

SORBONNE UNIVERSITY AND ENS RENNES

2025

M2 Internship Report

---

# A POSTERIORI ERROR ESTIMATOR IN DISCONTINUOUS GALERKIN METHODS

---



Université  
de Rennes



Internship report by  
Raphaël LECOQ

Supervised by  
Andrew PEITAVY  
Erell JAMELOT

This report is part of my 3rd year in the Mathematics Department at the École Normale Supérieure (ENS) of Rennes<sup>1</sup> and my M2 “Mathematics of Modeling” program<sup>2</sup> at Sorbonne University<sup>3</sup>.

Although this internship does not mark the end of my studies at ENS, it does mark the end of my university education at the Master’s level and, therefore, of academic theses. In this sense, this report is special to me, and I would like to express my gratitude to several people.

I would first like to thank my two supervisors, Andrew and Erell. Thank you, Andrew, for your kindness, humor, and patience. Thank you, Erell, for your advice, rigor, and good mood. Thanks to both of you for your availability and for the opportunity you offered me at CEA Saclay. It would be hard to quantify everything I have learned from you, but after these (almost) six months, I feel I have grown both as a scientist and as a collaborator.

I thank the teaching staff of the Jacques-Louis Lions Laboratory<sup>4</sup> for the excellent quality of education provided at Sorbonne University. The variety of courses and numerous valuable recommendations this year have allowed me to fully thrive in numerical mathematics.

I also thank the Mathematics Department at ENS Rennes for allowing such freedom in their curriculum while maintaining supportive guidance for their students. The solid foundation I gained there gave me confidence in my projects and skills, for which I am deeply grateful.

Of course, no work exists without colleagues, and I would like to greet all members of the Local Scale Development Laboratory (LDEL) at CEA Saclay<sup>5</sup>. The warm welcome at the lab allowed for a quick and pleasant integration. I wish you to maintain the great working atmosphere that is so valuable for long-term research.

A special thanks to the interns of office 49, who made me start each morning with a smile and finish each day without noticing the time pass.

Finally, thanks to my family, who support me in all my projects, even the craziest ones, and who let me dream of new adventures and experiences.

I am aware of my luck in having had a simple, carefree childhood, and I owe everything to you for allowing me to reach this point with serenity.

---

<sup>1</sup><https://www.ens-rennes.fr/>

<sup>2</sup><https://www.ljll.fr/MathModel/>

<sup>3</sup><https://www.sorbonne-universite.fr/>

<sup>4</sup><https://www.ljll.fr/>

<sup>5</sup><https://www.cea.fr/paris-saclay/>

## Abstract

This study focuses on a posteriori error estimation for the isotropic heterogeneous diffusion operator within the framework of discontinuous Galerkin methods.

The study is divided into three main parts and one appendix:

Part 1 places the research project in the context of industrial applications. The first concepts of a posteriori error estimation are introduced to build intuition about the mathematical objects and their purpose.

Part 2 presents the numerical method used, called the *Discontinuous Galerkin Method*, corresponding to discontinuous finite elements. The considered mesh is defined, and the method is constructed. Finally, a priori error analysis is given for both homogeneous and heterogeneous diffusion cases.

Part 3 summarizes an article by Mark Ainsworth, 2007 [AIN07], which constructs an a posteriori error estimate. A modification of the article is added to fit the heterogeneous diffusion case. A second section presents numerical results of the estimator to verify its behavior, serving as a basis for benchmarking another estimator.

In Appendix 4, auxiliary computations for reproducing test cases and justifying numerical results are provided.

For any question, error, or typo:

[raphael.lecoq@ens-rennes.fr](mailto:raphael.lecoq@ens-rennes.fr)

# Contents

---

---

## Part 1 Industrial Context

---

---

<b>I -</b>	<b>The CEA</b>	<b>6</b>
	1) The Paris-Saclay Center . . . . .	6
	2) The STMF and the LDEL . . . . .	6
<b>II -</b>	<b>Theoretical and Industrial Motivation</b>	<b>7</b>

---

---

## Part 2 Discontinuous Galerkin Methods

---

---

<b>I -</b>	<b>Domain, mesh, and broken spaces</b>	<b>10</b>
	1) Homogeneous diffusion problem . . . . .	10
	2) Domain discretization . . . . .	11
	3) Space of discontinuous finite elements . . . . .	13
<b>II -</b>	<b>Variational Formulations</b>	<b>18</b>
	1) Symmetric Interior Penalty . . . . .	18
	2) Symmetric Weighted Interior Penalty . . . . .	22

---

---

## Part 3 Posteriori Error Estimators

---

---

<b>I -</b>	<b>Posteriori Estimation</b>	<b>26</b>
	1) Problem and Definition of the Estimator . . . . .	26
	2) Non-conforming Estimator . . . . .	28
	3) Conforming Estimator . . . . .	28
	4) Local Equivalence between Error and Estimator . . . . .	32
	5) Modification for SWIP . . . . .	32
<b>II -</b>	<b>Numerical Results</b>	<b>33</b>
	1) Homogeneous diffusion problem . . . . .	33
	2) Heterogeneous diffusion problem . . . . .	43
	<b>Conclusion</b>	<b>47</b>

---

---

## Part 4 Appendix

---

---

<b>A</b>	<b>SWIP formulation for nonhomogeneous conditions</b>	<b>48</b>
	1) Pure Dirichlet . . . . .	48
	2) Pure Neumann . . . . .	49
	3) Mixed boundary conditions . . . . .	50
<b>B</b>	<b>Equivalent formulation</b>	<b>52</b>
<b>C</b>	<b>Gradients and regularity of test cases</b>	<b>54</b>
	1) Spectral problem on the square domain . . . . .	54
	2) Harmonic problem . . . . .	54

<b>D</b>	<b><a href="#">SWIP solution</a></b>	<b>55</b>
	<b><a href="#">Glossary</a></b>	<b>59</b>
	<b><a href="#">Figures list</a></b>	<b>62</b>
	<b><a href="#">Bibliography</a></b>	<b>65</b>

## Part 1

# Industrial Context

## I - The CEA

The CEA (French Alternative Energies and Atomic Energy Commission) is a major player in French research, active in many fields related to energy, health, defense, and advanced technologies.

### 1) The Paris-Saclay Center

The Paris-Saclay site is one of the nine CEA centers. It brings together about 7,000 people, including CEA employees and scientific partners. The center focuses on a large share of CEA's civil activities, particularly in low-carbon energy, environment, life sciences, materials, and information technologies.

Located mainly in Saclay, Fontenay-aux-Roses, and Évry, it also has facilities in Orsay, Jouy-en-Josas, Paris, and Caen. It is a strategic partner of the University of Paris-Saclay, which alone accounts for more than 13% of French research.

### 2) The STMF and the LDEL

The Thermo-Hydraulics and Fluid Mechanics Service (STMF) is an applied research unit within the Department of Modeling for Systems and Structures (DM2S), part of the ISAS Institute (Applied Sciences and Simulation for Low-Carbon Energy), under the Energy Division (DES) of the CEA. Its director is Nicolas DORVILLE.

The STMF specializes in the development and validation of simulation software for fluid mechanics and thermo-hydraulics. These tools are used mainly in nuclear applications but also in other energy fields. The service's scientific approach relies on multi-scale modeling, from detailed multiphase flow simulations to large-scale representation of circuits and energy systems.

STMF also contributes to safety studies, particularly related to hydrogen, both in the nuclear sector and in emerging energy technologies (such as transport).

The Local Scale Development Laboratory (LDEL) is part of STMF, with 25 permanent researchers and 25 non-permanent members (PhD students, postdocs, apprentices, interns) under the direction of Julie DARONA.

A key focus of the laboratory is the development and sustainability of its industrial simulation platform TRUST<sup>6</sup>, which supports the open-source code TrioCFD<sup>7</sup>, dedicated to numerical simulation of multiphase flows.

---

<sup>6</sup><https://cea-trust-platform.github.io/>

<sup>7</sup><https://triocfd.cea.fr/>

## II - Theoretical and Industrial Motivation

In the analysis of numerical methods, a crucial step is establishing the convergence of the scheme. This is done via a priori error estimates such as Corollary 2.9 or Theorem 2.10.

The goal of a posteriori error estimation is to accurately estimate the numerical error using the data and the computed solution when the exact solution is unknown.

During this internship, we studied and implemented an a posteriori error estimator to integrate it into TrioCFD within the framework of Discontinuous Galerkin Methods (DGM).

The LDEL is interested in DGM for several reasons:

- Mesh flexibility:  
DGM allows complex meshes and supports cells of different shapes within the same mesh.

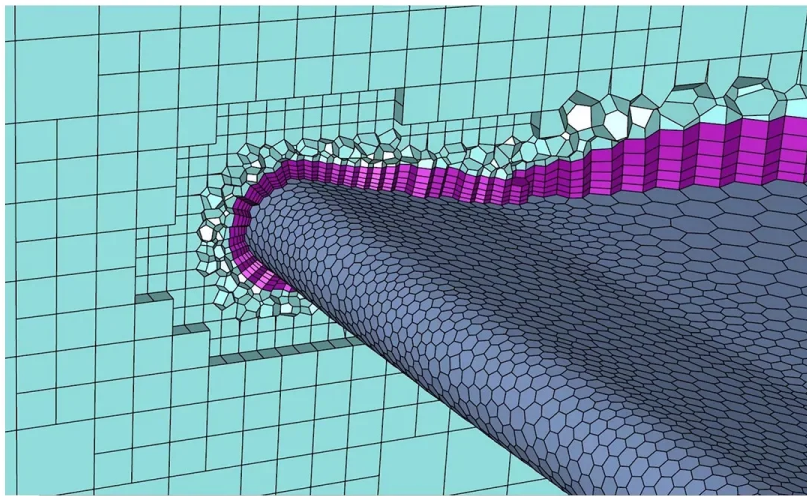


Figure 1: Mesh presented by ANSYS<sup>8</sup> showing several cell shapes within the same grid.

- GPU suitability:  
New, faster GPU architectures are increasingly used in simulation research. DGM benefits greatly from GPU computation.
- Parallelization:  
DGM is easy to parallelize, enabling efficient use of supercomputers and multi-processor simulations.

Applications of a posteriori estimation include:

- Determining convergence order without over-refined meshes.
- $hp$ -adaptive mesh refinement [D11].
- Model coupling [FAD22].

The study was conducted using the MATLAB prototype of TrioCFD, as well as C++ with FreeFEM++, and the method will later be integrated into the official TrioCFD version in C++.

<sup>8</sup><https://www.ansys.com>

We denote by  $\epsilon_h$  the exact simulation error in an energy norm, which may correspond to the broken flux norm (Theorem 3.1), the SIP norm (Definition 2.6), the SWIP norm (Equation (16)), or any other norm defined for a Discontinuous Galerkin method.

We consider a variational formulation obtained by a DGM with a source term represented by  $f$  and  $g = (g_D, g_N)$ , on a mesh  $\mathcal{T}_h$  and for a solution  $u_h$ .

### Definition 1.1: A posteriori error estimator

An a posteriori error estimator  $S_h(u_h, f, g, \mathcal{T}_h)$  is defined by an analogy to a norm equivalence:

$$\exists C > 0, \forall h > 0, \forall T \in \mathcal{T}_h, \quad CS_h|_T \leq \epsilon_h|_T \leq S_h|_T.$$

$S_h$  is *asymptotically correct* if

$$\frac{\epsilon_h - S_h}{\epsilon_h} \xrightarrow{h \rightarrow 0} 0 \quad \Longleftrightarrow \quad \xi_h := \frac{S_h}{\epsilon_h} \xrightarrow{h \rightarrow 0} 1,$$

where  $\xi_h \geq 1$  is the *effectivity* of the estimator. This means  $S_h$  tends toward  $e_h$  faster than  $e_h$  tends to 0.

We then use this estimator to construct *local error maps*, allowing visualization of mesh regions where the discretization is less accurate:

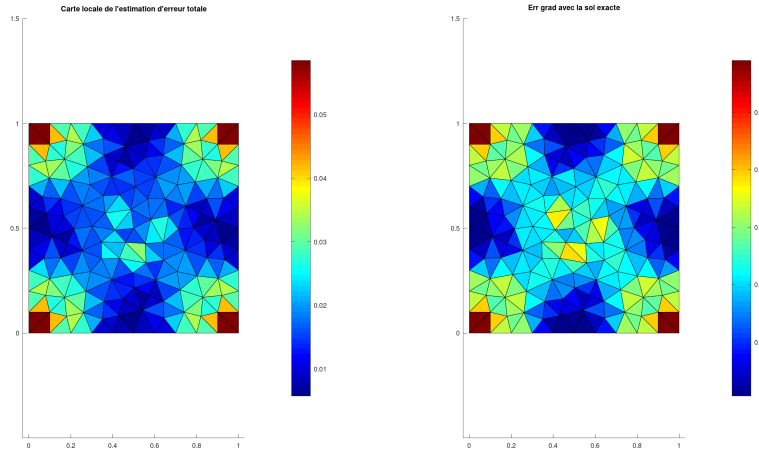


Figure 2: Local map of the a posteriori error estimate  $S_h$  (left) and the exact local error  $\epsilon_h$  (right) for the problem  $-\Delta u = \pi^2 \sin(\pi x) \sin(\pi y)$  with homogeneous Dirichlet conditions,  $h = 0.1, v_h \in \mathbb{P}^1$ .

The error estimate can then be used to locally refine (see Fig.3) the mesh only in regions where the estimate has high amplitude.

This allows obtaining superconvergence results, as observed in Fig. 28.

This superconvergence—recovering optimal convergence for problems with low-regularity solutions—helps reduce computation times.



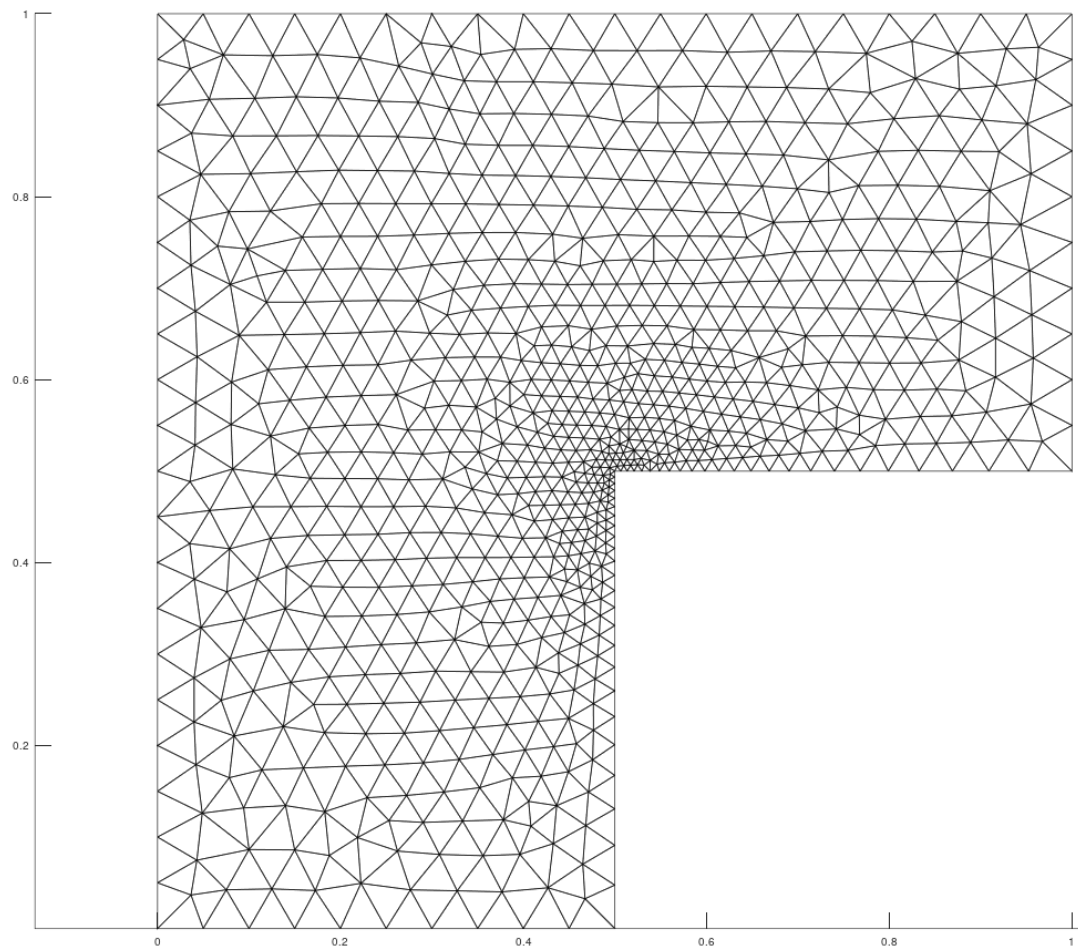


Figure 3: Example of an L-shaped mesh locally refined near the re-entrant corner.

## Part 2

# Discontinuous Galerkin Methods

This part is mainly based on the book *Mathematical Aspects of Discontinuous Galerkin Methods* [DE12, Section 1.2, Section 1.4, Section 4.1, Section 4.2, Section 4.5].

## I - Domain, mesh, and broken spaces

### 1) Homogeneous diffusion problem

We first consider the Poisson problem. This problem is an elementary building block for Stokes and Navier-Stokes equations.

It also naturally introduces all necessary concepts for properly defining discontinuous Galerkin methods, including the descriptive elements of the mesh.

Let  $\Omega \subset \mathbb{R}^d$  be a bounded open set with Lipschitz boundary,  $\partial\Omega = \overline{\Gamma_D} \sqcup \overline{\Gamma_N}$ , with data  $f \in L^2(\Omega)$ ,  $g_D \in H^{1/2}(\Gamma_D)$ , and  $g_N \in L^2(\Gamma_N)$ .

We are interested in the following problem:

$$\text{Find } u \in H^1(\Omega) \text{ such that } \begin{cases} -\Delta u = f & \text{in } \Omega, \\ u = g_D & \text{on } \Gamma_D, \\ \nabla u \cdot \mathbf{n} = g_N & \text{on } \Gamma_N. \end{cases} \quad (1)$$

To simplify the study, we assume  $g_D \equiv 0$  and  $\Gamma_N = \emptyset$ , the rest of the analysis being reproducible in the case of non-homogeneous data; see Appendix A.

By integration by parts, (1) is equivalent to

$$\text{Find } u \in H_0^1(\Omega), \quad \int_{\Omega} \nabla u \cdot \nabla w = \int_{\Omega} f w, \quad \forall w \in H_0^1(\Omega). \quad (2)$$

The well-posedness of the problem follows from the Lax-Milgram theorem [BRE11, Corollary 5.8]. Indeed, defining the bilinear form

$$a(v, w) = \int_{\Omega} \nabla v \cdot \nabla w = \langle v | w \rangle_{H_0^1(\Omega)},$$

$a(\cdot, \cdot)$  is a continuous bilinear form on  $H_0^1(\Omega) \times H_0^1(\Omega)$  by Cauchy-Schwarz [BRE11, Definition p.131] and coercive by Poincaré [BRE11, Proposition 8.13].

We define the linear form

$$\ell(w) = \int_{\Omega} f w = \langle f | w \rangle_{L^2(\Omega)},$$

which is continuous on  $H_0^1(\Omega)$  by Cauchy-Schwarz and Poincaré.

The diffusion problem can also be expressed in a mixed formulation involving fluxes:

$$\text{Find } \boldsymbol{\sigma}(u) \in \mathbf{H}(\text{div}; \Omega), u \in H_0^1(\Omega) \text{ such that } \begin{cases} -\text{div } \boldsymbol{\sigma} = f & \text{in } \Omega, \\ \boldsymbol{\sigma}(u) = \nabla u & \text{in } \Omega. \end{cases} \quad (3)$$

with appropriate boundary conditions.

## 2) Domain discretization

### Remark :

To ensure that  $\forall T \in \mathcal{T}_h, \nabla u \in \mathbf{H}^s(T)$ ,  $s > 1/2$ , which will later allow us to define its trace by Sobolev embedding [TAR07, Lemma 16.1], we must assume that  $\Omega$  is a sufficiently smooth manifold [HIR12, Chapter 1] so that  $u \in H^{3/2+\varepsilon}(\Omega)$ . According to [GRI11; DAU88]:

- If  $\Omega$  is of class  $\mathcal{C}^2$  or convex, then  $u \in H^2(\Omega)$ .
- If  $\Omega$  is a non-convex polyhedron, there exists  $\varepsilon \in ]0; 1/2[$  such that  $u \in \bigcap_{0 < s < \varepsilon} H^{3/2+\varepsilon}(\Omega)$ .

These assumptions therefore ensure sufficient regularity of the solution  $u$ .

DGM allow the use of very general meshes; see [DE12, Section 1.2, Section 1.4]:

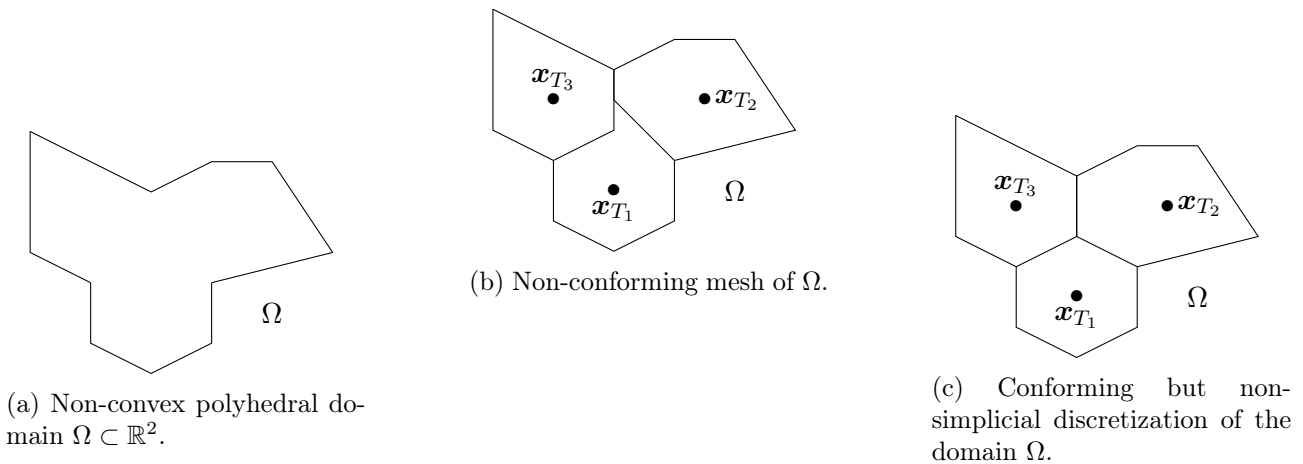


Figure 4: Example of a polyhedral domain and its mesh.

For our study, we restrict the domain discretization to a mesh made of simplices, whose definition is recalled in [DE12, Section 1.2.2], and whose properties are studied in [BMR04].

We assume the mesh is conforming in the following sense: a face  $F$  is shared by a single pair  $(T_1, T_2)$  if  $F \subset \Omega$ , or by a single  $T_1$  if  $F \subset \partial\Omega$ .

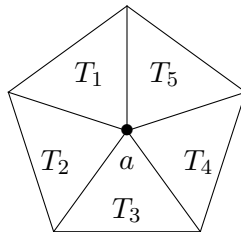


Figure 5: Simplicial discretization of a pentagon in 2D (triangulation).

We denote  $\mathcal{T}_h = \{T \mid T \text{ is a mesh element}\}$ .

We denote  $\mathcal{F}_T$  the set of faces of  $T$ ,  $\mathcal{F}_h$  the set of all mesh faces, and define  $\mathcal{F}_h^b := \bigcup_{T \in \mathcal{T}_h} \mathcal{F}_T \cap \Omega$ ,

$\mathcal{F}_h^i = \bigcup_{T \in \mathcal{T}_h} \mathcal{F}_T \cap \partial\Omega$  as the sets of interior and boundary faces, respectively. We have  $\mathcal{F}_h = \mathcal{F}_h^b \sqcup \mathcal{F}_h^i$ .

We denote  $\mathcal{A}_h(T) := \{a \mid a \text{ is a vertex of } T\}$  the set of vertices. We sometimes write  $a_i \in \mathcal{A}_h$  if  $a_i$  is the  $i$ -th vertex of an element and we are interested in elements sharing that vertex.

Finally, we define  $\mathcal{T}_F := \{T \mid F \subset \partial T\}$  the elements having  $F$  as a face and  $\Omega_i := \{T \mid a_i \in T\}$  the

set of elements sharing the vertex  $a_i$ .

For an interior face  $F \in \mathcal{F}_h^i$  and its two elements in  $\mathcal{T}_F$ , we fix a unique numbering  $T_1$  and  $T_2$  such that the unit normal  $\mathbf{n}_F$  is directed from  $T_1$  to  $T_2$ , and  $\mathbf{t}_F$  the unit tangent vector directed to form a direct basis.

The unnormalized normal vector is denoted  $\mathbf{S}_F = |F| \mathbf{n}_F$  and the tangential vector  $\boldsymbol{\tau}_F = |F| \mathbf{t}_F$ .

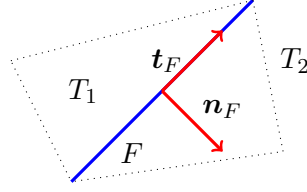
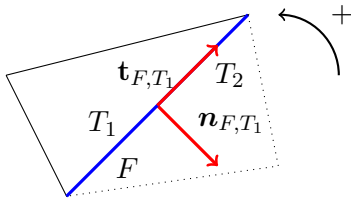


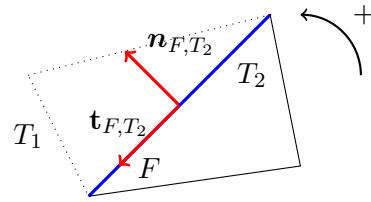
Figure 6: In blue, a face  $F$  shared by two elements  $T_1$  and  $T_2$ ,  $\mathbf{n}_F$  its unit normal vector and  $\mathbf{t}_F$  its tangential vector.

When focusing on a particular element  $T$ , the unit normal is oriented outward from  $T$ , and the unit tangent is oriented counterclockwise.

We denote  $\mathbf{n}_{\partial T}$  and  $\mathbf{t}_{\partial T}$  when considering only an element  $T \in \mathcal{T}_h$ , and  $\mathbf{n}_{T,F}$ ,  $\mathbf{t}_{T,F}$  when considering them on  $F \in \mathcal{F}_T$ .



(a) Normal and tangent unit vector of facet  $F$  seen from  $T_1$ .



(b) Normal and tangent unit vector of facet  $F$  seen from  $T_2$ .

Figure 7: Orientation of normal and tangent unit vectors depending on the chosen element.

We will sometimes need to define quantities with respect to a given vertex. The convention is as follows: on the vertex  $a_i$  of a triangle  $T$ , the face  $F_i$  considered will be the opposite face. The numbering of the faces is fixed arbitrarily in the counterclockwise direction.

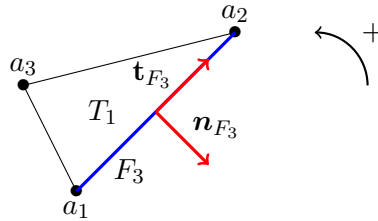
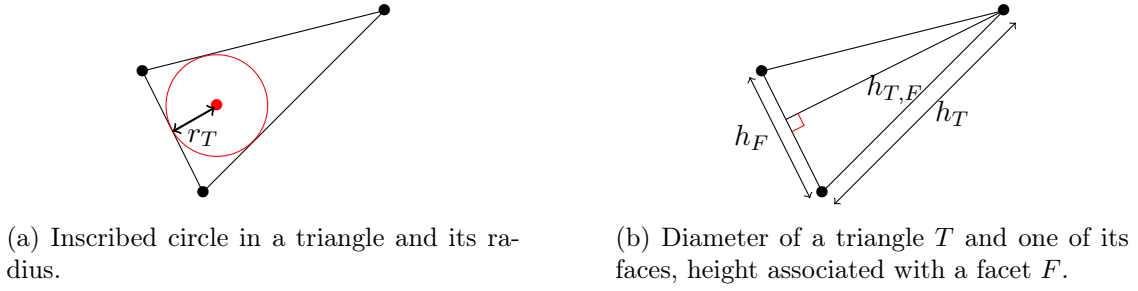


Figure 8: In blue, the face  $F_3$  opposite to vertex  $a_3$  (also in blue),  $\mathbf{n}_{F_3}$  its unit normal vector, and  $\mathbf{t}_{F_3}$  its unit tangential vector, with  $T_1$  and  $T_2$  being the elements sharing  $F_3$ .

In the following, some properties will depend on the mesh regularity. In our framework, this notion is much simpler than for a more general mesh; see [DE12, Section 1.4.1].

We define  $r_T$  as the radius of the inscribed circle (ball) in  $T$ ,  $h_T := \max_{x,y \in T} \|x - y\|$  as the diameter of the simplex  $T$ , and  $h_F := |F|$  as the length of a face for  $d \geq 2$ , with  $h_F = \min(h_{T_1}, h_{T_2})$  for  $d = 1$ . We denote by  $h_{T,F}$  the height of the vertex opposite to  $F$ . We define  $h = \max_{T \in \mathcal{T}_h} h_T$  as the mesh diameter, and we denote by  $(\mathcal{T}_h)_{h>0}$  the formal sequence of meshes as the refinement tends to 0. We define the

Figure 9: Geometric quantities for simplices of dimension  $d = 2$ .

mesh regularity  $\sigma \in ]0; +\infty]$  as the largest  $\hat{\sigma}$  independent of  $h > 0$  such that for all  $h > 0$ :

$$\forall T \in \mathcal{T}_h, \quad \sigma \leq \frac{r_T}{h_T}. \quad (4)$$

In what follows, we assume  $\sigma < +\infty$ , meaning that the diameters of the simplices are comparable to the radii of their inscribed balls.

### 3) Space of discontinuous finite elements

#### a) Broken polynomials

We denote by  $\mathbb{P}_d^k(\Omega)$  the space of polynomials of total degree  $k \in \mathbb{N}$ :

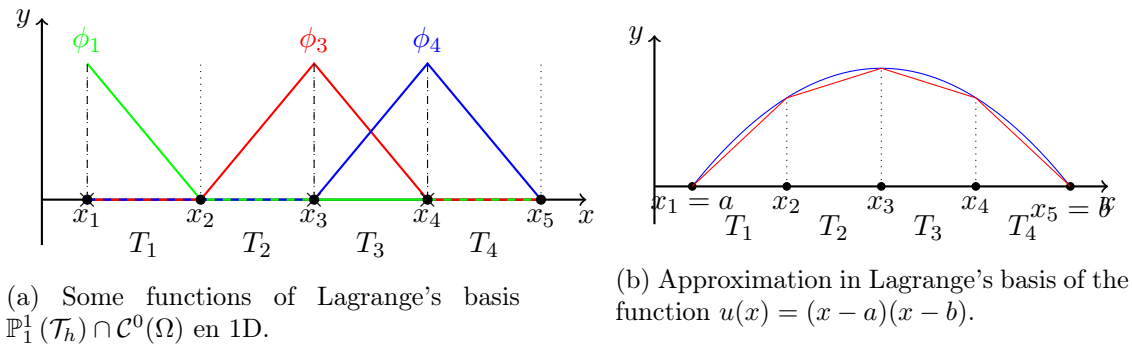
$$\mathbb{P}_d^k(\Omega) := \left\{ P : \Omega \rightarrow \mathbb{R} \mid \exists n \leq k, \quad P(x) = \sum_{|\alpha| \leq n} a_\alpha x^\alpha, \quad \forall \alpha, a_\alpha \in \mathbb{R} \right\},$$

where  $\alpha$  is a multi-index and  $x^\alpha = x_1^{\alpha_1} \dots x_d^{\alpha_d}$ . We then define the space of broken polynomials:

$$\mathbb{P}_d^k(\mathcal{T}_h) := \left\{ V \in L^2(\Omega) \mid \forall T \in \mathcal{T}_h, \quad v|_T \in \mathbb{P}_d^k(T) \right\}.$$

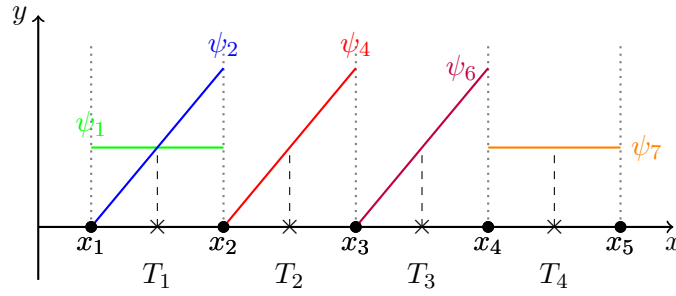
A standard method for the discrete solution of a variational formulation of PDEs is the finite element method. For example, Lagrange finite elements seek the discrete solution  $u_h$  in

$$V_h := \mathbb{P}_d^k(\mathcal{T}_h) \cap \mathcal{C}^0(\Omega).$$

Figure 10: Basis and approximation of the continuous finite elements  $\mathbb{P}_1^1(\mathcal{T}_h) \cap \mathcal{C}^0(\Omega)$ .

In the discontinuous finite element method, discontinuities of the discrete solution are allowed within the mesh itself. This notably reduces the stencil size and thus yields sparser matrices. The discrete solution  $u_h$  is then sought in

$$V_h = \mathbb{P}_d^k(\mathcal{T}_h).$$

Figure 11: Some basis functions of  $\mathbb{P}_1^1(\mathcal{T}_h)$  en 1D.

### b) Jumps and averages inside the mesh

Discontinuities within the mesh are characterized by the jump and the average at interfaces:

#### Definition 2.1: Jump and average

Consider two elements  $T_1$  and  $T_2$ , and let  $F \in \mathcal{F}_{T_1} \cap \mathcal{F}_{T_2}$  be their common face. The termwise average of  $v$  on  $F$  is defined by

$$\llbracket v \rrbracket_F(x) := \frac{1}{2} \left( \text{Tr}_F(v|_{T_1})(x) + \text{Tr}_F(v|_{T_2})(x) \right),$$

and the termwise jump of  $v$  on  $F$  is defined by

$$[[v]]_F(x) := \text{Tr}_F(v|_{T_1})(x) - \text{Tr}_F(v|_{T_2})(x).$$

In the following, we will write  $\llbracket \cdot \rrbracket$  and  $[[ \cdot ]]$  when there is no ambiguity about the considered face.

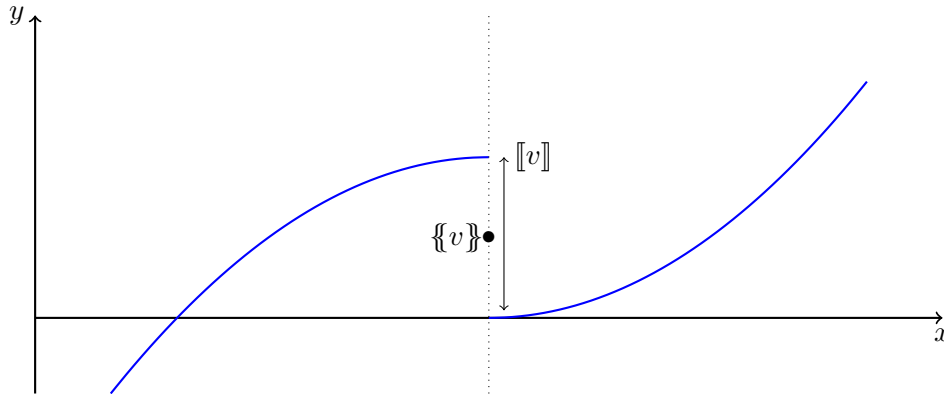


Figure 12: Jump and average of a discontinuous function.

If  $F \in \mathcal{F}_T^b$ , i.e.  $F$  is a boundary face on  $\partial\Omega$ , there is no element  $T_2$  to define  $\llbracket v \rrbracket$  or  $[[v]]$ ; we then set by convention:

$$\llbracket v \rrbracket = [[v]] := \text{Tr}_F(v).$$

#### Remark :

If  $v \in H^s(T)$  with  $s \leq 1/2$ , it is not always possible to define a trace  $L^2$ , which motivates the assumption  $v \in H^s(T)$  with  $s > 1/2$  for all considered functionals, see [DE12, Remark 2.16].

If  $u \in H^s(\Omega)$ ,  $s > 3/2$ , then  $u$  and  $\nabla u$  satisfy  $[[u]] = 0$  and  $[[\nabla_h u]] = 0_{\mathbb{R}^d}$ , see [DE12, Lemma 1.23].

### c) Broken Sobolev spaces

Allowing discontinuous discrete solutions makes the variational formulation (2) invalid!

Indeed, if  $u_h$  is a discontinuous function, its gradient satisfies  $\nabla u_h = \delta_x \notin \mathbf{L}^2$  at the discontinuities.

To overcome this issue, it is necessary to redefine Sobolev spaces on the mesh to give meaning to the functions we manipulate and to define a new gradient: the broken gradient.

We define the broken Sobolev spaces

$$\mathbf{H}^s(\mathcal{T}_h) := \{v \in \mathbf{L}^2(\Omega) \mid \forall T \in \mathcal{T}_h, v|_T \in \mathbf{H}^s(T)\},$$

which then allows us to define the broken gradient for  $v \in \mathbf{H}^s(\mathcal{T}_h)$  such that for any  $T \in \mathcal{T}_h$

$$(\nabla_h v)|_T := \nabla(v|_T).$$

In the following, when there is no ambiguity, we will write inside  $T$ :  $\nabla_h v = \nabla v$ .

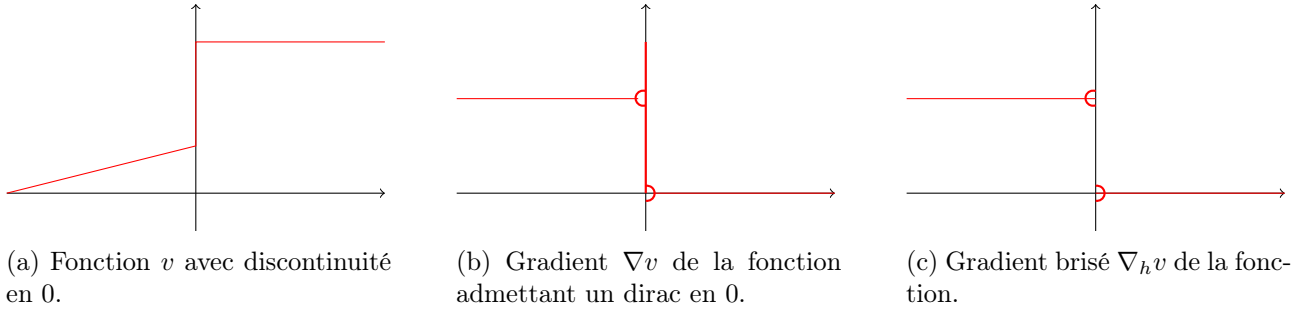


Figure 13: Comparison between the gradient and the broken gradient.

We define analogously the broken space of  $\mathbf{H}(\text{div}; \Omega)$ :

$$\mathbf{H}(\text{div}; \mathcal{T}_h) := \{\tau \in \mathbf{L}^2(\Omega) \mid \forall T \in \mathcal{T}_h, \text{div}(\tau|_T) \in \mathbf{L}^2(T)\},$$

which will be useful later for defining fluxes.

The algebraic properties and the strong relationships between these spaces and  $\mathbf{H}^s(\Omega)$  and  $\overline{\mathbf{H}}^s(\text{div}; \Omega) = \mathbf{H}(\text{div}; \Omega) \cap \mathbf{H}^s(\Omega)$  are well studied in [DE12, Sections 1.2.5, 1.2.6 and 5.1].

### d) Inequalities

We recall here the inequalities essential for the analysis of discretization errors.

#### Remark :

When the constants in inequalities are independent of  $h, \sigma$  and are not required for the method analysis, they will be omitted using the symbol  $\lesssim$ :

$$\exists C > 0, X \leq CY \iff X \lesssim Y.$$

**Lemma 2.2: Discrete trace inequality**

Let  $k \in \mathbb{N}$ . There exists  $C_{\text{tr}} = C_{\text{tr}}(d, k) > 0$  such that for all  $h > 0$ , for all  $v \in \mathbb{P}_d^k(\mathcal{T}_h)$ , and for all  $T \in \mathcal{T}_h$  we have

$$h_T^{1/2} \|v_h\|_{L^2(F)} \leq C_{\text{tr}} \sigma^{-1/2} \|v_h\|_{L^2(T)}.$$

**D: Lemma 2.2**

The proof in [WH03, Theorems 4 and 5] shows the existence of  $C = C(d, k) > 0$  such that

$$\|v_h\|_{L^2(F)} \leq C \left( \frac{|\partial T|}{|T|} \right)^{1/2} \|v_h\|_{L^2(T)}.$$

Using [CR73, Inequality (3.17)], we obtain

$$\|v_h\|_{L^2(F)} \leq C(r_T)^{-1/2} \|v_h\|_{L^2(T)}.$$

Then, using the definition of  $\sigma$  from Equation (4), we have

$$\|v_h\|_{L^2(F)} \leq C(\sigma h_T)^{-1/2} \|v_h\|_{L^2(T)},$$

which gives the result.  $\square$

**Lemma 2.3: Inverse inequality**

Let  $k \in \mathbb{N}$ . There exists  $C_{\text{inv}} = C_{\text{inv}}(d, k) > 0$  such that for all  $h > 0$ , for all  $v_h \in \mathbb{P}_d^k(\mathcal{T}_h)$ , and for all  $T \in \mathcal{T}_h$  we have

$$\|\nabla v_h\|_{L^2(T)} \leq C_{\text{inv}} \sigma^{-1} h_T^{-1} \|v_h\|_{L^2(T)}.$$

**D: Lemma 2.3**

We prove it in the case  $v \in \mathbb{P}^1(T)$  following [ABJ25, Proposition 3]:

$$\begin{aligned} \|\nabla v\|_{L^2(T)}^2 &= \int_T |\nabla v|^2 = |T| |\nabla v|^2 && \text{(since } \nabla v \text{ is constant)} \\ &= |T| \left| |T|^{-1} \int_T \nabla v \right|^2 && \text{(since } \nabla v \text{ is constant)} \\ &= |T|^{-1} \left| \int_{\partial T} v \mathbf{n}_{\partial T} \right|^2 && \text{(by integration by parts)} \\ &\leq |T|^{-1} \left| \int_{\partial T} 1 \times v \right|^2 && \text{(since } \|\mathbf{n}_T\| \leq 1) \\ &\leq |T|^{-1} |\partial T| \|v\|_{L^2(\partial T)}^2 && \text{(by Cauchy-Schwarz)} \\ &\leq |T|^{-1} |\partial T| \sigma^{-1} h_T^{-1} \|v\|_T^2 && \text{(by Lemma 2.2)} \\ &\leq C_{\text{inv}}^2 \sigma^{-2} h_T^{-2} \|v\|_T^2 && \text{(by [CR73, Inequality (3.17)])}. \end{aligned}$$

$\square$



**Lemma 2.4: Approximation by  $L^2$ -orthogonal projection**

Let  $\pi_h$  be the  $L^2$ -orthogonal projection of  $v \in H^{k+1}(\mathcal{T}_h)$  onto  $\mathbb{P}_d^k(\mathcal{T}_h)$ . There exists a constant  $C_{\text{app}} > 0$  independent of  $T$  and  $h$  such that

$$|v - \pi_h v|_{H^m(T)} \leq C_{\text{app}} h_T^{k+1-m} \sigma^m |v|_{H^{k+1}(T)},$$

where  $|v|_{H^k(\Omega)}$  is the  $H^k(\Omega)$  semi-norm.

**D: Lemma 2.4**

We rely on [EG21, Lemma 11.9], known as the  $\mathbb{P}^k$  Bramble-Hilbert/Deny-Lions lemma, which states that for  $S$  a Lipschitz domain in  $\mathbb{R}^d$ ,  $k \in \mathbb{N}$ , there exists  $c > 0$  such that

$$\inf_{q \in \mathbb{P}_d^k(S)} \|v - q\|_{H^{k+1}(S)} \leq c |v|_{H^{k+1}(S)}$$

and on the property of affine transformations from the reference element  $\hat{T}$ .

Let  $\psi_T : \hat{T} \rightarrow T$  be the affine function transforming the reference element  $\hat{T}$  into  $T$  in an invertible manner, so that a functional operator can be written as

$$\phi(v) = \mathbb{A}_T(v \circ \psi_T),$$

where  $\mathbb{A}_T$  is an invertible matrix. We define  $\mathbb{J} = D\psi_T$  as the Jacobian matrix of  $\psi$ . According to [EG21, Lemma 11]:

$$\|\mathbb{J}\|_{\ell^2} \|\mathbb{J}^{-1}\|_{\ell^2} \lesssim \sigma.^a$$

Note that the  $L^2$  projection on  $T$ , denoted  $\pi_T$ , satisfies

$$\pi_T = \psi_T^{-1} \circ \pi_{\hat{T}} \circ \psi_T.$$

The proof of [EG21, Theorem 11.13] then shows that for  $v \in H^{k+1}(T)$

$$|v - \pi_h v|_{H^{k+1}(T)} \leq \|\mathbb{J}^{-1}\|_{\ell^2}^m |v - \pi_{\hat{T}}(v)|_{H^{k+1}(T)} \lesssim \|\mathbb{J}\|_{\ell^2}^{k+1} \|\mathbb{J}^{-1}\|_{\ell^2}^m |v|_{H^{k+1}(T)} \lesssim \sigma^m h^{k+1-m} |v|_{H^{k+1}(T)}.$$

□

---

<sup>a</sup> $\|\cdot\|_{\ell^2}$  is the norm subordinate to the 2-norm in  $\mathbb{R}^d$ .

## II - Variational Formulations

### 1) Symmetric Interior Penalty

We start here from the variational formulation (2) by replacing the gradients with broken gradients: for  $v_h, w_h \in V_h = \mathbb{P}_d^k(\mathcal{T}_h)$ ,

$$a_h^{(0)}(v_h, w_h) = \int_{\Omega} \nabla_h v_h \cdot \nabla_h w_h.$$

The idea is then to check whether this formulation is consistent with (2). We have:

$$\begin{aligned}
a_h^{(0)}(v_h, w_h) &= \int_{\Omega} \nabla_h v_h \cdot \nabla_h w_h \\
&= \sum_{T \in \mathcal{T}_h} \int_T \nabla v_h \cdot \nabla w_h \\
&= \sum_{T \in \mathcal{T}_h} \int_{\partial T} (\nabla v_h) \cdot \mathbf{n}_{\partial T} w_h - \sum_{T \in \mathcal{T}_h} \int_T \Delta v_h w_h && \text{(by IBP)} \\
&= \sum_{T \in \mathcal{T}_h} \sum_{F \in \mathcal{F}_T} \int_F (\nabla v_h) \cdot \mathbf{n}_{F,T} w_h - \sum_{T \in \mathcal{T}_h} \int_T \Delta v_h w_h \\
&= \sum_{F \in \mathcal{F}_h} \sum_{T \in \mathcal{T}_F} \int_F (\nabla v_h) \cdot \mathbf{n}_{F,T} w_h - \sum_{T \in \mathcal{T}_h} \int_T \Delta v_h w_h && \text{(looking at each face by triangle or each triangle by face is equivalent)} \\
&= \sum_{F \in \mathcal{F}_h^i} \sum_{T \in \mathcal{T}_F} \int_F (\nabla v_h) \cdot \mathbf{n}_{F,T} w_h \\
&\quad + \sum_{F \in \mathcal{F}_h^b} \sum_{T \in \mathcal{T}_F} \int_F (\nabla v_h) \cdot \mathbf{n}_F w_h - \sum_{T \in \mathcal{T}_h} \int_T \Delta v_h w_h \\
&= \sum_{F \in \mathcal{F}_h^i} \int_F (\nabla v_h) \cdot \mathbf{n}_F w_h + \int_F (\nabla v_h) \cdot (-\mathbf{n}_F) w_h \\
&\quad + \sum_{F \in \mathcal{F}_h^b} \int_F \{\nabla_h v \cdot \mathbf{n}_F\} \llbracket w_h \rrbracket - \sum_{T \in \mathcal{T}_h} \int_T \Delta v_h w_h && \text{(by Fig. 7, Def. 2.1)} \\
&= \sum_{F \in \mathcal{F}_h^i} \int_F \llbracket (\nabla_h v_h) w_h \rrbracket \cdot \mathbf{n}_F + \sum_{F \in \mathcal{F}_h^b} \int_F \{\nabla_h v \cdot \mathbf{n}_F\} \llbracket w_h \rrbracket \\
&\quad - \sum_{T \in \mathcal{T}_h} \int_T \Delta v_h w_h \\
&= \sum_{F \in \mathcal{F}_h^i} \int_F \{\nabla_h v_h \cdot \mathbf{n}_F\} \llbracket w_h \rrbracket + \int_F \llbracket \nabla_h v_h \cdot \mathbf{n}_F \rrbracket \{w_h\} \\
&\quad + \sum_{F \in \mathcal{F}_h^b} \int_F \{\nabla_h v \cdot \mathbf{n}_F\} \llbracket w_h \rrbracket - \sum_{T \in \mathcal{T}_h} \int_T \Delta v_h w_h && \text{(by identity (5))} \\
&= \sum_{F \in \mathcal{F}_h} \int_F \{\nabla_h v_h \cdot \mathbf{n}_F\} \llbracket w_h \rrbracket + \sum_{F \in \mathcal{F}_h^i} \int_F \llbracket \nabla_h v_h \cdot \mathbf{n}_F \rrbracket \{w_h\} \\
&\quad - \sum_{T \in \mathcal{T}_h} \int_T \Delta v_h w_h.
\end{aligned}$$

where we used the identity

$$\llbracket fg \rrbracket = \{f\} \llbracket g \rrbracket + \llbracket f \rrbracket \{g\}. \quad (5)$$

We would then like to evaluate  $a_h^{(0)}(\cdot, w_h)$  at the solution  $u \in H_0^1(\Omega) \cap H^s(\Omega)$ , however  $a_h^{(0)}$  is defined only for functions defined locally on each mesh element: in other words  $u \notin V_h$ !

We assume  $a_h^{(0)}$  can be extended to  $V := H^{3/2+\varepsilon}(\Omega) \cap H_0^1(\Omega)$  for  $\varepsilon > 0$ , and in preparation for the error analysis  $u - u_h$ , we consider the enriched space  $V_{h,\star} := V + V_h$ . Then for  $v \in V_{h,\star}$ :

$$a_h^{(0)} : (v, w_h) \in V_{h,\star} \times V_h \longmapsto \sum_{F \in \mathcal{F}_h} \int_F \{\{\nabla_h v \cdot \mathbf{n}_F\}\} [w_h] + \sum_{F \in \mathcal{F}_h^i} \int_F \{\{\nabla_h v \cdot \mathbf{n}_F\}\} [w_h] - \sum_{T \in \mathcal{T}_h} \int_T \Delta v w_h. \quad (6)$$

### Definition 2.5: Consistency

Let  $a_h : V_{h,\star} \times V_h \longrightarrow \mathbb{R}$  be a *discrete* bilinear form derived from problem (2). The bilinear form is said to be consistent with the problem if for the solution  $u \in V$ :

$$a_h(u, w_h) = \int_{\Omega} f w_h, \quad \forall w_h \in V_h,$$

i.e., the discrete formulation gives the same equation as the continuous formulation when evaluated at the solution  $u$  of the problem.

We then evaluate at the solution of the Poisson problem  $u \in V_{h,\star}$  using Remark p.14:

$$a_h^{(0)}(u, w_h) = \sum_{F \in \mathcal{F}_h} \int_F \nabla u \cdot \mathbf{n}_F [w_h] - \sum_{T \in \mathcal{T}_h} \int_T \Delta u w_h \quad (7)$$

$$= \sum_{F \in \mathcal{F}_h} \int_F \nabla u \cdot \mathbf{n}_F [w_h] - \int_{\Omega} \Delta u w_h \quad (8)$$

$$= \sum_{F \in \mathcal{F}_h} \int_F \nabla u \cdot \mathbf{n}_F [w_h] + \int_{\Omega} f w_h. \quad (9)$$

We notice in (9) that the bilinear form  $a_h^{(0)}$  is not consistent. To make it consistent, we modify it by removing the excess term:

$$a_h^c(v, w_h) = \int_{\Omega} \nabla_h v \cdot \nabla_h w_h - \sum_{F \in \mathcal{F}_h} \int_F \{\{\nabla_h v \cdot \mathbf{n}_F\}\} [w_h].$$

The bilinear form  $a_h^c$  is then **consistent**, however note that if  $v_h, w_h \in V_h$ :

$$a_h^c(w_h, v_h) := \int_{\Omega} \nabla_h v_h \cdot \nabla_h w_h - \sum_{F \in \mathcal{F}_h} \int_F \{\{\nabla_h w_h \cdot \mathbf{n}_F\}\} [v_h] \neq a_h^c(v_h, w_h).$$

The bilinear form  $a_h^c$  is not symmetric. Since linear system solvers behave better on symmetric systems, and for error analysis, we define a new consistent symmetric bilinear form  $a_h^{cs}$  by adding a consistent term:

$$a_h^{cs}(v, w_h) = \int_{\Omega} \nabla_h v \cdot \nabla_h w_h - \sum_{F \in \mathcal{F}_h} \int_F (\{\{\nabla_h v \cdot \mathbf{n}_F\}\} [w_h] + \{\{\nabla_h w_h \cdot \mathbf{n}_F\}\} [v]).$$

Then for all  $v_h, w_h \in V_h$  and for  $u \in V$  the solution of the problem:

$$a_h^{cs}(u, w_h) = \int_{\Omega} f w_h \text{ and } a_h^{cs}(v_h, w_h) = a_h^{cs}(w_h, v_h).$$

Finally, we want discrete coercivity on  $V_h \times V_h$  for the existence of the discrete solution  $u_h$ , but:

$$a_h^{cs}(v_h, v_h) = \|\nabla_h v_h\|_{L^2(\Omega)}^2 - 2 \sum_{F \in \mathcal{F}_h} \int_F \{\{\nabla_h v_h \cdot \mathbf{n}_F\}\} [v_h] < 0 \text{ for some } v_h.$$

A good approach is first to define a norm on  $V_{h,\star}$ . Indeed, the semi-norm of the broken gradient is not a norm as it does not account for the magnitude of the jumps at internal mesh interfaces.

**Definition 2.6: SIP Norm**

We define a mesh-dependent norm on  $V_{h,\star}$ , which will be called the SIP norm:

$$\|v\|_{\text{sip}}^2 := \|\nabla_h v\|_{\mathbf{L}^2(\Omega)}^2 + |v|_J^2, \quad (10)$$

with

$$|v|_J^2 := \sum_{F \in \mathcal{F}_h} \frac{1}{h_F} \|\llbracket v \rrbracket\|_{\mathbf{L}^2(F)}^2. \quad (11)$$

The semi-norm  $|v|_J$  is the jump norm and quantifies the size of the function discontinuities.

**D: The norm is well-defined**

It is easy to show that  $\|v\|_{\text{sip}}$  is homogeneous and satisfies the triangle inequality. The only difficulty lies in definiteness:

- If  $\|v\|_{\text{sip}} = 0$  then  $\|\nabla_h v\|_{\mathbf{L}^2(\Omega)} = 0$  and  $|v|_J = 0$ .
- If  $\|\nabla_h v\|_{\mathbf{L}^2(\Omega)} = 0$  then  $\nabla v = 0$  on each element  $T$ , so  $v$  is piecewise constant.
- $|v|_J = 0$  implies  $\|\llbracket v_h \rrbracket\|_{\mathbf{L}^2(F)} = 0$  on all faces  $F$ , i.e.,  $v$  is continuous.
- But  $v = 0$  on  $\partial\Omega$ , hence  $v \equiv 0$  in  $\Omega$ .

Thus the norm is definite. We conclude that  $\|\cdot\|_{\text{sip}}$  is indeed a norm.  $\square$

We then add a stabilization bilinear form that remains consistent and symmetric

$$s_h(v, w_h) = \sum_{F \in \mathcal{F}_h} \frac{\gamma_F}{h_F} \int_F \llbracket v \rrbracket \llbracket w_h \rrbracket.$$

The penalty  $\gamma$  is a locally defined function for each  $F \in \mathcal{F}_h$  such that  $\gamma_F = \gamma|_F > 0$ , a constant depending on the mesh and user choice [DE12, Lemma 4.12] and [WH03].

We notice that  $s_h(v_h, v_h) = \gamma |v_h|_J^2$ ,  $s_h$  is actually a penalty on interior jumps, since the sought solution has no internal mesh discontinuities.

We thus define the bilinear form  $a_h^{\text{sip}}$  called **Symmetric Interior Penalty**, abbreviated SIP:

$$a_h^{\text{sip}}(v, w_h) = \int_{\Omega} \nabla_h v \cdot \nabla_h w_h - \sum_{F \in \mathcal{F}_h} \int_F \left( \underbrace{\{\nabla_h v \cdot \mathbf{n}_F\} \llbracket w_h \rrbracket}_{\text{consistency}} + \underbrace{\{\nabla_h w_h \cdot \mathbf{n}_F\} \llbracket v \rrbracket}_{\text{symmetry}} \right) + \underbrace{s_h(v, w_h)}_{\text{interior penalty}} \quad (12)$$

We then have the following lemma:

**Lemma 2.7: Discrete coercivity of SIP**

For all  $\gamma \geq \underline{\gamma} := C_{\text{tr}}^2 N_{\partial}$ , we have

$$\forall v_h \in V_h, \quad a_h^{\text{sip}}(v_h, v_h) \geq C_{\gamma} \|v_h\|_{\text{sip}}^2,$$

with  $C_{\gamma} := (\gamma - C_{\text{tr}}^2 N_{\partial})(1 + \gamma)^{-1}$  and  $N_{\partial} := \max_{T \in \mathcal{T}_h} \text{Card}(\mathcal{F}_T) = d + 1$  for simplices.

**D: Lemma 2.7**[DE12, Lemma 4.12]. □

We then study the continuity of  $a_h^{\text{sip}}$  for this norm. Using the previous optimal inequalities:

$$\left| a_h^{\text{sip}}(v, w_h) \right| \leq \left( \|v\|_{\text{sip}}^2 + \sum_{T \in \mathcal{T}_h} h_T \|\nabla v|_T \cdot \mathbf{n}_T\|_{L^2(\partial T)}^2 \right)^{1/2} \|w_h\|_{\text{sip}}.$$

We enrich the SIP norm on  $V_{h,\star}$  to a norm  $\|v\|_{\text{sip},\star}^2 := \|v\|_{\text{sip}}^2 + \sum_{T \in \mathcal{T}_h} h_T \|\nabla v|_T \cdot \mathbf{n}_T\|_{L^2(\partial T)}^2$  so as to make the bilinear form  $a_h^{\text{sip}}$  continuous.

The bilinear form  $a_h^{\text{sip}} : (V_{h,\star}, \|\cdot\|_{\text{sip},\star}) \times (V_h, \|\cdot\|_{\text{sip}}) \mapsto \mathbb{R}$  is therefore:

- continuous,
- coercive,
- consistent.

We note that  $\ell : w_h \in V_h \mapsto \langle f | w_h \rangle_{L^2(\Omega)}$  remains continuous because  $\|\cdot\|_{L^2(\Omega)} \lesssim \|\cdot\|_{\text{sip}} \leq \|\cdot\|_{\text{sip},\star}$ .

**Theorem 2.8: Quasi-optimality**

If  $u$  is the solution of the Poisson problem,  $u_h$  the discrete solution. If the penalty satisfies  $\gamma \geq \underline{\gamma}$ :

$$\|u - u_h\|_{\text{sip}} \lesssim \inf_{v_h \in V_h} \|u - v_h\|_{\text{sip},\star} \leq \|u - u_h\|_{\text{sip}}.$$

The scheme is quasi-optimal in the sense that the error is equivalent to the optimal error.

**D: Theorem 2.8**[DE12, Théorème 4.17]. □

The following corollary then directly follows:

**Corollary 2.9: A priori error estimate**

Under the hypotheses of Theorem 2.8 and  $u \in H^{k+1}(\Omega)$ :

$$\|u - u_h\|_{\text{sip}} \lesssim \|u\|_{H^{k+1}(\Omega)} h^k.$$

**D: Corollary 2.9**[DE12, Corollaire 4.18]. □

We notice that the SIP method does not converge if  $k = 0$ , so in numerical computations we will impose  $k \geq 1$ .

## 2) Symmetric Weighted Interior Penalty

### a) Generalization of the continuous problem

We consider an open set  $\Omega \subset \mathbb{R}^d$  with a Lipschitz boundary, data  $f \in L^2(\Omega)$  and  $\kappa \in L^\infty(\Omega)$ ,  $\kappa \geq \alpha > 0$  a scalar.

We are interested in the generalized diffusion problem:

$$\text{Find } u \in H^1(\Omega) \text{ such that } \begin{cases} -\operatorname{div}(\kappa \nabla u) = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \quad (13)$$

By integration by parts, (13) is equivalent to:

$$\text{Find } u \in H^1(\Omega), \quad \int_{\Omega} (\kappa \nabla u) \cdot \nabla w = \int_{\Omega} f w, \quad \forall w \in H_0^1(\Omega). \quad (14)$$

We can define the bilinear and linear forms associated with the problem:

$$a(v, w) = \int_{\Omega} (\kappa \nabla v) \cdot \nabla w, \quad \ell(w) = \int_{\Omega} f w = \langle f | w \rangle_{L^2(\Omega)}.$$

Then by Cauchy-Schwarz

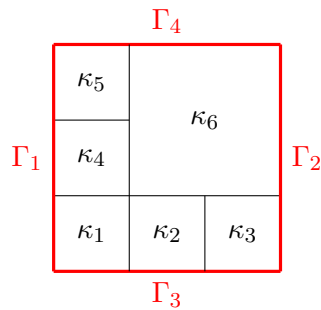
$$|a(v, w)| \leq \|\kappa\|_{L^\infty} \|v\|_{L^2(\Omega)} \|w\|_{L^2(\Omega)}.$$

and by Poincaré

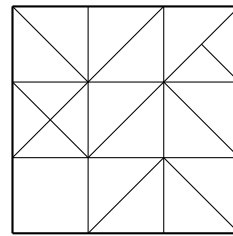
$$a(v, v) \geq \alpha \|\nabla v\|_{L^2(\Omega)}^2 \gtrsim \alpha \|v\|_V^2.$$

thus by the Lax-Milgram theorem, the problem is well-posed.

In practice,  $\kappa$  will be more than  $L^\infty$  and we will assume, possibly approximating  $\kappa$ , that it is a piecewise constant function on a polyhedral partition of  $\Omega$ :



(a) Square domain  $\Omega$  partitionned in subdomains.



(b) Admissible mesh of  $\Omega$ .

Figure 14: Example of partitioned domain according to  $\kappa$  and a possible mesh

We can assume  $\kappa$  is piecewise smooth at the cost of some additional technicalities, but we cannot allow a mesh without a partition of  $\Omega$  because the solution  $u$  will not be regular at the jumps of  $\kappa$ .

The heterogeneous diffusion problem can also be expressed in a mixed formulation involving fluxes:

$$\text{Find } \sigma(u) \in \mathbf{H}(\operatorname{div}; \Omega), u \in H(\Omega) \text{ such that } \begin{cases} -\operatorname{div} \sigma = f & \text{in } \Omega, \\ \sigma(u) = \kappa \nabla(u) & \text{in } \Omega. \end{cases} \quad (15)$$

### b) Weighted averages

At the interfaces, it is necessary to weight the averages to handle cases where  $\kappa_i \gg \kappa_j$ .

The expected behavior is as follows: if  $\kappa_1 \gg \kappa_2$  then the values from  $T_2$  tend to diffuse into  $T_1$ , so the

average value should be more weighted towards  $T_2$ , i.e.  $\omega_1 \xrightarrow[\kappa_1 \gg \kappa_2]{} 0$  and  $\omega_2 \xrightarrow[\kappa_1 \gg \kappa_2]{} 1$ .

Thus, we define the weighted averages on a face  $F \in \mathcal{F}_h^i$  separating two elements  $T_1$  and  $T_2$ :

$$\{v\}_{F,\omega}(x) = \omega_1 v|_{T_1}(x) + \omega_2 v|_{T_2}(x),$$

with

$$\omega_1 + \omega_2 = 1.$$

If  $T \in \mathcal{T}_h$  and  $F \in \mathcal{F}_h^b \cap \mathcal{F}_T$ , we will always write  $\{v\}_{F,\omega}(x) = v|_T$ . If there is no confusion about the face  $F \in \mathcal{F}_h$ , we will write  $\{v\}_\omega(x)$ .

To account for the physics and thanks to the error analysis, a good choice is

$$\omega_1 := \frac{\kappa_2}{\kappa_1 + \kappa_2} \quad \omega_2 := \frac{\kappa_1}{\kappa_1 + \kappa_2}.$$

If  $\kappa_1 = \kappa_2$ , we recover the arithmetic mean, which corresponds to using the SIP method.

However, when modifying the averages, we also need to modify the way we penalize to maintain a coercive method with the  $\|\cdot\|_{\text{dG}}$  norm corresponding to the energy norm of the variational formulation, see [DE12, Section 4.5.3]. We then set

$$s_h^\kappa(v, w_h) = \sum_{F \in \mathcal{F}_h} \frac{\gamma_F}{h_F} \gamma_\kappa \int_F \llbracket v \rrbracket \llbracket w_h \rrbracket,$$

where

$$\gamma_\kappa := \frac{2\kappa_1\kappa_2}{\kappa_1 + \kappa_2}.$$

which also recovers the SIP penalization if  $\kappa_1 = \kappa_2$ . For  $F \subset \partial\Omega$ , we set  $\gamma_\kappa = \kappa_T$ .

The jumps must be controlled relative to  $\gamma_\kappa$ , so we introduce a new jump norm:

$$|v|_{J,\kappa}^2 := \sum_{F \in \mathcal{F}_h} \frac{1}{h_F} \gamma_\kappa \|v\|_{L^2(F)}^2.$$

This notably implies modifying the SIP norm into a SWIP norm

$$\|v\|_{\text{swip}}^2 := \left\| \kappa^{1/2} \nabla_h v \right\|_{L^2(\Omega)}^2 + |v|_{J,\kappa}^2. \quad (16)$$

We then define the bilinear form **Symmetric Weighted Interior Penalty** abbreviated SWIP

$$a_h^{\text{swip}}(v, w_h) = \int_\Omega \kappa \nabla_h v \cdot \nabla_h w_h - \sum_{F \in \mathcal{F}_h} \int_F (\{ \nabla_h v \cdot \mathbf{n}_F \}_\omega \llbracket w_h \rrbracket + \{ \nabla_h w_h \cdot \mathbf{n}_F \}_\omega \llbracket v \rrbracket) + s_h^\kappa(v, w_h) \quad (17)$$

### Theorem 2.10: A priori error

If  $u \in H^{k+1}(\Omega)$  is the solution of the heterogeneous diffusion problem then

$$\|u - u_h\|_{\text{swip}} \lesssim \|\kappa\|_{L^\infty(\Omega)}^{1/2} \|u\|_{H^{k+1}(\Omega)} h^k. \quad (18)$$

### D: Theorem 2.10

[DE12, Theorem 4.53]. □

[DE12, Remark 4.19] gives the equivalence between the SWIP norm and the broken gradient norm, i.e. for a sufficiently large penalization coefficient  $\gamma > 0$ , there exists  $C > 0$  such that if  $v_h \in H_0^1(\mathcal{T}_h)$  then

$$\sum_{T \in \mathcal{T}_h} \left\| \kappa^{1/2} \nabla v \right\|_{\mathbf{L}^2(T)}^2 \leq \|v\|_{\text{swip}}^2 \leq C \sum_{T \in \mathcal{T}_h} \left\| \kappa^{1/2} \nabla v \right\|_{\mathbf{L}^2(T)}^2$$

and

$$|v_h|_{J,\kappa}^2 \xrightarrow{h \rightarrow 0} 0.$$

It is therefore sufficient to estimate  $\sum_{T \in \mathcal{T}_h} \left\| \kappa^{1/2} \nabla v \right\|_{\mathbf{L}^2(T)}^2$  to obtain an error estimate for a discontinuous Galerkin method, up to data oscillation (which is zero for  $v_h \in H_0^1(\mathcal{T}_h)$ ).

**Remark :**

[DE12, Section 4.5.1.2] indicates that a regularity of the solution  $u \in H^{1+s}(\Omega)$ ,  $s > 0$ , is expected and sufficient for the SWIP method.

### Theorem 2.11: A priori error estimate for low-regularity solutions

If the solution  $u \in H^{1+\alpha}(\Omega)$  with  $\alpha < 1$  and  $u_h \in \mathbb{P}_1^1(\mathcal{T}_h)$  then there exists  $C_u > 0$  such that:

$$\|u - u_h\|_{\text{swip}} \leq C_u h^\alpha.$$

### $\mathcal{D}$ : Theorem 2.11

[DE12, Section 4.5.4], [PE11, Theorem 3.6] and [CYZ11]. □



## Part 3

# Posteriori Error Estimators

We consider  $\Omega \subset \mathbb{R}^2$  in the following, but the results can also be adapted to  $\mathbb{R}^d$ .

We define the broken flux  $\boldsymbol{\sigma}_h(v)|_T := \boldsymbol{\sigma}(v|_T) = \kappa \nabla(v|_T)$  and  $\boldsymbol{\sigma}_h(e_h)$  the broken flux of the total error  $e_h := u - u_h$ .

Since the error  $e_h$  does not belong to the solution space  $V$  as explained for equation (6), we want to split the error into two parts:

- The non-conforming error  $\epsilon_{h,\text{nonconforming}}$  which is due to the fact that the error does not belong to  $V$ , the space in which the solution exists.
- The conforming error  $\epsilon_{h,\text{conforming}}$  which corresponds to the numerical precision between the discrete solution projected into the space of continuous polynomials and the exact solution.

As specified in equation (18), we then want to estimate  $\|\boldsymbol{\sigma}_h(e_h)\|_{\mathbf{L}^2(T)}$  for  $T \in \mathcal{T}_h$  a mesh element.

Suppose there exists an  $L^2$ -orthogonal projection from the solution space  $V$  into the space of broken polynomials  $V_h = \mathbb{P}_2^k(\mathcal{T}_h)$ , then by noting:

- $u \in V$  the exact solution.
- $u_h \in V_h$  the discrete solution.
- $u_{h,\text{conforming}} \in V_h$  the  $L^2$ -orthogonal projection of  $u$  from  $V$  into  $V_h$ .
- $u_{h,\text{nonconforming}} := u_h - u_{h,\text{conforming}} \in V_h \setminus V$  the non-conforming part of  $u$ .

We write in the case of a flux:

$$\begin{aligned} \epsilon_T &:= \|\boldsymbol{\sigma}(e_h)\|_{\mathbf{L}^2(T)}^2 = \|\boldsymbol{\sigma}(u - u_{h,\text{conforming}}) + \boldsymbol{\sigma}(u_{h,\text{conforming}} - u_h)\|_{\mathbf{L}^2(T)}^2 \\ &= \|\boldsymbol{\sigma}(u - u_{h,\text{conforming}})\|_{\mathbf{L}^2(T)}^2 + \|\boldsymbol{\sigma}(u_{h,\text{conforming}} - u_h)\|_{\mathbf{L}^2(T)}^2. \end{aligned}$$

We would then have obtained:

$$\left\| \kappa^{1/2} \nabla e_h \right\|_{\mathbf{L}^2(T)}^2 = \epsilon_{h,\text{conforming}}^2 + \epsilon_{h,\text{nonconforming}}^2.$$

This would allow us to estimate  $\epsilon_{h,\text{conforming}}$  and  $\epsilon_{h,\text{nonconforming}}$  independently, i.e., to find  $\eta_{\text{conforming}} > 0$  and  $\eta_{\text{nonconforming}} > 0$  such that

$$\begin{aligned} \epsilon_{T,\text{conforming}} &\leq \eta_{\text{conforming}}, \\ \epsilon_{T,\text{nonconforming}} &\leq \eta_{\text{nonconforming}}. \end{aligned}$$

with  $\eta$  a quantity depending only on the mesh geometry, the data, and the discrete solution.

For better readability in the equations, we will use the following notation:

- $\eta_{\text{conforming}} = \eta_{\text{CF}}$ ,
- $\eta_{\text{nonconforming}} = \eta_{\text{NC}}$ .

For a fixed element  $T \in \mathcal{T}_h$ , we will use:

$$\begin{aligned} \eta_{\text{CF}}|_T &:= \eta_{\text{CF},T}, \\ \eta_{\text{NC}}|_T &:= \eta_{\text{NC},T}. \end{aligned}$$

## I - Posteriori Estimation

We first consider a posteriori error estimator proposed by Mark AINSWORTH in 2007 [AIN07], who introduced the idea of the numerical flux equilibrium method.

### 1) Problem and Definition of the Estimator

Let  $\Gamma_N, \Gamma_D$  be such that  $\partial\Omega = \overline{\Gamma_N} \cup \overline{\Gamma_D}$  and  $\Gamma_N \cap \Gamma_D = \emptyset$ .

We consider data  $f \in L^2(\Omega)$ ,  $g_D \in H^{1/2}(\Gamma_D)$  and  $g_N \in L^2(\Gamma_N)$ . We start from the mixed formulation (15):

$$\text{Find } \sigma(u) \in \mathbf{H}(\text{div}; \Omega), u \in H^1(\Omega) \text{ such that } \begin{cases} -\text{div } \sigma = f & \text{in } \Omega, \\ \sigma(u) = \kappa \nabla u & \text{in } \Omega, \\ u = g_D & \text{on } \Gamma_D, \\ \sigma \cdot \mathbf{n} = g_N & \text{on } \Gamma_N. \end{cases}$$

Let the Dirichlet solution space be  $H_{g_D}^1(\Omega) := \{u \in H^1(\Omega) \mid u = g_D \text{ on } \Gamma_D\}$ .

We can write the continuous variational formulation of this problem:

$$\text{Find } u \in H_{g_D}^1(\Omega), \quad \langle \kappa \nabla u | \nabla v \rangle_{L^2(\Omega)} = \langle f | v \rangle_{L^2(\Omega)} + \int_{\Gamma_N} g_N v, \quad \forall v \in H_{g_D}^1(\Omega). \quad (19)$$

We then use a discrete variational formulation based on the bilinear form  $a_h^{\text{mixed}}$  defined in Equation (51):

$$\text{Find } u \in H_{g_D}^1(\mathcal{T}_h), \quad a_h^{\text{mixed}} = \ell_h^{\text{mixed}}(v), \quad \forall v \in H_{g_D}^1(\mathcal{T}_h).$$

where  $g_D^h = \Pi_H^1(g_D)$  is the **projection onto the space of piecewise linear continuous functions** on the set of boundary faces of the domain denoted  $\mathbb{P}_c^1(\mathcal{F}_h^b)$ .

#### Remark :

If the problem is solved with  $g_D \notin \mathbb{P}_c^1(\mathcal{F}_h^b)$ , then equation (27) is not exactly satisfied.

[AIN07, Theorem 2] then shows the following a posteriori error result:

#### Theorem 3.1: Decomposed Posteriori Estimation

For every element  $T \in \mathcal{T}_h$  there exist  $\eta_{CF,T} > 0$  and  $\eta_{NC,T} > 0$  depending only on the mesh geometry, the data, and the discrete solution such that

$$\sum_{T \in \mathcal{T}_h} \left\| \kappa^{1/2} \nabla e_h \right\|_T^2 \leq \sum_{T \in \mathcal{T}_h} \underbrace{\eta_{CF,T}^2}_{\text{Conforming Error}} + \sum_{T \in \mathcal{T}_h} \underbrace{\eta_{NC,T}^2}_{\text{Non-conforming Error}} =: \sum_{T \in \mathcal{T}_h} \eta_T^2, \quad (20)$$

where  $\sum_{T \in \mathcal{T}_h} \left\| \kappa^{1/2} \nabla e_h \right\|_T^2$  is the **broken flux norm** and

$$\|e_h\|_{\text{swip}}^2 \leq \sum_{T \in \mathcal{T}_h} (\eta_{CF,T}^2 + \eta_{NC,T}^2) + \sum_{F \in \mathcal{F}_h} \frac{\gamma_F}{h_F} \|\llbracket u_h \rrbracket\|_{L^2(\Omega)}^2. \quad (21)$$

Moreover, for every  $T \in \mathcal{T}_h$  there exist  $c_1(T), c_2(T) > 0$  such that:

$$c_1(T) \eta_{CF,T} + c_2(T) \eta_{NC,T} \leq \left\| \kappa^{1/2} \nabla e_h \right\|_T \leq \eta_{CF,T} + \eta_{NC,T}.$$

<sup>a</sup>  $\|\cdot\|_{\text{swip}}$  is defined in equation (16).

The proof of [E D+96, Theorem 3.1] separates the flux into two orthogonal parts:

$$\boldsymbol{\sigma}_h(e_h) = \boldsymbol{\sigma}(\chi) + \mathbf{curl}(\psi), \quad (22)$$

where  $\chi \in V_D = \{v \in H^1(\Omega) \mid v = 0 \text{ on } \Gamma_D\}$  satisfies

$$\langle \kappa \nabla \chi | \nabla v \rangle_{\mathbf{L}^2(\Omega)} = \langle \kappa \nabla_h e_h | \nabla v \rangle_{\mathbf{L}^2(\Omega)}, \quad \forall v \in V_D$$

and  $\psi \in V_N := \{v \in H^1(\Omega) \mid \partial_n v = 0 \text{ on } \Gamma_N\}$  satisfies

$$\langle \kappa^{-1} \mathbf{curl}(\psi) | \mathbf{curl}(w) \rangle_{\mathbf{L}^2(\Omega)} = \langle \kappa^{-1} \boldsymbol{\sigma}(e_h) | \mathbf{curl}(w) \rangle_{\mathbf{L}^2(\Omega)} = \langle \nabla_h e_h | \mathbf{curl}(w) \rangle_{\mathbf{L}^2(\Omega)}, \quad \forall w \in V_N.$$

This decomposition is orthogonal in the sense that:

$$\sum_{T \in \mathcal{T}_h} \left\| \kappa^{1/2} \nabla e_h \right\|_{\mathbf{L}^2(T)}^2 = \langle \kappa^{-1} \boldsymbol{\sigma}_h(e_h) | \boldsymbol{\sigma}_h(e_h) \rangle_{\mathbf{L}^2(\Omega)} = \langle \kappa^{-1} \boldsymbol{\sigma}(\chi) | \boldsymbol{\sigma}(\chi) \rangle_{\mathbf{L}^2(\Omega)} + \langle \kappa^{-1} \mathbf{curl}(\psi) | \mathbf{curl}(\psi) \rangle_{\mathbf{L}^2(\Omega)}.$$

The article [AIN07] then shows that we can construct

$$\eta_{NC,T} = \left\| \kappa^{1/2} \nabla (u_h^* - u_h) \right\|_{\mathbf{L}^2(T)}, \quad (23)$$

$$\eta_{CF,T} = \kappa_T^{-1/2} \sqrt{\|\boldsymbol{\rho}_T\|_{\mathbf{L}^2(T)}^2 - \mathcal{C}_T^{*2} |T| \|\mathbf{curl}(\boldsymbol{\rho}_T)\|_{\mathbf{L}^2(T)}^2} \quad (24)$$

$$+ C_P(T) \kappa_T^{-1/2} \|f - \bar{f}\|_{\mathbf{L}^2(T)} + \kappa_T^{-1/2} \sum_{F \in \mathcal{F}_T \cap \mathcal{F}_h^b} C_t(T, F) \|g_N - \bar{g}_{N,F}\|_{\mathbf{L}^2(F)}, \quad (25)$$

where  $u_h^*$  is defined in equation (26),  $\rho_T$  is defined in (36) and

$$\left\{ \begin{array}{l} (\mathcal{C}_T^*)^2 := \frac{\kappa_T}{20 \text{Tr}(\mathbf{S}_T)}, \text{ where } \mathbf{S}_T \text{ is the local stiffness matrix of } T. \\ C_P(T) := \frac{1}{\pi} \max_{\mathbf{x}, \mathbf{y} \in T} \|\mathbf{x} - \mathbf{y}\|_{\mathbb{R}^2} = \frac{1}{\pi} h_T \text{ where } h_T \text{ is the diameter of } T. \\ \bar{f}_T := \frac{1}{|T|} \int_T f. \\ \bar{g}_{N,F} := \frac{1}{|F|} \int_F g_N. \\ L_F := \max_{\mathbf{x} \in F} \|\mathbf{x} - \mathbf{x}_F\|_{\mathbb{R}^2} = \max\{|F_1|, |F_2|\}, \text{ where } F_i \text{ are the other two faces.} \\ l_F := \min_{\mathbf{x} \in F} \|\mathbf{x} - \mathbf{x}_F\|_{\mathbb{R}^2} = h_{T,F}, \text{ where } h_{T,F} \text{ is the height from the vertex opposite to } F. \\ C_t(T, F) := \frac{2}{l_F} C_P(C_P + L_F). \end{array} \right.$$

### Proposition 3.2: Estimation of the Two Components

$$\langle \kappa^{-1} \mathbf{curl}(\psi) | \mathbf{curl}(\psi) \rangle_{\mathbf{L}^2(\Omega)} = \left\| \kappa^{-1/2} \mathbf{curl}(\psi) \right\|_{\mathbf{L}^2(\Omega)}^2 \leq \sum_{T \in \mathcal{T}_h} \eta_{NC,T}^2$$

and

$$\langle \kappa^{-1} \boldsymbol{\sigma}(\chi) | \boldsymbol{\sigma}(\chi) \rangle_{\mathbf{L}^2(\Omega)} = \left\| \kappa^{1/2} \nabla \chi \right\|_{\mathbf{L}^2(\Omega)}^2 \leq \sum_{T \in \mathcal{T}_h} \eta_{CF,T}^2.$$

## 2) Non-conforming Estimator

This section corresponds to [AIN07, Section 6]. According to [AIN05, Lemma 3.1],

$$\langle \kappa^{-1} \mathbf{curl}(\psi) | \mathbf{curl}(\psi) \rangle_{\mathbf{L}^2(\Omega)} = \min_{u^* \in H_{g_D}^1} \langle \kappa \nabla(u^* - u_h) | \nabla(u^* - u_h) \rangle_{\mathbf{L}^2(\Omega)}.$$

The goal is then to construct a  $u_h^* \in H_{g_D}^1$  that can be computed from  $u_h$  and is sufficiently accurate. Let  $T \in \mathcal{T}_h$  be an element,  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \in \mathbb{R}^2$  its vertices, and  $\mathbf{x}_G$  the barycenter of  $T$ . We define  $u_h^* \in \mathbb{P}_2^1(\mathcal{T}_h)$  by the so-called Oswald interpolation [OSW94] of  $u_h$  at the mesh vertices in the Lagrange basis  $\mathbb{P}_2^1(\mathcal{T}_h)$  with Dirichlet conditions such that

$$u_h^*(\mathbf{x}_i) := \begin{cases} \frac{1}{\text{Card}(\Omega_i)} \sum_{T \in \Omega_i} u_h|_T(\mathbf{x}_i) & \text{if } \mathbf{x}_i \notin \Gamma_D, \\ g_D(\mathbf{x}_i) & \text{if } \mathbf{x}_i \in \Gamma_D. \end{cases} \quad (26)$$

where the set of elements sharing  $\mathbf{x}_i$  denoted  $\Omega_i$  is defined in Fig. 5.

We then have a representative  $u_h^* \in H_{g_D}^1(\Omega)$  in the space of broken polynomials, hence

$$\begin{aligned} \langle \kappa \nabla(u_h^* - u_h) | \nabla(u_h^* - u_h) \rangle_{\mathbf{L}^2(\Omega)} &= \sum_{T \in \mathcal{T}_h} \langle \kappa \nabla(u_h^* - u_h) | \nabla(u_h^* - u_h) \rangle_{\mathbf{L}^2(T)} \\ &= \sum_{T \in \mathcal{T}_h} \left\| \kappa^{1/2} \nabla(u_h^* - u_h) \right\|_{\mathbf{L}^2(T)}^2 \\ &\geq \langle \kappa^{-1} \mathbf{curl}(\psi) | \mathbf{curl}(\psi) \rangle_{\mathbf{L}^2(\Omega)}. \end{aligned}$$

We can then define

$$\eta_{NC,T}^2 := \left\| \kappa^{1/2} \nabla(u_h^* - u_h) \right\|_{\mathbf{L}^2(\Omega)}^2.$$

## 3) Conforming Estimator

This section corresponds to [AIN07, Section 5]. The section on equilibrated numerical fluxes corresponds to the main result of the article, as it presents only well-determined constants [ESV10], whereas they are underdetermined in [BHL03; KP03; CGJ09].

### a) Equilibrated Numerical Fluxes

Let  $T \in \mathcal{T}_h$ , we seek to define a piecewise linear function  $\Phi_T : \partial T \rightarrow \mathbb{R}$  such that

$$\int_T f + \int_{\partial T} \Phi_T = 0. \quad (27)$$

i.e., we want to reconstruct an approximation of the flux associated with the numerical scheme from the discrete solution data.

Following [AIN07, Lemma 5], we note that  $\int_T f = \int_\Omega f \mathbf{1}_T = a^{\text{swip}}(u_h, \mathbf{1}_T)$ .

We then seek  $\Phi_T$  such that

$$\int_T f + \int_{\partial T} \Phi_T = a^{\text{swip}}(u, \chi_T) + \sum_{F \in \mathcal{F}_h} \int_F \Phi_T \mathbf{1}_T = a^{\text{swip}}(u, \mathbf{1}_T) - \ell_h(\mathbf{1}_T) = 0. \quad (28)$$

For  $T \in \mathcal{T}_h$  and  $F \in \mathcal{F}_T$ , we define  $\mu_{T,F} := \mathbf{n}_{T,F} \cdot \mathbf{n}_F \in \{-1, 1\}$  and the piecewise affine function **on the faces**  $\Phi_T : \partial T \rightarrow \mathbb{R} \in \mathbb{P}_2^1(\mathcal{F}_T)$  by

$$\Phi_T := \begin{cases} \mu_{T,F} \left( \left\{ \left\{ \boldsymbol{\sigma}_h(u_h) \cdot \mathbf{n}_F \right\} - \frac{\gamma_F}{|F|} \llbracket u_h \rrbracket \right\} \right) & \text{if } F \in \mathcal{F}_h^i, \\ \boldsymbol{\sigma}(u_h) \cdot \mathbf{n}_{T,F} - \frac{\gamma_F}{|F|} (u_h - g_D) & \text{if } F \in \mathcal{F}_h^b. \end{cases} \quad (29)$$

Using the fact that  $\llbracket \mathbf{1}_T \rrbracket_F = \mu_{T,F} \mathbf{1}_T$  and that  $\nabla_h(\mathbf{1}_T) \cdot \mathbf{n}_F = \{\{\sigma_h(\mathbf{1}_T)\}\} \cdot \mathbf{n}_F = 0$ , we obtain (28). We also note, by observing that  $g_{T_1} + g_{T_2} = 0$ , that for all  $v \in V_N$ :

$$\sum_{\mathcal{T}_h} \int_{\partial T} \Phi_T v = \int_{\Gamma_N} g_N v - \sum_{F \in \mathcal{F}_h \cap \Gamma_N} \int_F (g_N - \bar{g}_{N,F}) v, \quad (30)$$

where  $\bar{g}_{N,F} = \frac{1}{|F|} \int_F g_N|_F$ .

### b) Local Representative

We then want to estimate the part on  $\nabla \chi$  of the estimator, recalling the equation:

$$\langle \kappa \nabla \chi | \nabla v \rangle_{\mathbf{L}^2(\Omega)} = \langle \kappa \nabla_h e_h | \nabla v \rangle_{\mathbf{L}^2(\Omega)}, \quad \forall v \in V_D.$$

We choose  $v \in V_D$ :

$$\begin{aligned} \langle \kappa \nabla \chi | \nabla v \rangle_{\mathbf{L}^2(\Omega)} &= \langle \kappa \nabla_h e_h | \nabla v \rangle_{\mathbf{L}^2(\Omega)} \\ \iff \langle \kappa \nabla \chi | \nabla v \rangle_{\mathbf{L}^2(\Omega)} + \langle \kappa \nabla u_h | \nabla v \rangle_{\mathbf{L}^2(\Omega)} &= \langle \kappa \nabla u | \nabla v \rangle_{\mathbf{L}^2(\Omega)} \end{aligned}$$

We then have, thanks to the continuous variational formulation of the problem (19):

$$\langle \kappa \nabla \chi | \nabla v \rangle_{\mathbf{L}^2(\Omega)} = \langle f | v \rangle_{\mathbf{L}^2(\Omega)} + \int_{\Gamma_N} g_N v - \langle \kappa \nabla u_h | \nabla v \rangle_{\mathbf{L}^2(\Omega)}. \quad (31)$$

Then, by injecting (30) into the above equation, we obtain

$$\begin{aligned} \langle \kappa \nabla \chi | \nabla v \rangle_{\mathbf{L}^2(\Omega)} &= \sum_{T \in \mathcal{T}_h} \left\{ \langle \bar{f}_T | v \rangle_{\mathbf{L}^2(T)} + \int_{\partial T} \Phi_T v - \langle \kappa \nabla u_h | \nabla v \rangle_{\mathbf{L}^2(T)} \right\} \\ &\quad + \sum_{T \in \mathcal{T}_h} \langle f - \bar{f}_T | v \rangle_{\mathbf{L}^2(T)} + \sum_{F \in \mathcal{F}_h \cap \Gamma_N} \int_F (g_N - \bar{g}_{N,F}) v, \quad \forall v \in V_D, \end{aligned}$$

where  $\bar{f}_T = \frac{1}{|T|} \int_T f$ .

The part involving  $f$ ,  $\bar{f}$ ,  $g_N$ ,  $\bar{g}$  depends only on the data and the mesh geometry. It can be easily bounded.

The objective is then to estimate the part in braces. We then seek a local representative  $\rho_T$  in  $\mathbb{P}^1(T)$  such that

$$\langle \rho_T | \nabla v \rangle_{\mathbf{L}^2(\Omega)} = \langle \bar{f}_T | v \rangle_{\mathbf{L}^2(T)} + \int_{\partial T} \Phi_T v - \langle \kappa \nabla u_h | \nabla v \rangle_{\mathbf{L}^2(T)}, \quad \forall v \in V_N. \quad (32)$$

But for  $v \in V_D$ :

$$\begin{aligned} \langle \rho_T | \nabla v \rangle_{\mathbf{L}^2(\Omega)} &= \int_T \rho_T \cdot \nabla v \\ &= \int_{\partial T} \rho_T \cdot \mathbf{n}_T v - \int_{\partial T} \operatorname{div}(\rho_T) \quad (\text{by IPP}) \end{aligned}$$

and on the other hand, by performing an IPP on the last term on the right of (32)

$$- \langle \sigma(u_h) | \nabla v \rangle_{\mathbf{L}^2(\Omega)} = - \int_{\partial T} \sigma(u_h) \cdot \mathbf{n}_{\partial T} v + \int_T \operatorname{div}(\sigma(u_h)) v,$$

hence, by combining the two equations

$$\int_{\partial T} \rho_T \cdot \mathbf{n}_{\partial T} v - \int_T \operatorname{div}(\rho_T) v = \int_{\partial T} (\Phi_T - \sigma(u_h) \cdot \mathbf{n}_{\partial T}) v + \int_T (\bar{f}_T + \operatorname{div}(\sigma(u_h))) v.$$

We can then identify  $\boldsymbol{\rho}_T \cdot \mathbf{n}_{\partial T}$  and  $\text{div}(\boldsymbol{\rho}_T)$ :

$$\boldsymbol{\rho}_T \cdot \mathbf{n}_{\partial T} = \Phi_T - \boldsymbol{\sigma}(u_h) \cdot \mathbf{n}_{\partial T}, \quad (33)$$

$$\text{div}(\boldsymbol{\rho}_T) = \bar{f}_T + \text{div}(\boldsymbol{\sigma}(u_h)). \quad (34)$$

In the following, we denote by  $\lambda_j$  the local barycentric function of  $T \in \mathcal{T}_h$  such that if  $\mathbf{x}_i$  is the coordinate of vertex  $a_i$  of  $T$ , then  $\lambda_j(\mathbf{x}_i) = \delta_{ij}$ .

The proof [AIN07, Lemma 6] then shows that by defining

$$\begin{cases} \boldsymbol{\rho}_1^{(T)} = |F_3| \Delta_3^{(T)}(\mathbf{x}_1) \boldsymbol{\tau}_2 - |F_2| \Delta_2^{(T)}(\mathbf{x}_1) \boldsymbol{\tau}_3, \\ \boldsymbol{\rho}_2^{(T)} = |F_1| \Delta_1^{(T)}(\mathbf{x}_2) \boldsymbol{\tau}_3 - |F_3| \Delta_3^{(T)}(\mathbf{x}_2) \boldsymbol{\tau}_1, \\ \boldsymbol{\rho}_3^{(T)} = |F_2| \Delta_2^{(T)}(\mathbf{x}_3) \boldsymbol{\tau}_1 - |F_1| \Delta_1^{(T)}(\mathbf{x}_3) \boldsymbol{\tau}_2. \end{cases} \quad (35)$$

where for a face  $F_n$  and a vertex  $S_j$ ,  $j \neq n$  such that the coordinate of vertex  $\mathbf{x}_j$  is on  $F_n$  (see Figure 8):

$$\Delta_n^{(T)} = \Phi_T - \boldsymbol{\sigma}(u_h) \cdot \mathbf{n}_T|_{F_n} \in \mathbb{P}^1(F_n),$$

the function  $\mathbb{P}^1(T)$  defined by

$$\boldsymbol{\rho}_T = \frac{1}{2|T|} \sum_{n=1}^3 \boldsymbol{\rho}_n^{(T)} \lambda_n \quad (36)$$

then satisfies (33) and (34). By retracing the calculations, we finally show that  $\boldsymbol{\rho}_T$  satisfies (32).

The function  $\Delta_n^{(T)}$  then corresponds to the barycentric coordinate of  $\boldsymbol{\rho}_T \cdot \mathbf{n}_{\partial T}$  on the face  $F_n$ , i.e., for  $n \in \{1, 2, 3\}$ ,  $i, j \in \{1, 2, 3\} \setminus \{n\}$  with always  $i \neq j$

$$\boldsymbol{\rho}_T \cdot \mathbf{n}_{\partial T}|_{F_n} = \Delta_n^{(T)}(x_i) \lambda_i + \Delta_n^{(T)}(x_j) \lambda_j, \quad (37)$$

or again, if  $\mathbf{x}_i$  is the coordinate of a vertex on the face  $F_n$

$$(\boldsymbol{\rho}_T \cdot \mathbf{n}_{T, F_n})(x_i) = \Delta_n^{(T)}(x_i).$$

We then want to calculate the norm of  $\boldsymbol{\rho}_T$  on a triangle, but

$$\int_T \lambda_i \lambda_j = \frac{1}{12} |T| (1 + \delta_{ij}).$$

We therefore have

$$\|\boldsymbol{\rho}_T\|_{\mathbf{L}^2(T)}^2 = \frac{1}{48|T|} \sum_{i=1}^3 \sum_{j=1}^3 (1 + \delta_{ij}) \boldsymbol{\rho}_i^{(T)} \cdot \boldsymbol{\rho}_j^{(T)}. \quad (38)$$

And finally, for  $v \in V_D$ , by applying Cauchy-Schwarz

$$\begin{aligned} \langle \kappa \nabla \chi | \nabla \chi \rangle_{\mathbf{L}^2(\Omega)} &= \sum_{T \in \mathcal{T}_h} \left\{ \langle \bar{f}_T | \chi \rangle_{\mathbf{L}^2(T)} + \int_{\partial T} \Phi_T \chi - \langle \kappa \nabla u_h | \nabla \chi \rangle_{\mathbf{L}^2(T)} \right\} \\ &\quad + \sum_{T \in \mathcal{T}_h} \langle f - \bar{f}_T | \chi \rangle_{\mathbf{L}^2(T)} + \sum_{F \in \mathcal{F}_h \cap \Gamma_N} \int_F (g_N - \bar{g}_{N,F}) \chi \\ &= \sum_{T \in \mathcal{T}_h} \langle \boldsymbol{\rho}_T | \nabla \chi \rangle_{\mathbf{L}^2(T)} + \sum_{T \in \mathcal{T}_h} \langle f - \bar{f}_T | \chi \rangle_{\mathbf{L}^2(T)} + \sum_{T \in \mathcal{T}_h} \sum_{F \in \mathcal{F}_T \cap \Gamma_N} \int_F (g_N - \bar{g}_{N,F}) \chi \\ \implies \left\| \kappa^{1/2} \nabla \chi \right\|_{\mathbf{L}^2(\Omega)}^2 &\leq \sum_{T \in \mathcal{T}_h} \kappa_T^{-1/2} \|\boldsymbol{\rho}_T\|_{\mathbf{L}^2(T)} \left\| \kappa_T^{1/2} \nabla \chi \right\|_{\mathbf{L}^2(T)} + \sum_{T \in \mathcal{T}_h} \kappa_T^{-1/2} \|f - \bar{f}_T\|_{\mathbf{L}^2(T)} \left\| \kappa_T^{-1/2} \chi \right\|_{\mathbf{L}^2(T)} \\ &\quad + \sum_{T \in \mathcal{T}_h} \sum_{F \in \mathcal{F}_h \cap \Gamma_N} \kappa_T^{-1/2} \|g_N - \bar{g}_{N,F}\|_{\mathbf{L}^2(F)} \left\| \kappa_T^{1/2} \chi \right\|_{\mathbf{L}^2(F)}. \end{aligned}$$

where we take advantage of the constant nature of  $\kappa_T = \kappa|_T$  to make it appear in the norms.

To obtain an estimate of  $\|\kappa^{1/2}\nabla\chi\|_{\mathbf{L}^2(\Omega)}$ , we will transform the terms  $\|\kappa_T^{-1/2}\chi\|_{\mathbf{L}^2(T)}$  and  $\|\kappa_T^{1/2}\chi\|_{\mathbf{L}^2(F)}$  by  $C\|\kappa^{1/2}\nabla\chi\|_{\mathbf{L}^2(\Omega)}$ .

The optimal Poincaré inequality given in [PW60, Equation (1.9)] allows us to state

$$\|\chi\|_{\mathbf{L}^2(T)} \leq C_P \|\nabla\chi\|_{\mathbf{L}^2(T)},$$

where

$$C_P = \frac{1}{\pi} \max_{\mathbf{x}, \mathbf{y} \in T} \|\mathbf{x} - \mathbf{y}\|_{\mathbb{R}^2}.$$

Then, the proof [AIN07, Lemma 10] (based on Stokes' formula) proves that by defining

$$L_F = \max_{\mathbf{x} \in F} \|\mathbf{x} - \mathbf{x}_F\|_{\mathbb{R}^2}, \quad (39)$$

$$l_F = \min_{\mathbf{x} \in F} \|\mathbf{x} - \mathbf{x}_F\|_{\mathbb{R}^2}, \quad (40)$$

we can write:

$$\|\chi\|_{\mathbf{L}^2(F)} = \frac{2}{l_F} \|\chi\|_{\mathbf{L}^2(T)} \left( \|\chi\|_{\mathbf{L}^2(T)} + L_F \|\nabla\chi\|_{\mathbf{L}^2(T)} \right).$$

By reapplying the Poincaré inequality, we can factor out the  $\|\nabla\chi\|_{\mathbf{L}^2(T)}$  by the trace constant

$$C_t(T, F) = \frac{2}{l_F} C_P (C_P + L_F),$$

and we deduce the estimate

$$\|\chi\|_{\mathbf{L}^2(F)} \leq C \|\kappa^{1/2}\nabla\chi\|_{\mathbf{L}^2(T)}.$$

Hence, finally

$$\begin{aligned} \|\kappa^{1/2}\nabla\chi\|_{\mathbf{L}^2(\Omega)} &\leq \sum_{T \in \mathcal{T}_h} \kappa_T^{-1/2} \|\rho_T\|_{\mathbf{L}^2(T)} + \sum_{T \in \mathcal{T}_h} \kappa_T^{-1/2} C_P \|f - \bar{f}_T\|_{\mathbf{L}^2(T)} \\ &\quad + \sum_{T \in \mathcal{T}_h} \kappa_T^{-1/2} \sum_{F \in \mathcal{F}_T} C_t(T, F) \|g_N - \bar{g}_{N,F}\|_{\mathbf{L}^2(F)}. \end{aligned}$$

We can then set for  $T \in \mathcal{T}_h$ :

$$\eta_{CF,T} := \kappa_T^{-1/2} \left( \|\rho_T\|_{\mathbf{L}^2(T)} + C_P \|f - \bar{f}_T\|_{\mathbf{L}^2(T)} + \sum_{F \in \mathcal{F}_T} C_t(T, F) \|g_N - \bar{g}_{N,F}\|_{\mathbf{L}^2(F)} \right). \quad (41)$$

The article then shows that it is possible to improve the accuracy of this estimate using bubble functions  $\beta$  solutions of:

$$\begin{cases} -\Delta\beta &= 1 & \text{in } T, \\ \beta &= 0 & \text{on } \partial T. \end{cases} \quad (42)$$

Following [AIN07, Eq (20) and paragraph 3.2], we approximate this function by a third-order polynomial such that by noting

$$(\mathcal{C}_T^*)^2 = \frac{\kappa_T}{20\text{Tr}(\mathbf{S}_T)}$$

with  $\mathbf{S}_T$  the local stiffness matrix of the triangle, we then obtain by [AIN07, Lemma 7]

$$\begin{aligned} \rho_T^* &= \rho_T - \mathcal{C}_T^* \text{curl}(\rho_T), \\ \|\rho^*\|_{\mathbf{L}^2(T)}^2 &= \|\rho\|_{\mathbf{L}^2(T)}^2 - (\mathcal{C}_T^*)^2 |T| \|\text{curl}(\rho_T)\|_{\mathbf{L}^2(T)}^2. \end{aligned}$$

and by replacing  $\rho_T$  with  $\rho_T^*$  we have the estimate:

$$\langle \kappa \nabla \chi | \chi \rangle_{\mathbf{L}^2(\Omega)} \leq \sum_{T \in \mathcal{T}_h} (\eta_{CF,T}^*)^2 \leq \sum_{T \in \mathcal{T}_h} \eta_{CF,T}^2. \quad (43)$$

#### 4) Local Equivalence between Error and Estimator

We consider  $T \in \mathcal{T}_h$  and  $F = \partial T \cap \Gamma_N$  a face of  $T$  on the Neumann boundary if it exists.

Let  $\tilde{T} := \{T' / \mathcal{F}_T \cap \mathcal{F}_{T'} \neq \emptyset\}$  be the set of elements neighboring  $T$ .

Let  $T^\star = \{T' / \mathcal{A}_h(T) \cap \mathcal{A}_h(T') \neq \emptyset\}$  be the set of elements  $T' \in \mathcal{T}_h$  that share a vertex with  $T$ .

We define the data oscillations as:

$$\begin{cases} \text{Osc}(f, T)^2 = |T| \|f - \bar{f}\|_{\mathbf{L}^2(T)}^2, \\ \text{Osc}(g_N, F)^2 = |F| \|g_N - \bar{g}_N\|_{\mathbf{L}^2(F)}^2. \end{cases}$$

[AIN07, Section 4, Lemma 2, Lemma 3, Lemma 8] proves the following inequality:

##### Proposition 3.3: Norm Equivalence

If  $T \in \mathcal{T}_h$ , there exists  $c = c(T) > 0$  independent of  $h_T$  such that

$$c\eta_{CF,T} \leq \left\| \kappa^{1/2} \nabla \xi \right\|_{\mathbf{L}^2(\tilde{T})} + \left\| \kappa^{-1/2} \mathbf{curl} \, \psi \right\|_{\mathbf{L}^2(\tilde{T})} + \text{Osc}(g_N ; \{F \in \mathcal{F}_h^N \cap T\}) + \text{Osc}(f, \tilde{T}).$$

Similarly, there exists  $C = C(T) > 0$  independent of  $h_T$  such that

$$C\eta_{NC,T} \leq \left\| \kappa^{1/2} \nabla \xi \right\|_{\mathbf{L}^2(T^\star)} + \left\| \kappa^{-1/2} \mathbf{curl} \, \psi \right\|_{\mathbf{L}^2(T^\star)} + \text{Osc}(g_N ; \{F \in \mathcal{F}_h^N \cap T\}) + \text{Osc}(f, T^\star).$$

We then deduce the equivalence between  $\eta_T$  and the norm  $\|\kappa \nabla u_h\|_{\mathbf{L}^2(T)}$  up to the data oscillation.

#### 5) Modification for SWIP

To adapt this article in the context of the SWIP method, it is appropriate to modify the estimator so that

$$\{\!\{v_h\}\!\}_\omega = \omega_1 v_1 + \omega_2 v_2.$$

We theoretically and numerically verify that the numerical flux is indeed conservative.



## II - Numerical Results

### 1) Homogeneous diffusion problem

#### a) Homogeneous Dirichlet in the unit square

We consider the square  $\Omega = ]0; 1[^2$ .

Let  $\lambda \in \mathbb{N}^*$ . We consider the following test case:

$$\text{Find } \boldsymbol{\sigma}(u) \in \mathbf{H}(\text{div}; \Omega), u \in H^1(\Omega) \text{ such that } \begin{cases} -\text{div } \boldsymbol{\sigma} = 2(\lambda\pi)^2 \sin(\lambda\pi x) \sin(\lambda\pi y) & \text{in } \Omega, \\ \boldsymbol{\sigma}(u) = \nabla u & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \quad (44)$$

i.e.

$$\begin{cases} f(x, y) = 2(\lambda\pi)^2 \sin(\lambda\pi x) \sin(\lambda\pi y) \in \mathcal{C}^\infty(\Omega), \\ \Gamma_D = \partial\Omega, \\ g_D = 0 \in \mathcal{C}^\infty(\partial\Omega), \\ \kappa \equiv 1. \end{cases}$$

This test case corresponds to the spectral problem of the Laplacian on the unit square. The solution of this problem is known:

$$u_\lambda : (x, y) \in \mathbb{R}^2 \mapsto \sin(\lambda\pi x) \sin(\lambda\pi y) \in \mathcal{C}^\infty(\Omega).$$

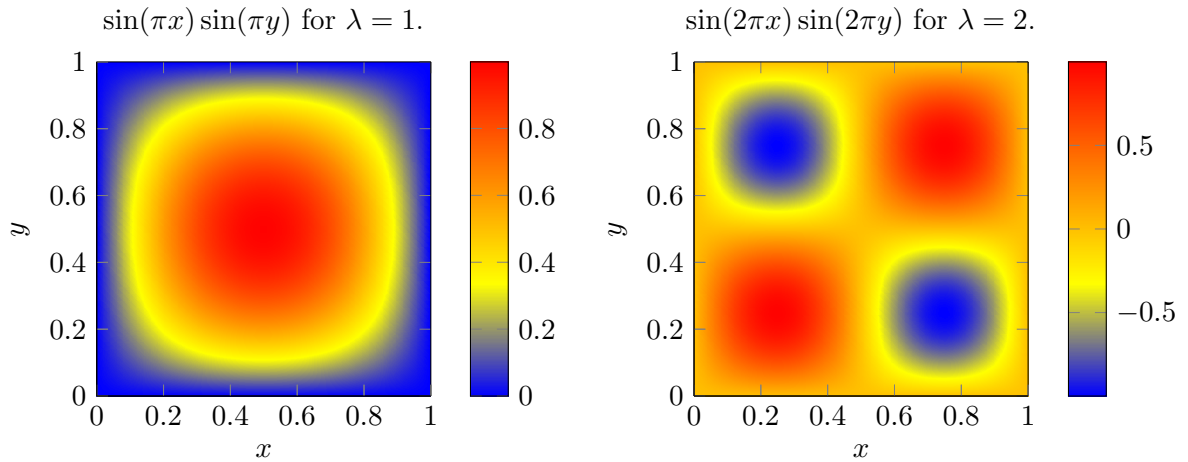


Figure 15: Solution of the spectral problem on  $[0; 1]^2$  for  $\lambda = 1, 2$ .

The broken flux restricted to an element  $T \in \mathcal{T}_h$  is thus

$$\boldsymbol{\sigma}(u_\lambda)|_T = \nabla u_\lambda,$$

and we can easily compute

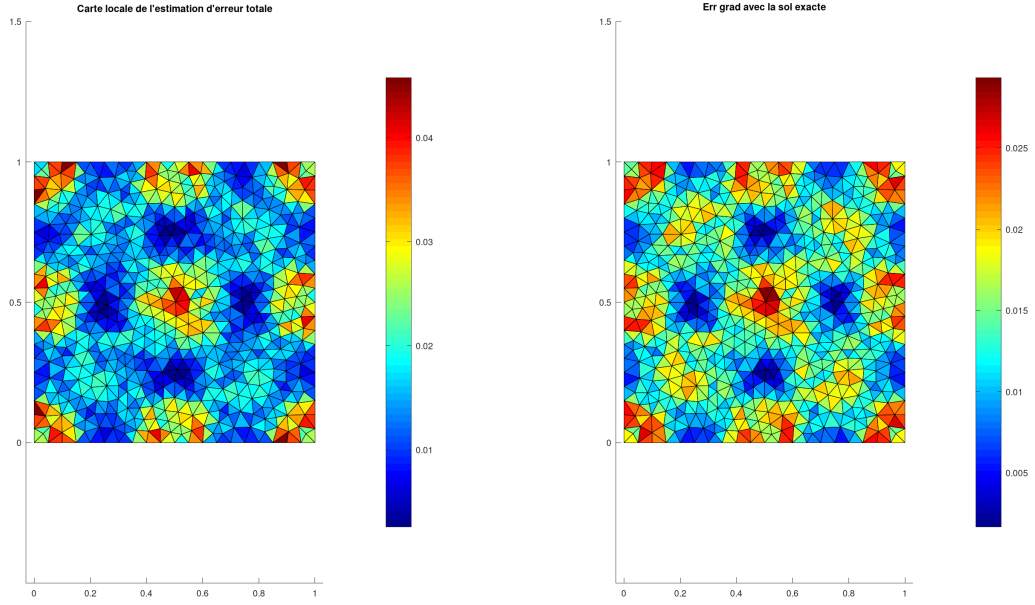
$$\epsilon_T := \|\boldsymbol{\sigma}(u_\lambda) - \boldsymbol{\sigma}(u_h)\|_{\mathbf{L}^2(T)},$$

$$\epsilon_h := \sqrt{\sum_{T \in \mathcal{T}_h} \epsilon_T^2},$$

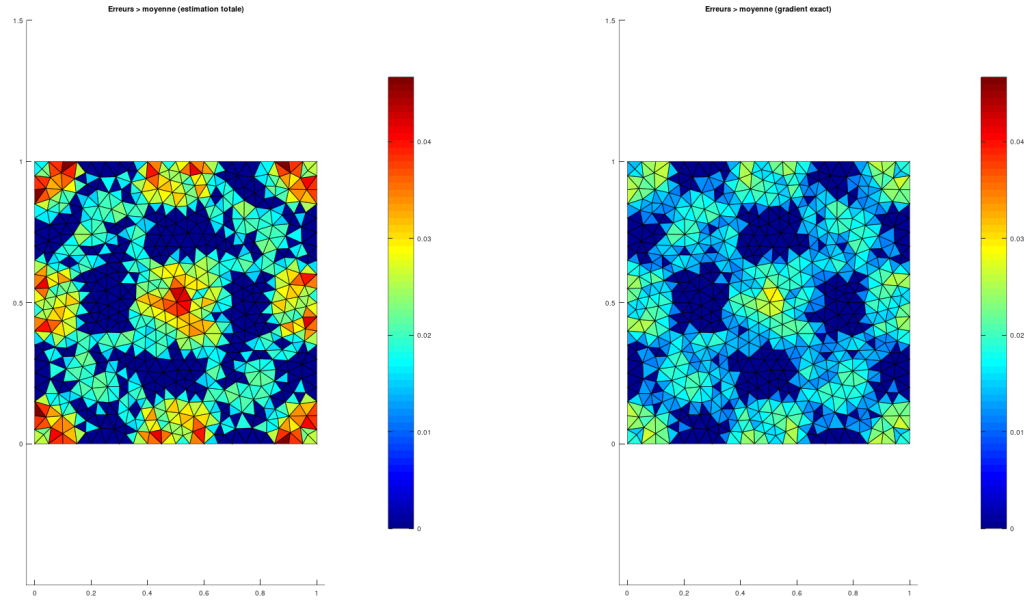
for each simulation. We will next compare the results of  $\epsilon_h$  with those of  $\eta_T$ , where  $\eta_T$  is the a posteriori error estimator defined in Theorem 3.1.

The local error map is obtained in Figure 2 for  $\lambda = 1$ .

We focus on the case  $\lambda = 2$  because the solution variations are more significant. We will denote the solution by  $u$  without specifying the dependence on  $\lambda = 2$ .



(a) A posteriori error estimator (left) and broken flux error (right) on their relative value scales.



(b) A posteriori error estimator (left) and broken flux error (right) on the same value scale, showing only values above the mean  $\langle \eta \rangle_{\mathcal{T}_h}$  (left) and  $\langle \epsilon \rangle_{\mathcal{T}_h}$  (right).<sup>9</sup>

Figure 16: Local map of a posteriori error estimator  $\eta_T$  (left) and broken flux error  $\epsilon_T$  (right) for  $\lambda = 2$ ,  $h = 0.05$  and  $v_h \in \mathbb{P}^1$ .

Figure 16 shows that the error estimator is locally faithful to the true error in the sense that their maxima and minima coincide; it slightly overestimates but the relative values are of the same order as the true error.

---

<sup>9</sup>  $\langle X \rangle_{\mathcal{T}_h} := \frac{1}{\text{Card}(\mathcal{T}_h)} \sum_{T \in \mathcal{T}_h} X_T$

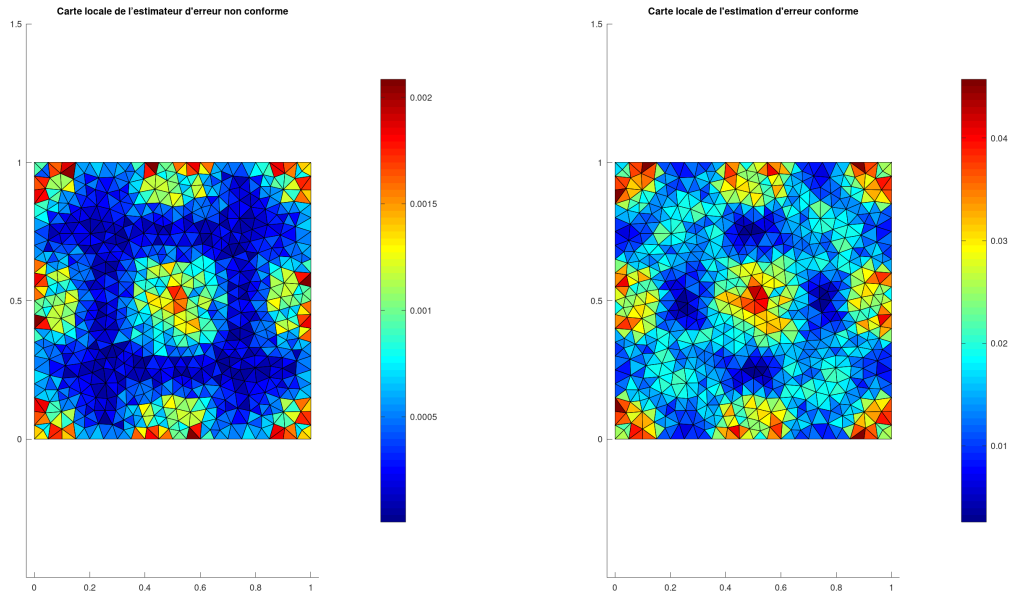
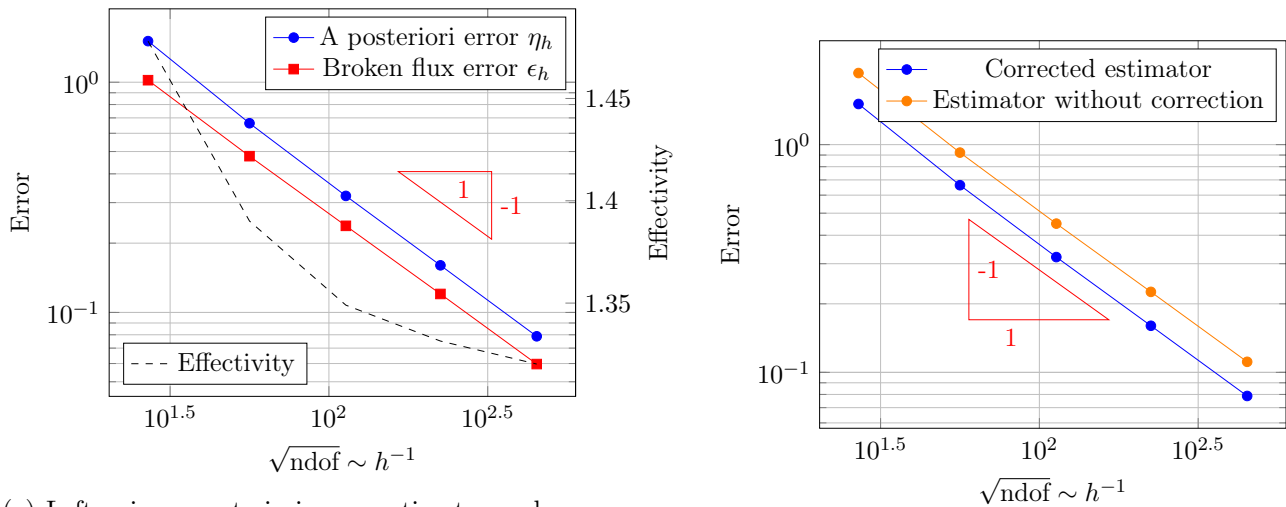


Figure 17: Left: nonconforming error estimator  $\eta_{NC,T}$ , right: conforming error estimator  $\eta_{CF,T}$  for  $\lambda = 2$ ,  $h = 0.05$  and  $v_h \in \mathbb{P}^1$ .

Figure 17 shows that the nonconforming error is small in magnitude and localized at the gradient variations. This is the expected behavior since the jump penalization is optimized in our numerical implementation.



(a) Left axis: a posteriori error estimator and broken flux error in log-log scale. Right axis: effectivity in log-linear scale. Abscissa: number of degrees of freedom.  $\lambda = 2$ ,  $v_h \in \mathbb{P}^1$ .

(b) Log-log comparison between the corrected and uncorrected error estimators.  $\lambda = 2$ ,  $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom.

Figure 18: Comparison of the global estimated error and true broken flux error (left), comparison of the error estimation with and without bubble function correction (right).

From Figure 19, it is observed that the nonconforming error estimate is several orders of magnitude smaller than the conforming error, which agrees with Figure 17.

The data oscillation decreases as  $h^{-2}$  as shown in [AIN07].

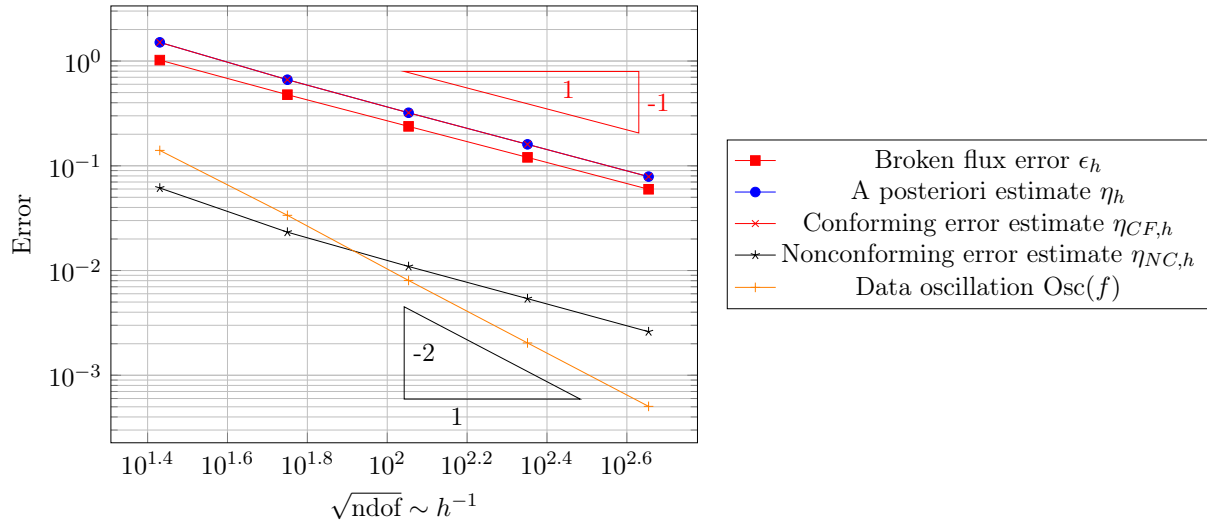


Figure 19: Comparison of the different components of the error estimator and the error in log-log scale for  $\lambda = 2$  and  $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom.

In the following, the local error estimator map is used to refine the mesh in regions where the local estimator is largest. The goal is to refine only critical areas and keep a coarse mesh in regions where the error is already small.

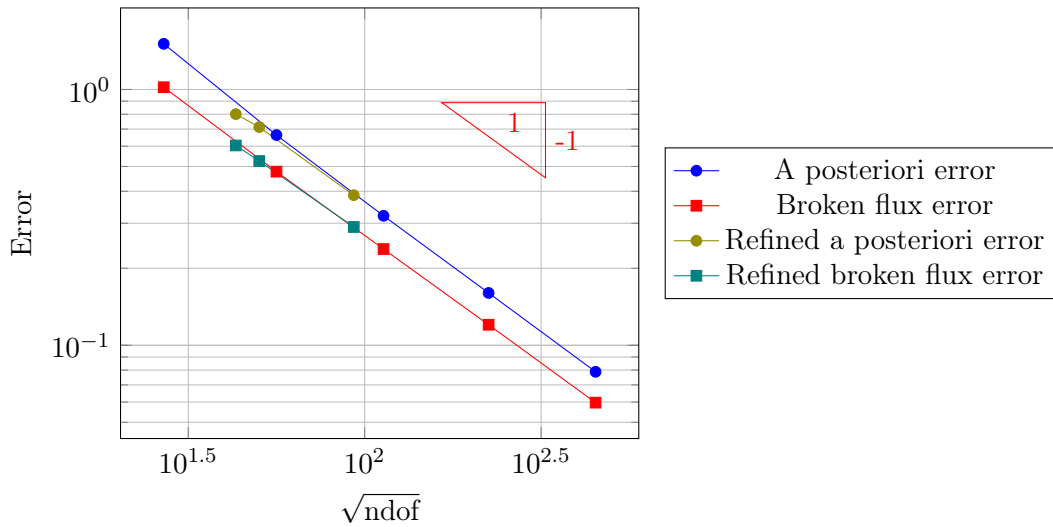


Figure 20: A posteriori error and broken flux error in the case of global refinement (all cells have equal size) and in the case of local refinement (mesh refined in regions where the error is large). Log-log scale. Abscissa: number of degrees of freedom.

It can be seen that the error for the locally refined mesh is smaller for the same number of degrees of freedom, but no superconvergence or significantly smaller error is obtained.

This is not surprising since the solution is smooth and the domain convex, thus there is no regularity issue. Nevertheless, the error remains smaller when refining locally in regions where the error estimate is largest.

One can also study the behavior of the error when varying  $\gamma$  to highlight the effect of jump penalization on the conforming and nonconforming errors.

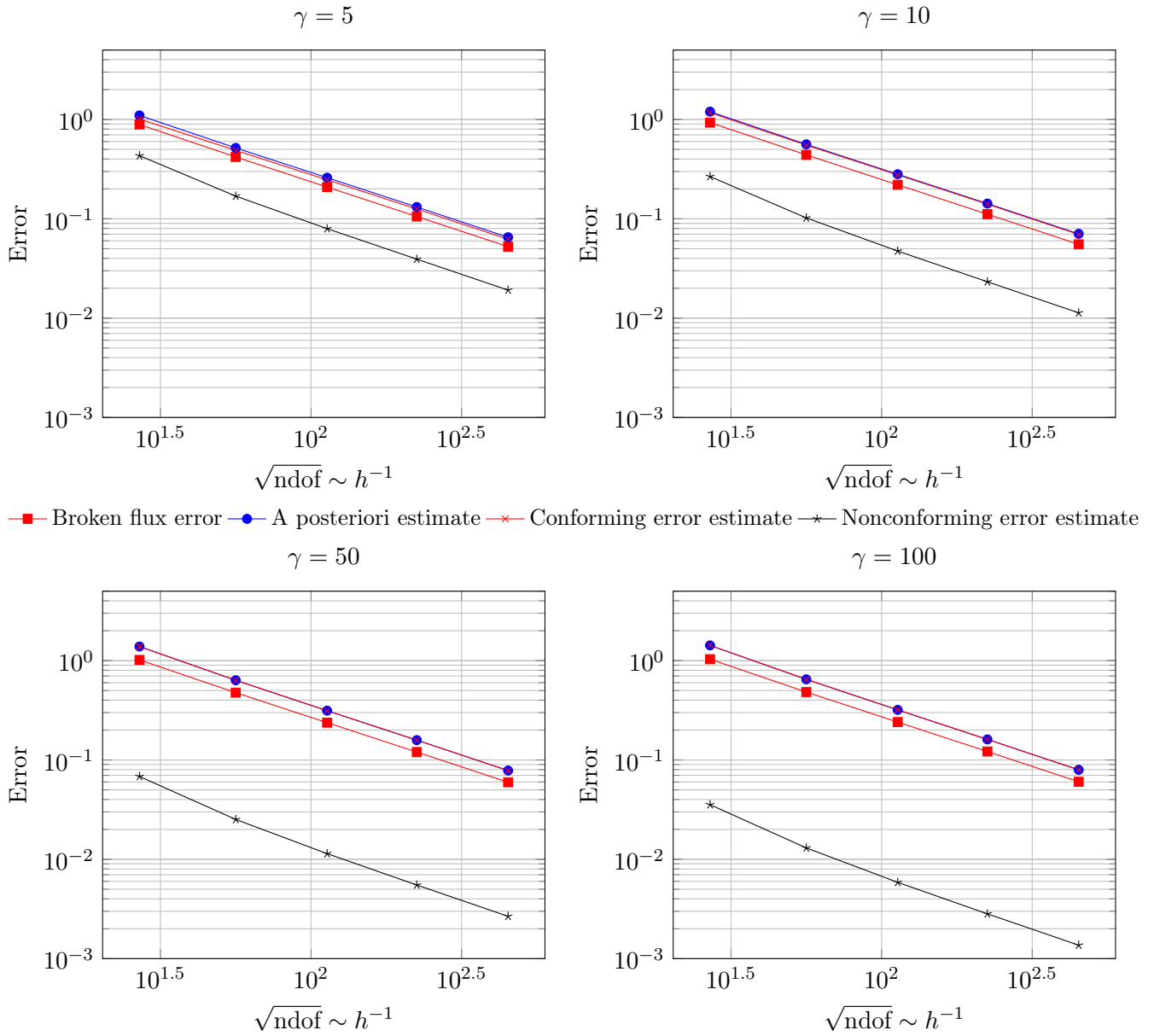


Figure 21: Comparison of the different components of the error estimator and the actual error when  $\gamma_F$  varies in  $[5, 10, 50, 100]$  for  $\lambda = 2$  and  $v_h \in \mathbb{P}^1$ , log-log scale. Abscissa: number of degrees of freedom.

The discrete solution obtained by the discontinuous Galerkin method becomes increasingly conforming, in the sense that the amplitude of its nonconforming part decreases as the jump penalization increases. This is the expected behavior since a higher penalization enforces continuity more strongly.

In conclusion, the estimator is correctly implemented for homogeneous Dirichlet conditions and its estimates reflect the local and global behavior of the error. The next step is to verify its behavior for non-homogeneous boundary conditions and less regular solutions in a nonconvex domain.

### b) Non-homogeneous Neumann condition in the unit square

We consider the square  $\Omega = ]0; 1[^2$ .

Let  $\lambda \in \mathbb{N}^*$ . We consider the following test case:

$$\text{Find } \boldsymbol{\sigma}(u) \in \mathbf{H}(\text{div}; \Omega), u \in H^1(\Omega) \text{ such that } \begin{cases} -\text{div } \boldsymbol{\sigma} = 2(\lambda\pi)^2 \sin(\lambda\pi x) \sin(\lambda\pi y) & \text{in } \Omega, \\ \boldsymbol{\sigma}(u) = \nabla u & \text{in } \Omega, \\ \boldsymbol{\sigma}(u) \cdot \mathbf{n} = \lambda\pi \begin{bmatrix} \cos(\lambda\pi x) \sin(\lambda\pi y) \\ \sin(\lambda\pi x) \cos(\lambda\pi y) \end{bmatrix} \cdot \mathbf{n} & \text{on } \partial\Omega. \end{cases} \quad (45)$$

i.e.

$$\begin{cases} f(x, y) = 2(\lambda\pi)^2 \sin(\lambda\pi x) \sin(\lambda\pi y) \in \mathcal{C}^\infty(\Omega), \\ \Gamma_N = \partial\Omega, \\ g_N = \lambda\pi \begin{bmatrix} \cos(\lambda\pi x) \sin(\lambda\pi y) \\ \sin(\lambda\pi x) \cos(\lambda\pi y) \end{bmatrix} \cdot \mathbf{n}_{\partial\Omega}^{10} \in L^2(\partial\Omega), \\ \kappa \equiv 1. \end{cases}$$

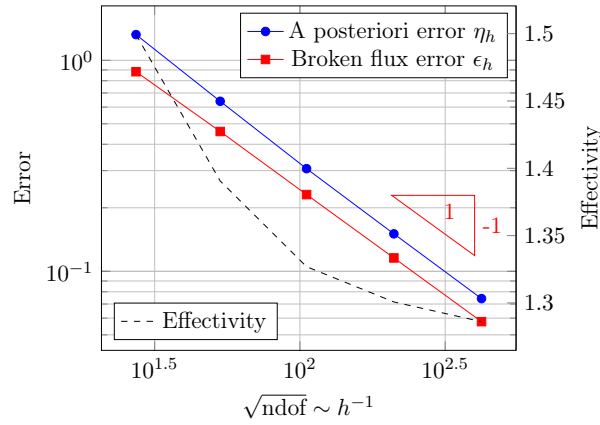


Figure 22: A posteriori error estimator and broken flux error in log-log scale. Abscissa: number of degrees of freedom.  $\lambda = 2$  and  $v_h \in \mathbb{P}^1$ .

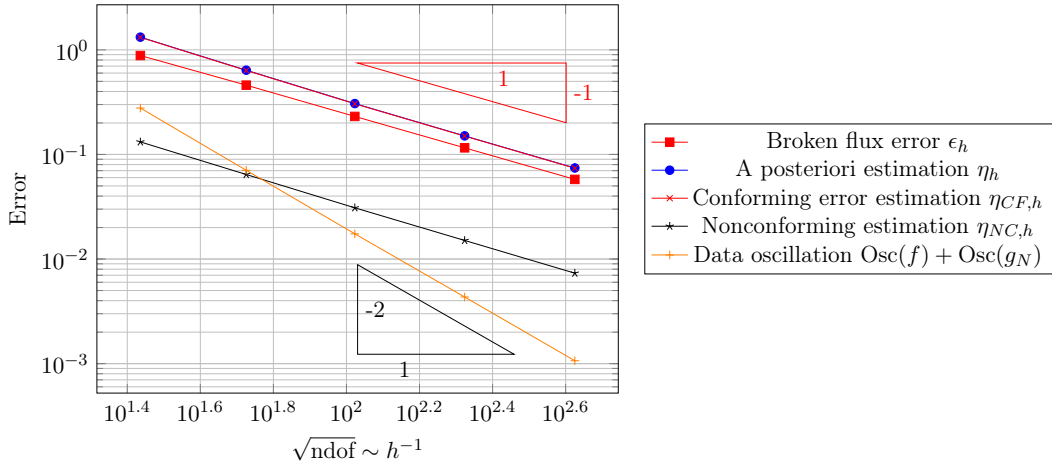
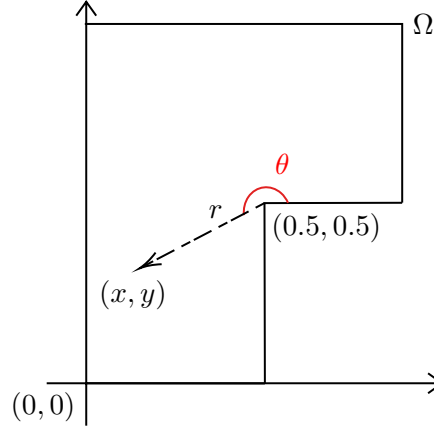


Figure 23: Comparison of the different components of the error estimator and the error in log-log scale for  $\lambda = 2$  and  $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom.

<sup>10</sup>Outward unit normal vector on the boundary of  $\Omega$ .

## c) Dirichlet in a nonconvex L-shaped bend

Figure 24: Nonconvex L-shaped domain  $\Omega$  and its associated polar coordinates.

We consider the polar coordinates  $(r, \theta) \in \mathbb{R}_+^* \times [0; \frac{3}{2}\pi]$  centered at  $(0.5, 0.5)$  which describe the domain and its boundary in a bijective manner:

$$\overline{\Omega} = \Omega \sqcup \partial\Omega = \left\{ (x, y) \in \mathbb{R}^2 \mid x \in [0; \frac{1}{2}], y \in [0; 1] \right\} \cup \left\{ (x, y) \in \mathbb{R}^2 \mid x \in [\frac{1}{2}; 1], y \in [\frac{1}{2}; 1] \right\}.$$

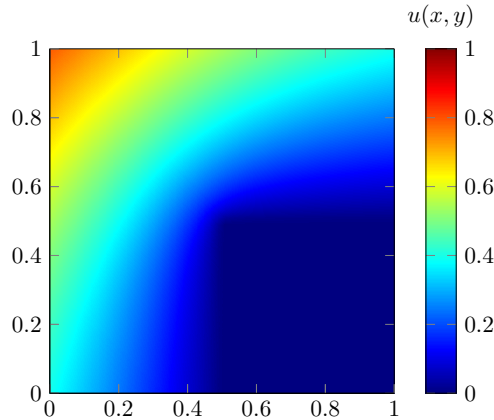
We want to solve the following harmonic problem:

$$\text{Find } u \in H^1(\Omega), \quad \begin{cases} -\Delta_x u = 0 & \text{in } \Omega, \\ u(r, \theta) = r^\alpha \sin(\frac{2}{3}\theta) & \text{on } \partial\Omega. \end{cases} \quad (46)$$

where  $\Delta_x$  is the Laplacian in Cartesian coordinates.

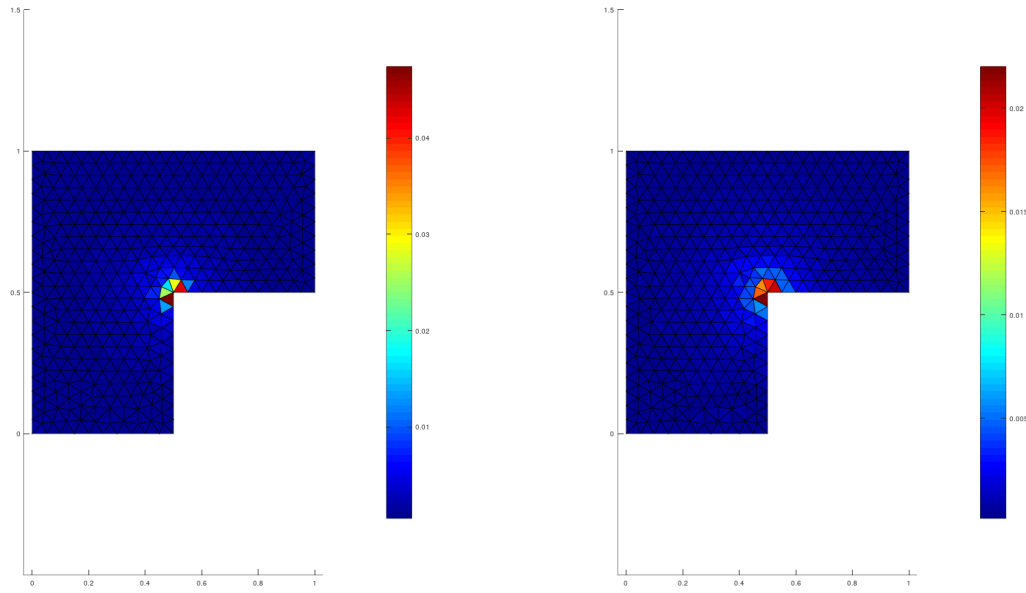
The solution is the function  $u(x, y) = f \circ \phi(x, y)$ , where  $f(r, \theta) = r^{2/3} \sin(\frac{2}{3}\theta) \in H^{1+2/3}(\Omega)$ .

$\phi(x, y) = (r(x, y), \theta(x, y))$  is the diffeomorphism transforming Cartesian coordinates into polar coordinates with  $\theta \in [0; 2\pi[$ .

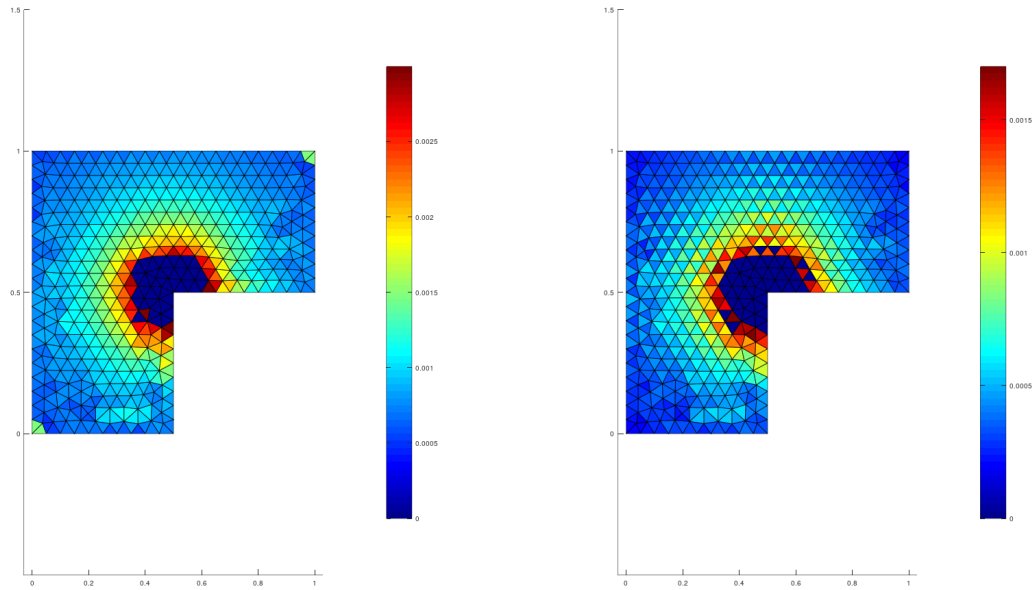
Figure 25: Solution  $f(r, \theta) = r^{2/3} \sin(\frac{2}{3}\theta)$ .**Remark :**

Following Remark p.26, the Dirichlet boundary data should be continuous and piecewise affine on the boundary ( $g_D^h \in \mathbb{P}_c^1(\mathcal{F}_h^b)$ ). Therefore, the Dirichlet data  $g_D \in H^{1/2}(\partial\Omega)$  must be projected onto  $\mathbb{P}^1(\mathcal{F}_h^b)$  to compute the estimator, taking into account the error between the solution computed with  $g_D$  and that computed with  $g_D^h$ .

According to Appendix C, a loss of regularity is expected near the reentrant corner due to the blow-up of the gradient norm at the point  $(0,0)$ . This results in a concentration of the error at this point, as well as a reduction in the convergence rate of the error.



(a) A posteriori error estimation (left) and broken flux error (right) on their respective relative value scales.



(b) A posteriori error estimation (left) and broken flux error (right) on their respective relative value scales, showing only values smaller than  $2 < \eta >_{\tau_h}$  (left) and  $2 < \epsilon >_{\tau_h}$  (right).<sup>11</sup>

Figure 26: Local map of the a posteriori error estimation  $\eta_T$  (left) and broken flux error  $\epsilon_T$  (right) for  $h = 0.05$  and  $v_h \in \mathbb{P}^1$ .

The error is clearly concentrated around the reentrant corner, and the error estimation outside this point remains consistent when considering the regions where the error is dominated by the singularity at  $(0,0)$ .

In the error estimation, some artifacts can be observed at the domain corners; these disappear when the problem is solved using boundary data  $g_D^h \in \mathbb{P}_c^1(\mathcal{F}_h^b)$  instead of the exact data  $g_D \in L^2(\Gamma_D)$ .

<sup>11</sup> $\langle X \rangle_{\tau_h} := \frac{1}{\text{Card}(\mathcal{T}_h)} \sum_{T \in \mathcal{T}_h} X_T$ .



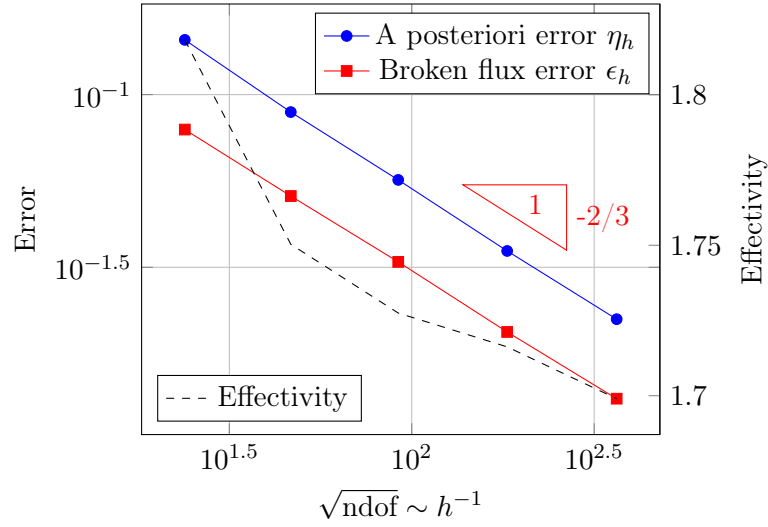


Figure 27: Comparison of the different components of the error estimator and the true error in log-log scale for the L-shaped domain and  $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom.

The error decay is at  $h^{-2/3}$ , which is predicted by the a priori error estimate in Appendix C.

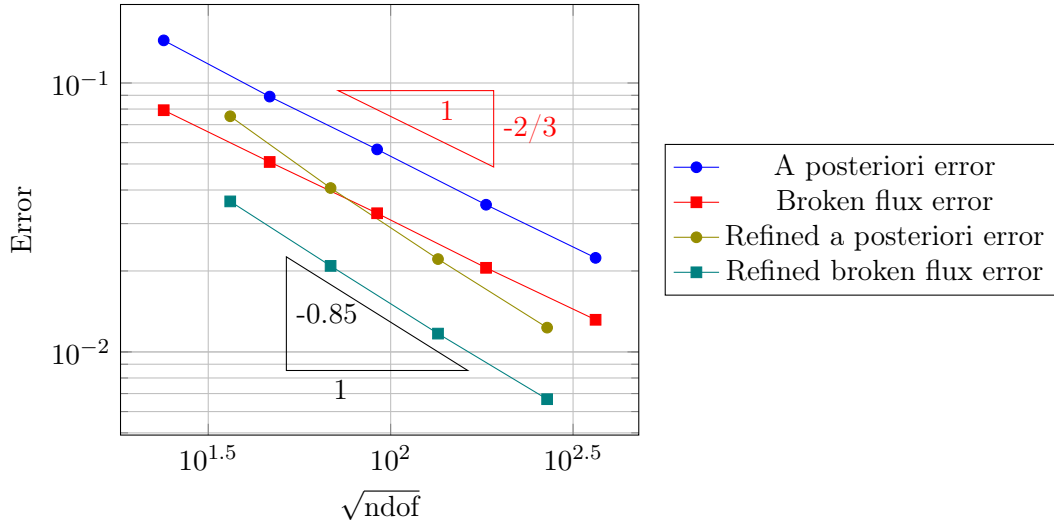


Figure 28: A posteriori error and broken flux error in the case of global refinement (all cells are of equivalent size) and in the case of local refinement (mesh refined in zones where the error is large). Log-log scale. Abscissa: number of degrees of freedom.

During refinement, we observe super-convergence of order  $r \simeq 0.85 > 2/3$ .

Since the domain is convex and the solution less regular, we can then numerically obtain faster convergence without paying more in terms of degrees of freedom, provided that we refine locally (here in the reentrant corner).

#### d) Neumann in an L-shaped bend

We still consider  $\Omega$  as in Fig.24 and the problem whose solution is given in Fig. 25:

$$\text{Find } u \in H^1(\Omega), \quad \begin{cases} -\Delta_x u = 0 & \text{in } \Omega, \\ \sigma(u) \cdot \mathbf{n} = \frac{2}{3} r^{-\alpha-1} \begin{bmatrix} \sin(\frac{2}{3}\theta) \\ \cos(\frac{2}{3}\theta) \end{bmatrix} \cdot \mathbf{n} & \text{on } \partial\Omega. \end{cases} \quad (47)$$

The continuous problem is well-posed but numerically ill-posed at  $r = 0$  because  $\|\sigma(u)\|_{\ell^2} \xrightarrow[r \rightarrow 0]{} +\infty$ <sup>12</sup>. However, the quadrature points for numerical integration at the boundary always evaluate  $g_N$  at  $(x, y) = (0, 0)$ , at the point where the values are infinite.

We then approximate the value at  $(x, y) = (0, 0)$  by the average over a circle around this point:

$$\tilde{g}_N(0, 0) = \int_{r=0}^{\varepsilon} \int_{\theta=0}^{2\pi} g_N,$$

where  $\varepsilon \lesssim h$ . This choice being arbitrary, the data oscillation and convergence results are degraded in Fig.29.

We nevertheless observe a convergence order similar to the pure Dirichlet case, which shows the robustness of the method despite the numerical approximations.

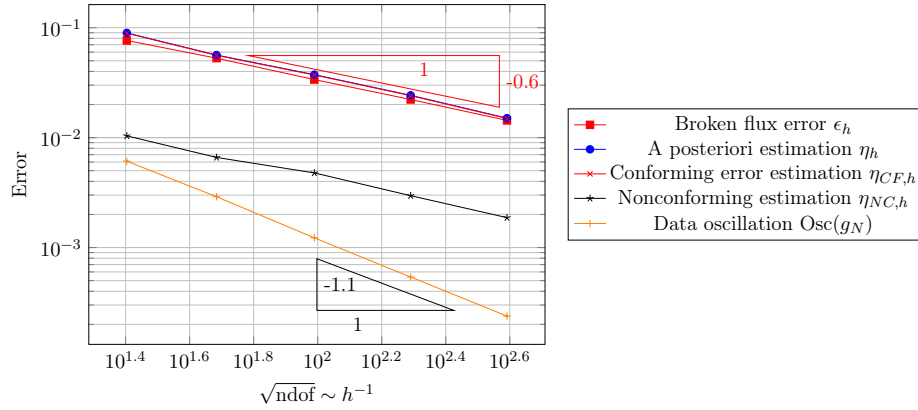


Figure 29: Comparison of the different components of the error estimator and the error in log-log scale for  $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom.

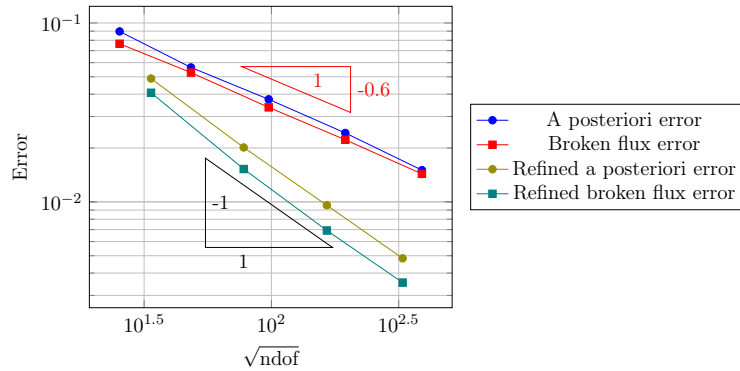


Figure 30: A posteriori error and broken flux error for global and local refinement. Log-log scale. Abscissa: number of degrees of freedom.

We can nevertheless note that it is still possible to obtain super-convergence of the simulation, and that we have reached the optimal case which is convergence of order 1 according to the appendix.

<sup>12</sup> $\ell^2$  is the matrix norm subordinate to the 2-norm in  $\mathbb{R}^2$

## 2) Heterogeneous diffusion problem

We consider the problem on  $\Omega = ]0; 1[^2$  partitioned into four subdomains

$$\begin{cases} \Omega_1 = ]1/2; 1[^2, \\ \Omega_2 = ]0; 1/2[ \times ]1/2; 1[, \\ \Omega_3 = ]0; 1/2[^2, \\ \Omega_4 = ]1/2; 1[ \times ]0; 1/2[. \end{cases}$$

We equip  $\Omega$  with polar coordinates defined as in Fig.31.

We define the interfaces  $\Gamma_{ij} := \overline{\Omega_i} \cap \overline{\Omega_j}$ .

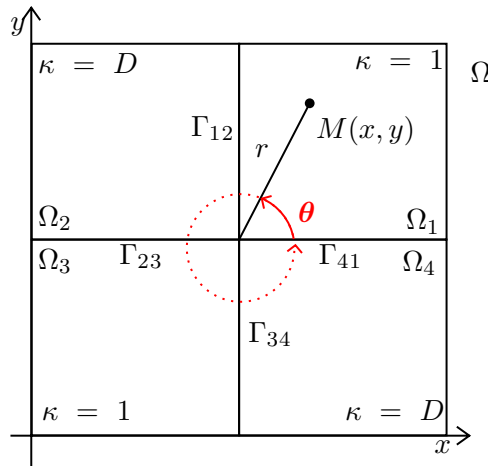


Figure 31: Domain  $\Omega$  considered and its polar coordinates.

On  $\Omega_i$  we denote  $\kappa_i$  the diffusion in the domain, and in the following we will always have  $\kappa_1 = \kappa_3 = 1$ . We seek the solution to the harmonic problem:

$$\text{Find } u \in H^1(\Omega), \quad \begin{cases} \operatorname{div}_x \cdot (\kappa \nabla_x u) = 0 & \text{in } \Omega, \\ u = r^\alpha (a \cos(\alpha\theta) + b \sin(\alpha\theta)) & \text{on } \partial\Omega. \end{cases}$$

where  $a = a(D, \alpha) \in L^\infty(\Omega)$  and  $b = b(D, \alpha) \in L^\infty(\Omega)$  are piecewise constant functions on the  $\Omega_i$ . The calculation method to obtain the value of  $\alpha$  and  $a, b$  as a function of  $D$  or to obtain  $D$  and  $a, b$  as a function of  $\alpha$  is detailed in Appendix D. We fix here  $\alpha = 1/4$  and we find numerically  $D = 0.0395661$ .

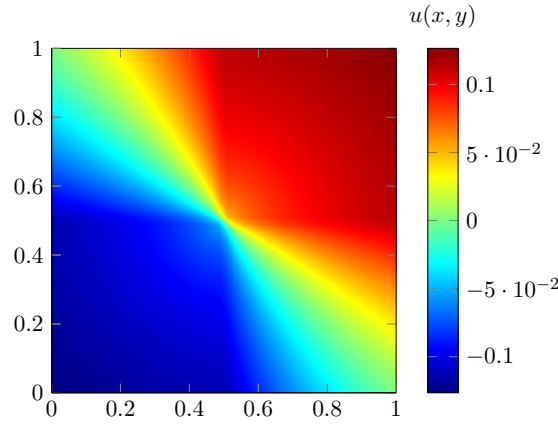


Figure 32: Solution  $f(r, \theta) = r^{1/4}(a \cos(\theta/4) + b \sin(\theta/4))$  for  $\alpha = 1/4$  and  $D = 0.0395661$ .

### a) SIP method

We propose here the analysis of results in the case of heterogeneous diffusion without weighting of averages.

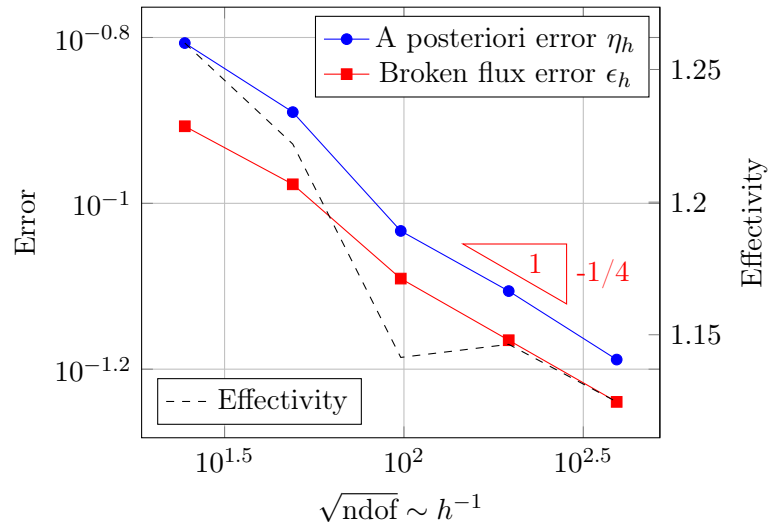


Figure 33: Comparison of the different components of the error estimator and the error in log-log scale for the L-shaped domain and  $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom.

We note that the error is indeed at  $h^{1/4}$  which is the convergence order predicted by the estimate in Theorem 2.11. The effectivity very close to 1 is an indicator of very good quality of the error estimation! We notice a change in behavior at the level of the nonconforming error. The nonconforming error estimation has large amplitude compared to the homogeneous case. This is probably explained by the sudden changes in variations at the interfaces.

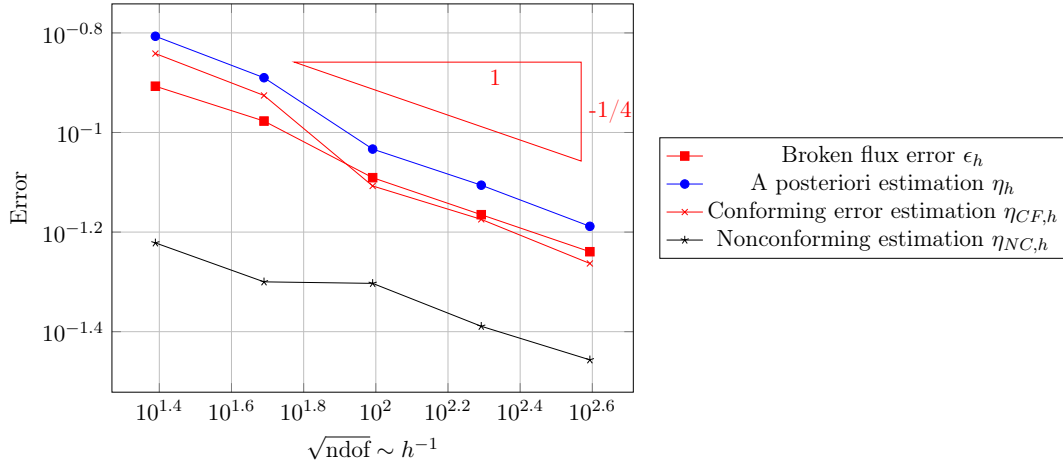


Figure 34: Comparison of the different components of the error estimator and the error in log-log scale for  $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom.

### b) SWIP method

Here, the estimations are made with the weightings presented in the SWIP section.

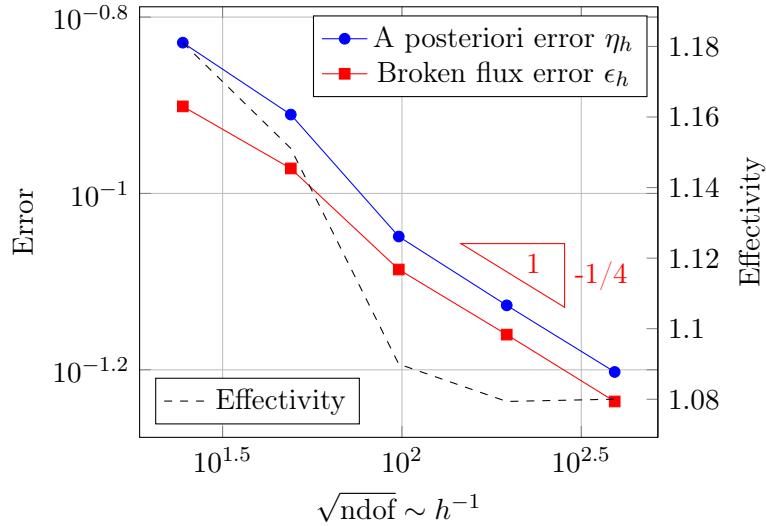


Figure 35: Comparison of the different components of the error estimator and the error in log-log scale for the L-shaped domain and  $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom.

Once again, the error is indeed at  $h^{1/4}$  which is the convergence order predicted by the estimate in Theorem 2.11. The effectivity very close to 1 is again an indicator of very good quality of the error estimation!

We observe the same behavior at the level of the nonconforming error as in Fig. 34, which confirms the good behavior in both the SIP case and the SWIP case of the estimator.

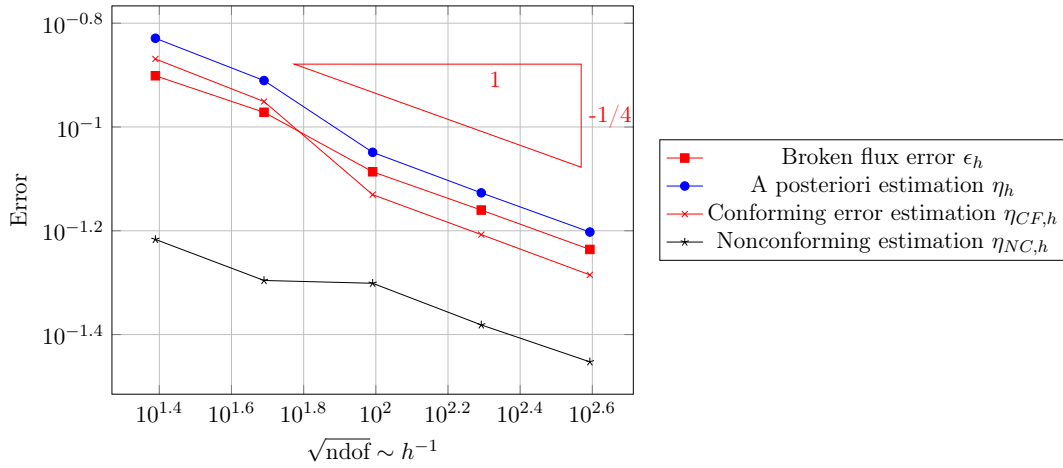


Figure 36: Comparison of the different components of the error estimator and the error in log-log scale for  $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom.

### c) Comparison of the two methods

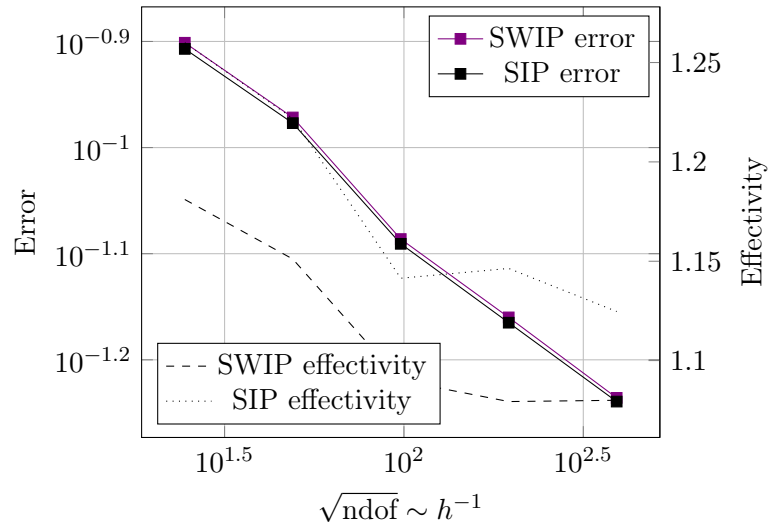


Figure 37: Comparison of the different components of the error estimator and the error in log-log scale for the L-shaped domain and  $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom.

We compare here the error in the SIP case and in the SWIP case. It is interesting to note that the SWIP error is slightly higher than the SIP error. No test case has been implemented to verify that SWIP allows a better approximation than SIP.

However, the difference is negligible and the behavior of the estimator in the SWIP case is of better quality than in the SIP case as evidenced by the effectivity.

## Conclusion

During this internship, an a posteriori error estimator for Discontinuous Galerkin Methods proposed by Mark AINSWORTH in 2007 [AIN07] was implemented in the Matlab prototype of TrioCFD. Many test cases were considered, but only those that allowed for a complete analysis are presented here.

The estimator allows the evaluation of the error for isotropic heterogeneous diffusion problems on various domains: L-shape, square, and perforated domain (not shown in this report).

It has demonstrated good performance for polynomial spaces  $\mathbb{P}_1^1(\mathcal{T}_h)$ .

The estimation led to superconvergence in irregular cases, which will eventually reduce computational costs.

However, the current estimator is not suitable for higher-order polynomials.

Moreover, it is only applicable to the isotropic diffusion operator, and the computations are implemented on simplicial meshes.

Several improvements can therefore be considered:

- Build the estimation on general meshes [EV09; CYZ11].
- Extend the estimation to higher-order polynomials [EV09; ESV10].
- Develop an estimator with continuous Dirichlet data.
- Design an estimator for Diffusion–Advection–Reaction problems [DE12, Section 4.6, 5.6] and [ESV10].

These references, among others, also provide opportunities to construct estimators with a priori estimates on their effectivity and thus on the quality of the estimation.

In parallel, Discontinuous Galerkin methods for diffusion with mixed boundary conditions on general affine domains were implemented.

The next step will be to implement Discontinuous Galerkin methods for systems such as Stokes [DE12, Section 6.1] and then Navier–Stokes [DE12, Section 6.2].

## Part 4

## Appendix

## A SWIP formulation for nonhomogeneous conditions

We present here the variational formulations for pure Dirichlet problems, pure Neumann problems and a formulation with mixed conditions.

**Remark :**

Unlike the case where  $\omega_1 = \omega_2 = 1/2$ , we cannot express the jump of a product by a symmetric quantity. We have indeed

$$[[ab]] = \{a\}_\omega [b] + [a] \{b\}_{\bar{\omega}},$$

where  $\{\cdot\}_{\bar{\omega}}$  is the *antisymmetric average*

$$\{X\}_{\bar{\omega}} = \omega_2 X_1 + \omega_1 X_2.$$

**1) Pure Dirichlet**

Let  $f \in L^2(\Omega)$ ,  $g_D \in H^{1/2}(\partial\Omega)$  and  $\kappa \in L^\infty(\Omega)$ ,  $\kappa \geq \alpha > 0$  piecewise constant.

$$\text{Find } u \in H^1(\Omega) \text{ such that } \begin{cases} -\operatorname{div}(\kappa \nabla u) = f & \text{in } \Omega, \\ u = g_D & \text{on } \partial\Omega. \end{cases} \quad (48)$$

We define  $H_{g_D}^1(\Omega) := \{v \in H^1(\Omega) \mid v = g_D \text{ on } \Gamma_D\}$ .

We set  $V := H_{g_D}^1(\Omega) \cap H^{3/2}(\Omega)$  and recall that  $V_{h,\star} := V + V_h$ .

To derive the discrete variational problem, we will verify consistency at all steps of the heuristic derivation of the SWIP method.

Let  $v \in V_{h,\star}$ ,  $w_h \in V_h$ :

$$\begin{aligned} a^{(0)}(u, w_h) &= \int_{\Omega} \kappa \nabla_h v \cdot \nabla_h w_h \\ &= \sum_{T \in \mathcal{T}_h} \int_T \kappa \nabla v \cdot \nabla w_h \\ &= \sum_{T \in \mathcal{T}_h} \sum_{F \in \mathcal{F}_T} \int_F \kappa \nabla v \cdot \mathbf{n}_{F,T} w_h - \int_T \operatorname{div}(\kappa \nabla_h v) w_h \\ &= \sum_{F \in \mathcal{F}_h} \int_F [[\kappa \nabla_h v \cdot \mathbf{n}_F w_h]] - \int_{\Omega} \operatorname{div}(\kappa \nabla v) w_h \\ &= \sum_{F \in \mathcal{F}_h^i} \int_F [[\kappa \nabla_h v \cdot \mathbf{n}_F]] \{w_h\}_{\bar{\omega}} + \{[\kappa \nabla_h v \cdot \mathbf{n}_F]\}_{\omega} [w_h] + \sum_{F \in \mathcal{F}_h^b} \kappa \nabla v \cdot \mathbf{n}_w - \int_{\Omega} \operatorname{div}(\kappa \nabla_h v) w_h \\ &= \sum_{F \in \mathcal{F}_h^i} \int_F [[\kappa \nabla_h v \cdot \mathbf{n}_F]] \{w_h\}_{\bar{\omega}} + \sum_{F \in \mathcal{F}_h} \{[\kappa \nabla_h v \cdot \mathbf{n}_F]\}_{\omega} [w_h] - \int_{\Omega} \operatorname{div}(\kappa \nabla_h v) w_h. \end{aligned}$$

By evaluating for the solution  $u \in V$  we notice the same consistency problem appears, we make the formula consistent by subtracting

$$\sum_{F \in \mathcal{F}_h} \int_F \{[\kappa \nabla_h v \cdot \mathbf{n}_F]\}_{\omega} [w_h].$$



Then we wish to symmetrize by subtracting the following sum:

$$\begin{aligned} \sum_{F \in \mathcal{F}_h} \int_F \{\kappa \nabla_h w_h \cdot \mathbf{n}_F\}_\omega [v] &= \sum_{F \in \mathcal{F}_h^i} \int_F \{\kappa \nabla_h w_h \cdot \mathbf{n}_F\}_\omega [v] + \sum_{F \in \mathcal{F}_h^b} \int_F \kappa \nabla_h w_h \cdot \mathbf{n}_F [v] \\ &= \underbrace{\sum_{F \in \mathcal{F}_h^i} \int_F \{\kappa \nabla_h w_h \cdot \mathbf{n}_F\}_\omega [v]}_{\text{bilinear part}} + \underbrace{\sum_{F \in \mathcal{F}_h^b} \int_F \kappa \nabla_h w_h \cdot \mathbf{n}_F g_D}_{\text{linear part}}. \end{aligned}$$

We notice that by applying the Dirichlet conditions at the boundary, we lose symmetry on the bilinear part because we make  $[\nabla_h w_h \cdot \mathbf{n}_F]_\omega [v_h]$  appear only for interior faces. We then rather subtract

$$\sum_{F \in \mathcal{F}_h} \int_F \{\kappa \nabla_h w_h \cdot \mathbf{n}_F\}_\omega [v] + \sum_{F \in \mathcal{F}_h^b} \int_F \kappa \nabla_h w_h \cdot \mathbf{n}_F g_D.$$

We then penalize to make the formulation coercive with  $s^\kappa(v, w_h)$  which is not consistent for the solution  $u$ :

$$s^\kappa(u, w_h) = \sum_{F \in \mathcal{F}_h^b} \frac{\gamma_F \gamma_\kappa}{|F|} \int_F g_D w_h.$$

It is therefore necessary to subtract this part to finally obtain the exact variational formulation:

$$a_h^{\text{swip}}(u, w_h) - \left( - \sum_{F \in \mathcal{F}_h^b} \int_F \kappa \nabla_h w_h \cdot \mathbf{n}_F g_D + \sum_{F \in \mathcal{F}_h^b} \frac{\gamma_F \gamma_\kappa}{|F|} \int_F g_D w_h \right) = \int_\Omega f w_h,$$

where  $a_h^{\text{swip}}$  is the bilinear form defined in Equation (17).

However, since the part in parentheses is not bilinear in  $(v_h, w_h)$  but rather linear in  $w_h$ , we define a new linear form

$$\ell_h(w_h ; g_D) := \int_\Omega f w_h - \int_{\partial\Omega} \kappa \nabla_h w_h \cdot \mathbf{n}_F g_D + \sum_{F \in \mathcal{F}_h^b} \frac{\gamma_F}{h_F} \gamma_\kappa \int_F g_D w_h.$$

And the discrete problem is equivalent to finding a solution  $v_h \in V_h$  of

$$a_h^{\text{swip}}(v_h, w_h) = \ell_h(w_h ; g_D).$$

## 2) Pure Neumann

Let  $f \in L^2(\Omega)$ ,  $g_N \in H^{1/2}(\partial\Omega)$ , and  $\kappa \in L^\infty(\Omega)$ ,  $\kappa \geq \alpha > 0$  piecewise constant.

$$\text{Find } u \in H^1(\Omega) \text{ such that } \begin{cases} -\text{div}(\kappa \nabla u) = f & \text{in } \Omega, \\ \kappa \nabla u \cdot \mathbf{n} = g_N & \text{on } \partial\Omega. \end{cases} \quad (49)$$

### Remark :

For the problem to be well-posed, the data  $f$  and  $g_N$  must satisfy a **compatibility condition** obtained with Stokes' formula:

$$\int_\Omega f + \int_{\partial\Omega} g = 0.$$

and the solution is defined up to an additive constant [DE12, Section 4.2 p.127-128].

We define  $V_N := \{v \in H^1(\Omega) \mid \nabla v = g_D \text{ on } \Gamma_N\}$ .

We set  $V := V_N \cap H^{3/2}(\Omega)$  and recall that  $V_{h,\star} := V + V_h$ .

In the same way as before, we will start from the non-consistent formula and modify the variational

formulation to ensure consistency.

Let  $v \in V_{h,\star}$ ,  $w_h \in V_h$ .

$$\begin{aligned}
a_h^{(0)}(v, w_h) &= \sum_{T \in \mathcal{T}_h} \int_T \kappa \nabla v \cdot \nabla w_h \\
&= \sum_{T \in \mathcal{T}_h} \sum_{F \in \mathcal{F}_T} \int_F \kappa \nabla v \cdot \mathbf{n}_{F,T} w_h - \int_T \operatorname{div}(\kappa \nabla_h v) w_h \\
&= \sum_{F \in \mathcal{F}_h^i} \int_F \llbracket \kappa \nabla_h v \cdot \mathbf{n}_F \rrbracket \llbracket w_h \rrbracket_{\bar{\omega}} + \llbracket \kappa \nabla_h v \cdot \mathbf{n}_F \rrbracket_{\omega} \llbracket w_h \rrbracket + \sum_{F \in \mathcal{F}_h^b} \int_F \kappa \nabla v \cdot \mathbf{n}_F w_h - \int_{\Omega} \operatorname{div}(\kappa \nabla_h v) w_h \\
&= \sum_{F \in \mathcal{F}_h^i} \int_F \llbracket \kappa \nabla_h v \cdot \mathbf{n}_F \rrbracket \llbracket w_h \rrbracket_{\bar{\omega}} + \llbracket \kappa \nabla_h v \cdot \mathbf{n}_F \rrbracket_{\omega} \llbracket w_h \rrbracket + \int_{\partial\Omega} g_N \llbracket w_h \rrbracket - \int_{\Omega} \operatorname{div}(\kappa \nabla_h v) w_h.
\end{aligned}$$

By evaluating at the solution  $u \in V$  we obtain the following consistency formula:

$$\sum_{F \in \mathcal{F}_h^i} \int_F \llbracket \kappa \nabla_h v \cdot \mathbf{n}_F \rrbracket_{\omega} \llbracket w_h \rrbracket + \underbrace{\sum_{F \in \mathcal{F}_h^b} \int_F g_N w_h + \int_{\Omega} f w_h}_{\text{linear form } \ell_h^R(w_h; g_N)}.$$

We make it consistent and symmetric by subtracting

$$\sum_{F \in \mathcal{F}_h^i} \int_F \llbracket \kappa \nabla_h v \cdot \mathbf{n}_F \rrbracket_{\omega} \llbracket w_h \rrbracket + \sum_{F \in \mathcal{F}_h^i} \int_F \llbracket \kappa \nabla_h w_h \cdot \mathbf{n}_F \rrbracket_{\omega} \llbracket v \rrbracket.$$

The numerical analysis of coercivity shows that it suffices to penalize only inside the mesh, we therefore add the consistent penalization

$$s^{\kappa}(v, w_h) = \sum_{F \in \mathcal{F}_h^i} \frac{\gamma_F \gamma_{\kappa}}{|F|} \int_F \llbracket v \rrbracket \llbracket w_h \rrbracket.$$

To finally obtain the following Neumann variational formulation:

$$a_h^N(v, w_h) := \sum_{T \in \mathcal{T}_h} \int_T \kappa \nabla v \cdot \nabla w_h + \sum_{F \in \mathcal{F}_h^i} \int_F \llbracket \kappa \nabla_h v \cdot \mathbf{n} \rrbracket_{\omega} \llbracket w_h \rrbracket + \llbracket \kappa \nabla_h w_h \cdot \mathbf{n} \rrbracket_{\omega} \llbracket v \rrbracket + \sum_{F \in \mathcal{F}_h^i} \frac{\gamma_F \gamma_{\kappa}}{|F|} \int_F \llbracket v \rrbracket \llbracket w_h \rrbracket,$$

so that the variational problem becomes

$$\text{Find } u \in V, \quad a_h^N(u, w_h) = \ell_h^N(w_h; g_N), \quad \forall w_h \in V_h.$$

Finally, to impose a constant to the problem, we perform a lifting so that

$$\int_{\Omega} u = 0.$$

### 3) Mixed boundary conditions

Let  $\Gamma_N, \Gamma_D$  such that  $\partial\Omega = \overline{\Gamma_N} \cup \overline{\Gamma_D}$  and  $\Gamma_N \cap \Gamma_D = \emptyset$ .

Let  $f \in L^2(\Omega)$ ,  $g_N \in L^2(\Gamma_N)$ ,  $g_D \in H^{1/2}(\Gamma_D)$  and  $\kappa \in L^{\infty}(\Omega)$ ,  $\kappa \geq \alpha > 0$  piecewise constant.

$$\text{Find } u \in H^1(\Omega) \text{ such that } \begin{cases} -\operatorname{div}(\kappa \nabla u) = f & \text{in } \Omega, \\ u = g_D & \text{on } \Gamma_D, \\ \kappa \nabla u \cdot \mathbf{n} = g_N & \text{on } \Gamma_N. \end{cases} \quad (50)$$

Once again, the idea is to start from the formulation  $a^{(0)}(v, w_h) = \int_{\Omega} \kappa \nabla v \cdot \nabla w_h$ .

We denote in what follows  $\mathcal{F}_h^D := \mathcal{F}_h \cap \Gamma_D$  and  $\mathcal{F}_h^N := \mathcal{F}_h \cap \Gamma_N$ .

We can notice that it suffices to separate the Neumann part and the Dirichlet part at the boundary as follows:

$$\begin{aligned} a_h^{(0)}(v, w_h) &= \sum_{F \in \mathcal{F}_h^i} \int_F \llbracket \kappa \nabla_h v \cdot \mathbf{n}_F \rrbracket \llbracket w_h \rrbracket_{\bar{\omega}} + \sum_{F \in \mathcal{F}_h^i \sqcup \mathcal{F}_h^D} \int_F \llbracket \kappa \nabla_h v \cdot \mathbf{n}_F \rrbracket_{\omega} \llbracket w_h \rrbracket \\ &\quad + \sum_{F \in \mathcal{F}_h^N} \int_F g_N w_h - \int_{\Omega} \operatorname{div}(\kappa \nabla_h v) w_h. \end{aligned}$$

To obtain consistency of the bilinear part, we must then subtract

$$\sum_{F \in \mathcal{F}_h^i \sqcup \mathcal{F}_h^D} \int_F \llbracket \kappa \nabla_h v \cdot \mathbf{n}_F \rrbracket_{\omega} \llbracket w_h \rrbracket.$$

Then we symmetrize by subtracting

$$\sum_{F \in \mathcal{F}_h^i \sqcup \mathcal{F}_h^D} \int_F \llbracket \kappa \nabla_h v \cdot \mathbf{n}_F \rrbracket_{\omega} \llbracket w_h \rrbracket = \sum_{F \in \mathcal{F}_h^i} \int_F \llbracket \kappa \nabla_h v \cdot \mathbf{n}_F \rrbracket_{\omega} \llbracket w_h \rrbracket + \sum_{F \in \mathcal{F}_h^D} \int_F \kappa \nabla w_h \cdot \mathbf{n}_F g_D.$$

As we noticed in the study of pure Dirichlet, we must still subtract by

$$\sum_{F \in \mathcal{F}_h^D} \int_F \kappa \nabla w_h \cdot \mathbf{n}_F v,$$

to maintain the symmetry of the bilinear part.

We then verify that the consistent penalization that makes the formulation coercive is of the form

$$\tilde{s}^{\kappa}(u, w_h) = \underbrace{\sum_{F \in \mathcal{F}_h^i \sqcup \mathcal{F}_h^D} \frac{\gamma_F \gamma_{\kappa}}{|F|} \int_F \llbracket v \rrbracket \llbracket w_h \rrbracket}_{\text{consistent and bilinear}} + \underbrace{\sum_{F \in \mathcal{F}_h^D} \frac{\gamma_F \gamma_{\kappa}}{|F|} \int_F w_h g_D}_{\text{linear}}.$$

We must then still subtract

$$\sum_{F \in \mathcal{F}_h^D} \frac{\gamma_F \gamma_{\kappa}}{|F|} \int_F w_h g_D,$$

so that the penalization becomes

$$s_h^{\kappa}(v, w_h) = \sum_{F \in \mathcal{F}_h^i \sqcup \mathcal{F}_h^D} \frac{\gamma_F \gamma_{\kappa}}{|F|} \int_F \llbracket v \rrbracket \llbracket w_h \rrbracket.$$

We finally obtain the continuous and coercive bilinear form

$$a_h^{\text{mixed}}(v, w_h) := \sum_{T \in \mathcal{T}_h} \int_T \kappa \nabla v \cdot \nabla w_h - \sum_{F \in \mathcal{F}_h^i \sqcup \mathcal{F}_h^D} \int_F \llbracket \kappa \nabla_h v \cdot \mathbf{n}_F \rrbracket_{\omega} \llbracket w_h \rrbracket + \llbracket \kappa \nabla_h w_h \cdot \mathbf{n}_F \rrbracket_{\omega} \llbracket v \rrbracket + s_h^{\kappa}(v, w_h)$$

and the linear form

$$\ell_h^{\text{mixed}}(w_h) := \int_{\Omega} f w_h + \sum_{F \in \mathcal{F}_h^D} \frac{\gamma_F \gamma_{\kappa}}{|F|} \int_F w_h g_D - \int_{\Gamma_D} \kappa \nabla w_h \cdot \mathbf{n}_F g_D + \int_{\Gamma_N} g_N w_h,$$

so that the problem is then

$$\text{Find } u \in V, \quad a_h^{\text{mixed}}(u, w_h) = \ell_h^{\text{mixed}}(w_h), \quad \forall w_h \in V_h. \quad (51)$$

**Remark :**

These formulations amount to weakly imposing the Dirichlet condition by penalizing the jump between  $u_h$  and  $g_D$  at the domain boundary, which are generally imposed in a strong manner in continuous Finite Elements.

## B Equivalent formulation

In this part, we rewrite the formulation (51) with an equivalent bilinear form that was used for test cases without weights (SIP). On one hand, by performing integration by parts on  $w_h$ :

$$\begin{aligned}
a_h^{\text{mixed}}(v_h, w_h) &= \sum_{T \in \mathcal{T}_h} \int_T \kappa \nabla v_h \cdot \nabla w_h - \sum_{F \in \mathcal{F}_h^i \sqcup \mathcal{F}_h^D} \int_F \{ \kappa \nabla_h v_h \cdot \mathbf{n}_F \}_\omega [w_h] + \{ \kappa \nabla_h w_h \cdot \mathbf{n}_F \}_\omega [v_h] + s^\kappa(v_h, w_h) \\
&= \sum_{T \in \mathcal{T}_h} \int_T -\text{div}(\kappa \nabla v_h) w_h + \int_{\partial T} \kappa \nabla v_h \cdot \mathbf{n}_{\partial T} w_h \\
&\quad - \sum_{F \in \mathcal{F}_h^i \sqcup \mathcal{F}_h^D} \int_F \{ \kappa \nabla_h v_h \cdot \mathbf{n}_F \}_\omega [w_h] + \{ \kappa \nabla_h w_h \cdot \mathbf{n}_F \}_\omega [v_h] + s^\kappa(v_h, w_h) \\
&= \sum_{T \in \mathcal{T}_h} \int_T -\text{div}(\kappa \nabla v_h) w_h + \sum_{T \in \mathcal{T}_h} \sum_{F \in \mathcal{F}_T} \int_F \kappa \nabla v_h \cdot \mathbf{n}_{T,F} w_h \\
&\quad - \sum_{F \in \mathcal{F}_h^i \sqcup \mathcal{F}_h^D} \int_F \{ \kappa \nabla_h v_h \cdot \mathbf{n}_F \}_\omega [w_h] + \{ \kappa \nabla_h w_h \cdot \mathbf{n}_F \}_\omega [v_h] + \sum_{F \in \mathcal{F}_h^i \sqcup \mathcal{F}_h^D} \frac{\gamma_F \gamma_\kappa}{|F|} \int_F [v_h] [w_h] \\
&= \sum_{T \in \mathcal{T}_h} \int_T -\text{div}(\kappa \nabla v_h) w_h + \sum_{F \in \mathcal{F}_h^i} \int_F [ \kappa \nabla_h v_h \cdot \mathbf{n}_F w_h ] + \sum_{F \in \mathcal{F}_h^b} \kappa \nabla v_h \cdot \mathbf{n} w_h \\
&\quad - \sum_{F \in \mathcal{F}_h^i \sqcup \mathcal{F}_h^D} \int_F \{ \kappa \nabla_h v_h \cdot \mathbf{n}_F \}_\omega [w_h] + \{ \kappa \nabla_h w_h \cdot \mathbf{n}_F \}_\omega [v_h] + s^\kappa(v_h, w_h) \\
&= \sum_{T \in \mathcal{T}_h} \int_T -\text{div}(\kappa \nabla v_h) w_h + \sum_{F \in \mathcal{F}_h^i} \int_F \{ \kappa \nabla_h v_h \cdot \mathbf{n}_F \}_\omega [w_h] + [ \kappa \nabla_h v_h \cdot \mathbf{n}_F ] \{ w_h \}_\omega \\
&\quad + \sum_{F \in \mathcal{F}_h^b} \kappa \nabla v_h \cdot \mathbf{n} w_h - \sum_{F \in \mathcal{F}_h^i \sqcup \mathcal{F}_h^D} \int_F \{ \kappa \nabla_h v_h \cdot \mathbf{n}_F \}_\omega [w_h] + \{ \kappa \nabla_h w_h \cdot \mathbf{n}_F \}_\omega [v_h] + s^\kappa(v_h, w_h) \\
&= \sum_{T \in \mathcal{T}_h} \int_T -\text{div}(\kappa_T \nabla v_h) w_h + \sum_{F \in \mathcal{F}_h^N} \int_F \{ \kappa \nabla_h v_h \cdot \mathbf{n}_F \}_\omega [w_h] \\
&\quad - \sum_{F \in \mathcal{F}_h^i \sqcup \mathcal{F}_h^D} \int_F \{ \kappa \nabla_h w_h \cdot \mathbf{n}_F \}_\omega [v_h] + \sum_{F \in \mathcal{F}_h^i} [ \kappa \nabla_h v_h \cdot \mathbf{n}_F ] \{ w_h \}_\omega + s_h^\kappa(v_h, w_h).
\end{aligned}$$

By performing integration by parts on  $v_h$  on the other hand:

$$\begin{aligned}
a_h^{\text{mixed}}(v_h, w_h) &= \sum_{T \in \mathcal{T}_h} \int_T -\text{div}(\kappa_T \nabla w_h) v_h + \sum_{F \in \mathcal{F}_h^N} \int_F \{ \kappa \nabla_h w_h \cdot \mathbf{n}_F \}_\omega [v_h] - \sum_{F \in \mathcal{F}_h^i \sqcup \mathcal{F}_h^D} \int_F \{ \kappa \nabla_h v_h \cdot \mathbf{n}_F \}_\omega [w_h] \\
&\quad + \sum_{F \in \mathcal{F}_h^i} [ \kappa \nabla_h w_h \cdot \mathbf{n}_F ] \{ v_h \}_\omega + s_h^\kappa(w_h, v_h).
\end{aligned}$$

To maintain symmetry, we take the average of the two expressions and obtain:

$$\begin{aligned}
a_h^{\text{mixed}}(v_h, w_h) &= -\frac{1}{2} \sum_{T \in \mathcal{T}_h} \int_T \kappa_T (\Delta w_h v_h + w_h \Delta v_h) + \frac{1}{2} \sum_{F \in \mathcal{F}_h^N} \int_F \{ \kappa \nabla_h w_h \cdot \mathbf{n}_F \}_\omega [v_h] + \{ \kappa \nabla_h v_h \cdot \mathbf{n}_F \}_\omega [w_h] \\
&\quad - \frac{1}{2} \sum_{F \in \mathcal{F}_h^i \sqcup \mathcal{F}_h^D} \int_F \{ \kappa \nabla_h v_h \cdot \mathbf{n}_F \}_\omega [w_h] + \{ \kappa \nabla_h w_h \cdot \mathbf{n}_F \}_\omega [v_h] \\
&\quad + \frac{1}{2} \sum_{F \in \mathcal{F}_h^i} [ \kappa \nabla_h w_h \cdot \mathbf{n}_F ] \{ v_h \}_\omega + [ \kappa \nabla_h v_h \cdot \mathbf{n}_F ] \{ w_h \}_\omega \\
&\quad + s_h^\kappa(w_h, v_h).
\end{aligned}$$

To reduce to a linear system, we will consider  $T_i \in \mathcal{T}_h$  and  $T_j \in \mathcal{T}_h$  two *neighboring* elements with common face  $F$ .

For a polynomial approximation in  $\mathbb{P}_d^k(\mathcal{T}_h)$  and  $T \in \mathcal{T}_h$ , we denote  $K := \dim(\mathbb{P}_d^k(T)) = \binom{k+d}{k}$ .

We denote  $(\varphi_m)_{m \in \llbracket 1;K \rrbracket}$  the elements of the local basis of  $T_i$  and  $(\psi_m)_{m \in \llbracket 1;K \rrbracket}$  those of  $T_j$ .

If  $F$  is not on the boundary and if  $T_i \neq T_j$ :

$$\begin{aligned} a_h^{\text{mixed}}(\varphi_i, \psi_j) &= -\frac{1}{2} \int_F \kappa_1 \omega_1 \nabla \varphi_i \cdot \mathbf{n}_{T_i, F} (-\psi_j) + \kappa_2 \omega_2 \nabla \psi_j \cdot \mathbf{n}_{T_i, F} \varphi_i \\ &\quad + \frac{1}{2} \int_F -\kappa_2 \omega_2 \nabla \psi_j \cdot \mathbf{n}_{T_i, F} \varphi_i + \kappa_1 \omega_1 \nabla \varphi_i \cdot \mathbf{n}_{T_i, F} \psi_j + \frac{\gamma_F \gamma_\kappa}{|F|} \int_F \varphi_i \psi_j \\ &= \int_F \omega_1 \kappa_1 \nabla \varphi_i \cdot \mathbf{n}_{T_i, F} \psi_j - \omega_2 \kappa_2 \nabla \psi_j \cdot \mathbf{n}_{T_i, F} \varphi_i + \frac{\gamma_F \gamma_\kappa}{|F|} \int_F \varphi_i \psi_j. \end{aligned}$$

If  $F$  is not on the boundary and  $T_i = T_j = T$  then  $\omega_1 = \omega_2 = 1/2$  and:

$$a_h^{\text{mixed}}(\varphi_i, \varphi_j) = -\frac{1}{2} \int_T \kappa_T \Delta \varphi_i \varphi_j - \Delta \varphi_j \varphi_i - \frac{1}{2} \int_F \kappa_T \nabla(\varphi_i \varphi_j) \cdot \mathbf{n}_{T_i, F} + \frac{\gamma_F \gamma_\kappa}{|F|} \int_F \varphi_i \varphi_j.$$

If  $F$  is on the Dirichlet boundary then  $T_i = T_j = T$  then  $\omega_1 = \omega_2 = 1/2$  and:

$$\begin{aligned} a_h^{\text{mixed}}(\varphi_i, \varphi_j) &= -\frac{1}{2} \int_{T_i} \kappa_T (\Delta \varphi_i \varphi_j - \Delta \varphi_j \varphi_i) + \frac{1}{2} \frac{\gamma_F \kappa_T}{|F|} \int_F \varphi_i \varphi_j \\ &\quad - \frac{1}{2} \int_F \kappa_T (\nabla \varphi_i \cdot \mathbf{n}_F \varphi_j + \nabla \varphi_j \cdot \mathbf{n}_F \varphi_i) + \frac{1}{2} \int_F \kappa_T (\nabla \varphi_j \cdot \mathbf{n}_F \varphi_i + \nabla \varphi_i \cdot \mathbf{n}_F \varphi_j) \\ &= -\frac{1}{2} \int_{T_i} -\kappa_T (\Delta \varphi_i \varphi_j - \Delta \varphi_j \varphi_i) + \frac{\gamma_F \kappa_T}{|F|} \int_F \varphi_i \varphi_j - \frac{1}{2} \int_F \kappa_T \nabla(\varphi_i \varphi_j) \cdot \mathbf{n}_F. \end{aligned}$$

If  $F$  is on the Neumann boundary then  $T_i = T_j = T$  and:

$$a_h^{\text{mixed}}(\varphi_i, \varphi_j) = -\frac{1}{2} \int_{T_i} -\kappa_T (\Delta \varphi_i \varphi_j - \Delta \varphi_j \varphi_i) + \frac{1}{2} \frac{\gamma_F \kappa_T}{|F|} \int_F \varphi_i \varphi_j + \frac{1}{2} \int_F \kappa_T \nabla(\varphi_i \varphi_j) \cdot \mathbf{n}_F.$$

## C Gradients and regularity of test cases

### 1) Spectral problem on the square domain

We recall that in the case of the spectral problem in the square domain, for  $\lambda \in \mathbb{N}^*$  fixed, we seek to solve the following problem:

$$\text{Find } \boldsymbol{\sigma}(u) \in \mathbf{H}(\text{div}; \Omega), u \in H^1(\Omega) \text{ such that } \begin{cases} -\text{div } \boldsymbol{\sigma} = 2(\lambda\pi)^2 \sin(\lambda\pi x) \sin(\lambda\pi y) & \text{in } \Omega, \\ \boldsymbol{\sigma}(u) = \nabla u & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \quad (44)$$

with solution

$$u_\lambda : (x, y) \in \mathbb{R}^2 \mapsto \sin(\lambda\pi x) \sin(\lambda\pi y).$$

Its gradient is then:

$$\nabla u_\lambda = \lambda\pi \begin{bmatrix} \cos(\lambda\pi x) \sin(\lambda\pi y) \\ \sin(\lambda\pi x) \cos(\lambda\pi y) \end{bmatrix}.$$

And  $u$  is  $\mathcal{C}^\infty(\Omega)$  because  $\sin$  is a  $\mathcal{C}^\infty(\mathbb{R}^2)$  function therefore  $\mathcal{C}^\infty(\Omega)$ .

According to Corollary 2.9, we therefore expect convergence at  $h^1$  for polynomials  $v_h \in \mathbb{P}_1^1(\mathcal{T}_h)$ . Réessayer

### 2) Harmonic problem

We recall that in the case of the harmonic problem we solve:

$$\text{Find } u \in H^1(\Omega), \quad \begin{cases} -\Delta_x u = 0 & \text{in } \Omega, \\ u(r, \theta) = r^\alpha \sin(\alpha\theta) & \text{on } \partial\Omega. \end{cases}$$

The gradient of the function is calculated as follows:

$$\begin{aligned} \nabla_x f(r, \theta) &= \frac{\partial f}{\partial r} \mathbf{u}_r + \frac{1}{r} \frac{\partial f}{\partial \theta} \mathbf{u}_\theta \\ &= \frac{\partial f}{\partial r} (\cos(\theta) \mathbf{u}_x + \sin(\theta) \mathbf{u}_y) + \frac{1}{r} \frac{\partial f}{\partial \theta} (-\sin(\theta) \mathbf{u}_x + \cos(\theta) \mathbf{u}_y) \\ &= \left( \cos(\theta) \frac{\partial f}{\partial r} - \sin(\theta) \frac{1}{r} \frac{\partial f}{\partial \theta} \right) \mathbf{u}_x + \left( \sin(\theta) \frac{\partial f}{\partial r} + \cos(\theta) \frac{1}{r} \frac{\partial f}{\partial \theta} \right) \mathbf{u}_y. \end{aligned}$$

However, we have

$$\begin{aligned} \frac{\partial f}{\partial r} &= \alpha r^{\alpha-1} \sin(\alpha\theta), \\ \frac{1}{r} \frac{\partial f}{\partial \theta} &= \alpha r^{\alpha-1} \cos(\alpha\theta). \end{aligned}$$

We note that  $\|\nabla_x f\|_{\ell^2} = Cr^{\alpha-1}$  is indeed integrable at 0 therefore  $f \in H^1(\Omega)$ .

However

$$\Delta_x f \underset{r \rightarrow 0}{\sim} Cr^{\alpha-2},$$

which is not square integrable at 0 if  $\alpha < 1$ . In other words,  $f \notin H^2(\Omega)$ .

The task is then to find  $s \in ]0; 1[$  such that  $f \in H^{1+s}(\Omega) \iff \nabla_x f \in \mathbf{H}^s(\Omega)$ .

Suppose  $f \in H^s(\Omega)$  for  $s > 0$ . We denote:

$$f_\lambda(x) = f(\lambda x) = \lambda^\alpha f(x).$$

However

$$\|f_\lambda\|_{H^s(\Omega)} = \|\lambda^\alpha f\|_{H^s(\Omega)} = \lambda^\alpha \|f\|_{H^s(\Omega)}.$$

By extending  $f$  by 0 on  $\mathbb{R}^2 \setminus \Omega$ :

$$\begin{aligned}
\|f_\lambda\|_{H^s(\Omega)}^2 &= \int_{\mathbb{R}^2} \frac{|f_\lambda(\mathbf{x}_1) - f_\lambda(\mathbf{x}_2)|}{\|\mathbf{x}_1 - \mathbf{x}_2\|_{\ell^2}^{2+2s}} \\
&= \int_{\mathbb{R}^2} \frac{|f(\lambda\mathbf{x}_1) - f(\lambda\mathbf{x}_2)|}{\|\mathbf{x}_1 - \mathbf{x}_2\|_{\ell^2}^{2+2s}} \\
&= \int_{\mathbb{R}^2} \frac{|f(\mathbf{X}_1) - f(\mathbf{X}_2)|}{\lambda^{-2-2s} \|\mathbf{X}_1 - \mathbf{X}_2\|_{\ell^2}^{2+2s}} \lambda^{-4} \\
&= \lambda^{2(s-1)} \int_{\mathbb{R}^2} \frac{|f(\mathbf{X}_1) - f(\mathbf{X}_2)|}{\|\mathbf{X}_1 - \mathbf{X}_2\|_{\ell^2}^{2+2s}}
\end{aligned}$$

Finally

$$\|f_\lambda\|_{H^s(\Omega)} = \lambda^{s-1} \|f\|_{H^s(\Omega)} = \lambda^\alpha \|f_\lambda\|_{H^s(\Omega)}.$$

We deduce that the regularity  $s$  satisfies:

$$s = \alpha + 1 > 1.$$

And  $f \in H^{\alpha+1}(\Omega)$ . According to the error analysis stated in Thm. 2.11, we expect convergence of the broken flux norm at order  $\min(\alpha, 1)$  for a first-order approximation ( $v_h \in \mathbb{P}_1^1(\mathcal{T}_h)$ ).

## D SWIP solution

This part mainly corresponds to the calculations by Erell JAMELOT which have not been published.

We consider the problem on  $\Omega = ]0; 1[^2$  separated into 4 domains  $\left\{ \begin{array}{l} \Omega_1 = ]1/2; 1[^2, \\ \Omega_2 = ]0; 1/2[ \times ]1/2; 1[, \\ \Omega_3 = ]0; 1/2[^2, \\ \Omega_4 = ]1/2; 1[ \times ]0; 1/2[. \end{array} \right.$

We equip  $\Omega$  with polar coordinates defined as in Fig.31.

We define the boundaries  $\Gamma_{ij} := \overline{\Omega_i} \cap \overline{\Omega_j}$ .

On  $\Omega_i$  we denote by  $\kappa_i$  the diffusion in the domain, and in the following we will always have  $\kappa_1 =$

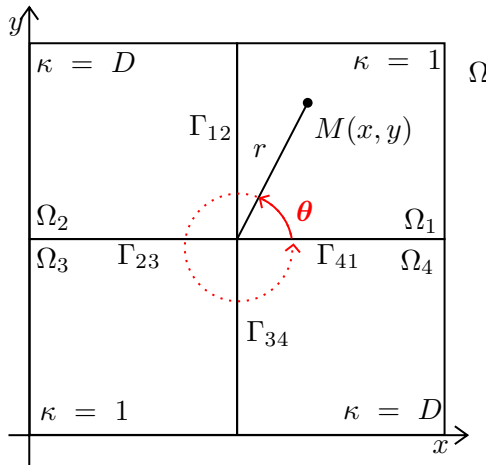


Figure 38: Domain  $\Omega$  and its polar coordinates.

$\kappa_3 = 1$ .

We seek the solution to the harmonic problem:

$$\text{Find } \text{div}_x(\kappa \nabla_x u) = 0. \quad (52)$$

where  $a = a(D, \alpha) \in L^\infty(\Omega)$  and  $b = b(D, \alpha) \in L^\infty(\Omega)$  are piecewise constant functions on the  $\Omega_i$ . We seek  $u$  in the form:

$$u = r^\alpha (a \cos(\alpha\theta) + b \sin(\alpha\theta)).$$

To ensure  $u \in H^1(\Omega)$ , Neumann and Dirichlet transmission conditions must be satisfied at the interfaces:

$$\begin{aligned} u|_{\Omega_i} &= u|_{\Omega_j} && \text{on } \Omega_i \cap \Omega_j, \\ \kappa|_{\Omega_i} \nabla(u|_{\Omega_i}) \cdot \mathbf{n}_{ij} &= \kappa|_{\Omega_j} \nabla(u|_{\Omega_j}) \cdot \mathbf{n}_{ij} && \text{on } \Omega_i \cap \Omega_j. \end{aligned} \quad (53)$$

where  $\mathbf{n}_{ij}$  is the unit normal directed from  $\Omega_i$  to  $\Omega_j$ .

In the following, we define  $x_k = \cos(\alpha\theta_k)$  where  $\theta_k = k\frac{\pi}{2}$ .

Similarly, we define  $y_k = \sin(\alpha\theta_k)$ .

The condition (53) then translates to:

$$\begin{aligned} c_4 &= c_1, & s_4 &= D s_1, \\ c_1 x_1 + s_1 y_1 &= c_2 x_1 + s_2 y_1, & D c_1 y_1 - D s_1 x_1 &= c_2 y_1 - s_2 x_1, \\ c_2 x_2 + s_2 y_2 &= c_3 x_2 - s_3 y_2, & c_2 y_2 - s_2 x_2 &= -D c_3 y_2 - D s_3 x_2, \\ c_3 x_1 - s_3 y_1 &= c_4 x_1 - s_4 y_1, & D c_3 y_1 + D s_3 x_1 &= c_4 y_1 + s_4 x_1. \end{aligned} \quad (54)$$

We can then express these conditions as a linear system:

For  $k = 0, 1, 2$  and  $k' = k + 1$ :

$$\begin{pmatrix} x_{k'} & y_{k'} \\ -D_k y_{k'} & D_k x_{k'} \end{pmatrix} \begin{pmatrix} c_k \\ s_k \end{pmatrix} - \begin{pmatrix} x_{k'} & y_{k'} \\ -D_{k'} y_{k'} & D_{k'} x_{k'} \end{pmatrix} \begin{pmatrix} c_{k'} \\ s_{k'} \end{pmatrix} = 0. \quad (55)$$

On  $\Gamma_{41}$ :

$$\begin{pmatrix} x_4 & y_4 \\ -D y_4 & D x_4 \end{pmatrix} \begin{pmatrix} c_3 \\ s_3 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} c_0 \\ s_0 \end{pmatrix} = 0.$$

Let us set  $\mathbf{A}_n := \begin{pmatrix} x_n & y_n \\ -y_n & x_n \end{pmatrix}$  and  $\mathbf{B}_n := -\begin{pmatrix} x_n & y_n \\ -D y_n & D x_n \end{pmatrix}$ , so that  $|\mathbf{A}_n| = 1$  and  $|\mathbf{B}_n| = D$ .

Equation (55) can be rewritten as:  $\mathbf{M} \mathbf{x} = 0$ , where  $\mathbf{x} := (c_0, s_0, c_1, s_1, c_2, s_2, c_3, s_3)^T$ , and  $\mathbf{M} \in \mathbb{R}^{8 \times 8}$  is the square matrix:

$$\mathbf{M} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{B}_1 & 0 & 0 \\ 0 & \mathbf{B}_2 & \mathbf{A}_2 & 0 \\ 0 & 0 & \mathbf{A}_3 & \mathbf{B}_3 \\ \mathbf{A}_0 & 0 & 0 & \mathbf{B}_4 \end{pmatrix}.$$

To solve  $\mathbf{M} \mathbf{x} = 0$ , we seek the pairs  $(D, \nu)$  that cancel the determinant of  $\mathbf{M}$ .

By expanding  $|\mathbf{M}|$  with respect to the first column:

$$\begin{aligned} |\mathbf{M}| &= x_1 \begin{vmatrix} x_1 & (\mathbf{B}_1)_{2,:} & 0 & 0 \\ 0 & \mathbf{B}_2 & \mathbf{A}_2 & 0 \\ 0 & 0 & \mathbf{A}_3 & \mathbf{B}_3 \\ \begin{pmatrix} 0 \\ 1 \end{pmatrix} & 0 & 0 & \mathbf{B}_4 \end{vmatrix} + y_1 \begin{vmatrix} y_1 & (\mathbf{B}_1)_{1,:} & 0 & 0 \\ 0 & \mathbf{B}_2 & \mathbf{A}_2 & 0 \\ 0 & 0 & \mathbf{A}_3 & \mathbf{B}_3 \\ \begin{pmatrix} 0 \\ 1 \end{pmatrix} & 0 & 0 & \mathbf{B}_4 \end{vmatrix} \\ &+ \begin{vmatrix} (\mathbf{A}_1)_{:,2} & \mathbf{B}_1 & 0 & 0 \\ 0 & \mathbf{B}_2 & \mathbf{A}_2 & 0 \\ 0 & 0 & \mathbf{A}_3 & \mathbf{B}_3 \\ 1 & 0 & 0 & (\mathbf{B}_4)_{2,:} \end{vmatrix} \end{aligned}$$



By further expanding with respect to the first column:

$$\begin{aligned}
|\mathbf{M}| &= x_1^2 \begin{vmatrix} \mathbf{B}_2 & \mathbf{A}_2 & 0 \\ 0 & \mathbf{A}_3 & \mathbf{B}_3 \\ 0 & 0 & \mathbf{B}_4 \end{vmatrix} + x_1 \begin{vmatrix} (\mathbf{B}_1)_{2,:} & 0 & 0 \\ \mathbf{B}_2 & \mathbf{A}_2 & 0 \\ 0 & \mathbf{A}_3 & \mathbf{B}_3 \\ 0 & 0 & (\mathbf{B}_4)_{1,:} \end{vmatrix} \\
&+ y_1^2 \begin{vmatrix} \mathbf{B}_2 & \mathbf{A}_2 & 0 \\ 0 & \mathbf{A}_3 & \mathbf{B}_3 \\ 0 & 0 & \mathbf{B}_4 \end{vmatrix} + y_1 \begin{vmatrix} (\mathbf{B}_1)_{1,:} & 0 & 0 \\ \mathbf{B}_2 & \mathbf{A}_2 & 0 \\ 0 & \mathbf{A}_3 & \mathbf{B}_3 \\ 0 & 0 & (\mathbf{B}_4)_{1,:} \end{vmatrix} \\
&+ y_1 \begin{vmatrix} (\mathbf{B}_1)_{2,:} & 0 & 0 \\ \mathbf{B}_2 & \mathbf{A}_2 & 0 \\ 0 & \mathbf{A}_3 & \mathbf{B}_3 \\ 0 & 0 & (\mathbf{B}_4)_{2,:} \end{vmatrix} - x_1 \begin{vmatrix} (\mathbf{B}_1)_{1,:} & 0 & 0 \\ \mathbf{B}_2 & \mathbf{A}_2 & 0 \\ 0 & \mathbf{A}_3 & \mathbf{B}_3 \\ 0 & 0 & (\mathbf{B}_4)_{2,:} \end{vmatrix} \\
&+ \begin{vmatrix} \mathbf{B}_1 & 0 & 0 \\ \mathbf{B}_2 & \mathbf{A}_2 & 0 \\ 0 & \mathbf{A}_3 & \mathbf{B}_3 \end{vmatrix}
\end{aligned}$$

By defining:

$$\mathbf{M}_{ij} := \begin{pmatrix} (\mathbf{B}_1)_{i,:} & 0 & 0 \\ \mathbf{B}_2 & \mathbf{A}_2 & 0 \\ 0 & \mathbf{A}_3 & \mathbf{B}_3 \\ 0 & 0 & (\mathbf{B}_4)_{j,:} \end{pmatrix}$$

We deduce:  $|\mathbf{M}| = y_1 (|\mathbf{M}_{11}| + |\mathbf{M}_{22}|) + x_1 (|\mathbf{M}_{21}| - |\mathbf{M}_{12}|) + 2 D^2$ .

By expanding  $|\mathbf{M}_{ij}|$  with respect to the first row, we deduce:

$$|\mathbf{M}_{ij}| = (\mathbf{B}_1)_{i,1} |\mathbf{L}_{2j}| - (\mathbf{B}_1)_{i,2} |\mathbf{L}_{1j}|,$$

where for  $i, j \in \{1, 2\}$ :

$$\mathbf{L}_{ij} := \begin{pmatrix} (\mathbf{B}_2)_{:,i} & \mathbf{A}_2 & 0 \\ 0 & \mathbf{A}_3 & \mathbf{B}_3 \\ 0 & 0 & (\mathbf{B}_4)_{j,:} \end{pmatrix}.$$

We then obtain:

$$\begin{aligned}
|\mathbf{M}| &= (D x_1^2 + y_1^2) |\mathbf{L}_{11}| + (x_1^2 + D y_1^2) |\mathbf{L}_{22}| \\
&+ x_1 y_1 (D - 1) (|\mathbf{L}_{12}| + |\mathbf{L}_{21}|) + 2 D^2.
\end{aligned} \tag{56}$$

We can expand with respect to the first column:

$$|\mathbf{L}_{ij}| = (\mathbf{B}_2)_{1,i} |\mathbf{K}_{2j}| - (\mathbf{B}_2)_{2,i} |\mathbf{K}_{1j}|,$$

where  $i, j \in \{1, 2\}$ :

$$\mathbf{K}_{ij} := \begin{pmatrix} (\mathbf{A}_2)_{i,:} & 0 \\ \mathbf{A}_3 & \mathbf{B}_3 \\ 0 & (\mathbf{B}_4)_{j,:} \end{pmatrix}.$$

We expand  $|\mathbf{K}_{ij}|$  with respect to the first row:

$$|\mathbf{K}_{ij}| = (\mathbf{A}_2)_{i,1} |\mathbf{J}_{2j}| - (\mathbf{A}_2)_{i,2} |\mathbf{J}_{1j}|,$$

where  $i, j \in \{1, 2\}$ :

$$\mathbf{J}_{ij} := \begin{pmatrix} (\mathbf{A}_3)_{:,i} & \mathbf{B}_3 \\ 0 & (\mathbf{B}_4)_{j,:} \end{pmatrix}.$$

To compute  $|\mathbf{J}_{ij}|$  and  $|\mathbf{K}_{ij}|$ , we recall:

$$\begin{aligned} y_{i+1} x_i - y_{i+1} x_i &= y_1, \\ x_{i+1} x_i + y_{i+1} y_i &= x_1. \end{aligned}$$

We obtain:

$$\begin{aligned} |\mathbf{J}_{11}| &= -D x_1 x_3 + y_1 y_3, & |\mathbf{K}_{11}| &= -(D+1) x_1 y_1, \\ |\mathbf{J}_{12}| &= D^2 y_1 x_3 + D x_1 y_3, & |\mathbf{K}_{12}| &= D^2 y_1^2 - D x_1^2, \\ &\text{and} & \\ |\mathbf{J}_{21}| &= -D x_1 y_3 - y_1 x_3, & |\mathbf{K}_{21}| &= D x_1^2 - y_1^2, \\ |\mathbf{J}_{22}| &= D^2 y_1 y_3 - D x_1 x_3, & |\mathbf{K}_{22}| &= -D(D+1) x_1 y_1. \end{aligned}$$

To compute  $|\mathbf{L}_{ij}|$ , we recall:

$$\begin{aligned} y_2 x_1 + y_1 x_2 &= y_3, \\ x_2 x_1 - y_2 y_1 &= x_3. \end{aligned}$$

We then obtain the following values for  $|\mathbf{L}_{ij}|$ :

$$\begin{aligned} |\mathbf{L}_{11}| &= D^2 x_1 y_1 y_2 - D x_1 x_3 + x_2 y_1^2, \\ |\mathbf{L}_{12}| &= -D^3 y_1^2 y_2 + D^2 x_1 y_3 + D x_2 x_1 y_1, \\ |\mathbf{L}_{21}| &= -D^2 x_1 x_2 y_1 - D x_1 y_3 + y_2 y_1^2, \\ |\mathbf{L}_{22}| &= D^3 y_1^2 x_2 - D^2 x_1 x_3 + D y_2 x_1 y_1. \end{aligned}$$

Using Equation (56), we obtain:

$$\begin{aligned} |\mathbf{M}| &= -D^4 y_1^4 + D^3 (2 x_1 y_1)^2 \\ &\quad + 2 D^2 y_1^2 (y_1^2 + 4 x_1^2) \\ &\quad + D (2 x_1 y_1)^2 - y_1^4. \end{aligned}$$

We note that  $-1$  is an obvious root.

Moreover, if  $D_r$  is a root, then  $1/D_r$  is also a root. We deduce that  $-1$  is a double root. By defining  $z_1 := \frac{x_1}{y_1}$ . The two other roots are:

$$D_{\pm} = (2 z_1^2 + 1) \pm 2 z_1 \sqrt{z_1^2 + 1},$$

Finally:  $|\mathbf{M}| = -y_1^2 (D+1)^2 (D-D_+) (D-D_-)$ .

## Glossary

$\mathbb{P}_d^k(\mathcal{T}_h)$  Set of broken polynomials on  $\mathcal{T}_h$ .

$C_P$  Poincaré constant.

$F$  Face of an element.

$H, V$  Hilbert spaces.

$H'$  Topological dual space of  $H$ .

$T$  Mesh element.

$V_h$  Space of discretized functions on  $\mathcal{T}_h$ .

$C_{\text{app}}$  Constant of  $L^2$ -orthogonal projection approximation.

$C_{\text{inv}}$  Inverse inequality constant.

$C_{\text{tr}}$  Discrete trace inequality constant.

$\Omega$  Open subset of  $\mathbb{R}^d$ .

$\{\!\!\{v\}\!\!\}_F$  Component-wise average of  $v$  at face  $F$ .

$\mathbf{H}^s(\text{div}; \mathcal{T}_h)$  Broken divergence-conforming Sobolev space.

$\mathbf{n}_F$  Unit normal vector of the face.

$\mathbf{n}_{T,F}$  Unit normal vector of the face, directed from the interior to the exterior of the element.

$\mathbf{n}_{\partial\Omega}$  Unit outward normal vector of the surface, directed outward from the domain.

$\ell(\cdot)$  Linear form.

$\ell_h(\cdot)$  Discrete linear form.

$\text{Tr}_\gamma u$  Trace of  $u$  on the boundary  $\gamma$ .

$\epsilon_h$  Error between the real solution and the discrete solution.

$\gamma$  Interior penalty parameter.

$\llbracket v \rrbracket$  Component-wise jump of  $v$  at face  $F$ .

$\mathbb{P}_d^k(\Omega)$  Polynomials of total degree  $k$  in  $\Omega \subset \mathbb{R}^d$ .

$\mathbf{H}(\text{div}; \Omega)$  Vector-valued Sobolev space.

$\mathbf{L}^2(\Omega)$  Set of square-integrable vector-valued functions from  $\Omega$  to  $\mathbb{R}^d$ .

$\mathbf{t}_F$  Tangent at a point of  $F$ .

$\mathcal{F}_T$  Set of faces of  $T$ .

$\mathcal{F}_h$  Set of faces of the mesh.

$\mathcal{F}_h^b$  Set of mesh faces included in  $\partial\Omega$ .

$\mathcal{F}_h^i$  Set of mesh faces included in  $\Omega$ .

$\mathcal{T}_F$  Set of triangles that admit  $F$  as a face.

$\mathcal{T}_h$  Set of elements of the triangulation.

$\nabla_h v$  Broken gradient of  $v$ .

$\nabla u$  Gradient (Cartesian) of  $u$ .

$\omega$  Weighting for the average in the SWIP method.

$\overline{\Omega}$  Closure of  $\Omega$ .

$\partial\Omega$  Boundary of  $\Omega$ .

$\partial^k u$  If  $u : \Omega \rightarrow \mathbb{R}^d$ ,  $k = (k_1, \dots, k_d)$ , then  $\partial_x^k u = \partial_{x_1}^{k_1} \dots \partial_{x_d}^{k_d} u$ .

$\sigma$  Mesh regularity index.

$H_{g_D}^1(\Omega)$   $u \in H^1(\Omega)$  such that  $\text{Tr}_{\partial\Omega}(u) \equiv g_D$ .

$H_0^p(\Omega)$   $u \in H^p(\Omega)$  such that  $\text{Tr}_{\partial\Omega}(\partial_x^k u) \equiv 0$  for all  $0 \leq |k| \leq p-1$ .

$H^s(\Omega)$  Sobolev space.

$H^s(\mathcal{T}_h)$  Broken Sobolev space.

$L^2(\Omega)$  Set of square-integrable functions from  $\Omega$  to  $\mathbb{R}$ .

$L^\infty(\Omega)$  Set of essentially bounded functions from  $\Omega$  to  $\mathbb{R}$ .

$\text{div}$  Divergence of  $u$ .

$\text{div}_h v$  Broken divergence of  $v$ .

$\varrho$  Regularity of the mesh.

$a(\cdot, \cdot)$  Bilinear form.

$a_h(\cdot, \cdot)$  Discrete bilinear form.

$a_i$  Vertex  $i$  of an element.

$f$  Function, often the source term of the problem.

$h$  Diameter of the mesh.

$h_F$  Diameter of  $F$ .

$h_T$  Diameter of  $T$ .

$h_{T,F}$  Height of a triangle from the vertex opposite to face  $F$ .

$r_T$  Radius of the inscribed circle in  $T$ .

$u_h$  Discrete solution.

$(\mathcal{C}_T^\star)^2$   $(\mathcal{C}_T^\star)^2 := \frac{\kappa_T}{20\text{Tr}(\mathbf{S}_T)}$ , where  $\mathbf{S}_T$  is the local stiffness matrix of  $T$ .

$T^\star$   $T^\star = \{T' / \mathcal{A}_h(T) \cap \mathcal{A}_h(T') \neq \emptyset\}$  is the set of elements  $T' \in \mathcal{T}_h$  that share a vertex with  $T$ .

$\Delta u$  (Cartesian) Laplacian of  $u$ .

$\Gamma_N, \Gamma_D$  Neumann and Dirichlet boundaries.

$\Omega_i$  Set of elements  $T$  that share a common vertex  $\mathbf{x}_i$ .

$|u|_J$  Norm of the jumps of  $u$ .

$\mathbf{x}_G$  Barycenter of an element.

$\boldsymbol{\sigma}(u)$  Flux of  $u$ ,  $\boldsymbol{\sigma}(u) = \kappa \nabla u$ .

$\boldsymbol{\sigma}_h(u)$  Broken flux of  $u$ ,  $\boldsymbol{\sigma}_h(u) = \kappa_h \nabla u$ .

$\epsilon_{T,\text{conforming}}$  Conforming error on  $T$ .

$\epsilon_{T,\text{non-conforming}}$  Non-conforming error on  $T$ .

$\epsilon_T$  Error in  $T$ .

$\eta_T$   $\eta_T^2 = \eta_{\text{NC},T}^2 + \eta_{\text{CF},T}^2$  is the a posteriori error estimator on  $T$ .

$\eta_h$   $\eta_h$  is the a posteriori error estimator.

$\eta_{\text{CF},T}$  Conforming error estimator.

$\eta_{\text{NC},T}$  Non-conforming error estimator.

$\kappa$  Diffusion coefficient.

$\lambda_i$  Barycentric coordinate in an element.

$\mathbb{P}_c^1(\mathcal{F}_h)$  Piecewise continuous affine functions on  $\mathcal{F}_h$ .

$\mathbf{S}_F$  Non-normalized normal vector of face  $F$ .

$C^k(\Omega)$  Functions  $k$ -times differentiable with continuous  $k$ -th derivative on  $\Omega$ .

$\mathcal{F}_h^N, \mathcal{F}_h^D$  Set of Neumann and Dirichlet faces.

$\bar{f}_T$  Average of  $f$  on an element,  $\bar{f}_T := \frac{1}{|T|} \int_T f$ .

$\bar{g}_F$  Average of  $g$  on a face,  $\bar{g}_F := \frac{1}{|F|} \int_F g$ .

$\phi_T$  Equilibrated numerical flux.

$\mathbf{L}_F$   $\mathbf{L}_F := \max_{\mathbf{x} \in F} \|\mathbf{x} - \mathbf{x}_F\|_{\mathbb{R}^2}$ .

$\text{Osc}(f, T)$   $\text{Osc}(f, T)^2 = |T| \|f - \bar{f}\|_{L^2(T)}^2$  is the oscillation of the data  $f$  on  $T$ .

$\text{Osc}(g, F)$   $\text{Osc}(g, F)^2 = |F| \|g_N - \bar{g}_N\|_{L^2(F)}^2$  is the oscillation of the data  $g$  on  $F$ .

$\mathbf{l}_F$   $\mathbf{l}_F := \min_{\mathbf{x} \in F} \|\mathbf{x} - \mathbf{x}_F\|_{\mathbb{R}^2}$ .

$\text{curl}(u)$  Curl of  $u$ .

$\tilde{T}$   $\tilde{T} := \{T' / \mathcal{F}_T \cap \mathcal{F}_{T'} \neq \emptyset\}$  is the set of neighboring elements of  $T$ .

$e_h$   $e_h := u - u_h$  is the error in  $V_h$ .

$g = (g_N, g_D)$  Dirichlet or Neumann boundary data.

$u_h^\star$  Ostwald projection of  $u_h$ .

## List of Figures

2	Local map of the a posteriori error estimate $S_h$ (left) and the exact local error $\epsilon_h$ (right) for the problem $-\Delta u = \pi^2 \sin(\pi x) \sin(\pi y)$ with homogeneous Dirichlet conditions, $h = 0.1, v_h \in \mathbb{P}^1$ .	8
3	Example of an L-shaped mesh locally refined near the re-entrant corner.	9
4	Example of a polyhedral domain and its mesh.	11
5	Simplicial discretization of a pentagon in 2D (triangulation).	11
6	In blue, a face $F$ shared by two elements $T_1$ and $T_2$ , $\mathbf{n}_F$ its unit normal vector and $\mathbf{t}_F$ its tangential vector.	12
7	Orientation of normal and tangent unit vectors depending on the chosen element.	12
8	In blue, the face $F_3$ opposite to vertex $a_3$ (also in blue), $\mathbf{n}_{F_3}$ its unit normal vector, and $\mathbf{t}_{F_3}$ its unit tangential vector, with $T_1$ and $T_2$ being the elements sharing $F_3$ .	12
9	Geometric quantities for simplices of dimension $d = 2$ .	13
10	Basis and approximation of the continuous finite elements $\mathbb{P}_1^1(\mathcal{T}_h) \cap \mathcal{C}^0(\Omega)$ .	13
11	Some basis functions of $\mathbb{P}_1^1(\mathcal{T}_h)$ en 1D.	14
12	Jump and average of a discontinuous function.	14
13	Comparison between the gradient and the broken gradient.	15
14	Example of partitioned domain according to $\kappa$ and a possible mesh	22
15	Solution of the spectral problem on $[0; 1]^2$ for $\lambda = 1, 2$ .	33
16	Local map of a posteriori error estimator $\eta_T$ (left) and broken flux error $\epsilon_T$ (right) for $\lambda = 2, h = 0.05$ and $v_h \in \mathbb{P}^1$ .	34
17	Left: nonconforming error estimator $\eta_{NC,T}$ , right: conforming error estimator $\eta_{CF,T}$ for $\lambda = 2, h = 0.05$ and $v_h \in \mathbb{P}^1$ .	35
18	Comparison of the global estimated error and true broken flux error (left), comparison of the error estimation with and without bubble function correction (right).	35
19	Comparison of the different components of the error estimator and the error in log-log scale for $\lambda = 2$ and $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom.	36
20	A posteriori error and broken flux error in the case of global refinement (all cells have equal size) and in the case of local refinement (mesh refined in regions where the error is large). Log-log scale. Abscissa: number of degrees of freedom.	36
21	Comparison of the different components of the error estimator and the actual error when $\gamma_F$ varies in $[5, 10, 50, 100]$ for $\lambda = 2$ and $v_h \in \mathbb{P}^1$ , log-log scale. Abscissa: number of degrees of freedom.	37
22	A posteriori error estimator and broken flux error in log-log scale. Abscissa: number of degrees of freedom. $\lambda = 2$ and $v_h \in \mathbb{P}^1$ .	38
23	Comparison of the different components of the error estimator and the error in log-log scale for $\lambda = 2$ and $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom.	38
24	Nonconvex L-shaped domain $\Omega$ and its associated polar coordinates.	39
25	Solution $f(r, \theta) = r^{2/3} \sin(\frac{2}{3}\theta)$ .	39
26	Local map of the a posteriori error estimation $\eta_T$ (left) and broken flux error $\epsilon_T$ (right) for $h = 0.05$ and $v_h \in \mathbb{P}^1$ .	40
27	Comparison of the different components of the error estimator and the true error in log-log scale for the L-shaped domain and $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom.	41
28	A posteriori error and broken flux error in the case of global refinement (all cells are of equivalent size) and in the case of local refinement (mesh refined in zones where the error is large). Log-log scale. Abscissa: number of degrees of freedom.	41
29	Comparison of the different components of the error estimator and the error in log-log scale for $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom.	42
30	A posteriori error and broken flux error for global and local refinement. Log-log scale. Abscissa: number of degrees of freedom.	42

31	Domain $\Omega$ considered and its polar coordinates. . . . .	43
32	Solution $f(r, \theta) = r^{1/4}(a \cos(\theta/4) + b \sin(\theta/4))$ for $\alpha = 1/4$ and $D = 0.0395661$ . . . . .	44
33	Comparison of the different components of the error estimator and the error in log-log scale for the L-shaped domain and $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom. . . . .	44
34	Comparison of the different components of the error estimator and the error in log-log scale for $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom. . . . .	45
35	Comparison of the different components of the error estimator and the error in log-log scale for the L-shaped domain and $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom. . . . .	45
36	Comparison of the different components of the error estimator and the error in log-log scale for $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom. . . . .	46
37	Comparison of the different components of the error estimator and the error in log-log scale for the L-shaped domain and $v_h \in \mathbb{P}^1$ . Abscissa: number of degrees of freedom. . . . .	46
38	Domain $\Omega$ and its polar coordinates. . . . .	55

## References

- [AIN05] M. AINSOWRTH. “Robust A Posteriori Error Estimation for Nonconforming Finite Element Approximation”. In: (2005). DOI: [10.1137/S0036142903425112](https://doi.org/10.1137/S0036142903425112).
- [AIN07] M. AINSWORTH. “A Posteriori Error Estimation for Discontinuous Galerkin Finite Element Approximation”. In: (2007). DOI: [10.1137/060665993](https://doi.org/10.1137/060665993).
- [ABJ25] P.-E. ANGELI, H. BERTRAND, and E. JAMELOT. “Stability and convergence of the  $P_{nc}^1 \times (P^0 + P^1)$  discretization  $\star$ ”. 2025. URL: <https://hal.archives-ouvertes.fr/hal-05071391>.
- [BHL03] R. BECKER, P. HANSBO, and M. G. LARSON. “Energy norm a posteriori error estimation for discontinuous Galerkin methods”. In: (2003). DOI: [https://doi.org/10.1016/S0045-7825\(02\)00593-5](https://doi.org/10.1016/S0045-7825(02)00593-5).
- [BMR04] C BERNADI, Y. MADAY, and F. RAPETTI. *Discrétisations variationnelles de problèmes aux limites elliptiques*. 2004. URL: <https://hal.archives-ouvertes.fr/hal-00020651>.
- [BRE11] H. BREZIS. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. 2011. DOI: [10.1007/978-0-387-70914-7](https://doi.org/10.1007/978-0-387-70914-7).
- [CYZ11] Zhiqiang Cai, Xiu Ye, and Shun Zhang. “Discontinuous Galerkin Finite Element Methods for Interface Problems: A Priori and A Posteriori Error Estimations”. In: (2011). DOI: [10.1137/100805133](https://doi.org/10.1137/100805133).
- [CGJ09] C. CARSTENSEN, T. GUDI, and M. JENSEN. “A unifying theory of a posteriori error control for discontinuous Galerkin FEM”. In: (2009). DOI: [10.1007/s00211-009-0223-9](https://doi.org/10.1007/s00211-009-0223-9).
- [CR73] M. CROUZEIX and P.-A. RAVIART. “Conforming and nonconforming finite element methods for solving the stationary Stokes equations I”. In: (1973). DOI: [10.1051/m2an/197307R300331](https://doi.org/10.1051/m2an/197307R300331).
- [D11] FOURNIER D. “Analyse et développement de méthodes de raffinement hp en espace pour l’équation de transport des neutrons”. Thèse de doctorat dirigée par R. RAPHAËLE. PhD thesis. 2011. URL: <http://www.theses.fr/2011AIX10085/document>.
- [DAU88] M. DAUGE. *Elliptic Boundary Value Problems on Corner Domains*. 1988. DOI: <https://doi.org/10.1007/BFb0086682>.
- [DE12] D. Di PIETRO and A. ERN. *Mathematical Aspects of Discontinuous Galerkin Methods*. 2012. DOI: [10.1007/978-3-642-22980-0](https://doi.org/10.1007/978-3-642-22980-0).
- [E D+96] E., DARI et al. “A posteriori error estimators for nonconforming finite element methods”. In: (1996). DOI: [10.1051/m2an/1996300403851](https://doi.org/10.1051/m2an/1996300403851).
- [EG21] A. ERN and J.-L. GUERMOND. *Finite Elements I: Approximation and Interpolation*. 2021. DOI: [10.1007/978-3-030-56341-7](https://doi.org/10.1007/978-3-030-56341-7).
- [ESV10] A. ERN, A. F. STEPHANSEN, and M. VOHRALÍK. “Guaranteed and robust discontinuous Galerkin a posteriori error estimates for convection–diffusion–reaction problems”. In: (2010). DOI: [10.1016/j.cam.2009.12.009](https://doi.org/10.1016/j.cam.2009.12.009).
- [EV09] A. ERN and M. VOHRALÍK. “Flux reconstruction and a posteriori error estimation for discontinuous Galerkin methods on general nonmatching grids”. In: (2009). DOI: <https://doi.org/10.1016/j.crma.2009.01.017>.
- [FAD22] J. FADDOUL. “Estimations d’erreur a posteriori pour le couplage des équations de Navier-Stokes avec l’équation de convection–diffusion–réaction”. Theses. 2022. URL: <https://hal.archives-ouvertes.fr/tel-04066549>.
- [GRI11] P. GRISVARD. *Elliptic Problems in Nonsmooth Domains*. SIAM, 2011. DOI: [10.1137/1.9781611972030](https://doi.org/10.1137/1.9781611972030).
- [HIR12] M. W. HIRSCH. *Differential Topology*. 2012. DOI: [10.1007/978-1-4684-9449-5](https://doi.org/10.1007/978-1-4684-9449-5).



- [KP03] O. A. KARAKASHIAN and F. PASCAL. “A Posteriori Error Estimates for a Discontinuous Galerkin Approximation of Second-Order Elliptic Problems”. In: (2003). DOI: [10.1137/S0036142902405217](https://doi.org/10.1137/S0036142902405217).
- [LEC25] R. LECOQ. *Code*. 2025. URL: <https://github.com/RaphaelLcq/CEA-DGM-A-posteriori->.
- [OSW94] P. OSWALD. *Multilevel Finite Element Approximation*. 1994. DOI: [10.1007/978-3-322-91215-2](https://doi.org/10.1007/978-3-322-91215-2).
- [PW60] L.E. PAYNE and H. F. WEINBERGER. “An optimal Poincaré inequality for convex domains”. In: (1960). DOI: [10.1007/BF00252910](https://doi.org/10.1007/BF00252910).
- [PE11] D.A. DI PIETRO and A. ERN. “Analysis of a discontinuous Galerkin method for heterogeneous diffusion problems with low-regularity solutions”. In: (2011). DOI: [10.1002/num.20675](https://doi.org/10.1002/num.20675).
- [TAR07] L. TARTAR. *An Introduction to Sobolev Spaces and Interpolation Spaces*. 2007. DOI: [10.1007/978-1-4684-9449-5](https://doi.org/10.1007/978-1-4684-9449-5).
- [VOH24] M. VOHRALÍK. *A posteriori numerical analysis based on the method of equilibrated fluxes course*. 2024.
- [WH03] T. Warburton and J.S. Hesthaven. “On the constants in hp-finite element trace inverse inequalities”. In: *Computer Methods in Applied Mechanics and Engineering* (2003). DOI: [10.1016/S0045-7825\(03\)00294-9](https://doi.org/10.1016/S0045-7825(03)00294-9).