UNIVERSIDAD DE COSTA RICA

SISTEMA DE ESTUDIOS DE POSGRADO

A two level overlapping Schwarz preconditioner for Discontinuous Galerkin Methods

Tesis sometida a la consideración de la Comisión del Programa de Posgrado en matemática pura, para optar al grado y título de Maestría Académica en matemática pura.

por

Moisés Solano Córdoba

Ciudad Universitaria Rodrigo Facio, Costa Rica Febrero, 2024

Quisiera dedicar este trabajo a mis padres, porque nunca han dejado de apoyarme en todas las etapas de mi vida y me ayudan en todo lo que puedan, siempre estaré agradecido. A mis hermanos y sobrinos, que también me han apoyado, se merecen parte de la dedicatoria, un abrazo fraternal para todos.

Agradecimientos

Agradezco a la Universidad de Costa Rica por brindarme la oportunidad de estudiar y trabajar al mismo tiempo, lo que ha permitido enriquecer mi formación académica y el desarrollo de mis capacidades al máximo. Mi agradecimiento se extiende a la Escuela de Matemática, que siempre ha facilitado los procesos académicos por lo que he pasado y han sido un apoyo constante a lo largo de mi carrera universitaria. Quiero expresar mi gratitud a todos los profesores que compartieron sus conocimientos conmigo en este camino. Además, quiero reconocer y agradecer a todas las personas que forman parte de la Escuela de Matemática, incluyendo compañeros administrativos y misceláneos, por su contribución, de una u otra manera, a mi formación académica. Su apoyo y colaboración han sido muy importantes. Por último, no puedo dejar de mencionar a mi director de tesis, Juan Gabriel Calvo, por su constante apoyo y disponibilidad. Siempre estuvo dispuesto a aclarar mis dudas, trabajar en conjunto y proporcionarme una guía sólida durante este tiempo. Estoy profundamente agradecido por su compromiso y orientación. Su ayuda fue fundamental para el éxito de este trabajo.

Esta Tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en matemáticas de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Matemática Pura.

Dr. Javier Trejos Zelaya
Representante del SEP

Dr. Juan G. Calvo
Director de tesis

Dr. Mario Alvarez Guadamuz
Miembro de tribunal

Dr. Esteban Segura Ugalde
Miembro de tribunal

Dr. Dario Mena Arias
Presidente del tribunal

Moisés Solano Córdoba
Estudiante

iv

Contents

Abstract (spanish)

Se usarán dos herramientas para resolver el problema de Poisson con condiciones homogéneas de Dirichlet: un precondicionador de Schwarz con traslape de dos niveles para un problema generado con el método discontinuo de Galerkin. Se dará una descripción básica del DGM. Se darán algunos consejos sobre las herramientas matemáticas básicas que normalmente se utilizan para aplicación del método discontinuo, además estableceremos propiedades de aproximación y mostraremos cómo derivar estimaciones a priori. A la hora de definir el precondicionador, introducimos un espacio global de manera que la cota del número de condición del sistema precondicionado no dependa del número de subdominios. Se plantea una conjetura de que el número de condición se mantiene a lo sumo de orden logarítmico en función de los diámetros de las mallas $\frac{H}{h}$, mientras que el crecimiento es lineal en función de $\frac{H}{\delta}$, como ocurre con estudios anteriores. Se incluyen resultados numéricos que apoyan nuestra conjetura, así como casos específicos con discontinuidades en los coeficientes de la ecuación diferencial.

Abstract

Two tools will be used to solve the Poisson problem with homogeneous Dirichlet boundary conditions: a two-level overlapping Schwarz preconditioner for a problem generated using the discontinuous Galerkin method (DGM). A basic description of DGM will be provided. Some advice on the fundamental mathematical tools typically used for the application of the discontinuous method will also be given, and we will establish approximation properties and show how to derive a priori estimates. When defining the preconditioner, we introduce a global space such that the bound on the condition number of the preconditioned system does not depend on the number of subdomains. A conjecture is proposed that the condition number remains at most of logarithmic order with respect to the mesh diameters $\frac{H}{h}$, while the growth is linear concerning $\frac{H}{\delta}$, as observed in previous studies. Numerical results supporting our conjecture are included, as well as specific cases involving coefficient discontinuities in the differential equation.

List of Figures

List of Tables

UNIVERSIDAD DE COSTA RICA

SEP Sistema de Estudios de Posgrado

Autorización para digitalización y comunicación pública de Trabajos Finales de Graduación del Sistema de Estudios de Posgrado en el Repositorio Institucional de la Universidad de Costa Rica.

Yo, Moisés Solano Córdoba, con cédula de identidad 303950371, en mi condición de autor del TFG titulado A two level overlapping Schwarz preconditioner for Discontinuous Galerkin Method

Autorizo a la Universidad de Costa Rica para digitalizar y hacer divulgación pública de forma gratuita de dicho TFG a través del Repositorio Institucional u otro medio electrónico, para ser puesto a disposición del público según lo que establezca el Sistema de Estudios de Posgrado. **SI** [X] **NO** * [ ]

***En caso de la negativa favor indicar el tiempo de restricción:** ________ **año (s).**

Este Trabajo Final de Graduación será publicado en formato PDF, o en el formato que en el momento se establezca, de tal forma que el acceso al mismo sea libre, con el fin de permitir la consulta e impresión, pero no su modificación.

Manifiesto que mi Trabajo Final de Graduación fue debidamente subido al sistema digital Kerwá y su contenido corresponde al documento original que sirvió para la obtención de mi título, y que su información no infringe ni violenta ningún derecho a terceros. El TFG además cuenta con el visto bueno de mi Director (a) de Tesis o Tutor (a) y cumplió con lo establecido en la revisión del Formato por parte del Sistema de Estudios de Posgrado.

Moisés Solano C.

FIRMA ESTUDIANTE

Nota: El presente documento constituye una declaración jurada, cuyos alcances aseguran a la Universidad, que su contenido sea tomado como cierto. Su importancia radica en que permite abreviar procedimientos administrativos, y al mismo tiempo genera una responsabilidad legal para que quien declare contrario a la verdad de lo que manifiesta, puede como consecuencia, enfrentar un proceso penal por delito de perjurio, tipificado en el artículo 318 de nuestro Código Penal. Lo anterior implica que el estudiante se vea forzado a realizar su mayor esfuerzo para que no sólo incluya información veraz en la Licencia de Publicación, sino que también realice diligentemente la gestión de subir el documento correcto en la plataforma digital Kerwá.

1 — Introduction

To begin our discontinuous finite element analysis and our domain decomposition method, we must first revisit important concepts. Let Ω denote an open subset of $\mathbb{R}^2$. An expression of the form

$$\sum_{|\alpha|\leq 2} a_\alpha(x) D^\alpha u = f(x),$$

is called a second-order linear partial differential equation (PDE), where $u : \Omega \longrightarrow \mathbb{R}$ is the unknown and $\alpha = (\alpha_1, \alpha_2)$ is a multiindex. This equation can be classified as elliptic, hyperbolic or parabolic. Consider a second order differential operator L, as a bounded linear operator,

$$Lu = \sum_{|\alpha|\leq 2} a_\alpha(x) D^\alpha u.$$

As a model problem, we consider Poisson's equation with homogeneous Dirichlet boundary conditions:

$$\begin{cases} -\Delta u = f(x) & x \in \Omega, \\ u(x) = 0 & x \in \partial\Omega. \end{cases} \tag{1.1}$$

This problem is a self-adjoint partial differential equation and we will focus on designing preconditioners for this problem. It is well-know that (1.1) is well-posed; see, e.g., [Brenner and Scott, 2008, page 129, chapter 5].

The finite element theory has emerged as a powerful tool in solving this and many other problems. The method guarantees the existence and uniqueness of the numerical solution, and also that allows finding the explicit solution, considering an approximation space of finite dimension; see, e.g., [Brenner and Scott, 2008, page 57, chapter 2].

Finite Element Methods (FEM) have shown their utility in the numerical approximation of solutions to self-adjoint elliptic PDEs problems. This is due in part to the application of the method in very general computational geometries and the availability of tools for error analysis. Nevertheless, employing classical FEM for solving hyperbolic problems is generally inadequate.

Discontinuous Galerkin Methods (DGM) demonstrate appealing properties for numerical approximation of hyperbolic problems and exhibit enhanced stability characteristics near steep gradients. In general, DGM represents a highly attractive class of arbitrary order methods for numerically solving various types of PDE problems where classical FEMs are inapplicable [Georgoulis, 2011].

The main purpose of the first part of the thesis is to present a basic description of DGM for elliptic problems, with homogeneous boundary conditions. For simplicity, we consider Poisson's equation; however, the best results are given for numerical solutions of hyperbolic problems, as we mentioned before. We will give some hints on the basic mathematical tools typically used to study and will state approximation properties and show how to derive *a-priori* estimates.

On the other hand, to solve the resulting linear system after applying DGM, we will develop a preconditioner based on Domain Decomposition Methods (DDM). This method generally refers to the splitting of a domain of a PDE into coupled problems on smaller subdomains forming a (finite) partition of the original domain. The algebraic systems arising from the approximation of the partial differential equation:

$$Au = f, \tag{1.2}$$

are large for real-life applications, and direct methods are not an option. Hence, we consider iterative methods with techniques of preconditioning. There are two categories of DDM: structuring method based on a partition of the domain into disjoint regions, and overlapping methods. The best known example in this category is the Schwarz method, which is the method we will consider. Given an initial estimate u_0, which vanishes on $\partial\Omega$, once the iteration u_n is known, the next iteration u_{n+1} can be determined in first instance, in a number of steps that is equal to the number of subdomains. We will incorporate a coarse space, which will be necessary for good convergence, which will not depend on the number of subdomains.

The general idea of a preconditioner method is that, instead of solving system (1.2), we will solve

$$M^{-1}Au = M^{-1}f,$$

where M is an invertible operator such that

$$\kappa(M^{-1}A) \ll \kappa(A).$$

We recall that the condition number κ of a matrix is a measure that tells us how sensitive the solution of a system of linear equations is to small changes in the input. A high condition number indicates that small perturbations in the matrix data can lead to large changes in the solution.

A large part of the work in DDM focuses on the choice of coarse spaces. We will introduce a coarse subspace for a two-level overlapping Schwarz method. Furthermore, in order to prove bounds for these two Schwarz operators, it is enough to make two assumptions, that will produce a condition number bound proportional to

$$\left(1 + \frac{H}{\delta}\right),$$

where H is the maximum diameter of the subdomains and δ is the minimum of the measurements of the overlaps. We will focus in the construction of the preconditioner and the validation of results; the theoretical analysis is future (ongoing) work.

All this information is organized by chapters as follows. In Chapter 2, we will give a complete review of DGM. In Chapter 3, we will describe the family of two-level overlapping Schwarz methods. In Chapter 4, we will define a new preconditioner for our model problem and validate its performance with several numerical experiments. Finally, in Chapter 5, we include some conclusions, final remarks and future work.

1.1 Objectives

General Objective

Design and implement a two-level overlapping additive Schwarz preconditioner for the Discontinuous Galerkin method and elliptic problems.

Specific Objectives

1. Compose a theoretical framework that includes essential content on DGM and the two-level overlapping Schwarz method to support the research.

2. Explain the DGM, completing details in the theory.

3. Explain the two-level overlapping Schwarz method.

4. Implement the DGM in Matlab.

5. Verify the convergence of the DGM using toy examples.

6. Implement the two-level overlapping Schwarz method in Matlab.

7. Numerically validate the performance of the two-level overlapping preconditioner.

2 — Discontinuous Galerkin Methods

In this section we describe the basic DGM theory. A theoretical framework is given on which we can work and the basic theory of Sobolev spaces is used. We define the norms on which the necessary definitions will be applied. We then define the space on which we will work, in addition to the respective bilinear form, which will give us the matrix system. Once we have the norm, the basis space and the bilinear form, we can guarantee the existence and uniqueness of a solution for the given problem. We end this chapter by providing results on convergence and computing the matrix system.

2.1 Preliminaries and notation

Let $\Omega \subset \mathbb{R}^2$ be a bounded open set with non-empty interior and piecewise smooth boundaries. We start our work by imposing some restrictions on the geometry of triangulations of Ω. We divide Ω into triangles with the following characteristics:

Definition 2.1. (Admissible triangulation) A partition $\mathcal{T} = \{K_1, \ldots, K_m\}$ of Ω into open triangles is admissible if:

i) $\overline{\Omega} = \displaystyle\bigcup_{i=1}^{m} \overline{K}_i$.

ii) If $\overline{K}_i \cap \overline{K}_j$ consists of a point x, then x is a vertex of K_i and K_j.

iii) If $\overline{K}_i \cap \overline{K}_j$ has more than one point, then $K_i \cap K_j$ is a common edge of K_i and K_j.

We refer to [Brenner and Scott, 2008, p.79] for more details. We write $\mathcal{T}_h$ if $\mathrm{diam}(K_i) \leq 2h$, where we define the diameter of a triangle as the measurement of the longest side of the triangle. We have the following definition regarding the shape of the triangles:

Definition 2.2. (Regular and uniform triangulation)

1. A family $\{\mathcal{T}_h\}_h$ is called shape regular if there exists $t > 0$ such that, $\forall K \in \mathcal{T}_h$, K contains a circle of radius ρ_K with

$$\rho_K \geq \frac{h_K}{t},$$

where $h_K = \frac{1}{2}\mathrm{diam}(K)$.

2. A family $\{\mathcal{T}_h\}$ is called uniform if

$$\rho_K \geq \frac{h}{t},$$

where $h = \max_K(\mathrm{diam}(K))$.

We refer to [Braess, 2007, page 61, Def. 5.1] for more details. We remark that is essential that no angle is sufficiently close to 180 degrees, as shown in [Babuska and Aziz, 1976].

Definition 2.3. Given a triangulation $\mathcal{T}_h$ of Ω, denote by $\mathcal{E}_h^{\mathfrak{b}}$ the set of boundary edges, and by $\mathcal{E}_h^{\mathfrak{i}}$ the set of interior edges. Let

$$\mathcal{E}_h = \mathcal{E}_h^{\mathfrak{i}} \cup \mathcal{E}_h^{\mathfrak{b}}$$

denote the set of edges of the triangulation. For each element $K \in \mathcal{T}_h$, it holds that

$$\partial K \backslash \partial\Omega \in \mathcal{E}_h^{\mathfrak{i}} \quad \text{and} \quad \partial K \cap \partial\Omega \in \mathcal{E}_h^{\mathfrak{b}}.$$

We will consider discontinuous function spaces that approximate a continuous function. Therefore, we need the following definitions in order to "control" discontinuities across the edges of a triangulation.

Definition 2.4. (Averages and jumps) Consider two elements K_i, $K_j \in \mathcal{T}_h$ that share a common edge e; i.e.,

$$e := \partial K_i \cap \partial K_j \in \mathcal{E}_h^{\mathfrak{i}}.$$

Let $v : \overline{K_i \cup K_j} \to \mathbb{R}$ and $\boldsymbol{q} : \overline{K_i \cup K_j} \to \mathbb{R}^2$, be functions such that v and $\boldsymbol{q}$ are continuous functions in K_i and K_j.
Denote by $v_{K_n}, \boldsymbol{q}_{K_n}$ the restrictions of v and $\boldsymbol{q}$ to element K_n ($n \in \{i, j\}$). The averages of v and $\boldsymbol{q}$ for a given $x \in e$ are given by

$$\{\{v\}\} := \frac{1}{2}(v_{K_i} + v_{K_j}) \quad \text{and} \quad \{\{\boldsymbol{q}\}\} := \frac{1}{2}(\boldsymbol{q}_{K_i} + \boldsymbol{q}_{K_j}),$$

respectively. The jumps of v and $\boldsymbol{q}$ in x are defined as

$$[\![v]\!] := v_{K_i}\boldsymbol{n}_{K_i} + v_{K_j}\boldsymbol{n}_{K_j} \quad \text{and} \quad [\![\boldsymbol{q}]\!] := \boldsymbol{q}_{K_i} \cdot \boldsymbol{n}_{K_i} + \boldsymbol{q}_{K_j} \cdot \boldsymbol{n}_{K_j},$$

respectively, where $\boldsymbol{n}_{K_i}$, $\boldsymbol{n}_{K_j}$ are the unit outward normal vectors of the elements K_i and K_j, respectively; see Figure 2.1. If $e \in \mathcal{E}_h^{\mathfrak{b}}$, with $e \subset \partial K$, $K \in \mathcal{T}_h$, we let

$$\{\{v\}\} := v_K, \quad \{\{\boldsymbol{q}\}\} := \boldsymbol{n}_K, \quad [\![v]\!] := v_K\boldsymbol{n}_K, \quad [\![\boldsymbol{q}]\!] := \boldsymbol{q}_K \cdot \boldsymbol{n}_K.$$

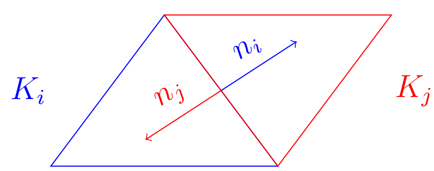


Figure 2.1: Two adjacent triangles K_i and K_j share edge e with corresponding unit outward normal vectors n_i, n_j.

Definition 2.5. (Products) Let $\boldsymbol{f}$ and $\boldsymbol{g}$ be two functions such that $\boldsymbol{f} \cdot \boldsymbol{g} \in L^1(e)$ and $\boldsymbol{f} \cdot \boldsymbol{g} \in L^1(K)$, for all $e \in \mathcal{E}_h$ and for all $K \in \mathcal{T}_h$. We define

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle_{\mathcal{E}_h} := \sum_{e \in \mathcal{E}_h} \int_e \boldsymbol{f} \cdot \boldsymbol{g} \; ds, \quad (\boldsymbol{f}, \boldsymbol{g})_{\mathcal{T}_h} := \sum_{K \in \mathcal{T}_h} \int_K \boldsymbol{f} \cdot \boldsymbol{g} \; dx,$$

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle_{\mathcal{E}_h^{\mathfrak{b}}} := \sum_{e \in \mathcal{E}_h^{\mathfrak{b}}} \int_e \boldsymbol{f} \cdot \boldsymbol{g} \; ds, \quad \langle \boldsymbol{f}, \boldsymbol{g} \rangle_{\mathcal{E}_h^{\mathfrak{i}}} := \sum_{e \in \mathcal{E}_h^{\mathfrak{i}}} \int_e \boldsymbol{f} \cdot \boldsymbol{g} \; ds.$$

We will use the following formula, based on [Brezzi and Marini, 2014], which allows us to rewrite the sum over the elements, such as the sum of a product over all edges and another product over the internal edges. This formula allows us to defined the bilinear form for the variational formulation in Section 2.2.

Theorem 2.6. *Let v and $\boldsymbol{q}$ a scalar and vector functions, respectively, such that $v\boldsymbol{q} \in (L^1(\partial K))^2$, $\forall K \in \mathcal{T}_h$, $[\![v]\!] \cdot \{\{\boldsymbol{q}\}\} \in L^1(e)$, for all edges $e \in \mathcal{E}_h$, and*

$\{\{v\}\}[\![\boldsymbol{q}]\!] \in L^1(e)$, *for all edges* $e \in \mathcal{E}_h^{\mathrm{i}}$. *Then, it holds that:*

$$\sum_{K\in\mathcal{T}_h} \int_{\partial K} v\boldsymbol{q}\cdot\boldsymbol{n}_K ds = \langle [\![v]\!], \{\{\boldsymbol{q}\}\}\rangle_{\mathcal{E}_h} + \langle \{\{v\}\}, [\![\boldsymbol{q}]\!]\rangle_{\mathcal{E}_h^{\mathrm{i}}}.$$

Proof. According to

$$\begin{aligned}
\langle [\![v]\!], &\{\{\boldsymbol{q}\}\}\rangle_{\mathcal{E}_h} + \langle \{\{v\}\}, [\![\boldsymbol{q}]\!]\rangle_{\mathcal{E}_h^{\mathrm{i}}} \\
&= \sum_{e\in\mathcal{E}_h} \int_e [\![v]\!]\cdot\{\{\boldsymbol{q}\}\} ds + \sum_{e\in\mathcal{E}_h^{\mathrm{i}}} \int_e \{\{v\}\}[\![\boldsymbol{q}]\!] ds \\
&= \sum_{e\in\mathcal{E}_h^{\mathrm{i}}} \int_e \frac{1}{2}\big(v_{K_i}\boldsymbol{n}_{K_i} + v_{K_j}\boldsymbol{n}_{K_j}\big)\cdot\big(\boldsymbol{q}_{K_i}+\boldsymbol{q}_{K_j}\big) ds + \sum_{e\in\mathcal{E}_h^{\mathfrak{b}}} \int_e v\boldsymbol{q}\cdot\boldsymbol{n} ds \\
&\quad + \sum_{e\in\mathcal{E}_h^{\mathrm{i}}} \int_e \frac{1}{2}\big(v_{K_i}+v_{K_j}\big)\big(\boldsymbol{q}_{K_i}\cdot\boldsymbol{n}_{K_i} + \boldsymbol{q}_{K_j}\cdot\boldsymbol{n}_{K_j}\big) ds \\
&= \sum_{e\in\mathcal{E}_h^{\mathrm{i}}} \int_e \big(v_{K_i}\boldsymbol{n}_{K_i}\cdot\boldsymbol{q}_{K_i} + v_{K_j}\boldsymbol{n}_{K_j}\cdot\boldsymbol{q}_{K_j}\big) ds + \sum_{e\in\mathcal{E}_h^{\mathfrak{b}}} \int_e v\boldsymbol{q}\cdot\boldsymbol{n} ds \\
&= \sum_{K\in\mathcal{T}_h} \int_{\partial K} v\boldsymbol{q}\cdot\boldsymbol{n}_K ds,
\end{aligned}$$

where we have distributed and added the products on the integrals related to interior edges, as follows

$$\left\{\begin{array}{l}
\frac{1}{2}v_{K_i}\boldsymbol{n}_{K_i}\cdot\boldsymbol{q}_{K_i} + \frac{1}{2}v_{K_i}\boldsymbol{n}_{K_i}\cdot\boldsymbol{q}_{K_j} + \frac{1}{2}v_{K_j}\boldsymbol{n}_{K_j}\cdot\boldsymbol{q}_{K_i} + \frac{1}{2}v_{K_j}\boldsymbol{n}_{K_j}\cdot\boldsymbol{q}_{K_j} \\
\underline{+\frac{1}{2}v_{K_i}\boldsymbol{n}_{K_i}\cdot\boldsymbol{q}_{K_i} + \frac{1}{2}v_{K_i}\boldsymbol{n}_{K_j}\cdot\boldsymbol{q}_{K_j} + \frac{1}{2}v_{K_j}\boldsymbol{n}_{K_i}\cdot\boldsymbol{q}_{K_i} + \frac{1}{2}v_{K_j}\boldsymbol{n}_{K_j}\cdot\boldsymbol{q}_{K_j}}
\end{array}\right.$$

$$= v_{K_i}\boldsymbol{n}_{K_i}\cdot\boldsymbol{q}_{K_i} + v_{K_j}\boldsymbol{n}_{K_j}\cdot\boldsymbol{q}_{K_j}.$$

□

In order to define a required norm for the DGM, we recall the definitions of relevant norms, seminorms, weak derivatives and Sobolev spaces. The definition of this DG norm includes a component that measures the effect of discontinuities, and the rest of the components are quite traditional, as we will see.

Definition 2.7. Given a geometric object $\mathcal{O}$ (an edge, an element or a general domain) and a sufficiently smooth function v defined over $\mathcal{O}$, we define the norm in $L^2(\mathcal{O})$

$$\|v\|_{0,\mathcal{O}}^2 = |v|_{0,\mathcal{O}}^2 = \int_{\mathcal{O}} v^2 d\mathcal{O}.$$

Definition 2.8. For a domain $\mathcal{O} \subset \mathbb{R}^2$, as in the definition above, denote by $C_0^\infty(\mathcal{O})$ the set of infinitely differentiable functions with compact support in $\mathcal{O}$.

Definition 2.9. (Locally integrable) Given a domain $\mathcal{O}$, and a compact subset $\mathcal{K}$ of the interior of $\mathcal{O}$, the set of locally integrable functions is denoted by

$$L^1_{loc}(\mathcal{O}) := \{f : f \in L^1(\mathcal{K}) \quad \forall \mathcal{K} \subset \mathring{\mathcal{O}}\}.$$

Definition 2.10. (Weak derivate) Let a function $f \in L^1_{loc}(\mathcal{O})$ and a multi-index α. We say that f has a weak derivative, D_w^α, if there exists a function $g \in L^1_{loc}(\mathcal{O})$ such that

$$\int_{\mathcal{O}} g(x)\phi(x)dx = (-1)^{|\alpha|} \int_{\mathcal{O}} f(x)\phi^{(\alpha)}(x)dx, \qquad \forall \phi \in C_0^\infty(\mathcal{O}),$$

If such a g exists, we write $D_w^\alpha f = g$.

Definition 2.11. (Seminorm, Sobolev norm and Sobolev Space) Let $k \in \mathbb{N}$, $v \in L^1_{loc}(\mathcal{O})$. Suppose that $D_w^\alpha v$ exists for all $|\alpha| \leq k$. Define the seminorm

$$|v|_{k,\mathcal{O}}^2 := \sum_{|\alpha|=k} \|D_w^\alpha v\|_{0,\mathcal{O}}^2,$$

and the Sobolev norm

$$\|v\|_{W_2^k(\mathcal{O})}^2 := \sum_{|\alpha|\leq k} \|D_w^\alpha v\|_{0,\mathcal{O}}^2.$$

Define also the Sobolev space

$$W_2^k(\mathcal{O}) := \{v \in L^1_{loc}(\mathcal{O}) : \|v\|_{W_2^k(\mathcal{O})} < \infty\},$$

which is also denoted by $H^k(\mathcal{O}) = W_2^k(\mathcal{O})$.

Definition 2.12. (Norm jump) For a function v such that $[\![v]\!] \in (L^2(e))^2 \; \forall e \in \mathcal{E}_h$, define

$$\|v\|_j^2 = \|v\|_{jump}^2 := \sum_{e\in\mathcal{E}_h} \frac{1}{h} \|[\![v]\!]\|_{0,e}^2.$$

For $v \in H^k(K)$ for all $K \in \mathcal{T}_h$ and $k \in \mathbb{N}$, define

$$|\nabla v|_{k,h}^2 := \sum_{K \in \mathcal{T}_h} |\nabla v|_{k,K}^2, \qquad |v|_{k+1,h}^2 := |\nabla v|_{k,h}^2.$$

Definition 2.13. (DG norm) Let $v \in H^2(K)$ for all $K \in \mathcal{T}_h$, such that $[\![v]\!] \in (L^2(e))^2$ for all $e \in \mathcal{E}_h$. We define the DG norm

$$\|v\|_{DG}^2 := \|v\|_j^2 + |v|_{1,h}^2 + h^2 |v|_{2,h}^2 = \|v\|_j^2 + |\nabla v|_{0,h}^2 + h^2 |\nabla v|_{1,h}^2. \tag{2.1}$$

We present two useful lemmas: (i) a Poincaré inequality in two dimensions, and (ii) the inequality of Ciarlet. Poincaré inequality allows to bound the energy of a function on the border of an element K, in terms of the sum of the energy of the function, on K and the energy of the gradient function on the element K (where each term is proportional to the length of the element and its multiplicative inverse). A way, in general, to understand how Poincaré's lemma works is: if the value of the function is large at a point, we have two possibilities: it has a large integral or it has a large derivative. In the case of a large integral, the value remains large and we multiply it by the multiplicative inverse of the diameter. In the case of a large derivative, the values of the functions decay rapidly and the integral of the derivative is multiplied by the diameter; see Figure 2.2

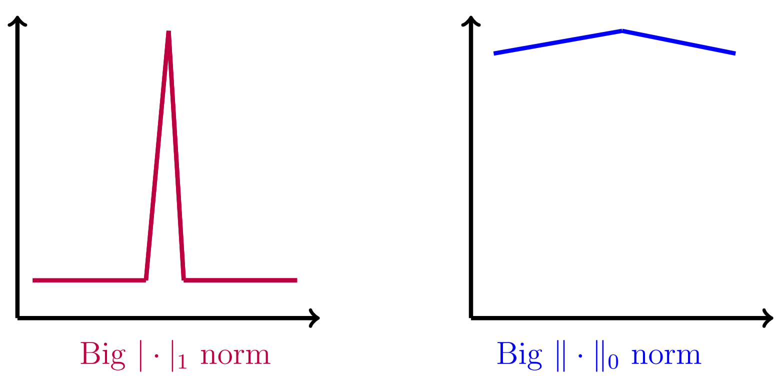


Figure 2.2: Idea in one dimension of Poincaré inequality.

Lemma 2.14 (Poincaré inequality in two dimensions)**.** *Let v a function such that $v \in L^2(\partial K)$ and $v \in H^1(K)$, with $K \in \mathcal{T}_h$. Then, it holds that*

$$\|v\|_{0,\partial K}^2 \leq C\Big(l^{-1}\|v\|_{0,K}^2 + l|v|_{1,K}^2\Big),$$

where the constant C and the characteristic length l can depend on several geometric features, but the constant C does not depend on the size of K.

Proof. See [Brezzi and Marini, 2014]. $\square$

The second important lemma is the inequality of Ciarlet, which allows to bound the seminorm of order s, by the seminorm of any value k less than s, proportional to h^{-2}. This will be important in demonstrating the continuity of the bilinear form (which will be defined later), in addition to being used in the search for convergence estimates. However, first we will use Lemma 2.15 in the proof of Theorem 2.16.

Lemma 2.15 (Ciarlet inequality)**.** *Let $K \in \mathcal{T}_h$ and $v \in H^s(K)$. Furthermore, suppose that v is a polynomial of degree at most r. It holds that*

$$|v|_{s,K}^2 \leq Ch^{-2(s-k)}\|v\|_{k,K}^2,$$

for s, k integers with $s \geq k$ and $h =$ diam (K).

Proof. See [Brenner and Scott, 2008, Lemma 4.5.3, pag. 111]. $\square$

The following theorem tells us that, in the case of functions that are polynomial of degree less than or equal to r, the norm $\|\cdot\|_{DG}$ is equivalent to the same norm without the third term, corresponding to the seminorm of order 2.

Theorem 2.16. *Let $v \in H^2(K)$, $K \in \mathcal{T}_h$, and $[\![v]\!] \in L^2(e)$, $e \in \mathcal{E}_h$, such that v is a piecewise polynomial of degree at most r. We have that*

$$\|v\|_{DG}^2 \simeq \|v\|_{jump}^2 + |v|_{1,h}^2.$$

Proof. We begin by using Lemma 2.15 for $s = 2$ and $k = 1$, it follows that

$$|v|_{2,K}^2 \leq Ch^{-2}|v|_{1,K}^2,$$

and then

$$h^2|v|_{2,h}^2 = h^2 \sum_{K\in\mathcal{T}_h} |v|_{2,K}^2 \leq Ch^2h^{-2} \sum_{K\in\mathcal{T}_h} |v|_{1,K}^2 = C|v|_{1,h}^2,$$

we now combine the inequalities to obtain

$$\|v\|_{jump}^2 + |v|_{1,h}^2 \leq \|v\|_{DG}^2$$

$$
\begin{aligned}
&= \|v\|_{jump}^2 + |v|_{1,h}^2 + h^2|v|_{2,h}^2 \\
&\le \|v\|_{jump}^2 + |v|_{1,h}^2 + C|v|_{1,h}^2 \\
&= \|v\|_{jump}^2 + (1+C)|v|_{1,h}^2 \\
&\le (1+C)(\|v\|_{jump}^2 + |v|_{1,h}^2).
\end{aligned}
$$

In summary,

$$
\|v\|_{jump}^2 + |v|_{1,h}^2 \le \|v\|_{DG}^2 \le (1+C)(\|v\|_{jump}^2 + |v|_{1,h}^2),
$$

which completes the proof. □

Finally, to finish this subsection, we demonstrate that the DG norm is actually a norm. The hypotheses that are needed are included in Definition 2.13. However, we will see in the next section that when defining our approximation space, the hypotheses of the theorem are trivially satisfied and we can associate the DG norm to our space.

Theorem 2.17. *The DG norm defined in (2.1) is a norm with the hypothesis of Definition 2.13.*

Proof. We begin by proof the homogeneity property. For the first component of the DG norm, we have that

$$
\begin{aligned}
\|\alpha v\|_j^2 &= \sum_{e\in\mathcal{E}_h} \frac{1}{h}\|[\![\alpha v]\!]\|_{0,e}^2 \\
&= \sum_{e\in\mathcal{E}_h} \frac{1}{h}\|\alpha v_i \boldsymbol{n}_i + \alpha v_j \boldsymbol{n}_j\|_{0,e}^2 \\
&= \sum_{e\in\mathcal{E}_h} \frac{1}{h}|\alpha|^2\|[\![v]\!]\|_{0,e}^2 \\
&= |\alpha|^2\|v\|_j^2 \\
&\implies \|\alpha v\|_j = |\alpha|\|v\|_j.
\end{aligned}
$$

For the second term,

$$
\begin{aligned}
|\alpha v|_{1,h}^2 = |\nabla(\alpha v)|_{0,h}^2 &= \sum_{K\in\mathcal{T}_h} |\nabla(\alpha v)|_{0,K}^2 = \sum_{K\in\mathcal{T}_h} |\alpha\nabla v|_{0,K}^2 \\
&= \sum_{K\in\mathcal{T}_h} |\alpha|^2|\nabla v|_{0,K}^2 = |\alpha|^2 \sum_{K\in\mathcal{T}_h} |\nabla v|_{0,K}^2
\end{aligned}
$$

$$
\begin{aligned}
&= |\alpha|^2 |\nabla v|^2_{0,h} = |\alpha|^2 |v|^2_{1,h} \\
&\Longrightarrow |\alpha v|_{1,h} = |\alpha||v|_{1,h}.
\end{aligned}
$$

Similarly, for the third term,

$$|\alpha v|_{2,h} = |\alpha||v|_{2,h}.$$

We then obtain that

$$
\begin{aligned}
\|\alpha v\|^2_{DG} &= |\alpha|^2 \|v\|^2_j + |\alpha|^2 |v|^2_{1,h} + h^2 |\alpha|^2 |v|^2_{2,h} \\
&= |\alpha|^2 \|v\|^2_{DG}.
\end{aligned}
$$

We now prove the triangle inequality property. By using the triangular inequality for the norm in L^2, and Cauchy-Schwarz inequality, we deduce that

$$
\begin{aligned}
\|u+v\|^2_j &= \sum_{e\in\mathcal{E}_h} \frac{1}{h} \|[\![u+v]\!]\|^2_{0,e} \\
&= \sum_{e\in\mathcal{E}_h} \frac{1}{h} \|(u+v)_i \boldsymbol{n}_i - (u+v)_j \boldsymbol{n}_i\|^2_{0,e} \\
&= \sum_{e\in\mathcal{E}_h} \frac{1}{h} \|u_i \boldsymbol{n}_i - u_j \boldsymbol{n}_i + v_i \boldsymbol{n}_i - v_j \boldsymbol{n}_i\|^2_{0,e} \\
&\leq \sum_{e\in\mathcal{E}_h} \frac{1}{h} (\|u_i \boldsymbol{n}_i - u_j \boldsymbol{n}_i\|_{0,e} + \|v_i \boldsymbol{n}_i - v_j \boldsymbol{n}_i\|_{0,e})^2 \\
&= \sum_{e\in\mathcal{E}_h} \frac{1}{h} (\|[\![u]\!]\|_{0,e} + \|[\![v]\!]\|_{0,e})^2 \\
&\leq \|u\|^2_j + \|v\|^2_j + 2\frac{1}{h} \left(\sum_{e\in\mathcal{E}_h} \|[\![u]\!]\|^2_{0,e} \right)^{\frac{1}{2}} \left(\sum_{e\in\mathcal{E}_h} \|[\![v]\!]\|^2_{0,e} \right)^{\frac{1}{2}} \\
&= \|u\|^2_j + \|v\|^2_j + 2 \left(\sum_{e\in\mathcal{E}_h} \frac{1}{h} \|[\![u]\!]\|^2_{0,e} \right)^{\frac{1}{2}} \left(\sum_{e\in\mathcal{E}_h} \frac{1}{h} \|[\![v]\!]\|^2_{0,e} \right)^{\frac{1}{2}} \\
&= \|u\|^2_j + \|v\|^2_j + 2\|u\|_j \|v\|_j = (\|u\|_j + \|v\|_j)^2,
\end{aligned}
$$

and we conclude that $\|u+v\|_j \leq \|u\|_j + \|v\|_j$. Moreover,

$$|\nabla(u+v)|^2_{0,h} = \sum_{K\in\mathcal{T}_h} |\nabla u + \nabla v|^2_{0,K} \leq \sum_{K\in\mathcal{T}_h} \big(|\nabla u|_{0,K} + |\nabla v|_{0,K}\big)^2$$

$$\leq |\nabla u|_{0,h}^2 + |\nabla v|_{0,h}^2 + 2\Big(\sum_{K\in\mathcal{T}_h} |\nabla u|_{0,K}^2\Big)^{\frac{1}{2}} \Big(\sum_{K\in\mathcal{T}_h} |\nabla v|_{0,K}\Big)^{\frac{1}{2}}$$
$$= (|\nabla u|_{0,h} + |\nabla v|_{0,h})^2.$$

The case of the third term is analogous.

Finally, we prove the third property of a norm (positivity). If $\|v\|_{DG} = 0$, then $\|v\|_j = 0$ and $|v|_{1,h} = 0$. If

$$|\nabla v|_{0,h} = |v|_{1,h} = 0 \Longrightarrow \|\nabla v\|_{0,K} = 0 \Longrightarrow \nabla v = 0 \text{ a.e. on } K,$$

that is, c is constant a.e. on K. On the other hand,

$$\begin{aligned}\|v\|_j = 0 &\Longrightarrow \|[\![v]\!]\|_{0,e} = 0 \ \forall e \in \mathcal{E}_h\\ &\Longrightarrow [\![v]\!] = 0 \text{ ae on } e,\ \forall e \in \mathcal{E}_h\\ &\Longrightarrow v_i\boldsymbol{n}_i = v_j\boldsymbol{n}_i \text{ ae on } e,\ \forall e \in \mathcal{E}_h\\ &\Longrightarrow v_i = v_j \text{ ae on } e,\ \forall e \in \mathcal{E}_h,\end{aligned}$$

and we conclude that the function v is continuous from one triangle to another, and since the value of v at the boundary is zero, then

$$v \equiv 0 \text{ ae on } \Omega.$$

□

2.2 DGM for Poisson equation

The problem to solve is the following: Given a two-dimensional domain Ω and $f \in L^2(\Omega)$, we look for u such that

$$-\Delta u = f \text{ in } \Omega \text{ and } u = 0 \text{ on } \partial\Omega. \tag{2.2}$$

Given a triangulation of Ω, we define the degree $r \geq 1$ of the local polynomials, and we define V_h as the space of functions over Ω that are piecewise polynomials of degree r on each triangle (that can be discontinuous). We define the space V_h as

$$V_h = \{v_h \in L^1_{loc}(\Omega) : v_h|_K \in P_r(K) \text{ for all } K \in \mathcal{T}_h\}.$$

For $v_h \in V_h$, we have

$$\int_\Omega -\Delta u v_h dx = \sum_{K\in\mathcal{T}_h} \Big(\int_K \nabla u \cdot \nabla v_h dx - \int_{\partial K} v_h \nabla u \cdot \boldsymbol{n}_K ds \Big)$$
$$= (\nabla u, \nabla v_h)_{\mathcal{T}_h} - \langle \{\{\nabla u\}\}, [\![v_h]\!]\rangle_{\mathcal{E}_h} - \langle [\![\nabla u]\!]_s, \{\{v_h\}\}\rangle_{\mathcal{E}_h^{\mathrm{i}}}.$$

where we have used Theorem 2.6 in the last step. Note that $[\![\nabla u]\!]_s = 0$ for the solution u, and therefore

$$(-\Delta u, v_h)_{\mathcal{T}_h} = (\nabla u, \nabla v_h)_{\mathcal{T}_h} - \langle \{\{\nabla u\}\}, [\![v_h]\!]\rangle_{\mathcal{E}_h} \quad \forall v_h \in V_h, \tag{2.3}$$

For every function $v_h \in V_h$, for $\delta \in \{-1, 1, 0\}$, and $\alpha_{sta} > 0$ (this value will allow you to "control" the bilinear form in the proof that it is coercive), we define the discrete problem as follows:
Find $u_h \in V_h$ such that

$$a_\delta(u_h.v_h) = (f, v_h)_{\mathcal{T}_h}, \forall v_h \in V_h, \tag{2.4}$$

where

$$a_\delta(u_h, v_h) := (\nabla u_h, \nabla v_h)_{\mathcal{T}_h} - \langle \{\{\nabla u_h\}\}, [\![v_h]\!]\rangle_{\mathcal{E}_h}$$
$$+\delta\langle \{\{\nabla v_h\}\}, [\![u_h]\!]\rangle_{\mathcal{E}_h} + \frac{\alpha_{sta}}{h}\langle [\![u_h]\!], [\![v_h]\!]\rangle_{\mathcal{E}_h}. \tag{2.5}$$

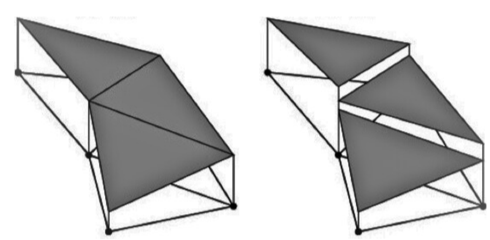


Figure 2.3: Continuous vs discontinuous Galerkin's method

Remark 2.18. Equation (2.3) is not a coercive bilinear form. We then add two extra terms to this expansion; one term will transform (2.3) into a coercive bilinear form and the other term will give us some flexibility to be able to define three different bilinear forms: a symmetric form, a stable form for all α, or a simpler expression ([Brezzi and Marini, 2014]). The resulting expression (2.5) is our final bilinear form.

Remark 2.19. Taking into consideration the variational Equation (2.3), then (2.4) is consistent. Thus, the convergence of these methods depends on their stability, which involves boundedness and coercivity of the bilinear form $a_\delta(\cdot,\cdot)$; see [Brenner and Scott, 2008, page 291].

Remark 2.20. Note that the elements of V_h are piecewise polynomials of degree at most r on Ω and can be discontinuous. This situation produces that

$$V_h \not\subset H_0^1(\Omega) =: V.$$

We must therefore define the bilinear form on $V_h + V$. It will be enough to prove the properties of coercivity and continuity on the space V_h; see [Brenner and Scott, 2008, Chapter 10, page 271].

Theorem 2.21. *The bilinear form a_δ is continuos and coercive on V_h with respect to the DG-norm; this is, there exists M, $\alpha^* > 0$ such that*

$$|a_\delta(u_h, v_h)| \leq M\|u_h\|_{DG}\|v_h\|_{DG} \quad \forall u_h, v_h \in V_h$$

and

$$a_\delta(v_h, v_h) \geq \alpha^*\|v_h\|_{DG}^2 \quad \forall v_h \in V_h.$$

Proof. We first proof the continuity property. For $u_h, v_h \in V_h$,

$$|\{\{\nabla u_h\}\}| \leq \frac{|\nabla u_{K_i}| + |\nabla u_{K_j}|}{2}$$

on each edge. Applying Hölder's inequality, first for integrals, and then for the sum

$$\begin{aligned}
\langle [\![v_h]\!], \{\{\nabla u_h\}\}\rangle_{\mathcal{E}_h} &\leq \sum_{e\in\mathcal{E}_h} \int_e |[\![v_h]\!]||\{\{\nabla u_h\}\}|ds \\
&\leq \sum_{e\in\mathcal{E}_h} \Big(\int_e [\![v_h]\!]^2 ds\Big)^{\frac{1}{2}} \Big(\int_e \{\{\nabla u_h\}\}^2 ds\Big)^{\frac{1}{2}} \\
&\leq \Big(\sum_{e\in\mathcal{E}_h} \frac{1}{h}\|[\![v_h]\!]\|_{0,e}^2\Big)^{\frac{1}{2}} \Big(\sum_{e\in\mathcal{E}_h} h\|\{\{\nabla u_h\}\}\|_{0,e}^2\Big)^{\frac{1}{2}}
\end{aligned}$$

$$= \Big(\sum_{e\in\mathcal{E}_h} \frac{1}{h}\|[\![v_h]\!]\|_{0,e}^2\Big)^{\frac{1}{2}} \Big(\sum_{e\in\mathcal{E}_h^{\mathfrak{i}}} h\|\{\{\nabla u_h\}\}\|_{0,e}^2 + \sum_{e\in\mathcal{E}_h^{\mathfrak{b}}} h\|\nabla u_h\|_{0,e}^2\Big)^{\frac{1}{2}}. \tag{2.6}$$

Now, adding the following expression over interior edges, in the second factor of (2.6)

$$\begin{aligned}\sum_{e\in\mathcal{E}_h^{\mathfrak{i}}} h\int_e \{\{\nabla u_h\}\}^2 ds &\leq \sum_{e\in\mathcal{E}_h^{\mathfrak{i}}} h\int_e \frac{(|\nabla u_{K_i}|+|\nabla u_{K_j}|)^2}{4}ds\\ &\leq \sum_{e\in\mathcal{E}_h^{\mathfrak{i}}} h\int_e (|\nabla u_{K_i}|^2+|\nabla u_{K_j}|^2)ds.\end{aligned}$$

Adding the contributions of the boundary edges in the last expression, and using Lemma 2.14, we deduce that

$$\begin{aligned}\sum_{K\in\mathcal{T}_h} h\|\nabla u_h\|^2_{0,\partial K} &= \sum_{e\in\mathcal{E}_h^{\mathfrak{i}}} h\|\nabla u_{K_i}\|^2_{0,e} + h\|\nabla u_{K_j}\|^2_{0,e} + \sum_{e\in\mathcal{E}_h^{\mathfrak{b}}} h\|\nabla u_h\|^2_{0,e}\\ &\leq C\sum_{K\in\mathcal{T}_h} h(h^{-1}\|\nabla u_h\|^2_{0,K} + h|\nabla u_h|^2_{1,K})\\ &= C\sum_{K\in\mathcal{T}_h} \|\nabla u_h\|^2_{0,K} + h^2|\nabla u_h|^2_{1,K}.\end{aligned} \tag{2.7}$$

Finally, combining (2.6) and (2.7) we get

$$\langle [\![v_h]\!], \{\{\nabla u_h\}\}\rangle_{\mathcal{E}_h} \leq \Big(\sum_{e\in\mathcal{E}_h} \frac{1}{h}\|[\![v_h]\!],\|^2_{0,e}\Big)^{\frac{1}{2}}\Big(C\sum_{k\in\mathcal{T}_h}\|\nabla u_h\|^2_{0,k} + h^2|\nabla u_h|^2_{1,k}\Big)^{\frac{1}{2}}. \tag{2.8}$$

In case that u_h is a piecewise polynomial of degree at most r, we can rewrite the expression on the right of (2.8) in such a way that we can "eliminate" the term from the seminorm $|\nabla u_h|^2_{1,k}$ and the factor h^2 then, for $s=1$ and $k=0$ in Lemma 2.15, we get

$$|\nabla u_h|^2_{1,k} \leq \hat{C}h^{-2}\|\nabla u_h\|^2_{0,k}.$$

Applying this to the second factor in (2.8)

$$\begin{aligned}\|\nabla u_h\|^2_{0,k} + h^2|\nabla u_h|^2_{1,k} &\leq \|\nabla u_h\|^2_{0,k} + \hat{C}\|\nabla u_h\|^2_{0,k}\\ &= (1+\hat{C})||\nabla u_h||^2_{0,k},\end{aligned}$$

so that, retaking (2.8), we get

$$\langle [\![v_h]\!], \{\{\nabla u_h\}\}\rangle_{\mathcal{E}_h} \leq C_r\Big(\sum_{e\in\mathcal{E}_h}\frac{1}{h}\|[\![v_h]\!]\|_{0,e}^2\Big)^{\frac{1}{2}}\Big(\sum_{k\in\mathcal{T}_h}\|\nabla u_h\|_{0,k}^2\Big)^{\frac{1}{2}}, \tag{2.9}$$

and inequalities (2.8) and (2.9) become, respectively,

$$\begin{aligned}\langle [\![v_h]\!], \{\{\nabla u_h\}\}\rangle_{\mathcal{E}_h} &\leq C\|v_h\|_j\Big[\Big(\sum_{k\in\mathcal{T}_h}|\nabla u_h|_{0,k}^2\Big)^{\frac{1}{2}} + \Big(\sum_{k\in\mathcal{T}_h}h^2|\nabla u_h|_{1,k}^2\Big)^{\frac{1}{2}}\Big]\\ &= C\|v_h\|_j\Big(|u_h|_{1,h} + h^2|u_h|_{2,h}\Big)\\ &\leq C\|v_h\|_{DG}\|u_h\|_{DG},\end{aligned}$$

and

$$\langle [\![v_h]\!], \{\{\nabla u_h\}\}\rangle_{\mathcal{E}_h} \leq C_r\|v_h\|_j|u_h|_{1,h} \leq C_r\|v_h\|_{DG}\|u_h\|_{DG}. \tag{2.10}$$

Furthermore, in a simpler way it is concluded that

$$\begin{aligned}\langle [\![u_h]\!], [\![v_h]\!]\rangle_{\mathcal{E}_h} &\leq \sum_{e\in\mathcal{E}_h}\int_e |[\![u_h]\!]||[\![v_h]\!]|ds\\ &\leq \sum_{e\in\mathcal{E}_h}\Big(\int_e [\![u_h]\!]^2 ds\Big)^{\frac{1}{2}}\Big(\int_e [\![v_h]\!]^2 ds\Big)^{\frac{1}{2}}\\ &\leq \Big(\sum_{e\in\mathcal{E}_h}\int_e [\![u_h]\!]^2 ds\Big)^{\frac{1}{2}}\Big(\sum_{k\in\mathcal{T}_h}\int_e [\![v_h]\!]^2 ds\Big)^{\frac{1}{2}}\\ &= \sqrt{h}\|u_h\|_j\sqrt{h}\|u_h\|_j\\ &\leq h\|u_h\|_{DG}\|v_h\|_{DG}.\end{aligned}$$

Finally,

$$\begin{aligned}(\nabla u_h, \nabla v_h) &\leq \sum_{K\in\mathcal{T}_h}\int_K |\nabla u_h||\nabla v_h|ds\\ &\leq \sum_{K\in\mathcal{T}_h}\Big(\int_K \nabla v_h^2 ds\Big)^{\frac{1}{2}}\Big(\int_K \nabla u_h^2 ds\Big)^{\frac{1}{2}}\\ &\leq \Big(\sum_{K\in\mathcal{T}_h}\int_K \nabla v_h^2 ds\Big)^{\frac{1}{2}}\Big(\sum_{K\in\mathcal{T}_h}\int_K \nabla u_h^2 ds\Big)^{\frac{1}{2}}\end{aligned}$$

$$
\begin{aligned}
&= |\nabla u_h|_{0,h} |\nabla v_h|_{0,h} \\
&\leq ||u_h||_{DG} ||v_h||_{DG}.
\end{aligned}
$$

In summary, recalling that, for u_h and v_h in V_h, we have

$$
\begin{aligned}
|a_\delta(u_h, v_h)| \leq & \left|(\nabla v_h, \nabla v_h)_{\mathcal{T}_h}\right| + \left|\langle \{\{\nabla u_h\}\}, [\![v_h]\!]\rangle_{\mathcal{E}_h}\right| \\
& + \left|\delta \langle \{\{\nabla v_h\}\}, [\![u_h]\!]\rangle_{\mathcal{E}_h}\right| + \left|\frac{\alpha_{sta}}{h} \langle [\![u_h]\!], [\![v_h]\!]\rangle_{\mathcal{E}_h}\right| \\
& \leq \|u_h\|_{DG} \|v_h\|_{DG} + C\|u_h\|_{DG}\|v_h\|_{DG} + C\delta \|u_h\|_{DG}\|v_h\|_{DG} \\
& + \frac{\alpha_{sta}}{h} h \|u_h\|_{DG} \|v_h\|_{DG} \\
& \leq (1 + C + \delta C + \alpha_{sta}) \|u_h\|_{DG} \|v_h\|_{DG}.
\end{aligned}
$$

As the second part of the proof, we now prove the coercivity of the bilinear form. Remember that in the space V_h, applying Theorem 2.16, we get

$$
\|v_h\|_{DG}^2 = \|v_h\|_j^2 + |v_h|_{1,h}^2 + h^2 \|\nabla v_h\|_{1,h}^2 \simeq \|v_h\|_j^2 + |v_h|_{1,h}^2.
$$

Moreover,

$$
a_\delta(v_h, v_h) = (\nabla v_h, \nabla v_h)_{\mathcal{T}_h} + (\delta - 1)\langle \{\{\nabla v_h\}\}, [\![v_h]\!]\rangle_{\mathcal{E}_h} + \frac{\alpha_{sta}}{h} \langle [\![v_h]\!], [\![v_h]\!]\rangle_{\mathcal{E}_h},
$$

and since the following expressions can be rewritten as

$$
\begin{aligned}
(\nabla v_h, \nabla v_h)_{\mathcal{T}_h} &= \sum_{K \in \mathcal{T}_h} \int_K \nabla v_h^2 dx = \sum_{K \in \mathcal{T}_h} |\nabla v_h|_{0,K}^2 = |\nabla v_h|_{0,h}^2 = |v|_{1,h}^2, \\
\langle [\![v_h]\!], [\![v_h]\!]\rangle_{\mathcal{E}_h} &= \sum_{e \in \mathcal{E}_h} \int_e [\![v_h]\!]^2 ds = \sum_{e \in \mathcal{E}_h} \|[\![v_h]\!]\|_{0,e}^2 = h \|v_h\|_j^2,
\end{aligned}
$$

we get

$$
a_\delta(v_h, v_h) = |v_h|_{1,h}^2 + (\delta - 1)\langle \{\{\nabla v_h\}\}, [\![v_h]\!]\rangle_{\mathcal{E}_h} + \alpha_{sta} \|v_h\|_j^2. \tag{2.11}
$$

We recall that for a given a quadratic form $x^2 + \alpha_{sta} y^2 - 2\beta xy$, the associated matrix

$$
B = \begin{pmatrix} 1 & -\beta \\ -\beta & \alpha_{sta} \end{pmatrix}
$$

is positive definite if and only if $\alpha_{sta} > \beta^2$, since the roots of the characteristic polynomial

$$\det(B - \lambda I) = \lambda^2 - (1 + \alpha_{sta})\lambda - \beta^2 + \alpha_{sta},$$

are given by

$$\lambda_1 = \frac{1 + \alpha_s + \sqrt{(1 - \alpha_{sta})^2 + 4\beta^2}}{2} > 0, \quad \lambda_2 = \frac{1 + \alpha_{sta} - \sqrt{(1 - \alpha_{sta})^2 + 4\beta^2}}{2}.$$

It holds that λ_2 is positive if and only if $(1+\alpha_{sta})^2 > (1-\alpha_{sta})^2 + 4\beta^2$ if and only if $\alpha_{sta} > \beta^2$. Then, for β fixed, it holds that

$$x^2 + \alpha_{sta} y^2 - 2\beta xy \geq \alpha^*(x^2 + y^2), \tag{2.12}$$

for some constant $\alpha^* > 0$, whenever α_{sta} is big enough.

It holds that

$$(\delta - 1)\langle \{\{\nabla v_h\}\}, [\![v_h]\!]\rangle_{\mathcal{E}_h} \geq -|\delta - 1| C |v_h|_{1,h} \|v_h\|_j, \tag{2.13}$$

for the three following cases:

1. if $\delta = 1$ it is clear.

2. if $\delta = 0$, we have that

   $$-\langle \{\{\nabla v_h\}\}, [\![v_h]\!]\rangle_{\mathcal{E}_h} \geq -C|v_h|_{1,h}\|v_h\|_j \Leftrightarrow \langle \{\{\nabla v_h\}\}, [\![v_h]\!]\rangle_{\mathcal{E}_h} \leq C_r |v_h|_{1,h}\|v_h\|_j,$$

   which is a consequence of (2.10), for piecewise polynomials of degree at most r, with $C = C_r$.

3. if $\delta = -1$, we use the same argument to deduce that

   $$-2\langle \{\{\nabla v_h\}\}, [\![v_h]\!]\rangle_{\mathcal{E}_h} \geq -2C|v_h|_{1,h}\|v_h\|_j.$$

Getting back to (2.11), and from (2.13), we obtain that

$$\begin{aligned} a_\delta(v_h, v_h) &\geq |v_h|_{1,h}^2 + \alpha_{sta}\|v_h\|_j^2 - |\delta - 1|C|v_h|_{1,h}||v_h||_j \\ &\geq \alpha^*(|v_h|_{1,h}^2 + \|v_h\|_j^2) \\ &= \alpha^*\|v_h\|_{DG}^2, \end{aligned}$$

for some constant $\alpha^* > 0$, whenever α_{sta} is big enough. Note that the inequality of the quadratic form (2.12) is applied for $\beta = \dfrac{|\delta - 1|C}{2}$.

$\square$

2.2.1 Convergence estimates and Galerkin orthogonality

For the sake of completeness, we include results related to *a-priory* estimates, and refer to [F. Brezzi, 2014] for more details. First we prove a Galerkin orthogonality property, which esentially means that the solution u_h is the orthogonal projection onto the space V_h, with respect to the bilinear form. Hence, it is the best approximation w.r.t. the finite dimensional space. In the second theorem, we give a bound for the approximation error in terms of the error of approximating u with any element of V_h. Next, we prove a bound for the error of approximating u with any approximation in V_h, in terms of h and the half-norm of u. Finally, we finish this subsection by bounding the error of approximating with u_h, using the DG norm, in terms of h and the sum of the seminorms of u.

Theorem 2.22. ***(Galerkin Orthogonality)*** *Let u be the exact solution for the continuous problem (2.3) and u_h the solution for the discrete problem (2.4). It holds that*

$$a_\delta(u - u_h, v_h) = 0 \ \forall v_h \in V_h.$$

Proof. Note that, when u is the exact solution we have, for all $v_h \in V_h$:

$$a_\delta(u, v_h) = (\nabla u, \nabla v_h)_{\mathcal{T}_h} - \langle \{\{\nabla u\}\}, [\![v_h]\!] \rangle_{\mathcal{E}_h} = (f, v_h).$$

Hence, if u_h solves $a_\delta(u_h, v_h) = (f, v_h)_{\mathcal{T}_h}$ for all $v_h \in V_h$, we get the desired result. $\square$

Theorem 2.23. *Let u be the exact solution for the continuous problem (2.3) and u_h the solution for the discrete problem (2.4). Then,*

$$\|u - u_h\|_{DG} \leq \left(1 + \frac{M}{\alpha^*}\right) \|u - u_I\|_{DG},$$

where u_I is any aproximation of u in V_h.

Proof. Setting $\delta_h := u_h - u_I$ and using Galerkin Orthogonality Theorem 2.22, we have

$$\alpha^* \|\delta_h\|^2_{DG} \leq a_\delta(\delta_h, \delta_h)$$

$$
\begin{aligned}
&= a_\delta(u_h - u_I, \delta_h) \\
&= a_\delta(u_h - u + u - u_I, \delta_h) \\
&= a_\delta(u - u_I, \delta_h) \\
&\leq M\|u - u_I\|_{DG}\|\delta_h\|_{DG},
\end{aligned}
$$

where M is the continuity constant.Thus

$$\|\delta_h\|_{DG} \leq \frac{M}{\alpha^*}\|u - u_I\|_{DG}.$$

Finally

$$\|u - u_h\|_{DG} \leq \|u - u_I\|_{DG} + \|\delta_h\|_{DG} \leq \left(1 + \frac{M}{\alpha^*}\right)\|u - u_I\|_{DG}.$$

□

Assumption 2.24. We assume that u_I is an approximation of u in V_h with the property: There exists an integer r (the degree of the local polynomials) and a constant C such that

$$|u_I - u|_{s,K} \leq Ch^{r+1-s}|u|_{r+1,K}, \tag{2.14}$$

for all integers s with $0 \leq s \leq r$, for all element $K \in \mathcal{T}_h$.

Theorem 2.25. *Let u be the exact solution for the continuous problem (2.3). It holds that*

$$\|u_I - u\|_{DG}^2 \leq Ch^{2r} \sum_{K \in \mathcal{T}_h} |u|_{r+1,K}^2,$$

where $u_I \in V_h$ is any approximation of u with the Assumption 2.24.

Proof. Using Assumption 2.24, we bound first the jump norm:

$$
\begin{aligned}
\|u_I - u\|_j^2 &= \sum_{e \in \mathcal{E}_h} \frac{1}{h}\|[\![u_I - u]\!]\|_{0,e}^2 \\
&= \sum_{e \in \mathcal{E}_h^i} \frac{1}{h}\int_e [\![u_I - u]\!]^2 ds + \sum_{e \in \mathcal{E}_h^b} \frac{1}{h}\int_e (u_I - u)^2\|\boldsymbol{n}\|^2 ds \\
&= \sum_{e \in \mathcal{E}_h^i} \frac{1}{h}\int_e [(u_{I_i} - u_i)\boldsymbol{n}_i + (u_{I_j} - u_j)\boldsymbol{n}_j]^2 ds + \sum_{e \in \mathcal{E}_h^b} \frac{1}{h}\int_e (u_I - u)^2 ds
\end{aligned}
$$

$$
\begin{aligned}
&\le 2\sum_{e\in\mathcal{E}_h^i}\frac{1}{h}\int_e (u_{I_i}-u_i)^2\|\boldsymbol{n}_i\|^2+(u_{I_j}-u_j)^2\|\boldsymbol{n}_j\|^2 ds+\sum_{e\in\mathcal{E}_h^b}\frac{1}{h}\int_e (u_I-u)^2 ds\\
&= 2\sum_{e\in\mathcal{E}_h^i}\frac{1}{h}\int_e (u_{I_i}-u_i)^2 ds+2\sum_{e\in\mathcal{E}_h^i}\frac{1}{h}\int_e (u_{I_j}-u_j)^2 ds+\sum_{e\in\mathcal{E}_h^b}\frac{1}{h}\int_e (u_I-u)^2 ds\\
&= 2\sum_{e\in\mathcal{E}_h^i}\frac{1}{h}\int_e (u_{I_i}-u_i)^2 ds+\sum_{K\in\mathcal{T}_h}\sum_{e\in\partial K}\frac{1}{h}\int_e (u_I-u)^2 ds\\
&\le 2\sum_{K\in\mathcal{T}_h}\sum_{e\in\partial K}\frac{1}{h}\int_e (u_I-u)^2 ds\\
&= 2\sum_{K\in\mathcal{T}_h}\sum_{e\in\partial K}\frac{1}{h}\|u_I-u\|_{0,e}^2\\
&= 2\sum_{K\in\mathcal{T}_h}\frac{1}{h}\|u_I-u\|_{0,\partial K}^2\\
&\le C_1\sum_{K\in\mathcal{T}_h}\left(h^{-2}|u_I-u|_{0,K}^2+|u_I-u|_{1,K}^2\right)\quad\text{(Lemma 2.14)}\\
&\le C_1\sum_{K\in\mathcal{T}_h}\left(h^{-2}C^2h^{2r+2}|u|_{r+1,K}^2+C^2h^{2r}|u|_{r+1,K}^2\right)\quad\text{(applying Assumption 2.24 twice)}\\
&= 2C_1C^2h^{2r}\sum_{K\in\mathcal{T}_h}|u|_{r+1,K}^2.
\end{aligned}
$$

Now, taking $v=u_I-u$, we bound the second part of the DG norm:

$$
\begin{aligned}
|\nabla v|_{0,h}^2+h^2|\nabla v|_{1,h}^2 &= \sum_{K\in\mathcal{T}_h}\left(|\nabla(u_I-u)|_{0,K}^2+h^2|\nabla(u_I-u)|_{1,K}^2\right)\\
&\le \sum_{K\in\mathcal{T}_h}\left(|u_I-u|_{1,K}^2+2h^2|u_I-u|_{2,K}^2\right)\\
&\le \sum_{K\in\mathcal{T}_h}\left(C^2h^{2r}|u|_{r+1,K}^2+2h^2C^2h^{2r-2}|u|_{r+1,K}^2\right)\\
&= 3C^2h^{2r}\sum_{K\in\mathcal{T}_h}|u|_{r+1,K}^2.
\end{aligned}
$$

We conclude that under Assumption 2.24 we have

$$
\|u_I-u\|_{DG}^2=\|u_I-u\|_j^2+\sum_{K\in\mathcal{T}_h}\left(|\nabla(u_I-u)|_{0,K}^2+h^2|\nabla(u_I-u)|_{1,K}^2\right)
$$

$$
\begin{aligned}
&\leq (2C_1+3)C^2h^{2r}\sum_{K\in\mathcal{T}_h}|u|^2_{r+1,K}\\
&= \tilde{C}h^{2r}\sum_{K\in\mathcal{T}_h}|u|^2_{r+1,K}.
\end{aligned}
$$

$\square$

Combining Theorems 2.23 and 2.25, we have the final result:

Theorem 2.26. *Let u be the exact solution for the continuous Problem (2.3) and $u_h \in V_h$ the solution for the discrete Problem (2.4). Then*

$$\|u-u_h\|^2_{DG} \leq C\left(1+\frac{M}{\alpha^*}\right)^2 h^{2r}\sum_{K\in\mathcal{T}_h}|u|^2_{r+1,K}.$$

2.2.2 Linear system

As the next step in our analysis, we describe some considerations related to the linear system that is obtained from the bilinear form. We will state the theorem that guarantees existence of a unique solution of the discrete problem. Finally, we include some considerations for solving the corresponding linear system. We remark that we focus on the case $\delta = -1$ (and we will omit to write it) that generates a symmetric bilinear form.

We consider a finite dimensional Hilbert space V_h (the space of functions that are piecewise polynomials, of degree r over Ω and that can be discontinuous from one triangle to another). Given a positive definite bilinear form,

$$a(\cdot,\cdot) : V_h \times V_h \longrightarrow \mathbb{R},$$

and an element $f \in V_h'$ and, we consider the problem of finding $u_h \in V_h$, such that

$$a(u_h, v_h) = f(v_h),\ v_h \in V_h, \tag{2.15}$$

and let $V_h = span\{\psi_1,\ldots,\psi_N\}$. Problem (2.15) is equivalent to

$$a(u_h,\psi_j) = (f,\psi_j),\ \forall j \in \{1,\ldots,N\}. \tag{2.16}$$

Since $u_h \in V_h$, we can write $u_h = \sum_{i=1}^{N} u_i \psi_i$ and substituting in (2.16), we have

$$a\left(\sum_i u_i \psi_i, \psi_j\right) = (f, \psi_j),\ \forall j \in \{1, \dots, N\},$$

which can be rewritten as

$$\sum_i u_i a\big(\psi_i, \psi_j\big) = (f, \psi_j),\ \forall j \in \{1, \dots, N\}.$$

Let

$$b = \begin{pmatrix} (f, \psi_1) \\ (f, \psi_2) \\ \vdots \\ (f, \psi_N) \end{pmatrix},\ u \longleftrightarrow \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix} \text{ and } a_{ji} = a(\psi_i, \psi_j),$$

and A denotes the stiffness matrix:

$$A = (a_{ij})_{N \times N} \text{ with } a_{ij} = a(\psi_j, \psi_i).$$

Problem (2.15) is equivalent to the linear system

$$Au = b. \tag{2.17}$$

It holds that

$$z^T A z = \sum_{i,k} z_i z_k a_{ik} = a\left(\sum_k z_k \psi_k, \sum_i z_i \psi_i\right) = a(z, z) \geq \alpha^* \|z\|_{V_h}^2.$$

If z is different from zero, then

$$z^T A z > 0,$$

and A is a positive definite matrix.

Theorem 2.27. *It A is a symmetric positive definite matrix, then (2.17) has a unique solution.*

Proof. It is enough to prove that the matrix A is invertible. By contradiction, if

A is not invertible, there exists $x \neq 0$ such $Ax = 0$, so $x^T Ax = 0$ for $x \neq 0$, which is a contradiction. $\square$

Remark 2.28. We recall two difficulties related to solving the linear system $Au = b$ obtained in (2.17):

1. When using direct solvers, it is well know that the exact solution u and the approximated solution $\widetilde{u}$ satisfy the bound

$$\|u - \widetilde{u}\| \leq C\epsilon_m \kappa_2(A),$$

   for some constant $C > 0$, where $\epsilon_m \approx 10^{-16}$, $\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2 \sim \frac{1}{h^2}$. Therefore, for large values of κ_2 we might have loss of significant digits in the solution.

2. Execution time: if we apply direct Gaussian elimination to solve the linear system, it requires $\mathcal{O}(N^3)$ operations for a dense matrix, and even though the matrix is sparse, on each step we might lose the sparsity of the matrix. Hence, problems are solved by using iterative methods or specific methods for factoring sparse matrices. We will focus on preconditioned iterative methods, where we compute a sequence of vectors

$$u_0, u_1, u_2, \ldots, u_k \to u = A \backslash b,$$

   that will converge (in a few steps) to the exact solution of the linear system. The iterative method that we will use to deal with this problem is a two-level overlapping Schwarz method.

3 — Two-level overlapping Schwarz methods

This part of the thesis is based on the work of [Toselli and Widlund, 2004], specifically Chapters 1, 2 and 3. In this section, we describe the theoretical framework of Schwarz methods. Given spaces of local functions and a coarse space of functions, the goal is to define extension operators from these spaces to the DG space, in such a way that the base space can be decomposed as the sum of the extensions of the local spaces and the extension of the coarse space. Such decomposition requires certain hypothesis; see Assumptions 3.5, 3.6. We define a two-level overlapping Schwarz method by using the local stiffness matrices, from which we define the Schwarz operators from certain projection operators. From here, and taking two assumptions, we find bounds for the condition number of the additive operator.

3.1 Abstract theory

In this section, we follow Chapter 2 [Toselli and Widlund, 2004]. Given $S \in \mathbb{N}^*$ and a finite dimensional space $V_h = \langle \psi_n \rangle_{n=1}^N$, with $N = \dim V_h$, we consider a family of spaces $\{V_i\}_{i=0}^S$, and interpolation and extension operators

$$R_i^T : V_i \longrightarrow V_h, \tag{3.1}$$

such that V_h admits the decomposition

$$V_h = R_0^T V_0 + \sum_{i=1}^S R_i^T V_i,$$

which is not necessarily a direct sum of subspaces; i.e., the representation of an element of V_h as sum of terms of components in V_i is not unique. The spaces V_i do not need to be subspaces of V_h, but we refer to them as 'subspaces' or 'local spaces' for $i \in \{1, \ldots, S\}$, and 'coarse space' for V_0.

We next introduce local positive definite bilinear form on the subspaces,

$$a_i(\cdot, \cdot) : V_i \times V_i \longrightarrow \mathbb{R},\, i \in \{0, \ldots, S\}.$$

For this, we use the original bilinear form and the extension operators. This is referred as 'exact local solvers':

$$a_i(u_i, v_i) = a(R_i^T u_i, R_i^T v_i) \quad \text{ for } u_i, v_i \in V_i. \tag{3.2}$$

Let $N_i = \dim V_i$, for $i \in \{1, \ldots, S\}$. There is a family of functions $\varphi_i : \{1, \ldots, N_i\} \rightarrow \{1, \ldots, N\}$ such that $\{\psi_{\varphi_i(k)}\}_{k=1}^{N_i}$ is a basis of V_i. The associated local stiffness matrices are then given by

$$A_i = (a_{lk})_{N_i \times N_i} \text{ , with } a_{lk} = a_i\left(\psi_{\varphi_i(k)}, \psi_{\varphi_i(l)}\right) \;\; (l, k \in \{1, \ldots, N_i\}).$$

Remark 3.1. One of the characteristics of a finite-dimensional function space is that we can define a isomorphism $V_h \cong \mathbb{R}^N$ defined by

$$v = \sum_{l=1}^{N} v_l \psi_l \longleftrightarrow \mathbf{v} = (v_1, \ldots, v_N)^T.$$

The same happens with the spaces V_i and the extensions. We therefore denote with v the function and its vector of coordinates.

Lemma 3.2. *With the above notations, we have that*

$$A_i = R_i A R_i^T,$$

where $R_i = (R_i^T)^T$.

Proof. Since $R_i^T u_i, R_i^T v_i \in V_h$, we can write

$$R_i^T u_i = \sum_{k=1}^{N} (R_i^T u_i)_k \psi_k \quad \text{and} \quad R_i^T v_i = \sum_{l=1}^{N} (R_i^T v_i)_l \psi_l,$$

where $(R_i^T u_i)_k$ denotes the k-th entry of $R_i^T u_i$. Thus,

$$a\left(\sum_{k=1}^{N} (R_i^T u_i)_k \psi_k, \sum_{l=1}^{N} (R_i^T v_i)_l \psi_l\right) = \sum_{l=1}^{N} \sum_{k=1}^{N} (R_i^T u_i)_k (R_i^T v_i)_l a\big(\psi_k, \psi_l\big)$$

$$= \big((R_i^T v_i)_1 \quad \cdots \quad (R_i^T v_i)_N\big)^T \begin{pmatrix} a\big(\psi_1,\psi_1\big) & \cdots & a\big(\psi_N,\psi_1\big) \\ \vdots & & \vdots \\ a\big(\psi_1,\psi_N\big) & \cdots & a\big(\psi_N,\psi_N\big) \end{pmatrix} \begin{pmatrix} (R_i^T u_i)_1 \\ \vdots \\ (R_i^T u_i)_N \end{pmatrix}. \tag{3.3}$$

Operators R_i^T are matrices of size $N \times N_i$, and it holds that

$$\big((R_i^T v_i)_1 \quad \cdots \quad (R_i^T v_i)_N\big)^T = \big(R_i^T v_i\big)^T = v_i^T (R_i^T)^T.$$

Since $a_{lk} = a(\psi_k, \psi_l)$, from (3.3) we get that

$$a(R_i^T u_i, R_i^T v_i) = v_i^T R_i A R_i^T u_i. \tag{3.4}$$

Secondly, since $u_i, v_i \in V_i$, we can write

$$u_i = \sum_{k=1}^{N_i} (u_i)_k \psi_{\varphi_i(k)} \quad \text{and} \quad v_i = \sum_{l=1}^{N_i} (v_i)_l \psi_{\varphi_i(l)},$$

and therefore

$$a_i\left(\sum_{k=1}^{N_i} (u_i)_k \psi_{\varphi_i(k)}, \sum_{l=1}^{N_i} (v_i)_l \psi_{\varphi_i(l)}\right) = \sum_{l=1}^{N_i}\sum_{k=1}^{N_i} (u_i)_k (v_i)_l a_i\big(\psi_{\varphi_i(k)}, \psi_{\varphi_i(l)}\big)$$
$$= \big((v_i)_1 \quad \cdots \quad (v_i)_{N_i}\big)^T \begin{pmatrix} a_i\big(\psi_{\varphi_i(1)},\psi_{\varphi_i(1)}\big) & \cdots & a_i\big(\psi_{\varphi_i(N_i)},\psi_{\varphi_i(1)}\big) \\ \vdots & & \vdots \\ a_i\big(\psi_{\varphi_i(1)},\psi_{\varphi_i(N_i)}\big) & \cdots & a_i\big(\psi_{\varphi_i(N_i)},\psi_{\varphi_i(N_i)}\big) \end{pmatrix} \begin{pmatrix} (u_i)_1 \\ \vdots \\ (u_i)_{N_i} \end{pmatrix},$$

which is the same as

$$a_i(u_i, v_i) = v_i^T A_i u_i. \tag{3.5}$$

From (3.2), (3.4) and (3.5), we conclude that

$$A_i = R_i A R_i^T.$$

□

Schwarz operators are defined in terms of projection-like operators

$$P_i = R_i^T \widetilde{P}_i : V_h \longrightarrow R_i^T V_i \subset V_h,\ i = 0, \ldots, S, \tag{3.6}$$

where $\widetilde{P}_i : V_h \longrightarrow V_i$ is defined by

$$a_i(\widetilde{P}_i u, v_i) = a(u, R_i^T v_i),\ v_i \in V_i. \tag{3.7}$$

We note that $\widetilde{P}_i$ is well defined since the local bilinear forms are coercive. Schwarz operators have very interesting and useful properties.

Lemma 3.3 (Projection)**.** *The operators P_i can be written as*

$$P_i = R_i^T A_i^{-1} R_i A,\ 0 \leq i \leq S.$$

In addition, the operators P_i are selfadjoint with respect to the scalar product induced by $a(\cdot,\cdot)$ and positive semi-definite. Moreover, P_i is a projection, i.e., $P_i^2 = P_i$; see [Toselli and Widlund, 2004, Chapter 2, page 37].

Proof. First, consider the operator $\widetilde{P}_i$ defined in (3.7). Let $u \in V_h$, $\widetilde{P}_i u \in V_i$ and $v_i \in V_i$, $i \in \{1, \ldots, S\}$. We then write

$$\widetilde{P}_i u = \sum_{k=1}^{N_i} (\widetilde{P}_i u)_k \psi_{\varphi_i(k)}, \quad v_i = \sum_{k=1}^{N_i} (v_i)_k \psi_{\varphi_i(k)} \quad \text{and} \quad u = \sum_{k=1}^{N} (u)_k \psi_k. \tag{3.8}$$

Substituting (3.8) in (3.7), we obtain that

$$\begin{aligned}
\sum_{l=1}^{N_i} \sum_{k=1}^{N_i} (\widetilde{P}_i u)_k (v_i)_l a_i(\psi_{\varphi_i(k)}, \psi_{\varphi_i(l)}) &= \sum_{l=1}^{N} \sum_{k=1}^{N} (u)_k (R_i^T v_i)_l a(\psi_k, \psi_l) \\
= \begin{pmatrix} (v_i)_1 & \cdots & (v_i)_{N_i} \end{pmatrix}^T &\begin{pmatrix} a_i(\psi_{\varphi_i(1)}, \psi_{\varphi_i(1)}) & \cdots & a_i(\psi_{\varphi_i(N_i)}, \psi_{\varphi_i(1)}) \\ \vdots & & \vdots \\ a_i(\psi_{\varphi_i(1)}, \psi_{\varphi_i(N_i)}) & \cdots & a_i(\psi_{\varphi_i(N_i)}, \psi_{\varphi_i(N_i)}) \end{pmatrix} \begin{pmatrix} (\widetilde{P} u)_1 \\ \vdots \\ (\widetilde{P} u)_{N_i} \end{pmatrix} \\
= \begin{pmatrix} (R_i^T v_i)_1 & \cdots & (R_i^T v_i)_N \end{pmatrix}^T &\begin{pmatrix} a(\psi_1, \psi_1) & \cdots & a(\psi_N, \psi_1) \\ \vdots & & \vdots \\ a(\psi_1, \psi_N) & \cdots & a(\psi_N, \psi_N) \end{pmatrix} \begin{pmatrix} (u)_1 \\ \vdots \\ (u)_N \end{pmatrix}.
\end{aligned} \tag{3.9}$$

Notice that

$$\begin{pmatrix} (R_i^T v_i)_1 & \cdots & (R_i^T v_i)_N \end{pmatrix}^T = \left(R_i^T v_i\right)^T = v_i^T (R_i^T)^T,$$

and (3.9) becomes

$$v_i^T A_i \widetilde{P}_i u = v_i^T (R_i^T)^T A u, \quad \forall u \in V_h \text{ and } \forall v_i \in V_i.$$

We can then conclude that

$$A_i \widetilde{P}_i = R_i A.$$

Finally, from (3.6) we deduce that

$$P_i = R_i^T \widetilde{P}_i = R_i^T A_i^{-1} R_i A. \tag{3.10}$$

To prove that P_i is selfadjoint, let $u, v \in V_h$. Using (3.10), it holds that

$$\begin{aligned} a(P_i u, v) &= v^T A (P_i u) \\ &= v^T A (R_i^T A_i^{-1} R_i A u) \\ &= (v^T A R_i^T A_i^{-1} R_i) A u \\ &= ((A R_i^T A_i^{-1} R_i)^T v)^T A u \\ &= (R_i^T A_i^{-1} R_i A v)^T A u \\ &= (P_i v)^T A u \\ &= a(u, P_i v). \end{aligned}$$

The positive semi-definiteness of P_i is a consequence of the coercivity of the bilinear form:

$$a(P_i u, u) = u^T A P_i u = u^T A R_i^T A_i^{-1} R_i A u = w_i^T A_i^{-1} w_i \geq 0,$$

with $w_i = R_i A u$. Finally, P_i is a projection:

$$\begin{aligned} P_i^2 &= P_i P_i \\ &= R_i^T A_i^{-1} R_i A R_i^T A_i^{-1} R_i A \\ &= R_i^T A_i^{-1} A_i A_i^{-1} R_i A \\ &= R_i^T A_i^{-1} R_i A \\ &= P_i. \end{aligned}$$

□

We note that we can always generate the appropriate right hand side for the resulting operator equation since we can compute $P_i u = R_i^T \widetilde{P}_i u$ when u is the finite element solution, by noting that

$$a_i(\widetilde{P}_i u, v_i) = a(u, R_i^T v_i) = f(R_i^T v_i).$$

Similarly, we can find $P_j P_i u$, once $P_i u$ has been computed. We will work with additive preconditioned operators of the form

$$P_{ad} = \sum_{i=0}^{S} P_i = \sum_{i=0}^{S} R_i^T A_i^{-1} R_i A.$$

We end this section by noting that the Schwarz operator introduced here is a preconditioned operator for the original operator A and can be written as the product of a suitable preconditioner and A, where the former only involves extensions $\{R_i^T\}$, restrictions $\{R_i\}$, local operators $\{A_i^{-1}\}$, and A. This is straightforward for the additive method, where

$$P_{ad} = A_{ad}^{-1} A, \quad A_{ad}^{-1} = \sum_{i=0}^{S} R_i^T A_i^{-1} R_i. \tag{3.11}$$

3.2 Convergence Theory

In this section, we present the necessary results that allows us to find a bound for the condition number of the preconditioned linear system. We consider two assumptions, which will allow us to demonstrate that the additive operator is invertible, and to obtain an upper bound for the preconditioned system. We consider then the conjugate gradient algorithm [Braess, 2007, Chapter 3, page 201] for the solution of

$$P_{ad} u = g_{ad},$$

with $g_{ad} := A_{ad}^{-1} f$ and $P_{ad} := A_{ad}^{-1} A$. We show how to estimate the condition number of P_{ad}

$$\kappa(P_{ad}) = \frac{\lambda_{max}(P_{ad})}{\lambda_{min}(P_{ad})},$$

with the objective to obtain

$$\kappa(P_{ad}) = \kappa(A_{ad}^{-1} A) \ll \kappa(A).$$

Remark 3.4. In the case of continuous Galerkin method, the approximation error depends on the condition number, which is why the new condition number of the preconditioned system is intended to be much smaller.

Assumption 3.5. (Stable decomposition) There exists a constant C_0, such that every $u \in V_h$ admits a decomposition

$$u = \sum_{i=0}^{S} R_i^T u_i = R_0^T u_0 + \sum_{i=1}^{S} R_i^T u_i,\ u_i \in V_i,$$

that satisfies

$$\sum_{i=0}^{S} a_i(u_i, u_i) \leq C_0^2 a(u, u).$$

Assumption 3.6. (Strengthened Cauchy-Schwarz Inequalities) There exists constants $0 \leq \epsilon_{ij} \leq 1$ for every $i, j \in \{1, \ldots, S\}$, such that

$$|a(R_i^T u_i, R_j^T u_j)| \leq \epsilon_{ij} a(R_i^T u_i, R_i^T u_i)^{\frac{1}{2}}\ a(R_j^T u_j, R_j^T u_j)^{\frac{1}{2}},$$

for $u_i \in V_i$ and $u_j \in V_j$. We denote the spectral radius of the matrix E with entries $E_{ij} = (\epsilon_{ij})$ by $\rho(E)$.

Remark 3.7. (Local Stability) Tipically, it is required that there exists a constant ω such that

$$a(R_i^T u_i, R_i^T u_i) \leq \omega a_i(u_i, u_i),\ u_i \in \text{ range } (\widetilde{P}_i) \subset V_i,\ 0 \leq i \leq S.$$

In our case, since we are using exact solvers, this assumption holds for $\omega = 1$. Remark 3.7 ensures that the local bilinear forms are coercive and gives a one-sided measure of their approximation properties. We note that Assumption 3.5 ensures a weak type of continuity for the $a_i(\cdot, \cdot)$.

Lemma 3.8. *[Invertible] Let Assumption 3.5 be satisfied. Then,*

$$a(P_{ad} u, u) \geq C_0^{-2} a(u, u)\ \forall u \in V_h, \tag{3.12}$$

and consequently P_{ad} is invertible. In addition,

$$a(P_{ad}^{-1} u, u) = \min_{\substack{u_i \in V_i \\ u = \sum R_i^T u_i}} \sum_{i=0}^{S} a_i(u_i, u_i); \tag{3.13}$$

see [Toselli and Widlund, 2004, Chapter 2, page 42].

Proof. By Assumption 3.5 and the definition of the operators P_j, we have that

$$\begin{aligned} a(u,u) &= \sum_{i=0}^{S} a(u, R_i^T u_i) \\ &= \sum_{i=0}^{S} a_i(\widetilde{P}_i u, u_i) \\ &\leq \left(\sum_{i=0}^{S} a_i(\widetilde{P}_i u, \widetilde{P}_i u) \right)^{\frac{1}{2}} \cdot \left(\sum_{i=0}^{S} a_i(u_i, u_i) \right)^{\frac{1}{2}} \\ &\leq \left(\sum_{i=0}^{S} a_i(\widetilde{P}_i u, \widetilde{P}_i u) \right)^{\frac{1}{2}} \cdot C_0 (a(u,u))^{\frac{1}{2}}. \end{aligned}$$

Here we used the fact that the bilinear form a_i is defined by means of the original bilinear form, so it is defined by means of integrals, which, applied for functions in V_i, satisfies the Cauchy-Schwarz inequality; and we used the fact that the local bilinear forms are symmetric and positive definite. Squaring, canceling a common factor and using the definition of the $\{\widetilde{P}_i\}$, we find

$$a(u,u) \leq C_0^2 \sum_{i=0}^{S} a_i(\widetilde{P}_i u, \widetilde{P}_i u) = C_0^2 \sum_{i=0}^{S} a(u, R_i^T \widetilde{P}_i u) = C_0^2 a(u, P_{ad} u),$$

which proves (3.12) and thus the invertibility of P_{ad}. Now, because P_{ad} is invertible, we can choose

$$u_i = \widetilde{P}_i P_{ad}^{-1} u, \; 0 \leq i \leq S, \tag{3.14}$$

which implies that

$$u = \sum_{i=0}^{S} R_i^T u_i,$$

and

$$\sum_{i=0}^{S} a_i(u_i, u_i) = \sum_{i=0}^{S} a_i(\widetilde{P}_i P_{ad}^{-1} u, \widetilde{P}_i P_{ad}^{-1} u) = \sum_{i=0}^{S} a(P_{ad}^{-1} u, R_i^T \widetilde{P}_i P_{ad}^{-1} u)$$

$$= a\left(P_{ad}^{-1}u, \sum_{i=0}^{S} R_i^T \widetilde{P}_i P_{ad}^{-1}u\right) = a\left(P_{ad}^{-1}u, \Big(\sum_{i=0}^{S} P_i\Big) P_{ad}^{-1}u\right)$$
$$= a(P_{ad}^{-1}u, P_{ad}P_{ad}^{-1}u) = a(P_{ad}^{-1}u, u).$$

Consider now, any decomposition

$$u = \sum_{i=0}^{S} R_i^T u_i,\ v_i \in V_i.$$

Then

$$\begin{aligned}
a(P_{ad}^{-1}u, u) &= \sum_{i=0}^{S} a(P_{ad}^{-1}u, R_i^T u_i) = \sum_{i=0}^{S} a_i(\widetilde{P}_i P_{ad}^{-1}u, u_i) \\
&\leq \left(\sum_{i=0}^{S} a_i(\widetilde{P}_i P_{ad}^{-1}u, \widetilde{P}_i P_{ad}^{-1}u)\right)^{\frac{1}{2}} \left(\sum_{i=0}^{S} a_i(u_i, u_i)\right)^{\frac{1}{2}} \\
&= \left(\sum_{i=0}^{S} a(P_{ad}^{-1}u, R_i^T \widetilde{P}_i P_{ad}^{-1}u)\right)^{\frac{1}{2}} \left(\sum_{i=0}^{S} a_i(u_i, u_i)\right)^{\frac{1}{2}} \\
&= a(P_{ad}^{-1}u, u)^{\frac{1}{2}} \left(\sum_{i=0}^{S} a_i(u_i, u_i)\right)^{\frac{1}{2}}.
\end{aligned}$$

Then we have that, for every decomposition of u,

$$a(P_{ad}^{-1}u, u) \leq \sum_{i=0}^{S} a_i(u_i, u_i).$$

Since the choice of the decomposition (3.14) gives us the minimum, we conclude the proof. □

Lemma 3.9. *Let Assumption 3.6 be satisfied. Then, for $i = 0, \ldots, S$,*

$$\|P_i\|_a \leq 1.$$

In addition,

$$a(P_{ad}u, u) \leq (\rho(\varepsilon) + 1)a(u, u),$$

where $\|\cdot\|_a$ is the usual norm associated to the bilinear form $a(\cdot,\cdot)$; see [Toselli and Widlund, 2004, Chapter 2, page 43].

Proof. It holds that

$$\begin{aligned} a(P_i u, P_i u) &= a(R_i^T \widetilde{P}_i u, R_i^T \widetilde{P}_i u) \le a_i(\widetilde{P}_i u, \widetilde{P}_i u) \\ &= a(u, R_i^T \widetilde{P}_i u) = a(u, P_i u); \end{aligned}$$

i.e,

$$a(P_i u, P_i u) \le a(u, P_i u),\ u \in V_h.$$

After applying the process again, using the fact that the bilinear form is symmetric, we get

$$a(P_i u, P_i u) \le a(u, u),\ u \in V_h,$$

and therefore

$$\|P_i u\|_a = \sqrt{a(P_i u, P_i u)} \le \sqrt{a(u,u)} = \|u\|_a.$$

In order to prove the second inequality, we first consider the operator

$$\hat{P} = \sum_{i=1}^{S} P_i.$$

Recall that, if ε is a symmetric matrix, then

$$\|\varepsilon\|_2 = \rho(\varepsilon^t \varepsilon)^{\frac{1}{2}} = \rho(\varepsilon),$$

and for all $v \neq 0$, we have

$$\frac{v^t \varepsilon v}{v^t v} \le \|\varepsilon\|_2. \tag{3.15}$$

From Assumption 3.6 and using (3.15), we deduce that

$$\begin{aligned} a(\hat{P}u, \hat{P}u) &= \sum_{1 \le i,j \le S} a(P_i u, P_j u) \\ &\le \sum_{1 \le i,j \le S} a(R_i^T \widetilde{P}_i u, R_j^T \widetilde{P}_j u) \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{1\leq i,j\leq S} \epsilon_{ij} a(P_i u, P_i u)^{\frac{1}{2}} a(P_j u, P_j u)^{\frac{1}{2}} \\
&\leq \sum_{1\leq i,j\leq S} \epsilon_{ij} a(u, P_i u)^{\frac{1}{2}} a(u, P_j u)^{\frac{1}{2}} \\
&\leq \rho(\varepsilon) \sum_{1\leq i\leq S} a(u, P_i u) \quad (3.15) \\
&= \rho(\varepsilon) a(u, \hat{P} u) \\
&\leq \rho(\varepsilon) a(u,u)^{\frac{1}{2}} a(\hat{P}u, \hat{P}u)^{\frac{1}{2}}.
\end{aligned}
$$

Thus,

$$
a(\hat{P}u, u) \leq a(\hat{P}u, \hat{P}u)^{\frac{1}{2}} a(u,u)^{\frac{1}{2}} \leq \rho(\varepsilon) a(u,u). \tag{3.16}
$$

Now, we use (3.16) and the bound for $\|P_0\|_a$, to conclude that

$$
\begin{aligned}
a(P_{ad}u, u) &= a(P_0 u, u) + a(\hat{P}u, u) \\
&\leq a(P_0 u, P_0 u)^{\frac{1}{2}} a(u,u)^{\frac{1}{2}} + \rho(\varepsilon) a(u,u) \\
&\leq a(u,u)^{\frac{1}{2}} a(u,u)^{\frac{1}{2}} + \rho(\varepsilon) a(u,u) \\
&= (\rho(\varepsilon) + 1) a(u,u).
\end{aligned}
$$

□

Theorem 3.10. *Let Assumptions 3.5 and 3.6 be satisfied. Then, the condition number of the additive Schwarz operador satisfies*

$$
\kappa(P_{ad}) \leq C_0^2 \left(1 + \rho(\varepsilon)\right). \tag{3.17}
$$

Proof. Recall that

$$
\kappa(P_{ad}) = \frac{\lambda_{max}(P_{ad})}{\lambda_{min}(P_{ad})},
$$

where

$$
\lambda_{max}(P_{ad}) = \sup_{u\in V} \frac{a(P_{ad}u, u)}{a(u,u)}, \quad \lambda_{min}(P_{ad}) = \inf_{u\in V} \frac{a(P_{ad}u, u)}{a(u,u)}.
$$

The result follows from Lemmas 3.8 and 3.9, since

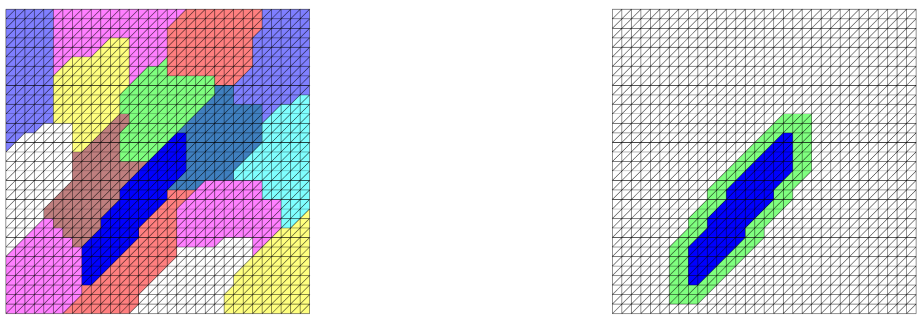
$$
\frac{a(P_{ad}u, u)}{a(u,u)} \leq \rho(\varepsilon) + 1, \quad \frac{a(P_{ad}u, u)}{a(u,u)} \geq C_0^{-2},\ u \in V.
$$

$\square$

3.3 Local Solvers

To finish this chapter, we present the basic design of local solvers, typically used in the overlapping Schwarz methods.

We consider the partition of Ω into non-overlapping subdomains $\{\Omega_i : 1 \leq i \leq S\}$, with diameter H_i; see Figure 3.1(a). The idea is to extend the regions $\{\Omega_i\}$ into larger regions $\Omega'_i \supset \Omega_i$, by adding new layers of elements of the fine triangulation; see Figure 3.1(b).



(a) (b)

Figure 3.1: (a) A triangulation with $S = 16$ irregular subdomains. (b) A subdomain Ω_i (in blue) and layers that are considered for defining Ω'_i (in green). The set Ω'_i is the union of blue and green elements.

Assumption 3.11. Given $i \in \{1, \ldots, S\}$, there exists $\delta_i > 0$ such that, if x belongs to Ω'_i, then

$$dist(x, \partial\Omega'_j \backslash \partial\Omega) \geq \delta_i,$$

for j that depends on x, with $x \in \Omega'_j$. The maximum of the ratios H_i/δ_i is denoted by

$$\frac{H}{\delta} = \max_{1 \leq i \leq S} \left\{ \frac{H_i}{\delta_i} \right\};$$

see [Toselli and Widlund, 2004, Section 3.2, page 56].

Remark 3.12. Roughly speaking, the values δ_i measure the width of the $\Omega'_i \backslash \Omega_i$ regions.

We also require the following assumption.

Assumption 3.13. (Finite Covering) The partion $\{\Omega'_i\}$ can be colored using at most N^c colors, in such a way that subregions with the same color are disjoint. It is possible to show that if $x \in \Omega$, then it belongs to at most N^c subdomains in $\{\Omega'_i\}$; see [Toselli and Widlund, 2004, Section 3.2, page 56].

Lemma 3.14. *Let $\{\Omega'_i\}$ be an overlapping partition satisfying Assumption 3.11 and Assumption 3.13. Then, there exists a family of functions $\{\tilde{\theta}_i : 1 \leq i \leq S\} \subset W^{1,\infty}(\Omega)$, such that*

$$
\begin{aligned}
&0 \leq \tilde{\theta}_i(x) \leq 1 \quad , x \in \overline{\Omega}, \\
&supp(\tilde{\theta}_i) \subset \overline{\Omega'_i}, \\
&\sum_{i=1}^{S} \tilde{\theta}_i(x) = 1 \quad , x \in \overline{\Omega},
\end{aligned}
$$

and

$$
\|\nabla \tilde{\theta}_i\|_\infty \leq C/\delta_i \quad , 1 \leq i \leq S,
$$

where C is a constant independent of δ_i and H_i. We refer to $\{\tilde{\theta}_i\}$ as a partition of unity related to the subdomains $\{\Omega'_i\}$.

Proof. See [Toselli and Widlund, 2004, Lemma 3.4.]. □

Since we have to work on a finite dimensional space, we need to use a modified partition of unity, obtained by interpolation of the local functions of Lemma 3.14. We define

$$
\theta_i = I^h(\tilde{\theta}_i) \quad , 1 \leq i \leq S,
$$

where I^h is the nodal piecewise linear interpolant on the fine mesh; i.e., $I^h : L^1_{loc}(\overline{\Omega}) \longrightarrow V_h$ with

$$
I^h \tilde{\theta}_i = \sum_{k=1}^{N} \tilde{\theta}_i(x_k)\psi_k.
$$

Note that since

$$\|\nabla I^h(\tilde{\theta}_i)\|_\infty \leq \|\nabla \tilde{\theta}_i\|_\infty,$$

the family $\{\theta_i\}$ also satisfies Lemma 3.14.

4 An overlapping two-level Schwarz preconditioner

In this section, we describe the construction of a preconditioner tailored to the discretization derived in (3.11) through DGM. Our approach encompasses overlapping domains, wherein we establish the definition of local spaces across these overlapping regions. Subsequently, we introduce the zero extension operators to the global domain Ω.

On the other hand, we define the coarse space as the space generated by specific basis functions. When considering the intersection between two disjoint subdomains, with the extreme points as references, our objective is to replicate a linear behavior between them. However, these intersections may exhibit irregular shapes. To achieve this, we project the vertices of the fine mesh belonging to the intersection of two subdomains onto the unit vector connecting the endpoints of these intersections, as detailed in [Calvo, 2018]. It is essential to note that while these intersections encompass numerous vertices of the fine mesh, not all of them contribute to defining the basis functions on the coarse mesh. Only the values at the endpoints of the intersection of two subdomains are considered. This crucial point in defining the coarse space enables the removal of many vertices from the fine mesh, ensuring that the convergence of the method remains independent of the number of subdomains, as highlighted in [Toselli and Widlund, 2004]. Therefore, the definition of a coarse function relies on its nodal values at the vertices of the subdomains. To extend the boundary values to the interior nodes of each subdomain, we consider discrete harmonic extensions.

In the case of triangular elements, functions are linear, in particular they are harmonic. For the general case, we can solve a Laplace problem with given boundary values over each subdomain. We will also define the extension operator by interpolating the coarse functions onto the fine mesh.

Finally, we present our algorithm and main result, a bound for the condition number of the additive operator, which will be proportional to

$$(S^c + 1)\left(1 + \frac{H}{\delta}\right),$$

with $S^c \ll S$ and $\frac{H}{\delta} = \max_{i \in \{1,\ldots,S\}} \frac{H_i}{\delta_i}$, and we provide numerical results that support the conjecture.

4.1 Local spaces

We start by defining the local spaces for the preconditioner given in (3.11). Consider the partition of Ω into overlapping subdomains $\{\Omega_i' : 1 \leq i \leq S\}$, as we discussed in the previous section. We define the local space V_i as the space of functions that are piecewise-linear polynomials on Ω_i' that vanish on $\partial\Omega_i'$. We recall that a function $v_i \in V_i$ can be discontinuous; see Figure 4.2.

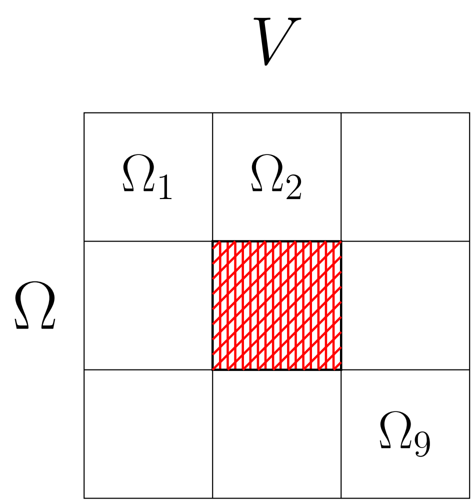


Figure 4.1: Non-overlapping subdomains. The function space V is defined on a triangulation $\mathcal{T}_h$ of Ω.

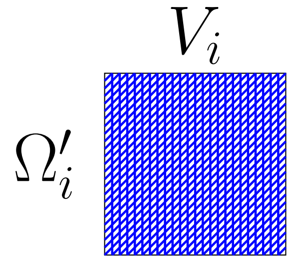


Figure 4.2: Specific subdomain. The function space V_i is defined on Ω_i', with zero boundary values.

Let $N_i = \dim(V_i)$, with $\dim(V_i) \ll \dim(V_h)$. There is a family of functions $\varphi_i : \{1, \ldots, N_i\} \to \{1, \ldots, N\}$ such that $\{\psi_{\varphi_i(k)}\}_{k=1}^{N_i}$ is a basis of V_i. For the local spaces $V_1, V_2, \ldots, V_S$, we define extensions by zero from Ω_i' to Ω; i.e., consider

$$R_i^T : V_i \longrightarrow V_h, (i \in \{1, \ldots, S\})$$

defined as follows: if $v_i \in V_i$,

$$(R_i^T v_i)(x) = \begin{cases} \displaystyle\sum_{k=1}^{N_i} v_i(x_{\varphi_i(k)})\psi_{\varphi_i(k)}(x) & \text{if} \quad x \in \Omega_i', \\ \\ 0 & \text{if} \quad \text{otherwise}, \end{cases}$$

4.2 Coarse space

We now define the space V_0. Define the **subdomain edge** associated to subdomains Ω_i and Ω_j as

$$\mathcal{E}^{ij} = (\overline{\Omega}_i \cap \overline{\Omega}_j)^\circ, \quad i, j \in \{1, \ldots, S\}.$$

The set of **subdomains vertices** is defined then as the of endpoints of all subdomain edges, excluding those on $\partial\Omega$. This is, nodes on subdomain edges belong to the boundary of two subdomains, and subdomain vertices belong to at least three subdomains. Let $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ be the end points of $\mathcal{E}^{ij}$, and let $\boldsymbol{d}_\mathcal{E}$ be the unit vector from $\boldsymbol{v}_i$ to $\boldsymbol{v}_j$. We want the function to be 1 in $\boldsymbol{v}_i$, 0 in $\boldsymbol{v}_j$, and have a linear variation between them. Of course the vertices that we want to evaluate, in general, do not have this linear behavior, so we will use their projection on the vector $\boldsymbol{d}_{\mathcal{E}}$, and the function would look like this:

$$c_{v_i}(\bar{\boldsymbol{x}}) = \begin{cases} 0 & \text{if} \quad \frac{(\bar{\boldsymbol{x}}-\boldsymbol{v}_i)\cdot \boldsymbol{d}_{\mathcal{E}}}{\|\boldsymbol{v}_j-\boldsymbol{v}_i\|} > 1, \\ 1 - \frac{(\bar{\boldsymbol{x}}-\boldsymbol{v}_i)\cdot \boldsymbol{d}_{\mathcal{E}}}{\|\boldsymbol{v}_j-\boldsymbol{v}_i\|} & \text{if} \quad 0 \leq (\bar{\boldsymbol{x}}-\boldsymbol{v}_i)\cdot \boldsymbol{d}_{\mathcal{E}} \leq \|\boldsymbol{v}_j-\boldsymbol{v}_i\|, \\ 1 & \text{if} \quad (\bar{\boldsymbol{x}}-\boldsymbol{v}_i)\cdot \boldsymbol{d}_{\mathcal{E}} < 0, \end{cases}$$

where $\bar{\boldsymbol{x}} \in \mathcal{E}^{ij}$ is a vertex of $\mathcal{T}_h$.

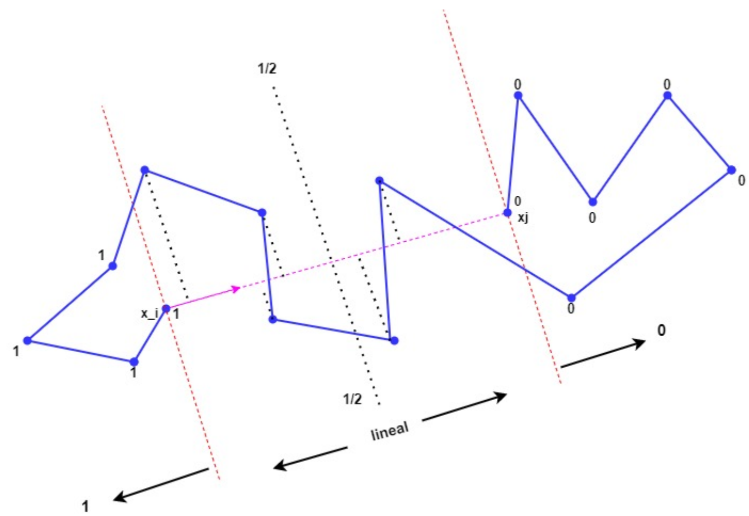


Figure 4.3: Basis functions for coarse space

In this way, using the Ω_i for $i = \{1, \ldots, S\}$, we can define a set of M vertices and the same number of basis functions. Now, we extend the values of c_v at $\partial\Omega_i$ to the interior vertices $\hat{x} \in \Omega_i$ of $\mathcal{T}_h$, with a discrete harmonic extension (note that linear functions are harmonic, so we can see this as the natural generalization). For each Ω_i we solve

$$\begin{cases} \Delta c_v = 0 & \text{on} \quad \Omega_i, \\ c_v = g & \text{on} \quad \partial\Omega_i. \end{cases}$$

In general, for all subdomain vertex v, we define its corresponding base function c_v, and finally the space V_0 is defined as

$$V_0 = \langle c_v \rangle.$$

Remember that, if $v \in V_h$, then $v(x) = \sum_{k=1}^{N} v(x_k)\psi_k(x)$, where $\{x_k\}$ denote the vertices of $\mathcal{T}_h$. Therefore we define

$$R_0^T : V_0 \longrightarrow V_h,$$

for $c \in V_0$, as

$$\begin{aligned}(R_0^T c)(x) &= \sum_{k=1}^{N} c(x_k)\psi_k(x)\\ &= \sum_{k=1}^{N}\sum_{j=1}^{M} c(y_j)c_{v_j}(x_k)\psi_k(x),\end{aligned}$$

where $\{y_j\}$ are the vertices associated with the mesh applied on Ω and V_0 (note that the y_j are some of the x_i, but we will call them differently to maintain an order).

4.3 Main result

We finally present the main result of this tesis. Once we have built the local spaces with their respective extension operators, the coarse space and it's extension operator. The selecting of this coarse space allows to reduce the aproximation error, to a point where it will not depend of the number of subdomains. Hence, we are ready to ennunciate the main results for a bound for the condition number of the additive operator, and to then corroborate the results with experiment in the final section. We will see some basic results in the basic structure of a two-level method with overlap.

Lemma 4.1. *Let $\rho(\varepsilon) = \{\epsilon_{ij}\}$ be given as in Assumption 3.13. Suppose that are at most S^c nonzeros in each row of ε. Then,*

$$\rho(\varepsilon) \leq S^c.$$

Proof.

$$\rho(\varepsilon) \leq \|\varepsilon\|_\infty = \max_{1\leq i\leq S}\sum_{j=1}^{S}|\epsilon_{ij}| = \max_{1\leq i\leq S}\sum_{j=1}^{S^c}|\epsilon_{ij}| \leq \max_{1\leq i\leq S}\sum_{j=1}^{S^c} 1 = S^c.$$

$\square$

Lemma 4.2. *There holds*

$$a(P_{ad}u,u) \le (S^c+1)a(u,u), \quad u \in V_h.$$

Proof. It is a straightforward consequence of Lemma 3.9 and Lemma 4.1, we have

$$a(P_{ad}u,u) \le (\rho(\varepsilon)+1)a(u,u) \le (S^c+1)a(u,u).$$

$\square$

The following result is work in process, but we have numerical evidence to support this claim.

Conjecture 4.3. *There exists a constant C, independent of h, H, and δ, such that*

$$a(P_{ad}^{-1}u,u) \le C\left(1+\frac{H}{\delta}\right)a(u,u), \quad u \in V_h.$$

Lemma 4.4. *If exact solvers are employed on all the subspaces, then there exists a constant C, independent of h, H, and δ, such that*

$$\kappa(P_{ad}) \le C\left(1+\frac{H}{\delta}\right). \tag{4.1}$$

Proof. For the finite-dimensional case we have

$$\lambda_{min}^{-1}(P_{ad}) = \lambda_{max}(P_{ad}^{-1}),$$

then, using Lemma 4.2 and Conjecture 4.3, we get

$$\begin{aligned}\kappa(P_{ad}) &= \frac{\lambda_{max}(P_{ad})}{\lambda_{min}(P_{ad})}\\ &= \lambda_{max}(P_{ad})\cdot\lambda_{max}(P_{ad}^{-1})\\ &\le C(S^c+1)\cdot\left(1+\frac{H}{\delta}\right).\end{aligned}$$

$\square$

4.4 Implementation details

We present relevant remarks on the construction of the stiffness matrix A which are not found tipically in literature; these `MATLAB` routines can be downloaded from [Calvo, 2024]. We separate the details for each component of the bilinear form (5.2), giving explicit formulas for the local matrices.

4.4.1 Integral over elements

For a triangle E with vertices $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\boldsymbol{v}_3$ and associated local basis functions ϕ_1, ϕ_2, ϕ_3, it is known that the local matrix associated to the term

$$\int_E \nabla\phi_i \cdot \nabla\phi_j, \quad i, j \in \{1, 2, 3\},$$

can be written exactly as

$$A_1^{loc} = \frac{1}{4|E|}\begin{pmatrix} \boldsymbol{y}'\boldsymbol{y} & \boldsymbol{y}'\boldsymbol{z} & \boldsymbol{y}'\boldsymbol{x} \\ \boldsymbol{z}'\boldsymbol{y} & \boldsymbol{z}'\boldsymbol{z} & \boldsymbol{z}'\boldsymbol{x} \\ \boldsymbol{x}'\boldsymbol{y} & \boldsymbol{x}'\boldsymbol{z} & \boldsymbol{x}'\boldsymbol{x} \end{pmatrix}, \tag{4.2}$$

where $\boldsymbol{x} = \boldsymbol{v}_2 - \boldsymbol{v}_1$, $\boldsymbol{y} = \boldsymbol{v}_3 - \boldsymbol{v}_2$, $\boldsymbol{z} = \boldsymbol{v}_1 - \boldsymbol{v}_3$; equation (4.2) follows from straightforward computations, given that ϕ_i, ϕ_j are linear functions with $\phi_i(\boldsymbol{x}_k) = \delta_{ik}$, $\phi_j(\boldsymbol{x}_k) = \delta_{jk}$ ($k \in \{1, 2, 3\}$). We will show results for the more general PDE

$$-div(\rho\nabla u) = f$$

in Ω. In this case, the contribution of an element E is given by the integral

$$\int_E \rho\nabla\phi_i \cdot \nabla\phi_j = \nabla\phi_i \cdot \nabla\phi_j \int_E \rho,$$

and we use a one-point quadrature rule to approximate $\int_E \rho \approx \rho(\boldsymbol{x}_b^E)|E|$, where $\boldsymbol{x}_b^E$ is the barycenter of E. Therefore, the matrix A_1^{loc} given in (4.2) has to be multiplied by $\rho(\boldsymbol{x}_b^E)$.

4.4.2 Integrals over edges

Consider now the integrals over the edges of the bilinear form. Given an edge e shared by elements E_1^e and E_2^e, let $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ be the endpoints of e, and $\boldsymbol{v}_3^+, \boldsymbol{v}_3^-$ the third vertex of E_1^e and E_2^e, respectively; see Figure 4.4, where we also present the order of the local ordering for the degrees of freedom. Define the vectors $\boldsymbol{x} = \boldsymbol{v}_2 - \boldsymbol{v}_1, \boldsymbol{y} = \boldsymbol{v}_3^+ - \boldsymbol{v}_2, \boldsymbol{z} = \boldsymbol{v}_1 - \boldsymbol{v}_3^+, \boldsymbol{a} = \boldsymbol{v}_3^- - \boldsymbol{v}_2, \boldsymbol{b} = \boldsymbol{v}_1 - \boldsymbol{v}_3^-$.

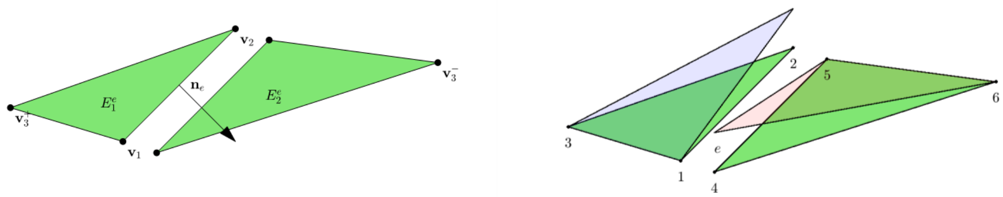


Figure 4.4: (left) An internal edge shared by triangles E_1^e and E_2^e, with vertices $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3^+$ and $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3^-$, respectively (right) Local index for degrees of freedom and two local basis functions ϕ_2, ϕ_4.

For the bilinear form $b_1(\cdot,\cdot)$ (see equation (5.2)) we need to compute

$$\frac{1}{2}\int_e ((\rho\nabla\phi_i)|_{E_1^e} + (\rho\nabla\phi_i)|_{E_2^e}) \cdot \boldsymbol{n}_e(\phi_j|_{E_1^e} - \phi_j|_{E_2^e})$$

for any internal edge e. Since $\nabla\phi_i$ is constant on each element, terms of the form $\int_e \rho\phi_j$ are approximated with a one-point quadrature, using the midpoint of e. Define

$$M = \begin{pmatrix} \dfrac{\boldsymbol{y}\cdot\boldsymbol{x}}{\boldsymbol{y}^\perp\cdot\boldsymbol{x}} & \dfrac{\boldsymbol{y}\cdot\boldsymbol{x}}{\boldsymbol{y}^\perp\cdot\boldsymbol{x}} & 0 \\ \dfrac{\boldsymbol{x}\cdot\boldsymbol{z}}{\boldsymbol{x}^\perp\cdot\boldsymbol{z}} & \dfrac{\boldsymbol{x}\cdot\boldsymbol{z}}{\boldsymbol{x}^\perp\cdot\boldsymbol{z}} & 0 \\ \dfrac{\boldsymbol{x}\cdot\boldsymbol{x}}{\boldsymbol{y}^\perp\cdot\boldsymbol{x}} & \dfrac{\boldsymbol{x}\cdot\boldsymbol{x}}{\boldsymbol{y}^\perp\cdot\boldsymbol{x}} & 0 \end{pmatrix}, N = \begin{pmatrix} \dfrac{\boldsymbol{x}\cdot\boldsymbol{a}}{\boldsymbol{x}^\perp\cdot\boldsymbol{a}} & \dfrac{\boldsymbol{x}\cdot\boldsymbol{a}}{\boldsymbol{x}^\perp\cdot\boldsymbol{a}} & 0 \\ \dfrac{\boldsymbol{b}\cdot\boldsymbol{x}}{\boldsymbol{b}^\perp\cdot\boldsymbol{x}} & \dfrac{\boldsymbol{b}\cdot\boldsymbol{x}}{\boldsymbol{b}^\perp\cdot\boldsymbol{x}} & 0 \\ \dfrac{\boldsymbol{x}\cdot\boldsymbol{x}}{\boldsymbol{x}^\perp\cdot\boldsymbol{a}} & \dfrac{\boldsymbol{x}\cdot\boldsymbol{x}}{\boldsymbol{x}^\perp\cdot\boldsymbol{a}} & 0 \end{pmatrix},$$

where $\boldsymbol{v}^\perp := (-v_2, v_1)$ for any vector $\boldsymbol{v} = (v_1, v_2)$. The local matrix associated to the bilinear form $b_1(\cdot,\cdot)$ for the edge e can be approximated in block form as

$$B_{e,1}^{loc} = \frac{1}{4}\begin{bmatrix} \rho^+ M & -\rho^+ M \\ -\rho^- N & \rho^- N \end{bmatrix},$$

where $\rho^+ = \rho|_{E_1^e}$, $\rho^- = \rho|_{E_2^e}$ are the evaluations of ρ at the midpoint of e. For the last bilinear form $b_2(\cdot,\cdot)$, (see equation (5.2)) we need to compute

$$\frac{\alpha_{sta}}{|e|}\int_e (\phi_i|_{K^+} - \phi_i|_{K^-}) \cdot (\phi_j|_{K^+} - \phi_j|_{K^-}),$$

for any interior edge e. The associated local matrix for an interior edge e can be

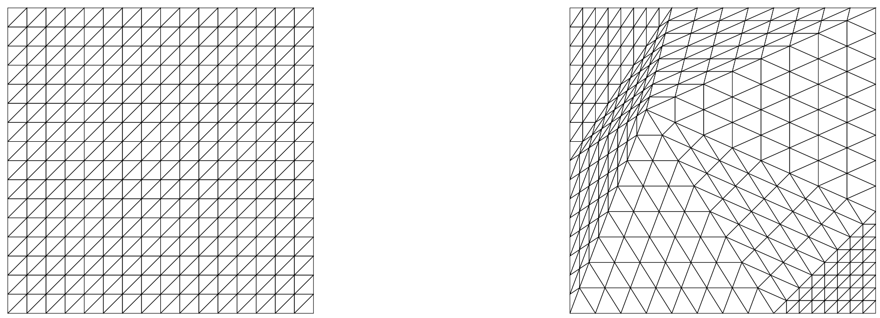


Figure 4.5: (left) Structured and (right) unstructured triangular mesh.

computed exactly as

$$B_{e,2}^{loc} = \frac{\alpha_{stb}}{6}\begin{bmatrix} 2 & 1 & 0 & -2 & -1 & 0 \\ 1 & 2 & 0 & -1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & -1 & 0 & 2 & 1 & 0 \\ -1 & -2 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

If we denote by A_1, B_1, B_2 the global matrices obtained by assembling the local matrices $A_1^{loc}, B_{e,1}^{loc}, B_{e,2}^{loc}$, respectively, then the global stiffness matrix A is given by

$$A = A_1 - B_1 - B_1^T + B_2.$$

4.4.3 Convergence of DGM

For the sake of completeness, we first confirm the convergence of our DGM implementation for structured and unstructured triangular meshes as shown in Figure 4.5. We use $\rho(x_1, x_2) = 1$ and $\rho(x_1, x_2) = x_1x_2 + 1$ for the exact solution $u(x_1, x_2) = \sin(\pi x_1)\sin(\pi x_2)$. We confirm convergence of order h^2 as shown in Figure 4.6.

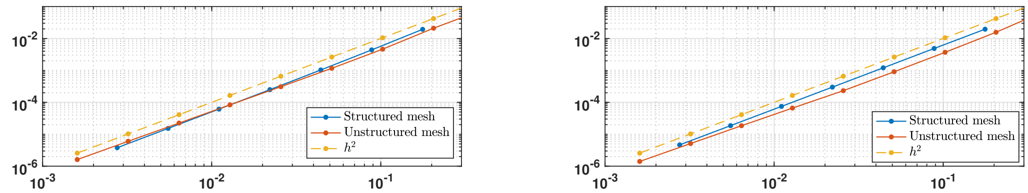


Figure 4.6: Inf error $\|u - u_h\|_\infty$ as a function of h for DGM, for a structured and unstructured triangular mesh as shown in Figure 4.5, for (left) $\rho(x_1, x_2) = 1$ and (right) $\rho(x_1, x_2) = 1 + x_1 x_2$.

4.5 Numerical results

In this section we discuss some implementation details and present numerical results for our two-level overlapping Schwarz preconditioner for discontinuous Galerkin methods with $\Omega = [0, 1]^2$. We present numerical examples to support the bound that we assume as Conjecture 4.3 and Lemma 4.4. We compare the behavior using square ('ic') subdomains and irregular ('metis') subdomains; see [Karypis and Kumar, 1998] and Figure (4.7). We incorporate the data corresponding to the generalization of Poisson´s equation (see [Calvo, 2018])

$$-\nabla(\rho(x)\nabla u(x)) = f(x), \tag{4.3}$$

where $\rho : \Omega \longrightarrow \mathbb{R}^+$ is a given function and we include experiments for two different choices of ρ: first a piecewise constant function on each subdomain and a new experiment with a channel distribution. We present some graphs to see the behavior of our approximations, as well as results with different partitions of the subdomains. We estimate the condition number κ, compute the number of iterations I and the dimension of the coarse space M for each experiment. We solve the corresponding preconditioned linear system with the Conjugate Gradient method with a relative tolerance of 10^{-6}; see Figure 4.8 for a numerical solution with $h = 1/16$. We verify that our algorithm is scalable (bound is independent of the number of subdomains) and confirm the linear growth in the condition number as we increase H/δ.

4.5.1 Square and METIS subdomains with $\rho = 1$

We first present experiments with square (regular) subdomains and measure the effect of irregular (METIS) subdomains; see Figure 4.7.

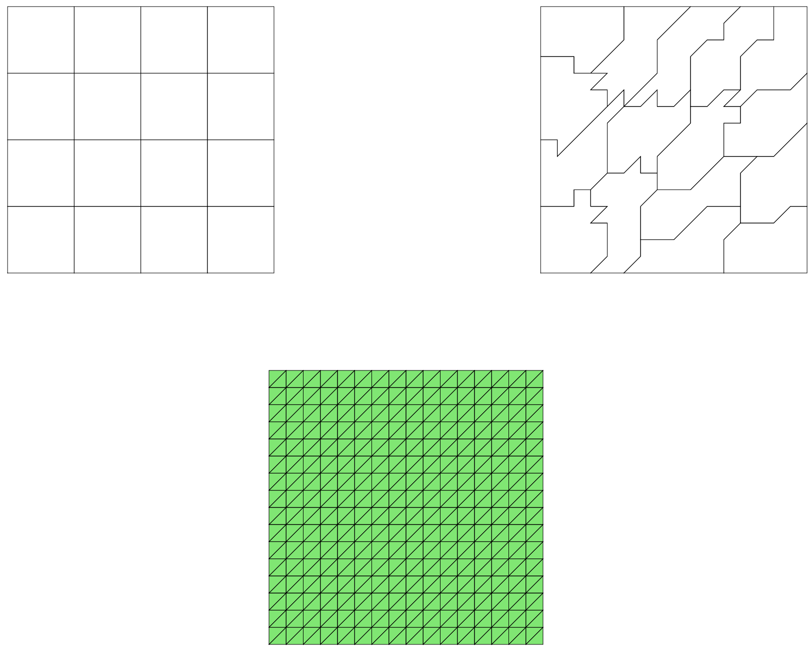


Figure 4.7: (top) Two coarse meshes: (left) a square and a (right) METIS decomposition, both for the (bottom) triangulation with $h = 1/16$.

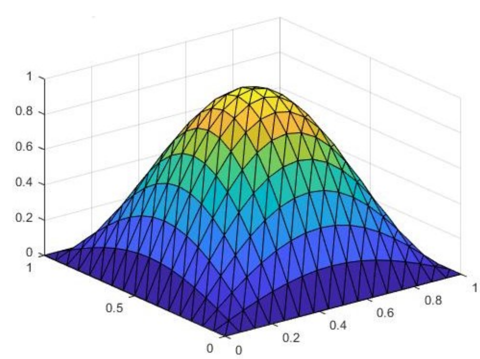


Figure 4.8: Numerical solution with $h = 1/16$.

1. We verify the dependence on the number of subdomains S with minimum overlap ($\delta/h = 1$) and generous overlap ($\delta/h = 4$) with $H/h = 16$; see results in Table 4.1. For square subdomains it is clear that κ remains bounded, and in the presence of irregular subdomains, we require a generous overlap in

order to obtain robust results.

2. We verify that our bound is independent of H/h (roughly speaking, the number of degrees of freedom per subdomain); see Table 4.2. We observe no dependence on this parameter, where we fix $S = 36$ and $\delta/h = 4$, and therefore, the dimension M of V_0 is fixed.

3. We verify the linear growth of our bound as a function of $1+H/\delta$; see Figure 4.9. It is clear that in the case of METIS subdomains, the condition number increases due to the irregularity of the subdomains.

We note that when we increase the number of layers to create the overlapping subdomains Ω_i', the condition number decreases but numerically we increase the size of the local matrices $A_i (i \in \{1, \ldots, S\})$. Hence, a balance has to be established to determine the number of layers. In the case of irregular subdomains, the geometry of the subdomains deteriorates the constant C in our main result. However, results maintain a moderate growth.

	Square subdomains		METIS subdomains	
	$I(\kappa)$		$I(\kappa)$	
S	$H/h = 16$, $\delta/h = 1$	M	$H/h = 16$, $\delta/h = 1$	M
64	20(14.9)	49	37(20.3)	97
144	20(14.1)	121	40(25.0)	243
256	20(13.5)	225	40(24.7)	447
400	18(13.8)	361	48(38.6)	711
S	$H/h = 16$, $\delta/h = 4$	M	$H/h = 16$, $\delta/h = 4$	M
64	15(6.2)	49	25(9.3)	97
144	14(6.0)	121	26(9.8)	243
256	14(5.8)	225	26(11.0)	447
400	14(5.7)	361	29(13.2)	711

Table 4.1: Condition number (κ) for the preconditioned system $A_{ad}^{-1}A$ and number of PGC iterations (I) as a function of the number of square and irregular subdomains N; M is the coarse space dimension and $H/h = 16$. The case $H/\delta = 16$ corresponds to the minimal overlap case $\delta = h$.

	Square subdomains		METIS subdomains	
	$I(\kappa)$		$I(\kappa)$	
H/h	$S = 36,\ \delta/h = 4$	N_V	$S = 36,\ \delta/h = 4$	N_V
8	15(5.3)	25	24(10.9)	48
16	15(5.8)	25	25(10.8)	52
32	16(6.3)	25	25(10.8)	49
64	17(6.4)	25	26(10.7)	48

Table 4.2: Condition number (κ) for the preconditioned system $A_{ad}^{-1}A$ and number of PCG iterations (I) as a function of H/h for square and irregular subdomains, $N = 36$ and $H/\delta = 4$; M is the coarse space dimension.

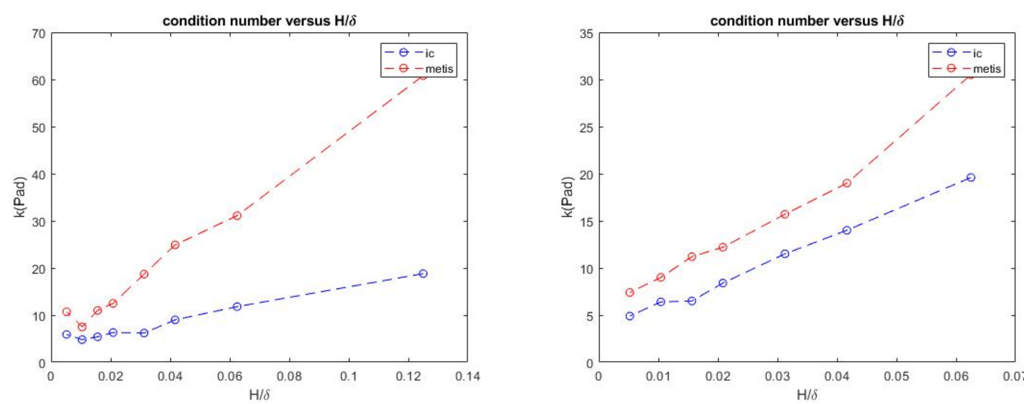


Figure 4.9: Condition number for regular and irregular subdomains as a function of H/δ, with (left) $H/h = 24$ and (right) $H/h = 48$.

4.5.2 Experiments with discontinuous values of $\rho = 1 + xy$

We now consider problem (4.3) with different values for the coefficient ρ. We first use a piecewise constant function over each subdomain, and second we consider a *channel* distribution, as shown in Figure 4.10. We remark that in the second case, these channels intersect several subdomains. We refer to these two cases as piecewise discontinuous coefficient and discontinuous channels, respectively. These two functions ρ are used in the presence of regular and irregular subdomains. Results are shown in Table 4.3, Table 4.4 and Table 4.5. Compared to the case $\rho = 1$, we observe a expected increment in the number of iterations and condition number estimates. We recall that bounds typically depend on the contrast, defined

	Square subdomains		METIS subdomains	
	$I(\kappa)$		$I(\kappa)$	
S	$H/h = 16,\ H/\delta{=}4$	N_V	$H/h = 16,\ H/\delta = 4$	N_V
64	24(8.7)	49	29(10.5)	97
144	28(10.8)	121	26(8.6)	243
256	32(12.1)	225	28(8.9)	447
400	32(12.4)	361	31(11.3)	711

Table 4.3: Condition number (κ) for the preconditioned system $A_{ad}^{-1}A$ and number of PGC iterations (I) as a function of the number of square and irregular subdomains N; M is the coarse space dimension, $H/h = 16$, $H/\delta = 4$ and ρ is piecewise constant

as $\eta = \dfrac{\max \rho}{\min \rho}$; see [Calvo and Galvis, 2024].

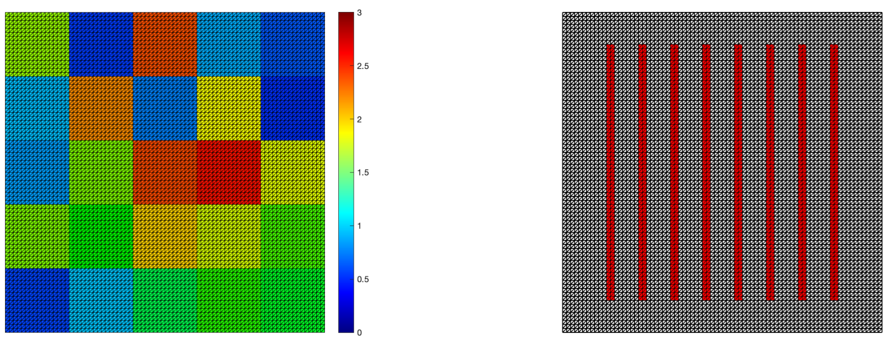


Figure 4.10: Two coefficients ρ: (left) $\rho \in [10^{-3}, 10^{3}]$ is constant on each subdomain (right), $\rho = 10^3$ in the red channels and $\rho = 1$ otherwise.

4.5.3 Sparsity considerations

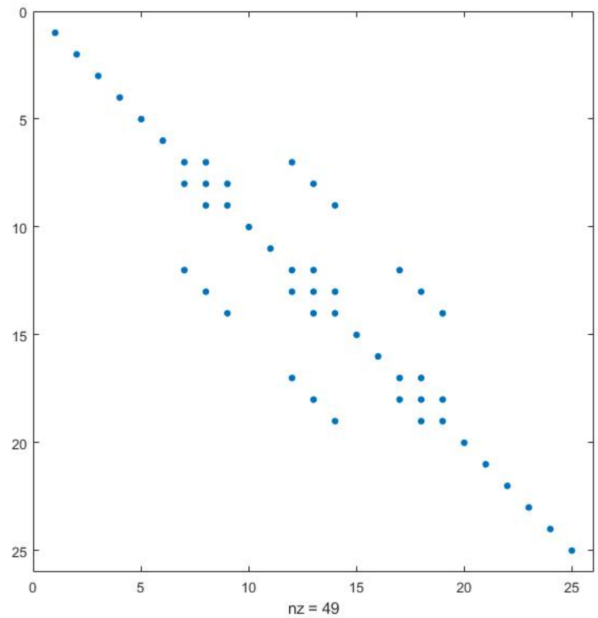
We conclude this chapter with visual representations of the sparse stiffness matrix for FEM and DGM; see Figure 4.11, where we compare the numbers of non-zero entries for $h = \frac{1}{4}$.

	Square subdomains		METIS subdomains	
	$I(\kappa)$		$I(\kappa)$	
H/h	$S = 36,\ H/\delta{=}4$	N_V	$S = 36,\ H/\delta = 4$	N_V
8	22(6.8)	25	29(11.0)	48
16	22(7.0)	25	26(11.6)	48
32	22(6.7)	25	27(11.2)	48
64	24(8.4)	25	29(11.0)	48

Table 4.4: Condition number (κ) for the preconditioned system $A_{ad}^{-1}A$ and number of PCG iterations (I) as a function of H/h for square and irregular subdomains, $N = 36$ and $H/\delta = 4$; M is the coarse space dimension. We use a piecewise constant coefficient ρ.

	Square subdomains		METIS subdomains	
	$I(\kappa)$		$I(\kappa)$	
S	$H/h = 16,\ H/\delta{=}4$	N_V	$H/h = 16,\ H/\delta = 4$	N_V
64	51(220)	49	79(243)	97
144	53(221)	121	66(204)	243
256	52(179)	225	41(19.6)	447
400	65(214)	361	39(16.8)	711

Table 4.5: Condition number (κ) for the preconditioned system $A_{ad}^{-1}A$ and number of PGC iterations (I) as a function of the number of square and irregular subdomains N; M is the coarse space dimension, $H/h = 16$, $H/\delta = 4$ and ρ has a channel distribution.



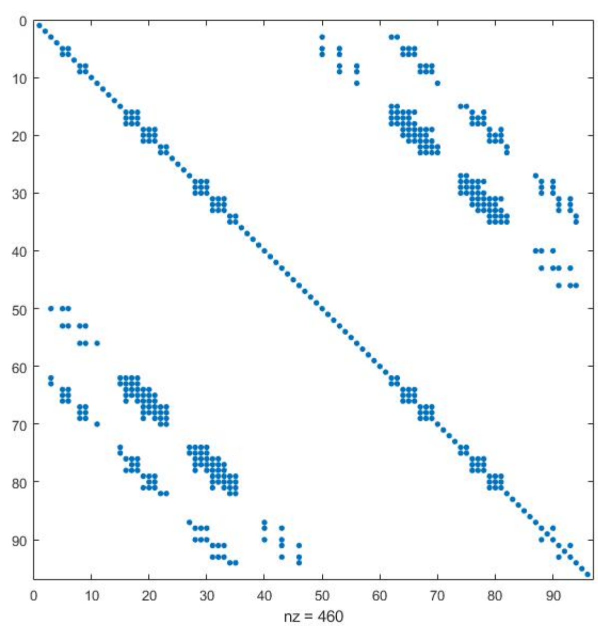


Figure 4.11: Sparsity patters for stiffness matrices for (up) FEM and (bottom) DGM. We have 49 and 460 nonzero entries, respectively, for $h = \frac{1}{4}$.

5 — Conclusions and future work

In the course of this research, we have studied and solved Poisson's equation with homogeneous Dirichlet boundary conditions. Our focus was on the formulation of a bilinear form and on the application of the discontinuous Galerkin method to obtain accurate and efficient solutions. In the first part of our investigation, we defined the bilinear form for the discontinuous problem and demonstrated the continuity and coercivity of this formulation. Furthermore, we explored the convergence of this method in Poisson's equation. These theoretical foundations are essential to understand the effectiveness of our approach and lay the foundation for subsequent stages of the research. Once we obtained a symmetric bilinear form for the problem, we proceeded to generate the matrix linear system. Here, our attention was focused on the solution of this system, where we implemented an iterative Schwarz preconditioner method with overlap. This method showed promising results, and its effectiveness lies in the definition of local spaces and extension operators, as well as in understanding the coarse space and the corresponding extension operator. Additionally, we narrowed down the condition number of the additive operator, which significantly increases the efficiency and stability of our approach. Although we still need to formally prove our conjecture regarding this bound, the experimental results support our claims. In the last stage of our work, we developed the construction of local spaces and the construction of the coarse space, which allowed us to make the condition number of the additive operator independent of the number of subdomains. This innovation is crucial for the scalability of our approach and its applicability to two-dimensional problems. As the main result of this investigation, we obtained a bound for the condition number of the additive operator. Although, this result is still considered a conjecture awaiting formal demonstration, the preliminary results and experiments performed support our hypothesis. Moving forward, two important directions for research emerge: First of all, it is essential to work on the formal proof of the conjecture on the condition number of the additive operator. This validation will further strengthen the robustness of our approach. Second, expanding our work to address problems in three dimensions, represents an opportunity-rich area for future research. In summary, this master's thesis has addressed the solution of Poisson´s equation with zero Dirichlet boundary conditions, by using an innovative approach that combines the formulation of a bilinear form, the discontinuous Galerkin method and the Schwarz preconditioner, with overlap in the presence of irregular subdomains. Although challenges remain, our preliminary results support the feasibility and effectiveness of this approach and pave the way for future research in this exciting field.

Bibliography

[Babuska and Aziz, 1976] Babuska, I. and Aziz, A. K. (1976). On the angle condition in the finite element method. *SIAM Journal on Numerical Analysis*, 13(2):214–226.

[Braess, 2007] Braess, D. (2007). *Finite elements: theory, fast solvers, and applications in elasticity theory*. Cambridge University Press.

[Brenner and Scott, 2008] Brenner, S. C. and Scott, L. R. (2008). *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. Springer.

[Brezzi and Marini, 2014] Brezzi, F. and Marini, L. D. (2014). A quick tutorial on DG methods for elliptic problems. In *Recent developments in discontinuous Galerkin finite element methods for partial differential equations*, volume 157 of *IMA Vol. Math. Appl.*, pages 1–24. Springer, Cham.

[Calvo, 2018] Calvo, J. G. (2018). On the approximation of a virtual coarse space for domain decomposition methods in two dimensions. *Math. Models Methods Appl. Sci.*, 28(7):1267–1289.

[Calvo, 2024] Calvo, J. G. (2024). DGM Library. `https://github.com/jgcalvo/DGM`.

[Calvo and Galvis, 2024] Calvo, J. G. and Galvis, J. (2024). Robust domain decomposition methods for high-contrast multiscale problems on irregular domains with virtual element discretizations. *Journal of Computational Physics*, 505:112909.

[F. Brezzi, 2014] F. Brezzi, L. D. M. (2014). *Recent Developments in Discontinuous Galerkin Finite Element Methods for Partial Differential Equations*. Springer International Publishing.

[Georgoulis, 2011] Georgoulis, E. H. (2011). Discontinuous Galerkin methods for linear problems: an introduction. In *Approximation algorithms for complex systems*, volume 3 of *Springer Proc. Math.*, pages 91–126. Springer, Heidelberg.

[Karypis and Kumar, 1998] Karypis, G. and Kumar, V. (1998). *METIS: A Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, and Computing Fill-Reducing Orderings of Sparse Matrices.*

[Toselli and Widlund, 2004] Toselli, A. and Widlund, O. (2004). *Domain Decomposition Methods - Algorithms and Theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer.

Appendix

5.1 Weak formulation

In this appendix we present the computations required for assembling the stiffness matrix associated to the discontinuous formulation. We focus on Poisson's equation on a given polygonal domain $\Omega \subset \mathbb{R}^2$ with homogeneous Dirichlet boundary conditions:

$$\begin{aligned} -\mathrm{div}(\rho \nabla u) &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{5.1}$$

where the coefficient $\rho : \overline{\Omega} \to \mathbb{R}^+$ is a given uniformly positive bounded function. As usual, we multiply (5.1) by a test function $v \in V_h$ and integrate to obtain

$$\sum_{K \in \mathcal{T}_h} \int_K \rho \nabla u \nabla v - \sum_{K \in \mathcal{T}_h} \int_{\partial K} \rho v \nabla u \cdot \boldsymbol{n}_k = \sum_{K \in \mathcal{T}_h} \int_K f v,$$

which is the same as

$$(\rho \nabla u, \nabla v)_{\mathcal{T}_h} - \langle [\![v]\!], \{\{\rho \nabla u\}\} \rangle_{\mathcal{E}_h} - \langle \{\{v\}\}, [\![\rho \nabla u]\!] \rangle_{\mathcal{E}_h^i} = (f, v)_{\mathcal{T}_h}.$$

For smooth solutions, the last term to the left of the equals sign is zero. Finally, we can define the symmetric bilinear form by

$$a_h(u, v) = (\rho \nabla u, \nabla v)_{\mathcal{T}_h} - \langle \{\{\rho \nabla u\}\}, [\![v]\!] \rangle_{\mathcal{E}_h} - \langle \{\{\rho \nabla v\}\}, [\![u]\!] \rangle_{\mathcal{E}_h} + \frac{\alpha_{sta}}{h} \langle [\![u]\!], [\![v]\!] \rangle_{\mathcal{E}_h}$$

for some $\alpha_{sta} > 0$. We have used that the jumps $[\![u]\!]$ are zero for the solution u. We then write

$$a_h(u,v) = a(u,v) - b_1(u,v) - b_1(v,u) + b_2(u,v), \tag{5.2}$$

where

$$b_1(u,v) := \sum_{e\in\mathcal{E}_h} \int_e \{\{\rho\nabla u\}\} \cdot [\![v]\!] ds, \quad b_2(u,v) := \sum_{e\in\mathcal{E}_h} \frac{\alpha_{sta}}{h} \int_e [\![u]\!][\![v]\!] ds.$$

5.1.1 Integrals for edges on the border

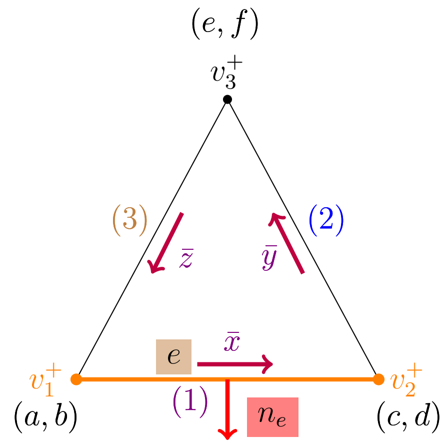


Figure 5.1: Data of a general triangle

Given a triangle E with vertices (a,b), (c,d), (e,f), as shown in Figure 5.1, we compute the equations of the edges of E:

(1) $y - d = \hat{m}(x - c)$ with $\hat{m} = \dfrac{d-b}{c-a}$.

(2) $y - d = m(x - c)$ with $m = \dfrac{f-d}{e-c}$.

(3) $y - b = \tilde{m}(x - a)$ with $\tilde{m} = \dfrac{f-b}{e-a}$.

The normal vector for edge e is $n_e = \dfrac{1}{|e|}\begin{bmatrix} d-b \\ c-a \end{bmatrix}$.
On each triangle, the local basis functions are defined by

$$\begin{aligned}
&1) \quad \phi_1 = \tfrac{1}{M_1}[y-d-m(x-c)] \quad \Rightarrow \nabla\phi_1 = \begin{bmatrix} -m/M_1 \\ 1/M_1 \end{bmatrix}, \\
&\text{with } M_1 = b-d-m(a-c) \\
&2) \quad \phi_2 = \tfrac{1}{M_2}[y-b-\tilde{m}(x-a)] \quad \Rightarrow \nabla\phi_2 = \begin{bmatrix} -\tilde{m}/M_2 \\ 1/M_2 \end{bmatrix}, \\
&\text{with } M_2 = d-b-\tilde{m}(c-a) \\
&3) \quad \phi_3 = \tfrac{1}{M_3}[y-d-\hat{m}(x-c)] \quad \Rightarrow \nabla\phi_3 = \begin{bmatrix} -\hat{m}/M_3 \\ 1/M_3 \end{bmatrix}. \\
&\text{with } M_3 = f-d-\hat{m}(e-c)
\end{aligned}$$

Define

$$\bar{\boldsymbol{x}} = \begin{bmatrix} a-c \\ d-b \end{bmatrix}, \quad \bar{\boldsymbol{y}} = \begin{bmatrix} e-c \\ f-d \end{bmatrix}, \quad \bar{\boldsymbol{z}} = \begin{bmatrix} a-e \\ b-f \end{bmatrix}.$$

We first compute b_{11}. We note that $\phi_1^- = 0$ and therefore, using that $\nabla\phi_1^- \cdot \boldsymbol{n}^+$ is constant, we obtain

$$\begin{aligned}
b_{11} &= \frac{1}{2}\int_e \left(\rho^+\nabla\phi_1^+ + \cancel{\rho^-\nabla\phi_1^-}\right)\cdot\left(\phi_1^+\boldsymbol{n}^+ + \cancel{\phi_1^-\boldsymbol{n}}\right) = \frac{1}{2}\nabla\phi_1^+\cdot\boldsymbol{n}^+\int_e \rho^+\phi_1^+ \\
&= \frac{1}{2}\frac{1}{M_1}\begin{bmatrix} -m \\ 1 \end{bmatrix}\cdot\frac{1}{|e|}\begin{bmatrix} d-b \\ a-c \end{bmatrix}\frac{|e|\rho^+(x_m^e)}{2} \\
&= \frac{a-c-m(d-b)}{4(b-d-m(a-c))}\rho^+(x_m^e) \\
&= \frac{(a-c)(e-c)-(f-d)(d-b)}{4[(b-d)(e-c)-(f-d)(a-c)]}\rho^+(x_m^e) \\
&= \frac{1}{4}\frac{\bar{\boldsymbol{y}}\cdot\bar{\boldsymbol{x}}}{\bar{\boldsymbol{y}}^\perp\cdot\bar{\boldsymbol{x}}}\rho^+(x_m^e).
\end{aligned}$$

Since $\phi_1^+(x_m^e) = \frac{1}{2}$, where x_m^e is middle point and $\rho^+ = \rho|_{K_e^+}$.
Since $\displaystyle\int_e \phi_2^+ = \frac{|e|}{2}$, we obtain $b_{12} = b_1(\phi_1,\phi_2) = b_{11}$.
Moreover,

$$b_{13} = \frac{1}{2}\int_e \left(\rho^+\nabla\phi_1^+ + \cancel{\rho^-\nabla\phi_1^-}\right)\cdot\left(\cancel{\phi_3^+ n^+} + \cancel{\phi_3^- n}\right) = 0,$$

since $\phi_3^+ = 0$ on e.
Following the same process,

$$\begin{aligned}
b_{21} &= \frac{1}{2}\int_e \left(\rho^+\nabla\phi_2^+ + \cancel{\rho^-\nabla\phi_2^-}\right)\cdot\left(\phi_1^+\boldsymbol{n}^+ + \cancel{\phi_1^-\bar{\boldsymbol{n}}}\right) = \frac{1}{2}\nabla\phi_2^+\cdot\boldsymbol{n}^+\frac{|e|}{2}\rho^+(x_m^e)\\
&= \frac{a-c-\tilde{m}(d-b)}{4(d-b-\tilde{m}(c-a))}\rho^+(x_m^e)\\
&= \frac{(c-a)(a-e)-(b-f)(d-b)}{4[(c-a)(b-f)-(d-b)(a-e)]}\rho^+(x_m^e)\\
&= \frac{1}{4}\frac{\bar{\boldsymbol{x}}\cdot\bar{\boldsymbol{z}}}{\bar{\boldsymbol{x}}^\perp\cdot\bar{\boldsymbol{z}}}\rho^+(x_m^e).
\end{aligned}$$

Similary $\boxed{b_{22} = b_{21}}$ because $\displaystyle\int_e \phi_2^+ = \frac{|e|}{2}$.
Also

$$b_{23} = \frac{1}{2}\int_e \left(\rho^+\nabla\phi_2^+ + \cancel{\rho^-\nabla\phi_2^-}\right)\cdot\left(\cancel{\phi_3^+ n^+} + \cancel{\phi_3^-\bar{n}}\right) = 0,$$

since $\phi_3^+ = 0$ on e.
Finally,

$$\begin{aligned}
b_{31} &= \frac{1}{2}\int_e \left(\rho^+\nabla\phi_3^+ + \cancel{\rho^-\nabla\phi_3^-}\right)\cdot\left(\phi_1^+\boldsymbol{n}^+ + \cancel{\phi_1^-\bar{n}}\right) = \frac{1}{2}\nabla\phi_3^+\cdot\boldsymbol{n}^+\frac{|e|}{2}\rho^+(x_m^e)\\
&= \frac{d-b-\hat{m}(a-c)}{4(f-d-\hat{m}(e-c))}\rho^+(x_m^e)\\
&= \frac{(c-a)(a-c)-(d-b)(d-b)}{4[(c-a)(f-d)-(d-b)(e-c)]}\rho^+(x_m^e)\\
&= \frac{1}{4}\frac{\bar{\boldsymbol{x}}\cdot\bar{\boldsymbol{x}}}{\bar{\boldsymbol{y}}^\perp\cdot\bar{\boldsymbol{x}}}\rho^+(x_m^e).
\end{aligned}$$

Again, notice that $\boxed{b_{32} = b_{31}}$ because $\displaystyle\int_e \phi_2^+ = \frac{|e|}{2}$.
We have

$$b_{33} = \frac{1}{2}\int_e \left(\rho^+\nabla\phi_3^+ + \cancel{\rho^-\nabla\phi_3^-}\right)\cdot\left(\cancel{\phi_3^+ n^+} + \cancel{\phi_3^-\bar{n}}\right) = 0,$$

since $\phi_3^+ = 0$ on e. We then have the local matriz associated to $b_1(\cdot,\cdot)$:

$$B_1^{loc} = \frac{1}{4}\begin{bmatrix} \frac{y\cdot x}{y\otimes x} & \frac{y\cdot x}{y\otimes x} & 0 \\ \frac{x\cdot z}{x\otimes z} & \frac{x\cdot z}{x\otimes z} & 0 \\ \frac{x\cdot x}{y\otimes x} & \frac{x\cdot x}{y\otimes x} & 0 \end{bmatrix}.$$

Consider the edge e parameterized by $\boldsymbol{r}(t)$

$$\boldsymbol{r}(t) = (1-t)\begin{pmatrix} a \\ b \end{pmatrix} + t\begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} a + (c-a)t \\ b + (d-b)t \end{pmatrix},$$

and then

$$\begin{aligned} \phi_1(\boldsymbol{r}(t)) =& \frac{1}{M_1}(y - d - m(x-c))|_{\boldsymbol{r}(t)} \\ =& \frac{(1-t)}{M_1}\left(\frac{(b-d)(e-c)-(f-d)(a-c)}{e-c}\right) = 1-t, \end{aligned}$$

and

$$\begin{aligned} \phi_2(\boldsymbol{r}(t)) =& \frac{1}{M_2}(y - b - \tilde{m}(x-a))|_{\boldsymbol{r}(t)} \\ =& \frac{t}{M_2}\left(\frac{(d-b)(e-a)-(f-d)(c-a)}{e-a}\right) = t. \end{aligned}$$

Finally, we get

$$\int_e \phi_1 = |e|\int_0^1 \phi_1(\boldsymbol{r}(t))dt = \int_0^1 (1-t)dt = \frac{|e|}{2},$$

$$\int_e \phi_2 = |e|\int_0^1 \phi_2(\boldsymbol{r}(t))dt = \int_0^1 tdt = \frac{|e|}{2}.$$

And for the bilinear form b_2:

$$b_2(\phi_i, \phi_j) = \int_e \phi_i^+ \boldsymbol{n}^+ \cdot \phi_j^+ \boldsymbol{n}^+ = \int_e \phi_i^+ \phi_j^+.$$

We deduce that

$$
\begin{aligned}
&1) \quad b_{31} = b_{32} = b_{33} = b_{13} = b_{23} = 0 \text{ since } \phi_3^+ = 0 \text{ on } e. \\
&2) \quad b_{11} = \int_e (\phi_1^+)^2 = |e| \int_0^1 (1-t)^2 dt = |e| \frac{-(1-t)^3}{3}\Big|_0^1 = \frac{|e|}{3}, \\
&3) \quad b_{12} = b_{21} = |e| \int_0^1 t(1-t) dt = |e| \left(\frac{t^2}{2} - \frac{t^3}{3} \right) \Big|_0^1 = \frac{|e|}{6}, \\
&4) \quad b_{22} = \int_e (\phi_2^+)^2 = |e| \int_0^1 t^2 dt = \frac{|e|}{3}.
\end{aligned}
$$

And the local stability matrix is

$$
B_2^{loc} = |e| \begin{bmatrix} \frac{1}{3} & \frac{1}{6} & 0 \\ \frac{1}{6} & \frac{1}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix}.
$$

5.1.2 Integrals for inner edges

Consider an interior edge e, shared by elements E^+ y E^-; see Figure (5.2)

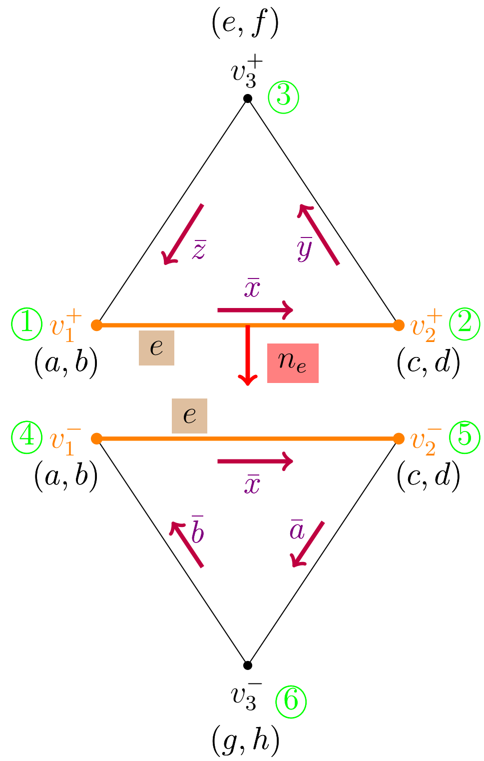


Figure 5.2: Data from two adjacent triangles

With the same notation as before, on each triangle, the basis functions are defined by

1) $\phi_1 = \frac{1}{M_1}[y - d - m(x-c)] \quad \Rightarrow \nabla\phi_1 = \begin{bmatrix} -m/M_1 \\ 1/M_1 \end{bmatrix}$,
with $M_1 = b - d - m(a-c)$

2) $\phi_2 = \frac{1}{M_2}[y - b - \tilde{m}(x-a)] \quad \Rightarrow \nabla\phi_2 = \begin{bmatrix} -\tilde{m}/M_2 \\ 1/M_2 \end{bmatrix}$,
with $M_2 = d - b - \tilde{m}(c-a)$

3) $\phi_3 = \frac{1}{M_3}[y - d - \hat{m}(x-c)] \quad \Rightarrow \nabla\phi_3 = \begin{bmatrix} -\hat{m}/M_3 \\ 1/M_3 \end{bmatrix}$,
with $M_3 = f - d - \hat{m}(e-c)$

with

$$m = \frac{f-d}{e-c}, \quad \tilde{m} = \frac{f-d}{e-a}, \quad \hat{m} = \frac{d-b}{c-a}, \text{ and}$$

4) $\quad \phi_4 = \frac{1}{M_4}[y - d - \hat{n}(x-c)] \quad \Rightarrow \nabla\phi_4 = \begin{bmatrix} -\hat{n}/M_4 \\ 1/M_4 \end{bmatrix},$

with $M_4 = b - d - \hat{n}(a-c)$

5) $\quad \phi_5 = \frac{1}{M_5}[y - b - \tilde{n}(x-a)] \quad \Rightarrow \nabla\phi_5 = \begin{bmatrix} -\tilde{n}/M_5 \\ 1/M_5 \end{bmatrix},$

with $M_5 = d - b - \tilde{n}(c-a)$

6) $\quad \phi_6 = \frac{1}{M_6}[y - d - \hat{m}(x-c)] \quad \Rightarrow \nabla\phi_6 = \begin{bmatrix} -\hat{m}/M_6 \\ 1/M_6 \end{bmatrix},$

with $M_6 = h - d - \hat{m}(g-c)$

with

$$\hat{n} = \frac{h-d}{g-c}, \quad \tilde{n} = \frac{h-b}{g-a}, \quad \hat{m} = \frac{d-b}{c-a}.$$

Consider

$$\bar{\boldsymbol{x}} = \begin{bmatrix} c-a \\ d-b \end{bmatrix}, \quad \bar{\boldsymbol{a}} = \begin{bmatrix} g-c \\ h-d \end{bmatrix}, \quad \bar{\boldsymbol{b}} = \begin{bmatrix} a-g \\ b-h \end{bmatrix}, \quad \boldsymbol{n_e} = \frac{1}{|e|} \cdot \begin{bmatrix} b-d \\ c-a \end{bmatrix}.$$

For computing

$$b_1(\phi_i, \phi_j) = \frac{1}{2} \int_e \left(\nabla\phi_i^+ + \nabla\phi_i^-\right) \cdot \left(\phi_j^+ n^+ + \phi_j^- \bar{n}\right),$$

similarly we derive formulas for b_{11}, b_{12}, …, b_{33}. We continue calculating

$$\begin{aligned}
b_{14} &= \frac{1}{2}\int_e \left(\rho^+\nabla\phi_1^+ + \cancel{\rho^-\nabla\phi_1^-}\right) \cdot \left(\cancel{\phi_4^+ n^+} + \phi_4^- \bar{n}\right) = \frac{1}{2}\nabla\phi_1^+ \cdot \bar{n} \int_e \phi_4^- = -b_{11}, \\
b_{15} &= \frac{1}{2}\int_e \left(\rho^+\nabla\phi_1^+ + \cancel{\rho^-\nabla\phi_1^-}\right) \cdot \left(\cancel{\phi_5^+ n^+} + \phi_5^- \bar{n}\right) = \frac{1}{2}\nabla\phi_1^+ \cdot \bar{n} \int_e \phi_5^- = -b_{12}, \\
b_{16} &= \frac{1}{2}\int_e \left(\rho^+\nabla\phi_1^+ + \cancel{\rho^-\nabla\phi_1^-}\right) \cdot \left(\cancel{\phi_6^+ n^+} + \cancel{\phi_6^- \bar{n}}\right) = 0. \\
b_{24} &= \frac{1}{2}\int_e \left(\rho^+\nabla\phi_2^+ + \cancel{\rho^-\nabla\phi_2}\right) \cdot \left(\cancel{\phi_4^+ n^+} + \phi_4^- \bar{n}\right) = \frac{1}{2}\nabla\phi_2^+ \cdot n^- \int_e \phi_4^- = -b_{21}, \\
b_{25} &= \frac{1}{2}\int_e \left(\rho^+\nabla\phi_2^+ + \cancel{\rho^-\nabla\phi_2^-}\right) \cdot \left(\cancel{\phi_5^+ n^+} + \phi_5^- \bar{n}\right) = \frac{1}{2}\nabla\phi_2^+ \cdot \bar{n} \int_e \phi_5^- = -b_{22}, \\
b_{26} &= \frac{1}{2}\int_e \left(\rho^+\nabla\phi_2^+ + \cancel{\rho^-\nabla\phi_2^-}\right) \cdot \left(\cancel{\phi_6^+ n^+} + \cancel{\phi_6^- \bar{n}}\right) = 0. \\
b_{34} &= \frac{1}{2}\int_e \left(\rho^+\nabla\phi_3^+ + \cancel{\rho^-\nabla\phi_3^-}\right) \cdot \left(\cancel{\phi_4^+ n^+} + \phi_4^- \bar{n}\right) = \frac{1}{2}\nabla\phi_3^+ \cdot \bar{n} \int_e \phi_4^- = -b_{31},
\end{aligned}$$

$$
\begin{aligned}
b_{35} &= \frac{1}{2}\int_e \left(\rho^+\nabla\phi_3^+ + \cancel{\rho^-\nabla\phi_3^-}\right)\cdot\left(\cancel{\phi_5^+ n^+} + \phi_5^-\bar{n}\right) = \frac{1}{2}\nabla\phi_3^+\cdot\bar{n}\int_e \phi_5^- = -b_{32},\\
b_{36} &= \frac{1}{2}\int_e \left(\rho^+\nabla\phi_3^+ + \cancel{\rho^-\nabla\phi_3^-}\right)\cdot\left(\cancel{\phi_6^+ n^+} + \cancel{\phi_6^-\bar{n}}\right) = 0.\\
b_{41} &= \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_4^+} + \rho^-\nabla\phi_4^-\right)\cdot\left(\phi_1^+ n^+ + \cancel{\phi_1^-\bar{n}}\right) = -b_{44},\\
b_{42} &= \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_4^+} + \rho^-\nabla\phi_4^-\right)\cdot\left(\phi_2^+ n^+ + \cancel{\phi_2^-\bar{n}}\right) = -b_{45},\\
b_{43} &= \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_4^+} + \rho^-\nabla\phi_4^-\right)\cdot\left(\cancel{\phi_3^+ n^+} + \cancel{\phi_3^-\bar{n}}\right) = 0.\\
b_{51} &= \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_5^+} + \rho^-\nabla\phi_5^-\right)\cdot\left(\phi_1^+ n^+ + \cancel{\phi_1^-\bar{n}}\right) = -b_{54},\\
b_{52} &= \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_5^+} + \rho^-\nabla\phi_5^-\right)\cdot\left(\phi_2^+ n^+ + \cancel{\phi_2^-\bar{n}}\right) = -b_{55},\\
b_{53} &= \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_5^+} + \rho^-\nabla\phi_5^-\right)\cdot\left(\cancel{\phi_3^+ n^+} + \cancel{\phi_3\bar{n}}\right) = 0.\\
b_{61} &= \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_6^+} + \rho^-\nabla\phi_6^-\right)\cdot\left(\phi_1^+ n^+ + \cancel{\phi_1^- n^-}\right) = -b_{64},\\
b_{62} &= \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_6^+} + \rho^-\nabla\phi_6^-\right)\cdot\left(\phi_2^+ n^+ + \cancel{\phi_2^-\bar{n}}\right) = -b_{65},\\
b_{63} &= \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_6^+} + \rho^-\nabla\phi_6^-\right)\cdot\left(\cancel{\phi_3^+ n^+} + \cancel{\phi_3\bar{n}}\right) = 0.
\end{aligned}
$$

$$
\begin{aligned}
b_{44} &= \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_4^+} + \nabla\phi_4^-\right)\cdot\left(\cancel{\phi_4^+\bar{n}^+} + \phi_4^-\bar{n}\right) = \frac{1}{2}\nabla\phi_4^-\cdot\bar{n}\int_e \phi_4^- = \frac{1}{4}\frac{\bar{\boldsymbol{x}}\cdot\bar{\boldsymbol{a}}}{\bar{\boldsymbol{x}}^{\perp}\cdot\bar{\boldsymbol{a}}},\\
b_{45} &= \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_4^+} + \rho^-\nabla\phi_4^-\right)\cdot\left(\cancel{\phi_5^+ n^+} + \phi_5^-\bar{n}\right) = b_{44},\\
b_{46} &= \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_4^+} + \rho^-\nabla\phi_4^-\right)\cdot\left(\cancel{\phi_6^+ n^+} + \cancel{\phi_6^+ n^-}\right) = 0.\\
b_{54} &= \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_5^+} + \rho^-\nabla\phi_5^-\right)\cdot\left(\cancel{\phi_4^+ n^+} + \phi_4^-\bar{n}\right) = \frac{1}{2}\nabla\phi_5^-\cdot\bar{n}\int_e \phi_4^- = \frac{1}{4}\frac{\bar{\boldsymbol{b}}\cdot\bar{\boldsymbol{x}}}{\bar{\boldsymbol{b}}^{\perp}\cdot\bar{\boldsymbol{x}}},\\
b_{55} &= \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_5^+} + \nabla\phi_5^-\right)\cdot\left(\cancel{\phi_5^+ n^+} + \phi_5^-\bar{n}\right) = b_{54},\\
b_{56} &= \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_5^+} + \rho^-\nabla\phi_5^-\right)\cdot\left(\cancel{\phi_6^+ n^+} + \cancel{\phi_6^+ n^-}\right) = 0.
\end{aligned}
$$

$$b_{64} = \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_6^+} + \rho^-\nabla\phi_6^-\right)\cdot\left(\cancel{\phi_4^+ n^+} + \phi_4^-\bar{n}\right) = \frac{1}{2}\nabla\phi_6^-\cdot\bar{n}\int_e \phi_4^- = \frac{1}{4}\frac{\bar{\boldsymbol{x}}\cdot\bar{\boldsymbol{x}}}{\bar{\boldsymbol{x}}^{\perp}\cdot\bar{\boldsymbol{a}}},$$

$$b_{65} = \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_6^+} + \nabla\phi_6^-\right)\cdot\left(\cancel{\phi_5^+ n^+} + \phi_5^-\bar{n}\right) = b_{64},$$

$$b_{66} = \frac{1}{2}\int_e \left(\cancel{\rho^+\nabla\phi_6^+} + \rho^-\nabla\phi_6^-\right)\cdot\left(\cancel{\phi_6^+ n^+} + \cancel{\phi_6^+ n^-}\right) = 0.$$

Finally, the local matriz associated to $b_1(\cdot,\cdot)$ is

$$B_1^{loc} = \frac{1}{4}\left[\begin{array}{ccc|ccc}
\frac{y\cdot x}{y\otimes x} & \frac{y\cdot x}{y\otimes x} & 0 & -\frac{y\cdot x}{y\otimes x} & -\frac{y\cdot x}{y\otimes x} & 0 \\
\frac{x\cdot z}{x\otimes z} & \frac{x\cdot z}{x\otimes z} & 0 & -\frac{x\cdot z}{x\otimes z} & -\frac{x\cdot z}{x\otimes z} & 0 \\
\frac{x\cdot x}{y\otimes x} & \frac{x\cdot x}{y\otimes x} & 0 & -\frac{x\cdot x}{y\otimes x} & -\frac{x\cdot x}{y\otimes x} & 0 \\
\hline
-\frac{\bar{x}\cdot\bar{a}}{\bar{x}\otimes\bar{a}} & -\frac{\bar{x}\cdot\bar{a}}{\bar{x}\otimes\bar{a}} & 0 & \frac{\bar{x}\cdot\bar{a}}{\bar{x}\otimes\bar{a}} & \frac{\bar{x}\cdot\bar{a}}{\bar{x}\otimes\bar{a}} & 0 \\
-\frac{b\cdot\bar{x}}{\bar{b}\otimes\bar{x}} & -\frac{b\cdot\bar{x}}{\bar{b}\otimes\bar{x}} & 0 & \frac{b\cdot\bar{x}}{\bar{b}\otimes\bar{x}} & \frac{b\cdot\bar{x}}{\bar{b}\otimes\bar{x}} & 0 \\
-\frac{\bar{x}\cdot\bar{x}}{\bar{x}\otimes\bar{a}} & -\frac{\bar{x}\cdot\bar{x}}{\bar{x}\otimes\bar{a}} & 0 & \frac{\bar{x}\cdot\bar{x}}{\bar{x}\otimes\bar{a}} & \frac{\bar{x}\cdot\bar{x}}{\bar{x}\otimes\bar{a}} & 0
\end{array}\right].$$

In the case of

$$b_2(\phi_i,\phi_j) = \int_e \left(\phi_i^+ n^+ + \phi_i^- n^-\right)\cdot\left(\phi_j^+ n^+ + \phi_j^- n^-\right),$$

we compute

$$1)\quad b_{44} = \int_e \left(\phi_4^-\right)^2 = |e|\int_0^1 (1-t)^2 = \frac{|e|}{3},$$

$$2)\quad b_{45} = b_{54} = \int_e \phi_4^-\cdot\phi_5^- = |e|\int_0^1 t(1-t) = \frac{|e|}{6},$$

$$3)\quad b_{55} = \int_e \left(\phi_5^-\right)^2 = \frac{|e|}{3}.$$

$$4)\quad b_{41} = b_{14} = -\int_e \phi_1^+\phi_4^- = -|e|\int_0^1 (1-t)^2 = -\frac{|e|}{3},$$

$$5)\quad b_{51} = b_{15} = -\int_e \phi_1^+\phi_5^- = -|e|\int_0^1 (1-t)t = -\frac{|e|}{6},$$

$$6)\quad b_{42} = b_{24} = -\int_e \phi_2^+\phi_4^- = -|e|\int_0^1 t(1-t) = -\frac{|e|}{6},$$

$$7)\quad b_{52} = b_{25} = -\int_e \phi_2^+\phi_5^- = -|e|\int_0^1 t\cdot t = -\frac{|e|}{3}.$$

Thus, the local matriz associated to $b_2(\cdot,\cdot)$ is

$$B_2^{loc} = |e| \left[\begin{array}{ccc|ccc} 1/3 & 1/6 & 0 & -1/3 & -1/6 & 0 \\ 1/6 & 1/3 & 0 & -1/6 & -1/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline -1/3 & -1/6 & 0 & 1/3 & 1/6 & 0 \\ -1/6 & -1/3 & 0 & 1/6 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right].$$