



**HAL**  
open science

**Stabilité de l'équation d'advection-diffusion et stabilité  
de l'équation d'advection pour la solution du problème  
approché, obtenue par la méthode upwind  
d'éléments-finis et de volumes-finis avec des éléments de  
Crouzeix-Raviart**

Marcus Mildner

► **To cite this version:**

Marcus Mildner. Stabilité de l'équation d'advection-diffusion et stabilité de l'équation d'advection pour la solution du problème approché, obtenue par la méthode upwind d'éléments-finis et de volumes-finis avec des éléments de Crouzeix-Raviart. Mathématiques générales [math.GM]. Université du Littoral Côte d'Opale, 2013. Français. NNT : 2013DUNK0316 . tel-00839524

**HAL Id: tel-00839524**

**<https://theses.hal.science/tel-00839524v1>**

Submitted on 28 Jun 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE

Présentée pour obtenir

LE GRADE DE DOCTEUR EN SCIENCES  
DE L'UNIVERSITÉ DU LITTORAL CÔTE  
D'OPALE

Spécialité: Mathématiques

par

Marcus MILDNER

**Stabilité de l'équation d'advection-diffusion et stabilité de  
l'équation d'advection pour la solution du problème approché  
obtenue par la méthode upwind d'éléments-finis et de volumes-finis  
avec des éléments de Crouzeix-Raviart**

Soutenue le jeudi 30 mai 2013 devant la Commission d'examen:

- M. Jean-Paul CHEHAB (Président du jury)
- M. Paul DEURING (Directeur de thèse)
- M. Robert EYMARD
- M. Jean-Luc GUERMOND
- M. Hassane SADOK

Rapporteurs:

- M. Robert Eymard
- M. Jean-Luc Guermond

**Laboratoire de Mathématiques**  
**Pures et Appliquées**  
**Joseph Liouville**



J. Liouville 1809-1882

Thèse préparée au  
**Laboratoire de Mathématiques Pures et Appliquées**  
**Joseph Liouville**  
Centre Universitaire de la Mi-Voix  
Maison de la Recherche Blaise Pascal  
50 rue F.Buisson B.P. 699  
62 228 Calais Cedex

## Résumé

On considère le problème d'advection-diffusion stationnaire  $\nu(\nabla u, \nabla v) + (\beta \cdot \nabla u, v) = (f, v)$  et non stationnaire  $\frac{d}{dt}(u(t), v) + \nu(\nabla u, \nabla v) + (\beta \cdot \nabla u, v) = (g(t), v)$ , ainsi que le problème d'advection  $(\beta \cdot \nabla u, v) = (f, v)$  sur un domaine polygonal borné du plan.

Le terme de diffusion est approché par des éléments de Crouzeix Raviart et le terme de convection par une méthode upwind sur des volumes barycentriques finis avec un maillage triangulaire.

Pour le problème stationnaire d'advection-diffusion, la  $L^2$ -stabilité (c'est-à-dire indépendante du coefficient de diffusion  $\nu$ ) est démontrée pour la solution du problème approché obtenue par cette méthode d'éléments finis et de volumes finis. Pour cela une condition sur la géométrie doit être satisfaite. Des exemples de maillages sont donnés.

Toujours avec cette condition géométrique sur le maillage, une inégalité de stabilité (où la discrétisation en temps n'est pas couplée à une condition sur la finesse du maillage) est obtenue pour le cas non-stationnaire. La discrétisation en temps  $y$  est faite par un schéma d'Euler implicite.

Une majoration de l'erreur, proportionnelle au pas en temps et à la finesse du maillage, est ensuite proposée et exprimée explicitement en fonction des données du problème.

Pour le problème d'advection, une approche utilisant la théorie des graphes est utilisée pour obtenir l'existence et l'unicité de la solution, ainsi que le résultat de stabilité. Comme pour la stabilité du problème d'advection-diffusion, une condition géométrique - qui est équivalente pour les points intérieurs du maillage à celle du problème d'advection-diffusion - est nécessaire.

**Mots-clefs** : équation d'advection-diffusion, méthode d'éléments et de volumes finis, éléments finis de Crouzeix-Raviart, volume fini barycentrique, méthode upwind, estimation de l'erreur, équation d'advection, graphe orienté, existence, unicité, stabilité.

STABILITY FOR THE CONVECTION-DIFFUSION PROBLEM AND STABILITY FOR THE  
CONVECTION PROBLEM DISCRETIZED BY CROUZEIX-RAVIART FINITE ELEMENT  
USING UPWIND FINITE VOLUME-FINITE ELEMENT METHOD

**Abstract**

We consider the stationary linear convection-diffusion equation  $\nu(\nabla u, \nabla v) + (\beta \cdot \nabla u, v) = (f, v)$ , the time dependent  $\frac{d}{dt}(u(t), v) + \nu(\nabla u, \nabla v) + (\beta \cdot \nabla u, v) = (g(t), v)$  equation and the linear advection equation  $(\beta \cdot \nabla u, v) = (f, v)$  on a two dimensional bounded polygonal domain  $\Omega$ . The diffusion term is discretized by Crouzeix-Raviart piecewise linear finite elements, and the convection term by upwind barycentric finite volumes on a triangular grid.

For the stationary convection-diffusion problem,  $L^2$ -stability (i.e. independent of the diffusion coefficient  $\nu$ ) is proven for the approximate solution obtained by this combined finite-element finite-volume method. This result holds if the underlying grid satisfies a condition that is fulfilled, for example, by some structured meshes.

Using again this condition on the grid, stability is shown for the time dependent convection-diffusion equation (without any link between mesh size and time step). An implicit Euler approach is used for the time discretization. It is shown that the error associated with this scheme decays linearly with the mesh size and the time step. This result holds without any link between mesh size and time step. The dependence of the corresponding error bound on the diffusion coefficient is completely explicit.

For the stationary advection equation, an approach using graph theory is used to obtain existence, uniqueness and stability. As in the stationary linear convection-diffusion equation, the underlying grid must satisfy some geometric condition.

**Keywords** : convection-diffusion equation, combined finite element-finite volume method, Crouzeix-Raviart finite elements, barycentric finite volumes, upwind method, error estimates, advection equation, directed graph, existence, uniqueness, stability.

STABILITÄT DES DIFFUSIONS-KONVEKTIONS-PROBLEMS UND STABILITÄT DES  
KONVEKTIONS-PROBLEMS FÜR DIE LÖSUNG MITTELS UPWIND FINITE-ELEMENTE  
FINITE-VOLUME METHODEN MIT CROUZEIX-RAVIART ELEMENTE

**Zusammenfassung**

Gegenstand der Arbeit ist die zweidimensionale stationäre Konvektion-Diffusionsgleichung  $\nu(\nabla u, \nabla v) + (\beta \cdot \nabla u, v) = (f, v)$ , die zeitabhängige Konvektion-Diffusionsgleichung  $\frac{d}{dt}(u(t), v) + \nu(\nabla u, \nabla v) + (\beta \cdot \nabla u, v) = (g(t), v)$ , sowie die Konvektionsgleichung  $(\beta \cdot \nabla u, v) = (f, v)$ . Der Diffusionsterm ist diskretisiert mittels Crouzeix-Raviart stückweise lineare Finite Elemente. Das Gebiet ist in Dreiecke unterteilt und der Konvektionsterm ist mittels einer upwind Methode auf Baryzentrische Finite Volumenelemente definiert.

Für die stationäre Konvektion-Diffusionsgleichung, wird (d.h. von  $\nu$  unabhängige)  $L^2$ -Stabilität der numerischen Lösung bewiesen. Voraussetzung dafür, ist die Erfüllung gewisser geometrischer Bedingungen an die Unterteilung des Gebiets. Beispiele von Unterteilungen die diese Bedingungen erfüllen, werden gegeben. Wieder an dieser geometrischen Bedingung geknüpft, wird Stabilität (d.h. die Zeitdiskretisierung ist entkoppelt von der Netzweite) für die zeitabhängige Konvektion-Diffusionsgleichung, bewiesen. Für die Zeitableitung wird dabei eine Implizite Euler Diskretisierung verwendet. Eine obere Schranke für den Diskretisierungsfehler, proportional zum Zeitdiskretisierungsparameter und zur Netzfeinheit, ausgedrückt als Funktion der Daten der Differenzialgleichung, wird gezeigt.

Für die Konvektionsgleichung wird ein graphentheoretischer Zugang verwendet, der es ermöglicht Existenz, Eindeutigkeit und Stabilität, zu bekommen. Für die Stabilität, werden ähnliche geometrische Bedingungen an die Unterteilung des Gebiets gestellt, wie beim stationären Konvektion-Diffusionsproblem.

**Schlagworte** : Diffusions-Konvektions-Gleichung, finite Elemente-finite Volumen Methoden, Crouzeix-Raviart finite Elemente, Baryzentrische finite Volumenelemente, upwind-Methode, Fehlerabschätzung, Konvektions-Gleichung, gerichteter Graph, Existenz, Eindeutigkeit, Stabilität.



# Remerciements

Cette thèse n'aurait pu se faire sans l'orientation, l'aide et le soutien continu et patient de mon directeur de thèse Paul Deuring. Je lui en suis profondément reconnaissant.

Merci au laboratoire de Laboratoire de Mathématiques Pures et Appliquées Joseph Liouville qui m'a accueilli. Très particulièrement, je souhaite y remercier pour leur aide Marc Prévost, Philippe Marion et Isabelle Buchard.

Merci Philippe pour avoir permis le calcul de certaines simulations numériques importantes.

Je souhaite aussi remercier l'École Doctorale Sciences pour l'Ingénieur Université Lille Nord-de-France pour le cadre et les conditions favorables au déroulement de cette thèse.

Merci à ma famille, mes amis et, surtout, ma femme pour leur longue patience, le fait d'avoir accepté et permis ces longues et innombrables heures consacrées à ce travail mathématique. Merci beaucoup à Christine Laigle pour la patiente relecture du manuscrit.

Toute cette énergie, n'aurait pas possible, sans l'inspiration, le plaisir et la curiosité transmis par mon ancien professeur, Hubertus Th. Jongen, pour l'activité mathématique.

Enfin, je souhaite remercier toutes les personnes et les organisations qui contribuent à faciliter le travail de recherche en mettant à disposition des outils et des logiciels de plus en plus sophistiqués. En particulier, et sans les nommer tous : Latex<sup>1</sup>, bien sur, mais aussi Scilab<sup>2</sup>, Sage<sup>3</sup>, Geogebra<sup>4</sup> et freeFEM++<sup>5</sup>.

---

1. <http://www.latex-project.org/>

2. <http://www.scilab.org/>

3. [http://fr.wikipedia.org/wiki/Sage\\_%28logiciel\\_de\\_calcul\\_formel%29](http://fr.wikipedia.org/wiki/Sage_%28logiciel_de_calcul_formel%29)

4. <http://www.geogebra.org/cms/>

5. <http://www.freefem.org/>





# Table des matières

<b>Introduction</b>	<b>12</b>
<b>Organisation du livre</b>	<b>17</b>
<b>I Introduction</b>	<b>19</b>
<b>1 Cadre pour les problèmes exacts</b>	<b>20</b>
1.1 Notations, conventions, préliminaires . . . . .	20
1.2 Formulation faible du problème d'advection-diffusion stationnaire . . . . .	23
1.3 Formulation faible du problème d'advection-diffusion non stationnaire . . . . .	23
1.4 Problème d'advection stationnaire . . . . .	24
<b>2 Cadre pour les problèmes approchés</b>	<b>29</b>
2.1 Famille de maillage $(\mathcal{T}_h)_h$ : . . . . .	29
2.2 Éléments finis de Crouzeix-Raviart . . . . .	30
2.3 Volumes finis barycentriques, voisinage d'un point $Q_i$ , points en amont/aval de $Q_i$ : . . . . .	32
2.4 Formulation du terme de convection $b_h$ . . . . .	34
2.5 Opérateur d'interpolation $I_h$ . . . . .	35
2.6 Coordonnées locales d'un triangle $T \in \mathcal{T}_h$ : . . . . .	39
2.7 Opérateur gradient local $\nabla_h$ . . . . .	39
2.8 Opérateur local de projection $\rho_h$ . . . . .	40
2.9 Coordonnées locales upwind d'un triangle $T \in \mathcal{T}_h$ : . . . . .	42
2.10 Quelques propriétés de $b_h$ : . . . . .	44
2.11 Formulation faible du problème approché d'advection-diffusion . . . . .	47
2.12 Formulation faible du problème approché d'advection-diffusion non sta- tionnaire . . . . .	47
2.13 Formulation faible du problème approché d'advection . . . . .	47
<b>II Stabilité de l'équation d'advection-diffusion stationnaire</b>	<b>49</b>
<b>3 Introduction</b>	<b>50</b>

4	Conditions géométriques pour le maillage	51
5	Stabilité du problème approché	53
6	Étude de familles de maillage vérifiant les conditions de stabilité	54
6.1	Etude de la condition $\delta - S^+$	54
6.2	Etude des conditions $\delta - S_\beta^+$ -faible et $\delta - S^+ - faible$	61
6.3	Simulations numériques	66
<b>III Stabilité de l'équation d'advection-diffusion non stationnaire</b>		<b>71</b>
7	Introduction	72
8	Stabilité du problème approché	76
<b>IV Estimation de l'erreur pour l'équation d'advection-diffusion non stationnaire</b>		<b>81</b>
9	Introduction	82
9.1	Existence, unicité, conditions de compatibilité pour le problème continu	82
9.2	Estimation de l'erreur	84
10	Estimation de l'erreur	86
11	Simulations numériques	91
<b>V Équation d'advection stationnaire</b>		<b>94</b>
12	Notations, définitions du graphe orienté $\mathcal{G}_h$ associé au maillage $\mathcal{T}_h$	96
13	Existence et unicité de la solution $u_h$	102
14	Encapsulation de l'information dans $\mathcal{G}_h$	107
15	Étapes et idées principales pour les démonstrations	123
16	Calcul de quelques types de cellules	126

<b>17 Stabilité du problème approché</b>	<b>138</b>
17.1 Majoration de $\ u_h\ _{L_h^\infty(\Omega)}$	138
17.2 Majoration de $\ u_h\ _{L_h^1(\Omega)}$ et $\ u_h\ _{L_h^2(\Omega)}$	143
17.3 Majoration de $\ \beta \cdot \nabla_h u_h\ _{L^1(\Omega)}$	162
17.4 Conclusion :	187
<b>VI Annexes</b>	<b>188</b>
<b>A Existence, unicité et estimation d'erreur pour le problème d'advection-diffusion stationnaire</b>	<b>189</b>
<b>B Existence et unicité du problème exact d'advection</b>	<b>191</b>
<b>C Quelques propriétés de <math>d_k = \frac{1}{\ \beta\ ^2} \sum_{j \in S^-(k)} [\beta \cdot (Q_k - Q_j)] \theta_{j,k}</math></b>	<b>196</b>
C.1 Interprétation géométrique de $d_k$	196
C.2 Encapsulation de l'information dans $\mathcal{G}_h$ et forme linéaire $(\cdot, \cdot)_\Pi$	197
<b>D Liste des symboles</b>	<b>201</b>
<b>Références</b>	<b>210</b>

## Introduction

On considère trois types de problèmes :

- a) Le problèmes d'advection-diffusion stationnaire

$$-\nu \Delta u + \beta \cdot \nabla u = f, \quad (1)$$

avec

$$u|_{\partial\Omega} = 0, \quad (1.b)$$

- b) Le problème d'évolution

$$\partial_t u - \nu \Delta u + \beta \cdot \nabla u = g \quad \text{dans } \Omega \times ]0, T[, \quad (2)$$

Les conditions initiales et au bord sont, respectivement,

$$u(x, 0) = u^0(x) \quad \text{pour } x \in \Omega, \quad u|_{\partial\Omega} \times (0, T) = 0,$$

- c) Le problème d'advection

$$\beta \cdot \nabla u = f \quad \text{avec } u|_{\Gamma^-} = g \quad (3)$$

(où  $\Gamma^-$  est le bord entrant de  $\Omega$ ) ,

sur un domaine polygonal borné du plan.

Pour les problèmes approchés, l'approximation se fait par des éléments de Crouzeix Raviart sur un maillage triangulaire pour le terme de diffusion et, pour le terme de convection, on utilise une méthode upwind sur des volumes barycentriques finis.

Pour le problème non-stationnaire, la discrétisation en  $t$  est faite par une méthode implicite d'Euler.

Ces problèmes sont particulièrement intéressants dans le cas où l'advection est dominante ( $\nu \ll |\beta|$ ). Il est généralement admis qu'un lien (même si très distant) relie le problème d'évolution d'advection diffusion au problème d'écoulement compressible de Navier-Stokes.

Dans cet esprit, l'obtention de méthodes numériques stables pour le modèle très simplifié qu'est le problème d'évolution d'advection-diffusion, semble être un bon critère pour le choix des méthodes à retenir pour le problème compressible de Navier-Stokes.

Ainsi, depuis quelques années, des méthodes combinant Éléments Finis (EF) et Volumes Finis (VF) ont gagné en popularité dans les équations de Navier Stokes compressibles à grandes vitesses. Elles ont fait la preuve d'une grande efficacité et robustesse dans la capture d'ondes de chocs et la résolution des couches limites. ([sFFK02], [FFS03, Section 4.4], [HRB97], [DDMN98]).

Dans le contexte de la recherche sur les équations de Navier-Stokes, l'étude se dirige fréquemment vers des équations plus accessibles à une approche théorique.

Ainsi Feistauer et coauteurs [AsFF98], [sFFK02] considèrent le problème non linéaire

$$\partial_t u - \nu \Delta_x u + \sum_{s=1}^2 \partial_{x_s} f_s(u) = g \quad \text{in } \Omega \times (0, T). \quad (4)$$

Ici  $\nu$  et  $T$  sont des réels positifs,  $\Omega \subset \mathbb{R}^2$  est un domaine polygonal borné, et  $f_1, f_2 : \mathbb{R} \mapsto \mathbb{R}$ ,  $g : \Omega \times (0, T) \mapsto \mathbb{R}$  sont des fonctions données. Les conditions initiales et au bord sont respectivement

$$u(x, 0) = u^0(x) \text{ pour } x \in \Omega, \quad \text{et} \quad u|_{\partial\Omega \times (0, T)} = 0.$$

Le problème approché est obtenu par une méthode EF – VF avec une discrétisation d'Euler semi-implicite en  $t$ . Le flux numérique  $H$ , utilisé dans le terme de convection contient, comme cas particulier, la méthode upwind utilisée. Il doit être consistant, conservatif et monotone.<sup>6</sup>

Un exposé de ces résultats est présenté dans [FFS03, Chapter 4.5.5].

Dans ce cadre, ils ont établi la convergence,  $L^\infty$ -stabilité ([AsFF98]), ainsi que des estimations d'erreur ([sFFK02]). Les résultats de stabilité dans [AsFF98] contiennent des majorations indépendantes de  $\nu$ , mais requièrent des maillages de type faiblement aigus. De plus, une dépendance entre la finesse du maillage et le pas en  $t$  empêchent de pouvoir parler de stabilité inconditionnelle. C'est l'utilisation d'un principe du maximum discret, avec le terme de réaction provenant de la discrétisation dans le temps, qui induit ce couplage.

Dans [EHV06] la même combinaison d'Éléments Finis et de Volumes Finis est utilisée pour résoudre une équation d'évolution advection-diffusion non linéaire dégénérée dans deux et trois dimensions. On y montre que la méthode converge si la discrétisation en  $t$  est implicite et une discrétisation upwind adaptée est utilisée pour le terme convectif.

Pour le résultat de stabilité présenté ici, l'approche est différente : dans le cas d'advection-diffusion stationnaire et non stationnaire, une approche variationnelle est utilisée. Elle prend la forme d'une condition inf-sup pour le terme de convection discret de la forme

$$\|v_h\|_2 \leq C \cdot \sup\{b_h(v_h, w_h)/\|w_h\|_2 : w_h \in V_h \setminus \{0\}\} \text{ pour tout } v_h \in V_h,$$

où  $b_h$  est le terme discret de convection et  $V_h$  est l'espace des éléments finis de Crouzeix Raviart qui s'annulent sur le bord.

Cette condition ne sera pas utilisée sous cette forme explicite, mais plutôt sous la forme d'une inégalité dérivée d'une condition géométrique (condition  $\delta - S^+$ ) pour la famille de maillage  $(\mathcal{T}_h)_h$  :

Il existe une constante  $\delta > 0$  indépendante de  $h$  et de  $\beta$  telle que

$$\sum_{j \in S^+(i)} \theta_{i,j}^\beta [\beta \cdot (Q_j - Q_i)] \geq \delta \|\beta\|^2 |S_i|^2 \quad \forall i \in J_0 \text{ et } \beta \in \mathbb{R}^2$$

où  $S_i$  est le côté du triangle dont  $Q_i$  est le milieu.

---

6. En choisissant

$$H(a, b, n) = (n \cdot \beta) a \text{ if } n \cdot \beta > 0, \quad H(a, b, n) = (n \cdot \beta) b \text{ if } n \cdot \beta \leq 0,$$

pour  $a, b \in \mathbb{R}$ ,  $n \in \mathbb{R}^2$  avec  $|n| = 1$ , on retrouve la méthode upwind considérée ici.

Cette condition géométrique est vérifiée par certains maillages structurés, ainsi qu'une petite perturbation de ces maillages. Des exemples concrets sont donnés au chapitre 6.1, page 54.

L'idée, pour exploiter l'inégalité ci-dessus, est d'employer une fonction test de poids exponentiel. Elle a déjà été utilisée par C. Johnson et J. Piktäranta dans [JP86].

La preuve peut être simplifiée techniquement, en utilisant une fonction test affine inspirée de la fonction test exponentielle.

Le terme de réaction qui apparaît avec la discrétisation en  $t$  n'intervient pas spécifiquement dans la preuve. C'est le caractère implicite de la discrétisation, qui permet de traiter ce terme en opérant des majorations entre elles. Voir la preuve des théorèmes 5.1, page 53 et théorèmes 8.1, page 76.

Le théorème 5.1 montre la  $L^2$ -stabilité par rapport au paramètre de diffusion. On retrouve cette stabilité dans le problème d'évolution (théorème 8.1).

Les estimations d'erreur du problème d'évolution dans les normes discrètes de  $L^\infty$ - $L^2$  et  $L^2$ - $H^1$  respectivement, dépendent linéairement de la finesse du maillage et du pas en temps (sans couplage entre les deux). La dépendance de cette estimation en fonction du coefficient de diffusion est explicite.

Dans [sFFK02] les estimations d'erreurs dépendent exponentiellement de  $1/\nu$  et s'expriment dans une certaine norme de la solution continue. Comme remarqué dans [sFFK02, p. 329], [FFcMW99, p. 1546], cette situation n'est pas pleinement satisfaisante. En effet, les estimations d'erreurs sont particulièrement intéressantes pour  $\nu$  petit, car c'est dans cette configuration que l'on peut relever une certaine affinité entre (4) et les équations d'écoulement compressible de Navier-Stokes. Or la solution continue dépend elle-même de  $\nu$  et cette dépendance n'étant pas explicitée, il est donc pas possible de déduire de quelle manière l'erreur dépend de  $\nu$ . Bien sur, on aimerait éviter une estimation qui dépende exponentiellement de  $1/\nu$ .

Le résultat présenté (théorème 9.4, page 84) mesure l'erreur dans une norme discrète correspondant à  $L^\infty(0, T, L^2(\Omega))$  et  $L^2(0, T, H^1(\Omega))$ , respectivement. Ces normes pour l'erreur sont majorées par le produit d'une constante et la somme de la finesse du maillage avec le pas en  $t$ . Cette constante s'exprimant en fonction de  $u_0, g, T, \nu$ .

Le paramètre  $\nu$  y apparaît comme un facteur de la forme  $\nu^{1/2} + \nu^{-5/2}$ . En particulier, il n'y a pas de dépendance exponentielle en  $1/\nu$ , ni couplage entre la finesse du maillage et le pas en  $t$ .

Il est peut être intéressant de remarquer que dans [sFFK02], la dépendance exponentielle n'apparaît pas comme un effet de la non linéarité; elle apparaîtrait aussi dans [DM10], si pour notre problème d'évolution, nous appliquions l'approche qui y est utilisée.

Dans [FFS03, Chapter 4.6], Feistauer et coauteurs considèrent (4) pour une méthode de Galerkin discontinue.

Le point de départ pour le problème d'advection est l'article de Tidiri [Tid02]. Il discrétise (3) (cas où  $\beta$  est constant) par une méthode conforme upwind EF-FV sur des éléments  $P_1$  avec un maillage triangulaire. Pour cette méthode il prouve un résultat de

stabilité de la forme

$$\|u_h\|_2^2 + \|\beta \cdot \nabla u_h\|_2^2 + \int_{\Gamma^+} u_h^2 \beta \cdot n \, d\sigma \leq C \left( \|\nabla \cdot (\beta u)\|_2^2 - \int_{\Gamma^-} g^2 \beta \cdot n \, d\sigma \right)$$

Cependant, il n'est pas clair pourquoi l'opérateur  $P_h$ , introduit dans [Tid02, p. 327], possède bien toutes les propriétés qui lui sont attribuées.

Des résultats de stabilité de la forme

$$\|u_h\|_2^2 + h \|\beta \cdot \nabla u_h\|_2^2 \leq C \left( \|f\|_2^2 - \int_{\Gamma^-} g^2 \beta \cdot n \, d\sigma \right)$$

ont été obtenus dans [JP86], [JNP84] et [LR74].

L'approximation utilisée est de type Galerkin Discontinu (les fonctions sont des polynômes de degré inférieur ou égal à  $p$  sur chaque maille). La forme bilinéaire utilise une pénalisation au sens des moindres carrés des sauts aux interfaces. Dans [EG] la propriété de stabilité prend la forme d'une condition infsup ( $C > 0$  indépendant de  $h$ ) :

$$\forall v \in W_h \quad |||v_h||| \leq C \sup_{w_h \in W_h \setminus \{0\}} \frac{a_h(v_h, w_h)}{|||w_h|||}$$

avec

$$|||v|||^2 = \|v\|_{L^2(\Omega)}^2 + \sum_{T \in \mathcal{T}_h} h_T \|\beta \cdot \nabla v\|_{L^2(\Omega)}^2 + \int_{L^2(\partial\Omega)} v^2 |\beta \cdot n| \, dx + \sum_{S \in \mathcal{F}_h^i} (v^+ - v^-)^2 |\beta \cdot n| \, dx$$

Ici  $v^+ - v^-$  exprime le saut à l'interface  $S$  et  $\mathcal{F}_h^i$  est l'ensemble de ces interfaces à l'intérieur de  $\Omega$ .

L'estimation de l'erreur prend la forme

$$|||u - u_h||| \leq ch^{k+\frac{1}{2}} \|u\|_{H^{k+1}(\Omega)}$$

avec  $c > 0$  indépendant de  $h$ .

Deux hypothèses de régularité sont faites pour la solution :

- a) La solution est suffisamment régulière à l'intérieur de chaque maille;
- b) La solution exacte ne peut sauter à travers une face que si le champ advectif est tangent à cette face. Le dernier terme dans la norme  $||| \cdot |||$  disparaît donc pour la solution du problème exact.

La condition b) nécessite donc d'adapter le maillage à la singularité de la solution exacte.<sup>7</sup>

Pour le problème d'advection étudié ici, la régularité des données  $f$  et  $g$  assure une régularité suffisante pour la solution exacte.

L'approche utilisée est celle de la théorie des graphes. Elle permet de formuler des conditions nécessaires pour l'existence et l'unicité de la solution approchée et aussi

---

7. Voir, par exemple le fichier <http://www.ljll.math.upmc.fr/MathModel/polycopies/ern-galerkin.pdf>



d'obtenir des estimations pour  $\|u_h\|_{L_h^1(\Omega)}$  et  $\|u_h\|_{L_h^2(\Omega)}$ , ainsi que le résultat de stabilité suivant :

$$\begin{aligned} \|u_h\|_{Y_{1,h}(\Omega)} &= \|u_h\|_{L_h^1(\Omega)} + h \|\beta \cdot \nabla_h u_h\|_{L^1(\Omega)} = \|u_h\|_{L_h^1(\Omega)} + h \sum_{T \in \mathcal{T}_h} |\beta \cdot \nabla_h u_h|_T |T| \\ &\leq K \left[ \|g_h\|_{L_h^1(\Gamma^-)} + h_0 \|\beta\| \|g_h\|_{\text{Var}_h(\Gamma^-)} + \left(1 + \frac{1}{\|\beta\|}\right) \|f_h\|_{L_h^1(\Omega)} \right] \quad \forall h \in ]0; h_0[ \end{aligned}$$

avec  $K > 0$  et  $h_0 > 0$ , des constantes indépendantes de  $h$  et  $\|\beta\|$ . Comme pour la stabilité du problème (1), une condition géométrique - qui est équivalente pour les points intérieurs à celle du problème (1) - est nécessaire sur le maillage.

## Autres liens avec la théorie existante

Ohmori, Ushijima [OU84] ont étudié le cas stationnaire (1) dans la situation plus générale où  $\beta$  est une fonction de  $C^1(\overline{\Omega})^2$  (et donc pas nécessairement constant). La discrétisation utilisée pour (1) utilise une méthode upwind avec des éléments de Crouzeix-Raviart, laquelle, pour  $\beta$  constant, correspond à la méthode EF-VF utilisée ici pour (1),(1.b).

Celle-ci est donc comprise dans la théorie de [OU84]. En particulier, d'après [OU84, Theorem 1], notre solution approchée de (1), (1.b) satisfait un principe de maximum discret (et est donc  $L^\infty$ -stable indépendamment de  $\nu$ ) si le membre droit  $f$  de (1) vérifie une condition de signe et si le maillage est de type aigu.

Cependant, une autre situation est étudiée ici :  $f$  doit être dans  $H^1(\Omega)$ , sans condition de signe pour  $f$ .

De plus, un maillage de type aigu, n'est pas nécessaire. La référence [OU84] contient aussi des estimations d'erreur en norme  $H^1(\Omega)$ , mais sans indication sur comment les constantes dans ces estimations dépendent de  $\nu$ . Parmi les autres méthodes de discrétisation de l'équation d'advection-diffusion, Feistauer et coauteurs [FFcM97], [FSS99], [FFcMW99] considèrent une méthode EF-VF pour (4) combinant des éléments conformes  $P_1$  avec volumes duaux finis sur un maillage triangulaire.

La convergence de ce schéma numérique a été établie dans [FFcM97] pour le cas semi-implicite et [FSS99] pour le cas explicite. Des estimations d'erreurs semblables à celles de [sFFK02] ont été établies dans [FFcMW99]. Un aperçu de la théorie développée dans ces références est présenté dans [FFS03, Chapter 4.5]. Enfin, Feistauer et coauteurs considèrent dans [FFS03, Chapter 4.6] des approximations basées sur la méthode EF de Galerkin discontinue.

Tidiri [Tid03] discrétise (1) (cas où  $\nu$  et  $\beta$  constants) par une méthode conforme upwind EF-FV sur des éléments  $P_1$  avec un maillage triangulaire. Pour cette méthode il prouve des estimations d'erreur indépendantes de  $\nu$ , hormis pour les termes impliquant la solution du problème continu. Cependant, il n'est pas clair pourquoi l'opérateur  $P_h$ , introduit dans [Tid02, p. 327], possède bien toutes les propriétés qui lui sont attribuées.

Buffa, Hughes, Sangalli [BHS06] appliquent une méthode d'éléments finis de Galerkin Discontinue multi-échelles à (1). Pour des fonctions  $\nu$  et  $\beta$  constantes par morceaux,

l'erreur est mesurée avec une norme pondérée par  $\nu$  et ces auteurs obtiennent des estimations qui ne dépendent pas de  $\nu$ , sauf à travers la solution du problème continu.

Dans [Li09], l'équation (1) est étudiée sous la condition que  $\beta \in W^{1,\infty}(\Omega)^2$  avec la caractéristique d'un terme de réaction à coefficient non négatif. Cette équation est transformée en un système d'équations du premier degré, qui est discrétisé par une méthode mixte upwind basée sur des éléments de Raviart-Thomas avec un maillage structuré. Les estimations de l'erreur contiennent des constantes dépendant explicitement de  $\nu$  et certaines normes du problème continu.

Plusieurs méthodes des moindres carrés appliquées à (1) – équation réécrite comme système de premier ordre – ont été étudiées par Hsieh et Yang [HY], qui ont obtenu des estimations de l'erreur explicite en  $\nu$  et implicite à travers certaines normes de la solution du problème continu.

Des principes de maximum discrets pour (1) et (2) sont présentés dans [Ike83], ainsi que des méthodes d'éléments finis adaptés aux équations d'advection-diffusion.

Dans [GHL09], [EGHL10], une méthode combinée EF-VF avec des éléments de Crouzeix-Raviart est appliquée à une version non linéaire du système compressible de Navier-Stokes stationnaire.

Une équation dégénérée d'advection-diffusion réaction comme dans [EHV06] est étudiée dans [EHV10]. Les solutions approchées sont obtenues par une méthode combinée EF-VF utilisant des éléments conformes et des volumes de contrôle centrés sur chaque maille. Les solutions approchées obtenues convergent vers la solution exacte du problème continu.

Une très large littérature traite de l'équation stationnaire d'advection-diffusion réaction

$$-\nu \Delta w + (\beta \cdot \nabla)w + \mu w = G \quad \text{in } \Omega \quad (5)$$

ainsi que de l'équation d'évolution correspondante. Ces études sont faites sous l'hypothèse qu'il existe  $\gamma_0 > 0$  avec

$$\mu - \operatorname{div} \beta / 2 \geq \gamma_0, \quad (6)$$

où  $\mu$  et  $\beta$  ne sont pas nécessairement constants. Voir, par exemple [QV97, Chapter 8 and 12], [EG, Section 5.2.3, 5.4.4], [RST08] et les références citées.

La condition (6) n'est, à l'évidence, pas respectée dans notre cas.

Des méthodes d'EF-VF pour (5) sous l'hypothèse (6) ont été développées.

Par exemple, des estimations d'erreur ont été obtenues dans [WL03] pour des maillages mixtes (triangles et rectangles). Des méthodes mixtes de volumes finis sont étudiés dans [BMM<sup>+</sup>06] (stabilité et estimation d'erreur). D'autres articles récents sur ce thème sont [Rui07] et [sFcS09].

## Organisation du livre

La partie introductive précise certaines notations et rappelle un certain nombre de résultats utiles pour les problèmes exacts (chapitre 1, page 20) et leur variante approchée

(chapitre 2, page 29).

Au chapitre II, page 50, on considère le problème stationnaire d'advection-diffusion approché pour lequel un résultat de  $L^2$ -stabilité est montré (théorème 5.1, page 53) pourvu que le maillage satisfasse une condition géométrique (définition 4.1, page 52). Cette condition se trouve être vérifiée par certains maillages structurés (voir le chapitre 4), y compris en cas de perturbation de ces maillages.

Le chapitre III, page 72 présente un résultat de stabilité pour le problème d'évolution advection-diffusion approché. Ce résultat est possible si le maillage remplit la même condition géométrique (définition 4.1, page 52) que dans le problème stationnaire approché.

Une majoration de l'erreur pour le problème d'évolution est possible sans la condition géométrique (définition 4.1, page 52). Ce résultat est démontrée au chapitre IV.

Le chapitre V, qui est grandement indépendant des chapitres précédent, propose une étude de stabilité pour le problème d'advection stationnaire  $b_h(u_h, v_h) = (f_h, v_h)$ . Ici aussi, le maillage doit satisfaire une condition géométrique presque identique à celle du problème stationnaire d'advection-diffusion approché.

Les annexes rappellent quelques résultats connus.

---

## Première partie

# Introduction

## Sommaire

---

<b>1</b>	<b>Cadre pour les problèmes exacts</b>	<b>20</b>
1.1	Notations, conventions, préliminaires . . . . .	20
1.2	Formulation faible du problème d'advection-diffusion stationnaire .	23
1.3	Formulation faible du problème d'advection-diffusion non stationnaire	23
1.4	Problème d'advection stationnaire . . . . .	24
<b>2</b>	<b>Cadre pour les problèmes approchés</b>	<b>29</b>
2.1	Famille de maillage $(\mathcal{T}_h)_h$ : . . . . .	29
2.2	Éléments finis de Crouzeix-Raviart . . . . .	30
2.3	Volumes finis barycentriques, voisinage d'un point $Q_i$ , points en amont/aval de $Q_i$ : . . . . .	32
2.4	Formulation du terme de convection $b_h$ . . . . .	34
2.5	Opérateur d'interpolation $I_h$ . . . . .	35
2.6	Coordonnées locales d'un triangle $T \in \mathcal{T}_h$ : . . . . .	39
2.7	Opérateur gradient local $\nabla_h$ . . . . .	39
2.8	Opérateur local de projection $\rho_h$ . . . . .	40
2.9	Coordonnées locales upwind d'un triangle $T \in \mathcal{T}_h$ : . . . . .	42
2.10	Quelques propriétés de $b_h$ : . . . . .	44
2.11	Formulation faible du problème approché d'advection-diffusion . . .	47
2.12	Formulation faible du problème approché d'advection-diffusion non stationnaire . . . . .	47
2.13	Formulation faible du problème approché d'advection . . . . .	47

---

# 1 Cadre pour les problèmes exacts

## 1.1 Notations, conventions, préliminaires

Dans l'ensemble du texte les notations suivantes seront utilisées<sup>8</sup>

La plupart des propriétés données ici sans preuves sont extraites de [LM68a], [LM68b], [Ada75] [QV97] et [EG].

Le produit scalaire de deux vecteurs  $u, v \in \mathbb{R}^m$  est noté  $u \cdot v$ .

Pour  $1 \leq p < \infty$  et  $u = (u_1, \dots, u_m)$  on définit  $\|u\|_p = (\sum_{i=1}^m |u_i|^p)^{1/p}$  et  $\|u\|_\infty = \max_{i \in \{1, \dots, m\}} |u_i|$ . On a  $\|u\|_2 = \sqrt{u \cdot u}$ . Pour  $p = 2$  on écrit aussi  $\|\cdot\| = \|\cdot\|_2$ .

Si  $u$  et  $v$  sont des éléments d'un espace fonctionnel  $F$  munit d'un produit scalaire, celui-ci sera noté  $(u, v)_F$  et la norme associée  $\|\cdot\|_V$ . Une semi-norme sur  $V$  sera notée  $|\cdot|_V$  : Pour  $u, v \in L^2(\Omega)$  on note ainsi  $(u, v)_{L^2(\Omega)} = \int_\Omega u v dx$ .

Si le domaine de définition des fonctions est  $\Omega$ , une écriture plus concise sera utilisée :

Pour  $u \in L^p(\Omega)$  avec  $1 \leq p < \infty$  on note  $\|u\|_p = \|u\|_{L^p(\Omega)} = (\int_\Omega |u|^p dx)^{1/p}$  la norme de  $u$  dans  $L^p(\Omega)$ . Si  $p = \infty$  alors  $\|u\|_\infty = \|u\|_{L^\infty(\Omega)} = \text{ess sup}_{x \in \Omega} |u(x)|$ .

On note  $\partial_i u = \frac{\partial u}{\partial x_i}$ , la dérivée partielle par rapport à  $x_i$ .

Les espaces de Sobolev suivants seront utilisés ( $\Omega \subset \mathbb{R}^2$ ) :

$$H^1(\Omega) = \{u \in L^2(\Omega) : \partial_1 u, \partial_2 u \in L^2(\Omega)\}$$

munit de la norme

$$\|u\|_{H^1(\Omega)} (= \|u\|_{1,2}) = \left( \int_\Omega |u|^2 + (\partial_1 u)^2 + (\partial_2 u)^2 dx \right)^{1/2},$$

de la semi-norme

$$|u|_{H^1(\Omega)} (= |u|_{1,2}) = \left( \int_\Omega (\partial_1 u)^2 + (\partial_2 u)^2 dx \right)^{1/2},$$

et du produit scalaire

$$(u, v)_{H^1(\Omega)} = (u, v) + \int_\Omega (\partial_1 u \partial_1 v + \partial_2 u \partial_2 v) dx = (u, v) + \int_\Omega \nabla u \cdot \nabla v dx$$

où  $\nabla u = (\partial_1 u, \partial_2 u)$  est le gradient de  $u$ . On définit  $H_0^1(\Omega)$  comme la fermeture de l'ensemble  $C_c^\infty(\Omega)$  des fonctions infiniment dérivables de support compact dans  $\Omega$  dans la topologie définie par la norme de  $H^1(\Omega)$ .

$$H_0^1(\Omega) = \overline{C_c^\infty(\Omega)}^{H^1(\Omega)}$$

On suppose que  $\Omega$  est un domaine polygonal borné de  $\mathbb{R}^2$ .

---

8. Une liste détaillée des notations utilisées se trouve en annexe D

**Définition 1.1 (domaine polygonal)** *En dimension deux, un domaine polygonal borné est un ouvert borné, non vide et connexe de  $\mathbb{R}^2$  tel que sa frontière  $\Gamma = \partial\Omega$  est une union finie de segments.*

On définit  $L^2(\Gamma) = \{g : \int_{\Gamma} g^2 d\sigma < +\infty\}$  où  $d\sigma$  est mesure superficielle. C'est un espace de Hilbert pour le produit scalaire

$$(g_1, g_2)_{L^2(\Gamma)} = \int_{\Gamma} g_1 g_2 d\sigma$$

Ici  $\Gamma$  est de dimension 1 (car  $\Omega \subset \mathbb{R}^2$ ) et on définit  $H^{1/2}(\Gamma)$  par

$$H^{1/2}(\Gamma) = \left\{ g \in L^2(\Gamma) : \int_{\Gamma} \int_{\Gamma} \frac{|g(x) - g(y)|^2}{|x - y|^2} d\sigma d\sigma < \infty \right\}$$

C'est un espace de Hilbert pour le produit scalaire

$$(g_1, g_2)_{H^{1/2}(\Gamma)} = (g_1, g_2)_{L^2(\Gamma)} + \int_{\Gamma} \int_{\Gamma} \frac{|g_1(x) - g_1(y)| |g_2(x) - g_2(y)|}{|x - y|^2} d\sigma d\sigma$$

**Théorème 1.1 (trace)** *Si  $\Omega$  est un domaine polygonal borné, alors*

- a) *Il existe une unique application linéaire et continue  $\gamma_0 : H^1(\Omega) \rightarrow H^{1/2}(\Gamma)$  telle que  $\gamma_0(v) = v|_{\Gamma}$  pour tout  $v \in H^1(\Omega) \cap C(\overline{\Omega})$ .*
- b) *Pour tout  $\phi \in H^{1/2}(\Gamma)$ , il existe un relèvement  $R_0\phi \in H^1(\Omega)$  tel que  $\gamma_0(R_0\phi) = \phi$ .*

**Remarque 1.2** *On a  $C^{0,1}(\Gamma) \subset H^{1/2}(\Gamma)$  où  $C^{0,1}(\Gamma)$  est l'espace des fonctions lipschitziennes sur  $\Gamma$ . Muni de la norme*

$$\|\phi\|_{C^{0,1}(\Gamma)} = \|\phi\|_{L^\infty(\Gamma)} + \sup_{x,y \in \Gamma} \frac{|\phi(x) - \phi(y)|}{|x - y|}$$

*c'est un espace de Banach. Par contre, il existe des fonctions continues sur  $\Gamma$  qui n'appartiennent pas à  $H^{1/2}(\Gamma)$*

On définit aussi  $C^{0,1}(\Omega) \subset H^1(\Omega)$  où  $C^{0,1}(\Omega)$  est l'espace des fonctions lipschitziennes sur  $\Omega$ . Muni de la norme

$$\|v\|_{C^{0,1}(\Omega)} = \|v\|_{L^\infty(\Omega)} + \sup_{x,y \in \Omega} \frac{|v(x) - v(y)|}{|x - y|} \quad (1.1)$$

c'est un espace de Banach.

Si  $\Omega$  est un domaine polygonal borné, alors pour toute fonction  $u \in C^{0,1}(\Omega) \subset H^1(\Omega)$  on a  $\gamma_0(u) \in C^{0,1}(\Gamma)$ .

Si  $\Omega$  est un domaine polygonal borné, on a la caractérisation

$$H_0^1(\Omega) = \{u \in H^1(\Omega) : \gamma_0(u) = 0\}$$

Par la suite on omettra l'écriture du symbole  $\gamma_0$ . Pour  $\Sigma \subset \Gamma$  de mesure strictement positive, on définit l'espace fonctionnel

$$H_{\Sigma}^1(\Omega) = \{v \in H^1(\Omega) : v = 0 \text{ sur } \Sigma\}$$

**Théorème 1.2 (inégalité de Poincaré)** *Si  $\Omega$  est un domaine polygonal borné et  $\Sigma \subset \Gamma$  de mesure strictement positive, alors il existe une constante  $C_{\Omega} > 0$  telle que*

$$\forall v \in H_{\Sigma}^1(\Omega), \quad C_{\Omega} \int_{\Omega} v^2 dx \leq \int_{\Omega} \nabla v \cdot \nabla v dx$$

L'inégalité de Poincaré montre que la semi-norme  $v \mapsto |v|_{H^1(\Omega)}$  est une norme pour  $V = H_0^1(\Omega)$ , équivalente à sa norme usuelle.

On définit

$$\|v\|_{*V} = |v|_{H^1(\Omega)} \text{ avec le produit scalaire } ((u, v)) = \int_{\Omega} \nabla u \cdot \nabla v dx \quad \forall u, v \in V$$

Il existe donc des constantes  $c_1$  et  $c_*$  telles que

$$\frac{1}{c_1} \|v\|_{*V} \leq |v|_{H^1(\Omega)} \leq c_1 \|v\|_{*V} \quad \forall v \in V \quad (1.2)$$

et

$$\|v\|_{L^2(\Omega)} \leq c_* \|v\|_{*V} \quad \forall v \in V \quad (1.3)$$

L'écriture  $\|\nabla u\|_{L^2(\Omega)}$  sera parfois (abusivement) utilisée pour  $\|u\|_{*V}$ .

Soit

$$H^2(\Omega) = \{u \in L^2(\Omega) : \partial_1 u, \partial_2 u, \partial_i \partial_j u \in L^2(\Omega), \forall i, j \in 1, 2\}$$

avec norme

$$\|u\|_{H^2(\Omega)} = \|u\|_{2,2} = \left( \int_{\Omega} |u|^2 + |\partial_1 u|^2 + |\partial_2 u|^2 + \sum_{i,j=1}^2 |\partial_i \partial_j u|^2 dx \right)^{1/2}$$

Pour un domaine  $K$  de  $\mathbb{R}^2$ ,  $n_K$  est la normale extérieure à  $K$ .

Si  $n$  apparaît dans une intégrale  $\int_{\partial T} f \beta \cdot n d\sigma$ , alors  $n = n_T$  désignera (sans ambiguïté) la normale extérieure à  $T$ .  $d\sigma$  est la mesure superficielle sur  $\partial T$  induite par la mesure de Lebesgue  $dx$  définie sur  $\mathbb{R}^2$  et  $f$  est une fonction scalaire.

On note  $n(x) = (n_1(x), n_2(x))$  la normale extérieure à  $\Omega$  au point  $x \in \Gamma$ .

Soit  $\Omega$  un domaine polygonal borné, on dispose de la formule de Green :

$$\int_{\Omega} (\partial_j w) v dx = - \int_{\Omega} w \partial_j v dx + \int_{\partial \Omega} w v n_j d\sigma, \quad j = 1, 2 \quad \forall w, v \in H^1(\Omega)$$

ou encore

$$\forall u \in H^2(\Omega) \forall v \in H^1(\Omega), \quad - \int_{\Omega} \Delta u v \, dx = \int_{\Omega} \nabla u \cdot \nabla v \, dx - \int_{\Gamma} v \nabla u \cdot n \, d\sigma$$

où  $\Delta u = (\partial_1 \partial_1 + \partial_2 \partial_2)(u)$  est le laplacien de  $u$ .

## 1.2 Formulation faible du problème d'advection-diffusion stationnaire

Soit  $0 \neq \beta \in \mathbb{R}^2$ ,  $\Omega$  un domaine polygonal borné et  $f \in L^2(\Omega)$ . Au chapitre II, on considère le problème stationnaire d'advection-diffusion

$$-\nu \Delta u + \beta \cdot \nabla u = f \text{ dans } \Omega \quad (1.4)$$

$$u = 0 \text{ sur } \partial\Omega \quad (1.5)$$

Une formulation faible de (1.4) est obtenue par multiplication de cette équation par  $v \in V = H_0^1(\Omega)$  et intégration sur  $\Omega$ .

Avec la formule de Green, on obtient

**Définition 1.3** Une fonction  $u \in V$  est une solution faible du problème (1.4), si

$$\nu((u, v)) + b(u, v) = \int_{\Omega} f v \, dx, \quad \forall v \in V \quad (1.6)$$

où

$$b(u, v) = \int_{\Omega} \beta \cdot \nabla u v \, dx \text{ pour } u, v \in V \quad (1.7)$$

## 1.3 Formulation faible du problème d'advection-diffusion non stationnaire

Au chapitre III, on étudie le problème d'évolution dans le domaine espace/temps  $Q_T := \Omega \times ]0, T[$  ( $0 < T < \infty$ ). Trouver  $u : \overline{Q_T} \rightarrow \mathbb{R}$ , tel que

$$\frac{\partial u}{\partial t} - \nu \Delta u + \beta \cdot \nabla u = g \text{ in } Q_T \quad (1.8)$$

$$u|_{\partial\Omega \times ]0, T[} = 0 \quad (1.9)$$

$$u(x, 0) = u^0(x), \forall x \in \Omega \quad (1.10)$$

avec  $\nu > 0$  et  $g : Q_T \rightarrow \mathbb{R}$ ,  $u^0 : \Omega \rightarrow \mathbb{R}$  des fonctions données.

Avant de déduire une formulation faible de (1.8), il est nécessaire d'introduire les espaces de Banach suivants :

Soit  $X$  un espace de Banach, alors  $C([0, T], X)$  est l'espace des fonctions continues bornées  $t \mapsto u(t) = u(t, \cdot) \in X$  muni de la norme  $\|u\|_{C([0, T], X)} = \sup_{t \in [0, T]} \|u(t)\|_X$ . L'espace  $C^n([0, T], X)$  est l'espace des fonctions dont les dérivées par rapport au temps jusqu'à l'ordre  $n$  existent et sont dans  $C([0, T], X)$ .



$L^2(]0, T[, X)$ ,  $L^\infty(]0, T[, X)$  sont les fermetures de  $C(]0, T[, X)$  par rapport à la topologie définie par les normes  $\|u\|_{L^2(]0, T[, X)} = \left(\int_0^T \|u(t)\|_X^2 dt\right)^{1/2}$ ,  $\|u\|_{L^\infty(]0, T[, X)} = \sup \operatorname{ess}_{t \in ]0, T[} \|u(t)\|_X$ .

L'espace de Sobolev  $H^1(]0, T[, X)$  est l'ensemble des fonctions  $u \in L^2(]0, T[, X)$  telles que  $u' = \frac{du}{dt} \in L^2(]0, T[, X)$ . Cet espace est muni de la norme

$$\|u\|_{H^1(]0, T[, X)} = \left(\int_0^T \|u\|_X^2 + \|u'\|_X^2 dt\right)^{1/2}$$

On suppose que

$$g \in C(]0, T[, H^1(\Omega)) \text{ and } u^0 \in V \quad (1.11)$$

Supposons que  $u$  soit une solution au sens classique, alors en multipliant (1.8) par  $v \in V$  et intégrant sur  $\Omega$ , on obtient avec la formule de Green (en ayant permuté intégration sur  $\Omega$  et la dérivée par rapport à  $t$ ) l'expression suivante

$$\begin{aligned} & \frac{d}{dt} \int_{\Omega} u(t) v dx + \nu \int_{\Omega} \nabla u(t) \cdot \nabla v dx + \int_{\Omega} \beta \cdot \nabla u(t) v dx \\ &= \int_{\Omega} g(t) v dx \end{aligned} \quad (1.12)$$

Reprenant la terminologie de [sFFK02], on définit :

**Définition 1.4** Une fonction  $u$  est une solution faible de (1.8), si et seulement si :

- a)  $u \in L^2(]0, T[, V) \cap L^\infty(Q_T)$  (1.13)
- b)  $\frac{d}{dt}(u(t), v) + \nu((u, v)) + b(u, v) = (g(t), v), \forall v \in V$   
pris dans le sens des distributions sur  $]0, T[$
- c)  $u(0) = u^0$

## 1.4 Problème d'advection stationnaire

Pour  $0 \neq \beta \in \mathbb{R}^2$  on définit respectivement

$$\Gamma^+ = \{x \in \Gamma : n(x) \cdot \beta > 0\} \quad (1.14)$$

$$\Gamma^- = \{x \in \Gamma : n(x) \cdot \beta < 0\} \quad (1.15)$$

$$\Gamma^b = \{x \in \Gamma : n(x) \cdot \beta = 0\} \quad (1.16)$$

les parties de la frontière où le champ d'advection  $\beta$  est sortant, entrant et longeant. On a  $\Gamma^+ \cup \Gamma^- \cup \Gamma^b = \Gamma$  et on note  $n_{\Gamma^+}$ ,  $n_{\Gamma^-}$  et  $n_{\Gamma^b}$  les restrictions de  $n$  respectivement à ces ensembles.

L'espace des solutions :

$$Y_1 = \{v \in L^1(\Omega) : \beta \cdot \nabla v \in L^1(\Omega)\} \quad (1.17)$$

L'espace  $Y_1$ , avec la semi-norme  $|v|_{Y_1} = |\beta \cdot \nabla v|_{L^1(\Omega)}$  et la norme  $\|v\|_{Y_1} = \|v\|_{L^1(\Omega)} + |v|_{Y_1}$ , est un espace de Banach.

On considère le problème d'advection suivant :

**Définition 1.5 (Problème advection)** *Trouver  $u \in Y_1$  tel que*

$$\begin{cases} \beta \cdot \nabla u = f \text{ sur } \Omega \\ \text{avec } u|_{\Gamma^-} = g \end{cases} \quad (1.18)$$

avec  $g : \Gamma^- \rightarrow \mathbb{R}$  et  $f \in L^1(\Omega)$ .

La méthode des caractéristiques permet maintenant de montrer l'existence et l'unicité de la solution du problème (1.18) pour  $f \in W^{1,1}(\Omega) \subset L^1(\Omega)$  et  $g \in W^{1,1}(\Gamma^-)$  :

**Théorème 1.3 (Existence et unicité du problème d'advection)**

*Soit  $\emptyset \neq \Omega \subset \mathbb{R}^2$  un domaine polygonal borné, tel que le bord entrant  $\Gamma^-$  soit connexe,  $f \in W^{1,1}(\Omega)$  et  $g \in W^{1,1}(\Gamma^-)$ , alors il existe exactement une solution  $u \in W^{1,1}(\Omega) \subset Y_1$  de (1.18) et on a les inégalités*

$$\|u\|_{L^1(\Omega)} \leq \text{diam}(\Omega) \|g\|_{L^1(\Gamma^-)} + \frac{\text{diam}(\Omega)}{\|\beta\|} \|f\|_{L^1(\Omega)} \quad (1.19)$$

$$\|\beta \cdot \nabla u\|_{L^1(\Omega)} = \|f\|_{L^1(\Omega)} \quad (1.20)$$

$$\begin{aligned} \|\nabla u\|_{L^1(\Omega)} &\leq (C_{\Gamma^-, \beta} + 1) \text{diam}(\Omega) \|g\|_{L^1(\Gamma^-)} + \frac{C_{\Gamma^-, \beta}}{\|\beta\|} \|f\|_{L^1(\Omega)} \\ &\quad + \frac{\text{diam}(\Omega)}{\|\beta\|} (C_{\Gamma^-, \beta} + 1) \|\nabla f\|_{L^1(\Omega)} \end{aligned} \quad (1.21)$$

$$\text{avec } C_{\Gamma^-, \beta} = \max_{x \in \Gamma^-} \frac{\|\beta\|}{|\beta \cdot n(x)|}. \quad (1.22)$$

$$(1.23)$$

Ainsi

$$\|u\|_{Y_1} \leq \text{diam}(\Omega) \|g\|_{L^1(\Gamma^-)} + \left(1 + \frac{\text{diam}(\Omega)}{\|\beta\|}\right) \|f\|_{L^1(\Omega)} \quad (1.24)$$

$$\|u\|_{W^{1,1}(\Omega)} \leq \tilde{C} (\|g\|_{W^{1,1}(\Gamma^-)} + \|f\|_{W^{1,1}(\Omega)}) \text{ et} \quad (1.25)$$

$$\tilde{C} \text{ est une constante qui dépend de } \beta, \Gamma^- \text{ et } \Omega. \quad (1.26)$$

**Preuve :** Voir l'annexe page 191. □

Cela motive la définition du sous-espace vectoriel :

$$Y_{\text{sol}} = \{w \in Y_1 : w \in W^{1,1}(\Omega), w|_{\Gamma^-} \in W^{1,1}(\Gamma^-) \text{ et } \beta \cdot \nabla w \in W^{1,1}(\Omega)\} \subset W^{1,1}(\Omega) \cap Y_1 \quad (1.27)$$

**Remarque 1.6** Dans le sous-espace vectoriel  $Y_{\text{sol}}$  la norme  $\|\cdot\|_{Y_1}$  peut être remplacée par la norme  $\|u\|_{Y_{\text{sol}}} = \|\beta \cdot \nabla u\|_{L^1(\Omega)} + \|u|_{\Gamma^-}\|_{L^1(\Gamma^-)}$ , car avec (1.19) et (1.20) on a  $\|u\|_{Y_1} \leq \text{diam}(\Omega) \|u|_{\Gamma^-}\|_{L^1(\Gamma^-)} \left( \frac{\text{diam}(\Omega)}{\|\beta\|} + 1 \right) \|\beta \cdot \nabla u\|_{L^1(\Omega)}$ .

Les domaines polygonaux bornés vérifient la condition lipschitzienne forte au bord<sup>9</sup> et donc aussi la propriété du cône<sup>10</sup> Le théorème d'injection suivant s'applique (voir [Ada75], page 97) :

**Théorème 1.4 (théorème d'injection de Sobolev)** Soit  $\Omega$  un domaine de  $\mathbb{R}^n$  et soit  $\Omega^k$  le domaine de dimension  $k$  obtenu par l'intersection de  $\Omega$  avec une surface de dimension  $k$  dans  $\mathbb{R}^n$ ,  $1 \leq k \leq n$  (donc  $\Omega^n = \Omega$ ). Soient  $j$  et  $m$  des entiers positifs ou nuls et soit  $1 \leq p < \infty$ .

**Partie I** Si  $\Omega$  a la propriété du cône, alors il existe les injections continues suivantes :

**Cas A** Si  $mp < n$  et  $n - mp < k \leq n$ , alors

$$W^{j+m,p}(\Omega) \rightarrow W^{j,q}(\Omega^k), \quad \text{pour } p \leq q \leq kp/(n - mp), \quad (1.28)$$

et en particulier ,

$$W^{j+m,p}(\Omega) \rightarrow W^{j,q}(\Omega), \quad \text{pour } p \leq q \leq np/(n - mp), \quad (1.29)$$

ou encore

$$W^{m,p}(\Omega) \rightarrow L^q(\Omega), \quad \text{pour } p \leq q \leq np/(n - mp). \quad (1.30)$$

De plus, si  $p = 1$  et  $m < n$ , l'injection continue (1.28) existe aussi pour  $k = n - m$ .

**Cas B** Supposons que  $mp = n$ . Alors pour  $k$  avec  $1 \leq k \leq n$ ,

$$W^{j+m,p}(\Omega) \rightarrow W^{j,q}(\Omega^k), \quad \text{pour } p \leq q < \infty, \quad (1.31)$$

et en particulier ,

$$W^{m,p}(\Omega) \rightarrow L^q(\Omega), \quad \text{pour } p \leq q < \infty, \quad (1.32)$$

De plus, si  $p = 1$  et  $m = n$ , les injections continues (1.31) et (1.32) existent aussi pour  $q = \infty$ . On a

$$W^{j+n,1}(\Omega) \rightarrow C_B^j(\Omega) = \{u \in C^j(\Omega) : D^\alpha u \text{ est borné dans } \Omega \text{ pour } |\alpha| \leq j\}$$

(Avec la norme  $\|u\|_{C_B^j(\Omega)} = \max_{0 \leq |\alpha| \leq j} \sup_{x \in \Omega} |D^\alpha u(x)|$ ,  $C_B^j(\Omega)$  est un espace de Banach.)

**Cas C** Si  $mp > n$ , alors

$$W^{j+m,p}(\Omega) \rightarrow C_B^j(\Omega) \quad (1.33)$$

**Partie II** Si  $\Omega$  vérifie la condition lipschitzienne forte au bord, alors le cas C de la partie I peut être remplacé par :

**Cas C'** Si  $mp > n > (m - 1)p$ , alors

$$W^{j+m,p}(\Omega) \rightarrow C^{j,\lambda}(\overline{\Omega}), \quad 0 < \lambda \leq m - (n/p). \quad (1.34)$$

9. "strong local Lipschitz property", voir [Ada75], page 66/67

10. "cone property", voir [Ada75], page 66/67.

**Cas C''** Si  $n = (m - 1)p$ , alors

$$W^{j+m,p}(\Omega) \rightarrow C^{j,\lambda}(\overline{\Omega}), \quad 0 < \lambda < 1. \quad (1.35)$$

Si maintenant  $n = m - 1$  et  $p = 1$ , alors (1.35) est valide aussi pour  $\lambda = 1$  ( $C^{j,\lambda}(\overline{\Omega})$  est le sous espace de  $C^j(\overline{\Omega})$  défini par les fonctions  $\phi$ , telles que  $D^\alpha \phi$  soit  $\lambda$ -Hölderienne.  $C^{j,\lambda}(\overline{\Omega})$  est un espace de Banach pour la norme  $\|\phi\|_{C^{j,\lambda}(\overline{\Omega})} = \|\phi\|_{C^j(\overline{\Omega})} + \max_{0 \leq |\alpha| \leq j} \sup_{x,y \in \Omega \text{ avec } x \neq y} \frac{|D^\alpha u(x) - D^\alpha u(y)|}{\|x-y\|^\lambda}$ )

### Remarque 1.7

- 1) Les fonctions dans  $W^{1,1}(\Omega)$  ne sont pas nécessairement continues et peuvent contenir des singularités. Cependant, le cas A du théorème ci-dessus garantit l'existence, pour  $p = 1$ ,  $m = 1$ ,  $k = 1$  et  $j = 0$ , de l'intégrale  $\int_{[AB]} w(x) d\sigma$  pour tout segment  $[AB] \subset \Omega$  et  $w \in W^{1,1}(\Omega)$ . En particulier  $I_h(w)$  (voir définition 2.3, page 35 pour la définition de l'opérateur d'interpolation  $I_h$ ) est donc bien défini pour  $w \in W^{1,1}(\Omega)$ .
- 2) Du cas B (partie I) on déduit que  $W^{2,1}(\Omega) \rightarrow C_B^0(\Omega)$ , c'est-à-dire que  $f \in C_B^0(\Omega)$ . La solution  $u$  du problème d'advection est alors continue, c'est le théorème 1.5.

### Théorème 1.5 (Existence et unicité du problème d'advection 2)

Soit  $\emptyset \neq \Omega \subset \mathbb{R}^2$  un domaine polygonal borné, tel que le bord entrant  $\Gamma^-$  soit connexe,  $f \in W^{2,1}(\Omega)$  et  $g \in W^{1,1}(\Gamma^-)$ , alors il existe exactement une solution  $u \in W^{1,1}(\Omega) \cap C(\overline{\Omega})$  de (1.18). Les conclusions du théorème 1.18 restent valides et on a<sup>11</sup>

$$\|u\|_{L^\infty(\Omega)} \leq \|g\|_{L^\infty(\Gamma^-)} + \frac{\text{diam}(\Omega)}{\|\beta\|} \|f\|_{L^\infty(\Omega)} \quad (1.36)$$

**Preuve :** Voir l'annexe page 191. □

### Remarque 1.8

- a) L'existence et l'unicité ne sont pas obtenues pas des méthodes d'analyse fonctionnelle. Les espaces  $L^1(\Omega)$  et  $L^\infty(\Omega)$  ne sont pas réflexifs et le théorème de Banach-Nečas-Babuška ([BA72], p. 112, [EG], p. 85 ou [QV97], p 139) ne s'applique pas.
- b) La situation est similaire avec le problème approché : Le problème upwind de Galerkin Discontinue utilisé (avec des éléments de Crouseix-Raviart non conformes) va être interprété comme un problème de calcul de valeurs aux sommets d'un graphe orienté. Ainsi, l'existence et l'unicité de la solution du problème approché sont assurées, si le schéma itératif (plus précisément le graphe associé) qui permet de calculer les valeurs de  $u_h(Q_i)$ , ne contient pas de circuit élémentaire.
- c) L'objectif du chapitre V est de prouver les inégalités correspondantes à (1.20) et (1.36) pour le problème approché (voir chapitre 2.13, page 47 pour la formulation variationnelle du problème approché).

11. Les fonctions  $g$  de  $W^{1,1}(\Gamma^-)$  sont continues (ou plus exactement la classe d'équivalence correspondant à  $g$  possède un représentant  $\hat{g}$  continu), ce qui justifie  $\|g\|_{L^\infty(\Gamma^-)}$ . Voir par exemple [Bre83], page 122.

La forme variationnelle de (1.18) est obtenue en multipliant (1.18) par  $v \in L^\infty(\Omega)$  et en intégrant sur  $\Omega$  :

**Définition 1.9 (Formulation variationnelle du problème d'advection)**

*Trouver  $u \in Y_1$  tel que*

$$\int_{\Omega} \beta \cdot \nabla u \, v \, dx = (f, v) \quad \forall v \in L^\infty(\Omega) \quad \text{et } u|_{\Gamma^-} = g \quad (1.37)$$

## 2 Cadre pour les problèmes approchés

### 2.1 Famille de maillage $(\mathcal{T}_h)_h$ :

Soit  $\emptyset \neq \Omega \subset \mathbb{R}^2$  un domaine polygonal borné, on considère une famille de maillage  $(\mathcal{T}_h)_h$  de  $\Omega$  avec  $h \in ]0; h_0[$ , dont les éléments (mailles)  $T \in \mathcal{T}_h$  sont des triangles fermés.

$$\forall h \in ]0; h_0[, \quad \overline{\Omega} = \cup_{T \in \mathcal{T}_h} T$$

On note  $I$  l'ensemble des indices (positifs)  $i$  permettant d'indexer les triangles  $T_i$  de  $\mathcal{T}_h$ .

$$\mathcal{T}_h = \{T_i : i \in I\}$$

Pour deux triangles  $T_k$  et  $T_l$  de  $\mathcal{T}_h$  les contraintes de dispositions suivantes doivent être satisfaites :

- a)  $T_k \cap T_l = \emptyset$  ou
- b) Les triangles  $T_k$  et  $T_l$  ont un côté en commun
- c) Les triangles  $T_k$  et  $T_l$  ont un sommet en commun

Avec  $h_T$ ,  $\rho_T$  et  $h$  on note respectivement le diamètre de  $T$ , le diamètre du cercle inscrit dans  $T$  et la finesse du maillage

$$h_T = \text{diam}(T) = \max_{x_1, x_2 \in T} \|x_1 - x_2\|_2 \quad h = \max_{T \in \mathcal{T}_h} h_T$$

$$\rho_T = \sup\{\text{diam}(B) : B \text{ boule de } \mathbb{R}^2, B \subset T\}$$

**Définition 2.1 (régularité)** Une famille de maillage  $(\mathcal{T}_h)_h$  est régulière s'il existe une constante  $\sigma$  telle que

$$\forall h \in ]0; h_0[ \quad \forall T \in \mathcal{T}_h \quad \sigma_T = \frac{h_T}{\rho_T} \leq \sigma \quad (2.38)$$

**Définition 2.2 (quasi-uniformité)** Une famille de maillage  $(\mathcal{T}_h)_h$  est quasi-uniforme si elle est régulière et s'il existe une constante  $c_0$  telle que

$$\forall h, \quad \forall T \in \mathcal{T}_h, \quad \frac{h}{h_T} \leq c_0 \quad (2.39)$$

Des inégalités (2.39) et (2.38) on déduit

**Lemme 2.1** Il existe  $\hat{c}_0 > 0$  indépendant de  $h$  et  $T$  avec

$$h^2 \leq \hat{c}_0 |T| \quad \forall T \in \mathcal{T}_h \quad \forall h \in ]0; h_0[ \quad (2.40)$$

Soit  $h \in ]0; h_0[$ , alors l'ensemble  $J$  définit une indexation de l'ensemble des côtés  $\{S_i : i \in J\}$  des triangles de  $\mathcal{T}_h$ . On note  $Q_i$  le milieu du segment  $S_i$ . L'ensemble des indices des points qui ne sont pas sur le bord est donné par :

$$J_0 = \{j \in J : Q_j \notin \partial\Omega\} \quad (2.41)$$

Pour le problème d'advection défini par  $0 \neq \beta \in \mathbb{R}^2$ , on aura besoin des ensembles suivants :

$$J^+ = \{j \in J : Q_j \in \Gamma^+\} \quad (2.42)$$

$$J^- = \{j \in J : Q_j \in \Gamma^-\} \quad (2.43)$$

$$J^b = \{j \in J : Q_j \in \Gamma^b\} \quad (2.44)$$

où  $J^+$ ,  $J^-$  et  $J^b$  sont respectivement les ensembles des indices des points  $Q_i$  sur la partie sortante, entrante et longeant du bord. Ces ensembles interviennent quand le terme de diffusion est absent de l'équation.

**Lemme 2.2 (Relations d'Euler)** *Soit  $\Omega$  un domaine polygonal de  $\mathbb{R}^2$ . soit  $\mathcal{T}_h$  un maillage de  $\Omega$ . Soit  $\kappa$  le degré de multiple connexité<sup>12</sup>, alors*

*$\text{card}(I) - \text{card}(J) + N_{\text{nombre de sommet}} = 1 - \kappa$  et  $N_{\text{nombre de sommet sur } \Gamma} = N_{\text{nombre d'arête sur } \Gamma}$*

*et*

$$2 \text{card}(J) - N_{\text{nombre d'arête sur } \Gamma} = 3 \text{card}(I)$$

*où  $\text{card}(I)$  est le cardinal de  $I$ .*

## 2.2 Éléments finis de Crouzeix-Raviart

Les éléments finis non conformes de Crouzeix-Raviart sont utilisés pour la résolution du problème approché.

La régularité et la quasi-uniformité de la famille de maillage permettent de retrouver l'essentiel des propriétés données dans le cadre du problème exact.

Tout d'abord les espaces utilisés dans la résolution des problèmes approchés :

Soit un  $(\mathcal{T}_h)_h$  une famille de maillage triangulaire, régulière et quasi-uniforme. On définit

**Espace pour le problème approché d'advection-diffusion :**

$$X_h = \{v_h \in L^2(\Omega) : v_h|_T \text{ est affine } \forall T \in \mathcal{T}_h, v_h \text{ est continu en } Q_j, \forall j \in J\} \quad (2.45)$$

$X_h$  est un espace de Hilbert équipé du produit scalaire

$$(w_h, v_h)_{X_h} = (w_h, v_h) + ((w_h, v_h))_h \text{ où } (w_h, v_h) = \int_{\Omega} w_h v_h dx \quad (2.46)$$

$$\text{et } ((w_h, v_h))_h = \sum_{T \in \mathcal{T}_h} \int_T \nabla w_h \cdot \nabla v_h dx \quad (2.47)$$

---

12. Le nombre de "trous" dans  $\Omega$

On a ainsi pour la norme et la semi-norme induite par le produit scalaire :

$$\|v_h\|_{X_h} = (v_h, v_h)_{X_h}^{1/2} \text{ et } |v_h|_{X_h} = ((v_h, v_h))_h^{1/2}$$

L'ensemble  $\{w_{h,i} : i \in J\}$  des fonctions  $w_{h,i} \in X_h$  telles que  $w_{h,i}(Q_j) = \delta_{ij} = \begin{cases} 1 & \text{si } i=j \\ 0 & \text{si } i \neq j \end{cases}$  forme une base de  $X_h$ .

Avec le lemme de Friedrichs 2.12 on dispose de la version discrète de  $V$  :

$$V_h = \{v_h \in X_h : v_h(Q_i) = 0, \forall i \in J - J_0\} \quad (2.48)$$

est un sous-espace vectoriel de  $X_h$  pour lequel la semi-norme  $|\cdot|_{X_h}$  est une norme équivalente à  $\|\cdot\|_{X_h}$ . On note  $\|\cdot\|_{*V_h} = |\cdot|_{X_h}$  cette norme induite par  $((\cdot, \cdot))_h$  (qui est maintenant un produit scalaire sur  $V_h$ ).

**Espaces  $Y_h$  pour le problème approché d'advection :**

$$Y_h = \{v_h : \bar{\Omega} \rightarrow \mathbb{R} : v_h|_T \text{ est affine } \forall T \in \mathcal{T}_h, v_h \text{ est continu en } Q_j, \forall j \in J\} \quad (2.49)$$

Pour  $Y_h$  on considère différentes normes et semi-normes :

$$\|w_h\|_{L_h^p(\Omega)} = \left( \sum_{k \in J} |w_h(Q_k)|^p |D_k| \right)^{1/p} \text{ pour } p \in [1; \infty[ \text{ et} \quad (2.50)$$

$$\|w_h\|_{L_h^\infty(\Omega)} = \max_{k \in J} |w_h(Q_k)| \quad (2.51)$$

$$|w_h|_{Y_{p,h}} = \left( \sum_{T \in \mathcal{T}_h} \|\beta \cdot \nabla_h w_h\|_{L^p(T)}^p \right)^{1/p} \text{ pour } p \in [1; \infty[ \text{ et} \quad (2.52)$$

$$|w_h|_{Y_{\infty,h}} = \max_{T \in \mathcal{T}_h} |\beta \cdot \nabla_h w_h|_T \quad (2.53)$$

$$\|w_h\|_{Y_{p,h}} = \left( \|w_h\|_{L_h^p(\Omega)}^p + h |w_h|_{Y_{p,h}}^p \right)^{1/p} \text{ pour } p \in [1; \infty[ \text{ et} \quad (2.54)$$

$$\|w_h\|_{Y_{\infty,h}} = \max\{\|w_h\|_{L_h^\infty(\Omega)}, h |w_h|_{Y_{\infty,h}}\} \quad (2.55)$$

où  $|D_k|$  est la mesure du volume barycentrique  $D_k$  associé à  $Q_k$ .

On considère aussi le sous-espace des fonctions de  $Y_h$  qui s'annulent pour  $Q_i \in \Gamma^-$  :

$$Y_h^0 = \{w_h \in Y_h : w_h(Q_j) = 0, \forall j \in J^-\} \quad (2.56)$$

L'espace des fonctions tests pour la forme bilinéaire du problème approché d'advection sera  $Y_h^0$ .



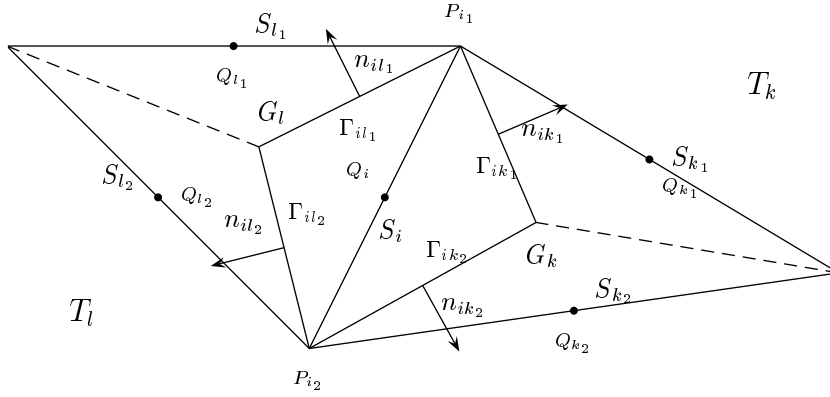


FIGURE 1 – Maillage dual barycentrique

### 2.3 Volumes finis barycentriques, voisinage d'un point $Q_i$ , points en amont/aval de $Q_i$ :

#### Volumes finis barycentriques

A partir du maillage  $(\mathcal{T}_h)$  se construit un système  $(\mathcal{D}_h) = \{D_i : i \in J\}$  de volumes finis s'appuyant sur les arêtes et les centres de gravité (maillage dual barycentrique).

A chaque point  $Q_i$  avec  $i \in J$  on associe un volume barycentrique  $D_i$  de la manière suivante :

- a) Si  $i \in J_0$ , c'est-à-dire  $Q_i \in S_i := [P_{i_1}, P_{i_2}]$  est un point intérieur de  $\Omega$ , alors il existe deux triangles  $T_k$  et  $T_l$  de côtés respectivement  $S_{k_1}, S_{k_2}, S_i$  et  $S_{l_1}, S_{l_2}, S_i$ , tels que  $P_1 \in S_{k_1} \cap S_{l_1}$  et  $P_2 \in S_{k_2} \cap S_{l_2}$ . Soient  $G_k, G_l$  les centres de gravité de  $T_k, T_l$ , alors  $D_i$  est le quadrilatère (fermé) incluant  $S_i$  et défini par les segments  $\Gamma_{k_1,i} := [G_k, P_1]$ ,  $\Gamma_{k_2,i} := [G_k, P_2]$ ,  $\Gamma_{l_1,i} := [G_l, P_1]$  et  $\Gamma_{l_2,i} := [G_l, P_2]$ .
- b) Si  $Q_i \in S_i \subset \partial\Omega$ , c'est-à-dire  $Q_i \in S_i := [P_1, P_2]$  est un point du bord  $\Gamma$ , alors il existe un seul triangle  $T_k$  ayant les côtés  $S_{k_1}, S_{k_2}, S_i$ . Le volume  $D_i$  est alors le triangle (fermé) défini par les segments  $S_i$ ,  $\Gamma_{k_1,i} := [G_k, P_1]$  et  $\Gamma_{k_2,i} := [G_k, P_2]$ .

$$\bar{\Omega} := \bigcup_{i \in J} D_i \quad (2.57)$$

On note  $|T|$  et  $|D_i|$  les mesures (aires) respectives de  $T \in \mathcal{T}_h$  et  $D_i \in \mathcal{D}_h$ , ainsi que  $|\Gamma_{i,k}|$  et  $|S_i|$  les mesures (longueurs) respectives de  $\Gamma_{i,k}$  et  $S_i$ .

On rappelle que  $|D_i \cap T| = \frac{1}{3}|T|$  pour  $Q_i \in T$ .

**Voisinage de  $Q_i$ , points en amont/aval de  $Q_i$** 

On commence avec la situation la plus simple :

**Cas  $Q_i \in \Omega$  :** C'est-à-dire que  $i \in J_0$ . Les points  $Q_{k_1}, Q_{k_2}, Q_{l_1}, Q_{l_2}$  sont les voisins de  $Q_i$  et on définit  $S(i) = \{k_1, k_2, l_1, l_2\}$  pour  $Q_i \in \Omega$ .

On a ainsi

$$\partial D_i := \bigcup_{j \in S(i)} \Gamma_{i,j} \quad (2.58)$$

On note  $n_{i,k_1}, n_{i,k_2}, n_{i,l_1}$  et  $n_{i,l_2}$  les restrictions de  $n_{D_i}$  (vecteur normal orienté vers l'extérieur de  $D_i$ ) à respectivement  $\Gamma_{i,k_1}, \Gamma_{i,k_2}, \Gamma_{i,l_1}$  et  $\Gamma_{i,l_2}$ .

Dans ces notations :  $\Gamma_{i,k} = \Gamma_{k,i}$  et  $n_{i,k} = -n_{k,i}$

Pour  $i \in J_0$  et  $k \in S(i)$  on définit

$$\theta_{i,k} = \int_{\Gamma_{i,k}} \beta \cdot n_{i,k} d\sigma$$

Le voisinage de  $Q_i$  en amont (upwind) de  $\beta$  est défini par

$$S^-(i) = \{k \in S(i) : \theta_{i,k} \leq 0\}$$

Le voisinage de  $Q_i$  en aval de  $\beta$  est défini par

$$S^+(i) = \{k \in S(i) : \theta_{i,k} \geq 0\}$$

Bien sur  $k \in S^+(i) \cap S^-(i) \iff \theta_{i,k} = 0$

De façon générale, on dira que  $Q_k$  est en amont (en aval de  $Q_i$ ) si  $\theta_{i,k} \leq 0$  ( $\theta_{i,k} \geq 0$ ).

Avec cette définition on a ( $S^+(i) \cup S^-(i) = S(i)$ ) :

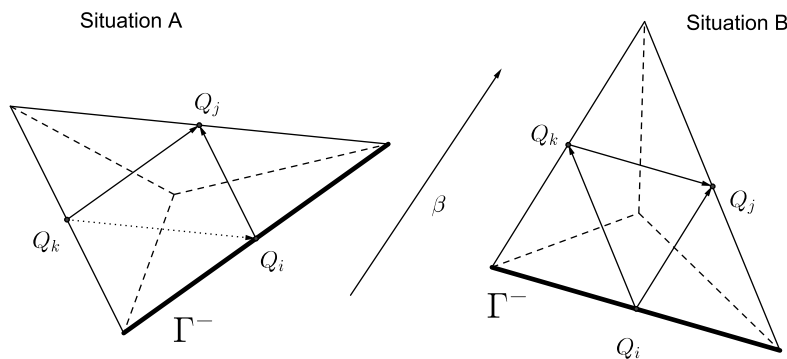
$$\sum_{j \in S^-(i)} \theta_{i,j} + \sum_{j \in S^+(i)} \theta_{i,j} = \int_{\partial D_i} \beta \cdot n d\sigma = \int_{D_i} \nabla \cdot \beta dx = 0 \quad \forall i \in J \quad (2.59)$$

**Cas  $Q_i \in \partial\Omega$  :** Ce cas concerne surtout le problème d'advection stationnaire.

**Si  $Q_i \in J^-$**  Le point  $Q_i$  est sur le bord entrant et ne doit pas avoir de voisin en amont.

Ainsi, avoir  $S^-(i) = \{-\infty, j\}$  n'est pas acceptable. On définit :

**Situation A :**  $S^-(i) = \emptyset$  et  $S^+(i) = \{j\}$ , **Situation B :**  $S^-(i) = \emptyset$  et  $S^+(i) = \{j, k\}$



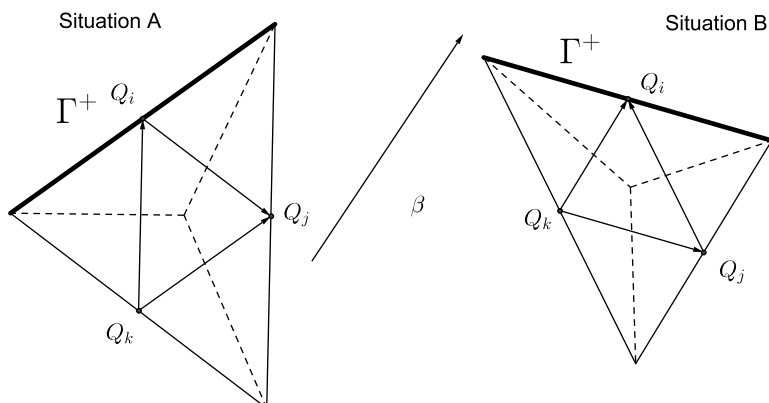
Si  $Q_i \in J^+$  Le point  $Q_i$  est sur le bord sortant et ne doit pas avoir de voisin en aval dans  $\Omega$ . Pour des raisons de notation, on introduit un indice virtuel  $+\infty$  pour lequel on définit

$$\Gamma_{i,+\infty} = S_i \text{ et } \theta_{i,+\infty} = \int_{\Gamma_{i,+\infty}} (\beta \cdot n_{D_i}) d\sigma \quad (2.60)$$

avec

Situation A :  $S^-(i) = \{k\}$  et  $S^+(i) = \{j, +\infty\}$  et on a  $\theta_{i,+\infty} + \theta_{i,j} = \theta_{k,i}$

Situation B :  $S^-(i) = \{j, k\}$  et  $S^+(i) = \{+\infty\}$  et on a  $\theta_{i,+\infty} = \theta_{j,i} + \theta_{k,i}$



Si  $Q_i \in J^b$  Le point  $Q_i$  est sur le bord longeant et il a exactement un point  $Q_k$  en amont  $\{k\} = S^-(i)$  et un point  $Q_j$  en aval  $\{j\} = S^+(i)$ . La normale du segment  $S_i$  étant orthogonale à  $\beta$  on a  $\int_{S_i} (\beta \cdot n_{D_i}) d\sigma = 0$ , donc  $\theta_{k,i} = \theta_{i,j}$ .

Il est parfois nécessaire de rappeler explicitement la dépendance à  $\beta$  de certaines quantités. On écrira ainsi  $S_{\beta}^-(i)$ ,  $S_{\beta}^+(i)$ ,  $\theta_{i,k}^{\beta}$  etc...

## 2.4 Formulation du terme de convection $b_h$

Le terme de convection  $b$  défini en (1.7) sera approché par une méthode upwind de volumes finis :

Soit  $u \in H^1(\Omega) \cap C(\overline{\Omega})$  et  $v_h \in X_h$ , alors

$$\begin{aligned} b(u, v_h) &= \int_{\Omega} \beta \cdot \nabla u v_h dx \\ &= \sum_{i \in J} \int_{D_i} \beta \cdot \nabla u v_h dx \\ &\approx \sum_{i \in J} \int_{D_i} \beta \cdot \nabla u v_h(Q_i) dx \\ &\approx \sum_{i \in J} v_h(Q_i) \int_{D_i} \nabla \cdot (\beta u) dx \\ &\approx \sum_{i \in J} v_h(Q_i) \int_{\partial D_i} u \beta \cdot n d\sigma \end{aligned}$$

$$\begin{aligned}
& n \text{ est le vecteur normal extérieur à } \partial D_i \\
& \approx \sum_{i \in J} v_h(Q_i) \sum_{j \in S(i)} \int_{\Gamma_{i,j}} u \beta \cdot n_{i,j} d\sigma
\end{aligned} \tag{2.61}$$

Pour chaque  $j \in S(i)$  les choix  $u(Q_j)$  ou  $u(Q_i)$  sont possibles pour une approximation de  $u$  sur  $\Gamma_{ij}$ . La méthode upwind consiste à choisir le point en "amont" de  $\beta$ , c'est-à-dire  $u(Q_i)$  si  $\theta_{i,j} \geq 0$ , ( $Q_i$  est upwind à  $Q_j$ ), sinon de prendre  $u(Q_j)$  dans le cas  $\theta_{i,j} \leq 0$ . Ainsi

$$b(u, v_h) \approx \sum_{i \in J} v_h(Q_i) \left[ \sum_{j \in S^+(i)} \theta_{i,j} u(Q_i) + \sum_{j \in S^-(i)} \theta_{i,j} u(Q_j) \right] \tag{2.62}$$

Et on pose

$$b_h(u_h, v_h) = \sum_{i \in J} v_h(Q_i) \left[ \sum_{j \in S^+(i)} \theta_{i,j} u_h(Q_i) + \sum_{j \in S^-(i)} \theta_{i,j} u_h(Q_j) \right], \forall u_h, v_h \in X_h \tag{2.63}$$

## 2.5 Opérateur d'interpolation $I_h$

**Interpolation dans  $H^1(\Omega)$  pour le problème d'advection-diffusion :**

Comme  $X_h \not\subset H^1(\Omega)$  et  $V_h \not\subset H_0^1(\Omega)$ , l'approximation est non conforme. Pour l'étude de stabilité et d'estimation d'erreur, il est nécessaire de pouvoir étendre la forme bilinéaire (2.63) à une forme bilinéaire continue sur  $(X_h + H^1(\Omega)) \times X_h$ .

Cela sera fait via l'opérateur d'interpolation  $I_h$  :

**Définition 2.3 (opérateur d'interpolation  $I_h$ )** Si

$v : H^1(\Omega) + X_h = \{v + v_h : v \in H^1(\Omega), v_h \in X_h\} \rightarrow \mathbb{R}$ , alors on définit

$$I_h v \in X_h, \text{ par } (I_h v)(Q_i) = \frac{1}{|S_i|} \int_{S_i} v d\sigma, \quad i \in J \tag{2.64}$$

Cette intégrale existe (théorème de trace dans  $H^1(T)$  et  $L^2(S_i) \subset L^1(S_i)$ ). On pose

$$\tilde{b}_h(u, v_h) = b_h(I_h u, v_h), \quad \forall u \in H^1(\Omega) + X_h, v_h \in V_h \tag{2.65}$$

**Lemme 2.3** L'opérateur  $I_h$  défini par (2.3) a les propriétés suivantes :

$$\text{Si } \varphi \in V \text{ alors } I_h \varphi \in V_h \tag{2.66}$$

Soit  $\varphi \in H^{k+1}(\Omega)$ , avec  $k = 0$  ou  $1$ . Alors pour  $h \in ]0, h_0[$  on a

$$\|\varphi - I_h\varphi\|_{X_h} \leq c_{I1} h^k \|\varphi\|_{H^{k+1}(\Omega)} \quad (2.67)$$

$$\|\varphi - I_h\varphi\|_{L^2(\Omega)} \leq c_{I2} h^{k+1} \|\varphi\|_{H^{k+1}(\Omega)}, \quad (2.68)$$

$$\|I_h\varphi\|_{X_h} \leq c_{I3} \|\varphi\|_{H^1(\Omega)}, \quad (2.69)$$

$$\varphi \in H^1(\Omega) \Rightarrow \|\varphi - I_h\varphi\|_{X_h} \rightarrow 0 \text{ pour } h \rightarrow 0 \quad (2.70)$$

avec  $c_{I1} > 0$ ,  $c_{I2} > 0$  et  $c_{I3} > 0$  indépendants de  $\varphi$  et  $h$ .

**Preuve :** Voir [Fei93], Lemme 8.9.81 □

### Interpolation dans $W^{1,p}(\Omega)$ pour le problème d'advection :

Comme  $Y_h \not\subset Y_{\text{sol}} \subset W^{1,1}(\Omega) \cap Y_1$ , l'approximation est non conforme.

On étend ainsi la forme bilinéaire (2.63) à une forme bilinéaire continue sur  $(Y_h + Y_{\text{sol}}) \times Y_h^0$ .

Cela sera fait<sup>13</sup> via l'opérateur d'interpolation  $I_h$  :

#### Définition 2.4 (opérateur d'interpolation $I_h$ )

*Si  $v : Y_h + W^{1,p}(\Omega) = \{v + v_h : v \in W^{1,p}(\Omega), v_h \in Y_h\} \rightarrow \mathbb{R}$ , alors on définit*

$$I_h v \in Y_h, \text{ par } (I_h v)(Q_i) = \frac{1}{|S_i|} \int_{S_i} v \, d\sigma, \quad i \in J \quad (2.71)$$

Cette intégrale existe (voir la remarque 1.7). On pose

$$\check{b}_h(u, v_h) = b_h(I_h u, v_h), \quad \forall u \in Y_{\text{sol}} + Y_h, v_h \in Y_h^0 \quad (2.72)$$

Soit  $(\mathcal{T}_h)_h$  une famille de maillage et  $T \in \mathcal{T}_h$ , alors

**Lemme 2.4** Pour  $w \in W^{1,1}(\Omega)$  et  $w_h = I_h(w)$  on a

a) Pour  $T \in \mathcal{T}_h$   $n_{T|S_i} = n_i$  et  $w_i = w_h(Q_i)$ ,  $i \in \{1, 2, 3\}$  (voir figure ci-dessous) :

$$\int_T w \, dx + \frac{1}{2} \int_T [(x - P_i) \cdot \nabla w] \, dx = w_i |T|$$

Si  $\tilde{T}_i$  est le triangle défini par  $G$  et les extrémités du segment  $S_i$ , alors

$$\int_{\tilde{T}_i} w \, dx + \frac{1}{2} \int_{\tilde{T}_i} [(x - G) \cdot \nabla w] \, dx = w_i |\tilde{T}_i|$$

b) Pour  $T \in \mathcal{T}_h$

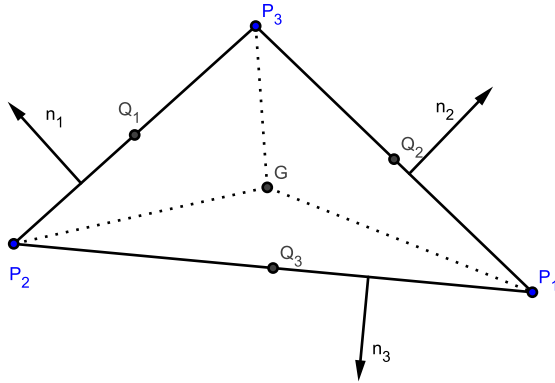
$$\left| \int_T w \, dx - \int_T I_h(w) \, dx \right| \leq h \int_T \|\nabla w\| \, dx$$

---

13. La régularité des données  $f \in W^{1,1}(\Omega)$ ,  $g \in W^{1,1}(\Gamma^-)$  permet d'obtenir une solution  $u \in W^{1,1} \subset Y_1$  du problème exact. Nous intéressent les estimations dans les normes  $\|\cdot\|_{Y_1}$  et  $\|\cdot\|_{Y_{1,h}}$  les solutions exactes et approchées.

c)

$$\|I_h(w)\|_{L_h^1(\Omega)} \leq \|w\|_{L^1(\Omega)} + h \underbrace{\int_{\Omega} \|\nabla w\| dx}_{=\|\nabla w\|_{L^1(\Omega)}}$$



**Preuve :** Pour  $T \in \mathcal{T}_h$  (voir la figure ci-dessus) et  $i \in \{1, 2, 3\}$  :

$$\begin{aligned} \frac{1}{2} \int_T \nabla \cdot [(x - P_i)w] dx &= \frac{1}{2} \int_{\partial T} w(x - P_i) \cdot n_T d\sigma \\ \int_T w dx + \frac{1}{2} \int_T [(x - P_i) \cdot \nabla w] dx &= \frac{1}{2} \int_{S_i} w(x - P_i) \cdot n_i d\sigma \end{aligned}$$

Maintenant pour  $x \in S_i$  :  $(x - P_i) \cdot n_i = \underbrace{(x - Q_i) \cdot n_i}_{=0} + \underbrace{(Q_i - P_i) \cdot n_i}_{=\text{hauteur issue de } P_i} = \frac{|T|}{2|S_i|}$ , ce qui prouve la première égalité de a).

Le même raisonnement permet de montrer la seconde égalité (remplacer  $P_i$  par  $G$ ).

Avec  $\int_T I_h(w) dx = \frac{|T|}{3}(w_1 + w_2 + w_3)$  on déduit b).

Avec a) :

$$|w_i| \frac{|T|}{3} \leq \frac{1}{3} \int_T |w| dx + \frac{1}{2} h \frac{1}{3} \int_T \|\nabla w\| dx$$

Sommation sur  $i$ , puis sur les triangles  $T$  livre c). □

**Lemme 2.5** Pour tous  $p, q \in [1; \infty]$  et  $\varphi \in W^{1,p}(T)$  on a

$$\|\partial_j I_h \varphi\|_{L^q(T)} \leq |T|^{1/q-1/p} \|\partial_j \varphi\|_{L^p(T)} \text{ pour } j = 1, 2 \quad (2.73)$$

**Preuve :** [ANS01], lemme 2 page 200. □

**Lemme 2.6** Pour tous  $p, q \in [1; \infty]$ ,  $p \geq q$  et  $\varphi \in W^{1,p}(T)$ , il existe une constante  $c_{I4}$  ne dépendant pas de  $h$ , telle que

$$\|\partial_j(\varphi - I_h \varphi)\|_{L^q(T)} \leq c_{I4} |T|^{1/q-1/p} \|\partial_j \varphi\|_{L^p(T)} \text{ pour } j = 1, 2 \quad (2.74)$$

**Preuve :** [ANS01], corollaire 1 page 200.  $\square$

**Remarque 2.5** *Les deux lemmes ci-dessus sont valides sans condition d'angle au maillage.*

**Lemme 2.7** *Pour tous  $p, q \in [1; \infty]$ ,  $p \geq q$  et  $\varphi \in W^{2,p}(T)$ , il existe une constante  $c_{I5}$  ne dépendant pas de  $h$ , telle que*

$$\|\partial_j(\varphi - I_h\varphi)\|_{L^q(T)} \leq h c_{I5} |T|^{1/q-1/p} \sum_{i=1}^2 \|\partial_i \partial_j \varphi\|_{L^p(T)} \text{ pour } j = 1, 2 \quad (2.75)$$

**Preuve :** On applique le lemme 3 page 200 dans [ANS01].  $\square$

On suppose que  $\Gamma^-$  est connexe, et peut donc s'écrire comme la réunion de  $k$  segments  $\Gamma^- = ]A_0; A_1] \cup ]A_1; A_2] \cup \dots \cup ]A_i; A_{i+1}] \cup \dots \cup ]A_{k-1}; A_k[$ .

Sur le segment  $]A_i; A_{i+1}[$  on note  $n_i$  la normale extérieure à  $\Gamma^-$  et  $n_i^T = \frac{\vec{A_i A_{i+1}}}{\|A_i A_{i+1}\|}$  un vecteur orthogonal à  $n_i$ .

On suppose que  $\Gamma^- = \cup_{l=1}^{|J^-|} S_{i_l}$  avec  $S_{i_l} \cap S_{i_{l+1}} \neq \emptyset$  et  $i_l \in J^-$  pour  $l = 1, \dots, |J^-|$ .

L'espace des conditions au bord au bord  $Y_h(\Gamma^-)$  pour le problème approché est défini de la manière suivante :

$$Y_h(\Gamma^-) = \{ \varphi_h : \text{il existe } w_h \in Y_h \text{ avec } w_h|_{\Gamma^-} = \varphi_h \} \quad (2.76)$$

On définit la semi-norme  $|\cdot|_{Var_h(\Gamma^-)}$  et la norme  $\|\cdot\|_{L_h^p(\Gamma^-)}$  pour  $\varphi_h \in Var_h(\Gamma^-)$  par

$$|\varphi_h|_{Var_h(\Gamma^-)} = \sum_{l=1}^{|J^-|-1} |\varphi_h(Q_{i_{l+1}}) - \varphi_h(Q_{i_l})| \text{ et} \quad (2.77)$$

$$\|\varphi_h\|_{L_h^p(\Gamma^-)} = \left[ \sum_{l=1}^{|J^-|} |\varphi_h(Q_{i_l})|^p |S_{i_l}| \right]^{1/p} \quad (2.78)$$

Pour la condition au bord  $g$  on aura besoin du lemme suivant

**Lemme 2.8** *Pour  $g \in W^{1,1}(\Gamma^-)$ , on a*

$$|I_h g|_{Var_h(\Gamma^-)} \leq 2 \|g'\|_{L^1(\Gamma^-)} \quad (2.79)$$

**Preuve :** Soit  $l \in \{1, \dots, |J^-| - 1\}$ , et  $A = S_{i_l} \cap S_{i_{l+1}}$ , alors pour  $y_1 \in S_{i_l}$  et  $y_2 \in S_{i_{l+1}}$  on a

$$\begin{aligned} - \left[ \int_{]y_1; A[} |g'(x)| d\sigma + \int_{]A; y_2[} |g'(x)| d\sigma \right] &\leq g(y_1) - g(y_2) \leq \int_{]y_1; A[} |g'(x)| d\sigma + \int_{]A; y_2[} |g'(x)| d\sigma \\ &- \int_{S_{i_l} \cup S_{i_{l+1}}} |g'(x)| d\sigma \leq g(y_1) - g(y_2) \leq \int_{S_{i_l} \cup S_{i_{l+1}}} |g'(x)| d\sigma \end{aligned}$$

En intégrant maintenant  $g(y_1)$  sur  $S_i$ , puis  $g(y_2)$  sur  $S_{i+1}$ , on obtient

$$-|S_i| |S_{i+1}| \int_{S_i \cup S_{i+1}} |g'(x)| d\sigma \leq |S_{i+1}| \int_{S_i} g d\sigma - |S_i| \int_{S_{i+1}} g d\sigma \leq |S_i| |S_{i+1}| \int_{S_i \cup S_{i+1}} |g'(x)| d\sigma$$

En divisant cette inégalité par  $|S_i| |S_{i+1}|$  et en sommant sur  $l$  on obtient l'inéquation désirée.  $\square$

## 2.6 Coordonnées locales d'un triangle $T \in \mathcal{T}_h$ :

Pour un triangle  $T \in \mathcal{T}_h$  on note  $S_1^T$ ,  $S_2^T$  et  $S_3^T$  les côtés du triangle  $T \in \mathcal{T}_h$ .

Le point  $Q_r^T$  est le milieu de  $S_r^T$  pour  $r \in \{1, 2, 3\}$ . Ainsi

$$\partial T = \bigcup_{r=1,2,3} S_r^T$$

**Lemme 2.9** Pour  $v_h \in X_h$ ,  $T \in \mathcal{T}_h$ , on a  $\rho_h(v_h) = v_h$  et (en coordonnées locales)

$$\frac{1}{\hat{c}_2} \left( \sum_{r=1}^3 v_h(Q_r^T)^2 h_T^2 \right)^{1/2} \leq \|v_h\|_{L^2(T)} \leq \hat{c}_2 \left( \sum_{r=1}^3 v_h(Q_r^T)^2 h_T^2 \right)^{1/2} \quad (2.80)$$

avec  $\hat{c}_2 > 0$  indépendant de  $h$ ,  $v_h$  et  $T$ .

**Preuve :** Appliquer lemme 2.10 et (2.40).  $\square$

**Lemme 2.10** Le produit scalaire  $(\cdot, \cdot)$  peut s'exprimer par une formule de quadrature utilisant les milieux  $Q_i^T$  des côtés des triangles  $T \in \mathcal{T}_h$  :

$$(u_h, v_h) = \frac{1}{3} \sum_{T \in \mathcal{T}_h} |T| \sum_{j=1}^3 u_h(Q_j^T) v_h(Q_j^T) = (L_h u_h, L_h v_h) \quad u_h, v_h \in X_h \quad (2.81)$$

$$\|v_h\|_{L^2(\Omega)} = \|L_h v_h\|_{L^2(\Omega)} \quad v_h \in X_h \quad (2.82)$$

## 2.7 Opérateur gradient local $\nabla_h$

Pour  $v_h \in X_h$ ,  $\nabla v_h$  n'est pas défini, contrairement à la restriction  $v_h|_T$  pour  $T \in \mathcal{T}_h$ .

**Définition 2.6** Pour  $v \in X_h + H^1(\Omega)$  on définit l'opérateur gradient local  $\nabla_h : X_h + H^1(\Omega) \rightarrow L^2(\Omega)$  par

$$(\nabla_h v)|_T = \nabla v|_T, \quad \forall T \in \mathcal{T}_h \quad (2.83)$$

**Lemme 2.11** 1. pour  $u, v \in H^1(\Omega)$  :

$$((u, v)) = ((u, v))_h = (\nabla_h u, \nabla_h v)_{L^2(\Omega)}, \quad (2.84)$$



2. Pour  $u_h, v_h \in X_h$  :

$$((I_h u_h, I_h v_h))_h = ((u_h, v_h))_h \text{ et } |((u_h, v_h))_h| \leq \|u_h\|_{*V_h} \|v_h\|_{*V_h} \quad (2.85)$$

3. Soit  $i \in J$ ,  $w_{h,i} \in X_h$  défini par  $w_{h,i}(Q_k) = \delta_{ik}$  et  $T \in \mathcal{T}_h$  un triangle ayant  $S_i$  comme côté, alors

$$|\nabla w_{h,i}|_T |T| = |S_i| \quad (2.86)$$

**Preuve :** Il suffit de montrer (2.86) :

Soit  $Q_l$  et  $Q_k$  les milieux des deux autres côtés de  $T$  et soit  $Q \in (Q_l Q_k)$  tel que  $(Q - Q_i) \cdot (Q_l - Q_k) = 0$ , alors

$$\begin{aligned} |\nabla w_{h,i}|_T |T| &= |\nabla w_{h,i}|_T \underbrace{|Q - Q_i|}_{|T|} |S_i| \\ &= |\nabla w_{h,i}|_T \cdot (Q - Q_i) |S_i| \quad \text{car } \nabla w_{h,i}|_T \perp S_i \\ &= |w_{h,i}(Q) - w_{h,i}(Q_i)| |S_i| \\ &= |w_{h,i}(Q)| |S_i| \quad \text{car } w_{h,i}(Q_i) = 1, \text{ donc } w_{h,i}(Q) = 0 \\ &= |S_i| \end{aligned}$$

□

La version non conforme du théorème de Poincaré (théorème 1.2) est :

**Lemme 2.12 (Inégalité de Friedrichs)** *Il existe une constante  $\hat{c}_*$  indépendante de  $h$  telle que*

$$\|u_h\|_{L^2(\Omega)} \leq \hat{c}_* \|u_h\|_{*V_h} (= ((u_h, u_h))_h^{1/2}), \quad u_h \in V_h, \quad h \in (0, h_0). \quad (2.87)$$

**Preuve :** Voir [sFF99]

□

## 2.8 Opérateur local de projection $\rho_h$

**Définition 2.7** (Opérateur local de projection  $\rho_h$ )

Soit  $v \in L^2(\Omega)$  tel que  $v|_T \in C^0(T)$  pour  $T \in \mathcal{T}_h$  et  $v$  est continu en  $Q_i$  pour  $i \in J$ , alors on définit  $\rho_h(v) \in X_h$  par

$$\rho_h(v)(Q_i) = v(Q_i) \text{ pour } i \in J \quad (2.88)$$

**Lemme 2.13** Soit  $u_h \in X_h$ ,  $\varphi \in C^1(\overline{\Omega})$ , alors

$$\|\nabla_h \rho_h(u_h)\|_2 \leq \hat{c}_3 (\|\nabla_h u_h\|_2 \|\varphi\|_\infty + \|u_h\|_2 \|\nabla \varphi\|_\infty) \quad (2.89)$$

avec  $\hat{c}_3$  indépendant de  $h$ ,  $u_h$  et  $\varphi$

**Preuve :** Soit  $T \in \mathcal{T}_h$ . Soit  $M_T \in T$  un point quelconque mais fixe. Soient  $i_1, i_2, i_3 \in J$  tels que  $Q_{i_1}, Q_{i_2}, Q_{i_3}$  sont les milieux des côtés de  $T$ . On a  $\rho(u_h\varphi)|_T = \sum_{\nu=1}^3 (u_h\varphi)(Q_{i_\nu}) w_{h,i_\nu}$ . Ici  $w_{h,i_\nu} \in V_h$  est un élément de la base de  $V_h$  (voir page 31). On a alors

$$\begin{aligned}
\|\nabla(u_h\varphi)|_T\|_{L^2(T)}^2 &= \int_T \left\| \sum_{\nu=1}^3 (u_h\varphi)(Q_{i_\nu}) \nabla w_{h,i_\nu} \right\|^2 dx \\
&\leq 2 \int_T \left\| \sum_{\nu=1}^3 u_h(Q_{i_\nu}) (\varphi(Q_{i_\nu}) - \varphi(Q_{M_T})) \nabla w_{h,i_\nu} \right\|^2 dx \\
&\quad + 2\varphi(Q_{M_T})^2 \int_T \left\| \sum_{\nu=1}^3 u_h(Q_{i_\nu}) \nabla w_{h,i_\nu} \right\|^2 dx \\
&\leq 2 \int_T \left\| \sum_{\nu=1}^3 u_h(Q_{i_\nu}) \left[ \int_0^1 \nabla\varphi(M_T + \theta(Q_{i_\nu} - M_T)) \cdot (Q_{i_\nu} - M_T) d\theta \right] \nabla w_{h,i_\nu} \right\|^2 dx \\
&\quad + 2\varphi(Q_{M_T})^2 \|\nabla u_h|_T\|_{L^2(T)}^2 \\
&\leq 6 \int_T \sum_{\nu=1}^3 \|u_h(Q_{i_\nu}) \left[ \int_0^1 \nabla\varphi(M_T + \theta(Q_{i_\nu} - M_T)) \cdot (Q_{i_\nu} - M_T) d\theta \right] \nabla w_{h,i_\nu}\|^2 dx \\
&\quad + 2\varphi(Q_{M_T})^2 \|\nabla u_h|_T\|_{L^2(T)}^2 \\
&\leq 6 \int_T \sum_{\nu=1}^3 \|u_h(Q_{i_\nu}) \nabla w_{h,i_\nu}\|^2 dx \|\nabla\varphi\|_{L^\infty(\Omega)}^2 h_T^2 + 2\varphi(Q_{M_T})^2 \|\nabla u_h|_T\|_{L^2(T)}^2 \\
&\leq 6 \sum_{\nu=1}^3 \|u_h(Q_{i_\nu}) \nabla w_{h,i_\nu}\|^2 |T| \|\nabla\varphi\|_{L^\infty(\Omega)}^2 h_T^2 + 2\|\varphi\|_{L^\infty(\Omega)}^2 \|\nabla u_h|_T\|_{L^2(T)}^2 \\
(2.86) \text{ et } (2.40) &\leq 6\hat{c}_0 \sum_{\nu=1}^3 u_h(Q_{i_\nu})^2 |S_i|^2 \|\nabla\varphi\|_\infty^2 + 2\|\varphi\|_\infty^2 \|\nabla u_h|_T\|_{L^2(T)}^2 \\
&\stackrel{(2.80)}{\leq} 6\hat{c}_0\hat{c}_2 \|u_h|_T\|_{L^2(T)}^2 \|\nabla\varphi\|_\infty^2 + 2\|\varphi\|_\infty^2 \|\nabla u_h|_T\|_{L^2(T)}^2
\end{aligned}$$

On conclut la preuve du lemme par sommation sur  $T \in \mathcal{T}_h$ .

□

**Lemme 2.14** Soit  $u_h \in X_h$ ,  $\varphi \in C^0(\bar{\Omega})$ , alors

$$\|\rho_h(u_h\varphi)\|_2 \leq \|u_h\|_2 \|\varphi\|_\infty \quad (2.90)$$

**Preuve :**  $\rho_h(u_h\varphi) \in X_h$ , donc avec le lemme 2.10 on obtient

$$\begin{aligned}
\|\rho_h(u_h\varphi)\|_2^2 &= (\rho_h(u_h\varphi), \rho_h(u_h\varphi)) \\
&= \frac{1}{3} \sum_{T \in \mathcal{T}_h} |T| \sum_{j=1}^3 u_h(Q_T^j)^2 \varphi(Q_T^j)^2 \\
&\leq \|\varphi\|_\infty^2 \frac{1}{3} \sum_{T \in \mathcal{T}_h} |T| \sum_{j=1}^3 u_h(Q_T^j)^2 \\
&\leq \|\varphi\|_\infty^2 (u_h, u_h) \text{ comme } \rho_h(u_h) = u_h \in X_h.
\end{aligned}$$

□

**Lemme 2.15 (formule de divergence dans  $V_h$ )** Soit  $u_h, v_h \in V_h$ , alors

$$\left| \int_{\Omega} \beta \cdot \nabla_h u_h v_h dx + \int_{\Omega} u_h \beta \cdot \nabla_h v_h dx \right| \leq 3 \|\beta\| \hat{c}_0 h \|u_h\|_{*V_h} \|v_h\|_{*V_h} \quad (2.91)$$

**Preuve :**

$$\begin{aligned} \int_{\Omega} \beta \cdot \nabla_h u_h v_h dx &= \sum_{T \in \mathcal{T}_h} \int_T \beta \cdot \nabla_h u_h v_h dx \\ &= - \sum_{T \in \mathcal{T}_h} \int_T u_h \beta \cdot \nabla_h v_h dx + \sum_{T \in \mathcal{T}_h} \int_{\partial T} u_h v_h \beta \cdot n_T d\sigma \\ &\quad \text{En coordonnées locales : } \partial T = \cup_{r=1}^3 S_T^r \\ \sum_{T \in \mathcal{T}_h} \int_{\partial T} u_h v_h \beta \cdot n_T d\sigma &= \sum_{T \in \mathcal{T}_h} \sum_{r=1}^3 \int_{S_T^r} u_h v_h \beta \cdot n_T^r d\sigma \end{aligned} \quad (2.92)$$

Si  $S_T^r \subset \partial\Omega$ , alors  $\int_{S_T^r} u_h v_h \beta \cdot n_T^r d\sigma = 0$ .

Si  $S_T^r \cap \partial\Omega = \emptyset$ , alors il existe un autre triangle  $T^*$  et  $r^* \in \{1, 2, 3\}$  tel que  $S_T^r = T \cap T^* = S_{T^*}^{r^*}$  et  $n_T^r = -n_{T^*}^{r^*}$  sur  $S_T^r = S_{T^*}^{r^*}$ . Soient  $T_{Q_i}$  et  $T_{Q_i}^*$  les deux triangles ayant  $S_i$  comme côté commun ( $Q_i$  milieu de  $S_i$ ), alors

$$\sum_{T \in \mathcal{T}_h} \sum_{r=1}^3 \int_{S_T^r} u_h v_h \beta \cdot n_T^r d\sigma = \sum_{i \in J_0} \left( \int_{S_i} u_h v_h \beta \cdot n_{T_{Q_i}} d\sigma + \int_{S_i} u_h v_h \beta \cdot n_{T_{Q_i}^*} d\sigma \right) \quad (2.93)$$

Comme  $u_h$  et  $v_h$  sont continus en  $Q_i^r$  et comme  $u_{h|T}(Q) = \nabla u_{h|T}(Q - Q_i) + u_h(Q_i)$ ,  $u_{h|T^*}(Q) = \nabla u_{h|T^*}(Q - Q_i) + u_h(Q_i)$  (avec les mêmes formules pour  $v_h$ ), on obtient

$$\begin{aligned} \sum_{T \in \mathcal{T}_h} \sum_{r=1}^3 \int_{S_T^r} u_h v_h \beta \cdot n_T^r d\sigma &\leq \sum_{i \in J_0} |\beta| |S_i| h_T^2 (\|\nabla u_{h|T}\| \|\nabla v_{h|T}\| + \|\nabla u_{h|T^*}\| \|\nabla v_{h|T^*}\|) \\ &\leq 3 \|\beta\| \hat{c}_0 h \sum_{T \in \mathcal{T}_h} |T| \|\nabla u_{h|T}\| \|\nabla v_{h|T}\| \end{aligned}$$

□

## 2.9 Coordonnées locales upwind d'un triangle $T \in \mathcal{T}_h$ :

Soit  $T \in \mathcal{T}_h$  avec  $Q_a, Q_b, Q_c$  comme milieux des côtés de  $T$ . Les coordonnées locales upwind  $Q_0^{\beta, T}$ ,  $Q_1^{\beta, T}$  et  $Q_2^{\beta, T}$ , sont une manière de renommer  $Q_a, Q_b, Q_c$  de façon à disposer de l'information permettant de savoir quel point est en amont de quel autre dans le triangle  $T$ .

Les coordonnées locales upwind  $Q_0^{\beta,T}$ ,  $Q_1^{\beta,T}$  et  $Q_2^{\beta,T}$ , à la différence des coordonnées locales présentées ci-dessus, permettent donc de savoir quel point est en amont de quel autre.

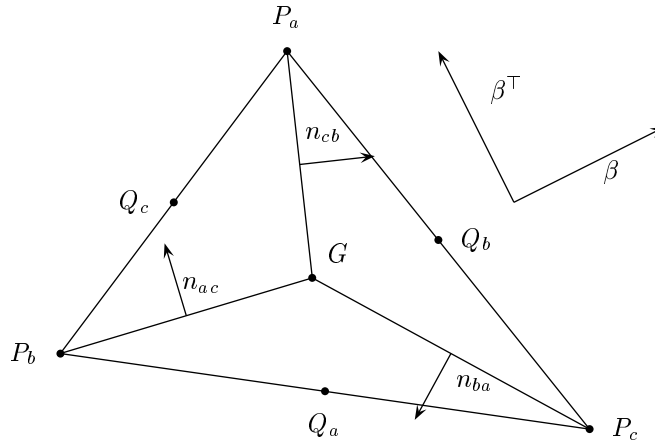
Ainsi  $Q_0^{\beta,T}$  est en amont des points  $Q_1^{\beta,T}$  et  $Q_2^{\beta,T}$ ,  $Q_1^{\beta,T}$  est en amont de  $Q_2^{\beta,T}$ . Le point  $Q_2^{\beta,T}$  est donc en aval des points  $Q_0^{\beta,T}$  et  $Q_1^{\beta,T}$ . Il faut donc s'assurer de l'absence de "boucle" dans  $T$  où les points  $Q_a, Q_b, Q_c$  seraient en même temps en aval et en amont des deux autres points.

L'existence de telles coordonnées doit être démontrée.

**Définition 2.8 (coordonnées locales upwind)** Soient  $S_a, S_b, S_c$  les côtés de  $T \in \mathcal{T}_h$ . Soit  $G$  le centre de gravité de  $T$  et  $P_a, P_b$  et  $P_c$  les sommets de  $T$  tels que  $G \in [Q_a, P_a] \cap [Q_b, P_b] \cap [Q_c, P_c]$ .

Il est alors toujours possible de renommer  $Q_a, Q_b, Q_c$  en  $Q_0^{\beta,T}, Q_1^{\beta,T}, Q_2^{\beta,T}$  (pour une certaine permutation de ces points) de telle manière, que pour  $\theta_{i,j}^{\beta,T} = \theta_{Q_i^{\beta,T}, Q_j^{\beta,T}}$  on ait  $\theta_{0,1}^{\beta,T} \geq 0$ ,  $\theta_{1,2}^{\beta,T} \geq 0$  et  $\theta_{0,2}^{\beta,T} \geq 0$ .

**Preuve :** Supposons que de telles coordonnées ne soient pas possibles, alors sans restriction on peut supposer que pour une certaine permutation des points  $Q_a, Q_b, Q_c$  on ait  $\theta_{0,1}^{\beta,T} \geq 0$ ,  $\theta_{1,2}^{\beta,T} \geq 0$  et  $\theta_{2,0}^{\beta,T} \geq 0$  avec au moins un de ces nombres strictement



positifs.

Supposons maintenant que les points  $Q_0^{\beta,T}, Q_1^{\beta,T}, Q_2^{\beta,T}$  soient arrangés dans le sens des aiguilles d'une montre autour de  $G$  (le cas où les points sont arrangés dans le sens contraire est analogue), alors  $\theta_{0,1}^{\beta,T} = (P_2^{\beta,T} - G)\beta^t$ ,  $\theta_{1,2}^{\beta,T} = (P_0^{\beta,T} - G)\beta^t$ ,  $\theta_{2,0}^{\beta,T} = (P_1^{\beta,T} - G)\beta^t$  où  $\beta^t$  est l'image de  $\beta$  par la rotation de  $\frac{\pi}{2}$  dans le sens contraire des aiguilles d'une montre.

$G$  est le centre de gravité du triangle  $T$ , donc

$$0 = \overrightarrow{GP_2^{\beta,T}} + \overrightarrow{GP_0^{\beta,T}} + \overrightarrow{GP_1^{\beta,T}}$$

On a alors la contradiction

$$0 = (\overrightarrow{GP_2^{\beta,T}} + \overrightarrow{GP_0^{\beta,T}} + \overrightarrow{GP_1^{\beta,T}})\beta^t = \theta_{0,1}^{\beta,T} + \theta_{1,2}^{\beta,T} + \theta_{2,0}^{\beta,T} > 0$$

□

## 2.10 Quelques propriétés de $b_h$ :

**Lemme 2.16** *On utilise les coordonnées locales upwind :  $u_{h,i}^T = u_h(Q_i^T)$ ,  $v_{h,i}^T = v_h(Q_i^T)$  pour  $i = 0, 1, 2$ .*

a) *Soit  $v_h \in V_h$ , alors*

$$|T|\beta \cdot \nabla v_h|_T = (\theta_{0,2}^T + \theta_{1,2}^T)(v_{h,2}^T - v_{h,1}^T) + (\theta_{0,1}^T + \theta_{0,2}^T)(v_{h,1}^T - v_{h,0}^T) \quad \forall T \in \mathcal{T}_h \quad (2.94)$$

b) *Soit  $u_h, v_h \in V_h$ , alors, exprimé dans des coordonnées locales upwind, on a*

$$b_h(u_h, v_h) = \sum_{T \in \mathcal{T}_h} [-u_{h,0}^T |T|\beta \cdot \nabla v_h|_T + (u_{h,0}^T - u_{h,1}^T) \theta_{1,2}^T (v_{h,1}^T - v_{h,2}^T)] \quad (2.95)$$

c) *Soit  $u_h, v_h \in V_h$ , alors*

$$|b_h(u_h, v_h) + (u_h, \beta \cdot \nabla v_h)| \leq \|\beta\| h(1 + \hat{c}_0) \|u_h\|_{*V_h} \|v_h\|_{*V_h} \quad (2.96)$$

d) *Soit  $u_h, v_h \in V_h$ , alors*

$$|b_h(u_h, v_h)| \leq \|\beta\| (\sqrt{3}\hat{c}_* + h_0\hat{c}_0) \|u_h\|_{*V_h} \|v_h\|_{*V_h}$$

**Preuve :**

a) Soit  $n_T$  un vecteur normal à  $\partial\Omega$  orienté vers l'extérieur et  $n_{i,j}^T$  un vecteur normal à  $\Gamma_{i,j}^T$  tel que  $\beta \cdot n_{i,j}^T \geq 0$ , alors

$$\begin{aligned} 0 &= \int_{S_0^T} \beta \cdot n_T d\sigma + \underbrace{\int_{\Gamma_{0,1}^T} \beta \cdot n_{0,1}^T d\sigma}_{=\theta_{0,1}^T} + \underbrace{\int_{\Gamma_{0,2}^T} \beta \cdot n_{0,2}^T d\sigma}_{=\theta_{0,2}^T} \\ 0 &= \int_{S_2^T} \beta \cdot n_T d\sigma - \underbrace{\int_{\Gamma_{1,2}^T} \beta \cdot n_{1,2}^T d\sigma}_{=\theta_{1,2}^T} - \underbrace{\int_{\Gamma_{0,2}^T} \beta \cdot n_{0,2}^T d\sigma}_{=\theta_{0,2}^T} \end{aligned}$$

et

$$0 = \int_{S_0^T} \beta \cdot n_T d\sigma + \int_{S_1^T} \beta \cdot n_T d\sigma + \int_{S_2^T} \beta \cdot n_T d\sigma$$

Ainsi (dans les coordonnées locales upwind) :  $\int_{S_0^T} \beta \cdot n_T d\sigma = -(\theta_{0,1}^T + \theta_{0,2}^T)$ ,  $\int_{S_2^T} \beta \cdot n_T d\sigma = (\theta_{1,2}^T + \theta_{0,2}^T)$  et  $\int_{S_1^T} \beta \cdot n_T d\sigma = (\theta_{0,1}^T + \theta_{0,2}^T) - (\theta_{1,2}^T + \theta_{0,2}^T)$

On a

$$|T|\beta \cdot \nabla v_h|_T = \int_T \beta \cdot \nabla v_h dx = \int_{\partial T} v_h \beta \cdot n d\sigma = \sum_{l=0}^2 \int_{S_l^T} v_h \beta \cdot n_T d\sigma.$$

Pour que cette dernière intégrale (sur le bord) soit définie, on considère qu'elle s'effectue sur l'extension continue de  $v_h|_T$  à  $\bar{T}$  (nommée  $v_h$  aussi). La continuité des éléments

de Crouzeix-Raviart entraîne que, pour le côté commun  $S$  de deux mailles  $T, \hat{T} \in \mathcal{T}_h$ , l'intégrale  $\int_S v_h \beta \cdot n \, d\sigma$  ne dépend pas du choix de la maille pour l'extension de  $v_h$ .

$$\begin{aligned} \int_{S_0^T} v_h \beta \cdot n_T \, d\sigma &= (\beta \cdot n_T) \int_{S_0^T} v_h \, d\sigma = (\beta \cdot n_T) |S_0^T| v_h(Q_0^T) = -v_h(Q_0^T)(\theta_{0,1}^T + \theta_{0,2}^T) \\ \int_{S_2^T} v_h \beta \cdot n_T \, d\sigma &= (\beta \cdot n_T) \int_{S_2^T} v_h \, d\sigma = (\beta \cdot n_T) |S_2^T| v_h(Q_2^T) = +v_h(Q_2^T)(\theta_{0,2}^T + \theta_{1,2}^T) \\ \int_{S_1^T} v_h \beta \cdot n_T \, d\sigma &= (\beta \cdot n_T) |S_1^T| v_h(Q_1^T) = v_h(Q_1^T)[(\theta_{0,1}^T + \theta_{0,2}^T) - (\theta_{0,1}^T + \theta_{0,2}^T)] \end{aligned}$$

La preuve se conclut en sommant les intégrales ci-dessus.

b) Avec (2.63) on obtient (somme sur  $J_0$  car  $u_h(Q_i) = 0$  et  $v_h(Q_i) = 0$  pour  $Q_i \in \Gamma$ ) :

$$\begin{aligned} b_h(u_h, v_h) &= \sum_{i \in J_0} \sum_{j \in S^+(i)} v_h(Q_i) \theta_{ij} u_h(Q_i) + \sum_{l \in J_0} \sum_{k \in S^-(l)} v_h(Q_l) \theta_{lk} u_h(Q_k) \\ &\quad k \in S^-(l) \iff l \in S^+(k) \\ &= \sum_{i \in J_0} \sum_{j \in S^+(i)} v_h(Q_i) \theta_{ij} u_h(Q_i) + \sum_{k \in J_0} \sum_{l \in S^+(k)} v_h(Q_l) \theta_{lk} u_h(Q_k) \\ &\stackrel{i:=k, j:=l}{=} \sum_{i \in J_0} \sum_{j \in S^+(i)} (v_h(Q_i) - v_h(Q_j)) \theta_{ij} u_h(Q_i) \\ &= \sum_{T \in \mathcal{T}_h} [(v_{h,0}^T - v_{h,1}^T) \theta_{0,1}^T u_{h,0}^T + (v_{h,1}^T - v_{h,2}^T) \theta_{1,2}^T u_{h,1}^T + (v_{h,0}^T - v_{h,2}^T) \theta_{0,2}^T u_{h,0}^T] \\ &= \sum_{T \in \mathcal{T}_h} \{ [(v_{h,0}^T - v_{h,1}^T) \theta_{0,1}^T + (v_{h,0}^T - v_{h,2}^T) \theta_{0,2}^T + (v_{h,1}^T - v_{h,2}^T) \theta_{1,2}^T] u_{h,0}^T + (v_{h,1}^T - v_{h,2}^T) \theta_{1,2}^T u_{h,1}^T \} \\ &= \sum_{T \in \mathcal{T}_h} [-u_{h,0}^T |T| \beta \cdot \nabla v_h|_T + (u_{h,0}^T - u_{h,1}^T) \theta_{1,2}^T (v_{h,1}^T - v_{h,2}^T)] \end{aligned}$$

c) Soit  $u_h, v_h \in V_h$  and  $T \in \mathcal{T}_h$ , alors (utilisation des coordonnées locales upwind)

$$\begin{aligned} &|b_h(u_h, v_h) + (u_h, \beta \cdot \nabla_h v_h)| \\ &= \left| \sum_{T \in \mathcal{T}_h} [-u_{h,0}^T |T| \beta \cdot \nabla v_h|_T + (u_{h,0}^T - u_{h,1}^T) \theta_{1,2}^T (v_{h,1}^T - v_{h,2}^T)] + \sum_{T \in \mathcal{T}_h} \int_T u_h \beta \cdot \nabla v_h|_T \, dx \right| \\ &\quad \text{Maintenant } u_h(Q) = u_{h,0}^T + \nabla u_h|_T \cdot (Q - Q_0^T) \text{ et } v_h(Q) = v_{h,1}^T + \nabla v_h|_T \cdot (Q - Q_1^T) \quad \forall Q \in T \\ &\leq \sum_{T \in \mathcal{T}_h} \int_T |\nabla u_h|_T h_T |\beta \cdot \nabla v_h|_T \, dx + \sum_{T \in \mathcal{T}_h} \|\nabla u_h|_T\| h_T h_T |\beta| \|\nabla v_h|_T\| h_T \\ &\quad h_T \leq h \text{ et } h_T^2 \leq h^2 \leq \hat{c}_0 |T| \text{ avec (2.40)} \\ &\leq \|\beta\| h (1 + \hat{c}_0) \sum_{T \in \mathcal{T}_h} |T| \|\nabla u_h|_T\| \|\nabla v_h|_T\| \end{aligned}$$

d) Soit  $u_h, v_h \in V_h$  and  $T \in \mathcal{T}_h$ , alors (utilisation des coordonnées locales upwind)

$$\begin{aligned} u_h(Q) &= u_{h,0}^T + \nabla u_{h|T}(Q - Q_O^T) \text{ et } v_h(Q) = v_{h,1}^T + \nabla v_{h|T}(Q - Q_1^T) \quad \forall Q \in T \\ |u_h(Q) - u_{h,0}^T| &\leq |\nabla u_{h|T}| h_T \text{ et } |v_{h,1}^T - v_{h,2}^T| \leq |\nabla v_{h|T}| h_T. \end{aligned}$$

Avec (2.95) et  $0 \leq \theta_{1,2} \leq \|\beta\| h_T$  on a :

$$|b_h(u_h, v_h)| \leq \sum_{T \in \mathcal{T}_h} \left[ |u_{h,0|T}| |T| \|\beta\| \|\nabla v_{h|T}\| + \|\nabla u_{h|T}\| h_T \|\beta\| h_T \|\nabla v_{h|T}\| h_T \right]$$

Maintenant  $h_T^2 \leq \hat{c}_0 |T|$  et  $h_T \leq h_0$ , donc

$$|b_h(u_h, v_h)| \leq \|\beta\| \sum_{T \in \mathcal{T}_h} |u_{h,0|T}| |T| \|\nabla v_{h|T}\| + h_0 \|\beta\| \hat{c}_0 \sum_{T \in \mathcal{T}_h} |\nabla u_{h|T}| |T| \|\nabla v_{h|T}\|$$

$$|b_h(u_h, v_h)| \leq \|\beta\| \left( \sum_{T \in \mathcal{T}_h} |u_{h,0|T}|^2 |T| \right)^{1/2} \|v_h\|_{*V_h} + h_0 \|\beta\| \hat{c}_0 \sum_{T \in \mathcal{T}_h} |\nabla u_{h|T}| |T| \|\nabla v_{h|T}\|$$

$$|b_h(u_h, v_h)| \stackrel{\text{lemme 2.10}}{\leq} \|\beta\| \sqrt{3} \|u_h\|_{L^2(\Omega)} \|v_h\|_{*V_h} + h_0 \|\beta\| \hat{c}_0 \|u_h\|_{*V_h} \|v_h\|_{*V_h}$$

$$|b_h(u_h, v_h)| \stackrel{(2.87)}{\leq} \sqrt{3} \|\beta\| \hat{c}_* \|u_h\|_{*V_h} \|v_h\|_{*V_h} + h_0 \|\beta\| \hat{c}_0 \|u_h\|_{*V_h} \|v_h\|_{*V_h}$$

□

**Lemme 2.17** Soit  $w \in V_h$  et  $\varphi \in C^1(\Omega)$  avec  $\varphi \geq 0$ , alors

$$b_h(w_h, \rho_h(w_h \varphi)) \geq \frac{1}{2} \sum_{i \in J_0} \sum_{j \in S^+(i)} \theta_{i,j} w_h(Q_i)^2 (\varphi(Q_i) - \varphi(Q_j)) \quad (2.97)$$

**Preuve :** Soit  $w_{h,i} = w_h(Q_i)$  et  $\varphi_i = \varphi(Q_i)$ , alors (sommation sur  $J_0$  car  $w_h(Q_i) = 0$  pour  $Q_i \in \Gamma$ ) :

$$\begin{aligned} b_h(w_h, \rho_h(w_h \varphi)) &= \sum_{i \in J_0} \left[ \sum_{j \in S^+(i)} \theta_{i,j} w_{h,i} + \sum_{j \in S^-(i)} \theta_{i,j} w_{h,j} \right] w_{h,i} \varphi_i \\ b_h(w_h, \rho_h(w_h \varphi)) &= \sum_{i \in J_0} \sum_{j \in S^+(i)} \theta_{i,j} w_{h,i}^2 \varphi_i + \sum_{i \in J} \sum_{j \in S^-(i)} \theta_{i,j} w_{h,i} w_{h,j} \varphi_i \end{aligned}$$

Pour  $j \in S^-(i)$  on a  $\theta_{i,j} \leq 0$ , donc  $\theta_{i,j} w_{h,i} w_{h,j} \varphi_i \geq \theta_{i,j} \varphi_i \frac{1}{2} (w_{h,i}^2 + w_{h,j}^2)$

$$\begin{aligned} b_h(w_h, \rho_h(w_h \varphi)) &\geq \sum_{i \in J_0} \sum_{j \in S^+(i)} \theta_{i,j} w_{h,i}^2 \varphi_i + \sum_{i \in J_0} \sum_{j \in S^-(i)} \theta_{i,j} \varphi_i \frac{1}{2} w_{h,i}^2 + \sum_{l \in J_0} \sum_{k \in S^-(l)} \theta_{l,k} \varphi_l \frac{1}{2} w_{h,k}^2 \\ &\quad \text{Maintenant } k \in S^-(l) \iff l \in S^+(k) \text{ et} \\ \sum_{l \in J_0} \sum_{k \in S^-(l)} \theta_{l,k} \varphi_l \frac{1}{2} w_{h,k}^2 &= \sum_{k \in J_0} \sum_{l \in S^+(k)} \theta_{l,k} \varphi_l \frac{1}{2} w_{h,k}^2 = \sum_{i \in J_0} \sum_{j \in S^+(i)} -\theta_{i,j} \varphi_j \frac{1}{2} w_{h,i}^2 \\ b_h(w_h, \rho_h(w_h \varphi)) &\geq \frac{1}{2} \sum_{i \in J_0} \sum_{j \in S^+(i)} \theta_{i,j} w_{h,i}^2 (\varphi_i - \varphi_j) + \sum_{i \in J_0} \varphi_i \frac{1}{2} w_{h,i}^2 \left[ \sum_{j \in S^-(i)} \theta_{i,j} + \sum_{j \in S^+(i)} \theta_{i,j} \right] \\ &\geq \frac{1}{2} \sum_{i \in J_0} \sum_{j \in S^+(i)} \theta_{i,j} w_{h,i}^2 (\varphi_i - \varphi_j) \text{ avec (2.59)} \end{aligned}$$

□

## 2.11 Formulation faible du problème approché d'advection-diffusion

Le problème approché correspondant à (1.6) est :

**Définition 2.9** *Formulation faible du problème approché d'advection-diffusion.*  
 Trouver  $u_h \in V_h$  tel que

$$a_h(u_h, v_h) = (f, v_h) \quad \forall v_h \in V_h \quad (2.98)$$

$$\text{où } a_h(u_h, v_h) = \nu((u_h, v_h))_h + b_h(u_h, v_h) \quad (2.99)$$

## 2.12 Formulation faible du problème approché d'advection-diffusion non stationnaire

Le problème approché correspondant à (1.6) nécessite une discrétisation de la variable de temps :

Soit  $N \in \mathbb{N}$ , le pas pour  $t$  est défini par  $\tau = \frac{T}{N+1}$  et on pose  $g_h^k(x) = I_h(g(\cdot, \tau k))(x)$  pour  $k \in \{0, \dots, N+1\}$ .

**Définition 2.10** *Formulation faible du problème approché d'advection-diffusion non stationnaire.*

Trouver  $u_h^k \in V_h$  pour  $k \in \{1, \dots, N+1\}$  tels que  $u_h^0 = I_h(u^0)$  et

$$\frac{1}{\tau}(u_h^{k+1} - u_h^k, v_h) + \nu((u_h^{k+1}, v_h))_h + b_h(u_h^{k+1}, v_h) = (g_h^{k+1}, v_h) \quad \forall v_h \in V_h \quad \forall k \in \{1, \dots, N+1\} \quad (2.100)$$

La discrétisation est implicite, car les termes de diffusion et de convection sont pris en  $t = \tau(k+1)$ .

Dans [AsFF98], une discrétisation semi-implicite est considérée, dans laquelle le terme convectif est pris en  $k\tau$  et le terme de diffusion en  $(k+1)\tau$ .

## 2.13 Formulation faible du problème approché d'advection

On définit l'espace des solutions par  $Y_h$  et l'espace des fonctions tests par  $Y_h^0$

Les fonctions  $f \in W^{l,1}(\Omega)$ ,  $l \in \{1, 2\}$  et  $g \in W^{1,1}(\Gamma^-)$  sont les données du problème exact (voir Théorème 1.3 et 1.5).

**Pour**  $g_h$  :  $g_h = I_h|_{\Gamma^-}(g)$  défini par  $g_h(Q_i) = \frac{1}{|S_i|} \int_{S_i} g \, d\sigma$ .

**Pour**  $f_h$  :  $f_h = I_h f$



**Remarque 2.11** Pour  $f \in W^{2,1}(\Omega)$ , il est possible aussi d'approcher  $f$  par une interpolation  $W^{1,1}$ -conforme, car  $f$  est continue (voir par exemple le théorème d'injection de Sobolev 1.4). L'approche ci-dessus permet de traiter les cas  $f \in W^{1,1}(\Omega)$  et  $f \in W^{2,1}(\Omega)$  ensemble.

La formulation variationnelle ( $b_h$  définie par (2.63)) du problème approché d'advection I<sup>14</sup> définie par

**Définition 2.12 (Formulation variationnelle I du problème approché d'advection)**

Trouver  $u_h \in Y_h$  tel que

$$\hat{b}_h(u_h, v_h) = (f_h, v_h) \quad \forall v_h \in Y_h^0 \text{ avec } u_h(Q_i) = g_h(Q_i) \text{ pour } i \in J^-$$

avec<sup>15</sup>

$$\hat{b}_h(u_h, v_h) = \sum_{i \in J \setminus J^-} v_h(Q_i) \left[ \sum_{j \in S^+(i)} \theta_{ij} u_h(Q_i) + \sum_{j \in S^-(i)} \theta_{ij} u_h(Q_j) \right], \quad \forall u_h \in Y_h, v_h \in Y_h^0 \quad (2.101)$$

Rappelons que l'ensemble  $\{w_{h,i} : i \in J \setminus J^-\}$  des fonctions  $w_{h,i} \in Y_h$  telles que  $w_{h,i}(Q_j) = \delta_{ij} = \begin{cases} 1 & \text{si } i=j \\ 0 & \text{si } i \neq j \end{cases}$  forme une base de  $Y_h^0$  et, dans cette base, on a

$$(f_h, v_h) = \sum_{k \in J \setminus J^-} f_h(Q_k) v_h(Q_k) |D_k| \quad (2.102)$$

où  $D_k \in \mathcal{D}_h$  est le volume barycentrique associé à  $Q_k$ .

---

14. Au chapitre 17.3, pour des raisons techniques, une forme variationnelle II va être introduite.

15. Remarquons au passage que la définition de  $\hat{b}$  diffère légèrement de celle de  $b$  : pour  $b$  on somme sur  $J_0$  (les points intérieurs de  $\Omega$ ) et pour  $\hat{b}$  la sommation s'effectue sur  $J \setminus J^-$ , tous les points saufs ceux du bord entrant.

---

## Deuxième partie

# Stabilité de l'équation d'advection-diffusion stationnaire

## Sommaire

---

<b>3</b>	<b>Introduction</b>	<b>50</b>
<b>4</b>	<b>Conditions géométriques pour le maillage</b>	<b>51</b>
<b>5</b>	<b>Stabilité du problème approché</b>	<b>53</b>
<b>6</b>	<b>Étude de familles de maillage vérifiant les conditions de stabilité</b>	<b>54</b>
6.1	Etude de la condition $\delta - S^+$ . . . . .	54
6.2	Etude des conditions $\delta - S_\beta^+$ -faible et $\delta - S^+ - faible$ . . . . .	61
6.3	Simulations numériques . . . . .	66

---

### 3 Introduction

Soit  $0 \neq \beta \in \mathbb{R}^2$ ,  $\Omega$  un domaine polygonal borné et  $f \in L^2(\Omega)$ . Au chapitre II, on considère le problème stationnaire d'advection-diffusion

$$\begin{aligned} -\nu \Delta u + \beta \cdot \nabla u &= f \text{ dans } \Omega \\ u &= 0 \text{ sur } \partial\Omega \end{aligned} \quad (3.103)$$

Une formulation faible de (3.103) est obtenue par multiplication de cette équation par  $v \in V = H_0^1(\Omega)$  et intégration sur  $\Omega$ . Avec la formule de Green, on obtient

**Définition 3.1** *Une fonction  $u \in V$  est une solution faible du problème (3.103), si*

$$\nu((u, v)) + b(u, v) = \int_{\Omega} f v \, dx, \quad \forall v \in V \quad (3.104)$$

où

$$b(u, v) = \int_{\Omega} \beta \cdot \nabla u v \, dx \text{ pour } u, v \in V \quad (3.105)$$

Le problème approché pour (3.104) est

**Définition 3.2** *Formulation faible du problème approché d'advection-diffusion*

*Trouver  $u_h \in V_h$  tel que*

$$\begin{aligned} a_h(u_h, v_h) &= (f, v_h) \, dx \quad \forall v_h \in V_h \\ \text{où } a_h(u_h, v_h) &= \nu((u_h, v_h))_h + b_h(u_h, v_h) \end{aligned} \quad (3.106)$$

où  $b_h$  est défini par (2.63).

Pour le problème continu, le résultat de stabilité  $L^2$  (c'est-à-dire indépendant de  $\nu$ ) est bien connu :

**Théorème 3.1 ( $L_2$ -Estimation  $\nu$  indépendante)** *Soit  $u \in V$  la solution du problème (1.6), alors*

$$\|u\|_{L^2(\Omega)} \leq \frac{\text{diam}(\Omega)}{\|\beta\|} \|f\|_{L^2(\Omega)} \text{ et } \sqrt{\nu} \|u\|_{*V} \leq \sqrt{\frac{\text{diam}(\Omega)}{\|\beta\|}} \|f\|_2 \quad (3.107)$$

et  $\text{diam}(\Omega)$  est le diamètre<sup>16</sup> de  $\Omega$ .

Le résultat correspondant pour le problème approché va être établi. Sa preuve nécessite d'ajouter aux conditions standard de la famille de maillage  $(\mathcal{T}_h)_h$  présentées au chapitre 2.1, une condition supplémentaire (4.112).

La preuve du résultat de stabilité pour le problème approché est très proche de celle du problème continu. Il est donc intéressant de la présenter :

---

16. c'est-à-dire le diamètre de la plus petite boule contenant  $\Omega$

**Preuve :** Pour  $v = u$  dans (1.6) on a

$$\nu \|u\|_{*V}^2 + \underbrace{\int_{\Omega} (\beta \cdot \nabla u) u \, dx}_{= \frac{1}{2} \int_{\Omega} \nabla \cdot (\beta u^2) \, dx = 0 \text{ (*)}} = (f, u) \quad (3.108)$$

Soit  $\varphi(x) = \|\beta\| \text{diam}(\Omega) - \beta \cdot (x - x_0)$  avec  $x_0 \in \partial\Omega$  tel que  $0 \leq \varphi(x) \leq \|\beta\| \text{diam}(\Omega)$  pour  $x \in \Omega$ .

Posons  $v = u\varphi$ , on a alors

$$\begin{aligned} \int_{\Omega} (\beta \cdot \nabla u) u \varphi \, dx &= (f, u\varphi) - \nu \int_{\Omega} \nabla u \cdot [\nabla u \varphi + u\beta] \, dx \\ \int_{\Omega} (\beta \cdot \nabla u) u \varphi \, dx &\stackrel{\text{avec (*)}}{\leq} (f, u\varphi) - \nu \int_{\Omega} \nabla u \cdot \nabla u \varphi \, dx \\ 0 = \int_{\Omega} \nabla \cdot (\beta u^2 \varphi) \, dx &= 2 \int_{\Omega} (\beta \cdot \nabla u) u \varphi \, dx - \int_{\Omega} \|\beta\|^2 u^2 \, dx \\ \text{Donc } \frac{1}{2} \int_{\Omega} \|\beta\|^2 u^2 \, dx &\leq \int_{\Omega} (\beta \cdot \nabla u) u \varphi \, dx \end{aligned} \quad (3.109)$$

Maintenant :

$$\begin{aligned} (f, u\varphi) &\leq \|f\|_2 \|u\|_2 \|\varphi\|_{\infty} \leq \|f\|_2 \|u\|_2 \|\beta\| \text{diam}(\Omega) \\ -\nu \int_{\Omega} \nabla u \cdot \nabla u \varphi \, dx &\leq \nu \int_{\Omega} \nabla u \cdot \nabla u \varphi \, dx \\ &\leq \|\beta\| \text{diam}(\Omega) \nu \int_{\Omega} \nabla u \cdot \nabla u \, dx \\ &\stackrel{(3.108)}{\leq} \|\beta\| \text{diam}(\Omega) \|f\|_2 \|u\|_2 \end{aligned}$$

Ainsi

$$\frac{1}{2} \int_{\Omega} \|\beta\|^2 u^2 \, dx \stackrel{(3.109)}{\leq} \|f\|_2 \|u\|_2 \|\beta\| \text{diam}(\Omega) + \|\beta\| \text{diam}(\Omega) \|f\|_2 \|u\|_2$$

Donc  $\|u\|_2 \leq 4 \frac{\text{diam}(\Omega)}{\|\beta\|} \|f\|_2$  et, avec (3.109) :  $\nu \|u\|_{*V}^2 \leq 4 \frac{\text{diam}(\Omega)}{\|\beta\|} \|f\|_2^2$   $\square$

## 4 Conditions géométriques pour le maillage

La preuve de stabilité pour le problème approché, qui sera présentée au chapitre suivant, utilise les mêmes ingrédients que le théorème 3.1. En particulier, est aussi utilisée une fonction test  $\varphi$  semblable à celle de la preuve du cas continu. Si dans le cas continu, le choix d'un autre vecteur que  $\beta$  dans la définition de  $\varphi$  n'est pas intéressant, il n'en est pas de même pour le problème approché. En effet, l'utilisation d'un autre vecteur  $\tilde{\beta}$  pour  $\varphi$  permet d'affaiblir la condition supplémentaire demandée à la famille de maillage. Ci-dessous sont ainsi présentées les différentes conditions sur la famille de maillage :

**Condition  $\delta - S_{\beta}^+$ -faible** Dans la condition  $\delta - S_{\beta}^+$ -faible, on considère que  $\beta$  est fixe et le théorème 5.1 donne un résultat de stabilité indépendant de  $\nu$  et  $h$ .

**Condition  $\delta - S^+$ -faible** Dans la condition  $\delta - S^+$ -faible<sup>17</sup>, on formule une condition sur la famille de maillage indépendante de la direction de  $\beta$ . Avec la condition  $\delta - S^+$ -faible, le théorème 5.1 donne un résultat de stabilité indépendant de  $\nu$ ,  $h$  et de la direction de  $\beta$ .

**Condition  $\delta - S^+$**  Cette condition est la plus forte des conditions pour la famille de maillage, car elle formule une condition sur le maillage indépendante de la direction de  $\beta$  et correspond au choix de  $\tilde{\beta} = \beta$  dans  $\varphi$ . Avec la condition  $\delta - S^+$  le théorème 5.1 donne un résultat de stabilité indépendant de  $\nu$ ,  $h$  et de la direction de  $\beta$ . C'est seulement pour cette dernière condition, qu'une étude explicite a été possible montrant qu'elle est vérifiée par certains types de maillages structurés (voir le lemme 6.1).

La dépendance de  $\theta_{i,j}^\beta$  à  $\beta$  va être rendue explicite dans tout ce chapitre :

$$\theta_{i,j}^\beta = \int_{S_i} \beta \cdot n_{i,j} d\sigma$$

**Définition 4.1 (Condition  $\delta - S_\beta^+$ -faible pour  $(\mathcal{T}_h)_h$ )** Soit  $0 \neq \beta \in \mathbb{R}^2$ ,  $e = \frac{\beta}{\|\beta\|}$  et  $(\mathcal{T}_h)_h$  une famille de maillage. On dira que la condition  $\delta - S_\beta^+$ -faible est satisfaite pour cette famille, si il existe une constante  $\delta_e > 0$  tel que

$\forall h > 0 \exists \tilde{\beta} \in \mathbb{R}^2 \neq 0$  avec  $\|\tilde{\beta}\| = \|\beta\|$  et

$$\forall i \in J_0 \sum_{j \in S^+(i)} \theta_{i,j}^\beta [\tilde{\beta} \cdot (Q_j - Q_i)] \geq \delta_e \|\beta\|^2 |S_i|^2 \quad (4.110)$$

Si un résultat de stabilité veut être obtenu indépendamment de la direction de  $\beta$ , la condition suivante est suffisante :

**Définition 4.2 (Condition  $\delta - S^+$ -faible pour  $(\mathcal{T}_h)_h$ )** Soit  $(\mathcal{T}_h)_h$  une famille de maillage. On dira que la condition  $\delta - S^+$ -faible est satisfaite pour cette famille, si il existe une constante  $\delta > 0$ , telle que

$\forall \beta \in \mathbb{R}^2 \forall h > 0 \exists \tilde{\beta} \in \mathbb{R}^2 \neq 0$  avec  $\|\tilde{\beta}\| = \|\beta\|$  et

$$\forall i \in J_0 \sum_{j \in S^+(i)} \theta_{i,j}^\beta [\tilde{\beta} \cdot (Q_j - Q_i)] \geq \delta \|\beta\|^2 |S_i|^2 \quad (4.111)$$

Pour le choix  $\tilde{\beta} = \beta$ , la condition ci-dessus devient

**Définition 4.3 (Condition  $\delta - S^+$  pour  $(\mathcal{T}_h)_h$ )** Soit  $(\mathcal{T}_h)_h$  une famille de maillage. On dira que la condition  $\delta - S^+$  est satisfaite pour cette famille, si il existe une constante  $\delta > 0$  indépendante de  $h$  et de  $\beta$  telle que

$$\sum_{j \in S^+(i)} \theta_{i,j}^\beta [\beta \cdot (Q_j - Q_i)] \geq \delta \|\beta\|^2 |S_i|^2 \quad \forall i \in J_0 \text{ et } \beta \in \mathbb{R}^2 \quad (4.112)$$

---

17. cette condition est plus forte que la condition  $\delta - S_\beta^+$ -faible

## 5 Stabilité du problème approché

**Théorème 5.1** ( $L^2$ -stabilité pour  $u_h$  (indépendante de  $\nu$  et  $h$ )) *On suppose les conditions standard définies au chapitre 2.1 vérifiées pour la famille de maillage  $(\mathcal{T}_h)_h$ , ainsi que la condition  $\delta - S^+$ -faible (4.111).*

Soit  $u_h \in V_h$  la solution du problème approché (3.106) alors

$$\|u_h\|_2 \leq K \frac{1}{\delta \|\beta\|} \|f\|_2 \text{ et } \sqrt{\nu} \|u_h\|_{*V_h} \leq K \frac{1}{\sqrt{\delta \|\beta\|}} \|f\|_2 \quad (5.113)$$

où  $K > 0$  est une constante ne dépendant pas de  $\nu$ ,  $\delta$  et  $\beta$ .

**Preuve :** La preuve est proche de celle du problème exact.

Soit  $u_h \in V_h$  la solution du problème approché (3.106), alors avec le lemme 2.17 et  $\varphi = 1$  on a  $b_h(u_h, u_h) \geq 0$ , donc

$$\nu \|u_h\|_{*V_h}^2 \leq (f, u_h) \quad (5.114)$$

Comme la condition  $\delta - S_\beta^+$ -faible est vérifiée, il existe  $\tilde{\beta}$  avec  $\|\tilde{\beta}\| = \|\beta\|$  et (4.110).

Soit  $\varphi(x) = \|\tilde{\beta}\| \text{diam}(\Omega) - \tilde{\beta} \cdot (x - x_0)$  avec  $x_0 \in \partial\Omega$  tel que  $0 \leq \varphi(x) \leq \|\beta\| \text{diam}(\Omega)$  pour  $x \in \Omega$ . Pour  $v_h = \rho_h(u_h \varphi)$  on obtient

$$-\nu((u_h, \rho_h(u_h \varphi)))_h + (f, \rho_h(u_h \varphi)) = b_h(u_h, \rho_h(u_h \varphi)) \quad (5.115)$$

Avec lemme 2.17 et l'inégalité (4.110) on a<sup>18</sup>

$$\begin{aligned} b_h(u_h, \rho_h(u_h \varphi)) &\geq \frac{1}{2} \sum_{i \in J_0} \sum_{j \in S^+(i)} \theta_{i,j}^\beta u_h(Q_i)^2 \tilde{\beta} \cdot (Q_j - Q_i) \\ &\geq \frac{1}{2} \delta \frac{\beta}{\|\beta\|} \|\beta\|^2 \sum_{i \in J_0} u_h(Q_i)^2 |S_i|^2 \end{aligned} \quad (5.116)$$

$$\text{Avec lemme 2.10 et (2.38) } |T| \leq h_T^2 \leq \frac{1}{\sigma^2} |S_i|^2$$

$$\geq \frac{\delta \frac{\beta}{\|\beta\|}}{2\sigma^2} \|\beta\|^2 \|u_h\|_2^2$$

Avec lemme 2.13 :

$$\begin{aligned} -\nu((u_h, \rho_h(u_h \varphi)))_h &\leq \nu \|u_h\|_{*V_h} \|\rho_h(u_h \varphi)\|_{*V_h} \\ &\leq \nu \hat{c}_3 \|u_h\|_{*V_h} (\|u_h\|_{*V_h} \|\varphi\|_\infty + \|u_h\|_2 \|\nabla \varphi\|_\infty) \\ &\quad \text{Avec l'inégalité de Friedrichs (2.87) et la définition de } \varphi : \\ &\leq \nu \hat{c}_3 \|u_h\|_{*V_h}^2 \|\beta\| \text{diam}(\Omega) + \nu \hat{c}_3 \|u_h\|_{*V_h} \hat{c}_* \|u_h\|_{*V_h} \|\beta\| \\ &\leq \nu \hat{c}_3 \|\beta\| \|u_h\|_{*V_h}^2 [\text{diam}(\Omega) + \hat{c}_*] \\ (5.114) &\leq \hat{c}_3 \|\beta\| [\text{diam}(\Omega) + \hat{c}_*] (f, u_h) \\ &\leq \hat{c}_3 \|\beta\| [\text{diam}(\Omega) + \hat{c}_*] \|u_h\|_2 \|f\|_2 \\ (2.90) &\leq \|f\|_2 \|u_h\|_2 \|\varphi\|_\infty \\ &\leq \|f\|_2 \|u_h\|_2 \|\beta\| \text{diam}(\Omega) \end{aligned}$$

18. C'est le seul endroit où la condition  $\delta - S_\beta^+$ -faible intervient.

En combinant ces inégalités avec (5.115) on a :

$$\begin{aligned} \frac{\delta}{2\sigma^2} \|\beta\|^2 \|u_h\|_2^2 &\leq \hat{c}_3 \|\beta\| [ \text{diam}(\Omega) + \hat{c}_* ] \|u_h\|_2 \|f\|_2 + \|f\|_2 \|u_h\|_2 \|\beta\| \text{diam}(\Omega) \\ \|u_h\|_2 &\leq \frac{2\sigma^2}{\delta \|\beta\|} [ \hat{c}_3 ( \text{diam}(\Omega) + \hat{c}_* ) + \text{diam}(\Omega) ] \|f\|_2 \end{aligned}$$

Avec cette inégalité et (5.114), on obtient

$$\nu \|u_h\|_{*V_h}^2 \leq \frac{2\sigma^2}{\delta \|\beta\|} [ \hat{c}_3 ( \text{diam}(\Omega) + \hat{c}_* ) + \text{diam}(\Omega) ] \|f\|_2^2 \quad (5.117)$$

□

## 6 Étude de familles de maillage vérifiant les conditions de stabilité

La vérification des conditions formulées ci-dessus pour une famille de maillage n'est pas facile.

La condition  $\delta - S^+$  est la seule pour laquelle une approche purement algébrique a été obtenue. Et cela, uniquement dans le cas d'un maillage symétrique. C'est ce résultat qui sera présenté en premier :

### 6.1 Etude de la condition $\delta - S^+$

**Définition 6.1 (Condition  $\delta - S^+$  pour  $(\mathcal{T}_h)_h$ )** Soit  $(\mathcal{T}_h)_h$  une famille de maillage. On dira que la condition  $\delta - S^+$  est satisfaite pour cette famille, si il existe une constante  $\delta > 0$  indépendante de  $h$  telle que<sup>19</sup>

$$\sum_{j \in S^+(i)} \theta_{i,j}^\beta [\beta \cdot (Q_j - Q_i)] \geq \delta \|\beta\|^2 |S_i|^2 \quad \forall i \in J_0 \text{ et } \beta \in \mathbb{R}^2 \quad (6.118)$$

#### Conditions suffisantes pour un maillage symétrique uniforme

On se place dans le cas d'un maillage symétrique, c'est-à-dire où deux mailles  $T_i^+$  et  $T_i^-$  ayant un côté en commun  $S_i$  sont symétriques par rapport à  $Q_i$ , le milieu de  $S_i$  :

**Lemme 6.1** Soit  $(\mathcal{T}_h)_h$  une famille de maillage telle que

- i) Pour tout  $i \in J_0$ , les deux triangles  $T_i^+$  et  $T_i^-$  ayant  $S_i$  comme côté commun sont symétriques par rapport à  $Q_i$ , le milieu du segment  $S_i$ . (Voir figure 2.)
- ii) Pour tout  $i \in J_0$ , on considère le repère orthonormal d'origine  $Q_i$  et d'axes  $x$  et  $y$ , tel que l'axe  $x$  est perpendiculaire à  $S_i$  et pointant vers  $T_i^+$ . Alors pour le sommet  $R_i^+(x_i^+, y_i^+)$  de  $T_i^+$  (opposé à  $S_i$ ) on demande

---

19.  $\theta_{i,j}^\beta = \int_{S_i} \beta \cdot n_{i,j} d\sigma$

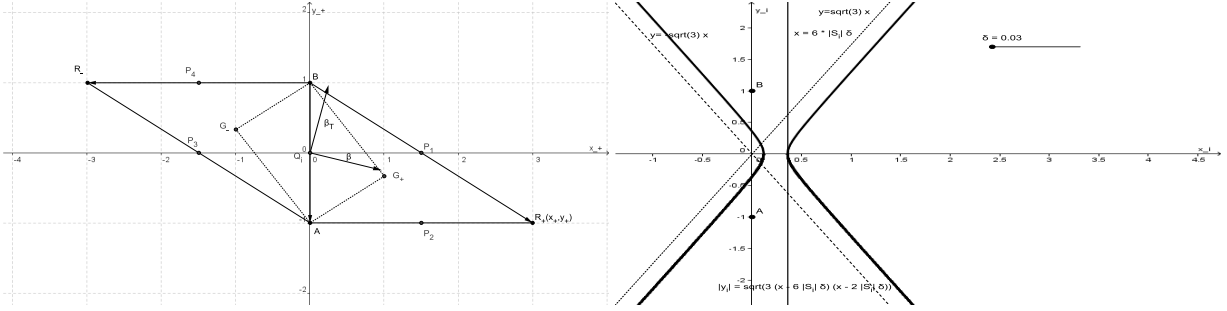


FIGURE 2 – Maillage symétrique par rapport au milieu  $Q_i$  du segment commun  $S_i$ . Les positions de  $R_+$  qui correspondent à  $\delta > 0$  sont celles contenues dans le cône  $|y_i^+| < \sqrt{3}x^+$ .

$$\begin{cases} \frac{1}{6|S_i|} \left( 2x_i^+ - \sqrt{(x_i^+)^2 + (y_i^+)^2} \right) \geq \delta \\ \frac{1}{6|S_i|} \left( 2x_i^+ + \sqrt{(x_i^+)^2 + (y_i^+)^2} \right) \geq \delta \end{cases} \quad \forall i \in J_0 \quad (6.119)$$

Cette condition est équivalente à

$$\begin{cases} x_i^+ \geq 6\delta|S_i| \\ \sqrt{3(x_i^+ - 6\delta|S_i|)(x_i^+ - 2\delta|S_i|)} \geq |y_i^+| \end{cases} \quad \forall i \in J_0 \quad (6.120)$$

Alors les conditions *i*) and *ii*) impliquent (4.112).

**Preuve :** Soit  $i \in J_0$ ,  $T_i^+$  et  $T_i^-$  ayant le côté  $S_i$  en commun et symétriques par rapport à  $Q_i$ , le milieu de  $S_i$ . Les points  $R_i^+$  et  $R_i^-$  sont les sommets opposés à  $S_i$  de, respectivement,  $T_i^+$  et  $T_i^-$ .

Soit  $A$  et  $B$  les extrémités de  $S_i$ , et  $P_1, P_2, P_3, P_4$  les milieux respectivement des segments  $[A, R_i^+]$ ,  $[R_i^+, B]$ ,  $[B, R_i^-]$ ,  $[R_i^-, A]$ , alors  $A(0, \frac{|S_i|}{2})$ ,  $B(0, -\frac{|S_i|}{2})$ ,  $R_i^+(x_i^+, y_i^+)$  et  $R_i^-(-x_i^+, -y_i^+)$ .

Pour le vecteur  $0 \neq \beta \in \mathbb{R}^2$  on a :  $\theta_{Q_i, P_1}^\beta = AG_i^+ \cdot \beta \cdot n_{i,1} = \beta \cdot AG_i^{+T}$ , où  $AG_i^{+T}$  est la rotation d'angle  $\frac{\pi}{2}$  du vecteur  $AG_i^+$ .

Exprimé en fonction des coordonnées de  $R_i^+$  on a  $\theta_{Q_i, P_1}^\beta = \beta \cdot \begin{pmatrix} \frac{|S_i|}{2} - \frac{y_i^+}{3} \\ \frac{x_i^+}{3} \end{pmatrix}$ .

De même pour  $\theta_{Q_i, P_2}^\beta, \theta_{Q_i, P_3}^\beta, \theta_{Q_i, P_4}^\beta$  on obtient

$$\begin{aligned} \theta_{Q_i, P_1}^\beta &= \begin{pmatrix} \frac{|S_i|}{2} - \frac{y_i^+}{3} \\ \frac{x_i^+}{3} \end{pmatrix} \cdot \beta = -\theta_{Q_i, P_3}^\beta \quad \text{et} \quad \beta \cdot (P_1 - Q_i) = \begin{pmatrix} \frac{x_i^+}{2} \\ \frac{y_i^+ + |S_i|}{2} \end{pmatrix} \cdot \beta = -\beta \cdot (P_3 - Q_i) \\ \theta_{Q_i, P_2}^\beta &= \begin{pmatrix} \frac{|S_i|}{2} + \frac{y_i^+}{3} \\ -\frac{x_i^+}{3} \end{pmatrix} \cdot \beta = -\theta_{Q_i, P_4}^\beta \quad \text{et} \quad \beta \cdot (P_2 - Q_i) = \begin{pmatrix} \frac{x_i^+}{2} \\ \frac{y_i^+ - |S_i|}{2} \end{pmatrix} \cdot \beta = -\beta \cdot (P_4 - Q_i) \end{aligned} \quad (6.121)$$



Soit  $s_j = \theta_{Q_i, P_j}^\beta [\beta \cdot (P_j - Q_i)]$ ,  $j \in \{1, 2, 3, 4\}$ .

Pour la permutation  $m(1) = 3, m(2) = 4, m(3) = 1, m(4) = 2$ , de 1, 2, 3, 4, on constate

$$s_{m(j)} = s_j \text{ et } j \in S^+(i) \iff m(j) \in S^-(i) \quad \forall j \in \{1, 2, 3, 4\}$$

Maintenant

$$\sum_{j \in S^+(i)} \theta_{i,j}^\beta [\beta \cdot (P_j - Q_i)] = \sum_{j \in S^+(i)} s_j = \sum_{j \in S^+(i)} s_{m(j)} = \sum_{m(j) \in S^-(i)} s_{m(j)} = \sum_{k \in S^-(i)} s_k$$

Ainsi

$$\sum_{j \in S^+(i)} \theta_{i,j}^\beta [\beta \cdot (P_j - Q_i)] = \frac{1}{2} \left( \sum_{j \in S^+(i)} s_j + \sum_{j \in S^-(i)} s_j \right) = \frac{1}{2} \sum_{j \in S(i)} s_j \quad (6.122)$$

Afin de simplifier les calculs, un changement d'échelle est effectué sur les coordonnées afin d'avoir  $|S_i| = 2$  (voir figure 2). Soit  $(X, Y) = \frac{2}{|S_i|}(x_i^+, y_i^+)$  les coordonnées de  $R_i^+$  dans repère renormé, la condition (4.112) est équivalente à

$$\frac{4}{|S_i|^2} \sum_{j \in S^+(i)} \theta_{i,j}^\beta [\beta \cdot (P_j - Q_i)] = (\beta_1, \beta_2) \underbrace{\begin{pmatrix} X & \frac{1}{3}Y \\ \frac{1}{3}Y & \frac{1}{3}Y \end{pmatrix}}_{=M} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \geq 4\delta|\beta|^2 \quad (6.123)$$

Les valeurs propres de la matrice (symétrique)  $M$  sont

$$\lambda_1 = \frac{2}{3}X + \sqrt{\left(\frac{1}{3}X\right)^2 + \left(\frac{1}{3}Y\right)^2} \text{ and } \lambda_2 = \frac{2}{3}X - \sqrt{\left(\frac{1}{3}X\right)^2 + \left(\frac{1}{3}Y\right)^2}$$

Les inégalités  $\lambda_1, \lambda_2 \geq 4\delta$  sont satisfaites si

$$\begin{cases} \frac{1}{6}X + \frac{1}{12}\sqrt{X^2 + Y^2} \geq \delta \\ \frac{1}{6}X - \frac{1}{12}\sqrt{X^2 + Y^2} \geq \delta \end{cases} \quad (6.124)$$

ou encore (calcul direct)

$$\begin{cases} X \geq 12\delta \\ \sqrt{3(X - 12\delta)(X - 4\delta)} \geq |Y| \end{cases} \quad (6.125)$$

Exprimées dans les coordonnées d'origine, les conditions (6.125) et (6.124) sont équivalentes à (6.120) et (6.119)

□

**Exemple 6.2** *L'application du lemme 6.1 à un maillage uniforme fournit l'exemple suivant :*

*Soit  $\Omega = (0, a) \times (0, b)$  et  $(\mathcal{T}_h)_h$  une famille de maillage uniforme définie par des triangles rectangles de côtés  $h_a = h$ ,  $h_b = \kappa h$  où  $\kappa > 0$  est une constante.*

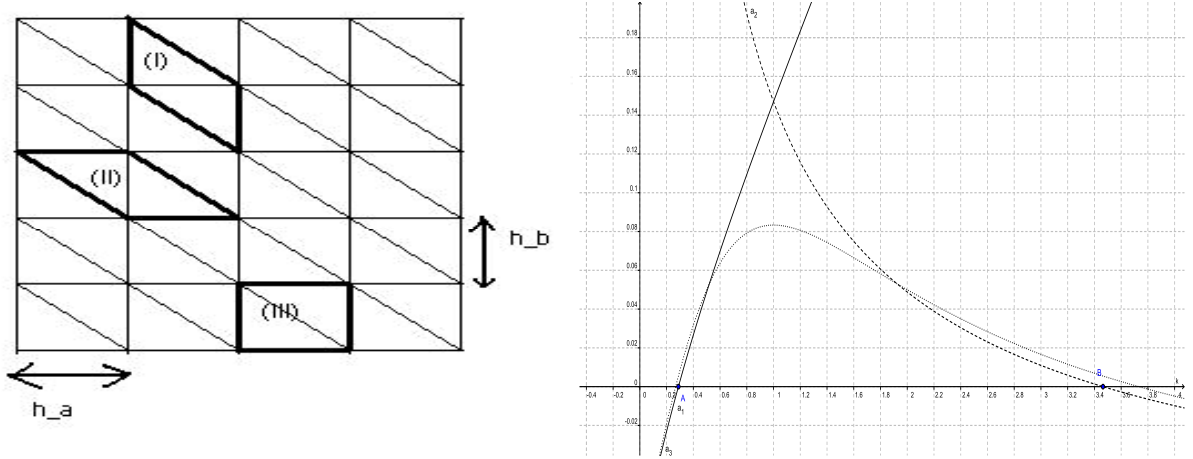


FIGURE 3 – Maillage uniforme où deux triangles ayant un côté en commun sont symétriques par rapport au milieu de ce côté.  $A(\frac{1}{2\sqrt{3}})$  et  $B(2\sqrt{3})$ . Représentation graphique de  $a_I(\kappa) = \frac{\kappa}{3} - \frac{\sqrt{1+4\kappa^2}}{12}$ ,  $a_{II}(\kappa) = \frac{1}{3\kappa} - \frac{\sqrt{1+\frac{4}{\kappa^2}}}{12}$  et  $a_{III}(\kappa) = \frac{\kappa}{3(1+\kappa^2)} - \frac{1}{12}$

La figure 3 montre que deux triangles, qui ont un côté commun  $S_i$ , sont symétriques par rapport au milieu de ce segment. Trois cas apparaissent alors, et l'inégalité (6.119) appliquée à chacun de ces cas permet de conclure que la condition  $\delta - S^+$  est satisfaite si  $2\sqrt{3} > \kappa > \frac{1}{2\sqrt{3}}$ . Pour  $\delta$  on a alors

$$0 < \delta \leq \min\left\{\frac{\kappa}{3} - \frac{\sqrt{1+4\kappa^2}}{12}, \frac{1}{3\kappa} - \frac{\sqrt{1+\frac{4}{\kappa^2}}}{12}, \frac{\kappa}{3(1+\kappa^2)} - \frac{1}{12}\right\} \quad (6.126)$$

Dans la figure 3 sont représentées graphiquement les trois expressions présentes dans le membre gauche de (6.126). Pour un maillage uniforme, la condition  $\delta - S^+$  impose donc

$$2\sqrt{3} > \frac{h_b}{h_a} > \frac{1}{2\sqrt{3}}$$

### Perturbation du maillage, condition nécessaire pour $\delta - S^+$

Si le maillage n'est pas fait de triangles symétriques par rapport au milieu du côté commun, alors le lemme suivant permet de justifier pourquoi une perturbation d'un maillage qui vérifie la condition (6.118) garde cette propriété (pour une autre valeur de  $\delta$ ).

La seconde partie du lemme est une condition nécessaire pour qu'un maillage (quelconque) vérifie la condition (6.118).

**Lemme 6.2 1)** Soient  $\mathcal{T}_h$  et  $\tilde{\mathcal{T}}_h$  deux maillages de  $\Omega$ , tels que  $\tilde{\mathcal{T}}_h$  est une  $\epsilon$  perturbation de  $\mathcal{T}_h$  dans le sens où, à chaque maille  $T \in \mathcal{T}_h$ , correspond exactement une maille  $\tilde{T} \in \tilde{\mathcal{T}}_h$ , telle que la distance entre les sommets de  $T$  et  $\tilde{T}$  soit inférieure à  $h\epsilon$ . (Sans restriction, on considère que les maillages partagent les mêmes ensembles

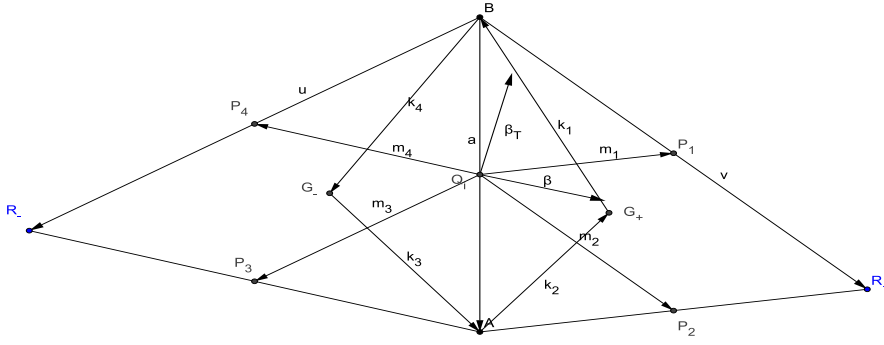


FIGURE 4 – Situation de deux triangles du maillage ayant un côté en commun.

d'indices  $I$  et  $J$  et  $J_0$ .) Alors pour tout  $0 \neq \beta \in \mathbb{R}^2$  et  $i \in J_0$  on a :

$$\left| \sum_{j \in S^+(i)} \theta_{i,j}^\beta [\beta \cdot (Q_j - Q_i)] - \sum_{j \in S^+(i)} \tilde{\theta}_{i,j}^\beta [\beta \cdot (\tilde{Q}_j - \tilde{Q}_i)] \right| \leq 12\epsilon(1 + 2\epsilon)h^2 \|\beta\|^2 \quad (6.127)$$

où  $\tilde{Q}_i$  (milieu de  $\tilde{S}_i$ ) est la perturbation de  $Q_i$  (milieu de  $S_i$ ) et

$$\theta_{i,j}^\beta = \int_{S_i} \beta \cdot n_{i,j} d\sigma \quad \text{et} \quad \tilde{\theta}_{i,j}^\beta = \int_{\tilde{S}_i} \beta \cdot n_{i,j} d\sigma$$

2) Soit  $(\mathcal{T}_h)_h$  une famille de maillage de  $\Omega$  qui vérifie la condition (6.118) pour un certain  $\delta > 0$ , alors pour tout  $i \in J_0$  on a

$$\frac{|T^-| + |T^+|}{2\|a\|^2} - \frac{\|v - u\|}{12\|a\|} (1 + \sin(\widehat{v - u, a})) \geq \delta \quad (6.128)$$

où  $T^- = ABR^-$ ,  $T^+ = ABR^+$  sont les triangles (mailles) de côté commun  $S_i = [AB]$ , avec  $u, v$  et  $a$  les vecteurs de  $\mathbb{R}^2$  définis par  $u = \overrightarrow{BR^-}$ ,  $v = \overrightarrow{BR^+}$  et  $a = \overrightarrow{BA}$ . (Voir figure 4.)

**Preuve :** On considère deux triangles  $T^+$  et  $T^-$  ayant le côté  $S_i = [AB]$  en commun (voir figure 4).  $G_-$  et  $G_+$  sont les centres de gravité respectivement de  $T^-$  et  $T^+$ . Les points  $P_1, P_2, P_3$  et  $P_4$  sont les milieux<sup>20</sup> respectivement des segments  $[BR^+]$ ,  $[R^+A]$ ,  $[AR^-]$  et  $[R^-B]$ . On note les vecteurs  $k_1 = \overrightarrow{G^+B}$ ,  $k_2 = \overrightarrow{AG^+}$ ,  $k_3 = \overrightarrow{G^-A}$ ,  $k_4 = \overrightarrow{BG^-}$ , ainsi que  $m_1 = \overrightarrow{Q_iP_1}$ ,  $m_2 = \overrightarrow{Q_iP_2}$ ,  $m_3 = \overrightarrow{Q_iP_3}$  et  $m_4 = \overrightarrow{Q_iP_4}$ .  $\beta^T$  est la rotation d'angle  $\frac{\pi}{2}$  de  $\beta$

20. Toutes ces notations dépendent en principe de  $i$ , mais, pour ne pas alourdir inutilement les notations, ces dépendances n'ont pas été explicitées.

dans le sens direct. Alors

$$\begin{aligned} \sum_{j \in S^+(i)} \theta_{i,j}^\beta [\beta \cdot (Q_j - Q_i)] &= \sum_{j \in S(i)} \eta(\theta_{i,j}^\beta) [\beta \cdot (Q_j - Q_i)] \text{ où } \eta(x) = \frac{x + |x|}{2} \\ &= \sum_{j \in S(i)} \eta(\beta^T \cdot k_j) (\beta \cdot m_j) \end{aligned}$$

$$\begin{aligned} \text{Maintenant } k_1 &= \frac{-a - v}{3}, \quad k_2 = \frac{v - 2a}{3}, \quad k_3 = \frac{-u + a}{3}, \quad k_4 = \frac{u + a}{3} \\ \text{et } m_1 &= \frac{-a + v}{2}, \quad m_2 = \frac{v}{2}, \quad m_3 = \frac{u}{2}, \quad m_4 = \frac{u - a}{2} \end{aligned}$$

Preuve du 1) : Pour le maillage perturbé, on reprend les mêmes notations que ci-dessus en ajoutant un tilde.

$$\begin{aligned} \text{On a donc } & \left| \sum_{j \in S^+(i)} \theta_{i,j}^\beta [\beta \cdot (Q_j - Q_i)] - \sum_{j \in S^+(i)} \tilde{\theta}_{i,j}^\beta [\beta \cdot (\tilde{Q}_j - \tilde{Q}_i)] \right| \\ &= \left| \sum_{j \in S(i)} \eta(\beta^T \cdot k_j) (\beta \cdot m_j) - \eta(\beta^T \cdot \tilde{k}_j) (\beta \cdot \tilde{m}_j) \right| \\ &= \left| \sum_{j \in S(i)} [\eta(\beta^T \cdot k_j) - \eta(\beta^T \cdot \tilde{k}_j)] (\beta \cdot m_j) + \eta(\beta^T \cdot \tilde{k}_j) [(\beta \cdot m_j) - (\beta \cdot \tilde{m}_j)] \right| \end{aligned}$$

$$\text{On a } |\eta(x) - \eta(y)| \leq |x - y| \text{ et } \|a - \tilde{a}\| = \|\vec{\tilde{B}\tilde{A}} - \vec{B\tilde{A}}\| \leq \|\vec{\tilde{B}B} + \vec{A\tilde{A}}\| \leq 2h\epsilon$$

$$\text{Idem pour } \|u - \tilde{u}\| \leq 2h\epsilon \text{ et } \|v - \tilde{v}\| \leq 2h\epsilon$$

$$\text{Enfin } \|a\| \leq h, \quad \|u\| \leq h, \quad \|v\| \leq h, \quad \|\tilde{a}\| \leq h(1 + 2\epsilon), \quad \|\tilde{u}\| \leq h(1 + 2\epsilon), \quad \|\tilde{v}\| \leq h(1 + 2\epsilon)$$

$$\begin{aligned} \text{Donc } &\leq 2 \|\beta\|^2 h (1 + 2\epsilon) (\|a - \tilde{a}\| + \|u - \tilde{u}\| + \|v - \tilde{v}\|) \\ &\leq 2 \|\beta\|^2 h (1 + 2\epsilon) 6 \epsilon h \end{aligned}$$

Preuve du 2) : Comme la condition  $\delta - S^+$  est vérifiée pour le maillage, alors

$$\delta \|\beta\|^2 |S_i|^2 \leq \sum_{j \in S(i)} \eta(\theta_{i,j}^\beta) [\beta \cdot (Q_j - Q_i)]$$

Pour  $-\beta$  cette inégalité est aussi vérifiée, donc

$$\delta \|\beta\|^2 |S_i|^2 \leq \sum_{j \in S(i)} \eta(\theta_{i,j}^{-\beta}) [-\beta \cdot (Q_j - Q_i)]$$

On a  $\theta_{i,j}^{-\beta} = -\theta_{i,j}^\beta$ . En ajoutant ces deux inégalités, en divisant par 2 :

$$\begin{aligned} \delta \|\beta\|^2 |S_i|^2 &\leq \frac{1}{2} \sum_{j \in S(i)} \frac{\theta_{i,j}^\beta + |\theta_{i,j}^\beta|}{2} [\beta \cdot (Q_j - Q_i)] + \frac{-\theta_{i,j}^\beta + |-\theta_{i,j}^\beta|}{2} [-\beta \cdot (Q_j - Q_i)] \\ &\leq \frac{1}{2} \sum_{j \in S(i)} \theta_{i,j}^\beta [\beta \cdot (Q_j - Q_i)] \end{aligned} \tag{6.129}$$

Avec les notations introduites ci-dessus

$$\leq \frac{1}{2} \sum_{j \in S(i)} (\beta^T \cdot k_j) (\beta \cdot m_j)$$

$$\leq \frac{1}{12} [-3(a \cdot \beta^T)(v \cdot \beta) + (v \cdot \beta^T)(a \cdot \beta) + 3(a \cdot \beta^T)(u \cdot \beta) - (u \cdot \beta^T)(a \cdot \beta)]$$

$$\leq \frac{\|\beta\|^2}{12} \left[ 3 \det \begin{vmatrix} a \cdot \frac{\beta}{\|\beta\|} & v \cdot \frac{\beta}{\|\beta\|} \\ a \cdot \frac{\beta^T}{\|\beta\|} & v \cdot \frac{\beta^T}{\|\beta\|} \end{vmatrix} + 3 \det \begin{vmatrix} u \cdot \frac{\beta}{\|\beta\|} & a \cdot \frac{\beta}{\|\beta\|} \\ u \cdot \frac{\beta^T}{\|\beta\|} & a \cdot \frac{\beta^T}{\|\beta\|} \end{vmatrix} \right. \\ \left. + 2((v - u) \cdot \frac{\beta^T}{\|\beta\|})(a \cdot \frac{\beta}{\|\beta\|}) \right]$$

Les déterminants ci-dessus sont ceux de deux vecteurs exprimés dans le repère orthonormal direct  $(\frac{\beta}{\|\beta\|}, \frac{\beta^T}{\|\beta\|})$ . Donc

$$\delta \|\beta\|^2 \|a\|^2 \leq \frac{\|\beta\|^2}{12} \left[ 3 \times 2|T^-| + 3 \times 2|T^+| + 2((v - u) \cdot \frac{\beta^T}{\|\beta\|})(a \cdot \frac{\beta}{\|\beta\|}) \right] \quad (6.130)$$

En travaillant en coordonnées polaires dans le repère  $(\frac{\beta}{\|\beta\|}, \frac{\beta^T}{\|\beta\|})$ , il est aisé de trouver la valeur minimale prise par le second terme du membre droit :

$$\delta \|a\|^2 \leq \frac{1}{12} \left[ 3 \times 2|T^-| + 3 \times 2|T^+| - 2\|v - u\| \|a\| \frac{1}{2}(1 + \sin(\widehat{v - u, a})) \right]$$

□

**Remarque 6.3 a)** *La preuve ci-dessus fait apparaître un lien étroit entre la condition  $\delta - S^+$  et un calcul d'aire. Ce lien est plus explicite au b) de la remarque page 66 et en particulier à la remarque 14.10, page 122.*

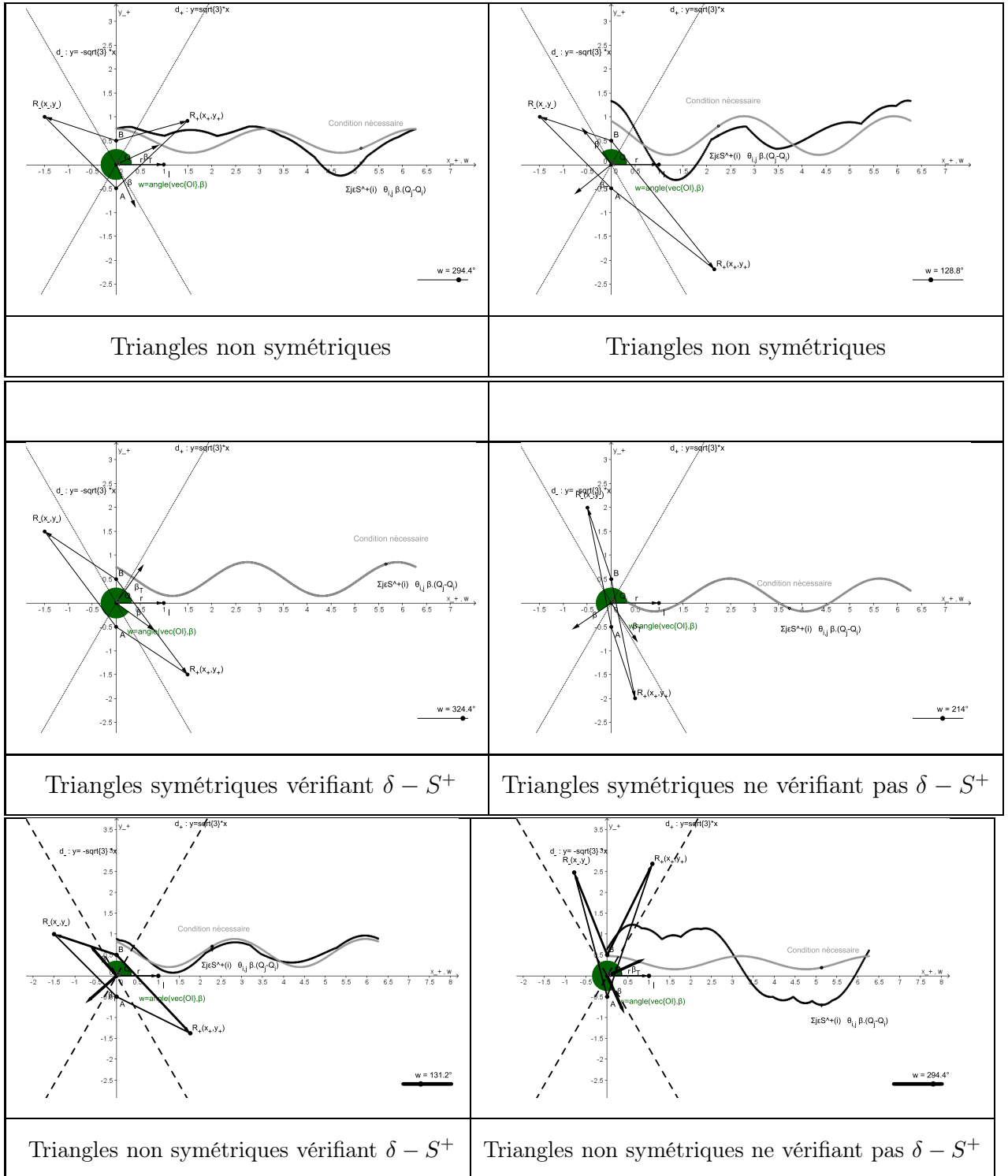
**b)** *Pour le cas d'un maillage symétrique, la condition nécessaire est aussi suffisante, car les formules (6.129) et (6.122) sont les mêmes.*

**Exemples de représentation graphique de  $\beta \mapsto \sum_{j \in S^+(i)} \theta_{i,j}^\beta [\beta \cdot (Q_j - Q_i)]$**

Dans les figures ci-dessous<sup>21</sup> sont représentées différentes situations géométriques de deux triangles  $T^- = ABR^-$  et  $T^+ = ABR^+$  ayant le côté  $S_i = [AB]$  en commun.  $Q_i$  est le milieu de  $S_i$ . Sans restriction on peut choisir  $AB = |S_i| = 1$ ,  $\|\beta\| = 1$  et représenter en fonction de  $w = (\vec{OI}, \beta)$  les quantités  $\sum_{j \in S^+(i)} \theta_{i,j}^\beta [\beta \cdot (Q_j - Q_i)]$  et  $\frac{1}{2} \sum_{j \in S(i)} \theta_{i,j}^\beta [\beta \cdot (Q_j - Q_i)]$ .

Avec les données ci-dessus, la condition  $\delta - S_\beta^+$  signifie que  $\sum_{j \in S^+(i)} \theta_{i,j}^\beta [\beta \cdot (Q_j - Q_i)] \geq \delta > 0$ . Pour que cette condition soit satisfaite, il est nécessaire (mais pas suffisant, sauf pour un maillage symétrique) que  $\frac{1}{2} \sum_{j \in S(i)} \theta_{i,j}^\beta [\beta \cdot (Q_j - Q_i)] \geq \delta > 0$  (attention, sommation ici sur tous les voisins de  $i$  dans la formule). La valeur minimale de cette dernière expression (indépendante de la direction de  $\beta$ ) est donnée par (6.128) où  $\vec{u} = \vec{BR}^-$ ,  $\vec{v} = \vec{BR}^+$  et  $\vec{a} = \vec{BA}$ . On constate que la condition nécessaire n'a d'intérêt que pour des triangles proches d'une situation de symétrie.

21. Utilisation de GeoGebra <http://www.geogebra.org/>



## 6.2 Etude des conditions $\delta - S^+_{\beta}$ -faible et $\delta - S^+$ - faible

Il semble compliqué de montrer que les conditions  $\delta - S^+_{\beta}$ -faible et  $\delta - S^+$  - faible s'appliquent à un grand choix de familles de maillage (symétrique ou non). Les approches numériques faites, confortent l'idée que la réalisation de ces deux conditions n'est pas une très forte restriction supplémentaire pour des familles de maillage vérifiant les conditions standard présentées au chapitre 2.1.

Soit  $(\mathcal{T}_h)_h$  une famille de maillage et  $h > 0$ .

Supposons  $0 \neq \beta \in \mathbb{R}^2$  quelconque, mais ensuite fixe, et notons  $t = \widehat{(\beta, \tilde{\beta}_t)}$  l'angle orienté entre  $\beta$  et  $\tilde{\beta}_t$  (avec  $\|\tilde{\beta}_t\| = \|\beta\|$ ).

Pour  $i \in J_0$ , notons  $\eta_i(\beta, \tilde{\beta}_t) = \sum_{j \in S^+(i)} \theta_{i,j}^\beta [\tilde{\beta}_t \cdot (Q_j - Q_i)]$ .

Pour  $i \in J_0$ , la fonction  $t \mapsto \eta_i(\beta, \tilde{\beta}_t)$  est périodique de période  $2\pi$  et aura toujours des valeurs strictement positives<sup>22</sup> dans un certain intervalle ouvert  $]t_i^-; t_i^+[$  :

$$\sum_{j \in S^+(i)} \theta_{i,j}^\beta [\tilde{\beta}_t \cdot (Q_j - Q_i)] > 0 \iff t \in ]t_i^-; t_i^+[$$

La difficulté principale est de prouver l'existence de  $\delta > 0$  et de  $\bar{t} \in \cap_{i \in J_0} ]t_i^-; t_i^+[$ , tel que  $\min_{i \in J_0} \frac{\eta_i(\beta, \tilde{\beta}_t)}{\|\beta\|^2 |S_i|^2} \geq \delta > 0$ .

**Remarque 6.4** *Un tel résultat pour une certaine classe de maillage non uniforme serait très intéressant, mais serait aussi intéressant un exemple de maillage ne vérifiant pas cette condition : En effet, si une famille de maillage ne vérifie pas la condition  $\delta - S_\beta^+$ -faible, existe-il alors un exemple concret de fonction  $f$ , de domaine  $\Omega$  et  $(\mathcal{T}_h)_h$  illustrant une perte de stabilité telle que*

$$\lim_{n \rightarrow \infty} \|u_{h_n, \nu_n}\|_{L^2(\Omega)} = \infty$$

pour des suites  $h_n \rightarrow 0$ ,  $\nu_n \rightarrow 0$  et  $u_{h_n, \nu_n}$ , solutions du problème approché (3.106) ?

Cette question reste encore ouverte, mais un tel exemple doit être difficile à trouver, car il faut chercher cet exemple dans des familles de maillage non uniformes et non symétriques (voir le chapitre suivant).

## Maillage symétrique

**Lemme 6.3** *Soit  $(\mathcal{T}_h)_h$  une famille de maillage symétrique, c'est-à-dire que pour tout  $i \in J_0$  on a la situation de la figure 6.  $G_+$  et  $G_-$  sont les centres de gravité des deux triangles ayant  $Q_i$  en commun. On pose  $a = \overrightarrow{Q_i A}$ ,  $u = \overrightarrow{Q_i G_+}$ ,  $S_i = [AB]$ . Soit  $0 \neq \tilde{\beta}_t \in \mathbb{R}^2$  un vecteur quelconque, tel que  $\|\tilde{\beta}_t\| = \|\beta\|$ . Le vecteur  $\beta^T$  est l'image de  $\beta$  par la rotation d'angle  $\pi/2$ . Alors*

$$\sum_{j \in S^+(i)} \theta_{i,j}^\beta [\tilde{\beta}_t \cdot (Q_j - Q_i)] = -(u \cdot \beta^T) (\tilde{\beta}_t \cdot a) + 3(a \cdot \beta^T) (\tilde{\beta}_t \cdot u) \quad (6.131)$$

**Preuve :** On a (voir figure 6) :

$$\begin{aligned} v_1 = \overrightarrow{G_+ A} = -u + a, \quad v_2 = \overrightarrow{B G_+} = u + a, \quad v_3 = \overrightarrow{G_- B} = u - a = -v_1, \\ v_4 = \overrightarrow{A G_-} = -u - a = -v_2 \text{ et } m_1 = \overrightarrow{Q_i P_1} = \frac{1}{2}a + \frac{3}{2}u, \quad m_2 = \overrightarrow{Q_i P_2} = -\frac{1}{2}a + \frac{3}{2}u, \\ m_3 = \overrightarrow{Q_i P_3} = -\frac{1}{2}a - \frac{3}{2}u = -m_1, \quad m_4 = \overrightarrow{Q_i P_4} = \frac{1}{2}a - \frac{3}{2}u = -m_2. \end{aligned}$$

22. Voir par exemple la figure 5.

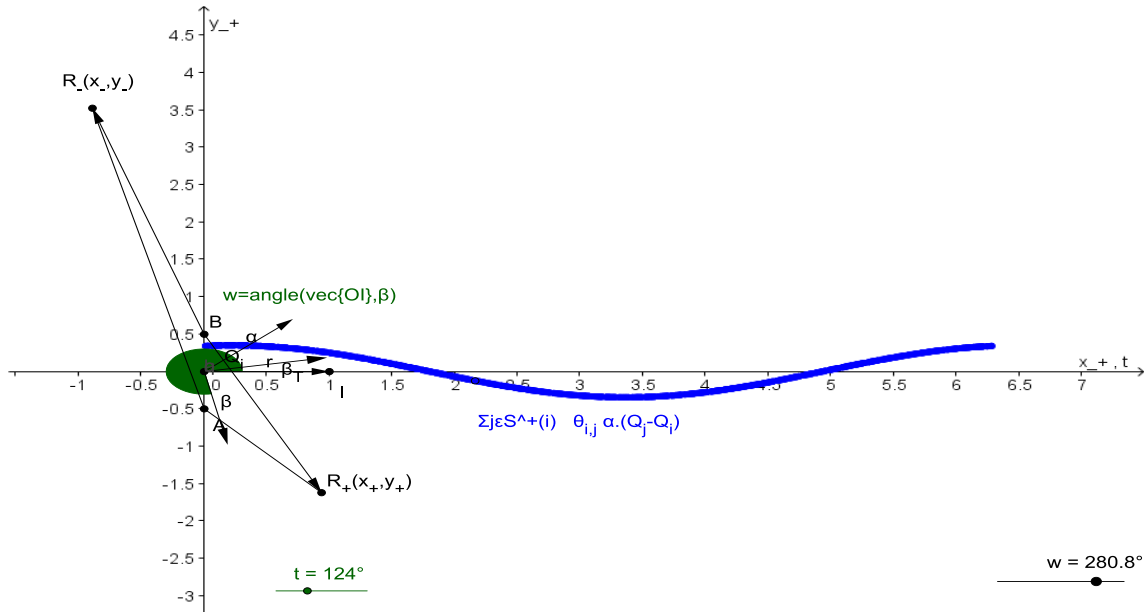


FIGURE 5 – Représentation graphique de  $t \mapsto \sum_{j \in S^+(i)} \theta_{i,j}^\beta [\alpha_t \cdot (Q_j - Q_i)]$  avec  $t = (\widehat{\beta}, \widehat{\alpha_t})$ ,  $\|\beta\| = \|\alpha\| = 1$  et  $|S_i| = 1$ .

Donc (avec  $S(i) = \{j_1, \dots, j_4\}$  indices des points  $Q_{j_1} = P_1, \dots, Q_{j_4} = P_4$ ) :

$$\theta_{i,j_1}^\beta (\tilde{\beta}_t \cdot (Q_{j_1} - Q_i)) = (v_1 \cdot \beta^T) (\tilde{\beta}_t \cdot m_1) = (v_3 \cdot \beta^T) (\tilde{\beta}_t \cdot m_3) = \theta_{i,j_2}^\beta (\tilde{\beta}_t \cdot (Q_{j_2} - Q_i))$$

$$\theta_{i,j_2}^\beta (\tilde{\beta}_t \cdot (Q_{j_2} - Q_i)) = (v_2 \cdot \beta^T) (\tilde{\beta}_t \cdot m_2) = (v_4 \cdot \beta^T) (\tilde{\beta}_t \cdot m_4) = \theta_{i,j_4}^\beta (\tilde{\beta}_t \cdot (Q_{j_4} - Q_i))$$

Par symétrie par rapport à  $Q_i$  on a :  $j_1 \in S^+ \iff j_3 \in S^-$ ,  $j_2 \in S^+ \iff j_4 \in S^-$ ,  $j_3 \in S^+ \iff j_1 \in S^-$ ,  $j_4 \in S^+ \iff j_2 \in S^-$ .

Ainsi  $\sum_{j \in S^+(i)} \theta_{i,j}^\beta (\tilde{\beta}_t \cdot (Q_j - Q_i)) = \sum_{j \in S^-(i)} \theta_{i,j}^\beta (\tilde{\beta}_t \cdot (Q_j - Q_i))$ .

Avec les notations définies ci-dessus et en calculant un peu, on obtient :

$$\begin{aligned} \sum_{j \in S^+(i)} \theta_{i,j}^\beta (\tilde{\beta}_t \cdot (Q_j - Q_i)) &= \frac{1}{2} \sum_{j \in S(i)} \theta_{i,j}^\beta (\tilde{\beta}_t \cdot (Q_j - Q_i)) \\ &= \frac{1}{2} \sum_{j \in S(i)} (v_j \cdot \beta^T) (\tilde{\beta}_t \cdot m_j) \\ &= -(u \cdot \beta^T) (\tilde{\beta}_t \cdot a) + 3 (a \cdot \beta^T) (\tilde{\beta}_t \cdot u) \end{aligned}$$

□

### Application à un maillage symétrique et uniforme

L'application de ce lemme au maillage uniforme de la figure 3 se fait en posant  $\kappa = h_b/h_a$ . Sans restriction soit  $\|\beta\| = 1$  et  $\|\tilde{\beta}_t\| = 1$ . Notons  $\beta_w = \beta = \begin{pmatrix} \cos(w) \\ \sin(w) \end{pmatrix}$  et  $\beta_{w+t} = \tilde{\beta}_t =$



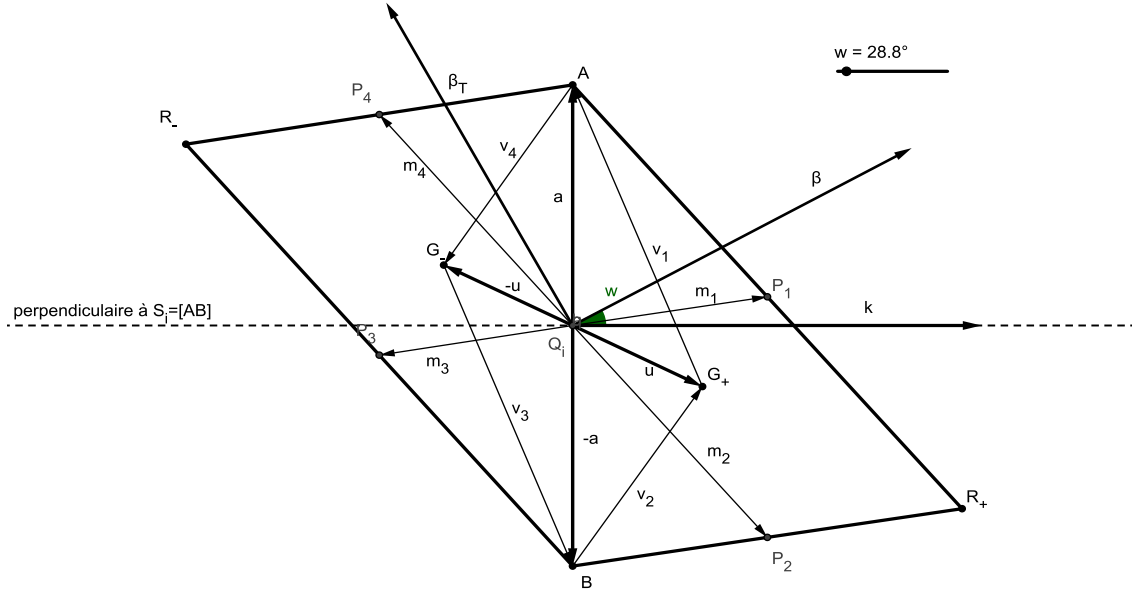


FIGURE 6 –  $G_+$  et  $G_-$  sont les centres de gravité des deux triangles ayant  $Q_i$  en commun.  
 $a = \vec{Q_i A}$ ,  $u = \vec{Q_i G_+}$ ,  $S_i = [AB]$

$\left(\frac{\cos(w+t)}{\sin(w+t)}\right)$ . Pour les trois cas qui se présentent on a :

$$\begin{aligned} \text{Cas I : } & a = \begin{pmatrix} -h_a/2 \\ 0 \end{pmatrix}, u = \begin{pmatrix} -h_a/6 \\ h_b/3 \end{pmatrix} \\ \text{Cas II : } & a = \begin{pmatrix} 0 \\ h_b/2 \end{pmatrix}, u = \begin{pmatrix} h_a/3 \\ -h_b/6 \end{pmatrix} \\ \text{Cas III : } & a = \begin{pmatrix} -h_a/2 \\ h_b/2 \end{pmatrix}, u = \begin{pmatrix} h_a/6 \\ h_b/6 \end{pmatrix} \end{aligned}$$

Avec le lemme ci-dessus, on obtient

$$\begin{aligned} \frac{1}{\|\beta\|^2} \eta_{\text{cas I}}(\beta_w, \beta_{w+t}) &= -\frac{h_a^2}{6} \sin(w) \cos(w+t) + \frac{h_a h_b}{6} [\cos(w) \cos(w+t) + 3 \sin(w) \sin(w+t)] \\ \frac{1}{\|\beta\|^2} \eta_{\text{cas II}}(\beta_w, \beta_{w+t}) &= -\frac{h_b^2}{6} \cos(w) \sin(w+t) + \frac{h_a h_b}{6} [\sin(w) \sin(w+t) + 3 \cos(w) \cos(w+t)] \\ \frac{1}{\|\beta\|^2} \eta_{\text{cas III}}(\beta_w, \beta_{w+t}) &= \frac{h_a^2}{6} \sin(w) \cos(w+t) + \frac{h_b^2}{6} \cos(w) \sin(w+t) + \frac{h_a h_b}{3} \cos(t) \\ \text{avec } \eta_i(\beta_w, \beta_{w+t}) &= \sum_{j \in S^+(i)} \theta_{i,j}^{\beta_w} [\beta_{w+t} \cdot (Q_j - Q_i)] \end{aligned}$$

Ainsi

$$\begin{aligned} \frac{\eta_{\text{cas I}}(\beta_w, \beta_{w+t})}{\|\beta_w\|^2 |S_I|^2} &= -\frac{1}{6} \sin(w) \cos(w+t) + \frac{\kappa}{6} [\cos(w) \cos(w+t) + 3 \sin(w) \sin(w+t)] \\ \frac{\eta_{\text{cas II}}(\beta_w, \beta_{w+t})}{\|\beta_w\|^2 |S_{II}|^2} &= -\frac{1}{6\kappa} \cos(w) \sin(w+t) + \frac{\kappa}{6} [\sin(w) \sin(w+t) + 3 \cos(w) \cos(w+t)] \\ \frac{\eta_{\text{cas III}}(\beta_w, \beta_{w+t})}{\|\beta_w\|^2 |S_{III}|^2} &= \frac{1}{6(1+\kappa^2)} \sin(w) \cos(w+t) + \frac{\kappa^2}{6(1+\kappa^2)} \cos(w) \sin(w+t) + \frac{\kappa}{3(1+\kappa^2)} \cos(t) \end{aligned}$$

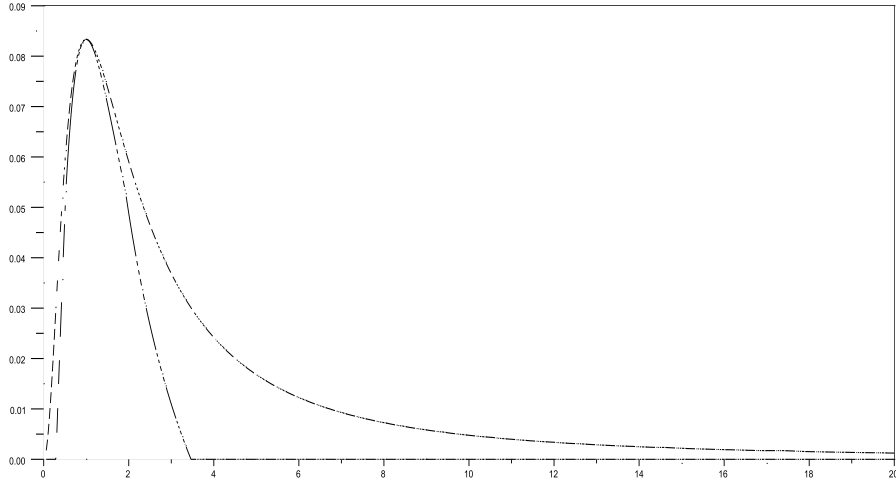


FIGURE 7 – Maillage uniforme : En trait continu, est représenté graphiquement, en fonction de  $\kappa = h_b/h_a$ , la valeur minimale possible pour  $\delta$  pour la condition  $\delta - S^+$ . C'est la fonction  $\kappa \mapsto \min_{w \in [0; 2\pi]} \left( \min \left\{ \frac{\eta_{\text{cas I}}(\beta_w, \beta_{w+0})}{\|\beta_w\|^2 |S_{\text{II}}|^2}, \frac{\eta_{\text{cas II}}(\beta_w, \beta_{w+0})}{\|\beta_w\|^2 |S_{\text{II}}|^2}, \frac{\eta_{\text{cas III}}(\beta_w, \beta_{w+0})}{\|\beta_w\|^2 |S_{\text{III}}|^2} \right\} \right)$ . Cette fonction est négative en dehors de  $[\frac{1}{2\sqrt{3}}; 2\sqrt{3}]$ . En pointillés est représenté la valeur minimale pour  $\delta$  pour la condition  $\delta - S^+$ -faible. C'est la fonction  $\kappa \mapsto \min_{w \in [0; 2\pi]} \left[ \max_{t \in [0; 2\pi]} \left( \min \left\{ \frac{\eta_{\text{cas I}}(\beta_w, \beta_{w+t})}{\|\beta_w\|^2 |S_{\text{II}}|^2}, \frac{\eta_{\text{cas II}}(\beta_w, \beta_{w+t})}{\|\beta_w\|^2 |S_{\text{II}}|^2}, \frac{\eta_{\text{cas III}}(\beta_w, \beta_{w+t})}{\|\beta_w\|^2 |S_{\text{III}}|^2} \right\} \right) \right]$

Sur la figure 7 est représentée la fonction

$$\kappa \mapsto \min_{w \in [0; 2\pi]} \left[ \max_{t \in [0; 2\pi]} \left( \min \left\{ \frac{\eta_{\text{cas I}}(\beta_w, \beta_{w+t})}{\|\beta_w\|^2 |S_{\text{II}}|^2}, \frac{\eta_{\text{cas II}}(\beta_w, \beta_{w+t})}{\|\beta_w\|^2 |S_{\text{II}}|^2}, \frac{\eta_{\text{cas III}}(\beta_w, \beta_{w+t})}{\|\beta_w\|^2 |S_{\text{III}}|^2} \right\} \right) \right]$$

Des évaluations numériques de cette fonction pour  $\kappa \in [\frac{1}{20}; 20]$ , montrent qu'elle peut être minorée par  $\delta_{[\frac{1}{20}; 20]} > 0$ . Ainsi

$\forall \kappa \in [\frac{1}{20}; 20] \quad \forall w \in [0; 2\pi] \quad \exists t \in [0; 2\pi]$  tel que

$$\delta_\kappa := \min \left\{ \frac{\eta_{\text{cas I}}(\beta_w, \beta_{w+t})}{\|\beta_w\|^2 |S_{\text{II}}|^2}, \frac{\eta_{\text{cas II}}(\beta_w, \beta_{w+t})}{\|\beta_w\|^2 |S_{\text{II}}|^2}, \frac{\eta_{\text{cas III}}(\beta_w, \beta_{w+t})}{\|\beta_w\|^2 |S_{\text{III}}|^2} \right\} \geq \delta_{[\frac{1}{20}; 20]} > 0$$

**Conclusion :** En comparaison avec la condition  $\delta - S^+$ , la figure 7 montre que la condition  $\delta - S^+$ -faible élargit considérablement les familles de maillage uniforme possibles. On note cependant que  $\delta_\kappa$  décroît rapidement vers 0 quand  $\kappa$  croît. Une preuve montrant que  $\delta_\kappa > 0$  pour tout  $\kappa > 0$  manque encore.

Avant de procéder à quelques simulations numériques, quelques remarques qui seront utiles pour la suite :

- a) Pour une valeur de  $h = \sqrt{h_a^2 + h_b^2}$  donnée, les maillages obtenus pour  $\kappa = h_b/h_a$  et  $\tilde{\kappa} = \tilde{h}_b/\tilde{h}_a$  (avec  $\tilde{h}_a = h_b$  et  $\tilde{h}_b = h_a$ ) sont symétriques. En posant  $\tilde{w} = 3\pi/2 - w$  et  $\tilde{t} = -t$ , il est possible de montrer que  $\delta_\kappa = \delta_{\frac{1}{\kappa}}$ . L'étude de  $\delta_\kappa$  se restreint donc à  $]0; 1]$  ou  $[1; +\infty[$ .

b) On constate, par un simple calcul, que

$$\frac{1}{\|\beta\|^2} [\eta_{\text{cas I}}(\beta_w, \beta_{w+t}) + \eta_{\text{cas II}}(\beta_w, \beta_{w+t}) + \eta_{\text{cas III}}(\beta_w, \beta_{w+t})] = h_a h_b \cos(t) \quad (6.132)$$

Ce résultat ne dépend plus de  $w$  et donne une condition nécessaire pour  $\delta_\kappa > 0$ . Elle permet aussi de restreindre l'intervalle de recherche pour  $t$  à  $] -\pi/2; \pi/2[$ .

$$\text{Si } \delta_\kappa > 0 \text{ alors } t \in ] -\pi/2; \pi/2[$$

### 6.3 Simulations numériques

L'objectif des simulations numériques était de constater la sensibilité de  $\|u_{h,\nu}\|_{L^2(\Omega)}$  à

$$\delta_{\kappa,w} = \max_{t \in [0; 2\pi]} \left( \min \left\{ \frac{\eta_{\text{cas I}}(\beta_w, \beta_{w+t})}{\|\beta_w\|^2 |S_I|^2}, \frac{\eta_{\text{cas II}}(\beta_w, \beta_{w+t})}{\|\beta_w\|^2 |S_{II}|^2}, \frac{\eta_{\text{cas III}}(\beta_w, \beta_{w+t})}{\|\beta_w\|^2 |S_{III}|^2} \right\} \right) \quad (6.133)$$

(valeur optimale dans la condition  $\delta - S_\beta^+$ -faible) suggérée par l'inégalité (5.113).

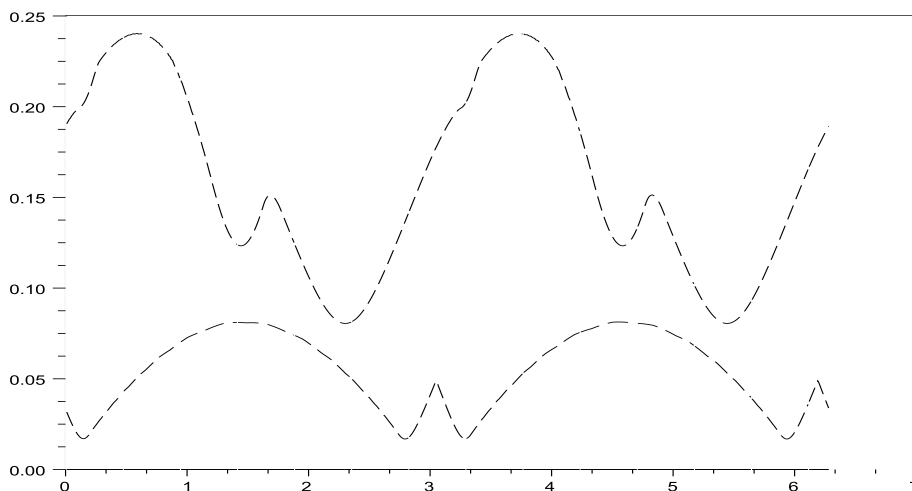
Dans toutes les simulation numériques<sup>23</sup> qui suivent :  $\Omega = [0; 1] \times [0; 1]$ , et  $f_r : \Omega \rightarrow \mathbb{R}$  est définie par  $f_r(x, y) = \frac{1}{\sqrt{\pi \ln(1 + \frac{1}{4r})}} \frac{1}{\sqrt{r + (x-0,5)^2 + (y-0,5)^2}}$ .

On a  $\|f\|_{L^2(\Omega)} \leq 1$  pour tout  $r > 0$  et  $\lim_{r \rightarrow 0^+} f(1/2, 1/2) = +\infty$ .

Les deux maillages uniformes utilisés pour  $\Omega$  sont :

- a)  $N_1 = 31$  et  $M_1 = 25$ . Donc  $h_1 \approx 0.0513866$  et  $\kappa_1 = 31/25$ .
- b)  $N_2 = 20$  et  $M_2 = 100$ . Donc  $h_2 \approx 0.0509902$  et  $\kappa_2 = 1/5$ .

La figure ci-dessous représente respectivement  $\delta_{\kappa_1,w}$  (courbe du haut) et  $\delta_{\kappa_2,w}$  (courbe du bas) en fonction de  $w$  :



Pour  $r = 10^{-6}$  et  $\nu = 2^{-10}$  on obtient le tableau suivant, où  $t_{\kappa_1, w}$  et  $t_{\kappa_2, w}$  sont les valeurs optimales pour  $t$  dans (6.133) pour les deux maillages considérés.

$w$ (en radians)	$\delta_{\kappa_1, w}$	$t_{\kappa_1, w}$ (en degrés)	$\ u_{h, w}\ _2$	$\delta_{\kappa_2, w}$	$t_{\kappa_2, w}$ (en degrés)	$\ u_{h, w}\ _2$
0.1	0.1986915	14	0.4333798	0.0193089	25	0.3829098
0.2	0.2094766	24	0.4343529	0.0198605	77	0.3841150
0.3	0.2269816	21	0.4380228	0.0288026	70	0.3804997
0.4	0.2344583	14	0.4441043	0.0370890	64	0.3784233
0.5	0.2388066	6	0.4520128	0.0444809	57	0.3762041
0.6	0.2400414	0	0.4606768	0.0515559	51	0.3723050
0.7	0.2381148	-8	0.4684794	0.0575345	45	0.3661238
0.8	0.2325566	-15	0.4736409	0.0632921	39	0.3581320
0.9	0.2220726	-24	0.4747762	0.0679584	33	0.3493205
1.0	0.2055112	-30	0.4682596	0.0726142	27	0.3406635
1.1	0.1843315	-37	0.4607783	0.0755073	21	0.3328901
1.2	0.1598694	-44	0.4541875	0.0781941	15	0.3264717
1.3	0.1372839	-36	0.4496402	0.0804763	9	0.3216875
1.4	0.1246559	-23	0.4476915	0.0810494	3	0.3186968
1.5	0.1255645	-9	0.4485566	0.0808779	-2	0.3175933

On constate que les valeurs de  $\|u_{h, w}\|_2$  restent relativement peu sensibles aux variations de  $\delta_{\kappa, w}$ , qui, elles, sont assez fortes. Une explication possible de ce phénomène utilise (6.132) et pourrait être la suivante :

La minoration vers  $\|u_{h, w}\|_2$  dans la preuve de stabilité s'effectue pour (5.116). Le terme à minorer est

$$\begin{aligned} & \sum_{i \in J_0} \sum_{j \in S^+(i)} \theta_{i,j}^\beta v_h(Q_i)^2 \tilde{\beta} \cdot (Q_j - Q_i) \\ &= \sum_{T \in \mathcal{T}_h} \left[ \eta_{\text{cas I}}(\beta_w, \beta_{w+t}) v_h(Q_{\text{I}}^T)^2 + \eta_{\text{cas II}}(\beta_w, \beta_{w+t}) v_h(Q_{\text{II}}^T)^2 \right. \\ & \quad \left. + \eta_{\text{cas III}}(\beta_w, \beta_{w+t}) v_h(Q_{\text{III}}^T)^2 \right] \end{aligned}$$

On a le résultat suivant :

**Lemme 6.4** *Dans les conditions du maillage uniforme décrites ci-dessus, on a*

$$\forall v_h \in V_h \quad \sum_{i \in J_0} \sum_{j \in S^+(i)} \theta_{i,j}^\beta v_h(Q_i)^2 \tilde{\beta} \cdot (Q_j - Q_i) \geq (\|v_h\|_2^2 - 2h^2 \|\nabla v_h\|_2^2) \cos(t) \quad (6.134)$$

**Preuve :** Pour tout triangle  $T \in \mathcal{T}_h$ , les milieux des côtés sont les points  $Q_{\text{I}}^T$ ,  $Q_{\text{II}}^T$  et  $Q_{\text{III}}^T$  qui correspondent respectivement aux cas I, II et III.

Pour chaque triangle  $T \in \mathcal{T}_h$ , il existe  $\bar{Q}^T \in T$ , tel que  $v_h(\bar{Q}^T)^2 |T| = \|v_h\|_{L^2(T)}^2$ .

En effet : il existe  $\tilde{Q}^T \in [Q_I^T; Q_{II}^T]$  tel que  $v_h(\tilde{Q}^T)^2 = \frac{v_h(Q_I^T)^2 + v_h(Q_{II}^T)^2}{2}$ .

Comme  $\min\{v_h(\tilde{Q}^T)^2, v_h(Q_{III}^T)^2\} \leq \frac{v_h(Q_{III}^T)^2 + 2v_h(\tilde{Q}^T)^2}{3} \leq \max\{v_h(\tilde{Q}^T)^2, v_h(Q_{III}^T)^2\}$ ,  
alors il existe  $\bar{Q}^T \in [\tilde{Q}^T, Q_{III}^T]$  avec  $v_h(\bar{Q}^T)^2 = \frac{v_h(Q_I^T)^2 + v_h(Q_{II}^T)^2 + v_h(Q_{III}^T)^2}{3}$ .

La preuve est faite en appliquant le lemme 2.10.

Notons maintenant  $\eta_I^T = \eta_{\text{cas I}}(\beta_w, \beta_{w+t})$ ,  $\eta_{II}^T = \eta_{\text{cas II}}(\beta_w, \beta_{w+t})$ ,  $\eta_{III}^T = \eta_{\text{cas III}}(\beta_w, \beta_{w+t})$   
et  $v_{h|T}(Q) = v_{h|T}(\bar{Q}) + \nabla v_{h|T} \cdot (Q - \bar{Q})$  pour  $Q \in T$ .

Alors

$$\begin{aligned}
& \sum_{i \in J_0} \sum_{j \in S^+(i)} \theta_{i,j}^\beta v_h(Q_i)^2 \tilde{\beta} \cdot (Q_j - Q_i) \\
= & \sum_{T \in \mathcal{T}_h} \left[ \eta_I^T \left\{ v_{h|T}(\bar{Q}^T) + \nabla v_{h|T} \cdot (Q_I^T - \bar{Q}^T) \right\}^2 + \eta_{II}^T \left\{ v_{h|T}(\bar{Q}^T) + \nabla v_{h|T} \cdot (Q_{II}^T - \bar{Q}^T) \right\}^2 \right. \\
& \left. + \eta_{III}^T \left\{ v_{h|T}(\bar{Q}^T) + \nabla v_{h|T} \cdot (Q_{III}^T - \bar{Q}^T) \right\}^2 \right] \\
& (q+r)^2 \geq \frac{1}{2}q^2 - r^2, \text{ et } h \geq \|Q - \bar{Q}\| \text{ pour } Q \in T, \text{ donc} \\
\geq & \sum_{T \in \mathcal{T}_h} \left[ \frac{1}{2}(\eta_I^T + \eta_{II}^T + \eta_{III}^T) v_{h|T}(\bar{Q})^2 - (\eta_I^T + \eta_{II}^T + \eta_{III}^T) h^2 \|\nabla v_{h|T}\|^2 \right] \\
\stackrel{(6.132)}{\geq} & \sum_{T \in \mathcal{T}_h} 2|T| \cos(t) \left[ \frac{1}{2} v_{h|T}(\bar{Q})^2 - h^2 \|\nabla v_{h|T}\|^2 \right] \\
\geq & (\|v_h\|_2^2 - 2h^2 \|\nabla v_h\|_2^2) \cos(t)
\end{aligned}$$

□

Pour un maillage quasi-uniforme, comme c'est le cas ici, le théorème des inégalités inverses s'applique, et il existe une constante  $c_{\text{inv}} > 0$ , indépendante de  $h$ , telle que

$$\forall h > 0 \quad \forall v_h \in V_h \quad \|\nabla v_h\|_2 \leq c_{\text{inv}} \frac{1}{h} \|v_h\|_2 \iff \frac{h \|\nabla u_{h,\nu}\|_2}{\|u_{h,\nu}\|_2} \leq c_{\text{inv}}$$

Cette inégalité laisse peu de place pour un exemple de perte de stabilité pour le problème approché, tel qu'il est décrit au b) de la remarque 6.4 page 62.

En effet, si on accepte -un instant- ce que les simulations numériques semblent suggérer, c'est-à-dire, pour la solution  $u_{h,\nu}$ , l'existence de  $\epsilon > 0$  et  $C > 0$  tels que

$$\forall h > 0 \quad \forall \nu > 0 \quad \frac{h^{1-\epsilon} \|\nabla u_{h,\nu}\|_2}{\|u_{h,\nu}\|_2} \leq C, \tag{6.135}$$

alors, sous l'hypothèse (6.135), on aurait

$$\sum_{i \in J_0} \sum_{j \in S^+(i)} \theta_{i,j}^\beta u_{h,\nu}(Q_i)^2 \tilde{\beta} \cdot (Q_j - Q_i) \geq (1 - 2C^2 h^{2\epsilon}) \cos(t) \|u_{h,\nu}\|_2^2 \tag{6.136}$$

Ainsi, pour  $h$  suffisamment petit, on aurait en (5.116) une minoration insensible aux valeurs proches de 0 de  $\delta_{\kappa,w}$ .

Pour conclure ce chapitre, quelques simulations numériques où  $r = 10^{-6}$ ,  $w = 0.2$  et  $\kappa = \kappa_2 = 1/5$ .<sup>24</sup>

$\nu$		$h_0 \approx 0.2039608$	$h_0/2$	$h_0/4$
$\nu = 2^{-4}$	$\ \nabla u_{h,\nu}\ _2 \approx$	4.0526168	1.439616	1.526972
	$\ u_{h,\nu}\ _2 \approx$	0.5909599	0.2618141	0.2687291
	$\frac{h\ \nabla u_{h,\nu}\ _2}{\ u_{h,\nu}\ _2} \approx$	1.3986988	0.5607513	0.2897364
$\nu = 2^{-8}$	$\ \nabla u_{h,\nu}\ _2 \approx$	14.599702	3.1833264	3.9200821
	$\ u_{h,\nu}\ _2 \approx$	0.9695239	0.3694875	0.3737188
	$\frac{h\ \nabla u_{h,\nu}\ _2}{\ u_{h,\nu}\ _2} \approx$	3.0713699	0.8786139	0.5348560
$\nu = 2^{-12}$	$\ \nabla u_{h,\nu}\ _2 \approx$	22.677829	4.3912254	5.5017487
	$\ u_{h,\nu}\ _2 \approx$	1.0564197	0.3838080	0.3876843
	$\frac{h\ \nabla u_{h,\nu}\ _2}{\ u_{h,\nu}\ _2} \approx$	4.3783618	1.1667782	0.7236178
$\nu = 2^{-16}$	$\ \nabla u_{h,\nu}\ _2 \approx$	23.815766	4.7428655	6.1619076
	$\ u_{h,\nu}\ _2 \approx$	1.062731	0.3853672	0.3891689
	$\frac{h\ \nabla u_{h,\nu}\ _2}{\ u_{h,\nu}\ _2} \approx$	4.570754	1.2551127	0.8073535
$\nu = 2^{-20}$	$\ \nabla u_{h,\nu}\ _2 \approx$	23.893451	4.7706812	6.2248814
	$\ u_{h,\nu}\ _2 \approx$	1.0631275	0.3854766	0.3892777
	$\frac{h\ \nabla u_{h,\nu}\ _2}{\ u_{h,\nu}\ _2} \approx$	4.5839535	1.2621155	0.8153765

### Conclusions :

- 1) La condition (6.132) permet de contrôler la répartition des valeurs de  $Q_i \mapsto \sum_{j \in S^+(i)} \theta_{i,j}^\beta (\beta_t \cdot (Q_j - Q_i))$  dans le maillage en donnant une interprétation géométrique (comme aire) à ces valeurs. Cette condition trouve une généralisation à des maillages non uniformes (voir les chapitres suivants, en particulier le lemme C.1) et la remarque 14.10, page 122).
- 2) L'hypothèse (6.135) faite pour la solution  $u_{h,\nu}$ , pourrait être une propriété qu'il serait intéressant de pouvoir prouver. On aurait alors un résultat plus en accord avec les simulations numériques, qui montrent que  $\|u_{h,\nu}\|_2$  reste stable même pour des valeurs de  $\delta_{\kappa,w}$  très proches de 0.
- 3) Des remarques précédentes, il semblerait que certaines structures, qui jouent un rôle dans la stabilité, sont encore à découvrir. Si cela se confirme, la recherche d'un exemple tel que décrit au b) de la remarque 6.4, s'annoncerait difficile.

24. Les capacités de calculs disponibles ne permettent pas d'atteindre le régime asymptotique nécessaire pour une estimation de  $C$  et  $\epsilon$ . La dernière ligne du tableau ci-dessus donne comme valeurs (très approximatives)  $C \approx 12$  et  $\epsilon \approx 0,8$ .

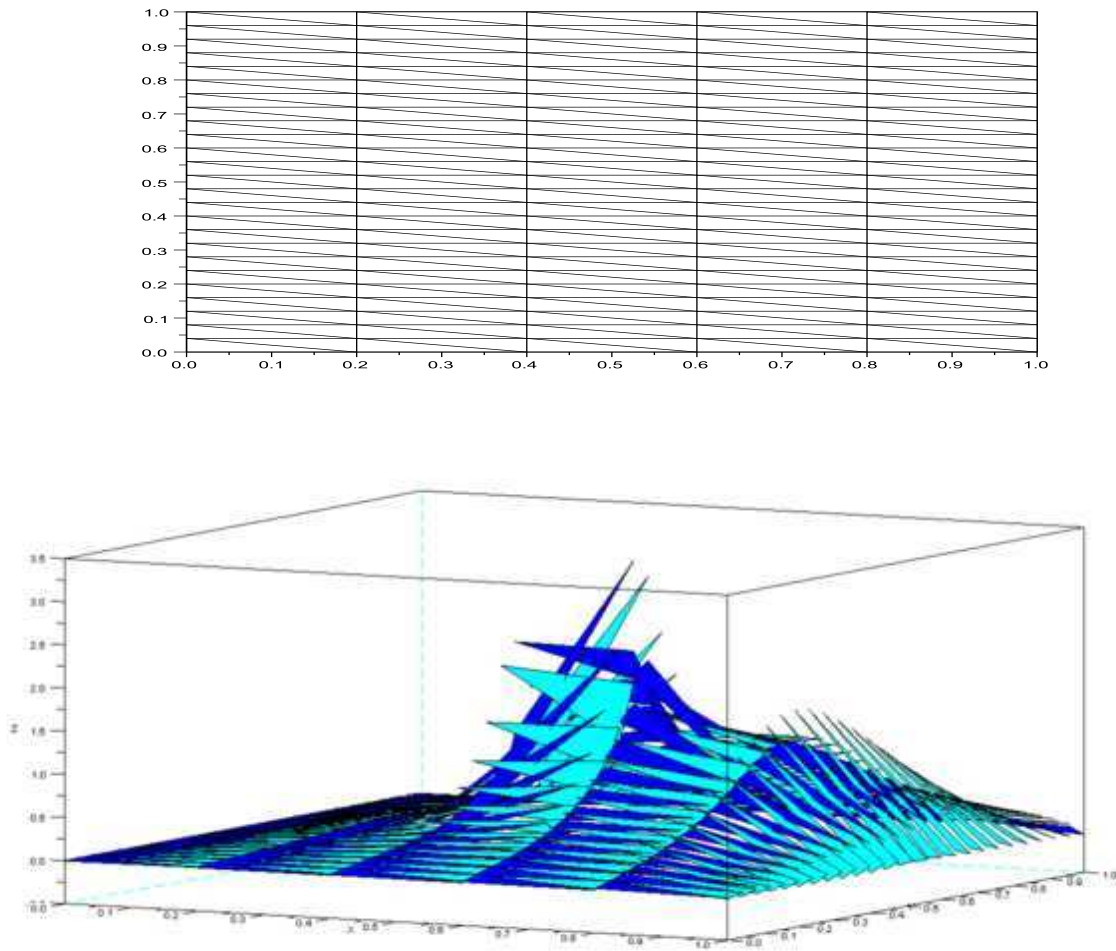


FIGURE 8 – Représentation graphique du maillage et de  $u_{h,\nu}$  pour  $N = 5$ ,  $M = 25$ ,  $\nu = 2^{-4}$ ,  $w = 0,2$  ( $\beta = (\cos(w), \sin(w))$ ),  $r = 10^{-6}$

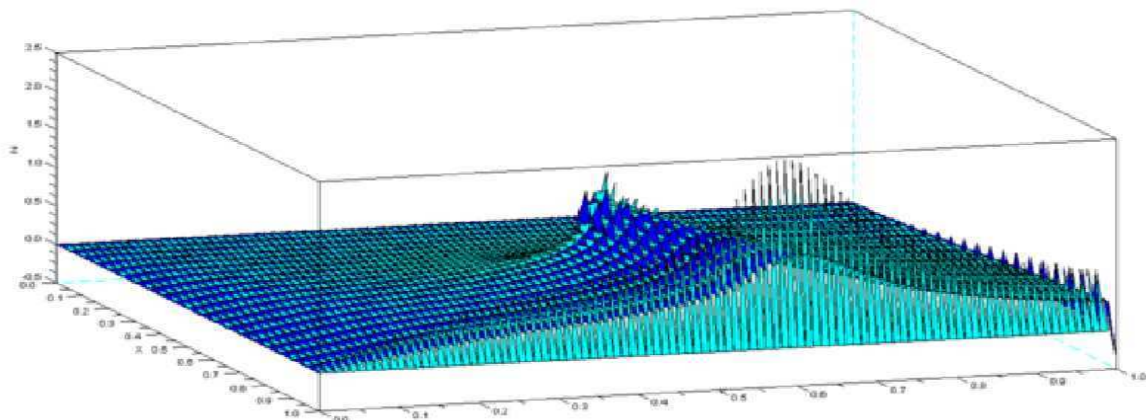


FIGURE 9 – Représentation graphique de  $u_{h,\nu}$  pour  $N = 20$ ,  $M = 100$ ,  $\nu = 2^{-20}$ ,  $w = 0,2$  ( $\beta = (\cos(w), \sin(w))$ ),  $r = 10^{-6}$

Troisième partie

# Stabilité de l'équation d'advection-diffusion non stationnaire

## Sommaire

---

7	Introduction	72
8	Stabilité du problème approché	76

---



## 7 Introduction

Soit  $0 \neq \beta \in \mathbb{R}^2$ ,  $\Omega$  un domaine polygonal borné et  $f \in L^2(\Omega)$ . Au chapitre 1.3, on a introduit la formulation faible du problème d'advection-diffusion non stationnaire (1.8) :

**Définition 7.1** Une fonction  $u$  est une solution faible de (1.8), si et seulement si :

$$\begin{aligned} a) \quad & u \in L^2(]0, T[, V) \cap L^\infty(Q_T) & (7.137) \\ b) \quad & \frac{d}{dt}(u(t), v) + \nu((u, v)) + b(u, v) = (g(t), v), \quad \forall v \in V \\ & \text{pris dans le sens des distributions sur } ]0, T[ \\ c) \quad & u(0) = u^0 \end{aligned}$$

où

$$b(u, v) = \int_{\Omega} \beta \cdot \nabla u v \, dx \quad \text{pour } u, v \in V \quad (7.138)$$

Soit  $N \in \mathbb{N}$ , le pas pour  $t$  est défini par  $\tau = \frac{T}{N+1}$  et on pose  $g_h^k(x) = I_h(g(\cdot, \tau k))(x)$  pour  $k \in \{0, \dots, N+1\}$ .

Le problème approché pour (7.137) est :

**Définition 7.2** Formulation faible du problème approché d'advection-diffusion non stationnaire

Trouver  $u_h^k \in V_h$  pour  $k \in \{1, \dots, N+1\}$  tels que  $u_h^0 = I_h(u^0)$  et

$$\frac{1}{\tau}(u_h^{k+1} - u_h^k, v_h) + \nu((u_h^{k+1}, v_h))_h + b_h(u_h^{k+1}, v_h) = (g_h^{k+1}, v_h) \quad \forall v_h \in V_h \quad \forall k \in \{1, \dots, N+1\} \quad (7.139)$$

La discrétisation est implicite, car les termes de diffusion et de convection sont pris en  $t = \tau(k+1)$ .

La preuve du résultat pour le problème approché est très proche de celle du problème continu discrétisé en  $t$  :

**Théorème 7.1** Soit  $N \in \mathbb{N}$ , le pas pour  $t$  est défini par  $\tau = \frac{T}{N+1}$  et on pose  $g^k(x) = g(x, \tau k)$  pour  $k \in \{0, \dots, N+1\}$  et  $u^1, u^2, \dots, u^{N+1} \in V$  définis par

$$\frac{1}{\tau}(u^{k+1} - u^k, v) + \nu((u^{k+1}, v)) + b(u^{k+1}, v) = (g^{k+1}, v) \quad \forall v \in V \quad \forall k \in \{1, \dots, N+1\} \quad (7.140)$$

Alors il existe une constante  $C > 0$  ne dépendant ni de  $\text{diam}(\Omega)$  et ni  $\|\beta\|$ , telle que

$$\max_{0 \leq k \leq N} \|u^{k+1}\|_2 + \sqrt{\nu} \left( \tau \sum_{l=0}^N \|u^{l+1}\|_{*V}^2 \right)^{1/2} \leq C \left[ \sqrt{\frac{\text{diam}(\Omega)}{\|\beta\|}} \left( \tau \sum_{l=0}^N \|g^{l+1}\|_2^2 \right)^{1/2} + \|u^0\|_2 \right] \quad (7.141)$$

Remarquons au passage que  $\max_{0 \leq k \leq N} \|u^{k+1}\|_2$ ,  $\left(\tau \sum_{l=0}^N \|u^{l+1}\|_{*V}^2\right)^{1/2}$  et  $\left(\tau \sum_{l=0}^N \|g^{l+1}\|_2^2\right)^{1/2}$  sont les variantes discrétisées dans le temps des normes respectivement des espaces  $L^\infty([0, T], L^2(\Omega))$ ,  $L^2([0, T], V)$  et  $L^2([0, T], L^2(\Omega))$ .

**Preuve :**

**Étape 1 :** Pour  $v = u^{k+1}$  dans (7.140) on a

$$\begin{aligned} \frac{1}{\tau}(u^{k+1} - u^k, u^{k+1}) + \nu \|u^{k+1}\|_{*V}^2 + \int_{\Omega} (\beta \cdot \nabla u^{k+1}) u^{k+1} dx &= (g^{k+1}, u^{k+1}) \\ \frac{1}{\tau}(u^{k+1} - u^k, u^{k+1}) + \nu \|u^{k+1}\|_{*V}^2 &= (g^{k+1}, u^{k+1}) \end{aligned}$$

On utilise ici :

$$\int_{\Omega} (\beta \cdot \nabla u^{k+1}) u^{k+1} dx = \frac{1}{2} \int_{\Omega} \nabla \cdot (\beta (u^{k+1})^2) dx = 0 \quad (7.142)$$

Ainsi (en remplaçant l'indice  $k$  par  $l$ )

$$\frac{1}{\tau}(u^{l+1}, u^{l+1}) + \nu \|u^{l+1}\|_{*V}^2 = (g^{l+1}, u^{l+1}) + \frac{1}{\tau}(u^l, u^{l+1})$$

Avec l'inégalité  $(u^l, u^{l+1}) \leq \frac{1}{2}\|u^l\|^2 + \frac{1}{2}\|u^{l+1}\|^2$ , simplification de  $\frac{1}{2}\|u^{l+1}\|^2$  et sommation sur  $l$  :

$$\frac{1}{2\tau} \sum_{l=0}^s \|u^{l+1}\|_2^2 + \nu \sum_{l=0}^s \|u^{l+1}\|_{*V}^2 \leq \sum_{l=0}^s (g^{l+1}, u^{l+1}) + \frac{1}{2\tau} \sum_{l=0}^s \|u^l\|_2^2$$

Multiplication des deux membres par  $\tau$ , soustraction de  $\frac{1}{2} \sum_{l=1}^s \|u^l\|_2^2$  et utilisation de l'inégalité  $(g^{l+1}, u^{l+1}) \leq \|g^{l+1}\| \|u^{l+1}\| \leq \frac{1}{4\kappa} \|g^{l+1}\|_2^2 + \kappa \|u^{l+1}\|_2^2$  et  $s = N$  :

$$\frac{1}{2} \|u^{N+1}\|_2^2 + \tau \nu \sum_{l=0}^N \|u^{l+1}\|_{*V}^2 \leq \frac{1}{4\kappa} \tau \sum_{l=0}^N \|g^{l+1}\|_2^2 + \kappa \tau \sum_{l=0}^N \|u^{l+1}\|_2^2 + \frac{1}{2} \|u^0\|_2^2$$

Ainsi

$$\tau \nu \sum_{l=0}^N \|u^{l+1}\|_{*V}^2 \leq \frac{1}{4\kappa} \tau \sum_{l=0}^N \|g^{l+1}\|_2^2 + \kappa \tau \sum_{l=0}^N \|u^{l+1}\|_2^2 + \frac{1}{2} \|u^0\|_2^2 \quad (7.143)$$

**Étape 2 :** Soit  $\varphi$  défini par  $\varphi(x) = R - \beta \cdot (x - x_0)$  avec  $x_0 \in \partial\Omega$  tel que  $\beta \cdot (x - x_0) \geq 0$  pour  $x \in \Omega$  et  $R > \|\beta\| \text{diam}(\Omega)$ , une constante qui sera déterminée plus tard. On a donc  $0 < R - \|\beta\| \text{diam}(\Omega) \leq \varphi(x) \leq R$ . Prenons  $v = u^{k+1}\varphi$  dans (7.140) :

$$\begin{aligned} \frac{1}{\tau}(u^{k+1} - u^k, u^{k+1}\varphi) + \nu \int_{\Omega} \nabla u^{k+1} \cdot [\nabla u^{k+1}\varphi - u^{k+1}\beta] dx + \\ b(u^{k+1}, u^{k+1}\varphi) &= (g^{k+1}, u^{k+1}\varphi) \\ \frac{1}{\tau}(u^{k+1} - u^k, u^{k+1}\varphi) + \nu \int_{\Omega} \|\nabla u^{k+1}\|^2 \varphi dx + b(u^{k+1}, u^{k+1}\varphi) &\stackrel{(7.142)}{=} (g^{k+1}, u^{k+1}\varphi). \end{aligned}$$

Maintenant

$$0 = \int_{\Omega} \nabla \cdot (\beta (u^{k+1})^2 \varphi) dx = 2 \int_{\Omega} (\beta \cdot \nabla u^{k+1}) u^{k+1} \varphi dx - \int_{\Omega} \|\beta\|^2 (u^{k+1})^2 dx$$

Ainsi

$$b(u^{k+1}, u^{k+1}\varphi) = \int_{\Omega} (\beta \cdot \nabla u^{k+1}) u^{k+1}\varphi \, dx = \frac{1}{2} \|\beta\|^2 \int_{\Omega} (u^{k+1})^2 \, dx = \frac{1}{2} \|\beta\|^2 \|u^{k+1}\|_2^2 \quad (7.144)$$

$$\begin{aligned} \frac{1}{\tau}(u^{k+1}, u^{k+1}\varphi) + \frac{1}{2} \|\beta\|^2 \|u^{k+1}\|_2^2 &= (g^{k+1}, u^{k+1}\varphi) - \nu \int_{\Omega} \|\nabla u^{k+1}\|^2 \varphi \, dx + \frac{1}{\tau}(u^k, u^{k+1}\varphi) \\ &\leq (g^{k+1}, u^{k+1}\varphi) + \nu \int_{\Omega} \|\nabla u^{k+1}\|^2 \varphi \, dx + \frac{1}{\tau}(u^k, u^{k+1}\varphi) \end{aligned} \quad (7.145)$$

Maintenant :

$$0 < (R - \|\beta\| \operatorname{diam}(\Omega))(w, w) \leq (w, w\varphi) \leq R(w, w) \quad \forall w \in V, \quad (7.146)$$

$$\begin{aligned} \text{On a aussi } (g^{k+1}, u^{k+1}\varphi) &= (g^{k+1}\sqrt{\varphi}, u^{k+1}\sqrt{\varphi}) \\ &\leq \|g^{k+1}\sqrt{\varphi}\|_2 \|u^{k+1}\sqrt{\varphi}\|_2 \\ &\leq (g^{k+1}, g^{k+1}\varphi)^{1/2} (u^{k+1}, u^{k+1}\varphi)^{1/2} \end{aligned} \quad (7.147)$$

$ab \leq \kappa a^2 + \frac{1}{4\kappa} b^2$  pour  $a, b, \kappa > 0$ , donc

$$\begin{aligned} (g^{k+1}, u^{k+1}\varphi) &\leq (g^{k+1}, g^{k+1}\varphi)^{1/2} (u^{k+1}, u^{k+1}\varphi)^{1/2} \\ &\leq \kappa (g^{k+1}, g^{k+1}\varphi) + \frac{1}{4\kappa} (u^{k+1}, u^{k+1}\varphi) \end{aligned}$$

De la même manière (avec  $\kappa = \frac{1}{2}$ ) :

$$(u^k, u^{k+1}\varphi) \leq \frac{1}{2}(u^k, u^k\varphi) + \frac{1}{2}(u^{k+1}, u^{k+1}\varphi)$$

En appliquant ces inégalités à (7.145), en soustrayant  $\frac{1}{2}(u^{k+1}, u^{k+1}\varphi)$  aux deux membres de l'inégalité et en remplaçant  $k$  par  $l$  on obtient

$$\begin{aligned} \frac{1}{2\tau}(u^{l+1}, u^{l+1}\varphi) + \frac{1}{2} \|\beta\|^2 \|u^{l+1}\|_2^2 \\ \leq \kappa (g^{l+1}, g^{l+1}\varphi) + \frac{1}{4\kappa} (u^{l+1}, u^{l+1}\varphi) + \frac{1}{2\tau} (u^l, u^l\varphi) + \nu \int_{\Omega} \|\nabla u^{l+1}\|^2 \varphi \, dx \end{aligned}$$

Sommation de 0 à  $s$  et soustraction de  $\sum_{l=1}^s \frac{1}{2\tau} (u^l, u^l\varphi)$  :

$$\begin{aligned} &\frac{1}{2\tau} (u^{s+1}, u^{s+1}\varphi) + \frac{1}{2} \|\beta\|^2 \sum_{l=0}^s \|u^{l+1}\|_2^2 \\ &\leq \kappa \sum_{l=0}^s (g^{l+1}, g^{l+1}\varphi) + \sum_{l=0}^s \frac{1}{4\kappa} (u^{l+1}, u^{l+1}\varphi) + \frac{1}{2\tau} (u^0, u^0\varphi) + \nu \sum_{l=0}^s \int_{\Omega} \|\nabla u^{l+1}\|^2 \varphi \, dx \\ &\leq \kappa \sum_{l=0}^s (g^{l+1}, g^{l+1}\varphi) + \frac{R}{4\kappa} \sum_{l=0}^s \|u^{l+1}\|_2^2 + \frac{1}{2\tau} (u^0, u^0\varphi) + \nu \sum_{l=0}^s \int_{\Omega} \|\nabla u^{l+1}\|^2 \varphi \, dx \end{aligned}$$

En choisissant  $\kappa = \frac{R}{\|\beta\|^2}$  et soustraction de  $\frac{1}{4} \|\beta\|^2 \sum_{l=0}^s \|u^{l+1}\|_2^2$  on a :

$$\begin{aligned} &\frac{1}{2\tau} (u^{s+1}, u^{s+1}\varphi) + \frac{1}{4} \|\beta\|^2 \sum_{l=0}^s \|u^{l+1}\|_2^2 \\ &\leq \frac{R}{\|\beta\|^2} \sum_{l=0}^s (g^{l+1}, g^{l+1}\varphi) + \frac{R}{2\tau} \|u^0\|_2^2 + \nu \sum_{l=0}^s \int_{\Omega} \|\nabla u^{l+1}\|^2 \varphi \, dx \end{aligned}$$

Les termes du membre gauche de l'inégalité étant positifs, on obtient après multiplication de l'inégalité par  $\tau$

$$\begin{aligned} \frac{1}{2} \max_{0 \leq k \leq N} (u^{k+1}, u^{k+1} \varphi) &\leq \frac{R}{\|\beta\|^2} \tau \sum_{l=0}^N (g^{l+1}, g^{l+1} \varphi) + \frac{R}{2} \|u^0\|_2^2 + \nu \tau \sum_{l=0}^N \int_{\Omega} \|\nabla u^{l+1}\|^2 \varphi \, dx \\ \frac{1}{4} \|\beta\|^2 \tau \sum_{l=0}^N \|u^{l+1}\|_2^2 &\leq \frac{R}{\|\beta\|^2} \tau \sum_{l=0}^N (g^{l+1}, g^{l+1} \varphi) + \frac{R}{2} \|u^0\|_2^2 + \nu \tau \sum_{l=0}^N \int_{\Omega} \|\nabla u^{l+1}\|^2 \varphi \, dx \end{aligned}$$

Il faut maintenant se débarrasser de  $\varphi$ . Avec (7.146) et  $\int_{\Omega} \|\nabla u^{l+1}\|^2 \varphi \, dx \leq R \|u^{l+1}\|_{*V}^2$  :

$$\max_{0 \leq k \leq N} \|u^{k+1}\|_2^2 \leq \left[ \frac{R}{\|\beta\|^2} \tau \sum_{l=0}^N \|g^{l+1}\|_2^2 + \frac{1}{2} \|u^0\|_2^2 + 2\nu \tau \sum_{l=0}^N \|u^{l+1}\|_{*V}^2 \right] \frac{2R}{R - \|\beta\| \operatorname{diam}(\Omega)} \quad (7.148)$$

$$\tau \sum_{l=0}^N \|u^{l+1}\|_2^2 \leq \left[ \frac{R}{\|\beta\|^2} \tau \sum_{l=0}^N \|g^{l+1}\|_2^2 + \frac{1}{2} \|u^0\|_2^2 + \nu \tau \sum_{l=0}^N \|u^{l+1}\|_{*V}^2 \right] \frac{4R}{\|\beta\|^2} \quad (7.149)$$

**Étape 3 :** On commence par majorer  $\nu \tau \sum_{l=0}^N \|u^{l+1}\|_{*V}^2$  :

$$\begin{aligned} \tau \nu \sum_{l=0}^N \|u^{l+1}\|_{*V}^2 &\stackrel{(7.143)}{\leq} \frac{1}{4\kappa} \tau \sum_{l=0}^N \|g^{l+1}\|_2^2 + \kappa \tau \sum_{l=0}^N \|u^{l+1}\|_2^2 + \frac{1}{2} \|u^0\|_2^2 \\ &\stackrel{(7.149)}{\leq} \frac{1}{4\kappa} \tau \sum_{l=0}^N \|g^{l+1}\|_2^2 + \frac{1}{2} \|u^0\|_2^2 \\ &+ \kappa \left[ \frac{R}{\|\beta\|^2} \tau \sum_{l=0}^N \|g^{l+1}\|_2^2 + \frac{1}{2} \|u^0\|_2^2 + \nu \tau \sum_{l=0}^N \|u^{l+1}\|_{*V}^2 \right] \frac{4R}{\|\beta\|^2} \end{aligned} \quad (7.150)$$

En choisissant  $\kappa \frac{4R}{\|\beta\|^2} = \frac{1}{2}$  et en soustrayant  $\frac{1}{2} \tau \nu \sum_{l=0}^N \|u^{l+1}\|_{*V}^2$  à chaque membre de l'inégalité, on obtient après multiplication par 2 :

$$\tau \nu \sum_{l=0}^N \|u^{l+1}\|_{*V}^2 \leq \left( \frac{4R}{\|\beta\|^2} + \frac{R}{\|\beta\|^2} \right) \tau \sum_{l=0}^N \|g^{l+1}\|_2^2 + \frac{3}{2} \|u^0\|_2^2 \quad (7.151)$$

Avec  $R = 2\|\beta\| \operatorname{diam}(\Omega)$  et en combinant (7.148) et (7.151), on peut maintenant conclure à l'existence d'une constante  $\tilde{C} > 0$  ne dépendant ni de  $\operatorname{diam}(\Omega)$  et ni de  $\|\beta\|$ , telle que

$$\max_{0 \leq k \leq N} \|u^{k+1}\|_2^2 + \nu \tau \sum_{l=0}^N \|u^{l+1}\|_{*V}^2 \leq \tilde{C} \left[ \frac{\operatorname{diam}(\Omega)}{\|\beta\|} \tau \sum_{l=0}^N \|g^{l+1}\|_2^2 + \|u^0\|_2^2 \right]$$

□

## 8 Stabilité du problème approché

La preuve du théorème de stabilité pour le problème suit les mêmes lignes que le théorème 7.1. Avant de procéder à sa formulation et sa preuve, il est nécessaire de formuler les versions discrètes de certaines inégalités utilisées dans la preuve du problème continu. La condition sur la famille de maillage  $(\mathcal{T}_h)$  est la même que pour le problème stationnaire :

**Définition 8.1 (Condition  $\delta - S^+$ -faible pour  $(\mathcal{T}_h)_h$ )** Soit  $(\mathcal{T}_h)_h$  une famille de maillage. On dira que la condition  $\delta - S^+$ -faible est satisfaite pour cette famille, si il existe une constante  $\delta > 0$ , telle que

$$\forall \beta \in \mathbb{R}^2 \quad \forall h > 0 \quad \exists \tilde{\beta} \in \mathbb{R}^2 \neq 0 \text{ avec } \|\tilde{\beta}\| = \|\beta\| \text{ et} \quad (8.152)$$

$$\forall i \in J_0 \quad \sum_{j \in S^+(i)} \theta_{i,j}^\beta [\tilde{\beta} \cdot (Q_j - Q_i)] \geq \delta \|\beta\|^2 |S_i|^2$$

Si le choix  $\beta = \tilde{\beta}$  est possible, on retrouve la condition  $\delta - S^+$  de la définition 4.3, page 52. Cette condition sur le maillage est étudiée au chapitre 6.1.

**Théorème 8.1** Soit  $N \in \mathbb{N}$ , le pas pour  $t$  est défini par  $\tau = \frac{T}{N+1}$ . On pose  $g_h^k(x) = I_h(g(\cdot, \tau k)(x))$  pour  $k \in \{0, \dots, N+1\}$ ,  $u_h^0 = I_h(u^0)$  et  $u_h^1, u_h^2, \dots, u_h^{N+1} \in V_h$  définis par

$$\frac{1}{\tau}(u_h^{k+1} - u_h^k, v_h) + \nu((u_h^{k+1}, v_h))_h + b_h(u_h^{k+1}, v_h) = (g_h^{k+1}, v_h) \quad \forall v_h \in V_h \quad \forall k \in \{1, \dots, N+1\} \quad (8.153)$$

Si la condition (8.152) est satisfaite, alors il existe  $h_0 > 0$ , une constante  $C > 0$  ne dépendant ni de  $h$  et ni de  $\|\beta\|$ , telle que pour tout  $h \in ]0; h_0[$  :

$$\max_{0 \leq k \leq N} \|u_h^{k+1}\|_2 + \sqrt{\nu} \left( \tau \sum_{l=0}^N \|u_h^{l+1}\|_{*V_h}^2 \right)^{1/2} \leq C \left[ \frac{1}{\sqrt{\delta \|\beta\|}} \left( \tau \sum_{l=0}^N \|g^{l+1}\|_{H^1(\Omega)}^2 \right)^{1/2} + \|u^0\|_{H^1(\Omega)} \right] \quad (8.154)$$

**Preuve :** La valeur de  $h_0$  est déterminée par les lemmes utilisés dans la preuve, ainsi que les conditions géométriques sur le maillage (régularité et quasi-uniformité).

**Étape 1 :** Pour  $v_h = u_h^{k+1}$  dans (8.153) on a

$$\begin{aligned} \frac{1}{\tau}(u_h^{k+1} - u_h^k, u_h^{k+1}) + \nu \|u_h^{k+1}\|_{*V_h}^2 + b_h(u_h^{k+1}, u_h^{k+1}) &= (g_h^{k+1}, u_h^{k+1}) \\ \frac{1}{\tau}(u_h^{k+1} - u_h^k, u_h^{k+1}) + \nu \|u_h^{k+1}\|_{*V_h}^2 &\leq (g_h^{k+1}, u_h^{k+1}) \end{aligned}$$

On utilise ici<sup>25</sup> le théorème 2.17, page 46 avec  $\varphi = 1 : 0 \leq b_h(u_h^{k+1}, u_h^{k+1})$ .

Ainsi (en remplaçant l'indice  $k$  par  $l$ )

$$\frac{1}{\tau}(u_h^{l+1}, u_h^{l+1}) + \nu \|u_h^{l+1}\|_{*V_h}^2 \leq (g_h^{l+1}, u_h^{l+1}) + \frac{1}{\tau}(u_h^l, u_h^{l+1})$$

25. Version discrète de (7.142). Ce résultat apparaît déjà dans [Ike83], lemme 3

Avec l'inégalité  $(u_h^l, u_h^{l+1}) \leq \frac{1}{2}\|u_h^l\|_2^2 + \frac{1}{2}\|u_h^{l+1}\|_2^2$ , simplification de  $\frac{1}{2}\|u_h^{l+1}\|_2^2$  et sommation sur  $l$  :

$$\frac{1}{2\tau} \sum_{l=0}^s \|u_h^{l+1}\|_2^2 + \nu \sum_{l=0}^s \|u_h^{l+1}\|_{*V_h}^2 \leq \sum_{l=0}^s (g_h^{l+1}, u_h^{l+1}) + \frac{1}{2\tau} \sum_{l=0}^s \|u_h^l\|_2^2$$

Multiplication des deux membres par  $\tau$ , soustraction de  $\frac{1}{2} \sum_{l=1}^s \|u_h^l\|_2^2$  et utilisation de l'inégalité  $(g_h^{l+1}, u_h^{l+1}) \leq \|g_h^{l+1}\| \|u_h^{l+1}\| \leq \frac{1}{4\kappa} \|g_h^{l+1}\|_2^2 + \kappa \|u_h^{l+1}\|_2^2$  et  $s = N$  :

$$\frac{1}{2} \|u_h^{N+1}\|_2^2 + \tau\nu \sum_{l=0}^N \|u_h^{l+1}\|_{*V_h}^2 \leq \frac{1}{4\kappa} \tau \sum_{l=0}^N \|g_h^{l+1}\|_2^2 + \kappa\tau \sum_{l=0}^N \|u_h^{l+1}\|_2^2 + \frac{1}{2} \|u_h^0\|_2^2$$

Ainsi, en particulier :

$$\tau\nu \sum_{l=0}^N \|u_h^{l+1}\|_{*V_h}^2 \leq \frac{1}{4\kappa} \tau \sum_{l=0}^N \|g_h^{l+1}\|_2^2 + \kappa\tau \sum_{l=0}^N \|u_h^{l+1}\|_2^2 + \frac{1}{2} \|u_h^0\|_2^2 \quad (8.155)$$

**Étape 2 :** Soit  $\tilde{\beta}$  défini comme dans (8.152) (donc  $\|\beta\| = \|\tilde{\beta}\|$ ) et  $\varphi$  par  $\varphi(x) = R - \tilde{\beta} \cdot (x - x_0)$ , avec  $x_0 \in \partial\Omega$ , tel que  $\tilde{\beta} \cdot (x - x_0) \geq 0$  pour  $x \in \Omega$ .  $R > \|\beta\| \text{diam}(\Omega)$  est une constante qui sera déterminée plus tard. On a donc  $0 < R - \|\beta\| \text{diam}(\Omega) \leq \varphi(x) \leq R$ . Prenons  $v_h = \rho_h(u_h^{k+1}\varphi)$  dans (8.153) :

$$\begin{aligned} \frac{1}{\tau} (u_h^{k+1} - u_h^k, \rho_h(u_h^{k+1}\varphi)) + \nu((u_h^{k+1}, \rho_h(u_h^{k+1}\varphi)))_h + \\ b(u_h^{k+1}, \rho_h(u_h^{k+1}\varphi)) = (g^{k+1}, \rho_h(u_h^{k+1}\varphi)) \end{aligned} \quad (8.156)$$

Maintenant avec le théorème 2.17 :

$$\begin{aligned} \frac{1}{2} \sum_{i \in J_0} \sum_{j \in S^+(i)} u_h^{k+1}(Q_i)^2 \theta_{i,j}^\beta (\varphi(Q_i) - \varphi(Q_j)) &\leq b(u_h^{k+1}, \rho_h(u_h^{k+1}\varphi)) \\ \frac{1}{2} \sum_{i \in J_0} \sum_{j \in S^+(i)} u_h^{k+1}(Q_i)^2 \theta_{i,j}^\beta [\tilde{\beta} \cdot (Q_j - Q_i)] &\leq b(u_h^{k+1}, \rho_h(u_h^{k+1}\varphi)) \end{aligned}$$

Avec (8.152) :

$$\delta \frac{1}{2} \sum_{i \in J_0} u_h^{k+1}(Q_i)^2 \|\beta\|^2 |S_i|^2 \leq b(u_h^{k+1}, \rho_h(u_h^{k+1}\varphi))$$

$$\text{Avec lemme 2.10 et (2.38) : } |T| \leq h_T^2 \leq \frac{1}{\sigma^2} |S_i|^2$$

$$\frac{\delta}{2\sigma^2} \|\beta\|^2 \|u_h^{k+1}\|_2^2 \leq b(u_h^{k+1}, \rho_h(u_h^{k+1}\varphi)) \quad (8.157)$$

Ce qui donne la version discrète de (7.144). Avec (8.156) on a ainsi

$$\begin{aligned} \frac{1}{\tau} (u_h^{k+1}, \rho_h(u_h^{k+1}\varphi)) + \frac{\delta}{2\sigma^2} \|\beta\|^2 \|u_h\|_2^2 &\leq -\nu((u_h^{k+1}, \rho_h(u_h^{k+1}\varphi)))_h + (g_h^{k+1}, \rho_h(u_h^{k+1}\varphi)) \\ &\quad + \frac{1}{\tau} (u_h^k, \rho_h(u_h^{k+1}\varphi)) \end{aligned} \quad (8.158)$$

Avec le lemme 2.13, lemme 2.11 et l'inégalité de Friedrichs :

$$\begin{aligned} -\nu((u_h^{k+1}, \rho_h(u_h^{k+1}\varphi)))_h &\leq \nu \|u_h^{k+1}\|_{*V_h} \hat{c}_3 (\|u_h^{k+1}\|_{*V_h} \|\varphi\|_\infty + \hat{c}_* \|u_h^{k+1}\|_{*V_h} \|\nabla \varphi\|_\infty) \\ &\leq \nu \hat{c}_3 (R + \hat{c}_* \|\beta\|) \|u_h^{k+1}\|_{*V_h}^2 \end{aligned} \quad (8.159)$$

Maintenant :  $\rho_h(w_h\varphi) \in V_h \subset X_h$  pour  $w_h \in V_h$ , donc avec le lemme 2.10 on obtient

$$\begin{aligned} \|w_h\|_2^2 &= (w_h, w_h) = \frac{1}{3} \sum_{T \in \mathcal{T}_h} |T| \sum_{j=1}^3 w_h(Q_T^j)^2 \frac{\varphi(Q_T^j)}{\varphi(Q_T^j)} \\ &\leq \frac{1}{R - \text{diam}(\Omega) \|\beta\|} \frac{1}{3} \sum_{T \in \mathcal{T}_h} |T| \sum_{j=1}^3 w_h(Q_T^j)^2 \varphi(Q_T^j) \\ \|w_h\|_2^2 &\leq \frac{1}{R - \text{diam}(\Omega) \|\beta\|} (w_h, \rho_h(w_h\varphi)) \text{ et aussi} \\ (w_h, \rho_h(w_h\varphi)) &\leq \|w_h\|_2 \|\rho_h(w_h\varphi)\|_2 \stackrel{\text{lemme 2.14}}{\leq} R \|w_h\|_2^2 \end{aligned}$$

Donc (version discrète de (7.146)) :

$$(R - \text{diam}(\Omega) \|\beta\|) \leq (w_h, \rho_h(w_h\varphi)) \leq R \|w_h\|_2^2 \quad \forall w_h \in V_h, \quad (8.160)$$

Pour  $v_h, w_h \in V_h$  :

$$\begin{aligned} (v_h, \rho_h(w_h\varphi)) &= \frac{1}{3} \sum_{T \in \mathcal{T}_h} |T| \sum_{j=1}^3 v_h(Q_T^j) w_h(Q_T^j) \varphi(Q_T^j) \\ &\stackrel{\text{inégalité de Hölder}}{\leq} (v_h, \rho_h(v_h\varphi))^{1/2} (w_h, \rho_h(w_h\varphi))^{1/2} \end{aligned} \quad (8.161)$$

$ab \leq \kappa a^2 + \frac{1}{4\kappa} b^2$  pour  $a, b, \kappa > 0$ . Ainsi :

$$(v_h, \rho_h(w_h\varphi)) \leq \kappa (v_h, \rho_h(v_h\varphi)) + \frac{1}{4\kappa} (w_h, \rho_h(w_h\varphi)) \quad (8.162)$$

Pour  $v_h = u_h^k, w_h = u_h^{k+1}$  et  $\kappa = \frac{1}{2}$  on obtient :

$$(u^k, \rho_h(u^{k+1}\varphi)) \leq \frac{1}{2} (u_h^k, \rho_h(u_h^k\varphi)) + \frac{1}{2} (u_h^{k+1}, \rho_h(u_h^{k+1}\varphi))$$

Pour  $v_h = g_h^{k+1}, w_h = u_h^{k+1}$  et  $\kappa > 0$  on obtient :

$$(g_h^{k+1}, \rho_h(u_h^{k+1}\varphi)) \leq \kappa (g_h^{k+1}, \rho_h(g_h^{k+1}\varphi)) + \frac{1}{4\kappa} (u_h^{k+1}, \rho_h(u_h^{k+1}\varphi))$$

En appliquant les inégalités ci-dessus, ainsi que (8.159) à (8.158), en soustrayant  $\frac{1}{2} (u_h^{k+1}, \rho_h(u_h^{k+1}\varphi))$  aux deux membres de l'inégalité et en remplaçant  $k$  par  $l$  on obtient

$$\begin{aligned} \frac{1}{2\tau} (u_h^{l+1}, \rho_h(u_h^{l+1}\varphi)) &+ \frac{\delta}{2\sigma^2} \|\beta\|^2 \|u_h^{l+1}\|_2^2 \\ &\leq \nu \hat{c}_3 (R + \hat{c}_* \|\beta\|) \|u_h^{l+1}\|_{*V_h}^2 + \kappa (g_h^{l+1}, \rho_h(g_h^{l+1}\varphi)) + \\ &\quad \frac{1}{4\kappa} (u_h^{l+1}, \rho_h(u_h^{l+1}\varphi)) + \frac{1}{2\tau} (u_h^l, \rho_h(u_h^l\varphi)) \end{aligned} \quad (8.163)$$

Sommation de 0 à  $s$ , soustraction de  $\sum_{l=1}^s \frac{1}{2\tau} (u_h^l, \rho_h(u_h^l\varphi))$  et (8.160) :

$$\begin{aligned} \frac{1}{2\tau} (u_h^{s+1}, \rho_h(u_h^{s+1}\varphi)) &+ \frac{\delta}{2\sigma^2} \|\beta\|^2 \sum_{l=0}^s \|u_h^{l+1}\|_2^2 \\ &\leq \nu \hat{c}_3 (R + \hat{c}_* \|\beta\|) \sum_{l=0}^s \|u_h^{l+1}\|_{*V_h}^2 + \kappa \sum_{l=0}^s (g_h^{l+1}, \rho_h(g_h^{l+1}\varphi)) + \\ &\quad R \frac{1}{4\kappa} \sum_{l=0}^s \|u_h^{l+1}\|_2^2 + \frac{1}{2\tau} (u_h^0, \rho_h(u_h^0\varphi)) \end{aligned}$$

En choisissant  $\kappa = \frac{R\sigma^2}{\delta\|\beta\|^2}$  et soustraction de  $\frac{\delta}{4\sigma^2}\|\beta\|^2 \sum_{l=0}^s \|u^{l+1}\|_2^2$  on a :

$$\begin{aligned} & \frac{1}{2\tau}(u_h^{s+1}, \rho_h(u_h^{s+1}\varphi)) + \frac{\delta}{4\sigma^2}\|\beta\|^2 \sum_{l=0}^s \|u_h^{l+1}\|_2^2 \\ & \leq \nu\hat{c}_3(R + \hat{c}_*\|\beta\|) \sum_{l=0}^s \|u_h^{l+1}\|_{*V_h}^2 + \frac{R\sigma^2}{\delta\|\beta\|^2} \sum_{l=0}^s (g_h^{l+1}, \rho_h(g_h^{l+1}\varphi)) + \frac{1}{2\tau}(u_h^0, \rho_h(u_h^0\varphi)) \end{aligned}$$

Les termes du membre gauche de l'inégalité étant positifs, on obtient après multiplication de l'inégalité par  $\tau$

$$\begin{aligned} & \frac{1}{2} \max_{0 \leq k \leq N} (u_h^{k+1}, \rho_h(u_h^{k+1}\varphi)) \\ & \leq \hat{c}_3(R + \hat{c}_*\|\beta\|) \nu\tau \sum_{l=0}^N \|u_h^{l+1}\|_{*V_h}^2 + \frac{R\sigma^2}{\delta\|\beta\|^2} \tau \sum_{l=0}^N (g_h^{l+1}, \rho_h(g_h^{l+1}\varphi)) + \frac{1}{2}(u_h^0, \rho_h(u_h^0\varphi)) \\ & \quad \frac{\delta}{4\sigma^2} \|\beta\|^2 \tau \sum_{l=0}^N \|u_h^{l+1}\|_2^2 \\ & \leq \hat{c}_3(R + \hat{c}_*\|\beta\|) \nu\tau \sum_{l=0}^N \|u_h^{l+1}\|_{*V_h}^2 + \frac{R\sigma^2}{\delta\|\beta\|^2} \tau \sum_{l=0}^N (g_h^{l+1}, \rho_h(g_h^{l+1}\varphi)) + \frac{1}{2}(u_h^0, \rho_h(u_h^0\varphi)) \end{aligned}$$

Il faut maintenant se débarrasser de  $\varphi$ . Avec (8.160) :

$$\begin{aligned} & \max_{0 \leq k \leq N} \|u_h^{k+1}\|_2^2 \tag{8.164} \\ & \leq \left[ 2\hat{c}_3(R + \hat{c}_*\|\beta\|) \nu\tau \sum_{l=0}^N \|u_h^{l+1}\|_{*V_h}^2 + 2\frac{R^2\sigma^2}{\delta\|\beta\|^2} \tau \sum_{l=0}^N \|g_h^{l+1}\|_2^2 + R\|u_h^0\|_2^2 \right] \frac{1}{R - \text{diam}(\Omega)\|\beta\|} \\ & \quad \tau \sum_{l=0}^N \|u_h^{l+1}\|_2^2 \\ & \leq \left[ \hat{c}_3(R + \hat{c}_*\|\beta\|) \nu\tau \sum_{l=0}^N \|u_h^{l+1}\|_{*V_h}^2 + \frac{R^2\sigma^2}{\delta\|\beta\|^2} \tau \sum_{l=0}^N \|g_h^{l+1}\|_2^2 + \frac{R}{2}\|u_h^0\|_2^2 \right] \frac{4\sigma^2}{\delta\|\beta\|^2} \tag{8.165} \end{aligned}$$

**Etape 3 :** On commence par majorer  $\nu\tau \sum_{l=0}^N \|u_h^{l+1}\|_{*V_h}^2$  :

$$\begin{aligned} \tau\nu \sum_{l=0}^N \|u_h^{l+1}\|_{*V_h}^2 & \stackrel{(8.155)}{\leq} \frac{1}{4\kappa} \tau \sum_{l=0}^N \|g_h^{l+1}\|_2^2 + \kappa\tau \sum_{l=0}^N \|u_h^{l+1}\|_2^2 + \frac{1}{2}\|u_h^0\|_2^2 \\ & \stackrel{(8.165)}{\leq} \frac{1}{4\kappa} \tau \sum_{l=0}^N \|g_h^{l+1}\|_2^2 + \frac{1}{2}\|u^0\|_2^2 \\ & + \kappa \left[ \hat{c}_3(R + \hat{c}_*\|\beta\|) \nu\tau \sum_{l=0}^N \|u_h^{l+1}\|_{*V_h}^2 + \frac{R^2\sigma^2}{\delta\|\beta\|^2} \tau \sum_{l=0}^N \|g_h^{l+1}\|_2^2 + \frac{R}{2}\|u_h^0\|_2^2 \right] \frac{4\sigma^2}{\delta\|\beta\|^2} \end{aligned} \tag{8.166}$$



En choisissant  $\kappa \frac{4\sigma^2}{\delta \|\beta\|^2} \hat{c}_3 (R + \hat{c}_* \|\beta\|) = \frac{1}{2}$  et en soustrayant  $\frac{1}{2} \tau \nu \sum_{l=0}^N \|u^{l+1}\|_{*V}^2$  à chaque membre de l'inégalité, on obtient après multiplication par 2 :

$$\begin{aligned} \tau \nu \sum_{l=0}^N \|u_h^{l+1}\|_{*V}^2 &\leq \frac{4\sigma^2 \hat{c}_3 (R + \hat{c}_* \|\beta\|)}{\delta \|\beta\|^2} \tau \sum_{l=0}^N \|g_h^{l+1}\|_2^2 + \|u_h^0\|_2^2 \\ &+ \frac{R^2 \sigma^2}{\delta \|\beta\|^2 \hat{c}_3 (R + \hat{c}_* \|\beta\|)} \tau \sum_{l=0}^N \|g_h^{l+1}\|_2^2 + \frac{R}{\hat{c}_3 (R + \hat{c}_* \|\beta\|)} \|u_h^0\|_2^2 \end{aligned}$$

Avec  $R = 2\|\beta\| \text{diam}(\Omega)$  :

$$\begin{aligned} \tau \nu \sum_{l=0}^N \|u_h^{l+1}\|_{*V}^2 &\leq \frac{4\sigma^2 \hat{c}_3 (2 \text{diam}(\Omega) + \hat{c}_*)}{\delta \|\beta\|} \tau \sum_{l=0}^N \|g_h^{l+1}\|_2^2 + \|u_h^0\|_2^2 \\ &+ \frac{(2 \text{diam}(\Omega) + \hat{c}_*) \sigma^2}{\delta \|\beta\| \hat{c}_3} \tau \sum_{l=0}^N \|g_h^{l+1}\|_2^2 + \frac{1}{\hat{c}_3} \|u_h^0\|_2^2 \end{aligned} \quad (8.167)$$

et en combinant (8.164) et (8.167), on peut maintenant conclure à l'existence d'une constante  $\hat{C} > 0$  ne dépendant ni de  $h$  et ni de  $\|\beta\|$ , telle que

$$\begin{aligned} \max_{0 \leq k \leq N} \|u_h^{k+1}\|_2^2 + \nu \tau \sum_{l=0}^N \|u_h^{l+1}\|_{*V}^2 &\leq \hat{C} \left[ \frac{1}{\delta \|\beta\|} \tau \sum_{l=0}^N \|g_h^{l+1}\|_2^2 + \|u_h^0\|_2^2 \right] \\ (2.69) \text{ du lemme 2.3, page 35} &\leq \hat{C} c_{T3}^2 \left[ \frac{1}{\delta \|\beta\|} \tau \sum_{l=0}^N \|g^{l+1}\|_{H^1(\Omega)}^2 + \|u^0\|_{H^1(\Omega)}^2 \right] \end{aligned}$$

□

---

## Quatrième partie

# Estimation de l'erreur pour l'équation d'advection-diffusion non stationnaire

## Sommaire

---

<b>9</b>	<b>Introduction</b>	<b>82</b>
9.1	Existence, unicité, conditions de compatibilité pour le problème continu	82
9.2	Estimation de l'erreur . . . . .	84
<b>10</b>	<b>Estimation de l'erreur</b>	<b>86</b>
<b>11</b>	<b>Simulations numériques</b>	<b>91</b>

---

## 9 Introduction

Soit  $0 \neq \beta \in \mathbb{R}^2$ ,  $\Omega$  un domaine polygonal convexe borné et  $T, \nu > 0$ .

On considère le problème d'advection-diffusion non stationnaire

$$\partial_t u - \nu \Delta_x u + (\beta \cdot \nabla_x) u = g \quad \text{dans } \Omega \times ]0, T[, \quad (9.168)$$

complété des conditions initiales et au bord :

$$u |_{\partial\Omega \times ]0, T[} = 0, \quad u(x, 0) = u^0(x) \quad \text{pour } x \in \Omega, \quad (9.169)$$

Sans restriction, on suppose que  $\|\beta\| = 1$ .

Pour le problème approché, une discrétisation implicite en temps est utilisée :

Soit  $N \in \mathbb{N}$ , le pas pour  $t$  est défini par  $\tau = \frac{T}{N+1}$  et on pose  $g_h^k(x) = I_h(g(\cdot, \tau k))(x)$  pour  $k \in \{0, \dots, N+1\}$ .

**Définition 9.1** *Formulation faible du problème approché d'advection-diffusion non stationnaire*

Trouver  $u_h^k \in V_h$  pour  $k \in \{1, \dots, N+1\}$  tels que  $u_h^0 = I_h(u^0)$  et

$$\frac{1}{\tau}(u_h^{k+1} - u_h^k, v_h) + \nu((u_h^{k+1}, v_h))_h + b_h(u_h^{k+1}, v_h) = (g_h^{k+1}, v_h) \quad \forall v_h \in V_h \quad \forall k \in \{1, \dots, N+1\} \quad (9.170)$$

Les conditions standard pour la famille de maillage  $(\mathcal{T}_h)_h$  sont présentées au chapitre 2.1.

La discrétisation ici est implicite, car les termes de diffusion et de convection sont pris en  $t = \tau(k+1)$ .<sup>26</sup>

### 9.1 Existence, unicité, conditions de compatibilité pour le problème continu

Pour l'espace de Banach  $\mathfrak{B}$ , pour  $a, b \in \mathbb{R}$  avec  $a < b$ , et pour  $p \in [1, \infty]$ ,  $m \in \mathbb{N}$ , les définitions habituelles des espaces  $C^0([a, b], \mathfrak{B})$ ,  $C^m([a, b], \mathfrak{B})$ ,  $L^p(a, b, \mathfrak{B})$  et  $H^m(a, b, \mathfrak{B})$  sont utilisées.

Les normes de  $L^p(a, b, \mathfrak{B})$  et  $H^m(a, b, \mathfrak{B})$  sont respectivement notées  $\|\cdot\|_{L^p(a,b,\mathfrak{B})}$  et  $\|\cdot\|_{H^m(a,b,\mathfrak{B})}$ .

Utilisé sera le

---

<sup>26</sup> Dans [AsFF98], une discrétisation semi-implicite est considérée, dans laquelle le terme convectif est pris en  $k\tau$  et le terme de diffusion en  $(k+1)\tau$ .

**Théorème 9.1** Soit  $a, b \in \mathbb{R}$  avec  $a < b$ , et soit  $v \in H^1(a, b, \mathfrak{B})$ . Alors  $v \in C^0([a, b], \mathfrak{B})$  et

$$v(t) = \int_a^t v'(s) ds + v(a) \quad \text{for } t \in [a, b],$$

où l'intégrale ci-dessus doit être comprise comme une intégrale au sens de Bochner dans  $\mathfrak{B}$ . De plus, la fonction  $v$  est dérivable presque partout dans  $]a; b[$  et sa dérivée en  $t$  est égale à la dérivée au sens des distributions de  $v'$  en  $t$ .

**Preuve :** Voir [Tem79, Lemma 3.1.1]. □

Pour  $\mathfrak{B}$ , on note  $\mathfrak{B}'$  le dual de  $\mathfrak{B}$ , équipé de la norme duale habituelle. Si  $v \in \mathfrak{B}$  et  $F \in \mathfrak{B}'$  ( $F$  est donc un opérateur linéaire borné de  $\mathfrak{B}$  dans  $\mathbb{R}$ ), on note  $F(v) = \langle F, v \rangle_{\mathfrak{B}', \mathfrak{B}}$ , où  $\langle F, v \rangle_{\mathfrak{B}', \mathfrak{B}}$  est le produit scalaire dans la dualité  $\mathfrak{B}', \mathfrak{B}$ .

À chaque fonction  $\gamma \in L^2(\Omega)$  on peut associer une forme linéaire  $F_\gamma \in H_0^1(\Omega)'$  définie par  $F_\gamma(v) := \int_\Omega \gamma v dx$  pour  $v \in H_0^1(\Omega)$ .

On identifiera  $\gamma$  et  $F_\gamma$ , c'est-à-dire que l'on écrira  $\gamma$  à la place de  $F_\gamma$ .

Inversement, si  $S$  est une partie de  $L^2(\Omega)$ , alors on dira que la forme linéaire  $l \in H_0^1(\Omega)'$  appartient à  $S$  si il existe une fonction  $\gamma \in S$  telle que  $l(v) = \int_\Omega \gamma v dx$  pour  $v \in H_0^1(\Omega)$ .

Ici aussi, on identifiera  $l$  et  $\gamma$  et on écrira  $l$  à la place de  $\gamma$ .<sup>27</sup>

Soit  $a : H_0^1(\Omega) \times H_0^1(\Omega) \mapsto \mathbb{R}$ , la forme bilinéaire définie par

$$a(v, w) := \int_\Omega (\nu \nabla v \cdot \nabla w + (\beta \cdot \nabla)v w) dx \quad \text{pour } v, w \in H_0^1(\Omega). \quad (9.171)$$

De façon évidente, il existe  $C > 0$  avec

$$|a(v, w)| \leq C \|v\|_{1,2} \|w\|_{1,2} \quad \text{pour } v, w \in H_0^1(\Omega),$$

Ainsi  $a(v, \cdot) \in H_0^1(\Omega)'$  et on peut donc définir un opérateur  $A : H_0^1(\Omega) \mapsto H_0^1(\Omega)'$  en posant  $Av := a(v, \cdot)$  pour  $v \in H_0^1(\Omega)$ . Si  $Av \in L^2(\Omega)$  (dans le sens précisé plus haut) pour un  $v \in H_0^1(\Omega)$ , alors avec  $f := Av - (\beta \cdot \nabla)v \in L^2(\Omega)$  on a  $\int_\Omega \nu \nabla v \cdot \nabla w dx = \int_\Omega f w dx$  pour  $w \in H_0^1(\Omega)$ .

Dans des domaines convexes, on a le théorème de régularité :

**Théorème 9.2** Si  $v \in H_0^1(\Omega)$  et  $F \in L^2(\Omega)$  avec  $\int_\Omega \nu \nabla v \cdot \nabla w dx = \int_\Omega F w dx$  pour  $w \in H_0^1(\Omega)$ , alors  $v \in H^2(\Omega)$  et  $-\Delta v = F$ .

De plus, il existe une constante  $C_2 = C_2(\Omega) > 0$  telle que

$$\|v\|_{2,2} \leq C_2 \|F\|_2 \quad \text{pour } v, F \text{ comme ci-dessus.} \quad (9.172)$$

**Preuve :** [Gri85, Theorem 3.1.2.1, Lemma 3.2.1.1] et la preuve [Gri85, Theorem 3.2.1.2].

□

<sup>27</sup>.  $L^2(\Omega)$  est l'espace pivot où les injections canoniques pour le schéma  $H_0^1(\Omega) \subset L^2(\Omega) = L^2(\Omega)' \subset H_0^1(\Omega)'$  sont continues et denses. On a  $\langle F_\gamma, v \rangle_{H_0^1(\Omega)', H_0^1(\Omega)} = (\gamma, v)$  pour  $\gamma \in L^2(\Omega)$  et  $v \in H_0^1(\Omega)$ .

Avec ce théorème on a  $v \in H^2(\Omega)$  et  $Av = -\nu \Delta v + (\beta \cdot \nabla)v$ . L'opérateur  $A$  peut être interprété comme une généralisation de l'opérateur différentiel  $-\nu \Delta + (\beta \cdot \nabla)$ .

Cela permet peut-être de mieux comprendre les conditions qui, à partir de (9.168) et (9.169), seront imposées à  $g$  et  $u_0$  :

$$g \in H^2(0, T, H_0^1(\Omega)') \cap L^\infty(0, T, H^1(\Omega)), \quad (9.173)$$

$$u_0 \in V, \quad g(0) - Au_0 \in V, \quad g'(0) - A(g(0) - Au_0) \in L^2(\Omega). \quad (9.174)$$

Les deux conditions dans (9.174) sont les relations de compatibilité.

Avec (9.173) et (9.174) on a le théorème de théorème d'existence, unicité et régularité des solutions de (9.168), (9.169) :

**Théorème 9.3** *Il existe une application unique*

$$u \in L^2(0, T, H_0^1(\Omega)) \cap H^1(0, T, H_0^1(\Omega)') \cap C^0([0, T], L^2(\Omega))$$

telle que

$$u'(t)(v) + \int_{\Omega} (\nu \nabla u(t) \cdot \nabla v + (\beta \cdot \nabla)u(t) v) dx = \int_{\Omega} g(t) v dx \quad (9.175)$$

$$\text{pour } v \in H_0^1(\Omega), \quad t \in ]0, T[, \quad u(0) = u_0.$$

Cette application appartient à  $C^0([0, T], H^2(\Omega)) \cap C^1([0, T], H_0^1(\Omega)) \cap C^2([0, T], L^2(\Omega))$  et est solution de

$$u'(t) - \nu \Delta u(t) + (\beta \cdot \nabla)u(t) = g(t) \quad \text{for } t \in [0, T]. \quad (9.176)$$

De plus, on a l'estimation

$$\begin{aligned} & \|u\|_{L^\infty(0, T, H^2(\Omega))} + \|u'\|_{L^\infty(0, T, H_0^1(\Omega))} + \|u''\|_{L^\infty(0, T, L^2(\Omega))} \\ & \leq C (1 + T^{1/2}) (\min\{1, \nu\})^{-2} \left( \|g'(0) - A(g(0) - Au_0)\|_2 \right. \\ & \quad \left. + \|g(0) - Au_0\|_{1,2} + \|u_0\|_{1,2} + \|g\|_{H^2(0, T, H_0^1(\Omega)')} + \|g\|_{L^\infty(0, T, L^2(\Omega))} \right). \end{aligned} \quad (9.177)$$

**Preuve :** Annexe A de [DM10] □

## 9.2 Estimation de l'erreur

Le résultat principal pour l'estimation de l'erreur est

**Théorème 9.4** *Quelque soit  $0 \leq s \leq N + 1$ , on a l'inégalité*

$$\begin{aligned} & \nu^{1/2} \left( \tau \sum_{k=0}^{N+1} \|u(t_k) - u_h^{(k)}\|_{*V_h}^2 \right)^{1/2} + \|u(t_s) - u_h^{(s)}\|_2 \\ & \leq \mathfrak{C} (1 + T) (\nu^{1/2} + \nu^{-5/2}) \mathfrak{F}(u_0, g) (h + \tau) \end{aligned}$$

où

$$\begin{aligned} \mathfrak{F}(u_0, g) := & \|g\|_{L^\infty(0,T,H^1(\Omega))} + \|g\|_{H^2(0,T,H_0^1(\Omega)')} \\ & + \|u_0\|_{1,2} + \|g(0) - Au_0\|_{1,2} + \|g'(0) - A(g(0) - Au_0)\|_2, \end{aligned}$$

et  $\mathfrak{C}$  représente une constante indépendante de  $\tau$  et  $h$ .

La preuve du résultat se fait en deux temps. Dans un premier temps, l'erreur est majorée (théorème 10.5) par certaines données du problème et la quantité

$$\|u\|_{L^\infty(0,T,H^2(\Omega))} + \|u'\|_{L^\infty(0,T,H_0^1(\Omega))} + \|u''\|_{L^\infty(0,T,L^2(\Omega))} \quad (9.178)$$

où  $u$  est la solution du problème continu (9.168), (9.169). Dans un second temps, le théorème 9.3, donne une majoration de la somme (9.178) en fonction des données du problème continu.

**Remarque 9.2** *Le paramètre  $\nu$  contribue à la majoration de l'erreur via un facteur  $\nu^{1/2} + \nu^{-5/2}$ . En particulier il n'y a pas de dépendance exponentielle par rapport à  $1/\nu$  (voir aussi [sFFK02]). Il n'y a pas de couplage entre la finesse du maillage et le pas en temps. Il est peut-être intéressant de remarquer que la dépendance exponentielle de l'erreur en  $1/\nu$  n'est pas une conséquence de la non linéarité du problème étudié dans [sFFK02], mais qu'elle apparaîtrait aussi si l'approche utilisée s'appliquait à notre problème linéaire.*

Les hypothèses pour  $g$  et  $u_0$  ne peuvent pas être affaiblies de manière significative, car la dépendance linéaire du pas dans le temps nécessite de majorer  $\|u'\|_{L^\infty(0,T,H_0^1(\Omega))}$  et  $\|u''\|_{L^\infty(0,T,L^2(\Omega))}$  (voir, par exemple la preuve du lemme 10.4).

Dans [Wlo87, §27], la théorie de la régularité des équations paraboliques nécessite que  $u$  appartienne à  $H^2(0,T,H_0^1(\Omega)) \cap H^3(0,T,H_0^1(\Omega)')$ . Les hypothèses de (9.173) et (9.174) permettent exactement cela ([Wlo87, Theorem 27.2]). (Exception faite de la condition additionnelle  $g \in L^\infty(0,T,H^1(\Omega))$ , qui est imposée pour permettre une inégalité d'interpolation sur  $g$  (voir la preuve du lemme 10.2) et avoir  $u \in C^0([0,T],H^2(\Omega))$ ).

Avec  $u \in H^2(0,T,H_0^1(\Omega)) \cap H^3(0,T,H_0^1(\Omega)')$  on a  $u \in C^1([0,T],H_0^1(\Omega)) \cap C^2([0,T],L^2(\Omega))$ . Ce dernier espace étant particulièrement adapté dans notre cas, il remplace  $u \in H^2(0,T,H_0^1(\Omega)) \cap H^3(0,T,H_0^1(\Omega)')$  dans le théorème 9.3.

On a

**Lemme 9.5** *Soit  $t_1, t_2 \in [0, T]$  avec  $t_1 < t_2$ . Alors*

$$\|u(t_2) - u(t_1) - (t_2 - t_1) u'(t_2)\|_2 \leq (t_2 - t_1)^2 \|u''\|_{L^\infty(0,T,L^2(\Omega))}. \quad (9.179)$$

**Preuve :** Théorème 9.1 ; voir aussi [sFFK02, p. 317]. □

Dans ce qui suit,  $\mathfrak{C}$  notera une constante qui ne dépendra que de  $\Omega$  et des constantes de régularité et de quasi-uniformité de la famille de maillage  $(\mathcal{T}_h)_{h>0}$ .

## 10 Estimation de l'erreur

### Rappel de quelques lemmes utiles

Pour  $v_h \in X_h$ ,  $K \in \mathcal{T}_h$ , soit  $E(v_h|K)$  l'extension continue de  $v_h|K$  à  $\bar{K}$ . Pour  $v_h, w_h \in X_h \oplus H^1(\Omega)$ , on pose

$$((v_h, w_h))_h := \sum_{K \in \mathcal{T}_h} \int_K \nabla(v_h|K) \cdot \nabla(w_h|K) dx, \quad \|v_h\|_{X_h} := (v_h, v_h) + ((v_h, v_h))_h^{1/2}.$$

Pour  $V_h$ , la semi-norme  $|v_h|_{X_h} = ((v_h, v_h))_h^{1/2}$  est une norme. Avec [Fei93, Lemma 8.9.92] ou encore [sFF99], il existe une constante  $h_0 > 0$ , telle que l'on ait l'inégalité discrète de Poincaré :

$$\|v_h\|_2 \leq \mathfrak{C} |v_h|_{X_h} (= \mathfrak{C} \|v_h\|_{*V_h}) \quad \text{pour } v_h \in V_h, \quad \text{si } h \leq h_0. \quad (10.180)$$

Avec [Cia79, §4.1, (4.1.17)], on a

$$(|K|/3) \sum_{i \in J, Q_i \in \bar{K}} w_h(Q_i) = \int_K w_h dx \quad \text{pour } K \in \mathcal{T}_h, w_h \in X_h. \quad (10.181)$$

Soit  $I_h : H^1(\Omega) \oplus X_h \mapsto X_h$  l'opérateur d'interpolation introduit dans [Fei93, 8.9.79, a)] et défini par

$$I_h(v)(Q_i) := |S_i|^{-1} \int_{S_i} v do_x \quad \text{for } v \in H^1(\Omega) \oplus X_h, i \in J.$$

On a  $I_h(v) \in V_h$  for  $v \in H_0^1(\Omega)$  et

$$\|v - I_h(v)\|_2 \leq \mathfrak{C} h \|v\|_{1,2} \quad \text{pour } v \in H^1(\Omega), \quad (10.182)$$

$$\|v - I_h(v)\|_{X_h} \leq \mathfrak{C} h \|v\|_{2,2} \quad \text{pour } v \in H^2(\Omega), \quad \|I_h(v)\|_{X_h} \leq \|v\|_{X_h} \quad \text{pour } v \in H^1(\Omega), \quad (10.183)$$

L'inégalité (10.182) est montrée dans [sFFK02, Lemma 9 a)], et les inégalités (10.183) dans [Fei93, (8.9.83),  $\alpha$ ) et  $\gamma$ ]. L'inégalité suivante, voir [Fei93, Lemma 8.9.85], nous sera aussi utile :

$$\left| \sum_{K \in \mathcal{T}_h} \int_{\partial K} v n_{K,i} E(w_h|K) do_x \right| \leq \mathfrak{C} h \|v\|_{1,2} \|w_h\|_{X_h} \quad (10.184)$$

for  $v \in H^1(\Omega)$ ,  $w_h \in V_h$ ,  $i \in \{1, 2\}$ , si  $h \leq h_0$ , avec  $h_0$  de (10.180).

### Preuve du théorème 9.4

La fonction  $u$  est la solution de (9.168), (9.169) comme définie par le théorème 9.3.

On suppose que  $h \leq h_0$ , avec  $h_0$  défini dans (10.180). On note

$$u_h(t) := I_h(u(t)) \quad \text{for } t \in [0, T], \quad t_k := k \tau, \quad w_h^{(k)} := u_h(t_k) - u_h^{(k)} \quad \text{for } k \in \{0, \dots, N+1\}.$$

L'objectif est de majorer l'erreur  $w_h^{(k+1)}$  for  $0 \leq k \leq N$ . Le point de départ est le

**Lemme 10.1** Soit  $k \in \{0, \dots, N\}$ . Alors

$$\nu \|w_h^{(k+1)}\|_{*V_h}^2 + \tau^{-1} \|w_h^{(k+1)}\|_2^2 \leq \sum_{s=1}^5 \mathfrak{A}_s^{(k+1)},$$

avec

$$\begin{aligned} \mathfrak{A}_1^{(k+1)} &:= \nu \left( (u_h(t_{k+1}) - u(t_{k+1}), w_h^{(k+1)}) \right)_h, \\ \mathfrak{A}_2^{(k+1)} &:= \nu \left( (u(t_{k+1}), w_h^{(k+1)}) \right)_h + \int_{\Omega} \nu \Delta u(t_{k+1}) w_h^{(k+1)} dx, \\ \mathfrak{A}_3^{(k+1)} &:= \int_{\Omega} g(t_{k+1}) w_h^{(k+1)} dx - (g_h^{(k+1)}, w_h^{(k+1)}), \\ \mathfrak{A}_4^{(k+1)} &:= - \int_{\Omega} (\beta \cdot \nabla) u(t_{k+1}) w_h^{(k+1)} dx + b_h(u_h(t_{k+1}), w_h^{(k+1)}), \\ \mathfrak{A}_5^{(k+1)} &:= \tau^{-1} \left( (u_h(t_{k+1}), w_h^{(k+1)}) - (u_h^{(k)}, w_h^{(k+1)}) \right) - \int_{\Omega} u'(t_{k+1}) w_h^{(k+1)} dx. \end{aligned}$$

**Preuve :** En posant  $w := w_h^{(k+1)}$ , on trouve avec le lemme 2.17, page 46 (en prenant  $\varphi = 1$ ), (9.170) et (9.176) que

$$\begin{aligned} \nu \|w\|_{*V_h}^2 + \tau^{-1} \|w\|_2^2 &\leq \nu \|w\|_{*V_h}^2 + \tau^{-1} \|w\|_2^2 + b_h(w, w) \\ &= \nu \left( (u_h(t_{k+1}), w) \right)_h + \tau^{-1} (u_h(t_{k+1}), w) + b_h(u_h(t_{k+1}), w) - (g_h^{(k+1)}, w) \\ &\quad - \tau^{-1} (u_h^{(k)}, w) \\ &= \sum_{s=1}^5 \mathfrak{A}_s^{(k+1)}. \end{aligned}$$

□

On commence par majorer  $\mathfrak{A}_1^{(k+1)}$ ,  $\mathfrak{A}_2^{(k+1)}$  et  $\mathfrak{A}_3^{(k+1)}$  :

**Lemme 10.2** Pour  $k \in \{0, \dots, N\}$ ,  $\epsilon \in (0, \infty)$ , on a

$$\begin{aligned} |\mathfrak{A}_s^{(k+1)}| &\leq \mathfrak{C} \nu \epsilon^{-1} h^2 \|u(t_{k+1})\|_{2,2}^2 + \epsilon \nu \|w_h^{(k+1)}\|_{*V_h}^2 \quad \text{for } s \in \{1, 2\}, \\ |\mathfrak{A}_3^{(k+1)}| &\leq \mathfrak{C} (\epsilon \nu)^{-1} h^2 \|g(t_{k+1})\|_{1,2}^2 + \epsilon \nu \|w_h^{(k+1)}\|_{*V_h}^2. \end{aligned}$$

**Preuve :** Notons  $w := w_h^{(k+1)}$ , on a

$$\begin{aligned} |\mathfrak{A}_1^{(k+1)}| &\leq \nu \left( \epsilon^{-1} \|u_h(t_{k+1}) - u(t_{k+1})\|_{*V_h}^2 + \epsilon \|w\|_{*V_h}^2 \right) \\ &\leq \mathfrak{C} \nu \epsilon^{-1} h^2 \|u(t_{k+1})\|_{2,2}^2 + \epsilon \nu \|w\|_{*V_h}^2, \end{aligned}$$

où la dernière inégalité vient de (10.183). Intégration par parties et (10.184) :

$$\begin{aligned} |\mathfrak{A}_2^{(k+1)}| &= \left| \nu \sum_{K \in \mathcal{T}_h} \int_{\partial K} (\nabla u(t_{k+1}) \cdot n_K) E(w|K) do_x \right| \leq \mathfrak{C} \nu h \|u(t_{k+1})\|_{2,2} \|w\|_{*V_h} \\ &\leq \mathfrak{C} \nu h^2 \epsilon^{-1} \|u(t_{k+1})\|_{2,2}^2 + \nu \epsilon \|w\|_{*V_h}^2. \end{aligned}$$



Avec (10.182) et (10.180),

$$\begin{aligned} |\mathfrak{A}_3^{(k+1)}| &\leq \|g(t_{k+1}) - I_h(g(t_{k+1}))\|_2 \|w\|_2 \leq \mathfrak{C} h \|g(t_{k+1})\|_{1,2} \|w\|_{*V_h} \\ &\leq \mathfrak{C} (\epsilon \nu)^{-1} h^2 \|g(t_{k+1})\|_{1,2}^2 + \nu \epsilon \|w\|_{*V_h}^2. \end{aligned}$$

□

Estimation de  $\mathfrak{A}_4^{(k+1)}$  :

**Lemme 10.3** *Soit  $k \in \{0, \dots, N\}$ ,  $\epsilon \in (0, \infty)$ . Alors*

$$|\mathfrak{A}_4^{(k+1)}| \leq \mathfrak{C} (\nu \cdot \epsilon)^{-1} h^2 \|u(t_{k+1})\|_{2,2}^2 + \epsilon \nu \|w_h^{(k+1)}\|_{*V_h}^2.$$

**Preuve :**

$$\begin{aligned} \mathfrak{A}_4^{(k+1)} &= - \int_{\Omega} \beta \cdot [\nabla_h u(t_{k+1}) - \nabla_h u_h(t_{k+1})] w_h^{(k+1)} dx \\ &\quad - \int_{\Omega} \beta \cdot \nabla_h u_h(t_{k+1}) w_h^{(k+1)} dx - \int_{\Omega} u_h(t_{k+1}) \beta \cdot \nabla_h w_h^{(k+1)} dx \\ &\quad + \int_{\Omega} u_h(t_{k+1}) \beta \cdot \nabla_h w_h^{(k+1)} dx + b_h(u_h(t_{k+1}), w_h^{(k+1)}) \end{aligned}$$

Maintenant

$$\begin{aligned} \left| - \int_{\Omega} \beta \cdot [\nabla_h u(t_{k+1}) - \nabla_h u_h(t_{k+1})] w_h^{(k+1)} dx \right| &\leq \|u(t_{k+1}) - I_h u(t_{k+1})\|_{*V_h} \|w_h^{(k+1)}\|_2 \\ &\leq \mathfrak{C} h \|u(t_{k+1})\|_{2,2} \|w_h^{(k+1)}\|_{*V_h} \quad \text{avec (10.183) et (10.180)} \end{aligned}$$

$$\left| - \int_{\Omega} \beta \cdot \nabla_h u_h(t_{k+1}) w_h^{(k+1)} dx - \int_{\Omega} u_h(t_{k+1}) \beta \cdot \nabla_h w_h^{(k+1)} dx \right| \leq \mathfrak{C} h \|u_h(t_{k+1})\|_{*V_h} \|w_h^{(k+1)}\|_{*V_h}$$

avec le lemme 2.15, page 42.

Enfin

$$\left| \int_{\Omega} u_h(t_{k+1}) \beta \cdot \nabla_h w_h^{(k+1)} dx + b_h(u_h(t_{k+1}), w_h^{(k+1)}) \right| \leq \mathfrak{C} h \|u_h(t_{k+1})\|_{*V_h} \|w_h^{(k+1)}\|_{*V_h}$$

avec (2.96), page 44. Avec (10.183) :  $\|u_h(t_{k+1})\|_{*V_h} \leq \|u(t_{k+1})\|_{*V_h} \leq \|u(t_{k+1})\|_{2,2}$ , donc

$$|\mathfrak{A}_4^{(k+1)}| \leq \mathfrak{C} h \|u(t_{k+1})\|_{2,2} \|w_h^{(k+1)}\|_{*V_h} \leq \mathfrak{C} (\nu \cdot \epsilon)^{-1} h^2 \|u(t_{k+1})\|_{2,2}^2 + \epsilon \nu \|w_h^{(k+1)}\|_{*V_h}^2.$$

$$\text{car } \mathfrak{C} a b = \mathfrak{C} \left( \frac{\sqrt{\mathfrak{C} h a}}{\sqrt{\epsilon \nu}} \right) \left( \frac{\sqrt{\epsilon \nu} b}{\sqrt{\mathfrak{C}}} \right) \leq \mathfrak{C} \left( \frac{\sqrt{\mathfrak{C} h a}}{\sqrt{\epsilon \nu}} \right)^2 + \mathfrak{C} \left( \frac{\sqrt{\epsilon \nu} b}{\sqrt{\mathfrak{C}}} \right)^2 = \mathfrak{C}^2 (\nu \cdot \epsilon)^{-1} h^2 a^2 + \epsilon \nu b^2 \quad \square$$

Estimation de  $\mathfrak{A}_5^{(k+1)}$  :

**Lemme 10.4** *Soit  $k \in \{0, \dots, N\}$ ,  $\epsilon \in (0, \infty)$ . Alors*

$$\begin{aligned} |\mathfrak{A}_5^{(k+1)}| &\leq \mathfrak{C} (\nu \epsilon)^{-1} (h^2 + \tau^2) \left( \|u'\|_{L^\infty(0,T,H_0^1(\Omega))} + \|u''\|_{L^\infty(0,T,L^2(\Omega))} \right)^2 \\ &\quad + 2 \epsilon \nu \|w_h^{(k+1)}\|_{*V_h}^2 + (2 \tau)^{-1} \|w_h^{(k)}\|_2^2 + (2 \tau)^{-1} \|w_h^{(k+1)}\|_{*V_h}^2. \end{aligned}$$

**Preuve :** On remarque que  $\mathfrak{A}_5^{(k+1)} = \sum_{s=1}^4 \mathfrak{B}_s$ , avec

$$\begin{aligned}\mathfrak{B}_1 &:= \tau^{-1} \left( (u_h(t_{k+1}), w) - (u(t_{k+1}), w) \right), \\ \mathfrak{B}_2 &:= \tau^{-1} \left( (u(t_{k+1}), w) - (u(t_k), w) \right) - (u'(t_{k+1}), w), \\ \mathfrak{B}_3 &:= \tau^{-1} \left( (u(t_k), w) - (u_h(t_k), w) \right), \quad \mathfrak{B}_4 := \tau^{-1} \left( (u_h(t_k), w) - (u_h^{(k)}, w) \right),\end{aligned}$$

où  $w := w_h^{(k+1)}$ . Comme  $u \in C^1([0, T], H_0^1(\Omega))$  (théorème 9.3), on a avec le théorème 9.1 que

$$u(t_{k+1}) - u(t_k) = \int_{t_k}^{t_{k+1}} u' ds,$$

l'intégrale ci-dessus étant comprise dans le sens d'une intégrale de Bochner dans  $H_0^1(\Omega)$ . Ainsi

$$\|u(t_{k+1}) - u(t_k)\|_{1,2} \leq \int_{t_k}^{t_{k+1}} \|u'(s)\|_{1,2} ds \leq \tau \|u'\|_{L^\infty(0,T,H_0^1(\Omega))}.$$

De (10.182) et (10.180) on peut déduire que

$$\begin{aligned}|\mathfrak{B}_1 + \mathfrak{B}_3| &\leq \tau^{-1} \left\| u_h(t_{k+1}) - u_h(t_k) - (u(t_{k+1}) - u(t_k)) \right\|_2 \|w\|_2 \quad (10.185) \\ &= \tau^{-1} \left\| I_h(u_h(t_{k+1}) - u_h(t_k)) - (u(t_{k+1}) - u(t_k)) \right\|_2 \|w\|_2 \\ &\leq \mathfrak{C} \tau^{-1} h \|u(t_{k+1}) - u(t_k)\|_{1,2} \|w\|_{*V_h} \leq \mathfrak{C} h \|u'\|_{L^\infty(0,T,H_0^1(\Omega))} \|w\|_{*V_h} \\ &\leq \mathfrak{C} (\epsilon \nu)^{-1} h^2 \|u'\|_{L^\infty(0,T,H_0^1(\Omega))}^2 + \epsilon \nu \|w\|_{*V_h}^2.\end{aligned}$$

De plus, on sait de (9.179) et (10.180) que

$$\begin{aligned}|\mathfrak{B}_2| &\leq \tau^{-1} \|u(t_{k+1}) - u(t_k) - \tau u'(t_{k+1})\|_2 \|w\|_2 \quad (10.186) \\ &\leq \mathfrak{C} \tau \|u''\|_{L^\infty(0,T,L^2(\Omega))} \|w\|_{*V_h} \leq \mathfrak{C} (\epsilon \nu)^{-1} \tau^2 \|u''\|_{L^\infty(0,T,L^2(\Omega))}^2 + \epsilon \nu \|w\|_{*V_h}^2.\end{aligned}$$

Il est évident que

$$|\mathfrak{B}_4| \leq \tau^{-1} \|w_h^{(k)}\|_2 \|w\|_2 \leq (2\tau)^{-1} \|w_h^{(k)}\|_2^2 + (2\tau)^{-1} \|w\|_2^2. \quad (10.187)$$

Le lemme 10.4 se déduit de (10.185) – (10.187) et de l'égalité  $\mathfrak{A}_5^{(k+1)} = \mathfrak{B}_1 + \dots + \mathfrak{B}_4$ .  $\square$

On peut maintenant obtenir une estimation d'erreur qui dépend de  $u$  :

**Théorème 10.5** *On a l'inégalité*

$$\begin{aligned}\nu^{1/2} \left( \tau \sum_{k=0}^{N+1} \|u(t_k) - u_h^{(k)}\|_{*V_h}^2 \right)^{1/2} + \|u(t_s) - u_h^{(s)}\|_2 \quad (10.188) \\ \leq \mathfrak{C} (1 + T^{1/2}) (\nu + \nu^{-1})^{1/2} \mathfrak{B}(u, g) (h + \tau)\end{aligned}$$

pour  $0 \leq s \leq N + 1$ , avec

$$\mathfrak{B}(u, g) := \|u\|_{L^\infty(0,T,H^2(\Omega))} + \|u'\|_{L^\infty(0,T,H_0^1(\Omega))} + \|u''\|_{L^\infty(0,T,L^2(\Omega))} + \|g\|_{L^\infty(0,T,H^1(\Omega))}.$$

**Preuve :**

Avec les lemmes 10.1, 10.2, 10.3 et 10.4, on a pour  $\epsilon \in (0, \infty)$  et  $k \in \{0, \dots, N\}$  que

$$\begin{aligned} & \nu \|w_h^{(k+1)}\|_{*V_h}^2 + \tau^{-1} \|w_h^{(k+1)}\|_2^2 \\ & \leq 6 \epsilon \nu \|w_h^{(k+1)}\|_{*V_h}^2 + \mathfrak{C} \epsilon^{-1} (\nu + \nu^{-1}) \mathfrak{B}(u, g)^2 (h^2 + \tau^2) \\ & \quad + \mathfrak{C} (\epsilon \nu)^{-1} h^2 \|g\|_{L^\infty(0, T, H^1(\Omega))} + (2 \tau)^{-1} \|w_h^{(k)}\|_2^2 + (2 \tau)^{-1} \|w_h^{(k+1)}\|_2^2. \end{aligned} \quad (10.189)$$

Choisissons maintenant  $\epsilon = 1/12$  et soustrayons  $6 \epsilon \nu \|w_h^{(k+1)}\|_{*V_h}^2$  et  $(2 \tau)^{-1} \|w_h^{(k+1)}\|_2^2$  des membres gauches et droits de (10.189). On obtient

$$\begin{aligned} & \nu \|w_h^{(k+1)}\|_{*V_h}^2 + \tau^{-1} \|w_h^{(k+1)}\|_2^2 \\ & \leq \mathfrak{C} (\nu + \nu^{-1}) \mathfrak{B}(u, g)^2 (h^2 + \tau^2) + \tau^{-1} \|w_h^{(k)}\|_2^2 \end{aligned} \quad (10.190)$$

pour  $k \in \{0, \dots, N\}$ . Pour  $s \in \{0, \dots, N\}$ , on effectue une sommation des deux côtés de (10.190) par rapport à  $k \in \{0, \dots, s\}$ , pour obtenir

$$\begin{aligned} & \nu \sum_{l=1}^{s+1} \|w_h^{(l)}\|_{*V_h}^2 + \tau^{-1} \sum_{l=1}^{s+1} \|w_h^{(l)}\|_2^2 \\ & \leq (s+1) \mathfrak{C} (\nu + \nu^{-1}) \mathfrak{B}(u, g)^2 (h^2 + \tau^2) + \tau^{-1} \sum_{k=0}^s \|w_h^{(k)}\|_2^2. \end{aligned}$$

Remarquons que  $w_h^{(0)} = u_h(0) - u_h^{(0)} = I_h(u(0)) - u_h^{(0)} = 0$  (voir (9.175), (9.170)), nous obtenons

$$\tau \nu \sum_{l=0}^{s+1} \|w_h^{(l)}\|_{*V_h}^2 + \|w_h^{(s+1)}\|_2^2 \leq \tau (s+1) \mathfrak{C} (\nu + \nu^{-1}) \mathfrak{B}(u, g)^2 (h^2 + \tau^2) \quad (10.191)$$

pour  $s \in \{0, \dots, N\}$ . Comme  $\tau (s+1) \leq \tau (N+1) = T$ , on obtient de (10.191) que

$$\|w_h^{(s+1)}\|_2^2 \leq \mathfrak{C} T (\nu + \nu^{-1}) \mathfrak{B}(u, g)^2 (h^2 + \tau^2) \quad (0 \leq s \leq N),$$

et avec  $s = N$ ,

$$\tau \nu \sum_{l=0}^{N+1} \|w_h^{(l)}\|_{*V_h}^2 \leq \mathfrak{C} T (\nu + \nu^{-1}) \mathfrak{B}(u, g)^2 (h^2 + \tau^2).$$

Ainsi

$$\begin{aligned} & \nu^{1/2} \left( \tau \sum_{l=0}^{N+1} \|u_h(t_l) - u_h^{(l)}\|_{*V_h}^2 \right)^{1/2} + \|u_h(t_s) - u_h^{(s)}\|_2 \\ & \leq \mathfrak{C} T^{1/2} (\nu + \nu^{-1})^{1/2} \mathfrak{B}(u, g) (h + \tau) \quad (0 \leq s \leq N+1). \end{aligned} \quad (10.192)$$

Avec (10.182), (10.183) :

$$\begin{aligned} & \|u_h(t_k) - u(t_k)\|_{*V_h} \leq \mathfrak{C} h \|u(t_k)\|_{2,2} \leq \mathfrak{C} h \mathfrak{B}(u, g), \\ & \|u_h(t_k) - u_h^{(k)}\|_2 \leq \mathfrak{C} h \|u(t_k)\|_{1,2} \leq \mathfrak{C} h \mathfrak{B}(u, g) \quad (0 \leq k \leq N+1), \end{aligned}$$

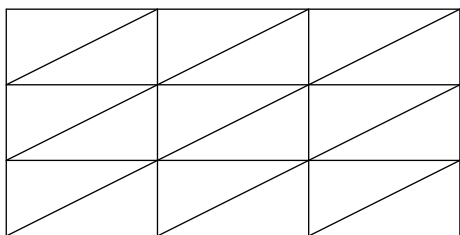


FIGURE 10 – Partie du maillage fait de rectangles divisés par deux.

En particulier, comme  $\tau(N+1) = T$ ,

$$\left( \tau \sum_{l=0}^{N+1} \|u_h(t_k) - u(t_k)\|_{*V_h}^2 \right)^{1/2} \leq \mathfrak{C} T^{1/2} \mathfrak{B}(u, g) h.$$

L'inégalité (10.188) se déduit ainsi de (10.192). □

Les normes de  $u$  qui apparaissent dans l'estimation d'erreur dans (10.188) peuvent être majorées par (9.177), ce qui conclut la preuve du théorème 9.4.

## 11 Simulations numériques

La méthode EF-VF a été l'objet de simulations numériques intenses ([sFFK02], [FFS03, Section 4.4], [HRB97], [DDMN98],...). Cependant, afin de tester numériquement les estimations d'erreurs obtenues, une solution pour (9.170) est trouvée pour un choix simple de  $\Omega$ ,  $f$  et  $u_0$ . Les  $L^\infty$ - $L^2$ - et  $L^2$ - $H^1$  erreurs discrètes sont ensuite calculées.

On considère le cas

$\Omega = ]0, a[ \times ]0, b[$  et

$$u(x, t) = \exp(-\nu \cdot (k_1^2/a^2 + k_2^2/b^2) \cdot \pi^2 \cdot t) \cdot \sin(k_1 \cdot \pi \cdot x_1/a - t \cdot \beta_1 \cdot k_1 \cdot \pi/a) \cdot \sin(k_2 \cdot \pi \cdot x_2/b - t \cdot \beta_2 \cdot k_2 \cdot \pi/b) \quad (11.193)$$

pour  $t \in [0, T]$ ,  $x \in \overline{\Omega}$ , avec

$$\beta = (\cos(\pi/3), \sin(\pi/3)), \quad a = 2.5, \quad b = 1, \quad k_1 = 2, \quad k_2 = 1, \quad T = 1.$$

En particulier  $u$  est la solution de (9.168), (9.169) avec

$$f = 0, \quad u_0(x) = \sin(k_1 \cdot \pi \cdot x_1/a) \cdot \sin(k_2 \cdot \pi \cdot x_2/b) \quad (x \in \overline{\Omega}). \quad (11.194)$$

Pour la famille de maillage  $(\mathcal{T}_{h_n})_{n>0}$ , procède de la façon suivante : on commence par diviser  $\Omega$  en quatre rectangles identiques ( $n = 1$ ), puis on répète la procédure  $n$  fois (chaque rectangle en quatre rectangles identiques) pour obtenir les itérations  $n = 2, 3, 4$ .

Le maillage  $\mathcal{T}_{h_n}$  est ensuite obtenu en divisant chaque rectangle en deux triangles, comme indiqué par la figure 10.

La finesse du maillage  $\mathcal{T}_{h_n}$  est donné par la diagonale du triangle rectangle qui est congruent à toutes les  $2 \cdot 4^n$  mailles de  $\mathcal{T}_{h_n}$ .

Ainsi  $h_n = 2^{-n} \cdot \sqrt{7.25}$ .

Choisissons comme pas en temps  $\tau_n := 2^{-n-2}$ .

Pour  $n \in \{1, \dots, 5\}$ , et pour  $\nu_r = 2^{-r}$  avec  $r \in \{1, \dots, 5\}$ , la solution  $u_{n,r}$  de (9.170) avec  $\mathcal{T}_{h_n}$ ,  $h_n$ ,  $\tau_n$  a été calculée.  $u_r$  est la solution du problème exact pour  $\nu_r = 2^{-r}$ .

Les erreurs considérées sont

$$\max_{1 \leq k \leq 2^{n+2}} \|u_r(\cdot, k \cdot 2^{-n-2}) - u_{n,r}(\cdot, k \cdot 2^{-n-2})\|_2^*$$

et

$$\left[ 2^{-n-2} \cdot \sum_{k=1}^{2^{n+2}} \left( \|u_r(\cdot, k \cdot 2^{-n-2}) - u_{n,r}(\cdot, k \cdot 2^{-n-2})\|_{*V_{h_n}}^* \right)^2 \right]^{1/2},$$

avec  $\|\cdot\|_2^*$  et  $\|\cdot\|_{*V_{h_n}}^*$  définis par

$$\|g\|_2^* := \left( \sum_{K \in \mathfrak{G}_n} |K| \cdot \sum_{i=1}^3 g(P_{K,i})^2 / 3 \right)^{1/2}$$

et

$$\|g\|_{*V_{h_n}}^* := \left( \sum_{K \in \mathcal{T}_{h_n}} |K| \cdot \sum_{i=1}^3 |\nabla(g|K)(P_{K,i})|^2 / 3 \right)^{1/2},$$

respectivement.

( $n \in \mathbb{N}$ ,  $g \in V_{h_n} \oplus H_0^1(\Omega)$ , et  $(P_{K,1}, P_{K,2}, P_{K,3})$  les milieux des côtés de  $K \in \mathcal{T}_{h_n}$ .)

Les formules  $\|\cdot\|_2^*$  et  $\|\cdot\|_{*V_{h_n}}^*$  sont, bien sur, des approximations des normes  $\|\cdot\|_2$  and  $\|\cdot\|_{*V_{h_n}}$ , respectivement.

Les tableaux 1 et 2 ci-dessous contiennent les résultats obtenus. Ils correspondent bien aux majorations obtenues dans le théorème 9.4.

	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
$r = 1$	0.08677280	0.04342670	0.02155141	0.01075595	0.00537527
$r = 2$	0.07341962	0.04172808	0.02262559	0.01197149	0.00619091
$r = 3$	0.15657676	0.09610852	0.05406534	0.02889056	0.01495716
$r = 4$	0.24539138	0.15756216	0.09057426	0.04876265	0.02529267
$r = 5$	0.32471112	0.21346850	0.12563057	0.06827972	0.03551312

TABLE 1 – Erreur  $L^\infty$ - $L^2$ -discrète  $\max_{1 \leq k \leq 2^{n+2}} \|u_r(\cdot, k \cdot 2^{-n-2}) - u_{n,r}(\cdot, k \cdot 2^{-n-2})\|_2^*$ .

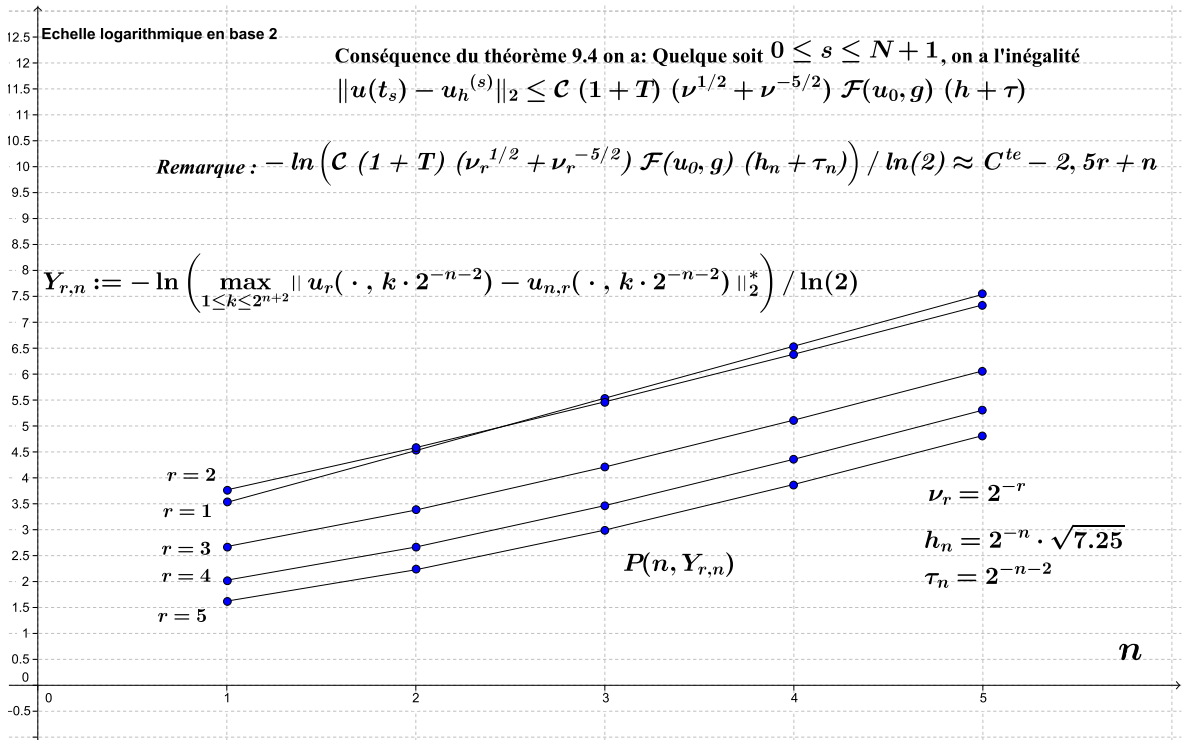


FIGURE 11 – Erreur  $L^\infty$ - $L^2$ -discrète  $\max_{1 \leq k \leq 2^{n+2}} \|u_r(\cdot, k \cdot 2^{-n-2}) - u_{n,r}(\cdot, k \cdot 2^{-n-2})\|_2^*$ .

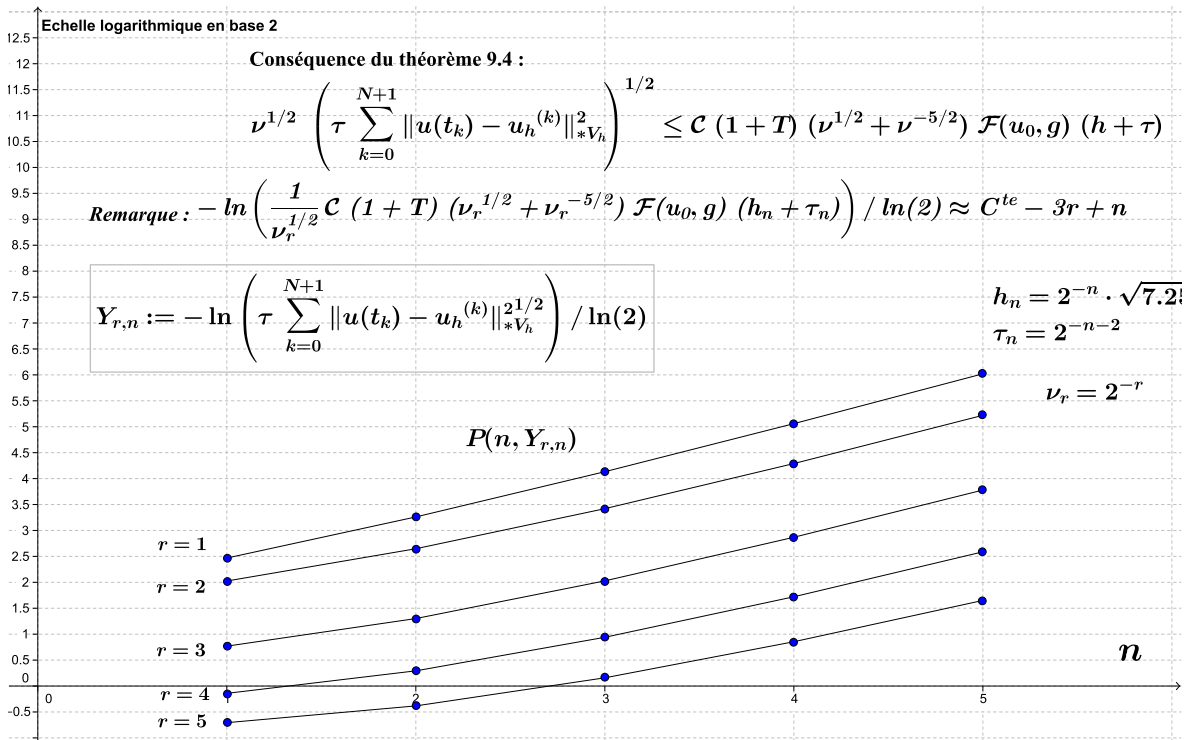


FIGURE 12 – Erreur  $L^2$ - $H^1$ -discrète  $\left[ 2^{-n-2} \cdot \sum_{k=1}^{2^{n+2}} \left( \|u_r(\cdot, k \cdot 2^{-n-2}) - u_{n,r}(\cdot, k \cdot 2^{-n-2})\|_{X_{h_n}}^* \right)^2 \right]^{1/2}$ .

	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
$r = 1$	0.18078114	0.10432310	0.05699424	0.02998321	0.01540407
$r = 2$	0.24571509	0.15958957	0.09342069	0.05111642	0.02683319
$r = 3$	0.58689486	0.40577022	0.24523998	0.13689838	0.07283282
$r = 4$	1.10291847	0.81614571	0.52128707	0.30326142	0.16630205
$r = 5$	1.63203918	1.30402557	0.89593016	0.55312044	0.31848283

TABLE 2 – Erreur  $L^2$ - $H^1$ -discrète  $\left[2^{-n-2} \cdot \sum_{k=1}^{2^{n+2}} \left( \|u_r(\cdot, k \cdot 2^{-n-2}) - u_{n,r}(\cdot, k \cdot 2^{-n-2})\|_{X_{h_n}}^* \right)^2 \right]^{1/2}$ .

## Cinquième partie

# Équation d'advection stationnaire

## Sommaire

---

<b>12</b>	<b>Notations, définitions du graphe orienté <math>\mathcal{G}_h</math> associé au maillage <math>\mathcal{T}_h</math></b>	<b>96</b>
<b>13</b>	<b>Existence et unicité de la solution <math>u_h</math></b>	<b>102</b>
<b>14</b>	<b>Encapsulation de l'information dans <math>\mathcal{G}_h</math></b>	<b>107</b>
<b>15</b>	<b>Étapes et idées principales pour les démonstrations</b>	<b>123</b>
<b>16</b>	<b>Calcul de quelques types de cellules</b>	<b>126</b>
<b>17</b>	<b>Stabilité du problème approché</b>	<b>138</b>
17.1	Majoration de $\ u_h\ _{L_h^\infty(\Omega)}$ . . . . .	138
17.2	Majoration de $\ u_h\ _{L_h^1(\Omega)}$ et $\ u_h\ _{L_h^2(\Omega)}$ . . . . .	143
17.3	Majoration de $\ \beta \cdot \nabla_h u_h\ _{L^1(\Omega)}$ . . . . .	162
17.4	Conclusion : . . . . .	187

---

L'objectif de ce chapitre est de mettre en place le cadre de la théorie des graphes qui sera nécessaire pour la théorie d'existence et de stabilité du problème approché.

Soit  $\emptyset \neq \Omega \subset \mathbb{R}^2$  un domaine polygonal borné, on considère une famille de maillage  $(\mathcal{T}_h)_h$  de  $\Omega$  avec  $h \in ]0; h_0[$ , dont les éléments (mailles)  $T \in \mathcal{T}_h$  sont des triangles fermés.

Soit  $h \in ]0; h_0[$ , alors l'ensemble  $J$  définit une indexation de l'ensemble des côtés  $\{S_i : i \in J\}$  des triangles de  $\mathcal{T}_h$ . On note  $Q_i$  le milieu du segment  $S_i$ . L'ensemble des indices des points qui ne sont pas sur le bord est donné par :

$$J_0 = \{j \in J : Q_j \notin \partial\Omega\} \quad (11.195)$$

Pour le problème d'advection défini par  $0 \neq \beta \in \mathbb{R}^2$ , on aura besoin des ensembles suivants :

$$J^+ = \{j \in J : Q_j \in \Gamma^+\} \quad (11.196)$$

$$J^- = \{j \in J : Q_j \in \Gamma^-\} \quad (11.197)$$

$$J^b = \{j \in J : Q_j \in \Gamma^b\} \quad (11.198)$$

où  $J^+$ ,  $J^-$  et  $J^b$  sont respectivement les ensembles des indices des points  $Q_i$  sur la partie sortante, entrante et longeant du bord.

Pour  $i \in J$  et  $k \in S(i)$  on définit

$$\theta_{i,k} = \int_{\Gamma_{i,k}} \beta \cdot n_{i,k} d\sigma$$

Pour  $i \in J \setminus J^+$ , le voisinage de  $Q_i$  en aval de  $\beta$  est défini par

$$S^+(i) = \{k \in S(i) : \theta_{i,k} \geq 0\}$$

Pour  $i \in J \setminus J^-$ , le voisinage de  $Q_i$  en amont (upwind) de  $\beta$ , est défini par

$$S^-(i) = \{k \in S(i) : \theta_{i,k} \leq 0\}$$

Bien sur  $k \in S^+(i) \cap S^-(i) \iff \theta_{i,k} = 0$

De façon générale, on dira que  $Q_k$  est en amont (en aval de  $Q_i$ ) si  $\theta_{i,k} \leq 0$  ( $\theta_{i,k} \geq 0$ ).

La théorie développée concerne le problème variationnels I défini par

**Définition 11.1 (Formulation variationnelle I du problème approché d'advection)**

Trouver  $u_h \in Y_h$  tel que

$$\hat{b}_h(u_h, v_h) = (f_h, v_h) \quad \forall v_h \in Y_h^0 \text{ avec } u_h(Q_i) = g_h(Q_i) \text{ pour } i \in J^- \quad (11.199)$$

avec

$$(f_h, v_h) = \sum_{k \in J \setminus J^-} f_h(Q_k) v_h(Q_k) |D_k| \quad (11.200)$$

où  $D_k \in \mathcal{D}_h$  est le volume barycentrique associé à  $Q_k$ ,

Pour des raisons techniques, nous aurons besoin aussi de



**Définition 11.2 (Formulation variationnelle II du problème approché d'advection)**

Trouver  $u_h \in Y_h$  tel que

$$\hat{b}_h(u_h, v_h) = (f_h, v_h)_\Pi \quad \forall v_h \in Y_h^0 \quad (11.201)$$

avec  $u_h(Q_i) = g_h(Q_i)$  pour  $i \in J^-$  et

$$(f_h, v_h)_\Pi = \sum_{k \in J \setminus J^-} f_h(Q_k) v_h(Q_k) d_k \quad , \quad d_k = \frac{1}{\|\beta\|^2} \sum_{j \in S^-(k)} [\beta \cdot (Q_k - Q_j)] \theta_{j,k} \quad (11.202)$$

Avec

$$\hat{b}_h(u_h, v_h) = \sum_{i \in J \setminus J^-} v_h(Q_i) \left[ \sum_{j \in S^+(i)} \theta_{ij} u_h(Q_i) + \sum_{j \in S^-(i)} \theta_{ij} u_h(Q_j) \right], \quad \forall u_h \in Y_h, v_h \in Y_h^0 \quad (11.203)$$

## 12 Notations, définitions du graphe orienté $\mathcal{G}_h$ associé au maillage $\mathcal{T}_h$

On reprend les conventions du chapitre 2.3, page 33. De la définition des ensembles  $S^+(i)$  et  $S^-(i)$  on a  $S^+(i) \cup S^-(i) = S(i)$ . Ainsi

$$\sum_{j \in S^-(i)} \theta_{i,j} + \sum_{j \in S^+(i)} \theta_{i,j} = \int_{\partial D_i} \beta \cdot n \, d\sigma = \int_{D_i} \nabla \cdot \beta \, dx = 0 \quad \forall i \in J \quad (12.204)$$

On pose<sup>28</sup> :

**Définition 12.1** ( $\bar{\theta}_i$  et  $\lambda_{j,i}$ ) Pour  $i \in J$  on définit

$$\bar{\theta}_i = \sum_{k \in S^+(i)} \theta_{i,k} \quad (12.205)$$

Pour  $i \in J \setminus J^-$  on définit

$$\lambda_{j,i} = \frac{\theta_{j,i}}{\bar{\theta}_i} \geq 0 \quad \forall j \in S^-(i) \quad (12.206)$$

**Remarque 12.2** Avec les conventions du chapitre 2.3, page 33, on a

$$\bar{\theta}_i = \sum_{k \in S^+(i)} \theta_{i,k} = \sum_{k \in S^-(i)} \theta_{k,i} > 0 \quad \forall i \in J_0 \cup J^b \cup J^+ \quad (12.207)$$

De (12.204) on déduit :

$$\sum_{j \in S^-(i)} \lambda_{j,i} = 1 \quad \text{et} \quad \lambda_{j,i} \bar{\theta}_i = \theta_{j,i} \quad (12.208)$$

---

28.  $\theta_{k,l} = -\theta_{l,k}$ ,  $\theta_{k,l} \geq 0$  si  $k \in S^-(l)$

Pour obtenir le schéma de calcul par itération, on choisit pour  $i \in J \setminus J^-$ , la fonction test  $v_h = w_{h,i} \in Y_h$  définie par  $w_{h,i}(Q_j) = \delta_{ij} = \begin{cases} 1 & \text{si } i=j \\ 0 & \text{si } i \neq j \end{cases}$ . On obtient alors, avec les notations ci-dessus, les récurrences :

$$u_h(Q_i) = \sum_{j \in S^-(i)} u_h(Q_j) \lambda_{j,i} + f_h(Q_i) \frac{|D_i|}{\sum_{j \in S^+(i)} \theta_{i,j}} \quad (12.209)$$

à partir de la formulation I, et

$$u_h(Q_i) = \sum_{j \in S^-(i)} u_h(Q_j) \lambda_{j,i} + f_h(Q_i) \frac{1}{\|\beta\|^2} \sum_{j \in S^-(i)} \beta \cdot (Q_i - Q_j) \lambda_{j,i} \quad (12.210)$$

à partir de la formulation II. Avec ces relations un calcul de  $u_h(Q_i)$  pour  $j \in J \setminus J^-$  par itération semble possible.

Pour unifier l'étude des formulations I et II, posons  $\alpha_i \in \{\alpha_i^I, \alpha_i^{II}\}$  où

$$\alpha_i^I = f_h(Q_i) \frac{|D_i|}{\sum_{j \in S^+(i)} \theta_{i,j}} \quad (12.211)$$

$$\alpha_i^{II} = f_h(Q_i) \frac{1}{\|\beta\|^2} \sum_{j \in S^-(i)} \beta \cdot (Q_i - Q_j) \lambda_{j,i} \quad (12.212)$$

De la régularité et de la quasi-uniformité de la famille de maillage  $(\mathcal{T}_h)_h$ , on peut déduire l'existence d'une constante  $\hat{\sigma} \in ]0; 1[$  ne dépendant pas de  $h \in ]0; h_0[$  telle que

$$h \hat{\sigma} \leq \rho_D = \sup\{\text{diam}(B) \quad : \quad B \subset D\} \quad \forall D \in \mathcal{D}_h \quad (12.213)$$

**Lemme 12.1** *Si  $i \in J$  alors*

$$\bar{\theta}_i = \sum_{j \in S^+(i)} \theta_{i,j} \geq \rho_{D_i} \|\beta\| \quad (12.214)$$

**Preuve :** La figure ci-dessous, qui représente les deux cas possibles pour le volume  $D_i$ , permet de conclure que  $\bar{\theta}_i = \sum_{j \in S^+(i)} \theta_{i,j} \geq \rho_{D_i} \|\beta\|$ .  $\square$

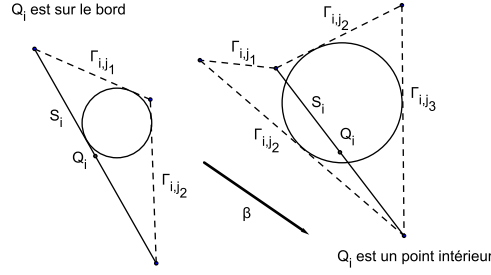
On a alors la majoration :

**Lemme 12.2** *Si  $\alpha_i \in \{\alpha_i^I, \alpha_i^{II}\}$  alors*

$$|\alpha_i| \leq |f_h(Q_i)| \frac{h}{\|\beta\| \hat{\sigma}} \quad (12.215)$$

**Preuve :**

**Cas  $\alpha_i = \alpha_i^I$  :** On a  $\bar{\theta}_i = \sum_{j \in S^+(i)} \theta_{i,j} \geq \rho_{D_i} \|\beta\|$ . Ainsi, avec (12.213) et  $|D_i| \leq h^2$  on



a donc  $|\alpha_i| \leq |f_h(Q_i)| \frac{h}{\|\beta\| \hat{\sigma}}$ .

**Cas  $\alpha_i = \alpha_i^{II}$  :**

$$|\alpha_i^{II}| \leq |f_h(Q_i)| \frac{1}{\|\beta\|^2} \sum_{j \in S^-(i)} \|\beta\| h \lambda_{j,i} \leq |f_h(Q_i)| \frac{h}{\|\beta\| \hat{\sigma}} \text{ car } \sum_{j \in S^-(i)} \lambda_{j,i} = 1 \text{ et } 1 \leq \frac{1}{\hat{\sigma}}.$$

□

La preuve, qui sera faite au chapitre suivant, nécessite encore quelques notations :

### Coordonnées locales upwind

Soit  $T \in \mathcal{T}_h$  avec  $Q_a, Q_b, Q_c$  comme milieux des côtés de  $T$ . Les coordonnées locales upwind  $Q_0^T, Q_1^T$  et  $Q_2^T$ , sont une manière de renommer  $Q_a, Q_b, Q_c$  de façon à disposer de l'information permettant de savoir quel point est en amont de quel autre dans le triangle  $T$ .

Ainsi  $Q_0^T$  est en amont des points  $Q_1^T$  et  $Q_2^T$ ,  $Q_1^T$  est en amont de  $Q_2^T$ . Le point  $Q_2^T$  est donc en aval des points  $Q_0^T$  et  $Q_1^T$ .

Cette notation n'a de sens que si on peut garantir l'absence de "circuit élémentaire"<sup>29</sup> dans  $T$ .

L'existence de telles coordonnées doit être démontrée.

**Définition 12.3 (coordonnées locales upwind)** Soient  $S_a, S_b, S_c$  les côtés de  $T \in \mathcal{T}_h$ . Soit  $G$  le de centre de gravité de  $T$  et  $P_a, P_b$  et  $P_c$  les sommets de  $T$  tels que  $G \in [Q_a, P_a] \cap [Q_b, P_b] \cap [Q_c, P_c]$ .

Il est alors toujours possible de renommer  $Q_a, Q_b, Q_c$  en  $Q_0^T, Q_1^T, Q_2^T$  (pour une certaine permutation de ces points) de telle manière que, pour  $\theta_{i,j}^T = \theta_{Q_i^T, Q_j^T}$ , on ait  $\theta_{0,1}^T \geq 0$ ,  $\theta_{1,2}^T \geq 0$  et  $\theta_{0,2}^T \geq 0$ .

**Preuve :** Voir page 43

□

**Remarque 12.4** La notion de point en amont ou en aval doit être prise au sens large. Si  $\theta_{k,l} = 0$ , les coordonnées locales upwind ne sont pas nécessairement uniques :  $Q_k$  peut

29. Un circuit élémentaire dans le triangle  $T$  signifierait que les points  $Q_a, Q_b, Q_c$  seraient en même temps en aval et en amont des deux autres points.

être considéré comme étant en aval ou en amont de  $Q_l$ . Cela n'a aucune conséquence sur la théorie, pourvu que le graphe  $\mathcal{G}_h$  associé à  $\mathcal{T}_h$  reflète fidèlement les choix opérés lors de la notation upwind locale utilisée.

**Définition 12.5 (Graphe orienté  $\mathcal{G}_h$  associé à  $J$ )** On définit le graphe orienté  $\mathcal{G}_h = (J, \mathcal{F})$  où  $J$  représente l'ensemble des indices des points  $Q_i$  et  $\mathcal{F} \subset J \times J$  l'ensemble des arcs orientés ("flèches") tels que l'arc  $(k, l) \in \mathcal{F}$  si et seulement si

- a)  $Q_l \notin \Gamma^-$  (aucun arc pointe vers un point du bord entant  $\Gamma^-$ ) et
  - b) si  $Q_k = Q_i^T$ ,  $Q_l = Q_j^T$  sont les coordonnées locales upwind alors  $0 \leq i < j \leq 2$  ( $Q_k$  est en amont de  $Q_l$ )
- (12.216)

**Exemple 12.6** L'exemple de maillage ci-dessous montre le graphe  $\mathcal{G}_h$  associé au maillage  $\mathcal{T}_h$ . Les coordonnées locales upwind 0, 1 ou 2 sont indiquées près du point  $Q_i$  concerné. Dans le triangle  $T = P_a P_b P_c$ , les coordonnées locales de  $Q_a$  sont  $Q_2^T$  (donc 2 est écrit dans  $T$  près de  $Q_a$ ), de  $Q_b$  sont  $Q_1^T$  (donc 1 est écrit dans  $T$  près de  $Q_a$ ) et de  $Q_c$  sont  $Q_0^T$  (donc 0 est écrit dans  $T$  près de  $Q_a$ ). Les points de  $\Gamma^-$  sont représentés par des cercles  $\circ$ . On remarquera qu'un arc (représenté en pointillés) a été retiré du graphe  $\mathcal{G}_h$ .

Notation habituelle	notation locale upwind
$Q_a$	$Q_2^T$
$Q_b$	$Q_1^T$
$Q_c$	$Q_0^T$
$\theta_{b,a}$	$\theta_{1,2}^T$
$\theta_{c,a}$	$\theta_{0,2}^T$
$\theta_{c,b}$	$\theta_{0,1}^T$
$\lambda_{b,a}$	$\lambda_{1,2}^T$
$\lambda_{c,a}$	$\lambda_{0,2}^T$
$\lambda_{c,b}$	$\lambda_{0,1}^T$

**Définition 12.7 (chaîne (ou chemin) dans  $\mathcal{G}_h$ )** Une chaîne (ou un chemin) est une suite de sommets telle que chaque sommet est relié au suivant par une arête. La longueur d'une chaîne est le nombre d'arêtes composant cette chaîne.

**Définition 12.8 (circuit élémentaire dans  $\mathcal{G}_h$ )** Un circuit élémentaire<sup>30</sup> est une chaîne fermée (c'est à dire dont l'origine et l'extrémité sont identiques) et dont toutes les arêtes sont distinctes

30. Dans un graphe non orienté, un cycle est une suite d'arêtes consécutives (chaîne) dont les deux sommets extrémités sont identiques. Si la chaîne est élémentaire, c'est-à-dire ne passe pas deux fois par un même sommet, alors on parle de circuit élémentaire. Un circuit élémentaire ne contient pas d'autre cycle. Dans les graphes orientés, comme ici, la notion équivalente est celle de circuit, même si on parle parfois aussi de cycle (par exemple dans l'expression graphe acyclique orienté). L'utilisation du mot boucle pour désigner un circuit est malheureusement à éviter, car en théorie des graphes, une boucle est une arête d'un graphe ayant pour extrémités le même sommet.

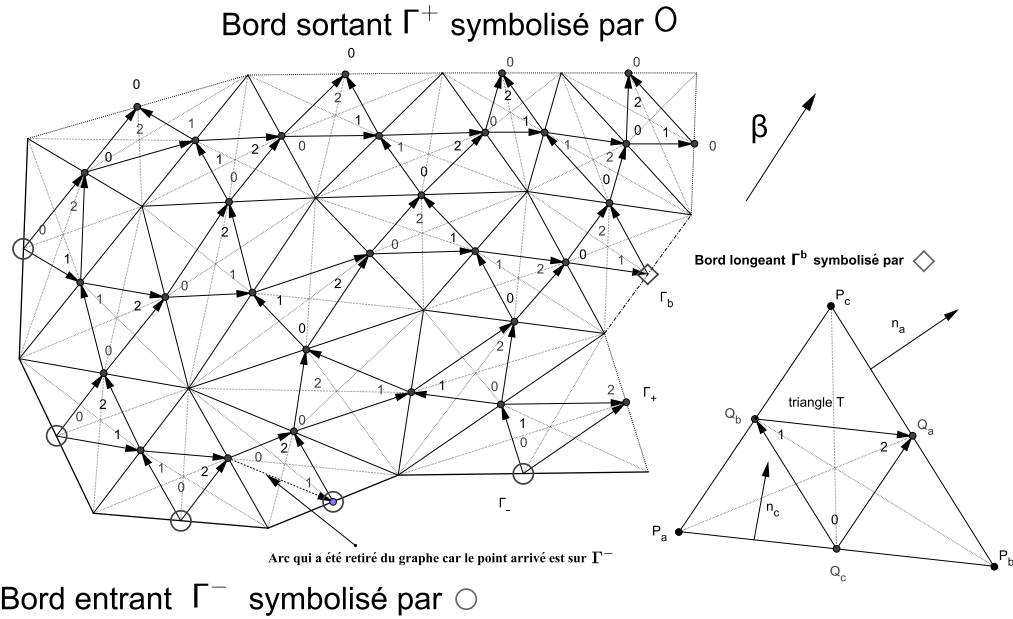
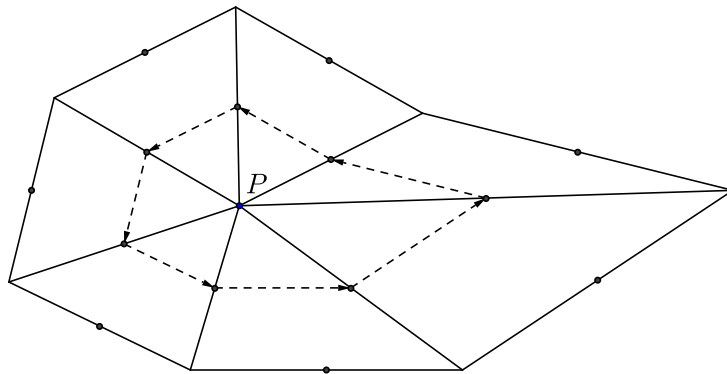


FIGURE 13 –  $\circ$  symbolisent les points de  $\Gamma^-$  (bord entrant),  $\square$  symbolisent les points de  $\Gamma^b$  (bord longéant) et 0 symbolisent les points de  $\Gamma^+$  (bord sortant).

**Définition 12.9 (circuit élémentaire local (ou circuit local) dans  $\mathcal{G}_h$ )** *Un circuit élémentaire local (ou circuit local) est une chaîne fermée autour d'un point, qui est le sommet commun de mailles de  $\mathcal{T}_h$ .*



Circuit élémentaire local (ou circuit local) autour du point  $P$ , sommet commun de plusieurs mailles de  $\mathcal{T}_h$

**Remarque 12.10 (Types de points et prolongement d'un chemin)** *Soit  $\mathcal{C}$ , un chemin dans  $\mathcal{G}_h$ , donné par*

$$Q_{l_0} \rightarrow Q_{l_1} \rightarrow \cdots \overbrace{Q_{l_{k-1}} \rightarrow Q_{l_k}}^{\text{dans le triangle } T_{l_k}} \rightarrow Q_{l_{k+1}} \rightarrow \cdots Q_{l_{s-1}} \rightarrow Q_{l_s}$$

*Ce chemin commence en  $Q_{l_0}$  et finit en  $Q_{l_s}$ .  $\mathcal{C}$  traverse les triangles  $T_{l_1}, T_{l_1}, \dots, T_{l_s}$ . Pour  $1 \leq k < s$ , le point  $Q_{l_k}$  est donc sur l'intersection des triangles  $T_{l_k}$  et  $T_{l_{k+1}}$ . En*

utilisant les coordonnées locales upwind, on a  $Q_{i_k}^{T_{i_k}} = Q_{l_k} = Q_{i_{k+1}}^{T_{i_{k+1}}}$  avec  $i_k \in \{0, 1, 2\}$  et  $i_{k+1} \in \{0, 1, 2\}$ . On dira alors que  $Q_{l_k}$  est de type  $i_k - i_{k+1}$ .

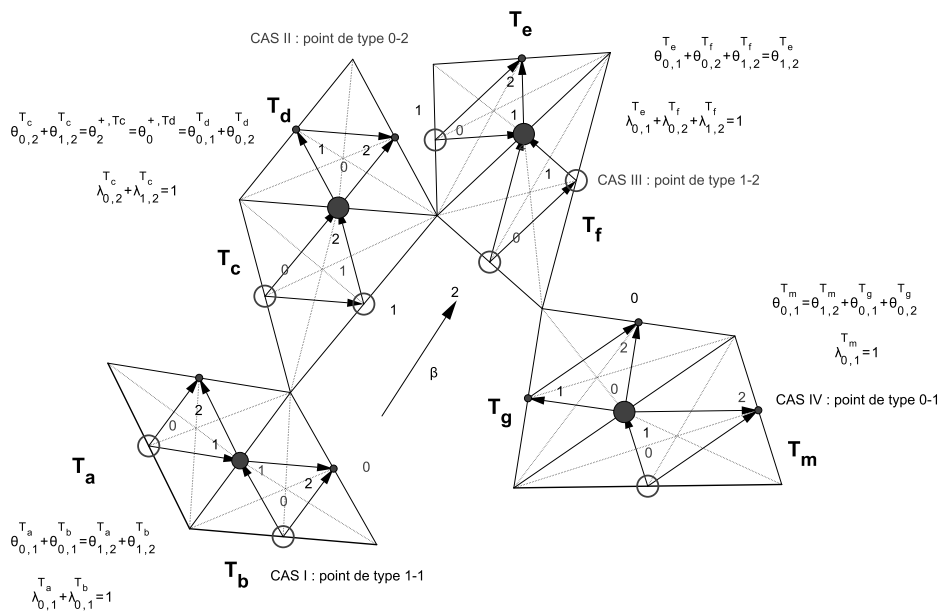
Tous les types de points sont possibles sauf  $0 - 0$  et  $2 - 2$ .

En effet, le types  $0 - 0$  (qui correspond à un puits) et  $2 - 2$  (qui correspond à une source) ne sont pas possibles, car cela contredirait (12.204).

Il est fondamental que ces types ne soient pas possibles, car alors on ne pourrait pas toujours prolonger le chemin  $\mathcal{C}$  en amont de  $Q_{l_0}$  ou en aval de  $Q_{l_s}$ , quand ceux-ci ne sont pas sur le bord.

Dans le problème exact, l'information est transmise le long des courbes caractéristiques et, dans le problème approché, c'est les chemins  $\mathcal{C}$  qui transmettent l'information. Dans le cas du problème exact, tout point de  $\Omega$  doit être atteint par exactement un point de  $\Gamma^-$  en temps fini. Les coordonnées upwind locales et la non-existence des types  $0 - 0$  et  $2 - 2$  sont des conséquences locales du champs (ici constant)  $\beta$ . Ces conclusions devraient être généralisables pour un champs remplissant<sup>31</sup> non constant et suffisamment régulier.

Si  $Q_i$  n'est pas sur le bord de  $\Omega$ , alors  $Q_i$  est sur l'intersection de deux triangles et les différentes combinaisons possibles sont représentées ci-dessous et on distingue quatre types de points :



En conclusion de cette remarque :

**Cas**  $Q_i \in \Omega$  : Alors  $i \in J_0$ ,  $Q_i$  possède au moins un point voisin en amont, et au moins un voisin en aval. Il est donc toujours possible de prolonger  $\mathcal{C}$  en amont de  $Q_i = Q_{l_0}$  et en aval de  $Q_i = Q_{l_s}$ .

**Cas bord entrant** : Si  $Q_i \in \Gamma^-$  ( $i \in J^-$ ), alors  $Q_i$  ne possède pas de voisin en amont. C'est-à-dire que  $\mathcal{C}$  ne peut pas être prolongé en amont de  $Q_i = Q_{l_0}$ , car ces arêtes ont été enlevées de  $\mathcal{G}$ . (Bien sur  $Q_i$  possède toujours un voisin en aval.)

31. c'est-à-dire si pour presque tout  $x \in \Omega$ , il existe une courbe caractéristique du champs de vecteur  $\beta$  partant de  $\Gamma^-$  qui atteint  $x$  en temps fini

**Cas bord longeant :** Si  $Q_i \in \Gamma^b$  ( $i \in J^b$ ), alors  $Q_i$  aura exactement un voisin en aval et un voisin en amont (voir la figure page 100).

**Cas bord sortant :** Si  $Q_i \in \Gamma^+$  ( $i \in J^+$ ), alors  $Q_i$  aura toujours au moins un voisin en amont. Il est cependant possible qu'existe un voisin en aval (voir dans la figure page 100 le point de  $\Gamma^+$  au dessus de  $\square$ ).

### 13 Existence et unicité de la solution $u_h$

L'existence de  $u_h$  est garantie si le graphe  $\mathcal{G}_h$  ne contient pas de circuit élémentaire (voir le théorème 13.3).

Il est donc important de pouvoir préciser sous quelles conditions géométriques du maillage  $\mathcal{T}_h$ , un circuit élémentaire est possible. L'argument qui va être développé est le suivant :

Si un circuit élémentaire existe dans  $\mathcal{G}_h$ , alors il existe un circuit local.

Ainsi, si aucun circuit local existe, il ne peut y avoir de circuit élémentaire dans  $\mathcal{G}_h$ .

Le lemme suivant en précise les conditions d'existence :

**Lemme 13.1 (Conditions d'existence de circuit local dans  $\mathcal{G}_h$ )** Soit  $\beta^T$  la rotation de  $\beta$  de  $\pi/2$  dans le sens contraire des aiguilles d'une montre.

Si  $\mathcal{G}_h$  contient un circuit élémentaire, alors il existe un circuit local, c'est-à-dire un point  $P \in \Omega$ , sommet d'un triangle  $T \in \mathcal{T}_h$ , tel que pour l'ensemble des triangles  $T_{i_0}, T_{i_1}, \dots, T_{i_{n-2}}$  ayant aussi  $P$  comme sommet et numérotés dans le sens contraire des aiguilles d'une montre, on ait :

$$\theta_{i_j, i_{j+1}} = \vec{\beta}^T \cdot P\vec{G}_{i_j} < 0 \quad \forall j \in \{0, 1, \dots, n-2\},$$

*(pour un circuit élémentaire orientée dans le sens des aiguilles d'une montre)*

ou

$$\theta_{i_j, i_{j+1}} = \vec{\beta}^T \cdot P\vec{G}_{i_j} > 0 \quad \forall j \in \{0, 1, \dots, n-2\},$$

*(pour un circuit élémentaire orientée dans le sens contraire des aiguilles d'une montre)*

(  $G_j$  étant le centre de gravité du triangle  $T_j = PP_{i_j}P_{i_{j+1}}$ . Voir figure 14 )

C'est-à-dire que  $\cup_{j \in \{i_0, i_1, \dots, i_{n-1}\}} T_j$  contient un circuit élémentaire.

**Preuve :** Les relations ci-dessus ne font qu'exprimer que :

a) soit  $Q_{i_j}$  est en aval de  $Q_{i_{j+1}}$  (c'est-à-dire  $\theta_{i_j, i_{j+1}} < 0$ ) pour  $j \in \{0, 1, \dots, n-2\}$

b) soit  $Q_{i_j}$  est en amont de  $Q_{i_{j+1}}$  (c'est-à-dire  $\theta_{i_j, i_{j+1}} > 0$ ) pour  $j \in \{0, 1, \dots, n-2\}$ ,

L'égalité  $\theta_{i_j, i_{j+1}} = \vec{\beta}^T \cdot P\vec{G}_{i_j}$  s'obtient en prenant la définition de

$$\theta_{i_j, i_{j+1}} = \beta \cdot n_{i_j, i_{j+1}} |PG_{i_j}| = \vec{\beta}^T \cdot P\vec{G}_{i_j}$$

et en utilisant le fait que le produit scalaire est invariant par la rotation d'angle  $\pi/2$ .  $\square$

Il suffit donc de s'assurer que pour tous les points  $P \in \Omega$ , les conditions locales (13.217) et (13.217) ne soient jamais vérifiées pour garantir que  $\mathcal{G}_h$  ne contient pas de circuit élémentaire.

La figure 14 illustre géométriquement la condition (13.217).

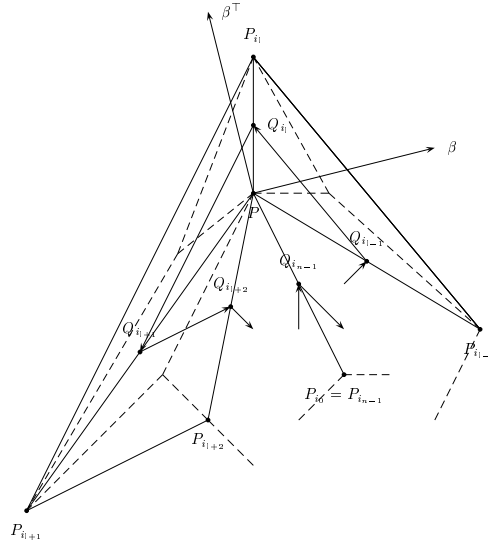


FIGURE 14 – Représentation géométrique de (13.217) avec apparition d'un circuit élémentaire dans le sens positif autour de  $P$

**Lemme 13.2 (Existence de circuit élémentaire dans  $\mathcal{G}_h$ )** *Si  $\mathcal{G}_h$  contient un circuit élémentaire, alors il existe un circuit local.*

**Preuve du lemme :**

La preuve se fait en trois étapes et s'appuie sur la remarque 12.10 qui précise les conditions sous lesquelles un chemin peut être prolongé en amont ou en aval :

**Etape 1 :** Soit  $Q_{i_0} \rightarrow Q_{i_1} \rightarrow Q_{i_2} \rightarrow \dots \rightarrow Q_{i_l} \rightarrow Q_{i_{l+1}} \rightarrow \dots \rightarrow Q_{i_{n-1}} = Q_{i_0}$  un circuit élémentaire  $\mathcal{B}$  dans  $\Omega$  (donc  $(i_{l-1}, i_l) \in \mathcal{F}$ ). Sans restriction, on peut supposer que les points  $Q_{i_l}$  sont distincts deux à deux (car sinon il existerait un circuit élémentaire plus petit, qu'il serait possible de choisir). Supposons qu'il existe un triangle  $T$  à l'intérieur de ce circuit élémentaire (dans le sens d'une courbe de Jordan) et qui ne soit pas traversé par lui (aucun sommet de  $\mathcal{B}$  est dans  $T$ ), alors il existerait, sur ce triangle, un point  $Q_k$  qui ne serait pas sur  $\mathcal{B}$  et qui ne serait pas sur le bord de  $\Omega$ . Si on prolonge  $Q_k$  dans les deux directions par des points en amont et par des points en aval, en une courbe  $\mathcal{C}$ , on a pour  $\mathcal{C}$  les trois possibilités suivantes :



1. La courbe  $\mathcal{C}$  ne coupe jamais  $\mathcal{B}$  et contient donc un circuit élémentaire strictement plus petit contenu à l'intérieur de  $\mathcal{B}$  (il n'y a qu'un nombre fini de points  $Q_s$  avec  $s \in J_0$  dans le domaine défini par l'intérieur de  $\mathcal{B}$ ). Dans ce cas on répète l'étape 1 jusqu'à ce qu'il n'y ait plus de triangle à l'intérieur de ce circuit élémentaire qui ne soit pas un triangle traversé par celui-ci.
2. La courbe coupe  $\mathcal{C}$  ne coupe jamais  $\mathcal{B}$  en amont de  $Q_k$ . En appliquant le raisonnement ci-dessus à cette partie de  $\mathcal{C}$ , on peut conclure, qu'elle doit contenir un circuit élémentaire et on se retrouve dans la situation d'un circuit élémentaire strictement plus petit contenu dans l'intérieur de  $\mathcal{B}$ .
3. La courbe coupe  $\mathcal{C}$  ne coupe jamais  $\mathcal{B}$  en aval de  $Q_k$ . En appliquant le raisonnement ci-dessus à cette partie de  $\mathcal{C}$ , on peut conclure, qu'elle doit contenir un circuit élémentaire et on se retrouve dans la situation d'un circuit élémentaire strictement plus petit contenu dans l'intérieur de  $\mathcal{B}$ .
4. La courbe  $\mathcal{C}$  coupe  $\mathcal{B}$  en deux points (en amont et en aval de  $Q_k$ ). Il est alors toujours possible de trouver par collage d'une partie de  $\mathcal{C}$  avec une partie de  $\mathcal{B}$  une courbe strictement plus petite que  $\mathcal{B}$ . (Voir la figure 15)

A la fin de la première étape on a donc trouvé un circuit élémentaire  $\mathcal{B}$  dont l'intérieur ne contient aucun triangle qui ne soit pas un triangle traversé par ce circuit. (Voir la figure 16)

**Étape 2 :** Prenons au hasard deux triangles traversés par  $\mathcal{B}$ , ayant un côté en commun  $S_k$  qui ne soit pas traversé par  $\mathcal{B}$ .<sup>32</sup> Voir la figure 16.

$Q_k$  est le milieu de du côté  $S_k$ . Comme  $Q_k$  ne peut pas être de type  $0 - 0$  ou  $2 - 2$ ,  $Q_k$  possède un voisin en aval et un voisin en amont de  $Q_k$ . Ces voisins sont des points de  $\mathcal{B}$ . Il est possible de trouver  $r \in \{i_1, i_2\}$  et  $s \in \{i_3, i_4\}$  tel que  $Q_s \rightarrow Q_k \rightarrow Q_r$  ou  $Q_r \rightarrow Q_k \rightarrow Q_s$ .

Ainsi, il est donc possible de trouver un circuit élémentaire strictement plus petit que  $\mathcal{B}$  et ce processus peut être répété autant de fois qu'un tel segment  $S_k$  existe.

A la fin de cette opération, on obtient un circuit élémentaire autour d'un point  $P \in \Omega$  du maillage. C'est exactement la situation de la figure 14.

cqfd

**Remarque 13.1** *On dispose d'un critère local pour tester si  $\mathcal{G}_h$  contient un circuit élémentaire ou non. Une condition suffisante pour ne pas avoir de circuit élémentaire dans  $\mathcal{G}_h$  serait de demander que les volumes barycentriques  $D_i$  soient convexes pour tout  $i \in J_0$ .*

Nous allons conclure cette partie par un résultat qui relie l'absence de circuit élémentaire avec l'existence d'un schéma de calcul par itérations. Ce théorème est l'équivalent du théorème d'existence (utilisant la méthode des caractéristiques) pour le problème exact :

**Théorème 13.3 (Existence et unicité de  $u_h$ )** *Soit  $(\mathcal{T}_h)_h$  une famille de maillage régulière telle que  $\mathcal{G}_h$  ne contient pas de circuit élémentaire alors pour le problème (11.199), on a :*

---

<sup>32</sup>. Ces segments non traversés par  $\mathcal{B}$  sont en pointillés dans la figure 16.

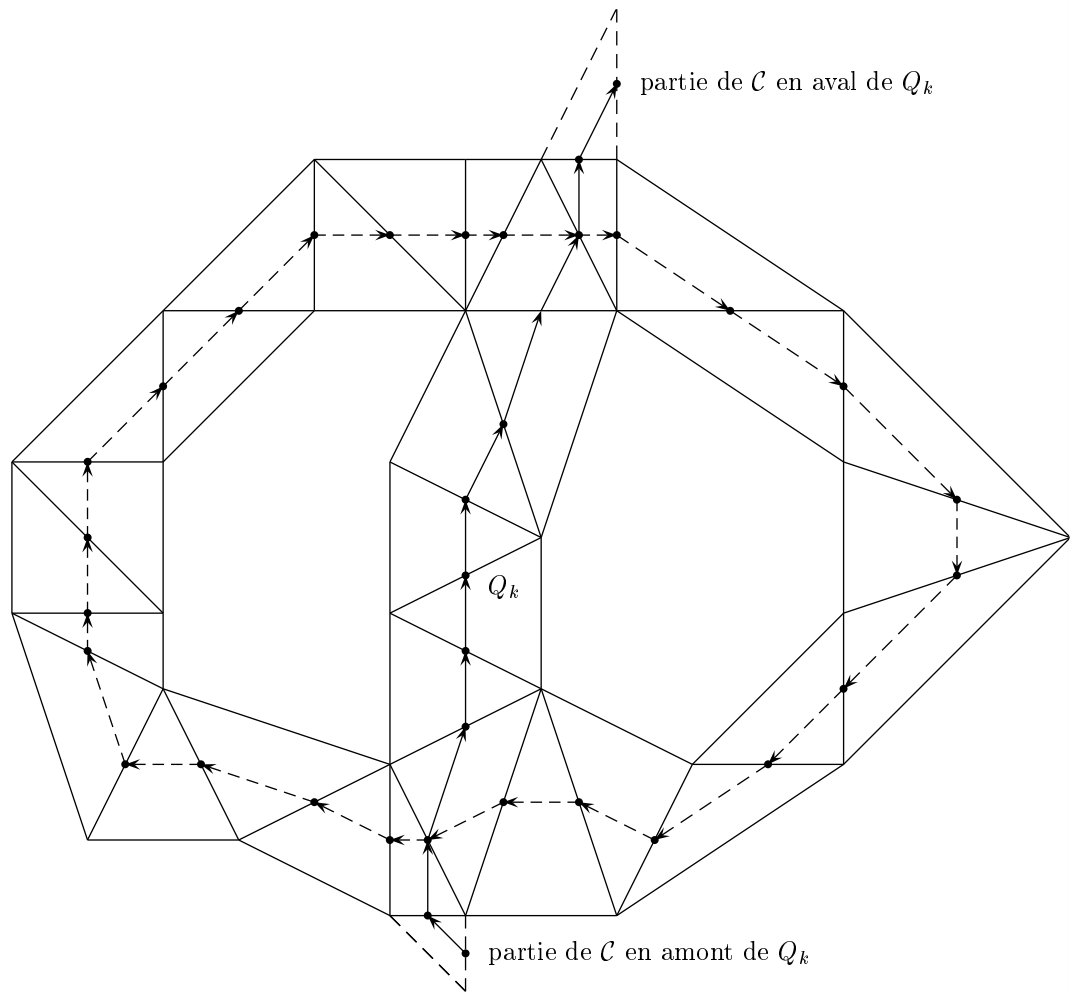


FIGURE 15 – Situation où  $\mathcal{C}$  a deux intersections en amont et en aval de  $Q_k$  avec  $\mathcal{B}$

**Existence :** *Il existe un schéma itératif permettant de calculer  $u_h$ ,*

**Unicité :**  *$u_h$  est alors l'unique solution de (11.199)*

**Preuve de l'existence de  $u_h$  :**

Nous allons montrer qu'il existe une numérotation  $i_1, i_2, \dots, i_k, \dots, i_{|J|}$  de  $J$  telle que

$J^- = \{i_1, i_2, \dots, i_k, \dots, i_{|J|}\}$  et

$$S^-(i_k) \subset \{i_1, \dots, i_{k-1}\}, \forall k \in \{|J^-| + 1, \dots, |J|\} \quad (13.217)$$

Avec cette numérotation, (12.209) livre un schéma itératif permettant de calculer  $u_h$ .

La preuve va se faire par induction.

**Début de l'induction :** Supposons qu'il n'existe pas de point  $Q_t$  tel que  $S^-(t) \subset \{i_1, i_2, \dots, i_k, \dots, i_{|J|}\} = J^-$ . Alors pour  $t_0 \in J \setminus J^-$  quelconque, il existe

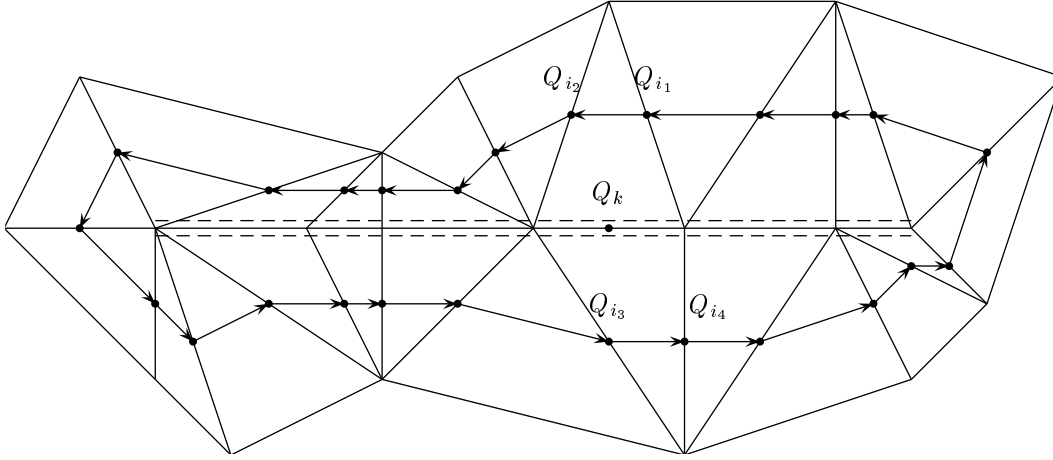


FIGURE 16 – Les pointillés indiquent les faces intérieures des triangles qui ont un côté en commun, mais dont le milieu n'est pas un point de  $\mathcal{B}$ .

$t_{-1}$  avec  $t_{-1} \in (J \setminus J^-) \cap S^-(t_0)$ . (On utilise ici le fait que tout point de  $J \setminus J^-$  possède un voisin en amont.) De même pour  $t_{-1}$ , il existe  $t_{-2} \in (J \setminus J^-) \cap S^-(t_{-1})$ , etc ... Comme  $|J \setminus J^-| < \infty$ , le chemin ainsi construit va avoir une intersection avec lui-même en  $Q_{t-k} = Q_{t-l}$  pour deux valeurs  $k, l \in \{1, \dots, |J \setminus J^-|\}$ . Sans restriction, soit  $k > l$ , alors le chemin  $t_{-k} \rightarrow t_{-k+1} \rightarrow \dots \rightarrow t_{-l-1} \rightarrow t_{-l}$  est un circuit élémentaire, ce qui contredit l'hypothèse d'absence de circuit élémentaire. Il existe donc bien un point  $Q_t \in J \setminus J^-$  tel que  $S^-(t) \subset J^-$ . Numérotons  $i_{|J^-|+1}$  ce point.

**Induction sur  $n$  :**

Supposons que les points  $i_1, \dots, i_n$  aient déjà été numérotés pour  $n > |J^-|$ . Donc  $S^-(i_k) \subset \{i_1, \dots, i_{k-1}\}$  pour  $k \in \{|J^-| + 1, \dots, n\}$ .

Le raisonnement est maintenant le même que pour le début de l'induction : On fait de nouveau l'hypothèse que pour tout point  $Q_t$  avec  $t \in J \setminus \{i_1, \dots, i_n\}$  on ait  $S^-(t) \not\subset \{i_1, \dots, i_n\}$  et on peut conclure alors à l'existence d'un circuit élémentaire dans  $\mathcal{G}_h$ , ce qui est une contradiction.

**Fin :** Bien sûr si  $n = |J|$  l'algorithme s'arrête et on a fini.

**Preuve de l'unicité de  $u_h$  :**

Si  $u_h$  et  $\tilde{u}_h$  sont deux solutions de (12.209), alors  $v_h = u_h - \tilde{u}_h$  est une solution de (12.209) avec  $f_h = 0$  et  $v_h(Q_l) = 0$  pour  $l \in J^-$ .

Comme  $\mathcal{G}_h$  ne contient pas de circuit élémentaire,  $v_h$  existe et (12.209) devient

$$v_h(Q_i) = \sum_{j \in S^-(i)} v_h(Q_j) \lambda_{j,i} \text{ avec } v_h(Q_l) = 0 \text{ pour } l \in J^- \quad (13.218)$$

Donc  $v_h(Q_t) = 0$  pour  $t \in J$ , c'est-à-dire  $u_h = \tilde{u}_h$ .

**cqfd**

**Remarque 13.2**

Même pour un champs constant, une mauvaise configuration géométrique peut donc induire l'existence d'un circuit élémentaire, et pourrait ainsi compromettre l'existence et l'unicité de la solution  $u_h$  du problème approché. Il n'est cependant pas évident pourquoi la condition suffisante – l'absence de circuit élémentaire – est aussi une condition nécessaire.

**14 Encapsulation de l'information dans  $\mathcal{G}_h$** 

Les coordonnées locales upwind jouent ici un rôle central, car, pour chaque triangle du maillage, elles contiennent des informations sur le graphe  $\mathcal{G}_h$ . Si  $Q_i$  est le milieu du segment commun à deux triangles, alors  $Q_i$  a deux coordonnées locales (une coordonnée  $\{0, 1, 2\}$  par triangle).

Se dégage ainsi la typologie des points du graphe  $\mathcal{G}_h$  présentée à la remarque 12.10. Dans cette typologie (0 – 1, 1 – 1, 1 – 2, 0 – 2) les types 0 – 1 et 0 – 2 sont particulièrement intéressants.

En effet, si  $Q_i$  est un point de type 0 – 1 ou 0 – 2, cela signifie que

**$Q_i$  et les voisins en amont de  $Q_i$  sont sur un même triangle (maille de  $\mathcal{T}_h$ .)**

Ceci est bien sûr faux pour les points de type 1 – 2 et 1 – 1 où les valeurs des points de deux triangles différents sont nécessaires au calcul de  $u_h(Q_i)$ .

Les points de types 0 – 1 et 0 – 2 permettent donc de faire apparaître l'encapsulation de l'information due à la structure de  $\mathcal{G}_h$ .

Cette encapsulation peut être rendue visible en traçant en gras les segments des points  $Q_i$  de type 0 – 1 et 0 – 2, comme cela est représenté dans la figure 17, qui reprend l'exemple 12.6. A ces segments, il faut ajouter, par convention :

- a) les segments sur  $\Gamma^-$  qui sont considérés comme étant de type 0 – 2,
- b) les segments de  $\Gamma^b$  de type 0 – 1,
- c) les segments sur  $\Gamma^+$  de type 0 – 1 ou 0 – 2 selon les coordonnées upwind locales du point.

**Remarque 14.1 (Triangle ayant deux coordonnées locales 0)** *Si pour un point  $Q_i \in \Gamma^-$ , on a retiré de  $\mathcal{G}_h$  l'arc adjacent (orienté) qui pointe vers  $Q_i$  - comme le demande la définition de  $\mathcal{G}_h$  - on reste cohérent avec a). Le triangle auquel appartient cet arc représente la seule exception d'un triangle ayant deux points de coordonnées upwind locales 0. La coordonnée locale modifiée sera alors notée  $\mathbb{X}0$ .*

Apparaissent ainsi des réunions de 1 ou plusieurs triangles (dans la figure 17 on a au plus deux triangles, mais davantage sont possibles) dont le bord est constitué de segments dont les milieux sont des points de types 0 – 1 ou 0 – 2.

Ces réunions de triangles qui encapsulent l'information localement vont jouer un rôle central dans ce qui suit.

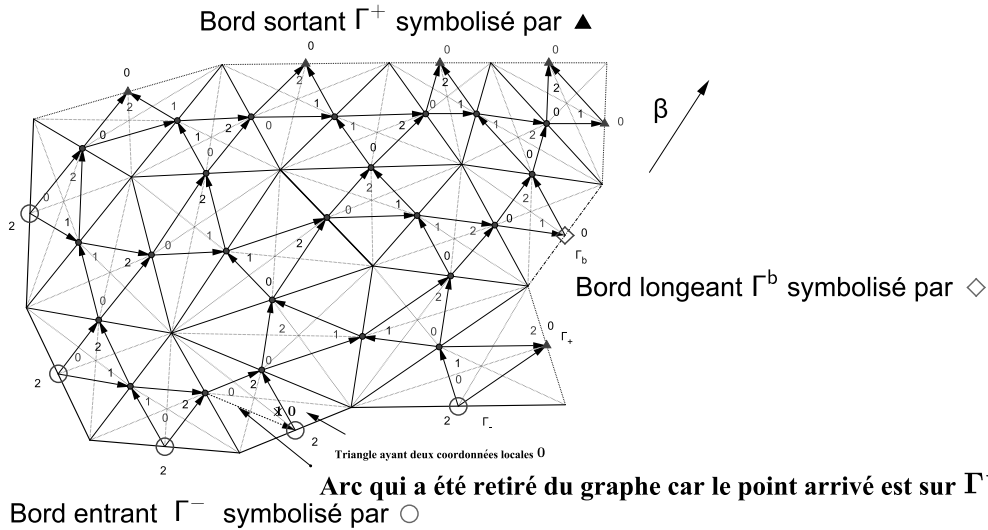


FIGURE 17 – En traits épais, les segments correspondants aux points de types 0 – 1 et 0 – 2. A ces segments, il faut ajouter : a) Les segments sur  $\Gamma^-$  qui sont considérés comme étant de type 0 – 2. (Cela est cohérent avec le point  $Q_i$  de  $\Gamma^-$  auquel un arc adjacent a été retiré, car il pointait sur le bord. Le triangle auquel appartient cet arc représente la seule exception d'un triangle ayant deux points de coordonnées upwind locales 0), b) les segments de  $\Gamma^b$  de type 0 – 1, et c) les segments sur  $\Gamma^+$  de type 0 – 1 ou 0 – 2 selon les coordonnées upwind locales du point.

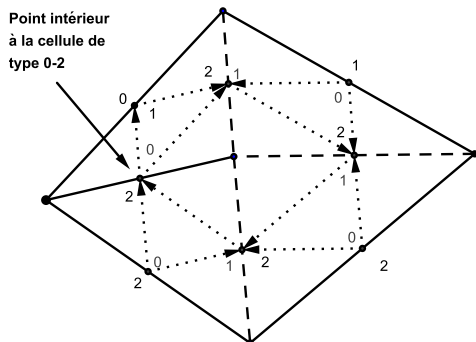
**Définition 14.2 (cellule)** Une réunion  $C$  de triangles de  $\mathcal{T}_h$ , qui forme un ensemble connexe, est appelé une cellule, si le bord de  $C$  est la réunion de segments  $S_i$  ( $i \in J$ ) dont les milieux  $Q_i$  sont des points de type 0 – 1 ou 0 – 2. On note  $\mathcal{C}_h$  l'ensemble des cellules qui composent  $\Omega$ , donc

$$\Omega = \cup_{C \in \mathcal{C}_h} C$$

On note  $N_C$  le nombre de triangles de  $\mathcal{T}_h$  dans  $C$ .

**Remarque 14.3** 1) L'ensemble  $\mathcal{C}_h$  est donc, par construction, unique.

2) Le terme "point intérieur" est pris au sens topologique. Par exemple, la cellule  $C$  représentée ci-dessous est formée de quatre mailles. Le bord topologique consiste en quatre segments de type 0 – 1 ou 0 – 2. Cependant, on ne peut pas en général conclure que les points intérieurs d'une cellule sont de type 1 – 1 ou 1 – 2, comme le montre la figure ci-dessous.



Les triangles du maillage sont délimités par les segments du bord.  
 Si on retire les segments dont les milieux sont de type 1-1 (marqués par des tirés - - - -) on obtient des assemblages connexes de triangles (les cellules) dont le bord topologique est de type 0-1 ou 0-2.  
 Dans la figure ci-contre, une cellule est représentée dont un point intérieur est de type 0-2.  
 Cette situation n'est pas typique des cellules obtenues par des maillages réguliers.  
 Une étude systématique des cellules qui ne contiennent pas de point intérieur de type 0-1 et 0-2 sera menée ci-dessous.

Dans l'étude des cellules, l'hypothèse de travail par la suite sera **d'exclure des points intérieurs de type 0 – 1 et 0 – 2 dans les cellules.**

Avec cette hypothèse, il est possible de procéder à une classification et un calcul de tous les types de cellules possibles et, de plus, le graphe  $\mathcal{G}_h$  n'a alors plus de circuit élémentaire.

3) L'exemple ci-dessus a un caractère hypothétique dans le sens où sa configuration géométrique (qui respecte les coordonnées locales données) n'est peut-être pas réalisable. Les quelques tentatives pour construire (ou invalider) un tel cas "atypique" n'ont pas abouties. Un tel exemple permettrait peut-être de savoir si les questions suivantes sont reliées :

- a) Il existe un circuit local, mais pas de circuit élémentaire plus grand (alors qu'un circuit élémentaire entraîne toujours l'existence d'un circuit local).
- b) Il existe une cellule avec un point intérieur de type 0 – 1 ou 0 – 2.

Pour exprimer la conservation de certaines quantités dans une cellule  $C \in \mathcal{C}_h$ , il est nécessaire d'introduire les quantités  $\hat{\theta}_i$  pour les points  $Q_i$  sur le bord de  $C$  :

**Définition 14.4** ( $\hat{\theta}_i$ ) Soit  $Q_i$  un point de type 0–1 ou 0–2 et  $T \in \mathcal{T}_h$ , tel que  $Q_i = Q_0^T$  (on choisit le triangle où  $Q_i$  a la coordonnée locale upwind 0), alors, par définition de la coordonnée locale upwind,  $\{Q_j : j \in S^-(i)\} \cap \bar{T} = \emptyset$ , et on définit

$$\hat{\theta}_i = \sum_{\substack{j \in S^+(i) \\ \text{et } Q_j \in \bar{T}}} \theta_{i,j} \quad (14.219)$$

Si  $Q_i \in \Gamma^+$ , alors  $\hat{\theta}_i = \theta_{i,+\infty} = \int_{S_i} (\beta \cdot n_\Omega) d\sigma$ .

En coordonnées locales upwind on notera  $\hat{\theta}_0^T = \hat{\theta}_i$ .

**Définition 14.5 (points d'entrée (ou de contrôle), points de sortie, d'une cellule)**

Soit  $C$  une cellule de  $\mathcal{C}_h$ , on note  $J^C = \{i \in J : Q_i \in C\}$ .

Pour  $i \in J^C$  et  $Q_i \in \partial C$  un point du bord de  $C$ , on définit

- a)  $Q_i$  est un point entrant (ou de contrôle) si  $S^-(i) \cap J^C = \emptyset$ . Tous les points en amont de  $Q_i$  sont en dehors de  $C$  ou n'existent pas (si  $Q_i \in \Gamma^-$ ). La coordonnée locale upwind 0 se trouve sur le triangle inclus dans  $C$ .
- b)  $Q_i$  est un point sortant si  $S^+(i) \setminus J^C \neq \emptyset$ . Au moins un point en aval n'appartient pas à  $C$ , mais aucun point en amont de  $Q_i$  ne peut être à l'extérieur de la cellule  $C$ . La coordonnée locale upwind 0 se trouve sur le triangle extérieur à  $C$ .

L'ensemble des indices des points entrants est noté  $J_{\text{entrant}}^C$  et celui des indices des points sortants, est noté  $J_{\text{sortant}}^C$ . On note  $J_{\text{intérieur}}^C = J^C \setminus (J_{\text{entrant}}^C \cup J_{\text{sortant}}^C)$  les points intérieurs à  $C$ .

**Remarque 14.6 a)** Si  $\mathcal{G}_h$  ne contient pas de circuit élémentaire, alors les valeurs de  $u_h$  pour des points entrants  $J_{\text{entrant}}^C$ , permettent de calculer toutes les autres valeurs de  $u_h$  dans  $C$ . C'est le sens même de l'encapsulation de l'information permise par une cellule<sup>33</sup>.

33. On peut aussi appliquer le théorème d'existence à la cellule  $C$ .

b) Pour les points  $Q_i \in \Gamma^b \cap \partial C$  du bord longeant, on a donc  $i \in J_{\text{intérieur}}^C$ .

c) •  $\{Q_l : l \in J_{\text{entrant}}^C \cup J_{\text{sortant}}^C\} \subset \partial C$ ,

• si  $l \in J_{\text{entrant}}^C$  alors  $\beta \cdot n_C(Q_l) < 0$  (voir la figure page 113),

• si  $l \in J_{\text{sortant}}^C$  alors  $\beta \cdot n_C(Q_l) \geq 0$ ,

Cette encapsulation a des conséquences sur le bilan de certaines intégrales sur le bord :

d) Les points de  $J_{\text{entrant}}^C$  sont de type 0 – 1 ou 0 – 2. Cela signifie pour le triangle  $T$  (inclus dans  $C$ ) de la figure ci-dessous, que (en coordonnées locales upwind) :

– Si  $T$  contient deux coordonnées locales upwind 0, alors

$$\hat{\theta}_{\mathbf{x}_0}^T = \theta_{\mathbf{x}_0,2}^T \leq - \int_{S_{\mathbf{x}_0}^T} \beta \cdot n_C d\sigma \quad \text{et} \quad \hat{\theta}_0^T = \theta_{0,2}^T \leq - \int_{S_0^T} \beta \cdot n_C d\sigma \quad (14.220)$$

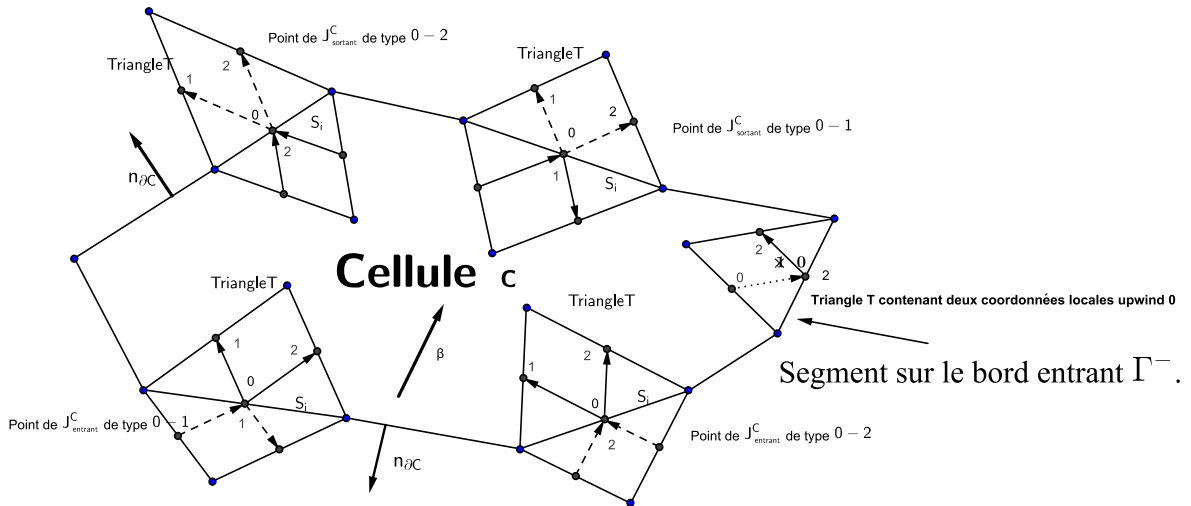
– Sinon

$$\hat{\theta}_0^T = \theta_{0,1}^T + \theta_{0,2}^T = - \int_{S_i} \beta \cdot n_C d\sigma \quad (14.221)$$

e) Les points de  $J_{\text{sortant}}^C$  sont de type 0 – 1 ou 0 – 2. Cela signifie, pour le triangle  $T$  (non inclus dans  $C$ ) de la figure ci-dessous, que (en coordonnées locales upwind) :

$$\hat{\theta}_0^T = \theta_{0,1}^T + \theta_{0,2}^T = \int_{S_i} \beta \cdot n_C d\sigma \quad (14.222)$$

Excepté le cas de  $T$  ayant deux coordonnées locales upwind 0, on a ,  $\hat{\theta}_i = \left| \int_{S_i} \beta \cdot n_C d\sigma \right|$ .



Le lemme suivant montre encore une conséquence intéressante de l'encapsulation au niveau d'une cellule :

**Lemme 14.1 (encapsulation)** Soit  $C_h$  l'ensemble des cellules définies à partir de  $\mathcal{T}_h$  et si  $\mathcal{G}_h$  ne contient pas de circuit élémentaire alors

**Formulation I** pour  $u_h$  solution de (11.199), on a

$$\sum_{k \in J^C \setminus J_{entrant}^C} |D_k| f_k = - \sum_{l \in J_{entrant}^C} u_l \hat{\theta}_l + \sum_{l \in J_{sortant}^C} u_l \hat{\theta}_l \quad (14.223)$$

En particulier, si  $C \in \mathcal{C}_h$  est une cellule qui ne contient pas de triangle ayant deux coordonnées locales 0 on a (voir la remarque 14.1) :

$$\sum_{k \in J^C \setminus J_{entrant}^C} |D_k| f_k = \sum_{l \in J_{entrant}^C \cup J_{sortant}^C} u_l \int_{S_l} \beta \cdot n_C d\sigma \quad (14.224)$$

**Formulation II** pour  $u_h$  solution de (11.201), on a

$$\sum_{k \in J^C \setminus J_{entrant}^C} d_k f_k = - \sum_{l \in J_{entrant}^C} u_l \hat{\theta}_l + \sum_{l \in J_{sortant}^C} u_l \hat{\theta}_l \quad (14.225)$$

En particulier, si  $C \in \mathcal{C}_h$  est une cellule qui ne contient pas de triangle ayant deux coordonnées locales 0 on a (voir la remarque 14.1) :

$$\sum_{k \in J^C \setminus J_{entrant}^C} d_k f_k = \sum_{l \in J_{entrant}^C \cup J_{sortant}^C} u_l \int_{S_l} \beta \cdot n_C d\sigma \quad (14.226)$$

où  $u_i = u_h(Q_i)$ , pour  $i \in J$ .

**Preuve :**

La preuve, pour les deux formulations, est presque identique : il suffit de remplacer le symbole  $|D_k|$  par  $d_k$  dans la preuve ci-dessous pour avoir le résultat de la formulation I. On prouve la formule pour la formulation II.

Soit  $\omega = \{(l, i) \in \mathcal{F} : l \in J^C \text{ et } i \in J^C\}$  l'ensemble des arcs  $(l, i)$  du graphe  $\mathcal{G}_h$  qui sont dans la cellule  $C$ .

Soit

$$\omega^+ = \{(i, l) \in \mathcal{F} : i \in J^C \setminus J_{entrant}^C \text{ et } l \in S^+(i)\} \cup \{(i, l) \in \mathcal{F} : i \in J_{entrant}^C \text{ et } l \in S^+(i) \cap J^C\}$$

Il s'agit donc de l'ensemble des arcs  $(i, l)$  du graphe  $\mathcal{G}_h$  qui

- partent d'un point de  $J^C \setminus J_{entrant}^C$ , et
- partent d'un point de  $J_{entrant}^C$ , mais qui restent dans  $C$ .

On a  $\omega \subset \omega^+$  et  $\omega^+ \setminus \omega$  est l'ensemble des arcs  $(i, l)$  du graphe  $\mathcal{G}_h$  avec  $i \in J_{sortant}^C$  qui quittent  $C$  ( $l \notin J^C$ ).

• Pour  $\omega^+ \setminus \omega$  :

$$\begin{aligned} \sum_{(i,l) \in \omega^+ \setminus \omega} u_i \theta_{i,l} &= \sum_{i \in J_{sortant}^C} \sum_{l \in S^+(i) \setminus J^C} u_i \theta_{i,l} \\ &= \sum_{i \in J_{sortant}^C} u_i \left( \sum_{l \in S^+(i) \setminus J^C} \theta_{i,l} \right) \\ &\stackrel{\text{remarque 14.6, e)}}{=} \sum_{i \in J_{sortant}^C} u_i \hat{\theta}_i \end{aligned} \quad (14.227)$$



- Pour  $\omega^+$  :

$$\begin{aligned} \sum_{(i,l) \in \omega^+} u_i \theta_{i,l} &= \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{l \in S^+(i)} u_i \theta_{i,l} + \sum_{i \in J_{\text{entrant}}^C} \sum_{l \in S^+(i) \cap J^C} u_i \theta_{i,l} \\ &\stackrel{\text{remarque 14.6, d)}}{=} \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{l \in S^+(i)} u_i \theta_{i,l} + \sum_{i \in J_{\text{entrant}}^C} u_i \left( -\hat{\theta}_i \right) \end{aligned} \quad (14.228)$$

- Pour  $\omega$  on a  $\omega = \tilde{\omega}$  avec :

$$\begin{aligned} \omega &= \{(l, i) \in \mathcal{F} : i \in J^C \text{ et } l \in J^C\} \\ \tilde{\omega} &= \{(k, i) \in \mathcal{F} : i \in J^C \setminus J_{\text{entrant}}^C \text{ et } k \in S^-(i)\} \end{aligned} \quad (14.229)$$

On effectue la preuve par double inclusion :

" $\subset$ " : Soit  $(l, i) \in \omega$ , alors  $i$  ne peut pas être dans  $J_{\text{entrant}}^C$ , car sinon  $k \in S^-(i)$  ne peut pas être<sup>34</sup> dans  $J^C$ . Ainsi  $i \in J^C \setminus J_{\text{entrant}}^C$  et  $S^-(i) \subset J^C$ , ce qui prouve " $\subset$ ".

" $\supset$ " : Soit  $(l, i) \in \{(k, i) \in \mathcal{F} : i \in J^C \setminus J_{\text{entrant}}^C \text{ et } k \in S^-(i)\}$ .

Si  $i \notin J_{\text{sortant}}^C$ , alors il n'y a rien à montrer.

Si  $i \in J_{\text{sortant}}^C$ , alors  $l \in J^C$  car  $Q_i$  est de type 0-1 ou 0-2 ce qui signifie qu'aucun point en amont de  $Q_i$  ne peut être à l'extérieur de  $C$ .

Avec " $\supset$ " et " $\subset$ " l'égalité (14.229) est démontrée.

$u_h$  est solution de (11.201). Soit  $i \in J \setminus J^-$ , alors pour  $v_h(Q_j) = 1$  si  $j = i$  et 0 sinon, on obtient avec (11.203) et (11.201) :

$$\sum_{j \in S^+(i)} \theta_{ij} u_i + \sum_{j \in S^-(i)} \theta_{ij} u_j = f_i |D_i| \quad (14.230)$$

Maintenant  $\sum_{j \in S^+(i)} \theta_{ij} + \sum_{j \in S^-(i)} \theta_{ij} = 0$ , donc (multiplication par  $u_i$  et  $\theta_{ij} = -\theta_{ji}$ )  $\sum_{j \in S^+(i)} u_i \theta_{ij} = -\sum_{j \in S^-(i)} u_i \theta_{ij} = \sum_{j \in S^-(i)} u_i \theta_{ji}$ , et (14.230) devient

$$\sum_{j \in S^-(i)} (u_i - u_j) \theta_{j,i} = f_i |D_i| \quad (14.231)$$

---

34. On utilise ici le fait que  $i$  est alors de type 0-1 ou 0-2, ce qui signifie qu'aucun point en amont de  $Q_i$  ne peut être à l'intérieur de  $C$ .

$$\begin{aligned}
\sum_{i \in J^C \setminus J_{\text{entrant}}^C} f_i |D_i| &= \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{k \in S^-(i)} [u_i - u_k] \theta_{k,i} \\
&= \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{k \in S^-(i)} u_i \theta_{k,i} - \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{k \in S^-(i)} u_k \theta_{k,i} \\
&\quad \text{avec } \sum_{k \in S^-(i)} \theta_{k,i} = - \sum_{k \in S^-(i)} \theta_{i,k} = \sum_{k \in S^+(i)} \theta_{i,k} \text{ on a :} \\
&= \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{k \in S^+(i)} u_i \theta_{i,k} - \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{k \in S^-(i)} u_k \theta_{k,i} \\
&\stackrel{(14.229)}{=} \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{k \in S^+(i)} u_i \theta_{i,k} - \sum_{(k,i) \in \omega} u_k \theta_{k,i} \\
&\stackrel{(14.228)}{=} \sum_{(i,l) \in \omega^+} u_i \theta_{i,l} + \sum_{i \in J_{\text{entrant}}^C} u_i (-\hat{\theta}_i) - \sum_{(k,i) \in \omega} u_k \theta_{k,i} \\
&\stackrel{(14.227)}{=} \sum_{i \in J_{\text{sortant}}^C} u_i \hat{\theta}_i + \sum_{i \in J_{\text{entrant}}^C} u_i (-\hat{\theta}_i) \\
&= - \sum_{i \in J_{\text{entrant}}^C} u_i \hat{\theta}_i + \sum_{i \in J_{\text{sortant}}^C} u_i \hat{\theta}_i
\end{aligned}$$

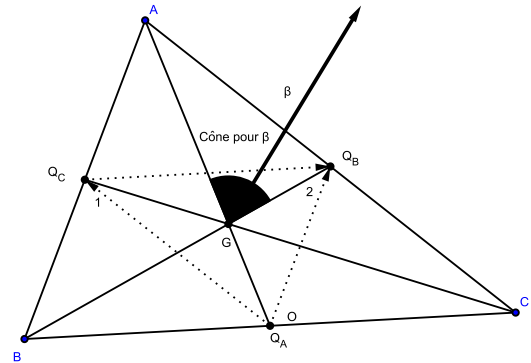
□

### Structure des cellules :

La structure des cellules est restreinte par :

**Contrainte 1 :** l'impossibilité des types  $0 - 0$  et  $2 - 2$ , et le fait que chaque triangle (sauf ceux du bord  $\Gamma^-$  dont un arc a été enlevé et qui ont comme coordonnées locales upwind deux 0 et un 2) ait 0, 1 et 2 comme coordonnées locales upwind.

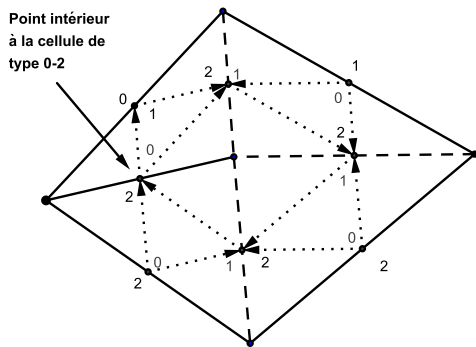
**Contrainte 2 :** la disposition possible du triangle  $T \in \mathcal{T}_h$ , due à  $\beta$ . La figure ci-contre montre le cône de directions possibles pour  $\beta$ , si les coordonnées locales sont données.



Pour chercher les combinaisons possibles pour une cellule, le plus simple est de commencer par la contrainte 1 pour faire un premier tri de ce qui est possible. Ensuite, sur ce qui reste, d'utiliser la contrainte 2.

Cependant, une étude complète qui permettrait de classer tous les assemblages possibles de triangles en fonction de la géométrie du maillage (condition d'angle) n'a pas pu être

obtenue. Par exemple la cellule ci-dessous respecte la contrainte 1, mais il n'est pas évident de prouver si, oui ou non, une réalisation qui vérifierait la contrainte 2 soit possible.



Les triangles du maillage sont délimités par les segments du bord.  
Si on retire les segments dont les milieux sont de type 1-1 (marqués par des tirés - - - - -) on obtient des assemblages connexes de triangles (les cellules) dont le bord topologique est de type 0-1 ou 0-2.

Dans la figure ci-contre, une cellule est représentée dont un point intérieur est de type 0-2.

Cette situation n'est pas typique des cellules obtenues par des maillages réguliers.

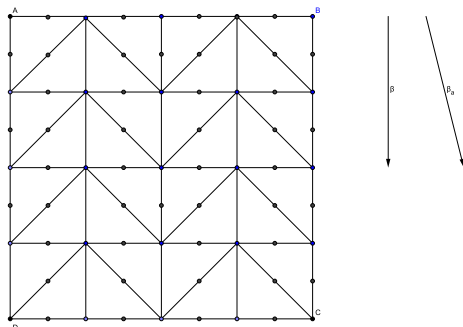
Une étude systématique des cellules qui ne contiennent pas de point intérieur de type 0-1 et 0-2 sera menée ci-dessous.

En l'occurrence, la figure ci-dessus n'est pas possible, si le graphe  $\mathcal{G}_h$ , associé à  $\mathcal{T}_h$ , ne contient pas de circuit élémentaire.

Une classification plus facile, et suffisante pour la suite, est possible, si on suppose que  $J_{\text{intérieur}}^C = J^C \setminus (J_{\text{sortant}}^C \cup J_{\text{entrant}}^C)$  - les points donc, qui ne sont pas sur le bord de  $C$  - ne contient aucun point de type 0 - 1 ou 0 - 2, c'est-à-dire que tous les points intérieurs de  $C$  sont de type 1 - 1 ou 1 - 2 (ce qui n'est pas le cas de la figure ci-dessus).

Avant de présenter le lemme, un exemple pour illustrer les différentes étapes pour, à partir du maillage, aboutir aux différents type de cellules :

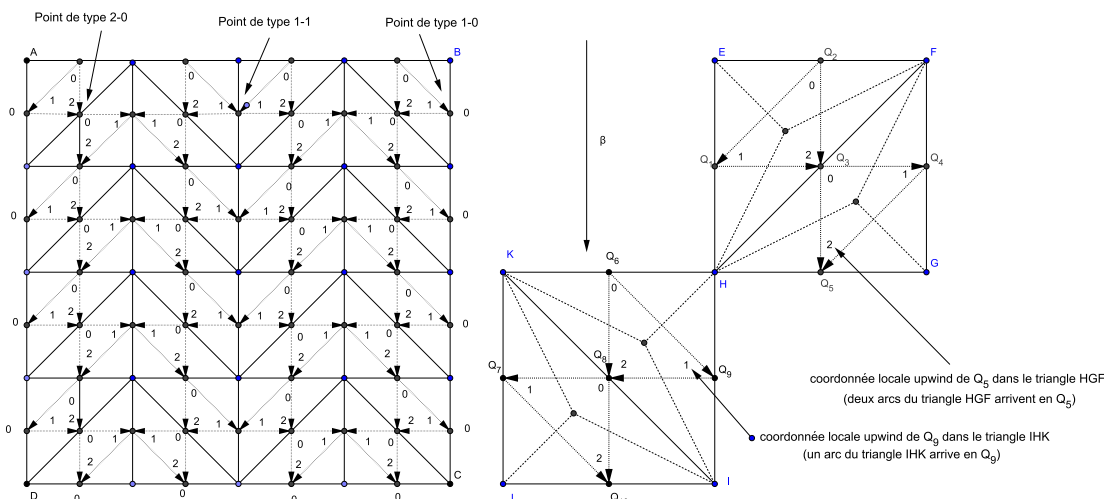
**Exemple 14.7** On considère  $\Omega = ]0; 1[ \times ]0; 1[$  et la famille de maillage  $(\mathcal{T}_h)_{h>0}$  obtenue en décomposant le carré  $\Omega$  en  $(2N) \times (2N)$  petits carrés, puis chaque carré en deux triangles, comme le montre la figure ci-dessous pour  $N = 2$ .



La finesse du maillage est alors  $h = \frac{\sqrt{2}}{2N}$ . Deux vecteurs  $\beta$  et  $\beta_a$  représentés sur la figure sont considérés ici.

**Cas du vecteur  $\beta$  longeant le bord AD :** Le bord entrant est donné par les points du segment  $[AB]$  (donc  $\Gamma^- = [AB]$ ).

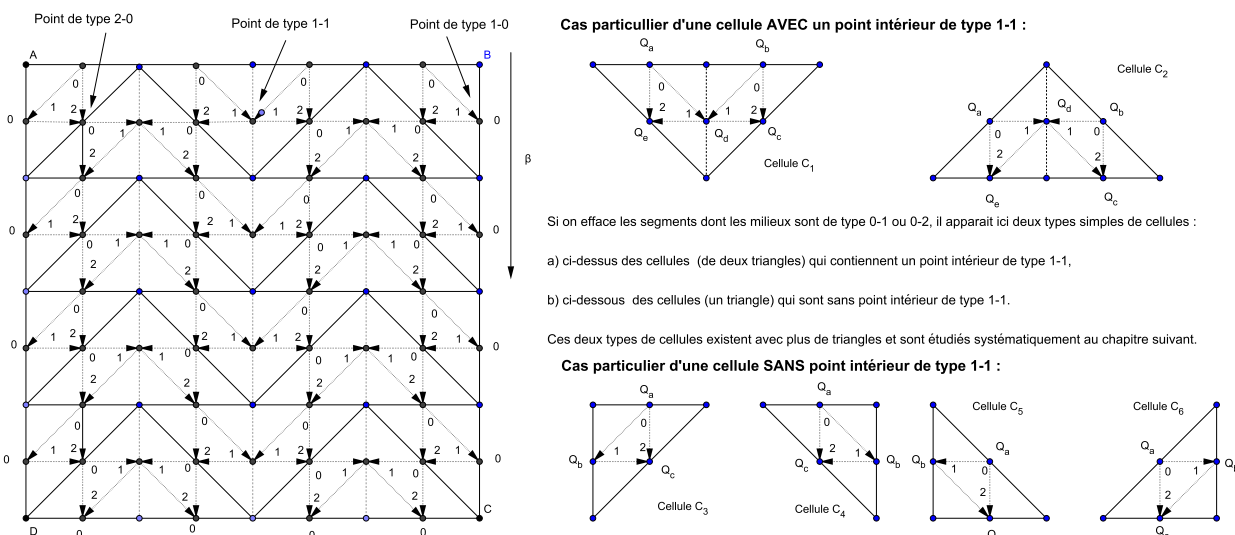
**Étape 1** La première étape consiste à faire apparaître le graphe  $\mathcal{G}_h$  associé à  $(\mathcal{T}_h)$ . On obtient le graphe orienté suivant :



**Etape 2** Faire apparaître les coordonnées locales upwind : Pour un triangle  $T$  du maillage donné, le nombre d'arcs orientés (dans  $T$ ) qui pointent sur le milieu d'un des côtés  $Q_i$  de  $T$  représente la coordonnée locale upwind de  $Q_i$  dans  $T$ . Ce nombre est placé dans  $T$  à côté de  $Q_i$  (voir la figure ci-dessus). On calcule facilement ( $h = EF = EH$ ) :  $\theta_{Q_2, Q_1} = \frac{h\|\beta\|}{3}$ ,  $\theta_{Q_1, Q_3} = \frac{h\|\beta\|}{3}$  et  $\theta_{Q_2, Q_3} = \frac{2h\|\beta\|}{3}$ .

**Etape 2** Faire apparaître le type de chaque point : Les points intérieurs  $Q_i$ , qui sont donc communs à deux triangles de  $T_h$ , ont donc deux coordonnées locales upwind (une pour chaque triangle). Ces deux coordonnées locales définissent le type du point  $Q_i$ . Les combinaisons possibles étant 0 – 2 (ou 2 – 0), 0 – 1 (ou 1 – 0), 1 – 1 et 1 – 2 (ou 2 – 1). Le type 1 – 2 n'apparaît pas ici, mais est parfaitement concevable.

**Etape 3** Faire apparaître les cellules  $C$  de  $\mathcal{C}_h$  en effaçant les segments de type 1 – 1 (et de type 1 – 2 si il y en avait) . On obtient la figure suivante :



Pour les points du bord longeant ( $[AD]$  et  $[BC]$ ) et sortant ( $[DC]$ ), la coordonnée locale upwind est complétée par 0, ce qui nous donne deux types possibles sur le bord ( 0 – 2 et 0 – 1).

Il est possible de calculer explicitement  $d_k = \frac{1}{\|\beta\|^2} \sum_{j \in S^-(k)} \theta_{j,k} [\beta \cdot (Q_k - Q_j)]$  :

Cellule $C_i =$	$C_1$	$C_2$	$C_3$ et $C_4$	$C_5$ et $C_6$
	$J^{C_i} = \{a, b, c, d, e\}$	$J^{C_i} = \{a, b, c, d, e\}$	$J^{C_i} = \{a, b, c\}$	$J^{C_i} = \{a, b, c\}$
	$J_{\text{entrant}}^{C_i} = \{a, b\}$	$J_{\text{entrant}}^{C_i} = \{a, b\}$	$J_{\text{entrant}}^{C_i} = \{a\}$	$J_{\text{entrant}}^{C_i} = \{a\}$
$J^{C_1} \setminus J_{\text{entrant}}^{C_1} =$	$\{c, d, e\}$	$\{c, d, e\}$	$\{b, c\}$	$\{b, c\}$
Pour $Q_b :$	-	-	$S^-(b) = \{a\}$	$S^-(b) = \{a\}$
Pour $Q_c :$	$S^-(c) = \{d, b\}$	$S^-(c) = \{d, b\}$	$S^-(c) = \{a, b\}$	$S^-(c) = \{a, b\}$
Pour $Q_d :$	$S^-(d) = \{a, b\}$	$S^-(d) = \{a, b\}$	-	-
Pour $Q_e :$	$S^-(e) = \{a, d\}$	$S^-(e) = \{a, d\}$	-	-
$d_b =$	-	-	$\frac{1}{\ \beta\ ^2} \frac{h\ \beta\ }{3} \frac{h\ \beta\ }{2}$	0
$d_c =$	$\frac{1}{\ \beta\ ^2} \frac{2h\ \beta\ }{3} \frac{h\ \beta\ }{2}$	$\frac{1}{\ \beta\ ^2} \left( \frac{h\ \beta\ }{3} + \frac{2h\ \beta\ }{3} \right) \frac{h\ \beta\ }{2}$	$\frac{1}{\ \beta\ ^2} \frac{2h\ \beta\ }{3} \frac{h\ \beta\ }{2}$	$\frac{1}{\ \beta\ ^2} \left( \frac{h\ \beta\ }{3} + \frac{2h\ \beta\ }{3} \right) \frac{h\ \beta\ }{2}$
$d_d =$	$\frac{1}{\ \beta\ ^2} \left( \frac{h\ \beta\ }{3} + \frac{h\ \beta\ }{3} \right) \frac{h\ \beta\ }{2}$	0	-	-
$d_e =$	$\frac{1}{\ \beta\ ^2} \frac{2h\ \beta\ }{3} \frac{h\ \beta\ }{2}$	$\frac{1}{\ \beta\ ^2} \left( \frac{h\ \beta\ }{3} + \frac{2h\ \beta\ }{3} \right) \frac{h\ \beta\ }{2}$	-	-
$\sum_{k \in J^{C_1} \setminus J_{\text{entrant}}^{C_1}} d_k =$	$h^2 =  C_i $	$h^2 =  C_i $	$\frac{1}{2} h^2 =  C_i $	$\frac{1}{2} h^2 =  C_i $

### Conclusions :

- a) Les cellules du maillages ci-dessus ne contiennent pas de point intérieur de type 0 – 1 ou 0 – 2 (contrairement à la situation de la figure page 114).

Cette situation est typique des cellules obtenues par des maillages réguliers et dont une étude systématique sera menée ci-dessous. Le maillage présenté ici entre parfaitement dans le cadre de la théorie de stabilité qui sera développée dans les chapitres suivants.

- b) On retrouve sur la dernière ligne du tableau ci-dessus le cas particulier du lemme C.1.

### Remarque 14.8 (Forme linéaire $(\cdot, \cdot)_{\text{II}}$ , produit scalaire et condition $\delta - S^+$ )

On constate dans l'exemple ci-dessus que beaucoup de coefficients s'annulent. La forme linéaire n'est donc pas un produit scalaire (d'ailleurs rien n'empêcherait  $d_k$  de prendre des valeurs négatives).

**Si la fonction  $f_h$  est quelconque, il est possible d'avoir une solution qualitativement incohérente avec l'équation (1.4) en prenant la formulation variationnelle II) :**

Soit  $a > 0$  et  $f_{h,a} \in Y_h$  définit par  $f_{h,a}(Q_i) = a$  si  $Q_i$  est le point  $Q_d$  dans les cellules  $C_2$  et sinon  $f_h(Q_i) = 0$ . Soit  $g = g_h = 0$ . Alors,  $(f_{h,a}, v_h)_{\text{II}} = 0$  pour toute fonction  $v_h \in Y_h^0$  (car  $d_d = 0$ ), et on déduit que  $u_{h,a} = 0$  est la solution du problème variationnel II défini par (17.331).

Soit  $U_{h,a}$  la solution du problème exact (1.4), alors  $\beta \cdot \nabla U_{h,a} = f_{h,a}$  et

$$|U_{h,a}|_{Y_{1,h}} = \|\beta \cdot \nabla U_{h,a}\|_{L^1(\Omega)} = \|f_{h,a}\|_{L^1(\Omega)} = a \|f_{h,1}\|_{L^1(\Omega)} \rightarrow +\infty$$

pour  $a \rightarrow +\infty$ .

Cette situation peut être évitée on demande qu'il existe  $\delta > 0$ , indépendant de la famille de maillage, tel que

$$\sum_{j \in S_{\beta}^-(i)} \theta_{j,i}^{\beta} [\beta \cdot (Q_i - Q_j)] \geq \delta \|\beta\|^2 |S_i|^2 (> 0) \quad \forall j \in J \setminus J^- \quad (14.232)$$

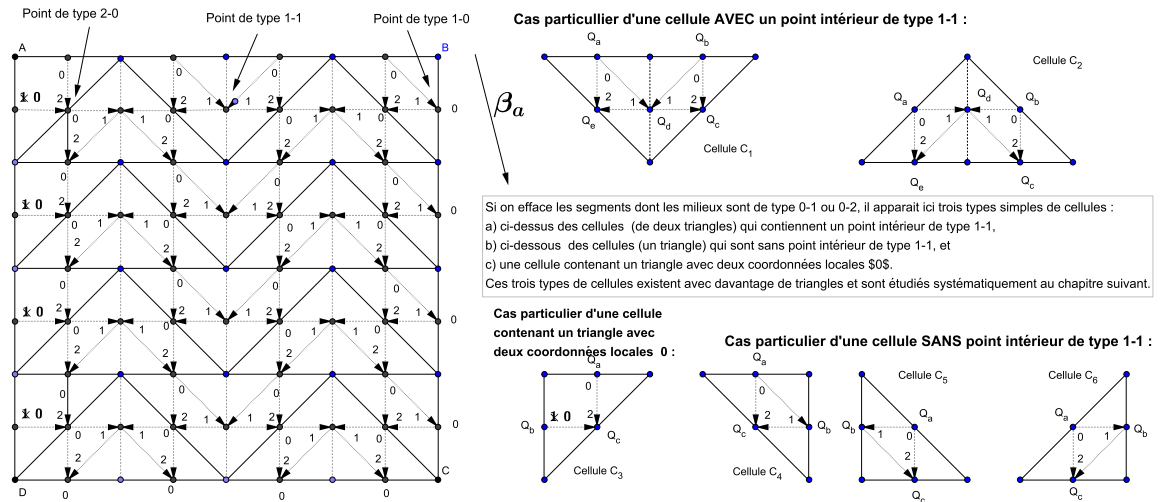
Ce sera la condition  $\delta - S^-$ , qui interviendra de façon cruciale au chapitre 17.3. Pour les points intérieurs, c'est-à-dire  $Q_i$  avec  $i \in J_0$ , la condition  $\delta - S^-$  est équivalente à la condition  $\delta - S^+$  (voir la remarque 17.4) et ne nécessite donc pas d'étude particulière.

**Cas du vecteur  $\beta_a$  faisant un petit angle entrant avec le bord AD :** Dans le

cas précédent ( $\beta$  longeant le côté  $[AD]$ ) deux types de cellules sont apparus<sup>35</sup>. Ce sont :

- a) Les cellules AVEC un point intérieur de type 1 – 1 (cellules  $C_1$  et  $C_2$ ), dont une étude complète sera faite au chapitre 16, et
- b) Les cellules SANS point intérieur de type 1 – 1 (cellules  $C_3$  à  $C_6$ ), dont une étude complète sera faite au chapitre 16.

Le choix du vecteur  $\beta_a$  comme champ de vecteur à pour conséquence de changer  $\Gamma^- = [AB] \cup [AD]$ . Les valeurs de  $u_h(Q_i)$ , pour  $Q_i \in [AD]$  sont donc connues par les conditions au bord entrant et, pour respecter cela, il est nécessaire de retirer de  $G_h$  les arcs orientés pointant vers  $Q_i \in [AD]$ , comme le montre la figure ci-dessous



On voit donc apparaître un troisième type de cellules :

- c) Les cellules contenant un triangle avec deux coordonnées locales 0, dont une étude complète sera faite au chapitre 16.

**Conclusion :** Si les cellules  $C \in \mathcal{C}_h$  ne contiennent pas (au sens topologique) de point intérieur de type 0 – 1 ou 0 – 2, alors on constate l'existence de trois types de cellules a), b) et c). Le lemme ci-dessous prouve que ce sont les seuls possibles.

**Lemme 14.2 (Classification des cellules sans point intérieur de type 0 – 1 ou 0 – 2)**

Soit  $C = \cup_{j=1, \dots, k} T_j \in \mathcal{C}_h$  une cellule, avec au moins deux triangles, telle que  $J_{intérieure}^C$  ne contient aucun point de type 0 – 1 ou 0 – 2, alors

- 1)  $|J_{sortant}^C| = 2$ ,
- 2) Tous les triangles  $T_j$  ont un sommet  $P = \cap_{j=1, \dots, k} T_j$  en commun,
- 3) Il y a au plus un point de type 1 – 1 dans  $J_{intérieure}^C$ .

(Voir la figure 18, page 120)

**Preuve :** A) On suppose qu'il existe un point de type 1 – 1 dans  $C$  :

35. Ces cellules ne contiennent pas de point intérieur de type 0 – 1 ou 0 – 2. Contrairement à ce qui se passe dans la cellule représentée à la page 114

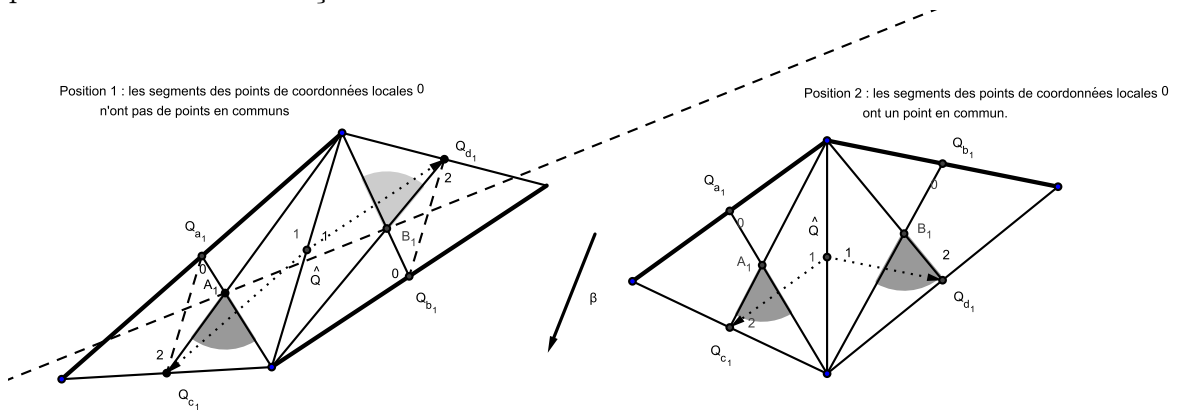
**Initialisation :** Supposons que  $J_{\text{intérieur}}^C$  possède un point de type 1–1. Notons ce point  $\hat{Q}$ , et les deux triangles qui ont ce point en commun  $T_{j_1}$  et  $T_{k_1}$ .

$$\underbrace{? \leftarrow 1}_{T_{j_1}} - \hat{Q} - \underbrace{1 \rightarrow ?}_{T_{k_1}}$$

**Pas 1 :** Les extrémités des arcs orientés issus de  $\hat{Q}$  ne peuvent être que des points de coordonnées locales 2 (car il ne peut y avoir deux points de coordonnées locales upwind 0 dans un même triangle). Donc

$$\underbrace{2 \leftarrow 1}_{T_{j_1}} - \hat{Q} - \underbrace{1 \rightarrow 2}_{T_{k_1}}$$

Les segments contenant les points  $Q_{a_1}$  et  $Q_{b_1}$  de coordonnées locales 0 peuvent se positionner de deux façons :



La position 1, n'est pas possible, car l'intersection des deux cônes (qui doivent contenir la direction de  $\beta$ ) est vide. La seule position possible est la position 2. On a :  $a_1, b_1 \in J_{\text{entrant}}^C$  car les points  $Q_{a_1}$  et  $Q_{b_1}$  sont de type 0–?, les coordonnées locale 0 sont sur des triangles inclus dans  $C$ , et  $J_{\text{intérieur}}^C$  ne contient que des points de type 1–2 ou 1–1.

**Itération :** Quatre cas peuvent se présenter :

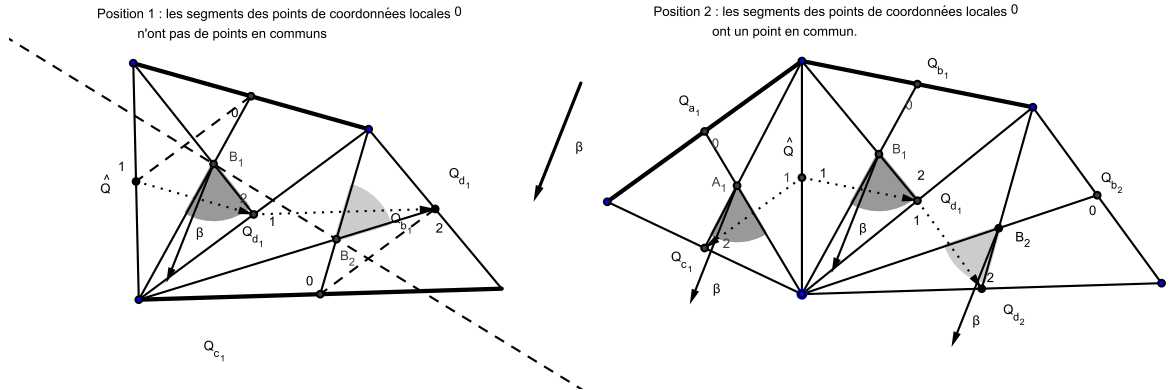
**Cas I :**  $Q_{c_1}$  et  $Q_{d_1}$  sont de type 2–0. Alors  $c_1, d_1 \in J_{\text{sortant}}^C$  et on a exploré tous les triangles de la cellule  $C$ , les affirmations 1) 2) et 3) du lemme sont vérifiés et on a terminé.

**Cas IIa :**  $Q_{c_1}$  est de type 2–0, mais pas  $Q_{d_1}$  (qui est donc de type 2–1). Alors  $c_1 \in J_{\text{sortant}}^C$  et il reste à explorer le triangle contenant l'arc issu de  $Q_{d_1}$  (qui apportera le second élément de  $J_{\text{sortant}}^C$ ).

**Cas IIb :**  $Q_{d_1}$  est de type 2–0, mais pas  $Q_{c_1}$  (qui est donc de type 2–1). Alors  $d_1 \in J_{\text{sortant}}^C$  et il reste à explorer le triangle contenant l'arc issu de  $Q_{c_1}$  (qui apportera le second élément de  $J_{\text{sortant}}^C$ ).

**Cas III :**  $Q_{c_1}$  et  $Q_{d_1}$  sont de type 2–1. Aucun élément de  $J_{\text{sortant}}^C$  n'a encore été découvert. Supposons, sans restriction, que  $Q_{c_1}$  soit de type 2–1 (rappelons que le type 2–2 n'est pas possible), alors il existe un arc orienté issu de  $Q_{c_1}$ , d'extrémité  $Q_{c_2}$  de coordonnée locale 2.

Supposons que l'on soit dans les cas IIa ou III, alors on a la situation suivante :



Pour le nouveau triangle, les milieux des côtés  $Q_{b_2}$  et  $Q_{d_2}$  vont avoir comme coordonnées locales 0 ou 2. Deux choix sont possibles, mais la position 1 est incompatible avec les cônes de directions imposés par  $\beta$ .

Le cas IIb se traite de la même façon. Cette itération va se poursuivre jusqu'à ce que l'extrémité de l'arc issu de la coordonnée locale 1 du dernier triangle ajouté soit de type 2 – 0.

**Fin :** A la fin du processus d'exploration de  $C$ , on a trouvé exactement deux points dans  $J_{\text{sortant}}^C$ , ce qui prouve 1). Ces deux segments ont un sommet en commun, ce qui prouve 2). Aucun point autre que  $\hat{Q}$  est de type 1 – 1, ce qui prouve 3). Les points  $\dots, Q_{a_1}, Q_{b_1}, \dots$  sont de type 0–?, et, comme  $O$  est la coordonnée locale du triangle dans  $C$ , on a  $\dots, a_1, b_1, \in J_{\text{entrant}}^C$ .

**A) On suppose qu'il n'existe pas de point de type 1 – 1 dans  $C$  :**

Soit  $\hat{T}$  un triangle de  $C$ , alors  $Q_0^{\hat{T}}$  est de type 0–?, et  $Q_0^{\hat{T}}$  (son indice) est dans  $J_{\text{entrant}}^C$ .

On doit maintenant explorer  $C$  en amont du point  $Q_1^{\hat{T}}$  et en aval de  $Q_2^{\hat{T}}$ .

Si  $Q_1^{\hat{T}}$  et  $Q_2^{\hat{T}}$  sont de type 1 – 2, cela est toujours possible et on procède comme ci-dessus (itération). A la fin, on se retrouve avec deux points dans  $J_{\text{sortant}}^C$ , dont les segments ont un point en commun.

Jamais on ne rencontre donc de points dans  $J_{\text{intérieur}}^C$  de type 1 – 1 et  $J_{\text{intérieur}}^C$  ne contient aucun point de type 1 – 1.

□

Le nombre de triangles nécessaires pour former une cellule dépend de la géométrie du maillage. Il est essentiel, pour la suite de la théorie, que le nombre de triangles de chaque cellule  $C \in \mathcal{C}_h$  n'augmente pas quand  $h \rightarrow 0$ .

**Définition 14.9 (Condition sur la taille des cellules)** Soit  $\mathcal{C}_h$  l'ensemble des cellules de  $\mathcal{T}_h$ , alors les cellules sont appelées de tailles finies, si il existe une constante  $K_{\text{cell}} > 0$ , ne dépendant pas de  $h$ , ni de  $C \in \mathcal{C}_h$ , telle que

$$\text{card}(C) \leq K_{\text{cell}} \quad \forall C \in \mathcal{C}_h \text{ et } h > 0 \tag{14.233}$$



Le lemme suivant précise la longueur maximale possible d'une suite de triangle de  $C$  dont les points intérieurs sont de type 1 – 2 ou 1 – 1 :

**Lemme 14.3 (Nombre maximal de triangles dans une cellule)** *Soit  $\beta \neq 0$  un champs constant,  $(\mathcal{T}_h)_h$  une famille de maillage régulière et quasi-uniforme. Soit  $\mathcal{G}_h$  le graphe orienté associé à  $\mathcal{T}_h$  et  $\mathcal{C}_h$  l'ensemble des cellules. Si toutes les cellules  $C \in \mathcal{C}_h$  ne contiennent pas de point intérieur de type 0 – 1 ou 0 – 2, alors pour le nombre maximal de triangles  $N_C$  dans  $C$  on a :*

$$N_C \leq K_{cell} \text{ avec } K_{cell} \leq 2 \frac{\ln(4\sigma c_0)}{\ln 2} \tag{14.234}$$

où  $\sigma$  et  $c_0$  sont les constantes, respectivement, telles que

$$\forall h \in ]0; h_0[ \forall T \in \mathcal{T}_h \quad \sigma_T = \frac{h_T}{\rho_T} \leq \sigma \text{ et } \forall h, \forall T \in \mathcal{T}_h, \quad \frac{h}{h_T} \leq c_0 \tag{14.235}$$

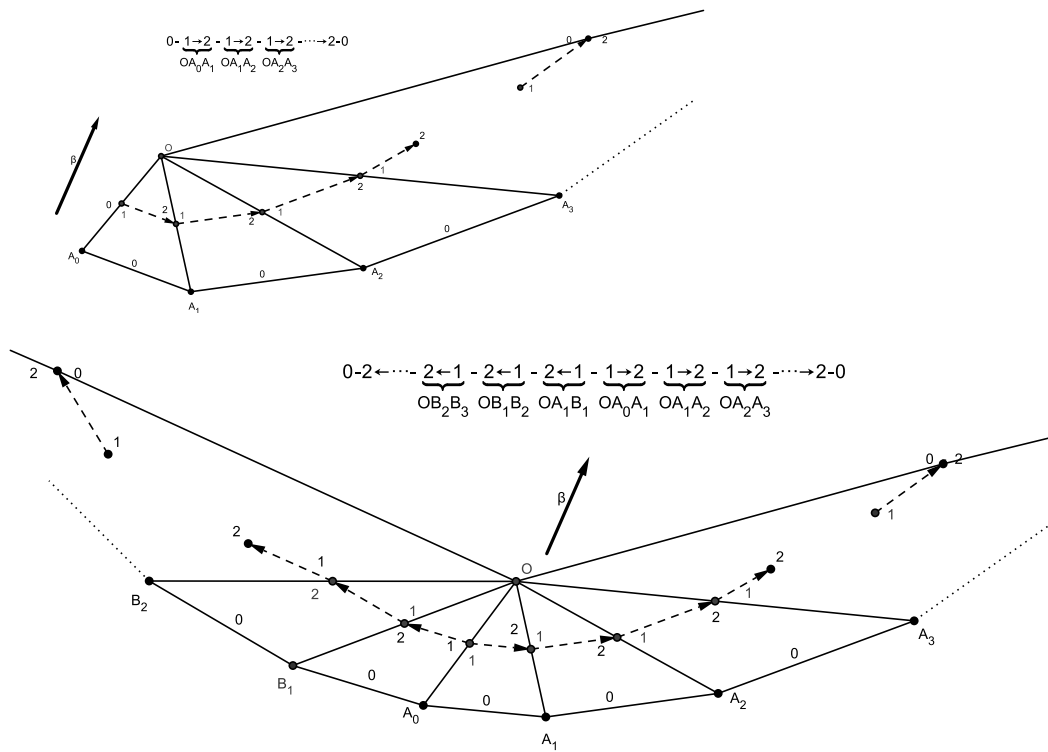


FIGURE 18 – Soit  $C \in \mathcal{C}_h$  une cellule telle que  $J_{int\acute{e}rieur}^C$  ne contienne aucun point de type 0 – 1 ou 0 – 2, alors les deux cas possibles : a)  $C$  ne contient pas de point intérieur de type 1 – 1, et b)  $C$  contient un point (qui est alors unique) intérieur de type 1 – 1, sont représentés ci-dessus. (N’ont été représentés que les arcs contribuant à l’exploration de la cellule.)

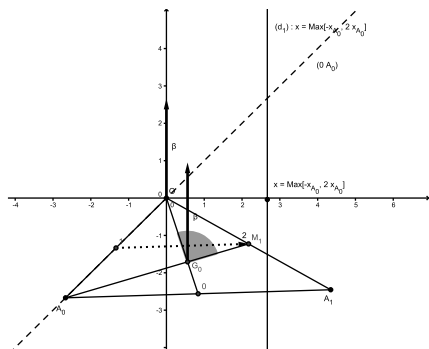
**Preuve :** Soit  $C \in \mathcal{C}_h$ . La structure de  $C$  est parfaitement connue :

a) soit il n’y a pas de point de type 1 – 1 dans  $J_{int\acute{e}rieur}^C$ , alors  $C$  consiste en une séquence de triangles dont les points intérieurs sont de type 1 – 2 (voir la figure 18).

b) soit  $C$  contient un point intérieur de type 1 – 1, alors on a de part et d'autre de ce point, deux séquences de triangles de points intérieurs de type 1 – 2 (voir la figure 18).

On est donc amené à étudier la chaîne de triangles qui commence par le triangle  $OA_0A_1$ .

Soit  $(O, \vec{i}, \vec{j})$  un repère orthonormal tel que  $\beta = \|\beta\| \vec{i}$ . Sans restriction, on peut supposer<sup>36</sup> que l'angle  $\overrightarrow{OA_0}, \overrightarrow{OA_1} \in ]0; \pi[$ . On a la figure suivante :



dire

Posons  $A_k \begin{pmatrix} x_k \\ y_k \end{pmatrix}$ . Alors  $G_k \begin{pmatrix} \frac{x_k + x_{k+1}}{3} \\ \frac{y_k + y_{k+1}}{3} \end{pmatrix}$  est le centre de gravité du triangle  $OA_kA_{k+1}$  et  $M_k \begin{pmatrix} \frac{x_k}{2} \\ \frac{y_k}{2} \end{pmatrix}$  est le milieu de  $[OA_k]$ . Comme  $M_k$  est de type 2 – 1, alors  $\beta$  doit être dans le cône de directions  $(\overrightarrow{G_kM_{k+1}}, \overrightarrow{G_kO})$ . Cette condition est équivalente à la représentation  $\beta = s \overrightarrow{G_kM_{k+1}} + t \overrightarrow{G_kO}$  avec  $s \geq 0$  (et  $t \geq 0$ ). C'est-à-dire

$$0 \leq s = \frac{\|\beta\|(x_{k+1}/2 - x_k)}{\det \begin{pmatrix} x_k & x_{k+1} \\ x_k & x_{k+1} \end{pmatrix}} \text{ et } 0 \leq t = \frac{2\|\beta\|(x_{k+1} + x_k)}{\det \begin{pmatrix} x_k & x_{k+1} \\ x_k & x_{k+1} \end{pmatrix}}$$

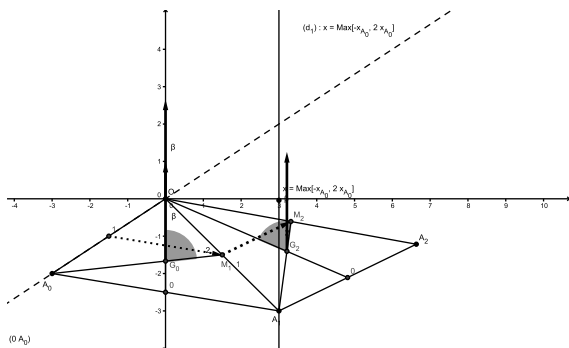
**Triangle  $OA_0A_1$  :**  $M_1$  est de type 2 – 1, donc  $\beta$  doit être dans le cône de directions  $(\overrightarrow{G_0M_1}, \overrightarrow{G_0O})$ . Comme  $\overrightarrow{OA_0}, \overrightarrow{OA_1} \in ]0; \pi[$ , alors  $\det \begin{pmatrix} x_k & x_{k+1} \\ x_k & x_{k+1} \end{pmatrix} > 0$  et cette condition est équivalente à  $x_1 \geq \max\{-x_0, 2x_0\} > 0$ .

Si  $x_0 \leq 0$ , la plus petite valeur possible pour  $x_1$  est  $x_1 = -x_0$  et  $x_1 - x_0 = -2x_0$ .  
Si  $x_0 > 0$ , la plus petite valeur possible pour  $x_1$  est  $x_1 = 2x_0$  et  $x_1 - x_0 = x_0$ .

Le triangle  $T = OA_0A_1$  doit contenir une boule de rayon  $\rho_T$ , donc  $x_1 - x_0 \geq \rho_T \geq \frac{h_T}{\sigma} \geq \frac{h}{\sigma c_0}$ .

Si  $x_0 \leq 0$ , alors avec la plus petite valeur possible pour  $x_1 - x_0$  on a  $-2x_0 = x_1 - x_0 \geq \frac{h}{\sigma c_0}$ . De même, si  $x_0 > 0$ , on obtient  $x_0 = x_1 - x_0 \geq \frac{h}{\sigma c_0}$ . Donc  $|x_0| \geq \frac{h}{2\sigma c_0}$  et ainsi  $x_1 \geq \max\{-x_0, 2x_0\} \geq \frac{h}{2\sigma c_0}$ .

**Pas de  $k$  à  $k + 1$  pour  $k \geq 1$  :** Soit  $k$  tel que  $x_k > 0$  (c'est possible, car  $x_1 \geq \frac{h}{2\sigma c_0}$ )



Le milieu du segment  $[A_k; A_{k+1}]$  est un point de coordonnée locale 0, donc  $\overrightarrow{OA_k}, \overrightarrow{OA_{k+1}} \in ]0; \pi[$  et  $\det \begin{pmatrix} x_k & x_{k+1} \\ x_k & x_{k+1} \end{pmatrix} > 0$ . Ainsi  $s, t \geq 0$  est équivalent à  $x_{k+1} \geq \max\{-x_k, 2x_k\} = 2x_k$  et  $x_{k+1} > 0$ .

36. sinon, on effectue une symétrie axiale par rapport à la droite passant par  $O$  et de direction  $\beta$

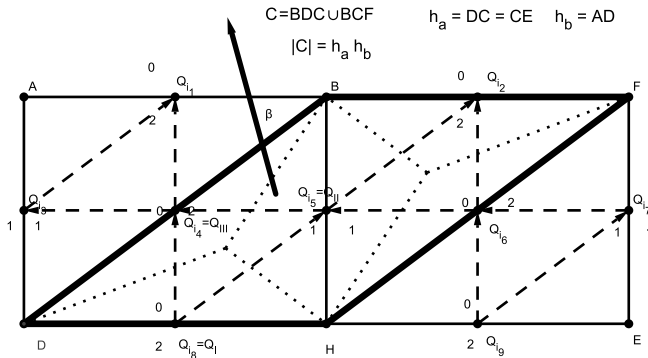
**Conclusion :** Pour  $k \geq 1$  on a donc  $x_k \geq \frac{h}{2\sigma c_0} 2^{k-1}$ . La taille du triangle  $OA_{k-1}A_k$  doit être inférieure à  $h$ , donc  $h \geq OA_k \geq \frac{h}{2\sigma c_0} 2^{k-1}$ . Ainsi  $\frac{\ln(4\sigma c_0)}{\ln 2} \geq k$ .

Si la cellule  $C$  contient un point intérieur de type  $1 - 1$ , alors dans le cas le plus défavorable on a deux séquences de tailles maximales, et le nombre maximal possible de triangles dans  $C$  est  $2 \frac{\ln(4\sigma c_0)}{\ln 2}$ .

□

**Remarque 14.10 (passage de  $\beta$  à  $-\beta$ )** a) La structure des cellules dépend étroitement de la direction de  $\beta$ . Le passage de  $\beta$  à  $-\beta$  inverse l'orientation des arcs dans  $\mathcal{G}_h$  :  $S_{-\beta}^-(i) = S_{-\beta}^+(i)$  et une coordonnée locale 0 devient 2, 2 devient 0 et 1 reste 1. Si un point dans  $J_{entrant}^C \cup J_{sortant}^C$  est de type  $0 - 1$  pour  $\beta$ , alors il sera de type  $2 - 1$  pour  $-\beta$ , ce qui signifie qu'il appartiendra à  $J_{intérieur}^C$  pour une certaine cellule.

b) Dans cette remarque, une autre preuve de (6.132), page 66 est donnée pour  $t = 0$  : On considère un maillage uniforme symétrique (dans le sens où deux triangles ayant un côté en commun sont symétriques par rapport au milieu de ce côté) et on a la situation suivante<sup>37</sup> :



Alors pour le triangle  $DHB$  :

$$\begin{aligned}
& \frac{1}{\|\beta\|^2} \sum_{j \in S_{\beta}^+(I)} [\beta \cdot (Q_j - Q_I)] \theta_{I,j}^{\beta} + \frac{1}{\|\beta\|^2} \sum_{j \in S_{\beta}^+(II)} [\beta \cdot (Q_j - Q_{II})] \theta_{II,j}^{\beta} + \\
& \frac{1}{\|\beta\|^2} \sum_{j \in S_{\beta}^+(III)} [\beta \cdot (Q_j - Q_{III})] \theta_{III,j}^{\beta} \text{ Passage de } \beta \text{ à } -\beta : \\
= & \frac{1}{\|\beta\|^2} \sum_{j \in S_{-\beta}^-(I)} [(-\beta) \cdot (Q_I - Q_j)] \theta_{I,j}^{-\beta} + \frac{1}{\|\beta\|^2} \sum_{j \in S_{-\beta}^-(II)} [(-\beta) \cdot (Q_{II} - Q_j)] \theta_{II,j}^{-\beta} + \\
& \frac{1}{\|\beta\|^2} \sum_{j \in S_{-\beta}^-(III)} [(-\beta) \cdot (Q_{III} - Q_j)] \theta_{III,j}^{-\beta} \text{ Le maillage est uniforme, on peut remplacer } i_4 \text{ par } i_6 :
\end{aligned}$$

37. Les indices (ou exposants)  $\beta$  et  $-\beta$  indiquent le vecteur utilisé. La cellule  $C$  ne change pas lors du passage de  $\beta$  à  $-\beta$ , car le type  $0 - 2$  passera à  $2 - 0$  et  $1 - 1$  reste inchangé.

$$\begin{aligned}
&= \frac{1}{\|\beta\|^2} \sum_{j \in S_{-\beta}^-(i_8)} [(-\beta) \cdot (Q_{i_8} - Q_j)] \theta_{i_8,j}^{-\beta} + \frac{1}{\|\beta\|^2} \sum_{j \in S_{-\beta}^-(i_5)} [(-\beta) \cdot (Q_{i_5} - Q_j)] \theta_{i_5,j}^{-\beta} + \\
&\quad \frac{1}{\|\beta\|^2} \sum_{j \in S_{-\beta}^-(i_6)} [(-\beta) \cdot (Q_{i_6} - Q_j)] \theta_{i_6,j}^{-\beta} \text{ Pour } C = BHD \cup BHF, \quad J^C \setminus J_{-\beta,entrant}^C = \{i_8, i_5, i_6\} : \\
&= \sum_{i \in J^C \setminus J_{-\beta,entrant}^C} \frac{1}{\|\beta\|^2} \sum_{j \in S_{-\beta}^-(i)} [-\beta \cdot (Q_k - Q_j)] \theta_{j,k}^{-\beta} \\
&= |C| = h_a h_b
\end{aligned}$$

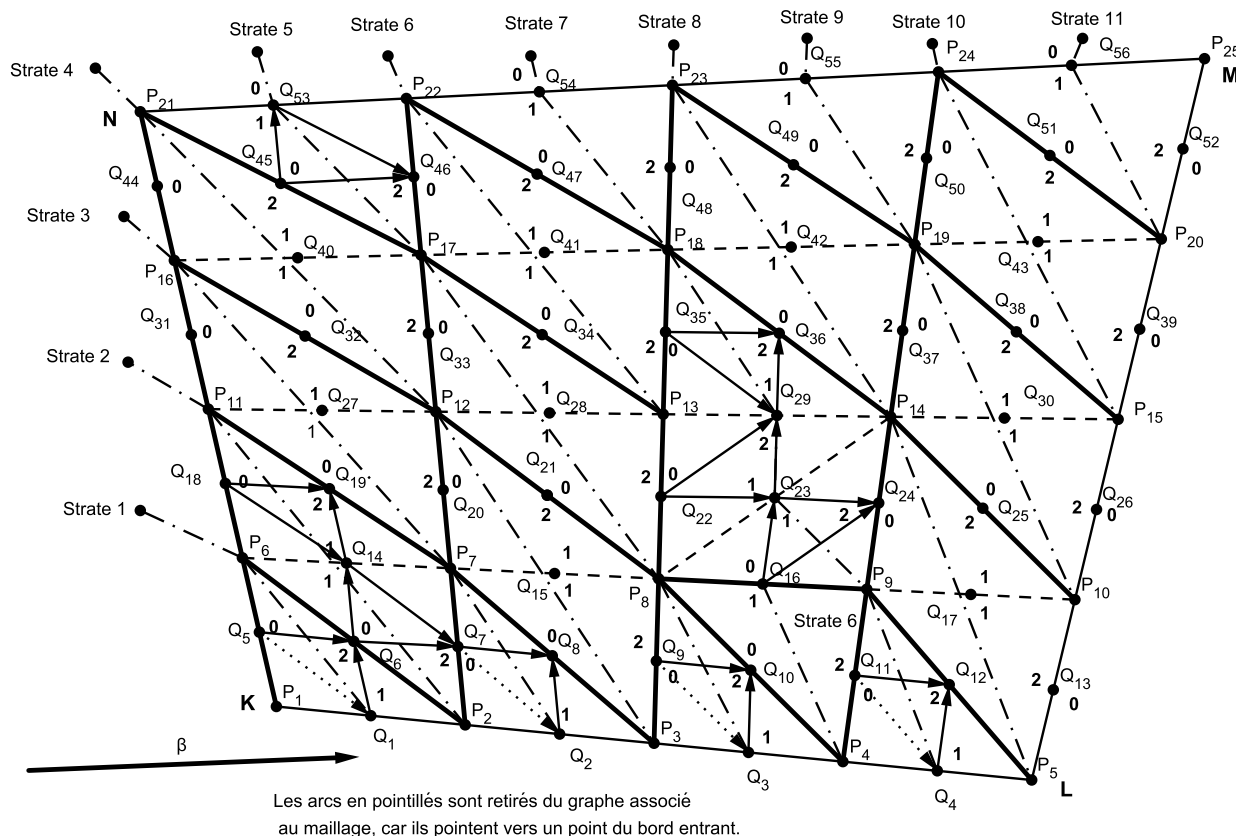
## 15 Étapes et idées principales pour les démonstrations

### Organisation des cellules en strates

L'étude locale se fait au niveau des cellules via leurs points d'entrée. La connaissance de  $u_h$  aux points d'entrée d'une cellule permet de calculer "en block" tous les points de cette cellule par des formules explicites qui vont être données au chapitre suivant.

A contrario, tant qu'une des valeurs pour un point d'entré d'une cellule n'est pas connue, cette cellule ne peut être calculée.

Il faut donc se pencher sur l'organisation des cellules entre elles.



La figure ci-dessus montre la décomposition en mailles triangulaires de  $\Omega = KLMN$ .

Le bord entrant est  $\Gamma^- = ]NK] \cup [KL[$ , le bord longeant<sup>38</sup> est  $\Gamma^b = ]NM[$  et le bord sortant est  $\Gamma^+ = ]ML[$ .  $J^- = \{44, 33, 18, 5, 1, 2, 3, 4\}$ ,  $J^b = \{53, 54, 55, 56\}$  et  $J^+ = \{13, 26, 39, 52\}$ .

Certains arcs orientés ont été retirés de  $\mathcal{G}_h$  :  $(5, 1)$ ,  $(7, 2)$ ,  $(9, 3)$  et  $(11, 4)$ , car ils pointent vers le bord entrant. Les triangles  $P_6P_1P_2$ ,  $P_7P_2P_3$ ,  $P_8P_3P_4$  et  $P_9P_4P_5$  ont donc deux coordonnées locales 0 : il faudrait donc barrer la coordonnée locale 1 et la remplacer par 0, ce qui n'a pas été fait sur la figure ci-dessus pour ne pas l'alourdir inutilement les notations.

Certains arcs orientés ont été représentés, mais pas tous. Les segments contenant les points de type 0 – 1 et 0 – 2 ont été représentés en traits continus. On a ainsi une décomposition de  $\Omega$  en cellules contenant 1, 2 ou 3 triangles.

Les cellules, obtenues ici, sont celles étudiées au chapitre précédent. En effet, aucune d'elles ne contient un point intérieur de type 0 – 1 ou 0 – 2. On remarquera, que le graphe  $\mathcal{G}_h$  associé à un tel maillage ne peut pas avoir de circuit élémentaire.

Le schéma itératif, qui permet de calculer  $u_h$  pour les points  $\{Q_1, Q_2, \dots, Q_{56}\} \setminus \{Q_i : i \in J^-\}$ , commence avec  $Q_6$ . Ce sera notre première strate

$$\Xi_1 = \{C_{1,1}\} \text{ avec } C_{1,1} = P_6P_1P_2$$

Ici,  $C_{s,k}$  désigne la  $k$ -ème cellule dans la strate  $s$ . On notera  $n_{\Xi,s}$  le nombre de cellules de la strate  $s$  et  $N_{\Xi}$  le nombre de strates. Donc  $n_{\Xi,1} = 1$

La strate suivante  $\Xi_2$ , ne contient, elle aussi, qu'une cellule  $C_{2,1} = P_{11}P_6P_7 \cup P_6P_2P_7$

$$\Xi_2 = \{C_{2,1}\} \text{ avec } n_{\Xi,2} = 1$$

Pour la strate  $\Xi_3$  :  $n_{\Xi,3} = 2$  et

$$\Xi_3 = \{C_{3,1}, C_{3,2}\} \text{ avec } C_{3,1} = P_{16}P_{11}P_{12} \cup P_{11}P_7P_{12} \text{ et } C_{3,2} = P_7P_2P_3$$

En effet, les cellules  $C_{3,1}$  et  $C_{3,2}$  se calculent - indépendamment l'une de l'autre - de manière parallèle.

Pour la strate  $\Xi_4$  :  $n_{\Xi,4} = 2$  et

$$\Xi_4 = \{C_{4,1}, C_{4,2}\} \text{ avec } C_{4,1} = P_{21}P_{16}P_{17} \cup P_{16}P_{12}P_{17} \text{ et } C_{4,2} = P_{12}P_7P_8 \cup P_7P_3P_8$$

Ici aussi, les cellules  $C_{4,1}$  et  $C_{4,2}$  se calculent en parallèle.

Pour la strate  $\Xi_5$  :  $n_{\Xi,5} = 3$  et

$$\Xi_5 = \{C_{5,1}, C_{5,2}, C_{5,3}\} \text{ avec } C_{5,1} = P_{21}P_{17}P_{22}, C_{5,2} = P_{17}P_{12}P_{13} \cup P_{12}P_8P_{13} \text{ et } C_{5,3} = P_8P_3P_4$$

Pour la strate  $\Xi_6$  :  $n_{\Xi,6} = 2$  et

$$\Xi_6 = \{C_{6,1}, C_{6,2}, C_{6,3}\} \text{ avec } C_{6,1} = P_{22}P_{17}P_{18} \cup P_{17}P_{13}P_{18} \text{ et } C_{6,2} = P_8P_4P_9$$

---

38.  $(NM)$  est donc parallèle à  $\beta$

Pour la strate  $\Xi_7 : n_{\Xi,7} = 3$  et

$$\Xi_7 = \{C_{7,1}, C_{7,2}, C_{7,3}\} \text{ avec}$$

$$C_{7,1} = P_{22}P_{18}P_{23}, C_{7,2} = P_{18}P_{13}P_{14} \cup P_{13}P_8P_{14} \cup P_8P_9P_{14} \text{ et } C_{7,3} = P_9P_4P_5$$

Etc, etc...

La dernière strate étant

$$\Xi_{11} = \{C_{11,1} \text{ avec } C_{11,1} = \{P_{24}P_{20}P_{25}\}$$

Donc  $N_{\Xi} = 11$ .

Précisons exactement la notion de strates :

**Définition 15.1 (strates)** Soit  $\mathcal{G}_h$  le graphe associé à un maillage  $\mathcal{T}_h$  qui ne contient pas de circuit élémentaire et  $\mathcal{C}_h$  l'ensemble des cellules de  $\Omega$ , alors  $\Xi_1, \dots, \Xi_l, \Xi_{l+1}, \dots, \Xi_N$  est une décomposition disjointe de  $\mathcal{C}_h$  en strates si :

- a) La réunion de toutes les strates  $\Xi_1$  à  $\Xi_N$  est  $\Omega$ ,
- b) Les points d'entrée de la strate  $\Xi_{l+1}$  sont inclus<sup>39</sup> dans la réunion des points de sortie des strates inférieurs  $\Xi_1$  à  $\Xi_l$ .
- c) Les cellules d'une même strate se calculent indépendamment : aucun point de sortie d'une cellule de  $\Xi_l$  ne peut être un point d'entrée d'une autre cellule de la même strate  $\Xi_l$  ou d'une strate inférieure.

On note  $N_{\Xi} = N$  le nombre de strates.

**Remarque 15.2 i)** Si  $\mathcal{G}_h$  contient un circuit élémentaire, alors la notion de strate est problématique. On peut bien sur imaginer qu'un circuit élémentaire soit présent à l'intérieur d'une cellule d'une des strates et qu'il soit possible de calculer cette cellule en fonction de ses points d'entrée. En effet, si l'existence d'un circuit élémentaire dans  $\mathcal{G}_h$  entraîne l'existence d'un circuit local (autour d'un sommet d'un triangle du maillage), la réciproque peut ne pas être vraie. Cette question n'est pas encore tranchée.

ii) Tous les points d'entrée de la première strate  $\Xi_1$  sont sur le bord entrant.

Le nombre de strates augmente avec la finesse  $h$  du maillage, mais aussi le nombre de cellules dans chaque strate  $\Xi_s$ .

Dans les expériences numériques faites, on constate que le nombre moyen de cellules dans chaque strate augmente comme  $O(1/h)$ . Cela semble logique, car la taille d'une cellule est  $O(h)$ .

Comme le nombre de valeurs  $u_h(Q_i)$  à déterminer est  $|J \setminus J^-| = O(1/h^2)$ , l'hypothèse que le nombre de strates augmente comme  $O(1/h)$ , semble raisonnable :

**Définition 15.3 (Hypothèse sur le nombre de strates  $N_{\Xi}$ )** Soit  $(\mathcal{T}_h)_h$  une famille de maillage, alors on suppose qu'il existe une constante  $K_{\Xi}$ , indépendante de  $h$ , telle que

$$N_{\Xi} \leq K_{\Xi} \frac{1}{h} \tag{15.236}$$

39. Autrement dit, le calcul de la strate  $\Xi_{l+1}$ , nécessite de calculer avant  $\Xi_1$  à  $\Xi_l$ .

## Étapes et idées principales dans les preuves de stabilité

Les preuves de stabilité se font en deux étapes :

**Étape 1 :** Les strates sont contrôlées par leurs points d'entrés. La première étape consiste donc à établir des estimations sur les valeurs aux points d'entrés des cellules. Les points intérieurs d'une cellule n'interviennent pas ici !

**Un point fondamental pour cette première étape, est le fait, que pour un point d'entrée  $j$  de la strate  $\Xi_{t+1}$ , deux possibilités se présentent :**

- a) **Le point  $Q_j \in \Gamma^-$  est sur le bord entrant  $\Gamma^-$  et alors  $u_h(Q_j) = g_h(Q_j)$  est connu,**
- b) **Le point  $Q_j \notin \Gamma^-$  est un point sortant d'une cellule  $C \in \Xi_q$  d'une strate  $\Xi_q$  inférieure à  $\Xi_{t+1}$  (donc  $t \geq q \geq 1$ ) et il sera possible d'exprimer  $u_h(Q_j)$  en fonction des points d'entrés de  $C$ .**

**Étape 2 :** L'ensemble des valeurs d'une cellule se détermine à partir des valeurs ses points d'entrés. Il reste donc à obtenir une majoration des quantités aux points intérieurs et sortants des cellules en fonction des valeurs aux points d'entrés pour obtenir l'estimation désirée.

Les inégalités suivantes sont nécessaires pour établir les estimations nécessaires dans les étapes 1 et 2.

## 16 Calcul de quelques types de cellules

La mise en place de la matrice  $M$  nécessite de calculer, pour chaque cellule  $C \in \mathcal{C}_h$ , les valeurs aux points  $J^C \setminus J_{\text{entrant}}^C$  en fonction des valeurs en  $J_{\text{entrant}}^C$ .

On se place dans le cadre du lemme 14.2 où les cellules prises en compte ici n'ont pas de points intérieurs de type 0 – 1 ou 1 – 2. Pour les trois types de cellules possibles, il est aisé d'établir des formules explicites de calcul -exprimées en coordonnées locales- pour ces cellules.

Ces formules sont nécessaires pour les deux formulations variationnelles. Afin d'unifier ces résultats pour les formulations I et II, posons  $\alpha_i \in \{\alpha_i^I, \alpha_i^{II}\}$  où

$$\alpha_i^I = f_h(Q_i) \frac{|D_i|}{\sum_{j \in S^+(i)} \theta_{i,j}} \quad (16.237)$$

$$\alpha_i^{II} = f_h(Q_i) \frac{1}{\|\beta\|^2} \sum_{j \in S^-(i)} \beta \cdot (Q_i - Q_j) \lambda_{j,i} \quad (16.238)$$

Avec le lemme 12.2, on a aussi :

Si  $\alpha_i \in \{\alpha_i^I, \alpha_i^{II}\}$  alors

$$|\alpha_i| \leq |f_h(Q_i)| \frac{h}{\|\beta\| \hat{\sigma}} \quad (16.239)$$

On rappelle la formule de récurrence<sup>40</sup> :

---

40. Voir (12.209), page 97.

$$u_i = \sum_{j \in S^-(i)} u_j \lambda_{j,i} + \alpha_i \text{ et on posera } f_i = f_h(Q_i)$$

Notations :

- Rappelons que  $0 \leq \lambda_{j,i} \leq 1$  et  $\sum_{j \in S^-(i)} \lambda_{j,i} = 1$ ,
- On écrira  $Q_0^r, u_0^r, \dots$ , au lieu de  $Q_0^{Tr}, u_0^{Tr}, \dots$ ,
- Les coordonnées locales upwind seront aussi utilisés pour les segments des côtés des triangles, par exemple, si  $T$  est un triangle d'une cellule  $C$  et  $Q_j = Q_2^T$  (respectivement  $Q_j = Q_0^T$ ) est un point sortant (respectivement entrant) de  $C$ , alors le segment  $S_j$  de milieu  $Q_j$  sera aussi noté  $S_2^T$  (respectivement  $S_0^T$ ) en coordonnées locales upwind. En coordonnées locales upwind  $\hat{\theta}_i$  devient (voir la définition 14.4, donnée à la page 109) :

**Définition 16.1** ( $\hat{\theta}_{\mathbf{x}0}^T$ ,  $\hat{\theta}_0^T$  et  $\hat{\theta}_2^T$ )

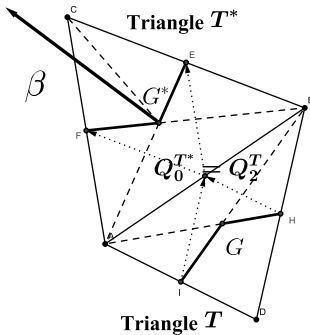
Si  $Q_0^T$  est un point entrant sans arc orienté pointant sur ce point ou si  $Q_0^T$  est un point sortant :

$$\hat{\theta}_0^T = - \int_{S_0^T} \beta \cdot n_{\partial C} d\sigma = \theta_{0,1}^T + \theta_{0,2}^T > 0$$

Si  $Q_1^T$  est un point entrant où un arc orienté pointant sur ce point a été supprimé : L'arc orienté  $(Q_0^T, Q_1^T)$  pointant sur  $\Gamma^-$  a été retiré,  $Q_1^T$  est renommé  $Q_{\mathbf{x}0}^T$  et on pose

$$\hat{\theta}_0^T = \theta_{0,2}^T \text{ et } \hat{\theta}_{\mathbf{x}0}^T = \theta_{1,2}^T \quad (16.240)$$

Si  $Q_2^T$  est un point sortant : On a la situation suivante



Le vecteur  $\beta$  est dans le cône de direction défini par  $\widehat{FG^*E} \cap \widehat{AGB}$ , donc

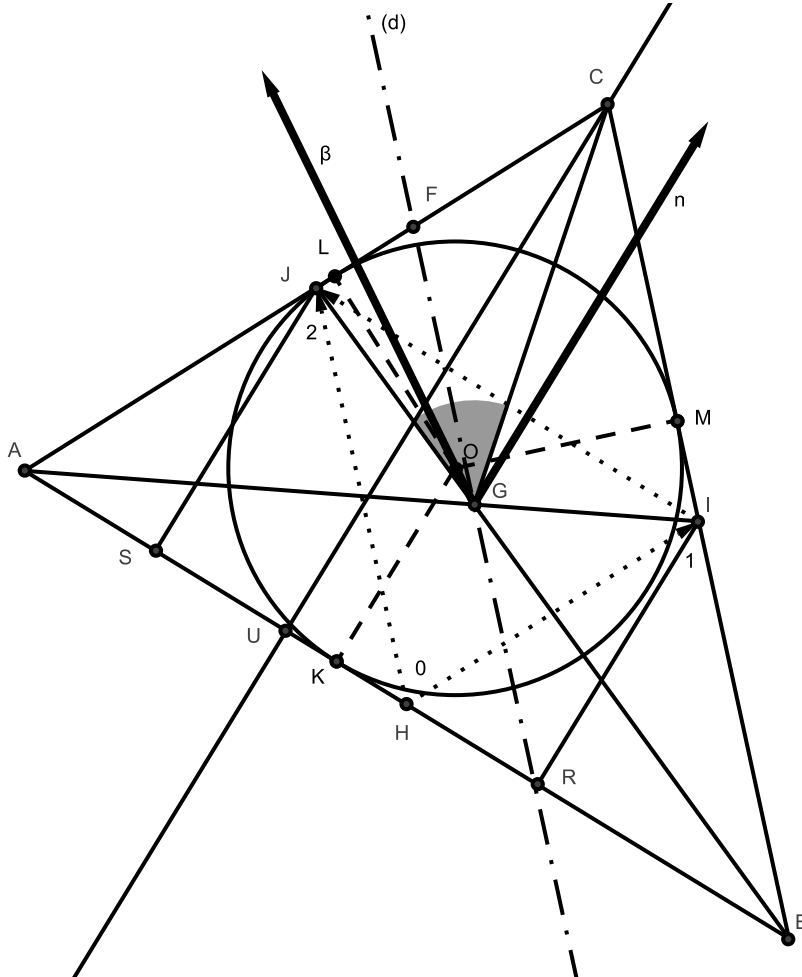
$$\hat{\theta}_2^T = \int_{S_2^T} \beta \cdot n_{\partial C} d\sigma = \theta_{0,2}^T + \theta_{1,2}^T > 0$$



**Lemme 16.1 (Minoration de  $\hat{\theta}_j$ )** *Il existe une constante  $c_{\hat{\theta}} > 0$ , indépendante de  $h$ , telle que*

$$h \|\beta\| c_{\hat{\theta}} \leq \hat{\theta}_j \text{ pour tout point } Q_j, j \in J \text{ de type } 0-1 \text{ ou } 0-2 \quad (16.241)$$

**Preuve :**



Dans le triangle  $T = ABC$  :  $O$  est le centre du cercle inscrit, donc  $\rho_T = OL = OM = OK$  ;  $G$  est le centre de gravité ;  $H, J$  et  $I$  sont respectivement les milieux des côtés  $[AB], [AC]$  et  $[BC]$ . En coordonnées locales upwind :  $H = Q_0^T, I = Q_1^T$  et  $J = Q_2^T$ , donc  $\beta$  est dans le cône de directions défini par  $\widehat{JGC}$ .

Dans le triangle  $OAK$  on a  $\frac{OK}{OA} = \sin(A/2)$ , donc  $\sin(A/2) = \frac{OK}{OA} \geq \frac{\rho_T}{h_T} \geq \frac{1}{\sigma}$  et ainsi

$$\theta_0 := 2 \sin^{-1}(1/\sigma) \leq \hat{A}.$$

Dans le triangles  $OBK$  on a de même :  $\theta_0 := 2 \sin^{-1}(1/\sigma) \leq \hat{B}$ .

Soient  $S, R$  et  $U$ , les points de  $[AB]$ , tels que les segments  $[IR], [JS]$  et  $[CU]$  soient perpendiculaires à  $[AB]$ , donc  $(CU = 2IR, AC \geq \rho_T$  et  $h_T \geq AI$ ) :

$$\sin(\widehat{IAB}) = \frac{IR}{AI} = \frac{CU}{2IA} = \frac{CU}{AC} \times \frac{AC}{2IA} \geq \sin(A) \frac{\rho_T}{2h_T} \geq \sin(\theta_0) \frac{1}{2\sigma} \quad (*)$$

et (même raisonnement) :  $\sin(\widehat{JBA}) \geq \sin(\theta_0) \frac{1}{2\sigma} \quad (**)$

Le vecteur  $\widehat{n}$  est le vecteur normal à  $[AB]$  et tel que  $n \cdot \beta > 0$ .  $\beta$  est dans le cône de directions  $\widehat{JGC} \subset \widehat{JGI}$ , donc

$$\begin{aligned} \min\left\{n \cdot \frac{\overrightarrow{AI}}{\|\overrightarrow{AI}\|}, n \cdot \frac{\overrightarrow{BJ}}{\|\overrightarrow{BJ}\|}\right\} &\leq n \cdot \frac{\beta}{\|\beta\|} \\ \min\{\sin(\widehat{BAI}), \sin(\widehat{ABJ})\} &\leq n \cdot \frac{\beta}{\|\beta\|} \\ \text{avec (*) et (**): } \sin(\theta_0) \frac{1}{2\sigma} &\leq n \cdot \frac{\beta}{\|\beta\|} \text{ Ainsi} \\ \frac{1}{\sigma} \frac{1}{c_0} \frac{1}{2\sigma} \sin(\theta_0) &\leq \frac{\rho_T h_T (n \cdot \beta)}{h_T h \|\beta\|} \leq \frac{AB (n \cdot \beta)}{\|\beta\| h} = \frac{\hat{\theta}_0^T}{\|\beta\| h} \end{aligned}$$

Ce qui prouve (16.241) pour les triangles où l'arc orienté  $(H, I) = (Q_0^T, Q_1^T)$  n'a pas été retiré.

Le cas  $\hat{\theta}_2^T$  est analogue (il suffit de passer de  $\beta$  à  $-\beta$  (voir la figure dans la définition 16.1 de  $\hat{\theta}_2^T$  ci-dessus) et constater que  $-\beta$  est dans le cône de direction  $\widehat{IGH}$ ).

Il reste à étudier la situation du triangle  $T$ , où l'arc  $(H, I) = (Q_0^T, Q_1^T)$  a été retiré. Le cône de direction pour  $\beta$  se restreint maintenant à  $\widehat{FGJ}$  (la droite  $(d)$  est parallèle à  $[BC] \subset \Gamma^-$ ). Deux points sont à considérer :

$I = Q_1^T = Q_{\mathbf{x}0}^T \in \Gamma^-$  est un point du bord entrant : Comme  $G$  est le centre de gravité du triangle  $ABC$ , alors  $FC = \frac{1}{3}AC$  et on a  $\hat{\theta}_{\mathbf{x}0} = \theta_{1,2}^T \geq \frac{1}{3}\hat{\theta}_2^T$  et la preuve est faite pour  $\hat{\theta}_{\mathbf{x}0}$ .

$H = Q_0^T$  : On a  $\hat{\theta}_0^T = \theta_{0,2}^T \geq \frac{1}{2}\hat{\theta}_2^T$ .

□

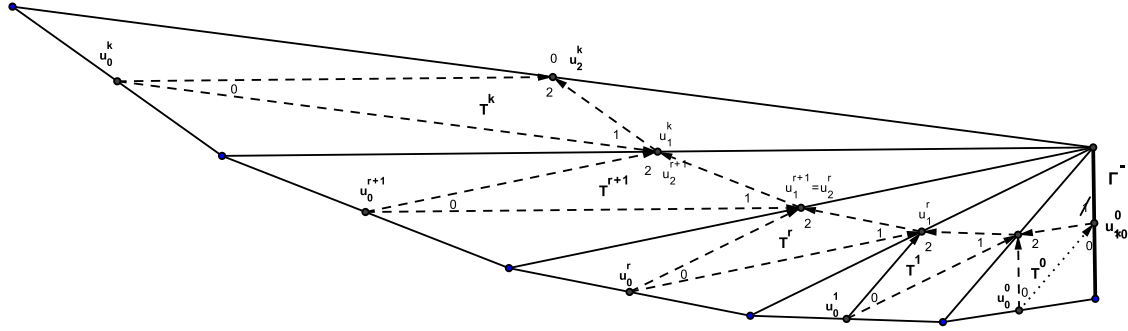
On utilisera aussi les majorations suivantes pour  $\alpha_i \in \{\alpha_i^I, \alpha_i^{II}\}$  (voir le lemme 12.2, page 97) :

$$|\alpha_i| \leq \frac{h}{\hat{\sigma} \|\beta\|} |f_i|, \text{ ainsi que} \quad (16.242)$$

$$|d_i| \leq h^2, \quad |D_i| \leq h^2 \text{ On définit } \omega^T = \lambda_{0,2}^T + \lambda_{0,1}^T \lambda_{1,2}^T \text{ et on a} \quad (16.243)$$

$$0 \leq \omega^T \leq \lambda_{0,2}^T + \lambda_{1,2}^T \leq 1 \text{ donc } |1 - \omega^T| = 1 - \omega^T \text{ et } |\omega^T| = \omega^T \quad (16.244)$$

## Cellule contenant un triangle avec deux coordonnées locales 0



Dans le triangle  $T^0$ , le point (en coordonnées locales upwind)  $Q_1^0 \in \Gamma^-$ . L'arc orienté  $(Q_1^0, Q_0^0)$  a donc été retiré de  $\mathcal{G}_h$  et  $T^0$  contient deux points de coordonnées locales  $O$  :  $Q_0^0$  et  $Q_{\mathbf{x}0}^0$ .

La cellule représentée ci-dessus compte  $k + 1$  triangles  $T^0, T^1, \dots, T^r, T^{r+1}, \dots, T^k$ .

Les  $k + 2$  points entrants sont  $Q_{\mathbf{x}0}^0, Q_0^0, \dots, Q_0^k$ , le point sortant est  $Q_2^k$ .

A) On a les formules de récurrences :

Pour  $r = k$  :

$$u_2^k = u_0^k \lambda_{0,2}^k + u_1^k \lambda_{1,2}^k + \alpha_2^k \text{ avec } \lambda_{0,2}^k + \lambda_{1,2}^k = 1 \quad (16.245)$$

Pour  $r \in \{0, \dots, k-1\}$  :

$$u_2^r = u_1^{r+1} = u_0^{r+1} \lambda_{0,1}^{r+1} + u_0^r \lambda_{0,2}^r + u_1^r \lambda_{1,2}^r + \alpha_2^r \text{ avec } \lambda_{0,1}^{r+1} + \lambda_{0,2}^r + \lambda_{1,2}^r = 1 \quad (16.246)$$

B) Alors<sup>41</sup> avec  $\omega^r = \lambda_{0,2}^r + \lambda_{0,1}^r \lambda_{1,2}^r$  on obtient les formules explicites :

**Point sortant :** Pour  $r = k$  on a

$$u_2^k = \sum_{r=1}^k u_0^r \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l + (u_0^0 \lambda_{0,2}^0 + u_{\mathbf{x}0}^0 \lambda_{1,2}^0) \prod_{l=1}^k \lambda_{1,2}^l + \sum_{r=0}^k \alpha_2^r \prod_{l=r+1}^k \lambda_{1,2}^l \quad (16.247)$$

$$1 = \sum_{r=1}^k \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l + (\lambda_{0,2}^0 + \lambda_{1,2}^0) \prod_{l=1}^k \lambda_{1,2}^l \quad (16.248)$$

De (16.247) on déduit l'inégalité :

$$|u_2^k| \leq \sum_{r=1}^k |u_0^r| \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l + (|u_0^0| \lambda_{0,2}^0 + |u_{\mathbf{x}0}^0| \lambda_{1,2}^0) \prod_{l=1}^k \lambda_{1,2}^l + \frac{h}{\hat{\sigma} \|\beta\|} \sum_{r=0}^k |f_2^r| \quad (16.249)$$

**Points intérieurs :** Pour  $r \in \{0, \dots, k-1\}$  on a

$$u_2^r = u_0^{r+1} \lambda_{0,1}^{r+1} + \sum_{s=1}^r u_0^s \omega^s \prod_{l=s+1}^r \lambda_{1,2}^l + (u_0^0 \lambda_{0,2}^0 + u_{\mathbf{x}0}^0 \lambda_{1,2}^0) \prod_{l=1}^r \lambda_{1,2}^l + \sum_{s=0}^r \alpha_2^s \prod_{l=s+1}^r \lambda_{1,2}^l \quad (16.250)$$

$$1 = \lambda_{0,1}^{r+1} + \sum_{s=1}^r \omega^s \prod_{l=s+1}^r \lambda_{1,2}^l + (\lambda_{0,2}^0 + \lambda_{1,2}^0) \prod_{l=1}^r \lambda_{1,2}^l \text{ pour } r \in \{0, \dots, k-1\} \quad (16.251)$$

41. avec la convention qu'une somme vide est égale à 0 et un produit vide est égal à 1

De (16.250) on déduit l'inégalité :

$$|u_2^r| \leq |u_0^{r+1}| \lambda_{0,1}^{r+1} + \sum_{s=1}^r |u_0^s| \omega^s \prod_{l=s+1}^r \lambda_{1,2}^l + (|u_0^0| \lambda_{0,2}^0 + |u_{\mathbf{x}0}^0| \lambda_{1,2}^0) \prod_{l=1}^r \lambda_{1,2}^l + \frac{h}{\delta \|\beta\|} \sum_{s=0}^r |f_2^s| \quad (16.252)$$

C) De plus on a l'égalité suivante :

$$\hat{\theta}_2^k u_2^k = \sum_{r=1}^k u_0^r \hat{\theta}_0^r + u_0^0 \hat{\theta}_0^0 + u_{\mathbf{x}0}^0 \hat{\theta}_{\mathbf{x}0}^0 + \sum_{r=0}^k \begin{cases} |D_2^r| f_2^r & \text{pour formulation I} \\ d_2^r f_2^r & \text{pour formulation II} \end{cases} \quad (16.253)$$

$$\text{où } \hat{\theta}_2^k = \int_{S_2^k} \beta \cdot n_{\partial C} d\sigma = \theta_{0,2}^k + \theta_{1,2}^k > 0 ,$$

$$\hat{\theta}_0^r = - \int_{S_0^r} \beta \cdot n_{\partial C} d\sigma = \theta_{0,1}^r + \theta_{0,2}^r > 0 \text{ pour } 1 \leq r \leq k \text{ et } \hat{\theta}_0^0 = \theta_{0,2}^0 > 0, \hat{\theta}_{\mathbf{x}0}^0 = \theta_{1,2}^0 \geq 0,$$

(La définition de  $\hat{\theta}_0^0$  et  $\hat{\theta}_{\mathbf{x}0}^0$  est particulière, car l'arc orienté  $(Q_0^0, Q_1^0)$  a été retiré de  $\mathcal{G}_h$ .)

**Preuve :** La preuve des équations (16.247) à (16.251) se fait par récurrence sur  $r$  où, à chaque étape, deux cas sont à étudier :

- a) le nouveau point est un point intérieur (c'est-à-dire  $r < k$ ),
- b) le nouveau point est un point sortant (c'est-à-dire  $r = k$  et la récurrence s'arrête).

La preuve sera faite pour la formulation II, pour la formulation I il suffit de remplacer  $d_k$  par  $|D_k|$  dans ce qui suit. <sup>42</sup>

**Début de l'induction :**  $r = 0$  Deux cas sont à étudier (rappel :  $u_1^0 = u_{\mathbf{x}0}^0$ )

**Cas  $r = k$  (point sortant)** Alors les formules (16.245) et (16.247) sont identiques. Et (16.248) se déduit de (16.245).

**Cas  $r < k$  (point intérieur)** Alors les formules (16.246) et (16.250) sont identiques. Et (16.251) se déduit de (16.246).

**Induction de  $r$  à  $r + 1$**  On suppose que les formules (16.250) et (16.251) sont vraies pour un certain  $r < k$ . Deux cas sont à étudier

**Cas  $r + 1 = k$  (point sortant)** Alors pour  $r + 1$  (16.245) devient

$$u_2^k = u_0^k \lambda_{0,2}^k + u_1^k \lambda_{1,2}^k + \alpha_2^k \text{ avec } \lambda_{0,2}^k + \lambda_{1,2}^k = 1$$

42. Rappelons que  $\alpha_i = \alpha_i^I$  dans la formulation I et  $\alpha_i = \alpha_i^{II}$  dans la formulation II, voir page 97

Comme  $u_1^k = u_2^r$ , on peut remplacer (16.250) dans l'équation ci-dessus :

$$\begin{aligned}
u_2^k &= u_0^k \lambda_{0,2}^k + \left[ u_0^{r+1} \lambda_{0,1}^{r+1} + \sum_{s=1}^r u_0^s \omega^s \prod_{l=s+1}^r \lambda_{1,2}^l + (u_0^0 \lambda_{0,2}^0 + u_{\mathbf{x}0}^0 \lambda_{1,2}^0) \prod_{l=1}^r \lambda_{1,2}^l \right. \\
&\quad \left. + \sum_{s=0}^r \alpha_2^s \prod_{l=s+1}^r \lambda_{1,2}^l \right] \lambda_{1,2}^k + \alpha_2^k \\
&= u_0^k [\lambda_{0,2}^k + \lambda_{0,1}^{r+1} \lambda_{1,2}^k] + \sum_{s=1}^r u_0^s \omega^s \prod_{l=s+1}^k \lambda_{1,2}^l + (u_0^0 \lambda_{0,2}^0 + u_{\mathbf{x}0}^0 \lambda_{1,2}^0) \prod_{l=1}^k \lambda_{1,2}^l \\
&\quad + \sum_{s=0}^k \alpha_2^s \prod_{l=s+1}^k \lambda_{1,2}^l \\
&= \sum_{s=1}^k u_0^s \omega^s \prod_{l=s+1}^k \lambda_{1,2}^l + (u_0^0 \lambda_{0,2}^0 + u_{\mathbf{x}0}^0 \lambda_{1,2}^0) \prod_{l=1}^k \lambda_{1,2}^l + \sum_{s=0}^k \alpha_2^s \prod_{l=s+1}^k \lambda_{1,2}^l
\end{aligned}$$

Ce qui nous donne (16.247). Avec (16.245) et (16.251) on a ( $k = r + 1$ ) :

$$\begin{aligned}
1 &= \lambda_{0,2}^k + \lambda_{1,2}^k \\
1 &= \lambda_{0,2}^k + \left[ \lambda_{0,1}^{r+1} + \sum_{s=1}^r \omega^s \prod_{l=s+1}^r \lambda_{1,2}^l + (\lambda_{0,2}^0 + \lambda_{1,2}^0) \prod_{l=1}^r \lambda_{1,2}^l \right] \lambda_{1,2}^k \\
1 &= \lambda_{0,2}^k + \lambda_{0,1}^k \lambda_{1,2}^k + \sum_{s=1}^r \omega^s \prod_{l=s+1}^k \lambda_{1,2}^l + (\lambda_{0,2}^0 + \lambda_{1,2}^0) \prod_{l=1}^k \lambda_{1,2}^l \\
1 &= \sum_{s=1}^k \omega^s \prod_{l=s+1}^k \lambda_{1,2}^l + (\lambda_{0,2}^0 + \lambda_{1,2}^0) \prod_{l=1}^k \lambda_{1,2}^l
\end{aligned}$$

Ce qui nous donne (16.248).

**Cas  $r + 1 < k$  (point intérieur)** Alors pour  $r + 1$  (16.246) devient ( $u_1^{r+1} = u_2^r$ )

$$\begin{aligned}
u_1^{r+2} &= u_0^{r+2} \lambda_{0,1}^{r+2} + u_0^{r+1} \lambda_{0,2}^{r+1} + u_1^{r+1} \lambda_{1,2}^{r+1} + \alpha_2^{r+1} \\
u_1^{r+2} &= u_0^{r+2} \lambda_{0,1}^{r+2} + u_0^{r+1} \lambda_{0,2}^{r+1} + \\
&\quad \left[ u_0^{r+1} \lambda_{0,1}^{r+1} + \sum_{s=1}^r u_0^s \omega^s \prod_{l=s+1}^r \lambda_{1,2}^l + (u_0^0 \lambda_{0,2}^0 + u_{\mathbf{x}0}^0 \lambda_{1,2}^0) \prod_{l=1}^r \lambda_{1,2}^l + \sum_{s=0}^r \alpha_2^s \prod_{l=s+1}^r \lambda_{1,2}^l \right] \\
&\quad \lambda_{1,2}^{r+1} + \alpha_2^{r+1} \\
&= u_0^{r+2} \lambda_{0,1}^{r+2} + u_0^{r+1} [\lambda_{0,2}^{r+1} + \lambda_{0,1}^{r+1} \lambda_{1,2}^{r+1}] \\
&\quad + \sum_{s=1}^r u_0^s \omega^s \prod_{l=s+1}^{r+1} \lambda_{1,2}^l + (u_0^0 \lambda_{0,2}^0 + u_{\mathbf{x}0}^0 \lambda_{1,2}^0) \prod_{l=1}^{r+1} \lambda_{1,2}^l + \sum_{s=0}^r \alpha_2^s \prod_{l=s+1}^{r+1} \lambda_{1,2}^l + \alpha_2^{r+1} \\
u_2^{r+1} &= u_0^{r+2} \lambda_{0,1}^{r+2} + \sum_{s=1}^{r+1} u_0^s \omega^s \prod_{l=s+1}^{r+1} \lambda_{1,2}^l + (u_0^0 \lambda_{0,2}^0 + u_{\mathbf{x}0}^0 \lambda_{1,2}^0) \prod_{l=1}^{r+1} \lambda_{1,2}^l + \sum_{s=0}^{r+1} \alpha_2^s \prod_{l=s+1}^{r+1} \lambda_{1,2}^l
\end{aligned}$$

Ce qui donne (16.250) pour  $r + 1$ .

Pour  $r + 1$  dans (16.246) et (16.251) on a

$$\begin{aligned}
1 &= \lambda_{0,1}^{r+2} + \lambda_{0,2}^{r+1} + \lambda_{1,2}^{r+1} \\
&= \lambda_{0,1}^{r+2} + \lambda_{0,2}^{r+1} + \left[ \lambda_{0,1}^{r+1} + \sum_{s=1}^r \omega^s \prod_{l=s+1}^r \lambda_{1,2}^l + (\lambda_{0,2}^0 + \lambda_{1,2}^0) \prod_{l=1}^r \lambda_{1,2}^l \right] \lambda_{1,2}^{r+1} \\
&= \lambda_{0,1}^{r+2} + \lambda_{0,2}^{r+1} + \lambda_{0,1}^{r+1} \lambda_{1,2}^{r+1} + \sum_{s=1}^r \omega^s \prod_{l=s+1}^{r+1} \lambda_{1,2}^l + (\lambda_{0,2}^0 + \lambda_{1,2}^0) \prod_{l=1}^{r+1} \lambda_{1,2}^l \\
&= \lambda_{0,1}^{r+2} + \sum_{s=1}^{r+1} \omega^s \prod_{l=s+1}^{r+1} \lambda_{1,2}^l + (\lambda_{0,2}^0 + \lambda_{1,2}^0) \prod_{l=1}^{r+1} \lambda_{1,2}^l
\end{aligned}$$

Ce qui donne (16.251) pour  $r + 1$ .

Il reste à prouver<sup>43</sup> (16.253). La preuve se fera, par récurrence sur  $r$  de  $k - 1$  à  $0$ , en montrant que

$$u_2^r \hat{\theta}_2^k = \sum_{s=r+1}^k u_0^s \hat{\theta}_0^s + u_0^r \theta_{0,2}^r + u_1^r \theta_{1,2}^r + \sum_{s=r}^k d_2^s f_2^s \quad (**)$$

$$u_i \sum_{j \in S^+(i)} \theta_{i,j} = \sum_{j \in S^-(i)} u_j \theta_{i,j} + d_i f_i \quad (*)$$

Si on applique (\*) pour  $Q_i = Q_2^k$  et  $Q_i = Q_1^r = Q_2^{r-1}$ , on obtient :

$$u_2^k \hat{\theta}_2^k = u_0^k \theta_{0,2}^k + u_1^k \theta_{1,2}^k + d_2^k f_2^k \quad (16.255)$$

$$u_1^r \theta_{1,2}^r = u_0^r \theta_{0,1}^r + u_1^{r-1} \theta_{0,2}^{r-1} + u_1^{r-1} \theta_{1,2}^{r-1} + d_2^{r-1} f_2^{r-1} \quad (16.256)$$

**Début de l'induction :** Si  $k = 0$  il n'y a rien à montrer, car (16.255) et (16.253) sont identiques ( $u_1^0 = u_{\mathbf{x}}^0$ ). Si  $k > 1$ , alors pour  $r = k$  on obtient :

$$\begin{aligned}
u_1^k \theta_{1,2}^k &= u_0^k \theta_{0,1}^k + u_1^{k-1} \theta_{0,2}^{k-1} + u_1^{k-1} \theta_{1,2}^{k-1} + d_2^{k-1} f_2^{k-1} \text{ donc} \\
u_2^k \hat{\theta}_2^k &= u_0^k \theta_{0,2}^k + [u_0^k \theta_{0,1}^k + u_1^{k-1} \theta_{0,2}^{k-1} + u_1^{k-1} \theta_{1,2}^{k-1} + d_2^{k-1} f_2^{k-1}] + d_2^k f_2^k \\
u_2^k \hat{\theta}_2^k &= u_0^k \hat{\theta}_0^k + u_1^{k-1} \theta_{0,2}^{k-1} + u_1^{k-1} \theta_{1,2}^{k-1} + d_2^{k-1} f_2^{k-1} + d_2^k f_2^k
\end{aligned}$$

Ce qui prouve (\*\*) pour  $r = k - 1$ .

**Pas de  $r$  à  $r - 1$  :** Supposons (\*\*) vraie pour un certain  $r$ . Si  $r = 0$ , il n'y a rien à faire, car (\*\*) est équivalent à (16.253) ( $u_1^0 = u_{\mathbf{x}}^0$ ). Si  $r > 0$ , alors

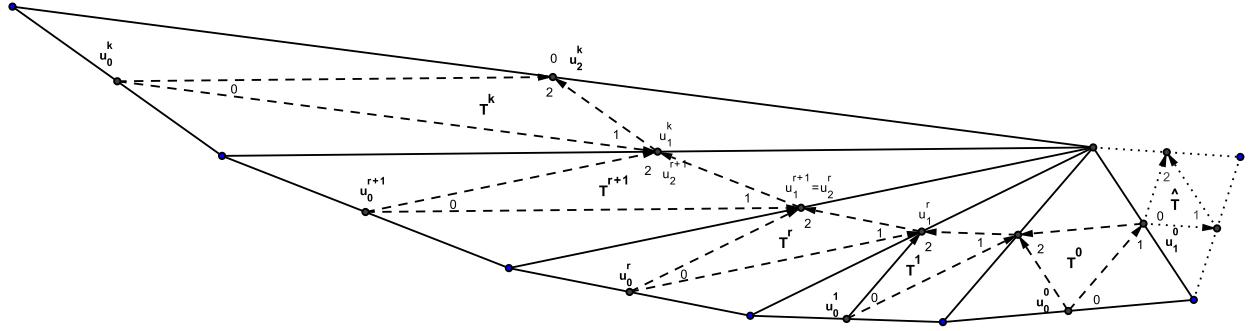
$$\begin{aligned}
u_2^r \hat{\theta}_2^k &= \sum_{s=r+1}^k u_0^s \hat{\theta}_0^s + u_0^r \theta_{0,2}^r + u_1^r \theta_{1,2}^r + \sum_{s=r}^k d_2^s f_2^s \\
u_2^r \hat{\theta}_2^k &= \sum_{s=r+1}^k u_0^s \hat{\theta}_0^s + u_0^r \theta_{0,2}^r + [u_0^r \theta_{0,1}^r + u_1^{r-1} \theta_{0,2}^{r-1} + u_1^{r-1} \theta_{1,2}^{r-1} + d_2^{r-1} f_2^{r-1}] + \sum_{s=r}^k d_2^s f_2^s \\
u_2^r \hat{\theta}_2^k &= \sum_{s=r}^k u_0^s \hat{\theta}_0^s + u_1^{r-1} \theta_{0,2}^{r-1} + u_1^{r-1} \theta_{1,2}^{r-1} + \sum_{s=r-1}^k d_2^s f_2^s
\end{aligned}$$

Ce qui prouve (\*\*) pour  $r = k - 1$ .

□

43. On peut aussi utiliser le lemme 14.1, page 110.

## Cellule sans point intérieur de type 1 – 1



La cellule représentée ci-dessus compte  $k + 1$  triangles  $T^0, T^1, \dots, T^r, T^{r+1}, \dots, T^k$ .

Les  $k + 1$  points entrants sont  $Q_0^0, \dots, Q_0^k$ , les deux points sortants sont  $Q_1^0$  et  $Q_2^k$ .

A) On a les formules de récurrences :

**Pour**  $u_1^0$  : on a  $\lambda_{0,1}^0 = 1$ , car  $Q_1^0$  ne possède qu'un seul voisin en amont (le point  $Q_0^0$ )

et  $\lambda_{0,1}^0 = \frac{\theta_{0,1}^0}{\theta_{1,2}^0 + \theta_{0,1}^r + \theta_{0,2}^r}$ . On a ainsi

$$u_1^0 = u_0^0 \lambda_{0,1}^0 + \alpha_1^0 = u_0^0 + \alpha_1^0 \quad (16.257)$$

**Pour**  $r \in \{0, \dots, k-1\}$  :

$$u_2^r = u_1^{r+1} = u_0^{r+1} \lambda_{0,1}^{r+1} + u_0^r \lambda_{0,2}^r + u_1^r \lambda_{1,2}^r + \alpha_2^r \text{ avec } \lambda_{0,1}^{r+1} + \lambda_{0,2}^r + \lambda_{1,2}^r = 1 \quad (16.258)$$

**Pour**  $r = k$  :

$$u_2^k = u_0^k \lambda_{0,2}^k + u_1^k \lambda_{1,2}^k + \alpha_2^k \text{ avec } \lambda_{0,2}^k + \lambda_{1,2}^k = 1 \quad (16.259)$$

B) Alors avec  $\omega^r = \lambda_{0,2}^r + \lambda_{0,1}^r \lambda_{1,2}^r$  on obtient les formules explicites :

**Points sortants :**

$$u_1^0 = u_0^0 + \alpha_1^0 \text{ et} \quad (16.260)$$

$$u_2^k = \sum_{r=0}^k u_0^r \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l + \sum_{r=0}^k \alpha_2^r \prod_{l=r+1}^k \lambda_{1,2}^l \quad (16.261)$$

$$1 = \sum_{s=0}^k \omega^s \prod_{l=s+1}^k \lambda_{1,2}^l \quad (16.262)$$

De (16.260) et (16.261) on déduit les inégalités :

$$|u_1^0| \leq |u_0^0| + \frac{h}{\hat{\sigma} \|\beta\|} |f_1^0| \quad \text{et} \quad |u_2^k| \leq \sum_{r=0}^k |u_0^r| \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l + \frac{h}{\hat{\sigma} \|\beta\|} \sum_{r=0}^k |f_2^r| \quad (16.263)$$

**Points intérieurs :** Pour  $r \in \{0, \dots, k-1\}$

$$u_2^r = u_0^{r+1} \lambda_{0,1}^{r+1} + \sum_{s=0}^r u_0^s \omega^s \prod_{l=s+1}^r \lambda_{1,2}^l + \sum_{s=0}^r \alpha_2^s \prod_{l=s+1}^r \lambda_{1,2}^l \quad (16.264)$$

$$1 = \lambda_{0,1}^{r+1} + \sum_{s=0}^r \omega^s \prod_{l=s+1}^k \lambda_{1,2}^l \quad (16.265)$$

De (16.264) on déduit l'inégalité :

$$|u_2^r| \leq |u_0^{r+1}| \lambda_{0,1}^{r+1} + \sum_{s=0}^r |u_0^s| \omega^s \prod_{l=s+1}^r \lambda_{1,2}^l + \frac{h}{\hat{\sigma} \|\beta\|} \sum_{s=0}^r |f_2^s| \tag{16.266}$$

C) De plus on a les égalités suivantes :

$$\hat{\theta}_0^0 = \hat{\theta}_2^k \omega^0 \prod_{l=1}^k \lambda_{1,2}^l + \hat{\theta}_1^0 \text{ et} \tag{16.267}$$

$$\hat{\theta}_0^r = \hat{\theta}_2^k \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l \text{ pour } r \in \{1, \dots, k\} \tag{16.268}$$

**Preuve :** La preuve des égalités (16.260) à (16.265) se fait par récurrence sur  $r$  comme au chapitre 16.

La preuve sera faite pour la formulation II, pour la formulation I il suffit de remplacer  $d_k$  par  $|D_k|$  dans ce qui suit.<sup>44</sup>

Il reste à montrer (16.267) et (16.268) : Avec le lemme lemme 14.1, page 110 on a

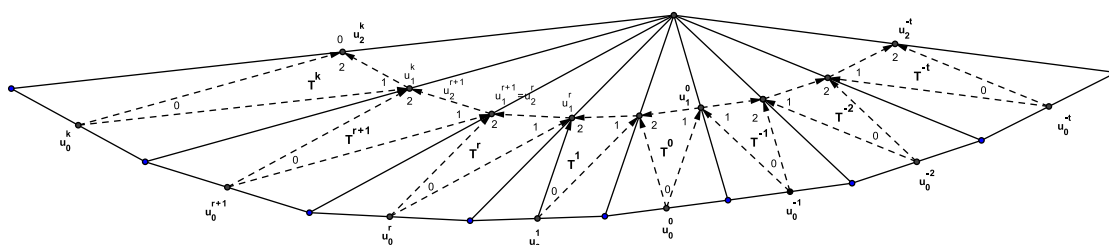
$$\sum_{r=0}^k u_0^r \hat{\theta}_0^r + \sum_{r=0}^k d_2^r f_2^r + d_1^0 f_1^0 = u_1^0 \hat{\theta}_1^0 + u_2^k \hat{\theta}_2^k \tag{16.269}$$

Si on substitue (16.260) et (16.261) dans (16.269), alors on obtient

$$\sum_{r=0}^k u_0^r \hat{\theta}_0^r + \sum_{r=0}^k d_2^r f_2^r + d_1^0 f_1^0 = [u_0^0 + \alpha_1^0] \hat{\theta}_1^0 + \left[ \sum_{r=0}^k u_0^r \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l + \sum_{r=0}^k \alpha_2^r \prod_{l=r+1}^k \lambda_{1,2}^l \right] \hat{\theta}_2^k$$

Cette égalité est toujours vraie, quelque soient les valeurs des points d'entrés  $u_0^0, \dots, u_0^k$  que l'on peut considérer comme des variables indépendantes. En identifiant dans les membres gauches et droites les coefficients de  $u_0^0, \dots, u_0^k$ , on obtient (16.267) et (16.268). □

### Cellule avec un point intérieur de type 1 – 1



La cellule représentée ci-dessus compte  $k + t + 1$  triangles  $T^k, \dots, T^0, T^{-1}, \dots, T^{-t}$  ( $k \geq 0, t \geq 1$ ).

Les  $k + t + 1$  points entrants sont  $Q_0^k, \dots, Q_0^0, Q_0^{-1}, \dots, Q_0^{-t}$ , les deux points sortants sont  $Q_2^k$  et  $Q_2^{-t}$ .

A) On a les formules de récurrences :

---

44. Rappelons que  $\alpha_i = \alpha_i^I$  dans la formulation I et  $\alpha_i = \alpha_i^{II}$  dans la formulation II, voir page 97



Pour  $r = k$  :

$$u_k^2 = u_0^k \lambda_{0,2}^k + u_1^k \lambda_{1,2}^k + \alpha_2^k \text{ avec } 1 = \lambda_{0,2}^k + \lambda_{1,2}^k \quad (16.270)$$

Pour  $r \in \{0, \dots, k-1\}$  :

$$u_2^r = u_1^{r+1} = u_0^{r+1} \lambda_{0,1}^{r+1} + u_0^r \lambda_{0,2}^r + u_1^r \lambda_{1,2}^r + \alpha_2^r \text{ avec } 1 = \lambda_{0,1}^{r+1} + \lambda_{0,2}^r + \lambda_{1,2}^r \quad (16.271)$$

Pour  $r = 0$  :

$$u_1^0 = u_1^{-1} = u_0^0 \lambda_{0,1}^0 + u_0^{-1} \lambda_{0,1}^{-1} + \alpha_1^0 \text{ avec } 1 = \lambda_{0,1}^0 + \lambda_{0,1}^{-1} \quad (16.272)$$

Pour  $r \in \{1, \dots, t-1\}$  :

$$u_2^{-r} = u_1^{-(r+1)} = u_0^{-(r+1)} \lambda_{0,1}^{-(r+1)} + u_0^{-r} \lambda_{0,2}^{-r} + u_1^{-r} \lambda_{1,2}^{-r} + \alpha_2^{-r} \text{ avec } 1 = \lambda_{0,1}^{-(r+1)} + \lambda_{0,2}^{-r} + \lambda_{1,2}^{-r} \quad (16.273)$$

Pour  $r = t$  :

$$u_2^{-t} = u_0^{-t} \lambda_{0,2}^{-t} + u_1^{-t} \lambda_{1,2}^{-t} + \alpha_2^{-t} \text{ avec } 1 = \lambda_{0,2}^{-t} + \lambda_{1,2}^{-t} \quad (16.274)$$

B) Alors avec  $\omega^r = \lambda_{0,2}^r + \lambda_{0,1}^r \lambda_{1,2}^r$  on obtient les formules explicites :

**Points sortants :**

$$\begin{aligned} u_2^k &= \sum_{r=0}^k u_0^r \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l + u_0^{-1} \lambda_{0,1}^{-1} \prod_{l=0}^k \lambda_{1,2}^l \\ &\quad + \sum_{r=0}^k \alpha_2^r \prod_{l=r+1}^k \lambda_{1,2}^l + \alpha_1^0 \prod_{l=0}^k \lambda_{1,2}^l \end{aligned} \quad (16.275)$$

$$1 = \sum_{r=0}^k \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l + \lambda_{0,1}^{-1} \prod_{l=0}^k \lambda_{1,2}^l \quad (16.276)$$

$$\begin{aligned} u_2^{-t} &= \sum_{r=1}^t u_0^{-r} \omega^{-r} \prod_{l=r+1}^t \lambda_{1,2}^{-l} + u_0^0 \lambda_{0,1}^0 \prod_{l=1}^t \lambda_{1,2}^{-l} \\ &\quad + \sum_{r=1}^t \alpha_2^{-r} \prod_{l=r+1}^t \lambda_{1,2}^{-l} + \alpha_1^0 \prod_{l=1}^t \lambda_{1,2}^{-l} \end{aligned} \quad (16.277)$$

$$1 = \sum_{r=1}^t \omega^{-r} \prod_{l=r+1}^t \lambda_{1,2}^{-l} + \lambda_{0,1}^0 \prod_{l=1}^t \lambda_{1,2}^{-l} \quad (16.278)$$

De (16.275) et (16.277) on déduit les inégalités :

$$\begin{aligned} |u_2^k| &\leq \sum_{r=0}^k |u_0^r| \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l + |u_0^{-1}| \lambda_{0,1}^{-1} \prod_{l=0}^k \lambda_{1,2}^l \\ &\quad + \frac{h}{\hat{\sigma} \|\beta\|} \sum_{r=0}^k |f_2^r| + \frac{h}{\hat{\sigma} \|\beta\|} |f_1^0| \quad \text{et} \end{aligned} \quad (16.279)$$

$$\begin{aligned} |u_2^{-t}| &\leq \sum_{r=1}^t |u_0^{-r}| \omega^{-r} \prod_{l=r+1}^t \lambda_{1,2}^{-l} + |u_0^0| \lambda_{0,1}^0 \prod_{l=1}^t \lambda_{1,2}^{-l} \\ &\quad + \frac{h}{\hat{\sigma} \|\beta\|} \sum_{r=1}^t |f_2^{-r}| + \frac{h}{\hat{\sigma} \|\beta\|} |f_1^0| \end{aligned} \quad (16.280)$$

Points intérieurs : Pour  $r \in \{0, \dots, k-1\}$  :

$$\begin{aligned} u_2^r &= u_0^{r+1} \lambda_{0,1}^{r+1} + \sum_{s=0}^r u_0^s \omega^s \prod_{l=s+1}^r \lambda_{1,2}^l + u_0^{-1} \lambda_{0,1}^{-1} \prod_{l=0}^r \lambda_{1,2}^l \\ &\quad + \sum_{s=0}^r \alpha_2^s \prod_{l=s+1}^r \lambda_{1,2}^l + \alpha_1^0 \prod_{l=0}^r \lambda_{1,2}^l \end{aligned} \quad (16.281)$$

$$1 = \lambda_{0,1}^{r+1} + \sum_{s=0}^r \omega^s \prod_{l=s+1}^r \lambda_{1,2}^l + \lambda_{0,1}^{-1} \prod_{l=0}^r \lambda_{1,2}^l \quad (16.282)$$

$$(16.283)$$

Pour  $r = 0$  :

$$u_1^0 = u_1^{-1} = u_0^0 \lambda_{0,1}^0 + u_0^{-1} \lambda_{0,1}^{-1} + \alpha_1^0 \quad (16.284)$$

$$1 = \lambda_{0,1}^0 + \lambda_{0,1}^{-1} \quad (16.285)$$

Pour  $r \in \{1, \dots, t-1\}$  :

$$\begin{aligned} u_2^{-r} &= u_0^{-(r+1)} \lambda_{0,1}^{-(r+1)} + \sum_{s=1}^r u_0^{-s} \omega^{-s} \prod_{l=s+1}^r \lambda_{1,2}^{-l} + u_0^0 \lambda_{0,1}^0 \prod_{l=1}^r \lambda_{1,2}^{-l} \\ &\quad + \sum_{s=1}^r \alpha_2^{-s} \prod_{l=s+1}^r \lambda_{1,2}^{-l} + \alpha_1^0 \prod_{l=1}^r \lambda_{1,2}^{-l} \end{aligned} \quad (16.286)$$

$$1 = \lambda_{0,1}^{-(r+1)} + \sum_{s=1}^r \omega^{-s} \prod_{l=s+1}^r \lambda_{1,2}^{-l} + \lambda_{0,1}^0 \prod_{l=1}^r \lambda_{1,2}^{-l} \quad (16.287)$$

De (16.281), (16.284) et (16.286) on déduit les inégalités :

Pour  $r \in \{0, \dots, k-1\}$  :

$$\begin{aligned} |u_2^r| &\leq |u_0^{r+1}| \lambda_{0,1}^{r+1} + \sum_{s=0}^r |u_0^s| \omega^s \prod_{l=s+1}^r \lambda_{1,2}^l + |u_0^{-1}| \lambda_{0,1}^{-1} \prod_{l=0}^r \lambda_{1,2}^l \\ &\quad + \frac{h}{\hat{\sigma} \|\beta\|} \sum_{s=0}^r |f_2^s| + \frac{h}{\hat{\sigma} \|\beta\|} |f_1^0| \end{aligned} \quad (16.288)$$

Pour  $r = 0$  :

$$|u_1^0| = |u_1^{-1}| \leq |u_0^0| \lambda_{0,1}^0 + |u_0^{-1}| \lambda_{0,1}^{-1} + \frac{h}{\hat{\sigma} \|\beta\|} |f_1^0| \quad (16.289)$$

Pour  $r \in \{1, \dots, t-1\}$  :

$$\begin{aligned} |u_2^{-r}| &\leq |u_0^{-(r+1)}| \lambda_{0,1}^{-(r+1)} + \sum_{s=1}^r |u_0^{-s}| \omega^{-s} \prod_{l=s+1}^r \lambda_{1,2}^{-l} + |u_0^0| \lambda_{0,1}^0 \prod_{l=1}^r \lambda_{1,2}^{-l} \\ &\quad + \frac{h}{\hat{\sigma} \|\beta\|} \sum_{s=1}^r |f_2^{-s}| + \frac{h}{\hat{\sigma} \|\beta\|} |f_1^0| \end{aligned} \quad (16.290)$$

C) De plus on a les égalités suivantes :

$$\hat{\theta}_0^r = \hat{\theta}_2^k \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l \text{ pour } r \in \{1, \dots, k\} \quad (16.291)$$

$$\hat{\theta}_0^0 = \hat{\theta}_2^k \omega^0 \prod_{l=1}^k \lambda_{1,2}^l + \hat{\theta}_2^{-t} \lambda_{0,1}^0 \prod_{l=2}^t \lambda_{1,2}^{-l} \quad (16.292)$$

$$\hat{\theta}_0^{-1} = \hat{\theta}_2^k \lambda_{0,1}^{-1} \prod_{l=1}^k \lambda_{1,2}^l + \hat{\theta}_2^{-t} \omega^{-1} \prod_{l=2}^t \lambda_{1,2}^{-l} \quad (16.293)$$

$$\hat{\theta}_0^{-r} = \hat{\theta}_2^{-t} \omega^{-r} \prod_{l=r+1}^t \lambda_{1,2}^{-l} \text{ pour } r \in \{2, \dots, t\} \quad (16.294)$$

**Preuve :** La preuve des égalités (16.275) à (16.269) se fait par deux récurrences sur  $r$  ( $r$  allant de 0 à  $k$ , puis  $r$  allant de 1 à  $t$ ) comme au chapitre 16.

La preuve sera faite pour la formulation I, pour la formulation II il suffit de remplacer  $|D_k|$  par  $d_k$  dans ce qui suit.<sup>45</sup>

Il reste à montrer les égalités (16.291) à (16.294) : Avec le lemme lemme 14.1, page 110 on a

$$\sum_{r=-t}^k u_0^r \hat{\theta}_0^r + \sum_{r=0}^k d_2^r f_2^r + f_1^0 |D_1^0| + \sum_{r=1}^t |D_2^{-r}| f_2^{-r} = u_2^{-t} \hat{\theta}_2^{-t} + u_2^k \hat{\theta}_2^k \quad (16.295)$$

Si on substitue (16.275), et (16.277) dans (16.295), alors on obtient une égalité qui reste toujours vraie, quelque soient les valeurs des points d'entrés  $u_0^{-t}, \dots, u_0^k$ , qui peuvent être considérés comme des variables indépendantes. En identifiant dans les membres gauches et droites les coefficients de  $u_0^{-t}, \dots, u_0^k$ , on obtient (16.291) à (16.294).

□

## 17 Stabilité du problème approché

L'objectif est de montrer les équivalents de (1.36) et (1.19) , page 25 et 27, pour le problème approché.

### 17.1 Majoration de $\|u_h\|_{L_h^\infty(\Omega)}$

La majoration qui suit est l'équivalent de (1.36) pour le problème approché.

**Théorème 17.1** ( $\|u_h\|_{L_h^\infty(\Omega)}$ -stabilité) *Soit  $(\mathcal{T}_h)$  une famille de maillage quasi-uniforme et régulière, soit  $\mathcal{G}_h$  le graphe associé à  $\mathcal{T}_h$ . Soit  $\mathcal{C}_h$  l'ensemble des cellules obtenues à partir de  $\mathcal{G}_h$ . On suppose que :*

---

45. Rappelons que  $\alpha_i = \alpha_i^I$  dans la formulation I et  $\alpha_i = \alpha_i^{II}$  dans la formulation II, voir page 97

- a)  $\mathcal{C}_h$  contient pas de point intérieur de type 0 – 1 ou 0 – 2 (donc  $\mathcal{G}_h$  ne contient pas de boucles).
- b) Il existe  $K_\Xi > 0$ , ne dépendant pas de  $h$ , tel que pour le nombre de strates  $N_\Xi$  on ait  $N = N_\Xi \leq K_\Xi \frac{1}{h}$ ,

Alors pour la solution  $u_h$  du problème variationnel I (voir page 95), on a :

$$\|u_h\|_{L_h^\infty(\Omega)} \leq \|u_h\|_{L_h^\infty(\Gamma^-)} + 2K_\Xi K_{cell} \frac{1}{\hat{\sigma} \|\beta\|} \|f\|_{L_h^\infty} \quad (17.296)$$

avec  $K_{cell} \leq 2 \frac{\ln(4\sigma c_0)}{\ln 2}$ .

**Preuve :**

**Etape 1 : Estimation sur les points d'entrés (points de contrôles)**

On commence par définir le vecteur colonne  $\xi$  qui contient les valeurs sur les points d'entrés des cellules de  $C \in \mathcal{C}_h$ .

Soit  $m = \text{card}(\bigcup_{C \in \mathcal{C}_h} J_{\text{entrant}}^C)$  le nombre total de points d'entrés et  $m_l = \text{card}(\bigcup_{C \in \Xi_l} J_{\text{entrant}}^C \setminus J^-)$  le nombre de points d'entrés dans la strate  $\Xi_l$  qui ne sont pas sur  $\Gamma^-$ . Le transposé du vecteur  $\xi \in \mathbb{R}^m$  est le vecteur ligne défini<sup>46</sup> par

$$\begin{aligned} \xi = & \left( \underbrace{\overset{\xi_1}{\downarrow} \cdots, |u_h(Q_j)|, \cdots}_{j \in \bigcup_{C \in \Xi_N} J_{\text{entrant}}^C \setminus J^-}, \underbrace{\overset{\xi_{\tau_N}}{\downarrow} \cdots, |u_h(Q_j)|, \cdots}_{j \in \bigcup_{C \in \Xi_{N-1}} J_{\text{entrant}}^C \setminus J^-}, \cdots, \right. \\ & \underbrace{\overset{\xi_{\tau_{l+2}+1}}{\downarrow} \cdots, |u_h(Q_j)|, \cdots}_{j \in \bigcup_{C \in \Xi_{l+1}} J_{\text{entrant}}^C \setminus J^-}, \underbrace{\overset{\xi_{\tau_{l+1}+1}}{\downarrow} \cdots, |u_h(Q_j)|, \cdots}_{j \in \bigcup_{C \in \Xi_l} J_{\text{entrant}}^C \setminus J^-}, \\ & \left. \cdots, \underbrace{\overset{\xi_{\tau_3+1}}{\downarrow} \cdots, |u_h(Q_j)|, \cdots}_{j \in \bigcup_{C \in \Xi_2} J_{\text{entrant}}^C \setminus J^-}, \underbrace{\overset{\xi_{\tau_2}}{\downarrow} \cdots, |u_h(Q_j)|, \cdots}_{j \in \bigcup_{C \in \Xi_1} J_{\text{entrant}}^C \setminus J^-}, \underbrace{\overset{\xi_{\tau_1}}{\downarrow} \cdots, |u_h(Q_j)|, \cdots}_{j \in J^-} \right)^{tr} \end{aligned} \quad (17.297)$$

où  $\tau_N = m_N$ ,  $\tau_{N-1} = m_N + m_{N-1}$  et  $\tau_l = \sum_{k=l}^N m_k$  indique la position du dernier point entrant de la strate  $\Xi_l$ .

Les coordonnées de  $\xi^{tr}$  sont donc groupées par strates (sans les points de  $\Gamma^-$ ), où la strate la plus haute ( $\Xi_N$ ,  $N = N_\Xi$ ) est à gauche et la plus basse ( $\Xi_1$ ) est à droite. Les points de  $\Gamma^-$  étant regroupés à part en derniers.

Une inégalité de la forme

$$\xi \leq M\xi + \gamma \quad (17.298)$$

va maintenant être établie pour  $\xi$  où  $M$  est une matrice positive  $M = (m_{d,e})_{1 \leq d \leq m, 1 \leq e \leq m}$  et  $\gamma \in \mathbb{R}^m$  est un vecteur positif qui correspond à la contribution de  $f_h$ .

L'espace vectoriel  $\mathbb{R}^m$  est muni de la norme  $\|\xi\|_\infty = \max_{m \geq d \geq 1} |\xi_d|$ .

46.  $\xi^{tr}$  est le transposé de  $\xi$

La norme matricielle induite est alors  $\|M\|_\infty = \max_{m \geq d \geq 1} \sum_{e=1}^m |m_{d,e}|$ .

Soit  $l \in \{1, 2, \dots, N\}$  le numéro d'une strate et  $j \in \bigcup_{C \in \Xi_l} J_{\text{entrant}}^C$  un point entrant quelconque de la strate  $\Xi_l$ , alors

**Cas  $Q_j \in \Gamma^-$  :** On a  $\xi_d = |u_h(Q_j)| = |g_h(Q_j)|$  pour  $d \in \{\tau_1 + 1, \tau_1 + 2, \dots, m\}$ .

On définit alors la ligne  $d$  de la matrice  $M$  par  $m_{d,e} = \begin{cases} 1 & \text{si } e=d \\ 0 & \text{si } e \neq d \end{cases}$  et on pose  $\gamma_d = 0$ .

**Cas  $Q_j \notin \Gamma^-$  :** Alors il existe exactement une strate inférieure  $\Xi_q$  avec<sup>47</sup>  $l > q \geq 1$  et une cellule  $C \in \Xi_q$ , telle que  $Q_j$  soit un point sortant de  $C$  ( $j \in J_{\text{sortant}}^C$ ) et  $\xi_d = |u_h(Q_j)|$  pour  $d \in \{\tau_{q+1} + 1, \dots, \tau_q\}$ .  $\Xi_d$  peut être majorée explicitement – voir ci-dessous le calcul pour les différents types de cellules – par les valeurs des points d'entrés de  $C \in \Xi_d$ , c'est-à-dire par des valeurs de  $\xi_e$  avec  $\tau_l + 1 \geq \tau_{q+1} + 1 \geq e \geq \tau_{q-1}$ .

Ainsi, pour la ligne  $d$  de  $M$  on a  $m_{d,1} = 0, m_{d,2} = 0, \dots, m_{d,\tau_l} = 0$ .

$$\xi \leq \left( \begin{array}{cccccccccccccccccccc} \xi_1 & \dots & \xi_{\tau_N} & \xi_{\tau_N+1} & \dots & \xi_{\tau_{N-1}} & \dots & \dots & \xi_{\tau_{l+1}+1} & \dots & \xi_d & \dots & \xi_{\tau_l} & \dots & \dots & \xi_{\tau_1} & \dots & \xi_{\tau_1+1} & \dots & \xi_m \\ 0 & \dots & 0 & * & \dots & * & \dots & \dots & * & \dots & * & \dots & * & \dots & \dots & * & \dots & * & \dots & * & \dots \\ \vdots & \vdots \\ 0 & \dots & 0 & * & \dots & * & * & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & * \\ 0 & \dots & 0 & 0 & \dots & 0 & * & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & * \\ \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & * & \dots & * & \dots & * & * & * & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & * \\ \vdots & \vdots \\ 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & * & \dots & * & \dots & * & \dots & \dots & \dots & \dots & \dots & \dots & \dots & * \\ \vdots & \vdots \\ 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 & \dots & 0 & \dots & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & * \\ \vdots & \vdots \\ 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 & \dots & 0 & \dots & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & * \\ \vdots & \vdots \\ 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 & \dots & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & * \\ \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \end{array} \right) + \gamma$$

avec

$$\gamma^{tr} = (\gamma_1 \dots \gamma_{\tau_N} \gamma_{\tau_N+1} \dots \gamma_{\tau_{N-1}} \dots \gamma_{\tau_{l+1}+1} \cdot \gamma_d \cdot \gamma_{\tau_l} \dots \gamma_{\tau_{q+1}+1} \dots \gamma_{\tau_q} \dots \gamma_{\tau_2+1} \dots \gamma_{\tau_1} 0 \dots 0)$$

(17.299)

Le vecteur  $\gamma$  a la structure  $\gamma = \begin{pmatrix} c \\ 0_{m-\tau_1} \end{pmatrix}$  et pour la matrice  $M$  on a

$$M = \left( \begin{array}{c|c} A_{\tau_1 \times \tau_1} & B_{\tau_1 \times (m-\tau_1)} \\ \hline 0_{(m-\tau_1) \times \tau_1} & Id_{(m-\tau_1) \times (m-\tau_1)} \end{array} \right) \tag{17.300}$$

où  $A_{\tau_1 \times \tau_1}$  est une matrice nilpotente :  $A^N = 0_{\tau_1 \times \tau_1}$ . Cette dernière propriété est une conséquence directe de la structure en strates des cellules. Pour  $l \in \mathbb{N}$  on calcule :

47. Tous les points entrants de la strate  $\Xi_1$  sont dans  $\Gamma^-$ , donc  $l > 1$ .

$$M^l = \left( \frac{A_{\tau_1 \times \tau_1}^l \mid \sum_{\mu=0}^{l-1} A_{\tau_1 \times \tau_1}^\mu B_{\tau_1 \times (m-\tau_1)}}{0_{(m-\tau_1) \times \tau_1} \mid Id_{(m-\tau_1) \times (m-\tau_1)}} \right) \quad (17.301)$$

En particulier pour  $l = N$  on a

$$M^N = \left( \frac{0_{\tau_1 \times \tau_1} \mid \sum_{\mu=0}^{N-1} A_{\tau_1 \times \tau_1}^\mu B_{\tau_1 \times (m-\tau_1)}}{0_{(m-\tau_1) \times \tau_1} \mid Id_{(m-\tau_1) \times (m-\tau_1)}} \right) \quad (17.302)$$

Soit  $Q_j \notin \Gamma^-$  un point sortant d'une cellule  $C \in \Xi_q$ .

**Majorations pour les trois types de cellules :**

**Avec deux coordonnées locales 0 :** Cette cellule n'a qu'un point sortant  $Q_j = Q_2^k$  et  $\gamma_d = |u_2^k|$ . Il existe  $\xi_{\tilde{e}}, \xi_{e_0}, \xi_{e_1}, \dots, \xi_{e_k}$ , avec  $\tau_l > \tilde{e}, \tau_l > e_0, \dots, \tau_1 > e_k$  tels que  $\xi_{\tilde{e}} = |u_{\tilde{x}}^0|, \xi_{e_0} = |u_0^r|, \dots, \xi_{e_k} = |u_0^r|$ . Alors avec

$$m_{d,e} = \begin{cases} \lambda_{1,2}^0 \prod_{l=1}^k \lambda_{1,2}^l & \text{pour } e = \tilde{e} \\ \lambda_{0,2}^0 \prod_{l=1}^k \lambda_{1,2}^l & \text{pour } e = \tilde{e} \\ \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l & \text{pour } e \in \{e_1, \dots, e_k\} \\ 0 & \text{sinon.} \end{cases} \quad \text{et } \gamma_d = \frac{h}{\hat{\sigma} \|\beta\|} (k+1) \|f\|_{L_h^\infty} \quad (17.303)$$

et l'inégalité (16.249) on a :

- a)  $m_{d,e} \geq 0$  pour  $e \in \{1, \dots, m\}$ ,
- b)  $\xi_d \leq \sum_{e=1}^k m_{d,e} \xi_e + \gamma_d$ ,
- b)  $\sum_{e=1}^k m_{d,e} = 1$  avec l'égalité (16.248)

**Sans point intérieur de type 1 – 1 :** Les mêmes conclusions s'obtiennent avec les inégalités (16.263) et l'égalité (16.262). (Donc  $\xi_d = |u_2^k|$  ou  $\xi_d = |u_1^0|$ .)

**Avec un point intérieur de type 1 – 1 :** Avec les inégalités (16.279), (16.280) et les égalités (16.276) et (16.278) on obtient  $\gamma_d = \frac{h}{\hat{\sigma} \|\beta\|} (k+t+1) \|f\|_{L_h^\infty}$  et  $m_{d,e} \geq 0$  pour  $e \in \{1, \dots, m\}$ . (Donc  $\xi_d = |u_2^k|$  ou  $\xi_d = |u_2^{-t}|$ .) Des égalités (16.276) et (16.278) on déduit que  $\sum_{e=1}^m m_{d,e} = 1$ .

Le lemme 14.3 donne une majoration du nombre maximal  $N_C$  de triangles possibles dans une cellule  $C$  :

$$N_C \leq K_{cell} \quad \text{où } K_{cell} \leq 2 \frac{\ln(4\sigma c_0)}{\ln 2} \quad (17.304)$$

où  $\sigma$  et  $c_0$  sont les constantes liées à la régularité et la quasi-uniformité du maillage. La constante  $K_{cell}$  ne dépend ni de  $h$ , ni de  $C \in \mathcal{C}_h$ .

**Conclusion : Quelque soit le type de cellule, on déduit que**

- i)  $m_{d,e} \geq 0$  pour  $e \in \{1, \dots, m\}$ ,
- ii)  $\sum_{e=1}^m m_{d,e} = 1$ ,
- ii)  $\gamma_d \leq \frac{h}{\hat{\sigma} \|\beta\|} K_{cell} \|f\|_{L_h^\infty}$

Dans la norme induite on a donc  $\|M\|_{L_h^\infty} \leq 1$ .

Il est maintenant possible de trouver une majoration de  $\|\xi\|_\infty$  :

$$\begin{aligned}
\xi &\leq M\xi + \gamma \text{ donc} \\
\underline{0} &\leq -\xi + M\xi + \gamma \text{ la matrice est positive, donc} \\
\underline{0} = M\underline{0} &\leq -M\xi + M^2\xi + M\gamma \text{ donc} \\
M\xi &\leq M^2\xi + M\gamma \text{ Maintenant, on ajoute } \gamma \text{ de chaque côté, donc} \\
\xi &\leq M^2\xi + M\gamma + \gamma \text{ En procédant par itération, on obtient} \\
\xi &\leq M^N\xi + \sum_{s=0}^{N-1} M^s\gamma \text{ Donc :} \\
\xi &\leq M^N\xi + \sum_{s=0}^{N-1} \left( \frac{A_{\tau_1 \times \tau_1}^s \mid \sum_{\mu=0}^{\mu-1} A_{\tau_1 \times \tau_1}^\mu B_{\tau_1 \times (m-\tau_1)}}{0_{(m-\tau_1) \times \tau_1} \mid Id_{(m-\tau_1) \times (m-\tau_1)}} \right) \begin{pmatrix} c \\ 0_{m-\tau_1} \end{pmatrix} \text{ avec} \\
M^N\xi &= \left( \frac{0_{\tau_1 \times \tau_1} \mid \sum_{\mu=0}^{N-1} A_{\tau_1 \times \tau_1}^\mu B_{\tau_1 \times (m-\tau_1)}}{0_{(m-\tau_1) \times \tau_1} \mid Id_{(m-\tau_1) \times (m-\tau_1)}} \right) \xi \tag{17.305}
\end{aligned}$$

Les coordonnées  $\xi_{\tau_1+1}$  à  $\xi_m$  correspondent aux données du bord entrant et sont connues. De (17.305) on déduit que  $M^N\xi$  ne dépend que de  $\xi_{\tau_1+1}$  à  $\xi_m$ . On a  $\|M^N\|_\infty \leq \|M\|_\infty^N \leq 1$ , donc

$$\|M^N\xi\|_\infty \leq \max_{\mu \in \{\tau_1+1, \dots, m\}} \xi_\mu = \max_{j \in J^-} |u_h(Q_j)| = \|u_h\|_{L_h^\infty(\Gamma^-)}$$

Ainsi :

$$\|\xi\|_\infty \leq \|u_h\|_{L_h^\infty(\Gamma^-)} + \left\| \sum_{s=0}^{N-1} A_{\tau_1 \times \tau_1}^s c \right\|_\infty$$

Pour la sous matrice  $A_{\tau_1 \times \tau_1}$  extraite de  $M$  on a aussi  $\|A_{\tau_1 \times \tau_1}\|_\infty \leq 1$  (sur le sous-espace vectoriel  $\mathbb{R}^{\tau_1}$ ), donc

$$\begin{aligned}
\|\xi\|_\infty &\leq \|u_h\|_{L_h^\infty(\Gamma^-)} + \sum_{s=0}^{N-1} 1^s \|c\|_\infty \\
&\leq \|u_h\|_{L_h^\infty(\Gamma^-)} + K_\Xi \frac{1}{h} \frac{h}{\hat{\sigma} \|\beta\|} K_{cell} \|f\|_{L_h^\infty} \\
\|\xi\|_\infty &\leq \|u_h\|_{L_h^\infty(\Gamma^-)} + K_\Xi K_{cell} \frac{1}{\hat{\sigma} \|\beta\|} \|f\|_{L_h^\infty}
\end{aligned}$$

## Etape 2 : Estimation sur les points intérieurs et les points sortants des cellules

De nouveau trois types de cellules  $C \in \mathcal{C}_h$  sont à étudier :

**Avec deux coordonnées locales 0 :** En combinant l'inégalité (16.252) et l'égalité (16.251), on déduit pour les points intérieurs  $r \in \{0, \dots, k-1\}$  ( $1 > h > 0$ ) :

$$\begin{aligned}
|u_2^r| &\leq \|\xi\|_\infty + \frac{h}{\hat{\sigma} \|\beta\|} \sum_{s=0}^r |f_2^s| \text{ Donc} \\
|u_2^r| &\leq \|u_h\|_{L_h^\infty(\Gamma^-)} + K_\Xi K_{cell} \frac{h}{\hat{\sigma} \|\beta\|} \|f\|_{L_h^\infty}
\end{aligned}$$

Pour le point sortant on a :

$$|u_2^k| \leq \|\xi\|_\infty + K_{cell} K_{cell} \frac{h}{\hat{\sigma} \|\beta\|} \|f\|_{L_h^\infty} \tag{17.306}$$

**Sans point intérieur de type 1 – 1 :** Les mêmes conclusions s'obtiennent avec les inégalités (16.266) et l'égalité (16.265).

**Avec un point intérieur de type 1 – 1 :** Avec les inégalités (16.288), (16.289) et (16.290) combinées avec les égalités (16.282), (16.285) et (16.287), on peut conclure de la même façon.

Conclusion

$$\|u_h\|_{L_h^\infty(\Omega)} \leq \|u_h\|_{L_h^\infty(\Gamma^-)} + 2K_\Xi K_{cell} \frac{1}{\hat{\sigma} \|\beta\|} \|f\|_{L_h^\infty}$$

□

## 17.2 Majoration de $\|u_h\|_{L_h^1(\Omega)}$ et $\|u_h\|_{L_h^2(\Omega)}$

La majoration qui suit est l'équivalent de (1.19) pour le problème approché.

**Théorème 17.2 ( $\|u_h\|_{L_h^1(\Omega)}$ -stabilité)** Soit  $(\mathcal{T}_h)$  une famille de maillage quasi-uniforme et régulière, soit  $\mathcal{G}_h$  le graphe associé à  $\mathcal{T}_h$ . Soit  $\mathcal{C}_h$  l'ensemble des cellules obtenues à partir de  $\mathcal{G}_h$ . On suppose que :

- a)  $\mathcal{C}_h$  contient pas de point intérieur de type 0 – 1 ou 0 – 2 (donc  $\mathcal{G}_h$  ne contient pas de circuit élémentaire).
- b) Il existe  $K_\Xi > 0$ , ne dépendant pas de  $h$ , tel que pour le nombre de strates  $N_\Xi$  on ait  $N = N_\Xi \leq K_\Xi \frac{1}{h}$ ,

Alors pour la solution  $u_h$  du problème variationnel I (voir page 95), on a :

$$\|u_h\|_{L_h^1(\Omega)} \leq (K_{cell} + 2) \frac{1}{c_{\hat{\theta}}} (K_\Xi + 1) \|g_h\|_{L_h^1(\Gamma^-)} + \frac{1}{\|\beta\|} (K_{cell} + 2) 6\sigma^2 c_0^2 \frac{1}{\hat{\sigma}} \left[ \frac{1}{c_{\hat{\theta}}} K_\Xi + 1 \right] \|f_h\|_{L_h^1(\Omega)} \quad (17.307)$$

### Preuve du théorème de $L_h^1$ -stabilité

#### Etape 1 : Estimation sur les points d'entrées (points de contrôles)

On commence par définir le vecteur colonne  $\xi$  qui contient les valeurs sur les points d'entrées des cellules de  $C \in \mathcal{C}_h$ .

Soit  $m = \text{card}\left(\bigcup_{C \in \mathcal{C}_h} J_{\text{entrant}}^C\right)$  le nombre total de points d'entrées et  $m_t = \text{card}\left(\bigcup_{C \in \Xi_t} J_{\text{entrant}}^C \setminus J^-\right)$  le nombre de points d'entrées dans la strate  $\Xi_t$  qui ne



sont pas sur  $\Gamma^-$ . Le transposé du vecteur  $\xi \in \mathbb{R}^m$  est le vecteur ligne défini<sup>48</sup> par

$$\begin{aligned} \xi = & \left( \underbrace{\overset{\xi_1}{\downarrow} \cdots, |u_h(Q_j)| h\hat{\theta}_j, \cdots \overset{\xi_{\tau_N}}{\downarrow}}_{j \in \bigcup_{C \in \Xi_N} J_{\text{entrant}}^C \setminus J^-}, \underbrace{\overset{\xi_{\tau_N+1}}{\downarrow} \cdots, |u_h(Q_j)| h\hat{\theta}_j, \cdots \overset{\xi_{\tau_{N-1}}}{\downarrow}}_{j \in \bigcup_{C \in \Xi_{N-1}} J_{\text{entrant}}^C \setminus J^-}, \cdots, \right. \\ & \left. \underbrace{\overset{\xi_{\tau_t+2+1}}{\downarrow} \cdots, |u_h(Q_j)| h\hat{\theta}_j, \cdots \overset{\xi_{\tau_t+1}}{\downarrow}}_{j \in \bigcup_{C \in \Xi_{t+1}} J_{\text{entrant}}^C \setminus J^-}, \underbrace{\overset{\xi_{\tau_t+1+1}}{\downarrow} \cdots, |u_h(Q_j)| h\hat{\theta}_j, \cdots \overset{\xi_{\tau_t}}{\downarrow}}_{j \in \bigcup_{C \in \Xi_t} J_{\text{entrant}}^C \setminus J^-}, \cdots, \right. \\ & \left. \underbrace{\overset{\xi_{\tau_3+1}}{\downarrow} \cdots, |u_h(Q_j)| h\hat{\theta}_j, \cdots \overset{\xi_{\tau_2}}{\downarrow}}_{j \in \bigcup_{C \in \Xi_2} J_{\text{entrant}}^C \setminus J^-}, \underbrace{\overset{\xi_{\tau_2+1}}{\downarrow} \cdots, |u_h(Q_j)| h\hat{\theta}_j, \cdots \overset{\xi_{\tau_1}}{\downarrow}}_{j \in \bigcup_{C \in \Xi_1} J_{\text{entrant}}^C \setminus J^-}, \underbrace{\overset{\xi_{\tau_1+1}}{\downarrow} \cdots, |u_h(Q_j)| h\hat{\theta}_j, \cdots \overset{\xi_m}{\downarrow}}_{j \in J^-} \right)^{tr} \end{aligned} \quad (17.308)$$

où  $\tau_N = m_N$ ,  $\tau_{N-1} = m_N + m_{N-1}$  et  $\tau_t = \sum_{k=t}^N m_k$  indique la position du dernier point entrant de la strate  $\Xi_t$ .

Les coordonnées de  $\xi^{tr}$  sont donc groupées par strates (sans les points de  $\Gamma^-$ ), où la strate la plus haute ( $\Xi_N$ ,  $N = N_\Xi$ ) est à gauche et la plus basse ( $\Xi_1$ ) est à droite. Les points de  $\Gamma^-$  étant regroupés à part en derniers.

Une inégalité de la forme

$$\xi \leq M\xi + \gamma \quad (17.309)$$

va maintenant être établie pour  $\xi$  où  $M$  est une matrice positive  $M = (m_{d,e})_{1 \leq d \leq m, 1 \leq e \leq m}$  et  $\gamma \in \mathbb{R}^m$  est un vecteur positif qui correspond à la contribution de  $f_h$ .

L'espace vectoriel  $\mathbb{R}^m$  est muni de la norme  $\|\xi\|_1 = \sum_{d=1}^m |\xi_d|$ .

La norme matricielle induite est alors  $\|M\|_1 = \max_{m \geq e \geq 1} \sum_{d=1}^m |m_{d,e}|$ .

Soit  $t \in \{1, 2, \dots, N\}$  le numéro d'une strate et  $j \in \bigcup_{C \in \Xi_{t+1}} J_{\text{entrant}}^C$  un point entrant quelconque de la strate  $\Xi_t$ , alors

**Cas**  $Q_j \in \Gamma^-$  : On a  $\xi_d = |u_h(Q_j)| h\hat{\theta}_j = |g_h(Q_j)| h\hat{\theta}_j$  pour  $d \in \{\tau_1 + 1, \tau_1 + 2, \dots, m\}$ .

On définit alors la ligne  $d$  de la matrice  $M$  par  $m_{d,e} = \begin{cases} 1 & \text{si } e=d \\ 0 & \text{si } e \neq d \end{cases}$  et on pose  $\gamma_d = 0$ .

**Cas**  $Q_j \notin \Gamma^-$  : Alors il existe exactement une strate inférieure  $\Xi_q$  avec<sup>49</sup>  $t > q \geq 1$  et une cellule  $C \in \Xi_q$ , telle que  $Q_j$  soit un point sortant de  $C$  ( $j \in J_{\text{sortant}}^C$ ) et  $\xi_d = |u_h(Q_j)| h\hat{\theta}_j$  pour  $d \in \{\tau_{q+1} + 1, \dots, \tau_q\}$ .  $\Xi_d$  peut être majoré explicitement – voir le calcul des différentes cellules ci-dessous – par les valeurs des points d'entrés de  $C \in \Xi_d$ , c'est-à-dire par des valeurs de  $\xi_e$  avec  $\tau_t + 1 \geq \tau_{q+1} + 1 \geq e \geq \tau_{q-1}$ .

Ainsi, pour la ligne  $d$  de  $M$  on a  $m_{d,1} = 0, m_{d,2} = 0, \dots, m_{d,\tau_t} = 0$ .

Voir aussi (17.299), page 140 pour représentation de la matrice  $M$ .

Le vecteur  $\gamma$  a la structure  $\gamma = \begin{pmatrix} c \\ 0_{m-\tau_1} \end{pmatrix}$  et pour la matrice  $M$  on a

$$M = \left( \begin{array}{c|c} A_{\tau_1 \times \tau_1} & B_{\tau_1 \times (m-\tau_1)} \\ \hline 0_{(m-\tau_1) \times \tau_1} & Id_{(m-\tau_1) \times (m-\tau_1)} \end{array} \right) \quad (17.310)$$

48.  $\xi^{tr}$  est le transposé de  $\xi$

49. Tous les points entrants de la strate  $\Xi_1$  sont dans  $\Gamma^-$ , donc  $t > 1$ .

où  $A_{\tau_1 \times \tau_1}$  est une matrice nilpotente :  $A^N = 0_{\tau_1 \times \tau_1}$ . Cette dernière propriété est une conséquence directe de la structure en strates des cellules. Pour  $t \in \mathbb{N}$  on calcule :

$$M^t = \left( \begin{array}{c|c} A_{\tau_1 \times \tau_1}^t & \sum_{\mu=0}^{t-1} A_{\tau_1 \times \tau_1}^\mu B_{\tau_1 \times (m-\tau_1)} \\ \hline 0_{(m-\tau_1) \times \tau_1} & Id_{(m-\tau_1) \times (m-\tau_1)} \end{array} \right) \quad (17.311)$$

En particulier pour  $t = N$  on a

$$M^N = \left( \begin{array}{c|c} 0_{\tau_1 \times \tau_1} & \sum_{\mu=0}^{N-1} A_{\tau_1 \times \tau_1}^\mu B_{\tau_1 \times (m-\tau_1)} \\ \hline 0_{(m-\tau_1) \times \tau_1} & Id_{(m-\tau_1) \times (m-\tau_1)} \end{array} \right) \quad (17.312)$$

A la différence de l'estimation en norme  $\infty$ , il est nécessaire de considérer ensemble tous les points sortants d'une même cellule.

### Majorations pour les trois types de cellules :

**Avec deux coordonnées locales 0 :** Cette cellule n'a qu'un point sortant  $Q_j = Q_2^k$  et, avec (16.253), on a

$$\xi_d = |u_2^k| h \hat{\theta}_2^k = \sum_{r=1}^k u_0^r h \hat{\theta}_0^r + u_0^0 h \hat{\theta}_0^0 + u_{\mathbf{x}0}^0 h \hat{\theta}_{\mathbf{x}0}^0 + \sum_{r=0}^k |D_2^r| h f_2^r$$

Il existe  $\xi_{\tilde{e}}, \xi_{e_0}, \xi_{e_1}, \dots, \xi_{e_k}$ , avec  $\tau_l > \tilde{e}, \tau_l > e_0, \dots, \tau_1 > e_k$  tels que  $\xi_{\tilde{e}} = |u_{\mathbf{x}0}^0| h \hat{\theta}_{\mathbf{x}0}^0$ ,  $\xi_{e_0} = |u_0^r| h \hat{\theta}_0^0, \dots, \xi_{e_k} = |u_0^r| h \hat{\theta}_0^k$ .

On définit maintenant  $m_{d,e} = 1$  pour  $e \in \{\tilde{e}, e_0, \dots, e_k\}$  et  $m_{d,e} = 0$  sinon. Maintenant

$$\left| \sum_{r=0}^k h |D_2^r| f_2^r \right| \leq \sum_{r=0}^k h h^2 |f_2^r| \leq \frac{1}{\hat{\sigma}} \sum_{r=0}^k h h^2 |f_2^r| \text{ car } 0 < \hat{\sigma} \leq 1$$

et on définit  $\gamma_d = \frac{1}{\hat{\sigma}} \sum_{r=0}^k h h^2 |f_2^r|$ . On a ainsi :

$$\xi_d \leq \sum_{e=1}^m m_{d,e} \xi_e + \gamma_d$$

Les points entrants correspondants à  $\xi_{\tilde{e}}, \xi_{e_0}, \dots, \xi_{e_k}$  de la cellule  $C$  n'interviennent que dans  $C$  même (dans le cadre du calcul de  $\xi_d$ ). Ainsi, pour  $e \in \{\tilde{e}, e_0, e_1, \dots, e_k\}$ , on a donc  $m_{\mu,e} = 1$  pour  $\mu = d$ , et  $m_{\mu,e} = 0$  pour  $e \in \{1, \dots, \tau_1\}$ . Conclusion :

$$\sum_{\mu=1}^{\tau_1} m_{\mu,e} = 1 \text{ pour } e \in \{\tilde{e}, e_0, e_1, \dots, e_k\}$$

**Sans point intérieur de type 1 – 1 :** Cette cellule possède deux points sortants

$Q_{\tilde{j}} = Q_2^k$  et  $Q_{\tilde{j}} = Q_1^0$ , qui correspondent respectivement à  $\xi_{\tilde{d}} = |u_{\tilde{j}}| h \hat{\theta}_{\tilde{j}}^k = |u_2^k| h \hat{\theta}_2^k$  et  $\xi_{\tilde{d}} = |u_{\tilde{j}}| h \hat{\theta}_{\tilde{j}}^0 = |u_1^0| \hat{\theta}_1^0$ . Il existe  $\xi_{e_0}, \xi_{e_1}, \dots, \xi_{e_k}$ , avec  $\tau_l > e_0, \dots, \tau_1 > e_k$  tels que

50. Attention  $\xi_e$  pourrait être une valeur du bord, c'est-à-dire  $m \geq e > \tau_1$ , et on aurait aussi  $m_{e,e} = 1$

$\xi_{e_0} = |u_0^r| h\hat{\theta}_0^0, \dots, \dots, \xi_{e_r} = |u_0^r| h\hat{\theta}_0^r, \dots, \xi_{e_k} = |u_0^k| h\hat{\theta}_0^k$ . Avec (16.263) on obtient, si on regroupe les lignes  $\check{d}$  et  $\tilde{d}$ , l'inégalité suivante :

$$\begin{pmatrix} \xi_{\check{d}} \\ \xi_{\tilde{d}} \end{pmatrix} \leq \begin{pmatrix} \frac{\hat{\theta}_1^0}{\hat{\theta}_0^0} \omega^0 \prod_{l=1}^k \lambda_{1,2}^l & \frac{\hat{\theta}_2^k}{\hat{\theta}_1^0} \omega^1 \prod_{l=2}^k \lambda_{1,2}^l & \dots & \frac{\hat{\theta}_2^k}{\hat{\theta}_0^r} \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l & \dots & \frac{\hat{\theta}_2^k}{\hat{\theta}_0^k} \omega^k \\ \frac{\hat{\theta}_2^k}{\hat{\theta}_0^0} & 0 & \dots & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} \xi_{e_0} \\ \xi_{e_1} \\ \vdots \\ \xi_{e_r} \\ \vdots \\ \xi_{e_k} \end{pmatrix} + \begin{pmatrix} h\hat{\theta}_2^k \frac{h}{\hat{\sigma} \|\beta\|} \sum_{r=0}^k |f_2^r| \\ h\hat{\theta}_1^0 \frac{h}{\hat{\sigma} \|\beta\|} |f_1^0| \end{pmatrix}$$

On définit maintenant les lignes  $\check{d}$  et  $\tilde{d}$  de la matrice  $M$  par

$$m_{\check{d},e} = \begin{cases} \frac{\hat{\theta}_2^k}{\hat{\theta}_0^r} \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l & \text{pour } e \in \{e_0, \dots, e_k\} \\ 0 & \text{sinon} \end{cases}, \quad m_{\tilde{d},e} = \begin{cases} \frac{\hat{\theta}_2^k}{\hat{\theta}_0^0} & \text{pour } e = e_0, \\ 0 & \text{sinon} \end{cases} \quad (17.313)$$

Les points entrants correspondants à  $\xi_{\check{e}}, \xi_{e_0}, \dots, \xi_{e_k}$  de la cellule  $C$  n'interviennent que dans celle-ci pour calculer  $\xi_{\check{d}}$  et  $\xi_{\tilde{d}}$ . Pour  $e \in \{e_0, e_1, \dots, e_k\}$ , on a donc  $m_{\mu,e} = 0$  si  $\mu \neq \check{d}$  ou  $\mu \neq \tilde{d}$ . Avec les égalités (16.267) et (16.268) on a donc :

$$\sum_{\mu=1}^{\tau_1} m_{\mu,e} = 1 \text{ pour } e \in \{e_0, e_1, \dots, e_k\}$$

Enfin, avec  $\hat{\theta}_1^0 \leq h\|\beta\|$  et  $\hat{\theta}_2^k \leq h\|\beta\|$  :

$$h\hat{\theta}_2^k \frac{h}{\hat{\sigma} \|\beta\|} \sum_{r=0}^k |f_2^r| \leq h^3 \frac{1}{\hat{\sigma}} \sum_{r=0}^k |f_2^r| =: \gamma_{\check{d}} \text{ et } h\hat{\theta}_1^0 \frac{h}{\hat{\sigma} \|\beta\|} |f_1^0| \leq h^3 \frac{1}{\hat{\sigma}} |f_1^0| =: \gamma_{\tilde{d}}$$

On a ainsi :

$$\xi_d \leq \sum_{e=1}^m m_{d,e} \xi_e + \gamma_d \text{ pour } d \in \{\check{d}, \tilde{d}\}$$

**Avec un point intérieur de type 1 – 1 :** Cette cellule possède deux points sortants  $Q_{\check{j}} = Q_2^k$  et  $Q_{\tilde{j}} = Q_2^{-t}$ , qui correspondent respectivement à  $\xi_{\check{d}} = |u_{\check{j}}| h\hat{\theta}_{\check{j}}^k = |u_2^k| h\hat{\theta}_2^k$  et  $\xi_{\tilde{d}} = |u_{\tilde{j}}| h\hat{\theta}_{\tilde{j}}^k = |u_2^{-t}| \hat{\theta}_2^{-t}$ . Il existe  $\xi_{e_k}, \dots, \xi_{e_0}, \dots, \xi_{e_{-t}}$ , avec  $\tau_1 > e_k, \dots, \tau_1 > e_{-t}$  tels que  $\xi_{e_k} = |u_0^k| h\hat{\theta}_0^k, \dots, \dots, \xi_{e_r} = |u_0^r| h\hat{\theta}_0^r, \dots, \xi_{e_{-t}} = |u_0^{-t}| h\hat{\theta}_0^{-t}$ .

Un raisonnement parfaitement identique au cas ci-dessus permet d'obtenir :

a)  $m_{\check{d},e} \geq 0$  et  $m_{\tilde{d},e} \geq 0$  pour  $e \in \{1, \dots, m\}$ ,

b)

$$\sum_{\mu=1}^{\tau_1} m_{\mu,e} = 1 \text{ pour } e \in \{e_k, \dots, e_0, \dots, e_{-t}\}$$

c)

$$\xi_d \leq \sum_{e=1}^m m_{d,e} \xi_e + \gamma_d \text{ pour } d \in \{\check{d}, \tilde{d}\}$$

$$\text{où } \gamma_{\check{d}} = h^3 \frac{1}{\hat{\sigma}} \sum_{r=0}^k |f_2^r| + h^3 \frac{1}{\hat{\sigma}} |f_1^0| \text{ et } \gamma_{\tilde{d}} = h^3 \frac{1}{\hat{\sigma}} \sum_{r=1}^t |f_2^{-r}| + h^3 \frac{1}{\hat{\sigma}} |f_1^0|$$

**Conclusion :** Quelque soit le type de cellule  $C \in \mathcal{C}_h$ , on déduit pour  $j \in J_{\text{sortant}}^C$ , et  $\xi_d = |u_h(Q_j)| h \hat{\theta}_j$  :

i)  $m_{d,e} \geq 0$  pour  $e \in \{1, \dots, m\}$ ,

ii)  $\sum_{e=1}^{\tau_1} m_{d,e} = 1$ ,

ii)  $\gamma_d \leq h^3 \frac{1}{\hat{\sigma}} \sum_{r \in J^C \setminus J_{\text{entrant}}^C} |f_r|$

En particulier i) et ii) signifient que  $\|A_{\tau_1 \times \tau_1}\|_1 \leq 1$  et  $\|B_{\tau_1 \times (m-\tau_1)}\|_1 \leq 1$ .

$$\begin{aligned} \xi &\leq M\xi + \gamma \text{ donc} \\ \underline{0} &\leq -\xi + M\xi + \gamma \text{ la matrice est positive, donc} \\ \underline{0} = M\underline{0} &\leq -M\xi + M^2\xi + M\gamma \text{ donc} \\ M\xi &\leq M^2\xi + M\gamma \text{ Maintenant, on ajoute } \gamma \text{ de chaque côté, donc} \\ \xi &\leq M^2\xi + M\gamma + \gamma \text{ En procédant par itération, on obtient} \\ \xi &\leq M^N \xi + \sum_{s=0}^{N-1} M^s \gamma \text{ Donc :} \\ \xi &\leq M^N \xi + \sum_{s=0}^{N-1} \left( \frac{A_{\tau_1 \times \tau_1}^s}{0_{(m-\tau_1) \times \tau_1}} \mid \frac{\sum_{\mu=0}^{\mu-1} A_{\tau_1 \times (m-\tau_1)}^s}{Id_{(m-\tau_1) \times (m-\tau_1)}} \right) \begin{pmatrix} c \\ 0_{m-\tau_1} \end{pmatrix} \text{ avec} \\ M^N \xi &= \left( \frac{0_{\tau_1 \times \tau_1}}{0_{(m-\tau_1) \times \tau_1}} \mid \frac{\sum_{\mu=0}^{N-1} A_{\tau_1 \times \tau_1}^\mu B_{\tau_1 \times (m-\tau_1)}}{Id_{(m-\tau_1) \times (m-\tau_1)}} \right) \xi \end{aligned} \quad (17.314)$$

Les coordonnées  $\xi_{\tau_1+1}$  à  $\xi_m$  correspondent aux données du bord entrant et sont connues. Posons  $\xi_{\text{entrant}} = (\xi_{\tau_1+1}, \dots, \xi_m)^{tr}$ , alors :

$$\begin{aligned} \|\xi\|_1 &\leq \left\| \sum_{\mu=0}^{N-1} A_{\tau_1 \times \tau_1}^\mu B_{\tau_1 \times (m-\tau_1)} \xi_{\text{entrant}} \right\|_1 + \|Id_{(m-\tau_1) \times (m-\tau_1)} \xi_{\text{entrant}}\|_1 + \left\| \sum_{s=0}^{N-1} A_{\tau_1 \times \tau_1}^s c \right\|_1 \\ \|\xi\|_1 &\leq N \|\xi_{\text{entrant}}\|_1 + \|\xi_{\text{entrant}}\|_1 + N \|c\|_1 \\ &\leq (K_{\Xi} \frac{1}{h} h + h) \sum_{j \in J^-} |g_h(Q_j)| \hat{\theta}_j + K_{\Xi} \frac{1}{h} h \sum_{j \in J \setminus J^-} \frac{1}{\hat{\sigma}} h^2 |f_j| \\ &\quad \text{Avec } \frac{2}{3} \frac{h^2}{(2\sigma c_0)^2} \leq |D_j| \text{ et } h < 1 \\ \|\xi\|_1 &\leq (K_{\Xi} + 1) \sum_{j \in J^-} |g_h(Q_j)| \int_{S_j} |\beta \cdot n| d\sigma + 6K_{\Xi} \sigma^2 c_0^2 \frac{1}{\hat{\sigma}} \sum_{j \in J \setminus J^-} |D_j| |f_j| \\ &\leq \|\beta\| (K_{\Xi} + 1) \sum_{j \in J^-} |g_h(Q_j)| |S_j| + 6K_{\Xi} \sigma^2 c_0^2 \frac{1}{\hat{\sigma}} \sum_{j \in J \setminus J^-} |D_j| |f_j| \\ &\leq \|\beta\| (K_{\Xi} + 1) \|g_h\|_{L_h^1(\Gamma^-)} + 6K_{\Xi} \sigma^2 c_0^2 \frac{1}{\hat{\sigma}} \|f_h\|_{L_h^1(\Omega)} \end{aligned}$$

### Étape 2 : Estimation sur les points intérieurs et sortants des cellules

De nouveau trois types de cellules  $C \in \mathcal{C}_h$  sont à étudier ( $N_C$  est le nombre de triangles dans  $C$ ) :

**Avec deux coordonnées locales 0 :** Avec (16.252) et (16.249) on obtient pour les points intérieurs et le point sortant :

$$\begin{aligned}
\sum_{j \in J_{\text{intérieur}}^C \cup J_{\text{sortant}}^C} |u_h(Q_j)| &= \sum_{r=0}^{k-1} |u_2^r| + |u_2^k| \\
&\leq \sum_{r=0}^{k-1} |u_0^{r+1}| \lambda_{0,1}^{r+1} + \sum_{r=0}^{k-1} \sum_{s=1}^r |u_0^s| \omega^s \prod_{l=s+1}^r \lambda_{1,2}^l + \\
&\quad \sum_{r=0}^{k-1} (|u_0^0| \lambda_{0,2}^0 + |u_{\star 0}^0| \lambda_{1,2}^0) \prod_{l=1}^r \lambda_{1,2}^l + \frac{h}{\|\beta\|} \sum_{r=0}^{k-1} \sum_{s=0}^r |f_2^s| \\
&\quad + \sum_{r=1}^k |u_0^r| \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l + (|u_0^0| \lambda_{0,2}^0 + |u_{\star 0}^0| \lambda_{1,2}^0) \prod_{l=1}^k \lambda_{1,2}^l \\
&\quad + \frac{h}{\hat{\sigma} \|\beta\|} \sum_{r=0}^k |f_2^r| \\
&\leq \sum_{r=0}^{k-1} |u_0^{r+1}| + \sum_{r=0}^{k-1} \sum_{s=1}^r |u_0^s| + \sum_{r=0}^{k-1} (|u_0^0| + |u_{\star 0}^0|) + \frac{h}{\hat{\sigma} \|\beta\|} \sum_{r=0}^{k-1} \sum_{s=0}^r |f_2^s| \\
\sum_{j \in J_{\text{intérieur}}^C \cup J_{\text{sortant}}^C} |u_h(Q_j)| &\leq N_C \sum_{j \in J_{\text{entrant}}^C} |u_h(Q_j)| + \frac{h}{\hat{\sigma} \|\beta\|} N_C \sum_{j \in J_{\text{intérieur}}^C \cup J_{\text{sortant}}^C} |f_h(Q_j)| \\
&\quad \text{avec } N_C \leq K_{\text{cell}} \leq 2 \frac{\ln(4\sigma c_0)}{\ln 2}
\end{aligned}$$

**Sans point intérieur de type 1 – 1 et avec intérieur de type 1 – 1 :** Avec (16.266) et (16.288) à (16.290) on obtient l'estimation

$$\sum_{j \in J_{\text{intérieur}}^C \cup J_{\text{sortant}}^C} |u_h(Q_j)| \leq (N_C + 1) \sum_{j \in J_{\text{entrant}}^C} |u_h(Q_j)| + \frac{h}{\hat{\sigma} \|\beta\|} (N_C + 1) \sum_{j \in J_{\text{intérieur}}^C \cup J_{\text{sortant}}^C} |f_h(Q_j)|$$

Ainsi, en sommant sur toutes les cellules et en utilisant la majoration  $N_C \leq K_{\text{cell}}$  on a :

$$\begin{aligned}
\sum_{C \in \mathcal{C}_h} \sum_{j \in J_{\text{intérieur}}^C \cup J_{\text{sortant}}^C} |u_h(Q_j)| &\leq (K_{\text{cell}} + 1) \sum_{C \in \mathcal{C}_h} \sum_{j \in J_{\text{entrant}}^C} |u_h(Q_j)| + \\
&\quad \frac{h}{\hat{\sigma} \|\beta\|} (K_{\text{cell}} + 1) \sum_{C \in \mathcal{C}_h} \sum_{j \in J_{\text{intérieur}}^C \cup J_{\text{sortant}}^C} |f_h(Q_j)|
\end{aligned}$$

En multipliant cette inégalité par  $h^2\|\beta\|$  et en appliquant le lemme 16.1 on obtient

$$\begin{aligned}
\sum_{C \in \mathcal{C}_h} \sum_{j \in J_{\text{intérieur}}^C \cup J_{\text{sortant}}^C} |u_h(Q_j)| & \leq h^2 \|\beta\| \\
& \leq (K_{\text{cell}} + 1) \frac{1}{c_{\hat{\theta}}} \sum_{C \in \mathcal{C}_h} \sum_{j \in J_{\text{entrant}}^C} |u_h(Q_j)| h \hat{\theta}_j \\
& \quad + h^3 \frac{1}{\hat{\sigma}} (K_{\text{cell}} + 1) \sum_{C \in \mathcal{C}_h} \sum_{j \in J_{\text{intérieur}}^C \cup J_{\text{sortant}}^C} |f_h(Q_j)| \\
& \leq (K_{\text{cell}} + 1) \frac{1}{c_{\hat{\theta}}} \|\xi\|_1 + h^3 \frac{1}{\hat{\sigma}} (K_{\text{cell}} + 1) \sum_{C \in \mathcal{C}_h} \sum_{j \in J_{\text{intérieur}}^C \cup J_{\text{sortant}}^C} |f_h(Q_j)| \\
& \leq (K_{\text{cell}} + 1) \frac{1}{c_{\hat{\theta}}} \|\xi\|_1 + \frac{6h\sigma^2 c_0^2}{\hat{\sigma}} (K_{\text{cell}} + 1) \|f_h\|_{L_h^1(\Omega)}
\end{aligned}$$

Maintenant

$$\begin{aligned}
\|u_h\|_{L_h^1} & = \sum_{C \in \mathcal{C}_h} \sum_{j \in J_{\text{entrant}}^C} |u_h(Q_j)| |D_j| + \sum_{C \in \mathcal{C}_h} \sum_{j \in J_{\text{intérieur}}^C \cup J_{\text{sortant}}^C} |u_h(Q_j)| |D_j| \\
& \leq \frac{1}{\|\beta\|} \sum_{C \in \mathcal{C}_h} \sum_{j \in J_{\text{entrant}}^C} |u_h(Q_j)| \|\beta\| h^2 + \frac{1}{\|\beta\|} \sum_{C \in \mathcal{C}_h} \sum_{j \in J_{\text{intérieur}}^C \cup J_{\text{sortant}}^C} |u_h(Q_j)| \|\beta\| h^2 \\
& \leq \frac{1}{\|\beta\|} \frac{1}{c_{\hat{\theta}}} \sum_{C \in \mathcal{C}_h} \sum_{j \in J_{\text{entrant}}^C} |u_h(Q_j)| h \hat{\theta}_j + \frac{1}{\|\beta\|} \sum_{C \in \mathcal{C}_h} \sum_{j \in J_{\text{intérieur}}^C \cup J_{\text{sortant}}^C} |u_h(Q_j)| \|\beta\| h^2 \\
& \leq \frac{1}{\|\beta\|} \frac{1}{c_{\hat{\theta}}} \|\xi\|_1 + \frac{1}{\|\beta\|} (K_{\text{cell}} + 1) \frac{1}{c_{\hat{\theta}}} \|\xi\|_1 + \frac{1}{\|\beta\|} \frac{6h\sigma^2 c_0^2}{\hat{\sigma}} (K_{\text{cell}} + 1) \|f_h\|_{L_h^1(\Omega)} \\
& \leq \frac{1}{\|\beta\|} (K_{\text{cell}} + 2) \frac{1}{c_{\hat{\theta}}} \left[ \|\beta\| (K_{\Xi} + 1) \|g_h\|_{L_h^1(\Gamma^-)} + 6K_{\Xi} \sigma^2 c_0^2 \frac{1}{\hat{\sigma}} \|f_h\|_{L_h^1(\Omega)} \right] \\
& \quad + \frac{1}{\|\beta\|} \frac{6h\sigma^2 c_0^2}{\hat{\sigma}} (K_{\text{cell}} + 1) \|f_h\|_{L_h^1(\Omega)} \\
& \leq (K_{\text{cell}} + 2) \frac{1}{c_{\hat{\theta}}} (K_{\Xi} + 1) \|g_h\|_{L_h^1(\Gamma^-)} \\
& \quad + \frac{1}{\|\beta\|} (K_{\text{cell}} + 2) \frac{1}{c_{\hat{\theta}}} 6K_{\Xi} \sigma^2 c_0^2 \frac{1}{\hat{\sigma}} \|f_h\|_{L_h^1(\Omega)} + \frac{1}{\|\beta\|} \frac{6h\sigma^2 c_0^2}{\hat{\sigma}} (K_{\text{cell}} + 1) \|f_h\|_{L_h^1(\Omega)} \\
& \leq (K_{\text{cell}} + 2) \frac{1}{c_{\hat{\theta}}} (K_{\Xi} + 1) \|g_h\|_{L_h^1(\Gamma^-)} + \frac{1}{\|\beta\|} (K_{\text{cell}} + 2) 6\sigma^2 c_0^2 \frac{1}{\hat{\sigma}} \left[ \frac{1}{c_{\hat{\theta}}} K_{\Xi} + 1 \right] \|f_h\|_{L_h^1(\Omega)}
\end{aligned}$$

□

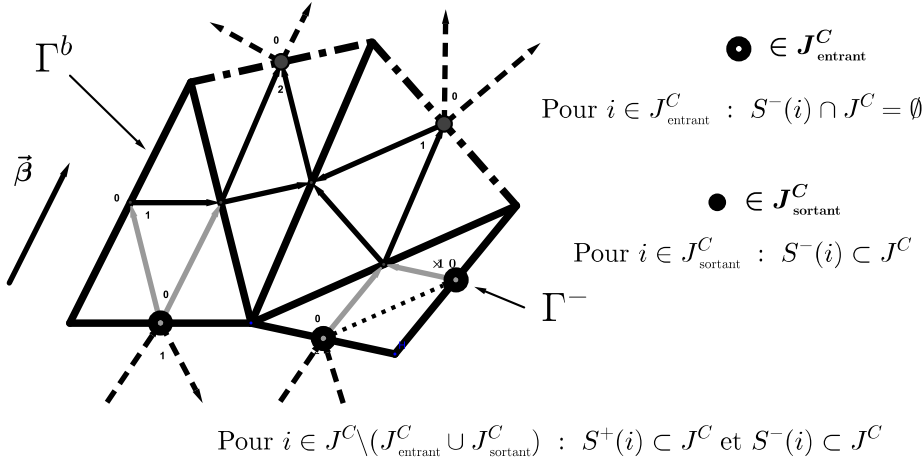
**Remarque 17.1** La preuve du théorème suivant diffère dans sa structure de celle du théorème précédent, car elle procède par une estimation ad hoc et ne nécessite pas l'utilisation d'une matrice. L'inégalité de récurrence qui y est contenue, est adaptable pour avoir une autre preuve du théorème précédent. Un exemple en fin de chapitre montre comment.

Avant de procéder au théorème suivant, un lemme important :

**Lemme 17.3 (Estimation clé pour  $\|u_h\|_{L_h^2(\Omega)}$ )** . Soit  $(\mathcal{T}_h)$  une famille de maillage et  $0 \neq \beta \in \mathbb{R}^2$ , soit  $\mathcal{G}_h$  le graphe associé à  $\mathcal{T}_h$ . Soit  $E$  une réunion connexe de triangles de  $\mathcal{T}_h$  telle que le bord topologique de  $E$  ne contienne que des points de type 0 – 1 ou 0 – 2.<sup>51</sup> Si  $u_h$  est une solution de  $\hat{b}_h(u_h, v_h) = (f_h, v_h)$  pour  $v_h \in Y_0^h$  quelconque, alors

$$\sum_{i \in J_{\text{sortant}}^E} u_i^2 \hat{\theta}_i \leq \sum_{i \in J_{\text{entrant}}^E} u_i^2 \hat{\theta}_i + \sum_{i \in J^E \setminus J_{\text{entrant}}^E} u_i^2 |D_i| + \sum_{i \in J^E \setminus J_{\text{entrant}}^E} f_i^2 |D_i| \quad (17.315)$$

Le cadre de ce lemme est donc plus large que celui des trois types de cellules autorisés par la famille de maillage dans le théorème de stabilité énoncé ci-dessous. La situation ci-dessous permet de mieux comprendre la preuve :



**Preuve :** on définit la fonction test  $v_h(Q_i) = u_h(Q_i)$  si  $i \in J^E \setminus J_{\text{entrant}}^E$  et  $v_h(Q_i) = 0$  sinon. Alors  $v_h \in Y_0^h$  et  $\hat{b}_h(u_h, v_h) = (f_h, v_h)$ . Ainsi

$$\begin{aligned} & \sum_{i \in J^E \setminus J_{\text{entrant}}^E} u_i \left[ \sum_{j \in S^+(i)} \theta_{i,j} u_j + \sum_{j \in S^-(i)} \theta_{i,j} u_j \right] = \sum_{k \in J^E \setminus J_{\text{entrant}}^E} f_h(Q_k) u_h(Q_k) |D_k| \\ & \sum_{i \in J^E \setminus J_{\text{entrant}}^E} \sum_{j \in S^+(i)} u_i^2 \theta_{i,j} = \sum_{i \in J^E \setminus J_{\text{entrant}}^E} \sum_{j \in S^-(i)} \theta_{j,i} u_i u_j + \sum_{k \in J^E \setminus J_{\text{entrant}}^E} f_h(Q_k) u_h(Q_k) |D_k| \\ & \leq \frac{1}{2} \sum_{i \in J^E \setminus J_{\text{entrant}}^E} \sum_{j \in S^-(i)} \theta_{j,i} u_i^2 + \frac{1}{2} \sum_{i \in J^E \setminus J_{\text{entrant}}^E} \sum_{j \in S^-(i)} \theta_{j,i} u_j^2 + \frac{1}{2} \sum_{k \in J^E \setminus J_{\text{entrant}}^E} \frac{1}{2} f_k^2 |D_k| + \\ & \frac{1}{2} \sum_{k \in J^E \setminus J_{\text{entrant}}^E} \frac{1}{2} u_k^2 |D_k| \end{aligned}$$

On considère maintenant les différentes sommes ci-dessus :

51. Le cas d'un triangle  $T$  de  $E$  contenant deux points de type 0, c'est-à-dire contenant un point sur le bord entrant  $\Gamma^-$  vers lequel l'arc orienté  $(Q_0^T Q_1^T)$  a été retiré, est donc possible.

**Membre de gauche :**

$$\begin{aligned} & \sum_{i \in J^E \setminus J^E_{\text{entrant}}} \sum_{j \in S^+(i)} u_i^2 \theta_{i,j} \\ = & \sum_{i \in J^E \setminus J^E_{\text{entrant}} \setminus J^E_{\text{sortant}}} \sum_{j \in S^+(i)} u_i^2 \theta_{i,j} + \sum_{i \in J^E_{\text{sortant}}} \sum_{j \in S^+(i) \cap J^E} u_i^2 \theta_{i,j} + \sum_{i \in J^E_{\text{sortant}}} \sum_{j \in S^+(i) \setminus J^E} u_i^2 \theta_{i,j} \end{aligned}$$

Comme les points sortants sont de type 0 – 1 ou 0 – 2, pour  $i \in J^E_{\text{sortant}}$  on a  $\sum_{j \in S^+(i) \setminus J^E} \theta_{i,j} = \int_{S_i} \beta \cdot n_E d\sigma = \hat{\theta}_i$  et  $\sum_{i \in J^E_{\text{sortant}}} \sum_{j \in S^+(i) \setminus J^E} u_i^2 \theta_{i,j} = \sum_{i \in J^E_{\text{sortant}}} u_i^2 \hat{\theta}_i$ .

Comme les points entrants sont de type 0 – 1 ou 0 – 2, pour  $i \in J^E_{\text{entrant}}$  on a  $\sum_{j \in S^+(i) \cap J^E} \theta_{i,j} = \hat{\theta}_i$  et  $\sum_{i \in J^E_{\text{entrant}}} \sum_{j \in S^+(i) \cap J^E} u_i^2 \theta_{i,j} = \sum_{i \in J^E_{\text{entrant}}} u_i^2 \hat{\theta}_i$ .

Ainsi

$$\begin{aligned} & \sum_{i \in J^E \setminus J^E_{\text{entrant}}} \sum_{j \in S^+(i)} u_i^2 \theta_{i,j} = \sum_{i \in J^E \setminus J^E_{\text{sortant}}} \sum_{j \in S^+(i) \cap J^E} u_i^2 \theta_{i,j} - \sum_{i \in J^E_{\text{entrant}} \cap J^E} \sum_{j \in S^+(i)} u_i^2 \theta_{i,j} \\ & \quad + \sum_{i \in J^E_{\text{sortant}}} \sum_{j \in S^+(i) \cap J^E} u_i^2 \theta_{i,j} + \sum_{i \in J^E_{\text{sortant}}} u_i^2 \hat{\theta}_i \\ = & \sum_{i \in J^E \setminus J^E_{\text{sortant}}} \sum_{j \in S^+(i) \cap J^E} u_i^2 \theta_{i,j} - \sum_{i \in J^E_{\text{entrant}}} u_i^2 \hat{\theta}_i + \sum_{i \in J^E_{\text{sortant}}} \sum_{j \in S^+(i) \cap J^E} u_i^2 \theta_{i,j} + \sum_{i \in J^E_{\text{sortant}}} u_i^2 \hat{\theta}_i \\ & \sum_{i \in J^E \setminus J^E_{\text{entrant}}} \sum_{j \in S^+(i)} u_i^2 \theta_{i,j} = \sum_{i \in J^E} \sum_{j \in S^+(i) \cap J^E} u_i^2 \theta_{i,j} - \sum_{i \in J^E_{\text{entrant}}} u_i^2 \hat{\theta}_i + \sum_{i \in J^E_{\text{sortant}}} u_i^2 \hat{\theta}_i \end{aligned}$$

**Premier membre de droite :** On a  $\sum_{j \in S^-(i)} \theta_{j,i} = \bar{\theta}_i = \sum_{j \in S^+(i)} \theta_{i,j}$ , donc

$$\begin{aligned} & \frac{1}{2} \sum_{i \in J^E \setminus J^E_{\text{entrant}}} \sum_{j \in S^-(i)} \theta_{j,i} u_i^2 = \frac{1}{2} \sum_{i \in J^E \setminus J^E_{\text{entrant}}} \sum_{j \in S^+(i)} \theta_{i,j} u_i^2 \\ = & \frac{1}{2} \sum_{i \in J^E \setminus J^E_{\text{entrant}} \setminus J^E_{\text{sortant}}} \sum_{j \in S^+(i)} \theta_{i,j} u_i^2 + \frac{1}{2} \sum_{i \in J^E_{\text{sortant}}} \sum_{j \in S^+(i) \cap J^E} \theta_{i,j} u_i^2 + \frac{1}{2} \sum_{i \in J^E_{\text{sortant}}} \sum_{j \in S^+(i) \setminus J^E} \theta_{i,j} u_i^2 \end{aligned}$$

Pour  $i \in J^E \setminus J^E_{\text{entrant}} \setminus J^E_{\text{sortant}}$ ,  $S^+(i) \subset J^E$ , donc

$$= \frac{1}{2} \sum_{i \in J^E \setminus J^E_{\text{entrant}}} \sum_{j \in S^+(i) \cap J^E} \theta_{i,j} u_i^2 + \frac{1}{2} \sum_{i \in J^E_{\text{sortant}}} \sum_{j \in S^+(i) \setminus J^E} \theta_{i,j} u_i^2$$

Les points entrants sont de type 0 – 1 ou 0 – 2, donc  $\sum_{i \in J^E_{\text{entrant}}} \sum_{j \in S^+(i) \cap J^E} u_i^2 \theta_{i,j} = \sum_{i \in J^E_{\text{entrant}}} u_i^2 \hat{\theta}_i$ . Les points sortants sont de type 0 – 1 ou 0 – 2, on a ainsi

$$\frac{1}{2} \sum_{i \in J^E \setminus J^E_{\text{entrant}}} \sum_{j \in S^-(i)} \theta_{j,i} u_i^2 = -\frac{1}{2} \sum_{i \in J^E_{\text{entrant}}} u_i^2 \hat{\theta}_i + \frac{1}{2} \sum_{i \in J^E} \sum_{j \in S^+(i) \cap J^E} \theta_{i,j} u_i^2 + \frac{1}{2} \sum_{i \in J^E_{\text{sortant}}} \hat{\theta}_i u_i^2$$

**Deuxième membre de droite :** Les points sortants sont de type 0 – 1 ou 0 – 2, donc

$$\frac{1}{2} \sum_{i \in J^E \setminus J^E_{\text{entrant}}} \sum_{j \in S^-(i)} \theta_{j,i} u_j^2 = \frac{1}{2} \sum_{k \in J^E} \sum_{l \in S^+(i) \cap J^E} \theta_{k,l} u_k^2$$



**Conclusion** : On remplace les différentes égalités et on simplifie  $\sum_{k \in J^E} \sum_{l \in S^+(i) \cap J^E} \theta_{k,l} u_k^2$  :

$$\begin{aligned} & - \sum_{i \in J_{\text{entrant}}^E} u_i^2 \hat{\theta}_i + \sum_{i \in J_{\text{sortant}}^E} u_i^2 \hat{\theta}_i \\ & \leq -\frac{1}{2} \sum_{i \in J_{\text{entrant}}^E} u_i^2 \hat{\theta}_i + \frac{1}{2} \sum_{i \in J_{\text{sortant}}^E} \hat{\theta}_i u_i^2 + \frac{1}{2} \sum_{k \in J^E \setminus J_{\text{entrant}}^E} \frac{1}{2} f_k^2 |D_k| + \frac{1}{2} \sum_{k \in J^E \setminus J_{\text{entrant}}^E} \frac{1}{2} u_k^2 |D_k| \end{aligned}$$

□

**Théorème 17.4** ( $\|u_h\|_{L_h^2(\Omega)}$ -stabilité) Soit  $(\mathcal{T}_h)$  une famille de maillage quasi-uniforme et régulière, soit  $\mathcal{G}_h$  le graphe associé à  $\mathcal{T}_h$ . Soit  $\mathcal{C}_h$  l'ensemble des cellules obtenues à partir de  $\mathcal{G}_h$ . On suppose que :

- $\mathcal{C}_h$  contient pas de point intérieur de type 0 – 1 ou 0 – 2 (donc  $\mathcal{G}_h$  ne contient pas de circuit élémentaire).
- Il existe  $K_{\Xi} > 0$ , ne dépendant pas de  $h$ , tel que pour le nombre de strates  $N_{\Xi}$  on ait  $N = N_{\Xi} \leq K_{\Xi} \frac{1}{h}$ ,

Alors pour la solution  $u_h$  du problème variationnel  $I$  (voir page 95), il existe des constantes  $C > 0$  et  $h_0 > 0$ , qui ne dépendent pas de  $h$  et  $\|\beta\|$ , telles que pour  $h \in ]0; h_0[$  on ait :

$$\|u_h\|_{L_h^2(\Omega)} \leq C e^{\frac{C}{\|\beta\|}} \left[ \left( 1 + \frac{h_0}{\|\beta\|} \right) \|f_h\|_{L_h^2(\Omega)} + (\|\beta\| + h_0)^{1/2} \|g_h\|_{L_h^2(\Gamma^-)} \right] \quad (17.316)$$

**Preuve du théorème de  $L_h^2$ -stabilité** Les étapes sont les mêmes que dans la preuve de l'exemple vu en début de chapitre.

**Étape préparatoire : Notations** Trois types de cellules sont possibles et les notations utilisées pour  $\hat{\theta}_i$  sont celles de la définition 16.1.

**Calcul de la cellule  $C$**  : Voir le chapitre 16.

**Estimation clé pour  $\|u_h\|_{L_h^2(\Omega)}$**  : Soit  $C \in \mathcal{C}_h$ , on commence par le résultat du lemme 17.3

$$\sum_{i \in J_{\text{sortant}}^C} u_i^2 \hat{\theta}_i \leq \sum_{i \in J_{\text{entrant}}^C} u_i^2 \hat{\theta}_i + \sum_{i \in J^C \setminus J_{\text{entrant}}^C} u_i^2 |D_i| + \sum_{i \in J^C \setminus J_{\text{entrant}}^C} f_i^2 |D_i|$$

Notre objectif étant une majoration du membre gauche pour les points d'entrés, il faut aussi majorer le deuxième terme du membre de droite par les points d'entrés.

Pour les trois types de cellules (voir le chapitre 16), on a pour  $i \in J^C \setminus J_{\text{entrant}}^C$  l'inégalité

$$|u_i| \leq \sum_{k \in J_{\text{entrant}}^C} |u_k| s_k + \sum_{l \in J^C \setminus J_{\text{entrant}}^C} \frac{h}{\hat{\sigma} \|\beta\|} |f_l| t_l$$

avec  $s_k \in [0; 1]$  et  $t_l \in [0; 1]$  – des valeurs connues qui dépendent de  $C$  – pour  $k \in J_{\text{entrant}}^C$  et  $l \in J^C \setminus J_{\text{entrant}}^C$ .

On a donc  $|u_i| \leq \sum_{i \in J^C} a_i$  pour certaines quantités  $a_i$ ,  $i \in J^C$ . On a maintenant l'inégalité

$$\left( \sum_{i \in J^C} a_i \right)^2 = \left( \sum_{i \in J^C} a_i \right) \left( \sum_{j \in J^C} a_j \right) \leq \sum_{i \in J^C} \sum_{j \in J^C} \left( \frac{1}{2} a_i^2 + \frac{1}{2} a_j^2 \right) \leq |J^C| \sum_{i \in J^C} a_i^2$$

$N_C$  désigne le nombre maximal de triangles dans  $C$ , et le lemme 14.3 montre que  $N_C \leq K_{cell}$  où  $K_{cell}$  est une constante qui ne dépend pas de  $h$ . Ainsi avec  $|J^C| \leq 3K_{cell}$  et le lemme 16.1<sup>52</sup>

$$\begin{aligned} \sum_{i \in J^C \setminus J_{\text{entrant}}^C} u_i^2 |D_i| &\leq \sum_{i \in J^C \setminus J_{\text{entrant}}^C} 3K_{cell} \left[ \sum_{k \in J_{\text{entrant}}^C} u_k^2 + \sum_{l \in J^C \setminus J_{\text{entrant}}^C} \frac{h^2}{\hat{\sigma}^2 \|\beta\|^2} f_l^2 \right] h^2 \\ &\leq 6K_{cell}^2 \sum_{k \in J_{\text{entrant}}^C} u_k^2 h \frac{\hat{\theta}_k}{\|\beta\| c_{\hat{\theta}}} + 6K_{cell}^2 \sum_{l \in J^C \setminus J_{\text{entrant}}^C} \frac{h^2}{\hat{\sigma}^2 \|\beta\|^2} f_l^2 h^2 \end{aligned} \quad (17.317)$$

Ainsi (en multipliant l'inégalité par  $h$ ) :

$$\sum_{i \in J_{\text{sortant}}^C} u_i^2 h \hat{\theta}_i \leq \sum_{i \in J_{\text{entrant}}^C} u_i^2 h \hat{\theta}_i \left( 1 + h \frac{6K_{cell}^2}{\|\beta\| c_{\hat{\theta}}} \right) + \sum_{i \in J^C \setminus J_{\text{entrant}}^C} f_i^2 h^3 \left( 1 + \frac{6K_{cell}^2 h^2}{\hat{\sigma}^2 \|\beta\|^2} \right) \quad (17.318)$$

**Estimation des points d'entrés (points de contrôles) :** Nous supposons les strates  $\Xi^1, \Xi^2, \dots, \xi^s, \Xi^{s+1}, \dots, \Xi^{N-1}, \Xi^N$  de  $\mathcal{C}_h$  numérotées de manière à ce que

$$\bigcup_{m=1}^s \bigcup_{C \in \Xi^m} J_{\text{sortant}}^C \subset \bigcup_{m=1}^{s+1} \bigcup_{C \in \Xi^m} J_{\text{entrant}}^C \text{ avec } \mathcal{C}_h = \bigcup_{m=1}^N \Xi^m$$

Cette inclusion exprime que les points sortants d'une strate sont inclus dans les points entrants d'une strate supérieure. Les points  $Q_i$  sur le bord entrant  $\Gamma^-$  sont exactement les points qui ne peuvent être pas un point sortant d'une strate.

Pour  $n \in \{1, \dots, N\}$ , posons

$$K_n = \sum_{s=1}^n \sum_{C \in \Xi^s} \sum_{i \in J_{\text{entrant}}^C} u_i^2 h \hat{\theta}_i$$

Pour la strate  $\Xi^1$  on a, bien sur,  $J_{\text{entrant}}^C \subset J^-$ , car cette première strate se calcule directement à partir de certaines valeurs du bord, donc

$$K_1 \leq h L \text{ où } L = \sum_{i \in J^-} g_i^2 \hat{\theta}_i \quad (17.319)$$

Nous voulons estimer  $K_N$ . Pour cela une inégalité de récurrence sera établie pour  $K_n$ .

---

52.  $h \|\beta\| c_{\hat{\theta}} \leq \hat{\theta}_j$

$$K_n = \sum_{s=1}^{n-1} \sum_{C \in \Xi^s} \sum_{i \in J_{\text{sortant}}^C} u_i^2 h \hat{\theta}_i + \sum_{i \in (\cup_{m=1}^n \cup_{C \in \Xi^m} J_{\text{entrant}}^C) \setminus (\cup_{m=1}^{n-1} \cup_{C \in \Xi^m} J_{\text{sortant}}^C)} u_i^2 h \hat{\theta}_i$$

Les points  $Q_i$  sur le bord entrant  $\Gamma^-$  sont exactement les points qui ne peuvent être pas un point sortant d'une strate, donc

$$(\cup_{m=1}^n \cup_{C \in \Xi^m} J_{\text{entrant}}^C) \setminus (\cup_{m=1}^{n-1} \cup_{C \in \Xi^m} J_{\text{sortant}}^C) \subset J^-$$

Ainsi

$$\begin{aligned} K_n &\leq \sum_{s=1}^{n-1} \sum_{C \in \Xi^s} \sum_{i \in J_{\text{sortant}}^C} u_i^2 h \hat{\theta}_i + h L \\ (17.318) \quad &\leq h L + \sum_{s=1}^{n-1} \sum_{C \in \Xi^s} \sum_{i \in J_{\text{entrant}}^C} u_i^2 h \hat{\theta}_i \left( 1 + h \frac{6K_{\text{cell}}^2}{\|\beta\| c_{\hat{\theta}}} \right) \\ &\quad + \sum_{s=1}^{n-1} \sum_{C \in \Xi^s} \sum_{i \in J^C \setminus J_{\text{entrant}}^C} f_i^2 h^3 \left( 1 + \frac{6K_{\text{cell}}^2 h^2}{\hat{\sigma}^2 \|\beta\|^2} \right) h^2 \\ K_n &\leq K_{n-1} q + h L + h M \text{ où } q = \left( 1 + h \frac{6K_{\text{cell}}^2}{\|\beta\| c_{\hat{\theta}}} \right) \text{ et } M = \sum_{i \in J \setminus J^-} f_i^2 h^2 \left( 1 + \frac{6K_{\text{cell}}^2 h^2}{\hat{\sigma}^2 \|\beta\|^2} \right) \end{aligned}$$

En procédant par récurrence sur  $n$  on obtient

$$K_n \leq h L q^{n-1} + \frac{q^{n-1} - 1}{q - 1} h(L + M) \quad (17.320)$$

La famille de maillage étant quasi-uniforme et régulière, il existe  $h_0 > 0$ , tel que pour  $h \in ]0; h_0[$  on ait  $\frac{h}{\sigma c_0} \leq |S_i| \leq h$ , donc  $L \leq \sigma c_0 \|\beta\| \|g_h\|_{L_h^2(\Gamma^-)}$ .

De plus  $h^2 \leq \hat{c}_0 |T|$ , donc  $M \leq 3\hat{c}_0 \left( 1 + \frac{6K_{\text{cell}}^2 h_0^2}{\hat{\sigma}^2 \|\beta\|^2} \right) \|f_h\|_{L_h^2(\Omega)}$

Maintenant  $N \leq K_{\Xi} \frac{1}{h}$ , donc  $q^{N-1} \leq q^N \leq \left[ e^{\frac{h 6K_{\text{cell}}^2}{\|\beta\| c_{\hat{\theta}}}} \right]^N \leq e^{K_{\Xi} \frac{6K_{\text{cell}}^2}{\|\beta\| c_{\hat{\theta}}}}$  et  $\frac{h}{q-1} \leq \frac{\|\beta\| c_{\hat{\theta}}}{6K_{\text{cell}}^2}$

Ainsi

$$\begin{aligned} K_N &\leq h_0 \sigma c_0 \|\beta\| \|g_h\|_{L_h^2(\Gamma^-)}^2 e^{K_{\Xi} \frac{6K_{\text{cell}}^2}{\|\beta\| c_{\hat{\theta}}}} + \\ &\quad e^{K_{\Xi} \frac{6K_{\text{cell}}^2}{\|\beta\| c_{\hat{\theta}}}} \frac{\|\beta\| c_{\hat{\theta}}}{6K_{\text{cell}}^2} \left[ \sigma c_0 \|\beta\| \|g_h\|_{L_h^2(\Gamma^-)}^2 + 3\hat{c}_0 \left( 1 + \frac{6K_{\text{cell}}^2 h^2}{\hat{\sigma}^2 \|\beta\|^2} \right) \|f_h\|_{L_h^2(\Omega)}^2 \right] \end{aligned}$$

$$K_N \leq \|\beta\| e^{K_{\Xi} \frac{6K_{\text{cell}}^2}{\|\beta\| c_{\hat{\theta}}}} \left[ (h_0 + \|\beta\|) \sigma c_0 \|g_h\|_{L_h^2(\Gamma^-)}^2 + \frac{\hat{c}_0 c_{\hat{\theta}}}{2K_{\text{cell}}^2} \left( 1 + \frac{6K_{\text{cell}}^2 h^2}{\hat{\sigma}^2 \|\beta\|^2} \right) \|f_h\|_{L_h^2(\Omega)}^2 \right]$$

**Estimation des points intérieurs, sortants des cellules et conclusion**

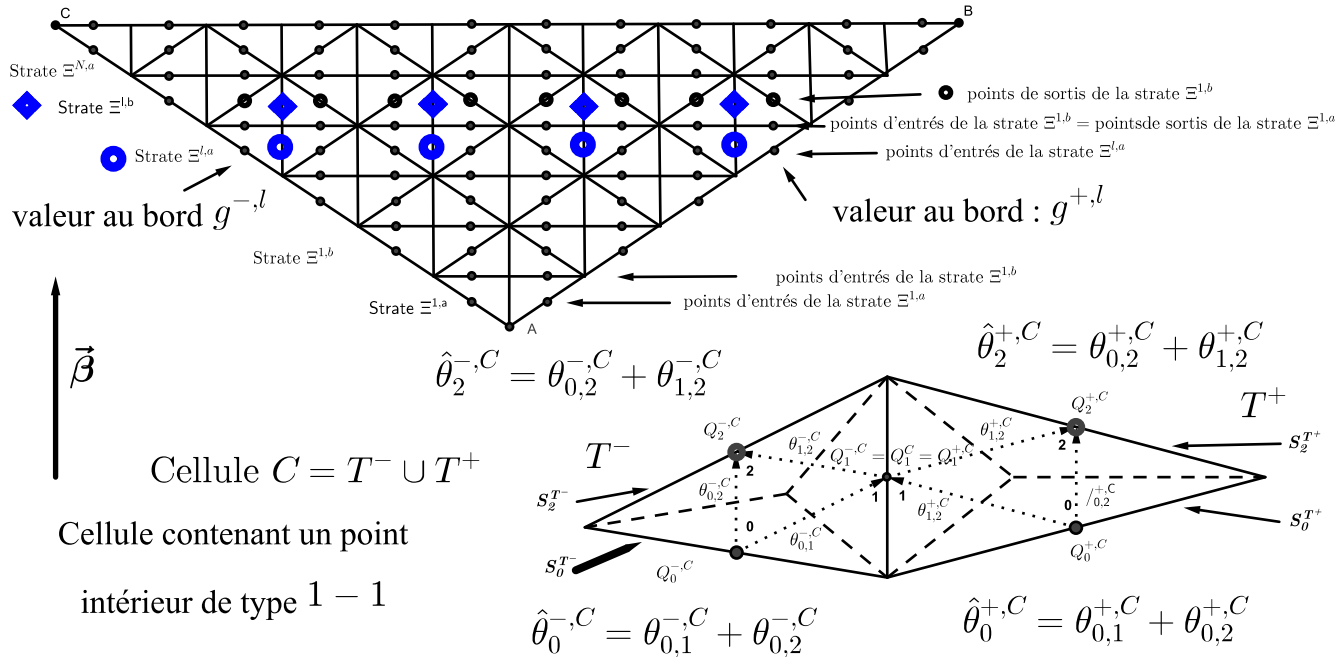
$$\begin{aligned}
 & \|u_h\|_{L_h^2(\Omega)}^2 \\
 = & \sum_{s=1}^N \sum_{C \in \Xi^s} \sum_{i \in J^C \setminus J_{\text{entrant}}^C} u_i^2 |D_i| + \sum_{i \in J^-} g_i^2 |D_i| \\
 (17.317) \leq & \sum_{s=1}^N \sum_{C \in \Xi^s} \sum_{i \in J^C \setminus J_{\text{entrant}}^C} 6K_{\text{cell}}^2 \sum_{k \in J_{\text{entrant}}^C} u_k^2 h \frac{\hat{\theta}_k}{\|\beta\| c_{\hat{\theta}}} + \\
 & \sum_{s=1}^N \sum_{C \in \Xi^s} \sum_{i \in J^C \setminus J_{\text{entrant}}^C} 6K_{\text{cell}}^2 \sum_{l \in J^C \setminus J_{\text{entrant}}^C} \frac{h^2}{\hat{\sigma}^2 \|\beta\|^2} f_l^2 h^2 + \sum_{i \in J^-} g_i^2 h^2 \\
 \leq & \frac{6K_{\text{cell}}^2}{c_{\hat{\theta}}} e^{K_{\Xi} \frac{6K_{\text{cell}}^2}{\|\beta\| c_{\hat{\theta}}}} \left[ (h_0 + \|\beta\|) \sigma c_0 \|g_h\|_{L_h^2(\Gamma^-)}^2 + \frac{\hat{c}_0 c_{\hat{\theta}}}{2K_{\text{cell}}^2} \left( 1 + \frac{6K_{\text{cell}}^2 h_0^2}{\hat{\sigma}^2 \|\beta\|^2} \right) \|f_h\|_{L_h^2(\Omega)}^2 \right] + \\
 & 6K_{\text{cell}}^2 \frac{h_0^2}{\hat{\sigma}^2 \|\beta\|^2} 3\hat{c}_0 \|f_h\|_{L_h^2(\Omega)}^2 + h_0 \sigma c_0 \|g_h\|_{L_h^2(\Gamma^-)}^2
 \end{aligned}$$

On peut donc conclure à l'existence d'une constante  $C > 0$ , qui ne dépend ni de  $h$ , ni de  $\|\beta\|$ , telle que

$$\|u_h\|_{L_h^2(\Omega)}^2 \leq C e^{C/\|\beta\|} \left[ (h_0 + \|\beta\|) \|g_h\|_{L_h^2(\Gamma^-)}^2 + \left( 1 + \frac{h_0^2}{\|\beta\|^2} \right) \|f_h\|_{L_h^2(\Omega)}^2 \right]$$

Pour conclure ce chapitre, un exemple :

### Exemple 17.2



Le domaine  $\Omega$  est donné par le triangle  $ABC$ . Le maillage  $\mathcal{T}_h$  est défini comme dans la figure ci-dessus avec  $h = O(1/N)$ . Le cas  $N = 6$  y est représenté. Le maillage n'est pas nécessairement fait de triangles isométriques, mais toutes les cellules  $C$  sont même type et contiennent un point intérieur 1 – 1 (voir ci-dessus).

L'objectif de cet exemple est de montrer l'existence de constantes  $C > 0$  et  $h_0 > 0$ , qui ne dépendent pas de  $h$  et  $\|\beta\|$ , telles que pour  $h \in ]0; h_0[$  :

$$\text{a) } \|u_h\|_{L_h^1(\Omega)} \leq C \left[ \frac{1}{\|\beta\|} \|f_h\|_{L_h^1(\Omega)} + \|g_h\|_{L_h^1(\Gamma^-)} \right]$$

$$\text{b) } \|u_h\|_{L_h^2(\Omega)} \leq C e^{\frac{C}{\|\beta\|}} \left[ \left(1 + \frac{h_0}{\|\beta\|}\right) \|f_h\|_{L_h^2(\Omega)} + (\|\beta\| + h_0)^{1/2} \|g_h\|_{L_h^2(\Gamma^-)} \right]$$

**Étape préparatoire : Notations** Pour une cellule  $C = T^- \cup T^+ \in \mathcal{C}_h$  on introduit des notations (en coordonnées locales) relatives à la cellule  $C$  :  $\theta_{i,j}^{-,C} = \theta_{i,j}^{T^-}$  et  $\theta_{i,j}^{+,C} = \theta_{i,j}^{T^+}$ . Ainsi

$$\begin{aligned} \hat{\theta}_2^{+,C} &= \int_{S_2^{T^+}} \beta \cdot n_{\partial C} d\sigma = \theta_{0,2}^{+,C} + \theta_{1,2}^{+,C} > 0, \quad \hat{\theta}_2^{-,C} = \int_{S_2^{T^-}} \beta \cdot n_{\partial C} d\sigma = \theta_{0,2}^{-,C} + \theta_{1,2}^{-,C} > 0 \\ \hat{\theta}_0^{+,C} &= - \int_{S_0^{T^+}} \beta \cdot n_{\partial C} d\sigma = \theta_{0,1}^{+,C} + \theta_{0,2}^{+,C} > 0, \quad \hat{\theta}_0^{-,C} = - \int_{S_0^{T^-}} \beta \cdot n_{\partial C} d\sigma = \theta_{0,1}^{-,C} + \theta_{0,2}^{-,C} > 0 \end{aligned}$$

Pour  $Q_i$  posons  $\alpha_i = f_h(Q_i) \frac{|D_i|}{\sum_{j \in S^+(i)} \theta_{i,j}}$

Alors, avec le lemme 12.2, on a  $|\alpha_i| \leq |f_h(Q_i)| \frac{h}{\|\beta\| \bar{\sigma}}$

On rappelle la formule de récurrence<sup>53</sup> :  $u_i = \sum_{j \in S^-(i)} u_j \lambda_{j,i} + \alpha_i$  et on posera  $f_i = f_h(Q_i)$

où  $\lambda_{j,i} = \frac{\theta_{i,j}}{\sum_{k \in S^+(i)} \theta_{i,k}} (= \frac{\theta_{k,i}}{\sum_{k \in S^-(i)} \theta_{k,i}} = \frac{\theta_{i,j}}{\bar{\theta}_i})$ .

**Calcul de la cellule  $C$  :**

$$\begin{aligned} u_1^C &= u_0^{-,C} \lambda_{0,1}^{-,C} + u_0^{+,C} \lambda_{0,1}^{+,C} + \alpha_1^C \\ u_2^{-,C} &= u_0^{-,C} \lambda_{0,2}^{-,C} + u_1^C \lambda_{1,2}^{-,C} + \alpha_2^{-,C} \\ u_2^{+,C} &= u_0^{+,C} \lambda_{0,2}^{+,C} + u_1^C \lambda_{1,2}^{+,C} + \alpha_2^{+,C} \end{aligned}$$

Donc (en remplaçant  $u_1^C$  dans les formules suivantes) :

$$\begin{aligned} u_2^{-,C} &= u_0^{-,C} \omega^{-,C} + u_0^{+,C} (1 - \omega^{-,C}) + \alpha_1^C \lambda_{1,2}^{-,C} + \alpha_2^{-,C} \\ \text{avec } \omega^{-,C} &= \lambda_{0,2}^{-,C} + \lambda_{0,1}^{-,C} \lambda_{1,2}^{-,C} \in ]0, 1] \text{ et } 1 - \omega^{-,C} = \lambda_{0,1}^{+,C} \lambda_{1,2}^{-,C} \geq 0 \\ (\text{car } \lambda_{0,2}^{-,C} + \lambda_{0,1}^{-,C} \lambda_{1,2}^{-,C} + \lambda_{0,1}^{+,C} \lambda_{1,2}^{-,C} &= \lambda_{0,2}^{-,C} + \underbrace{(\lambda_{0,1}^{-,C} + \lambda_{0,1}^{+,C})}_{=1} \lambda_{1,2}^{-,C} = \lambda_{0,2}^{-,C} + \lambda_{1,2}^{-,C} = 1) \end{aligned}$$

$$\begin{aligned} u_2^{+,C} &= u_0^{+,C} \omega^{+,C} + u_0^{-,C} (1 - \omega^{+,C}) + \alpha_1^C \lambda_{1,2}^{+,C} + \alpha_2^{+,C} \\ \text{avec } \omega^{+,C} &= \lambda_{0,2}^{+,C} + \lambda_{0,1}^{+,C} \lambda_{1,2}^{+,C} \in ]0, 1] \text{ et } 1 - \omega^{+,C} = \lambda_{0,1}^{-,C} \lambda_{1,2}^{+,C} \geq 0 \end{aligned}$$

De plus on a les relations

$$\begin{aligned} \omega^{-,C} \hat{\theta}_2^{-,C} + (1 - \omega^{+,C}) \hat{\theta}_2^{+,C} &= (\lambda_{0,2}^{-,C} + \lambda_{0,1}^{-,C} \lambda_{1,2}^{-,C}) \hat{\theta}_2^{-,C} + \lambda_{0,1}^{-,C} \lambda_{1,2}^{+,C} \hat{\theta}_2^{+,C} \\ &= \theta_{0,2}^{-,C} + \lambda_{0,1}^{-,C} \theta_{1,2}^{-,C} + \lambda_{0,1}^{-,C} \theta_{1,2}^{+,C} \\ &= \theta_{0,2}^{-,C} + \lambda_{0,1}^{-,C} (\theta_{1,2}^{-,C} + \theta_{1,2}^{+,C}) \\ &= \theta_{0,2}^{-,C} + \theta_{0,1}^{-,C} = \hat{\theta}_0^{-,C} \end{aligned}$$

et (preuve analogue)  $\omega^{+,C} \hat{\theta}_2^{+,C} + (1 - \omega^{-,C}) \hat{\theta}_2^{-,C} = \hat{\theta}_0^{+,C}$

53. Voir (12.209), page 97.

**Estimation clé pour  $\|u_h\|_{L_h^1(\Omega)}$  :**

En combinant les inégalités obtenues ci-dessus ( $0 < \hat{\theta}_2^{+,C}$ ,  $\hat{\theta}_2^{-,C} \leq \|\beta\| h$ ) on obtient la relation :

$$\begin{aligned} |u_2^{-,C} \hat{\theta}_2^{-,C} + |u_2^{+,C} \hat{\theta}_2^{+,C} &\leq [|u_0^{-,C} \omega^{-,C} + |u_0^{+,C} (1 - \omega^{-,C}) + |\alpha_1^C \lambda_{1,2}^{-,C} + |\alpha_2^{-,C} |] \hat{\theta}_2^{-,C} \\ &\quad + [|u_0^{+,C} \omega^{+,C} + |u_0^{-,C} (1 - \omega^{+,C}) + |\alpha_1^C \lambda_{1,2}^{+,C} + |\alpha_2^{+,C} |] \hat{\theta}_2^{+,C} \\ |u_2^{-,C} \hat{\theta}_2^{-,C} + |u_2^{+,C} \hat{\theta}_2^{+,C} &\leq |u_0^{-,C} \hat{\theta}_0^{-,C} + |u_0^{+,C} \hat{\theta}_0^{+,C} + [2|f_1^C| + |f_2^{+,C}| + |f_2^{-,C}|] \frac{1}{\delta} h^2 \end{aligned} \quad (17.321)$$

**Estimation clé pour  $\|u_h\|_{L_h^2(\Omega)}$  :**

Il est nécessaire de revenir vers la forme bilinéaire du terme de convection. Soit  $C \in \mathcal{C}_h$  une cellule quelconque, alors on définit la fonction test  $v_h(Q_i) = u_h(Q_i)$  si  $i \in \mathcal{J}^C \setminus \mathcal{J}_{entrant}^C$  et  $v_h(Q_i) = 0$  sinon. Alors  $v_h \in Y_0^h$  et  $\hat{b}_h(u_h, v_h) = (f_h, v_h)$ . Donc<sup>54</sup>

$$\begin{aligned} \sum_{i \in \mathcal{J} \setminus \mathcal{J}^-} v_h(Q_i) \left[ \sum_{j \in \mathcal{S}^+(i)} \theta_{ij} u_h(Q_j) + \sum_{j \in \mathcal{S}^-(i)} \theta_{ij} u_h(Q_j) \right] &= \sum_{k \in \mathcal{J} \setminus \mathcal{J}^-} f_h(Q_k) v_h(Q_k) |D_k| \\ &\quad u_2^{-,C} [u_2^{-,C} \hat{\theta}_2^{-,C} - u_0^{-,C} \theta_{0,2}^{-,C} - u_1^C \theta_{1,2}^{-,C}] + u_1^C [u_1^C \bar{\theta}_1^C - u_0^{-,C} \theta_{0,1}^{-,C} - u_0^{+,C} \theta_{0,1}^{+,C}] \\ + u_2^{+,C} [u_2^{+,C} \hat{\theta}_2^{+,C} - u_0^{+,C} \theta_{0,2}^{+,C} - u_1^C \theta_{1,2}^{+,C}] &= u_2^{-,C} f_2^{-,C} |D_2^{-,C}| + u_1^C f_1^C |D_1^C| + u_2^{+,C} f_2^{+,C} |D_2^{+,C}| \end{aligned}$$

Donc

$$\begin{aligned} [u_2^{-,C}]^2 \hat{\theta}_2^{-,C} + [u_1^C]^2 \bar{\theta}_1^C + [u_2^{+,C}]^2 \hat{\theta}_2^{+,C} &= \\ u_2^{-,C} u_0^{-,C} \theta_{0,2}^{-,C} + u_2^{-,C} u_1^C \theta_{1,2}^{-,C} + u_1^C u_0^{-,C} \theta_{0,1}^{-,C} + u_1^C u_0^{+,C} \theta_{0,1}^{+,C} + u_2^{+,C} u_0^{+,C} \theta_{0,2}^{+,C} + u_2^{+,C} u_1^C \theta_{1,2}^{+,C} \\ + u_2^{-,C} f_2^{-,C} |D_2^{-,C}| + u_1^C f_1^C |D_1^C| + u_2^{+,C} f_2^{+,C} |D_2^{+,C}| \end{aligned}$$

Comme  $\theta_{0,2}^{-,C}, \theta_{1,2}^{+,C}, \theta_{0,2}^{+,C} \geq 0$ , avec  $ab \leq \frac{1}{2}(a^2 + b^2)$  (\*) pour  $a, b \in \mathbb{R}$ , et les formules explicites pour  $u_2^{-,C}$ ,  $u_1^C$  et  $u_2^{+,C}$  :

$$\begin{aligned} [u_2^{-,C}]^2 \hat{\theta}_2^{-,C} + [u_1^C]^2 \bar{\theta}_1^C + [u_2^{+,C}]^2 \hat{\theta}_2^{+,C} &\leq \\ \frac{1}{2} [u_2^{-,C}]^2 \hat{\theta}_2^{-,C} + \frac{1}{2} [u_2^{+,C}]^2 \hat{\theta}_2^{+,C} + [u_1^C]^2 \bar{\theta}_1^C + \frac{1}{2} [u_0^{-,C}]^2 \hat{\theta}_0^{-,C} + \frac{1}{2} [u_0^{+,C}]^2 \hat{\theta}_0^{+,C} \\ + [u_0^{-,C} \omega^{-,C} + u_0^{+,C} (1 - \omega^{-,C}) + \alpha_1^C \lambda_{1,2}^{-,C} + \alpha_2^{-,C}] f_2^{-,C} |D_2^{-,C}| \\ + [u_0^{-,C} \lambda_{0,1}^{-,C} + u_0^{+,C} \lambda_{0,1}^{+,C} + \alpha_1^C] f_1^C |D_1^C| \\ + [u_0^{+,C} \omega^{+,C} + u_0^{-,C} (1 - \omega^{+,C}) + \alpha_1^C \lambda_{1,2}^{+,C} + \alpha_2^{+,C}] f_2^{+,C} |D_2^{+,C}| \end{aligned}$$

Ainsi (soustraction de  $\frac{1}{2}[u_2^{-,C}]^2 \hat{\theta}_2^{-,C} + [u_0^{-,C}]^2 \hat{\theta}_0^{-,C} + \frac{1}{2}[u_2^{+,C}]^2 \hat{\theta}_2^{+,C}$ , multiplication par 2, inégalité (\*),  $1 \geq \lambda_{0,1}^{\pm,C}, \lambda_{0,2}^{\pm,C}, \lambda_{1,2}^{\pm,C} \geq 0$ ,  $1 \geq \omega^{\pm,C} \geq 0$ ,  $h^2 \geq |D_1^C|, |D_2^{\pm,C}|$  et  $|\alpha_i| \leq |f_h(Q_i)| \frac{h}{\|\beta\| \delta}$ ) :

$$\begin{aligned} [u_2^{-,C}]^2 \hat{\theta}_2^{-,C} + [u_2^{+,C}]^2 \hat{\theta}_2^{+,C} &\leq [u_0^{-,C}]^2 \hat{\theta}_0^{-,C} + [u_0^{+,C}]^2 \hat{\theta}_0^{+,C} + 3[u_0^{-,C}]^2 h^2 + 3[u_0^{+,C}]^2 h^2 \\ &\quad + 3 \left[ (f_2^{-,C})^2 + (f_1^C)^2 + (f_2^{+,C})^2 \right] \left( h^2 + \frac{h^3}{\|\beta\| \delta} \right) \end{aligned}$$

54. Rappelons que  $\bar{\theta}_i = \sum_{j \in \mathcal{S}^+(i)} \theta_{i,j} = - \sum_{j \in \mathcal{S}^-(i)} \theta_{j,i}$ . De plus, comme les points d'entrées et de sorties des cellules sont de type 0-2, on a  $\bar{\theta}_2^{-,C} = \hat{\theta}_2^{-,C}$  et  $\bar{\theta}_2^{+,C} = \hat{\theta}_2^{+,C}$ .

$$\begin{aligned}
[u_2^-, C]^2 \hat{\theta}_2^-, C + [u_2^+, C]^2 \hat{\theta}_2^+, C &\leq \left[ [u_0^-, C]^2 \hat{\theta}_0^-, C + [u_0^+, C]^2 \hat{\theta}_0^+, C \right] \left( 1 + h \frac{3}{\|\beta\| c_{\hat{\theta}}} \right) + \\
&+ 3 \left[ (f_2^-, C)^2 + (f_1^C)^2 + (f_2^+, C)^2 \right] h^2 \left( 1 + \frac{h}{\|\beta\| \hat{\sigma}} \right)
\end{aligned} \tag{17.322}$$

**Estimation des points d'entrées (points de contrôles) :** *La situation simple de notre maillage permet de se passer de l'utilisation de la matrice qui est nécessaire dans la preuve du cas général :*

Dans les notations introduites ci-dessus  $\mathcal{C}_h = (\cup_{l=1}^N \Xi^{a,l}) \cup (\cup_{l=1}^{N-1} \Xi^{b,l})$ .

Pour  $n \in \{1, \dots, N\}$ , posons

$$K_n^{L^1} = \sum_{l=1}^n \sum_{C \in \Xi^{a,l}} \left( |u_0^-, C| h \hat{\theta}_0^-, C + |u_0^+, C| h \hat{\theta}_0^+, C \right) + \sum_{l=1}^{n-1} \sum_{C \in \Xi^{b,l}} \left( |u_0^-, C| h \hat{\theta}_0^-, C + |u_0^+, C| h \hat{\theta}_0^+, C \right)$$

et

$$K_n^{L^2} = \sum_{l=1}^n \sum_{C \in \Xi^{a,l}} \left( [u_0^-, C]^2 h \hat{\theta}_0^-, C + [u_0^+, C]^2 h \hat{\theta}_0^+, C \right) + \sum_{l=1}^{n-1} \sum_{C \in \Xi^{b,l}} \left( [u_0^-, C]^2 h \hat{\theta}_0^-, C + [u_0^+, C]^2 h \hat{\theta}_0^+, C \right)$$

Nous voulons estimer  $K_N^{L^1}$  et  $K_N^{L^2}$ . Pour cela une inégalité de récurrence sera établie pour  $K_n^{L^1}$  et  $K_n^{L^2}$  :

**Pour  $K_n^{L^1}$  :**

Pour  $l \in \{1, \dots, N-1\}$  et  $C \in \Xi^{b,l}$  on a (les points d'entrées de  $\Xi^{b,l}$  sont les points de sortis de  $\Xi^{a,l}$ )

$$\begin{aligned}
\sum_{C \in \Xi^{b,l}} \left( |u_0^-, C| h \hat{\theta}_0^-, C + |u_0^+, C| h \hat{\theta}_0^+, C \right) &= \sum_{C \in \Xi^{a,l}} \left( |u_2^-, C| h \hat{\theta}_2^-, C + |u_2^+, C| h \hat{\theta}_2^+, C \right) \\
&\stackrel{(17.321)}{\leq} \sum_{C \in \Xi^{a,l}} \left( |u_0^-, C| h \hat{\theta}_0^-, C + |u_0^+, C| h \hat{\theta}_0^+, C \right) + \sum_{C \in \Xi^{a,l}} [2|f_1^C| + |f_2^+, C| + |f_2^-, C|] \frac{1}{\hat{\sigma}} h^3
\end{aligned} \tag{17.323}$$

Maintenant, pour  $l \in \{2, \dots, N\}$  et  $C \in \Xi^{a,l}$ , les points d'entrées de  $\Xi^{a,l}$  sont les points de sortis de  $\Xi^{b,l-1}$ , sauf pour deux points sur le bord entrant dont les valeurs  $g^{-,l-1}$  et  $g^{+,l-1}$  sont connues. Ainsi

$$\begin{aligned}
\sum_{C \in \Xi^{a,l}} \left( |u_0^-, C| h \hat{\theta}_0^-, C + |u_0^+, C| h \hat{\theta}_0^+, C \right) &= \sum_{C \in \Xi^{b,l-1}} \left( |u_2^-, C| h \hat{\theta}_2^-, C + |u_2^+, C| h \hat{\theta}_2^+, C \right) + \\
&|g^{-,l-1}| \hat{\theta}^{-,l-1} + |g^{+,l-1}| \hat{\theta}^{+,l-1} \\
&\stackrel{(17.321)}{\leq} \sum_{C \in \Xi^{b,l-1}} \left( |u_0^-, C| h \hat{\theta}_0^-, C + |u_0^+, C| h \hat{\theta}_0^+, C \right) + \sum_{C \in \Xi^{b,l-1}} [2|f_1^C| + |f_2^+, C| + |f_2^-, C|] \frac{1}{\hat{\sigma}} h^3 \\
&+ |g^{-,l-1}| h^2 \|\beta\| + |g^{+,l-1}| h^2 \|\beta\|
\end{aligned} \tag{17.324}$$

Ainsi

$$\begin{aligned}
K_n^{L^1} &\leq \sum_{l=1}^n \sum_{C \in \Xi^{b,l-1}} \left( |u_0^{-,C}| h \hat{\theta}_0^{-,C} + |u_0^{+,C}| h \hat{\theta}_0^{+,C} \right) + \\
&\sum_{l=1}^n \sum_{C \in \Xi^{b,l-1}} [2|f_1^C| + |f_2^{+,C}| + |f_2^{-,C}|] \frac{1}{\hat{\sigma}} h^3 + \sum_{l=1}^n |g^{-,l-1}| h^2 \|\beta\| + |g^{+,l-1}| h^2 \|\beta\| \\
&+ \sum_{l=1}^{n-1} \sum_{C \in \Xi^{a,l}} \left( |u_0^{-,C}| h \hat{\theta}_0^{-,C} + |u_0^{+,C}| h \hat{\theta}_0^{+,C} \right) \\
&+ \sum_{l=1}^{n-1} \sum_{C \in \Xi^{a,l}} [2|f_1^C| + |f_2^{+,C}| + |f_2^{-,C}|] \frac{1}{\hat{\sigma}} h^3 \tag{17.325}
\end{aligned}$$

En appliquant encore une fois les inégalités (17.323) et (17.324), on obtient finalement

$$\begin{aligned}
K_n^{L^1} &\leq K_{n-1}^{L^1} + \sum_{l=1}^n \sum_{C \in \Xi^{b,l-1} \cup \Xi^{a,l-1}} [2|f_1^C| + |f_2^{+,C}| + |f_2^{-,C}|] \frac{1}{\hat{\sigma}} h^3 \\
&+ \sum_{l=1}^n |g^{-,l-1}| h^2 \|\beta\| + |g^{+,l-1}| h^2 \|\beta\| \\
&+ \sum_{l=1}^{n-1} \sum_{C \in \Xi^{a,l} \cup \Xi^{b,l-1}} [2|f_1^C| + |f_2^{+,C}| + |f_2^{-,C}|] \frac{1}{\hat{\sigma}} h^3 \\
&+ \sum_{l=1}^n |g^{-,l-2}| h^2 \|\beta\| + |g^{+,l-2}| h^2 \|\beta\|
\end{aligned}$$

Donc

$$K_n^{L^1} \leq K_{n-1}^{L^1} + 4 \sum_{i \in J \setminus J^-} |f_i| \frac{1}{\hat{\sigma}} h^3 + 2 \sum_{i \in J^-} |g_i| h^2 \|\beta\| \tag{17.326}$$

avec  $K_1^{L^1} = \sum_{C \in \Xi^{a,1}} \left( |u_0^{-,C}| h \hat{\theta}_0^{-,C} + |u_0^{+,C}| h \hat{\theta}_0^{+,C} \right) \leq |g^{-,1}| h^2 \|\beta\| + |g^{+,1}| h^2 \|\beta\| \leq \sum_{i \in J^-} |g_i| h^2 \|\beta\|$  (car la strate  $\Xi^{a,1}$  ne contient qu'une cellule dont les points d'entrés sont sur le bord entrant). Ainsi (en raisonnant par récurrence sur  $n$ ) :

$$K_N \leq 4(Nh) \frac{1}{\hat{\sigma}} \sum_{i \in J \setminus J^-} |f_i| h^2 + 4(hN) \|\beta\| \sum_{i \in J^-} |g_i| h$$

**Pour  $K_n^{L^2}$  :** La procédure est la même, à la différence notable du facteur  $1 + h \frac{3}{\|\beta\| c_{\hat{\theta}}}$  qui apparaît dans (17.322). On obtient :

$$\begin{aligned}
K_1^{L^2} &\leq \sum_{i \in J^-} |g_i|^2 h^2 \|\beta\| \text{ et } K_n^{L^2} \leq K_{n-1}^{L^2} q(h) + h p(h) \text{ avec } q(h) = \left( 1 + h \frac{3}{\|\beta\| c_{\hat{\theta}}} \right)^2 \\
\text{et } p(h) &= \left( 1 + h \frac{3}{\|\beta\| c_{\hat{\theta}}} \right) \left\{ 6 \left( 1 + \frac{h}{\|\beta\| \hat{\sigma}} \right) \sum_{i \in J \setminus J^-} |f_i|^2 h^2 + 2 \sum_{i \in J^-} |g_i|^2 h \|\beta\| \right\} \tag{17.327}
\end{aligned}$$



En procédant par itération sur  $n$  on obtient :

$$K_N^{L^2} \leq K_1^{L^2} q(h)^N + hp(h) \frac{q(h)^N - 1}{q(h) - 1} \leq e^{\frac{6Nh}{\|\beta\| c_{\hat{\theta}}}} \left[ K_1^{L^2} + p(h) \frac{\|\beta\| c_{\hat{\theta}}}{6} \right] \quad (17.328)$$

$$\text{car } q(h)^N \leq \left( e^{\frac{h}{\|\beta\| c_{\hat{\theta}}}} \right)^{2N} \text{ et } h \frac{q(h)^N - 1}{q(h) - 1} \leq e^{2Nh \frac{3}{\|\beta\| c_{\hat{\theta}}}} \frac{\|\beta\| c_{\hat{\theta}}}{6}.$$

**Estimation des points intérieurs et sortants des cellules**  
(ce qui inclus donc les points de  $\Gamma^+ = [BC]$ )

Pour la norme  $L_h^1$  :

$$\begin{aligned} & \sum_{C \in (\cup_{i=1}^N \Xi^{a,l}) \cup (\cup_{i=1}^{N-1} \Xi^{b,l})} |u_2^{-,C}| |D_2^{-,C}| + |u_1^{-,C}| |D_1^{-,C}| + |u_2^{+,C}| |D_2^{+,C}| \\ \leq & \sum_{C \in (\cup_{i=1}^N \Xi^{a,l}) \cup (\cup_{i=1}^{N-1} \Xi^{b,l})} [|u_0^{-,C}| \omega^{-,C} + |u_0^{+,C}| (1 - \omega^{-,C}) + |\alpha_1^C| \lambda_{1,2}^{-,C} + |\alpha_2^{-,C}|] h^2 + \\ & \sum_{C \in (\cup_{i=1}^N \Xi^{a,l}) \cup (\cup_{i=1}^{N-1} \Xi^{b,l})} [|u_0^{-,C}| \lambda_{0,1}^{-,C} + |u_0^{+,C}| \lambda_{0,1}^{+,C} + |\alpha_1^C|] h^2 \\ & \sum_{C \in (\cup_{i=1}^N \Xi^{a,l}) \cup (\cup_{i=1}^{N-1} \Xi^{b,l})} [|u_0^{+,C}| \omega^{+,C} + |u_0^{-,C}| (1 - \omega^{+,C}) + |\alpha_1^C| \lambda_{1,2}^{+,C} + |\alpha_2^{+,C}|] h^2 \\ \leq & 3 \sum_{C \in (\cup_{i=1}^N \Xi^{a,l}) \cup (\cup_{i=1}^{N-1} \Xi^{b,l})} [|u_0^{-,C}| + |u_0^{+,C}| + \frac{h}{\|\beta\| \hat{\sigma}} (|f_2^{-,C}| + |f_1^C| + |f_2^{+,C}|)] h^2 \\ & \text{Avec le lemme 16.1, on a} \\ \leq & \frac{3}{\|\beta\| c_{\hat{\theta}}} \sum_{C \in (\cup_{i=1}^N \Xi^{a,l}) \cup (\cup_{i=1}^{N-1} \Xi^{b,l})} (|u_0^{-,C}| h \hat{\theta}_0^{-,C} + |u_0^{+,C}| h \hat{\theta}_0^{+,C}) \\ & + \frac{3}{\|\beta\| \hat{\sigma}} h \sum_{i \in J \setminus J^-} |f_h(Q_i)| h^2 \leq \frac{3}{\|\beta\| c_{\hat{\theta}}} K_N^{L^2} + \frac{3}{\|\beta\| \hat{\sigma}} h \sum_{i \in J \setminus J^-} |f_h(Q_i)| h^2 \end{aligned}$$

Pour la norme  $L_h^2$  :

$$\begin{aligned} & \sum_{C \in (\cup_{i=1}^N \Xi^{a,l}) \cup (\cup_{i=1}^{N-1} \Xi^{b,l})} |u_2^{-,C}|^2 |D_2^{-,C}| + |u_1^{-,C}|^2 |D_1^{-,C}| + |u_2^{+,C}|^2 |D_2^{+,C}| \\ \leq & \sum_{C \in (\cup_{i=1}^N \Xi^{a,l}) \cup (\cup_{i=1}^{N-1} \Xi^{b,l})} [|u_0^{-,C}| \omega^{-,C} + |u_0^{+,C}| (1 - \omega^{-,C}) + |\alpha_1^C| \lambda_{1,2}^{-,C} + |\alpha_2^{-,C}|]^2 h^2 + \\ & \sum_{C \in (\cup_{i=1}^N \Xi^{a,l}) \cup (\cup_{i=1}^{N-1} \Xi^{b,l})} [|u_0^{-,C}| \lambda_{0,1}^{-,C} + |u_0^{+,C}| \lambda_{0,1}^{+,C} + |\alpha_1^C|]^2 h^2 \\ & \sum_{C \in (\cup_{i=1}^N \Xi^{a,l}) \cup (\cup_{i=1}^{N-1} \Xi^{b,l})} [|u_0^{+,C}| \omega^{+,C} + |u_0^{-,C}| (1 - \omega^{+,C}) + |\alpha_1^C| \lambda_{1,2}^{+,C} + |\alpha_2^{+,C}|]^2 h^2 \end{aligned}$$

Avec  $(a + b + c + d)^2 \leq 4(a^2 + b^2 + c^2 + d^2)$  :

$$\begin{aligned} &\leq 12 \sum_{C \in (\cup_{i=1}^N \Xi^{a,l}) \cup (\cup_{i=1}^{N-1} \Xi^{b,l})} \left[ |u_0^{-,C}|^2 + |u_0^{+,C}|^2 + \frac{h^2}{\|\beta\|^2 \hat{\sigma}^2} \left( |f_2^{-,C}|^2 + |f_1^C|^2 + |f_2^{+,C}|^2 \right) \right] h^2 \\ &\quad \text{Avec le lemme 16.1, il on a} \\ &\leq \frac{12}{\|\beta\| c_{\hat{\theta}}} \sum_{C \in (\cup_{i=1}^N \Xi^{a,l}) \cup (\cup_{i=1}^{N-1} \Xi^{b,l})} \left( |u_0^{-,C}|^2 h \hat{\theta}_0^{-,C} + |u_0^{+,C}|^2 h \hat{\theta}_0^{+,C} \right) \\ &\quad + \frac{12}{\|\beta\|^2 \hat{\sigma}^2} h^2 \sum_{i \in J \setminus J^-} |f_h(Q_i)| h^2 \leq \frac{12}{\|\beta\| c_{\hat{\theta}}} K_N^{L^2} + \frac{12}{\|\beta\|^2 \hat{\sigma}^2} h^2 \sum_{i \in J \setminus J^-} |f_h(Q_i)|^2 h^2 \end{aligned}$$

### Conclusion

**Estimation de  $\|u_h\|_{L_h^1(\Omega)}$  :**

$$\begin{aligned} \|u_h\|_{L_h^1(\Omega)} &= \sum_{i \in J \setminus J^-} |u_i| |D_i| + \sum_{i \in J^-} |u_i| |D_i| \\ &\leq 4 \frac{1}{\|\beta\| c_{\hat{\theta}}} K_N^{L^1} + 3 \frac{1}{\|\beta\| \hat{\sigma}} h \sum_{i \in J \setminus J^-} |f_h(Q_i)| h^2 + \sum_{i \in J^-} |g_i| |D_i| \\ &\leq 4 \frac{1}{\|\beta\| c_{\hat{\theta}}} \left[ 4(Nh) \frac{1}{\hat{\sigma}} \sum_{i \in J \setminus J^-} |f_i| h^2 + 4(hN) \|\beta\| \sum_{i \in J^-} |g_i| h \right] + \\ &\quad 3 \frac{1}{\|\beta\| \hat{\sigma}} h \sum_{i \in J \setminus J^-} |f_h(Q_i)| h^2 + \sum_{i \in J^-} |g_i| |D_i| \end{aligned}$$

Maintenant  $h = O(1/N)$ , et avec les hypothèses de régularité et de quasi-uniformité sur le maillage, on peut conclure à l'existence de constantes  $C > 0$  et  $h_0 > 0$ , qui ne dépendent pas de  $h$  et  $\|\beta\|$ , telles que pour  $h \in ]0; h_0[$  :

$$\|u_h\|_{L_h^1(\Omega)} \leq C \left[ \frac{1}{\|\beta\|} \|f_h\|_{L_h^1(\Omega)} + \|g_h\|_{L_h^1(\Gamma^-)} \right]$$

**Estimation de  $\|u_h\|_{L_h^2(\Omega)}$  :**

$$\begin{aligned} \|u_h\|_{L_h^2(\Omega)}^2 &= \sum_{i \in J \setminus J^-} |u_i|^2 |D_i| + \sum_{i \in J^-} |u_i|^2 |D_i| \\ &\leq \frac{13}{\|\beta\| c_{\hat{\theta}}} K_N^{L^2} + \frac{12}{\|\beta\|^2 \hat{\sigma}^2} h^2 \sum_{i \in J \setminus J^-} |f_h(Q_i)|^2 h^2 + \sum_{i \in J^-} |g_i|^2 |D_i| \\ &\leq \frac{13}{\|\beta\| c_{\hat{\theta}}} e^{\frac{6Nh}{\|\beta\| c_{\hat{\theta}}}} \left[ K_1^{L^2} + p(h) \frac{\|\beta\| c_{\hat{\theta}}}{6} \right] + \\ &\quad \frac{12}{\|\beta\|^2 \hat{\sigma}^2} h^2 \sum_{i \in J \setminus J^-} |f_h(Q_i)|^2 h^2 + \sum_{i \in J^-} |g_i|^2 |D_i| \end{aligned}$$

Maintenant  $h = O(1/N)$ , et avec les hypothèses de régularité et de quasi-uniformité sur le maillage, on peut conclure à l'existence de constantes  $C > 0$  et  $h_0 > 0$ , qui ne

dépendent pas de  $h$  et  $\|\beta\|$ , telles que pour  $h \in ]0; h_0[$  :

$$\|u_h\|_{L_h^2(\Omega)}^2 \leq C e^{\frac{C}{\|\beta\|}} \left[ \left(1 + \frac{h_0}{\|\beta\|}\right)^2 \|f_h\|_{L_h^2(\Omega)}^2 + (\|\beta\| + h_0) \|g_h\|_{L_h^2(\Gamma^-)}^2 \right]$$

□

### 17.3 Majoration de $\|\beta \cdot \nabla_h u_h\|_{L^1(\Omega)}$

**Définition 17.3 (Condition  $\delta - S^-$  pour  $(\mathcal{T}_h)_h$ )** Soit  $(\mathcal{T}_h)_h$  une famille de maillage. On dira que la condition  $\delta - S^-$  est satisfaite pour cette famille, si il existe une constante  $\delta > 0$  indépendante de  $h$  et de  $\beta$  telle que

$$\sum_{j \in S_{\beta}^-(i)} \theta_{j,i}^{\beta} [\beta \cdot (Q_i - Q_j)] \geq \delta \|\beta\|^2 |S_i|^2 \quad \forall i \in J \setminus J^- \text{ et } \beta \in \mathbb{R}^2 \quad (17.329)$$

Le résultat central de ce chapitre est :

**Théorème 17.5 (Majoration  $\|\beta \cdot \nabla_h u_h\|_{L^1(\Omega)}$ )** Soit  $(\mathcal{T}_h)$  une famille de maillage quasi-uniforme et régulière, soit  $\mathcal{G}_h$  le graphe associé à  $\mathcal{T}_h$ . Soit  $\mathcal{C}_h$  l'ensemble des cellules obtenues à partir de  $\mathcal{G}_h$ . On suppose que :

- $\mathcal{C}_h$  ne contient pas de point intérieur de type 0 – 1 ou 0 – 2 (donc  $\mathcal{G}_h$  ne contient pas de circuit élémentaire).
- Il existe  $K_{\Xi} > 0$ , ne dépendant pas de  $h$ , tel que pour le nombre de strates  $N_{\Xi}$  on ait  $N = N_{\Xi} \leq K_{\Xi} \frac{1}{h}$ ,
- Le bord entrant  $\Gamma^-$  est connexe.
- La condition  $\delta - S^-$  est satisfaite pour la famille de maillage  $(\mathcal{T}_h)_h$

Alors il existe une constante  $C > 0$  et  $h_0 > 0$ , ne dépendant pas de  $h$ , telle que pour la solution  $u_h \in Y_h$  du problème variationnel approché I (11.199), on a :

$$\underbrace{\|u_h\|_{Y_{1,h}}}_{=\|\beta \cdot \nabla_h u_h\|_{L^1(\Omega)}} \leq C \left[ \|\beta\| \|g_h\|_{\text{Var}_h(\Gamma^-)} + \left(1 + \frac{1}{h}\right) \|f_h\|_{L_h^1(\Omega)} \right] \quad \forall h \in ]0; h_0[ \quad (17.330)$$

**Remarque 17.4** Pour les points intérieurs, c'est-à-dire  $Q_i$  avec  $i \in J_0$ , la condition  $\delta - S^-$  est équivalente à la condition  $\delta - S^+$  et ne nécessite donc pas d'étude particulière. En effet soit  $i \in J_0$  ( $Q_i$  n'est pas sur le bord), alors

$$j \in S_{\beta}^-(i) = \{k \in J : \theta_{k,i}^{\beta} \geq 0\} \iff j \in S_{-\beta}^+(i) = \{k \in J : \underbrace{\theta_{k,i}^{-\beta}}_{=-\theta_{k,i}^{\beta} = \theta_{i,k}^{\beta}} \leq 0\}$$

Ainsi

$$\begin{aligned}
\sum_{j \in S_\beta^-(i)} \theta_{j,i}^\beta [\beta \cdot (Q_i - Q_j)] &= \sum_{j \in S_\beta^-(i)} \theta_{j,i}^\beta [(-\beta) \cdot (Q_j - Q_i)] \\
&= \sum_{j \in S_\beta^-(i)} -\theta_{i,j}^\beta [(-\beta) \cdot (Q_j - Q_i)] \\
&= \sum_{j \in S_\beta^-(i)} \theta_{i,j}^{-\beta} [(-\beta) \cdot (Q_j - Q_i)] \\
&= \sum_{j \in S_{-\beta}^+(i)} \theta_{i,j}^{-\beta} [(-\beta) \cdot (Q_j - Q_i)] \\
&\geq \delta \|\beta\|^2 |S_i|^2 (> 0) \text{ (condition } \delta - S^+ \text{ )}
\end{aligned}$$

Avant de procéder à la preuve, qui sera donnée ci-dessous, le théorème technique utilisant la formulation II est nécessaire :

**Définition 17.5 (Formulation variationnelle II du problème approché d'advection)**

Trouver  $u_h \in Y_h$  tel que

$$\hat{b}_h(u_h, v_h) = (f_h, v_h)_\Pi \quad \forall v_h \in Y_h^0 \quad (17.331)$$

avec  $u_h(Q_i) = g_h(Q_i)$  pour  $i \in J^-$  et

$$(f_h, v_h)_\Pi = \sum_{k \in J \setminus J^-} f_h(Q_k) v_h(Q_k) d_k \quad , \quad d_k = \frac{1}{\|\beta\|^2} \sum_{j \in S^-(k)} [\beta \cdot (Q_k - Q_j)] \theta_{j,k} \quad (17.332)$$

**Théorème 17.6 (Théorème technique pour la formulation II)** Soit  $(\mathcal{T}_h)$  une famille de maillage quasi-uniforme et régulière, soit  $\mathcal{G}_h$  le graphe associé à  $\mathcal{T}_h$ . Soit  $\mathcal{C}_h$  l'ensemble des cellules obtenues à partir de  $\mathcal{G}_h$ . On suppose que :

- $\mathcal{C}_h$  ne contient pas de point intérieur de type 0 – 1 ou 0 – 2 (donc  $\mathcal{G}_h$  ne contient pas de circuit élémentaire).
- Il existe  $K_\Xi > 0$ , ne dépendant pas de  $h$ , tel que pour le nombre de strates  $N_\Xi$  on ait  $N = N_\Xi \leq K_\Xi \frac{1}{h}$ ,

Alors pour la solution  $u_h \in Y_h$  du problème variationnel approché II (11.201), il existe une constante  $C > 0$  et  $h_0 > 0$  ne dépendants ni de  $h$ , ni de  $\|\beta\|$  tels que :

$$\begin{aligned}
|u_h|_{Y_{1,h}} &= \sum_{T \in \mathcal{T}_h} |\beta \cdot \nabla_h u_h|_T |T| \\
&\leq C \left[ \sum_{T \in \mathcal{T}_h} |T| \|\nabla_h f_h|_T\| + \|\beta\| \|g_h\|_{Var_h(\Gamma^-)} + \|f_h\|_{L_h^1(\Gamma^-)} + \|f_h\|_{L_h^1(\Omega)} \right] \quad \forall h \in ]0, h_0[
\end{aligned} \quad (17.333)$$

Avant de commencer la preuve, quelques résultats et remarques préliminaires :

Les arguments utilisés dans la preuve sont proches de ceux utilisés au chapitre précédent, mais les points de contrôle sont ici d'une toute autre nature :

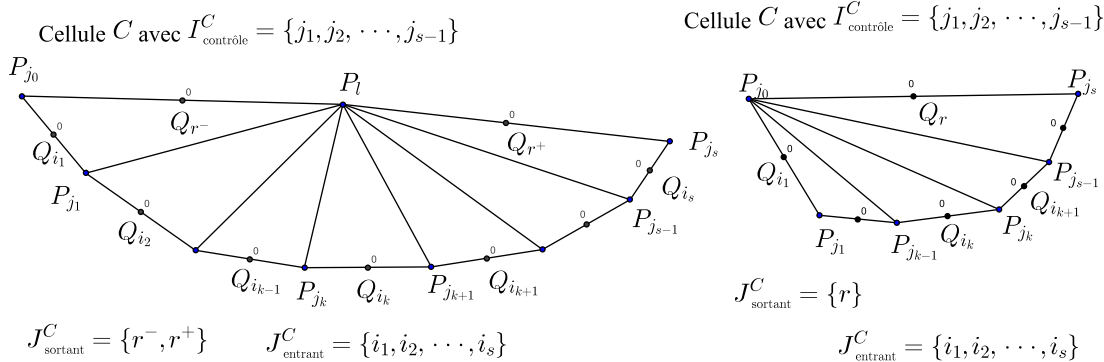
**Définition 17.6 (Points de contrôle  $I_{\text{contrôle}}^C$  de la cellule  $C$  et  $I_{\text{contrôle}}^\Xi$  de la strate  $\Xi$ )**  
 On appelle point de contrôle de la cellule  $C$ , un point  $P_j \in C$ ,  $j \in I$ , tel que  $P_j$  est le point commun de deux segments dont les milieux sont des points de  $J_{\text{entrant}}^C$ . L'ensemble des points de contrôle de  $C$  est noté  $I_{\text{contrôle}}^C$ .

Pour une strate  $\Xi$ , on définit

$$I_{\text{contrôle}}^\Xi = \cup_{C \in \Xi} I_{\text{contrôle}}^C$$

L'ensemble des points de contrôle de  $\Omega$  est noté  $I_{\text{contrôle}}^\Omega = \cup_{C \in \mathcal{C}_h} I_{\text{contrôle}}^C$ .

L'ensemble des points de contrôle sur  $\Gamma^-$  est noté  $I_{\text{contrôle}}^{\Gamma^-} = \{d \in I_{\text{contrôle}}^\Omega : P_q \in \Gamma^-\}$ .



**Motivation :** Pour une cellule  $C$ , les valeurs de  $u_h(Q_i)$ ,  $i \in J_{\text{entrant}}^C$  permettent de calculer complètement  $C$ . Pour le calcul de  $\beta \cdot \nabla_h u_T$ ,  $T \in C$ , la situation est différente : il suffit de connaître les différences  $u_h(Q_{i_{k+1}}) - u_h(Q_{i_k})$ ,  $k \in \{1, \dots, s - 1\}$  pour les points entrants de la cellule.

A la base de ce calcul, il y a la formule (en coordonnées locales upwind) :

$$|T| \beta \cdot \nabla_h u_{h|_T} = \underbrace{(\theta_{0,2}^T + \theta_{1,2}^T)}_{=\hat{\theta}_2^T} (u_2^T - u_1^T) + \underbrace{(\theta_{0,1}^T + \theta_{0,2}^T)}_{=\hat{\theta}_0^T} (u_1^T - u_0^T) (*)$$

La figure page 123 donne un exemple de maillage, pour lequel on retrouve des cellules avec

**deux points entrants et deux points sortants :** Par exemple la cellule définie par le polygone  $P_{11}P_6P_2P_7$ ,

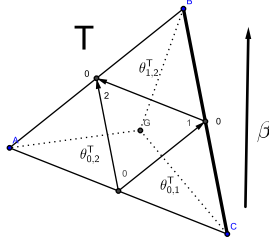
**trois points entrants et deux points sortants :** Par exemple la cellule définie par le polygone  $P_{18}P_{13}P_8P_9P_{14}$ ,

**deux points entrants et un point sortant :** Par exemple la cellule définie par le polygone  $T = P_6P_1P_2$  (ici  $Q_1 = Q_x^T$ , car l'arc orienté  $(Q_5, Q_1)$  a été retiré),

**un point entrant et deux points sortants :** Par exemple la cellule définie par les triangles  $P_{21}P_{17}P_{22}$  ou encore  $P_8P_4P_9$ . Ces cas semblent, à première vue, rendre problématique la définition des points de contrôle ci-dessus, car deux points entrants (au minimum) sont nécessaires pour qu'elle ait un sens. Ce paradoxe apparent est réglé par le fait que ces cellules, qui ne comportent qu'un triangle  $T$ ,

n'ont pas de point de contrôle. Pour ces cellules,  $|T| \beta \cdot \nabla_h u_{h|_T}$  se calcule indépendamment de la valeur de  $u_0^T$  et les différences  $u_2^T - u_1^T$ ,  $u_2^T - u_1^T$  et  $u_1^T - u_0^T$  ne dépendent que de  $f_2^T$  et  $f_1^T$  :

**Exemple 17.7 (Calcul d'une cellule sans point de contrôle)**



Une cellule qui ne contient pas de point de contrôle n'a qu'un seul point entrant. Cela exclut les cellules contenant un triangle avec deux coordonnées locales 0 et les cellules ayant un point intérieur de type 1 - 1. Les cellules sans point de contrôle sont donc uniquement des cellules, sans point intérieur de type 1 - 1, et n'ayant qu'un triangle  $T$ .

On a :

$$Q_0^T \text{ est de type } 0 - 1 \text{ ou } 0 - 2, \text{ donc } \hat{\theta}_0^T = \int_{S_0^T} |(\beta \cdot n)| d\sigma = \theta_{0,2}^T + \theta_{0,1}^T.$$

$$Q_2^T \text{ est de type } 0 - 2, \text{ donc } \hat{\theta}_2^T = \int_{S_2^T} |(\beta \cdot n)| d\sigma = \theta_{0,2}^T + \theta_{1,2}^T \text{ et } \bar{\theta}_2^T = \theta_{0,2}^T + \theta_{1,2}^T.$$

$$Q_1^T \text{ est de type } 1 - 0, \text{ donc } \hat{\theta}_1^T = \int_{S_1^T} |(\beta \cdot n)| d\sigma = \hat{\theta}_0^T - \hat{\theta}_2^T \geq 0 \text{ et } \bar{\theta}_1^T = \hat{\theta}_1^T + \theta_{1,2}^T = \theta_{0,1}^T.$$

	Formulation I	Formulation II
Ainsi :	$u_1^T \bar{\theta}_1^T - u_0^T \theta_{0,1}^T =  D_1^T  f_1^T$	$u_1^T \bar{\theta}_1^T - u_0^T \theta_{0,1}^T = d_1^T f_1^T$
	$u_2^T \bar{\theta}_2^T - u_0^T \theta_{0,2}^T - u_1^T \theta_{1,2}^T =  D_2^T  f_2^T$	$u_2^T \bar{\theta}_2^T - u_0^T \theta_{0,2}^T - u_1^T \theta_{1,2}^T = d_2^T f_2^T$

Prenons, par exemple, la formulation II (le raisonnement est parfaitement identique pour la formulation I), on obtient, en sommant ces deux équations :

$$u_1^T \bar{\theta}_1^T + u_2^T \bar{\theta}_2^T - u_0^T \theta_{0,1}^T - u_0^T \theta_{0,2}^T - u_1^T \theta_{1,2}^T = d_1^T f_1^T + d_2^T f_2^T$$

$$u_1^T (\bar{\theta}_1^T - \theta_{1,2}^T) + u_2^T \bar{\theta}_2^T - u_0^T (\theta_{0,1}^T + \theta_{0,2}^T) = d_1^T f_1^T + d_2^T f_2^T$$

$$\text{Avec } \hat{\theta}_0^T - \hat{\theta}_2^T = \hat{\theta}_1^T = \bar{\theta}_1^T - \theta_{1,2}^T :$$

$$-u_1^T (\hat{\theta}_2^T - \hat{\theta}_0^T) + u_2^T \bar{\theta}_2^T - u_0^T \hat{\theta}_0^T = d_1^T f_1^T + d_2^T f_2^T$$

Ainsi, avec (\*) ci-dessus :

$$|T| \beta \cdot \nabla_h u_{h|_T} = (u_2^T - u_1^T) \hat{\theta}_2^T + (u_1^T - u_0^T) \hat{\theta}_0^T = d_1^T f_1^T + d_2^T f_2^T$$

Conclusion :

Formulation I	Formulation II
$ T  \beta \cdot \nabla_h u_{h _T} =  D_1^T  f_1^T +  D_2^T  f_2^T$	$ T  \beta \cdot \nabla_h u_{h _T} = d_1^T f_1^T + d_2^T f_2^T$

(17.334)

$|T| \beta \cdot \nabla_h u_{h|_T}$  ne dépend donc pas de la valeur du point entrant  $u_0^T$ , mais seulement de  $f_h(Q_1^T)$  et  $f_h(Q_2^T)$ .

$$u_1^T = u_0^T \lambda_{0,1}^T + \alpha_1^T \text{ et } u_2^T = u_0^T \lambda_{0,2}^T + u_1^T \lambda_{1,2}^T + \alpha_2^T$$

Formules de récurrences : Soit  $\alpha_i \in \{\alpha_i^I, \alpha_i^{II}\}$ , alors

pour  $Q_1^T$  :  $\bar{\theta}_{0,1}^T = \bar{\theta}_1^T (= \theta_{1,2}^T)$ , donc  $\lambda_{0,1}^T = 1$ ,

pour  $Q_2^T$  :  $\lambda_{0,2}^T + \lambda_{1,2}^T = 1$ ,

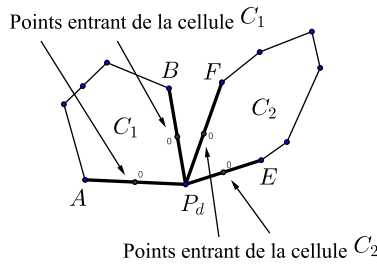
Ainsi

$$u_1^T - u_0^T = \alpha_1^T \quad , \quad u_2^T - u_1^T = \alpha_2^T - \lambda_{0,2}^T \alpha_1^T \text{ et } u_2^T - u_0^T = \alpha_2^T + \lambda_{1,2}^T \alpha_1^T \quad (17.335)$$

Pour la suite, il est nécessaire de s'assurer que deux cellules différentes ne peuvent pas avoir un point de contrôle en commun :

**Lemme 17.7 (unicité des points de contrôle)** *Quelque soit  $d \in I_{\text{contrôle}}^\Omega$ , il existe exactement une cellule  $C \in \mathcal{C}_h$ , telle que  $d \in I_{\text{contrôle}}^C$ .*

**Preuve :** Supposons que l'affirmation soit fautive. Il existerait alors deux cellules  $C_1$  et  $C_2$ , telles que  $P_d$  soit un point de contrôle de  $C_1$  et  $C_2$ . La figure ci-dessous illustre la situation :



Comme les points d'entrées des cellules sont de type  $0-1$  ou  $0-2$ ,  $\beta$  doivent être dans le cône de directions  $\widehat{AP_dB}$  et  $\widehat{FP_dE}$ . Or l'intersection de ces deux cônes est vide!

□

Le théorème technique qui doit être démontré s'appuie sur la formulation II du problème d'advection stationnaire. Afin de pouvoir exploiter cette formulation on introduit

**Définition 17.8 (Fonctions auxiliaires  $w_h(Q_i, \kappa)$  et  $\mathcal{L}(Q_a, Q_b, \kappa)$ )** *Soit  $u_h \in Y_h$  la solution du problème variationnel II, alors pour  $i \in J$  et  $\kappa \in \mathbb{R}$  on définit*

$$w_h(Q_i, \kappa) = u_h(Q_i) - \frac{\kappa}{\|\beta\|^2} (\beta \cdot Q_i)$$

et

$$\mathcal{L}(Q_a, Q_b, \kappa) = h |w_h(Q_a, \kappa) - w_h(Q_b, \kappa)| \quad \forall a, b \in J$$

On est maintenant en mesure de définir les quantités qu'il faut majorer pour pouvoir obtenir une estimation de  $\sum_{T \in \mathcal{T}_h} |T| |\beta \cdot \nabla_h u_{h|_T}|$ .

**Définition 17.9 (Valeur de contrôle  $L_d$  et fonction de contrôle  $\underline{\mathcal{L}}(P_d, \kappa)$ )**

Soit  $P_d$ ,  $d \in I_{\text{contrôle}}^\Omega$ , un point de contrôle et  $C \in \mathcal{C}_h$  la cellule qui contient ce point (donc  $d \in I_{\text{contrôle}}^C$ ), alors on définit la valeur de contrôle<sup>55</sup>

$$\begin{aligned} L_d &= \mathcal{L}(Q_a, Q_b, f_h^C) \text{ avec } a, b \in J_{\text{entrant}}^C \text{ tel que } P_d \in S_a \cap S_b \text{ et} \\ f_h^C &= f_h(Q_j) \text{ avec } j \in J^C \text{ tel que } |f_h(Q_j)| = \min_{r \in J^C} |f_h(Q_r)| \end{aligned} \quad (17.336)$$

et la fonction de contrôle

$$\underline{\mathcal{L}}(P_d, \kappa) = \mathcal{L}(Q_a, Q_b, \kappa) \quad \forall \kappa \in \mathbb{R}$$

Il est essentiel de garder la même valeur  $\kappa = f^C$  pour tous les points de contrôle d'une même cellule. Comme un point de contrôle ne peut appartenir qu'à une cellule, à chaque point de contrôle est donc associé une unique valeur de contrôle. Bien sur  $L_d = \underline{\mathcal{L}}(P_d, f_h^C)$ .

**Lemme 17.8 (Propriétés de la fonction auxiliaire) a)** Pour  $a, b, c \in J$ , on a

$$\mathcal{L}(Q_a, Q_c, \kappa) \leq \mathcal{L}(Q_a, Q_b, \kappa) + \mathcal{L}(Q_b, Q_c, \kappa) \text{ (inégalité triangulaire)}$$

b) Pour  $\kappa_1, \kappa_2 \in \mathbb{R}$  et  $a, b \in J$  on a :

$$\mathcal{L}(Q_a, Q_b, \kappa_1) \leq \mathcal{L}(Q_a, Q_b, \kappa_2) + h^2 \frac{|\kappa_2 - \kappa_1|}{\|\beta\|}$$

c) En particulier pour  $d \in I_{\text{contrôle}}^C$  :

$$L_d - h^2 \frac{|f_h^C - \kappa|}{\|\beta\|} \leq \underline{\mathcal{L}}(P_d, \kappa) \leq L_d + h^2 \frac{|f_h^C - \kappa|}{\|\beta\|} \quad (17.337)$$

**Preuve :** Pour a) il n'y a rien à montrer. Pour b) :

$$\begin{aligned} & \mathcal{L}(Q_a, Q_b, \kappa_2) \\ &= h \left| u_h(Q_b) - \frac{\kappa_2}{\|\beta\|^2} (\beta \cdot Q_b) - [u_h(Q_a) - \frac{\kappa_2}{\|\beta\|^2} (\beta \cdot Q_a)] \right| \\ &= h \left| u_h(Q_b) - \frac{\kappa_1 + (\kappa_2 - \kappa_1)}{\|\beta\|^2} (\beta \cdot Q_b) - [u_h(Q_a) - \frac{\kappa_1 + (\kappa_2 - \kappa_1)}{\|\beta\|^2} (\beta \cdot Q_a)] \right| \\ &= h \left| u_h(Q_b) - \frac{\kappa_1}{\|\beta\|^2} (\beta \cdot Q_b) - [u_h(Q_a) - \frac{\kappa_1}{\|\beta\|^2} (\beta \cdot Q_a)] + \frac{(\kappa_2 - \kappa_1)}{\|\beta\|^2} [\beta \cdot (Q_a - Q_b)] \right| \\ &\leq \mathcal{L}(Q_a, Q_b, \kappa_1) + h^2 \frac{|\kappa_2 - \kappa_1|}{\|\beta\|} \end{aligned} \quad (17.338)$$

□

Soit  $u_h \in Y_h$  la solution du problème approché (11.201), alors

$$u_h(Q_i) = \sum_{j \in S^-(i)} u_h(Q_j) \lambda_{j,i} + \frac{f_h(Q_i)}{\|\beta\|^2} \sum_{j \in S^-(i)} \beta \cdot (Q_i - Q_j) \lambda_{j,i} \quad (17.339)$$

Presque la même relation peut être obtenue pour  $w_h(Q_i, \kappa)$  :

55.  $Q_a$  et  $Q_b$  sont donc les milieux, respectivement, des segments  $S_a$  et  $S_b$



**Lemme 17.9** Soit  $\kappa \in \mathbb{R}$  et  $w_h(Q_l, \kappa) = u_h(Q_l) - \kappa \frac{1}{\|\beta\|^2}(\beta \cdot Q_l)$  pour  $l \in J$ , alors

$$w_h(Q_i, \kappa) = \sum_{j \in S^-(i)} w(Q_j, \kappa) \lambda_{j,i} + \alpha_h(Q_i, \kappa) \quad (17.340)$$

$$\text{avec } \alpha_h(Q_i, \kappa) = \frac{f_h(Q_i) - \kappa}{\|\beta\|^2} \sum_{j \in S^-(i)} \beta \cdot (Q_i - Q_j) \lambda_{j,i} \text{ et } |\alpha_h(Q_i, \kappa)| \leq \frac{h}{\|\beta\|} |f_h(Q_i) - \kappa| \quad (17.341)$$

**Preuve :**

$$\begin{aligned} w_h(Q_i, \kappa) &= u_h(Q_i) - \kappa \frac{1}{\|\beta\|^2}(\beta \cdot Q_i) \\ &= \sum_{j \in S^-(i)} u_h(Q_j) \lambda_{j,i} - \kappa \frac{1}{\|\beta\|^2}(\beta \cdot Q_i) + \frac{f_h(Q_i)}{\|\beta\|^2} \sum_{j \in S^-(i)} \beta \cdot (Q_i - Q_j) \lambda_{j,i} \\ &\quad \text{Maintenant } \sum_{j \in S^-(i)} \lambda_{j,i} = 1 \text{ donc} \\ w_h(Q_i, \kappa) &= \sum_{j \in S^-(i)} [u_h(Q_j) - \kappa \frac{1}{\|\beta\|^2}(\beta \cdot Q_j)] \lambda_{j,i} - \sum_{j \in S^-(i)} \kappa \frac{1}{\|\beta\|^2} \beta \cdot (Q_i - Q_j) \lambda_{j,i} \\ &\quad + \frac{f_h(Q_i)}{\|\beta\|^2} \sum_{j \in S^-(i)} \beta \cdot (Q_i - Q_j) \lambda_{j,i} \\ w_h(Q_i, \kappa) &= \sum_{j \in S^-(i)} w(Q_j, \kappa) \lambda_{j,i} + \frac{f_h(Q_i) - \kappa}{\|\beta\|^2} \sum_{j \in S^-(i)} \beta \cdot (Q_i - Q_j) \lambda_{j,i} \end{aligned}$$

□

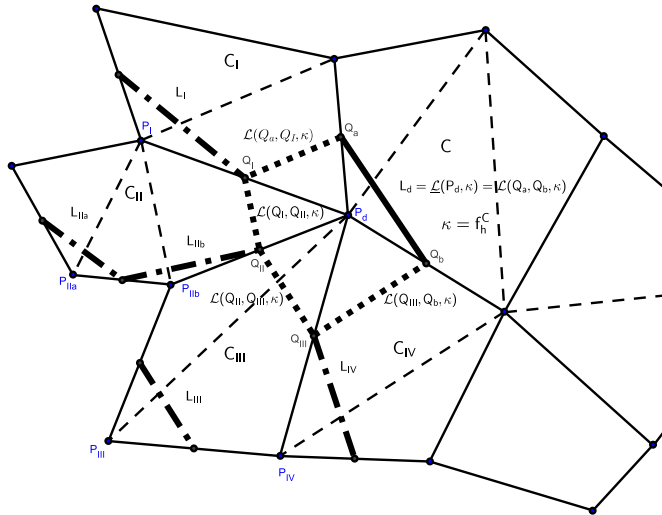
**Toutes les formules explicites de calcul des cellules du chapitre 16 s'appliquent donc aussi pour  $(w_h(Q_l, \kappa))_{l \in J}$  si on remplace le symbole  $\alpha_i$  par  $\alpha_h(Q_i, \kappa)$ .**

La valeur de  $\kappa$ , si elle varie, sera néanmoins toujours identique pour tous les points  $J^C$  d'une même cellule. C'est sous cette condition seulement qu'il est possible d'exploiter la formule (17.340).

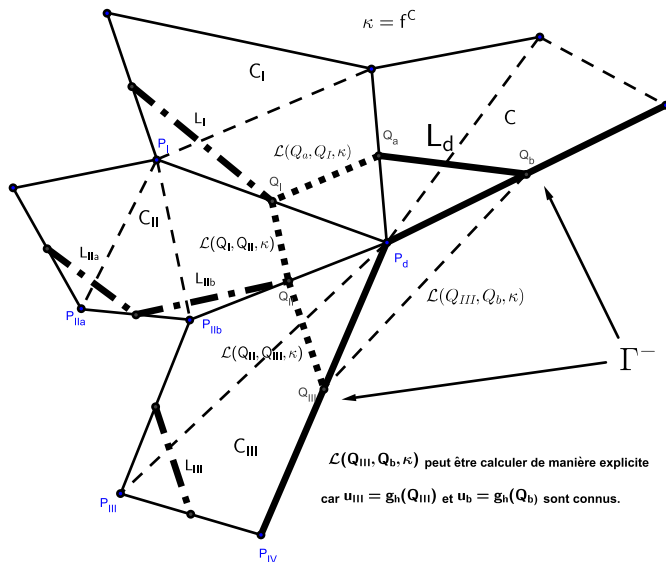
**Idées principales :** La figure suivante permet de comprendre la suite des opérations. Plusieurs cellules y sont représentées, dont les quatre cellules  $C_I$ ,  $C_{II}$ ,  $C_{III}$  et  $C_{IV}$  de strates inférieures à celle de  $C$  et qui ont en commun le point  $P_d$ . L'idée principal pour majorer  $L_d$  se résume à utiliser l'inégalité triangulaire (pour  $\kappa = f_h^C$ ) :

$$\underline{\mathcal{L}}(P_d, \kappa) = \mathcal{L}(Q_a, Q_b, \kappa) \leq \mathcal{L}(Q_a, Q_I, \kappa) + \mathcal{L}(Q_I, Q_{II}, \kappa) + \mathcal{L}(Q_{II}, Q_{III}, \kappa) + \mathcal{L}(Q_{III}, Q_b, \kappa),$$

puis à majorer  $\mathcal{L}(Q_a, Q_I, \kappa)$  en utilisant la valeur de contrôle  $L_I$ ,  $\mathcal{L}(Q_I, Q_{II}, \kappa)$  en utilisant les valeurs de contrôles  $L_{IIa}$  et  $L_{IIb}$ ,  $\mathcal{L}(Q_{II}, Q_{III}, \kappa)$  la valeur de contrôle  $L_{III}$  et, enfin,  $\mathcal{L}(Q_{III}, Q_b, \kappa)$  par la valeur de contrôle  $L_{IV}$ .



La même opération peut se faire sur le bord :



$$\underline{\mathcal{L}}(P_d, \kappa) = \mathcal{L}(Q_a, Q_b, \kappa) \leq \mathcal{L}(Q_a, Q_I, \kappa) + \mathcal{L}(Q_I, Q_{II}, \kappa) + \mathcal{L}(Q_{II}, Q_{III}, \kappa) + \mathcal{L}(Q_{III}, Q_b, \kappa),$$

avec la différence que  $\mathcal{L}(Q_{III}, Q_b, \kappa)$  est connu, car  $u_{III} = g_h(Q_{III})$  et  $u_b = g_h(Q_b)$ .

L'inégalité triangulaire s'applique sur certains couples de points seulement (voir les segments en pointillés) qui vont être maintenant définis :

**Définition 17.10** (couples de points  $\varpi_+^C, \varpi_-^C, \varpi_0^C, \varpi_d^{\Gamma^-}$  et ensembles  $\Upsilon^{\Gamma^-}, \Upsilon^\Omega$  et  $\Upsilon$ )  
 Soit  $C \in \mathcal{C}_h$  une cellule, alors :

**Cellule contenant un triangle avec deux coordonnées locales 0** : Voir chapitre 16, page 130 :

$$\varpi_+^C = \{Q_2^k, Q_0^k\}, \varpi_-^C = \{Q_2^k, Q_{\#0}^0\} \text{ et } \varpi_0^C = \{Q_2^k, Q_2^k\} \text{ ( et donc } \mathcal{L}(Q_2^k, Q_2^k, \kappa) = 0), \tag{17.342}$$

**Cellule sans point intérieur de type 1 - 1** : Voir chapitre 16, page 134 :

$$\varpi_+^C = \{Q_2^k, Q_0^k\}, \varpi_-^C = \{Q_1^0, Q_0^0\} \text{ et } \varpi_0^C = \{Q_2^k, Q_1^0\} \tag{17.343}$$

Cellule avec un point intérieur de type 1 – 1 : Voir chapitre 16, page 135 :

$$\varpi_+^C = \{Q_2^k, Q_0^k\}, \varpi_-^C = \{Q_2^{-t}, Q_0^{-t}\} \text{ et } \varpi_0^C = \{Q_2^k, Q_2^{-t}\} \quad (17.344)$$

Cellule  $T$  sans point de contrôle : Voir ci-dessus :

$$\varpi_+^C = \{Q_2^T, Q_0^T\}, \varpi_-^C = \{Q_2^T, Q_1^T\} \text{ et } \varpi_0^C = \{Q_1^T, Q_0^T\} \quad (17.345)$$

On définit

$$\Upsilon^\emptyset = \cup_{\substack{T \in \mathcal{C}_h \\ \Gamma_{\text{contrôle}}^T = \emptyset}} \{\varpi_-^T, \varpi_+^T, \varpi_0^T\}$$

Point de contrôle  $P_d$  sur le bord entrant  $\Gamma^-$  : Soit  $P_d \in \Gamma^-$  le point commun des segments  $S_k \subset \Gamma^-$  et  $S_l \subset \Gamma^-$ , alors on définit

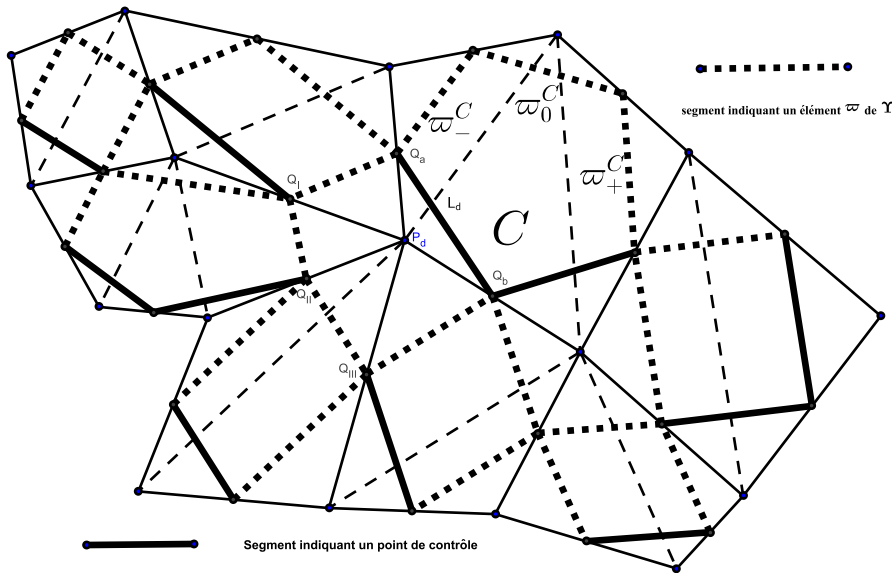
$$\varpi_d^{\Gamma^-} = \{Q_s, Q_k\} \text{ et } \Upsilon^{\Gamma^-} = \{\varpi_d^{\Gamma^-} : d \text{ est un point de contrôle sur } \Gamma^-\} \quad (17.346)$$

On pose

$$\Upsilon^\Omega = \cup_{C \in \mathcal{C}_h} \{\varpi_+^C, \varpi_-^C, \varpi_0^C\} \text{ et } \Upsilon = \Upsilon^\Omega \cup \Upsilon^{\Gamma^-}$$

Bien sur  $\Upsilon^\Omega \cap \Upsilon^{\Gamma^-} = \emptyset$ .

**Remarque 17.11 (exemple/notations/propriétés)** Dans la figure page 169, on a  $\varpi_+^{C_I} = \{Q_I, Q_a\}$ ,  $\varpi_+^{C_{II}} = \{Q_I, Q_{II}\}$ ,  $\varpi_0^{C_{III}} = \{Q_{II}, Q_{III}\}$  et  $\varpi_0^{C_{IV}} = \{Q_{III}, Q_b\}$ .



**Notations :** Pour  $\varpi = \{Q_k, Q_l\}$ , on utilisera la notation  $\mathcal{L}(\varpi, \kappa) = \mathcal{L}(Q_k, Q_l, \kappa)$  (ce qui est possible, car  $\mathcal{L}(Q_k, Q_l, \kappa) = \mathcal{L}(Q_l, Q_k, \kappa)$ ).

Autour de chaque point de contrôle  $P_d$ , il est possible d'obtenir une majoration de  $\mathcal{L}(\varpi, \kappa) = \mathcal{L}(Q_a, Q_b, \kappa)$  avec des inégalités triangulaires utilisant certains éléments  $\varpi$  de  $\Upsilon$ . On constate que cette décomposition existe toujours et qu'elle est unique. Il est en particulier impossible que deux segments indiquant un point de contrôle apparaissent dans cette décomposition, car cela contredit le fait qu'un point de contrôle ne peut pas appartenir à deux cellules différentes.

On note  $\odot_d \subset \Upsilon$  cette décomposition. Ainsi, dans l'exemple page 169

$$\odot_d = \{\varpi_+^{CI}, \varpi_+^{CII}, \varpi_0^{CIII}, \varpi_-^{CIV}\}$$

**Propriétés :**

a) On a l'inégalité : Pour  $d \in I_{\text{contrôle}}^\Omega$  quelconque

$$\underline{\mathcal{L}}(P_d, \kappa) \leq \sum_{\varpi \in \odot_d} \mathcal{L}(\varpi, \kappa) \quad (\text{inégalité triangulaire}) \quad (17.347)$$

Le choix de  $\varpi_-^C$ ,  $\varpi_+^C$  et  $\varpi_0^C$  (qui correspondent respectivement aux côtés "gauche", "droit" et "haut" des cellules, voir la figure ci-dessus), entraîne

$$\text{Pour } d, e \in I_{\text{contrôle}}^\Omega \text{ quelconques, si } d \neq e \text{ alors } \odot_d \cap \odot_e = \emptyset \quad (17.348)$$

Un élément  $\varpi \in \Upsilon$  ne peut pas apparaître dans les inégalités triangulaires de deux points de contrôle.

b) Enfin, si  $d \in I_{\text{contrôle}}^{\Xi_s}$  est un point de contrôle de la strate  $\Xi_s$ , alors pour  $\varpi \in \odot_d$  quelconque, on a  $\varpi \subset J^C$  où  $C \in \Xi_k$  avec  $k < s$ . Donc :

$$\text{Si } d \in I_{\text{contrôle}}^{\Xi_s} \text{ alors } \varpi \subset \cup_{k=1 \dots s-1} (\cup_{C \in J^C}) \text{ , } \forall \varpi \in \odot_d \quad (17.349)$$

Avant de passer aux estimations des fonctions  $\mathcal{L}(\varpi_-^C \kappa)$ ,  $\mathcal{L}(\varpi_+^C \kappa)$  et  $\mathcal{L}(\varpi_0^C \kappa)$  pour une cellule  $C \in \mathcal{C}_h$ , un petit lemme technique :

**Lemme 17.10 (formule)** Soit  $y_2 = \sum_{s=r_1}^{r_2} a_s x_s + \bar{a}$ ,  $y_1 = \sum_{s=r_1}^{r_2} b_s x_s + \bar{b}$ , avec  $\sum_{s=r_1}^{r_2} a_s = 1$ ,  $\sum_{s=r_1}^{r_2} b_s = 1$  et  $r_1 < r_2$ , alors :

$$y_1 - x_{r_1} = \sum_{l=r_1+1}^{r_2} \left( \sum_{s=l}^{r_2} b_s \right) (x_l - x_{l-1}) + \bar{b} \quad (17.350)$$

$$x_{r_2} - y_2 = \sum_{l=r_1+1}^{r_2} \left( \sum_{s=r_1}^{l-1} a_s \right) (x_l - x_{l-1}) - \bar{a} \quad (17.351)$$

$$x_{r_2} - x_{r_1} = \sum_{l=r_1+1}^{r_2} \left[ \sum_{s=r_1}^{l-1} a_s + \sum_{s=l}^{r_2} b_s - 1 \right] (x_l - x_{l-1}) + \bar{b} - \bar{a} \quad (17.352)$$

**Preuve :**

$$\begin{aligned}
y_1 - x_{r_1} &= \sum_{s=r_1}^{r_2} b_s x_s + \bar{b} - \sum_{s=r_1}^{r_2} b_s x_{r_1} \\
&= \sum_{s=r_1}^{r_2} b_s (x_s - x_{r_1}) + \bar{b} \\
&= \sum_{s=r_1}^{r_2} b_s \sum_{l=r_1+1}^s (x_l - x_{l-1}) + \bar{b} \\
&= \sum_{l=r_1+1}^{r_2} \left( \sum_{s=l}^{r_2} b_s \right) (x_l - x_{l-1}) + \bar{b}
\end{aligned}$$

$$\begin{aligned}
x_{r_2} - y_2 &= \sum_{s=r_1}^{r_2} a_s x_{r_2} - \sum_{s=r_1}^{r_2} a_s x_s - \bar{a} \\
&= \sum_{s=r_1}^{r_2-1} a_s (x_{r_2} - x_s) - \bar{a} \\
&= \sum_{s=r_1}^{r_2-1} a_s \sum_{l=s+1}^{r_2} (x_l - x_{l-1}) - \bar{a} \\
&= \sum_{l=r_1+1}^{r_2} \left( \sum_{s=r_1}^{l-1} a_s \right) (x_l - x_{l-1}) - \bar{a}
\end{aligned}$$

$$\begin{aligned}
y_1 - y_2 &= y_1 - x_{r_1} - \sum_{l=r_1+1}^{r_2} (x_l - x_{l-1}) + x_{r_2} - y_2 \\
&= \sum_{l=r_1+1}^{r_2} \left[ \sum_{s=r_1}^{l-1} b_s + \sum_{s=l}^{r_2} a_s - 1 \right] (x_l - x_{l-1}) + \bar{b} - \bar{a}
\end{aligned}$$

□

**Lemme 17.11 (Estimations des fonctions auxiliaires  $\mathcal{L}(\varpi_-^C \kappa)$ ,  $\mathcal{L}(\varpi_+^C \kappa)$  et  $\mathcal{L}(\varpi_0^C \kappa)$ )**

Soit  $I_{\text{contrôle}}^C$  l'ensemble des points de contrôle de la cellule  $C$ , alors

**Cellule contenant un triangle avec deux coordonnées locales 0 :** <sup>56</sup>

$$\mathcal{L}(\varpi_-^C \kappa) \leq \sum_{d \in I_{\text{contrôle}}^C} \eta_{-,d}^C \underline{\mathcal{L}}(P_d, \kappa) + \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j) - \kappa| \quad (17.353)$$

$$\mathcal{L}(\varpi_+^C \kappa) \leq \sum_{d \in I_{\text{contrôle}}^C} \eta_{+,d}^C \underline{\mathcal{L}}(P_d, \kappa) + \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j) - \kappa| \quad (17.354)$$

$$\mathcal{L}(\varpi_0^C \kappa) = 0 \quad (\text{donc } \eta_{0,d}^C = 0) \quad (17.355)$$

avec  $0 \leq \eta_{-,d}^C$ ,  $0 \leq \eta_{+,d}^C$  et  $\eta_{-,d}^C + \eta_{+,d}^C + \eta_{0,d}^C = 1$  pour  $d \in I_{\text{contrôle}}^C$ .

56. Cette cellule contient toujours au moins un point de contrôle

**Cellule avec point intérieur de type 1 – 1** <sup>57</sup> \ sans point intérieur de type 1 – 1 et avec au moins un point de contrôle <sup>58</sup> : Voir chapitre 16, page 134 et 16, page 135 :

$$\mathcal{L}(\varpi_-^C \kappa) \leq \sum_{d \in I_{\text{contrôle}}^C} \eta_{-,d}^C \underline{\mathcal{L}}(P_d, \kappa) + \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j) - \kappa| \quad (17.356)$$

$$\mathcal{L}(\varpi_+^C \kappa) \leq \sum_{d \in I_{\text{contrôle}}^C} \eta_{+,d}^C \underline{\mathcal{L}}(P_d, \kappa) + \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j) - \kappa| \quad (17.357)$$

$$\mathcal{L}(\varpi_0^C \kappa) \leq \sum_{d \in I_{\text{contrôle}}^C} \eta_{0,d}^C \underline{\mathcal{L}}(P_d, \kappa) + 2 \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j) - \kappa| \quad (17.358)$$

avec  $0 \leq \eta_{-,d}^C$ ,  $0 \leq \eta_{+,d}^C$ ,  $0 \leq \eta_{0,d}^C$  et  $\eta_{-,d}^C + \eta_{+,d}^C + \eta_{0,d}^C = 1$  pour  $d \in I_{\text{contrôle}}^C$ .

**Cellule sans point intérieur de type 1 – 1 et sans point de contrôle** <sup>59</sup>

$$\mathcal{L}(\varpi_+^T, \kappa) \leq \frac{h^2}{\|\beta\|} [\|f_2^T - \kappa\| + |f_1^T - f_2^T|] \quad (17.359)$$

$$\mathcal{L}(\varpi_-^T, \kappa) \leq \frac{h^2}{\|\beta\|} [\|f_2^T - \kappa\| + |f_1^T - f_2^T|] \text{ et } \mathcal{L}(\varpi_0^T, \kappa) \leq \frac{h^2}{\|\beta\|} |f_1^T - \kappa| \quad (17.360)$$

**Preuve :**

Posons  $w_i^{\kappa,r} = w_h(Q_i^r, \kappa)$ ,  $\alpha_i^{r,\kappa} = \alpha_h(Q_i, \kappa)$  pour  $i \in \{0, \neq 0, 1, 2, \}$  et  $\underline{\mathcal{L}}_r(\kappa) = \mathcal{L}(Q_0^r, Q_0^{r-1}, \kappa) = h|w_0^{\kappa,r} - w_0^{\kappa,r-1}|$ ,  $\underline{\mathcal{L}}_0(\kappa) = \mathcal{L}(Q_0^0, Q_{\neq 0}^0, \kappa) = h|w_0^{\kappa,0} - w_0^{\kappa,\neq 0}|$ . (Rappelons que les formules du chapitre 16 restent valables si on y remplace  $u_h(Q_r)$  par  $w_h(Q^r, \kappa)$  et  $\alpha_i$  par  $\alpha_h(Q_i, \kappa)$  pour  $\kappa \in \mathbb{R}$  quelconque (voir le lemme 17.9).

**Cellule contenant un triangle avec deux coordonnées locales 0 :** Posons

$\underline{\mathcal{L}}_r(\kappa) = h|w_0^{\kappa,r} - w_0^{\kappa,r-1}|$  pour  $r \in \{1, \dots, k\}$  et  $\underline{\mathcal{L}}_0(\kappa) = h|w_0^{\kappa,0} - w_{\neq 0}^{\kappa,0}|$ , alors avec (16.247) on a :

$$\underbrace{w_2^{\kappa,k}}_{y_2, y_1} = \sum_{r=1}^k \underbrace{w_0^{\kappa,r}}_{x_r} \omega^r \underbrace{\prod_{l=r+1}^k \lambda_{1,2}^l}_{a_r, b_r} + \underbrace{w_0^{\kappa,0}}_{x_0} \lambda_{0,2}^0 \underbrace{\prod_{l=1}^k \lambda_{1,2}^l}_{a_0, b_0} + \underbrace{w_{\neq 0}^{\kappa,0}}_{x_{-1}} \lambda_{1,2}^0 \underbrace{\prod_{l=1}^k \lambda_{1,2}^l}_{a_{-1}, b_{-1}} + \sum_{r=0}^k \alpha_2^{\kappa,r} \underbrace{\prod_{l=r+1}^k \lambda_{1,2}^l}_{\bar{a}, \bar{b}}$$

$r_1 = -1$ ,  $r_2 = k$ . L'égalité (16.248) signifie  $\sum_{s=r_1}^{r_2} a_s = 1$ .

Avec le lemme ci-dessus ( $a_s = b_s$ ,  $\bar{a} = \bar{b}$ ) :

$$\underbrace{w_0^{\kappa,k}}_{x_k} - \underbrace{w_2^{\kappa,k}}_{y_2} = \sum_{l=0}^k \left( \sum_{s=-1}^{l-1} a_s \right) (w_0^{\kappa,l} - w_0^{\kappa,l-1}) - \bar{a}$$

$$\underbrace{w_2^{\kappa,k}}_{y_1} - \underbrace{w_{\neq 0}^{\kappa,k}}_{x_{-1}} = \sum_{l=0}^k \left( \sum_{s=l}^k a_s \right) (w_0^{\kappa,l} - w_0^{\kappa,l-1}) + \bar{a}$$

57. Cette cellule contient toujours au moins un point de contrôle

58. Cette cellule contient au moins un point de contrôle si elle est la réunion de deux triangles ou plus

59. Cette cellule est faite d'un triangle uniquement

Multiplication par  $h$  et majoration (valeur absolue et inégalité triangulaire) :

$$\begin{aligned} \mathcal{L}(\varpi_-^C \kappa) &\leq \sum_{l=0}^k \underbrace{\left( \sum_{s=-1}^{l-1} a_s \right)}_{=\eta_{-,l}^C} \mathcal{L}_l(\kappa) + \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j) - \kappa| \\ \mathcal{L}(\varpi_+^C \kappa) &\leq \sum_{l=0}^k \underbrace{\left( \sum_{s=l}^k a_s \right)}_{=\eta_{+,l}^C} \mathcal{L}_l(\kappa) + \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j) - \kappa| \end{aligned}$$

et  $0 \leq \eta_{-,r}^C$ ,  $0 \leq \eta_{+,r}^C$  et  $\eta_{-,r}^C + \eta_{+,r}^C = 1$  pour  $r \in \{0, \dots, k\}$ .

Pour conclure, il suffit de remarquer que le point commun des segments contenant  $Q_0^r$  et  $Q_0^{r-1}$ , ainsi que  $Q_0^0$  et  $Q_{\mathbb{X}0}^0$  définit de façon bijective un point de contrôle de  $I_{\text{contrôle}}^C$ .

**Cellule sans point intérieur de type 1 – 1 et avec au moins deux triangles :**

Posons  $\underline{\mathcal{L}}_r(\kappa) = h|w_0^{\kappa,r} - w_0^{\kappa,r-1}|$  pour  $r \in \{1, \dots, k\}$  alors avec (16.260) et (16.261) on a :

$$\begin{aligned} \underbrace{w_1^{\kappa,0}}_{y_1} &= \underbrace{1}_{b_0} \times \underbrace{w_0^{\kappa,0}}_{x_0} + \underbrace{\alpha_{\kappa,1}^0}_{\bar{b}} \text{ et } a_1 = a_2 = \dots = a_k = 0 \\ \underbrace{w_2^{\kappa,k}}_{y_2} &= \sum_{r=0}^k \underbrace{w_0^{\kappa,r}}_{x_r} \omega^r \underbrace{\prod_{l=r+1}^k \lambda_{1,2}^l}_{a_r} + \sum_{r=0}^k \underbrace{\alpha_2^{\kappa,r}}_{\bar{a}} \underbrace{\prod_{l=r+1}^k \lambda_{1,2}^l}_{\bar{a}} \end{aligned}$$

$r_1 = 0$ ,  $r_2 = k$ . L'égalité (16.262) signifie  $\sum_{s=0}^k b_s = 1$ .  
Avec le lemme ci-dessus ( $a_s = b_s$ ,  $\bar{a} = \bar{b}$ ) :

$$\begin{aligned} \underbrace{w_1^{\kappa,0}}_{y_1} - \underbrace{w_0^{\kappa,0}}_{x_0} &= \sum_{l=1}^k \underbrace{\left( \sum_{s=l}^k b_s \right)}_{=0} (w_0^{\kappa,l} - w_0^{\kappa,l-1}) + \bar{b} \\ \underbrace{w_0^{\kappa,k}}_{x_k} - \underbrace{w_2^{\kappa,k}}_{y_2} &= \sum_{l=0}^k \left( \sum_{s=1}^{l-1} a_s \right) (w_0^{\kappa,l} - w_0^{\kappa,l-1}) - \bar{a} \\ \underbrace{w_1^{\kappa,0}}_{y_1} - \underbrace{w_0^{\kappa,k}}_{y_2} &= - \sum_{l=1}^k \underbrace{\left[ 1 - \sum_{s=l}^k b_s - \sum_{s=1}^{l-1} a_s \right]}_{\geq 0} (w_0^{\kappa,l} - w_0^{\kappa,l-1}) + \bar{b} - \bar{a} \end{aligned}$$

Multiplication par  $h$  et majoration (valeur absolue et inégalité triangulaire) :

$$\begin{aligned}\mathcal{L}(\varpi_{-}^C \kappa) &\leq \sum_{l=1}^k \underbrace{\left( \sum_{s=-1}^{l-1} b_s \right)}_{=\eta_{-,l}^C} \underline{\mathcal{L}}_l(\kappa) + \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j) - \kappa| \\ \mathcal{L}(\varpi_{+}^C \kappa) &\leq \sum_{l=1}^k \underbrace{\left( \sum_{s=l}^k a_s \right)}_{=\eta_{+,l}^C} \underline{\mathcal{L}}_l(\kappa) + \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j) - \kappa| \\ \mathcal{L}(\varpi_0^C \kappa) &\leq \sum_{l=1}^k \underbrace{\left[ 1 - 0 - \sum_{s=0}^{l-1} a_s \right]}_{=\eta_{0,l}^C} \underline{\mathcal{L}}_l(\kappa) + 2 \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j) - \kappa|\end{aligned}$$

et  $0 \leq \eta_{-,r}^C$ ,  $0 \leq \eta_{+,r}^C$ ,  $0 \leq \eta_{0,r}^C$  et  $\eta_{-,r}^C + \eta_{+,r}^C + \eta_{0,r}^C = 1$  pour  $r \in \{0 \dots, k\}$ .

Pour conclure, il suffit de remarquer que le point commun des segments contenant  $Q_0^r$  et  $Q_0^{r-1}$  définit de façon bijective un point de contrôle de  $I_{\text{contrôle}}^C$ .

**Cellule avec point intérieur de type 1 – 1 :** Posons  $\underline{\mathcal{L}}_r(\kappa) = h|w_0^{\kappa,r} - w_0^{\kappa,r-1}|$  pour  $r \in \{-t+1, \dots, k\}$  alors avec (16.275) et (16.277) on a :

$$\begin{aligned}\underbrace{w_2^{\kappa,k}}_{y_2} &= \sum_{r=0}^k \underbrace{w_0^{\kappa,r}}_{x_r} \omega^r \underbrace{\prod_{l=r+1}^k \lambda_{1,2}^l}_{a_r} + \underbrace{w_0^{\kappa,-1}}_{x_{-1}} \lambda_{0,1}^{-1} \underbrace{\prod_{l=0}^k \lambda_{1,2}^l}_{a_{-1}} \\ &+ \underbrace{\sum_{r=0}^k \alpha_2^{\kappa,r} \prod_{l=r+1}^k \lambda_{1,2}^l + \alpha_1^{\kappa,0} \prod_{l=0}^k \lambda_{1,2}^l}_{\bar{a}}\end{aligned}\quad (17.361)$$

$$1 = \sum_{r=0}^k \omega^r \prod_{l=r+1}^k \lambda_{1,2}^l + \lambda_{0,1}^{-1} \prod_{l=0}^k \lambda_{1,2}^l \quad (17.362)$$

$$\begin{aligned}\underbrace{w_2^{\kappa,-t}}_{y_1} &= \sum_{r=1}^t \underbrace{w_0^{\kappa,-r}}_{x_{-r}} \omega^{-r} \underbrace{\prod_{l=r+1}^t \lambda_{1,2}^{-l}}_{b_{-r}} + \underbrace{w_0^{\kappa,0}}_{x_0} \lambda_{0,1}^0 \underbrace{\prod_{l=1}^t \lambda_{1,2}^{-l}}_{b_0} \\ &+ \underbrace{\sum_{r=1}^t \alpha_2^{\kappa,-r} \prod_{l=r+1}^t \lambda_{1,2}^{-l} + \alpha_1^{\kappa,0} \prod_{l=1}^t \lambda_{1,2}^{-l}}_{\bar{b}}\end{aligned}\quad (17.363)$$

$$1 = \sum_{r=1}^t \omega^{-r} \prod_{l=r+1}^t \lambda_{1,2}^{-l} + \lambda_{0,1}^0 \prod_{l=1}^t \lambda_{1,2}^{-l} \quad (17.364)$$

$r_1 = -t$ ,  $r_2 = k$ . On pose  $a_{-2} = \dots = a_{-t} = 0$  et  $b_k = \dots = b_1 = 0$ .

Les égalités (17.362) et (17.364) signifient que  $\sum_{s=-1}^k a_s = 1$  et  $\sum_{s=-t}^0 b_s = 1$ .

Avec le lemme ci-dessus :



$$\begin{aligned}
\underbrace{w_2^{\kappa,-t}}_{y_1} - \underbrace{w_0^{\kappa,-t}}_{x_{-t}} &= \sum_{l=-t+1}^k \left( \sum_{s=l}^k b_s \right) (w_0^{\kappa,l} - w_0^{\kappa,l-1}) + \bar{b} \\
\underbrace{w_0^{\kappa,k}}_{x_k} - \underbrace{w_2^{\kappa,k}}_{y_2} &= \sum_{l=0}^k \left( \sum_{s=-t}^{l-1} a_s \right) (w_0^{\kappa,l} - w_0^{\kappa,l-1}) - \bar{a} \\
\underbrace{w_1^{\kappa,0}}_{y_1} - \underbrace{w_0^{\kappa,k}}_{y_k} &= - \sum_{l=-t+1}^k \left[ 1 - \sum_{s=l}^k b_s - \sum_{s=-t}^{l-1} a_s \right] (w_0^{\kappa,l} - w_0^{\kappa,l-1}) + \bar{b} - \bar{a}
\end{aligned}$$

Il reste à étudier  $\eta_{0,l}^C = 1 - \sum_{s=l}^k b_s - \sum_{s=-t}^{l-1} a_s$  :

- a) Si  $l \geq 1$  ou  $l \leq -1$ , alors  $\eta_{0,l}^C \geq 0$ , car  $a_{-2} = \dots = a_{-t} = 0$  et  $b_k = \dots = b_1 = 0$ .  
b) Si  $l = 0$ , alors

$$\begin{aligned}
\eta_{0,l}^C &= 1 - \sum_{s=0}^k b_s - \sum_{s=-t}^{-1} a_s = 1 - b_0 - a_{-1} \\
&= 1 - \lambda_{0,1}^0 \prod_{l=1}^t \lambda_{1,2}^{-l} - \lambda_{0,1}^{-1} \prod_{l=0}^k \lambda_{1,2}^l \\
&\quad \text{On a } \lambda_{0,1}^0 + \lambda_{0,1}^{-1} = 1 \text{ et } 0 \leq \lambda_{1,2}^l \leq 1, \text{ donc} \\
\eta_{0,l}^C &\geq 1 - (\lambda_{0,1}^0 + \lambda_{0,1}^{-1}) = 0
\end{aligned}$$

Multiplication par  $h$  et majoration (valeur absolue et inégalité triangulaire) :

$$\begin{aligned}
\mathcal{L}(\varpi_-^C \kappa) &\leq \sum_{l=1}^k \underbrace{\left( \sum_{s=-1}^{l-1} b_s \right)}_{=\eta_{-,l}^C} \mathcal{L}_l(\kappa) + \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j) - \kappa| \\
\mathcal{L}(\varpi_+^C \kappa) &\leq \sum_{l=1}^k \underbrace{\left( \sum_{s=l}^k a_s \right)}_{=\eta_{+,l}^C} \mathcal{L}_l(\kappa) + \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j) - \kappa| \\
\mathcal{L}(\varpi_0^C \kappa) &\leq \sum_{l=1}^k \underbrace{\left[ 1 - \sum_{s=0}^{l-1} b_s - \sum_{s=l}^k a_s \right]}_{=\eta_{0,l}^C} \mathcal{L}_l(\kappa) + 2 \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j) - \kappa|
\end{aligned}$$

et  $0 \leq \eta_{-,r}^C$ ,  $0 \leq \eta_{+,r}^C$ ,  $0 \leq \eta_{0,r}^C$  et  $\eta_{-,r}^C + \eta_{+,r}^C + \eta_{0,r}^C = 1$  pour  $r \in \{0 \dots, k\}$ .

Pour conclure, il suffit de remarquer que le point commun des segments contenant  $Q_0^r$  et  $Q_0^{r-1}$  définit de façon bijective un point de contrôle de  $I_{\text{contrôle}}^C$ .

**Cellule  $T$  sans point intérieur de type 1 – 1 et sans point de contrôle :**

Il faut exploiter les formules explicites (17.335) de l'exemple 17.7 :

$$u_1^T - u_0^T = \alpha_1^T, \quad u_2^T - u_1^T = \alpha_2^T - \lambda_{0,2}^T \alpha_1^T \text{ et } u_2^T - u_0^T = \alpha_2^T + \lambda_{1,2}^T \alpha_1^T$$

$$\varpi_+^T = \{Q_2^T, Q_0^T\}, \varpi_-^T = \{Q_2^T, Q_1^T\} \text{ et } \varpi_0^T = \{Q_1^T, Q_0^T\}$$

$$\begin{aligned} \mathcal{L}(\varpi_+^T, \kappa) &= \mathcal{L}(Q_2^T, Q_0^T, \kappa) = h|u_2^T - u_0^T - \kappa \frac{1}{\|\beta\|^2} \beta \cdot (Q_2^T - Q_0^T)| \\ &= h \left| f_2^T \frac{1}{\|\beta\|^2} [\lambda_{0,2}^T \beta \cdot (Q_2^T - Q_0^T) + \lambda_{1,2}^T \beta \cdot (Q_2^T - Q_1^T)] + \right. \\ &\quad \left. f_1^T \lambda_{1,2}^T \frac{1}{\|\beta\|^2} \beta \cdot (Q_1^T - Q_0^T) - \kappa \frac{1}{\|\beta\|^2} \beta \cdot (Q_2^T - Q_0^T) \right| \\ &= \frac{h}{\|\beta\|^2} |(f_2^T - \kappa) \beta \cdot (Q_2^T - Q_0^T) + \lambda_{1,2}^T (f_1^T - f_2^T) \beta \cdot (Q_1^T - Q_0^T)| \end{aligned}$$

car  $\lambda_{1,2}^T + \lambda_{0,2}^T = 1$ , donc  $(0 \leq \lambda_{1,2}^T \leq 1)$  :

$$\mathcal{L}(\varpi_+^T, \kappa) \leq \frac{h^2}{\|\beta\|} [|f_2^T - \kappa| + |f_1^T - f_2^T|]$$

Par un calcul analogue, on obtient :

$$\mathcal{L}(\varpi_-^T, \kappa) \leq \frac{h^2}{\|\beta\|} [|f_2^T - \kappa| + |f_1^T - f_2^T|] \text{ et } \mathcal{L}(\varpi_0^T, \kappa) \leq \frac{h^2}{\|\beta\|} |f_1^T - \kappa|$$

□

On est maintenant en mesure de majorer les valeurs de contrôle :

**Lemme 17.12 (Majoration des valeurs de contrôle)** Soit  $C \in \mathcal{C}_h$  et  $d \in I_{\text{contrôle}}^C \setminus I_{\text{contrôle}}^{\Xi_1}$ , alors

$$L_d \leq \sum_{\varpi \in \odot_d \cap (\Upsilon^\Omega \setminus \Upsilon^\emptyset)} \sum_{e \in I_{\text{contrôle}}^{\text{cell}(\varpi)}} L_e \eta_{\text{sign}(\varpi), e}^{\text{cell}(\varpi)} + \frac{h^3}{\|\beta\|} 3(1+K_{\text{cell}}) \frac{2\pi}{\theta_0} \mathcal{M}_d + \underbrace{\sum_{\varpi \in \odot_d \cap \Upsilon^{\Gamma^-}} \left[ h|\delta g_h|(\varpi) + \frac{h^2}{\|\beta\|} f_h^C \right]}_{=0 \text{ si } P_d \notin \Gamma^-} \quad (17.365)$$

où  $\text{sign}(\varpi) \in \{-, +, 0\}$  et  $\text{cell}(\varpi) \in \mathcal{C}_h$  représentent, respectivement, le signe affecté à  $\varpi$  selon sa position dans la cellule (voir la définition 17.10) et la cellule correspondant à  $\varpi$ .

On note  $\mathcal{M}_d = \sum_{T \in \cup_{C \in \mathcal{O}_d} C} \|\nabla_h f_h|_T\|$  la somme des gradients de tous les triangles des cellules qui ont le point de contrôle  $P_d$  en commun.

Cette réunion de cellules est notée  $\mathcal{O}_d$ .

On note  $|\delta g_h|(\varpi) = |g_h(Q_r) - g_h(Q_t)|$  où  $\varpi = \{Q_r, Q_t\}$ .

**Remarque :** Si  $d \in I_{\text{contrôle}}^{\Xi_1}$ , alors les valeurs de tous les points entrants  $J_{\text{entrant}}^{\Xi_1} \subset J^-$  de la première strate  $\Xi_1$  sont connues, car données par  $g_h$  et on a  $\odot_d \subset \Upsilon^{\Gamma^-}$ .

**Preuve :** Avec les lemmes 17.11 et 17.8 ( $\Upsilon^\theta$  est l'ensemble des éléments  $\varpi$  qui appartiennent à des cellules sans point de contrôle) :

$$\begin{aligned}
L_d &\leq \underline{\mathcal{L}}(P_d, f_h^C) \leq \sum_{\varpi \in \mathcal{O}_d} \mathcal{L}(\varpi, f_h^C) \\
&\leq \sum_{\varpi \in \mathcal{O}_d \cap (\Upsilon^\Omega \setminus \Upsilon^\theta)} \mathcal{L}(\varpi, f_h^C) + \sum_{\varpi \in \mathcal{O}_d \cap \Upsilon^\theta} \mathcal{L}(\varpi, f_h^C) + \sum_{\varpi \in \mathcal{O}_d \cap \Upsilon^{\Gamma^-}} \mathcal{L}(\varpi, f_h^C) \\
&\leq \sum_{\varpi \in \mathcal{O}_d \cap (\Upsilon^\Omega \setminus \Upsilon^\theta)} \sum_{e \in I_{\text{contrôle}}^{\text{cell}(\varpi)}} \mathcal{L}(P_e, f_h^C) \eta_{\text{sign}(\varpi), e}^{\text{cell}(\varpi)} + \frac{h^2}{\|\beta\|} \sum_{\varpi \in \mathcal{O}_d \cap (\Upsilon^\Omega \setminus \Upsilon^\theta)} \sum_{j \in J^{\text{cell}(\varpi)} \setminus J_{\text{entrant}}^{\text{cell}(\varpi)}} |f_h(Q_j) - f_h^C| \\
&\quad + \sum_{\varpi \in \mathcal{O}_d \cap \Upsilon^\theta} \mathcal{L}(\varpi, f_h^C) + \sum_{\varpi \in \mathcal{O}_d \cap \Upsilon^{\Gamma^-}} \mathcal{L}(\varpi, f_h^C) \\
&\leq \sum_{\varpi \in \mathcal{O}_d \cap (\Upsilon^\Omega \setminus \Upsilon^\theta)} \sum_{e \in I_{\text{contrôle}}^{\text{cell}(\varpi)}} \left[ L_e + \frac{h^2}{\|\beta\|} |f_h^{\text{cell}(\varpi)} - f_h^C| \right] \eta_{\text{sign}(\varpi), e}^{\text{cell}(\varpi)} \\
&\quad + \frac{h^2}{\|\beta\|} \sum_{\varpi \in \mathcal{O}_d \cap (\Upsilon^\Omega \setminus \Upsilon^\theta)} \sum_{j \in J^{\text{cell}(\varpi)} \setminus J_{\text{entrant}}^{\text{cell}(\varpi)}} |f_h(Q_j) - f_h^C| + \sum_{\varpi \in \mathcal{O}_d \cap \Upsilon^{\Gamma^-}} \mathcal{L}(\varpi, f_h^C) + \sum_{\varpi \in \mathcal{O}_d \cap \Upsilon^\theta} \mathcal{L}(\varpi, f_h^C) \\
0 &\leq \eta_{\text{sign}(\varpi), e}^{\text{cell}(\varpi)} \leq 1 \text{ (avec, pour } e \in I_{\text{contrôle}}^{\text{cell}(\varpi)} : \eta_{-, e}^{\text{cell}(\varpi)} + \eta_{-, e}^{\text{cell}(\varpi)} + \eta_{0, e}^{\text{cell}(\varpi)} = 1) \text{, donc}
\end{aligned}$$

$$L_d \leq \sum_{\varpi \in \mathcal{O}_d \cap (\Upsilon^\Omega \setminus \Upsilon^\theta)} \sum_{e \in I_{\text{contrôle}}^{\text{cell}(\varpi)}} L_e \eta_{\text{sign}(\varpi), e}^{\text{cell}(\varpi)} + \mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3 + \mathcal{A}_4$$

avec

$$\begin{aligned}
\mathcal{A}_1 &= \sum_{\varpi \in \mathcal{O}_d \cap (\Upsilon^\Omega \setminus \Upsilon^\theta)} \sum_{e \in I_{\text{contrôle}}^{\text{cell}(\varpi)}} \frac{h^2}{\|\beta\|} |f_h^{\text{cell}(\varpi)} - f_h^C| \\
\mathcal{A}_2 &= \frac{h^2}{\|\beta\|} \sum_{\varpi \in \mathcal{O}_d \cap (\Upsilon^\Omega \setminus \Upsilon^\theta)} \sum_{j \in J^{\text{cell}(\varpi)} \setminus J_{\text{entrant}}^{\text{cell}(\varpi)}} |f_h(Q_j) - f_h^C| \\
\mathcal{A}_3 &= \sum_{\varpi \in \mathcal{O}_d \cap \Upsilon^\theta} \mathcal{L}(\varpi, f_h^C) \\
\mathcal{A}_4 &= \sum_{\varpi \in \mathcal{O}_d \cap \Upsilon^{\Gamma^-}} \mathcal{L}(\varpi, f_h^C)
\end{aligned}$$

Notons  $\mathcal{O}_d$  la réunion des cellules qui ont le sommet  $P_d$  en commun.

Soient  $Q_r, Q_s \in \cup_{C \in \mathcal{O}_d} C$  deux points quelconques, alors  $Q_r$  et  $Q_s$  peuvent être reliés par un chemin fait de segments  $[Q_i, Q_j]$  pour lequel on a l'inégalité  $|f_h(Q_i) - f_h(Q_j)| \leq \|\nabla_h f_h|_T\| h$  pour  $Q_i, Q_j \in T$ . Donc

$$|f_h(Q_r) - f_h(Q_s)| \leq \underbrace{\sum_{T \in \cup_{C \in \mathcal{O}_d} C}}_{=\mathcal{M}_d} \|\nabla_h f_h|_T\| h \quad (*)$$

**Pour  $\mathcal{A}_1$  :** Maintenant le nombre de points de contrôle d'une cellule est inférieur ou égal au nombre de triangles qu'elle contient, donc  $\text{card}(I_{\text{contrôle}}^{\text{cell}(\varpi)}) \leq K_{\text{cell}}$  où  $K_{\text{cell}}$  est majorée par une constante qui ne dépend pas de  $h$  (voir le lemme 14.3), donc

$$\mathcal{A}_1 \leq K_{\text{cell}} \frac{h^2}{\|\beta\|} \sum_{\varpi \in \odot_d \cap (\Upsilon^\Omega \setminus \Upsilon^\emptyset)} |f_h^{\text{cell}(\varpi)} - f_h^C| \stackrel{(*)}{\leq} \frac{h^3}{\|\beta\|} K_{\text{cell}} \sum_{\varpi \in \odot_d \cap (\Upsilon^\Omega \setminus \Upsilon^\emptyset)} \mathcal{M}_d$$

Un élément  $\varpi \in \odot_d$  ne peut pas appartenir à deux cellules distinctes.  $\text{card}(\odot_d)$  est donc inférieur ou égal au nombre de cellules qui peuvent se rejoindre en  $P_d$ . Ce nombre reste fini, car les triangles de ces cellules qui ont  $P_d$  en commun, ont un angle minimal  $\theta_0$  (voir le lemme 16.1, page 128). Donc  $\text{card}(\odot_d) \leq \frac{2\pi}{\theta_0}$ . Ainsi

$$\mathcal{A}_1 \leq \frac{h^3}{\|\beta\|} K_{\text{cell}} \frac{2\pi}{\theta_0} \mathcal{M}_d$$

**Pour  $\mathcal{A}_2$  :** Le raisonnement est le même :  $\text{card}(J^{\text{cell}(\varpi)} \setminus J^{\text{cell}(\varpi)}) \leq 2K_{\text{cell}}$  (3 points par triangle moins le(s) point(s) entrant(s)) et on a

$$\mathcal{A}_2 \leq \frac{h^3}{\|\beta\|} 2K_{\text{cell}} \frac{2\pi}{\theta_0} \mathcal{M}_d$$

**Pour  $\mathcal{A}_3$  :** Si  $\varpi \in \Upsilon^\emptyset$ , alors la cellule  $\text{cell}(\varpi) = T$  se réduit à un triangle, et on a avec (17.359) et (17.360) :

$$\begin{aligned} \mathcal{L}(\varpi, f_h^C) &\leq \max\{\mathcal{L}(\varpi_+, f_h^C), \mathcal{L}(\varpi_-, f_h^C), \mathcal{L}(\varpi_0^T, f_h^C)\} \\ &\leq \frac{h^2}{\|\beta\|} [|f_2^T - f_h^C| + |f_1^T - f_2^T| + |f_1^T - f_h^C|] \\ &\leq 3 \frac{h^3}{\|\beta\|} \frac{2\pi}{\theta_0} \mathcal{M}_d \end{aligned}$$

**Pour  $\mathcal{A}_4$  :** Si  $P_d \in \Gamma^-$ , il ne peut avoir qu'un seul élément  $\varpi \in \odot_d \cap \Upsilon^\emptyset$ . Soit  $\varpi = \{Q_r, Q_t\}$  où  $Q_r$  et  $Q_t$  sont les milieux de deux segments  $S_r \subset \Gamma^-$ ,  $S_t \subset \Gamma^-$  ayant  $P_d$  en commun. Alors

$$\mathcal{A}_4 \leq h |g(Q_r) - g(Q_t)| + |f_h^C| \frac{|\beta \cdot (Q_r - Q_t)|}{\|\beta\|^2} \leq h \underbrace{|g(Q_r) - g(Q_t)|}_{=|\delta g|(\varpi)} + \frac{h^2}{\|\beta\|} |f_h^C|$$

Ainsi

$$L_d \leq \sum_{\varpi \in \odot_d \cap (\Upsilon^\Omega \setminus \Upsilon^\emptyset)} \sum_{e \in I_{\text{contrôle}}^{\text{cell}(\varpi)}} L_e \eta_{\text{sign}(\varpi), e} + \frac{h^3}{\|\beta\|^2} 3(1 + K_{\text{cell}}) \frac{2\pi}{\theta_0} \mathcal{M}_d + \sum_{\varpi \in \odot_d \cap \Upsilon^{\Gamma^-}} \left[ h |\delta g|(\varpi) + \frac{h^2}{\|\beta\|} |f_h^C| \right]$$

□

**Preuve du théorème technique :**

**Etape 1 : Estimation sur les points de contrôle** Soient  $\Xi_s$ ,  $s \in \{1, \dots, N\}$  l'ensemble des strates de  $\mathcal{C}_h$ .

L'objectif est de majorer la quantité  $\mathcal{K}_N$ , avec

$$\mathcal{K}_n = \sum_{d \in \cup_{k=1}^n I_{\text{contrôle}}^{\Xi_k}} L_d \text{ pour } n \in \{1, \dots, N\}$$

Pour  $n = 1$ ,  $J_{\text{entrant}}^{\Xi_1} \subset J^-$ , car les points entrants de la première strate ne dépendent pas d'une strate inférieure, mais des valeurs du bord entrant données par  $g_h$ . Donc

$$\mathcal{K}_1 \leq \sum_{d \in I_{\text{contrôle}}^{\Xi_1}} \sum_{\varpi \in \odot_d \cap \Upsilon^{\Gamma^-}} \left[ h |\delta g|(\varpi) + \frac{h^2}{\|\beta\|} |f_h^C| \right]$$

Pour cela une inégalité de récurrence sur  $n$  va être établie pour  $\mathcal{K}_n$ ,  $n \in \{2, \dots, N\}$

$$\begin{aligned} \mathcal{K}_n &= \sum_{d \in \cup_{k=1}^n I_{\text{contrôle}}^{\Xi_k}} L_d = \sum_{d \in I_{\text{contrôle}}^{\Xi_1}} L_d + \sum_{d \in \cup_{k=2}^n I_{\text{contrôle}}^{\Xi_k}} L_d \\ &\leq \mathcal{K}_1 + \underbrace{\sum_{d \in \cup_{k=2}^n I_{\text{contrôle}}^{\Xi_k}} \sum_{\varpi \in \odot_d \cap (\Upsilon^\Omega \setminus \Upsilon^\emptyset)} \sum_{e \in I_{\text{contrôle}}^{\text{cell}(\varpi)}} L_e \eta_{\text{sign}(\varpi), e}^{\text{cell}(\varpi)}}_{\mathcal{A}} + \\ &+ \sum_{d \in \cup_{k=2}^n I_{\text{contrôle}}^{\Xi_k}} \left[ \frac{h^3}{\|\beta\|} 3(1 + K_{\text{cell}}) \frac{2\pi}{\theta_0} \mathcal{M}_d + \sum_{\varpi \in \odot_d \cap \Upsilon^{\Gamma^-}} \left[ h |\delta g|(\varpi) + \frac{h^2}{\|\beta\|} |f_h^C| \right] \right] \end{aligned}$$

Pour le terme  $\mathcal{A}$ , considérons les ensembles :

$$\mathcal{Y} = \{(d, \varpi) : d \in \cup_{k=2}^n I_{\text{contrôle}}^{\Xi_k} \text{ et } \varpi \in \odot_d \cap (\Upsilon^\Omega \setminus \Upsilon^\emptyset)\}$$

et

$$\mathcal{Z} = \{(\text{cell}(\varpi), \text{sign}(\varpi)) : \varpi \in I_{\text{contrôle}}^{\cup_{k=1}^{n-1} \Xi_k} \setminus \Upsilon^\emptyset\}$$

Ces deux ensembles sont en bijection. En effet, le couple  $(d, \varpi)$  définit de façon unique, une cellule  $C \in \cup_{k=1}^{n-1} \Xi_k \setminus \Upsilon^\emptyset$  et un élément dans  $\{\varpi_-^C, \varpi_+^C, \varpi_0^C\}$ . La somme  $\mathcal{A}$  peut donc s'écrire<sup>60</sup> :

$$\begin{aligned} \mathcal{A} &= \sum_{\substack{C \in \cup_{k=1}^{n-1} \Xi_k \text{ contrôle} \\ \text{et } I_{\text{contrôle}}^C \neq \emptyset}} \sum_{\varpi \in \{\varpi_-^C, \varpi_+^C, \varpi_0^C\}} \sum_{e \in I_{\text{contrôle}}^C} L_e \eta_{\text{sign}(\varpi), e}^C \\ &= \sum_{\substack{C \in \cup_{k=1}^{n-1} \Xi_k \text{ contrôle} \\ \text{et } I_{\text{contrôle}}^C \neq \emptyset}} \sum_{e \in I_{\text{contrôle}}^C} \sum_{\varpi \in \{\varpi_-^C, \varpi_+^C, \varpi_0^C\}} L_e \eta_{\text{sign}(\varpi), e}^C \end{aligned}$$

60. Les inégalités triangulaires de deux points de contrôle  $d_1$  et  $d_2$  distincts ne peuvent pas partager un même élément  $\varpi : \odot_{d_1} \cap \odot_{d_2} = \emptyset$ . Voir les figures pages 170, 169 et 169.

Maintenant, quelque soit le point de contrôle  $e$ , on a  $\eta_{-,e}^C + \eta_{+,e}^C + \eta_{0,e}^C = 1$ , donc

$$\mathcal{A} = \sum_{\substack{C \in \cup_{k=1}^{n-1} \Xi_k^{\text{contrôle}} \\ \text{et } I_{\text{contrôle}}^C \neq \emptyset}} \sum_{e \in I_{\text{contrôle}}^C} L_e = \mathcal{K}_{n-1}$$

Ainsi

$$\begin{aligned} \mathcal{K}_n &\leq \mathcal{K}_{n-1} + \sum_{d \in \cup_{k=2}^N I_{\text{contrôle}}^{\Xi_k}} \frac{h^3}{\|\beta\|} 3(1 + K_{\text{cell}}) \frac{2\pi}{\theta_0} \mathcal{M}_d \\ &\quad + \sum_{d \in \cup_{k=1}^N I_{\text{contrôle}}^{\Xi_k}} \sum_{\varpi \in \odot_d \cap \Gamma^{\Gamma^-}} \left[ h|\delta g|(\varpi) + \frac{h^2}{\|\beta\|} |f_h^C| \right] \end{aligned}$$

$$\mathcal{K}_1 \leq \sum_{d \in \cup_{k=1}^N I_{\text{contrôle}}^{\Xi_k}} \sum_{\varpi \in \odot_d \cap \Gamma^{\Gamma^-}} \left[ h|\delta g|(\varpi) + \frac{h^2}{\|\beta\|} |f_h^C| \right], \text{ donc :}$$

$$\begin{aligned} \mathcal{K}_N &\leq (Nh) \frac{1}{\|\beta\|} 3(1 + K_{\text{cell}}) \frac{2\pi}{\theta_0} \underbrace{\sum_{d \in \cup_{k=2}^N I_{\text{contrôle}}^{\Xi_k}} h^2 \mathcal{M}_d}_{=\mathcal{G}_1} \\ &\quad + (Nh) \underbrace{\sum_{d \in \cup_{k=1}^N I_{\text{contrôle}}^{\Xi_k}} \sum_{\varpi \in \odot_d \cap \Gamma^{\Gamma^-}} |\delta g|(\varpi) + \frac{1}{\|\beta\|}}_{=\mathcal{G}_2} (Nh) \underbrace{\sum_{\varpi \in \odot_d \cap \Gamma^{\Gamma^-}} h|f_h^C|}_{\mathcal{G}_3} \end{aligned}$$

**Pour  $\mathcal{G}_1$  :** Dans la somme définie par  $\mathcal{G}_1$ , le terme  $\|\nabla_h f_{h|T}\|$ , pour  $T \in \mathcal{T}_h$  ne peut apparaître qu'autant de fois qu'il y a de points de contrôle de  $I_{\text{contrôle}}^\Omega$  sur les sommets du polygone définissant la cellule contenant  $T$ . Une cellule quelconque ayant au plus  $K_{\text{cell}} + 2$  sommets, on a donc  $(h^2 \leq \hat{c}_0|T|$  et  $N \leq \frac{1}{h}K_\Xi$ )

$$\mathcal{G}_1 \leq \hat{c}_0(K_{\text{cell}} + 2) \sum_{T \in \mathcal{T}_h} |T| \|\nabla_h f_{h|T}\|$$

**Pour  $\mathcal{G}_2$  :**

$$\mathcal{G}_2 = |g_h|_{\text{Var}_h(\Gamma^-)}$$

**Pour  $\mathcal{G}_3$  :** Soit  $S_i \subset \Gamma^-$ , alors  $\frac{h}{c_0\sigma} \leq |S_i|$  et, avec le choix de  $f_h^C$ , on a

$$\mathcal{G}_3 \leq c_0\sigma \sum_{b \in J^-} |f_h(Q_b)| |S_b| \leq c_0\sigma \|f_h\|_{L_h^1(\Gamma^-)}$$

**Conclusion :**

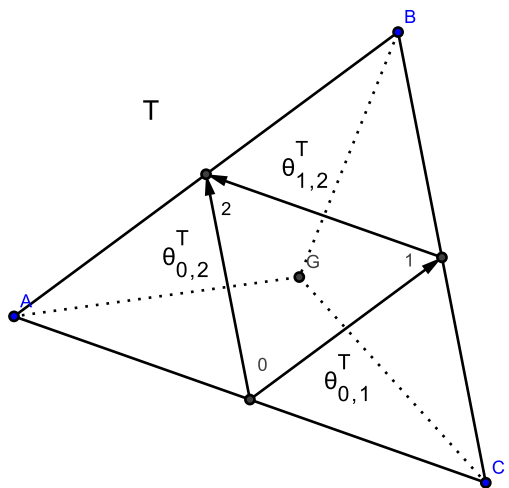
$$\mathcal{K}_N \leq \frac{1}{\|\beta\|} K_\Xi \frac{6\pi}{\theta_0} \hat{c}_0 (K_{\text{cell}} + 2)^2 \sum_{T \in \mathcal{T}_h} |T| \|\nabla_h f_{h|T}\| + K_\Xi |g_h|_{\text{Var}_h(\Gamma^-)} + \frac{1}{\|\beta\|} K_\Xi c_0\sigma \|f_h\|_{L_h^1(\Gamma^-)} \quad (17.366)$$

**Etape 2 : Estimation de  $|\beta \cdot \nabla_h u_{h|_T}| |T|$**

1) Soit  $C \in \mathcal{C}_h$  une cellule,  $d \in I_{\text{contrôle}}^C$  et  $a, b \in J_{\text{entrant}}^C$  avec  $S_a \cap S_b = \{P_d\}$ , alors

$$\begin{aligned}
 L_d &= h \left| u_h(Q_b) - f_h^C \frac{1}{\|\beta\|^2} (\beta \cdot Q_b) - [u_h(Q_a) - f_h^C \frac{1}{\|\beta\|^2} (\beta \cdot Q_a)] \right| \\
 &\geq h |u_h(Q_b) - u_h(Q_a)| - \frac{h^2}{\|\beta\|} |f_h^C| \text{ donc} \\
 L_d + \frac{h^2}{\|\beta\|} |f_h^C| &\geq h |u_h(Q_b) - u_h(Q_a)|
 \end{aligned}$$

2) Avec  $h |u_h(Q_b) - u_h(Q_a)|$  il est possible de minorer  $|\beta \cdot \nabla_h u_h| |T|$  pour  $T \in C$ .



Pour cela on utilise la formule suivante :  
Soit  $T \in \mathcal{T}_h$ , alors

$$\beta \cdot \nabla_h u_{h|_T} |T| = \hat{\theta}_2^T (u_2^T - u_1^T) + \hat{\theta}_0^T (u_2^T - u_0^T)$$

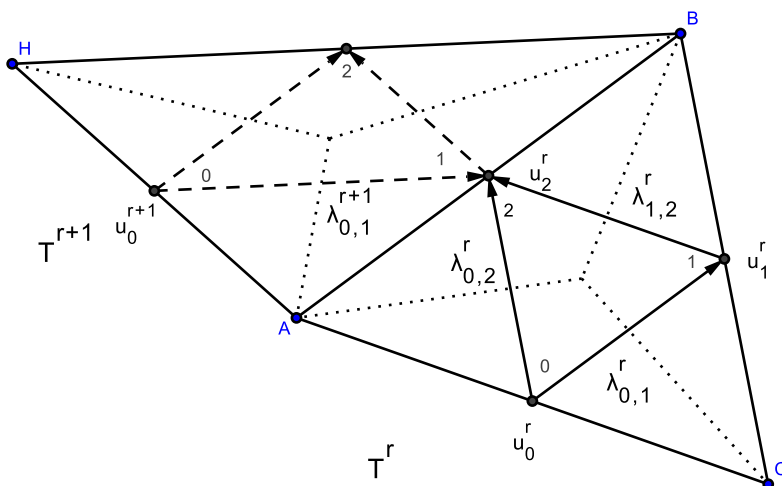
où

$$u_0^T = u_h(Q_0^T), u_1^T = u_h(Q_1^T), u_2^T = u_h(Q_2^T)$$

et

$$\hat{\theta}_2^T = \theta_{0,2}^T + \theta_{1,2}^T, \hat{\theta}_0^T = \theta_{0,2}^T + \theta_{0,1}^T$$

La figure ci-dessous représente, pour une cellule  $C \in \mathcal{C}_h$ , la situation du triangle  $T^r$  avec deux cas possibles :



a)  $Q_2^r$  est un point intérieur de  $C$  et on a les formules :

$$\begin{aligned} u_2^r &= u_0^{r+1}\lambda_{0,1}^{r+1} + u_0^r\lambda_{0,2}^r + u_1^r\lambda_{1,2}^r + \alpha_2^r \text{ avec } \lambda_{0,1}^{r+1} + \lambda_{0,1}^r + \lambda_{1,2}^r = 1 \\ \beta \cdot \nabla_h u_{hT^r} |T^r| &= \hat{\theta}_2^r(u_2^r - u_1^r) + \hat{\theta}_0^r(u_1^r - u_0^r) \\ &= \hat{\theta}_2^r[u_0^{r+1}\lambda_{0,1}^{r+1} + u_0^r\lambda_{0,2}^r + u_1^r\lambda_{1,2}^r + \alpha_2^r - u_1^r] + \hat{\theta}_0^r(u_1^r - u_0^r) \\ &\quad \text{Avec } u_1^r - u_0^r = u_1^r - u_0^{r+1} + u_0^{r+1} - u_0^r \text{ et } \lambda_{0,1}^{r+1} + \lambda_{0,1}^r + \lambda_{1,2}^r = 1 : \\ \beta \cdot \nabla_h u_{hT^r} |T^r| &= \hat{\theta}_2^r\alpha_2^r + (u_0^{r+1} - u_1^r)[\hat{\theta}_2^r(\lambda_{0,1}^{r+1} + \lambda_{0,2}^r) - \hat{\theta}_0^r] + (u_0^{r+1} - u_0^r)[\hat{\theta}_0^r - \hat{\theta}_2^r\lambda_{0,2}^r] \end{aligned}$$

Avec  $|\hat{\theta}_2^r| \leq h\|\beta\|$ ,  $|\hat{\theta}_0^r| \leq h\|\beta\|$  et  $|\alpha_2^r| \leq \frac{h}{\|\beta\|}f_2^r$  on obtient :

$$|\beta \cdot \nabla_h u_{hT^r} |T^r| \leq h^2|f_2^r| + 2h\|\beta\||u_0^{r+1} - u_1^r| + 2h\|\beta\||u_0^{r+1} - u_0^r| \quad (17.367)$$

b)  $Q_2^r$  est un point sortant de  $C$  (alors il n'y pas de triangle  $T^{r+1}$ ) et on a les formules :

$$u_2^r = u_0^r\lambda_{0,2}^r + u_1^r\lambda_{1,2}^r + \alpha_2^r \text{ avec } \lambda_{0,1}^r + \lambda_{1,2}^r = 1$$

De façon identique on obtient :

$$|\beta \cdot \nabla_h u_{hT^r} |T^r| \leq h^2|f_2^r| + 2h\|\beta\||u_0^r - u_1^r| \quad (17.368)$$

Pour calculer  $\beta \cdot \nabla_h u_h|_{T^r}|T|$ .

Le lemme 17.10 s'applique aussi pour les points intérieurs, car on dispose de l'égalité (16.251), page 130 pour les cellules avec deux coordonnées locales 0, de l'égalité (16.265), page 134 pour les cellules sans point intérieur de type 1 – 1 et des l'égalité (16.282) à (16.287), page 137 pour les cellules avec un point intérieur de type 1 – 1.

Posons  $\mathcal{E}_{d_r} = h|u_h(Q_0^r) - u_h(Q_0^{r-1})|$  et  $d_r \in I_{\text{contrôle}}^C$  l'élément de contrôle qui correspond à l'intersection des segments  $S_0^r$  et  $S_0^{r-1}$  de  $C$ .

On procède maintenant exactement comme dans la preuve du lemme 17.11 (estimations des fonctions auxiliaires  $\mathcal{L}(\varpi_-^C \kappa)$ ,  $\mathcal{L}(\varpi_+^C \kappa)$ ), pour obtenir :

**Points intérieurs :**

$$h|u_0^{r+1} - u_1^r| \leq \sum_{d \in I_{\text{contrôle}}^C} \mathcal{E}_d + \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j)| \quad (17.369)$$

**Points sortants :**

$$h|u_0^r - u_1^r| \leq \sum_{d \in I_{\text{contrôle}}^C} \mathcal{E}_d + \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j)| \quad (17.370)$$

Ainsi, avec (17.368) et (17.367) on a :

**Points intérieurs :**

$$\begin{aligned} |\beta \cdot \nabla_h u_{hT^r} |T^r| &\leq h^2|f_2^r| + 2\|\beta\| \left[ \sum_{d \in I_{\text{contrôle}}^C} \mathcal{E}_d + \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j)| \right] + 2\|\beta\| \mathcal{E}_{d_{r+1}} \\ &\leq h^2|f_2^r| + 2\|\beta\| \left( \sum_{d \in I_{\text{contrôle}}^C} \mathcal{E}_d + \mathcal{E}_{d_{r+1}} \right) + 2h^2 \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j)| \end{aligned}$$



Points sortants :

$$\begin{aligned} |\beta \cdot \nabla_h u_{hT^r}| |T^r| &\leq h^2 |f_2^r| + 2 \|\beta\| \left[ \sum_{d \in I_{\text{contrôle}}^C} \mathcal{E}_d + \frac{h^2}{\|\beta\|} \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j)| \right] \\ &\leq h^2 |f_2^r| + 2 \|\beta\| \sum_{d \in I_{\text{contrôle}}^C} \mathcal{E}_d + 2h^2 \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j)| \end{aligned}$$

En sommant sur tous les éléments  $T$  de la cellule  $C$  on a :

$$\sum_{T \in C} |\beta \cdot \nabla_h u_{hT}| |T| \leq (2K_{\text{cell}} + 1) \|\beta\| \sum_{d \in I_{\text{contrôle}}^C} \mathcal{E}_d + (2K_{\text{cell}} + 1) h^2 \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j)| \quad (17.371)$$

On a maintenant de 1) :

$$\begin{aligned} \mathcal{E}_d &\leq L_d + \frac{h^2}{\|\beta\|} |f^C| \text{ donc} \\ \sum_{d \in I_{\text{contrôle}}^C} \mathcal{E}_d &\leq \sum_{d \in I_{\text{contrôle}}^C} L_d + \sum_{d \in I_{\text{contrôle}}^C} \frac{h^2}{\|\beta\|} |f^C| \\ &\text{Avec le choix de } f^C \text{ et } \text{card}(I_{\text{contrôle}}^C) \leq \text{card}(J_{\text{entrant}}^C) \text{ on a :} \end{aligned}$$

$$\begin{aligned} \sum_{d \in I_{\text{contrôle}}^C} \mathcal{E}_d &\leq \sum_{d \in I_{\text{contrôle}}^C} \xi_{j_d} + \frac{h^2}{\|\beta\|} \sum_{j \in J_{\text{entrant}}^C} |f_h(Q_j)| \text{ Donc :} \\ \sum_{T \in C} |\beta \cdot \nabla_h u_{hT}| |T| &\leq (2K_{\text{cell}} + 1) \|\beta\| \left[ \sum_{d \in I_{\text{contrôle}}^C} L_d + \frac{h^2}{\|\beta\|} \sum_{j \in J_{\text{entrant}}^C} |f_h(Q_j)| \right] \\ &\quad + (2K_{\text{cell}} + 1) h^2 \sum_{j \in J^C \setminus J_{\text{entrant}}^C} |f_h(Q_j)| \\ \sum_{T \in C} |\beta \cdot \nabla_h u_{hT}| |T| &\leq (2K_{\text{cell}} + 1) \|\beta\| \sum_{d \in I_{\text{contrôle}}^C} L_d + h^2 (2K_{\text{cell}} + 1) \sum_{j \in J^C} |f_h(Q_j)| \end{aligned}$$

En sommant sur l'ensemble des cellules on a maintenant ( $h^2 \leq \frac{3}{2} \hat{c}_0 |D_i|$ ) :

$$\begin{aligned} |u_h|_{Y_{1,h}} &= \sum_{\substack{C \in \mathcal{C}_h \\ \text{et } I_{\text{contrôle}}^C \neq \emptyset}} \sum_{T \in C} |\beta \cdot \nabla_h u_{hT}| |T| + \sum_{\substack{C \in \mathcal{C}_h \\ \text{et } I_{\text{contrôle}}^C = \emptyset}} \sum_{T \in C} |\beta \cdot \nabla_h u_{hT}| |T| \\ |u_h|_{Y_{1,h}} &\stackrel{(17.334)}{\leq} (2K_{\text{cell}} + 1) \left[ \|\beta\| \mathcal{K}_N + \frac{3}{2} \hat{c}_0 \|f_h\|_{L_h^1(\Omega)} \right] + \|f_h\|_{L_h^1(\Omega)} \\ |u_h|_{Y_{1,h}} &\leq (2K_{\text{cell}} + 1) \left[ K_{\Xi} \frac{6\pi}{\theta_0} \hat{c}_0 (K_{\text{cell}} + 2)^2 \sum_{T \in \mathcal{T}_h} |T| \|\nabla_h f_h|_T\| + \right. \\ &\quad \left. \|\beta\| K_{\Xi} |g_h|_{\text{Var}_h(\Gamma^-)} + K_{\Xi} c_0 \sigma \|f_h\|_{L_h^1(\Gamma^-)} + \frac{3}{2} \hat{c}_0 \|f_h\|_{L_h^1(\Omega)} \right] + \|f_h\|_{L_h^1(\Omega)} \end{aligned}$$

□

Avant de procéder à la preuve du théorème 17.5, quelques estimations utiles :

**Lemme 17.13** *Soit  $\varphi_h \in Y_h$ , alors :*

a)

$$|\varphi_h|_{Var_h(\Gamma^-)} \leq \frac{1}{h} 2c_0\sigma \|\varphi_h\|_{L_h^1(\Gamma^-)} \quad (17.372)$$

b)

$$\|\nabla_h \varphi_h\|_{L^1(\Omega)} \leq \frac{1}{h} 6\hat{c}_0\sigma \|\varphi_h\|_{L_h^1(\Omega)} \quad (17.373)$$

c) *Si la condition  $\delta - S^-$  est satisfaite pour la famille de maillage  $(\mathcal{T}_h)_h$ , alors*

$$0 < \frac{|D_i|}{d_i} \leq \frac{(c_0\sigma)^2}{\delta} \quad \forall i \in J \setminus J^- \quad (17.374)$$

**Preuve :**

**Pour a) :** Soit  $i_1, i_2, \dots, i_{|J^-|}$  une numérotation de  $|J^-|$  telle que  $Q_{i_1}, Q_{i_2}, \dots, Q_{i_{|J^-|}}$  "longent" le bord entrant  $\Gamma^-$  (c'est-à-dire  $S_{i_k} \cap S_{i_{k+1}} \neq \emptyset$ ), alors

$$\begin{aligned} |\varphi_h|_{Var_h(\Gamma^-)} &= \sum_{k=1}^{|J^-|-1} |\varphi_h(Q_{i_{k+1}}) - \varphi_h(Q_{i_k})| \\ &\leq \frac{1}{h} \sum_{k=1}^{|J^-|-1} |\varphi_h(Q_{i_{k+1}})| h + \frac{1}{h} \sum_{k=1}^{|J^-|-1} |\varphi_h(Q_{i_k})| h \end{aligned}$$

Pour  $T \in \mathcal{T}_h$  quelconque, on a  $h \leq c_0\sigma\rho_T \leq c_0\sigma|S|$  où  $S \subset \partial T$  est un côté quelconque de  $T$ . Donc

$$\begin{aligned} |\varphi_h|_{Var_h(\Gamma^-)} &\leq \frac{1}{h} c_0\sigma \sum_{k=1}^{|J^-|-1} |\varphi_h(Q_{i_{k+1}})| |S_{i_{k+1}}| + \frac{1}{h} c_0\sigma \sum_{k=1}^{|J^-|-1} |\varphi_h(Q_{i_k})| |S_{i_k}| \\ &\leq \frac{1}{h} 2c_0\sigma \sum_{k=1}^{|J^-|} |\varphi_h(Q_{i_k})| |S_{i_k}| \end{aligned}$$

**Pour b) :** Soit  $T \in \mathcal{T}_h$  quelconque et  $S_0^T, S_1^T, S_2^T$  sont les côtés de  $T$  ( $\partial T = S_0^T \cup S_1^T \cup S_2^T$ ) de milieu, respectivement,  $Q_0^T, Q_1^T, Q_2^T$ . Alors avec  $m = \nabla_h \varphi_h|_T$  on a

$$\begin{aligned} \|\nabla_h \varphi_h|_T\|^2 |T| &= \int_T \|\nabla_h \varphi_h|_T\|^2 dx \\ &= \int_T m \cdot \nabla_h \varphi_h|_T dx \\ &= \int_{\partial T} \varphi_h|_T m \cdot n d\sigma \end{aligned}$$

$$\begin{aligned}
&= \sum_{l=0}^2 \varphi_{h|T}(Q_l^T) |S_l^T| (m \cdot n_{S_l^T}) \\
&\leq \|m\| \sum_{l=0}^2 |\varphi_{h|T}(Q_l^T)| h
\end{aligned}$$

En divisant cette inégalité par  $\|m\| = \|\nabla_h \varphi_{h|T}\|$  :

$$\|\nabla_h \varphi_{h|T}\| |T| \leq \frac{1}{h} \sum_{l=0}^2 |\varphi_{h|T}(Q_l^T)| h^2$$

Notons  $D_{Q_l^T}$  le volume fini de  $Q_l^T$ , alors  $\frac{1}{3} \frac{1}{\hat{c}_0} h^2 \leq \frac{1}{3} |T| \leq |D_{Q_l^T}|$ , donc

$$\|\nabla_h \varphi_{h|T}\| |T| \leq \frac{1}{h} 3\hat{c}_0 \sum_{l=0}^2 |\varphi_{h|T}(Q_l^T)| |D_{Q_l^T}|$$

Ainsi

$$\sum_{T \in \mathcal{T}_h} \|\nabla_h \varphi_{h|T}\| |T| \leq \frac{1}{h} 3\hat{c}_0 \sum_{T \in \mathcal{T}_h} \sum_{l=0}^2 |\varphi_{h|T}(Q_l^T)| |D_{Q_l^T}| \leq \frac{1}{h} 6\hat{c}_0 \|\varphi_h\|_{L_h^1(\Omega)}$$

**Pour c) :** On a  $d_i \geq \frac{1}{\|\beta\|^2} \delta \|\beta\|^2 |S_i|^2$  et (voir a))  $h \leq c_0 \sigma |S_i|$  donc

$$\frac{|D_i|}{d_i} \leq \frac{h^2}{\delta |S_i|^2} \leq \frac{(c_0 \sigma)^2}{\delta}$$

□

**Preuve du théorème 17.5 :** Soit  $u_h \in Y_h$  la solution du problème variationnel approché I (11.199), alors

$$\begin{aligned}
b_h(u_h, v_h) &= \sum_{k \in J \setminus J^-} f_h(Q_k) v_h(Q_k) |D_k| \\
&= \sum_{k \in J \setminus J^-} d_k \left[ \frac{|D_k|}{d_k} f_h(Q_k) \right] v_h(Q_k) \\
&= \sum_{k \in J \setminus J^-} d_k \tilde{f}_h(Q_k) v_h(Q_k)
\end{aligned}$$

Posons  $\tilde{f}_h(Q_k) = \frac{|D_k|}{d_k} f_h(Q_k)$  pour  $k \in J \setminus J^-$  et  $\tilde{f}_h(Q_k) = 0$  pour  $k \in J^-$ , alors

$$b_h(u_h, v_h) = (\tilde{f}_h, v_h)_{\Pi} \quad \forall v_h \in Y_h^0 \text{ et } \|\tilde{f}_h\|_{L_h^1(\Gamma^-)} = 0$$

Avec l'estimation b) ci-dessus, pour  $\varphi_h = \tilde{f}_h$ , on a

$$\|\nabla_h \tilde{f}_h\|_{L^1(\Omega)} \leq \frac{1}{h} 6\hat{c}_0 \sigma \|\tilde{f}_h\|_{L_h^1(\Omega)} \leq \frac{1}{h} 6\hat{c}_0 \sigma \left[ 1 + \frac{(c_0 \sigma)^2}{\delta} \right] \|f_h\|_{L_h^1(\Omega)}$$

Le théorème technique 17.6 s'applique et on obtient la preuve du théorème 17.5. □

## 17.4 Conclusion :

En combinant (17.330) et (17.307), on obtient le résultat de stabilité suivant :

**Théorème 17.14 (Stabilité)** *Soit  $(\mathcal{T}_h)$  une famille de maillage quasi-uniforme et régulière, soit  $\mathcal{G}_h$  le graphe associé à  $\mathcal{T}_h$ . Soit  $\mathcal{C}_h$  l'ensemble des cellules obtenues à partir de  $\mathcal{G}_h$ . On suppose que :*

- a)  $\mathcal{C}_h$  ne contient pas de point intérieur de type 0 – 1 ou 0 – 2 (donc  $\mathcal{G}_h$  ne contient pas de circuit élémentaire).
- b) Il existe  $K_{\Xi} > 0$ , ne dépendant pas de  $h$ , tel que pour le nombre de strates  $N_{\Xi}$  on ait  $N = N_{\Xi} \leq K_{\Xi} \frac{1}{h}$ ,
- c) La condition  $\delta - S^-$  est satisfaite pour la famille de maillage  $(\mathcal{T}_h)_h$

Alors pour la solution  $u_h \in Y_h$  du problème variationnel approché I (11.199), il existe une constante  $K > 0$  et  $h_0 > 0$  ne dépendant pas de  $h$ , ni de  $\|\beta\|$ , telle que :

$$\begin{aligned} \|u_h\|_{Y_{1,h}(\Omega)} &= \|u_h\|_{L_h^1(\Omega)} + h \|\beta \cdot \nabla_h u_h\|_{L^1(\Omega)} = \|u_h\|_{L_h^1(\Omega)} + h \sum_{T \in \mathcal{T}_h} |\beta \cdot \nabla_h u_h|_T |T| \\ &\leq K \left[ \|g_h\|_{L_h^1(\Gamma^-)} + h_0 \|\beta\| \|g_h\|_{Var_h(\Gamma^-)} + \left(1 + \frac{1}{\|\beta\|}\right) \|f_h\|_{L_h^1(\Omega)} \right] \quad \forall h \in ]0; h_0[ \end{aligned} \tag{17.375}$$

## Sixième partie

## Annexes

---

Sommaire

---

<b>A</b>	<b>Existence, unicité et estimation d'erreur pour le problème d'advection-diffusion stationnaire</b>	<b>189</b>
<b>B</b>	<b>Existence et unicité du problème exact d'advection</b>	<b>191</b>
<b>C</b>	<b>Quelques propriétés de <math>d_k = \frac{1}{\ \beta\ ^2} \sum_{j \in S^-(k)} [\beta \cdot (Q_k - Q_j)] \theta_{j,k}</math></b>	<b>196</b>
C.1	Interprétation géométrique de $d_k$ . . . . .	196
C.2	Encapsulation de l'information dans $\mathcal{G}_h$ et forme linéaire $(\cdot, \cdot)_\Pi$ . . .	197
<b>D</b>	<b>Liste des symboles</b>	<b>201</b>
	<b>Références</b>	<b>210</b>

---

## A Existence, unicité et estimation d'erreur pour le problème d'advection-diffusion stationnaire

Les résultats suivants sont bien connus. L'intérêt de certains de ces résultats est de montrer l'efficacité des coordonnées adaptées upwind dans le travail avec  $b_h$ .

**Théorème A.1 (Existence et unicité du problème continu)** *Le problème (1.6) est bien défini. Il existe une solution unique  $u \in V$  avec*

$$\|u\|_{*V} \leq \frac{1}{\nu} \|f\|_2 \quad (1.376)$$

**Preuve :** Soit  $a(u, v) = \nu((u, v)) + b(u, v)$  pour  $u, v \in V$ , alors la forme bilinéaire est continue et coercitive

$$\begin{aligned} |a(u, v)| &\leq (\nu + \|\beta\|) \|u\|_{*V} \|v\|_{1,2} \leq \frac{1}{c_1} (\nu + \|\beta\|) \|u\|_{*V} \|v\|_{*V} \quad \forall u, v \in V \text{ et} \\ a(v, v) &= \nu \|v\|_{*V}^2 + \frac{1}{2} \int_{\Omega} \nabla \cdot (\beta v^2) dx \\ &= \nu \|v\|_{*V}^2 \text{ car } v = 0 \text{ sur } \partial\Omega. \end{aligned}$$

Le théorème de Lax-Milgram s'applique. Il existe une solution unique de (1.6) avec (1.376)  $\square$

L'utilisation des coordonnées locales upwind permet maintenant de faire apparaître les similitudes des preuves avec le problème approché.

**Théorème A.2 (Existence et unicité du problème approché)** *Le problème approché (2.98) a exactement une solution  $u_h \in V_h$  et*

$$\|u_h\|_{*V_h} \leq \frac{1}{\nu} \|f\|_2 \quad (1.377)$$

**Preuve :**

La forme bilinéaire  $a_h(\cdot, \cdot) : V_h \times V_h \rightarrow \mathbb{R}$  définie par (2.99) est continue :

Soit  $w_h, v_h \in V_h$  and  $T \in \mathcal{T}_h$ , alors avec (2.97), on a

$$\begin{aligned} |a_h(w_h, v_h)| &\leq \nu |((w_h, v_h))_h| + \|\beta\| (\sqrt{3}\hat{c}_* + h_0\hat{c}_0) \|u_h\|_{*V_h} \|v_h\|_{*V_h} \\ |a_h(w_h, v_h)| &\leq [\nu + \|\beta\| (\sqrt{3}\hat{c}_* + h_0\hat{c}_0)] \|w_h\|_{*V_h} \|v_h\|_{*V_h} \end{aligned} \quad (1.378)$$

La forme bilinéaire est coercitive :

Soit  $\varphi = 1$ , alors avec le lemme 2.17 on obtient  $a_h(w_h, w_h) \geq \nu \|w_h\|_{*V_h}^2$ .

Le théorème de Lax-Milgram s'applique et on a l'inégalité de stabilité (1.377).  $\square$

**Théorème A.3 (estimation d'erreur)** *Soit  $\Omega \subset \mathbb{R}^2$  un domaine polynomial convexe. Soit  $u \in V$  la solution du problème continu (1.6) et  $u_h \in V_h$  la solution du problème approché (2.98). Alors*

$$\|u - u_h\|_{*V_h} \leq h \frac{1}{\nu^2} \left[ (2\nu + \|\beta\|) c_{I1} c_2 (1 + \frac{1}{\nu} \|\beta\|) + \|\beta\| (4\hat{c}_0 + 1) c_{I3} c_1 \right] \|f\|_2 \quad (1.379)$$

**Preuve :**

Comme le domaine est polynomial et convexe, la solution  $w$  de  $-\Delta w = \hat{f}$ ,  $\hat{f} \in L^2(\Omega)$  est dans  $w \in H^2(\Omega) \cap H_0^1(\Omega)$  (voir [Gri85]). Il existe (résultat de régularité elliptique)  $c_2 > 0$  telle que

$$\|w\|_{2,2} \leq c_2 \|\Delta w\|_2 \text{ for } w \in H^2(\Omega) \cap H_0^1(\Omega) \quad (1.380)$$

Soit  $u \in V$  la solution du problème continu  $\hat{f} = \frac{1}{\nu}(f - \beta \cdot \nabla u) \in L^2(\Omega)$ . Alors  $-\Delta w = \hat{f}$  pour  $w = u$ . Ainsi

$$\|u\|_{H^2(\Omega)} \leq c_2 \frac{1}{\nu} (\|f\|_{L^2(\Omega)} + \|\beta\| \|\nabla u\|_{L^2(\Omega)}) \stackrel{(1.376)}{\leq} c_2 \frac{1}{\nu} (1 + \frac{1}{\nu} \|\beta\|) \|f\|_{L^2(\Omega)} \quad (1.381)$$

Le lemme 2.3 s'applique et, avec (1.381), on a

$$\|u - I_h u\|_{*V_h} \leq c_{T1} c_2 h \frac{1}{\nu} (1 + \frac{1}{\nu} \|\beta\|) \|f\|_2 \quad (1.382)$$

Soit  $u_h$  la solution du problème approché et  $v_h, w_h \in V_h$ , alors

$$\begin{aligned} & \nu((u_h - w_h, v_h))_h + b_h(u_h - w_h, v_h) \\ = & \underbrace{\nu((u_h, v_h))_h + b_h(u_h, v_h)}_{=(f, v_h)} - \nu((w_h, v_h))_h - b_h(w_h, v_h) \\ = & \underbrace{\nu((u, v_h)) + (\beta \cdot \nabla u, v_h)} - \nu((w_h, v_h))_h - b_h(w_h, v_h) \\ = & \nu((u - w_h, v_h))_h + (\beta \cdot \nabla u - \beta \cdot \nabla_h(I_h u), v_h) \\ & + (\beta \cdot \nabla_h(I_h u), v_h) + (I_h u, \beta \cdot \nabla_h v_h) + (-I_h u, \beta \cdot \nabla_h v_h) - b_h(I_h u, v_h) \\ & + b_h(I_h u, v_h) - b_h(w_h, v_h) \end{aligned} \quad (1.383)$$

Maintenant

$$\begin{aligned} \nu((u - w_h, v_h))_h & \leq \nu \|u - w_h\|_{*V_h} \|v_h\|_{*V_h} \\ (\beta \cdot \nabla u - \beta \cdot \nabla_h(I_h u), v_h) & \leq \|\beta\| \|u - I_h u\|_{*V_h} \|v_h\|_{*V_h} \\ (\beta \cdot \nabla_h(I_h u), v_h) + (I_h u, \beta \cdot \nabla_h v_h) & \stackrel{\text{lemme 2.15}}{\leq} 3 \|\beta\| \hat{c}_0 h \|I_h u\|_{*V_h} \|v_h\|_{*V_h} \\ |(-I_h u, \beta \cdot \nabla_h v_h) - b_h(I_h u, v_h)| & \stackrel{(2.96)}{\leq} \|\beta\| h (1 + \hat{c}_0) \|I_h u\|_{*V_h} \|v_h\|_{*V_h} \\ b_h(I_h u, v_h) - b_h(w_h, v_h) & \stackrel{(2.97)}{\leq} \|\beta\| (\sqrt{3} \hat{c}_* + h_0 \hat{c}_0) \|I_h u - w_h\|_{*V_h} \|v_h\|_{*V_h} \end{aligned}$$

En choisissant  $v_h = u_h - w_h$  on obtient

$$\begin{aligned} \nu \|u_h - w_h\|_{*V_h}^2 &\leq \nu((u_h - w_h, v_h))_h + b_h(u_h - w_h, v_h) \\ &\leq \left[ \nu \|u - w_h\|_{*V_h} + \|\beta\| \|u - I_h u\|_{*V_h} + 3\|\beta\| \hat{c}_0 h \|I_h u\|_{*V_h} + \right. \\ &\quad \left. \|\beta\| h(1 + \hat{c}_0) \|I_h u\|_{*V_h} + \|\beta\| (\sqrt{3}\hat{c}_* + h_0 \hat{c}_0) \|I_h u - w_h\|_{*V_h} \right] \|v_h\|_{*V_h} \end{aligned}$$

En divisant par  $\|v_h\|_{*V_h}$ , et en prenant  $w_h = I_h u$  on obtient

$$\begin{aligned} \nu \|u_h - I_h u\|_{*V_h} &\leq \nu \|u - I_h u\|_{*V_h} + \|\beta\| \|u - I_h u\|_{*V_h} + \\ &\quad 3\|\beta\| \hat{c}_0 h \|I_h u\|_{*V_h} + \|\beta\| h(1 + \hat{c}_0) \|I_h u\|_{*V_h} \end{aligned}$$

Avec le lemme 2.3, (1.382) et (1.3) :

$$\begin{aligned} \nu \|u - u_h\|_{*V_h} &\leq \nu (\|u - I_h u\|_{*V_h} + \|u_h - I_h u\|_{*V_h}) \\ &\leq (2\nu + \|\beta\|) c_{I1} c_2 h \frac{1}{\nu} (1 + \frac{1}{\nu} \|\beta\|) \|f\|_2 + h \|\beta\| (4\hat{c}_0 + 1) c_{I3} c_1 \frac{1}{\nu} \|f\|_2 \\ \nu \|u - u_h\|_{*V_h} &\leq h \frac{1}{\nu} \left[ (2\nu + \|\beta\|) c_{I1} c_2 (1 + \frac{1}{\nu} \|\beta\|) + \|\beta\| (4\hat{c}_0 + 1) c_{I3} c_1 \right] \|f\|_2 \end{aligned}$$

□

## B Existence et unicité du problème exact d'advection

**Théorème B.1 (Existence et unicité du problème d'advection)** *Soit*

$\emptyset \neq \Omega \subset \mathbb{R}^2$  un domaine polygonal borné, tel que le bord entrant  $\Gamma^-$  soit connexe et  $g \in W^{1,1}(\Gamma^-)$

A) Si  $f \in W^{1,1}(\Omega)$  alors il existe exactement une solution  $u \in W^{1,1}(\Omega) \subset Y_1$  de (1.18) et on a les inégalités

$$\|u\|_{L^1(\Omega)} \leq \text{diam}(\Omega) \|g\|_{L^1(\Gamma^-)} + \frac{\text{diam}(\Omega)}{\|\beta\|} \|f\|_{L^1(\Omega)} \quad (2.384)$$

$$\|\beta \cdot \nabla u\|_{L^1(\Omega)} = \|f\|_{L^1(\Omega)} \quad (2.385)$$

$$\begin{aligned} \|\nabla u\|_{L^1(\Omega)} &\leq (C_{\Gamma^-, \beta} + 1) \text{diam}(\Omega) \|g'\|_{L^1(\Gamma^-)} + \frac{C_{\Gamma^-, \beta}}{\|\beta\|} \|f\|_{L^1(\Omega)} \\ &\quad + \frac{\text{diam}(\Omega)}{\|\beta\|} (C_{\Gamma^-, \beta} + 1) \|\nabla f\|_{L^1(\Omega)} \end{aligned} \quad (2.386)$$

$$\text{avec } C_{\Gamma^-, \beta} = \max_{x \in \Gamma^-} \frac{\|\beta\|}{|\beta \cdot n(x)|}. \quad (2.387)$$

$$(2.388)$$

Ainsi

$$\|u\|_{Y_1} \leq \text{diam}(\Omega) \|g\|_{L^1(\Gamma^-)} + \left(1 + \frac{\text{diam}(\Omega)}{\|\beta\|}\right) \|f\|_{L^1(\Omega)} \quad (2.389)$$

$$\|u\|_{W^{1,1}(\Omega)} \leq \tilde{C} (\|g\|_{W^{1,1}(\Gamma^-)} + \|f\|_{W^{1,1}(\Omega)}) \text{ et} \quad (2.390)$$

$$\tilde{C} \text{ est une constante qui dépend de } \beta, \Gamma^- \text{ et } \Omega. \quad (2.391)$$



B) Si  $f \in W^{2,1}(\Omega)$  alors il existe exactement une solution  $u \in W^{1,1}(\Omega) \cap C(\overline{\Omega})$  de (1.18). Les inégalités ci-dessus restent valides et on a

$$\|u\|_{L^\infty(\Omega)} \leq \|g\|_{L^\infty(\Gamma^-)} + \frac{\text{diam}(\Omega)}{\|\beta\|} \|f\|_{L^\infty(\Omega)} \quad (2.392)$$

**Preuve :**

Soit  $\tilde{\Omega} = \cup_{x \in \Gamma^-} \{x + \beta t : t \in \mathbb{R}\}$ . Soit  $P : \tilde{\Omega} \rightarrow \Gamma^-$  la projection parallèlement à  $\beta$  sur  $\Gamma^-$ .  $\Gamma^-$  est connexe et peut donc s'écrire comme la réunion de  $k$  segments  $\Gamma^- = ]A_0; A_1] \cup [A_1; A_2] \cup \dots \cup [A_i; A_{i+1}] \cup \dots \cup [A_{k-1}; A_k[$ .

Sur le segment  $[A_i; A_{i+1}]$  on note  $n_i$  la normale extérieure à  $\Gamma^-$  et  $n_i^T = \frac{\overrightarrow{A_i A_{i+1}}}{\|A_i A_{i+1}\|}$  un vecteur orthogonal à  $n_i$ . Soit  $\omega_i = \{x \in \Omega : P(x) \in ]A_i; A_{i+1}[ \}$  l'ensemble des points de  $\Omega$  projetés sur le segment ouvert  $]A_i; A_{i+1}[$ , alors

$$a) \quad \bigcup_{i \in \{0,1,\dots,k-1\}} \overline{\omega_i} = \overline{\Omega} \quad (2.393)$$

$$b) \quad P(x) = x - \frac{(x - A_i) \cdot n_i}{\beta \cdot n_i} \beta \quad (2.394)$$

$$c) \quad P : \Omega \rightarrow \Gamma^- \text{ est continue et } P|_{\omega_i} \text{ est dérivable et } DP \in C^\infty(\omega_i).$$

$$i) \quad \forall x \in \omega_i \quad \forall \eta \in \mathbb{R}^2 \quad DP(x)\eta = \eta - \frac{\eta \cdot n_i}{\beta \cdot n_i} \beta \quad (2.395)$$

$$ii) \quad \forall x \in \omega_i \quad DP(x)\beta = 0 \text{ et} \quad (2.396)$$

$$iii) \quad \forall x \in \omega_i \quad \forall \eta \in \mathbb{R}^2 \quad (DP(x)\eta) \cdot n_i = 0 \quad (2.397)$$

$$DP(x)\eta \text{ et } n_i^T \text{ sont donc colinéaires.} \quad (2.398)$$

$$(2.399)$$

Le fait que  $n(x) \cdot \beta < 0$  pour  $x \in \Gamma^-$  et, bien sur, que  $\Gamma^-$  soit une réunion finie de segments, permet d'assurer l'existence de  $C_{\Gamma^-, \beta} = \max_{x \in \Gamma^-} \frac{\|\beta\|}{|\beta \cdot n(x)|}$ .

$g \in W^{1,1}(\Gamma^-)$  et  $\Gamma^-$  est une réunion finie de segments, donc il existe une suite  $(g_l)_{l \geq 0}$  de fonctions définies sur  $\Gamma^-$  avec  $\lim_{l \rightarrow \infty} \|g_l - g\|_{W^{1,1}(\Gamma^-)} = 0$  et telle que  $g_l \in C(\Gamma^-)$  et  $g_l|_{]A_i; A_{i+1}[} \in C^\infty(]A_i; A_{i+1}[)$  pour  $i \in \{0, \dots, k\}$ .

Le domaine  $\Omega$  est polygonal et borné, donc il existe une suite  $(f_l)_{l \geq 0}$  de fonctions définies sur  $\Omega$  telles de  $\lim_{l \rightarrow \infty} \|f_l - f\|_{W^{1,1}(\Omega)} = 0$  et  $f_l \in C^\infty(\Omega)$ .

La preuve passe par les étapes suivantes :

**Étape 1 :** Construire explicitement la solution  $u_l$  du problème  $\beta \cdot \nabla u_l = f_l$  avec  $u_l|_{\Gamma^-} = g_l$ . Cette solution est continue sur  $\overline{\Omega}$  et  $u_l|_{\omega_i} \in C^\infty(\omega_i)$ . Une intégration par parties permet maintenant de vérifier que  $\int_\Omega \frac{\partial_j u_l}{\partial x_j} \varphi \, dx = - \int_\Omega u_l \frac{\partial_j \varphi}{\partial x_j} \, dx$  pour toutes les fonctions tests  $\varphi \in \mathcal{D}(\Omega)$  et  $j \in \{1, 2\}$ . Donc  $u_l \in W^{1,1}(\Omega)$ .

**Étape 2 :** On établit ici les inégalités du théorème.

**Étape 3 :**  $u_l - u_m$  est solution du problème  $\beta \cdot \nabla u = f_l - f_m$  avec  $(u_l - u_m)|_{\Gamma^-} = g_l - g_m$ , on a donc maintenant

$$\|u_l - u_m\|_{W^{1,1}(\Omega)} \leq \tilde{C}(\|g_l - g_m\|_{W^{1,1}(\Gamma^-)} + \|f_l - f_m\|_{W^{1,1}(\Omega)})$$

pour  $l, m \geq 0$ . La suite  $(u_l)_{n \geq 0}$  est donc une suite de Cauchy, qui converge vers une fonction  $u \in W^{1,1}(\Omega)$ . Enfin

$$\|\beta \cdot \nabla u - f\|_{L^1(\Omega)} \leq \underbrace{\|\beta \cdot \nabla(u - u_l) - (f - f_l)\|_{L^1(\Omega)}}_{\leq \|\beta\| \|u - u_l\|_{W^{1,1}(\Omega)} + \|f - f_l\|_{L^1(\Omega)}} + \underbrace{\|\beta \cdot \nabla u_l - f_l\|_{L^1(\Omega)}}_{=0} \rightarrow 0 \text{ pour } l \rightarrow \infty$$

### Preuve de l'étape 1 :

On note  $D_{n_i} g_l(\bar{x}) = \lim_{r \rightarrow 0} \frac{g_l(\bar{x} + n_i r) - g_l(\bar{x})}{r}$  pour  $\bar{x} \in ]A_i; A_{i+1}[$ .

Alors pour  $x \in \omega_i$  et  $\eta \in \mathbb{R}^2$  (en utilisant (2.395) et (2.398) )

$$D_\eta(g_l \circ P)(x) = \lim_{r \rightarrow 0} \frac{g_l(P(x + \eta r)) - g_l(P(x))}{r} = D_{n_i} g_l(P(x)) \left( \eta - \frac{\eta \cdot n_i}{\beta \cdot n_i} \beta \right) \cdot n_i^T \quad (2.400)$$

Pour  $x \in \bar{\omega}_i$  on pose (méthode des caractéristiques) :

$$u_l(x) = g_l \circ P(x) + \int_0^{\frac{(x - A_i) \cdot n_i}{\beta \cdot n_i}} f_l(P(x) + \beta \tau) d\tau \quad (2.401)$$

On vérifie facilement que  $u_l \in C(\bar{\Omega})$  (continuité de  $P$  et  $g_l$  sur  $\Gamma^-$ ),  $u_l \in C^\infty(\omega_i)$  et  $u_l(x) = g_l(x)$  pour  $x \in \Gamma^-$ .

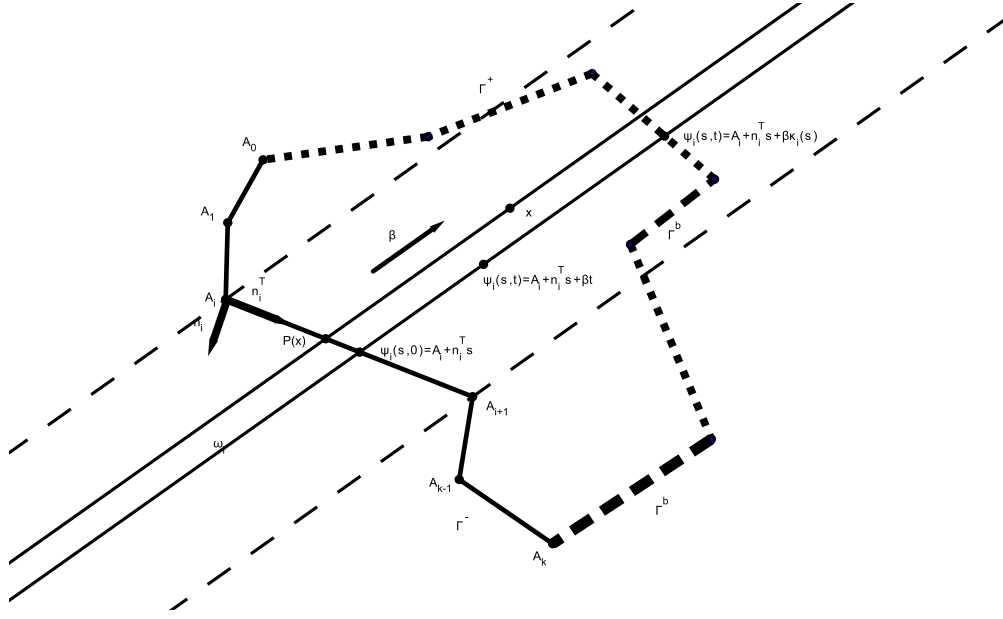
Maintenant pour  $x \in \omega_i$  et  $\eta \in \mathbb{R}^2$  :

$$\begin{aligned} D_\eta u_l(x) &= \lim_{r \rightarrow 0} \frac{u_l(x + \eta r) - u_l(x)}{r} \\ &= D_{n_i} g_l(P(x)) \left( \eta - \frac{\eta \cdot n_i}{\beta \cdot n_i} \beta \right) \cdot n_i^T + f_l \left( \frac{(x - A_i) \cdot n_i}{\beta \cdot n_i} + P(x) \right) \frac{\eta \cdot n_i}{\beta \cdot n_i} \\ &\quad + \int_0^{\frac{(x - A_i) \cdot n_i}{\beta \cdot n_i}} D f_l(P(x) + \beta \tau) \left( \eta - \frac{\eta \cdot n_i}{\beta \cdot n_i} \beta \right) d\tau \end{aligned}$$

Avec (2.394) on a  $\frac{(x - A_i) \cdot n_i}{\beta \cdot n_i} + P(x) = x$  et pour  $\eta = \beta$  on obtient immédiatement  $\beta \cdot \nabla u_l = D_\beta u_l(x) = f_l(x)$ .

$$\begin{aligned} \|D u_l(x)\| &= \max_{\|\eta\|=1} |D_\eta u_l(x)| \\ &\leq |D_{n_i} g_l(P(x))| (1 + C_{\Gamma^-, \beta}) + |f_l|(x) \frac{C_{\Gamma^-, \beta}}{\|\beta\|} \\ &\quad + \int_0^{\frac{(x - A_i) \cdot n_i}{\beta \cdot n_i}} \|D f_l(P(x) + \beta \tau)\| (1 + C_{\Gamma^-, \beta}) d\tau \end{aligned}$$

Soit  $\varphi \in \mathcal{D}(\Omega)$  et  $j \in \{1, 2\}$ . Le support de  $\varphi$  peut croiser les caractéristiques  $\{A_i + \beta t : t \geq 0\}$  issues des points  $A_i$  et décomposer le support de  $\varphi$  en des ensembles disjoints  $\omega_i \cap \text{support}(\varphi)$ . En procédant à des intégrations par parties sur chacun de ces ensembles disjoints et en utilisant la continuité de  $u_l$ , on obtient  $\int_\Omega \frac{\partial_j u_l}{\partial x_j} \varphi dx = - \int_\Omega u_l \frac{\partial_j \varphi}{\partial x_j} dx$ . Donc  $u_l \in W^{1,1}(\Omega)$ .



**Preuve de l'étape 2 :** On considère l'application

$$\Psi_i : \hat{\omega}_i \rightarrow \omega_i \text{ avec } \hat{\omega}_i = \{(s, t) \in \mathbb{R}^2 : 0 < s < |[A_i; A_{i+1}]| \text{ et } 0 < t < \kappa_i(s)\}$$

et  $\Psi_i$  définie par  $\Psi_i(s, t) = A_i + n_i^T s + \beta t$  et où  $\kappa_i(s)$  est le temps que met la caractéristique issue de  $\psi(s, 0) = A_i + n_i^T s \in [A_i; A_{i+1}]$  pour "traverser"  $\Omega$ , c'est-à-dire que<sup>61</sup>  $\kappa_i(s) = \inf\{t \geq 0 : \psi_i(s, t) \notin \Omega\}$ . On a la majoration  $\kappa_i(s) \leq \frac{\text{diam}(\Omega)}{\|\beta\|}$ .  $\Psi_i$  est un difféomorphisme de classe  $C^\infty$  de  $\hat{\omega}_i$  dans  $\omega_i$ .

On a  $D\Psi_i(s, t) = (n_i^T | \beta)$  et son déterminant est  $\det(D\Psi_i(s, t)) = \det(n_i^T | \beta) = \pm n_i \cdot \beta$ , donc  $|\det(D\Psi_i(s, t))| \leq \|\beta\|$ .

Dans les majorations qui suivent on utilisera aussi les relations suivantes :

$$P(\Psi_i(s, t)) = \Psi_i(s, 0) \text{ donc } \beta\tau + P \circ \Psi_i(s, t) = \Psi_i(s, \tau)$$

$$\frac{(\Psi_i(s, t) - A_i) \cdot n_i}{\beta \cdot n_i} = t \text{ donc } P(\Psi_i(s, \kappa_i(s))) + \beta\kappa_i(s) = \frac{(\Psi_i(s, \kappa_i(s)) - A_i) \cdot n_i}{\beta \cdot n_i} = \kappa_i(s)$$

Norme  $L^1$  de  $u_l$  sur  $\omega_i$  :

61. En regardant la figure ci-dessus, on constatera qu'il est parfaitement possible, pour la caractéristique issue de  $\psi_i(s, 0)$ , de traverser une partie de  $\Gamma^b$  et d'y "construire" la solution  $u_l$ . Ce temps n'est pas compté ici, car notre objectif est une intégration sur l'ouvert  $\Omega$ .

$$\begin{aligned}
\|u_l\|_{L^1(\omega_i)} &= \int_{\omega_i} |u_l|(x) dx = \int_0^{A_i A_{i+1}} \int_0^{\kappa_i(s)} |u_l| \circ \Psi_i(s, t) |det(n_i^T | \beta)| dt ds \\
&\leq \int_0^{A_i A_{i+1}} \int_0^{\kappa_i(s)} |g_l \circ P \circ \Psi_i(s, t)| \underbrace{|det(n_i^T | \beta)|}_{\leq \|\beta\|} dt ds \\
&\quad + \int_0^{A_i A_{i+1}} \int_0^{\kappa_i(s)} \left[ \int_0^{\frac{(\Psi_i(s, t) - A_i) \cdot n_i}{\beta \cdot n_i}} |f_l| \underbrace{(P(\Psi(s, t)) + \beta \tau)}_{=\Psi(s, \tau)} d\tau \right] |det(n_i^T | \beta)| dt ds \\
&\leq \int_0^{A_i A_{i+1}} \frac{\text{diam}(\Omega)}{\|\beta\|} |g_l \circ \Psi(s, 0)| \|\beta\| ds \\
&\quad + \int_0^{A_i A_{i+1}} \int_0^{\kappa_i(s)} \left[ \int_0^{\kappa_i(s)} |f_l|(\Psi(s, \tau)) d\tau \right] |det(n_i^T | \beta)| dt ds \\
&\leq \text{diam}(\Omega) \int_0^{A_i A_{i+1}} |g_l \circ \Psi(s, 0)| ds \\
&\quad + \underbrace{\frac{\text{diam}(\Omega)}{\|\beta\|} \int_0^{A_i A_{i+1}} \left[ \int_0^{\kappa_i(s)} |f_l|(\Psi(s, \tau)) d\tau \right] |det(n_i^T | \beta)| ds}_{=\|f_l\|_{L^1(\omega_i)}} \\
&\leq \text{diam}(\Omega) \|g_l\|_{L^1(\{A_i; A_{i+1}\})} + \frac{\text{diam}(\Omega)}{\|\beta\|} \|f_l\|_{L^1(\omega_i)}
\end{aligned} \tag{2.402}$$

En sommant maintenant sur les  $\omega_i$  pour  $i$  allant de 0 à  $k - 1$  on obtient (2.384)

On procède de même pour

$$\begin{aligned}
\|Du_l\|_{L^1(\omega_i)} &= \int_{\omega_i} \|Du_l\|(x) dx = \int_0^{A_i A_{i+1}} \int_0^{\kappa_i(s)} \|Du_l\| \circ \Psi_i(s, t) |det(n_i^T | \beta)| dt ds \\
&\leq (1 + C_{\Gamma^-, \beta}) \int_0^{A_i A_{i+1}} \int_0^{\kappa_i(s)} |D_\eta g_l \circ P \circ \Psi_i(s, t)| \underbrace{|det(n_i^T | \beta)|}_{\leq \|\beta\|} dt ds \\
&\quad + \frac{C_{\Gamma^-, \beta}}{\|\beta\|} \int_0^{A_i A_{i+1}} \int_0^{\kappa_i(s)} |f_l \circ \Psi_i(s, t)| |det(n_i^T | \beta)| dt ds \\
&\quad + (1 + C_{\Gamma^-, \beta}) \int_0^{A_i A_{i+1}} \int_0^{\kappa_i(s)} \left[ \int_0^{\frac{(\Psi_i(s, t) - A_i) \cdot n_i}{\beta \cdot n_i}} \|Df_l\| \underbrace{(P(\Psi(s, t)) + \beta \tau)}_{=\Psi(s, \tau)} d\tau \right] |det(n_i^T | \beta)| dt ds
\end{aligned}$$

En procédant aux mêmes estimations que pour  $\|u_l\|_{L^1(\omega_i)}$  on obtient :

$$\begin{aligned}
&\leq (1 + C_{\Gamma^-, \beta}) \text{diam}(\Omega) \|D_\eta g_l\|_{L^1(\{A_i; A_{i+1}\})} + \frac{C_{\Gamma^-, \beta}}{\|\beta\|} \|f_l\|_{L^1(\omega_i)} \\
&\quad + (1 + C_{\Gamma^-, \beta}) \frac{\text{diam}(\Omega)}{\|\beta\|} \|Df_l\|_{L^1(\omega_i)}
\end{aligned}$$

En sommant maintenant sur les  $\omega_i$  pour  $i$  allant de 0 à  $k - 1$  on obtient (2.386)

L'étape 3 a déjà été démontrée. Il reste à établir (2.392).  $f$  est maintenant continue, donc l'application

$$u(x) = g \circ P(x) + \int_0^{\frac{(x-A_i) \cdot n_i}{\beta \cdot n_i}} f(P(x) + \beta \tau) d\tau$$

l'est aussi, et on a<sup>62</sup> :

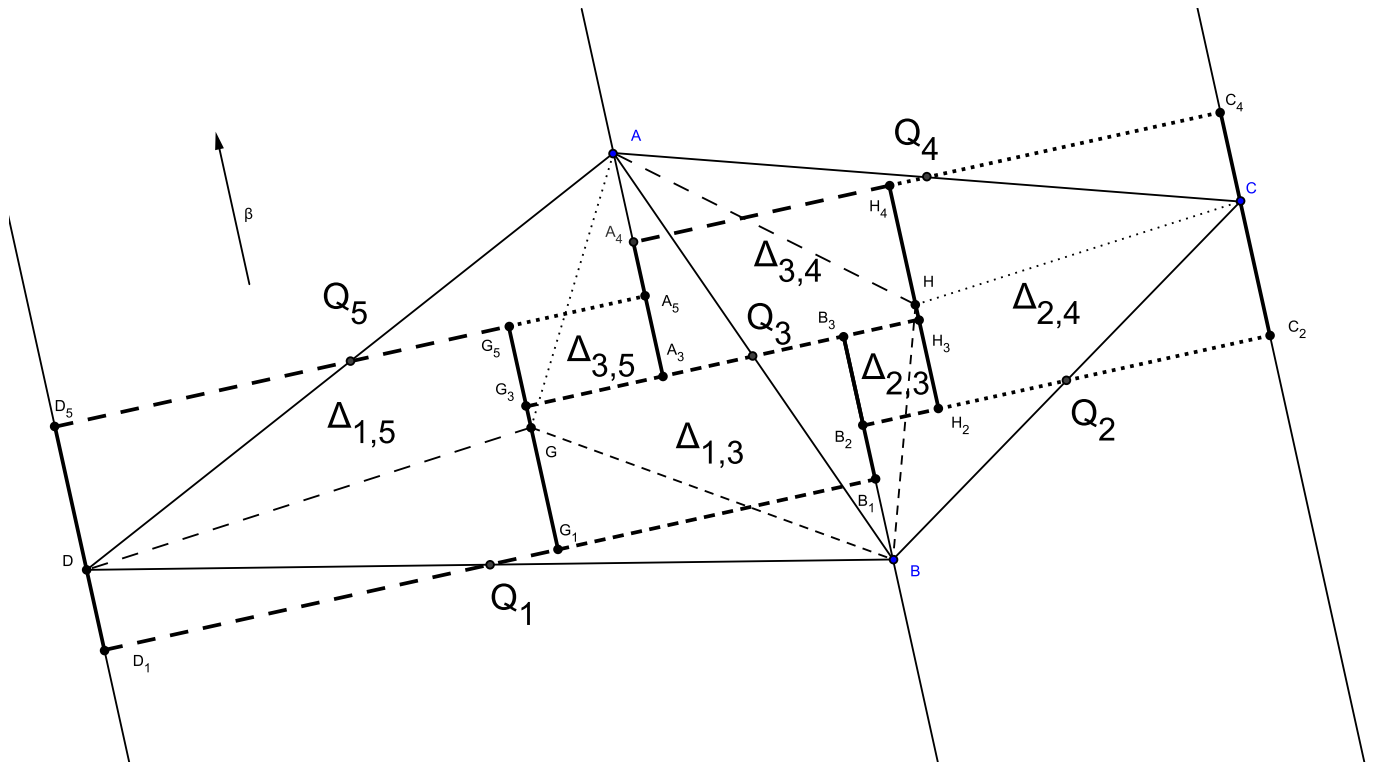
$$\begin{aligned} |u(x)| &\leq |g \circ P(x)| + \int_0^{\frac{(x-A_i) \cdot n_i}{\beta \cdot n_i}} |f(P(x) + \beta \tau)| d\tau \\ &\leq \|g\|_{L^\infty(\Gamma^-)} + \frac{\text{diam}(\Omega)}{\|\beta\|} \|f\|_{L^\infty(\Omega)} \end{aligned}$$

□

## C Quelques propriétés de $d_k = \frac{1}{\|\beta\|^2} \sum_{j \in S^-(k)} [\beta \cdot (Q_k - Q_j)] \theta_{j,k}$

### C.1 Interprétation géométrique de $d_k$

Il est possible de donner une interprétation géométrique à  $d_k$ , et plus particulièrement à  $\Delta_{k,l} = \frac{1}{\|\beta\|^2} \theta_{k,l}^\beta [\beta \cdot (Q_l - Q_k)]$ . Pour cela considérons la situation géométrique typique suivante :



62. Les fonctions  $g$  de  $W^{1,1}(\Gamma^-)$  sont continues (ou plus exactement la classe d'équivalence correspondant à  $g$  possède un représentant  $\hat{g}$  continu), ce qui justifie  $\|g\|_{L^\infty(\Gamma^-)}$ . Voir par exemple [Bre83], page 122.

Deux mailles (les triangles  $ABC$  et  $ABD$  ont un côté  $[AB]$  en commun. Les points  $Q_1, Q_2, Q_3, Q_4$  et  $Q_5$  sont les milieux respectivement de  $[BD], [BC], [BA], [CA]$  et  $[AD]$ . Dans la situation ci-dessus  $S^-(3) = \{1, 2\}$  (les points  $Q_1$  et  $Q_2$  sont en amont de  $Q_3$ ),  $S^-(5) = \{1, 3\}$  (les points  $Q_1$  et  $Q_3$  sont en amont de  $Q_4$ ) et  $S^-(3) = \{3, 2\}$  (les points  $Q_2$  et  $Q_3$  sont en amont de  $Q_4$ ). Les points  $G$  et  $H$  sont respectivement les centres de gravité de  $ABC$  et  $ABD$ .

En terme de schéma de calcul itératif pour la solution  $u_h$ , cela implique que  $u_h(Q_3), u_h(Q_4)$  et  $u_h(Q_5)$  se calculent à partir de  $u_h(Q_1)$  et  $u_h(Q_2)$  (et  $f_h$ ) uniquement. Cette situation est étudiée plus systématiquement ci-après sous la notion d'encapsulation de l'information : Les deux triangles  $ABC$  et  $ABD$  forment une unité qui sera appelée cellule, pour laquelle  $Q_1$  et  $Q_2$  sont des points de contrôle (ou points d'entrées). Le lemme C.1, page 199, appliqué à notre exemple, affirme que la somme des  $d_k$  pour tous les points  $Q_k$ , qui ne sont pas des points de contrôle, est égale à la mesure de la cellule :

$$|ABC| + |ABD| = d_3 + d_4 + d_5$$

Dans la figure ci-dessus, les droites  $(D_1D_5), (A_3A), (H_4H_2)$  et  $(C_4C_2)$  ont la même direction que le vecteur  $\beta$ . Les segments en pointillés sont perpendiculaires à ces droites.

Soit  $\Delta_{k,l} = \frac{1}{\|\beta\|^2} \theta_{k,l}^\beta [\beta \cdot (Q_l - Q_k)]$ , avec  $(k, l) \in \{(1; 5), (1; 3), (2; 3), (3; 5), (3; 4), (2; 4)\}$ .

Pour  $k = 1$  et  $l = 3$  on a  $\theta_{1,3}^\beta = \int_{[GB]} \beta \cdot n_{1,3} d\sigma = \|\beta\| |G_1B_1|$  et  $\beta \cdot (Q_3 - Q_1) = \|\beta\| |G_1G_3|$ , donc  $\Delta_{1,3} = \frac{1}{\|\beta\|^2} \theta_{1,3}^\beta [\beta \cdot (Q_3 - Q_1)] = G_1B_1 \times G_1G_3 = |G_1B_1B_3G_3|$ .

En procédant de façon identique pour les autres valeurs de  $\Delta_{k,l}$ , on obtient :

$$\begin{aligned} d_3 + d_4 + d_5 &= (\Delta_{1,3} + \Delta_{2,3}) + (\Delta_{2,4} + \Delta_{3,4}) + (\Delta_{1,5} + \Delta_{3,5}) \\ &= (|G_1B_1B_3G_3| + |B_2H_2H_3B_3|) + (|H_2C_2C_4H_4| + |A_2H_3H_4A_4|) \\ &\quad + (|D_1G_1G_5D_5| + |G_3A_3A_5G_5|) \\ &= |D_1D_5A_5A_4C_4C_2B_2B_1| \\ &= |ABC| + |ABD| \end{aligned}$$

## C.2 Encapsulation de l'information dans $\mathcal{G}_h$ et forme linéaire $(\cdot, \cdot)_{\Pi}$

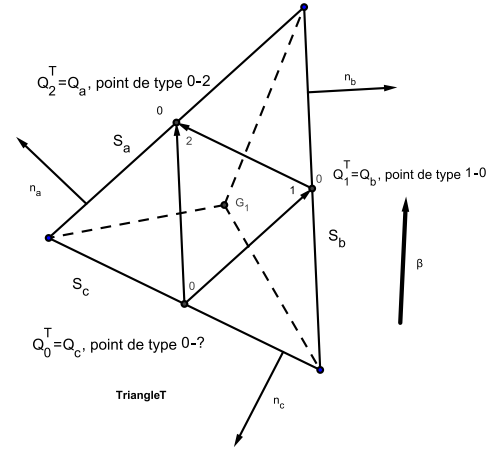
L'objectif est d'établir un lien entre le graphe  $\mathcal{G}_h$ , le maillage  $\mathcal{T}_h$  et la forme linéaire  $v_h \mapsto (f_h, v_h)_{\Pi}$ .

### Relation entre $d_k$ et $|C|$ :

Avant de formuler le lemme général qui relie  $d_k$  et  $|C|$ , deux exemples pour illustrer l'utilisation des coordonnées locales dans le calcul.

#### Exemple C.1 (Exemple 1)

Exemple 1 : C=T



La cellule  $C$  n'est constituée que d'un triangle  $T$  (cela peut se produire pour un "coin" de  $\Omega$ , comme on peut le voir sur la figure 17 en haut à droite).  $Q_a = Q_2^T$ ,  $Q_b = Q_1^T$ ,  $Q_c = Q_0^T$  et  $J^C = \{a, b, c\}$ .  $S^-(a) = \{b, c\}$ ,  $S^-(b) = \{c\}$  et  $S^-(c) \cap J^C = \emptyset$ . Donc  $J^C_{entrant} = \{c\}$ ,  $J^C_{sortant} = \{a, b\}$ . Maintenant  $\|\beta\|^2 \sum_{i \in J^C \setminus J^C_{entrant}} d_i = d_a + d_b$

$$= \underbrace{\theta_{0,2}^T [\beta \cdot (Q_2^T - Q_0^T)]}_{=d_a} + \underbrace{\theta_{1,2}^T [\beta \cdot (Q_2^T - Q_1^T)] + \theta_{0,1}^T [\beta \cdot (Q_1^T - Q_0^T)]}_{=d_b}$$

$$= \theta_{0,2}^T [\beta \cdot Q_2^T] - \theta_{0,2}^T [\beta \cdot Q_0^T] + \theta_{1,2}^T [\beta \cdot Q_2^T] - \theta_{1,2}^T [\beta \cdot Q_1^T] + \theta_{0,1}^T [\beta \cdot Q_1^T] - \theta_{0,1}^T [\beta \cdot Q_0^T]$$

$$= (\theta_{0,2}^T + \theta_{1,2}^T) [\beta \cdot Q_2^T] - (\theta_{0,2}^T + \theta_{0,1}^T) [\beta \cdot Q_0^T] - \theta_{1,2}^T [\beta \cdot Q_1^T] + \theta_{0,1}^T [\beta \cdot Q_1^T]$$

$\theta_{0,2}^T + \theta_{1,2}^T = \int_{S_a} \beta \cdot n_a d\sigma$ ,  $-(\theta_{0,2}^T + \theta_{0,1}^T) = \int_{S_c} \beta \cdot n_c d\sigma$  et  $-\theta_{1,2}^T + \theta_{0,1}^T = \theta_{2,1}^T + \theta_{0,1}^T = \int_{S_b} \beta \cdot n_b d\sigma$   
La fonction  $Q \mapsto \beta \cdot Q$  est linéaire, donc

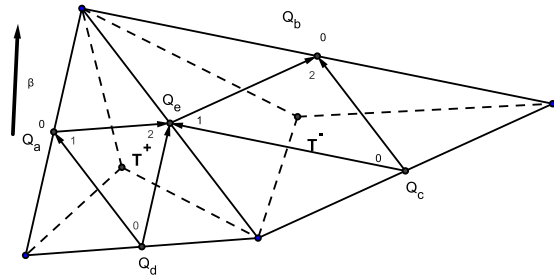
$$[\beta \cdot Q_a] \int_{S_a} \beta \cdot n_a d\sigma = |S_a| [\beta \cdot Q_a] \beta \cdot n_a = \int_{S_a} [\beta \cdot Q] \beta \cdot n_a d\sigma$$

La même relation existe pour les côtés  $S_b$  et  $S_c$ , et on a

$$\|\beta\|^2 \sum_{i \in J^C \setminus J^C_{entrant}} d_i = \int_{\partial C} [\beta \cdot Q] \beta \cdot n d\sigma = \int_C \nabla \cdot ([\beta \cdot Q] \beta) dx = \|\beta\|^2 |C| \text{ donc } \sum_{i \in J^C \setminus J^C_{entrant}} d_i = |C|$$

**Exemple C.2 (Exemple 2)**

Exemple 2 : C=T<sup>+</sup> ∪ T<sup>-</sup>



La cellule  $C$  est constituée de la réunion de deux triangles  $T^+$  et  $T^-$ . En procédant comme dans l'exemple 1, on obtient  $Q_a = Q_1^{T^+}$ ,  $Q_b = Q_2^{T^-}$ ,  $Q_c = Q_0^{T^-}$ ,  $Q_d = Q_0^{T^+}$  et  $Q_e = Q_2^{T^+} = Q_1^{T^-}$ , donc  $J^C = \{a, b, c, d, e\}$  et  $J^C_{entrant} = \{c, d\}$ ,  $J^C_{sortant} = \{a, b\}$ ,  $J^C_{intérieur} = \{e\}$ .

$$\|\beta\|^2 \sum_{i \in J^C \setminus J^C_{entrant}} d_i = d_a + d_b + d_e =$$

$$\begin{aligned}
&= \underbrace{\theta_{0,1}^{T^+} [\beta \cdot (Q_1^{T^+} - Q_0^{T^+})]}_{=d_a} + \underbrace{\theta_{1,2}^{T^-} [\beta \cdot (Q_2^{T^-} - Q_1^{T^-})]}_{=d_b} + \theta_{0,2}^{T^-} [\beta \cdot (Q_2^{T^-} - Q_0^{T^-})] + \\
&+ \underbrace{\theta_{0,1}^{T^-} [\beta \cdot (Q_1^{T^-} - Q_0^{T^-})]}_{=d_c} + \theta_{0,2}^{T^+} [\beta \cdot (Q_2^{T^+} - Q_0^{T^+})] + \theta_{1,2}^{T^+} [\beta \cdot (Q_2^{T^+} - Q_1^{T^+})] \\
&= (\theta_{0,1}^{T^+} - \theta_{1,2}^{T^+}) [\beta \cdot Q_1^{T^+}] + (-\theta_{0,1}^{T^+} - \theta_{0,2}^{T^+}) [\beta \cdot Q_0^{T^+}] + (\theta_{0,2}^{T^+} + \theta_{1,2}^{T^+}) [\beta \cdot Q_2^{T^+}] \\
&+ (\theta_{0,1}^{T^-} - \theta_{1,2}^{T^-}) [\beta \cdot Q_1^{T^-}] + (-\theta_{0,2}^{T^-} - \theta_{0,1}^{T^-}) [\beta \cdot Q_0^{T^-}] + (\theta_{1,2}^{T^-} + \theta_{0,2}^{T^-}) [\beta \cdot Q_2^{T^-}]
\end{aligned}$$

Maintenant :

$$\begin{aligned}
(\theta_{0,1}^{T^+} - \theta_{1,2}^{T^+}) [\beta \cdot Q_1^{T^+}] &= (\theta_{0,1}^{T^+} + \theta_{2,1}^{T^+}) [\beta \cdot Q_1^{T^+}] = \int_{S_a} [\beta \cdot Q] \beta \cdot n_a \, d\sigma \\
(-\theta_{0,2}^{T^+} - \theta_{0,1}^{T^+}) [\beta \cdot Q_0^{T^+}] &= \int_{S_d} [\beta \cdot Q] \beta \cdot n_d \, d\sigma
\end{aligned}$$

$Q_e = Q_2^{T^+} = Q_1^{T^-}$  donc

$$(\theta_{0,2}^{T^+} + \theta_{1,2}^{T^+}) [\beta \cdot Q_2^{T^+}] + (\theta_{0,1}^{T^-} - \theta_{1,2}^{T^-}) [\beta \cdot Q_1^{T^-}] = \underbrace{(\theta_{0,2}^{T^+} + \theta_{1,2}^{T^+} + \theta_{0,1}^{T^-} + \theta_{2,1}^{T^-})}_{-(\theta_{e,a} + \theta_{e,b} + \theta_{e,c} + \theta_{e,a})} [\beta \cdot Q_e] = 0$$

Enfin

$$(\theta_{1,2}^{T^-} + \theta_{0,2}^{T^-}) [\beta \cdot Q_2^{T^-}] = \int_{S_b} [\beta \cdot Q] \beta \cdot n_b \, d\sigma$$

Conclusion :

$$\|\beta\|^2 \sum_{i \in J^C \setminus J_{\text{entrant}}^C} d_i = \int_{\partial C} [\beta \cdot Q] \beta \cdot n \, d\sigma = \int_C \nabla \cdot ([\beta \cdot Q] \beta) \, dx = \|\beta\|^2 |C| \text{ donc } \sum_{i \in J^C \setminus J_{\text{entrant}}^C} d_i = |C|$$

Le lemme suivant généralise ce résultat :

**Lemme C.1 (Mesure de  $|C|$ )** Soit  $\mathcal{C}_h$  l'ensemble des cellules définies à partir de  $\mathcal{T}_h$  et  $C \in \mathcal{C}_h$  une cellule qui ne contient pas de triangle ayant deux coordonnées locales 0 (voir la remarque 14.1), alors

$$\sum_{k \in J^C \setminus J_{\text{entrant}}^C} d_k = |C| \text{ avec } d_k = \frac{1}{\|\beta\|^2} \sum_{j \in S^-(k)} [\beta \cdot (Q_k - Q_j)] \theta_{j,k} \quad (3.403)$$

**Preuve :**

Soit  $\omega = \{(l, i) \in \mathcal{G}_h : l \in J^C \text{ et } i \in J^C\}$  l'ensemble des arcs  $(l, i)$  du graphe  $\mathcal{G}_h$  qui sont dans la cellule  $C$ .

Soit

$$\omega^+ = \{(i, l) \in \mathcal{G}_h : i \in J^C \setminus J_{\text{entrant}}^C \text{ et } l \in S^+(i)\} \cup \{(i, l) \in \mathcal{G}_h : i \in J_{\text{entrant}}^C \text{ et } l \in S^+(i) \cap J^C\}$$

Il s'agit donc de l'ensemble des arcs  $(i, l)$  du graphe  $\mathcal{G}_h$  qui

- partent d'un point de  $J^C \setminus J_{\text{entrant}}^C$ , et
- partent d'un point de  $J_{\text{entrant}}^C$ , mais qui restent dans  $C$ .



On a  $\omega \subset \omega^+$  et  $\omega^+ \setminus \omega$  est l'ensemble des arcs  $(i, l)$  du graphe  $\mathcal{G}_h$  avec  $i \in J_{\text{sortant}}^C$  qui quittent  $C$  ( $l \notin J^C$ ).

- Pour  $\omega^+ \setminus \omega$  :

$$\begin{aligned} \sum_{(i,l) \in \omega^+ \setminus \omega} \theta_{i,l} [\beta \cdot Q_i] &= \sum_{i \in J_{\text{sortant}}^C} \sum_{l \in S^+(i) \setminus J^C} \theta_{i,l} [\beta \cdot Q_i] \\ &= \sum_{i \in J_{\text{sortant}}^C} \left( \sum_{l \in S^+(i) \setminus J^C} \theta_{i,l} \right) [\beta \cdot Q_i] \\ &\stackrel{(14.222)}{=} \sum_{i \in J_{\text{sortant}}^C} \int_{S_i} \beta \cdot n_C d\sigma [\beta \cdot Q_i] \end{aligned} \quad (3.404)$$

- Pour  $\omega^+$  :

$$\begin{aligned} \sum_{(i,l) \in \omega^+} \theta_{i,l} [\beta \cdot Q_i] &= \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{l \in S^+(i)} \theta_{i,l} [\beta \cdot Q_i] + \sum_{i \in J_{\text{entrant}}^C} \sum_{l \in S^+(i) \cap J^C} \theta_{i,l} [\beta \cdot Q_i] \\ &\stackrel{(14.221)}{=} \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{l \in S^+(i)} \theta_{i,l} [\beta \cdot Q_i] + \sum_{i \in J_{\text{entrant}}^C} \left( - \int_{S_i} \beta \cdot n_C d\sigma \right) [\beta \cdot Q_i] \end{aligned} \quad (3.405)$$

- Pour  $\omega$  on a  $\omega = \tilde{\omega}$  avec :

$$\begin{aligned} \omega &= \{(l, i) \in \mathcal{G}_h : i \in J^C \text{ et } l \in J^C\} \\ \tilde{\omega} &= \{(k, i) \in \mathcal{G}_h : i \in J^C \setminus J_{\text{entrant}}^C \text{ et } k \in S^-(i)\} \end{aligned} \quad (3.406)$$

On effectue la preuve par double inclusion :

" $\subset$ " : Soit  $(l, i) \in \omega$ , alors  $i$  ne peut pas être dans  $J_{\text{entrant}}^C$ , car sinon  $k \in S^-(i)$  ne peut pas être<sup>63</sup> dans  $J^C$ . Ainsi  $i \in J^C \setminus J_{\text{entrant}}^C$  et  $S^-(i) \subset J^C$ , ce qui prouve " $\subset$ ".

" $\supset$ " : Soit  $(l, i) \in \{(k, i) \in \mathcal{G}_h : i \in J^C \setminus J_{\text{entrant}}^C \text{ et } k \in S^-(i)\}$ .

Si  $i \notin J_{\text{sortant}}^C$  et il n'y a rien à montrer.

Si  $i \in J_{\text{sortant}}^C$ , alors  $l \in J^C$  car  $Q_i$  est de type 0-1 ou 0-2 ce qui signifie qu'aucun point en amont de  $Q_i$  ne peut être à l'extérieur de  $C$ .

Avec " $\supset$ " et " $\subset$ " l'égalité (3.406) est démontrée.

<sup>63</sup>. On utilise ici le fait que  $i$  est alors de type 0-1 ou 0-2, ce qui signifie qu'aucun point en amont de  $Q_i$  ne peut être à l'intérieur de  $C$ .

$$\begin{aligned}
\sum_{i \in J^C \setminus J_{\text{entrant}}^C} \|\beta\|^2 d_i &= \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{k \in S^-(i)} \theta_{k,i} [\beta \cdot (Q_i - Q_k)] \\
&= \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{k \in S^-(i)} \theta_{k,i} [\beta \cdot Q_i] - \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{k \in S^-(i)} \theta_{k,i} [\beta \cdot Q_k] \\
&\quad \text{avec } \sum_{k \in S^-(i)} \theta_{k,i} = - \sum_{k \in S^-(i)} \theta_{i,k} = \sum_{k \in S^+(i)} \theta_{i,k} \text{ on a :} \\
&= \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{k \in S^+(i)} \theta_{i,k} [\beta \cdot Q_i] - \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{k \in S^-(i)} \theta_{k,i} [\beta \cdot Q_k] \\
&\stackrel{(3.406)}{=} \sum_{i \in J^C \setminus J_{\text{entrant}}^C} \sum_{k \in S^+(i)} \theta_{i,k} [\beta \cdot Q_i] - \sum_{(k,i) \in \omega} \theta_{k,i} [\beta \cdot Q_k] \\
&\stackrel{(3.405)}{=} \sum_{(i,l) \in \omega^+} \theta_{i,l} [\beta \cdot Q_i] + \sum_{i \in J_{\text{entrant}}^C} \left( \int_{S_i} \beta \cdot n_C d\sigma \right) [\beta \cdot Q_i] - \sum_{(k,i) \in \omega} \theta_{k,i} [\beta \cdot Q_k] \\
&\stackrel{(3.404)}{=} \sum_{i \in J_{\text{entrant}}^C} \int_{S_i} \beta \cdot n_C d\sigma [\beta \cdot Q_i] + \sum_{i \in J_{\text{entrant}}^C} \left( \int_{S_i} \beta \cdot n_C d\sigma \right) [\beta \cdot Q_i] \\
&= \sum_{i \in J_{\text{entrant}}^C \cup J_{\text{sortant}}^C} \int_{S_i} \beta \cdot n_C d\sigma [\beta \cdot Q_i]
\end{aligned}$$

La fonction  $Q \mapsto \beta \cdot Q$  est linéaire, donc

$$\begin{aligned}
\sum_{i \in J^C \setminus J_{\text{entrant}}^C} \|\beta\|^2 d_i &= \sum_{i \in J_{\text{entrant}}^C \cup J_{\text{sortant}}^C} \int_{S_i} [\beta \cdot Q] \beta \cdot n_C d\sigma \\
&= \int_{\partial C} [\beta \cdot Q] \beta \cdot n_C d\sigma = \int_C \nabla \cdot ([\beta \cdot x] \beta) dx = \|\beta\|^2 |C|
\end{aligned}$$

□

Si  $C$  contient un triangle ayant deux coordonnées locales 0, alors il suffit de corriger la formule du lemme est enlevant la (les) contribution(s)  $\frac{1}{\|\beta\|^2} \theta_{k,i} [\beta \cdot (Q_i - Q_k)]$  de(s) arc(s) enlevé(s). Sur l'ensemble des cellules, cette correction modifie  $|\Omega|$  de  $O(h)$  (si on considère que  $|J^-| = O(1/h)$ ).

## D Liste des symboles

$\text{card}(I)$	Cardinal de $I$
$u _E$	Restriction de $u$ à l'ensemble $E$
$\dim(V)$	Dimension de l'espace vectoriel $V$
$\delta_{ij}$	Symbole de Kronecker : $\delta_{ij} = 1$ si $i = j$ , sinon 0.
$ \Gamma ,  T $	mesure du segment $\Gamma$ , de l'ensemble mesurable $T$ .
$B_r(x)$	disque ouvert de centre $x$ et de rayon $r$ .
$\text{diam}(A) = \sup\{ x - y  \mid x, y \in A\}$	diamètre du plus petit disque contenant $A$ .

## Vecteurs et matrices

$I$	matrice identité
$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$	vecteur de $\mathbb{R}^n$
$x^t = (x_1, \dots, x_n)$	transposée du vecteur $x$
$A^t$	transposée de la matrice $A$
$x \cdot y = (x, y) = \sum_{i=1}^n x_i y_i = xy^t$	produit scalaire de $x$ et $y$
$\ x\  = (x, x)^{1/2}$	norme euclidienne de $x$

## Opérateurs différentiels

$u : \Omega \rightarrow \mathbb{R}^n$ ,  $\beta, U : \Omega \rightarrow \mathbb{R}^n$  et  $\Omega \subset \mathbb{R}^n$ .  $\alpha = (\alpha_1, \dots, \alpha_n)$  multi-indice où les  $\alpha_i$ , pour  $1 \leq i \leq n$ , sont des positifs entiers.

$ \alpha  = \sum_{i=1}^n \alpha_i$	
$\partial_i u = \frac{\partial u}{\partial x_i}$	dérivée partielle dans la $i$ -ième coordonnée
$\partial_{ij} u = \frac{\partial^2 u}{\partial x_i \partial x_j}$	dérivée seconde dans les $i$ -ième et $j$ -ième coordonnées
$\partial^\alpha = \partial_1^{\alpha_1} \dots \partial_n^{\alpha_n}$	dérivées partielles en dimension $n$
$\nabla u = (\partial_i u)_{i \in \{1, \dots, n\}} = \begin{pmatrix} \partial_1 u \\ \vdots \\ \partial_n u \end{pmatrix}$	Opérateur gradient
$\nabla \cdot U = \sum_{i=1}^n \partial_i U_i$	Opérateur divergence
$\Delta u = \sum_{i=1}^n \partial_{ii} u$	Opérateur laplacien

## Espaces utilisés

Soit  $m \geq 0$  un entier et  $p$  un réel avec  $1 \leq p \leq \infty$ , alors  $W^{m,p}(\Omega)$  est l'espace de Sobolev des fonctions  $v \in L^p(\Omega)$  telles que les dérivées partielles  $\partial^\alpha v$  (prises au sens des distributions) appartiennent à  $L^p(\Omega)$  avec  $|\alpha| \leq m$ .

Cet espace est muni de la norme :

$$\|v\|_{m,p,\Omega} = \left( \sum_{|\alpha| \leq m} \int_{\Omega} |\partial^\alpha v|^p dx \right)^{1/p} \quad \text{si } 1 \leq p < \infty$$

$$\|v\|_{m,\infty,\Omega} = \max_{|\alpha| \leq m} \left\{ \text{ess} \cdot \sup_{x \in \Omega} |\partial^\alpha v| \right\} \quad \text{si } p = \infty$$

Semi-normes :

$$|v|_{m,p,\Omega} = \left( \sum_{|\alpha|=m} \int_{\Omega} |\partial^\alpha v|^p dx \right)^{1/p} \quad \text{si } 1 \leq p < \infty$$

$$|v|_{m,\infty,\Omega} = \max_{|\alpha|=m} \left\{ \text{ess} \cdot \sup_{x \in \Omega} |\partial^\alpha v| \right\} \quad \text{si } p = \infty$$

Avec les notations :

$$\begin{aligned} \|v\|_{s,\Omega'} &= \|v\|_{s,2,\Omega'} \text{ chaque fois que } \Omega' \text{ est différent de } \Omega \\ |v|_s &= |v|_{s,2,\Omega} \\ |v|_{s,\Omega'} &= |v|_{s,2,\Omega'} \text{ chaque fois que } \Omega' \text{ est différent de } \Omega \end{aligned}$$

Si  $p = 2$ , on utilisera aussi  $H^m(\Omega)$  pour  $W^{m,2}(\Omega)$ , et  $H_0^m(\Omega)$  pour la fermeture de  $C_c^\infty(\Omega)$  (fonctions  $C^\infty$  à support compact) dans  $H^m(\Omega)$ .

$$\begin{aligned} (u, v) &= \int_{\Omega} u v dx \text{ produit scalaire dans } H^0(\Omega) (= L^2(\Omega)) \\ ((u, v)) &= \int_{\Omega} \nabla u \cdot \nabla v dx \\ (u, v)_{H^1(\Omega)} &= (u, v) + ((u, v)) \text{ produit scalaire dans } H^1(\Omega) \end{aligned}$$

Soit  $X$  un espace de Banach, alors  $C([0, T], X)$  est l'espace des fonctions continues bornées  $t \mapsto u(t) = u(t, \cdot) \in X$  muni de la norme  $\|u\|_{C([0, T], X)} = \sup_{t \in [0, T]} \|u(t)\|_X$ . L'espace  $C^n([0, T], X)$  est l'espace des fonctions dont les dérivées par rapport au temps jusqu'à l'ordre  $n$  existent et sont dans  $C([0, T], X)$ .

$L^2(]0, T[, X)$ ,  $L^\infty(]0, T[, X)$  sont les fermetures de  $C(]0, T[, X)$  par rapport à la topologie définie par les normes  $\|u\|_{L^2(]0, T[, X)} = \left( \int_0^T \|u(t)\|_X^2 dt \right)^{1/2}$ ,  $\|u\|_{L^\infty(]0, T[, X)} = \sup \text{ess}_{t \in ]0, T[} \|u(t)\|_X$ .

L'espace de Sobolev  $H^1(]0, T[, X)$  est l'ensemble des fonctions  $u \in L^2(]0, T[, X)$  telles que  $u' = \frac{du}{dt} \in L^2(]0, T[, X)$ . Cet espace est muni de la norme

$$\|u\|_{H^1(]0, T[, X)} = \left( \int_0^T \|u\|_X^2 + \|u'\|_X^2 dt \right)^{1/2}$$

Espaces utilisés dans la résolution des problèmes approchés :

Soit un  $(\mathcal{T}_h)_h$  une famille de maillage triangulaire, régulière et quasi-uniforme. On définit

$$X_h = \{v_h \in L^2(\Omega) : v_h|_T \text{ est affine } \forall T \in \mathcal{T}_h, v_h \text{ est continu en } Q_j, \forall j \in J\} \quad (4.407)$$

$X_h$  est un espace de Hilbert équipé du produit scalaire

$$(w_h, v_h)_{X_h} = (w_h, v_h) + ((w_h, v_h))_h \text{ où } (w_h, v_h) = \int_{\Omega} w_h v_h dx \quad (4.408)$$

$$\text{et } ((w_h, v_h))_h = \sum_{T \in \mathcal{T}_h} \int_T \nabla w_h \cdot \nabla v_h dx \quad (4.409)$$

On a ainsi pour la norme et la semi-norme induite par le produit scalaire :

$$\|v_h\|_{X_h} = (v_h, v_h)_{X_h}^{1/2} \text{ et } |v_h|_{X_h} = ((v_h, v_h))_h^{1/2}$$

$$V_h = \{v_h \in X_h : v_h(Q_i) = 0, \forall i \in J - J_0\} \quad (4.410)$$

est un sous-espace vectoriel de  $X_h$  pour lequel la semi-norme  $|\cdot|_{X_h}$  est une norme équivalente à  $\|\cdot\|_{X_h}$ . On note  $\|\cdot\|_{*V_h} = |\cdot|_{X_h}$  cette norme induite par  $((\cdot, \cdot))_h$  (qui est maintenant un produit scalaire sur  $V_h$ ).

Espaces  $Y_h$  pour le problème approché d'advection :

$$Y_h = \{v_h : \bar{\Omega} \rightarrow \mathbb{R} : v_h|_T \text{ est affine } \forall T \in \mathcal{T}_h, v_h \text{ est continu en } Q_j, \forall j \in J\} \quad (4.411)$$

Pour  $Y_h$  on considère différentes normes et semi-normes :

$$\|w_h\|_{L_h^p(\Omega)} = \left( \sum_{k \in J} |w_h(Q_k)|^p |D_k| \right) \text{ pour } p \in [1; \infty[ \text{ et} \quad (4.412)$$

$$\|w_h\|_{L_h^\infty(\Omega)} = \max_{k \in J} |w_h(Q_k)| \quad (4.413)$$

$$|w_h|_{Y_{p,h}} = \left( \sum_{T \in \mathcal{T}_h} \|\beta \cdot \nabla_h w_h\|_{L^p(T)}^p \right)^{1/p} \text{ pour } p \in [1; \infty[ \text{ et} \quad (4.414)$$

$$|w_h|_{Y_{\infty,h}} = \max_{T \in \mathcal{T}_h} |\beta \cdot \nabla_h w_h|_T \quad (4.415)$$

$$\|w_h\|_{Y_{p,h}} = \left( \|w_h\|_{L_h^p(\Omega)}^p + h |w_h|_{Y_{p,h}}^p \right)^{1/p} \text{ pour } p \in [1; \infty[ \text{ et} \quad (4.416)$$

$$\|w_h\|_{Y_{\infty,h}} = \max\{\|w_h\|_{L_h^\infty(\Omega)}, h |w_h|_{Y_{\infty,h}}\} \quad (4.417)$$

où  $|D_k|$  est la mesure du volume barycentrique  $D_k$  associé à  $Q_k$ .

On considère aussi le sous-espace des fonctions de  $Y_h$  qui s'annulent pour  $Q_i \in \Gamma^-$  :

$$Y_h^0 = \{w_h \in Y_h : w_h(Q_j) = 0, \forall j \in J^-\} \quad (4.418)$$

## Table des constantes

Nom	Inégalité	page
$C_{\Omega}$	$\forall v \in H_{\Sigma}^1(\Omega), C_{\Omega} \int_{\Omega} v^2 dx \leq \int_{\Omega} \nabla v \cdot \nabla v dx$	119
$K_C$	$\text{card}(C) \leq K_C \quad \forall C \in \mathcal{C}_h \text{ et } h > 0$	120
$K_{\Xi}$	$N_{\Xi} \leq K_{\Xi} \frac{1}{h}$	125
$c_{I1}$	$\ \varphi - I_h \varphi\ _{X_h} \leq c_{I1} h^k \ \varphi\ _{H^{k+1}(\Omega)}$	36
$c_{I2}$	$\ \varphi - I_h \varphi\ _{L^2(\Omega)} \leq c_{I2} h^{k+1}, \ \varphi\ _{H^{k+1}(\Omega)}$	36
$c_{I3}$	$\ I_h \varphi\ _{X_h} \leq c_{I3} \ \varphi\ _{H^1(\Omega)}, \varphi \in H^1(\Omega) \Rightarrow \ \varphi - I_h \varphi\ _{X_h} \rightarrow 0 \text{ pour } h \rightarrow 0$	36
$c_{I4}$	$\ \partial_j(\varphi - I_h \varphi)\ _{L^q(T)} \leq c_{I4}  T ^{1/q-1/p} \ \partial_j \varphi\ _{L^p(T)} \text{ pour } j = 1, 2$	37
$c_{I5}$	$\ \partial_j(\varphi - I_h \varphi)\ _{L^q(T)} \leq h c_{I5}  T ^{1/q-1/p} \sum_{i=1}^2 \ \partial_i \partial_j \varphi\ _{L^p(T)} \text{ pour } j = 1, 2$	38
$\hat{c}_*$	$\ u_h\ _{L^2(\Omega)} \leq \hat{c}_* \ u_h\ _{*V_h} (= ((u_h, u_h))_h^{1/2}), u_h \in V_h, h \in (0, h_0)$	40
$\hat{c}_0$	$h^2 \leq \hat{c}_0  T  \quad \forall T \in \mathcal{T}_h \quad \forall h \in ]0, h_0[$	29
$\hat{c}_2$	$\frac{1}{\hat{c}_2} \left( \sum_{r=1}^3 v_h(Q_r^T)^2 h_T^2 \right)^{1/2} \leq \ v\ _{L^2(T)} \leq \hat{c}_2 \left( \sum_{r=1}^3 v_h(Q_r^T)^2 h_T^2 \right)^{1/2}$	39
$\hat{c}_3$	$\ \nabla_h \rho_h(u_h)\ _2 \leq \hat{c}_3 (\ \nabla_h u_h\ _2 \ \varphi\ _{\infty} + \ u_h\ _2 \ \nabla \varphi\ _{\infty})$	40
$\sigma$	$\forall h \in ]0, h_0[ \quad \forall T \in \mathcal{T}_h \quad \sigma_T = \frac{h_T}{\rho_T} \leq \sigma$	29
$\hat{\sigma}$	$h \hat{\sigma} \leq \rho_D = \sup\{\text{diam}(B) : B \subset D\} \quad \forall D \in \mathcal{D}_h \quad \forall h \in ]0, h_0[$	97
$c_0$	$\forall h, \forall T \in \mathcal{T}_h, \frac{h}{h_T} \leq c_0$	29
$c_1$	$\frac{1}{c_1} \ v\ _{*V} \leq \ v\ _{H^1(\Omega)} \leq c_1 \ v\ _{*V} \quad \forall v \in V$	22
$c_{\hat{\theta}}$	$h \ \beta\  c_{\hat{\theta}} \leq \hat{\theta}_j \text{ pour tout point } Q_j, j \in J \text{ de type } 0-1 \text{ ou } 0-2$	128
$c_{\text{inv}}$	$\forall h > 0 \quad \forall v_h \in V_h \quad \ \nabla v_h\ _2 \leq c_{\text{inv}} \frac{1}{h} \ v_h\ _2$	68
$c_*$	$\ v\ _{L^2(\Omega)} \leq c_* \ v\ _{*V} \quad \forall v \in V$	22

# Index

## condition

- $\delta - S^+$ , 52
- $\delta - S^+$ -faible, 52, 76
- $\delta - S_\beta^+$ -faible, 52
- $\delta - S^-$ , 161
- compatibilité, 84
- taille des cellules, 118

## constante

- $C_\Omega$ , 22
- $K_\Xi$ , 124
- $K_{cell}$ , 118
- $\hat{\sigma}$ , 96
- $\hat{c}_*$  (inégalité de Friedrichs), 40
- $\hat{c}_0$ , 29
- $\hat{c}_2$  (quadrature), 39
- $\hat{c}_3$ , 40
- $\sigma$  (régularité), 29
- $c_*$ , 22
- $c_0$  (quasi-uniformité), 29
- $c_1$ , 22
- $c_{I1}$ , 36
- $c_{I2}$ , 36
- $c_{I3}$ , 36
- $c_{I4}$ , 38
- $c_{I5}$ , 38
- $c_{\hat{\theta}}$ , 127
- $c_{inv}$  (inégalité inverse), 68
- $h_0$ , 29

## coordonnées locales, 39

- coordonnées locales upwind, 42, 97
- Crouzeix-Raviart (éléments de), 30

## définition

- $\hat{\theta}_i$ , 108
- problème advection-diffusion non stationnaire approché, 47, 82
- problème advection-diffusion stationnaire approché, 47
- cellule, 107
- chaîne (ou chemin) dans  $\mathcal{G}_h$ , 98
- circuit élémentaire dans  $\mathcal{G}_h$ , 98
- circuit local dans  $\mathcal{G}_h$ , 98
- coordonnées locales, 39
- coordonnées locales upwind, 43
- fonctions auxiliaires  $w_h(Q_i, \kappa)$  et  $\mathcal{L}(Q_a, Q_b, \kappa)$ , 165

formulation variationnelle II du problème approché d'advection, 95, 162

graphe orienté  $\mathcal{G}_h = (J, \mathcal{F})$  associé à  $\mathcal{T}_h$ , 98

points d'entrée (ou de contrôle), 107

points de contrôle  $I_{\text{contrôle}}^C, I_{\text{contrôle}}^\Xi, I_{\text{contrôle}}^\Omega, I_{\text{contrôle}}^{\Gamma^-}$ , 163

points de sortie, 107

points intérieurs, 107

problème approché d'advection II, 94

problème approché d'advection I, 48

problème d'advection, 25

problème d'advection-diffusion, 23

strates, 124

taille d'une cellule, 107

types de points, 98

volumes finis barycentriques, 32

domaine polygonal borné, 21

encapsulation de l'information dans  $\mathcal{G}_h$ , 106

ensemble

$I$ , 29

$J, J_0, J^+, J^-, J^b$ , 30, 94

$\mathcal{F}$  des arcs orientés, 98

$\Gamma^b$ , 24

$\Gamma^-$ , 24

$\Gamma^+$ , 24

$\Upsilon^\Omega, \Upsilon^{\Gamma^-}, \Upsilon^\Xi, \Upsilon^\emptyset$ , 170

points de contrôle  $I_{\text{contrôle}}^C, I_{\text{contrôle}}^\Xi, I_{\text{contrôle}}^\Omega, I_{\text{contrôle}}^{\Gamma^-}$ , 163

espace

$H_0^1(\Omega)$ , 21

$H_\Sigma^1(\Omega)$ , 22

$L^p(\Omega)$ , 20

$C([0, T], X)$ , 202

$C^n([0, T], X)$ , 202

$C^{0,1}(\Gamma)$ , 21

$C^{0,1}(\Omega)$ , 21

$H^1(\Omega)$ , 20

$H^1(]0, T[, X)$ , 202

$H_0^1(\Omega)$ , 20

$H^{1/2}(\Gamma)$ , 21

$L^2(\Gamma)$ , 21

$L^2(]0, T[, X)$ , 202

$L^\infty(]0, T[, X)$ , 202

- $V = H_0^1(\Omega)$ , 22  
 $V_h$ , 31, 203  
 $X_h$ , 31, 203  
 $Y_1$ , 24  
 $Y_h, Y_h^0$ , 31, 203  
 sous-espace vectoriel  $Y_{\text{sol}} \subset Y_1$ , 25  
 Euler, 30  
 exemple  
   cellule sans point de contrôle, 164  
  
 forme bilinéaire  
    $\tilde{b}_h : (Y_h + Y_{\text{sol}}) \times Y_h^0 \rightarrow \mathbb{R}$ , 36  
    $\hat{b}_h : Y_h \times Y_h^0 \rightarrow \mathbb{R}$ , 48  
    $\tilde{b}_h : (H^1(\Omega) + X_h) \times V_h \rightarrow \mathbb{R}$ , 35  
    $b_h : V_h \times V_h \rightarrow \mathbb{R}$ , 35  
 forme linéaire  $(\cdot, \cdot)_{\Pi}$ , 95  
 formule de récurrence, 96  
 frontière  
   frontière entrante  $\Gamma^-$ , 24  
   frontière longeant  $\Gamma^b$ , 24  
   frontière sortante  $\Gamma^+$ , 24  
  
 interpolation  
   opérateur  $I_h$ , 36  
  
 lemme  
   classification des cellules, 116  
   encapsulation, 109  
   existence de circuit élémentaire dans  $\mathcal{G}_h$ , 102  
   existence de circuit local dans  $\mathcal{G}_h$ , 101  
   majoration de  $\alpha_i \in \{\alpha_i^I, \alpha_i^{II}\}$ , 96  
   minoration de  $\hat{\theta}_j$ , 127  
   minoration de  $\theta_i$ , 96  
   nombre maximal de triangles dans une cellule, 119  
   taille des cellules ne contenant aucun point intérieur de type 0–1 ou 0–2, 116  
  
 maillage  
   contraintes de disposition, 29  
   dual barycentrique, 32  
   quasi-uniforme, 29  
   régularité, 29  
 mesure  
   de Lebesgue  $dx$ , 22  
   superficielle  $d\sigma$ , 22  
  
 normale  
    $n_{i,k}$  à  $\Gamma_{i,k}$ , 33  
   extérieure  $n$ , 24  
   extérieure  $n_{\Gamma^+}, n_{\Gamma^-}, n_{\Gamma^b}$ , 24  
   extérieure  $n_K$ , 22  
  
 norme  
    $\|\cdot\|_{m,p,\Omega}$ , 201  
    $\|\cdot\|_p$ , 20  
    $\|\cdot\|_{*V} = |\cdot|_{H^1(\Omega)}$ , 22  
    $\|\cdot\|_{C([0,T],X)}$ , 202  
    $\|\cdot\|_{C^n([0,T],X)}$ , 202  
    $\|\cdot\|_{H^1([0,T],X)}$ , 202  
    $\|\cdot\|_{L^2([0,T],X)}$ , 202  
    $\|\cdot\|_{L^\infty([0,T],X)}$ , 202  
    $\|\cdot\|_{L_h^p(\Gamma^-)}$ , 38  
    $\|\cdot\|_{*V_h} = |\cdot|_{X_h}$  sur  $V_h$ , 31, 203  
    $\|\cdot\|_{Y_{p,h}}$  sur  $Y_h$ , 31, 203  
    $\|\cdot\|_{L_h^p(\Omega)}$  sur  $Y_h$ , 31, 203  
   semi-norme  $|\cdot|_{Y_{p,h}}$  sur  $Y_h$ , 31, 203  
   semi-norme  $|v|_{m,p,\Omega}$ , 202  
   semi-norme  $|\cdot|_{Var_h(\Gamma^-)}$ , 38  
  
 produit scalaire  
    $((\cdot, \cdot))$  sur  $V$ , 22  
    $((\cdot, \cdot))_h$  sur  $V_h$ , 31, 203  
    $(\cdot, \cdot)_F$  dans l'espace fonctionnel  $F$ , 20  
    $(\cdot, \cdot)_{H^1(\Omega)}$ , 20  
    $(\cdot, \cdot)_{H^{1/2}(\Gamma)}$ , 21  
    $(\cdot, \cdot)_{L^2(\Omega)}$ , 20  
    $(\cdot, \cdot)_{X_h}$ , 31, 203  
   de vecteurs de  $\mathbb{R}^s$ , 20  
  
 structure des cellules, 112  
 symbole  
    $S^-(i), S^+(i)$ , 33, 94  
    $B_r(x)$ , 200  
    $D_i$  (volume fini de  $Q_i$ ), 33  
    $I$ , 29  
    $I_{\text{contrôle}}^C, I_{\text{contrôle}}^\Xi, I_{\text{contrôle}}^\Omega$  (points de contrôle), 163  
    $J, J_0, J^+, J^-, J^b$ , 30  
    $J^C$ , 107  
    $J_{\text{entrant}}^C, J_{\text{intérieur}}^C, J_{\text{sortant}}^C$ , 107  
    $N_C$ , 107  
    $N_\Xi$ , 124  
    $Q_i$  (point), 30, 94  
    $R_0$ , 21  
    $S(i)$  (voisinage de  $Q_i$ ), 33  
    $S_i$  (segment), 30, 94  
    $\mathcal{C}_h$ , 107



$\Delta$ , 201  
 $\nabla \cdot U$ , 201  
 $\mathcal{F}$  (ensemble des arcs orientés), 98  
 $\Gamma_{i,k}$ , 33  
 $\Gamma^b$ , 24  
 $\Gamma^-$ , 24  
 $\Gamma^+$ , 24  
 $\nabla u$  gradient de  $u$ , 20  
 $\nabla_h$  (opérateur gradient local), 39  
 $S_\beta^-(i)$ ,  $S_\beta^+(i)$ ,  $\theta_{i,k}^\beta$ , 34  
 $\Upsilon^\Omega$ ,  $\Upsilon^{\Gamma^-}$ ,  $\Upsilon^\Xi$ ,  $\Upsilon^\emptyset$ , 170  
 $\text{card}(I)$ , 30, 200  
 $\gamma_0$ , 21  
 $\hat{\theta}_2^T$ ,  $\hat{\theta}_0^T$ , 126  
 $\hat{\theta}_i$ , 108  
 $\hat{\theta}_j$ , 126  
 $\lambda_{j,i}$ , 95  
 $\text{diam}(\cdot)$  diamètre, 50  
 $\dim(V)$ , 200  
 $|T|$ ,  $|D_i|$ ,  $|\Gamma_{i,k}|$  et  $|S_i|$ , 33  
 $\odot_d$  (odot), 170  
 $\omega^T$ , 128  
 $\bar{\theta}_i = \sum_{k \in S^+(i)} \theta_{i,k} = \sum_{k \in S^-(i)} \theta_{k,i} > 0$ ,  
 96  
 $\partial^\alpha$ , 201  
 $\partial_i u$ , 20  
 $\partial_{ij} u$ , 201  
 $\varpi$ , 170  
 $d_k$ , 94, 162  
 $h_T$ ,  $\rho_T$  et  $h$ , 29  
 $u|_E$ , 200  
 $w_h(Q_i, \kappa)$ ,  $\mathcal{L}(Q_a, Q_b, \kappa)$  (fonctions auxiliaires), 165  
 $\mathcal{O}_d$ , 177

théorème

$L_h^1(\Omega)$ -stabilité, 142  
 $L_h^2(\Omega)$ -stabilité, 151  
 $L_h^\infty(\Omega)$ -stabilité, 137  
 d'injection de Sobolev, 26  
 de Friedrichs, 40  
 de trace, 21  
 estimation d'erreur pour l'équation  
 d'advection diffusion non station-  
 naire, 85  
 existence et unicité du problème d'ad-  
 vection, 105  
 existence, unicité

et stabilité pour l'équation d'advec-  
 tion diffusion non stationnaire, 84  
 inégalité de Poincaré, 22  
 régularité, 83  
 stabilité advection-diffusion non sta-  
 tionnaire, 76  
 stabilité advection-diffusion station-  
 naire, 53



## Références

- [Ada75] Robert A. Adams. *Sobolev Spaces*. Academic Press, New York San Francisco London, 1975.
- [ANS01] Th. Apel, S. Nicaise, and J. Schoberl. Crouzeix-raviart type finite elements on anisotropic meshes. *Numer. Math.*, 89 :193–223., 2001.
- [AsFF98] Ph. Angot, V. Dolejší, M. Feistauer, and J. Felcman. Analysis of a combined barycentric finite volume-nonconforming finite element method for nonlinear convection-diffusion problems. *Appl. Math.*, 43 :263–310, 1998.
- [BA72] I Babuška and A Aziz. Survey lectures on the mathematical foundation of the finite element method. In *The mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations*, pages 1–359. Academic Press, New-York, 1972.
- [BGP05] Daniel Bouché, Jean-Michel Ghidaglia, and Frédéric Pascal. Error estimate and the geometric corector for the upwind finite volume method applied to the linear advection equation. *SIAM J. Numer. Anal.*, 2005.
- [BHS06] A. Buffa, T. J. R. Hughes, and G. Sangalli. Analysis of multiscale discontinuous galerkin method for convection-diffusion problems. *SIAM J. Numer. Anal.*, 44 :1420–1440, 2006.
- [BMM<sup>+</sup>06] F. Brezzi, L. D. Marini, S. Micheletti, P. Pietra, and R. Sacco. Stability and error analysis of mixed finite-volume methods for advection-dominated problems. *Comp. Math. Appl.*, 51 :681–696, 2006.
- [Bre83] Haïm Brezis. *Analyse fonctionnelle (Théorie et applications)*. Masson, Paris, 1983.
- [CH68] R. Courant and D. Hilbert. *Methoden der mathematischen Physik Bd 1 u. 2*. Springer Verlag, 1968.
- [Cho69] Gustave Choquet. *Cours de Topologie*. Dunod, deuxième édition (2000) edition, 1969.
- [Cia79] P. G. Ciarlet. *The finite element method for elliptic problems*. North-Holland, Amsterdam, 1979.
- [CSB05] Paola Causin, Ricardo Sacco, and Carlo L. Bottasso. Flux-upwind stabilization of the discontinuous petrov-galerkin formulation with lagrange multipliers for advection-diffusion problems. *ESSAIM : M2AN*, 2005.
- [DDMN98] C. Debiez, A. Dervieux, K. Mer, and B. Nkonga. Computation of unsteady flows with mixed finite volume/finite element upwind methods. *Int. J. Numer. Methods Fluids*, 27 :193–206, 1998.
- [DM10] Paul Deuring and Marcus Mildner. Error estimates for a finite element-finite volume discretization of convection-diffusion equations. *Numerical Methods for Partial Differential Equations*, october 2010.
- [DM11] Paul Deuring and Marcus Mildner. Stability of a combined finite element – finite volume discretization of convection-diffusion equations. *Applied Numerical Mathematics*, 61, june 2011.
- [EG] A. Ern and J.L. Guermond. *Theory and Practice of Finite Elements*.

- [EG06] A. Ern and J.L. Guermond. Discontinuous galerkin methods for friedrichs' systems. i. general theory. *SIAM J. Numer. Anal.*, 44(2) :753–778, 2006.
- [EGHL10] R. Eymard, T. Gallouët, R. Herbin, and J.-C. Latché. A convergent finite element - finite volume scheme for the compressible stokes problem. part ii : the isentropic case. *Math. Comp.*, 270 :649–675, 2010.
- [EHV06] R. Eymard, D. Hilhorst, and M. Vohralik. A combined finite volume-nonconforming/mixed-hybrid finite element scheme for degenerate parabolic problems. *Numer. Math.*, 105 :73–131, 2006.
- [EHV10] R. Eymard, D. Hilhorst, and M. Vohralik. A combined finite volume - finite element scheme for the discretization of strongly nonlinear convection-diffusion-reaction problems on nonmatching grids. *Numer. Meth. Partial Diff. Equ.*, 26 :612–646, 2010.
- [Fei93] M. Feistauer. *Mathematical Methods in Fluid Dynamics*. Longman Scientific & technical, Harlow, 1993.
- [FFcM97] M. Feistauer, J. Felcman, and M. Lukáčová Medvid'ová. On the convergence of a combined finite volume – finite element method for nonlinear convection-diffusion problems. *Numer. Meth. Partial Diff. Equ.*, 13 :163–190, 1997.
- [FFcMW99] M. Feistauer, J. Felcman, M. Lukáčová Medvid'ová, and G. Warnecke. Error estimates for a combined finite volume – finite element method for nonlinear convection-diffusion problems. *SIAM J. Math. Anal.*, 36 :1528–1548, 1999.
- [FFS03] M. Feistauer, J. Felcman, and I. Straskraba. *Mathematical and Computational Methods for Compressible Flow*. Clarendon Press, Oxford, 2003.
- [FSS99] M. Feistauer, J. Slavík, and P. Stupka. On the convergence of a combined finite volume – finite element method for nonlinear convection-diffusion problems. explicit schemes. *Numer. Meth. Partial Diff. Equ.*, 15 :215–235, 1999.
- [GHL09] T. Gallouët, R. Herbin, and J.-C. Latché. A convergent finite element - finite volume scheme for the compressible stokes problem. part i : the isothermal case. *Math. Comp.*, 267 :1333–1352, 2009.
- [Gri] Grisvard. *Nonsmooth boundary values problems*.
- [Gri85] P. Grisvard. *Elliptic Problems in Non-Smooth Domains*. Pitman, 1985.
- [GS99] P. M. Gresho and R. L. Sani. *Incompressible flow and the finite element method. Volume one : advection-diffusion*. Wiley, Chichester e.a, 1999.
- [HRB97] L. Hallo, C. Le Ribault, and M. Buffat. An implicit mixed finite-volume-finite-element method for solving 3d turbulent compressible flows. *Int. J. Numer. Meth. Fluids*, 25 :1241–1261, 1997.
- [HY] Po-Wen Hsieh and Suh-Yuh Yang. On efficient least-squares finite element methods for convection-dominated problems. *Comp. Meth. Appl. Mech. Eng.*, 199 :2009.
- [Ike83] T. Ikeda. *Maximum principle in finite element models for convection-diffusion phenomena*. North Holland, Amsterdam, 1983.
- [JNP84] C. Johnson, U. Navert, and J. Piktäranta. Finite element methods for linear hyperbolic problems. *Comp. Methods Appl. Mech. Eng.*, 45 :285–312, 1984.

- [JP86] C. Johnson and J. Piktäranta. An analysis of the discontinuous galerkin method for a scalar hyperbolic equation. *Mathematics of Computation*, 46 :1–26, 1986.
- [Li09] Zhitao Li. Convergence analysis of an upwind mixed element method for advection diffusion problems. *Appl. Math. Comp.*, 212 :318–326, 2009.
- [LL91] L.D. Landau and E.M. Lifschitz. *Lehrbuch der theoretischen Physik Bd VI*. Akademie Verlag, 1991.
- [LM68a] J.-L. Lions and E. Magènes. *Problèmes aux Limites non Homogènes et Applications, 1*. Dunod, Paris, 1968.
- [LM68b] J.-L. Lions and E. Magènes. *Problèmes aux Limites non Homogènes et Applications, 2*. Dunod, Paris, 1968.
- [LR74] P. Lesaint and P.A. Raviart. On a finite element method for solving the neutron transport equation. In in C. de Boor (Ed), editor, *Mathematical Aspects of Finite Elements in Partial Differential Equations*. Academic Press, New York, 1974.
- [Mic96] Robert C. Rogers Michael Renardy. *An Introduction to Partial Differential Equations*. Springer, 1996.
- [M.S] M.Stynes. Numerical methods for convection-diffusion problems or the 30 years war. In *20th Biennial Conference on Numerical Analysis*.
- [Nec67] J. Necas. *Les méthodes directes en théorie des équations elliptiques*. Masson, 1967.
- [OU84] K. Ohmori and T. Ushijima. A technique of upstream type applied to a linear nonconforming finite element approximation of convective diffusion equations. *RAIRO An. num.*, 18 :309–332, 1984.
- [QV97] A. Quateroni and A. Valli. *Numerical Approximation of Partial Differential Equations*. Springer, 1997.
- [RST08] H.-G. Roos, M. Stynes, and L. Tobiska. *Robust numerical methods for singularly perturbed differential equations (2nd ed.)*, volume 24 of *Springer Series in Computational Mathematics*. Springer, New York e.a., 2008.
- [Rud95] Walter Rudin. *Analyse fonctionnelle*. Ediscience, 1995.
- [Rui07] Hongxing Rui. Convergence of an upwind control-volume mixed finite element method for convection-diffusion problems. *Computing*, 81, 2007.
- [Ser01] Denis Serre. *Les Matrices*. Dunod, 2001.
- [sFcS09] V. Dolejší, M. Feistauer, V. Kučera, and V. Sobotíková.  $l^\infty(l^2)$ -error estimates for the dgfem applied to convection-diffusion problems on nonconforming meshes. *J. Numer. Math.*, 17 :45–65, 2009.
- [sFF99] V. Dolejší, M. Feistauer, and J. Felcman. On the discrete friedrichs inequality for nonconforming finite elements. *Numer. Funct. Anal. Optim.*, 20 :437–447, 1999.
- [sFFK02] V. Dolejší, M. Feistauer, J. Felcman, and A. Klikova. Error estimates for barycentric finite volumes combined with nonconforming finite elements applied to nonlinear convection-diffusion problems. *Applications of Mathematics*, 47 :301–340, 2002.
- [Tem79] R. Temam. *Navier-Stokes Equations*. Amsterdam New-York Oxford, 1979.

- [Tid02] M.D. Tidriri. Analysis of the hybrid finite element/finite volume methods for linear hyperbolic and convection-dominated convection-difusion problems. *Journal of computational and applied mathematics*, 139 :323–350, 2002.
- [Tid03] M.D. Tidriri. Error estimates for the hybrid finite element/finite volume methods for linear hyperbolic and convection-dominated problems. *Journal of computational and applied mathematics*, 156 :77–92, 2003.
- [WL03] Song Wang and Zi-Cai Li. A nonconforming combination of the finite element and volume methods with an anisotropic mesh refinement for a singularly perturbed convection-diffusion equation. *Math. Comp.*, 72 :1689–1709, 2003.
- [Wlo87] J. Wloka. *Partial differential equations*. Cambridge University Press, Cambridge e.a., 1987.