Commissariat à l'énergie atomique et aux énergies alternatives

Study of a finite volume scheme with nodal fluxes for parabolic problems on unstructured meshes

Internship report
april - september 2022

Axelle DROUARD
Nantes Université
UFR Sciences et techniques
Master 2 MACS
September 2022

Tutors: Christophe BUET
Emmanuel LABOURASSE

Acknowledgements

I would like to express my gratitude to all the people that helped me during this internship and in the writing of this report.

I especially wish to thank the CEA research engineers Mr. BUET Christophe and Mr. LABOURASSE Emmanuel, who were my tutors for the internship. I learned many things in the mathematical and computational field thanks to them. They offered me their knowledge and precious advises throughout this period, always with patience and pedagogy. I would also like to thank Mr. DEL PINO Stéphane for making himself available whenever I had a question even though he was not one of my supervisors.

Then, I would like to thank the educational team of the Sciences and Techniques pole of Nantes University for the work environment they provided. Finally, I especially thank my Master's degree professors who brought me the necessary tools for this internship, and above all the will to continue along this path.

Contents

1 Introduction

In this document we will study a finite volume scheme with nodal fluxes for the following diffusion equation:

$$\begin{cases} \partial_t p(\mathbf{x},t) - \nabla.(\kappa(\mathbf{x})\mathbf{u}(\mathbf{x},t)) = f(\mathbf{x},t) & \forall\ \mathbf{x}\in\Omega,\ t\in[0,T] \\ p(\mathbf{x},0) = p_0(\mathbf{x}) & \forall\ \mathbf{x}\in\Omega \\ p(\mathbf{x},t) = h(\mathbf{x},t) & \forall\ \mathbf{x}\in\Gamma_D,\ t\in[0,T] \\ \langle\kappa(\mathbf{x})\mathbf{u}(\mathbf{x},t),\mathbf{n}_{\Gamma_N}\rangle = g(\mathbf{x},t) & \forall\ \mathbf{x}\in\Gamma_N,\ t\in[0,T] \end{cases} \tag{1}$$

where

- $\Omega \subset \mathbb{R}^d$, with d the space dimension;
- $p : \Omega \times [0,T] \to \mathbb{R}$ is a scalar unknown;
- $\mathbf{u} = \nabla p : \Omega \times [0,T] \to \mathbb{R}^d$ is a vectorial unknown;
- $f : \Omega \times [0,T] \to \mathbb{R}$ is the source function;
- $p_0 : \Omega \to \mathbb{R}$ is a given function that is the initial condition of the problem;
- $\kappa : \Omega \to \mathbb{R}^{d\times d}$ is the diffusion coefficient.
  We are working in an anisotropic medium, which means the diffusion coefficient is a symmetric and positive tensor that depends on space. Another property of κ is that it is bounded: there exists $M > 0$ a constant such that $|||\kappa|||_{\Omega,\infty} = \max\limits_{\mathbf{x}\in\Omega}(|||\kappa(\mathbf{x})|||_\infty) < M$, where $|||A|||_\infty = \max\limits_{1\leq i\leq m}\sum_{j=1}^n |a_{ij}|$, for $A \in \mathbb{R}^{m\times n}$.
- $\mathbf{n}_{\Gamma_N}$ is the normal to the boundary $\Gamma = \partial\Omega$, directed torwards the outside of the domain;
- $g : \Gamma_N \times [0,T] \to \mathbb{R}$. We will consider $g = 0$ for homogeneous Neumann boundary conditions;
- $h : \Gamma_D \times [0,T] \to \mathbb{R}$. For homogeneous Dirichlet boundary conditions one has $h = 0$.

Proposition 1 (Uniqueness of the solution)**.** *Let Ω be a bounded domain of $\mathbb{R}^2$. For convenience, let us consider $\Omega \in C^\infty$ as in [4]. Let $f \in L^2([0,T], L^2(\Omega))$, $p_0 \in H^1(\Omega)$, $\kappa \in L^\infty(\Omega)$, $g, h \in H^{1/2}([0,T], L^2(\Gamma))$, then the system (1) has a unique weak solution $p \in C^0([0,T]; H^1(\Omega)) \cap C^1([0,T], L^2(\Omega))$. The proof is obtained with Lions' Theorem from [4].*

Proposition 2 (Conservation.)**.** *The system (1) with no source ($f = 0$) and periodical or Neumann homogeneous boundary conditions is conservative, which means one has:*

$$\partial_t \int_\Omega p\ d\mathbf{x} = 0 \tag{2}$$

Proof can be found in Appendix B.1.

Proposition 3 (L^2 stability)**.** *The system (1) with no source and homogeneous boundary conditions is stable in L^2 norm, which means:*

$$\partial_t \|p\|_{L_2} \leq 0 \tag{3}$$

Proof can be found in Appendix B.2.

The choice of a finite volume scheme was made to ensure that the scheme is natively locally conservative which would not necessarily be the case with a finite element method. Moreover the scheme has to be consistent on unstructured meshes in order to be coupled with Lagrangian hydrodynamics codes which can produce very distorted meshes.

The origin of the problem comes from the hyperbolic heat equation (4), or P_1 equation, which is an approximation of a transport equation.

$$\begin{cases} \partial_t p + \dfrac{1}{\varepsilon}\nabla.\mathbf{u} = 0 \\ \partial_t \mathbf{u} + \dfrac{1}{\varepsilon}\nabla p + \dfrac{\sigma}{\varepsilon^2}\mathbf{u} = \mathbf{0} \end{cases} \tag{4}$$

This model can be approximated by the following diffusion equation when ε tends to 0, thanks to a simple Hilbert's expansion:

$$\partial_t p - \nabla.\Big(\frac{1}{\sigma}\nabla p\Big) = 0 \tag{5}$$

In this work we will study a numerical scheme for the system (1). This scheme has been obtained in Emmanuel Franck's thesis [13] as the asymptotic limit of a scheme for the hyperbolic heat system.

It leads to the choice of nodal fluxes, this way the scheme remains consistent on unstructured meshes in its diffusion limit: it is asymptotic preserving (AP). Indeed with classical face fluxes the diffusion limit of the P_1 scheme would have given a TPFA scheme [12] which is not consistent on distorted meshes. The nodal scheme presents other advantages compared to other existing schemes with face fluxes: unlike Diamond, Mimetic and MPFA [7, 16, 1] the nodal scheme does not need any auxiliary unknowns. Moreover there are no dual diffusion problem to solve with the nodal scheme, as opposed to DDFV [14].

In E. Franck's thesis and in [5] the limit diffusion scheme is built and studied in its semi-discrete version, with periodical boundary conditions and a constant scalar diffusion coefficient which is insufficient for applications. The objective of this document is to study more in depth the nodal diffusion scheme by adding to it a time discretization, boundary conditions and a tensorial diffusion coefficient in order to simulate an anisotropic domain.

The organization of this work is as follows: in Section 2 one can find the definitions, notations and properties of the mesh that are needed in the rest of the document. In Section 3 we built the scheme with Neumann, Dirichlet and mixed boundary conditions, added an Euler implicit time discretization to it and gave the conditions on the mesh for the invertibility of β_r. In Section 4 properties of the scheme are given: uniqueness, exactness, stability, consistency and convergence, for every boundary conditions and Euler implicit method. In Section 5 we extended the time discretization to a second-order in time Crank-Nicolson method, and proved the properties of stability, consistency and convergence in that case. This nodal scheme (in Euler implicit and Crank-Nicolson version) gave us some surprising results depending on the mesh as one can see on the numerical results in Section 7. In order to fix the problems encountered with the nodal scheme, in Section 6 we built a hybrid face-node scheme that is a correction of the purely nodal scheme and does not require any auxiliary unknown either. Finally, all the numerical results are given in Section 7 for linear, quadratic, trigonometric, non-stationary solutions and with a non-constant coefficient diffusion on several meshes of quad and triangle and for the Euler implicit and Crank-Nicolson versions of the nodal and hybrid scheme, with Neumann and Dirichlet boundary conditions.

2 Notations and properties

All over the document, some definitions, notations and general properties will be used that are listed and proved in this section.

2.1 Definitions

Definition 1. *Let us define the scalar product of two vectors* $\mathbf{u} = (u_1, \dots, u_d)$ *and* $\mathbf{v} = (v_1, \dots, v_d)$ *in dimension* d*, noted* $\langle \mathbf{u}, \mathbf{v} \rangle$ *as the sum of the product of their coefficients:*

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{d} u_i v_i \tag{6}$$

Definition 2. *Let us define the classical norm of a vector* $\mathbf{u}$ *associated to the classical scalar product of Definition 1 as*

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle} = \sqrt{\sum_{i=1}^{d} u_i^2} \tag{7}$$

Definition 3. *Let* $\mathbf{u} = (u_1, \ldots, u_d)$ *be any vector in* $\mathbb{R}^d$*.*

Then

$$\langle \mathbf{u}, \mathbf{u}^{\perp} \rangle = 0 \quad \text{and} \quad \mathbf{u} \wedge \mathbf{u}^{\perp} = \|\mathbf{u}\|^2 \tag{8}$$

In particular, if $d = 2$*, we chose*

$$\mathbf{u}^{\perp} = (-u_2, u_1) \tag{9}$$

Then

$$\langle \mathbf{u}, \mathbf{u}^{\perp} \rangle = -u_1 u_2 + u_2 u_1 = 0 \quad \text{and} \quad \mathbf{u} \wedge \mathbf{u}^{\perp} = u_1^2 + u_2^2 = \|\mathbf{u}\|^2$$

Definition 4. *Let us introduce* $A \in \mathbb{R}^{d \times d}$*, and* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ *such that* $A = \mathbf{u} \otimes \mathbf{v}$*. Then*

$$A_{i,j} = u_i v_j, \quad \forall\ i, j = 1, \ldots, d \tag{10}$$

Definition 5. *Let us introduce* $A \in \mathbb{R}^{d \times d}$*. Then its transpose matrix* A^t *is defined by*

$$A^t_{i,j} = A_{j,i} \quad \forall\ i, j = 1, \ldots d \tag{11}$$

Definition 6. *Let us introduce* $A \in \mathbb{R}^{d \times d}$*. Then its symmetric part* A^s *is defined by*

$$A^s = \frac{A + A^t}{2} \tag{12}$$

Definition 7. *Let us introduce* $A \in \mathbb{R}^{2 \times 2}$*. The determinant of the matrix* A *is given by*

$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{13}$$

Definition 8. *Let us introduce* $A \in \mathbb{R}^{d \times d}$*. The trace of the matrix* A *is given by*

$$\mathrm{tr}(A) = \sum_{i=1}^{d} A_{i,i} \tag{14}$$

2.2 Construction of the mesh

2.2.1 Notations

Let us define:

- Ω_j a cell of the mesh;
- $\mathbf{x}_j$ the center of a cell Ω_j;

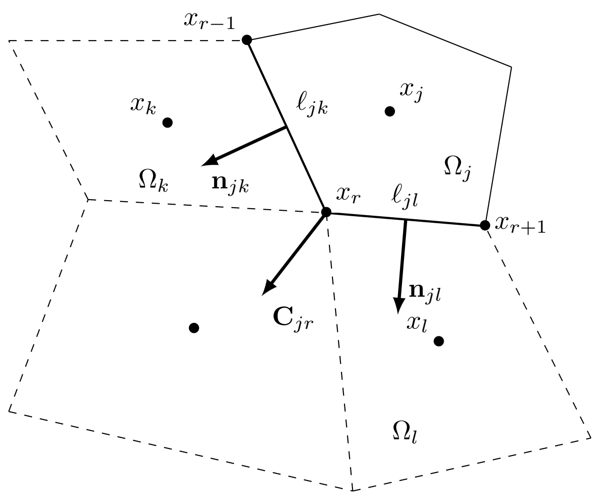


Figure 1: Unstructured mesh on Ω

- $\mathbf{x}_r$ a vertex of the mesh;
- $\mathcal{C}$ the set of cells of the mesh;
- $\mathcal{V}$ the set of vertices of the mesh;
- $\ell_{jk} = |\partial\Omega_{jk}|$, the length of $\Omega_j \cap \Omega_k$;
- $\mathbf{n}_{jk}$ the normal exterior to Ω_j.

Remark 1. We can add one or two indices to the previous sets in order to restrain them:

- $\mathcal{C}_r$ is the set of cells having $\mathbf{x}_r$ for common vertex;
- $\mathcal{C}_j$ is the set of cells that share a common face or node with Ω_j;
- $\mathcal{L}_j$ is the set of cells at the edge of Ω_j;
- $\mathcal{V}_j$ is the set of vertices of the cell Ω_j;
- $\mathcal{C}_{jr} = \mathcal{L}_j \cap \mathcal{C}_r$;
- $\mathcal{V}_{jk} = \mathcal{V}_j \cap \mathcal{V}_k$;
- $\mathcal{V}_j^i = \mathcal{V}_j \backslash \Gamma$ is the set of vertices of Ω_j inside the domain (not on the boundary);
- $\mathcal{V}_j^b = \mathcal{V}_j \cap \Gamma$ is the set of vertices of Ω_j on the boundary of the domain, except for the corners;
- $\mathcal{V}_j^c$ is the set of "corner nodes", that is the set of nodes on the boundary belonging to a unique cell as one can see in Figure 2;
- $\mathcal{V}_j^{i,b} = \mathcal{V}_j^b \cup \mathcal{V}_j^i$.

Definition 9. *We call corner a node belonging to a single cell (ie. $|\mathcal{C}_r| = 1$) so that two edges sharing this node belong to the boundary. One can see an example in Figure 2.*

Definition 10. *We call angle a node belonging to a several cells (ie. $|\mathcal{C}_r| > 1$) so that two of the edges sharing this node belong to the boundary. One can see an example in Figure 3.*

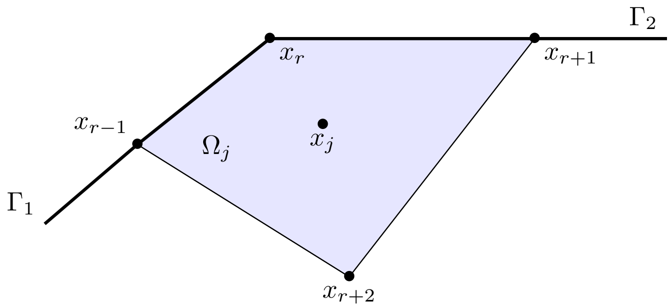


Figure 2: Construction of a corner $\mathbf{x}_r \in \mathcal{V}_j^c$

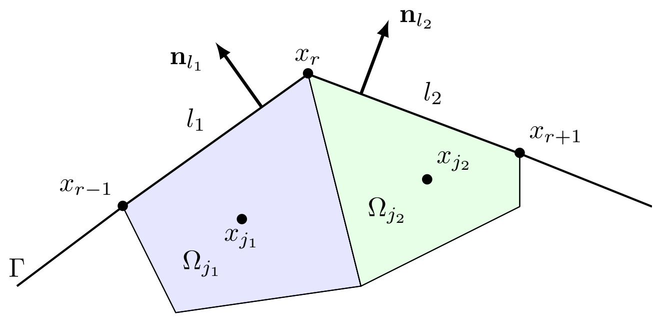


Figure 3: Construction of an angle $\mathbf{x}_r$

Remark 2 (Notation for the sums). For the whole document, for any q:

- $\sum_{j\in\mathcal{C}} q_j$ is the sum of q_j over all the cells of the mesh;
- $\sum_{r\in\mathcal{V}} q_r$ is the sum of q_r over all the vertices of the mesh;
- $\sum_{r\in\mathcal{V}_j} q_{jr}$ is the sum of q_{jr} over all the vertices of the cell Ω_j;
- $\sum_{j\in\mathcal{C}_r} q_{jr}$ is the sum of q_{jr} over all the cells that have $\mathbf{x}_r$ for common vertex.
- Note that for all q_{jr}:

$$\sum_{j\in\mathbf{C}}\sum_{r\in\mathcal{V}_j} q_{jr} = \sum_{r\in\mathcal{V}}\sum_{j\in\mathcal{C}_r} q_{jr} \tag{15}$$

Definition 11 (Control volumes V_r and V_{jr})**.** *The control volume is defined by the closed loop:* $\ldots, \mathbf{x}_{j-\frac{1}{2}}, \mathbf{x}_j, \mathbf{x}_{j+\frac{1}{2}}, \ldots$. *Where* $\mathbf{x}_{j+\frac{1}{2}}$ *are the center of the edges around* $\mathbf{x}_r$*. We can see the control volume* V_r *in Figure 4.*

One can divide V_r *into smaller areas delimited by the points* $\mathbf{x}_r$*,* $\mathbf{x}_{j+\frac{1}{2}}$*,* $\mathbf{x}_j$*,* $\mathbf{x}_{j-\frac{1}{2}}$*. These areas are named* V_{jr} *and they verify:* $\bigcup_{j\in\mathcal{C}_r} V_{jr} = V_r$*. They are represented in Figure 4*

2.2.2 Properties of the mesh

Proposition 4. *One has*

$$|\Omega_j|\widehat{I_d} = \sum_{r\in\mathcal{V}_j} \mathbf{x}_r \otimes \mathbf{C}_{jr} \tag{16}$$

where

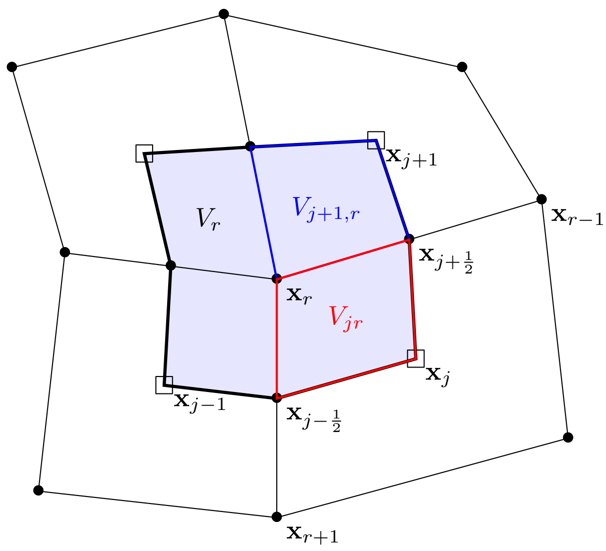


Figure 4: Control volume V_r on a mesh of quads

- $|\Omega_j|$ *is the volume of the cell* Ω_j*;*
- $\widehat{I_d}$ *is the identity matrix;*
- $\mathbf{C}_{jr} = \frac{1}{2}(\ell_{jl}\mathbf{n}_{jl} + \ell_{jk}\mathbf{n}_{jk})$*.*

Proof. By definition of the volume and thanks to Green's theorem, one has the following identity:

$$|\Omega_j|\widehat{I_d} = \int_{\Omega_j} \nabla \mathbf{x}\ d\mathbf{x} = \int_{\partial\Omega_j} \mathbf{x} \otimes \mathbf{n}_j\ d\sigma$$

where $\mathbf{n}_j$ is the normal to the boundary of Ω_j, directed towards the outside of the cell.

Let us divide the integration on each side of $\partial\Omega_j$, called $\partial\Omega_{jk}$

$$\int_{\partial\Omega_j} \mathbf{x} \otimes \mathbf{n}_j\ d\sigma = \sum_{k\in\mathcal{L}_j} \int_{\partial\Omega_{jk}} \mathbf{x} \otimes \mathbf{n}_j\ d\sigma$$

where $\mathcal{L}_j$ represents the cell at the edge of Ω_j.

Knowing that $\partial\Omega_{jk}$ is a straight line, one has

$$\int_{\partial\Omega_{jk}} \mathbf{x} \otimes \mathbf{n}_j\ d\sigma = \mathbf{x}_{jk} \otimes \mathbf{n}_{jk}\ell_{jk}$$

where

- $\mathbf{x}_{jk}$ is the middle of the face $\partial\Omega_{jk}$;
- $\mathbf{n}_{jk}$ is its exterior normal;
- $\ell_{jk} = |\partial\Omega_{jk}|$ is its length.

Then

$$\sum_{k\in\mathcal{L}_j} \mathbf{x}_{jk} \otimes \mathbf{n}_{jk} = \sum_{k\in\mathcal{L}_j} \frac{1}{2} \sum_{r\in\mathcal{V}_{jk}} \mathbf{x}_r \otimes \mathbf{n}_{jk}\ell_{jk} = \sum_{r\in\mathcal{V}_j} \mathbf{x}_r \otimes \frac{1}{2} \sum_{k\in\mathcal{C}_{jr}} \mathbf{n}_{jk}\ell_{jk}$$

with $\mathcal{V}_{jk}$ the vertices of $\partial\Omega_{jk}$ and $\mathcal{C}_{jr}$ the cells adjacent to Ω_j for which $\mathbf{x}_r$ is a vertice.

Let us introduce $\mathbf{C}_{jr}$, that is represented in Figure 1, such that

$$\mathbf{C}_{jr} = \frac{1}{2}\sum_{k\in\mathcal{C}_{jr}} \mathbf{n}_{jk}\ell_{jk} = \frac{1}{2}\big(\mathbf{n}_{jl}\ell_{jl} + \mathbf{n}_{jk}\ell_{jk}\big) \tag{17}$$

where Ω_l and Ω_k are adjacent to Ω_j, having $\mathbf{x}_r$ as a common vertice.

Therefore we retrieve the identity (16) and it ends the proof.

□

Remark 3. Let us consider the trace of the equation (16)

$$\mathrm{tr}(|\Omega_j|\widehat{I_d}) = \mathrm{tr}(\sum_{r\in\mathcal{V}_j} \mathbf{x}_r \otimes \mathbf{C}_{jr})$$

Knowing that the trace function is linear and considering the property Proposition 35, one has in dimension 2:

$$2|\Omega_j| = \sum_{r\in\mathcal{V}_j} \langle \mathbf{x}_r, \mathbf{C}_{jr}\rangle$$

Which leads to

$$|\Omega_j| = \frac{1}{2}\sum_{r\in\mathcal{V}_j} \langle \mathbf{x}_r, \mathbf{C}_{jr}\rangle$$

Proposition 5. *The vector $\mathbf{C}_{jr} \in \mathbb{R}^d$ can also be defined as the gradient of the volume of the cell Ω_j with respect to the vertex $\mathbf{x}_r$*

$$\mathbf{C}_{jr} = \nabla_{\mathbf{x}_r}|\Omega_j| = \frac{1}{2}(\mathbf{x}_{r-1} - \mathbf{x}_{r+1})^{\perp} \tag{18}$$

Proof. This proof is based on the definition of $\mathbf{C}_{jr}$ in [9].
In dimension 2, the oriented area of the triangle defined by the points $(\mathbf{O}, \mathbf{x}_r, \mathbf{x}_{r+1})$, where $\mathbf{O} = (0,0)$ is the origin of the mesh is: $\frac{1}{2}\Big(x_r y_{r+1} - x_{r+1}y_r\Big)$. The sum of these oriented areas on $r \in \mathcal{V}_j$ gives the surface $|\Omega_j|$

$$|\Omega_j| = \sum_{r\in\mathcal{V}_j} \frac{1}{2}\Big(x_r y_{r+1} - x_{r+1}y_r\Big)$$

Let us now derivate this formula according to $\mathbf{x}_r$

$$\begin{aligned}
\nabla_{\mathbf{x}_r}|\Omega_j| &= \nabla_{\mathbf{x}_r}\sum_{r\in\mathcal{V}_j} \frac{1}{2}\Big(x_r y_{r+1} - x_{r+1}y_r\Big) \\
&= \frac{1}{2}\begin{pmatrix} y_{r+1} - y_{r-1} \\ -x_{r+1} + x_{r-1}\end{pmatrix} \\
&= \frac{1}{2}\begin{pmatrix} -(y_{r-1} - y_{r+1}) \\ (x_{r-1} - x_{r+1})\end{pmatrix} \\
&= \frac{1}{2}(\mathbf{x}_{r-1} - \mathbf{x}_{r+1})^{\perp}
\end{aligned}$$

Let us get back to the definition of $\mathbf{C}_{jr}$ given in (17):

$$\mathbf{C}_{jr} = \frac{1}{2}\big(\ell_{jl}\mathbf{n}_{jl} + \ell_{jk}\mathbf{n}_{jk}\big)$$

As we can see on Figure 1, $\partial\Omega_{jl} = [\mathbf{x}_r, \mathbf{x}_{r+1}]$ and $\partial\Omega_{jk} = [\mathbf{x}_{r-1}, \mathbf{x}_r]$ and

$$\ell_{jl} = \|\mathbf{x}_{r+1} - \mathbf{x}_r\|_2 \quad \text{and} \quad \mathbf{n}_{jl} = \frac{(\mathbf{x}_r - \mathbf{x}_{r+1})^{\perp}}{\|\mathbf{x}_{r+1} - \mathbf{x}_r\|_2}$$

$$\ell_{jk} = \|\mathbf{x}_{r-1} - \mathbf{x}_r\|_2 \quad \text{and} \quad \mathbf{n}_{jk} = \frac{(\mathbf{x}_{r-1} - \mathbf{x}_r)^{\perp}}{\|\mathbf{x}_{r-1} - \mathbf{x}_r\|_2}$$

Thus

$$\mathbf{C}_{jr} = \frac{1}{2}\Big((\mathbf{x}_r - \mathbf{x}_{r+1})^\perp + (\mathbf{x}_{r-1} - \mathbf{x}_r)^\perp\Big) = \frac{1}{2}\big(\mathbf{x}_{r-1} - \mathbf{x}_{r+1}\big)^\perp$$

Finally we obtain $\mathbf{C}_{jr} = \nabla_{\mathbf{x}_r}|\Omega_j|$ and this ends the proof.

□

Proposition 6. *The mesh satisfies*

$$\sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} = 0, \forall r \in \mathcal{V}^i \quad \textit{and} \quad \sum_{r\in\mathcal{V}_j} \mathbf{C}_{jr} = 0 \tag{19}$$

Proof. Let us begin by proving the second equation of (19).

We showed earlier that $|\Omega_j| = \frac{1}{2}\sum_{r\in\mathcal{V}_j}\langle \mathbf{x}_r, \mathbf{C}_{jr}\rangle$. In the case where all the $\mathbf{x}_r$, coincide (ie $\exists\, \mathbf{x}$ such that $\mathbf{x} = \mathbf{x}_r,\ \forall r \in \mathcal{V}_j$), then $|\Omega_j| = 0$. Therefore

$$0 = \frac{1}{2}\sum_{r\in\mathcal{V}_j}\langle \mathbf{x}_r, \mathbf{C}_{jr}\rangle = \frac{1}{2}\langle \mathbf{x}, \sum_{r\in\mathcal{V}_j}\mathbf{C}_{jr}\rangle$$

This formula is true for any $\mathbf{x}$, hence $\sum_{r\in\mathcal{V}_j}\mathbf{C}_{jr} = 0$.

For the first equality of (19), let us get back to the definition of $\mathbf{C}_{jr}$:

$$\mathbf{C}_{jr} = \frac{1}{2}(\ell_{jl}\mathbf{n}_{jl} + \ell_{jk}\mathbf{n}_{jk})$$

where l and k are the indices of the neighbor cells of Ω_j.

Summing this equation over $j \in \mathcal{C}_r$ will create a loop and each term will cancel out with its neighbor: $\ell_{jl}\mathbf{n}_{jl} + \ell_{lj}\mathbf{n}_{lj} = 0$ for all $j, l \in \mathcal{C}_r$.

Thus, as all the terms of the sum cancel out each other, one has $\sum_{j\in\mathcal{C}_r}\mathbf{C}_{jr} = 0$.

□

Remark 4 (Assumptions on the mesh)*.* We assume the caracteristic length of the mesh is

$$h = \max_j(diam[\Omega_j]) \tag{20}$$

By assumption, there exists a constant C such that

$$\|\mathbf{C}_{jr}\|_2 \leq Ch \tag{21}$$

We make the assumption that the mesh is regular in the sens that there are two constants C_1 and C_2 such that

$$C_1h^2 \leq |\Omega_j| \leq C_2h^2, \quad \forall j \quad \text{uniformly with respect to h.} \tag{22}$$

and that

$$C_1h^2 \leq |V_r| \leq C_2h^2, \quad \forall r \quad \text{uniformly with respect to h.} \tag{23}$$

Remark 5 (Notations)*.* In this document, C, C_1 and C_2 are undefined positive constants.

3 Construction of the nodal scheme

3.1 Space discretization

Let us consider the problem (1), which we integrate over the cells Ω_j.

$$\partial_t \int_{\Omega_j} p \; d\mathbf{x} - \int_{\partial\Omega_j} \langle \kappa\mathbf{u}, \mathbf{n}\rangle \; d\sigma = \int_{\Omega_j} f \; d\mathbf{x}$$

Let us introduce $p_j = \dfrac{1}{|\Omega_j|}\displaystyle\int_{\Omega_j} p \; d\mathbf{x}$ and $f_j = \dfrac{1}{|\Omega_j|}\displaystyle\int_{\Omega_j} f \; d\mathbf{x}$. The equation becomes:

$$|\Omega_j|\partial_t p_j - \int_{\partial\Omega_j} \langle \kappa\mathbf{u}, \mathbf{n}\rangle \; d\sigma = |\Omega_j| f_j$$

In nodal formulation, we approach the integral over the boundary of Ω_j by the sum over its vertices $\mathbf{x}_r$:

$$\int_{\partial\Omega_j} \langle \kappa\mathbf{u}, \mathbf{n}\rangle \; d\sigma \approx \sum_{r\in\mathcal{V}_j} \langle \kappa_r \mathbf{u}(\mathbf{x}_r), \mathbf{C}_{jr}\rangle$$

Knowing that $\mathcal{V} = \mathcal{V}^i \cup \mathcal{V}^b \cup \mathcal{V}^c$, and since $\kappa_r = \kappa(\mathbf{x}_r)$ is symmetric, the problem becomes:

$$|\Omega_j|\partial_t p_j - \Big(\sum_{r\in\mathcal{V}_j^i} \langle \mathbf{u}(\mathbf{x}_r), \kappa_r\mathbf{C}_{jr}\rangle + \sum_{r\in\mathcal{V}_j^b} \langle \mathbf{u}(\mathbf{x}_r), \kappa_r\mathbf{C}_{jr}\rangle + \sum_{r\in\mathcal{V}_j^c} \langle \mathbf{u}(\mathbf{x}_r), \kappa_r\mathbf{C}_{jr}\rangle \Big) = |\Omega_j| f_j \tag{24}$$

In the case of a periodic domain, there are no boundary conditions to the problem (ie. $\mathcal{V}^{b,c} = \emptyset$). We will study the boundaries and the corners of the domain for different boundary conditions later in this section.

In order to approach $\mathbf{u}(\mathbf{x}_r)$, we consider Taylor's expansion of $p(\mathbf{x}_j)$:

$$p(\mathbf{x}_j) = p(\mathbf{x}_r) + \langle \mathbf{u}(\mathbf{x}_r), \mathbf{x}_j - \mathbf{x}_r\rangle + O(h^2)$$

Which means

$$\begin{aligned} &\sum_{j\in\mathcal{C}_r} p(\mathbf{x}_j)\mathbf{C}_{jr} = \sum_{j\in\mathcal{C}_r} p(\mathbf{x}_r)\mathbf{C}_{jr} - \sum_{j\in\mathcal{C}_r} \langle \mathbf{u}(\mathbf{x}_r), \mathbf{x}_r - \mathbf{x}_j\rangle \mathbf{C}_{jr} + O(h^2) \\ \Leftrightarrow\; &\sum_{j\in\mathcal{C}_r} (p(\mathbf{x}_r) - p(\mathbf{x}_j))\mathbf{C}_{jr} = \sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} \otimes (\mathbf{x}_r - \mathbf{x}_j)\mathbf{u}(\mathbf{x}_r) + O(h^2) \\ \Leftrightarrow\; &\sum_{j\in\mathcal{C}_r} (p(\mathbf{x}_r) - p(\mathbf{x}_j))\mathbf{C}_{jr} = \beta_r \mathbf{u}(\mathbf{x}_r) + O(h^2) \end{aligned} \tag{25}$$

with $\beta_r = \displaystyle\sum_{j\in\mathcal{C}_r} \beta_{jr} = \sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} \otimes (\mathbf{x}_r - \mathbf{x}_j)$.

Using Proposition 6 on $r \in \mathcal{V}^i$, one has $\displaystyle\sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} = 0$. This way, we obtain:

$$\beta_r \mathbf{u}(\mathbf{x}_r) \approx -\sum_{j\in\mathcal{C}_r} p(\mathbf{x}_j)\mathbf{C}_{jr} \quad \forall\, r \in \mathcal{V}^i$$

Then the semi-discrete diffusion scheme, in nodal formulation on a periodic domain writes:

$$\begin{cases} |\Omega_j|\partial_t p_j - \displaystyle\sum_{r\in\mathcal{V}_j^i} \langle \mathbf{u}_r, \kappa_r\mathbf{C}_{jr}\rangle = |\Omega_j| f_j \\ \qquad \beta_r \mathbf{u}_r = -\displaystyle\sum_{j\in\mathcal{C}_r} p_j \mathbf{C}_{jr} \quad \forall\, r \in \mathcal{V}^i \end{cases} \tag{26}$$

Remark 6. For the nodal flux solver to be well posed, we need the matrix β_r to be invertible, which will bring some conditions on the mesh. These conditions are given in Section 3.4.

Remark 7. The scheme (26) finds its origin in the nodal scheme for the P_1 equation (4) built by E. Franck in his thesis [13]. This scheme has the advantage of being consistent on unstructured meshes. It is given in ((27)-(28)).

$$\begin{cases} |\Omega_j|\partial_t p_j + \dfrac{1}{\varepsilon}\displaystyle\sum_r \langle \mathbf{u}_r, \mathbf{C}_{jr}\rangle = 0 \\ |\Omega_j|\partial_t \mathbf{u}_j + \dfrac{1}{\varepsilon}\displaystyle\sum_r p_{jr}\mathbf{C}_{jr} = -|\Omega_j|\dfrac{\sigma}{\varepsilon^2}\mathbf{u}_j \end{cases} \tag{27}$$

with the fluxes

$$\begin{cases} p_{jr} = p_j + \langle \mathbf{u}_j - \mathbf{u}_r, \mathbf{n}_{jr}\rangle - \dfrac{\sigma}{\varepsilon}\langle \mathbf{u}_r, \mathbf{x}_r - \mathbf{x}_j\rangle \\ \Big(\displaystyle\sum_j \mathbf{C}_{jr}\otimes\mathbf{n}_{jr} + \dfrac{\sigma}{\varepsilon}\sum_j \mathbf{C}_{jr}\otimes(\mathbf{x}_r - \mathbf{x}_j)\Big)\mathbf{u}_r = \sum_j p_j\mathbf{C}_{jr} + \sum_j \mathbf{C}_{jr}\otimes\mathbf{n}_{jr}\mathbf{u}_j \end{cases} \tag{28}$$

The Hilbert expansion of the P_1 scheme ((27)-(28)) gives the limit diffusion scheme (26).

The P_1 scheme was built to be asymptotic preserving, which means the solution of its diffusion limit (26) converges towards the solution of the diffusion equation (1) with periodic boundary conditions, even on unstructured meshes.

3.2 Boundary conditions

Let us now study the same problem with Neumann or Dirichlet boundary conditions, which leads to adding respectively (29) or (30) to (1):

$$\kappa(\mathbf{x})\mathbf{u}(\mathbf{x},t).\mathbf{n}_\Gamma = g(\mathbf{x},t) \qquad \forall\ \mathbf{x}\in\Gamma,\ t\in[0,T] \tag{29}$$

$$p(\mathbf{x},t) = h(\mathbf{x},t) \qquad \forall\ \mathbf{x}\in\Gamma,\ t\in[0,T] \tag{30}$$

3.2.1 Neumann boundary conditions

Let us get back to (24):

$$|\Omega_j|\partial_t p_j - \Big(\sum_{r\in\mathcal{V}_j^i}\langle \mathbf{u}(\mathbf{x}_r), \kappa_r\mathbf{C}_{jr}\rangle + \sum_{r\in\mathcal{V}_j^b}\langle \mathbf{u}(\mathbf{x}_r), \kappa_r\mathbf{C}_{jr}\rangle + \sum_{r\in\mathcal{V}_j^c}\langle \mathbf{u}(\mathbf{x}_r), \kappa_r\mathbf{C}_{jr}\rangle\Big) = |\Omega_j|f_j$$

On the boundary ($r\in\mathcal{V}_j^b$), one can decompose $\langle\kappa_r\mathbf{u}(\mathbf{x}_r), \mathbf{C}_{jr}\rangle$ in the direction of $\mathbf{n}_r$, the normal to the boundary of the domain at the point $\mathbf{x}_r$, and $\mathbf{t}_r = \mathbf{n}_r^\perp$:

$$\begin{aligned} \sum_{r\in\mathcal{V}_j^b}\langle\kappa_r\mathbf{u}(\mathbf{x}_r), \mathbf{C}_{jr}\rangle &= \sum_{r\in\mathcal{V}_j^b}\Big(\langle\kappa_r\mathbf{u}(\mathbf{x}_r), \mathbf{C}_{jr}.\mathbf{n}_r\ \mathbf{n}_r\rangle + \langle\kappa_r\mathbf{u}(\mathbf{x}_r), \mathbf{C}_{jr}.\mathbf{t}_r\ \mathbf{t}_r\rangle\Big) \\ &= \sum_{r\in\mathcal{V}_j^b}\Big(\langle\mathbf{C}_{jr}, \mathbf{n}_r\rangle\ \langle\kappa_r\mathbf{u}(\mathbf{x}_r), \mathbf{n}_r\rangle + \langle\kappa_r\mathbf{u}(\mathbf{x}_r), (\mathbf{t}_r\otimes\mathbf{t}_r)\mathbf{C}_{jr}\rangle\Big) \end{aligned}$$

Thanks to Neumann boundary condition (29) we obtain, $\forall\ r\in\mathcal{V}_j^b$

$$\begin{aligned} \sum_{r\in\mathcal{V}_j^b}\langle\kappa_r\mathbf{u}(\mathbf{x}_r), \mathbf{C}_{jr}\rangle &= \sum_{r\in\mathcal{V}_j^b}\Big(\langle\mathbf{C}_{jr}, \mathbf{n}_r\rangle g_r + \langle\kappa_r\mathbf{u}(\mathbf{x}_r), (\mathbf{t}_r\otimes\mathbf{t}_r)\mathbf{C}_{jr}\rangle\Big) \\ &= \sum_{r\in\mathcal{V}_j^b}\Big(\langle\mathbf{C}_{jr}, \mathbf{n}_r\rangle g_r + \langle\kappa_r\mathbf{u}(\mathbf{x}_r), (\widehat{I_d} - \mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\rangle\Big) \end{aligned} \tag{31}$$

where $g_r = g(\mathbf{x}_r)$.

Remark 8. The normal vector $\mathbf{n}_r$ is used on the nodes of the boundary. Therefore, at the "angles" defined in Definition 10, (ie. if the three adjacent nodes of this boundary are not aligned), the definition of $\mathbf{n}_r$ is not obvious.

Let us define, on the node $\mathbf{x}_r \in \Gamma$

$$\mathbf{n}_r = \frac{\displaystyle\sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr}}{\left\|\displaystyle\sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr}\right\|} = \frac{\displaystyle\sum_{j\in\mathcal{C}_r} \frac{1}{2}(\ell_{jh}\mathbf{n}_{jh} + \ell_{jk}\mathbf{n}_{jk})}{\left\|\displaystyle\sum_{j\in\mathcal{C}_r} \frac{1}{2}(\ell_{jh}\mathbf{n}_{jh} + \ell_{jk}\mathbf{n}_{jk})\right\|} = \frac{\mathbf{n}_{l_1}\ell_{l_1} + \mathbf{n}_{l_2}\ell_{l_2}}{\|\mathbf{n}_{l_1}\ell_{l_1} + \mathbf{n}_{l_2}\ell_{l_2}\|} \tag{32}$$

because inside the domain the sum on the adjacent cells is null (ie. $\displaystyle\sum_{j_1,j_2\notin\Gamma} \mathbf{n}_{j_1j_2}\ell_{j_1j_2} = 0$). One can see it on Figure 5.

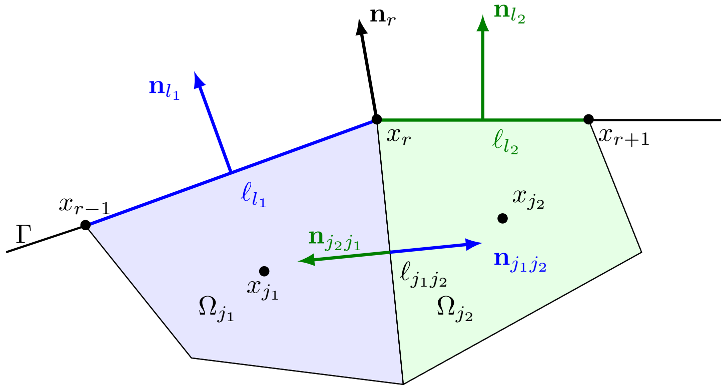


Figure 5: Construction of $\mathbf{n}_r$ on a node

For the corners defined above ($r \in \mathcal{V}_j^c$), one has $\beta_r = \mathbf{C}_{jr} \otimes (\mathbf{x}_r - \mathbf{x}_j)$ which is non invertible due to Proposition 36. Therefore we have to manipulate the formula as follows so that we don't have to invert the matrix β_r.

By definition, at a corner $\mathbf{x}_r \in \mathcal{V}_j^c$

$$\mathbf{C}_{jr} = \frac{\ell_{l_1}}{2}\mathbf{n}_{l_1} + \frac{\ell_{l_2}}{2}\mathbf{n}_{l_2}$$

Then, thanks to Neumann boundary condition (29) one has

$$\begin{aligned}\sum_{r\in\mathcal{V}_j^c} \langle \kappa_r \mathbf{u}(\mathbf{x}_r), \mathbf{C}_{jr}\rangle &= \sum_{r\in\mathcal{V}_j^c} \frac{\ell_{l_1}}{2}\langle \kappa_r \mathbf{u}(\mathbf{x}_r), \mathbf{n}_{l_1}\rangle + \sum_{r\in\mathcal{V}_j^c} \frac{\ell_{l_2}}{2}\langle \kappa_r \mathbf{u}(\mathbf{x}_r), \mathbf{n}_{l_2}\rangle \\ &= \sum_{r\in\mathcal{V}_j^c} \frac{1}{2}(\ell_{l_1} g_{r_1} + \ell_{l_2} g_{r_2}) \end{aligned} \tag{33}$$

where g_{r_1}, g_{r_2} are the boundary conditions at the point $\mathbf{x}_r$, respectively in the direction $\mathbf{n}_{l_1}$ and $\mathbf{n}_{l_2}$, and ℓ_{l_1} and ℓ_{l_2} are the lengths of Ω_j's faces, respectively on the boundaries Γ_1 and Γ_2.

Remark 9. In the code, $g_{r_{1,2}}$ the boundary conditions at the nodes of the boundary are approximated by their value at the faces of the boundary $g_{l_{1,2}} = \langle \kappa_{l_{1,2}} \mathbf{u}(\mathbf{x}_{l_{1,2}}), \mathbf{n}_{l_{1,2}}\rangle$, which are given by the boundary conditions.

One can replace (31) and (33) into (24) and obtain the first equation of the semi-discrete scheme for Neumann Boundary conditions

$$|\Omega_j|\partial_t p_j - \Big(\sum_{r\in\mathcal{V}_j^i}\langle\kappa_r\mathbf{u}_r,\mathbf{C}_{jr}\rangle + \sum_{r\in\mathcal{V}_j^b}\langle\kappa_r\mathbf{u}_r,(\widehat{I_d}-\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\rangle\Big) = |\Omega_j|f_j$$
$$+\sum_{r\in\mathcal{V}_j^b}\langle\mathbf{C}_{jr},\mathbf{n}_r\rangle g_r + \sum_{r\in\mathcal{V}_j^c}\frac{1}{2}(\ell_{l_1}g_{l_1}+\ell_{l_2}g_{l_2})$$

We have already built an equation for $\mathbf{u}_r, \forall\ r\in\mathcal{V}^i$ in (26) thus we have to determine $\mathbf{u}_r, \forall\ r\in\mathcal{V}^b$.

Repeating the process of the previous section, we consider Taylor's expansion of $p(\mathbf{x}_j)$ (25), but this time Proposition 6 does not stand, therefore we are left with a third equation for the scheme:

$$\beta_r\mathbf{u}_r = \sum_{j\in\mathcal{C}_r}(p_r-p_j)\mathbf{C}_{jr} \quad \forall r\in\mathcal{V}^b \tag{34}$$

where we have to determine p_r, a discrete approximation of $p(\mathbf{x}_r)$ for $r\in\mathcal{V}^b$.

To do so, let us introduce $\mathbf{v}_r = \dfrac{\kappa_r\mathbf{n}_r}{\|\kappa_r\mathbf{n}_r\|}$ and $\mathbf{w}_r = \mathbf{v}_r^\perp$.

We can consider the scalar product of (34) against $\mathbf{v}_r$, for all $r\in\mathcal{V}^b$:

$$\langle\mathbf{u}(\mathbf{x}_r),\mathbf{v}_r\rangle \approx \langle\beta_r^{-1}\sum_{j\in\mathcal{C}_r}(p(\mathbf{x}_r)-p(\mathbf{x}_j))\mathbf{C}_{jr},\mathbf{v}_r\rangle \tag{35}$$

Let us now consider Neumann boundary condition (29)

$$\langle\mathbf{u}(\mathbf{x}_r),\mathbf{v}_r\rangle = \frac{g_r}{\|\kappa_r\mathbf{n}_r\|} \tag{36}$$

Combining (35) and (36) and isolating p_r, one gets

$$\langle\beta_r^{-1}\sum_{j\in\mathcal{C}_r}(p(\mathbf{x}_r)-p(\mathbf{x}_j))\mathbf{C}_{jr},\mathbf{v}_r\rangle \approx \frac{g_r}{\|\kappa_r\mathbf{n}_r\|}$$
$$\Leftrightarrow \quad p(\mathbf{x}_r)\sum_{j\in\mathcal{C}_r}\theta_{jr} \approx \sum_{j\in\mathcal{C}_r}\theta_{jr}p(\mathbf{x}_j) + \frac{g_r}{\|\kappa_r\mathbf{n}_r\|}$$

where $\theta_{jr} = \langle\beta_r^{-1}\mathbf{C}_{jr},\mathbf{v}_r\rangle$. Let us assume that $\sum_{j\in\mathcal{C}_r}\theta_{jr}\neq 0$:

$$\Leftrightarrow \quad p(\mathbf{x}_r) \approx \Big(\sum_{j\in\mathcal{C}_r}\theta_{jr}\Big)^{-1}\Big(\sum_{j\in\mathcal{C}_r}\theta_{jr}p(\mathbf{x}_j) + \frac{g_r}{\|\kappa_r\mathbf{n}_r\|}\Big)$$

Thus we obtain the discrete approximation for p_r:

$$\begin{aligned} p_r &= \Big(\sum_{j\in\mathcal{C}_r}\theta_{jr}\Big)^{-1}\Big(\sum_{j\in\mathcal{C}_r}\theta_{jr}p_j + \frac{g_r}{\|\kappa_r\mathbf{n}_r\|}\Big) \\ &= \theta_r^{-1}\Big(\sum_{j\in\mathcal{C}_r}\theta_{jr}p_j + \frac{g_r}{\|\kappa_r\mathbf{n}_r\|}\Big) \end{aligned} \tag{37}$$

with $\theta_r = \sum_{j\in\mathcal{C}_r}\theta_{jr}$.

Then the semi-discrete scheme for Neumann boundary conditions writes:

$$\begin{cases} |\Omega_j|\partial_t p_j - \Big(\sum_{r\in\mathcal{V}_j^i}\langle\kappa_r\mathbf{u}_r,\mathbf{C}_{jr}\rangle + \sum_{r\in\mathcal{V}_j^b}\langle\kappa_r\mathbf{u}_r,(\widehat{I_d}-\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\rangle\Big) = B_j \\ \beta_r\mathbf{u}_r = -\sum_{j\in\mathcal{C}_r}p_j\mathbf{C}_{jr} \quad \forall r\in\mathcal{V}^i \\ \beta_r\mathbf{u}_r = \sum_{j\in\mathcal{C}_r}(p_r-p_j)\mathbf{C}_{jr} \quad \forall r\in\mathcal{V}^b \end{cases} \tag{38}$$

where

- $B_j = |\Omega_j| f_j + \sum_{r \in \mathcal{V}_j^b} \langle \mathbf{C}_{jr}, \mathbf{n}_r \rangle g_r + \sum_{r \in \mathcal{V}_j^c} \frac{1}{2}(\ell_{l_1} g_{l_1} + \ell_{l_2} g_{l_2})$;
- $p_r = \theta_r^{-1} \Big(\sum_{j \in \mathcal{C}_r} \theta_{jr} p_j + \frac{g_r}{\|\kappa_r \mathbf{n}_r\|} \Big)$;
- $\theta_{jr} = \langle \beta_r^{-1} \mathbf{C}_{jr}, \mathbf{v}_r \rangle$.

3.2.2 Dirichlet boundary conditions

Let us get back to (24), with Dirichlet boundary conditions (30), which gives us immediately the first equation of the scheme:

$$\partial_t p_j - \frac{1}{|\Omega_j|} \Big(\sum_{r \in \mathcal{V}_j^i} \langle \mathbf{u}(\mathbf{x}_r), \kappa_r \mathbf{C}_{jr} \rangle + \sum_{r \in \mathcal{V}_j^b} \langle \mathbf{u}(\mathbf{x}_r), \kappa_r \mathbf{C}_{jr} \rangle + \sum_{r \in \mathcal{V}_j^c} \langle \mathbf{u}(\mathbf{x}_r), \kappa_r \mathbf{C}_{jr} \rangle \Big) = f_j \tag{39}$$

Let us now determine $\mathbf{u}_r$ for all $r \in \mathcal{V}^b$. To do so we will consider Taylor's expansion of p (25) and replace $p(\mathbf{x}_r)$ with Dirichlet boundary condition (30)

$$\beta_r \mathbf{u}(\mathbf{x}_r) \approx \sum_{j \in \mathcal{C}_r} (h_r - p(\mathbf{x}_j)) \mathbf{C}_{jr} \quad \forall r \in \mathcal{V}^b$$

It gives us the third equation of the semi-discrete scheme for Dirichlet boundary conditions

$$\beta_r \mathbf{u}_r = \sum_{j \in \mathcal{C}_r} (h_r - p_j) \mathbf{C}_{jr} \quad \forall r \in \mathcal{V}^b$$

Let us now consider $r \in \mathcal{V}^c$ (ie. $\mathbf{x}_r$ is on a corner with a unique cell), as stated in Proposition 36, in that case β_r is non invertible. We will try to approximate β_r with an invertible matrix. To do so, let us divide the corner cell into two as one can see on Figure 6. In that case β_r rewrites

$$\beta_r^c = \mathbf{C}_{j_1 r} \otimes (\mathbf{x}_r - \mathbf{x}_{j_1}) + \mathbf{C}_{j_2 r} \otimes (\mathbf{x}_r - \mathbf{x}_{j_2})$$

with $\mathbf{C}_{j_1 r} = \frac{1}{2}\big(\mathbf{x}_{r-1} - \mathbf{x}_{r+2}\big)^{\perp}$, $\mathbf{C}_{j_2 r} = \frac{1}{2}\big(\mathbf{x}_{r+2} - \mathbf{x}_{r+1}\big)^{\perp}$ and $\mathbf{x}_{j_1}$, $\mathbf{x}_{j_2}$ respectively the barycenters of the newly divided cells Ω_{j_1} and Ω_{j_2}.

Therefore the problem is well-posed (ie. β_r^c is invertible) if $\mathbf{C}_{j_1 r}$ and $\mathbf{C}_{j_2 r}$ are non collinear, which means $\mathbf{x}_{r-1}, \mathbf{x}_{r+1}$ and $\mathbf{x}_{r+2}$ are not aligned. One can see the forbidden configuration in Figure 7.

Thus we obtain the fourth equation of the scheme

$$\beta_r^c \mathbf{u}_r = \sum_{j \in \mathcal{C}_r} (h_r - p_j) \mathbf{C}_{jr} \quad \forall r \in \mathcal{V}^c$$

Finally it gives us the following semi-discrete scheme for Dirichlet boundary conditions:

$$\begin{cases} |\Omega_j| \partial_t p_j - \sum_{r \in \mathcal{V}_j} \langle \kappa_r \mathbf{u}_r, \mathbf{C}_{jr} \rangle = |\Omega_j| f_j \\ \beta_r \mathbf{u}_r = - \sum_{j \in \mathcal{C}_r} p_j \mathbf{C}_{jr} & \text{if } r \in \mathcal{V}_j^i \\ \beta_r \mathbf{u}_r = \sum_{j \in \mathcal{C}_r} (h_r - p_j) \mathbf{C}_{jr} & \text{if } r \in \mathcal{V}_j^b \\ \beta_r^c \mathbf{u}_r = \sum_{j \in \mathcal{C}_r} (h_r - p_j) \mathbf{C}_{jr} & \text{if } r \in \mathcal{V}_j^c \end{cases} \tag{40}$$

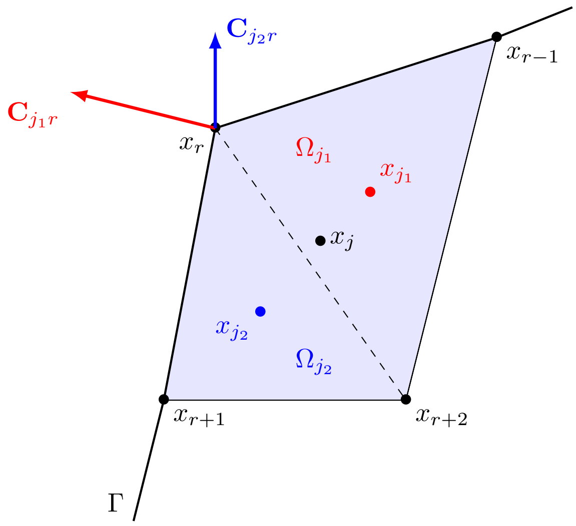


Figure 6: Construction of β_r^c on a corner of Ω with a single cell

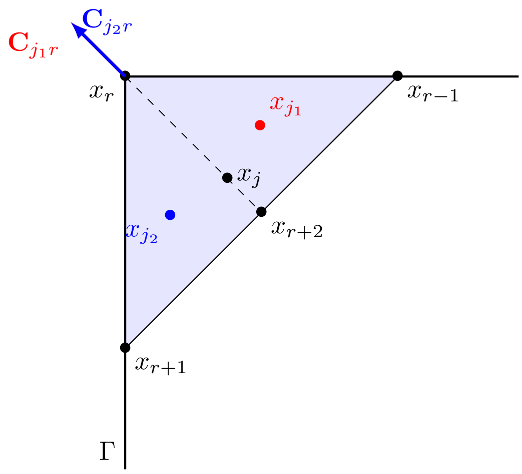


Figure 7: Example of forbidden mesh at the corner

3.2.3 Mixed boundary conditions

Let us now consider the case where some of the boundaries have a Neumann boundary condition and some other have a Dirichlet one. In practice the boundary conditions are implemented on every face of the boundary, but we are working on the boundary vertices. Each one of them is linked to two faces, therefore it can have several boundary conditions, descripted below and visible on Figure 8:

- Say r_1 is linked to two Neumann faces, let us call it Neumann, then $r \in \mathcal{V}^N$;
- Say r_2 is linked to two Dirichlet faces, it is considered Dirichlet and $r \in \mathcal{V}^D$;
- Say r_3 is linked to one Neumann face and one Dirichlet face, then the Dirichlet boundary condition prevails and the node is considered Dirichlet: $r \in \mathcal{V}^D$.

Then the semi -discrete scheme is a mix of the two previous ones:

$$
\begin{cases}
|\Omega_j|\partial_t p_j - \Big(\sum_{r\in\mathcal{V}_j^i} \langle \kappa_r \mathbf{u}_r, \mathbf{C}_{jr}\rangle + \sum_{r\in\mathcal{V}_j^{N,b}} \langle \kappa_r \mathbf{u}_r, (\widehat{I_d} - \mathbf{n}_r \otimes \mathbf{n}_r)\mathbf{C}_{jr}\rangle \Big) = B_j \\
\beta_r \mathbf{u}_r = -\sum_{j\in\mathcal{C}_r} p_j \mathbf{C}_{jr} \qquad \text{if } r \in \mathcal{V}_j^i \\
\beta_r \mathbf{u}_r = \sum_{j\in\mathcal{C}_r} (p_r - p_j)\mathbf{C}_{jr} \quad \text{if } r \in \mathcal{V}^{N,b} \\
\beta_r \mathbf{u}_r = \sum_{j\in\mathcal{C}_r} (h_r - p_j)\mathbf{C}_{jr} \quad \text{if } r \in \mathcal{V}_j^{D,b} \\
\beta_r^c \mathbf{u}_r = \sum_{j\in\mathcal{C}_r} (h_r - p_j)\mathbf{C}_{jr} \quad \text{if } r \in \mathcal{V}_j^{D,c}
\end{cases}
\tag{41}
$$

where

- $B_j = |\Omega_j| f_j + \sum_{r\in\mathcal{V}_j^{N,b}} \langle \mathbf{C}_{jr}, \mathbf{n}_r\rangle g_r + \sum_{r\in\mathcal{V}_j^{N,c}} \frac{1}{2}(\ell_{l_1} g_{l_1} + \ell_{l_2} g_{l_2});$
- $p_r = \theta_r^{-1}\Big(\sum_{j\in\mathcal{C}_r} \theta_{jr} p_j + \frac{g_r}{\|\kappa_r \mathbf{n}_r\|} \Big)$ for all $r \in \mathcal{V}^{N,b}$;
- $\theta_{jr} = \langle \beta_r^{-1} \mathbf{C}_{jr}, \mathbf{v}_r\rangle$ for all $j \in \mathcal{C}$, $r \in \mathcal{V}^{N,b}$.

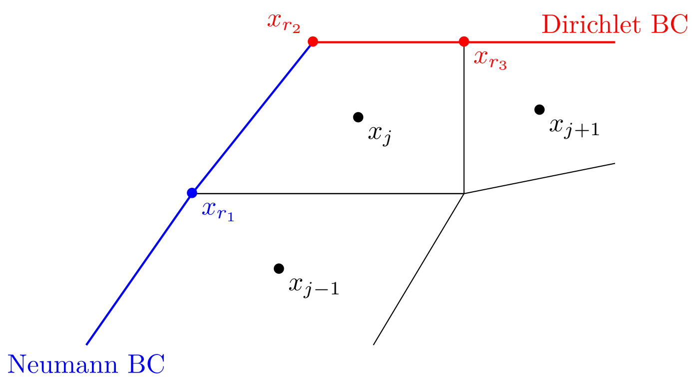


Figure 8: Representation of the nodes boundary conditions

3.3 Time discretization

In order to discretize the scheme in time, let us use the backward Euler method (or implicit Euler method). We chose the implicit scheme because it will give us unconditional stability in Section 4. Let us introduce the time step $\Delta t > 0$, such that $t^n = n\Delta t \le T$, with T the final time.

The scheme writes, for periodic boundary conditions

$$
\begin{cases}
|\Omega_j| \dfrac{p_j^{n+1} - p_j^n}{\Delta t} - \sum_{r\in\mathcal{V}_j} \langle \mathbf{u}_r^{n+1}, \kappa_r \mathbf{C}_{jr}\rangle = |\Omega_j| f_j^{n+1} \\
\beta_r \mathbf{u}_r^{n+1} = -\sum_{j\in\mathcal{C}_r} p_j^{n+1} \mathbf{C}_{jr}
\end{cases}
\tag{42}
$$

For Neumann boundary conditions

$$\begin{cases} |\Omega_j|\dfrac{p_j^{n+1}-p_j^n}{\Delta t} - \Big(\displaystyle\sum_{r\in\mathcal{V}_j^i}\langle\kappa_r\mathbf{u}_r^{n+1},\mathbf{C}_{jr}\rangle + \sum_{r\in\mathcal{V}_j^b}\langle\kappa_r\mathbf{u}_r^{n+1},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\rangle\Big) = B_j^{n+1} \\ \beta_r\mathbf{u}_r^{n+1} = -\displaystyle\sum_{j\in\mathcal{C}_r} p_j^{n+1}\mathbf{C}_{jr} \quad \forall r\in\mathcal{V}^i \\ \beta_r\mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r}(p_r^{n+1}-p_j^{n+1})\mathbf{C}_{jr} \quad \forall r\in\mathcal{V}^b \end{cases} \tag{43}$$

where

- $B_j^{n+1} = |\Omega_j|f_j^{n+1} + \displaystyle\sum_{r\in\mathcal{V}_j^b}\langle\mathbf{C}_{jr},\mathbf{n}_r\rangle g_r + \sum_{r\in\mathcal{V}_j^c}\frac{1}{2}(\ell_{l_1}g_{l_1}+\ell_{l_2}g_{l_2})$;
- $p_r^{n+1} = \theta_r^{-1}\Big(\displaystyle\sum_{j\in\mathcal{C}_r}\theta_{jr}p_j^{n+1} + \frac{g_r}{\|\kappa_r\mathbf{n}_r\|}\Big)$;
- $\theta_{jr} = \langle\beta_r^{-1}\mathbf{C}_{jr},\mathbf{v}_r\rangle$.

For Dirichlet boundary conditions

$$\begin{cases} |\Omega_j|\dfrac{p_j^{n+1}-p_j^n}{\Delta t} - \displaystyle\sum_{r\in\mathcal{V}_j}\langle\kappa_r\mathbf{u}_r^{n+1},\mathbf{C}_{jr}\rangle \quad = |\Omega_j|f_j^{n+1} \\ \beta_r\mathbf{u}_r^{n+1} = -\displaystyle\sum_{j\in\mathcal{C}_r}p_j^{n+1}\mathbf{C}_{jr} \qquad \text{if } r\in\mathcal{V}_j^i \\ \beta_r\mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r}(h_r-p_j^{n+1})\mathbf{C}_{jr} \qquad \text{if } r\in\mathcal{V}_j^b \\ \beta_r^c\mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r}(h_r-p_j^{n+1})\mathbf{C}_{jr} \qquad \text{if } r\in\mathcal{V}_j^c \end{cases} \tag{44}$$

And for mixed boundary conditions

$$\begin{cases} |\Omega_j|\dfrac{p_j^{n+1}-p_j^n}{\Delta t} - \Big(\displaystyle\sum_{r\in\mathcal{V}_j^i}\langle\kappa_r\mathbf{u}_r^{n+1},\mathbf{C}_{jr}\rangle + \sum_{r\in\mathcal{V}_j^{N,b}}\langle\kappa_r\mathbf{u}_r^{n+1},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\rangle + \sum_{r\in\mathcal{V}_j^{D,b}}\langle\kappa_r\mathbf{u}_r^{n+1},\mathbf{C}_{jr}\rangle \\ \qquad\qquad + \displaystyle\sum_{r\in\mathcal{V}_j^{D,c}}\langle\kappa_r\mathbf{u}_r^{n+1},\mathbf{C}_{jr}\rangle\Big) = B_j^{n+1} \\ \beta_r\mathbf{u}_r^{n+1} = -\displaystyle\sum_{j\in\mathcal{C}_r}p_j^{n+1}\mathbf{C}_{jr} \qquad \text{if } r\in\mathcal{V}_j^i \\ \beta_r\mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r}(p_r^{n+1}-p_j^{n+1})\mathbf{C}_{jr} \qquad \text{if } r\in\mathcal{V}^{N,b} \\ \beta_r\mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r}(h_r-p_j^{n+1})\mathbf{C}_{jr} \qquad \text{if } r\in\mathcal{V}_j^{D,b} \\ \beta_r^c\mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r}(h_r-p_j^{n+1})\mathbf{C}_{jr} \qquad \text{if } r\in\mathcal{V}_j^{D,c} \end{cases} \tag{45}$$

where

- $B_j^{n+1} = |\Omega_j|f_j^{n+1} + \displaystyle\sum_{r\in\mathcal{V}_j^{N,b}}\langle\mathbf{C}_{jr},\mathbf{n}_r\rangle g_r + \sum_{r\in\mathcal{V}_j^{N,c}}\frac{1}{2}(\ell_{l_1}g_{l_1}+\ell_{l_2}g_{l_2})$;
- $p_r^{n+1} = \theta_r^{-1}\Big(\displaystyle\sum_{j\in\mathcal{C}_r}\theta_{jr}p_j^{n+1} + \frac{g_r}{\|\kappa_r\mathbf{n}_r\|}\Big)$ for all $r\in\mathcal{V}^{N,b}$;

- $\theta_{jr} = \langle \beta_r^{-1}\mathbf{C}_{jr}, \mathbf{v}_r \rangle$ for all $j \in \mathcal{C}$, $r \in \mathcal{V}^{N,b}$.

3.4 Invertibility of β_r

In this section, we will study the conditions on the mesh such that β_r is positive, that is

$$\langle \mathbf{x}, \beta_r \mathbf{x} \rangle > 0 \quad \forall \mathbf{x} \in \mathbb{R}, \quad \mathbf{x} \neq 0.$$

Proposition 7 (Decomposition of β_r)**.** *The matrix β_r satisfies*

$$\beta_r = |V_r|\widehat{I}_d + P \tag{46}$$

where

- $P = \sum_{j\in\mathcal{C}_r} \frac{1}{2}\left(\mathbf{v}^{\perp}_{j+\frac{1}{2}} \otimes \mathbf{v}_{j+\frac{1}{2}} - \mathbf{w}^{\perp}_{j-\frac{1}{2}} \otimes \mathbf{w}_{j-\frac{1}{2}}\right)$
- $\mathbf{w}_{j-\frac{1}{2}} = \mathbf{x}_j - \mathbf{x}_{j-\frac{1}{2}}$
- $\mathbf{v}_{j+\frac{1}{2}} = \mathbf{x}_{j+\frac{1}{2}} - \mathbf{x}_j$
- $\widehat{I}_d$ *is the identity matrix.*

One has the formula

$$\mathrm{tr}(\beta_r) = 2|V_r| \tag{47}$$

with V_r the control volume given in Figure 4.

Proof. Using the same principle as previously for the volume $|\Omega_j|$, one can write

$$|V_r|\widehat{I}_d = \int_{V_r} \nabla \mathbf{x} \; d\mathbf{x} = \int_{\partial V_r} \mathbf{x} \otimes \mathbf{n} \; d\sigma$$

with $\mathbf{n}$ the exterior normal to the control volume V_r.

We can decompose the integral on each cell of the control volume:

$$\int_{\partial V_r} \mathbf{x} \otimes \mathbf{n} \; d\sigma = \sum_{j\in\mathcal{C}_r} \int_{\partial V_{jr}} \mathbf{x} \otimes \mathbf{n} \; d\sigma$$

where V_{jr} is defined in Definition 11. As a reminder it is the area delimited by the points $\mathbf{x}_r, \mathbf{x}_{j+\frac{1}{2}}, \mathbf{x}_j, \mathbf{x}_{j-\frac{1}{2}}$. One can see it in Figure 4.

Knowing that the identity is linear and the boundaries of V_{jr} are straight lines, one can write

$$\int_{\partial V_{jr}} \mathbf{x} \otimes \mathbf{n} \; d\sigma = \frac{1}{2}\left(\mathbf{x}_{j+\frac{1}{2}} + \mathbf{x}_j\right) \otimes \left(\mathbf{x}_j - \mathbf{x}_{j+\frac{1}{2}}\right)^{\perp} + \frac{1}{2}\left(\mathbf{x}_{j-\frac{1}{2}} + \mathbf{x}_j\right) \otimes \left(\mathbf{x}_{j-\frac{1}{2}} - \mathbf{x}_j\right)^{\perp}$$

Which leads to

$$|V_r|\widehat{I}_d = \sum_{j\in\mathcal{C}_r} \frac{1}{2}\left(\mathbf{x}_{j+\frac{1}{2}} + \mathbf{x}_j\right) \otimes \left(\mathbf{x}_j - \mathbf{x}_{j+\frac{1}{2}}\right)^{\perp} + \frac{1}{2}\left(\mathbf{x}_{j-\frac{1}{2}} + \mathbf{x}_j\right) \otimes \left(\mathbf{x}_{j-\frac{1}{2}} - \mathbf{x}_j\right)^{\perp} \tag{48}$$

Since $j \in \mathcal{C}_r$ forms a loop, one has $\sum_{j\in\mathcal{C}_r} \left(\mathbf{x}_{j-\frac{1}{2}} - \mathbf{x}_{j+\frac{1}{2}}\right) = 0$, therefore

$$\sum_{j\in\mathcal{C}_r} \left(\mathbf{x}_{j-\frac{1}{2}} - \mathbf{x}_{j+\frac{1}{2}}\right)^{\perp} \otimes \mathbf{x}_r = \sum_{j\in\mathcal{C}_r} \left(\mathbf{x}_j - \mathbf{x}_{j+\frac{1}{2}}\right)^{\perp} \otimes \mathbf{x}_r + \left(\mathbf{x}_{j-\frac{1}{2}} - \mathbf{x}_j\right)^{\perp} \otimes \mathbf{x}_r = 0$$

We can deduct this expression from the equation (3.4) and we obtain

$$\begin{aligned}
|V_r|\widehat{I}_d &= \sum_{j\in\mathcal{C}_r} \Big(\frac{1}{2}(\mathbf{x}_{j+\frac{1}{2}}+\mathbf{x}_j)-\mathbf{x}_r\Big)\otimes\big(\mathbf{x}_j-\mathbf{x}_{j+\frac{1}{2}}\big)^\perp + \Big(\frac{1}{2}(\mathbf{x}_{j-\frac{1}{2}}+\mathbf{x}_j)-\mathbf{x}_r\Big)\otimes\big(\mathbf{x}_{j-\frac{1}{2}}-\mathbf{x}_j\big)^\perp \\
&= \sum_{j\in\mathcal{C}_r} \Big(\frac{1}{2}(\mathbf{x}_{j+\frac{1}{2}}-\mathbf{x}_j)+\mathbf{x}_j-\mathbf{x}_r\Big)\otimes\big(\mathbf{x}_j-\mathbf{x}_{j+\frac{1}{2}}\big)^\perp + \Big(\frac{1}{2}(\mathbf{x}_{j-\frac{1}{2}}-\mathbf{x}_j)+\mathbf{x}_j-\mathbf{x}_r\Big)\otimes\big(\mathbf{x}_{j-\frac{1}{2}}-\mathbf{x}_j\big)^\perp \\
&= \sum_{j\in\mathcal{C}_r} \Big((\mathbf{x}_j-\mathbf{x}_r)\otimes\big((\mathbf{x}_j-\mathbf{x}_{j+\frac{1}{2}})^\perp+(\mathbf{x}_{j-\frac{1}{2}}-\mathbf{x}_j)^\perp\big)\Big) + \sum_{j\in\mathcal{C}_r}\Big(\frac{1}{2}(\mathbf{x}_{j+\frac{1}{2}}-\mathbf{x}_j)\otimes(\mathbf{x}_j-\mathbf{x}_{j+\frac{1}{2}})^\perp \\
&\qquad + \frac{1}{2}(\mathbf{x}_{j-\frac{1}{2}}-\mathbf{x}_j)\otimes(\mathbf{x}_{j-\frac{1}{2}}-\mathbf{x}_j)^\perp\Big)
\end{aligned} \tag{49}$$

Let us get back to the definition of $\mathbf{C}_{jr}$:

$$\mathbf{C}_{jr} = \frac{1}{2}\big(\mathbf{x}_{r-1}-\mathbf{x}_{r+1}\big)^\perp = (\mathbf{x}_{j+\frac{1}{2}}-\mathbf{x}_{j-\frac{1}{2}})^\perp$$

We can replace it into (49) and we obtain

$$\begin{aligned}
|V_r|\widehat{I}_d &= \sum_{j\in\mathcal{C}_r}(\mathbf{x}_r-\mathbf{x}_j)\otimes\mathbf{C}_{jr} + \sum_{j\in\mathcal{C}_r}\Big(\frac{1}{2}(\mathbf{x}_{j+\frac{1}{2}}-\mathbf{x}_j)\otimes(\mathbf{x}_j-\mathbf{x}_{j+\frac{1}{2}})^\perp + \frac{1}{2}(\mathbf{x}_{j-\frac{1}{2}}-\mathbf{x}_j)\otimes(\mathbf{x}_{j-\frac{1}{2}}-\mathbf{x}_j)^\perp\Big) \\
&= \sum_{j\in\mathcal{C}_r}(\mathbf{x}_r-\mathbf{x}_j)\otimes\mathbf{C}_{jr} + \sum_{j\in\mathcal{C}_r}\frac{1}{2}\Big(-\mathbf{v}_{j+\frac{1}{2}}\otimes\mathbf{v}^\perp_{j+\frac{1}{2}} + \mathbf{w}_{j-\frac{1}{2}}\otimes\mathbf{w}^\perp_{j-\frac{1}{2}}\Big)
\end{aligned}$$

We know thanks to Proposition 33 that $\mathbf{x}\otimes\mathbf{y} = (\mathbf{y}\otimes\mathbf{x})^t$.

Thus

$$\begin{aligned}
|V_r|\widehat{I}_d = (|V_r|\widehat{I}_d)^t &= \sum_{j\in\mathcal{C}_r}\mathbf{C}_{jr}\otimes(\mathbf{x}_r-\mathbf{x}_j) - \sum_{j\in\mathcal{C}_r}\frac{1}{2}\Big(\mathbf{v}^\perp_{j+\frac{1}{2}}\otimes\mathbf{v}_{j+\frac{1}{2}} - \mathbf{w}^\perp_{j-\frac{1}{2}}\otimes\mathbf{w}_{j-\frac{1}{2}}\Big) \\
&= \beta_r - P
\end{aligned}$$

with $P, \mathbf{v}_{j+\frac{1}{2}}$ and $\mathbf{w}_{j-\frac{1}{2}}$ defined earlier.

That way, we obtain the equality (46).

Thanks to know that, for all vector $\mathbf{v}$, $\mathrm{tr}(\mathbf{v}^\perp\otimes\mathbf{v}) = \langle\mathbf{v}^\perp,\mathbf{v}\rangle = 0$. Which means

$$\mathrm{tr}(\beta_r) = \mathrm{tr}(|V_r|\widehat{I}_d) + \mathrm{tr}(P) = d|V_r|$$

with $d = 2$ because we are in a two-dimensional space.
Therefore we obtain (47) and it ends the proof.

□

Remark 10. We would like to study β_r's invertibility with its determinant. One has:

$$\det(\beta_r) = \det(|V_r|\widehat{I}_d + P)$$

Unfortunately there is no formula that can help us determine the determinant of a sum, therefore we will study β_r's positivity in order to obtain its invertibility. This property will be useful as a hypothesis for the stability of the scheme when $\kappa = I_d$.

Because of the circular numbering around the control volume V_r, one can write

$$\sum_{j\in\mathcal{C}_r}\big(\mathbf{x}_j-\mathbf{x}_{j-\frac{1}{2}}\big)^\perp\otimes\big(\mathbf{x}_{j-1}-\mathbf{x}_{j-\frac{1}{2}}\big) = \sum_{j\in\mathcal{C}_r}\big(\mathbf{x}_{j+1}-\mathbf{x}_{j+\frac{1}{2}}\big)^\perp\otimes\big(\mathbf{x}_{j+1}-\mathbf{x}_{j+\frac{1}{2}}\big)$$

Therefore

$$P = \frac{1}{2} \sum_{j \in \mathcal{C}_r} \left(\mathbf{v}^{\perp}_{j+\frac{1}{2}} \otimes \mathbf{v}_{j+\frac{1}{2}} - \mathbf{w}^{\perp}_{j+\frac{1}{2}} \otimes \mathbf{w}_{j+\frac{1}{2}} \right) = \widetilde{P}$$

Let us write

$$\beta_r = \sum_{j \in \mathcal{C}_r} \left(|V_{jr}| \widehat{I}_d + P_j \right) \tag{50}$$

$$= \sum_{j \in \mathcal{C}_r} \left(|\widetilde{V}_{jr}| \widehat{I}_d + \widetilde{P}_j \right) \tag{51}$$

where

- $\widetilde{V}_{jr}$ is defined by $(\mathbf{x}_r, \mathbf{x}_j, \mathbf{x}_{j+\frac{1}{2}}, \mathbf{x}_{j+1})$:
- V_{jr} is defined by $(\mathbf{x}_r, \mathbf{x}_{j-\frac{1}{2}}, \mathbf{x}_j, \mathbf{x}_{j+\frac{1}{2}})$:
- $P_j = \frac{1}{2} \left(\mathbf{v}^{\perp}_{j+\frac{1}{2}} \otimes \mathbf{v}_{j+\frac{1}{2}} - \mathbf{w}^{\perp}_{j-\frac{1}{2}} \otimes \mathbf{w}_{j-\frac{1}{2}} \right)$;
- $\widetilde{P}_j = \frac{1}{2} \left(\mathbf{v}^{\perp}_{j+\frac{1}{2}} \otimes \mathbf{v}_{j+\frac{1}{2}} - \mathbf{w}^{\perp}_{j+\frac{1}{2}} \otimes \mathbf{w}_{j+\frac{1}{2}} \right)$;

Proposition 8 (β_r's positivity)**.** *β_r is positive under one of the following conditions:*
For all $j \in \mathcal{C}$,

$$|V_{jr}| > \left\| \mathbf{x}_{j+\frac{1}{2}} - \mathbf{x}_{j-\frac{1}{2}} \right\|_2 \left\| \mathbf{x}_j - \frac{1}{2} \left(\mathbf{x}_{j+\frac{1}{2}} + \mathbf{x}_{j-\frac{1}{2}} \right) \right\|_2 \tag{52}$$

Or

$$|\widetilde{V}_{jr}| > \left\| \mathbf{x}_{j+1} - \mathbf{x}_j \right\|_2 \left\| \mathbf{x}_{j+\frac{1}{2}} - \frac{1}{2} \left(\mathbf{x}_{j+1} + \mathbf{x}_j \right) \right\|_2 \tag{53}$$

with $\|\mathbf{x}\|_2 = \sqrt{x^2 + y^2}$, for all $\mathbf{x} = (x, y)$.

Proof. Based on (50), one can write

$$\langle \mathbf{x}, \beta_r \mathbf{x} \rangle = \sum_{j \in \mathcal{C}_r} \left(|V_{jr}| \|\mathbf{x}\|^2 + \langle \mathbf{x}, P^s_j \mathbf{x} \rangle \right) \geq \sum_{j \in \mathcal{C}_r} \left(|V_{jr}| - \rho(P^s_j) \right) \|\mathbf{x}\|^2$$

with

- P^s_j the symmetric part of the matrix P_j: $P^s_j = \frac{1}{2}(P_j + P^t_j)$;
- $\rho(A)$ the spectral radius of the matrix A. Let us introduce $\lambda_1, \ldots, \lambda_n$ the eigenvalues of A, then its spectral radius is $\rho(A) = \max_{i \in 1, \ldots, n} |\lambda_i|$;
- $\|.\|$ the norm corresponding to the euclidian scalar product $\langle ., . \rangle$.

This formula remains true if we consider (51) instead of (50).

Proving that β_r is positive is equivalent to showing that for all $j \in \mathcal{C}_r$, $|V_{jr}| - \rho(P^s_j) > 0$.

Let us now introduce $C = \mathbf{v}^{\perp} \otimes \mathbf{v} - \mathbf{w}^{\perp} \otimes \mathbf{w}$, with $\mathbf{v} = (\tilde{a}, \tilde{b})$, $\mathbf{v}^{\perp} = (-\tilde{b}, \tilde{a})$ and $\mathbf{w} = (a, b)$, $\mathbf{w}^{\perp} = (-b, a)$.

Then

$$C^s = \frac{C + C^t}{2} = \begin{pmatrix} ab - \tilde{a}\tilde{b} & \dfrac{\tilde{a}^2 - a^2 - \tilde{b}^2 + b^2}{2} \\ \dfrac{\tilde{a}^2 - a^2 - \tilde{b}^2 + b^2}{2} & \tilde{a}\tilde{b} - ab \end{pmatrix}$$

We note $x = (a - \tilde{a})$, $y = (b - \tilde{b})$, $\alpha = \frac{1}{2}(b + \tilde{b})$ and $\beta = \frac{1}{2}(a + \tilde{a})$.

Knowing that for all $\mathbf{v}$, $\mathrm{tr}(\mathbf{v}^\perp \otimes \mathbf{v}) = 0$, we have $\mathrm{tr}(C^s) = \dfrac{\mathrm{tr}(C) + \mathrm{tr}(C^t)}{2} = 0$.

One can see that $ab - \tilde{a}\tilde{b} = x\alpha + y\beta$ and $\dfrac{\tilde{a}^2 - a^2 - \tilde{b}^2 + b^2}{2} = -\beta x + \alpha y$.

Then

$$\begin{aligned}\det(C^s) &= -(ab - \tilde{a}\tilde{b})^2 - \Big(\frac{\tilde{a}^2 - a^2 - \tilde{b}^2 + b^2}{2}\Big)^2 = -(x\alpha + y\beta)^2 - (\beta x - \alpha y)^2 \\ &= -(y^2 + x^2)^2(\alpha^2 + \beta^2)^2 = -\Big\|\mathbf{w} - \mathbf{v}\Big\|_2^2 \Big\|\frac{1}{2}(\mathbf{w} + \mathbf{v})\Big\|_2^2\end{aligned}$$

with $\|.\|_2^2 = \langle ., . \rangle$.

Since C^s is a two-dimensional matrix, Proposition 32 says that its characteristic polynomial writes

$$\lambda^2 - \mathrm{tr}(C^s)\lambda + \det(C^s) = 0$$

with λ its eigenvalues.

Hence

$$\rho(C^s) = \lambda_{\max} = \frac{\mathrm{tr}(C^s) + \sqrt{\mathrm{tr}(C^s)^2 - 4\det(C^s)}}{2} = \sqrt{-\det(C^s)} = \Big\|\mathbf{v} - \mathbf{w}\Big\|_2 \Big\|\frac{1}{2}(\mathbf{v} + \mathbf{w})\Big\|_2$$

Replacing C with P_j, the positivity condition for β_r rewrites

$$|V_{jr}| > \rho(P_j^s) = \Big\|\mathbf{v}_{j+\frac{1}{2}} - \mathbf{w}_{j-\frac{1}{2}}\Big\|_2 \Big\|\frac{1}{2}(\mathbf{v}_{j+\frac{1}{2}} + \mathbf{w}_{j-\frac{1}{2}})\Big\|_2$$

With $\mathbf{w}_{j-\frac{1}{2}} = \mathbf{x}_j - \mathbf{x}_{j-\frac{1}{2}}$ and $\mathbf{v}_{j+\frac{1}{2}} = \mathbf{x}_{j+\frac{1}{2}} - \mathbf{x}_j$ the previous inequation becomes

$$|V_{jr}| > \Big\|\mathbf{x}_j - \frac{1}{2}(\mathbf{x}_{j+\frac{1}{2}} + \mathbf{x}_{j-\frac{1}{2}})\Big\|_2 \Big\|\mathbf{x}_{j+\frac{1}{2}} - \mathbf{x}_{j-\frac{1}{2}}\Big\|_2$$

Doing the same for (51) leads to

$$|\widehat{V}_{jr}| > \Big\|\mathbf{x}_{j+\frac{1}{2}} - \frac{1}{2}(\mathbf{x}_{j+} + \mathbf{x}_{j+1})\Big\|_2 \Big\|\mathbf{x}_{j+1} - \mathbf{x}_j\Big\|_2$$

Therefore we obtain (52) and (53), the conditions on the mesh so that β_r is positive thus invertible. □

Remark 11. When considering non periodic boundary conditions, the problem is ill-posed if there is only one cell in the angles of the mesh (as we can see on Figure 9), then the matrix $\beta_r = \mathbf{C}_{jr} \otimes (\mathbf{x}_r - \mathbf{x}_j)$ is a rank one matrix, therefore is is non invertible. It will be necessary to study specifically the corners of the mesh, for Neumann, Dirichlet and mixed boundary conditions.

4 Properties of the scheme

4.1 Uniqueness of the solution

In order to prove the uniqueness of the discrete solution, we want for the matrix of the problem to be invertible.

Proposition 9. *Let us consider the discrete diffusion scheme (44) with homogeneous Dirichlet boundary conditions.*

Let us make the assumption that $\kappa_r\beta_r$ is positive:

$$\forall\ \mathbf{x} \in \Omega,\ \langle \mathbf{x}, \kappa_r\beta_r\mathbf{x}\rangle > 0 \quad \forall r \in \mathcal{V} \tag{54}$$

Under the condition that $\mathbf{u}_r = 0 \quad \forall\ r \in \mathcal{V} \Leftrightarrow p_j = 0 \quad \forall\ j \in \mathcal{C}$, the diffusion matrix M is invertible and the discrete diffusion problem admits a unique solution.

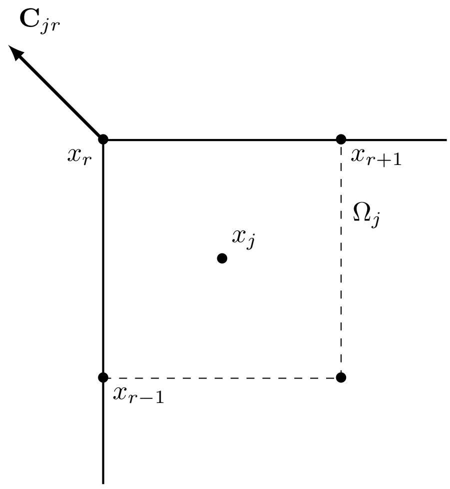


Figure 9: Construction of $\mathbf{C}_{jr}$ on a corner of Ω

Proof. Let us consider $P^TMP = 0$.

In the first place, one has

$$(MP)_j = \sum_{k\in\mathcal{C}} M_{jk}p_k = \sum_{r\in\mathcal{V}_j} \langle \kappa_r \mathbf{u}_r, \mathbf{C}_{jr}\rangle$$

Then

$$\begin{aligned} PMP &= \sum_{j\in\mathcal{C}} p_j \sum_{k\in\mathcal{C}} M_{jk}p_k = \sum_{j\in\mathcal{C}} p_j \sum_{r\in\mathcal{V}_j} \langle \kappa_r \mathbf{u}_r, \mathbf{C}_{jr}\rangle \\ &= \sum_{r\in\mathcal{V}} \langle \kappa_r \mathbf{u}_r, \sum_{j\in\mathcal{C}_r} p_j \mathbf{C}_{jr}\rangle = -\sum_{r\in\mathcal{V}} \langle \kappa_r \mathbf{u}_r, \beta_r \mathbf{u}_r\rangle \end{aligned}$$

Thanks to the assumption (54), we know that $\kappa_r\beta_r$ is positive, hence $\langle \kappa_r\beta_r\mathbf{u}_r, \mathbf{u}_r\rangle \geq 0 \quad \forall\, r \in \mathcal{V}$. Therefore $P^TMP = 0 \;\Leftrightarrow\; u_r = 0 \quad \forall\, r \in \mathcal{V}$.

Knowing that the matrix M is invertible under the condition that $P^TMP = 0$ if and only if $P = \mathbf{0}$ (ie. $p_j = 0 \quad \forall j \in \mathcal{C}$), one can deduce that $\mathbf{u}_r = 0 \quad \forall\, r \in \mathcal{V} \;\Leftrightarrow\; p_j = 0 \quad \forall\, j \in \mathcal{C}$ implies that M is invertible. And that ends the proof.

□

Let us now study the conditions such that

$$\mathbf{u}_r = 0 \quad \forall\, r \in \mathcal{V} \Leftrightarrow \begin{cases} \sum_{j\in\mathcal{C}_r} p_j \mathbf{C}_{jr} = 0 & \forall\, r \in \mathcal{V}^i \\ \sum_{j\in\mathcal{C}_r} (h_r - p_j)\mathbf{C}_{jr} = 0 & \forall\, r \in \mathcal{V}^{b,c} \end{cases}$$

Since we consider homogeneous Dirichlet boundary conditions, $h_r = 0$ for all $r \in \mathcal{V}^{b,c}$, then we are left with $\sum_{j\in\mathcal{C}_r} p_j \mathbf{C}_{jr} = 0 \quad \forall\, r \in \mathcal{V}^{b,c}$.

- In dimension 1, let us number the vertices from $\mathbf{x}_{r_0}$ to $\mathbf{x}_{r_N}$ and the cells from Ω_{j_1} to Ω_{j_N}. We know that $\mathbf{u}_r = 0$ for $r \in \{r_0, \ldots, r_N\}$ and we want to determine p_j, $j \in \{j_1, \ldots, j_N\}$. Since in dimension 1, $\mathbf{C}_{jr} = \pm 1$, one has $\mathbf{u}_{r_k} = p_{j_{k+1}} - p_{j_k} = 0$, for all $k \in \{1, \ldots, N-1\}$ and $\mathbf{u}_{r_0} = p_{j_1} = 0$, $\mathbf{u}_{r_N} = p_{j_N} = 0$. A simple induction gives $p_j = 0$ for all $j \in \{1, \ldots, N\}$.

- In dimension 2, let us consider the structured mesh presented in Figure 10. One can see that the point $\mathbf{x}_{r_1}$ is linked to two cells, therefore the system $\sum_{j\in\mathcal{C}_{r_1}} p_j \mathbf{C}_{jr_1} = 0$ has two equations for two unknowns, thus it has a unique solution which is $(p_{j_1}, p_{j_2}) = (0,0)$. Then we can do the same thing with the vertice $\mathbf{x}_{r_2}$, linked to four cells with two of them for which p_j is known. Once again we can deduce $(p_{j_3}, p_{j_4}) = (0,0)$. This method can be extended to the whole domain, in every direction. This result is also true for non Cartesian structured meshes.

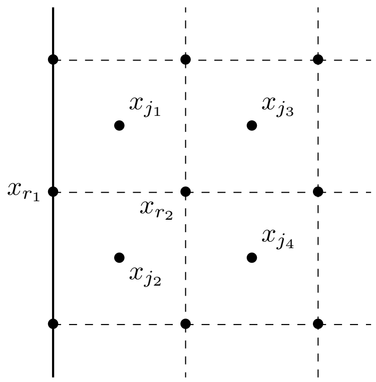


Figure 10: Structured mesh

- This method can not be extended to most of the unstructured meshed, especially with triangles. Indeed for a vertice $\mathbf{x}_r$ if the number of neighbor cells for which p_j is unknown is greater than 2, then it is impossible to assert that $p_j = 0 \quad \forall\, j \in \mathcal{C}$.

4.2 Exactness of the scheme for linear solutions

In this section, we will study the exactness of the scheme with Neumann and Dirichlet boundary conditions for different types of solutions. We do not study it for periodic boundary conditions because we haven't implemented it but it can be easily extended to this case.

4.2.1 Neumann boundary condition

Proposition 10. *Let us consider the scheme (43), with no source ($f_j = 0,\ \forall\, j \in \mathcal{C}$) and with homogeneous Neumann boundary conditions ($g_r = 0,\ \forall\, r \in \mathcal{V}^{b,c}$).*

Then the scheme is exact for a constant solution $p(\mathbf{x}) = C$, for all $j \in \mathcal{C}$.

Proof. Let us write the scheme (43) with no source and homogeneous boundary conditions:

$$\begin{cases} |\Omega_j| \dfrac{p_j^{n+1} - p_j^n}{\Delta t} - \Big(\sum\limits_{r\in\mathcal{V}_j^i} \langle \kappa_r \mathbf{u}_r^{n+1}, \mathbf{C}_{jr}\rangle + \sum\limits_{r\in\mathcal{V}_j^b} \langle \kappa_r \mathbf{u}_r^{n+1}, (\widehat{I_d} - \mathbf{n}_r \otimes \mathbf{n}_r)\mathbf{C}_{jr}\rangle \Big) = 0 \\ -\beta_r \mathbf{u}_r^{n+1} = \sum\limits_{j\in\mathcal{C}_r} p_j^{n+1} \mathbf{C}_{jr} \quad \forall\, r \in \mathcal{V}^i \\ \beta_r \mathbf{u}_r^{n+1} = \sum\limits_{j\in\mathcal{C}_r} (p_r^{n+1} - p_j^{n+1}) \mathbf{C}_{jr} \quad \forall\, r \in \mathcal{V}^b \end{cases} \tag{55}$$

with $p_r^{n+1} = \Big(\sum\limits_{j\in\mathcal{C}_r} \theta_{jr} \Big)^{-1} \sum\limits_{j\in\mathcal{C}_r} \theta_{jr} p_j^{n+1}$ and $\theta_{jr} = \langle \beta_r^{-1} \mathbf{C}_{jr}, \mathbf{v}_r \rangle$.

Let us now consider that p is a constant at the time t^{n+1}, which means that we inject $p_j^{n+1} = C$ in the scheme. Then the second equation of (55) becomes, $\forall\, r \in \mathcal{V}^i$

$$\begin{aligned} -\beta_r \mathbf{u}_r^{n+1} &= \sum_{j\in\mathcal{C}_r} C\ \mathbf{C}_{jr} = C \sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} = 0 \\ \Leftrightarrow\quad \mathbf{u}_r^{n+1} &= 0 \quad \forall\, r \in \mathcal{V}^i \end{aligned}$$

due to Proposition 6.

On the other side, when $r \in \mathcal{V}^b$, knowing that $p_j^{n+1} = C$ one has

$$p_r^{n+1} = \Big(\sum_{j\in\mathcal{C}_r} \theta_{jr}\Big)^{-1} \sum_{j\in\mathcal{C}_r} \theta_{jr} C = C$$

Hence the third equation of (55) becomes

$$\begin{aligned} -\beta_r \mathbf{u}_r^{n+1} &= \sum_{j\in\mathcal{C}_r} (C - C)\ \mathbf{C}_{jr} = 0 \\ \Leftrightarrow\quad \mathbf{u}_r^{n+1} &= 0 \quad \forall\, r \in \mathcal{V}^b \end{aligned}$$

One can now replace the values of $\mathbf{u}_r^{n+1}$ inside the first equation of (55) and obtain

$$|\Omega_j| \frac{p_j^{n+1} - p_j^n}{\Delta t} = 0$$

Then the scheme is constant in time and a constant is an exact solution.

□

Proposition 11. *Let us consider the scheme (43), with no source ($f_j = 0,\ \forall\, j \in \mathcal{C}$). Let us chose for the boundary condition $g(\mathbf{x}) = \langle \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \kappa \mathbf{n}_\Gamma \rangle$. Assume that κ is a constant.*

Then the scheme is exact for a linear solution $p(\mathbf{x}) = \alpha x + \beta y + \gamma$.

Proof. Knowing that the scheme is linear, we can split the solution p defined above in the base $(1, x, y)$. We already proved in Proposition 10 that the scheme is exact when p is a constant. Thus we are left with $p(\mathbf{x}) = \alpha x$ (by symmetry the proof will be the same for $p(\mathbf{x}) = \beta y$). The corresponding boundary condition is $g_r = \langle \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \kappa_r \mathbf{n}_r \rangle$ (respectively $g_r = \langle \begin{pmatrix} 0 \\ \beta \end{pmatrix}, \kappa_r \mathbf{n}_r \rangle$).

Let us now inject the linear solution $p(\mathbf{x}) = \alpha x$ into the scheme. It means for all $j \in \mathcal{C}$, one can replace p_j^{n+1} with αx_j.

Let us consider the second equation of (43) with $p_j^{n+1} = \alpha x_j$:

$$-\beta_r \mathbf{u}_r^{n+1} = \sum_{j\in\mathcal{C}_r} \alpha x_j \mathbf{C}_{jr} \quad \forall\, r \in \mathcal{V}^i \tag{56}$$

This is a system of two equations for two unknowns, therefore it has a unique solution. Let us now calculate $-\beta_r \mathbf{u}_r^{n+1}$ with $\mathbf{u}_r^{n+1} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}$. Thanks to the definition of β_r one has

$$\begin{aligned} -\beta_r \begin{pmatrix} \alpha \\ 0 \end{pmatrix} &= -\sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} \otimes (\mathbf{x}_r - \mathbf{x}_j) \begin{pmatrix} \alpha \\ 0 \end{pmatrix} = \sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} \otimes \mathbf{x}_j \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \\ &= \sum_{j\in\mathcal{C}_r} \begin{pmatrix} C_{jr}^{(1)} x_j & C_{jr}^{(1)} y_j \\ C_{jr}^{(2)} x_j & C_{jr}^{(2)} y_j \end{pmatrix} \begin{pmatrix} \alpha \\ 0 \end{pmatrix} = \sum_{j\in\mathcal{C}_r} (C_{jr}^{(1)} x_j + C_{jr}^{(2)} x_j)\alpha \\ &= \sum_{j\in\mathcal{C}_r} \alpha x_j \mathbf{C}_{jr} \end{aligned}$$

Thus $\mathbf{u}_r^{n+1} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}$ is the unique solution of this system for all $r \in \mathcal{V}^i$.

Let us now study the problem at the boundary of the domain: $r \in \mathcal{V}^b$. The third equation of (43), associated with the Neumann boundary condition gives the following system

$$\begin{cases} \beta_r \mathbf{u}_r = \sum_{j \in \mathcal{C}_r} (p_r - p_j)\mathbf{C}_{jr} \\ \langle \mathbf{u}_r, \kappa_r \mathbf{n}_r \rangle = g_r \end{cases} \tag{57}$$

with $g_r = \left\langle \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \kappa_r \mathbf{n}_r \right\rangle$ and $p_j = \alpha x_j$ as stated in the beginning of the proof.

Knowing that β_r is invertible for all $r \in \mathcal{V}^b$ and this system has 3 equations for 3 unknowns (p_r of dimension 1 and $\mathbf{u}_r$ of dimension 2), it admits a unique solution for the couple $(p_r, \mathbf{u}_r)$, for all $r \in \mathcal{V}^b$.

Let us inject $\mathbf{u}_r = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}$ into the system (57).

The second equation gives

$$\left\langle \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \kappa_r \mathbf{n}_r \right\rangle = g_r$$

which is true by definition of g_r.

Let us now consider the left term of the first equation of the system (57), with $\mathbf{u} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}$:

$$\begin{aligned} \beta_r \begin{pmatrix} \alpha \\ 0 \end{pmatrix} &= \sum_{j \in \mathcal{C}_r} \mathbf{C}_{jr} \otimes (\mathbf{x}_r - \mathbf{x}_j) \begin{pmatrix} \alpha \\ 0 \end{pmatrix} = \sum_{j \in \mathcal{C}_r} \langle (\mathbf{x}_r - \mathbf{x}_j), \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \rangle \mathbf{C}_{jr} \\ &= \sum_{j \in \mathcal{C}_r} \alpha (x_r - x_j) \mathbf{C}_{jr} = \sum_{j \in \mathcal{C}_r} (\alpha x_r - p_j) \mathbf{C}_{jr} \end{aligned}$$

The first equation of (57) is verified if $p_r = \alpha x_r$.

Knowing that the system has a unique solution, $(p_r, \mathbf{u}_r) = (\alpha x_r, \begin{pmatrix} \alpha & 0 \end{pmatrix}^t)$ is the unique solution of this system, for all $r \in \mathcal{V}^b$.

We obtained $\mathbf{u}_r^{n+1} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}$ for all $r \in \mathcal{V}^{i,b}$, hence we can inject this value inside the first equation of the scheme (43).

$$|\Omega_j| \frac{p_j^{n+1} - p_j^n}{\Delta t} - \Big(\sum_{r \in \mathcal{V}_j^i} \langle \kappa_r \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \mathbf{C}_{jr} \rangle + \sum_{r \in \mathcal{V}_j^b} \langle \kappa_r \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, (\widehat{I}_d - \mathbf{n}_r \otimes \mathbf{n}_r) \mathbf{C}_{jr} \rangle \Big) = B_j^{n+1}$$

with $B_j^{n+1} = \sum_{r \in \mathcal{V}_j^b} \langle \mathbf{C}_{jr}, \mathbf{n}_r \rangle g_r + \sum_{r \in \mathcal{V}_j^c} \frac{1}{2} (\ell_{l_1} g_{l_1} + \ell_{l_2} g_{l_2})$.

Let us assume that κ_r is a constant, then thanks to Proposition 6 one has

$$\begin{aligned} & \sum_{r \in \mathcal{V}_j} \langle \kappa \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \mathbf{C}_{jr} \rangle = 0 \\ \Leftrightarrow \quad & \sum_{r \in \mathcal{V}_j^{i,b}} \langle \kappa \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \mathbf{C}_{jr} \rangle + \sum_{r \in \mathcal{V}_j^c} \langle \kappa \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \mathbf{C}_{jr} \rangle = 0 \end{aligned}$$

Moreover

$$\begin{aligned}
\sum_{r\in\mathcal{V}_j^b}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\rangle &= \sum_{r\in\mathcal{V}_j^b}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},\mathbf{C}_{jr}\rangle-\sum_{r\in\mathcal{V}_j^b}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},(\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\rangle\\
&= \sum_{r\in\mathcal{V}_j^b}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},\mathbf{C}_{jr}\rangle-\sum_{r\in\mathcal{V}_j^b}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},\langle\mathbf{n}_r,\mathbf{C}_{jr}\rangle\mathbf{n}_r\rangle\\
&= \sum_{r\in\mathcal{V}_j^b}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},\mathbf{C}_{jr}\rangle-\sum_{r\in\mathcal{V}_j^b}\langle\mathbf{n}_r,\mathbf{C}_{jr}\rangle\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},\mathbf{n}_r\rangle\\
&= \sum_{r\in\mathcal{V}_j^b}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},\mathbf{C}_{jr}\rangle-\sum_{r\in\mathcal{V}_j^b}\langle\mathbf{C}_{jr},\mathbf{n}_r\rangle g_r
\end{aligned}$$

Then

$$\begin{aligned}
\sum_{r\in\mathcal{V}_j^i}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},\mathbf{C}_{jr}\rangle+\sum_{r\in\mathcal{V}_j^b}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\rangle &= \sum_{r\in\mathcal{V}_j^{i,b}}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},\mathbf{C}_{jr}\rangle-\sum_{r\in\mathcal{V}_j^b}\langle\mathbf{C}_{jr},\mathbf{n}_r\rangle g_r\\
&= -\sum_{r\in\mathcal{V}_j^c}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},\mathbf{C}_{jr}\rangle-\sum_{r\in\mathcal{V}_j^b}\langle\mathbf{C}_{jr},\mathbf{n}_r\rangle g_r
\end{aligned}$$

Finally, we are left with

$$|\Omega_j|\frac{p_j^{n+1}-p_j^n}{\Delta t}+\sum_{r\in\mathcal{V}_j^c}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},\mathbf{C}_{jr}\rangle+\sum_{r\in\mathcal{V}_j^b}\langle\mathbf{C}_{jr},\mathbf{n}_r\rangle g_r=\sum_{r\in\mathcal{V}_j^b}\langle\mathbf{C}_{jr},\mathbf{n}_r\rangle g_r+\sum_{r\in\mathcal{V}_j^c}\frac{1}{2}(\ell_{l_1}g_{l_1}+\ell_{l_2}g_{l_2})$$

Thus the solution is constant in time (ie. $|\Omega_j|\frac{p_j^{n+1}-p_j^n}{\Delta t}=0$) if

$$\begin{aligned}
\sum_{r\in\mathcal{V}_j^c}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},\mathbf{C}_{jr}\rangle &= \sum_{r\in\mathcal{V}_j^c}\frac{1}{2}(\ell_{l_1}g_{l_1}+\ell_{l_2}g_{l_2})\\
&= \sum_{r\in\mathcal{V}_j^c}\frac{1}{2}(\ell_{l_1}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},\mathbf{n}_{l_1}\rangle+\ell_{l_2}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},\mathbf{n}_{\Gamma_2}\rangle)\\
&= \sum_{r\in\mathcal{V}_j^c}\langle\kappa\begin{pmatrix}\alpha\\0\end{pmatrix},\frac{1}{2}(\ell_{l_1}\mathbf{n}_{l_1}+\ell_{l_2}\mathbf{n}_{l_2})\rangle
\end{aligned}$$

By definition of $\mathbf{C}_{jr}$, on the corner one has $\mathbf{C}_{jr}=\frac{1}{2}(\ell_{l_1}\mathbf{n}_{l_1}+\ell_{l_2}\mathbf{n}_{l_2})$.

Remark 12. The approximation $g_{r_{1,2}}\approx g_{l_{1,2}}$ we made in Section 3.2.1 is exact in this case since the boundary condition g is a constant for the linear solution. Therefore this approximation doesn't compromise the exactness of the scheme.

In conclusion, the solution is constant in time, which means that the scheme is exact for $p(\mathbf{x})=\alpha x$. As said in the beginning of the proof, the method is the same for $p(\mathbf{x})=\beta y$, and we have already proved that the scheme is exact for a constant solution.
Therefore thanks to the linearity of the scheme, it is exact for any linear function $p(\mathbf{x})=\alpha x+\beta y+\gamma$. □

4.2.2 Dirichlet boundary condition

Proposition 12. *Let us consider the scheme (44), with no source ($f_j=0,\ \forall\ j\in\mathcal{C}$) and with constant Dirichlet boundary conditions ($h_r=C,\ \forall\ r\in\mathcal{V}^{b,c}$).*

Then the scheme is exact for a constant solution $p(\mathbf{x})=C$, for all $j\in\mathcal{C}$.

Proof. Let us write the scheme (44) with no source and with $h_r = C,\ \forall\, r \in \mathcal{V}^{b,c}$

$$\begin{cases} |\Omega_j| \dfrac{p_j^{n+1} - p_j^n}{\Delta t} - \sum_{r \in \mathcal{V}_j} \langle \kappa_r \mathbf{u}_r^{n+1}, \mathbf{C}_{jr} \rangle = 0 \\ \beta_r \mathbf{u}_r^{n+1} = - \sum_{j \in \mathcal{C}_r} p_j^{n+1} \mathbf{C}_{jr} & \text{if } r \in \mathcal{V}_j^i \\ \beta_r \mathbf{u}_r^{n+1} = \sum_{j \in \mathcal{C}_r} (C - p_j^{n+1}) \mathbf{C}_{jr} & \text{if } r \in \mathcal{V}_j^b \\ \beta_r^c \mathbf{u}_r^{n+1} = \sum_{j \in \mathcal{C}_r} (C - p_j^{n+1}) \mathbf{C}_{jr} & \text{if } r \in \mathcal{V}_j^c \end{cases}$$

Let us now inject the constant solution $p_j^{n+1} = C$ in the second, third and fourth equations of the system. Then on $\mathcal{V}^b$ and $\mathcal{V}^c$ we immediately obtain that $\mathbf{u}_r^{n+1} = 0$ and thanks to Proposition 6 one has $\mathbf{u}_r^{n+1} = 0\ \forall\, r \in \mathcal{V}^i$.

Then one can replace $\mathbf{u}_r^{n+1}$ by its value inside the first equation of the system and obtain:

$$|\Omega_j| \frac{p_j^{n+1} - p_j^n}{\Delta t} = 0$$

Which means if one injects a constant solution in the scheme with the correct boundary conditions, it stays constant. And that ends the proof.

□

Proposition 13. *Let us consider the scheme (44), with no source ($f_j = 0,\ \forall\, j \in \mathcal{C}$). Let us chose for the boundary condition $h(\mathbf{x}) = \alpha x + \beta y + \gamma$. Assume that the mesh respects the following statement: there is no corner with a unique cell (ie.$\mathcal{V}^c = \emptyset$). Assume that κ is a constant.*

Then the scheme is exact for a linear solution $p(\mathbf{x}) = \alpha x + \beta y + \gamma$ if the mesh respects the condition mentioned above.

Proof. In the same way as Proposition 11, one can decompose the solution and the boundary condition in the base $(1, x, y)$. It gives: $p(\mathbf{x}) = \alpha x$ and $h(\mathbf{x}) = \alpha x$.

One can rewrite the scheme (44) with the source and boundary conditions of Proposition 13, and according to the condition on the mesh with the discretization of h: $h_r = \alpha x_r$

$$\begin{cases} |\Omega_j| \dfrac{p_j^{n+1} - p_j^n}{\Delta t} - \sum_{r \in \mathcal{V}_j} \langle \kappa_r \mathbf{u}_r^{n+1}, \mathbf{C}_{jr} \rangle = 0 \\ \beta_r \mathbf{u}_r^{n+1} = - \sum_{j \in \mathcal{C}_r} p_j^{n+1} \mathbf{C}_{jr} & \text{if } r \in \mathcal{V}_j^i \\ \beta_r \mathbf{u}_r^{n+1} = \sum_{j \in \mathcal{C}_r} (\alpha x_r - p_j^{n+1}) \mathbf{C}_{jr} & \text{if } r \in \mathcal{V}_j^b \end{cases}$$

Let us now inject the linear solution into the scheme, by replacing p_j^{n+1} with αx_j.

From the second equation, one gets

$$\beta_r \mathbf{u}_r^{n+1} = - \sum_{j \in \mathcal{C}_r} p_j^{n+1} \mathbf{C}_{jr} \quad \text{if } r \in \mathcal{V}_j^i$$

which is the same as in Proposition 11. Therefore, using the same principle of proof we obtain that this system of two equations for two unknowns admits a unique solution, which is $\mathbf{u}_r^{n+1} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}$, for all $r \in \mathcal{V}^i$.

The third equation of (44) writes

$$\beta_r \mathbf{u}_r^{n+1} = \sum_{j \in \mathcal{C}_r} (\alpha x_r - \alpha x_j)\mathbf{C}_{jr} \quad \text{if } r \in \mathcal{V}_j^b$$

Knowing that β_r is invertible for $r \in \mathcal{V}_j^b$, the equation admits a unique solution for $\mathbf{u}_r^{n+1}$.

Let $\mathbf{u}_r^{n+1} \begin{pmatrix} \alpha \\ 0 \end{pmatrix}$. Then the left hand side of the equation gives

$$\begin{aligned} \beta_r \begin{pmatrix} \alpha \\ 0 \end{pmatrix} &= \sum_{j \in \mathcal{C}_r} \mathbf{C}_{jr} \otimes (\mathbf{x}_r - \mathbf{x}_j) \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \\ &= \sum_{j \in \mathcal{C}_r} \langle \mathbf{x}_r - \mathbf{x}_j, \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \rangle \mathbf{C}_{jr} \\ &= \sum_{j \in \mathcal{C}_r} \alpha(x_r - x_j)\mathbf{C}_{jr} \end{aligned}$$

Hence the third equation of the scheme is verified by the solution $\mathbf{u}_r^{n+1} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}$, $\forall\, r \in \mathcal{V}^b$.

Thanks to the condition on the mesh that states $\mathcal{V}^c = \emptyset$, one has $\mathcal{V} = \mathcal{V}^i \cup \mathcal{V}^b$. Thus

$$\mathbf{u}_r^{n+1} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \quad \forall\, r \in \mathcal{V}$$

We can replace this value inside the first equation of the scheme (44)

$$|\Omega_j| \frac{p_j^{n+1} - p_j^n}{\Delta t} - \sum_{r \in \mathcal{V}_j} \langle \kappa_r \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \mathbf{C}_{jr} \rangle = 0$$

Once again under the condition that $\kappa_r = \kappa$ is a constant, the solution is constant in time when we inject a linear function in the scheme with the corresponding Dirichlet boundary condition. The result is the same for the two other elements of the base $(1, x, y)$. Thus the scheme is exact for a linear solution. □

Remark 13. If the condition on the mesh is not respected, which means that $\mathcal{V}^c \neq \emptyset$, then the scheme is not exact for a linear solution.

Indeed on the corners with a unique cell, we approached β_r with β_r^c so that it becomes invertible. But then if we try to test the fourth equation of the scheme (44) with $\mathbf{u}_r^{n+1} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}$ in writes

$$\begin{aligned} \beta_r^c \begin{pmatrix} \alpha \\ 0 \end{pmatrix} &= (\mathbf{C}_{j_1 r} \otimes (\mathbf{x}_r - \mathbf{x}_{j_1}) + \mathbf{C}_{j_2 r} \otimes (\mathbf{x}_r - \mathbf{x}_{j_2})) \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \\ &= \langle \mathbf{x}_r - \mathbf{x}_{j_1}, \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \rangle \mathbf{C}_{j_1 r} + \langle \mathbf{x}_r - \mathbf{x}_{j_2}, \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \rangle \mathbf{C}_{j_2 r} \\ &= \alpha(x_r - x_{j_1})\mathbf{C}_{j_1 r} + \alpha(x_r - x_{j_2})\mathbf{C}_{j_2 r} \\ &= \alpha x_r \mathbf{C}_{jr} - \alpha(x_{j_1}\mathbf{C}_{j_1 r} + x_{j_2}\mathbf{C}_{j_2 r}) \end{aligned}$$

since by definition $\mathbf{C}_{jr} = \mathbf{C}_{j_1 r} + \mathbf{C}_{j_2 r}$.

In order to retrieve the right hand side of the fourth equation of the scheme (44), we need

$$x_{j_1}\mathbf{C}_{j_1 r} + x_{j_2}\mathbf{C}_{j_2 r} = \sum_{j \in \mathcal{V}^c} x_j \mathbf{C}_{jr} \Leftrightarrow x_{j_1} = x_{j_2} = x_j$$

because in a corner with a unique cell there is only one index $j \in \mathcal{V}^c$, therefore the sum over j disappears.

This condition is impossible because it leads to $\beta_r^c = \beta_r$ which is not invertible, then the scheme is ill-posed. Thus the solution is not exact for a linear solution as soon as there are some unique cells at the corners of the mesh.

4.3 Stability

In order to study the stability, consistency and convergence of the scheme, let us consider the problem with no source: $f(\mathbf{x},t) = 0$ and homogeneous boundary conditions: $g(\mathbf{x}) = 0$ and $h(\mathbf{x}) = 0$. The generalization to the case where $f \neq 0$ and non-homogeneous boundary conditions is straightforward.

This section as well as the two next ones (consistency and convergence) are inspired by the class of B. Després [10].

Proposition 14 (L^2 Stability of the periodic scheme). *Let us keep on with the assumption that $\kappa_r\beta_r$ is positive:*

$$\forall\ \mathbf{x} \in \Omega,\ \langle \mathbf{x}, \kappa_r\beta_r\mathbf{x}\rangle > 0 \quad \forall r \in \mathcal{V} \tag{58}$$

Then the implicit diffusion scheme (42) with no source ($f(\mathbf{x},t) = 0$) and periodic boundary conditions is unconditionally stable in L^2 norm, which means the following inequality stands

$$\|p_h^{n+1}\|_{L^2(\Omega)} \leq \|p_h^n\|_{L^2(\Omega)} \quad \forall\ \Delta t \geq 0 \tag{59}$$

Remark 14. Saying that the scheme is unconditionally stable means that no CFL condition is needed on Δt for the scheme to be stable.

Proof. The scheme (42) rewrites:

$$\begin{cases} p_j^{n+1} - \dfrac{\Delta t}{|\Omega_j|}\displaystyle\sum_{r\in\mathcal{V}_j}\langle \mathbf{u}_r^{n+1}, \kappa_r\mathbf{C}_{jr}\rangle = p_j^n \\ \qquad -\beta_r\mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r} p_j^{n+1}\mathbf{C}_{jr} \end{cases} \tag{60}$$

We can multiply the first equation of (60) by p_j^{n+1}, and sum it over $j \in \mathcal{C}$:

$$\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^{n+1})^2 - \Delta t\sum_{j\in\mathcal{C}}\left(p_j^{n+1}\sum_{r\in\mathcal{V}_j}\langle \mathbf{u}_r^{n+1}, \kappa_r\mathbf{C}_{jr}\rangle\right) = \sum_{j\in\mathcal{C}}|\Omega_j|p_j^{n+1}p_j^n$$

By permuting the sums on r and j, and using (15), we obtain:

$$\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^{n+1})^2 - \Delta t\sum_{r\in\mathcal{V}}\langle \mathbf{u}_r^{n+1}, \kappa_r\sum_{j\in\mathcal{C}_r}p_j^{n+1}\mathbf{C}_{jr}\rangle = \frac{1}{2}\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^{n+1})^2 + \frac{1}{2}\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^n)^2 - \frac{1}{2}\sum_{j\in\mathcal{C}_r}|\Omega_j|(p_j^{n+1} - p_j^n)^2$$

Using the second equation of (60) the equation becomes

$$\frac{1}{2}\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^{n+1})^2 + \Delta t\sum_{r\in\mathcal{V}}\langle \mathbf{u}_r^{n+1}, \kappa_r\beta_r\mathbf{u}_r^{n+1}\rangle + \frac{1}{2}\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^{n+1} - p_j^n)^2 = \frac{1}{2}\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^n)^2$$

Obviously,

$$\frac{1}{2}\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^{n+1} - p_j^n)^2 \geq 0$$

Thanks to the assumption (58), we have

$$\Delta t\sum_{r\in\mathcal{V}}\langle \mathbf{u}_r^{n+1}, \kappa_r\beta_r\mathbf{u}_r^{n+1}\rangle \geq 0$$

Therefore

$$\|p_h^{n+1}\|_{L^2(\Omega)} \leq \|p_h^n\|_{L^2(\Omega)} \quad \forall\ \Delta t \geq 0$$

and this ends the proof.

□

Proposition 15 (L^2 Stability with Neumann boundary conditions)**.** *Let us make the two following assumptions:*

$$\langle \mathbf{x}, \kappa_r \beta_r \mathbf{x} \rangle > 0 \quad \forall r \in \mathcal{V}^i \tag{61}$$

$$\langle \mathbf{x}, \kappa_r (\widehat{I}_d - \mathbf{n}_r \otimes \mathbf{n}_r) \beta_r \mathbf{x} \rangle > 0 \quad \forall r \in \mathcal{V}^b \tag{62}$$

Then the implicit diffusion scheme (43) with no source ($f(\mathbf{x}, t) = 0$) and homogeneous Neumann boundary conditions ($g(\mathbf{x}) = 0$) is unconditionally stable in L^2 norm, which means the following inequality stands

$$\|p_h^{n+1}\|_{L^2(\Omega)} \leq \|p_h^n\|_{L^2(\Omega)} \quad \forall\ \Delta t \geq 0 \tag{63}$$

Proof. The scheme (43) rewrites:

$$\begin{cases} |\Omega_j| p_j^{n+1} - \Delta t \Big(\displaystyle\sum_{r \in \mathcal{V}_j^i} \langle \kappa_r \mathbf{u}_r^{n+1}, \mathbf{C}_{jr} \rangle + \sum_{r \in \mathcal{V}_j^b} \langle \kappa_r \mathbf{u}_r^{n+1}, (\widehat{I}_d - \mathbf{n}_r \otimes \mathbf{n}_r) \mathbf{C}_{jr} \rangle \Big) = |\Omega_j| p_j^n \\ -\beta_r \mathbf{u}_r^{n+1} = \displaystyle\sum_{j \in \mathcal{C}_r} p_j^{n+1} \mathbf{C}_{jr} \quad \forall r \in \mathcal{V}^i \\ \beta_r \mathbf{u}_r^{n+1} = \displaystyle\sum_{j \in \mathcal{C}_r} (p_r^{n+1} - p_j^{n+1}) \mathbf{C}_{jr} \quad \forall r \in \mathcal{V}^b \end{cases} \tag{64}$$

We can multiply the first equation of (64) by p_j^{n+1}, and sum it over $j \in \mathcal{C}$:

$$\sum_{j \in \mathcal{C}} |\Omega_j| (p_j^{n+1})^2 - \Delta t \sum_{j \in \mathcal{C}} p_j^{n+1} \Big(\sum_{r \in \mathcal{V}_j^i} \langle \kappa_r \mathbf{u}_r^{n+1}, \mathbf{C}_{jr} \rangle + \sum_{r \in \mathcal{V}_j^b} \langle \kappa_r \mathbf{u}_r^{n+1}, (\widehat{I}_d - \mathbf{n}_r \otimes \mathbf{n}_r) \mathbf{C}_{jr} \rangle \Big) = \sum_{j \in \mathcal{C}} |\Omega_j| p_j^{n+1} p_j^n$$

By permuting the sums on r and j, and using (15), we obtain:

$$\begin{aligned} \sum_{j \in \mathcal{C}} |\Omega_j| (p_j^{n+1})^2 - \Delta t \sum_{r \in \mathcal{V}^i} \langle \kappa_r \mathbf{u}_r^{n+1}, \sum_{j \in \mathcal{C}} p_j^{n+1} \mathbf{C}_{jr} \rangle - \Delta t \sum_{r \in \mathcal{V}^b} \langle \kappa_r \mathbf{u}_r^{n+1}, (\widehat{I}_d - \mathbf{n}_r \otimes \mathbf{n}_r) \sum_{j \in \mathcal{C}} p_j^{n+1} \mathbf{C}_{jr} \rangle \\ = \frac{1}{2} \sum_{j \in \mathcal{C}} |\Omega_j| (p_j^{n+1})^2 + \frac{1}{2} \sum_{j \in \mathcal{C}} |\Omega_j| (p_j^n)^2 - \frac{1}{2} \sum_{j \in \mathcal{C}} |\Omega_j| (p_j^{n+1} - p_j^n)^2 \end{aligned}$$

Using the second equation of (64) the equation becomes

$$\begin{aligned} &\frac{1}{2} \sum_{j \in \mathcal{C}} |\Omega_j| (p_j^{n+1})^2 + \Delta t \sum_{r \in \mathcal{V}^i} \langle \kappa_r \mathbf{u}_r^{n+1}, \beta_r \mathbf{u}_r^{n+1} \rangle + \Delta t \sum_{r \in \mathcal{V}^b} \langle \kappa_r \mathbf{u}_r^{n+1}, (\widehat{I}_d - \mathbf{n}_r \otimes \mathbf{n}_r) \beta_r \mathbf{u}_r^{n+1} \mathbf{C}_{jr} \rangle \\ &-\Delta t \sum_{r \in \mathcal{V}^b} \langle \kappa_r \mathbf{u}_r^{n+1}, (\widehat{I}_d - \mathbf{n}_r \otimes \mathbf{n}_r) \sum_{j \in \mathcal{C}_r} p_r^{n+1} \mathbf{C}_{jr} \rangle + \frac{1}{2} \sum_{j \in \mathcal{C}} |\Omega_j| (p_j^{n+1} - p_j^n)^2 = \frac{1}{2} \sum_{j \in \mathcal{C}} |\Omega_j| (p_j^n)^2 \end{aligned} \tag{65}$$

By definition of $\mathbf{n}_r$, one has

$$\sum_{j \in \mathcal{C}_r} p_r^{n+1} \mathbf{C}_{jr} = p_r^{n+1} \sum_{j \in \mathcal{C}_r} \mathbf{C}_{jr} = p_r^{n+1} \mathbf{n}_r \Big\| \sum_{j \in \mathcal{C}_r} \mathbf{C}_{jr} \Big\|$$

Then, thanks to Proposition 38, (65) becomes

$$\begin{aligned} \frac{1}{2} \sum_{j \in \mathcal{C}} |\Omega_j| (p_j^{n+1})^2 + \Delta t \sum_{r \in \mathcal{V}^i} \langle \kappa_r \mathbf{u}_r^{n+1}, \beta_r \mathbf{u}_r^{n+1} \rangle + \Delta t \sum_{r \in \mathcal{V}^b} \langle \kappa_r \mathbf{u}_r^{n+1}, (\widehat{I}_d - \mathbf{n}_r \otimes \mathbf{n}_r) \beta_r \mathbf{u}_r^{n+1} \mathbf{C}_{jr} \rangle \\ + \frac{1}{2} \sum_{j \in \mathcal{C}} |\Omega_j| (p_j^{n+1} - p_j^n)^2 = \frac{1}{2} \sum_{j \in \mathcal{C}} |\Omega_j| (p_j^n)^2 \end{aligned} \tag{66}$$

Obviously,

$$\frac{1}{2} \sum_{j \in \mathcal{C}} |\Omega_j| (p_j^{n+1} - p_j^n)^2 \geq 0$$

And thanks to the assumptions (61) and (62), we have

$$\Delta t \sum_{r\in\mathcal{V}^i} \langle \mathbf{u}_r^{n+1}, \kappa_r\beta_r\mathbf{u}_r^{n+1}\rangle \geq 0 \quad \text{and} \quad \Delta t \sum_{r\in\mathcal{V}^b} \langle \mathbf{u}_r^{n+1}, \kappa_r(\widehat{I}_d - \mathbf{n}_r\otimes\mathbf{n}_r)\beta_r\mathbf{u}_r^{n+1}\rangle \geq 0$$

Therefore

$$\|p_h^{n+1}\|_{L^2(\Omega)} \leq \|p_h^n\|_{L^2(\Omega)},\ \forall\ \Delta t \geq 0$$

and this ends the proof.

□

Proposition 16 (L^2 Stability with Dirichlet boundary conditions)**.** *Let us make the assumption that $\kappa_r\beta_r$ and $\kappa_r\beta_r^c$ are positive:*

$$\forall\ \mathbf{x}\in\Omega,\ \langle \mathbf{x}, \kappa_r\beta_r\mathbf{x}\rangle > 0 \quad \forall r\in\mathcal{V}^{i,b} \tag{67}$$

and

$$\forall\ \mathbf{x}\in\Omega,\ \langle \mathbf{x}, \kappa_r\beta_r^c\mathbf{x}\rangle > 0 \quad \forall r\in\mathcal{V}^{c} \tag{68}$$

Then the implicit diffusion scheme (44) with no source ($f(\mathbf{x},t)=0$) and homogeneous Dirichlet boundary conditions ($h(\mathbf{x})=0$) is unconditionally stable in L^2 norm, which means the following inequality stands

$$\|p_h^{n+1}\|_{L^2(\Omega)} \leq \|p_h^n\|_{L^2(\Omega)} \quad \forall\ \Delta t \geq 0 \tag{69}$$

Proof. The scheme (44) rewrites:

$$\begin{cases} p_j^{n+1} - \dfrac{\Delta t}{|\Omega_j|}\displaystyle\sum_{r\in\mathcal{V}_j}\langle \mathbf{u}_r^{n+1}, \kappa_r\mathbf{C}_{jr}\rangle = p_j^n \\ \beta_r\mathbf{u}_r^{n+1} = -\displaystyle\sum_{j\in\mathcal{C}_r} p_j^{n+1}\mathbf{C}_{jr} \quad \forall r\in\mathcal{V}^{i,b} \\ \beta_r^c\mathbf{u}_r^{n+1} = -\displaystyle\sum_{j\in\mathcal{C}_r} p_j^{n+1}\mathbf{C}_{jr} \quad \forall r\in\mathcal{V}^{c} \end{cases} \tag{70}$$

We can multiply the first equation of (70) by p_j^{n+1}, and sum it over $j\in\mathcal{C}$:

$$\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^{n+1})^2 - \Delta t\sum_{j\in\mathcal{C}}\Big(p_j^{n+1}\sum_{r\in\mathcal{V}_j}\langle \mathbf{u}_r^{n+1}, \kappa_r\mathbf{C}_{jr}\rangle\Big) = \sum_{j\in\mathcal{C}}|\Omega_j|p_j^{n+1}p_j^n$$

By permuting the sums on r and j, and using (15), we obtain:

$$\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^{n+1})^2 - \Delta t\sum_{r\in\mathcal{V}}\langle \mathbf{u}_r^{n+1}, \kappa_r\sum_{j\in\mathcal{C}_r}p_j^{n+1}\mathbf{C}_{jr}\rangle = \frac{1}{2}\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^{n+1})^2 + \frac{1}{2}\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^n)^2 - \frac{1}{2}\sum_{j\in\mathcal{C}_r}|\Omega_j|(p_j^{n+1}-p_j^n)^2$$

Using the second and third equations of (70) the previous equation becomes

$$\frac{1}{2}\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^{n+1})^2 + \Delta t\sum_{r\in\mathcal{V}^{i,b}}\langle \mathbf{u}_r^{n+1}, \kappa_r\beta_r\mathbf{u}_r^{n+1}\rangle + \Delta t\sum_{r\in\mathcal{V}^c}\langle \mathbf{u}_r^{n+1}, \kappa_r\beta_r^c\mathbf{u}_r^{n+1}\rangle + \frac{1}{2}\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^{n+1}-p_j^n)^2 = \frac{1}{2}\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^n)^2$$

Obviously,

$$\frac{1}{2}\sum_{j\in\mathcal{C}}|\Omega_j|(p_j^{n+1}-p_j^n)^2 \geq 0$$

Thanks to the assumptions (67) and (68), we have

$$\Delta t\sum_{r\in\mathcal{V}^{i,b}}\langle \mathbf{u}_r^{n+1}, \kappa_r\beta_r\mathbf{u}_r^{n+1}\rangle \geq 0$$

and

$$\Delta t \sum_{r\in\mathcal{V}^c} \langle \mathbf{u}_r^{n+1}, \kappa_r \beta_r^c \mathbf{u}_r^{n+1}\rangle \geq 0$$

Therefore

$$\|p_h^{n+1}\|_{L^2(\Omega)} \leq \|p_h^n\|_{L^2(\Omega)} \quad \forall\ \Delta t \geq 0$$

and this ends the proof.

□

4.4 Consistency

For this whole section, let us assume that the solution is smooth: $p \in W^{2,\infty}([0,T],\Omega)$. Let us introduce $\overline{p_j^n} = p(\mathbf{x}_j, t^n)$, $\overline{\mathbf{u}_r^n} = \mathbf{u}(\mathbf{x}_r, t^n)$ and $\kappa_r = \kappa(\mathbf{x}_r)$. We will study the consistency errors a_h^n and $\mathbf{b}_h^n$ of the scheme.

Let us define the discrete L^2 norm of these consistency errors

$$\|a_h^n\|^2_{L^2(\Omega)} = \sum_{j\in\mathcal{C}} |\Omega_j|(a_j^n)^2 \quad \text{ and } \quad \|\mathbf{b}_h^n\|^2_{L^2(\Omega)} = \sum_{r\in\mathcal{V}} |V_r|\|\mathbf{b}_r^n\|^2$$

Where $\|.\|$ is any norm over the coefficients of $\mathbf{b}_r$.

For the scheme with periodic boundary conditions (42) the consistency errors are defined as follows

$$\begin{cases} a_j^n = \dfrac{\overline{p_j^{n+1}} - \overline{p_j^n}}{\Delta t} - \dfrac{1}{|\Omega_j|} \displaystyle\sum_{r\in\mathcal{V}_j} \langle \overline{\mathbf{u}_r^{n+1}}, \kappa_r \mathbf{C}_{jr}\rangle \\ \mathbf{b}_r^n = \dfrac{1}{|V_r|}\Big(-\beta_r \overline{\mathbf{u}_r^{n+1}} - \displaystyle\sum_{j\in\mathcal{C}_r} \overline{p_j^{n+1}} \mathbf{C}_{jr}\Big) \end{cases} \tag{71}$$

Proposition 17 (Consistency of the periodic scheme)**.** *Let us consider the scheme (42) with no source ($f(\mathbf{x},t) = 0$) and periodic boundary conditions and the consistency errors (71). There exists a constant C such that, if $n\Delta t \leq T$*

$$\|a_h^n\|_{L^2(\Omega)} \leq C(\Delta t + h) \tag{72}$$

and

$$\|\mathbf{b}_h^n\|_{L^2(\Omega)} \leq Ch \tag{73}$$

Proof. We have

$$a_j^n = \frac{\overline{p_j^{n+1}} - \overline{p_j^n}}{\Delta t} - \frac{1}{|\Omega_j|} \sum_{r\in\mathcal{V}_j} \langle \overline{\mathbf{u}_r^{n+1}}, \kappa_r \mathbf{C}_{jr}\rangle$$

Knowing that

$$\frac{\overline{p_j^{n+1}} - \overline{p_j^n}}{\Delta t} = \partial_t \overline{p_j^{n+1}} + O(\Delta t)$$

and since $\overline{p_j}$ is $\overline{p}$ evaluated on the center of the cell Ω_j,

$$\begin{aligned} \partial_t \overline{p_j^{n+1}} &= \frac{1}{|\Omega_j|} \int_{\Omega_j} \partial_t \overline{p^{n+1}}(\mathbf{x})\ dx + O(h) \\ &= \frac{1}{|\Omega_j|} \int_{\Omega_j} \nabla.\kappa(\mathbf{x})\overline{\mathbf{u}^{n+1}}(\mathbf{x})\ dx + O(h) \\ &= \frac{1}{|\Omega_j|} \int_{\partial\Omega_j} \langle \kappa(\mathbf{x})\overline{\mathbf{u}^{n+1}}(\mathbf{x}), \mathbf{n}_j\rangle\ d\sigma + O(h) \end{aligned}$$

By definition of $\mathbf{C}_{jr}$, and based on Figure 1, one has

$$\mathbf{C}_{jr} = \frac{\mathbf{n}_{jl}\ell_{jl} + \mathbf{n}_{jk}\ell_{jk}}{2} = \frac{\tilde{\mathbf{n}}_{jr}|\Gamma_{jr}| + \tilde{\mathbf{n}}_{jr-1}|\Gamma_{jr-1}|}{2}$$

with $\tilde{\mathbf{n}}_{jr} = \mathbf{n}_{jl}$, $|\Gamma_{jr}| = \ell_{jl}$ and $\tilde{\mathbf{n}}_{jr-1} = \mathbf{n}_{jk}$, $|\Gamma_{jr-1}| = \ell_{jk}$.

This way, one can rewrite

$$\begin{aligned}\sum_{r\in\mathcal{V}_j} \langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, \mathbf{C}_{jr}\rangle &= \sum_{r\in\mathcal{V}_j} \langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, \frac{\tilde{\mathbf{n}}_{jr}|\Gamma_{jr}| + \tilde{\mathbf{n}}_{jr-1}|\Gamma_{jr-1}|}{2}\rangle \\ &= \frac{1}{2}\Big(\sum_{r\in\mathcal{V}_j} \langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, \tilde{\mathbf{n}}_{jr}|\Gamma_{jr}|\rangle + \sum_{r\in\mathcal{V}_j} \langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, \tilde{\mathbf{n}}_{jr-1}|\Gamma_{jr-1}|\rangle\Big)\end{aligned}$$

As $\mathcal{V}_j$ is a loop, we can change the indices of the second sum:

$$\begin{aligned}&= \frac{1}{2}\Big(\sum_{r\in\mathcal{V}_j} \langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, \tilde{\mathbf{n}}_{jr}|\Gamma_{jr}|\rangle + \sum_{r\in\mathcal{V}_j} \langle \kappa_{r+1} \overline{\mathbf{u}_{r+1}^{n+1}}, \tilde{\mathbf{n}}_{jr}|\Gamma_{jr}|\rangle\Big) \\ &= \sum_{r\in\mathcal{V}_j} \Big\langle \frac{\kappa_{r+1}\overline{\mathbf{u}_{r+1}^{n+1}} + \kappa_r \overline{\mathbf{u}_r^{n+1}}}{2}, \tilde{\mathbf{n}}_{jr}|\Gamma_{jr}|\Big\rangle\end{aligned}$$

Finally, by transforming the sum over the vertices of Ω_j into a sum over its edges, we obtain the formula

$$\sum_{r\in\mathcal{V}_j} \langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, \mathbf{C}_{jr}\rangle = \sum_{l\in\mathcal{C}_j} |\partial\Omega_{jl}| \Big\langle \frac{\kappa(\mathbf{x}_{jl}^+)\overline{\mathbf{u}^{n+1}}(\mathbf{x}_{jl}^+) + \kappa(\mathbf{x}_{jl}^-)\overline{\mathbf{u}^{n+1}}(\mathbf{x}_{jl}^-)}{2}, \mathbf{n}_{jl}\Big\rangle$$

where $\mathbf{x}_{jl}^+$ and $\mathbf{x}_{jl}^-$ are the ends of $\partial\Omega_{jl}$, the common face of Ω_j and Ω_l.

Therefore

$$a_j^n = O(\Delta t) + O(h) + \frac{1}{|\Omega_j|}\sum_{l\in\mathcal{C}_j} \Big\langle \int_{\partial\Omega_{jl}} \kappa(\mathbf{x})\overline{\mathbf{u}^{n+1}}(\mathbf{x})\, d\sigma - |\partial\Omega_{jl}| \frac{\kappa(\mathbf{x}_{jl}^+)\overline{\mathbf{u}^{n+1}}(\mathbf{x}_{jl}^+) + \kappa(\mathbf{x}_{jl}^-)\overline{\mathbf{u}^{n+1}}(\mathbf{x}_{jl}^-)}{2}, \mathbf{n}_{jl}\Big\rangle$$

Since the function under the integral is approximated by the trapezoidal rule, the error of integration is $O(h^2)$, which means there exists a $C > 0$ such that

$$\Big|\Big\langle \int_{\partial\Omega_{jl}} \kappa(\mathbf{x})\overline{\mathbf{u}^{n+1}}(\mathbf{x})\, d\sigma - |\partial\Omega_{jl}| \frac{\kappa(\mathbf{x}_{jl}^+)\overline{\mathbf{u}^{n+1}}(\mathbf{x}_{jl}^+) + \kappa(\mathbf{x}_{jl}^-)\overline{\mathbf{u}^{n+1}}(\mathbf{x}_{jl}^-)}{2}, \mathbf{n}_{jl}\Big\rangle\Big| \le Ch^2|\partial\Omega_{jl}| \le Ch^3$$

Dividing by $|\Omega_j|$ and using the lower bound of (22), we obtain $a_j^n \le C(\Delta t + h)$.

Therefore we have

$$\begin{aligned}&\|a_h^n\|_{L^2(\Omega)}^2 = \sum_{j\in\mathcal{C}} |\Omega_j||a_j^n|^2 \le (C(\Delta t + h))^2 \\ \Rightarrow\quad &\|a_h^n\|_{L^2(\Omega)} \le C(\Delta t + h)\end{aligned}$$

Which gives us (72).

Let us now consider $\mathbf{b}_r^n$.

$$\begin{aligned}
\mathbf{b}_r^n &= \frac{1}{|V_r|}\Big(- \beta_r \overline{\mathbf{u}_r^{n+1}} - \sum_{j\in\mathcal{C}_r} \overline{p_j^{n+1}} \mathbf{C}_{jr}\Big) \\
&= \frac{1}{|V_r|}\Big(- \sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} \otimes (\mathbf{x}_r - \mathbf{x}_j)\overline{\mathbf{u}_r^{n+1}} - \sum_{j\in\mathcal{C}_r} \overline{p_j^{n+1}} \mathbf{C}_{jr}\Big) \\
&= \frac{1}{|V_r|}\Big(- \sum_{j\in\mathcal{C}_r} \langle \mathbf{x}_r - \mathbf{x}_j, \overline{\mathbf{u}_r^{n+1}}\rangle \mathbf{C}_{jr} - \sum_{j\in\mathcal{C}_r} \overline{p_j^{n+1}} \mathbf{C}_{jr}\Big) \\
&= \frac{1}{|V_r|}\Big(- \sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} \big(\langle \mathbf{x}_r - \mathbf{x}_j, \overline{\mathbf{u}_r^{n+1}}\rangle + \overline{p_j^{n+1}}\big)\Big)
\end{aligned}$$

Let us write Taylor's expansion of p

$$\overline{p_j^{n+1}} = \overline{p_r^{n+1}} + \langle \mathbf{x}_j - \mathbf{x}_r, \overline{\mathbf{u}_r^{n+1}}\rangle + O(h^2)$$

Now we have

$$\begin{aligned}
\mathbf{b}_r^n &= \frac{1}{|V_r|}\Big(- \sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} \big(\overline{p_r^{n+1}} + O(h^2)\big)\Big) \\
&= \frac{-1}{|V_r|}\Big(\sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} O(h^2)\Big)
\end{aligned}$$

because

$$\sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} \overline{p_r^{n+1}} = \overline{p_r^{n+1}} \sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} = 0$$

Thanks to (23) and (21), we obtain

$$\mathbf{b}_r^n \leq \frac{-1}{C_2 h^2} \sum_{j\in\mathcal{C}_r} Ch\ O(h^2) = O(h)$$

Which means $\|\mathbf{b}_r^n\| = O(h)$, and

$$|||\mathbf{b}_h^n|||_{L^2(\Omega)} = \Big(\sum_{r\in\mathcal{V}} |V_r| \|\mathbf{b}_r^n\|^2\Big)^{1/2} = O(h)$$

And that ends the proof.

□

Let us now define the consistency error of the scheme with Neumann boundary conditions (43)

$$\begin{cases}
a_j^n = \dfrac{\overline{p_j^{n+1}} - \overline{p_j^n}}{\Delta t} - \dfrac{1}{|\Omega_j|} \displaystyle\sum_{r\in\mathcal{V}_j^i} \langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, \mathbf{C}_{jr}\rangle - \dfrac{1}{|\Omega_j|} \displaystyle\sum_{r\in\mathcal{V}_j^b} \langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, (\widehat{I_d} - \mathbf{n}_r \otimes \mathbf{n}_r)\mathbf{C}_{jr}\rangle \\
\mathbf{b}_r^n = \dfrac{1}{|V_r|}\Big(- \beta_r \overline{\mathbf{u}_r^{n+1}} - \displaystyle\sum_{j\in\mathcal{C}_r} \overline{p_j^{n+1}} \mathbf{C}_{jr}\Big) \quad \forall r \in \mathcal{V}^i \\
\mathbf{b}_r^n = \dfrac{1}{|V_r|}\Big(- \beta_r \overline{\mathbf{u}_r^{n+1}} + \displaystyle\sum_{j\in\mathcal{C}_r} (\overline{p_r^{n+1}} - \overline{p_j^{n+1}}) \mathbf{C}_{jr}\Big) \quad \forall r \in \mathcal{V}^b
\end{cases} \tag{74}$$

Proposition 18 (Consistency with Neumann boundary conditions)**.** *Let us consider the scheme (43) with no source ($f(\mathbf{x},t) = 0$) and homogeneous Neumann boundary conditions ($g(\mathbf{x}) = 0$) and the consistency errors (74). Then there exists a constant C such that, if $n\Delta t \leq T$*

$$\|a_h^n\|_{L^2(\Omega)} \leq C(\Delta t + h) \tag{75}$$

and

$$|||\mathbf{b}_h^n|||_{L^2(\Omega)} \leq Ch \tag{76}$$

Proof. Let us consider

$$
\begin{aligned}
a_j^n &= \frac{\overline{p_j^{n+1}} - \overline{p_j^n}}{\Delta t} - \frac{1}{|\Omega_j|}\sum_{r\in\mathcal{V}_j^i}\langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, \mathbf{C}_{jr}\rangle - \frac{1}{|\Omega_j|}\sum_{r\in\mathcal{V}_j^b}\langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, (\widehat{I_d} - \mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\rangle \\
&= \frac{\overline{p_j^{n+1}} - \overline{p_j^n}}{\Delta t} - \frac{1}{|\Omega_j|}\sum_{r\in\mathcal{V}_j^i}\langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, \mathbf{C}_{jr}\rangle - \frac{1}{|\Omega_j|}\sum_{r\in\mathcal{V}_j^b}\langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, \mathbf{C}_{jr}\rangle \\
&\qquad + \frac{1}{|\Omega_j|}\sum_{r\in\mathcal{V}_j^b}\langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, (\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\rangle \\
&= \frac{\overline{p_j^{n+1}} - \overline{p_j^n}}{\Delta t} - \frac{1}{|\Omega_j|}\sum_{r\in\mathcal{V}_j}\langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, \mathbf{C}_{jr}\rangle + \frac{1}{|\Omega_j|}\sum_{r\in\mathcal{V}_j^b}\langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, (\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\rangle
\end{aligned}
$$

Thanks to Proposition 34, we have

$$
\langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, (\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\rangle = \langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, \langle\mathbf{C}_{jr},\mathbf{n}_r\rangle\mathbf{n}_r\rangle = \langle\mathbf{C}_{jr},\mathbf{n}_r\rangle\langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, \mathbf{n}_r\rangle
$$

The Neumann boundary condition states that $\langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, \mathbf{n}_r\rangle = 0$.

Then, we obtain

$$
\begin{cases}
a_j^n = \dfrac{\overline{p_j^{n+1}} - \overline{p_j^n}}{\Delta t} - \dfrac{1}{|\Omega_j|}\displaystyle\sum_{r\in\mathcal{V}_j}\langle \kappa_r \overline{\mathbf{u}_r^{n+1}}, \mathbf{C}_{jr}\rangle \\
\mathbf{b}_r^n = \dfrac{1}{|V_r|}\Big(-\beta_r\overline{\mathbf{u}_r^{n+1}} - \displaystyle\sum_{j\in\mathcal{C}_r}\overline{p_j^{n+1}}\mathbf{C}_{jr}\Big) \quad \forall r\in\mathcal{V}^i \\
\mathbf{b}_r^n = \dfrac{1}{|V_r|}\Big(-\beta_r\overline{\mathbf{u}_r^{n+1}} + \displaystyle\sum_{j\in\mathcal{C}_r}(\overline{p_r^{n+1}} - \overline{p_j^{n+1}})\mathbf{C}_{jr}\Big) \quad \forall r\in\mathcal{V}^b
\end{cases} \tag{77}
$$

Note that the first equation of (77) is the same as in Proposition 17. Knowing that no property of the periodic domain has been used in its proof, the method to prove the consistency for a_h with Neumann boundary conditions is identical. However, we have to work a little bit more with $\mathbf{b}_h$.

Inside the domain, when $r\in\mathcal{V}^i$, the proof is the same as is Proposition 17, thus we obtain

$$
\|\mathbf{b}_r\| = O(h) \quad \forall\, r\in\mathcal{V}^i \tag{78}
$$

Let us now consider $\mathbf{b}_r^n$ on the boundary of the domain.

$$
\begin{aligned}
\mathbf{b}_r^n &= \frac{1}{|V_r|}\Big(-\beta_r\overline{\mathbf{u}_r^{n+1}} + \sum_{j\in\mathcal{C}_r}(\overline{p_r^{n+1}} - \overline{p_j^{n+1}})\mathbf{C}_{jr}\Big) \\
&= \frac{1}{|V_r|}\Big(-\sum_{j\in\mathcal{C}_r}\mathbf{C}_{jr}\otimes(\mathbf{x}_r - \mathbf{x}_j)\overline{\mathbf{u}_r^{n+1}} + \sum_{j\in\mathcal{C}_r}(\overline{p_r^{n+1}} - \overline{p_j^{n+1}})\mathbf{C}_{jr}\Big) \\
&= \frac{1}{|V_r|}\Big(-\sum_{j\in\mathcal{C}_r}\langle\mathbf{x}_r - \mathbf{x}_j, \overline{\mathbf{u}_r^{n+1}}\rangle\mathbf{C}_{jr} + \sum_{j\in\mathcal{C}_r}(\overline{p_r^{n+1}} - \overline{p_j^{n+1}})\mathbf{C}_{jr}\Big) \\
&= \frac{1}{|V_r|}\Big(-\sum_{j\in\mathcal{C}_r}\mathbf{C}_{jr}\big(\langle\mathbf{x}_r - \mathbf{x}_j, \overline{\mathbf{u}_r^{n+1}}\rangle + \overline{p_j^{n+1}}\big) + \sum_{j\in\mathcal{C}_r}\overline{p_r^{n+1}}\mathbf{C}_{jr}\Big)
\end{aligned}
$$

Let us write Taylor's expansion of p

$$
\overline{p_j^{n+1}} = \overline{p_r^{n+1}} + \langle\mathbf{x}_j - \mathbf{x}_r, \overline{\mathbf{u}_r^{n+1}}\rangle + O(h^2)
$$

Now we have

$$\begin{aligned}\mathbf{b}_r^n &= \frac{1}{|V_r|}\Big(- \sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr}\big(\overline{p_r^{n+1}} + O(h^2)\big) + \sum_{j\in\mathcal{C}_r} \overline{p_r^{n+1}}\mathbf{C}_{jr}\Big)\\ &= \frac{-1}{|V_r|}\Big(\sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} O(h^2)\Big)\end{aligned}$$

Therefore on the boundary, when $r \in \mathcal{V}^b$, we can do the same as in the proof of Proposition 17, and we obtain

$$\|\mathbf{b}_r\| = O(h) \quad \forall\ r \in \mathcal{V}^b \tag{79}$$

Thanks to (78) and (79), we obtain $\|\mathbf{b}_r^n\| = O(h)\forall\ r \in \mathcal{V}$, thus

$$|||\mathbf{b}_h^n|||_{L^2(\Omega)} = \Big(\sum_{r\in\mathcal{V}} |V_r| \|\mathbf{b}_r^n\|^2\Big)^{1/2} = O(h)$$

And that ends the proof.

□

Let us define the consistency errors of the scheme with Dirichlet boundary conditions (44)

$$\begin{cases} a_j^n = \dfrac{\overline{p_j^{n+1}} - \overline{p_j^n}}{\Delta t} - \dfrac{1}{|\Omega_j|}\displaystyle\sum_{r\in\mathcal{V}_j} \langle \overline{\mathbf{u}_r^{n+1}}, \kappa_r \mathbf{C}_{jr}\rangle \\ \mathbf{b}_r^n = \dfrac{1}{|V_r|}\Big(- \beta_r \overline{\mathbf{u}_r^{n+1}} - \displaystyle\sum_{j\in\mathcal{C}_r} \overline{p_j^{n+1}}\mathbf{C}_{jr}\Big) \quad \forall r \in \mathcal{V}^{i,b} \\ \mathbf{b}_r^n = \dfrac{1}{|V_r|}\Big(- \beta_r^c \overline{\mathbf{u}_r^{n+1}} - \displaystyle\sum_{j\in\mathcal{C}_r} \overline{p_j^{n+1}}\mathbf{C}_{jr}\Big) \quad \forall r \in \mathcal{V}^{c} \end{cases} \tag{80}$$

Proposition 19 (Consistency with Dirichlet boundary conditions)**.** *Let us consider the scheme (44) with no source ($f(\mathbf{x}, t) = 0$) and homogeneous Dirichlet boundary conditions ($h(\mathbf{x}) = 0$) and the consistency errors (80). Then there exists a constant C such that, if $n\Delta t \le T$*

$$\|a_h^n\|_{L^2(\Omega)} \le C(\Delta t + h) \tag{81}$$

and

$$|||\mathbf{b}_h^n|||_{L^2(\Omega)} \le Ch \tag{82}$$

Proof. In this section, the formula for a_j^n is the same as in Proposition 17 and in its proof not any property of the periodic boundary conditions have been used. Hence (81) is proved immediately.

One can notice that the third equation of (80) is the same as the second one, with $\beta_r = \beta_r^c$. Then we will study the two equations in one with $\beta_r = \beta_r^c$ on the corners.

Let us now consider $\mathbf{b}_r^n$ for $r \in \mathcal{V}^{b,c}$.

$$\begin{aligned}\mathbf{b}_r^n &= \frac{1}{|V_r|}\Big(- \beta_r \overline{\mathbf{u}_r^{n+1}} - \sum_{j\in\mathcal{C}_r} \overline{p_j^{n+1}}\mathbf{C}_{jr}\Big)\\ &= \frac{1}{|V_r|}\Big(- \sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} \otimes (\mathbf{x}_r - \mathbf{x}_j)\overline{\mathbf{u}_r^{n+1}} - \sum_{j\in\mathcal{C}_r} \overline{p_j^{n+1}}\mathbf{C}_{jr}\Big)\\ &= \frac{1}{|V_r|}\Big(- \sum_{j\in\mathcal{C}_r} \langle \mathbf{x}_r - \mathbf{x}_j, \overline{\mathbf{u}_r^{n+1}}\rangle \mathbf{C}_{jr} - \sum_{j\in\mathcal{C}_r} \overline{p_j^{n+1}}\mathbf{C}_{jr}\Big)\\ &= \frac{1}{|V_r|}\Big(- \sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr}\big(\langle \mathbf{x}_r - \mathbf{x}_j, \overline{\mathbf{u}_r^{n+1}}\rangle + \overline{p_j^{n+1}}\big)\Big)\end{aligned}$$

Let us write Taylor's expansion of p

$$\overline{p_r^{n+1}} = \overline{p_j^{n+1}} + \langle \mathbf{x}_r - \mathbf{x}_j, \overline{\mathbf{u}_r^{n+1}} \rangle + O(h^2)$$

Now we have

$$\begin{aligned} \mathbf{b}_r^n &= \frac{-1}{|V_r|}\Big(\sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} \big(\overline{p_r^{n+1}} + O(h^2)\big)\Big) \\ &= \frac{-1}{|V_r|}\Big(\sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} O(h^2)\Big) \end{aligned}$$

because

$$\sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr}\overline{p_r^{n+1}} = \overline{p_r^{n+1}} \sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr} \quad \forall r \in \mathcal{V}^{b,c}$$

and $p_r^{n+1} = 0 \quad \forall r \in \mathcal{V}^{b,c}$, thanks to the homogeneous Dirichlet boundary conditions.

Thanks to (23) and (21), we obtain

$$\mathbf{b}_r^n \leq \frac{-1}{C_2 h^2} \sum_{j\in\mathcal{C}_r} Ch\, O(h^2) = O(h) \quad \forall r \in \mathcal{V}$$

Which means $\|\mathbf{b}_r^n\| = O(h)$, and

$$|||\mathbf{b}_h^n|||_{L^2(\Omega)} = \Big(\sum_{r\in\mathcal{V}} |V_r| \|\mathbf{b}_r^n\|^2 \Big)^{1/2} = O(h)$$

And that ends the proof.

□

Remark 15. We showed that the scheme is consistent and stable for each boundary condition. Now we would like to use Lax theorem as in [8] which states that if a linear scheme is stable and consistent in a certain norm, then it is convergent in the same norm. Unfortunately we can't because the scheme is not consistent in the sense of the finite differences. Indeed in the proofs of the previous propositions $\mathbf{u}_r^{n+1}$ is considered to be exact in the expression of a_j^n while in reality it should be $\overline{\mathbf{u}_r^{n+1}} + O(h)$. Therefore in a_j^n one has

$$\frac{1}{|\Omega_j|} \sum_{r\in\mathcal{V}_j} \langle \kappa_r O(h), \mathbf{C}_{jr} \rangle \approx \frac{1}{|\Omega_j|} O(h^2) \approx O(1)$$

since $|\Omega_j|$ is homogeneous to h^2.

Therefore a_j^n is not consistent anymore.

4.5 Convergence

Let us define two error variables

$$e_j^n = p_j^n - p(\mathbf{x}_j, t^n) \quad \text{and} \quad \mathbf{f}_r^n = \mathbf{u}_r^n - \mathbf{u}(\mathbf{x}_r, t^n)$$

and their norm

$$\|e^n\|_{L^2(\Omega)} = \Big(\sum_{j\in\mathcal{C}} |\Omega_j| |e_j^n|^2 \Big)^{1/2} \quad \text{and} \quad |||\mathbf{f}^n|||_{L^2(\Omega)} = \Big(\sum_{r\in\mathcal{V}} |V_r| \|\mathbf{f}_r^n\|^2 \Big)^{1/2}$$

Proposition 20 (Convergence of the periodic scheme)**.** *Assume that $p \in W^{3,\infty}$ and the periodic scheme is consistent:* (72)-(73) *are verified. Assume there exists a constant $\alpha > 0$ such that*

$$\forall \mathbf{x} \in \Omega,\ \forall r \in \mathcal{V},\ \langle \mathbf{x}, (\kappa_r \beta_r)^s \mathbf{x} \rangle \geq \alpha |V_r| \|\mathbf{x}\|^2 \tag{83}$$

with A^s the symmetric part of A defined in Definition 6.

Then the scheme (42) is convergent for all final time $T > 0$: there exist a constant C such that

$$\|e^n\|_{L^2(\Omega)} \leq C(\Delta t + h) \tag{84}$$

Proof. By construction

$$\begin{cases} \dfrac{e_j^{n+1} - e_j^n}{\Delta t} - \dfrac{1}{|\Omega_j|} \displaystyle\sum_{r \in \mathcal{V}_j} \langle \kappa_r \mathbf{f}_r^{n+1}, \mathbf{C}_{jr} \rangle = -a_j^n \\ \beta_r \mathbf{f}_r^{n+1} + \displaystyle\sum_{j \in \mathcal{C}_r} \mathbf{C}_{jr} e_j^{n+1} = b_r^n |V_r| \end{cases} \tag{85}$$

We can multiply the first equation of (85) by $\Delta t |\Omega_j| e_j^{n+1}$ and sum it over $j \in \mathcal{C}$:

$$\begin{aligned} \sum_{j \in \mathcal{C}} |\Omega_j| |e_j^{n+1}|^2 - \sum_{j \in \mathcal{C}} |\Omega_j| e_j^{n+1} e_j^n &= \Delta t \sum_{j \in \mathcal{C}} e_j^{n+1} \sum_{r \in \mathcal{V}_j} \langle \kappa_r \mathbf{f}_r^{n+1}, \mathbf{C}_{jr} \rangle - \Delta t \sum_{j \in \mathcal{C}} |\Omega_j| e_j^{n+1} a_j^n \\ \frac{1}{2} \sum_{j \in \mathcal{C}} |\Omega_j| |e_j^{n+1}|^2 - \frac{1}{2} \sum_{j \in \mathcal{C}} |\Omega_j| |e_j^n|^2 &= -\frac{1}{2} \sum_{j \in \mathcal{C}} |\Omega_j| |e_j^{n+1} - e_j^n|^2 + \Delta t \sum_{r \in \mathcal{V}} \langle \kappa_r \mathbf{f}_r^{n+1}, \sum_{j \in \mathcal{C}_r} e_j^{n+1} \mathbf{C}_{jr} \rangle \\ &\quad - \Delta t \sum_{j \in \mathcal{C}} |\Omega_j| e_j^{n+1} a_j^n \end{aligned}$$

We can now replace the sum in the scalar product with the second line of the equation (85)

$$\begin{aligned} \sum_{j \in \mathcal{C}} |\Omega_j| |e_j^{n+1}|^2 - \sum_{j \in \mathcal{C}} |\Omega_j| |e_j^n|^2 &= -\sum_{j \in \mathcal{C}} |\Omega_j| |e_j^{n+1} - e_j^n|^2 + 2\Delta t \sum_{r \in \mathcal{V}} \langle \kappa_r \mathbf{f}_r^{n+1}, b_r^n |V_r| \rangle \\ &\quad - 2\Delta t \sum_{r \in \mathcal{V}} \langle \kappa_r \mathbf{f}_r^{n+1}, \beta_r \mathbf{f}_r^{n+1} \rangle - 2\Delta t \sum_{j \in \mathcal{C}} |\Omega_j| e_j^{n+1} a_j^n \end{aligned}$$

Thanks to Proposition 39 and the assumption (83), we know that there exists $\alpha > 0$ such that

$$-2\Delta t \sum_{r \in \mathcal{V}} \langle \kappa_r \mathbf{f}_r^{n+1}, \beta_r \mathbf{f}_r^{n+1} \rangle = -2\Delta t \sum_{r \in \mathcal{V}} \langle \mathbf{f}_r^{n+1}, (\kappa_r \beta_r)^s \mathbf{f}_r^{n+1} \rangle \leq -2\Delta t \alpha \left|\left|\left| \mathbf{f}^{n+1} \right|\right|\right|_{L^2(\Omega)}^2$$

One can write

$$-2\Delta t \sum_{j \in \mathcal{C}} |\Omega_j| e_j^{n+1} a_j^n = -2\Delta t \sum_{j \in \mathcal{C}} |\Omega_j| (e_j^{n+1} - e_j^n) a_j^n - 2\Delta t \sum_{j \in \mathcal{C}} |\Omega_j| e_j^n a_j^n$$

Young's inequality gives us the following results:

$$\begin{aligned} -2\Delta t \sum_{j \in \mathcal{C}} |\Omega_j| (e_j^{n+1} - e_j^n) a_j^n &\leq \sum_{j \in \mathcal{C}} |\Omega_j| |e_j^{n+1} - e_j^n|^2 + \Delta t^2 \sum_{j \in \mathcal{C}} |\Omega_j| |a_j^n|^2 \\ -2\Delta t \sum_{j \in \mathcal{C}} |\Omega_j| e_j^n a_j^n &\leq \Delta t \sum_{j \in \mathcal{C}} |\Omega_j| |e_j^n|^2 + \Delta t \sum_{j \in \mathcal{C}} |\Omega_j| |a_j^n|^2 \\ 2\Delta t \sum_{r \in \mathcal{V}} |V_r| \langle \mathbf{f}_r^{n+1}, \kappa_r b_r^n \rangle &\leq \Delta t \varepsilon \sum_{r \in \mathcal{V}} |V_r| \|\mathbf{f}_r^{n+1}\|^2 + \Delta t \frac{1}{\varepsilon} \sum_{r \in \mathcal{V}} |V_r| \|\kappa_r \mathbf{b}_r^n\|^2 \end{aligned}$$

Finally, we obtain the following inequation, with $\varepsilon > 0$

$$\begin{aligned} \|e^{n+1}\|_{L^2(\Omega)}^2 - \|e^n\|_{L^2(\Omega)}^2 &\leq \Delta t \varepsilon \left|\left|\left| \mathbf{f}^{n+1} \right|\right|\right|_{L^2(\Omega)}^2 + \Delta t \frac{|||\kappa|||_{\Omega,\infty}^2}{\varepsilon} |||\mathbf{b}^n|||_{L^2(\Omega)}^2 - \Delta t 2\alpha \left|\left|\left| \mathbf{f}^{n+1} \right|\right|\right|_{L^2(\Omega)}^2 + \Delta t^2 \|a^n\|_{L^2(\Omega)}^2 \\ &\quad + \Delta t \|a^n\|_{L^2(\Omega)}^2 + \Delta t \|e^n\|_{L^2(\Omega)}^2 \end{aligned}$$

where $|||\kappa|||^2_{\Omega,\infty} = \max_{r\in\mathcal{V}} |||\kappa_r|||_\infty$ which is a discretization of the infinity norm defined in the statement of the problem. Therefore the majoration of κ still holds with this discrete norm.

One can chose $\varepsilon = 2\alpha$ and rewrite

$$\|e^{n+1}\|^2_{L^2(\Omega)} \le (1+\Delta t)\|e^n\|^2_{L^2(\Omega)} + \Delta t\Big((1+\Delta t)\|a^n\|^2_{L^2(\Omega)} + \frac{|||\kappa|||^2_{\Omega,\infty}}{2\alpha}|||\mathbf{b}^n|||^2_{L^2(\Omega)}\Big)$$

Using the consistency estimates (72)-(73), one can write

$$\begin{aligned}\|e^{n+1}\|^2_{L^2(\Omega)} &\le (1+\Delta t)\|e^n\|^2_{L^2(\Omega)} + \Delta t\Big((1+\Delta t)C_1(\Delta t+h)^2 + \frac{|||\kappa|||^2_{\Omega,\infty}}{2\alpha}C_2h^2\Big)\\ &\le (1+\Delta t)\|e^n\|^2_{L^2(\Omega)} + K'\Delta t((1+\Delta t)(\Delta t+h)^2+h^2)\\ &\le (1+\Delta t)\|e^n\|^2_{L^2(\Omega)} + K\Delta t(\Delta t+h)^2\end{aligned}$$

where

$$K = 3K' = 3\max\Big(C_1, \frac{M}{2\alpha}C_2\Big)$$

with M the upper bound of $|||\kappa|||_{\Omega,\infty}$.

Thanks to Grönwall's lemma, we obtain

$$\|e^{n+1}\|^2_{L^2(\Omega)} \le \sum_{p=0}^{n-1} e^{p\Delta t}K\Delta t(\Delta t+h)^2$$

Which means, for $n\Delta t \le T$

$$\|e^{n+1}\|^2_{L^2(\Omega)} \le Q(\Delta t+h)^2$$

where

$$Q = KTe^T$$

And that ends the proof. □

Proposition 21 (Convergence with Neumann boundary conditions)**.** *Assume that $p \in W^{3,\infty}$ and the scheme with Neumann boundary conditions is consistent: (75)-(76) are verified. Assume there exists two constants $\alpha_1, \alpha_2 > 0$ such that*

$$\forall \mathbf{x} \in \Omega,\ \forall r \in \mathcal{V}^i,\ \langle \mathbf{x}, (\kappa_r\beta_r)^s\mathbf{x}\rangle \ge \alpha_1|V_r|\|\mathbf{x}\|^2 \tag{86}$$

and

$$\forall \mathbf{x} \in \Omega,\ \forall r \in \mathcal{V}^b,\ \langle \mathbf{x}, (\kappa_r(\widehat{I_d} - \mathbf{n}_r\otimes\mathbf{n}_r)\beta_r)^s\mathbf{x}\rangle \ge \alpha_2|V_r|\|\mathbf{x}\|^2 \tag{87}$$

with A^s the symmetric part of A defined in Definition 6.

Then the scheme (43) is convergent for all final time $T > 0$: there exists a constant C such that

$$\|e^n\|_{L^2(\Omega)} \le C(\Delta t + h) \tag{88}$$

Proof. By construction

$$\begin{cases}\dfrac{e_j^{n+1}-e_j^n}{\Delta t} - \dfrac{1}{|\Omega_j|}\displaystyle\sum_{r\in\mathcal{V}_j^i}\langle\kappa_r\mathbf{f}_r^{n+1},\mathbf{C}_{jr}\rangle - \dfrac{1}{|\Omega_j|}\sum_{r\in\mathcal{V}_j^b}\langle\kappa_r\mathbf{f}_r^{n+1},(\widehat{I_d}-\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\rangle = -a_j^n\\ \beta_r\mathbf{f}_r^{n+1} + \displaystyle\sum_{j\in\mathcal{C}_r}\mathbf{C}_{jr}e_j^{n+1} = b_r^n|V_r|\\ \beta_r\mathbf{f}_r^{n+1} - \displaystyle\sum_{j\in\mathcal{C}_r}\mathbf{C}_{jr}(e_r^{n+1}-e_j^{n+1}) = b_r^n|V_r|\end{cases} \tag{89}$$

We can multiply the first equation of (89) by $\Delta t|\Omega_j|e_j^{n+1}$ and sum it over $j \in \mathcal{C}$:

$$\sum_{j\in\mathcal{C}}|\Omega_j||e_j^{n+1}|^2 - \sum_{j\in\mathcal{C}}|\Omega_j|e_j^{n+1}e_j^n = \Delta t\sum_{j\in\mathcal{C}}e_j^{n+1}\sum_{r\in\mathcal{V}_j^i}\langle\kappa_r\mathbf{f}_r^{n+1},\mathbf{C}_{jr}\rangle + \Delta t\sum_{j\in\mathcal{C}}e_j^{n+1}\sum_{r\in\mathcal{V}_j^b}\langle\kappa_r\mathbf{f}_r^{n+1},$$
$$(\widehat{I}_d - \mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\rangle - \Delta t\sum_{j\in\mathcal{C}}|\Omega_j|e_j^{n+1}a_j^n$$

$$\frac{1}{2}\sum_{j\in\mathcal{C}}|\Omega_j||e_j^{n+1}|^2 - \frac{1}{2}\sum_{j\in\mathcal{C}}|\Omega_j||e_j^n|^2 = -\frac{1}{2}\sum_{j\in\mathcal{C}}|\Omega_j||e_j^{n+1}-e_j^n|^2 + \Delta t\sum_{r\in\mathcal{V}^i}\langle\kappa_r\mathbf{f}_r^{n+1},\sum_{j\in\mathcal{C}_r}e_j^{n+1}\mathbf{C}_{jr}\rangle$$
$$+\Delta t\sum_{r\in\mathcal{V}^b}\langle\kappa_r\mathbf{f}_r^{n+1},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\sum_{j\in\mathcal{C}_r}e_j^{n+1}\mathbf{C}_{jr}\rangle - \Delta t\sum_{j\in\mathcal{C}}|\Omega_j|e_j^{n+1}a_j^n$$

We can now replace the sums in the scalar products with the second and third lines of the equation (89)

$$\sum_{j\in\mathcal{C}}|\Omega_j||e_j^{n+1}|^2 - \sum_{j\in\mathcal{C}}|\Omega_j||e_j^n|^2 = -\sum_{j\in\mathcal{C}}|\Omega_j||e_j^{n+1}-e_j^n|^2 + 2\Delta t\sum_{r\in\mathcal{V}^i}\langle\kappa_r\mathbf{f}_r^{n+1},b_r^n|V_r|\rangle - 2\Delta t\sum_{r\in\mathcal{V}^i}\langle\kappa_r\mathbf{f}_r^{n+1},\beta_r\mathbf{f}_r^{n+1}\rangle$$
$$+2\Delta t\sum_{r\in\mathcal{V}^b}\langle\kappa_r\mathbf{f}_r^{n+1},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)b_r^n|V_r|\rangle + 2\Delta t\sum_{r\in\mathcal{V}^b}\langle\kappa_r\mathbf{f}_r^{n+1},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\sum_{j\in\mathcal{C}_r}e_r^{n+1}\mathbf{C}_{jr}\rangle$$
$$-2\Delta t\sum_{r\in\mathcal{V}^b}\langle\kappa_r\mathbf{f}_r^{n+1},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\beta_r\mathbf{f}_r^{n+1}\rangle - 2\Delta t\sum_{j\in\mathcal{C}}|\Omega_j|e_j^{n+1}a_j^n$$

By definition of $\mathbf{n}_r$ and thanks to Proposition 38, one has

$$2\Delta t\sum_{r\in\mathcal{V}^b}\langle\kappa_r\mathbf{f}_r^{n+1},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\sum_{j\in\mathcal{C}_r}e_r^{n+1}\mathbf{C}_{jr}\rangle = 2\Delta t\sum_{r\in\mathcal{V}^b}\langle\kappa_r\mathbf{f}_r^{n+1},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)e_r^{n+1}\Big\|\sum_{j\in\mathcal{C}_r}\mathbf{C}_{jr}\Big\|\mathbf{n}_r\rangle = 0$$

Which leads to

$$\sum_{j\in\mathcal{C}}|\Omega_j||e_j^{n+1}|^2 - \sum_{j\in\mathcal{C}}|\Omega_j||e_j^n|^2 = -\sum_{j\in\mathcal{C}}|\Omega_j||e_j^{n+1}-e_j^n|^2 + 2\Delta t\sum_{r\in\mathcal{V}^i}\langle\kappa_r\mathbf{f}_r^{n+1},b_r^n|V_r|\rangle - 2\Delta t\sum_{r\in\mathcal{V}^i}\langle\kappa_r\mathbf{f}_r^{n+1},\beta_r\mathbf{f}_r^{n+1}\rangle$$
$$+2\Delta t\sum_{r\in\mathcal{V}^b}\langle\kappa_r\mathbf{f}_r^{n+1},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)b_r^n|V_r|\rangle - 2\Delta t\sum_{r\in\mathcal{V}^b}\langle\kappa_r\mathbf{f}_r^{n+1},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_\Gamma)\beta_r\mathbf{f}_r^{n+1}\rangle - 2\Delta t\sum_{j\in\mathcal{C}}|\Omega_j|e_j^{n+1}a_j^n$$

Thanks to Proposition 39 and the assumptions (86) and (87), we know that there exists $\alpha_1, \alpha_2 > 0$ such that

$$-2\Delta t\sum_{r\in\mathcal{V}^i}\langle\kappa_r\mathbf{f}_r^{n+1},\beta_r\mathbf{f}_r^{n+1}\rangle = -2\Delta t\sum_{r\in\mathcal{V}^i}\langle\mathbf{f}_r^{n+1},(\kappa_r\beta_r)^s\mathbf{f}_r^{n+1}\rangle \leq -2\Delta t\alpha_1\sum_{r\in\mathcal{V}^i}|V_r|\|\mathbf{f}^{n+1}\|^2$$

and

$$-2\Delta t\sum_{r\in\mathcal{V}^b}\langle\kappa_r\mathbf{f}_r^{n+1},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\beta_r\mathbf{f}_r^{n+1}\rangle = -2\Delta t\sum_{r\in\mathcal{V}^b}\langle\mathbf{f}_r^{n+1},(\kappa_r(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\beta_r)^s\mathbf{f}_r^{n+1}\rangle \leq -2\Delta t\alpha_2\sum_{r\in\mathcal{V}^b}|V_r|\|\mathbf{f}^{n+1}\|^2$$

Thus

$$-2\Delta t\sum_{r\in\mathcal{V}^b}\langle\kappa_r\mathbf{f}_r^{n+1},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\beta_r\mathbf{f}_r^{n+1}\rangle - 2\Delta t\sum_{r\in\mathcal{V}^i}\langle\kappa_r\mathbf{f}_r^{n+1},\beta_r\mathbf{f}_r^{n+1}\rangle \leq -2\Delta t\min(\alpha_1,\alpha_2)\sum_{r\in\mathcal{V}}|V_r|\|\mathbf{f}^{n+1}\|^2$$

One can write

$$-2\Delta t\sum_{j\in\mathcal{C}}|\Omega_j|e_j^{n+1}a_j^n = -2\Delta t\sum_{j\in\mathcal{C}}|\Omega_j|(e_j^{n+1}-e_j^n)a_j^n - 2\Delta t\sum_{j\in\mathcal{C}}|\Omega_j|e_j^n a_j^n$$

Young's inequality gives us the following results:

$$-2\Delta t \sum_{j\in\mathcal{C}} |\Omega_j|(e_j^{n+1}-e_j^n)a_j^n \leq \sum_{j\in\mathcal{C}} |\Omega_j||e_j^{n+1}-e_j^n|^2 + \Delta t^2 \sum_{j\in\mathcal{C}} |\Omega_j||a_j^n|^2$$

$$-2\Delta t \sum_{j\in\mathcal{C}} |\Omega_j| e_j^n a_j^n \leq \Delta t \sum_{j\in\mathcal{C}} |\Omega_j||e_j^n|^2 + \Delta t \sum_{j\in\mathcal{C}} |\Omega_j||a_j^n|^2$$

$$2\Delta t \sum_{r\in\mathcal{V}^i} |V_r|\langle \mathbf{f}_r^{n+1}, \kappa_r b_r^n\rangle \leq \Delta t\varepsilon \sum_{r\in\mathcal{V}^i} |V_r|\|\mathbf{f}_r^{n+1}\|^2 + \Delta t\frac{1}{\varepsilon} \sum_{r\in\mathcal{V}^i} |V_r|\|\kappa_r \mathbf{b}_r^n\|^2$$

$$2\Delta t \sum_{r\in\mathcal{V}^b} |V_r|\langle \mathbf{f}_r^{n+1}, \kappa_r(\widehat{I}_d - \mathbf{n}_r\otimes\mathbf{n}_r) b_r^n\rangle \leq \Delta t\mu \sum_{r\in\mathcal{V}^b} |V_r|\|\mathbf{f}_r^{n+1}\|^2 + \Delta t\frac{1}{\mu} \sum_{r\in\mathcal{V}^b} |V_r|\|\kappa_r(\widehat{I}_d - \mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{b}_r^n\|^2$$

Let us now gather the previous inequalities

$$\begin{aligned}\|e^{n+1}\|^2_{L^2(\Omega)} - \|e^n\|^2_{L^2(\Omega)} \leq \Delta t\varepsilon \sum_{r\in\mathcal{V}^i} |V_r|\|\mathbf{f}_r^{n+1}\|^2 + \Delta t\frac{1}{\varepsilon} \sum_{r\in\mathcal{V}^i} |V_r|\|\kappa_r \mathbf{b}_r^n\|^2 - 2\Delta t \min(\alpha_1,\alpha_2)\left|\left|\left|\mathbf{f}^{n+1}\right|\right|\right|^2_{L^2(\Omega)} \\ +\Delta t\mu \sum_{r\in\mathcal{V}^b} |V_r|\|\mathbf{f}_r^{n+1}\|^2 + \Delta t\frac{1}{\mu} \sum_{r\in\mathcal{V}^b} |V_r|\|\kappa_r(\widehat{I}_d - \mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{b}_r^n\|^2 \\ +\Delta t^2\|a^n\|^2_{L^2(\Omega)} + \Delta t\|e^n\|^2_{L^2(\Omega)} + \Delta t\|a^n\|^2_{L^2(\Omega)}\end{aligned}$$

Let us chose $\mu=\varepsilon=2\min(\alpha_1,\alpha_2)$, therefore some of the terms simplify:

Firstly

$$\Delta t\varepsilon \sum_{r\in\mathcal{V}^i} |V_r|\|\mathbf{f}_r^{n+1}\|^2 + \Delta t\mu \sum_{r\in\mathcal{V}^b} |V_r|\|\mathbf{f}_r^{n+1}\|^2 = \Delta t 2\min(\alpha_1,\alpha_2) \sum_{r\in\mathcal{V}} |V_r|\|\mathbf{f}_r^{n+1}\|^2 = \Delta t 2\min(\alpha_1,\alpha_2)\left|\left|\left|\mathbf{f}^{n+1}\right|\right|\right|^2_{L^2(\Omega)}$$

Secondly

$$\begin{aligned}&\Delta t\frac{1}{\varepsilon} \sum_{r\in\mathcal{V}^i} |V_r|\|\kappa_r \mathbf{b}_r^n\|^2 \Delta t\frac{1}{\mu} \sum_{r\in\mathcal{V}^b} |V_r|\|\kappa_r(\widehat{I}_d - \mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{b}_r^n\|^2 \\ &= \Delta t\frac{1}{2\min(\alpha_1,\alpha_2)}|||\kappa|||^2_{\Omega,\infty} \sum_{r\in\mathcal{V}^i} |V_r|\|\mathbf{b}_r^n\|^2 + \Delta t\frac{1}{2\min(\alpha_1,\alpha_2)}|||\kappa|||^2_{\Omega,\infty}\left|\left|\left|(\widehat{I}_d - \mathbf{n}_r\otimes\mathbf{n}_r)\right|\right|\right|^2_\infty \sum_{r\in\mathcal{V}^b} |V_r|\|\mathbf{b}_r^n\|^2 \\ &= \Delta t\frac{1}{2\min(\alpha_1,\alpha_2)}|||\kappa|||^2_{\Omega,\infty} \sum_{r\in\mathcal{V}} |V_r|\|\mathbf{b}_r^n\|^2 = \Delta t\frac{1}{2\min(\alpha_1,\alpha_2)}|||\kappa|||^2_{\Omega,\infty}|||\mathbf{b}_r^n|||^2_{\Omega,\infty}\end{aligned}$$

because $\left|\left|\left|(\widehat{I}_d - \mathbf{n}_r\otimes\mathbf{n}_r)\right|\right|\right|^2_\infty = 1$

Finally, we obtain

$$\|e^{n+1}\|^2_{L^2(\Omega)} \leq (1+\Delta t)\|e^n\|^2_{L^2(\Omega)} + \Delta t\Big(\frac{|||\kappa|||^2_{\Omega,\infty}}{2\min(\alpha_1,\alpha_2)}|||\mathbf{b}^n|||^2_{L^2(\Omega)} + (1+\Delta t)\|a^n\|^2_{L^2(\Omega)}\Big)$$

Using the consistency estimates (75)-(76), one can write

$$\begin{aligned}\|e^{n+1}\|^2_{L^2(\Omega)} &\leq (1+\Delta t)\|e^n\|^2_{L^2(\Omega)} + \Delta t\Big((1+\Delta t)C_1(\Delta t+h)^2 + \frac{|||\kappa|||^2_{\Omega,\infty}}{2\min(\alpha_1,\alpha_2)2\alpha}C_2 h^2\Big) \\ &\leq (1+\Delta t)\|e^n\|^2_{L^2(\Omega)} + K'\Delta t((1+\Delta t)(\Delta t+h)^2 + h^2) \\ &\leq (1+\Delta t)\|e^n\|^2_{L^2(\Omega)} + K\Delta t(\Delta t+h)^2\end{aligned}$$

where

$$K = 3K' = 3\max\left(C_1, \frac{M}{2\min(\alpha_1, \alpha_2)}C_2\right)$$

with M the upper bound of $\|\kappa\|_{\Omega,\infty}$.

Thanks to Grönwall's lemma, we obtain

$$\|e^{n+1}\|^2_{L^2(\Omega)} \leq \sum_{p=0}^{n-1} e^{p\Delta t} K\Delta t(\Delta t + h)^2$$

Which means, for $n\Delta t \leq T$

$$\|e^{n+1}\|^2_{L^2(\Omega)} \leq Q(\Delta t + h)^2$$

where

$$Q = KTe^T$$

And that ends the proof. □

Proposition 22 (Convergence with Dirichlet boundary conditions)**.** *Assume that $p \in W^{3,\infty}$ and the scheme (44) with Dirichlet boundary conditions is consistent: (81)-(82) are verified. Assume there exists two constants $\alpha_1, \alpha_2 > 0$ such that*

$$\forall \mathbf{x} \in \Omega,\ \forall r \in \mathcal{V}^{i,b},\ \langle \mathbf{x}, (\kappa_r\beta_r)^s\mathbf{x}\rangle \geq \alpha_1|V_r|\|\mathbf{x}\|^2 \tag{90}$$

and

$$\forall \mathbf{x} \in \Omega,\ \forall r \in \mathcal{V}^{c},\ \langle \mathbf{x}, (\kappa_r\beta_r^c)^s\mathbf{x}\rangle \geq \alpha_2|V_r|\|\mathbf{x}\|^2 \tag{91}$$

with A^s the symmetric part of A defined in Definition 6.

Then the scheme (44) is convergent for all final time $T > 0$: there exists a constant C such that

$$\|e^n\|_{L^2(\Omega)} \leq C(\Delta t + h) \tag{92}$$

Proof. By construction, and choosing to write the third equation of (44) as the second one with $\beta_r = \beta_r^c$ when $r \in \mathcal{V}^c$, one has

$$\begin{cases} \dfrac{e_j^{n+1} - e_j^n}{\Delta t} - \dfrac{1}{|\Omega_j|}\displaystyle\sum_{r\in\mathcal{V}_j} \langle \kappa_r \mathbf{f}_r^{n+1}, \mathbf{C}_{jr}\rangle = -a_j^n \\ \beta_r\mathbf{f}_r^{n+1} + \displaystyle\sum_{j\in\mathcal{C}_r} \mathbf{C}_{jr}e_j^{n+1} = b_r^n|V_r| \quad \forall r \in \mathcal{V} \end{cases} \tag{93}$$

The rest of the proof is the same as in the one of Equation 84, choosing $\alpha = \min(\alpha_1, \alpha_2)$ in order to have the following inequality

$$\forall \mathbf{x} \in \Omega,\ \forall r \in \mathcal{V},\ \langle \mathbf{x}, (\kappa_r\beta_r)^s\mathbf{x}\rangle \geq \alpha|V_r|\|\mathbf{x}\|^2$$

Finally we obtain (92) with

$$C = KTe^T$$

where $K = 3\max\left(C_1, \frac{M}{2\alpha}C_2\right)$. And that ends the proof.

□

5 Second order in time

The scheme as we built it is theoretically order 1 in space and time. We will see in Section 7 that its numerical order in space is 2 and we want to obtain order 2 in time too.

5.1 Crank-Nicolson method

5.1.1 Construction of the scheme

Crank-Nicolson method is based on the trapezoidal rule, let us introduce $\mathbf{u}_r^{n+\frac{1}{2}} = \dfrac{1}{2}(\mathbf{u}_r^{n+1} + \mathbf{u}_r^n)$.

Therefore for periodic boundary conditions the new scheme writes

$$\begin{cases} |\Omega_j|\dfrac{p_j^{n+1} - p_j^n}{\Delta t} - \displaystyle\sum_{r\in\mathcal{V}_j} \langle \mathbf{u}_r^{n+\frac{1}{2}}, \kappa_r \mathbf{C}_{jr}\rangle = \dfrac{1}{2}|\Omega_j|(f_j^{n+1} + f_j^n) \\ \beta_r \mathbf{u}_r^{n+1} = -\displaystyle\sum_{j\in\mathcal{C}_r} p_j^{n+1}\mathbf{C}_{jr} \end{cases} \tag{94}$$

The schemes for Neumann, Dirichlet and Mixed boundary condition are the same as respectively (43), (44) and (45) replacing each time $\mathbf{u}_r^{n+\frac{1}{2}}$ by $\frac{1}{2}(\mathbf{u}_r^{n+1} + \mathbf{u}_r^n)$ in the first line of the system.

5.1.2 Matrix-vector form

In order to solve the problem with Crank-Nicolson method, let us get back to the scheme in vectorial form

$$I_\Omega \frac{P^{n+1} - P^n}{\Delta t} - \frac{1}{2}(MP^{n+1} + MP^n) = \frac{1}{2}(B^{n+1} + B^n) \tag{95}$$

where

- $P^n = \begin{pmatrix} p_1^n & \cdots & p_N^n \end{pmatrix}^t$ is known at each iteration, initialized with P^0 in the statement;
- B^n, I_Ω and M are the same as in Section C.1, known for all $t^n \in [0, T]$.

Therefore the problem becomes:

Find P^{n+1} such that

$$(I_\Omega - \frac{\Delta t}{2}M)P^{n+1} = \frac{\Delta t}{2}(B^{n+1} + B^n) + (I_\Omega + \frac{\Delta t}{2}M)P^n \tag{96}$$

5.2 Properties of the scheme

In order to study the stability, consistency and convergence of the scheme, let us consider the problem with no source: $f(\mathbf{x}, t) = 0$ and homogeneous boundary conditions: $g(\mathbf{x}) = 0$ and $h(\mathbf{x}) = 0$.

5.2.1 Stability

Proposition 23 (L^2 Stability of the periodic scheme)**.** *Let us make the assumption that $\kappa_r\beta_r$ is positive:*

$$\forall\ \mathbf{x} \in \Omega,\ \langle \mathbf{x}, \kappa_r\beta_r\mathbf{x}\rangle > 0 \quad \forall r \in \mathcal{V} \tag{97}$$

Then the diffusion scheme with Crank-Nicolson method (94) with no source ($f(\mathbf{x}, t) = 0$) and periodic boundary conditions is unconditionally stable in L^2 norm, which means the following inequality stands

$$\|p_h^{n+1}\|_{L^2(\Omega)} \leq \|p_h^n\|_{L^2(\Omega)} \quad \forall\ \Delta t \geq 0 \tag{98}$$

Proof. The scheme (94) rewrites:

$$\begin{cases} p_j^{n+1} - \dfrac{\Delta t}{|\Omega_j|}\displaystyle\sum_{r\in\mathcal{V}_j} \langle \frac{1}{2}(\mathbf{u}_r^{n+1} + \mathbf{u}_r^n), \kappa_r\mathbf{C}_{jr}\rangle = p_j^n \\ -\beta_r\mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r} p_j^{n+1}\mathbf{C}_{jr} \end{cases} \tag{99}$$

We can multiply the first equation of (99) by $(p_j^{n+1} + p_j^n)$, and sum it over $j \in \mathcal{C}$:

$$\sum_{j\in\mathcal{C}} |\Omega_j| p_j^{n+1}(p_j^{n+1} + p_j^n) - \Delta t \sum_{j\in\mathcal{C}} \Big((p_j^{n+1} + p_j^n) \sum_{r\in\mathcal{V}_j} \langle \frac{1}{2}(\mathbf{u}_r^{n+1} + \mathbf{u}_r^n), \kappa_r \mathbf{C}_{jr} \rangle \Big) = \sum_{j\in\mathcal{C}} |\Omega_j| p_j^n (p_j^{n+1} + p_j^n)$$

By permuting the sums on r and j, and using (15), we obtain:

$$\sum_{j\in\mathcal{C}} |\Omega_j| (p_j^{n+1})^2 - \frac{\Delta t}{2} \sum_{r\in\mathcal{V}} \langle (\mathbf{u}_r^{n+1} + \mathbf{u}_r^n), \kappa_r \sum_{j\in\mathcal{C}_r} (p_j^{n+1} + p_j^n) \mathbf{C}_{jr} \rangle = \sum_{j\in\mathcal{C}} |\Omega_j| (p_j^n)^2$$

Using the second equation of (99) the equation becomes

$$\frac{1}{2} \sum_{j\in\mathcal{C}} |\Omega_j| (p_j^{n+1})^2 + \frac{\Delta t}{2} \sum_{r\in\mathcal{V}} \langle (\mathbf{u}_r^{n+1} + \mathbf{u}_r^n), \kappa_r \beta_r (\mathbf{u}_r^{n+1} + \mathbf{u}_r^n) \rangle = \frac{1}{2} \sum_{j\in\mathcal{C}} |\Omega_j| (p_j^n)^2$$

Thanks to the assumption (97), we have

$$\frac{\Delta t}{2} \sum_{r\in\mathcal{V}} \langle (\mathbf{u}_r^{n+1} + \mathbf{u}_r^n), \kappa_r \beta_r (\mathbf{u}_r^{n+1} + \mathbf{u}_r^n) \rangle \geq 0$$

Therefore

$$\|p_h^{n+1}\|_{L^2(\Omega)} \leq \|p_h^n\|_{L^2(\Omega)} \quad \forall\ \Delta t \geq 0$$

and this ends the proof.

□

Proposition 24 (L^2 Stability with Neumann boundary conditions)**.** *Let us make the two following assumptions:*

$$\kappa_r \beta_r > 0 \quad \forall r \in \mathcal{V}^i \tag{100}$$

$$\kappa_r (\widehat{I_d} - \mathbf{n}_r \otimes \mathbf{n}_r) \beta_r > 0 \quad \forall r \in \mathcal{V}^b \tag{101}$$

Then the diffusion scheme with Crank-Nicolson method with no source ($f(\mathbf{x}, t) = 0$) and homogeneous Neumann boundary conditions ($g(\mathbf{x}) = 0$) is unconditionally stable in L^2 norm, which means the following inequality stands

$$\|p_h^{n+1}\|_{L^2(\Omega)} \leq \|p_h^n\|_{L^2(\Omega)} \quad \forall\ \Delta t \geq 0 \tag{102}$$

Proof. The scheme rewrites:

$$\begin{cases} |\Omega_j| p_j^{n+1} - \dfrac{\Delta t}{2} \Big(\displaystyle\sum_{r\in\mathcal{V}_j^i} \langle \kappa_r (\mathbf{u}_r^{n+1} + \mathbf{u}_r^n), \mathbf{C}_{jr} \rangle + \sum_{r\in\mathcal{V}_j^b} \langle \kappa_r (\mathbf{u}_r^{n+1} + \mathbf{u}_r^n), (\widehat{I_d} - \mathbf{n}_r \otimes \mathbf{n}_r) \mathbf{C}_{jr} \rangle \Big) = |\Omega_j| p_j^n \\ -\beta_r \mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r} p_j^{n+1} \mathbf{C}_{jr} \quad \forall r \in \mathcal{V}^i \\ \beta_r \mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r} (p_r^{n+1} - p_j^{n+1}) \mathbf{C}_{jr} \quad \forall r \in \mathcal{V}^b \end{cases} \tag{103}$$

In order to prove the stability of the scheme with Neumann boundary conditions with Crank-Nicolson method, we will repeat the steps of the proof for the implicit scheme (Proposition 15) adapted as in Proposition 23 for Crank-Nicolson method.

□

Proposition 25 (L^2 Stability with Dirichlet boundary conditions)**.** *Let us make the assumption that $\kappa_r \beta_r$ and $\kappa_r \beta_r^c$ are positive:*

$$\forall\ \mathbf{x} \in \Omega,\ \langle \mathbf{x}, \kappa_r \beta_r \mathbf{x} \rangle > 0 \quad \forall r \in \mathcal{V}^{i,b} \tag{104}$$

and

$$\forall\ \mathbf{x} \in \Omega,\ \langle \mathbf{x}, \kappa_r \beta_r^c \mathbf{x} \rangle > 0 \quad \forall r \in \mathcal{V}^c \tag{105}$$

Then the diffusion scheme with Crank-Nicolson method with no source ($f(\mathbf{x}, t) = 0$) and homogeneous Dirichlet boundary conditions ($h(\mathbf{x}) = 0$) is unconditionally stable in L^2 norm, which means the following inequality stands

$$\|p_h^{n+1}\|_{L^2(\Omega)} \leq \|p_h^n\|_{L^2(\Omega)} \quad \forall\ \Delta t \geq 0 \tag{106}$$

Proof. The scheme with Crank-Nicolson method rewrites:

$$\begin{cases} |\Omega_j| p_j^{n+1} - \dfrac{\Delta t}{2} \displaystyle\sum_{r \in \mathcal{V}_j} \langle (\mathbf{u}_r^{n+1} + \mathbf{u}_r^n), \kappa_r \mathbf{C}_{jr} \rangle = |\Omega_j| p_j^n \\ \beta_r \mathbf{u}_r^{n+1} = -\displaystyle\sum_{j \in \mathcal{C}_r} p_j^{n+1} \mathbf{C}_{jr} \quad \forall r \in \mathcal{V}^{i,b} \\ \beta_r^c \mathbf{u}_r^{n+1} = -\displaystyle\sum_{j \in \mathcal{C}_r} p_j^{n+1} \mathbf{C}_{jr} \quad \forall r \in \mathcal{V}^c \end{cases} \tag{107}$$

Once again the proof of stability for this scheme is the same as the one for the implicit scheme (Proposition 16) adapted as in Proposition 23 for Crank-Nicolson method.

□

5.2.2 Consistency

For the scheme with periodic boundary conditions and with Crank-Nicolson method, the consistency errors are defined as follows

$$\begin{cases} a_j^n = \dfrac{\overline{p_j^{n+1}} - \overline{p_j^n}}{\Delta t} - \dfrac{1}{|\Omega_j|} \displaystyle\sum_{r \in \mathcal{V}_j} \langle \overline{\mathbf{u}_r^{n+\frac{1}{2}}}, \kappa_r \mathbf{C}_{jr} \rangle \\ \mathbf{b}_r^n = \dfrac{1}{|V_r|} \Big(-\beta_r \overline{\mathbf{u}_r^{n+1}} - \displaystyle\sum_{j \in \mathcal{C}_r} \overline{p_j^{n+1}} \mathbf{C}_{jr} \Big) \end{cases} \tag{108}$$

Proposition 26 (Consistency of the periodic scheme with Crank-Nicolson method)**.** *Let us consider the scheme (94) with no source ($f(\mathbf{x}, t) = 0$) and periodic boundary conditions and the consistency errors (108). There exists a constant C such that, if $n\Delta t \leq T$*

$$\|a_h^n\|_{L^2(\Omega)} \leq C(\Delta t^2 + h) \tag{109}$$

and

$$\|\|\mathbf{b}_h^n\|\|_{L^2(\Omega)} \leq Ch \tag{110}$$

Proof. We have

$$a_j^n = \frac{\overline{p_j^{n+1}} - \overline{p_j^n}}{\Delta t} - \frac{1}{|\Omega_j|} \sum_{r \in \mathcal{V}_j} \langle \overline{\mathbf{u}_r^{n+\frac{1}{2}}}, \kappa_r \mathbf{C}_{jr} \rangle$$

Let us consider Taylor's expansion of $\overline{p_j^n}$ and $\overline{p_j^{n+1}}$

$$\overline{p_j^n} = \overline{p_j^{n+\frac{1}{2}}} - \frac{\Delta t}{2} \partial_t \overline{p_j^{n+\frac{1}{2}}} + O(\Delta t^2)$$

and

$$\overline{p_j^{n+1}} = \overline{p_j^{n+\frac{1}{2}}} + \frac{\Delta t}{2} \partial_t \overline{p_j^{n+\frac{1}{2}}} + O(\Delta t^2)$$

Then

$$\frac{\overline{p_j^{n+1}} - \overline{p_j^n}}{\Delta t} = \partial_t \overline{p_j^{n+\frac{1}{2}}} + O(\Delta t^2)$$

This time, let us consider Taylor's expansion of $\partial_t \overline{p_j^n}$ and $\partial_t \overline{p_j^{n+1}}$

$$\partial_t \overline{p_j^n} = \partial_t \overline{p_j^{n+\frac{1}{2}}} - \frac{\Delta t}{2}\partial_t^2 \overline{p_j^{n+\frac{1}{2}}} + O(\Delta t^2)$$

and

$$\partial_t \overline{p_j^{n+1}} = \partial_t \overline{p_j^{n+\frac{1}{2}}} + \frac{\Delta t}{2}\partial_t^2 \overline{p_j^{n+\frac{1}{2}}} + O(\Delta t^2)$$

Then

$$\frac{\overline{p_j^{n+1}} - \overline{p_j^n}}{\Delta t} = \frac{\partial_t \overline{p_j^{n+1}} + \partial_t \overline{p_j^n}}{2} + O(\Delta t^2)$$

Since $\overline{p_j}$ is $\overline{p}$ evaluated on the center of the cell Ω_j, one has

$$\begin{aligned}
\partial_t \overline{p_j^{n+1}} &= \frac{1}{|\Omega_j|}\int_{\Omega_j} \partial_t \overline{p^{n+1}}(\mathbf{x})\ dx + O(h) \\
&= \frac{1}{|\Omega_j|}\int_{\Omega_j} \nabla.\kappa(\mathbf{x})\overline{\mathbf{u}^{n+1}}(\mathbf{x})\ dx + O(h) \\
&= \frac{1}{|\Omega_j|}\int_{\partial\Omega_j} \langle \kappa(\mathbf{x})\overline{\mathbf{u}^{n+1}}(\mathbf{x}), \mathbf{n}_j\rangle\ d\sigma + O(h)
\end{aligned}$$

and

$$\partial_t \overline{p_j^n} = \frac{1}{|\Omega_j|}\int_{\partial\Omega_j} \langle \kappa(\mathbf{x})\overline{\mathbf{u}^{n}}(\mathbf{x}), \mathbf{n}_j\rangle\ d\sigma + O(h)$$

Then we use the same manipulation of $\mathbf{C}_{jr}$ as in Proposition 17 and we obtain

$$\sum_{r\in\mathcal{V}_j} \langle \kappa_r \overline{\mathbf{u}_r^{n+\frac{1}{2}}}, \mathbf{C}_{jr}\rangle = \sum_{l\in\mathcal{C}_j} |\partial\Omega_{jl}| \Big\langle \frac{\kappa(\mathbf{x}_{jl}^+)\overline{\mathbf{u}^{n+\frac{1}{2}}}(\mathbf{x}_{jl}^+) + \kappa(\mathbf{x}_{jl}^-)\overline{\mathbf{u}^{n+\frac{1}{2}}}(\mathbf{x}_{jl}^-)}{2}, \mathbf{n}_{jl}\Big\rangle$$

where $\mathbf{x}_{jl}^+$ and $\mathbf{x}_{jl}^-$ are the ends of $\partial\Omega_{jl}$, the common face of Ω_j and Ω_l.

Therefore

$$a_j^n = O(\Delta t) + O(h) + \frac{1}{|\Omega_j|}\sum_{l\in\mathcal{C}_j}\Big\langle \int_{\partial\Omega_{jl}} \kappa(\mathbf{x})\overline{\mathbf{u}^{n+\frac{1}{2}}}(\mathbf{x})\ d\sigma - |\partial\Omega_{jl}| \frac{\kappa(\mathbf{x}_{jl}^+)\overline{\mathbf{u}^{n+\frac{1}{2}}}(\mathbf{x}_{jl}^+) + \kappa(\mathbf{x}_{jl}^-)\overline{\mathbf{u}^{n+\frac{1}{2}}}(\mathbf{x}_{jl}^-)}{2}, \mathbf{n}_{jl}\Big\rangle$$

Since the function under the integral is approximated by the trapezoidal rule, the error of integration is $O(h^2)$, which means there exists a $C > 0$ such that

$$\Big|\Big\langle \int_{\partial\Omega_{jl}} \kappa(\mathbf{x})\overline{\mathbf{u}^{n+1}}(\mathbf{x})\ d\sigma - |\partial\Omega_{jl}| \frac{\kappa(\mathbf{x}_{jl}^+)\overline{\mathbf{u}^{n+1}}(\mathbf{x}_{jl}^+) + \kappa(\mathbf{x}_{jl}^-)\overline{\mathbf{u}^{n+1}}(\mathbf{x}_{jl}^-)}{2}, \mathbf{n}_{jl}\Big\rangle\Big| \le Ch^2|\partial\Omega_{jl}| \le Ch^3$$

Dividing by $|\Omega_j|$ and using the lower bound of (22), we obtain $a_j^n \le C(\Delta t^2 + h)$.

Therefore we have

$$\begin{aligned}
& \|a_h^n\|_{L^2(\Omega)}^2 = \sum_{j\in\mathcal{C}} |\Omega_j||a_j^n|^2 \le (C(\Delta t^2 + h))^2 \\
\Rightarrow\ & \|a_h^n\|_{L^2(\Omega)} \le C(\Delta t^2 + h)
\end{aligned}$$

Which gives us (109).

The consistency error $\mathbf{b}_r^n$ is the same as in Proposition 17, therefore the consistency error is the same. Thus we obtain (110) and that ends the proof. □

Let us now define the consistency error of the scheme with Neumann boundary conditions and Crank-Nicolson method

$$\begin{cases} a_j^n = \dfrac{\overline{p_j^{n+1}} - \overline{p_j^n}}{\Delta t} - \dfrac{1}{|\Omega_j|} \displaystyle\sum_{r \in \mathcal{V}_j^i} \langle \kappa_r \overline{\mathbf{u}_r^{n+\frac{1}{2}}}, \mathbf{C}_{jr} \rangle - \dfrac{1}{|\Omega_j|} \displaystyle\sum_{r \in \mathcal{V}_j^b} \langle \kappa_r \overline{\mathbf{u}_r^{n+\frac{1}{2}}}, (\widehat{I_d} - \mathbf{n}_r \otimes \mathbf{n}_r) \mathbf{C}_{jr} \rangle \\ \mathbf{b}_r^n = \dfrac{1}{|V_r|} \Big(- \beta_r \overline{\mathbf{u}_r^{n+1}} - \displaystyle\sum_{j \in \mathcal{C}_r} \overline{p_j^{n+1}} \mathbf{C}_{jr} \Big) \quad \forall r \in \mathcal{V}^i \\ \mathbf{b}_r^n = \dfrac{1}{|V_r|} \Big(- \beta_r \overline{\mathbf{u}_r^{n+1}} + \displaystyle\sum_{j \in \mathcal{C}_r} (\overline{p_r^{n+1}} - \overline{p_j^{n+1}}) \mathbf{C}_{jr} \Big) \quad \forall r \in \mathcal{V}^b \end{cases} \tag{111}$$

Proposition 27 (Consistency with Neumann boundary conditions and Crank-Nicolson method)**.** *Let us consider the scheme with no source ($f(\mathbf{x}, t) = 0$) and homogeneous Neumann boundary conditions ($g(\mathbf{x}) = 0$) and the consistency errors (74). Then there exists a constant C such that, if $n\Delta t \leq T$*

$$\|a_h^n\|_{L^2(\Omega)} \leq C(\Delta t^2 + h) \tag{112}$$

and

$$|||\mathbf{b}_h^n|||_{L^2(\Omega)} \leq Ch \tag{113}$$

Proof. The beginning of the proof is the same as for Proposition 18 replacing $\overline{\mathbf{u}_r^{n+1}}$ with $\overline{\mathbf{u}_r^{n+\frac{1}{2}}}$. Once we transformed the consistency error of (111) into (108), one has to follow the steps of Proposition 26 in order to prove (112) and (113).

□

Let us define the consistency errors of the scheme with Dirichlet boundary conditions and Crank-Nicolson method

$$\begin{cases} a_j^n = \dfrac{\overline{p_j^{n+1}} - \overline{p_j^n}}{\Delta t} - \dfrac{1}{|\Omega_j|} \displaystyle\sum_{r \in \mathcal{V}_j} \langle \overline{\mathbf{u}_r^{n+\frac{1}{2}}}, \kappa_r \mathbf{C}_{jr} \rangle \\ \mathbf{b}_r^n = \dfrac{1}{|V_r|} \Big(- \beta_r \overline{\mathbf{u}_r^{n+1}} - \displaystyle\sum_{j \in \mathcal{C}_r} \overline{p_j^{n+1}} \mathbf{C}_{jr} \Big) \quad \forall r \in \mathcal{V}^{i,b} \\ \mathbf{b}_r^n = \dfrac{1}{|V_r|} \Big(- \beta_r^c \overline{\mathbf{u}_r^{n+1}} - \displaystyle\sum_{j \in \mathcal{C}_r} \overline{p_j^{n+1}} \mathbf{C}_{jr} \Big) \quad \forall r \in \mathcal{V}^c \end{cases} \tag{114}$$

Proposition 28 (Consistency with Dirichlet boundary conditions and Crank-Nicolson method)**.** *Let us consider the scheme with no source ($f(\mathbf{x}, t) = 0$) and homogeneous Dirichlet boundary conditions ($h(\mathbf{x}) = 0$) and the consistency errors (114). Then there exists a constant C such that, if $n\Delta t \leq T$*

$$\|a_h^n\|_{L^2(\Omega)} \leq C(\Delta t^2 + h) \tag{115}$$

and

$$|||\mathbf{b}_h^n|||_{L^2(\Omega)} \leq Ch \tag{116}$$

Proof. The proof for the consistency error a_j^n is the same as in Proposition 26 and the one for $\mathbf{b}_r^n$ is the same as in Proposition 19. Therefore (115) and (116) are immediately proved. □

5.2.3 Convergence

Let us use the same convergence errors as in Section 4.5 and the norm associated to them.

Proposition 29 (Convergence of the periodic scheme with Crank-Nicolson method)**.** *Assume that $p \in W^{3,\infty}$ and the periodic scheme with Crank-Nicolson method is consistent: (109)-(110) are verified. Assume that $\Delta t \leq 1$ and there exists a constant $\alpha > 0$ such that*

$$\forall \mathbf{x} \in \Omega,\ \forall r \in \mathcal{V},\ \langle \mathbf{x}, (\kappa_r \beta_r)^s \mathbf{x}\rangle \geq \alpha |V_r| \|\mathbf{x}\|^2 \tag{117}$$

with A^s the symmetric part of A defined in Definition 6.

Then the scheme (94) is convergent for all final time $T > 0$: there exist a constant C such that

$$\|e^n\|_{L^2(\Omega)} \leq C(\Delta t^2 + h) \tag{118}$$

Proof. By construction

$$\begin{cases} \dfrac{e_j^{n+1} - e_j^n}{\Delta t} - \dfrac{1}{|\Omega_j|} \displaystyle\sum_{r \in \mathcal{V}_j} \langle \kappa_r \mathbf{f}_r^{n+\frac{1}{2}}, \mathbf{C}_{jr}\rangle = -a_j^n \\ \beta_r \mathbf{f}_r^{n+1} + \displaystyle\sum_{j \in \mathcal{C}_r} \mathbf{C}_{jr} e_j^{n+1} = b_r^n |V_r| \end{cases} \tag{119}$$

where $\mathbf{f}_r^{n+\frac{1}{2}} = \dfrac{1}{2}(\mathbf{f}_r^{n+1} + \mathbf{f}_r^n)$.

We can multiply the first equation of (119) by $\Delta t |\Omega_j| e_j^{n+\frac{1}{2}}$ with $e_j^{n+\frac{1}{2}} = \frac{1}{2}(e_j^{n+1} + e_j^n)$ and sum it over $j \in \mathcal{C}$:

$$\sum_{j \in \mathcal{C}} |\Omega_j| e_j^{n+1} e_j^{n+\frac{1}{2}} - \sum_{j \in \mathcal{C}} |\Omega_j| e_j^n e_j^{n+\frac{1}{2}} = \Delta t \sum_{j \in \mathcal{C}} e_j^{n+\frac{1}{2}} \sum_{r \in \mathcal{V}_j} \langle \kappa_r \mathbf{f}_r^{n+\frac{1}{2}}, \mathbf{C}_{jr}\rangle - \Delta t \sum_{j \in \mathcal{C}} |\Omega_j| e_j^{n+\frac{1}{2}} a_j^n$$
$$\frac{1}{2} \sum_{j \in \mathcal{C}} |\Omega_j| |e_j^{n+1}|^2 - \frac{1}{2} \sum_{j \in \mathcal{C}} |\Omega_j| |e_j^n|^2 = \Delta t \sum_{r \in \mathcal{V}} \langle \kappa_r \mathbf{f}_r^{n+\frac{1}{2}}, \sum_{j \in \mathcal{C}_r} e_j^{n+\frac{1}{2}} \mathbf{C}_{jr}\rangle - \Delta t \sum_{j \in \mathcal{C}} |\Omega_j| e_j^{n+\frac{1}{2}} a_j^n$$

We can now replace the sum in the scalar product with the second line of the equation (119)

$$\sum_{j \in \mathcal{C}} |\Omega_j| |e_j^{n+1}|^2 - \sum_{j \in \mathcal{C}} |\Omega_j| |e_j^n|^2 = 2\Delta t \sum_{r \in \mathcal{V}} \langle \kappa_r \mathbf{f}_r^{n+\frac{1}{2}}, \frac{\mathbf{b}_r^n + \mathbf{b}_r^{n-1}}{2} |V_r|\rangle - 2\Delta t \sum_{r \in \mathcal{V}} \langle \kappa_r \mathbf{f}_r^{n+\frac{1}{2}}, \beta_r \mathbf{f}_r^{n+\frac{1}{2}}\rangle$$
$$-2\Delta t \sum_{j \in \mathcal{C}} |\Omega_j| e_j^{n+\frac{1}{2}} a_j^n$$

Thanks to Proposition 39 and the assumption (117), we know that there exists $\alpha > 0$ such that

$$-2\Delta t \sum_{r \in \mathcal{V}} \langle \kappa_r \mathbf{f}_r^{n+\frac{1}{2}}, \beta_r \mathbf{f}_r^{n+\frac{1}{2}}\rangle = -2\Delta t \sum_{r \in \mathcal{V}} \langle \mathbf{f}_r^{n+\frac{1}{2}}, (\kappa_r \beta_r)^s \mathbf{f}_r^{n+\frac{1}{2}}\rangle \leq -2\Delta t \alpha \left\|\left| \mathbf{f}^{n+\frac{1}{2}} \right\|\right|^2_{L^2(\Omega)}$$

One can write

$$-2\Delta t \sum_{j \in \mathcal{C}} |\Omega_j| e_j^{n+\frac{1}{2}} a_j^n = -\Delta t \sum_{j \in \mathcal{C}} |\Omega_j| e_j^{n+1} a_j^n - \Delta t \sum_{j \in \mathcal{C}} |\Omega_j| e_j^n a_j^n$$

and

$$2\Delta t \sum_{r \in \mathcal{V}} \langle \kappa_r \mathbf{f}_r^{n+\frac{1}{2}}, \frac{\mathbf{b}_r^n + \mathbf{b}_r^{n-1}}{2} |V_r|\rangle = \Delta t \sum_{r \in \mathcal{V}} \langle \kappa_r \mathbf{f}_r^{n+\frac{1}{2}}, \mathbf{b}_r^n |V_r|\rangle + \Delta t \sum_{r \in \mathcal{V}} \langle \kappa_r \mathbf{f}_r^{n+\frac{1}{2}}, \mathbf{b}_r^{n-1} |V_r|\rangle$$

Young's inequality gives us the following results:

$$-\Delta t \sum_{j\in\mathcal{C}} |\Omega_j| e_j^{n+1} a_j^n \leq \frac{\Delta t}{2} \sum_{j\in\mathcal{C}} |\Omega_j||e_j^{n+1}|^2 + \frac{\Delta t}{2} \sum_{j\in\mathcal{C}} |\Omega_j||a_j^n|^2$$

$$-\Delta t \sum_{j\in\mathcal{C}} |\Omega_j| e_j^{n} a_j^n \leq \frac{\Delta t}{2} \sum_{j\in\mathcal{C}} |\Omega_j||e_j^{n}|^2 + \frac{\Delta t}{2} \sum_{j\in\mathcal{C}} |\Omega_j||a_j^n|^2$$

$$\Delta t \sum_{r\in\mathcal{V}} |V_r| \langle \mathbf{f}_r^{n+\frac{1}{2}}, \kappa_r \mathbf{b}_r^n \rangle \leq \frac{\Delta t}{2}\varepsilon \sum_{r\in\mathcal{V}} |V_r| \|\mathbf{f}_r^{n+\frac{1}{2}}\|^2 + \frac{\Delta t}{2\varepsilon} \sum_{r\in\mathcal{V}} |V_r| \|\kappa_r \mathbf{b}_r^n\|^2$$

$$\Delta t \sum_{r\in\mathcal{V}} |V_r| \langle \mathbf{f}_r^{n+\frac{1}{2}}, \kappa_r \mathbf{b}_r^{n-1} \rangle \leq \frac{\Delta t}{2}\varepsilon \sum_{r\in\mathcal{V}} |V_r| \|\mathbf{f}_r^{n+\frac{1}{2}}\|^2 + \frac{\Delta t}{2\varepsilon} \sum_{r\in\mathcal{V}} |V_r| \|\kappa_r \mathbf{b}_r^{n-1}\|^2$$

Finally, we obtain the following inequation, with $\varepsilon > 0$

$$\|e^{n+1}\|^2_{L^2(\Omega)} - \|e^n\|^2_{L^2(\Omega)} \leq \Delta t \varepsilon \left\| \left| \mathbf{f}^{n+\frac{1}{2}} \right| \right\|^2_{L^2(\Omega)} + \Delta t \frac{|\!|\!|\kappa|\!|\!|^2_{\Omega,\infty}}{\varepsilon} \frac{|\!|\!|\mathbf{b}^n|\!|\!|^2_{L^2(\Omega)} + |\!|\!|\mathbf{b}^{n-1}|\!|\!|^2_{L^2(\Omega)}}{2} - 2\Delta t \alpha \left\| \left| \mathbf{f}^{n+\frac{1}{2}} \right| \right\|^2_{L^2(\Omega)}$$
$$+ \Delta t \|a^n\|^2_{L^2(\Omega)} + \frac{\Delta t}{2} \|e^n\|^2_{L^2(\Omega)} + \frac{\Delta t}{2} \|e^{n+1}\|^2_{L^2(\Omega)}$$

where $|\!|\!|\kappa|\!|\!|^2_{\Omega,\infty} = \max_{r\in\mathcal{V}} |\!|\!|\kappa_r|\!|\!|_\infty$ which is a discretization of the infinity norm defined in the statement of the problem. Therefore the majoration of κ still holds with this discrete norm.

One can chose $\varepsilon = 2\alpha$ and rewrite

$$\Big(1 - \frac{\Delta t}{2}\Big)\|e^{n+1}\|^2_{L^2(\Omega)} \leq \Big(1 + \frac{\Delta t}{2}\Big)\|e^n\|^2_{L^2(\Omega)} + \Delta t\Big(\|a^n\|^2_{L^2(\Omega)} + \frac{|\!|\!|\kappa|\!|\!|^2_{\Omega,\infty}}{2\alpha} \frac{|\!|\!|\mathbf{b}^n|\!|\!|^2_{L^2(\Omega)} + |\!|\!|\mathbf{b}^{n-1}|\!|\!|^2_{L^2(\Omega)}}{2}\Big)$$

$$\Leftrightarrow \quad \|e^{n+1}\|^2_{L^2(\Omega)} \leq \Big(1 - \frac{\Delta t}{2}\Big)^{-1}\Big(1 + \frac{\Delta t}{2}\Big)\|e^n\|^2_{L^2(\Omega)} + \Delta t \Big(1 - \frac{\Delta t}{2}\Big)^{-1}\Big(\|a^n\|^2_{L^2(\Omega)} + \frac{|\!|\!|\kappa|\!|\!|^2_{\Omega,\infty}}{2\alpha} \frac{|\!|\!|\mathbf{b}^n|\!|\!|^2_{L^2(\Omega)} + |\!|\!|\mathbf{b}^{n-1}|\!|\!|^2_{L^2(\Omega)}}{2}\Big)$$

Let us now use the method found in [17] in order to bound $\|e^{n+1}\|^2_{L^2(\Omega)}$.

Thanks to the assumption that $\Delta t \leq 1$, then one has

$$\Big(1 - \frac{\Delta t}{2}\Big) \geq \frac{1}{2} \quad \Rightarrow \Big(1 - \frac{\Delta t}{2}\Big)^{-1} \leq 2$$

Remark 16. Assuming that $\Delta t \leq 1$ is not a problem since we are considering the convergence of the scheme when Δt tends to 0 which means it will eventually be as small as we need it to be.

Moreover

$$\begin{aligned}
\Big(1 - \frac{\Delta t}{2}\Big)^{-1}\Big(1 + \frac{\Delta t}{2}\Big) &= \Big(1 - \frac{\Delta t}{2}\Big)^{-1} + \frac{\Delta t}{2}\Big(1 - \frac{\Delta t}{2}\Big)^{-1} \\
&= \Big(1 - \frac{\Delta t}{2} + \frac{\Delta t}{2}\Big)\Big(1 - \frac{\Delta t}{2}\Big)^{-1} + \frac{\Delta t}{2}\Big(1 - \frac{\Delta t}{2}\Big)^{-1} \\
&= 1 + \Delta t\Big(1 - \frac{\Delta t}{2}\Big)^{-1} \\
&\leq 1 + 2\Delta t
\end{aligned}$$

Then the previous inequality on the error of convergence becomes

$$\|e^{n+1}\|^2_{L^2(\Omega)} \leq (1 + 2\Delta t)\|e^n\|^2_{L^2(\Omega)} + 2\Delta t\Big(\|a^n\|^2_{L^2(\Omega)} + \frac{|\!|\!|\kappa|\!|\!|^2_{\Omega,\infty}}{2\alpha} \frac{|\!|\!|\mathbf{b}^n|\!|\!|^2_{L^2(\Omega)} + |\!|\!|\mathbf{b}^{n-1}|\!|\!|^2_{L^2(\Omega)}}{2}\Big)$$

Using the consistency estimates (109)-(110), one can write

$$\begin{aligned}\|e^{n+1}\|^2_{L^2(\Omega)} &\le (1+2\Delta t)\|e^n\|^2_{L^2(\Omega)} + 2\Delta t\Big(C_1(\Delta t^2+h)^2 + \frac{|||\kappa|||^2_{\Omega,\infty}}{2\alpha}C_2h^2\Big)\\ &\le (1+2\Delta t)\|e^n\|^2_{L^2(\Omega)} + K'\Delta t((\Delta t^2+h)^2+h^2)\\ &\le (1+2\Delta t)\|e^n\|^2_{L^2(\Omega)} + K\Delta t(\Delta t^2+h)^2\end{aligned}$$

where

$$K = 3K' = 2\max\left(2C_1, \frac{M}{\alpha}C_2\right)$$

with M the upper bound of $|||\kappa|||_{\Omega,\infty}$.

Thanks to Grönwall's lemma, we obtain

$$\|e^{n+1}\|^2_{L^2(\Omega)} \le \sum_{p=0}^{n-1} e^{2p\Delta t}K\Delta t(\Delta t^2+h)^2$$

Which means, for $n\Delta t \le T$

$$\|e^n\|^2_{L^2(\Omega)} \le Q(\Delta t^2+h)^2$$

where

$$Q = KTe^{2T}$$

And that ends the proof.

□

Proposition 30 (Convergence with Neumann boundary conditions). *Assume that $p \in W^{3,\infty}$ and the scheme with Neumann boundary and Crank-Nicolson method conditions is consistent: (112)-(113) are verified. Assume there exists two constants $\alpha_1, \alpha_2 > 0$ such that*

$$\forall \mathbf{x} \in \Omega,\ \forall r \in \mathcal{V}^i,\ \langle \mathbf{x}, (\kappa_r\beta_r)^s\mathbf{x}\rangle \ge \alpha_1|V_r|\|\mathbf{x}\| \tag{120}$$

and

$$\forall \mathbf{x} \in \Omega,\ \forall r \in \mathcal{V}^b,\ \langle \mathbf{x}, (\kappa_r(\widehat{I_d} - \mathbf{n}_r \otimes \mathbf{n}_r)\beta_r)^s\mathbf{x}\rangle \ge \alpha_2|V_r|\|\mathbf{x}\| \tag{121}$$

with A^s the symmetric part of A defined in Definition 6.

Then the scheme with Neumann boundary conditions and Crank-Nicolson method is convergent for all final time $T > 0$: there exists a constant C such that

$$\|e^n\|_{L^2(\Omega)} \le C(\Delta t^2 + h) \tag{122}$$

Proof. In order to prove this proposition, let us repeat the steps of Proposition 21 adapting it to the Crank-Nicolson method as in the proof of Proposition 29.

□

Proposition 31 (Convergence with Dirichlet boundary conditions). *Assume that $p \in W^{3,\infty}$ and the scheme with Dirichlet boundary conditions and Crank-Nicolson method is consistent: (115)-(116) are verified. Assume there exists two constants $\alpha_1, \alpha_2 > 0$ such that*

$$\forall \mathbf{x} \in \Omega,\ \forall r \in \mathcal{V}^{i,b},\ \langle \mathbf{x}, (\kappa_r\beta_r)^s\mathbf{x}\rangle \ge \alpha_1|V_r|\|\mathbf{x}\|^2 \tag{123}$$

and

$$\forall \mathbf{x} \in \Omega,\ \forall r \in \mathcal{V}^c,\ \langle \mathbf{x}, (\kappa_r\beta_r^c)^s\mathbf{x}\rangle \ge \alpha_2|V_r|\|\mathbf{x}\|^2 \tag{124}$$

with A^s the symmetric part of A defined in Definition 6.

Then the scheme with Dirichlet boundary conditions and Crank-Nicolson method is convergent for all final time $T > 0$: there exists a constant C such that

$$\|e^n\|_{L^2(\Omega)} \le C(\Delta t^2 + h) \tag{125}$$

Proof. Once again to prove this proposition, one can repeat the steps of Proposition 22 adapting it to the Crank-Nicolson method as in the proof of Proposition 29.

□

6 Hybrid face-node scheme

We saw in Section 7 that the nodal scheme has its limits in the case of Cartesian quad or triangle meshes: that is when spurious modes appear. In order to fix that, we propose to use an hybrid scheme with fluxes on both faces and nodes of the mesh. This way each cell will communicate with all of its neighbors. Similar construction for the gradient at the middle of the faces can be found in [2, 3].

6.1 Space discretization

The idea is to get back to the integration of the continuous problem as seen in Section 3.1 and discretize the gradient with a nodal and a face formulation.

$$|\Omega_j|\partial_t p_j - \int_{\partial\Omega_j} \langle \kappa\mathbf{u}, \mathbf{n}\rangle \, d\sigma = |\Omega_j| f_j$$

As a reminder in nodal formulation, we approached the gradient as follows:

$$\int_{\partial\Omega_j} \langle \kappa\mathbf{u}, \mathbf{n}\rangle \, d\sigma \approx \sum_{r\in\mathcal{V}_j} \langle \kappa_r \mathbf{u}(\mathbf{x}_r), \mathbf{C}_{jr}\rangle$$

We now introduce the gradient discretized at the faces:

$$\int_{\partial\Omega_j} \langle \kappa\mathbf{u}, \mathbf{n}\rangle \, d\sigma \approx \sum_{l\in\mathcal{F}_j} \ell_l \langle \kappa_l \mathbf{u}(\mathbf{x}_l), \mathbf{n}_l\rangle$$

Knowing that $\mathcal{F} = \mathcal{F}^i \cup \mathcal{F}^b$, and since $\kappa_l = \kappa(\mathbf{x}_l)$ is symmetric, one has:

$$\sum_{l\in\mathcal{F}_j} \ell_l \langle \kappa_l \mathbf{u}(\mathbf{x}_l), \mathbf{n}_l\rangle = \sum_{l\in\mathcal{F}_j^i} \ell_l \langle \mathbf{u}(\mathbf{x}_l), \kappa_l \mathbf{n}_l\rangle + \sum_{l\in\mathcal{F}_j^b} \ell_l \langle \mathbf{u}(\mathbf{x}_l), \kappa_l \mathbf{n}_l\rangle \tag{126}$$

In the case of a periodic domain, there are no boundary conditions to the problem (ie. $\mathcal{F}^b = \emptyset$). We will study the boundaries of the domain for different boundary conditions later in this section.

In order to approach $\mathbf{u}(\mathbf{x}_l)$, let us introduce $j_{1,2}$, the indices of the cells that share the face l. We consider Taylor's expansion of p at the points $\mathbf{x}_{j_1}$ and $\mathbf{x}_{j_2}$:

$$p(\mathbf{x}_{j_1}) = p(\mathbf{x}_l) + \langle \mathbf{u}(\mathbf{x}_l), \mathbf{x}_{j_1} - \mathbf{x}_l\rangle + O(h^2)$$
$$p(\mathbf{x}_{j_2}) = p(\mathbf{x}_l) + \langle \mathbf{u}(\mathbf{x}_l), \mathbf{x}_{j_2} - \mathbf{x}_l\rangle + O(h^2)$$

Then

$$\langle \mathbf{u}(\mathbf{x}_l), \mathbf{x}_{j_1} - \mathbf{x}_{j_2}\rangle = p(\mathbf{x}_{j_1}) - p(\mathbf{x}_{j_2}) + O(h^2) \tag{127}$$

Now we have a formula for $\mathbf{u}(\mathbf{x}_l)$ in the direction $\mathbf{x}_{j_1} - \mathbf{x}_{j_2}$ and we want one in the orthogonal direction.

In order to approach $\mathbf{u}(\mathbf{x}_l)$ in the direction $(\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp$, let us define an approximation of $\mathbf{u}(\mathbf{x}_l)$:

$$\langle \mathbf{u}(\mathbf{x}_l), (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp\rangle \approx \langle \sum_{r\in\mathcal{V}_l} \frac{\mathbf{u}(\mathbf{x}_r)}{2}, (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp\rangle \tag{128}$$

Then the approximation of $\mathbf{u}$ at the point $\mathbf{x}_l$ is entirely defined.

Let us now introduce $(\alpha_l, \delta_l) \in \mathbb{R}^2$ such that

$$\kappa_l \mathbf{n}_l = \alpha_l (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp + \delta_l (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})$$

Then from (128) and (127)

$$\langle \mathbf{u}(\mathbf{x}_l), \kappa_l \mathbf{n}_l \rangle = \alpha_l \langle \mathbf{u}(\mathbf{x}_l), (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp \rangle + \delta_l \langle \mathbf{u}(\mathbf{x}_l), \mathbf{x}_{j_1} - \mathbf{x}_{j_2} \rangle$$
$$\approx \alpha_l \sum_{r \in \mathcal{V}_l} \langle \frac{\mathbf{u}(\mathbf{x}_r)}{2}, (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp \rangle + \delta_l (p(\mathbf{x}_{j_1}) - p(\mathbf{x}_{j_2}))$$

Finally, the face flux for the hybrid scheme writes

$$\langle \mathbf{u}_l, \kappa_l \mathbf{n}_l \rangle = \alpha_l \sum_{r \in \mathcal{V}_l} \langle \frac{\mathbf{u}_r}{2}, (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp \rangle + \delta_l (p_{j_1} - p_{j_2}) \tag{129}$$

One can combine the nodal flux from Section 3.1 and the face flux from (129) with a coefficient $\lambda \in [0, 1]$ in order to build the semi-discrete hybrid scheme on a periodic domain:

$$\begin{cases} |\Omega_j| \partial_t p_j - \lambda \sum_{l \in \mathcal{F}_j^i} \ell_l \langle \mathbf{u}_l, \kappa_l \mathbf{n}_l \rangle - (1 - \lambda) \sum_{r \in \mathcal{V}_j^i} \langle \kappa_r \mathbf{u}_r, \mathbf{C}_{jr} \rangle = |\Omega_j| f_j \\ \langle \mathbf{u}_l, \kappa_l \mathbf{n}_l \rangle = \alpha_l \sum_{r \in \mathcal{V}_l^i} \langle \frac{\mathbf{u}_r}{2}, (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp \rangle + \delta_l (p_{j_1} - p_{j_2}) \quad \forall\, l \in \mathcal{F}^i \\ \beta_r \mathbf{u}_r = - \sum_{j \in \mathcal{C}_r} p_j \mathbf{C}_{jr} \quad \forall\, r \in \mathcal{V}^i \end{cases} \tag{130}$$

As a reminder, j_1 and j_2 are the indices of the cells that share the face l as one can see in Figure 11. Since the face l belongs to the cells Ω_j, Ω_{j_1} and Ω_{j_2}, one has $\Omega_j = \Omega_{j_1}$ or $\Omega_j = \Omega_{j_2}$. We chose $\Omega_j = \Omega_{j_2}$ so that the vector $\mathbf{x}_{j_2}\mathbf{x}_{j_1}$ is in the same direction as $\mathbf{n}_l$, thus α_l and δ_l are positive.

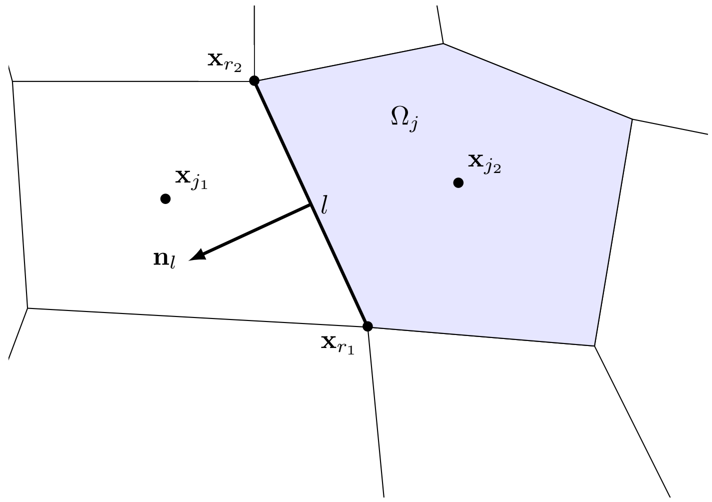


Figure 11: Construction of the flux at the face l

Remark 17. For the nodal flux solver to be well posed, we need the matrix β_r to be invertible, which will bring some conditions on the mesh. These conditions are given in Section 3.4.

Remark 18. What is interesting with the hybrid scheme when $\lambda = 1$ is that on Cartesian meshes and when $\kappa = I_d$, it degenerates towards the most intuitive scheme with fluxes at the faces, named TPFA [12], which we know is order 2 in space and even preserves the maximum principle.
Indeed, on Cartesian meshes the vector $\mathbf{x}_{j_1} - \mathbf{x}_{j_2}$ has the same direction as $\mathbf{n}_l$, therefore the scalar product of $\mathbf{u}_l$ with $(\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp$ disappears. We are then left with the most intuitive gradient formula: $\langle \mathbf{u}_l, \mathbf{n}_l \rangle = p_{j_1} - p_{j_2}$.

Remark 19. The choice of λ is not obvious. Ideally we would like to chose $\lambda << 1$ so that the scheme stays as close as possible from the original nodal solver, with some of the fluxes that pass through the faces. We would like for the hybrid scheme to correct the issues encountered with the nodal scheme (spurious modes).
Unfortunately, even though we have good reasons to believe that the hybrid scheme will be efficient with $\lambda = 1$, we have no certainty that it will work for $0 < \lambda < 1$. That is why we will implement the method in a way that makes it possible to chose $\lambda \in [0, 1]$ and make numerical tests to determine which are a sufficient value for λ.

6.2 Boundary conditions

Let us now build the hybrid scheme with Neumann and Dirichlet boundary conditions. As a reminder, working with Neumann or Dirichlet boundary conditions leads to adding respectively (131) or (132) to the problem:

$$\kappa(\mathbf{x})\mathbf{u}(\mathbf{x},t).\mathbf{n}_\Gamma = g(\mathbf{x},t) \qquad \forall\ \mathbf{x} \in \Gamma,\ t \in [0,T] \tag{131}$$

$$p(\mathbf{x},t) = h(\mathbf{x},t) \qquad \forall\ \mathbf{x} \in \Gamma,\ t \in [0,T] \tag{132}$$

6.2.1 Neumann boundary conditions

Let us get back to the first equation of (130) with $\lambda = 1$ so that the nodal flux disappears since we already built its boundary conditions in Section 3.2:

$$\begin{aligned} &|\Omega_j|\partial_t p_j - \Big(\sum_{l\in\mathcal{F}_j^i} \ell_l\langle \mathbf{u}_l, \kappa_l\mathbf{n}_l\rangle + \sum_{l\in\mathcal{F}_j^b} \ell_l\langle \mathbf{u}_l, \kappa_l\mathbf{n}_l\rangle\Big) = |\Omega_j|f_j \\ \Leftrightarrow\quad &|\Omega_j|\partial_t p_j - \sum_{l\in\mathcal{F}_j^i} \ell_l\langle \mathbf{u}_l, \kappa_l\mathbf{n}_l\rangle = |\Omega_j|f_j + \sum_{l\in\mathcal{F}_j^b} \ell_l g_l \end{aligned}$$

where $g_l = g(\mathbf{x}_l, t)$ which is given in the statement of the problem.

Therefore the semi-discrete hybrid scheme with Neumann boundary conditions writes, for $\lambda = 1$:

$$\begin{cases} |\Omega_j|\partial_t p_j - \sum_{l\in\mathcal{F}_j^i} \ell_l\langle \mathbf{u}_l, \kappa_l\mathbf{n}_l\rangle = |\Omega_j|f_j + \sum_{l\in\mathcal{F}_j^b} \ell_l g_l \\ \langle \mathbf{u}_l, \kappa_l\mathbf{n}_l\rangle = \alpha_l \sum_{r\in\mathcal{V}_l} \langle \frac{\mathbf{u}_r}{2}, (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp\rangle + \delta_l(p_{j_1} - p_{j_2}) \quad \forall\ l \in \mathcal{F}^i \\ \beta_r\mathbf{u}_r = -\sum_{j\in\mathcal{C}_r} p_j\mathbf{C}_{jr} \quad \forall\ r \in \mathcal{V}^i \\ \beta_r\mathbf{u}_r = \sum_{j\in\mathcal{C}_r} (p_r - p_j)\mathbf{C}_{jr} \quad \forall r \in \mathcal{V}^b \end{cases} \tag{133}$$

where

- $p_r = \theta_r^{-1}\Big(\sum_{j\in\mathcal{C}_r} \theta_{jr}p_j + \frac{g_r}{\|\kappa_r\mathbf{n}_r\|}\Big)$;
- $\theta_{jr} = \langle \beta_r^{-1}\mathbf{C}_{jr}, \mathbf{v}_r\rangle$;
- $\mathbf{v}_r = \frac{\kappa_r\mathbf{n}_r}{\|\kappa_r\mathbf{n}_r\|}$.

Remark 20. We do not have to consider $r \in \mathcal{V}^c$ since when $l \in \mathcal{F}^b$, we immediately replace $\langle \mathbf{u}_l, \kappa_l\mathbf{n}_l\rangle$ by the Neumann boundary condition g_l and on the other hand when $l \in \mathcal{F}^i$, $\mathcal{V}_l \cap \mathcal{V}^c = \emptyset$.

6.2.2 Dirichlet boundary conditions

With Dirichlet boundary conditions, we saw earlier in Section 3.2.2 that the nodal flux writes as follows

$$\beta_r \mathbf{u}_r = \sum_{j\in\mathcal{C}_r} (h_r - p_j)\mathbf{C}_{jr} \quad \forall r \in \mathcal{V}^b \tag{134}$$

with $h_r = h(\mathbf{x}_r)$ which is given in the statement.

Let us consider the second equation of the scheme (130) and inject $\mathbf{u}_r$ from (134) in it.

$$\langle \mathbf{u}_l, \kappa_l \mathbf{n}_l\rangle = \alpha_l \sum_{r\in\mathcal{V}_l^b} \langle \frac{1}{2}\beta_r^{-1} \sum_{k\in\mathcal{C}_r} (h_r - p_k)\mathbf{C}_{kr}, (\mathbf{x}_l - \mathbf{x}_{j_2})^\perp\rangle + \delta_l(p_l - p_{j_2}) \tag{135}$$

Since we consider the boundary of the domain, the point $\mathbf{x}_{j_1}$ does not exist anymore, it would be outside the domain. Therefore we chose $\mathbf{x}_l$, which is the middle of the face l, to replace it.

As the boundary values for Dirichlet boundary conditions are known at the vertices of the boundary in the code, one can approximate p_l as follows

$$p_l = h_l = \sum_{r\in\mathcal{V}_l} \frac{h_r}{2} \quad \forall\, l \in \mathcal{F}^b$$

Therefore the semi-discrete hybrid scheme with Dirichlet boundary conditions writes, with $\lambda = 1$

$$\begin{cases} |\Omega_j|\partial_t p_j - \sum_{l\in\mathcal{F}_j} \ell_l \langle \mathbf{u}_l, \kappa_l \mathbf{n}_l\rangle = |\Omega_j| f_j \\ \langle \mathbf{u}_l, \kappa_l \mathbf{n}_l\rangle = \alpha_l \sum_{r\in\mathcal{V}_l} \langle \frac{\mathbf{u}_r}{2}, (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp\rangle + \delta_l(p_{j_1} - p_{j_2}) \quad \forall\, l\in\mathcal{F}^i \\ \langle \mathbf{u}_l, \kappa_l \mathbf{n}_l\rangle = \alpha_l \sum_{r\in\mathcal{V}_l^b} \langle \frac{\mathbf{u}_r}{2}, (\mathbf{x}_l - \mathbf{x}_{j_2})^\perp\rangle + \delta_l(h_l - p_{j_2}) \quad \forall\, l\in\mathcal{F}^b \\ \beta_r \mathbf{u}_r = -\sum_{j\in\mathcal{C}_r} p_j \mathbf{C}_{jr} \qquad \forall\, r\in\mathcal{V}^i \\ \beta_r \mathbf{u}_r = \sum_{j\in\mathcal{C}_r} (h_r - p_j)\mathbf{C}_{jr} \quad \forall\, r\in\mathcal{V}^b \\ \beta_r^c \mathbf{u}_r = \sum_{j\in\mathcal{C}_r} (h_r - p_j)\mathbf{C}_{jr} \quad \forall\, r\in\mathcal{V}^b \end{cases} \tag{136}$$

6.2.3 Mixed boundary conditions

The semi-discrete hybrid scheme for mixed boundary conditions with $\lambda = 1$ is directly built from the schemes (133) and (136):

$$\begin{cases} |\Omega_j|\partial_t p_j - \Big(\sum_{l\in\mathcal{F}_j^i} \ell_l \langle \mathbf{u}_l, \kappa_l \mathbf{n}_l\rangle + \sum_{l\in\mathcal{F}_j^{D,b}} \ell_l \langle \mathbf{u}_l, \kappa_l \mathbf{n}_l\rangle \Big) = |\Omega_j| f_j + \sum_{l\in\mathcal{F}_j^{N,b}} \ell_l g_l \\ \langle \mathbf{u}_l, \kappa_l \mathbf{n}_l\rangle = \alpha_l \sum_{r\in\mathcal{V}_l} \langle \frac{\mathbf{u}_r}{2}, (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp\rangle + \delta_l (p_{j_1} - p_{j_2}) \qquad \forall\, l \in \mathcal{F}^i \\ \langle \mathbf{u}_l, \kappa_l \mathbf{n}_l\rangle = \alpha_l \sum_{r\in\mathcal{V}_l^b} \langle \frac{\mathbf{u}_r}{2}, (\mathbf{x}_{l} - \mathbf{x}_{j_2})^\perp\rangle + \delta_l (h_l^{n+1} - p_{j_2}^{n+1}) \quad \forall\, l \in \mathcal{F}^{D,b} \\ \beta_r \mathbf{u}_r = -\sum_{j\in\mathcal{C}_r} p_j \mathbf{C}_{jr} \qquad \forall\, r \in \mathcal{V}^i \\ \beta_r \mathbf{u}_r = \sum_{j\in\mathcal{C}_r} (p_r - p_j)\mathbf{C}_{jr} \quad \forall r \in \mathcal{V}^{N,b} \\ \beta_r \mathbf{u}_r = \sum_{j\in\mathcal{C}_r} (h_r - p_j)\mathbf{C}_{jr} \quad \forall r \in \mathcal{V}^{D,b} \\ \beta_r^c \mathbf{u}_r = \sum_{j\in\mathcal{C}_r} (h_r - p_j)\mathbf{C}_{jr} \quad \forall r \in \mathcal{V}^{D,c} \end{cases} \tag{137}$$

where

- $p_r = \theta_r^{-1}\Big(\sum_{j\in\mathcal{C}_r} \theta_{jr} p_j + \frac{g_r}{\|\kappa_r \mathbf{n}_r\|}\Big)$;
- $\theta_{jr} = \langle \beta_r^{-1}\mathbf{C}_{jr}, \mathbf{v}_r\rangle$;
- $\mathbf{v}_r = \frac{\kappa_r \mathbf{n}_r}{\|\kappa_r \mathbf{n}_r\|}$.

6.3 Time discretization

In order to discretize the scheme in time, let us use the same method as in Section 3.3 which is the backward Euler method (or implicit Euler method). Let us reuse the time step $\Delta t > 0$, such that $t^n = n\Delta t \le T$, with T the final time.

The hybrid scheme writes, for periodic boundary conditions

$$\begin{cases} |\Omega_j| \frac{p_j^{n+1} - p_j^n}{\Delta t} - \lambda \sum_{l\in\mathcal{F}_j^i} \ell_l \langle \mathbf{u}_l^{n+1}, \kappa_l \mathbf{n}_l\rangle - (1-\lambda) \sum_{r\in\mathcal{V}_j} \langle \kappa_r \mathbf{u}_r^{n+1}, \mathbf{C}_{jr}\rangle = |\Omega_j| f_j \\ \langle \mathbf{u}_l^{n+1}, \kappa_l \mathbf{n}_l\rangle = \alpha_l \sum_{r\in\mathcal{V}_l} \langle \frac{\mathbf{u}_r^{n+1}}{2}, (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp\rangle + \delta_l (p_{j_1}^{n+1} - p_{j_2}^{n+1}) \quad \forall\, l \in \mathcal{F}^i \\ \beta_r \mathbf{u}_r^{n+1} = -\sum_{j\in\mathcal{C}_r} p_j^{n+1} \mathbf{C}_{jr} \quad \forall\, r \in \mathcal{V}^i \end{cases} \tag{138}$$

In the rest of the section, for the sake of clarity we will consider that $\lambda = 1$, which leads to the disappearance of the nodal part in the first equation of the scheme.

For Neumann boundary conditions one has

$$\begin{cases} |\Omega_j|\dfrac{p_j^{n+1}-p_j^n}{\Delta t} - \displaystyle\sum_{l\in\mathcal{F}_j^i} \ell_l \langle \mathbf{u}_l^{n+1}, \kappa_l \mathbf{n}_l\rangle = |\Omega_j| f_j + \sum_{l\in\mathcal{F}_j^b} \ell_l g_l^{n+1} \\ \langle \mathbf{u}_l^{n+1}, \kappa_l \mathbf{n}_l\rangle = \alpha_l \displaystyle\sum_{r\in\mathcal{V}_l} \langle \frac{\mathbf{u}_r^{n+1}}{2}, (\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle + \delta_l(p_{j_1}^{n+1}-p_{j_2}^{n+1}) \quad \forall\, l\in\mathcal{F}^i \\ \beta_r \mathbf{u}_r^{n+1} = -\displaystyle\sum_{j\in\mathcal{C}_r} p_j^{n+1}\mathbf{C}_{jr} \quad \forall\, r\in\mathcal{V}^i \\ \beta_r \mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r} (p_r^{n+1}-p_j^{n+1})\mathbf{C}_{jr} \quad \forall r\in\mathcal{V}^b \end{cases} \tag{139}$$

where

- $p_r^{n+1} = \theta_r^{-1}\Big(\displaystyle\sum_{j\in\mathcal{C}_r}\theta_{jr}p_j^{n+1} + \frac{g_r^{n+1}}{\|\kappa_r\mathbf{n}_r\|}\Big);$
- $\theta_{jr} = \langle \beta_r^{-1}\mathbf{C}_{jr}, \mathbf{v}_r\rangle;$
- $\mathbf{v}_r = \dfrac{\kappa_r\mathbf{n}_r}{\|\kappa_r\mathbf{n}_r\|}.$

And for Dirichlet boundary conditions:

$$\begin{cases} |\Omega_j|\dfrac{p_j^{n+1}-p_j^n}{\Delta t} - \displaystyle\sum_{l\in\mathcal{F}_j} \ell_l \langle \mathbf{u}_l^{n+1}, \kappa_l \mathbf{n}_l\rangle = |\Omega_j| f_j \\ \langle \mathbf{u}_l^{n+1}, \kappa_l \mathbf{n}_l\rangle = \alpha_l \displaystyle\sum_{r\in\mathcal{V}_l} \langle \frac{\mathbf{u}_r^{n+1}}{2}, (\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle + \delta_l(p_{j_1}^{n+1}-p_{j_2}^{n+1}) \quad \forall\, l\in\mathcal{F}^i \\ \langle \mathbf{u}_l^{n+1}, \kappa_l \mathbf{n}_l\rangle = \alpha_l \displaystyle\sum_{r\in\mathcal{V}_l^b} \langle \frac{\mathbf{u}_r^{n+1}}{2}, (\mathbf{x}_{l}-\mathbf{x}_{j_2})^\perp\rangle + \delta_l(h_{l}^{n+1}-p_{j_2}^{n+1}) \quad \forall\, l\in\mathcal{F}^b \\ \beta_r \mathbf{u}_r^{n+1} = -\displaystyle\sum_{j\in\mathcal{C}_r} p_j^{n+1}\mathbf{C}_{jr} \quad \forall\, r\in\mathcal{V}^i \\ \beta_r \mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r} (h_r^{n+1}-p_j^{n+1})\mathbf{C}_{jr} \quad \forall r\in\mathcal{V}^b \\ \beta_r^c \mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r} (h_r^{n+1}-p_j^{n+1})\mathbf{C}_{jr} \quad \forall r\in\mathcal{V}^c \end{cases} \tag{140}$$

Finally, for mixed boundary conditions the implicit hybrid scheme writes:

$$\begin{cases} |\Omega_j|\dfrac{p_j^{n+1}-p_j^n}{\Delta t} - \Big(\displaystyle\sum_{l\in\mathcal{F}_j^i} \ell_l\langle \mathbf{u}_l^{n+1},\kappa_l\mathbf{n}_l\rangle + \sum_{l\in\mathcal{F}_j^{D,b}} \ell_l\langle \mathbf{u}_l^{n+1},\kappa_l\mathbf{n}_l\rangle\Big) = |\Omega_j|f_j + \sum_{l\in\mathcal{F}_j^{N,b}} \ell_l g_l^{n+1} \\ \langle \mathbf{u}_l^{n+1},\kappa_l\mathbf{n}_l\rangle = \alpha_l \displaystyle\sum_{r\in\mathcal{V}_l} \langle \frac{\mathbf{u}_r^{n+1}}{2},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle + \delta_l(p_{j_1}^{n+1}-p_{j_2}^{n+1}) \quad \forall\, l\in\mathcal{F}^i \\ \langle \mathbf{u}_l^{n+1},\kappa_l\mathbf{n}_l\rangle = \alpha_l \displaystyle\sum_{r\in\mathcal{V}_l^{D,b}} \langle \frac{\mathbf{u}_r^{n+1}}{2},(\mathbf{x}_{l}-\mathbf{x}_{j_2})^\perp\rangle + \delta_l(h_l^{n+1}-p_{j_2}^{n+1}) \quad \forall\, l\in\mathcal{F}^{D,b} \\ \beta_r\mathbf{u}_r^{n+1} = -\displaystyle\sum_{j\in\mathcal{C}_r} p_j^{n+1}\mathbf{C}_{jr} \qquad \forall\, r\in\mathcal{V}^i \\ \beta_r\mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r} (p_r^{n+1}-p_j^{n+1})\mathbf{C}_{jr} \quad \forall r\in\mathcal{V}^{N,b} \\ \beta_r\mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r} (h_r^{n+1}-p_j^{n+1})\mathbf{C}_{jr} \quad \forall r\in\mathcal{V}^{D,b} \\ \beta_r^c\mathbf{u}_r^{n+1} = \displaystyle\sum_{j\in\mathcal{C}_r} (h_r^{n+1}-p_j^{n+1})\mathbf{C}_{jr} \quad \forall r\in\mathcal{V}^{D,c} \end{cases} \tag{141}$$

where

- $p_r^{n+1} = \theta_r^{-1}\Big(\displaystyle\sum_{j\in\mathcal{C}_r}\theta_{jr}p_j^{n+1} + \frac{g_r^{n+1}}{\|\kappa_r\mathbf{n}_r\|}\Big);$
- $\theta_{jr} = \langle \beta_r^{-1}\mathbf{C}_{jr},\mathbf{v}_r\rangle;$
- $\mathbf{v}_r = \dfrac{\kappa_r\mathbf{n}_r}{\|\kappa_r\mathbf{n}_r\|}.$

7 Numerical results

Let us consider the following problem:

Find p such that

$$\begin{cases} \partial_t p(\mathbf{x},t) - \nabla.(\kappa(\mathbf{x})\mathbf{u}(\mathbf{x},t)) = f(\mathbf{x},t) & \forall\, \mathbf{x}\in\Omega,\ t\in[0,T] \\ p(\mathbf{x},0) = p_0(\mathbf{x}) & \forall\, \mathbf{x}\in\Omega \\ \kappa(\mathbf{x})\mathbf{u}(\mathbf{x},t).\mathbf{n}_\Gamma = g(\mathbf{x},t) & \forall\, \mathbf{x}\in\Gamma_D,\ t\in[0,T] \\ p(\mathbf{x},t) = h(\mathbf{x},t) & \forall\, \mathbf{x}\in\Gamma_N,\ t\in[0,T] \end{cases} \tag{142}$$

where Γ_N and Γ_D represent the boundaries of Ω with respectively Neumann and Dirichlet boundary conditions.

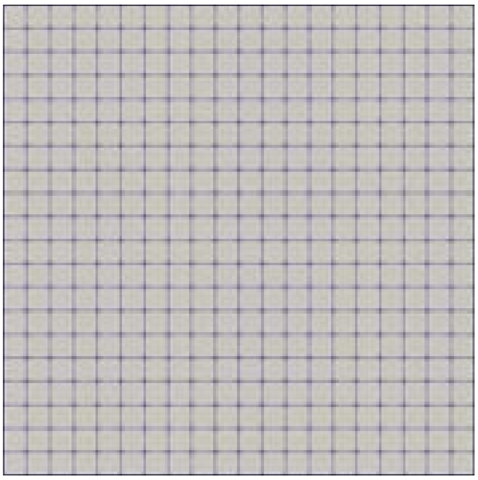
We chose to study the convergence of the scheme on the domain $\Omega = [0,1]\times[0,1]$ discretized with four non-cartesian meshes of quads, going from $N = 10$ cells per side to $N = 80$. We chose $T = 1$ as final time. As we consider the backward Euler method, let us chose $\Delta t = a\times h^2$, with $h = \frac{1}{N}$ the space step. We chose $a = 10$ so that the time step does not become too small when we refine the mesh.

We shall study the L^2 error of the scheme at the time T, calculated as follows

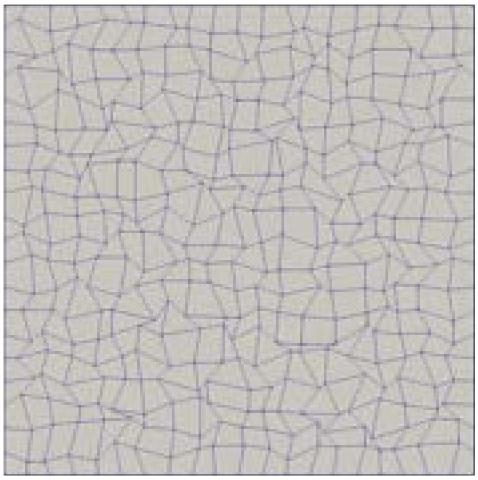
$$Err = \Big(\int_\Omega |p_{ex}(\mathbf{x},T) - p_h(\mathbf{x},T)|^2\ d\mathbf{x}\Big)^{\frac{1}{2}}$$

where the exact solution p_{ex} is given for each test-case. The boundary condition and the source are constructed using this exact solution, this way we do not need the computer to calculate a reference solution of the problems.

In this section, non Cartesian meshes of quads are made from Cartesian meshes in which the position of the vertices is randomly shifted of a length between 0 and h. We obtain a mesh of quadrangular cells that one can see on Figure 12b, with 20 cell per side.



(a) Example of Cartesian mesh of squares



(b) Non Cartesian mesh of quadrangles

Figure 12: Randomization of a mesh for $N = 20$ cell per side

7.1 Stationary solutions

7.1.1 Linear solution

Let us consider the following linear solution

$$p(\mathbf{x}, t) = 5x + y + 2 \tag{143}$$

with a constant diffusion coefficient $\kappa = I_d$.

Then we have

$$f(\mathbf{x}, t) = 0 \quad \text{and} \quad \begin{cases} g(\mathbf{x}, t) = \begin{pmatrix} 5 \\ 1 \end{pmatrix} .\mathbf{n}_\Gamma & \forall\ \mathbf{x} \in \Gamma_N \\ h(\mathbf{x}, t) = 5x + y + 2 & \forall\ \mathbf{x} \in \Gamma_D \end{cases}$$

The convergence order of the scheme with Neumann boundary conditions is not given here because the scheme is exact for this type of solution. As stated in Proposition 13 the scheme is not exact for Dirichlet and mixed boundary conditions here because there are corners with a unique cell in our mesh. Therefore one can see on Figure 13 that for Dirichlet and mixed boundary conditions the numerical order of convergence in space is 2.

7.1.2 Trigonometric solution

Let us consider the following stationary solution

$$p(\mathbf{x}, t) = \sin(\pi x) \sin(\pi y) \tag{144}$$

with a constant diffusion coefficient $\kappa = I_d$.

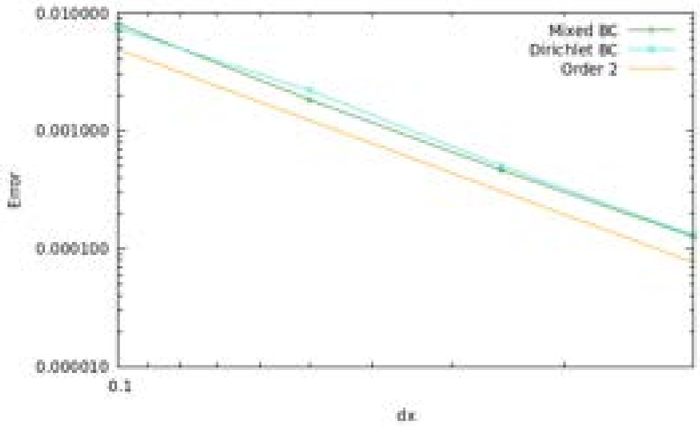


Figure 13: Convergence graph of the linear solution (143) at $T = 1$

Then we have

$$f(\mathbf{x},t) = 2\pi \sin(\pi x)\sin(\pi y) \quad \text{and} \quad \begin{cases} g(\mathbf{x},t) = \begin{pmatrix} \pi\cos(\pi x)\sin(\pi y) \\ \pi \sin(\pi x)\cos(\pi y) \end{pmatrix} .\mathbf{n}_\Gamma & \forall\ \mathbf{x} \in \Gamma_D \\ h(\mathbf{x},t) = \sin(\pi x)\sin(\pi y) & \forall\ \mathbf{x} \in \Gamma_N \end{cases}$$

One can see on Figure 14 that for every boundary condition the numerical order of convergence in space is 2.

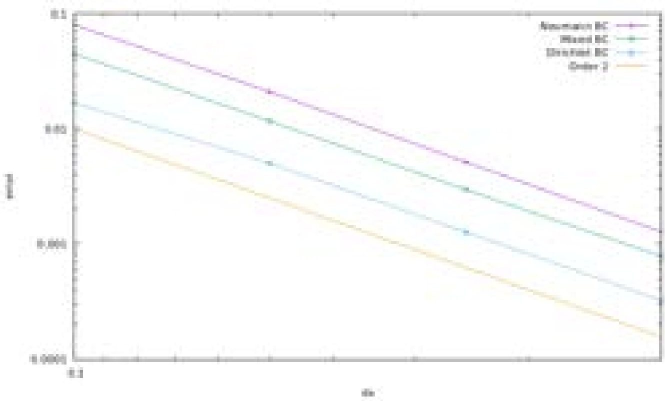
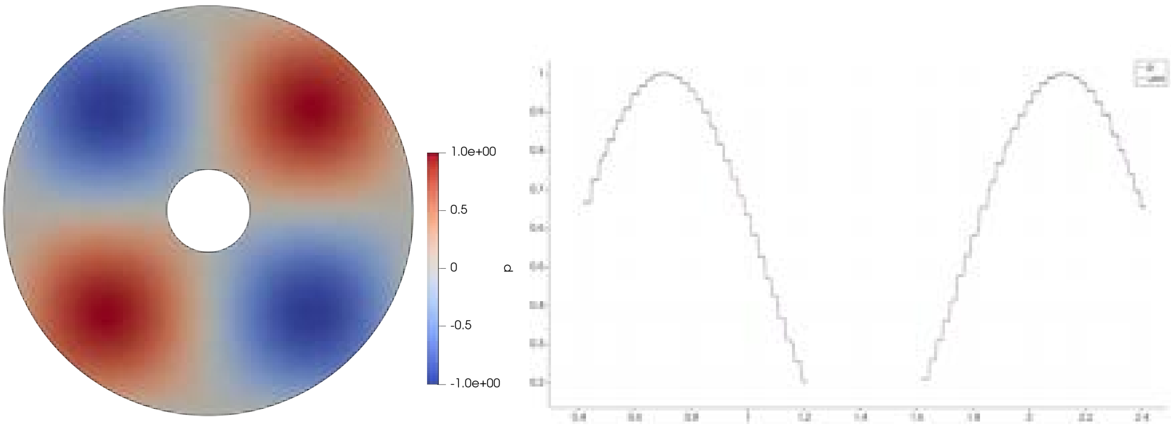


Figure 14: Convergence graph of the constant solution (144) at $T = 1$

Let us now represent the trigonometric solution (144) and its diagonal section on an annulus of center $O = (1, 1)$ and radius $R = 1$. The solution is approximated with Neumann boundary conditions. This way we can verify the definition of the normal $\mathbf{n}_r$ at the "angle" nodes from Definition 10. The formula for the normal at the "angles" is given in remark 8. With this test, we want to verify that this formula is true when three adjacent points of the boundary are not aligned and when they do not form a right angle. As one can see on Figure 15 the solution and its diagonal section on the annulus are correct. Thus the formula for $\mathbf{n}_r$ at the angles is verified in that case.

Remark 21. Note that the linear solution (143) was also tested on this domain and the scheme gave the exact solution for Neumann and Dirichlet boundary conditions. For Dirichlet boundary conditions this result is explained by the fact that there are no "corner" nodes that belong to only one cell in that mesh, thus there is no approximation of β_r at the corner nodes that breaks the exactness of the scheme. Moreover this result shows that the definition of $\mathbf{n}_r$ at the "angle" nodes preserves the exactness of the scheme.



(a) Allure of the solution (b) Diagonal section of the solution

Figure 15: Solution (144) on an annulus mesh of quads

7.2 Non stationary solutions

For this section we found the two following test-cases in Andrei D. Polyanin's Handbook of linear partial differential equations [18].

7.2.1 Linear solution

Let us consider the following stationary solution

$$p(\mathbf{x}, t) = Ax^2 + By^2 + 2(A + B)t \tag{145}$$

with a constant diffusion coefficient $\kappa = I_d$ and $A = 2$, $B = 1$.

Then we have

$$f(\mathbf{x}, t) = 0 \quad \text{and} \quad \begin{cases} g(\mathbf{x}, t) = \begin{pmatrix} 2Ax \\ 2By \end{pmatrix} .\mathbf{n}_\Gamma & \forall\ \mathbf{x} \in \Gamma_D \\ h(\mathbf{x}, t) = Ax^2 + By^2 + 2(A + B)t & \forall\ \mathbf{x} \in \Gamma_N \end{cases}$$

Once again the numerical order of convergence in space is 2 for every boundary conditions as one can see on Figure 16.

7.2.2 Non-linear solution

Let us consider the following stationary solution

$$p(\mathbf{x}, t) = Ae^{-\mu x - \lambda y} \cos(\mu x - 2\mu^2 t + C_1) \sin(\lambda y - 2\lambda^2 t + C_2) \tag{146}$$

with a constant diffusion coefficient $\kappa = I_d$ and $A = 2$, $C_1 = \lambda = 1$, $C_2 = 1.5$ and $\mu = 0.5$.

Then we have

$$f(\mathbf{x}, t) = 0$$

and

$$\begin{cases} g(\mathbf{x}, t) = \begin{pmatrix} -\mu Ae^{-\mu x - \lambda y} \cos(\lambda y - 2\lambda^2 t + C_2)\big(\cos(\mu x - 2\mu^2 t + C_1) + \sin(\mu x - 2\mu^2 t + C_1)\big) \\ -\lambda Ae^{-\mu x - \lambda y} \cos(\mu x - 2\mu^2 t + C_1)\big(\cos(\lambda y - 2\lambda^2 t + C_2) + \sin(\lambda y - 2\lambda^2 t + C_2)\big) \end{pmatrix} .\mathbf{n}_\Gamma & \forall\ \mathbf{x} \in \Gamma_D \\ h(\mathbf{x}, t) = Ae^{-\mu x - \lambda y} \cos(\mu x - 2\mu^2 t + C_1) \sin(\lambda y - 2\lambda^2 t + C_2) & \forall\ \mathbf{x} \in \Gamma_N \end{cases}$$

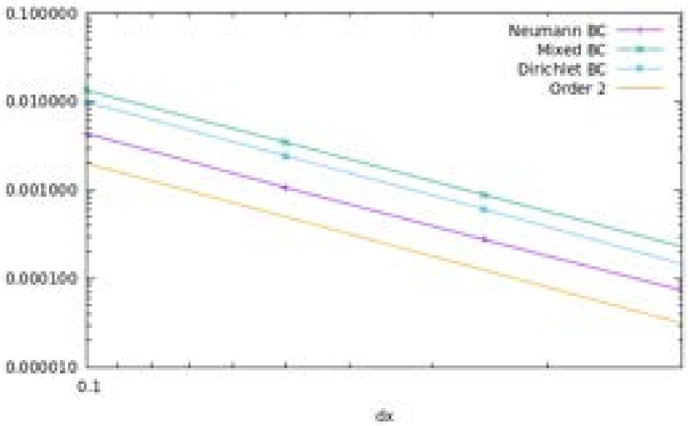


Figure 16: Convergence graph of the linear unstationary solution (145) at $T = 1$

Once again the numerical order of convergence in space is 2 for every boundary conditions as one can see on Figure 17.

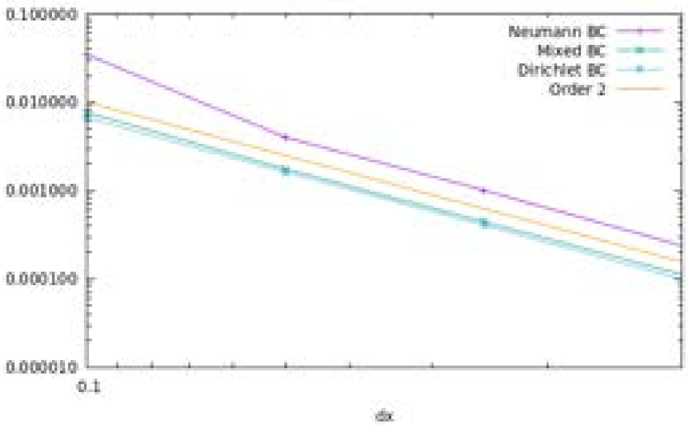


Figure 17: Convergence graph of the non linear unstationary solution (146) at $T = 1$

7.3 Non-constant and tensorial diffusion coefficient

Let us now consider the following solution

$$p(\mathbf{x}, t) = t + \sin(\pi x)\sin(\pi y) \tag{147}$$

With a diffusion coefficient that depends on space:

$$\kappa(\mathbf{x}) = \begin{pmatrix} \cos^2(x+y) + 2\sin^2(x+y) & -\cos(x+y)\sin(x+y) \\ -\cos(x+y)\sin(x+y) & \sin^2(x+y) + 2\cos^2(x+y) \end{pmatrix} \tag{148}$$

Remark 22. Note that $\kappa = ODO^t$ with 0 an orthogonal matrix and D a diagonal matrix defined as follows

$$O = \begin{pmatrix} \cos(x+y) & -\sin(x+y) \\ \sin(x+y) & \cos(x+y) \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

This way, κ has for eigenvalues 1 and 2, thus it is positive. Moreover it is symmetric and bounded: $|||\kappa(\mathbf{x})|||_\infty = \max(|\cos^2(x+y)+2\sin^2(x+y)|+|-\cos(x+y)\sin(x+y)|, |\cos(x+y)\sin(x+y)|+|\sin^2(x+y)+2\cos^2(x+y)|) \leq 4$, thus $|||\kappa|||_{\Omega,\infty} = \max_{\mathbf{x}\in\Omega}(|||\kappa(\mathbf{x})|||_\infty) \leq 4$.

Then we can build the second member and the boundary conditions

$$\begin{aligned} f(\mathbf{x},t) = 1 + 3\pi^2 \sin(\pi x)\sin(\pi y) - \pi\cos(\pi x)\sin(\pi y)(2\cos(x+y)\sin(x+y) + \sin^2(x+y) - \cos^2(x+y)) \\ +2\pi^2\cos(\pi x)\cos(\pi y)\cos(x+y)\sin(x+y) + \pi\sin(\pi x)\cos(\pi y)(\sin^2(x+y) \\ - \cos^2(x+y) - 2\sin(x+y)\cos(x+y)) \end{aligned}$$

and

$$\begin{cases} g(\mathbf{x},t) = \kappa(\mathbf{x}) \begin{pmatrix} \pi\cos(\pi x)\sin(\pi y) \\ \pi\sin(\pi x)\cos(\pi y) \end{pmatrix} .\mathbf{n}_\Gamma & \forall\ \mathbf{x} \in \Gamma_D \\ h(\mathbf{x},t) = t + \sin(\pi x)\sin(\pi y) & \forall\ \mathbf{x} \in \Gamma_N \end{cases}$$

One can see on Figure 18 that the numerical order of convergence in space is 2 for every boundary conditions. Note that the Dirichlet and mixed boundary conditions graphs are overlaid on each other.

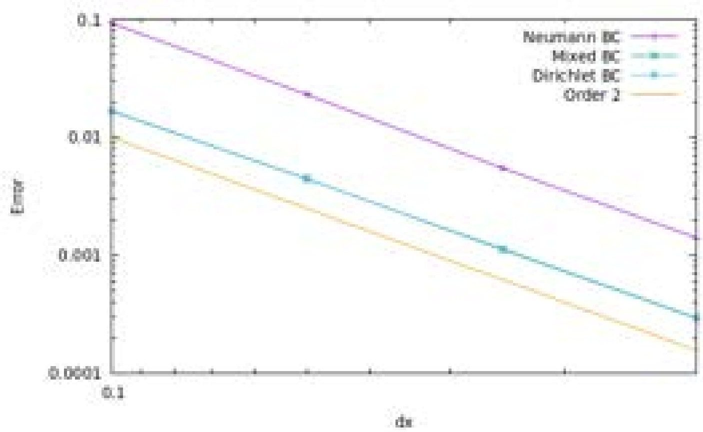


Figure 18: Convergence graph of the solution (147) with a diffusion coefficient that depends on space at $T = 1$

7.4 Crank-Nicolson method

Thanks to Crank-Nicolson method, the time order of error of the scheme is supposed to be the same as its space order (Order 2). Let us see if this property is numerically verified.

Let us chose $\Delta t \approx h$ and study the L^2 errors of the solutions on meshes going from $N = 10$ to 160 cells per side.

In this section, we will use the solutions mentioned earlier and their corresponding boundary conditions, defined above.

First of all with the linear stationary solution

$$p(\mathbf{x},t) = 5x + y + 2 \tag{149}$$

with a constant diffusion coefficient $\kappa = I_d$, the scheme with Neumann boundary conditions is exact, so we represent its convergence order with Dirichlet and mixed boundary conditions. One can see on Figure 19 that the convergence graph tends to the order 2.

Let us continue with another stationary solution, but this time trigonometric:

$$p(\mathbf{x},t) = \sin(\pi x)\sin(\pi y) \tag{150}$$

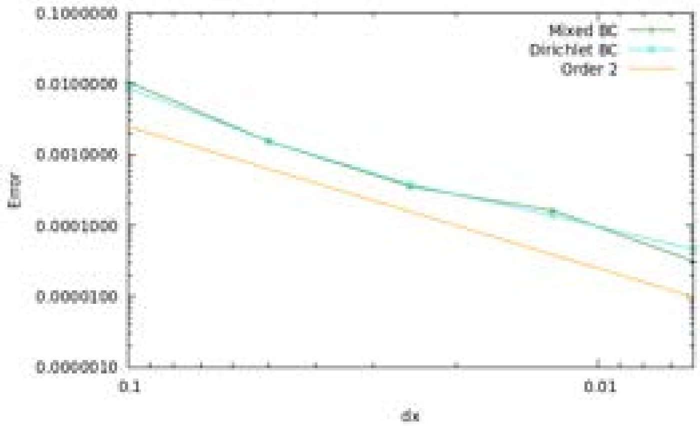


Figure 19: Convergence graph for the linear solution (149)

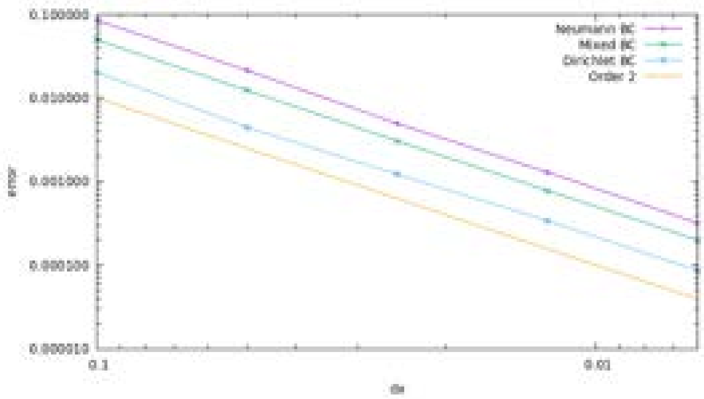


Figure 20: Convergence graph for the trigonometric solution (150)

with a constant diffusion coefficient $\kappa = I_d$.

Once again, as expected order 2 is found in time and space on Figure 20.
Let us now consider the following unstationary linear solution

$$p(\mathbf{x}, t) = Ax^2 + By^2 + 2(A + B)t \tag{151}$$

with a constant diffusion coefficient $\kappa = I_d$ and $A = 2$, $B = 1$.

One has order 2 in time and space as one can see on Figure 21.

For the second unstationary solution, we will study:

$$p(\mathbf{x}, t) = Ae^{-\mu x - \lambda y} \cos(\mu x - 2\mu^2 t + C_1) \sin(\lambda y - 2\lambda^2 t + C_2) \tag{152}$$

with a constant diffusion coefficient $\kappa = I_d$ and $A = 2$, $C_1 = \lambda = 1$, $C_2 = 1.5$ and $\mu = 0.5$.

One has order 2 in time and space as one can see on Figure 22.

Finally, let us consider another unstationary solution

$$p(\mathbf{x}, t) = t + \sin(\pi x) \sin(\pi y) \tag{153}$$

with a diffusion coefficient that depends on space:

$$\kappa(\mathbf{x}) = \begin{pmatrix} \cos^2(x + y) + 2\sin^2(x + y) & -\cos(x + y)\sin(x + y) \\ -\cos(x + y)\sin(x + y) & \sin^2(x + y) + 2\cos^2(x + y) \end{pmatrix} \tag{154}$$

Once again order 2 is found in time and space on Figure 23.

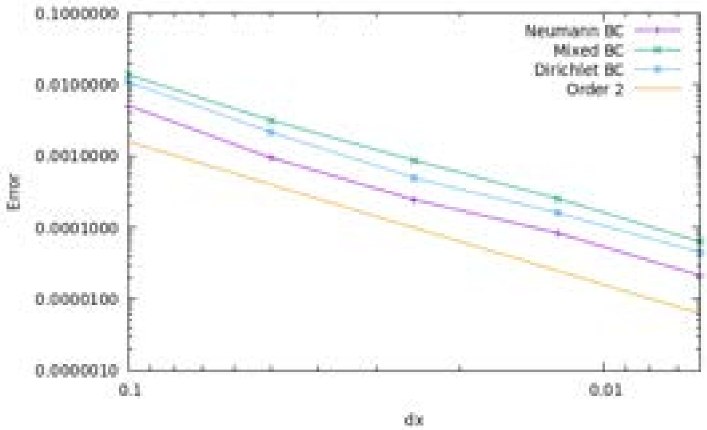


Figure 21: Convergence graph for the unstationary linear solution (151)

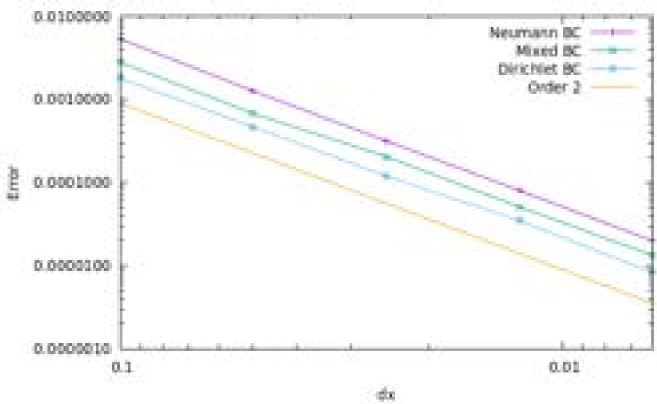


Figure 22: Convergence graph for the unstationary trigonometric solution (152)

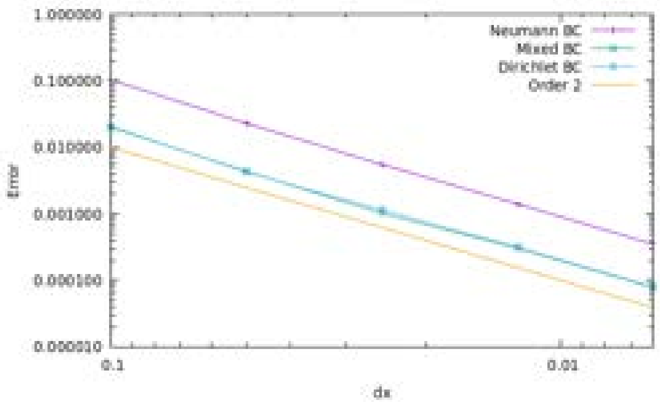


Figure 23: Convergence graph for the solution with a non constant κ (153)

7.5 Spurious modes

As it is done in [5], let us study the scheme on Cartesian mesh. In that case the fluxes $\mathbf{C}_{jr}$ point in the direction of the diagonals of the square Ω_j, therefore the adjacent cells are not impacted by Ω_j, only the diagonal ones are.

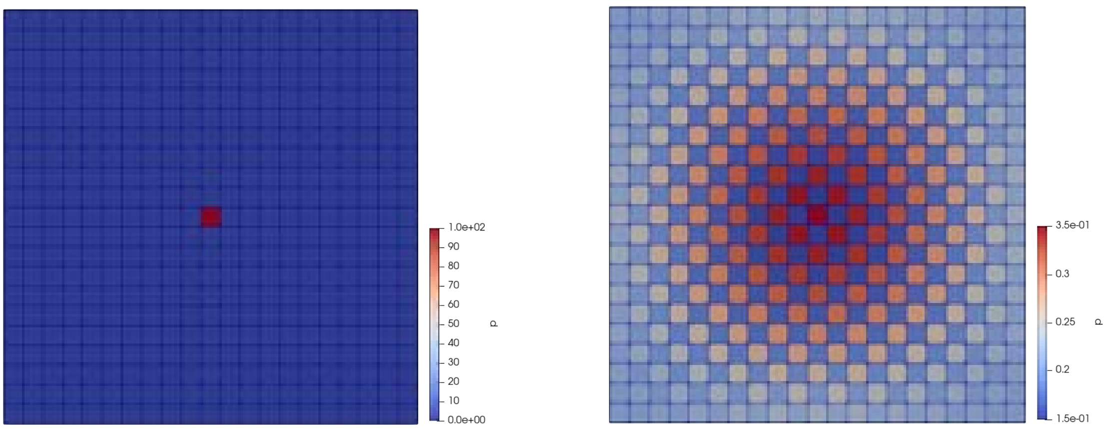
We can easily spot this problem by choosing a Dirac mass for initial condition. In Figure 24a we see the initialization of the problem where only the center cell Ω_j is non-null. Let us chose the following initial condition:

$$p_0(\mathbf{x}) = 100 \times \mathbb{1}_{\mathbf{x} \in \Omega_j}$$

with homogeneous Neumann boundary conditions, on a Cartesian mesh with 21 cell per side.

With this initial data we should obtain the fundamental solution of the heat equation. Its formula is given in [18].

However as one can see on Figure 24b, spurious modes appear immediately at the beginning of the simulation and they only fade away when the non-null values reach the boundary of the domain.



(a) Initialization (b) T = 0.14

Figure 24: Spurious mode on a Cartesian mesh

Let us now study the same problem on a non Cartesian mesh.

The initialization is the same as for the Cartesian mesh, but on Figure 25b to Figure 25d, notice that the situation is fixed quickly (after about 6 time iterations).

Another way to fix this problem is given in [5].

It consists in slightly changing the definition of $\mathbf{C}_{jr}$: one does not consider the middle of the edges $(\mathbf{x}_{r-1}, \mathbf{x}_r)$ and $(\mathbf{x}_r, \mathbf{x}_{r+1})$ called respectively $\mathbf{x}_{r-\frac{1}{2}}$ and $\mathbf{x}_{r+\frac{1}{2}}$ but points shifted by $\pm\alpha$ where $\alpha = \frac{1}{6}h$ with h the length of each edge. The choice of point on an edge must be the same for two cells having the edge in common. Then for quadrangular meshes there are 16 groups of $\mathbf{C}_{jr}$ that can be chosen.

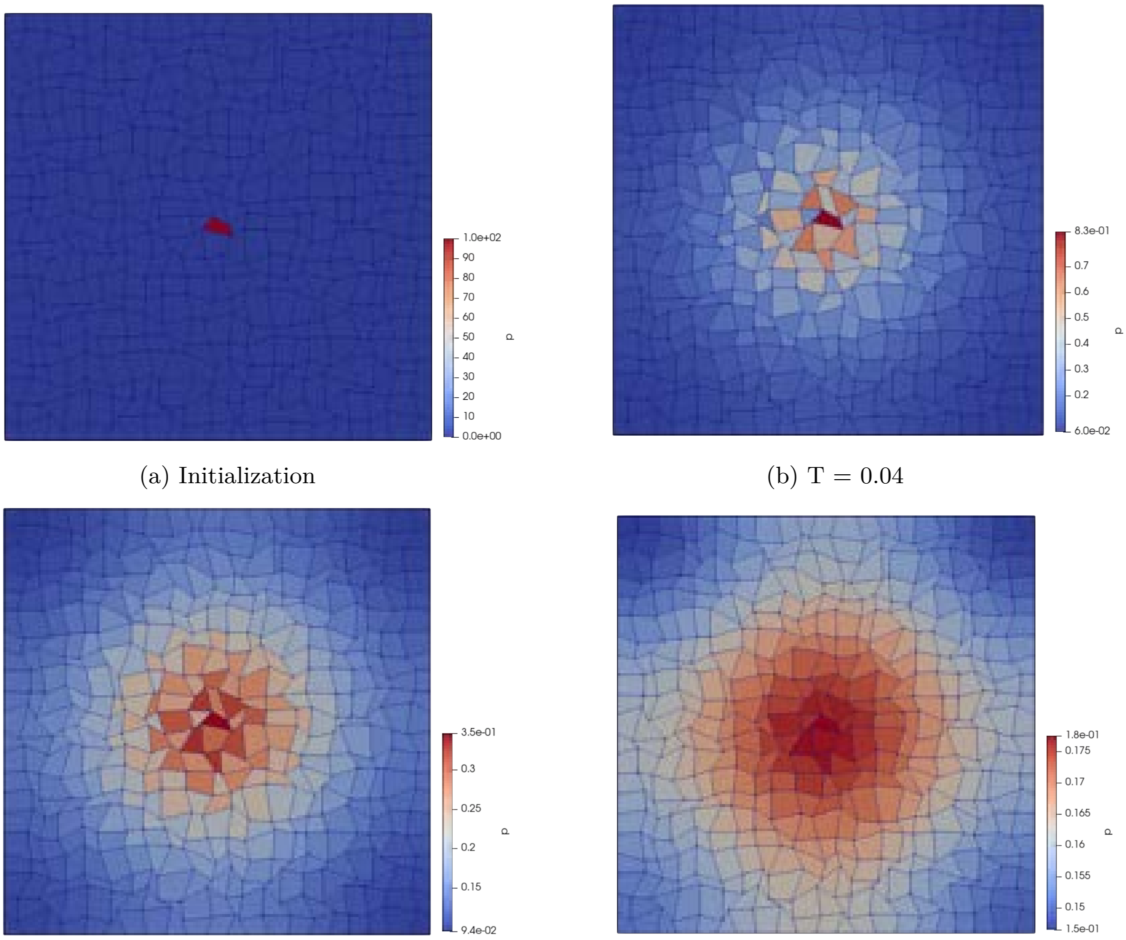
In practice let $(a, b) \in \{\frac{1}{3}, \frac{2}{3}\}^2$. Let us now redefine the middle of the edges: $\mathbf{x}_{r-\frac{1}{2}}$ becomes $\mathbf{x}_r + a(\mathbf{x}_{r-1} - \mathbf{x}_r)$ and $\mathbf{x}_{r+\frac{1}{2}}$ becomes $\mathbf{x}_r + b(\mathbf{x}_{r+1} - \mathbf{x}_r)$. This way the two following identities are preserved inside the domain

$$\sum_{j \in \mathcal{C}_r} \mathbf{C}_{jr} = 0 \quad \text{and} \quad \sum_{r \in \mathcal{V}_j} \mathbf{C}_{jr} = 0 \tag{155}$$

Finally the scheme is defined by the average over the four choices for (a, b).

7.6 Triangle meshes

After studying this scheme on a square domain with meshes of quads, we are interested in testing the method on meshes of triangles. To do so, we will consider the following meshes of triangles, cartesian and random. Note that for the Cartesian mesh, the corner cells are quads in order to avoid the forbidden mesh seen on Figure 7.



(a) Initialization

(b) T = 0.04

(c) T = 0.07

(d) T = 0.14

Figure 25: Spurious mode on a non-Cartesian mesh

In this section the solutions are represented at the time $T = 10$ with a constant time step of $\Delta t = 0.1$. Tests have been made and reducing the time step at every refinement of the mesh doesn't impact the convergence order, hence we chose to keep Δt constant so that the computation time doesn't explode for the finest mesh.

The domain remains the same as presented in the introduction of the section, such as the L^2 error formula.

7.6.1 Linear solutions

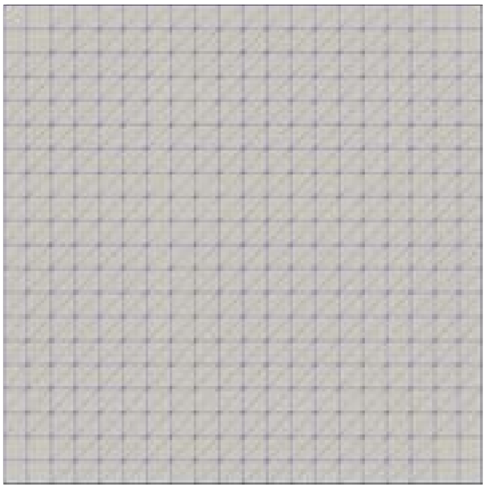
Let us once again consider the linear solution (143):

$$p(\mathbf{x}, t) = 5x + y + 2 \tag{156}$$

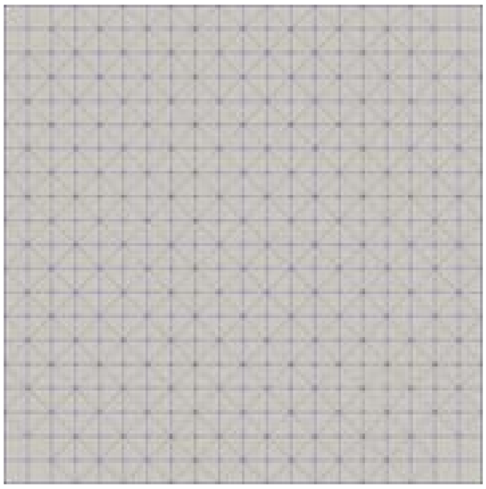
with a constant diffusion coefficient $\kappa = I_d$ and the same second member and boundary conditions as in Section 7.1.1.

In this case the scheme is exact, therefore we won't calculate its order of convergence but one can see its L^2 error on different meshes and for different boundary conditions on Table 1.

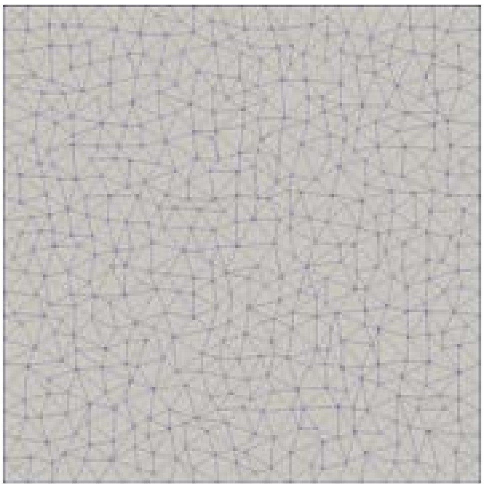
Remark 23. At the first iteration the errors are on a 10^{-15} scale which is the machine precision. The slight augmentation of this error at the final time is due to approximations errors at every time step. We



(a) Cartesian left triangle mesh



(b) Cartesian alternated triangle mesh



(c) Random triangle mesh

Figure 26: Examples of triangle meshes for $N = 20$ triangles per side

	Cartesian left triangles	Cartesian alternated triangles	Random triangles
Dirichlet BC	1.95×10^{-14}	4.3×10^{-13}	9.3×10^{-14}
Neumann BC	1.2×10^{-13}	3.89×10^{-13}	1.41×10^{-13}
Mixed BC	5.83×10^{-14}	3.96×10^{-13}	6.33×10^{-14}

Table 1: L^2 error of the linear solution at $T = 10$ for $N = 20$ cell per side

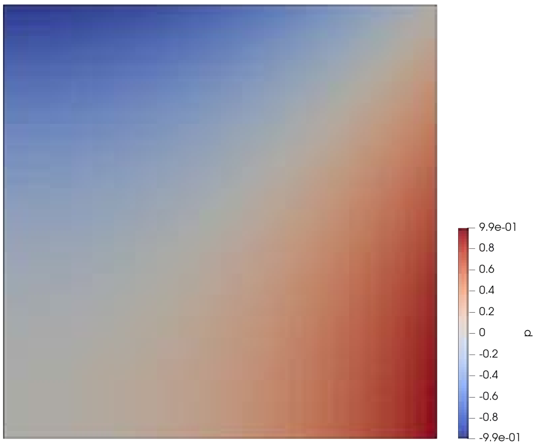
can still assert that the scheme is exact in this case.

7.6.2 Quadratic solution

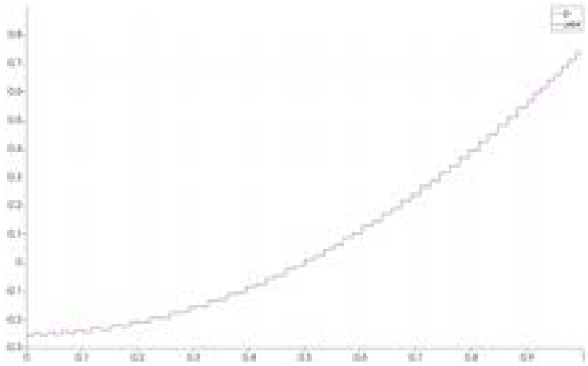
Let us consider the quadratic solution:

$$p(\mathbf{x}, t) = x^2 - y^2 \tag{157}$$

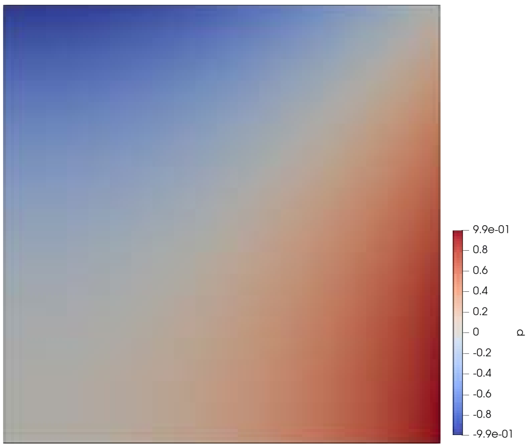
In Figure 27, one can see that the quadratic solution is very well approximated by the scheme as the horizontal sections on the left sided triangle and the random triangle meshes show. Indeed the approximation and the exact solution overlay each other, and the final error with Dirichlet boundary conditions and $T = 10$ is of order 10^{-4}, which is satisfying.



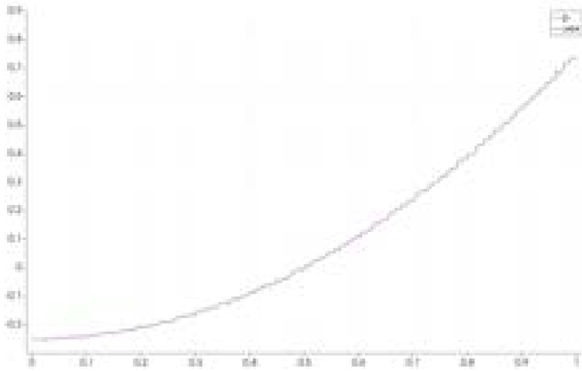
(a) Cell value on the left-sided triangle mesh - $N = 80$



(b) Horizontal section on the left-sided triangle mesh



(c) Cell value on the random triangle mesh - $N = 80$



(d) Horizontal section on the random triangle mesh

Figure 27: Allure of the solution (157) and its horizontal section at $T = 10$ with $N = 80$

7.6.3 Trigonometric solution

Let us now consider the trigonometric solution

$$p(\mathbf{x}, t) = \sin(\pi x)\sin(\pi y) \tag{158}$$

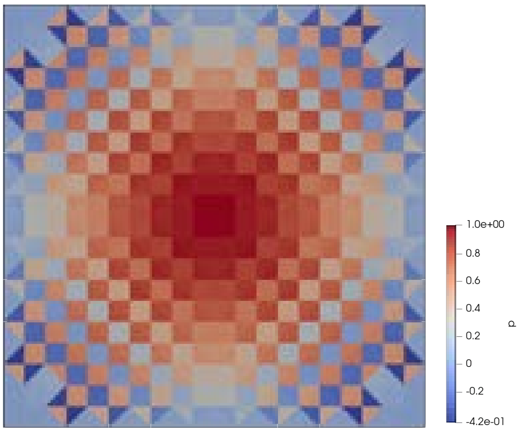
with a constant diffusion coefficient $\kappa = I_d$.

Once again the second member and the boundary conditions are the same as in Section 7.1.2.

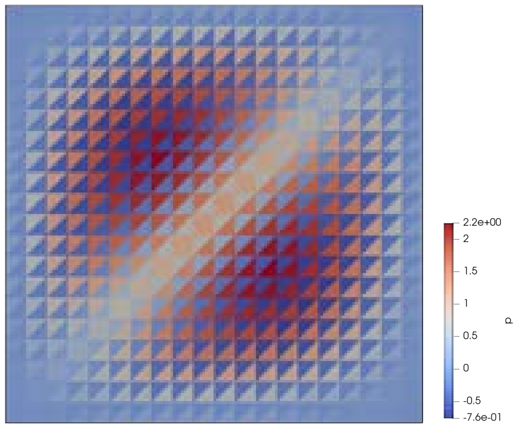
In the first place, let us observe the solution obtained on a Cartesian mesh with alternated or left-sided triangles and Dirichlet Boundary conditions. One can see on Figure 28a and Figure 28b that some kind of spurious modes appear even though the initial condition is regular (in opposition with the spurious modes on the quad mesh, where the initial condition had to be a Dirac mass for them to appear).

Let us consider the average solution at the nodes of the mesh, calculated with the values of their neighbour cells. To do so we use a median dual mesh on which the center of each cell corresponds to the nodes of the primal mesh. The solutions obtained can be seen on Figure 28c and Figure 28d. On that new mesh the spurious modes disappear.

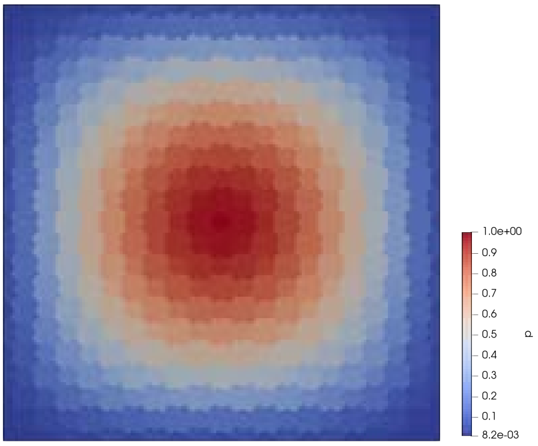
Refining the solution on a mesh of $N = 80$ cell per side makes the spurious modes fade away in the case of the alternated triangles (see Figure 28e), but it worsens them on the refined left sided triangles mesh (Figure 28f).



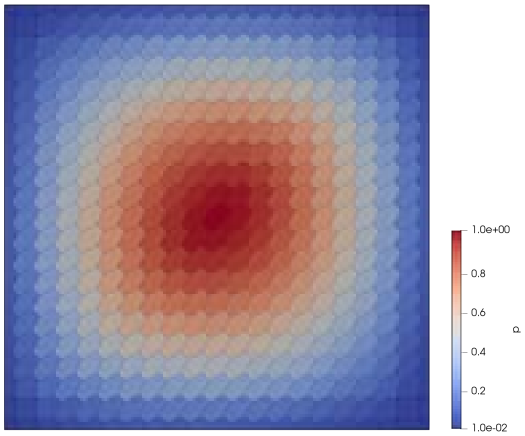
(a) Cell value on the alternated triangle mesh - $N = 20$



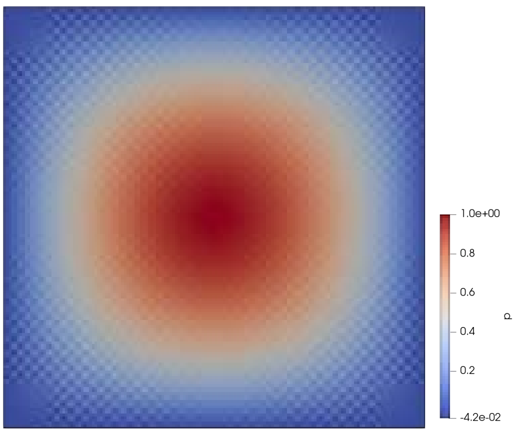
(b) Cell value on the left-sided triangle mesh - $N = 20$



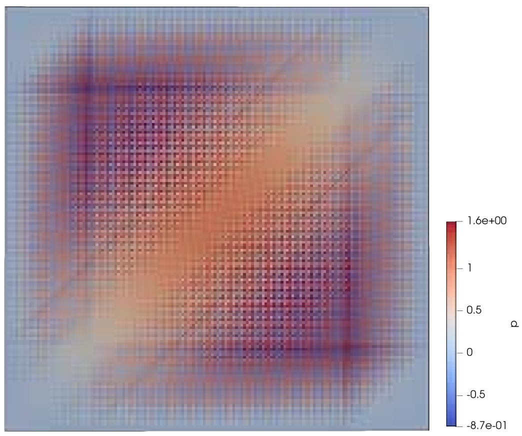
(c) Node mean-value on the alternated triangle dual mesh - $N = 20$



(d) Node mean-value on the left-sided triangle dual mesh - $N = 20$



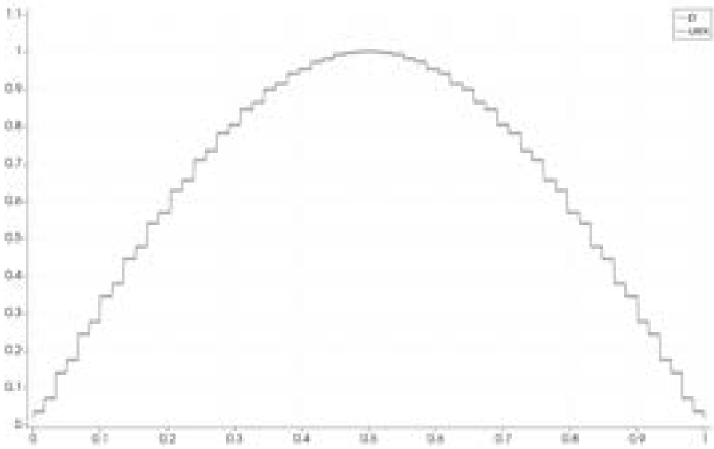
(e) Cell value on the alternated triangle mesh - $N = 80$



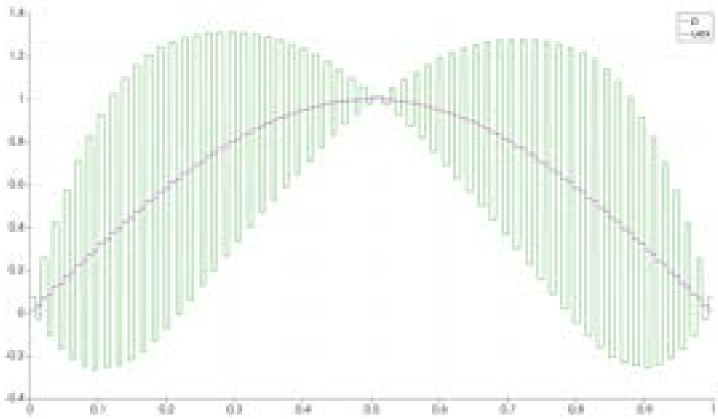
(f) Cell value on the left-sided triangle mesh - $N = 80$

Figure 28: Allure of the solution (158) on the cells and its node mean-value at $T = 10$

On Figure 29a and Figure 29b one can see the shape of the solutions on the horizontal section of the domain $\{x = [0, 1],\ y = 0.5\}$. The solution on the alternated triangle mesh is almost overlaid with the exact solution while on the left sided triangles mesh the approximated solution has a completely different shape from the exact solution.



(a) Alternated triangle mesh



(b) Left sided triangle mesh

Figure 29: Horizontal section of the solution (158) at $T = 10$ with $N = 80$

Remark 24. In order to explain this result, let us first redefine the node values β_r. In the particular case of a Cartesian triangle mesh, one can rewrite $\beta_r = |V_r| I_d$. V_r is defined in Definition 11 and can be seen in Figure 30 in the particular case of the Cartesian left-sided triangle mesh.
This new writing of β_r comes from Proposition 7 where we stated that

$$\beta_r = |V_r| I_d + P$$

where

- $P = \sum_{j \in \mathcal{C}_r} \frac{1}{2} \left(\mathbf{v}^{\perp}_{j+\frac{1}{2}} \otimes \mathbf{v}_{j+\frac{1}{2}} - \mathbf{w}^{\perp}_{j-\frac{1}{2}} \otimes \mathbf{w}_{j-\frac{1}{2}} \right)$
- $\mathbf{w}_{j-\frac{1}{2}} = \mathbf{x}_j - \mathbf{x}_{j-\frac{1}{2}}$
- $\mathbf{v}_{j+\frac{1}{2}} = \mathbf{x}_{j+\frac{1}{2}} - \mathbf{x}_j$
- I_d is the identity matrix.

In the Cartesian meshes of triangles, one has $\mathbf{v}_{j+\frac{1}{2}} = \mathbf{w}_{j+\frac{1}{2}}$, therefore the terms of P cancel each other two by two. We are left with $P = 0$ and we retrieve the new definition of β_r.

Tanks to this new definition, it is immediate to see that the flux between some cells that share an edge in a Cartesian triangle mesh is null in the case of an isotropic diffusion coefficient (take $\kappa_r = \lambda I_d$ for example). Indeed, as one can see in Figure 30, $\mathbf{C}_{jr}$ and $\mathbf{C}_{kr}$ are orthogonal, therefore the flux in the matrix M is null between the cell j and k:

$$(M^i)_{jk} = -\sum_{r\in\mathcal{V}^i_{jk}} \langle \kappa_r \beta_r^{-1} \mathbf{C}_{kr}, \mathbf{C}_{jr} \rangle = -\sum_{r\in\mathcal{V}^i_{jk}} \frac{\lambda}{|V_r|} \langle \mathbf{C}_{kr}, \mathbf{C}_{jr} \rangle = 0$$

The same phenomenon appears at the boundary of the mesh where some cells aren't impacted by the boundary condition at one of their nodes that is on the boundary.

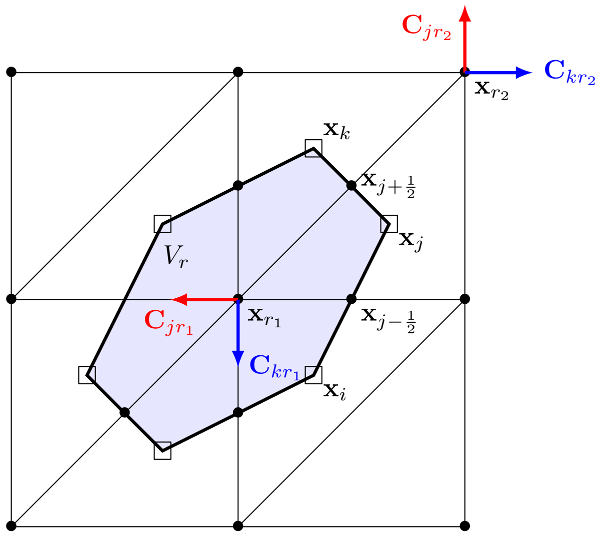


Figure 30: Control volume V_r on a left-sided triangle mesh and orthogonal $\mathbf{C}_{jr}$'s

Finally, let us observe the convergence graphs of the scheme on the primal (cell value) and the dual (node mean-value) meshes. On Figure 31a one can see that on a mesh with alternated triangles the L^2 error of the solution with Dirichlet or Neumann boundary conditions are the same. The order of convergence is 2 with the cell values and the node mean-values, even though the average solution at the nodes is six to sixty times more precise than the cell value one.
On the other hand, on the left-sided triangle mesh of Figure 31b, the cell value solution converges with an order 1 towards the exact solution (for Neumann and Dirichlet Boundary conditions), while the average value at the nodes converges with an order between 1.5 and 2 for the Dirichlet boundary conditions and 2 with the Neumann ones.

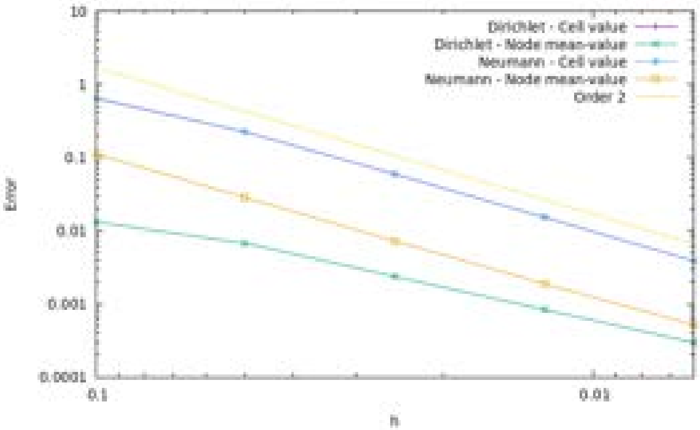
Let us now study the same solution on the random triangle mesh.

One can see on Figure 32a that the solution on the random triangle mesh presents a lot of incorrect patterns. However when we consider the mean value of the solution at the nodes it seems to be better. On this figure the solution is represented with Neumann boundary conditions but the result is the same with Dirichlet ones.

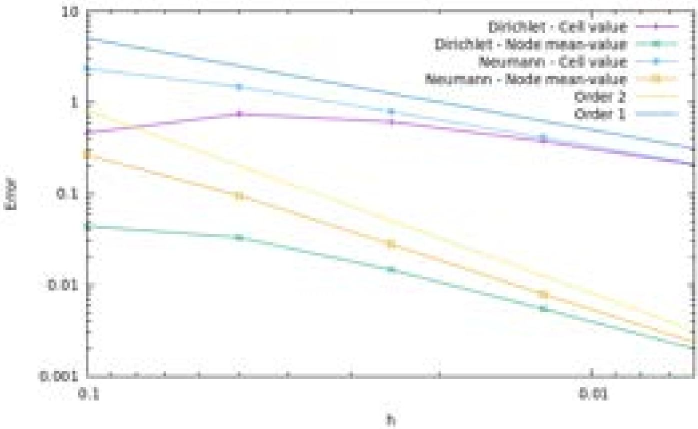
Let us now focus on the convergence graph of this problem with Neumann and Dirichlet boundary conditions. One can see on Figure 32c that the order of convergence of the solution with Dirichlet boundary conditions is 1 and the one for Neumann boundary conditions is 1.4.

As one can see on Figure 33a the solution on the refined mesh ($N = 80$ cell per side of the domain) still presents the same patterns as in Figure 32a. Therefore even though it converges with an order 1.4, the solution on a refined mesh is not smooth and the conclusion is the same for Dirichlet boundary conditions.

One can see the difference between the nodal scheme on a triangle mesh, the same scheme on a random quad mesh and the Diamond scheme on a the same triangle mesh represented in Figure 33b and Fig-



(a) Alternated triangle mesh



(b) Left-sided triangle mesh

Figure 31: L^2 error graphs for the solution (158)

ure 33c. The irregularities of the solution only appear in the case of the triangle mesh for the nodal scheme.

The convergence graphs of these three problems is presented in Figure 33d. The order of convergence of the Diamond scheme on triangles and the Nodal scheme on quads is 2 while it is 1 for the Nodal scheme on triangles.

On Figure 34a and Figure 34b one can see the shape of the solutions on the horizontal section of the domain $\{x = [0,1],\ y = 0.5\}$. The solution on the random triangle mesh is full of variations while the one on a random quad mesh is more steady and it almost overlays the exact solution.

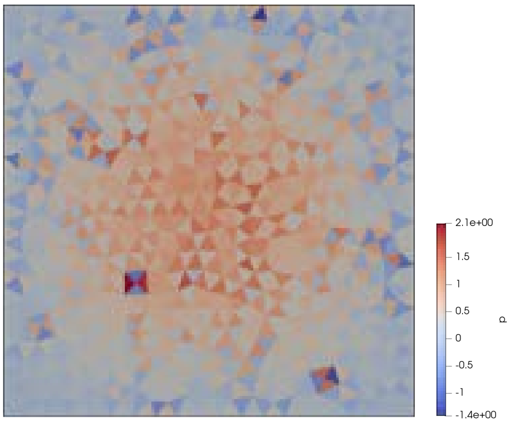
Finally, let us summarize the convergence graphs obtained for the trigonometric solution on the three triangle meshes in Figure 35. One has order 1 on the left-aligned triangle mesh, order 1.4 on the random triangle mesh and order 2 on the alternated triangle mesh.

Remark 25. As a reminder we did not prove in this document that the solution of the scheme is unique in this particular case (unstructured mesh and non-null source term).

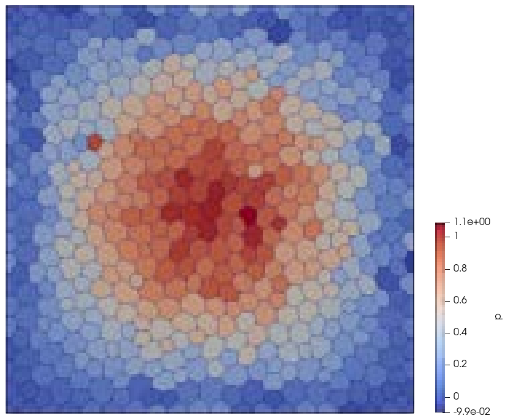
7.6.4 Non stationary solutions

Let us study the following unstationary solution:

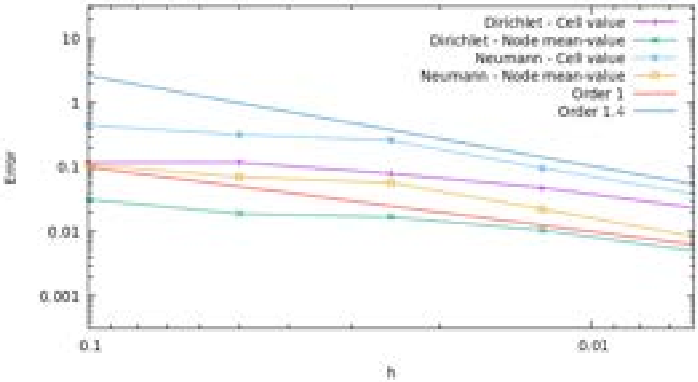
$$p(\mathbf{x}, t) = Ae^{-\mu x - \lambda y} \cos(\mu x - 2\mu^2 t + C_1) \sin(\lambda y - 2\lambda^2 t + C_2) \tag{159}$$



(a) Cell value - Primal mesh



(b) Average node value - Dual mesh



(c) L^2 error graphs for the solution (158) on a random triangle mesh

Figure 32: Allure of the solution (158) at $T = 10$ with $N = 20$ and Neumann BC

with a constant diffusion coefficient $\kappa = I_d$ and $A = 2$, $C_1 = \lambda = 1$, $C_2 = 1.5$ and $\mu = 0.5$.

The second member and the boundary conditions are the same as in Section 7.1.2.

One can see on Figure 36b that the unstationary solution (159) does not present the spurious modes as clearly as they appear on the same mesh for the trigonometric solution, and its diagonal section in Figure 36b confirms that the approximation is correct. Likewise, the same unstationary solution on the random triangle mesh (Figure 36c) is smooth and the diagonal section is of the approximation is close to the one of the exact solution (see Figure 36d).

One can see in Figure 36e that the order of convergence of the nodal scheme is 1 for the left-aligned triangle mesh, 1.4 on the random triangle mesh and 2 on the alternated triangle mesh, as for the trigonometric solution.

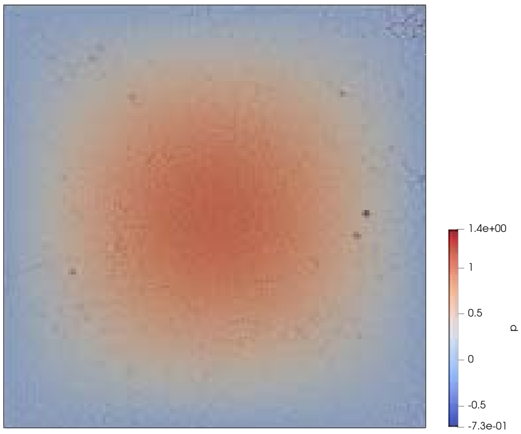
Let us now consider a second unstationary solution:

$$p(\mathbf{x}, t) = x^3 - 6xt \tag{160}$$

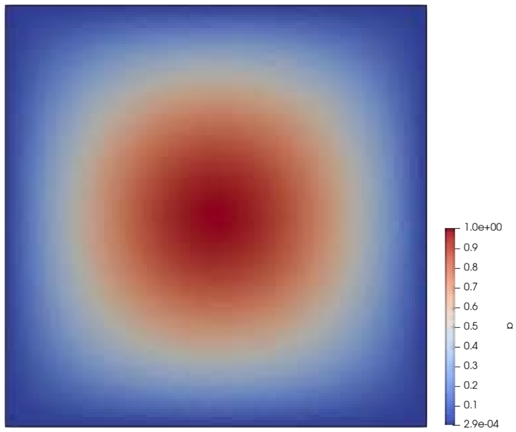
with $\kappa = I_d$ and

$$f(\mathbf{x}, t) = -12x \quad \text{and} \quad \begin{cases} g(\mathbf{x}, t) = \begin{pmatrix} 3x^2 - 6t \\ 0 \end{pmatrix} .\mathbf{n}_\Gamma & \forall\, \mathbf{x} \in \Gamma_D \\ h(\mathbf{x}, t) = x^3 - 6xt & \forall\, \mathbf{x} \in \Gamma_N \end{cases}$$

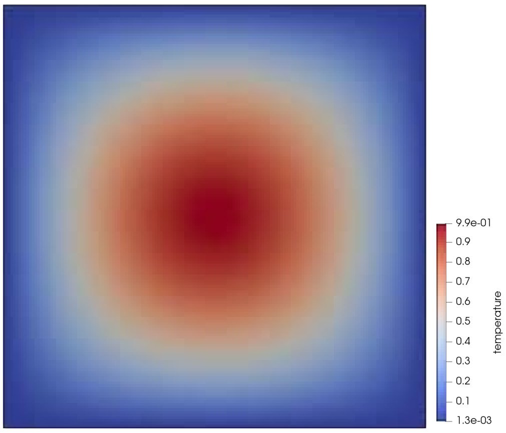
The second unstationary solution does not present any spurious modes and the approximation is correct on the left-sided and random triangle meshes as one can see in Figure 37.



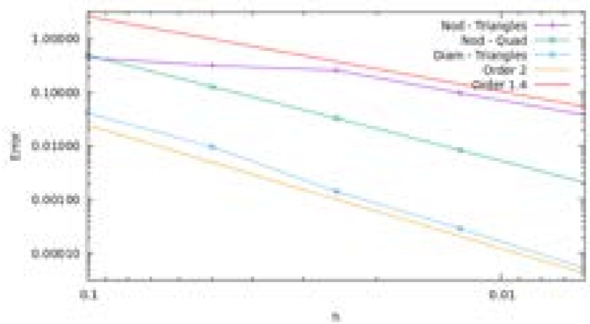
(a) Nodal scheme on random triangles



(b) Nodal scheme on random quads



(c) Diamond method on random triangles



(d) Comparison of the convergence graphs

Figure 33: Allure of the solution (158) on different meshes with different methods

Let us now represent the solution (160) and its diagonal section on an annulus discretized with a random triangle mesh. As one can see on Figure 38 the solution and its diagonal section with Neumann boundary conditions on the annulus are correct.

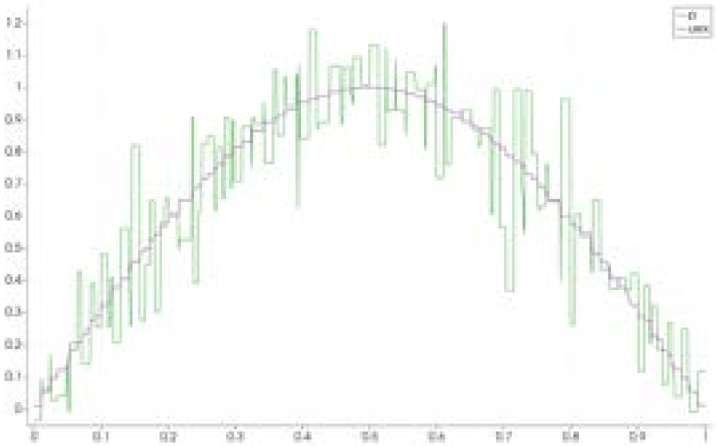
7.7 Hybrid method

Once the hybrid method from Section 6 is implemented, we want to test its numerical performances. To do so, we will begin with a constant $\lambda = 1$, testing the scheme on quad and triangle meshes for the same test cases as above. Then we will try it with $\lambda = h^n$, with different values of n so that the hybrid scheme converges at different speeds towards the nodal scheme when h tends to 0.

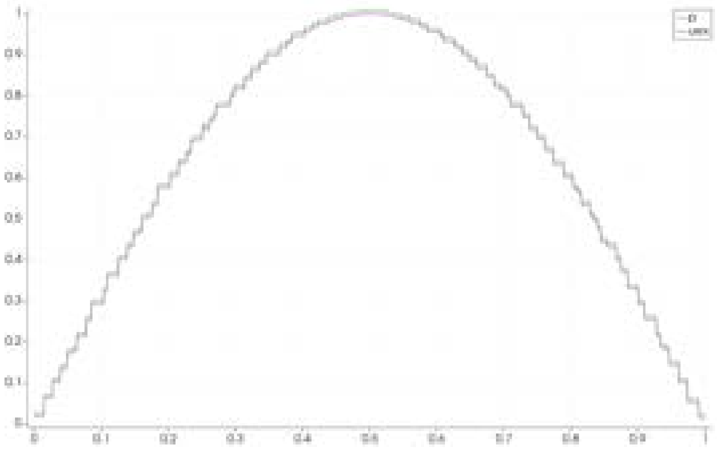
7.7.1 Hybrid scheme with $\lambda = 1$

For the linear solution, as one can see on Table 2 the hybrid scheme is exact for $\lambda = 1$ on the Cartesian quad mesh, the left aligned and the random triangle meshes for Neumann and Dirichlet boundary conditions. The scheme is not exact for Dirichlet boundary conditions on random quad meshes because of the approximation $\beta_r \approx \beta_r^c$ at the corners.

Let us now consider the allure of the trigonometric solution and the order of convergence of the scheme on triangle meshes, which was causing trouble on triangle meshes with the purely nodal scheme.



(a) Random triangle mesh



(b) Random quad mesh

Figure 34: Horizontal section of the solution (158) at $T = 10$ with $N = 80$

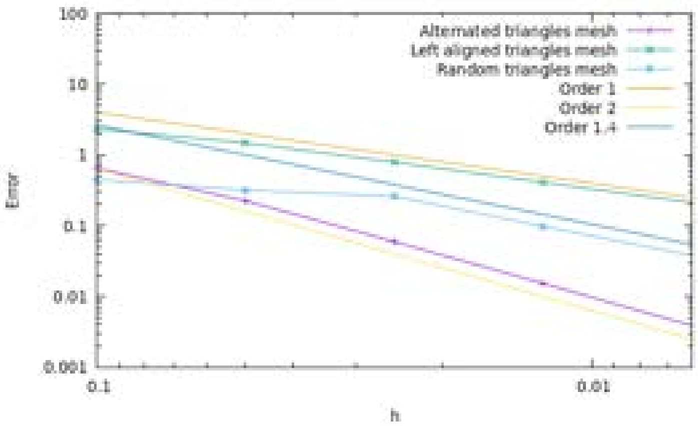
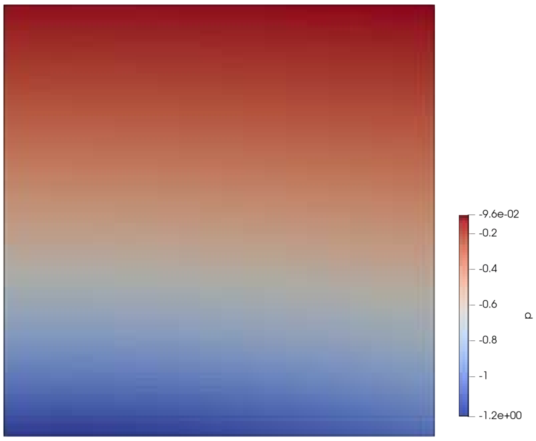


Figure 35: Convergence graph of the trigonometric solution on triangle meshes

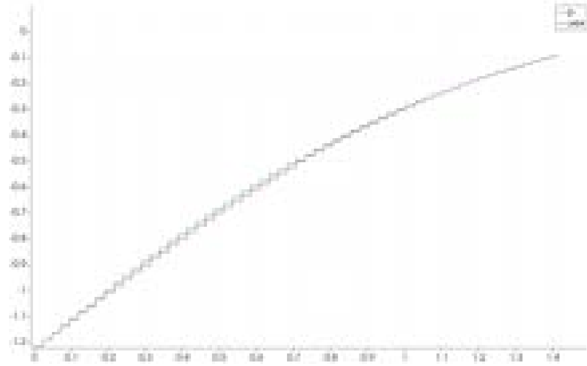
As a reminder, the trigonometric test-case is the following:

$$p(\mathbf{x}, t) = \sin(\pi x)\sin(\pi y) \tag{161}$$

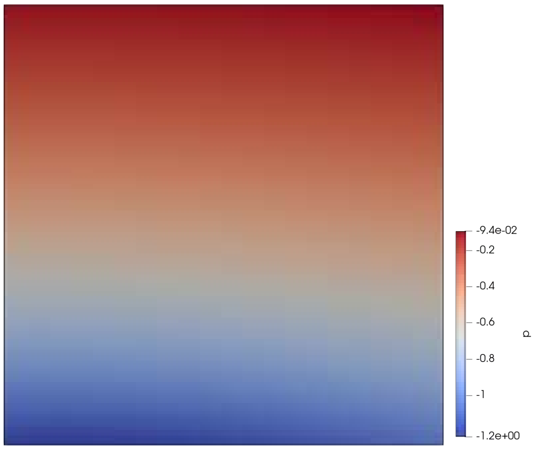
with a constant diffusion coefficient $\kappa = I_d$.



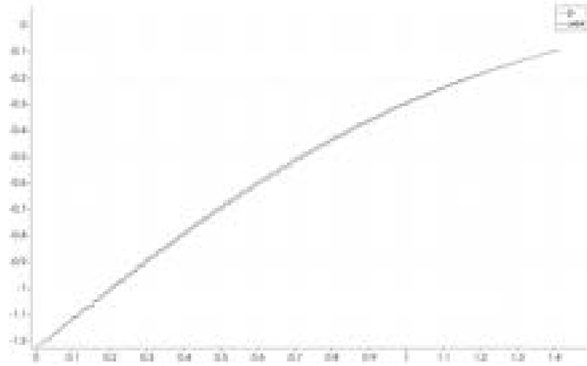
(a) Cell value on the left-sided triangle mesh - $N = 80$



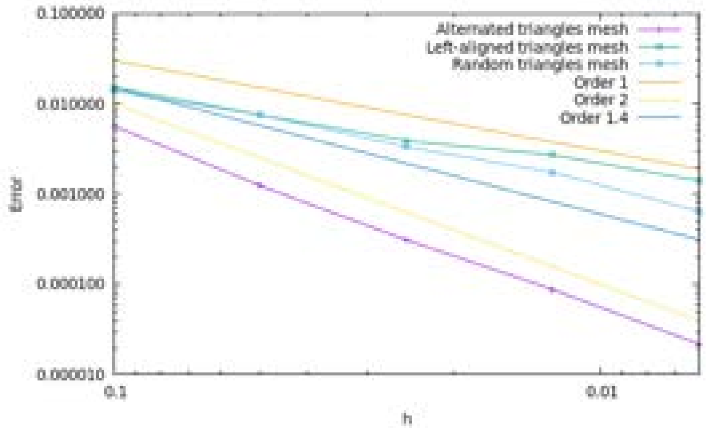
(b) Diagonal section on the left-sided triangle mesh



(c) Cell value on the random triangle mesh - $N = 80$



(d) Diagonal section on the random triangle mesh

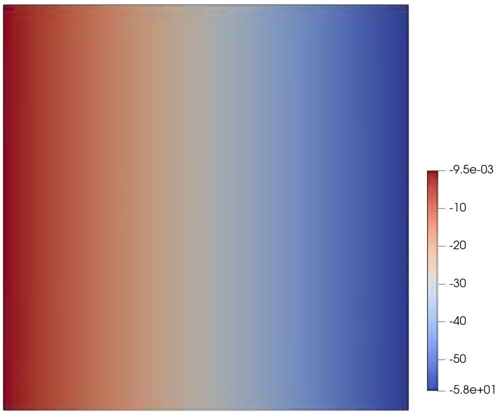


(e) Convergence graph of the non stationary solution on triangle meshes

Figure 36: Allure of the solution (159) and its diagonal section at $T = 10$ with $N = 80$

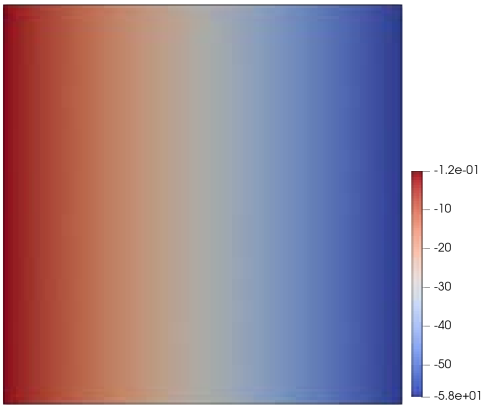
The second member and the boundary conditions are the same as in Section 7.1.2.

One can see on Figure 39 that the hybrid scheme with $\lambda = 1$ gives us smooth solutions on triangle meshes and the horizontal sections of the approximated solution overlay the exact solution. One does not see the oscillations from the nodal scheme anymore. Moreover one can see on Figure 40 that the order

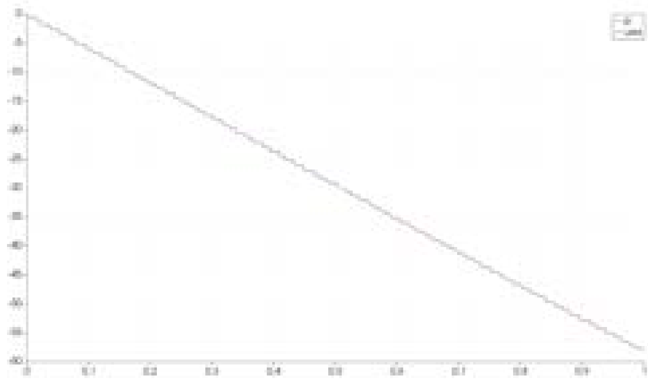


(a) Cell value on the left-sided triangle mesh - $N = 80$

(b) Horizontal section on the left-sided triangle mesh

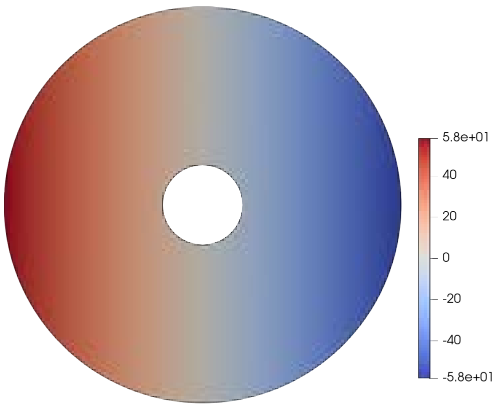


(c) Cell value on the random triangle mesh - $N = 80$

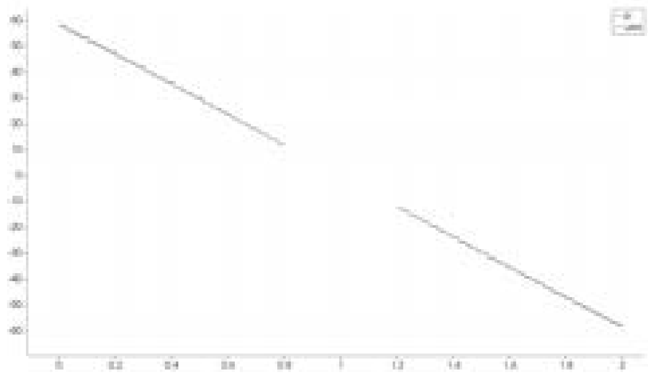


(d) Horizontal section on the random triangle mesh

Figure 37: Allure of the solution (160) and its horizontal section at $T = 10$ with $N = 80$



(a) Allure of the solution



(b) Diagonal section of the solution

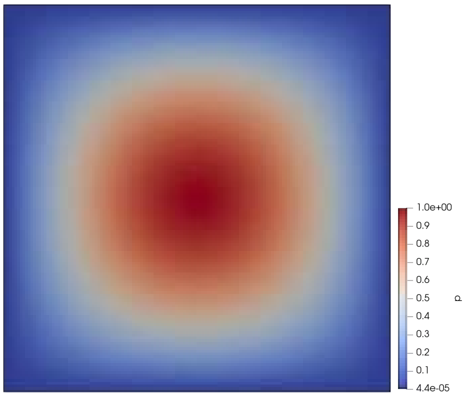
Figure 38: Solution (160) on an annulus mesh of triangles

of convergence of this scheme is 2 for any quad or triangle mesh.

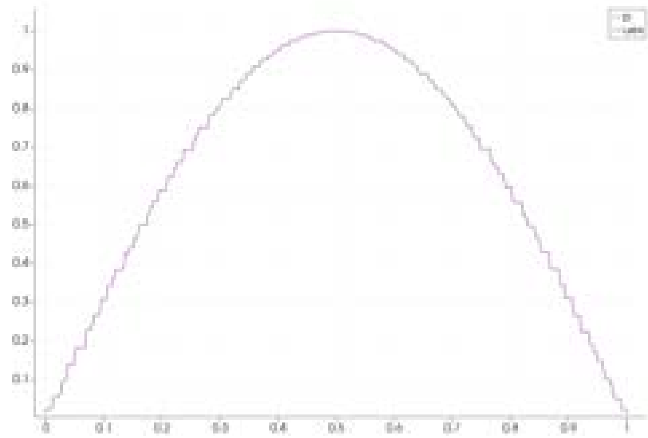
Remark 26. We chose to only represent the solution and the order of convergence of the scheme for the trigonometric problem on random and left aligned triangle meshes since they were the ones causing the most trouble with the nodal scheme. However tests have been made for the hybrid scheme with $\lambda = 1$ on all the quad and triangle meshes mentioned earlier and on all the test cases (linear, quadratic, non

	Cartesian quad mesh	Left aligned triangle mesh	Random triangle mesh
Dirichlet BC	4.05×10^{-16}	1.91×10^{-15}	1.3×10^{-15}
Neumann BC	1.82×10^{-14}	2.16×10^{-13}	1.96×10^{-13}
Mixed BC	5.32×10^{-16}	4.54×10^{-15}	2.08×10^{-15}

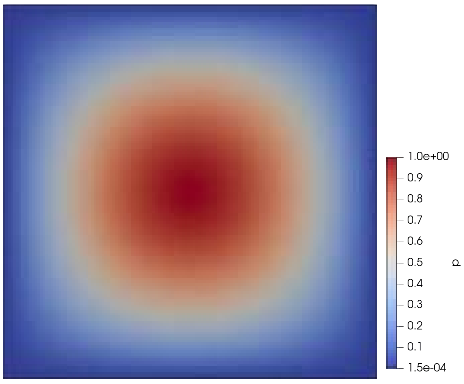
Table 2: L^2 error of the linear solution for the hybrid scheme at $T = 10$ for $N = 20$ cell per side



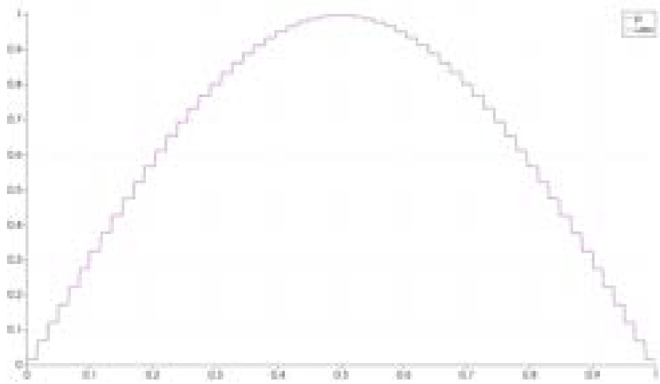
(a) Allure of the solution on the random triangle mesh



(b) Horizontal section on the random triangle mesh



(c) Allure of the solution on the left aligned triangle mesh



(d) Horizontal section on the left aligned triangle mesh

Figure 39: Allure of the solution (161) and its horizontal section at $T = 10$ with $N = 80$ obtained with the hybrid scheme

stationary...) and solutions are smooth and the orders of convergence are always 2.

7.7.2 Hybrid scheme with $\lambda = h^n$

Let us now consider the hybrid scheme with $\lambda = h^n$ with $n \geq 0$. As a reminder, when $\lambda = 0$ the scheme is purely nodal which means when $\lambda = h^n$ the scheme converges towards the nodal scheme at the speed of n as the mesh is refined.

Let us begin with the linear solution. The hybrid scheme is exact for any value of λ since it is a convex combination of two exact schemes for the linear solutions. For the same reason the scheme with $\lambda = h^n$ is order 2 in space for the quadratic and non stationary solutions on any type of quad or triangle mesh.

Let us now go back to the spurious modes problem and see if the hybrid scheme resolves it. Let us

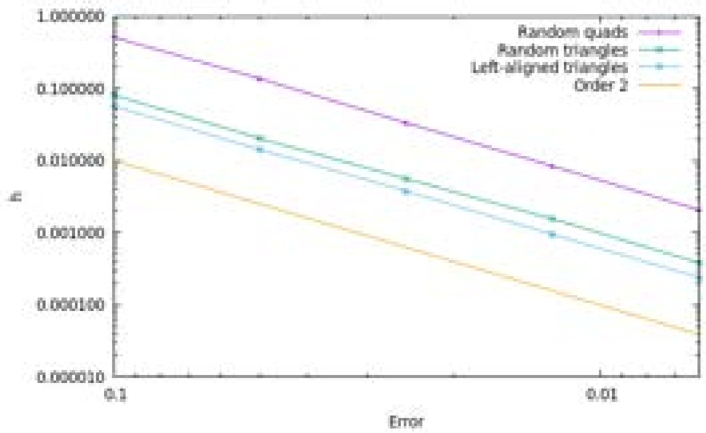


Figure 40: Convergence graph of the hybrid scheme for the trigonometric solution (161) at $T = 10$ for Neumann boundary conditions

consider a Cartesian quad mesh, with for initialization

$$p_0(\mathbf{x}) = 100 \times \mathbb{1}_{\mathbf{x}\in\Omega_j}$$

and homogeneous Neumann boundary conditions.

One can see without any surprise on Figure 41b that the spurious modes disappear when $\lambda = 1$ since all the fluxes are spread through the faces of the cells. When $\lambda = h^{\frac{1}{2}}$, there are still no spurious modes appearing (see Figure 41c). However when $\lambda = h$ one can slightly see them on Figure 41d at $T = 0.14$ but they will disappear as T increases. Finally on Figure 41e when $\lambda = h^2$, the spurious modes are almost as bad as with the purely nodal scheme (Figure 41f) and they don't disappear with time.

Therefore one can consider that the hybrid scheme corrects the spurious modes up to $\lambda = h^n$ where $n \in [0, \dfrac{1}{2}]$.

Let us now focus on the trigonometric solution

$$p(\mathbf{x}, t) = \sin(\pi x)\sin(\pi y)$$

on triangle meshes for different values of λ.

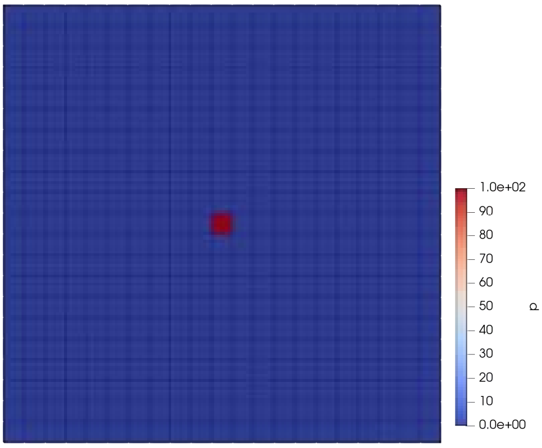
We know that for $\lambda = 1$ the solution is smooth and the problems of the purely nodal scheme do not appear. We will visualise the solution of the scheme with $\lambda = h^n$, $n > 0$ on Figure 42. Let us begin with $n = \frac{1}{2}$ so that the hybrid scheme tends to the nodal one but not too fast. One can see on Figure 42a and Figure 42b that the solution is smooth and we can not spot any pattern or oscillation on the horizontal section of the solution. The result is the same with $\lambda = h$ and $\lambda = h^2$ as seen in Figure 42c and Figure 42e. However when $\lambda = h^3$ the parasitic patterns appear once again in Figure 42g and the horizontal section in Figure 42h is full of oscillations.

One can deduce from these results that when $\lambda = h^3$ the scheme converges too fast towards the purely nodal scheme. Therefore we will study the order of convergence of the scheme for $\lambda = h^n$, $n \in \{0, \frac{1}{2}, 1, 2\}$ in Table 3 for the random triangle mesh and in Table 4 for the left aligned triangle mesh.

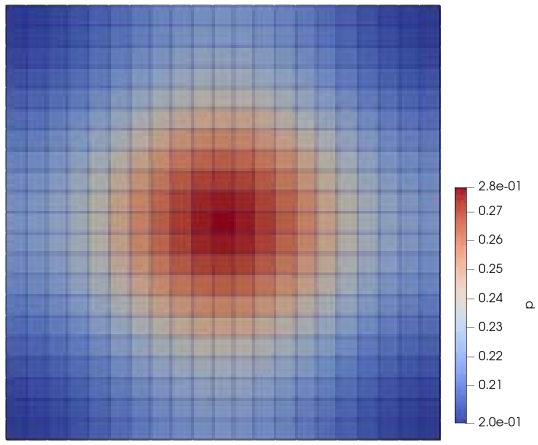
In order to calculate the order of convergence let us use the following formula:

$$q = \frac{log(e_1/e_2)}{log(N_2/N_1)}$$

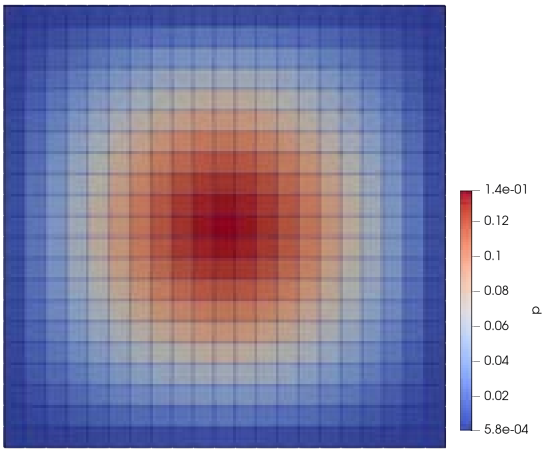
where $e_{1,2}$ are respectively the errors associated with the meshes with N_1 and N_2 cells per side.



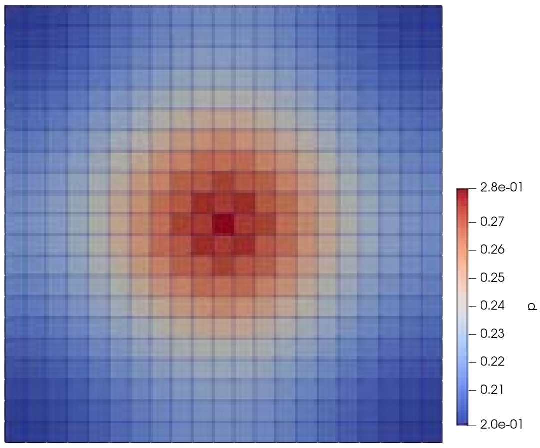
(a) Initialization



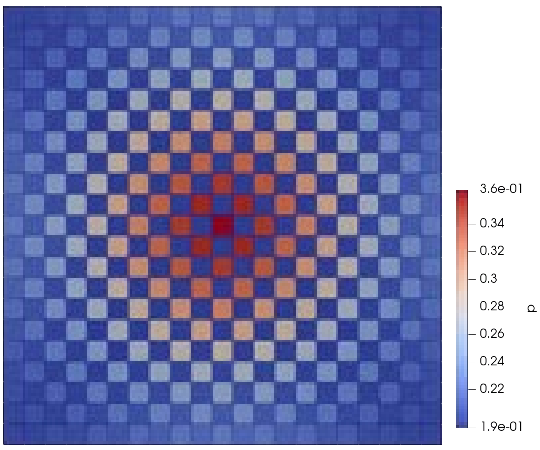
(b) $\lambda = 1$ - $T = 0.14$



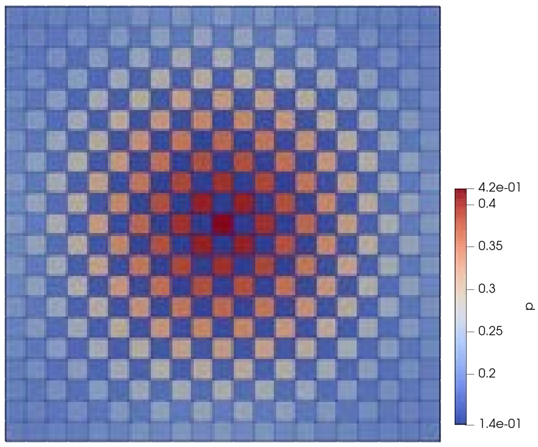
(c) $\lambda = h^{\frac{1}{2}}$ - $T = 0.14$



(d) $\lambda = h$ - $T = 0.14$



(e) $\lambda = h^2$ - $T = 0.14$



(f) $\lambda = 0$ - $T = 0.14$

Figure 41: Spurious modes appearing depending the value of λ

As one can notice in Table 3, the order of convergence of the hybrid scheme on random triangle mesh tends to 2 when $\lambda \to 1$. That is what we expected since the scheme is order 2 in space when $\lambda = 1$. From $\lambda = h$ the order of convergence is superior to 1.8 which means the hybrid scheme raises significantly the order of the purely nodal scheme.

	Dirichlet BC	Neumann BC
$\lambda = 0$	1.19	1.34
$\lambda = h^2$	1.27	1.42
$\lambda = h$	1.8	1.91
$\lambda = h^{\frac{1}{2}}$	1.93	1.98
$\lambda = 1$	1.95	2

Table 3: Order of convergence between $N_1 = 80$ and $N_2 = 160$ for the trigonometric solution on the random triangle mesh

On the left-aligned triangle mesh (Table 4) the results are even more extreme: the order of convergence of the hybrid scheme is 1 for $\lambda \in \{0, h\}$ and it switches to 2 (or almost 2) as soon as $\lambda = h^n,\ n \leq 1$.

	Dirichlet BC	Neumann BC
$\lambda = 0$	0.86	0.95
$\lambda = h^2$	0.97	1.14
$\lambda = h$	1.97	2
$\lambda = h^{\frac{1}{2}}$	1.99	2
$\lambda = 1$	1.99	2

Table 4: Order of convergence between $N_1 = 80$ and $N_2 = 160$ for the trigonometric solution on the left-aligned triangle mesh

Remark 27. The order of convergence of the scheme for the trigonometric solution is only 0.86 between $N = 80$ and $N = 160$ cell per side but it becomes 0.93 with one more refinement. Thus in that case the order of convergence still tends to 1.

8 Conclusion

In this document we have built the nodal diffusion scheme with a non constant tensorial diffusion coefficient and Neumann, Dirichlet and mixed boundary conditions that preserve the order of the periodical scheme. It was not an easy task because of the nodal fluxes. We have also built the Euler implicit time discretization and the Crank-Nicolson one in order to extend it to second order in time. We have also proved that the solution of the scheme is unique on structured meshes with homogeneous Dirichlet boundary conditions. Then we proved that the scheme with no source is exact for linear solutions with Dirichlet and Neumann boundary conditions. Finally, we proved under some conditions of positivity and coercivity of $\kappa_r \beta_r$, the stability, consistency and convergence of the scheme with no source and periodic, homogeneous Dirichlet and homogeneous Neumann boundary conditions with Euler implicit and Crank-Nicolson time discretization. The theoretical convergence order obtained was 1 in space and up to 2 in time.

After implementing and testing this scheme on various test cases (linear, quadratic, trigonometric, non-stationary solutions and with a non constant tensorial diffusion coefficient) we obtained on the meshes of quads a numerical order of convergence of 2 in time and space. The only problem on quad meshes is the spurious modes appearing on Cartesian meshes when initializing the problem with the fundamental solution of the heat equation, but it is a very specific test because the initial condition is not a function (it is a measure). However when testing the scheme on meshes of triangle we had many more surprises. First of all the order of convergence in space depends on the considered mesh, but it stays between 1 and 2 so it remains coherent with the theoretical proof of convergence. However the scheme presents serious oscillations with the trigonometric test-case on triangle meshes even for regular initial conditions. Unfortunately we can't explain this behavior but we noticed that there was no flux between some cells that share an edge.

The idea was then to use an hybrid scheme that allowed some of the flux to pass through the faces and to make it converge towards the nodal scheme when the mesh is refined. This hybrid scheme gave us a good correction of the nodal scheme and we retrieved order 2 in space when $\lambda = h^n,\ n \leq 1$, where λ is the parameter of the convex combination between the nodal and the face fluxes.

A perspective of this work could be to finish the study of the hybrid scheme (theoretical study). Another idea would be to take interest in the the M_1 model. This non linear model approximates the transport equation and it is studied by B. Dubroca in [11] and C. D. Levermore in [15]. It has the particularity to be entropic, to preserve the maximum principle and to converge towards a diffusion equation. In his thesis [13], E. Franck has built and studied an AP scheme for the M_1 model with nodal fluxes. Theoretical and numerical properties of this scheme are also studied in [6]. Since it is asymptotic preserving it would be interesting to study the limit diffusion of the Lagrangian part of this nodal scheme with an implicit time discretization and non periodic boundary conditions, which means trying to prove that it is positive, conservative, stable, consistent and convergent.

A Useful formulas

General formulas and properties used in the document are listed and proved in this section.

Proposition 32. *Let A be a square matrix of size $d = 2$, and λ its eigenvalues. Then its characteristic polynomial writes*

$$P(\lambda) = \lambda^2 - \mathrm{tr}(A)\lambda + \det(A) \tag{162}$$

Proof. For $d = 2$, one can express the matrix A as $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Its characteristic polynomial writes

$$\begin{aligned}\det(A - \lambda I_d) &= (a-\lambda)(d-\lambda) - bc \\ &= ad - a\lambda - d\lambda + \lambda^2 - bc \\ &= \lambda^2 - (a+d)\lambda + (ad - bc) \\ &= \lambda^2 - \mathrm{tr}(A)\lambda + \det(A)\end{aligned}$$

and that ends the proof.

□

Proposition 33. *For any vector $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, one has*

$$\mathbf{u} \otimes \mathbf{v} = (\mathbf{v} \otimes \mathbf{u})^t \tag{163}$$

Proof. Thanks to Definition 4, one can write

$$(\mathbf{u} \otimes \mathbf{v})_{i,j} = u_i v_j = v_j u_i = (\mathbf{v} \otimes \mathbf{u})_{j,i}$$

Thus $\mathbf{u} \otimes \mathbf{v} = (\mathbf{v} \otimes \mathbf{u})^t$.

□

Proposition 34. *For any vector $\mathbf{u}, \mathbf{v}$ and $\mathbf{w} \in \mathbb{R}^d$, one has*

$$\langle \mathbf{u}, \mathbf{v} \rangle \mathbf{w} = (\mathbf{w} \otimes \mathbf{u})\mathbf{v} \tag{164}$$

Proof. By definition of the scalar product one has

$$\big(\langle \mathbf{u}, \mathbf{v} \rangle \mathbf{w}\big)_{i=1,\dots,d} = \Big(\sum_{j=1}^{d} u_j v_j w_i\Big)_{i=1,\dots,d}$$

On the other hand by definition of the tensor product and the matrix-vector product

$$\big((\mathbf{w} \otimes \mathbf{u})\mathbf{v}\big)_{i=1,\dots,d} = \Big(\sum_{i=1}^{d} (\mathbf{w} \otimes \mathbf{u})_{i,j} v_j\Big)_{i=1,\dots,d} = \Big(\sum_{i=1}^{d} w_i u_j v_j\Big)_{i=1,\dots,d}$$

Hence (164) is verified.

□

Proposition 35. *For any vector* $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^d$*, one has*

$$\operatorname{tr}(\mathbf{u} \otimes \mathbf{v}) = \langle \mathbf{u}, \mathbf{v} \rangle \tag{165}$$

Proof. By definition of the tensor product, one has

$$\operatorname{tr}(\mathbf{u} \otimes \mathbf{v}) = \sum_{i=1}^{d} (\mathbf{u} \otimes \mathbf{v})_{i,i} = \sum_{i=1}^{d} u_i v_i = \langle \mathbf{u}, \mathbf{v} \rangle$$

which gives us (165).

□

Proposition 36. *For any vector* $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^d$*, the tensor product* $\mathbf{u} \otimes \mathbf{v}$ *is a rank one matrix.*

Proof. Let us define $\mathbf{x}$ the kernel of $\mathbf{u} \otimes \mathbf{v}$ such that

$$\begin{aligned} & \mathbf{u} \otimes \mathbf{v}\mathbf{x} = 0 \\ \Leftrightarrow \quad & \langle \mathbf{v}, \mathbf{x} \rangle \mathbf{u} = 0 \end{aligned}$$

Then the space spanned by the orthogonal vectors to $\mathbf{v}$ is in the kernel of $\mathbf{u} \otimes \mathbf{v}$. Thus $\mathbf{u} \otimes \mathbf{v}$ is a rank one matrix.

□

In this part, let $\mathbf{n}$ be any vector in $\mathbb{R}^d$, and we chose $\mathbf{t}$ such that $\mathbf{t} = \mathbf{n}^\perp$, which means

$$\langle \mathbf{t}, \mathbf{n} \rangle = 0 \tag{166}$$

Proposition 37. *For any vector* $\mathbf{u} \in \mathbb{R}^d$ *and any matrix* $A \in \mathbb{R}^{d \times d}$*, one has*

$$\langle (\mathbf{t} \otimes \mathbf{t}) A (\mathbf{t} \otimes \mathbf{t}) \mathbf{u}, \mathbf{n} \rangle = 0 \tag{167}$$

Proof. Let us consider the left hand of (167). Using Proposition 34 two times in a row and thanks to (166), we obtain

$$\begin{aligned} \langle (\mathbf{t} \otimes \mathbf{t}) A (\mathbf{t} \otimes \mathbf{t}) \mathbf{u}, \mathbf{n} \rangle &= \langle (\mathbf{t} \otimes \mathbf{t}) A \langle \mathbf{t}, \mathbf{u} \rangle \mathbf{t}, \mathbf{n} \rangle \\ &= \langle \mathbf{t}, \mathbf{u} \rangle \langle (\mathbf{t} \otimes \mathbf{t}) A \mathbf{t}, \mathbf{n} \rangle \\ &= \langle \mathbf{t}, \mathbf{u} \rangle \langle \mathbf{t}, A\mathbf{t} \rangle \langle \mathbf{t}, \mathbf{n} \rangle \\ &= 0 \end{aligned}$$

And this ends the proof. □

Proposition 38. *For any vector* $\mathbf{u} \in \mathbb{R}^d$*, one has*

$$\langle (\mathbf{t} \otimes \mathbf{t}) \mathbf{u}, \mathbf{n} \rangle = \langle \mathbf{u}, (\mathbf{t} \otimes \mathbf{t}) \mathbf{n} \rangle = 0 \tag{168}$$

Proof. Using Proposition 34 and thanks to (166), one obtains

$$\langle (\mathbf{t} \otimes \mathbf{t}) \mathbf{u}, \mathbf{n} \rangle = \langle \mathbf{t}, \mathbf{u} \rangle \langle \mathbf{t}, \mathbf{n} \rangle = 0$$

And this ends the proof.

□

Proposition 39. *For any vector* $\mathbf{u} \in \mathbb{R}^d$ *and any matrix* $A \in \mathbb{R}^{d \times d}$*, one has*

$$\langle \mathbf{u}, A\mathbf{u} \rangle = \langle \mathbf{u}, A^s \mathbf{u} \rangle \tag{169}$$

with A^s *the symmetric part of* A *defined in Definition 6.*

Proof. Let us consider

$$\langle \mathbf{u}, A\mathbf{u} \rangle = \langle \mathbf{u}, \Big(\frac{A + A^t}{2} + \frac{A - A^t}{2} \Big) \mathbf{u} \rangle = \langle \mathbf{u}, \frac{A + A^t}{2} \mathbf{u} \rangle + \frac{1}{2} \langle \mathbf{u}, A\mathbf{u} \rangle - \frac{1}{2} \langle \mathbf{u}, A^t \mathbf{u} \rangle = \langle \mathbf{u}, A^s \mathbf{u} \rangle$$

because $\langle \mathbf{u}, A\mathbf{u} \rangle = \langle \mathbf{u}, A^t \mathbf{u} \rangle$ □

B Proofs

B.1 Proof of Proposition 2

Let us integrate the first equation of (1) over the domain Ω. It gives

$$\int_\Omega \partial_t p \, d\mathbf{x} - \int_\Omega \nabla.(\kappa \mathbf{u}) \, d\mathbf{x} = 0$$

Using the divergence formula this equation becomes:

$$\partial_t \int_\Omega p \, d\mathbf{x} - \int_\Gamma \langle \kappa \mathbf{u}, \mathbf{n} \rangle \, d\sigma = 0$$

Since we consider the scheme with homogeneous Neumann boundary conditions we obtain:

$$\partial_t \int_\Omega p \, d\mathbf{x} = 0$$

And that ends the proof.

B.2 Proof of Proposition 3

Let us begin with a multiplication of the first equation of (1) by p and an integration over Ω:

$$\int_\Omega p \partial_t p \, d\mathbf{x} - \int_\Omega p \nabla.(\kappa \mathbf{u}) \, d\mathbf{x} = 0$$

Using Green's formula, and since $\mathbf{u} = \nabla p$, we obtain:

$$\int_\Omega \frac{1}{2} \partial_t p^2 \, d\mathbf{x} + \int_\Omega \langle \mathbf{u}, \kappa \mathbf{u} \rangle \, d\mathbf{x} + \int_\Gamma p \langle \kappa \mathbf{u}, \mathbf{n}_\Gamma \rangle \, d\sigma = 0$$

Knowing that κ is positive by construction, one has

$$\int_\Omega \langle \mathbf{u}, \kappa \mathbf{u} \rangle \, d\mathbf{x} \geq 0$$

Then with the homogeneous boundary conditions one has $p|_\Gamma = 0$ or $\langle \kappa \mathbf{u}|_\Gamma, \mathbf{n}_\Gamma \rangle = 0$, therefore

$$\int_\Gamma p \langle \kappa \mathbf{u}, \mathbf{n}_\Gamma \rangle \, d\sigma = 0$$

Thus we obtain

$$\int_\Omega \frac{1}{2} \partial_t p^2 \, d\mathbf{x} = \frac{1}{2} \partial_t \|p\|_{L_2}^2 \leq 0$$

And that ends the proof.

C Construction of the matrices

C.1 Nodal scheme matrix

Let us rewrite the scheme in vectorial form with backward Euler method:

$$I_\Omega \frac{P^{n+1} - P^n}{\Delta t} - M P^{n+1} = B^{n+1} \tag{170}$$

where

- $P^n = \begin{pmatrix} p_1^n & \cdots & p_N^n \end{pmatrix}^t$ is known at each iteration, initialized with P^0 in the statement;
- $(I_\Omega)_{jk} = |\Omega_j| \mathbb{1}_{j=k}$ is the diagonal matrix with the volume of each cell j;

- $B^n = \begin{pmatrix} B_1^{n+1} & \cdots & B_N^{n+1} \end{pmatrix}^t$ is given in the statement of the problem;
- M is a matrix of size $N \times N$ to determine, with $N = |\mathcal{C}|$, the number of cells in the mesh.

In order to resolve this problem, we want to write it under the form

$$(I_\Omega - \Delta t M) P^{n+1} = \Delta t B^{n+1} + P^n \tag{171}$$

Let us now construct the matrix M.

Let us write $M = M^i + M^b + M^c$ with M^i the contribution of the vertices inside the domain to the matrix M, M^b the contribution of the boundary vertices to M and M^c for the corner vertices contribution. We can decompose the vector B with the same method.

C.1.1 Inside the domain

The second equation of (45) gives, for all $r \in \mathcal{V}^i$

$$\mathbf{u}_r^{n+1} = -\beta_r^{-1} \sum_{j \in \mathcal{C}_r} p_j^{n+1} \mathbf{C}_{jr}$$

Let us replace $\mathbf{u}_r^{n+1}$ in the first equation of (42), it becomes

$$|\Omega_j| \frac{p_j^{n+1} - p_j^n}{\Delta t} + \sum_{r \in \mathcal{V}_j^i} \left\langle \kappa_r \beta_r^{-1} \sum_{k \in \mathcal{C}_r} p_k^{n+1} \mathbf{C}_{kr}, \mathbf{C}_{jr} \right\rangle = |\Omega_j| f_j^{n+1}$$

Now, we want to invert the sums over r and k so that we obtain the matrix-vector product.

$$\begin{aligned} \sum_{r \in \mathcal{V}_j^i} \left\langle \kappa_r \beta_r^{-1} \sum_{k \in \mathcal{C}_r} p_k^{n+1} \mathbf{C}_{kr}, \mathbf{C}_{jr} \right\rangle &= \sum_{k \in \mathcal{C}_j} \sum_{r \in \mathcal{V}_{jk}^i} \left\langle \kappa_r \beta_r^{-1} \mathbf{C}_{kr} p_k^{n+1}, \mathbf{C}_{jr} \right\rangle \\ &= \sum_{k \in \mathcal{C}} \sum_{r \in \mathcal{V}_{jk}^i} \left\langle \kappa_r \beta_r^{-1} \mathbf{C}_{kr} p_k^{n+1}, \mathbf{C}_{jr} \right\rangle \end{aligned}$$

Thus the matrix of contribution of the vertices inside the domain writes, for all $j, k \in \mathcal{C}$

$$(M^i)_{jk} = -\sum_{r \in \mathcal{V}_{jk}^i} \left\langle \kappa_r \beta_r^{-1} \mathbf{C}_{kr}, \mathbf{C}_{jr} \right\rangle \tag{172}$$

And the second member gives

$$(B^i)_j^{n+1} = |\Omega_j| f_j^{n+1} \tag{173}$$

C.1.2 On the boundary

Let us now focus on the boundary of the domain. First of all for Neumann boundary conditions we can build $M^{N,b}$ and $B^{N,b}$.

The third equation of (45) gives, for all $r \in \mathcal{V}^{N,b}$

$$\mathbf{u}_r^{n+1} = \beta_r^{-1} \sum_{j \in \mathcal{C}_r} (p_r^{n+1} - p_j^{n+1}) \mathbf{C}_{jr}$$

Replacing $\mathbf{u}_r^{n+1}$ in the first equation of (45) gives us on the boundary with Neumann boundary conditions:

$$\begin{aligned} &- \sum_{r \in \mathcal{V}_j^{N,b}} \left\langle \kappa_r \beta_r^{-1} \sum_{k \in \mathcal{C}_r} (p_r^{n+1} - p_k^{n+1}) \mathbf{C}_{kr}, (\widehat{I}_d - \mathbf{n}_r \otimes \mathbf{n}_r) \mathbf{C}_{jr} \right\rangle \\ &= - \sum_{r \in \mathcal{V}_j^{N,b}} \left\langle \kappa_r \beta_r^{-1} p_r^{n+1} \sum_{k \in \mathcal{C}_r} \mathbf{C}_{kr}, (\widehat{I}_d - \mathbf{n}_r \otimes \mathbf{n}_r) \mathbf{C}_{jr} \right\rangle \\ &+ \sum_{r \in \mathcal{V}_j^{N,b}} \left\langle \kappa_r \beta_r^{-1} \sum_{k \in \mathcal{C}_r} p_k^{n+1} \mathbf{C}_{kr}, (\widehat{I}_d - \mathbf{n}_r \otimes \mathbf{n}_r) \mathbf{C}_{jr} \right\rangle \end{aligned} \tag{174}$$

We can replace p_r^{n+1} given in (37) in the first term of the right hand of (174)

$$\begin{aligned}
&-\sum_{r\in\mathcal{V}_j^{N,b}}\Big\langle\kappa_r\beta_r^{-1}\theta_r^{-1}\Big(\sum_{k\in\mathcal{C}_r}\theta_{kr}p_k^{n+1}+\frac{g_r}{\|\kappa_r\mathbf{n}_r\|}\Big)\sum_{l\in\mathcal{C}_r}\mathbf{C}_{lr},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\Big\rangle\\
&=-\sum_{r\in\mathcal{V}_j^{N,b}}\Big\langle\kappa_r\beta_r^{-1}\theta_r^{-1}\sum_{l\in\mathcal{C}_r}\mathbf{C}_{lr}\sum_{k\in\mathcal{C}_r}\theta_{kr}p_k^{n+1},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\Big\rangle\\
&\quad-\sum_{r\in\mathcal{V}_j^{N,b}}\Big\langle\kappa_r\beta_r^{-1}\theta_r^{-1}\sum_{l\in\mathcal{C}_r}\mathbf{C}_{lr}\frac{g_r}{\|\kappa_r\mathbf{n}_r\|},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\Big\rangle
\end{aligned}\tag{175}$$

with $\theta_{kr}=\langle\beta_r^{-1}\mathbf{C}_{kr},\mathbf{v}_r\rangle$.

Let us replace (175) into (174) and invert the sums on r and k

$$-\sum_{r\in\mathcal{V}_j^{N,b}}\Big\langle\kappa_r\beta_r^{-1}\sum_{k\in\mathcal{C}_r}(p_r^{n+1}-p_k^{n+1})\mathbf{C}_{kr},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\Big\rangle\tag{176}$$

$$\begin{aligned}
&=\sum_{k\in\mathcal{C}_j}\sum_{r\in\mathcal{V}_{jk}^{N,b}}\Big\langle\kappa_r\beta_r^{-1}\Big(\mathbf{C}_{kr}-\theta_r^{-1}\theta_{kr}\sum_{l\in\mathcal{C}_r}\mathbf{C}_{lr}\Big)p_k^{n+1},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\Big\rangle\\
&\quad-\sum_{r\in\mathcal{V}_j^{N,b}}\Big\langle\kappa_r\beta_r^{-1}\theta_r^{-1}\sum_{l\in\mathcal{C}_r}\mathbf{C}_{lr}\frac{g_r}{\|\kappa_r\mathbf{n}_r\|},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\Big\rangle
\end{aligned}\tag{177}$$

One can notice that that the second line of (177) doesn't depend on p_k thus it will go into the right-hand side $(B^{N,b})^{n+1}$.

This way we have built the matrix $M^{N,b}$ with, for all $j,k\in\mathcal{C}$

$$(M^{N,b})_{jk}=-\sum_{r\in\mathcal{V}_{jk}^{N,b}}\Big\langle\kappa_r\beta_r^{-1}\Big(\mathbf{C}_{kr}-\theta_r^{-1}\theta_{kr}\sum_{l\in\mathcal{C}_r}\mathbf{C}_{lr}\Big),(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\Big\rangle\tag{178}$$

And the contribution to the right hand side vector is for all $j\in\mathcal{C}$

$$(B^{N,b})_j^{n+1}=\sum_{r\in\mathcal{V}_j^{N,b}}g_r\Big(\frac{1}{\|\kappa_r\mathbf{n}_r\|}\theta_r^{-1}\Big\langle\kappa_r\beta_r^{-1}\sum_{l\in\mathcal{C}_r}\mathbf{C}_{lr},(\widehat{I}_d-\mathbf{n}_r\otimes\mathbf{n}_r)\mathbf{C}_{jr}\Big\rangle+\langle\mathbf{C}_{jr},\mathbf{n}_r\rangle\Big)\tag{179}$$

After that, let us build $M^{D,b}$ and $B^{D,b}$ for Dirichlet boundary conditions.

Let us consider the fourth equation of (45)

$$\mathbf{u}_r^{n+1}=\beta_r^{-1}\sum_{j\in\mathcal{C}_r}(h_r-p_j^{n+1})\mathbf{C}_{jr}\quad\forall\ r\in\mathcal{V}_j^{D,b}$$

Let us replace this formula into the first equation of (45) and invert the sums on r and k

$$\begin{aligned}
-\sum_{r\in\mathcal{V}_j^{D,b}}\Big\langle\kappa_r\beta_r^{-1}\sum_{k\in\mathcal{C}_r}(h_r-p_k^{n+1})\mathbf{C}_{kr},\mathbf{C}_{jr}\Big\rangle&=-\sum_{r\in\mathcal{V}_j^{D,b}}\Big\langle\kappa_r\beta_r^{-1}\sum_{k\in\mathcal{C}_r}h_r\mathbf{C}_{kr},\mathbf{C}_{jr}\Big\rangle+\\
&\qquad\sum_{r\in\mathcal{V}_j^{D,b}}\Big\langle\kappa_r\beta_r^{-1}\sum_{k\in\mathcal{C}_r}p_k^{n+1}\mathbf{C}_{kr},\mathbf{C}_{jr}\Big\rangle\\
=\sum_{k\in\mathcal{C}}\sum_{r\in\mathcal{V}_{jk}^{D,b}}\langle\kappa_r\beta_r^{-1}\mathbf{C}_{kr}p_k^{n+1},\mathbf{C}_{jr}\rangle&-\sum_{r\in\mathcal{V}_j^{D,b}}\Big\langle\kappa_r\beta_r^{-1}\sum_{k\in\mathcal{C}_r}h_r\mathbf{C}_{kr},\mathbf{C}_{jr}\Big\rangle
\end{aligned}$$

The second term of the right hand side of this equation does not depend on p_k^{n+1}, therefore it will go into $(B^{D,b})^{n+1}$.

Thus, we obtain the matrix $M^{D,b}$ with for all $j, k \in \mathcal{C}$

$$(M^{D,b})_{jk} = - \sum_{r \in \mathcal{V}_{jk}^{D,b}} \langle \kappa_r \beta_r^{-1} \mathbf{C}_{kr}, \mathbf{C}_{jr} \rangle \tag{180}$$

And the second member writes for all $j \in \mathcal{C}$

$$(B^{D,b})_j^{n+1} = \sum_{r \in \mathcal{V}_j^{D,b}} \left\langle \kappa_r \beta_r^{-1} \sum_{k \in \mathcal{C}_r} h_r \mathbf{C}_{kr}, \mathbf{C}_{jr} \right\rangle \tag{181}$$

Thus we have built he matrix $M^b = M^{N,b} + M^{D,b}$ and the second member $B^b = B^{N,b} + B^{D,b}$.

C.1.3 On the corners

Finally, we will focus on the corner values $r \in \mathcal{V}^c$.

For the Neumann boundary condition vertices $r \in \mathcal{V}^{N,c}$, their contribution only appears in the second member of the problem, thus one has

$$M^{N,c} = 0 \tag{182}$$

and for all $j \in \mathcal{C}$

$$(B^{N,c})_j^{n+1} = \sum_{r \in \mathcal{V}_j^{N,c}} \frac{1}{2} (\ell_{l_1} g_{l_1} + \ell_{l_2} g_{l_2}) \tag{183}$$

Let us finish with the Dirichlet boundary condition on the corners $r \in \mathcal{V}^{D,c}$.

Let us consider the last equation of (45)

$$\mathbf{u}_r^{n+1} = (\beta_r^c)^{-1} \sum_{j \in \mathcal{C}_r} (h_r - p_j^{n+1}) \mathbf{C}_{jr} \qquad \text{if } r \in \mathcal{V}_j^{D,c}$$

We will repeat the same steps as for the Neumann boundary condition on the boundaries and we obtain for all $j, k \in \mathcal{C}$

$$(M^{D,c})_{jk} = - \sum_{r \in \mathcal{V}_{jk}^{D,c}} \langle \kappa_r (\beta_r^c)^{-1} \mathbf{C}_{kr}, \mathbf{C}_{jr} \rangle \tag{184}$$

And the second member writes for all $j \in \mathcal{C}$

$$(B^{D,c})_j^{n+1} = \sum_{r \in \mathcal{V}_j^{D,c}} \left\langle \kappa_r (\beta_r^c)^{-1} \sum_{k \in \mathcal{C}_r} h_r \mathbf{C}_{kr}, \mathbf{C}_{jr} \right\rangle \tag{185}$$

Thus we have built he matrix $M^c = M^{N,c} + M^{D,c}$ and the second member $B^c = B^{N,c} + B^{D,c}$.

Thus every component of the matrix M and the second member B is entirely built.

Remark 28. In the scheme one can see that the boundary conditions are taken at the nodes of the boundary which is also true in the code for Dirichlet boundary conditions. However in the case of Neumann boundary conditions, the values are given at the middle of the boundary faces.

In order to retrieve the node values on straight borders, let us use the following formula where the boundary node r is linked to the cells j_1 and j_2 with their face on a boundary called respectively l_1 and l_2:

$$g_r = \frac{g_{l_1} \ell_{l_1} + g_{l_2} \ell_{l_2}}{\ell_{l_1} + \ell_{l_2}}$$

where g stands for the Neumann boundary condition, $g_{l_{1,2}}$ and $\ell_{l_{1,2}}$ are respectively the value of the boundary condition and the length of the faces l_1 and l_2.

In order to preserve the exactness of the scheme for linear solutions (proved in Section 4.2), we need to change the definition of the Neumann boundary values at the angles of the mesh. As a reminder, the definition of an "angle" of the mesh is given in Definition 10.

Let us begin with finding the two coefficients α, β such that $\mathbf{n}_r = \alpha \ell_{l_1} \mathbf{n}_{l_1} + \beta \ell_{l_2} \mathbf{n}_{l_2}$. This is easily done by finding the solution of the following system

$$A \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \mathbf{n}_r^{(1)} \\ \mathbf{n}_r^{(2)} \end{pmatrix} \quad \text{where} \quad A = \begin{pmatrix} \mathbf{n}_{l_1}^{(1)} \ell_{l_1} & \mathbf{n}_{l_2}^{(1)} \ell_{l_2} \\ \mathbf{n}_{l_1}^{(2)} \ell_{l_1} & \mathbf{n}_{l_2}^{(2)} \ell_{l_2} \end{pmatrix}$$

Then we know that

$$g_r = \langle \kappa_r \nabla p_r, \mathbf{n}_r \rangle = \langle \kappa_r \nabla p_r, \mathbf{n}_{l_1} \rangle \alpha \ell_{l_1} + \langle \kappa_r \nabla p_r, \mathbf{n}_{l_2} \rangle \alpha \ell_{l_2} \approx \alpha \ell_{l_1} g_{l_1} + \beta \ell_{l_2} g_{l_2}$$

Note that this approximation is exact when κ and ∇p are constant, which are the hypothesis made to prove the exactness of the scheme in Section 4.2.

If the node is linked to two different boundary conditions then the Dirichlet one prevails.

C.2 Hybrid scheme matrix

Using the same method as in Section C.1, we want to write the scheme under the form

$$(I_\Omega - \Delta t M) P^{n+1} = \Delta t B^{n+1} + P^n \tag{186}$$

with $M = \lambda M_H + (1-\lambda) M_N$ and $B = \lambda B_H + (1-\lambda) B_N$, where M_N and B_N are respectively the matrix and the second member of the nodal scheme built in Section C.1.

Let us now construct the matrix M_H and the second member B_H.

Let us write $M_H = M_H^i + M_H^b$ with M_H^i the contribution of the faces inside the domain and M_H^b the contribution of the boundary faces. We can decompose the vector B_H with the same method.

C.2.1 Inside the domain faces

Let us inject the second equation of (141) into the sum on the faces of the first equation which becomes

$$\sum_{l \in \mathcal{F}_j^i} \ell_l \langle \mathbf{u}_l^{n+1}, \kappa_l \mathbf{n}_l \rangle = \sum_{l \in \mathcal{F}_j^i} \left(\frac{\ell_l}{2} \alpha_l \sum_{r \in \mathcal{V}_l} \langle \mathbf{u}_r^{n+1}, (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp \rangle + \ell_l \delta_l (p_{j_1}^{n+1} - p_{j_2}^{n+1}) \right) \tag{187}$$

One can now replace $\mathbf{u_r}^{n+1}$ into (187) given in the fourth to the sixth equations of (141):

$$= - \sum_{l \in \mathcal{F}_j^i} \frac{\ell_l}{2} \alpha_l \sum_{r \in \mathcal{V}_l^i} \langle \beta_r^{-1} \sum_{k \in \mathcal{C}_r} p_k^{n+1} \mathbf{C}_{kr}, (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp \rangle \tag{188}$$

$$+ \sum_{l \in \mathcal{F}_j^i} \frac{\ell_l}{2} \alpha_l \sum_{r \in \mathcal{V}_l^{N,b}} \langle \beta_r^{-1} \sum_{k \in \mathcal{C}_r} (p_r^{n+1} - p_k^{n+1}) \mathbf{C}_{kr}, (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp \rangle \tag{189}$$

$$+ \sum_{l \in \mathcal{F}_j^i} \frac{\ell_l}{2} \alpha_l \sum_{r \in \mathcal{V}_l^{D,b}} \langle \beta_r^{-1} \sum_{k \in \mathcal{C}_r} (h_r^{n+1} - p_k^{n+1}) \mathbf{C}_{kr}, (\mathbf{x}_{j_1} - \mathbf{x}_{j_2})^\perp \rangle \tag{190}$$

$$+ \sum_{l \in \mathcal{F}_j^i} \ell_l \delta_l (p_{j_1}^{n+1} - p_{j_2}^{n+1}) \tag{191}$$

Let us now consider the first sum over l (188). We want to bring the sum over k outside of the other sums so that we obtain a matrix-vector product. It can be decomposed into three steps. The first one consists in swapping the sums over l and r. Secondly one can invert the sums over l and k since they do not depend on each other. Finally one can swap the sum over r and k, so that the sum over k is the most external one:

$$\begin{aligned}
-\sum_{l\in\mathcal{F}_j^i}\frac{\ell_l}{2}\alpha_l\sum_{r\in\mathcal{V}_l^i}\langle\beta_r^{-1}\sum_{k\in\mathcal{C}_r}p_k^{n+1}\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle &= -\sum_{r\in\mathcal{V}_j^i}\sum_{l\in\mathcal{F}_{jr}^i}\frac{\ell_l}{2}\alpha_l\langle\beta_r^{-1}\sum_{k\in\mathcal{C}_r}p_k^{n+1}\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle \\
&= -\sum_{r\in\mathcal{V}_j^i}\sum_{k\in\mathcal{C}_r}\sum_{l\in\mathcal{F}_{jr}^i}\frac{\ell_l}{2}\alpha_l\langle\beta_r^{-1}p_k^{n+1}\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle \\
&= -\sum_{k\in\mathcal{C}_j}\sum_{r\in\mathcal{V}_{jk}^i}\sum_{l\in\mathcal{F}_{jr}^i}\frac{\ell_l}{2}\alpha_l\langle\beta_r^{-1}p_k^{n+1}\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle \\
&= -\sum_{k\in\mathcal{C}}\sum_{r\in\mathcal{V}_{jk}^i}\sum_{l\in\mathcal{F}_{jr}^i}\frac{\ell_l}{2}\alpha_l\langle\beta_r^{-1}p_k^{n+1}\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle
\end{aligned} \tag{192}$$

This contribution will be incorporated into the matrix M_H^i.

Thus the matrix of contribution of the faces inside the domain writes from (192) and (191), for all $j,k\in\mathcal{C}$

$$(M_H^i)_{jk} = -\sum_{r\in\mathcal{V}_{jk}^i}\sum_{l\in\mathcal{F}_{jr}^i}\frac{\ell_l}{2}\alpha_l\langle\beta_r^{-1}\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle+\sum_{l\in\mathcal{F}_j^i}\ell_l\delta_l(\mathbb{1}_{k=j_1}-\mathbb{1}_{k=j_2}) \tag{193}$$

And the second member gives

$$(B_H^i)_j^{n+1} = |\Omega_j|f_j^{n+1} \tag{194}$$

Let us now consider the second sum over l (189).

$$\begin{aligned}
-\sum_{l\in\mathcal{F}_j^i}\frac{\ell_l}{2}\alpha_l\sum_{r\in\mathcal{V}_l^{N,b}}\langle\beta_r^{-1}\sum_{k\in\mathcal{C}_r}(p_r^{n+1}-p_j^{n+1})\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle &= -\sum_{l\in\mathcal{F}_j^i}\frac{\ell_l}{2}\alpha_l\sum_{r\in\mathcal{V}_l^{N,b}}\langle\beta_r^{-1}\sum_{k\in\mathcal{C}_r}p_r^{n+1}\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle \\
&\quad+\sum_{l\in\mathcal{F}_j^i}\frac{\ell_l}{2}\alpha_l\sum_{r\in\mathcal{V}_l^{N,b}}\langle\beta_r^{-1}\sum_{k\in\mathcal{C}_r}p_k^{n+1}\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle
\end{aligned} \tag{195}$$

One can replace p_r^{n+1} by its formula into the first sum of (195):

$$\begin{aligned}
-\sum_{l\in\mathcal{F}_j^i}\frac{\ell_l}{2}\alpha_l\sum_{r\in\mathcal{V}_l^{N,b}}\langle\beta_r^{-1}\sum_{k\in\mathcal{C}_r}p_r^{n+1}\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle &= -\sum_{l\in\mathcal{F}_j^i}\frac{\ell_l}{2}\alpha_l\sum_{r\in\mathcal{V}_l^{N,b}}\langle\beta_r^{-1}\theta_r^{-1}\Big(\sum_{k\in\mathcal{C}_r}\theta_{kr}p_k^{n+1}+\frac{g_r^{n+1}}{\|\kappa_r\mathbf{n}_r\|}\Big) \\
&\qquad\sum_{k'\in\mathcal{C}_r}\mathbf{C}_{k'r},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle \\
&= -\sum_{l\in\mathcal{F}_j^i}\frac{\ell_l}{2}\alpha_l\sum_{r\in\mathcal{V}_l^{N,b}}\langle\beta_r^{-1}\theta_r^{-1}\sum_{k'\in\mathcal{C}_r}\mathbf{C}_{k'r}\sum_{k\in\mathcal{C}_r}\theta_{kr}p_k^{n+1},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle \\
&\quad-\sum_{l\in\mathcal{F}_j^i}\frac{\ell_l}{2}\alpha_l\sum_{r\in\mathcal{V}_l^{N,b}}\langle\beta_r^{-1}\theta_r^{-1}\frac{g_r^{n+1}}{\|\kappa_r\mathbf{n}_r\|}\sum_{k'\in\mathcal{C}_r}\mathbf{C}_{k'r},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle
\end{aligned} \tag{196}$$

One can now replace (196) into (195) and invert the sums over l, r and k just like above in order to obtain the matrix-vector product:

$$
\begin{aligned}
&-\sum_{l\in\mathcal{F}_j^i}\frac{\ell_l}{2}\alpha_l\sum_{r\in\mathcal{V}_l^{N,b}}\langle\beta_r^{-1}\sum_{k\in\mathcal{C}_r}(p_r^{n+1}-p_j^{n+1})\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle\\
&=-\sum_{k\in\mathcal{C}}\sum_{r\in\mathcal{V}_{jk}^{N,b}}\sum_{l\in\mathcal{F}_{jr}^i}\frac{\ell_l}{2}\alpha_l\langle\beta_r^{-1}\Big(\mathbf{C}_{kr}-\theta_r^{-1}\sum_{k'\in\mathcal{C}_r}\mathbf{C}_{k'r}\Big)p_k^{n+1},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle\\
&\qquad-\sum_{r\in\mathcal{V}_j^{N,b}}\sum_{l\in\mathcal{F}_{jr}^i}\frac{\ell_l}{2}\alpha_l\langle\beta_r^{-1}\theta_r^{-1}\frac{g_r^{n+1}}{\|\kappa_r\mathbf{n}_r\|}\sum_{k'\in\mathcal{C}_r}\mathbf{C}_{k'r},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle
\end{aligned}
\tag{197}
$$

The sum over k will be a contribution to the matrix $M_H^{N,b}$ and the second sum will be added to $(B_h^{N,b})^{n+1}$.

For now one has

$$
(M_H^{N,b})_{jk}=-\sum_{r\in\mathcal{V}_{jk}^{N,b}}\sum_{l\in\mathcal{F}_{jr}^i}\frac{\ell_l}{2}\alpha_l\langle\beta_r^{-1}\Big(\mathbf{C}_{kr}-\theta_r^{-1}\sum_{k'\in\mathcal{C}_r}\mathbf{C}_{k'r}\Big),(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle
\tag{198}
$$

and

$$
(B_H^{N,b})_j^{n+1}=\sum_{r\in\mathcal{V}_j^{N,b}}\sum_{l\in\mathcal{F}_{jr}^i}\frac{\ell_l}{2}\alpha_l\langle\beta_r^{-1}\theta_r^{-1}\frac{g_r^{n+1}}{\|\kappa_r\mathbf{n}_r\|}\sum_{k'\in\mathcal{C}_r}\mathbf{C}_{k'r},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle
\tag{199}
$$

Finally, one can consider the third sum over l (190) and invert the sums as before:

$$
\begin{aligned}
-\sum_{l\in\mathcal{F}_j^i}\frac{\ell_l}{2}\alpha_l\sum_{r\in\mathcal{V}_l^{D,b}}\langle\beta_r^{-1}\sum_{k\in\mathcal{C}_r}(h_r^{n+1}-p_k^{n+1})\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle&=\sum_{l\in\mathcal{F}_j^i}\frac{\ell_l}{2}\alpha_l\sum_{r\in\mathcal{V}_l^{D,b}}\langle\beta_r^{-1}\sum_{k\in\mathcal{C}_r}p_k^{n+1}\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle\\
&-\sum_{l\in\mathcal{F}_j^i}\frac{\ell_l}{2}\alpha_l\sum_{r\in\mathcal{V}_l^{D,b}}\langle\beta_r^{-1}\sum_{k\in\mathcal{C}_r}h_r^{n+1}\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle\\
&=\sum_{k\in\mathcal{C}}\sum_{r\in\mathcal{V}_{jk}^{D,b}}\sum_{l\in\mathcal{F}_{jr}^i}\frac{\ell_l}{2}\alpha_l\langle\beta_r^{-1}p_k^{n+1}\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle\\
&-\sum_{r\in\mathcal{V}_j^{D,b}}\sum_{l\in\mathcal{F}_{jr}^i}\frac{\ell_l}{2}\alpha_l\langle\beta_r^{-1}\sum_{k\in\mathcal{C}_r}h_r^{n+1}\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle
\end{aligned}
$$

The sum over k will be a contribution to the matrix $M_H^{D,b}$ and the second sum will be added to $(B_h^{D,b})^{n+1}$.

Therefore for now one has

$$
(M_H^{D,b})_{jk}=-\sum_{r\in\mathcal{V}_{jk}^{D,b}}\sum_{l\in\mathcal{F}_{jr}^i}\frac{\ell_l}{2}\alpha_l\langle\beta_r^{-1}\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle
\tag{200}
$$

and

$$
(B_H^{D,b})_j^{n+1}=\sum_{r\in\mathcal{V}_j^{D,b}}\sum_{l\in\mathcal{F}_{jr}^i}\frac{\ell_l}{2}\alpha_l\langle\beta_r^{-1}\sum_{k\in\mathcal{C}_r}h_r^{n+1}\mathbf{C}_{kr},(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle
\tag{201}
$$

C.2.2 Boundary faces

In the case of the Neumann boundary conditions, since the boundary faces only appear in the second member of the hybrid scheme, then it is obvious that nothing is added to the matrix (198), therefore for all j,k in $\mathcal{C}$ one has:

$$
(M_H^{N,b})_{jk}=-\sum_{r\in\mathcal{V}_{jk}^{N,b}}\sum_{l\in\mathcal{F}_{jr}^i}\frac{\ell_l}{2}\alpha_l\langle\beta_r^{-1}\Big(\mathbf{C}_{kr}-\theta_r^{-1}\sum_{k'\in\mathcal{C}_r}\mathbf{C}_{k'r}\Big),(\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^\perp\rangle
\tag{202}
$$

And we add the sum over the boundary faces to the second member (199), therefore for all j in $\mathcal{C}$:

$$(B_H^{N,b})_j = \sum_{r\in\mathcal{V}_j^{N,b}} \sum_{l\in\mathcal{F}_{jr}^i} \frac{\ell_l}{2}\alpha_l \langle \beta_r^{-1}\theta_r^{-1}\frac{g_r^{n+1}}{\|\kappa_r \mathbf{n}_r\|}\sum_{k'\in\mathcal{C}_r}\mathbf{C}_{k'r}, (\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^{\perp}\rangle + \sum_{l\in\mathcal{F}_j^{N,b}} \ell_l g_l^{n+1} \tag{203}$$

On the other hand, with Dirichlet boundary conditions, one has

$$\mathbf{u}_r^{n+1} = \beta_r^{-1}\sum_{j\in\mathcal{C}_r}(h_r^{n+1}-p_j^{n+1})\mathbf{C}_{jr} \quad \forall\, r\in\mathcal{V}^{D,b}$$

Let us inject this definition of $\mathbf{u}_r$ into the third equation of (141) so that we obtain for the Dirichlet boundary faces

$$\begin{aligned}\langle \mathbf{u}_l^{n+1}, \kappa_l \mathbf{n}_l\rangle &= \alpha_l \sum_{r\in\mathcal{V}_l^{D,b}} \langle \frac{\mathbf{u}_r^{n+1}}{2}, (\mathbf{x}_l-\mathbf{x}_{j_2})^{\perp}\rangle + \delta_l(h_l^{n+1}-p_{j_2}^{n+1}) \\ &= \alpha_l \sum_{r\in\mathcal{V}_l^{D,b}} \langle \frac{1}{2}\beta_r^{-1}\sum_{k\in\mathcal{C}_r}(h_r^{n+1}-p_k^{n+1})\mathbf{C}_{kr}, (\mathbf{x}_l-\mathbf{x}_{j_2})^{\perp}\rangle + \delta_l(h_l^{n+1}-p_{j_2}^{n+1})\end{aligned}$$

One can now split this formula into two: one part that depends on the unknown p and the other one only depending on the Dirichlet boundary condition h. Then it can be replaced inside the the sum on the boundary faces of the first equation of (141):

$$\begin{aligned}-\sum_{l\in\mathcal{F}_j^{D,b}} \ell_l\langle \mathbf{u}_l^{n+1}, \kappa_l\mathbf{n}_l\rangle = &\sum_{l\in\mathcal{F}_j^{D,b}} \ell_l\Big(\alpha_l \sum_{r\in\mathcal{V}_l^{D,b}} \langle \frac{1}{2}\beta_r^{-1}\sum_{k\in\mathcal{C}_r} p_k^{n+1}\mathbf{C}_{kr}, (\mathbf{x}_l-\mathbf{x}_{j_2})^{\perp}\rangle - \delta_l p_{j_2}^{n+1}\Big) \\ &- \sum_{l\in\mathcal{F}_j^{D,b}} \ell_l\Big(\alpha_l \sum_{r\in\mathcal{V}_l^{D,b}} \langle \frac{1}{2}\beta_r^{-1}\sum_{k\in\mathcal{C}_r} h_r^{n+1}\mathbf{C}_{kr}, (\mathbf{x}_l-\mathbf{x}_{j_2})^{\perp}\rangle + \delta_l h_l^{n+1}\Big)\end{aligned}$$

Using the same sum inversion as for the inside of the domain on the part including the unknown p and adding it to (200), one gets for all $j,k\in\mathcal{C}$:

$$(M_H^{D,b})_{jk} = -\sum_{r\in\mathcal{V}_{jk}^{D,b}} \sum_{l\in\mathcal{F}_{jr}^i\cup\mathcal{F}_{jr}^{D,b}} \frac{\ell_l}{2}\alpha_l\langle \beta_r^{-1}\mathbf{C}_{kr}, (\mathbf{x}_{j_1}-\mathbf{x}_{j_2})^{\perp}\rangle - \sum_{l\in\mathcal{F}_j^{D,b}} \ell_l\delta_l \mathbb{1}_{k=j_2} \tag{204}$$

And the part that only depends on the boundary conditions gives, by adding the previous equation to (201), for all $j\in\mathcal{C}$:

$$(B_H^D)_j^{n+1} = \sum_{r\in\mathcal{V}_{jk}^{D,b}} \sum_{l\in\mathcal{F}_{jr}^i\cup\mathcal{F}_{jr}^{D,b}} \frac{\ell_l}{2}\alpha_l h_r^{n+1}\langle \beta_r^{-1}\sum_{k\in\mathcal{C}_r}\mathbf{C}_{kr}, (\mathbf{x}_l-\mathbf{x}_{j_2})^{\perp}\rangle + \sum_{l\in\mathcal{F}_j^{D,b}} \ell_l\delta_l h_l^{n+1} \tag{205}$$

References

[1] I. Aavatsmark, G.T. Eigestad, R.A. Klausen, M.F. Wheeler, and I. Yotov. Convergence of a symmetric MPFA method on quadrilateral grids. *Comput. Geosci.*, 11(4):333–345, 2007.

[2] Xavier Blanc, Vincent Delmas, and Philippe Hoch. Asymptotic preserving schemes on conical unstructured 2d meshes. *International Journal for Numerical Methods in Fluids*, 93(8):2763–2802, 2021.

[3] Xavier Blanc, Philippe Hoch, and Clément Lasuen. Positive composite finite volume schemes for the anisotropic diffusion equation on unstructured meshes. In preparation, 2022.

[4] Haim Brezis. *Functional analysis, Sobolev spaces and partial differential equations.* Universitext. New York, NY: Springer, 2011.

[5] Christophe Buet, Bruno Després, and Emmanuel Franck. Design of asymptotic preserving finite volume schemes for the hyperbolic heat equation on unstructured meshes. *Numer. Math.*, 122(2):227–278, 2012.

[6] Christophe Buet, Bruno Després, and Emmanuel Franck. An asymptotic preserving scheme with the maximum principle for the m1 model on distorded meshes. *Comptes Rendus Mathematique*, 350(11):633–638, 2012.

[7] Yves Coudi'ere, Jean-Paul Vila, and Philippe Villedieu. Convergence rate of a finite volume scheme for a two dimensional convection-diffusion problem. *Math. Model. Numer. Anal.*, 33(3):493–516, 1999.

[8] Bruno Després. Lax theorem and finite volume schemes. *Math. Comput.*, 73(247):1203–1234, 2004.

[9] Bruno Després. Weak consistency of the cell-centered Lagrangian GLACE scheme on general meshes in any dimension. *Comput. Methods Appl. Mech. Eng.*, 199(41-44):2669–2679, 2010.

[10] Bruno Després. Différences Finies et Volumes Finis - Cours de Master Mathématiques et Application. LJLL, Sorbonne Université, 2021. Private communication.

[11] Bruno Dubroca and Jean-Luc Feugeas. Theoretical and numerical study on a moment closure hierarchy for the radiative transfer equation. *C. R. Acad. Sci., Paris, Sér. I, Math.*, 329(10):915–920, 1999.

[12] Robert Eymard, Thierry Gallouët, Cindy Guichard, Rapha'ele Herbin, and Roland Masson. TP or not TP, that is the question. *Computational Geosciences*, 18(3-4):285–296, 2014.

[13] Emmanuel Franck. *Construction et analyse numérique de schéma asymptotic preserving sur maillages non structurés. Application au transport linéaire et aux systèmes de Friedrichs.* Thèse, Université Pierre et Marie Curie - Paris VI, October 2012.

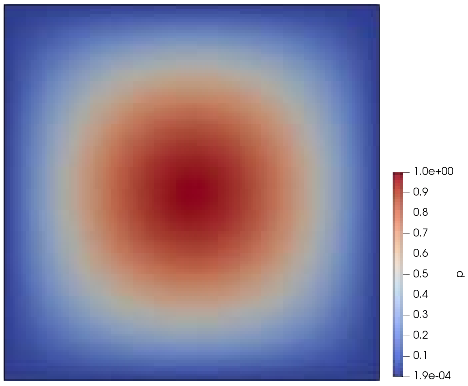
[14] F. Hermeline. A finite volume method for the approximation of diffusion operators on distorted meshes. *J. Comput. Phys.*, 160(2):481–499, 2000.

[15] C. David Levermore. Entropy-based moment closures for kinetic equations. *Transp. Theory Stat. Phys.*, 26(4-5):591–606, 1997.

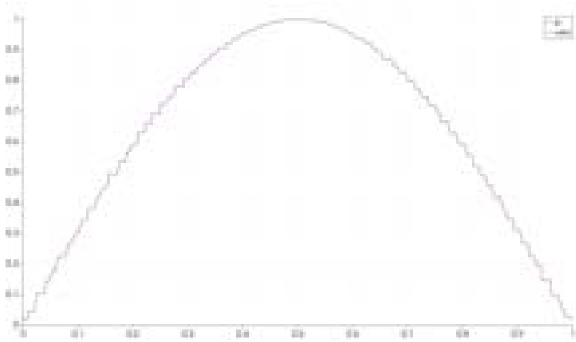
[16] Konstantin Lipnikov, Mikhail Shashkov, and Daniil Svyatskiy. The mimetic finite difference discretization of diffusion problem on unstructured polyhedral meshes. *J. Comput. Phys.*, 211(2):473–491, 2006.

[17] Panagiotis Paraschis and Georgios E Zouraris. On the convergence of the crank-nicolson method for the logarithmic schrödinger equation. working paper or preprint, February 2022.

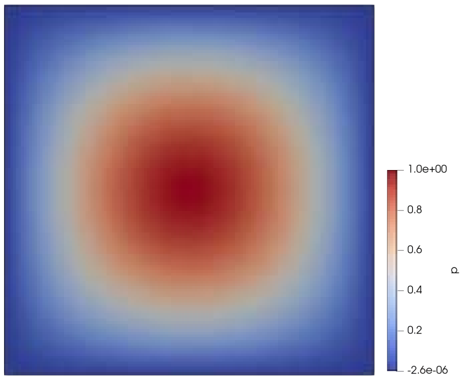
[18] Andrei D. Polyanin and Vladimir E. Nazaikinskii. *Handbook of linear partial differential equations for engineers and scientists.* Boca Raton, FL: CRC Press, 2nd edition edition, 2016.



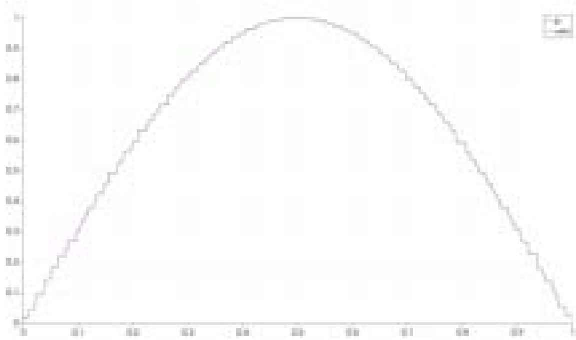
(a) Allure of the solution for $\lambda = h^{\frac{1}{2}}$



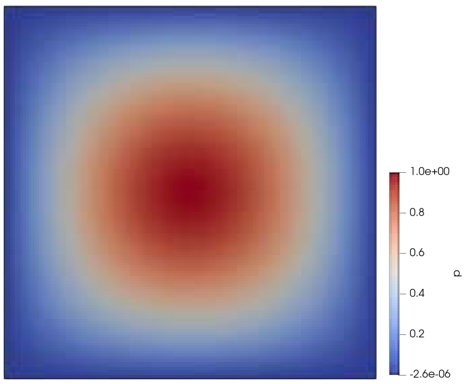
(b) Horizontal section of the solution for $\lambda = h^{\frac{1}{2}}$



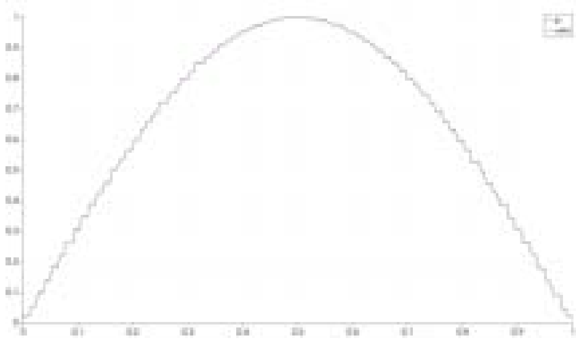
(c) Allure of the solution for $\lambda = h$



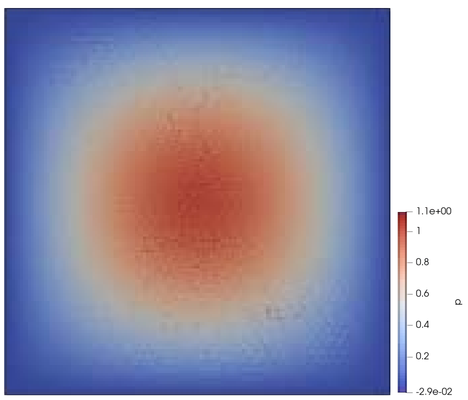
(d) Horizontal section of the solution for $\lambda = h$



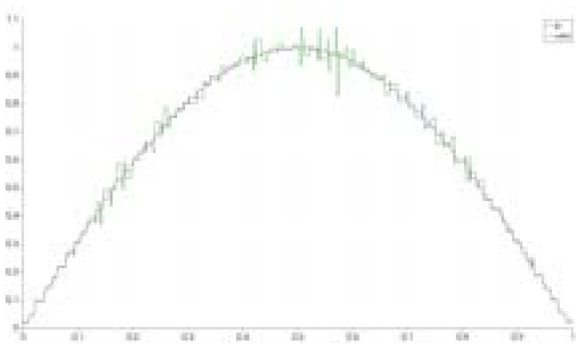
(e) Allure of the solution for $\lambda = h^2$



(f) Horizontal section of the solution for $\lambda = h^2$



(g) Allure of the solution for $\lambda = h^3$



(h) Horizontal section of the solution for $\lambda = h^3$

Figure 42: Allure and horizontal sections of the trigonometric solution on a random triangle mesh depending on the value of λ for $N = 80$