	$(1,0)^\top$
R_UPPER	$(1,0)^\top$
LOWER	$\left(b_2^{\mathfrak{M}},-b_1^{\mathfrak{M}}\right)^\top / \| \left(-b_2^{\mathfrak{M}},b_1^{\mathfrak{M}}\right)^\top \|_2$
UPPER	$\left(-b_2^{\mathfrak{M}},b_1^{\mathfrak{M}}\right)^\top / \| \left(-b_2^{\mathfrak{M}},b_1^{\mathfrak{M}}\right)^\top \|_2$

TABLE 5.4: *Unit outer normals of interfaces.*

- `Elem_To_KL`: $2 \times N_\Sigma$, see (5.34)
  - 1^{st} dimension: contains: (k,l)-index
  - 2^{nd} dimension: global cell index i_e
- `Interface_To_Points`: 2×6, see Figure 5.2
  - 1^{st} dimension: contains: local start (1) and end (2) point index
  - 2^{nd} dimension: local interface index
- `XCoordsElem`: $2 \times 6 \times N_\Sigma$, see Figure 5.2 and Table 5.1
  - 1^{st} dimension: contains: (x,y)-coordinates
  - 2^{nd} dimension: local point index
  - 3^{rd} dimension: global cell index i_e
- `LocalInterfaceID`: 2×6, see Table 5.2
  - 1^{st} dimension: contains: local interface index in main cell , neighbour cell
  - 2^{nd} dimension: local interface index in main cell

- `InterfaceInfo`: $4 \times 6 \times N_\Sigma$, see Figure 5.2, Tables 5.2 and 5.3
  - 1^{st} dimension: contains: global interface index, local interface index, cell index of neighbour, local interface index of neighbour
  - 2^{nd} dimension: local interface index in main cell
  - 3^{rd} dimension: global cell index i_e
- `LocalIntLimits`: $2 \times 4 \times 6 \times N_\Sigma$, see Figure 5.2 and Table 5.1
  - 1^{st} dimension: contains: (ξ, η)-coordinates
  - 2^{nd} dimension: start $(1, 3)$, end $(2, 4)$ point of interface in main, neighbouring cell
  - 3^{rd} dimension: local interface index in main cell
  - 4^{th} dimension: global cell index i_e
- `UnitOuterNormals`: $2 \times 6 \times N_\Sigma$, see Table 5.4
  - 1^{st} dimension: contains: unit outer normal $\boldsymbol{n}$
  - 2^{nd} dimension: local interface index
  - 3^{rd} dimension: global cell index i_e

For the assignment of the global interface index of `InterfaceInfo`, we start with

$$\texttt{InterfaceInfo}(1,1,1) = 1 = \texttt{InterfaceInfo}(1,i,j) \tag{5.35}$$

where j is the index of the neighbour of cell 1 sharing the local interface 1 in cell 1 and i the local position of the interface within cell j. After an increment of the global interface index by 1, this process is repeated for each unset global interface component of `InterfaceInfo`.
We note that the data structure for `UnitOuterNormals` is redundant in its current form as the normals are the same for each cell. However, the proposed structure allows generalizing $\boldsymbol{b}$.
The Jacobian matrices and Jacobians for the volume and surface integrals are the same for each cell. Hence we can store (5.5) and (5.6) once and (5.9) once for each interface. For the extension to general $\boldsymbol{b}$ the associated data structure can be generalized in the same manner as for `UnitOuterNormals`.

5.4 Fourier postprocessing

As Fourier modes are exact eigenfunctions of (3.1) and (4.1), we are interested in associating them to the discrete eigenvalues and eigenvectors. After computing the eigensystem of the non-reduced generalized (3.67), (4.51), reduced generalized (3.41), (3.59), (3.99), (4.55) or standard (3.43), (3.69)

eigenvalue problem which share the form

$$A\boldsymbol{V} = \omega^2 B\boldsymbol{V} \qquad , \qquad AM^{\frac{1}{2}}\boldsymbol{\Phi} = \omega^2 M^{\frac{1}{2}}\boldsymbol{\Phi} \tag{5.36}$$

where the generalized eigenvalue problem is shown on the left and the standard eigenvalue problem is shown on the right, we project the obtained eigenvectors, which are multiplied by $M^{-\frac{1}{2}}$ in the standard eigenvalue problem case, onto a truncated space of Fourier modes. Note that in the case of non-reduced generalized eigenvalue problems, we are only interested in projecting the components of the eigenvector associated to the coefficient vector $\boldsymbol{\Phi}$.
The space of Fourier modes is defined as

$$\left\{ \phi_{m,n}(x,y) = \exp\left(\mathrm{i}\left(mx+ny\right)\right) \,\middle|\, |m| \leqslant m_{\max}, |n| \leqslant n_{\max} \right\} \tag{5.37}$$

with maximal mode numbers $m_{\max}, n_{\max}$. We recall

$$\phi(x,y) = \sum_{K\in\mathcal{K}} \sum_{k} \Phi_k^K \phi_k^K(x,y) \qquad , \qquad \Phi_k^K \in \mathbb{R}\,, \boldsymbol{\Phi} = \left(\Phi_k^K\right)_{k,K} \tag{5.38}$$

where the basis functions ϕ_k^K are defined in (5.25) and eigenvector $\boldsymbol{\Phi}$. The Fourier coefficients of the basis functions are given by

$$\widehat{\phi}_{k,K}^{m,n} = \int_{\mathbb{R}^2} \phi_k^K(x,y) \exp\left(\mathrm{i}\left(mx+ny\right)\right) \,\mathrm{d}\boldsymbol{x}\,. \tag{5.39}$$

Using

$$\phi_k^K(x,y) \approx \sum_{m=-m_{\max}}^{m_{\max}} \sum_{n=-n_{\max}}^{n_{\max}} \widehat{\phi}_{k,K}^{m,n} \exp\left(\mathrm{i}\left(mx+ny\right)\right)\,, \tag{5.40}$$

we obtain the Fourier expansion of ϕ via

$$\begin{aligned} \phi(x,y) &= \sum_{K\in\mathcal{K}} \sum_{k} \Phi_k^K \phi_k^K(x,y) \\ &\approx \sum_{K\in\mathcal{K}} \sum_{k} \Phi_k^K \sum_{m=-m_{\max}}^{m_{\max}} \sum_{n=-n_{\max}}^{n_{\max}} \widehat{\phi}_{k,K}^{m,n} \exp\left(\mathrm{i}\left(mx+ny\right)\right) \\ &= \sum_{m=-m_{\max}}^{m_{\max}} \sum_{n=-n_{\max}}^{n_{\max}} \sum_{K\in\mathcal{K}} \sum_{k} \Phi_k^K \widehat{\phi}_{k,K}^{m,n} \exp\left(\mathrm{i}\left(mx+ny\right)\right) \end{aligned} \tag{5.41}$$

which yields the Fourier coefficients of the eigenfunction ϕ

$$\widehat{\phi}^{m,n} = \sum_{K\in\mathcal{K}} \sum_{k} \Phi_k^K \widehat{\phi}_{k,K}^{m,n}\,. \tag{5.42}$$

We determine the maximal amplitude of Fourier modes over all eigenvectors given by

$$A_\Phi := \max_{|m|\leqslant m_{\max}} \max_{|n|\leqslant n_{\max}} \left|\widehat{\phi}^{m,n}\right| \qquad , \qquad A_{\max} := \max_{\boldsymbol{\Phi}\text{ eigenvector}} A_{\boldsymbol{\Phi}} \tag{5.43}$$

and associate the eigenvectors whose maximal amplitude is larger than a certain fraction of $A_{\max}$ which we set as

$$A_\Phi \geqslant \frac{1}{40} A_{\max} \, . \tag{5.44}$$

We argue that eigenvectors with a significantly smaller, i.e., with less than a factor of $1/40$, amplitude are either associated to a mode number larger than $m_{\max}$ or $n_{\max}$ or are not sufficiently converged yet. In either case we want to discard these results in the evaluation of the eigensystem. Setting this constant to $1/40$ is debatable. We experienced the choice of $1/4$ as too restrictive and $1/100$ as too generous.
We then associate the Fourier mode $(m_{\mathbf{\Phi}}, n_{\mathbf{\Phi}})$ to the eigenvector $\mathbf{\Phi}$ with

$$(m_{\mathbf{\Phi}}, n_{\mathbf{\Phi}}) = \underset{|m| \leqslant m_{\max}, |n| \leqslant n_{\max}}{\operatorname{argmax}} \left| \widehat{\phi}^{m,n} \right| \, . \tag{5.45}$$

If this is not unique, we choose the representative with $m \geqslant 0$. If there is still more than one, we choose one representative with $m \geqslant 0$ at random.
Dependent on the choice of quadrature points given in Section 5.2 for evaluating the Fourier-coefficients, we might end up with aliasing effects in the Fourier representation of eigenvectors. In the case of constant $\boldsymbol{b}$, Section 3.9.1 considers the locally aligned mesh and the primal variational form of Section 3.5. Therein, the matrices collapse to systems of the size of the basis with powers of the roots of unity $\omega_{N_x} = \exp(2\pi \mathrm{i}/N_x)$, $\omega_{N_y} = \exp\left(2\pi \mathrm{i}/N_y\right)$ as scalar factors. As

$$\omega_{N_x}^{cm} = \omega_{N_x}^{cm+\tilde{m}N_x} \qquad , \qquad \omega_{N_y}^{cn} = \omega_{N_y}^{cn+\tilde{n}N_y} \qquad \forall c, m, n, \tilde{m}, \tilde{n} \in \mathbb{Z} \tag{5.46}$$

the Fourier coefficients of a mode (m, n) and a mode number with multiples in the mesh resolution $(m + \tilde{m}N_x, n + \tilde{n}N_y)$ cannot be distinguished on a discrete level. Therefore, if the mesh resolution is too low, we expect less precise results for the Fourier postprocessing. This can be prevented by choosing

$$m_{\max} \leqslant \frac{N_x - 1}{2} \qquad , \qquad n_{\max} \leqslant \frac{N_y - 1}{2} \tag{5.47}$$

for a given mesh.

5.5 Linked libraries

This section provides a short overview of and motivation for the libraries linked in the `FORTRAN`-code. Section 5.5.1 presents the `FEAST` eigenvalue solver which is particularly suited for computing all eigenvalues within a certain interval. Section 5.5.2 provides an overview of used formats for sparse matrices and libraries for sparse matrix operations. For moving to three-dimensional geometries, Section 5.5.3 introduces a library providing the metric factors of the MHD equilibrium geometry necessary for the anisotropic wave equation with metric terms.

5.5.1 FEAST

We compute the eigensystem of the non-reduced generalized (3.67), (4.51), reduced generalized (3.41), (3.59), (3.99), (4.55) or standard (3.43), (3.69) eigenvalue problem which share the form

$$A\boldsymbol{V} = \omega^2 B\boldsymbol{V} \qquad , \qquad AM^{\frac{1}{2}}\boldsymbol{\Phi} = \omega^2 M^{\frac{1}{2}}\boldsymbol{\Phi} \tag{5.48}$$

where the left hand side system matrix A is symmetric positive semidefinite matrix and B is symmetric positive definite for the reduced system of the anisotropic wave equation presented in Chapter 3. Note, that B is symmetric positive semidefinite for the non-reduced system of the anisotropic wave equation of Chapter 3 and the 4^{th}-order problem of Chapter 4. Therefore, we need an eigenvalue solver which is capable of dealing with generalized eigenvalue problems where both system matrices might be symmetric positive semidefinite.
Furthermore, as outlined in Section 2.7, we are interested in small eigenvalues located in a fixed interval which we define as $[0, \omega^2_{\text{max}}]$. As the amount of discrete eigenvalues within this interval is unknown, commonly used methods like the Arnoldi iteration or the Lanczos algorithm are not as well-suited for this problem.
The `FEAST` algorithm "contains elements from complex analysis, numerical linear algebra and approximation theory, to produce an optimal subspace iteration method using spectral projectors" [80, Section 2.1]. The full documentation of `FEAST` can be found in [80]. Amongst other techniques, complex contour integrations are used for solving the eigenvalue problem. For each contour integration point, a linear system has to be solved.
We provide a list of input parameters for the `FEAST` eigenvalue solver which can be specified in the parameter-file of the `FORTRAN`-code and provide the used default configuration.

- [`Emin`, `Emax`]: Search interval for the eigenvalues.
- `M0`: Upper bound for the number of eigenvalues in the search interval.
- `eigsolveFlag`: Type of eigenvalue solver used
  - `R`: Real-value based eigenvalue solver for B symmetric positive definite; calls `dfeast_srcix`. Default for anisotropic wave equation.
  - `C`: Complex-value based eigenvalue solver for B symmetric positive semidefinite; calls `zfeast_srcix`. Default for 4^{th}-order equation.
- `feast_nCP/fpm(2)`: Number of contour integration points on a half contour; addresses the accuracy of the algorithm within one iteration. Default: 16.
- `feast_epsexp/fpm(3)`: Abort criterion; residue $\leqslant 10^{-\texttt{feast\_epsexp}}$. Default: 10
- `feast_maxNumLoops/fpm(4)`: Maximal number of iterations. Default: 6

As a default for `M0`, we use the system size for calculating all eigenvalues and a quarter of the system size for calculating eigenvalues in an interval. For large system matrices, this number can be reduced by a larger margin.
We remark that in the case of the complex eigenvalue solver `zfeast_srcix`, the convergence of `FEAST` is highly dependent on the initial guess for the number of eigenvalues `M0` as well as the accuracy of the eigenvalue solver set by `feast_nCP`. Small variations of these numbers decide whether the algorithm converges.

5.5.2 `SPARSKIT` and `MUMPS`

For a cell-based assembly of the building blocks of the system matrices, see for example the individual matrices of (3.70), we use the coordinate based COO format [81, Section 2.1.1] storing

$$(\text{row-indices, column-indices, values}) \ . \tag{5.49}$$

For combining these parts to system matrices, we transform them to the compressed sparse row (CSR) format [81, Section 2.1.3] storing

$$(\text{values, non-zero-indices, column-indices}) \tag{5.50}$$

where the non-zero-indices of a $M \times N$ matrix is an array of length $M+1$ recursively defined $\forall k = 2, \dots, M+1$ by

- non-zero-indices$[1] = 0$
- non-zero-indices$[k]$ = non-zero-indices$[k-1]$+ number of non-zeroes in row $k-1$

For performing matrix sums and multiplications of CSR-matrices, we use the `SPARSKIT` library [81].
Using the symmetry of the resulting system matrices, we convert them to lower triangular COO matrices to reduce disk space. These matrices are then used by the eigenvalue solver. For performing matrix multiplications and factorizations as well as the solution of linear systems needed by the `FEAST` eigenvalue solver, we use the `MUMPS` library [82, 83].

5.5.3 `VMEC`

For solving the anisotropic wave equation on a flux surface of a magnetohydrodynamic equilibrium as outlined in Section 3.7, we need the metric terms (3.111) for the transformation to the fully periodic reference domain Ω.
The magnetohydrodynamic equilibrium is provided by the 3D-Variational Moments Equilibrium Code [84, 85], short `VMEC`, which "uses a variational method to find a minimum in the total energy of the system" [86].

This precalculated VMEC-equilibrium can be evaluated at a specified position (s, θ, φ) in straight field line coordinates yielding the constant direction of the magnetic field $\bar{\boldsymbol{b}} = (\iota, 1)^\top$ of the associated flux surface. Furthermore, given a flux surface coordinate s and quadrature nodes $(x, y) \hat{=} (\theta, \varphi)$ of Section 5.2, we can compute all metric factors needed in (3.111).

5.6 Input parameters

The FORTRAN-code offers the specification of a broad range of input parameters for its setup. Several of these parameters have a default configuration. Parameters for the eigenvalue solver are listed in Section 5.5.1. Whenever we deviate from either of these settings in the numerical evaluation of Chapter 6, we highlight the changes appropriately.

- `Beq`: Sets $\boldsymbol{b}$.
- `Bmesh`: Sets the direction of the mesh as outlined at the beginning of this chapter. Default: `Bmesh = Beq`.
- `nElemsX`, `nElemsY`: Sets N_x and N_y, i.e., the parallel and perpendicular resolution of the mesh.
- `stabilizationEtaV`: `V` $\in$ { "", `BR2`, `2`, `Phi`, `E` }. Sets the respective stabilization constants $\eta_{\mathtt{V}}$. Default: 6; Default for `V` $= 2$: 0
- `degXiV, degEtaV`: `V` $\in$ { `Phi`, `U`, `Q`, `U1`, `U2`, `U4` }. Sets the degrees p_ξ, p_η of the respective bases (5.25).
- `degXiEval, degEtaEval`: Specifies the number of quadrature nodes -1 for integral evaluation, see (5.17).
- `nodeTypeInterpol`: Sets the type of nodes for building the Legendre polynomials for the basis functions (5.25), 1 are Legendre-Gauss nodes, 2 are Legendre-Gauss-Lobatto nodes, see Section 5.2. Default for anisotropic wave equation: 2, Default for 4^{th}-order equation: 1
- `noteTypeEval`: Same as `nodeTypeInterpol`, but instead sets the type of nodes for the quadrature (5.17). Default: 1

As outlined at the end of Section 5.2, we choose the same value for all `degXiV` and a not necessarily different value for all `degEtaV`, `V` $\in$ { `Phi`, `U`, `Q`, `U1`, `U2`, `U4` }. To account for the proper evaluation of terms not related to the bases, e.g., metric terms and Fourier postprocessing, we choose higher values for `degXiEval` and `degEtaEval`, namely

$$\texttt{degXiEval} = \lceil 1.5\ \texttt{degXiPhi} \rceil \quad , \quad \texttt{degEtaEval} = \lceil 1.5\ \texttt{degEtaPhi} \rceil \, . \tag{5.51}$$

In the following we refer to `degXiPhi` as p_ξ and to `degEtaPhi` as p_η.
The parallel and perpendicular resolution (degrees of freedom) is defined as

$$\mathrm{DoF}_\parallel := \left(p_\xi + 1\right) N_x \qquad , \qquad \mathrm{DoF}_\perp := \left(p_\eta + 1\right) N_y \tag{5.52}$$

yielding the total resolution

$$\mathrm{DoF} := \mathrm{DoF}_\parallel \, \mathrm{DoF}_\perp \tag{5.53}$$

which is the size of the system matrices for reduced generalized and standard eigenvalue problems. The system size doubles for the non-reduced generalized eigenvalue problem using LDG fluxes for the anisotropic wave equation (3.67) and quadruples when considering the non-reduced generalized eigenvalue problem of the 4^{th}-order equation (4.51).
For setting up the eigensystem, we offer

- `useReducedSystem`: Determines whether the system matrices for mixed variational forms should be reduced in size. Default: `TRUE`
- `generalizedEV`: Cannot be `FALSE` if `useReducedSystem` is `FALSE`
  - `TRUE`: Set up a generalized eigenvalue problem (default)
  - `FALSE`: Set up a standard eigenvalue problem
- `EVmaxMode`: Sets $m_{\max}, n_{\max}$ for the Fourier postprocessing of Section 5.4

For `VMEC`-equilibria, we offer

- `readBFromFluxSurface`: Determines whether to use a `VMEC` equilibrium as an input. Default: `FALSE`
- `radiusFluxSurface`: Normalized flux surface coordinate s
- `vmecWoutFile`: Directory of the `VMEC` equilibrium
- `VMEC_swap_theta_zeta`: Addresses the alignment of interfaces
  - `TRUE`: aligned left and right, i.e., toroidally non-conforming, interfaces (default)
  - `FALSE`: aligned upper and bottom, i.e., poloidally non-conforming, interfaces

Additionally, we offer to visualize the mesh as well as the eigenvectors by setting `visualizeMesh` and `visualizeEV`. For visualizing the mesh on a flux surface, we offer `mapToXYZ`. `nvisu_mesh` and `nvisu_EV` are arrays of length 2 which specify the number of visualization points in ξ, η-direction.

Chapter 6

NUMERICAL RESULTS
Evaluating methods

This chapter discusses the numerical results of the schemes presented in Chapters 3 and 4 using the implementation outlined in Chapter 5. Independent of the concrete choice of variational form, flux and equation, we refer to the method as locally field-aligned discontinuous Galerkin method (ADG). The basic ingredients of ADG consist of the local field-alignment of mesh and basis and a distribution of resolution with $\mathrm{DoF}_\perp > \mathrm{DoF}_\parallel$. We explicitly allow for $\boldsymbol{b}_{\mathrm{mesh}} \approx \boldsymbol{b}$.
For the constant coefficient anisotropic wave equation (3.1) and the 4^{th}-order equation (4.1) with constant $\boldsymbol{b}$, analytical solutions are known. This allows us to study errors of the eigenvalues of the numerical approximations. Throughout this chapter, we present relative errors whenever we speak of errors if not marked explicitly otherwise. If the exact eigenvalue of a mode is zero, we use absolute errors instead. This is particularly the case for the constant mode.
Furthermore, we consider the anisotropic wave equation with metric terms (3.110) on the flux surface of a three-dimensional MHD equilibrium where no analytical solution is known. Hence, we compare the results with existing codes.
For all equations, we examine the impact of the local alignment of mesh and basis. In particular, we compare to non-aligned cartesian meshes. We further investigate different ratios $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel$ for the distribution of resolution by varying the parallel and perpendicular degrees of the bases as well as the resolution of the locally field-aligned mesh. We study the convergence of ADG and, if possible, the dependence of the error on different directions of $\boldsymbol{b}$.
We remark that system matrices are stored as lower triangular sparse matrices as mentioned in Section 5.5.2. Therefore, we investigate ADG regarding the numbers of non-zeroes of the lower triangle of the left hand side system matrices in relation to the total number of matrix entries and abbreviate this percentage by nnzA. All eigenvalues are calculated from generalized eigenvalue problems.
The outline of this chapter is as follows: A reference configuration with constant $\boldsymbol{b}$ for the com-

parison of results is presented in Section 6.1. Section 6.2 studies ADG applied to the constant coefficient anisotropic wave equation whereas Section 6.3 discusses the results for the 4^{th}-order equation. In Section 6.4, we consider the anisotropic wave equation with metric terms from the flux surface of a three-dimensional MHD equilibrium and compare ADG with existing codes.

6.1 Reference case

For the purpose of a first evaluation and comparison of ADG, this section defines a reference test case by fixing the magnetic field direction $\boldsymbol{b} = (\iota_{\text{LHD}}(s), 1)^\top$ at $s = 0.9$. The LHD-like-ι-profile [87], is defined by

$$\begin{aligned}\iota_{\text{LHD}}(s) = 0.47262 + 0.32392\, s + 0.49604\, s^2 + 5.3991 \times 10^{-6}\, s^3 \\ - 3.165 \times 10^{-5}\, s^4 + 4.7963 \times 10^{-5}\, s^5 - 2.2824 \times 10^{-5}\, s^6\end{aligned} \tag{6.1}$$

yielding $\boldsymbol{b} = (\iota, 1)^\top := (1.165939762441386, 1)^\top$. We choose this value of ι to be physically meaningful and result in a locally aligned mesh with low shear and non-conforming interfaces of different lengths. Furthermore, we avoid a low rational number such that all considered eigenvalues differ from zero except for the constant mode.
Since $\boldsymbol{b}$ is constant, the exact eigenvalues of the constant coefficient anisotropic wave equation and the 4^{th}-order equation are given by $\omega_{m,n}^2 = (\iota m + n)^2$ as shown in Theorems 3.1 and 4.1.
In the reference case, we limit the space of the Fourier postprocessing of Section 5.4 to maximal mode numbers $m_{\max} = n_{\max} = 20$. Figure 3.2 shows the distribution of eigenvalues for a logarithmically scaled color map. For improving readability and comparability, Figure 6.1 is the same as Figure 3.2. As outlined at the end of Section 3.1, we emphasize that small eigenvalues are gathered along the direction perpendicular to $\boldsymbol{b}$.
As modes with small parallel gradient and therefore small associated eigenvalues are of interest as outlined in Section 2.7, we aim to resolve all modes with associated eigenvalue $\omega_{m,n}^2 \leqslant 0.2 = \omega_{\max}^2$. We call these modes a band of modes with eigenvalues smaller than $\omega_{\max}^2$.
Note, that if all eigenvalues have to be computed, we set up the search interval of the eigenvalue solver to $[\texttt{Emin}, \texttt{Emax}] = [-0.01, 2000]$.

6.2 Constant coefficient anisotropic wave equation

In this section, we study the impact of the mesh alignment and different discretization parameters of ADG onto the accuracy of the eigenvalues in the reference case for the constant coefficient anisotropic wave equation (3.1). For this, the construction of the method is outlined in Chapter 3. For exploring the capabilities of and a useful setup for ADG, we first examine the mixed variational form using LDG fluxes, as presented in Sections 3.6.2 and 3.6.3, as a reduced, generalized eigenvalue problem.

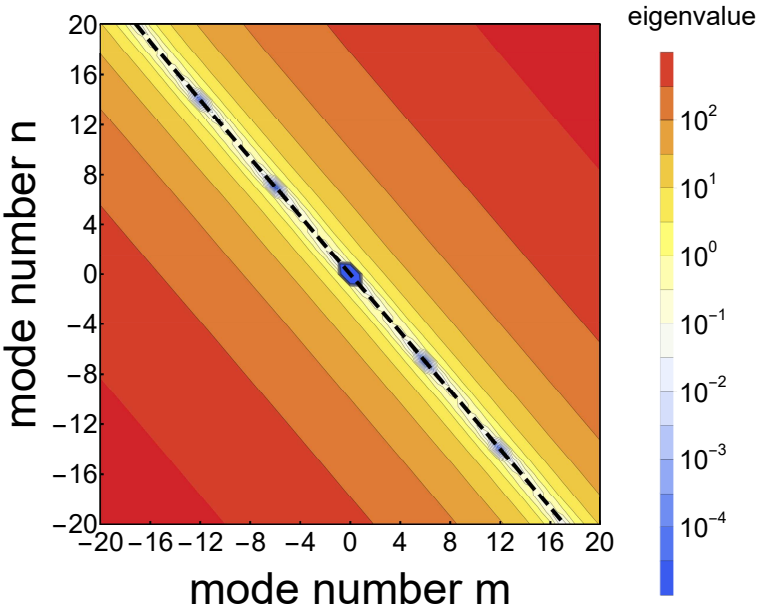
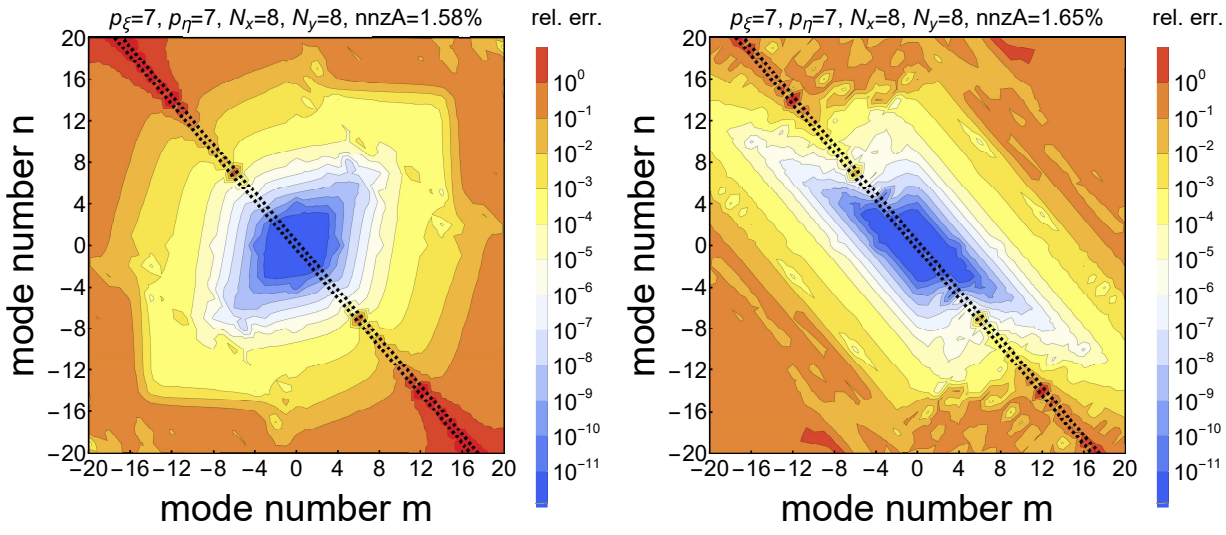


FIGURE 6.1: *Contour plot of the size of the exact eigenvalues for* $\boldsymbol{b} = (1.1659398, 1)^\top$ *with associated mode numbers up to* 20*. The black dashed line indicates the direction perpendicular to* $\boldsymbol{b}$*.*

The outline of this section is as follows: Section 6.2.1 examines the impact of the local alignment of mesh and basis. Section 6.2.2 discusses a variety of different distributions of parallel and perpendicular resolution $\mathrm{DoF}_\parallel$ and $\mathrm{DoF}_\perp$ regarding the choice of different degrees for the bases as well as different mesh resolutions. We follow with a study of the convergence by refining the mesh in Section 6.2.3. Section 6.2.4 explores the dependence of the accuracy on $\boldsymbol{b}$. The different fluxes for the mixed variational form of Chapter 3 are compared in Section 6.2.5. We close with a summary in Section 6.2.6.

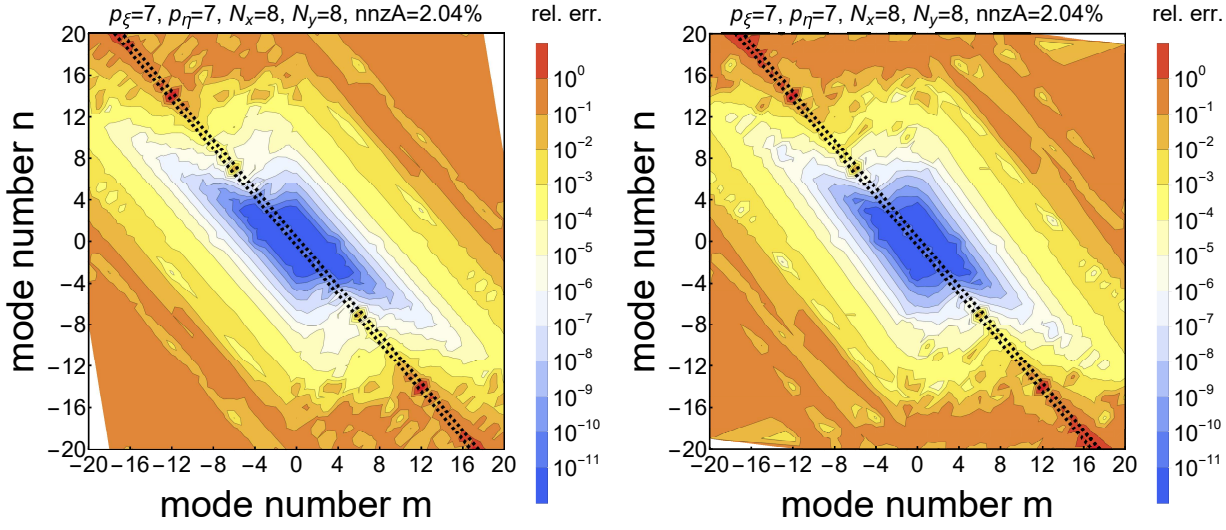
6.2.1 Impact of the local alignment

This section compares a non-aligned cartesian mesh, see Figure 3.4, with ADG. Furthermore, we examine the impact of the local alignment of the mesh by varying $\boldsymbol{b}_{\text{mesh}}$ such that the conformity condition (3.14) is fulfilled and a fully aligned mesh could be constructed. The proposed values are the two closest rational numbers for the considered case of a mesh with $N_y = 8$.
We plot the eigenvalue errors for all Fourier modes considered in the reference case. Figure 6.2 shows these as contour line plots for various choices of $\boldsymbol{b}_{\text{mesh}}$. We observe in the case of the cartesian mesh of Figure 6.2(a) that the error increases for higher mode numbers. We observe this dependence for the locally aligned cases in Figure 6.2(b)-(d) as well, but additionally a coupling of the error to the magnitude of the eigenvalue is introduced. The well resolved region is tilted towards the direction perpendicular to $\boldsymbol{b}_{\text{mesh}}$, which is the region where small eigenvalues reside for $\boldsymbol{b}_{\text{mesh}} = \boldsymbol{b}$. Figure 6.2(d) illustrates that the well resolved region is indeed tilted towards the



(a) cartesian mesh, $\boldsymbol{b}_{\text{mesh}} = (1,0)^\top$

(b) locally aligned mesh, $\boldsymbol{b}_{\text{mesh}} = \boldsymbol{b}$



(c) locally aligned mesh $\boldsymbol{b}_{\text{mesh}} = (8/7,1)^\top$

(d) locally aligned mesh $\boldsymbol{b}_{\text{mesh}} = (8/6,1)^\top$

FIGURE 6.2: *Errors of a non-aligned cartesian case, i.e.,* $\boldsymbol{b}_{\text{mesh}} = (1,0)^\top$*, ADG with* $\boldsymbol{b}_{\text{mesh}} = \boldsymbol{b}$ *and ADG using an almost aligned mesh fulfilling the conformity condition* (3.14)*, in the reference case with* $\text{DoF} = 2^{12}$ *for the constant coefficient anisotropic wave equation. In between the black dashed lines resides the band of modes with eigenvalues* $\omega^2 \leqslant \omega^2_{\text{max}} = 0.2$*.*

direction perpendicular to $\boldsymbol{b}_{\text{mesh}} = (8/6,1)^\top$. Taking a closer look at the band of modes with eigenvalues $\omega^2 \leqslant \omega^2_{\text{max}}$ as plotted in Figure 6.3, we observe that the errors of eigenvalues with large mode numbers modes are smaller by 1.5 to 2 orders of magnitude for ADG in comparison to a non-aligned discontinuous Galerkin method. For this discretization, i.e., $p_\xi = p_\eta = 7$ and $N_x = N_y = 8$,

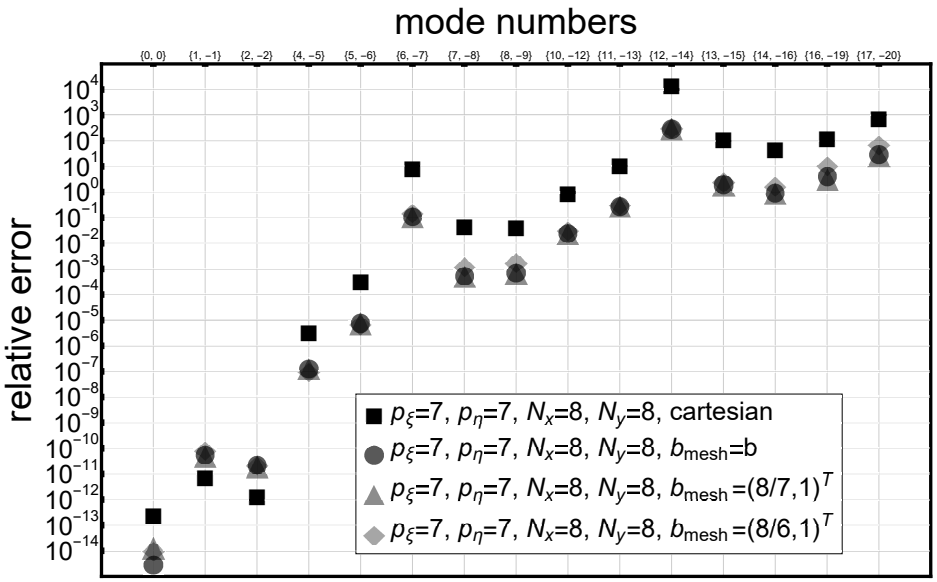


FIGURE 6.3: *Comparison of errors of a non-aligned cartesian case with ADG for different alignments in the reference case with* $\text{DoF} = 2^{12}$ *for the constant coefficient anisotropic wave equation.*

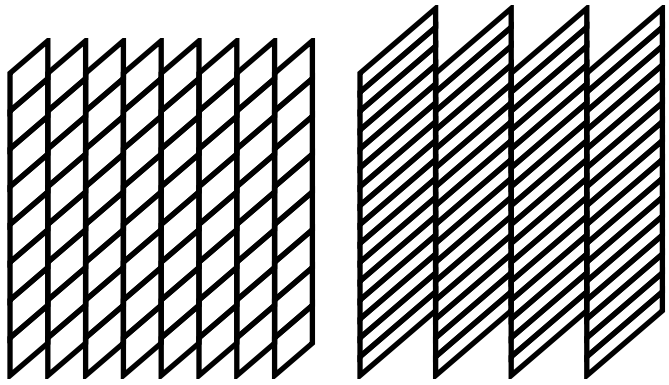
the almost aligned meshes yield roughly the same results as for $\boldsymbol{b}_{\text{mesh}} = \boldsymbol{b}$. However, we observe that for very large mode numbers, i.e., $(16, -19)$ and $(17, 20)$, ADG with $\boldsymbol{b}_{\text{mesh}} = (8/6, 1)^\top$ yields slightly worse results than the other two configurations with $\boldsymbol{b}_{\text{mesh}}$ closer aligned to $\boldsymbol{b}$. This effect is discussed in more detail in Section 6.2.2. We conclude that the mesh alignment improves the accuracy.

The white regions in the bottom left and top right corners of Figure 6.2(c) depict modes to which no eigenvalue is associated which can be seen in a variety of upcoming contour plots. These regions are characterized by large eigenvalues with large associated mode numbers which we don't aim at resolving.

6.2.2 Distribution of resolution

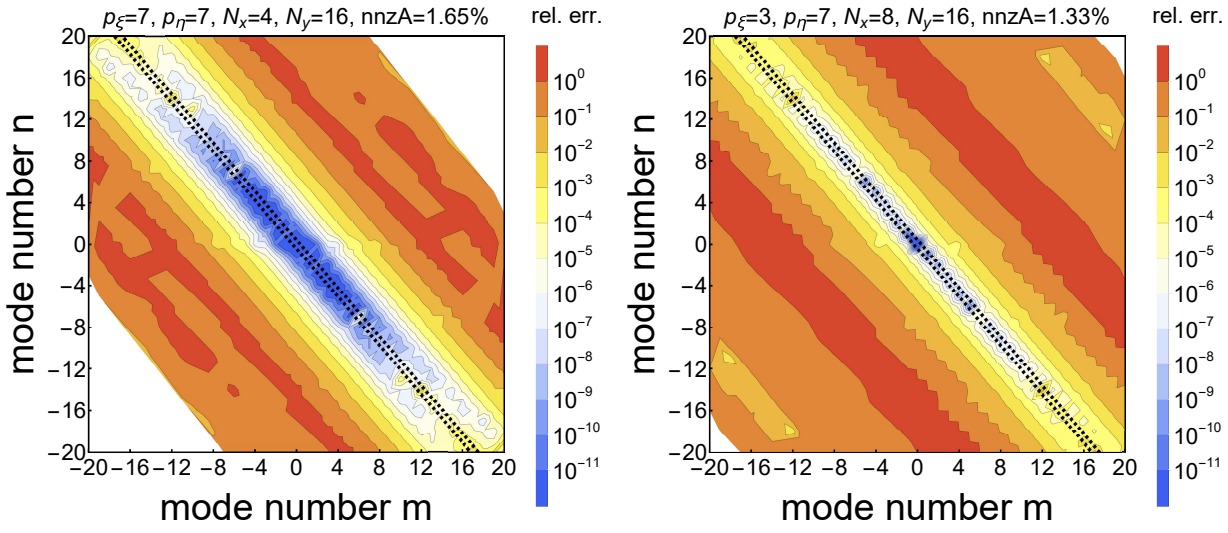
Considering the band of modes for $\omega^2 \leqslant \omega^2_{\text{max}}$ in Figure 6.2 and the distribution of exact eigenvalues in Figure 6.1, we observe that too much effort is spent on resolving mode numbers outside this band. Remembering the discussion in Section 3.1 that eigenfunctions with small eigenvalues have a small parallel gradient, we aim to distribute the resolution $\text{DoF}_\parallel$ and $\text{DoF}_\perp$ of the method such that $\text{DoF}_\perp > \text{DoF}_\parallel$. This section explores different choices for this distribution dependent on ω^2_{max} and $m_{\text{max}}, n_{\text{max}}$.

As a first measure, we keep the total number of cells constant but change the cell distribution by refining N_y and coarsening N_x as shown in Figure 6.4. The effects are shown in Figure 6.5(a) where we used the mesh of Figure 6.4(b). In comparison to Figure 6.2(b), we observe that the



(a) $N_x = 8, N_y = 8$ (b) $N_x = 4, N_y = 16$

FIGURE 6.4: *Meshes of ADG with different resolutions in the reference case.*



(a) high degree p_ξ (b) low degree p_ξ

FIGURE 6.5: *Errors of ADG for high and low degree p_ξ but same parallel resolution* $\mathrm{DoF}_\parallel$ *with* $\mathrm{DoF} = 2^{12}$*, in the reference case for the constant coefficient anisotropic wave equation. In between the black dashed lines resides the band of modes with eigenvalues* $\omega^2 \leqslant \omega^2_{\max} = 0.2$.

well-resolved region with errors smaller than 10^{-4}, indicated in blue and white gathers narrower around the interesting band of modes and extends into regions of larger mode numbers.
Taking a closer look at the band of modes in Figure 6.6, we observe that changing the cell distribution yields an increase in accuracy of 4 to 5 orders of magnitude for mode numbers larger than 4. Combining this with the results of the previous section, we gain 5.5 to 7 orders of magnitude by aligning the mesh and distribute its resolution such that $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 4$ in comparison to the

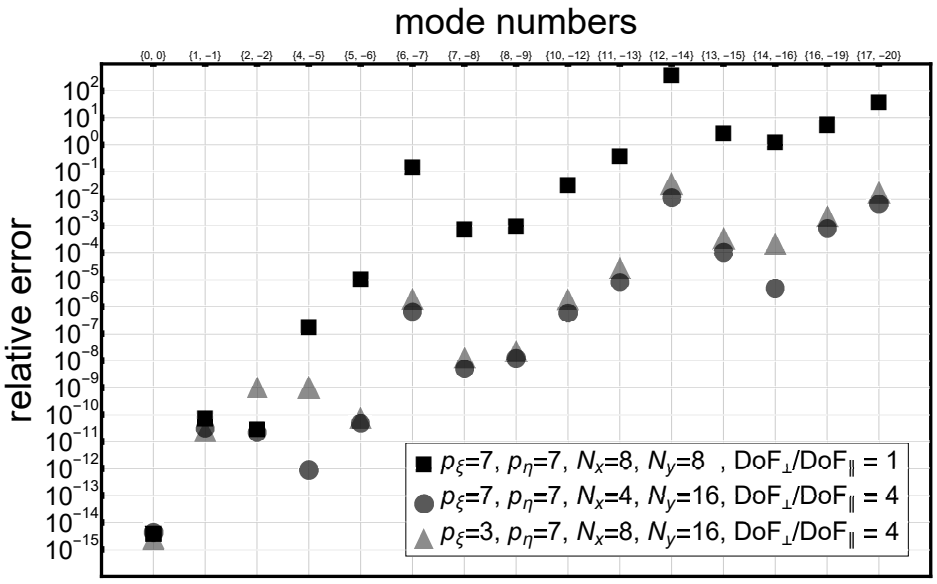


FIGURE 6.6: *Comparison of errors of ADG for high and low degree in p_ξ but same parallel resolution* $\mathrm{DoF}_\parallel$ *with* $\mathrm{DoF} = 2^{12}$*, in the reference case for the constant coefficient anisotropic wave equation. A high degree configuration with* $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 1$ *is plotted as a reference.*

non-aligned cartesian case with the same total resolution $\mathrm{DoF} = 2^{12}$.
Another way to adapt the ratio $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel$ is to modify the degree of the basis functions. Figure 6.5(b) shows a configuration with $p_\xi = 3$, while keeping $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 4$. In comparison to Figure 6.5(a) we observe that the well resolved region gathers even closer around the interesting band of modes whereas the errors related to the mode number are of the same magnitude. This is confirmed in Figure 6.6 where the errors roughly have the same order of magnitude for most modes when comparing $p_\xi = 7$ and $p_\xi = 3$ with $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 4$. However, when choosing a lower degree p_ξ, we obtain system matrices of higher sparsity with 1.33% instead of 1.65%, so we trade some accuracy for higher sparsity.
For creating a framework for the comparison of different configurations regarding the discretization parameters p_ξ, p_η, N_x and N_y, we set $\omega^2_{\max} = 0.2$ and $m_{\max} = n_{\max} = 10$. As a quality measure we opt for minimization of the biggest error among all these modes and compare only parameter combinations yielding the same number of DoF. As the eigenvalues are close to zero, we investigate both the absolute and relative error

$$A_{\omega^2_{\max},m_{\max},n_{\max}} := \max_{\substack{|m|\leqslant m_{\max},\ |n|\leqslant n_{\max} \\ \omega^2_{m,n}\leqslant\omega^2_{\max}}} \left|\widetilde{\omega}^2_{m,n} - \omega^2_{m,n}\right| \tag{6.2}$$

$$R_{\omega^2_{\max},m_{\max},n_{\max}} := \max\left\{ \max_{\substack{|m|\leqslant m_{\max},\ |n|\leqslant n_{\max} \\ 0\neq\omega^2_{m,n}\leqslant\omega^2_{\max}}} \frac{\left|\widetilde{\omega}^2_{m,n} - \omega^2_{m,n}\right|}{\omega^2_{m,n}}, \max_{\substack{|m|\leqslant m_{\max},\ |n|\leqslant n_{\max} \\ \omega^2_{m,n}=0}} \left|\widetilde{\omega}^2_{m,n}\right| \right\} \tag{6.3}$$

where $\omega^2_{m,n}$ is the exact eigenvalue and $\widetilde{\omega}^2_{m,n}$ the approximated eigenvalue associated to mode (m, n).

basis		mesh		resolution	sparsity	$\log_{10}$(errors)	
p_ξ	p_η	N_x	N_y	$\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel$	nnzA	$A_{0.2,10,10}$	$R_{0.2,10,10}$
3	3	2^5	2^5	1	0.168%	-3.21	$+0.69$
3	3	2^4	2^6	4	0.168%	-5.45	-1.61
3	3	2^3	2^7	16	0.168%	-8.18	-4.34
7	3	2^4	2^5	1	0.208%	-3.17	$+0.69$
7	3	2^3	2^6	4	0.208%	-6.42	-2.58
7	3	2^2	2^7	16	0.208%	-8.27	-4.43
3	7	2^5	2^4	1	0.333%	-7.80	-4.79
3	7	2^4	2^5	4	0.333%	-11.0	-6.89
3	7	2^3	2^6	16	0.333%	-9.82	-6.62
7	7	2^4	2^4	1	0.412%	-7.83	-4.85
7	7	2^3	2^5	4	0.412%	$\mathbf{-11.7}$	$\mathbf{-8.58}$
7	7	2^2	2^6	16	0.412%	-11.4	-7.62

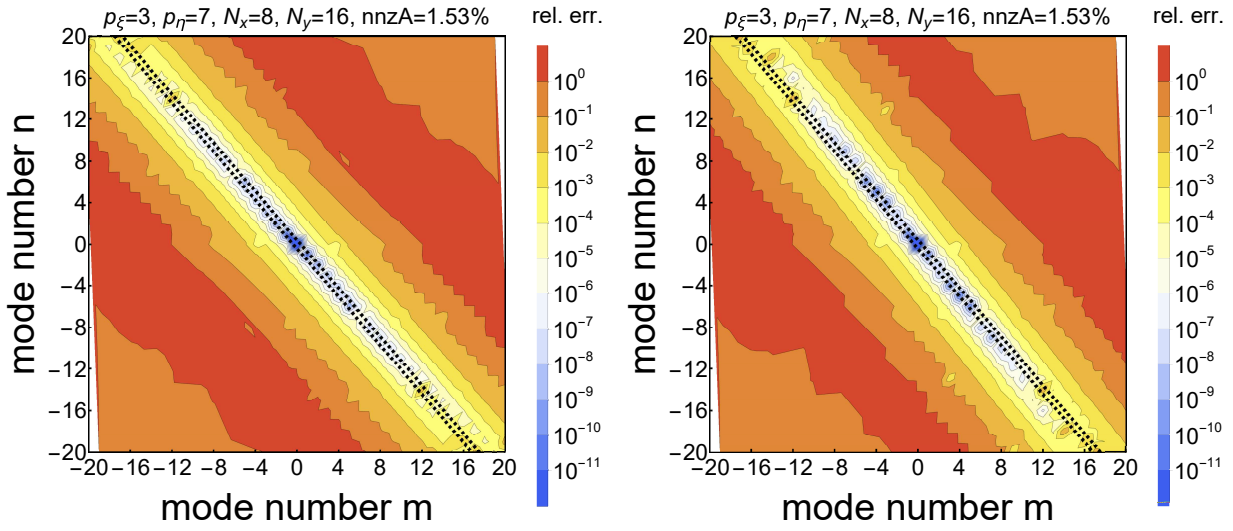
TABLE 6.1: *Results of configurations of ADG with* $\mathrm{DoF} = 2^{14}$ *and* $\omega^2_{\max} = 0.2$, $m_{\max} = n_{\max} = 10$, *in the reference case for the constant coefficient anisotropic wave equation.*

basis		mesh		resolution	sparsity	$\log_{10}$(errors)	
p_ξ	p_η	N_x	N_y	$\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel$	nnzA	$A_{0.2,10,10}$	$R_{0.2,10,10}$
3	3	2^4	2^5	2	0.336%	-3.17	$+0.69$
3	3	2^3	2^6	8	0.336%	-6.42	-2.58
7	3	2^3	2^5	2	0.415%	-3.64	$+0.21$
7	3	2^2	2^6	8	0.415%	-6.38	-2.48
3	7	2^4	2^4	2	0.665%	-7.83	-4.86
3	7	2^3	2^5	8	0.665%	-9.85	-6.14
7	7	2^3	2^4	2	0.824%	-8.53	-5.60
7	7	2^2	2^5	8	0.824%	$\mathbf{-11.2}$	$\mathbf{-7.48}$

TABLE 6.2: *Results of configurations of ADG with* $\mathrm{DoF} = 2^{13}$ *and* $\omega^2_{max} = 0.2$, $m_{max} = n_{max} = 10$, *in the reference case for the constant coefficient anisotropic wave equation.*

The results are summarized in Tables 6.1 and 6.2. First we observe that $p_\xi > p_\eta$ yields worse results than $p_\xi \leqslant p_\eta$. Hence, when aiming for systems of higher sparsity, p_η should be increased prior to

$p_{\tilde{\xi}}$. Furthermore, we observe that the errors are generally the smaller the bigger $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel$ is. However, for $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 16$ and $p_\eta = 7$, the parallel resolution becomes too small as the errors are bigger than for $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 4$ in Table 6.1. Hence, for this setting of $\omega^2_{\max}, m_{\max}, n_{\max}$ we propose to choose $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel \in \{4, 8\}$. Whether to choose $p_{\tilde{\xi}} = 3, 7$ depends on the requirements on the sparsity of the system and the desired accuracy of the results.
In the parameter choices of Tables 6.1 and 6.2, $(6, -7)$ is the mode yielding the largest relative error. This is due to the fact that for $m_{\max} = n_{\max} = 10$, $\omega^2_{6,-7} = 1.90 \times 10^{-5}$ is the smallest non-zero eigenvalue among the considered ones.
Revisiting the almost aligned meshes discussed in Section 6.2.1 for distributed resolution, the alignment of the well resolved regions perpendicular to $\boldsymbol{b}_{\mathrm{mesh}}$ is again observed, see Figure 6.7. Taking a closer look at the interesting band of modes in Figure 6.8, we observe that the different mesh alignments overall perform very similar which is particularly indicated by the errors for the mode numbers $10 \leqslant |m|, |n|$ where the largest deviations had to be expected due to the biggest deviation of the well resolved region from the fully aligned case.



(a) $\boldsymbol{b}_{\mathrm{mesh}} = (16/14, 1)^\top$ (b) $\boldsymbol{b}_{\mathrm{mesh}} = (16/13, 1)^\top$

FIGURE 6.7: *Errors of ADG using an almost aligned mesh fulfilling the conformity condition* (3.14) *with* $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 4$, $\mathrm{DoF} = 2^{12}$, *in the reference case for the constant coefficient anisotropic wave equation. In between the black dashed lines resides the band of modes with eigenvalues* $\omega^2 \leqslant \omega^2_{\max} = 0.2$.

Whether to choose an almost aligned over an aligned mesh is an intricate question. As the well resolved regions are tilted perpendicularly to $\boldsymbol{b}_{\mathrm{mesh}}$, the approximation properties of almost aligned meshes overall perform worse for a broad band of interesting modes, i.e., large $\omega^2_{\max}$, as well as for high frequency modes, i.e., large $m_{\max}, n_{\max}$, as the wrong tilt of the region affects the accuracy more and more. However, this can be thoroughly balanced with a broadening of the well resolved band by choosing a higher degree $p_{\tilde{\xi}}$ as indicated by Figure 6.5(a) or also by choosing a higher

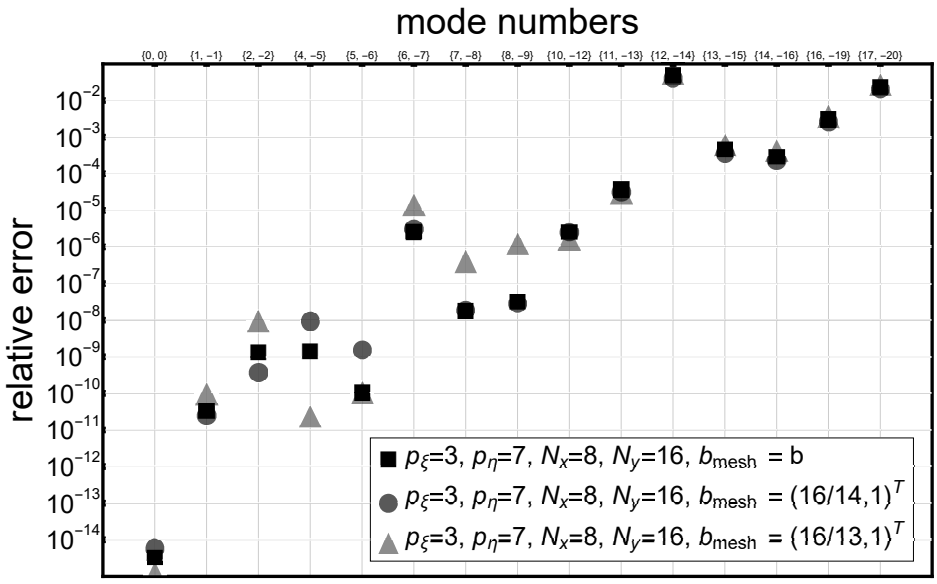


FIGURE 6.8: *Comparison of errors of ADG for different alignments with* $\text{DoF}_\perp$ / $\text{DoF}_\parallel = 4$, $\text{DoF} = 2^{12}$, *in the reference case for the constant coefficient anisotropic wave equation.*

mesh resolution which then allows the choice of a closer rational approximation of $\boldsymbol{b}$ as indicated by (3.14).

Analogously, for a broad band of interesting modes, we suggest to use a high parallel degree p_ξ in addition to the high perpendicular degree p_η. As the norm of the parallel gradient gets bigger for large ω_{max}^2 as shown in Theorem 3.1, more resolution is needed along the parallel direction. Hence, the ratio $\text{DoF}_\perp$ / $\text{DoF}_\parallel$ can be chosen large for small ω_{max}^2 and vice versa. From now on we focus on the case where $\boldsymbol{b}_{\text{mesh}} = \boldsymbol{b}$.

In the mixed form (3.45), the basis of the parallel gradient u can be chosen differently than the basis of ϕ. We investigate the impact of choosing a parallel degree of $p_\xi - 1$ for all $V_{K,U}$ if p_ξ is the parallel degree for all $V_{K,\Phi}$.

Figure 6.9 shows two sets of configurations of ADG with $\text{DoF}_\perp$ / $\text{DoF}_\parallel = 4, 8$. We observe that both configurations of the basis of $V_{K,U}$ yield similar results. As we consider reduced system matrices, a decrease in the size of the basis for u does not yield an increase in the sparsity of the system matrices. Hence, we opt for keeping the same degrees for all bases of the discretization.

6.2.3 Convergence

In this section, we analyze the behaviour of ADG when refining the locally aligned mesh and investigate the convergence of ADG for different ratios $\text{DoF}_\perp$ / $\text{DoF}_\parallel$ in the framework of the previous section and the reference case.

Figure 6.10 shows contour line plots for various mesh resolutions. When refining the parallel

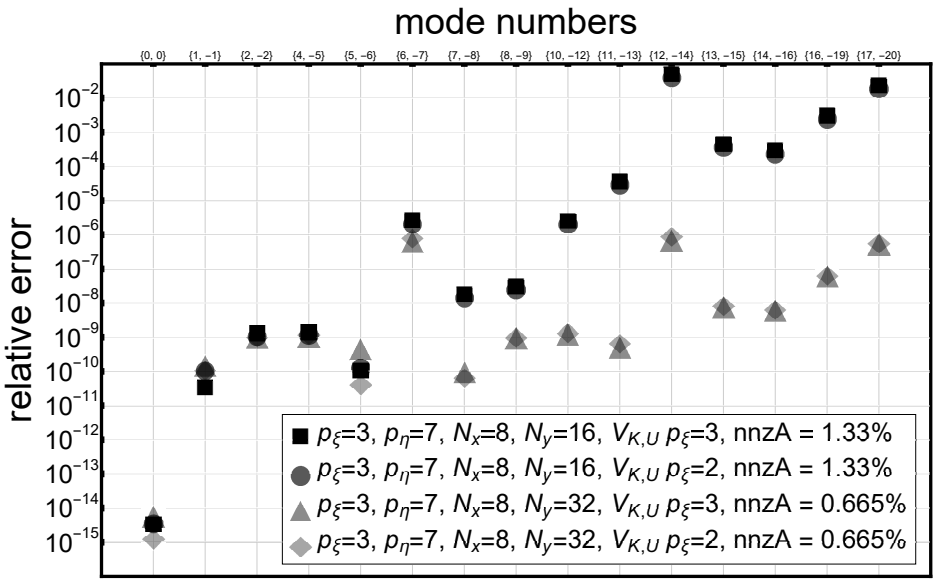
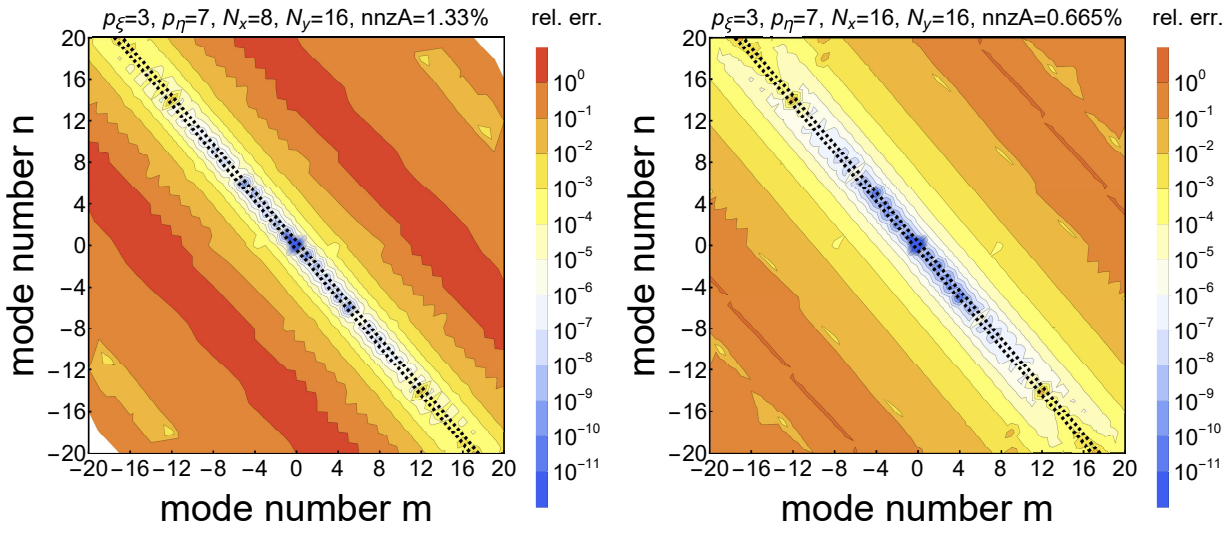


FIGURE 6.9: *Comparison of errors of ADG with $p_\xi = 3, p_\eta = 7$ for $V_{K,\Phi}$ and $p_\xi = 2, 3, p_\eta = 7$ for $V_{K,U}$, in the reference case for the constant coefficient anisotropic wave equation.*

resolution $\text{DoF}_\parallel$ by increasing N_x, we observe by comparing Figure 6.10(a) with Figure 6.10(b) that the region of well resolved eigenvalues broadens along $\boldsymbol{b}$, i.e., modes with larger eigenvalues $\omega^2_{m,n}$ are better resolved, whereas the magnitude of errors related to the maximal mode number which is indicated by the length of the well resolved region in the direction perpendicular to $\boldsymbol{b}$ roughly stays the same. Refining the perpendicular resolution $\text{DoF}_\perp$ by increasing N_y, we observe the exact opposite by comparing Figure 6.10(a) with Figure 6.10(c): The width of the well resolved region along $\boldsymbol{b}$ stays the same whereas the length in the direction perpendicular to $\boldsymbol{b}$ increases and eigenvalues associated to large mode numbers are better resolved.

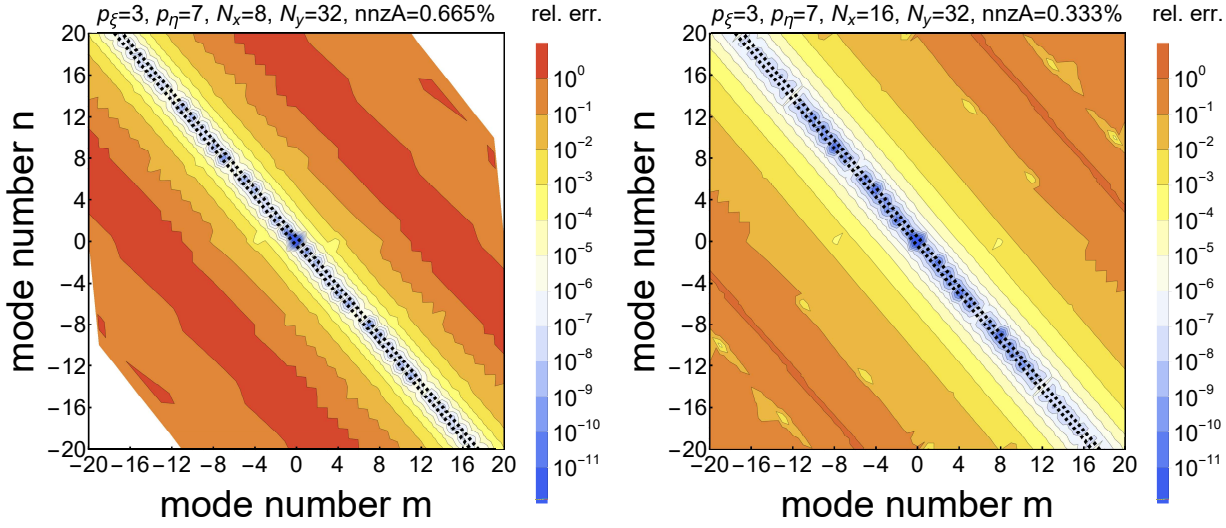
This fits well with the developed theory where we aimed at a discretization which allows to address the parallel and perpendicular resolution individually. The parallel resolution $\text{DoF}_\parallel$ mainly controls the error related to the size of the eigenvalues $\omega^2_{m,n}$ as these are related to the norm of the parallel gradient as shown in Theorem 3.1. The perpendicular resolution $\text{DoF}_\perp$ mainly addresses the error related to the mode number (m, n) of the eigenmode as the mode is highly oscillatory in the direction perpendicular to $\boldsymbol{b}$.

For closely examining the convergence behaviour of ADG, we use the framework of Section 6.2.2 and trace the errors (6.2), (6.3) for $\omega^2_{\max} = 0.2$, $m_{\max} = 20, n_{\max} = 20$ in the reference case . For one setup we use degrees $p_\xi = 3, p_\eta = 7$ and $p_\xi = 7, p_\eta = 7$ and different choices for $\text{DoF}_\perp$ / $\text{DoF}_\parallel$. The mesh is refined simultaneously in N_x and N_y by a factor of two to keep $\text{DoF}_\perp$ / $\text{DoF}_\parallel$ fixed throughout the convergence process. Figure 6.11 summarizes these results. As a second setup, we compare $p_\eta = 3, 7$ as well as a different refinement strategy of the mesh, shown in Figure 6.12. We set the search interval for these figures to [Emin,Emax] = $[-0.01, 0.4]$ to account for approximated



(a) $N_x = 8, N_y = 16$

(b) $N_x = 16, N_y = 16$



(c) $N_x = 8, N_y = 32$

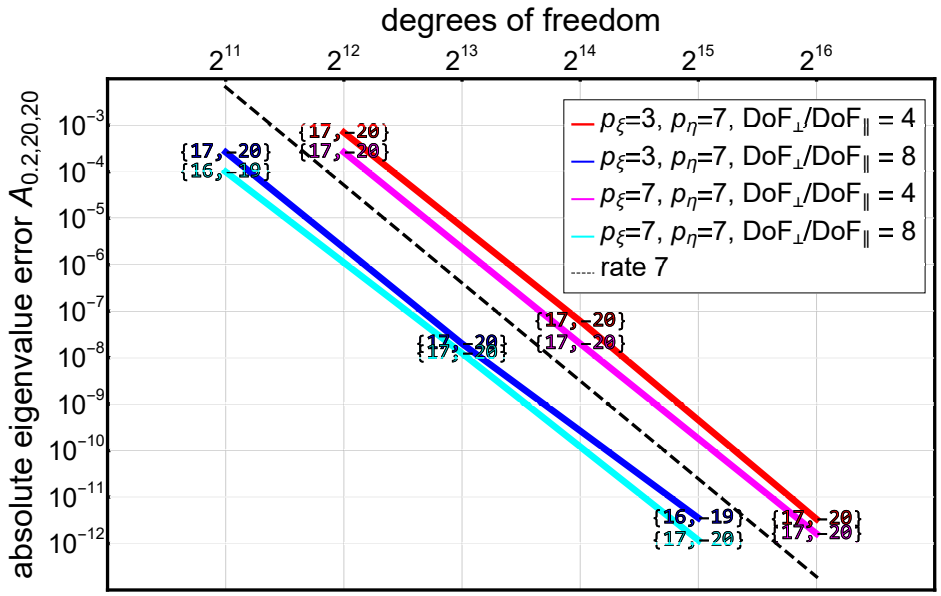
(d) $N_x = 16, N_y = 32$

FIGURE 6.10: *Errors of ADG with $p_\xi = 3, p_\eta = 7$ and different mesh resolutions, in the reference case for the constant coefficient anisotropic wave equation. In between the black dashed lines resides the band of modes with eigenvalues $\omega^2 \leqslant \omega^2_{\max} = 0.2$.*

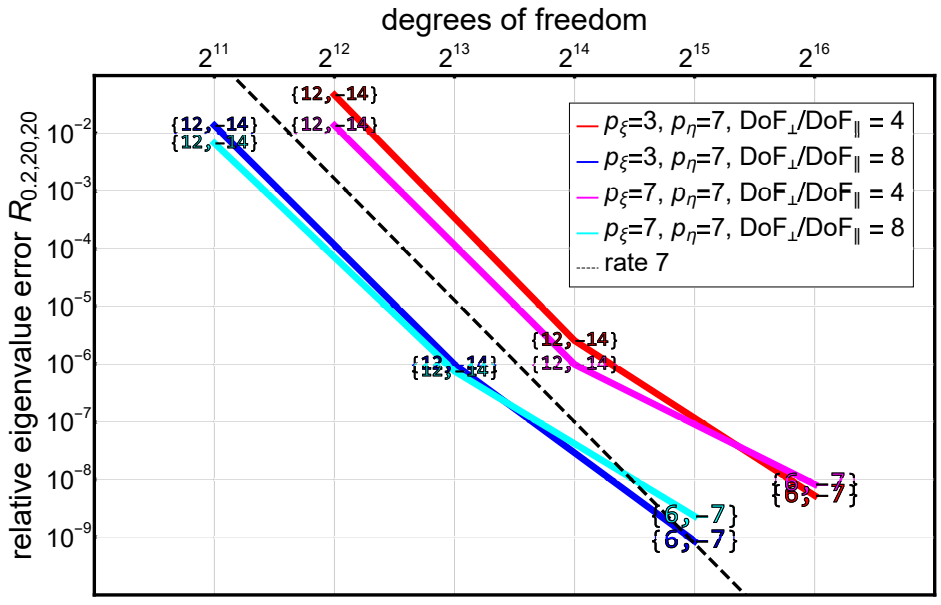
eigenvalues which might end up bigger than 0.2 due to errors of the approximation.

In Figure 6.11, we observe that the parallel degree p_ξ has minor impact on the convergence properties. For the same resolution, $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 8$ yields an improvement of 2.5 to 3 orders of magnitude compared to $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 4$.

Figure 6.12 shows that the numerically observed rate of convergence depends on p_η. This encour-

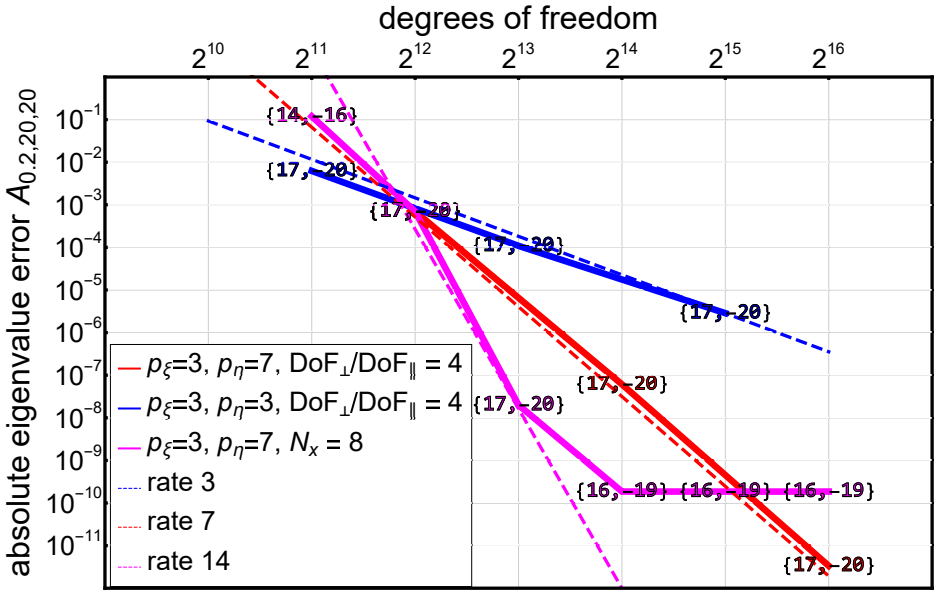


(a) $A_{0.2,20,20}$

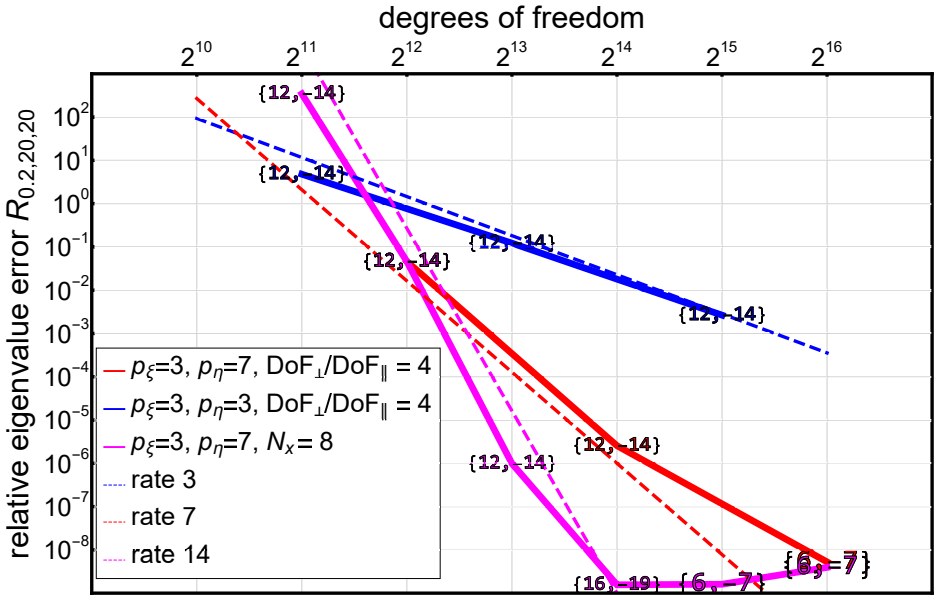


(b) $R_{0.2,20,20}$

FIGURE 6.11: *Convergence of the maximal absolute and relative errors* (6.2) *and* (6.3) *keeping* p_ξ, p_η *fixed and increasing the mesh resolution in* N_x, N_y *simultaneously, in the reference case for the constant coefficient anisotropic wave equation. The black dashed line shows a theoretical convergence rate of* $\mathcal{O}\left(\mathrm{DoF}^{-7}\right)$ *as reference. Nodes on the curves are labeled with the mode producing the largest error.*



(a) $A_{0.2,20,20}$



(b) $R_{0.2,20,20}$

FIGURE 6.12: *Convergence of the maximal absolute and relative errors* (6.2) *and* (6.3) *for different setups of ADG, in the reference case for the constant coefficient anisotropic wave equation. For a specified ratio* $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel$, *the mesh refinement is simultaneous in* N_x *and* N_y. *For* $N_x = 8$, *the refinement is done in* N_y *exclusively. Dashed lines show a theoretical convergence rate of* $\mathcal{O}\left(\mathrm{DoF}^{-\,\mathrm{rate}}\right)$ *as reference. Nodes on the curves are labeled with the mode producing the largest error.*

ages the use of a high perpendicular degree p_η as indicated by Tables 6.1 and 6.2. We further observe in Figure 6.12 that we obtain rapid convergence when keeping the parallel resolution N_x fixed and just improve the perpendicular resolution $\mathrm{DoF}_\perp$ by increasing N_y. This convergence process lasts until the given parallel resolution is too coarse to obtain better results. The discussion of Section 6.2.2 of how to choose $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel$ therefore extends. The optimal choice of $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel$ strongly depends on the total resolution DoF of the discretization.
For the simultaneous refinement of the mesh in N_x and N_y, the experimentally observed convergence is of the order of the perpendicular degree p_η of the underlying basis, i.e.,

$$A_{0.2,20,20}, R_{0.2,20,20} = \mathcal{O}\left(\mathrm{DoF}^{-p_\eta}\right) \tag{6.4}$$

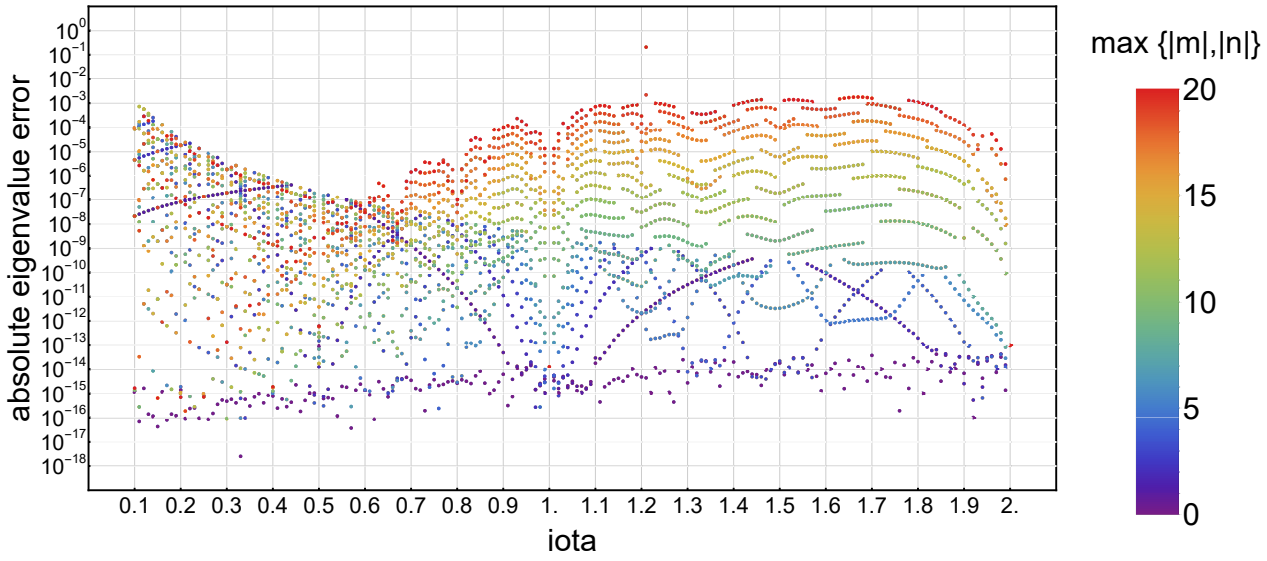
as indicated by the dashed lines in Figures 6.11 and 6.12. Note, that this property is independent of $p_{\tilde{\xi}}$. As we consider the convergence of multiple eigenvalues at once, we leave this formula as a bare observation of the numerical results rather than stating it as a general property of ADG for arbitrary $\omega^2_{\max}, m_{\max}, n_{\max}$.
We remark that the curves flatten in the case of relative errors due to the abort tolerance of the eigenvalue solver and round-off errors. The smallest eigenvalue except zero for the constant mode is approximately 1.9×10^{-5} for the mode $(6, -7)$ in the reference case. A relative error of order 10^{-8} for this eigenvalue for a total resolution of 2^{15} and 2^{16} implies absolute deviations of order 10^{-13}.

6.2.4 b-dependence

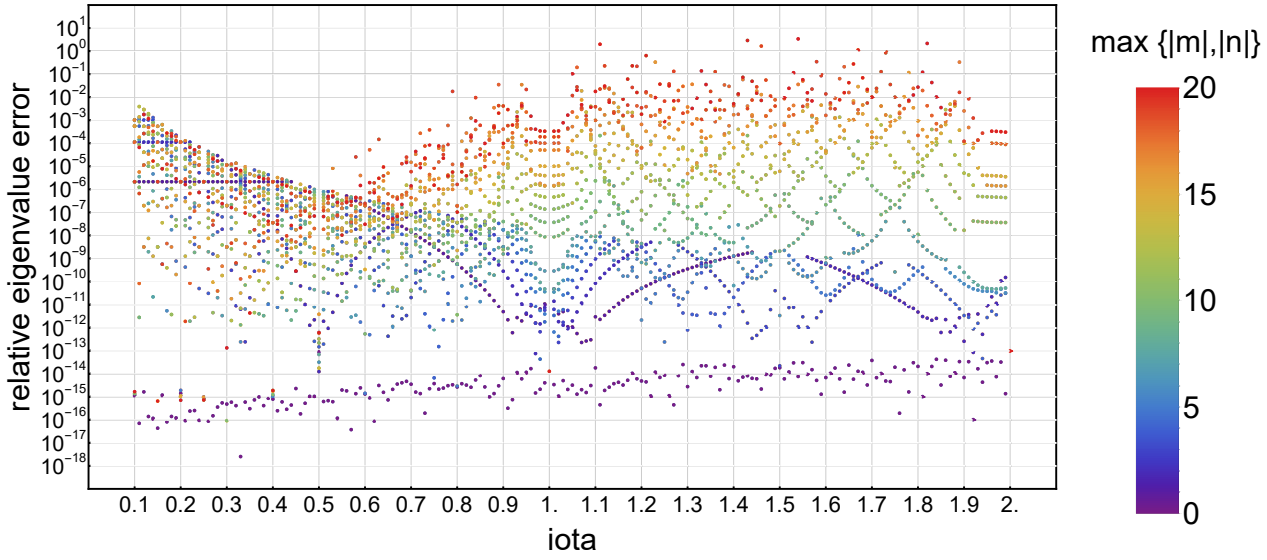
For investigating the impact of different directions of $\boldsymbol{b}$, we run ADG using $\boldsymbol{b} = (\iota, 1)^\top$ for

$$\iota \in \left\{ \frac{10}{100}, \frac{11}{100}, \dots, \frac{200}{100} \right\} \tag{6.5}$$

and $p_{\tilde{\xi}} = 3, p_\eta = 7, N_x = 8, N_y = 16$ such that $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 4$. For small ι, we deal with strongly sheared meshes. For properly evaluating the Fourier postprocessing outlined in Section 5.4 along the above-average long interfaces parallel to $\boldsymbol{b}$, we increase `degXiEval` to 90 for all ι to ensure comparability. However, this increase in evaluation points is only needed for small ι when aligning the upper and bottom interface of the locally aligned mesh.
Figure 6.13 shows absolute and relative errors for different values of ι. We observe that for $\iota \in [0.7, 2]$, the dominant error contribution is given by the maximal mode number as indicated by Figures 6.6 and 6.8 for the respective configuration of ADG. For $\iota \leqslant 0.7$ we observe that this coupling dissolves and the maximal error starts to increase for decreasing ι. This increase in the error can be traced back to the huge shear of the mesh, e.g., for $\iota = 1/10$, an aligned interface has the length $\pi\sqrt{101}/4 \approx 7.9$. However, ADG yields results with relative errors of the order $10^{-2.5}$ for this configuration. We suggest to use a mesh with aligned left and right interfaces for small values of ι as outlined in Section 3.3.3.



(a) absolute errors



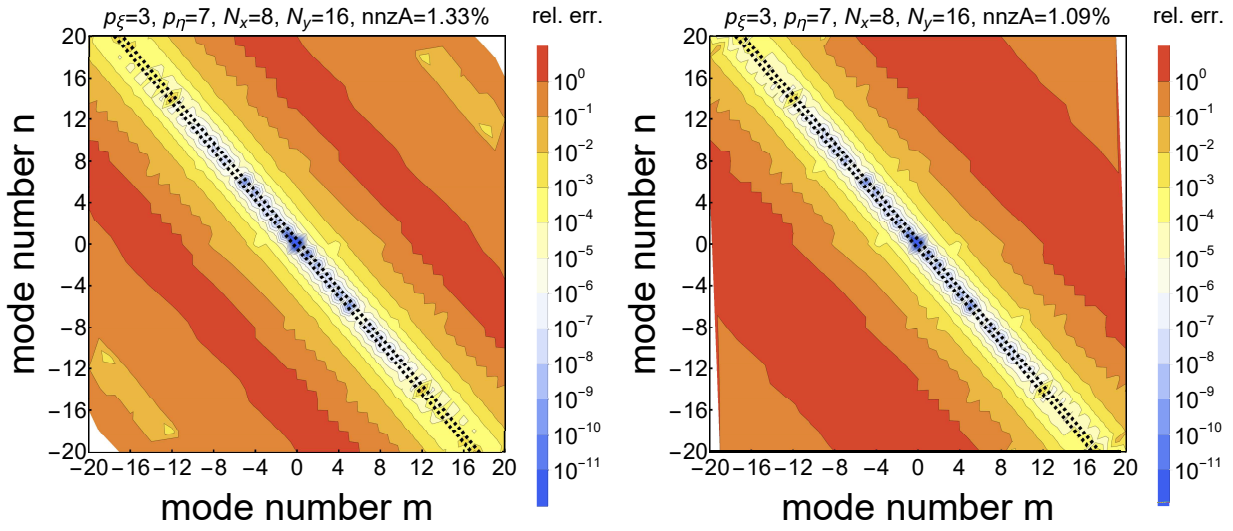
(b) relative errors

FIGURE 6.13: *Absolute and relative errors of ADG with* $p_\xi = 3, p_\eta = 7, N_x = 8, N_y = 16$ *and* $\boldsymbol{b} = (\iota, 1)^\top$ *for the constant coefficient anisotropic wave equation. Colors indicate the maximal mode number associated to the eigenvalue.*

For $\iota \geqslant 2$ we expect results of similar quality as the mesh converges to a cartesian mesh. Considering absolute errors, ADG yields results of similar order for the presented range of ι. Considering relative errors, the deviations can be larger as certain ι-configurations have exact eigenvalues very close to zero which yield large relative errors.

6.2.5 Choice of fluxes

In this section, we evaluate the choice between local discontinuous Galerkin (LDG) and Bassi-Rebay 2 (BR2) fluxes for the mixed variational form of Section 3.6.



(a) LDG (b) BR2

FIGURE 6.14: *Errors of ADG with $p_\xi = 3, p_\eta = 7, N_x = 8, N_y = 16$ and different fluxes for the mixed variational form of Section 3.6, in the reference case for the constant coefficient anisotropic wave equation. In between the black dashed lines resides the band of modes with eigenvalues $\omega^2 \leqslant \omega^2_{\max} = 0.2$.*

Figure 6.14 shows contour line plots for the two flux choices for a fixed configuration of ADG with $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 4$. We observe on the global scale that both fluxes yield structurally equivalent results. However, we remark that system matrices using BR2 fluxes exhibit a higher sparsity than for LDG fluxes as reported in Section 3.6.5 and deduced in [77].
Taking a closer look at the band of modes in Figure 6.15, we observe that BR2 fluxes yield superior results by a little less than one order of magnitude for $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 4$ whereas there is no such trend identifiable for high parallel degree p_ξ and $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 8$. In addition, BR2 exhibits system matrices of higher sparsity.
Hence, we conclude that BR2 fluxes should be used preferably.
Remark: The primal variational form of Section 3.5 was only evaluated for theoretical investigations using a prototype of the presented locally field-aligned discontinuous Galerkin method written in `MATHEMATICA`. Since the primal form therein yields results very similar to the mixed variational form using LDG fluxes, it is not incorporated in the `FORTRAN`-code.

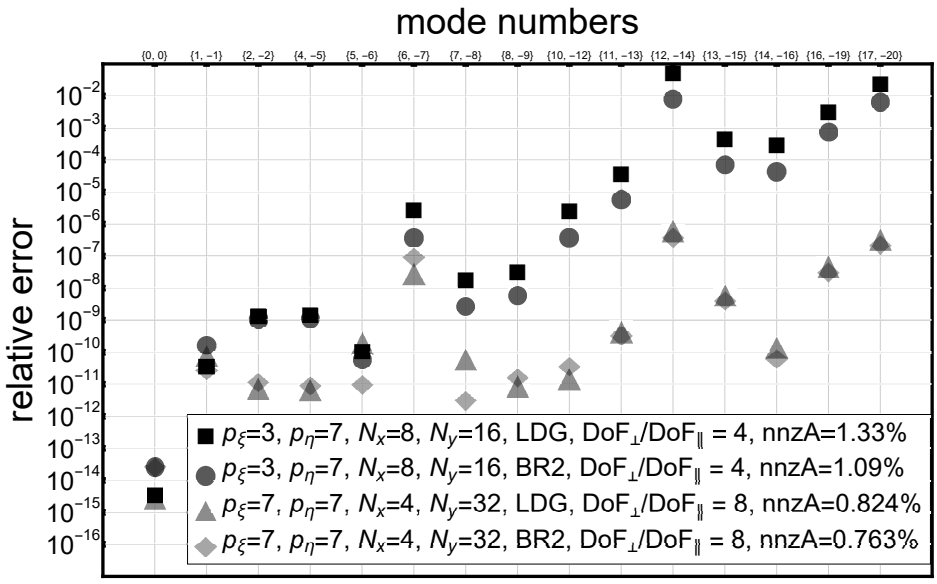


FIGURE *6.15: Comparison of errors of ADG for different fluxes of the mixed variational form of Section 3.6, in the reference case for the constant coefficient anisotropic wave equation.*

6.2.6 Summary

We summarize the results of Section 6.2. Section 6.2.1 illustrates that the local alignment of mesh and basis improves the approximation of eigenvalues by 1.5 to 2 orders of magnitude compared to a cartesian mesh, when using $N_x = N_y = 8$ and $p_\xi = p_\eta = 7$. The distribution of resolution into more $\text{DoF}_\perp$ and less $\text{DoF}_\parallel$ yields an improvement of 4 to 5 orders of magnitude, in particular for large mode numbers as discussed in Section 6.2.2. These findings are summarized in Figure 6.16. In the reference case, a ratio of $\text{DoF}_\perp$ / $\text{DoF}_\parallel = 4$ yields 5 to 6.5 orders of magnitude for mode numbers larger than 6 compared to the cartesian case with the same $\text{DoF} = 2^{12}$.
Section 6.2.1 further yields that small deviations from the local alignment yield similar results. However, larger bounds $\omega^2_{\max}$, $m_{\max}$, $n_{\max}$ narrow the range of these deviations. Section 6.2.3 shows the convergence of ADG for various configurations. Given a sufficient resolution $\text{DoF}_\parallel$, the refinement of the mesh can be done by increasing N_y exclusively. To generalize the findings of the reference case, we show in Section 6.2.4 that the accuracy of ADG remains on the same level for a broad range of $\boldsymbol{b}$. However, the quality of the results decreases when $\boldsymbol{b}$ produces strongly sheared meshes in the case of small ι which can be circumvented by aligning left and right interfaces instead of upper and bottom interfaces. Section 6.2.5 shows that using BR2 fluxes yields slightly superior results in comparison to LDG fluxes and also generates reduced system matrices of higher sparsity.

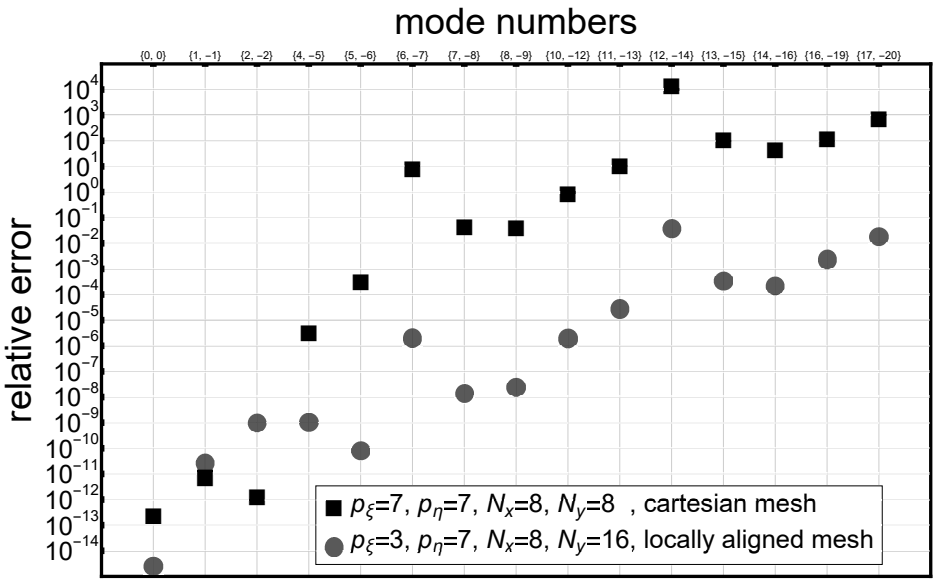


FIGURE 6.16: *Comparison of errors of a non-aligned cartesian case to ADG with* $\mathrm{DoF} = 2^{12}$*, in the reference case for the constant coefficient anisotropic wave equation.*

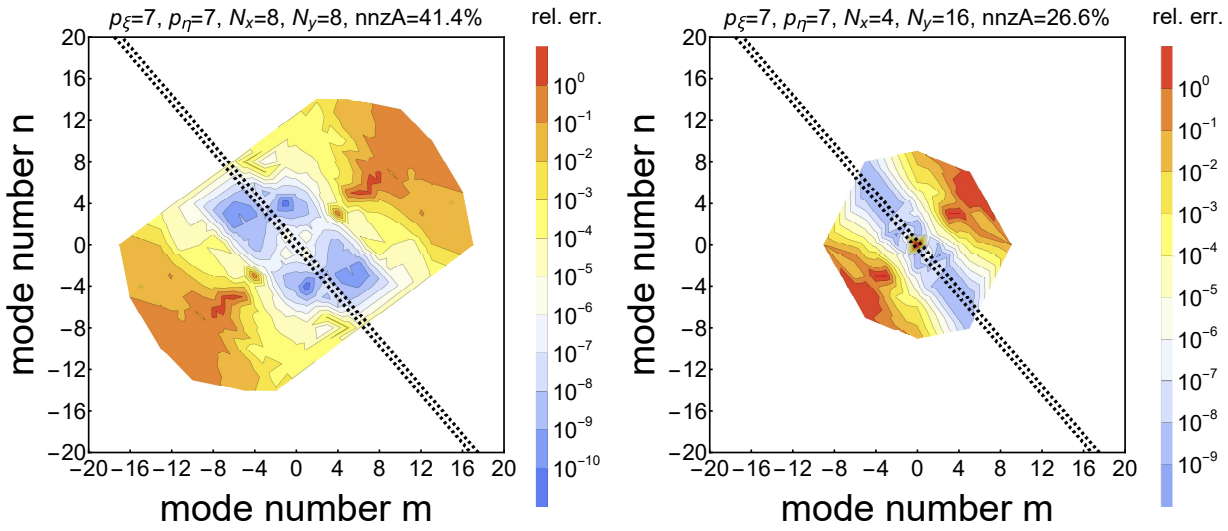
6.3 4th-order equation

In this section, we analyze ADG for the 4th-order equation (4.1) constructed in Chapter 4 and evaluate the impact of different discretization parameters as well as the dependence on the maximal eigenvalue of interest $\omega^2_{\max}$ and the maximal mode numbers $m_{\max}, n_{\max}$. This evaluation takes results of Section 6.2 into consideration as the underlying equations share structural similarities. As outlined in Section 4.1, we expect difficulties in approximating modes for which $b_\perp \cdot (m,n)^\top = 0$ or $\omega^2_{\max} \geqslant 1$. In particular, problems are expected when approximating the constant mode. Note, that the constant mode is therefore for the most part excluded in the upcoming error evaluations. Accounting for these expected issues for the eigenvalue calculation, we lower the accuracy goal of the eigenvalue solver to 10^{-7} (`feast_epsexp`$= 7$).
The outline of this section is as follows: Section 6.3.1 examines the impact of the local alignment of mesh and basis. Section 6.3.2 discusses a variety of different distributions of parallel and perpendicular resolution $\mathrm{DoF}_\parallel$ and $\mathrm{DoF}_\perp$ regarding the choice of different polynomial degrees for the bases as well as different mesh resolutions. We follow with a study of the convergence by refining the mesh in Section 6.3.3. Section 6.3.4 explores the dependence of the approximation on $\boldsymbol{b}$. We close with a summary in Section 6.3.5.

6.3.1 Impact of the local alignment

In this section, we compare a non-aligned discontinuous Galerkin method using a cartesian mesh, see Figure 3.4, with ADG.

In comparison to Section 6.2.1, we cannot rely on contour line plots such as Figure 6.2 as the approximation of eigenvalues $\omega^2 \geqslant 1$ is unreliable as outlined at the beginning of Section 6.3. Indeed, `FEAST` did not convergence for $p_\xi = p_\eta = 7$, $N_x = N_y = 8$ using a cartesian mesh when calculating all eigenvalues of the associated system matrices even when using high values for the accuracy `feast_nCP`. However, for aligned meshes we were able to produce results.



(a) $N_x = 8, N_y = 8$ (b) $N_x = 4, N_y = 16$

FIGURE 6.17: *Errors of a ADG with $p_\xi = p_\eta = 7$ and different mesh resolutions with* DoF $= 2^{12}$*, in the reference case for the* 4^{th}*-order equation. In between the black dashed lines resides the band of modes with eigenvalues* $\omega^2 \leqslant \omega^2_{\text{max}} = 0.2$.

Figure 6.17 shows contour line plots of the errors of two configurations of ADG among the domain of modes. The white regions surrounding the colored parts depict modes to which no eigenvalue is associated. Aiming at the calculation of all eigenvalues results in a reduced region of modes in which eigenvalues are associated. We experienced that the size of this domain strongly varies with different configurations of ADG. For some configurations using aligned meshes, no convergence was observed. Hence, we advise to ensure $\omega^2_{\text{max}} < 1$. Nevertheless, Figure 6.17(a) indicates the same behaviour as Figure 6.2(b), namely a coupling of the error to the magnitude of the eigenvalue as the well resolved region is tilted towards the direction perpendicular to b_{mesh}.

In the following, we set the search interval for `FEAST` to $[\texttt{Emin}, \texttt{Emax}] = [-0.01, 0.4]$ in the reference case to account for errors of eigenvalues in the desired interval $[0, 0.2]$. We obtain the results shown in Figure 6.18. First, we observe that the constant mode has a comparably large error as expected and argued at the beginning of Section 6.3. For the cartesian mesh, we increased the accuracy of `FEAST` to `feast_nCP`$= 32$. We observe that many modes still have no associated eigenvalue for the cartesian mesh. The local alignment of the mesh yields an increase of 1 to 2.5 orders of magnitude

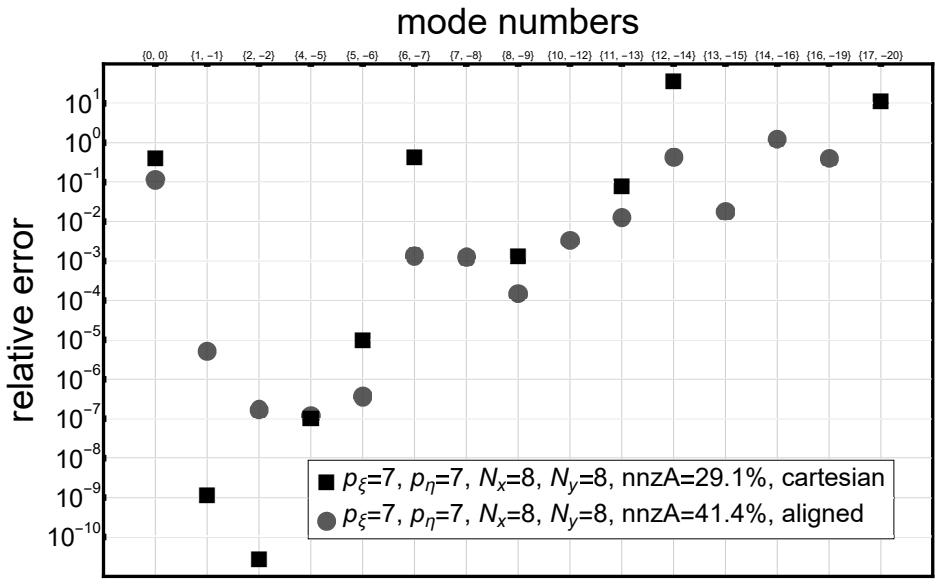


FIGURE *6.18: Comparison of errors of a non-aligned discontinuous cartesian case to ADG with* $\mathrm{DoF} = 2^{12}$*, in the reference case for the* 4^{th}*-order equation.*

for mode numbers larger than 5 and all modes but $(-17, 20)$ are associated. As we used the default accuracy `feast_nCP` $= 16$ for the locally aligned mesh, Figure 6.18 indicates that the discretization using a locally aligned mesh is easier to handle by `FEAST`.

6.3.2 Distribution of resolution

In this section, we analyze the distribution of resolution $\mathrm{DoF}_{\parallel}$ and $\mathrm{DoF}_{\perp}$. As outlined in Section 6.2.2, we aim at discretizations with $\mathrm{DoF}_{\perp} \,/\, \mathrm{DoF}_{\parallel} > 1$.
Comparing Figure 6.17(a) with Figure 6.17(b) shows that increasing the perpendicular resolution in comparison to the parallel resolution yields a narrower well-resolved region around the interesting band of modes which is the same result as in the case of the constant coefficient anisotropic wave equation as depicted by Figures 6.2(b) and 6.5(a). This provides first evidence that $\mathrm{DoF}_{\perp} \,/\, \mathrm{DoF}_{\parallel} > 1$ again emphasizes the resolution of small eigenvalues. Considering the band of modes of the reference case, distributing the resolution such that $\mathrm{DoF}_{\perp} \,/\, \mathrm{DoF}_{\parallel} = 4$ yields an increase of 2 to 5 orders of magnitude for mode numbers larger than 7 as shown in Figure 6.19. Comparing the parallel degree $p_{\tilde{\xi}}$, we observe that $p_{\tilde{\xi}} = 3$ overall yields better results than $p_{\tilde{\xi}} = 7$. Furthermore, the resulting system matrices exhibit a higher sparsity.
For further insight, we use the framework introduced in Section 6.2.1 and consider the errors (6.2), (6.3). Tables 6.3 and 6.4 summarize the results in the reference case for $\omega^2_{\max} = 0.2$, $m_{\max} = n_{\max} = 10$.

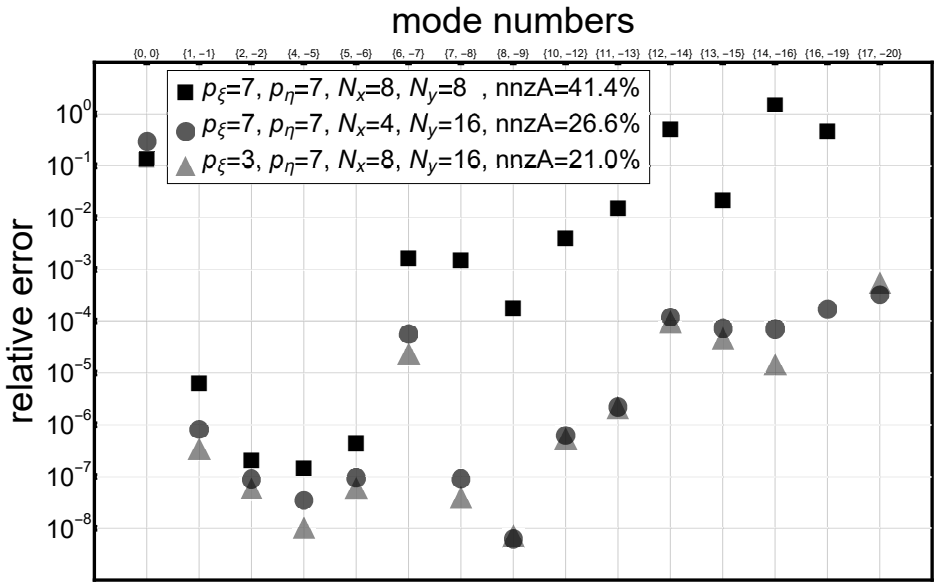


FIGURE 6.19: *Comparison of errors of ADG for high and low degree in p_ξ but same parallel resolution* $\mathrm{DoF}_\parallel$ *with* $\mathrm{DoF} = 2^{12}$*, in the reference case for the* 4^{th}*-order equation. A high degree configuration with* $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 1$ *is plotted as a reference.*

basis		mesh		resolution	sparsity	$\log_{10}$(errors)	
p_ξ	p_η	N_x	N_y	$\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel$	nnzA	$A_{0.2,10,10}$	$R_{0.2,10,10}$
3	7	2^4	2^3	1	21.5%	−4.69	−2.44
3	7	2^3	2^4	4	21.0%	**−7.97**	**−4.57**
3	7	2^2	2^5	16	20.8%	−7.05	−3.98
7	7	2^3	2^3	1	41.4%	−4.43	−2.81
7	7	2^2	2^4	4	26.6%	−7.61	−4.20
7	7	2^1	2^5	16	35.1%	−7.32	−3.40

TABLE 6.3: *Results of configurations of ADG with* $\mathrm{DoF} = 2^{12}$ *,* $\omega^2_{\max} = 0.2$*,* $m_{\max} = n_{\max} = 10$*, in the reference case for the* 4^{th}*-order equation.*

basis		mesh		resolution	sparsity	$\log_{10}$(errors)	
p_ξ	p_η	N_x	N_y	$\text{DoF}_\perp$ / $\text{DoF}_\parallel$	nnzA	$A_{0.2,10,10}$	$R_{0.2,10,10}$
3	7	2^4	2^4	2	10.7%	$\mathbf{-6.81}$	-3.45
3	7	2^3	2^5	8	10.5%	-6.64	$\mathbf{-3.47}$
7	7	2^3	2^4	2	21.0%	-6.43	-3.10
7	7	2^2	2^5	8	20.8%	-6.26	-3.32

TABLE 6.4: *Results of configurations of ADG with* $\text{DoF} = 2^{13}$, $\omega^2_{\max} = 0.2$, $m_{\max} = n_{\max} = 10$, *in the reference case for the* 4^{th}*-order equation.*

We observe an increase of 1.5 to 2 orders of magnitude for the relative error when comparing $\text{DoF}_\perp$ / $\text{DoF}_\parallel = 1$ to 4 in Table 6.3. Again, $\text{DoF}_\perp$ / $\text{DoF}_\parallel = 16$ lacks parallel resolution as the results for $\text{DoF}_\perp$ / $\text{DoF}_\parallel = 4$ are better by 0.5 to 1 order of magnitude. For producing results for the configuration $(3, 7, 2^4, 2^3)$ we used `feast_nCp` $= 20$ and for $(7, 7, 2^1, 2^5)$ we used `feast_nCp` $= 18$. $p_\xi = 3$ yields higher sparsity than $p_\xi = 7$. As hinted by Figure 6.19, we obtain worse approximations for a high parallel degree $p_\xi = 7$. The choice of the same basis for all substituted variables in (4.59) due to symmetry requirements might yield too big ansatz spaces for the resolution of these functions. Furthermore, using high degree bases without a stabilization introduces numerical dissipation.
Considering Table 6.4, no evidence whether to choose $\text{DoF}_\perp$ / $\text{DoF}_\parallel = 2, 8$ is provided. Again $p_\xi = 3$ performs superior to $p_\xi = 7$ in regard of sparsity and approximation quality.
Despite using half of the resolution of Table 6.4, the results of Table 6.3 for $\text{DoF}_\perp$ / $\text{DoF}_\parallel = 4$ are better by roughly 1 order of magnitude. This is closer examined in Section 6.3.3.
In the parameter choices of Tables 6.3 and 6.4, $(6, -7)$ is the mode yielding the largest relative error. This is due to the fact that for $m_{\max} = n_{\max} = 10$, $\omega^2_{6,-7} = 1.90 \times 10^{-5}$ is the smallest non-zero eigenvalue among the considered ones.

6.3.3 Convergence

In this section, we analyze the behaviour of ADG when refining the underlying locally aligned mesh and investigate the convergence of ADG for different ratios $\text{DoF}_\perp$ / $\text{DoF}_\parallel$.
Figure 6.20 shows errors for various mesh resolutions. Results for the constant mode are excluded. When refining the parallel resolution $\text{DoF}_\parallel$ by increasing N_x, we observe by comparing $N_x = 8, N_y = 16$ with $N_x = 16, N_y = 16$ that we obtain worse results despite having a higher resolution. Increasing the perpendicular resolution $\text{DoF}_\perp$ by increasing N_y and comparing $N_x = 8, N_y = 16$ with $N_x = 8, N_y = 32$, we observe a loss of up to 1.5 orders of magnitude for mode numbers smaller than 9 and a gain of 1.5 to 4 orders of magnitude for mode numbers larger than 10. Increasing the parallel resolution from $N_x = 8, N_y = 32$ to $N_x = 16, N_y = 32$ again worsens the

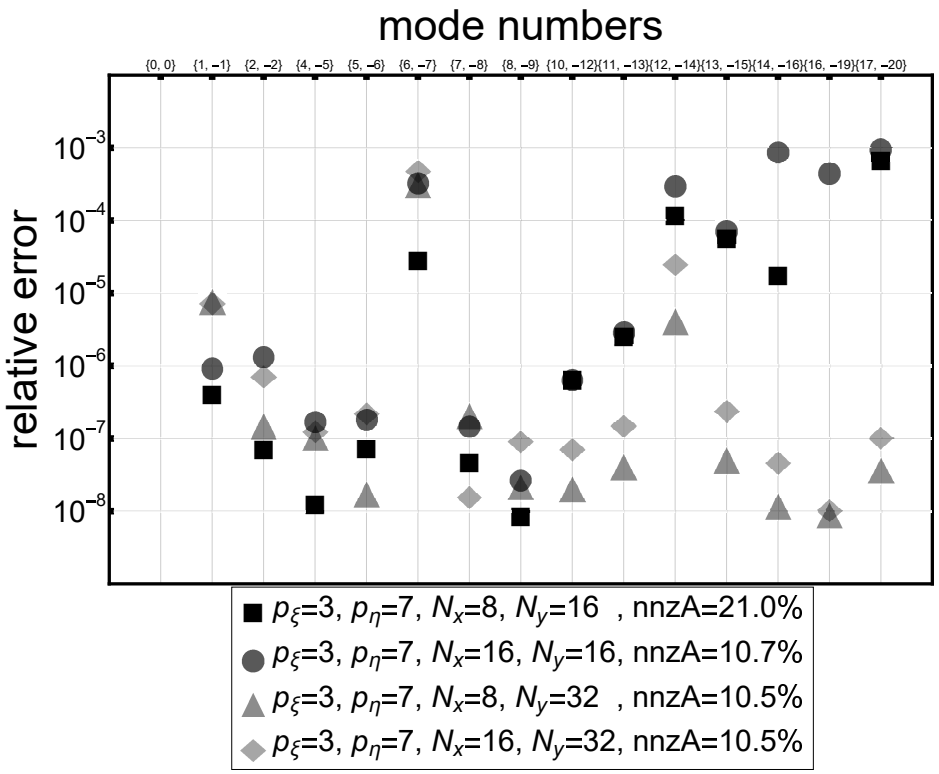


FIGURE 6.20: *Errors of ADG with $p_\xi = 3, p_\eta = 7$ and different mesh resolutions, in the reference case for the 4^{th}-order equation.*

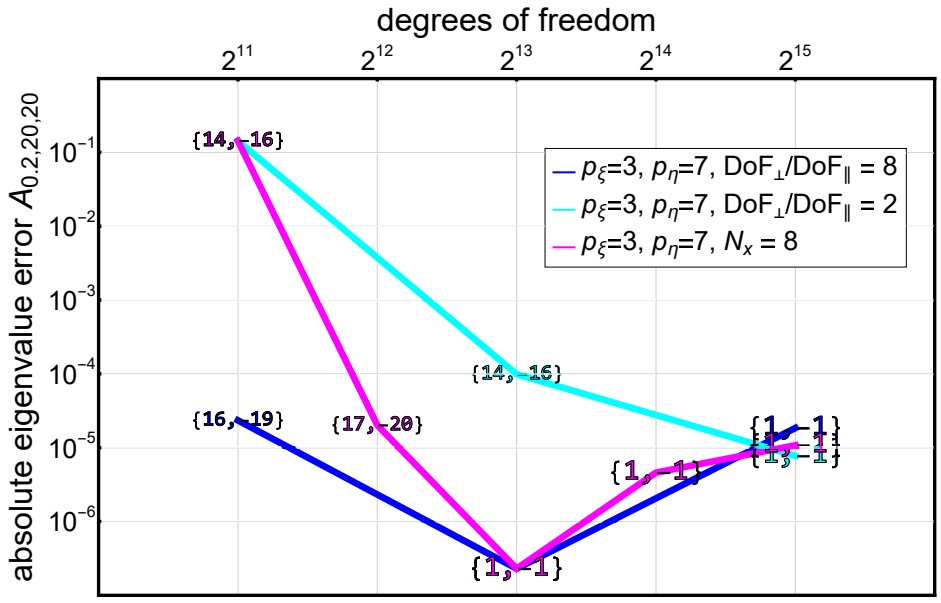
results by up to 1 order of magnitude.
This indicates that the method doesn't converge when increasing the parallel resolution.
For examining the convergence behaviour of ADG for the 4^{th}-order equation more closely, we use the framework of Section 6.2.2 and trace the errors (6.2), (6.3) with $\omega^2_{\max} = 0.2$, $m_{\max} = 20, n_{\max} = 20$ in the reference case. For one setup we use degrees $p_\xi = 3, p_\eta = 7$ and $\text{DoF}_\perp$ / $\text{DoF}_\parallel = 2, 8$. The mesh is refined simultaneously in N_x and N_y by a factor of two to keep $\text{DoF}_\perp$ / $\text{DoF}_\parallel$ fixed throughout the convergence process. As a second setup, we keep the parallel resolution fixed and refine the mesh by increasing N_y. The results are summarized in Figure 6.21.
We observe that we obtain no convergence for either configuration.

6.3.4 b-dependence

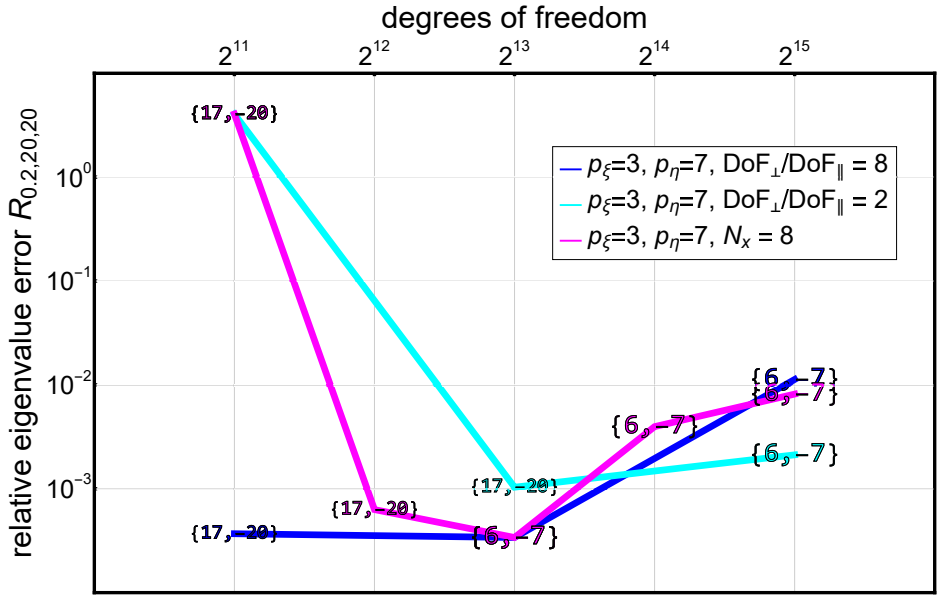
For investigating the impact of different directions of $\boldsymbol{b}$, we run ADG using $\boldsymbol{b} = (\iota, 1)^\top$ for

$$\iota \in \left\{ \frac{10}{100}, \frac{11}{100}, \dots, \frac{200}{100} \right\} \tag{6.6}$$

and $p_\xi = 3, p_\eta = 7, N_x = 8, N_y = 16$ such that $\text{DoF}_\perp$ / $\text{DoF}_\parallel = 4$. For small ι, we deal with strongly sheared meshes. For properly evaluating the Fourier postprocessing outlined in Section 5.4 along



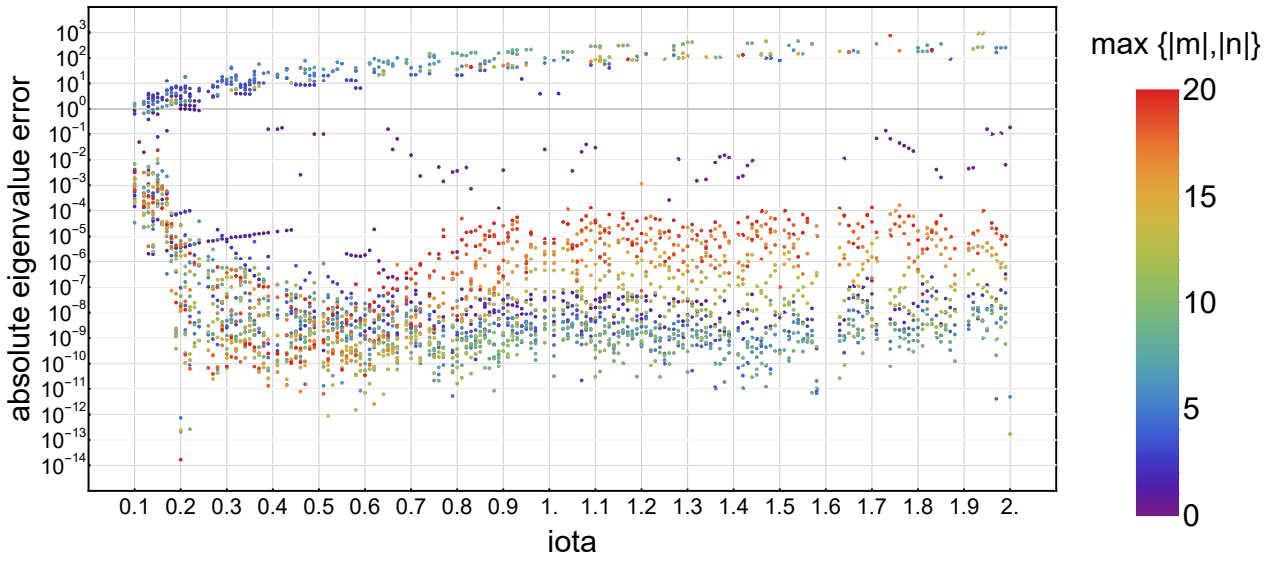
(a) $A_{0.2,20,20}$



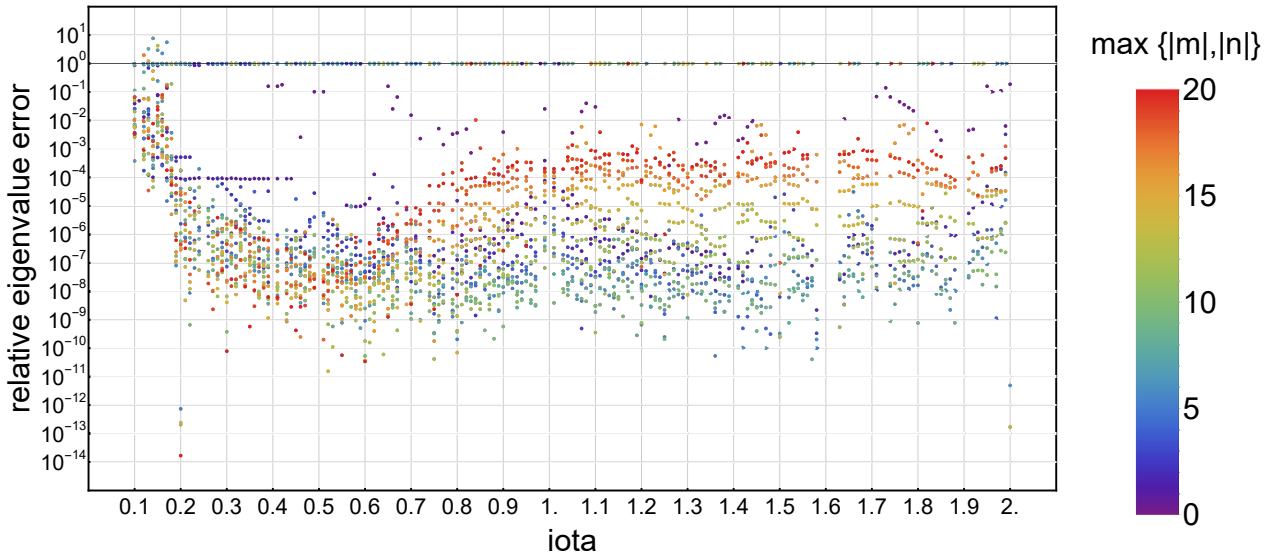
(b) $R_{0.2,20,20}$

FIGURE 6.21: *Convergence of the maximal absolute and relative errors* (6.2) *and* (6.3) *keeping* p_ξ, p_η *fixed, in the reference case for the* 4th*-order equation. For fixed* DoF$_\perp$ / DoF$_\parallel$*, the mesh resolution is increased in* N_x, N_y *simultaneously. For fixed* N_x*, the mesh is refined in* N_y *exclusively. Nodes on the curves are labeled with the mode producing the largest error.*

the above-average long interfaces parallel to $\boldsymbol{b}$, we increase `degXiEval` to 90 for all ι to ensure comparability. However, this increase in evaluation points is only needed for small ι when aligning the upper and bottom interface of the locally aligned mesh.



(a) absolute errors



(b) relative errors

FIGURE 6.22: *Absolute and relative errors of ADG with* $p_\xi = 3, p_\eta = 7, N_x = 8, N_y = 16$ *and* $\boldsymbol{b} = (\iota, 1)^\top$ *for the* 4^{th}*-order equation. Colors indicate the maximal mode number associated to the eigenvalue.*

Figure 6.22 shows absolute and relative errors for different values of ι. We observe that the full ι-spectrum exhibits falsely associated modes which are indicated by dots with relative error

10^0 in Figure 6.22(b). This can be traced back to FEAST reporting that the resulting subspace of eigenvectors is not biorthonormal. Hence, eigenvectors might not be accurately associated to a mode.
The purple dots in $[0.5, 2] \times [10^{-5}, 10^0]$ in Figure 6.22(a) are associated to the $(0, 0)$ mode. As remarked at the beginning of Section 6.3, we observe difficulties when approximating the constant mode.
For some ι, FEAST did not converge which is indicated by missing data points. This is in particular the case for $\iota \approx 1.6$ and fractions with small numerator and denominator. In the former case, running FEAST with a higher accuracy or by adapting the expected number of eigenvalues in the search interval M0 can fill in the gaps as remarked in Section 5.5.1. In the latter case, it holds $\boldsymbol{b}_\perp \cdot (m, n)^\top = 0$ for many (m, n) in the considered range. Hence, we expect difficulties for solving the associated eigenvalue system.
For $\iota \in [0.7, 2]$, we observe as in Section 6.2.4 that the dominant error contribution is given by the maximal mode number as indicated by Figures 6.19 and 6.20 for the respective configuration of ADG. For $\iota \leqslant 0.7$ we observe that this coupling dissolves and the maximal error starts to increase for decreasing ι. This increase in the error can be traced back to the huge shear of the mesh, e.g., for $\iota = 1/10$, an aligned interface has the length $\pi\sqrt{101}/4 \approx 7.9$. We suggest to use a mesh with aligned left and right interfaces for small values of ι as outlined in Section 3.3.3.
For $\iota \geqslant 2$ we expect results of similar quality as the mesh converges to a cartesian mesh. For converged modes, we observe that the magnitude of the error is independent of $\boldsymbol{b}$ for $\iota \geqslant 0.2$ and strongly worsens for smaller values.

6.3.5 Summary

We summarize the results of Section 6.3. First, we remark that for all results where convergence of the eigenvalue solver was achieved, FEAST warned that the resulting subspace of eigenvectors is not biorthonormal. For some settings of input parameters, FEAST did not converge. The inclusion of SLEPC [88] as another eigenvalue solver did not improve the convergence properties of the eigenvalue solver. It is known that generalized eigenvalue problems with semidefinite matrices are hard to solve [89]. Furthermore, the process of deriving the 4^{th}-order equation in Section 2.5 is very coarse and important features of the reduced MHD equations might not be well represented or the equation itself is too artificial. The problem of the semidefinite right hand side can be tackled by including metric terms as in the derivation of the anisotropic wave equation in Section 2.6 to diminish the space of zero eigenvalues. Furthermore, the extension to three dimensions and inclusion of radial boundary conditions as for the structurally equivalent reduced MHD shear Alfvén wave equation (2.69) might assist in the evaluation. However, the previous sections illustrate the following:
Section 6.3.1 shows that the local alignment of mesh and basis improves the approximation of eigenvalues by 1 to 2.5 orders of magnitude. Furthermore, FEAST is able to resolve more modes

when using an aligned mesh. The distribution of resolution yields another 2 to 5 orders of magnitude, in particular for large mode numbers as discussed in Section 6.3.2. These results are summarized in Figure 6.23. In the reference case, a ratio of $\mathrm{DoF}_{\perp} / \mathrm{DoF}_{\parallel} = 4$ yields an improvement of 4 to 5.5 orders of magnitude for mode numbers larger than 6 when compared to the cartesian case with the same $\mathrm{DoF} = 2^{12}$.

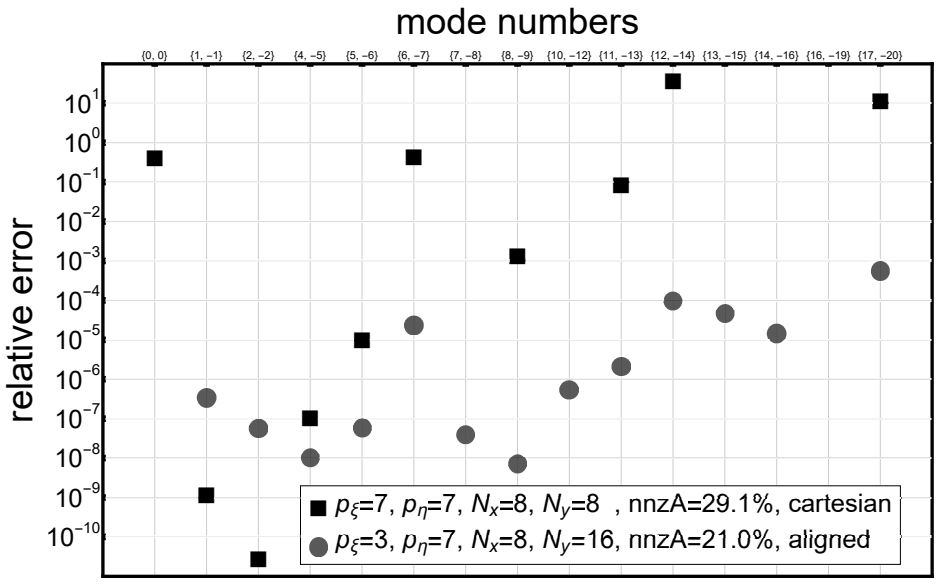


FIGURE 6.23: *Comparison of errors of a non-aligned cartesian case to ADG for* $\mathrm{DoF} = 2^{12}$*, in the reference case for the* 4^{th}*-order equation.*

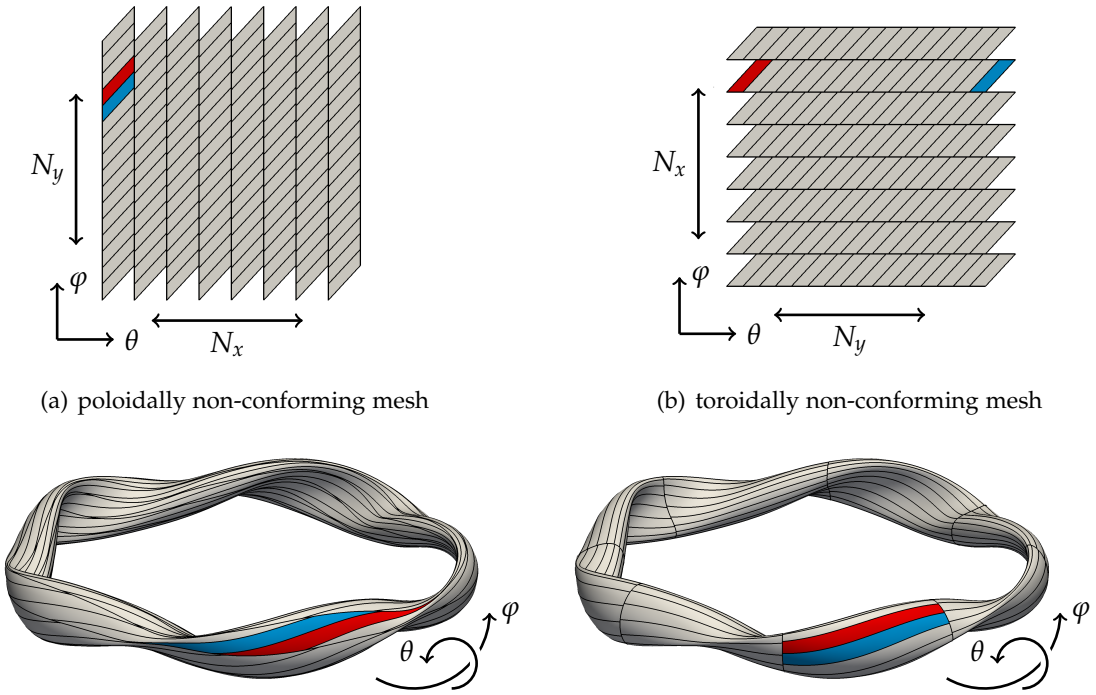
Section 6.3.3 shows that the method is not converging. The magnitude of the results of ADG stays the same for a broad range of $\boldsymbol{b}$ as discussed in Section 6.3.4. The quality of the results decreases when $\boldsymbol{b}$ yields strongly sheared meshes which can be circumvented by aligning left and right interfaces instead of upper and bottom interfaces. However, modes might be wrongly associated and problems occur for $\boldsymbol{b} = (\iota, 1)^{\top}$ where ι is a fraction with small numerator and denominator. We conclude that the theoretically discovered difficulties when dealing with the 4^{th}-order equation (4.1) prevail in the numerical method. Nevertheless, for small accuracy goals, the local alignment approach of ADG provides results accurate to 3 orders of magnitude in the reference case for $\mathrm{DoF} = 2^{12}$.

6.4 Anisotropic wave equation for MHD equilibria

In this section, we analyze ADG for solving the anisotropic wave equation (3.110) with metric terms from a flux surface of a three-dimensional MHD equilibrium.
The results of this section are obtained using the `VMEC`-equilibrium, see Section 5.5.3, of the W7-X high-mirror case [90, Table IV], with units normalized in tesla and meter.
Throughout this section, we use $p_\xi = 3$, $p_\eta = 7$ and `[Emin,Emax]` = $[-0.01, 0.4]$ for finding eigenvalues with $\omega^2 \leqslant \omega^2_{\max} = 0.2$. Furthermore, we set $m_{\max} = n_{\max} = 25$ for associating mode numbers (m, n). m denotes the poloidal and n the toroidal mode number. As no analytic results exist for the eigenvalues of the given MHD equilibrium, we present the discrete results for the eigenvalues and their associated mode numbers over the normalized flux surface coordinate s instead of errors. We leave out flux surfaces close to the magnetic axis, consider $s \in [0.1, 1]$ and choose a stepsize of 0.01 in s-direction. On each flux surface at position s, we solve the two-dimensional eigenvalue problem of the anisotropic wave equation with the corresponding metric terms of the flux surface.
The outline of this section is as follows: We first investigate the impact of aligning upper and bottom interfaces in comparison to left and right interfaces and compare to a non-aligned cartesian mesh in Section 6.4.1. In Section 6.4.2, we show the convergence of the eigenvalue spectrum when increasing the mesh resolution. A comparison with the codes CONTI and CKA regarding the eigenvalue spectrum is performed in Section 6.4.3. The results are summarized in Section 6.4.4.

6.4.1 Choice of cell alignment

So far, we discussed the alignment of upper and bottom and left and right interfaces for constant $\boldsymbol{b} = (\iota, 1)^\top$, where the accuracy just depends on ι and the ratio N_x/N_y. Now, including metric terms and three-dimensional geometries, we reevaluate the choice of the cell alignment.
Figure 6.24 shows locally aligned meshes with aligned upper and bottom and aligned left and right interfaces mapped onto a flux surface of the MHD equilibrium. Aligned upper and bottom interfaces yield non-conforming interfaces in poloidal direction θ whereas aligned left and right interfaces yield non-conforming interfaces in toroidal direction φ. For aligned upper and bottom interfaces, N_x is associated to poloidal and N_y is associated to toroidal resolution. For aligned left and right interfaces, N_x is associated to toroidal and N_y is associated to poloidal resolution. The mesh discretization is chosen such that N_x is the parallel and N_y the perpendicular mesh resolution as summarized in Section 5.6.
In the case of the W7-X-like equilibrium, a five-star symmetry is used [91], meaning that the geometry is defined on a fifth of the toroidal direction and then repeated periodically as shown in Figure 6.24(c),(d). Such a toroidal subsection which uniquely defines the geometry is called a field period. Hence, the number of field periods of the W7-X-like equilibrium is $M = 5$. Simulations can be carried out only on a single field period, but then boundary conditions have then to be taken into account or only modes with a toroidal mode number of a multiple of M can be considered.



(a) poloidally non-conforming mesh

(b) toroidally non-conforming mesh

(c) mapped mesh, poloidally non-conforming

(d) mapped mesh, toroidally non-conforming

FIGURE 6.24: *Meshes with $N_x = 8$, $N_y = 16$ and toroidally or poloidally non-conforming interfaces for ADG on a flux surface of a W7-X-like MHD equilibrium. For this representation in N_x, N_y, field periods were not taken into account. The mapping between logical and physical domain is highlighted for two colored cells.*

For simplicity, we choose to simulate the full domain and include M in the total cell count. For a given field period resolution (N_x, N_y), the setup of ADG uses (MN_x, N_y) cells for toroidally non-conforming meshes and (N_x, MN_y) for poloidally non-conforming meshes.
For a selection of mode numbers, Figure 6.25 compares results of toroidally and poloidally non-conforming meshes for $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 4$ within a field period. Note, that due to the inclusion of metric terms, eigenfunctions of the anisotropic wave equation with metric terms are no longer Fourier modes. We recall that the Fourier postprocessing of Section 5.4 associates the eigenvalue to the Fourier mode with the largest amplitude. In the transition from one flux surface to another, this association may shift from one mode to another. Therefore, when tracing a single Fourier mode along the normalized flux surface coordinate, jumps may appear within its eigenvalue spectrum. The represented modes were selected such that multiple jumps are present and the whole eigenvalue spectrum of $[0, 0.2]$ is represented.
In Figure 6.25, modes are distinguished by the shape of markers. The usage of poloidally and toroidally non-conforming meshes is indicated by the filling of markers where empty markers represent poloidal non-conformity. Results of the coarse mesh are depicted in red and refined

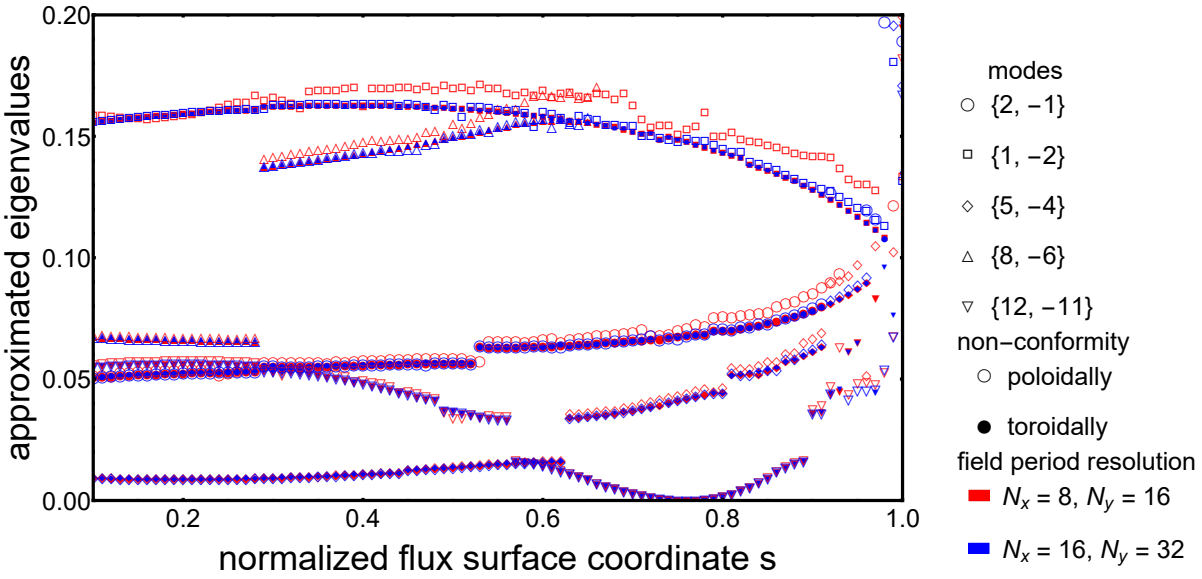


FIGURE 6.25: *Results of ADG for $p_\xi = 3, p_\eta = 7$ for the anisotropic wave equation with metric terms. The two cell alignments are compared for a coarse (red) and fine (blue) mesh. Modes are indicated by their shape and toroidal and poloidal non-conformities by fillings.* $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 4$ *within a field period is fulfilled.*

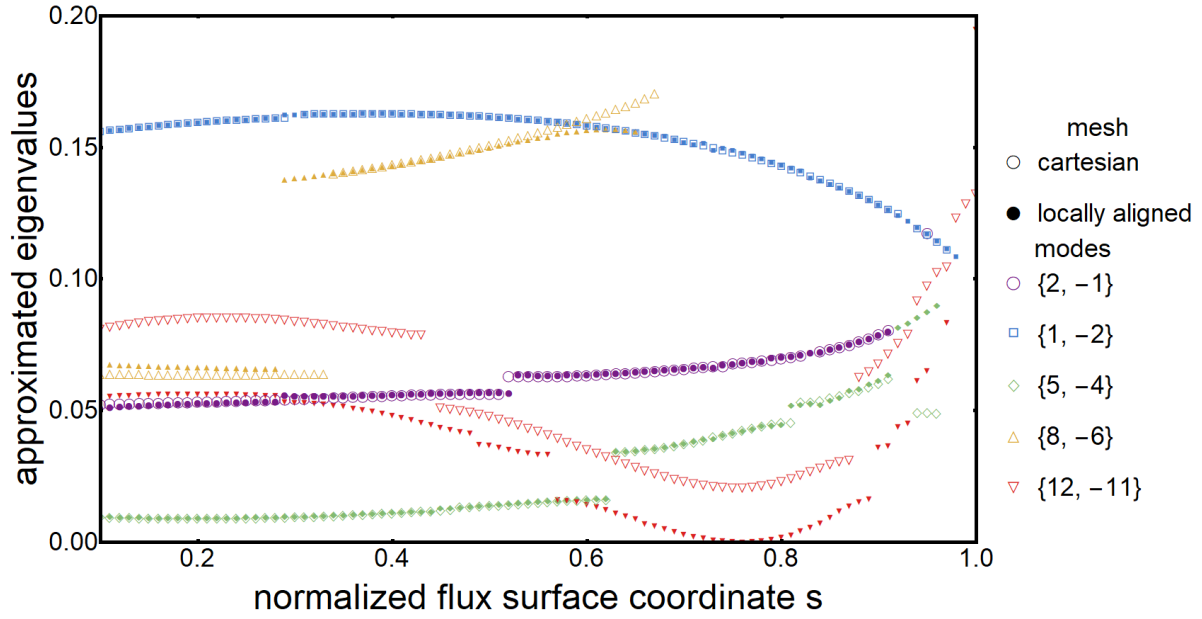


FIGURE 6.26: *Results of ADG with $p_\xi = 3, p_\eta = 7, N_x = 8, N_y = 16$ using toroidally non-conforming interfaces in comparison to a non-aligned cartesian case with $p_\xi = 7, p_\eta = 7, N_x = 8, N_y = 8$ for the anisotropic wave equation with metric terms. Both fulfill* $\mathrm{DoF} = 2^{12}$ *within a field period. Modes are indicated by shape and color, meshes are indicated by fillings.*

mesh results in blue. We observe that some eigenvalues coincide for coarse and fine mesh which means that these eigenvalues are already converged. Further, we observe that more eigenvalues are already converged for the toroidally non-conforming mesh. In particular, the poloidally non-conforming coarse mesh shows deviations for the modes $(8, -6)$ for $s \geqslant 0.3$, $(1, -2)$ for $s \geqslant 0.25$ and $(2, -1)$ for $s \geqslant 0.6$. Overall, all results differ by a larger margin due to strong metrics at the boundary of the MHD equilibrium for $s \geqslant 0.95$. We therefore conclude superior approximation properties when using toroidally non-conforming meshes and focus on these in the following.
In Figure 6.26, we compare ADG using a toroidally non-conforming mesh with a non-aligned cartesian mesh using the same number of DoF for the selection of modes of Figure 6.25. We observe that both agree for small mode numbers. However, the higher the mode number, the more the results differ for the cartesian mesh. This is indicated by the modes $(8, -6)$ and $(12, -11)$ which suggests that eigenvalues are not converged yet for high mode numbers in the cartesian case. Compared to the poloidally non-conforming mesh investigated in Figure 6.25, the cartesian mesh produces lager deviations.

6.4.2 Convergence

We investigate the convergence of ADG for the anisotropic wave equation with metric terms using toroidally non-conforming meshes.

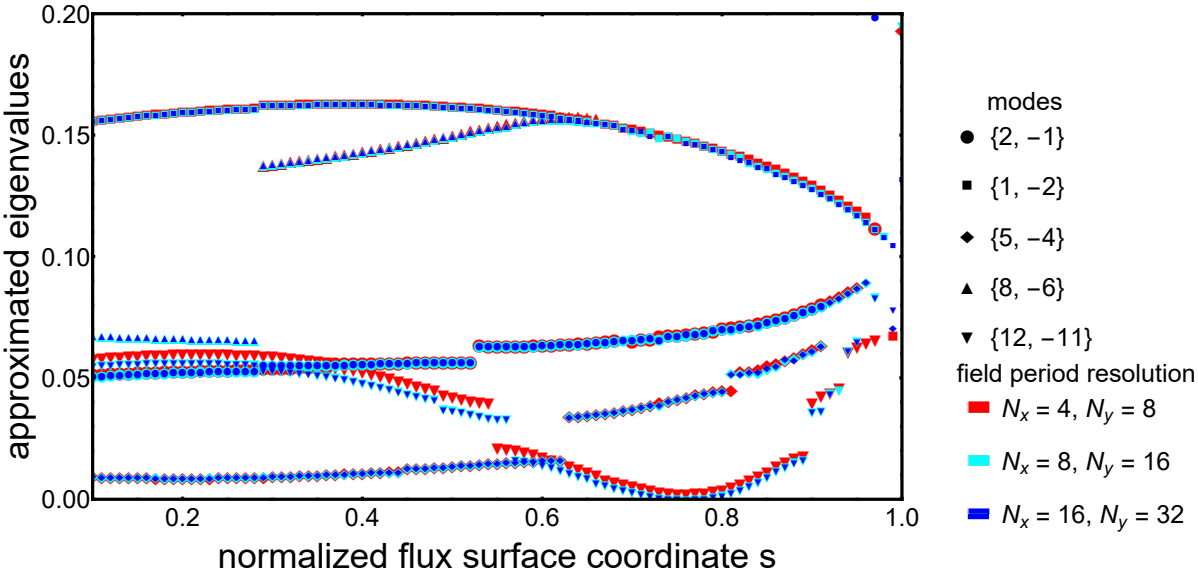
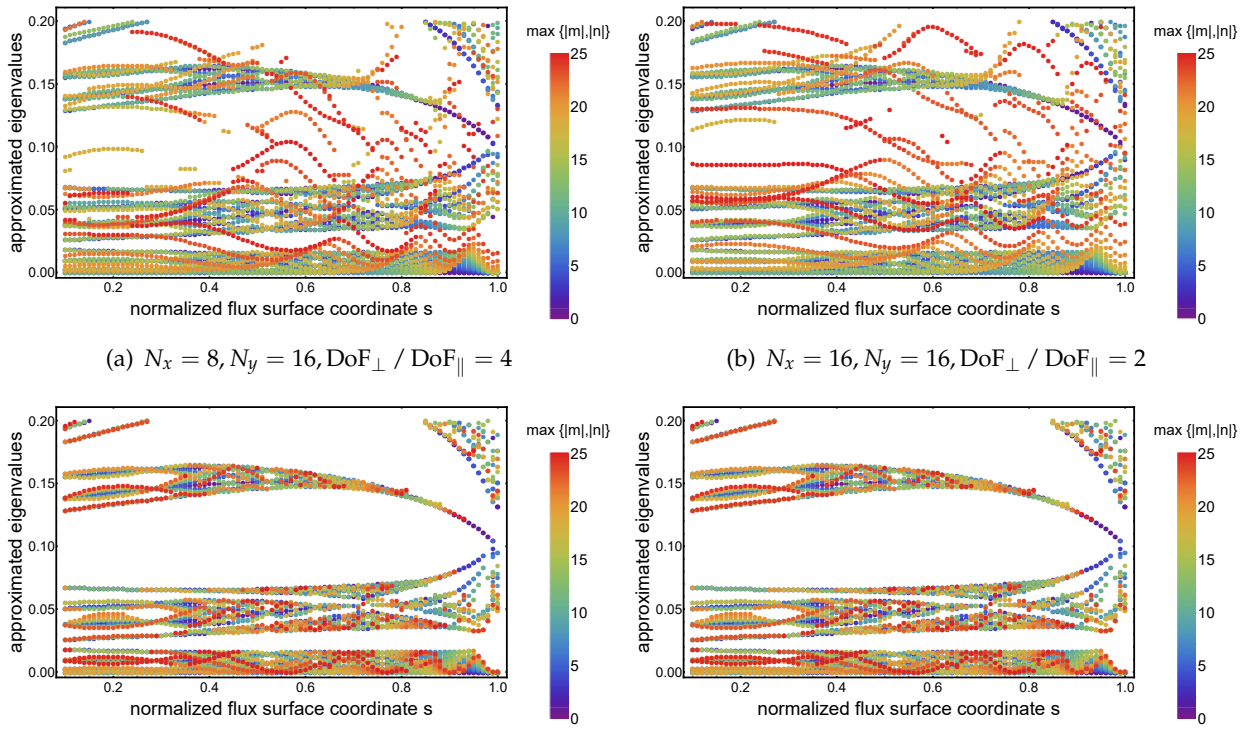


FIGURE 6.27: *Convergence results of ADG with $p_\xi = 3, p_\eta = 7$ for the anisotropic wave equation with metric terms. Toroidally non-conforming meshes of increasing resolution are used. Modes are indicated by their shape, and resolution by colour.* $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 4$ *within a field period is fulfilled.*

Figure 6.27 shows results for increasing mesh resolutions fulfilling $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 4$ within a field period for the selection of modes of Figure 6.25. We observe that the spectra of $N_x = 8, N_y = 16$

and $N_x = 16, N_y = 32$ overall coincide. Therefore, we deduce that further refinement of the mesh yields roughly the same results. For $N_x = 4, N_y = 8$, eigenvalues of higher mode numbers, namely $(12, -11)$ for all s and $(8, -6)$ for $s < 0.3$, are not converged yet. The eigenvalues of low mode numbers have converged as the results coincide for all meshes.
We now consider a broader range of mode numbers up to $m_{\max}, n_{\max} = 25$.



(a) $N_x = 8, N_y = 16, \mathrm{DoF}_\perp \,/\, \mathrm{DoF}_\parallel = 4$

(b) $N_x = 16, N_y = 16, \mathrm{DoF}_\perp \,/\, \mathrm{DoF}_\parallel = 2$

(c) $N_x = 8, N_y = 32, \mathrm{DoF}_\perp \,/\, \mathrm{DoF}_\parallel = 8$

(d) $N_x = 16, N_y = 32, \mathrm{DoF}_\perp \,/\, \mathrm{DoF}_\parallel = 4$

FIGURE 6.28: *Results of ADG with $p_\xi = 3, p_\eta = 7$ using toroidally non-conforming meshes for the anisotropic wave equation with metric terms. Colors indicate the maximal mode number associated to the eigenvalue.*

Figure 6.28 shows the spectra of ADG for different meshes. High perpendicular resolution of $N_y = 32$ yields two clean gaps within the eigenvalue spectrum located in a neighbourhood of $\omega^2 \approx 0.02$ and $\omega^2 \approx 0.1$ where no eigenvalues reside for all s as indicated by Figures 6.28(c),(d). The lower gap is the so-called toroidicity-induced Alfvén eigenmode gap (TAE gap), the upper gap the so-called ellipticity-induced Alfvén eigenmode gap (EAE gap) and both are related to plasma instabilities [92]. Therefore, we aim at resolving these gaps accurately and assume that the results are converged if the gaps are clean. For $N_y = 16$, mode numbers larger than 17 exhibit eigenvalues within these gaps as shown in Figures 6.28(a),(b). These modes are not converged yet. We observe that only an increase in perpendicular resolution clears the gaps as given by Figures 6.28(b),(c). Given high perpendicular resolution, a further increase in parallel resolution yields similar results as Figures 6.28(c) and (d) are structurally equivalent.

6.4.3 Comparison

This section compares results of ADG to different codes for the same or structurally equivalent problems to determine whether the same physical behaviour depicted by the eigenvalue spectrum can be observed. We setup ADG using a toroidally non-conforming mesh with the resolution on a single field period being $N_x = 8, N_y = 32$ and basis degrees $p_\xi = 3, p_\eta = 7$ which yields $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 8$.

As a first comparison, we choose CONTI [93] solving the same equation as ADG, namely the anisotropic wave equation with metric terms (3.110) on flux surfaces. CONTI uses a Fourier approach for discretizing functions on a field period. To account for different families of toroidal mode numbers which otherwise cannot be resolved when discretizing a single field period, a phase factor shift $\sigma \in \{0, 1, \ldots, M-1\}$ is introduced where M is the number of field periods. The setup uses poloidal mode numbers with $n_{\max} = 58$ and toroidal mode numbers

$$m \in \{-45, -40, -35, -30, \ldots, 30, 35, 40, 45\} - \sigma \tag{6.7}$$

as $M = 5$ for the underlying W7-X-like equilibrium. For evaluating metric terms, we use a resolution of 240×80 points in poloidal, toroidal direction. The results of CONTI were kindly provided by Axel Könies.

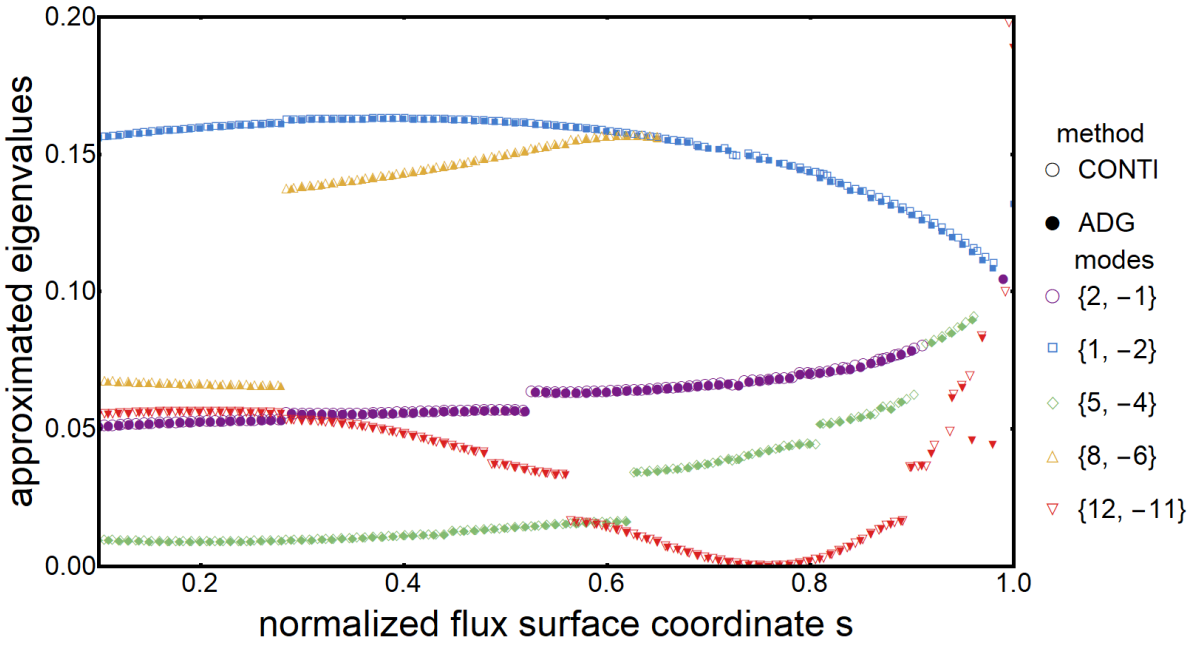
Secondly, we compare with the code for the calculation of kinetic Alfvén waves in three-dimensional geometries (CKA) [94] which solves the reduced MHD shear Alfvén wave equation (2.69) in three dimensions. As this problem is structurally equivalent to the anisotropic wave equation with metric terms as deduced in Section 2.6, we expect similar results. CKA uses B-splines for the discretization of all three dimensions. The setup uses 100×25 splines for the discretization of the poloidal, toroidal direction without an alignment of the mesh. For the radial direction, 150 splines are used and a zero Dirichlet boundary condition is enforced at $s = 1$. As the computations are performed on only one field period as for CONTI, a phase factor shift is used to account for different families of toroidal mode numbers. The results of CKA were kindly provided by Tamás Béla Fehér.

Figure 6.29 shows results for the selection of modes of Figure 6.25 for CONTI, CKA and ADG. For ADG we use filled markers and for CONTI and CKA empty markers. We observe that the results of all codes overall agree with deviations at the boundary of the equilibrium, i.e., $s \geqslant 0.9$, where strong metric terms reside and ι tends towards 1. In particular, the values of CONTI and ADG coincide with few exceptions for mode $(12, -11)$ and $s \geqslant 0.9$. For $\iota = 1$, multiple modes couple and their proper resolution is difficult. Considering the constant coefficient anisotropic wave equation (3.1) for $\boldsymbol{b} = (1, 1)^\top$ and its analytical solution given by Theorem 3.1, we note that the zero-eigenspace has infinite dimensions. This is the case for all $\iota \in \mathbb{Q}$ but $\iota = 1$ produces the largest zero-eigenspace within the truncated space of Fourier modes for fixed $m_{\max} = n_{\max}$. We observe multiple outliers in the results of CKA for all modes but $(12, -11)$ in Figure 6.29(b) which pollute the spectrum.

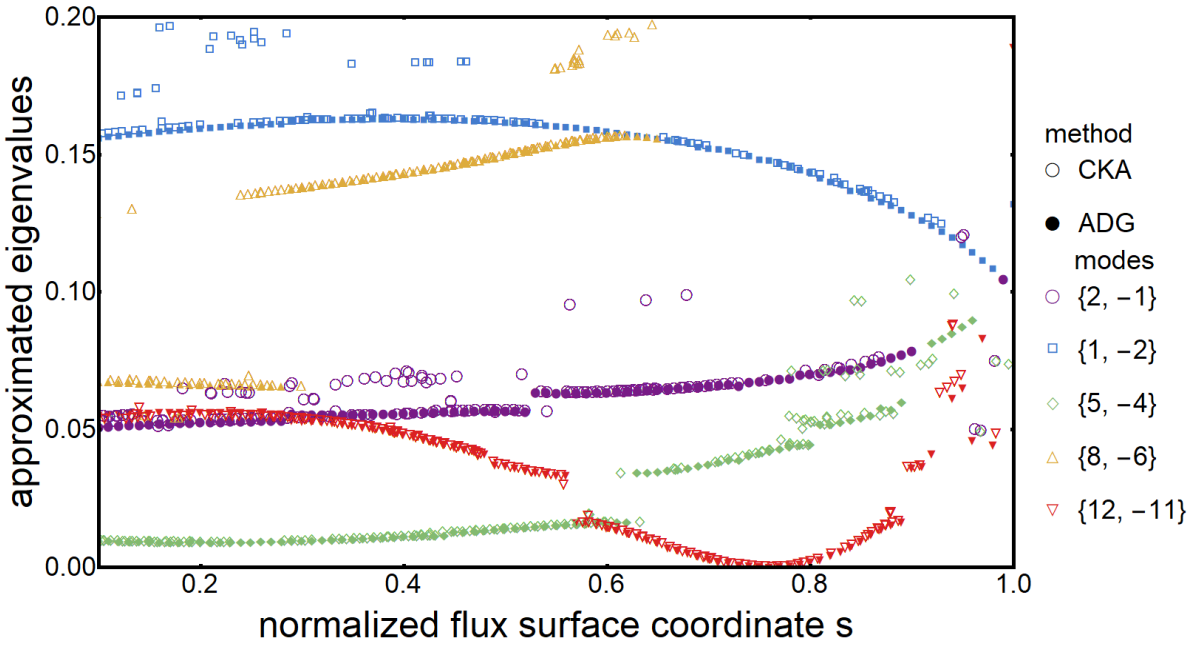
Figure 6.30 shows a collection of eigenvalues for all resolved modes. Modes for CONTI are determined by its setup. The CKA results show mode associations up to maximal mode numbers $m_{\max} = n_{\max} = 14$. For ADG, the maximal mode numbers are set to $m_{\max} = n_{\max} = 25$. We find the gaps at the same positions as in Figure 6.28(c),(d) around the toroidicity-induced gap at $\omega^2 \approx 0.02$ and the ellipticity-induced gap at $\omega^2 \approx 0.1$ but observe a certain amount of outliers in the toroidicity-induced gap and multiple modes with eigenvalues narrowing the ellipticity-induced gap for CKA. The spectra of ADG and CONTI overall coincide.

6.4.4 Summary

We summarize the results of Section 6.4. Section 6.4.1 illustrates that toroidally non-conforming meshes perform superior to poloidally non-conforming meshes. Furthermore, locally aligned meshes with toroidally non-conforming interfaces yield more precise results than a cartesian mesh in particular for higher mode numbers which proves the impact of locally aligning the mesh. Section 6.4.2 shows the convergence of ADG and the benefits of distributing resolution. The formation of toroidicity-induced gaps and ellipticity-induced gaps is observed. Comparisons with CONTI and CKA in Section 6.4.3 shows that ADG reproduces physically important properties.

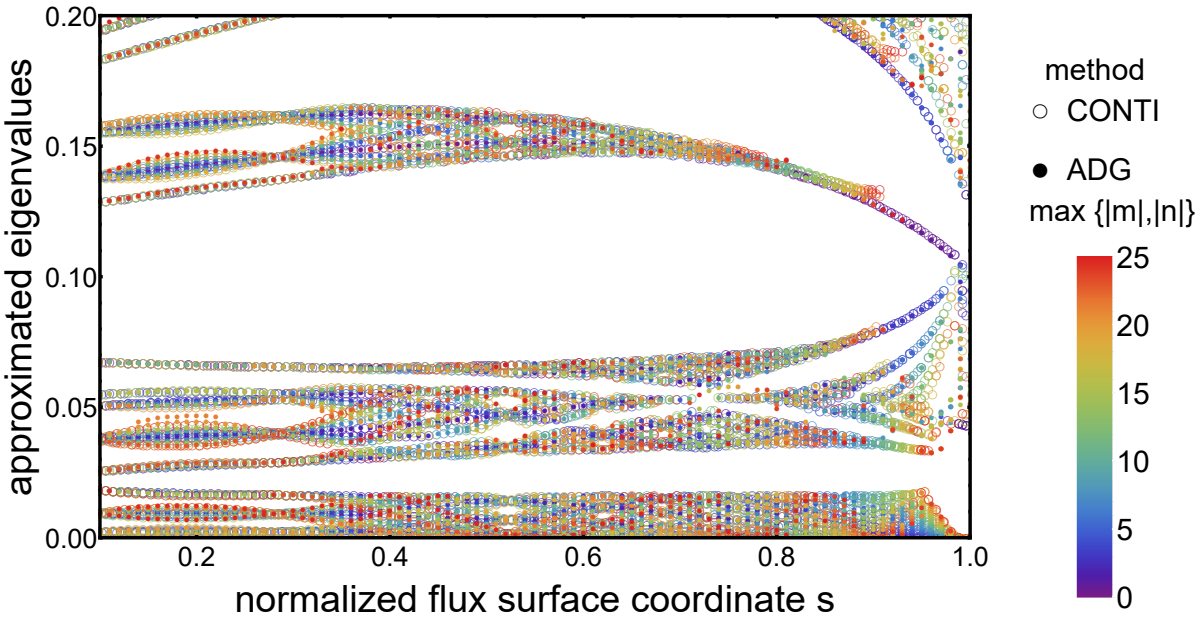


(a) CONTI vs. ADG

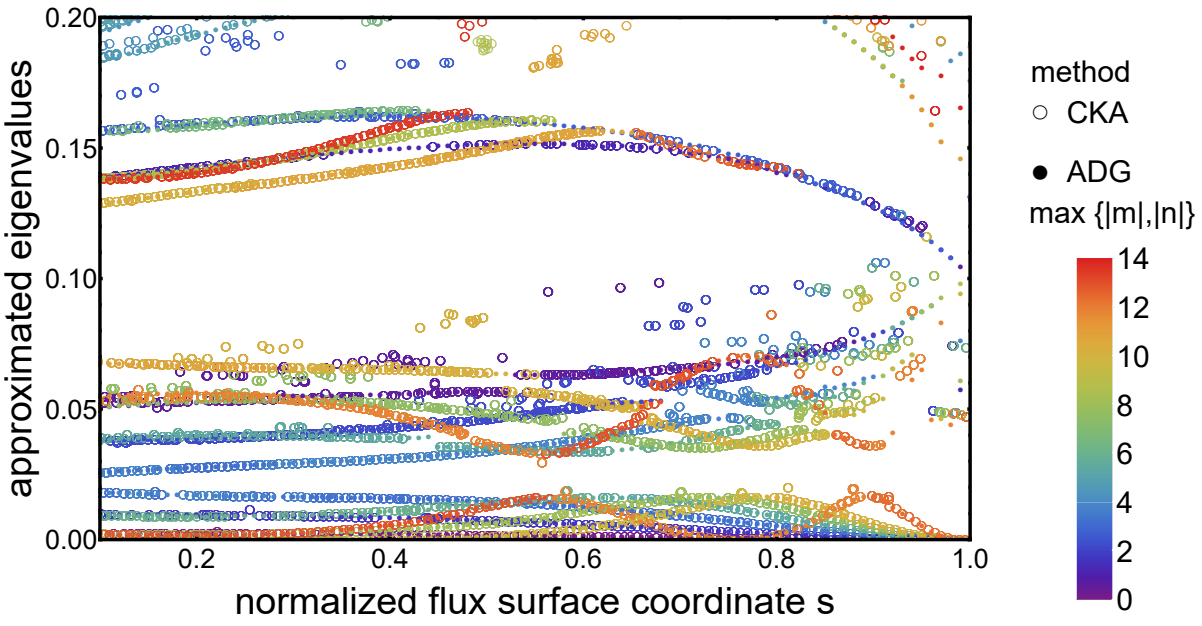


(b) CKA vs. ADG

FIGURE 6.29: *Comparison of ADG using a mesh with toroidally non-conforming meshes and* $p_\xi = 3, p_\eta = 7, N_x = 8, N_y = 32$ *such that* $\mathrm{DoF}_\perp \,/\, \mathrm{DoF}_\parallel = 8$ *for the anisotropic wave equation with metric terms. For CKA, the underlying equation is the reduced MHD shear Alfvén wave equation* (2.69). *Modes are indicated by shape and color whereas methods are indicated by fillings.*



(a) CONTI vs. ADG



(b) CKA vs. ADG

FIGURE 6.30: *Comparison of ADG using a mesh with toroidally non-conforming meshes and* $p_\xi = 3, p_\eta = 7, N_x = 8, N_y = 32$ *such that* $\mathrm{DoF}_\perp$ / $\mathrm{DoF}_\parallel = 8$ *for the anisotropic wave equation with metric terms. For CKA, the underlying equation is the reduced MHD shear Alfvén wave equation* (2.69)*. Colors depict the maximal mode number associated to the eigenvalue. For ADG small dots were used to give orientation.*

Chapter 7

CONCLUSIONS AND PROSPECTS

Wrapping up

Starting from the equations of ideal magnetohydrodynamics (MHD) linearized around an equilibrium state in Chapter 2, we deduce a 4^{th}-order equation, the anisotropic wave equation with metric terms from three-dimensional geometries and the constant coefficient anisotropic wave equation. The analysis of analytical solutions for constant magnetic field $\boldsymbol{b}$ yields benefits when decoupling and distributing the resolution in parallel and perpendicular direction. In Chapters 3 and 4, we develop a discontinuous Galerkin method with a locally field-aligned mesh and basis for all model problems and highlight the case of constant $\boldsymbol{b}$ where matrix assembly simplifies and asymptotic behaviour of the method can be studied. In Chapter 5, we emphasize the implementation of the locally field-aligned discontinuous Galerkin method (ADG) on non-conforming two-dimensional meshes, considering the numerical evaluation of variational forms as well as the association of discrete eigenfunctions to Fourier modes.
The study of numerical results in Chapter 6 shows the following: We assert that the local alignment of mesh and basis indeed allows to decouple the resolution in parallel and perpendicular direction. Therefore, the resolution can be distributed such that highly oscillatory functions are well resolved while providing the possibility of a coarse discretization of close to constant parts. Furthermore, the size of eigenvalue errors no longer exclusively depends on the magnitude of the mode numbers of the associated Fourier eigenmode but also on the size of the eigenvalue itself. This supports the accurate calculation of the spectrum of small eigenvalues which is relevant for plasma heating and stability considerations, see Section 2.7.
For all model problems, we examine the impact of the local alignment of mesh and basis. For the constant coefficient anisotropic wave equation and the 4^{th}-order equation for constant $\boldsymbol{b}$, ADG respectively yields an improvement of up to 6.5 and 5.5 orders of magnitude in accuracy when comparing to a non-aligned cartesian case with the same number of degrees of freedom. In particular, a large gain in accuracy is found for high mode numbers. We study the anisotropic

wave equation with metric terms from a three-dimensional MHD equilibrium for meshes with toroidally and poloidally non-conforming interfaces as well as for the cartesian case. Toroidally non-conforming meshes yield the best results. Poloidally non-conforming meshes are superior to the cartesian case. Furthermore, we investigate the convergence of ADG for all model problems. For the 4^{th}-order equation, no convergence was found due to the ill-posedness of the problem and cancellations shown in the analytical derivation of the solution. For the anisotropic wave equations with constant coefficients and with metric terms, we observe rapid convergence of the results. The converged result of ADG for the eigenvalue spectrum of a MHD equilibrium displays the same physical behaviour as and good agreement with the one provided by existing codes. Furthermore, we prove the distribution of resolution to be impactful even in three-dimensional geometries.

So far, ADG relies on two-dimensional meshes such that wave solutions can only be studied on isolated flux surfaces of an MHD equilibrium. Considering further development, three-dimensional equations such as the reduced MHD shear Alfvén wave equation (2.69) or the normal mode formulation of the linearized MHD stability problem for general three-dimensional equilibria (2.32) couple over different flux surfaces. Hence, the mesh has to be extended to also discretize the normalized flux surface coordinate s. As flux surfaces are nested, the inner flux surfaces are smaller than the outer flux surfaces with the innermost flux surface, namely the magnetic axis, degenerating to a line. We therefore aim at a coarser discretization of the inner part and a finer discretization of the outer part of an MHD equilibrium, i.e., a distribution of resolution in s-direction, which can be achieved using non-conforming interfaces in s-direction. The choice of a non-conforming discontinuous Galerkin method readily establishes the mathematical methodology in two dimensions and a radially extended tensor-product approach can be introduced. However, the local alignment on the different flux surfaces varies with s and the treatment of two-dimensional non-conforming interfaces is numerically challenging. As small deviations from the alignment of cells are possible as shown in Sections 6.2.1 and 6.2.2, the construction of a three-dimensional non-conforming mesh could be simplified.

We conclude that the local alignment of mesh and basis allows the distribution of resolution and therefore improves the accuracy by multiple orders of magnitude in comparison to non-aligned meshes for all model problems. ADG offers the flexibility to focus the computational effort on the numerically challenging and structurally important aspects of the physical problem.

ACKNOWLEDGEMENTS

Several people contributed to this thesis on the scientific level. I would like to thank Axel Könies and Ralf Kleiber for uncovering the initial problem statement and for introducing myself to the physical background. Omar Maj and Marco Restelli enlightened this background in various discussions from a mathematician's point of view. Tamás Fehér gave an insight into the CKA method and processed its data. Further, I thank Roman Hatzky for providing best practise examples for scientific writing as well as general insight into the scientific community.
I would like to express my gratitude to my supervisor Florian Hindenlang for his untiring commitment and work input particularly regarding the implementation of ADG. Last but not least, I deeply acknowledge my advisor Eric Sonnendrücker for his open door policy, his continuous interest for and contributions to the project.

Bibliography

[1] U.S. Energy Information Administration. International Energy Outlook 2017. https://www.eia.gov/outlooks/ieo/pdf/0484(2017).pdf, 2017. Accessed: 2018-03-20.

[2] International Energy Agency. World Energy Outlook 2017. https://www.iea.org/weo2017, 2017. Accessed: 2018-03-20.

[3] UNFCCC. Paris Agreement. https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7-d&chapter=27&clang=_en, 2016. C.N.92.2016.TREATIES-XXVII.7.d, Accessed: 2018-03-20.

[4] Bundeskabinett. Klimaschutzplan 2050 – Klimaschutzpolitische Grundsätze und Ziele der Bundesregierung. https://www.bmub.bund.de/fileadmin/Daten_BMU/Download_PDF/Klimaschutz/klimaschutzplan_2050_bf.pdf, 2016. Accessed: 2018-03-20.

[5] CDU, CSU, SPD. Ein neuer Aufbruch für Europa, Eine neue Dynamik für Deutschland, Ein neuer Zusammenhalt für unser Land, Koalitionsvertrag zwischen CDU, CSU und SPD. https://www.bundesregierung.de/Content/DE/_Anlagen/2018/03/2018-03-14-koalitionsvertrag.pdf?__blob=publicationFile&v=1, 2018. Accessed: 2018-03-20.

[6] acatech/Leopoldina/Akademienunion. Sektorkopplung – Optionen für die nächste Phase der Energiewende. http://energiesysteme-zukunft.de/fileadmin/user_upload/Publikationen/pdf/ESYS_Stellungnahme_Sektorkopplung.pdf, 2017. Accessed: 2018-03-20.

[7] Friedrich Wagner. Surplus from and storage of electricity generated by intermittent sources. *The European Physical Journal Plus*, 131(12):445, 2016.

[8] D. Grand, Ch. Le Brun, R. Vidil, and F. Wagner. Electricity production by intermittent renewable sources: A synthesis of French and German studies. *The European Physical Journal Plus*, 131(9):329, 2016.

[9] Francesco Romanelli. Strategies for the integration of intermittent renewable energy sources in the electrical system. *The European Physical Journal Plus*, 131(3):53, 2016.

[10] Friedrich Wagner. Considerations for an EU-wide use of renewable energies for electricity generation. *The European Physical Journal Plus*, 129(10):219, 2014.

[11] eStorage. Overview of potential locations for new Pumped Storage Plants in EU 15, Switzerland and Norway. `http://www.estorage-project.eu/wp-content/uploads/2013/06/eStorage_D4.2-Overview-of-potential-locations-for-new-variable-PSP-in-Europe.pdf`, 2015. Accessed: 2018-03-20.

[12] Hans-Werner Sinn. Buffering Volatility: A Study on the Limits of Germany's Energy Revolution. *European Economic Review*, 2017.

[13] Niamh Troy, Eleanor Denny, and Mark O'Malley. Base-load cycling on a system with significant wind penetration. *IEEE Transactions on Power Systems*, 25(2):1088–1097, 2010.

[14] Bundeskabinett. Dreizehntes Gesetz zur Änderung des Atomgesetzes. `http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl111s1704.pdf`, 2011. Accessed: 2018-03-20.

[15] Arthur S. Eddington. The internal constitution of the stars. *The Scientific Monthly*, 11(4):297–303, 1920.

[16] Hans Albrecht Bethe. Energy production in stars. *Physical Review*, 55(5):434, 1939.

[17] Richard Rhodes. *The making of the atomic bomb*. Simon and Schuster, 2012.

[18] Lyman Spitzer Jr. The stellarator concept. *The Physics of Fluids*, 1(4):253–264, 1958.

[19] Boris Dmitrievich Bondarenko. Role played by OA Lavrent'ev in the formulation of the problem and the initiation of research into controlled nuclear fusion in the USSR. *Physics-Uspekhi*, 44(8):844–851, 2001.

[20] Jeffrey P. Freidberg. *Plasma physics and fusion energy*. Cambridge university press, 2008.

[21] Susanne Pfalzner. *An introduction to inertial confinement fusion*. CRC Press, 2006.

[22] Robert L. Hirsch. Inertial-electrostatic confinement of ionized fusion gases. *Journal of Applied Physics*, 38(11):4522–4534, 1967.

[23] W.H. Breunlich, P. Kammel, J.S. Cohen, and M. Leon. Muon-catalyzed fusion. *Annual Review of Nuclear and Particle Science*, 39(1):311–356, 1989.

[24] N.J. Peacock, D.C. Robinson, M.J. Forrest, P.D. Wilcock, and V.V. Sannikov. Measurement of the electron temperature by Thomson scattering in tokamak T3. *Nature*, 224(5218):488–490, 1969.

[25] Michael Kenward. Fusion research-The temperature rises. *New Scientist*, 82:626–630, 1979.

[26] Daniel Clery. After ITER, many other obstacles for fusion power. *Science Insider*, 2013.

[27] Chris Lee. Wibbly-wobbly magnetic fusion stuff: The return of the stellarator. https://arstechnica.com/science/2017/06/wibbly-wobbly-magnetic-fusion-stuff-the-return-of-the-stellarator/, 2017. Accessed: 2018-03-20.

[28] EUROfusion. Joint European Torus. https://www.euro-fusion.org/JET/. Accessed: 2018-03-20.

[29] Max-Planck-Institut für Plasmaphysik. Introduction: The ASDEX Upgrade tokamak. https://www.ipp.mpg.de/16208/einfuehrung. Accessed: 2018-03-20.

[30] ITER Organization. What is ITER? https://www.iter.org/proj/inafewlines. Accessed: 2018-03-20.

[31] National Institute for Fusion Science. Large Helical Device Project. http://www.lhd.nifs.ac.jp/en/home/lhd.html. Accessed: 2018-03-20.

[32] Max-Planck-Institut für Plasmaphysik. Introduction – The Wendelstein 7-X stellarator. https://www.ipp.mpg.de/16931/einfuehrung. Accessed: 2018-03-20.

[33] D. Maisonnier, I. Cook, P. Sardain, R. Andreani, L. Di Pace, R. Forrest, L. Giancarli, S. Hermsmeyer, P. Norajitra, N. Taylor, et al. A conceptual study of commercial fusion power plants. Final Report of the European Fusion Power Plant Conceptual Study (PPCS). *European Fusion Development Agreement*, 2005.

[34] J. Kenneth Shultis and Richard E. Faw. *Fundamentals of Nuclear Science and Engineering Third Edition*. CRC press, 2016.

[35] Wolf Häfele. *Energy in a Finite World: A Global Systems Analysis (Volume 2)*. Ballinger, 1981.

[36] David Bodansky. *Nuclear energy: principles, practices, and prospects*. Springer Science & Business Media, 2007.

[37] Graham M. Keyser, David L. Mader, and James A. O'neill. Method for isotope replenishment in an exchange liquid used in a laser induced isotope enrichment process. https://patentimages.storage.googleapis.com/99/9f/6e/7f270d14d7bc47/US4620909.pdf, November 4 1986. US Patent 4,620,909 A, Accessed: 2018-03-20.

[38] Hisham Zerriffi. Tritium: The environmental, health, budgetary, and strategic effects of the Department of Energy's decision to produce tritium. *IEER. January*, 1996.

[39] EUROPEAN COMMISSION. Questions and Answers: Signature of the European Fusion Joint Programme – 'EUROfusion'. `http://europa.eu/rapid/press-release_MEMO-14-570_en.pdf`. Accessed: 2018-03-20.

[40] Daniele Antonio Di Pietro and Alexandre Ern. *Mathematical aspects of discontinuous Galerkin methods*, volume 69. Springer Science & Business Media, 2011.

[41] William H. Reed and T.R. Hill. Triangular mesh methods for the neutron transport equation. Technical report, Los Alamos Scientific Lab., N. Mex.(USA), 1973.

[42] Joachim Nitsche. Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. In *Abhandlungen aus dem mathematischen Seminar der Universität Hamburg*, volume 36, pages 9–15. Springer, 1971.

[43] Ivo Babuška. The finite element method with penalty. *Mathematics of computation*, 27(122):221–228, 1973.

[44] Ivo Babuška and Miloš Zlámal. Nonconforming elements in the finite element method with penalty. *SIAM Journal on Numerical Analysis*, 10(5):863–875, 1973.

[45] Jim Douglas and Todd Dupont. Interior penalty procedures for elliptic and parabolic Galerkin methods. *Computing methods in applied sciences*, pages 207–216, 1976.

[46] Mary Fanett Wheeler. An elliptic collocation-finite element method with interior penalties. *SIAM Journal on Numerical Analysis*, 15(1):152–161, 1978.

[47] Douglas N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAM journal on numerical analysis*, 19(4):742–760, 1982.

[48] Francesco Bassi and Stefano Rebay. A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier–Stokes equations. *Journal of computational physics*, 131(2):267–279, 1997.

[49] Franco Brezzi, Gianmarco Manzini, Donatella Marini, Paola Pietra, and Alessandro Russo. Discontinuous Galerkin approximations for elliptic problems. *Numerical Methods for Partial Differential Equations*, 16(4):365–378, 2000.

[50] Carlos Erik Baumann and J. Tinsley Oden. A discontinuous hp finite element method for convection—diffusion problems. *Computer Methods in Applied Mechanics and Engineering*, 175(3-4):311–341, 1999.

[51] Bernardo Cockburn and Chi-Wang Shu. The local discontinuous Galerkin method for time-dependent convection-diffusion systems. *SIAM Journal on Numerical Analysis*, 35(6):2440–2463, 1998.

[52] Douglas N. Arnold, Franco Brezzi, Bernardo Cockburn, and L. Donatella Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM journal on numerical analysis*, 39(5):1749–1779, 2002.

[53] Jeffrey P. Freidberg. *Ideal MHD*. Cambridge University Press, 2014.

[54] Allen H. Boozer. Establishment of magnetic coordinates for a given magnetic field. Technical report, Princeton Univ., NJ (USA). Plasma Physics Lab., 1981.

[55] R.C. Grimm, J.M. Greene, and J.L. Johnson. Methods in Computational Physics. *Academic Press*, (9):253 – 280, 1976.

[56] Richard P. Feynman. Feynman lectures on physics. Volume 2: Mainly electromagnetism and matter. *Reading, Ma.: Addison-Wesley, 1964, edited by Feynman, Richard P.; Leighton, Robert B.; Sands, Matthew*, 1964.

[57] Tamás Béla Fehér. *Simulation of the interaction between Alfvén waves and fast particles*. PhD thesis, Max-Planck-Institut für Plasmaphysik, 2014.

[58] Bruce Scott. The character of transport caused by E$\times$ B drift turbulence. *Physics of Plasmas*, 10(4):963–976, 2003.

[59] Joseph D. Huba. NRL: Plasma formulary. Technical report, Naval Research Laboratory Washington, DC, 2013.

[60] C.Z. Cheng and M.S. Chance. Low-n shear Alfvén spectra in axisymmetric toroidal plasmas. *The Physics of fluids*, 29(11):3695–3701, 1986.

[61] H.L. Berk, J.W. Van Dam, Z. Guo, and D.M. Lindberg. Continuum damping of low-n toroidicity-induced shear Alfvén eigenmodes. *Physics of Fluids B: Plasma Physics*, 4(7):1806–1835, 1992.

[62] Carolin Schwab. *Numerische Verfahren zur Untersuchung der magnetohydrodynamischen Stabilität dreidimensionaler Plasmagleichgewichte*. PhD thesis, Technischen Hochschule Darmstadt, 1991.

[63] Albert Salat and John A. Tataronis. Radial dependence of magnetohydrodynamic continuum modes in axisymmetric toroidal geometry. *Physics of Plasmas*, 6(8):3207–3216, 1999.

[64] Eliezer Hameiri. On the essential spectrum of ideal magnetohydrodynamics. Dedicated to Harold Grad on the occasion of his sixtieth birthday. *Communications on pure and applied mathematics*, 38(1):43–66, 1985.

[65] S. Rauf and J.A. Tataronis. Finite-amplitude Alfvén waves in a dissipative inhomogeneous plasma. *Physics of Plasmas*, 2(5):1453–1459, 1995.

[66] G.Y. Fu and J.W. Van Dam. Excitation of the toroidicity-induced shear Alfvén eigenmode by fusion alpha particles in an ignited tokamak. *Physics of Fluids B: Plasma Physics*, 1(10):1949–1952, 1989.

[67] Ya I. Kolesnichenko, V.V. Lutsenko, A. Weller, A. Werner, Yu V. Yakovenko, J. Geiger, and O.P. Fesenyuk. Conventional and nonconventional global Alfvén eigenmodes in stellarators. *Physics of Plasmas*, 14(10):102504, 2007.

[68] G.W. Hammett, M.A. Beer, W. Dorland, S.C. Cowley, and S.A. Smith. Developments in the gyrofluid approach to tokamak turbulence simulations. *Plasma physics and controlled fusion*, 35(8):973, 1993.

[69] Bruce Scott. Shifted metric procedure for flux tube treatments of toroidal geometry: Avoiding grid deformation. *Physics of Plasmas*, 8(2):447–458, 2001.

[70] F. Hariri and M. Ottaviani. A flux-coordinate independent field-aligned approach to plasma turbulence simulations. *Computer Physics Communications*, 184(11):2419–2429, 2013.

[71] Andreas Stegmeir, David Coster, Alexander Ross, Omar Maj, Karl Lackner, and Emanuele Poli. GRILLIX: a 3D turbulence code based on the flux-coordinate independent approach. *Plasma Physics and Controlled Fusion*, 60(3):035005, 2018.

[72] Faker Ben Belgacem. The mortar finite element method with Lagrange multipliers. *Numerische Mathematik*, 84(2):173–197, 1999.

[73] David A. Kopriva, Stephen L. Woodruff, and M. Yousuff Hussaini. Computation of electromagnetic scattering with a non-conforming discontinuous spectral element method. *International journal for numerical methods in engineering*, 53(1):105–122, 2002.

[74] Tiago Tamissa Ribeiro and Bruce D. Scott. Conformal tokamak geometry for turbulence computations. *IEEE Transactions on plasma science*, 38(9):2159–2168, 2010.

[75] Claude E. Shannon. Communication in the presence of noise. *Proceedings of the IEEE*, 72(9):1192–1201, 1984.

[76] Roger A. Hom and Charles R. Johnson. Topics in matrix analysis. *Cambridge UP, New York*, 1991.

[77] F. Bassi, S. Rebay, G. Mariotti, Savini Pedinotti, and M. Savini. A high-order accurate discontinuous finite element method for inviscid and viscous turbomachinery flows. In *Proceedings of the 2nd European Conference on Turbomachinery Fluid Dynamics and Thermodynamics*, pages 99–109. Technologisch Instituut, Antwerpen, Belgium, 1997.

[78] Robert M. Gray et al. Toeplitz and circulant matrices: A review. *Foundations and Trends® in Communications and Information Theory*, 2(3):155–239, 2006.

[79] Milton Abramowitz and Irene A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 9. Dover, New York, 1972.

[80] Eric Polizzi and James Kestyn. FEAST Eigenvalue Solver v3.0 User Guide. `https://arxiv.org/pdf/1203.4031`, 2015. Accessed: 2018-03-20.

[81] Youcef Saad. SPARSKIT: A basic tool kit for sparse matrix computations. `https://pdfs.semanticscholar.org/b56b/61fe2387cb9e64c1e249255e9372151438b8.pdf`, 1994. Accessed: 2018-03-20.

[82] P. R. Amestoy, I. S. Duff, J. Koster, and J.-Y. L'Excellent. A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM Journal on Matrix Analysis and Applications*, 23(1):15–41, 2001.

[83] P. R. Amestoy, A. Guermouche, J.-Y. L'Excellent, and S. Pralet. Hybrid scheduling for the parallel solution of linear systems. *Parallel Computing*, 32(2):136–156, 2006.

[84] Steven P. Hirshman and J.C. Whitson. Steepest-descent moment method for three-dimensional magnetohydrodynamic equilibria. *The Physics of fluids*, 26(12):3553–3568, 1983.

[85] S.P. Hirshman and O. Betancourt. Preconditioned descent algorithm for rapid calculations of magnetohydrodynamic equilibria. *Journal of Computational Physics*, 96(1):99–109, 1991.

[86] VMECwiki. VMEC. `http://vmecwiki.pppl.wikispaces.net/VMEC`, 1994. Accessed: 2018-03-20.

[87] Axel Könies. personal communication. 2015-11-19.

[88] Vicente Hernandez, Jose Roman, A. Tomás, and Vicent Vidal. SLEPc Users Manual, Scalable library for eigenvalue problem computations. Technical report, Tech. Rep. DISC-II/24/02, Universidad Politecnica de Valencia. See http://www. grycap. upv. es/slepc, 2006.

[89] Sheung Hun Cheng and Nicholas J. Higham. The nearest definite pair for the Hermitian generalized eigenvalue problem. *Linear Algebra and Its Applications*, 302:63–76, 1999.

[90] Carolin Nührenberg. Global ideal magnetohydrodynamic stability analysis for the configurational space of Wendelstein 7–X. *Physics of Plasmas*, 3(6):2401–2410, 1996.

[91] Craig Beidler, Günter Grieger, Franz Herrnegger, Ewald Harmeyer, Johann Kisslinger, Wolf Lotz, Henning Maassberg, Peter Merkel, Jürgen Nührenberg, Fritz Rau, et al. Physics and engineering design for Wendelstein VII-X. *Fusion Technology*, 17(1):148–168, 1990.

[92] Ricardo Betti and Jeffrey P. Freidberg. Stability of Alfvén gap modes in burning plasmas. *Physics of Fluids B: Plasma Physics*, 4(6):1465–1474, 1992.

[93] Axel Könies and Denis Eremin. Coupling of Alfvén and sound waves in stellarator plasmas. *Physics of Plasmas*, 17(1):012107, 2010.

[94] Axel Könies. A code for the calculation of kinetic Alfvén waves in threedimensional Geometry. *10th IAEA TM on Energetic Particles in Magnetic Confinement Systems*, 2007.