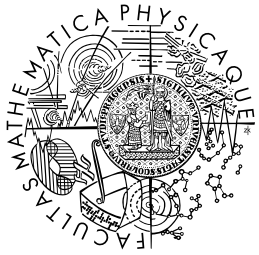

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Bc. Ivan Gálfy

Analysis of the numerical solution of the Forchheimer model

Mathematical Institute of Charles University

Supervisor of the master thesis:	prof. RNDr. Vít Dolejší, Ph.D., DSc.
Study programme:	Mathematics
Study branch:	Mathematical Modelling in Physics and Technology

Prague 2019

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ..... signature of the author

I would like to thank the supervisor of my thesis prof. RNDr. Vít Dolejší, Ph.D., DSc. for professional management, provided resources and the time dedicated during my works on this thesis. I would also like to thank my parents and my family for their continuous support during my studies.

Title: Analysis of the numerical solution of the Forchheimer model

Author: Bc. Ivan Gálfy

Institute: Mathematical Institute of Charles University

Supervisor: prof. RNDr. Vít Dolejší, Ph.D., DSc. , Department of Numerical Mathematics

Abstract: The thesis is dedicated to the study and numerical analysis of the nonlinear flows in the porous media, using general Forchheimer models. In the numerical analysis, the local discontinuous Galerkin method is chosen. The first part of the paper is dedicated to the derivation of the studied equations based on the physical motivation and summarizing the theory needed for the further analysis. Core of the thesis consists of the introduction of the chosen discretization method and the derivation of the main stability and a priory error estimates, optimal for the linear ansatz functions. At the end we present a couple of numerical experiments to verify the results.

Keywords: Numerical analysis , discontinuous Galerkin method , numerical solution , Forcheimer model.

Contents

Introduction **2**

1 Derivation of the model **4**
1.1 Notation 4
1.2 Generalization of the Darcy's law 4
1.3 Properties of the non linear function K 8
1.4 Arising challenges 10

2 Soboloev-Orlicz spaces and N-functions **12**
2.1 Introduction to Soboloev-Orlicz spaces 12
2.2 Relation between K and its associated N-function 17

3 Discretization of the domain and discrete function spaces **24**
3.1 Discretization of the domain 24
3.2 Function spaces 25

4 Auxiliary results **27**
4.1 The global distributional gradient generalization 27
4.2 The local L^2 projection 28
4.3 Scott-Zhang interpolation 31

5 Discontinuous Galerkin formulations **38**
5.1 Local DG formulation 38
5.2 The primal formulation 41

6 A priori stability estimates **43**
6.1 Estimate assuming $u_D^* = 0$ and $\Gamma_N = \emptyset$ 43
6.2 Estimate assuming time independent problem 44

7 A priori error estimates **49**
7.1 Time independent problem 49
7.2 The estimate with the choice $u_D^* = \Pi_{SZ} u$ 55
7.3 The estimate with the choice $u_D^* = u$ 56
7.4 Time dependent problem 58

8 Numerical Examples **61**

Conclusion **69**

Bibliography **70**

List of Figures **72**

Introduction

This paper is dedicated to the modelling and the numerical analysis of nonlinear flows in porous media, which is mostly used in hydrology, oil industry or environmental protection. While Darcy's law is commonly used in modelling of the flow in porous media, it is important to remember that it is derived under very specific assumptions, or by somewhat restricting simplification of the general conservation laws. This has been observed to be problematic in situations with high values of velocity or equivalently, the situations, where the Reynolds number exceeds a certain characteristic value.

These problems led to multiple generalizations of Darcy's law that were intended to capture the nonlinear nature of the flow shown in the experiments. The class of the generalizations studied in this paper are called Forhheimer equations, which in addition to simple Darcy's law $\frac{\mu}{k}u = -\nabla p$ contain polynomial dependencies on the velocity field. These equations can be written as $g(|u|)u = -\nabla p$, where g is a polynomial with positive coefficients and positive exponents of the given degree. Under some additional assumptions on the domain and the studied fluid these can be rewritten as the equation

$$u_t - \nabla \cdot (K(\nabla u)\nabla u) = f,$$

with the appropriate boundary and initial conditions, where K is a nonlinear function with attributes dependent on the polynomial g. Using the properties of K derived in the paper, this leads to a nonstationary quasilinear convection-diffusion problem, which generally degenerates for pressure values approaching infinity, making the analysis somewhat complicated and demands the use of non-standard techniques. This problem will be shown to be similar to the perturbed p-laplace problem with $p \in (1, 2)$.

In the numerical analysis of this equation we chose the discontinuous Galerkin approach, which is similar to standard finite elements methods, but does not prescribe continuous test functions on the edges of the triangulation. One of the advantages of this approach is that we are able to locally adapt the mesh and the polynomial degree of approximation, without impacting the rest of the computational domain. More precisely, the local Discontinuous Galerkin method was chosen, in which the equation is first rewritten as three equations of the first order and numerical fluxes are used to control the the jumps of the discrete solution on the boundaries of the triangulation. This is different to the use of additional terms that preserve the consistency of the discretization as in interior penalty discontinuous Galerkin methods. The main results of this paper consists of a priori stability estimates for the case with simplified boundary conditions and for the stationary case, and the a priori error estimates that show the convergence rate in the special norm $|| \cdot ||_{F,DG}$.

The paper is organised as follows. In chapter 1 we present the derivation of the studied equation from the generalized Darcy's law and show some properties of the nonlinear function K. In Chapter 2 we define Sobolev-Orlicz spaces, and derive some useful properties of N-functions, which can be linked to the nonlinear nature of the problem and are used in the further analysis. Chapter 3 consists of the proper formulation of the studied equations, discretization of the computational

domain and the introduction of the discontinuous spaces used in the local DG formulation. In chapter 4 we present additional results, which are needed in the local DG formulation and further numerical analysis of the problem. Chapter 5 consists of the local DG discretization and the primal formulation of the problem. In chapter 6 we present the a priori estimates for the numerical solution under the special conditions on the boundary, or in the stationary case. Finally in chapter 7 we show the a priori error estimates in both stationary and time dependent case. The paper closes with the numerical experiments in chapter 8.

1. Derivation of the model

1.1 Notation

First we establish some simplifying notation. Throughout the paper we will use in most estimates generic constants usually denoted c or C, meaning they can change from line to line, but never depend on important quantities, such as h the parameter of the mesh or the unknown functions u_h or u. We will also use the symbol for equivalence $\sim$, meaning $f \sim g$, iff there exist constants c_1 and c_2, such that $c_1 g \leq f \leq c_2 g$. It will be useful to simplify the notation for the normed integral over a domain Ω as $\langle f \rangle_\Omega = \fint_\Omega f dx = \frac{1}{|\Omega|} \int_\Omega f dx$, where $|\Omega|$ is the Lebesgue measure of Ω with appropriate dimension.

1.2 Generalization of the Darcy's law

In the first chapter we start with the physical motivation behind the analysed problem and the derivation of the most important equation studied in the thesis. We mostly follow the derivation in [1] and [2] for slightly compressible fluid, isothermal conditions and homogeneous domain.

It is common to describe the viscous fluid laminar flows in porous media by Darcy's law. This model can be derived in several ways, including homogenization, or simplification of the general balance equations governing the flow. Either way, several simplifying assumptions have to be made in order to derive Darcy's law. While it is easier to work with during computations, it proves insufficient to describe situations involving higher velocities and therefore large Reynolds numbers.

One of the ways to deal with this problem is the use of Forchheimer's models, which include nonlinear dependencies between the velocity and the pressure gradient to describe different phenomena like friction between the fluid and the solid in the porous media. The Darcy's law in the general setting can be written as

$$\alpha v = -\Pi \nabla p, \qquad \alpha = \frac{\mu}{k}, \tag{1.1}$$

where v is the velocity field, p is the pressure distribution, μ is the dynamic viscosity of the fluid, k is the permeability of the medium, which can be a function of the spacial variables and Π is a dimensionless normalized positive definite symmetric permeability tensor.

There are three different Forchheimer's laws commonly used to generalize this equation.

- Forchheimer two term law

$$\alpha v + \beta \sqrt{(Bv, v)} v = -\Pi \nabla p, \qquad \beta = \frac{\rho F \Phi}{l^{1/2}}, \tag{1.2}$$

  where ρ is the density of the fluid, F is the Forchheimer's coefficient, Φ is the porosity and B is a positive definite tensor, with bounded entries, which can depend on the spatial variable.

- The Forchheimer power law

$$av + c^n\sqrt{(Bv, v)^{n-1}}v = -\Pi\nabla p, \tag{1.3}$$

  where n is a number from interval $[1, 2]$ and the functions a and c, which are positive and bounded can be found empirically, or they can be taken as $c = (n-1)\beta^{1/2}$ and $a = \alpha$.

- The Forchheimer three term law

$$av + b\sqrt{(Bv, v)}v + c(Bv, v)v = -\Pi\nabla p, \tag{1.4}$$

  where a, b and c are empirical constants.

We can write these equations in the more general form, as follows

$$g(x, |v|_B)v = -\Pi\nabla p. \tag{1.5}$$

Using B the same as before, $g(x, s) > 0$ for $s \geq 0$ and $|v|_B = \sqrt{(Bv, v)}$. If we further assume that the porous media is homogeneous and isotropic and the function $g(s)$ is independent of the spatial variable, meaning that

$$\Pi(x) = I, \qquad B(x) = I, \qquad g(x, |v|_B) = g(|v|), \tag{1.6}$$

we arrive at

$$g(|v|)v = \nabla p. \tag{1.7}$$

By taking the norm of both sides and defining $G(s) \coloneqq g(s)s$ we get

$$g(|v|)\,|v| = |\nabla p|\,, \qquad G(|v|) = |\nabla p|\,. \tag{1.8}$$

In order for this problem to be solvable for $|v|$, we need G to be invertible. To guarantee this we place the following conditions on $g(s)$.

Definition 1.1. *The function $g(s)$ satisfies **G-Conditions** if*

- $g \in C([0, \infty)) \cup C^1((0, \infty))$,
- $g(0) > 0$ *and* $g'(s) \geq 0$ *for all* $s \geq 0$.

For the function G we have $G(0) = 0$ and if $g(s)$ satisfies the G-Conditions, which also means that it is growing on $[0, \infty)$, we also have $G'(s) = g'(s)s + g(s) \geq g(0) > 0$. This implies that G is a one to one mapping of the interval $[0, \infty)$ to itself. Therefore G is invertible on this interval and we can write

$$|v| = G^{-1}(|\nabla p|). \tag{1.9}$$

To check if these conditions on the function $g(s)$ are reasonable within our framework, we can verify if they are compatible with Forchheimer's laws. Under conditions (1.6) the three Forchheimer's laws reduce to

- Forchheimer two term law

$$av + b\,|v|\,v = -\nabla p. \tag{1.10}$$

- The Forchheimer power law

$$av + d\,|v|^{n-1}\,v = -\nabla p. \tag{1.11}$$

- The Forchheimer three term law

$$av + b\,|v|\,v + c\,|v|^2\,v = -\nabla p, \tag{1.12}$$

where a, b, c and d are positive constants. We can write the generalized form of these equations as follows.

$$\sum_{i=0}^{k} a_i\,|v|^{\alpha_i}\,v = a_0\,|v|^{\alpha_0}\,v + a_1\,|v|^{\alpha_1}\,v + \ldots + a_k\,|v|^{\alpha_k}\,v = -\nabla p, \tag{1.13}$$

for $k \geq 0$, positive coefficients $a_i, i = 0, \ldots, k$ and exponents satisfying $0 < \alpha_0 < \alpha_1 < \ldots < \alpha_k$. In this situation $g(s)$ is a polynomial with positive coefficients and positive exponents

$$g(s) = \sum_{i=0}^{k} a_i s^{\alpha_i}. \tag{1.14}$$

These types of functions are called g-Forchheimer polynomials of degree α_k and they trivially satisfy the G-Conditions.

Now we can plug (1.9) into (1.7) and get

$$v = \frac{\nabla p}{g(G^{-1}(|\nabla p|))} = -K(|\nabla p|)\nabla p, \tag{1.15}$$

where $K : [0, \infty) \to [0, \infty)$ is defined by

$$K(\xi) \coloneqq \frac{1}{g(G^{-1}(\xi))}. \tag{1.16}$$

We will further assume that we are working with slightly compressible fluid, which means it has small, but non zero constant compressibility $\frac{1}{\kappa}$ of magnitude between 10^{-5} to 10^{-6}. Gas free oil, or water can serve as examples for slightly compressible fluids. Now we want to use the state equation for slightly compressible fluid, and the continuity equation. In isothermal condition it holds $\rho = \rho(p)$ and the state equations for the fluid reads

$$\frac{1}{\rho}\frac{d\rho}{dp} = \frac{1}{\kappa}, \tag{1.17}$$

where $\frac{1}{\kappa}$ is the compressibility of the fluid and under the previous assumptions $\frac{1}{\kappa} = konst > 0$. If we solve this equation for ρ, we get

$$\rho = \rho_o exp(\frac{p - p_0}{\kappa}), \tag{1.18}$$

with ρ_0 and p_0 being the reference density and reference pressure respectively.

The continuity equation

$$\frac{d\rho}{dt} = -\nabla \cdot (\rho v) \tag{1.19}$$

can be rewritten assuming $\rho = \rho(p)$ as

$$\frac{d\rho}{dp}\frac{dp}{dt} = -\rho\nabla \cdot v - \frac{d\rho}{dp}v \cdot \nabla p. \tag{1.20}$$

Here we can substitute for ρ from the state equation (1.17)

$$\begin{aligned}\frac{d\rho}{dp}\frac{dp}{dt} &= -\kappa\frac{d\rho}{dp}\nabla \cdot v - \frac{d\rho}{dp}v \cdot \nabla p,\\ \frac{dp}{dt} &= -\kappa\nabla \cdot v - v \cdot \nabla p.\end{aligned}$$

Since for most slightly compressible fluids in porous media flows the constant κ is large, i.e of the magnitude between 10^5 and 10^6, the last term in this equation is often dropped and the following reduced equation is studied

$$\frac{dp}{dt} = -\kappa\nabla \cdot v. \tag{1.21}$$

We can substitute for v from the equation (1.15) to obtain

$$\frac{dp}{dt} = \kappa\nabla \cdot (K(|\nabla p|)\nabla p). \tag{1.22}$$

In order to get rid of the constant κ in this equation, we can transition into dimensionless variables.

We can take $\frac{1}{\kappa}, Q$ and $|\Omega|$ as the reference values for the compressibility, the total production of the fluid, and the volume of the domain. Therefore the reference length is $L = |\Omega|^{1/d}$, where $d = 2, 3$ is the dimension and reference time is $T = \frac{|\Omega|}{Q}$. The dimensionless pressure, velocity and time are defined as

$$p^* = \frac{p}{\kappa}, \quad v^* = \frac{L^{d-1}}{Q}v, \quad t^* = \frac{Q}{|\Omega|}t. \tag{1.23}$$

Further the dimensionless non linear function A^* can be defined as

$$A^*(\xi^*) = \frac{\kappa L^{d-2}K(\xi)}{Q} = \frac{\kappa L^{d-2}K(\frac{\kappa}{L}\xi^*)}{Q}. \tag{1.24}$$

The equation (1.15) can be rewritten as

$$\frac{Q}{L^{d-1}}v^* = -K(|\nabla^*(\kappa/Lp^*)|)\nabla^*(\kappa/Lp^*), \tag{1.25}$$

$$v^* = -\frac{L^{d-2}K(|\nabla^*(\kappa/Lp^*)|)}{Q}\nabla^* p^* = -A^*(|\nabla^* p^*|)\nabla^* p^*. \tag{1.26}$$

Similarly the equation (1.22) can be rewritten as

$$\frac{Q}{L^d}\kappa\frac{dp^*}{dt^*} = \nabla^* \cdot \frac{\kappa}{L}K(|\nabla^*(\kappa/Lp^*)|)\nabla^* p^*, \tag{1.27}$$

$$\frac{dp^*}{dt^*} = \nabla^* \cdot \frac{\kappa L^{d-2} K(|\nabla^*(\kappa/Lp^*)|)}{Q} \nabla^* p^* = \nabla^* \cdot (A^*(|\nabla^* p^*|)\nabla^* p^*). \tag{1.28}$$

Finally if we drop the $*$ in the notation we get the equations (1.15), (1.22) in the form without the constant κ

$$v = -K(|\nabla p|)\nabla p, \tag{1.29}$$

$$\frac{dp}{dt} = \nabla \cdot (K(|\nabla p|)\nabla p). \tag{1.30}$$

If we denote the unknown function as u and the right hand side as f, we arrive at the final equation

$$u_t - \nabla \cdot (K(|\nabla u|)\nabla u) = f, \tag{1.31}$$

which will be the subject of the analysis in the further chapters, assuming appropriate initial and boundary conditions.

$$\begin{aligned} &u|_{\delta\Omega_D \times (0,T)} = u_D, \\ &K(|\nabla u|)\nabla u \cdot \boldsymbol{n}|_{\delta\Omega_N \times (0,T)} = g_N, \\ &u(x,0) = u^0(x) \quad x \in \Omega. \end{aligned}$$

On the boundary of the domain $\Omega \times (0,T)$, we chose a combination of the Dirichlet and Neumann boundary conditions.

As we can see, we end up with the nonlinear convection-diffusion problem, where the nonlinearity stems from the function K, for which we can derive some additional properties based on the physical model.

1.3 Properties of the non linear function K

Since K is a nonlinear function, we will need to acquire some estimates on $K(\xi)$ and the derivative $K'(\xi)$, for $\xi \geq 0$, based on the definition $K(\xi) \coloneqq \frac{1}{g(G^{-1}(\xi))}$ and the properties of the polynomial g. We derive some of the properties of the nonlinear function K, which can be found in [1] and [2] with the similar proofs.

Lemma 1.1. *Let the function $g(s)$ satisfy the G-Conditions. Then the function K is well defined, belongs to $C^1([0,\infty))$ and is decreasing. Moreover if we use the notation $s = G^{-1}(\xi)$ for $\xi \geq 0$ (which will be used in the further results as well) we have*

$$K'(\xi) = -K(\xi)\frac{g'(s)}{\xi g'(x) + g^2(s)} \leq 0. \tag{1.32}$$

Proof. Since $g(s)$ satisfies the G-Conditions, G^{-1} is well defined and therefore also K is well defined. Straightforward calculation using the chain rule gives us

$$K(\xi) = \frac{1}{g(G^{-1}(\xi))},$$
$$K'(\xi) = -\frac{1}{g^2(s)} g'(s) \frac{1}{G'(s)} = -\frac{1}{g(G^{-1}(\xi)} \frac{g'(s)}{g(s)} \frac{1}{g'(s)s + g(s)}$$
$$= -K(\xi) \frac{g'(s)}{\xi g'(s) + g^2(s)} \leq 0.$$

Since $\xi \geq 0, g'(s) \geq (0)$ and $K(\xi) \geq 0$, the last inequality holds and K is decreasing. □

In order to acquire certain monotone properties for K that will be useful in the stability and error estimates down the road, we introduce an additional condition on $g(s)$.

Definition 1.2. *The function $g(s)$ as defined previously satisfies the **Lambda-Condition**, if there exists $\lambda > 0$ such that for all $s > 0$*

$$g(s) \geq \lambda s g'(s). \tag{1.33}$$

Note that g-Forchheimer polynomials of degree α_k satisfy this condition with $\lambda = \frac{1}{\alpha_k}$.

Lemma 1.2. *Let $g(s)$ satisfy the G-Condition and Lambda-Condition, then*

$$-\frac{1}{\lambda + 1} \frac{K(\xi)}{\xi} \leq K'(\xi) \leq 0. \tag{1.34}$$

Proof. If $g'(s) = 0$, then $K'(\xi) = 0$ and the inequality holds. Otherwise we can use the result from the previous lemma and the Lambda-Condition

$$K'(\xi) = -K(\xi) \frac{g'(s)}{\xi g'(s) + g^2(s)} \geq -K(\xi) \frac{g'(s)}{\xi g'(s) + g(s)\lambda s g'(s)}$$
$$= -K(\xi) \frac{g'(s)}{\xi g'(s) + \xi \lambda g'(s)} = -\frac{1}{\lambda + 1} \frac{K(\xi)}{\xi}.$$

□

Lemma 1.3. *Let the function $g(s)$ be a g-Forchheimer polynomial, then K satisfies the inequalities*

$$\frac{C_0}{(1+\xi)^\alpha} \leq K(\xi) \leq \frac{C_1}{(1+\xi)^\alpha}, \quad \xi \geq 0, \tag{1.35}$$

for $\alpha = \frac{\alpha_k}{\alpha_k + 1} \in [0, 1)$ and C_0, C_1 being positive constants.

Proof. For $x \geq 0$, $b_i \geq 0$, $i = 0, \ldots k$ and $0 \leq \beta_0 < \ldots < \beta_k$,we can use the general inequalities

$$\sum_{i=0}^{k} b_i x^{\beta_i} = b_0 x^{\beta_0} + b_1 x^{\beta_1} + \ldots + b_k x^{\beta_k} \leq C_2 (1 + x)^{\beta_k},$$
$$\sum_{i=0}^{k} b_i x^{\beta_i} = b_0 x^{\beta_0} + b_1 x^{\beta_1} + \ldots + b_k x^{\beta_k} \geq C_3 (1 + x)^{\beta_k}.$$

This implies

$$\xi + 1 = g(s)s + 1 = 1 + a_0 s + \ldots + a_k s^{\alpha_k + 1} \sim (1+s)^{\alpha_{k+1}}$$
$$\implies (1+s) \sim (\xi+1)^{\frac{1}{\alpha_k+1}},$$
$$g(s) = a_0 + \ldots + a_k s^{\alpha_k} \sim (1+s)^{\alpha_k},$$
$$K(\xi) = \frac{1}{g(s)} \sim \frac{1}{(1+s)^{\alpha_k}} \sim \frac{1}{(1+\xi)^{\frac{\alpha_k}{\alpha_k+1}}}.$$

□

Let us denote $p \coloneqq 2 - \alpha$. Then $p \in (1,2)$ and based on lemma 1.3 we have for $P \in \mathbb{R}^d$

$$K(|P|)P \sim (1+|P|)^{p-2}P. \tag{1.36}$$

We can write the weak formulation of the problem (1.31) for $v \in W_0^{1,p}(\Omega)$ and $t \in (0,T)$

$$\int_\Omega u_t v dx + \int_\Omega K(|\nabla u|)\nabla u \cdot \nabla v dx = \int_\Omega f v dx. \tag{1.37}$$

$$\begin{aligned} &u|_{\delta\Omega_D \times (0,T)} = u_D, \\ &K(|\nabla u|)\nabla u \cdot \boldsymbol{n}|_{\delta\Omega_N \times (0,T)} = g_N, \\ &u(x,0) = u^0(x) \quad x \in \Omega. \end{aligned}$$

Assuming that $f \in C([0,T]; L^2(\Omega))$, $u^0 \in W^{1,2}(\Omega) \cap W^{1,p}(\Omega)$, according to the results in [2] and [9], there exists a weak solution $u \in L^2_{loc}(0,\infty; W^{2,2}(\Omega))$ to this problem, which satisfies $u_t \in L^2_{loc}(0,\infty; W^{1,2}(\Omega))$ and $u|_{\delta\Omega_D \times (0,T)} = u_D$.

1.4 Arising challenges

The later chapters of the thesis are dedicated to the numerical analysis of the local discontinuous Galerkin method for the solution of (1.31). More precisely we will be concerned with the stability and error estimates of the method.

If we managed to get $c_1|\nabla u| \leq K(|\nabla u|)\nabla u \leq c_2|\nabla u|$, the stability estimates would resemble the case of the linear equation. The upper estimate follows from lemma 1.3, but the lower estimate does not hold. The best estimate we can use is $K(|\nabla u|)\nabla u \sim (1+|\nabla u|)^{p-2}\nabla u$, which resembles the perturbed p-Laplace problem, with the main difference being that we will have to use the properties of K to estimate the derivative.

In the error analysis of the chosen numerical method, we will also need to estimate for $P, Q \in \mathbb{R}^d$ the terms $(K(|P|)P - K(|Q|)Q) \cdot (P - Q)$ and $|K(|P|)P - K(|Q|)Q|$. Ideally we would look for the estimates of the type

$$(K(|P|)P - K(|Q|)Q) \cdot (P - Q) \geq C_1|P - Q|^2, \tag{1.38}$$

$$|K(|P|)P - K(|Q|)Q| \leq C_2|P - Q|. \tag{1.39}$$

This with the Cauchy-Schwarz inequality would imply $(K(|P|)P - K(|Q|)Q) \cdot (P - Q) \sim |P - Q|^2$ and $|K(|P|)P - K(|Q|)Q| \sim |P - Q|$, which would allow us to work with basic Hilbert-Sobolev spaces, again resembling the case of the linear equation.

The second estimate holds as proven in [4, Lemma 2.4], which we will not show here, since we will need stronger result later. The best estimate of the first term we can get is in the following lemma, using similar steps as in [2, Lemma III.6].

Lemma 1.4. *Let $g(s)$ satisfy the G-Conditions and the Lambda-Conditions, then for $P, Q \in \mathbb{R}^d$*

$$(K(|P|)P - K(|Q|)Q) \cdot (P - Q) \geq K(max\{|P|, |Q|\})|P - Q|^2 \frac{\lambda}{1+\lambda}. \quad (1.40)$$

Proof. Let us first consider that the zero vector does not belong to the line segment connecting P and Q. Define $\gamma(t) = (tP + (1-t)Q)$,for $t \in [0,1]$. We also define $h(t) = (K(|\gamma(t)|)\gamma(t)) \cdot (P - Q)$, for $t \in [0,1]$. Then using the Mean Value Theorem, we have $t_0 \in [0,1]$ and $P_0 = \gamma(t_0) \neq 0$, such that

$$\begin{aligned}(K(|P|)P - K(|Q|)Q) \cdot (P - Q) &= h(1) - h(0) = h'(t_0) \\ &= (\nabla(K(|P_0|)P_0)(P - Q)) \cdot (P - Q)\end{aligned}$$

Here in the compound derivative we can use the result from lemma 1.1, where $s = G^{-1}(|P_0|)$.

$$\begin{aligned}&= K(|P_0|)|P - Q|^2 - K(|P_0|)\frac{g'(s)}{|P_0|}\frac{\sum_{i,j} P_{0i}P_{0,j}(P_j - Q_j)(P_i - Q_i)}{|P_0|g'(s) + g^2(s)} \\ &= K(|P_0|)|P - Q|^2 - K(|P_0|)\frac{g'(s)}{|P_0|}\frac{|P_0 \cdot (P - Q)|^2}{|P_0|g'(s) + g^2(s)}\end{aligned}$$

Applying the Cauchy-Schwarz inequality to $|P_0 \cdot (P - Q)|^2$ and the Lambda-Condition we arrive at

$$\begin{aligned}&\geq K(|P_0|)|P - Q|^2(1 - \frac{|P_0|g'(s)}{|P_0|g'(s) + g^2(s)}) \\ \geq K(|P_0|)|P - Q|^2\frac{\lambda}{1+\lambda} &\geq K(max\{|P|, |Q|\})|P - Q|^2\frac{\lambda}{1+\lambda},\end{aligned}$$

since K is decreasing.

In case that origin lies on the line segment connecting P and Q, replace Q by $Q_\epsilon \neq 0$, such that this is not the case and $Q_\epsilon \to 0$ for $\epsilon \to 0$. Then we can apply the result on P and Q_ϵ and let $\epsilon \to 0$.

□

The proven estimate is weaker, than what we would like to have and also does not allow us to estimate $|K(|P|)P - K(|Q|)Q|$ from the bottom. This will make the error analysis significantly more complicated than the linear case, forcing us to work in different function spaces, which are generally not Hilbert spaces.

In order to achieve estimates of a similar form as mentioned here, we will have to introduce some additional theory, mainly the theory of N-functions and Orlicz spaces.

2. Soboloev-Orlicz spaces and N-functions

As explained at the end of chapter one, we are not able to achieve the estimates, we would want

$$c_1|P| \leq K(|P|)P \leq c_2|P|,$$
$$(K(|P|)P - K(|Q|)Q) \cdot (P - Q) \geq C_1|P - Q|^2,$$
$$|K(|P|)P - K(|Q|)Q| \leq C_2|P - Q|,$$

for $P, Q \in \mathbb{R}^d$.

In this chapter we will try to derive similar estimates, using the theory of N-functions, where instead of constants c_1, c_2, C_1, C_2, the estimators will depend on a nonlinear N-function φ, which will be defined later. The estimates we aim for will look like

$$K(|P|)P \sim \varphi'(|P|)\frac{P}{|P|},$$
$$(K(|P|)P - K(|Q|)Q) \cdot (P - Q) \geq c\varphi'_{|P|}(|P - Q|)|P - Q|,$$
$$|K(|P|)P - K(|Q|)Q| \leq c\varphi'_{|P|}(|P - Q|),$$

for $P, Q \in \mathbb{R}^d$.

To achieve this we will need to use the theory Sobolev-Orlicz spaces and shifted N-functions. We start by introducing a collection of needed preliminary results based on the works [7], [10], [11] and mainly [13].

2.1 Introduction to Soboloev-Orlicz spaces

Definition 2.1. *A function $\psi : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ is called an $N - function$ if it is convex, continuous, positive on $\mathbb{R}^{>0}$, $\psi(0) = 0$ and*

$$lim_{t\to 0}\frac{\psi(t)}{t} = 0, \qquad lim_{t\to\infty}\frac{\psi(t)}{t} = \infty. \tag{2.1}$$

The definition of ψ implies existence of ψ' the right derivative of ψ, which is continuous, nondecreasing, positive on $\mathbb{R}^{>0}$ and

$$\psi'(0) = 0, \qquad lim_{t\to\infty}\psi'(t) = \infty. \tag{2.2}$$

This allows us to represent ψ as

$$\psi(t) = \int_0^t \psi'(s)ds. \tag{2.3}$$

We can also define the right inverse of ψ' as $(\psi')^{-1} : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ satisfying

$$(\psi')^{-1} = sup\{s \in \mathbb{R}^{\geq 0} | \psi'(s) \leq t\}. \tag{2.4}$$

Assuming ψ' is strictly monotone, $(\psi')^{-1}$ is reduced to the normal inverse function. The N-functions, in which we will be interested throughout this paper will have this property, so we can use the normal inverse function from now on.

Definition 2.2. *For an N-function ψ we define the associated Orlicz space $L^{\psi}(\Omega)$ and the Orlicz-Sobolev space $W^{1,\psi}(\Omega)$, where $f \in L^{\psi}(\Omega)$, if $\int_{\Omega} \psi(|f|)dx < \infty$ and $f \in W^{1,\psi}(\Omega)$ if $f, \nabla f \in L^{\psi}(\Omega)$.*

$L^{\psi}(\Omega)$ is equipped with the norm $\|f\|_{\psi} = inf\{\lambda > 0; \int_{\Omega} \psi(\frac{f}{\lambda})dx \leq 1\}$, and $W^{1,\psi}(\Omega)$ is equipped with the norm $\|f\|_{\psi} + \|\nabla f\|_{\psi}$.

Basic theory of Sobolev-Orlicz spaces implies that $L^{\psi}(\Omega)$ and $W^{1,\psi}(\Omega)$ are Banach spaces.

Definition 2.3. *For the N-function ψ we define the complementary function ψ^* as*

$$\psi^*(t) = \int_0^t (\psi')^{-1}(s)ds. \tag{2.5}$$

The complementary function can be equivalently defined by

$$\psi^*(t) = sup\{st - \psi(s) | s \in \mathbb{R}^{\geq 0}\}. \tag{2.6}$$

Definition 2.4. *The $N - function$ ψ satisfies $\Delta_2 - condition$ if $\psi(2t) \leq C\psi(t)$ for all $t \geq 0$, where C is a constant. The $\Delta_2 - constant$ is the smallest C satisfying this property.*

Note that since ψ is increasing, if it satisfies $\Delta_2 - condition$, then

$$\psi(t) \sim \psi(2t).$$

From the representation (2.3) and the fact that ψ' is nondecreasing we have

$$\psi(t) \leq t\psi'(t), \qquad \psi(2t) \geq \int_t^{2t} \psi'(s)ds \geq t\psi(t). \tag{2.7}$$

This together with ψ satisfying the $\Delta_2 - condition$ implies

$$\psi(t) \sim t\psi'(t). \tag{2.8}$$

From which we can automatically deduce that if ψ satisfies the $\Delta_2 - condition$, then ψ' also satisfies the $\Delta_2 - condition$ and vice-versa. We also get the equivalence

$$(\psi')^{-1}(\frac{\psi(t)}{t}) \sim t. \tag{2.9}$$

Using the first definition of ψ^* we have

$$\psi^*(\frac{\psi(t)}{t}) = \int_0^{\frac{\psi(t)}{t}} (\psi')^{-1}(s)ds \leq \frac{\psi(t)}{t}(\psi')^{-1}(\frac{\psi(t)}{t}) \leq \psi(t).$$

Similarly we can get

$$\psi(t) \leq \psi^*(\frac{2\psi(t)}{t}).$$

Putting this together we have for ψ^* satisfying the $\Delta_2 - condition$

$$\psi(t) \sim \psi^*(\frac{\psi(t)}{t}). \tag{2.10}$$

Also if ψ^* satisfies the $\Delta_2 - condition$ we have

$$\psi^*(\psi'(t)) \sim \psi'(t)(\psi^*)'(\psi'(t)) = \psi'(t)t \sim \psi(t). \tag{2.11}$$

From the second definition of ψ^* we automatically get the following Young-type inequality, which can be extended, assuming ψ and ψ^* satisfy the $\Delta_2 - condition$.

Lemma 2.1. *let ψ be and N-function and ψ^* be its conjugate, then for all $s, t \geq 0$*

$$ts \leq \psi(t) + \psi^*(s). \tag{2.12}$$

If further ψ and ψ^* satisfy the $\Delta_2 - condition$, then for $\epsilon \in (0, 1)$

$$ts \leq \epsilon\psi(t) + c_\epsilon\psi^*(s). \tag{2.13}$$

This inequality is very useful in the numerical analysis and will be heavily used throughout the paper. We will also need to introduce the shifted N-functions, explaining the notation $\varphi'_{|P|}$ in the introduction to this chapter.

Definition 2.5. *For the N-function ψ we define the set of shifted N-functions $\{\psi_a\}_{a\geq 0}$ by*

$$\psi_a = \int_0^t \psi'_a(s)ds, \quad \psi'_a(t) = \psi'(a+t)\frac{t}{a+t}. \tag{2.14}$$

From definition, it is obvious that ψ_a is also an N-function. Furthermore if ψ satisfies the $\Delta_2 - condition$, then also ψ_a satisfies the $\Delta_2 - condition$

$$\psi'_a(2t) = \frac{\psi'(a+2t)2t}{a+2t} \leq \frac{\psi'(2a+2t)2t}{a+t} \leq c\psi'_a(t). \tag{2.15}$$

Lemma 2.2. *Let ψ be an N-function satisfying the $\Delta_2 - condition$, then for all $P, Q \in \mathbb{R}^d$*

$$\psi'_{|P|}(|P-Q|) \sim \psi'_{|Q|}(|P-Q|), \tag{2.16}$$

$$\psi_{|P|}(|P-Q|) \sim \psi_{|Q|}(|P-Q|). \tag{2.17}$$

Proof. For $P = Q$ the assertion is trivial. We can assume that $|P-Q| > 0$. Since $(|Q| + |P-Q|) \sim (|P| + |P-Q|)$,

$$\begin{aligned}\psi'_{|P|}(|P-Q|) &= \psi'(|P| + |P-Q|)\frac{|P-Q}{|P| - |P-Q|} \\ \sim \psi'(|Q| + |P-Q|)&\frac{|P-Q}{|P| - |P-Q|} = \psi'_{|Q|}(|P-Q|).\end{aligned}$$

□

Lemma 2.3. *Let ψ be an N-function and $M \in \mathbb{N}$. Then for all $t \geq a(2^M - 1)^{-1}$*

$$\frac{1}{2^M}\psi'(t) \leq \psi'_a(t) \leq \psi'(2^M t). \tag{2.18}$$

Proof.

$$\psi_a'(t) = \frac{\psi'(a+t)}{a+t}t \leq \frac{\psi'(2^M t)}{t}t = \psi'(2^M t),$$
$$\psi_a'(t) = \frac{\psi'(a+t)}{a+t}t \geq \frac{\psi'(t)}{2^M t}t = \frac{1}{2^M}\psi'(t).$$

□

Lemma 2.4. *Let ψ be an N-function and $M \in \mathbb{N}$. Then for all $0 \leq t \leq a(2^M - 1)^{-1}$, it holds*

$$\frac{1}{2^M}\frac{\psi'(a)}{a}t \leq \psi_a'(t) \leq \frac{\psi'(2^M a)}{a}t. \tag{2.19}$$

Proof.

$$\psi_a'(t) = \frac{\psi'(a+t)}{a+t}t \leq \frac{\psi'(2^M a)}{a}t,$$
$$\psi_a'(t) = \frac{\psi'(a+t)}{a+t}t \geq \frac{\psi'(a)}{2^M a}t.$$

□

Lemma 2.5. *Let ψ and ψ^* be N-fucntions satisfying the $\Delta_2 -$ condition. Then for all $a, t \geq 0$ it holds*

$$(\psi^*)'_{\psi'(a)}(\psi_a'(t)) \leq ct \tag{2.20}$$

Proof. For $t = 0$ or $a = 0$ the inequality is trivial. First let $t \geq a > 0$. Then $\psi_a'(t) \leq \psi'(2t)$.

$$(\psi^*)'_{\psi'(a)}(\psi_a'(t)) \leq (\psi^*)'_{\psi'(a)}(\psi'(2t)) \leq (\psi^*)'(2\psi'(2t)) \leq ct.$$

Where we used lemma 2.3 with $M = 1$ twice. For $0 < t \leq \frac{a}{\Delta_2(\psi')}$ we have $\psi_a'(t) \leq \frac{\psi'(2a)t}{a}$. Therefore

$$(\psi^*)'_{\psi'(a)}(\psi_a'(t)) \leq (\psi^*)'_{\psi'(a)}(\frac{\psi'(2a)t}{a}) \leq \frac{(\psi^*)'(2\psi'(a))}{\psi'(a)}\psi'(a)\frac{t}{a} \leq ct.$$

Where we use lemma 2.4 with $M = 1$. Finally for $\frac{a}{\Delta_2(\psi')} \leq t \leq a$, it holds $\psi_a'(t) \leq \psi_a'(a)$ and therefore

$$(\psi^*)'_{\psi'(a)}(\psi_a'(t)) \leq (\psi^*)'_{\psi'(a)}(\psi_a'(a)) \leq ca \leq ct.$$

□

Lemma 2.6. *Let ψ and ψ^* be N-functions satisfying the $\Delta_2 -$ condition. Then for all $a, t \geq 0$ it holds*

$$(\psi^*)'_{\psi'(a)}(\psi_a'(t)) \geq ct. \tag{2.21}$$

Proof. For $t \geq a$, $\psi'(t) \geq \psi'(a)$ and we have

$$\begin{aligned}(\psi^*)'_{\psi'(a)}(\psi_a'(t)) &\geq (\psi^*)'_{\psi'(a)}(\frac{1}{2}\psi'(t)) \geq c(\psi^*)'_{\psi'(a)}(\psi'(t)) \\ &\geq c(\psi^*)'(\psi'(t)) = ct.\end{aligned}$$

Where we used lemma 2.3 with $M = 1$ twice. For $t \leq a$, $\frac{\psi'(a)t}{a} \leq \psi'(a)$ and we have

$$(\psi^*)'_{\psi'(a)}(\psi'_a(t)) \geq (\psi^*)'_{\psi'(a)}(\frac{\psi'(a)t}{2a}) \geq c(\psi^*)'_{\psi'(a)}(\frac{\psi'(a)t}{a})$$
$$\geq c\frac{(\psi^*)'(\psi'(a))}{2\psi'(a)}\frac{\psi'(a)t}{a} = ct.$$

Where we used lemma 2.4 with $M = 1$ twice. □

Lemma 2.7. *Let ψ and ψ^* be N-functions satisfying the $\Delta_2 -$ condition. Then for all $a, t \geq 0$ it holds*

$$((\psi_a)^*)'(t) \sim (\psi^*)'_{\psi'(a)}(t). \tag{2.22}$$

Proof. First we deal with the inequality $\leq$. Taking $((\psi_a)^*)'(t)$ instead of t in the lemma 2.6 we have

$$((\psi_a)^*)'(t) \leq c(\psi^*)'_{\psi'(a)}(\psi'_a(((\psi_a)^*)'(t))) = c(\psi^*)'_{\psi'(a)}(t).$$

Now for the inequality $\geq$, we take $((\psi_a)^*)'(t)$ instead of t in lemma 2.5.

$$(\psi^*)'(t) \leq (\psi^*)'_{\psi'(a)}(\psi'_a(((\psi_a)^*)'(t)) \leq c((\psi_a)^*)'(t).$$

□

Lemma 2.8. *Let ψ be an N-function satisfying the $\Delta_2 -$ condition, then for all $P, Q \in \mathbb{R}^d$ and $t \geq 0$*

$$\psi'_{|P|}(t) \leq c\psi'_{|Q|}(t) + \psi'_{|P|}(|P - Q|). \tag{2.23}$$

Proof. If $|P - Q| \geq t$, then

$$\psi'_{|P|}(t) \leq \psi'_{|P|}(|P - Q|).$$

In the other case $|P-Q| \leq t$, the following inequalities hold. $0 \leq \frac{1}{2}(|Q|+t) \leq |P| + t \leq 2(|Q| + t)$ and therefore

$$\psi'_{|P|}(t) = \frac{\psi'(|P| + t)}{|P| + t}t \leq \frac{\psi'(2(|Q| + t))}{\frac{1}{2}|P| + t}t \leq c\frac{\psi'(|Q| + t)}{|P| + t}t = c\psi'_{|Q|}(t).$$

□

Lemma 2.9. *Let ψ be an N-functions such that ψ and ψ^* satisfy the $\Delta_2 -$ condition, then for $\delta \in (0, 1)$, $P, Q \in \mathbb{R}^d$ and $t \geq 0$ we have*

$$\psi_{|P|} \leq c_\delta\psi_{|Q|}(t) + \delta\psi_{|Q|}(|P - Q|). \tag{2.24}$$

Proof. Using the previous lemma, we have

$$\psi_{|P|}(t) \leq \psi'_{|P|}(t)t \leq c\psi'_{|Q|}(t)t + c\psi'_{|Q|}(|P - Q|)t = I_1 + I_2.$$

Afterwards we use the Young-type inequality 2.1

$$I_1 \leq c(\psi_{|Q|})^*(\psi'_{|Q|}(t)) + \psi_{|Q|}(t) \leq c\psi_{|Q|}(t),$$

$$I_2 \leq (\psi_{|Q|})^*(\delta c\psi'_{|Q|}(|P-Q|)) + c\psi_{|Q|}(t) \leq \delta\psi_{|Q|}(|P-Q|) + c_\delta\psi_{|Q|}(t).$$

□

In order to derive further results, we will demand an additional property from our N-functions.

Definition 2.6. *An N-function ψ satisfies the continuity-condition, if for all $s, t \geq 0$, there exits a constant C, such that the following inequality holds.*

$$|\psi'(s+t) - \psi'(t)| \leq C\psi'_t(s). \tag{2.25}$$

Lemma 2.10. *Let ψ be an N-functions such that ψ and ψ^* satisfy the Δ_2 − condition and ψ satisfies the continuity-condition, then for $P, Q \in \mathbb{R}^d$ and $t \geq 0$ the following inequality holds*

$$((\psi_{|P|})^*)'(t) \leq c(((\psi_{|Q|})^*)'(t) + ||P| - |Q||). \tag{2.26}$$

Proof. First note that the result of lemma 2.8 holds for the $||P| - |Q||$ in place of $|P - Q|$ in the last term on the right hand side of the inequality. This can be derived by using $P = |P|R$, $Q = |Q|R$, with $|R| = 1$ in the said lemma. In the following set of inequalities we use in order, lemma 2.7, lemma 2.8, the continuity-condition for ψ and lemma 2.7 again.

$$\begin{aligned}
((\psi_{|P|})^*)'(t) &\leq c(\psi^*)'_{\psi'(|P|)}(t) \\
&\leq c((\psi^*)'_{\psi'(|Q|)}(t)) + (\psi^*)'_{\psi'(|P|)}(|\psi'(|P|) - \psi'(|Q|)|) \\
&\leq c((\psi^*)'_{\psi'(|Q|)}(t)) + (\psi^*)'_{\psi'(|P|)}(C\psi'_{|P|}(||P| - |Q||)) \\
&\leq c(((\psi_{|Q|})^*)'(t) + ((\psi_{|P|})^*)'(\psi'_{|P|}(||P| - |Q||))) \\
&= c(((\psi_{|Q|})^*)'(t) + ||P| - |Q|| \leq c(((\psi_{|Q|})^*)'(t) + |P - Q|).
\end{aligned}$$

□

Lemma 2.11. *Under the assumptions of lemma 2.10, for $\delta \in (0, 1)$, $P, Q \in \mathbb{R}^d$ and $t \geq 0$ we have*

$$(\psi_{|P|})^* \leq c_\delta(\psi_{|Q|})^*(t) + \delta\psi_{|P|}(|P - Q|). \tag{2.27}$$

Proof. The proof of this lemma follows the same steps as the proof of lemma 2.9, but lemma 2.10 is used instead of lemma 2.8 in the first step. □

2.2 Relation between K and its associated N-function

First, for the simplicity of the notation, we denote for $P \in \mathbb{R}^d$

$$A(P) = K(|P|)P, \tag{2.28}$$

since this is the nonlinear function at the core of our equations.

Taking note of lemma 1.3 and setting $p = 2 - \alpha \in (1, 2)$ we define the N-function associated with A, for $t \geq 0$ as

$$\begin{aligned} \varphi(t) &= \int_0^t (1+s)^{p-2} s ds, \\ \varphi'(t) &= (1+t)^{p-2} t. \end{aligned} \tag{2.29}$$

Using lemma 1.3 we can express the relation between A and φ in the following equivalence

$$A(\nabla u) \sim \varphi'(|\nabla u|) \frac{\nabla u}{|\nabla u|}. \tag{2.30}$$

Let us check, if φ is a well defined N-function that satisfies all the requirements defined in the first section. Function φ is clearly an N-function with a strictly increasing derivative, which implies existence of $(\varphi')^{-1}$ as a conventional inverse of φ'. Note that with this definition $L^{\varphi}(\Omega)$ is isomorphic to $L^p(\Omega)$ and $W^{1,\varphi}(\Omega)$ is isomorphic to $W^{1,p}(\Omega)$, with the constant depending only on p. φ also satisfies the previously defined continuity-condition, defined in (2.6). For $t, s \geq 0$, it holds

$$\begin{aligned} |\varphi'(s+t) - \varphi'(t)| &= |(1+t+s)^{p-2}(t+s) - (1+t)^{p-2} t| \\ &\leq |(1+t+s)^{p-2}(t+s) - (1+t+s)^{p-2} t| \\ &= |(1+t+s)^{p-2} s| = \frac{\varphi'(s+t)}{s+t} s = \varphi_t'(s). \end{aligned} \tag{2.31}$$

Since $\varphi''(t) = (p-2)(1+t)^{p-3} t + (1+t)^{p-2}$ and $min\{1, p-1\}(1+t)^{p-2} \leq \varphi''(t) \leq max\{1, p-1\}(1+t)^{p-2}$, φ also satisfies

$$\varphi''(t) t \sim \varphi'(t). \tag{2.32}$$

The $\Delta_2 - condition$ for $\varphi(t)$ is also satisfied with $\Delta_2(\varphi) \leq c 2^{max\{2,p\}}$. The conjugate function φ^* satisfies for q such that $\frac{1}{p} + \frac{1}{q} = 1$

$$\varphi^*(t) \sim (1+t)^{q-2} t^2, \tag{2.33}$$

and therefore $\Delta_2(\varphi^*) \leq c 2^{max\{2,q\}}$.

Concerning the shifted versions of φ and φ^* we have for $a \geq 0$

$$\begin{aligned} \varphi_a(t) &\sim (1+a+t)^{p-2} t^2, \\ (\varphi_a)^*(t) &\sim ((1+a)^{p-1} + t)^{q-2} t^2. \end{aligned} \tag{2.34}$$

Therefore we have $\Delta_2(\varphi_a) \leq c 2^{max\{2,p\}}$ and $\Delta_2((\varphi_a)^*) \leq c 2^{max\{2,q\}}$, for all $a \geq 0$. Note that these constants are independent of a.

Lemma 2.12. *Let the function $g(s)$ be a g-Forchheimer polynomial, then A satisfies the inequality*

$$(A(P) - A(Q)) \cdot (P - Q) \geq c \varphi'_{|P|}(|P - Q|)|P - Q|. \tag{2.35}$$

Proof. Using the definition of shifted N-function

$$\begin{aligned} \varphi'_{|P|}(|P-Q|)|P-Q| &= \varphi'(|P| + |P-Q|) \frac{|P-Q|^2}{|P| + |P-Q|} \\ &= (1 + |P| + |P-Q|)^{p-2} |P-Q|^2. \end{aligned}$$

Also since K is decreasing and $|P-Q| \geq |Q| - |P|$ we have

$$K(max\{|P|,|Q|\}) \geq K(|P| + |P-Q|) \geq c(1 + |P| + |P-Q|)^{p-2},$$

where we used lemma 1.3. Combining this with the result of lemma 1.4 we reach

$$(K(|P|)P - K(|Q|)Q) \cdot (P-Q) \geq c\varphi'_{|P|}(|P-Q|)|P-Q|.$$

□

Lemma 2.13. *Let Ψ be and N-function such that Ψ and Ψ^* satisfy Δ_2-condition, then for $P, Q \in \mathbb{R}^d$*

$$\frac{\psi'(|P| + |Q|)}{|P| + |Q|} \sim \int_0^1 \frac{\psi'(tP + (1-t)Q)}{tP + (1-t)Q} dt. \tag{2.36}$$

Proof. In order to avoid having to introduce additional theory, which does not relate to any other part of the paper, we refer this proof to [13, Lemma 6.6]. □

Lemma 2.14. *Let A be the previously defined nonlinear function and φ its associated N-function, then*

$$|A(P) - A(Q)| \leq c\varphi'_{|P|}(|P-Q|). \tag{2.37}$$

Proof. First we calculate the following compound derivative, using similar steps as in lemma 1.2. Assuming $|P| \neq 0$,

$$\begin{aligned}
\frac{\partial}{\partial P_j}(K(|P|)P_i) &= K(|P|)(\delta_{i,j} - \frac{g'(s)}{|P|}\frac{P_iP_j}{|P|g'(s) + g^2(s)}) \\
&\geq K(|P|)(\delta_{i,j} - \frac{g'(s)}{|P|}\frac{P_iP_j}{|P|g'(s) + g(s)\lambda s g'(s)}) \\
&\geq K(|P|)(\delta_{i,j} - \frac{1}{|P|}\frac{P_iP_j}{|P| + \lambda|P|}) \geq -cK(|P|).
\end{aligned}$$

Taking the norm

$$\left|\frac{\partial}{\partial P_i}(K(|P|)P_j)\right| \leq cK(|P|) \leq c(1 + |P|)^{p-2} = c\frac{\varphi'(|P|)}{|P|},$$

and using the following equality we have

$$\begin{aligned}
K(|P|)P_j - K(|Q|)Q_j &= \int_0^1 \frac{\partial(K(|tP + (1-t)Q|)|tP_j + (1-t)Q_j|)}{\partial t} dt \\
&= \sum_{i=1}^d \int_0^1 \frac{\partial(K(|tP + (1-t)Q|)|tP_j + (1-t)Q_j|)}{\partial P_i}(P_i - Q_i) dt.
\end{aligned}$$

Finally using the estimate for the derivative, lemma 2.13, the inequality $\frac{1}{2}(|P| + |Q|) \leq |P| + |P-Q| \leq 2(|P| + |Q|)$ and the definition of $\varphi'_{|P|}$, we arrive at

$$|K(|P|)P - K(|Q|)Q| \leq c\int_0^1 \frac{\varphi'(|tP+(1-t)Q|)}{|tP+(1-t)Q|}dt|P-Q|$$
$$\leq c\frac{\varphi'(|P|+|Q|)}{|P|+|Q|}|P-Q| \leq c\frac{\varphi'(|P|+|P-Q|)}{|P|+|P-Q|}|P-Q| = c\varphi'_{|P|}(|P-Q|).$$

□

Note that most of the proofs presented here followed the steps in [13], which treats the similar problem with $A(P)$ directly equal to $\varphi'(P)\frac{P}{|P|}$, but in this particular lemma [13, Lemma 6.7], we had to estimate the derivative differently and use the properties of K from the first chapter.

Using together lemma 2.12 and lemma 2.14 and Cauchy-Schwarz inequality we get an important result for $P, Q \in \mathbb{R}^d$

$$\begin{aligned} |A(P) - A(Q)| &\sim \varphi'_{|P|}(|P-Q|), \\ (A(P) - A(Q)) \cdot (P-Q) &\sim \varphi'_{|P|}(|P-Q|)|P-Q|. \end{aligned} \tag{2.38}$$

We finally achieved the estimates outlined in the beginning of this section. Lastly we introduce the pair of nonlinear functions F and F^*, which will serve us in expressing the error of the studied DG method in the proper norm. We define for $P \in \mathbb{R}^d$ functions F and F^* related to A as follows

$$\begin{aligned} F(P) &= (1+|P|)^{p-2/2}P, \\ F^*(P) &= (1+|P|)^{q-2/2}P. \end{aligned} \tag{2.39}$$

The function F has an associated N-function ϕ, where

$$\phi(t) = \int_0^t \phi'(s)ds, \qquad \phi'(t) = \sqrt{\varphi'(t)t}. \tag{2.40}$$

More precisely $\phi'(t) = (1+t)^{\frac{p-2}{2}}t$, and ϕ satisfies all the same important conditions as φ. It also holds for $P \in \mathbb{R}^d$

$$F(P) = \phi'(|P|)\frac{P}{|P|}. \tag{2.41}$$

In order to show the proper relation between F and A we will need similar results as lemmas 2.12 and 2.14 for the function F.

Lemma 2.15. *Let $P, Q \in \mathbb{R}^d$ and $F(P) = \phi'(|P|)\frac{P}{|P|}$. Then*

$$\begin{aligned} (F(P) - F(Q)) \cdot (P-Q) &\geq c\phi'_{|P|}(|P-Q|)|P-Q| \\ |F(P) - F(Q)| &\leq c\phi'_{|P|}(|P-Q|). \end{aligned} \tag{2.42}$$

Proof. To prove the second inequality we use similar steps as in the proof of lemma 2.14. Assuming $|P| \neq 0$

$$\frac{\partial F_j(P)}{\partial P_l} = \frac{\phi'(|P|)}{|P|}\delta_{jl} - \frac{\phi'(|P|)}{|P|^3}P_jP_l + \phi''(|P|)\frac{P_jP_l}{|P|^2}.$$

Taking the norm and using $\phi''(t)t \sim \phi'(t)$, we have

$$\left|\frac{\partial F_j(P)}{\partial P_l}\right| \leq 2\frac{\phi'(|P|)}{|P|} + \phi''(|P|) \leq c\frac{\phi'(|P|)}{|P|}.$$

Using the equality

$$F_j(P) - F_j(Q) = \sum_{l=1}^{d} \int_0^1 \frac{\partial(F_j(tP + (1-t)Q))}{\partial P_l}(P_l - Q_l)dt,$$

we derive similarly as in the proof of lemma 2.14, with the use of derivative estimate, lemma 2.13 and the fact $|P| + |Q| \sim |P| + |P - Q|$ the following

$$\begin{aligned}|F(P) - F(Q)| &\leq c\int_0^1 \frac{\phi'(|tP + (1-t)Q)|}{|tP + (1-t)Q|}|P - Q|dt \\ &\leq c\frac{\phi'(|P| + |Q|}{|P| + |Q|}|P - Q| \leq c\phi'_{|P|}(|P - Q|).\end{aligned}$$

For the first inequality we need the following estimate for the derivative, assuming $P \neq 0$,

$$\begin{aligned}\sum_{j,l=1}^{d} \frac{\partial F_j(P)}{\partial P_l}Q_lQ_j &= \frac{\phi'(|P|)}{|P|}(|Q|^2 - \frac{|PQ|^2}{|Q|^2}) + \phi''(|Q|)\frac{|PQ|^2}{|Q|^2} \\ \geq \frac{\phi'(|P|)}{|P|}(|Q|^2 - \frac{|PQ|^2}{|Q|^2}) &+ c\frac{\phi'(|P|)}{|P|}\frac{|PQ|^2}{|Q|^2} = c\frac{\phi'(|P|)}{|P|}|Q|^2.\end{aligned}$$

Therefore we can use the equality

$$(F(P) - F(Q)) \cdot (P - Q) = \sum_{j,l=1}^{d} \int_0^1 \frac{\partial F_j(|tP + (1-t)Q|)}{|tP + (1-t)Q|}(P_l - Q_l)(P_j - Q_j)dt.$$

Using this, estimate for the derivative with $P = (tP + (1-t)Q)$ and $Q = (P - Q)$, and lemma 2.13 we have

$$\begin{aligned}(F(P) - F(Q)) \cdot (P - Q) &\geq c\int_0^1 \frac{\phi'(|tP + (1-t)Q|)}{|tP + (1-t)Q|}|P - Q|^2dt \\ \geq c\frac{\phi'(|P| + |Q|)}{|P| + |Q|}|P - Q|^2 &\geq c\phi'_{|P|}(|P - Q|)|P - Q|.\end{aligned}$$

□

Lemma 2.16. *Let $P, Q \in \mathbb{R}^d$ and $F(P) = \phi'(|P|)\frac{P}{|P|}$. Then*

$$(A(P) - A(Q)) \cdot (P - Q) \sim |F(P) - F(Q)|^2. \tag{2.43}$$

Proof. Using the result of lemma 2.15 and Cauchy-Schwarz inequality we get

$$|F(P)-F(Q)|^2 \sim \phi'_{|P|}(|P-Q|) = \varphi'_{|P|}(|P-Q|)|P-Q| \sim (A(P)-A(Q))\cdot(P-Q),$$

where in the equality we used the definition of ϕ and in the second equivalence we used (2.38). □

To sum up all the important relations between A, F and φ we recall (2.38), use (2.8) and lemma 2.16. Thus we get for all $P, Q \in \mathbb{R}^d$

$$|A(P) - A(Q)| \sim \varphi'_{|P|}(|P - Q|), \tag{2.44}$$

$$\begin{aligned}(A(P) - A(Q)) \cdot (P - Q) &\sim \varphi'_{|P|}(|P - Q|)|P - Q|, \\ &\sim \varphi_{|P|}(|P - Q|) \sim |F(P) - F(Q)|^2.\end{aligned} \tag{2.45}$$

conjugate function φ^* and F^* also satisfy the assumptions of lemma 2.16, and therefore

$$|F^*(P) - F^*(Q)|^2 \sim (\varphi^*)_{|P|}(|P - Q|). \tag{2.46}$$

Choosing $P = 0$ we also get for all $Q \in \mathbb{R}^d$

$$A(Q) \cdot Q \sim |F(Q)|^2 \sim \varphi(|Q|), \tag{2.47}$$

$$|A(Q)| \sim \varphi'(|Q|). \tag{2.48}$$

From lemma 2.9 and (2.45) it follows that

$$\varphi_{|P|}(t) \leq c(\varphi_{|Q|}(t) + \varphi_{|Q|}(|P - Q|)) \leq c(\varphi_{|Q|}(t) + |F(P) - F(Q)|^2). \tag{2.49}$$

We can apply this to φ^* and get

$$(\varphi_{|P|})^*(t) \leq c((\varphi_{|Q|})^*(t) + |F^*(P) - F^*(Q)|^2). \tag{2.50}$$

Another useful equivalence can be derived from (2.46) with $P = 0$, (2.48) together with (2.11) and (2.45) with $P = 0$

$$|F^*(A(Q))|^2 \sim \varphi^*(|A(Q)|) \sim \varphi(|Q|) \sim |F(Q)|^2. \tag{2.51}$$

From the definition of shifted N-functions it holds that if $a \sim b$, then $\varphi_a(t) \sim \varphi_b(t)$. Further lemma 2.7 implies that $((\psi_a)^*)(t) \sim (\psi^*)_{\psi'(a)}(t)$. Combining these two relations we can get

$$(\varphi^*)_{|A(P)|}(t) \sim (\varphi^*)_{\varphi'|P|}(t) \sim (\varphi_|P|)^*(t). \tag{2.52}$$

Using this for $t = |A(P) - A(Q)|$ together with (2.44) and (2.11) we arrive at

$$(\varphi^*)_{|A(P)|}(|A(P) - A(Q)|) \sim \varphi_{|P|}(|P - Q|). \tag{2.53}$$

Finally using (2.45)

$$|F^*(A(Q)) - F^*(A(P))|^2 \sim |F(Q) - F(P)|^2. \tag{2.54}$$

We finish this section with one more technical lemma, which will allow us to manipulate the normed integrals over a domain in the error estimates. Note that we will be using both notations for the normed integral introduced at the start of chapter 1. The following lemma can be found in [12, Lemma A.2].

Lemma 2.17. *For $K \in \mathcal{T}_h$ and $P : \Omega \to \mathbb{R}^d$ it holds*

$$\begin{aligned} \fint_K |F(P) - \langle F(P)\rangle_K|^2 dx &\sim \fint_K |F(P) - F(\langle P\rangle_K|^2 dx \\ &\sim \fint_K |F^*(A(P)) - F^*(\langle A(P)\rangle_K)|^2 dx. \end{aligned} \tag{2.55}$$

Proof. In the first equivalence, the inequality $\leq$ follows from

$$\fint_K |F(P) - \langle F(P)\rangle_K|^2 dx = inf_{Q\in\mathbb{R}^d} \fint_K |F(P) - Q|^2 dx.$$

Let us denote P_F the function that satisfies $F(P_F) = \langle F(P)\rangle_K$. Thanks to (2.45) we have

$$\fint_K |F(P) - F(\langle P\rangle_K|^2 dx \sim \fint_K (A(P) - A(\langle P\rangle_K) \cdot (P - \langle P\rangle_K) dx.$$

Since $\langle P\rangle_K$ is constant on K and $\fint_K (P - \langle P\rangle_K) dx = 0$, we can replace it by P_F, which is also constant,

$$\sim \fint_K (A(P) - A(P_F)) \cdot (P - \langle P\rangle_K) dx.$$

Using Young inequality with $\varphi_{|P|}$, (2.44) together with (2.11) and finally (2.45)

$$\begin{aligned} &\leq c_\epsilon \fint_K \varphi_{|P|})^*(|(A(P) - A(P_F))|) dx + \epsilon \fint_K \varphi_{|P|}(|(P - \langle P\rangle_K)|) dx \\ &\sim c_\epsilon \fint_K \varphi_{|P|}(|P - P_F|) dx + \epsilon \fint_K \varphi_{|P|}(|(P - \langle P\rangle_K)|) dx \\ &\sim c_\epsilon \fint_K |F(P) - \langle F(P)\rangle_K|^2 dx + \epsilon \fint_K |F(P) - F(\langle P\rangle_K)|^2 dx. \end{aligned}$$

This implies the inequality $\geq$.

For the second equivalence we use the fact that first one holds for $F = F^*$ and $P = A(P)$ and (2.51)

$$\begin{aligned} \fint_K |F^*(A(P)) - \langle F^*(A(P))\rangle_K|^2 dx &\sim \fint_K |F^*(A(P)) - F^*(\langle A(P)\rangle_K)|^2 dx \\ &\sim \fint_K |F(P) - \langle F(P)\rangle_K|^2 dx. \end{aligned}$$

□

3. Discretization of the domain and discrete function spaces

Let $\Omega \subset \mathbb{R}^d$ be our domain, where $d = 2, 3$ is the dimension. We assume that Ω is bounded, open and polygonal (in case $d = 2$) or polyhedral (in case $d = 3$) with the Lipschitz-continuous boundary $\delta\Omega = \delta\Omega_D \cup \delta\Omega_N$ and $\delta\Omega_D \cap \delta\Omega_N = \emptyset$. For $T > 0$ we define $Q_T = \Omega \times (0, T)$. We are studying the following equation, which follows from (1.29), with $u : Q_T \to \mathbb{R}$ as the solution.

$$\begin{aligned} &u_t - \nabla \cdot (K(|\nabla u|)\nabla u) = f \quad x \in Q_T, \\ &u|_{\delta\Omega_D \times (0,T)} = u_D, \\ &(K(|\nabla u|)\nabla u) \cdot \boldsymbol{n}|_{\delta\Omega_N \times (0,T)} = g_N, \\ &u(x, 0) = u^0(x) \quad x \in \Omega, \end{aligned} \tag{3.1}$$

for given data $f : Q_T \to \mathbb{R}$, $u_D : \delta\Omega_D \times (0, T) \to \mathbb{R}$, $g_N : \delta\Omega_N \times (0, T) \to \mathbb{R}$ and $u^0 : \Omega \to \mathbb{R}$. Here $\mathbf{n}$ is the outer normal to $\delta\Omega$ and K is a nonlinear function that represents the model.

The goal of this paper lies in the selection of the suitable numerical discontinuous Galrekin (DG) method for the solution of this problem and its numerical analysis. DG methods are similar to the classic Finite element methods, with the main difference being that we do not require the conforming properties, meaning that our test functions do not need to be continuous on the edges of the triangulation.

3.1 Discretization of the domain

Let $h > 0$ and $\mathcal{T}_h$ be a partition of the closure of Ω into finite number of closed simplexes K with mutually disjoint interiors. We call $\mathcal{T}_h$ a triangulation of Ω. For each $K \in \mathcal{T}_h$ we denote $h_K = diam(K)$ and $h = max_{K \in \mathcal{T}_h}$. For simplicity we will assume that $h \leq 1$. By ρ_K we denote the radius of the largest d-dimensional ball inscribed into K and by $|K|$ we denote the d-dimensional Lebesgue measure of K. In order to avoid the elements K having certain undesired shapes, like drastically unproportional lengths of the sides, we require the following property

$$\frac{h_K}{\rho_K} \leq C_R, \qquad K \in \mathcal{T}_h, \tag{3.2}$$

where C_R is a positive constant. This is one of the constants that many of the future estimates will depend on, but will not be included explicitly. For a simplex K in $\mathcal{T}_h$ denote S_K the neighbourhood of K, meaning the union of all simplices touching K. We will assume that in our triangulation each S_K has a connected interior. By F_h we denote the set of all $(d-1)$-dimensional faces of all elements $K \in \mathcal{T}_h$ (edges in case $d = 2$ and faces in case $d = 3$). For an edge $\Gamma \in F_h$ we denote $S_\Gamma = K \cup K'$, if $\Gamma = \delta K \cap \delta K'$, or $S_\Gamma = K$, if Γ is an edge on the boundary of Ω. Further we divide F_h into F_h^I, representing the interior faces of F_h and F_h^D, F_h^N, representing the faces belonging to the Dirichlet and the Neumann part of

the boundary respectively. Sometimes simplified notation will be used, combining the subscripts, for example $F_h^{DN} = F_h^D \cup F_h^N$. Also, we will use the simplified notation for integrals over a set of edges, for example $\sum_{\Gamma \in F_h^I} \int_\Gamma \dots ds \coloneqq \int_{F_h^I} \dots ds$.

We will also need for each $\Gamma \in F_h$ the unit normal vector $\mathbf{n_\Gamma}$. For $\Gamma \in F_h^{DN}$, $\mathbf{n_\Gamma}$ is the outer normal to $\delta\Omega$. For $\Gamma \in F_h^I$, the orientation of $\mathbf{n_\Gamma}$ is arbitrary, but fixed for each face.

It is useful to require one more condition on the triangulation $F_h, h > 0$. We introduce the quantity $h_\Gamma > 0$, which represents a "one dimensional" size of the face Γ. We require that h_Γ satisfy the following equivalence condition with h_K, for each $K \in \mathcal{T}_h$ and its face Γ

$$h_K \sim h_\Gamma. \tag{3.3}$$

The properties of the mesh $\mathcal{T}_h$ imply that for each $K \in \mathcal{T}_h$ and its face Γ, it holds

$$|K| \sim h_\Gamma |\Gamma|, \tag{3.4}$$

where $|\Gamma|$ is $d-1$ dimensional Lebesgue measure of Γ.

3.2 Function spaces

In numerical analysis of this problem we will use so called broken Sobolev spaces. Over a triangulation $\mathcal{T}_h$ we define for $q \geq 0$, the broken Sobolev space as

$$W_{DG}^{l,q}(\Omega) \coloneqq \{v; v \in L^2(\Omega), v|_K \in W^{l,q}(K); \forall K \in \mathcal{T}_h\},$$

with the seminorm $|v|_{W_{DG}^{l,q}(\Omega)} = (\sum_{K \in \mathcal{T}_h} |v|_W^{l,q}(K))^{\frac{1}{2}}$, where $|v|_W^{l,q}(K)$ is the standard Sobolev seminorm on $W^{l,q}(K)$, $K \in \mathcal{T}_h$. We can analogically define the vector valued broken Sobolev space $W_{DG}^{l,q}(\Omega, \mathbb{R}^d)$.

Let $k \in \mathbb{N}$ denote the degree of the polynomial approximation, then we denote $P_k(K)$ and $P_k(K, \mathbb{R}^d)$ the spaces of scalar and vector polynomial functions on K, of degree $\leq k$, $K \in \mathcal{T}_h$. Now we can define the finite dimensional subspaces of $W^{1,q}(\Omega, \mathcal{T}_h)$ and $W^{1,q}(\Omega, \mathcal{T}_h, \mathbb{R}^d)$ by

$$V_h^k = V_h^k(\Omega) \coloneqq \{v; v \in L^2(\Omega), v|_K \in P_k(K); \forall K \in \mathcal{T}_h\}, \tag{3.5}$$

$$X_h^k = X_h^k(\Omega) \coloneqq \{\boldsymbol{v}; \boldsymbol{v} \in L^2(\Omega, \mathbb{R}^d), \boldsymbol{v}|_K \in P_k(K, \mathbb{R}^d); \forall K \in \mathcal{T}_h\}. \tag{3.6}$$

We will also work with the broken Sobolev-Orlicz space

$$W_{DG}^{1,\psi} = W_{DG}^{1,\psi}(\Omega) \coloneqq \{v; v \in L^2(\Omega, \mathbb{R}^d), \boldsymbol{v}|_K \in W^{1,\psi}(K); \forall K \in \mathcal{T}_h\}. \tag{3.7}$$

Sometimes we will use the same notation for scalar valued functions in the analogous space. Note that both $W^{1,\psi}(\Omega)$ and $V_h^k(\Omega)$ are subspaces of $W_{DG}^{1,\psi}(\Omega)$.

Since our test functions from V_h^k or X_h^k are not continuous , we will need to define jumps and averages of these functions on the edges of the triangulation.

In case $\Gamma \in F_h^I$ there always exist two elements K^+ and $K^- \in \mathcal{T}_h$ such that $\Gamma \subset K^+ \cap K^-$ and K^- lies in the direction of $\boldsymbol{n}_\Gamma$. For $\Gamma \in F_h^{DN}$ there exists K^+, such that $\Gamma \subset K^+ \cap \delta\Omega$.

Now we can define jumps and averages on the edges of the triangulation, for $v \in V_h^k$. For each $\Gamma \in F_h^I$ denote $v|_\Gamma^+$ as the trace $v|_{K^+}$ and $v|_\Gamma^-$ as the trace $v|_{K^-}$. The mean value of v on $\Gamma \in F_h^I$ is defined as

$$\{v\}_\Gamma = (v|_\Gamma^+ + v|_\Gamma^-)/2,$$

and the jump of v on $\Gamma \in F_h^I$ is defined as

$$[v]_\Gamma = v|_\Gamma^+ - v|_\Gamma^-.$$

Note that $[v]$ depends on the orientation of $\boldsymbol{n}_\Gamma$, but $[v]\boldsymbol{n}_\Gamma$ does not. In case $\Gamma \in F_h^{DN}$ the definition of $v|_\Gamma^+$ is the same and we set $\{v\}_\Gamma = [v]_\Gamma = v|_{K^+}$. When there is no doubt about to which edge $\Gamma \in F_h$ symbols $\boldsymbol{n}_\Gamma, \{v\}_\Gamma$ and $[v]_\Gamma$ belong, for example if they are arguments of $\int_\Gamma \dots ds$, the subscript Γ is omitted.

In the vectorial case $\boldsymbol{v} \in X_h^k$, $\boldsymbol{v}|_\Gamma^+$, $\boldsymbol{v}|_\Gamma^-$,the mean value and the jump are defined analogically. For example $\{\boldsymbol{v}\}_\Gamma \coloneqq (\boldsymbol{v}|_\Gamma^+ + (\boldsymbol{v}|_\Gamma^-)/2$ and $[\boldsymbol{v}]_\Gamma \coloneqq (\boldsymbol{v}|_\Gamma^+ - (\boldsymbol{v}|_\Gamma^-$, $\Gamma \in F_h^I$.

Before we proceed with the discretization of the main equation, we will need to define the generalization of the distributional gradient for discontinuous functions and two projections onto the finite dimensional spaces V_h^k. In the next next chapter we will also derive some useful results concerning these definitions.

4. Auxiliary results

4.1 The global distributional gradient generalization

First we would like to generalize the global distributional gradient to the DG setting, using similar construction as in [10, Appendix 2]. For $g \in W^{1,\psi}_{DG}(\Omega)$ the local distributional gradient will be denoted $\nabla_h g$, meaning that for each $K \in \mathcal{T}_h$, $\nabla_h(g)$ only depends on values of g on K. Since functions from $W^{1,\psi}_{DG}(\Omega)$ are not continuous across Ω, the global distributional gradient will contain the terms with their jumps on the inner edges of the triangulation. In further analysis, it will be beneficial for us if it also included the jumps on the edges from $\mathcal{F}^D_h$.

In order to achieve this, we will need the following construction. Let Ω' be an extension of the domain Ω, such that Ω is a strict subset of Ω' and $\delta\Omega \setminus \delta\Omega' = \mathcal{F}^D_h$, $\delta\Omega \cap \delta\Omega' = \mathcal{F}^N_h$. Also denote $\mathcal{T}'_h$ the extended triangulation of $\mathcal{T}_h$ to Ω' having the same properties as $\mathcal{T}_h$. All notation associated with $\mathcal{T}'_h$ will be differentiated by the addition of $'$, for example simplex $K' \in \mathcal{T}'_h$ or S'_K being the neighbourhood of K' in $\mathcal{T}'_h$. We define the space of functions from $W^{1,\psi}_{DG}(\Omega)$ extended by 0 to Ω' by

$$W^{1,\psi}_{DG,D}(\Omega) = \{g \in W^{1,\psi}_{DG}(\Omega); g|_{\Omega' \setminus \Omega} = 0\}. \tag{4.1}$$

Using the definitions of the jumps on the boundaries of elements K in the triangulation $\mathcal{T}_h$ we have for $g \in W^{1,\psi}_{DG,D}$ and $\boldsymbol{x} \in C^\infty_0(\Omega, \mathbb{R}^d)$

$$\sum_{K \in \mathcal{T}_h} \int_{\delta K} g\boldsymbol{x}\boldsymbol{n}dx = \int_{\mathcal{F}^{ID}_h} [g]\boldsymbol{x}\boldsymbol{n}ds. \tag{4.2}$$

Therefore the global distributional gradient on Ω', for $g \in W^{1,\psi}_{DG,D}(\Omega)$ satisfies for $\boldsymbol{x} \in C^{\infty,\mathbb{R}^d}_0(\Omega')$

$$(\nabla g, \boldsymbol{x})_{D'(\Omega'),D(\Omega')} = \int_{\Omega'} \nabla_h g\boldsymbol{x}dx - \int_{\mathcal{F}^{ID}_h} [g]\boldsymbol{x}\boldsymbol{n}ds. \tag{4.3}$$

This motivates for $g \in W^{1,\psi}_{DG}(\Omega)$ the definition of ∇g, extended as a functional for discontinuous functions $\boldsymbol{x}_h \in X^k_h(\Omega)$ by

$$(\nabla g, \boldsymbol{x}_h)_{D'(\Omega'),D(\Omega')} = \int_{\Omega'} \nabla_h g\boldsymbol{x}_h dx - \int_{\mathcal{F}^{ID}_h} [g]\{\boldsymbol{x}_h\}\boldsymbol{n}ds. \tag{4.4}$$

This functional is continuous and therefore we can use Reisz theorem and define its representation for all $\boldsymbol{x}_h \in X^k_h(\Omega)$ by

$$\int_{\Omega'} \nabla^h_{DG} g\boldsymbol{x}_h dx = \int_{\Omega'} \nabla_h g\boldsymbol{x}_h dx - \int_{\mathcal{F}^{ID}_h} [g]\{\boldsymbol{x}_h\}\boldsymbol{n}ds. \tag{4.5}$$

In the same way we can represent for $\Gamma \in \mathcal{F}^{ID}_h$ only the second term on the right hand side as

$$\int_{\Omega'} R^\Gamma_h g\boldsymbol{x}_h dx = \int_\Gamma [g]\{\boldsymbol{x_h}\}\boldsymbol{n}ds, \tag{4.6}$$

and denoting $R_h = \sum_{\Gamma \in \mathcal{F}_h^{ID}} R_h^\Gamma$ we have

$$\int_{\Omega'} R_h g \boldsymbol{x}_h dx = \int_{\mathcal{F}_h^{ID}} [g]\{\boldsymbol{x_h}\}\boldsymbol{n} ds. \tag{4.7}$$

In some literature the functionals R_h are called jump functionals. With this definition we can write

$$\nabla_{DG}^h g = \nabla_h g - R_h g. \tag{4.8}$$

Note that the same holds for $g_h \in V_h^k(\Omega) \subset W_{DG}^{1,\psi}(\Omega)$. For $\Gamma \in F_h^{ID}$ and $g \in W_{DG,D}^{1,\psi}$ we have by equivalence of norms on the finite dimensional $X_h^k(S_\Gamma)$

$$\begin{aligned} \left\| R_h^\Gamma g \right\|_{L^\infty(S_\Gamma)} &\leq c \fint_{S_\Gamma} |R_h^\Gamma g| dx = c sup_{\boldsymbol{x}_h \in X_h^k(S_\Gamma); \|\boldsymbol{x}_h\|_\infty \leq 1} \frac{1}{|S_\Gamma|} (R_h^\Gamma g, \boldsymbol{x}_h) \\ &\leq \frac{1}{h_\Gamma |\Gamma|} \int_\Gamma |[g]\boldsymbol{n}| ds \, \|\boldsymbol{x}_h\|_\infty \leq c \fint_\Gamma h_\Gamma^{-1} |[g]\boldsymbol{n}| ds, \end{aligned} \tag{4.9}$$

where we also used $|S_\Gamma| \sim h_\Gamma |\Gamma|$. Therefore we also have the pointwise inequality

$$|R_h^\Gamma g| \leq c \fint_\Gamma h_\Gamma^{-1} |[g]\boldsymbol{n}| ds. \tag{4.10}$$

From convexity of an N-function ψ, by Jensen's inequality we get

$$\psi(|R_h^\Gamma g|) \leq c \fint_\Gamma \psi(h_\Gamma^{-1} |[g]\boldsymbol{n}|) ds. \tag{4.11}$$

Integrating this on S_Γ, we get

$$\int_{S_\Gamma} \psi(|R_h^\Gamma g|) dx \leq c h_\Gamma \int_\Gamma \psi(h_\Gamma^{-1} |[g]\boldsymbol{n}|) ds. \tag{4.12}$$

And finally if we sum this through all $\Gamma \in \mathcal{F}_h^{ID}$, we arrive at

$$\int_\Omega \psi(|R_h g|) dx \leq c h \int_{\mathcal{F}_h^{ID}} \psi(h^{-1} |[g]\boldsymbol{n}|) ds. \tag{4.13}$$

4.2 The local L^2 projection

We will also need to introduce the local L^2 projection $\Pi : L^1(\Omega) \to V_h^k(\Omega)$ and derive some estimates for the interaction between Π and N-functions. Most of the presented estimates are the chosen results from [10, Appendix 1] and [12].

Definition 4.1. *We define the local L^2 projection $\Pi : L^1(\Omega) \to V_h^k(\Omega)$ by*

$$\int_\Omega \Pi g z_h dx = \int_\Omega g z_h dx \quad \forall z_h \in V_h^k(\Omega). \tag{4.14}$$

The same projection can be analogously defined for $L^1(\Omega, \mathbb{R}^d)$ functions as $\Pi : L^1(\Omega, \mathbb{R}^d) \to \boldsymbol{x}_h^k(\Omega)$.

From the nature of the test functions z_h, there also holds a local version

$$\int_K \Pi g z_h dx = \int_K g z_h dx \quad \forall z_h \in P_k(K). \tag{4.15}$$

Since Π is a local $L^2 - projection$, for $g \in L^2(K)$ it holds

$$⨍_K |\Pi g|^2 dx \leq ⨍_K |g|^2 dx, \tag{4.16}$$

where we used the notation $⨍_M g dx = \frac{1}{|M|}\int g dx$. Since $P_k(K)$ is finite dimensional it follow by the equivalence of norms that for $g \in L^1(\Omega)$

$$\|\Pi g\|_{L^\infty(K)} \leq c⨍_K |\Pi g| dx \leq c sup_{z_h \in P_k(K); \|z_h\|_\infty \leq 1} ⨍_K |\Pi g z_h| dx \leq c⨍_K |g| dx. \tag{4.17}$$

Since N-functions are convex, we have by Jensen's inequality that for an N-function ψ, it holds

$$⨍_K \psi(|\Pi g|) dx \leq c\psi(\|\Pi g\|_{L^\infty(K)}) \leq c\psi(⨍_K |g| dx) \leq c⨍_K \psi(|g|) dx. \tag{4.18}$$

In order to get further estimates concerning the projection Π, we will need the following lemma, called the inverse theorem for polynomials.

Lemma 4.1. *Let $K \in \mathcal{T}_h$ and $p \in P_k(K)$. Then*

$$|p|_{H^1(K)} \leq c h_K^{-1} \|p\|_{L^2(K)}. \tag{4.19}$$

Proof. Let K' be a reference simplex with $h_{K'} = 1$ and $G : K' \to K$ be an affine mapping of K' onto K.For $x' \in K'$ denote $p'(x') = p(x)$, where $x = G(x')$. Then by substitution theorem it holds that

$$\begin{aligned} |p|_{H^1(K)} &\leq c h_K^{d/2-1} |p'|_{H^1(K')}, \\ \|p'\|_{L^2(K')} &\leq c h_K^{d/2} \|p\|_{L^2(K)}. \end{aligned}$$

Then by the equivalence of norms on finite dimensional space

$$|p|_{H^1(K)} \leq c h_K^{d/2-1} |p'|_{H^1(K')} \leq c h_K^{d/2-1} \|p'\|_{L^2(K')} \leq c h_K^{-1} \|p\|_{L^2(K)}. \tag{4.20}$$

□

Note that by equivalence of norms the same holds for the L^1 norm of ∇p and p.

In order to derive most of the important estimates for the projection Π that will be used in the derivation of the stability of the numerical solution to the problem (3.1), we will need the following construction. Let $0 \leq j \leq l \leq k+1$. For all $p \in P_k(K)$ it holds that $\Pi p = p$. Thus

$$\begin{aligned} ⨍_K \psi(h_K^j |\nabla_h^j (g - \Pi g)|) dx &= ⨍_K \psi(h_K^j |\nabla_h^j (g - p + p - \Pi g)|) dx \\ &\leq c(⨍_K \psi(h_K^j |\nabla_h^j (g - p)|) dx + ⨍_K \psi(h_K^j |\nabla_h^j \Pi (g - p)|) dx). \end{aligned} \tag{4.21}$$

For the second term on the right hand side, we use in order the equivalence of L_∞ and L_1 norms for p , the inverse inequality for polynomials (4.1), the Jensen's inequality and (4.18).

$$\begin{aligned} \fint_K \psi(h_K^j|\nabla_h^j \Pi(g-p)|)dx \leq \fint_K \psi(ch_K^j|\fint_K |\nabla_h^j \Pi(g-p)dx|)dx \\ \leq \fint_K \psi(c\fint_K |\Pi(g-p)|)dx \leq c\fint_K \psi(|\Pi(g-p)|)dx \leq c\fint_K \psi(|g-p|)dx. \end{aligned} \tag{4.22}$$

Substituting this back into previous estimate, we have

$$\begin{aligned} \fint_K \psi(h_K^j|\nabla_h^j(g-\Pi g)|)dx = \fint_K \psi(h_K^j|\nabla_h^j(g-p+p\Pi g)|)dx \\ \leq c(\fint_K \psi(h_K^j|\nabla_h^j(g-p)|)dx + \fint_K \psi(|g-p|)dx). \end{aligned} \tag{4.23}$$

We will further need the following lemma

Lemma 4.2. *Let $0 \leq j \leq l \leq k+1$, $K \in \mathcal{T}_h$ and $g \in W^{1,\psi}(\Omega)$, then there exists a polynomial $q \in P_{l-1}(\Omega)$, such that*

$$\sum_{i=0}^{l} \fint_K \psi(h_K^i|\nabla^i(g-q)|)dx \leq c\fint_K \psi(h_K^l|\nabla^l g|)dx. \tag{4.24}$$

Proof. The proof to this lemma can be found in [17, Corollary 3.3]. The polynomial, for which the assertion of the lemma holds is the averaged Taylor polynomial of g. □

Since $p \in P_k(K)$ was arbitrary, we can choose the polynomial from lemma 4.2 in (4.23) and get for all $K \in \mathcal{T}_h$ and $g \in W^{1,\psi}(K)$

$$\fint_K \psi(h_K^j|\nabla_h^j(g-\Pi g)|)dx \leq c\fint_K \psi(h_K^l|\nabla_h^l g|)dx. \tag{4.25}$$

Now we sum through all $K \in \mathcal{T}_h$ and using the cases $j=0$, $l=0$ and $j=0$, $l=1$ and finally $j=1$, $l=1$ we get the following set of estimates

$$\int_\Omega \psi(|g-\Pi g|) \leq c\int_\Omega \psi(|g|), \tag{4.26}$$

$$\int_\Omega \psi(|g-\Pi g|) \leq c\int_\Omega \psi(h|\nabla_h g|), \tag{4.27}$$

$$\int_\Omega \psi(|\nabla_h(g-\Pi g)|) \leq c\int_\Omega \psi(|\nabla_h g|). \tag{4.28}$$

Also using triangle inequality we get

$$\int_\Omega \psi(|\Pi g|) \leq c\int_\Omega \psi(|g|), \tag{4.29}$$

$$\int_\Omega \psi(|\nabla_h(\Pi g)|) \leq c\int_\Omega \psi(|\nabla_h g|). \tag{4.30}$$

Lemma 4.3. *Let Γ be a face of $K \in \mathcal{T}_h$, then for all $g \in W^{1,\Psi}(K)$ it holds*

$$\fint_\Gamma \psi(|g|)ds \leq c(\fint_K \psi(|g|)dx + \fint_K \psi(h_\Gamma|\nabla g|)dx). \tag{4.31}$$

Proof. From the theory of Sobolev-Orlitz spaces we have the embedding $W^{1,\psi}(K) \hookrightarrow L^{\psi}(\Gamma)$ and therefore

$$\int_{\Gamma} \psi(|g|)ds \leq c(\int_{K} \psi(|g|)dx + \int_{K} \psi(|\nabla g|)dx.$$

By multiplying both sides by $\frac{1}{|\Gamma|} \sim \frac{h_{\Gamma}}{|K|}$ we get the original assertion. □

Due to the equivalence of norms on the finite dimensional space, if $g \in P_k(K)$, the previous lemma implies

$$⨍_{\Gamma} \psi(|g|)ds \leq c⨍_{K} \psi(|g|)dx. \tag{4.32}$$

For the final set of results, let $g \in W^{1,psi}(K)$ and Γ be a face of $K \in \mathcal{T}_h$, then by lemma 4.3 and (4.25) for $j = 0, 1$ and $l = 1$

$$\begin{aligned} h_{\Gamma} \int_{\Gamma} \psi(h_{\Gamma}^{-1}|g - \Pi g|)ds &\leq c(h_{\Gamma} \int_{K} \psi(h_{\Gamma}^{-1}|g - \Pi g|)dx + h_{\Gamma} \int_{K} \psi(\nabla_h|g - \Pi g|)dx) \\ &\leq c \int_{K} \psi(|\nabla_h g|). \end{aligned} \tag{4.33}$$

Summing this through all $K \in \mathcal{T}_h$ we have

$$h \int_{\mathcal{F}_h^{ID}} \psi(h^{-1}[|g - \Pi g]\boldsymbol{n}|)ds \leq c \int_{\Omega} \psi(|\nabla_h g|). \tag{4.34}$$

For the sake of simplification of the notation, let us denote for $g \in W^{1,\psi}(\Omega)$

$$M_{\psi,h}(g) \coloneqq \int_{\Omega} \psi(|\nabla_h g|) + h \int_{\mathcal{F}_h^{ID}} \psi(h^{-1}[|g]\boldsymbol{n}|)ds. \tag{4.35}$$

Using estimate (4.28) we get the final estimate for the projection Π

$$M_{\psi,h}(g - \Pi g) = \int_{\Omega} \psi(|\nabla_h(g - \Pi g)|) + h \int_{\mathcal{F}_h^{ID}} \psi(h^{-1}[|g - \Pi g]\boldsymbol{n}|)ds \leq c \int_{\Omega} \psi(|\nabla_h g|). \tag{4.36}$$

4.3 Scott-Zhang interpolation

The second functional, we need to introduce is a generalized Scott-Zhang interpolation and its interactions with the N-functions, based on the results in [10, Appendix 3], [18] and [19].

In order to define an analogy for a classic Scott-Zhang interpolation in DG setting we will need to denote the following discontinuous function spaces

$$\begin{aligned} V_h^{k,1}(\Omega) &\coloneqq V_h^k(\Omega) \cap W^{1,1}(\Omega), \\ V_h^{k,1}(\Omega') &\coloneqq V_h^k(\Omega') \cap W^{1,1}(\Omega'). \end{aligned} \tag{4.37}$$

Len N' be the set of all nodes in the triangulation $\mathcal{T}_h'$ and $\{\phi_a\}_{a \in N'}$ be the Lagrange basis of $V_h^{k,1}(\Omega')$, for example in each $a \in N'$, ϕ_a is locally a polynomial of degree k, $\phi_a(a) = 1$ and ϕ_a is zero in all the nodes in the same simplex K as

the node a. Number of these nodes is equal to the degrees of freedom allowed by the given degree of the polynomial approximation. This is one of the advantages of the discontinuous approach allowing the test functions to be supported only on exactly one simplex of the triangulation.

With $a \in N'$ we associate either a simplex $K = K_a$, if $a \in intK$ or a face $\Gamma = \Gamma_a$, if $a \in \overline{\Gamma}$. We will not differentiate the notation and denote both as σ_a.

Let us denote $\{\phi_{a,i}\}_i$ the local basis of $[\phi_b|_{\sigma_a} | b \in N']$, such that $\phi_{a,1} = \phi_a|_{\sigma_a}$ is the base function that is nonzero in a. Let $\{\beta_{a,i}\}_i$ be the dual basis to ϕ_{a,i_i} with respect to scalar product $(f,g)_{\sigma_a} = \int_{\sigma_a} fgdx$. Therefore $(\phi_{a,i}, \beta_{a,j}) = \delta_{i,j}$. Now we can finally define the Scott-Zhang interpolation Π_{SZ} for a smooth function $g \in W^{1,1}(\Omega')$ by

$$\Pi_{SZ} g = \sum_{a \in N'} (g, \beta_{a,1})_{\sigma_a} \phi_a. \tag{4.38}$$

Let us now extend this definition to discontinuous functions on Ω. If σ_a is a face, we arbitrarily choose exactly one of the two simplices, of which it is a part of and denote it K_a. Whenever, there is not clear, which trace of g should we work with on a face σ_a, we take the trace $g|_{K_a}$, more precisely

$$\Pi_{SZ} g = \sum_{a \in N'} (g|_{K_a}, \beta_{a,1})_{\sigma_a} \phi_a. \tag{4.39}$$

By this extended definition, the functional $\Pi_{SZ} : W^{1,1}_{DG}(\Omega') \to V^{k,1}_h(\Omega')$ is a linear mapping and a projection. For K outside S'_{K_b}, $\phi_b|_K = 0$ and therefore $(\Pi_{SZ} g)|_K$ only depends on values of g on S'_K.

In order to ensure that functions extended by zero outside Ω, stay that way after being projected by Π_{SZ}, we require that, if σ_a is a face and $\sigma_a \in F^D_h$ then K_a assigned to σ_a is a simplex outside Ω. This requirement implies that $\Pi_{SZ} : W^{1,1}_{DG,D}(\Omega) \to V^k_h(\Omega') \cap W^{1,1}_D(\Omega')$.

It is useful to note the standard result for Scott-Zhang interpolation

$$\|\phi_{a,i}\|_\infty \leq 1, \qquad \|\beta_{a,i}\|_\infty \leq \frac{c}{|\sigma_a|}. \tag{4.40}$$

Lemma 4.4. *Let $g \in W^{1,\psi}(\Omega)$, ψ be an N-function, $K \in \mathcal{T}_h$ and $0 \leq j \leq l \leq k+1$, then*

$$⨍_K \psi(h^j_K |\nabla^j_h (g - \Pi_{SZ} g)|) dx \leq c ⨍_{S_K} \psi(h^l_K |\nabla^l g|) dx. \tag{4.41}$$

Proof. for $p \in P_k(K)$ the interpolation Π_{SZ} also satisfies $\Pi_{SZ}(p) = p$ and therefore we can use the similar steps as in derivation of the result for Π in (4.25). □

Using the lemma 4.3 for $|g - \Pi_{SZ} g|$ we have

$$h_\Gamma \int_\Gamma \psi(h^{-1}_\Gamma |g - \Pi_{SZ} g|) ds \leq c\left(\int_K \psi(h^{-1}_\Gamma |g - \Pi_{SZ} g|) dx + \int_K \psi(|\nabla_h (g - \Pi_{SZ} g)|) dx\right). \tag{4.42}$$

and using lemma (4.4) for $j = 0, 1$ and $l = 1, 2$, the last term can be estimated by

$$\leq c \int_{S_K} \psi(|\nabla g|)dx,$$
$$\leq c \int_{S_K} \psi(h|\nabla^2 g|)dx.$$

Putting these together and summing through all $K \in \mathcal{T}_h$ we have the following estimates for $g \in W^{1,\psi}(\Omega)$ and $g \in W^{2,\psi}(\Omega)$ respectively

$$\begin{aligned} h \int_{\mathcal{F}_h^{ID}} \psi(h^-1|[g - \Pi_{SZ} g]\boldsymbol{n}|)ds + \int_\Omega \psi(h^-1|g - \Pi_{SZ} g|)dx \\ + \int_\Omega \psi(|\nabla_h \Pi_{SZ} g|)dx \leq c \int_\Omega \psi(|\nabla g|)dx, \end{aligned} \tag{4.43}$$

$$\begin{aligned} h \int_{\mathcal{F}_h^{ID}} \psi(h^-1|[g - \Pi_{SZ} g]\boldsymbol{n}|)ds + \int_\Omega \psi(h^-1|g - \Pi_{SZ} g|)dx \\ + \int_\Omega \psi(|\nabla_h (g - \Pi_{SZ} g)|)dx \leq c \int_\Omega \psi(|h\nabla^2 g|)dx, \end{aligned} \tag{4.44}$$

where in the first estimate we used the triangle inequality on the last term on the left hand side.

Lemma 4.5. *Let $\Gamma \in F_h^I$, such that $S_\Gamma = K^1 \cup K^2$. Then for all $g \in W_{DG}^{1,1}(S_\Gamma)$*

$$|\langle g \rangle_{K^1} - \langle g \rangle_{K^2}| \leq c \fint_{S_\Gamma} h_\Gamma |\nabla_h g| dx + c \fint_{S_\Gamma} |[g]\boldsymbol{n}| ds. \tag{4.45}$$

Proof. We only need to use classic Poincaré-Friedrichs's inequality on $W^{1,1}(S_\Gamma)$ in the third inequality of the following set of estimates

$$\begin{aligned} |\langle g \rangle_{K^1} - \langle g \rangle_{K^2}| &= |\langle g \rangle_{K^1} - \langle g|_{K^1} \rangle_{K^1} + \langle g|_{K^1} \rangle_{K^1} - \langle g|_{K^2} \rangle_{K^2} + \langle g|_{K^2} \rangle_{K^2} - \langle g \rangle_{K^2}| \\ &\leq |\langle g \rangle_{K^1} - \langle g|_{K^1} \rangle_{K^1}| + \langle [g]\boldsymbol{n} \rangle_\Gamma + |\langle g|_{K^2} \rangle_{K^2} - \langle g \rangle_{K^2}| \\ &\leq \fint_{K^1} |g - \langle g|_{K^1} \rangle_{K^1}| dx + \fint_{K^2} |g - \langle g|_{K^2} \rangle_{K^2}| dx + \fint_\Gamma |[g]\boldsymbol{n}| ds \\ &\leq c \leq \fint_{K^1} h_\gamma |\nabla_h g| dx + \fint_{K^2} h_\gamma |\nabla_h g| dx + \fint_\Gamma |[g]\boldsymbol{n}| ds. \end{aligned}$$

□

In the further estimates we will need to use the Poincaré inequality in L^p spaces.

Lemma 4.6. *Let M be a Lipschitz domain and $g \in W^{1,p}$, for $1 \leq p \leq \infty$, then there exists a constant c, depending on M and p, such that*

$$\|g - \langle g \rangle_K\|_{L^p(M)}^p \leq c\, diam(M) \|\nabla g\|_{L^p(M)}^p. \tag{4.46}$$

Lemma 4.7. *Let $K \in \mathcal{T}_h$ and $g \in W_{DG}^{1,1}(\Omega')$ and denote $F_h(S'_K)$ the interior faces in S'_K, then*

$$\begin{aligned} \|\nabla \Pi_{SZ} g\|_{L^\infty(K)} \leq c h_K^{-1} \|\Pi_{SZ} g - \langle g \rangle_K\|_{L^\infty(K)} \leq c h_K^{-1} \fint_K |\Pi_{SZ} g - \langle g \rangle_K| dx, \\ \|\Pi_{SZ} g - \langle g \rangle_K\|_{L^\infty(K)} \leq c \fint_{S'_K} h_K dx + c \sum_{\Gamma \in F_h(S'_K)} \fint_\Gamma |[g]\boldsymbol{n}| ds. \end{aligned} \tag{4.47}$$

Proof. For the first row of estimates we use the fact that $\nabla\langle g\rangle_K = 0$ and the inverse inequality for polynomials, i.e lemma (4.1).

$$\|\nabla\Pi_{SZ}g\|_{L^\infty(K)} = \|\nabla(\Pi_{SZ}g - \langle g\rangle_K)\|_{L^\infty(K)}$$
$$\leq ch_K^{-1}\|\Pi_{SZ}g - \langle g\rangle_K\|_{L^\infty(K)} \leq ch_K^{-1}\fint_K |\Pi_{SZ}g - \langle g\rangle_K|dx.$$

For the second row it follows from (4.40)

$$\|\Pi_{SZ}g - \langle g\rangle_K\|_{L^\infty(K)} = \|\Pi_{SZ}(g - \langle g\rangle_K)\|_{L^\infty(K)} \leq c\sum_{\alpha\in N';K_\alpha\subset S'_K}\langle|g|_{K_\alpha} - \langle g\rangle_K|)\rangle_{\sigma_\alpha}.$$

Using lemma 4.3, if σ_α is a face, or trivial estimate otherwise

$$\langle|g|_{K_\alpha} - \langle g\rangle_K|)\rangle_{\sigma_\alpha} \leq c\fint_{K_\alpha}|g|_{K_\alpha} - \langle g\rangle_K|dx + c\fint_{K_\alpha} h_{K_\alpha}|\nabla_h g|dx.$$

Now due to the fact that interior of S'_K is connected we use lemma 4.5 for each pair of a sequence of neighbouring simplices connecting K and K_α.

$$\fint_{K_\alpha}|g|_{K_\alpha} - \langle g\rangle_K|dx \leq \fint_{K_\alpha}|g - \langle g\rangle_K|dx + c\sum_{\Gamma\in F_h(S'_K)}(\fint_\Gamma|[g]\boldsymbol{n}|ds + \fint_{S_\Gamma} h_\Gamma|\nabla_h g|dx).$$

Finally using Poincaré lemma 4.6 we arrive at

$$\leq c\sum_{\Gamma\in F_h(S'_K)}\fint_\Gamma|[g]\boldsymbol{n}|ds + \fint_{S'_K} h_K|\nabla_h g|dx.$$

Putting everything together we get the second inequality. □

Using the fact that $|K| \sim h_\Gamma|\Gamma|$ and Jensen's inequality, the lemma 4.7 implies

$$\begin{aligned}&\int_K\psi(|\nabla\Pi_{SZ}g|)dx + \int_K\psi(h_K^{-1}|\Pi_{SZ}g - \langle g\rangle_K|)dx\\ &\leq c\int_{S'_K}\psi(|\nabla_h g|)dx + c\sum_{\Gamma\in F_h(S'_K)} h_\Gamma\int_\Gamma\psi(h_\Gamma^{-1}|[g]\boldsymbol{n}|)ds.\end{aligned} \tag{4.48}$$

In further estimates we will need the following Poincaré inequality extended to N-functions.

Lemma 4.8. *Let ψ and ψ^* be N-functions satisfying Δ_2 − condition and $g \in W^{1,\psi}(K)$ for $K \in \mathcal{T}_h$, then*

$$\fint_K\psi(|g - \langle g\rangle_K|)dx \leq c\fint_K\psi(h_K|\nabla g|). \tag{4.49}$$

Proof. The proof to this lemma can be found in [18, Theorem 7]. □

Using (4.48) and lemma 4.8 we have

$$\begin{aligned}&\int_K\psi(h_K^{-1}|g - \Pi_{SZ}g|)dx\\ &\leq c\int_K\psi(h_K^{-1}|\Pi_{SZ}g - \langle g\rangle_K|)dx + c\int_K\psi(h_K^{-1}|g - \langle g\rangle_K|)dx\\ &\leq c\int_{S'_K}\psi(|\nabla_h g|)dx + c\sum_{\Gamma\in F_h(S'_K)} h_\Gamma\int_\Gamma\psi(h_\Gamma^{-1}|[g]\boldsymbol{n}|)ds.\end{aligned} \tag{4.50}$$

Summation of (4.48) and (4.50) over $K \in \mathcal{T}'_h$ implies

$$\begin{gathered}\int_{\Omega'} \psi(h^{-1}|g - \Pi_{SZ} g|)dx + \int_{\Omega'} \psi(|\nabla \Pi_{SZ} g|)dx \\ \leq c \int_{\Omega'} \psi(|\nabla_h g|)dx + \sum_{\Gamma \in F_h^I(\Omega')} h_\Gamma \int_\Gamma \psi(h_\Gamma^{-1}|[g]\boldsymbol{n}|)ds = cM_{\psi,h,\Omega'}(g).\end{gathered} \tag{4.51}$$

Since $g \in W^{1,\psi}_{DG,D}(\Omega)$ is extended by zero outside Ω the same estimate holds on Ω

$$\begin{gathered}\int_{\Omega} \psi(h^{-1}|g - \Pi_{SZ} g|)dx + \int_{\Omega} \psi(|\nabla \Pi_{SZ} g|)dx \\ \leq c \int_{\Omega} \psi(|\nabla_h g|)dx + \sum_{\Gamma \in F_h^I(\Omega')} h_\Gamma \int_\Gamma \psi(h_\Gamma^{-1}|[g]\boldsymbol{n}|)ds = cM_{\psi,h,\Omega}(g).\end{gathered} \tag{4.52}$$

Lemma 4.9. *Let Ω be a domain with Lipschitz boundary and $g \in W_0^{1,1}(\Omega)$, then*

$$|g(x)| \leq c \int_\Omega \frac{|\nabla g(y)|}{|x-y|^{d-1}} dy. \tag{4.53}$$

a.e in Ω.

Proof. function g extended by zero to Ω' satisfies the assumptions of the representation lemma in [19, Lemma 8.2.1b], proof of which we refer to the original work. □

Lemma 4.10. *Let ψ be an N-function satisfying Δ_2-condition and the function $g \in W^{1,\psi}_{DG,D}(\Omega)$, then it holds*

$$\int_\Omega \psi(|g|)dx \leq c \int_\Omega \psi(|diam(\Omega)\nabla g|)dx. \tag{4.54}$$

Proof. Let c_0 be a constant, such that $\int_\Omega diam(\Omega)^{-1}|x-y|^{d-1}dy \leq c_0$. Function g is extended by 0 on Ω' and $g \in W^{1,\psi}(\Omega')$. Using lemma (4.9), and Jensen inequality with respect to measure $\chi_\Omega c_0^{-1} diam(\Omega)^{-1}|x-y|^{d-1}$ we have

$$\begin{gathered}\int_\Omega \psi(|g|)dx \leq \int_\Omega c\psi(\int_\Omega \frac{|\nabla g(y)|}{|x-y|^{d-1}} dy)dx \\ \leq c \int_\Omega \int_\Omega \psi(diam(\Omega))|\nabla g(y)|)diam(\Omega)^{-1}|x-y|^{d-1}dx \\ \leq c \int_\Omega \psi(|diam(\Omega)\nabla g|)dx.\end{gathered}$$

□

Lemma 4.11. *Let ψ be an N-function satisfying Δ_2-condition, then there exists an N-function ρ and $\theta \in (0,1)$, such that $\psi(\rho^{-1}(t)) \sim t^{1/\theta}$.*

Proof. In order to avoid having to introduce additional theory of Orlicz spaces, we refer this proof from [20, Lemma 1.2.2, Lemma 1.2.3] □

Lemma 4.12. *Let ψ, ψ^* be N-functions satisfying $\Delta_2 -$ condition, then if we denote $R = diam(\Omega)$, for all $g \in W^{1,\psi}_{DG,D}(\Omega)$ it holds*

$$\int_{F_h^N} \psi(|g|) \leq cR^{-1} \int_\Omega \psi(|R\nabla g|). \tag{4.55}$$

Proof. Using lemma (4.9) and Jensen inequality applied to ρ and measure $\chi_\Omega c_0^{-1} diam(\Omega)^{-1}|x-y|^{d-1}$.

$$\begin{aligned}\int_{F_h^N} \psi(|g|) &\leq \int_{F_h^N} c\psi(\int_\Omega \frac{|\nabla g(y)|}{|x-y|^{d-1}} dy)dx \\ &\leq c\int_{F_h^N} \psi(\rho^{-1}(\int_\Omega \rho(R|\nabla g(y)|)R^{-1}|x-y|^{1-d}dy))ds(x) \\ &\leq c\int_{F_h^N}\int_\Omega \rho(R|\nabla g(y)|)R^{-1}|x-y|^{1-d}dyds(x) = c\int_{F_h^N} I_1 ds(x).\end{aligned}$$

For $\alpha > 0$ we multiply the inside of the integral I_1 by $1 = |x-y|^\alpha |x-y|^{-\alpha}$. Further we use Hölder inequality with $p = \frac{1}{\theta}$, $q = \frac{1}{1-\theta}$ and measure $|x-y|^{d-1}dy$.

$$I_1 \leq R^{-1/\theta}\int_\Omega \rho^{1/\theta}(R|\nabla g(y)|)|x-y|^{\alpha/\theta+1-d}dy(\int_\Omega |x-y|^{-\alpha/(1-\theta)+1-d}dy)^{1-\theta/\theta dx}.$$

Since we did not specify α previously, let $-\alpha/(1-\theta)+1-d > 1-d$. Then

$$I_1 \leq R^{-1/\theta}\int_\Omega \rho^{1/\theta}(R|\nabla g(y)|)|x-y|^{\alpha/\theta+1-d}dy(R^{-\alpha/(1-\theta)+1-d})^{1-\theta/\theta}.$$

Therefore

$$\begin{aligned}\int_{F_h^N} \psi(|g|) &\leq cR^{-1-\alpha/\theta}\int_{F_h^N}\int_\Omega \psi(R|\nabla g(y)|)|x-y|^{\alpha/\theta+1-d}dyds(x) \\ &\leq cR^{-1-\alpha/\theta}\int_\Omega \psi(R|\nabla g(y)|)\int_{F_h^N}|x-y|^{\alpha/\theta+1-d}ds(x)dy \\ &\leq cR^{-1}\int_\Omega \psi(|R\nabla g|)dy,\end{aligned}$$

since F_h^N is $d-1$ dimensional and

$$\int_{F_h^N}|x-y|^{\alpha/\theta+1-d}ds(x) \leq cR^{\alpha/\theta}.$$

□

Lemma 4.13. *Let ψ, ψ^* be N-functions satisfying $\Delta_2 -$ condition , the for all $g \in W^{1,\psi}_{DG,D}(\Omega)$*

$$\int_\Omega \psi(|g|)dx \leq cM_{\psi,h}(diam(\Omega)g). \tag{4.56}$$

Proof. For $g \in W^{1,\psi}_{DG,D}(\Omega)$, the projection $\Pi_{SZ}g \in W^{1,\psi}_{DG}(\Omega)$. From this, lemma 4.10 and (4.52), with $h_K \leq diam(\Omega)$ follows

$$\begin{aligned}\int_\Omega \psi(|g|)dx &\leq c\int_\Omega \psi(|g-\Pi_{SZ}g|)dx + c\int_\Omega \psi(|\Pi_{SZ}g|)dx \\ &\leq c\int_\Omega \psi(|g-\Pi_{SZ}g|)dx + c\int_\Omega \psi(|diam(\Omega)\Pi_{SZ}g|)dx \leq cM_{\psi,h}(diam(\Omega)g).\end{aligned}$$

□

For Γ face of K, using similar steps as in 4.12 we can get the estimate

$$⨍_{\Gamma} \psi(|g - \langle g \rangle_K|)dx \leq c⨍_K \psi(h_K|\nabla g|)dx. \tag{4.57}$$

Lemma 4.14. *Let ψ, ψ^* be N-functions satisfying $\Delta_2 -$ condition , the for all $g \in W^{1,\psi}_{DG,D}(\Omega)$*

$$\int_{F_h^N} \psi(|g|)ds \leq c\, diam(\Omega)^{-1} M_{\psi,h}(diam(\Omega)g). \tag{4.58}$$

Proof.

$$\int_{F_h^N} \psi(|g|)ds \leq c\int_{F_h^N} \psi(|g - \Pi_{SZ}g|)ds + c\int_{F_h^N} \psi(|\Pi_{SZ}g|)ds = I_1 + I_2.$$

For the estimate of I_1 we use (4.57), lemma 4.3 and (4.48)

$$I_1 = \sum_{\Gamma \in F_h^N} \int_\Gamma \psi(|g - \Pi_{SZ}g|)ds \leq c \sum_{\Gamma \in F_h^N} \int_\Gamma \psi(|g - \langle g \rangle_K|) + \psi(|\langle g \rangle_K - \Pi_{SZ}g|)ds$$

$$\leq c \sum_{K;\delta K \in F_h^N} h_K^{-1} \int_K \psi(h_K|\nabla_h g|) + h_K^{-1} \int_K \psi(|\langle g \rangle_K - \Pi_{SZ}g|) + \psi(h_K|\nabla \Pi_{SZ}g|)dx$$

$$\leq c \sum_{K \in \mathcal{T}_h} h_K^{-1} \int_{S_K} \psi(h_K|\nabla_h g|)dx + c \sum_{K;\delta K \in F_h^N} \int_\Gamma \psi(|[g]\boldsymbol{n}|)ds$$

$$\leq c\, diam(\Omega)^{-1} M_{\psi,h}(diam(\Omega)g).$$

From lemma 4.12 and (4.52) we have

$$I_2 \leq c\, diam(\Omega)^{-1} M_{\psi,h}(diam(\Omega)g).$$

□

5. Discontinuous Galerkin formulations

There were a couple of discontinuous Galerkin methods considered for the discretization of the problem, mainly interior penalty discontinuous Galerkin (IPDG) and local discontinuous Galerkin (LDG) method. In the end we chose the LDG method, described for the general case in [8], due to the easier error analysis. For the discretization itself, we follow the approach outlined in [10] and [11].

5.1 Local DG formulation

Using the definition $A(P) \coloneqq K(|P|)P$, for $P \in \mathbb{R}^d$ to simplify the notation, the main set of equations 3.1 we derived in chapter one can be rewritten as

$$\begin{aligned} &u_t - \nabla \cdot A(\nabla u) = f \quad x \in Q_T, \\ &u|_{\delta\Omega_D \times (0,T)} = u_D, \\ &A(\nabla u) \cdot \boldsymbol{n}|_{\delta\Omega_N \times (0,T)} = g_N, \\ &u(x,0) = u^0(x) \quad x \in \Omega. \end{aligned} \tag{5.1}$$

K is the nonlinear function defined in the first chapter motivated by the physical model and has all the properties we derived there.

For the problem data, we will assume that $u_D \in W^{1-\frac{1}{p},p}(F_h^D)$, $f \in L^p(\Omega)$ and $g_N \in L^q(F_h^N)$, where $\frac{1}{p} + \frac{1}{q} = 1$. Under these assumptions, by the theory of monotone operators there exists a weak solution $u \in W^{1,\varphi}(\Omega)$, $u - u_D \in W_D^{1,\varphi}(\Omega)$ satisfying for all $z \in W_D^{1,\varphi}(\Omega)$

$$\int_\Omega A(\nabla u) \cdot \nabla z dx = \int_\Omega f z dx + \int_{F_h^N} g_N z ds. \tag{5.2}$$

The local DG formulation has some similarities to classic Interior Penalty methods. First we rewrite the original equation (5.1) , into three equations of the first order. Then the equations are multiplied by the appropriate test functions and integration by parts is used. Instead of adding the interior penalty terms like in a IP method, the jumps on the edges of the triangulation are controlled by the appropriate choice of the numerical fluxes. These are the chosen approximations of the discrete solution on the edges of the triangulation.

Equation (5.1) can be rewritten as a system of first order equations for unknowns $u, \boldsymbol{l}, \boldsymbol{a}$.

$$
\begin{aligned}
&\boldsymbol{l} = \nabla u, \\
&\boldsymbol{a} = A(\boldsymbol{l}), \\
&u_t - \nabla \cdot \boldsymbol{a} = f,
\end{aligned}
\tag{5.3}
$$

$$
\begin{aligned}
&u|_{\mathcal{F}_h^D \times (0,T)} = u_D, \\
&\boldsymbol{a} \cdot \boldsymbol{n}|_{\mathcal{F}_h^N \times (0,T)} = g_N, \\
&u(x,0) = u^0(x) \quad x \in \Omega.
\end{aligned}
$$

Multiplying these equations by $\boldsymbol{x}_h, \boldsymbol{y}_h \in X_h^k, z_h \in V_h^k$ respectively, integrating over $K \in \mathcal{T}_h$ and using integration by parts we have

$$
\begin{aligned}
&\int_K \boldsymbol{l} \cdot \boldsymbol{x}_h dx = -\int_K u \nabla \cdot \boldsymbol{x}_h dx + \int_{\delta K} u \boldsymbol{x}_h \cdot \boldsymbol{n} ds, \\
&\int_K \boldsymbol{a} \cdot \boldsymbol{y}_h dx = \int_K A(\boldsymbol{l}) \cdot \boldsymbol{y}_h dx, \\
&\int_K u_t z_h dx + \int_K \boldsymbol{a} \cdot \nabla z_h dx = \int_K f z_h dx + \int_{\delta K} z_h \boldsymbol{a} \cdot \boldsymbol{n} ds.
\end{aligned}
$$

By replacing $u, \boldsymbol{l}, \boldsymbol{a}$ by their discrete versions $u_h \in V_h$; $\boldsymbol{l}_h, \boldsymbol{a}_h \in X_h$ in the volume integrals and by replacing u and $\boldsymbol{a}$ by $\hat{u}_h := \hat{u}(u_h)$ and $\hat{\boldsymbol{a}}_h := \hat{\boldsymbol{a}}(u_h, \boldsymbol{a}_h)$ in the surface integrals we get

$$
\begin{aligned}
&\int_K \boldsymbol{l}_h \cdot \boldsymbol{x}_h dx = -\int_K u_h \nabla \cdot \boldsymbol{x}_h dx + \int_{\delta K} \hat{u}_h \boldsymbol{x}_h \cdot \boldsymbol{n} ds, \\
&\int_K \boldsymbol{a}_h \cdot \boldsymbol{y}_h dx = \int_K A(\boldsymbol{l}_h) \cdot \boldsymbol{y}_h dx, \\
&\int_K u_{ht} z_h dx + \int_K \boldsymbol{a}_h \cdot \nabla z_h dx = \int_K f z_h dx + \int_{\delta K} z_h \hat{\boldsymbol{a}}_h \cdot \boldsymbol{n} ds.
\end{aligned}
\tag{5.4}
$$

Our definitions of the numerical fluxes are

$$
\hat{u}(u_h) := \begin{cases} \{u_h\} & , \Gamma \in \mathcal{F}_h^I, \\ u_D^* & , \Gamma \in \mathcal{F}_h^D, \\ u_h & , \Gamma \in \mathcal{F}_h^N, \end{cases}
$$

$$
\hat{\boldsymbol{a}}(u_h, \boldsymbol{a}_h) := \begin{cases} \{\boldsymbol{a}_h\} - \sigma A(h^{-1}[u_h]\boldsymbol{n}) & , \Gamma \in \mathcal{F}_h^I, \\ \boldsymbol{a}_h - \sigma A(h^{-1}(u - u_D^*)\boldsymbol{n}) & , \Gamma \in \mathcal{F}_h^D, \\ g_N \boldsymbol{n} & , \Gamma \in \mathcal{F}_h^N. \end{cases}
$$

Here $\sigma > 0$ is a constant and $u_D^* \in W^{1,\varphi}(\Omega)$ is an approximation of u_D. It will be defined either as u or $\Pi_{SZ} u$. The choice of u_D^* will be important for the error estimates. The parameter σ has the role of fine tuning the method during the implementation.

Definition 5.1. *The numerical fluxes $\hat{u}_h$and and $\hat{\boldsymbol{a}}_h$ are*

- *consistent, if* $\hat{u}_h(v) = v|_\Gamma$ *and* $\hat{\boldsymbol{a}}_h(v, \nabla v) = \nabla v|_\Gamma$ *for* $\Gamma \in \mathcal{F}_h$ *and* v *a smooth function satisfying the given boundary conditions.*
- *conservative, if* $\hat{u}_h(\cdot)$ *and* $\hat{\boldsymbol{a}}_h(\cdot,\cdot)$ *are single valued on* $\Gamma \in \mathcal{F}_h$.

We can easily see that our choice for $\hat{u}_h$ and $\hat{\boldsymbol{a}}_h$ is by definition both consistent and conservative, since jumps $[\cdot]$ of smooth functions vanish and both jumps and averages are single valued on the edges of the triangulation.

In order to obtain the formulation for the whole domain Ω, we want to sum trough all $K \in \mathcal{T}_h$. First we need to rewrite the surface terms similar to $\sum_{K\in\mathcal{T}_h} \int_{\delta K} v\boldsymbol{x}\cdot\boldsymbol{n}ds$, $v \in W^{1,\varphi}_{DG}(\Omega), \boldsymbol{x} \in W^{1,\varphi^*}_{DG}(\Omega, \mathbb{R}^d)$. For $\Gamma \in \mathcal{F}_h^I$ and K^+ and K^- from $\mathcal{T}_h$ such that the edge $\Gamma \subset K^+ \cap K^-$ we have

$$\begin{aligned}
\int_\Gamma (v|^+_\Gamma \boldsymbol{x}|^+_\Gamma \cdot \boldsymbol{n}^+ ds + \int_\Gamma v|^-_\Gamma \boldsymbol{x}|^-_\Gamma \cdot \boldsymbol{n}^- ds &= \int_\Gamma (v|^+_\Gamma \boldsymbol{x}|^+_\Gamma - v|^-_\Gamma \boldsymbol{x}|^-_\Gamma)\cdot \boldsymbol{n} ds \\
&= \int_\Gamma (1/2(v|^+_\Gamma \boldsymbol{x}|^+_\Gamma - v|^+_\Gamma \boldsymbol{x}|^-_\Gamma + v|^-_\Gamma \boldsymbol{x}|^+_\Gamma - v|^-_\Gamma \boldsymbol{x}|^-_\Gamma) \\
&\quad + 1/2(v|^+_\Gamma \boldsymbol{x}|^+_\Gamma + v|^+_\Gamma \boldsymbol{x}|^-_\Gamma - v|^-_\Gamma \boldsymbol{x}|^+_\Gamma - v|^-_\Gamma \boldsymbol{x}|^-_\Gamma))\cdot \boldsymbol{n} ds \\
= \int_\Gamma 1/2(v|^+_\Gamma + v|^-_\Gamma)(\boldsymbol{x}|^+_\Gamma - \boldsymbol{x}|^-_\Gamma)\cdot \boldsymbol{n} ds + \int_\Gamma (v|^+_\Gamma - v|^-_\Gamma) 1/2(\boldsymbol{x}|^+_\Gamma + \boldsymbol{x}|^-_\Gamma)\cdot \boldsymbol{n} ds \\
&= \int_\Gamma \{v\}[\boldsymbol{x}]\cdot\boldsymbol{n} + [v]\boldsymbol{n}\{\boldsymbol{x}\}\cdot\boldsymbol{n} ds. \quad (5.5)
\end{aligned}$$

Therefore, due to definition of jumps and averages on the boundary of the domain we can write

$$\begin{aligned}
\sum_{K\in\mathcal{T}_h} \int_{\delta K} v\boldsymbol{x}\cdot\boldsymbol{n} ds &= \int_{\mathcal{F}_h^I} \{v\}[\boldsymbol{x}]\cdot\boldsymbol{n} ds + \int_{\mathcal{F}_h^I} [v]\{\boldsymbol{x}\}\cdot\boldsymbol{n} ds + \int_{\mathcal{F}_h^{DN}} v\boldsymbol{x}\cdot\boldsymbol{n} ds \\
&= \int_{\mathcal{F}_h^I} \{v\}[\boldsymbol{x}]\cdot\boldsymbol{n} ds + \int_{\mathcal{F}_h^{ID}} [v]\{\boldsymbol{x}\}\cdot\boldsymbol{n} ds + \int_{\mathcal{F}_h^N} v\boldsymbol{x}\cdot\boldsymbol{n} ds. \quad (5.6)
\end{aligned}$$

Now we can sum (5.4) through all $K \in \mathcal{T}_h$ and use the definitions of the fluxes

$$\begin{aligned}
&\int_\Omega \boldsymbol{l}_h\cdot\boldsymbol{x}_h dx = -\int_\Omega u_h \nabla_h\cdot\boldsymbol{x}_h dx + \int_{\mathcal{F}_h^I} \{u_h\}[\boldsymbol{x}_h]\cdot\boldsymbol{n} ds + \int_{\mathcal{F}_h^D} u_D^* \boldsymbol{x}_h\cdot\boldsymbol{n} ds \\
&+ \int_{\mathcal{F}_h^N} u_h \boldsymbol{x}_h\cdot\boldsymbol{n} ds, \\
&\int_\Omega \boldsymbol{a}_h\cdot\boldsymbol{y}_h = \int_\Omega A(\boldsymbol{l}_h)\cdot\boldsymbol{y}_h dx, \\
&\int_\Omega u_{ht} z_h dx + \int_\Omega \boldsymbol{a}_h\cdot\nabla z_h dx = \int_\Omega f z_h dx + \int_{\mathcal{F}_h^I} \{\boldsymbol{a}_h\}\cdot[z_h]\boldsymbol{n} ds + \int_{\mathcal{F}_h^D} \boldsymbol{a}_h\cdot z_h\boldsymbol{n} ds \\
&+ \int_{\mathcal{F}_h^N} g_N z_h ds - \sigma\int_{\mathcal{F}_h^I} A(h^{-1}[u_h]\boldsymbol{n})\cdot[z_h]\boldsymbol{n} ds - \sigma\int_{\mathcal{F}_h^D} A(h^{-1}(u-u_D^*)\boldsymbol{n})z_h\cdot\boldsymbol{n} ds.
\end{aligned}$$

The equation (5.6) also implies the following version of integration by parts for our discontinuous functions

$$\int_\Omega \nabla_h u_h \boldsymbol{x}_h dx = -\int_\Omega u_h \nabla_h \cdot \boldsymbol{x}_h dx + \int_{\mathcal{F}_h^I} \{u_h\}[\boldsymbol{x}_h] \cdot \boldsymbol{n} ds + \int_{\mathcal{F}_h^I} [u_h]\{\boldsymbol{x}_h\} \cdot \boldsymbol{n} ds - \int_{\mathcal{F}_h^{DN}} u_h \boldsymbol{x}_h \cdot \boldsymbol{n} ds. \quad (5.7)$$

We can use this in the first equation to obtain the local DG formulation of our problem

$$\begin{aligned}
&\int_\Omega \boldsymbol{l}_h \cdot \boldsymbol{x}_h dx = \int_\Omega \nabla_h u_h \cdot \boldsymbol{x}_h dx + \int_{\mathcal{F}_h^I} [u_h]\boldsymbol{n} \cdot \{\boldsymbol{x}_h\} ds - \int_{\mathcal{F}_h^D} (u_h - u_D^*)\boldsymbol{x}_h \cdot \boldsymbol{n} ds, \\
&\int_\Omega \boldsymbol{a}_h \cdot \boldsymbol{y}_h dx = \int_\Omega A(\boldsymbol{l}_h) \cdot \boldsymbol{y}_h dx, \quad (5.8) \\
&\int_\Omega u_{ht} z_h dx + \int_\Omega \boldsymbol{a}_h \cdot \nabla z_h dx = \int_\Omega f z_h dx + \int_{\mathcal{F}_h^I} \{\boldsymbol{a}_h\} \cdot [z_h]\boldsymbol{n} ds + \int_{\mathcal{F}_h^D} \boldsymbol{a}_h \cdot z_h \boldsymbol{n} ds \\
&+ \int_{\mathcal{F}_h^N} g_N z_h ds - \sigma \int_{\mathcal{F}_h^I} A(h^{-1}[u_h]\boldsymbol{n}) \cdot [z_h]\boldsymbol{n} ds - \sigma \int_{\mathcal{F}_h^D} A(h^{-1}(u_h - u_D^*)\boldsymbol{n}) z_h \cdot \boldsymbol{n} ds.
\end{aligned}$$

5.2 The primal formulation

It will also be useful to have a discrete formulation in a single equation, called the primal formulation.

Recall that for $g \in V_h$ and $\boldsymbol{x}_h \in X_h$ we have by the definition of $\nabla_{DG}^h g$ and $\boldsymbol{R}_h g$

$$\int_\Omega \nabla_{DG}^h g \cdot \boldsymbol{x}_h dx = \int_\Omega \nabla_h g \cdot \boldsymbol{x}_h dx - \int_{\mathcal{F}_h^{ID}} [g]\boldsymbol{n} \cdot \{\boldsymbol{x_n}\} ds, \quad (5.9)$$

$$\int_\Omega \boldsymbol{R}_h g \cdot \boldsymbol{x}_h dx = \int_{\mathcal{F}_h^{ID}} [g]\boldsymbol{n} \cdot \{\boldsymbol{x_n}\} ds. \quad (5.10)$$

We can use this to eliminate the unknowns $\boldsymbol{l}_h$ and $\boldsymbol{a}_h$ in our equations

$$\begin{aligned}
&\int_\Omega \boldsymbol{l}_h \cdot \boldsymbol{x}_h dx = \int_\Omega (\nabla_{DG}^h u_h + \boldsymbol{R}_h u_D^*) \cdot \boldsymbol{x}_h dx, \\
&\int_\Omega \boldsymbol{a}_h \cdot \boldsymbol{y}_h dx = \int_\Omega A(\boldsymbol{l}_h) \cdot \boldsymbol{y}_h dx, \\
&\int_\Omega u_{ht} z_h dx + \int_\Omega \boldsymbol{a}_h \cdot \nabla_{DG}^h z_h dx = \int_\Omega f z_h dx + \int_{\mathcal{F}_h^N} g_N z_h ds \\
&- \sigma \int_{\mathcal{F}_h^I} A(h^{-1}[u_h]\boldsymbol{n}) \cdot [z_h]\boldsymbol{n} ds - \sigma \int_{\mathcal{F}_h^D} A(h^{-1}(u - u_D^*)\boldsymbol{n}) z_h \cdot \boldsymbol{n} ds.
\end{aligned}$$

This implies that

$$\boldsymbol{l}_h = \nabla_{DG}^h u_h + \boldsymbol{R}_h u_D^*, \quad (5.11)$$

$$\boldsymbol{a}_h = \Pi A(\boldsymbol{l}_h). \quad (5.12)$$

We can plug this into the third equation and obtain the primal formulation of the problem

$$\int_\Omega u_{ht} z_h dx + \int_\Omega A(\nabla^h_{DG} u_h + \boldsymbol{R}_h u^*_D) \cdot \nabla^h_{DG} z_h dx = \int_\Omega f z_h dx + \int_{\mathcal{F}^N_h} g_N z_h ds$$
$$- \sigma \int_{\mathcal{F}^I_h} A(h^{-1}[u_h]\boldsymbol{n}) \cdot [z_h]\boldsymbol{n} ds - \sigma \int_{\mathcal{F}^D_h} A(h^{-1}(u_h - u^*_D)\boldsymbol{n}) \cdot z_h \boldsymbol{n} ds. \quad (5.13)$$

By standard methods, it can be proven that the solution u_h exists. This also implies the existence of $\boldsymbol{l}_h$ and $\boldsymbol{a}_h$.

6. A priori stability estimates

Due to the complicated nature of our problem we are only going to present the a priori estimates in two simpler cases. In the first case we consider only trivial boundary conditions and in the second case we assume the time independent problem.

6.1 Estimate assuming $u_D^* = 0$ and $\Gamma_N = \emptyset$

In this case, we follow the approach from [11], considering that the main nonlinear function at the core of the equations here is a slightly more complicated and the estimate requires the theory from Chapter 2.

If we choose $z_h = u_h, \boldsymbol{x}_h = \boldsymbol{a}_h, \boldsymbol{y}_h = \boldsymbol{l}_h$ for our test function in (5.8), we get

$$\begin{aligned}
&\int_\Omega \boldsymbol{l}_h \cdot \boldsymbol{a}_h dx = \int_\Omega \nabla_h u_h \cdot \boldsymbol{a}_h dx + \int_{\mathcal{F}_h^I} [u_h]\boldsymbol{n} \cdot \{\boldsymbol{a}_h\} ds - \int_{\mathcal{F}_h^D} (u_h - u_D^*)\boldsymbol{a}_h \cdot \boldsymbol{n} ds,\\
&\int_\Omega \boldsymbol{a}_h \cdot \boldsymbol{l}_h dx = \int_\Omega A(\boldsymbol{l}_h) \cdot \boldsymbol{l}_h dx,\\
&\int_\Omega u_{ht} u_h dx + \int_\Omega \boldsymbol{a}_h \cdot \nabla u_h dx = \int_\Omega f u_h dx + \int_{\mathcal{F}_h^I} \{\boldsymbol{a}_h\} \cdot [u_h]\boldsymbol{n} ds + \int_{\mathcal{F}_h^D} \boldsymbol{a}_h \cdot u_h \boldsymbol{n} ds\\
&+ \int_{\mathcal{F}_h^N} g_N u_h ds - \sigma \int_{\mathcal{F}_h^I} A(h^{-1}[u_h]\boldsymbol{n}) \cdot [u_h]\boldsymbol{n} ds - \sigma \int_{\mathcal{F}_h^D} A(h^{-1}(u_h - u_D^*)\boldsymbol{n}) u_h \cdot \boldsymbol{n} ds.
\end{aligned}$$

If we assume that $u_D^* = 0$ and $\mathcal{F}_h^N = \emptyset$ the equations simplify to

$$\begin{aligned}
&\int_\Omega \boldsymbol{l}_h \cdot \boldsymbol{a}_h dx = \int_\Omega \nabla_h u_h \cdot \boldsymbol{a}_h dx + \int_{\mathcal{F}_h^I} [u_h]\boldsymbol{n} \cdot \{\boldsymbol{a}_h\} ds - \int_{\mathcal{F}_h^D} u_h \boldsymbol{a}_h \cdot \boldsymbol{n} ds,\\
&\int_\Omega \boldsymbol{a}_h \cdot \boldsymbol{l}_h dx = \int_\Omega A(\boldsymbol{l}_h) \cdot \boldsymbol{l}_h dx,\\
&\int_\Omega u_{ht} u_h dx + \int_\Omega \boldsymbol{a}_h \cdot \nabla u_h dx = \int_\Omega f u_h dx + \int_{\mathcal{F}_h^I} \{\boldsymbol{a}_h\} \cdot [u_h]\boldsymbol{n} ds + \int_{\mathcal{F}_h^D} \boldsymbol{a}_h \cdot u_h \boldsymbol{n} ds\\
&- \sigma \int_{\mathcal{F}_h^I} A(h^{-1}[u_h]\boldsymbol{n}) \cdot [u_h]\boldsymbol{n} ds - \sigma \int_{\mathcal{F}_h^D} A(h^{-1} u_h \boldsymbol{n}) u_h \cdot \boldsymbol{n} ds.
\end{aligned}$$

Now it its possible to eliminate the unknowns $\boldsymbol{l}_h$ and $\boldsymbol{a}_h$.Combining these equation, we arrive at

$$\int_\Omega u_{ht} u_h dx + \int_\Omega A(\boldsymbol{l}_h) \cdot \boldsymbol{l}_h dx + \sigma \int_{\mathcal{F}_h^{ID}} A(h^{-1}[u_h]\boldsymbol{n}) \cdot [u_h]\boldsymbol{n} ds = \int_\Omega f u_h dx. \quad (6.1)$$

Now we use the result concerning the relation between A and φ (2.47) in the following relations

$$A(\boldsymbol{l}_h) \cdot \boldsymbol{l}_h \sim \varphi(|\boldsymbol{l}_h|), \quad (6.2)$$

$$\sigma A(h^{-1}[u_h]\boldsymbol{n}) \cdot [u_h]\boldsymbol{n} \sim \sigma h \varphi(|h^{-1}[u_h]\boldsymbol{n}|), \quad (6.3)$$

We also use the Cauchy-Schwartz inequality on right hand side and the fact that

$$u_{ht}u_h = \frac{1}{2}\frac{\delta}{\delta t}u_h^2. \tag{6.4}$$

Putting all of these together we get

$$\frac{1}{2}\frac{\delta}{\delta t}\left\|u_h^2\right\|_{L^2(\Omega)}^2 + \int_\Omega \varphi(|\boldsymbol{l}|_h)dx + \sigma h\int_\Omega \varphi(h^{-1}[u_h]\boldsymbol{n})dx \leq \frac{1}{2}\left\|f\right\|_{L^2(\Omega)}^2 + \frac{1}{2}\left\|u_h\right\|_{L^2(\Omega)}^2. \tag{6.5}$$

By integrating from 0 to $t > 0$ and multiplying by 2 we get

$$\begin{aligned}&\|u_h(t)\|_{L^2(\Omega)}^2 + 2\int_0^t(\int_\Omega(\varphi(|\boldsymbol{l}|_h(\tau))dx + \sigma h\int_\Omega \varphi(h^{-1}[u_h(\tau)]\boldsymbol{n})dx)d\tau \\ &\leq \|u_h(0)\|_{L^2(\Omega)}^2 + \int_0^t(\|f(\tau)\|_{L^2(\Omega)}^2 + \|u_h(\tau)\|_{L^2(\Omega)}^2)d\tau.\end{aligned} \tag{6.6}$$

Here we will need the following version of the Gronwall inequality

Lemma 6.1. *Let $y : [0,T] \to \mathbb{R}$ be a nonnegative, measurable function, $r : [0,T] \to \mathbb{R}$ a nonnegative integrable function and $q, z \geq 0$ .If the following inequality holds for $t \in [0,T]$*

$$y(t) + q \leq z + \int_0^t r(s)y(s)ds, \tag{6.7}$$

then also

$$y(t) + q \leq z\, exp\Big(\int_0^t r(s)dsd\tau\Big). \tag{6.8}$$

Applying this to (6.6) we get the final estimate

Theorem 6.1. *Let $(u_h, \boldsymbol{l}_h, \boldsymbol{a}_h) \in V_h \times X_h \times X_h$ be a solution to (5.8) for some $\sigma > 0$, while $u_D^* = u$ or $u_D^* = \Pi_{SZ}u$. Then this solution satisfies*

$$\begin{aligned}&\|u_h(t)\|_{L^2(\Omega)}^2 + 2\int_0^t(\int_\Omega(\varphi(|\boldsymbol{l}|_h(\tau))dx + \sigma h\int_\Omega \varphi(h^{-1}[u_h(\tau)]\boldsymbol{n})dx)d\tau \\ &\leq e^t\Big(\|u_h(0)\|_{L^2(\Omega)}^2 + \int_0^t \|f(\tau)\|_{L^2(\Omega)}^2 d\tau\Big).\end{aligned}$$

6.2 Estimate assuming time independent problem

Stability estimate in the stationary case is based on [10, Theorem 3.2]. We start with the primal formulation

$$\begin{aligned}\int_\Omega A(\nabla_{DG}^h u_h + \boldsymbol{R}_h u_D^*)\cdot\nabla_{DG}^h z_h dx &= \int_\Omega f z_h dx + \int_{\mathcal{F}_h^N} g_N z_h ds \\ &- \sigma\int_{\mathcal{F}_h^I} A(h^{-1}[u_h]\boldsymbol{n})\cdot[z_h]\boldsymbol{n}ds - \sigma\int_{\mathcal{F}_h^D} A(h^{-1}(u_h - u_D^*)\boldsymbol{n})\cdot z_h\boldsymbol{n}ds.\end{aligned} \tag{6.9}$$

In this case we choose the trial function differently as $z_h = u_h - \Pi u$. It will also be useful to split the first term using the following.

$$\nabla^h_{DG}(u_h - \Pi u) = \nabla^h_{DG} u_h - \nabla^h_{DG}\Pi u = \nabla^h_{DG} u_h + R_h u^*_D - R_h u^*_D - \nabla_h \Pi u + R_h \Pi u. \tag{6.10}$$

By substituting for z_h and rearranging the terms we get

$$\begin{aligned}
&\int_\Omega A(\nabla^h_{DG} u_h + R_h u^*_D) \cdot (\nabla^h_{DG} u_h + R_h u^*_D) - \sigma \int_{\mathcal{F}^I_h} A(h^{-1}[u_h]\boldsymbol{n}) \cdot [u_h]\boldsymbol{n} ds \qquad (6.11)\\
&- \sigma \int_{\mathcal{F}^D_h} A(h^{-1}(u_h - u^*_D)\boldsymbol{n}) \cdot (u_h - u^*_D)\boldsymbol{n} ds\\
&= \int_\Omega A(\nabla^h_{DG} u_h + R_h u^*_D) \cdot (\nabla_h \Pi u) + \int_\Omega A(\nabla^h_{DG} u_h + R_h u^*_D) \cdot (R_h(u^*_D - \Pi u))\\
&+ \int_\Omega f(u_h - \Pi u) dx + \int_{\mathcal{F}^N_h} g_N(u_h - \Pi u) dx + \sigma \int_{\mathcal{F}^I_h} A(h^{-1}[u_h]\boldsymbol{n}) \cdot [\Pi u]\boldsymbol{n} ds\\
&- \sigma \int_{\mathcal{F}^D_h} A(h^{-1}(u_h - u^*_D)\boldsymbol{n}) \cdot (u^*_D - \Pi u)\boldsymbol{n} ds =: I_1 + I_2 + I_3 + I_4 + I_5 + I_6.
\end{aligned}$$

First we treat the left hand side. Using (5.11) and the fact that

$$A(\boldsymbol{l}_h) \cdot \boldsymbol{l}_h \sim \varphi(|\boldsymbol{l}_h|), \tag{6.12}$$

$$A(\nabla^h_{DG} u_h + R_h u^*_D) \cdot (\nabla^h_{DG} u_h + R_h u^*_D) \sim \varphi(|\nabla^h_{DG} u_h + R_h u^*_D|), \tag{6.13}$$

$$A(h^{-1}[u_h]\boldsymbol{n}) \cdot [u_h]\boldsymbol{n} \sim h\varphi(|h^{-1}[u_h]\boldsymbol{n}|), \tag{6.14}$$

$$A(h^{-1}(u_h - u^*_D)\boldsymbol{n}) \cdot (u_h - u^*_D)\boldsymbol{n} \sim h\varphi(|h^{-1}(u_h - u^*_D)\boldsymbol{n}|). \tag{6.15}$$

due to (2.47) we can see that the left hand side is equivalent to

$$\begin{aligned}
&\int_\Omega \varphi(|\boldsymbol{l}_h|) + \varphi(|\nabla^h_{DG} u_h + R_h u^*_D|) dx + \sigma h \int_{\mathcal{F}^I_h} \varphi(|h^{-1}[u_h]\boldsymbol{n}|) ds\\
&+ \sigma h \int_{\mathcal{F}^D_h} \varphi(|h^{-1}(u_h - u^*_D)\boldsymbol{n}|) ds\\
&= \int_\Omega \varphi(|\boldsymbol{l}_h|) + \varphi(|\nabla^h_{DG} u_h + R_h u^*_D|) dx + \sigma h \int_{\mathcal{F}^{ID}_h} \varphi(|h^{-1}[u_h - u^*_D]\boldsymbol{n}|) ds,
\end{aligned}$$

since $[u^*_D]\boldsymbol{n} = 0$ on $\Gamma \in \mathcal{F}^I_h$.

It is possible to estimate few other terms on the left hand side with a clever use of the previously derived results.

- Using (2.11), (2.48), (4.29) and (5.12) in this order for each of the following inequalities

$$\begin{aligned}
&\int_\Omega \varphi(|\boldsymbol{l}_h|) \sim \int_\Omega \varphi^*(\varphi'(|\boldsymbol{l}_h|)) \sim \int_\Omega \varphi^*(A(|\boldsymbol{l}_h|))\\
&\geq c \int_\Omega \varphi^*(\Pi A(|\boldsymbol{l}_h|)) = \int_\Omega \varphi^*(\boldsymbol{a}_h).
\end{aligned}$$

- From (5.11) we have

$$\boldsymbol{l}_h = \nabla^h_{DG} u_h + \boldsymbol{R}_h u^*_D = \nabla_h u_h - \boldsymbol{R}_h(u_h - u^*_D). \tag{6.16}$$

This implies

$$\begin{aligned}\int_\Omega \varphi(|\nabla_h u_h|) &\le c \int_\Omega \varphi(|\boldsymbol{l}_h|) + \varphi(|\boldsymbol{R}_h(u_h - u^*_D)|)dx \\ &\le c \int_\Omega \varphi(|\boldsymbol{l}_h|)dx + ch \int_{\mathcal{F}_h^{ID}} \varphi(|h^{-1}[u_h - u^*_D]\boldsymbol{n}|),\end{aligned}$$

where in the last inequality (4.13) was used.

- Using the fact that $\int_\Omega \varphi(|\nabla_h u_h|)$ is already controlled, another term can be estimated with the use of the following lemma.

Lemma 6.2. *For $u^*_D = u$ or $u^*_D = \Pi_{SZ} u$ we have*

$$h \int_{\mathcal{F}_h^{ID}} \varphi(|h^{-1}[u_h - u^*_D]|)ds \le c \int_\Omega \varphi(|\nabla u|)dx. \tag{6.17}$$

Proof. If $u^*_D = u$ then the left hand side is 0. In the other case the assertion follows from the estimate of the first term in (4.43). □

$$\begin{aligned}M_{\varphi,h}(u_h - u) &= \int_\Omega \varphi(|\nabla_h(u_h - u)|)dx + h \int_{\mathcal{F}_h^{ID}} \varphi(|h^{-1}[u_h - u]\boldsymbol{n}|)ds \\ &= \int_\Omega \varphi(|\nabla_h u_h - \nabla u|)dx + h \int_{\mathcal{F}_h^{ID}} \varphi(|h^{-1}[u_h - u^*_D + u^*_D - u]\boldsymbol{n}|)ds \\ &\le c \int_\Omega \varphi(|\nabla_h u_h|) + \varphi(|\nabla u|)dx \\ &+ ch \int_{\mathcal{F}_h^{ID}} \varphi(|h^{-1}[u_h - u^*_D]|) + \varphi(|h^{-1}[u^*_D - u]\boldsymbol{n}|)ds \\ &\le c \int_\Omega \varphi(|\nabla_h u_h|) + \varphi(|\nabla u|)dx + ch \int_{\mathcal{F}_h^{ID}} \varphi(|h^{-1}[u_h - u^*_D]|)ds.\end{aligned}$$

Here 6.2 was used in the last inequality.

$$\begin{aligned}M_{\varphi,h}(u_h - u) - c \int_\Omega \varphi(|\nabla u|)dx \\ \le c \int_\Omega \varphi(|\nabla_h u_h|)dx + ch \int_{\mathcal{F}_h^{ID}} \varphi(|h^{-1}[u_h - u^*_D]|)ds.\end{aligned} \tag{6.18}$$

- Finally we can estimate one last term, using lemma 4.13

$$M_{\varphi,h}(u_h - u) \ge c M_{\varphi,h}(diam(\Omega)(u_h - u)) \ge c \int_\Omega \varphi(|u_h - u|)dx. \tag{6.19}$$

Putting everything together, the left hand side is greater or equal to

$$\begin{aligned} & c(\int_\Omega \varphi(|\boldsymbol{l}_h|)dx + \varphi(|\nabla^h_{DG}u_h + R_h u^*_D|)dx \\ & +\sigma h \int_{\mathcal{F}_h^{ID}} \varphi(|h^{-1}[u_h - u^*_D]\boldsymbol{n}|)ds + \int_\Omega \varphi^*(\boldsymbol{a}_h)ds \\ & +min\{1,\sigma\}(\int_\Omega \varphi(|\nabla_h u_h|)dx + M_{\varphi,h}(u_h - u) \\ & - \int_\Omega \varphi(|\nabla u|)dx + \int_\Omega \varphi(|u_h - u|)dx)). \end{aligned} \tag{6.20}$$

To deal with the right hand side of (6.11), we estimate the terms $I_1 \dots I_6$ one by one using the modified Young's inequality, i.e. lemma 2.1 to split the integrals and move the terms multiplied by ϵ to the left hand side.

- For the first integral we have

$$\begin{aligned} |I_1| &\leq \epsilon \int_\Omega \varphi^*(|A(\nabla^h_{DG}u_h + R_h u^*_D)|) + c_\epsilon \int_\Omega \varphi(|\nabla_h \Pi u|) \\ &\leq \epsilon \int_\Omega \varphi^*(\varphi'(|\nabla^h_{DG}u_h + R_h u^*_D)|)) + c_\epsilon \int_\Omega \varphi(|\nabla_h \Pi u|) \\ &\leq \epsilon \int_\Omega \varphi(|\nabla^h_{DG}u_h + R_h u^*_D)|) + c_\epsilon \int_\Omega \varphi(|\nabla_h \Pi u|), \end{aligned}$$

  using (2.48) in the first inequality and (2.11) and (4.30) in the second.

- After the same two steps, we have

$$\begin{aligned} |I_2| &\leq \epsilon \int_\Omega \varphi(|\nabla^h_{DG}u_h + R_h u^*_D|)dx + c_\epsilon \int_\Omega \varphi(|R_h(u^*_D - \Pi u)|)dx \\ &\leq \epsilon \int_\Omega \varphi(|\nabla^h_{DG}u_h + R_h u^*_D|)dx + c_\epsilon h \int_{\mathcal{F}_h^{ID}} \varphi(|h^{-1}[u^*_D - \Pi u]\boldsymbol{n}|)ds \\ &\leq \epsilon \int_\Omega \varphi(|\nabla^h_{DG}u_h + R_h u^*_D|)dx \\ &+ c_\epsilon h \int_{\mathcal{F}_h^{ID}} \varphi(|h^{-1}[u^*_D - u]\boldsymbol{n}|) + \varphi(|h^{-1}[u - \Pi u]\boldsymbol{n}|)ds \\ &\leq \epsilon \int_\Omega \varphi(|\nabla^h_{DG}u_h + R_h u^*_D|)dx + c_\epsilon \int_\Omega \varphi(|\nabla u|), \end{aligned}$$

  using additionally (4.13) in the second inequality and 4.3, (4.34) in the fourth.

-

$$\begin{aligned} |I_3| &\leq \epsilon \int_\Omega \varphi(|u_h - u|) + \varphi(|u - \Pi u|)dx + c_\epsilon \int_\Omega \varphi^*(|f|)dx \\ &\leq \epsilon \int_\Omega \varphi(|u_h - u|)dx + c \int_\Omega \varphi(|h\nabla u|)dx + c_\epsilon \int_\Omega \varphi^*(|f|)dx \\ &\leq \epsilon \int_\Omega \varphi(|u_h - u|)dx + c \int_\Omega \varphi(|\nabla u|)dx + c_\epsilon \int_\Omega \varphi^*(|f|)dx, \end{aligned}$$

  using (4.27) in the second inequality.

•

$$
\begin{aligned}
|I_4| &\leq c_\epsilon \int_{\mathcal{F}_h^N} \varphi^*(|g_N|)dx + \epsilon \int_{\mathcal{F}_h^N} \varphi(|u_h - \Pi u|)dx \\
&\leq c_\epsilon \int_{\mathcal{F}_h^N} \varphi^*(|g_N|)dx + \epsilon \int_{\mathcal{F}_h^N} \varphi(|u_h - u|) + \varphi(|u - \Pi u|)dx \\
&\leq c_\epsilon \int_{\mathcal{F}_h^N} \varphi^*(|g_N|)dx + \epsilon(M_{\varphi,h}(u_h - u) + M_{\varphi,h}(u - \Pi u)) \\
&\leq c_\epsilon \int_{\mathcal{F}_h^N} \varphi^*(|g_N|)dx + \epsilon M_{\varphi,h}(u_h - u) + \epsilon \int_\Omega \varphi(|\nabla u|)dx,
\end{aligned}
$$

using lemma 4.14 in the third inequality and (4.36) in the fourth.

- It is possible to combine the estimate of I_5 and I_6 using that $[u_D^*]\boldsymbol{n} = 0$.

$$
\begin{aligned}
|I_5 + I_6| &= \sigma h| \int_{\mathcal{F}_h^{ID}} A(h^{-1}[u_h - u_D^*]\boldsymbol{n})h^{-1}[u_D^* - \Pi u]\boldsymbol{n}ds| \\
&\leq \epsilon\alpha h \int_{\mathcal{F}_h^{ID}} \varphi(|h^{-1}[u_h - u_D^*]|)ds + c_\epsilon \alpha h \int_{\mathcal{F}_h^{ID}} \varphi(|u_D^* - \Pi u|)ds \\
&\leq \epsilon\alpha h \int_{\mathcal{F}_h^{ID}} \varphi(|h^{-1}[u_h - u_D^*]|)ds \\
&+ c_\epsilon \alpha h \int_{\mathcal{F}_h^{ID}} \varphi(|u_D^* - u|)ds + c_\epsilon \alpha h \int_{\mathcal{F}_h^{ID}} \varphi(|u - \Pi u|)ds \\
&\leq \epsilon\alpha h \int_{\mathcal{F}_h^{ID}} \varphi(|h^{-1}[u_h - u_D^*]|)ds + c_\epsilon \alpha \int_\Omega \varphi(|\nabla u|)dx,
\end{aligned}
$$

using (2.48) and Young inequality in the first estimate, (2.11) in the second and (4.34) in the last inequality similarly to I_2 estimate.

By choosing ϵ sufficiently small we can put all the terms multiplied by ϵ to the left hand side and obtain the final result

Theorem 6.2. *Let $u_h \in V_h^k, \boldsymbol{l}_h \in X_h^k, \boldsymbol{a}_h \in X_h^k$ be the DG solution of (6.9) and $u_D^* = u$ or $u_D^* = \Pi_{SZ} u$. Then for $\sigma > 0$ we have the a priori estimate*

$$
\begin{aligned}
&\int_\Omega \varphi(|\boldsymbol{l}_h|) + \varphi(|\nabla_{DG}^h u_h + R_h u_D^*|)dx \\
&+ \sigma h \int_{\mathcal{F}_h^{ID}} \varphi(|h^{-1}[u_h - u_D^*]\boldsymbol{n}|)ds + \int_\Omega \varphi^*(\boldsymbol{a}_h)ds \\
&+ min\{1, \sigma\}(\int_\Omega \varphi(|\nabla_h u_h|)dx + M_{\varphi,h}(u_h - u) + \int_\Omega \varphi(|u_h - u|)dx) \\
&\leq c(\int_\Omega \varphi^*(|f|)dx + \int_{\mathcal{F}_h^N} \varphi^*(|g_N|)dx + \int_\Omega \varphi(|\nabla u|)dx).
\end{aligned}
$$

7. A priori error estimates

In this chapter we derive the a priori error estimates for our LDG method, inspired by the works [9] and [10].

7.1 Time independent problem

Let us start for simplicity with the stationary case, following similar steps as in [10]. First we want to derive an equation similar to the LDG formulation for the exact solution. We begin with the original equations

$$\begin{aligned} \boldsymbol{l} &= \nabla u, \\ \boldsymbol{a} &= A(\boldsymbol{l}), \\ -\nabla \cdot \boldsymbol{a} &= f. \end{aligned} \tag{7.1}$$

Using the standard procedure we multiply the third equation by $z_h, \in V_h^k$, integrate over Ω and use integration by parts combined with (5.6) and the fact that $[\boldsymbol{a}] = 0$.

$$\begin{aligned} &\int_\Omega -\nabla \cdot \boldsymbol{a} z_h dx = \int_\Omega f z_h dx, \\ &\int_\Omega \boldsymbol{a} \cdot \nabla_h z_h dx - \int_{\mathcal{F}_h^{ID}} \{\boldsymbol{a}\} \cdot [z_h] \boldsymbol{n} ds - \int_{\mathcal{F}_h^N} \boldsymbol{a} \cdot z_h \boldsymbol{n} ds = \int_\Omega f z_h dx, \\ &\int_\Omega \boldsymbol{a} \cdot \nabla_h z_h dx - \int_{\mathcal{F}_h^{ID}} \{\boldsymbol{a}\} \cdot [z_h] \boldsymbol{n} ds - \int_{\mathcal{F}_h^N} g_N z_h ds = \int_\Omega f z_h dx. \end{aligned}$$

It is beneficial to rewrite this in terms of ∇_{DG}^h. Using the properties of the projection Π, we have

$$\begin{aligned} &\int_\Omega \boldsymbol{a} \cdot \nabla_h z_h dx = \int_\Omega \Pi \boldsymbol{a} \cdot \nabla_h z_h dx \\ &= \int_\Omega \Pi \boldsymbol{a} \cdot \nabla_{DG}^h z_h dx + \int_\Omega \Pi \boldsymbol{a} \cdot R_h z_h dx \\ &= \int_\Omega \Pi \boldsymbol{a} \cdot \nabla_{DG}^h z_h dx + \int_{\mathcal{F}_h^{ID}} \{\Pi \boldsymbol{a}\} \cdot [z_h] \boldsymbol{n} ds. \end{aligned}$$

Using this together with the first and second equation multiplied by appropriate test functions we arrive at the following formulation of the original problem, satisfied by the exact solution $(u, \boldsymbol{l}, \boldsymbol{a})$, for all $\boldsymbol{x}_h, \boldsymbol{y}_h \in X_h^k, z_h \in V_h^k$.

$$\begin{aligned} &\int_\Omega \boldsymbol{l} \cdot \boldsymbol{x}_h dx = \int_\Omega \nabla u \cdot \boldsymbol{x}_h dx, \\ &\int_\Omega \boldsymbol{a} \cdot \boldsymbol{y}_h dx = \int_\Omega A(\boldsymbol{l}) \cdot \boldsymbol{y}_h dx, \\ &\int_\Omega \boldsymbol{a} \cdot \nabla_{DG}^h z_h dx = \int_\Omega f z_h dx + \int_{\mathcal{F}_h^N} g_N z_h ds + \int_{\mathcal{F}_h^{ID}} (\{\boldsymbol{a}\} - \{\Pi \boldsymbol{a}\}) \cdot [z_h] \boldsymbol{n} ds. \end{aligned} \tag{7.2}$$

We can use this together with the primal formulation (6.9), which reads

$$\int_\Omega A(\nabla^h_{DG} u_h + \boldsymbol{R}_h u^*_D) \cdot \nabla^h_{DG} z_h dx = \int_\Omega f z_h dx + \int_{\mathcal{F}^N_h} g_N z_h ds$$
$$- \sigma \int_{\mathcal{F}^I_h} A(h^{-1}[u_h]\boldsymbol{n}) \cdot [z_h]\boldsymbol{n} ds - \sigma \int_{\mathcal{F}^D_h} A(h^{-1}(u_h - u^*_D)\boldsymbol{n}) \cdot z_h \boldsymbol{n} ds. \quad (7.3)$$

In order to obtain the error equation, we subtract the third equation of (7.2) from the primal formulation.

$$\int_\Omega (A(\nabla^h_{DG} u_h + R_h u^*_D) - A(\nabla u)) \cdot \nabla^h_{DG} z_h dx + \sigma \int_{\mathcal{F}^I_h} A(h^{-1}[u_h]\boldsymbol{n}) \cdot [z_h]\boldsymbol{n} ds$$
$$+ \sigma \int_{\mathcal{F}^D_h} A(h^{-1}(u_h - u^*_D)\boldsymbol{n}) \cdot z_h \boldsymbol{n} ds$$
$$= \int_{\mathcal{F}^{ID}_h} (\{\Pi \boldsymbol{a}\} - \{\boldsymbol{a}\}) \cdot [z_h]\boldsymbol{n} ds.$$

For $w_h \in V^k_h$ we choose the test functions as $z_h = u_h - w_h$. We can rewrite $\nabla^h_{DG} z_h$ as follows

$$\nabla^h_{DG} z_h = (\nabla^h_{DG} u_h + R_h u^*_D - \nabla u) - (\nabla^h_{DG} w_h + R_h u^*_D - \nabla u). \quad (7.4)$$

With this we can rewrite the error equation as

$$\int_\Omega (A(\nabla^h_{DG} u_h + R_h u^*_D) - A(\nabla u)) \cdot (\nabla^h_{DG} u_h + R_h u^*_D - \nabla u) dx$$
$$+ \sigma \int_{\mathcal{F}^I_h} A(h^{-1}[u_h]\boldsymbol{n}) \cdot [u_h]\boldsymbol{n} ds + \sigma \int_{\mathcal{F}^D_h} A(h^{-1}(u_h - u^*_D)\boldsymbol{n}) \cdot (u_h - u^*_D)\boldsymbol{n} ds$$
$$= \int_\Omega (A(\nabla^h_{DG} u_h + R_h u^*_D) - A(\nabla u)) \cdot (\nabla^h_{DG} w_h + R_h u^*_D - \nabla u) dx$$
$$+ \sigma \int_{\mathcal{F}^I_h} A(h^{-1}[u_h]\boldsymbol{n}) \cdot [w_h]\boldsymbol{n} ds + \sigma \int_{\mathcal{F}^D_h} A(h^{-1}(u_h - u^*_D)\boldsymbol{n}) \cdot (w_h - u^*_D)\boldsymbol{n} ds$$
$$- \int_{\mathcal{F}^I_h} (\{\Pi \boldsymbol{a}\} - \{\boldsymbol{a}\}) \cdot [w_h]\boldsymbol{n} ds - \int_{\mathcal{F}^D_h} (\{\Pi \boldsymbol{a}\} - \{\boldsymbol{a}\}) \cdot [w_h - u^*_D]\boldsymbol{n} ds$$
$$+ \int_{\mathcal{F}^I_h} (\{\Pi \boldsymbol{a}\} - \{\boldsymbol{a}\}) \cdot [u_h]\boldsymbol{n} d + \int_{\mathcal{F}^D_h} (\{\Pi \boldsymbol{a}\} - \{\boldsymbol{a}\}) \cdot [u_h - u^*_D]\boldsymbol{n} ds,$$

using that $[u^*_h] = 0$ on Γ_D. To estimate the left hand side we use the equivalence results for the N-function φ (2.45) and (2.47)

$$\int_\Omega (A(\nabla^h_{DG} u_h + R_h u^*_D) - A(\nabla u)) \cdot (\nabla^h_{DG} u_h + R_h u^*_D - \nabla u) dx$$
$$\sim \left\| F(\nabla^h_{DG} u_h + R_h u^*_D) - F(\nabla u) \right\|^2_{L^2(\Omega)},$$
$$\int_{\mathcal{F}^I_h} A(h^{-1}[u_h]\boldsymbol{n}) \cdot [u_h]\boldsymbol{n} ds \sim h \int_{\mathcal{F}^I_h} \varphi(|h^- 1[u_h]\boldsymbol{n}|) ds,$$
$$\sigma \int_{\mathcal{F}^D_h} A(h^{-1}(u_h - u^*_D)\boldsymbol{n}) \cdot (u_h - u^*_D)\boldsymbol{n} ds \sim \sigma h \int_{\mathcal{F}^I_h} \varphi(|h^- 1(u_h - u^*_D)\boldsymbol{n}|) ds.$$

Let us now estimate the right hand side. In order to deal with the first term we need the Young inequality, i.e. lemma 2.1, for N-functions $\varphi_{|\nabla u|}$ and (2.44), (2.45), (2.11) and (2.45) again.

$$
\begin{aligned}
&\int_\Omega (A(\nabla^h_{DG} u_h + R_h u^*_D) - A(\nabla u)) \cdot (\nabla^h_{DG} w_h + R_h u^*_D - \nabla u) dx \\
&\leq \epsilon \int_\Omega \varphi^*_{|\nabla u|}(|A(\nabla^h_{DG} + R_h u^*_D) - A(\nabla u)|) dx \\
&+ c_\epsilon \int_\Omega \varphi_{|\nabla u|}(|\nabla^h_{DG} + R_h u^*_D - \nabla u|) dx \\
&\leq c\epsilon \int_\Omega \varphi^*_{|\nabla u|}(\varphi'_{|\nabla u|}(|\nabla^h_{DG} + R_h u^*_D - \nabla u|)) dx \\
&+ c_\epsilon \left\| F(\nabla^h_{DG} w_h + R_h u^*_D) - F(\nabla u) \right\|^2_{L^2(\Omega)} \\
&\leq c\epsilon \int_\Omega \varphi_{|\nabla u|}(|\nabla^h_{DG} + R_h u^*_D - \nabla u|) dx + c_\epsilon \left\| F(\nabla^h_{DG} w_h + R_h u^*_D) - F(\nabla u) \right\|^2_{L^2(\Omega)} \\
&\leq c\epsilon \left\| F(\nabla^h_{DG} u_h + R_h u^*_D) - F(\nabla u) \right\|^2_{L^2(\Omega)} \\
&+ c_\epsilon \left\| F(\nabla^h_{DG} w_h + R_h u^*_D) - F(\nabla u) \right\|^2_{L^2(\Omega)}.
\end{aligned}
$$

Using again Young inequality for φ, (2.48) and (2.11) in the estimate of the next two terms, we obtain

$$
\begin{aligned}
&\int_{\mathcal{F}^I_h} A(h^{-1}[u_h]\boldsymbol{n}) \cdot [w_h]\boldsymbol{n} ds + \int_{\mathcal{F}^D_h} A(h^{-1}(u_h - u^*_D)\boldsymbol{n}) \cdot (w_h - u^*_D)\boldsymbol{n} ds \\
&\leq \epsilon(h \int_{\mathcal{F}^I_h} \varphi(|h^{-1}[u_h]\boldsymbol{n}|) ds + h \int_{\mathcal{F}^D_h} \varphi(|h^{-1}(u_h - u^*_D)\boldsymbol{n}|) ds) \\
&+ c_\epsilon(h \int_{\mathcal{F}^I_h} \varphi(|h^{-1}[w_h]\boldsymbol{n}|) ds + h \int_{\mathcal{F}^D_h} \varphi(|h^{-1}(w_h - u^*_D)\boldsymbol{n}|) ds).
\end{aligned}
$$

The last four terms are dealt with only using Young inequality for φ, but taking care, which term on the right hand side gets multiplied by ϵ.

$$
\begin{aligned}
&\int_{\mathcal{F}^I_h} (\{\Pi \boldsymbol{a}\} - \{\boldsymbol{a}\}) \cdot [w_h]\boldsymbol{n} ds = h \int_{\mathcal{F}^I_h} (\{\Pi \boldsymbol{a}\} - \{\boldsymbol{a}\}) \cdot h^{-1}[w_h]\boldsymbol{n} ds \\
&\leq \epsilon h \int_{\mathcal{F}^I_h} \varphi^*(|\{\Pi \boldsymbol{a}\} - \{\boldsymbol{a}\}|) ds + c_\epsilon h \int_{\mathcal{F}^I_h} \varphi(|h^{-1}[w_h]\boldsymbol{n}|) ds \\
&\int_{\mathcal{F}^I_h} (\{\Pi \boldsymbol{a}\} - \{\boldsymbol{a}\}) \cdot (w_h - u^*_D)\boldsymbol{n} ds \\
&\leq \epsilon h \int_{\mathcal{F}^I_h} \varphi^*(|\{\Pi \boldsymbol{a}\} - \{\boldsymbol{a}\}|) ds + c_\epsilon h \int_{\mathcal{F}^I_h} \varphi(|h^{-1}(w_h - u^*_D)\boldsymbol{n}|) ds \\
&\int_{\mathcal{F}^I_h} (\{\Pi \boldsymbol{a}\} - \{\boldsymbol{a}\}) \cdot [u_h]\boldsymbol{n} ds \\
&\leq c_\epsilon h \int_{\mathcal{F}^I_h} \varphi^*(|\{\Pi \boldsymbol{a}\} - \{\boldsymbol{a}\}|) ds + \epsilon h \int_{\mathcal{F}^I_h} \varphi(|h^{-1}[u_h]\boldsymbol{n}|) ds \\
&\int_{\mathcal{F}^I_h} (\{\Pi \boldsymbol{a}\} - \{\boldsymbol{a}\}) \cdot (u_h - u^*_D)\boldsymbol{n} ds \\
&\leq c_\epsilon h \int_{\mathcal{F}^I_h} \varphi^*(|\{\Pi \boldsymbol{a}\} - \{\boldsymbol{a}\}|) ds + \epsilon h \int_{\mathcal{F}^I_h} \varphi(|h^{-1}(u_h - u^*_D)\boldsymbol{n}|) ds.
\end{aligned}
$$

Putting everything together and transferring the terms on the right hand side multiplied by ϵ, for ϵ small enough we arrive at the following theorem.

Theorem 7.1. *Let $u \in W^{1,\varphi}(\Omega)$, $\boldsymbol{l} \in L^{\varphi}(\Omega)$, $\boldsymbol{a} \in W^{1,\varphi^*}(\Omega)$ be the solution of the problem (5.3) and $u_h, \boldsymbol{l_h}, \boldsymbol{a_h}$ the DG solution of (6.9), from V_h^k, X_h^k, X_h^k respectively, then for all $w_h \in V_h^k$ it holds*

$$\begin{aligned}
&\left\|F(\nabla^h_{DG}u_h + R_h u_D^*) - F(\nabla u)\right\|^2_{L^2(\Omega)} \\
&+ \sigma h \int_{\mathcal{F}_h^I} \varphi(|h^{-}1[u_h]\boldsymbol{n}|)ds + \sigma h \int_{\mathcal{F}_h^D} \varphi(|h^{-}1(u_h - u_D^*)\boldsymbol{n}|)ds \\
&\leq c(\left\|F(\nabla^h_{DG}w_h + R_h u_D^*) - F(\nabla u)\right\|^2_{L^2(\Omega)} \\
&+ \sigma h \int_{\mathcal{F}_h^I} \varphi(|h^{-}1[w_h]\boldsymbol{n}|)ds + \sigma h \int_{\mathcal{F}_h^D} \varphi(|h^{-}1(w_h - u_D^*)\boldsymbol{n}|)ds \\
&+ ch \int_{\mathcal{F}_h^{ID}} \varphi^*(|\{\Pi\boldsymbol{a}\} - \{\boldsymbol{a}\}|)ds.
\end{aligned}$$

- From the derivation of the primal formulation (5.11) we have the following equality for the first term on the left hand side

$$\left\|F(\nabla^h_{DG}u_h + R_h u_D^*) - F(\nabla u)\right\|^2_{L^2(\Omega)} = \|F(\boldsymbol{l_h}) - F(\boldsymbol{l})\|^2_{L^2(\Omega)}\,. \tag{7.5}$$

- We are also able to include the estimate of the error between $\boldsymbol{a}$ and $\boldsymbol{a}_h$ expressed in the form $\|F^*(\boldsymbol{a}) - F^*(\boldsymbol{a}_h)\|^2_{L^2(\Omega)}$. From (2.46) and (5.12) we have

$$\begin{aligned}
&\|F^*(\boldsymbol{a}) - F^*(\boldsymbol{a}_h)\|^2_{L^2(\Omega)} \leq c \sum_{K\in\mathcal{T}_h} \int_K \varphi^*_{|\boldsymbol{a}|}(|\boldsymbol{a} - \boldsymbol{a}_h|)dx \\
&= c \sum_{K\in\mathcal{T}_h} \int_K \varphi^*_{|\boldsymbol{a}|}(|\boldsymbol{a} - \Pi\boldsymbol{a} + \Pi\boldsymbol{a} - \boldsymbol{a}_h|)dx \leq \\
&c \sum_{K\in\mathcal{T}_h} (\int_K \varphi^*_{|\boldsymbol{a}|}(|\boldsymbol{a} - \Pi\boldsymbol{a}|)dx + \int_K \varphi^*_{|\boldsymbol{a}|}(|\Pi(\boldsymbol{a} - A(\boldsymbol{l}_h))|)dx) = c \sum_{K\in\mathcal{T}_h} (I_1 + I_2).
\end{aligned}$$

  In the estimate of I_1 we use (2.50)

$$I_1 \leq c \int_K \varphi^*_{|\langle\boldsymbol{a}\rangle_K|}(|\boldsymbol{a} - \Pi\boldsymbol{a}|)dx + c \int_K |F^*(\boldsymbol{a}) - F^*(\langle\boldsymbol{a}\rangle_K)|^2 dx = I_3 + I_4. \tag{7.6}$$

  For I_4 we use the fact that $\boldsymbol{a} = A(\boldsymbol{l})$, lemma 2.17 and finally lemma 4.6

$$\begin{aligned}
I_4 &= \int_K |F^*(A(\boldsymbol{l})) - F^*(\langle A(\boldsymbol{l})\rangle_K)| \\
&\leq c \int_K |F(\boldsymbol{l}) - \langle F(\boldsymbol{l})\rangle_K|^2 dx \leq c h_K^2 \int_K |\nabla F(\boldsymbol{l})|^2 dx.
\end{aligned} \tag{7.7}$$

  We will estimate I_3 by I_4, using the fact that $\Pi A(\langle\boldsymbol{l}\rangle_K) = A(\langle\boldsymbol{l}\rangle_K)$, (4.18) and (2.50)

$$I_3 \leq c\int_K \varphi^*_{|\langle \boldsymbol{a}\rangle_K|}(|\boldsymbol{a}-\langle \boldsymbol{a}\rangle_K|)dx + c\int_K \varphi^*_{|\langle \boldsymbol{a}\rangle_K|}(|\Pi(\boldsymbol{a}-\langle \boldsymbol{a}\rangle_K)|)dx$$
$$\leq c\int_K \varphi^*_{|\langle \boldsymbol{a}\rangle_K|}(|\boldsymbol{a}-\langle \boldsymbol{a}\rangle_K|)dx \leq c\int_K |F^*(\boldsymbol{a})-F^*(\langle \boldsymbol{a}\rangle_K)|^2dx = cI_4.$$

In the estimate of I_2 we use in order (2.50), (4.18) and (2.45)

$$I_2 \leq c\int_K \varphi^*_{\langle|\boldsymbol{a}|\rangle_K}(|\Pi(\boldsymbol{a}-A(\boldsymbol{l}_h))|)dx + c\int_K |F^*(\boldsymbol{a})-F^*(\langle \boldsymbol{a}\rangle_K)|^2dx$$
$$\leq c\int_K \varphi^*_{\langle|\boldsymbol{a}|\rangle_K}(|\boldsymbol{a}-A(\boldsymbol{l}_h)|)dx + cI_4$$
$$\leq c\int_K \varphi^*_{|\boldsymbol{a}|}(|\boldsymbol{a}-A(\boldsymbol{l}_h)|)dx + cI_4 + +c\int_K |F^*(\boldsymbol{a})-F^*(\langle \boldsymbol{a}\rangle_K)|^2dx$$
$$\leq c\int_K |F(\boldsymbol{l})-F(\boldsymbol{l_h})|^2dx + cI_4.$$

Putting these estimates together, we have

$$\|F^*(\boldsymbol{a})-F^*(\boldsymbol{a}_h)\|^2_{L^2(\Omega)} \leq c\,\|F(\boldsymbol{l})-F(\boldsymbol{l}_h)\|^2_{L^2(\Omega)} + ch^2\,\|\nabla F(\boldsymbol{l})\|^2_{L^2(\Omega)}\,. \quad (7.8)$$

- Now we choose $w_h = \Pi_{SZ}u$. This immediately reduces the right hand side.

$$\int_{\mathcal{F}_h^I} \varphi(|h^-1[\Pi_{SZ}u]\boldsymbol{n}|)ds = 0. \quad (7.9)$$

- Estimate of the term $h\int_{\mathcal{F}_h^{ID}} \varphi^*(|\{\Pi\boldsymbol{a}\}-\{\boldsymbol{a}\}|)ds$ can be done in a following way. Using the Young inequality and the fact that $\Pi A(\langle \boldsymbol{l}\rangle_K) = A(\langle \boldsymbol{l}\rangle_K)$, for $\Gamma \in \mathcal{F}_h^{ID}$ the edge of some element $K \in \mathcal{T}_h$ we have

$$h\int_\Gamma \varphi^*(|\Pi\boldsymbol{a}-\boldsymbol{a}|)ds = h\int_\Gamma \varphi^*(|\Pi A(\boldsymbol{l})-\Pi A(\langle \boldsymbol{l}\rangle_K)+A(\langle \boldsymbol{l}\rangle_K)-A(\boldsymbol{l})|)ds$$
$$\leq ch\int_\Gamma \varphi^*(|\Pi(A(\boldsymbol{l})-A(\langle \boldsymbol{l}\rangle_K)|)ds + ch\int_\Gamma \varphi^*(|A(\boldsymbol{l})-A(\langle \boldsymbol{l}\rangle_K)|)ds =$$
$$I_1 + I_2.$$

For the first integral we use in order (4.32), (4.18), the fact that for $a \geq 0$ it holds $\varphi^*(t) \leq (\varphi^*)_a(t)$, since $q-2 \geq 0$, (2.46) togeher with (2.54), lemma 2.17 and finally Poincaré inequality, i.e lemma 4.6 gives us

$$I_1 \leq c\int_K \varphi^*(|\Pi(A(\boldsymbol{l})-A(\langle \boldsymbol{l}\rangle_K))|)dx \leq c\int_K \varphi^*(|A(\boldsymbol{l})-A(\langle \boldsymbol{l}\rangle_K)|)dx$$
$$\leq c\int_K \varphi^*_{|A(\langle \boldsymbol{l}\rangle_K)|}(|A(\boldsymbol{l})-A(\langle \boldsymbol{l}\rangle_K)|)dx \leq c\,\|F(\nabla u)-F(\langle \nabla u\rangle_K)\|^2_{L^2_K}$$
$$\leq c\,\|F(\nabla u)-\langle F(\nabla u)\rangle_K\|^2_{L^2_K} \leq ch^2\,\|\nabla F(\nabla u)\|^2_{L^2_K}\,.$$

And for I_2 we again use in order $\varphi^*(t) \leq (\varphi^*)_a(t)$, (2.54), lemma 4.3 for $\psi(t) = t^2$, lemma 2.17 and Poincaré lemma 4.6.

$$\begin{aligned} I_2 &\leq ch \int_\Gamma \varphi^*_{|A(\langle \boldsymbol{l} \rangle_K)|}(|A(\boldsymbol{l}) - A(\langle \boldsymbol{l} \rangle_K)|)ds \leq ch \left\| F(\nabla u) - F(\langle \nabla u \rangle_K) \right\|^2_{L^2_\Gamma} \\ &\leq c \left\| F(\nabla u) - F(\langle \nabla u \rangle_K) \right\|^2_{L^2(K)} + ch^2 \left\| \nabla F(\nabla u) \right\|^2_{L^2(K)} \\ &\leq ch^2 \left\| \nabla F(\nabla u) \right\|^2_{L(K)} . \end{aligned}$$

Therefore

$$\begin{aligned} h \int_{\mathcal{F}_h^{ID}} \varphi^*(|\{\Pi \boldsymbol{a}\} - \{\boldsymbol{a}\}|)ds &\leq c \sum_{K \in \mathcal{T}_h} h \int_{\delta K \cap int\Omega'} (|\Pi \boldsymbol{a} - \boldsymbol{a}|)ds \\ &\leq ch^2 \left\| \nabla F(\nabla u) \right\|^2_{L^2(\Omega)} . \end{aligned} \tag{7.10}$$

In the further estimates, we will need the following lemma, based on [17, Theorem 5.7].

Lemma 7.1. *Let $k \geq 1$ and the function F be defined as before, i.e $F(\nabla u) \in W^{1,2}(\Omega)$, then*

$$\left\| F(\nabla_h \Pi_{SZ} u) - F(\nabla u) \right\|^2_{L^2(\Omega)} \leq ch^2 \left\| \nabla F(\nabla u) \right\|^2_{L^2(\Omega)} . \tag{7.11}$$

Proof. Let $K \in \mathcal{T}_h$, $Q \in \mathbb{R}^d$ and $r \in P^1(S_K)$ be a polynomial, such that $\nabla_h r = Q$. Then $\Pi_{SZ}(r) = r$ and $Q = \nabla_h \Pi_{SZ} r$.

$$\begin{aligned} &\fint_K |F(\nabla_h u) - F(\nabla_h \Pi_{SZ} u)|^2 dx \\ &\leq c(\fint_K |F(\nabla_h u) - F(Q)|^2 dx + \fint_K |F(\nabla_h \Pi_{SZ} u) - F(Q)|^2 dx) = I_1 + I_2. \end{aligned} \tag{7.12}$$

Using (2.45), the verzion of (4.43) local to K and (2.45) again

$$\begin{aligned} I_2 &\leq c \fint_K \varphi_{|Q|}(|\nabla_h \Pi_{SZ} u - Q|)dx = c \fint_K \varphi_{|Q|}(|\nabla_h \Pi_{SZ} u - \nabla_h \Pi_{SZ} r|)dx \\ &= c \fint_K \varphi_{|Q|}(|\nabla_h (\Pi_{SZ} u - \Pi_{SZ} r)|)dx \leq c \fint_{S_K} \varphi_{|Q|}(|\nabla_h (u - r)|)dx \\ &= c \fint_{S_K} \varphi_{|Q|}(|\nabla_h (u - Q)|)dx \leq c \fint_{S_K} |F(\nabla_h u) - F(Q)|^2 dx. \end{aligned} \tag{7.13}$$

I_1 is easily estimated by

$$I_1 \leq c \fint_{S_K} |F(\nabla_h u) - F(Q)|^2 dx. \tag{7.14}$$

F is strictly monotone and therefore there exists $P \in \mathbb{R}^d$, such that $F(Q) = P$. Since Q was arbitrary, we have

$$\fint_K |F(\nabla_h u) - F(\nabla_h \Pi_{SZ} u)|^2 dx \leq c \, inf_{P \in \mathbb{R}^d} \fint_{S_K} |F(\nabla_h u) - F(Q)|^2 dx. \tag{7.15}$$

Since $W^{1,2}$ is a Hilbert space, the P reaching the infimum is $\langle F(\nabla_h u)\rangle_K$. This together with Poincaré lemma 4.6 implies

$$\fint_K |F(\nabla_h u) - F(\nabla_h \Pi_{SZ} u)|^2 dx \leq c h_K^2 \fint_K |\nabla F(\nabla u)|^2 dx. \tag{7.16}$$

Finally summing trough all $K \in \mathcal{T}_h$ we get the original assertion.

□

The final estimate can have a slightly different form depending on the choice of u_D^*, which affects the first and the third term on the right hand side of the Theorem 7.1.

7.2 The estimate with the choice $u_D^* = \Pi_{SZ} u$

- It holds that

$$\nabla_{DG}^h \Pi_{SZ} u + R_h u_D^* = \nabla_h \Pi_{SZ} u + R_h(u_D^* - \Pi_{SZ} u) = \nabla_h \Pi_{SZ} u. \tag{7.17}$$

  Therefore with lemma (7.1), it holds

$$\begin{aligned} &\left\| F(\nabla_{DG}^h w_h + R_h u_D^*) - F(\nabla u) \right\|_{L^2(\Omega)}^2 \\ &= \left\| F(\nabla_h \Pi_{SZ} u_h) - F(\nabla u) \right\|_{L^2(\Omega)}^2 \leq c h^2 \left\| \nabla F(\nabla u) \right\|_{L^2(\Omega)}^2. \end{aligned}$$

- Thanks to the choice of u_D^* we also have

$$\int_{\mathcal{F}_h^D} \varphi(|h^- 1(w_h - u_D^*)\boldsymbol{n}|) ds = \int_{\mathcal{F}_h^D} \varphi(|h^- 1(\Pi_{SZ} u - \Pi_{SZ} u)\boldsymbol{n}|) ds = 0. \tag{7.18}$$

Putting all the estimates together, we have the following theorem.

Theorem 7.2. *Let $u \in W^{1,\varphi}(\Omega)$, $\boldsymbol{l} \in L^{\varphi}(\Omega)$, $\boldsymbol{a} \in W^{1,\varphi^*}(\Omega)$ be the solution of the problem (5.3) and $u_h, \boldsymbol{l_h}, \boldsymbol{a_h}$ the DG solution of (6.9) from V_h^k, X_h^k, X_h^k respectively. Then for $\sigma > 0$, $u_D^* = \Pi_{SZ} u$ and $F(\nabla u) \in W^{1,2}(\Omega)$ it holds*

$$\begin{aligned} &\|F(\boldsymbol{l}) - F(\boldsymbol{l}_h)\|_{L^2(\Omega)}^2 + \|F^*(\boldsymbol{a}) - F^*(\boldsymbol{a}_h)\|_{L^2(\Omega)}^2 \\ &+ \sigma h \int_{\mathcal{F}_h^I} \varphi(|h^- 1[u_h]\boldsymbol{n}|) ds + \sigma h \int_{\mathcal{F}_h^D} \varphi(|h^- 1(u_h - u_D^*)\boldsymbol{n}|) ds \\ &\leq c h^2 \left\| \nabla F(\nabla u) \right\|_{L^2(\Omega)}^2. \end{aligned}$$

7.3 The estimate with the choice $u_D^* = u$

- Since

$$\nabla_{DG}^h \Pi_{SZ} u + R_h u = \nabla_h \Pi_{SZ} u + R_h(u - \Pi_{SZ} u), \tag{7.19}$$

we can estimate the first term on the right hand of Theorem 7.1 in the following way. First we use (2.45) and triangle inequality

$$\begin{aligned}
&\left\| F(\nabla_{DG}^h \Pi_{SZ} u + R_h u) - F(\nabla u) \right\|_{L^2(\Omega)}^2 \\
&\leq c \int_\Omega \varphi_{|\nabla u|}(|\nabla_h \Pi_{SZ} u + R_h(u - \Pi_{SZ} u) - \nabla u|) dx \\
&\leq c \int_\Omega \varphi_{|\nabla u|}(|\nabla_h \Pi_{SZ} u - \nabla u|) dx + c \int_\Omega \varphi_{|\nabla u|}(|R_h(u - \Pi_{SZ} u)|) dx = I_1 + I_2.
\end{aligned}$$

For I_1 we use (2.45) again, together with lemma (7.1)

$$I_1 \leq \|F(\nabla_h \Pi_{SZ} u) - F(\nabla u)\|_{L^2(\Omega)}^2 \leq ch^2 \|\nabla F(\nabla u)\|_{L^2(\Omega)}^2 . \tag{7.20}$$

Concerning I_2 we use in order the definition of R_h, (2.49), (4.12), lemma 2.17 and finally Poincaré lemma 4.6

$$\begin{aligned}
I_2 &= \sum_{\Gamma \in F_h^D} \int_{S_\Gamma} \varphi_{|\nabla u|}(|R_h^\Gamma(u - \Pi_{SZ} u)|) dx \\
&\leq c \sum_{\Gamma \in F_h^D} \int_{S_\Gamma} \varphi_{|\langle \nabla u \rangle_{S_\Gamma}|}(|R_h^\Gamma(u - \Pi_{SZ} u)|) + |F(\nabla u) - F(\langle \nabla u \rangle_{S_\Gamma})|^2 dx \\
&\leq c \sum_{\Gamma \in F_h^D} h_\Gamma \int_\Gamma \varphi_{|\langle \nabla u \rangle_{S_\Gamma}|}(|h_\Gamma^{-1}(u - \Pi_{SZ} u)|\boldsymbol{n}) ds \\
&+ \int_{S_\Gamma} |F(\nabla u) - F(\langle \nabla u \rangle_{S_\Gamma})|^2 dx \\
&\leq c \sum_{\Gamma \in F_h^D} h_\Gamma \int_\Gamma \varphi_{|\langle \nabla u \rangle_{S_\Gamma}|}(|h_\Gamma^{-1}(u - \Pi_{SZ} u)|\boldsymbol{n}) ds + ch^2 \|\nabla F(\nabla u)\|_{L^2(\Omega)}^2 .
\end{aligned}$$

Next we use in order the local version of (4.43) on K for $g = h^{-1}(u - \Pi_{SZ} u)$, (2.49) and (2.45) together with Poincaré lemma 4.6

$$\begin{aligned}
&\sum_{\Gamma \in F_h^D} h_\Gamma \int_\Gamma \varphi_{|\langle \nabla u \rangle_{S_\Gamma}|}(|h_\Gamma^{-1}(u - \Pi_{SZ} u)|) ds \\
&\leq \sum_{\Gamma \in F_h^D} c \int_{S_\Gamma} \varphi_{|\langle \nabla u \rangle_{S_\Gamma}|}(|\nabla u - \nabla \Pi_{SZ} u|) dx \\
&\leq c \sum_{\Gamma \in F_h^D} \Big(\int_{S_\Gamma} \varphi_{|\nabla u|}(|\nabla u - \nabla \Pi_{SZ} u|) dx + c \int_{S_\Gamma} |F(\nabla u) - F(\langle \nabla u \rangle_{S_\Gamma})|^2 dx \Big) \\
&\leq c \sum_{\Gamma \in F_h^D} \Big(\int_{S_\Gamma} |F(\nabla u) - F(\nabla \Pi_{SZ} u)|^2 dx + ch_\Gamma^2 \int_{S_\Gamma} |\nabla F(\nabla u)|^2 dx \Big) \\
&\leq ch^2 \|\nabla F(\nabla u)\|_{L^2(\Omega)}^2 .
\end{aligned}$$

- For this choice of u_D^* the term $h\int_{\mathcal{F}_h^D}\varphi(|h^{-1}(u_D^*-\Pi_{SZ}u)|)dx$ doest not vanish. In order to estimate it, we will need the following lemma.

  Lemma 7.2. *Under the assumptions of Theorem 7.1 it holds that*

  $$\varphi(|\nabla^2 u|) \leq c\,\|\nabla F(\nabla u)\|^2_{L^2(\Omega)} + c\int_\Omega \varphi(|\nabla u|)dx \int_\Omega. \tag{7.21}$$

  Proof. First we compute the derivative of $F(\nabla u)$

  $$\frac{\partial F_j(\nabla u)}{\partial_l P} = \frac{p-2}{2}(1+|\nabla u|)^{\frac{p-4}{2}}\frac{\partial u}{\partial x_j}\frac{\partial|\nabla u|}{\partial x_l} + (1+|\nabla u|)^{\frac{p-2}{2}}\frac{\partial^2 u}{\partial x_j \partial x_l}. \tag{7.22}$$

  Therefore

  $$|\nabla F(\nabla u)|^2 = |A|^2 + 2A\cdot B + |B|^2, \tag{7.23}$$

  where

  $$\begin{aligned} |B|^2 &= (1+|\nabla u|)^{p-2}|\nabla^2 u|^2 \\ 2A\cdot B &= (p-2)(1+|\nabla u|)^{p-3}|\nabla u||\nabla^2 u|^2 \\ |A|^2 = (\frac{p-2}{2})^2(1+|\nabla u|)^{p-4}|\nabla u|^2|\nabla^2 u|^2 &\leq (\frac{p-2}{2})^2|B|^2. \end{aligned} \tag{7.24}$$

  Putting this together , we have the estimate

  $$|\nabla F(\nabla u)|^2 \geq |B|^2 \tag{7.25}$$

  now for $q\in[1,2], a\geq 0$ and $b\geq 1$ we have

  $$a^q = (a^2 b^{q-2})^{\frac{q}{2}}(b^{\frac{(2-q)q)}{2}}) \leq a^2 b^{q-2} + b^q, \tag{7.26}$$

  where we used basic Young inequality. Using this for $a=|\nabla^2 u|$ and $b=(1+|\nabla u|)$ we have

  $$|\nabla^2 u|^p \leq (1+|\nabla u|)^{p-2}|\nabla^2 u|^2 + (1+|\nabla u|)^p. \tag{7.27}$$

  Finally using that left hand side is $\geq c\varphi(|\nabla^2 u|)$ and right hand side is $\leq c(|B|^2 + \int_\Omega \varphi(|\nabla u|)dx)$, we have the original assertion. □

  Now we can estimate using (4.44), the fact that $\varphi(ht)\leq ch^p\varphi(t)$ and lemma (7.2).

  $$\begin{aligned} &h\int_{\mathcal{F}_h^D}\varphi(|h^{-1}(u-\Pi_{SZ}u)|)dx \leq c\int_\Omega \varphi(|h\nabla^2 u|)dx \\ &\leq ch^p\int_\Omega \varphi(|\nabla^2 u|)dx \leq ch^p(\|\nabla F(\nabla u)\|^2_{L^2(\Omega)} + \int_\Omega \varphi(|\nabla u|)dx). \end{aligned}$$

Putting all the estimates together we arrive at

Theorem 7.3. *Let $u \in W^{1,\varphi}(\Omega)$, $\boldsymbol{l} \in L^{\varphi}(\Omega)$, $\boldsymbol{a} \in W^{1,\varphi^*}(\Omega)$ be the solution of the problem (5.3) and $u_h, \boldsymbol{l_h}, \boldsymbol{a_h}$ the DG solution of (6.9) V_h^k, X_h^k, X_h^k respectively. Then for $\sigma > 0$, $u_D^* = u$ and $F(\nabla u) \in W^{1,2}(\Omega)$*

$$\begin{aligned}
&\|F(\boldsymbol{l}) - F(\boldsymbol{l}_h)\|^2_{L^2(\Omega)} + \|F^*(\boldsymbol{a}) - F^*(\boldsymbol{a}_h)\|^2_{L^2(\Omega)} \\
&+ \sigma h \int_{\mathcal{F}_h^I} \varphi(|h^- 1[u_h]\boldsymbol{n}|)ds + \sigma h \int_{\mathcal{F}_h^D} \varphi(|h^- 1(u_h - u_D^*)\boldsymbol{n}|)ds \\
&\leq c(h^p \|\nabla F(\nabla u)\|^2_{L^2(\Omega)} + \int_\Omega \varphi(|\nabla u|)dx).
\end{aligned}$$

It is important to note that the constants in the estimates of the Theorem 7.2 and the Theorem 7.3 depend only on the characteristics of the domain Ω, the mesh $\mathcal{T}_H$ and the function A.

7.4 Time dependent problem

Definition 7.1. *For u satisfying the conditions of the Theorem 7.1 we define the following $F-$norm as*

$$\begin{aligned}
\|u - u_h\|^2_{F,DG} &:= \|F(\boldsymbol{l}) - F(\boldsymbol{l}_h)\|^2_{L^2(\Omega)} + \|F^*(\boldsymbol{a}) - F^*(\boldsymbol{a}_h)\|^2_{L^2(\Omega)} \\
&+\sigma h \int_{\mathcal{F}_h^I} \varphi(|h^- 1[u_h]\boldsymbol{n}|)ds + \sigma h \int_{\mathcal{F}_h^D} \varphi(|h^- 1(u_h - u_D^*)\boldsymbol{n}|)ds.
\end{aligned} \tag{7.28}$$

In case of time dependent equation (5.3), we can use the results from the time independent problem for fixed $t \in (0,T)$ in the error estimates. This will give us the analogy of the Theorem 7.2 or 7.3 for fixed $t \in (0,T)$, with only difference being that the terms with the time derivatives u_{ht} and u_t will be present. Using this in combination with the procedure for getting the time dependent estimates inspired by the result from [9], where it is used for a different problem, we can arrive at the final error estimates.

Following the Theorem 7.2, we have

$$\int_\Omega (u_{ht} - u_t) z_h dx + \|u - u_h\|^2_{F,DG} \leq ch^2 \|\nabla F(\nabla u)\|^2_{L^2(\Omega)}, \tag{7.29}$$

where $z_h = u_h - \Pi_{SZ} u$. We substract $\int_\Omega \partial_t \Pi_{SZ} u$ from both sides.

$$\int_\Omega z_{ht} z_h dx + \|u - u_h\|^2_{F,DG} \leq \int_\Omega \partial_t (u - \Pi_{SZ} u) z_h dx + ch^2 \|\nabla F(\nabla u)\|^2_{L^2(\Omega)}. \tag{7.30}$$

Applying Cauchy-Schwarz inequality on the first term on the right hand side

$$\begin{aligned}
&\frac{1}{2}\frac{\partial}{\partial t} \|z_h\|^2_{L^2(\Omega)} + \|u - u_h\|^2_{F,DG} \\
&\leq \frac{1}{2} \|\partial_t (u - \Pi_{SZ} u)\|^2_{L^2(\Omega)} + \frac{1}{2} \|z_h\|^2_{L^2(\Omega)} + ch^2 \|\nabla F(\nabla u)\|^2_{L^2(\Omega)}.
\end{aligned} \tag{7.31}$$

Integrating from 0 to $t \in (0,T)$ we have

$$\begin{aligned} &\|z_h(t)\|^2_{L^2(\Omega)} + 2\int_0^t \|u(\tau) - u_h(\tau)\|^2_{F,DG}\, d\tau \\ &\leq \int_0^t \|\partial_t(u(\tau) - \Pi_{SZ}u(\tau))\|^2_{L^2(\Omega)} + ch^2 \|\nabla F(\nabla u(\tau))\|^2_{L^2(\Omega)}\, d\tau + \\ &\int_0^t \|z_h(\tau)\|^2_{L^2(\Omega)}\, d\tau, \end{aligned} \tag{7.32}$$

since $u_h(0) = \Pi_{SZ}u(0)$. Applying Gronwall inequality, i,e lemma 6.1, we get

$$\begin{aligned} &\|u_h(t) - \Pi_{SZ}u(t)\|^2_{L^2(\Omega)} + \int_0^t \|u(\tau) - u_h(\tau)\|^2_{F,DG}\, d\tau \\ &\leq ce^t(\int_0^t \|\partial_t(u(\tau) - \Pi_{SZ}u(\tau))\|^2_{L^2(\Omega)} + h^2 \|\nabla F(\nabla u(\tau))\|^2_{L^2(\Omega)}\, d\tau). \end{aligned} \tag{7.33}$$

Now we use the triangle inequality for

$$\|u_h(t) - u(t)\|^2_{L^2(\Omega)} \leq \|u_h(t) - \Pi_{SZ}u(t)\|^2_{L^2(\Omega)} + \|u(t) - \Pi_{SZ}u(t)\|^2_{L^2(\Omega)}. \tag{7.34}$$

Putting the term independent of u_h to the right hand side we arrive at

$$\begin{aligned} &\|u_h(t) - u(t)\|^2_{L^2(\Omega)} + \int_0^t \|u(\tau) - u_h(\tau)\|^2_{F,DG}\, d\tau \\ &\leq ce^t(\int_0^t \|\partial_t(u(\tau) - \Pi_{SZ}u(\tau))\|^2_{L^2(\Omega)} + h^2 \|\nabla F(\nabla u(\tau))\|^2_{L^2(\Omega)}\, d\tau) \\ &+ \|u(t) - \Pi_{SZ}u(t)\|^2_{L^2(\Omega)}. \end{aligned} \tag{7.35}$$

Due to the definition of Π_{SZ} it holds that $\partial_t \Pi_{SZ} u = \Pi_{SZ}\partial_t u$. Further the proof of lemma 7.1 holds up for F_2 being defined the same way as F, with $p = 2$ and φ_2 defined analogically, meaning we get the estimate in L^2 norm. Therefore

$$\begin{aligned} \|\partial_t(u(\tau) - \Pi_{SZ}u(\tau))\|^2_{L^2(\Omega)} &\leq ch^2 \left\|\nabla^2 u_t(\tau)\right\|^2_{L^2(\Omega)}, \\ \|u(t) - \Pi_{SZ}u(t)\|^2_{L^2(\Omega)} &\leq ch^2 \left\|\nabla^2 u(t)\right\|^2_{L^2(\Omega)}. \end{aligned} \tag{7.36}$$

Combining all this we finally arrive at the following result.

Theorem 7.4. *Let u, u_h be the solutions to (5.1) and (5.3) respectively that satisfy the assumptions of Theorem 7.1 and $u_D^* = \Pi_{SZ}u$. If further $u(t) \in W^{2,2}(\Omega)$, for all $t \in (0,T)$, $u_t \in L^2(0,T,W^{2,2}(\Omega))$ and $F(\nabla(u)) \in L^2(0,T,W^{1,2}(\Omega))$, then*

$$\begin{aligned} &\|u_h(t) - u(t)\|^2_{L^2(\Omega)} + \int_0^t \|u(\tau) - u_h(\tau)\|^2_{F,DG}\, d\tau \\ &\leq ch^2 e^t(\left\|\nabla^2 u(t)\right\|^2_{L^2(\Omega)} + \left\|\nabla^2 u_t(\tau)\right\|^2_{L^2(0,T,L^2(\Omega))} + \|\nabla F(\nabla u(\tau))\|^2_{L^2(0,T,L^2(\Omega))}), \end{aligned} \tag{7.37}$$

for all $t \in (0,T)$.

Via same steps we can arrive at the analogy of the Theorem 7.4, which follows from the Theorem 7.3.

Theorem 7.5. *Let u, u_h be the solutions to (5.1) and (5.3) respectively that satisfy the assumptions of the Theorem 7.1 and $u_D^* = u$. If further $u(t) \in W^{2,2}(\Omega)$, for all $t \in (0,T)$, $u_t \in L^2(0,T,W^{2,2}(\Omega))$, $F(\nabla(u)) \in L^2(0,T,W^{1,2}(\Omega))$ and $u \in L^1(0,T,W^{1,\varphi}(\Omega))$, then*

$$\begin{aligned} &\|u_h(t) - u(t)\|^2_{L^2(\Omega)} + \int_0^t \|u(\tau) - u_h(\tau)\|^2_{F,DG}\, d\tau \\ &\leq ch^p e^t(\left\|\nabla^2 u(t)\right\|^2_{L^2(\Omega)} + \left\|\nabla^2 u_t(\tau)\right\|^2_{L^2(0,T,L^2(\Omega))} + \|\nabla F(\nabla u(\tau))\|^2_{L^2(0,T,L^2(\Omega))}) \\ &\qquad + ce^t \int_0^t \int_\Omega \varphi(|\nabla u(\tau)|) dx d\tau, \end{aligned} \tag{7.38}$$

for all $t \in (0,T)$.

8. Numerical Examples

In the Numerical Examples we use the code package ADGFEM developed in Charles University Prague for the numerical solution of nonlinear convection-diffusion equations. Specifically, we use the part of the package modified for the solution of Forchheimer equations. The implemented method uses space-time discontinuous Galerkin method with the adaptive step choice, in first degree time discretization. The solver for the system of nonlinear equations is Newton-like, using the linearization, instead of calculating the Jacobi matrix. The linear systems of equations are solved using GMRES with the block ILU(0) preconditioner. The whole method is described in much broader detail in [21].

Due to the complexity of the implementation, we are forced to use IPDG method for the numerical experiments instead of local DG method. Discontinuous Galerkin formulation of both methods end up with the very similar form, differing only in the minor technical terms and therefore both methods have the same expected asymptotic convergence rate.

The program works with the IPDG formulation of the original problem

$$\begin{aligned}\int_\Omega \frac{\partial u_h}{\partial t} v_h + \sum_{K\in\mathcal{T}_h}\int_K K(|\nabla u_h|)\nabla u_h\cdot\nabla v_h dx - \sum_{\Gamma\in F_h^{ID}}\int_\Gamma \{K(|\nabla u_h|)\nabla u_h\}\cdot \boldsymbol{n}[v_h]ds \\ -\sigma\sum_{\Gamma\in F_h^I}\int_\Gamma \{K(h_\Gamma^{-1}|[u_h]|)\nabla v_h\}\cdot\boldsymbol{n}[u_h]ds \\ -\sigma\sum_{\Gamma\in F_h^D}\int_\Gamma \{K(h_\Gamma^{-1}|[u_h-u_D]|)\nabla v_h\}\cdot\boldsymbol{n}(u_h-u_D)ds \\ +\sigma\sum_{\Gamma\in F_h^I}\int_\Gamma h_\Gamma^{-1}K(h_\Gamma^{-1}|[u_h]|)[u_h][v_h]ds \\ +\sigma\sum_{\Gamma\in F_h^D}\int_\Gamma h_\Gamma^{-1}K(h_\Gamma^{-1}|u_h-u_D|)(u_h-u_D)v_h ds \\ = \int_\Omega f v_h dx + \sum_{\Gamma\in F_h^N}\int_\Gamma g_N v_h ds.\end{aligned} \tag{8.1}$$

In the experiments we use one of the simpler versions of the Frochheimer models

$$(a_0 + a_1|v|)v = -\nabla p, \tag{8.2}$$

where $a_0 = \frac{\mu}{k}$ and $a_1 = \frac{0.55\rho}{\sqrt{k}}$. Taking the norm of both sides we have

$$(a_0 + a_1|v|)|v| = |\nabla p|. \tag{8.3}$$

If we want to get the equation for v, we proceed as follows

$$a_1|v|^2 + a_0|v| - |\nabla p| = 0.$$

This quadratic equation has a positive root

$$|v| = \frac{1}{2a_1}(-a_0 + \sqrt{a_0^2 + 4a_1|\nabla p|}).$$

Substituting this back to 8.2 we have

$$v = \frac{\nabla p}{a_0 + \frac{a_1}{2a_1}(-a_0 + \sqrt{a_0^2 + 4a_1|\nabla p|})} = -\frac{2\nabla p}{a_0 + \sqrt{a_0^2 + 4a_1|\nabla p|}}. \tag{8.4}$$

We can use the relation 1.15 from chapter 1 to get the form of K

$$K(|\nabla p|) = \frac{2}{a_0 + \sqrt{a_0^2 + 4a_1|\nabla p|}}, \tag{8.5}$$

that can be substituted into the final equation based on (1.21)

$$\frac{\partial p}{\partial t} - \kappa \nabla \cdot (K(|\nabla p|)\nabla p) = 0. \tag{8.6}$$

Note that here we did not use the dimensionless variant of the equation and therefore κ is not eliminated.

The values of the physical parameters of the fluid are

$$\begin{aligned} k &= 10^{-12} m^{-2}, \\ \kappa &= 5 \cdot 10^{-10} N^{-} 1 m^2, \\ \mu &= 1.310^{-3} N s m^{-} 2, \\ \rho &= 10^3 kg m^{-3}. \end{aligned}$$

The first example is computed on the simple square domain $(0,1) \times (0,1)$ and time scale $t \in (0,1)$, with parameter $\sigma = 1$. The right hand side $f = 0$, boundary and initial conditions are chosen in such a way that there exists a nontrivial exact solution

$$u = e^{-2t} x_1 x_2 (1 - x_1)(1 - x_2).$$

The calculations are done for the polynomial degree of test functions $1, 2$ and 3. The numerical error is computed in the norms $||\cdot||_{L^2(\Omega)}$ and $||\cdot||_{H^1(\Omega)}$. Assuming the numerical error has the form

$$||e_h|| = Ch^{EOC}, \tag{8.7}$$

where EOC is the experimental order of convergence. Since we have the exact solution we know e_h. Using these two facts we can determine the EOC from the computations on two subsequently refined meshes. Following figures show the experimental error ranges, with the values of h on the horizontal axis and the values of the computed error on the vertical axis.

$$EOC = \frac{log(||e_{h_1}||/||e_{h_2}||)}{log(h_1/h_2)}. \tag{8.8}$$

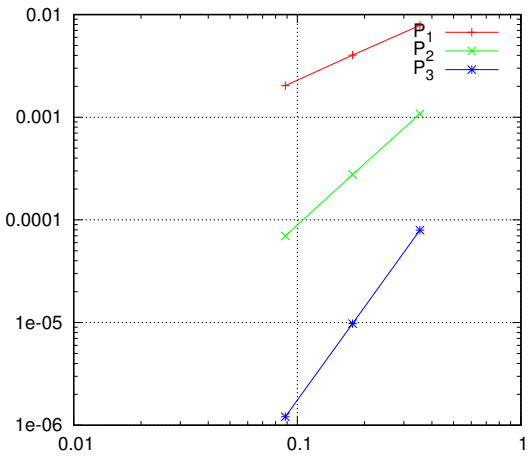

Figure 8.1: Error estimates in H^1 norm, Example 1

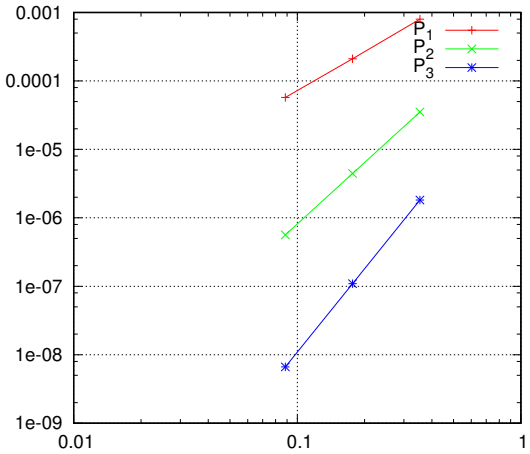

Figure 8.2: Error estimates in L^2 norm, Example 1

We can see that the the method behaves as per the theoretical results. In case of higher degree of polynomial approximation, the numerical results are even better suggesting that the theoretical results are not generally optimal.

The second example shows the solution of the equation 8.6 on a little more complicated domain $\Omega = (-1,1) \times (-1,1) \cup (-0.3,0) \times (1,1.1)$ that consists of two subdomains Ω_1 and Ω_2. Ω_2 consists of $(-1,-0.1) \times (-0.25,0.25) \cup (0.1,1) \times (-0.25,0.25)$ and Ω_1 consists of the rest of the domain. This represents the seepage through a hole in the subsurface of the different permeability, with the two subdomains being easily visible in the Figures with the solution. Two different permeabilities are prescribed for Ω_1 and Ω_2.

$$k_1 = 10^{-12} m^{-2},$$
$$k_2 = 10^{-15} m^{-2}.$$

Initial conditions are set to $p = 0.1Pa$ in $(-1,1) \times (-1,1)$ and $p = 1000Pa$ in the source $(-0.3,0) \times (1,1.1)$. Analogously, there is a Dirichlet boundary condition $p = 1000Pa$ prescribed on the part of the boundary of the source, meaning part of the boundary, where $x_2 > 1$ and Neumann boundary condition

$\nabla p \cdot \boldsymbol{n} = 0$ on the rest of the boundary. The computations is carried over the time interval $(0, T)$, with $T = 10s$.

The following results are shown for the polynomial approximation of degree 1 and values of time in order $t = 1.5\,, 3\,, 5\,, 7.5$ and $10s$

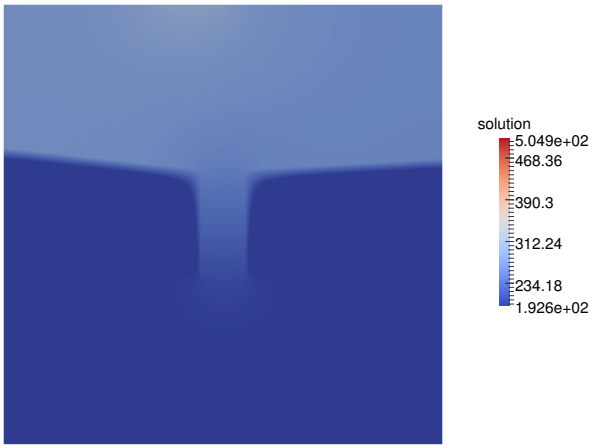

Figure 8.3: Seepage through the hole in the subsurface, Forchheimer equation, result at $1.5s$

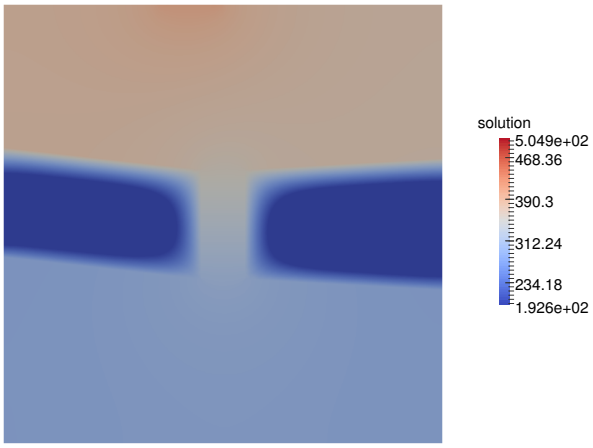

Figure 8.4: Seepage through the hole in the subsurface, Forchheimer equation, result at $3s$

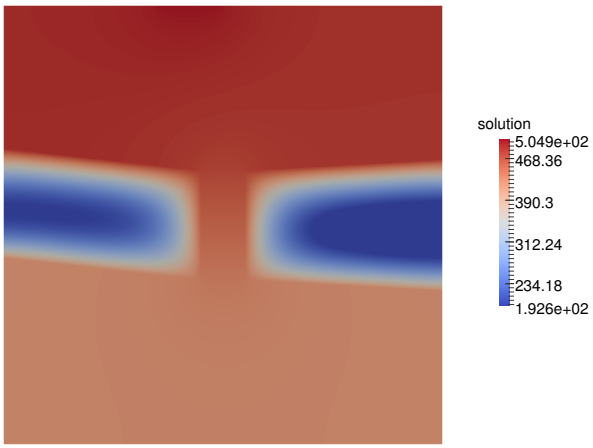

Figure 8.5: Seepage through the hole in the subsurface, Forchheimer equation, result at $5s$

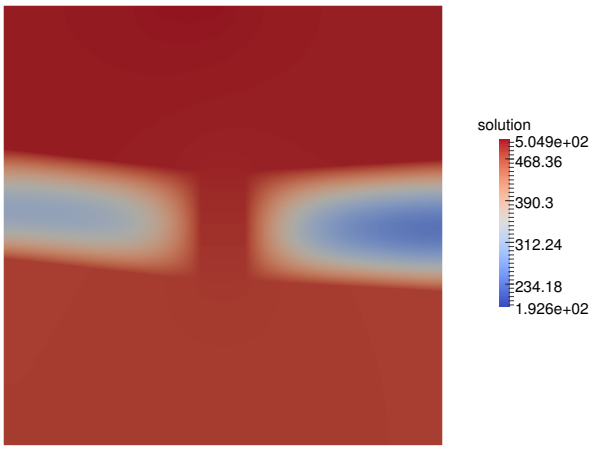

Figure 8.6: Seepage through the hole in the subsurface, Forchheimer equation, result at $7.5s$

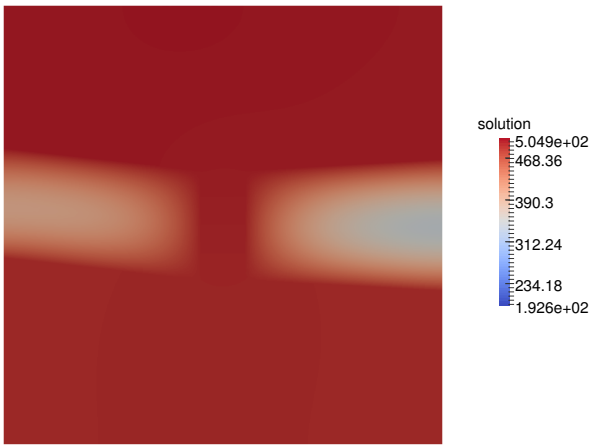

Figure 8.7: Seepage through the hole in the subsurface, Forchheimer equation, result at $10s$

The figures show that the flow behaves as expected on a relatively complicated domain. The flow behaves similarly as in the case $a_1 = 0$, which corresponds to standard Darcy's law. For the sake of comparison we show the results for the linear case in the same time intervals.

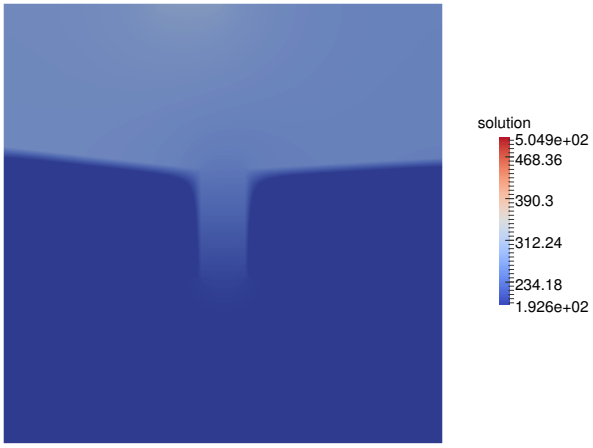

Figure 8.8: Seepage through the hole in the subsurface, Darcy equation, result at $1.5s$

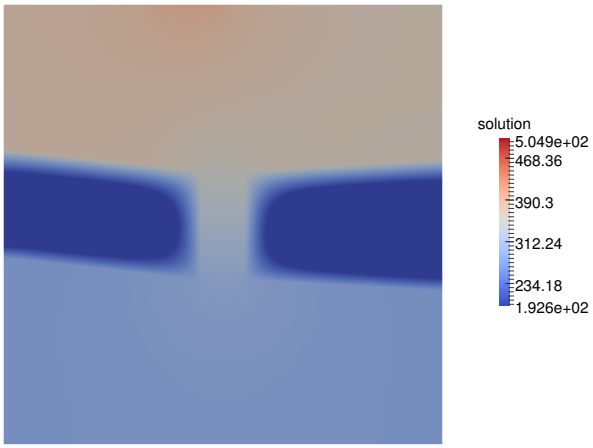

Figure 8.9: Seepage through the hole in the subsurface, Darcy equation, result at $3s$

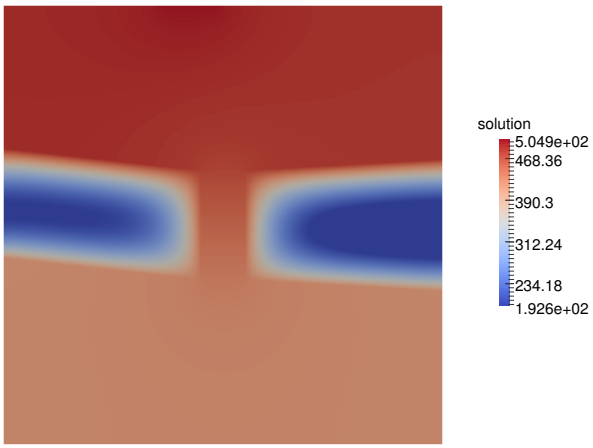

Figure 8.10: Seepage through the hole in the subsurface, Darcy equation, result at $5s$

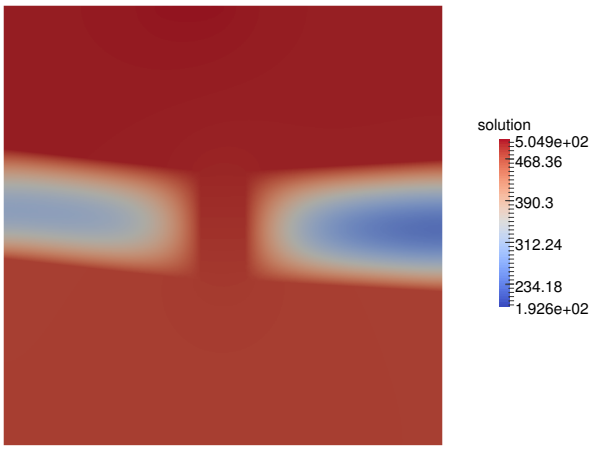

Figure 8.11: Seepage through the hole in the subsurface, Darcy equation, result at $7.5s$

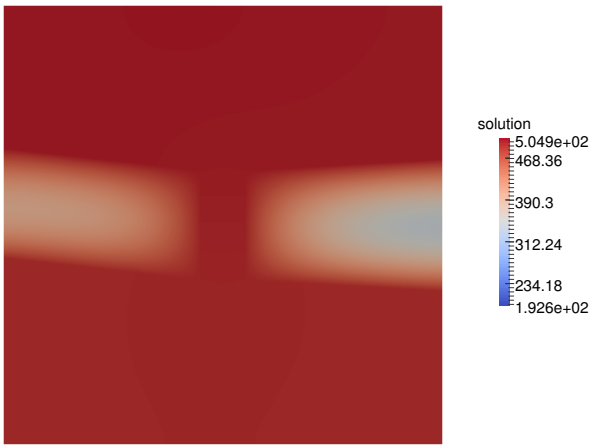

Figure 8.12: Seepage through the hole in the subsurface, Darcy equation, result at $10s$

Large difference probably cannot be seen due to the fact that the nonlinear term contributes to the equation on the scale of 10^{10} as opposed to the linear term, which contributes on the scale of 10^{12}. If the type of the model was chosen, in which both contributions are of the same order of magnitude and the velocity of the flow was higher, an improvement in the nonlinear solution is expected.

Conclusion

The first goal of the thesis was to study the nonlinear flows in the porous media and suggest the fitting numerical method for the given problem. The way we derived the partial differential equation studied in the rest of the thesis is not unique and therefore the analysis ends up having a broader use in the field. The final equations are also very similar to the p-Laplace problem with $p \in (1, 2)$, which only differs in replacing the nonlinear function K bounded from both sides by $(1 + |\nabla u|)^{p-2}$, with the exact formula $(\mu + |\nabla u|)^{p-2}$. More precisely this is the formula for a perturbed p-laplace problem, with the parameter $\mu > 0$, which does not make the analysis much different as opposed to the case, where $\mu = 1$.

The case, where $p \in (1, 2)$ is much more complicated, than the one with $p \geq 2$ and the final results can be easily extended to this case as well. Considering all these generalizations, the results have a more general application, such as in aerodynamics, plasticity and glaciology.

In the numerical analysis of the equations, first the interior penalty discontinuous Galerkin method was considered. For IPDG method, it is possible to derive the general stability estimate, but derivation of any kind of a priori error estimate has proven difficult. In the end we chose the local DG formulation, which caused some complications in the stability estimates, but we were able to derive the a priori error estimates of the method. The derived local DG formulation is written in (5.8), primal formulation is in (5.13) and the main stability results are summed up in the Theorem 6.1 and the Theorem 6.2. In the end, we achieved the linear rate of convergence of the error, estimated in the norm $||\cdot||_{F,DG}$, which also provides an estimate for the jumps of the numerical solution on the edges of the triangulation and even the estimate for the difference between the terms $K(|\nabla u|)\nabla u$ and $\Pi(K(|\nabla^h_{DG} u_h + R_h u^*_D|)(\nabla^h_{DG} u_h + R_h u^*_D))$. The main error estimate results are summed up in Theorems 7.2 and 7.3, for the stationary case and Theorems 7.4 and 7.5, for the time dependent case. The proven rate of convergence is optimal for linear ansatz functions. We were not able to get better estimate for approximation functions of higher polynomial degree, due to the complex nature of the problem and the fact that the error does not depend solely on the approximation error between the exact solution u and the projection $\Pi_{SZ} u$.

The use of the local DG method also required us to implement the theory of Sobolev-Orlicz spaces and N-functions, which had to be introduced in the chapter 2, and some further results concerning the generalized local gradient, the local L^2 projection and Scott-Zhang projection and their interactions with N-functions in chapter 3. In chapter 8 we have shown numerical experiments for one of the simpler Forchheimer's models, verifying the derived results and suggesting they are not optimal for higher degrees of polynomial approximation.

Both IPDG and local DG methods are very similar, only differing in certain technical terms that do not prove to be significant in the numerical experiments and therefore are expected to have similar properties and convergence rates.

Bibliography

[1] Emine Celik, Luan Hoang, Thinh Kieu. *Generalized Forchheimer flows of isentropic gases*. arXiv:1504.00742v1 [math.AP] 3 Apr 2015.

[2] Eugenio Aulisa, Lidia Bloshanskaya, Luan Hoang, Akif Ibragimov *Analysis of generalized Forchheimer flows of compressible fluids in porous media*. Journal of Mathematical Physics, 50(10), 2009.

[3] Luan Hoang and Akif Ibragimov. *Structural stability of generalized Forchheimer. equations for compressible fluids in porous media*. Ltd and London Mathematical Society , 25 November 2010.

[4] Thinh T. Kieu. *Analysis of expanded mixed finite element methods. for the generalized Forchheimer equations*. arXiv:1409.7821v1 [math.NA] 27 Sep 2014.

[5] Thinh T. Kieu. *Numerical analysis for generalised Forchheimer flows of slightly compressible fluids in porous media*. arXiv:1512.03951v1 [math.NA] 12 Dec 2015.

[6] L. Ridgway Scott and Shangyou Zhang. *finite element interpolation of nonsmooth functions satysfying boundary conditions*. Article in Mathematics of Computation, April 1990. 10.1090/S0025-5718-1990-1011446-7.

[7] Petteri Harjulehto, Peter Hästö. *Orlicz spaces and Generalized Orlicz spaces*. Springer International Publishing, June 27, 2018. ISBN 978-3-030-15100-3.

[8] Douglas N. Arnold, Franco Brezzi, Bernardo Cockburn, and L. Donatella Marini. *Unified analysis of discontinuous galerkin methods for elliptic problems*. Society for Industrial and Applied Mathematics 2002.

[9] Ioannis Toulopoulos. *An interiorpenaltydiscontinuousGalerkin finite elementmethod for quasilinearparabolicproblems*. Elsevier B.V, 2014. 10.1016/j.finel.2014.11.001.

[10] Lars Diening, Dietmar Kröner, Michael Růžička and Ioannis Toulopoulos. *A local discontinuous Galerkin approximation for systems with p-structure*. IMA Journal of Numerical Analysis (2014) 34, 1447–1488. 10.1093/imanum/drt040.

[11] Dietmar Kröner, Michael Růžička and Ioannis Toulopoulos. *Numerical solutions of systems with (p,delta)-structure using local discontinuous Galerkin finite element methods*. Int. J. Numer. Meth. Fluids 2014; 76:855–874. 10.1002/f ld.3955.

[12] L. Diening, P. Kaplický, S. Schwarzacher. *BMO estimates for the p-Laplacian*. Elsevier Ltd, 2011.

[13] Michael Růžička, Lars Diening. *Non-Newtonian fluids and function spaces*. Institute of Mathematics AS CR, 2007.

[14] Vít Dolejší , Michal Kuraz, Pavel Solin. *Adaptive higher-order space-time discontinuous Galerkin method for the computer simulation of variably-saturated porous media flows*. Elsevier B.V, 2019.

[15] P. G. Ciarlet. *The Finite Elements Method for Elliptic Problems*. North-Holland, Amsterdam, New York, Oxford, 1979.

[16] V. Dolejší and M. Feistauer. *Discontinuous Galerkin Method - Analysis and Applications to Compressible Flow*. Springer Series in Computational Mathematics 48. Springer, Cham, 2015.

[17] L. Diening, M. Růžička. *Interpolation operators in Orlicz-Sobolev spaces*. Springer-Verlag, 2007.

[18] L. Diening, F. Ettwein. *Fractional estimates for non-differentiable elliptic systems with general growth, 523–556.*. Forum Mathematicum 20, 2008.

[19] Diening, Harjulehto, Hästö, Růžička. *Lebesgue and Sobolev Spaces with Variable Exponents*. Heidelberg: Springer, 2011.

[20] V. Kokilashvili, M. Krbec. *Weighted Inequalities in Lorentz and Orlicz Spaces*. River Edge, NJ:World Scientific pp.233, 1991.

[21] Vít Dolejší , Filip Roskovec, Miloslav Vlasák. *Residual based error estimates for the space–time discontinuous Galerkin method applied to the compressible flows*. Elsevier B.V, 2015.

List of Figures

8.1 Error estimates in H^1 norm, Example 1 63
8.2 Error estimates in L^2 norm, Example 1 63
8.3 Seepage through the hole in the subsurface, Forchheimer equation, result at $1.5s$ 64
8.4 Seepage through the hole in the subsurface, Forchheimer equation, result at $3s$ 64
8.5 Seepage through the hole in the subsurface, Forchheimer equation, result at $5s$ 65
8.6 Seepage through the hole in the subsurface, Forchheimer equation, result at $7.5s$ 65
8.7 Seepage through the hole in the subsurface, Forchheimer equation, result at $10s$ 65
8.8 Seepage through the hole in the subsurface, Darcy equation, result at $1.5s$ 66
8.9 Seepage through the hole in the subsurface, Darcy equation, result at $3s$ 66
8.10 Seepage through the hole in the subsurface, Darcy equation, result at $5s$ 67
8.11 Seepage through the hole in the subsurface, Darcy equation, result at $7.5s$ 67
8.12 Seepage through the hole in the subsurface, Darcy equation, result at $10s$ 67