FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Tomáš Hammerbauer

Domain decomposition methods for the solution of partial differential equations using discontinuous Galerkin method

Department of Numerical Mathematics

Supervisor of the master thesis:	prof. RNDr. Vít Dolejší, Ph.D., DSc.
Study programme:	Mathematics
Study branch:	Computational Mathematics

Prague 2024

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

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 ....
Author's signature

I would like to express my gratitude to prof. Vít Dolejší for his help a valuable advice. I would also like to thank my family and my girlfriend for their help and support during my study.

Title: Domain decomposition methods for the solution of partial differential equations using discontinuous Galerkin method

Author: Tomáš Hammerbauer

Department: Department of Numerical Mathematics

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

Abstract: This thesis deals with the analysis and numerical study of the domain decomposition method based preconditioner for algebraic systems arising from the discontinuous Galerkin (DG) discretization of the linear elliptic problems. We introduce the DG discretization of the model problem. We derive from the properties of the bilinear form the spectral bounds of corresponding forms and matrices. Moreover, we present the Additive Schwarz method and its application as a preconditioner for the system of algebraic equations. We derive the spectral bounds of the preconditioned system of algebraic equations. Finally, we present the numerical results that support the theoretical results and demonstrate the potential of this approach.

Keywords: domain decomposition method, elliptic partial differential equations, additive Schwarz preconditioner

Contents

Introduction **2**

List of used notation **3**

1 Discontinuous Galerkin Method **4**

1.1 Model Problem 4

1.2 Discontinuous Galerkin discretization 4

1.2.1 Partitioning of domain Ω 4

1.2.2 Discontinuous Galerkin method based on primal formulation 7

1.2.3 Basic properties of DGM 8

1.2.4 Equivalence with the system of linear algebraic equations 13

1.3 *hp* - condition number estimates 13

2 Domain decomposition method **17**

2.1 Subdomain partitioning, Local and Coarse solvers 17

2.1.1 Local Solvers 17

2.1.2 Coarse solver 18

2.1.3 Variational formulation 18

2.2 Convergence analysis 19

2.3 Implementation 25

3 Numerical experiments **28**

3.1 Laplace equation 28

3.2 Symmetric linear elliptic equation 31

3.3 Symmetric nonlinear elliptic equation 35

Conclusion **39**

Bibliography **40**

Introduction

Discontinuous Galerkin method is method developed for solving the partial differential equations. The principle of the method is to use partitioning (mesh) of the computational domain into finite number of elements, where we approximate the solution using polynomials of some degree. The standard conforming finite element method approximates the solution by functions, that are piecewise polynomial on the mesh and that are continuous in the whole domain. This is not the case of the DGM, which is based on a piecewise polynomial but discontinuous approximation. In order to guarantee the well-posedness of the numerical scheme, we have to introduce some term, which mimics the continuity on the boundaries of the elements. This term is called the interior penalty.

The discretization of the partial differential equation leads to a large sparse algebraic system, which is usually solved by a suitable iterative solver. The solution of algebraic systems exhibits usually the most time consuming part of the whole computational process.

The domain decomposition method was developed in the end of 19th century for the computation of partial differential equations, see the article Schwarz [1870]. Nowadays it is being studied again to use with modern numerical methods, where we must deal with large problems and the use of the supercomputers need the parallelization of the computations to be efficient. The DDM decomposes the computational domain into smaller subdomains, where the problem is computed separately and then it is put back together. There are also many versions of DDM, so we will only focus on the Additive Schwarz (AS) method. In particular, we analyze and numerically verify the AS preconditioners, which significantly reduce the computational costs of iterative solvers.

In the first chapter we introduce the DGM on a model problem. The analysis can be extended to more complicated problems with different boundary conditions. We prove the continuity and coercivity of the billinear form from DGM, since these two properties are necessary for the condition number bounds. In the second chapter, we formulate the Additive Schwarz method and using three assumptions on the local solvers we obtain the condition bounds for the preconditioned system arising from DGM. In the final chapter we introduce the results of numerical experiments performed to back up the analysis.

List of used notation

$\mathbb{N}$	natural numbers
$\mathbb{R}$	real numbers
∇	gradient
$\partial\Omega$	boundary of domain Ω
$\overline{\Omega}$	closure of domain Ω
$\lambda(\mathcal{A})$	eigenvalue of bilinear form $\mathcal{A}$
$\lambda(\boldsymbol{A})$	eigenvalue of matrix $\boldsymbol{A}$
$\kappa(\mathcal{A})$	condition number of billinear form $\mathcal{A}$
$\kappa(\boldsymbol{A})$	condition number of matrix $\boldsymbol{A}$
$\boldsymbol{x} \cdot \boldsymbol{y}$	scalar product of two vectors $\boldsymbol{x}$, $\boldsymbol{y}$

1. Discontinuous Galerkin Method

In this chapter, we introduce Discontinuous Galerkin method for the numerical solution of a model problem. We slightly extend the results from the monograph (Dolejší and Feistauer [2015]) which will be used later in the analysis of the domain decomposition method.

1.1 Model Problem

By $\Omega \subseteq \mathbb{R}^d, d = 2, 3$, we denote a bounded domain with polygonal, Lipschitz boundary $\partial\Omega$. We use the notation $L^2(\Omega)$ for the space of square integrable Lebesgue functions and $W^{s,p}(\Omega)$, $1 \leq p \leq \infty$, $s \in \mathbb{N}$ for the Sobolev spaces. In particular we set $H^s(\Omega) := W^{s,2}(\Omega)\,, s > 0$. We denote by $|\cdot|_{s,\Omega}$ and $\|\cdot\|_{s,\Omega}$ the standard Sobolev seminorm and norm, respectively, defined on $H^s(\Omega)$. Moreover, we denote by $H_0^1(\Omega)$ the space of functions from $H^1(\Omega)$ that have zero trace on the boundary $\partial\Omega$. Finally, we denote by $(\cdot,\cdot)_\Omega$ the standard inner product in $[L^2(\Omega)]^d$.

We consider the following problem. Let $f = f(x) \in L^2(\Omega)$ be given, we seek $u = u(x)$ such that

$$-\mathrm{div}(\boldsymbol{K}\nabla u) = f \quad \text{in } \Omega \tag{1.1}$$

$$u = 0 \quad \text{on } \partial\Omega, \tag{1.2}$$

where $\boldsymbol{K} = \boldsymbol{K}(x)$ is a symmetric positive definite matrix in $\mathbb{R}^{d\times d}$. We assume that $\exists k_0,\, k_1 > 0$ such that,

$$k_0|\xi| \leq |\boldsymbol{K}\xi| \leq k_1|\xi| \qquad \forall \xi \in \mathbb{R}^d. \tag{1.3}$$

For simplicity, we consider homogeneous Dirichlet boundary condition on $\partial\Omega$, but the results can by simply extended to a more general case. For the completeness we state the definition of the weak solution.

Definition 1 (Weak solution). *A weak solution of problem* (1.1) *is function* $u \in H_0^1(\Omega)$*, that satisfies the following identity*

$$\int_\Omega \boldsymbol{K}\nabla u \cdot \nabla v \; dx = \int_\Omega fv \; dx \qquad \forall v \in H_0^1(\Omega). \tag{1.4}$$

The existence of the weak solution can be proven by the Lax-Milgram lemma.

1.2 Discontinuous Galerkin discretization

1.2.1 Partitioning of domain Ω

Let $\mathcal{T}_h, h > 0$ be a partition of $\overline{\Omega}$ into a finite number of closed d-dimensional non-overlapping simplexes K, such that

$$\bigcup_{K\in\mathcal{T}_h} \overline{K} = \overline{\Omega}. \tag{1.5}$$

Every element $K \in \mathcal{T}_h$ is an image of fixed master element $\hat{K}$, and $\hat{K}$ is the open unit d-simplex in $\mathbb{R}^d$. In $d = 3$, K is a tetrahedron, but we call it triangle. The partition $\mathcal{T}_h$ is called a *triangulation* of Ω, we do not assume the standard conforming properties, so we are allowing hanging nodes in the triangulation. Further we will use following notation. We denote by ∂K the boundary of element K and h_K as its diameter. We set $h = \max_{K \in \mathcal{T}_h} h_K$.

Let $K, K' \in \mathcal{T}_h$. We say that K and K' are *neighboring elements* of the triangulation $\mathcal{T}_h$, if $\partial K \cap \partial K'$ has a positive $d-1$ dimensional measure. We say that $\gamma \subset \partial K$ is a *face* of element K, if γ is maximal connected subset of either $\partial K \cap \partial K'$, for K' neighboring element of K, or $\partial K \cap \partial \Omega$. The $(d-1)$ dimensional Lebesgue measure of γ we denote as $|\gamma|$ and the same notation will also be used for the d dimensional Lebesgue measure of the simplex K.

Further, let $\mathcal{F}_h$ denote the union of all faces ($d = 3$) or edges ($d = 2$) of all triangles in the triangulation $\mathcal{T}_h$. We will use the term "face" even for $d = 2$ in the following text for simplicity. Furthermore we will distinguish between interior and boundary faces as

- the set of boundary faces denoted by
$$\mathcal{F}_h^B = \{\gamma \in \mathcal{F}_h : \gamma \subset \partial\Omega\}, \tag{1.6}$$
- and the set of inner faces denoted by
$$\mathcal{F}_h^I = \mathcal{F}_h \setminus \mathcal{F}_h^B. \tag{1.7}$$

Let $\boldsymbol{p} := \{p_K : K \in \mathcal{T}_h\}$ be a set of integers that assigns to each element of triangulation its polynomial degree of approximation. We also assume that the polynomial order has *local bounded variant.* Which means that there exists constant $C_V > 0$, such that
$$\frac{p_{K_1}}{p_{K_2}} \leq C_V, \tag{1.8}$$
for any pair of elements K_1 and K_2 sharing a face.

In some theorems we will use the *broken Sobolev space* defined by the following for $s \in \mathbb{N}$
$$H^s(\Omega, \mathcal{T}_h) := \{v \in L^2(\Omega) : v|_K \in H^s(K)\, \forall K \in \mathcal{T}_h\}. \tag{1.9}$$
For $v \in H^s(\Omega, \mathcal{T}_h)$, we define the the norm
$$\|v\|^2_{H^s(\Omega,\mathcal{T}_h)} = \sum_{K \in \mathcal{T}_h} \|v\|^2_{s,K} \tag{1.10}$$
and the seminorm
$$|v|^2_{H^s(\Omega,\mathcal{T}_h)} = \sum_{K \in \mathcal{T}_h} |v|^2_{s,K}. \tag{1.11}$$
We define the space of discontinuous piecewise polynomial functions as
$$S_{hp} := \{v \in L^2(\Omega) : v|_K \in P_{p_K}(K) \forall K \in \mathcal{T}_h\}, \tag{1.12}$$
where $P_{p_K}(K)$ denotes the space of polynomials of degree less than p_K on K. Furthermore we have that
$$S_{hp} \subset H^s(\Omega, \mathcal{T}_h), \qquad s \in \mathbb{N}. \tag{1.13}$$
For the sake of simplicity we will use in text the generic constant $C > 0$, that does not depend on h and p. The constants that are somewhat important, will have assigned index.

Jump notation

Let $v \in H^s(\Omega, \mathcal{T}_h)$ and $\boldsymbol{q} \in [H^s(\Omega, \mathcal{T}_h)]^d$, $s \in \mathbb{N}$, be a vector and scalar valued functions smooth in the interior of $K \in \mathcal{T}_h$. We denote by $v^{\pm}$ and $\boldsymbol{q}^{\pm}$ the trace of functions v and $\boldsymbol{q}$ on two neighboring elements $K^+, K^- \in \mathcal{T}_h$ sharing interior face γ. Using this notation we will introduce definition of jump $[\cdot]$ and mean value $\langle\cdot\rangle$ of function on faces as

$$\begin{aligned} [v] &= v^+\boldsymbol{n}_+ + v^-\boldsymbol{n}_-, \qquad & [\boldsymbol{q}] &= \boldsymbol{q}^+ \cdot \boldsymbol{n}_+ + \boldsymbol{q}^- \cdot \boldsymbol{n}_-, \\ \langle v\rangle &= \frac{1}{2}(v^+ + v^-), & \langle \boldsymbol{q}\rangle &= \frac{1}{2}(\boldsymbol{q}^+ + \boldsymbol{q}^-), \end{aligned} \tag{1.14}$$

where $\boldsymbol{n}_{\pm}$ denotes unit outward normal for an element $K_{\pm}$ respectively. On a boundary face $\gamma \in \mathcal{F}^B$ we define jump and mean value in the following way $[v] = v\boldsymbol{n}$, $[\boldsymbol{q}] = \boldsymbol{q}\cdot\boldsymbol{n}$, $\langle v\rangle = v$, $\langle\boldsymbol{q}\rangle = \boldsymbol{q}$.

Asumption on meshes

We consider a system of triangulation $\{\mathcal{T}_h\}_{h\in(0,\bar{h})}$, $\bar{h} \geq 0$. The following assumption are used for the continuity of bilinear form, that we will get from DGM discretization.

- The system of triangulation $\{\mathcal{T}_h\}_{h\in(0,\bar{h})}$, $\bar{h} \geq 0$ is *shape - regular* if there exist constant C_R such that

$$\frac{h_K}{\rho_K} \leq C_R \quad \forall K \in \mathcal{T}_h \,\forall h \in (0, \bar{h}). \tag{1.15}$$

Moreover we need to introduce the quantity h_γ for $\gamma \in \mathcal{F}_h$, which is a counterpart to h_K on the faces. For this quantity we will assume the equivalence condition.

- The system of triangulations $\{\mathcal{T}_h\}_{h\in(0,\bar{h})}$ satisfy the *equivalence condition*, if there exists constants $C_U \geq 0$ and $C_L \geq 0$ such that

$$C_L h_K \leq h_\gamma \leq C_U h_K \quad \forall K \in \mathcal{T}_h \,\forall \gamma \in \mathcal{F}_h, \gamma \subset \partial K \,\forall h \in (0, \bar{h}). \tag{1.16}$$

The equivalency condition (1.16) can be fulfilled by the suitable choice of h_γ based on additional assumption on the family of triangulations $\{\mathcal{T}_h\}_{h\in(0,\bar{h})}$. For example:

- If faces $\gamma \subset \partial K$ do not degenerate with respect to the diameter of K, $h \to 0$, we can rewrite it as:

$$\exists C_d \geq 0 \text{ such that } \frac{h_K}{\operatorname{diam}(\gamma)} \leq C_d \quad \forall K \in \mathcal{T}_h, \forall \gamma \in \mathcal{F}_h,\, \gamma \subset \partial K, \tag{1.17}$$

  then

$$h_\gamma = \operatorname{diam}(\gamma). \tag{1.18}$$

- The family of triangulations $\{\mathcal{T}_h\}_{h\in(0,\bar{h})}$ is *locally quasi − uniform*

$$\exists C_Q \geq 0 \quad \forall K,\, K' \in \mathcal{T}_h,\, K,\, K' neighbors, \forall h \in (0, \bar{h}): \quad h_K \leq C_Q h_{K'}. \tag{1.19}$$

- The family of triangulations $\{\mathcal{T}_h\}_{h\in(0,\bar{h})}$ is called *quasi-uniform* if:
$$\exists C_U \geq 0\, \forall K \in \mathcal{T}_h \quad h \leq C_U h_K. \tag{1.20}$$

We can set the parameter h_γ as follows.

- If $\mathcal{T}_h$ is conforming (no hanging nodes are allowed), then we can set
$$h_\gamma := \mathrm{diam}(\gamma). \tag{1.21}$$
- If $\mathcal{T}_h$ is local quasi-uniform, then we can set
$$h_\gamma := \max_{\substack{K,K'\in\mathcal{T}_h \\ \gamma\subset\partial K\cap\partial K'}} (h_K, h_{K'}) \tag{1.22}$$

Finally for an edge γ we can define the polynomial degree p_γ by:

$$p_\gamma := \begin{cases} \max\{p_{K'}, p_K\} & \text{if } \gamma \subset \partial K' \cap \partial K, \\ p_K & \text{if } \gamma \subset \partial K \cap \partial\Omega. \end{cases} \tag{1.23}$$

In our experiments, we will use the quasi-uniform mesh with same polynomial degree on every element K.

1.2.2 Discontinuous Galerkin method based on primal formulation

Now we are ready to introduce the symmetric interior penalty variant of DGM. Multiplying the probem (1.1) with function $v \in H^2(\Omega, \mathcal{T}_h)$, summing over K and using Green's theorem, we obtain the identity

$$\mathcal{A}_h(u, v) = (f, v)_\Omega \quad \forall v \in H^2(\Omega, \mathcal{T}_h), \tag{1.24}$$

where $\mathcal{A}_h : H^2(\Omega, \mathcal{T}_h) \times H^2(\Omega, \mathcal{T}_h) \to \mathbb{R}$ is bilinear form given by

$$\mathcal{A}_h(u, v) := a_h(u, v) + \mathrm{J}_h^\sigma(u, v), \qquad u, v \in H^2(\Omega, \mathcal{T}_h), \tag{1.25}$$

where

$$a_h(u, v) := \sum_{K\in\mathcal{T}_h} \int_K \boldsymbol{K}\nabla u \cdot \nabla v \ \mathrm{dx} - \sum_{\gamma\in\mathcal{F}_h^I} \int_\gamma (\langle \boldsymbol{K}\nabla u\rangle \cdot [v] + \langle \boldsymbol{K}\nabla v\rangle \cdot [u]) \ \mathrm{dS} \tag{1.26}$$

and

$$\mathrm{J}_h^\sigma(u, v) := \sum_{\gamma\in\mathcal{F}_h^I} \int_\gamma \sigma\, [u]\, [v] \ \mathrm{dS} \quad v \in H^2(\Omega, \mathcal{T}_h). \tag{1.27}$$

The form J_h^σ is called the interior and boundary penalty bilinear form, that we introduced in order to mimic the continuity of conforming FEM. The parameter $\sigma \geq 0$ is chosen arbitrary. For our analysis we will use the following representation. For $\sigma : \bigcup_{\gamma\in\mathcal{F}_h^I} \to \mathbb{R}$ we use

$$\sigma|_\gamma = \sigma_\gamma = \alpha \frac{k_0 p_\gamma^2}{h_\gamma}, \quad \gamma \in \mathcal{F}_h^I, \tag{1.28}$$

where k_0 is from (1.3) and $\alpha > 0$ is some positive constant. In the analysis, we will show some bound for it.

Finally, we can formulate the definition of the approximate solution u_h.

Definition 2. *The function $u_h \in S_{hp}$ is called an approximate solution of* (1.4) *if*

$$\mathcal{A}_h(u_h, v) = (f, v)_\Omega \quad \forall v \in S_{hp}. \tag{1.29}$$

This scheme is called the symmetric interior penalty Galerkin (SUPG) method.

Now that we have formulated the form $\mathcal{A}_h$ we discuss its properties.

1.2.3 Basic properties of DGM

We introduce a few inequalities that will be useful in the numerical analysis. First, we prove some properties of the norms in the space S_{hp} and form $\mathcal{A}_h$. We define the DG-norm by the following

$$|||u|||_{\mathcal{T}_h} := (k_0|u|^2_{H^1(\Omega,\mathcal{T}_h)} + \mathrm{J}^\sigma_h(u,u))^{1/2}, \qquad \text{where}\, k_0 \text{ is from (1.3)}, \tag{1.30}$$

In the following text we will shall omit the subscript $\mathcal{T}_h$ and write just $|||\cdot|||$.

Inverse inequality and multiplicative trace inequality

For the analysis, it is important to show the relations between the norms that we use. The following results are taken from Dolejší and Feistauer [2015] and Antonietti and Houston [2011]. We will need them to prove the continuity and coercivity of the bilinear form $\mathcal{A}_h$.

Lemma 1 (Inverse inequality)**.** *Let the shape-regularity assumption* (1.15) *be satisfied. Then there exists a constant $C_I \geq 0$ independent of v, h, p and K such that*

$$|v|^2_{1,K} \leq C_I \frac{p_K^4}{h_K^2} \|v\|^2_{0,K} \quad \forall v \in P_{p_K}(K), \forall K \in \mathcal{T}_h, \forall h \in (0, \tilde{h}). \tag{1.31}$$

Lemma 2 (Multiplicative trace inequality 1)**.** *Let the shape-regularity assumption* (1.15) *hold. Then there exists a constant $C_{\tilde{M}} > 0$ independent of v, h and K such that*

$$\|v\|^2_{0,\partial K} \leq C_{\tilde{M}} \left(\|v\|_{0,K}|v|_{1,K} + h_K^{-1}\|v\|^2_{0,K} \right), \quad \forall K \in \mathcal{T}_h, \\ \forall v \in H^1(K),\ \forall h \in (0, \tilde{h}). \tag{1.32}$$

Proof can be found in (Dolejší and Feistauer [2015]).

Lemma 3 (Multiplicative trace inequality 2)**.** *Let the shape-regularity assumption* (1.15) *hold. Then there exists a constant $C_M > 0$, that is independent of v, h and K*

$$\|v\|^2_{0,\partial K} \leq C_M \frac{p_K^2}{h_K} \|v\|^2_{0,K} \quad \forall v \in P_{p_K}(K), \forall K \in \mathcal{T}_h, \forall h \in (0, \tilde{h}). \tag{1.33}$$

Proof. The proof easily follows from (1.32) and (1.31). Let $v \in P_{p_K}(K)$ for some $K \in \mathcal{T}_h$, we have that

$$\begin{aligned} \|v\|^2_{0,\partial K} &\leq C_{\tilde{M}} \left(\|v\|_{0,K}|v|_{1,K} + h_K^{-1}\|v\|^2_{0,K} \right) \\ &\leq C_{\tilde{M}} \left(\sqrt{C_I} \frac{p_K^2}{h_K} \|v\|^2_{0,K} + \frac{1}{h_K}\|v\|^2_{0,K} \right) \\ &\leq C_{\tilde{M}} (1 + \sqrt{C_I}) \frac{p_K^2}{h_K} \|v\|^2_{0,K}. \end{aligned} \tag{1.34}$$

Setting $C_M = C_{\tilde{M}}(1 + \sqrt{C_I})$ is the proof completed. □

For the norm of the jump we will need the following result, which follows from the definition of the mean value (1.14) and the equivalence condition (1.16). For $v \in H^1(\Omega, \mathcal{T}_h)$ we have

$$\sum_{\gamma \in \mathcal{F}_h^I} h_\gamma \|\langle v \rangle\|_{0,\gamma}^2 \leq C_U \sum_{K \in \mathcal{T}_h} h_K \|v\|_{0,\partial K}^2. \tag{1.35}$$

Continuity of bilinear form for SIPG

First in our analysis we will show the continuity of the form $\mathcal{A}_h$. For this we adopt the analysis done in (Dolejší and Feistauer [2015]). The proofs of the theorems that we mention, are modification of the proofs mentioned in the book. We formulate auxillary lemmas that we change to fit our model problem and that will help us prove the continuity in DG-norm.

First we will use the Corollary 1.33. from (Dolejší and Feistauer [2015]).

Lemma 4. *Let the system of triangulation $\{\mathcal{T}_h\}_{h \in (0,\bar{h})}$ be shape-regular* (1.15) *and let the quantity h_γ, $\gamma \in \mathcal{F}_h$, $h \in (0, \bar{h})$, satisfy the equivalence condition* (1.16)*. Let the parameter σ be defined by* (1.28)*. Then the form $\mathcal{A}_h$, defined above, satisfies the estimate*

$$|\mathcal{A}_h(u,v)| \leq 2\frac{k_1}{k_0} \|u\|_{1,\sigma} \|v\|_{1,\sigma} \quad \forall u, v \in H^2(\Omega, \mathcal{T}_h), \tag{1.36}$$

where

$$\|v\|_{1,\sigma}^2 = \|\!|v|\!\|^2 + \sum_{\gamma \in \mathcal{F}_h^I} \int_\gamma k_0^2 \sigma^{-1} (\boldsymbol{n} \cdot \langle \nabla v \rangle)^2 dS. \tag{1.37}$$

Proof. We will prove the bounds for the specific parts of the billinear form $\mathcal{A}_h$. First

$$\begin{aligned} |a_h(u,v)| \leq & \underbrace{\sum_{K \in \mathcal{T}_h} \int_K |\boldsymbol{K} \nabla u \cdot \nabla v| \ \mathrm{dx}}_{\delta_1} \\ & + \underbrace{\sum_{\gamma \in \mathcal{F}_h^I} \int_\gamma |\boldsymbol{n} \cdot \langle \boldsymbol{K} \nabla u, \nabla v \rangle| \ \mathrm{dS}}_{\delta_2} + \underbrace{\sum_{\gamma \in \mathcal{F}_h^I} \int_\gamma |\boldsymbol{n} \cdot \langle \boldsymbol{K} \nabla v, \nabla u \rangle| \ \mathrm{dS}}_{\delta_3}. \end{aligned} \tag{1.38}$$

We will start with bound to δ_1 using a Cauchy-Schwarz inequality for integral and then for series and using (1.3).

$$\delta_1 \leq \sum_{K \in \mathcal{T}_h} \frac{k_1}{k_0} k_0^{\frac{1}{2}} |u|_{1,K} k_0^{\frac{1}{2}} |v|_{1,K} \leq \frac{k_1}{k_0} k_0^{\frac{1}{2}} |u|_{H^1(\Omega, \mathcal{T}_h)} k_0^{\frac{1}{2}} |v|_{H^1(\Omega, \mathcal{T}_h)}. \tag{1.39}$$

For the δ_2 we will again use the Cauchy-Schwarz ineguality and the (1.3)

$$\begin{aligned} \delta_2 \leq & \sum_{\gamma \in \mathcal{F}_h^I} \left(\frac{k_1^2}{k_0^2} \int_\gamma k_0^2 \sigma^{-1} (\boldsymbol{n} \cdot \langle \nabla u \rangle)^2 \ \mathrm{dS} \right)^{\frac{1}{2}} \left(\int_\gamma \sigma [v]^2 \ \mathrm{dS} \right)^{\frac{1}{2}} \\ \leq & \left(\frac{k_1}{k_0} \sum_{\gamma \in \mathcal{F}_h^I} \int_\gamma k_0^2 \sigma^{-1} (\boldsymbol{n} \cdot \langle \nabla u \rangle)^2 \ \mathrm{dS} \right)^{\frac{1}{2}} \left(\frac{k_1}{k_0} \sum_{\gamma \in \mathcal{F}_h^I} \int_\gamma \sigma [v]^2 \ \mathrm{dS} \right)^{\frac{1}{2}}, \end{aligned} \tag{1.40}$$

and we will do the same manipulation for δ_3. From that we get

$$\delta_3 \leq \left(\frac{k_1}{k_0} \sum_{\gamma \in \mathcal{F}_h^I} \int_\gamma k_0^2 \sigma^{-1} (\boldsymbol{n} \cdot \langle \nabla v \rangle)^2 \text{ dS} \right)^{\frac{1}{2}} \left(\frac{k_1}{k_0} \sum_{\gamma \in \mathcal{F}_h^I} \int_\gamma \sigma [u]^2 \text{ dS} \right)^{\frac{1}{2}}. \tag{1.41}$$

Now putting (1.38) - (1.41) together and using the discrete Cauchy-Schwarz inequlity we obtain,

$$\begin{aligned} a_h(u,v) &\leq \frac{k_1}{k_0} k_0^{\frac{1}{2}} |u|_{H^1(\Omega,\mathcal{T}_h)} k_0^{\frac{1}{2}} |v|_{H^1(\Omega,\mathcal{T}_h)} \\ &+ \left(\frac{k_1}{k_0} \sum_{\gamma \in \mathcal{F}_h^I} \int_\gamma k_0^2 \sigma^{-1} (\boldsymbol{n} \cdot \langle \nabla u \rangle)^2 \text{ dS} \right)^{\frac{1}{2}} \left(\frac{k_1}{k_0} \sum_{\gamma \in \mathcal{F}_h^I} \int_\gamma \sigma [v]^2 \text{ dS} \right)^{\frac{1}{2}} \\ &+ \left(\frac{k_1}{k_0} \sum_{\gamma \in \mathcal{F}_h^I} \int_\gamma k_0^2 \sigma^{-1} (\boldsymbol{n} \cdot \langle \nabla v \rangle)^2 \text{ dS} \right)^{\frac{1}{2}} \left(\frac{k_1}{k_0} \sum_{\gamma \in \mathcal{F}_h^I} \int_\gamma \sigma [u]^2 \text{ dS} \right)^{\frac{1}{2}} \\ &\leq \frac{k_1}{k_0} \left(k_0 |u|^2_{H^1(\Omega,\mathcal{T}_h)} + \sum_{\gamma \in \mathcal{F}_h^I} \int_\gamma k_0^2 \sigma^{-1} (\boldsymbol{n} \cdot \langle \nabla u \rangle)^2 \text{ dS} + J_h^\sigma(u,u) \right)^{\frac{1}{2}} \\ &\times \left(k_0 |v|^2_{H^1(\Omega,\mathcal{T}_h)} + \sum_{\gamma \in \mathcal{F}_h^I} \int_\gamma k_0^2 \sigma^{-1} (\boldsymbol{n} \cdot \langle \nabla v \rangle)^2 \text{ dS} + J_h^\sigma(v,v) \right)^{\frac{1}{2}} \\ &\leq \frac{k_1}{k_0} \|u\|_{1,\sigma} \|v\|_{1,\sigma}. \end{aligned} \tag{1.42}$$

Now for the term $J_h^\sigma(u,v)$ we have from Cauchy-Schwarz, that

$$|J_h^\sigma(u,v)| \leq J_h^\sigma(u,u)^{\frac{1}{2}} J_h^\sigma(v,v)^{\frac{1}{2}}. \tag{1.43}$$

Putting (1.38) - (1.43) together and using the definition of $\||\cdot\||$ norm, we get

$$\begin{aligned} |\mathcal{A}_h(u,v)| &\leq |a_h(u,v)| + |\mathrm{J}_h^\sigma(u,v)| \leq \frac{k_1}{k_0} \|u\|_{1,\sigma} \|v\|_{1,\sigma} + \mathrm{J}_h^\sigma(u,u)^{\frac{1}{2}} \mathrm{J}_h^\sigma(v,v)^{\frac{1}{2}} \\ &\leq \frac{k_1}{k_0} \|u\|_{1,\sigma} \|v\|_{1,\sigma} + \|u\|_{1,\sigma} \|v\|_{1,\sigma} = 2 \frac{k_1}{k_0} \|u\|_{1,\sigma} \|v\|_{1,\sigma} \end{aligned} \tag{1.44}$$

□

The next lemma will be dealing with bounding of the jump term $J_h^\sigma(u,u)^{\frac{1}{2}}$. We will also get the bound for the norm $\|\cdot\|_{1,\sigma}$, that we will need for the continuity of the bilinear form in the DG-norm.

Lemma 5. *Under the assumption as in Lemma 4, there exists constant $C_\sigma \geq 0$ such that*

$$J_h^\sigma(u,u)^{\frac{1}{2}} \leq \||u\|| \leq \|u\|_{1,\sigma} \quad \forall u \in H^2(\Omega, \mathcal{T}_h),\ h \in (0, \bar{h}), \tag{1.45}$$

and

$$J_h^\sigma(v_h,v_h)^{\frac{1}{2}} \leq \||v_h\|| \leq \|v_h\|_{1,\sigma} \leq C_\sigma \||v_h\|| \quad \forall v_h \in S_{hp},\ h \in (0, \bar{h}). \tag{1.46}$$

Proof. All of the inequalities are trivial except for the last inequality $\|v_h\|_{1,\sigma} \leq C_\sigma \|\!|v_h|\!\|$. For this inequality we need the following bound, using estimate (1.35) and (1.8) we get

$$\begin{aligned}\sum_{\gamma \in \mathcal{F}_h^I} \int_\gamma k_0^2 \sigma^{-1} (\boldsymbol{n} \cdot \langle \nabla v \rangle)^2 \mathrm{d}S &\leq \frac{Ck_0}{\alpha} \sum_{K \in \mathcal{T}_h} \frac{h_K}{p_K^2} \|\nabla v\|_{0,\partial K} \\ &\leq \frac{Ck_0}{\alpha} \sum_{K \in \mathcal{T}_h} \frac{h_K}{p_K^2} C_M \frac{p_K^2}{h_K} |v|_{1,K} \\ &\leq \frac{CC_M k_0}{\alpha} \sum_{K \in \mathcal{T}_h} |v|_{1,K}.\end{aligned} \tag{1.47}$$

Hence we can add this term into the DG-norm and we get the inequality with constant C_σ independent of p_K. □

Now for the we have the auxillary estimates prepared for proving the continuity of form $\mathcal{A}_h$ in the DG-norm.

Theorem 6 (Continuity of the billinear form)**.** *Let the system of triangulation* $\{\mathcal{T}_h\}_{h \in (0,\bar{h})}$ *be shape regular and let the quantity* h_γ*,* $\gamma \in \mathcal{F}_h$*,* $h \in (0, \bar{h})$*, satisfy the equivalence condition. Let the parameter* σ *be defined by* (1.28)*. Then there exists a positive constant* $C_\sigma \geq 0$ *such that*

$$|\mathcal{A}_h(u, v)| \leq 2C_\sigma \frac{k_1}{k_0} \|\!|u|\!\| \, \|\!|v|\!\| \qquad \forall u,\, v \in S_{hp}. \tag{1.48}$$

Proof. The proof of this theorem follows from the auxillary lemmas, that we stated before. More precisely Lemma 4 and Lemma 5. □

Coercivity of bilinear form for SIPG

The following result will be needed in the proof of Theorems 8 and 11.

Theorem 7 (Coercivity)**.** *Let the system of triangulation* $\{\mathcal{T}_h\}_{h \in (0,\bar{h})}$ *be shape regular and let the quantity* h_γ*,* $\gamma \in \mathcal{F}_h$*,* $h \in (0, \bar{h})$*, satisfy the equivalence condition. Let*

$$\alpha \geq \frac{k_1^2 4 C_U C_M C_V}{k_0^2}, \tag{1.49}$$

where C_U*,* C_M *and* C_V *are constant from* (1.16)*,* (1.33) *and* (1.8)*, respectively, and let the penalty parameter be given by* (1.28) *for all* $\gamma \in \mathcal{F}_h$*. Then*

$$\mathcal{A}_h(v_h, v_h) \geq \frac{1}{2} \|\!|v_h|\!\|^2 \quad \forall v_h \in S_{hp} \, \forall h \in (0, \bar{h}). \tag{1.50}$$

Proof. Let $\delta \geq 0$ be a constant that we will specify later. Then from definition of the form $a_h(\cdot,\cdot)$, boundedness of eigenvalues of $\boldsymbol{K}$ and from Cauchy and Young's inequality it follows that

$$\begin{aligned}
a_h(v_h, v_h) &\geq k_0 |v_h|^2_{H^1(\Omega,\mathcal{T}_h)} - 2 \sum_{\gamma\in\mathcal{F}_h^I} \int_\gamma \boldsymbol{n}\cdot\langle \boldsymbol{K}\nabla v_h\rangle [v_h] \text{ dS} \\
&\geq k_0 |v_h|^2_{H^1(\Omega,\mathcal{T}_h)} - 2\left\{ \frac{k_1}{\delta} \sum_{\gamma\in\mathcal{F}_h^I} \int_\gamma h_\gamma (\boldsymbol{n}\cdot\langle\nabla v_h\rangle)^2 \text{ dS} \right\}^{\frac{1}{2}} \left\{ k_1\delta \sum_{\gamma\in\mathcal{F}_h^I} \int_\gamma \frac{1}{h_\gamma}[v_h]^2 \text{ dS}\right\}^{\frac{1}{2}} \\
&\geq k_0 |v_h|^2_{H^1(\Omega,\mathcal{T}_h)} - \omega - \frac{k_1\delta}{k_0\alpha}\mathrm{J}_h^\sigma(v_h, v_h),
\end{aligned} \tag{1.51}$$

where

$$\omega := \frac{k_1}{\delta} \sum_{\gamma\in\mathcal{F}_h^I} \int_\gamma \frac{h_\gamma}{p_\gamma^2} |\langle\nabla v_h\rangle|^2 \text{ dS}. \tag{1.52}$$

Now we use the multiplicative trace inequality, equivalence condition for h_K and h_γ , bound (1.35) , assumption (1.8) and we get

$$\begin{aligned}
\omega &\leq \frac{C_U C_V k_1}{\delta} \sum_{K\in\mathcal{T}_h} \frac{h_K}{p_K^2} \|\nabla v_h\|^2_{0,\partial K} \\
&\leq \frac{C_M C_U C_V k_1}{\delta} \sum_{K\in\mathcal{T}_h} |v_h|^2_{1,K} \\
&\leq \frac{C_M C_U C_V k_1}{\delta} |v_h|^2_{H^1(\Omega,\mathcal{T}_h)}.
\end{aligned} \tag{1.53}$$

Now, we choose

$$\delta = \frac{2 C_M C_U C_V k_1}{k_0}. \tag{1.54}$$

Then it follows

$$\begin{aligned}
a_h(v_h, v_h) &\geq \frac{1}{2}\left(k_0 |v_h|^2_{H^1(\Omega,\mathcal{T}_h)} - \frac{4 C_M C_U C_V k_1^2}{\alpha k_0^2} \mathrm{J}_h^\sigma(v_h, v_h)\right) \\
&\geq \frac{1}{2}\left(k_0 |v_h|^2_{H^1(\Omega,\mathcal{T}_h)} - \mathrm{J}_h^\sigma(v_h, v_h)\right)
\end{aligned} \tag{1.55}$$

Now if we use this in the form $\mathcal{A}_h$ we get

$$\begin{aligned}
\mathcal{A}_h(v_h, v_h) &= a_h(v_h, v_h) + \mathrm{J}_h^\sigma(v_h, v_h) \\
&\geq \frac{1}{2}\left(k_0 |v_h|^2_{H^1(\Omega,\mathcal{T}_h)} + \mathrm{J}_h^\sigma(v_h, v_h)\right) \\
&\geq \frac{1}{2} |\!|\!| v_h |\!|\!|^2
\end{aligned} \tag{1.56}$$

□

For the completeness of the analysis, we introduce the theorems about the existence and uniqueness of the approximate solution and about the bounds of the error. We refer to (Dolejší and Feistauer [2015]) for the Lax-Milgram lemma and for the proofs of the following theorems.

Theorem 8. *There exists only one approximate solution of problem* (1.29).

Proof. By the Lax-Milgram lemma, the coercivity and boundedness of the bilinear form $\mathcal{A}_h$ implies the existence and uniqueness of the solution of the discrete problem (1.29). □

Theorem 9. *Let Ω be bounded, convex polygonal domain and let $u \in H^s(\Omega)$ is the solution of (1.1). Let $\mathcal{T}_h$ be the triangulation of Ω with all the assumptions in Section 1.2.1.Let $\boldsymbol{s}$ and $\boldsymbol{p}$ be the vectors defined in Dolejší and Feistauer [2015], such that $s_K \geq 2$ and $p_K \geq 1$, $\forall K \in \mathcal{T}_h$. Moreover, let all the assumptions for the boundedness and coercivity of the bilinear form $\mathcal{A}_h$ be satisfied then*

$$\|\!|u_h - u|\!\| \leq C \left(\sum_{K \in \mathcal{T}_h} \frac{h_K^{2(\mu_K - 1)}}{p_K^{2s_K - 3}} \|u\|_{s_K,K} \right), \tag{1.57}$$

where $\mu_K = \min\{p_K + 1, s_K\}$.

1.2.4 Equivalence with the system of linear algebraic equations

Choosing a suitable basis of a function space S_{hp}, we can get the equivalence of the problem (1.29), with the system of linear algebraic equations

$$\boldsymbol{A}\boldsymbol{u} = \boldsymbol{F}. \tag{1.58}$$

This system can be solved using a numerous numerical methods, for example GMRE, MINRES, Conjugate Gradients, etc.. Usually the system is very large and the traditional methods are very slow. Hence we use the preconditioning to increase the speed of the method. One of the things that can be done to increase the speed of convergence is to decrease the condition number. This doesn't mean that, if we lower the condition number that we get faster convergence all the time, but it is the usually true.

1.3 *hp* - condition number estimates

Now that we have proven essential properties of the form $\mathcal{A}_h(u, v)$ we will look at the estimates for the form. First we will define the spectrum of bilinear form $\mathcal{A}_h$.We will use the definition presented in Rynne and Youngson [2007].

Definition 3. *Let $\mathcal{A}_h$ be a bilinear form defined on the space S_{hp}, $dim(S_{hp}) = N$ and w_j be a nonzero function from S_{hp} such that*

$$\mathcal{A}_h(w_j, v) = \lambda_j \, (w_j, v)_\Omega \quad \forall v \in S_{hp},\ j = 1, \dots, N. \tag{1.59}$$

The functions w_j are called eigenfunctions and the scalars λ_j are called eigenvalues.

For the condition numbers are important the largest and smallest eigenvalues given by the following definitions

$$\lambda_{max}(\mathcal{A}_h) := \max_{v \neq 0} \frac{\mathcal{A}_h(v, v)}{(v, v)_\Omega}, \qquad \lambda_{min}(\mathcal{A}_h) := \min_{v \neq 0} \frac{\mathcal{A}_h(v, v)}{(v, v)_\Omega}. \tag{1.60}$$

Now we can formulate the definition of the condition number of bilinear form $\mathcal{A}_h$.

Definition 4. *The condition number $\kappa(\mathcal{A}_h)$ of the bilinear form $\mathcal{A}_h$ is defined by the following*

$$\kappa(\mathcal{A}_h) := \frac{\lambda_{max}}{\lambda_{min}}. \tag{1.61}$$

For the completeness of the analysis we, also state the definitions for matrices.

Definition 5. *Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be a real valued matrix and let there be a vector $v \in \mathbb{R}^n$, $\|\boldsymbol{v}\| = 1$ such, that*

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}. \tag{1.62}$$

The $\lambda \in \mathbb{R}$ is called eigenvalue of matrix $\boldsymbol{A}$ and the vector $\boldsymbol{v}$ is called a eigenvector. The condition number of symmetric matrix $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ is defined by the following

$$\kappa(\boldsymbol{B}) := \frac{|\lambda_{max}(\boldsymbol{B})|}{|\lambda_{min}(\boldsymbol{B})|}, \tag{1.63}$$

where $\lambda_{max}(\boldsymbol{B})$ and $\lambda_{min}(\boldsymbol{B})$ are maximal and minimal (by moduli) eigenvalues of $\boldsymbol{B}$ respectively.

Now let the functions $\{\varphi_j\}_{j=1}^N$ be the basis functions of the space S_{hp}, $\dim(S_{hp}) = N$. Then

$$S_{hp} = \operatorname{span}\{\varphi_1, \ldots, \varphi_N\}, \tag{1.64}$$

and for all functions $v \in S_{hp}$ we have the following representation

$$v = \sum_{j=1}^{N} v_j \varphi_j, \tag{1.65}$$

where $v_i = \mathbb{R}$, $i = 1, \ldots, N$. We use this representation to define stiffness matrix $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ and mass matrix $\boldsymbol{M} \in \mathbb{R}^{N\times N}$, that corresponds to the bilinear form $\mathcal{A}_h$ and scalar product in $L^2(\Omega)$ respectively. The entries of those matrices are given by

$$A_{ij} := \mathcal{A}_h(\varphi_j, \varphi_i), \qquad M_{ij} := (\varphi_j, \varphi_i)_\Omega \quad i, j = 1, \ldots, N. \tag{1.66}$$

Using these matrices we can rewrite the action of bilinear form and scalar product on two functions u, v by the following identities

$$\mathcal{A}_h(u, v) = \mathbf{u}^T \mathbf{A} \mathbf{v} \quad (u, v)_\Omega = \mathbf{u}^T \mathbf{M} \mathbf{v}, \tag{1.67}$$

where $\mathbf{u} = (u_1, \ldots, u_N)^T$ and $\mathbf{v} = (v_1, \ldots, v_N)$ are vectors of coefficients from the representation of the functions in the basis of S_{hp}, cf. (1.65). Then we deduce the relations between $\kappa(\mathbf{M})$, $\kappa(\mathbf{A})$ and $\kappa(\mathcal{A}_h)$. These relations are the following inequalities:

$$\lambda_{max}(\mathbf{A}) \le \lambda_{max}(\mathcal{A}_h)\lambda_{max}(\mathbf{M}), \qquad \lambda_{min}(\mathcal{A}_h)\lambda_{min}(\mathbf{M}) \le \lambda_{min}(\mathbf{A}). \tag{1.68}$$

The proof of the first inequality can be as follows

$$\begin{aligned} \lambda_{max}(\mathbf{A}) = \boldsymbol{u}_m^T \boldsymbol{A} \boldsymbol{u}_m = \boldsymbol{u}_m^T \boldsymbol{M} \boldsymbol{u}_m \frac{\boldsymbol{u}_m^T \boldsymbol{A} \boldsymbol{u}_m}{\boldsymbol{u}_m^T \boldsymbol{M} \boldsymbol{u}_m} \\ &\le \sup_{|\boldsymbol{u}|=1} (\boldsymbol{u}^T \boldsymbol{M} \boldsymbol{u}) \lambda_{max}(\mathcal{A}_h) \\ &\le \lambda_{max}(\mathbf{M}) \lambda_{max}(\mathcal{A}_h), \end{aligned} \tag{1.69}$$

where $\boldsymbol{u}_m$ is the eigenvector of $\boldsymbol{A}$ corresponding to λ_{max}. The other inequalities can be proven the same way.

Moreover we state auxillary lemma.

Lemma 10 (Friedrich-Poincare inequality)**.** *Let $M \subset \Omega$ be an open connected polyhedral domain such that M is an union of elements from $\mathcal{T}_h$. Let the diameter of M be H_M. Then for any function $v \in H^1(\Omega, \mathcal{T}_h)$ the following holds*

$$\|v\|_{0,M}^2 \leq CH_M^2 \left(\sum_{\substack{K\in\mathcal{T}_h\\K\subset M}} |v|_{1,K}^2 + \sum_{\substack{\gamma\in\mathcal{F}_h^I\\\gamma\subset M}} \|h_\gamma^{-\frac{1}{2}} [v] \|_{0,\gamma}^2 + \sum_{\substack{\gamma\in\mathcal{F}_h^I\\\gamma\subset\partial M}} \|h_\gamma^{-\frac{1}{2}} v\|_{0,\gamma}^2 \right), \tag{1.70}$$

where $C > 0$. Moreover if v has a zero average over M, then we have

$$\|v\|_{0,M}^2 \leq CH_M^2 \left(\sum_{\substack{K\in\mathcal{T}_h\\K\subset M}} |v|_{1,K}^2 + \sum_{\substack{\gamma\in\mathcal{F}_h^I\\\gamma\subset M}} \|h_\gamma^{-\frac{1}{2}} [v] \|_{0,\gamma}^2 \right) \tag{1.71}$$

The proof can be found in Antonietti and Houston [2011]. There is also the Broken Poincare inequality introduced in Dolejší and Feistauer [2015], that can be also used in the analysis. Now the hp-bounds of the condition number $\kappa(\mathcal{A}_h)$ are given in next theorem.

Theorem 11. *For any $v \in S_{hp}$, we get*

$$Ck_0 \sum_{K\in\mathcal{T}_h} \|v\|_{0,K}^2 \leq \mathcal{A}_h(v,v) \leq Ck_1 \sum_{K\in\mathcal{T}_h} \frac{p_K^4}{h_K^2} \|v\|_{0,K}^2 \tag{1.72}$$

Proof. For the lower bound we can use the Lemma 10 with $M = \Omega$ and the coercivity of the form $\mathcal{A}_h$:

$$\begin{aligned} k_0\|v\|_{0,\Omega}^2 &\leq k_0 CH_\Omega^2 \left(\sum_{K\in\mathcal{T}_h} |v|_{1,K}^2 + \sum_{\gamma\in\mathcal{F}_h^I} \|h_\gamma^{-\frac{1}{2}} [v] \|_{0,\gamma}^2 \right) \\ &\leq CH_\Omega^2 \left(\sum_{K\in\mathcal{T}_h} k_0 |v|_{1,K}^2 + \sum_{\gamma\in\mathcal{F}_h^I} \frac{1}{\alpha p_\gamma^2} \|\sigma^{\frac{1}{2}} [v] \|_{0,\gamma}^2 \right) \\ &\leq CH_\Omega^2 \max\left\{ 1, \frac{1}{\alpha \min_{\gamma\in\mathcal{F}_h^I} p_\gamma^2} \right\} \|\!|v|\!\|^2 \\ &\leq 2CH_\Omega^2 \mathcal{A}_h(v,v). \end{aligned} \tag{1.73}$$

For the upper bound of the first term in the DG-norm we can use the inverse inequality (1.31) and we get:

$$k_0 \sum_{K\in\mathcal{T}_h} \|\nabla v\|_{0,K}^2 = k_0 \sum_{K\in\mathcal{T}_h} |v|_{1,K}^2 \leq k_0 C_I \sum_{K\in\mathcal{T}_h} \frac{p_K^4}{h_K^2} \|v\|_{0,K}^2. \tag{1.74}$$

For the jump term $\mathrm{J}_h^\sigma(u,u)$ we use the definition of σ (1.28) and the inverse inequality (1.31):

$$\mathrm{J}_h^\sigma(u,u) = \sum_{\gamma\in\mathcal{F}_h^I} \|\sigma^{\frac{1}{2}} [v] \|_{0,\gamma}^2 = \sum_{\gamma\in\mathcal{F}_h^I} \frac{\alpha p_\gamma^2 k_0}{h_\gamma} \| [v] \|_{0,\gamma}^2 \leq k_0 C \sum_{K\in\mathcal{T}_h} \frac{p_K^4}{h_K^2} \|v\|_{0,K}^2 \tag{1.75}$$

Now if we use the continuity of the form $\mathcal{A}_h$, cf. (1.48). we get the result we needed. $\square$

Corollary 12. *The condition number of $\boldsymbol{A}$ can be estimated using the inequality* (1.68) *and previous theorem to get the result:*

$$\kappa(\boldsymbol{A}) \leq C \frac{k_1 \max_{K \in \mathcal{T}_h} p_K^4}{k_0 \min_{K \in \mathcal{T}_h} h_K^4} \kappa(\boldsymbol{M}). \tag{1.76}$$

Hence in order to bound the condition number of $\mathbf{A}$ it is sufficient to bound the condition number of $\mathbf{M}$, in other words bound the eigenvalues of $\mathbf{M}$. This problem depends on the choice of basis. We will suppose that the basis $\{\varphi_j\}_{j=0}^N$ is orthogonal, which is easy for DGM to construct in practice. Then we get the diagonal matrix, and then the following result from Quarteroni and Valli [2008] can be proven

Lemma 13. *Let $\{\varphi_i\}_{i=1}^N$ be the orthogonal basis of the space S_{hp}. Then for any $u \in S_{hp}$, let $\boldsymbol{u}$ be the vector of coefficients of u in the basis $\{\varphi_i\}_{i=1}^N$, then*

$$C \min_{K \in \mathcal{T}_h} h_K^d \boldsymbol{u}^T \boldsymbol{u} \leq \boldsymbol{u}^T \boldsymbol{M} \boldsymbol{u} \leq C \max_{K \in \mathcal{T}_h} h_K^d \boldsymbol{u}^T \boldsymbol{u}. \tag{1.77}$$

Hence we get the following corollary combining the two previous results.

Corollary 14. *For a set of orthogonal functions $\{\varphi_i\}_{i=1}^N$, which form the basis of the space S_{hp}, we have the following bound on the condition number $\kappa(\boldsymbol{A})$ of a system matrix $\boldsymbol{A}$*

$$\kappa(\boldsymbol{A}) \leq C \frac{k_1 \max_{K \in \mathcal{T}_h} p_K^4}{k_0 \min_{K \in \mathcal{T}_h} h_K^2} \frac{\max_{K \in \mathcal{T}_h} h_K^d}{\min_{K \in \mathcal{T}_h} h_K^d}. \tag{1.78}$$

Therefore, if the triangulation $\mathcal{T}_h$ has a globally quasi-uniform polynomial approximation, then

$$\kappa(\boldsymbol{A}) \leq C \frac{k_1}{k_0} p^4 h^{-2}. \tag{1.79}$$

Therefore, for $h \to 0$ and $p \to \infty$ we have that $\kappa(\mathcal{A}_h) \to \infty$. Hence the computation cost is increasing with the usage of finer mesh and higher polynomial approximation. This leads us to using the preconditions to increase the speed of the computation. As we already said, this is not the most reliable way, since in some cases this does not have to lead to increased speed of computation, but it is all we can do.

2. Domain decomposition method

Domain decomposition method is a method for solving partial differential equations by partitioning of the domain Ω into subdomains Ω_i and solving the problem there. We will be mostly interested in the non-overlapping Schwarz preconditioners. The preconditioning is important for solving large systems with a lot of unknowns where the traditional methods will be slow and inefficient. More details on the domain decomposition method can be found in Dolean et al. [2015] or Toselli and Widlund [2004].

2.1 Subdomain partitioning, Local and Coarse solvers

We consider an non-overlapping domain partitioning of Ω onto the finite set of open domains Ω_i such that

$$\overline{\Omega} = \bigcup_{i=1}^{N} \overline{\Omega_i}, \tag{2.1}$$

and

$$\Omega_i \cap \Omega_j = \emptyset \quad i,\, j = 1, \dots, N,\, i \neq j. \tag{2.2}$$

We assume that the domains Ω_i are unions of elements of triangulation $\mathcal{T}_h$. In addition, we consider the coarse partitioning of Ω. We will use the notation $\mathcal{T}_h$ for the fine partitioning and the $\mathcal{T}_H$ for the coarse partitioning. We will assume that these partitioning are *nested* i.e. the element of a coarse mesh is union of elements of the fine mesh. By Γ_{ij} we denote the set of all faces $\gamma \in \mathcal{F}_h^I$ such that $\gamma \subset \partial\Omega_i \cap \partial\Omega_j$, $i, j = 1, \dots, N$.

Next, we will introduce the local solvers in Ω_i and the coarse solver on the coarse mesh.

2.1.1 Local Solvers

The local solvers on Ω_i, $i = 1, \dots, N$ are defined using a space of piecewise polynomial functions. We will introduce the spaces S_{hp}^i, that are the restriction of space S_{hp} on the domains Ω_i. More precisely,

$$S_{hp}^i = \{u \in L^2(\Omega_i) : u|_K \in P_{p_K}\, \forall K \in \mathcal{T}_h\, K \subset \Omega_i\}. \tag{2.3}$$

The space is generally discontinuous, and the functions generally do not vanish on $\partial\Omega_i$. Then the bilinear from $\mathcal{A}_h(u, v)$ will reduce to $\mathcal{A}_h^i : S_{hp}^i \times S_{hp}^i \to \mathbb{R}$ using the *extension operators*

$$R_i^T \,:\, S_{hp}^i \to S_{hp} \qquad i = 1, \dots, N, \tag{2.4}$$

where the transpose of the restriction operator R_i is with respect to $L^2(\Omega_i)$ inner product. The extension operators give us the representation of the space S_{hp} as the sum of spaces S_{hp}^i,

$$S_{hp} = R_1^T S_{hp}^1 \oplus \cdots \oplus R_N^T S_{hp}^N. \tag{2.5}$$

The sum here is interpreted as a linear combination of extended function from S_{hp}^i by zero on $\Omega \setminus \Omega_i$. Now we can finally write the definition of the local solvers as

$$\mathcal{A}_h^i(u_i, v_i) := \mathcal{A}_h(R_i^T u_i, R_i^T v_i) \qquad \forall u_i, v_i \in S_{hp}^i,\ i = 1, \dots N. \tag{2.6}$$

2.1.2 Coarse solver

With the coarse solver we will first need to deal with the inconsistency of the polynomial degree. We will denote by q_D the polynomial degree of the coarse element $D \in \mathcal{T}_H$. We set q_D such that

$$0 \le q_D \le \min_{K \subset D} p_K. \tag{2.7}$$

Then we define the finite dimensional space for coarse solver as

$$S_{Hp}^0 := \{v \in L^2(\Omega) : v|_D \in P_{q_D}(D), D \in \mathcal{T}_H\} \tag{2.8}$$

We can define the the extension operator $R_0^T : S_{Hp}^0 \to S_{hp}$ as the classical injection of S_{Hp}^0 in S_{hp}. Then we can use this extension to define the coarse solver as

$$\mathcal{A}_h^0(u_0, v_0) := \mathcal{A}_h(R_0^T u_0, R_0^T v_0) \qquad \forall u_0, v_0 \in S_{Hp}^0. \tag{2.9}$$

The coarse solver is used to speed up the transition of information among the subdomains. In general, if we use a higher polynomial degree, then more information will travel, but the computation will be more costly. In general, it is chosen to balance the cost and efficiency of the computation. For our analysis, we will assume that the coarse solver is piecewise constant approximation.

2.1.3 Variational formulation

Now we will define the *local projection operators* $\tilde{P}_i$, which we use for projection of u to the space S_{hp}^i, $i = 1, \dots, N$. The operators are defined as

$$\tilde{P}_i : S_{hp} \to S_{hp}^i, \qquad \mathcal{A}_i(\tilde{P}_i u, v_i) = \mathcal{A}_h(u, R_i^T v_i), \quad \forall v_i \in S_{hp}^i. \tag{2.10}$$

Moreover, we define the projection operators on the space S_{hp}

$$P_i := R_i^T \tilde{P}_i : S_{hp} \to S_{hp} \quad \forall i = 1, \dots, N. \tag{2.11}$$

The same definition of projection operators can be formulated for the coarse space S_{Hp}^0, i. e.,

$$\tilde{P}_0 : S_{hp} \to S_{Hp}^0, \qquad \mathcal{A}_0(\tilde{P}_0 u, v_0) := \mathcal{A}_h(u, R_0^T v_0), \quad \forall v_0 \in S_{Hp}^0. \tag{2.12}$$

and

$$P_0 := R_0^T \tilde{P}_0 : S_{hp} \to S_{hp} \tag{2.13}$$

Then we can formulate the operator of additive Schwarz method by

$$P_{ad} := \sum_{i=0}^{N} P_i. \tag{2.14}$$

2.2 Convergence analysis

The convergence analysis is based on the general framework of Toselli and Widlund [2004]. Three assumptions on the local solvers are needed for the formal analysis. We will follow the analysis done in Antonietti and Houston [2011] , but the proof of the auxillary lemma is new and the proof of the last theorem is a modification of the proof done in the article.

Assumption 1 (Stable decomposition) Let $C_0 > 0$ be a constant such that for every $u \in S_{hp}$ we have the decomposition

$$u = \sum_{i=0}^{N} R_i^T u_i, \tag{2.15}$$

with $u_0 \in S_{Hp}^0$, $u_i \in S_{hp}^i$, $i = 1, \ldots, N$, such that

$$\sum_{i=0}^{N} \mathcal{A}_h^i(u_i, u_i) \leq C_0^2 \mathcal{A}_h(u, u). \tag{2.16}$$

Assumption 2 (Local stability) There exist a constant $0 \leq \omega \leq 2$, such that

$$\begin{aligned} \mathcal{A}_h(R_i^T u_i, R_i^T u_i) &\leq \omega \mathcal{A}_h^i(u_i, u_i) \quad \forall u_i \in S_{hp}^i,\ i = 1, \ldots, N, \\ \mathcal{A}_h(R_0^T u_0, R_0^T u_0) &\leq \omega \mathcal{A}_h^0(u_0, u_0) \quad \forall u_0 \in S_{Hp}^0. \end{aligned} \tag{2.17}$$

Assumption 3 (Strengthened Cauchy-Schwarz inequalities) There exist constant $0 \leq \epsilon_{ij} \leq 1$, $i, j = 1, \ldots, N$, such that

$$|\mathcal{A}_h(R_i^T u_i, R_j^T u_j)| \leq \epsilon_{ij} \mathcal{A}_h(R_i^T u_i, R_i^T u_i)^{\frac{1}{2}} \mathcal{A}_h(R_j^T u_j, R_j^T u_j)^{\frac{1}{2}}, \quad i, j = 1, \ldots N, \tag{2.18}$$

for all $u_i \in S_{hp}^i$, $u_j \in S_{hp}^j$. By $\rho(\boldsymbol{\varepsilon})$ we denote the spectral radius of $\boldsymbol{\varepsilon} = \{\epsilon_{ij}\}_{i,j=0}^N$

The spectral bounds for the Schwarz operator, follows from Assumptions 1, 2 and 3. The result is taken from Toselli and Widlund [2004] is formulated the following theorem.

Theorem 15. *Let the Assumptions 1 - 3 be satisfied. Then the condition number of the additive Schwarz operator can be bounded as*

$$\kappa(P_{ad}) \leq C_0^2\, \omega\, (\rho(\boldsymbol{\varepsilon}) + 1). \tag{2.19}$$

Proof. The proof can be found in Toselli and Widlund [2004]. □

The aim of the remaining part of the section is to show that for the operators defined in Section 2.1 satisfies Assumptions 1 - 3. Then we obtain the bounds of condition number of the preconditioned system.

First we can see that Assumption 2 is trivially satisfied, from the definition of the local projections and local solvers. In fact, it is an identity with $\omega = 1$.

Moreover, for the assumption 3, it can be seen that $\epsilon_{ii} = 1$ for $i = 1, \cdots, N$. Since the form $\mathcal{A}_h$ defines a symmetric and positive definite matrix, the Cauchy-Schwarz inequality implies

$$\mathcal{A}_h(u, v) \leq \mathcal{A}_h(u, u)^{\frac{1}{2}} \mathcal{A}_h(v, v)^{\frac{1}{2}}, \quad \forall u, v \in H_0^1(\Omega, \mathcal{T}_h). \tag{2.20}$$

Moreover, if $\partial\Omega_i \cap \partial\Omega_j = \emptyset$, then

$$\mathcal{A}_h(R_i^T u_i, R_j^T u_j) = 0. \tag{2.21}$$

Therefore, we have $\epsilon_{ij} = 1$, when $\partial\Omega_i \cap \partial\Omega_j \neq \emptyset$ and $\epsilon_{ij} = 0$ otherwise. Then we can bound the spectral radius $\rho(\boldsymbol{\varepsilon})$ as

$$\rho(\boldsymbol{\varepsilon}) \leq \max_{i=1,\dots,N} \sum_{j=1}^{N} |\epsilon_{ij}| \leq 1 + N_C, \tag{2.22}$$

where N_C is the maximum number of the adjacent subdomains that a given subdomain can have and is typically independent of h.

Before verifying Assumption 1, we prove the following result. Obviously any $u \in S_{hp}$ can be uniquely represented as

$$u = \sum_{i=1}^{N} R_i^T u_i, \qquad u_i \in S_{hp}^i, \quad i = 1, \dots, N. \tag{2.23}$$

Hence the following identity holds

$$\mathcal{A}_h(u, u) = \sum_{i=1}^{N} \mathcal{A}_h^i(u_i, u_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \mathcal{A}_h(R_i^T u_i, R_j^T u_j). \tag{2.24}$$

The following result deals with the upper bound of the second term of the identity (2.24).

Lemma 16. *Let $u \in S_{hp}$, we have the following upper bound*

$$\left| \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \mathcal{A}_h(R_i^T u_i, R_j^T u_j) \right| \leq C \frac{k_1}{k_0} \left(|\!|\!| u |\!|\!|^2 + \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \sum_{\gamma \in \Gamma_{ij}} \left(\|\sigma^{\frac{1}{2}} u_i\|_{0,\gamma}^2 + \|\sigma^{\frac{1}{2}} u_j\|_{0,\gamma}^2 \right) \right), \tag{2.25}$$

where $u_i \in S_{hp}^i$, $i = 1, \dots, N$ are given by (2.23) *and $C > 0$ independent of h, H, p_K, k_0, k_1.*

Proof. Let $u \in S_{hp}$. From the definition of the form $\mathcal{A}_h$ we have that for $\partial\Omega_i \cap \partial\Omega_j = \emptyset$, $\mathcal{A}_h(R_i^T u_i, R_j^T u_j) = 0$. Also, we have the triangle inequality for the sum

$$\left| \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \mathcal{A}_h(R_i^T u_i, R_j^T u_j) \right| \leq \sum_{\substack{i,j=1 \\ i \neq j}}^{N} |\mathcal{A}_h(R_i^T u_i, R_j^T u_j)|. \tag{2.26}$$

For simplicity we will use the notation $\tilde{u}_i$ and $\tilde{u}_j$ for $R_i^T u_i$ and $R_j^T u_j$ respectively. Let Ω_i and Ω_j be neighbouring subdomains. We have

$$\mathcal{A}_h(\tilde{u}_i, \tilde{u}_j) = a_h(\tilde{u}_i, \tilde{u}_j) + \mathrm{J}_h^{\sigma}(\tilde{u}_i, \tilde{u}_j). \tag{2.27}$$

We bound every term on the right-hand side of (2.27). First, we deal with the form

$$a_h(\tilde{u}_i,\tilde{u}_j) = \sum_{K\in\mathcal{T}_h}\int_K \boldsymbol{K}\nabla\tilde{u}_i\cdot\nabla\tilde{u}_j \ \mathrm{dx} - \sum_{\gamma\in\mathcal{F}_h^I}\int_\gamma (\langle\boldsymbol{K}\nabla\tilde{u}_i\rangle\cdot[\tilde{u}_j] + \langle\boldsymbol{K}\nabla\tilde{u}_j\rangle\cdot[\tilde{u}_i]) \ \mathrm{dS}. \tag{2.28}$$

The first term on the right is easy to deal with, since $\int_K \boldsymbol{K}\nabla\tilde{u}_i\cdot\nabla\tilde{u}_j \ \mathrm{dx} = 0$. This is due to the fact that the functions $\tilde{u}_i$ and $\tilde{u}_j$ are zero on all subdomains except on Ω_i and Ω_j respectively. For the next bound, we need to use the notation K_γ^i for the element $K\in\mathcal{T}_h$ such that $\gamma\subset\partial K$ and $K\subset\Omega_i$. Using the Cauchy-Schwarz inequality, the multiplicative trace inequality (1.33) and the bound (1.35) and the assumption (1.8), we estimate the middle term in (2.28) as

$$\begin{aligned}
\sum_{\gamma\in\mathcal{F}_h^I}\int_\gamma \langle\boldsymbol{K}\nabla\tilde{u}_i\rangle\cdot[\tilde{u}_j] \ \mathrm{dS} &\le Ck_1\sum_{\gamma\in\Gamma_{ij}}\int_\gamma\langle\nabla\tilde{u}_i\rangle\cdot[\tilde{u}_j] \ \mathrm{dS}\\
&\le C\frac{k_1}{k_0}\sum_{\gamma\in\Gamma_{ij}}\int_\gamma k_0\langle\nabla\tilde{u}_i\rangle\cdot[\tilde{u}_j] \ \mathrm{dS}\\
&\le C\frac{k_1}{k_0}\sum_{\gamma\in\Gamma_{ij}} k_0^{\frac{1}{2}}\|\langle\nabla\tilde{u}_i\rangle\|_{0,\gamma}k_0^{\frac{1}{2}}\|[\tilde{u}_j]\|_{0,\gamma}\\
&\le C\frac{k_1}{k_0}\sum_{\gamma\in\Gamma_{ij}} k_0^{\frac{1}{2}}\|\nabla\tilde{u}_i\|_{0,\partial K_\gamma^i}k_0^{\frac{1}{2}}\|[\tilde{u}_j]\|_{0,\gamma}\\
&\le C\frac{k_1}{k_0\sqrt{\alpha}}\sum_{\gamma\in\Gamma_{ij}} k_0^{\frac{1}{2}}\|\nabla\tilde{u}_i\|_{0,K_\gamma^i}\left(k_0\frac{p_\gamma^2\alpha}{h_\gamma}\right)^{\frac{1}{2}}\|[\tilde{u}_j]\|_{0,\gamma}\\
&\le C\frac{k_1}{k_0\sqrt{\alpha}}\left(\sum_{\substack{K\in\mathcal{T}_h\\ K\subset\Omega_i}} k_0\|\nabla\tilde{u}_i\|_{0,K}^2\right)^{\frac{1}{2}}\left(\sum_{\substack{\gamma\in\mathcal{F}_h^I\\ \gamma\subset\bar{\Omega}_j}}\|\sigma^{\frac{1}{2}}[\tilde{u}_j]\|_{0,\gamma}^2\right)^{\frac{1}{2}}\\
&\le C\frac{k_1}{k_0\sqrt{\alpha}}\left(\sum_{\substack{K\in\mathcal{T}_h\\ K\subset\Omega_i}} k_0\|\nabla\tilde{u}_i\|_{0,K}^2 + \sum_{\substack{\gamma\in\mathcal{F}_h^I\\ \gamma\subset\bar{\Omega}_j}}\|\sigma^{\frac{1}{2}}[\tilde{u}_j]\|_{0,\gamma}^2\right).
\end{aligned} \tag{2.29}$$

The same can be done for the third term, and we get

$$\sum_{\gamma\in\mathcal{F}_h^I}\int_\gamma\langle\boldsymbol{K}\nabla\tilde{u}_j\rangle\cdot[\tilde{u}_i] \ \mathrm{dS} \le C\frac{k_1}{k_0\sqrt{\alpha}}\left(\sum_{\substack{K\in\mathcal{T}_h\\ K\subset\Omega_j}} k_0\|\nabla\tilde{u}_j\|_{0,K}^2 + \sum_{\substack{\gamma\in\mathcal{F}_h^I\\ \gamma\subset\bar{\Omega}_i}}\|\sigma^{\frac{1}{2}}[\tilde{u}_i]\|_{0,\gamma}^2\right). \tag{2.30}$$

Now for the term J_h^σ we use the Cauchy-Schwarz inequality and we get

$$\begin{aligned}
\mathrm{J}_h^\sigma(\tilde{u}_i,\tilde{u}_j) &\le C\sum_{\gamma\in\Gamma_{ij}}\|\sigma^{\frac{1}{2}}\tilde{u}_i\|_{0,\gamma}\|\sigma^{\frac{1}{2}}\tilde{u}_j\|_{0,\gamma}\\
&\le C\left(\sum_{\gamma\in\Gamma_{ij}}\|\sigma^{\frac{1}{2}}\tilde{u}_i\|_{0,\gamma}^2 + \sum_{\gamma\in\Gamma_{ij}}\|\sigma^{\frac{1}{2}}\tilde{u}_j\|_{0,\gamma}^2\right).
\end{aligned} \tag{2.31}$$

Now adding the previous formulas together, we get

$$\begin{aligned}\mathcal{A}_h(\tilde{u}_i,\tilde{u}_j) \leq C\frac{k_1}{k_0}\Bigg(&\sum_{\substack{K\in\mathcal{T}_h\\K\subset\Omega_i}} k_0\|\nabla\tilde{u}_i\|^2_{0,K} + \sum_{\substack{\gamma\in\mathcal{F}_h^I\\\gamma\subset\bar{\Omega}_j}}\|\sigma^{\frac{1}{2}}[\tilde{u}_j]\|^2_{0,\gamma}+\\
&+\sum_{\substack{K\in\mathcal{T}_h\\K\subset\Omega_j}} k_0\|\nabla\tilde{u}_j\|^2_{0,K} + \sum_{\substack{\gamma\in\mathcal{F}_h^I\\\gamma\subset\bar{\Omega}_i}}\|\sigma^{\frac{1}{2}}[\tilde{u}_i]\|^2_{0,\gamma}\\
&+\sum_{\gamma\in\Gamma_{ij}}\|\sigma^{\frac{1}{2}}\tilde{u}_i\|^2_{0,\gamma}+\sum_{\gamma\in\Gamma_{ij}}\|\sigma^{\frac{1}{2}}\tilde{u}_j\|^2_{0,\gamma}\Bigg)\end{aligned}\tag{2.32}$$

Finally, using the fact that $u|_{\Omega_i} = \tilde{u}_i$ and on each $\gamma \in \Gamma_{ij}$, $[\tilde{u}_i] = u_i\boldsymbol{n}_i$ we get

$$\begin{aligned}\mathcal{A}_h(\tilde{u}_i,\tilde{u}_j) \leq C\frac{k_1}{k_0}\Bigg(&\sum_{\substack{K\in\mathcal{T}_h\\K\subset\Omega_i\cup\Omega_j}} k_0\|\nabla u\|^2_{0,K} + \sum_{\substack{\gamma\in\mathcal{F}_h^I\\\gamma\subset\bar{\Omega}_j\cup\bar{\Omega}_i\\\gamma\not\subset\Gamma_{ij}}}\|\sigma^{\frac{1}{2}}[u]\|^2_{0,\gamma}\\
&+\Bigg(\sum_{\gamma\in\Gamma_{ij}}\|\sigma^{\frac{1}{2}}u_i\|^2_{0,\gamma}+\|\sigma^{\frac{1}{2}}u_j\|^2_{0,\gamma}\Bigg)\Bigg)\end{aligned}\tag{2.33}$$

Hence, summing over all subdomains Ω_i we get the result. □

The key point in the analysis of the domain decomposition preconditioners is the *stable splitting*, that can be found for a family of subspaces and the corresponding bilinear form.

Theorem 17 (Stable decomposition)**.** *For any $u \in S_{hp}$, there exists a decomposition of the form $u = \sum_{i=0}^N R_i^T u_i$, with $u_0 \in S_{Hp}^0$ and $u_i \in S_{hp}^i$, $i = 1,\dots,N$, such that*

$$\sum_{i=0}^N \mathcal{A}_i(u_i,u_i) \leq CC_0^2\mathcal{A}_h(u,u), \quad C_0^2 = \alpha\frac{k_1}{k_0}\max_{D\in\mathcal{T}_H} H_D\frac{\max_{\substack{K\in\mathcal{T}_h\\K\subset D}} p_K^2}{\min_{\substack{K\in\mathcal{T}_h\\K\subset D}} h_K},\tag{2.34}$$

where C is independent of h, H, p_K, k_0, k_1.

Proof. Given $u \in S_{hp}$, let $u_0 \in S_{hp}^0$ be defined by

$$u_0|_D := \frac{1}{|D|}\int_D u\mathrm{dx}, \qquad \forall D \in \mathcal{T}_H.\tag{2.35}$$

Next, we decompose uniquely $u - R_0^T u_0 = \sum_{i=1}^N R_i^T u_i$. From the unique decomposition and from identity (2.24), we have

$$\mathcal{A}_h(u - R_0^Tu_0, u - R_0^Tu_0) = \sum_{i=1}^N \mathcal{A}_h(R_i^Tu_i, R_i^Tu_i) + \sum_{\substack{i,j=1\\i\neq j}}^N \mathcal{A}_h(R_i^Tu_i, R_j^Tu_j).\tag{2.36}$$

Now if we add to both sides $\mathcal{A}_0(u_0,u_0) = \mathcal{A}_h(R_0^Tu_0, R_0^Tu_0)$, putting the second term on the right-hand side on the other side, taking the absolute value of the equality and using the triangle inequality we get

$$\left|\sum_{i=0}^{N} \mathcal{A}_h(R_i^T u_i, R_i^T u_i)\right| \leq C \Bigg(|\mathcal{A}_h(u - R_0^T u_0, u - R_0^T u_0)| + |\mathcal{A}_h(R_0^T u_0, R_0^T u_0)| + \Bigg|\sum_{\substack{i,j=1\\ i\neq j}}^{N} \mathcal{A}_h(R_i^T u_i, R_j^T u_j)\Bigg| \Bigg). \tag{2.37}$$

Now will will try to bound every term on the right-hand side. For the first term, we have for $C > 0$ using (2.20) and Young inequality

$$|\mathcal{A}_h(u - R_0^T u_0, u - R_0^T u_0)| \leq C|\mathcal{A}_h(u, u)| + |\mathcal{A}_h(R_0^T u_0, R_0^T u_0)|. \tag{2.38}$$

We then get

$$\left|\sum_{i=0}^{N} \mathcal{A}_h(R_i^T u_i, R_i^T u_i)\right| \leq C \Bigg(|\mathcal{A}_h(u, u)| + |\mathcal{A}_h(R_0^T u_0, R_0^T u_0)| + \Bigg|\sum_{\substack{i,j=1\\ i\neq j}}^{N} \mathcal{A}_h(R_i^T u_i, R_j^T u_j)\Bigg| \Bigg). \tag{2.39}$$

First term in (2.39) is estimated by the continuity of the form $\mathcal{A}_h$ (1.48). Now we focus on the bounding of the second term in (2.39). We show the following bound

$$|\mathcal{A}_h(R_0^T u_0, R_0^T u_0)| \leq C \left(\|\!|u|\!\|^2 + \sum_{D\in\mathcal{T}_H} \eta_D \|u - R_0^T u_0\|_{0,\partial D}^2 \right), \tag{2.40}$$

where constant η_D is defined by

$$\eta_D := \alpha \frac{\max_{\substack{K\in\mathcal{T}_h\\ K\subset D}} p_K^2}{\min_{\substack{K\in\mathcal{T}_h\\ K\subset D}} h_K}. \tag{2.41}$$

Now, for the bound, we use the continuity of the form $\mathcal{A}_h$, the we add and subtract u and use the triangle inequality. Due to (2.35), u_0 is a piecewise constant, so we have

$$\nabla R_0^T u_0 = 0 \qquad \forall K \in \mathcal{T}_h. \tag{2.42}$$

Using (1.25), triangle inequality and the definition of the norm (1.30), we get

$$\begin{aligned} |\mathcal{A}_h(R_0^T u_0, R_0^T u_0)| &\leq C\frac{k_1}{k_0} \sum_{\gamma\in\mathcal{F}_h^I} \|\sigma^{\frac{1}{2}} \left[R_0^T u_0\right] \|_{0,\gamma}^2 \\ &\leq C\frac{k_1}{k_0} \left(\sum_{\gamma\in\mathcal{F}_h^I} \|\sigma^{\frac{1}{2}} \left[u - R_0^T u_0\right] \|_{0,\gamma}^2 + \|\!|u|\!\|^2 \right) \end{aligned} \tag{2.43}$$

We estimate the first term on the right-hand side. We use the fact that the jump of u_0 is zero on the edges in D. Hence, we get

$$\begin{aligned}\sum_{\gamma\in\mathcal{F}_h^I}\|\sigma^{\frac{1}{2}}\left[u-R_0^Tu_0\right]\|_{0,\gamma}^2 &= \sum_{D\in\mathcal{T}_H}\sum_{\substack{\gamma\in\mathcal{F}_h^I\\ \gamma\in D}}\|\sigma^{\frac{1}{2}}\left[u\right]\|_{0,\gamma}^2+\sum_{D\in\mathcal{T}_H}\sum_{\substack{\gamma\in\mathcal{F}_h^I\\ \gamma\in\partial D}}\|\sigma^{\frac{1}{2}}\left[u-R_0^Tu_0\right]\|_{0,\gamma}^2\\ &\le |||u|||^2+\sum_{D\in\mathcal{T}_H}\sum_{\substack{\gamma\in\mathcal{F}_h^I\\ \gamma\in\partial D}}\|\sigma^{\frac{1}{2}}\left[u-R_0^Tu_0\right]\|_{0,\gamma}^2\\ &\le |||u|||^2+k_0\sum_{D\in\mathcal{T}_H}\eta_D\|\left[u-R_0^Tu_0\right]\|_{0,\partial D}^2.\end{aligned}\tag{2.44}$$

If we add together (2.43) and (2.44) we get the bound (2.40).

Now we will use Lemma 16 for the bound of the third term in (2.39). Since each Ω_i is a union of elements of $\mathcal{T}_H$ we get that

$$\sum_{\substack{i,j=1\\ i\neq j}}^{N}\sum_{\gamma\in\Gamma_{ij}}\left(\|\sigma^{\frac{1}{2}}u_i\|_{0,\gamma}^2+\|\sigma^{\frac{1}{2}}u_j\|_{0,\gamma}^2\right)\le Ck_0\sum_{D\in\mathcal{T}_H}\eta_D\|u-R_0^Tu_0\|_{\partial D}^2.\tag{2.45}$$

We note that u_i, $i=1,\dots,N$ in (2.45) come from the unique decomposition $u-R_0^Tu_0=\sum_{i=1}^{N}R_i^Tu_i$. Using Lemma 16, we get

$$\left|\sum_{\substack{i,j=1\\ i\neq j}}^{N}\mathcal{A}_h(R_i^Tu,R_j^Tu)\right|\le C\frac{k_1}{k_0}\left(|||u-R_0^Tu_0|||^2+k_0\sum_{D\in\mathcal{T}_H}\eta_D\|u-R_0^Tu_0\|_{\partial D}^2\right).\tag{2.46}$$

We simplify equation (2.39). We will use all these bounds, which we just did now, and we will use the triangle inequality and the continuity of the form $\mathcal{A}_h$ and we get

$$\begin{aligned}\left|\sum_{i=0}^{N}\mathcal{A}_h(R_i^Tu_i,R_i^Tu_i)\right| &\le C\frac{k_1}{k_0}\left(|||u|||^2+|||u-R_0^Tu_0|||^2+k_0\sum_{D\in\mathcal{T}_H}\eta_D\|u-R_0^Tu_0\|_{\partial D}^2\right)\\ &\le C\frac{k_1}{k_0}\left(|||u|||^2+k_0\sum_{D\in\mathcal{T}_H}\eta_D\|u-R_0^Tu_0\|_{\partial D}^2\right).\end{aligned}\tag{2.47}$$

Now we will need the trace inequality shown in [Feng and Karashian, 2002, Lemma 3.1], which is for $v\in H^1(\Omega,\mathcal{T}_h)$ and $D\in\mathcal{T}_H$ as follows:

$$\|v\|_{0,\partial D}^2\le C\left[H_D^{-1}\|v\|_{0,D}^2+H_D\left(\sum_{\substack{K\in\mathcal{T}_h\\ K\subset D}}|v|_{1,K}^2+\sum_{\substack{\gamma\in\mathcal{F}_h^I\\ \gamma\subset D}}h_\gamma^{-1}\|v\|_{0,\gamma}^2\right)\right]\tag{2.48}$$

Now we can use this inequality for the last term in our bound and we also use the fact that $R_0^Tu_0$ is a constant on D and we get

$$\|u-R_0^Tu_0\|_{0,\partial D}^2\le C\left[H_D^{-1}\|u-R_0^Tu_0\|_{0,D}^2+H_D\left(\sum_{\substack{K\in\mathcal{T}_h\\ K\subset D}}|u|_{1,K}^2+\sum_{\substack{\gamma\in\mathcal{F}_h^I\\ \gamma\subset D}}h_\gamma^{-1}\|u\|_{0,\gamma}^2\right)\right].\tag{2.49}$$

Hence, using (2.47) and (2.49), we get

$$\left|\sum_{i=0}^{N} \mathcal{A}_h(R_i^T u_i, R_i^T u_i)\right| \leq C\frac{k_1}{k_0}\left(\|\!|u|\!\|^2 + k_0 \sum_{D\in\mathcal{T}_H} \eta_D \|u - R_0^T u_0\|^2_{\partial D}\right)$$

$$\leq C\frac{k_1}{k_0}\Bigg\{ \|\!|u|\!\|^2 +$$

$$+ k_0 \sum_{D\in\mathcal{T}_H} \eta_D \left[H_D^{-1}\|u - R_0^T u_0\|^2_{0,D} + H_D\Big(\sum_{\substack{K\in\mathcal{T}_h\\K\subset D}} |u|^2_{1,K} + \sum_{\substack{\gamma\in\mathcal{F}_h^I\\ \gamma\subset D}} h_\gamma^{-1}\|u\|^2_{0,\gamma}\Big)\right]\Bigg\}. \tag{2.50}$$

Now, using the Friedrich-Poincare inequality (1.71) on the term $\|u - R_0^T u_0\|^2_{0,D}$, we get

$$\left|\sum_{i=0}^{N} \mathcal{A}_h(R_i^T u_i, R_i^T u_i)\right| \leq C\frac{k_1}{k_0}\left(\|\!|u|\!\|^2 + \sum_{D\in\mathcal{T}_H} C_D \left(\sum_{\substack{K\in\mathcal{T}_h\\K\subset D}} |u|^2_{1,K} + \sum_{\substack{\gamma\in\mathcal{F}_h^I\\ \gamma\subset D}} h_\gamma^{-1}\|u\|^2_{0,\gamma}\right)\right), \tag{2.51}$$

where $C_D = k_0\eta_D H_D$. Now using the coercivity (1.50) of $\mathcal{A}_h$, we get the desired result. □

If we now use assumptions (1.8) and (1.16), we get the estimate

$$\sum_{i=0}^{N} \mathcal{A}_i(u_i, u_i) \leq C\alpha\frac{p^2 H k_1}{h k_0}\mathcal{A}_h(u, u). \tag{2.52}$$

Finally, Theorem 15 gives

$$\kappa(P_{ad}) \leq C\alpha\frac{p^2 H k_1}{h k_0}. \tag{2.53}$$

This shows us the dependence of the condition number on the coarse mesh size and its elements. However, we do not see the role of the coarse mesh polynomial degree. This aspect was analysed in Antonietti et al. [2016].

2.3 Implementation

In this section, we briefly describe the implementation of the two-level additive Schwarz method defined in Section 2.1 - 2.2. The implementation is inspired by the method done in Antonietti et al. [2014]. Let $n \in \mathbb{N}$ be the dimension of the space S_{hp}. The DG discretization (1.29) is equivalent to the following linear algebraic problem. Find $\boldsymbol{u} \in \mathbb{R}^n$ such that

$$\boldsymbol{A}\boldsymbol{u} = \boldsymbol{F}, \tag{2.54}$$

where $\boldsymbol{F} \in \mathbb{R}^n$ is the vector of right hand side and matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is representation of the bilinear form $\mathcal{A}_h$ in a suitable basis. Now we use the matrix representation $\boldsymbol{R}_i^T$ $i = 0, \dots, N$ of operator (2.4) and $\tilde{\boldsymbol{P}}_i \, i = 0, \dots, N$ of an operator (2.11), for the restriction and prolongation operators, respectively. Using the notation $n_i = \dim(S^i_{hp})$ we know that $\boldsymbol{R}_i^T \in \mathbb{R}^{n \times n_i}$, $i = 1, \dots, N$ and from the construction of the operators it is clear that $\tilde{\boldsymbol{P}}_i = (\boldsymbol{R}_i^T)^T$.

Now we can see that the local billinear form $\boldsymbol{\mathcal{A}}_i$ will have

$$\boldsymbol{A}_i = \boldsymbol{R}_i \boldsymbol{A} \boldsymbol{R}_i^T \ \in \mathbb{R}^{n_i \times n_i} \tag{2.55}$$

Now if we want to write the projection operators P_i in the matrix representation we have for $i = 0, \dots n$ that

$$\boldsymbol{P}_i = \boldsymbol{R}_i^T \boldsymbol{A}_i^{-1} \boldsymbol{R}_i \boldsymbol{A}. \tag{2.56}$$

Then we can rewrite the operator P_{AD} as follows

$$\boldsymbol{P}_{AD} = \sum_{i=0}^{N} \boldsymbol{P}_i = \sum_{i=0}^{N} \boldsymbol{R}_i^T \boldsymbol{A}_i^{-1} \boldsymbol{R}_i \boldsymbol{A} = \boldsymbol{M}_{AD}^{-1} \boldsymbol{A} \tag{2.57}$$

Hence in computation we will use the matrix $\boldsymbol{M}_{AD}$ as a preconditioner. The computation of the system (2.54) will be done using the preconditioned GMRES. The algorithm we present is taken from Golub and Van Loan [2013].

Preconditioned m-step GMRES If $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{M} \in \mathbb{R}^{n \times n}$ are nonsingular, $\boldsymbol{b} \in \mathbb{R}^n$ and $\boldsymbol{x}_0 \in \mathbb{R}^n$ is an initial vector and m is an positive integer setting the maximum number of steps. Then we obtain the approximate solution of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ using the algorithm:

Algorithm 1 Preconditioned m-step GMRES

```
$k = 0$, $\boldsymbol{r}_0 = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_0$, Solve $\boldsymbol{M}\boldsymbol{z}_0 = \boldsymbol{r}_0$, $\beta_0 = \|\boldsymbol{z}_0\|_2$
while $\beta_k > 0$ and $k < m$ do
    $\boldsymbol{q}_{k+1} = \frac{1}{\beta_k}\boldsymbol{z}_k$
    $k = k + 1$
    Solve $\boldsymbol{M}\boldsymbol{z}_k = \boldsymbol{A}\boldsymbol{q}_k$
    for $i = 1 : k$ do
        $h_{ik} = \boldsymbol{q}_i^T \boldsymbol{z}_k$
        $\boldsymbol{z}_k = \boldsymbol{z}_k - h_{ik}\boldsymbol{q}_i$
    end for
    $\beta_k = \|\boldsymbol{z}_k\|_2$, $h_{k+1k} = \beta_k$
    Apply $\boldsymbol{G}_1, \dots, \boldsymbol{G}_{k-1}$ matrices of Givens rotation to $\boldsymbol{H}_{k+1,k} = \{h_{ij}\}_{i=1,\dots,k+1}^{j=1,\dots,k}$,
and determine $\boldsymbol{G}_k$, $\boldsymbol{R}_k$, $\boldsymbol{p}_k$ and $\boldsymbol{\rho}_k = \|\boldsymbol{M}^{-1}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_k)\|_2$.
end while
Solve $\boldsymbol{R}_k\boldsymbol{y}_k = \boldsymbol{p}_k$ and set $\tilde{\boldsymbol{x}} = \boldsymbol{x}_0 + \boldsymbol{Q}_k\boldsymbol{y}_k$, where $\boldsymbol{Q}_k = [\boldsymbol{q}_1, \dots, \boldsymbol{q}_k]$.
```

In our implementation the Solve step can be done using direct solvers, since we have a special structure of the matrix $\boldsymbol{M}_{AD}$. The action of the matrix $\boldsymbol{M}_{AD}$ on some vector $\boldsymbol{z}$ is described in Algorithm 1 in Antonietti et al. [2014]. For the completeness of the implementation method we will also mention it.

We build the coarse mesh using the partition of the elements Ω_i into k macroelements, that are union of elements of the fine mesh. Subdomains Ω_i

Algorithm 2 Action of the precondition matrix $\boldsymbol{M}_{AD}$

```
set $\boldsymbol{z} = 0$
**for** i = 1:N **do**
    $\boldsymbol{x}_i = \boldsymbol{R}_i\boldsymbol{z}$
    Solve $\boldsymbol{A}_i\boldsymbol{u}_i = \boldsymbol{x}_i$
    $\boldsymbol{z}_i = \boldsymbol{R}_i^T\boldsymbol{u}_i$
    $\boldsymbol{z} = \boldsymbol{z} + \boldsymbol{z}_i$
**end for**
$\boldsymbol{x}_0 = \boldsymbol{R}_0\boldsymbol{z}$
Solve $\boldsymbol{A}_0\boldsymbol{u}_0 = \boldsymbol{x}_0$
$\boldsymbol{z}_0 = \boldsymbol{R}_0^T\boldsymbol{u}_0$
$\boldsymbol{z} = \boldsymbol{z} + \boldsymbol{z}_0$
```

are generated from the fine mesh using the library METIS. For the direct solvers used in the preconditioning we use the Multifrontal Massively Parallel Solver (MUMPS).

From the implementation may arise numerical errors that can have an impact on our results. We state few numerical difficulties, that can lead to errors.

- The assembling of matrix A in the code is done by multiplication by a canoniacal vector.

- The performance of the preconditioning can also be problematic. We use MUMPS to compute the solution using $\boldsymbol{A}_i^{-1}$, $i = 0, \ldots, N$.

- The partition of the triangulation into subdomains Ω_i, $i = 1, \ldots, N$ can also have an impact.

- The precision of the function used in MATLAB can also lead to numerical errors.

3. Numerical experiments

We present several numerical experiments supporting the theoretical results from Chapter 2. The first problem that we discuss is the Laplace equation with homogeneous Dirichlet boundary condition. It is the simplest version of problem (1.1). In the following example we will use more nontrivial problem dealing with the magnetic field in the alternator.

3.1 Laplace equation

We consider the problem

$$-\Delta u = f \quad \text{in } \Omega \tag{3.1}$$

$$u = 0 \quad \text{on } \partial\Omega, \tag{3.2}$$

where $\Omega = (0,1)^2$. Since $k_0 = k_1 = 1$ all results depend only on h, H and p. We use the ADGFEM code (Dolejší [2020]) to generate the problem matrix (3.1). The ADGFEM code assembles the preconditioned matrices which are exported to MATLAB which estimate the condition number using `codest` function. The assembling by ADGFEM is carried out by multiplying the generated matrix by a canonical vector. The function `codest` is the recommended function for estimation of the condition number of large sparse matrices. We investigate the dependence of the condition number $\kappa(\boldsymbol{A})$ and $\kappa(\boldsymbol{M}_{Ad}^{-1}\boldsymbol{A})$ on the parameters h, H and p as we have seen (1.79) and (2.53).

First, we evaluate the condition number of a non-preconditioned system (2.54) and verify bound (1.79). We carried out computations using a sequence of uniform meshes having (approximately) 125, 250, 500 and 1000 elements and a fixed polynomial approximation p up to degree 5. Based on the result (1.79), we expect that the condition number will increase with the polynomial degree and also will increase with a finer partition of Ω. This is also very natural for the problem to be more costly, since we want a better approximation. The results can be found in Figures 3.1 and 3.2, respectively.

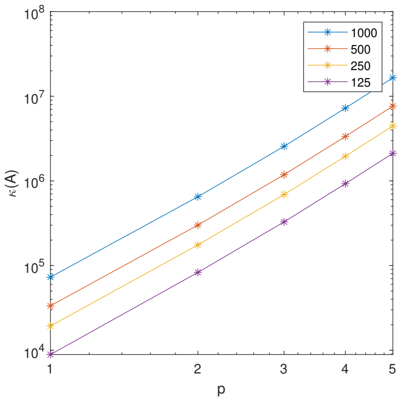

Figure 3.1: Dependecy of the condition number of the system matrix on the polynomial approximation for different meshes.

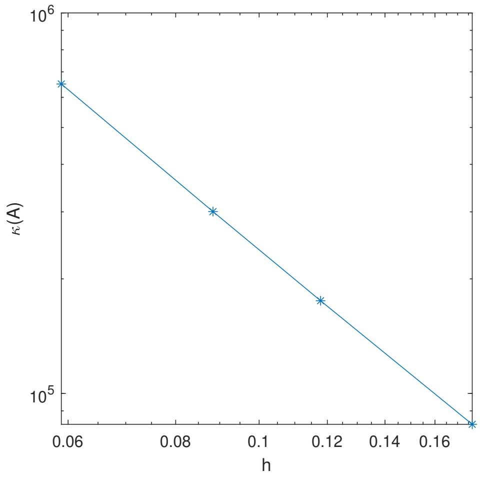

Figure 3.2: Dependecy of the condition number of the system matrix on the size of elements for $p = 2$.

Moreover, we deal with the evaluation of the condition number of the preconditioned system by the two-level additive Schwarz method. Figure 3.3 illustrates the mesh partition with subdomains and the coarse mesh.

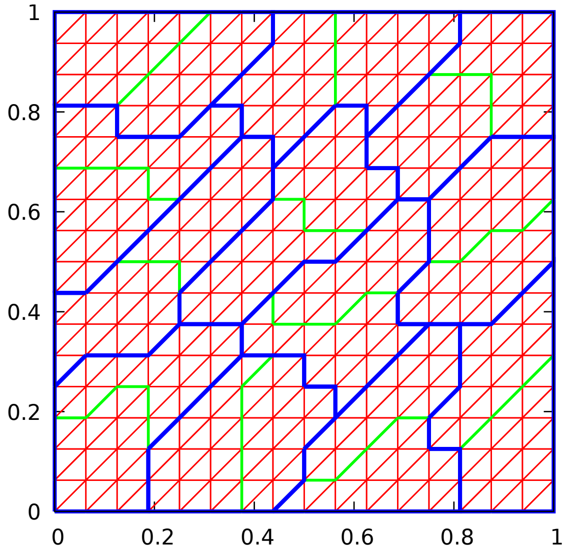

Figure 3.3: Mesh of the square $(0,1)^2$ with highlighted subdomains (blue) and coarse mesh (green).

We expect the results proven in 2.34. We will now expect to have some kind of dependence on the coarse mesh size. Hence from our method of discretization subdomains into coarse elements we also have dependence on the number of subdomains. From the implementation, we know that this will be a hard task and we can expect different results using different implementation methods for some subtask in our code. For example the method of choosing the prologation operators can be tricky and also the definition of the coarse solver. We present results with $q = p$, since we are using the constant polynomial approximation on all elements of the fine mesh.

The dependence of the preconditioned system on the polynomial degree can be seen in Figure 3.4. where we can already see that we have some computational errors. However, we still see the increase of the condition number with increasing polynomial approximation.

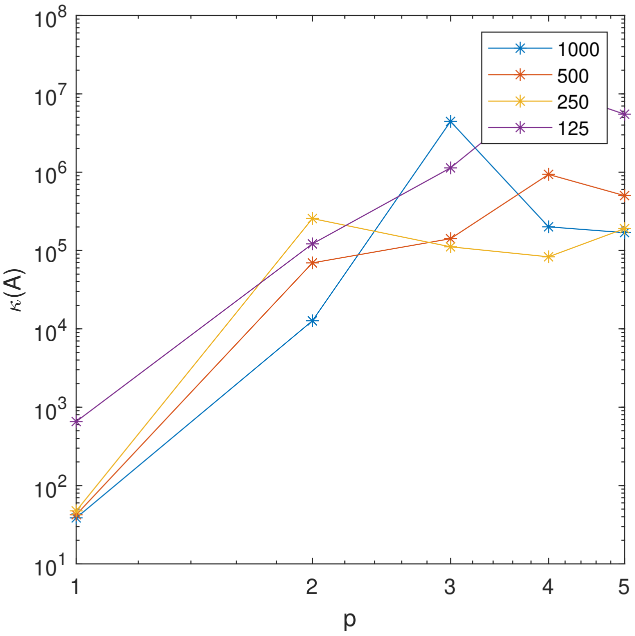

Figure 3.4: Dependency of the condition number of the preconditioned system matrix on the polynomial approximation.

Finally, the dependence of the condition number on the coarse mesh size can be seen in Figure 3.5. We can see that the condition number has decreased compared to the condition number without preconditioning, but we have behavior that we did not expect. Still, for the mesh with 500 elements, we can see the increase of the condition number, with respect to the increase of the coarse mesh elements. As we already said, this is the more tricky part to implement, and one has more options how to do some steps in the computation. Therefore, we can say that this probably is due to some implementation inaccuracies, e.g., in the setting of the coarse solver, performing on the solution of the local and coarse systems or due to the fact that we are far from asymptotic regime. However, the application of the additive Schwarz preconditionier reduces the condition number by several orders.

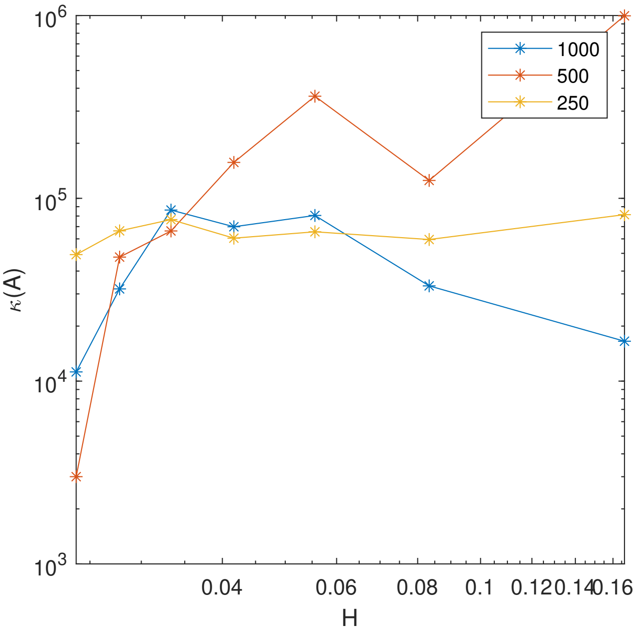

Figure 3.5: Dependency of the condition number of the preconditioned system matrix on the coarse element size H.

3.2 Symmetric linear elliptic equation

The second example is a simplified variant of the example from Dolejší and Congreve [2023]. It deals with the magneto-static field of an alternator. We consider as domain Ω only the quarter of the alternator, this is due to the symmetry of the problem. We also consider the partitioning $\Omega = \Omega_s \cup \Omega_r \cup \Omega_a$, where Ω_s represents the stator, Ω_r represents the rotor, and Ω_a represents the gap that is filled by air, see Figure 3.6.

The problem is described by the Maxwell equations for the stationary magnetic field in the form

$$\begin{aligned} \mathrm{rot} H &= f \quad \text{in } \Omega \\ \mathrm{div} B &= 0 \quad \text{in } \Omega, \end{aligned} \tag{3.3}$$

where $H = (H_1, H_2)$, is the magnetic intensity field, $B = (B_1, B_2)$ is the magnetic induction field and f is the current density. The operator rot is defined by the following $\mathrm{rot}(H) = (\frac{\partial H_2}{\partial x_1}, \frac{\partial H_1}{\partial x_2})$. We consider the constitutive relation

$$H(x) = \nu(x, |B(x)|^2)B(x), \quad x \in \Omega, \tag{3.4}$$

where ν is given by

$$\nu(x,r) = \begin{cases} \frac{1}{\mu_0} & \text{for } x \in \Omega_a \\ \frac{1}{\mu_0}\left(\alpha + (1-\alpha)\frac{r^4}{\beta + r^4}\right) & \text{for } x \in \Omega_s \cup \Omega_a. \end{cases} \tag{3.5}$$

The symbols μ_0, α, β are taken from Glowinski and Marrocco [1974]. Consequently, we get the problem

$$-\mathrm{div}(\mu(x, |\nabla u(x)|^2)\nabla u(x)) = f \quad \text{in } \Omega. \tag{3.6}$$

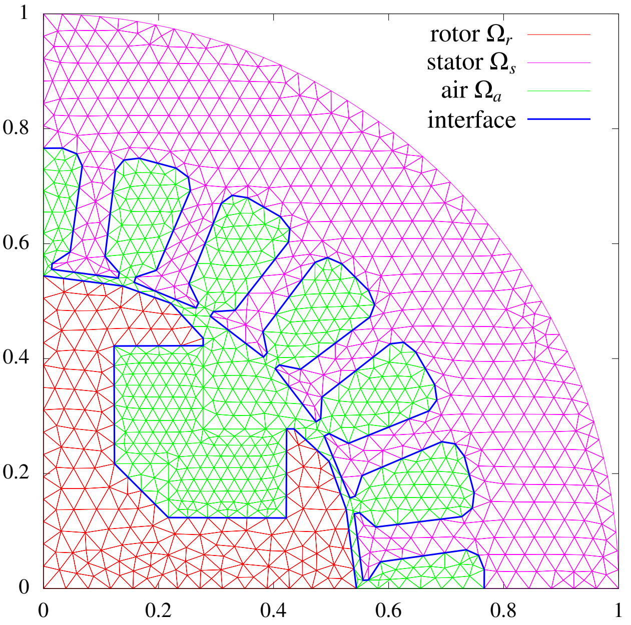

Figure 3.6: Geometry of the alternator

.

In this section, we consider the linearized variant of (3.5) – (3.6), namely we replace (3.5) by

$$\nu(x,r) = \begin{cases} \frac{1}{\mu_0} & \text{for } x \in \Omega_a \\ \frac{100}{\mu_0} & \text{for } x \in \Omega_s \cup \Omega_a, \end{cases} \tag{3.7}$$

where $\mu_0 = 1.256 \cdot 10^{-6}$.

We follow the steps done for the Laplace equation. First, we generate the system matrix with the ADGFEM code and then we use MATLAB to compute the condition number. The assembly is done again by the multiplication by a canonical vector. We use the mesh seen in Figure 3.7, where we have the fine mesh, coarse mesh, and domain decomposition partitioning. We use 20 subdomains and only look at the condition number of the system matrix, when we change the coarse mesh partitionig (partitioning of the subdomains Ω_i into macroelements H) and the polynomial degree of the approximation. The polynomial degree of the coarse solver q is taken as maximum of the polynomial degrees p_K, where $K \in \mathcal{T}_h$, $K \subset H$. We look at the polynomial degrees $p = 1, 2, 3, 4, 5$ and set $p = p_K$, $\forall K \in \mathcal{T}_h$.

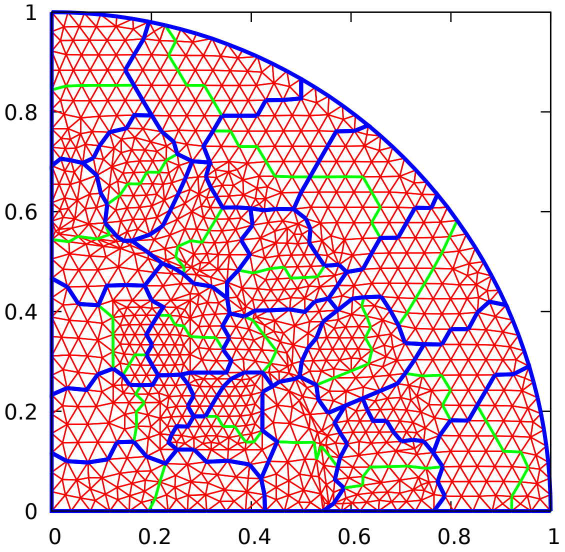

Figure 3.7: Triangular mesh of the alternator domain (red), with Ω_i partitioning (blue) and coarse mesh (green).

We examine the dependence of the condition number on the parameters h, H, p, k_0 and k_1. First, we look at the matrix of the non-preconditioned system $\boldsymbol{A}$ and its condition number. We see the results in Figure 3.8, where is the plot of the dependence of the condition number on the polynomial degree. We see that the results are in compliance with our analysis. The results are similar to the results obtained for the Laplace problem.

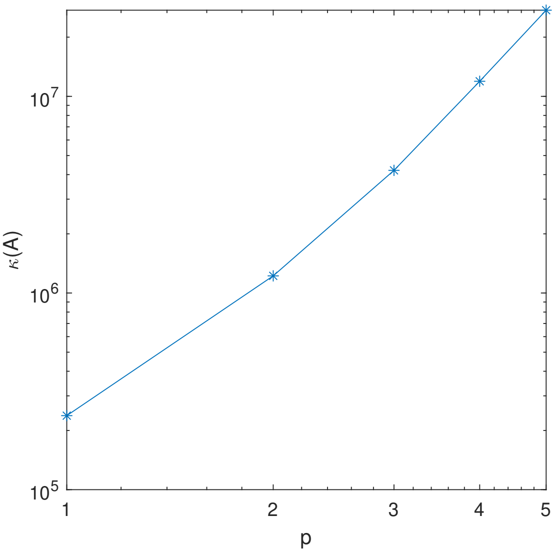

Figure 3.8: Dependence of the condition number $\kappa(\boldsymbol{A})$ (non-preconditioned system) on the polynomial approximation.

Figure 3.9 shows the dependence of the condition number of the preconditioned system $\boldsymbol{M}_{AD}^{-1}\boldsymbol{A}$ on the polynomial order p. Again we can see, that the condition number is growing with the degree of the polynomial approximation as we expected, but we can see the implmentation inaccuracies that we described earlier.

For the coarse mesh refinement we get the result that is expected by the analysis. Here we see a better graph than what we have seen in the case of the Laplace equation. This can be due to the mesh geometry and also due to the implementation done in ADGFEM. The graph for the polynomial degree $p = 2$ can be seen in Figure 3.10.

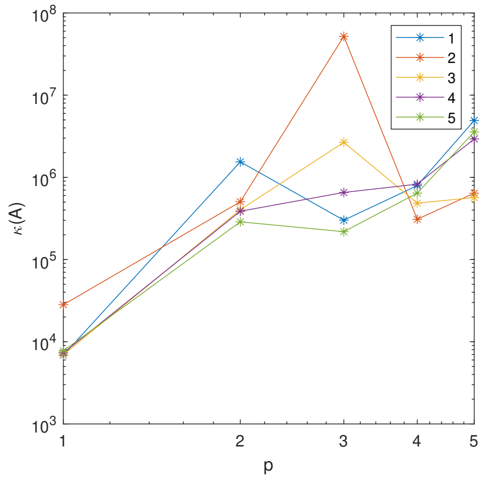

Figure 3.9: Dependence of the condition number of the preconditioned system $\kappa(\boldsymbol{M}_{AD}^{-1}\boldsymbol{A})$ on the polynomial approximation for different coarse mesh partitioning Ω_i.

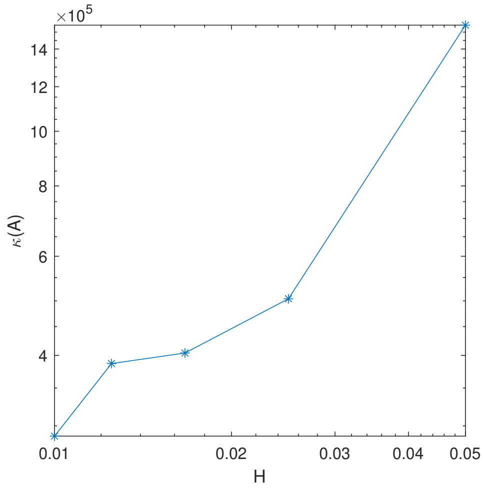

Figure 3.10: Dependence of the condition number of the preconditioned system on the coarse element size H, for $p = 2$

Finally, we look at the results we get for different ratios of k_1 and k_0. We use the ratios 10, 100, 500 and 1000 and generate the matrix for $p = 1, 2, 3, 4$. The results for the non-preconditioned matrix $\boldsymbol{A}$ can be seen in Figure 3.11 and the results for the preconditioned system can be seen in Figure 3.12. We see that our results are in agreement with the results (1.79) and (2.53). The results have some errors, which might be due to the implementation, as we discussed in the previous.

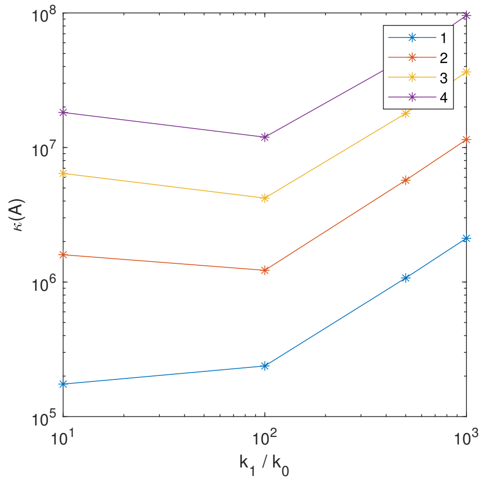

Figure 3.11: Dependence of the condition number of matrix $\boldsymbol{A}$ on the bounds k_1 and k_0, for different polynomial approximations.

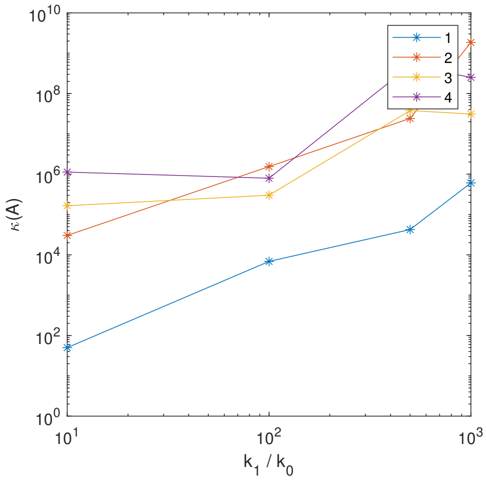

Figure 3.12: Graph of condition number of the preconditioned system depending on bounds k_1 and k_0, for different p.

Generally we can say, that the condition number of the preconditioned system is lower by few orders, than the condition number of the non-preconditioned system, despite some inaccuracies in the implementation.

3.3 Symmetric nonlinear elliptic equation

For the non-linear case of the alternator problem (3.6) we have the equations

$$-\mathrm{div}(\nu(x, |\nabla u|)\nabla u) = f, \quad \text{in } \Omega, \tag{3.8}$$

where ν is given by the nonlinear formula (3.5). The nonlinear problem (3.8) is solved as a sequence of linear ones, namely

$$-\mathrm{div}(\nu(x, |\nabla u_{k-1}|)\nabla u_k) = f, \quad \text{in } \Omega, \tag{3.9}$$

where $k = 1, 2, \ldots$ is the index of iteration.

Moreover, we are not interested in the solution itself but in the value of the *quantity of interest*, in our case the magnetic energy given by

$$J(u) := \frac{1}{2}\int_{\Omega} \nu(x, |\nabla u(x)|^2)|\nabla u(x)|^2 \mathrm{dx}. \tag{3.10}$$

The error of this quantity can be estimated by the goal-oriented techniques and based on this estimates we performed several level of the anisotropic *hp*-mesh adaptation. This approach optimizes the size and shape of mesh elements and also the polynomial approximation degrees. For more details, we refer to Dolejší and Congreve [2023], Dolejší and May [2022].

The linearized problem for each $k = 1, 2, \dots$ is solved by GMRES method with two-level additive Schwarz preconditioner. Figure 3.13 shows the generated meshes for the first and last mesh adaptation level together with the corresponding domain decomposition. The solution and *hp*-mesh are shown in Figure 3.14.

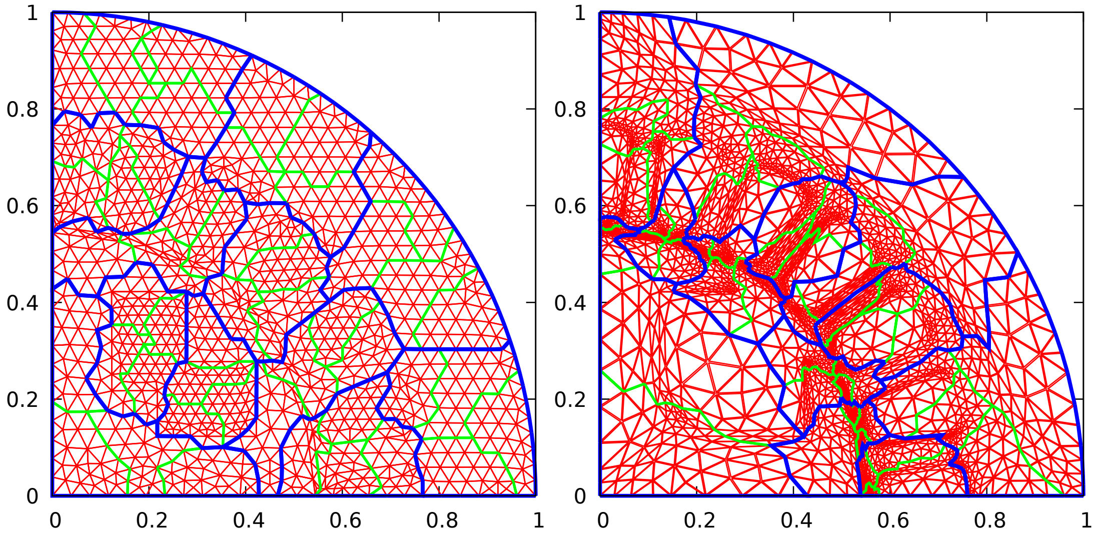

Figure 3.13: Initial mesh of the alternator on left and the final adaptation of the mesh on the right.

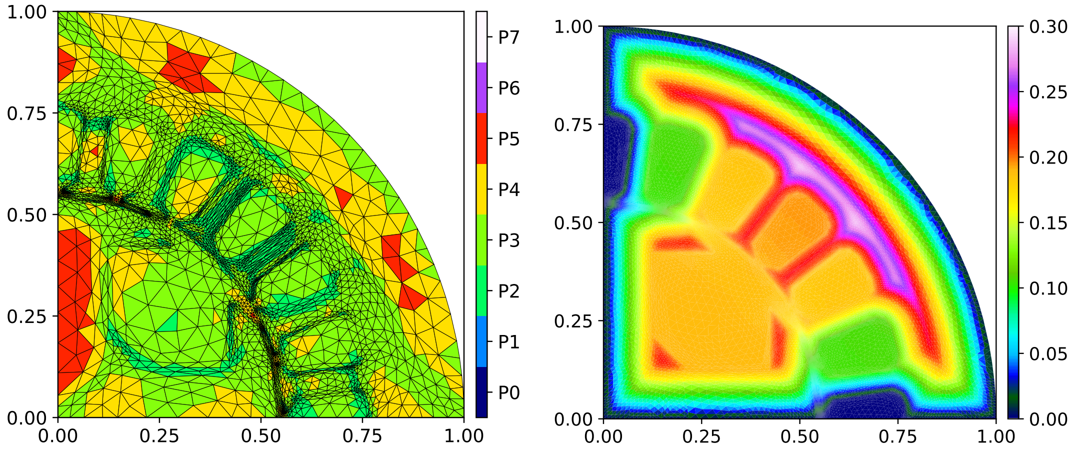

Figure 3.14: Final *hp*-mesh on the left and the corresponding magnetic potential on the right

Moreover, Figure 3.15 shows the convergence of the error of the quantity of interest $e_h = |J(u) - J(u_h)|$ (cf. (3.10)) and its estimate. The "exact" value $J(u)$ has been calculated by an overkill using a sufficiently refined grid. In this figure, we plot two types of error estimators: $\eta^{\mathrm{I}}(u_h)$, which is the residual of the primal problem tested by the interpolation error of the dusolution,on, and $\eta^{\mathrm{I}}(u_h)$, which is the bound of $\eta^{\mathrm{I}}(u_h)$ arising from the Cauchy inequality, for more details, we refer to Dolejší and Congreve [2023], Dolejší and May [2022]. We note that the mesh adaptation is carried out with respect to the estimate $\eta^{\mathrm{I}}(u_h)$.

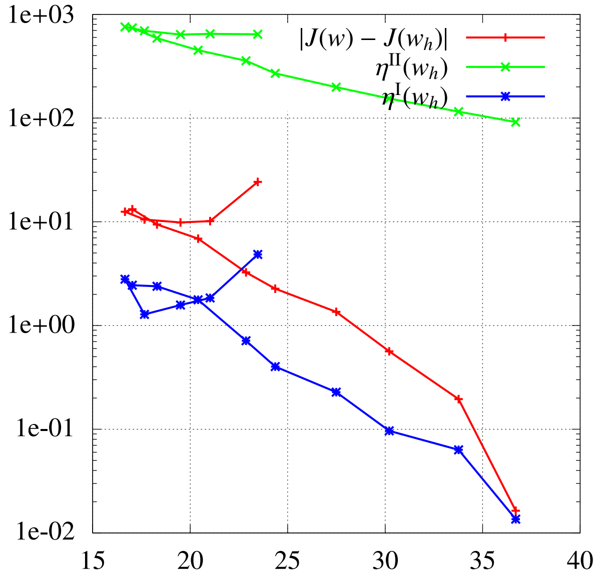

Figure 3.15: Convergence of the non-linear alternator problem. (3.8).

Finally, Figure 3.16 demonstrates the convergence of the nonlinear solver, each line corresponds to one mesh adaptation and each nodes corresponds to residual of the linearized problem (3.9) for one k.

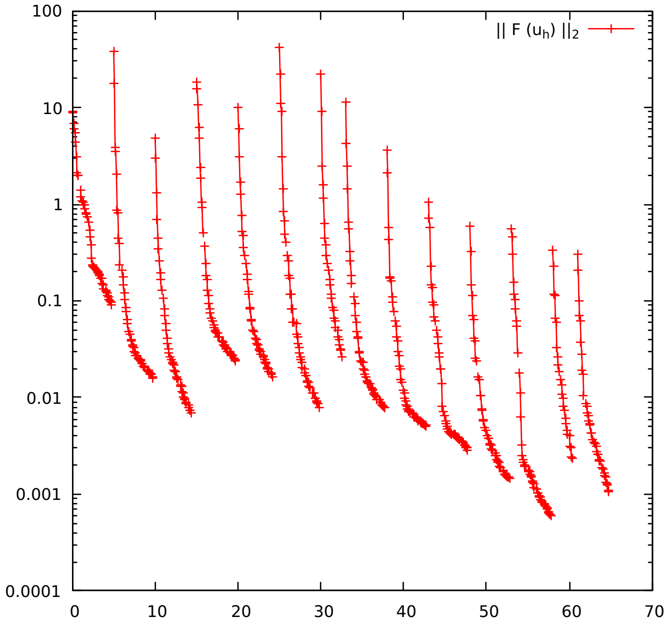

Figure 3.16: Convergence of the iterative solvers for non-linear alternator problem.

Conclusion

We dealt with the numerical solution of linear elliptic problem (1.1) by the symmetric interior penalty variant of the discontinuous Galerkin method (SIPG scheme). We proved the coercivity and continuity properties of the bilinear form arising from the SIPG discretization. These properties guarantee the existence of the discrete solution and they are important for the analysis of the bounds of the condition number of the equivalent system of algebraic equations.

In the second chapter we presented the two level non-overlapping Schwarz method for the DGM. We formulated bounds on the condition number of the preconditioned system using three assumption. We then presented new proof of the auxiliary lemma and modified proof from Antonietti and Houston [2011] of the stable decomposition assumption of our form $\mathcal{A}_h$. Moreover, we shown bounds of the condition number depending on the fine mesh size h, coarse mesh size H, data bounds k_1 and k_0 and on the polynomial degree of the approximation p. We briefly discussed the implementation of the method and presented some algorithms used for the computation, as well as some implementation problems that can arise.

In the third chapter, we presented the numerical study and verification of the theoretical results. First we considered the Laplace equation on the unit square. We observed a relative good agreement with the theoretical results, some inconsitencies were discussed. Moreover, we dealt with the numerical simulation of a linearized magneto-static field in the alternator, where the condition number of the preconditioned operator (2.53) was investigated. Finally, we dealt with the original non-linear version of the alternator problem and demonstrated the potential of the two-level additive Schwarz preconditioning.

For further research, an interesting question is the investigation of the dependence of the condition number on the number of subdomains. This aspect was investigated in Krzyżanowski [2016] but for a different setting where the subdomains are smaller than the elements of the coarse mesh.

Bibliography

P. Antonietti and P. Houston. A class of domain decomposition preconditioners for *hp*-discontinuous Galerkin finite element methods. *Journal of Scientific Computing*, 46, 01 2011. doi: 10.1007/s10915-010-9390-1.

P. Antonietti, S. Giani, and P. Houston. Domain decomposition preconditioners for discontinuous Galerkin methods for elliptic problems on complicated domains. *Journal of Scientific Computing*, 60:203–227, 2014.

Paola F. Antonietti, Paul Houston, and Iain Smears. A note on optimal spectral bounds for nonoverlapping domain decomposition preconditioners for *hp*-version discontinuous Galerkin methods. *International Journal of Numerical Analysis and Modeling*, 13(4):513–524, 2016.

V. Dolean, P. Jolivet, and F. Nataf. *An Introduction to Domain Decomposition Methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2015. doi: 10.1137/1.9781611974065. URL `https://epubs.siam.org/doi/abs/10.1137/1.9781611974065`.

V. Dolejší. *ADGFEM – Adaptive discontinuous Galerkin finite element method, in-house code*. Charles University, Prague, Faculty of Mathematics and Physics, 2020. `https://msekce.karlin.mff.cuni.cz/~dolejsi/adgfem/index.html`.

V. Dolejší and M. Feistauer. *Discontinuous Galerkin Method: Analysis and Applications to Compressible Flow*. Springer Series in Computational Mathematics. Springer International Publishing, 2015. ISBN 9783319192673. URL `https://books.google.cz/books?id=Pj4wCgAAQBAJ`.

V. Dolejší and G. May. *Anisotropic hp-Mesh Adaptation Methods*. Birkhäuser, 2022.

V. Dolejší and S. Congreve. Goal-oriented error analysis of iterative Galerkin discretizations for nonlinear problems including linearization and algebraic errors. *Journal of Computational and Applied Mathematics*, 427:115134, 2023. ISSN 0377-0427. doi: https://doi.org/10.1016/j.cam.2023.115134. URL `https://www.sciencedirect.com/science/article/pii/S037704272300078X`.

X. Feng and O. Karashian. Two-level additive Schwarz methods for a discontinuous Galerkin approximation of second order elliptic problems. *SIAM J. Numer. Anal.*, 39:1343–1365, 01 2002. doi: 10.1137/S0036142900378480.

R. Glowinski and A. Marrocco. Analyse numerique du champ magnetique d'un alternateur par elements finis et sur-relaxation ponctuelle non lineaire. *Computer Methods in Applied Mechanics and Engineering*, 3(1):55–85, 1974. ISSN 0045-7825. doi: https://doi.org/10.1016/0045-7825(74)90042-5. URL `https://www.sciencedirect.com/science/article/pii/0045782574900425`.

G. Golub and C. Van Loan. *Matrix computations*. JHU press, 2013.

P. Krzyżanowski. On a nonoverlapping additive Schwarz method for *h-p* discontinuous Galerkin discretization of elliptic problems. *Numerical Methods for Partial Differential Equations*, 32(6):1572–1590, 2016. doi: https://doi.org/10.1002/num.22063. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/num.22063`.

A. Quarteroni and A. Valli. *Numerical approximation of partial differential equations*, volume 23. Springer Science & Business Media, 2008.

B. Rynne and M. Youngson. *Linear functional analysis.* Springer Science & Business Media, 2007.

Hermann Amandus Schwarz. *Ueber einen Grenzübergang durch alternirendes Verfahren.* Zürcher u. Furrer, 1870.

A. Toselli and O. Widlund. *Domain decomposition methods-algorithms and theory*, volume 34. Springer Science & Business Media, 2004.