

Space-Time Methods for Acoustic Waves with Applications to Full Waveform Inversion

Zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften

von der KIT-Fakultät für Mathematik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

Dissertation

von
Johannes Ernesti

Tag der mündlichen Prüfung: 20. Dezember 2017

Referent: Prof. Dr. Christian Wieners

Korreferent: Prof. Dr. Andreas Rieder



This document is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0): <https://creativecommons.org/licenses/by-sa/4.0/deed.en>

Contents

1	Introduction	1
1.1	Outline	4
1.2	Acknowledgement	5
1.3	Notation and basic terms	5
2	Theory of linear variational problems	7
2.1	Variational problems in BANACH spaces	7
2.1.1	Relations to BANACH's theory	8
2.2	Approximation of variational problems	14
2.3	Saddle point problems in HILBERT spaces	16
2.4	Approximation of saddle point problems	19
3	Mathematical modeling of acoustic waves	23
3.1	The acoustic wave equation	23
3.1.1	The semigroup setting	24
3.1.2	Duality, adjoint operators and the HILBERT adjoint	25
3.2	A variational space-time setting	25
3.2.1	A space-time HILBERT space setting	26
3.2.2	The closure of the space-time operator (L, \mathcal{V})	29
3.3	Space-time substructuring	31
4	Space-time minimal residual methods	33
4.1	Least-Squares Finite Elements	33
4.2	Weakly conforming Least-Squares Finite Elements	37
4.2.1	A weakly conforming approximation space	37
4.2.2	Saddle point reformulation	40
4.2.3	Discrete ellipticity and inf-sup stability	43
4.2.4	The skeleton reduction	45
4.2.5	Discussion	49

4.3	The Discontinuous PETROV-GALERKIN method	50
4.3.1	Continuous and discrete well-posedness	50
4.3.2	Optimal test functions – the trial-to-test operator	51
4.3.3	Breaking the test space – the ideal DPG method	52
4.3.4	Approximate optimal testing – the practical DPG method	54
4.3.5	DPG as a minimal residual method	55
4.3.6	Built-in residual error-estimator	55
4.3.7	Skeleton reduction	56
4.4	Acoustic waves – the ideal DPG method	57
4.5	Acoustic waves – the practical DPG method	60
4.5.1	Local construction of the FORTIN operator	60
4.5.2	A scaling argument	64
4.5.3	An a-priori error estimate for the practical DPG method	65
4.6	Acoustic waves – the simplified DPG method	65
4.6.1	The construction of the FORTIN Operator	69
4.6.2	Skeleton reduction	70
4.7	A-priori error estimates for smooth solutions	72
5	Numerical Experiments	77
5.1	Numerical setup	77
5.1.1	Discretizations and error quantities	77
5.1.2	Mesh refinements	78
5.1.3	Local SCHUR complements and problem sizes	79
5.1.4	A remark on the upcoming sections	83
5.2	Numerical examples (1D)	84
5.2.1	A smooth example	84
5.2.2	A traveling wave in a homogeneous medium	90
5.2.3	A low regularity model problem	96
5.3	Numerical examples (2D)	103
5.3.1	A smooth example	103
5.3.2	A traveling wave-front in an inhomogeneous medium	107
5.3.3	A double slit experiment in a homogeneous medium	113
5.4	Adaptivity with respect to the polynomial degree	115
5.5	Summary	117
6	Full Waveform Inversion (FWI)	119
6.1	Mathematical setting and notation	120

6.1.1	The forward problem and the adjoint problem	121
6.1.2	Observations and the parameter-to-seismogram map	122
6.1.3	The problem in FWI	123
6.2	Root-finding approach	124
6.3	The derivative of \mathcal{F} and its adjoint	125
6.3.1	The derivative $\mathcal{F}'(m^k)$	126
6.3.2	The adjoint $\mathcal{F}'(m^k)^*$ of the derivative $\mathcal{F}'(m^k)$	126
6.3.3	The CG-REGINN algorithm	128
6.4	Minimization approach	129
6.4.1	A gradient descent method	130
6.4.2	NEWTON-type methods for minimization	132
6.4.3	GAUSS-NEWTON approximation of J''	133
6.5	A numerical example for FWI	134
6.5.1	The experimental setup	134
6.5.2	Application of CG-REGINN	138
6.5.3	A remark on the computational effort	144
7	Summary and outlook	147
A	Explicit estimates using symmetric eigenvalue problems	151
B	Evaluate J'' by the adjoint state method	153
C	An L_1-setting in 1D	155
	Bibliography	161
	Danksagung	167

Chapter 1

Introduction

Wave phenomena play an important role in our everyday life as well as in science and engineering. For instance, in the form of light and sound, waves provide the most important source of information for humans and animals to interact with their environment. Another example are electro-magnetic waves which represent the foundation for modern data transmission techniques, e.g., used for television, radio, mobile phones, and for high speed data transfer via intercontinental optical fibers. Moreover, ultra-sound waves and X-rays are important tools for medical diagnosis as well as for therapeutic applications such as cancer treatment. In order to get a deeper understanding of the universe, telescopes collect electro-magnetic waves that have been radiated by interstellar objects. Furthermore, seismic waves excited by earthquakes or special devices provide an important source for our knowledge about the interior of the earth. Hence, extending our knowledge of wave phenomena also leads to a better understanding of the world.

In this work, we will focus on mathematical aspects concerning the simulation of acoustic waves in heterogeneous media using modern parallel computers. Since in our applications the amplitudes are small, we restrict ourselves to linear models which provide a reasonable approximation of the physical processes in this setting. The linear acoustic wave equation models the interaction of pressure waves with gas, liquid and soil. Thus, it acts as a prototype for a variety of different wave phenomena. Mathematically, acoustic waves can be described using time-dependent partial differential equations (PDEs). In practical applications, the solution of these PDEs is not accessible in closed form and has to be approximated numerically. A classical numerical method to deal with time-dependent PDEs is the method of lines. Here, the problem is first discretized in space leading to a system of finitely many ordinary differential equations (ODEs). This system can be treated with classi-

cal time-integration schemes. Another classical approach for time-dependent problems is ROTHE’s method. Here, first time is discretized, e.g. using discrete difference quotients. This yields a sequence of stationary problems for the chosen time-discretization. Both approaches, however, yield an inherently iterative procedure in time that is challenging to parallelize, see e.g. [32] for an overview. Additionally, flexible methods that allow for adaptivity are not straight-forward to implement using the method of lines as well as ROTHE’s method.

To overcome these issues, we consider *space-time discretizations*. In this approach, the time variable is treated in the same way as every other space variable. As a result, a d -dimensional evolutionary problem in space becomes a $(d + 1)$ -dimensional *stationary* problem in space-time, cf. FIGURE 1.1. In case of a linear equation, this

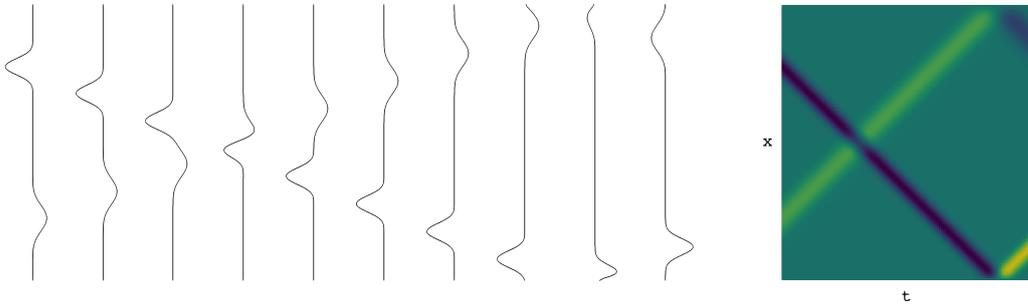


Figure 1.1: Two waves in one space dimension traveling up and down with reflections on the boundaries. On the left, ten snapshots for different times are shown that have been calculated successively using the method of lines from left to right. On the right, a space-time plot of the same waves is depicted. Here, the solution has been calculated *at once* for all times.

procedure leads to a huge linear system containing *all* space-time degrees of freedom to be solved for in one shot. At a first glance, this procedure makes the problem more challenging than before, since a much larger system has to be solved. However, assuming that a well-scaling parallel algorithm for large linear systems is available, the space-time approach yields a scheme that is not only parallel in space but also *parallel in time*. Since the classical methods do not parallelize well for large numbers of processes on huge super-computers space-time approaches can be more efficient with respect to wall-clock time. In addition to the promising aspects concerning parallelization, space-time methods are also appealing for adaptivity. Since the time-dependent evolutionary problem becomes stationary in space-time, all tools that are available for adaptivity in case of stationary problems can be applied.

In this work, we consider families of space-time discretizations originating from Least-Squares approaches, also called *minimal residual methods*. In particular, we construct two novel Least-Squares methods that minimize the residual norm in space-

time, one of them being an application of the discontinuous PETROV-GALERKIN method (DPG) introduced by DEMKOWICZ et. al. [19]. The stability of Least-Squares techniques renders them interesting candidates for applications with space-time adaptivity.

As an application of the space-time discretization schemes, we consider an inverse problem originating from seismic imaging. The idea of seismic imaging is to exploit that a wave after having traveled through a heterogeneous medium contains information about the medium's spatially varying structure. For instance, by using surface measurements of waves that traveled through the earth's crust, one can try to reconstruct the spatially changing material properties, see FIGURE 1.2. Then, this information can be used, e.g. to locate mineral resources without the need of drilling holes. Especially for areas that are hard to access such as structures below the seabed, non-invasive methods are appealing, because large areas can be examined without the need of expensive drilling. A technique to tackle seismic imaging

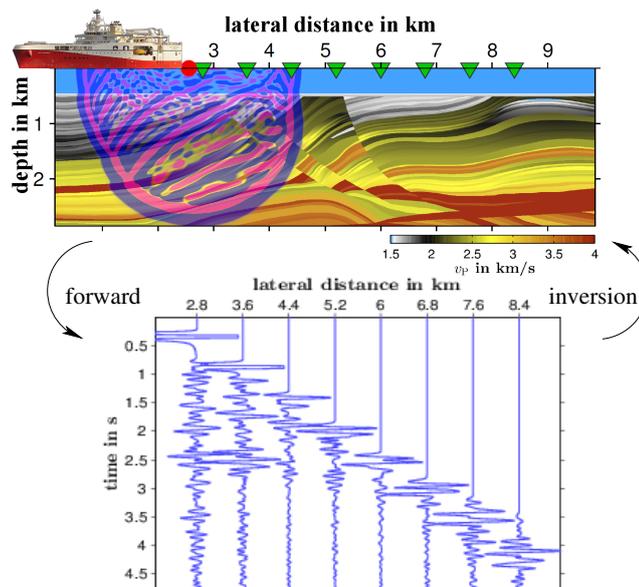


Figure 1.2: A survey ship exciting a wave that travels through the water and then the earth below the seabed. The receivers located on top of the sea record the reflected signals, the *seismograms*, which are depicted at the bottom. In applications, only these seismograms are available in order to reconstruct the corresponding material distribution. (Image by Thomas Bohlen, personal communication.)

problems is Full Waveform Inversion (FWI) which, in contrast to other methods, uses the *full* information contained in the measurements in order to achieve high accuracy reconstructions. It is well-known that implementing numerical schemes for FWI is challenging for two reasons: on the one hand the inverse problem is ill-posed meaning

that small errors in the measurements can significantly change the reconstruction; on the other hand it is computationally expensive since the state-of-the-art algorithms are of NEWTON-type and require high-accuracy solutions of the wave-equation in every step. Typically, during the inversion dozens to hundreds of wave equations have to be solved. To address the first issue, a sensible regularization strategy has to be chosen. In order to improve the computational efficiency, well-scaling parallel algorithms to solve the required wave equations are of interest.

1.1 Outline

In order to set up a solid framework for the following considerations, we start with a mostly self-contained overview of linear variational problems in CHAPTER 2 by summarizing well-known and also some more recent findings for variational problems in HILBERT spaces.

In CHAPTER 3, we introduce a space-time HILBERT space setting for acoustic waves that allows for solutions with low regularity such as space-time discontinuities. Treating time as an additional space dimension, we consider a space-time differential operator and construct a suitable domain of definition. Furthermore, we provide a well-defined notion of generalized traces for functions in this enlarged closure.

In CHAPTER 4, we introduce two new families of space-time minimal residual methods for acoustic waves. Starting from conforming Finite Element Methods of Least-Squares type in space-time, we introduce a non-conforming generalization using the framework introduced in CHAPTER 3. Moreover, we describe variants of the Discontinuous PETROV-GALERKIN (DPG) method in space-time including a non-conforming scheme that features appealing properties from an implementation point of view. For both methods, we present numerical analysis results including convergence estimates.

Complementing the theoretical considerations in CHAPTER 4, we present an extensive numerical study in CHAPTER 5. For different variants of methods introduced in CHAPTER 4, we compare the convergence properties by considering examples in one and two spatial dimensions.

In CHAPTER 6, we consider the FWI problem using a space-time setting for waves. We present two points of view, a root-finding point of view originating from the inverse-problems community and an optimization point that has been used in many applications. For both, we set up NEWTON-type algorithms on an abstract level and explain their relation to each other. To reduce the numerical effort, we employ the adjoint-state method to tackle linearized problems. Here, using the space-time

framework, we make extensive use of the accessibility to the adjoint equation that can be interpreted as an evolution problem backward in time. Finally, we show numerical results for the regularized inexact NEWTON scheme, CG-REGINN [57], applied to the FWI problem in a simple test setting. To handle the wave equations in this algorithm, we employ the space-time DPG method introduced in CHAPTER 4.

Parts of CHAPTER 3 and CHAPTER 4 have been submitted in [27]. Based on the *Python Seismic Imaging Toolbox, PySIT* [34], we developed a software to experiment with the principles of seismic imaging. The program can be downloaded for free from our website <http://www.math.kit.edu/ianm3/seite/seismicimaging/en>.

1.2 Acknowledgement

The author gratefully acknowledges the support of the German Research Foundation (DFG) by CRC 1173. This work has been created as part of joint research projects in this CRC on space-time discretizations and Full Waveform Inversion in cooperation with the geophysics department (GPI) at KIT.

Some of the numerical experiments were performed on the computational resource ForHLR II funded by the Ministry of Science, Research and the Arts Baden-Württemberg and DFG.

1.3 Notation and basic terms

To emphasize that an identifier B equals an expression E by definition we write $B := E$. We use the KRONECKER delta with $\delta_{ij} = 1$, $i = j$, and $\delta_{ij} = 0$, $i \neq j$. By \mathbb{R} , we denote the set of real numbers and $\mathbb{N} := \{1, 2, 3, \dots\}$ is used for the set of natural numbers. Furthermore, we write $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$ for the natural numbers including 0. The set of integers is denoted by $\mathbb{Z} = \{0, 1, -1, 2, -2, \dots\}$.

For two sets A, B , we denote by A^B the set of all maps $f: B \rightarrow A$. For instance, the set of all sequences in \mathbb{R} with natural indices is denoted by $\mathbb{R}^{\mathbb{N}}$.

For $n \in \mathbb{N}$ we denote the space of real vectors with n components by \mathbb{R}^n . If $x \in \mathbb{R}^n$ we write $x_d \in \mathbb{R}$, $d = 1, \dots, n$, for the d -th component of x .

For a normed vector space X , we usually write $\|x\|_X$ for the norm of $x \in X$. Moreover, if there is an inner product on X that induces $\|\cdot\|_X$, we write $(x, y)_X$ for the inner product of $x, y \in X$. In case that $U \in \mathbb{R}^n$ is an open set, we also write $(x, y)_U$ for the inner product of $x, y \in L_2(U, \mathbb{R}^m)$. The identity in X is $\text{id}_X: X \rightarrow X$, $x \mapsto x$. In this work, we only consider vector spaces over the field \mathbb{R} .

For $x \in X$, $r > 0$, we denote the *open ball with radius r around x* by $B(r, x) := \{y \in X : \|x - y\|_X < r\}$. We call $A \subset X$ a *neighborhood of $x \in A$* if there is $r > 0$ with $B(r, x) \subset A$. Further, in this case x is called an *inner point of A* . $A \subset X$ is called *closed* if for each convergent sequence $(x_n)_{n \in \mathbb{N}} \in A^{\mathbb{N}}$, i.e. $\exists x \in X : \|x_n - x\|_X \rightarrow 0$ for $n \rightarrow \infty$, we have $x \in A$. We call the smallest closed set that contains A the *closure of A* and denote it by $\bar{A} := \bigcap \{B \subset X : A \subset B, B \text{ closed}\}$.

By $\partial A := \{x \in \bar{A} : x \text{ is not an interior point of } A\}$, the *boundary of A* is denoted.

If $f: D \rightarrow Y$ is a mapping from $D \subset X$ to another vector space Y , we write $\text{supp } f := \overline{\{x \in D : f(x) \neq 0\}}$ for the *support of f* . In particular, we denote by $C_c^1(U) := \{f \in C^1(U) : \text{supp } f \subset U\}$ the set of compactly supported functions defined on $U \subset \mathbb{R}^n$, where $C^1(U)$ are the continuously differentiable functions on U .

The space of all $m \times n$ -matrices is denoted by $\mathbb{R}^{m \times n}$ and we write M_{ij} for the element of M located at the i -th row and the j -th column. For a matrix $M \in \mathbb{R}^{m \times n}$, we denote the transposed matrix with $M_{ij}^T := M_{ji}$ by $M^T \in \mathbb{R}^{n \times m}$. We write $I_n := (\delta_{kl})_{kl}$ for the identity matrix in $\mathbb{R}^{n \times n}$.

Given a subset $A \subset X$ of a vector space X , we define the *linear hull of A*

$$\text{span } A := \left\{ \sum_{i=1}^k \alpha_i a_i : \alpha_i \in \mathbb{R}, a_i \in A, k \in \mathbb{N} \right\}.$$

For linear operators between normed vector spaces X and Y , we use the following notation. We write $\mathcal{L}(X, Y) := \{T: X \rightarrow Y : T \text{ linear and bounded}\}$ and $\mathcal{L}(X) := \mathcal{L}(X, X)$. For a linear operator $T: X \rightarrow Y$, we denote the image of $x \in X$ under T by $Tx := T[x] := T(x)$. For the special case that X is a normed vector space and $Y = \mathbb{R}$, we also use the *dual-pairing notation* $\langle T, x \rangle := \langle T, x \rangle_{X' \times X} := Tx$ if $T \in X'$ is a bounded linear functional. Here, by $X' := \mathcal{L}(X, \mathbb{R})$ the topological dual-space of X is denoted.

For a map $f: A \rightarrow B$, A, B sets, we denote for a subset $C \subset A$ the restriction of f to C by $f|_C : C \rightarrow B$, $c \mapsto f(c)$, $c \in C$.

The LEBESGUE measure of a set $B \subset \mathbb{R}^n$ is symbolized by $|B|$.

Often, we consider suprema or infima of quotients silently assuming that the denominator does not vanish, e.g. $\sup_{y \in \mathbb{R}^n} \frac{x^\top y}{\|y\|_2}$, $x \in \mathbb{R}^n$.

For an open set $O \subset \mathbb{R}^D$, we denote by $L_2(O)$ the set of measurable and square integrable functions on O . The set of vector fields on O into \mathbb{R}^N that are componentwise in $L_2(O)$ is denoted by $L_2(O, \mathbb{R}^N)$.

Chapter 2

Theory of variational problems in HILBERT spaces

In this work, we consider a variational space-time framework for acoustic waves and, based on this framework, different discretization schemes. Because we make extensive use of variational formulations for partial differential equations, we summarize well-known results from Finite Element theory and describe links to linear Functional Analysis for the convenience of the reader. Self-contained textbook references on this subject include [24], [9] and [7].

2.1 Variational problems in BANACH spaces

We start by establishing well-posedness results for linear variational problems.

Definition 2.1. *Let X, Y be normed vector spaces. We call $b: X \times Y \rightarrow \mathbb{R}$ bilinear if for every $x \in X$ the map $b(\cdot, y)$ is a linear form on X and for every $y \in Y$ the map $b(x, \cdot)$ is a linear form on Y . We define the set of bounded bilinear maps on $X \times Y$*

$$\mathcal{B}(X \times Y, \mathbb{R}) := \left\{ b: X \times Y \rightarrow \mathbb{R}: \|b\| := \sup_{(x,y) \in X \times Y} \frac{b(x,y)}{\|x\|_X \|y\|_Y} < \infty \right\}.$$

Now, given BANACH spaces X, Y , a bounded bilinear form $b \in \mathcal{B}(X \times Y, \mathbb{R})$, and a functional $\ell \in Y'$, we consider the following problem.

$$\begin{cases} \text{Find } x \in X \text{ such that} \\ b(x, y) = \ell(y) \quad \text{for all } y \in Y. \end{cases} \quad (\text{VP})$$

Definition 2.2. *We say that problem (VP) is well-posed if for all right-hand sides $\ell \in Y'$, there is a unique solution $x = x(\ell) \in X$ and if this solution depends continuously on ℓ .*

In order to characterize the well-posedness of (VP), we set up an operator framework. Sometimes it is convenient to consider (VP) as an equation in the dual space Y' . To this end, we introduce two operators induced by the bilinear form b .

Definition 2.3. For $b \in \mathcal{B}(X \times Y, \mathbb{R})$, we define the operators $B \in \mathcal{L}(X, Y')$ and $B' \in \mathcal{L}(Y, X')$ by

$$\langle Bx, y \rangle = b(x, y) = \langle B'y, x \rangle, \quad x \in X, y \in Y.$$

We write $\mathcal{N}(B) = \{x \in X : Bx = 0\}$ and $\mathcal{R}(B) = \{Bx : x \in X\}$ for the kernel and the range of B .

As a result, (VP) can be reformulated by $Bx = \ell$ as an equation in Y' .

In the following, using the close relation between B' and the BANACH adjoint of B , we apply the closed range theorem and the open mapping theorem to characterize the solvability of (VP) by duality properties of B and B' .

Sometimes we only need unique solvability of (VP) up to the kernel of B . To this end, we introduce variants of B and B' as well as the quotient space.

Definition 2.4. Let X be a BANACH space and $U \subset X$ a closed subspace. We consider the quotient space $X/U = \{x + U : x \in X\}$ equipped with the quotient norm

$$\|x + U\|_{X/U} := \inf_{u \in U} \|x + u\|_X, \quad x + U \in X/U,$$

which itself is a BANACH space, cf. e.g. [72, Section I.II].

Definition 2.5. For a bilinear form $b \in \mathcal{B}(X \times Y, \mathbb{R})$ define $\hat{B} : X/\mathcal{N}(B) \rightarrow \mathcal{R}(B)$ and $\hat{B}' : Y/\mathcal{N}(B') \rightarrow \mathcal{R}(B')$ by

$$\langle \hat{B}(x + \mathcal{N}(B)), y \rangle = b(x, y) = \langle \hat{B}'(y + \mathcal{N}(B')), x \rangle, \quad x \in X, y \in Y,$$

both of which are well-defined and bijective by construction.

This notation yields the variant $\hat{B}(x + \mathcal{N}(B)) = \ell$ in Y' of (VP) having solutions that are only unique up to a difference in $\mathcal{N}(B)$. In other words, the solution lives in the factor space $X/\mathcal{N}(B)$.

2.1.1 Relations to BANACH's theory

The characterization of the well-posedness of (VP) relies on classical results from Functional Analysis. As a first step, we introduce some notation and provide elementary results.

Definition 2.6 (Annihilators). [60, Chap. 4] Let X be a normed space and $U \subset X$, $Z \subset X'$. Then we define the closed subspaces

$$U^\perp = \{\ell \in X' : \ell(u) = 0 \text{ for all } u \in U\} \subset X',$$

$${}^\perp Z = \{x \in X : z'(x) = 0 \text{ for all } z' \in Z\} \subset X.$$

In particular, ${}^\perp(U^\perp)$ is the closure of U in X if $U \subset X$ is a linear space, cf. [60, Thm. 4.7].

Definition 2.7 (BANACH adjoint). For BANACH spaces X, Y and $L \in \mathcal{L}(X, Y)$, we define the BANACH adjoint $L^\circledast : Y' \rightarrow X'$ of L by

$$\langle L^\circledast y', x \rangle = y'(Lx) = \langle y', Lx \rangle, \quad x \in X, y' \in Y'$$

see [72, section VII].

Remark 2.8. It holds $L^\circledast \in \mathcal{L}(Y', X')$ with $\|L^\circledast\| = \|L\|$, cf. [72, Thm. VII.4].

If L is bijective then L^\circledast also is and we have $(L^{-1})^\circledast = (L^\circledast)^{-1}$, since

$$(L^\circledast(L^{-1})^\circledast x')(x) = \langle (L^{-1})^\circledast x', Lx \rangle = \langle x', x \rangle = x'(x),$$

$$((L^{-1})^\circledast(L^\circledast y'))(y) = \langle L^\circledast y', L^{-1}y \rangle = \langle y', y \rangle = y'(y),$$

for all $x' \in X'$, $x \in X$, $y' \in Y'$, $y \in Y$.

Definition 2.9. Let X be a BANACH space. We say that X is reflexive if the canonical embedding $\iota_X \in \mathcal{L}(X, X'')$ defined by

$$\iota_X(x)(x') = x'(x), \quad x \in X, x' \in X',$$

is an isomorphism. Here, $X'' := (X')'$ is the bi-dual of X .

Proposition 2.10. Let X, Y be BANACH spaces and $b \in \mathcal{B}(X \times Y, \mathbb{R})$. Then, we have $B' = B^\circledast \circ \iota_Y$ for B' as in DEFINITION 2.3.

For reflexive Y , we have $\mathcal{R}(B') = \mathcal{R}(B^\circledast)$ and $\mathcal{N}(B') = {}^\perp \mathcal{R}(B)$.

Proof. It holds $B^\circledast \in \mathcal{L}(Y'', X')$ and for all $x \in X$, $y \in Y$

$$\langle B^\circledast \iota_Y(y), x \rangle = \langle \iota_Y(y), Bx \rangle = \iota_Y(y)(Bx) = \langle Bx, y \rangle = b(x, y) = \langle B'y, x \rangle.$$

In case that Y is reflexive, we have $Y'' = \iota_Y(Y)$ and therefore

$$\mathcal{R}(B') = B'(Y) = B^\circledast(\iota_Y(Y)) = B^\circledast(Y'') = \mathcal{R}(B^\circledast).$$

Using $\langle Bx, y \rangle = b(x, y) = \langle B'y, x \rangle$ for $x \in X$, $y \in Y$, we have

$$\begin{aligned} {}^\perp \mathcal{R}(B) &= \{y \in Y : \langle Bx, y \rangle = 0 \ \forall x \in X\} \\ &= \{y \in Y : \langle B'y, x \rangle = 0 \ \forall x \in X\} = \mathcal{N}(B'). \end{aligned} \quad \square$$

Remark 2.11. *Considering the operator B' instead of B^\circledast can be interpreted as identifying $Y \cong Y''$ through ι_Y which is very common in the literature. In the following, however, we treat this identification explicitly hoping that this leads to arguments that are easier to comprehend.*

Now, we cite the fundamental results to characterize the well-posedness of (VP).

Theorem 2.12 (Open Mapping, S. BANACH). *Let X, Y be BANACH spaces and let $L \in \mathcal{L}(X, Y)$ be onto. Then, for every open set $O \subset X$ in X , the image $L(O) \subset Y$ is open in Y .*

Proof. See [72, section II.5]. □

Theorem 2.13 (Closed Range, S. BANACH). *Let X, Y be BANACH spaces and $L \in \mathcal{L}(X, Y)$. Then the following assertions are equivalent:*

1. $\mathcal{R}(L)$ is closed in Y .
2. $\mathcal{R}(L^\circledast)$ is closed in X' .
3. $\mathcal{R}(L) = {}^\perp \mathcal{N}(L^\circledast)$.
4. $\mathcal{R}(L^\circledast) = \mathcal{N}(L)^\perp$.

Here, we use the notation of annihilators as introduced in DEFINITION 2.6.

Proof. See [72, section VII.5]. □

Lemma 2.14. *Let X, Y be BANACH spaces and $L \in \mathcal{L}(X, Y)$. Then, $\mathcal{R}(L)$ is closed in Y if and only if there is $\alpha > 0$ with $\|Lx\|_Y \geq \alpha \|x + \mathcal{N}(L)\|_{X/\mathcal{N}(L)}$ for all $x \in X$.*

Proof. See also [24, Lemma A.36]. Since $\mathcal{N}(L) \subset X$ is a closed subspace, the quotient space $X/\mathcal{N}(L)$ is a BANACH space for the quotient norm, see DEFINITION 2.4.

If $\mathcal{R}(L)$ is closed in Y , then $\hat{L}: X/\mathcal{N}(L) \rightarrow \mathcal{R}(L), x + \mathcal{N}(L) \mapsto Lx$ is linear, bounded and bijective between two BANACH spaces. Thus, $\hat{L}^{-1}: \mathcal{R}(L) \rightarrow X/\mathcal{N}(L)$ exists and is bounded by THEOREM 2.12. Setting $\alpha = \|\hat{L}^{-1}\|_{Y, X/\mathcal{N}(L)}^{-1}$, the first implication follows from

$$\|x + \mathcal{N}(L)\|_{X/\mathcal{N}(L)} = \|\hat{L}^{-1}Lx\|_{X/\mathcal{N}(L)} \leq \|\hat{L}^{-1}\|_{Y, X/\mathcal{N}(L)} \|Lx\|_Y, \quad x \in X.$$

Conversly, let $y_n = Lx_n \in \mathcal{R}(L)$, $x_n \in X$, be a sequence in $\mathcal{R}(L)$ such that $\lim_{n \rightarrow \infty} y_n = y \in Y$ exists. Then, $(x_n + \mathcal{N}(L))_n$ is a CAUCHY sequence in $X/\mathcal{N}(L)$ since by assumption

$$\alpha \|(x_n + \mathcal{N}(L)) - (x_m + \mathcal{N}(L))\|_{X/\mathcal{N}(L)} \leq \|Lx_n - Lx_m\|_Y, \quad n, m \in \mathbb{N}.$$

Thus, there is $x_\infty \in X$ such that $x_n + \mathcal{N}(L) \rightarrow x_\infty + \mathcal{N}(L)$ in $X/\mathcal{N}(L)$ and

$$Lx_\infty = \hat{L}(x_\infty + \mathcal{N}(L)) = \lim_{n \rightarrow \infty} \hat{L}(x_n + \mathcal{N}(L)) = \lim_{n \rightarrow \infty} Lx_n = y,$$

by the continuity of \hat{L} . Finally, this implies $y \in \mathcal{R}(L)$. \square

Using LEMMA 2.14, we obtain the following reformulation of THEOREM 2.13.

Theorem 2.15. *Let X, Y be reflexive BANACH spaces and $b \in \mathcal{B}(X \times Y, \mathbb{R})$. Then, for B, B' as in DEFINITION 2.3 and \hat{B}, \hat{B}' as in DEFINITION 2.5, the following assertions are equivalent:*

1. *There is $\beta_1 > 0$ with $\|Bx\|_{Y'} \geq \beta_1 \|x + \mathcal{N}(B)\|_{X/\mathcal{N}(B)}$ for all $x \in X$.*
2. *There is $\beta_2 > 0$ with $\|B'y\|_{X'} \geq \beta_2 \|y + \mathcal{N}(B')\|_{Y/\mathcal{N}(B')}$ for all $y \in Y$.*
3. *There is $\beta_3 > 0$ such that $\hat{B}: X/\mathcal{N}(B) \rightarrow \mathcal{N}(B')^\perp$ is an isomorphism and*

$$\|\hat{B}^{-1}\|_{Y', X/\mathcal{N}(B)} \leq \frac{1}{\beta_3}.$$

4. *There is $\beta_4 > 0$ such that $\hat{B}': Y/\mathcal{N}(B') \rightarrow \mathcal{N}(B)^\perp$ is an isomorphism and*

$$\|(\hat{B}')^{-1}\|_{Y', Y/\mathcal{N}(B')} \leq \frac{1}{\beta_4}.$$

Proof. Using the reflexivity of Y , straight-forward calculations show $\mathcal{N}(B^\circledast) = \iota_Y(\mathcal{N}(B'))$ and ${}^\perp\mathcal{N}(B^\circledast) = \mathcal{N}(B')^\perp$. Thus, we see by LEMMA 2.14 and THEOREM 2.13

$$1. \iff \mathcal{R}(B) \text{ closed in } Y' \iff \mathcal{R}(B) = {}^\perp\mathcal{N}(B^\circledast) = \mathcal{N}(B')^\perp.$$

For $y' = Bx = \hat{B}(x + \mathcal{N}(B)) \in \mathcal{R}(B)$, $x \in X$, we have for $\beta > 0$

$$\|Bx\|_{Y'} \geq \beta \|x + \mathcal{N}(B)\|_{X/\mathcal{N}(B)} \iff \|y'\|_{Y'} \geq \beta \|\hat{B}^{-1}y'\|_{X/\mathcal{N}(B)}$$

since \hat{B} is bijective by definition.

This yields 1. \iff 3. and, repeating the arguments for B' , 2. \iff 4..

Using $\mathcal{R}(B') = \mathcal{R}(B^\circledast)$, PROPOSITION 2.10, we see 1. \iff 2. by THEOREM 2.13 and LEMMA 2.14. \square

In the following, we characterize the solvability of (VP) using operator notation as well as using the bilinear form itself.

Proposition 2.16. *Let X, Y be reflexive BANACH spaces, $b \in \mathcal{B}(X \times Y, \mathbb{R})$. Then $B: X \rightarrow Y'$ as in DEFINITION 2.3 is an isomorphism if and only if*

$$\exists \beta > 0: \quad \|Bx\|_{Y'} \geq \beta \|x\|_X \quad \forall x \in X \quad \text{and} \quad \mathcal{N}(B') = \{0\}. \quad (2.1)$$

Given that $\beta > 0$ as in (2.1) exists, we have $\|B^{-1}\|_{Y', X} \leq \frac{1}{\beta}$.

Proof. If B is an isomorphism, we have $\|x\|_X = \|B^{-1}Bx\|_X \leq \|B^{-1}\|_{Y',X} \|Bx\|_{Y'}$ for all $x \in X$ and the first implication follows by $\mathcal{N}(B') = {}^\perp\mathcal{R}(B) = \{0\}$, see PROPOSITION 2.10.

Conversely, assuming $\mathcal{N}(B') = \{0\}$ gives $\mathcal{N}(B')^\perp = \{0\}^\perp = Y'$, and implication (1. \implies 3.) in THEOREM 2.15 yields that $\hat{B}: X/\{0\} \rightarrow Y'$ is an isomorphism with $\|\hat{B}^{-1}\|_{Y',X/\{0\}} \leq \frac{1}{\beta}$.

Since the quotient map $T_{\{0\}}: X \rightarrow X/\{0\}$, $x \mapsto x + \{0\}$ is an isometric isomorphism, by $\hat{B} \circ T_{\{0\}} = B$, we obtain that B is an isomorphism as well with $\|B^{-1}\| \leq \frac{1}{\beta}$. \square

Theorem 2.17 (BANACH-NEČAS-BABUŠKA). *For reflexive BANACH spaces X, Y consider $b \in \mathcal{B}(X \times Y, \mathbb{R})$ and $\ell \in Y'$. Then, we have:*

1. (VP) is well-posed if and only if

$$\exists \beta > 0: \quad \inf_{x \in X} \sup_{y \in Y} \frac{b(x, y)}{\|x\|_X \|y\|_Y} \geq \beta, \quad (\text{BNB1})$$

and

$$\forall y \in Y: \quad (\forall x \in X: b(x, y) = 0) \implies (y = 0). \quad (\text{BNB2})$$

2. If (BNB1) holds, a solution $x \in X$ of (VP) fulfills $\|x\|_X \leq \frac{1}{\beta} \|\ell\|_{Y'}$.

Proof. See [24, Thm. 2.6]. Considering the operators B, B' as in DEFINITION 2.3, the equivalence statement is a reformulation of PROPOSITION 2.16:

$$(\text{BNB1}) \iff (\|Bx\|_{Y'} \geq \beta \|x\|_X \quad \forall x \in X), \quad (\text{BNB2}) \iff \mathcal{N}(B') = \{0\}$$

Finally, the solution $x \in X$ of (VP) fulfills by (BNB1)

$$\|x\|_X \leq \frac{1}{\beta} \sup_{y \in Y} \frac{b(x, y)}{\|y\|_Y} = \frac{1}{\beta} \sup_{y \in Y} \frac{\ell(y)}{\|y\|_Y} = \frac{1}{\beta} \|\ell\|_{Y'}. \quad \square$$

An important consequence of THEOREM 2.17 is the well-known LAX-MILGRAM Lemma for HILBERT spaces.

Corollary 2.18 (LAX-MILGRAM Lemma). *Let V be a HILBERT space, $\ell \in V'$ and let $a: V \times V \rightarrow \mathbb{R}$ be a bounded and coercive bilinear form, i.e.*

$$\exists \alpha > 0: \quad \inf_{v \in V} \frac{a(v, v)}{\|v\|_V^2} \geq \alpha. \quad (2.2)$$

Then for $X = Y = V$ and $b = a$, problem (VP) is well-posed and its solution $u \in V$ fulfills $\|u\|_V \leq \frac{1}{\alpha} \|\ell\|_{V'}$.

Proof. By the coercivity of a , we obtain (BNB1), since for $u \in U$

$$\inf_{v \in V} \sup_{\tilde{v} \in V} \frac{a(v, \tilde{v})}{\|v\|_V \|\tilde{v}\|_V} \geq \inf_{v \in V} \frac{a(v, v)}{\|v\|_V^2} \geq \alpha.$$

Further, $a(v, v) \geq \alpha \|v\|_V^2$, $v \in V$, yields (BNB2) and THEOREM 2.17 implies the assertion. \square

Remark 2.19. *Given a bounded and coercive bilinear form $a: X \times X \rightarrow \mathbb{R}$ for a BANACH space X , defining $(\cdot, \cdot) = a(\cdot, \cdot) + a(\cdot, \cdot)$ gives an inner product that yields the same topology as $\|\cdot\|_X$. Thus, the restriction to HILBERT spaces in COROLLARY 2.18 is natural.*

Lemma 2.20. *Let V be a HILBERT space and $V_0 \subset V$ be a closed subspace. Then for every $\hat{v} \in V/V_0$ there is a unique $v_{\min} \in \hat{v}$ such that*

$$\|v_{\min}\|_V = \inf_{v \in \hat{v}} \|v\|_V. \quad (2.3)$$

Proof. This is a consequence of the orthogonal projection theorem in HILBERT spaces, see e.g. [60, Thm. 4.11]. \square

Before continuing to the approximation theory, we provide a useful result to find the constants β_i from THEOREM 2.15 in practice.

Remark 2.21. *The proof of THEOREM 2.15 shows that $\beta_1 = \beta_3$ and $\beta_2 = \beta_4$ if they exist. In case that $\mathcal{N}(B') = \{0\}$, PROPOSITION 2.22 shows that all β_i coincide.*

Proposition 2.22. *Let X, Y be reflexive BANACH spaces, $b \in \mathcal{B}(X \times Y, \mathbb{R})$ such that $B: X \rightarrow Y'$ from DEFINITION 2.3 is an isomorphism.*

Then, $B': X \rightarrow Y'$ is an isomorphism as well and both are bounded below by the same constant, i.e. the largest lower bounds $\beta_1^{\text{opt}}, \beta_3^{\text{opt}}$ coincide:

$$\inf_{x \in X} \frac{\|Bx\|_{Y'}}{\|x\|_X} =: \beta_1^{\text{opt}} = \beta_3^{\text{opt}} := \inf_{y \in Y} \frac{\|B'y\|_{X'}}{\|y\|_Y}.$$

Proof. By PROPOSITION 2.16, B is bounded below and B' is injective. Using implication (1. \implies 2.) of THEOREM 2.15 and repeating the arguments in the proof of PROPOSITION 2.16, we see that B' is an isomorphism and bounded below as well.

Since $B' = B^{\otimes} \circ \iota_Y$ and ι_Y is an isometric isomorphism, we have

$$\beta_1^{\text{opt}} = \|B^{-1}\| = \|(B^{\otimes})^{-1}\| = \|\iota_Y \circ (B')^{-1}\| = \|(B')^{-1}\| = \beta_3^{\text{opt}}. \quad \square$$

Remark 2.23. *For practical applications, PROPOSITION 2.22 gives freedom of choice whether to prove that B or B' are bounded below in order to determine a bound for the constant β . Using this fact is often referred to as by duality in the literature.*

2.2 Approximation of variational problems

To approximate the solution $x \in X$ of (VP), we restrict ourselves to discrete subspaces $X_h \subset X$, $Y_h \subset Y$ and consider the following finite dimensional problem.

$$\begin{cases} \text{Find } x_h \in X_h \text{ such that} \\ b(x_h, y_h) = \ell(y_h) \quad \text{for all } y_h \in Y_h. \end{cases} \quad (\text{VP}_h)$$

Applying THEOREM 2.17, problem (VP_h) is uniquely solvable if and only if

$$\exists \beta_h > 0: \quad \inf_{x_h \in X_h} \sup_{y_h \in Y_h} \frac{b(x_h, y_h)}{\|x_h\|_X \|y_h\|_Y} \geq \beta_h, \quad (\text{BNB1}_h)$$

and

$$\forall y_h \in Y_h: \quad (\forall x_h \in X_h: b(x_h, y_h) = 0) \implies (y_h = 0). \quad (\text{BNB2}_h)$$

Since the supremum is taken over a smaller space Y_h instead of Y , (BNB1) *does not* imply (BNB1_h) in general. Also the implication (BNB2) \implies (BNB2_h) is not true in general.

As a result, we need to verify (BNB1_h) and (BNB2_h) in order to guarantee the well-posedness of (VP_h). However, there is a link to the rank theorem from linear algebra for finite dimensional spaces. Let $N, M \in \mathbb{N}$ such that $N = \dim X_h$ and $M = \dim Y_h$. Using ordered bases (x_1, \dots, x_N) of X_h and (y_1, \dots, y_M) of Y_h , we define $\underline{B} \in \mathbb{R}^{N \times M}$ by

$$\underline{B}_{nm} = b(x_n, y_m), \quad n = 1, \dots, N, \quad m = 1, \dots, M. \quad (2.4)$$

In case that $N = M$, \underline{B} is a square matrix and both, (BNB1_h) and (BNB2_h), imply that \underline{B} has full rank.

Proposition 2.24. *Let X_h, Y_h be finite dimensional normed spaces of the same dimension, i.e. $\dim X_h = \dim Y_h < \infty$. Then, we have $(\text{BNB1}_h) \iff (\text{BNB2}_h)$.*

Proof. See [24, Proposition 2.21]. □

Remark 2.25. *By PROPOSITION 2.24, as soon as we ensure $\dim X_h = \dim Y_h$, we only need to check (BNB1_h) in order to guarantee well-posedness of (VP_h).*

Given a variational problem (VP) and approximation spaces X_h, Y_h , the truth of (BNB1_h) can be characterized by the existence of a linear operator, the so-called FORTIN operator.

Lemma 2.26 (FORTIN criterium). *Let X, Y be reflexive BANACH spaces, $X_h \subset X$, $Y_h \subset Y$ closed subspaces and let $b \in \mathcal{B}(X \times Y, \mathbb{R})$ fulfill (BNB1).*

Then, (BNB1_h) holds true if there is $\Pi_h \in \mathcal{L}(Y, Y_h)$ with

$$b(x_h, \Pi_h y) = b(x_h, y) \quad \text{for all } x_h \in X_h, y \in Y. \quad (2.5)$$

Proof. Given $\Pi_h \in \mathcal{L}(Y, Y_h)$ with (2.5), we obtain for $x_h \in X_h$ by $\Pi_h(Y) \subset Y_h$

$$\sup_{y_h \in Y_h} \frac{b(x_h, y_h)}{\|y_h\|_Y} \geq \sup_{y \in Y} \frac{b(x_h, \Pi_h y)}{\|\Pi_h y\|_Y} \geq \frac{1}{\|\Pi_h\|} \sup_{y \in Y} \frac{b(x_h, y)}{\|y\|_Y} \geq \frac{\beta}{\|\Pi_h\|} \|x_h\|_X.$$

This yields (BNB1_h) with $\beta_h = \frac{\beta}{\|\Pi_h\|}$. \square

Remark 2.27. *In case that Y is a HILBERT space, the converse of LEMMA 2.26 holds true, see [15] and [25].*

The following lemma provides an easy to prove variant of the converse statement for BANACH spaces. Here, additionally to (BNB1), we assume that (BNB2) is fulfilled and that we have $\dim X_h = \dim Y_h$.

Lemma 2.28. *Let X, Y be reflexive BANACH spaces, $X_h \subset X$, $Y_h \subset Y_h$ finite dimensional subspaces with $\dim X_h = \dim Y_h$ and let $b \in \mathcal{B}(X \times Y, \mathbb{R})$ fulfill (BNB1) and (BNB2). Then, (BNB1_h) holds true if and only if there is $\Pi_h \in \mathcal{L}(Y, Y_h)$ with*

$$b(x_h, \Pi_h y) = b(x_h, y) \quad \text{for all } x_h \in X_h, y \in Y.$$

Proof. Assuming that (BNB1_h) holds true, we observe by duality (REMARK 2.23) that also the dual discrete problem

$$\begin{cases} \text{Find } y_h \in Y_h \text{ such that} \\ b(x_h, y_h) = \ell(x_h) \quad \text{for all } x_h \in X_h, \end{cases} \quad (\text{VP}'_h)$$

is well-posed for every right-hand side $\ell \in X'$.

For fixed $y \in Y$ we have $\ell := b(\cdot, y) \in X'$ and we define $\Pi_h y := y_h(y)$ using the unique solution $y_h(y) \in Y_h$ of (VP'_h). Then, Π_h is a linear operator and

$$\|\Pi_h y\|_Y = \|y_h(y)\|_Y \leq \frac{1}{\beta_h} \sup_{x_h \in X_h} \frac{b(x_h, y_h)}{\|x_h\|_X} \leq \frac{\|b\|}{\beta_h} \|y\|_Y. \quad (2.6)$$

The remaining implication is a special case of LEMMA 2.26. \square

Now, we prove the main approximation result of this section to estimate the approximation error when instead of (VP) the discrete problem (VP_h) is solved. To this end, we define the operator that maps the continuous solution x^{sol} of (VP) to the approximate solution x_h^{sol} of (VP_h).

Definition 2.29. *Assuming that (VP_h) is well-posed, we define $P_h \in \mathcal{L}(X, X_h)$ by solving (VP_h) with right-hand side $\ell := b(x, \cdot) \in Y'$ for fixed $x \in X$. Then, P_h fulfills $\|P_h\|_{X, X_h} \leq \frac{\|b\|}{\beta_h}$ by the same calculation as in (2.6).*

Theorem 2.30. *Let X, Y be reflexive BANACH spaces and $X_h \subset X, Y_h \subset Y$ finite dimensional subspaces and let $b \in \mathcal{B}(X \times Y, \mathbb{R})$ such that (VP) and (VP_h) are well-posed. Let $x^{\text{sol}} \in X$ solve (VP) and $x_h^{\text{sol}} \in X_h$ solve (VP_h) . Then, it holds*

$$\|x^{\text{sol}} - x_h^{\text{sol}}\|_X \leq \|I_X - P_h\|_{X, X} \inf_{\tilde{x}_h \in X_h} \|x^{\text{sol}} - \tilde{x}_h\|_X. \quad (2.7)$$

Proof. See [71]. Since $b(x^{\text{sol}}, y_h) = \ell(y_h) = b(x_h^{\text{sol}}, y_h)$ for all $y_h \in Y_h$, we observe $P_h x = x_h$. Thus, for $\tilde{x}_h \in X_h$, we have $P_h \tilde{x}_h = \tilde{x}_h$ and we conclude

$$\begin{aligned} \|x^{\text{sol}} - x_h^{\text{sol}}\|_X &= \|(I_X - P_h)x^{\text{sol}}\|_X = \|(I_X - P_h)(x^{\text{sol}} - \tilde{x}_h)\|_X \\ &\leq \|I_X - P_h\|_{X, X} \|x^{\text{sol}} - \tilde{x}_h\|_X. \quad \square \end{aligned}$$

The proof of THEOREM 2.30 shows that P_h is a projection onto X_h . Combined with a result due to KATO, this gives an improved variant of (2.7).

Lemma 2.31 (KATO). *If X is a HILBERT space and $P \in \mathcal{L}(X)$ is a projection with $P \notin \{I_X, 0\}$, we have $\|P_h\|_{X, X} = \|I_X - P_h\|_{X, X}$.*

Proof. See [62] and [71]. □

Remark 2.32. *If X is a HILBERT space, we can improve estimate (2.7) by LEMMA 2.31, since $\|I_X - P_h\|_{X, X} = \|P_h\|_{X, X} \leq \frac{\|b\|}{\beta_h}$:*

$$\|x^{\text{sol}} - x_h^{\text{sol}}\|_X \leq \frac{\|b\|}{\beta_h} \inf_{\tilde{x}_h \in X_h} \|x^{\text{sol}} - \tilde{x}_h\|_X.$$

For BANACH spaces, we get the classical result

$$\begin{aligned} \|x^{\text{sol}} - x_h^{\text{sol}}\|_X &\leq (\|I_X\|_{X, X} + \|P_h\|_{X, X}) \inf_{\tilde{x}_h \in X_h} \|x^{\text{sol}} - \tilde{x}_h\|_X \\ &\leq \left(1 + \frac{\|b\|}{\beta_h}\right) \inf_{\tilde{x}_h \in X_h} \|x^{\text{sol}} - \tilde{x}_h\|_X. \end{aligned}$$

The assertion of THEOREM 2.30 is a generalized version of CÉA's lemma, e.g., see [24, Lem. 2.8].

2.3 Saddle point problems in HILBERT spaces

In the applications later on, we consider variants of (VP) having a particular structure, so-called saddle-point problems. For simplicity, we restrict ourselves to HILBERT spaces.

Definition 2.33 (RIESZ map). For a HILBERT space V , we define the RIESZ map $\Pi_V: V \rightarrow V'$ by $\langle \Pi_V(v), \tilde{v} \rangle := (v, \tilde{v})_V$ for all $v, \tilde{v} \in V$.

Remark 2.34. For a HILBERT space V , we have for $v \in V, v' \in V'$

$$\iota_V(v)(v') = v'(v) = (\Pi_V^{-1}v', v)_V = \langle \Pi_V v, \Pi_V^{-1}v' \rangle = \langle (\Pi_V^{-1})^{\otimes} \Pi_V v, v' \rangle,$$

and as a result $\iota_V = (\Pi_V^{-1})^{\otimes} \Pi_V$ which is an isometric isomorphism by the RIESZ representation theorem, cf. [72, Setion III.6]. So, every HILBERT space is reflexive.

Let V, W be HILBERT spaces and $a \in \mathcal{B}(V \times V, \mathbb{R})$ be a symmetric and positive bilinear form, i.e. $a(v, v) \geq 0$ for $v \in V$. Moreover, let $c \in \mathcal{B}(V \times W, \mathbb{R}), \ell_V \in V'$ and $\ell_W \in W'$. Consider the following variational problem.

$$\left\{ \begin{array}{ll} \text{Find } (v, w) \in V \times W \text{ such that} & \\ a(v, \tilde{v}) + c(\tilde{v}, w) = \ell_V(\tilde{v}), & \text{for all } \tilde{v} \in V, \\ c(v, \tilde{w}) = \ell_W(\tilde{w}) & \text{for all } \tilde{w} \in W. \end{array} \right. \quad (\text{SP})$$

Definition 2.35. To consider (SP), we introduce the following operators.

$$\begin{aligned} A &\in \mathcal{L}(V, V'), & \langle Av, \tilde{v} \rangle &= a(v, \tilde{v}), \\ C &\in \mathcal{L}(V, W'), & C' &\in \mathcal{L}(W, V'), & \langle Cv, w \rangle &= c(v, w) = \langle C'w, v \rangle \end{aligned}$$

for $v, \tilde{v} \in V, w \in W$. We define variants of C with respect to the factor spaces $V/\mathcal{N}(C)$ and $W/\mathcal{N}(C')$, i.e. $\hat{C} \in \mathcal{L}(V/\mathcal{N}(C), W')$, $\hat{C}' \in \mathcal{L}(W/\mathcal{N}(C'), V')$ with

$$\langle \hat{C}\hat{v}, w \rangle = c(v, w) = \langle \hat{C}'\hat{w}, v \rangle$$

for $\hat{v} = v + \mathcal{N}(C) \in V/\mathcal{N}(C), \hat{w} = w + \mathcal{N}(C') \in W/\mathcal{N}(C'), v \in V, w \in W$.

The problem (SP) is called a *saddle point* problem, because the solution is a saddle point of the corresponding LAGRANGE functional.

Proposition 2.36. Define the LAGRANGE function $\mathcal{L}: V \times W \rightarrow \mathbb{R}$ by

$$\mathcal{L}(v, w) = \frac{1}{2}a(v, v) + c(v, w) - \langle \ell_V, v \rangle - \langle \ell_W, w \rangle.$$

Then, $(u, \mu) \in V \times W$ solves (SP) if and only if it is a saddle point of \mathcal{L} , i.e.

$$\mathcal{L}(u, \tilde{w}) \leq \mathcal{L}(u, \mu) \leq \mathcal{L}(\tilde{v}, \mu) \quad \text{for all } (\tilde{v}, \tilde{w}) \in V \times W.$$

Proof. We refer to [24, Prop. 2.39]. □

Theorem 2.37. *Let $\ell_W(\mathcal{N}(C')) = \{0\}$ and $\alpha, \beta > 0$ with*

$$\inf_{v_0 \in \mathcal{N}(C)} \frac{a(v_0, v_0)}{\|v_0\|_V} \geq \alpha, \quad \inf_{w + \mathcal{N}(C') \in W/\mathcal{N}(C')} \sup_{v \in V} \frac{c(v, w)}{\|v\|_V \|w + \mathcal{N}(C')\|_{W/\mathcal{N}(C')}} \geq \beta.$$

Then, there is a unique solution $(u, \hat{\mu}) \in V \times (W/\mathcal{N}(C'))$ of (SP) with

$$\begin{aligned} \|u\|_V &\leq \frac{\|\ell_V\|_{V'}}{\alpha} + \left(\frac{1}{\beta} + \frac{\|a\|}{\alpha\beta} \right) \|\ell_W\|_{W'}, \\ \|\hat{\mu}\|_{W/\mathcal{N}(C')} &\leq \left(\frac{1}{\beta} + \frac{\|a\|}{\alpha} \right) \|\ell_V\|_{V'} + \left(\frac{\|a\|}{\beta^2} + \frac{\|a\|^2}{\alpha\beta^2} \right) \|\ell_W\|_{W'}. \end{aligned}$$

Proof. By assumption, $\ell_W \in \mathcal{N}(C')^\perp$ and $\|\hat{C}'w\|_{V'} \geq \beta\|w + \mathcal{N}(C')\|_{W/\mathcal{N}(C')}$ for $w \in W$. By THEOREM 2.15, there is $\hat{u}_1 \in V/\mathcal{N}(C)$ such that $\hat{C}'u_1 = -\ell_W$ and $\|u_1\|_{V/\mathcal{N}(C)} \leq \frac{1}{\beta}\|\ell_W\|_{W'}$. As a result, we have for all $u_1 \in \hat{u}_1$

$$c(u_1, \tilde{w}) = -\ell_W(\tilde{w}) \quad \forall \tilde{w} \in W, \quad \inf_{\tilde{v}_0 \in \mathcal{N}(C)} \|u_1 + \tilde{v}_0\|_V = \|\hat{u}_1\|_{V/\mathcal{N}(C)} \leq \frac{1}{\beta}\|\ell_W\|_{W'}.$$

Let $u_1 \in \hat{u}_1$ be the minimum norm representative, see LEMMA 2.20, and obtain

$$\|u_1\|_V = \inf_{\tilde{v}_0 \in \mathcal{N}(C)} \|u_1 + \tilde{v}_0\|_V \leq \frac{1}{\beta}\|\ell_W\|_{W'}.$$

Because the space $\mathcal{N}(C)$ is a closed subspace of V , $a|_{\mathcal{N}(C) \times \mathcal{N}(C)}$ fulfills the assumptions of COROLLARY 2.18 (LAX-MILGRAM). Thus, there is $u_0 \in \mathcal{N}(C)$ with

$$a(u_0, \tilde{v}_0) = a(u_1, \tilde{v}_0) - \ell_V(\tilde{v}_0) \quad \text{for all } \tilde{v}_0 \in \mathcal{N}(C).$$

Then, $\|u_0\|_V \leq \frac{\|a\|}{\alpha}\|u_1\|_V + \frac{\|\ell_V\|_{V'}}{\alpha}$ and $u = u_0 - u_1$ fulfills

$$c(u, \tilde{w}) = c(u_0 - u_1, \tilde{w}) = -c(u_1, \tilde{w}) = \ell_W(\tilde{w}) \quad \text{for all } \tilde{w} \in W.$$

By the boundedness of ℓ_V and a and by the construction of u , we conclude that $\ell_V(\cdot) - a(u, \cdot) \in \mathcal{N}(C)^\perp$. Applying THEOREM 2.15 again, we find $\hat{\mu} \in W/\mathcal{N}(C')$ such that $\hat{C}'\hat{\mu} = \ell_V - a(u, \cdot)$. This yields for all $\mu \in \hat{\mu}$

$$c(\tilde{v}, \mu) = \ell_V(\tilde{v}) - a(u, \tilde{v}) \quad \text{for all } \tilde{v} \in V, \quad \|\hat{\mu}\|_{W/\mathcal{N}(C')} \leq \frac{1}{\beta}(\|\ell_V\|_{V'} + \|a\|\|u\|_V).$$

The stability estimates follow from

$$\begin{aligned} \|u\|_V &\leq \|u_0\|_V + \|u_1\|_V \leq \left(1 + \frac{\|a\|}{\alpha}\right) \|u_1\|_V + \frac{\|\ell_V\|_{V'}}{\alpha}, \\ \|\hat{\mu}\|_{W/\mathcal{N}(C')} &\leq \frac{1}{\beta}\|\ell_V\|_{V'} + \frac{\|a\|}{\beta} \left(\left(1 + \frac{\|a\|}{\alpha}\right) \|u_1\|_V + \frac{\|\ell_V\|_{V'}}{\alpha} \right). \end{aligned}$$

This also gives uniqueness since the difference of two solutions of (SP) solves (SP) for zero right-hand side. \square

Remark 2.38. *The ellipticity condition for the bilinear form a can be replaced by a suitable inf-sup condition. In this work, however, the elliptic case is sufficient.*

2.4 Approximation of saddle point problems

Under the same assumptions as for (SP), let $V_h \subset V$ and $W_h \subset W$ be discrete subspaces and consider the variational problem

$$\left\{ \begin{array}{ll} \text{Find } (v_h, w_h) \in V_h \times W_h \text{ such that} & \\ a(v_h, \tilde{v}_h) + c(\tilde{v}_h, w_h) = \ell_V(\tilde{v}_h) & \text{for all } \tilde{v}_h \in V_h, \\ c(v_h, \tilde{w}_h) = \ell_W(\tilde{w}_h) & \text{for all } \tilde{w}_h \in W_h. \end{array} \right. \quad (\text{SP}_h)$$

Definition 2.39. Define the discrete kernels

$$\begin{aligned} V_{0,h} &= \{v_h \in V_h : c(v_h, \tilde{w}_h) = 0 \text{ for all } \tilde{w}_h \in W_h\}, \\ W_{0,h} &= \{w_h \in W_h : c(\tilde{v}_h, w_h) = 0 \text{ for all } \tilde{v}_h \in V_h\}. \end{aligned}$$

We have $\mathcal{N}(C') \subset W_{0,h}$ and $\mathcal{N}(C) \subset V_{0,h}$. However, in general the opposite inclusions are *not* fulfilled.

Theorem 2.40. Assume that $\ell_W(W_{0,h}) = \{0\}$ and $\alpha_h, \beta_h > 0$ exist such that

$$\inf_{v_0 \in V_0} \frac{a(v_0, v_0)}{\|v_0\|_V^2} \geq \alpha_h, \quad \inf_{w_h + W_{0,h} \in W_h/W_{0,h}} \sup_{v_h \in V_h} \frac{c(v_h, w_h)}{\|v_h\|_V \|w_h + W_{0,h}\|_{W_h/W_{0,h}}} \geq \beta_h.$$

Then, (SP_h) has a unique solution $(u_h, \hat{\mu}_h) \in V_h \times W_h/W_{0,h}$ satisfying

$$\begin{aligned} \|u_h\|_V &\leq \frac{\|\ell_V\|_{V'_h}}{\alpha_h} + \left(\frac{1}{\beta_h} + \frac{\|a\|}{\alpha_h \beta_h} \right) \|\ell_W\|_{W'_h}, \\ \|\hat{\mu}\|_{W/\mathcal{N}(C')} &\leq \left(\frac{1}{\beta} + \frac{\|a\|}{\alpha} \right) \|\ell_V\|_{V'_h} + \left(\frac{\|a\|}{\beta_h^2} + \frac{\|a\|^2}{\alpha_h \beta_h^2} \right) \|\ell_W\|_{W'_h}. \end{aligned}$$

Proof. This is a direct consequence of THEOREM 2.37. \square

Theorem 2.41. Let the assumptions of THEOREM 2.37 and THEOREM 2.40 be fulfilled. Then, we have for all $(v_h, w_h) \in V_h \times W_h$

$$\begin{aligned} \|u - u_h\|_V &\leq \left(1 + \frac{\|a\|}{\alpha_h} \right) \left(1 + \frac{\|c\|}{\beta_h} \right) \|v_h - u\|_V + \frac{1}{\alpha_h} \sup_{\tilde{v}_h \in V_{0,h}} \frac{c(\tilde{v}_h, \mu - w_h)}{\|\tilde{v}_h\|_V}, \\ \|\hat{\mu} - \hat{\mu}_h\|_{W/W_{0,h}} &\leq \left(1 + \frac{\|c\|}{\beta_h} \right) \|\hat{\mu} - \hat{w}_h\|_{W/W_{0,h}} + \frac{\|a\|}{\beta_h} \|u - u_h\|_V, \end{aligned}$$

where $(u, \hat{\mu}) \in V \times W/\mathcal{N}(C')$ and $(u_h, \hat{\mu}_h) \in V \times W_h/W_{0,h}$ are the solutions of (SP) and (SP_h), respectively.

Proof. Fix arbitrary $(v_h, w_h) \in V_h \times W_h$. We have $c(u - v_h, \cdot) \in W_{0,h}^\perp \subset W'_h$ since $c(u - v_h, \tilde{w}_h) = \ell_W(\tilde{w}_h) - c(v_h, \tilde{w}_h) = 0$ for $\tilde{w}_h \in W_{0,h}$ by (SP_h). Thus, we find $r_h \in V_h$ with

$$c(r_h, \tilde{w}_h) = c(u - v_h, \tilde{w}_h) \quad \text{for all } \tilde{w}_h \in W_h, \quad \|r_h\|_V \leq \frac{1}{\beta_h} \sup_{\tilde{w}_h \in W_h} \frac{c(u - v_h, \tilde{w}_h)}{\|\tilde{w}_h\|_W}.$$

As a result, we have for $\phi_h := r_h + v_h \in V_h$ that $\phi_h - u_h \in V_{0,h}$. This and

$$a(u - u_h, \tilde{v}_h) = (\ell_V(\tilde{v}_h) - c(\tilde{v}_h, \mu)) - (\ell_V(\tilde{v}_h) - c(\tilde{v}_h, \mu_h)) = c(\tilde{v}_h, \mu_h - \mu) \quad (2.8)$$

for $\mu \in \hat{\mu}$, $\mu_h \in \hat{\mu}_h$ and $\tilde{v}_h \in V_h$ imply

$$\begin{aligned} \alpha_h \|\phi_h - u_h\|_V &\leq \sup_{\tilde{v}_h \in V_{0,h}} \frac{a(\phi_h - u_h, \tilde{v}_h)}{\|\tilde{v}_h\|_V} = \sup_{\tilde{v}_h \in V_{0,h}} \frac{a(\phi_h - u, \tilde{v}_h) + a(u - u_h, \tilde{v}_h)}{\|\tilde{v}_h\|_V} \\ &= \sup_{\tilde{v}_h \in V_{0,h}} \frac{a(\phi_h - u, \tilde{v}_h) + c(\tilde{v}_h, \mu_h - \mu)}{\|\tilde{v}_h\|_V} \\ &\leq \|a\| \|\phi_h - u\|_V + \sup_{\tilde{v}_h \in V_{0,h}} \frac{c(\tilde{v}_h, w_h - \mu)}{\|\tilde{v}_h\|_V}. \end{aligned}$$

Using $\|\phi_h - u\|_V \leq \|v_h - u\|_V + \|r_h\|_V$, we conclude

$$\begin{aligned} \|u - u_h\|_V &\leq \|u - \phi_h\|_V + \|\phi_h - u_h\|_V \\ &\leq \left(1 + \frac{\|a\|}{\alpha_h}\right) \|\phi_h - u\|_V + \frac{1}{\alpha_h} \sup_{\tilde{v}_h \in V_{0,h}} \frac{c(\tilde{v}_h, w_h - \mu)}{\|\tilde{v}_h\|_V} \\ &\leq \left(1 + \frac{\|a\|}{\alpha_h}\right) \left(\|v_h - u\|_V + \frac{1}{\beta_h} \sup_{\tilde{w}_h \in W_h} \frac{c(u - v_h, \tilde{w}_h)}{\|\tilde{w}_h\|_W} \right) \\ &\quad + \frac{1}{\alpha_h} \sup_{\tilde{v}_h \in V_{0,h}} \frac{c(\tilde{v}_h, w_h - \mu)}{\|\tilde{v}_h\|_V} \\ &\leq \left(1 + \frac{\|a\|}{\alpha_h}\right) \left(1 + \frac{\|c\|}{\beta_h}\right) \|v_h - u\|_V + \frac{1}{\alpha_h} \sup_{\tilde{v}_h \in V_{0,h}} \frac{c(\tilde{v}_h, w_h - \mu)}{\|\tilde{v}_h\|_V}. \end{aligned}$$

To obtain the estimate for $\|\hat{\mu} - \hat{\mu}_h\|_{W/W_{0,h}}$, we again use (2.8) and get

$$c(v_h, w_h - \mu_h) = -a(u - u_h, v_h) - c(v_h, \mu - w_h)$$

for all $\mu \in \hat{\mu}$ and $\mu_h \in \hat{\mu}_h$. The discrete inf-sup stability of c yields

$$\begin{aligned} \beta_h \|w_h - \mu_h\|_W &\leq \sup_{v_h \in V_h} \frac{c(v_h, w_h - \mu_h)}{\|v_h\|_V} \\ &= \sup_{v_h \in V_h} \frac{c(v_h, \mu - w_h) + a(u - u_h, v_h)}{\|v_h\|_V} \\ &\leq \|c\| \|\mu - w_h\|_W + \|a\| \|u - u_h\|_V \end{aligned}$$

and the triangle inequality finishes the proof by

$$\|\mu - \mu_h\|_W \leq \|\mu - w_h\|_W + \|w_h - \mu_h\|_W. \quad \square$$

Theorem 2.42. *Assume that the assumptions of THEOREM 2.37 and THEOREM 2.40 are satisfied. Further, assume that $\ell_W = 0 \in W'$.*

Then, we have $u \in \mathcal{N}(C)$ and $u_h \in V_{h,0}$ and

$$\|u - u_h\|_V \leq \left(1 + \frac{\|a\|}{\alpha_h}\right) \inf_{v_{h,0} \in V_{h,0}} \|u - v_{h,0}\|_V + \frac{1}{\alpha_h} \sup_{v_{h,0} \in V_{h,0}} \frac{a(u, v_{h,0}) - \ell_V(v_{h,0})}{\|v_{h,0}\|_V}.$$

Proof. (SP) and (SP_h) directly imply $u \in \mathcal{N}(C)$ and $u_h \in V_{h,0}$ since

$$c(u, w) = 0 \quad \text{for all } w \in W \quad \text{and} \quad c(u_h, w_h) = 0 \quad \text{for all } w_h \in W_h.$$

For $v_{h,0} \in V_{h,0}$ by $a(u_h, v_{h,0}) = \ell_V(v_{h,0}) - c(v_{h,0}, w_h) = \ell_V(v_{h,0})$, $w_h \in W_h$, we obtain

$$\begin{aligned} \alpha_h \|u_h - v_{h,0}\|_V^2 &\leq a(u_h - v_{h,0}, u_h - v_{h,0}) \\ &= a(u - v_{h,0}, u_h - v_{h,0}) - a(u - u_h, u_h - v_{h,0}) \\ &= a(u - v_{h,0}, u_h - v_{h,0}) - a(u, u_h - v_{h,0}) + \ell_V(u_h - v_{h,0}) \\ &\leq \|a\| \|u - v_{h,0}\|_V \|u_h - v_{h,0}\|_V \\ &\quad + \|a(u, \cdot) - \ell_V(\cdot)\|_{V'_{h,0}} \|u_h - v_{h,0}\|_V \end{aligned}$$

which yields by the triangle inequality

$$\begin{aligned} \alpha_h \|u - u_h\|_V &\leq \alpha_h \|u - v_{h,0}\|_V + \alpha_h \|v_{h,0} - u_h\|_V \\ &\leq (\alpha_h + \|a\|) \|u - v_{h,0}\|_V + \sup_{v_{h,0} \in V_{h,0}} \frac{a(u, v_{h,0}) - \ell_V(v_{h,0})}{\|v_{h,0}\|_V}. \end{aligned}$$

This finishes the proof. □

Chapter 3

Mathematical modeling of acoustic waves

3.1 The acoustic wave equation

We consider a bounded LIPSCHITZ domain $\Omega \subset \mathbb{R}^d$ and a time interval $(0, T)$ yielding the space-time cylinder $Q = \Omega \times (0, T)$. For given right hand-side $\mathbf{b}(x, t) = (f, \mathbf{g})$, as well as density and compression modulus distributions $\rho(x)$, $\kappa(x)$, we search for solutions of the acoustic wave equation, i.e.

$$\begin{aligned}\kappa^{-1} \partial_t p - \operatorname{div} \mathbf{v} &= f, \\ \rho \partial_t \mathbf{v} - \nabla p &= \mathbf{g},\end{aligned}\tag{3.1}$$

where the unknown $\mathbf{y} = (p, \mathbf{v})$ is the space-time wavefield.

Introducing the block operators M and A by

$$M(x) = \begin{pmatrix} \kappa^{-1}(x) & 0 \\ 0 & \rho(x)I_d \end{pmatrix}, \quad A = \begin{pmatrix} 0 & \operatorname{div} \\ \nabla & 0 \end{pmatrix},$$

we can rewrite (3.1) as $L\mathbf{y} = \mathbf{b}$ for the space-time differential operator

$$L\mathbf{y} = M\partial_t \mathbf{y} - A\mathbf{y} = (\kappa^{-1}\partial_t p - \operatorname{div} \mathbf{v}, \rho \partial_t \mathbf{v} - \nabla p).$$

Now we establish an analytic HILBERT space setting for a unique solution of

$$L(p, \mathbf{v}) = (f, \mathbf{g})\tag{3.2}$$

(subject to initial and boundary conditions) which depends continuously on the data.

To keep our notation simple, we restrict ourselves to the case $\rho \equiv \kappa \equiv 1$ in the following, i.e. $M \equiv I_{1+d}$.

3.1.1 The semigroup setting

We consider the ODE

$$\partial_t(p, \mathbf{v}) = A(p, \mathbf{v}) + (f, \mathbf{g}), \quad A(p, \mathbf{v}) = (\nabla \cdot \mathbf{v}, \nabla p),$$

where the operator A is associated with a dense domain

$$\mathcal{D}(A) \subset L_2(\Omega; \mathbb{R} \times \mathbb{R}^d).$$

Here, we choose $\mathcal{D}(A) = H_0^1(\Omega) \times H(\operatorname{div}, \Omega)$ including homogeneous Dirichlet boundary conditions for the pressure on $\partial\Omega$.

We show that the operator A with domain $\mathcal{D}(A)$ generates a semigroup. Therefore, we check the requirements of the LUMER-PHILLIPS theorem.

Theorem 3.1 (LUMER-PHILLIPS). *Let Y be a HILBERT space and let A be a linear operator in Y satisfying the following conditions for an $\omega \in \mathbb{R}$*

1. $\mathcal{D}(A)$ is dense in Y .
2. $(y, Ay)_Y \leq \omega \|y\|_Y^2$ for every $y \in \mathcal{D}(A)$.
3. There exists $\lambda_0 > \omega$ such that $A - \lambda_0 \operatorname{id}$ is onto.

Then A generates a quasicontraction semigroup with $\|\exp(tA)\| \leq \exp(\omega t)$.

Proof. See e.g. [56, Thm. 12.22]. □

To check the requirements, we choose $\omega = 0$ in THEOREM 3.1 since

$$(A(p, \mathbf{v}), (p, \mathbf{v}))_\Omega = 0, \quad (p, \mathbf{v}) \in \mathcal{D}(A). \quad (3.3)$$

Then, we set $\lambda_0 = 1 > \omega$ and show that $\operatorname{id} - A$ is surjective. For a given right-hand side $(f, \mathbf{g}) \in L_2(\Omega; \mathbb{R} \times \mathbb{R}^d)$, we define $p \in H_0^1(\Omega)$ solving

$$(\nabla p, \nabla q)_\Omega + (p, q)_\Omega = (f, q)_\Omega - (\mathbf{g}, \nabla q)_\Omega, \quad q \in H_0^1(\Omega),$$

and then we define $\mathbf{v} = \mathbf{g} + \nabla p$. We observe

$$(\mathbf{v}, \nabla q)_\Omega = (f, q)_\Omega - (p, q)_\Omega, \quad q \in C_c^1(\Omega),$$

i.e., $\mathbf{v} \in H(\operatorname{div}, \Omega)$ and $\nabla \cdot \mathbf{v} = p - f$, so that together $(p, \mathbf{v}) - A(p, \mathbf{v}) = (f, \mathbf{g})$. This gives surjectivity.

According to THEOREM 3.1, the operator A generates a semigroup. See also [36, Sect. 2.2] and [43] for the application to general linear wave equations, in particular for the case of non-constant material parameters.

3.1.2 Duality, adjoint operators and the HILBERT adjoint

In the next section, many arguments rely on duality. For this purpose, we introduce the HILBERT adjoint A^* of the operator A with domain $\mathcal{D}(A^*)$, cf. [56, Sect. 8.4.2].

Remark 3.2. *In case of acoustic waves, we have $A^* = -A$. However, since the considerations also apply for operators with a different adjoint, we treat A^* explicitly.*

The adjoint operator is defined in the domain

$$\mathcal{D}(A^*) = \{(q, \mathbf{w}) \in L_2(\Omega; \mathbb{R} \times \mathbb{R}^d) : (f, \mathbf{g}) \in L_2(\Omega; \mathbb{R} \times \mathbb{R}^d) \text{ exists} \\ \text{such that } ((f, \mathbf{g}), (p, \mathbf{v}))_\Omega = ((q, \mathbf{w}), A(p, \mathbf{v}))_\Omega \text{ for } (p, \mathbf{v}) \in \mathcal{D}(A)\}.$$

For the acoustic wave equation we have $\mathcal{D}(A^*) = \mathbf{H}_0^1(\Omega) \times \mathbf{H}(\text{div}, \Omega) = \mathcal{D}(A)$.

Then, for $(q, \mathbf{w}) \in \mathcal{D}(A^*)$ we define A^* by

$$(A^*(q, \mathbf{w}), (p, \mathbf{v}))_\Omega = ((q, \mathbf{w}), A(p, \mathbf{v}))_\Omega, \quad (p, \mathbf{v}) \in \mathcal{D}(A), (q, \mathbf{w}) \in \mathcal{D}(A^*).$$

Since $\mathcal{D}(A) \subset L_2(\Omega; \mathbb{R} \times \mathbb{R}^d)$ is dense, this defines $A^*(q, \mathbf{w}) \in L_2(\Omega; \mathbb{R} \times \mathbb{R}^d)$.

Correspondingly, for the space-time operator $L = \partial_t - A$ the formal adjoint of the differential operator is given by $L^* = -\partial_t - A^*$, and we obtain in $Q = (0, T) \times \Omega$

$$(L^*(q, \mathbf{w}), (p, \mathbf{v}))_Q = ((q, \mathbf{w}), L(p, \mathbf{v}))_Q, \quad (p, \mathbf{v}), (q, \mathbf{w}) \in C_c^1(Q; \mathbb{R} \times \mathbb{R}^d).$$

In our application the adjoint problem describes a wave equation backward in time.

In the next section we define suitable domains for the operators L and L^* extending the domains $\mathcal{D}(A)$ and $\mathcal{D}(A^*)$ in $L_2(\Omega; \mathbb{R} \times \mathbb{R}^d)$ to domains of the space-time operators in $L_2((0, T) \times \Omega; \mathbb{R} \times \mathbb{R}^d)$, so that L^* is the Hilbert adjoint of L in this setting.

3.2 A variational space-time setting

We consider the ODE

$$\partial_t \mathbf{y} = A\mathbf{y} + \mathbf{b} \quad \text{in } [0, T], \quad \mathbf{y}(0) = \mathbf{0}, \quad (3.4)$$

where A is an operator with a dense domain $\mathcal{D}(A)$ in $Y = L_2(Q; \mathbb{R}^m)$. We assume that the operator A generates a semigroup. Then, for all

$$\mathbf{b} \in W^{1,1}((0, T); Y)$$

a solution $\mathbf{y} \in C^1([0, T]; Y) \cap C^0([0, T]; \mathcal{D}(A))$ of (3.4) exists and is of the form

$$\mathbf{y}(t) = \int_0^t \exp((t-s)A) [\mathbf{b}(s)] ds, \quad t \in [0, T].$$

This directly implies

$$\|\mathbf{y}(t)\|_{\Omega} \leq \int_0^t \|\exp((t-s)A)\|_{\Omega} \|\mathbf{b}(s)\|_{\Omega} ds, \quad t \in [0, T].$$

In case of hyperbolic operators satisfying (3.3) we have $\|\exp(tA)\|_{\Omega} = 1$, see, e.g., [56, Thm. 12.22]. Then, $\|\mathbf{y}(t)\|_{\Omega} \leq \int_0^t \|\mathbf{b}(s)\|_{\Omega} ds$ for $t \in (0, T)$ and integration in time yields

$$\begin{aligned} \|\mathbf{y}\|_{\Omega \times (0, T)} &\leq \left(\int_0^T \left(\int_0^t \|\mathbf{b}(s)\|_{\Omega} ds \right)^2 dt \right)^{1/2} \leq \left(\int_0^T t \|\mathbf{b}\|_{\Omega \times (0, t)}^2 dt \right)^{1/2} \\ &\leq \left(\int_0^T t dt \|\mathbf{b}\|_{\Omega \times (0, T)}^2 \right)^{1/2} \leq \frac{T}{\sqrt{2}} \|\mathbf{b}\|_{\Omega \times (0, T)}. \end{aligned} \quad (3.5)$$

The ODE solution (3.4) belongs to the Banach space

$$\mathcal{V} = \left\{ \mathbf{y} \in C^1([0, T]; Y) \cap C^0([0, T]; \mathcal{D}(A)) : \mathbf{y}(0) = \mathbf{0} \right\},$$

and we obtain for all $\mathbf{b} \in W^{1,1}((0, T); Y)$ a solution $\mathbf{y} \in \mathcal{V}$ with $L\mathbf{y} = \mathbf{b}$, see [56, Thm. 12.16]. Note that L is not a closed operator in \mathcal{V} .

Since $W^{1,1}((0, T); Y)$ is dense in $L_2((0, T); Y)$, we obtain the following result.

Lemma 3.3. *$L(\mathcal{V})$ is dense in $L_2((0, T); Y)$.*

In our application also the adjoint operator A^* generates a semigroup. Thus, this result transfers to the adjoint problem, given by the ODE backward in time

$$-\partial_t \mathbf{z} = A^* \mathbf{z} + \mathbf{c} \quad \text{in } [0, T], \quad \mathbf{z}(T) = \mathbf{0}. \quad (3.6)$$

Thus, for $\mathbf{c} \in W^{1,1}((0, T); Y)$ the solution of $L^* \mathbf{z} = \mathbf{c}$ is given by

$$\mathbf{z}(t) = \int_t^T \exp((s-t)A^*) [\mathbf{c}(s)] ds.$$

Defining

$$\mathcal{V}^* = \left\{ \mathbf{z} \in C^1([0, T]; Y) \cap C^0([0, T]; \mathcal{D}(A^*)) : \mathbf{z}(T) = \mathbf{0} \right\}$$

this shows that $L^*(\mathcal{V}^*)$ is dense in $L_2((0, T); Y)$, and we have

$$(L^*(q, \mathbf{w}), (p, \mathbf{v}))_Q = ((q, \mathbf{w}), L(p, \mathbf{v}))_Q, \quad (p, \mathbf{v}) \in \mathcal{V}, \quad (q, \mathbf{w}) \in \mathcal{V}^*.$$

3.2.1 A space-time HILBERT space setting

In $W = L_2((0, T); Y) = L_2(Q; \mathbb{R}^m)$ we define the space

$$\begin{aligned} \mathbb{H}(L, Q) &:= \left\{ \mathbf{y} \in W : L\mathbf{y} \in W \right\} \\ &:= \left\{ \mathbf{y} \in W : \mathbf{b} \in W \text{ exists such that} \right. \\ &\quad \left. (\mathbf{b}, \mathbf{z})_Q = (\mathbf{y}, L^* \mathbf{z})_Q \text{ for } \mathbf{z} \in C_c^1(Q; \mathbb{R}^m) \right\}. \end{aligned} \quad (3.7)$$

For $\mathbf{y} \in \mathbf{H}(L, Q)$, we define $L\mathbf{y} := \mathbf{b}$ with $\mathbf{b} \in W$ as in (3.7). Since $C_c^1(Q; \mathbb{R}^m)$ is dense in W , this uniquely defines $L\mathbf{y}$ and we have $(L\mathbf{y}, \mathbf{z})_Q = (\mathbf{y}, L^*\mathbf{z})_Q$ for $\mathbf{y} \in \mathbf{H}(L, Q)$, $\mathbf{z} \in C_c^1(Q, \mathbb{R}^m)$.

Proposition 3.4. $\mathbf{H}(L, Q)$ is a HILBERT space with respect to the graph norm

$$\|\mathbf{y}\|_{L, Q} = \sqrt{\|\mathbf{y}\|_Q^2 + \|L\mathbf{y}\|_Q^2}, \quad \mathbf{y} \in \mathbf{H}(L, Q).$$

Proof. Straight-forward calculations show that $\mathbf{H}(L, Q)$ is a vector space and that $(\mathbf{y}, \tilde{\mathbf{y}})_{L, Q} := (\mathbf{y}, \tilde{\mathbf{y}})_Q + (L\mathbf{y}, L\tilde{\mathbf{y}})_Q$, $\mathbf{y}, \tilde{\mathbf{y}} \in \mathbf{H}(L, Q)$, is an inner product.

In order to show that $\mathbf{H}(L, Q)$ is complete, we consider a CAUCHY sequence $(\mathbf{y}_n)_n \in \mathbf{H}(L, Q)^\mathbb{N}$. Then, $(\mathbf{y}_n)_n$ and $(L\mathbf{y}_n)_n$ are CAUCHY sequences in W possessing limits $\mathbf{y} \in W$ and $\mathbf{b} \in W$, respectively, by the completeness of W . Using the triangle inequality and $(L\mathbf{y}_n, \mathbf{z})_Q = (\mathbf{y}_n, L^*\mathbf{z})_Q$, $\mathbf{z} \in C_c^1(Q, \mathbb{R}^m)$, $n \in \mathbb{N}$, we conclude

$$\begin{aligned} |(\mathbf{b}, \mathbf{z})_Q - (\mathbf{y}, L^*\mathbf{z})_Q| &\leq |(\mathbf{b}, \mathbf{z})_Q - (L\mathbf{y}_n, \mathbf{z})_Q| + |(L\mathbf{y}_n, \mathbf{z})_Q - (\mathbf{y}, L^*\mathbf{z})_Q| \\ &\leq \|\mathbf{b} - L\mathbf{y}_n\|_Q \|\mathbf{z}\|_Q + \|\mathbf{y}_n - \mathbf{y}\|_Q \|L^*\mathbf{z}\|_Q \\ &\longrightarrow 0 \quad \text{for } n \longrightarrow \infty. \end{aligned}$$

Thus, $(\mathbf{b}, \mathbf{z})_Q = (\mathbf{y}, L^*\mathbf{z})_Q$ for all $\mathbf{z} \in C_c^1(Q, \mathbb{R}^m)$ implying $\mathbf{y} \in \mathbf{H}(L, Q)$. \square

Analogously, we define $\mathbf{H}(L^*, Q) = \{\mathbf{y} \in W : L^*\mathbf{y} \in W\}$ and let $\mathbf{H}(L^*, Q)'$ denote its dual space. We define the operator $D \in \mathcal{L}(\mathbf{H}(L, Q), \mathbf{H}(L^*, Q)')$ by

$$\langle D\mathbf{y}, \mathbf{z} \rangle = (L\mathbf{y}, \mathbf{z})_Q - (\mathbf{y}, L^*\mathbf{z})_Q, \quad \mathbf{y} \in \mathbf{H}(L, Q), \quad \mathbf{z} \in \mathbf{H}(L^*, Q),$$

and we denote the kernel of D by

$$\mathcal{N}(D) = \left\{ \mathbf{y} \in \mathbf{H}(L, Q) : D\mathbf{y} = \mathbf{0} \right\}.$$

By definition of the adjoint operator L^* , we have $C_c^1(Q; \mathbb{R}^m) \subset \mathcal{N}(D)$. Thus, the operator D describes traces obtained using integration by parts in abstract form.

Let $\mathbf{H}_0(L, Q) \subset \mathbf{H}(L, Q)$ be the closure of $C_c^1(Q; \mathbb{R}^m) \subset \mathcal{N}(D)$. Then, also $\mathbf{H}_0(L, Q) \subset \mathcal{N}(D)$.

In fact, we can establish equality. The proof is based on a duality argument using the operator $D' \in \mathcal{L}(\mathbf{H}(L^*, Q), \mathbf{H}(L, Q)')$ with

$$\langle D'\mathbf{z}, \mathbf{y} \rangle = (L\mathbf{y}, \mathbf{z})_Q - (\mathbf{y}, L^*\mathbf{z})_Q = \langle D\mathbf{y}, \mathbf{z} \rangle.$$

Theorem 3.5. We have

$$\mathbf{H}_0(L, Q) = \mathcal{N}(D).$$

Proof. We only have to show $\mathcal{N}(D) \subset \mathbf{H}_0(L, Q)$. Provided we have established $\mathbf{C}_c^1(Q; \mathbb{R}^m)^\perp \subset D'(\mathbf{H}(L^*, Q))$, see DEFINITION 2.6, the assertion follows from

$$\begin{aligned} \mathcal{N}(D) &= \left\{ \mathbf{y} \in \mathbf{H}(L, Q) : \langle D\mathbf{y}, \mathbf{z} \rangle = \mathbf{0} = \langle D'\mathbf{z}, \mathbf{y} \rangle \text{ for } \mathbf{z} \in \mathbf{H}(L^*, Q) \right\} \\ &= {}^\perp D'(\mathbf{H}(L^*, Q)) \subset {}^\perp (\mathbf{C}_c^1(Q; \mathbb{R}^m)^\perp) = \mathbf{H}_0(L, Q). \end{aligned}$$

The proof uses the technique in [26, Lem. 2.4], see also [14, Lem. 2.2] and [67, Lem. 1]. For a given functional $\ell \in \mathbf{C}_c^1(Q; \mathbb{R}^m)^\perp \subset \mathbf{H}(L, Q)'$, we construct $\mathbf{z} \in \mathbf{H}(L^*, Q)$ with $D'\mathbf{z} = \ell$. Therefore, we define $\mathbf{y} \in \mathbf{H}(L, Q)$ as the RIESZ representative of ℓ in $\mathbf{H}(L, Q)$ solving

$$(L\mathbf{y}, L\phi)_Q + (\mathbf{y}, \phi)_Q = \langle \ell, \phi \rangle, \quad \phi \in \mathbf{H}(L, Q). \quad (3.8)$$

Then, since $\langle \ell, \mathbf{w} \rangle = 0$ for test functions $\mathbf{w} \in \mathbf{C}_c^1(Q; \mathbb{R}^m)$, we observe

$$(\mathbf{y}, \phi)_Q = -(L\mathbf{y}, L\phi)_Q, \quad \phi \in \mathbf{C}_c^1(Q; \mathbb{R}^m).$$

Inserting $\mathbf{z} = L\mathbf{y}$ and using the definition of $\mathbf{H}(L^*, Q)$, we observe $\mathbf{z} \in \mathbf{H}(L^*, Q)$ and $L^*\mathbf{z} = -\mathbf{y}$. From (3.8), we now obtain

$$\begin{aligned} \langle D'\mathbf{z}, \phi \rangle &= (L\phi, \mathbf{z})_Q - (\phi, L^*\mathbf{z})_Q \\ &= (L\phi, L\mathbf{y})_Q + (\phi, \mathbf{y})_Q = \langle \ell, \phi \rangle, \quad \phi \in \mathbf{H}(L, Q), \end{aligned}$$

i.e., $D'\mathbf{z} = \ell$ in $\mathbf{H}(L, Q)'$. □

Example 3.6. *It is a natural question to ask whether the space $\mathbf{H}(L, Q)$ is larger than solution space $\mathcal{S} := \mathbf{C}^1([0, T]; Y) \cap \mathbf{C}^0([0, T]; \mathcal{D}(A))$ in the semigroup setting. Here, we provide an example $(p, \mathbf{v}) \in \mathbf{H}(L, Q) \setminus \mathcal{S}$.*

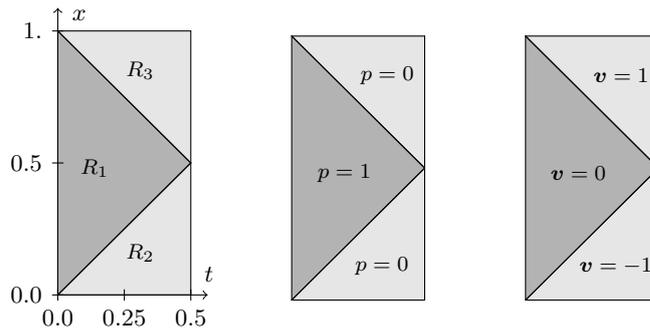


Figure 3.1: A function in $\mathbf{H}(L, Q) \setminus \mathbf{C}^1([0, T]; Y) \cap \mathbf{C}^0([0, T]; \mathcal{D}(A))$, $d = 1$.

Let $d = 1$, $\Omega = (0, 1)$, $T = \frac{1}{2}$ and $Q = (0, 1) \times (0, \frac{1}{2})$ the space-time cylinder. We partition Q into three triangular subdomains R_1, R_2, R_3 with

$$R_1 = \{(x, t) \in Q : x > t, x < 1 - t\}, \quad R_{2/3} = \{(x, t) \in Q \setminus R_1 : x \leq 1/2\},$$

see FIGURE 3.1. Following [45], we define $(p, \mathbf{v}) \in L_2(Q, \mathbb{R}^2)$ by

$$(p, \mathbf{v})(x, t) = \begin{cases} (1, 0) & (x, t) \in R_1, \\ (0, -1) & (x, t) \in R_2, \\ (0, 1) & (x, t) \in R_3. \end{cases}$$

Fixing $(\psi, \phi) \in C_c^1(Q, \mathbb{R}^2)$, we obtain by applying the product rule

$$(\partial_t, \partial_x) \cdot (f\psi, f\phi) = (\partial_t f, \partial_x f) \cdot (\psi, \phi) + f(\partial_t \psi + \partial_x \phi), \quad f \in C^1(Q).$$

as a point-wise equality. Using GAUSS' divergence theorem in every subdomain R_k , $k = 1, 2, 3$, and $L^*(\psi, \phi) = -\left(\frac{\partial_t \psi + \partial_x \phi}{\partial_t \phi + \partial_x \psi}\right)$, this yields

$$-((p, \mathbf{v}), L^*(\psi, \phi))_{R_k} = (p, (\psi, \phi) \cdot n_{R_k})_{\partial R_k} + (\mathbf{v}, (\phi, \psi) \cdot n_{R_k})_{\partial R_k}$$

since (p, \mathbf{v}) is subdomain-wise constant making the volume terms vanish. By the definition of (p, \mathbf{v}) , we have

$$\begin{aligned} (p, (\psi, \phi) \cdot n_{R_1})_{\partial R_1} &= \frac{1}{\sqrt{2}} \int_0^{1/2} (\psi(s, s) - \phi(s, s) + \psi(s, 1-s) + \phi(s, 1-s)) ds, \\ (\mathbf{v}, (\phi, \psi) \cdot n_{R_2})_{\partial R_2} &= \frac{1}{\sqrt{2}} \int_0^{1/2} (-\psi(s, s) + \phi(s, s)) ds, \\ (\mathbf{v}, (\phi, \psi) \cdot n_{R_3})_{\partial R_3} &= \frac{1}{\sqrt{2}} \int_0^{1/2} (-\phi(s, 1-s) - \psi(s, 1-s)) ds. \end{aligned}$$

All remaining integrals vanish by definition of (p, \mathbf{v}) giving $(p, \mathbf{v}) \in H(L, Q)$ by

$$((p, \mathbf{v}), L^*(\psi, \phi))_Q = \sum_{k=1}^3 ((p, \mathbf{v}), L^*(\psi, \phi))_{R_k} = 0 \quad \text{for all } (\psi, \phi) \in C_c^1(Q, \mathbb{R}^2),$$

which means $L(p, \mathbf{v}) = 0$.

On the other hand, for every $t \in (0, \frac{1}{2})$ we have $p(\cdot, t), \mathbf{v}(\cdot, t) \notin C(0, 1)$. However, by SOBOLEV's embedding theorem, it holds $H^1(0, 1) \subset C(0, 1)$ implying that $p(\cdot, t), \mathbf{v}(\cdot, t) \notin H^1(0, 1)$. Since we have for the domain $\mathcal{D}(A) = H_0^1(0, 1)^2$, we conclude $(p, \mathbf{v}) \notin \mathcal{S}$.

3.2.2 The closure of the space-time operator (L, \mathcal{V})

We assume that $C_L > 0$ exists such that

$$\|\mathbf{y}\|_Q \leq C_L \|L\mathbf{y}\|_Q, \quad \mathbf{y} \in \mathcal{V}. \quad (3.9)$$

In case of hyperbolic operators, this is obtained from (3.5) with $C_L = \frac{1}{\sqrt{2}}T$, see also [55, Thm. 3.1], [23, Lem. 1], and [68, Lem. 6].

In particular, L is injective on \mathcal{V} . Now, we define

$$V := {}^\perp(\mathcal{V}^\perp) \subset \mathbf{H}(L, Q),$$

i.e., V is the closure of \mathcal{V} in $\mathbf{H}(L, Q)$ with respect to the graph norm, see DEFINITION 2.6. By continuity, the estimate (3.9) also holds for the closure, i.e.,

$$\|\mathbf{y}\|_Q \leq C_L \|L\mathbf{y}\|_Q, \quad \mathbf{y} \in V. \quad (3.10)$$

Theorem 3.7. $L \in \mathcal{L}(V, W)$ is a bijection.

Proof. From (3.10) we observe that L is injective, and since $V \subset \mathbf{H}(L, Q)$ is closed, $L(V) \subset W$ has closed range. This is shown as follows: for any sequence $(\mathbf{y}_n)_n \in V^\mathbb{N}$ with $\lim_{n \rightarrow \infty} L\mathbf{y}_n = \mathbf{b} \in W$ we have

$$\|\mathbf{y}_n - \mathbf{y}_k\|_Q + \|L\mathbf{y}_n - L\mathbf{y}_k\|_Q \leq (C_L + 1)\|L\mathbf{y}_n - L\mathbf{y}_k\|_Q \rightarrow 0, \quad k, n \rightarrow \infty.$$

Thus, $(\mathbf{y}_n)_n$ is a Cauchy sequence in $\mathbf{H}(L, Q)$ and since $V \subset \mathbf{H}(L, Q)$ is closed, $\mathbf{y} := \lim_{n \rightarrow \infty} \mathbf{y}_n \in V$ with $L\mathbf{y} = \mathbf{b}$ exists. Since $L(\mathcal{V}) \subset W$ is dense (LEMMA 3.3), we obtain $L(V) = W$. \square

Remark 3.8. The assertion of THEOREM 3.7 is a general result for operators: if L satisfies (3.9) and $L(\mathcal{V}) \subset W$ is dense, then L extends to a bijection in the closure $V = {}^\perp(\mathcal{V}^\perp)$.

The estimate (3.9) transfers to the adjoint operator, i.e., we have for $\mathbf{z} \in \mathcal{V}^*$

$$\|\mathbf{z}\|_Q = \sup_{\mathbf{b} \in W} \frac{(\mathbf{z}, \mathbf{b})_Q}{\|\mathbf{b}\|_Q} = \sup_{\mathbf{y} \in \mathcal{V}} \frac{(\mathbf{z}, L\mathbf{y})_Q}{\|L\mathbf{y}\|_Q} = \sup_{\mathbf{y} \in \mathcal{V}} \frac{(L^*\mathbf{z}, \mathbf{y})_Q}{\|L\mathbf{y}\|_Q} \leq C_L \|L^*\mathbf{z}\|_Q$$

again using the density of $L(\mathcal{V})$ in W , and exploiting

$$\langle D'\mathbf{z}, \mathbf{y} \rangle = (L\mathbf{y}, \mathbf{z})_Q - (\mathbf{y}, L^*\mathbf{z})_Q = 0, \quad \mathbf{y} \in \mathcal{V}, \mathbf{z} \in \mathcal{V}^*, \quad (3.11)$$

which holds by construction of \mathcal{V} and \mathcal{V}^* . Defining $V^* := {}^\perp(\mathcal{V}^*)^\perp \subset \mathbf{H}(L^*, Q)$, the estimate corresponding to (3.10) also holds for the closure of the adjoint operator L^* , i.e.,

$$\|\mathbf{z}\|_Q \leq C_L \|L^*\mathbf{z}\|_Q, \quad \mathbf{z} \in V^*. \quad (3.12)$$

Theorem 3.9. We have

$$\begin{aligned} V &= {}^\perp D'(\mathcal{V}^*) \\ &= \left\{ \mathbf{y} \in \mathbf{H}(L, Q) : (L\mathbf{y}, \mathbf{z})_Q = (\mathbf{y}, L^*\mathbf{z})_Q \text{ for all } \mathbf{z} \in \mathcal{V}^* \right\}. \end{aligned}$$

Proof. We have $\mathcal{V} \subset {}^\perp D'(\mathcal{V}^*)$ by (3.11), so that $V \subset {}^\perp D'(\mathcal{V}^*)$, since ${}^\perp D'(\mathcal{V}^*)$ is closed in $\mathbb{H}(L, Q)$, see DEFINITION 2.6.

Now, for $\mathbf{w} \in {}^\perp D'(\mathcal{V}^*) \subset \mathbb{H}(L, Q)$ set $\mathbf{b} = L\mathbf{w} \in W$ and let $\mathbf{y} \in V$ be the unique solution of $L\mathbf{y} = \mathbf{b}$, cf. THEOREM 3.7, yielding $L(\mathbf{y} - \mathbf{w}) = \mathbf{0}$. Since $\mathbf{y} \in V \subset {}^\perp D'(\mathcal{V}^*)$, we have $\mathbf{y} - \mathbf{w} \in {}^\perp D'(\mathcal{V}^*)$, and we obtain for all $\mathbf{z} \in \mathcal{V}^*$

$$0 = \langle D'\mathbf{z}, \mathbf{y} - \mathbf{w} \rangle = (L(\mathbf{y} - \mathbf{w}), \mathbf{z})_Q - (\mathbf{y} - \mathbf{w}, L^*\mathbf{z})_Q = -(\mathbf{y} - \mathbf{w}, L^*\mathbf{z})_Q.$$

Since $L^*(\mathcal{V}^*) \subset W$ is dense, we obtain $\mathbf{w} = \mathbf{y} \in V$. \square

THEOREM 3.9 shows that the operator L with domain V is the HILBERT adjoint of the operator L^* with domain V^* in the sense of [56, Def. 8.58].

Corollary 3.10. $L: \mathbb{H}(L, Q) \supset V \rightarrow W$ and $L^*: \mathbb{H}(L^*, Q) \supset V^* \rightarrow W$ are densely defined closed surjective operators and we have

$$(L\mathbf{y}, \mathbf{z})_Q = (\mathbf{y}, L^*\mathbf{z})_Q, \quad \mathbf{y} \in V, \mathbf{z} \in V^*.$$

In the following section, we break the space $\mathbb{H}(L, Q)$ by considering functions that are piece-wise defined on subdomains of Q .

3.3 Space-time substructuring

For a decomposition $Q_h = \bigcup_{R \in \mathcal{R}_h} R$ into open disjoint space-time cells R , we consider the corresponding discontinuous space $\mathbb{H}(L, Q_h) = \prod_{R \in \mathcal{R}_h} \mathbb{H}(L, R)$.

Remark 3.11. For vector spaces X_1, \dots, X_N , we denote by $\prod_{n=1}^N X_n$ the CARTESIAN product of X_1, \dots, X_N . In the special case $X_R = L_2(R, \mathbb{R}^m)$, $R \in \mathcal{R}_h$, we identify $W = L_2(Q, \mathbb{R}^m)$ and $\prod_{R \in \mathcal{R}_h} L_2(R, \mathbb{R}^m)$.

Introducing local operators $D_R \in \mathcal{L}(\mathbb{H}(L, R), \mathbb{H}(L^*, R)')$, defined by

$$\langle D_R \mathbf{y}_R, \mathbf{z}_R \rangle = (L\mathbf{y}_R, \mathbf{z}_R)_R - (\mathbf{y}_R, L^*\mathbf{z}_R)_R, \quad \mathbf{y}_R \in \mathbb{H}(L, R), \mathbf{z}_R \in \mathbb{H}(L^*, R),$$

we extend the operator D to $D_h \in \mathcal{L}(\mathbb{H}(L, Q_h), \mathbb{H}(L^*, Q_h)')$ by

$$\langle D_h \mathbf{y}, \mathbf{z} \rangle = \sum_{R \in \mathcal{R}_h} \langle D_R \mathbf{y}_R, \mathbf{z}_R \rangle, \quad \mathbf{y} \in \mathbb{H}(L, Q_h), \mathbf{z} \in \mathbb{H}(L^*, Q_h),$$

with $\mathbf{y}_R = \mathbf{y}|_R$ and $\mathbf{z}_R = \mathbf{z}|_R$. In particular, we obtain

$$\langle D\mathbf{y}, \mathbf{z} \rangle = \sum_{R \in \mathcal{R}_h} ((L\mathbf{y})|_R, \mathbf{z}|_R)_R - (\mathbf{y}|_R, (L^*\mathbf{z})|_R)_R = \langle D_h \mathbf{y}, \mathbf{z} \rangle \quad (3.13)$$

for conforming functions $\mathbf{y} \in \mathbb{H}(L, Q)$ and $\mathbf{z} \in \mathbb{H}(L^*, Q)$.

Analogously, we define $D'_h \in \mathcal{L}(\mathbb{H}(L^*, Q_h), \mathbb{H}(L, Q_h)')$.

Remark 3.12. Note that we abuse notation, writing L for the operator defined in all of Q as well as for the operator that is subdomain-wise defined.

Lemma 3.13. We have

$$\begin{aligned} V &= {}^\perp D'_h(V^*) \\ &= \{\mathbf{y} \in \mathbf{H}(L, Q_h) : \langle D_h \mathbf{y}, \mathbf{z} \rangle = 0 \text{ for all } \mathbf{z} \in V^*\}. \end{aligned}$$

Proof. It is sufficient to show ${}^\perp D'_h(V^*) \subset \mathbf{H}(L, Q)$. Then, (3.13) yields the assertion by ${}^\perp D'_h(V^*) \cap \mathbf{H}(L, Q) = {}^\perp D'(V^*) = V$, cf. THEOREM 3.9.

For $\mathbf{y} \in {}^\perp D'_h(V^*) \subset \mathbf{H}(L, Q_h)$ and $\mathbf{b} = L\mathbf{y} \in W$, we have $\langle D_h \mathbf{y}, \mathbf{z} \rangle = 0$ for $\mathbf{z} \in C_c^1(Q, \mathbb{R}^m) \subset V^*$. Thus, we obtain

$$(\mathbf{b}, \mathbf{z})_Q = (L\mathbf{y}, \mathbf{z})_{Q_h} = (\mathbf{y}, L^* \mathbf{z})_{Q_h} = (\mathbf{y}, L^* \mathbf{z})_Q, \quad \mathbf{z} \in C_c^1(Q, \mathbb{R}^m),$$

so that indeed $\mathbf{y} \in \mathbf{H}(L, Q)$ by definition (3.7). \square

Lemma 3.14. We have

$$\mathbf{H}_0(L, Q_h) = \mathcal{N}(D_h).$$

Proof. We have $\mathbf{H}_0(L, R) = \mathcal{N}(D_R)$, cf. THEOREM 3.5. Thus, the assertion follows from

$$\mathbf{H}_0(L, Q_h) = \prod_{R \in \mathcal{R}_h} \mathbf{H}_0(L, R) = \prod_{R \in \mathcal{R}_h} \mathcal{N}(D_R) = \mathcal{N}(D_h). \quad \square$$

This shows that the operator D_h is well-defined on the quotient space associated with the quotient norm that is denoted by

$$\hat{\mathbf{H}}(L, Q_h) = \mathbf{H}(L, Q_h) / \mathbf{H}_0(L, Q_h), \quad \|\hat{\mathbf{y}}\|_{L; \partial Q_h} = \inf_{\hat{\mathbf{y}} = \mathbf{y} + \mathbf{H}_0(L, Q_h)} \|\mathbf{y}\|_{L, Q_h},$$

i.e., $\hat{D}_h \in \mathcal{L}(\hat{\mathbf{H}}(L, Q_h), \mathbf{H}(L^*, Q_h)')$ is well-defined with

$$\langle \hat{D}_h \hat{\mathbf{y}}, \mathbf{z} \rangle = \langle D_h \mathbf{y}, \mathbf{z} \rangle, \quad \hat{\mathbf{y}} = \mathbf{y} + \mathbf{H}_0(L, Q_h). \quad (3.14)$$

By construction, \hat{D}_h is injective, i.e., $\mathcal{N}(\hat{D}_h) = \{\mathbf{0}\}$.

Remark 3.15. Note that the roles of L and L^* in the above construction are symmetric to each other. Thus, following the proofs of the assertions above with L and L^* interchanged yields $V^* = {}^\perp D'_h(V)$ and $\mathbf{H}_0(L^*, Q_h) = \mathcal{N}(D'_h)$.

Chapter 4

Space-time minimal residual methods for waves

In this chapter, we consider different variants of Least-Squares Finite Element methods. Before we deal with the weakly conforming variant and the Discontinuous Petrov-Galerkin method, we briefly discuss the classical conforming situation in SECTION 4.1. The newly constructed methods presented in SECTION 4.2 and SECTION 4.3 generalize the classical situation in different ways.

Space-time discretizations yield promising schemes for exascale parallel computers since they allow by construction for parallelism in space and in time as soon as a well-scaling preconditioner for the full space-time system is available. For instance, a competitive space-time discretization for the heat-equation with outstanding parallel scaling properties has been constructed for instance in [51]. Moreover, since evolution problems become stationary when treated in space-time, adaptivity by locally refining the space-time mesh or by locally increasing polynomial degrees is easily accessible. See [30] for an example for space-time adaptivity applied to electromagnetic waves where the author considers a discontinuous GALERKIN method for hyperbolic evolution equations featuring a parallel multilevel preconditioner for the full space-time problem.

4.1 Least-Squares Finite Elements

This section summarizes well-known textbook contents that we provide for the convenience of the reader, see e.g. [6] for a self-contained reference. See also [1, 5, 13].

Before considering Least-Squares Finite Elements for acoustic waves, we start with a general operator equation $Lu = b$. For a real HILBERT space W and a subspace

$V \subset W$, we consider a bijective operator $L: V \rightarrow W$ and assume that $C_L > 0$ exists, such that

$$\|v\|_W \leq C_L \|Lv\|_W, \quad v \in V. \quad (4.1)$$

We select $\|v\|_V = \sqrt{\|v\|_W^2 + \|Lv\|_W^2}$, $v \in V$, and assume that V is closed in W with respect to $\|\cdot\|_V$. For acoustic waves, a setting like this is introduced in CHAPTER 3.

Since the operator L is a bijection, for any given $b \in W$ a unique $u \in V$ exists fulfilling $Lu = b$. Since $Lu - b = 0$, u is also characterized by the Least-Squares problem

$$u = \operatorname{argmin}_{v \in V} \frac{1}{2} \|Lv - b\|_W^2. \quad (4.2)$$

The functional $\tilde{J}: V \rightarrow \mathbb{R}$ with $v \mapsto \frac{1}{2} \|Lv - b\|_W^2$ can be rewritten as follows

$$\tilde{J}(v) = \frac{1}{2} (Lv - b, Lv - b)_W = \frac{1}{2} (Lv, Lv)_W - (b, Lv)_W + \frac{1}{2} \|b\|_W^2.$$

Minimizing \tilde{J} is equivalent to minimizing $J: V \rightarrow \mathbb{R}$ instead with

$$J(v) = \frac{1}{2} a(v, v) - \ell(v),$$

where $a: V \times V \rightarrow \mathbb{R}$ and $\ell: V \rightarrow \mathbb{R}$ are defined by

$$a(v, \tilde{v}) = (Lv, L\tilde{v})_W, \quad \ell(v) = (b, Lv)_W, \quad v, \tilde{v} \in V.$$

The minimizer $u \in V$ is a critical point of J and fulfills $J'(u) = 0$, i.e.

$$a(u, v) = \ell(v), \quad \text{for all } v \in V. \quad (4.3)$$

Lemma 4.1. *For HILBERT spaces $V \subset W$ let $L: V \rightarrow W$ be a linear operator with $\|v\|_W \leq C_L \|Lv\|_W$ for $v \in V$ and let $b \in W$.*

Then for $\|\cdot\|_V := \sqrt{\|\cdot\|_W^2 + \|L(\cdot)\|_W^2}$ we have

1. *The bilinear form $a \in \mathcal{B}(V \times V, \mathbb{R})$ with $a(v, \tilde{v}) = (Lv, L\tilde{v})_W$, $v, \tilde{v} \in V$, is bounded below by $\alpha = (C_L^2 + 1)^{-1}$, i.e. $a(v, v) \geq (C_L^2 + 1)^{-1} \|v\|_V^2$, $v \in V$, and it holds $\|a\| \leq 1$.*

2. *The linear form $\ell \in V'$ fulfills $\|\ell\|_{V'} \leq \|b\|_W$.*

Proof. For $v, \tilde{v} \in V$, it holds

$$\begin{aligned} |a(v, \tilde{v})| &\leq \|Lv\|_W \|L\tilde{v}\|_W \leq \|v\|_V \|\tilde{v}\|_V, \quad |\ell(v)| \leq \|b\|_W \|Lv\|_W \leq \|b\|_W \|v\|_V \\ \|v\|_V^2 &= \|v\|_W^2 + \|Lv\|_W^2 \leq (C_L^2 + 1) \|Lv\|_W^2 = (C_L^2 + 1) a(v, v). \quad \square \end{aligned}$$

Remark 4.2. In [68], estimates of the form $\|v\|_W \leq C_L \|Lv\|_W$ are established for various first-order systems, e.g. for diffusion-convection-reaction, variants of MAXWELL's equations, the HELMHOLTZ problem, linear elasticity, and STOKES equation.

As a result, the LAX-MILGRAM lemma (COROLLARY 2.18) yields the existence of a unique $u \in V$ solving (4.3) and fulfilling $\|u\|_V \leq \frac{1}{\alpha} \|\ell\|_{V'}$.

Selecting a conforming approximation space $V_h \subset V$, we obtain the discrete counterpart of (4.2) for the approximation $u_h \in V_h$ as

$$u_h = \operatorname{argmin}_{v_h \in V_h} \frac{1}{2} \|Lv_h - b\|_W^2. \quad (4.4)$$

This also yields the following variant of (4.3)

$$a(u_h, v_h) = \ell(v_h), \quad \text{for all } v_h \in V_h. \quad (4.5)$$

Since $V_h \subset V$, the restrictions $a_h: V_h \times V_h \rightarrow \mathbb{R}$ and $\ell_h: V_h \rightarrow \mathbb{R}$ fulfill the assumptions of the LAX-MILGRAM lemma with the same constants. Therefore, (4.5) uniquely determines the discrete solution $u_h \in V_h$ with $\|u_h\|_V \leq \frac{1}{\alpha} \|\ell_h\|_{V_h'} \leq \frac{1}{\alpha} \|\ell\|_{V'}$.

Now, (4.3) and (4.5) yield GALERKIN orthogonality

$$a(u - u_h, v_h) = 0, \quad v_h \in V_h,$$

and we obtain quasi best-approximation, since by

$$\alpha \|u - u_h\|_V^2 \leq a(u - u_h, u - u_h) = a(u - u_h, u - v_h) \leq \|u - u_h\|_V \|u - v_h\|_V$$

we conclude

$$\|u - u_h\|_V \leq \frac{1}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_V. \quad (4.6)$$

To solve (4.5), we select an ordered basis (v_h^1, \dots, v_h^n) of V_h , $n = \dim V_h$, and define $\underline{A} \in \mathbb{R}^{n \times n}$, $\underline{b} \in \mathbb{R}^n$ by

$$\underline{A}_{kl} = a(v_h^k, v_h^l), \quad \underline{b}_k = \ell(v_h^k), \quad k, l = 1, \dots, n.$$

Then finding $\underline{u} \in \mathbb{R}^n$ with $\underline{A}\underline{u} = \underline{b}$ yields $u_h = \sum_{k=1}^n \underline{u}_k v_h^k$ that is a solution of (4.5). Since $\underline{A}_{kl} = a(v_h^k, v_h^l) = a(v_h^l, v_h^k) = \underline{A}_{lk}$ and for $v_h = \sum_{k=1}^n \underline{v}_k v_h^k$, $\underline{v} \in \mathbb{R}^n$, it holds

$$\begin{aligned} \underline{v}^\top \underline{A} \underline{v} &= \sum_{k,l=1}^n \underline{v}_k \underline{A}_{kl} \underline{v}_l = \sum_{k,l=1}^n \underline{v}_k a(v_h^k, v_h^l) \underline{v}_l = a \left(\sum_{k=1}^n \underline{v}_k v_h^k, \sum_{l=1}^n \underline{v}_l v_h^l \right) \\ &= a(v_h, v_h) \geq \alpha \|v_h\|_V^2, \end{aligned}$$

the system matrix \underline{A} is symmetric and positive definite.

Discussion This short summary gives an insight, why Least-Squares Finite Elements are so powerful: given a very general setting as described in the beginning by (4.6), Least-Squares Finite Elements yield a convergent scheme for all sequences of approximation spaces $(V_h)_{h>0}$ such that $\bigcup_{h>0} V_h$ is dense in V with respect to $\|\cdot\|_V$. In addition, if $\dim V_h < \infty$ the finite dimensional linear system (4.5) yields a symmetric and positive definite system Matrix \underline{A} independently of the properties of the differential operator L . In particular, for hyperbolic problems like the acoustic wave equation, the Least-Squares system matrix is symmetric and positive definite.

However, since (4.3) is a normal equation, the condition number for (4.5) is squared compared to other approaches. Another disadvantage is that the discrete solution converges with respect to the norm $\|\cdot\|_V$ that may not be the norm we are interested in. In SECTION 4.3, we discuss a variant of the Least-Squares method for dual norms with in some sense optimal stability properties.

The discretization schemes presented in the following posses various appealing features. The weakly conforming Least-Squares variant discussed in the following chapter minimizes the residual in a larger discrete space compared to the conforming scheme described here. Since the approximation space contains discontinuous functions, we expect improved properties for solutions with low regularity.

All methods presented in the following use spaces that are coupled along the space-time cells only. Since we do not have nodal degrees of freedom and thus, no nodal coupling, the resulting system matrices have an appealing sparsity structure.

Furthermore, the methods allow for eliminating the interior degrees of freedom inside the cells yielding a system matrix of reduced size. The additional effort for this elimination procedure can be performed on each space-time cell separately. This is an appealing property for parallel implementations of the methods.

The Discontinuous PETROV-GALERKIN method has been proven to allow for robust schemes with respect to singular perturbations, see [22]. In the long run, we want to transfer this property to obtain robust approximations for the wave equation in case of jumping material coefficients.

In this work we focus on two variants of a larger family of space-time methods. It is beyond the scope of this thesis to provide an exhausting evaluation.

Remark 4.3 (Space-Time Least-Squares in scaled L_2 norms). *A simple generalization of the standard Least-Squares approach uses L_2 norms scaled by a weight. Let $W = L_2(Q, \mathbb{R}^m)$ and $\omega: Q \rightarrow \mathbb{R}$ a bounded function that fulfills $\omega > 0$ almost everywhere in $Q = \Omega \times (0, T)$. Then, we have for $v \in V$*

$$Lv = b \iff \omega \cdot (Lv - b) = 0 \iff v = \operatorname{argmin}_{\tilde{v} \in V} \frac{1}{2} \|\omega \cdot (L\tilde{v} - b)\|_{L_2(Q)}^2.$$

Choosing a rapidly decaying function ω in time like $\omega(t) = \alpha e^{-\alpha t/T}$, $\alpha > 0$, yields solutions that were less diffusive at material jumps compared to solutions of the unscaled variant. Furthermore, the iterative solver needs significantly less steps for solving the scaled linear system.

Finding a suitable scaling of the local norm for problems with large spatial variation of the material parameters is a promising future challenge.

4.2 Weakly conforming Least-Squares Finite Elements

To improve the conforming Least-Squares method presented in SECTION 4.1, we investigate a different choice of the approximation space V_h .

Here, we use the variational framework and the notation introduced in CHAPTER 3.

4.2.1 A weakly conforming approximation space

In SECTION 3.3, we introduced the operator $D_h \in \mathcal{L}(\mathbf{H}(L, Q_h), \mathbf{H}(L^*, Q_h)')$ using abstract integration by parts on a mesh Q_h of Q

$$\langle D_h \mathbf{y}, \mathbf{z} \rangle = \sum_{R \in \mathcal{R}_h} \langle D_R \mathbf{y}|_R, \mathbf{z}|_R \rangle, \quad \langle D_R \mathbf{y}_R, \mathbf{z}_R \rangle = (L \mathbf{y}_R, \mathbf{z}_R)_R - (\mathbf{y}_R, L^* \mathbf{z}_R)_R.$$

Using this operator, LEMMA 3.13 provides a characterization of what is needed for a cell-wise defined function to be conforming in V . More precisely, for a cell-wise defined $\mathbf{y} \in \mathbf{H}(L, Q_h)$ we have

$$\mathbf{y} \in V \iff \langle D_h \mathbf{y}, \mathbf{z} \rangle = 0 \quad \forall \mathbf{z} \in V^*. \quad (4.7)$$

In this section, we weaken (4.7) to obtain a larger approximation space V_h .

To illustrate the following construction, we consider an analogous situation for a stationary problem.

Example 4.4. We interpret (4.7) in case of the classical situation $L = \nabla$ and $L^* = -\operatorname{div}$. Here, the boundary of a connected LIPSCHITZ domain $\Omega \subset \mathbb{R}^d$ is partitioned into $\Gamma_D \cup \Gamma_N$ where Γ_D has non-zero $(d-1)$ -dimensional measure. Then, we have $\mathbf{H}(L, \Omega) = \mathbf{H}^1(\Omega)$ and choose $V = \mathbf{H}_0^1(\Omega)$, $V^* = \mathbf{H}_0(\operatorname{div}, \Omega)$ with

$$\begin{aligned} \mathbf{H}_0^1(\Omega) &= \{v \in \mathbf{H}^1(\Omega) : v|_{\Gamma_D} = 0\}, \\ \mathbf{H}_0(\operatorname{div}, \Omega) &= \{w \in \mathbf{H}(\operatorname{div}, \Omega) : (w \cdot \mathbf{n}_\Omega)|_{\Gamma_N} = 0\}. \end{aligned}$$

All restrictions to the boundary are understood in the sense of trace operators. In every cell K , integration by parts yields for $v \in \mathbf{H}^1(K)$, $w \in \mathbf{H}(\operatorname{div}, K)$

$$\langle D_K v, w \rangle := \int_K \nabla v \cdot w \, dx + \int_K v \operatorname{div} w \, dx = \int_{\partial K} v w \cdot n_K \, da.$$

For an inner face $f = \overline{K} \cap \overline{K'}$, we have $n_K|_f = -n_{K'}|_f$ and thus

$$\begin{aligned} \langle D_K v, w \rangle + \langle D_{K'} v, w \rangle &= \int_f (v_K - v_{K'}) w \cdot n_K \, da \\ &\quad + \int_{\partial K \setminus f} v w \cdot n_K \, da + \int_{\partial K' \setminus f} v w \cdot n_{K'} \, da. \end{aligned}$$

This implies for all $w \in \mathbf{H}(\operatorname{div}, \Omega)$ having vanishing normal traces on $\partial(K \cup K') \setminus f$

$$\langle D_K v, w \rangle + \langle D_{K'} v, w \rangle = \int_f (v_K - v_{K'}) w \cdot n_K \, da.$$

Therefore, the characterization (4.7) means that a cell-wise defined function $v \in \mathbf{H}^1(\Omega_h)$ is conforming in $\mathbf{H}^1(\Omega)$ if and only if the DIRICHLET traces coincide on every face between two cells when tested with normal traces of functions in $\mathbf{H}(\operatorname{div}, \Omega)$.

Following the spirit of EXAMPLE 4.4, we obtain for the acoustic wave operator L and cell-wise smooth functions $\mathbf{y} = (p, \mathbf{v}) \in V$, $\mathbf{z} = (\psi, \boldsymbol{\phi}) \in V^*$ on every cell $R = K \times (t_-, t^+)$ by GAUSS' divergence theorem

$$\begin{aligned} \langle D_R \mathbf{y}, \mathbf{z} \rangle &= \int_{K \times \{t^+\}} (p \psi + \mathbf{v} \cdot \boldsymbol{\phi}) \, dx - \int_{K \times \{t_-\}} (p \psi + \mathbf{v} \cdot \boldsymbol{\phi}) \, dx \\ &\quad + \int_{\partial K \times (t_-, t^+)} (p \boldsymbol{\phi} \cdot \mathbf{n}_R + \mathbf{v} \cdot \mathbf{n}_R \psi) \, d(a, t) \\ &= ((p, \mathbf{v}), (\psi, \boldsymbol{\phi}))_{K \times \{t^+\}} - ((p, \mathbf{v}), (\psi, \boldsymbol{\phi}))_{K \times \{t_-\}} \\ &\quad + ((p, \mathbf{v} \cdot \mathbf{n}_R), (\boldsymbol{\phi} \cdot \mathbf{n}_R, \psi))_{\partial K \times (t_-, t^+)}. \end{aligned} \tag{4.8}$$

Fixing a space-time face F with adjacent cells R and R' , we restrict the test space to functions vanishing on all other faces. This yields the following compatibility conditions, where we have to distinguish faces in time from faces in space.

- For a face in time having the form $F = K \times \{t\}$ continuity on p and of all components of \mathbf{v} is required.
- For a face $F = f \times (t_-, t^+)$ in space ($f \subset \partial K$ in space) the p component and the normal part of the \mathbf{v} component need to be continuous.

Using the variational characterization in (4.7), we can relax these compatibility conditions by testing with a smaller space than V^* . This yields our approximation space

V_h . Given locally conforming spaces V_R , i.e. $V_R \subset \mathbf{H}(L, R)$, and a globally conforming coupling space $V_h^* \subset V^*$, we set $V_{\mathcal{R}_h} = \prod_R V_R$, $V_{\mathcal{R}_h}^* = \prod_R (V_h^*)|_R$ and define the space of weakly conforming functions with respect to V_h^* by

$$V_h^{\text{wc}} = V_h^{\text{wc}}(V_h^*) = \left\{ \mathbf{y}_h \in V_{\mathcal{R}_h} : \langle D_h \mathbf{y}_h, \mathbf{z}_h \rangle = 0 \text{ for all } \mathbf{z}_h \in V_h^* \right\}. \quad (4.9)$$

By construction, we have $V \cap V_{\mathcal{R}_h} \subset V_h^{\text{wc}}$ but $V_h^{\text{wc}} \not\subset V$ in general, i.e. the approximation space V_h^{wc} is *larger* than a conforming space that is locally given by V_R . Thus, V_h^{wc} is *non-conforming* in V . Since we weakened the conformity condition (4.7), we say that V_h^{wc} is *weakly conforming*.

Example 4.5. *To illustrate the definition of V_h^{wc} , we continue the considerations from EXAMPLE 4.4. For a mesh \mathcal{K}_h of a domain $\Omega \subset \mathbb{R}^2$ consisting of rectangular cells K , we take locally bilinear functions $V_K := \mathbb{Q}_1(K) \subset \mathbf{H}^1(K)$ and the $\mathbf{H}(\text{div}, \Omega)$ conforming RAVIART-THOMAS space $V_h^* := \text{RT}_k(\Omega)$ as a test space, see e.g. [24, Sec. 1.26].*

This yields functions in $V_h^{\text{wc}}(V_h^)$ having DIRICHLET traces that coincide from both sides when tested with polynomials up to k -th order. In particular, if $k \geq l$ we end up with a conforming space $V_h^{\text{wc}} \subset V$. The special case $l = 1, k = 0$ yields the CROUZEIX-RAVIART approximation space (e.g. [7, Sec. 2.4.1]), see FIGURE 4.1. Note that not only the coupling along faces, but also the boundary conditions are enforced in a weak sense for functions in V_h^{wc} .*

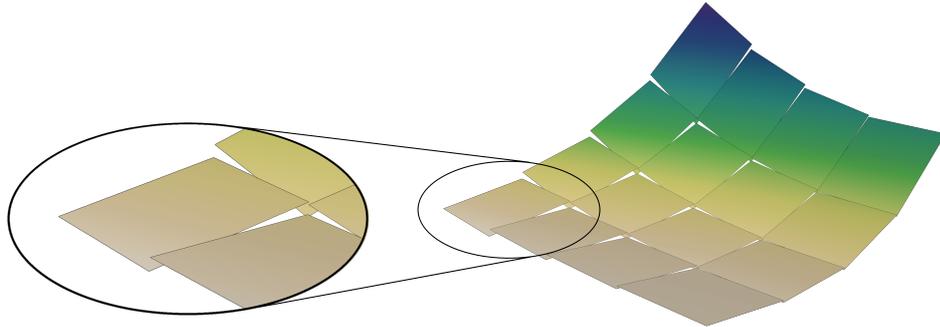


Figure 4.1: Plot of a weakly conforming function in $V_{\mathcal{R}_h} = \prod_R \mathbb{Q}_1(R)$ that is weakly coupled by $V_h^* = \text{RT}_0(Q)$ along the faces of rectangular cells. On the left, you can see that the functions coincide at the face midpoints, since their mean values are the same from both neighboring cells.

In the following, we consider the Least-Squares minimization problem (4.4) for $V_h = V_h^{\text{wc}}$ where we apply the operator L locally in every space-time cell $R \in \mathcal{R}_h$.

4.2.2 Saddle point reformulation

For the numerical analysis and also for the implementation, weakly conforming Least-Squares FEM feature additional challenges in contrast to the conforming case.

Before we consider the approximation of \mathbf{y} by a function in V_h^{wc} , we transform the analytic minimization problem (4.2) into cell-wise problems that are coupled along the faces. Since the approximation space is non-conforming, i.e. $V_h^{\text{wc}} \not\subset V$, we need to extend the bilinear form a and the right-hand side ℓ to V_h^{wc} .

To this end, we define for broken functions $\mathbf{y}, \mathbf{z} \in \mathbf{H}(L, Q_h)$ and $R \in \mathcal{R}_h$

$$\begin{aligned} a_h(\mathbf{y}, \mathbf{z}) &:= \sum_{R \in \mathcal{R}_h} a_R(\mathbf{y}|_R, \mathbf{z}|_R), & a_R(\mathbf{y}|_R, \mathbf{z}|_R) &:= (L_R \mathbf{y}|_R, L_R \mathbf{z}|_R)_R, \\ \ell_h(\mathbf{y}) &:= \sum_{R \in \mathcal{R}_h} \ell_R(\mathbf{y}|_R), & \ell_R(\mathbf{y}|_R) &:= (\mathbf{b}, L_R \mathbf{y}|_R)_R, \\ J_h(\mathbf{y}) &:= \sum_{R \in \mathcal{R}_h} J_R(\mathbf{y}|_R), & J_R(\mathbf{y}|_R) &:= \frac{1}{2} a_R(\mathbf{y}|_R, \mathbf{y}|_R) - \ell_R(\mathbf{y}|_R). \end{aligned}$$

Now, (4.2) can be reformulated as the constrained minimization problem

$$\min J_h(\mathbf{y}) \quad \text{subject to} \quad \mathbf{y} \in \mathbf{H}(L, Q_h): \langle D_h \mathbf{y}, \mathbf{z} \rangle = 0 \quad \forall \mathbf{z} \in V^*. \quad (4.10)$$

Lemma 4.6. *The problems (4.2) and (4.10) are equivalent:*

1. *If $\mathbf{y} \in V$ solves (4.2) then \mathbf{y} also solves (4.10).*
2. *If $\mathbf{y} \in \mathbf{H}(L, Q_h)$ solves (4.10), then $\mathbf{y} \in V$ and \mathbf{y} solves (4.2).*

Proof. We have $\mathcal{M} = \{\mathbf{y} \in \mathbf{H}(L, Q_h): \langle D_h \mathbf{y}, \tilde{\mathbf{y}} \rangle = 0, \tilde{\mathbf{y}} \in V^*\} = V$ for the admissible set \mathcal{M} of (4.10) by LEMMA 3.13. Now, $J_h|_V = J$ yields the assertion. \square

Problem (4.10) can be analyzed as a saddle point problem in $\mathbf{H}(L, Q_h) \times V^*$ as shown in LEMMA 4.7.

Lemma 4.7. *Let $(\mathbf{y}^{\text{sol}}, \mathbf{z}^{\text{sol}}) \in \mathbf{H}(L, Q_h) \times V^*$ be a saddle point of*

$$F_h(\mathbf{y}, \mathbf{z}) = J_h(\mathbf{y}) + \langle D_h \mathbf{y}, \mathbf{z} \rangle, \quad (\mathbf{y}, \mathbf{z}) \in \mathbf{H}(L, Q_h) \times V^*.$$

Then, we have $\mathbf{y}^{\text{sol}} \in V$ and $L\mathbf{y}^{\text{sol}} = \mathbf{b}$.

Proof. A saddle point $(\mathbf{y}^{\text{sol}}, \mathbf{z}^{\text{sol}}) \in \mathbf{H}(L, Q_h) \times V^*$ of F_h fulfills

$$F_h(\mathbf{y}^{\text{sol}}, \tilde{\mathbf{z}}) \leq F_h(\mathbf{y}^{\text{sol}}, \mathbf{z}^{\text{sol}}) \leq F_h(\tilde{\mathbf{y}}, \mathbf{z}^{\text{sol}}) \quad \text{for all } (\tilde{\mathbf{y}}, \tilde{\mathbf{z}}) \in \mathbf{H}(L, Q_h) \times V^*.$$

This yields $0 \geq F_h(\mathbf{y}^{\text{sol}}, \tilde{\mathbf{z}}) - F_h(\mathbf{y}^{\text{sol}}, \mathbf{z}^{\text{sol}}) = \langle D_h \mathbf{y}^{\text{sol}}, \tilde{\mathbf{z}} - \mathbf{z}^{\text{sol}} \rangle$ for all $\tilde{\mathbf{z}} \in V^*$. Since V^* is a vector space, LEMMA 3.13 implies $\mathbf{y}^{\text{sol}} \in V$.

On the other hand by $\mathbf{y} \in V$ and by LEMMA 3.13, we obtain for $\tilde{\mathbf{y}} \in V \subset \mathbf{H}(L, Q_h)$ the estimate $0 \geq F_h(\mathbf{y}^{\text{sol}}, \mathbf{z}^{\text{sol}}) - F_h(\tilde{\mathbf{y}}, \mathbf{z}^{\text{sol}}) = J(\mathbf{y}^{\text{sol}}) - J(\tilde{\mathbf{y}})$. So, $\mathbf{y}^{\text{sol}} \in V$ is the minimizer of $J = J_h|_V$. \square

Since a saddle-point $(\mathbf{y}^{\text{sol}}, \mathbf{z}^{\text{sol}}) \in \mathbf{H}(L, Q_h) \times V^*$ of F_h is also a critical point, we obtain the following linear system, cf. PROPOSITION 2.36

$$\begin{aligned} \langle A_h \mathbf{y}^{\text{sol}}, \tilde{\mathbf{y}} \rangle + \langle D'_h \mathbf{z}^{\text{sol}}, \tilde{\mathbf{y}} \rangle &= \ell_V(\tilde{\mathbf{y}}), & \text{for all } \tilde{\mathbf{y}} \in \mathbf{H}(L, Q_h), \\ \langle D_h \mathbf{y}^{\text{sol}}, \tilde{\mathbf{z}} \rangle &= 0, & \text{for all } \tilde{\mathbf{z}} \in V^*. \end{aligned} \quad (4.11)$$

We define the operators $A_R \in \mathcal{L}(\mathbf{H}(L, R), \mathbf{H}(L, R)')$ by $\langle A_R \mathbf{y}_R, \tilde{\mathbf{y}}_R \rangle = a_R(\mathbf{y}_R, \tilde{\mathbf{y}}_R)$ and $A_h \in \mathcal{L}(\mathbf{H}(L, Q_h), \mathbf{H}(L, Q_h)')$ by $\langle A_h \mathbf{y}, \tilde{\mathbf{y}} \rangle = \sum_{R \in \mathcal{R}_h} \langle A_R \mathbf{y}_R, \tilde{\mathbf{y}}_R \rangle$ for functions $\mathbf{y} = (\mathbf{y}_R)_R$, $\tilde{\mathbf{y}} = (\tilde{\mathbf{y}}_R)_R \in \mathbf{H}(L, Q_h)$.

Remark 4.8. *Note that the saddle point is not unique, since for $\mathbf{y} \in V$ and $\mathbf{z} \in V_h^*$, we have $F_h(\mathbf{y}, \mathbf{z}) = F_h(\mathbf{y}, \mathbf{z} + \mathbf{z}_0)$ for all $\mathbf{z}_0 \in \mathbf{H}_0(L^*, Q_h)$, in the kernel space*

$$\mathbf{H}_0(L^*, Q_h) = \left\{ \mathbf{z} \in \mathbf{H}(L^*, Q_h) : \langle D_h \tilde{\mathbf{y}}, \mathbf{z} \rangle = 0 \text{ for all } \tilde{\mathbf{y}} \in \mathbf{H}(L, Q_h) \right\}.$$

Due to LEMMA 3.13 and REMARK 3.15, we see $\mathbf{H}_0(L^*, Q_h) \subset V^*$.

Following the arguments in [14, Lem. 2.2], we can show that the saddle point problem is inf-sup stable in the quotient space

$$\hat{V}^* = V^* / \mathbf{H}_0(L^*, Q_h) \subset \hat{\mathbf{H}}(L^*, Q_h),$$

where $\hat{\mathbf{H}}(L^*, Q_h) = \mathbf{H}(L^*, Q_h) / \mathbf{H}_0(L^*, Q_h)$, see SECTION 3.3, normed by

$$\|\hat{\mathbf{z}}\|_{L; \partial Q_h} = \inf_{\hat{\mathbf{z}} = \mathbf{z} + \mathbf{H}_0(L^*, Q_h)} \|\mathbf{z}\|_{L^*, Q_h}, \quad \hat{\mathbf{z}} \in \hat{\mathbf{H}}(L^*, Q_h).$$

By REMARK 3.15, we have $\mathcal{N}(D'_h) = \mathbf{H}_0(L^*, Q_h)$ and furthermore we see that the operator $\hat{D}'_h \in \mathcal{L}(\hat{\mathbf{H}}(L^*, Q_h), \mathbf{H}(L, Q_h)')$ is well-defined with

$$\langle \hat{D}'_h \hat{\mathbf{z}}, \mathbf{y} \rangle = \langle D'_h \mathbf{z}, \mathbf{y} \rangle, \quad \hat{\mathbf{z}} = \mathbf{z} + \mathbf{H}_0(L^*, Q_h).$$

By construction, \hat{D}'_h is injective, i.e., $\mathcal{N}(\hat{D}'_h) = \{0\}$. We show that \hat{D}'_h is indeed bounded below with stability constant 1 using similar arguments as in THEOREM 3.5.

Lemma 4.9. *We have for $\mathbf{z} \in \mathbf{H}(L^*, Q_h)$*

$$\inf_{\mathbf{z}_0 \in \mathbf{H}_0(L^*, Q_h)} \|\mathbf{z} + \mathbf{z}_0\|_{L^*, Q_h} = \sup_{\mathbf{y} \in \mathbf{H}(L, Q_h)} \frac{\langle D'_h \mathbf{z}, \mathbf{y} \rangle}{\|\mathbf{y}\|_{L, Q_h}}.$$

Proof. For given $\mathbf{z} \in \mathbf{H}(L^*, Q_h)$ define $\mathbf{y}^* \in \mathbf{H}(L, Q_h)$ solving

$$(\mathbf{y}^*, \phi)_{L, Q_h} = \langle D_h \phi, \mathbf{z} \rangle \quad \text{for all } \phi \in \mathbf{H}(L, Q_h). \quad (4.12)$$

Then we set $\mathbf{z}^* = L\mathbf{y}^*$, and by $\langle D_h \phi, \mathbf{z} \rangle = 0$, $\phi \in C_c^1(Q_h, \mathbb{R}^m)$, (4.12) yields

$$0 = (\mathbf{y}^*, \phi)_{L, Q_h} = (\mathbf{y}^*, \phi)_{Q_h} + (\mathbf{z}^*, L\phi)_{Q_h}, \quad \phi \in C_c^1(Q_h, \mathbb{R}^m).$$

Thus, $\mathbf{z}^* \in \mathbf{H}(L^*, Q_h)$ with $L^*\mathbf{z}^* = -\mathbf{y}^*$ and $\|\mathbf{z}^*\|_{L^*, Q_h} = \|\mathbf{y}^*\|_{L, Q_h}$. By (4.12),

$$\begin{aligned} \langle D_h \phi, \mathbf{z} - \mathbf{z}^* \rangle &= \langle D_h \phi, \mathbf{z} \rangle - (L\phi, \mathbf{z}^*)_{Q_h} + (\phi, L^*\mathbf{z}^*)_{Q_h} \\ &= (\mathbf{y}^*, \phi)_{L, Q_h} - (L\phi, L\mathbf{y}^*)_{Q_h} - (\phi, \mathbf{y}^*)_{Q_h} = 0 \end{aligned}$$

for $\phi \in \mathbf{H}(L, Q_h)$, i.e., $\mathbf{z} - \mathbf{z}^* \in \mathbf{H}_0(L^*, Q_h)$. This finally yields

$$\begin{aligned} \inf_{\mathbf{z}_0 \in \mathbf{H}_0(L^*, Q_h)} \|\mathbf{z} + \mathbf{z}_0\|_{L^*, Q_h} &\leq \|\mathbf{z}^*\|_{L^*, Q_h} = \|\mathbf{y}^*\|_{L, Q_h} = \frac{\langle D_h \mathbf{y}^*, \mathbf{z} \rangle}{\|\mathbf{y}^*\|_{L, Q_h}} \\ &\leq \sup_{\phi \in \mathbf{H}(L, Q_h)} \frac{\langle D_h \phi, \mathbf{z} \rangle}{\|\phi\|_{L, Q_h}} \\ &= \inf_{\mathbf{z}_0 \in \mathbf{H}_0(L^*, Q_h)} \sup_{\phi \in \mathbf{H}(L, Q_h)} \frac{\langle D_h \phi, \mathbf{z} + \mathbf{z}_0 \rangle}{\|\phi\|_{L, Q_h}} \\ &\leq \inf_{\mathbf{z}_0 \in \mathbf{H}_0(L^*, Q_h)} \|\mathbf{z} + \mathbf{z}_0\|_{L^*, Q_h} \end{aligned}$$

where we use $|\langle D_h \phi, \psi \rangle| \leq \|\phi\|_{L, Q_h} \|\psi\|_{L, Q_h}$. \square

As an immediate consequence, we obtain

Corollary 4.10. *For $\hat{\mathbf{z}} \in \hat{\mathbf{H}}(L^*, Q_h)$, $\mathbf{y} \in \mathbf{H}(L, Q_h)$, we have*

$$\|\hat{\mathbf{z}}\|_{L^*; \partial Q_h} = \sup_{\mathbf{y} \in \mathbf{H}(L, Q_h)} \frac{\langle \hat{D}'_h \hat{\mathbf{z}}, \mathbf{y} \rangle}{\|\mathbf{y}\|_{L, Q_h}}.$$

Remark 4.11. COROLLARY 4.10 shows that

$$\hat{D}'_h : \hat{\mathbf{H}}(L^*, Q_h) \longrightarrow \hat{D}'_h(\hat{\mathbf{H}}(L^*, Q_h)) \subset \mathbf{H}(L, Q_h)'$$

is an isometry identifying the trace space $\hat{\mathbf{H}}(L^*, Q_h)$ with a subspace of $\mathbf{H}(L, Q_h)'$.

From the proof of LEMMA 4.9, we conclude that \mathbf{z}^* is the function in $\mathbf{z} + \mathbf{H}_0(L^*, Q_h)$ having the minimal $\|\cdot\|_{L^*, Q_h}$ norm. Since $\mathbf{z} + \mathbf{H}_0(L^*, Q_h)$ is the trace of \mathbf{z} , we interpret \mathbf{z}^* as the continuation of traces of \mathbf{z} having minimal norm.

Theorem 4.12. *There is a unique $(\mathbf{y}^{\text{sol}}, \hat{\mathbf{z}}^{\text{sol}}) \in \mathbf{H}(L, Q_h) \times \hat{V}^*$ solving (4.11) and fulfilling*

$$\|\mathbf{y}\|_{L, Q_h} \leq (1 + C_L^2) \|\ell_V\|_{V'}, \quad \|\hat{\mathbf{z}}^{\text{sol}}\|_{L^*; \partial Q_h} \leq (2 + C_L^2) \|\ell_V\|_{V'}.$$

Proof. We apply the theory in SECTION 2.3 to (4.11). To this end, we consider (4.11) on the product space $\mathbf{H}(L, Q_h) \times \hat{V}^*$ setting $\ell_{\hat{V}^*} = 0 \in (\hat{V}^*)'$ and

$$c \in \mathcal{B}(\mathbf{H}(L, Q_h) \times \hat{V}^*, \mathbb{R}), \quad c(\mathbf{y}, \hat{\mathbf{z}}) := \langle \hat{D}'_h \hat{\mathbf{z}}, \mathbf{y} \rangle, \quad \mathbf{y} \in \mathbf{H}(L, Q_h), \quad \hat{\mathbf{z}} \in \hat{V}^*.$$

By COROLLARY 4.10, c is inf-sup stable. We see by LEMMA 3.13 that

$$V = \left\{ \mathbf{y} \in \mathbf{H}(L, Q_h) : c(\mathbf{y}, \hat{\mathbf{z}}) = 0 \text{ for all } \hat{\mathbf{z}} \in \hat{V}^* \right\}$$

and since a is elliptic on V by LEMMA 4.1, the assumptions of THEOREM 2.37 are fulfilled.

This yields the existence of a unique saddle-point $(\mathbf{y}, \hat{\mathbf{z}}) \in \mathbf{H}(L, Q_h) \times \hat{V}^*$. Since $\|a\| \leq 1$, $\|\ell_V\|_{\mathbf{H}(L, Q_h)'} \leq \|\mathbf{b}\|_{Q_h}$ and $\alpha = (C_L^2 + 1)^{-1}$ by LEMMA 4.1, the stability bound follows. \square

4.2.3 Discrete ellipticity and inf-sup stability

In this section, we establish criteria for the stable approximation of the saddle point problem (4.11) using the weakly conforming approximation space V_h^{wc} from (4.9).

Using the extended bilinear form a_h and right-hand side ℓ_h from SECTION 4.2.2, we obtain the following variational problem for $\mathbf{y}_h^{\text{sol}} \in V_h^{\text{wc}}$:

$$a_h(\mathbf{y}_h^{\text{sol}}, \tilde{\mathbf{y}}_h) = \ell_h(\tilde{\mathbf{y}}_h) \quad \text{for all } \tilde{\mathbf{y}}_h \in V_h^{\text{wc}}. \quad (4.13)$$

Given that a_h is elliptic in V_h^{wc} , i.e. there is $\alpha_0 > 0$ such that

$$a_h(\mathbf{y}_h, \mathbf{y}_h) \geq \alpha_0 \|\mathbf{y}_h\|_{L, Q_h}^2, \quad \text{for all } \mathbf{y}_h \in V_h^{\text{wc}}, \quad (4.14)$$

we obtain the discrete counterpart of LEMMA 4.7 as in

Lemma 4.13. *Let $(\mathbf{y}_h^{\text{sol}}, \mathbf{z}_h^{\text{sol}}) \in V_{\mathcal{R}_h} \times V_h^*$ be a saddle point of $F_h|_{V_{\mathcal{R}_h} \times V_h^*}$. Then, we have $\mathbf{y}_h^{\text{sol}} \in V_h^{\text{wc}}$ and $\mathbf{y}_h^{\text{sol}}$ is a minimizer of J_h in V_h^{wc} .*

Proof. Repeat the arguments from LEMMA 4.7. \square

To compute the discrete solution $(\mathbf{y}_h^{\text{sol}}, \mathbf{z}_h^{\text{sol}}) \in V_{\mathcal{R}_h} \times V_h^*$, we consider the discrete counterpart of (4.11) given by

$$\begin{aligned} \langle A_h \mathbf{y}_h^{\text{sol}}, \tilde{\mathbf{y}} \rangle + \langle D'_h \mathbf{z}_h^{\text{sol}}, \tilde{\mathbf{y}} \rangle &= \ell_V(\tilde{\mathbf{y}}), & \text{for all } \tilde{\mathbf{y}} \in V_{\mathcal{R}_h}, \\ \langle D_h \mathbf{y}_h^{\text{sol}}, \tilde{\mathbf{z}} \rangle &= 0, & \text{for all } \tilde{\mathbf{z}} \in V_h^*. \end{aligned} \quad (4.15)$$

Again, the LAGRANGE multiplier $\mathbf{z}_h^{\text{sol}} \in V_h^*$ is not uniquely determined in general. Similarly to the continuous case, we have $F_h(\mathbf{y}_h, \mathbf{z}_h) = F_h(\mathbf{y}_h, \mathbf{z}_h + \mathbf{z}_{0,h})$ for all $\mathbf{z}_{0,h} \in V_{0,h}^*$, where $V_{0,h}^*$ is the discrete kernel space

$$V_{0,h}^* := \left\{ \mathbf{z}_h \in V_h^* : \langle D_h \tilde{\mathbf{y}}_h, \mathbf{z}_h \rangle = 0 \text{ for all } \tilde{\mathbf{y}}_h \in V_{\mathcal{R}_h} \right\}. \quad (4.16)$$

To ensure $V_{0,h}^* = \{0\}$, we aim to show discrete inf-sup stability of the constraint locally, i.e. there exists $\beta_0 > 0$ with

$$\sup_{\mathbf{y}_h \in V_{\mathcal{R}_h}} \frac{\langle D'_h \mathbf{z}_h, \mathbf{y}_h \rangle}{\|\mathbf{y}_h\|_{L, Q_h}} \geq \beta_0 \|\mathbf{z}_h\|_{L^*, Q_h}, \quad \text{for all } \mathbf{z}_h \in V_h^*. \quad (4.17)$$

Theorem 4.14. *Assume that $\alpha_0, \beta_0 > 0$ exist satisfying (4.14) and (4.17). Then, a unique saddle point $(\mathbf{y}_h^{\text{sol}}, \mathbf{z}_h^{\text{sol}}) \in V_{\mathcal{R}_h} \times V_h^*$ of $F_h|_{V_{\mathcal{R}_h} \times V_h^*}$ exists and we have the following error estimate with $C = (1 + \alpha_0^{-1})$*

$$\|\mathbf{y}^{\text{sol}} - \mathbf{y}_h^{\text{sol}}\|_{L, Q_h} \leq C \inf_{\mathbf{y}_h \in V_h^{\text{wc}}} \|\mathbf{y}^{\text{sol}} - \mathbf{y}_h\|_{L, Q_h} + \frac{1}{\alpha_0} \sup_{\mathbf{y}_h \in V_h^{\text{wc}}} \frac{a_h(\mathbf{y}^{\text{sol}}, \mathbf{y}_h) - \ell_h(\mathbf{y}_h)}{\|\mathbf{y}_h\|_{L, Q_h}}.$$

Proof. Since in (4.11) the second right-hand side is $0 \in (V^*)'$, we can apply THEOREM 2.40 and THEOREM 2.42. By LEMMA 4.1, we have $\|a_h\| \leq 1$. \square

The proof of the inf-sup stability in LEMMA 4.9 is a local construction, i.e.

$$\inf_{\mathbf{z} \in H(L^*, R)} \sup_{\mathbf{y} \in H(L, R)} \frac{\langle D_R \mathbf{y}_R, \mathbf{z}_R \rangle}{\|\mathbf{y}\|_{L, R} \|\mathbf{z}\|_{L^*, R}} = 1 \quad \text{for all } R \in \mathcal{R}_h.$$

Thus, we choose discrete spaces so that also the discrete local inf-sup constant is bounded independently of R .

Lemma 4.15. *For every $\mathbf{z}_h = (\mathbf{z}_R)_R \in V_{\mathcal{R}_h}^*$ there is $\mathbf{y}_h^* = (\mathbf{y}_R^*)_R \in V_{\mathcal{R}_h}$ with*

$$\sup_{\mathbf{y}_h \in V_{\mathcal{R}_h}} \frac{\langle D_h \mathbf{y}_h, \mathbf{z}_h \rangle}{\|\mathbf{y}_h\|_{L, Q_h}} = \|\mathbf{y}_h^*\|_{L, Q_h}, \quad \forall R \in \mathcal{R}_h: \sup_{\mathbf{y}_R \in V_R} \frac{\langle D_R \mathbf{y}_R, \mathbf{z}_R \rangle}{\|\mathbf{y}_R\|_{L, R}} = \|\mathbf{y}_R^*\|_{L, R}. \quad (4.18)$$

Proof. Let $\mathbf{z}_h = (\mathbf{z}_R)_R \in V_{\mathcal{R}_h}^*$. Then $D'_R \mathbf{z}_R \in V'_R$ and there is $\mathbf{y}_R^* \in V_R$ solving

$$(\mathbf{y}_R^*, \phi_R)_{L, R} = \langle D_R \phi_R, \mathbf{z}_R \rangle, \quad \phi_R \in V_R. \quad (4.19)$$

Summing all cells yields for $\mathbf{y}_h^* := (\mathbf{y}_R^*)_R \in V_{\mathcal{R}_h}$ and all $\phi_h = (\phi_R)_R \in V_{\mathcal{R}_h}$

$$(\mathbf{y}_h^*, \phi_h)_{L, Q_h} = \sum_{R \in \mathcal{R}_h} (\mathbf{y}_R^*, \phi_R)_{L, R} = \sum_{R \in \mathcal{R}_h} \langle D_R \phi_R, \mathbf{z}_R \rangle = \langle D_h \phi_h, \mathbf{z}_h \rangle. \quad (4.20)$$

Finally, by (4.19) and (4.20), we obtain for all $\phi_h = (\phi_R)_R \in V_{\mathcal{R}_h}$

$$\begin{aligned} \frac{\langle D_R \phi_R, \mathbf{z}_h \rangle}{\|\phi_R\|_{L, R}} &= \frac{(\mathbf{y}_R^*, \phi_R)_{L, R}}{\|\phi_R\|_{L, R}} \leq \|\mathbf{y}_R^*\|_{L, R}, \\ \frac{\langle D_h \phi_h, \mathbf{z}_h \rangle}{\|\phi_h\|_{L, Q_h}} &= \frac{(\mathbf{y}_h^*, \phi_h)_{L, Q_h}}{\|\phi_h\|_{L, Q_h}} \leq \|\mathbf{y}_h^*\|_{L, Q_h}, \end{aligned}$$

finishing the proof by choosing $\phi_h = \mathbf{y}_h^*$. \square

Lemma 4.16. *Assume that for all $R \in \mathcal{R}_h$ there is $\beta_R > 0$ satisfying*

$$\sup_{\mathbf{y}_R \in V_R} \frac{\langle D_R \mathbf{y}_R, \mathbf{z}_R \rangle}{\|\mathbf{y}_R\|_{L,R}} \geq \beta_R \|\mathbf{z}_R\|_{L^*,R}, \quad \mathbf{z}_R \in V_h^*|_R. \quad (4.21)$$

Then, the discrete inf-sup stability (4.17) holds with $\beta_0 = \min_{R \in \mathcal{R}_h} \beta_R > 0$.

Proof. By LEMMA 4.15, for $\mathbf{z}_h = (\mathbf{z}_R)_R \in V_h^*$ there is $\mathbf{y}_h^* = (\mathbf{y}_R^*)_R \in V_{\mathcal{R}_h}$ with (4.18).

This yields

$$\begin{aligned} \left(\sup_{\mathbf{y}_h \in V_{\mathcal{R}_h}} \frac{\langle D_h \mathbf{y}_h, \mathbf{z}_h \rangle}{\|\mathbf{y}_h\|_{L,Q_h}} \right)^2 &= \|\mathbf{y}_h^*\|_{L,Q_h}^2 = \sum_{R \in \mathcal{R}_h} \|\mathbf{y}_R^*\|_{L,R}^2 \\ &= \sum_{R \in \mathcal{R}_h} \left(\sup_{\mathbf{y}_R \in V_R} \frac{\langle D_R \mathbf{y}_R, \mathbf{z}_R \rangle}{\|\mathbf{y}_R\|_{L,R}} \right)^2 \\ &\geq \sum_{R \in \mathcal{R}_h} \left(\beta_R \|\mathbf{z}_R\|_{L^*,R} \right)^2 \geq \beta_0^2 \|\mathbf{z}_h\|_{L^*,Q_h}^2. \quad \square \end{aligned}$$

Remark 4.17. *Note that β_R can be calculated locally in each cell $R \in \mathcal{R}_h$ by a small eigenvalue problem, see (A.1) in APPENDIX A. Thus, LEMMA 4.16 provides a computationally accessible criterion to prove discrete inf-sup stability.*

4.2.4 The skeleton reduction

Addressing equation (4.15) for $(\mathbf{y}_h^{\text{sol}}, \mathbf{z}_h^{\text{sol}})$ directly, leads to an unnecessary large linear system. In the following, we use a technique that is known as SCHUR-complement reduction, see e.g. [24, Sec. 4.4.4], to reduce the size of the globally coupled linear system at the cost of many uncoupled local equations to be solved.

Similarly to (4.16), we define the discrete *local* kernel spaces for $R \in \mathcal{R}_h$

$$\begin{aligned} V_{0,R} &:= \left\{ \mathbf{y}_R \in V_R : \langle D_R \mathbf{y}_R, \boldsymbol{\psi}_R \rangle = 0 \text{ for all } \boldsymbol{\psi}_R \in V_h^*|_R \right\}, \\ V_{0,R}^* &:= \left\{ \mathbf{z}_R \in V_h^*|_R : \langle D_R \boldsymbol{\phi}_R, \mathbf{z}_R \rangle = 0 \text{ for all } \boldsymbol{\phi}_R \in V_R \right\} \end{aligned} \quad (4.22)$$

yielding the uncoupled spaces $V_{0,\mathcal{R}_h} = \prod_{R \in \mathcal{R}_h} V_{0,R}$ and $V_{0,\mathcal{R}_h}^* = \prod_{R \in \mathcal{R}_h} V_{0,R}^*$.

As a result, for $\mathbf{y}_R \in V_R$ and $\mathbf{y}_{0,R} \in V_{0,R}$, we have in every cell $R \in \mathcal{R}_h$

$$\langle D'_R \boldsymbol{\psi}_R, \mathbf{y}_R + \mathbf{y}_{0,R} \rangle = \langle D_R(\mathbf{y}_R + \mathbf{y}_{0,R}), \boldsymbol{\psi}_R \rangle, \quad \boldsymbol{\psi}_R \in V_h^*|_R.$$

Further, we define $\tilde{D}_R \in \mathcal{L}(V_R/V_{0,R}, (V_h^*|_R)')$ and $\tilde{D}'_R \in \mathcal{L}(V_h^*|_R, (V_R/V_{0,R})')$ by

$$\langle \tilde{D}_R \hat{\mathbf{y}}_R, \boldsymbol{\psi}_R \rangle := \langle \tilde{D}'_R \boldsymbol{\psi}_R, \hat{\mathbf{y}}_R \rangle := \langle D_R \mathbf{y}_R, \boldsymbol{\psi}_R \rangle, \quad \mathbf{y}_R \in \hat{\mathbf{y}}_R \in V_R/V_{0,R}, \quad \boldsymbol{\psi}_R \in V_h^*|_R,$$

and we obtain $\tilde{D}_h \in \mathcal{L}(V_h^{\text{wc}}/V_{0,\mathcal{R}_h}, (V_h^*)')$ and $\tilde{D}'_h \in \mathcal{L}(V_h^*, (V_h^{\text{wc}}/V_{0,\mathcal{R}_h})')$ with

$$\langle \tilde{D}_h \hat{\mathbf{y}}_h, \boldsymbol{\psi}_h \rangle := \langle \tilde{D}'_h \boldsymbol{\psi}_h, \hat{\mathbf{y}}_h \rangle := \sum_{R \in \mathcal{R}_h} \langle \tilde{D}_R \hat{\mathbf{y}}_R, \boldsymbol{\psi}_R \rangle$$

for $\mathbf{y}_h = (\mathbf{y}_R)_R \in \hat{\mathbf{y}}_h = (\hat{\mathbf{y}}_R)_R \in V_h^{\text{wc}}/V_{0,\mathcal{R}_h}$, $\boldsymbol{\psi}_h = (\boldsymbol{\psi}_R)_R \in V_h^*$.

Remark 4.18. Note that the operator \tilde{D}_R differs from \hat{D}_R , since a discrete space is factored out.

Following [68], we introduce a new unknown $\hat{\mathbf{y}}_h^{\text{sol}} \in V_h^{\text{wc}}/V_{0,\mathcal{R}_h}$ to approximate the skeleton trace of \mathbf{y}^{sol} . To this end, we consider the extended discrete saddle-point functional $\hat{F}_h: V_{\mathcal{R}_h} \times V_{\mathcal{R}_h}^* \times (V_h^{\text{wc}}/V_{0,\mathcal{R}_h}) \rightarrow \mathbb{R}$ given by

$$\begin{aligned} \hat{F}_h(\mathbf{y}_h, \mathbf{z}_h, \hat{\mathbf{y}}_h) &= \sum_{R \in \mathcal{R}_h} J_R(\mathbf{y}_R) + \langle D_R \mathbf{y}_R - \tilde{D}_R \hat{\mathbf{y}}_R, \mathbf{z}_R \rangle \\ &= \sum_{R \in \mathcal{R}_h} \frac{1}{2} \langle A_R \mathbf{y}_R, \mathbf{y}_R \rangle - \langle \ell_R, \mathbf{y}_R \rangle + \langle D_R \mathbf{y}_R - \tilde{D}_R \hat{\mathbf{y}}_R, \mathbf{z}_R \rangle. \end{aligned}$$

A saddle point $(\mathbf{y}_h^{\text{sol}}, \mathbf{z}_h^{\text{sol}}, \hat{\mathbf{y}}_h^{\text{sol}})$ of \hat{F}_h is characterized by

$$\forall R \in \mathcal{R}_h: \begin{cases} A_R \mathbf{y}_R^{\text{sol}} + D'_R \mathbf{z}_R^{\text{sol}} &= \ell_R & \in V'_R, \\ D_R \mathbf{y}_R^{\text{sol}} &= \tilde{D}_R \hat{\mathbf{y}}_R^{\text{sol}} & \in (V_h^*|_R)', \end{cases} \quad (4.23)$$

and

$$\tilde{D}'_h \mathbf{z}_h^{\text{sol}} = 0 \in (V_h^{\text{wc}}/V_{0,\mathcal{R}_h})'. \quad (4.24)$$

To ensure $V_{0,\mathcal{R}_h}^* = \{0\}$, we assume local inf-sup stability as formulated in

Lemma 4.19. Assume that for all cells $R \in \mathcal{R}_h$ there is $\bar{\beta}_R > 0$ such that

$$\sup_{\mathbf{y}_R \in V_R} \frac{\langle D_R \mathbf{y}_R, \mathbf{z}_R \rangle}{\|\mathbf{y}_R\|_{L,R}} \geq \bar{\beta}_R \|\mathbf{z}_R\|_{L^*,R}, \quad \mathbf{z}_R \in V_{\mathcal{R}_h}^*. \quad (4.25)$$

Then, with $\bar{\beta}_0 := \min_{R \in \mathcal{R}_h} \bar{\beta}_R$, we have for all $\mathbf{z}_h \in V_{\mathcal{R}_h}^*$

$$\sup_{\mathbf{y}_h \in V_{\mathcal{R}_h}} \frac{\langle D'_h \mathbf{z}_h, \mathbf{y}_h \rangle}{\|\mathbf{y}_h\|_{L,Q_h}} \geq \bar{\beta}_0 \|\mathbf{z}_h\|_{L^*,Q_h}.$$

Proof. Repeat the arguments in the proof of LEMMA 4.16. \square

Remark 4.20. Note that (4.25) implies (4.17).

Proposition 4.21. If $(\mathbf{y}_h^{\text{sol}}, \mathbf{z}_h^{\text{sol}}, \hat{\mathbf{y}}_h^{\text{sol}}) \in V_{\mathcal{R}_h} \times V_{\mathcal{R}_h}^* \times (V_h^{\text{wc}}/V_{0,\mathcal{R}_h})$ is a saddle point of \hat{F}_h , i.e. fulfilling (4.23) and (4.24), then we have $\mathbf{y}_h^{\text{sol}} \in V_h^{\text{wc}}$ and $\mathbf{y}_h^{\text{sol}}$ is a solution of (4.13).

Proof. Let $\mathbf{y}_h^0 \in \hat{\mathbf{y}}_h^{\text{sol}} \cap V_h^{\text{wc}}$, i.e. $\hat{\mathbf{y}}_h^{\text{sol}} = \mathbf{y}_h^0 + V_{0,\mathcal{R}_h}$. Then, by the second identity in (4.23) we obtain

$$\langle D_h \mathbf{y}_h^{\text{sol}}, \mathbf{z}_h \rangle = \langle \tilde{D}_h(\mathbf{y}_h^0 + V_{0,\mathcal{R}_h}), \mathbf{z}_h \rangle = \langle D_h \mathbf{y}_h^0, \mathbf{z}_h \rangle \quad \text{for all } \mathbf{z}_h \in V_h^*.$$

This implies $\mathbf{y}_h^{\text{sol}} - \mathbf{y}_h^0 \in V_h^{\text{wc}}$ yielding $\mathbf{y}_h^{\text{sol}} \in V_h^{\text{wc}}$. Finally, restricting the test space for the first equation in (4.23) and exploiting (4.24) imply for $\mathbf{y}_h \in V_h^{\text{wc}}$

$$\begin{aligned} a_h(\mathbf{y}_h^{\text{sol}}, \mathbf{y}_h) - \ell_h(\mathbf{y}_h) &= \sum_{R \in \mathcal{R}_h} \left(\langle A_R \mathbf{y}_R^{\text{sol}}, \mathbf{y}_R \rangle - \langle \ell_R, \mathbf{y}_R \rangle \right) + \langle \tilde{D}'_h \mathbf{z}_h^{\text{sol}}, \mathbf{y}_h + V_{0, \mathcal{R}_h} \rangle \\ &= \sum_{R \in \mathcal{R}_h} \langle A_R \mathbf{y}_R^{\text{sol}}, \mathbf{y}_R \rangle + \langle D'_R \mathbf{z}_R^{\text{sol}}, \mathbf{y}_R \rangle - \langle \ell_R, \mathbf{y}_R \rangle = 0 \end{aligned}$$

This yields the assertion. \square

In the following, we provide a construction to obtain a saddle point of \hat{F}_h which allows for the reduction of the saddle point system (4.23), (4.24) to $\hat{\mathbf{y}}_h^{\text{sol}}$.

Provided the operator $S_R := \begin{pmatrix} A_R & D'_R \\ D_R & 0 \end{pmatrix} \in \mathcal{L}(V_R \times V_h^*|_R, (V_R \times V_h^*|_R)')$ is invertible in every cell $R \in \mathcal{R}_h$, we can solve the local system (4.23) for $(\mathbf{y}_R^{\text{sol}}, \mathbf{z}_R^{\text{sol}})$ and obtain $(\mathbf{y}_R^{\text{sol}}, \mathbf{z}_R^{\text{sol}}) = S_R^{-1}(\ell_R, \tilde{D}_R \hat{\mathbf{y}}_R^{\text{sol}})$. Inserting into (4.24), this yields

$$\begin{aligned} 0 &= \sum_{R \in \mathcal{R}_h} \begin{pmatrix} 0 \\ \tilde{D}_R \end{pmatrix}' \begin{pmatrix} A_R & D'_R \\ D_R & 0 \end{pmatrix}^{-1} \begin{pmatrix} \ell_R \\ \tilde{D}_R \hat{\mathbf{y}}_R^{\text{sol}} \end{pmatrix} \\ &= \sum_{R \in \mathcal{R}_h} \left[\begin{pmatrix} 0 \\ \tilde{D}_R \end{pmatrix}' \begin{pmatrix} A_R & D'_R \\ D_R & 0 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \tilde{D}_R \end{pmatrix} \hat{\mathbf{y}}_R^{\text{sol}} + \begin{pmatrix} 0 \\ \tilde{D}_R \end{pmatrix}' \begin{pmatrix} A_R & D'_R \\ D_R & 0 \end{pmatrix}^{-1} \begin{pmatrix} \ell_R \\ 0 \end{pmatrix} \right] \end{aligned}$$

as an equation in $(V_h^{\text{wc}}/V_{0, \mathcal{R}_h})'$. Finally, we obtain the globally coupled system

$$\hat{S}_h \hat{\mathbf{y}}_h^{\text{sol}} = \hat{\ell}_h \quad \text{in } (V_h^{\text{wc}}/V_{0, \mathcal{R}_h})' \quad (4.26)$$

for the skeleton traces $\hat{\mathbf{y}}_h^{\text{sol}}$ with

$$\begin{aligned} \hat{S}_h &= - \sum_{R \in \mathcal{R}_h} \begin{pmatrix} 0 \\ \tilde{D}_R \end{pmatrix}' \begin{pmatrix} A_R & D'_R \\ D_R & 0 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \tilde{D}_R \end{pmatrix} \in \mathcal{L}(V_h^{\text{wc}}/V_{0, \mathcal{R}_h}, (V_h^{\text{wc}}/V_{0, \mathcal{R}_h})'), \\ \hat{\ell}_h &= \sum_{R \in \mathcal{R}_h} \begin{pmatrix} 0 \\ \tilde{D}_R \end{pmatrix}' \begin{pmatrix} A_R & D'_R \\ D_R & 0 \end{pmatrix}^{-1} \begin{pmatrix} \ell_R \\ 0 \end{pmatrix} \in (V_h^{\text{wc}}/V_{0, \mathcal{R}_h})'. \end{aligned}$$

Remark 4.22. *This reduction process gives rise to a solution procedure as formulated in ALGORITHM 1. This algorithm is appealing for a parallel implementation, since the dimension of the globally coupled system (4.26) is reduced from $\dim(V_{\mathcal{R}_h} \times V_h^*)$ to $\dim \hat{V}_h$ and all local contributions to \hat{S}_h and $\hat{\ell}_h$ can be assembled in parallel on each space-time cell $R \in \mathcal{R}_h$. Further, once the skeleton variable $\hat{\mathbf{y}}_h^{\text{sol}}$ has been calculated, also the local reconstruction of $\mathbf{y}_h^{\text{sol}}$ can be achieved efficiently in parallel. See CHAPTER 5 for practical computations.*

Algorithm 1 Find saddle point of \hat{F}_h

- 1: Assemble \hat{S}_h and $\hat{\ell}_h$ (no communication)
 - 2: Solve the system (4.26) for $\hat{\mathbf{y}}_h^{\text{sol}}$ (requires communication)
 - 3: For each $R \in \mathcal{R}_h$: solve (4.23) for $\mathbf{y}_R^{\text{sol}}$ (no communication)
-

Proposition 4.23. *Assume that $\alpha_0, \bar{\beta}_0 > 0$ satisfying (4.14) and (4.25) exist.*

Then, \hat{S}_h is well-defined and symmetric, i.e. $\langle \hat{S}_h \hat{\phi}_h, \hat{\mathbf{y}}_h \rangle = \langle \hat{S}_h \hat{\mathbf{y}}_h, \hat{\phi}_h \rangle$ for all $\hat{\mathbf{y}}_h, \hat{\phi}_h \in V_h^{\text{wc}}/V_{0,\mathcal{R}_h}$. Further, \hat{S}_h satisfies the spectral bounds

$$\alpha_0 \|\hat{\mathbf{y}}_h\|_{V_h^{\text{wc}}/V_{0,\mathcal{R}_h}}^2 \leq \langle \hat{S}_h \hat{\mathbf{y}}_h, \hat{\mathbf{y}}_h \rangle \leq \frac{1}{\alpha_0 \bar{\beta}_0^2} \|\hat{\mathbf{y}}_h\|_{V_h^{\text{wc}}/V_{0,\mathcal{R}_h}}^2, \quad \hat{\mathbf{y}}_h \in V_h^{\text{wc}}/V_{0,\mathcal{R}_h}.$$

Proof. We consider the linear operators $S_{\mathcal{R}_h} \in \mathcal{L}(V_{\mathcal{R}_h} \times V_{\mathcal{R}_h}^*, (V_{\mathcal{R}_h} \times V_{\mathcal{R}_h}^*)')$ and $G_h \in \mathcal{L}(V_h^{\text{wc}}/V_{0,h}, (V_{\mathcal{R}_h} \times V_{\mathcal{R}_h}^*)')$ defined by

$$\begin{aligned} \langle S_{\mathcal{R}_h}(\mathbf{y}_h, \mathbf{z}_h), (\phi_h, \psi_h) \rangle &:= \sum_{R \in \mathcal{R}_h} \langle A_R \mathbf{y}_R, \phi_R \rangle + \langle D'_R \mathbf{z}_R, \phi_R \rangle + \langle D_R \mathbf{y}_R, \psi_R \rangle, \\ \langle G_R \hat{\mathbf{y}}_R, (\phi_R, \psi_R) \rangle &:= \langle \tilde{D}_R \hat{\mathbf{y}}_R, \psi_R \rangle, \\ \langle G_h \hat{\mathbf{y}}_h, (\phi_h, \psi_h) \rangle &:= \sum_{R \in \mathcal{R}_h} \langle G_R \hat{\mathbf{y}}_R, (\phi_R, \psi_R) \rangle \end{aligned}$$

for $\mathbf{y}_h \in V_{\mathcal{R}_h}$, $\hat{\mathbf{y}}_h \in V_h^{\text{wc}}/V_{0,\mathcal{R}_h}$ and $(\phi_h, \psi_h) \in V_{\mathcal{R}_h} \times V_{\mathcal{R}_h}^*$. For fixed $\hat{\mathbf{y}}_h \in V_h^{\text{wc}}/V_{0,\mathcal{R}_h}$, a pair $(\mathbf{y}_h, \mathbf{z}_h) \in V_{\mathcal{R}_h} \times V_{\mathcal{R}_h}^*$ with $\mathbf{y}_h = (\mathbf{y}_R)_R$, $\mathbf{z}_h = (\mathbf{z}_R)_R$, fulfills the identity $S_{\mathcal{R}_h}(\mathbf{y}_h, \mathbf{z}_h) = G_h \hat{\mathbf{y}}_h$ if and only if

$$\forall R \in \mathcal{R}_h: \begin{cases} \langle A_R \mathbf{y}_R, \phi_R \rangle + \langle D'_R \mathbf{z}_R, \phi_R \rangle = 0, & \phi_R \in V_R, \\ \langle D_R \mathbf{y}_R, \psi_R \rangle = \langle \tilde{D}_R \hat{\mathbf{y}}_R, \psi_R \rangle, & \psi_R \in V_h^*|_R, \end{cases} \quad (4.27)$$

Since $V_h^* \subset V_{\mathcal{R}_h}^*$, we see $V_{0,\mathcal{R}_h} \subset V_h^{\text{wc}}$ by (4.22) and (4.9). Applying (4.14), this yields $a_h(\mathbf{y}_{0,h}, \mathbf{y}_{0,h}) \geq \alpha_0 \|\mathbf{y}_{0,h}\|_{L,Q_h}^2$ for $\mathbf{y}_{0,h} \in V_{0,\mathcal{R}_h}$. Using LEMMA 4.19, we obtain that the adjoint kernel space V_{0,\mathcal{R}_h}^* is trivial. Thus, THEOREM 2.37 yields the well-posedness of (4.27). For every cell $R \in \mathcal{R}_h$, we define $\mathbf{y}_R \in V_R$, $\mathbf{z}_R \in V_h^*|_R$ by

$$\begin{pmatrix} \mathbf{y}_R \\ \mathbf{z}_R \end{pmatrix} = \begin{pmatrix} A_R & D'_R \\ D_R & 0 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \tilde{D}_R \hat{\mathbf{y}}_R \end{pmatrix}.$$

By the second identity in (4.27), we see $\mathbf{y}_h := (\mathbf{y}_R)_R \in \hat{\mathbf{y}}_h$ and thus, $\mathbf{y}_h \in V_h^{\text{wc}}$. Using the definition of \hat{S}_h and (4.27), we obtain

$$\langle \hat{S}_h \hat{\mathbf{y}}_h, \hat{\mathbf{y}}_h \rangle = - \sum_{R \in \mathcal{R}_h} \langle \tilde{D}'_R \mathbf{z}_R, \hat{\mathbf{y}}_R \rangle = - \sum_{R \in \mathcal{R}_h} \langle D_R \hat{\mathbf{y}}_R, \mathbf{z}_R \rangle = \sum_{R \in \mathcal{R}_h} \langle A_R \mathbf{y}_R, \mathbf{y}_R \rangle.$$

Exploiting (4.14) and the definition of the quotient norm in $V_h^{\text{wc}}/V_{0,\mathcal{R}_h}$, we obtain the lower bound by $\langle \hat{S}_h \hat{\mathbf{y}}_h, \hat{\mathbf{y}}_h \rangle \geq \alpha_0 \|\mathbf{y}_h\|_{L,Q_h}^2 \geq \alpha_0 \|\hat{\mathbf{y}}_h\|_{V_h^{\text{wc}}/V_{0,\mathcal{R}_h}}^2$.

To show the upper bound, we calculate using (4.25)

$$\bar{\beta}_0 \|\mathbf{z}_h\|_{L^*,Q_h} \leq \sup_{\phi_h \in V_{\mathcal{R}_h}} \frac{\langle D'_h \mathbf{z}_h, \phi_h \rangle}{\|\phi_h\|_{L,Q_h}} = \sup_{\phi_h \in V_{\mathcal{R}_h}} \frac{\sum_{R \in \mathcal{R}_h} \langle A_R \mathbf{y}_R, \phi_R \rangle}{\|\phi_R\|_{L,Q_h}} \leq \|\mathbf{y}_h\|_{L,Q_h}$$

and by the definition of the quotient norm, we have

$$\begin{aligned} \alpha_0 \|\mathbf{y}_h\|_{L,Q_h}^2 &\leq a_h(\mathbf{y}_h, \mathbf{y}_h) = - \sum_{R \in \mathcal{R}_h} \langle D_R \mathbf{y}_R, \mathbf{z}_R \rangle = - \sum_{R \in \mathcal{R}_h} \langle \tilde{D}_R \hat{\mathbf{y}}_R, \mathbf{z}_R \rangle \\ &\leq \|\hat{\mathbf{y}}_h\|_{V_h^{\text{wc}}/V_{0,\mathcal{R}_h}} \|\mathbf{z}_h\|_{L^*,Q_h} \leq \frac{1}{\bar{\beta}_0} \|\hat{\mathbf{y}}_h\|_{V_h^{\text{wc}}/V_{0,\mathcal{R}_h}} \|\mathbf{y}_h\|_{L,Q_h}. \end{aligned}$$

This implies

$$\langle \hat{S}_h \hat{\mathbf{y}}_h, \hat{\mathbf{y}}_h \rangle = - \sum_{R \in \mathcal{R}_h} \langle \tilde{D}'_R \mathbf{z}_R, \hat{\mathbf{y}}_R \rangle \leq \|\hat{\mathbf{y}}_h\|_{V_h^{\text{wc}}/V_{0,\mathcal{R}_h}} \|\mathbf{z}_h\|_{L^*,Q_h} \leq \frac{1}{\alpha_0 \bar{\beta}_0^2} \|\hat{\mathbf{y}}_h\|_{V_h^{\text{wc}}/V_{0,\mathcal{R}_h}}^2.$$

Since \hat{S}_h is symmetric by its structure, this finishes the proof. \square

Remark 4.24. *The result in PROPOSITION 4.23 shows that the skeleton reduction procedure yields a symmetric and positive definite system.*

4.2.5 Discussion

In SECTION 5.2, we provide numerical experiments demonstrating the performance of this weakly conforming Least-Squares method. For one space dimension, we know various pairings $V_{\mathcal{R}_h}$ and V_h^* fulfilling (4.14), (4.25) and the numerical results are promising.

Unfortunately, in two space dimensions we were not able to find such a pairing, since either the local saddle point matrix S_R or the globally reduced system matrix \hat{S}_h are singular in all examples we tested.

It turns out that (4.14) and (4.25) can build up each other. On the one hand, (4.14) requires strong coupling along the faces meaning that for a given space $V_{\mathcal{R}_h}$ the test space V_h^* needs to be large. On the other hand, the space $V_{\mathcal{R}_h}$ has to be large enough to make the supremum in (4.25) positive.

In case that (4.14) holds but (4.25) is not fulfilled, there still might exist a saddle point of F_h . However, the LAGRANGE multiplier $\mathbf{z} \in V^*$ is not unique anymore.

Since we are interested in the primal unknown $\mathbf{y}_h \in V_h^{\text{wc}}$, any LAGRANGE multiplier would do the job. It might be possible to drop condition (4.25) and solve the singular problem, e.g. using Least-Squares approaches.

A similar strategy for an overdetermined problem is considered in [39].

4.3 The Discontinuous PETROV-GALERKIN method

In this section, we present a space-time DPG method that is another variant of a minimal residual method. Before we apply the Discontinuous PETROV-GALERKIN (DPG) methodology to the acoustic wave equation, we explain some key concepts on an abstract level using a similar notation as in SECTION 4.1.

There is vast literature on the DPG method covering the theory and wide ranges of different applications. The fundamental ideas are summarized in e.g. [11, 14, 20]. Also see [11, 52, 73] for applications to FRIEDRICHS systems and wave equations. More recently, the DPG method has been applied to space-time variational formulations. In [21], a space-time formulation for the SCHRÖDINGER equation is considered and in [33], the authors consider acoustic waves in space-time.

4.3.1 Continuous and discrete well-posedness

For HILBERT spaces V and Z , a bilinear form $b \in \mathcal{B}(V \times Z, \mathbb{R})$ and $\ell \in Z'$, we consider the general variational problem as in CHAPTER 2

$$\begin{cases} \text{Find } u \in V \text{ such that} \\ b(u, z) = \ell(z) \quad \text{for all } z \in Z, \end{cases} \quad (\text{VP})$$

and briefly revisit the well-posedness results.

THEOREM 2.17 tells us that (VP) is well-posed for all right-hand sides $\ell \in Z'$ if and only if we have inf-sup stability and definiteness, i.e.

$$\exists \beta > 0: \quad \inf_{v \in V} \sup_{z \in Z} \frac{b(v, z)}{\|v\|_V \|z\|_Z} \geq \beta, \quad (\text{BNB1})$$

and

$$\forall z \in Z: \quad (\forall v \in V: b(v, z) = 0) \implies (z = 0). \quad (\text{BNB2})$$

Assuming we are considering a well-posed problem, i.e. (BNB1) and (BNB2) are fulfilled, we can construct a discrete approximation by choosing finite-dimensional spaces $V_h \subset V$ and $Z_h \subset Z$. This yields a discrete counterpart of (VP) given by

$$\begin{cases} \text{Find } u_h \in V_h \text{ such that} \\ b(u_h, z_h) = \ell(z_h) \quad \text{for all } z_h \in Z_h. \end{cases} \quad (\text{VP}_h)$$

Again, THEOREM 2.17 guarantees that (VP_h) is well-posed if and only if we have discrete inf-sup stability and discrete definiteness, i.e.

$$\exists \beta_h > 0: \quad \inf_{v_h \in V_h} \sup_{z_h \in Z_h} \frac{b(v_h, z_h)}{\|v_h\|_V \|z_h\|_Z} \geq \beta_h, \quad (\text{BNB1}_h)$$

and

$$\forall z_h \in Z_h: \quad (\forall v_h \in V_h: b(v_h, z_h) = 0) \implies (z_h = 0). \quad (\text{BNB2}_h)$$

Since the supremum in (BNB1_h) is taken over the smaller set Z_h compared to Z in (BNB1), the inf-sup stability of the original problem does not carry over to the discrete problem in general. According to PROPOSITION 2.24, at least (BNB2_h) holds true as soon as (BNB1_h) is fulfilled and $\dim V_h = \dim Z_h$.

One key idea of the Discontinuous PETROV-GALERKIN (DPG) methodology arises from the following question: Given a trial space $V_h \subset V$, how to construct an *optimal* test space $Z_h \subset Z$ such that the discrete inf-sup constant β_h is not worse than the continuous constant β ?

4.3.2 Optimal test functions – the trial-to-test operator

The answer to that question leads to the trial-to-test operator $\mathcal{T}: V \rightarrow Z$ mapping each trial function $v \in V$ to the test function $\mathcal{T}v \in Z$ that realizes the supremum in (BNB1). We define \mathcal{T} as follows.

For $v \in V$, we calculate the RIESZ-representative $\mathcal{T}v \in Z$ of the linear form $b(v, \cdot) \in Z'$, i.e.

$$(\mathcal{T}v, \tilde{z})_Z = b(v, \tilde{z}) \quad \text{for all } \tilde{z} \in Z, \quad (4.28)$$

Then, we obtain the supremum using the CAUCHY-SCHWARZ inequality by

$$\sup_{z \in Z} \frac{b(v, z)}{\|z\|_Z} = \sup_{z \in Z} \frac{(\mathcal{T}v, z)_Z}{\|z\|_Z} = \frac{(\mathcal{T}v, \mathcal{T}v)_Z}{\|\mathcal{T}v\|_Z} = \frac{b(v, \mathcal{T}v)}{\|\mathcal{T}v\|_Z}. \quad (4.29)$$

Thus, choosing the *optimal* test space $Z_h := \mathcal{T}(V_h)$, we obtain for $v_h \in V_h$

$$\sup_{z_h \in Z_h} \frac{b(v_h, z_h)}{\|z_h\|_Z} = \frac{b(v_h, \mathcal{T}v_h)}{\|\mathcal{T}v_h\|_Z} = \sup_{z \in Z} \frac{b(v_h, z)}{\|z\|_Z} \geq \beta \|v_h\|_V. \quad (4.30)$$

Here, the first equality in (4.30) holds since for each $v_h \in V_h$ the maximizer $\mathcal{T}v_h$ is contained in Z_h . As a result, the PETROV-GALERKIN scheme (VP_h) using the optimal test space $Z_h = \mathcal{T}(V_h)$ is inf-sup stable by construction with the same or even a larger constant compared to the continuous problem. Furthermore, we have $\dim V_h = \dim Z_h$ resulting in the well-posedness of (VP_h). However, as soon as the norm in Z involves a differential operator, in general, solving (4.28) can be as hard as solving (VP) itself.

The second key idea of the DPG methodology comes up with a localized way to solve (4.28).

Remark 4.25. *Using the operator $B \in \mathcal{L}(V, Z')$ from DEFINITION 2.3 fulfilling $\langle Bv, z \rangle = b(v, z) = \langle B'z, v \rangle$ for $v \in V, z \in Z$, and inserting the RIESZ-isomorphism $\Pi_Z \in \mathcal{L}(Z, Z')$, we derive $\mathcal{T} = \Pi_Z^{-1}B$ from (4.28).*

4.3.3 Breaking the test space – the ideal DPG method

In the following, we assume that V and Z are function spaces on a space-time domain $Q \subset \mathbb{R}^{1+d}$. Furthermore, we assume that b , ℓ and $\|\cdot\|_Z$ are *localizable*, i.e. they can be written as sums of cell-wise defined local counterparts. More precisely, given a triangulation \mathcal{R}_h of Q , we assume that of cell-wise defined $b_R \in \mathcal{B}(V|_R \times Z|_R, \mathbb{R})$, $\ell_R \in (Z|_R)'$ and a norm $\|\cdot\|_{Z,R}: Z|_R \rightarrow \mathbb{R}$ on $Z|_R$ exist such that

$$b(v, z) = \sum_{R \in \mathcal{R}_h} b_R(v|_R, z|_R), \quad \ell(z) = \sum_{R \in \mathcal{R}_h} \ell_R(z|_R), \quad \|z\|_Z^2 = \sum_{R \in \mathcal{R}_h} \|z|_R\|_{Z,R}^2,$$

for all $v \in V$, $z \in Z$. Note that by summation, we can extend b and ℓ to the cell-wise defined space $Z_{\mathcal{R}_h} := \prod_{R \in \mathcal{R}_h} Z|_R$. These extensions are denoted by

$$b_{\mathcal{R}_h}(v, z_{\mathcal{R}_h}) := \sum_{R \in \mathcal{R}_h} b_R(v|_R, z_R), \quad \ell_{\mathcal{R}_h}(z_{\mathcal{R}_h}) := \sum_{R \in \mathcal{R}_h} \ell_R(z_R),$$

for $v \in V$, $z_{\mathcal{R}_h} = (z_R)_R \in Z_{\mathcal{R}_h}$.

Often, the local bilinear form b_R is constructed using integration by parts on the broken domain Q_h . To this end, additionally to the volume term b_R , usually a boundary pairing term occurs on every cell boundary. In this abstract outline, we assume that there is a space $\hat{V} = \prod_{R \in \mathcal{R}_h} \hat{V}_R$ representing traces of the solution on the cell boundaries and that a bilinear form $\hat{b}_R \in \mathcal{B}(\hat{V}_R \times Z|_R, \mathbb{R})$ exists such that the result of locally integrating by parts can be written as

$$b_R(v|_R, z_R) + \hat{b}_R(\gamma_R v|_R, z_R), \quad z_R \in Z|_R,$$

for sufficiently regular functions $v \in V^{\text{regular}} \subset V$. Here, $\gamma_R: V^{\text{regular}} \rightarrow \hat{V}_R$ is a suitable trace operator acting on the cell R .

Finally, we assume that problem (VP) remains unchanged, when testing with a cell-wise defined test space, i.e. $v \in V$ solves (VP) if and only if there exists a $\hat{v} = (\hat{v}_R)_R \in \hat{V}$ such that

$$\sum_{R \in \mathcal{R}_h} b_R(v|_R, z_R) + \hat{b}_R(\hat{v}_R, z_R) = \ell_{\mathcal{R}_h}(z_R) \quad \text{for all } (z_R)_R \in \prod_{R \in \mathcal{R}_h} Z|_R. \quad (4.31)$$

By assumption also the variational problem with broken test space is well-posed and therefore inf-sup stable with a constant $\beta_{\mathcal{R}_h} > 0$ by THEOREM 2.17. In [14, Thm. 3.1], criteria for the assumptions above are provided.

Before constructing the space-time DPG method for acoustic waves in the next section, we consider a static POISSON problem to illustrate the ideas.

Example 4.26. For a domain $\Omega \subset \mathbb{R}^d$ and $f \in L_2(\Omega)$, we consider the problem

$$\begin{cases} \boldsymbol{\sigma} - \nabla p = 0, \\ \operatorname{div} \boldsymbol{\sigma} = f, \end{cases} \quad \text{in } \Omega, \quad p = 0 \text{ on } \partial\Omega.$$

Assuming that $(\boldsymbol{\sigma}, p)$ is sufficiently smooth, we multiply by a smooth test function $(\boldsymbol{\phi}, \psi)$ and integrate by parts to obtain in every open subset $U \subset \Omega$

$$\begin{aligned} ((\boldsymbol{\sigma} - \nabla p, \operatorname{div} \boldsymbol{\sigma}), (\boldsymbol{\phi}, \psi))_U &= (\boldsymbol{\sigma}, \boldsymbol{\phi})_U + (p, \operatorname{div} \boldsymbol{\phi})_U - (\boldsymbol{\sigma}, \nabla \psi)_U \\ &\quad - \langle p, \boldsymbol{\phi} \cdot \boldsymbol{n}_U \rangle_{\partial U} + \langle \boldsymbol{\sigma} \cdot \boldsymbol{n}_U, \psi \rangle_{\partial U}. \end{aligned} \quad (4.32)$$

In particular the choice $U = \Omega$ in (4.32) yields using the boundary condition $p = 0$ on $\partial\Omega$ and testing with $Z := \mathbf{H}(\operatorname{div}, \Omega) \times \mathbf{H}_0^1(\Omega)$ a variational problem in the form of (VP) where $V := L_2(\Omega, \mathbb{R}^{d+1})$ and, for $(\boldsymbol{\sigma}, p) \in V$, $(\boldsymbol{\phi}, \psi) \in Z$,

$$\begin{aligned} b((\boldsymbol{\sigma}, p), (\boldsymbol{\phi}, \psi)) &= (\boldsymbol{\sigma}, \boldsymbol{\phi})_\Omega + (p, \operatorname{div} \boldsymbol{\phi})_\Omega - (\boldsymbol{\sigma}, \nabla \psi)_\Omega, \\ \ell((\boldsymbol{\phi}, \psi)) &= (f, \psi)_\Omega. \end{aligned}$$

This problem is discussed more in depth by [17].

In order to construct a variational problem with broken test space, we apply (4.32) in every cell $K \in \mathcal{K}_h$, where \mathcal{K}_h is a mesh of Ω , and obtain

$$\begin{aligned} b_K((\boldsymbol{\sigma}, p), (\boldsymbol{\phi}, \psi)) &= (\boldsymbol{\sigma}, \boldsymbol{\phi})_K + (p, \operatorname{div} \boldsymbol{\phi})_K - (\boldsymbol{\sigma}, \nabla \psi)_K, \\ \hat{b}_K((\hat{\boldsymbol{\sigma}}_n, \hat{p}), (\boldsymbol{\phi}, \psi)) &= -\langle \hat{p}, \boldsymbol{\phi} \cdot \boldsymbol{n}_K \rangle_{\partial K} + \langle \hat{\boldsymbol{\sigma}}_n, \psi \rangle_{\partial K}, \\ \ell_K((\boldsymbol{\phi}, \psi)) &= (f, \psi)_K \end{aligned}$$

for $(\hat{\boldsymbol{\sigma}}_n, \hat{p}) \in \hat{V}_K := \gamma_K(\mathbf{H}(\operatorname{div}, \Omega) \times \mathbf{H}_0^1(\Omega)) \subset \mathbf{H}^{-1/2}(\partial K) \times \mathbf{H}^{1/2}(\partial K)$. Here, $\gamma_K(\boldsymbol{\sigma}, p) = ((\boldsymbol{\sigma} \cdot \boldsymbol{n}_K)|_{\partial K}, p|_{\partial K})$ is the pair of normal and DIRICHLET traces in the sense of trace operators.

This gives rise to the broken problem as in (4.31)

$$\left\{ \begin{array}{l} \text{Find } ((\boldsymbol{\sigma}, p), (\hat{\boldsymbol{\sigma}}_n, \hat{p})) \in V \times \prod_{K \in \mathcal{K}_h} \hat{V}_K \text{ such that} \\ \sum_{K \in \mathcal{K}_h} b_K((\boldsymbol{\sigma}, p), (\boldsymbol{\phi}_K, \psi_K)) + \hat{b}_K((\hat{\boldsymbol{\sigma}}_n, \hat{p}), (\boldsymbol{\phi}_K, \psi_K)) = \ell_{\mathcal{K}_h}(\boldsymbol{\phi}_K, \psi_K) \\ \text{for all } (\boldsymbol{\phi}_K, \psi_K)_K \in \prod_{K \in \mathcal{K}_h} Z|_K. \end{array} \right.$$

See [14] where variants of this problem are considered.

Using the same construction as in the previous section, we define the trial-to-test operator $\mathcal{T}_{\mathcal{R}_h} : V \times \hat{V} \rightarrow Z_{\mathcal{R}_h}$ as in (4.28). However, since the new test space $Z_{\mathcal{R}_h}$

consists of discontinuous functions, for $(v, \hat{v}) \in V \times \hat{V}$ the optimal test function $\mathcal{T}_{\mathcal{R}_h}(v, \hat{v}) \in Z_{\mathcal{R}_h}$ is defined *locally* in every cell, i.e. introducing the local bilinear form $\bar{b}_R((v, \hat{v}), \tilde{z}_R) := b_R(v, \tilde{z}_R) + \hat{b}_R(\hat{v}, \tilde{z}_R)$, equation (4.28) becomes

$$(\mathcal{T}_{\mathcal{R}_h}(v, \hat{v})|_R, \tilde{z}_R)_{Z_{\mathcal{R}_h}} = \bar{b}_R((v, \hat{v}), \tilde{z}_R) \quad \text{for all } \tilde{z}_R \in Z|_R, R \in \mathcal{R}_h. \quad (4.33)$$

Using the broken variant, we repeat the reasoning in (4.29)

$$\begin{aligned} \sup_{z_{\mathcal{R}_h} \in Z_{\mathcal{R}_h}} \frac{\bar{b}_{\mathcal{R}_h}((v, \hat{v}), z_{\mathcal{R}_h})}{\|z_{\mathcal{R}_h}\|_{Z_{\mathcal{R}_h}}} &= \sup_{z_{\mathcal{R}_h} \in Z_{\mathcal{R}_h}} \frac{(\mathcal{T}_{\mathcal{R}_h}(v, \hat{v}), z_{\mathcal{R}_h})_{Z_{\mathcal{R}_h}}}{\|z_{\mathcal{R}_h}\|_{Z_{\mathcal{R}_h}}} \\ &= \frac{(\mathcal{T}_{\mathcal{R}_h}(v, \hat{v}), \mathcal{T}_{\mathcal{R}_h}(v, \hat{v}))_{Z_{\mathcal{R}_h}}}{\|\mathcal{T}_{\mathcal{R}_h}(v, \hat{v})\|_{Z_{\mathcal{R}_h}}} = \frac{\bar{b}_{\mathcal{R}_h}((v, \hat{v}), \mathcal{T}_{\mathcal{R}_h}(v, \hat{v}))}{\|\mathcal{T}_{\mathcal{R}_h}(v, \hat{v})\|_{Z_{\mathcal{R}_h}}}. \end{aligned}$$

Again, choosing the *optimal* test space $Z_{\mathcal{R}_h, h} := \mathcal{T}_{\mathcal{R}_h}(V_h \times \hat{V}_h)$ for a given ansatz space $V_h \times \hat{V}_h \subset V \times \hat{V}$ as in (4.30), we obtain for $(v_h, \hat{v}_h) \in V_h \times \hat{V}_h$

$$\begin{aligned} \sup_{z_{\mathcal{R}_h, h} \in Z_{\mathcal{R}_h, h}} \frac{b_{\mathcal{R}_h}((v_h, \hat{v}_h), z_{\mathcal{R}_h, h})}{\|z_{\mathcal{R}_h, h}\|_{Z_{\mathcal{R}_h, h}}} &= \sup_{z_{\mathcal{R}_h} \in Z_{\mathcal{R}_h}} \frac{b_{\mathcal{R}_h}((v_h, \hat{v}_h), z_{\mathcal{R}_h})}{\|z_{\mathcal{R}_h}\|_{Z_{\mathcal{R}_h}}} \\ &\geq \beta_{\mathcal{R}_h} \|(v_h, \hat{v}_h)\|_{V \times \hat{V}}, \end{aligned} \quad (4.34)$$

where $\beta_{\mathcal{R}_h} > 0$ is the inf-sup constant of the variational problem (4.31) with broken test space.

4.3.4 Approximate optimal testing – the practical DPG method

Breaking the test space at the cost of introducing new trace unknowns in the space \hat{V} yields a construction, where the optimal test function can be obtained *locally* in every cell. However, the variational problem (4.33) still searches in the infinite dimensional space $Z_{\mathcal{R}_h}$ and thus, is practically infeasible in most applications. For a simple example, where the ideal DPG method can be realized in practice, see [18].

To obtain a practical scheme, we solve (4.33) in a finite dimensional *enriched* test space $Z_{R, h}^{\text{enriched}} \subset Z|_R$, $R \in \mathcal{R}_h$, and $Z_h^{\text{enriched}} = \prod_{R \in \mathcal{R}_h} Z_{R, h}^{\text{enriched}}$ fulfilling $\dim Z_h^{\text{enriched}} > \dim(V_h \times \hat{V}_h)$. For the practical computation, we replace (4.33) by its GALERKIN approximation

$$(\mathcal{T}_{R, h}(v, \hat{v}), \tilde{z}_{R, h})_{Z_{R, h}} = \bar{b}_R((v, \hat{v}), \tilde{z}_{R, h}), \quad \text{for all } \tilde{z}_{R, h} \in Z_{R, h}^{\text{enriched}}, R \in \mathcal{R}_h,$$

searching for $\mathcal{T}_{R, h}(v, \hat{v}) \in Z_{R, h}^{\text{enriched}}$ for every pair $(v, \hat{v}) \in V \times \hat{V}$. This defines the approximate trial-to-test operator $\mathcal{T}_{R, h}: V \times \hat{V} \rightarrow Z_{R, h}^{\text{enriched}}$ and accordingly, the approximate optimal test space $Z_{R, h} := \mathcal{T}_{R, h}(V_h \times \hat{V}_h) \subset Z_{R, h}^{\text{enriched}}$.

It has been observed in many applications that for polynomial ansatz spaces of degree $p \in \mathbb{N}$, typically polynomials of degree $p + \Delta p$, $\Delta p \in \{2, 3\}$, in $Z_{R, h}^{\text{enriched}}$ are sufficient to guarantee the stability of the resulting DPG scheme.

4.3.5 DPG as a minimal residual method

The DPG method can also be interpreted as a minimal residual method in the dual space Z' . To see this, we reconsider the trial-to-test operator from (4.28) for $v \in V$

$$(\mathcal{T}v, \tilde{z})_Z = b(v, \tilde{z}) = \langle Bv, \tilde{z} \rangle, \quad \text{for all } \tilde{z} \in Z,$$

where we use the operator $B \in \mathcal{L}(V, Z')$ as in DEFINITION 2.3. Consequently, we have $\mathcal{T} = \Pi_Z^{-1}B$ for the RIESZ isomorphism $\Pi_Z \in \mathcal{L}(Z, Z')$ yielding for $v, \tilde{v} \in V$

$$b(v, \mathcal{T}\tilde{v}) - \ell(\mathcal{T}\tilde{v}) = \langle Bv - \ell, \Pi_Z^{-1}B\tilde{v} \rangle_Z = (\Pi_Z^{-1}(Bv - \ell), \Pi_Z^{-1}B\tilde{v})_Z. \quad (4.35)$$

Now, we select an approximation space $V_h \subset V$ and the corresponding optimal test space $Z_h = \mathcal{T}(V_h) \subset Z$. Using the same arguments as for the conforming Least-Squares method in SECTION 4.1, we see that $v_h \in V_h$ solves (VP_h) if and only if v_h is the minimizer of $J_h: V_h \rightarrow \mathbb{R}$ given by

$$\begin{aligned} J(v_h) &= \frac{1}{2}(\Pi_Z^{-1}(Bv_h - \ell), \Pi_Z^{-1}(Bv_h - \ell))_Z = \frac{1}{2}\|\Pi_Z^{-1}(Bv_h - \ell)\|_Z^2 \\ &= \frac{1}{2}\|Bv_h - \ell\|_{Z'}^2. \end{aligned} \quad (4.36)$$

Here, we exploit that the RIESZ isomorphism is an isometry. As a result, the DPG method can be interpreted as a generalized Least-Squares method in the dual space Z' .

4.3.6 Built-in residual error-estimator

The DPG method provides a built-in error estimator that can be evaluated numerically. To see this, for the solution $u_h \in V_h$ of (VP_h), we define the RIESZ representative of the residual by $\psi := \Pi_Z^{-1}(\ell - Bu_h)$, i.e. $(\psi, \tilde{z})_Z = \langle \ell - Bu_h, \tilde{z} \rangle = \ell(\tilde{z}) - b(u_h, \tilde{z})$ for $\tilde{z} \in Z$. Using the energy norm $\|v\|_E := \|Bv\|_{Z'}$, $v \in V$, this implies for the solution $u \in V$ of (VP)

$$\|u - u_h\|_E = \|Bu - Bu_h\|_{Z'} = \|\ell - Bu_h\|_{Z'} = \|\Pi_Z^{-1}(\ell - Bu_h)\|_Z = \|\psi\|_Z,$$

being the reason to call ψ the *error representing function*.

Then, (VP) and (4.35) yield the following system for the new group unknown $(u_h, \psi) \in V_h \times Z$ that is also called the *mixed formulation of the DPG method*

$$\begin{cases} b(\tilde{v}_h, \psi) &= 0, & \text{for all } \tilde{v}_h \in V_h, \\ (\psi, \tilde{z})_Z + b(u_h, \tilde{z}) &= \ell(\tilde{z}), & \text{for all } \tilde{z} \in Z. \end{cases} \quad (4.37)$$

Now, approximating (u_h, ψ) simultaneously in $V_h \times Z_h^{\text{enriched}}$ gives a residual error estimator.

4.3.7 Skeleton reduction

In [67] a reduction procedure for DPG similar to the skeleton reduction for weakly conforming Least-Squares, see SECTION 4.2.4, is described. We sketch the construction for the case of acoustic waves in SECTION 4.6.2.

4.4 Acoustic waves – the ideal DPG method

In the following, we use the notation introduced in CHAPTER 3.

To apply the DPG-method to acoustic waves, we select a finite decomposition $Q_h = \bigcup_{R \in \mathcal{R}_h} R$ of $Q = \Omega \times (0, T)$ into open disjoint space-time cells R .

Breaking the space $H(L, Q)$, we consider the corresponding discontinuous space $H(L, Q_h) = \prod_{R \in \mathcal{R}_h} H(L, R)$.

Now, the ideal DPG method is constructed as follows. With respect to the substructuring Q_h , we introduce a new unknown for the skeleton traces and represent the solution in the product space $W \times \hat{H}(L, Q_h)$ with $W = L_2(Q, \mathbb{R}^{1+d})$. For given $\mathbf{b} \in W$ let $\mathbf{y}^{\text{sol}} \in V$ be the unique solution of $L\mathbf{y} = \mathbf{b}$, and define its trace $\hat{\mathbf{y}}^{\text{sol}} = \mathbf{y}^{\text{sol}} + H_0(L, Q_h) \in \hat{H}(L, Q_h)$. Then, inserting \hat{D}_h , see (3.14), yields

$$(\mathbf{b}, \mathbf{z})_Q = (L\mathbf{y}^{\text{sol}}, \mathbf{z})_Q = (\mathbf{y}^{\text{sol}}, L^*\mathbf{z})_{Q_h} + \langle \hat{D}_h \hat{\mathbf{y}}^{\text{sol}}, \mathbf{z} \rangle, \quad \mathbf{z} \in H(L^*, Q_h).$$

For the corresponding PETROV-GALERKIN method in $W \times \hat{H}(L, Q_h)$, we define the operator

$$B_h \in \mathcal{L}(W \times \hat{H}(L, Q_h), H(L^*, Q_h)'), \quad \langle B_h(\mathbf{y}, \hat{\mathbf{y}}), \mathbf{z} \rangle = (\mathbf{y}, L^*\mathbf{z})_Q + \langle \hat{D}_h \hat{\mathbf{y}}, \mathbf{z} \rangle$$

for trial functions $(\mathbf{y}, \hat{\mathbf{y}}) \in W \times \hat{H}(L, Q_h)$ and test functions $\mathbf{z} \in H(L^*, Q_h)$.

As a result, the pair $(\mathbf{y}^{\text{sol}}, \hat{\mathbf{y}}^{\text{sol}})$ solves the equation

$$\langle B_h(\mathbf{y}, \hat{\mathbf{y}}), \mathbf{z} \rangle = (\mathbf{b}, \mathbf{z})_Q, \quad \mathbf{z} \in H(L^*, Q_h). \quad (4.38)$$

The norm in $W \times \hat{H}(L, Q_h)$ is denoted by $\|(\mathbf{y}, \hat{\mathbf{y}})\|_{Q;L,\partial Q_h} = \sqrt{\|\mathbf{y}\|_Q^2 + \|\hat{\mathbf{y}}\|_{L,\partial Q_h}^2}$.

Now, we show that (4.38) is well-posed by proving that the restriction of B_h to $W \times \hat{V}$ with $\hat{V} = V/H_0(L, Q_h) \subset \hat{H}(L, Q_h)$ is invertible.

Theorem 4.27. *For $(\mathbf{y}, \hat{\mathbf{y}}) \in W \times \hat{V}$, we have*

$$\sup_{\mathbf{z} \in H(L^*, Q_h)} \frac{\langle B_h(\mathbf{y}, \hat{\mathbf{y}}), \mathbf{z} \rangle}{\|\mathbf{z}\|_{L^*, Q_h}} \geq \frac{1}{\sqrt{4C_L^2 + 2}} \|(\mathbf{y}, \hat{\mathbf{y}})\|_{Q;L,\partial Q_h} \quad (4.39)$$

and (4.38) is well-posed for all $\mathbf{b} \in W$.

Proof. We apply THEOREM 2.15 and make use of PROPOSITION 2.22.

In the first step, we establish

$$\sup_{(\mathbf{y}, \hat{\mathbf{y}}) \in W \times \hat{V}} \frac{\langle B_h(\mathbf{y}, \hat{\mathbf{y}}), \mathbf{z} \rangle}{\|(\mathbf{y}, \hat{\mathbf{y}})\|_{Q;L,\partial Q_h}} \geq \frac{1}{\sqrt{4C_L^2 + 2}} \|\mathbf{z}\|_{L^*, Q_h}, \quad \mathbf{z} \in H(L^*, Q_h). \quad (4.40)$$

By THEOREM 3.7, for given $\mathbf{z} \in \mathbf{H}(L^*, Q_h) \subset W$ we find a unique $\mathbf{y}_0 \in V$ with $L\mathbf{y}_0 = \mathbf{z}$, and we set $\hat{\mathbf{y}}_0 = \mathbf{y}_0 + \mathbf{H}_0(L, Q_h) \in \hat{V}$. Then, it holds

$$\begin{aligned} \langle B_h(\mathbf{y}_0, \hat{\mathbf{y}}_0), \mathbf{z} \rangle &= (\mathbf{y}_0, L^* \mathbf{z})_{Q_h} + \langle \hat{D}_h \hat{\mathbf{y}}_0, \mathbf{z} \rangle \\ &= (\mathbf{y}_0, L^* \mathbf{z})_{Q_h} + (L\mathbf{y}_0, \mathbf{z})_{Q_h} - (\mathbf{y}_0, L^* \mathbf{z})_{Q_h} \\ &= (L\mathbf{y}_0, \mathbf{z})_Q = \|\mathbf{z}\|_Q^2, \end{aligned}$$

and exploiting (3.10) yields

$$\begin{aligned} \|(\mathbf{y}_0, \hat{\mathbf{y}}_0)\|_{Q;L,\partial Q_h}^2 &= \|\mathbf{y}_0\|_Q^2 + \|\hat{\mathbf{y}}_0\|_{L,\partial Q_h}^2 \\ &\leq \|\mathbf{y}_0\|_Q^2 + \|\mathbf{y}_0\|_{L,Q}^2 = 2\|\mathbf{y}_0\|_Q^2 + \|L\mathbf{y}_0\|_Q^2 \\ &\leq (2C_L^2 + 1)\|L\mathbf{y}_0\|_Q^2 = (2C_L^2 + 1)\|\mathbf{z}\|_Q^2, \end{aligned}$$

so that we obtain

$$\begin{aligned} \sup_{(\mathbf{y}, \hat{\mathbf{y}}) \in W \times \hat{V}} \frac{\langle B_h(\mathbf{y}, \hat{\mathbf{y}}), \mathbf{z} \rangle}{\|(\mathbf{y}, \hat{\mathbf{y}})\|_{Q;L,\partial Q_h}} &\geq \frac{\langle B_h(\mathbf{y}_0, \hat{\mathbf{y}}_0), \mathbf{z} \rangle}{\|(\mathbf{y}_0, \hat{\mathbf{y}}_0)\|_{Q;L,\partial Q_h}} = \frac{\|\mathbf{z}\|_Q^2}{\|(\mathbf{y}_0, \hat{\mathbf{y}}_0)\|_{Q;L,\partial Q_h}} \\ &\geq \frac{1}{\sqrt{2C_L^2 + 1}} \|\mathbf{z}\|_Q. \end{aligned}$$

Then, choosing $(\mathbf{y}, \hat{\mathbf{y}}) = (L^* \mathbf{z}, \mathbf{0})$ yields

$$\sup_{(\mathbf{y}, \hat{\mathbf{y}}) \in W \times \hat{V}} \frac{\langle B_h(\mathbf{y}, \hat{\mathbf{y}}), \mathbf{z} \rangle}{\|(\mathbf{y}, \hat{\mathbf{y}})\|_{Q;L,\partial Q_h}} \geq \frac{\langle B_h(L^* \mathbf{z}, \mathbf{0}), \mathbf{z} \rangle}{\|(L^* \mathbf{z}, \mathbf{0})\|_{Q;L,\partial Q_h}} = \|L^* \mathbf{z}\|_Q.$$

Now, (4.40) follows from $\|\mathbf{z}\|_{L^*, Q_h}^2 \leq 2 \max \{\|\mathbf{z}\|_Q^2, \|L^* \mathbf{z}\|_Q^2\}$, i.e.,

$$\max \{\|\mathbf{z}\|_Q, \|L^* \mathbf{z}\|_Q\} \geq \frac{1}{\sqrt{2}} \|\mathbf{z}\|_{L^*, Q_h}.$$

In the second step, we show that the operator B_h is injective in $W \times \hat{V}$. Then, the B_h is an isomorphism by THEOREM 2.15 and (4.39) is obtained by duality, see REMARK 2.23.

Therefore, we consider $(\mathbf{y}, \hat{\mathbf{y}}) \in W \times \hat{V}$ with

$$\langle B_h(\mathbf{y}, \hat{\mathbf{y}}), \mathbf{z} \rangle = 0, \quad \mathbf{z} \in \mathbf{H}(L^*, Q_h).$$

This yields

$$0 = \langle B_h(\mathbf{y}, \hat{\mathbf{y}}), \mathbf{z} \rangle = (\mathbf{y}, L^* \mathbf{z})_{Q_h}, \quad \mathbf{z} \in C_c^1(Q_h, \mathbb{R}^m),$$

i.e., $\mathbf{y} \in \mathbf{H}(L, Q_h)$ and $L\mathbf{y} = 0$. Thus, from $\hat{\mathbf{y}} \in \hat{V}$ and $\langle \hat{D}_h \hat{\mathbf{y}}, \mathbf{z} \rangle = 0$ for $\mathbf{z} \in V^*$, see COROLLARY 3.10, we conclude for all $\mathbf{z} \in V^*$

$$\begin{aligned} 0 &= \langle B_h(\mathbf{y}, \hat{\mathbf{y}}), \mathbf{z} \rangle - (L\mathbf{y}, \mathbf{z})_{Q_h} = (\mathbf{y}, L^* \mathbf{z})_{Q_h} + \langle \hat{D}_h \hat{\mathbf{y}}, \mathbf{z} \rangle - (L\mathbf{y}, \mathbf{z})_{Q_h} \\ &= \langle \hat{D}_h \hat{\mathbf{y}}, \mathbf{z} \rangle - \langle D_h \mathbf{y}, \mathbf{z} \rangle = -\langle D_h \mathbf{y}, \mathbf{z} \rangle, \end{aligned}$$

which shows $\mathbf{y} \in V$, cf. LEMMA 3.13. Together with $L\mathbf{y} = \mathbf{0}$ and (3.10) this implies $\mathbf{y} = \mathbf{0}$. Thus,

$$0 = \langle B_h(\mathbf{y}, \hat{\mathbf{y}}), \mathbf{z} \rangle = \langle \hat{D}_h \hat{\mathbf{y}}, \mathbf{z} \rangle, \quad \mathbf{z} \in \mathbf{H}(L^*, Q_h),$$

which finally yields $\hat{\mathbf{y}} = \mathbf{0}$, see LEMMA 3.14. \square

The following proposition provides an upper bound for B_h .

Proposition 4.28. *We have for all $(\mathbf{y}, \hat{\mathbf{y}}) \in W \times \hat{\mathbf{H}}(L, Q_h)$ and $\mathbf{z} \in \mathbf{H}(L^*, Q_h)$*

$$\langle B_h(\mathbf{y}, \hat{\mathbf{y}}), \mathbf{z} \rangle \leq \sqrt{2} \|(\mathbf{y}, \hat{\mathbf{y}})\|_{Q;L,\partial Q_h} \|\mathbf{z}\|_{L^*,Q_h}.$$

Proof. Fix $(\mathbf{y}, \hat{\mathbf{y}}) \in W \times \mathbf{H}(L, Q_h)/\mathbf{H}_0(L, Q_h)$ and $\mathbf{z} \in \mathbf{H}(L^*, Q_h)$. Then for $\bar{\mathbf{y}} \in \hat{\mathbf{y}}$, we have

$$\begin{aligned} \langle B_h(\mathbf{y}, \hat{\mathbf{y}}), \mathbf{z} \rangle &= (\mathbf{y}, L^* \mathbf{z})_Q + (L \bar{\mathbf{y}}, \mathbf{z})_Q - (\bar{\mathbf{y}}, L^* \mathbf{z})_Q \\ &\leq \|\mathbf{y}\|_Q \|L^* \mathbf{z}\|_Q + \|L \bar{\mathbf{y}}\|_Q \|\mathbf{z}\|_Q + \|\bar{\mathbf{y}}\|_Q \|L^* \mathbf{z}\|_Q \\ &\leq \sqrt{\|\mathbf{y}\|_Q^2 + \|L \bar{\mathbf{y}}\|_Q^2 + \|\bar{\mathbf{y}}\|_Q^2} \sqrt{2\|L^* \mathbf{z}\|_Q^2 + \|\mathbf{z}\|_Q^2} \\ &\leq \sqrt{2} \|(\mathbf{y}, \bar{\mathbf{y}})\|_{Q;L,Q_h} \|\mathbf{z}\|_{L^*,Q_h}. \end{aligned}$$

Taking the infimum for $\bar{\mathbf{y}} \in \hat{\mathbf{y}}$ yields the assertion. \square

Now, we obtain the convergence result for the ideal DPG method as follows. For a fixed discrete approximation space $W_h \times \hat{V}_h \subset W \times \hat{V}$, we choose the optimal test space $Z^{\text{opt}} = \Pi_{\mathbf{H}(L^*, Q_h)}^{-1} B_h(W_h \times \hat{V})$, see (4.34). Then, the continuous problem (4.38) and its discrete counterpart

$$\begin{cases} \text{find } (\mathbf{y}_h, \hat{\mathbf{y}}_h) \in W_h \times \hat{V}_h \text{ with} \\ \langle B_h(\mathbf{y}_h, \hat{\mathbf{y}}_h), \mathbf{z}_h \rangle = (\mathbf{b}, \mathbf{z}_h)_Q, \quad \mathbf{z}_h \in Z^{\text{opt}} \end{cases}$$

are well-posed by THEOREM 4.27 with inf-sup constant $\beta_h = \beta = (4C_L + 2)^{-\frac{1}{2}}$.

Exploiting PROPOSITION 4.28, REMARK 2.32 yields the quasi best approximation result

$$\|(\mathbf{y}_h^{\text{sol}}, \hat{\mathbf{y}}_h^{\text{sol}}) - (\mathbf{y}^{\text{sol}}, \hat{\mathbf{y}}^{\text{sol}})\|_{Q;L,\partial Q_h} \leq \frac{\sqrt{2}}{\beta_h} \inf_{(\mathbf{y}_h, \hat{\mathbf{y}}_h) \in W_h \times \hat{V}_h} \|(\mathbf{y}_h, \hat{\mathbf{y}}_h) - (\mathbf{y}^{\text{sol}}, \hat{\mathbf{y}}^{\text{sol}})\|_{Q;L,\partial Q_h}$$

Note that the optimal test space is not computationally accessible. Thus, in the following, we replace Z_h^{opt} by a discrete approximation yielding the practical DPG method.

4.5 Acoustic waves – the practical DPG method

Now we select a globally conforming discrete ansatz space $\hat{V}_h \subset \hat{V}$ on the skeleton and local ansatz and test spaces $W_{R,h} \subset L_2(R, \mathbb{R}^m)$ and $Z_{R,h} \subset H(L^*, R)$. We set $\hat{V}_{R,h} = \hat{V}_h|_{\partial R}$, $W_h = \prod_{R \in \mathcal{R}_h} W_{R,h}$ and $Z_h = \prod_{R \in \mathcal{R}_h} Z_{R,h}$.

For the practical DPG, the optimal test space Z^{opt} is replaced by the approximation $Z_h^{\text{opt}} = C_h^{-1} B_h(W_h \times \hat{V}_h)$, where C_h^{-1} is an approximation of the RIESZ operator in the test space Z . Thus, it is no longer guaranteed that the discrete stability constant equals the continuous constant β_h . In order to analyze this loss of stability, we construct a suitable local FORTIN operator $\Pi_{R,h} \in \mathcal{L}(H(L^*, R), Z_{R,h})$, see LEMMA 2.26, in every space-time cell R following the approach presented in [21, Sect. 3.1.4], see also the construction in [25, Thm. 1]. This yields a mesh-dependent estimate. Then, we show by scaling argument that it is sufficient to construct a local FORTIN operator on a reference cell, so that finally a mesh-independent a-priori bound for the DPG approximation is obtained.

4.5.1 Local construction of the FORTIN operator

We define $B_R \in \mathcal{L}(L_2(R, \mathbb{R}^m) \times \hat{H}(L, R), H(L^*, R)')$ by

$$\langle B_R(\mathbf{y}_R, \hat{\mathbf{y}}_R), \mathbf{z}_R \rangle = (\mathbf{y}_R, L^* \mathbf{z}_R)_R + \langle \hat{D}_R \hat{\mathbf{y}}_R, \mathbf{z}_R \rangle.$$

We assume that for given $\hat{V}_{R,h}$ and $W_{R,h}$ the local test spaces $Z_{R,h}$ are large enough, so that for all $\mathbf{z}_R \in H(L^*, R)$ the affine space

$$\mathcal{N}(\mathbf{z}_R) = \left\{ \mathbf{z}_{R,h} \in Z_{R,h} : \langle B_R(\mathbf{y}_{R,h}, \hat{\mathbf{y}}_{R,h}), \mathbf{z}_{R,h} \rangle = \langle B_R(\mathbf{y}_{R,h}, \hat{\mathbf{y}}_{R,h}), \mathbf{z}_R \rangle, \right. \\ \left. (\mathbf{y}_{R,h}, \hat{\mathbf{y}}_{R,h}) \in W_{R,h} \times \hat{V}_{R,h} \right\}$$

is not empty, cf. (2.5). Then, a FORTIN operator with $\Pi_{R,h} \mathbf{z}_R \in \mathcal{N}(\mathbf{z}_R)$ exists. For the scaling argument below we require the additional property

$$|\Pi_{R,h} \mathbf{z}_R|_{L^*, R} \leq |\mathbf{z}_R|_{L^*, R}, \quad \mathbf{z}_R \in H(L^*, R),$$

with respect to the semi-norm $|\mathbf{z}_R|_{L^*, R} = \|L^* \mathbf{z}_R\|_R$. This can easily be achieved by extending $W_{R,h}$ to $W_{R,h}^{\text{ext}} \supset W_{R,h} + L^*(Z_{R,h})$, since the orthogonality

$$0 = \langle B_R(\mathbf{y}_{R,h}, \mathbf{0}), \mathbf{z}_{R,h} - \mathbf{z}_R \rangle = (\mathbf{y}_{R,h}, L^*(\mathbf{z}_{R,h} - \mathbf{z}_R))_R, \quad \mathbf{y}_{R,h} \in W_{R,h}^{\text{ext}}$$

implies $|\mathbf{z}_{R,h}|_{L^*, R} \leq |\mathbf{z}_R|_{L^*, R}$ by choosing $\mathbf{y}_{R,h} = L^* \mathbf{z}_{R,h}$ and

$$|\mathbf{z}_{R,h}|_{L^*, R}^2 = (L^* \mathbf{z}_{R,h}, L^* \mathbf{z}_{R,h})_R = (L^* \mathbf{z}_{R,h}, L^* \mathbf{z}_R)_R \leq |\mathbf{z}_{R,h}|_{L^*, R} |\mathbf{z}_R|_{L^*, R}.$$

We assume that also $\mathcal{N}^{\text{ext}}(\mathbf{z}_R) \subset \mathcal{N}(\mathbf{z}_R)$ obtained by testing with the larger space $W_{R,h}^{\text{ext}} \supset W_{R,h}$ is not empty.

In order to compute a bound for the norm of $\Pi_{R,h}$ numerically, we assume that extensions $V_{R,h} \subset \mathbb{H}(L, R)$ of $\hat{V}_{R,h}$ exist with $\dim V_{R,h} = \dim \hat{V}_{R,h}$, so that for every trace function $\hat{\mathbf{y}}_{R,h} \in \hat{V}_{R,h}$ a unique extension $\bar{\mathbf{y}}_{R,h} \in V_{R,h}$ exists which can be locally evaluated in R and which satisfies

$$\langle D_R \bar{\mathbf{y}}_{R,h}, \mathbf{z}_R \rangle = \langle \hat{D}_R \hat{\mathbf{y}}_{R,h}, \mathbf{z}_R \rangle, \quad \mathbf{z}_R \in \mathbb{H}(L^*, R), \quad (4.41)$$

i.e., $\hat{\mathbf{y}}_{R,h} = \bar{\mathbf{y}}_{R,h} + \mathbb{H}_0(L, Q_h)$. This defines a well-defined bijection

$$\hat{I}_{R,h}: V_{R,h} \longrightarrow \hat{V}_{R,h}$$

such that $\hat{\mathbf{y}}_{R,h} = \hat{I}_{R,h} \bar{\mathbf{y}}_{R,h}$ satisfies (4.41). As a result, we have

$$\mathcal{N}^{\text{ext}}(\mathbf{z}_R) = \left\{ \mathbf{z}_{R,h} \in Z_{R,h}: \langle B_R(\mathbf{y}_{R,h}, \hat{I}_{R,h} \bar{\mathbf{y}}_{R,h}), \mathbf{z}_{R,h} \rangle = \langle B_R(\mathbf{y}_{R,h}, \hat{I}_{R,h} \bar{\mathbf{y}}_{R,h}), \mathbf{z}_R \rangle, \right. \\ \left. (\mathbf{y}_{R,h}, \hat{\mathbf{y}}_{R,h}) \in W_{R,h} \times V_{R,h} \right\}$$

The minimizer $\mathbf{z}_{R,h} = \Pi_{R,h} \mathbf{z}_R \in \mathcal{N}^{\text{ext}}(\mathbf{z}_R)$ with respect to the norm in $\mathbb{H}(L^*, R)$ can be computed by a discrete linear saddle point problem as follows. We define the discrete operators

$$B_{R,h} \in \mathcal{L}(W_{R,h}^{\text{ext}} \times V_{R,h}, Z'_{R,h}), \\ C_{R,h} \in \mathcal{L}(Z_{R,h}, Z'_{R,h}), \\ G_{R,h} \in \mathcal{L}(W_{R,h}^{\text{ext}} \times V_{R,h}, (W_{R,h}^{\text{ext}} \times V_{R,h})')$$

by

$$\langle B_{R,h}(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}), \mathbf{z}_{R,h} \rangle = \langle B_R(\mathbf{y}_{R,h}, \hat{I}_{R,h} \bar{\mathbf{y}}_{R,h}), \mathbf{z}_{R,h} \rangle, \\ \langle C_{R,h} \mathbf{z}_{R,h}, \boldsymbol{\psi}_{R,h} \rangle = (\mathbf{z}_{R,h}, \boldsymbol{\psi}_{R,h})_{L^*, R}, \\ \langle G_{R,h}(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}), (\boldsymbol{\phi}_{R,h}, \bar{\boldsymbol{\phi}}_{R,h}) \rangle = (\mathbf{y}_{R,h}, \boldsymbol{\phi}_{R,h})_R + (\bar{\mathbf{y}}_{R,h}, \bar{\boldsymbol{\phi}}_{R,h})_{L,R}, \quad (4.42)$$

and the embedding $E_{R,h} \in \mathcal{L}(W_{R,h}^{\text{ext}} \times V_{R,h}, W \times \hat{\mathbb{H}}(L, R))$ given by

$$E_{R,h}(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}) = (\mathbf{y}_{R,h}, \hat{I}_{R,h} \bar{\mathbf{y}}_{R,h}), \quad (\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}) \in W_{R,h}^{\text{ext}} \times V_{R,h}. \quad (4.43)$$

Then, $\mathbf{z}_{R,h} = \Pi_{R,h} \mathbf{z}_R$ solves the discrete saddle point problem

$$C_{R,h} \mathbf{z}_{R,h} + B_{R,h}(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}) = \mathbf{0}, \\ B'_{R,h} \mathbf{z}_{R,h} = E'_{R,h} B'_R \mathbf{z}_R, \quad (4.44)$$

where $(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}) \in W_{R,h}^{\text{ext}} \times V_{R,h}$ is the LAGRANGE multiplier.

Remark 4.29. *Inf-sup stability requires that the operator B_h is injective in $W_h \times \hat{V}_h$. However, locally we cannot expect that $B_{R,h}$ is injective, since $B_{R,h}(\mathbf{y}_{R,h}, -\bar{\mathbf{y}}_{R,h}) = \mathbf{0}$ for all functions $\mathbf{y}_{R,h} \in W_{R,h} \cap \mathcal{N}(L)$ and $\bar{\mathbf{y}}_{R,h} \in V_{R,h}$ having the same traces, i.e. $\mathbf{y}_{R,h} \in \bar{\mathbf{y}}_{R,h} + \mathbf{H}_0(L, Q_h)$. Thus, as soon as both, $W_{R,h}$ and $\hat{V}_{R,h}$, contain constant functions, we have $\mathcal{N}(B_{R,h}) \neq \{0\}$.*

On the other hand, since we assume that $\mathcal{N}^{\text{ext}}(\mathbf{z}_R)$ is not empty for all \mathbf{z}_R , (4.44) always has a solution, and since $C_{R,h}$ is positive definite, $\mathbf{z}_{R,h} = \Pi_{R,h}\mathbf{z}_R$ is the unique solution of the optimization problem. The Lagrange parameter $(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h})$ is only unique up to $\mathcal{N}(B_{R,h})$.

Inserting $\mathbf{z}_{R,h} = -C_{R,h}^{-1}B_{R,h}(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h})$ yields

$$S_{R,h}(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}) = -E'_{R,h}B'_R\mathbf{z}_R$$

with the SCHUR complement operator

$$S_{R,h} = B'_{R,h}C_{R,h}^{-1}B_{R,h} \in \mathcal{L}(W_{R,h}^{\text{ext}} \times V_{R,h}, (W_{R,h}^{\text{ext}} \times V_{R,h})').$$

Inserting the pseudo-inverse (with respect to the inner product in $W_{R,h}^{\text{ext}} \times V_{R,h}$)

$$S_{R,h}^+ = \lim_{\delta \rightarrow 0} (S_{R,h}G_{R,h}^{-1}S_{R,h} + \delta G_{R,h})^{-1}S_{R,h}G_{R,h}^{-1},$$

see [2, Thm. (3.4)], satisfying $S_{R,h}^+S_{R,h}S_{R,h}^+ = S_{R,h}^+$, defines

$$\Pi_{R,h} = C_{R,h}^{-1}B_{R,h}S_{R,h}^+E'_{R,h}B'_R. \quad (4.45)$$

We compute $\alpha_{R,h} > 0$ such that

$$\langle \ell_{R,h}, S_{R,h}^+\ell_{R,h} \rangle \leq \alpha_{R,h} \langle \ell_{R,h}, G_{R,h}^{-1}\ell_{R,h} \rangle, \quad \ell_{R,h} \in (W_{R,h}^{\text{ext}} \times V_{R,h})' \quad (4.46)$$

i.e., we determine the largest eigenvalue of a finite dimensional symmetric generalized eigenvalue problem, see APPENDIX A. For given $\mathbf{z}_R \in \mathbf{H}(L^*, R)$ we select the discrete functional $\ell_{R,h} = E'_{R,h}B'_R\mathbf{z}_R$, and the norm of the FORTIN operator is estimated by

$$\begin{aligned} \|\Pi_{R,h}\mathbf{z}_R\|_{L^*,R}^2 &= \langle B_{R,h}S_{R,h}^+\ell_{R,h}, C_{R,h}^{-1}B_{R,h}S_{R,h}^+\ell_{R,h} \rangle \\ &= \langle \ell_{R,h}, S_{R,h}^+S_{R,h}S_{R,h}^+\ell_{R,h} \rangle \\ &= \langle \ell_{R,h}, S_{R,h}^+\ell_{R,h} \rangle \\ &\leq \alpha_{R,h} \langle \ell_{R,h}, G_{R,h}^{-1}\ell_{R,h} \rangle \\ &\leq 2\alpha_{R,h} \|\mathbf{z}_R\|_{L^*,R}^2 \end{aligned}$$

using

$$\begin{aligned}
 \sqrt{\langle \ell_{R,h}, G_{R,h}^{-1} \ell_{R,h} \rangle} &\leq \sup_{(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}) \in W_{R,h}^{\text{ext}} \times V_{R,h}} \frac{\langle \ell_{R,h}, (\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}) \rangle}{\sqrt{\langle G_{R,h}(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}), (\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}) \rangle}} \\
 &= \sup_{(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}) \in W_{R,h}^{\text{ext}} \times V_{R,h}} \frac{\langle E'_{R,h} B'_R \mathbf{z}_R, (\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}) \rangle}{\|(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h})\|_{R;L,R}} \\
 &= \sup_{(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}) \in W_{R,h}^{\text{ext}} \times V_{R,h}} \frac{\langle B_R E_{R,h}(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}), \mathbf{z}_R \rangle}{\|(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h})\|_{R;L,R}} \leq \sqrt{2} \|\mathbf{z}_R\|_{L^*,R}
 \end{aligned}$$

with $\langle G_{R,h}(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}), (\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h}) \rangle = \|(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h})\|_{R;L,R}^2$ and

$$\begin{aligned}
 \langle B_R(\mathbf{y}_{R,h}, \hat{I}_{R,h} \bar{\mathbf{y}}_{R,h}), \mathbf{z}_R \rangle &= (\mathbf{y}_{R,h}, L^* \mathbf{z}_R)_R + (L \bar{\mathbf{y}}_{R,h}, \mathbf{z}_R)_R - (\bar{\mathbf{y}}_{R,h}, L^* \mathbf{z}_R)_R \\
 &\leq \|\mathbf{y}_{R,h}\|_R \|L^* \mathbf{z}_R\|_R + \|L \bar{\mathbf{y}}_{R,h}\|_R \|\mathbf{z}_R\|_R + \|\bar{\mathbf{y}}_{R,h}\|_R \|L^* \mathbf{z}_R\|_R \\
 &\leq \sqrt{\|\mathbf{y}_{R,h}\|_R^2 + \|L \bar{\mathbf{y}}_{R,h}\|_R^2 + \|\bar{\mathbf{y}}_{R,h}\|_R^2} \sqrt{2\|L^* \mathbf{z}_R\|_R^2 + \|\mathbf{z}_R\|_R^2} \\
 &\leq \sqrt{2} \|(\mathbf{y}_{R,h}, \bar{\mathbf{y}}_{R,h})\|_{R;L,R} \|\mathbf{z}_R\|_{L^*,R}.
 \end{aligned}$$

using (4.43). The construction is completely local, so that it extends to

$$\|\Pi_h \mathbf{z}\|_{L^*, Q_h} \leq \sqrt{2\alpha_h} \|\mathbf{z}\|_{L^*, Q_h}, \quad \mathbf{z} \in \mathbb{H}(L^*, Q_h) \quad (4.47)$$

with $\alpha_h = \max \alpha_{R,h}$. Since the continuous problem is inf-sup stable, cf. THEOREM 4.27, this implies discrete inf-sup stability using the FORTIN criterion, see LEMMA 2.26. As a result, we have

$$\sup_{\mathbf{z}_h \in Z_h} \frac{\langle B_h(\mathbf{y}_h, \hat{\mathbf{y}}_h), \mathbf{z}_h \rangle}{\|\mathbf{z}_h\|_{L^*, Q_h}} \geq \beta_h \|(\mathbf{y}_h, \hat{\mathbf{y}}_h)\|_{Q;L,\partial Q_h} \quad (4.48)$$

for $(\mathbf{y}_h, \hat{\mathbf{y}}_h) \in W_h \times \hat{V}_h$ with $\beta_h = \frac{1}{\sqrt{2\alpha_h} \sqrt{4C_L^2 + 2}}$, see LEMMA 2.26.

4.5.2 A scaling argument

Numerically, we observe that the eigenvalue estimate (4.46) is mesh-dependent. Thus, we compute $\alpha_0 = \alpha_{R_0, h_0}$ on a reference element $R_0 = (0, h_0)^d \times (0, ch_0)$, and we analyze the transformation $\varphi_R: R_0 \rightarrow R \in \mathcal{R}_h$. For simplicity, we only discuss an affine transformation of the form $\varphi_R(\mathbf{x}, t) = (\mathbf{x}_R, t_R) + (h/h_0)(\mathbf{x}, t)$ with $R = (\mathbf{x}_R, t_R) + (0, h)^d \times (0, ch)$.

Let Π_{R_0, h_0} be a local FORTIN operator on the reference cell R_0 . For the space-time L_2 norm, the semi-norm $|\mathbf{z}_R|_{L^*, R} = \|L^* \mathbf{z}_R\|_R$, and the operator B_R , we assume the scaling properties

$$\begin{aligned} h^{-d-1} \|\mathbf{z}_R\|_R^2 &= h_0^{-d-1} \|\mathbf{z}_R \circ \varphi_R\|_{R_0}^2, \\ h^{-d+1} |\mathbf{z}_R|_{L^*, R}^2 &= h_0^{-d+1} |\mathbf{z}_R \circ \varphi_R|_{L^*, R_0}^2, \\ h^{-d} \langle B_R(\mathbf{y}_{R, h}, \hat{\mathbf{y}}_{R, h}), \mathbf{z}_R \rangle &= h_0^{-d} \langle B_{R_0}(\mathbf{y}_{R, h} \circ \varphi_R, \hat{\mathbf{y}}_{R, h} \circ \varphi_R), \mathbf{z}_R \circ \varphi_R \rangle. \end{aligned}$$

By the integral transformation formula, this holds for acoustic waves with constant coefficients. Then, the transformed operator

$$\Pi_{R, h} \mathbf{z}_R = \left(\Pi_{R_0, h_0}(\mathbf{z}_R \circ \varphi_R) \right) \circ \varphi_R^{-1}, \quad \mathbf{z}_R \in \mathbb{H}(L^*, R),$$

defines a local FORTIN operator in R . By scaling we obtain for $h \leq h_0$

$$\begin{aligned} h^{-d-1} \|\Pi_{R, h} \mathbf{z}_R\|_R^2 &= h_0^{-d-1} \|(\Pi_{R, h} \mathbf{z}_R) \circ \varphi_R\|_{R_0}^2 \\ &= h_0^{-d-1} \|\Pi_{R_0, h_0}(\mathbf{z}_R \circ \varphi_R)\|_{R_0}^2 \\ &\leq h_0^{-d-1} \|\Pi_{R_0, h_0}(\mathbf{z}_R \circ \varphi_R)\|_{L^*, R_0}^2 \\ &\leq h_0^{-d-1} \|\Pi_{R_0, h_0}\|_{L^*, R_0}^2 \|\mathbf{z}_R \circ \varphi_R\|_{L^*, R_0}^2, \\ h_0^{-d-1} \|\mathbf{z}_R \circ \varphi_R\|_{L^*, R_0}^2 &= h^{-d-1} \|\mathbf{z}_R\|_R^2 + h_0^{-2} h^{-d+1} |\mathbf{z}_R|_{L^*, R}^2 \\ &\leq h^{-d-1} \|\mathbf{z}_R\|_{L^*, R}^2, \\ h^{-d+1} |\Pi_{R, h} \mathbf{z}_R|_{L^*, R}^2 &= h_0^{-d+1} |(\Pi_{R, h} \mathbf{z}_R) \circ \varphi_R|_{L^*, R_0}^2 \\ &= h_0^{-d+1} |\Pi_{R_0, h_0}(\mathbf{z}_R \circ \varphi_R)|_{L^*, R_0}^2 \\ &\leq h_0^{-d+1} |\mathbf{z}_R \circ \varphi_R|_{L^*, R_0}^2 = h^{-d+1} |\mathbf{z}_R|_{L^*, R}^2, \end{aligned}$$

which together yield

$$\|\Pi_{R, h} \mathbf{z}_R\|_{L^*, R_0} \leq \sqrt{1 + \|\Pi_{R_0, h_0}\|_{L^*, R_0}^2} \|\mathbf{z}_R\|_{L^*, R}.$$

For simple meshes, this results into the computable inf-sup constant

$$\beta_h = \frac{1}{\sqrt{1 + 2\alpha_{R_0, h_0}} \sqrt{4C_L^2 + 2}}. \quad (4.49)$$

4.5.3 An a-priori error estimate for the practical DPG method

To obtain the discrete solution corresponding to the practical DPG method, we replace the optimal test space $Z^{\text{opt}} = \Pi_{\mathbb{H}(L^*, Q_h)}^{-1} B_h(W_h \times \hat{V})$ by the approximation $Z_h^{\text{opt}} = C_h^{-1} B_h(W_h \times \hat{V}_h)$. Here,

$$C_h \in \mathcal{L}(Z_h, Z_h'), \quad \langle C_h \mathbf{z}_h, \tilde{\mathbf{z}}_h \rangle = \sum_{R \in \mathcal{R}_h} \langle C_{R,h} \mathbf{z}_{R,h}, \tilde{\mathbf{z}}_{R,h} \rangle$$

is the approximate RIESZ isomorphism in $Z_h = \prod_{R \in \mathcal{R}_h} Z_{R,h}$ with $C_{R,h}$ as in (4.42).

Then, the approximate solution $(\mathbf{y}_h^{\text{sol}}, \hat{\mathbf{y}}_h^{\text{sol}}) \in W_h \times \hat{V}_h$ is defined by

$$\langle B_h(\mathbf{y}_h^{\text{sol}}, \hat{\mathbf{y}}_h^{\text{sol}}), \mathbf{z}_h \rangle = (\mathbf{b}, \mathbf{z}_h)_Q, \quad \mathbf{z}_h \in Z_h^{\text{opt}}. \quad (4.50)$$

Since B_h is continuous and since we assume that Z_h is large enough (so that a computable but in general mesh dependent inf-sup constant exists), PETROV-GALERKIN estimates apply. In simple cases where the scaling argument applies, this yields a mesh-independent estimate for α_h and thus for the inf-sup constant β_h in (4.49). Summarizing our results, we obtain

Theorem 4.30. *Let $\mathbf{y}^{\text{sol}} \in V$ be the solution of $L\mathbf{y} = \mathbf{b}$ and define its trace by $\hat{\mathbf{y}}^{\text{sol}} = \mathbf{y}^{\text{sol}} + \mathbb{H}_0(L, Q_h) \in \hat{V}$. If a FORTIN operator can be constructed and bounded by (4.47), a unique solution $(\mathbf{y}_h^{\text{sol}}, \hat{\mathbf{y}}_h^{\text{sol}}) \in W_h \times \hat{V}_h$ of (4.50) exists and satisfies the a-priori error estimate*

$$\|(\mathbf{y}^{\text{sol}} - \mathbf{y}_h^{\text{sol}}, \hat{\mathbf{y}}^{\text{sol}} - \hat{\mathbf{y}}_h^{\text{sol}})\|_{Q;L,\partial Q_h} \leq \frac{\sqrt{2}}{\beta_h} \inf_{(\phi_h, \hat{\phi}_h) \in W_h \times \hat{V}_h} \|(\mathbf{y}^{\text{sol}} - \phi_h, \hat{\mathbf{y}}^{\text{sol}} - \hat{\phi}_h)\|_{Q;L,\partial Q_h}.$$

Proof. Apply THEOREM 2.30 and REMARK 2.32 using THEOREM 4.27, (4.49) and PROPOSITION 4.28. \square

4.6 Acoustic waves – the simplified DPG method

For the realization of the practical DPG method it is advantageous to use traces on the skeleton ∂Q_h . This process depends on the application and is now described for linear acoustic waves. For space-time tensor-product decompositions with space-time cells $R = K \times (a, b) \subset \Omega \times (0, T)$, we define a trace mapping $I_{\partial R}$ to $L_2(\partial R; \mathbb{R} \times \mathbb{R}^d)$ by

$$I_{\partial R}(p_R, \mathbf{v}_R) = \begin{cases} (p_R, \mathbf{v}_R)|_{K \times \{t\}} & \text{for traces at time } t \in \{a, b\}, \\ (p_R, (\mathbf{v}_R \cdot \mathbf{n}_F) \mathbf{n}_F)|_{F \times (a,b)} & \text{in space with } F \subset \partial K \end{cases}$$

for all sufficiently smooth functions (p_R, \mathbf{v}_R) . We define local the trace bilinear form obtained using integration by parts, cf. (4.8),

$$\begin{aligned} \gamma_R((\tilde{p}_R, \tilde{\mathbf{v}}_R), (q_R, \mathbf{w}_R)) &= ((\tilde{p}_R, \tilde{\mathbf{v}}_R), (q_R, \mathbf{w}_R))_{K \times \{b\}} - ((\tilde{p}_R, \tilde{\mathbf{v}}_R), (q_R, \mathbf{w}_R))_{K \times \{a\}} \\ &\quad + \sum_{F \subset \partial K} (\tilde{p}_R, \mathbf{w}_R \cdot \mathbf{n}_K)_{F \times (a,b)} + (\tilde{\mathbf{v}}_R \cdot \mathbf{n}_K, q_R)_{F \times (a,b)} \end{aligned}$$

for $(\tilde{p}_R, \tilde{\mathbf{v}}_R) \in L_2(\partial R; \mathbb{R} \times \mathbb{R}^d)$ and $(q_R, \mathbf{w}_R) \in H(L^*, R)$ sufficiently smooth with $I_{\partial R}(q_R, \mathbf{w}_R) \in L_2(\partial R; \mathbb{R} \times \mathbb{R}^d)$. Further, we define

$$b_h(((p, \mathbf{v}), (\tilde{p}, \tilde{\mathbf{v}})), (q, \mathbf{w})) = ((p, \mathbf{v}), L^*(q, \mathbf{w}))_{Q_h} + \gamma_h((\tilde{p}, \tilde{\mathbf{v}}), (q, \mathbf{w}))$$

for $(p, \mathbf{v}) \in L_2(Q; \mathbb{R} \times \mathbb{R}^d)$, $(\tilde{p}, \tilde{\mathbf{v}}) \in L_2(\partial Q_h; \mathbb{R} \times \mathbb{R}^d)$ and for $(q, \mathbf{w}) \in H(L^*, Q_h)$ with traces in L_2 , where $\gamma_h((\tilde{p}, \tilde{\mathbf{v}}), (q, \mathbf{w})) = \sum_{R \in \mathcal{R}_h} \gamma_R((\tilde{p}_R, \tilde{\mathbf{v}}_R), (q_R, \mathbf{w}_R))$.

By construction, we observe

$$\gamma_R(I_{\partial R}(p_R, \mathbf{v}_R), (q_R, \mathbf{w}_R)) = \langle D_R(p_R, \mathbf{v}_R), (q_R, \mathbf{w}_R) \rangle \quad (4.51)$$

for sufficiently smooth $(p_R, \mathbf{v}_R) \in H(L, R)$ and $(q_R, \mathbf{w}_R) \in H(L^*, R)$ both having traces in L_2 , and

$$b_h(((p, \mathbf{v}), I_{\partial Q_h}(\bar{p}, \bar{\mathbf{v}})), (q, \mathbf{w})) = \langle B_h((p, \mathbf{v}), (\bar{p}, \bar{\mathbf{v}}) + H(L, Q_h)), (q, \mathbf{w}) \rangle$$

for $(p, \mathbf{v}) \in L_2(Q; \mathbb{R} \times \mathbb{R}^d)$, and for $(\bar{p}, \bar{\mathbf{v}}) \in H(L, Q_h)$, and $(q, \mathbf{w}) \in H(L^*, Q_h)$ with traces in L_2 .

Thus, in the realization of the DPG method we can replace the operator B_h by the bilinear form $b_h(\cdot, \cdot)$, so that it is sufficient to represent \hat{V}_h by its trace values on ∂Q_h .

In the simplified DPG method, we select independently polynomial ansatz spaces for the traces on every space-time face of the skeleton ∂Q_h , i.e., we choose a discontinuous space

$$\tilde{V}_h = \prod_{K \times \{a\} \subset \partial Q_h} V_{K \times \{a\}, h} \times \prod_{F \times (a,b) \subset \partial Q_h} V_{F \times (a,b), h} \subset L_2(\partial Q_h; \mathbb{R} \times \mathbb{R}^d).$$

The representation of Neumann traces for $(\tilde{p}_h, \tilde{\mathbf{v}}_h) \in \tilde{V}_h$ requires to select an orientation $\mathbf{n}_F \in \{\pm \mathbf{n}_K\}$. Then, $\tilde{\mathbf{v}}_h|_{F \times (a,b)} = \tilde{v}_h \mathbf{n}_F$ with $\tilde{v}_h \in L_2(F \times (a, b))$.

In case that \tilde{V}_h is the trace of a conforming subspace $V_h \subset V$, i.e., $\tilde{V}_h = I_{\partial Q_h} V_h$, the simplified method coincides with a conforming DPG method. In general, the skeleton space \tilde{V}_h may be nonconforming. Then, we assume a weaker condition which is described in the following.

In order to obtain a well-defined method and to provide an a-priori error analysis, we assume that a conforming reconstruction $V_h \subset V$ of \tilde{V}_h exists such that for given $(\tilde{p}_h, \tilde{\mathbf{v}}_h) \in \tilde{V}_h$ the reconstruction $(\bar{p}_h, \bar{\mathbf{v}}_h) \in V_h$ is uniquely defined by

$$\gamma_R((\tilde{p}_{R,h}, \tilde{\mathbf{v}}_{R,h}), (q_{R,h}, \mathbf{w}_{R,h})) = \gamma_R(I_{\partial R,h}(\bar{p}_h, \bar{\mathbf{v}}_h), (q_{R,h}, \mathbf{w}_{R,h})) \quad (4.52)$$

for all $(q_{R,h}, \mathbf{w}_{R,h}) \in Z_{R,h}$ and all space-time cells $R \in \mathcal{R}_h$. In particular, this implies $\dim V_h = \dim \tilde{V}_h$. Note that the traces in V_h only coincide with functions \tilde{V}_h when tested with the finite dimensional space $Z_{R,h}$. See SECTION 4.6.1 for an example of this construction.

Then, by construction, the simplified method with ansatz space $W_h \times \tilde{V}_h$ and test space Z_h yields the same discrete linear system as the practical method with \tilde{V}_h replaced by $\hat{V}_h = V_h/H_0(L, Q_h)$. For the error analysis we introduce the discrete norm

$$\|(\tilde{p}_h, \tilde{\mathbf{v}}_h)\|_{Z'_h} := \sup_{(q_h, \mathbf{w}_h) \in Z_h} \frac{\gamma_h((\tilde{p}_h, \tilde{\mathbf{v}}_h), (q_h, \mathbf{w}_h))}{\|(q_h, \mathbf{w}_h)\|_{L^*, Q_h}}, \quad (\tilde{p}_h, \tilde{\mathbf{v}}_h) \in \tilde{V}_h. \quad (4.53)$$

This extends to a (mesh-dependent) semi-norm in $L_2(\partial Q_h; \mathbb{R} \times \mathbb{R}^d)$. Further, we obtain the following

Lemma 4.31. *For $(p, \mathbf{v}) \in V$ with trace $(\tilde{p}, \tilde{\mathbf{v}}) = I_{\partial Q_h}(p, \mathbf{v}) \in L_2(\partial Q_h; \mathbb{R} \times \mathbb{R}^d)$ and $(\hat{p}, \hat{\mathbf{v}}) = (p, \mathbf{v}) + H_0(L, Q_h) \in \hat{V}$, we have $\|(\tilde{p}, \tilde{\mathbf{v}})\|_{Z'_h} \leq \|(\hat{p}, \hat{\mathbf{v}})\|_{L, \partial Q_h}$.*

Proof. It holds using (4.51) and (4.52)

$$\begin{aligned} \|(\tilde{p}, \tilde{\mathbf{v}})\|_{Z'_h} &= \sup_{(q_h, \mathbf{w}_h) \in Z_h} \frac{\gamma_h((\tilde{p}, \tilde{\mathbf{v}}), (q_h, \mathbf{w}_h))}{\|(q_h, \mathbf{w}_h)\|_{L^*, Q_h}} \\ &= \sup_{(q_h, \mathbf{w}_h) \in Z_h} \inf_{(p_0, \mathbf{v}_0) \in H_0(L, Q_h)} \frac{\langle D_h(p + p_0, \mathbf{v} + \mathbf{v}_0), (q_h, \mathbf{w}_h) \rangle}{\|(q_h, \mathbf{w}_h)\|_{L^*, Q_h}} \\ &\leq \sup_{(q, \mathbf{w}) \in H(L^*, Q_h)} \inf_{(p_0, \mathbf{v}_0) \in H_0(L, Q_h)} \frac{\langle D_h(p + p_0, \mathbf{v} + \mathbf{v}_0), (q, \mathbf{w}) \rangle}{\|(q, \mathbf{w})\|_{L^*, Q_h}} \\ &\leq \inf_{(p_0, \mathbf{v}_0) \in H_0(L, Q_h)} \|(p + p_0, \mathbf{v} + \mathbf{v}_0)\|_{L, Q_h} = \|(\hat{p}, \hat{\mathbf{v}})\|_{L, \partial Q_h}. \quad \square \end{aligned}$$

With respect to the semi-norm (4.53), we can transfer the result in THEOREM 4.30 to the simplified DPG method.

Theorem 4.32. *Assume that a conforming reconstruction $V_h \subset V$ of \tilde{V}_h exists satisfying (4.52) and $\dim V_h = \dim \tilde{V}_h$.*

1. *If a FORTIN operator can be constructed and bounded by (4.47), a unique PETROV-GALERKIN approximation $((p_h, \mathbf{v}_h), (\tilde{p}_h, \tilde{\mathbf{v}}_h)) \in W_h \times \tilde{V}_h$ exists solving*

$$b_h(((p_h, \mathbf{v}_h), (\tilde{p}_h, \tilde{\mathbf{v}}_h)), (q_h, \mathbf{w}_h)) = ((f, \mathbf{g}), (q_h, \mathbf{w}_h))_Q, \quad (q_h, \mathbf{w}_h) \in Z_h^{\text{opt}}. \quad (4.54)$$

2. Let $(p, \mathbf{v}) \in V$ be the solution of (3.2), and assume that (p, \mathbf{v}) is sufficiently regular with traces $(\tilde{p}, \tilde{\mathbf{v}}) = I_{\partial Q_h}(p, \mathbf{v}) \in L_2(\partial Q_h; \mathbb{R} \times \mathbb{R}^d)$.

Then, the error can be bounded by

$$\begin{aligned} & \|((p, \mathbf{v}) - (p_h, \mathbf{v}_h)), ((\tilde{p}, \tilde{\mathbf{v}}) - (\tilde{p}_h, \tilde{\mathbf{v}}_h))\|_{W \times Z'_h} \\ & \leq (1 + \sqrt{2}\beta_h^{-1}) \\ & \quad \inf_{((\phi_h, \boldsymbol{\psi}_h), (\tilde{\phi}_h, \tilde{\boldsymbol{\psi}}_h)) \in W_h \times \tilde{V}_h} \|((p, \mathbf{v}) - (\phi_h, \boldsymbol{\psi}_h)), ((\tilde{p}, \tilde{\mathbf{v}}) - (\tilde{\phi}_h, \tilde{\boldsymbol{\psi}}_h))\|_{W \times Z'_h}. \end{aligned}$$

Proof. The first assertion is a direct consequence of THEOREM 4.30 since the discrete system for the simplified DPG method in (4.54) is the same as the system (4.50) for the practical DPG method with $\hat{V}_h = V_h/H_0(L, Q_h)$.

To prove the second assertion, and let $(p, \mathbf{v}) \in V$ be the solution of (3.2), let $(\tilde{p}, \tilde{\mathbf{v}}) = I_{\partial Q_h}(p, \mathbf{v}) \in L_2(\partial Q_h; \mathbb{R} \times \mathbb{R}^d)$ its trace, and set $(\hat{p}, \hat{\mathbf{v}}) = (p, \mathbf{v}) + H_0(L, Q_h)$, implying $(\hat{p}, \hat{\mathbf{v}}) \in \hat{V}$.

For the discrete solution $((p_h, \mathbf{v}_h), (\tilde{p}_h, \tilde{\mathbf{v}}_h)) \in W_h \times \tilde{V}_h$ let $(\bar{p}_h, \bar{\mathbf{v}}_h) \in V_h$ be the conforming reconstruction of $(\tilde{p}_h, \tilde{\mathbf{v}}_h)$ according to (4.52), and set $(\hat{p}_h, \hat{\mathbf{v}}_h) = (\bar{p}_h, \bar{\mathbf{v}}_h) + H_0(L, Q_h) \in \hat{V}_h$. Then, we have $\|(\tilde{p}_h, \tilde{\mathbf{v}}_h)\|_{Z'_h} = \|I_{\partial R, h}(\bar{p}_h, \bar{\mathbf{v}}_h)\|_{Z'_h} \leq \|(\hat{p}_h, \hat{\mathbf{v}}_h)\|_{L, \partial Q_h}$ by (4.52) and (4.53).

Now, for some $((\phi_h, \boldsymbol{\psi}_h), (\tilde{\phi}_h, \tilde{\boldsymbol{\psi}}_h)) \in W_h \times \tilde{V}_h$ let $(\bar{\phi}_h, \bar{\boldsymbol{\psi}}_h) \in V_h$ be the conforming reconstruction of $(\tilde{\phi}_h, \tilde{\boldsymbol{\psi}}_h)$ as in (4.52), and set $(\hat{\phi}_h, \hat{\boldsymbol{\psi}}_h) = (\bar{\phi}_h, \bar{\boldsymbol{\psi}}_h) + H_0(L, Q_h)$.

Then, using discrete inf-sup stability (4.48) it holds

$$\begin{aligned} & \beta_h \|((p_h, \mathbf{v}_h) - (\phi_h, \boldsymbol{\psi}_h), (\hat{p}_h, \hat{\mathbf{v}}_h) - (\hat{\phi}_h, \hat{\boldsymbol{\psi}}_h))\|_{Q; L, \partial Q_h} \\ & \leq \sup_{(q_h, \mathbf{w}_h) \in Z_h} \frac{\langle B_h((p_h, \mathbf{v}_h) - (\phi_h, \boldsymbol{\psi}_h), (\hat{p}_h, \hat{\mathbf{v}}_h) - (\hat{\phi}_h, \hat{\boldsymbol{\psi}}_h)), (q_h, \mathbf{w}_h) \rangle}{\|(q_h, \mathbf{w}_h)\|_{L^*, Q_h}} \\ & \leq \|(p, \mathbf{v}) - (\phi_h, \boldsymbol{\psi}_h)\|_W + \sup_{(q_h, \mathbf{w}_h) \in Z_h} \frac{\langle \hat{D}_h((\hat{p}, \hat{\mathbf{v}}) - (\hat{\phi}_h, \hat{\boldsymbol{\psi}}_h)), (q_h, \mathbf{w}_h) \rangle}{\|(q_h, \mathbf{w}_h)\|_{L^*, Q_h}} \\ & = \|(p, \mathbf{v}) - (\phi_h, \boldsymbol{\psi}_h)\|_W + \sup_{(q_h, \mathbf{w}_h) \in Z_h} \frac{\gamma_h((\tilde{p}, \tilde{\mathbf{v}}) - (\tilde{\phi}_h, \tilde{\boldsymbol{\psi}}_h), (q_h, \mathbf{w}_h))}{\|(q_h, \mathbf{w}_h)\|_{L^*, Q_h}} \\ & = \|(p, \mathbf{v}) - (\phi_h, \boldsymbol{\psi}_h)\|_W + \|(\tilde{p}, \tilde{\mathbf{v}}) - (\tilde{\phi}_h, \tilde{\boldsymbol{\psi}}_h)\|_{Z'_h} \\ & \leq \sqrt{2} \|((p, \mathbf{v}) - (\phi_h, \boldsymbol{\psi}_h), (\tilde{p}, \tilde{\mathbf{v}}) - (\tilde{\phi}_h, \tilde{\boldsymbol{\psi}}_h))\|_{W \times Z'_h}. \end{aligned}$$

By LEMMA 4.31, we have

$$\|(\tilde{\phi}_h, \tilde{\boldsymbol{\psi}}_h) - (\bar{p}_h, \bar{\mathbf{v}}_h)\|_{Z'_h} \leq \|(\hat{\phi}_h, \hat{\boldsymbol{\psi}}_h) - (\hat{p}_h, \hat{\mathbf{v}}_h)\|_{L, \partial Q_h}$$

which finally implies

$$\begin{aligned}
 & \left\| ((p, \mathbf{v}) - (p_h, \mathbf{v}_h), (\tilde{p}, \tilde{\mathbf{v}}) - (\tilde{p}_h, \tilde{\mathbf{v}}_h)) \right\|_{W \times Z'_h} \\
 & \leq \left\| ((p, \mathbf{v}) - (\phi_h, \boldsymbol{\psi}_h), (\tilde{p}, \tilde{\mathbf{v}}) - (\tilde{\phi}_h, \tilde{\boldsymbol{\psi}}_h)) \right\|_{W \times Z'_h} \\
 & \quad + \left\| ((\phi_h, \boldsymbol{\psi}_h) - (p_h, \mathbf{v}_h), (\tilde{\phi}_h, \tilde{\boldsymbol{\psi}}_h) - (\tilde{p}_h, \tilde{\mathbf{v}}_h)) \right\|_{W \times Z'_h} \\
 & \leq \left\| ((p, \mathbf{v}) - (\phi_h, \boldsymbol{\psi}_h), (\tilde{p}, \tilde{\mathbf{v}}) - (\tilde{\phi}_h, \tilde{\boldsymbol{\psi}}_h)) \right\|_{W \times Z'_h} \\
 & \quad + \left\| ((\phi_h, \boldsymbol{\psi}_h) - (p_h, \mathbf{v}_h), (\hat{\phi}_h, \hat{\boldsymbol{\psi}}_h) - (\hat{p}_h, \hat{\mathbf{v}}_h)) \right\|_{Q; L, \partial Q_h} \\
 & \leq \left(1 + \frac{\sqrt{2}}{\beta_h} \right) \left\| ((p, \mathbf{v}) - (\phi_h, \boldsymbol{\psi}_h), (\tilde{p}, \tilde{\mathbf{v}}) - (\tilde{\phi}_h, \tilde{\boldsymbol{\psi}}_h)) \right\|_{W \times Z'_h}. \quad \square
 \end{aligned}$$

On the one hand the reconstruction space V_h is completely virtual, since it is not required for the realization of the simplified DPG solution. On the other hand, one needs an explicit representation of V_h for the estimate of the discrete inf-sup constant as it is described in the previous section.

Remark 4.33 (Skeleton reduction). *In SECTION 4.2.4, we described a technique to eliminate the interior degrees of freedom for the weakly conforming Least-Squares method. For the numerical solution, the discrete PETROV-GALERKIN is reduced to a positive definite SCHUR complement problem for $(\tilde{p}_h, \tilde{\mathbf{v}}_h)$ using an analogous procedure; see [67, Lem. 9] for explicit estimates for the SCHUR complement depending on β_h and C_L .*

4.6.1 The construction of the FORTIN Operator

In case of conforming trace approximations \tilde{V}_h and simple meshes it is sufficient to construct the FORTIN operator in a reference element R_0 , and then the estimates for the FORTIN operator in $R \subset Q_h$ follows from the scaling argument in SECTION 4.5.2.

In the nonconforming case, a conforming reconstruction $V_h \subset V$ with (4.52) has to be computed. Therefore, we compute a minimum energy extension of trace functions in $\tilde{V}_{R,h}$. On each cell R we select a basis $\{(\tilde{p}_1, \tilde{\mathbf{v}}_1), \dots, (\tilde{p}_N, \tilde{\mathbf{v}}_N)\}$ of $\tilde{V}_{R,h}$ and an extension space $V_{R,h} \subset H(L, R)$. Then, we obtain $(\bar{p}_1, \bar{\mathbf{v}}_1), \dots, (\bar{p}_N, \bar{\mathbf{v}}_N) \in V_{R,h}$ by solving the discrete minimization problem

$$\min_{(\bar{p}_n, \bar{\mathbf{v}}_n) \in V_{R,h}(\tilde{p}_n, \tilde{\mathbf{v}}_n)} \|(\bar{p}_n, \bar{\mathbf{v}}_n)\|_{L,R}$$

in the affine space

$$\begin{aligned}
 V_{R,h}(\tilde{p}_n, \tilde{\mathbf{v}}_n) = \{(\bar{p}_n, \bar{\mathbf{v}}_n) \in V_{R,h} : \gamma_R((\bar{p}_n, \bar{\mathbf{v}}_n) - (\tilde{p}_n, \tilde{\mathbf{v}}_n), (q_{R,h}, \mathbf{w}_{R,h})) = 0 \\
 \text{for } (q_{R,h}, \mathbf{w}_{R,h}) \in Z_{R,h}\},
 \end{aligned}$$

see FIGURE 4.2 for an illustration. The resulting estimates for the FORTIN operator for different polynomial degrees are listed in TABLE 4.1.

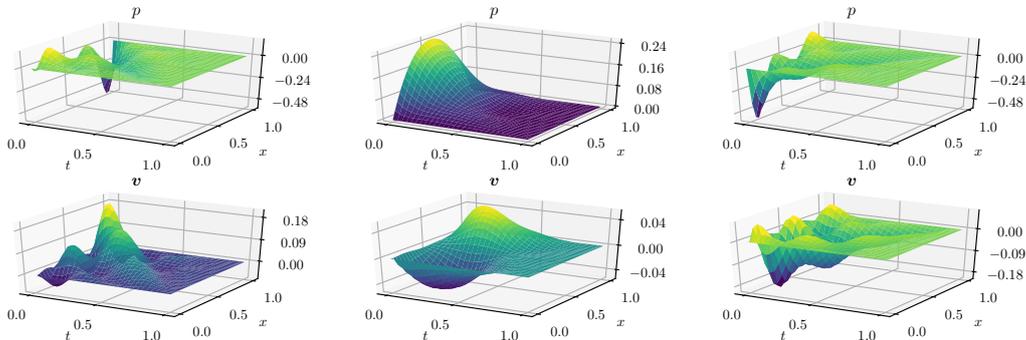


Figure 4.2: Conforming reconstructions in $V_{R,h} = \mathbb{Q}_6(R) \times \mathbb{Q}_6(R)$ for $d = 1$ of the trace space $\tilde{V}_K = \mathbb{P}_2 \times \mathbb{P}_2$ on a face $K \subset \partial R$, and test space $Z_{R,h} = \mathbb{Q}_4(R)^2$. We show the extensions \bar{p}_n and \tilde{v}_n for the three nodal basis functions in \mathbb{P}_2 .

$\ \Pi_{R,h}\ _{L^*,R}$	h_0	h_1	h_2	h_3	h_0
$p = 0$	2.067	2.161	2.182	2.19	2.91
$p = 1$	12.039	18.817	32.87	123.71	34.85
$p = 2$	35.861	64.140	116.78	239.71	144.78

Table 4.1: Considering $R = (0, a_1 h_k) \times (0, a_2 h_k) \times (0, c h_k)$ with $a_1 \approx a_2 \approx c \approx 1$, we present two upper bounds for $\|\Pi_{R,h}\|_{L^*,R}$ in two space-dimension.

Left: Numerical norm estimates with ansatz space $W_{R,h} = \mathbb{Q}_p(R)^3$, test space $Z_{R,h} = \mathbb{Q}_{p+2}(R)^3$, and extension space $\mathbb{Q}_{p+4}(R)^3 \supset \tilde{V}_{R,h}$. The estimates depend on the mesh size $h_k = 2^k$ and the polynomial degree p .

Right: Numerical estimate on the reference cell R_0 with $W_{R,h}^{\text{ext}} = \mathbb{Q}_{p+1}(R)^3$. This yields an inf-sup constant independent of h by the scaling argument in SECTION 4.5.2.

4.6.2 Skeleton reduction

Using a similar procedure as described in SECTION 4.2.4 for weakly conforming Least-Squares, also the DPG method allows for a reduction of the global linear system. In [67], the procedure is explained for variational problems resulting from first-order systems $L\mathbf{y} = \mathbf{b}$ such as the space-time HILBERT space setting that we consider in this chapter. Here, we restrict ourselves to a rough sketch of the procedure. We use

the following operators in every cell $R \in \mathcal{R}_h$

$$\begin{aligned}
 \langle \mathcal{B}_{R,h} v_R, z_R \rangle &= (\mathbf{y}_{R,h}, L^* z_{R,h}), & \mathbf{y}_{R,h} \in W_{R,h}, z_{R,h} \in Z_{R,h}, \\
 \langle C_{R,h} z_{R,h}, \tilde{z}_{R,h} \rangle &= (z_{R,h}, \tilde{z}_{R,h})_{L^*,R}, & z_{R,h}, \tilde{z}_{R,h} \in Z_{R,h}, \\
 \langle \hat{D}_{R,h} \mathbf{y}_{R,h}, z_{R,h} \rangle &= \langle \hat{D}_h \hat{\mathbf{y}}_{R,h}, z_{R,h} \rangle, & \hat{\mathbf{y}}_{R,h} \in \hat{V}_{R,h}, z_{R,h} \in Z_{R,h}, \\
 \langle \ell_{R,h}, z_{R,h} \rangle &= (\mathbf{b}, z_{R,h})_R, & z_{R,h} \in Z_R,
 \end{aligned}$$

to obtain the reduced system $\hat{S}_h \hat{v}_h = \hat{\ell}_h$ with

$$\begin{aligned}
 \hat{S}_h &= - \sum_{R \in \mathcal{R}_h} \begin{pmatrix} 0 \\ \hat{D}_{R,h} \end{pmatrix}' \begin{pmatrix} 0 & \mathcal{B}'_{R,h} \\ \mathcal{B}_{R,h} & C_{R,h} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \hat{D}_{R,h} \end{pmatrix}, \\
 \hat{\ell}_h &= \sum_{R \in \mathcal{R}_h} \begin{pmatrix} 0 \\ \hat{D}_R \end{pmatrix}' \begin{pmatrix} 0 & \mathcal{B}'_R \\ \mathcal{B}_R & C_R \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \ell_R \end{pmatrix}.
 \end{aligned}$$

This allows for an analogous implementation as in ALGORITHM 1, see SECTION 4.2.4. This procedure is well-defined if (4.48) is fulfilled, i.e. in case a FORTIN operator exists.

4.7 A-priori error estimates for smooth solutions

According to THEOREM 4.30, the approximation error for the DPG method can be bounded by an estimate of the following form

$$\|(\mathbf{y}_h^{\text{sol}}, \hat{\mathbf{y}}_h^{\text{sol}}) - (\mathbf{y}^{\text{sol}}, \hat{\mathbf{y}}^{\text{sol}})\|_{Q;L,\partial Q_h} \leq \frac{\sqrt{2}}{\beta_h} \inf_{(\mathbf{y}_h, \hat{\mathbf{y}}_h) \in W_h \times \hat{V}_h} \|(\mathbf{y}_h, \hat{\mathbf{y}}_h) - (\mathbf{y}^{\text{sol}}, \hat{\mathbf{y}}^{\text{sol}})\|_{Q;L,\partial Q_h} \quad (4.55)$$

where $\beta_h > 0$ is the stability constant obtained by the FORTIN operator.

For the weakly conforming Least-Squares method, we have the following error estimate according to THEOREM 4.14

$$\|\mathbf{y}^{\text{sol}} - \mathbf{y}_h^{\text{sol}}\|_{L,Q_h} \leq C \inf_{\mathbf{y}_h \in V_h^{\text{wc}}} \|\mathbf{y}^{\text{sol}} - \mathbf{y}_h\|_{L,Q_h} + \frac{1}{\alpha_0} \sup_{\mathbf{y}_h \in \hat{V}_h^{\text{wc}}} \frac{a_h(\mathbf{y}^{\text{sol}}, \mathbf{y}_h) - \ell_h(\mathbf{y}_h)}{\|\mathbf{y}_h\|_{L,Q_h}}. \quad (4.56)$$

In both estimates, (4.55) and (4.56), the best-approximation error in the graph norm $\|\cdot\|_{L,Q_h}^2 = \|\cdot\|_{Q_h}^2 + \|L(\cdot)\|_{Q_h}^2$ needs to be bounded.¹ To obtain a-priori error estimates, we use standard interpolation theory, see e.g. [24, Sec. 1.5]. To this end, we briefly recall the definition of the SOBOLEV semi-norm in the HILBERT spaces \mathbf{H}^s .

Definition 4.34 (SOBOLEV semi-norm). *Let $U \subset \mathbb{R}^D$ be open and $s \in \mathbb{N}$. For $v \in \mathbf{H}^s(U)$ we define the SOBOLEV semi-norm $|v|_{s,U}$ by*

$$|v|_{s,U}^2 := \sum_{|\alpha|=s} \|\partial^\alpha v\|_U^2.$$

Here, $\alpha \in \mathbb{N}_0^D$ is a multiindex, $|\alpha| := \sum_{r=1}^D \alpha_r$ and $\partial^\alpha := \partial_{x_1}^{\alpha_1} \dots \partial_{x_d}^{\alpha_d}$.

In particular, for $|\alpha| = 1$ there is $d \in \{1, \dots, D\}$ with $\partial^\alpha = \partial_{x_d}$, and therefore we have $|u|_{1,U}^2 = \sum_{d=1}^D \|\partial_{x_d} u\|_U^2$.

For vector fields $v \in \mathbf{H}^s(U, \mathbb{R}^M)$, we set $|v|_{s,U}^2 := \sum_{m=1}^M |v_m|_{s,U}^2$.

The following theorem is standard and a variant can be found in any textbook on Finite Element theory. We use [24] as a reference.

Theorem 4.35. *Let $\{\hat{R}, \hat{P}, \hat{\Sigma}\}$ be a finite element with associated normed vector space $V(\hat{R})$ and assume that $k \in \mathbb{N}$ exists with*

$$\mathbb{P}_k \subset \hat{P} \subset \mathbf{H}^{k+1}(\hat{R}) \subset V(\hat{R}).$$

Let $(\mathcal{R}_h)_h$ be a shape regular family of affine meshes of Q and let I_h^k be the cell-wise defined interpolation operator. Let $l \in [0, k]$ such that $\mathbf{H}^{l+1}(\hat{R}) \subset V(\hat{R})$ with continuous embedding.

¹In case of the DPG estimate, a part of the norm is hidden inside the norm for the skeleton trace.

Then, there is a constant $C > 0$ depending on the shape regularity of the mesh and k such that for all $v \in H^{l+1}(Q_h)$ we have

$$\|v - I_h^k v\|_{Q_h} + \sum_{m=1}^{l+1} h^m \left(\sum_{R \in \mathcal{R}_h} |v - I_h^k v|_{m,R}^2 \right)^{\frac{1}{2}} \leq Ch^{l+1} |v|_{l+1,Q}.$$

Proof. See [24, Thm. 1.103 and Cor. 1.109]. \square

Example 4.36. A typical example for this setting is $V(\hat{R}) = C^0(\hat{R})$, $\hat{P} = \mathbb{Q}_k$ for $k \in \mathbb{N}$. Where $I_h^k: C^0(Q) \rightarrow C^0(Q) \cap \prod_{R \in \mathcal{R}_h} \mathbb{Q}_k(R)$ is defined by

$$v \mapsto \sum_{z \in \mathcal{N}} v(z) \phi_z, \quad v \in C^0(Q).$$

Here, $\mathcal{N} \subset Q$ is a set of nodal points and the corresponding nodal basis of V_h is

$$\{\phi_z: z \in \mathcal{N}\} \subset V_h := \{\phi_z \in C^0(Q): \phi_z|_R \in \mathbb{Q}_k(R), R \in \mathcal{R}_h\}.$$

To apply THEOREM 4.35 we make use of the following lemma.

Lemma 4.37. Let $L: C^1(R, \mathbb{R}^M) \rightarrow C^0(R, \mathbb{R}^N)$ be a differential operator of first order in $R \subset \mathbb{R}^d$ of the form

$$L\mathbf{y} = \sum_{d=1}^D A^d \partial_{x_d} \mathbf{y}, \quad \mathbf{y} \in C^1(R, \mathbb{R}^M),$$

where $A^d \in L_\infty(R, \mathbb{R}^{N \times M})$, $d \in \{1, \dots, D\}$.

Then there is a real constant $C > 0$ such that for $\mathbf{y} \in C^1(R, \mathbb{R}^M)$

$$\|L\mathbf{y}\|_R \leq C |\mathbf{y}|_{1,R},$$

where C depends on $(A^d)_d$, M , N , D .

Proof. For $\mathbf{y} \in C^1(R, \mathbb{R}^M)$, $d \in \{1, \dots, D\}$ and $n \in \{1, \dots, N\}$, we get

$$\begin{aligned} \|(A^d \partial_{x_d} \mathbf{y})_n\|_R^2 &= \int_R \left(\sum_{m=1}^M (A^d)_{nm}(x) \partial_{x_d} \mathbf{y}_m(x) \right)^2 dx \\ &\leq \int_R \left(\sum_{m=1}^M (A^d)_{nm}(x)^2 \right) \left(\sum_{m=1}^M \partial_{x_d} \mathbf{y}_m(x)^2 \right) dx \\ &\leq M \sup_{m=1, \dots, M} \|(A^d)_{nm}\|_{R, \infty}^2 \|\partial_{x_d} \mathbf{y}\|_R^2 \end{aligned}$$

yielding

$$\|A^d \partial_{x_d} \mathbf{y}\|_R^2 = \sum_{n=1}^N \|(A^d \partial_{x_d} \mathbf{y})_n\|_R^2 \leq MN \sup_{\substack{n=1, \dots, N \\ m=1, \dots, M}} \|(A^d)_{nm}\|_{R, \infty}^2 \|\partial_{x_d} \mathbf{y}\|_R^2.$$

This finishes the proof, since

$$\|L\mathbf{y}\|_R \leq \sum_{d=1}^D \|A^d \partial_{x_d} \mathbf{y}\|_R \leq \sqrt{DMN} \sup_{\substack{n=1,\dots,N, \\ d=1,\dots,D}} \sup_{m=1,\dots,M} \|(A^d)_{nm}\|_{R,\infty} \sum_{d=1}^D \|\partial_{x_d} \mathbf{y}\|_R$$

giving $C = \sqrt{DMN} \sup_{\substack{n=1,\dots,N, \\ d=1,\dots,D}} \sup_{m=1,\dots,M} \|(A^d)_{mn}\|_{R,\infty}$. \square

Application the DPG-Estimate

Let the solution of $L\mathbf{y}^{\text{sol}} = \mathbf{b}$ fulfill $\mathbf{y} \in \mathbf{H}^{k+1}(Q, \mathbb{R}^{1+d})$ for a $k \in \mathbb{N}$. Applying the abstract result THEOREM 4.35 to (4.55) is done in two steps.

1. We use tensor-product elements $\hat{P} = \mathbb{Q}_{k-1}$ yielding $\mathbb{P}_{k-1} \subset \hat{P}$. We choose $V(\hat{R}) = C^0(\hat{R})$ and the standard LAGRANGE interpolation in every cell.

Then, setting $l = k - 1$, we obtain by using THEOREM 4.35 in every component

$$\|\mathbf{y}^{\text{sol}} - I_h^{k-1} \mathbf{y}^{\text{sol}}\|_{Q_h} \leq Ch^k |\mathbf{y}^{\text{sol}}|_{k,Q} \quad (4.57)$$

2. The quotient norm on the skeleton is given by

$$\|\hat{\mathbf{y}}\|_{L,\partial Q_h} = \inf_{\mathbf{y} \in \hat{\mathbf{y}}} \|\mathbf{y}\|_{L,Q_h} = \inf_{\mathbf{y} \in \hat{\mathbf{y}}} \left(\|\mathbf{y}\|_{Q_h}^2 + \|L\mathbf{y}\|_{Q_h}^2 \right)^{\frac{1}{2}}, \quad \hat{\mathbf{y}} \in \hat{\mathbf{H}}(L, Q_h).$$

Since $\hat{\mathbf{y}}^{\text{sol}} = \mathbf{y}^{\text{sol}} + \mathbf{H}_0(L, Q_h)$ by construction, we select the global LAGRANGE interpolation $I_h^k \mathbf{y}^{\text{sol}}$. Choosing $\mathbf{y}_0 = 0 \in \mathbf{H}_0(L, Q_h)$, we obtain for the interpolation's trace, $\hat{\mathbf{y}}_h := I_h^k \mathbf{y}^{\text{sol}} + \mathbf{H}_0(L, Q_h)$

$$\begin{aligned} \|\hat{\mathbf{y}}^{\text{sol}} - \hat{\mathbf{y}}_h\|_{L,\partial Q_h} &= \inf_{\mathbf{y}_0 \in \mathbf{H}_0(L, Q_h)} \|\mathbf{y}^{\text{sol}} - I_h^k \mathbf{y}^{\text{sol}} + \mathbf{y}_0\|_{L, Q_h} \\ &\leq \left(\|\mathbf{y}^{\text{sol}} - I_h^k \mathbf{y}^{\text{sol}}\|_{Q_h}^2 + \|L(\mathbf{y}^{\text{sol}} - I_h^k \mathbf{y}^{\text{sol}})\|_{Q_h}^2 \right)^{\frac{1}{2}}. \end{aligned}$$

As an approximation space \hat{V}_h of \hat{V} , we use traces of polynomials such that in every cell $R \in \mathcal{R}_h$, we have $\mathbb{Q}_k(R) + \mathbf{H}_0(L, Q_h) \subset \hat{V}_h(R)$. Again, choosing $V(R) = C^0(R)$ yields for $l = k$ as in step 1. by THEOREM 4.35

$$\|\mathbf{y}^{\text{sol}} - I_h^k \mathbf{y}^{\text{sol}}\|_{Q_h} \leq Ch^{k+1} |\mathbf{y}^{\text{sol}}|_{k+1,Q}. \quad (4.58)$$

For remaining part of the graph norm, we obtain using LEMMA 4.37

$$\|L(\mathbf{y}^{\text{sol}} - I_h^k \mathbf{y}^{\text{sol}})\|_{Q_h} \leq \tilde{C} \left(\sum_{R \in \mathcal{R}_h} |\mathbf{y}^{\text{sol}} - I_h^k \mathbf{y}^{\text{sol}}|_{1,R}^2 \right)^{\frac{1}{2}} \leq \hat{C} h^k |\mathbf{y}^{\text{sol}}|_{k+1,Q}. \quad (4.59)$$

Now we can insert (4.57), (4.58) and (4.59) into (4.55) yielding the expected convergence order k

$$\|\mathbf{y}^{\text{sol}} - \mathbf{y}_h^{\text{sol}}\|_{Q_h} + \|\hat{\mathbf{y}}^{\text{sol}} - \hat{\mathbf{y}}_h^{\text{sol}}\|_{L;\partial Q_h} \leq Ch^k |\mathbf{y}^{\text{sol}}|_{k+1,Q}. \quad (4.60)$$

Corollary 4.38. *For $k \in \mathbb{N}$, we expect convergence of order k of the DPG approximation to the analytical solution in $\|\cdot\|_{Q_h} + \|\cdot\|_{L;\partial Q_h}$ under the following conditions:*

1. *The solution of $L\mathbf{y}^{\text{sol}} = \mathbf{b}$ fulfills $\mathbf{y}^{\text{sol}} \in \mathbf{H}^{k+1}(Q, \mathbb{R}^{1+d})$,*
2. *$\hat{P} = \mathbb{Q}_{k-1}$,*
3. *$\mathbb{Q}_k(R) + \mathbf{H}_0(L, R) \subset \hat{V}_h(R)$ for all $R \in \mathcal{R}_h$.*

Application to the weakly conforming Least-Squares method

To obtain an estimate for the weakly conforming Least-Squares method, we assume that the second addend in (4.56) can be neglected.

Then, using the same arguments as for the DPG method in (4.59), we obtain the following result.

Corollary 4.39. *For $k \in \mathbb{N}$, we expect convergence of order k in $\|\cdot\|_{L,Q_h}$ for the weakly conforming Least-Squares method under the following conditions:*

1. *The solution of $L\mathbf{y} = \mathbf{b}$ fulfills $\mathbf{y} \in \mathbf{H}^{k+1}(Q, \mathbb{R}^{1+d})$,*
2. *$\hat{P} = \mathbb{Q}_k$,*
3. *The second addend in (4.56) decays at least with order k .*

Remark 4.40. *Note that by THEOREM 4.14, the test space V_h^* does not need to have any approximation quality since it just provides coupling conditions over cell interfaces.*

The convergence rate/reduction factor

Since we use families of meshes that result from dividing the mesh-width by factor 2 on each refinement, we obtain the expected reduction factor² $\theta_k = 2^k$ for

$$\theta_k := \frac{\|(\mathbf{y}, \hat{\mathbf{y}}) - (\mathbf{y}_{h_k}, \hat{\mathbf{y}}_{h_k})\|_{L;Q;\partial Q}}{\|(\mathbf{y}, \hat{\mathbf{y}}) - (\mathbf{y}_{h_{k+1}}, \hat{\mathbf{y}}_{h_{k+1}})\|_{L;Q;\partial Q}}, \quad h_l := \frac{h_0}{2^l}.$$

²We also call this quantity the (convergence) rate.

Chapter 5

Numerical Experiments

5.1 Numerical setup

The numerical experiments presented in the following sections have been implemented using the parallel Finite Element framework M++ described in [66]. A key feature of M++ is its parallel programming model which hides the details of parallelism from the developer. Furthermore, its modular structure is designed for the implementation of new FEM spaces and methods. This enables rapid development of parallel FEM software while being able to control the whole numerical algorithm, including mesh-refinement, load-balancing, FEM bases, quadrature formulas, preconditioners as well as linear and non-linear solvers.

5.1.1 Discretizations and error quantities

We consider discretizations originating from the simplified Discontinuous-PETROV-GALERKIN method described in SECTION 4.6 and the weakly conforming Least-Squares method that was introduced in SECTION 4.2.

Before comparing the performance of these methods with respect to different examples, we provide general remarks on the numerical setup.

Considered error quantities

Fitting our analytical setting, we consider the difference of the numerical approximation to the exact solution in component-wise $L_2(Q)$ norms.

However, for some configurations we observed convergence of increased order when looking at the cell-wise means. To this end, for $\mathbf{y} \in L_1(Q, \mathbb{R}^m)$, we define the cell-wise

mean value $\Pi_{Q_h}^0 \mathbf{y} : Q \rightarrow \mathbb{R}^m$ on a space-time mesh \mathcal{R}_h of Q by

$$(\Pi_{Q_h}^0 \mathbf{y})(x) := \frac{1}{|R|} \int_R \mathbf{y}(z) dz, \quad x \in R, R \in \mathcal{R}_h.$$

Straight-forward calculations show the bound $\|\Pi_{Q_h}^0 \mathbf{y}\|_{L_1(Q, \mathbb{R}^m)} \leq \|\mathbf{y}\|_{L_1(Q, \mathbb{R}^m)}$ and $(\Pi_{Q_h}^0)^2 = \Pi_{Q_h}^0$. For $\mathbf{y} \in L_2(Q, \mathbb{R}^m)$, one can show $\|\Pi_{Q_h}^0 \mathbf{y}\|_{L_2(Q, \mathbb{R}^m)} \leq \|\mathbf{y}\|_{L_2(Q, \mathbb{R}^m)}$.

As a result, the mapping $\Pi_{Q_h}^0 : L_l(Q, \mathbb{R}^m) \rightarrow L_l(Q, \mathbb{R}^m)$ is a bounded linear projection in $L_l(Q, \mathbb{R}^m)$, $l \in \{1, 2\}$.

In following convergence considerations, the L_l -mean error refers to the quantity $\|\Pi_{Q_h}^0 \mathbf{y} - \Pi_{Q_h}^0 \mathbf{y}_h\|_{L_l(Q, \mathbb{R}^m)}$, $l \in \{1, 2\}$, for the exact solution $\mathbf{y} \in H(L, Q)$ and the discrete numerical approximation \mathbf{y}_h .

5.1.2 Mesh refinements

The calculations are performed on sequences of space-time meshes. The coarsest mesh, we say *level 0*, is cell-wise refined using bisection of all edges yielding the mesh on *level 1*, *level 2* and so forth. In one spatial dimension, the space-time cells are rectangles. Thus, each cell is divided into 4 congruent rectangles on refinement. In two spatial dimensions, we consider space-time cells that are cuboids each of which is refined into 8 congruent cuboids.

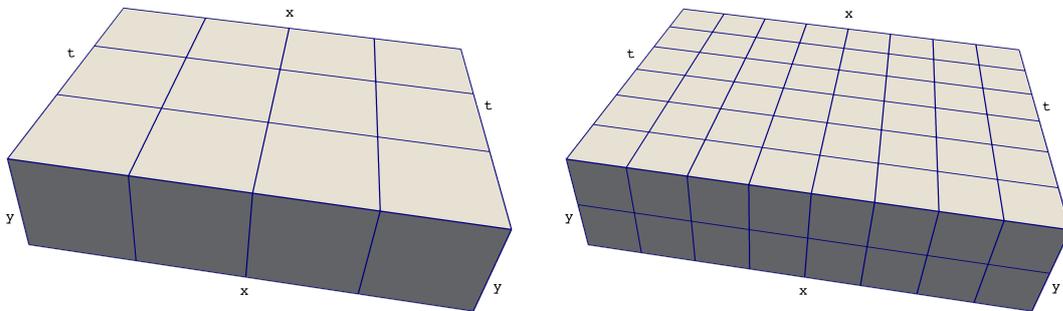


Figure 5.1: A space-time mesh in 2D on level ℓ (left) and the refined version using bisection of edges on level $\ell + 1$ (right).

Remark 5.1. *Note that for our demonstration-of-concept implementation, we restricted ourselves to rectangular meshes. However, the method can also be implemented using other types of meshes as long as each cell is of the form $K \times (a, b)$ for a spatial cell $K \subset \mathbb{R}^d$. Extending our implementation to more general meshes is a promising future challenge.*

5.1.3 Local SCHUR complements and problem sizes

On each level, we provide the number of global degrees of freedom, *DoFs* that remain after eliminating the interior cell degrees of freedom using SCHUR-complement reduction, see SECTION 4.2.4 for the weakly conforming Least-Squares method and SECTION 4.6.2 for the DPG method.

Furthermore, we provide the amount of total DoFs including the eliminated interior degrees of freedom in the column *all DoFs*.

For the SCHUR-complement process, we need to assemble and invert a local saddle-point matrix in each space-time cell $R \in \mathcal{R}_h$. According to SECTION 4.6.2, for the DPG-method, these matrices have the following structure

$$S_{R,h} = \begin{pmatrix} 0 & \mathcal{B}'_{R,h} \\ \mathcal{B}_{R,h} & C_{R,h} \end{pmatrix} \in \mathbb{R}^{(\dim W_{R,h} + \dim Z_{R,h}) \times (\dim W_{R,h} + \dim Z_{R,h})},$$

where $\mathcal{B}_{R,h}$, $C_{R,h}$ are defined by

$$\begin{aligned} \langle \mathcal{B}_{R,h} \mathbf{y}_{R,h}, \mathbf{z}_{R,h} \rangle &= (\mathbf{y}_{R,h}, L^* \mathbf{z}_{R,h})_R, & \mathbf{y}_{R,h} \in W_{R,h}, \mathbf{z}_{R,h}, \tilde{\mathbf{z}}_{R,h} \in Z_{R,h} \\ \langle C_{R,h} \mathbf{z}_{R,h}, \tilde{\mathbf{z}}_{R,h} \rangle &= (\mathbf{z}_{R,h}, \tilde{\mathbf{z}}_{R,h})_{L^*, R}, \end{aligned}$$

for the cell-wise defined ansatz space $W_{R,h} \subset L_2(R, \mathbb{R}^{1+d})$ and the cell-wise defined test space $Z_{R,h} \subset H(L^*, R)$. Since we use cell-wise tensor-product polynomial spaces in each of the $1 + d$ components, i.e.

$$W_{R,h} = \mathbb{Q}_k(R)^{1+d}, \quad Z_{R,h} = \mathbb{Q}_l(R)^{1+d}, \quad k, l \in \mathbb{N}_0,$$

and by using $\dim \mathbb{Q}_k(R) = (k + 1)^{1+d}$ we conclude that

$$\dim W_{R,h} = (1 + d) \cdot (k + 1)^{1+d}, \quad \dim Z_{R,h} = (1 + d) \cdot (l + 1)^{1+d},$$

and that $S_{R,h}$ is a square matrix with $(1 + d) \cdot ((k + 1)^{1+d} + (l + 1)^{1+d})$ rows, see TABLE 5.1.

As a result, at the expense of inverting a locally defined dense matrix, we can eliminate up to a few hundreds of local unknowns.

Analogous considerations also hold for the weakly conforming Least-Squares method. See SECTION 4.2.4 for the detailed structure of the local saddle-point matrix in this case.

Quadrature formulas – the curse of dimensions

The local saddle-point matrices for the DPG method and also for the weakly-conforming Least-Squares method contain cell-wise L_2 inner products of polynomials in each component.

d	k	l	$\dim W_{R,h}$	$\dim Z_{R,h}$	rows in $S_{R,h}$	d	k	l	$\dim W_{R,h}$	$\dim Z_{R,h}$	rows in $S_{R,h}$
1	0	3	2	32	34	2	0	3	3	192	195
1	1	4	8	50	58	2	1	4	24	375	399
1	2	5	18	72	90	2	2	5	81	648	729
1	3	6	32	98	130	2	3	6	192	1029	1221
1	4	7	50	128	178	2	4	7	375	1536	1911
1	5	8	72	162	234	2	5	8	648	2187	2835

Table 5.1: Dimensions of $W_{R,h}$, $Z_{R,h}$ and the number of rows for $S_{R,h}$ in one and two space-dimensions for different configurations of the DPG-method.

Choosing $Z_{R,h} = \mathbb{Q}_l(R)^{1+d}$, we need to integrate products of polynomials in $\mathbb{Q}_l(R) \cdot \mathbb{Q}_l(R) = \mathbb{Q}_{2l}(R)$ in order to assemble the matrix $C_{R,h}$.

Using a 1D GAUSS quadrature with $n \in \mathbb{N}$ nodes in all $d + 1$ space-time axes, we are able to integrate functions in $\mathbb{Q}_{2n-1}(R)$ exactly. Thus, to compute $L_2(R)$ inner products of functions in $\mathbb{Q}_l(R)$, we need $n \geq l + 1$ quadrature nodes in *every* space-time axis, resulting in at least $n_{\text{entry}} = (l + 1)^{1+d}$ operations to assemble a single entry of $C_{R,h}$. Thus, exploiting the symmetry $C_{R,h}$, we need at least

$$N = \frac{1}{2} \cdot n_{\text{entry}} \cdot (\dim Z_{R,h})^2 = \frac{1}{2} \cdot (l + 1)^{1+d} \cdot ((1 + d) \cdot (l + 1)^{1+d})^2$$

to assemble the dense matrix $C_{R,h}$ in a single cell when using a naive implementation with nested loops.

Although the assembling process can be done in parallel for every cell, it turned out that the resulting local costs grow significantly for high-order configurations, see TABLE 5.2.

d	1					
l	3	4	5	6	7	8
$N \geq$	8 192	31 250	93 312	235 298	524 288	1 062 882
d	2					
l	3	4	5	6	7	8
$N \geq$	1 179 648	8 789 062	45 349 632	181 591 232	603 979 776	1 743 392 200

Table 5.2: Lower bound for the amount of elementary operations needed to assemble $C_{R,h}$.

In the future, we would like to exploit the structure of our polynomial spaces to reduce these assembling costs. In [50, 70], approaches to handle this effect are considered.

Solving the global linear system

To solve the linear systems, for most examples we use a GMRES iterative solver preconditioned by a symmetric GAUSS-SEIDEL method in every parallel subdomain. To eliminate errors resulting from preliminary stopping of this iteration scheme, in some of the academic examples we make use of the parallel direct solver described in [47].

Both preconditioners perform suboptimal in our experiments, since the symmetric GAUSS-SEIDEL preconditioner needs by far too many steps (up to tens of thousands) for high-order variants on high levels and the parallel direct solver requires a large amount of memory due to fill-in effects.

It is a future challenge, to construct efficient preconditioners for the methods presented in this work.

Remark 5.2. *Since the SCHUR complement matrices for the weakly conforming Least-Squares method as well as for the DPG method are symmetric and positive definite, instead of the GMRES solver, a conjugate gradient (CG) scheme is a straightforward choice. However, in our experiments the CG algorithm combined with the GAUSS-SEIDEL preconditioner performs significantly worse than the GMRES solver.*

Configurations of the DPG method

As shown in SECTION 4.7, for the DPG method it is reasonable to select a polynomial space of degree $k \in \mathbb{N}$ on each space-time face and a space of degree $k - 1$ for the variables inside each cell. By COROLLARY 4.38, this yields a scheme converging with order k in the $L_2(Q)$ norm in case that the solution is smooth enough.

Remark 5.3. *In our calculations, see e.g. FIGURE 5.3 and FIGURE 5.4, we observe that using polynomials of degree k instead of $k - 1$ inside the cells increases the order of convergence by 1 for 1D examples. Thus, we also present numerical results for these configurations.*

Further enrichment of the local ansatz spaces does not lead to additional improvements in our examples. Therefore, we do not provide results for these configurations.

We select for both beforehand mentioned configurations polynomials of degree $k + 2$ as a test space which performed well in the experiments.

See TABLE 5.3 and TABLE 5.4 for the DPG configuration used in the numerical experiments. The configuration names are used to label the provided convergence results in the following sections.

configuration name	D1	D2	D3	D4	D5
cell polynomial degree $k - 1$	0	1	2	3	4
face polynomial degree k	1	2	3	4	5
test polynomial degree $k + 2$	3	4	5	6	7
expected order of convergence in $L_2(Q)$	1	2	3	4	5

Table 5.3: Considered configurations for the DPG method as suggested by SECTION 4.7.

configuration name	D1 ⁺	D2 ⁺	D3 ⁺	D4 ⁺	D5 ⁺
cell polynomial degree k	1	2	3	4	5
face polynomial degree k	1	2	3	4	5
test polynomial degree $k + 2$	3	4	5	6	7
expected order of convergence in $L_2(Q)$	1	2	3	4	5

Table 5.4: Considered configurations for the DPG method converging with increased order in our experiments.

Configurations of the weakly conforming Least-Squares method

By THEOREM 4.14 and the discussion SECTION 4.2.5, we note that for the weakly conforming Least-Squares method, the stability conditions (4.14) and (4.25) are sufficient to obtain a convergent method. However, these two conditions only hold for some well-balanced pairings of ansatz and test space.

We present numerical results for the following configurations in one spatial dimension that we found by numerical experiments.

For different choices $k \in \mathbb{N}$, we select polynomials in $\mathbb{Q}_k(R)$ for the pressure as well as for the velocity component in V_h^{wc} . For the discrete coupling space V_h^* , we distinguish faces in time having the form $F = (t_-, t^+) \times \{a\} \subset \partial Q_h$ and faces in space of the form $F = \{t\} \times (a, b) \subset \partial Q_h$. We choose spaces of face bubbles on each face, where we use the same coupling for the pressure and velocity component. More precisely, we set

$$V_h^*|_F = \text{span} \left\{ \begin{pmatrix} \phi_1 \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} \phi_{k_F} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \phi_1 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ \phi_{k_F} \end{pmatrix} \right\},$$

for $k_F \in \mathbb{N}$ that can be chosen for each face F individually. In TABLE 5.5, we list the configurations that have been used in the experiments.

Again, the configuration names are used to label the provided convergence results in the following section.

configuration name	W2	W3	W4	W5	W6	W7	W8
cell polynomial degree k	2	3	4	5	6	7	8
$k_{F,\text{space}}, F = (t_-, t^+) \times \{a\}$ (face in space)	1	2	3	3	4	4	5
$k_{F,\text{time}}, F = \{t\} \times (a, b)$ (face in time)	2	3	3	4	4	5	5
expected order of convergence in $L_2(Q)$	1	2	3	4	5	6	7

Table 5.5: Considered configurations for the weakly conforming Least-Squares method.

The basis functions ϕ_l are chosen as follows, where the choice of η_l ensures that $\|\phi_l\|_{L_1((0,1))} = 1$ for scaling reasons:¹

$$\phi_j(s) = \eta_j \cdot \prod_{i=0}^j (s - i/j)$$

5.1.4 A remark on the upcoming sections

In the following, we provide a large collection of numerical results for different problems and schemes. Due to its structure, the following sections are not intended for being read sequentially in detail. The reader may have a look at the summary in SECTION 5.5 before digging into the examples. He or she also may focus on the more interesting examples in two spatial dimensions, SECTION 5.3.2, SECTION 5.3.3, as well as the low-regularity example in 1D, SECTION 5.2.3, or the space-time adaptivity example in SECTION 5.4.

The author provides this extensive data set hoping that this is useful for further comparisons to other methods. Since a detailed description for each benchmark problem and numerical method is provided, the reader might reproduce the presented results using his or her own implementation.

¹Using this scaling leads to better conditioned system matrices in our experiments.

5.2 Numerical examples (1D)

In this section, we present different numerical examples in one spacial dimension. Some of these are designed to verify the convergence rates of the methods that follow from the theory.

5.2.1 A smooth example

To compare convergence rates, we consider a smooth example given by

$$p(x, t) = a \sin(\omega\pi(x + t)) = -\mathbf{v}(x, t), \quad (x, t) \in \Omega \times (0, T), \quad (5.1)$$

with $\Omega = (0, 1)$, $T = \frac{3}{\pi}$ and $a = 10$, $\omega = 4.124324523$. We select homogeneous material parameters $\rho \equiv \kappa \equiv 1$ in Ω and a space-time plot of this solution is presented in FIGURE 5.2.

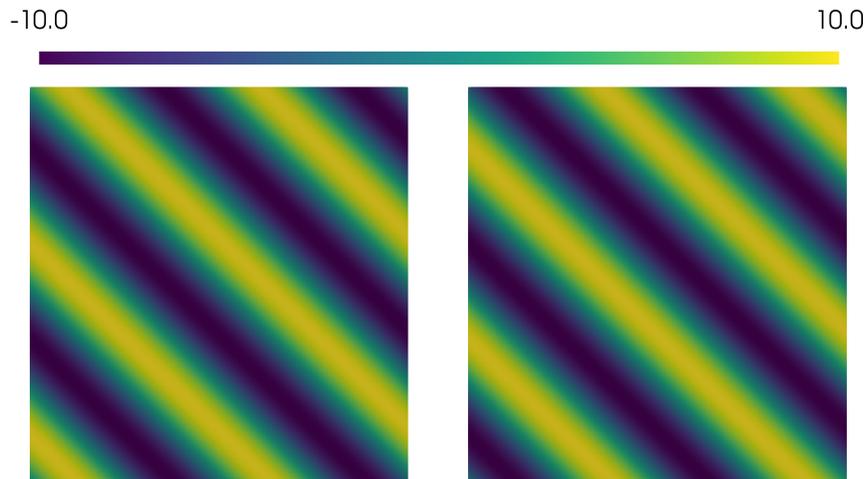


Figure 5.2: Plots of p (left) and \mathbf{v} (right) component of (5.1). The vertical axis corresponds to space and the horizontal axis is time, left to right.

Since we have $(p, \mathbf{v}) \in C^\infty(Q, \mathbb{R}^2)$, we expect to observe at least the convergence rates predicted by SECTION 4.7 that are maximal with respect to the used polynomial degree.

See FIGURE 5.3 to FIGURE 5.5 for a convergence study. A detailed collection of the results is provided in TABLE 5.6 to TABLE 5.8.

5.2. Numerical examples (1D)

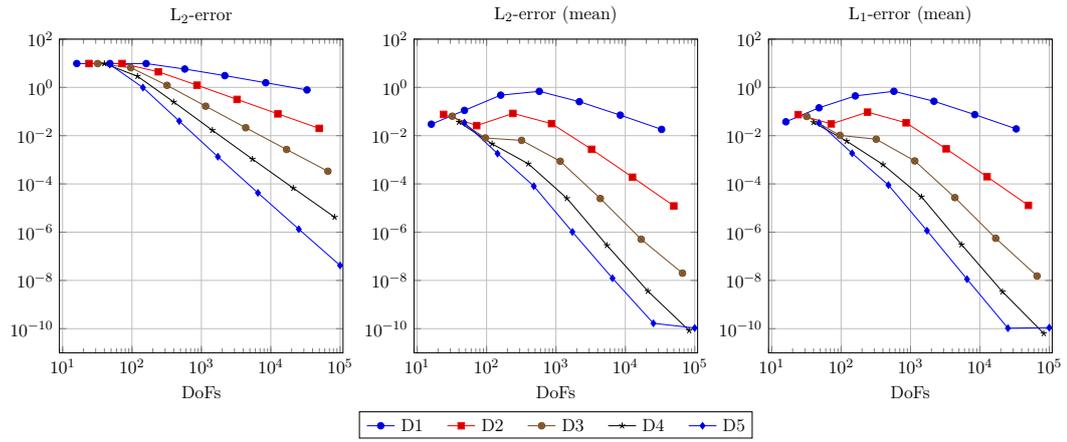


Figure 5.3: Convergence results for DPG with configurations according to TABLE 5.3.

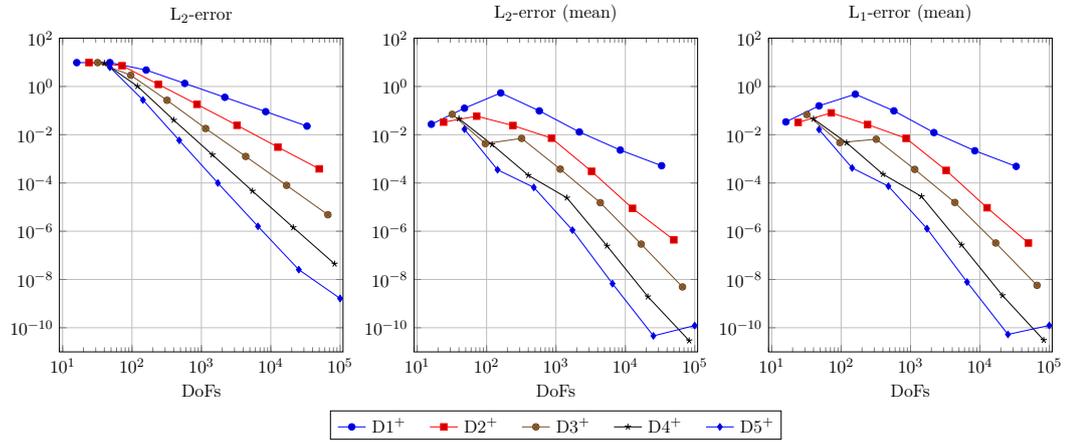


Figure 5.4: Convergence results for DPG with configurations according to TABLE 5.4.

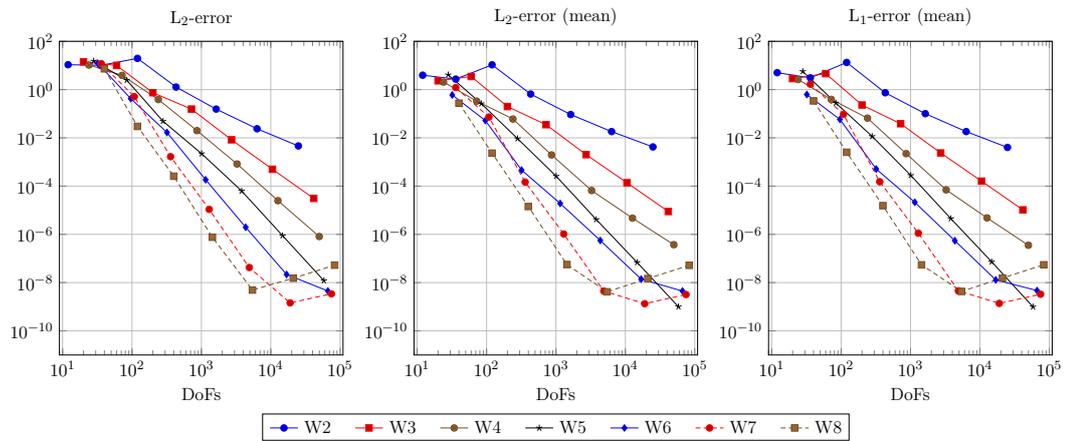


Figure 5.5: Convergence results for WC with configurations according to TABLE 5.5.

Discussion: DPG. The DPG method configured as described by TABLE 5.3 shows a convergence behaviour that is in agreement with the theoretical predictions provided by SECTION 4.7, see FIGURE 5.3. After leaving the pre-asymptotic regime, the convergence rates are matching the predictions precisely.

Interestingly, the convergence order of the cell-wise mean value in L_2 norms and also L_1 norms is increased, i.e. if the L_2 norm converges with order k in these experiments, we observe a convergence order of $k + \Delta k$ with $\Delta k \in [1, 3]$ for the cell-wise mean. Furthermore, the schemes display pre-asymptotic behavior of the mean error as well as round-off errors for high levels.

Discussion: DPG – increased cell degree. Increasing the polynomial degree in every cell by 1, we observe convergence rates that are higher than the theoretical predictions provided by SECTION 4.7. All configurations show a by approximately 1 increased convergence rate of the error in the $L_2(Q)$ norm.

Also the convergence of the cell-wise mean values benefits from the enriched polynomial spaces. While the absolute error is reduced for all configurations, only the low-order schemes show increased orders for the mean-values.

Discussion: weakly conforming Least-Squares. In this benchmark, the weakly conforming Least-Squares method converges in most examples with a increased order compared the theoretical prediction. For the lowest order case, see TABLE 5.8, the rate is oscillating.

Looking at the mean values, we do not observe increased convergence rates.

In comparison to the DPG method with enriched cell spaces, the required amount of global DoFs to achieve a certain accuracy is comparably large.

However, the weakly conforming Least-Squares method is more sensitive to round-off errors in the high-order variants in comparison to the DPG method.

5.2. Numerical examples (1D)

conf	level	cells	DoFs	all DoFs	L_2 -error	rate	order	L_2 -error (mean)	rate	order	L_1 -error (mean)	rate	order
D1	0	1	16	18	$9.746 \cdot 10^0$	–	–	$2.983 \cdot 10^{-2}$	–	–	$3.751 \cdot 10^{-2}$	–	–
D1	1	4	48	56	$9.747 \cdot 10^0$	1.00	0.00	$1.115 \cdot 10^{-1}$	0.27	–1.90	$1.426 \cdot 10^{-1}$	0.26	–1.93
D1	2	16	160	192	$9.725 \cdot 10^0$	1.00	0.00	$4.769 \cdot 10^{-1}$	0.23	–2.10	$4.433 \cdot 10^{-1}$	0.32	–1.64
D1	3	64	576	704	$5.769 \cdot 10^0$	1.69	0.75	$6.865 \cdot 10^{-1}$	0.69	–0.53	$6.897 \cdot 10^{-1}$	0.64	–0.64
D1	4	256	2176	2688	$3.078 \cdot 10^0$	1.87	0.91	$2.555 \cdot 10^{-1}$	2.69	1.43	$2.663 \cdot 10^{-1}$	2.59	1.37
D1	5	1024	8448	10496	$1.562 \cdot 10^0$	1.97	0.98	$7.079 \cdot 10^{-2}$	3.61	1.85	$7.438 \cdot 10^{-2}$	3.58	1.84
D1	6	4096	33280	41472	$7.84 \cdot 10^{-1}$	1.99	0.99	$1.817 \cdot 10^{-2}$	3.90	1.96	$1.907 \cdot 10^{-2}$	3.90	1.96
D2	0	1	24	32	$9.736 \cdot 10^0$	–	–	$7.597 \cdot 10^{-2}$	–	–	$7.429 \cdot 10^{-2}$	–	–
D2	1	4	72	104	$9.751 \cdot 10^0$	1.00	0.00	$2.604 \cdot 10^{-2}$	2.92	1.54	$3.07 \cdot 10^{-2}$	2.42	1.27
D2	2	16	240	368	$4.436 \cdot 10^0$	2.20	1.14	$8.344 \cdot 10^{-2}$	0.31	–1.68	$9.446 \cdot 10^{-2}$	0.33	–1.62
D2	3	64	864	1376	$1.237 \cdot 10^0$	3.59	1.84	$3.146 \cdot 10^{-2}$	2.65	1.41	$3.406 \cdot 10^{-2}$	2.77	1.47
D2	4	256	3264	5312	$3.178 \cdot 10^{-1}$	3.89	1.96	$2.715 \cdot 10^{-3}$	11.59	3.53	$2.847 \cdot 10^{-3}$	11.96	3.58
D2	5	1024	12672	20864	$7.999 \cdot 10^{-2}$	3.97	1.99	$1.896 \cdot 10^{-4}$	14.32	3.84	$1.989 \cdot 10^{-4}$	14.31	3.84
D2	6	4096	49920	82688	$2.003 \cdot 10^{-2}$	3.99	2.00	$1.223 \cdot 10^{-5}$	15.50	3.95	$1.289 \cdot 10^{-5}$	15.43	3.95
D3	0	1	32	50	$9.703 \cdot 10^0$	–	–	$6.4 \cdot 10^{-2}$	–	–	$6.236 \cdot 10^{-2}$	–	–
D3	1	4	96	168	$6.58 \cdot 10^0$	1.47	0.56	$8.017 \cdot 10^{-3}$	7.98	3.00	$1.03 \cdot 10^{-2}$	6.05	2.60
D3	2	16	320	608	$1.218 \cdot 10^0$	5.40	2.43	$6.364 \cdot 10^{-3}$	1.26	0.33	$7.119 \cdot 10^{-3}$	1.45	0.53
D3	3	64	1152	2304	$1.665 \cdot 10^{-1}$	7.32	2.87	$8.692 \cdot 10^{-4}$	7.32	2.87	$8.954 \cdot 10^{-4}$	7.95	2.99
D3	4	256	4352	8960	$2.129 \cdot 10^{-2}$	7.82	2.97	$2.489 \cdot 10^{-5}$	34.92	5.13	$2.713 \cdot 10^{-5}$	33.00	5.05
D3	5	1024	16896	35328	$2.675 \cdot 10^{-3}$	7.96	2.99	$5.086 \cdot 10^{-7}$	48.94	5.61	$5.576 \cdot 10^{-7}$	48.65	5.61
D3	6	4096	66560	140288	$3.349 \cdot 10^{-4}$	7.99	3.00	$2.002 \cdot 10^{-8}$	25.40	4.67	$1.521 \cdot 10^{-8}$	36.66	5.20
D4	0	1	40	72	$9.686 \cdot 10^0$	–	–	$3.636 \cdot 10^{-2}$	–	–	$3.542 \cdot 10^{-2}$	–	–
D4	1	4	120	248	$2.902 \cdot 10^0$	3.34	1.74	$4.578 \cdot 10^{-3}$	7.94	2.99	$5.947 \cdot 10^{-3}$	5.96	2.57
D4	2	16	400	912	$2.477 \cdot 10^{-1}$	11.71	3.55	$6.819 \cdot 10^{-4}$	6.71	2.75	$6.349 \cdot 10^{-4}$	9.37	3.23
D4	3	64	1440	3488	$1.672 \cdot 10^{-2}$	14.82	3.89	$2.519 \cdot 10^{-5}$	27.07	4.76	$2.913 \cdot 10^{-5}$	21.80	4.45
D4	4	256	5440	13632	$1.065 \cdot 10^{-3}$	15.70	3.97	$2.832 \cdot 10^{-7}$	88.95	6.48	$3.037 \cdot 10^{-7}$	95.92	6.58
D4	5	1024	21120	53888	$6.686 \cdot 10^{-5}$	15.93	3.99	$3.592 \cdot 10^{-9}$	78.84	6.30	$3.401 \cdot 10^{-9}$	89.30	6.48
D4	6	4096	83200	214272	$4.183 \cdot 10^{-6}$	15.98	4.00	$8.352 \cdot 10^{-11}$	43.01	5.43	$6.36 \cdot 10^{-11}$	53.47	5.74
D5	0	1	48	98	$8.902 \cdot 10^0$	–	–	$3.456 \cdot 10^{-2}$	–	–	$3.367 \cdot 10^{-2}$	–	–
D5	1	4	144	344	$9.818 \cdot 10^{-1}$	9.07	3.18	$1.803 \cdot 10^{-3}$	19.17	4.26	$1.856 \cdot 10^{-3}$	18.14	4.18
D5	2	16	480	1280	$4.011 \cdot 10^{-2}$	24.48	4.61	$8.074 \cdot 10^{-5}$	22.33	4.48	$8.981 \cdot 10^{-5}$	20.67	4.37
D5	3	64	1728	4928	$1.337 \cdot 10^{-3}$	30.00	4.91	$1.022 \cdot 10^{-6}$	79.00	6.30	$1.154 \cdot 10^{-6}$	77.82	6.28
D5	4	256	6528	19328	$4.248 \cdot 10^{-5}$	31.47	4.98	$1.231 \cdot 10^{-8}$	83.02	6.38	$1.126 \cdot 10^{-8}$	102.49	6.68
D5	5	1024	25344	76544	$1.333 \cdot 10^{-6}$	31.87	4.99	$1.671 \cdot 10^{-10}$	73.67	6.20	$1.045 \cdot 10^{-10}$	107.75	6.75
D5	6	4096	99840	304640	$4.17 \cdot 10^{-8}$	31.97	5.00	$1.07 \cdot 10^{-10}$	1.56	0.64	$1.108 \cdot 10^{-10}$	0.94	–0.08

Table 5.6: Convergence results for the DPG method according to TABLE 5.3.

5.2.1. A smooth example

conf	level	cells	DoFs	all DoFs	L ₂ -error	rate	order	L ₂ -error (mean)	rate	order	L ₁ -error (mean)	rate	order
D1 ⁺	0	1	16	24	9.738 · 10 ⁰	–	–	2.716 · 10 ⁻²	–	–	3.41 · 10 ⁻²	–	–
D1 ⁺	1	4	48	80	9.749 · 10 ⁰	1.00	0.00	1.255 · 10 ⁻¹	0.22	-2.21	1.575 · 10 ⁻¹	0.22	-2.21
D1 ⁺	2	16	160	288	4.791 · 10 ⁰	2.03	1.03	5.386 · 10 ⁻¹	0.23	-2.10	4.782 · 10 ⁻¹	0.33	-1.60
D1 ⁺	3	64	576	1088	1.34 · 10 ⁰	3.57	1.84	9.744 · 10 ⁻²	5.53	2.47	9.755 · 10 ⁻²	4.90	2.29
D1 ⁺	4	256	2176	4224	3.555 · 10 ⁻¹	3.77	1.91	1.299 · 10 ⁻²	7.50	2.91	1.226 · 10 ⁻²	7.96	2.99
D1 ⁺	5	1024	8448	16640	9.042 · 10 ⁻²	3.93	1.98	2.313 · 10 ⁻³	5.62	2.49	2.169 · 10 ⁻³	5.65	2.50
D1 ⁺	6	4096	33280	66048	2.271 · 10 ⁻²	3.98	1.99	5.199 · 10 ⁻⁴	4.45	2.15	4.867 · 10 ⁻⁴	4.46	2.16
D2 ⁺	0	1	24	42	9.779 · 10 ⁰	–	–	3.323 · 10 ⁻²	–	–	3.238 · 10 ⁻²	–	–
D2 ⁺	1	4	72	144	7.247 · 10 ⁰	1.35	0.43	5.875 · 10 ⁻²	0.57	-0.82	8.04 · 10 ⁻²	0.40	-1.31
D2 ⁺	2	16	240	528	1.22 · 10 ⁰	5.94	2.57	2.403 · 10 ⁻²	2.44	1.29	2.647 · 10 ⁻²	3.04	1.60
D2 ⁺	3	64	864	2016	1.817 · 10 ⁻¹	6.72	2.75	7.185 · 10 ⁻³	3.34	1.74	7.074 · 10 ⁻³	3.74	1.90
D2 ⁺	4	256	3264	7872	2.459 · 10 ⁻²	7.39	2.89	3.037 · 10 ⁻⁴	23.66	4.56	3.317 · 10 ⁻⁴	21.33	4.42
D2 ⁺	5	1024	12672	31104	3.06 · 10 ⁻³	8.04	3.01	8.906 · 10 ⁻⁶	34.10	5.09	9.428 · 10 ⁻⁶	35.18	5.14
D2 ⁺	6	4096	49920	123648	3.851 · 10 ⁻⁴	7.95	2.99	4.38 · 10 ⁻⁷	20.33	4.35	3.246 · 10 ⁻⁷	29.04	4.86
D3 ⁺	0	1	32	64	9.746 · 10 ⁰	–	–	7.02 · 10 ⁻²	–	–	6.84 · 10 ⁻²	–	–
D3 ⁺	1	4	96	224	2.913 · 10 ⁰	3.35	1.74	4.278 · 10 ⁻³	16.41	4.04	4.726 · 10 ⁻³	14.47	3.86
D3 ⁺	2	16	320	832	2.681 · 10 ⁻¹	10.87	3.44	7.057 · 10 ⁻³	0.61	-0.72	6.509 · 10 ⁻³	0.73	-0.46
D3 ⁺	3	64	1152	3200	1.809 · 10 ⁻²	14.82	3.89	3.75 · 10 ⁻⁴	18.82	4.23	3.659 · 10 ⁻⁴	17.79	4.15
D3 ⁺	4	256	4352	12544	1.258 · 10 ⁻³	14.38	3.85	1.526 · 10 ⁻⁵	24.57	4.62	1.556 · 10 ⁻⁵	23.52	4.56
D3 ⁺	5	1024	16896	49664	7.962 · 10 ⁻⁵	15.80	3.98	2.927 · 10 ⁻⁷	52.14	5.71	3.256 · 10 ⁻⁷	47.79	5.58
D3 ⁺	6	4096	66560	197632	4.834 · 10 ⁻⁶	16.47	4.04	4.919 · 10 ⁻⁹	59.51	5.90	5.717 · 10 ⁻⁹	56.95	5.83
D4 ⁺	0	1	40	90	9.252 · 10 ⁰	–	–	4.633 · 10 ⁻²	–	–	4.514 · 10 ⁻²	–	–
D4 ⁺	1	4	120	320	1.019 · 10 ⁰	9.08	3.18	3.921 · 10 ⁻³	11.82	3.56	4.64 · 10 ⁻³	9.73	3.28
D4 ⁺	2	16	400	1200	4.042 · 10 ⁻²	25.22	4.66	2.078 · 10 ⁻⁴	18.87	4.24	2.318 · 10 ⁻⁴	20.02	4.32
D4 ⁺	3	64	1440	4640	1.475 · 10 ⁻³	27.40	4.78	2.425 · 10 ⁻⁵	8.57	3.10	2.763 · 10 ⁻⁵	8.39	3.07
D4 ⁺	4	256	5440	18240	4.644 · 10 ⁻⁵	31.76	4.99	2.47 · 10 ⁻⁷	98.18	6.62	2.713 · 10 ⁻⁷	101.84	6.67
D4 ⁺	5	1024	21120	72320	1.427 · 10 ⁻⁶	32.54	5.02	1.937 · 10 ⁻⁹	127.51	7.00	2.174 · 10 ⁻⁹	124.79	6.96
D4 ⁺	6	4096	83200	288000	4.426 · 10 ⁻⁸	32.24	5.01	2.914 · 10 ⁻¹¹	66.47	6.06	3.057 · 10 ⁻¹¹	71.12	6.15
D5 ⁺	0	1	48	120	6.244 · 10 ⁰	–	–	1.689 · 10 ⁻²	–	–	1.646 · 10 ⁻²	–	–
D5 ⁺	1	4	144	432	2.752 · 10 ⁻¹	22.69	4.50	3.525 · 10 ⁻⁴	47.92	5.58	4.192 · 10 ⁻⁴	39.25	5.30
D5 ⁺	2	16	480	1632	5.851 · 10 ⁻³	47.03	5.56	6.564 · 10 ⁻⁵	5.37	2.43	7.358 · 10 ⁻⁵	5.70	2.51
D5 ⁺	3	64	1728	6336	9.958 · 10 ⁻⁵	58.76	5.88	1.113 · 10 ⁻⁶	58.98	5.88	1.28 · 10 ⁻⁶	57.48	5.85
D5 ⁺	4	256	6528	24960	1.598 · 10 ⁻⁶	62.31	5.96	6.692 · 10 ⁻⁹	166.32	7.38	7.84 · 10 ⁻⁹	163.26	7.35
D5 ⁺	5	1024	25344	99072	2.558 · 10 ⁻⁸	62.47	5.97	4.602 · 10 ⁻¹¹	145.42	7.18	5.283 · 10 ⁻¹¹	148.40	7.21
D5 ⁺	6	4096	99840	394752	1.634 · 10 ⁻⁹	15.65	3.97	1.231 · 10 ⁻¹⁰	0.37	-1.42	1.245 · 10 ⁻¹⁰	0.42	-1.24

Table 5.7: Convergence results for the DPG method according to TABLE 5.4.

5.2. Numerical examples (1D)

conf	level	cells	DoFs	all DoFs	L ₂ -error	rate	order	L ₂ -error (mean)	rate	order	L ₁ -error (mean)	rate	order
W2	0	1	12	30	$1.073 \cdot 10^1$	—	—	$3.952 \cdot 10^0$	—	—	$5.012 \cdot 10^0$	—	—
W2	1	4	36	108	$1.016 \cdot 10^1$	1.06	0.08	$2.714 \cdot 10^0$	1.46	0.54	$3.073 \cdot 10^0$	1.63	0.71
W2	2	16	120	408	$1.967 \cdot 10^1$	0.52	-0.95	$1.062 \cdot 10^1$	0.26	-1.97	$1.333 \cdot 10^1$	0.23	-2.12
W2	3	64	432	1584	$1.273 \cdot 10^0$	15.45	3.95	$6.515 \cdot 10^{-1}$	16.30	4.03	$7.39 \cdot 10^{-1}$	18.04	4.17
W2	4	256	1632	6240	$1.556 \cdot 10^{-1}$	8.18	3.03	$9.248 \cdot 10^{-2}$	7.04	2.82	$1.005 \cdot 10^{-1}$	7.35	2.88
W2	5	1024	6336	24768	$2.378 \cdot 10^{-2}$	6.54	2.71	$1.817 \cdot 10^{-2}$	5.09	2.35	$1.819 \cdot 10^{-2}$	5.53	2.47
W2	6	4096	24960	98688	$4.619 \cdot 10^{-3}$	5.15	2.36	$4.205 \cdot 10^{-3}$	4.32	2.11	$4.012 \cdot 10^{-3}$	4.53	2.18
W3	0	1	20	52	$1.407 \cdot 10^1$	—	—	$2.318 \cdot 10^0$	—	—	$2.859 \cdot 10^0$	—	—
W3	1	4	60	188	$1.005 \cdot 10^1$	1.40	0.49	$3.481 \cdot 10^0$	0.67	-0.59	$4.58 \cdot 10^0$	0.62	-0.68
W3	2	16	200	712	$7.433 \cdot 10^{-1}$	13.52	3.76	$1.98 \cdot 10^{-1}$	17.58	4.14	$2.293 \cdot 10^{-1}$	19.97	4.32
W3	3	64	720	2768	$1.556 \cdot 10^{-1}$	4.78	2.26	$3.515 \cdot 10^{-2}$	5.63	2.49	$3.847 \cdot 10^{-2}$	5.96	2.58
W3	4	256	2720	10912	$8.307 \cdot 10^{-3}$	18.73	4.23	$2.013 \cdot 10^{-3}$	17.46	4.13	$2.367 \cdot 10^{-3}$	16.25	4.02
W3	5	1024	10560	43328	$4.992 \cdot 10^{-4}$	16.64	4.06	$1.374 \cdot 10^{-4}$	14.65	3.87	$1.603 \cdot 10^{-4}$	14.77	3.88
W3	6	4096	41600	172672	$3.093 \cdot 10^{-5}$	16.14	4.01	$8.863 \cdot 10^{-6}$	15.50	3.95	$1.035 \cdot 10^{-5}$	15.49	3.95
W4	0	1	24	74	$1.043 \cdot 10^1$	—	—	$2.052 \cdot 10^0$	—	—	$2.618 \cdot 10^0$	—	—
W4	1	4	72	272	$3.885 \cdot 10^0$	2.69	1.43	$3.323 \cdot 10^{-1}$	6.17	2.63	$3.933 \cdot 10^{-1}$	6.66	2.74
W4	2	16	240	1040	$3.863 \cdot 10^{-1}$	10.06	3.33	$6.02 \cdot 10^{-2}$	5.52	2.47	$6.563 \cdot 10^{-2}$	5.99	2.58
W4	3	64	864	4064	$1.994 \cdot 10^{-2}$	19.38	4.28	$1.959 \cdot 10^{-3}$	30.73	4.94	$2.212 \cdot 10^{-3}$	29.67	4.89
W4	4	256	3264	16064	$8.245 \cdot 10^{-4}$	24.18	4.60	$6.562 \cdot 10^{-5}$	29.85	4.90	$6.984 \cdot 10^{-5}$	31.67	4.99
W4	5	1024	12672	63872	$2.484 \cdot 10^{-5}$	33.19	5.05	$4.696 \cdot 10^{-6}$	13.97	3.81	$4.811 \cdot 10^{-6}$	14.52	3.86
W4	6	4096	49920	254720	$8.151 \cdot 10^{-7}$	30.48	4.93	$3.739 \cdot 10^{-7}$	12.56	3.65	$3.54 \cdot 10^{-7}$	13.59	3.76
W5	0	1	28	100	$1.527 \cdot 10^1$	—	—	$4.109 \cdot 10^0$	—	—	$5.66 \cdot 10^0$	—	—
W5	1	4	84	372	$2.459 \cdot 10^0$	6.21	2.63	$2.609 \cdot 10^{-1}$	15.75	3.98	$2.775 \cdot 10^{-1}$	20.40	4.35
W5	2	16	280	1432	$4.868 \cdot 10^{-2}$	50.52	5.66	$9.257 \cdot 10^{-3}$	28.18	4.82	$1.147 \cdot 10^{-2}$	24.20	4.60
W5	3	64	1008	5616	$2.207 \cdot 10^{-3}$	22.06	4.46	$2.572 \cdot 10^{-4}$	35.99	5.17	$2.787 \cdot 10^{-4}$	41.14	5.36
W5	4	256	3808	22240	$6.265 \cdot 10^{-5}$	35.23	5.14	$4.124 \cdot 10^{-6}$	62.37	5.96	$4.498 \cdot 10^{-6}$	61.96	5.95
W5	5	1024	14784	88512	$8.828 \cdot 10^{-7}$	70.97	6.15	$6.84 \cdot 10^{-8}$	60.29	5.91	$7.268 \cdot 10^{-8}$	61.89	5.95
W5	6	4096	58240	353152	$1.21 \cdot 10^{-8}$	72.96	6.19	$1.002 \cdot 10^{-9}$	68.26	6.09	$9.9 \cdot 10^{-10}$	73.41	6.20
W6	0	1	32	130	$1.239 \cdot 10^1$	—	—	$5.993 \cdot 10^{-1}$	—	—	$6.25 \cdot 10^{-1}$	—	—
W6	1	4	96	488	$4.229 \cdot 10^{-1}$	29.29	4.87	$5.217 \cdot 10^{-2}$	11.49	3.52	$5.797 \cdot 10^{-2}$	10.78	3.43
W6	2	16	320	1888	$1.668 \cdot 10^{-2}$	25.35	4.66	$4.604 \cdot 10^{-4}$	113.32	6.83	$5.128 \cdot 10^{-4}$	113.05	6.82
W6	3	64	1152	7424	$1.83 \cdot 10^{-4}$	91.16	6.51	$1.906 \cdot 10^{-5}$	24.16	4.59	$2.136 \cdot 10^{-5}$	24.01	4.59
W6	4	256	4352	29440	$1.961 \cdot 10^{-6}$	93.32	6.55	$5.581 \cdot 10^{-7}$	34.15	5.09	$5.457 \cdot 10^{-7}$	39.14	5.29
W6	5	1024	16896	117248	$2.172 \cdot 10^{-8}$	90.29	6.50	$1.394 \cdot 10^{-8}$	40.03	5.32	$1.298 \cdot 10^{-8}$	42.04	5.39
W6	6	4096	66560	467968	$4.494 \cdot 10^{-9}$	4.83	2.27	$4.467 \cdot 10^{-9}$	3.12	1.64	$4.687 \cdot 10^{-9}$	2.77	1.47
W7	0	1	36	164	$1.191 \cdot 10^1$	—	—	$1.217 \cdot 10^0$	—	—	$1.676 \cdot 10^0$	—	—
W7	1	4	108	620	$5.091 \cdot 10^{-1}$	23.40	4.55	$7.255 \cdot 10^{-2}$	16.78	4.07	$9.339 \cdot 10^{-2}$	17.95	4.17
W7	2	16	360	2408	$1.656 \cdot 10^{-3}$	307.43	8.26	$1.472 \cdot 10^{-4}$	492.88	8.95	$1.517 \cdot 10^{-4}$	615.59	9.27
W7	3	64	1296	9488	$1.084 \cdot 10^{-5}$	152.77	7.26	$1.036 \cdot 10^{-6}$	142.08	7.15	$1.12 \cdot 10^{-6}$	135.45	7.08
W7	4	256	4896	37664	$4.226 \cdot 10^{-8}$	256.50	8.00	$4.436 \cdot 10^{-9}$	233.54	7.87	$4.557 \cdot 10^{-9}$	245.78	7.94
W7	5	1024	19008	150080	$1.44 \cdot 10^{-9}$	29.35	4.88	$1.338 \cdot 10^{-9}$	3.32	1.73	$1.375 \cdot 10^{-9}$	3.31	1.73
W7	6	4096	74880	599168	$3.435 \cdot 10^{-9}$	0.42	-1.25	$3.202 \cdot 10^{-9}$	0.42	-1.26	$3.301 \cdot 10^{-9}$	0.42	-1.26
W8	0	1	40	202	$7.269 \cdot 10^0$	—	—	$2.73 \cdot 10^{-1}$	—	—	$3.4 \cdot 10^{-1}$	—	—
W8	1	4	120	768	$2.984 \cdot 10^{-2}$	243.63	7.93	$2.32 \cdot 10^{-3}$	117.66	6.88	$2.522 \cdot 10^{-3}$	134.82	7.08
W8	2	16	400	2992	$2.58 \cdot 10^{-4}$	115.64	6.85	$1.425 \cdot 10^{-5}$	162.81	7.35	$1.544 \cdot 10^{-5}$	163.34	7.35
W8	3	64	1440	11808	$7.597 \cdot 10^{-7}$	339.60	8.41	$5.608 \cdot 10^{-8}$	254.10	7.99	$5.464 \cdot 10^{-8}$	282.58	8.14
W8	4	256	5440	46912	$4.923 \cdot 10^{-9}$	154.31	7.27	$4.152 \cdot 10^{-9}$	13.51	3.76	$4.306 \cdot 10^{-9}$	12.69	3.67
W8	5	1024	21120	187008	$1.524 \cdot 10^{-8}$	0.32	-1.63	$1.47 \cdot 10^{-8}$	0.28	-1.82	$1.547 \cdot 10^{-8}$	0.28	-1.85
W8	6	4096	83200	746752	$5.279 \cdot 10^{-8}$	0.29	-1.79	$5.207 \cdot 10^{-8}$	0.28	-1.82	$5.477 \cdot 10^{-8}$	0.28	-1.82

Table 5.8: Convergence results for the WC method according to TABLE 5.5.

5.2.2 A traveling wave in a homogeneous medium

In this example, we consider the superposition of two traveling wave fronts with opposite traveling directions, both of which are reflected five times at the boundaries of $\Omega = (0, 1)$ due to homogeneous DIRICHLET boundary values in p . The initial value $(p_0, \mathbf{v}_0) = (a_0^{\text{left}} - a_0^{\text{right}}, -a_0^{\text{left}} - a_0^{\text{right}})$ is given by

$$a_0^{\text{left}}(x) = \begin{cases} 2 \cos\left(\frac{x - m^{\text{left}}}{w^{\text{left}}}\right)^k, & x \in m^{\text{left}} + (-w^{\text{left}}, w^{\text{left}}), \\ 0, & \text{else,} \end{cases}$$

$$a_0^{\text{right}}(x) = \begin{cases} \cos\left(\frac{x - m^{\text{right}}}{w^{\text{right}}}\right)^k, & x \in m^{\text{right}} + (-w^{\text{right}}, w^{\text{right}}), \\ 0, & \text{else,} \end{cases}$$

with $k = 7$, $m^{\text{left}} = \frac{7}{9}$, $w^{\text{left}} = 0.25 \cdot \frac{3}{\pi}$ and $m^{\text{right}} = \frac{2}{9}$, $w^{\text{right}} = 0.4 \cdot \frac{3}{\pi}$, resulting in $(p_0, \mathbf{v}_0) \in C^6((0, 1), \mathbb{R}^2)$. By setting $\rho(x) = \kappa(x) = 1$, $x \in \Omega$, the wave travels with speed $c = 1$. To prevent alignment of the characteristics with the mesh, we used a grid with six cells congruent to $\Omega \times (0, \frac{3}{\pi})$ of the space-time domain $Q = \Omega \times (0, T)$, $T = 6 \cdot \frac{3}{\pi}$.

Using a periodic extension of the initial value, we obtain an analytical solution as depicted in FIGURE 5.6.

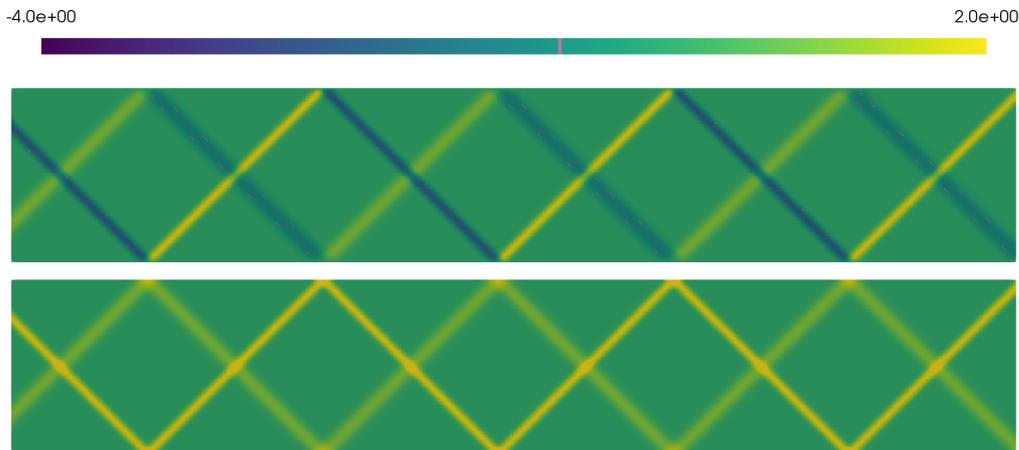


Figure 5.6: Two traveling waves with opposite directions. The top plot shows p and the plot at the bottom shows \mathbf{v} in space-time. The horizontal axis is time, left to right.

FIGURE 5.7 to FIGURE 5.9 visualize the convergence results. For a detailed comparison, the reader may refer to TABLE 5.9, TABLE 5.10 and TABLE 5.11.

5.2. Numerical examples (1D)

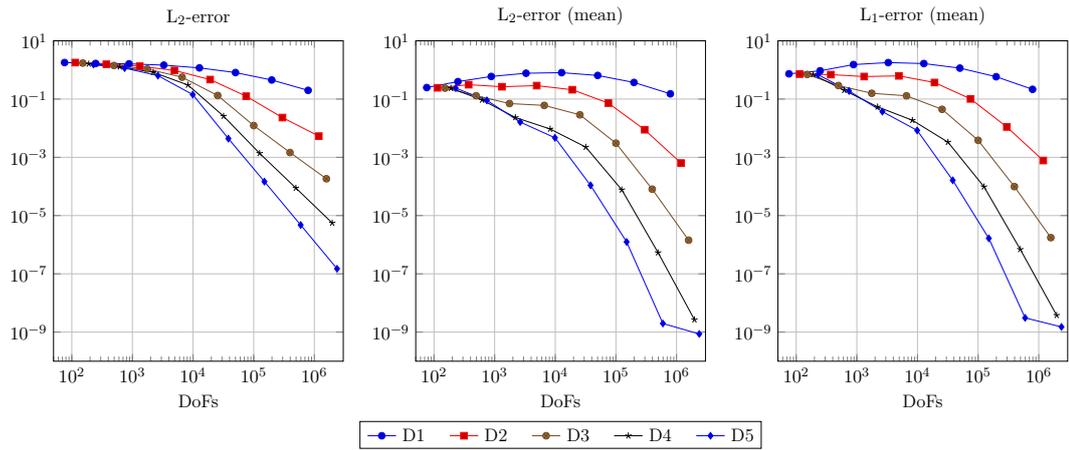


Figure 5.7: Convergence results for DPG with configurations according to TABLE 5.3.

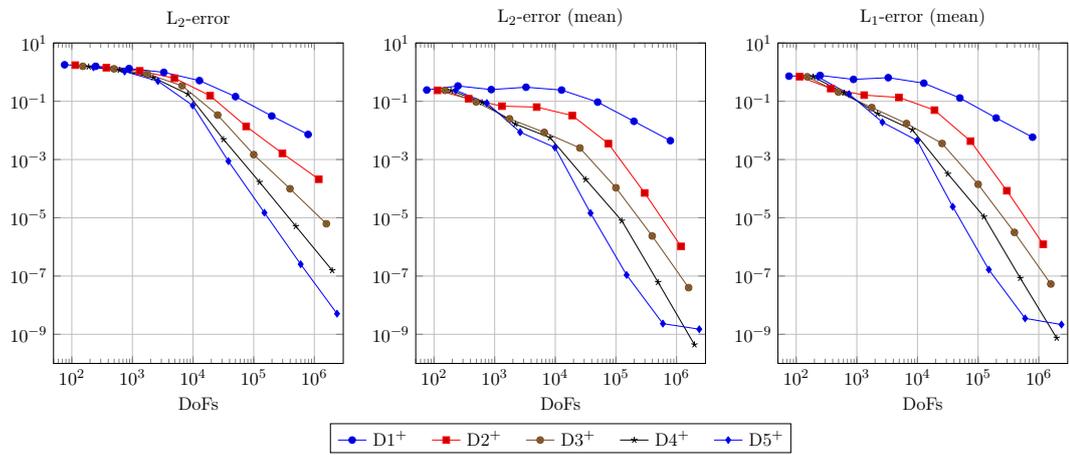


Figure 5.8: Convergence results for DPG with configurations according to TABLE 5.4.

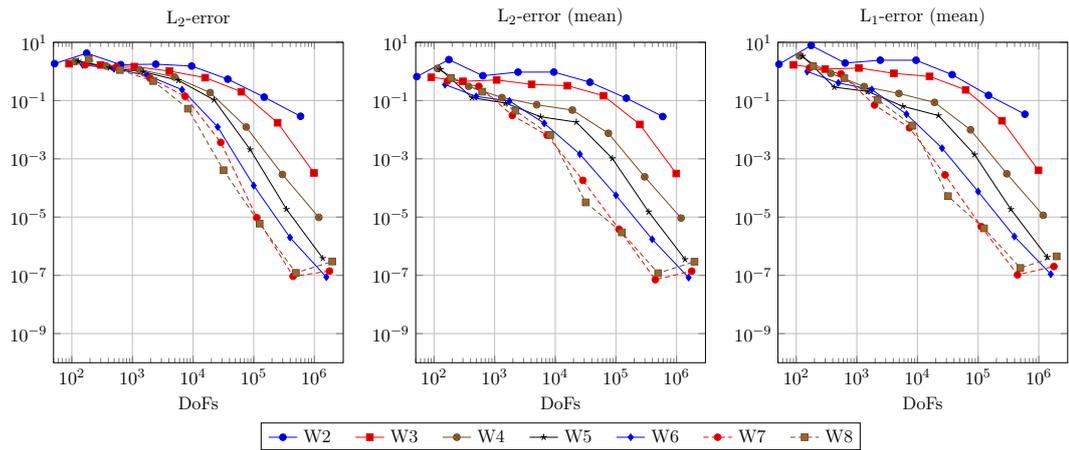


Figure 5.9: Convergence results for WC with configurations according to TABLE 5.5.

Discussion: DPG. From the experiments, we see that the predicted convergence rates are also achieved in practice for this non-trivial example with boundary reflections.

Since this solution features a fine structure, it is reasonable to expect that the error does not change heavily in a pre-asymptotic regime as long as the mesh is not fine enough to resolve the solution's structure. This is also what we observe in the experiment. However, after leaving this regime the method converges as predicted until the error stalls again due to round-off errors.

Again, we observe increased rates for the mean-values except for the lowest order method.

Discussion: DPG – increased cell degree. We observe by 1 increased convergence rates in this example with boundary reflections as well.

As in the configuration above, the convergence rate of the mean values is increased for the $L_1(Q)$ as well as the $L_2(Q)$ norm in all examples except for the lowest order scheme.

Interestingly, when looking at the convergence rates of the mean values, the schemes $D2^+$ and $D3^+$ show comparable rates despite using different polynomial degrees inside the cells.

Discussion: weakly conforming Least-Squares. In this example, the weakly conforming Least-Squares method does not behave as foreseeable as in the smooth example.

All configurations show oscillating convergence rates and for W5 to W8 we observe round-off problems occurring earlier in comparison to the DPG method.

Also the high-order schemes W7 and W8 do not show improved approximation qualities. However, since the solution is a $C^6(Q, \mathbb{R}^2)$ function only, we do not expect significant improvements for these schemes compared to the other variants.

5.2. Numerical examples (1D)

conf	level	cells	DoFs	all DoFs	L ₂ -error	rate	order	L ₂ -error (mean)	rate	order	L ₁ -error (mean)	rate	order
D1	0	6	76	88	$1.794 \cdot 10^0$	—	—	$2.483 \cdot 10^{-1}$	—	—	$7.364 \cdot 10^{-1}$	—	—
D1	1	24	248	296	$1.668 \cdot 10^0$	1.08	0.11	$3.975 \cdot 10^{-1}$	0.62	-0.68	$9.351 \cdot 10^{-1}$	0.79	-0.34
D1	2	96	880	1072	$1.617 \cdot 10^0$	1.03	0.04	$5.972 \cdot 10^{-1}$	0.67	-0.59	$1.526 \cdot 10^0$	0.61	-0.71
D1	3	384	3296	4064	$1.462 \cdot 10^0$	1.11	0.15	$7.706 \cdot 10^{-1}$	0.77	-0.37	$1.787 \cdot 10^0$	0.85	-0.23
D1	4	1536	12736	15808	$1.176 \cdot 10^0$	1.24	0.31	$8.063 \cdot 10^{-1}$	0.96	-0.07	$1.657 \cdot 10^0$	1.08	0.11
D1	5	6144	50048	62336	$8.119 \cdot 10^{-1}$	1.45	0.53	$6.472 \cdot 10^{-1}$	1.25	0.32	$1.154 \cdot 10^0$	1.44	0.52
D1	6	24576	198400	247552	$4.513 \cdot 10^{-1}$	1.80	0.85	$3.719 \cdot 10^{-1}$	1.74	0.80	$5.844 \cdot 10^{-1}$	1.97	0.98
D1	7	98304	790016	986624	$1.991 \cdot 10^{-1}$	2.27	1.18	$1.514 \cdot 10^{-1}$	2.46	1.30	$2.158 \cdot 10^{-1}$	2.71	1.44
D2	0	6	114	162	$1.79 \cdot 10^0$	—	—	$2.435 \cdot 10^{-1}$	—	—	$7.213 \cdot 10^{-1}$	—	—
D2	1	24	372	564	$1.566 \cdot 10^0$	1.14	0.19	$3.113 \cdot 10^{-1}$	0.78	-0.35	$6.915 \cdot 10^{-1}$	1.04	0.06
D2	2	96	1320	2088	$1.347 \cdot 10^0$	1.16	0.22	$2.652 \cdot 10^{-1}$	1.17	0.23	$5.935 \cdot 10^{-1}$	1.17	0.22
D2	3	384	4944	8016	$9.59 \cdot 10^{-1}$	1.40	0.49	$2.896 \cdot 10^{-1}$	0.92	-0.13	$6.26 \cdot 10^{-1}$	0.95	-0.08
D2	4	1536	19104	31392	$4.67 \cdot 10^{-1}$	2.05	1.04	$2.091 \cdot 10^{-1}$	1.38	0.47	$3.697 \cdot 10^{-1}$	1.69	0.76
D2	5	6144	75072	124224	$1.255 \cdot 10^{-1}$	3.72	1.90	$7.274 \cdot 10^{-2}$	2.87	1.52	$1.013 \cdot 10^{-1}$	3.65	1.87
D2	6	24576	297600	494208	$2.317 \cdot 10^{-2}$	5.42	2.44	$8.853 \cdot 10^{-3}$	8.22	3.04	$1.103 \cdot 10^{-2}$	9.19	3.20
D2	7	98304	1185024	1971456	$5.336 \cdot 10^{-3}$	4.34	2.12	$6.363 \cdot 10^{-4}$	13.91	3.80	$7.767 \cdot 10^{-4}$	14.20	3.83
D3	0	6	152	260	$1.722 \cdot 10^0$	—	—	$2.375 \cdot 10^{-1}$	—	—	$6.986 \cdot 10^{-1}$	—	—
D3	1	24	496	928	$1.42 \cdot 10^0$	1.21	0.28	$1.303 \cdot 10^{-1}$	1.82	0.87	$2.914 \cdot 10^{-1}$	2.40	1.26
D3	2	96	1760	3488	$1.074 \cdot 10^0$	1.32	0.40	$7.012 \cdot 10^{-2}$	1.86	0.89	$1.58 \cdot 10^{-1}$	1.84	0.88
D3	3	384	6592	13504	$5.633 \cdot 10^{-1}$	1.91	0.93	$6.058 \cdot 10^{-2}$	1.16	0.21	$1.294 \cdot 10^{-1}$	1.22	0.29
D3	4	1536	25472	53120	$1.327 \cdot 10^{-1}$	4.24	2.09	$2.882 \cdot 10^{-2}$	2.10	1.07	$4.442 \cdot 10^{-2}$	2.91	1.54
D3	5	6144	100096	210688	$1.231 \cdot 10^{-2}$	10.79	3.43	$3.057 \cdot 10^{-3}$	9.43	3.24	$3.845 \cdot 10^{-3}$	11.55	3.53
D3	6	24576	396800	839168	$1.461 \cdot 10^{-3}$	8.42	3.07	$8.062 \cdot 10^{-5}$	37.92	5.25	$9.867 \cdot 10^{-5}$	38.97	5.28
D3	7	98304	1580032	3349504	$1.838 \cdot 10^{-4}$	7.95	2.99	$1.431 \cdot 10^{-6}$	56.34	5.82	$1.75 \cdot 10^{-6}$	56.38	5.82
D4	0	6	190	382	$1.604 \cdot 10^0$	—	—	$2.373 \cdot 10^{-1}$	—	—	$6.928 \cdot 10^{-1}$	—	—
D4	1	24	620	1388	$1.278 \cdot 10^0$	1.26	0.33	$9.286 \cdot 10^{-2}$	2.56	1.35	$1.998 \cdot 10^{-1}$	3.47	1.79
D4	2	96	2200	5272	$8.309 \cdot 10^{-1}$	1.54	0.62	$2.33 \cdot 10^{-2}$	3.99	2.00	$5.396 \cdot 10^{-2}$	3.70	1.89
D4	3	384	8240	20528	$3.043 \cdot 10^{-1}$	2.73	1.45	$9.371 \cdot 10^{-3}$	2.49	1.31	$1.875 \cdot 10^{-2}$	2.88	1.53
D4	4	1536	31840	80992	$2.586 \cdot 10^{-2}$	11.77	3.56	$2.244 \cdot 10^{-3}$	4.18	2.06	$3.312 \cdot 10^{-3}$	5.66	2.50
D4	5	6144	125120	321728	$1.372 \cdot 10^{-3}$	18.85	4.24	$7.652 \cdot 10^{-5}$	29.32	4.87	$9.759 \cdot 10^{-5}$	33.94	5.09
D4	6	24576	496000	1282432	$8.828 \cdot 10^{-5}$	15.54	3.96	$5.336 \cdot 10^{-7}$	143.40	7.16	$6.903 \cdot 10^{-7}$	141.37	7.14
D4	7	98304	1975040	5120768	$5.565 \cdot 10^{-6}$	15.86	3.99	$2.666 \cdot 10^{-9}$	200.15	7.65	$3.773 \cdot 10^{-9}$	182.95	7.52
D5	0	6	228	528	$1.534 \cdot 10^0$	—	—	$2.374 \cdot 10^{-1}$	—	—	$6.923 \cdot 10^{-1}$	—	—
D5	1	24	744	1944	$1.157 \cdot 10^0$	1.33	0.41	$9.027 \cdot 10^{-2}$	2.63	1.40	$1.893 \cdot 10^{-1}$	3.66	1.87
D5	2	96	2640	7440	$6.375 \cdot 10^{-1}$	1.81	0.86	$1.623 \cdot 10^{-2}$	5.56	2.48	$3.704 \cdot 10^{-2}$	5.11	2.35
D5	3	384	9888	29088	$1.424 \cdot 10^{-1}$	4.48	2.16	$4.682 \cdot 10^{-3}$	3.47	1.79	$8.396 \cdot 10^{-3}$	4.41	2.14
D5	4	1536	38208	115008	$4.399 \cdot 10^{-3}$	32.37	5.02	$1.091 \cdot 10^{-4}$	42.91	5.42	$1.626 \cdot 10^{-4}$	51.64	5.69
D5	5	6144	150144	457344	$1.468 \cdot 10^{-4}$	29.97	4.91	$1.246 \cdot 10^{-6}$	87.56	6.45	$1.657 \cdot 10^{-6}$	98.13	6.62
D5	6	24576	595200	1824000	$4.751 \cdot 10^{-6}$	30.90	4.95	$1.971 \cdot 10^{-9}$	632.16	9.31	$3.083 \cdot 10^{-9}$	537.46	9.07
D5	7	98304	2370048	7285248	$1.498 \cdot 10^{-7}$	31.72	4.99	$8.633 \cdot 10^{-10}$	2.28	1.19	$1.505 \cdot 10^{-9}$	2.05	1.03

Table 5.9: Convergence results for the DPG method according to TABLE 5.3.

5.2.2. A traveling wave in a homogeneous medium

conf	level	cells	DoFs	all DoFs	L ₂ -error	rate	order	L ₂ -error (mean)	rate	order	L ₁ -error (mean)	rate	order
D1 ⁺	0	6	76	124	$1.791 \cdot 10^0$	—	—	$2.436 \cdot 10^{-1}$	—	—	$7.282 \cdot 10^{-1}$	—	—
D1 ⁺	1	24	248	440	$1.591 \cdot 10^0$	1.13	0.17	$3.355 \cdot 10^{-1}$	0.73	-0.46	$7.706 \cdot 10^{-1}$	0.94	-0.08
D1 ⁺	2	96	880	1648	$1.329 \cdot 10^0$	1.20	0.26	$2.519 \cdot 10^{-1}$	1.33	0.41	$5.641 \cdot 10^{-1}$	1.37	0.45
D1 ⁺	3	384	3296	6368	$9.855 \cdot 10^{-1}$	1.35	0.43	$3.037 \cdot 10^{-1}$	0.83	-0.27	$6.462 \cdot 10^{-1}$	0.87	-0.20
D1 ⁺	4	1536	12736	25024	$5.126 \cdot 10^{-1}$	1.92	0.94	$2.424 \cdot 10^{-1}$	1.25	0.33	$4.179 \cdot 10^{-1}$	1.55	0.63
D1 ⁺	5	6144	50048	99200	$1.448 \cdot 10^{-1}$	3.54	1.82	$9.341 \cdot 10^{-2}$	2.60	1.38	$1.296 \cdot 10^{-1}$	3.22	1.69
D1 ⁺	6	24576	198400	395008	$3.122 \cdot 10^{-2}$	4.64	2.21	$2.048 \cdot 10^{-2}$	4.56	2.19	$2.667 \cdot 10^{-2}$	4.86	2.28
D1 ⁺	7	98304	790016	1576448	$7.258 \cdot 10^{-3}$	4.30	2.11	$4.441 \cdot 10^{-3}$	4.61	2.21	$5.846 \cdot 10^{-3}$	4.56	2.19
D2 ⁺	0	6	114	222	$1.738 \cdot 10^0$	—	—	$2.376 \cdot 10^{-1}$	—	—	$7.006 \cdot 10^{-1}$	—	—
D2 ⁺	1	24	372	804	$1.42 \cdot 10^0$	1.22	0.29	$1.224 \cdot 10^{-1}$	1.94	0.96	$2.735 \cdot 10^{-1}$	2.56	1.36
D2 ⁺	2	96	1320	3048	$1.106 \cdot 10^0$	1.28	0.36	$6.793 \cdot 10^{-2}$	1.80	0.85	$1.625 \cdot 10^{-1}$	1.68	0.75
D2 ⁺	3	384	4944	11856	$6.193 \cdot 10^{-1}$	1.79	0.84	$6.312 \cdot 10^{-2}$	1.08	0.11	$1.355 \cdot 10^{-1}$	1.20	0.26
D2 ⁺	4	1536	19104	46752	$1.556 \cdot 10^{-1}$	3.98	1.99	$3.235 \cdot 10^{-2}$	1.95	0.96	$4.966 \cdot 10^{-2}$	2.73	1.45
D2 ⁺	5	6144	75072	185664	$1.371 \cdot 10^{-2}$	11.36	3.51	$3.514 \cdot 10^{-3}$	9.21	3.20	$4.278 \cdot 10^{-3}$	11.61	3.54
D2 ⁺	6	24576	297600	739968	$1.627 \cdot 10^{-3}$	8.42	3.08	$7.118 \cdot 10^{-5}$	49.37	5.63	$8.488 \cdot 10^{-5}$	50.40	5.66
D2 ⁺	7	98304	1185024	2954496	$2.103 \cdot 10^{-4}$	7.74	2.95	$1.041 \cdot 10^{-6}$	68.38	6.10	$1.233 \cdot 10^{-6}$	68.84	6.11
D3 ⁺	0	6	152	344	$1.587 \cdot 10^0$	—	—	$2.372 \cdot 10^{-1}$	—	—	$6.928 \cdot 10^{-1}$	—	—
D3 ⁺	1	24	496	1264	$1.3 \cdot 10^0$	1.22	0.29	$9.481 \cdot 10^{-2}$	2.50	1.32	$2.067 \cdot 10^{-1}$	3.35	1.75
D3 ⁺	2	96	1760	4832	$8.641 \cdot 10^{-1}$	1.50	0.59	$2.496 \cdot 10^{-2}$	3.80	1.93	$6.094 \cdot 10^{-2}$	3.39	1.76
D3 ⁺	3	384	6592	18880	$3.407 \cdot 10^{-1}$	2.54	1.34	$8.535 \cdot 10^{-3}$	2.92	1.55	$1.727 \cdot 10^{-2}$	3.53	1.82
D3 ⁺	4	1536	25472	74624	$3.357 \cdot 10^{-2}$	10.15	3.34	$2.491 \cdot 10^{-3}$	3.43	1.78	$3.548 \cdot 10^{-3}$	4.87	2.28
D3 ⁺	5	6144	100096	296704	$1.472 \cdot 10^{-3}$	22.80	4.51	$1.073 \cdot 10^{-4}$	23.22	4.54	$1.408 \cdot 10^{-4}$	25.20	4.66
D3 ⁺	6	24576	396800	1183232	$9.937 \cdot 10^{-5}$	14.81	3.89	$2.372 \cdot 10^{-6}$	45.24	5.50	$3.15 \cdot 10^{-6}$	44.70	5.48
D3 ⁺	7	98304	1580032	4725760	$6.243 \cdot 10^{-6}$	15.92	3.99	$3.995 \cdot 10^{-8}$	59.37	5.89	$5.323 \cdot 10^{-8}$	59.18	5.89
D4 ⁺	0	6	190	490	$1.527 \cdot 10^0$	—	—	$2.374 \cdot 10^{-1}$	—	—	$6.914 \cdot 10^{-1}$	—	—
D4 ⁺	1	24	620	1820	$1.177 \cdot 10^0$	1.30	0.38	$9.169 \cdot 10^{-2}$	2.59	1.37	$1.925 \cdot 10^{-1}$	3.59	1.84
D4 ⁺	2	96	2200	7000	$6.524 \cdot 10^{-1}$	1.80	0.85	$1.65 \cdot 10^{-2}$	5.56	2.47	$3.744 \cdot 10^{-2}$	5.14	2.36
D4 ⁺	3	384	8240	27440	$1.756 \cdot 10^{-1}$	3.72	1.89	$5.626 \cdot 10^{-3}$	2.93	1.55	$1.049 \cdot 10^{-2}$	3.57	1.84
D4 ⁺	4	1536	31840	108640	$4.9 \cdot 10^{-3}$	35.83	5.16	$2.068 \cdot 10^{-4}$	27.21	4.77	$3.204 \cdot 10^{-4}$	32.73	5.03
D4 ⁺	5	6144	125120	432320	$1.676 \cdot 10^{-4}$	29.24	4.87	$7.891 \cdot 10^{-6}$	26.21	4.71	$1.084 \cdot 10^{-5}$	29.56	4.89
D4 ⁺	6	24576	496000	1724800	$5.109 \cdot 10^{-6}$	32.80	5.04	$6.12 \cdot 10^{-8}$	128.94	7.01	$8.454 \cdot 10^{-8}$	128.22	7.00
D4 ⁺	7	98304	1975040	6890240	$1.56 \cdot 10^{-7}$	32.75	5.03	$4.405 \cdot 10^{-10}$	138.93	7.12	$7.467 \cdot 10^{-10}$	113.22	6.82
D5 ⁺	0	6	228	660	$1.462 \cdot 10^0$	—	—	$2.375 \cdot 10^{-1}$	—	—	$6.91 \cdot 10^{-1}$	—	—
D5 ⁺	1	24	744	2472	$1.052 \cdot 10^0$	1.39	0.47	$8.721 \cdot 10^{-2}$	2.72	1.45	$1.767 \cdot 10^{-1}$	3.91	1.97
D5 ⁺	2	96	2640	9552	$4.926 \cdot 10^{-1}$	2.14	1.09	$8.642 \cdot 10^{-3}$	10.09	3.34	$1.905 \cdot 10^{-2}$	9.28	3.21
D5 ⁺	3	384	9888	37536	$7.142 \cdot 10^{-2}$	6.90	2.79	$2.586 \cdot 10^{-3}$	3.34	1.74	$4.474 \cdot 10^{-3}$	4.26	2.09
D5 ⁺	4	1536	38208	148800	$8.781 \cdot 10^{-4}$	81.33	6.35	$1.444 \cdot 10^{-5}$	179.09	7.49	$2.405 \cdot 10^{-5}$	186.03	7.54
D5 ⁺	5	6144	150144	592512	$1.493 \cdot 10^{-5}$	58.81	5.88	$1.086 \cdot 10^{-7}$	132.96	7.06	$1.661 \cdot 10^{-7}$	144.79	7.18
D5 ⁺	6	24576	595200	2364672	$2.545 \cdot 10^{-7}$	58.66	5.88	$2.32 \cdot 10^{-9}$	46.81	5.55	$3.524 \cdot 10^{-9}$	47.13	5.56
D5 ⁺	7	98304	2370048	9447936	$5.091 \cdot 10^{-9}$	49.99	5.64	$1.496 \cdot 10^{-9}$	1.55	0.63	$2.155 \cdot 10^{-9}$	1.64	0.71

Table 5.10: Convergence results for the DPG method according to TABLE 5.4.

5.2. Numerical examples (1D)

conf	level	cells	DoFs	all DoFs	L_2 -error	rate	order	L_2 -error (mean)	rate	order	L_1 -error (mean)	rate	order
W2	0	6	52	160	$1.847 \cdot 10^0$	—	—	$6.635 \cdot 10^{-1}$	—	—	$1.767 \cdot 10^0$	—	—
W2	1	24	176	608	$4.23 \cdot 10^0$	0.44	-1.20	$2.542 \cdot 10^0$	0.26	-1.94	$7.777 \cdot 10^0$	0.23	-2.14
W2	2	96	640	2368	$1.708 \cdot 10^0$	2.48	1.31	$7.122 \cdot 10^{-1}$	3.57	1.84	$1.94 \cdot 10^0$	4.01	2.00
W2	3	384	2432	9344	$1.772 \cdot 10^0$	0.96	-0.05	$9.463 \cdot 10^{-1}$	0.75	-0.41	$2.43 \cdot 10^0$	0.80	-0.33
W2	4	1536	9472	37120	$1.531 \cdot 10^0$	1.16	0.21	$9.502 \cdot 10^{-1}$	1.00	-0.01	$2.423 \cdot 10^0$	1.00	0.00
W2	5	6144	37376	147968	$5.434 \cdot 10^{-1}$	2.82	1.49	$4.264 \cdot 10^{-1}$	2.23	1.16	$7.733 \cdot 10^{-1}$	3.13	1.65
W2	6	24576	148480	590848	$1.313 \cdot 10^{-1}$	4.14	2.05	$1.21 \cdot 10^{-1}$	3.52	1.82	$1.513 \cdot 10^{-1}$	5.11	2.35
W2	7	98304	591872	2361344	$2.886 \cdot 10^{-2}$	4.55	2.19	$2.828 \cdot 10^{-2}$	4.28	2.10	$3.398 \cdot 10^{-2}$	4.45	2.16
W3	0	6	90	282	$1.835 \cdot 10^0$	—	—	$6.29 \cdot 10^{-1}$	—	—	$1.684 \cdot 10^0$	—	—
W3	1	24	300	1068	$1.67 \cdot 10^0$	1.10	0.14	$4.564 \cdot 10^{-1}$	1.38	0.46	$1.203 \cdot 10^0$	1.40	0.48
W3	2	96	1080	4152	$1.462 \cdot 10^0$	1.14	0.19	$5.103 \cdot 10^{-1}$	0.89	-0.16	$1.313 \cdot 10^0$	0.92	-0.13
W3	3	384	4080	16368	$1.028 \cdot 10^0$	1.42	0.51	$3.593 \cdot 10^{-1}$	1.42	0.51	$8.569 \cdot 10^{-1}$	1.53	0.62
W3	4	1536	15840	64992	$6.116 \cdot 10^{-1}$	1.68	0.75	$3.244 \cdot 10^{-1}$	1.11	0.15	$6.742 \cdot 10^{-1}$	1.27	0.35
W3	5	6144	62400	259008	$2.006 \cdot 10^{-1}$	3.05	1.61	$1.467 \cdot 10^{-1}$	2.21	1.14	$2.331 \cdot 10^{-1}$	2.89	1.53
W3	6	24576	247680	1034112	$1.73 \cdot 10^{-2}$	11.59	3.54	$1.514 \cdot 10^{-2}$	9.69	3.28	$2.024 \cdot 10^{-2}$	11.52	3.53
W3	7	98304	986880	4132608	$3.268 \cdot 10^{-4}$	52.94	5.73	$3.126 \cdot 10^{-4}$	48.42	5.60	$4.027 \cdot 10^{-4}$	50.27	5.65
W4	0	6	114	414	$2.172 \cdot 10^0$	—	—	$1.263 \cdot 10^0$	—	—	$3.402 \cdot 10^0$	—	—
W4	1	24	372	1572	$1.531 \cdot 10^0$	1.42	0.50	$3.071 \cdot 10^{-1}$	4.11	2.04	$8.566 \cdot 10^{-1}$	3.97	1.99
W4	2	96	1320	6120	$1.149 \cdot 10^0$	1.33	0.41	$1.27 \cdot 10^{-1}$	2.42	1.27	$3.027 \cdot 10^{-1}$	2.83	1.50
W4	3	384	4944	24144	$6.418 \cdot 10^{-1}$	1.79	0.84	$7.182 \cdot 10^{-2}$	1.77	0.82	$1.752 \cdot 10^{-1}$	1.73	0.79
W4	4	1536	19104	95904	$1.876 \cdot 10^{-1}$	3.42	1.77	$4.713 \cdot 10^{-2}$	1.52	0.61	$8.648 \cdot 10^{-2}$	2.03	1.02
W4	5	6144	75072	382272	$1.235 \cdot 10^{-2}$	15.19	3.93	$7.536 \cdot 10^{-3}$	6.25	2.64	$9.887 \cdot 10^{-3}$	8.75	3.13
W4	6	24576	297600	1526400	$2.897 \cdot 10^{-4}$	42.62	5.41	$2.407 \cdot 10^{-4}$	31.31	4.97	$3.064 \cdot 10^{-4}$	32.27	5.01
W4	7	98304	1185024	6100224	$9.749 \cdot 10^{-6}$	29.72	4.89	$9.234 \cdot 10^{-6}$	26.07	4.70	$1.151 \cdot 10^{-5}$	26.62	4.74
W5	0	6	128	560	$2.271 \cdot 10^0$	—	—	$1.175 \cdot 10^0$	—	—	$3.359 \cdot 10^0$	—	—
W5	1	24	424	2152	$1.334 \cdot 10^0$	1.70	0.77	$1.254 \cdot 10^{-1}$	9.37	3.23	$2.886 \cdot 10^{-1}$	11.64	3.54
W5	2	96	1520	8432	$9.644 \cdot 10^{-1}$	1.38	0.47	$8.177 \cdot 10^{-2}$	1.53	0.62	$2.052 \cdot 10^{-1}$	1.41	0.49
W5	3	384	5728	33376	$4.932 \cdot 10^{-1}$	1.96	0.97	$2.764 \cdot 10^{-2}$	2.96	1.56	$6.356 \cdot 10^{-2}$	3.23	1.69
W5	4	1536	22208	132800	$1.053 \cdot 10^{-1}$	4.68	2.23	$1.823 \cdot 10^{-2}$	1.52	0.60	$3.064 \cdot 10^{-2}$	2.07	1.05
W5	5	6144	87424	529792	$2.076 \cdot 10^{-3}$	50.74	5.67	$1.025 \cdot 10^{-3}$	17.79	4.15	$1.395 \cdot 10^{-3}$	21.96	4.46
W5	6	24576	346880	2116352	$1.863 \cdot 10^{-5}$	111.44	6.80	$1.507 \cdot 10^{-5}$	68.02	6.09	$1.849 \cdot 10^{-5}$	75.45	6.24
W5	7	98304	1381888	8459776	$3.841 \cdot 10^{-7}$	48.50	5.60	$3.536 \cdot 10^{-7}$	42.62	5.41	$4.213 \cdot 10^{-7}$	43.89	5.46
W6	0	6	152	740	$1.675 \cdot 10^0$	—	—	$3.544 \cdot 10^{-1}$	—	—	$9.928 \cdot 10^{-1}$	—	—
W6	1	24	496	2848	$1.241 \cdot 10^0$	1.35	0.43	$1.385 \cdot 10^{-1}$	2.56	1.36	$4.036 \cdot 10^{-1}$	2.46	1.30
W6	2	96	1760	11168	$7.307 \cdot 10^{-1}$	1.70	0.76	$9.621 \cdot 10^{-2}$	1.44	0.53	$2.399 \cdot 10^{-1}$	1.68	0.75
W6	3	384	6592	44224	$2.359 \cdot 10^{-1}$	3.10	1.63	$1.653 \cdot 10^{-2}$	5.82	2.54	$3.403 \cdot 10^{-2}$	7.05	2.82
W6	4	1536	25472	176000	$1.238 \cdot 10^{-2}$	19.06	4.25	$1.446 \cdot 10^{-3}$	11.43	3.52	$2.321 \cdot 10^{-3}$	14.66	3.87
W6	5	6144	100096	702208	$1.202 \cdot 10^{-4}$	102.97	6.69	$5.599 \cdot 10^{-5}$	25.83	4.69	$7.524 \cdot 10^{-5}$	30.85	4.95
W6	6	24576	396800	2805248	$1.984 \cdot 10^{-6}$	60.58	5.92	$1.729 \cdot 10^{-6}$	32.38	5.02	$2.154 \cdot 10^{-6}$	34.93	5.13
W6	7	98304	1580032	11213824	$8.659 \cdot 10^{-8}$	22.91	4.52	$8.395 \cdot 10^{-8}$	20.60	4.36	$1.097 \cdot 10^{-7}$	19.64	4.30
W7	0	6	166	934	$1.678 \cdot 10^0$	—	—	$4.668 \cdot 10^{-1}$	—	—	$1.319 \cdot 10^0$	—	—
W7	1	24	548	3620	$1.448 \cdot 10^0$	1.16	0.21	$3.087 \cdot 10^{-1}$	1.51	0.60	$8.257 \cdot 10^{-1}$	1.60	0.68
W7	2	96	1960	14248	$5.837 \cdot 10^{-1}$	2.48	1.31	$3.086 \cdot 10^{-2}$	10.00	3.32	$7.041 \cdot 10^{-2}$	11.73	3.55
W7	3	384	7376	56528	$1.398 \cdot 10^{-1}$	4.18	2.06	$6.405 \cdot 10^{-3}$	4.82	2.27	$1.157 \cdot 10^{-2}$	6.08	2.61
W7	4	1536	28576	225184	$3.666 \cdot 10^{-3}$	38.12	5.25	$1.824 \cdot 10^{-4}$	35.11	5.13	$2.803 \cdot 10^{-4}$	41.29	5.37
W7	5	6144	112448	898880	$9.537 \cdot 10^{-6}$	384.40	8.59	$3.83 \cdot 10^{-6}$	47.62	5.57	$4.714 \cdot 10^{-6}$	59.46	5.89
W7	6	24576	446080	3591808	$9.248 \cdot 10^{-8}$	103.12	6.69	$7.113 \cdot 10^{-8}$	53.84	5.75	$1.043 \cdot 10^{-7}$	45.20	5.50
W7	7	98304	1776896	14359808	$1.398 \cdot 10^{-7}$	0.66	-0.60	$1.392 \cdot 10^{-7}$	0.51	-0.97	$2.013 \cdot 10^{-7}$	0.52	-0.95
W8	0	6	190	1162	$2.657 \cdot 10^0$	—	—	$5.951 \cdot 10^{-1}$	—	—	$1.52 \cdot 10^0$	—	—
W8	1	24	620	4508	$1.115 \cdot 10^0$	2.38	1.25	$2.015 \cdot 10^{-1}$	2.95	1.56	$5.849 \cdot 10^{-1}$	2.60	1.38
W8	2	96	2200	17752	$4.67 \cdot 10^{-1}$	2.39	1.26	$4.483 \cdot 10^{-2}$	4.49	2.17	$1.061 \cdot 10^{-1}$	5.51	2.46
W8	3	384	8240	70448	$5.257 \cdot 10^{-2}$	8.88	3.15	$6.587 \cdot 10^{-3}$	6.81	2.77	$1.396 \cdot 10^{-2}$	7.60	2.93
W8	4	1536	31840	280672	$4.056 \cdot 10^{-4}$	129.60	7.02	$3.184 \cdot 10^{-5}$	206.88	7.69	$5.233 \cdot 10^{-5}$	266.77	8.06
W8	5	6144	125120	1120448	$5.953 \cdot 10^{-6}$	68.13	6.09	$2.978 \cdot 10^{-6}$	10.69	3.42	$4.101 \cdot 10^{-6}$	12.76	3.67
W8	6	24576	496000	4477312	$1.216 \cdot 10^{-7}$	48.96	5.61	$1.185 \cdot 10^{-7}$	25.13	4.65	$1.791 \cdot 10^{-7}$	22.90	4.52
W8	7	98304	1975040	17900288	$2.982 \cdot 10^{-7}$	0.41	-1.29	$2.939 \cdot 10^{-7}$	0.40	-1.31	$4.489 \cdot 10^{-7}$	0.40	-1.33

Table 5.11: Convergence results for the WC method according to TABLE 5.5.

5.2.3 A low regularity model problem

The examples considered in the previous sections are smooth up to a certain degree to verify the predicted rates obtained in SECTION 4.7. Here, we consider a model problem that is not even continuous in order to evaluate the performance of the DPG method the weakly conforming Least-Squares method for low-regularity solutions.

Setting $\Omega = (0, 1)$ and $Q = \Omega \times (0, T)$, $T > 0$, we look at the problem

$$\begin{aligned} \partial_t p + \partial_x \mathbf{v} &= 0, & p(\cdot, 0) &= 1, \quad \mathbf{v}(\cdot, 0) = 0 \text{ in } \Omega, \\ \partial_t \mathbf{v} + \partial_x p &= 0, & p(0, \cdot) &= p(1, \cdot) = 0 \text{ in } (0, T), \end{aligned}$$

having a weak solution $(p, \mathbf{v}) \in H(L, Q)$ that can be obtained by extending the arguments in EXAMPLE 3.6, also see [45]. This solution has the following representation

$$p(x, t) = \frac{1}{2}(f_0(x+t) + f_0(x-t)), \quad \mathbf{v}(x, t) = \frac{1}{2}(-f_0(x+t) + f_0(x-t))$$

$$\text{with } f_0(x) = \begin{cases} 1 & x \in (0, 1) + 2\mathbb{Z}, \\ 0 & x \in \mathbb{Z}, \\ -1 & x \in (-1, 0) + 2\mathbb{Z}. \end{cases}$$

A space-time plot of p is given in FIGURE 5.10.

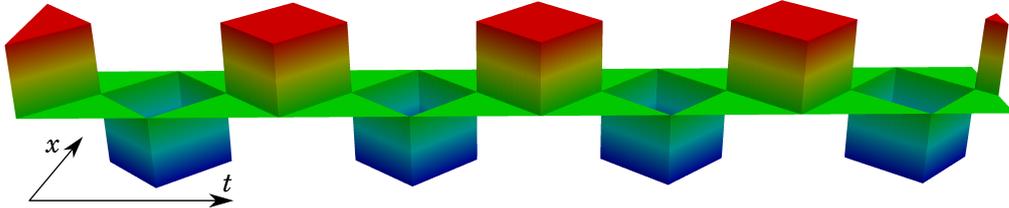


Figure 5.10: Space-time plot of the exact solution's pressure component.

Note that the solution fulfills $(p, \mathbf{v}) \in BV(Q, \mathbb{R}^2)$, see APPENDIX C, and that we have $(p, \mathbf{v}) \notin H^s(Q, \mathbb{R}^2)$, $s \geq \frac{1}{2}$. Due to its low regularity, this model problem is an interesting candidate to compare different discretization schemes. FIGURE 5.11 shows approximations by a finite difference scheme on staggered grids in comparison to a conforming Least-Squares Finite Element discretization as described in SECTION 4.1 and the weakly conforming Least-Squares method introduced in SECTION 4.2. The simulations are performed on the same grid as in the previous sections.

Leap-frog Finite Differences

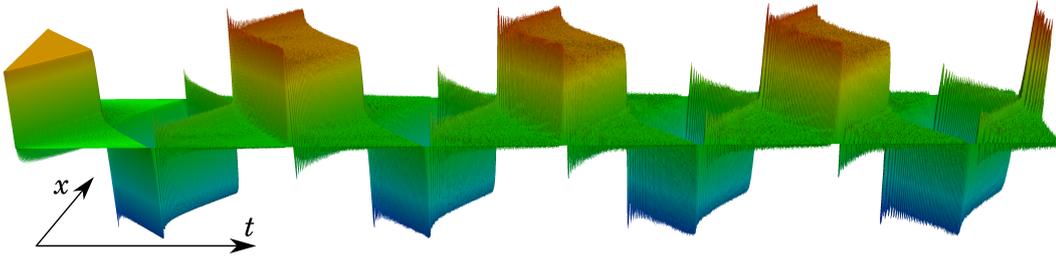
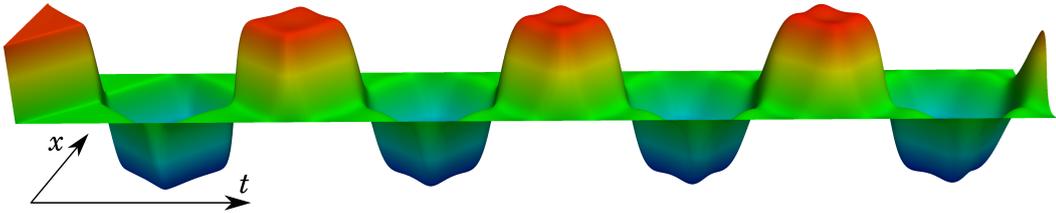
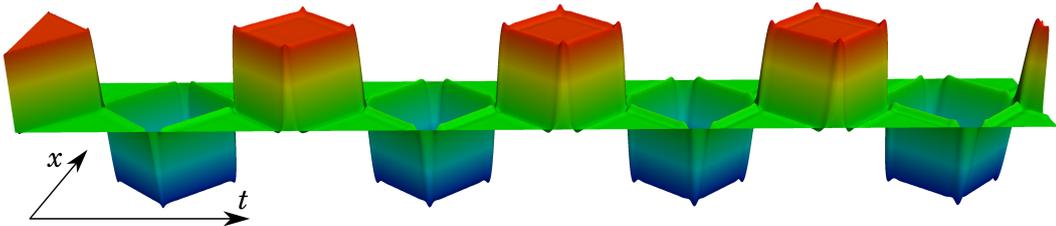
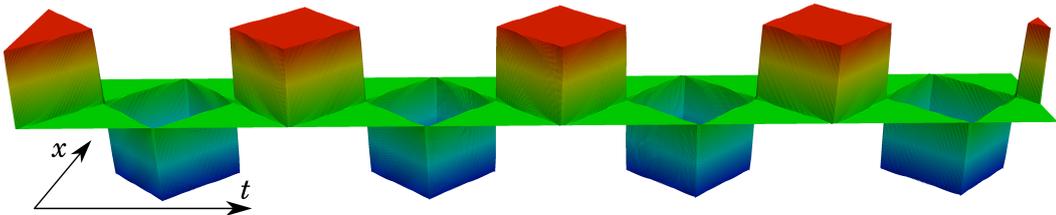
Conforming space-time Least-Squares FEM using cell-wise \mathbb{Q}_1 Conforming space-time Least-Squares FEM using cell-wise \mathbb{Q}_2 Weakly conforming Least-Squares using \mathbb{W}_4 

Figure 5.11: We choose $T \approx 8$ and wave speed $c = 1$ so that the jumps are not aligned with the mesh and plot the pressure component of the numerical solution for different schemes. All discretizations shown in the figures use approximately 150 000 global DoFs. The first picture shows that an explicit leap-frog Finite Difference method generates oscillations. The second and third plot correspond to the conforming Least-Squares FEM with \mathbb{Q}_1 and \mathbb{Q}_2 elements where the first order scheme is highly diffusive and the second order scheme develops over and undershoots at the jumps. The last picture shows the cell-wise mean value of the approximation for the weakly conforming Least-Squares method with \mathbb{Q}_4 polynomials in each cell. Visually, it is hard to distinguish this approximation from the exact solution.

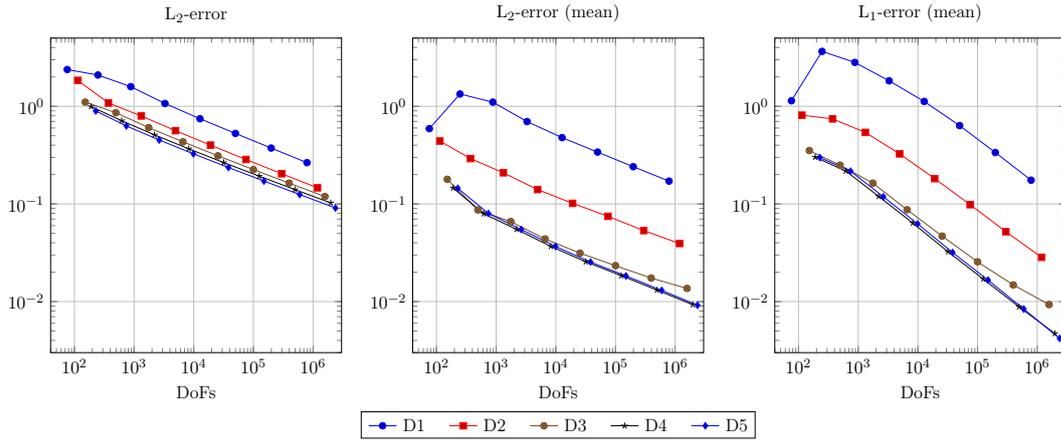


Figure 5.12: Convergence results for DPG with configurations according to TABLE 5.3.

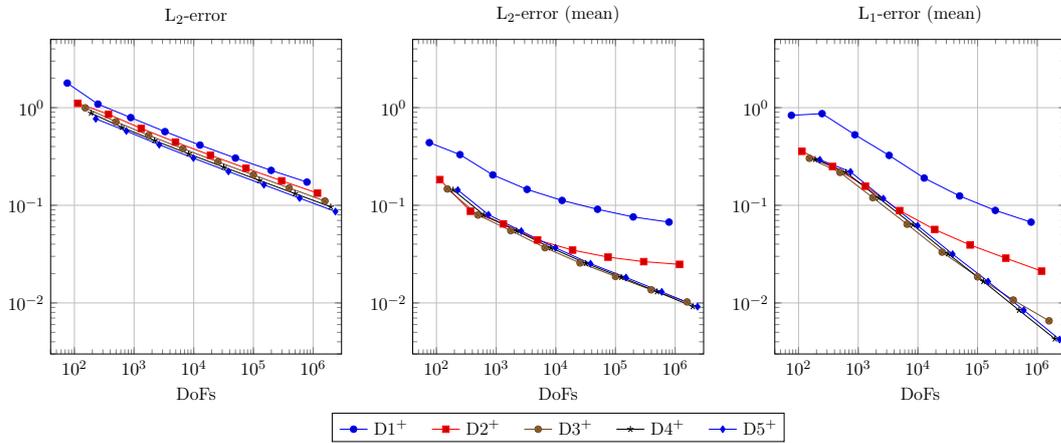


Figure 5.13: Convergence results for DPG with configurations according to TABLE 5.4.

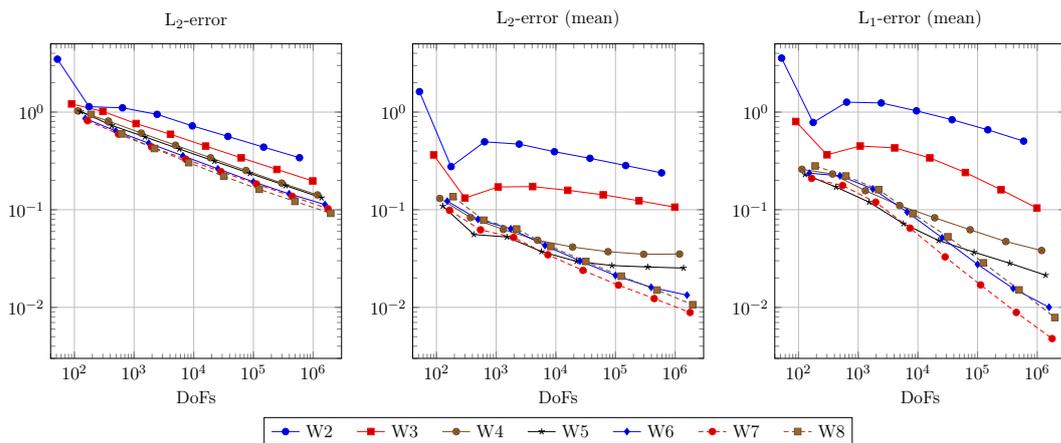


Figure 5.14: Convergence results for WC with configurations according to TABLE 5.5.

Discussion: DPG. The DPG method converges with a rate of about $\sqrt{2}$ with respect to the $L_2(Q)$ norm. When looking at the approximation error in $L_2(Q)$, there is no clear benefit for the high-order variants. Only the lowest order configuration D1 performs significantly worse than the others.

Remarkably, for all methods except D3, we observe a convergence rate of the cell-wise mean value with approximately order 1. Considering the cell-wise mean value, the high-order methods D3 to D5 outperform D1 and D2.

Discussion: DPG – increased cell degree. The results are mostly the same compared to the DPG configurations using polynomials of one degree less inside the cells. For most of the configurations, the errors coincide approximately except for the lowest order method $D1^+$ where a significant improvement can be observed.

However, the configurations $D1^+$, $D2^+$ and $D3^+$ indicate a loss of convergence for the cell-wise means. We did not observe this phenomenon for the DPG configurations without increased polynomial degree in the cells.

Discussion: weakly conforming Least-Squares. In this low-regularity example, the weakly conforming Least-Squares method converges with a rate of about $\frac{4}{3}$ in $L_2(Q)$ norms.

Considering the cell-wise means, we do not observe improved convergence for the low-order schemes in W2 to W5.

For the high-order methods in W6 to W8 however, the order of convergence in $L_1(Q)$ norm of the cell-wise mean appears to tend to 1, similarly as for the DPG method. The configurations shown in W5 and W6 suffer from a loss of convergence for the mean-values.

Comparing the accuracy per degree of freedom to the DPG method, the weakly conforming Least-Squares method delivers a comparable performance in this example.

5.2.3. A low regularity model problem

conf	level	cells	DoFs	all DoFs	L ₂ -error	rate	order	L ₂ -error (mean)	rate	order	L ₁ -error (mean)	rate	order
D1	0	6	76	88	$2.38 \cdot 10^0$	—	—	$5.903 \cdot 10^{-1}$	—	—	$1.14 \cdot 10^0$	—	—
D1	1	24	248	296	$2.091 \cdot 10^0$	1.14	0.19	$1.336 \cdot 10^0$	0.44	-1.18	$3.653 \cdot 10^0$	0.31	-1.68
D1	2	96	880	1072	$1.589 \cdot 10^0$	1.32	0.40	$1.104 \cdot 10^0$	1.21	0.28	$2.814 \cdot 10^0$	1.30	0.38
D1	3	384	3296	4064	$1.069 \cdot 10^0$	1.49	0.57	$6.981 \cdot 10^{-1}$	1.58	0.66	$1.825 \cdot 10^0$	1.54	0.62
D1	4	1536	12736	15808	$7.455 \cdot 10^{-1}$	1.43	0.52	$4.779 \cdot 10^{-1}$	1.46	0.55	$1.123 \cdot 10^0$	1.63	0.70
D1	5	6144	50048	62336	$5.276 \cdot 10^{-1}$	1.41	0.50	$3.397 \cdot 10^{-1}$	1.41	0.49	$6.342 \cdot 10^{-1}$	1.77	0.82
D1	6	24576	198400	247552	$3.737 \cdot 10^{-1}$	1.41	0.50	$2.408 \cdot 10^{-1}$	1.41	0.50	$3.365 \cdot 10^{-1}$	1.88	0.91
D1	7	98304	790016	986624	$2.649 \cdot 10^{-1}$	1.41	0.50	$1.714 \cdot 10^{-1}$	1.41	0.49	$1.752 \cdot 10^{-1}$	1.92	0.94
D2	0	6	114	162	$1.844 \cdot 10^0$	—	—	$4.411 \cdot 10^{-1}$	—	—	$8.113 \cdot 10^{-1}$	—	—
D2	1	24	372	564	$1.081 \cdot 10^0$	1.71	0.77	$2.915 \cdot 10^{-1}$	1.51	0.60	$7.434 \cdot 10^{-1}$	1.09	0.13
D2	2	96	1320	2088	$7.965 \cdot 10^{-1}$	1.36	0.44	$2.085 \cdot 10^{-1}$	1.40	0.48	$5.403 \cdot 10^{-1}$	1.38	0.46
D2	3	384	4944	8016	$5.634 \cdot 10^{-1}$	1.41	0.50	$1.405 \cdot 10^{-1}$	1.48	0.57	$3.259 \cdot 10^{-1}$	1.66	0.73
D2	4	1536	19104	31392	$4.006 \cdot 10^{-1}$	1.41	0.49	$1.015 \cdot 10^{-1}$	1.38	0.47	$1.817 \cdot 10^{-1}$	1.79	0.84
D2	5	6144	75072	124224	$2.855 \cdot 10^{-1}$	1.40	0.49	$7.47 \cdot 10^{-2}$	1.36	0.44	$9.833 \cdot 10^{-2}$	1.85	0.89
D2	6	24576	297600	494208	$2.04 \cdot 10^{-1}$	1.40	0.48	$5.334 \cdot 10^{-2}$	1.40	0.49	$5.197 \cdot 10^{-2}$	1.89	0.92
D2	7	98304	1185024	1971456	$1.463 \cdot 10^{-1}$	1.39	0.48	$3.935 \cdot 10^{-2}$	1.36	0.44	$2.84 \cdot 10^{-2}$	1.83	0.87
D3	0	6	152	260	$1.104 \cdot 10^0$	—	—	$1.787 \cdot 10^{-1}$	—	—	$3.521 \cdot 10^{-1}$	—	—
D3	1	24	496	928	$8.602 \cdot 10^{-1}$	1.28	0.36	$8.673 \cdot 10^{-2}$	2.06	1.04	$2.495 \cdot 10^{-1}$	1.41	0.50
D3	2	96	1760	3488	$6.052 \cdot 10^{-1}$	1.42	0.51	$6.589 \cdot 10^{-2}$	1.32	0.40	$1.633 \cdot 10^{-1}$	1.53	0.61
D3	3	384	6592	13504	$4.329 \cdot 10^{-1}$	1.40	0.48	$4.378 \cdot 10^{-2}$	1.50	0.59	$8.725 \cdot 10^{-2}$	1.87	0.90
D3	4	1536	25472	53120	$3.112 \cdot 10^{-1}$	1.39	0.48	$3.136 \cdot 10^{-2}$	1.40	0.48	$4.685 \cdot 10^{-2}$	1.86	0.90
D3	5	6144	100096	210688	$2.249 \cdot 10^{-1}$	1.38	0.47	$2.333 \cdot 10^{-2}$	1.34	0.43	$2.552 \cdot 10^{-2}$	1.84	0.88
D3	6	24576	396800	839168	$1.634 \cdot 10^{-1}$	1.38	0.46	$1.741 \cdot 10^{-2}$	1.34	0.42	$1.482 \cdot 10^{-2}$	1.72	0.78
D3	7	98304	1580032	3349504	$1.191 \cdot 10^{-1}$	1.37	0.46	$1.365 \cdot 10^{-2}$	1.28	0.35	$9.323 \cdot 10^{-3}$	1.59	0.67
D4	0	6	190	382	$9.94 \cdot 10^{-1}$	—	—	$1.459 \cdot 10^{-1}$	—	—	$3.011 \cdot 10^{-1}$	—	—
D4	1	24	620	1388	$7.12 \cdot 10^{-1}$	1.40	0.48	$7.935 \cdot 10^{-2}$	1.84	0.88	$2.173 \cdot 10^{-1}$	1.39	0.47
D4	2	96	2200	5272	$5.103 \cdot 10^{-1}$	1.40	0.48	$5.495 \cdot 10^{-2}$	1.44	0.53	$1.195 \cdot 10^{-1}$	1.82	0.86
D4	3	384	8240	20528	$3.67 \cdot 10^{-1}$	1.39	0.48	$3.676 \cdot 10^{-2}$	1.49	0.58	$6.362 \cdot 10^{-2}$	1.88	0.91
D4	4	1536	31840	80992	$2.653 \cdot 10^{-1}$	1.38	0.47	$2.556 \cdot 10^{-2}$	1.44	0.52	$3.246 \cdot 10^{-2}$	1.96	0.97
D4	5	6144	125120	321728	$1.927 \cdot 10^{-1}$	1.38	0.46	$1.844 \cdot 10^{-2}$	1.39	0.47	$1.698 \cdot 10^{-2}$	1.91	0.93
D4	6	24576	496000	1282432	$1.404 \cdot 10^{-1}$	1.37	0.46	$1.312 \cdot 10^{-2}$	1.41	0.49	$8.826 \cdot 10^{-3}$	1.92	0.94
D4	7	98304	1975040	5120768	$1.025 \cdot 10^{-1}$	1.37	0.45	$9.298 \cdot 10^{-3}$	1.41	0.50	$4.706 \cdot 10^{-3}$	1.88	0.91
D5	0	6	228	528	$8.961 \cdot 10^{-1}$	—	—	$1.446 \cdot 10^{-1}$	—	—	$2.982 \cdot 10^{-1}$	—	—
D5	1	24	744	1944	$6.296 \cdot 10^{-1}$	1.42	0.51	$7.98 \cdot 10^{-2}$	1.81	0.86	$2.169 \cdot 10^{-1}$	1.37	0.46
D5	2	96	2640	7440	$4.518 \cdot 10^{-1}$	1.39	0.48	$5.463 \cdot 10^{-2}$	1.46	0.55	$1.185 \cdot 10^{-1}$	1.83	0.87
D5	3	384	9888	29088	$3.26 \cdot 10^{-1}$	1.39	0.47	$3.67 \cdot 10^{-2}$	1.49	0.57	$6.253 \cdot 10^{-2}$	1.89	0.92
D5	4	1536	38208	115008	$2.362 \cdot 10^{-1}$	1.38	0.46	$2.539 \cdot 10^{-2}$	1.45	0.53	$3.182 \cdot 10^{-2}$	1.97	0.97
D5	5	6144	150144	457344	$1.717 \cdot 10^{-1}$	1.38	0.46	$1.831 \cdot 10^{-2}$	1.39	0.47	$1.658 \cdot 10^{-2}$	1.92	0.94
D5	6	24576	595200	1824000	$1.249 \cdot 10^{-1}$	1.37	0.46	$1.302 \cdot 10^{-2}$	1.41	0.49	$8.385 \cdot 10^{-3}$	1.98	0.98
D5	7	98304	2370048	7285248	$9.093 \cdot 10^{-2}$	1.37	0.46	$9.174 \cdot 10^{-3}$	1.42	0.51	$4.201 \cdot 10^{-3}$	2.00	1.00

Table 5.12: Convergence results for the DPG method according to TABLE 5.3.

5.2. Numerical examples (1D)

conf	level	cells	DoFs	all DoFs	L_2 -error	rate	order	L_2 -error (mean)	rate	order	L_1 -error (mean)	rate	order
D1 ⁺	0	6	76	124	$1.785 \cdot 10^0$	–	–	$4.385 \cdot 10^{-1}$	–	–	$8.342 \cdot 10^{-1}$	–	–
D1 ⁺	1	24	248	440	$1.088 \cdot 10^0$	1.64	0.71	$3.304 \cdot 10^{-1}$	1.33	0.41	$8.692 \cdot 10^{-1}$	0.96	–0.06
D1 ⁺	2	96	880	1 648	$7.915 \cdot 10^{-1}$	1.37	0.46	$2.048 \cdot 10^{-1}$	1.61	0.69	$5.279 \cdot 10^{-1}$	1.65	0.72
D1 ⁺	3	384	3 296	6 368	$5.684 \cdot 10^{-1}$	1.39	0.48	$1.455 \cdot 10^{-1}$	1.41	0.49	$3.246 \cdot 10^{-1}$	1.63	0.70
D1 ⁺	4	1 536	12 736	25 024	$4.134 \cdot 10^{-1}$	1.38	0.46	$1.122 \cdot 10^{-1}$	1.30	0.37	$1.903 \cdot 10^{-1}$	1.71	0.77
D1 ⁺	5	6 144	50 048	99 200	$3.046 \cdot 10^{-1}$	1.36	0.44	$9.106 \cdot 10^{-2}$	1.23	0.30	$1.247 \cdot 10^{-1}$	1.53	0.61
D1 ⁺	6	24 576	198 400	395 008	$2.278 \cdot 10^{-1}$	1.34	0.42	$7.598 \cdot 10^{-2}$	1.20	0.26	$8.866 \cdot 10^{-2}$	1.41	0.49
D1 ⁺	7	98 304	790 016	1 576 448	$1.729 \cdot 10^{-1}$	1.32	0.40	$6.729 \cdot 10^{-2}$	1.13	0.18	$6.723 \cdot 10^{-2}$	1.32	0.40
D2 ⁺	0	6	114	222	$1.108 \cdot 10^0$	–	–	$1.833 \cdot 10^{-1}$	–	–	$3.574 \cdot 10^{-1}$	–	–
D2 ⁺	1	24	372	804	$8.558 \cdot 10^{-1}$	1.30	0.37	$8.693 \cdot 10^{-2}$	2.11	1.08	$2.497 \cdot 10^{-1}$	1.43	0.52
D2 ⁺	2	96	1 320	3 048	$6.128 \cdot 10^{-1}$	1.40	0.48	$6.452 \cdot 10^{-2}$	1.35	0.43	$1.568 \cdot 10^{-1}$	1.59	0.67
D2 ⁺	3	384	4 944	11 856	$4.445 \cdot 10^{-1}$	1.38	0.46	$4.412 \cdot 10^{-2}$	1.46	0.55	$8.824 \cdot 10^{-2}$	1.78	0.83
D2 ⁺	4	1 536	19 104	46 752	$3.254 \cdot 10^{-1}$	1.37	0.45	$3.476 \cdot 10^{-2}$	1.27	0.34	$5.651 \cdot 10^{-2}$	1.56	0.64
D2 ⁺	5	6 144	75 072	185 664	$2.4 \cdot 10^{-1}$	1.36	0.44	$2.947 \cdot 10^{-2}$	1.18	0.24	$3.924 \cdot 10^{-2}$	1.44	0.53
D2 ⁺	6	24 576	297 600	739 968	$1.782 \cdot 10^{-1}$	1.35	0.43	$2.647 \cdot 10^{-2}$	1.11	0.15	$2.874 \cdot 10^{-2}$	1.37	0.45
D2 ⁺	7	98 304	1 185 024	2 954 496	$1.333 \cdot 10^{-1}$	1.34	0.42	$2.483 \cdot 10^{-2}$	1.07	0.09	$2.117 \cdot 10^{-2}$	1.36	0.44
D3 ⁺	0	6	152	344	$9.958 \cdot 10^{-1}$	–	–	$1.472 \cdot 10^{-1}$	–	–	$3.03 \cdot 10^{-1}$	–	–
D3 ⁺	1	24	496	1 264	$7.177 \cdot 10^{-1}$	1.39	0.47	$7.954 \cdot 10^{-2}$	1.85	0.89	$2.172 \cdot 10^{-1}$	1.40	0.48
D3 ⁺	2	96	1 760	4 832	$5.198 \cdot 10^{-1}$	1.38	0.47	$5.504 \cdot 10^{-2}$	1.45	0.53	$1.198 \cdot 10^{-1}$	1.81	0.86
D3 ⁺	3	384	6 592	18 880	$3.798 \cdot 10^{-1}$	1.37	0.45	$3.681 \cdot 10^{-2}$	1.50	0.58	$6.398 \cdot 10^{-2}$	1.87	0.90
D3 ⁺	4	1 536	25 472	74 624	$2.787 \cdot 10^{-1}$	1.36	0.45	$2.569 \cdot 10^{-2}$	1.43	0.52	$3.314 \cdot 10^{-2}$	1.93	0.95
D3 ⁺	5	6 144	100 096	296 704	$2.047 \cdot 10^{-1}$	1.36	0.44	$1.87 \cdot 10^{-2}$	1.37	0.46	$1.85 \cdot 10^{-2}$	1.79	0.84
D3 ⁺	6	24 576	396 800	1 183 232	$1.505 \cdot 10^{-1}$	1.36	0.44	$1.363 \cdot 10^{-2}$	1.37	0.46	$1.071 \cdot 10^{-2}$	1.73	0.79
D3 ⁺	7	98 304	1 580 032	4 725 760	$1.107 \cdot 10^{-1}$	1.36	0.44	$1.024 \cdot 10^{-2}$	1.33	0.41	$6.554 \cdot 10^{-3}$	1.63	0.71
D4 ⁺	0	6	190	490	$8.804 \cdot 10^{-1}$	–	–	$1.44 \cdot 10^{-1}$	–	–	$2.959 \cdot 10^{-1}$	–	–
D4 ⁺	1	24	620	1 820	$6.299 \cdot 10^{-1}$	1.40	0.48	$7.989 \cdot 10^{-2}$	1.80	0.85	$2.18 \cdot 10^{-1}$	1.36	0.44
D4 ⁺	2	96	2 200	7 000	$4.608 \cdot 10^{-1}$	1.37	0.45	$5.477 \cdot 10^{-2}$	1.46	0.54	$1.191 \cdot 10^{-1}$	1.83	0.87
D4 ⁺	3	384	8 240	27 440	$3.363 \cdot 10^{-1}$	1.37	0.45	$3.678 \cdot 10^{-2}$	1.49	0.57	$6.312 \cdot 10^{-2}$	1.89	0.92
D4 ⁺	4	1 536	31 840	108 640	$2.461 \cdot 10^{-1}$	1.37	0.45	$2.547 \cdot 10^{-2}$	1.44	0.53	$3.199 \cdot 10^{-2}$	1.97	0.98
D4 ⁺	5	6 144	125 120	432 320	$1.801 \cdot 10^{-1}$	1.37	0.45	$1.835 \cdot 10^{-2}$	1.39	0.47	$1.66 \cdot 10^{-2}$	1.93	0.95
D4 ⁺	6	24 576	496 000	1 724 800	$1.318 \cdot 10^{-1}$	1.37	0.45	$1.304 \cdot 10^{-2}$	1.41	0.49	$8.42 \cdot 10^{-3}$	1.97	0.98
D4 ⁺	7	98 304	1 975 040	6 890 240	$9.644 \cdot 10^{-2}$	1.37	0.45	$9.186 \cdot 10^{-3}$	1.42	0.51	$4.288 \cdot 10^{-3}$	1.96	0.97
D5 ⁺	0	6	228	660	$7.686 \cdot 10^{-1}$	–	–	$1.434 \cdot 10^{-1}$	–	–	$2.926 \cdot 10^{-1}$	–	–
D5 ⁺	1	24	744	2 472	$5.761 \cdot 10^{-1}$	1.33	0.42	$8.02 \cdot 10^{-2}$	1.79	0.84	$2.205 \cdot 10^{-1}$	1.33	0.41
D5 ⁺	2	96	2 640	9 552	$4.174 \cdot 10^{-1}$	1.38	0.46	$5.443 \cdot 10^{-2}$	1.47	0.56	$1.174 \cdot 10^{-1}$	1.88	0.91
D5 ⁺	3	384	9 888	37 536	$3.043 \cdot 10^{-1}$	1.37	0.46	$3.667 \cdot 10^{-2}$	1.48	0.57	$6.203 \cdot 10^{-2}$	1.89	0.92
D5 ⁺	4	1 536	38 208	148 800	$2.222 \cdot 10^{-1}$	1.37	0.45	$2.525 \cdot 10^{-2}$	1.45	0.54	$3.162 \cdot 10^{-2}$	1.96	0.97
D5 ⁺	5	6 144	150 144	592 512	$1.622 \cdot 10^{-1}$	1.37	0.45	$1.823 \cdot 10^{-2}$	1.39	0.47	$1.658 \cdot 10^{-2}$	1.91	0.93
D5 ⁺	6	24 576	595 200	2 364 672	$1.184 \cdot 10^{-1}$	1.37	0.45	$1.299 \cdot 10^{-2}$	1.40	0.49	$8.427 \cdot 10^{-3}$	1.97	0.98
D5 ⁺	7	98 304	2 370 048	9 447 936	$8.637 \cdot 10^{-2}$	1.37	0.46	$9.161 \cdot 10^{-3}$	1.42	0.50	$4.219 \cdot 10^{-3}$	2.00	1.00

Table 5.13: Convergence results for the DPG method according to TABLE 5.4.

5.2.3. A low regularity model problem

conf	level	cells	DofFs	all DofFs	L ₂ -error	rate	order	L ₂ -error (mean)	rate	order	L ₁ -error (mean)	rate	order
W2	0	6	52	160	$3.484 \cdot 10^0$	—	—	$1.619 \cdot 10^0$	—	—	$3.575 \cdot 10^0$	—	—
W2	1	24	176	608	$1.137 \cdot 10^0$	3.06	1.62	$2.764 \cdot 10^{-1}$	5.86	2.55	$7.831 \cdot 10^{-1}$	4.57	2.19
W2	2	96	640	2368	$1.106 \cdot 10^0$	1.03	0.04	$4.96 \cdot 10^{-1}$	0.56	-0.84	$1.264 \cdot 10^0$	0.62	-0.69
W2	3	384	2432	9344	$9.478 \cdot 10^{-1}$	1.17	0.22	$4.699 \cdot 10^{-1}$	1.06	0.08	$1.24 \cdot 10^0$	1.02	0.03
W2	4	1536	9472	37120	$7.224 \cdot 10^{-1}$	1.31	0.39	$3.93 \cdot 10^{-1}$	1.20	0.26	$1.03 \cdot 10^0$	1.20	0.27
W2	5	6144	37376	147968	$5.635 \cdot 10^{-1}$	1.28	0.36	$3.362 \cdot 10^{-1}$	1.17	0.23	$8.352 \cdot 10^{-1}$	1.23	0.30
W2	6	24576	148480	590848	$4.375 \cdot 10^{-1}$	1.29	0.37	$2.836 \cdot 10^{-1}$	1.19	0.25	$6.582 \cdot 10^{-1}$	1.27	0.34
W2	7	98304	591872	2361344	$3.41 \cdot 10^{-1}$	1.28	0.36	$2.385 \cdot 10^{-1}$	1.19	0.25	$5.041 \cdot 10^{-1}$	1.31	0.38
W3	0	6	90	282	$1.212 \cdot 10^0$	—	—	$3.642 \cdot 10^{-1}$	—	—	$8.008 \cdot 10^{-1}$	—	—
W3	1	24	300	1068	$1.017 \cdot 10^0$	1.19	0.25	$1.317 \cdot 10^{-1}$	2.77	1.47	$3.653 \cdot 10^{-1}$	2.19	1.13
W3	2	96	1080	4152	$7.629 \cdot 10^{-1}$	1.33	0.42	$1.71 \cdot 10^{-1}$	0.77	-0.38	$4.493 \cdot 10^{-1}$	0.81	-0.30
W3	3	384	4080	16368	$5.928 \cdot 10^{-1}$	1.29	0.36	$1.726 \cdot 10^{-1}$	0.99	-0.01	$4.297 \cdot 10^{-1}$	1.05	0.06
W3	4	1536	15840	64992	$4.486 \cdot 10^{-1}$	1.32	0.40	$1.581 \cdot 10^{-1}$	1.09	0.13	$3.396 \cdot 10^{-1}$	1.27	0.34
W3	5	6144	62400	259008	$3.401 \cdot 10^{-1}$	1.32	0.40	$1.421 \cdot 10^{-1}$	1.11	0.15	$2.406 \cdot 10^{-1}$	1.41	0.50
W3	6	24576	247680	1034112	$2.586 \cdot 10^{-1}$	1.32	0.40	$1.235 \cdot 10^{-1}$	1.15	0.20	$1.602 \cdot 10^{-1}$	1.50	0.59
W3	7	98304	986880	4132608	$1.971 \cdot 10^{-1}$	1.31	0.39	$1.062 \cdot 10^{-1}$	1.16	0.22	$1.04 \cdot 10^{-1}$	1.54	0.62
W4	0	6	114	414	$1.028 \cdot 10^0$	—	—	$1.305 \cdot 10^{-1}$	—	—	$2.594 \cdot 10^{-1}$	—	—
W4	1	24	372	1572	$8.099 \cdot 10^{-1}$	1.27	0.34	$8.306 \cdot 10^{-2}$	1.57	0.65	$2.32 \cdot 10^{-1}$	1.12	0.16
W4	2	96	1320	6120	$6.093 \cdot 10^{-1}$	1.33	0.41	$6.276 \cdot 10^{-2}$	1.32	0.40	$1.567 \cdot 10^{-1}$	1.48	0.57
W4	3	384	4944	24144	$4.579 \cdot 10^{-1}$	1.33	0.41	$4.871 \cdot 10^{-2}$	1.29	0.37	$1.103 \cdot 10^{-1}$	1.42	0.51
W4	4	1536	19104	95904	$3.405 \cdot 10^{-1}$	1.34	0.43	$4.145 \cdot 10^{-2}$	1.17	0.23	$8.265 \cdot 10^{-2}$	1.33	0.42
W4	5	6144	75072	382272	$2.523 \cdot 10^{-1}$	1.35	0.43	$3.717 \cdot 10^{-2}$	1.12	0.16	$6.234 \cdot 10^{-2}$	1.33	0.41
W4	6	24576	297600	1526400	$1.875 \cdot 10^{-1}$	1.35	0.43	$3.498 \cdot 10^{-2}$	1.06	0.09	$4.723 \cdot 10^{-2}$	1.32	0.40
W4	7	98304	1185024	6100224	$1.418 \cdot 10^{-1}$	1.32	0.40	$3.516 \cdot 10^{-2}$	0.99	-0.01	$3.823 \cdot 10^{-2}$	1.24	0.30
W5	0	6	128	560	$1.012 \cdot 10^0$	—	—	$1.088 \cdot 10^{-1}$	—	—	$2.297 \cdot 10^{-1}$	—	—
W5	1	24	424	2152	$7.296 \cdot 10^{-1}$	1.39	0.47	$5.572 \cdot 10^{-2}$	1.95	0.97	$1.708 \cdot 10^{-1}$	1.34	0.43
W5	2	96	1520	8432	$5.576 \cdot 10^{-1}$	1.31	0.39	$5.257 \cdot 10^{-2}$	1.06	0.08	$1.197 \cdot 10^{-1}$	1.43	0.51
W5	3	384	5728	33376	$4.188 \cdot 10^{-1}$	1.33	0.41	$3.727 \cdot 10^{-2}$	1.41	0.50	$7.197 \cdot 10^{-2}$	1.66	0.73
W5	4	1536	22208	132800	$3.141 \cdot 10^{-1}$	1.33	0.41	$2.958 \cdot 10^{-2}$	1.26	0.33	$4.841 \cdot 10^{-2}$	1.49	0.57
W5	5	6144	87424	529792	$2.356 \cdot 10^{-1}$	1.33	0.42	$2.681 \cdot 10^{-2}$	1.10	0.14	$3.675 \cdot 10^{-2}$	1.32	0.40
W5	6	24576	346880	2116352	$1.766 \cdot 10^{-1}$	1.33	0.42	$2.591 \cdot 10^{-2}$	1.03	0.05	$2.83 \cdot 10^{-2}$	1.30	0.38
W5	7	98304	1381888	8459776	$1.326 \cdot 10^{-1}$	1.33	0.41	$2.526 \cdot 10^{-2}$	1.03	0.04	$2.136 \cdot 10^{-2}$	1.32	0.41
W6	0	6	152	740	$8.597 \cdot 10^{-1}$	—	—	$1.226 \cdot 10^{-1}$	—	—	$2.369 \cdot 10^{-1}$	—	—
W6	1	24	496	2848	$6.405 \cdot 10^{-1}$	1.34	0.42	$7.967 \cdot 10^{-2}$	1.54	0.62	$2.226 \cdot 10^{-1}$	1.06	0.09
W6	2	96	1760	11168	$4.805 \cdot 10^{-1}$	1.33	0.41	$6.382 \cdot 10^{-2}$	1.25	0.32	$1.632 \cdot 10^{-1}$	1.36	0.45
W6	3	384	6592	44224	$3.566 \cdot 10^{-1}$	1.35	0.43	$4.305 \cdot 10^{-2}$	1.48	0.57	$9.475 \cdot 10^{-2}$	1.72	0.78
W6	4	1536	25472	176000	$2.623 \cdot 10^{-1}$	1.36	0.44	$2.988 \cdot 10^{-2}$	1.44	0.53	$5.154 \cdot 10^{-2}$	1.84	0.88
W6	5	6144	100096	702208	$1.929 \cdot 10^{-1}$	1.36	0.44	$2.12 \cdot 10^{-2}$	1.41	0.49	$2.753 \cdot 10^{-2}$	1.87	0.90
W6	6	24576	396800	2805248	$1.455 \cdot 10^{-1}$	1.33	0.41	$1.602 \cdot 10^{-2}$	1.32	0.40	$1.561 \cdot 10^{-2}$	1.76	0.82
W6	7	98304	1580032	11213824	$1.129 \cdot 10^{-1}$	1.29	0.37	$1.335 \cdot 10^{-2}$	1.20	0.26	$1.002 \cdot 10^{-2}$	1.56	0.64
W7	0	6	166	934	$8.167 \cdot 10^{-1}$	—	—	$9.858 \cdot 10^{-2}$	—	—	$2.094 \cdot 10^{-1}$	—	—
W7	1	24	548	3620	$5.957 \cdot 10^{-1}$	1.37	0.46	$6.217 \cdot 10^{-2}$	1.59	0.67	$1.771 \cdot 10^{-1}$	1.18	0.24
W7	2	96	1960	14248	$4.401 \cdot 10^{-1}$	1.35	0.44	$5.201 \cdot 10^{-2}$	1.20	0.26	$1.193 \cdot 10^{-1}$	1.48	0.57
W7	3	384	7376	56528	$3.287 \cdot 10^{-1}$	1.34	0.42	$3.457 \cdot 10^{-2}$	1.50	0.59	$6.481 \cdot 10^{-2}$	1.84	0.88
W7	4	1536	28576	225184	$2.457 \cdot 10^{-1}$	1.34	0.42	$2.39 \cdot 10^{-2}$	1.45	0.53	$3.287 \cdot 10^{-2}$	1.97	0.98
W7	5	6144	112448	898880	$1.835 \cdot 10^{-1}$	1.34	0.42	$1.69 \cdot 10^{-2}$	1.41	0.50	$1.697 \cdot 10^{-2}$	1.94	0.95
W7	6	24576	446080	3591808	$1.371 \cdot 10^{-1}$	1.34	0.42	$1.23 \cdot 10^{-2}$	1.37	0.46	$8.866 \cdot 10^{-3}$	1.91	0.94
W7	7	98304	1776896	14359808	$1.023 \cdot 10^{-1}$	1.34	0.42	$8.862 \cdot 10^{-3}$	1.39	0.47	$4.787 \cdot 10^{-3}$	1.85	0.89
W8	0	6	190	1162	$9.472 \cdot 10^{-1}$	—	—	$1.361 \cdot 10^{-1}$	—	—	$2.805 \cdot 10^{-1}$	—	—
W8	1	24	620	4508	$5.982 \cdot 10^{-1}$	1.58	0.66	$7.811 \cdot 10^{-2}$	1.74	0.80	$2.215 \cdot 10^{-1}$	1.27	0.34
W8	2	96	2200	17752	$4.237 \cdot 10^{-1}$	1.41	0.50	$6.348 \cdot 10^{-2}$	1.23	0.30	$1.606 \cdot 10^{-1}$	1.38	0.46
W8	3	384	8240	70448	$3.036 \cdot 10^{-1}$	1.40	0.48	$4.195 \cdot 10^{-2}$	1.51	0.60	$9.117 \cdot 10^{-2}$	1.76	0.82
W8	4	1536	31840	280672	$2.202 \cdot 10^{-1}$	1.38	0.46	$2.952 \cdot 10^{-2}$	1.42	0.51	$5.28 \cdot 10^{-2}$	1.73	0.79
W8	5	6144	125120	1120448	$1.619 \cdot 10^{-1}$	1.36	0.44	$2.088 \cdot 10^{-2}$	1.41	0.50	$2.859 \cdot 10^{-2}$	1.85	0.88
W8	6	24576	496000	4477312	$1.217 \cdot 10^{-1}$	1.33	0.41	$1.504 \cdot 10^{-2}$	1.39	0.47	$1.51 \cdot 10^{-2}$	1.89	0.92
W8	7	98304	1975040	17900288	$9.192 \cdot 10^{-2}$	1.32	0.41	$1.064 \cdot 10^{-2}$	1.41	0.50	$7.861 \cdot 10^{-3}$	1.92	0.94

Table 5.14: Convergence results for the WC method according to TABLE 5.5.

5.3 Numerical examples (2D)

In this section, we present different numerical examples in two spacial dimensions. Again, we start with an example that is designed to verify the convergence rates that are predicted by the theory. More realistic examples demonstrate the flexibility of the methods.

Remark 5.4. *Up until now, it is an open question whether stable pairing of test and ansatz spaces for the weakly conforming Least-Squares method exist in two spatial dimensions. Thus, no results for this method are presented here.*

5.3.1 A smooth example

We consider a smooth solution $\mathbf{y} = (p, \mathbf{v}) \in C^\infty(\mathbb{R}^{2+1}, \mathbb{R}^{1+2})$ of $L\mathbf{y} = \mathbf{b}$ with DIRICHLET boundary conditions in the pressure component. The solution and the right-hand side $\mathbf{b} = (b_p, \mathbf{b}_v)$ are given by

$$p(x, t) = \sin(\omega x) \sin(\omega y), \quad \mathbf{v}(x, t) = \begin{pmatrix} -(t-1)\omega \sin(\omega y) \cos(\omega x) \\ -(t-1)\omega \sin(\omega x) \cos(\omega y) \end{pmatrix},$$

$$b_p(x, t) = 2(t-1)\omega^2 \sin(\omega x) \sin(\omega y), \quad \mathbf{b}_v(x, t) = \mathbf{0},$$

see FIGURE 5.15. We choose $\omega = 0.6$ and $Q = (0, 3/\pi) \times (0, e/3) \times (0, 1)$ to prevent alignment of the characteristics with the mesh.

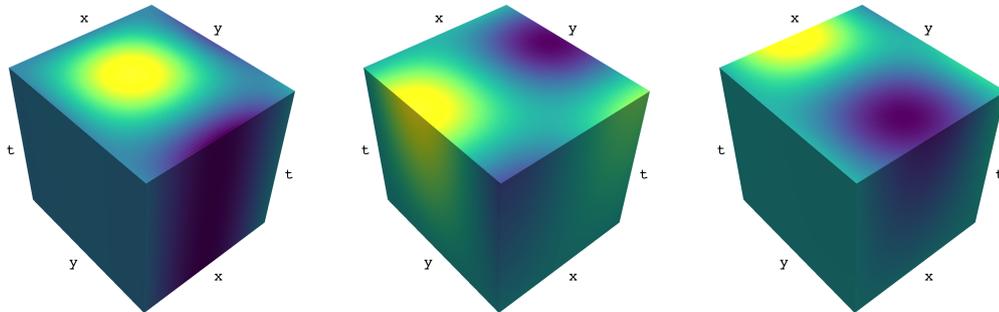


Figure 5.15: The pressure and velocity components of the analytic solution. Note that the color map of the pressure is scaled differently from the color map of two velocity components to improve the contrast of the plots.

See FIGURE 5.16, FIGURE 5.17 for a comparison of different DPG configurations. The detailed results can be found in TABLE 5.15 and TABLE 5.16.

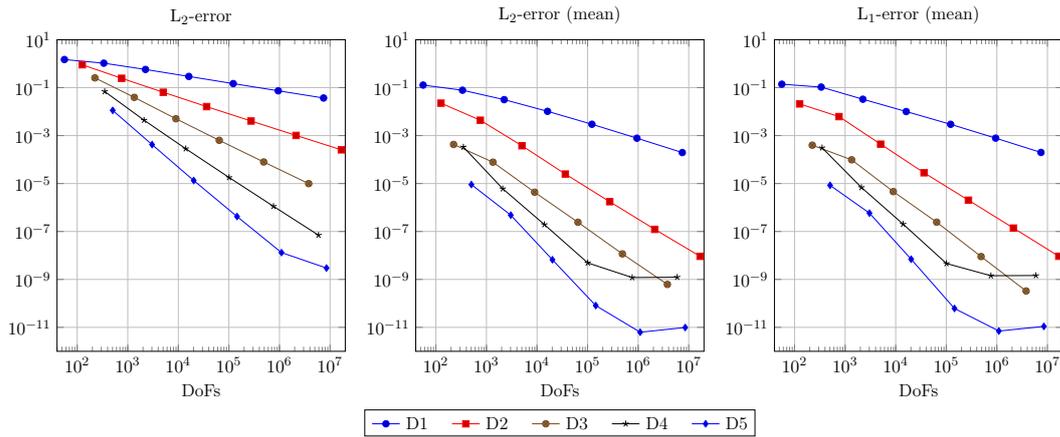


Figure 5.16: Convergence results for DPG with configurations according to TABLE 5.3.

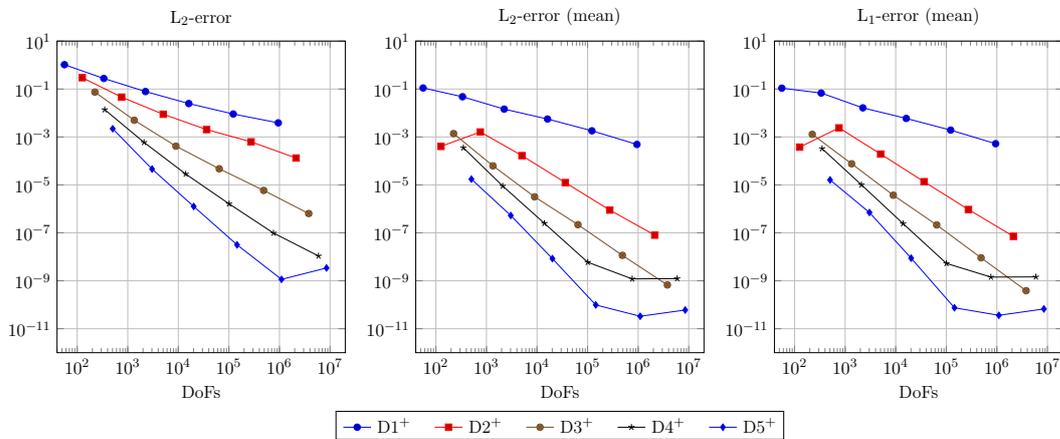


Figure 5.17: Convergence results for DPG with configurations according to TABLE 5.4.

Discussion: DPG. The numerical experiments demonstrate that also in two spatial dimensions, the space-time DPG method complies with the predictions made by SECTION 4.7, since the predicted rates are attained with high accuracy. For the cell-wise mean values, we observe an increased convergence rate where the rates are approximately the same as for the smooth example in one spatial dimension. For high levels, we observe round-off errors in the high-order configurations D4 and D5.

Discussion: DPG – increased cell degree Differently than from the one-dimensional case, increasing the polynomial degree inside the cells does not lead to a significant improvement of the convergence rate.

However, the approximation error in $L_2(Q)$ norms drops by about an order of magnitude compared to the configurations with lower polynomial degree. In contrast to the schemes D1-D5, for D1⁺-D5⁺ the convergence rates on $L_2(Q)$ oscillate. Interestingly, when considering the convergence of the cell-wise mean value, we ob-

5.3. Numerical examples (2D)

serve a loss in the rate as well as in the error itself.

conf	level	cells	DoFs	all DoFs	L_2 -error	rate	order	L_2 -error (mean)	rate	order	L_1 -error (mean)	rate	order
D1	0	1	56	59	$1.49 \cdot 10^0$	—	—	$1.278 \cdot 10^{-1}$	—	—	$1.392 \cdot 10^{-1}$	—	—
D1	1	8	336	360	$1.048 \cdot 10^0$	1.42	0.51	$7.806 \cdot 10^{-2}$	1.64	0.71	$1.056 \cdot 10^{-1}$	1.32	0.40
D1	2	64	2240	2432	$5.737 \cdot 10^{-1}$	1.83	0.87	$3.159 \cdot 10^{-2}$	2.47	1.31	$3.286 \cdot 10^{-2}$	3.21	1.68
D1	3	512	16128	17664	$2.936 \cdot 10^{-1}$	1.95	0.97	$1.029 \cdot 10^{-2}$	3.07	1.62	$1.007 \cdot 10^{-2}$	3.26	1.71
D1	4	4096	121856	134144	$1.476 \cdot 10^{-1}$	1.99	0.99	$2.973 \cdot 10^{-3}$	3.46	1.79	$2.952 \cdot 10^{-3}$	3.41	1.77
D1	5	32768	946176	1044480	$7.392 \cdot 10^{-2}$	2.00	1.00	$7.781 \cdot 10^{-4}$	3.82	1.93	$7.819 \cdot 10^{-4}$	3.78	1.92
D1	6	262144	7454720	8241152	$3.697 \cdot 10^{-2}$	2.00	1.00	$1.958 \cdot 10^{-4}$	3.97	1.99	$1.98 \cdot 10^{-4}$	3.95	1.98
D2	0	1	126	150	$9.006 \cdot 10^{-1}$	—	—	$2.252 \cdot 10^{-2}$	—	—	$2.095 \cdot 10^{-2}$	—	—
D2	1	8	756	948	$2.456 \cdot 10^{-1}$	3.67	1.87	$4.376 \cdot 10^{-3}$	5.15	2.36	$6.199 \cdot 10^{-3}$	3.38	1.76
D2	2	64	5040	6576	$6.417 \cdot 10^{-2}$	3.83	1.94	$3.763 \cdot 10^{-4}$	11.63	3.54	$4.382 \cdot 10^{-4}$	14.15	3.82
D2	3	512	36288	48576	$1.62 \cdot 10^{-2}$	3.96	1.99	$2.483 \cdot 10^{-5}$	15.15	3.92	$2.81 \cdot 10^{-5}$	15.59	3.96
D2	4	4096	274176	372480	$4.059 \cdot 10^{-3}$	3.99	2.00	$1.746 \cdot 10^{-6}$	14.22	3.83	$2.012 \cdot 10^{-6}$	13.97	3.80
D2	5	32768	2128896	2915328	$1.016 \cdot 10^{-3}$	4.00	2.00	$1.224 \cdot 10^{-7}$	14.26	3.83	$1.384 \cdot 10^{-7}$	14.54	3.86
D2	6	262144	16773120	23064576	$2.541 \cdot 10^{-4}$	4.00	2.00	$9.18 \cdot 10^{-9}$	13.33	3.74	$9.249 \cdot 10^{-9}$	14.96	3.90
D3	0	1	224	305	$2.557 \cdot 10^{-1}$	—	—	$4.272 \cdot 10^{-4}$	—	—	$3.974 \cdot 10^{-4}$	—	—
D3	1	8	1344	1992	$3.933 \cdot 10^{-2}$	6.50	2.70	$7.772 \cdot 10^{-5}$	5.50	2.46	$9.742 \cdot 10^{-5}$	4.08	2.03
D3	2	64	8960	14144	$5.035 \cdot 10^{-3}$	7.81	2.97	$4.343 \cdot 10^{-6}$	17.90	4.16	$4.603 \cdot 10^{-6}$	21.16	4.40
D3	3	512	64512	105984	$6.337 \cdot 10^{-4}$	7.95	2.99	$2.421 \cdot 10^{-7}$	17.94	4.17	$2.438 \cdot 10^{-7}$	18.88	4.24
D3	4	4096	487424	819200	$7.932 \cdot 10^{-5}$	7.99	3.00	$1.161 \cdot 10^{-8}$	20.85	4.38	$8.98 \cdot 10^{-9}$	27.15	4.76
D3	5	32768	3784704	6438912	$9.915 \cdot 10^{-6}$	8.00	3.00	$6.175 \cdot 10^{-10}$	18.80	4.23	$3.287 \cdot 10^{-10}$	27.32	4.77
D4	0	1	350	542	$6.898 \cdot 10^{-2}$	—	—	$3.329 \cdot 10^{-4}$	—	—	$3.097 \cdot 10^{-4}$	—	—
D4	1	8	2100	3636	$4.396 \cdot 10^{-3}$	15.69	3.97	$6.05 \cdot 10^{-6}$	55.02	5.78	$6.837 \cdot 10^{-6}$	45.30	5.50
D4	2	64	14000	26288	$2.837 \cdot 10^{-4}$	15.50	3.95	$1.935 \cdot 10^{-7}$	31.27	4.97	$2.019 \cdot 10^{-7}$	33.86	5.08
D4	3	512	100800	199104	$1.787 \cdot 10^{-5}$	15.88	3.99	$4.888 \cdot 10^{-9}$	39.59	5.31	$4.608 \cdot 10^{-9}$	43.81	5.45
D4	4	4096	761600	1548032	$1.119 \cdot 10^{-6}$	15.97	4.00	$1.197 \cdot 10^{-9}$	4.08	2.03	$1.426 \cdot 10^{-9}$	3.23	1.69
D4	5	32768	5913600	12205056	$7.004 \cdot 10^{-8}$	15.98	4.00	$1.235 \cdot 10^{-9}$	0.97	-0.05	$1.451 \cdot 10^{-9}$	0.98	-0.03
D5	0	1	504	879	$1.125 \cdot 10^{-2}$	—	—	$9.046 \cdot 10^{-6}$	—	—	$8.415 \cdot 10^{-6}$	—	—
D5	1	8	3024	6024	$4.208 \cdot 10^{-4}$	26.74	4.74	$4.793 \cdot 10^{-7}$	18.87	4.24	$5.822 \cdot 10^{-7}$	14.45	3.85
D5	2	64	20160	44160	$1.335 \cdot 10^{-5}$	31.52	4.98	$6.613 \cdot 10^{-9}$	72.48	6.18	$6.959 \cdot 10^{-9}$	83.66	6.39
D5	3	512	145152	337152	$4.19 \cdot 10^{-7}$	31.86	4.99	$8.091 \cdot 10^{-11}$	81.73	6.35	$6.131 \cdot 10^{-11}$	113.51	6.83
D5	4	4096	1096704	2632704	$1.312 \cdot 10^{-8}$	31.94	5.00	$6.261 \cdot 10^{-12}$	12.92	3.69	$7.017 \cdot 10^{-12}$	8.74	3.13
D5	5	32768	8515584	20803584	$2.982 \cdot 10^{-9}$	4.40	2.14	$9.847 \cdot 10^{-12}$	0.64	-0.65	$1.093 \cdot 10^{-11}$	0.64	-0.64

Table 5.15: Convergence results for the DPG method according to TABLE 5.3.

5.3.1. A smooth example

conf	level	cells	DoFs	all DoFs	L ₂ -error	rate	order	L ₂ -error (mean)	rate	order	L ₁ -error (mean)	rate	order
D1	0	1	56	80	$1.036 \cdot 10^0$	—	—	$1.106 \cdot 10^{-1}$	—	—	$1.087 \cdot 10^{-1}$	—	—
D1	1	8	336	528	$2.762 \cdot 10^{-1}$	3.75	1.91	$4.766 \cdot 10^{-2}$	2.32	1.22	$6.8 \cdot 10^{-2}$	1.60	0.68
D1	2	64	2240	3776	$7.861 \cdot 10^{-2}$	3.51	1.81	$1.453 \cdot 10^{-2}$	3.28	1.71	$1.639 \cdot 10^{-2}$	4.15	2.05
D1	3	512	16128	28416	$2.476 \cdot 10^{-2}$	3.17	1.67	$5.601 \cdot 10^{-3}$	2.59	1.38	$5.932 \cdot 10^{-3}$	2.76	1.47
D1	4	4096	121856	220160	$9.074 \cdot 10^{-3}$	2.73	1.45	$1.811 \cdot 10^{-3}$	3.09	1.63	$1.931 \cdot 10^{-3}$	3.07	1.62
D1	5	32768	946176	1732608	$3.886 \cdot 10^{-3}$	2.34	1.22	$4.877 \cdot 10^{-4}$	3.71	1.89	$5.248 \cdot 10^{-4}$	3.68	1.88
D2	0	1	126	207	$2.973 \cdot 10^{-1}$	—	—	$4.05 \cdot 10^{-4}$	—	—	$3.767 \cdot 10^{-4}$	—	—
D2	1	8	756	1404	$4.587 \cdot 10^{-2}$	6.48	2.70	$1.615 \cdot 10^{-3}$	0.25	-2.00	$2.395 \cdot 10^{-3}$	0.16	-2.67
D2	2	64	5040	10224	$8.878 \cdot 10^{-3}$	5.17	2.37	$1.666 \cdot 10^{-4}$	9.69	3.28	$1.934 \cdot 10^{-4}$	12.38	3.63
D2	3	512	36288	77760	$2.036 \cdot 10^{-3}$	4.36	2.12	$1.246 \cdot 10^{-5}$	13.37	3.74	$1.36 \cdot 10^{-5}$	14.22	3.83
D2	4	4096	274176	605952	$6.268 \cdot 10^{-4}$	3.25	1.70	$8.922 \cdot 10^{-7}$	13.97	3.80	$9.389 \cdot 10^{-7}$	14.48	3.86
D2	5	32768	2128896	4783104	$1.321 \cdot 10^{-4}$	4.74	2.25	$8.027 \cdot 10^{-8}$	11.12	3.47	$7.132 \cdot 10^{-8}$	13.16	3.72
D3	0	1	224	416	$7.438 \cdot 10^{-2}$	—	—	$1.382 \cdot 10^{-3}$	—	—	$1.286 \cdot 10^{-3}$	—	—
D3	1	8	1344	2880	$4.996 \cdot 10^{-3}$	14.89	3.90	$6.236 \cdot 10^{-5}$	22.16	4.47	$7.609 \cdot 10^{-5}$	16.90	4.08
D3	2	64	8960	21248	$4.131 \cdot 10^{-4}$	12.09	3.60	$3.178 \cdot 10^{-6}$	19.62	4.29	$3.722 \cdot 10^{-6}$	20.44	4.35
D3	3	512	64512	162816	$4.716 \cdot 10^{-5}$	8.76	3.13	$2.193 \cdot 10^{-7}$	14.49	3.86	$2.137 \cdot 10^{-7}$	17.42	4.12
D3	4	4096	487424	1273856	$5.875 \cdot 10^{-6}$	8.03	3.01	$1.15 \cdot 10^{-8}$	19.07	4.25	$9.089 \cdot 10^{-9}$	23.51	4.56
D3	5	32768	3784704	10076160	$6.397 \cdot 10^{-7}$	9.18	3.20	$6.691 \cdot 10^{-10}$	17.19	4.10	$3.865 \cdot 10^{-10}$	23.52	4.56
D4	0	1	350	725	$1.38 \cdot 10^{-2}$	—	—	$3.515 \cdot 10^{-4}$	—	—	$3.27 \cdot 10^{-4}$	—	—
D4	1	8	2100	5100	$5.952 \cdot 10^{-4}$	23.19	4.54	$8.951 \cdot 10^{-6}$	39.27	5.30	$1.008 \cdot 10^{-5}$	32.44	5.02
D4	2	64	14000	38000	$2.844 \cdot 10^{-5}$	20.93	4.39	$2.496 \cdot 10^{-7}$	35.86	5.16	$2.417 \cdot 10^{-7}$	41.70	5.38
D4	3	512	100800	292800	$1.622 \cdot 10^{-6}$	17.53	4.13	$5.929 \cdot 10^{-9}$	42.10	5.40	$5.302 \cdot 10^{-9}$	45.59	5.51
D4	4	4096	761600	2297600	$9.807 \cdot 10^{-8}$	16.54	4.05	$1.199 \cdot 10^{-9}$	4.94	2.31	$1.428 \cdot 10^{-9}$	3.71	1.89
D4	5	32768	5913600	18201600	$1.084 \cdot 10^{-8}$	9.05	3.18	$1.235 \cdot 10^{-9}$	0.97	-0.04	$1.451 \cdot 10^{-9}$	0.98	-0.02
D5	0	1	504	1152	$2.203 \cdot 10^{-3}$	—	—	$1.735 \cdot 10^{-5}$	—	—	$1.614 \cdot 10^{-5}$	—	—
D5	1	8	3024	8208	$4.582 \cdot 10^{-5}$	48.08	5.59	$5.291 \cdot 10^{-7}$	32.79	5.04	$7.044 \cdot 10^{-7}$	22.91	4.52
D5	2	64	20160	61632	$1.262 \cdot 10^{-6}$	36.31	5.18	$8.411 \cdot 10^{-9}$	62.90	5.98	$8.79 \cdot 10^{-9}$	80.14	6.33
D5	3	512	145152	476928	$3.162 \cdot 10^{-8}$	39.91	5.32	$9.811 \cdot 10^{-11}$	85.73	6.42	$7.479 \cdot 10^{-11}$	117.53	6.88
D5	4	4096	1096704	3750912	$1.147 \cdot 10^{-9}$	27.57	4.79	$3.306 \cdot 10^{-11}$	2.97	1.57	$3.624 \cdot 10^{-11}$	2.06	1.05
D5	5	32768	8515584	29749248	$3.421 \cdot 10^{-9}$	0.34	-1.58	$6.031 \cdot 10^{-11}$	0.55	-0.87	$6.67 \cdot 10^{-11}$	0.54	-0.88

Table 5.16: Convergence results for the DPG method according to TABLE 5.4.

5.3.2 A traveling wave-front in an inhomogeneous medium

As another example with analytically known solution, we consider a rectangular domain $\Omega = (-2, 2) \times (0, 1)$ with a non-homogeneous material distribution for the mass density ρ and the compression module κ given by

$$(\rho(x_1, x_2), \kappa(x_1, x_2)) = \begin{cases} (1, 1) & x_1 < 0, \\ (2, 0.5) & x_1 \in (0, 1), \\ (0.5, 2) & x_1 > 1. \end{cases}$$

such that the system

$$\partial_t p = \kappa \operatorname{div} \mathbf{v}, \quad \rho \partial_t \mathbf{v} = \nabla p$$

has a plane wave solution with amplitude $\alpha: \mathbb{R} \rightarrow \mathbb{R}$

$$\begin{pmatrix} p(x_1, x_2, t) \\ \mathbf{v}(x_1, x_2, t) \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \cdot \begin{cases} \alpha(x_1 - t) & x_1 < 0, \\ \alpha(2x_1 - t) & x_1 \in (0, 1), \\ \alpha(1.5 + 0.5x_1 - t) & x_1 > 1. \end{cases} \quad (5.2)$$

We test with $\alpha(s) = \cos(\pi s/2)^4$, $|s| < 1$, and $\alpha(s) = 0$ else, implying $\alpha \in C^3(\Omega)$. We use homogeneous NEUMANN boundary conditions $\mathbf{v} \cdot \mathbf{n} = 0$ for $y = 0$ and $y = 1$, and homogeneous DIRICHLET boundary conditions $p = 0$ for $x = \pm 2$.

Note that due to the special choice of material parameters, the analytical solution does not feature reflections at the material interfaces, cf. [31, Sec. 3.5]. FIGURE 5.18 illustrates the solution's evolution inside the space-time cylinder.

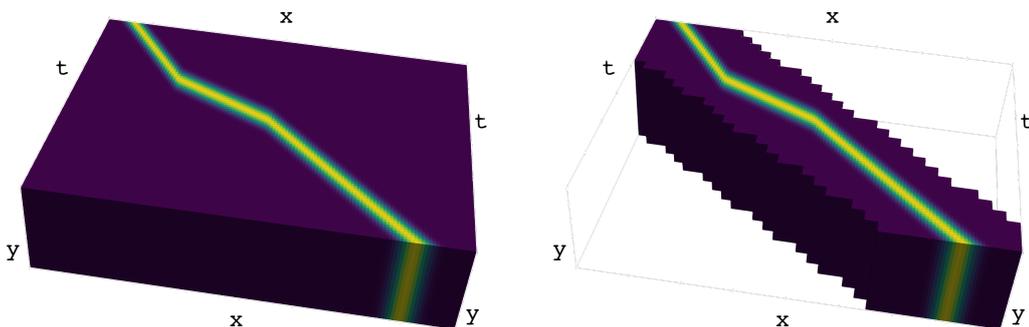


Figure 5.18: A wave front traveling from right to the left through three different materials. On the right, the mesh is truncated resulting in 1 284 984 DoFs compared to 3 193 344 DoFs in the full space-time mesh while the approximation quality remains unchanged, see TABLE 5.17 and TABLE 5.19

Locally increased polynomial degrees

Now, we exploit that the support of the solution is contained in a small fraction of the space-time cylinder Q . The analysis in SECTION 4.7 shows that the approximation error for the DPG method can be bounded by the interpolation error in the discrete space. Thus, we intend to choose an approximation space that contains polynomials of high degree only in areas where the solution does not vanish. Everywhere else, we want to select the lowest possible amount of degrees of freedom.

As a first step towards a space-time adaptive method, we demonstrate that an approximation space with locally increased polynomial degree yields the same approximation quality compared to the corresponding uniform space. To this end, we choose the lowest-order uniform configuration for the DPG method, i.e.

$$W_R = \mathbb{Q}_0(R)^2, \quad Z_R = \mathbb{Q}_2(R)^2, \quad \tilde{V}_{F,h} = \mathbb{Q}_0(F)^2,$$

in every space-time cell R and for all space-time faces F . We call this configuration D0, see TABLE 5.3 for the remaining DPG configurations. Then, we exploit that the solution in this case is known and increase the polynomial degree according the DPG configurations D1-D5 in the solution's space-time support, see FIGURE 5.19. The convergence results are presented in FIGURE 5.21 and TABLE 5.18.

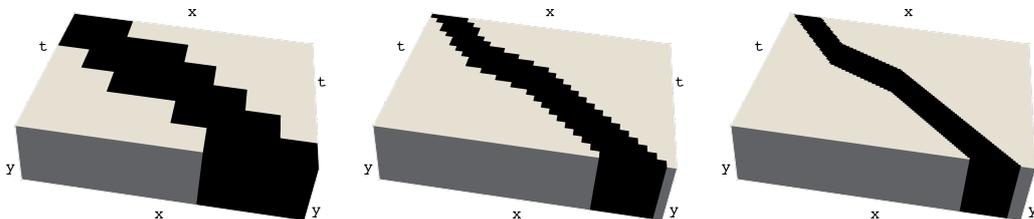


Figure 5.19: Space-time regions where the polynomial degree is increased on levels 1, 3, 5.

In practical applications, the solution's support is unknown. However, the same procedure can be used in combination with an error estimator, see SECTION 5.4 for an example.

Truncation of the space-time cylinder

Another possibility to reduce the degrees of freedom consists in truncating the space-time mesh by dropping cells where the solution vanishes. The resulting new space-time boundaries are equipped with zero boundary or initial conditions. As a result, we have reduced the amount of DoFs to approximate the solution while conserving the approximation quality. See FIGURE 5.18 for an example of this procedure and FIGURE 5.22, TABLE 5.19 for the convergence results.

5.3. Numerical examples (2D)

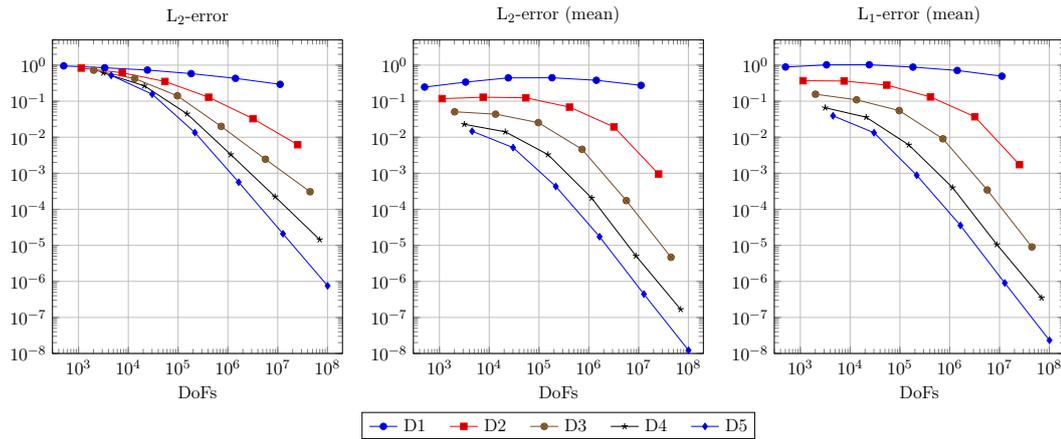


Figure 5.20: Convergence results for DPG with configurations as in TABLE 5.3 on the full mesh.

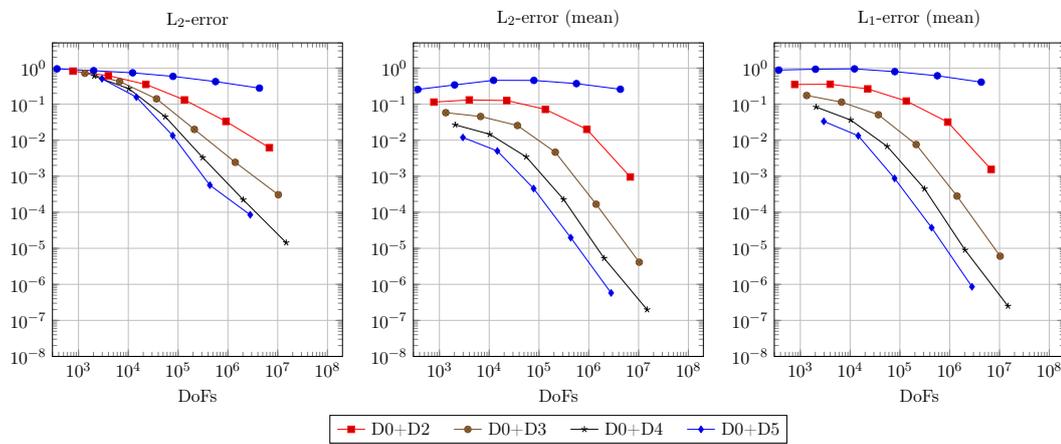


Figure 5.21: Convergence results for locally increased polynomial degrees on the full mesh.

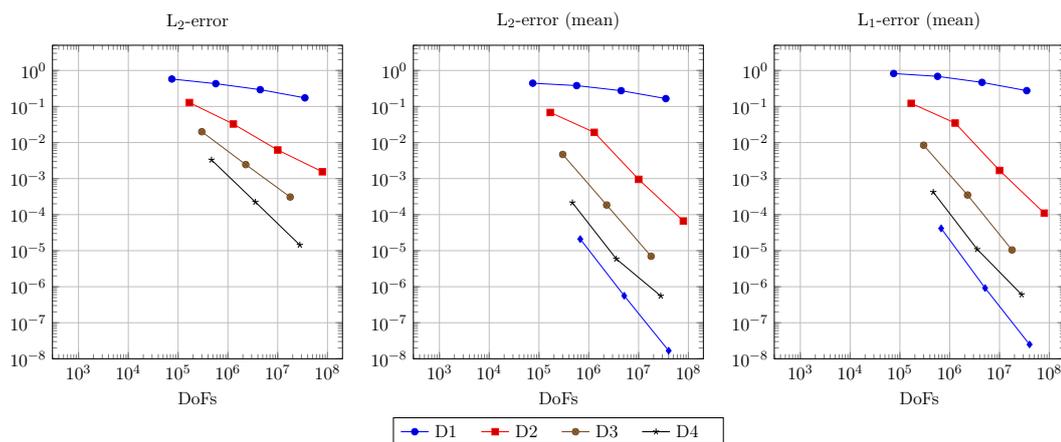


Figure 5.22: Convergence results for the truncated mesh. Level 0 corresponds to level 3 in FIGURE 5.20 with respect to the mesh width.

Discussion: DPG – full mesh In this example with non-homogeneous material distribution, we reproduce the convergence rates predicted by the theoretical considerations in SECTION 4.7 for the configurations D2 and D3 as well.

The lowest order method D1 does not show the expected behavior. Possibly, this happens because the pre-asymptotic regime is not left for these coarse meshes.

For the higher order methods in D4 and D5, the rates are better than the theoretical prediction, since the solution is C^3 only.

All methods except for D1 show improved convergence rates for the cell-wise mean values. However, differently than for the smooth example in the previous section, there is no further improvement for the highest-order methods. Thus, concerning the convergence of the mean-values, the methods in D3 to D5 behave similarly for small mesh-sizes.

Discussion: DPG – locally increased polynomial degrees The results show that the approximation quality stays the same when the polynomial degree is increased locally instead of everywhere in the mesh while the number of global degrees of freedom is reduced by a factor of about $\frac{1}{2}$ up to less than $\frac{1}{4}$.

Discussion: DPG – truncated mesh Since we solve the same problem on a truncated mesh, the results in FIGURE 5.22 extend those provided in FIGURE 5.20 where level 0 of the truncated mesh corresponds to level 3 of the full mesh.

Therefore, we expect the overlapping values to coincide for the different methods. This is true for the point-wise $L_2(Q)$ error and also for the difference of the cell-wise mean values up to small differences. For the $L_1(Q)$ norm of the cell-wise mean value, we observe larger discrepancies.

Considering the expected convergence rates, all methods except for the low-order scheme D1 behave as expected. However, since the rate in TABLE 5.19 approaches the expected rate 2 for finer grids, preasymptotic behavior remains as a possible explanation.

5.3. Numerical examples (2D)

conf	level	cells	DoFs	all DoFs	L_2 -error	rate	order	L_2 -error (mean)	rate	order	L_1 -error (mean)	rate	order
D1	0	12	504	540	$9.527 \cdot 10^{-1}$	–	–	$2.46 \cdot 10^{-1}$	–	–	$8.89 \cdot 10^{-1}$	–	–
D1	1	96	3360	3648	$8.451 \cdot 10^{-1}$	1.13	0.17	$3.382 \cdot 10^{-1}$	0.73	-0.46	$1.018 \cdot 10^0$	0.87	-0.20
D1	2	768	24192	26496	$7.31 \cdot 10^{-1}$	1.16	0.21	$4.437 \cdot 10^{-1}$	0.76	-0.39	$1.025 \cdot 10^0$	0.99	-0.01
D1	3	6144	182784	201216	$5.819 \cdot 10^{-1}$	1.26	0.33	$4.467 \cdot 10^{-1}$	0.99	-0.01	$8.817 \cdot 10^{-1}$	1.16	0.22
D1	4	49152	1419264	1566720	$4.3 \cdot 10^{-1}$	1.35	0.44	$3.802 \cdot 10^{-1}$	1.17	0.23	$7.12 \cdot 10^{-1}$	1.24	0.31
D1	5	393216	11182080	12361728	$2.933 \cdot 10^{-1}$	1.47	0.55	$2.748 \cdot 10^{-1}$	1.38	0.47	$4.932 \cdot 10^{-1}$	1.44	0.53
D2	0	12	1134	1422	$8.291 \cdot 10^{-1}$	–	–	$1.176 \cdot 10^{-1}$	–	–	$3.692 \cdot 10^{-1}$	–	–
D2	1	96	7560	9864	$6.155 \cdot 10^{-1}$	1.35	0.43	$1.287 \cdot 10^{-1}$	0.91	-0.13	$3.631 \cdot 10^{-1}$	1.02	0.02
D2	2	768	54432	72864	$3.521 \cdot 10^{-1}$	1.75	0.81	$1.242 \cdot 10^{-1}$	1.04	0.05	$2.794 \cdot 10^{-1}$	1.30	0.38
D2	3	6144	411264	558720	$1.286 \cdot 10^{-1}$	2.74	1.45	$6.844 \cdot 10^{-2}$	1.82	0.86	$1.31 \cdot 10^{-1}$	2.13	1.09
D2	4	49152	3193344	4372992	$3.267 \cdot 10^{-2}$	3.93	1.98	$1.928 \cdot 10^{-2}$	3.55	1.83	$3.698 \cdot 10^{-2}$	3.54	1.83
D2	5	393216	25159680	34596864	$6.196 \cdot 10^{-3}$	5.27	2.40	$9.529 \cdot 10^{-4}$	20.23	4.34	$1.735 \cdot 10^{-3}$	21.31	4.41
D3	0	12	2016	2988	$7.236 \cdot 10^{-1}$	–	–	$5.073 \cdot 10^{-2}$	–	–	$1.56 \cdot 10^{-1}$	–	–
D3	1	96	13440	21216	$4.24 \cdot 10^{-1}$	1.71	0.77	$4.366 \cdot 10^{-2}$	1.16	0.22	$1.091 \cdot 10^{-1}$	1.43	0.52
D3	2	768	96768	158976	$1.409 \cdot 10^{-1}$	3.01	1.59	$2.547 \cdot 10^{-2}$	1.71	0.78	$5.486 \cdot 10^{-2}$	1.99	0.99
D3	3	6144	731136	1228800	$1.998 \cdot 10^{-2}$	7.05	2.82	$4.627 \cdot 10^{-3}$	5.50	2.46	$9.024 \cdot 10^{-3}$	6.08	2.60
D3	4	49152	5677056	9658368	$2.447 \cdot 10^{-3}$	8.17	3.03	$1.742 \cdot 10^{-4}$	26.56	4.73	$3.44 \cdot 10^{-4}$	26.23	4.71
D3	5	393216	44728320	76578816	$3.073 \cdot 10^{-4}$	7.96	2.99	$4.674 \cdot 10^{-6}$	37.27	5.22	$9.03 \cdot 10^{-6}$	38.10	5.25
D4	0	12	3150	5454	$6.085 \cdot 10^{-1}$	–	–	$2.283 \cdot 10^{-2}$	–	–	$6.575 \cdot 10^{-2}$	–	–
D4	1	96	21000	39432	$2.673 \cdot 10^{-1}$	2.28	1.19	$1.414 \cdot 10^{-2}$	1.61	0.69	$3.588 \cdot 10^{-2}$	1.83	0.87
D4	2	768	151200	298656	$4.457 \cdot 10^{-2}$	6.00	2.58	$3.293 \cdot 10^{-3}$	4.29	2.10	$6.108 \cdot 10^{-3}$	5.87	2.55
D4	3	6144	1142400	2322048	$3.295 \cdot 10^{-3}$	13.53	3.76	$2.055 \cdot 10^{-4}$	16.02	4.00	$3.997 \cdot 10^{-4}$	15.28	3.93
D4	4	49152	8870400	18307584	$2.235 \cdot 10^{-4}$	14.74	3.88	$5.076 \cdot 10^{-6}$	40.48	5.34	$1.051 \cdot 10^{-5}$	38.03	5.25
D4	5	393216	69888000	145385472	$1.438 \cdot 10^{-5}$	15.54	3.96	$1.673 \cdot 10^{-7}$	30.34	4.92	$3.489 \cdot 10^{-7}$	30.12	4.91
D5	0	12	4536	9036	$5.197 \cdot 10^{-1}$	–	–	$1.465 \cdot 10^{-2}$	–	–	$3.921 \cdot 10^{-2}$	–	–
D5	1	96	30240	66240	$1.559 \cdot 10^{-1}$	3.33	1.74	$5.145 \cdot 10^{-3}$	2.85	1.51	$1.33 \cdot 10^{-2}$	2.95	1.56
D5	2	768	217728	505728	$1.344 \cdot 10^{-2}$	11.60	3.54	$4.306 \cdot 10^{-4}$	11.95	3.58	$8.811 \cdot 10^{-4}$	15.10	3.92
D5	3	6144	1645056	3949056	$5.657 \cdot 10^{-4}$	23.75	4.57	$1.737 \cdot 10^{-5}$	24.79	4.63	$3.565 \cdot 10^{-5}$	24.72	4.63
D5	4	49152	12773376	31205376	$2.096 \cdot 10^{-5}$	26.99	4.76	$4.431 \cdot 10^{-7}$	39.20	5.29	$9.007 \cdot 10^{-7}$	39.58	5.31
D5	5	393216	100638720	248094720	$7.554 \cdot 10^{-7}$	27.75	4.79	$1.222 \cdot 10^{-8}$	36.26	5.18	$2.323 \cdot 10^{-8}$	38.77	5.28

Table 5.17: Convergence results for the DPG method according to TABLE 5.3 for the full mesh.

5.3.2. A traveling wave-front in an inhomogeneous medium

conf	level	cells	DoFs	all DoFs	L ₂ -error	rate	order	L ₂ -error (mean)	rate	order	L ₁ -error (mean)	rate	order
D0+D1	0	12	369	405	$9.526 \cdot 10^{-1}$	—	—	$2.556 \cdot 10^{-1}$	—	—	$8.815 \cdot 10^{-1}$	—	—
D0+D1	1	96	2016	2304	$8.462 \cdot 10^{-1}$	1.13	0.17	$3.409 \cdot 10^{-1}$	0.75	-0.42	$9.296 \cdot 10^{-1}$	0.95	-0.08
D0+D1	2	768	12222	14526	$7.39 \cdot 10^{-1}$	1.14	0.20	$4.568 \cdot 10^{-1}$	0.75	-0.42	$9.501 \cdot 10^{-1}$	0.98	-0.03
D0+D1	3	6144	79008	97440	$5.878 \cdot 10^{-1}$	1.26	0.33	$4.544 \cdot 10^{-1}$	1.01	0.01	$7.966 \cdot 10^{-1}$	1.19	0.25
D0+D1	4	49152	564714	712170	$4.217 \cdot 10^{-1}$	1.39	0.48	$3.707 \cdot 10^{-1}$	1.23	0.29	$6.124 \cdot 10^{-1}$	1.30	0.38
D0+D1	5	393216	4297398	5477046	$2.78 \cdot 10^{-1}$	1.52	0.60	$2.584 \cdot 10^{-1}$	1.43	0.52	$4.077 \cdot 10^{-1}$	1.50	0.59
D0+D2	0	12	774	957	$8.228 \cdot 10^{-1}$	—	—	$1.138 \cdot 10^{-1}$	—	—	$3.52 \cdot 10^{-1}$	—	—
D0+D2	1	96	3976	5104	$6.157 \cdot 10^{-1}$	1.34	0.42	$1.309 \cdot 10^{-1}$	0.87	-0.20	$3.562 \cdot 10^{-1}$	0.99	-0.02
D0+D2	2	768	22512	29772	$3.526 \cdot 10^{-1}$	1.75	0.80	$1.257 \cdot 10^{-1}$	1.04	0.06	$2.645 \cdot 10^{-1}$	1.35	0.43
D0+D2	3	6144	134528	181856	$1.308 \cdot 10^{-1}$	2.70	1.43	$7.154 \cdot 10^{-2}$	1.76	0.81	$1.217 \cdot 10^{-1}$	2.17	1.12
D0+D2	4	49152	914544	1255200	$3.311 \cdot 10^{-2}$	3.95	1.98	$1.997 \cdot 10^{-2}$	3.58	1.84	$3.175 \cdot 10^{-2}$	3.83	1.94
D0+D2	5	393216	6800528	9416240	$6.186 \cdot 10^{-3}$	5.35	2.42	$9.526 \cdot 10^{-4}$	20.97	4.39	$1.544 \cdot 10^{-3}$	20.56	4.36
D0+D3	0	12	1341	1923	$7.182 \cdot 10^{-1}$	—	—	$5.789 \cdot 10^{-2}$	—	—	$1.736 \cdot 10^{-1}$	—	—
D0+D3	1	96	6720	10128	$4.233 \cdot 10^{-1}$	1.70	0.76	$4.536 \cdot 10^{-2}$	1.28	0.35	$1.136 \cdot 10^{-1}$	1.53	0.61
D0+D3	2	768	36918	57630	$1.4 \cdot 10^{-1}$	3.02	1.60	$2.548 \cdot 10^{-2}$	1.78	0.83	$5.074 \cdot 10^{-2}$	2.24	1.16
D0+D3	3	6144	212256	338016	$1.99 \cdot 10^{-2}$	7.04	2.81	$4.632 \cdot 10^{-3}$	5.50	2.46	$7.562 \cdot 10^{-3}$	6.71	2.75
D0+D3	4	49152	1404306	2269362	$2.426 \cdot 10^{-3}$	8.20	3.04	$1.671 \cdot 10^{-4}$	27.72	4.79	$2.805 \cdot 10^{-4}$	26.96	4.75
D0+D3	5	393216	10304910	16818510	$3.061 \cdot 10^{-4}$	7.93	2.99	$4.112 \cdot 10^{-6}$	40.64	5.35	$6.04 \cdot 10^{-6}$	46.44	5.54
D0+D4	0	12	2070	3429	$5.988 \cdot 10^{-1}$	—	—	$2.665 \cdot 10^{-2}$	—	—	$8.312 \cdot 10^{-2}$	—	—
D0+D4	1	96	10248	18096	$2.677 \cdot 10^{-1}$	2.24	1.16	$1.459 \cdot 10^{-2}$	1.83	0.87	$3.622 \cdot 10^{-2}$	2.30	1.20
D0+D4	2	768	55440	102348	$4.445 \cdot 10^{-2}$	6.02	2.59	$3.434 \cdot 10^{-3}$	4.25	2.09	$6.736 \cdot 10^{-3}$	5.38	2.43
D0+D4	3	6144	312192	590688	$3.287 \cdot 10^{-3}$	13.52	3.76	$2.271 \cdot 10^{-4}$	15.12	3.92	$4.476 \cdot 10^{-4}$	15.05	3.91
D0+D4	4	49152	2034000	3920256	$2.224 \cdot 10^{-4}$	14.78	3.89	$5.292 \cdot 10^{-6}$	42.91	5.42	$8.973 \cdot 10^{-6}$	49.88	5.64
D0+D4	5	393216	14810544	28914768	$1.43 \cdot 10^{-5}$	15.55	3.96	$1.986 \cdot 10^{-7}$	26.65	4.74	$2.49 \cdot 10^{-7}$	36.04	5.17
D0+D5	0	12	2961	5601	$5.047 \cdot 10^{-1}$	—	—	$1.189 \cdot 10^{-2}$	—	—	$3.319 \cdot 10^{-2}$	—	—
D0+D5	1	96	14560	29728	$1.561 \cdot 10^{-1}$	3.23	1.69	$5.01 \cdot 10^{-3}$	2.37	1.25	$1.324 \cdot 10^{-2}$	2.51	1.33
D0+D5	2	768	78078	168174	$1.338 \cdot 10^{-2}$	11.67	3.55	$4.545 \cdot 10^{-4}$	11.02	3.46	$8.704 \cdot 10^{-4}$	15.21	3.93
D0+D5	3	6144	434336	—	$5.626 \cdot 10^{-4}$	23.78	4.57	$1.972 \cdot 10^{-5}$	23.05	4.53	$3.723 \cdot 10^{-5}$	23.38	4.55
D0+D5	4	49152	2803626	—	$8.483 \cdot 10^{-5}$	6.63	2.73	$5.765 \cdot 10^{-7}$	34.20	5.10	$8.558 \cdot 10^{-7}$	43.50	5.44

Table 5.18: Convergence results for DPG with locally increased polynomial degree on the full mesh.

conf	level	cells	DoFs	all DoFs	L ₂ -error	rate	order	L ₂ -error (mean)	rate	order	L ₁ -error (mean)	rate	order
D1	0	2424	74904	82176	$5.789 \cdot 10^{-1}$	—	—	$4.427 \cdot 10^{-1}$	—	—	$8.222 \cdot 10^{-1}$	—	—
D1	1	19392	571104	629280	$4.301 \cdot 10^{-1}$	1.35	0.43	$3.802 \cdot 10^{-1}$	1.16	0.22	$6.852 \cdot 10^{-1}$	1.20	0.26
D1	2	155136	4456320	4921728	$2.932 \cdot 10^{-1}$	1.47	0.55	$2.747 \cdot 10^{-1}$	1.38	0.47	$4.667 \cdot 10^{-1}$	1.47	0.55
D1	3	1241088	35200512	38923776	$1.736 \cdot 10^{-1}$	1.69	0.76	$1.657 \cdot 10^{-1}$	1.66	0.73	$2.753 \cdot 10^{-1}$	1.70	0.76
D2	0	2424	168534	226710	$1.285 \cdot 10^{-1}$	—	—	$6.826 \cdot 10^{-2}$	—	—	$1.222 \cdot 10^{-1}$	—	—
D2	1	19392	1284984	1750392	$3.267 \cdot 10^{-2}$	3.93	1.98	$1.917 \cdot 10^{-2}$	3.56	1.83	$3.474 \cdot 10^{-2}$	3.52	1.82
D2	2	155136	10026720	13749984	$6.202 \cdot 10^{-3}$	5.27	2.40	$9.499 \cdot 10^{-4}$	20.18	4.34	$1.68 \cdot 10^{-3}$	20.68	4.37
D2	3	1241088	79201152	108987264	$1.544 \cdot 10^{-3}$	4.02	2.01	$6.656 \cdot 10^{-5}$	14.27	3.84	$1.099 \cdot 10^{-4}$	15.29	3.93
D3	0	2424	299616	495960	$1.999 \cdot 10^{-2}$	—	—	$4.67 \cdot 10^{-3}$	—	—	$8.41 \cdot 10^{-3}$	—	—
D3	1	19392	2284416	3855168	$2.443 \cdot 10^{-3}$	8.18	3.03	$1.847 \cdot 10^{-4}$	25.28	4.66	$3.497 \cdot 10^{-4}$	24.05	4.59
D3	2	155136	17825280	30391296	$3.079 \cdot 10^{-4}$	7.93	2.99	$6.995 \cdot 10^{-6}$	26.40	4.72	$1.044 \cdot 10^{-5}$	33.50	5.07
D4	0	2424	468150	933558	$3.297 \cdot 10^{-3}$	—	—	$2.146 \cdot 10^{-4}$	—	—	$4.203 \cdot 10^{-4}$	—	—
D4	1	19392	3569400	7292664	$2.229 \cdot 10^{-4}$	14.79	3.89	$5.888 \cdot 10^{-6}$	36.45	5.19	$1.091 \cdot 10^{-5}$	38.52	5.27
D4	2	155136	27852000	57638112	$1.436 \cdot 10^{-5}$	15.52	3.96	$5.543 \cdot 10^{-7}$	10.62	3.41	$6.049 \cdot 10^{-7}$	18.04	4.17
D5	0	2424	674136	1583136	$5.663 \cdot 10^{-4}$	—	—	$2.093 \cdot 10^{-5}$	—	—	$4.149 \cdot 10^{-5}$	—	—
D5	1	19392	5139936	12411936	$2.084 \cdot 10^{-5}$	27.17	4.76	$5.611 \cdot 10^{-7}$	37.30	5.22	$9.18 \cdot 10^{-7}$	45.20	5.50
D5	2	155136	40106880	98282880	$7.598 \cdot 10^{-7}$	27.43	4.78	$1.68 \cdot 10^{-8}$	33.40	5.06	$2.497 \cdot 10^{-8}$	36.76	5.20

Table 5.19: Results for DPG on the truncated mesh. Here, level 0 corresponds to level 3 in TABLE 5.17.

5.3.3 A double slit experiment in a homogeneous medium

Here, we consider a classical experiment from optical physics. See also [23] for applications of space-time Discontinuous-GALERKIN methods to a similar example.

Two coherent wave fronts enter the domain through a pair of small slits. By HUYGENS principle, a circular wave is propagated from each of the slits yielding a characteristic interference pattern, cf. FIGURE 5.23 for a description of the setup and FIGURE 5.24 for visualizations of the solution. The boundary $\partial\Omega = \Gamma_N \cup \Gamma_D$ is partitioned in a Neumann part Γ_N and a Dirichlet part Γ_D , where we use $\mathbf{v} \cdot \mathbf{n}_\Omega = 0$ on $\Gamma_N \times (0, T)$ and $p(x, t) = \sin(2\pi\omega(x - t))$ for $(x, t) \in \Gamma_D \times (0, T)$ with $\omega = 2$ and $T = 10$.

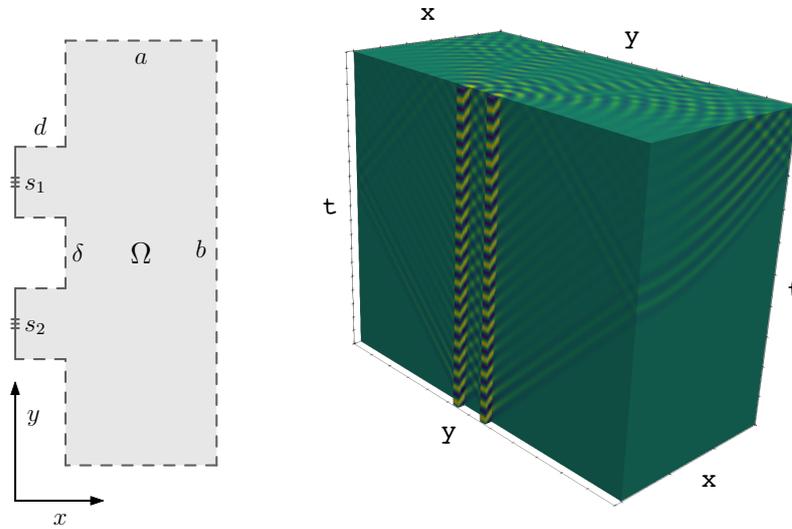


Figure 5.23: The spatial domain Ω is described on the left, where the slit dimensions are $d = s_1 = s_2 = 0.25$, their distance is $\delta = 1$, and the dimensions of the large rectangle are $a = 6$, $b = 12$. The domain Ω is substructured using a regular mesh Ω_h of squares with side lengths 0.25. The corresponding space-time cylinder $Q = \Omega \times (0, T)$ is discretized using tensor-product elements $R = K \times (t_{n-1}, t_n)$ for each cell $K \in \Omega_h$ and $t_n = T/N$, $n = 0, \dots, N$, with $T = 10$ and $N = 50$. The dashed portion of $\partial\Omega$ indicates Γ_N and the remaining faces, marked by three lines, represent Γ_D . On the right, a space-time plot of the solution is given on a two times refined version of this mesh featuring 3 692 800 space-time cells and 234 210 528 face DoFs.

This example demonstrates that the space-time DPG method is able to simulate complex waveforms in homogeneous media.

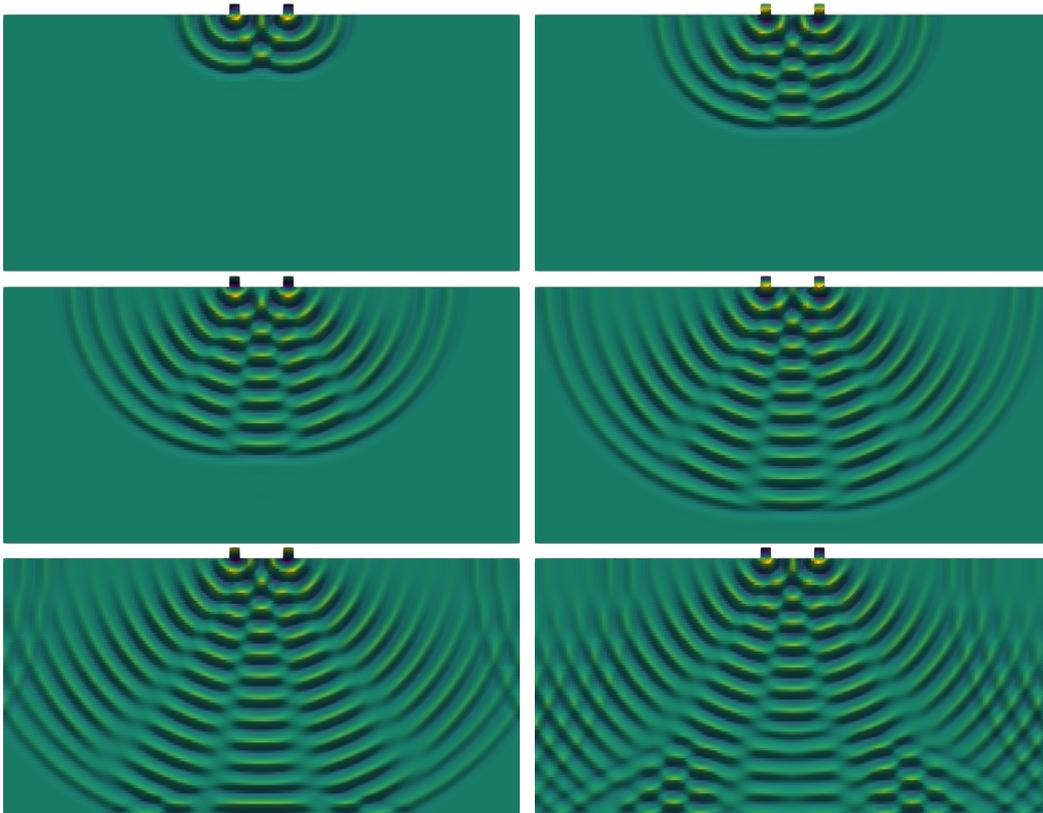


Figure 5.24: Snapshots of the pressure component at times $t = 0.6, 2.08, 3.56, 5.04, 6.52, 8$. These were obtained by slicing the space-time solution from Fig. 5.23 along planes that are orthogonal to the time direction.

5.4 Adaptivity with respect to the polynomial degree

In order to demonstrate the space-time adaptivity features of the DPG method, we revisit the example considered in SECTION 5.3.2. Since the solution given by (5.2) is invariant with respect to translations in y direction, for simplicity, we restrict ourselves to the one-dimensional projection, see FIGURE 5.25.

Since in practical applications, we do not know the solutions structure, we make use of DPG's built-in error estimator, cf. (4.37), to determine the parts of the mesh where high-degree polynomials are useful. As a second step, we adjust the local polynomial degrees similar to SECTION 5.3.2

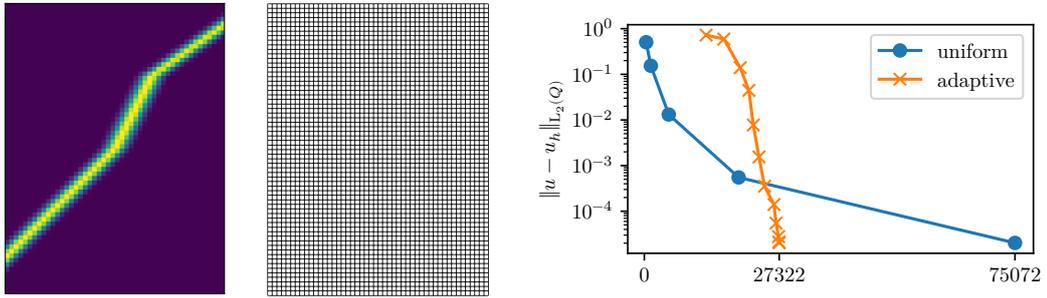


Figure 5.25: Left: space-time plot of the projected solution from (5.2) with time-axis from left to right. Center: the space-time mesh with 3072 cells used for adaptivity in the polynomial degree. Right: evolution of the approximation error for uniform and adaptive refinement where the horizontal axis corresponds to global degrees of freedom.

Then, we apply for fixed $\vartheta \in (0, 1)$, here $\vartheta = 0.98$, the iterative procedure described in ALGORITHM 2. The resulting refinements and corresponding solutions are shown in FIGURE 5.26 and the errors are shown in FIGURE 5.25.

Algorithm 2 Adaptive refinement

- 1: **while** not converged **do**
 - 2: Obtain the numerical solution $\mathbf{y}_h \in V_h$ for the current discretization.
 - 3: Calculate the error-representing function ψ and $(\|\psi\|_{L^*,R})_{R \in \mathcal{R}_h}$.
 - 4: Select $\tau: \{1, \dots, |\mathcal{R}_h|\} \rightarrow \mathcal{R}_h: \|\psi_{\tau(1)}\|_{L^*,\tau(1)} \leq \dots \leq \|\psi_{\tau(|\mathcal{R}_h|)}\|_{L^*,\tau(|\mathcal{R}_h|)}$.
 - 5: Identify the minimal $k \in \mathbb{N}$ such that $\sqrt{\sum_{l=1}^{k+1} \|\psi_{\tau(l)}\|_{L^*,\tau(l)}^2} \geq \vartheta \|\psi\|_{L^*,Q_h}$.
 - 6: For $R \in \tau(\{1, \dots, k\})$, change DPG configuration from Dl to $D(l+1)$ if $l < 4$.
 - 7: For adjacent cells of $\tau(\{1, \dots, k\})$ use the largest test space of their neighbors.
-

The error estimator identifies a superset of the solution's support following a marching-like pattern in time. After about 11 iterations, the error has reached the quality for uniform refinement. However, about a third of global degrees of freedom is needed. See [33, 52] for similar examples.

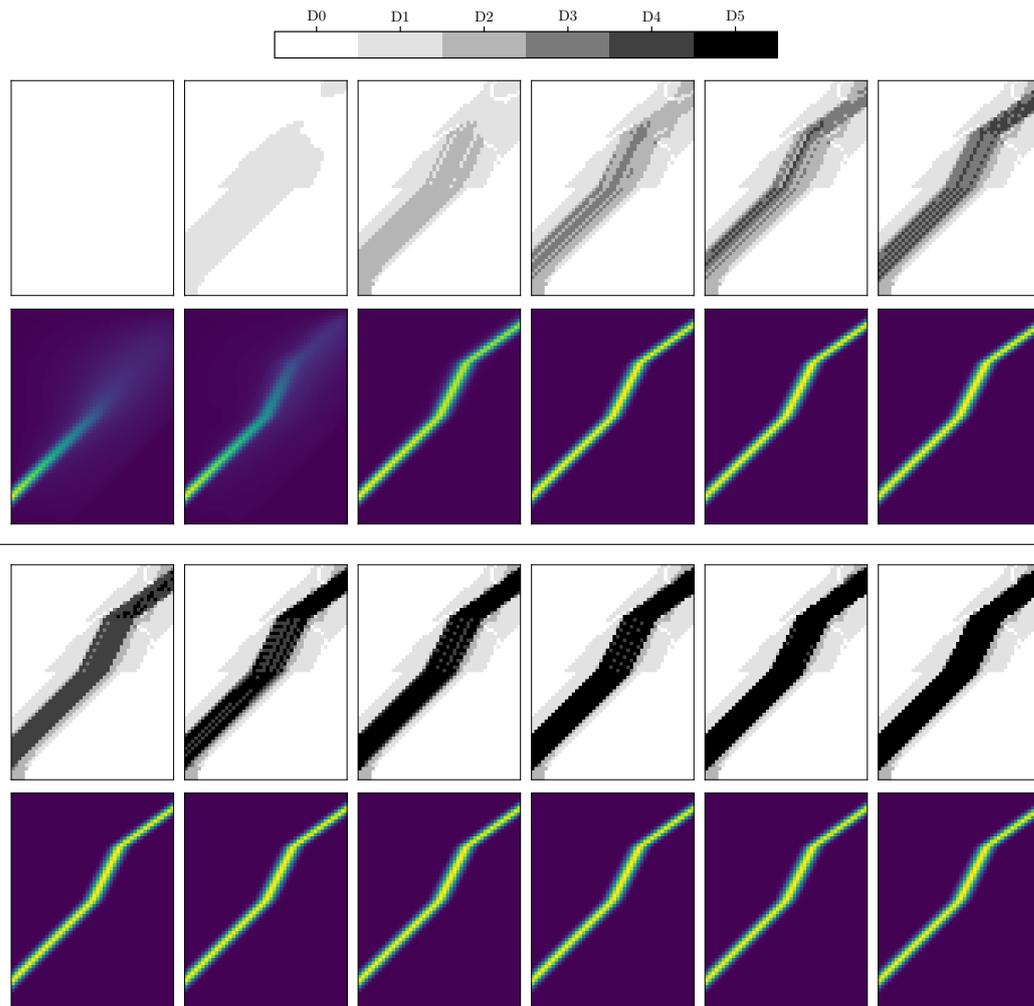


Figure 5.26: Space-time plots of areas with increased polynomial degree (rows 1, 3) and space time plot of the corresponding numerical solution's pressure component (rows 2, 4).

5.5 Summary

Considering the experiments, the results largely fit or are better than the theoretical predictions provided by SECTION 4.7.

For the DPG method, increasing the polynomial degree inside the cells by 1 improves the rate of convergence in one spatial dimension. In two spatial dimensions, the increased cell degree affects the order less strongly compared to the one-dimensional case. However, at least for smooth examples, the approximation quality for the same amount of global degrees of freedom improves while the convergence rates oscillate. Since we do not observe further improvements when raising the polynomial degree inside the cells by more than one, we do not provide examples for this case.

For one and two spatial dimensions, we observe an improved rate of convergence of the DPG approximation when considering the distance of cell-wise mean values. In particular, this can be observed for the low regularity example in one spatial dimension.

We demonstrate that the DPG method is well suited for space-time adaptivity. Both approaches, the locally increased polynomial degrees as well as the truncation of the space-time cylinder, yield the same approximation quality than a calculation on the full mesh with uniform polynomial degree distribution. Exploiting DPG's built-in error estimator, we present a way to automatically determine the space-time regions where refinement is advantageous.

The weakly conforming Least-Squares method delivers comparable performance considering the difference in $L_2(Q)$ norms, while being restricted to one spatial dimension in its current state, see SECTION 4.2.5.

For this method, however, we do not observe the effect of improved convergence rates for the cell-wise mean values in general. We only observe this effect for the low regularity example.

Chapter 6

Full Waveform Inversion (FWI)

In this chapter, we consider the problem of Full Waveform Inversion (FWI). The challenge in FWI consists in reconstructing spatial material properties from surface measurements of the wave field. FWI plays an important role to investigate the subsurface structure of the earth, see [61, 65] for an extensive overview and [63] for the foundation paper.

In the geophysics community, many strategies to tackle FWI have been evaluated. Usually, gradient-based iterative inversion schemes are the methods of choice, cf. [48] for a toolbox implementing various examples. Here, Finite Difference discretization are the standard method to solve the forward problems numerically, see e.g. [8, 59, 64].

Since gradient-based algorithms rely on linearization, the differentiability of the parameter-to-solution map is required and thus, has been investigated in [41, 43].

For practical implementations, usually a technique called adjoint-state method is used to efficiently handle the linearized problem, see e.g. [53].

Mathematically, FWI is an *inverse problem*, i.e. one tries to reconstruct the cause (material properties and the source signal) from its impacts (resulting wave field at the receiver positions). Typically, inverse problems are ill-posed in the sense that the unknown does not depend continuously on the data, see [43] for a proof that this also applies for FWI and e.g. [40, 58] for introductory monographs on inverse problems. Due to this ill-posedness, solving inverse problems using numerical methods is a challenging task.

Outline

We set up an abstract variational framework to derive different algorithms to solve the FWI problem numerically. In particular, we compare a root finding approach and an optimization approach both using NEWTON-type methods. We consider an inexact NEWTON-type method using a regularized CG-scheme as a solver for the linearized problem called CG-REGINN that has been considered in [57]. Here, *inexact* means that the linearized problem inside the NEWTON iteration is solved approximately. This preliminary stopping results in a regularization effect to handle the ill-posedness.

Under certain assumptions on the problem, there exists a convergence proof for the CG-REGINN algorithm, see [57]. This analysis is extended to special classes of BANACH spaces, see [46] where also KACZMARZ variants of this algorithm are studied.

CG-REGINN has been successfully applied to other inverse problems such as Electrical Impedance Tomography, see e.g. [69]. For defect detection of elastic structures, CG-REGINN has been applied in [44]. Here, we would like to start a discussion of regularized inexact NEWTON algorithms for FWI.

Although to the knowledge of the author it is an open question whether the convergence results for CG-REGINN can be transferred to the FWI problem, we provide a numerical example using CG-REGINN and the DPG method for acoustic waves as described in SECTION 4.3 to address a simple model problem.

6.1 Mathematical setting and notation

In this chapter, we drop mathematical rigor to focus on the formal derivation of the inversion algorithms. Beforehand, we introduce some notation that is needed in the following. In particular, we extend the language that has been introduced in CHAPTER 3.

We consider a space of material parameters \mathcal{P} and a subset of admissible material parameters $\mathcal{P}^{\text{adm}} \subset \mathcal{P}$. Further, let W be the space of right-hand sides for the wave equation and let $V \subset W$ be a W -dense subspace of wave-fields. For a linear operator $A: V \rightarrow W$ and fixed $m \in \mathcal{P}^{\text{adm}}$, we consider the operator

$$L_m \mathbf{y} := M(m) \partial_t \mathbf{y} + A \mathbf{y},$$

where $M: \mathcal{P} \rightarrow \mathcal{L}(W)$ is a differentiable mapping.

For the acoustic wave equation, we have

$$\begin{aligned} \mathcal{P} &:= L_\infty(\Omega, \mathbb{R}^2), & \mathcal{P}^{\text{adm}} &\subset L_\infty(\Omega, \mathbb{R}^2)^+, \\ W &:= L_2(Q, \mathbb{R}^{1+d}), & V &\subset H(L_m, Q) \end{aligned}$$

for the space

$$L_\infty(\Omega)^+ = \{f \in L_\infty(\Omega) : \exists c > 0 \text{ such that } f(x) \geq c \text{ for almost all } x \in \Omega\}$$

and the operators

$$M(\kappa(x), \rho(x)) = \begin{pmatrix} \kappa(x)^{-1} & 0 \\ 0 & \rho(x)I_d \end{pmatrix}, \quad x \in \Omega; \quad A = \begin{pmatrix} 0 & \text{div} \\ \nabla & 0 \end{pmatrix}. \quad (6.1)$$

We assume that the space V remains unchanged for different material parameters $m \in \mathcal{P}^{\text{adm}}$ and that $\partial_t \mathbf{y} \in W$ is well-defined for all $\mathbf{y} \in V$.

Remark 6.1. *To our knowledge, it is an open question, whether a unified domain for the family of operators $(L_m)_{m \in \mathcal{P}^{\text{adm}}}$ can be chosen in a HILBERT space setting as described in CHAPTER 3.*

In the simplest case, however, the following calculations can be performed after replacing the operator A by an approximation in a finite dimensional subspace.

6.1.1 The forward problem and the adjoint problem

For a given right hand side $\mathbf{b} \in W$ and material parameter $m \in \mathcal{P}^{\text{adm}}$ solving the forward problem means finding $\mathbf{y} \in V$ with

$$L_m \mathbf{y} = \mathbf{b} \quad \text{i. e.,} \quad M(m) \partial_t \mathbf{y} + A \mathbf{y} = \mathbf{b}. \quad (6.2)$$

This defines the parameter-to-solution map

$$\mathcal{F}: \mathcal{P}^{\text{adm}} \longrightarrow V, m \longmapsto \mathbf{y}^{\text{sol}}, \quad \text{where } \mathbf{y}^{\text{sol}} \text{ solves (6.2) for } m.$$

Further, we assume that there is an operator $L_m^*: V^* \longrightarrow W$ such that

$$(L_m \mathbf{y}, \mathbf{z})_Q = (\mathbf{y}, L_m^* \mathbf{z})_Q \quad \mathbf{y} \in V, \mathbf{z} \in V^*,$$

for a subspace $V^* \subset W$. We assume that both, $L_m: V \longrightarrow W$ and $L_m^*: V^* \longrightarrow W$, are isomorphisms. This ensures that the parameter-to-solution map is well-defined.

Example 6.2. *For the acoustic wave-equation, we have $L^* = -L$ defined on a space featuring final conditions instead of initial conditions, see CHAPTER 3.*

Remark 6.3. *Note that in practical applications, the right-hand side b is not known and has to be reconstructed as well, see [54]. However, for simplicity we assume that only the material parameters m have to be reconstructed.*

In practice, we do not have access to the full wavefield in space and time. Typically, finitely many measuring devices yield sequences of approximate point evaluations being the only accessible data. In the next section, we model this measurement procedure.

6.1.2 Observations and the parameter-to-seismogram map

For a finite set of space-time receiver points $\mathcal{M} \subset Q$, we consider the space of seismograms $\mathcal{S} := (\mathbb{R}^{1+d})^{\mathcal{M}}$ equipped with the L_2 inner product and corresponding norm

$$(s, \tilde{s})_{\mathcal{S}} := \sum_{r \in \mathcal{M}} \sum_{i=1}^{d+1} s(r)_i \tilde{s}(r)_i, \quad \|s\|_{\mathcal{S}} = \left(\sum_{r \in \mathcal{M}} |s(r)|_2^2 \right)^{\frac{1}{2}}, \quad s, \tilde{s} \in \mathcal{S}.$$

We consider a linear observation operator $\Psi: W \rightarrow \mathcal{S}$ mapping a wave $(p, \mathbf{v}) \in W$ to a seismogram $s = \Psi[(p, \mathbf{v})] \in \mathcal{S}$.

Often, the observation operator is realized by approximate point evaluations using measurement kernels $\varphi_r \in C_c^\infty(Q)$, $r \in \mathcal{M}$, by

$$\Psi[(p, \mathbf{v})] := \left((\varphi_r, p)_Q, (\varphi_r, \mathbf{v}_1)_Q, \dots, (\varphi_r, \mathbf{v}_d)_Q \right)_{r \in \mathcal{M}} \quad (6.3)$$

Here, the measurement kernels are chosen as representatives of a DIRAC sequence in space-time approximating the delta distribution.

For the inversion algorithm, we need the $L_2(Q)$ adjoint of Ψ . A straight-forward calculation shows that this adjoint $\Psi^* \in \mathcal{L}(\mathcal{S}, W)$ is given by

$$\Psi^* s = \sum_{r \in \mathcal{M}} \begin{pmatrix} s(r)_1 \varphi_r \\ \vdots \\ s(r)_{1+d} \varphi_r \end{pmatrix} \in W. \quad (6.4)$$

As a consequence, for given $s \in \mathcal{S}$, the object $\Psi^* s$ yields a right-hand side for the wave equation consisting of approximate point sources at all space-time receiver positions. These sources are scaled by the value of the seismogram s at the receivers.

Remark 6.4. *The observation operator Ψ in (6.3) measures all components of the wave field (p, \mathbf{v}) . In practical applications, the measurements might be restricted due to physical or technical reasons.*

Remark 6.5. *Note that for simplicity, all components in the inner product are scaled with a constant 1. To respect the physical scaling of the quantities, a differently scaled product can be chosen.*

6.1.3 The problem in FWI

Using the observation operator Ψ and the parameter-to-solution map \mathcal{F} , we define the parameter-to-seismogram operator by

$$\Phi: \mathcal{P}^{\text{adm}} \longrightarrow \mathcal{S}, \quad m \longmapsto \Psi[\mathcal{F}(m)].$$

In the application, we have access to a seismogram $s_{\text{obs}} \in \mathcal{S}$ generated by an unknown material parameter $m^{\text{sol}} \in \mathcal{P}$. Now, the problem of FWI consists in finding $m^{\text{sol}} \in \mathcal{P}$ such that $\Phi(m^{\text{sol}})$ fits the data s_{obs} . Formally, we define the problem of Full Waveform inversion by

$$\begin{cases} \text{Given } s_{\text{obs}} \in \mathcal{S}, \\ \text{find } m^{\text{sol}} \in \mathcal{P}^{\text{adm}} \text{ with } \Phi(m^{\text{sol}}) = s_{\text{obs}}. \end{cases} \quad (6.5)$$

Remark 6.6. *Often, the right-hand side b models a point source with a specific location. In this case a pair of source and corresponding seismograms, i.e. (b, s_{obs}) , is called a shot gather or a shot.*

Typically, the input data consist of more than a single shot. However, for the simplicity of notation, we restrict the presentation to the single-shot case.

It turned out that (6.5) is a challenging problem for two reasons. On the one hand, it is highly non-linear and iterative schemes require high-accuracy solutions of at least one wave equation in every step. On the other hand, numerical experiments showed that the robust reconstruction of m is a non-trivial task since typically spurious artifacts occur during the inversion. A mathematically satisfying explanation for this phenomenon was given in [42] by showing that (6.5) is locally ill-posed in the following sense.

Theorem 6.7. *The equation $\mathcal{F}(m) = \mathbf{y}$ is locally ill-posed in every $m \in \mathcal{P}^{\text{adm}}$, i.e. in any neighborhood of m there is a sequence $(m^k)_k \in (\mathcal{P}^{\text{adm}})^{\mathbb{N}}$ such that*

$$\lim_{k \rightarrow \infty} \|\mathcal{F}(m^k) - \mathcal{F}(m)\|_{L_2(Q, \mathbb{R}^3)} = 0, \quad \|m^k - m\|_{L_\infty(\Omega, \mathbb{R}^2)} \not\rightarrow 0, \quad k \rightarrow \infty.$$

Proof. See [42]. □

Since the result in THEOREM 6.7 assumes knowledge of the *whole* space-time wavefield and thus, does not depend on any observation operator, the ill-posedness is an inherent property of the wave equation and does not result from a lack of data.

Remark 6.8. In [43], the authors prove a variant of THEOREM 6.7 for abstract evolution equations. This result shows the ill-posedness of the FWI problem for the elastic wave equation as well as for MAXWELL's equations.

State of the art methods to solve non-linear ill-posed problems numerically are regularized NEWTON type methods. In the following, we consider variants of these methods for (6.5). We present two different points of view to tackle (6.5) numerically: a root-finding approach and an optimization approach.

6.2 Root-finding approach

Problem (6.5) can be considered as a root-finding procedure for the mapping

$$\Theta: \mathcal{P}^{\text{adm}} \longrightarrow \mathcal{S}, \quad m \longmapsto \Phi(m) - s_{\text{obs}}.$$

Then, a NEWTON-type method can be used to find the solution of $\Theta(m) = 0$. For the comfort of the reader, we briefly revisit the construction of NEWTON's method. Given a guess¹ $m^k \in \mathcal{P}$, we consider the TAYLOR expansion of $\Theta(m^{\text{sol}})$ in a neighborhood of m^k , i.e.

$$\Theta(m^{\text{sol}}) = \Theta(m^k) + \Theta'(m^k)[m^{\text{sol}} - m^k] + \mathcal{O}(\|m^{\text{sol}} - m^k\|^2).$$

Since $\Theta(m^{\text{sol}}) = 0$, the (unknown) update $h^k := m^{\text{sol}} - m^k$ fulfills approximately

$$\Phi'(m^k)[h^k] = \Theta'(m^k)[h^k] \approx -\Theta(m^k) = s_{\text{obs}} - \Phi(m^k). \quad (6.6)$$

The idea is to find an approximation for the update h^k by solving (6.6). Thus, NEWTON's method can be formulated as in ALGORITHM 3.

Algorithm 3 NEWTON's method for root finding

- 1: Choose $m^0 \in \mathcal{P}^{\text{adm}}$, $k \leftarrow 0$
 - 2: **while** not converged **do**
 - 3: $r^k \leftarrow s_{\text{obs}} - \Phi(m^k) \in \mathcal{S}$
 - 4: Find $h^k \in \mathcal{P}$ with $\Phi'(m^k)[h^k] = r^k$
 - 5: $m^{k+1} \leftarrow m^k + h^k$
 - 6: $k \leftarrow k + 1$
-

Often, the linearized problem that has to be solved in step 4 is also ill-posed and requires regularization, see e.g. [37, 38]. One way to do this is applying a linear regularization method to the corresponding normal equation

$$\Phi'(m^k)^* \Phi'(m^k)[h^k] = \Phi'(m^k)^* r^k, \quad (6.7)$$

¹The index k corresponds to the NEWTON iteration.

where the adjoint $\Phi'(m^k)^*: \mathcal{S} \rightarrow \mathcal{P}$ is chosen such that

$$(\Phi'(m^k)[\tilde{m}], s)_{\mathcal{S}} = (\tilde{m}, \Phi'(m^k)^*s)_{\Omega}, \quad \text{for all } \tilde{m} \in \mathcal{P}, s \in \mathcal{S}.$$

Since we have $\Phi = \Psi \circ \mathcal{F}$, we obtain by the linearity of Ψ

$$\Phi'(m^k)[\tilde{m}] = \Psi'(\mathcal{F}(m^k))[\mathcal{F}'(m^k)[\tilde{m}]] = \Psi[\mathcal{F}'(m^k)[\tilde{m}]], \quad m^k \in \mathcal{P}^{\text{adm}}, \tilde{m} \in \mathcal{P},$$

and thus

$$\Phi'(m^k)^*[s] = \mathcal{F}'(m^k)^*[\Psi^*[s]], \quad m^k \in \mathcal{P}, s \in \mathcal{S}.$$

Here, the adjoints Ψ^* and $\mathcal{F}'(m^k)^*$, $m^k \in \mathcal{P}^{\text{adm}}$, are chosen such that for all $\tilde{m} \in \mathcal{P}$, $s \in \mathcal{S}$ and $\mathbf{y} \in V$ it holds

$$(\Psi[\mathbf{y}], s)_{\mathcal{S}} = (\mathbf{y}, \Psi^*[s])_Q \quad \text{and} \quad (\mathcal{F}'(m^k)[\tilde{m}], \mathbf{y})_Q = (\tilde{m}, \mathcal{F}'(m^k)^*[\mathbf{y}])_{\Omega}.$$

Remark 6.9. *Note that the adjoint operators strongly depend on the inner products $(\cdot, \cdot)_{\mathcal{S}}$, $(\cdot, \cdot)_Q$ and $(\cdot, \cdot)_{\Omega}$.*

For our considerations in this work, we use standard unscaled L_2 products. However, it is an interesting future challenge to construct and evaluate other variants of these products. In [69], a scaled L_2 norm in the parameter space reduces spurious oscillations in the reconstructions for an inverse problem in electrical impedance tomography (EIT).

6.3 The derivative of \mathcal{F} and its adjoint

We assume that the regularization scheme only needs – apart from basic vector algebra – to evaluate

$$\Phi'(m^k)[h] \quad \text{and} \quad \Phi'(m^k)^*[s]$$

for different choices of $m^k \in \mathcal{P}^{\text{adm}}$, $h \in \mathcal{P}$ and $s \in \mathcal{S}$. Thus, in each NEWTON step, we need to evaluate $\mathcal{F}'(m^k)[\tilde{m}]$ for given $\tilde{m} \in \mathcal{P}$ at the current iterate $m^k \in \mathcal{P}^{\text{adm}}$. In the following, we formally construct algorithms to explicitly evaluate $\Phi'(m^k)[h]$ and $\Phi'(m^k)^*[s]$.

As described e.g. in [16], we define the state mapping $\mathcal{E}: \mathcal{P}^{\text{adm}} \times V \times V^* \rightarrow \mathbb{R}$,

$$\mathcal{E}(m, \mathbf{y}, \mathbf{z}) = (L_m \mathbf{y} - \mathbf{b}, \mathbf{z})_Q = (M(m) \partial_t \mathbf{y} + A \mathbf{y} - \mathbf{b}, \mathbf{z})_Q,$$

fulfilling $\mathcal{E}(m, \mathcal{F}(m), \mathbf{z}) = 0$ for all $m \in \mathcal{P}$, $\mathbf{z} \in V^*$ by the definition of \mathcal{F} .

6.3.1 The derivative $\mathcal{F}'(m^k)$

We carry out a formal derivation for a computable representation of the derivative. The justification that this indeed yields a proper derivative in a strict mathematical sense is a question that has been addressed for instance in [41], also see [35].

For arbitrary $m \in \mathcal{P}^{\text{adm}}$, $\tilde{m} \in \mathcal{P}$ and $\mathbf{z} \in V^*$, we formally have

$$\begin{aligned} 0 &= \left\langle \frac{\text{d}}{\text{d}m} (\mathcal{E}(m, \mathcal{F}(m), \mathbf{z})), \tilde{m} \right\rangle \\ &= \langle \partial_1 \mathcal{E}(m, \mathcal{F}(m), \mathbf{z}), \tilde{m} \rangle + \langle \partial_2 \mathcal{E}(m, \mathcal{F}(m), \mathbf{z}), \mathcal{F}'(m)[\tilde{m}] \rangle \\ &= (M'(m)[\tilde{m}] \partial_t \mathcal{F}(m), \mathbf{z})_Q + (L_m \mathcal{F}'(m)[\tilde{m}], \mathbf{z})_Q. \end{aligned} \quad (6.8)$$

As a result, for given $\tilde{m} \in \mathcal{P}$ the linearized parameter-to-solution map $\mathcal{F}'(m^k)[\tilde{m}]$ fulfills at the current iterate $m^k \in \mathcal{P}^{\text{adm}}$

$$L_{m^k} [\mathcal{F}'(m^k)[\tilde{m}]] = -M'(m^k)[\tilde{m}] \partial_t \mathcal{F}(m^k) \quad (6.9)$$

which means that solving the forward problem (6.9) yields $\mathcal{F}'(m^k)[\tilde{m}] \in V$, see ALGORITHM 4.

Algorithm 4 Evaluate the linearized parameter-to-wavefield map $\mathcal{F}'(m^k)[\tilde{m}]$

Input: $m^k \in \mathcal{P}^{\text{adm}}$, $\tilde{m} \in \mathcal{P}$, $\mathbf{y}_{m^k} := \mathcal{F}(m^k) \in V$

Output: $\mathbf{y} := \mathcal{F}'(m^k)[\tilde{m}] \in V$

1: Find $\mathbf{y} \in V$ with $L_{m^k} \mathbf{y} = -M'(m^k)[\tilde{m}] \partial_t \mathbf{y}_{m^k}$

Example 6.10 (Linear acoustics). *In case of the linear acoustic wave equation, we have for $m^k = (\kappa, \rho) \in \mathcal{P}^{\text{adm}}$*

$$M'(\kappa, \rho)[(\tilde{\kappa}, \tilde{\rho})] = \begin{pmatrix} -\kappa^{-2} \tilde{\kappa} & 0 \\ 0 & \tilde{\rho} I_d \end{pmatrix}. \quad (\tilde{\kappa}, \tilde{\rho}) \in \mathcal{P},$$

With $\mathbf{y} = (p, \mathbf{v}) = \mathcal{F}'(m^k)[(\tilde{\kappa}, \tilde{\rho})]$ and $\mathbf{y}_{m^k} = (p_{m^k}, \mathbf{v}_{m^k})$, this yields for step 1 in ALGORITHM 4

$$\begin{aligned} \frac{1}{\kappa} \partial_t p + \text{div } \mathbf{v} &= \frac{\tilde{\kappa}}{\kappa^2} \partial_t p_{m^k}, \\ \rho \partial_t \mathbf{v} + \nabla p &= -\tilde{\rho} \mathbf{v}_{m^k}, \end{aligned}$$

for all $(\tilde{\kappa}, \tilde{\rho}) \in \mathcal{P}$, cf. also [4, Thm. 8] and [43, Thm. 3.6].

6.3.2 The adjoint $\mathcal{F}'(m^k)^*$ of the derivative $\mathcal{F}'(m^k)$

Continuing the calculation in (6.8) for $m = m^k \in \mathcal{P}^{\text{adm}}$ yields

$$(\mathcal{F}'(m^k)^*[L_{m^k}^* \mathbf{z}], \tilde{m})_\Omega = -(M'(m^k)[\tilde{m}] \partial_t \mathcal{F}(m^k), \mathbf{z})_Q.$$

Thus, to obtain $\mathcal{F}'(m^k)^*[\tilde{\mathbf{b}}] \in \mathcal{P}$ for given $\tilde{\mathbf{b}} \in W$, find $\mathbf{z} \in V^*$ such that

$$L_{m^k}^* \mathbf{z} = \tilde{\mathbf{b}} \quad (6.10)$$

and in a second step find $\bar{m} \in \mathcal{P}$ solving

$$(\bar{m}, \tilde{m})_\Omega = -(M'(m)[\tilde{m}] \partial_t \mathcal{F}(m^k), \mathbf{z})_Q \quad \text{for all } \tilde{m} \in \mathcal{P}, \quad (6.11)$$

as shown in ALGORITHM 5.

Algorithm 5 Adjoint $\mathcal{F}'(m^k)^*[\tilde{\mathbf{b}}]$ of linearized parameter-to-wavefield map

Input: $m^k \in \mathcal{P}^{\text{adm}}$, $\mathbf{y}_{m^k} := \mathcal{F}(m^k) \in V$, $\tilde{\mathbf{b}} \in W$

Output: $\mathbf{z} \in V^*$ solving (6.10), $\bar{m} := \mathcal{F}'(m^k)^*[\tilde{\mathbf{b}}] \in \mathcal{P}$

1: Find $\mathbf{z} \in V^*$ with $L_{m^k}^* \mathbf{z} = \tilde{\mathbf{b}}$

2: Find $\bar{m} \in \mathcal{P}$ with $(\bar{m}, \tilde{m})_\Omega = -(M'(m^k)[\tilde{m}] \partial_t \mathbf{y}_{m^k}, \mathbf{z})_Q$ for all $\tilde{m} \in \mathcal{P}$

Solving the problem in (6.10) is referred to as solving the adjoint problem for the material parameter $m^k \in \mathcal{P}^{\text{adm}}$ with right-hand side $\tilde{\mathbf{b}} \in W$. Since equation (6.10) can be interpreted as solving the wave equation backward in time, it is also called the backward wave equation or back propagation problem, see SECTION 3.2.

Example 6.11 (Linear acoustics). *Continuing EXAMPLE 6.10, we obtain for step 2 in ALGORITHM 5 for $m^k = (\kappa^k, \rho^k) \in \mathcal{P}^{\text{adm}}$ and $\bar{m} = (\bar{\kappa}, \bar{\rho}) \in \mathcal{P}$*

$$\begin{aligned} ((\bar{\kappa}, \bar{\rho}), (\tilde{\kappa}, \tilde{\rho}))_\Omega &= - \int_0^T \int_\Omega \begin{pmatrix} -\kappa^k(x)^{-2} \tilde{\kappa}(x) \partial_t p_{m^k}(x, t) \\ \tilde{\rho}(x) \partial_t \mathbf{v}_{m^k}(x, t) \end{pmatrix} \cdot \begin{pmatrix} \mathbf{z}_p(x, t) \\ \mathbf{z}_v(x, t) \end{pmatrix} dx dt \\ &= \int_\Omega \begin{pmatrix} \tilde{\kappa}(x) \\ \tilde{\rho}(x) \end{pmatrix} \cdot \int_0^T \begin{pmatrix} \kappa^k(x)^{-2} \partial_t p_{m^k}(x, t) \mathbf{z}_p(x, t) \\ -\partial_t \mathbf{v}_{m^k}(x, t) \cdot \mathbf{z}_v(x, t) \end{pmatrix} dt dx \end{aligned}$$

for all $(\tilde{\kappa}, \tilde{\rho}) \in \mathcal{P}$. This implies

$$(\bar{\kappa}(x), \bar{\rho}(x)) = \int_0^T \begin{pmatrix} \kappa^k(x)^{-2} \partial_t p_{m^k}(x, t) \mathbf{z}_p(x, t) \\ -\partial_t \mathbf{v}_{m^k}(x, t) \cdot \mathbf{z}_v(x, t) \end{pmatrix} dt, \quad x \in \Omega,$$

see also [43, Thm. 3.8].

Remark 6.12. *In a space-time variational setting, the adjoint problem is easily available using integration by parts. This is an appealing feature from an implementation point of view, since with minimal modifications, the subroutines to handle the forward problem can be reused to solve the adjoint problem. Moreover, using a space-time distributed data structure for the forward and backward wavefields, step 2 in ALGORITHM 5 can be handled in parallel without extra communication overhead.*

6.3.3 The CG-REGINN algorithm

As an example of a regularized inexact NEWTON method, we consider the Conjugate Gradient REGularized INexact NEWTON (CG-REGINN) algorithm, see [57].

CG-REGINN is a special case of ALGORITHM 3 where the normal equation (6.7) is solved by a linear Conjugate Gradients (CG) algorithm, see ALGORITHM 6. It exploits that CG acts as a regularization scheme when stopped preliminary by a discrepancy principle, see [40], [58].

Algorithm 6 Conjugate Gradient algorithm for (6.7)

- 1: $l \leftarrow 0, \beta \leftarrow 0, \tilde{r}^0 \leftarrow r^k \in \mathcal{S}$
 - 2: $\tilde{p}^0, \tilde{h}^0 \leftarrow 0 \in \mathcal{P}$
 - 3: **while** not converged **do**
 - 4: $l \leftarrow l + 1$
 - 5: $d \leftarrow \Phi'(m^k)^*[\tilde{r}^{l-1}] \in \mathcal{P}$
 - 6: $p^l \leftarrow d + \beta \|d\|_{\Omega}^2 p_{l-1}$
 - 7: $q \leftarrow \Phi'(m^k)[p^l] \in \mathcal{S}$
 - 8: $\alpha \leftarrow \|d\|_{\Omega}^2 / \|q\|_{\mathcal{S}}^2$
 - 9: $\tilde{h}^l \leftarrow \tilde{h}^{l-1} + \alpha p^l$
 - 10: $\tilde{r}^l \leftarrow \tilde{r}^{l-1} - \alpha q$
 - 11: $\beta \leftarrow 1 / \|d\|_{\Omega}^2$
-

Most steps in ALGORITHM 6 involve basic linear algebra operations only. The main effort is concentrated in the highlighted steps 5 and 7, where wave equations have to be solved.

To obtain regularizing properties, the stopping criterion for the CG iteration is crucial. We use a heuristic proposed in [69] based on [57] that relies on an estimate for the local ill-posedness of the current iterate $m^k \in \mathcal{P}^{\text{adm}}$ that is obtained by comparing the iteration count of the CG iteration from previous non-linear iterations.

More precisely, for $k \in \mathbb{N}$ we denote the number of CG-iterations in the k -th iteration of NEWTON's algorithm until the inner iteration is stopped by $n_k \in \mathbb{N}$. Then, we choose $\vartheta_k \in (0, 1)$ according to the estimate of the local ill-posedness and stop the CG-iteration as soon as we have

$$\|\Phi'(m^k)[\tilde{h}^l] - r^k\|_{\mathcal{S}} \leq \vartheta_k \|r^k\|_{\mathcal{S}} \quad \text{or} \quad l_k \geq l_{k,\text{max}}. \quad (6.12)$$

To this end, we select $\gamma \in (0, 1)$, $\vartheta_0 \in (0, 1)$ and

$$\begin{aligned} \vartheta_k &= \begin{cases} 1, & n = 0, \\ \|r^1\|_{\mathcal{S}} / \|r^0\|_{\mathcal{S}}, & n = 1, \\ 1 - \frac{n_{k-2}}{n_{k-1}}(1 - \vartheta_{k-1}), & \text{ill-posedness became worse } (n_{k-1} \geq n_{k-2}), \\ \gamma\vartheta_{k-1}, & \text{ill-posedness became better,} \end{cases} \\ n_{k,\max} &= \begin{cases} 1, & k = 0, \\ 2, & k = 1, \\ n_{k-1} + n_{k-2}, & k \geq 2. \end{cases} \end{aligned} \tag{6.13}$$

A numerical example for this heuristic is presented in SECTION 6.5.

In the next section, we relate the CG-REGINN algorithm to another approach motivated from optimization theory.

Remark 6.13. *For the outer NEWTON loop, a discrepancy principle can be chosen as a stopping criterion, see e.g. [40], [58] or [57]. In this work, we focus on the inner iteration and leave systematic stopping of the outer loop as a future challenge.*

6.4 Minimization approach

Another approach to consider the FWI problem replaces (6.5) by an optimization problem as in (6.14).

$$\begin{cases} \text{Given } s_{\text{obs}} \in \mathcal{S}, \\ \text{find } m \in \mathcal{P} \text{ with } F(\Phi(m), m) \longrightarrow \min! \end{cases} \tag{6.14}$$

Here, the map $F: \mathcal{S} \times \mathcal{P} \longrightarrow \mathbb{R}$ is a measure of the misfit between the observed data s_{obs} and the data $\Phi(m)$ corresponding to the material m .

The direct dependence of F on the material m can encounter a-priori information on the material. Typical examples have the following structure

$$F(s, m) = f(s) + p(m), \quad s \in \mathcal{S}, m \in \mathcal{P},$$

with a penalty term $p(m)$ and straight-forward choices are

$$f(s) = \frac{1}{2} \|s - s_{\text{obs}}\|_{\mathcal{S}}^2 = \frac{1}{2} \sum_{r \in \mathcal{M}} |s(r) - s_{\text{obs}}(r)|_2^2, \tag{6.15}$$

$$p(m) = \frac{\gamma}{2} \|m\|_*^2, \tag{6.16}$$

where $\|\cdot\|_*: \mathcal{P} \rightarrow \mathbb{R}$ denotes a norm or a semi-norm on the material parameters and $\gamma > 0$ is a regularization parameter.² However, we set $p \equiv 0$ for simplicity in our considerations.

Using the function F and the parameter-to-seismogram operator Φ , we define the parameter-to-misfit function

$$J: \mathcal{P}^{\text{adm}} \rightarrow \mathbb{R}, \quad m \mapsto F(\Phi(m), m).$$

Then, problem (6.14) is equivalent to finding a minimizer of J .

In the simplest case, this minimization problem can be tackled using a *gradient method*.

6.4.1 A gradient descent method

The RIESZ representative $\nabla J(m^k) \in \mathcal{P}$ of the objective function's derivative $J'(m^k)$ yields a descent direction, i.e. there is $\alpha > 0$: $J(m^k - \alpha \nabla J(m^k)) < J(m^k)$. This motivates the algorithm shown in ALGORITHM 7.

Algorithm 7 A gradient descent method

- 1: Choose $m^0 \in \mathcal{P}^{\text{adm}}$, $k \leftarrow 0$
 - 2: **while** not converged **do**
 - 3: Find $\nabla J(m^k) \in \mathcal{P}$ with $(\nabla J(m^k), \tilde{m})_{\mathcal{P}} = \langle J'(m^k), \tilde{m} \rangle$ for all $\tilde{m} \in \mathcal{P}$
 - 4: Find step length $\alpha > 0$ such that $J(m^k - \alpha \nabla J(m^k)) < J(m^k)$
 - 5: $m^{k+1} \leftarrow m^k - \alpha \nabla J(m^k)$
 - 6: $k \leftarrow k + 1$
-

Since the problem in step 3 is a variational problem with the derivative $J'(m^k)$ as a right-hand side, ALGORITHM 7 relies on an efficient evaluation of the expression $\langle J'(m^k), \tilde{m} \rangle$ for all $\tilde{m} \in \mathcal{P}$. By the definition of J and F , we can rewrite the equation in step 3 as follows

$$\begin{aligned} (\nabla J(m^k), \tilde{m})_{\mathcal{P}} &= \langle J'(m^k), \tilde{m} \rangle \\ &= \langle f'(\Phi(m^k)), \Phi'(m^k)[\tilde{m}] \rangle \quad \tilde{m} \in \mathcal{P}. \end{aligned} \tag{6.17}$$

In order to assemble the linear system that has to be solved in (6.17), the expression $\Phi'(m^k)[\tilde{m}]$ has to be evaluated for each test material $\tilde{m} \in \mathcal{P}$. According to SECTION 6.3.1, this requires to solve $\dim \mathcal{P}$ forward problems to assemble the right-hand side in each step of ALGORITHM 7 which is practically infeasible.

²This is often referred to as TIKHONOV regularization.

However, assuming that the misfit-function's derivative is computationally accessible, we can find the gradient $\nabla f(\Phi(m^k)) \in \mathcal{S}$ of the misfit-function by

$$(\nabla f(\Phi(m^k)), \tilde{s})_{\mathcal{S}} = \langle f'(\Phi(m^k)), \tilde{s} \rangle, \quad \tilde{s} \in \mathcal{S}.$$

Then, continuing the calculation in (6.17), we obtain

$$\begin{aligned} (\nabla J(m^k), \tilde{m})_{\mathcal{P}} &= \left(\nabla f(\Phi(m^k)), \Phi'(m^k)[\tilde{m}] \right)_{\mathcal{S}} \\ &= \left(\Phi'(m^k)^* [\nabla f(\Phi(m^k))], \tilde{m} \right)_{\Omega}, \quad \tilde{m} \in \mathcal{P}, \end{aligned} \quad (6.18)$$

which means

$$\nabla J(m^k) = \Phi'(m^k)^* [\nabla f(\Phi(m^k))]. \quad (6.19)$$

Using the results in SECTION 6.3.2, after solving a single backward wave equation for the right-hand side $v = \Psi^* [\nabla f(\Phi(m^k))] \in H$ according to ALGORITHM 5 yields the expression on the right-hand side of (6.19).

Remark 6.14. *This procedure to efficiently evaluate the derivative of the misfit-function is usually referred to as the adjoint-state method. Here, the solution of the backward equation for the right-hand side $v = \Psi^* [\nabla f(\Phi(m^k))]$ is called the adjoint state, see e.g. [53] for an overview and further variants.*

Remark 6.15. *Note that in the infinite dimensional setting, the solution $\nabla J(m^k)$ in (6.18) fulfills $\nabla J(m^k) \in L_2(\Omega)$. Since $L_2(\Omega) \not\subset L_\infty(\Omega) = \mathcal{P}$, we cannot guarantee that the iterates obtained by $m^{k+1} = m^k - \alpha \nabla J(m^k)$ belong to the space \mathcal{P} . Furthermore, even in case that $m^{k+1} \in \mathcal{P}$, it might violate the positivity constraints in \mathcal{P}^{adm} , i.e. $m^{k+1} \notin \mathcal{P}^{\text{adm}}$. In our implementation, we assume that $m^{k+1} \in \mathcal{P}^{\text{adm}}$ on discrete level leaving a systematic treatment of these problems as a future challenge.*

Example 6.16 (L_2 misfit). *We consider the misfit function given in (6.15). In this case, we have the following derivative and gradient for all $s, \tilde{s} \in \mathcal{S}$*

$$\langle f'(s), \tilde{s} \rangle = (s - s_{\text{obs}}, \tilde{s})_{\mathcal{S}}, \quad \nabla f(s) = s - s_{\text{obs}}.$$

In this case, we obtain the gradient by

$$\nabla J(m^k) = \Phi'(m^k)^* [\Phi(m^k) - s_{\text{obs}}] = \mathcal{F}'(m^k)^* [\Psi^* [\Phi(m^k) - s_{\text{obs}}]]. \quad (6.20)$$

Note that $r^k := \Phi(m^k) - s_{\text{obs}} \in \mathcal{S}$ is the current non-linear residual in the seismograms space and by (6.4), $\Psi^ r^k \in W$ is a right-hand side consisting of point sources located at the receiver positions.*

Considering ALGORITHM 5 to evaluate $\mathcal{F}'(m^k)^$, the gradient $\nabla J(m^k)$ is obtained by propagating the residual backwards in time from the receiver positions.*

Remark 6.17. *In case that CG-REGINN is stopped after the first inner iteration, cf. ALGORITHM 3 and ALGORITHM 6, we obtain as a result of the conjugate gradient algorithm for the NEWTON update \tilde{h}^0*

$$\tilde{h}^0 = \alpha p^0 = \alpha d = \alpha \cdot \left(\Phi'(m^k)^* [s_{\text{obs}} - \Phi'(m^k)] \right), \quad \alpha = \frac{\|p^0\|_{\Omega}^2}{\|\Phi'(m^k)[p^0]\|_{\mathcal{S}}^2}.$$

Thus, CG-REGINN stopped after a single inner iteration yields a gradient method.

Finding a suitable step length

Given the descent direction $\nabla J(m^k)$, a proper scaling is required in order to obtain a convergent algorithm. A straight-forward approach is the *line-search* that try to find the minimizer $\alpha > 0$ in

$$\min_{\tilde{\alpha} > 0} J(m^k - \tilde{\alpha} \nabla J(m^k)).$$

Solving this minimization problem is practically infeasible in case of FWI and thus, approximation strategies are employed. See [65] for references.

Convergence properties

Gradient based optimization methods are known to suffer from slow convergence rates. Even in simple model examples, the descent directions can oscillate such that the algorithms yields a “zig-zag” path through the search space.

The NEWTON-type methods presented in the following are known to have more appealing convergence properties.

6.4.2 NEWTON-type methods for minimization

In this section, we consider NEWTON-type methods to solve (6.14) numerically. For the reader’s convenience, we provide a short derivation of NEWTON’s method for optimization problems. Using a second order Taylor-expansion of J , we obtain for the (unknown) exact update $h^k := m^{\text{sol}} - m^k$

$$J(m^{\text{sol}}) = J(m^k) + \langle J'(m^k), h^k \rangle + \frac{1}{2} J''(m^k)[h^k, h^k] + \mathcal{O}(\|h^k\|^3).$$

This yields an approximation of $J(m^{\text{sol}})$ by a quadratic function in h

$$h \mapsto J(m^k) + \langle J'(m^k), h \rangle + \frac{1}{2} J''(m^k)[h, h], \quad h \in \mathcal{P}.$$

In case that $h^k \in \mathcal{P}$ is a minimizer of this quadratic approximation, the corresponding derivative with respect to h is vanishing at h^k , i. e.

$$J''(m^k)[h^k, \tilde{h}] = -\langle J'(m^k), \tilde{h} \rangle \quad \text{for all } \tilde{h} \in \mathcal{P}. \quad (6.21)$$

This gives rise to NEWTON's method for optimization problems, see ALGORITHM 8.

Algorithm 8 NEWTON's method for optimization (abstract form)

- 1: Choose $m^0 \in \mathcal{P}^{\text{adm}}$, $k \leftarrow 0$
 - 2: **while** not converged **do**
 - 3: Find $h^k \in \mathcal{P}$ with $J''(m^k)[h^k, \tilde{h}] = -\langle J'(m^k), \tilde{h} \rangle$ for all $\tilde{h} \in \mathcal{P}$
 - 4: $m^{k+1} \leftarrow m^k + h^k$
 - 5: $k \leftarrow k + 1$
-

Comparing step 3 in ALGORITHM 7 and step 3 in ALGORITHM 8, we see that both linear problems have – up to the sign – the same right-hand sides. Thus, we can use the same technique as in the previous section to efficiently handle $J'(m^k)$.

Unlike the RIESZ map in ALGORITHM 7, the second derivative $J''(m^k)$ cannot be assembled efficiently as an operator in $\mathcal{B}(\mathcal{P} \times \mathcal{P}, \mathbb{R})$. Thus, practical algorithms rely on an approximation of $J''(m^k)$. Typical approaches use the GAUSS-NEWTON or variants of L-BFGS approximation, see [12], and apply an iterative linear solver for the resulting system.

Remark 6.18. *It has been evaluated in [29] that incorporating the second order derivative for the inversion can improve the reconstruction quality.*

Thus, in another approach, $J''(m^k)[\tilde{m}^1, \tilde{m}^2]$ is evaluated for given $\tilde{m}^1, \tilde{m}^2 \in \mathcal{P}$ exploiting second-order adjoint states. For completeness, we provide a derivation of this method in APPENDIX B using our space-time variational setting.

However, stopping strategies originating from optimization theory for the linear iteration have been employed, see e.g. [49]. It is a promising future challenge to implement and evaluate this method combined with regularizing stopping criteria.

Due to the relation to the CG-REGINN algorithm, in the following, we restrict ourselves to the GAUSS-NEWTON variant.

6.4.3 GAUSS-NEWTON approximation of J''

From (6.17) we obtain the following expression for the second derivative of J

$$\begin{aligned}
 J''(m^k)[\tilde{m}, \hat{m}] &= f''(\Phi(m^k))[\Phi'(m^k)[\tilde{m}], \Phi'(m^k)[\hat{m}]] \\
 &\quad + \langle f'(\Phi(m^k)), \Phi''(m^k)[\tilde{m}, \hat{m}] \rangle.
 \end{aligned}$$

We call the first addend the GAUSS-NEWTON approximation of $J''(m^k)[\tilde{m}, \hat{m}]$.

Remark 6.19. Let f be the simple L_2 functional from EXAMPLE 6.16. Then, using $f''(s)[\tilde{s}, \hat{s}] = (\tilde{s}, \hat{s})_{\mathcal{S}}$ for $s, \tilde{s}, \hat{s} \in \mathcal{S}$, we obtain for all $\tilde{m}, \hat{m} \in \mathcal{P}$

$$J''(m^k)[\tilde{m}, \hat{m}] \approx (\Phi'(m^k)[\tilde{m}], \Phi'(m^k)[\hat{m}])_{\mathcal{S}} = (\Phi'(m^k)^* \Phi'(m^k)[\tilde{m}], \hat{m})_{\mathcal{P}}.$$

Using the definition of $\nabla J(m^k)$ and (6.20), for the GAUSS-NEWTON approximation, the equation in step 3 of ALGORITHM 8 becomes

$$\Phi'(m^k)^* \Phi'(m^k)[h^k] = \Phi'(m^k)^* [s_{\text{obs}} - \Phi(m^k)].$$

Therefore, the normal equation (6.7) that is solved in the regularized root finding version of NEWTON's method can be interpreted as a result of the GAUSS-NEWTON approximation for the linearized system in NEWTON's method for optimization.

6.5 A numerical example for FWI

In this section, we provide a numerical experiment to demonstrate that CG-REGINN in combination with the space-time DPG method as a forward solver yields a working algorithm.

6.5.1 The experimental setup

We set $\Omega = (-\frac{9}{8}, \frac{3}{2}) \times (0, \frac{5}{4})$ and $T = 1.94$ yielding $Q = \Omega \times (0, T)$ as the space-time cylinder. As boundary conditions, we use $p = 0$ on the left, right and bottom boundaries of Ω . At the top boundary, we set $\mathbf{v} \cdot \mathbf{n}_{\Omega} = 0$ and in the beginning, the system is at rest, i.e. $p(0) = 0$, $\mathbf{v}(0) = \mathbf{0}$.

We add external energy to the system using a point-source at space-time position $(x_c, t_c) = ((-\frac{1}{8}, 0), 0.05)$ in the pressure component given by

$$b(x, t) = \begin{cases} a \cdot \exp\left(-6 \cdot \frac{|(x, t) - (x_c, t_c)|_2^2}{\delta^2}\right), & |(x, t) - (x_c, t_c)|_2 < \delta, \\ 0, & \text{else,} \end{cases}$$

where $\delta = 0.0625$ and $a = 10\,000$. The material distribution is homogeneous in Ω except for a horizontal layer at $H := (-\frac{9}{8}, \frac{3}{2}) \times (\frac{5}{8}, \frac{3}{4})$, cf. FIGURE 6.1. More precisely, we have $\rho(x) = 1$, $x \in \Omega$, and

$$\kappa(x) = v_{\text{back}}, \quad x \in \Omega \setminus H, \quad \kappa(x) = v_{\text{back}} + 0.55 \cdot \sin\left(\frac{x_2 - 5/8}{3/4 - 5/8}\right), \quad x \in H,$$

with $v_{\text{back}} = 1$.

The pressure component of the numerical solution for this setting is visualized in FIGURE 6.2 and FIGURE 6.3.

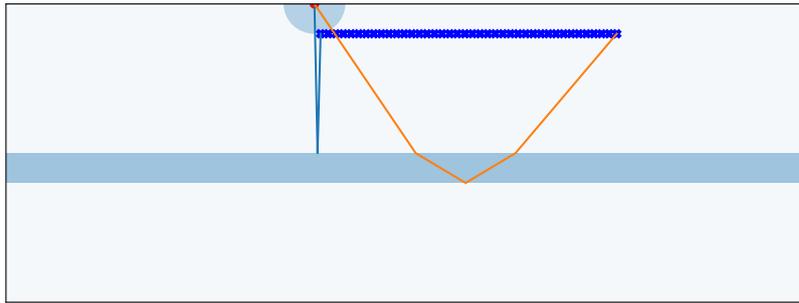


Figure 6.1: Experimental setup for the reconstruction of an inhomogeneous layer. On the top, the source location and receiver positions are depicted. The two lines illustrate parts of the signal being reflected at the top and the bottom of the inclusion. The area in between these lines indicates that part of the inclusion generating reflections which are recorded by the receivers. This is the only part of the inclusion, the seismograms contain information about.

Remark 6.20. *The material distribution in this example is inspired by the horizontal reflector example of the Python Toolbox for Seismic Imaging, PySIT, see [34].*

Remark 6.21. *Since we do not utilize absorbing boundary conditions or absorbing layers, we choose the size of Ω in such a way that boundary reflections cannot reach the receivers during the simulation time $(0, T)$ due to the finite speed of wave propagation. This is a technical simplification that we would like to drop in the future.*

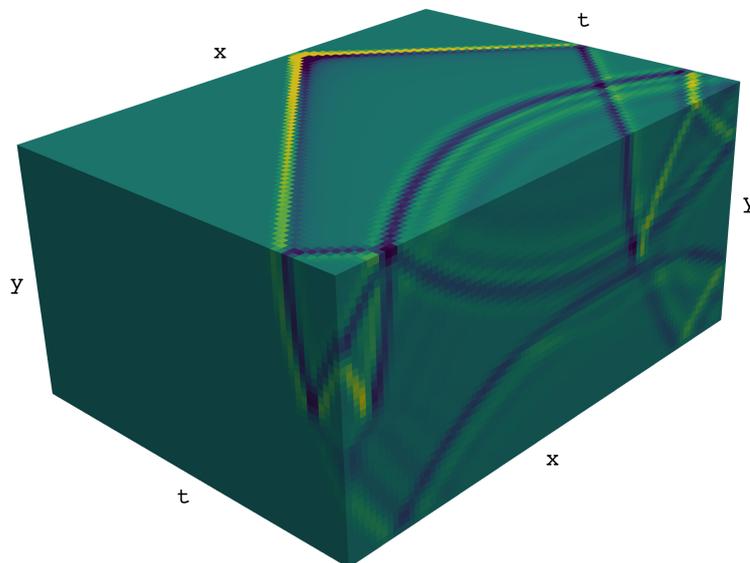


Figure 6.2: Space-time plot showing a numerical approximation of the solution's pressure component. Here, the time axis is pointing towards the reader. The simulation was performed using the DPG method setup as described in (6.24) and (6.25) on a mesh with $84 \cdot 40 \cdot 64 = 215\,040$ space-time cuboids yielding a linear system with 13 781 088 degrees of freedom in \hat{V}_h .

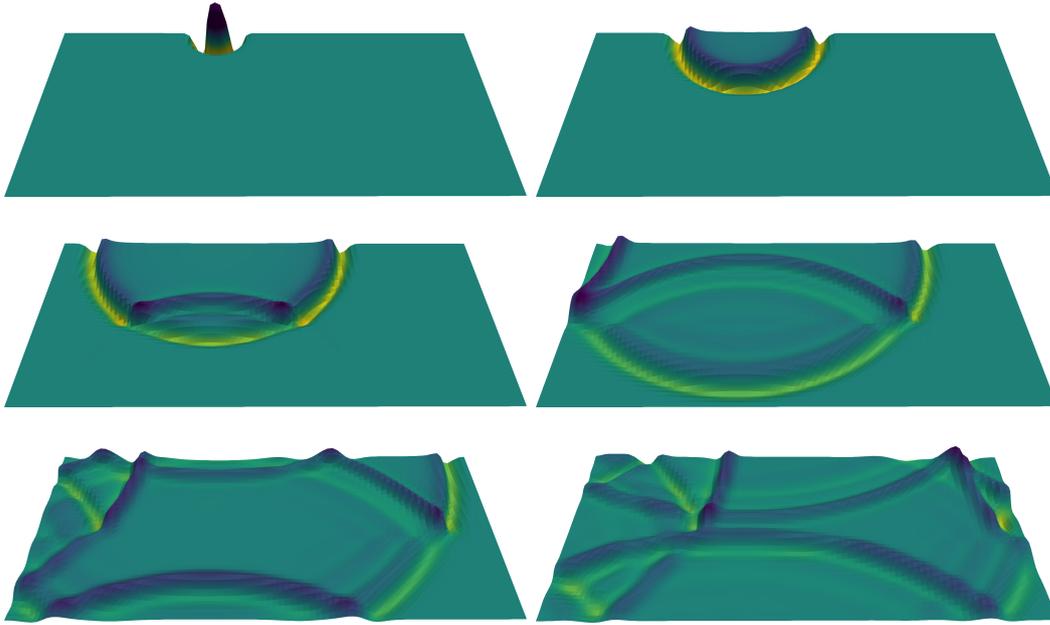


Figure 6.3: Pressure component at times $t = 0.15, 0.5, 0.85, 1.2, 1.55, 1.9$. These images were obtained by slicing the space-time cylinder from FIGURE 6.2. Here, reflections at the homogeneity layer as well as at the boundary of Ω can be observed.

The measurement setup

For the measurements, we select $N = 40$ equidistant spatial receivers, see FIGURE 6.1.

$$\mathcal{M}_\Omega := \left\{ -0.1 + n \cdot \frac{1.1 - (-0.1)}{N - 1} : n = 0, \dots, N - 1 \right\} \quad (6.22)$$

and $M = 92$ measurement times

$$\mathcal{M}_T := \left\{ 0.05 + m \cdot \frac{1.88 - 0.05}{M - 1} : m = 0, \dots, M - 1 \right\} \quad (6.23)$$

yielding the space-time receiver positions $\mathcal{M} := \mathcal{M}_\Omega \times \mathcal{M}_T$ and the total number of space-time receivers $|\mathcal{M}| = 40 \cdot 92 = 3680$. As a result, the whole measurement vector consists of $3 \cdot |\mathcal{M}| = 11\,040$ entries since we measure in every component of (p, \mathbf{v}) . To approximate point evaluations of the state $(p(x, t), \mathbf{v}(x, t))$, $(x, t) \in Q$, we use

$$\varphi_r(x, t) = \begin{cases} a \cdot \exp\left(-7 \cdot \frac{|(x, t) - r|_2^2}{\delta^2}\right), & |(x, t) - r|_2 < \delta, \\ 0, & \text{else,} \end{cases}$$

as measurement kernels, see (6.3). We choose $\delta = 0.05$ and select $a > 0$ such that $\|\varphi_r\|_{L_1(Q)} = 1$. In FIGURE 6.4 the observed seismogram is depicted.

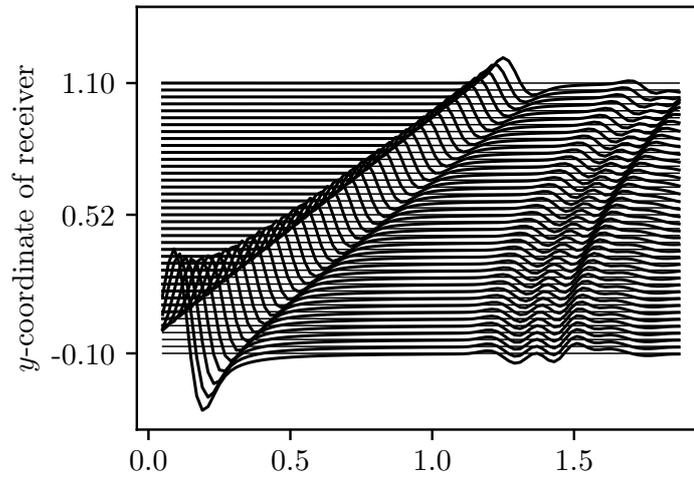


Figure 6.4: Pressure component of a seismogram obtained by averaged point measurements at the space-time receivers given in (6.22) and (6.23). The signal clearly shows the directly traveled wave followed by the reflections at the material inhomogeneity.

An example for the adjoint state

Inside the CG-loop of the algorithm, linearized forward and backward wave equations are solved one after another. In the backwards wave equation, the current residual seismogram acts as an array of sources located at the receiver positions, see EXAMPLE 6.16. Thus, the residual is propagated backwards in time, see FIGURE 6.5 and FIGURE 6.6 for an example.

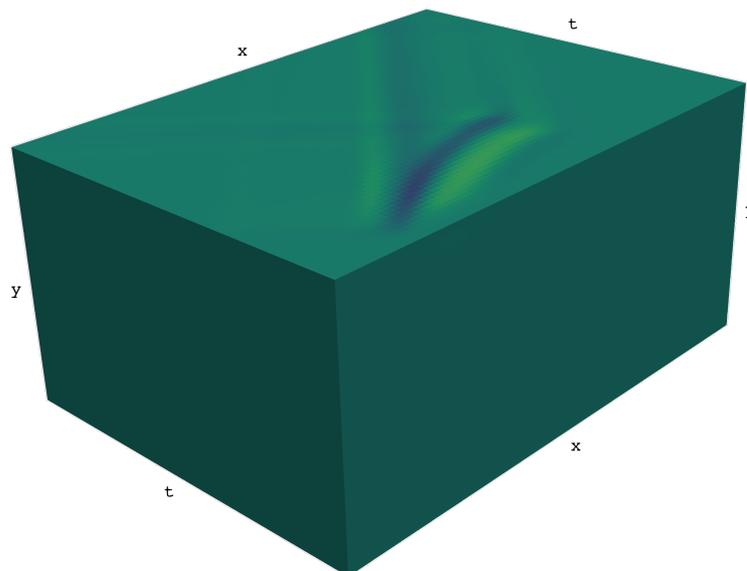


Figure 6.5: Space-time plot of an adjoint state's pressure component. Note that the time direction is pointing towards the reader. Also see the more detailed explanation given in FIGURE 6.6.



Figure 6.6: Numerical approximation of an adjoint state's pressure component plotted for $t = 0.15, 0.5, 0.85, 1.2, 1.55, 1.9$. It was obtained using the DPG method setup as in (6.24) and (6.25). The backpropagated residual resulted from a forward simulation for an homogeneous κ distribution.

6.5.2 Application of CG-REGINN

We apply the CG-REGINN algorithm described in SECTION 6.3.3 to the example described above. In this demonstration-of-concept example, we commit the inverse crime by using the same forward solver to generate the data that is also used during the inversion. Moreover, we did not add noise to the data.

Configuration of the DPG forward solver

To solve the forward and backward problems arising in steps 5 and 7 of ALGORITHM 6, we use the space-time Discontinuous PETROV-GALERKIN method as described in SECTION 4.3. For the face degrees of freedom, we apply the simplified DPG method as described in SECTION 4.6. Using a second-order variant D2 of the DPG method, see TABLE 5.3, we choose for the ansatz and test spaces

$$W_{R,h} = \mathbb{Q}_1(R) \times \mathbb{Q}_1(R)^2, \quad Z_{R,h} = \mathbb{Q}_4(R) \times \mathbb{Q}_4(R)^2 \quad (6.24)$$

inside the space-time cells $R = K \times (a, b)$, and on the skeleton ∂Q_h we use

$$\begin{aligned} \tilde{V}_{K \times \{t\}, h} &= \mathbb{Q}_2(K) \times \mathbb{Q}_2(K)^2 \text{ for faces in time, and} \\ \tilde{V}_{F \times (a,b), h} &= \mathbb{Q}_2(F \times (a, b)) \times \mathbb{Q}_2(F \times (a, b)) \mathbf{n}_F \text{ for faces in space.} \end{aligned} \quad (6.25)$$

According to SECTION 4.7, this yields a second order method in space-time.

Discretization of the material parameters

To discretize the material parameters, we use cell-wise constant functions as an approximation for $\kappa(x)$ and $\rho(x)$, $x \in \Omega$. Since we use $84 \cdot 40$ cells in space, this results in 3360 degrees of freedom for each of κ and ρ .

In order to simplify the implementation, we replace the map M from (6.1) by another map $\tilde{M}: \mathcal{P} \rightarrow \mathcal{M}$ given by

$$\tilde{M}(\rho(x), \tilde{\kappa}(x)) = \begin{pmatrix} \tilde{\kappa}(x) & 0 \\ 0 & \rho(x)I_d \end{pmatrix}, \quad x \in \Omega.$$

Here \tilde{M} is obtained by replacing $\kappa(x)^{-1}$ with $\tilde{\kappa}(x)$, $x \in \Omega$. Since the map

$$\mathcal{P}^{\text{adm}} \rightarrow \mathcal{P}^{\text{adm}}, \quad f \mapsto f^{-1}$$

is one-to-one, we can reconstruct $\tilde{\kappa}$ in first step and in a second step obtain κ itself.

Since \tilde{M} is a linear map, this simplification eliminates inner derivatives of κ in the implementation of the inversion scheme.

Protocol of a CG-REGINN iteration

For the application of CG-REGINN, we use the stopping criterion for the inner loop as described in (6.13), The parameters have are chosen as follows

$$\gamma = 0.9, \quad \vartheta_0 = 0.999.$$

Since we do not add noise to the data, instead of using a discrepancy principle to stop the outer loop, for simplicity, we limited the NEWTON iteration count by 11. As a starting value, we use $m^0 = (\rho^0, \kappa^0) \equiv (1, 1)$ in Ω .

Remark 6.22. *During the iteration, we fix $\rho \equiv 1$ such that only $\tilde{\kappa}$ is reconstructed.*

In FIGURE 6.8, the squared non-linear residual for each iteration is shown. Moreover, the count of CG iterations inside the inner loop, n_k , is plotted as well as the values of ϑ_k chosen by the heuristic in (6.13). According to this heuristic, the local ill-posedness increases in the beginning and then oscillates.

The evolution of the seismograms during the NEWTON iteration is visualized by FIGURE 6.9. After about 10 iterations, the reconstructed seismogram is in good match with the observation.

However, the ill-posedness of the FWI problem becomes obvious when looking at the iterates for the reconstructions of κ : although the seismograms coincide with high accuracy, the reconstruction of the material is far away from the real material corresponding to the observed seismogram. Here, we can distinguish two effects:

1. The only region where the inhomogeneity is reconstructed in an acceptable way is located at the horizontal center. The reason for that is the measurement geometry. Since our measurements were taken using a single source and a small array of receivers over a short period of time, we can only expect meaningful reconstructions in that part of the domain where the signal traveled through before being recorded. This part is located at the center of the domain where the algorithm obtained something meaningful, see FIGURE 6.1.
2. There are also reconstruction artifacts resulting from the ill-posedness. Especially, close to the source position these artifacts occur in our experiments.

The only way to tackle the first effect consists in investing more data. This can be done, e.g. by using a larger receiver array or more than one source-receivers pair.

Investigating the reduction of artifacts is a challenging task where many geophysicists and engineers have worked on for a long time. As a future challenge, we would like to transfer approaches from the inverse-problems community such as weighted norms or other regularization strategies to address this problem.

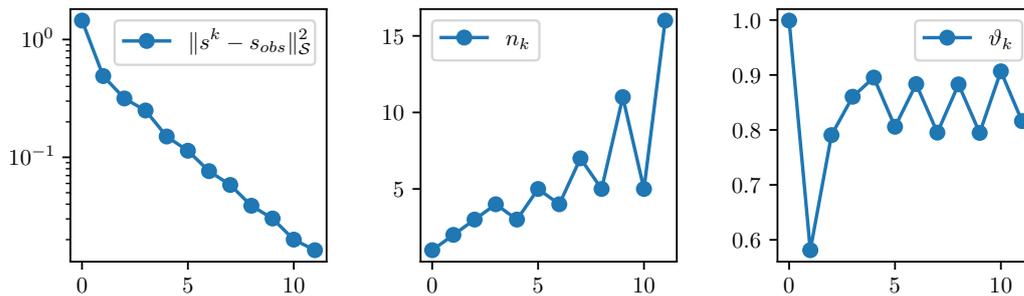


Figure 6.7: On the left, the squared non-linear residual (misfit) for each NEWTON-step k is shown. Here, k corresponds to the x -axis in all plots. The number of inner iterations n_k is depicted at the center and the chosen value for ϑ_k is shown by the right plot.

Remark 6.23. *In every NEWTON-step, at least one wave equation has to be solved to obtain the non-linear residual. Further, each CG iteration solves a linearized forward problem and another adjoint problem. In TABLE 6.1, the amounts of solved wave equations are listed.*

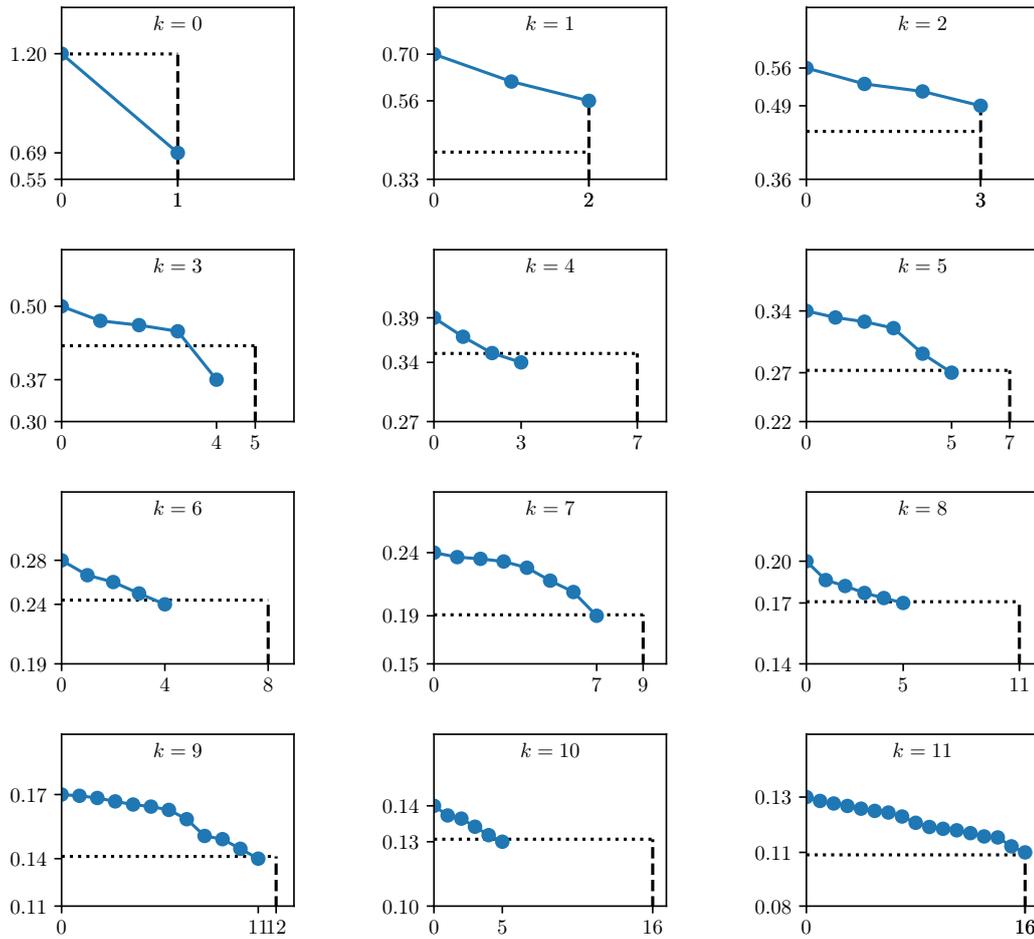


Figure 6.8: Evolution of linear residuals for each NEWTON-stop k . The vertical dashed line and the horizontal dotted line indicate the current choice of maximal linear iterations $n_{k,\max}$ and the stopping tolerance $\vartheta_k \|r^k\|_S$ for the residual according to (6.13).

k	0	1	2	3	4	5	6	7	8	9	10	11
n_k	1	2	3	4	3	5	4	7	5	11	5	16
c_k	3	5	7	9	7	11	9	15	11	23	11	33
C_k	3	8	15	24	31	42	51	66	77	100	111	144

Table 6.1: Number of wave equations solved in the k -th NEWTON step, c_k . By C_k the total number of wave equations after finishing the k -th NEWTON step is denoted.

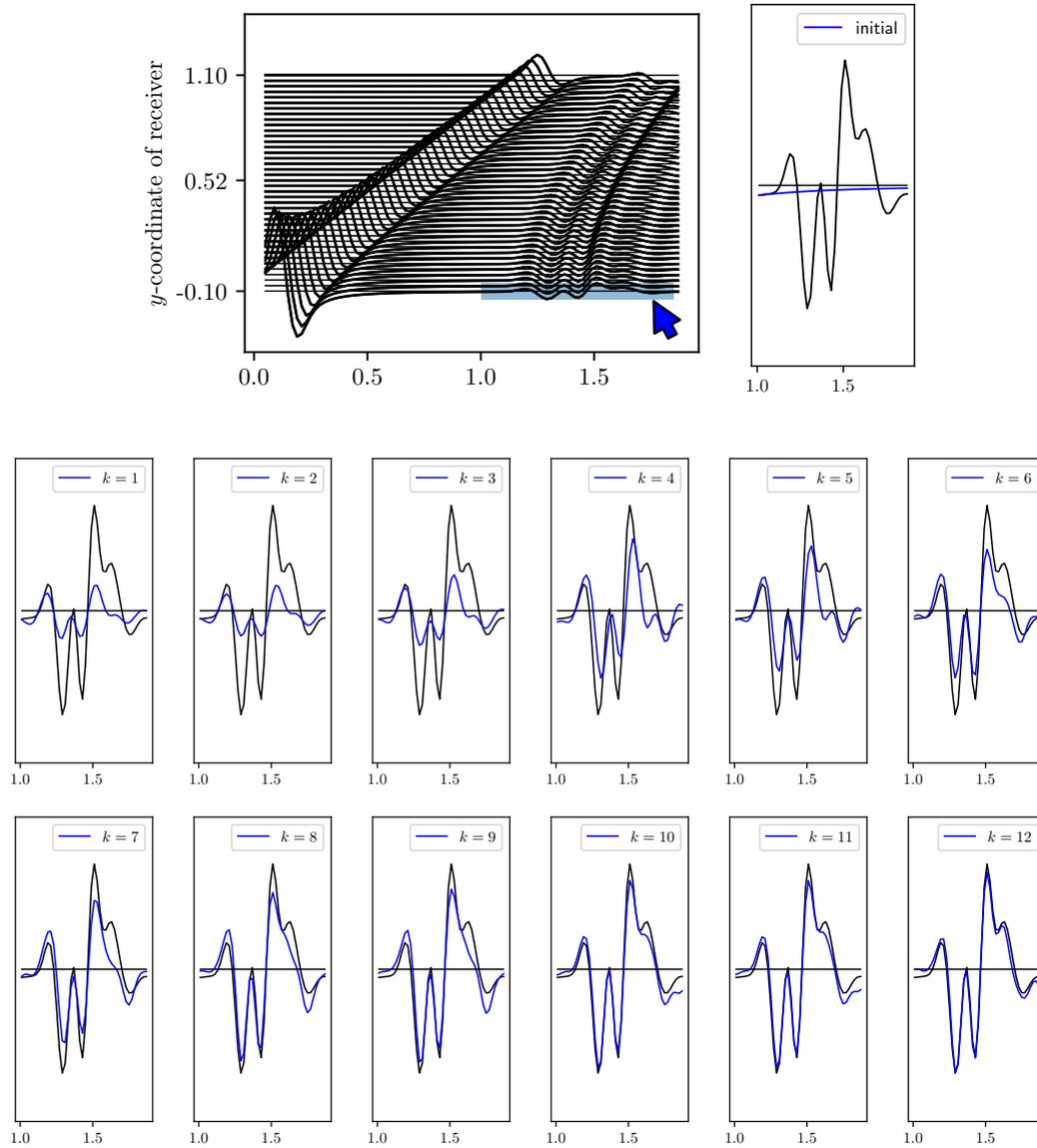


Figure 6.9: The plot at the top right shows the pressure component of the observed seismogram where the receiver at the bottom is highlighted. The remaining images show the observed signal for this receiver compared to the reconstruction in the k -th NEWTON step.

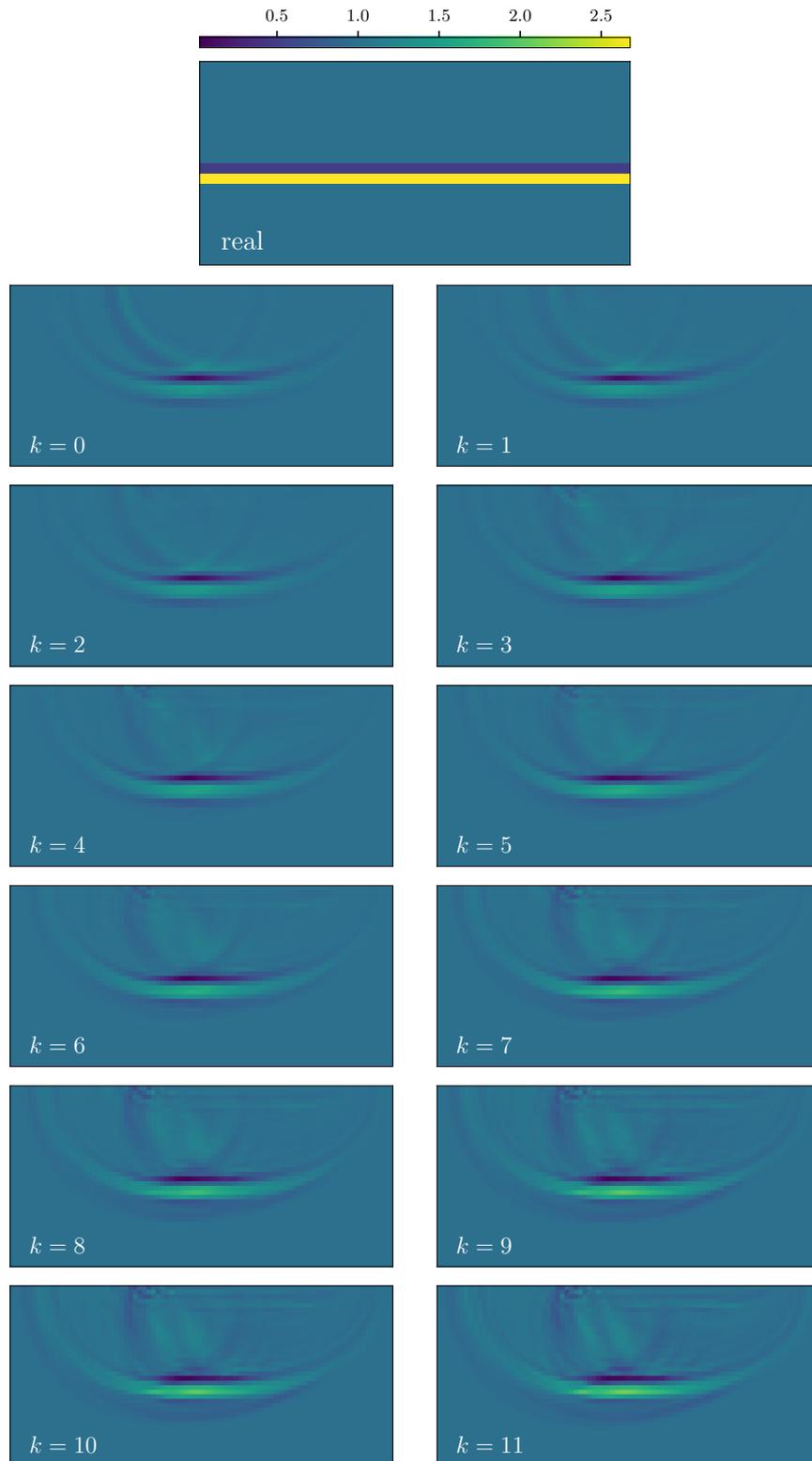


Figure 6.10: The top plot shows the real material used for data generation and the remaining images depict NEWTON iterates of $\tilde{\kappa} = \kappa^{-1}$. The inclusion is reconstructed only at the center as expected, see FIGURE 6.1.

6.5.3 A remark on the computational effort

Comparing the performance of the algorithm to Finite Difference based codes, e.g. PySIT [34], we draw the conclusion that in its current state, the space-time DPG method is by far not competitive to the established schemes.

Using the software PySIT, a comparable example can be solved within minutes on a standard desktop computer. The method using space-time DPG as a forward solver, however, takes about half an hour to solve a single wave equation using a parallel computer featuring two MD EPYC 7551 processors each of which having 32 cores. As a result, implementing CG-REGINN using more efficient discretization to solve the wave equation, remains an interesting challenge.

Possibilities for space-time adaptivity in FWI

In this demonstration-of-concept example, we use the space-time DPG method having the same ansatz and test spaces in every space-time cell. However, since in typical scenarios for FWI applications, we consider waves originating from point sources, the solution's support is contained in small fraction of the space-time cylinder due to the finite speed of wave propagation. This applies to the waves propagated forward in time, see FIGURE 6.2 and FIGURE 6.3, as well as to the solutions of the adjoint problem originating from sources at the receiver positions, see FIGURE 6.5 and FIGURE 6.6. Therefore, FWI is an application where space-time adaptivity as presented in SECTION 5.3.2 and SECTION 5.4 can be applied to reduce the computational effort.

Considering step 2 of ALGORITHM 5, we have to solve the variational problem

$$\text{Find } \bar{m} \in \mathcal{P} \quad \text{with} \quad (\bar{m}, \tilde{m})_{\Omega} = -(M'(m^k)[\tilde{m}]\partial_t \mathbf{y}_{m^k}, \mathbf{z})_Q, \quad \tilde{m} \in \mathcal{P}, \quad (6.26)$$

in order to find the update directions for the material reconstruction in all considered algorithms to address the FWI problem, see also EXAMPLE 6.11.

Since the right-hand side of (6.26) only takes into account space-time positions $(x, t) \in Q$, where both, $\mathbf{y}_{m^k}(x, t)$ and $\mathbf{z}(x, t)$, are non-zero, both \mathbf{y}_{m^k} and \mathbf{z} need to be approximated with high accuracy in the intersection of their space-time support, see FIGURE 6.11. Adaptively chosen approximation spaces as well as truncation techniques for the space-time cylinder as presented in SECTION 5.3.2 might be applied.

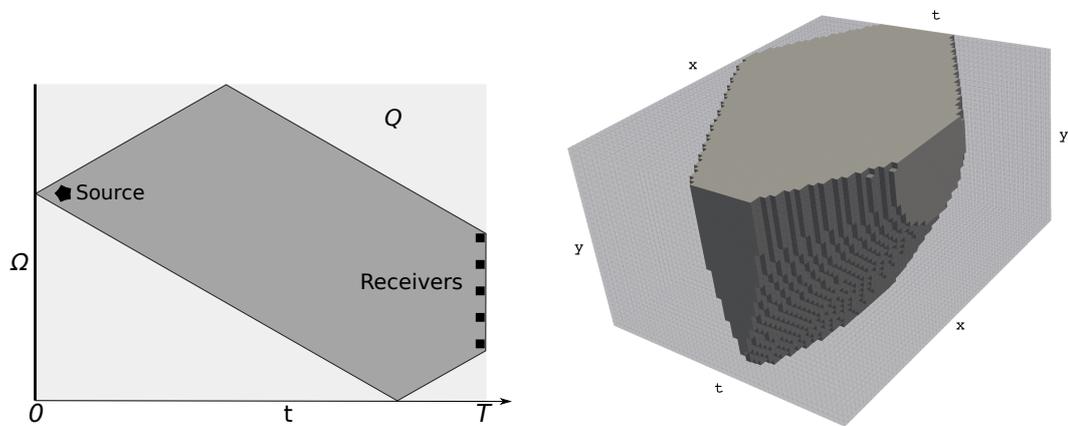


Figure 6.11: The space-time region where the both, the forward wave \mathbf{y}_{m^k} as well as the back propagated wave \mathbf{z} , are supported is highlighted in gray. The left picture is an illustration for the situation in 1D and the right picture corresponds to the solutions shown in FIGURE 6.2 and FIGURE 6.5.

Chapter 7

Summary and outlook

Classically, wave equations are considered as evolution equations where the derivative with respect to time is treated in a stronger way than the spatial differential operators. This results in an ordinary differential equation (ODE) with values in a function space, e.g. in a HILBERT space, with respect to the spatial variable. For instance, acoustic waves in a spatial domain $\Omega \subset \mathbb{R}^d$ for a given right-hand side \mathbf{b} can be considered in terms of the following ODE

$$\partial_t \mathbf{y} = A\mathbf{y} + \mathbf{b} \quad \text{in } [0, T], \quad \mathbf{y}(0) = \mathbf{0}, \quad A = \begin{pmatrix} 0 & \text{div} \\ \nabla & 0 \end{pmatrix},$$

where the solution \mathbf{y} is an element of the space $C^0(0, T; \mathcal{D}(A)) \cap C^1(0, T; L_2(\Omega))$ with $\mathcal{D}(A) \subset H^1(\Omega) \times H(\text{div}, \Omega)$. In order to analyze this ODE, space and time are treated separately and hence tools for partial differential equations are used in space and tools for ODEs are used in time. Typically, this separation carries over to the analysis of numerical schemes to approximate solutions of the equation.

By contrast, in this work, we consider the space-time operator

$$L(p, \mathbf{v}) = \begin{pmatrix} \partial_t p + \text{div } \mathbf{v} \\ \partial_t \mathbf{v} + \nabla p \end{pmatrix},$$

in $Q = (0, T) \times \Omega$ as a whole treating time and space dependence simultaneously in a variational manner. Using this approach, we constructed a space-time HILBERT space setting that allows for irregular solutions, e.g. with space-time discontinuities. In particular, we defined the HILBERT space $H(L, Q)$ and extended the operator L to $H(L, Q)$ in a space-time weak sense. Using generalized integration by parts, we established a notion of space-time traces for functions in $H(L, Q)$. By employing semi-group theory, we constructed a space $V \subset L_2(Q)$ featuring initial and boundary conditions, such that the full space-time operator $L: V \rightarrow L_2(Q)$ defines an

isomorphism. As a result, for every given right-hand side $\mathbf{b} \in L_2(Q)$, the problem of finding $\mathbf{y} \in V$ such that

$$L\mathbf{y} = \mathbf{b}$$

is well-posed in our framework.

Within this variational framework, we constructed and analyzed two classes of non-conforming discretization schemes for acoustic waves. On the one hand, we considered a weakly conforming Least-Squares method that finds the approximate solution by minimizing the residual in an enlarged approximation space allowing for space-time discontinuities. On the other hand, we present a variant of the discontinuous PETROV-GALERKIN method (DPG) for acoustic waves. The construction of both methods heavily relies on the generalized space-time traces that we introduced earlier. Expecting that an ansatz space containing functions of low regularity can lead to improved approximation quality for irregular solutions, we considered a non-conforming variant of the DPG method. This method allows for face-wise defined traces on the space-time skeleton.

For both methods, we provided a convergence analysis exploiting tools from classical Finite Element theory for space *and also time dependence*. By applying standard polynomial interpolation theory, we demonstrated how to design discretization with high order of convergence for both methods. These theoretical predictions are complemented by extensive numerical experiments showing that the high convergence rates are attained in practice. We compared schemes up to 8th order in case of the weakly conforming Least-Squares and up to 5th order for the DPG method solving various benchmark problems in one and two spatial dimensions. In particular, we considered a low-regularity example to explore the method's properties beyond the theory. Interestingly, we observed increased convergence rates for cell-wise average values in case of the DPG method.

Moreover, we demonstrated the flexibility of space-time methods with respect to adaptivity by varying the polynomial degree in the space-time cylinder. Considering an example with sparse space-time support, we used DPG's built-in error estimator to selectively increase the polynomial degree resulting in a severe reduction of degrees of freedom for the resulting linear system.

While considering the problem of Full Waveform Inversion (FWI), we focused on the derivation of NEWTON-type algorithms to tackle this inverse problem numerically. Here, we made extensive use of the space-time $L_2(Q)$ adjoint L^* that is easily accessible within our variational space-time framework. We implemented a regularized inexact NEWTON method, CG-REGINN, and provided a numerical example for

a benchmark problem.

Future challenges

The following paragraphs provide promising and open questions for the future.

While yielding promising numerical results, our considerations of the weakly conforming Least-Squares method are not fully satisfactory, since the construction of stable pairings of ansatz and test spaces in 2D remains an open question.

Currently, we use black-box solving techniques to handle the large linear system containing all space-time degrees of freedom for both discretization methods. More precisely, we use a restarted GMRES linear solver preconditioned by a subdomain-wise symmetric GAUSS-SEIDEL scheme. As a result, the number of needed iterations to obtain an acceptable accuracy increases with the refinement of the mesh. Even worse, since the preconditioner operates locally on every parallel subdomain, using more parallel processes decreases its efficiency. In order to render these space-time discretizations competitive to classical schemes, a preconditioner is necessary that scales well with respect to the mesh size as well as with the number of processes used. Promising candidates in this respect are multigrid-algorithms, see e.g. [23] for a space-time discontinuous GALERKIN method.

By numerical experiments which we do not provide in this work, we obtained numerical indication that using scaled L_2 in time can improve the performance of the iterative solver for minimal residual methods. A systematic investigation of this phenomenon remains as a future challenge.

To keep our first implementation simple, we restricted ourselves to rectangular meshes. We believe that the discussed methods also work for more general meshes, e.g. for triangular cells in space. Evaluating the performance of both methods with respect to adaptively refined meshes is a promising challenge. Furthermore, we would like to implement more efficient quadrature schemes in order to reduce the computational costs to assemble the local cell matrices.

In the computational experiments, we observed improved convergence rates when considering projections of the exact and numerical solution to cell-wise constant functions. Obtaining a better understanding of this effect may yield to other quantities converging with increased rates or even the construction of schemes that provide high-order convergence for specific quantities. This is a promising challenge since it may be applicable for low-regularity solutions where classical convergence estimates fail.

For the algorithmic considerations in CHAPTER 6, we assumed that a HILBERT

space V exists such that all differential operators L_m with $m \in \mathcal{P}^{\text{adm}}$, are isomorphisms from V to $L_2(Q)$. However, our variational framework gives a different domain V_m for each material parameter and thus cannot be applied directly. Finding a variational setting working around this problem could justify the formal considerations in CHAPTER 6.

Appendix A

Explicit estimates using generalized symmetric eigenvalue problems

Let X be a discrete space and $A, M \in \mathcal{L}(X, X')$ self-adjoint positive operators defining norms $\|x\|_A = \sqrt{\langle Ax, x \rangle}$ and $\|x\|_M = \sqrt{\langle Mx, x \rangle}$.

Let $\lambda_1, \dots, \lambda_{N_X}$ be the eigenvalues of $Ax = \lambda Mx$. Then, we obtain

$$\|x\|_M \leq \left(\min \lambda_n \right)^{-1/2} \|x\|_A, \quad x \in X.$$

Let Y be second discrete space and $D \in \mathcal{L}(Y, Y')$ a self-adjoint positive operator. We observe for $B \in \mathcal{L}(X, Y')$

$$\|A^{-1}B'y\|_A = \sup_{x \in X} \frac{\langle Bx, y \rangle}{\|x\|_A}.$$

Let μ_1, \dots, μ_{N_Y} be the eigenvalues of $BA^{-1}B'y = \mu Dy$. Then, we obtain

$$\sup_{x \in X} \frac{\langle Bx, y \rangle}{\|x\|_A} \geq \left(\min \mu_k \right)^{-1/2} \|y\|_D, \quad y \in Y. \quad (\text{A.1})$$

Moreover, we obtain

$$\langle Bx, y \rangle \leq \left(\max \mu_k \right)^{1/2} \|x\|_A \|y\|_D, \quad x \in X, y \in Y.$$



Appendix B

Evaluate J'' by the adjoint state method

We summarize how the second order derivative of the parameter-to-misfit map can be evaluated using adjoint states using the notation of SECTION 6.4, see [35, Sec. 1.65]. In [29], the authors report that taking into account the second derivative of Φ can increase the reconstruction quality, but assembling the discretized counterpart of $J''(m^k)$ is practically infeasible.

Since $\mathcal{E}(m, \mathcal{F}(m), \mathbf{z}) = 0$ for $m \in \mathcal{P}^{\text{adm}}$, $\mathbf{z} \in V^*$, we obtain by (6.8), (6.17)

$$\begin{aligned} \langle J'(m^k), \tilde{m} \rangle &= (M'(m^k)[\tilde{m}] \partial_t \mathcal{F}(m^k), \mathbf{z})_Q + (L_{M(m^k)} \mathcal{F}'(m^k)[\tilde{m}], \mathbf{z})_Q \\ &\quad + \langle f'(\Phi(m^k)), \Phi'(m^k)[\tilde{m}] \rangle. \end{aligned}$$

To calculate the second derivative of J , we use $\Phi''(m^k)[\tilde{m}, \hat{m}] = \Psi[\mathcal{F}''(m^k)[\tilde{m}, \hat{m}]]$. Exploiting $\mathbf{y}(0) = 0$, $\mathbf{z}(T) = 0$ for $\mathbf{y} \in V$, $\mathbf{z} \in V^*$ and writing H_f for the HESSE matrix of the misfit-function f yield

$$\begin{aligned} J''(m^k)[\tilde{m}, \hat{m}] &= (M''(m^k)[\tilde{m}, \hat{m}] \partial_t \mathcal{F}(m^k), \mathbf{z})_Q + (M'(m^k)[\tilde{m}] \partial_t \mathcal{F}'(m^k)[\hat{m}], \mathbf{z})_Q \\ &\quad + (M'(m^k)[\hat{m}] \partial_t \mathcal{F}'(m^k)[\tilde{m}], \mathbf{z})_Q + (L_{M(m^k)} \mathcal{F}''(m^k)[\tilde{m}, \hat{m}], \mathbf{z})_Q \\ &\quad + f''(\Phi(m^k))[\Phi'(m^k)[\tilde{m}], \Phi'(m^k)[\hat{m}]] \\ &\quad + \langle f'(\Phi(m^k)), \Phi''(m^k)[\tilde{m}, \hat{m}] \rangle \\ &= (M''(m^k)[\tilde{m}, \hat{m}] \partial_t \mathcal{F}(m^k), \mathbf{z})_Q - (\hat{m}, \mathcal{F}'(m^k)^* [M'(m^k)[\tilde{m}]^* \partial_t \mathbf{z}])_\Omega \\ &\quad + (M'(m^k)[\hat{m}] \partial_t \mathcal{F}'(m^k)[\tilde{m}], \mathbf{z})_Q + (\mathcal{F}''(m^k)[\tilde{m}, \hat{m}], L_{M(m^k)}^* \mathbf{z})_Q \\ &\quad + (\mathcal{F}'(m^k)^* [\Psi^* H_f(\Phi(m^k)) \Phi'(m^k)[\tilde{m}]], \hat{m})_\mathcal{S} \\ &\quad + (\Psi^* [\nabla f(\Phi(m^k))], \mathcal{F}''(m^k)[\tilde{m}, \hat{m}])_Q. \end{aligned}$$

Reordering finally yields

$$\begin{aligned}
J''(m^k)[\tilde{m}, \hat{m}] &= (L_{M(m^k)}^* \mathbf{z} + \Psi^* [\nabla f(\Phi(m^k))], \mathcal{F}''(m^k)[\tilde{m}, \hat{m}])_Q \\
&\quad + (\mathcal{F}'(m^k)^* [\Psi^* \mathbf{H}_f(\Phi(m^k)) \Phi'(m^k)[\tilde{m}] - M'(m^k)[\tilde{m}]^* \partial_t \mathbf{z}], \hat{m})_\Omega \\
&\quad + (M'(m^k)[\hat{m}] \partial_t \mathcal{F}'(m^k)[\tilde{m}], \mathbf{z})_Q + (M''(m^k)[\tilde{m}, \hat{m}] \partial_t \mathcal{F}(m^k), \mathbf{z})_Q
\end{aligned}$$

Given $\tilde{m} \in \mathcal{P}$, the second derivative applied to \tilde{m} , i.e. $J''(m^k)[\tilde{m}, \cdot]$ can be evaluated efficiently by the following procedure.

Solve an adjoint problem to get rid of the first addend by finding $\mathbf{z}_{\text{adjoint}} \in V^*$

$$L_{M(m^k)}^* \mathbf{z}_{\text{adjoint}} = -\Psi^* [\nabla f(\Phi(m^k))]. \quad (\text{B.1})$$

In the k -th NEWTON step, the non-linear residual has to be evaluated. Thus, the value of $\mathbf{y}_{m^k} := \mathcal{F}(m^k)$ is known already. As a result, also $\Phi(m^k) = \Psi[\mathcal{F}(m^k)]$ is known and thus, the right-hand side of (B.1) be evaluated efficiently without significant additional costs.

Then, calculate $\Phi'(m^k)[\tilde{m}] \in \mathcal{S}$, yielding also $\mathbf{y}_{\text{lin}} := \mathcal{F}'(m^k)[\tilde{m}] \in V$ and

$$m_{\text{adjoint}} := \mathcal{F}'(m^k)^* [\Psi^* \mathbf{H}_f(\Phi(m^k)) \Phi'(m^k)[\tilde{m}] - M'(m^k)[\tilde{m}]^* \partial_t \mathbf{z}_{\text{adjoint}}] \in \mathcal{P}.$$

As a result, we obtain the following representation of the second derivative that can be evaluated efficiently for every $\hat{m} \in \mathcal{P}$ by scalar products only:

$$\begin{aligned}
J''(m^k)[\tilde{m}, \hat{m}] &= (m_{\text{adjoint}}, \hat{m})_\Omega + (M'(m^k)[\hat{m}] \partial_t \mathbf{y}_{\text{lin}}, \mathbf{z}_{\text{adjoint}})_Q \\
&\quad + (M''(m^k)[\tilde{m}, \hat{m}] \partial_t \mathbf{y}_{m^k}, \mathbf{z}_{\text{adjoint}})_Q
\end{aligned} \quad (\text{B.2})$$

The whole procedure is summarized in ALGORITHM 9 and requires solving three additional wave equations to the non-linear residual.

Algorithm 9 Evaluate second order derivative $J''(m^k)[\tilde{m}, \cdot]$

Input: $m^k \in \mathcal{P}^{\text{adm}}$, $\tilde{m} \in \mathcal{P}$, $\mathbf{y}_{m^k} := \mathcal{F}(m^k) \in V$

Output: $\mathbf{z}_{\text{adjoint}} \in V^*$, $\mathbf{y}_{\text{lin}} \in V$, $m_{\text{adjoint}} \in \mathcal{P}$ fulfilling (B.2)

- 1: Set $s_{m^k} := \Psi[\mathbf{y}_{m^k}] \in \mathcal{S}$.
 - 2: Find $\mathbf{z}_{\text{adjoint}} \in V^*$ with $L_{M(m^k)}^* \mathbf{z}_{\text{adjoint}} = -\Psi^* [\nabla f(s_{m^k})]$.
 - 3: Set $\mathbf{y}_{\text{lin}} := \mathcal{F}'(m^k)[\tilde{m}] \in V$.
 - 4: Set $s_{\text{lin}} := \Psi[\mathbf{y}_{\text{lin}}] \in \mathcal{S}$.
 - 5: Assemble $b := \Psi^* \mathbf{H}_f(s_{m^k})[s_{\text{lin}}] - M'(m^k)[\tilde{m}]^* \partial_t \mathbf{z}_{\text{adjoint}} \in W$
 - 6: Set $m_{\text{adjoint}} := \mathcal{F}'(m^k)^*[b] \in \mathcal{P}$.
-

Note that in the steps 2, 3, and 6 a wave equation needs to be solved.

Appendix C

An L_1 -setting in 1D

In CHAPTER 5, we observed higher order convergence of the piecewise defined mean-value compared the norm convergence order. In particular, we observed an improved convergence rate in the $L_1(Q)$ norm even for a solution with low regularity, see SECTION 5.2.3. In order to systematically analyze this phenomenon, a non-HILBERT space setting might be required. In the following, we sketch an L_1 setting for waves using BV regularity. All considerations in the following are performed in one spatial dimension.

Furthermore, this setting might also help to set up a BANACH space framework to consider the FWI problem, see CHAPTER 6.

Functions with bounded variation

The *total variation* of a function $f: \Omega \rightarrow \mathbb{R}$, $\Omega = \mathbb{R}$, is defined as

$$|f|_{\text{BV}(\Omega)} := \sup \left\{ \sum_{k=1}^K |f(x_{k-1}) - f(x_k)| : x_0 < x_1 < \dots < x_K, K \in \mathbb{N} \right\}$$

and the space of functions with bounded variation in 1D is given by

$$\text{BV}(\Omega) := \left\{ f: \Omega \rightarrow \mathbb{R} : \|f\|_{L^1(\Omega)}, |f|_{\text{BV}(\Omega)} < \infty \right\}.$$

For a function of multiple variables $u: U \rightarrow \mathbb{R}$, $U \subset \mathbb{R}^n$ open, we define the *variation* as in [28, Sec. 5, p. 166] by

$$|u|_{\text{BV}(U)} := \sup \left\{ \int_U u \operatorname{div} \varphi \, dx : \varphi \in C_c^1(U, \mathbb{R}^n), |\varphi| \leq 1 \right\}.$$

The space of functions with bounded variation is then defined by

$$\text{BV}(U) := \left\{ u \in L^1(U) : |u|_{\text{BV}(U)} < \infty \right\}.$$

Remark C.1. In [28, Sec. 5.10, Thm. 1, p. 217], it is shown that for $n = 1$ both definitions of $|f|_{\text{BV}(\Omega)}$ are compatible.

By [28, Section 5.5, p. 185] the variation of u can also be written in terms of the level set $E_t := \{x \in U : u(x) > t\}$ of u , i.e.

$$|u|_{\text{BV}(U)} = \int_{-\infty}^{\infty} \|\partial E_t\|(U) dt, \quad (\text{C.1})$$

where $\|\partial E_t\|(U)$ is the perimeter of E_t in U , cf. [28, Sec. 5, p. 169].

Example C.2. Let $U = U_1 \cup U_2$ be a disjoint partition and $u(x) = \alpha_i$ for $x \in U_i$. Then, by (C.1), we have $|u|_{\text{BV}(U)} = \lambda^{n-1}(\partial U_1 \setminus \partial U) \cdot |\alpha_1 - \alpha_2|$.

Define for $k \in \frac{1}{2}\mathbb{N}_0$

$$\mathcal{S}^k = \begin{cases} \mathbb{Z}, & k \notin \mathbb{N}_0, \\ \frac{1}{2} + \mathbb{Z}, & k \in \mathbb{N}_0, \end{cases}$$

and consider a grid of staggered diamonds $D_n^k \subset Q$, $(k, n) \in \frac{1}{2}\mathbb{N}_0 \times \mathcal{S}^k$, with

$$D_n^k := \text{conv} \left((nh, k\Delta t) \pm \{(h/2, 0), (0, \Delta t/2)\} \right) \subset \Omega \times (0, T).$$

Here, $\Delta t = h/c$ for the speed of sound $c > 0$. A piece-wise constant weak solution (p, q) of the acoustic wave equation on the grid of these diamonds fulfills

$$\begin{aligned} p_n^k &= \frac{1}{2} \left(p_{n-1/2}^{k-1/2} + q_{n-1/2}^{k-1/2} + p_{n+1/2}^{k-1/2} - q_{n+1/2}^{k-1/2} \right), \\ q_n^k &= \frac{1}{2} \left(p_{n-1/2}^{k-1/2} + q_{n-1/2}^{k-1/2} - p_{n+1/2}^{k-1/2} + q_{n+1/2}^{k-1/2} \right), \end{aligned} \quad (\text{C.2})$$

for $k \in \frac{1}{2}\mathbb{N}$, $n \in \mathcal{S}^k$, see FIGURE C.1 and FIGURE C.2 for illustrations and EXAMPLE 3.6 for the idea of proof that this yields indeed a weak solution.

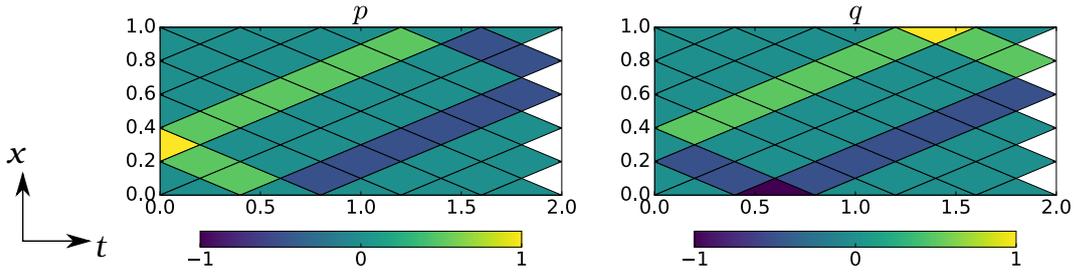


Figure C.1: Space-time plot of an ultra-weak solution that is piece-wise constant on the a diamond grid. The initial value is $(p_0, q_0) = (\mathbf{1}_{(0,2,0,4)}, 0)$

Lemma C.3. We have for $k \in \frac{1}{2}\mathbb{N}$, $n \in \mathcal{S}^k$

$$\begin{aligned} p_n^k &= \frac{1}{2} \left(p_{n-k}^0 + q_{n-k}^0 + p_{n+k}^0 - q_{n+k}^0 \right), \\ q_n^k &= \frac{1}{2} \left(p_{n-k}^0 + q_{n-k}^0 - p_{n+k}^0 + q_{n+k}^0 \right). \end{aligned}$$

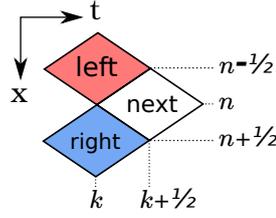


Figure C.2: Illustration of (C.2).

Proof. We only consider the p -component since the proof for q works analogously. We proof by induction that for $m \in \mathbb{N}$, $m \leq 2k$, it holds

$$p_n^k = \frac{1}{2} \left(p_{n-m/2}^{k-m/2} + q_{n-m/2}^{k-m/2} + p_{n+m/2}^{k-m/2} - q_{n+m/2}^{k-m/2} \right). \quad (\text{C.3})$$

For $m = 1$, (C.3) states the result from the weak formulation. Assuming that (C.3) is fulfilled for $m \in \mathbb{N}$, $m < 2k$, we obtain

$$\begin{aligned} p_n^k &= \frac{1}{2} \left(p_{n-m/2}^{k-m/2} + q_{n-m/2}^{k-m/2} + p_{n+m/2}^{k-m/2} - q_{n+m/2}^{k-m/2} \right) \\ &= \frac{1}{4} \left(p_{n-(m+1)/2}^{k-(m+1)/2} + q_{n-(m+1)/2}^{k-(m+1)/2} + p_{n-(m-1)/2}^{k-(m+1)/2} - q_{n-(m-1)/2}^{k-(m+1)/2} \right. \\ &\quad + p_{n-(m+1)/2}^{k-(m+1)/2} + q_{n-(m+1)/2}^{k-(m+1)/2} - p_{n-(m-1)/2}^{k-(m+1)/2} + q_{n-(m-1)/2}^{k-(m+1)/2} \\ &\quad + p_{n+(m-1)/2}^{k-(m+1)/2} + q_{n+(m-1)/2}^{k-(m+1)/2} + p_{n+(m+1)/2}^{k-(m+1)/2} - q_{n+(m+1)/2}^{k-(m+1)/2} \\ &\quad \left. - p_{n+(m-1)/2}^{k-(m+1)/2} - q_{n+(m-1)/2}^{k-(m+1)/2} + p_{n+(m+1)/2}^{k-(m+1)/2} - q_{n+(m+1)/2}^{k-(m+1)/2} \right) \\ &= \frac{1}{2} \left(p_{n-(m+1)/2}^{k-(m+1)/2} + q_{n-(m+1)/2}^{k-(m+1)/2} + p_{n+(m+1)/2}^{k-(m+1)/2} - q_{n+(m+1)/2}^{k-(m+1)/2} \right). \end{aligned}$$

Now, inserting $m = 2k$ yields the assertion. \square

Lemma C.4. For $k \in \frac{1}{2}\mathbb{N}_0$ it holds

$$\begin{aligned} \|p^k\|_{L^1(\Omega)}, \|q^k\|_{L^1(\Omega)} &\leq \|(p^0, q^0)\|_{L^1(\Omega)}, \\ |p^k|_{\text{BV}(\Omega)}, |q^k|_{\text{BV}(\Omega)} &\leq |(p^0, q^0)|_{\text{BV}(\Omega)}. \end{aligned}$$

Proof. By Lemma C.3 we obtain

$$\begin{aligned} \|p^k\|_{L^1(\Omega)} &= \sum_{n \in \mathcal{S}^k} h \cdot |p_n^k| = \sum_{n \in \mathcal{S}^k} \frac{h}{2} \cdot |p_{n-k}^0 + q_{n-k}^0 + p_{n+k}^0 - q_{n+k}^0| \\ &\leq \frac{h}{2} \left(\sum_{n \in \mathcal{S}^k} |p_n^0| + \sum_{n \in \mathcal{S}^k} |q_n^0| + \sum_{n \in \mathcal{S}^k} |p_n^0| + \sum_{n \in \mathcal{S}^k} |q_n^0| \right) \\ &= h \cdot \left(\sum_{n \in \mathcal{S}^0} |p_n^0| + \sum_{n \in \mathcal{S}^0} |q_n^0| \right) = \|p^0\|_{L^1(\Omega)} + \|q^0\|_{L^1(\Omega)}, \\ \|q^k\|_{L^1(\Omega)} &\leq \|p^0\|_{L^1(\Omega)} + \|q^0\|_{L^1(\Omega)}, \end{aligned}$$

proving the first assertion.

With $\llbracket \eta \rrbracket_n^k := \eta_{n+1/2}^k - \eta_{n-1/2}^k$, $\eta = p, q$ and $n \in \mathcal{S}^{k+1/2}$, we obtain for $n \in \mathcal{S}^{k+1/2}$ as a direct consequence of Lemma C.3

$$\begin{aligned}\llbracket p \rrbracket_n^k &= \frac{1}{2} \left(\llbracket p \rrbracket_{n-k}^0 + \llbracket q \rrbracket_{n-k}^0 + \llbracket p \rrbracket_{n+k}^0 - \llbracket q \rrbracket_{n+k}^0 \right), \\ \llbracket q \rrbracket_n^k &= \frac{1}{2} \left(\llbracket p \rrbracket_{n-k}^0 + \llbracket q \rrbracket_{n-k}^0 - \llbracket p \rrbracket_{n+k-1}^0 + \llbracket q \rrbracket_{n+k-1}^0 \right).\end{aligned}$$

yielding to

$$\begin{aligned}|p^k|_{\text{BV}(\Omega)} &= \sum_{n \in \mathcal{S}^{k+1/2}} |\llbracket p \rrbracket_n^k| = \sum_{n \in \mathcal{S}^{k+1/2}} \frac{1}{2} \left| \llbracket p \rrbracket_{n-k}^0 + \llbracket q \rrbracket_{n-k}^0 + \llbracket p \rrbracket_{n+k}^0 - \llbracket q \rrbracket_{n+k}^0 \right| \\ &\leq \sum_{n \in \mathcal{S}^0} |\llbracket p \rrbracket_n^0| + \sum_{n \in \mathcal{S}^0} |\llbracket q \rrbracket_n^0| = (|p^0|_{\text{BV}(\Omega)} + |q^0|_{\text{BV}(\Omega)}) \\ &= |(p^0, q^0)|_{\text{BV}(\Omega)}.\end{aligned}$$

The proof for $|q^k|_{\text{BV}(\Omega)} \leq |(p^0, q^0)|_{\text{BV}(\Omega)}$ is done using the same argument. \square

Lemma C.5. *With $\Delta t = \frac{h}{c}$, $T = K\Delta t$, $K \in \frac{1}{2}\mathbb{N}_0$ it holds*

$$\|p\|_{\text{L}^1(\Omega \times (0, T))}, \|q\|_{\text{L}^1(\Omega \times (0, T))} \leq T \cdot \|(p^0, q^0)\|_{\text{L}^1(\Omega)}$$

Proof. Set $I_k := (\frac{k-1}{2}\Delta t, \frac{k}{2}\Delta t)$, $\Omega_n := h \cdot (n - \frac{1}{2}, n + \frac{1}{2})$ for $k \in \mathbb{N}$, $n \in \frac{1}{2}\mathbb{Z}$.

$$\begin{aligned}\|p\|_{\text{L}^1(\Omega \times (0, T))} &= \int_{\Omega \times (0, K\Delta t)} |p(x, t)| = \sum_{k=1}^{2K} \int_{\Omega \times I_k} |p(x, t)| \\ &= \sum_{k=1}^{2K} \left(\sum_{n \in \mathcal{S}^{(k-1)/2}} \int_{\Omega_n \times I_k} |p^{(k-1)/2}(x, t)| + \sum_{n \in \mathcal{S}^{k/2}} \int_{\Omega_n \times I_k} |p^{k/2}(x, t)| \right) \\ &= \frac{\Delta t}{4} \sum_{k=1}^{2K} \left(\sum_{n \in \mathcal{S}^{(k-1)/2}} h \cdot |p_n^{(k-1)/2}| + \sum_{n \in \mathcal{S}^{k/2}} h \cdot |p_n^{k/2}| \right) \\ &= \frac{\Delta t}{4} \sum_{k=1}^{2K} \left(\|p^{(k-1)/2}\|_{\text{L}^1(\Omega)} + \|p^{k/2}\|_{\text{L}^1(\Omega)} \right) \\ &\leq \frac{\Delta t}{4} \cdot 2K \cdot 2 \|(p^0, q^0)\|_{\text{L}^1(\Omega)} = T \cdot \|(p^0, q^0)\|_{\text{L}^1(\Omega)}.\end{aligned}$$

The estimate for q is obtained in the same way. \square

Lemma C.6. *With $\Delta t = \frac{h}{c}$, $T = K\Delta t$, $K \in \frac{1}{2}\mathbb{N}_0$ it holds*

$$|p|_{\text{BV}(\Omega \times (0, T))}, |q|_{\text{BV}(\Omega \times (0, T))} \leq 2T \sqrt{1 + c^2} \|(p^0, q^0)\|_{\text{BV}(\Omega)}$$

Proof. Let $k \in \frac{1}{2}\mathbb{N}$, $n \in \mathcal{S}^k$. Then by Lemma C.3

$$\begin{aligned} \left| p_{n-1/2}^{k-1/2} - p_n^k \right| &= \left| p_{n-1/2}^{k-1/2} - \frac{1}{2} \left(p_{n-1/2}^{k-1/2} + p_{n+1/2}^{k-1/2} - q_{n+1/2}^{k-1/2} \right) \right| \\ &= \frac{1}{2} \left| p_{n-1/2}^{k-1/2} - p_{n+1/2}^{k-1/2} + q_{n+1/2}^{k-1/2} - q_{n-1/2}^{k-1/2} \right| \\ &\leq \frac{1}{2} \left(\left| \llbracket p \rrbracket_n^{k-1/2} \right| + \left| \llbracket q \rrbracket_n^{k-1/2} \right| \right) \end{aligned}$$

and analogously

$$\left| p_{n+1/2}^{k-1/2} - p_n^k \right| \leq \frac{1}{2} \left(\left| \llbracket p \rrbracket_n^{k-1/2} \right| + \left| \llbracket q \rrbracket_n^{k-1/2} \right| \right).$$

Setting $\delta := \sqrt{\frac{h^2}{4} + \frac{\Delta t^2}{4}} = \frac{h}{2} \sqrt{1 + c^{-2}}$ yields

$$\begin{aligned} |p|_{\text{BV}(\Omega \times I_k)} &= \delta \cdot \left(\sum_{n \in \mathcal{S}^k} \left| p_{n-1/2}^{k-1/2} - p_n^k \right| + \left| p_{n+1/2}^{k-1/2} - p_n^k \right| \right) \\ &\leq \delta \cdot \left(\sum_{n \in \mathcal{S}^k} \left| \llbracket p \rrbracket_n^{k-1/2} \right| + \left| \llbracket q \rrbracket_n^{k-1/2} \right| \right) \\ &= \delta \cdot \left(|p^{k-1/2}|_{\text{BV}(\Omega)} + |q^{k-1/2}|_{\text{BV}(\Omega)} \right). \end{aligned}$$

Thus, by Lemma C.4 we obtain

$$|p|_{\text{BV}(\Omega \times I_k)} \leq 2\delta \cdot \left(|p^0|_{\text{BV}(\Omega)} + |q^0|_{\text{BV}(\Omega)} \right).$$

Finally, we calculate

$$\begin{aligned} |p|_{\text{BV}(\Omega \times (0, K\Delta t))} &= \sum_{k=1}^{2K} |p|_{\text{BV}(\Omega \times I_k)} \leq \sum_{k=1}^{2K} 2\delta \cdot \left(|p^0|_{\text{BV}(\Omega)} + |q^0|_{\text{BV}(\Omega)} \right) \\ &= \left(|p^0|_{\text{BV}(\Omega)} + |q^0|_{\text{BV}(\Omega)} \right) \cdot 2Kh\sqrt{1 + c^{-2}} \\ &= \left(|p^0|_{\text{BV}(\Omega)} + |q^0|_{\text{BV}(\Omega)} \right) \cdot 2T\sqrt{c^2 + 1}. \end{aligned}$$

The estimate for q is obtained in the same way. \square

Theorem C.7. *For the piece-wise constant weak solution (p, q) of the acoustic wave equation with piece-wise constant initial value (p_0, q_0) , we have in $Q = \Omega \times (0, T)$*

$$\begin{aligned} \|(p, q)\|_{L_1(Q)} &\leq 2T \|(p^0, q^0)\|_{L_1(\Omega)}, \\ |(p, q)|_{\text{BV}(Q)} &\leq 2T\sqrt{1 + c^2} |(p^0, q^0)|_{\text{BV}(\Omega)}. \end{aligned}$$

Proof. Apply LEMMA C.6 and LEMMA C.5. \square

Approximation discussion

Consider the L_2 projection $\Pi_h^0: L_1(\Omega) \rightarrow \prod_K \mathbb{P}_0(K)$. We assume that there is a constant $C > 0$ such that for all $u \in \text{BV}(\Omega)$, we have

$$|\Pi_h^0 u|_{\text{BV}(\Omega)} \leq C |u|_{\text{BV}(\Omega)}. \quad (\text{C.4})$$

Similar interpolation estimates can be found in [3] and [10].

For $(p_0, q_0) \in L_1(\Omega)$, we have by LEBESGUE's differentiation theorem and the theorem of dominated convergence

$$\|(p_0, q_0) - (\Pi_h^0 p_0, \Pi_h^0 q_0)\|_{L_1(\Omega)} \rightarrow 0, \quad h \rightarrow 0.$$

Let $(p_h, q_h) \in \mathbb{P}_0(Q_h)$ be the weak solution of the wave equation with initial value $(\Pi_h^0 p_0, \Pi_h^0 q_0)$ obtained by (C.2).

Then, we conclude using THEOREM C.7 for $h, \tau > 0$

$$\begin{aligned} \|(p_h, q_h) - (p_\tau, q_\tau)\|_{L_1(Q)} &\leq 2T \left\| (\Pi_h^0 p_0, \Pi_h^0 q_0) - (\Pi_\tau^0 p_0, \Pi_\tau^0 q_0) \right\|_{L_1(\Omega)} \\ &\leq \left\| (\Pi_h^0 p_0, \Pi_h^0 q_0) - (p_0, q_0) \right\|_{L_1(\Omega)} \\ &\quad + \left\| (p_0, q_0) - (\Pi_\tau^0 p_0, \Pi_\tau^0 q_0) \right\|_{L_1(\Omega)} \end{aligned}$$

Thus, for every null sequence $(h_n)_n \in (0, \infty)^\mathbb{N}$, $((p_{h_n}, q_{h_n}))_n$ is a CAUCHY sequence having an $L_1(Q)$ limit $(p, q) \in L_1(Q)$.

Applying THEOREM C.7 again and setting $C := 2T\sqrt{1+c^2}$, we see that (p_h, q_h) is bounded in $\text{BV}(Q)$ since by (C.4)

$$|(p_h, q_h)|_{\text{BV}(Q_h)} \leq C |(\Pi_h^0 p_0, \Pi_h^0 q_0)|_{\text{BV}(\Omega)} \leq C |(p_0, q_0)|_{\text{BV}(\Omega)}.$$

Using the compact embedding of $\text{BV}(Q)$ into $L_1(Q)$, see [28, Thm. 4, Sec. 5.2.4, p. 176], this yields a function $(\tilde{p}, \tilde{q}) \in \text{BV}(Q)$ and a subsequence of (p_h, q_h) converging to (\tilde{p}, \tilde{q}) in $L_1(Q)$.

This implies $(p, q) = (\tilde{p}, \tilde{q}) \in \text{BV}(Q)$, since the whole sequence (p_h, q_h) converges to (p, q) in $L_1(Q)$.

Outlook

This setting provides a constructive proof that for every initial value in $\text{BV}(\Omega)$ a weak solution in $\text{BV}(Q)$ exists that is bounded by the norm of the initial value.

In the future, we would like to elaborate a similar setting for 2 or even 3 spatial dimensions using the same or similar arguments. However, in order to make this possible, we need a sensible generalization of the BV -space for vector-fields that does *not* treat all components separately.

Bibliography

- [1] ADLER, James H. ; BRANNICK, James J. ; LIU, Chun ; MANTEUFFEL, Thomas ; ZIKATANOV, Ludmil: First-order system least squares and the energetic variational approach for two-phase flow. In: *Journal of Computational Physics* 230 (2011), Nr. 17, S. 6647–6663
- [2] ALBERT, Arthur: *Regression and the Moore-Penrose pseudoinverse*. Elsevier, 1972
- [3] BARTELS, Sören: Total variation minimization with finite elements: convergence and iterative solution. In: *SIAM Journal on Numerical Analysis* 50 (2012), Nr. 3, S. 1162–1180
- [4] BLAZEK, Kirk D. ; STOLK, Christiaan ; SYMES, William W.: A mathematical framework for inverse wave problems in heterogeneous media. In: *Inverse Problems* 29 (2013), Nr. 6, S. 065001
- [5] BOCHEV, Pavel B. ; GUNZBURGER, Max D.: Finite element methods of least-squares type. In: *SIAM review* 40 (1998), Nr. 4, S. 789–837
- [6] BOCHEV, Pavel B. ; GUNZBURGER, Max D.: *Least-squares finite element methods*. Bd. 166. Springer Science & Business Media, 2009
- [7] BOFFI, Daniele ; BREZZI, Franco ; FORTIN, Michel: *Mixed finite element methods and applications*. Bd. 44. Springer Science & Business Media, 2013
- [8] BOHLEN, Thomas: Parallel 3-D viscoelastic finite difference seismic modelling. In: *Computers & Geosciences* 28 (2002), Nr. 8, S. 887–899
- [9] BRAESS, Dietrich: *Finite elements: Theory, fast solvers, and applications in solid mechanics*. Cambridge University Press, 2007
- [10] BUFFA, Annalisa ; ORTNER, Christoph: Compact embeddings of broken Sobolev spaces and applications. In: *IMA journal of numerical analysis* 29 (2008), Nr. 4, S. 827–855
- [11] BUI-THANH, Tan ; DEMKOWICZ, Leszek ; GHATTAS, Omar: A unified discontinuous Petrov–Galerkin method and its analysis for Friedrichs’ systems. In:

- SIAM Journal on Numerical Analysis* 51 (2013), Nr. 4, S. 1933–1958
- [12] BYRD, Richard H. ; LU, Peihuang ; NOCEDAL, Jorge ; ZHU, Ciyou: A limited memory algorithm for bound constrained optimization. In: *SIAM Journal on Scientific Computing* 16 (1995), Nr. 5, S. 1190–1208
- [13] CAI, Zhiqiang ; LAZAROV, R. ; MANTEUFFEL, Thomas A. ; MCCORMICK, Stephen F.: First-order system least squares for second-order partial differential equations: Part I. In: *SIAM Journal on Numerical Analysis* 31 (1994), Nr. 6, S. 1785–1799
- [14] CARSTENSEN, Carsten ; DEMKOWICZ, Leszek ; GOPALAKRISHNAN, Jay: Breaking spaces and forms for the DPG method and applications including Maxwell equations. In: *Computers & Mathematics with Applications* 72 (2016), Nr. 3, S. 494–522
- [15] CARSTENSEN, Carsten ; HELLWIG, Friederike: Low-order discontinuous Petrov–Galerkin finite element methods for linear elasticity. In: *SIAM Journal on Numerical Analysis* 54 (2016), Nr. 6, S. 3388–3410
- [16] CHAVENT, Guy: *Nonlinear least squares for inverse problems: theoretical foundations and step-by-step guide for applications*. Springer, 2009
- [17] DEMKOWICZ, Leszek: Various variational formulations and closed range theorem
- [18] DEMKOWICZ, Leszek ; GOPALAKRISHNAN, Jay: A class of discontinuous Petrov–Galerkin methods. II. Optimal test functions. In: *Numerical Methods for Partial Differential Equations* 27 (2011), Nr. 1, S. 70–105
- [19] DEMKOWICZ, Leszek ; GOPALAKRISHNAN, Jay: An overview of the discontinuous Petrov Galerkin method. In: *Recent developments in discontinuous Galerkin finite element methods for partial differential equations*. Springer, 2014, S. 149–180
- [20] DEMKOWICZ, Leszek ; GOPALAKRISHNAN, Jay: ICES REPORT 15-20. (2015)
- [21] DEMKOWICZ, Leszek ; GOPALAKRISHNAN, Jay ; NAGARAJ, Sriram ; SEPÚLVEDA, Paulina: A Spacetime DPG Method for the Schrödinger Equation. In: *SIAM Journal on Numerical Analysis* 55 (2017), Nr. 4, S. 1740–1759
- [22] DEMKOWICZ, Leszek ; HEUER, Norbert: Robust DPG method for convection-dominated diffusion problems. In: *SIAM Journal on Numerical Analysis* 51 (2013), Nr. 5, S. 2514–2537
- [23] DÖRFLER, Willy ; FINDEISEN, Stefan ; WIENERS, Christian: Space-time discontinuous Galerkin discretizations for linear first-order hyperbolic evolution

- systems. In: *Computational Methods in Applied Mathematics* 16 (2016), Nr. 3, S. 409–428
- [24] ERN, Alexandre ; GUERMOND, Jean-Luc: *Theory and practice of finite elements*. Bd. 159. Springer Science & Business Media, 2013
- [25] ERN, Alexandre ; GUERMOND, Jean-Luc: A converse to Fortin’s lemma in Banach spaces. In: *Comptes Rendus Mathématique* 354 (2016), Nr. 11, S. 1092–1095
- [26] ERN, Alexandre ; GUERMOND, Jean-Luc ; CAPLAIN, Gilbert: An Intrinsic criterion for the bijectivity of Hilbert operators related to Friedrich systems. In: *Communications in partial differential equations* 32 (2007), Nr. 2, S. 317–341
- [27] ERNESTI, Johannes ; WIENERS, Christian: A space-time DPG method for acoustic waves. In: *Radon Series on Computational and Applied Mathematics* (submitted)
- [28] EVANS, Lawrence C. ; GARIEPY, Ronald F.: *Measure theory and fine properties of functions*. CRC press, 2015
- [29] FICHTNER, Andreas ; TRAMPERT, Jeannot: Hessian kernels of seismic data functionals based upon adjoint techniques. In: *Geophys. J. Int* 185 (2011), S. 775–798
- [30] FINDEISEN, Stefan M.: *A parallel and adaptive space-Time method for Maxwell’s equations*, Karlsruhe Institut für Technologie (KIT), Diss., 2016
- [31] FOUQUE, Jean-Pierre ; GARNIER, Josselin ; PAPANICOLAOU, George ; SOLNA, Knut: *Wave propagation and time reversal in randomly layered media*. Bd. 56. Springer Science & Business Media, 2007
- [32] GANDER, Martin J.: 50 years of time parallel time integration. In: *Multiple Shooting and Time Domain Decomposition Methods*. Springer, 2015, S. 69–113
- [33] GOPALAKRISHNAN, Jay ; SEPULVEDA, Paulina: A spacetime DPG method for acoustic waves. In: *arXiv preprint arXiv:1709.08268* (2017)
- [34] HEWETT, Russell J. ; DEMANET, Laurent ; PYSIT TEAM the: *PySIT: Python Seismic Imaging Toolbox v0.5*. <http://www.pysit.org>. Version: 2013. – Release 0.5
- [35] HINZE, Michael ; PINNAU, René ; ULBRICH, Michael ; ULBRICH, Stefan: *Optimization with PDE constraints*. Bd. 23. Springer Science & Business Media, 2008
- [36] HOCHBRUCK, Marlis ; PAŽUR, Tomislav ; SCHULZ, Andreas ; THAWINAN, Ekkachai ; WIENERS, Christian: Efficient time integration for discontinuous

- Galerkin approximations of linear wave equations. In: *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik* 95 (2015), Nr. 3, S. 237–259
- [37] HOFMANN, Bernd ; SCHERZER, Otmar: Factors influencing the ill-posedness of nonlinear problems. In: *Inverse Problems* 10 (1994), Nr. 6, S. 1277
- [38] HOFMANN, Bernd ; SCHERZER, Otmar: Local ill-posedness and source conditions of operator equations in Hilbert spaces. In: *Inverse Problems* 14 (1998), Nr. 5, S. 1189
- [39] KEITH, Brendan ; PETRIDES, Socratis ; FUENTES, Federico ; DEMKOWICZ, Leszek: Discrete least-squares finite element methods. In: *ArXiv e-prints* (2017)
- [40] KIRSCH, Andreas: *An introduction to the mathematical theory of inverse problems*. Bd. 120. Springer Science & Business Media, 2011
- [41] KIRSCH, Andreas ; RIEDER, Andreas: On the linearization of operators related to the full waveform inversion in seismology. In: *Mathematical Methods in the Applied Sciences* 37 (2014), Nr. 18, S. 2995–3007
- [42] KIRSCH, Andreas ; RIEDER, Andreas: Seismic tomography is locally ill-posed. In: *Inverse Problems* 30 (2014), Nr. 12, S. 125001
- [43] KIRSCH, Andreas ; RIEDER, Andreas: Inverse problems for abstract evolution equations with applications in electrodynamics and elasticity. In: *Inverse Problems* 32 (2016), Nr. 8
- [44] LECHLEITER, Armin ; SCHLASCHKE, John W.: Identifying Lamé parameters from time-dependent elastic wave measurements. In: *Inverse Problems in Science and Engineering* 25 (2017), Nr. 1, S. 2–26
- [45] LEIS, Rolf: *Initial boundary value problems in mathematical physics*. Teubner, 1986
- [46] MARGOTTI, Fabio: *On inexact Newton methods for inverse problems in Banach spaces*, Karlsruhe Institut für Technologie (KIT), Diss., 2015
- [47] MAURER, Daniel: *Ein hochskalierbarer paralleler direkter Löser für Finite Elemente Diskretisierungen*, Karlsruhe Institut für Technologie (KIT), Diss., 2013
- [48] MÉTIVIER, Ludovic ; BROSSIER, Romain: The SEISCOPE optimization toolbox: A large-scale nonlinear optimization library based on reverse communication. In: *Geophysics* (2016)
- [49] MÉTIVIER, Ludovic ; BROSSIER, Romain ; VIRIEUX, Jean ; OPERTO, Stéphane: Full waveform inversion and the truncated Newton method. In: *SIAM Journal on Scientific Computing* 35 (2013), Nr. 2, S. B401–B437

- [50] MORA, Jaime ; DEMKOWICZ, Leszek: Fast integration of DPG matrices based on tensorization. (2017). – ICES REPORT 17-26
- [51] NEUMÜLLER, Martin: *Space-Time Methods: Fast Solvers and Applications*, Graz University of Technology, June 2013, Diss., 2013
- [52] PETRIDES, Socratis ; DEMKOWICZ, Leszek: An adaptive DPG method for high frequency time-harmonic wave propagation problems. In: *ICES Report*
- [53] PLESSIX, Rene-Edouard: A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. In: *Geophysical Journal International* 167 (2006), Nr. 2, S. 495–503
- [54] PRATT, Gerhard: Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model. In: *Geophysics* 64 (1999), Nr. 3, S. 888–901
- [55] RACKE, Reinhard: *Lectures on nonlinear evolution equations*. Birkhäuser, 1992. – Second Edition 2015
- [56] RENARDY, Michael ; ROGERS, Robert C.: *An introduction to partial differential equations*. Bd. 13. Springer Science & Business Media, 2006
- [57] RIEDER, Andreas: On the regularization of nonlinear ill-posed problems via inexact Newton iterations. In: *Inverse Problems* 15 (1999), Nr. 1, S. 309
- [58] RIEDER, Andreas: *Keine Probleme mit Inversen Problemen: Eine Einführung in ihre stabile Lösung*. Springer-Verlag, 2013
- [59] ROBERTSSON, Johan O. ; BLANCH, Joakim O. ; SYMES, William W.: Viscoelastic finite-difference modeling. In: *Geophysics* 59 (1994), Nr. 9, S. 1444–1456
- [60] RUDIN, Walter: *Functional Analysis*. McGraw-Hill, Inc., New York, 1991 (International Series in Pure and Applied Mathematics)
- [61] SYMES, William W.: The seismic reflection inverse problem. In: *Inverse problems* 25 (2009), Nr. 12, S. 123008
- [62] SZYLD, Daniel B.: The many proofs of an identity on the norm of oblique projections. In: *Numerical Algorithms* 42 (2006), Nr. 3, S. 309–323
- [63] TARANTOLA, Albert: Inversion of seismic reflection data in the acoustic approximation. In: *Geophysics* 49 (1984), Nr. 8, S. 1259–1266
- [64] VIRIEUX, Jean: P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method. In: *Geophysics* 51 (1986), Nr. 4, S. 889–901
- [65] VIRIEUX, Jean ; OPERTO, Stéphane: An overview of full-waveform inversion in exploration geophysics. In: *Geophysics* 74 (2009), Nr. 6, S. WCC1–WCC26

- [66] WIENERS, Christian: A geometric data structure for parallel finite elements and the application to multigrid methods with block smoothing. In: *Computing and visualization in science* 13 (2010), Nr. 4, S. 161–175
- [67] WIENERS, Christian: The skeleton reduction for finite element substructuring methods. In: *Numerical Mathematics and Advanced Applications ENUMATH 2015* Springer, 2016, S. 133–141
- [68] WIENERS, Christian ; WOHLMUTH, Barbara: Robust operator estimates and the application to substructuring methods for first-order systems. In: *ESAIM: Mathematical Modelling and Numerical Analysis* 48 (2014), Nr. 5, S. 1473–1494
- [69] WINKLER, Robert: *A model-aware inexact Newton scheme for electrical impedance tomography*, Karlsruhe Institut für Technologie (KIT), Diss., 2016
- [70] WOŹNIAK, Maciej ; ŁOŚ, Marcin ; PASZYŃSKI, Maciej ; DEMKOWICZ, Leszek: Fast parallel integration for three dimensional Discontinuous Petrov–Galerkin method. In: *Procedia Computer Science* 101 (2016), S. 8–17
- [71] XU, Jinchao ; ZIKATANOV, Ludmil: Some observations on Babuska and Brezzi theories. In: *Numerische Mathematik* 94 (2003), Nr. 1, S. 195–202
- [72] YOSIDA, Kosaku: Functional analysis. Reprint of the sixth (1980) edition. Classics in Mathematics. In: *Springer-Verlag, Berlin* 11 (1995), S. 14
- [73] ZITELLI, Jeff ; MUGA, Ignacio ; DEMKOWICZ, Leszek ; GOPALAKRISHNAN, Jayadeep ; PARDO, David ; CALO, Victor M.: A class of discontinuous Petrov–Galerkin methods. Part IV: The optimal test norm and time-harmonic wave propagation in 1D. In: *Journal of Computational Physics* 230 (2011), Nr. 7, S. 2406–2432

Danksagung

An dieser Stelle möchte ich mich bei den Menschen bedanken, ohne deren Unterstützung diese Arbeit nicht entstanden wäre und die meine Zeit am IANM zu etwas Besonderem gemacht haben.

Mein tiefer Dank gilt meinem Doktorvater Prof. Dr. Christian Wieners für das entgegengebrachte Vertrauen, die zahllosen fachlichen Diskussionen, die tatkräftige Unterstützung sowohl in der Theorie als auch beim Debugging, und vor allem für seine Geduld — Herr Wieners, ich denke Sie wissen, wie sehr ich unsere Zusammenarbeit schätze! Außerdem erinnere ich mich gerne an fliegende Fahnen an Quadrocoptern, mobile Touchscreens und Nachmittage auf Wasserski zurück. Für das gründliche Lesen möchte ich mich bei meinem Korreferenten Prof. Dr. Andreas Rieder bedanken — Herr Rieder, danke dafür, dass Sie sich nicht von der teils doch sehr technischen Notation haben abschrecken lassen und somit viele Unzulänglichkeiten der Arbeit aufdecken konnten! Christian Rheinbay danke ich für seine Diskussionsbereitschaft und unsere enge Zusammenarbeit — Christian, ohne dich hätte das FWI-Kapitel nicht die aktuelle Qualität erreicht und ich wüsste deutlich weniger über die Struktura von FWI!

Beim Jagen der (leider) unzähligen und für mich gegen Ende unsichtbaren Fehlerleufel im Manuskript wurde ich von Christian Rheinbay, Lydia Wagner, Carlos Hauser, Christine Gratwohl, Luca Hornung, Marcel Mikl, Julian Krämer, Mario Fernandez, Daniel Ziegler, Phillip Ernesti, Lina Ostermann-Schelleckes, Daniel Weiß, Daniele Corallo und Kevin Ganster unterstützt, wofür ich sehr dankbar bin.

Als ich 2014 meine Stelle als Doktorand am IANM angetreten habe, war eines meiner Ziele, möglichst viel über Mathematik zu lernen. Ich habe nicht damit gerechnet, dass die gemeinsamen Mittags- und Kaffeepausen mit derart vielfältigen, unterhaltenden und gewinnbringenden Gesprächen gefüllt sein würden. Für die schöne Zeit und die neuen Sichtweisen in verschiedensten Angelegenheiten bedanke ich mich bei allen meinen (ehemaligen wie aktuellen) Kollegen am IANM! Besonders hervorheben möchte ich Marcel Mikl, der mir durch seine breite Fachkenntnis und angenehme

Diskussionskultur in vielerlei Hinsicht die Augen geöffnet hat. (Außerdem entschuldige ich mich für die Behauptung, dass man weiße Gummibärchen in jedem Fall am Geschmack erkennen kann, und glaube weiterhin fest daran, dass vier Wochen Training für 100 Liegestütze ausreichen...)

Während meiner Zeit am IANM durfte ich mit Daniel Weiß eine Vorlesung über die Programmiersprache Python ausarbeiten und anbieten — Daniel, vielen Dank für den Mut zu diesem (in meinen Augen geglückten) Experiment! Auch möchte ich mich bei den Studenten bedanken, die im Rahmen dieser Veranstaltung und darüber hinaus ihre kreativen Projektideen umgesetzt haben.

Prof. Dr. Tamim Asfour danke ich für die vielen Jahre, die ich während meines Studiums an seinem Institut arbeiten durfte — Tamim, vielen Dank für die gebotenen Möglichkeiten und das entgegengebrachte Vertrauen insbesondere vor dem Vortrag auf der Humanoids! Im selben Atemzug möchte ich mich bei Martin Do und Ludovic Righetti für die angenehme Zusammenarbeit in dieser Zeit bedanken.

Abseits der Arbeit haben meine Freunde maßgeblich dazu beigetragen, mein Leben interessant und angenehm zu gestalten, wofür ich sehr dankbar bin. Besonderer Dank für die vielen gemeinsamen Stunden mit Kaffee im Schlosspark und für die damit verbundenen Diskussionen und Projekte gilt meinem langjährigen Freund Peter Kaiser — Peter, die Entscheidung gemeinsam mit dir nach Karlsruhe zu gehen habe ich nie bereut!

Nicht genügend danken kann ich meiner Freundin Lina Ostermann-Schelleckes — Lina, ohne deine Bereitschaft, monatelang mit einem Gespenst zusammenzuleben und zeitweise die Verantwortung für den Großteil der alltäglichen Aufgaben sowie die Organisation und Planung unserer Reise zu übernehmen, hätte ich es definitiv nicht geschafft!

Abschließend möchte ich meinen Eltern, Uschi Klug-Ernesti und Herbert Ernesti, sowie meinem Bruder, Phillip Ernesti, danken — ihr habt mich immer unterstützt und mich zu dem gemacht, was ich heute bin!

Karlsruhe, im Mai 2018

Johannes Ernesti