Schémas numériques d'ordre élevé en temps et en espace pour l'équation des ondes

Cyril Agut

ACADÉMIE DE BORDEAUX |_|_|_|_|_|_|_|_|_|_|

THÈSE DE DOCTORAT

présentée à

L'Université de Pau et des Pays de l'Adour
École doctorale des sciences et leurs applications - ED 211

par

Cyril AGUT

pour obtenir le grade de

DOCTEUR de l'Université de Pau et des Pays de l'Adour

Spécialité : Mathématiques Appliquées

SCHÉMAS NUMÉRIQUES D'ORDRE ÉLEVÉ EN ESPACE ET EN TEMPS POUR L'ÉQUATION DES ONDES

---

soutenue le 13 Décembre 2011

Après avis de :

M. Rémi ABGRALL	Professeur - Université de Bordeaux 1	Rapporteur
M. Philippe CHARTIER	Directeur de Recherche INRIA - INRIA Rennes Bretagne Atlantique	Rapporteur

Devant la commission d'examen formée de :

M. Rémi ABGRALL	Professeur - Université de Bordeaux 1	Rapporteur
M. Francois AUDEBERT	Géophysicien-Ingénieur R&D - centre CSTJF TOTAL	Examinateur
Mme. Hélène BARUCQ	Directrice de Recherche INRIA - INRIA Bordeaux Sud-Ouest	Directrice de thèse
M. Christophe BERTHON	Professeur - Université de Nantes	Examinateur
M. Philippe CHARTIER	Directeur de Recherche INRIA - INRIA Rennes Bretagne Atlantique	Rapporteur
M. Julien DIAZ	Chargé de Recherche INRIA - INRIA Bordeaux Sud-Ouest	Directeur de thèse
M. Sébastien PERNET	Chargé de Recherche - CERFACS Toulouse	Examinateur

Equipe Projet INRIA MAGIQUE-3D, Institut National de Recherche en Informatique et en Automatique (INRIA)
Laboratoire de Mathématiques et de leurs Applications de Pau, Unité Mixte de Recherche CNRS 5142, Université de Pau et des Pays de l'Adour (UPPA)

Remerciements

Je veux ici remercier toutes les personnes qui ont contribué de près ou de loin à l'aboutissement de ce travail.

Je tiens tout d'abord à remercier mes directeurs de thèse Hélène Barucq et Julien Diaz. Je les remercie de m'avoir orienté vers ce sujet lors de mon master ainsi que pour la confiance qu'ils m'ont accordée durant toutes ces années. Merci également pour leur disponibilité et leurs précieux conseils qui m'ont permis de progresser un peu plus chaque jour. Je retiendrai également les bons moments que l'on a pu avoir aux détours de diverses conférences (Vancouver, Concepcion,...).

Je remercie vivement Messieurs Rémi Abgrall et Philippe Chartier pour la lecture attentive et très constructive de ce manuscrit. Je les remercie pour l'intérêt qu'ils ont porté à mon travail ainsi que pour toutes leurs remarques qui ont sans aucun doute contribué à améliorer la qualité de ce manuscrit.

Je suis très reconnaissant envers Monsieur Christophe Berthon de m'avoir fait l'honneur de présider mon jury de thèse.

Je remercie Messieurs François Audebert et Sébastien Pernet d'avoir accepté de faire partie de mon jury de thèse et de l'intérêt qu'ils ont porté à mes travaux.

Je voudrais aussi remercier toutes les personnes de Magique 3D, ceux qui forment actuellement l'équipe comme les anciens membres. La liste étant trop longue, je ne citerai pas tout le monde ici mais sachez que je vous remercie pour tous ces moments partagés.

J'ai une pensée pour tous mes collègues et amis du basket et du foot. Ces moments avec vous ont été un défouloir sans pareil.

Je pense également aux petits jeunes. Mila et Vanessa, quand est-ce que vous nous enflammez le dancefloor à nouveau ? Jéjé, toi qui était mon étudiant, merci pour toutes ces blagues et histoires si... recherchées ! ! Je pense aussi aux trois petiots, bonne chance pour la fin et pour la suite.

Un petit mot aussi pour Pieyre et Guillaume, vous qui êtes "bien" loin aujourd'hui ! On a quand même bien rigolés tous ensemble, pourvu que ça dure...

Mon ancien et non moins illustre collègue de bureau Roro. Merci pour ta bonne humeur ainsi que tes énormes connaissances (informatique, musique, contrepèteries,...)

Je pense également à Melle Pédro... On a passé de très bons déjeuners ensemble ainsi que des parties de cartes mémorables. Merci pour tes conseils dans la dernière ligne droite.

Une pensée aussi pour notre JJ national. On a quand même bien rigolé ensemble, toutes ces danses endiablées que tu as pu nous montrer ! J'attends toujours de te voir en tenue complète de motard,.... mes plumes ne sont pas loin !

Je n'oublie pas Elies, mon guide parisien préféré... J'espère qu'on prendra encore un moment des cafés... avec un petit sandwich pourquoi pas ! Continue de nous faire rire surtout, ne change rien.

Caro, finalement je n'aurais pas répondu à pourquoi u ... J'ai vraiment apprécié tous ces instants passés avec toi (et que l'on partage encore ;-)). Comment oublier nos nombreuses es-

capades : Marseille et le petit train, Paris et notre petit Flo, le Canada... que de bons souvenirs ! !

Je ne t'oublie pas Gaga, ma "vieille" amie qui est pourtant si jeune... (et oui, je l'ai même écrit). Merci pour tous ces moments de rigolade et pour m'avoir amené un jour au basket avec toi... Je trouve qu'on est quand même de grands basketteurs :), non ? On t'attend toujours pour une petite danse, n'oublie pas...

Merci aussi à Nunie pour m'avoir soutenu et supporté (oui, on peut le dire) dans la dernière ligne droite. Je crois que notre culture générale est toujours à revoir mais bon, on s'est bien marré. Si tu as besoin d'un mécano, n'hésite pas...

Je voudrais également remercier Juanito. Quel plaisir cela a été de passer du temps en discutant (intelligemment ou pas...) avec toi et si souvent. Ce fut vraiment très sympa d'aller à la montagne ou sur la côte ensemble... à refaire et très vite. C'est tellement sympa un petit déjeuner au soleil à Irun....

J'ai aussi une grosse pensée pour mon DJ et clubber préféré. Merci à toi Bidou pour ces nombreuses sorties, pour ces voyages... ahlala, le Canada et Paris ! ! En tout cas, ce fut un plaisir de te "former" à Paris... A quand le prochain voyage Callaghan ? ?

Véro, le voyage a été long mais on est arrivé à bon port ! Je te remercie pour m'avoir toujours aidé et soutenu dans mes choix. J'espère que j'ai pu en faire autant pour toi. Comme toujours, on ne retiendra que les bons moments : un pisco sour à Concepcion, de la cannelle à Dresde parmi tant d'autres choses... A présent, je te souhaite le meilleur pour la suite. Merci pour tout.

Comment oublier Nono, mon ami, mon frère depuis si longtemps ! ! On pourrait compter depuis combien de temps mais on ne le fera pas... Merci de m'avoir écouté si souvent et pour tous les bons moments qu'on a passés et qu'on passera. Il me tarde de t'entendre râler la prochaine fois que tu perdras... comme d'habitude ;-)

Enfin, je tiens à remercier tout particulièrement mes parents sans qui je ne serai pas là aujourd'hui.... Merci de m'avoir toujours soutenu dans mes choix professionnels ou personnels, c'était plus qu'important. Les derniers mois n'ont pas dû être tout le temps très faciles, merci de m'avoir supporté. Merci pour tout et plus encore.

Je dis également merci à toutes les personnes que j'ai oublié mais qui ont contribué d'une façon ou d'une autre à l'écriture de ce manuscrit. Elles se reconnaîtront sans doute...

En conclusion, je ne sais pas si vous avez bien retenu le message mais MERCI A TOUS.

Table des matières

Introduction

Les travaux présentés dans ce manuscrit ont été réalisés au sein de l'équipe-projet INRIA Magique 3D (Modélisation avancée en géophysique 3D). La plupart des applications considérées par l'équipe nécessitent la résolution de l'équation des ondes. La simulation numérique de la propagation des ondes sismiques est essentielle pour modéliser les tremblements de terre et leurs répliques. Elle est également très utilisée pour la prospection pétrolière dont le but est de déterminer la composition précise du sous-sol à partir d'enregistrements d'ondes générées par des sources explosives qui sont réfléchies par les différentes couches géologiques. L'imagerie médicale peut aussi avoir recours à des techniques similaires pour détecter d'éventuelles anomalies à l'intérieure du corps humain. On utilise par exemple des ondes ultrasonores pour évaluer la densité des os et diagnostiquer l'ostéroporose [10]. Nous pouvons également citer l'imagerie radar qui utilise les ondes électromagnétiques pour la détection et la reconnaissance d'objets.

Toutes ces applications nécessitent une résolution précise et rapide de l'équation des ondes. Le cas de l'imagerie pétrolière est probablement l'un des plus significatifs. Sans rentrer dans les détails techniques, les méthodes les plus populaires comme la Reverse Time Migration (RTM) ou la Full Wave Inversion [16, 47] sont en effet basées sur un très grand nombre (plusieurs centaines) de résolution de cette équation. Les domaines tridimensionnels considérés étant de très grande taille et fortement hétérogènes, il est essentiel de disposer des méthodes de résolution précises et rapides. Les récents progrès en calcul scientifique avec notamment le développement de super-calculateurs rendent maintenant envisageables des simulations numériques réalistes mais ne suffisent pas pour résoudre tous les problèmes. Il faut donc encore améliorer les méthodes numériques pour gagner en précision sans augmenter les coûts de calcul (temps de calcul et place mémoire).
Le but de cette thèse est de contribuer à ces améliorations en proposant de nouveaux schémas de discrétisation de l'équation des ondes acoustiques qui peuvent être étendus sans difficulté aux équations d'ondes élastodynamiques et électromagnétiques.

De nombreuses méthodes existent déjà pour résoudre plus ou moins efficacement, suivant les cas, ce problème. La plupart des codes de calcul industriels reposent sur une approximation de la solution de l'équation des ondes par différences finies ou par éléments finis. Ces différentes techniques sont bien connues et très utilisées mais présentent parfois certains inconvénients. La méthode par différences finies est très facilement implémentable mais se base sur des grilles de calcul régulières, ce qui complique beaucoup, voire rend impossible la prise en compte de géométries quelconques ainsi que d'éventuelles hétérogénéités. Les méthodes d'éléments finis, qui permettent l'utilisation de maillages irréguliers, ne présentent pas ces inconvénients mais conduisent à une formulation implicite du problème alors qu'une formulation aux différences finies fournit une représentation explicite du champ d'onde. Une formulation aux éléments finis implique donc l'inversion d'une matrice de masse qui peut s'avérer très coûteuse dans des cas concrets et donc complexes. En effet, si on considère de très gros domaines en dimension trois, on peut être amené à

résoudre des systèmes comportant des millions, voire plus, d'inconnues et l'inversion de la matrice résultant de la discrétisation devient un réel problème. Il est bien entendu que des améliorations ont été apportées à ces deux techniques avec toujours quelque limitations. La méthode de condensation de masse [15, 37] permet ainsi de rendre diagonale la matrice de masse en utilisant une formule de quadrature adéquate rendant ainsi l'inversion triviale. Néanmoins, cette technique pénalise l'ordre de convergence du schéma car l'ordre de la formule de quadrature est faible devant l'ordre de la méthode d'éléments finis. La méthode des éléments spectraux (SEM) a permis de corriger ce défaut. En effet, en considérant une formule de quadrature de Gauss-Lobatto, on obtient une matrice de masse diagonale sans contraindre l'ordre de convergence de la méthode. Cette méthode a été développée dans le cas de la dynamique des fluides par Patera [44], elle a ensuite été utilisée en géophysique par Komatitsch et Tromp [40, 42, 41] ainsi que dans le cas de l'équation des ondes [21, 17]. Cette méthode n'est toutefois pas aussi flexible qu'une méthode d'éléments finis standards car elle s'appuie sur des maillages quadrangulaires en 2D et hexaédriques en 3D ce qui rend difficile la prise en compte de certains domaines à géométries complexes. Il est d'ailleurs difficile de trouver des mailleurs hexaédriques capables de mailler de telles géométries. Dans [20], les auteurs ont proposé une méthodologie pour pallier ce problème et considérer des éléments triangulaires en 2D mais cela pose certaines difficultés de mise en oeuvre et l'extension à la dimension trois est loin d'être triviale. A notre connaissance, elle n'a d'ailleurs pas été réalisée.
Dans les années 70, Reed et Hill [46] ont été les premiers à introduire une méthode de Galerkine Discontinue (DG) pour résoudre des équations hyperboliques. Cette méthode est une méthode par éléments finis qui utilise des fonctions de base discontinues. Elle possède ainsi les avantages des méthodes d'éléments finis et les matrices de masse sont diagonales par blocs, par construction. Elles sont donc facilement inversibles, ce qui conduit à une représentation quasi explicite de la solution. De plus, les méthodes DG peuvent être utilisées avec n'importe quel type de maillage et permettent de considérer facilement les variations des paramètres physiques à l'intérieur de chaque cellule du maillage (tout du moins pour des variations polynomiales). Elles sont aussi naturellement adaptées à la parallélisation puisque que toutes les intégrales de volume sont calculées localement et les communications entre les cellules sont assurées par des intégrales surfaciques sur chaque face des éléments. Les méthodes DG se retrouvent sous différentes formes dans la littérature et une étude complète et détaillée est proposée dans [5]. Dans cette thèse, on s'intéressera tout particulièrement à la méthode de Galerkine Discontinue avec Pénalité Intérieure (IPDG) (cf. [4]), aussi connue sous le nom de méthode avec Pénalité Interne Symétrique (SIP) [8]. Cette méthode est connue pour être stable et consistante, ce qui garantit l'ordre optimal de convergence du schéma. Cela explique pourquoi cette méthode a été successivement utilisée pour résoudre l'équation des ondes dans [33, 34] ou l'équation de Helmholtz [2, 9] pour ne citer qu'elles. En outre, dans [6] on montre que cette méthode est très prometteuse lorsqu'elle est utilisée pour la RTM, ce qui fait partie des objectifs visés par l'équipe Magique 3D. Dans [7], les auteurs ont comparé IPDG et SEM et ont conclu que les deux méthodes présentaient des performances similaires en termes de précision et de coûts de calcul, ce qui nous a conduit à privilégier IPDG car nous avons la possibilité d'utiliser des mailles triangulaires ou tétraédriques. Signalons toutefois que dans [23], les auteurs concluent que la SEM est plus performante, mais dans le cas particulier où IPDG est utilisée sur des maillages carrés.

Un point commun à toutes ces méthodes est que, pour améliorer la précision de la solution numérique, on doit considérablement réduire le pas d'espace, qui est la distance entre deux points du maillage représentant le domaine considéré. On augmente alors significativement le nombre d'inconnues du problème discret. De plus, le pas de temps, dont la valeur fixe le nombre d'itérations nécessaires pour résoudre le problème d'évolution, est lié au pas d'espace par la condition

CFL (Courant-Friedrichs-Lewy). Cette condition définit une borne supérieure pour le pas de temps de telle sorte que plus le pas d'espace est petit, plus le nombre d'itérations est important. Dans le cas 3D, le problème peut avoir plus de dix millions d'inconnues et celles-ci doivent être prises en compte à chaque itération en temps. Cependant, des méthodes numériques d'ordre élevé peuvent être utilisées pour obtenir des solutions précises en considérant de plus grands pas d'espace et de temps. Parmi toutes ces techniques, on peut trouver la Technique de l'Équation Modifiée (MET). Elle a été introduite par Shubin et Bell [50] et par Dablain [22] pour résoudre l'équation des ondes et utilisée ensuite par Cohen et Joly [19] dans le cas de domaines hétérogènes. Nous nous intéressons tout particulièrement à cette approche car elle ne nécessite de stocker la solution qu'en deux instants et car elle a l'avantage d'être conservative [35, 36]. Récemment, Joly et Gilbert dans [30] et Joly et Rodriguez dans [38] ont optimisé la Technique de l'Équation Modifiée. Néanmoins, si l'on souhaite utiliser cette méthode en considérant un ordre élevé en temps sur des cas où l'on doit considérer un grand nombre d'inconnues, celle-ci reste coûteuse à cause du nombre élevé de multiplications matricielles à prendre en compte. C'est pour cela que l'on s'est demandé s'il est possible d'optimiser cette méthode en réduisant le nombre de multiplications matricielles et en conservant, voire en améliorant, la condition CFL.

La plupart des travaux existants [21, 3, 54] appliquent la discrétisation en espace du système avant la discrétisation en temps. On se propose ici d'inverser ce processus en appliquant d'abord la discrétisation en temps via la MET puis en procédant à la discrétisation en espace. La discrétisation en temps fait apparaître des opérateurs d'ordre élevé (tels que des opérateurs p-harmoniques) et nous devons donc considérer des méthodes appropriées pour les discrétiser. Les méthodes d'éléments finis ne sont pas adaptées à cette discrétisation car elles font intervenir des fonctions de classe C^0 sur le domaine entier alors qu'ici nous voulons des fonctions de classe C^1. Les méthodes de type éléments finis d'Hermite peuvent être utilisées mais sont complexes à mettre en oeuvre en dimension deux et trois et sont limitées aux opérateurs d'ordre quatre. Les méthodes DG nous ont paru être une alternative intéressante car en plus de tous les avantages précédemment décrits elles permettent la discrétisation d'opérateurs d'ordre élevé par des fonctions discontinues aux interfaces des différentes mailles. Dans le cas de la méthode IPDG, la régularité est alors imposée faiblement via des fonctions de pénalisation judicieusement choisies et faciles à mettre en oeuvre (cf. par exemple [5, 2, 33] pour le Laplacien et [43, 51] pour l'opérateur biharmonique). Il est particulièrement intéressant de ne plus prendre en compte un grand nombre de multiplications matricielles et d'avoir recours simplement à une somme matricielle, ce qui ne représente quasiment aucun surcoût numérique. L'idée d'effectuer d'abord la discrétisation en temps n'est pas nouvelle. Elle a, par exemple, déjà été proposée dans [3] par Anné, Joly et Tran mais en discrétisant l'opérateur bilaplacien par différences finies. D'autres méthodes, comme par exemple la méthode ADER (Arbitrary high order, using high order DERivatives of polynomials) [39, 52] proposent d'effectuer la discrétisation en espace par éléments finis. Dans ce cas, les auteurs considérent une variable auxiliaire afin de ne pas avoir à considérer des opérateurs d'ordre élevé en espace.
Lorsque nous avons voulu étudier les propriétés numériques de ce nouveau schéma, nous nous sommes vite rendu compte qu'il faudrait préalablement procéder à l'analyse de la méthode IPDG. Il est bien connu que cette méthode présente deux difficultés principales : le choix du coefficient de pénalisation ainsi que son influence sur la condition CFL. En ce qui concerne le paramètre de pénalisation, plusieurs expressions sont proposées dans la littérature mais aucune analyse n'a jusqu'à maintenant permis de déterminer le choix le plus approprié. Dans [4], Arnold propose de le prendre inversement proportionnel à la longueur des arêtes du maillage (ou des surfaces des faces en 3D). Dans [33], les auteurs proposent plutôt d'utiliser les diamètres des mailles et finale-

ment, Shahbazi propose d'utiliser le rayon du cercle inscrit (resp. de la sphère inscrite en 3D) aux mailles (cf. [49]). La détermination du paramètre de pénalisation est importante puisqu'une valeur trop petite mène à des instabilités tandis qu'une valeur trop grande pénalise la condition CFL.
Dans [2], Ainsworth, Monk et Muniz ont conjecturé une valeur minimale du coefficient de pénalisation en fonction du degré polynomial des fonctions de base pour des maillages cartésiens 2D. Ils ont prouvé leur conjecture pour des degrés polynomiaux inférieurs ou égaux à 3. Il est à noter qu'aucun résultat n'a été proposé pour des maillages plus complexes (triangulaires par exemple) et aucune étude n'a encore permis d'expliciter le lien entre le coefficient de pénalisation et la condition CFL. Les schémas d'ordre élevé présentent les mêmes difficultés et nécessitent de plus l'analyse de paramètres de pénalisation supplémentaires liés aux opérateurs d'ordre élevé.

L'organisation de ce manuscrit de thèse est comme suit. Dans le premier chapitre, nous présentons une nouvelle famille de schémas d'ordre élevé en temps et en espace pour l'équation des ondes acoustiques. Nous présentons également une analyse de convergence (théorique et numérique) ainsi que des résultats numériques illustrant les performances de la méthode en dimension 2 et 3 d'espace. Les résultats présentés dans ce chapitre ont fait l'objet d'une publication [1].

Dans le chapitre deux, nous prouvons analytiquement la conjecture de Ainsworth, Monk et Muniz sur la valeur minimale du paramètre de pénalisation jusqu'à des degrés polynomiaux égaux à cinq dans le cas de maillages 1D, 2D et 3D composés respectivement de segments, carrés et de cubes. Nous proposons également une formule analytique liant la condition CFL au coefficient de pénalisation. En étendant notre analyse aux cas de mailles rectangulaires et parallélépipédiques, nous avons montré que le paramètre de pénalisation devait être inversement proportionnel au rayon des cercles inscrits des mailles.

Le but du chapitre trois est de donner une expression du paramètre de pénalisation la plus judicieuse sur des maillages plus complexes. Une étude numérique sur des maillages triangulaires réguliers confirme que le rayon du cercle inscrit est le plus adapté des trois choix présentés précédemment. Nous montrons également que la prise en compte d'autres paramètres géométriques tels que les angles des triangles permet d'améliorer ce choix. Finalement, par une étude sur des maillages triangulaires et tétraédriques quelconques nous montrons ensuite comment les différentes expressions influent sur la condition CFL.

Dans le chapitre quatre, nous étudions la stabilité du nouveau schéma avec opérateur biharmonique proposé au chapitre un. L'analyse est un peu moins complète car ce schéma fait intervenir trois paramètres de pénalisation. Il n'a donc pas été possible d'exprimer la condition CFL en fonction des trois paramètres et nous avons dû nous restreindre à des cas particuliers. Ces cas permettent cependant de comparer les performances de notre méthode à celle de l'équation modifiée.

Enfin, dans le chapitre cinq, nous nous intéressons à la p-adaptativité des nouveaux schémas avec opérateurs p-harmoniques. En effet, dans de nombreux cas pratiques, il est judicieux d'utiliser des méthodes permettant d'adapter les ordres de discrétisation en temps et en espace en divers endroits du domaine étudié. Cette technique a déjà été appliquée avec succés pour la résolution de l'équation des ondes [6, 39, 52] et permet de réduire sensiblement les coûts de calcul pour une précision souhaitée. Nous avons remarqué que nos schémas se prêtaient bien à la p-adaptativité et permettent naturellement d'adapter l'ordre à la fois en espace et en temps. Nous proposons un panel d'expériences numériques (en 1D et 2D) permettant d'analyser les propriétés d'adaptativité en espace et en temps de ces nouveaux schémas. Nous avons également comparé nos résultats à

ceux de la MET en s'appuyant sur le travail présenté dans [6]. Les premiers résultats que nous avons obtenus sont prometteurs et devraient être améliorés en utilisant également une technique de pas de temps local.

Chapitre 1

Schémas numériques d'ordre élevé en temps et en espace pour l'équation des ondes

Une résolution précise de l'équation des ondes génère des coûts de calcul importants. En effet, pour améliorer la précision de la solution numérique, on doit considérablement réduire le pas d'espace, qui est la distance entre deux points du maillage représentant le domaine considéré. On augmente donc significativement le nombre d'inconnues du problème discret. De plus, le pas de temps, dont la valeur fixe le nombre d'itérations nécessaires pour résoudre le problème d'évolution, est lié au pas d'espace par la condition CFL (Courant-Friedrichs-Lewy). La condition CFL définit une borne supérieure pour le pas de temps de telle sorte que plus le pas d'espace sera petit, plus le nombre d'itérations sera important. Dans le cas 3D, il est courant que le problème ait plus de dix millions d'inconnues et celles-ci doivent être prises en compte à chaque itération en temps. Des méthodes numériques d'ordre élevé peuvent être utilisées pour obtenir des solutions précises en considérant de plus grands pas d'espace et de temps. Récemment, Joly et Gilbert (cf. [30]) ont optimisé la Technique de l'Équation Modifiée (MET), qui avait été proposée par Shubin et Bell (cf. [50]) pour résoudre l'équation des ondes. Dans la plupart des travaux (cf. par exemple [21, 22, 50, 3]), on discrétise le système en espace avant de s'intéresser à la question de sa discrétisation en temps. On souhaite ici appliquer d'abord la discrétisation en temps via la MET ce qui entraîne l'apparition d'opérateurs d'ordre élevé (tels que des opérateurs p-harmoniques). Nous devons alors considérer des méthodes appropriées pour discrétiser ces opérateurs. Une première solution serait d'utiliser des méthodes d'éléments finis avec des fonctions de base suffisament régulières sur le domaine considéré (par exemple des fonctions de base d'Hermite pour l'opérateur biharmonique) mais ces méthodes ne sont pas facilement applicables en dimension élevée et ne permettent pas d'avoir des matrices de masse facilement inversibles. Nous avons donc plutôt privilégié les méthodes de Galerkine Discontinues qui ne nécessitent pas la continuité des fonctions de base, ni a fortiori de leurs dérivés, au travers des interfaces. Elles permettent donc l'utilisation des fonctions de base discontinues de Lagrange. Les continuités nécessaires à la discrétisation des opérateurs d'ordre élevé sont alors imposées faiblement par l'intermédiaire de fonctions de pénalisation. Parmi toutes les méthodes DG, nous nous sommes intéressés à la méthode de Galerkine discontinue avec Pénalité Intérieure (IPDG) (cf. par exemple [5, 2, 33] pour le Laplacien et [43] pour l'opérateur biharmonique).

Ce chapitre est organisé de la façon suivante. Dans la section 1.1, nous décrivons la technique classique de l'équation modifiée appliquée à l'équation des ondes semi-discrétisée en espace et

nous rappelons ses propriétés. Dans la section 1.2, nous dérivons des schémas d'ordre élevé en appliquant cette technique directement à l'équation des ondes continue et nous présentons la méthode numérique que nous avons choisie pour discrétiser les opérateurs d'ordre élevé. Nous montrons un résultat de convergence pour la méthode que nous proposons dans la section 1.3 et enfin nous présentons des résultats numériques illustrant les performances de la méthode en section 1.4.

1.1 La technique de l'Equation Modifiée

Dans cette section, nous présentons brièvement la technique de l'équation modifiée qui nous permet d'obtenir une approximation d'ordre pair en temps. Nous renvoyons à [22, 50, 30] pour plus de détails sur cette approche.
Nous considérons l'équation des ondes acoustiques dans un milieu borné hétérogène $\Omega \subset \mathbb{R}^d, d = 1, 2, 3$. Par souci de simplicité, nous imposons des conditions de bord de Dirichlet homogène sur la frontière $\Gamma := \partial\Omega$ mais cette étude peut être étendue sans difficultés majeures aux cas de conditions de bord de Neumann. Le problème s'écrit :

$$\begin{cases} \text{Trouver } u : \Omega \times [0,T] \mapsto \mathbb{R} \text{ tel que :} & \\ \dfrac{1}{\mu(x)}\dfrac{\partial^2 u}{\partial t^2} - \text{div}\left(\dfrac{1}{\rho(x)}\nabla u\right) = f & \text{dans } \Omega \times]0,T], \\ u(x,0) = u_0, \quad \dfrac{\partial u}{\partial t}(x,0) = u_1 & \text{dans } \Omega, \\ u = 0 & \text{sur } \partial\Omega. \end{cases} \tag{1.1}$$

où u désigne le déplacement, μ et ρ sont respectivement le module de compressibilité et la densité de Ω et f est le terme source suffisament régulier en temps. Nous supposons que μ et ρ sont non nuls et que $\dfrac{1}{\rho} \in W^{1,\infty}(\Omega)$ et $\dfrac{1}{\mu} \in L^\infty(\Omega)$. T représente le temps final de l'expérience, u_0 et u_1 sont les données initiales.
En considérant l'espace

$$\mathcal{D}_0(\Delta_\rho, \Omega) = \left\{ u \in H_0^1(\Omega) \,:\, \Delta_\rho = \text{div}\left(\frac{1}{\rho}\nabla u\right) \in L^2(\Omega) \right\},$$

on peut montrer, par la méthode des semi-groupes par exemple (cf. [12]), que si $f \in C^1\left(]0,T[, L^2(\Omega)\right)$, $u_0 \in \mathcal{D}_0(\Delta_\rho, \Omega)$ et $u_1 \in L^2(\Omega)$ alors le problème (1.1) admet une solution unique

$$u \in C^2\left(]0,T[, L^2(\Omega)\right) \cap C^1\left(]0,T[, H_0^1(\Omega)\right) \cap C^0\left(]0,T[, \mathcal{D}_0(\Delta_\rho, \Omega)\right). \tag{1.2}$$

1.1.1 Discrétisation

Si on applique à l'équation (1.1) une discrétisation en espace classique comme, par exemple, une méthode de type Différences Finies, Eléments Finis ou Galerkine Discontinue, on est amené à résoudre un système linéaire de la forme

$$M\frac{\partial^2 U}{\partial t^2} + KU = F, \tag{1.3}$$

où M est la matrice de masse, K la matrice de raideur, U le vecteur formé par les inconnues du problème et F le vecteur source. Dans la suite, nous supposerons que la discrétisation en espace

est telle que M est facilement inversible, c'est-à-dire que M est diagonale ou diagonale par blocs. Cela est le cas si l'on considère une méthode par Différences Finies, une méthode par Eléments Spectraux ou une méthode de Galerkine Discontinue.
L'équation (1.3) peut être facilement discrétisée par un schéma d'ordre $2p$ en utilisant un développement de Taylor d'ordre $2p$, si U est au moins de classe $C^{2(p+1)}$ en temps :

$$\frac{U\left(t+\Delta t\right)-2U\left(t\right)+U\left(t-\Delta t\right)}{\Delta t^{2}}=\sum_{i=1}^{p}c_{i}\frac{\partial^{2i}U}{\partial t^{2i}}\left(t\right)+O\left(\Delta t^{2i}\right) \tag{1.4}$$

où $c_i = \dfrac{2\Delta t^{2(i-1)}}{(2i)!}$ et Δt est le pas de temps.
Si de plus F est au moins de classe $C^{2(p-1)}$ en temps, on vérifie facilement, en utilisant (1.3), que

$$\frac{\partial^{2i}U}{\partial t^{2i}}\left(t\right)=M^{-1}\frac{\partial^{2(i-1)}F}{\partial t^{2(i-1)}}\left(t\right)-M^{-1}K\frac{\partial^{2(i-1)}U}{\partial t^{2(i-1)}}\left(t\right)=\left(-M^{-1}K\right)^{i}U\left(t\right)+\mathcal{F}_{2i}\left(t\right)$$

où $\mathcal{F}_{2i}$ est un terme source modifié tel que :

$$\begin{cases}\mathcal{F}_{2}\left(t\right)= & M^{-1}F\left(t\right)\\ \mathcal{F}_{2i}\left(t\right)= & M^{-1}\dfrac{\partial^{2(i-1)}F}{\partial t^{2(i-1)}}\left(t\right)-M^{-1}K\dfrac{\partial^{2}\mathcal{F}_{2i-1}}{\partial t^{2}}\left(t\right),i\geq2.\end{cases}$$

On peut ainsi obtenir un schéma de type Equation Modifiée d'ordre arbitraire $2p$ (MES-2p),

$$\frac{U^{n+1}-2U^{n}+U^{n-1}}{\Delta t^{2}}=\sum_{i=1}^{p}c_{i}\left(-1\right)^{i}\left(M^{-1}K\right)^{i}U^{n}+\sum_{i=1}^{p}c_{i}\mathcal{F}_{2i}\left(t^{n}\right). \tag{1.5}$$

Par la suite, par souci de simplicité, nous ne considèrerons que les cas où $1 \leq p \leq 3$, c'est-à-dire le schéma dit saute-moutons ($p = 1$) et les schémas MES-4 et MES-6 mais on peut étendre ce travail à des schémas d'ordre plus élevés sans aucune difficulté.

Remarque 1.1.1. *Le schéma MES-4 nécessite deux multiplications matricielles par $M^{-1}K$ et trois multiplications matricielles par $M^{-1}K$ pour le schéma MES-6, ce qui implique que les coûts de calculs pour une itération sont respectivement multipliés par deux et trois par rapport au schéma saute-moutons.*

1.1.2 Stabilité

Le théorème suivant ainsi que son corollaire garantissent la stabilité du schéma saute-moutons sous une certaine condition CFL.

Théorème 1.1.2. *Le schéma saute-moutons est stable si les matrices $M - \frac{\Delta t^2}{4}K$ et K sont des matrices positives.*

La démonstration de ce théorème est bien connue mais nous la rappelons ici car nous l'utiliserons fréquemment dans la suite de cette thèse.

Démonstration. Pour simplifier la présentation de la preuve, nous considèrerons la formulation (1.3) sans terme source, c'est-à-dire

$$M\frac{U^{n+1}-2U^{n}+U^{n-1}}{\Delta t^{2}}+KU^{n}=0. \tag{1.6}$$

Pour prouver ce résultat, nous allons utiliser une technique d'énergie. Pour cela, on multiplie le schéma (1.6) par $\dfrac{U^{n+1}-U^{n-1}}{2\Delta t}$ qui correspond à l'approximation centrée du terme $\dfrac{\partial u}{\partial t}$ à l'instant t^n. Ainsi, on obtient

$$\begin{aligned}\left(M\frac{U^{n+1}-2U^n+U^{n-1}}{\Delta t^2}+KU^n,\frac{U^{n+1}-U^{n-1}}{2\Delta t}\right) = \quad & \frac{1}{2\Delta t}\left(\left(M\frac{U^{n+1}-U^n}{\Delta t},\frac{U^{n+1}-U^n}{\Delta t}\right)\right.\\ & \left.-\left(M\frac{U^n-U^{n-1}}{\Delta t},\frac{U^n-U^{n-1}}{\Delta t}\right)\right)\\ & +\frac{1}{2\Delta t}\left(\left(KU^n,U^{n+1}\right)-\left(KU^{n-1},U^n\right)\right).\end{aligned} \tag{1.7}$$

On définit alors la quantité

$$E^{n+\frac{1}{2}}=\left(M\frac{U^{n+1}-U^n}{\Delta t},\frac{U^{n+1}-U^n}{\Delta t}\right)+\left(KU^n,U^{n+1}\right).$$

On peut constater que la quantité $E^{n+\frac{1}{2}}$ est conservée puisque

$$\frac{E^{n+\frac{1}{2}}-E^{n-\frac{1}{2}}}{2\Delta t}=0.$$

Ainsi $E^{n+\frac{1}{2}}=E^{\frac{1}{2}}$ où $E^{\frac{1}{2}}=\left(M\dfrac{U^1-U^0}{\Delta t},\dfrac{U^1-U^0}{\Delta t}\right)+\left(KU^0,U^1\right)$.
On cherche à présent à montrer que $E^{n+\frac{1}{2}}$ définit bien une énergie. Pour cela, on écrit $E^{n+\frac{1}{2}}$ sous la forme d'une somme de formes positives.
Si A est une matrice symétrique, on a l'égalité algèbrique

$$\frac{1}{4}\left(A\left(u+v\right),u+v\right)-\frac{1}{4}\left(A\left(u-v\right),u-v\right)=\left(Au,v\right).$$

Donc en utilisant cette égalité il vient que

$$E^{n+\frac{1}{2}}=\left(\left(M-\frac{\Delta t^2}{4}K\right)\frac{U^{n+1}-U^n}{\Delta t},\frac{U^{n+1}-U^n}{\Delta t}\right)+\left(K\frac{U^{n+1}+U^n}{2},\frac{U^{n+1}+U^n}{2}\right)$$

qui définit une énergie discrète si $M-\frac{\Delta t^2}{4}K$ et K sont deux matrices positives. Si ces conditions sont satisfaites, la stabilité du schéma sera garantie.

□

Corollaire 1.1.3. *Le schéma saute-moutons est stable sous une condition CFL (Courant-Friedrichs-Lewy),*

$$\Delta t\leq\Delta t_{LF}:=\alpha h,$$

où h est le pas d'espace du maillage et α est une constante dépendant uniquement de la méthode de discrétisation choisie ainsi que des paramètres physiques.

Démonstration. Puisque M et K sont des matrices positives, la positivité de $M-\dfrac{\Delta t^2}{4}K$ est équivalente à $\Delta t<\dfrac{2}{\sqrt{\lambda_{\max}}}$ où $\lambda_{\max}$ désigne la plus grande valeur propre de la matrice $M^{-1}K$. Il est connu que dans les méthodes de différences finies ou d'éléments finis, la plus grande valeur propre se comporte comme Ch^{-2} (cf. [30]) d'où le résultat. □

Corollaire 1.1.4. *Le schéma MES-4 est stable sous la condition CFL*

$$\Delta t \leq \Delta t_{MES-4} := \sqrt{3}\alpha h = \sqrt{3}\Delta t_{LF}.$$

Démonstration. Ce résultat a été montré dans [30]. En appliquant un raisonnement similaire à celui fait pour la preuve du théorème 1.1.2, il est clair que le schéma MES-4 est stable si les matrices $M - \dfrac{\Delta t^2}{4}K^*$ et K^* sont positives avec $K^* = K + \dfrac{\Delta t^2}{12}KM^{-1}K$.

- La positivité de $M - \dfrac{\Delta t^2}{4}K^*$ est toujours vérifiée. En effet, cela est équivalent à vérifier que $1 - \dfrac{\Delta t^2}{4}\lambda + \dfrac{\Delta t^4}{48}\lambda^2 \geq 0$ pour toute valeur propre λ de la matrice $M^{-1}K$. Or on peut réécrire, de manière équivalente, l'inégalité précédente sous la forme $\left(\Delta t^2\lambda - 6\right)^2 + 12 \geq 0$ ce qui est toujours vrai.
- Intéressons nous à présent à la positivité de K^*. Cela revient à vérifier que $M^{-1}K - \dfrac{\Delta t^2}{12}\left(M^{-1}K\right)^2 \geq 0$ puisque M^{-1} est définie positive. En se ramenant au problème aux valeurs propres, pour tout λ, valeur propre de $M^{-1}K$, on doit vérifier que $\left(\lambda - \frac{\Delta t^2}{12}\lambda^2\right) \geq 0$ ce qui revient à $\left(1 - \frac{\Delta t^2}{12}\lambda\right) \geq 0$. En effet, comme on sait que $\lambda \geq 0$, la condition est trivialement vérifiée si λ est nul, sinon on aboutit à $\left(1 - \frac{\Delta t^2}{12}\lambda_{\max}\right) \geq 0$ où $\lambda_{\max}$ est la plus grande valeur propre de $M^{-1}K$ ce qui se reformule par $\Delta t \leq \dfrac{2\sqrt{3}}{\sqrt{\lambda_{\max}}}$.
  On en déduit donc le résultat.

□

On peut également citer le résultat issu de [30] concernant le schéma MES-6 qui se démontre de la même manière que le corollaire précédent.

Corollaire 1.1.5. *Le schéma MES-6 est stable sous la condition CFL*

$$\Delta t \leq \Delta t_{MES-6} := 1.38\alpha h = 1.38\Delta t_{LF}.$$

Puisque les schémas MES-4 et MES-6 requièrent respectivement deux et trois multiplications matricielles par $M^{-1}K$ à chaque itération, les coûts de calcul s'élèvent respectivement aux coûts du schéma saute-moutons multipliés par $2/\sqrt{3} = 1.15$ et $3/1.38 = 2.17$. Le surcoût de calcul pour le schéma MES-4 est donc relativement faible tandis que celui du schéma MES-6 peut être prohibitif. Récemment, Gilbert et Joly dans [30] et Joly et Rodriguez dans [38] ont prouvé qu'il était possible d'améliorer la condition CFL de ces schémas, mais cette technique nécessite la prise en compte de multiplications supplémentaires par $M^{-1}K$ à chaque pas de temps ce qui représente au final un gain limité.

L'objet de ce chapitre est de proposer une technique permettant de diminuer le nombre de multiplications matricielles à prendre en compte. Pour cela, on propose d'adapter la technique de l'équation modifiée d'une manière originale, ce qui fera l'objet de la prochaine section.

1.2 Les Δ^p-schémas

Dans cette section, nous allons détailler la construction d'un nouveau schéma d'ordre quatre et nous présenterons brièvement celle d'un schéma d'ordre six. Néanmoins, il est à noter qu'une technique similaire pourra être utilisée pour obtenir des schémas d'ordres pairs plus élevés.

1.2.1 Le Δ^2-schéma

L'idée sur laquelle repose cette nouvelle méthode est d'inverser le processus classique de discrétisation en appliquant dans un premier temps la discrétisation en temps, via une technique de type équation modifiée, puis la discrétisation en espace. Au § 1.2.1.1, nous montrons que la discrétisation en temps d'ordre élevé provoque l'apparition d'un opérateur biharmonique qui peut être discrétisé par une méthode de Galerkine Discontinue présentée au § 1.2.1.2. La stabilité de ce nouveau schéma est discutée au § 1.2.1.3 et son coût est étudié au § 1.2.1.4.

1.2.1.1 La discrétisation en temps

Pour effectuer la discrétisation en temps de (1.1), nous considérons un développement de Taylor d'ordre quatre de la variation en temps du champ d'ondes

$$\frac{u\left(t+\Delta t\right)-2u\left(t\right)+u\left(t-\Delta t\right)}{\Delta t^2}=\frac{\partial^2 u\left(t\right)}{\partial t^2}+\frac{\Delta t^2}{12}\frac{\partial^4 u\left(t\right)}{\partial t^4}+O\left(\Delta t^4\right).$$

Comme u est solution de l'équation des ondes (1.1), nous avons

$$\frac{\partial^4 u}{\partial t^4}=\mu\text{div}\left(\frac{1}{\rho}\nabla\left[\mu\text{div}\left(\frac{1}{\rho}\nabla u\right)\right]\right)+\mu\frac{\partial^2 f}{\partial t^2}+\mu\text{div}\left(\frac{1}{\rho}\nabla\left(\mu f\right)\right).$$

Finalement, nous obtenons le schéma semi-discrétisé

$$\frac{1}{\mu}\frac{u^{n+1}-2u^n+u^{n-1}}{\Delta t^2}=\text{div}\left(\frac{1}{\rho}\nabla u^n\right)+\frac{\Delta t^2}{12}\text{div}\left(\frac{1}{\rho}\nabla\left[\mu\text{div}\left(\frac{1}{\rho}\nabla u^n\right)\right]\right)+f_4, \tag{1.8}$$

avec $f_4=f+\dfrac{\Delta t^2}{12}\left(\dfrac{\partial^2 f}{\partial t^2}+\text{div}\left(\dfrac{1}{\rho}\nabla\left(\mu f\right)\right)\right)$.

Remarque 1.2.1. *Dans le cas particulier où le domaine est homogène ($c^2=\mu/\rho$), le schéma devient :*

$$\frac{u^{n+1}-2u^n+u^{n-1}}{\Delta t^2}=c^2\Delta u^n+\frac{\Delta t^2}{12}c^4\Delta^2 u^n+\mu\left(f+\frac{\Delta t^2}{12}\left(\frac{\partial^2 f}{\partial t^2}+c^2\Delta f\right)\right). \tag{1.9}$$

Ce schéma sera appelé Δ^2-schéma et dans le cas d'un schéma d'ordre $2p$, ce dernier sera nommé Δ^p-schéma.

Remarque 1.2.2. *Les systèmes (1.8) et (1.9) sont en fait mal posés. En effet, pour tout Δt, il est possible de trouver une condition initiale telle que $|u^n|\geq Ce^{\alpha n\Delta t}$, où C est une constante strictement positive et $\alpha>0$. Cependant, une discrétisation en espace appropriée conduira à un schéma stable sous une certaine condition CFL (cf. [30]).*

L'idée d'effectuer d'abord la discrétisation en temps n'est pas nouvelle. Elle a, par exemple, déjà été proposée dans [3] par Anné, Joly et Tran. Cependant, les auteurs proposaient de discrétiser l'opérateur bilaplacien par différences finies. Le schéma obtenu est alors équivalent à celui qu'on obtient en effectuant d'abord la discrétisation en espace puis en utilisant l'équation modifiée en temps. D'autres méthodes, comme par exemple la méthode ADER (Arbitrary high order, using high order DERivatives of polynomials) [39, 52] proposent de discrétiser (1.9) par éléments finis en considérant une variable auxiliaire, afin de ne pas avoir à considérer des opérateurs d'ordre élevé,

$$\left\{\begin{array}{rcl}\dfrac{u^{n+1}-2u^n+u^{n-1}}{\Delta t^2} &=& c^2\Delta u^n+\dfrac{c^2\Delta t^2}{12}\Delta\varphi^n,\\ \varphi^n &=& c^2\Delta u^n.\end{array}\right.$$

Nous ignorons ici le terme source pour simplifier la présentation. Quelque soit la méthode d'éléments finis utilisée, on obtient alors un schéma totalement discrétisé de la forme

$$\begin{cases} M\dfrac{U^{n+1}-2U^n+U^{n-1}}{\Delta t^2} = & KU^n+\frac{\Delta t^2}{12}K\varphi^n, \\ M\varphi^n = & KU^n \end{cases}$$

qui peut se réécrire

$$M\frac{U^{n+1}-2U^n+U^{n-1}}{\Delta t^2} = KU^n+\frac{\Delta t^2}{12}KM^{-1}KU^n.$$

Le schéma ADER est donc équivalent à celui obtenu en effectuant le schéma MET classique.
L'originalité de notre approche consiste à discrétiser l'opérateur bilaplacien directement par des méthodes d'éléments finis adaptées aux opérateurs d'ordre élevé. C'est l'objet de la prochaine section.

1.2.1.2 La Discrétisation en Espace

Nous devons maintenant choisir une méthode d'éléments finis appropriée pour discrétiser l'opérateur d'ordre quatre. Si ρ et μ (ainsi que le terme source et les conditions initiales) sont assez réguliers, il est suffisant de considérer des discrétisations qui peuvent prendre en compte des quantités H^2, comme par exemple la méthode des Eléments Finis d'Hermite (HFEM). Si ρ et μ sont discontinus, alors la solution de l'équation des ondes n'est plus H^2 et cette méthode n'est donc pas appropriée. De plus, la HFEM n'est pas adaptée à la technique de condensation de masse et est relativement difficile à utiliser numériquement en dimension deux ou trois. C'est pourquoi nous proposons d'utiliser une méthode de Galerkine Discontinue avec Pénalité Intérieure (IPDG) [5, 2, 33] qui est adéquate pour considérer des milieux fortement hétérogènes grâce aux discontinuités des fonctions de base. De plus, contrairement à la HFEM, la méthode IPDG peut être facilement étendue pour discrétiser des opérateurs d'ordre plus élevé. Bien entendu, l'utilisation de cette méthode nécessite un choix judicieux des conditions de transmission entre chaque élément pour assurer la consistance de la discrétisation. Nous détaillerons ces conditions de transmission plus tard dans cette section.

Tout d'abord, introduisons une triangulation $\mathcal{T}_h$ de Ω par segments (en 1D), par triangles ou quadrilatères (en 2D) ou par tétraèdres ou hexaèdres (en 3D). L'ensemble des faces du maillage est noté $\mathcal{F}_h$ et est divisé en deux sous-espaces $\mathcal{F}_h^i$ et $\mathcal{F}_h^b$, correspondant respectivement aux faces internes et à celles situées sur la frontière du domaine. Pour $F \in \mathcal{F}_h^i$, nous noterons arbitrairement par K^+ et K^- les deux éléments partageant F et $\boldsymbol{\nu}$ la normale unitaire extérieure pointant de K^+ vers K^-.
De plus, en notant v^+ (resp. v^-) la restriction d'une fonction v à l'élément K^+ (resp. K^-), nous définissons le saut et la moyenne de v sur une face $F \in \mathcal{F}_h^i$ par

$$[\![v]\!] = v^+ - v^-, \qquad \{\!\{v\}\!\} = \frac{v^+ + v^-}{2}. \tag{1.10}$$

Pour une face extérieure $F \in \mathcal{F}_h^b$, nous définissons $[\![v]\!] = v$ et $\{\!\{v\}\!\} = v$ et $\boldsymbol{\nu}$ désigne le vecteur normal extérieur unitaire à la maille K contenant F.
Introduisons à présent l'espace d'approximation composé de fonctions polynomiales discontinues par morceaux

$$V_h := \left\{ v \in L^2(\Omega) \ : \ v_{|K} \in P^p(K), \forall K \in \mathcal{T}_h, \ p \geq 3 \right\}.$$

Remarque 1.2.3. *Remarquons que, contrairement à une méthode d'éléments finis classique, l'espace d'approximation V_h est un sous-espace de $L^2(\Omega)$ et non de $H^1(\Omega)$. Notons cependant que les fonction-test sont très régulières (polynomiales de degré p) par élément, nous effectuerons donc les intégrations sur chaque élément et non sur l'espace tout entier.*

Pour simplifier la présentation, on omettra l'élément de mesure dans les intégrales.
Par souci de simplicité, nous supposerons dans la suite que ρ et μ sont constants par éléments et nous écrivons le problème variationnel considéré de la manière suivante

$$\begin{cases} \text{Trouver } u_h^{n+1} \in V_h \text{ tel que,} \qquad \forall v \in V_h, \\ \displaystyle\sum_{K\in\mathcal{T}_h} \int_K \frac{1}{\mu} \frac{u_h^{n+1} - 2u_h^n + u_h^{n-1}}{\Delta t^2} v = \quad -a_{1h}\left(u_h^n, v\right) + \frac{\Delta t^2}{12} a_{2h}\left(u_h^n, v\right) + \sum_{K\in\mathcal{T}_h} \int_K f_4(., n\Delta t) v. \end{cases}$$

où a_{1h} et a_{2h} représentent respectivement les formes bilinéaires associées aux opérateurs d'ordre deux et quatre. La forme a_{1h} est définie par

$$a_{1h}\left(u_h^n, v\right) = B_{\mathcal{T}_{h_1}}\left(u_h^n, v\right) - \mathcal{I}_1\left(u_h^n, v\right) - \mathcal{I}_1\left(v, u_h^n\right) + B_{S1}\left(u_h^n, v\right),$$

avec

$$B_{\mathcal{T}_{h_1}}\left(u_h^n, v\right) = \sum_{K\in\mathcal{T}_h} \int_K \frac{1}{\rho} \nabla u_h^n \cdot \nabla v, \quad \mathcal{I}_1\left(u_h^n, v\right) = \sum_{F\in\mathcal{F}_h^i} \int_F [\![u_h^n]\!] \left\{\!\!\left\{ \frac{1}{\rho} \nabla v \cdot \boldsymbol{\nu} \right\}\!\!\right\},$$

et

$$B_{S_1}\left(u_h^n, v\right) = \sum_{F\in\mathcal{F}_h^i} \int_F \gamma_1 [\![u_h^n]\!] [\![v]\!].$$

La fonction de pénalisation γ_1 est introduite pour assurer la stabilité de la forme bilinéaire a_{1h}. Rappelons qu'une forme bilinéaire est stable si elle satisfait la condition de stabilité (cf. [5]) $a_{1h}(v, v) \geq C\|v\|^2, \forall v \in V_h$ avec $C > 0$. γ_1 est définie sur chaque face intérieure F par

$$\gamma_1 = \frac{\alpha_1}{\min(\xi_{F^+}, \xi_{F^-}) \min(\rho_{K^+}, \rho_{K^-})},$$

où α_1 est un paramètre positif dépendant uniquement du choix des fonctions de base de V_h. On peut trouver plusieurs expressions de la fonction ξ_F dans la littérature. Les plus fréquemment utilisées sont :

- $\xi_F = h(F)$ où $h(F)$ représente le rayon de F. Voir, par exemple, [5, 2, 34, 28]. Il est important de noter que ce choix n'a pas de sens en 1D.
- $\xi_F = \min(h(K^+), h(K^-))$ où $h(K^\pm)$ est le rayon de $K^\pm$. Cf. par exemple [33].
- $\xi_F = \min(\rho_{K^+}, \rho_{K^-})$ où $\rho_{K^\pm}$ est le rayon du cercle (ou de la sphère) inscrite dans $K^\pm$. Voir, par exemple, [49].

Quel que soit le choix de ξ_F, la coercivité de a_h est assurée pour $\alpha_1 \geq \alpha_0$. Bien entendu, le paramètre optimal α_0 dépend du choix des fonctions de base de V_h, mais aussi de ξ_F. Il a été montré par Shabazi dans [49] que le troisième choix est le plus approprié pour des maillages triangulaires.

Remarque 1.2.4. *Nous discuterons plus loin dans le manuscrit (cf. chapitre 2 et 3) du choix de cette fonction de pénalisation ainsi que du choix le plus approprié de ξ_F.*

En ce qui concerne la forme bilinéaire a_{2h}, elle est définie par

$$a_{2h}(u,v) = B_{\mathcal{T}_{h_2}}(u,v) + \mathcal{I}_2(u,v) + \mathcal{I}_2(v,u) + B_{S,2,1}(u,v) + B_{S,2,2}(u,v), \tag{1.11}$$

avec

$$\begin{cases} B_{\mathcal{T}_{h_2}}(u,v) = & \displaystyle\sum_{K\in\mathcal{T}_h}\int_K \mu \text{div}\left(\frac{1}{\rho}\nabla u\right)\text{div}\left(\frac{1}{\rho}\nabla v\right), \\ \mathcal{I}_2(u,v) = & -\mathcal{I}_{2,1}(u,v) + \mathcal{I}_{2,2}(u,v), \\ \mathcal{I}_{2,1}(u,v) = & \displaystyle\sum_{F\in\mathcal{F}_h}\int_F \left\{\left\{\mu\text{div}\left(\frac{1}{\rho}\nabla u\right)\right\}\right\}\left[\left[\frac{1}{\rho}\nabla v\cdot\boldsymbol{\nu}\right]\right], \\ \mathcal{I}_{2,2}(u,v) = & \displaystyle\sum_{F\in\mathcal{F}_h^i}\int_F \left\{\left\{\frac{1}{\rho}\nabla\left(\mu\text{div}\left(\frac{1}{\rho}\nabla u\right)\right)\cdot\boldsymbol{\nu}\right\}\right\}[\![v]\!], \\ B_{S,2,1}(u_h,v) = & \displaystyle\sum_{F\in\mathcal{F}_h}\int_F \gamma_{2,1}\left[\left[\frac{1}{\rho}\nabla u_h\cdot\boldsymbol{\nu}\right]\right]\left[\left[\frac{1}{\rho}\nabla v\cdot\boldsymbol{\nu}\right]\right], \\ B_{S,2,2}(u_h,v) = & \displaystyle\sum_{F\in\mathcal{F}_h^i}\int_F \gamma_{2,2}[\![u_h]\!][\![v]\!]. \end{cases}$$

Les fonctions de pénalisation $\gamma_{2,1}$ et $\gamma_{2,2}$ sont définies sur chaque face interne F par

$$\gamma_{2,1} = \alpha_{2,1}\frac{\max(\mu_{K^+},\mu_{K^-})}{\min(\xi_{F^+},\xi_{F^-})} \text{ et } \gamma_{2,2} = \frac{\alpha_{2,2}}{\min(\xi_{F^+}^3,\xi_{F^-}^3)}\max\left(\frac{\mu_{K^+}}{\rho_{K^+}^2},\frac{\mu_{K^-}}{\rho_{K^-}^2}\right).$$

et $\gamma_{2,1}$ est définie sur chaque face externe F par $\gamma_{2,1} = \frac{\alpha_{2,1}\mu_K}{\xi_F}$, où K est l'élément contenant F. On n'a pas besoin de définir $\gamma_{2,1}$ sur le bord car nous considérons le cas de conditions de bord de Dirichlet homogènes. Les paramètres $\alpha_{2,1}$ et $\alpha_{2,2}$ sont positifs et dépendent uniquement du choix des fonctions de base de V_h. Dans le cas d'un milieu homogène, on peut prouver (cf. [43]), en utilisant les inégalités de traces (1.31), que $\alpha_{2,1} > \alpha_{2,1}^0 \approx c_1 p^2$ et $\alpha_{2,2} > \alpha_{2,2}^0 \approx c_2 p^6$, où p désigne le degré des fonctions de base.

Intéressons nous à présent à la façon dont nous avons obtenu les formes a_{1h} et a_{2h}. La première étape consiste à multiplier (1.8) par une fonction-test v et à intégrer le résultat sur un élément K puis à sommer sur tous les éléments pour obtenir

$$\begin{aligned} \sum_{K\in\mathcal{T}_h}\int_K \frac{1}{\mu}\frac{u_h^{n+1}-2u_h^n+u_h^{n-1}}{\Delta t^2}v &= \sum_{K\in\mathcal{T}_h}\int_K \text{div}\left(\frac{1}{\rho}\nabla u_h^n\right)v + \sum_{K\in\mathcal{T}_h}\int_K f_4(\cdot,n\Delta t)\,v \\ &\quad + \frac{\Delta t^2}{12}\sum_{K\in\mathcal{T}_h}\int_K \text{div}\left(\frac{1}{\rho}\nabla\left[\mu\text{div}\left(\frac{1}{\rho}\nabla u_h^n\right)\right]\right)v \end{aligned} \tag{1.12}$$

Nous présentons tout d'abord l'utilisation de la méthode IPDG pour obtenir la forme bilinéaire correspondant à l'opérateur d'ordre deux. Cette méthode est présentée dans [5, 2, 33] et on renvoie à ces références pour plus de détails quant aux propriétés de cette forme bilinéaire.

En appliquant une formule de Green par élément sur l'opérateur d'ordre deux, nous obtenons la forme bilinéaire

$$a_{1h}(u,v) = \sum_{K\in\mathcal{T}_h}\int_K \frac{1}{\rho}\nabla u\cdot\nabla v - \sum_{K\in\mathcal{T}_h}\int_{\partial K}\left(\frac{1}{\rho}\nabla u\cdot\nu\right)v.$$

On note Q_1 le terme surfacique de a_{1h}. En sommant chaque intégrale par face, et non plus par élément, et en considérant les notations (1.10), on obtient

$$Q_1 = -\sum_{F\in\mathcal{F}_h}\int_F \left[\left[\left(\frac{1}{\rho}\nabla u\cdot\nu\right)v\right]\right].$$

En utilisant l'égalité $[\![u\,v]\!] = \{\!\{u\}\!\}\,[\![v]\!] + \{\!\{v\}\!\}\,[\![u]\!]$ sur les faces intérieures et le fait que $[\![u]\!] = \{\!\{u\}\!\} = u$ sur les face externes, on a $\forall v \in V_h$

$$Q_1 = -\sum_{F\in\mathcal{F}_h^i}\int_F \left(\left\{\left\{\frac{1}{\rho}\nabla u\cdot\nu\right\}\right\}[\![v]\!] - \{\!\{v\}\!\}\left[\left[\frac{1}{\rho}\nabla u\cdot\nu\right]\right]\right) - \sum_{F\in\mathcal{F}_h^b}\int_F \left\{\left\{\frac{1}{\rho}\nabla u\cdot\nu\right\}\right\}[\![v]\!].$$

Nous devons prendre en compte les conditions de transmission suivantes

$$\forall F\in\mathcal{F}_h^i,\ \begin{cases} [\![u]\!] &= 0 \quad \text{sur } F,\\ \left[\left[\frac{1}{\rho}\nabla u\cdot\boldsymbol{\nu}\right]\right] &= 0 \quad \text{sur } F,\end{cases} \tag{1.13}$$

qui sont faiblement satisfaites si $u \in H_0^1(\Omega)$ et $\text{div}\left(\frac{1}{\rho}\nabla u\right) \in L^2(\Omega)$ (cf. résultat (1.2)).
D'après ces conditions de transmission, il s'ensuit que

$$Q_1 = -\sum_{F\in\mathcal{F}_h}\int_F \left\{\left\{\frac{1}{\rho}\nabla u\cdot\nu\right\}\right\}[\![v]\!].$$

Ainsi,

$$a_{1h}(u,v) = B_{\mathcal{T}_{h_1}}(u,v) - \mathcal{I}_1(u,v).$$

La forme bilinéaire $a_{1h} : (u,v) \in V_h^2 \mapsto B_{\mathcal{T}_{h_1}}(u,v) - \mathcal{I}_1(u,v)$ n'est clairement pas symétrique. Nous ajoutons donc le terme $\mathcal{I}_1(v,u)$ qui ne nuit pas à la consistance de l'approximation car $\mathcal{I}_1(v,u) = 0$ d'après la seconde condition de transmission de (1.13). Cependant, quand u n'est pas continue ce qui est le cas pour l'approximation de u par des fonctions discontinues, $\mathcal{I}_1(v,u)$ ne s'annule plus et la coercivité de la forme n'est plus assurée. Nous ajoutons donc le terme $B_{S_1}(u,v)$ pour assurer la stabilité de a_{1h}.

Finalement, pour l'opérateur d'ordre deux, on obtient la forme bilinéaire

$$a_{1h}(u,v) = B_{\mathcal{T}_{h_1}}(u,v) - \mathcal{I}_1(u,v) - \mathcal{I}_1(v,u) + B_{S_1}(u,v).$$

Maintenant, considérant l'opérateur d'ordre quatre, nous appliquons une formule de Green deux fois sur chaque élément de l'équation (1.12) et nous obtenons la forme bilinéaire a_{2h} pour tout $v \in V_h$:

$$\begin{aligned} a_{2h}(u,v) = &\sum_{K\in\mathcal{T}_h}\int_K \mu\,\text{div}\left(\frac{1}{\rho}\nabla u\right)\text{div}\left(\frac{1}{\rho}\nabla v\right) - \sum_{K\in\mathcal{T}_h}\int_{\partial K}\mu\,\text{div}\left(\frac{1}{\rho}\nabla u\right)\left(\frac{1}{\rho}\nabla v\cdot\nu\right)\\ + &\sum_{K\in\mathcal{T}_h}\int_{\partial K}\frac{1}{\rho}\left(\nabla\left(\mu\,\text{div}\left(\frac{1}{\rho}\nabla u\right)\right)\cdot\nu\right)v \end{aligned}$$

Nous notons Q_2 le deuxième terme de cette expression et Q_3 le troisième. Q_2 se reformule, de la même façon que pour Q_1, en

$$Q_2 = -\sum_{F\in\mathcal{F}_h}\int_F \left[\left[\mu\,\text{div}\left(\frac{1}{\rho}\nabla u\right)\left(\frac{1}{\rho}\nabla v\cdot\boldsymbol{\nu}\right)\right]\right].$$

Utilisant à nouveau l'égalité $[\![u\,v]\!] = \{\!\{u\}\!\}\,[\![v]\!] + \{\!\{v\}\!\}\,[\![u]\!]$ sur les faces intérieures et le fait que $[\![u]\!] = \{\!\{u\}\!\} = u$ sur les face externes, on obtient, $\forall v \in V_h$

$$\begin{aligned} Q_2 = \; & -\sum_{F\in\mathcal{F}_h^i}\int_F \left(\left\{\left\{\mu\text{div}\left(\frac{1}{\rho}\nabla u\right)\right\}\right\}\left[\left[\frac{1}{\rho}\nabla v\cdot\boldsymbol{\nu}\right]\right] - \left\{\left\{\frac{1}{\rho}\nabla v\cdot\boldsymbol{\nu}\right\}\right\}\left[\left[\mu\text{div}\left(\frac{1}{\rho}\nabla u\right)\right]\right]\right) \\ & -\sum_{F\in\mathcal{F}_h^b}\int_F \left\{\left\{\mu\text{div}\left(\frac{1}{\rho}\nabla u\right)\right\}\right\}\left[\left[\frac{1}{\rho}\nabla v\cdot\boldsymbol{\nu}\right]\right]. \end{aligned}$$

Pour réécrire cette expression, nous devons considérer des conditions de transmission supplémentaires sur u. Remarquons que, si u est assez régulière en temps, les conditions de transmission (1.13) impliquent

$$\forall F\in\mathcal{F}_h^i,\ \begin{cases} \left[\left[\dfrac{\partial^2 u}{\partial t^2}\right]\right] = 0 \quad \text{sur } F, \\ \left[\left[\dfrac{1}{\rho}\nabla\dfrac{\partial^2 u}{\partial t^2}\cdot\boldsymbol{\nu}\right]\right] = 0 \quad \text{sur } F. \end{cases} \tag{1.14}$$

En utilisant l'équation des ondes (1.1), nous obtenons alors

$$\forall F\in\mathcal{F}_h^i,\ \begin{cases} \left[\left[\mu\text{div}\left(\frac{1}{\rho}\nabla u\right)\right]\right] = 0 \quad \text{sur } F, \\ \left[\left[\frac{1}{\rho}\nabla\left(\mu\text{div}\left(\frac{1}{\rho}\nabla u\right)\right)\cdot\boldsymbol{\nu}\right]\right] = 0 \quad \text{sur } F. \end{cases} \tag{1.15}$$

Remarquons à présent que si nous dérivons la condition de bord de Dirichlet deux fois par rapport au temps et que nous utilisons l'équation des ondes (1.1), on obtient une condition de bord supplémentaire

$$\mu\text{div}\left(\frac{1}{\rho}\nabla u\right) = 0 \text{ sur } F\in\mathcal{F}_h^b. \tag{1.16}$$

Ainsi, grâce à la première condition de (1.15), Q_2 s'écrit

$$Q_2 = -\sum_{F\in\mathcal{F}_h^i}\int_F \left\{\left\{\mu\text{div}\left(\frac{1}{\rho}\nabla u\right)\right\}\right\}\left[\left[\frac{1}{\rho}\nabla v\cdot\boldsymbol{\nu}\right]\right].$$

De la même manière, avec la seconde condition de (1.15), nous avons :

$$Q_3 = \sum_{F\in\mathcal{F}_h}\int_F \left\{\left\{\frac{1}{\rho}\nabla\left(\mu\text{div}\left(\frac{1}{\rho}\nabla u\right)\right)\cdot\boldsymbol{\nu}\right\}\right\}[\![v]\!].$$

Finalement, nous obtenons

$$a_{2h}(u,v) = \; B_{\mathcal{T}_{h_2}}(u,v) + \mathcal{I}_2(u,v).$$

Ici aussi, la forme bilinéaire $a_{2h} : (u_h, v)\in V_h^2 \mapsto B_{\mathcal{T}_{h_2}}(u_h,v) + \mathcal{I}_2(u_h,v)$ n'est clairement pas symétrique. Nous ajoutons donc le terme $\mathcal{I}_2(v,u_h)$ qui ne nuit pas à la consistance de l'approximation puisque les conditions de transmission (1.13) entraînent que $\mathcal{I}_2(v,u) = 0$. Cependant, quand u n'est pas continu, la forme a_{2h} n'est pas coercive. Ainsi pour assurer la stabilité de la forme, nous proposons d'ajouter les deux formes $B_{S,2,1}(u_h,v)$ et $B_{S,2,2}(u_h,v)$.
Finalement, a_{2h} est définie comme une forme symétrique et stable définie par

$$a_{2h}(u,v) = B_{\mathcal{T}_{h_2}}(u,v) + \mathcal{I}_2(u,v) + \mathcal{I}_2(v,u) + B_{S,2,1}(u,v) + B_{S,2,2}(u,v). \tag{1.17}$$

Dans un domaine homogène (*i.e.* ρ et μ constants), a_{2h} est similaire à la forme proposée par [43] pour la résolution du problème biharmonique.

Remarque 1.2.5. *Lors de toutes les expériences numériques que nous avons effectuées, nous avons choisi $\alpha_{2,2}=0$ et nous n'avons observé aucune instabilité. Cela est dû au fait que la forme $B_{S,2,2}$ est similaire à $B_{S,1}$, i.e. que le terme $B_{S,1}$ est suffisant pour assurer la stabilité des deux formes a_{1h} et a_{2h}. En effet, si nous supposons que $\alpha_{2,2}=0$, $\alpha_{2,1}>\alpha_{2,1}^0$ et $\alpha_{1,1}>\alpha_{1,1}^0$ alors, $\forall \alpha>\alpha_{2,2}^0$*

$$\begin{aligned} a_{1h}(u,v)-\frac{\Delta t^2}{12}a_{2h}(u,v) = & \quad a_{1h}(u,v)+\frac{\Delta t^2}{12}\frac{\alpha}{h^3}\sum_{F\in\mathcal{F}_h}\int_F [\![u]\!][\![v]\!] \\ & - \frac{\Delta t^2}{12}\left(a_{2h}(u,v)+\frac{\alpha}{h^3}\sum_{F\in\mathcal{F}_h}\int_F [\![u]\!][\![v]\!]\right) \\ = & \quad \widetilde{a}_{1h}(u,v)-\frac{\Delta t^2}{12}\widetilde{a}_{2h}(u,v) \end{aligned}$$

où $\widetilde{a}_{1h}$ (resp. $\widetilde{a}_{2h}$) est une forme bilinéaire dont le(s) coefficient(s) de pénalisation est (resp. sont) $\widetilde{\alpha}_{1,1}=\alpha_{1,1}+\frac{\Delta t^2}{12h^2}\alpha>\alpha_{1,1}^0$ (resp. $\widetilde{\alpha}_{2,1}=\alpha_{2,1}>\alpha_{2,1}^0$ et $\widetilde{\alpha}_{2,2}=\alpha>\alpha_{2,2}^0$). Par conséquent, la stabilité de $\widetilde{a}_{1h}$ et $\widetilde{a}_{2h}$ est assurée même si $\alpha_{2,2}=0$.

Intéressons-nous, à présent, à la consistance de la méthode. Tout d'abord, rappelons la définition suivante

Définition 1.2.6. *La méthode est consistante si les flux sont consistants. On dira que a_{1h} (resp. a_{2h}) est consistante avec l'opérateur Δ (resp. Δ^2) si $a_{1h}(u,v_h)=\int_\Omega \Delta u\, v_h$, $\forall v_h\in V_h$ (resp. $a_{2h}(u,v_h)=\int_\Omega \Delta^2 u\, v_h$, $\forall v_h\in V_h$) et pour toute fonction u de V_h vérifiant les conditions de Dirichlet et de transmission.*

Il s'en suit que, les formes a_{1h} et a_{2h} étant respectivement consistantes avec les opérateurs Δ et Δ^2, il est clair que $a_{1h}-\frac{\Delta t^2}{12}a_{2h}$ est consistante avec $\Delta+\frac{\Delta t^2}{12}\Delta^2$.

A présent, nous considérons $\{\varphi_i\}_{i=1,\dots,m}$, les fonctions de base de Lagrange discontinues de degré p de V_h, où m désigne le nombre de degrés de liberté du problème. Nous obtenons le système linéaire

$$\frac{U^{n+1}-2U^n+U^{n-1}}{\Delta t^2}+M^{-1}\left(K_1-\frac{\Delta t^2}{12}K_2\right)U^n=M^{-1}F^n, \tag{1.18}$$

où

$$(M)_{i,j}=\sum_{K\in\mathcal{T}_h}\int_K \varphi_i\varphi_j,\ (K_1)_{i,j}=a_{1h}(\varphi_i,\varphi_j),\ (K_2)_{i,j}=a_{2h}(\varphi_i,\varphi_j),$$

$$(F^n)_i=\sum_{K\in\mathcal{T}_h}\int_K f_4(\cdot,n\Delta t)\,\varphi_i.$$

La matrice de masse M est diagonale par blocs par construction et par conséquent, facilement inversible.
Les conditions initiales $U^0, U^1\in V_h$ sont données par

$$\begin{cases} U^0=P_h(u_0),\ \ V^0=P_h(v_0), \\ U^1=U^0+\Delta t V_0+\frac{\Delta t^2}{2}\widetilde{U}_0+\frac{\Delta t^3}{6}\widetilde{V}_0+\frac{\Delta t^4}{24}\widehat{U}_0 \end{cases}$$

où $P_h(u)$ est la projection L^2 de u sur V_h. $\widetilde{U}_0,\ \widetilde{V}_0,\ \widehat{U}_0 \in V_h$ sont les projections L^2 de $\dfrac{d^2u}{dt^2}(\cdot,0)$, $\dfrac{d^3u}{dt^3}(\cdot,0)$ et $\dfrac{d^4u}{dt^4}(\cdot,0)$ sur V_h. Elles sont donc telles que $\forall v \in V_h$

$$\begin{aligned}\left(\widetilde{U}_0, v\right) &= \left(\frac{d^2u}{dt^2}(\cdot,0), v\right),\\ \left(\widetilde{V}_0, v\right) &= \left(\frac{d^3u}{dt^3}(\cdot,0), v\right),\\ \left(\widehat{U}_0, v\right) &= \left(\frac{d^4u}{dt^4}(\cdot,0), v\right).\end{aligned} \tag{1.19}$$

Comme nous ne connaissons pas ces dérivées à l'intant $t = 0$, nous utilisons l'équation (1.1) pour les calculer :

$$\left(\frac{d^2u}{dt^2}(\cdot,0), v\right) = \left(\mu \text{div}\frac{1}{\rho}\nabla u(\cdot,0), v\right) + \left(f^0, v\right) = \left(\text{div}\frac{1}{\rho}\nabla u(\cdot,0), \mu v\right) + \left(f^0, v\right).$$

Comme a_{1h} est consistante avec $u \longmapsto \text{div}\frac{1}{\rho}\nabla u$, on a

$$\left(\frac{d^2u}{dt^2}(\cdot,0), v\right) = a_{1h}(u_0, \mu v) + \left(f^0, v\right)$$

et $\widetilde{U}_0$ est tel que $\forall v \in V_h$

$$\left(\widetilde{U}_0, v\right) = a_{1h}(u_0, \mu v) + \left(f^0, v\right).$$

On montre de même que, $\forall v \in V_h$

$$\begin{aligned}\left(\widetilde{V}_0, v\right) &= a_{1h}(v_0, v) + \left(\partial_t f^0, v\right),\\ \left(\widehat{U}_0, v\right) &= a_{2h}(u_0, v) + \left(\partial_t^2 f^0, v\right) + \left(\Delta f^0, v\right).\end{aligned} \tag{1.20}$$

Remarque 1.2.7. *Par souci de simplicité, nous noterons $\partial_t^k u(\cdot, t^n)$ par $\partial_t^k u^n$.*

1.2.1.3 Stabilité

Le théorème suivant ainsi que son corollaire garantissent la stabilité du schéma que nous proposons sous une certaine condition CFL.

Théorème 1.2.8. *Le Δ^2-schéma est stable si les matrices $M - \frac{\Delta t^2}{4}K_1$ et $K_1 - \frac{\Delta t^2}{12}K_2$ sont des matrices positives.*

Démonstration. Comme dans la preuve du théorème 1.1.2, on a conservation de l'énergie discrète suivante

$$E^{n+\frac{1}{2}} = \left(\left(M - \frac{\Delta t^2}{4}K^*\right)\frac{U^{n+1} - U^n}{\Delta t}, \frac{U^{n+1} - U^n}{\Delta t}\right) + \left(K^*\frac{U^{n+1} + U^n}{2}, \frac{U^{n+1} + U^n}{2}\right)$$

si $M - \frac{\Delta t^2}{4}K^*$ et K^* sont deux matrices positives. Puisque $K^* = K_1 - \frac{\Delta t^2}{12}K_2$, nous avons à assurer la positivité des matrices $M - \frac{\Delta t^2}{4}K_1 + \frac{\Delta t^4}{48}K_2$ et K^*. De plus, comme K_2 est positive, la positivité de $M - \frac{\Delta t^2}{4}K_1$ implique la positivité de $M - \frac{\Delta t^2}{4}K^*$. □

Corollaire 1.2.9. *Le Δ^2-schéma est stable sous une condition CFL.*

Démonstration. Puisque M, K_1 et K_2 sont des matrices positives, il est clair qu'il existe $(C_1,\, C_2) \in \mathbb{R}^+ \times \mathbb{R}^+$ tel que $M - \frac{\Delta t^2}{4} K_1$ est positive $\forall \Delta t < C_1$ et K^* est positive $\forall \Delta t < C_2$. Par conséquent, le Δ^2-schéma est stable pour tout $\Delta t < \min(C_1, C_2)$. □

Remarque 1.2.10. *Le paramètre C_1 est aussi la condition CFL du schéma saute-moutons. Il est bien connu que C_1 est une fonction décroissante de α_1 (cf. par exemple [34]). Cependant, il n'y a pas d'expression analytique de ce paramètre et nous devons l'évaluer numériquement. Nous avons observé que le paramètre C_2 est une fonction décroissante de $\alpha_{2,1}$ et $\alpha_{2,2}$. Dans toutes les expériences numériques que nous avons effectuées, C_2 était plus grand que C_1 c'est-à-dire que la stabilité du schéma saute-moutons semble être une condition suffisante pour assurer la stabilité du Δ^2-schéma.*
Nous proposons une étude plus approfondie de ce point dans le chapitre 2 où nous explicitons la dépendance de C_1 par rapport à α dans des configurations simples, et dans le chapitre 3 où nous étudions le comportement de C_2.

1.2.1.4 Coûts Numériques du schéma

A présent, comparons le coût de ce schéma au coût des schémas saute-moutons et MES-4. Nous supposons ici que la matrice K dans (1.3) a été obtenue en utilisant une méthode IPDG d'ordre p de telle sorte que $(K)_{ij} = a_{1h}(\varphi_i, \varphi_j) = (K_1)_{ij}$. En pratique, nous calculons $K^* := K_1 - \frac{\Delta t^2}{12} K_2$ de telle sorte que nous n'avons à effectuer qu'une seule multiplication matricielle par $M^{-1}K^*$ à chaque itération. De plus, il est clair que $a_{1h}(\varphi_i, \varphi_j) = a_{2h}(\varphi_i, \varphi_j) = 0$ dès que les degrés de liberté i et j sont associés à deux éléments qui ne partagent pas une même arête. Cela signifie que $M^{-1}K_1$ et $M^{-1}K_2$ ont le même nombre d'éléments non-nuls et que le coût d'une multiplication par $M^{-1}K^*$ est le même que le coût d'une multiplication par $M^{-1}K = M^{-1}K_1$. Par conséquent, le coût d'une itération du Δ^2-schéma est le même que le coût d'une itération du schéma saute-moutons et la moitié du coût d'une itération du schéma MES-4.
Le coût global de ces schémas est le coût d'une itération multiplié par le nombre d'itérations qui est imposé par la condition CFL. Les simulations numériques que nous avons effectuées (cf. section 1.4) montrent que cette condition est un peu plus large que la condition CFL du schéma saute-moutons, ce qui implique que le coût global du Δ^2-schéma est équivalent à celui du schéma saute-moutons. De plus, puisque le coût du schéma MES-4 est 1.15 fois plus élevé que le coût du schéma saute-moutons, nous pouvons en déduire que le coût global du Δ^2-schéma est inférieur à celui du schéma MES-4.

1.2.2 Le Δ^3-schéma

Pour obtenir des schémas d'ordre six en temps, on est amené à considérer le Δ^3-schéma.

1.2.2.1 Expression du schéma

Nous ne détaillerons pas ici la construction du Δ^3-schéma qui est très similaire à celle du Δ^2-schéma, nous donnerons seulement son expression. Le problème à résoudre est le suivant :

$$\left\{\begin{array}{l} \text{Trouver } u_h^{n+1} \in V_h \text{ tel que,} \qquad \forall v \in V_h : \\ \displaystyle\sum_{K\in\mathcal{T}_h}\int_K \frac{1}{\mu}\frac{u_h^{n+1}-2u_h^n+u_h^{n-1}}{\Delta t^2}v = \quad -a_{1h}\left(u_h^n,v\right)+\frac{\Delta t^2}{12}a_{2h}\left(u_h^n,v\right) \\ \qquad\qquad\qquad\qquad\qquad\qquad -\dfrac{\Delta t^4}{360}a_{3h}\left(u_h^n,v\right)+\displaystyle\sum_{K\in\mathcal{T}_h}\int_K f_6(.,n\Delta t)v, \end{array}\right.$$

où

$$f_6 = f_4 + \frac{\Delta t^4}{360}\left(\frac{\partial^4 f}{\partial t^4} + \text{div}\left(\frac{1}{\rho}\nabla\left(\mu\frac{\partial^2 f}{\partial t^2}\right)\right) + \text{div}\left(\frac{1}{\rho}\nabla\left(\mu\text{div}\left(\frac{1}{\rho}\nabla\left(\mu f\right)\right)\right)\right)\right)$$

et

$$a_{3h}\left(u_h,v_h\right) = B_{\mathcal{T}_{h_3}}\left(u_h,v_h\right)+\mathcal{I}_3\left(u_h,v_h\right)+\mathcal{I}_3\left(v_h,u_h\right)+B_{S_{3,1}}\left(u_h,v_h\right)+B_{S_{3,2}}\left(u_h,v_h\right)+B_{S_{3,3}}\left(u_h,v_h\right),$$

avec

$$\left\{\begin{array}{ll} B_{\mathcal{T}_{h_3}}\left(u,v\right) = & \displaystyle\sum_{K\in\mathcal{T}_h}\int_K \frac{1}{\rho}\nabla\left(\mu\text{div}\left(\frac{1}{\rho}\nabla u\right)\right)\nabla\left(\mu\text{div}\left(\frac{1}{\rho}\nabla v\right)\right), \\ \mathcal{I}_3\left(u,v\right) = & \mathcal{I}_{3,1}\left(u,v\right)-\mathcal{I}_{3,2}\left(u,v\right)+\mathcal{I}_{3,3}\left(u,v\right), \\ \mathcal{I}_{3,1}\left(u,v\right) = & \displaystyle\sum_{F\in\mathcal{F}_h^i}\int_F\left\{\left\{\frac{1}{\rho}\nabla\left(\mu\text{div}\left(\frac{1}{\rho}\nabla\left(\mu\text{div}\left(\frac{1}{\rho}\nabla u\right)\right)\right)\right)\right\}\right\}[\![v]\!], \\ \mathcal{I}_{3,2}\left(u,v\right) = & \displaystyle\sum_{F\in\mathcal{F}_h}\int_F\left\{\left\{\mu\text{div}\left(\frac{1}{\rho}\nabla\left(\mu\text{div}\left(\frac{1}{\rho}\nabla u\right)\right)\right)\right\}\right\}\left[\!\left[\frac{1}{\rho}\nabla v\cdot\boldsymbol{\nu}\right]\!\right], \\ \mathcal{I}_{3,3}\left(u,v\right) = & \displaystyle\sum_{F\in\mathcal{F}_h^i}\int_F\left\{\left\{\frac{1}{\rho}\nabla\left(\mu\text{div}\left(\frac{1}{\rho}\nabla u\right)\right)\right\}\right\}\left[\!\left[\mu\text{div}\left(\frac{1}{\rho}\nabla v\right)\right]\!\right], \\ B_{S_{3,1}}\left(u,v\right) = & \displaystyle\sum_{F\in\mathcal{F}_h^i}\int_F\gamma_{3,1}\left[\!\left[\mu\text{div}\left(\frac{1}{\rho}\nabla u\right)\right]\!\right]\left[\!\left[\mu\text{div}\left(\frac{1}{\rho}\nabla v\right)\right]\!\right]. \\ B_{S_{3,2}}\left(u,v\right) = & \displaystyle\sum_{F\in\mathcal{F}_h}\int_F\gamma_{3,2}\left[\!\left[\frac{1}{\rho}\nabla u\cdot\boldsymbol{\nu}\right]\!\right]\left[\!\left[\frac{1}{\rho}\nabla v\cdot\boldsymbol{\nu}\right]\!\right], \\ B_{S_{3,3}}\left(u,v\right) = & \displaystyle\sum_{F\in\mathcal{F}_h^i}\int_F\gamma_{3,3}[\![u]\!][\![v]\!], \end{array}\right.$$

Les fonctions de pénalisation $\gamma_{3,1}$, $\gamma_{3,2}$ et $\gamma_{3,3}$ sont définies sur chaque face interne F par

$$\gamma_{3,1} = \frac{\alpha_{3,1}}{\min(\rho_{K^+},\rho_{K^-})\min(\xi_{F^+},\xi_{F^-})}, \quad \gamma_{3,2} = \frac{\alpha_{3,2}}{\min(\xi^3_{F^+},\xi^3_{F^-})}\max\left(\frac{\mu^2_{K^+}}{\rho_{K^+}},\frac{\mu^2_{K^-}}{\rho_{K^-}}\right),$$

$$\text{and } \gamma_{3,3} = \frac{\alpha_{3,3}}{\min(\xi^5_{F^+},\xi^5_{F^-})}\max\left(\frac{\mu^2_{K^+}}{\rho^3_{K^+}},\frac{\mu^2_{K^-}}{\rho^3_{K^-}}\right),$$

et $\gamma_{3,2}$ est respectivement définie sur une face extérieure F par

$$\gamma_{3,2} = \frac{\alpha_{3,2}}{\xi_F^3} \frac{\mu_K^2}{\rho_K},$$

où K est l'élément contenant F. On n'a pas besoin de définir $\gamma_{3,1}$ et $\gamma_{3,3}$ sur les faces extérieures car on considère le cas de conditions de bord de Dirichlet homogènes. Les paramètres $\alpha_{3,1}$, $\alpha_{3,2}$ et $\alpha_{3,3}$ sont positifs et dépendent uniquement du choix des fonctions de base de V_h. Il peut être prouvé, de la même manière que dans le cas du Δ^2-schéma en utilisant les inégalités de trace 1.31, que $\alpha_{3,1} \geq \alpha_{3,1}^0 \approx c_1 p^2$, $\alpha_{3,2} \geq \alpha_{3,2}^0 \approx c_2 p^6$ et $\alpha_{3,3} \geq \alpha_{3,3}^0 \approx c_3 p^{10}$, où p représente le degré des fonctions de base. En pratique, comme pour le Δ^2-schéma, les paramètres $\alpha_{3,2}$ et $\alpha_{3,3}$ peuvent être égaux à zéro. Effectivement, les formes $B_{S,3,3}$ et $B_{S,3,2}$ sont respectivement similaires à $B_{S,1}$ et $B_{S,2,2}$. $B_{S,1}$ et $B_{S,2,2}$ sont donc suffisantes pour assurer la stabilité de a_{1h}, a_{2h} et a_{3h}.

Comme dans le cas du Δ^2-schéma, nous considérons $\{\varphi_i\}_{i=1,\dots,m}$, les fonctions de base de Lagrange discontinues de degré p de V_h, où m désigne le nombre de degrés de liberté du problème, et nous obtenons le système linéaire

$$\frac{U^{n+1} - 2U^n + U^{n-1}}{\Delta t^2} + M^{-1}\left(K_1 - \frac{\Delta t^2}{12}K_2 + \frac{\Delta t^4}{360}K_3\right)U^n = M^{-1}F^n, \tag{1.21}$$

avec

$$(K_3)_{i,j} = a_{3h}(\varphi_i, \varphi_j), \ (F^n)_i = \sum_{K \in \mathcal{T}_h} \int_K f_6(\cdot, n\Delta t)\,\varphi_i.$$

Les matrices K_1 et K_2 sont identiques aux matrices introduites à la section 1.2.1.2 dédiées au Δ^2-schéma. En utilisant les mêmes arguments que pour le Δ^2-schéma, on peut montrer que le coût d'une itération du Δ^3-schéma est le même que celui du schéma saute-moutons et est trois fois plus petit que celui du schéma MES-6. Les expériences numériques que nous avons effectuées montrent que la condition CFL du Δ^3-schéma est légèrement supérieure à celle du schéma saute-moutons. Par conséquent, le coût global des deux schémas sont équivalents. De la même manière, on peut dire que le coût global du Δ^3-schéma est beaucoup plus faible que celui du schéma MES-6.

1.2.2.2 Stabilité

Nous avons un résultat analogue au théorème de stabilité 1.2.8 pour le cas triharmonique.

Théorème 1.2.11. *Le Δ^3-schéma est stable si les matrices $M - \frac{\Delta t^2}{4}K_1$, $K_1 - \frac{\Delta t^2}{12}K_2$ et $K_2 - \frac{\Delta t^2}{30}K_3$ sont positives.*

Démonstration. De la même manière que dans la preuve du théorème 1.2.8, on a conservation d'une énergie discrète

$$E^{n+\frac{1}{2}} = \left(\left(M - \frac{\Delta t^2}{4}K^*\right)\frac{U^{n+1} - U^n}{\Delta t}, \frac{U^{n+1} - U^n}{\Delta t}\right) + \left(K^*\frac{U^{n+1} + U^n}{2}, \frac{U^{n+1} + U^n}{2}\right)$$

si $M - \frac{\Delta t^2}{4}K^*$ et K^* sont deux matrices positives avec $K^* = K_1 - \frac{\Delta t^2}{12}K_2 + \frac{\Delta t^4}{360}K_3$. Une

condition suffisante de stabilité est donc

$$\begin{cases} M - \dfrac{\Delta t^2}{4} K_1 & \text{est positive,} \\ K_1 - \dfrac{\Delta t^2}{12} K_2 & \text{est positive,} \\ K_2 - \dfrac{\Delta t^2}{30} K_3 & \text{est positive.} \end{cases} \tag{1.22}$$

En effet, comme K_3 est une matrice positive, alors la positivité de $K_1 - \dfrac{\Delta t^2}{12} K_2$ implique la positivité de K^* ce qui justifie la troisième condition dans (1.22). De plus, une condition suffisante pour la positivité de $M - \frac{\Delta t^2}{4} K^* = M - \frac{\Delta t^2}{4} K_1 + \frac{\Delta t^4}{48} \left(K_2 - \frac{\Delta t^2}{30} K_3 \right)$ est évidemment $M - \frac{\Delta t^2}{4} K_1$ et $K_2 - \frac{\Delta t^2}{30} K_3$ positives.

□

Corollaire 1.2.12. *Le Δ^3-schéma est stable sous une condition CFL.*

Démonstration. De la même manière que pour le Δ^2-schéma, étant donné que M, K_1, K_2 et K_3 sont des matrices positives, il existe $(C_1,\, C_2,\, C_3) \in \left(\mathbb{R}^+\right)^3$ tel que $M - \frac{\Delta t^2}{4} K_1$ (resp. $K_1 - \frac{\Delta t^2}{12} K_2$ et $K_2 - \frac{\Delta t^2}{30} K_3$) soit positive $\forall \Delta t < C_1$ (resp. C_2 et C_3). Par conséquent, le Δ^3-schéma est stable pour tout $\Delta t < \min\left(C_1, C_2, C_3\right)$. □

Remarque 1.2.13. *Ici aussi le paramètre Δt_1 est la condition CFL du schéma saute-moutons et il semblerait numériquement que la stabilité du schéma saute-moutons soit une condition suffisante pour assurer la stabilité du Δ^3-schéma.*

1.3 Un résultat de convergence pour le Δ^2-schéma

Dans cette partie, nous proposons une preuve de la convergence de la solution du système totalement discrétisé vers la solution continue à l'instant t^n. On trouve dans la littérature beaucoup de travaux sur des estimations d'erreur a priori et a posteriori pour les méthodes de Galerkine Discontinues (cf. par exemple [45, 4] pour les problèmes elliptiques et [29, 31] pour le problème biharmonique) mais seul le travail de Grote et al. dans [34] traite du cas de l'approximation du second ordre en temps pour l'équation des ondes. Nous nous inspirons ici de ce travail pour étudier la convergence du Δ^2-schéma. Nous supposons que ρ and μ sont constants pour simplifier la démonstration et pour fixer les idées nous nous restreignons aux cas de conditions de bord de Dirichlet homogènes.

Théorème 1.3.1. *Soit u la solution de l'équation des ondes (1.1) satisfaisant les hypothèses de régularité suivantes*

$$u \in C^2\left(\overline{J}; H^{p+1}(\Omega)\right), \quad \partial_t^5 u \in C\left(\overline{J}; L^2(\Omega)\right), \quad \partial_t^6 u \in L^1\left(J; L^2(\Omega)\right) \tag{1.23}$$

avec $J =]0, T[$. Soit $(U^n)_{n=0}^N$ la solution discrète définie par (1.18)-(1.19)-(1.20). Si Δt satisfait

$$\Delta t \leq \beta h \tag{1.24}$$

avec $\beta \in \mathbb{R}^+$ assez petit, et si $12/\Delta t^2$ n'est pas une valeur propre de l'opérateur $-\Delta$ sur Ω alors il existe une constante $C > 0$ indépendante de h et Δt telle que :

$$\max_{n=0,\dots,N} \|u^n - U^n\|_0 \leq C\left(h^{p+1} + \Delta t^4\right).$$

Les sous-sections suivantes présentent les principales étapes de la preuve du théorème précédent, et nous renvoyons à l'annexe 1.A pour les détails techniques. La première sous-section décrit les propriétés de la forme bilinéaire a_h et la deuxième concerne l'estimation d'erreur proprement dite.

1.3.1 Propriétés de la forme bilinéaire a_h

Tout d'abord, nous présentons les trois normes définies sur l'espace $V_h + H^4(\Omega)$ qui nous serviront pour établir les estimations d'erreur. De manière classique, les normes DG sont définies sur l'espace $V_h + H^2(\Omega)$ (cf. [5]) et si l'on souhaite monter en ordre (cf. [43] on est amené à considérer l'espace $V_h + H^4(\Omega)$ de façon à décomposer la solution en une partie régulière et en une partie discontinue.

$$\|u\|^2_{\mathrm{DG}_2} = |u|^2_{1,h} + \sum_K h_K^2 |u|^2_{2,K} + |\alpha_1^{1/2} u|^2_*$$
$$\|u\|^2_{\mathrm{DG}_4} = |u|^2_{2,h} + |\alpha_{2,2}^{1/2} u|^2_* + |\alpha_{2,1}^{1/2} \nabla u|^2_* + \|\alpha_{2,2}^{-1/2} \{\!\{\nabla(\Delta u)\cdot \boldsymbol{\nu}\}\!\}\|^2_{L^2(\Gamma)} + \|\alpha_{2,1}^{-1/2} \{\!\{\Delta u\}\!\}\|^2_{L^2(\Gamma)}$$
$$|\!|\!| u |\!|\!|^2 = \|u\|^2_{\mathrm{DG}_2} + \tfrac{\Delta t^2}{12} \|u\|^2_{\mathrm{DG}_4}$$

où $|u|^2_{1,h} = \sum_{K\in\mathcal{T}_h} (\nabla u, \nabla u)_{L^2(K)}$, $|u|^2_{2,h} = \sum_{K\in\mathcal{T}_h} (\Delta u, \Delta u)_{L^2(K)}$ et $|u|^2_* = \int_\Gamma [\![u]\!]^2 \, ds$.

Lemme 1.3.2. *Si $\gamma_1 > 0$, $\gamma_{2,1} > 0$ et $\gamma_{2,2} > 0$ alors $|\!|\!|\cdot|\!|\!|_{DG_2}$, $|\!|\!|\cdot|\!|\!|_{DG_4}$ et $|\!|\!|\cdot|\!|\!|$ sont des normes sur $V_h + H^4(\Omega)$.*

Remarque 1.3.3. *La preuve de ce lemme est une extension triviale des preuves données dans [5] et [43].*

Maintenant, regardons certaines propriétés de a_h dans les espaces V_h et $V_h + H^4(\Omega)$. Rappelons tout d'abord quelques résultats issus de [33] :

Lemme 1.3.4. *Il existe deux constantes $C_{coer1}, C_{cont1} > 0$ indépendantes de la taille du maillage telles que :*

$$a_{1h}(u,u) \geq C_{coer1} \|u\|^2_{DG_2}, \quad \forall u \in V_h$$

et

$$|a_{1h}(u,v)| \leq C_{cont1} \|u\|_{DG_2} \|v\|_{DG_2}, \quad \forall u, v \in V_h + H^4(\Omega)$$

De plus, pour des maillages quasi-uniformes :

$$a_{1h}(u,u) \leq C_S c^2_{max} h^{-2} \|u\|^2_0, \quad u \in V_h + H^4(\Omega).$$

où C_S est indépendante du maillage, de c^2 et de α_1.

Grâce à [43], nous avons un résultat de continuité pour la forme bilinéaire a_{2h} :

Lemme 1.3.5. *Il existe une constante $C_{cont2} > 0$ indépendante de h telle que :*

$$a_{2h}(u,v) \leq C_{cont2} \|u\|_{DG_4} \|v\|_{DG_4}, \quad \forall u, v \in V_h + H^4(\Omega).$$

Enfin, nous avons besoin du résultat suivant pour établir le théorème de convergence.

Théorème 1.3.6. *Sous la condition CFL (1.24), la forme bilinéaire a_h satisfait une condition de stabilité sur V_h, i.e. $\exists C > 0$ tel que*

$$a_h(u,u) \geq C\|u\|_{DG_2}^2, \quad \forall u \in V_h$$

et a_h est continue sur V_h, i.e. $\exists C > 0$ tel que

$$|a_h(u,v)| \leq C\|u\|_{DG_2}\|v\|_{DG_2}, \quad \forall u, v \in V_h.$$

La preuve de ce théorème est présentée dans la sous-section 1.A.1.

1.3.2 Majoration d'erreur

Tout d'abord, nous devons introduire la projection de Galerkine qui va nous permettre de projeter u sur V_h.

Définition 1.3.7. *Soit $u \in H^4(\Omega)$. Nous considérons la projection de Galerkine $p_h : u \to p_h(u) = u_h$, la projection de u sur V_h telle que :*

$$a_h(p_h(u), v) = a_h(u,v), \quad \forall v \in V_h.$$

Remarque 1.3.8. *L'existence et l'unicité de $p_h(u)$ sont garanties par la stabilité et la continuité de a_h dans V_h.*

La projection de Galerkine satisfait les estimations suivantes

Lemme 1.3.9. *Si $u \in H^{p+1}(\Omega)$ avec $p \geq 1$, alors il existe $C > 0$ indépendant de h tel que*

$$\begin{aligned}
\|u - p_h(u)\|_{DG_2} &\leq Ch^p\|u\|_{p+1}, \\
\|u - p_h(u)\|_{DG_4} &\leq Ch^{p-1}\|u\|_{p+1}, \\
\|u - p_h(u)\|_0 &\leq Ch^{p+1}\|u\|_{p+1}.
\end{aligned}$$

Nous renvoyons à la section 1.A.3 pour la preuve de ce lemme.

A présent, nous voulons obtenir une majoration de l'erreur entre la solution continue du problème à l'instant t^n et la solution du système totalement discrétisé au même instant. On note cette erreur $e^n = u^n - U^n$ qui peut être réécrite comme :

$$e^n = \eta^n + \phi^n, \quad n = 0, \ldots, N$$

où

$$\begin{cases}
\eta^n = u^n - w^n \\
\phi^n = w^n - U^n \\
w^n = p_h(u^n)
\end{cases}$$

Grâce à l'inégalité triangulaire, on a :

$$\max_{n=0,\ldots,N}\|e^n\|_0 \leq \max_{n=0,\ldots,N}\|\phi^n\|_0 + \max_{n=0,\ldots,N}\|\eta^n\|_0. \tag{1.25}$$

Intéressons-nous tout d'abord au terme $\max_{n=0,\dots,N}\|\phi^n\|_0$.
Nous introduisons :

$$r^n = \begin{cases} \delta^2 w^n - \partial_t^2 u^n - \frac{\Delta t^4}{12}\partial_t^4 u^n, & n = 1,\dots,N-1, \\ \dfrac{\phi^1 - \phi^0}{\Delta t^2}, & n = 0 \end{cases}$$

avec $\delta^2 w^n = \frac{w^{n+1}-2w^n+w^{n-1}}{\Delta t^2}$ et nous notons $R^n = \Delta t \sum_{m=0}^{n} r^m$. On a alors

Proposition 1.3.10. *Sous la condition CFL (1.24), il existe une constante $C > 0$ telle que*

$$\max_{n=0,\dots,N}\|\phi^n\|_0 \leq C\left(\|e^0\|_0 + \|\eta^0\|_0\right) + C^2\Delta t\sum_{n=0}^{N-1}\|R^n\|_0.$$

où $N = \frac{T}{\Delta t}$ est le nombre d'itérations et $C > 0$ est une constante indépendante de h, Δt et T.

On renvoie à la sous-section 1.A.4 pour la preuve de cette proposition.
En combinant la proposition 1.3.10 avec (1.25), nous obtenons

$$\max_{n=0,\dots,N}\|e^n\|_0 \leq C\left(\|e^0\|_0 + \max_{n=0,\dots,N}\|\eta^n\|_0 + \Delta t\sum_{n=0}^{N-1}\|R^n\|_0\right) \tag{1.26}$$

De plus, il est évident que

$$\Delta t\sum_{n=1}^{N-1}\|R^n\|_0 \leq T\max_{n=1,\dots,N-1}\|R^n\|_0,$$

de telle sorte que (1.26) peut être reformulé en

$$\max_{n=0,\dots,N}\|e^n\|_0 \leq C\left(\|e^0\|_0 + \max_{n=0,\dots,N}\|\eta^n\|_0 + T\max_{n=1,\dots,N-1}\|R^n\|_0\right). \tag{1.27}$$

Comme $\max_{n=0,\dots,N}\|\eta^n\|_0 = \max_{n=0,\dots,N}\|u^n - p_h\left(u^n\right)\|_0$, en utilisant le lemme 1.3.9, on obtient

$$\max_{n=0,\dots,N}\|\eta^n\|_0 \leq Ch^{p+1}\|u\|_{C\left(\overline{J};H^{p+1}(\Omega)\right)}. \tag{1.28}$$

De plus, grâce aux propriétés de la projection L^2 et au lemme 1.3.9, il vient :

$$\|e^0\|_0 = \|u^0 - P_h\left(u^0\right)\|_0 \leq Ch^{p+1}\|u_0\|_{p+1} \leq Ch^{p+1}\|u\|_{C\left(\overline{J};H^{p+1}(\Omega)\right)}. \tag{1.29}$$

Maintenant que nous disposons de ce résultat, nous devons borner $\|R^n\|_0$. On se doit de distinguer deux cas : le cas $n = 0$ et le cas $n \geq 1$. Dans le premier cas, nous avons le lemme suivant :

Lemme 1.3.11. *Il existe une constante $C > 0$ indépendante de h, Δt et T telle que*

$$\|r^0\|_0 \leq C\left(\Delta t^{-1}h^{p+1}\|\partial_t u\|_{C\left(\overline{J},H^{p+1}(\Omega)\right)} + \Delta t^3\|\partial_t^5 u\|_{C\left(\overline{J},L^2(\Omega)\right)}\right)$$

Dans le deuxième cas, nous avons

Lemme 1.3.12. *Pour $1 \leq n \leq N-1$, il existe $C > 0$ indépendante de h, Δt et T telle que*

$$\|r^n\|_0 \leq C \left(\frac{h^{p+1}}{\Delta t} \int_{t_{n-1}}^{t_{n+1}} \|\partial_t^2 u(\cdot, s)\|_{p+1} ds + \Delta t^3 \int_{t_{n-1}}^{t_{n+1}} \|\partial_t^6 u(\cdot, s)\|_0 ds \right)$$

Ces deux lemmes sont prouvés dans les sous-sections 1.A.5 et 1.A.6.
Par définition de R^n, et en appliquant l'inégalité triangulaire, on obtient l'inégalité

$$\|R^n\|_0 \leq \Delta t \|r^0\|_0 + \Delta t \sum_{m=1}^{N-1} \|r^m\|_0.$$

Grâce aux lemmes 1.3.11 et 1.3.12, $\forall n \in \{1 \dots N-1\}$, on sait qu'il existe une constante $C > 0$ indépendante de h, Δt et T telle que

$$\begin{aligned} \|R^n\|_0 \leq \quad & C\Delta t^4 \left(\|\partial_t^5 u\|_{C\left(\overline{J}, L^2(\Omega)\right)} + \|\partial_t^6 u\|_{L^1(J, L^2(\Omega))} \right) \\ & + Ch^{p+1} \left(\|\partial_t u\|_{C\left(\overline{J}, H^{p+1}(\Omega)\right)} + \|\partial_t^2 u\|_{C\left(\overline{J}, H^{p+1}(\Omega)\right)} \right) \end{aligned} \tag{1.30}$$

En combinant (1.27), (1.28), (1.29) et (1.30), on obtient donc

$$\max_{n=0,\dots,N} \|e^n\|_0 \leq Ch^{p+1} \|u\|_{C^2\left(\overline{J}; H^{p+1}(\Omega)\right)} + C\Delta t^4 \left(\|\partial_t^5 u\|_{C\left(\overline{J}; L^2(\Omega)\right)} + \|\partial_t^6 u\|_{L^1(J; L^2(\Omega))} \right),$$

ce qui prouve que

$$\max_{n=0,\dots,N} \|u^n - U^n\| \leq C \left(h^{p+1} + \Delta t^4 \right).$$

□

1.4 Résultats numériques

Dans cette section, nous présentons des résultats numériques en dimension un et deux pour comparer les performances des Δ^2- et Δ^3-schémas à celles des schémas MES-4 et MES-6. Nous comparons la précision et les coûts de calculs des deux techniques. Nous avons également mis en oeuvre la méthode des éléments finis de Hermite en 1D dans des milieux homogènes. Les détails sur cette méthode sont donnés en annexe 1.B.

1.4.1 Résultats 1D

Dans toute cette partie, nous considérons la simulation de la propagation d'une onde dans un domaine 1D homogène $\Omega = [0, 10]$ avec une vitesse $c = (\mu/\rho)^{1/2} = 1\text{ms}^{-1}$. Pour calculer de manière plus aisée la solution exacte, nous imposons des conditions de bord périodiques à chaque extrémité du domaine. Néanmoins des conditions de bords de Neumann et de Dirichlet seront considérées en dimensions d'espace supérieures. Le terme source est supposé nul et les données initiales sont

$$u_0(x) = \begin{cases} (x - x_0)\, e^{-\left(\frac{2\pi(x-x_0)}{r_0}\right)^2} & \text{si } |x - x_0| \leq r_0 \\ 0 & \text{sinon,} \end{cases}$$

et

$$u_1(x) = \begin{cases} \left(8\left(\frac{(x-x_0)\pi}{r_0}\right)^2 - 1\right) e^{-\left(\frac{2\pi(x-x_0)}{r_0}\right)^2} & \text{si } |x-x_0| \leq r_0 \\ 0 & \text{sinon} \end{cases}$$

de telle sorte que le signal soit à support strictement inclus dans le domaine de calcul.
Ainsi, la solution exacte est donnée par :

$$u^{\text{ex}}(x,t) = \sum_{i=0}^{+\infty} u_0(x + 10i - t).$$

Dans la suite, nous fixons $x_0 = 3$ et $r_0 = 4$.
Afin de discrétiser l'équation des ondes (1.1), nous avons considéré

1. le schéma MES-4, basé sur une discrétisation en espace utilisant des polynômes de Lagrange de degré 3 et un paramètre de pénalisation $\gamma_1 = 8$. Avec ces fonctions de base et ce paramètre, la condition CFL du schéma saute-moutons est (expérimentalement) $\Delta t_{LF_4} = 0.1533h$. Ainsi la condition CFL du schéma MES-4 est $\Delta t_{MES-4} = 0.1533\sqrt{3}h = 0.2655h$.
2. le schéma MES-6, basé sur une discrétisation en espace utilisant des polynômes de Lagrange de degré 5 et un paramètre de pénalisation $\gamma_1 = 20$. Avec ces fonctions de base et ce paramètre, la condition CFL du schéma saute-moutons est (expérimentalement) $\Delta t_{LF_6} = 0.073h$. Ainsi la condition CFL du schéma MES-6 est $\Delta t_{MES-6} = 1.38 \times 0.073h = 0.101h$.
3. Le Δ^2-schéma, avec des fonctions de base de Lagrange P^3 et des paramètres de pénalisation $\gamma_1 = 8$, $\gamma_{2,1} = 10$ et $\gamma_{2,2} = 0$. La condition CFL de ce schéma est (expérimentalement) $\Delta t_{\Delta^2} = 0.1821h$.
4. Le Δ^3-schéma, avec des fonctions de base de Lagrange P^5 et des paramètres de pénalisation $\gamma_1 = 20$, $\gamma_{2,1} = 20$, $\gamma_{2,2} = 0$, $\gamma_{3,1} = 20$, $\gamma_{3,2} = 0$ et $\gamma_{3,3} = 0$. Avec ces paramètres, la condition CFL est (expérimentalement) $\Delta t_{\Delta^3} = 0.077h$.
5. Le Δ^2-schéma avec des éléments finis de Hermite. La condition CFL est (expérimentalement) $\Delta t_{\text{Herm}} = 0.4471h$.

Remarque 1.4.1. *En suivant ce qui a été fait dans [2], nous choisissons $\alpha_1 > \alpha_1^0 = p(p+1)/2$. Puisque nous n'avons pas une expression explicite des autres coefficients de pénalisation, nous les évaluons numériquement dans le but d'obtenir une solution stable. Néanmoins, comme nous l'avons déjà dit, nous étudierons ces coefficients ainsi que les conditions CFL de ces schémas de manière plus précise dans les chapitres 2 et 3.*

Remarquons que les conditions CFL du Δ^2-schéma et du Δ^3-schéma sont légèrement plus grandes que la condition CFL des schémas saute-moutons Δt_{LF_4} et Δt_{LF_6}. Puisque les Δ^p-schémas nécessitent seulement une multiplication matricielle par itération, cela signifie que les coûts de calcul associés à ces schémas sont moins importants que ceux du schéma saute-moutons (tout du moins pour $p = 2$ et 3).

1.4.1.1 Evolution de l'erreur en fonction du pas d'espace

Dans cette partie, nous calculons l'erreur relative $L^2([0,T],\Omega)$, donnée par $\left(\int_0^T \left(\int_\Omega (u^{\text{ex}} - u_h)^2\, dx\right) dt\right)^{1/2}$ où u^{ex} et u_h représentent respectivement la solution exacte et l'approximation, pour $T = 100$ et

pour différents pas d'espace : $h = 0.25, 0.125, 0.0625, 0.03125$ pour les schémas d'ordre quatre et $h = 1, 0.5, 0.25, 0.125$ pour les schémas d'ordre six. Dans la Table 1.1 (resp. Table 1.2), nous présentons l'erreur $L^2([0,T],\Omega)$ de chaque schéma et dans la Fig. 1.1 (resp. Fig. 1.2) nous représentons l'erreur relative L^2 comme une fonction de la taille du maillage pour le schéma MES-4 (resp. MES-6) (courbe cyan avec losange) et le Δ^2-schéma (resp. Δ^3-schéma) (courbe verte avec carrés) en échelle logarithmique. On constate que tous les schémas convergent bien à l'ordre voulu et que les Δ^p-schémas donnent d'aussi bons résultats que le schéma d'ordre correspondant MES-p. Puisque la condition CFL des Δ^p-schémas est légèrement supérieure à la condition CFL des schémas saute-moutons et que cette méthode ne requiert qu'une multiplication matricielle à chaque itération, cela signifie qu'elle nous permet d'avoir une précision d'ordre élevé avec un coût plus petit que le schéma saute-moutons. A titre de comparaison, nous rappelons que les coûts numériques du schéma MES-4 et du schéma MES-6 sont respectivement 1.15 et 2.17 fois plus grands que le coût des schémas saute-moutons (cf. section 1.1.2).
A titre indicatif, nous donnons les résultats obtenus en considérant les éléments finis d'Hermite (HFEM) dans la Table 1.1. La convergence du schéma avec de tels éléments finis est bien d'ordre 4 mais on peut également constater que les résultats obtenus sont bien moins bons qu'avec les deux autres méthodes. Cela nous conforte donc dans le choix d'une discrétisation en espace de type Galerkine Discontinue.

h	MES-4	Δ^2-schéma	HFEM
0.25	$1.1\ 10^{-3}$	$2.0\ 10^{-3}$	$4.2\ 10^{-2}$
0.125	$7.4\ 10^{-5}$	$4.5\ 10^{-5}$	$2.1\ 10^{-3}$
0.0625	$5.6\ 10^{-6}$	$2.1\ 10^{-6}$	$1.2\ 10^{-4}$
0.03125	$3.7\ 10^{-7}$	$1.2\ 10^{-7}$	$7.9\ 10^{-6}$

TABLE 1.1 – Erreur relative $L^2([0,T],\Omega)$ au temps $T = 100$s pour les schémas d'ordre quatre.

h	MES-6	Δ^3-schéma
1	$2.4\ 10^{-1}$	$2.8\ 10^{-1}$
0.5	$2.3\ 10^{-4}$	$4.1\ 10^{-4}$
0.25	$2.1\ 10^{-6}$	$2.5\ 10^{-6}$
0.125	$4.6\ 10^{-8}$	$4.2\ 10^{-8}$

TABLE 1.2 – Erreur relative $L^2([0,T],\Omega)$ au temps $T = 100$s pour les schémas d'ordre six.

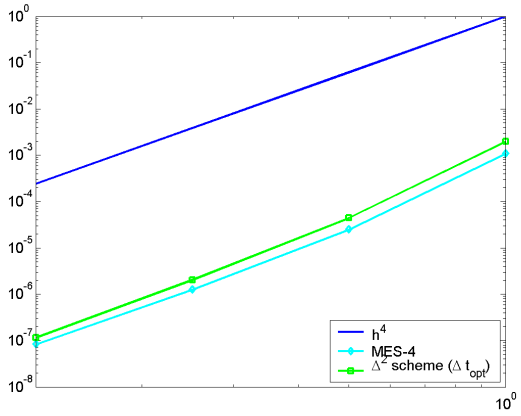

FIGURE 1.1 – Courbes de convergence pour les schémas d'ordre 4 en 1D.

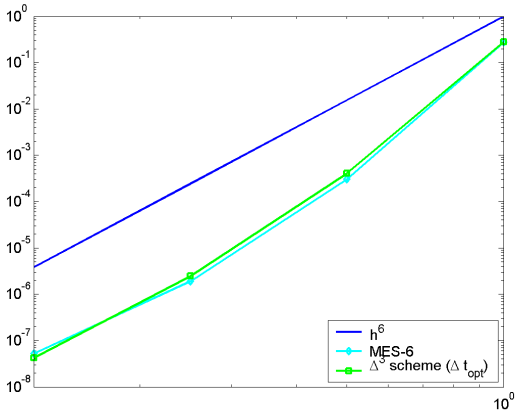

FIGURE 1.2 – Courbes de convergence pour les schémas d'ordre 6 en 1D.

1.4.1.2 Evolution de l'erreur en fonction du pas de temps

Dans cette partie, on va comparer les schémas MES aux Δ^p-schémas en fonction du pas de temps. Nous calculons l'erreur relative $L^2\left([0,T],\Omega\right)$ pour un pas d'espace fixé et pour différents pas de temps égaux à $a\Delta t$ où a est une constante et Δt le pas de temps utilisé dans la sous-section précédente. Les autres paramètres étant identiques à ceux de la partie précédente.
Dans la Table 1.3 (resp. Table 1.4 et Table 1.5), nous présentons l'erreur $L^2\left([0,T],\Omega\right)$ obtenue pour les schémas d'ordre quatre et un pas d'espace $h = 0.5$ (resp. $h = 0.25$ et $h = 0.125$).
On constate que le MES-4 et le Δ^2-schéma donnent des résultats similaires même si le MES-4 semblent donner de meilleurs résultats lorsque le pas d'espace est suffisamment raffiné. Lorsque l'on diminue le pas de temps les erreurs obtenues diminuent jusqu'à atteindre une certaine constante à partir de laqulle celles-ci augmentent très légérement, cela étant probablement dû au phénomène de dispersion numérique.

a	MES-4	Δ^2-schéma	a	MES-4	Δ^2-schéma
1	$5.27\ 10^{-2}$	$1.27\ 10^{-1}$	0.5	$3.60\ 10^{-2}$	$5.50\ 10^{-2}$
0.9	$4.59\ 10^{-2}$	$8.94\ 10^{-2}$	0.4	$3.57\ 10^{-2}$	$4.81\ 10^{-2}$
0.8	$4.16\ 10^{-2}$	$8.00\ 10^{-2}$	0.3	$3.54\ 10^{-2}$	$4.25\ 10^{-2}$
0.7	$3.89\ 10^{-2}$	$7.12\ 10^{-2}$	0.2	$3.52\ 10^{-2}$	$3.86\ 10^{-2}$
0.6	$3.73\ 10^{-2}$	$6.29\ 10^{-2}$	0.1	$3.53\ 10^{-2}$	$3.61\ 10^{-2}$

TABLE 1.3 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour les schémas d'ordre quatre et un pas d'espace $h = 0.5$.

a	MES-4	Δ^2-schéma	a	MES-4	Δ^2-schéma
1	$1.06\ 10^{-3}$	$1.97\ 10^{-3}$	0.5	$5.61\ 10^{-4}$	$4.42\ 10^{-4}$
0.9	$5.89\ 10^{-4}$	$1.36\ 10^{-3}$	0.4	$6.04\ 10^{-4}$	$4.81\ 10^{-4}$
0.8	$3.74\ 10^{-4}$	$9.30\ 10^{-4}$	0.3	$6.24\ 10^{-4}$	$5.40\ 10^{-4}$
0.7	$3.97\ 10^{-4}$	$6.38\ 10^{-4}$	0.2	$6.31\ 10^{-4}$	$5.90\ 10^{-4}$
0.6	$4.88\ 10^{-4}$	$4.78\ 10^{-4}$	0.1	$6.33\ 10^{-4}$	$6.22\ 10^{-4}$

TABLE 1.4 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour les schémas d'ordre quatre et un pas d'espace $h = 0.25$.

A présent nous présentons l'erreur $L^2\left([0,T],\Omega\right)$ obtenue pour les schémas d'ordre six et un pas d'espace $h = 0.5$ (resp. $h = 0.25$ et $h = 0.125$) dans la Table 1.6 (resp. Table 1.7 et Table 1.8). Les conclusions sont similaires à celles effectuées pour l'ordre quatre. En effet, les deux méthodes présentent un comportement similaire lorsque le pas de temps diminue.

1.4.2 Résultats 2D

Dans cette section, nous considérons la simulation de la propagation d'ondes dans un milieu bicouche 2D $\Omega = [-1,1]^2 = \Omega_t \cap \Omega_b$ où $\Omega_t = [-1,1] \times [0,1]$ et $\Omega_b = [-1,1] \times [-1,0]$ sont deux couches homogènes respectivement caractérisées par $\mu = 2$, $\rho = 2$ et $\mu = 8$, $\rho = 4$. Nous

a	MES-4	Δ^2-schéma	a	MES-4	Δ^2-schéma
1	$7.44\ 10^{-5}$	$4.48\ 10^{-5}$	0.5	$2.62\ 10^{-5}$	$2.62\ 10^{-5}$
0.9	$4.39\ 10^{-5}$	$2.94\ 10^{-5}$	0.4	$2.86\ 10^{-5}$	$2.78\ 10^{-5}$
0.8	$2.54\ 10^{-5}$	$2.34\ 10^{-5}$	0.3	$2.98\ 10^{-5}$	$2.90\ 10^{-5}$
0.7	$1.99\ 10^{-5}$	$2.28\ 10^{-5}$	0.2	$3.01\ 10^{-5}$	$2.97\ 10^{-5}$
0.6	$2.26\ 10^{-5}$	$2.43\ 10^{-5}$	0.1	$3.03\ 10^{-5}$	$3.01\ 10^{-5}$

TABLE 1.5 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour les schémas d'ordre quatre et un pas d'espace $h = 0.125$.

a	MES-6	Δ^3-schéma	a	MES-6	Δ^3-schéma
1	$2.29\ 10^{-4}$	$4.10\ 10^{-4}$	0.5	$6.61\ 10^{-5}$	$9.44\ 10^{-5}$
0.9	$1.11\ 10^{-4}$	$2.60\ 10^{-4}$	0.4	$6.35\ 10^{-5}$	$7.82\ 10^{-5}$
0.8	$8.73\ 10^{-5}$	$1.97\ 10^{-4}$	0.3	$6.18\ 10^{-5}$	$6.83\ 10^{-5}$
0.7	$7.63\ 10^{-5}$	$1.51\ 10^{-4}$	0.2	$6.08\ 10^{-5}$	$6.30\ 10^{-5}$
0.6	$7.00\ 10^{-5}$	$1.18\ 10^{-4}$	0.1	$6.02\ 10^{-5}$	$6.06\ 10^{-5}$

TABLE 1.6 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour les schémas d'ordre six et un pas d'espace $h = 0.5$.

a	MES-6	Δ^3-schéma	a	MES-6	Δ^3-schéma
1	$2.13\ 10^{-6}$	$2.46\ 10^{-6}$	0.5	$1.05\ 10^{-6}$	$1.09\ 10^{-6}$
0.9	$1.36\ 10^{-6}$	$1.68\ 10^{-6}$	0.4	$1.03\ 10^{-6}$	$1.05\ 10^{-6}$
0.8	$1.19\ 10^{-6}$	$1.40\ 10^{-6}$	0.3	$1.02\ 10^{-6}$	$1.03\ 10^{-6}$
0.7	$1.12\ 10^{-6}$	$1.24\ 10^{-6}$	0.2	$1.02\ 10^{-6}$	$1.02\ 10^{-6}$
0.6	$1.07\ 10^{-6}$	$1.14\ 10^{-6}$	0.1	$1.01\ 10^{-6}$	$1.01\ 10^{-6}$

TABLE 1.7 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour les schémas d'ordre six et un pas d'espace $h = 0.25$.

a	MES-6	Δ^3-schéma	a	MES-6	Δ^3-schéma
1	$4.65\ 10^{-8}$	$4.18\ 10^{-8}$	0.5	$4.54\ 10^{-8}$	$4.09\ 10^{-8}$
0.9	$4.49\ 10^{-8}$	$4.16\ 10^{-8}$	0.4	$4.57\ 10^{-8}$	$4.09\ 10^{-8}$
0.8	$4.48\ 10^{-8}$	$4.14\ 10^{-8}$	0.3	$4.60\ 10^{-8}$	$4.10\ 10^{-8}$
0.7	$4.49\ 10^{-8}$	$4.12\ 10^{-8}$	0.2	$4.63\ 10^{-8}$	$4.13\ 10^{-8}$
0.6	$4.51\ 10^{-8}$	$4.10\ 10^{-8}$	0.1	$4.66\ 10^{-8}$	$4.20\ 10^{-8}$

TABLE 1.8 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour les schémas d'ordre six et un pas d'espace $h = 0.125$.

considérons des données initiales nulles et une source qui est une dérivée seconde de Gaussienne en temps et un point source en espace :

$$f = \delta_{x_0} 2\lambda \left(\lambda \left(t - t_0\right)^2 - 1\right) e^{-\lambda (t-t_0)^2},$$

avec $x_0 = (0, 0.5)$, $\lambda = \pi^2 f_0^2$, $f_0 = 5$ et $t_0 = 1/f_0$.
Pour discrétiser l'équation des ondes (1.1), nous avons utilisé les deux méthodes suivantes :

1. le schéma MES-4, basé sur une discrétisation en espace utilisant des polynômes de Lagrange de degré 3 et un paramètre de pénalisation $\gamma_1 = 10$. Avec ces fonctions de base et ce paramètre, la condition CFL du schéma saute-moutons est (expérimentalement) $\Delta t_{LF_4} = 0.058h$. Ainsi la condition CFL du schéma MES-4 est $\Delta t_{MES-4} = 0.058\sqrt{3}h = 0.100h$.
2. le Δ^2-schéma, avec des fonctions de base de Lagrange P^3 et avec les paramètres de pénalisation $\gamma_1 = 10$, $\gamma_{2,1} = 10$ et $\gamma_{2,2} = 0$. La condition CFL de ce schéma est (expérimentalement) $\Delta t_{\Delta^2} = 0.061h$.

Comme pour la 1D, la condition CFL du Δ^2-schéma est légérement supérieure à la condition CFL du schéma saute-moutons Δt_{LF_4}.
Pour d'illustrer une telle simulation, on représente sur les figures 1.3 et 1.4 deux instantanés de propagation correspondant à des temps d'expérience $T \simeq 9.40\ 10^{-3}$s et $T \simeq 1.27\ 10^{-2}$s. Pour

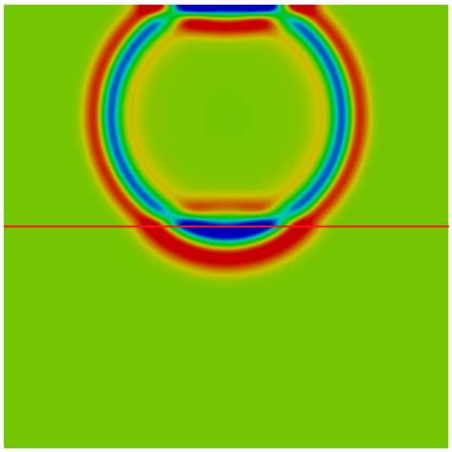

FIGURE 1.3 – Instantané de propagation à $T \simeq 9.40\ 10^{-3}$s.

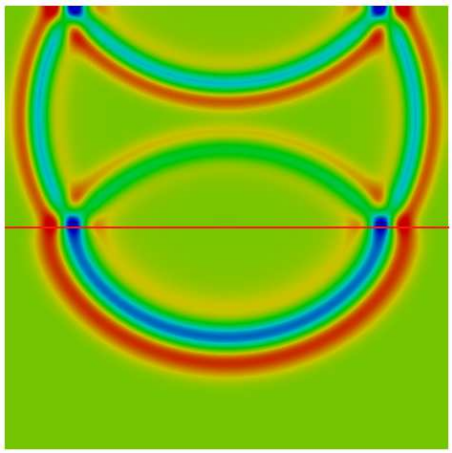

FIGURE 1.4 – Instantané de propagation à $T \simeq 1.27\ 10^{-2}$s.

comparer les performances des différentes méthodes, nous calculons la solution exacte en un récepteur situé au point $x_1 = (0.25, 0.25)$ et nous calculons l'erreur relative $L^2\left([0,T], x_1\right)$ pour différents pas d'espace moyen $h = 3\ 10^{-3}, 1.5\ 10^{-3}, 7.5\ 10^{-4}$ et un temps final approximativement égal à 1.33s. La solution analytique est calculée grâce à la méthode de Cagniard-De Hoop [14, 24] qui donne la solution du problème en un point du domaine au cours du temps. Les résultats sont présentés dans la Table 1.9 et les courbes de convergence en échelle logarithmique sont données à la figure 1.5.
Comme pour le cas 1D, on constate que les deux méthodes sont bien des approximations d'ordre quatre et qu'elles donnent des résultats similaires. Une fois de plus, nous pouvons donc conclure que le coût du Δ^2-schéma est plus petit que celui du schéma MES-4 pour une précision équivalente.

h	MES-4	Δ^2-schéma
$3.0\ 10^{-3}$	$2.5\ 10^{-2}$	$2.3\ 10^{-2}$
$1.5\ 10^{-3}$	$1.2\ 10^{-3}$	$1.1\ 10^{-3}$
$7.5\ 10^{-4}$	$6.6\ 10^{-5}$	$6.5\ 10^{-5}$

TABLE 1.9 – Erreur relative L^2 en temps au récepteur.

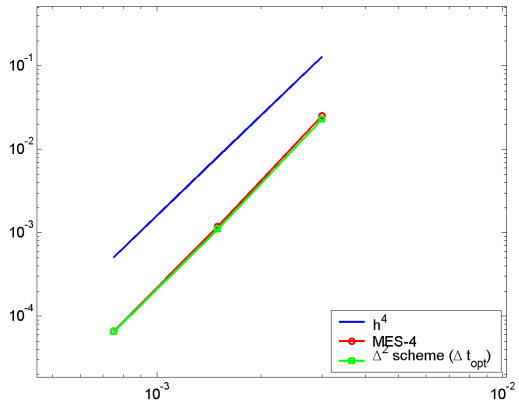

FIGURE 1.5 – Courbes de convergence pour les schémas d'ordre quatre en 2D

1.4.3 Résultats 3D

Dans cette section, nous considérons des configurations similaires à celles du cas 2D. En effet, nous allons prendre pour domaine de propagation le milieu bicouche 3D $\Omega = [-1,1]^3 = \Omega_1 \cap \Omega_2$ où $\Omega_1 = [-1,1]^2 \times [0,1]$ et $\Omega_2 = [-1,1]^2 \times [-1,0]$ sont deux couches homogènes caractérisées par des densités μ_i et des modules de compressibilité ρ_i pour $i = 1,2$. Dans toutes les expériences que nous avons effectuées, nous avons considéré des données initiales nulles et une source qui est une dérivée seconde de Gaussienne en temps et un point source en espace :

$$f = \delta_{x_0} 2\lambda \left(\lambda (t-t_0)^2 - 1\right) e^{-\lambda (t-t_0)^2},$$

avec $x_0 = (0,0,0.5)$, $\lambda = \pi^2 f_0^2$, $f_0 = 5$ et $t_0 = 1/f_0$.
Pour discrétiser l'équation des ondes (1.1), nous avons utilisé les deux méthodes suivantes :

1. le schéma MES-4, basé sur une discrétisation en espace utilisant des polynômes de Lagrange de degré 3 et un paramètre de pénalisation $\gamma_1 = 37$. Avec ces fonctions de base et ce paramètre, la condition CFL du schéma saute-moutons est (expérimentalement) $\Delta t_{LF_4} = 0.030h$ et ainsi, la condition CFL du schéma MES-4 est $\Delta t_{MES-4} = 0.030\sqrt{3}h = 0.052h$.
2. le Δ^2-schéma, avec des fonctions de base de Lagrange P^3 et avec les paramètres de pénalisation $\gamma_1 = 37$, $\gamma_{2,1} = 5.5$ et $\gamma_{2,2} = 0$. La condition CFL de ce schéma est (expérimentalement) $\Delta t_{\Delta^2} = 0.032h$.

Comme pour le cas en dimension un, la condition CFL du Δ^2-schéma est légérement supérieure à la condition CFL du schéma saute-moutons Δt_{LF_4}.
Pour comparer les performances des différentes méthodes, nous allons effectuer deux séries de tests, l'une sur un domaine homogène c'est-à-dire quand $\mu_1 = \mu_2 = 2$ et $\rho_1 = \rho_2 = 2$ et l'autre

sur un domaine hétérogène caractérisé par $\mu_1 = 2$, $\rho_1 = 2$, $\mu_2 = 8$ et $\rho_2 = 4$.
Afin d'illustrer une telle simulation, on représente sur la figure 1.6 (resp. la figure 1.7) un instantané de propagation correspondant à un temps d'expérience $T \simeq 1.94$s en ayant retiré artificiellement le cube $[0,1]^3$ (resp. le parallélépipède $[0,1] \times [-1,1]^2$). Pour des raisons de coût de calcul,

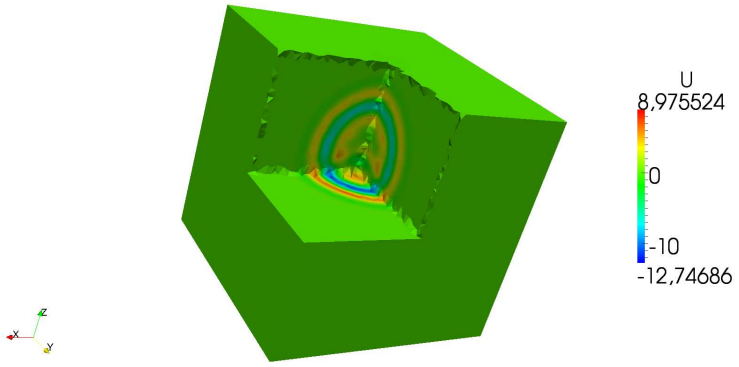

FIGURE 1.6 – Domaine privée du cube $[0,1]^3$.

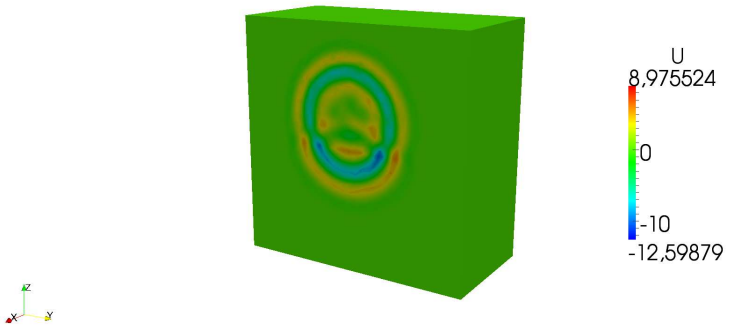

FIGURE 1.7 – Coupe transversale du domaine suivant la coordonnée z.

nous ne représentons pas ici les courbes de convergence de ces deux méthodes, nous allons juste présenter les résultats obtenus sur deux maillages différents du domaine Ω.
Nous calculons la solution exacte en un récepteur situé au point $x_1 = (0, 0, 0.4)$ et nous calculons l'erreur relative $L^2([0,T], x_1)$ pour un temps final approximativement égal à 2s. La solution analytique est calculée grâce à une méthode de Cagniard-De Hoop qui donne la solution du problème en un point du domaine au cours du temps. Les résultats obtenus pour le cas homogène sont présentés dans la Table 1.10 alors que ceux obtenus pour le cas hétérogène sont présentés dans la Table 1.11.

Les résultats des tables 1.10 et 1.11 montrent que les deux méthodes donnent des erreurs relatives très proches. Comme pour les dimensions inférieures, à précision fixée, le coût du Δ^2-schéma est inférieur à celui du schéma MES-4.

Nombre de Tetrahèdres	MES-4	Δ^2-schéma
12140	$2.17\ 10^{-2}$	$2.16\ 10^{-2}$
51400	$7.65\ 10^{-3}$	$7.56\ 10^{-3}$

TABLE 1.10 – Erreur relative L^2 en temps dans le cas homogène.

Nombre de Tetrahèdres	MES-4	Δ^2-schéma
12573	$2.54\ 10^{-1}$	$2.55\ 10^{-1}$
61176	$1.32\ 10^{-1}$	$1.32\ 10^{-1}$

TABLE 1.11 – Erreur relative L^2 en temps dans le cas hétérogène.

1.5 Conclusion

Dans ce chapitre, nous avons construit de nouveaux schémas d'ordre élevé en temps et en espace pour résoudre l'équation des ondes acoustiques et nous avons prouvé la convergence du Δ^2-schéma. Les résultats numériques que nous avons présentés illustrent le fait que les coûts de calculs de ces schémas sont équivalents à celui du schéma saute-moutons et sont donc plus petits que ceux des schémas MES-4 et MES-6 (respectivement 13% et 54%). Dans ce chapitre, nous nous sommes limités à la détermination empirique de la condition CFL et des paramètres de pénalisation. Il est important de noter que, même pour le schéma saute-moutons, le problème de la détermination du paramètre de pénalisation optimal n'a été résolu que dans des cas très particuliers. De plus, aucune étude analytique de la condition CFL n'a encore été proposée.

Une autre propriété très intéressante de ces schémas est le fait qu'ils semblent être appropriés à la p-adaptativité. En effet, si l'on applique par exemple le Δ^2-schéma avec un maillage composé de cellules P^1 , la forme $a_{2h}(\phi_i, \phi_j)$ s'annule pour toutes les fonctions de base et le schéma est uniquement d'ordre deux en espace et en temps. Par conséquent, en utilisant un maillage composé de cellules P^1 et P^3, on obtient facilement un schéma d'ordre quatre en espace et en temps sur les cellules P^3 et d'ordre deux en espace et en temps pour les cellules P^1.

1.A Preuves des théorèmes et lemmes auxiliaires

Tout d'abord, rappelons quelques inégalités de trace (cf. [53], [48]) et l'inégalité inverse (cf. [11]) qui nous serviront dans la suite

Proposition 1.A.1 (Inégalités de trace)**.** *Pour tout* $v \in P^p(K)$*,*

$$\|v\|^2_{L^2(\partial K)} \leq \frac{(p+1)(p+2)}{2h}\|v\|^2_{L^2(K)}, \quad \|\nabla v\|^2_{L^2(\partial K)} \leq \frac{\lambda(p)}{h^3}\|v\|^2_{L^2(K)} \tag{1.31}$$

avec $\lambda(p) \sim p^6$*.*

Proposition 1.A.2 (Inégalité inverse)**.** *Soient* $n \in \mathbb{N}$ *et* $j \geq 1$*,*

$$|v|_{n+j,K} \leq \alpha h^{-n}|v|_{j,K} \quad \forall v \in V_h \tag{1.32}$$

1.A.1 Preuve du Théorème 1.3.6

Théorème 4.5 *Sous la condition CFL (1.24), la forme bilinéaire a_h satisfait une condition de stabilité sur V_h, i.e. $\exists C > 0$ telle que*

$$a_h(u,u) \geq C\|u\|_{DG_2}^2, \quad \forall u \in V_h$$

et a_h est continue sur V_h i.e. $\exists C > 0$ telle que :

$$|a_h(u,v)| \leq C\|u\|_{DG_2}\|v\|_{DG_2}, \quad \forall u, v \in V_h.$$

Démonstration. D'après les lemmes 1.3.4 et 1.3.5, nous avons, $\forall u \in V_h$,

$$a_h(u,u) \geq C_{coer1}\|u\|_{DG_2}^2 - \frac{\Delta t^2}{12} C_{cont2}\|u\|_{DG_4}^2.$$

De plus, on a le résultat suivant qui sera prouvé dans la sous-section 1.A.2 :

Lemme 1.A.3. *Il existe $\gamma > 0$ tel que pour tout $u \in V_h$:*

$$\|u\|_{DG_4}^2 \leq \gamma h^{-2}\|u\|_{DG_2}^2$$

Par conséquent,

$$a_h(u,u) \geq C_{coer1}\|u\|_{DG_2}^2 - \frac{\Delta t^2}{12} C_{cont2}\|u\|_{DG_4}^2 \geq \left(C_{coer1} - \gamma\frac{\Delta t^2}{12}h^{-2}C_{cont2}\right)\|u\|_{DG_2}^2.$$

En utilisant la condition CFL (1.24), on a :

$$\left(C_{coer1} - \gamma\frac{\Delta t^2}{12}h^{-2}C_{cont2}\right)\|u\|_{DG_2}^2 \geq \left(C_{coer1} - \gamma C_{cont2}\frac{\beta^2}{12}\right)\|u\|_{DG_2}^2.$$

Alors, pour $\beta < \sqrt{\frac{12}{C\gamma C_{cont2}}}$, il existe une constante $C > 0$ telle que

$$a_h(u,u) \geq C\|u\|_{DG_2}^2, \quad \forall u \in V_h.$$

Concernant la seconde partie du théorème, les lemmes 1.3.4 et 1.3.5 impliquent que, $\forall u, v \in V_h$,

$$|a_h(u,v)| \leq C_{\text{cont1}}\|u\|_{DG_2}\|v\|_{DG_2} + \frac{\Delta t^2}{12} C_{\text{cont2}}\|u\|_{DG_4}\|v\|_{DG_4}.$$

En utilisant le lemme 1.A.3, nous avons

$$|a_h(u,v)| \leq \left(C_{\text{cont1}} + \gamma\frac{\Delta t^2}{12h^2}C_{\text{cont2}}\right)\|u\|_{DG_2}\|v\|_{DG_2}.$$

Dans ce cas, grâce à la condition CFL (1.24),

$$|a_h(u,v)| \leq C\|u\|_{DG_2}\|v\|_{DG_2}$$

avec $C = C_{\text{cont1}} + \gamma\frac{\beta^2}{12}C_{\text{cont2}}$. □

1.A.2 Preuve du lemme 1.A.3

Démonstration. Nous avons à majorer chaque terme contenu dans la norme $\|\cdot\|_{DG_4}$ par des termes contenus dans $\|\cdot\|_{DG_2}$. Etudions le terme $|u|_{2,h}$ pour tout $u \in V_h$.
Soit $u \in V_h$, grâce à l'inégalité inverse (1.32), on a :

$$|u|_{2,h}^2 = \sum_{K\in\mathcal{T}_h} |u|_{2,K}^2 \leq \alpha^2 \sum_{K\in\mathcal{T}_h} h_K^{-2}|u|_{1,K}^2 \leq \alpha^2 h^{-2}|u|_{1,h}^2$$

Maintenant, nous nous intéressons au terme $\|\alpha_{2,1}^{-1/2}\{\{\nabla(\Delta u)\cdot\boldsymbol{\nu}\}\}\|_{L^2(\Gamma)}^2$ et nous avons à considérer la restriction de cette norme à une arête unique $F \in \mathcal{F}_h$.
Soit $F \in \mathcal{F}_h$, en utilisant l'inégalité algébrique $(a+b)^2 \leq 2\left(a^2+b^2\right)$ et les inégalités de trace (1.31), nous avons :

$$\begin{aligned}
\|\alpha_{2,2}^{-1/2}\{\{\nabla(\Delta u)\cdot\boldsymbol{\nu}\}\}\|_{L^2(F)}^2 &\leq \tfrac{h^3}{p^6\gamma_{2,2}}\left(\left\|\tfrac{1}{2}\left(\nabla(\Delta u^+)\cdot\boldsymbol{\nu}+\nabla(\Delta u^-)\cdot\boldsymbol{\nu}\right)\right\|_{L^2(F)}^2\right)\\
&\leq \tfrac{h^3}{2p^6\gamma_{2,2}}\left(\|\nabla(\Delta u^+)\cdot\boldsymbol{\nu}\|_{L^2(F)}^2+\|\nabla(\Delta u^-)\cdot\boldsymbol{\nu}\|_{L^2(F)}^2\right)\\
&\leq \tfrac{\lambda(p)}{2p^6\gamma_{2,2}}\left(\|\Delta u\|_{L^2(K^+)}^2+\|\Delta u\|_{L^2(K^-)}^2\right).
\end{aligned}$$

Par conséquent, nous obtenons

$$\|\alpha_{2,2}^{-1/2}\{\{\nabla(\Delta u)\cdot\boldsymbol{\nu}\}\}\|_{L^2(\Gamma)}^2 \leq \frac{\lambda(p)}{p^6\gamma_{2,2}}|u|_{2,h}^2.$$

Par l'inégalité inverse (1.32), on a alors

$$\|\alpha_{2,2}^{-1/2}\{\{\nabla(\Delta u)\cdot\boldsymbol{\nu}\}\}\|_{L^2(\Gamma)}^2 \leq \frac{\alpha^2\lambda(p)\,h^{-2}}{p^6\gamma_{2,2}}|u|_{1,h}^2.$$

Avec un raisonnement similaire, l'autre terme faisant intervenir une moyenne vérifie :

$$\|\alpha_{2,1}^{-1/2}\{\{\Delta u\}\}\|_{L^2(\Gamma)}^2 \leq \frac{\alpha^2(p+1)(p+2)\,h^{-2}}{2p^2\gamma_{2,1}}|u|_{1,h}^2.$$

Maintenant, nous considérons $\|\alpha_{2,1}^{1/2}[\![\nabla u\cdot\boldsymbol{\nu}]\!]\|_{L^2(\Gamma)}^2$.
Soit $F \in \mathcal{F}_h$,

$$\begin{aligned}
\|\alpha_{2,1}^{1/2}[\![\nabla u\cdot\boldsymbol{\nu}]\!]\|_{L^2(F)}^2 &= \alpha_{2,1}\|\nabla u^+\cdot\boldsymbol{\nu}-\nabla u^-\cdot\boldsymbol{\nu}\|_{L^2(\Gamma)}^2\\
&= \alpha_{2,1}\int_F\left(\nabla v^+\cdot\boldsymbol{\nu}\right)^2+\left(\nabla v^-\cdot\boldsymbol{\nu}\right)^2-2\left(\nabla v^+\cdot\boldsymbol{\nu}\right)\left(\nabla v^-\cdot\boldsymbol{\nu}\right)ds\\
&\leq \alpha_{2,1}\left(\|\nabla v^+\|_{L^2(F)}^2+\|\nabla v^-\|_{L^2(F)}^2+2\int_F|\left(\nabla v^+\cdot\boldsymbol{\nu}\right)\left(\nabla v^-\cdot\boldsymbol{\nu}\right)|ds\right).
\end{aligned}$$

En utilisant l'inégalité de Cauchy-Schwarz, nous obtenons

$$\|\alpha_{2,1}^{1/2}[\![\nabla u\cdot\boldsymbol{\nu}]\!]\|_{L^2(F)}^2 \leq \alpha_{2,1}\left(\|\nabla v^+\|_{L^2(F)}^2+\|\nabla v^-\|_{L^2(F)}^2+2\|\nabla v^+\|_{L^2(F)}\|\nabla v^-\|_{L^2(F)}\right).$$

Nous en déduisons que

$$\|\alpha_{2,1}^{1/2}[\![\nabla u\cdot\boldsymbol{\nu}]\!]\|_{L^2(F)}^2 \leq \alpha_{2,1}\left(\|\nabla v^+\|_{L^2(F)}^2+\|\nabla v^-\|_{L^2(F)}^2+\|\nabla v^+\|_{L^2(F)}^2+\|\nabla v^-\|_{L^2(F)}^2\right)$$

ce qui amène à

$$\|\alpha_{2,1}^{1/2}[\![\nabla u\cdot\boldsymbol{\nu}]\!]\|_{L^2(F)}^2 \leq 2\alpha_{2,1}\left(\|\nabla v^+\|_{L^2(F)}^2+\|\nabla v^-\|_{L^2(F)}^2\right).$$

De plus, grâce aux inégalités de trace (1.31), on a

$$\|\alpha_{2,1}^{1/2}\left[\!\left[\nabla u\cdot\boldsymbol{\nu}\right]\!\right]\|_{L^2(\Gamma)}^2 \leq \sum_{F\in\mathcal{F}_h}\frac{2\gamma_{2,1}p^2}{h}\frac{(p+1)(p+2)}{2h}\left(\|\nabla v^+\|_{L^2(K^+)}^2+\|\nabla v^-\|_{L^2(K^-)}^2\right).$$

Ainsi, il vient

$$\|\alpha_{2,1}^{1/2}\left[\!\left[\nabla u\cdot\boldsymbol{\nu}\right]\!\right]\|_{L^2(\Gamma)}^2 \leq 2h^{-2}\gamma_{2,1}p^2(p+1)(p+2)|u|_{1,h}^2.$$

Finalement, nous considérons le terme $|\alpha_{2,2}^{1/2}u|_*^2$ qui peut être facilement réécrit comme

$$|\alpha_{2,2}^{1/2}u|_*^2=\frac{\gamma_1 p^2}{h}\left(\frac{\gamma_{2,2}p^4}{\gamma_1 h^2}\right)|u|_*^2=\frac{\gamma_{2,2}p^4}{\gamma_1 h^2}|\alpha_1^{1/2}u|_*^2$$

et donc

$$\|u\|_{DG_4}^2\leq\gamma h^{-2}\left(|u|_{1,h}^2+|\alpha_1^{1/2}u|_*^2\right)\leq\gamma h^{-2}\|u\|_\sim^2,$$

où $\|\cdot\|_\sim$ représente le terme :

$$\|v\|_\sim^2:=|v|_{1,h}^2+|\alpha_1^{1/2}v|_*^2.$$

Dans [5], on peut trouver l'équivalence sur V_h entre la norme $\|\cdot\|_{DG_2}$ et la norme $\|\cdot\|_\sim$.
Par conséquent,

$$\|u\|_{DG_4}^2\leq\gamma h^{-2}\|u\|_{DG_2}^2.$$

□

1.A.3 Preuve du lemme 1.3.9

Lemme 4.8 *Si $u\in H^{p+1}(\Omega)$ avec $p\geq 1$. Alors il existe $C>0$ indépendant de h telle que*

$$\begin{aligned}\|u-p_h(u)\|_{DG_2} &\leq Ch^p\|u\|_{p+1},\\ \|u-p_h(u)\|_{DG_4} &\leq Ch^{p-1}\|u\|_{p+1},\\ \|u-p_h(u)\|_0 &\leq Ch^{p+1}\|u\|_{p+1}.\end{aligned}$$

1.A.3.1 Preuve de la première estimation

Nous cherchons une majoration de la quantité $\|u-p_h(u)\|_{DG_2}$. Introduisons un projecteur $\pi_p^h u\in V_h$ de la solution exacte u, on a

$$\|u-p_h(u)\|_{DG_2}\leq\|u-\pi_p^h u\|_{DG_2}+\|\pi_p^h u-p_h(u)\|_{DG_2}. \tag{1.33}$$

L'existence d'un tel projecteur est garantie par le théorème suivant (cf. [51]).

Théorème 1.A.4. *On suppose que l'on a une partition $\mathcal{T}_h$ de Ω composée de simplexes ou de parallélépipèdes de dimension d. Alors, pour tous $u\in H^t(\Omega)$, t et $r\in\mathbb{N}$, il existe un projecteur*

$$\pi_r^h\,:\,H^t(\Omega)\longrightarrow V_h,\qquad\left(\pi_r^h u\right)_{|K}=\pi_r^h\left(u_{|K}\right)$$

tel que, pour $0\leq q\leq t$,

$$\|u-\pi_r^h u\|_{q,K}\leq C\frac{h^{s-q}}{r^{t-q}}\|u\|_{t,K}\quad\forall K\in\mathcal{T}_h,$$

et pour $0 \leq q \leq t-1$

$$\|D^{\alpha}\left(u-\pi_r^h u\right)\|_{0,\partial K} \leq C\frac{h^{s-q-\frac{1}{2}}}{r^{t-q-\frac{1}{2}}}\|u\|_{t,K}, \quad |\alpha|=q,\ \forall K \in \mathcal{T}_h,$$

où $s = min\,(r+1,t)$ *et* C *est une constante indépendante de* u*,* h *et* r*, mais dépendante de* t*.*

D'après ce théorème, en choisissant $r=p$, on en déduit que [5]

$$\begin{cases} \|u-\pi_p^h u\|_{DG_2} \leq Ch^p\|u\|_{p+1}, \\ \|u-\pi_p^h u\|_{DG_4} \leq Ch^{p-1}\|u\|_{p+1}. \end{cases} \tag{1.34}$$

A présent, nous devons trouver une majoration similaire de la quantité $\|\pi_p^h u - p_h(u)\|_{DG_2}$. Grâce aux propriétés de la projection de Galerkine et de la forme bilinéaire a_h, nous avons

$$\begin{aligned} a_h\left(\pi_p^h u - p_h(u), \pi_p^h u - p_h(u)\right) &= a_h\left(\pi_p^h u, \pi_p^h u - p_h(u)\right) - a_h\left(p_h(u), \pi_p^h u - p_h(u)\right) \\ &= a_h\left(\pi_p^h u - u, \pi_p^h u - p_h(u)\right). \end{aligned} \tag{1.35}$$

Puisque $\pi_p^h u - p_h(u) \in V_h$, d'après le théorème 1.3.6, il existe une constante C telle que

$$a_h\left(\pi_p^h u - p_h(u), \pi_p^h u - p_h(u)\right) \geq C\|\pi_p^h u - p_h(u)\|_{DG_2}^2. \tag{1.36}$$

De plus, nous cherchons une borne supérieure du terme $a_h\left(\pi_p^h u - u, \pi_p^h u - p_h(u)\right)$ mais comme $\pi_p^h u - u \in V_h + H^{p+1}(\Omega)$, nous ne pouvons pas utiliser la continuité de a_h dans V_h.
Soient $u, v \in V_h + H^{p+1}(\Omega)$, on a

$$|a_h(u,v)| \leq |a_{1h}(u,v)| + \frac{\Delta t^2}{12}|a_{2h}(u,v)|. \tag{1.37}$$

D'après le lemme 1.3.4, il existe une constante positive $C>0$ telle que

$$|a_{1h}(u,v)| \leq C\|u\|_{DG_2}\|v\|_{DG_2}, \qquad \forall u, v \in V_h + H^{p+1}(\Omega).$$

D'après la définition de la norme $|||\cdot|||$, il est clair que

$$|a_{1h}(u,v)| \leq C|||u|||\,|||v|||, \qquad \forall u, v \in V_h + H^{p+1}(\Omega). \tag{1.38}$$

De la même manière, il a été prouvé dans [43] que $\forall u, v \in H^{p+1}(\Omega)$, $\exists C>0$:

$$|a_{2h}(u,v)| \leq C\|u\|_{DG_4}\|v\|_{DG_4}$$

ce qui implique

$$\frac{\Delta t^2}{12}|a_{2h}(u,v)| \leq C|||u|||\,|||v|||, \qquad \forall u, v \in V_h + H^{p+1}(\Omega). \tag{1.39}$$

En utilisant (1.38) et (1.39), nous obtenons

$$a_h\left(\pi_p^h u - u, \pi_p^h u - p_h(u)\right) \leq C|||\pi_p^h u - u|||\,|||\pi_p^h u - p_h(u)|||.$$

Alors, en combinant cette inégalité avec (1.36) et (1.35), on a

$$\|\pi_p^h u - p_h(u)\|_{DG_2}^2 \leq C|||\pi_p^h u - u|||\,|||\pi_p^h u - p_h(u)|||. \tag{1.40}$$

Cependant, $\pi_p^h u - p_h(u) \in V_h$ et nous savons (cf. lemme 1.A.3) que $\forall v \in V_h$, $\exists C > 0$ telle que $\|v\|_{DG_4}^2 \leq Ch^{-2}\|v\|_{DG_2}^2$.
Par conséquent, $\forall v \in V_h$, la condition CFL (1.24) implique

$$|||v|||^2 \leq \|v\|_{DG_2}^2 + C\frac{\Delta t^2}{12h^2}\|v\|_{DG_2}^2 \leq \widetilde{C}\|v\|_{DG_2}^2$$

avec $\widetilde{C} = 1 + C\frac{\beta^2}{12}$, et ainsi

$$\|\pi_p^h u - p_h(u)\|_{DG_2}^2 \leq C|||\pi_p^h u - u|||\,\|\pi_p^h u - p_h(u)\|_{DG_2}, \tag{1.41}$$

ce qui se réécrit

$$\|\pi_p^h u - p_h(u)\|_{DG_2} \leq C|||\pi_p^h u - u|||. \tag{1.42}$$

De plus, en utilisant (1.34), nous avons, $\forall u \in H^{p+1}(\Omega)$

$$|||\pi_p^h u - u|||^2 \leq C\left(h^{2p} + \frac{\Delta t^2}{12}h^{2(p-1)}\right)\|u\|_{p+1,\Omega}^2$$

donc, sous la condition CFL (1.24),

$$|||\pi_p^h u - u|||^2 \leq C_1 h^{2p}\|u\|_{p+1,\Omega}^2 + \beta^2\frac{C_2}{12}h^{2p}\|u\|_{p+1,\Omega}^2$$

ce qui mène à

$$|||\pi_p^h u - u||| \leq Ch^p\|u\|_{p+1,\Omega}. \tag{1.43}$$

Finalement, en considérant (1.42), (1.43) et (1.34), nous obtenons

$$\|u - p_h(u)\|_{DG_2} \leq Ch^p\|u\|_{p+1,\Omega} \tag{1.44}$$

ce qui prouve la première estimation.

1.A.3.2 Preuve de la deuxième estimation

Comme dans la preuve de la première estimation, nous avons

$$\|u - p_h(u)\|_{DG_4} \leq \|u - \pi_p^h u\|_{DG_4} + \|\pi_p^h u - p_h(u)\|_{DG_4}. \tag{1.45}$$

et nous avons déjà vu que

$$\|u - \pi_p^h u\|_{DG_4} \leq Ch^{p-1}\|u\|_{p+1} \tag{1.46}$$

Pour majorer le second terme de (1.45), nous pouvons remarquer que $\pi_p^h u - p_h(u) \in V_h$. En appliquant le lemme 1.A.3, il vient alors

$$\|\pi_p^h u - p_h(u)\|_{DG_4} \leq \gamma h^{-1}\|\pi_p^h u - p_h(u)\|_{DG_2}.$$

En utilisant la première estimation du lemme 1.3.9, nous avons

$$\|\pi_p^h u - p_h(u)\|_{DG_4} \leq \gamma h^{p-1}\|u\|_{p+1}. \tag{1.47}$$

En combinant (1.45), (1.46) et (1.47), on a

$$\|u - p_h(u)\|_{DG_4} \leq Ch^{p-1}\|u\|_{p+1}.$$

1.A.3.3 Preuve de la troisième estimation

Considérons tout d'abord le problème auxiliaire suivant

$$\begin{cases} \Delta z + \dfrac{\Delta t^2}{12}\Delta^2 z &= u - p_h(u) \quad \text{dans } \Omega, \\ z &= 0 \quad \text{sur } \partial\Omega, \\ \Delta z &= 0 \quad \text{sur } \partial\Omega. \end{cases} \tag{1.48}$$

Tout d'abord, nous devons prouver qu'il existe une unique solution $z \in H^4(\Omega)$ à ce problème et obtenir une majoration des quantités $\|z\|_2$ et $\|z\|_4$.
Le problème (1.48) peut être réécrit comme deux problèmes couplés :

$$\begin{cases} \Delta z_1 &= u - p_h(u) \quad \text{dans } \Omega, \\ z_1 &= 0 \quad \text{sur } \partial\Omega, \end{cases} \tag{1.49}$$

et

$$\begin{cases} z + \dfrac{\Delta t^2}{12}\Delta z &= z_1 \quad \text{dans } \Omega, \\ z &= 0 \quad \text{sur } \partial\Omega. \end{cases} \tag{1.50}$$

Notons que la condition $\Delta z = 0$ sur $\partial\Omega$ est imposée implicitement car $z_1 = z + \dfrac{\Delta t^2}{12}\Delta z$ dans Ω et que l'on fixe $z_1 = 0$ sur $\partial\Omega$ ainsi que $z = 0$ sur $\partial\Omega$. Le domaine Ω est supposé convexe. $u - p_h(u)$ appartenant à $L^2(\Omega)$, la régularité elliptique implique ainsi que la solution z_1 du problème (1.49) appartient à $H^2(\Omega) \cap H_0^1(\Omega)$ et qu'il existe une constante C_1 telle que

$$\|z_1\|_2 \leq C_1 \|u - p_h(u)\|_0. \tag{1.51}$$

Maintenant, nous nous intéressons au problème (1.50) qui est un problème de Helmholtz.
Nous savons que, grâce à l'alternative de Fredholm, le problème

$$\begin{cases} Lu &= \lambda u + f \quad \text{dans } \Omega, \\ u &= 0 \quad \text{sur } \partial\Omega, \end{cases} \tag{1.52}$$

admet une unique solution $u \in H_0^1(\Omega)$ si $f \in L^2(\Omega)$ et $\lambda \notin Sp(L)$.
Ici, $L = -\Delta$, $\lambda = {}^{12}/_{\Delta t^2}$ et $f = ({}^{12}/_{\Delta t^2})\, z_1 \in H^2(\Omega)$ alors, si $\lambda \notin Sp(-\Delta)$ il existe un unique $z \in H_0^1(\Omega)$ solution du problème (1.52).
A présent, en appliquant le théorème de régularité H^m sur le bord (cf. chap. 6 dans [26]) il s'ensuit qu'il existe un unique $z \in H^4(\Omega)$ solution du problème (1.48) tel que

$$\|z\|_2 \leq C_2 \|z_1\|_0 \quad \text{et} \quad \|z\|_4 \leq C_2 \|z_1\|_2. \tag{1.53}$$

Par conséquent, en combinant (1.51) et (1.53), il vient que

$$\|z\|_2 \leq C \|u - p_h(u)\|_0 \tag{1.54}$$

avec une constante $C > 0$.

En multipliant la première équation de (1.48) par $u - p_h(u)$ et en l'intégrant sur Ω on a, en utilisant la consistance de a_h,

$$\|u - p_h(u)\|_0^2 = a_h(z, u - p_h(u)). \tag{1.55}$$

De plus, $\forall z_h \in V_h$ nous avons, d'après la définition de $p_h(u)$,

$$a_h(z, u - p_h(u)) = a_h(z - z_h, u - p_h(u))$$

et, grâce aux continuités de a_{1h} et a_{2h} dans $H^{p+1}(\Omega) + V_h$, il existe une constante C telle que

$$a_h(z, u - p_h(u)) \leq C\left(\|z - z_h\|_{DG_2}\|u - p_h(u)\|_{DG_2} + \frac{\Delta t^2}{12}\|z - z_h\|_{DG_4}\|u - p_h(u)\|_{DG_4}\right). \tag{1.56}$$

En choisissant $z_h = \pi_1^h z$, on a, d'après le théorème 1.A.4,

$$\|z - z_h\|_{DG_2} \leq Ch\|z\|_2$$

et , grâce à l'estimation (1.54), on obtient

$$\|z - z_h\|_{DG_2} \leq Ch\|u - p_h(u)\|_0. \tag{1.57}$$

A présent, de la même manière

$$\|z - z_h\|_{DG_4} \leq C\|z\|_4.$$

En utilisant la seconde inégalité de (1.53) et (1.51), il vient que

$$\|z - z_h\|_{DG_4} \leq C\|u - p_h(u)\|_0. \tag{1.58}$$

En combinant les résultats (1.55), (1.56), (1.57) et (1.58), nous concluons qu'il existe une constante C telle que

$$\|u - p_h(u)\|_0 \leq C\left(h\|u - p_h(u)\|_{DG_2} + \frac{\Delta t^2}{12}\|u - p_h(u)\|_{DG_4}\right).$$

D'après la condition CFL (1.24),

$$\|u - p_h(u)\|_0 \leq Ch\left(\|u - p_h(u)\|_{DG_2} + \frac{\beta^2 h}{12}\|u - p_h(u)\|_{DG_4}\right).$$

Finalement, nous concluons en utilisant les deux premières estimations du lemme 1.3.9.

1.A.4 Preuve de la proposition 1.3.10

Proposition 4.9 *Sous la condition CFL (1.24), il existe une constante $C > 0$ telle que*

$$\max_{n=0,\dots,N}\|\phi^n\|_0 \leq C\left(\|e^0\|_0 + \|\eta^0\|_0\right) + C^2\Delta t\sum_{n=0}^{N-1}\|R^n\|_0.$$

où $N = \frac{T}{\Delta t}$ est le nombre d'itérations et $C > 0$ est une constante indépendante de h, Δt et T.

Démonstration. Nous rappelons que nous avons

$$\max_{n=0,\dots,N}\|e^n\|_0 \leq \max_{n=0,\dots,N}\|\phi^n\|_0 + \max_{n=0,\dots,N}\|\eta^n\|_0.$$

On a déjà une majoration du deuxième terme (cf. lemme 1.3.9) donc nous devons juste borner le premier.
Grâce à l'équation des ondes, on a

$$\partial_t^2 u^n + \frac{\Delta t^2}{12}\Delta^2 u^n - \frac{\Delta t^2}{12}\Delta^2 u^n - \Delta u^n = f^n.$$

Puisque a_h est consistant, nous avons, $\forall v \in V_h$:

$$\left(\partial_t^2 u^n + \frac{\Delta t^2}{12}\Delta^2 u^n, v\right) + a_h\left(u^n, v\right) = \left(f^n, v\right). \tag{1.59}$$

De plus, (1.18) devient

$$\left(\delta U^n, v\right) + a_h\left(U^n, v\right) = \left(f^n + \frac{\Delta t^2}{12}\frac{\partial^2 f^n}{\partial t^2} + \Delta f^n, v\right)$$

La différence entre cette équation et (1.59) donne :

$$\left(\partial_t^2 u^n + \frac{\Delta t^2}{12}\Delta^2 u^n - \delta^2 U^n, v\right) + a_h\left(u^n - U^n, v\right) = -\frac{\Delta t^2}{12}\left(\frac{\partial^2 f^n}{\partial t^2} + \Delta f^n, v\right)$$

c'est-à-dire, $\forall v \in V_h$, $n = 1, \ldots, N-1$:

$$\left(\partial_t^2 u^n - \delta^2 w^n + \delta^2 w^n - \delta^2 U^n + \frac{\Delta t^2}{12}\Delta^2 u^n, v\right) + a_h\left(u^n - w^n + w^n - U^n, v\right) = -\frac{\Delta t^2}{12}\left(\frac{\partial^2 f^n}{\partial t^2} + \Delta f^n, v\right).$$

Or $a_h\left(u^n - w^n, v\right) = 0\ \forall v \in V_h$ par définition de w^n. Il vient donc :

$$\left(\delta^2\phi^n, v\right) + a_h\left(\phi^n, v\right) = \left(\delta^2 w^n - \partial_t^2 u^n - \frac{\Delta t^2}{12}\left(\Delta^2 u^n + \frac{\partial^2 f^n}{\partial t^2} + \Delta f^n\right), v\right).$$

En dérivant deux fois par rapport au temps l'équation des ondes, on a l'égalité suivante :

$$\Delta^2 u^n + \frac{\partial^2 f^n}{\partial t^2} + \Delta f^n = \partial_t^4 u^n$$

i.e. $\forall v \in V_h$, $n = 1, \ldots, N-1$:

$$\left(\delta^2\phi^n, v\right) + a_h\left(\phi^n, v\right) = \left(r^n, v\right).$$

Si nous sommons cette égalité de $n = 1$ à $n = m$, $1 \leq m \leq N-1$, nous avons :

$$\left(\frac{\phi^{m+1} - \phi^m}{\Delta t^2}, v\right) + \left(\frac{\phi^0 - \phi^1}{\Delta t^2}, v\right) + \sum_{n=1}^{m} a_h\left(\phi^n, v\right) = \sum_{n=1}^{m}\left(r^n, v\right)$$

ou encore

$$\left(\frac{\phi^{m+1} - \phi^m}{\Delta t}, v\right) + \Delta t\sum_{n=1}^{m} a_h\left(\phi^n, v\right) = \Delta t\sum_{n=1}^{m}\left(r^n, v\right) + \Delta t\left(r^0, v\right).$$

Nous notons $\xi^n = \Delta t\sum_{n=1}^{m}\phi^n$ avec $\xi^0 = 0$.
Alors, $\forall v \in V_h$, $0 \leq m \leq N-1$:

$$\left(\frac{\phi^{m+1} - \phi^m}{\Delta t}, v\right) + a_h\left(\xi^m, v\right) = \left(R^m, v\right)$$

où $\left(R^m, v\right) = \Delta t\sum_{n=0}^{m}\left(r^n, v\right)$.
Nous choisissons $v = \phi^{m+1} + \phi^m \in V_h$. On obtient donc $\forall m \in \{0 \ldots n-1\}$

$$\|\phi^{m+1}\|_0^2 - \|\phi^m\|_0^2 + \Delta t a_h\left(\xi^m, \phi^{m+1} + \phi^m\right) = \Delta t\left(R^m, \phi^{m+1} + \phi^m\right).$$

et nous sommons de $m = 0$ à $n-1$, $\forall n \in \{1 \dots N\}$ pour obtenir

$$\|\phi^n\|_0^2 - \|\phi^0\|_0^2 + \Delta t \sum_{m=0}^{n-1} a_h\left(\xi^m, \phi^{m+1} + \phi^m\right) = \Delta t \sum_{m=0}^{n-1} \left(R^m, \phi^{m+1} + \phi^m\right).$$

D'après la définition de ξ^m, on a, $\forall m \in \{1 \dots N-1\} : \xi^{m+1} - \xi^{m-1} = \Delta t\left(\phi^{m+1} + \phi^m\right)$. Ainsi

$$\begin{aligned}
\Delta t \sum_{m=0}^{n-1} a_h\left(\xi^m, \phi^{m+1} + \phi^m\right) &= \sum_{m=1}^{n-1} a_h\left(\xi^m, \xi^{m+1} - \xi^{m-1}\right) \\
&= \sum_{m=1}^{n-1} a_h\left(\xi^m, \xi^{m+1}\right) - \sum_{m=0}^{n-2} a_h\left(\xi^{m+1}, \xi^m\right) \\
&= a_h\left(\xi^{n-1}, \xi^n\right) - a_h\left(\xi^1, \xi^0\right), \\
&= a_h\left(\xi^{n-1}, \xi^n\right).
\end{aligned}$$

car $\xi^0 = 0$.
De plus, d'après la bilinéarité de a_h, nous avons

$$a_h\left(\xi^{n-1}, \xi^n\right) = a_h\left(\frac{\xi^{n-1} + \xi^n}{2}, \frac{\xi^{n-1} + \xi^n}{2}\right) - a_h\left(\frac{\xi^n - \xi^{n-1}}{2}, \frac{\xi^n - \xi^{n-1}}{2}\right).$$

En utilisant le théorème 1.3.6, comme $\xi^{n+1} - \xi^n = \Delta t \phi^{n+1}$, $\forall n \in \{0, \dots, N-1\}$, il vient que

$$a_h\left(\xi^{n-1}, \xi^n\right) \geq -\frac{\Delta t^2}{4} a_h\left(\phi^n, \phi^n\right).$$

Par conséquent, on obtient

$$\|\phi^n\|_0^2 - \frac{\Delta t^2}{4} a_h\left(\phi^n, \phi^n\right) \leq \|\phi^0\|_0^2 + \Delta t \sum_{m=0}^{n-1} \left(R^m, \phi^{m+1} + \phi^m\right). \tag{1.60}$$

Grâce à la continuité de a_{1h} et de a_{2h} mais aussi au lemme 1.A.3

$$\begin{aligned}
a_h\left(\phi^n, \phi^n\right) &\leq |a_{1h}\left(\phi^n, \phi^n\right)| + \tfrac{\Delta t^2}{12} |a_{2h}\left(\phi^n, \phi^n\right)| \\
&\leq C_{cont1} \|\phi^n\|_{DG_2}^2 + \tfrac{\Delta t^2}{12} C_{cont2} \|\phi^n\|_{DG_4}^2 \\
&\leq \left(C_{cont1} + \tfrac{\Delta t^2}{12} C_{cont2} \gamma h^{-2}\right) \|\phi^n\|_{DG_2}^2
\end{aligned}$$

Or, d'après le lemme 1.3.4 :

$$C_{coer1} \|\phi^n\|_{DG_2}^2 \leq C_S c_{\max}^2 h^{-2} \|\phi^n\|_0^2,$$

donc

$$a_h\left(\phi^n, \phi^n\right) \leq \left(C_{cont1} + \frac{\Delta t^2}{12} C_{cont2} \gamma h^{-2}\right) \frac{C_S}{C_{coer1}} c_{\max}^2 h^{-2} \|\phi^n\|_0^2.$$

En utilisant cette inégalité dans (1.60), on a, sous la condition CFL (1.24),

$$C^* \|\phi^n\| \leq \|\phi^0\|_0^2 + \Delta t \sum_{m=0}^{n-1} \left(R^m, \phi^{m+1} + \phi^m\right), \quad 0 \leq n \leq N,$$

avec

$$C^* := 1 - \frac{\beta^2}{4} \left(C_{cont1} + \frac{\beta^2}{12} C_{cont2} \gamma\right) \frac{C_S}{C_{coer1}} c_{\max}^2, \tag{1.61}$$

qui est évidemment positif pour β suffisament petit.
Nous appliquons l'inégalité de Cauchy-Schwarz

$$C^*\|\phi^n\|_0^2 \leq \|\phi^0\|_0^2 + \Delta t \sum_{m=0}^{n-1} \|R^m\|_0 \|\phi^{m+1} + \phi^m\|_0.$$

Alors, $\|\phi^m + \phi^{m+1}\|_0 \leq 2 \max_{m=0,\ldots,N} \|\phi^m\|_0$ implique

$$C^*\|\phi^n\|_0^2 \leq \|\phi^0\|_0^2 + 2\left(\Delta t \sum_{m=0}^{N-1} \|R^m\|_0\right)\left(\max_{m=0,\ldots,N} \|\phi^m\|_0\right)$$

En utilisant l'inégalité algébrique $2ab \leq \varepsilon^{-1}a^2 + \varepsilon b^2$, $\forall \epsilon > 0$, nous obtenons

$$C^*\|\phi^n\|_0^2 \leq \|\phi^0\|_0^2 + \frac{C^*}{2} \max_{m=0,\ldots,N} \|\phi^m\|_0^2 + \frac{2}{C^*}\left(\Delta t \sum_{m=0}^{N-1} \|R^m\|_0\right)^2.$$

Puisque le membre de droite est indépendant de n,

$$C^* \max_{n=0,\ldots,N} \|\phi^n\|_0^2 \leq \|\phi^0\|_0^2 + \frac{C^*}{2} \max_{m=0,\ldots,N} \|\phi^m\|_0^2 + \frac{2}{C^*}\left(\Delta t \sum_{m=0}^{N-1} \|R^m\|_0\right)^2$$

ce qui est équivalent à

$$C^* \max_{n=0,\ldots,N} \|\phi^n\|_0^2 \leq 2\|\phi^0\|_0^2 + \frac{4}{C^*}\left(\Delta t \sum_{m=0}^{N-1} \|R^m\|_0\right)^2.$$

Mais, pour $a, b \geq 0$, $(a+b)^2 \geq a^2 + b^2$. Ainsi,

$$C^* \max_{n=0,\ldots,N} \|\phi^n\|_0^2 \leq \left(\sqrt{2}\|\phi^0\|_0 + \frac{2\Delta t}{\sqrt{C^*}} \sum_{m=0}^{N-1} \|R^m\|_0\right)^2.$$

Par conséquent,

$$\max_{n=0,\ldots,N} \|\phi^n\|_0 \leq \sqrt{\frac{2}{C^*}}\|\phi^0\|_0 + \frac{2\Delta t}{C^*} \sum_{m=0}^{N-1} \|R^m\|_0.$$

De plus, $\|\phi^0\|_0 \leq \|e^0\|_0 + \|\eta^0\|_0$ puisque $e^0 = \phi^0 + \eta^0$, donc :

$$\max_{n=0,\ldots,N} \|\phi^n\|_0 \leq \sqrt{\frac{2}{C^*}}\left(\|e^0\|_0 + \|\eta^0\|_0\right) + \frac{2\Delta t}{C^*} \sum_{m=0}^{N-1} \|R^m\|_0.$$

□

1.A.5 Preuve du lemme 1.3.11

Lemme 4.10 *Il existe une constante $C > 0$ indépendante de h, Δt et T telle que*

$$\|r^0\|_0 \leq C\left(\Delta t^{-1} h^{p+1} \|\partial_t u\|_{C\left(\overline{J}, H^{p+1}(\Omega)\right)} + \Delta t^3 \|\partial_t^5 u\|_{C\left(\overline{J}, L^2(\Omega)\right)}\right)$$

Démonstration. Nous rappelons que $\Delta t^2 r^0 = \phi^1 - \phi^0$. Dans la suite, nous voulons majorer le terme $\|\phi^1 - \phi^0\|_0$. Soit $v \in V_h$. Puisque $\left(u^0 - U^0, v\right) = 0$, par définition de U^0, on a

$$\begin{aligned}\left(\phi^1 - \phi^0, v\right) &= \left(w^1 - U^1, v\right) - \left(w^0 - U^0, v\right) \\ &= \left(w^1 - u^1, v\right) + \left(u^1 - U^1, v\right) - \left(w^0 - u^0, v\right) - \left(u^0 - U^0, v\right) \\ &= \left(\left(p_h - I\right)\left(u^1 - u^0\right), v\right) + \left(u^1 - U^1, v\right)\end{aligned} \tag{1.62}$$

En utilisant un développement de Taylor avec reste intégral, le fait que $\partial_t\left(p_h\left(u\right)\right) = p_h\left(\partial_t u\right)$ et le lemme 1.3.9, on obtient, en posant $t_1 = \Delta t$

$$\begin{aligned}\left|\left(\left(p_h - I\right)\left(u^1 - u^0\right), v\right)\right| &\leq \int_0^{t_1} \left|\left(\partial_t\left(p_h - I\right)u, v\right)\right| dt \\ &\leq \int_0^{t_1} \left|\left(\left(p_h - I\right)u_t, v\right)\right| dt \\ &\leq C\Delta t h^{p+1}\|u_t\|_{C\left(\overline{J};H^{p+1}(\Omega)\right)}\|v\|_0\end{aligned}$$

A présent, nous devons étudier le deuxième terme. Grâce au développement de Taylor avec reste intégral, nous avons :

$$u^1 = u_0 + \Delta t v_0 + \frac{\Delta t^2}{2}\partial_t^2 u^0 + \frac{\Delta t^3}{6}\partial_t^3 u^0 + \frac{\Delta t^4}{24}\partial_t^4 u^0 + \frac{1}{24}\int_0^{t_1}\left(\Delta t - s\right)^4 \partial_t^5 u\left(\cdot, s\right) ds.$$

Alors,

$$\begin{aligned}\left(u^1 - U^1, v\right) = \; & \left(u_0 - P_h u_0, v\right) + \Delta t\left(v_0 - P_h v_0, v\right) + \tfrac{\Delta t^2}{2}\left(\partial_t^2 u^0 - \widetilde{U_0}, v\right) + \tfrac{\Delta t^3}{6}\left(\partial_t^3 u^0 - \widetilde{V_0}, v\right) \\ & + \tfrac{\Delta t^4}{24}\left(\partial_t^4 u^0 - \widehat{U_0}, v\right) + \tfrac{1}{24}\int_0^{t_1}\left(\Delta t - s\right)^4 \partial_t^5 u\left(\cdot, s\right) ds\end{aligned}$$

D'après la définition de la projection P_h, $\forall v \in V_h$

$$\left(u_0 - P_h u_0, v\right) = 0 \quad \text{et} \quad \left(v_0 - P_h v_0, v\right) = 0$$

La consistance de la méthode et les définitions de $\widetilde{U_0}$, $\widetilde{V_0}$ et $\widehat{U_0}$ en (1.19) et (1.20) donnent immédiatement

$$\begin{cases} \left(\partial_t^2 u^0 - \widetilde{U_0}, v\right) = 0, \\ \left(\partial_t^3 u^0 - \widetilde{V_0}, v\right) = 0, \\ \left(\partial_t^4 u^0 - \widehat{U_0}, v\right) = 0. \end{cases}$$

Ainsi,

$$\begin{aligned}\left|\left(u^1 - U^1, v\right)\right| &\leq \frac{1}{24}\int_0^{t_1}\left(\Delta t - s\right)^4 \left|\left(\partial_t^5 u\left(\cdot, s\right), v\right)\right| ds \\ &\leq C\Delta t^5 \|\partial_t^5 u\|_{C\left(\overline{J};L^2(\Omega)\right)}\|v\|_0.\end{aligned}$$

Puisque $\phi^1 - \phi^0 \in V_h$, on peut choisir $v = \phi^1 - \phi^0$ comme fonction-test dans (1.62) :

$$\|\phi^1 - \phi^0\|_0 \leq C\left(\Delta t h^{p+1}\|u_t\|_{C\left(\overline{J};H^{p+1}(\Omega)\right)} + \Delta t^5 \|\partial_t^5 u\|_{C\left(\overline{J};L^2(\Omega)\right)}\right).$$

Par conséquent, on obtient :

$$\|r^0\|_0 \leq C\left(\Delta t^{-1} h^{p+1}\|u_t\|_{C\left(\overline{J};H^{p+1}(\Omega)\right)} + \Delta t^3 \|\partial_t^5 u\|_{C\left(\overline{J};L^2(\Omega)\right)}\right).$$

□

1.A.6 Preuve du lemme 1.3.12

Lemme 4.11 *Pour $1 \leq n \leq N-1$, il existe $C > 0$ indépendante de h, Δt et T telle que*

$$\|r^n\|_0 \leq C \left(\frac{h^{p+1}}{\Delta t} \int_{t_{n-1}}^{t_{n+1}} \|\partial_t^2 u(\cdot, s)\|_{p+1} ds + \Delta t^3 \int_{t_{n-1}}^{t_{n+1}} \|\partial_t^6 u(\cdot, s)\|_0 ds \right)$$

Démonstration. D'après l'inégalité triangulaire, on a :

$$\begin{aligned} \|r^n\|_0 &= \|\delta^2 w^n - \partial_t^2 u^n - \tfrac{\Delta t^2}{12} \partial_t^4 u^n\|_0 \\ &\leq \|\delta^2 (p_h - I) u^n\|_0 + \|\delta^2 u^n - \partial_t^2 u^n - \tfrac{\Delta t^2}{12} \partial_t^4 u^n\|_0 \end{aligned}$$

Pour majorer le premier terme, on utilise :

$$v(\cdot, t_{n+1}) - 2v(\cdot, t_n) + v(\cdot, t_{n-1}) = \Delta t \int_{t_{n-1}}^{t_{n+1}} \left(1 - \frac{|s-t|}{\Delta t}\right) \partial_t^2 v(\cdot, s)\, ds.$$

En utilisant le fait que $\partial_t^i (u_h) = \pi_h \left(\partial_t^i u\right)$ avec $i = 0, \ldots, 2$, puis le lemme 1.3.9, nous obtenons

$$\begin{aligned} \|\delta^2 (p_h - I) u^n\|_0 &\leq \tfrac{1}{\Delta t} \int_{t_{n-1}}^{t_{n+1}} \left(1 - \frac{|s-t_n|}{\Delta t}\right) \|\partial_t^2 (p_h - I) u\|_0 ds \\ &\leq C \tfrac{h^{p+1}}{\Delta t} \int_{t_{n-1}}^{t_{n+1}} \|\partial_t^2 u(\cdot, s)\|_{p+1} ds. \end{aligned}$$

Pour le deuxième terme, en utilisant un développement de Taylor avec reste intégral, nous avons :

$$\delta^2 u^n - \partial_t^2 u^n - \frac{\Delta t^2}{12} \partial_t^4 u^n = \frac{\Delta t^{-2}}{120} \left(\int_{t_n}^{t_{n+1}} (t_{n+1} - s)^5 \partial_t^6 u(\cdot, s)\, ds + \int_{t_n}^{t_{n-1}} (t_{n-1} - s)^5 \partial_t^6 u(\cdot, s)\, ds \right).$$

Et, $\forall s \in [t_n, t_{n+1}]$, $t_{n+1} - s \leq \Delta t$. Alors :

$$\int_{t_n}^{t_{n+1}} (t_{n+1} - s)^5 \partial_t^6 u(\cdot, s)\, ds \leq \Delta t^5 \int_{t_n}^{t_{n+1}} \partial_t^6 u(\cdot, s)\, ds.$$

De la même manière, $\forall s \in [t_{n-1}, t_n]$, $s - t_{n-1} \leq \Delta t$ et $(t_{n-1} - s)^5 = -(s - t_{n-1})^5$ donc :

$$\int_{t_n}^{t_{n-1}} (t_{n-1} - s)^5 \partial_t^6 u(\cdot, s)\, ds \leq \Delta t^5 \int_{t_{n-1}}^{t_n} \partial_t^6 u(\cdot, s)\, ds.$$

Par conséquent,

$$\|\delta^2 u^n - \partial_t^2 u^n - \frac{\Delta t^2}{12} \partial_t^4 u^n\|_0 \leq \frac{\Delta t^3}{120} \int_{t_{n-1}}^{t_{n+1}} \|\partial_t^6 u(\cdot, s)\|_0 ds.$$

Finalement,

$$\|r^n\|_0 \leq C \left(\frac{h^{p+1}}{\Delta t} \int_{t_{n-1}}^{t_{n+1}} \|\partial_t^2 u(\cdot, s)\|_{p+1} ds + \Delta t^3 \int_{t_{n-1}}^{t_{n+1}} \|\partial_t^6 u(\cdot, s)\|_0 ds \right).$$

□

1.B Eléments finis d'Hermite

1.B.1 Discrétisation

Notons, dans un premier temps, que le problème (1.9) est considéré ici en dimension un avec des conditions de bord de Dirichlet homogènes.
Les éléments finis d'Hermite n'étant pas adaptés aux milieux hétérogènes, on s'est donc limité au cas homogène. Ainsi, en multipliant (1.9) par une fonction-test $v \in H^2(\Omega)$, en intégrant sur Ω et en utilisant la formule de Green, on a

$$\int_\Omega \left(\frac{u^{n+1} - 2u^n + u^{n-1}}{\Delta t^2} \right) v \, dx = b_h(u^n, v) + \int_\Omega f v \, dx$$

avec

$$\begin{cases} b_h(u^n, v) = & b_{1h}(u^n, v) + \dfrac{\Delta t^2}{12} b_{2h}(u^n, v), \\ b_{1h}(u^n, v) = & -c^2 \displaystyle\int_\Omega \nabla u^n \cdot \nabla v \, dx + \int_{\partial\Omega} (\nabla u^n \cdot \boldsymbol{\nu}) \, v \, d\sigma, \\ b_{2h}(u^n, v) = & \displaystyle\int_\Omega \Delta u^n \Delta v \, dx + \int_{\partial\Omega} (\nabla(\Delta u^n) \cdot \boldsymbol{\nu}) \, v \, d\sigma, \end{cases}$$

où $\boldsymbol{\nu}$ désigne la normale unitaire extérieure à Ω et f le terme source.
Nous cherchons donc à approcher numériquement une solution qui est supposée être $H^2(\Omega)$ donc on ne peut pas utiliser des éléments finis de type P^3. En effet, en 1D, on a l'injection $H^2(\Omega) \hookrightarrow C^1(\Omega)$ donc on va vouloir choisir l'espace de discrétisation comme approchant au mieux l'espace $C^1(\Omega)$ ce qui n'est pas le cas des éléments finis P^3 puisqu'ils n'offrent qu'une approximation de $H^1(\Omega)$. Pour obtenir un tel espace, nous allons considérer les éléments finis d'Hermite qui sont construits de façon à avoir la valeur de la fonction ainsi que la valeur de sa dérivée aux extrémités de chaque maille. Ainsi, on obtient le caractère C^1 grâce aux raccords des dérivées aux différents noeuds du maillage.
Les fonctions de base correspondantes sont définies de la manière suivante :

$$\forall 1 \leq i, j \leq n \begin{cases} \varphi_i(x_j) = \delta_{ij}, & \varphi_i'(x_j) = 0, \\ \overline{\varphi_i}(x_j) = 0, & \overline{\varphi_i}'(x_j) = \delta_{ij}, \end{cases}$$

où n désigne le nombre de noeuds x_j du maillage, δ est le symbole de Kronecker et les fonctions $\overline{\varphi_i}$ correspondent aux dérivées des fonctions φ_i comme cela est illustré sur la Figure 1.8.

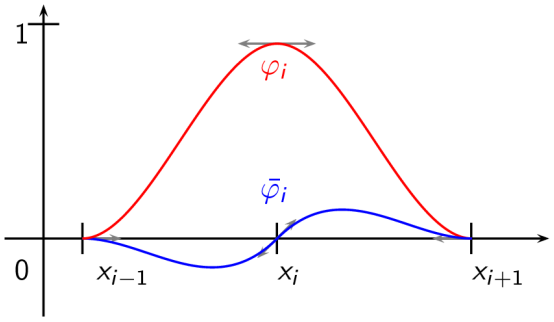

FIGURE 1.8 – Fonctions de base de Hermite 1D

En considérant ces fonctions de base et l'espace d'approximation associé, on peut reformuler le problème sous la forme des deux équations matricielles suivantes

$$A\left(\frac{U^{n+1}-2U^n+U^{n-1}}{\Delta t^2}\right)+B\left(\frac{\overline{U}^{n+1}-2\overline{U}^n+\overline{U}^{n-1}}{\Delta t^2}\right)+SU^n+R\overline{U}^n=0, \qquad (1.63)$$

$$B^T\left(\frac{U^{n+1}-2U^n+U^{n-1}}{\Delta t^2}\right)+C\left(\frac{\overline{U}^{n+1}-2\overline{U}^n+\overline{U}^{n-1}}{\Delta t^2}\right)+R^T+T\overline{U}^n=0, \qquad (1.64)$$

avec

$$\begin{cases} (A)_{i,j}=\displaystyle\int_\Omega \varphi_i\varphi_j\,dx, & (B)_{i,j}=\displaystyle\int_\Omega \varphi_i\overline{\varphi_j}\,dx, & (C)_{i,j}=\displaystyle\int_\Omega \overline{\varphi_i}\,\overline{\varphi_j}\,dx, \\ (S)_{i,j}=b_h\left(\varphi_i,\varphi_j\right) & (R)_{i,j}=b_h\left(\varphi_i,\overline{\varphi_j}\right), & (T)_{i,j}=b_h\left(\overline{\varphi_i},\overline{\varphi_j}\right). \end{cases} \qquad (1.65)$$

1.B.2 Stabilité

Nous avons également donné une condition nécessaire de stabilité pour le Δ^2-schéma en considérant une discrétisation en espace du type éléments finis d'Hermite. Ainsi, nous avons le théorème suivant :

Théorème 1.B.1. *Le Δ^2-schéma est L^2-stable sous la condition nécessaire suivante :*

$$c\frac{\Delta t}{h}\leq\frac{1}{\sqrt{5}}.$$

Démonstration. Tout d'abord, remarquons que si on fixe l'indice de ligne j dans toutes les matrices de (1.65), que l'on notera indifféremment X, alors les seuls termes non nuls sont $X_{j-1,j}$, $X_{j,j}$ et $X_{j,j+1}$. Ainsi l'équation (1.63) peut se reformuler, en notant $\delta_j U=U_j^{n+1}-2U_j^n+U_j^{n-1}$,

$$\sum_{i=1}^{3}\left(a_i\delta_{j-2+i}U+b_i\delta_{j-2+i}\overline{U}+s_iU^n_{j-2+i}+r_i\overline{U}^n_{j-2+i}\right)=0.$$

De même l'équation (1.64) se réécrit

$$\sum_{i=1}^{3}\left(b_i\delta_{j-2+i}U+c_i\delta_{j-2+i}\overline{U}+r_iU^n_{j-2+i}+t_i\overline{U}^n_{j-2+i}\right)=0.$$

où, pour tout $i\in\{1,2,3\}$,

$$\begin{cases} a_i=\dfrac{1}{\Delta t^2}\displaystyle\int_\Omega \varphi_{j-2+i}\varphi_j\,dx, & b_i=\dfrac{1}{\Delta t^2}\displaystyle\int_\Omega \overline{\varphi_{j-2+i}}\,\varphi_j\,dx, & c_i=\dfrac{1}{\Delta t^2}\displaystyle\int_\Omega \overline{\varphi_{j-2+i}}\,\overline{\varphi_j}\,dx, \\ s_i=b_h\left(\varphi_{j-2+i},\varphi_j\right), & r_i=b_h\left(\overline{\varphi_{j-2+i}},\varphi_j\right), & t_i=b_h\left(\overline{\varphi_{j-2+i}},\overline{\varphi_j}\right). \end{cases}$$

On précise, sans détailler les calculs, les valeurs de ces différents coefficients

$$
\begin{array}{lll}
a_1 = \dfrac{9h}{70\Delta t^2}, & a_2 = \dfrac{26h}{35\Delta t^2}, & a_3 = \dfrac{9h}{70\Delta t^2}, \\
b_1 = \dfrac{13h^2}{420\Delta t^2}, & b_2 = 0, & b_3 = \dfrac{-13h^2}{420\Delta t^2}, \\
c_1 = \dfrac{-h^3}{140\Delta t^2}, & b_2 = \dfrac{2h^3}{105\Delta t^2}, & c_3 = \dfrac{-h^3}{140\Delta t^2}, \\
r_1 = -\dfrac{c^2}{10} + \dfrac{c^4\Delta t^2}{2h^2}, & r_2 = 0, & r_3 = \dfrac{c^2}{10} - \dfrac{c^4\Delta t^2}{2h^2}, \\
s_1 = -\dfrac{6c^2}{5h} + \dfrac{c^4\Delta t^2}{h^3}, & s_2 = \dfrac{12c^2}{5h} - \dfrac{2c^4\Delta t^2}{h^3}, & s_3 = -\dfrac{6c^2}{5h} + \dfrac{c^4\Delta t^2}{h^3}, \\
t_1 = -\dfrac{hc^2}{30} - \dfrac{c^4\Delta t^2}{6h}, & t_2 = \dfrac{4hc^2}{15} - \dfrac{2c^4\Delta t^2}{3h}, & s_3 = -\dfrac{hc^2}{30} - \dfrac{c^4\Delta t^2}{6h}.
\end{array}
$$

Ainsi en appliquant une transformée de Fourier discrète sur les équations (1.63) et (1.64), on obtient le système

$$
\begin{cases}
A\hat{U}^{n+1} + B\hat{\overline{U}}^{n+1} + C\hat{U}^{n} + D\hat{\overline{U}}^{n} + A\hat{U}^{n-1} + B\hat{\overline{U}}^{n-1} = 0, \\
\overline{B}\hat{U}^{n+1} + E\hat{\overline{U}}^{n+1} + \overline{D}\hat{U}^{n} + F\hat{\overline{U}}^{n} + \overline{B}\hat{U}^{n-1} + E\hat{\overline{U}}^{n-1} = 0.
\end{cases}
\tag{1.66}
$$

avec

$$
\begin{cases}
A = \dfrac{h}{\Delta t^2}\left(1 - \dfrac{18}{35}\sin^2(\beta)\right), \\
B = -\dfrac{13h^2}{105\Delta t^2} i \sin^2(\beta), \\
C = \dfrac{2h^2}{\Delta t^2}\left(\dfrac{18}{35}\sin^2(\beta) - 1\right) + 4\left(\dfrac{6c^2}{5h} - \dfrac{c^4}{h^3}\Delta t^2\right)\sin^2(\beta), \\
D = \left(\dfrac{13h^2}{105\Delta t^2} + \dfrac{c^2}{5} - \dfrac{c^4\Delta t^2}{h^2}\right) i \sin^2(\beta), \\
E = \dfrac{h^3}{\Delta t^2}\left(\dfrac{1}{210} + \dfrac{\sin^2(\beta)}{35}\right), \\
F = \left(\dfrac{hc^2}{5} - \dfrac{c^4\Delta t^2}{h^2} - \dfrac{h^3}{105\Delta t^2}\right) + 2\left(\dfrac{hc^2}{15} + \dfrac{c^4\Delta t^2}{3h} - \dfrac{h^3}{35\Delta t^2}\right)
\end{cases}
$$

et $\beta = \dfrac{hk}{2}$ où k est la variable duale de la variable spatiale.
On peut reformuler le système (1.66) sous la forme matricielle

$$
\begin{pmatrix} A & B \\ \overline{B} & E \end{pmatrix}
\begin{pmatrix} \hat{U}^{n+1} \\ \hat{\overline{U}}^{n+1} \end{pmatrix}
= -
\begin{pmatrix} A & B & C & D \\ \overline{B} & E & \overline{D} & F \end{pmatrix}
\begin{pmatrix} \hat{U}^{n-1} \\ \hat{\overline{U}}^{n-1} \\ \hat{U}^{n} \\ \hat{\overline{U}}^{n} \end{pmatrix}.
\tag{1.67}
$$

En notant $W^n = \left(\hat{U}^{n-1}, \hat{\overline{U}}^{n-1}, \hat{U}^n, \hat{\overline{U}}^n\right)^T$, le système (1.67) devient

$$MW^{n+1} = -NW^n \tag{1.68}$$

où

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & A & B \\ 0 & 0 & \overline{B} & E \end{pmatrix} \qquad \text{et} \qquad N = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ A & B & C & D \\ \overline{B} & E & \overline{D} & F \end{pmatrix}.$$

Afin de justifier que M est inversible, il est utile de noter que

$$\det(M) = \left(\frac{1}{210} + \frac{119}{11025}\sin^2(\beta) + \frac{\sin^4(\beta)}{1575}\right)\frac{h^4}{\Delta t^2}$$

est clairement un terme positif.
En utilisant la notation $H(k) = M^{-1}N$, on arrive à la conclusion que le schéma est stable si et seulement si $\|H(k)\| \leq 1$ pour tout $k \in \mathbb{R}$, c'est-à-dire si et seulement si

$$\rho(H(k)) \leq 1, \ \forall k \in \mathbb{R}$$

où $\rho(H(k))$ désigne le rayon spectral de la matrice $H(k)$. En remarquant que l'inverse de M est de la forme

$$M^{-1} = \frac{1}{\det(M)} \begin{pmatrix} \det(M) & 0 & 0 & 0 \\ 0 & \det(M) & 0 & 0 \\ 0 & 0 & E & -B \\ 0 & 0 & -\overline{B} & A \end{pmatrix}$$

on a

$$H = -\begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & \dfrac{CE - B\overline{D}}{\det(M)} & \dfrac{DE - BF}{\det(M)} \\ 0 & 1 & \dfrac{A\overline{D} - \overline{B}C}{\det(M)} & \dfrac{AF - \overline{B}D}{\det(M)} \end{pmatrix}.$$

A présent, introduisons la notation suivante

$$\begin{pmatrix} \dfrac{CE - B\overline{D}}{\det(M)} & \dfrac{DE - BF}{\det(M)} \\ \dfrac{A\overline{D} - \overline{B}C}{\det(M)} & \dfrac{AF - \overline{B}D}{\det(M)} \end{pmatrix} := \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

et intéressons-nous au calcul des valeurs propres de H. On explicite le polynôme caractéristique associé à H :

$$p(\lambda) = \lambda^4 - (a+d)\lambda^3 + (ad - bc + 2)\lambda^2 - (a+d)\lambda + 1.$$

En posant $\Delta t = \dfrac{\sqrt{\alpha}h}{c}$ et en faisant l'hypothèse que $\beta = m\pi$, $m \in \mathbb{Z}$, c'est-à-dire le cas où $k = \dfrac{2m\pi}{h}$, $m \in \mathbb{Z}$, p se reformule

$$p(\lambda) = (\lambda+1)^2\left(\lambda^2 + 2\left(1 - 21\alpha + 105\alpha^2\right)\lambda + 1\right).$$

On en conclut que -1 est valeur propre double de H, on regarde alors les racines du second facteur. Le calcul du discriminant nous donne

$$\Delta = 84\alpha\,(5\alpha - 1)\left(105\alpha^2 - 21\alpha + 2\right).$$

On remarque que lorsque α décrit $\mathbb{R}^+$, $105\alpha^2 - 21\alpha + 2$ est toujours positif donc Δ est du signe de $5\alpha - 1$. On distingue donc trois cas :

- Si $5\alpha - 1 > 0$, p admet deux racines réelles distinctes λ_1 et λ_2 telles que $\lambda_1\lambda_2 = 1$. Ceci implique que l'une des deux est nécessairement plus grande que 1 et dans ce cas-là, le schéma est instable.
- Si $5\alpha - 1 = 0$, p admet $\frac{1}{5}$ pour racine double qui est bien inférieur à 1.
- Si $5\alpha - 1 < 0$, p admet deux racines complexes conjuguées dont le produit des racines est en module égal à 1 et qui ont même module. On en déduit donc que leur module est égal à 1.

Par conséquent, une condition nécessaire de stabilité est

$$c\frac{\Delta t}{h} \leq \frac{1}{\sqrt{5}}$$

ce qui achève la preuve du théorème. □

Chapitre 2

Analyse de stabilité pour la méthode IPDG appliquée à l'équation des ondes

Comme nous l'avons déjà souligné, les méthodes de Galerkine discontinues, utilisées pour discrétiser l'équation des ondes, mènent à des matrices de masse diagonales par bloc sans l'aide de formule de quadrature particulière. De plus, elles peuvent être utilisées avec n'importe quel type de maillage et permettent de prendre en compte facilement les variations des paramètres physiques à l'intérieur de chaque cellule du maillage (au moins des variations polynomiales). Les méthodes DG sont aussi naturellement adaptées à la parallélisation puisque que toutes les intégrales de volume sont calculées localement et les communications entre les cellules sont assurées par des intégrales surfaciques sur chaque face des éléments. Dans [5], les auteurs présentent une revue complète des différentes approximations de Galerkine discontinues pour l'opérateur de Laplace. Ils montrent en particulier que la méthode IPDG, également connue sous le nom de méthode avec Pénalité Interne Symétrique (SIP) [8], est l'une des plus appropriées puisqu'elle est stable et consistante, ce qui garantit un ordre optimal de convergence du schéma. Cela explique pourquoi cette méthode a été utilisée avec un vif intérêt pour résoudre l'équation de Helmholtz [2, 9] et l'équation des ondes [33, 34, 2].
Néanmoins, malgré ces propriétés intéressantes, la méthode IPDG pose deux difficultés. La première est la détermination du paramètre de pénalisation γ, introduit au chapitre précédent, qui pénalise les discontinuités de la solution au travers des faces internes du maillage. Ce paramètre s'exprime sous la forme $\frac{\alpha}{\xi_F}$ où ξ_F est un coefficient dépendant de la face considérée. En théorie, il devrait être tel que α soit uniquement dépendant du degré polynomial des fonctions de base et indépendant du maillage. En pratique, le choix de ξ_F reste une question ouverte. Le choix le plus classique (cf. [33]) est de considérer $\xi_F = \min\left(\rho_{\text{circ}}\left(K^+\right), \rho_{\text{circ}}\left(K^-\right)\right)$ où $\rho_{\text{circ}}\left(K\right)$ désigne le rayon du cercle circonscrit à l'élément K et K^+ et K^- sont les deux éléments partageant la face F. Cependant, ce choix nécessite soit de réajuster α pour chaque maillage, soit de choisir α très élevé et donc de sur-pénaliser la forme bilinéaire. Récemment, Shahbazi (cf. [49]) a proposé de plutôt considérer le cercle inscrit. Cependant, aucune étude n'a encore analysé l'intérêt d'un tel choix par rapport au choix classique. La question du choix de ce paramètre est cruciale puisqu'une valeur trop petite mène à des instabilités tandis qu'une valeur trop grande pénalise la condition CFL. Dans [2], les auteurs proposent une valeur minimale du paramètre de pénalisation, en fontion du degré polynomial p des fonctions de base et de la taille des éléments. Il s'agit d'une conjecture dont on trouve, dans [2], une preuve jusqu'à $p = 3$ mais l'extension de ce résultat à $p > 3$ et à des maillages non structurés reste à faire. La seconde difficulté est la détermination de la condition CFL. Il est connu que cette condition décroit quand le coefficient de pénalisation aug-

mente mais aucune étude analytique n'a encore été proposée. Dans ce chapitre, on se propose a) de prouver la conjecture de Ainsworth, Monk et Muniz jusqu'à $p = 5$; b) d'établir une méthodologie pour la prouver pour un p donné ; c) de fournir une formule analytique liant la condition CFL au coefficient de pénalisation. Les résultats obtenus, ainsi que la méthodologie proposée, nous serviront ensuite pour l'étude de la stabilité du Δ^2-schéma. Nous restreignons notre étude au cas de maillages structurés composés de segments (en 1D), de carrés (en 2D) ou de cubes (en 3D). Aux sections 2.1 et 2.2, nous établissons deux théorèmes en 1D, fournissant des conditions nécessaires de stabilité par rapport au coefficient de pénalisation et au pas de temps. La section 2.3 est consacrée à l'extension de ces résultats en dimension 2 et 3 sur des maillages structurés composés de carrés ou de cubes. La preuve en dimension 2 étant exactement similaire à celle en dimension 3, nous ne la présentons pas ici. Enfin, nous présenterons des résultats numériques en section 2.4 qui illustrent la validité de nos deux théorèmes.

2.1 Analyse de stabilité

Dans cette section, nous proposons des conditions nécessaires sur γ et Δt afin d'assurer la stabilité-L^2 du schéma

$$\frac{U^{n+1} - 2U^n + U^{n-1}}{\Delta t^2} = -M^{-1}KU^n + M^{-1}F^n. \tag{2.1}$$

Nous obtenons une condition nécessaire de stabilité exprimant Δt en fonction de γ et h. Numériquement, nous montrons que cette condition n'est pas suffisante. Ceci justifie le théorème 2.1.3 dans lequel nous montrons qu'il existe une façon de rendre cette condition quasi-suffisante en un sens que nous précisons plus tard. Nous supposons ici que le domaine Ω est infini $\left(\Omega = \mathbb{R}^d\right)$ et uniformément maillé par des segments (si $d = 1$), des carrés (si $d = 2$) ou des cubes (si $d = 3$). La longueur des côtés de ces éléments est notée indifféremment h.

Une condition nécessaire de stabilité est donnée par le théorème suivant :

Théorème 2.1.1. *Supposons $p \leq 5$ et posons $\alpha_{0,p} = \dfrac{p\,(p+1)}{2}$. Pour que le schéma (2.1) soit L^2-stable, il faut que*

$$\gamma \geq \frac{\alpha_{0,p}}{h} \tag{2.2}$$

et, si on écrit $\gamma = {}^{\alpha}/_{h}$,

$$\sqrt{d}\frac{c\Delta t}{h} \leq \begin{cases} C_{1,p} & \text{si } \alpha_{0,p} \leq \alpha \leq \alpha_{1,p} \\ C_{2,p}\,(\alpha) & \text{si } \alpha > \alpha_{1,p}. \end{cases} \tag{2.3}$$

où $\alpha_{1,p}$, $C_{1,p}$ et $C_{2,p}(\alpha)$ sont définis par :

p	$\alpha_{1,p}$	$C_{1,p}$	$C_{2,p}(\alpha)$
1	2	0.577	$\dfrac{1}{\sqrt{3(\alpha-1)}}$
2	5.4	0.258	$\sqrt{\dfrac{2}{-15+6\alpha+(405-240\alpha+36\alpha^2)^{1/2}}}$
3	9.65	0.153	$\sqrt{\dfrac{2}{-45+10\alpha+(4545-1320\alpha+100\alpha^2)^{1/2}}}$
4	14.7	0.103	$\sqrt{\dfrac{1}{2\sqrt{5}g_{4,1}(\alpha)\,g_{4,2}(\alpha)+5\alpha-35}}$
5	20.8	0.074	$\sqrt{\dfrac{1}{2\sqrt{7}g_{5,1}(\alpha)\,g_{5,2}(\alpha)+7(\alpha-10)}}$

avec, pour $p=4$,

$$\begin{cases} g_{4,1}(\alpha) = & \left(518-98\alpha+5\alpha^2\right)^{\frac{1}{2}}, \\ g_{4,2}(\alpha) = & \cos\left(\dfrac{1}{3}\arccos\left(\dfrac{1}{10}g_{4,3}(\alpha)\dfrac{\sqrt{5}}{g_{4,1}^3(\alpha)}\right)\right), \\ g_{4,3}(\alpha) = & -47705+14574\alpha-1470\alpha^2+50\alpha^3. \end{cases}$$

et, pour $p=5$,

$$\begin{cases} g_{5,1}(\alpha) = & \left(1555-200\alpha+7\alpha^2\right)^{\frac{1}{2}}, \\ g_{5,2}(\alpha) = & \cos\left(\dfrac{1}{3}\arccos\left(\dfrac{1}{14}g_{5,3}(\alpha)\dfrac{\sqrt{7}}{g_{5,1}^3(\alpha)}\right)\right), \\ g_{5,3}(\alpha) = & -299825+61440\alpha-4200\alpha^2+98\alpha^3. \end{cases}$$

Dans [2], les auteurs ont prouvé (2.2) pour $p=0,\ldots,3$ et conjecturé cette relation pour tout $p>3$. Le théorème (2.1.1) étend donc sa validité jusqu'à $p=5$.
La condition (2.2) ne permet pas de déterminer le choix le plus judicieux de ξ_F. En effet, dans le cas de mailles carrées ou cubiques, les rayons des cercles (ou sphères) inscrits et circonscrits sont proportionnels et la condition (2.2) peut s'écrire indifféremment sous la forme

$$\gamma \geq \frac{2\alpha_{0,p}}{\rho_{\text{ins}}}, \quad \text{ou} \quad \gamma \geq \frac{2\alpha_{0,p}\sqrt{d}}{\rho_{\text{circ}}}.$$

Le cas des maillages rectangulaires ou parallélépipèdiques, étudié à la section 2.3, nous donnera une meilleure indication du choix le plus approprié.

Remarque 2.1.2.

- *Dans [34], il a été montré que la condition de stabilité liant Δt à α se comporte comme $C/\sqrt{\alpha}$ pour α suffisamment grand, où plus précisément, $C=\sqrt{\dfrac{2}{(p+1)(p+2)}}$. Le théorème*

2.1.1 dit que pour α suffisamment grand, $\dfrac{\Delta t}{h} \leq C_{2,p}(\alpha)$ et $C_{2,p}(\alpha)$ se comporte comme $C/\sqrt{\alpha}$.

- *La condition de stabilité ne dépend pas de α pour $\alpha_{0,p} \leq \alpha \leq \alpha_{1,p}$. Il n'est donc pas nécessaire de choisir α trop proche de $\alpha_{0,p}$ pour améliorer la condition CFL.*

Le théorème suivant nous donne une condition quasi-suffisante de stabilité en un sens que nous préciserons plus loin dans ce chapitre.

Théorème 2.1.3. *Soit $V_{p,\alpha} = \left\{\lambda \in \mathbb{R}^+ : Q_{p,\alpha}(\lambda) = 0 \text{ et } |\tilde{Q}_{p,\alpha}(\lambda)| \leq 1\right\}$ où $Q_{p,\alpha}(\lambda)$ est un polynôme de degré $2p$, admettant au moins une racine réelle positive, et $\tilde{Q}_{p,\alpha}(\lambda)$ est une fraction rationnelle définis en annexe 2.B pour tout $p \leq 5$. Alors, si $p \leq 5$, pour que le schéma (2.1) soit L^2-stable, il faut que les conditions (2.2) et (2.3) soient satisfaites et que*

$$\begin{cases} V_{p,\alpha} = \varnothing \\ \\ \textit{ou} \\ \\ \sqrt{d}\dfrac{c\Delta t}{h} \leq C_{3,p}(\alpha) = \dfrac{1}{2\sqrt{\max V_{p,\alpha}}}. \end{cases} \tag{2.4}$$

Remarque 2.1.4.

- *Ce théorème ne donne pas une condition explicite. Néanmoins, elle peut être calculée numériquement via l'algorithme suivant :*
  1. *Calculer toutes les racines de $Q_{p,\alpha}$,*
  2. *Sélectionner toutes les racines réelles positives telles que $|\tilde{Q}_{p,\alpha}(\lambda)| \leq 1$,*
  3. *Choisir le maximum de ces racines.*
- *Les résultats numériques de la section 2.4 montrent que ce théorème fournit en pratique des conditions nécessaires et suffisantes.*
- *L'étude numérique de la condition (2.4) que nous présentons à la section 2.4 montre que l'ensemble $V_{p,\alpha}$ est en fait vide à l'exception du cas où α appartient à un petit voisinage de $\alpha_{1,p}$. Cela signifie que le théorème 2.1.1 fournit une condition nécessaire et suffisante de stabilité quand α n'est pas dans ce voisinage. De plus, les remarques 2.1.2 sont toujours valables.*

Malheureusement, nous n'avons pas pu établir ce théorème pour n'importe quel choix de p et nous nous sommes restreints aux cas $p \leq 5$, la technique étant similaire pour des degrés polynomiaux plus élevés. Les preuves dans le cas 1D sont données en section 2.2 alors que leur extension à la dimension $d = 3$ fait l'objet de la section 2.3. Comme nous l'avons déjà souligné, la preuve du cas $d = 2$ est exactement similaire à celle du cas $d = 3$ ce qui explique que nous ne la présentons pas ici.

2.2 Etude du cas mono-dimensionnel

Cette section contient les preuves des théorèmes 2.1.1 et 2.1.3 dans le cas mono-dimensionnel qui se décomposent en trois étapes. La première est une analyse de Fourier présentée en section 2.2.1 ; la seconde étape est consacrée à la preuve de la condition (2.2) qui est détaillée dans la section 2.2.2 ; la dernière étape concerne la preuve des conditions (2.3) et (2.4) en section 2.2.3. Les preuves sont détaillées pour $p = 3$ et sont adaptables sans difficulté aux cas $p = 1, 2, 4$ et 5.

Nous avons choisi de présenter ce cas, car pour $p = 1$ ou 2 certaines difficultés sont masquées et pour $p \geq 4$ les expressions des polynômes (et *a fortiori* de leurs racines) sont très lourdes et nuisent à la compréhension du mécanisme de preuve.
Ici, nous supposons que le domaine est $\Omega = \mathbb{R}$ et qu'il est maillé par des segments de longueur h. Nous considérons une vitesse $c^2 = {}^{\mu}/_{\rho} = 1$ mais nous pouvons étendre facilement la preuve à d'autres vitesses en posant $\Delta t' = {}^{\Delta t}/_{c}$. Nous considérons le schéma (2.1) sans terme source c'est-à-dire

$$M\frac{U^{n+1} - 2U^n + U^{n-1}}{\Delta t^2} + KU^n = 0. \tag{2.5}$$

Considérons l'équation sur un élément J du maillage. Nous avons, $\forall J \in \mathcal{T}_h$,

$$M_{1,p}\frac{U_J^{n+1} - 2U_J^n + U_J^{n-1}}{\Delta t^2} + \left(K_{1,p}^W\right)^T U_{J-1}^n + K_{1,p}U_J^n + K_{1,p}^W U_{J+1}^n = 0 \tag{2.6}$$

où U_J correspond au vecteur d'inconnues U restreint à l'élément J et $M_{1,p}$, $K_{1,p}$ et $K_{1,p}^W$ sont respectivement les matrices de masse et de raideur en dimension 1 obtenues en considérant des polynômes de degré p. Plus précisément, on a :

$$\begin{aligned}
M_{1,p}(i,j) = \;& h\int_{[0,1]} \hat{\varphi}_i(\hat{x})\,\hat{\varphi}_j(\hat{x})\,d\hat{x}, \\
K_{1,p}(i,j) = \;& \frac{1}{h}\int_{[0,1]} \frac{\partial\hat{\varphi}_i}{\partial\hat{x}}(\hat{x})\frac{\partial\hat{\varphi}_j}{\partial\hat{x}}(\hat{x})\,d\hat{x} + \frac{1}{2h}\hat{\varphi}_i(1)\frac{\partial\hat{\varphi}_j}{\partial\hat{x}}(1) + \frac{1}{2h}\hat{\varphi}_j(1)\frac{\partial\hat{\varphi}_i}{\partial\hat{x}}(1) \\
& + \gamma\hat{\varphi}_i(1)\,\hat{\varphi}_j(1) - \frac{1}{2h}\hat{\varphi}_i(0)\frac{\partial\hat{\varphi}_j}{\partial\hat{x}}(0) - \frac{1}{2h}\hat{\varphi}_j(0)\frac{\partial\hat{\varphi}_i}{\partial\hat{x}}(0) \\
& + \gamma\hat{\varphi}_i(0)\,\hat{\varphi}_j(0), \\
K_{1,p}^W(i,j) = \;& -\frac{1}{2h}\hat{\varphi}_i(1)\frac{\partial\hat{\varphi}_j}{\partial\hat{x}}(0) + \frac{1}{2h}\hat{\varphi}_j(0)\frac{\partial\hat{\varphi}_i}{\partial\hat{x}}(1) - \gamma\hat{\varphi}_i(1)\,\hat{\varphi}_j(0),
\end{aligned} \tag{2.7}$$

où $\{\hat{\varphi}_i\}_{i=1,\ldots,p+1}$ sont les fonctions de base discontinues de Lagrange sur l'élément de référence $[0,1]$.

2.2.1 Analyse de Fourier du schéma IPDG en 1D

Pour étudier la stabilité du schéma IPDG, nous introduisons la transformée de Fourier discrète

$$\begin{aligned}
\mathcal{F}_h \;:\; L_h^2 \to \;& L^2(K_h) \\
U \to \;& \tilde{U} = \mathcal{F}_h(U)(k) = \frac{h}{2\pi}\sum_{J\in\mathbb{Z}} U_J e^{-ikJh}
\end{aligned}$$

avec $K_h = \left[-\frac{\pi}{h}, \frac{\pi}{h}\right]$ et $L_h^2 = \left\{U = (U_J)_{J\in\mathbb{Z}}\,,\, \sum_{J\in\mathbb{Z}} \|U_J\|^2 < +\infty\right\}$.
A présent, en appliquant la transformée de Fourier discrète à (2.6), nous obtenons, $\forall\beta \in [-\pi, \pi]$

$$M_{1,p}\frac{\tilde{U}_J^{n+1}(\beta) - 2\tilde{U}_J^n(\beta) + \tilde{U}_J^{n-1}(\beta)}{\Delta t^2} + K_\beta\tilde{U}_J^n(\beta) = 0 \tag{2.8}$$

où $\beta = hk$ et $K_\beta = \left(K_{1,p}^W\right)^T e^{-i\beta} + K_{1,p} + K_{1,p}^W e^{i\beta}$.
La stabilité L^2 de (2.8), pour tout $\beta \in [-\pi, \pi]$, est équivalente à la stabilité L^2 de (2.5), grâce aux égalités de Parseval.

Puisque $M_{1,p}$ est une matrice définie positive et que K_β est hermitienne, toutes les valeurs propres de $N_\beta = M_{1,p}^{-1} K_\beta$ sont réelles. Une analyse de stabilité classique montre alors que (2.8) est stable si et seulement si

$$0 \leq \lambda \leq \frac{4}{\Delta t^2}$$

pour tout $\lambda \in \Lambda(\beta)$ où $\Lambda(\beta)$ représente l'ensemble des valeurs propres de N_β. Une condition nécessaire et suffisante de stabilité pour (2.5) est donc

$$\lambda_{\text{min}} \geq 0 \qquad \text{et} \qquad \Delta t \leq \frac{2}{\sqrt{\lambda_{\text{max}}}}$$

avec $\lambda_{\text{min}} = \min\limits_{\beta \in [-\pi,\pi]} [\min(\Lambda(\beta))]$ et $\lambda_{\text{max}} = \max\limits_{\beta \in [-\pi,\pi]} [\max(\Lambda(\beta))]$.
Pour étudier les valeurs propres de N_β, nous allons calculer son polynôme caractéristique q_α. En posant $\alpha = h\gamma$, il est de la forme

$$q_\alpha(\beta, \lambda) = (-1)^{p+1} \lambda^{p+1} + \sum_{i=0}^{p} c_i(\alpha, \beta) \lambda^i. \tag{2.9}$$

Les coefficients $c_i(\alpha, \beta)$ peuvent être calculés par un logiciel de calcul formel tel que Maple. Nous les présentons dans l'annexe 2.A pour $1 \leq p \leq 5$. Par exemple, pour $p = 3$, nous avons

$$\begin{cases} c_3(\alpha,\beta) = & \frac{8}{h^2}\left((15-\alpha)\cos(\beta) - 4\alpha\right) \\ c_2(\alpha,\beta) = & \frac{240}{h^4}\left(\cos^2(\beta) - (23+\alpha)\cos(\beta) + (18\alpha - 65)\right) \\ c_1(\alpha,\beta) = & \frac{2880}{h^6}\left(4\cos^2(\beta) + (65-3\alpha)\cos(\beta) + (141-32\alpha)\right) \\ c_0(\alpha,\beta) = & \frac{100800}{h^8}\left(3\cos^2(\beta) + 2(3-\alpha)\cos(\beta) + (2\alpha - 9)\right). \end{cases}$$

Comme toutes les valeurs propres de N_β sont réelles, il est évident que toutes les racines de $q_\alpha(\beta, \cdot)$ sont réelles.
Dans la section 2.2.2, nous montrons que la condition $\lambda_{\min} \geq 0$ est équivalente à (2.2) et dans la section 2.2.3, nous montrons que la condition $\lambda_{\max} \leq \frac{4}{\Delta t^2}$ implique (2.3) et (2.4).

2.2.2 Etude de la condition $\lambda_{\text{min}} \geq 0$

Pour montrer l'équivalence entre (2.2) et la condition $\lambda_{\text{min}} \geq 0$, nous cherchons une condition nécessaire et suffisante portant sur α pour que les racines du polynôme caractéristique q_α soient positives quel que soit β. Nous utiliserons le lemme suivant.

Lemme 2.2.1. *Soit P un polynôme de degré n avec n racines réelles tel que $P(Y) = \sum\limits_{i=0}^{n} c_i Y^i$. Toutes les racines de P sont positives si et seulement si*

$$(-1)^i c_i \geq 0.$$

Démonstration. Ce lemme est un corollaire des règles de Descartes. Nous en donnons cependant la preuve en annexe 2.D. □

Ainsi, nous devons trouver une condition sur α telle que, $\forall i \in \{0, \ldots, p\}$, $\forall \beta \in [-\pi, \pi]$,

$$(-1)^i c_i(\alpha, \beta) \geq 0.$$

Nous allons étudier chacun de ces coefficients, en commençant par le terme de plus haut degré. Nous ne présentons ici que le cas $p = 3$.

- Etudions tout d'abord la condition sur c_3. On a, $\forall \beta \in [-\pi, \pi]$,

$$-c_3(\alpha, \beta) \geq 0 \quad \Leftrightarrow \quad (\alpha - 15)\cos(\beta) + 4\alpha \geq 0.$$

  Il est clair que cette condition est satisfaite pour tout β si et seulement si

$$|\alpha - 15| + 4\alpha \geq 0 \tag{2.10}$$

  ce qui implique que $\alpha \geq 3$. Par conséquent, $-c_3(\alpha, \beta) \geq 0$, quel que soit le choix de β, si et seulement si

$$\alpha \geq 3. \tag{2.11}$$

- Considérons à présent la condition sur c_2.
  En posant $X = \cos(\beta)$, cette condition est équivalente à

$$f_\alpha(X) := X^2 - (23 + \alpha)X + (18\alpha - 65) \geq 0, \ \forall X \in [-1, 1]. \tag{2.12}$$

  Cette équation du second ordre admet deux racines :

$$\begin{cases} X_1 = \dfrac{1}{2}\left(23 + \alpha - \left((\alpha - 13)^2 + 620\right)^{1/2}\right), \\ X_2 = \dfrac{1}{2}\left(23 + \alpha + \left((\alpha - 13)^2 + 620\right)^{1/2}\right). \end{cases}$$

  Nous savons que $f_\alpha(X)$ est un polynôme du second degré et que son coefficient de plus haut degré est positif. Donc, f_α est positif $\forall X \in [-1; 1]$ si une des conditions suivantes est vérifiée :

  1. $X_1 = X_2$,
  2. $[-1, 1] \subset]-\infty, X_1]$,
  3. $[-1, 1] \subset]X_2, +\infty]$.

  Puisque $(\alpha - 13)^2 + 620 > 0$, $X_1 < X_2$ et 1. est impossible.
  Le cas $X_2 \leq -1$ est aussi impossible car $X_2 \geq 0$ quand $\alpha \geq 0$ donc nous avons juste à considérer le cas $X_1 \geq 1$, qui mène à l'inégalité

$$23 + \alpha - \left((\alpha - 13)^2 + 620\right)^{1/2} \geq 2,$$

  qui est équivalente à

$$\alpha \geq \frac{87}{17}. \tag{2.13}$$

  Finalement, $c_2(\alpha, \beta) \geq 0$, quel que soit le choix de β, si et seulement si (2.13) est vérifiée.

- Maintenant, étudions le signe de $c_1(\alpha,\beta)$.
  En effectuant le changement de variable $X=\cos(\beta)$, la condition $-c_1(\alpha,\beta)\geq 0, \forall\beta\in[-\pi,\pi]$ est équivalente à

$$f_\alpha(X) := -4X^2+(3\alpha-65)X+(32\alpha-141)\geq 0,\ \forall X\in[-1,1].$$

  Le polynôme f_α admet les deux racines suivantes :

$$\begin{cases} X_1=\dfrac{-1}{8}\left(65-3\alpha+\left(\left(3\alpha+\dfrac{61}{3}\right)^2+\dfrac{14000}{9}\right)^{1/2}\right),\\ X_2=\dfrac{-1}{8}\left(65-3\alpha-\left(\left(3\alpha+\dfrac{61}{3}\right)^2+\dfrac{14000}{9}\right)^{1/2}\right). \end{cases}$$

  Puisque le coefficient de tête du polynôme f_α est négatif, d'une manière analogue au cas précédent, f_α est positif $\forall X\in[-1;1]$ si $[-1;1]\subset[X_1;X_2]$.
  La condition $X_1\leq -1$ implique que

$$65-3\alpha+\left(\left(3\alpha+\frac{61}{3}\right)^2+\frac{14000}{9}\right)^{1/2}\geq 8$$

  ce qui mène à

$$\alpha\geq\frac{80}{29}. \tag{2.14}$$

  De la même manière, la condition $X_2\geq 1$ est équivalente à

$$\alpha\geq 6. \tag{2.15}$$

  Par conséquent, $-c_1(\alpha,\beta)\geq 0,\ \forall\beta\in[-\pi,\pi]$ si et seulement si $\alpha\geq 6$.
- Finalement, intéressons nous au signe de $c_0(\alpha,\beta),\forall\beta\in[-\pi,\pi]$.
  Ici aussi, en utilisant le changement de variable $X=\cos(\beta)$, nous avons

$$f_\alpha(X) := 3X^2+2(3-\alpha)X+2\alpha-9\geq 0.$$

  Cette fonction polynomiale f_α admet les deux racines suivantes :

$$\begin{cases} X_1=\dfrac{1}{3}(2\alpha-9),\\ X_2=1. \end{cases}$$

  De la même manière que précédemment, X_2 étant égal à 1, pour que f_α soit positif $\forall X\in[-1;1]$ il faut que $X_1\geq 1$ ce qui mène à la condition

$$\alpha\geq 6. \tag{2.16}$$

En conclusion, en prenant en compte les conditions (2.11), (2.13), (2.14), (2.15) et (2.16), nous avons

$$\lambda_{\min}\geq 0\Leftrightarrow\alpha\geq 6. \tag{2.17}$$

Cette démonstration dans le cas $p=3$ nous permet de retrouver d'une manière originale le résultat obtenu par Ainsworth, Monk et Muniz [2]. Notre technique a l'avantage de pouvoir être

p	c_0	c_1	c_2	c_3	c_4	c_5
1	$\alpha \geq 1$	$\alpha \geq 1$				
2	$\alpha \geq 3$	$\alpha \geq \frac{30}{11}$	$\alpha \geq 2$			
3	$\alpha \geq 6$	$\alpha \geq 6$	$\alpha \geq \frac{87}{17}$	$\alpha \geq 3$		
4	$\alpha \geq 10$	$\alpha \geq \frac{543}{55}$	$\alpha \geq \frac{325}{34}$	$\alpha \geq \frac{581}{73}$	$\alpha \geq 4$	
5	$\alpha \geq 15$	$\alpha \geq 15$	$\alpha \geq \frac{336}{23}$	$\alpha \geq \frac{1185}{86}$	$\alpha \geq \frac{124}{11}$	$\alpha \geq 5$

TABLE 2.1 – Conditions sur α pour chaque coefficient c_i et chaque degré polynomial p

étendue beaucoup plus facilement aux ordres plus élevés. Nous l'avons utilisée pour obtenir une condition sur α pour tout degré polynomial p de $p = 1$ à $p = 5$. Puisque les calculs sont très similaires, nous ne les détaillerons pas ici mais nous résumons les résultats dans le Tab. 2.1.

D'après ces résultats, nous pouvons facilement déduire quels sont les plus petits paramètres de pénalisation assurant la stabilité du schéma (cf. Tab. 2.2). Il est clair que, pour $1 \leq p \leq 5$, pour que la stabilité soit garantie, il faut que

$$\alpha \geq \frac{p\,(p+1)}{2}, \text{ ou, de manière équivalente, que } \gamma \geq \frac{p\,(p+1)}{2h}.$$

p	1	2	3	4	5
	$\alpha \geq 1$	$\alpha \geq 3$	$\alpha \geq 6$	$\alpha \geq 10$	$\alpha \geq 15$

TABLE 2.2 – Condition de stabilité sur α pour chaque degré polynomial p

2.2.3 La condition CFL

Ici, nous proposons de prouver que la condition

$$\Delta t \leq \frac{2}{\sqrt{\lambda_{\max}}}$$

implique (2.3) et (2.4).
Il est clair que pour tout $\beta \in [-\pi;\pi]$, le polynôme q_α admet $p+1$ racines, éventuellement multiples. Nous les notons, quitte à les réordonner, de la plus petite à la plus grande $\lambda_1(\beta), \ldots, \lambda_{p+1}(\beta)$. Ainsi, $\lambda_{\max} = \max\limits_{\beta\in[-\pi,\pi]} [\max(\Lambda(\beta))] = \max\limits_{\beta\in[-\pi;\pi]} \lambda_{p+1}(\beta)$.
L'intervalle $[-\pi;\pi]$ étant un fermé de $\mathbb{R}$, il existe $\beta_{\max}$ tel que $\lambda_{\max} = \lambda_{p+1}(\beta_{\max})$. De plus, le réel $\beta_{\max}$ est tel que l'une des conditions suivantes soit vérifiée :

(1) $\lambda'_{p+1}(\beta_{\max}) = 0$,
(2) $\beta_{\max} = \pm\pi$,
(3) λ_{p+1} n'est pas dérivable en $\beta_{\max}$.

Pour traiter les cas (1) et (3) nous utiliserons le théorème des fonctions implicites :

Théorème 2.2.2 (Théorème des fonctions implicites)**.** *Soit $f(x,y)$ une fonction de classe C^p définie sur un ouvert U de $\mathbb{R}^2$ et à valeurs dans $\mathbb{R}$. Soit (x_0,y_0) un point de U tel que $f(x_0,y_0)=0$ et tel que la dérivée partielle de f, par rapport à la deuxième variable, ne soit pas nulle en (x_0,y_0). Alors, il existe un ouvert V contenu dans U et contenant (x_0,y_0) et une fonction φ de classe C^p définie sur $\mathbb{R}$ à valeurs dans $\mathbb{R}$, tels que l'équivalence suivante soit vraie :*

$$\forall (x,y) \in V,\ f(x,y) = 0 \Leftrightarrow \varphi(x) = y.$$

De plus,

$$\frac{d\varphi}{dx}(x_0) = -\frac{\frac{\partial f}{\partial x}(x_0,y_0)}{\frac{\partial f}{\partial y}(x_0,y_0)}.$$

Le polynôme q_α vérifiant $q_\alpha(\beta, \lambda_{p+1}(\beta)) = 0$, le théorème des fonctions implicites nous permet de vérifier l'existence de $\lambda'_{p+1}(\beta)$ et de calculer sa valeur le cas échéant.
Ainsi, les conditions (1), (2) et (3) peuvent se reformuler de la manière suivante :
(1) $\dfrac{\partial q_\alpha}{\partial \beta}(\beta_{\max}, \lambda_{p+1}(\beta_{\max})) = 0$,
(2) $\beta_{\max} = \pm\pi$,
(3) $\dfrac{\partial q_\alpha}{\partial \lambda}(\beta_{\max}, \lambda_{p+1}(\beta_{\max})) = 0$.

Par conséquent, nous devons déterminer tous les β vérifiant (1), (2) ou (3). On a alors

$$\lambda_{\max} = \max_{\beta \in \mathcal{A}} \lambda_{p+1}(\beta)$$

où $\mathcal{A} = \{\beta \in [-\pi;\pi]$ tel que l'une des conditions (1), (2) ou (3) soit vérifiée$\}$.
Cependant, la troisième condition ne mène à rien d'exploitable. On se limite donc aux $\beta \in \tilde{\mathcal{A}}$ où $\tilde{\mathcal{A}} = \{\beta \in [-\pi;\pi]$: tel que l'une des conditions (1) ou (2) soit vérifiée$\}$.
On a alors $\lambda_{\max} \geq \max\limits_{\beta \in \tilde{\mathcal{A}}} \lambda_{p+1}(\beta)$ ce qui signifie que la condition de stabilité que l'on obtiendra sera seulement nécessaire.

Dans la suite, nous ne détaillerons que le cas $p = 3$, puisque les preuves sont similaires pour tout $p = 1, \ldots, 5$.

Pour la condition (1), nous cherchons $(\beta_0, \lambda(\beta_0))$ tel que $\dfrac{\partial q_\alpha}{\partial \beta}(\beta_0, \lambda(\beta_0)) = 0$. Comme

$$\begin{aligned}\frac{\partial q_\alpha}{\partial \beta}(\beta_0, \lambda(\beta_0)) = \ & \sin(\beta_0)\left[\frac{8}{h^2}(\alpha - 15)\lambda^3(\beta_0) + \frac{240}{h^4}(23 + \alpha - 2\cos(\beta_0))\lambda^2(\beta_0)\right. \\ & \left. + \frac{2880}{h^6}(3\alpha - 65 - 8\cos(\beta_0))\lambda(\beta_0) + \frac{100800}{h^8}(-6\cos(\beta_0) + 2(\alpha - 3))\right],\end{aligned}$$

on obtient les deux conditions suivantes :

$$\sin(\beta_0) = 0 \tag{2.18}$$

et

$$\begin{aligned}&\frac{8}{h^2}(\alpha-15)\lambda^3(\beta_0)+\frac{240}{h^4}(23+\alpha-2\cos(\beta_0))\lambda^2(\beta_0)\\&+\frac{2880}{h^6}(3\alpha-65-8\cos(\beta_0))\lambda(\beta_0)+\frac{100800}{h^8}(-6\cos(\beta_0)+2(\alpha-3))=0.\end{aligned}\tag{2.19}$$

Premièrement, nous considérons la condition (2.18).

- Si $\beta_0=0$, le polynôme q_α admet les racines suivantes :

$$0;\frac{60}{h^2};\frac{90+20\alpha+2g_1(\alpha)}{h^2};\frac{90+20\alpha-2g_1(\alpha)}{h^2}$$

où $g_1(\alpha)=\left(4545-1320\alpha+100\alpha^2\right)^{\frac{1}{2}}$.

Il est clair que, pour $\alpha\geq 0$, les deux racines les plus grandes sont $x_1=\frac{60}{h^2}$ et $x_2=\frac{1}{h^2}(90+20\alpha+2g_1(\alpha))$. En étudiant le signe de la quantité

$$h^2(x_1-x_2)=-150+20\alpha+2g_1(\alpha)$$

on peut facilement obtenir

$$\begin{cases}\lambda_4(0)=x_2 & \text{si } \alpha\geq 6\\ \lambda_4(0)=x_1 & \text{si } \alpha<6.\end{cases}\tag{2.20}$$

Nous avons prouvé en section 2.2.2 que la condition $\alpha\geq 6$ est une condition nécessaire de stabilité. On ne conserve donc que x_2.

- Si $\beta_0=\pm\pi$, les racines du polynôme q_α sont les suivantes :

$$\frac{2}{h^2}\left(45+\sqrt{1605}\right);\frac{2}{h^2}\left(45-\sqrt{1605}\right);\frac{2}{h^2}(-15+6\alpha+g_2(\alpha));\frac{2}{h^2}(-15+6\alpha-g_2(\alpha))$$

où $g_2(\alpha)=\left(405-240\alpha+36\alpha^2\right)^{\frac{1}{2}}$.
Les deux plus grandes racines sont $x_3=\frac{2}{h^2}\left(45+\sqrt{1605}\right)$ et $x_4=\frac{2}{h^2}(-15+6\alpha+g_2(\alpha))$.
L'étude du signe de x_3-x_4 implique

$$\begin{cases}\lambda_4(\pi)=x_3 & \text{si } \alpha\leq 10\\ \lambda_4(\pi)=x_4 & \text{si } \alpha\geq 10.\end{cases}\tag{2.21}$$

A présent, nous devons comparer $\lambda_4(0)$ et $\lambda_4(\pi)$. Nous pouvons facilement vérifier que

$$\begin{cases}\lambda_4(0)\geq\lambda_4(\pi) & \text{si } \alpha\geq\dfrac{2\sqrt{1605}+393}{49}\backsimeq 9.66,\\ \lambda_4(\pi)>\lambda_4(0) & \text{si } \dfrac{2\sqrt{1605}+393}{49}>\alpha.\end{cases}\tag{2.22}$$

On a donc $\max\limits_{\beta\in\tilde{\mathcal{A}}}\lambda_4(\beta)\geq\max\limits_{\beta=\{0,\pi\}}\lambda_4(\beta)$.
Par conséquent, en considérant (2.20) et (2.21), une condition nécessaire de stabilité est

$$\frac{\Delta t}{h}\leq\begin{cases}\dfrac{2}{\sqrt{\lambda_4(\pi)}} & \text{si } 6\leq\alpha\leq\alpha_{1,p},\\ \dfrac{2}{\sqrt{\lambda_4(0)}} & \text{si } \alpha_{1,p}<\alpha,\end{cases}\tag{2.23}$$

où $\alpha_{1,p} = \dfrac{2\sqrt{1605}+393}{49}$, ce qui correspond à la condition nécessaire (2.2). On remarque que la condition (2.18) implique (2) donc on n'aura pas besoin de la considérer une nouvelle fois.

Intéressons-nous maintenant à la condition (2.19), c'est-à-dire, trouvons $(\beta_0, \lambda(\beta_0))$ tel que la condition (2.19) soit vérifiée.
La condition (2.19) est équivalente à

$$\begin{aligned}\cos(\beta_0) &= \frac{(\alpha-15)h^6\lambda^3(\beta_0)+30(23+\alpha)h^4\lambda^2(\beta_0)+360(3\alpha-65)h^2\lambda(\beta_0)+25200(\alpha-3)}{60(h^4\lambda^2(\beta_0)+48h^2\lambda(\beta_0)+1260)}\\ &:= \tilde{Q}_{3,\alpha}(\lambda(\beta_0)).\end{aligned}$$

En injectant cette expression de $\cos(\beta_0)$ dans le polynôme caractéristique (2.9), nous obtenons que $\lambda(\beta_0)$ est solution de

$$q_\alpha(\beta_0, \lambda(\beta_0)) = -\frac{1}{15h^8(h^4\lambda^2(\beta_0)+48h^2\lambda(\beta_0)+1260)}\sum_{i=0}^{6}\lambda^i(\beta_0)h^{2i}\tilde{d}_i(\alpha) = 0$$

ou, de manière équivalente, solution de

$$Q_{3,\alpha}(\lambda(\beta_0)) = \sum_{i=0}^{6}\lambda^i(\beta_0)h^{2i}\tilde{d}_i(\alpha) = 0$$

avec

$$\begin{cases}\tilde{d}_0(\alpha) = & 635040000\left(\alpha^2-12\alpha+36\right),\\ \tilde{d}_1(\alpha) = & 3628800\left(15\alpha^2+70\alpha-96\right),\\ \tilde{d}_2(\alpha) = & 86400\left(31\alpha^2-447\alpha+5316\right),\\ \tilde{d}_3(\alpha) = & 14400\left(8\alpha^2-135\alpha-1728\right),\\ \tilde{d}_4(\alpha) = & 180\left(17\alpha^2-442\alpha+7740\right),\\ \tilde{d}_5(\alpha) = & 60\left(\alpha^2+16\alpha-357\right),\\ \tilde{d}_6(\alpha) = & \alpha^2-30\alpha+210.\end{cases}$$

Une fois les racines de $\tilde{Q}_{3,\alpha}$ calculées, il faut s'assurer que β_0 est bien défini, c'est-à-dire que $|\tilde{Q}_{3,\alpha}(\lambda)| \leq 1$. C'est pourquoi nous nous intéressons uniquement aux valeurs propres λ vérifiant $Q_{3,\alpha}(\lambda) = 0$ et $|\tilde{Q}_{3,\alpha}(\lambda)| \leq 1$, c'est-à-dire aux éléments de $V_{3,\alpha}$.
Finalement, on a $\max\limits_{\beta\in\tilde{\mathcal{A}}}\lambda_4(\beta) = \max\left[\max\left[\lambda_4(0), \lambda_4(\pi)\right], \max V_{3,\alpha}\right]$.
On en déduit alors le théorème 2.1.3. □

Avant de nous intéresser au cas de la dimension d, remarquons que la condition (2.2) ne dépend pas de la dimension d. Cela n'aurait pas été le cas si nous avions exprimé γ comme une fonction du rayon du cercle (ou de la sphère) circonscrit qui est $\sqrt{d}h$. Puisque h est le rayon du cercle (ou de la sphère) inscrit, nous conjecturons que le troisième choix du paramètre ξ_F introduit au chapitre 1 et intervenant dans la pénalisation est la plus approprié. Nous y reviendrons lorsque nous discuterons de l'extension de ce théorème à des maillages composés de rectangles ou de parallélépipèdes.

2.3 Le cas de la dimension d

Dans cette section, nous proposons d'adapter la technique proposée dans [27] pour étendre l'analyse du cas 1D au cas dD. Ici, nous détaillerons uniquement le cas 3D, puisque la technique est exactement la même pour le cas 2D.
Tout d'abord, nous considérons un domaine infini et homogène Ω, maillé uniformément avec des cubes dont la longueur d'arête est h.
A présent, nous rappelons les notations introduites dans [27].

- $\Omega = \bigcup_{K_J \in \mathcal{T}_h} K_J$ où $K_J = \prod_{k=1}^{3} S_{J_k} = \prod_{k=1}^{3} [J_k h, (J_k + 1)\, h]$ et $J = (J_k)_{k=1,\dots,3}$.
- Sur $\hat{K} = [0,1]^d$, nous définissons les fonctions de base de Lagrange $(\hat{\varphi}_{\mathbf{l}})_{\mathbf{l} \in \{1,\dots,p+1\}^3}$ par

$$\hat{\varphi}_{\mathbf{l}}(\mathbf{x}) = \prod_{k=1}^{3} \hat{\varphi}_{l_k}(x_k)$$

  où $\hat{\varphi}_{l_k}$ est une fonction de base de Lagrange 1D.
- Puisque le maillage est uniforme, les fonctions de base sont définies sur l'élément J grâce aux fonctions $(\hat{\varphi}_{\mathbf{l}})_{\mathbf{l} \in \{1,\dots,p+1\}^3}$ par

$$\varphi^J_{\mathbf{m}}(\mathbf{x}) = \hat{\varphi}_{\mathbf{m}}\left(\frac{x - Jh}{h}\right) 1_{K_{\mathbf{J}}}(\mathbf{x})$$

  où 1_{K_J} est la fonction indicatrice de K_J. Ces fonctions peuvent être écrites comme un produit de d fonctions de base 1D :

$$\varphi^J_{\mathbf{m}}(\mathbf{x}) \quad = \prod_{k=1}^{3} \hat{\varphi}_{m_k}\left(\frac{x - J_k h}{h}\right) 1_{S_{J_k}}(\mathbf{x}_k)\,.$$

- Les différentes faces de l'élément de référence $\hat{K}$ sont notées par un exposant C correspondant à l'orientation de la face : Nord (N), Sud (S), Est (E), Ouest (W), devant (F : front) et derrière (B : back) (cf. Fig 2.1).

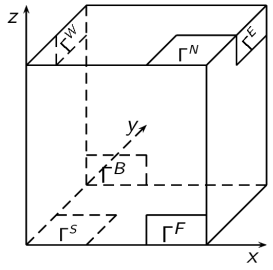

FIGURE 2.1 – Notations des faces en 3D

Puisque le maillage est uniforme, on peut réécrire le problème sur un élément $\mathbf{I} = \{I_1, I_2, I_3\}$:

$$\begin{aligned} M_{3,p}\delta^n U_{I_1,I_2,I_3} = \quad & K_{3,p}U_{I_1,I_2,I_3} + K^E_{3,p}U_{I_1+1,I_2,I_3} + K^W_{3,p}U_{I_1-1,I_2,I_3} \\ & + K^F_{3,p}U_{I_1,I_2+1,I_3} + K^B_{3,p}U_{I_1,I_2-1,I_3} + K^N_{3,p}U_{I_1,I_2,I_3+1} \\ & + K^S_{3,p}U_{I_1,I_2,I_3-1} \end{aligned} \tag{2.24}$$

où

- U_{I_1,I_2,I_3} correspond à la restriction de U sur l'élément $\mathbf{I}$.
- $\delta^n U_{I_1,I_2,I_3} = \dfrac{U^{n+1}_{I_1,I_2,I_3} - 2U^n_{I_1,I_2,I_3} + U^{n-1}_{I_1,I_2,I_3}}{\Delta t^2}$
- $M_{3,p}$ est un bloc de la matrice de masse M,

$$M_{3,p}(\mathbf{i},\mathbf{j}) = h^3 \int_{\hat{K}} \hat{\varphi}_{\mathbf{i}} \hat{\varphi}_{\mathbf{j}}\, d\hat{x}, \qquad \mathbf{i},\mathbf{j} \in \{1,\dots,p+1\}^3$$

- $K_{3,p}$ est un bloc diagonal de la matrice K

$$\begin{aligned} K_{3,p}(\mathbf{i},\mathbf{j}) &= h \int_{\hat{K}} \nabla\hat{\varphi}_{\mathbf{i}} \cdot \nabla\hat{\varphi}_{\mathbf{j}}\, d\hat{x} - \sum_{C\in\{N,S,E,W,B,F\}} \frac{h}{2} \int_{\Gamma^C} \left(\hat{\varphi}_{\mathbf{i}} \nabla\hat{\varphi}_{\mathbf{j}} + \hat{\varphi}_{\mathbf{j}} \nabla\hat{\varphi}_{\mathbf{i}}\right) \nu_C\, d\sigma \\ &\quad + \sum_{C\in\{N,S,E,W,B,F\}} h^2 \int_{\Gamma^C} \gamma \hat{\varphi}_{\mathbf{i}} \hat{\varphi}_{\mathbf{j}}\, d\sigma, \qquad \mathbf{i},\mathbf{j} \in \{1,\dots,p+1\}^3 \end{aligned}$$

  où ν_C est le vecteur normal unitaire extérieur à la face Γ^C.
- $K^C_{3,p}$ est un bloc de la matrice K correspondant aux interactions entre un élément I et son voisin par la face Γ^C :

$$\begin{aligned} K^E_{3,p}(\mathbf{i},\mathbf{j}) &= \int_{[0,1]^2} \frac{h}{2} \left(\hat{\varphi}_{\mathbf{i}}(1,x_2,x_3)\, \nabla\hat{\varphi}_{\mathbf{j}}(0,x_2,x_3) + \hat{\varphi}_{\mathbf{j}}(0,x_2,x_3)\, \nabla\hat{\varphi}_{\mathbf{i}}(1,x_2,x_3)\right) \nu_E \\ &\quad - h^2 \hat{\varphi}_{\mathbf{i}}(1,x_2,x_3)\, \hat{\varphi}_{\mathbf{j}}(0,x_2,x_3)\, dx_2 dx_3 \\ K^F_{3,p}(\mathbf{i},\mathbf{j}) &= \int_{[0,1]^2} \frac{h}{2} \left(\hat{\varphi}_{\mathbf{i}}(x_1,1,x_3)\, \nabla\hat{\varphi}_{\mathbf{j}}(x_1,0,x_3) + \hat{\varphi}_{\mathbf{j}}(x_1,0,x_3)\, \nabla\hat{\varphi}_{\mathbf{i}}(x_1,1,x_3)\right) \nu_F \\ &\quad - h^2 \hat{\varphi}_{\mathbf{i}}(x_1,1,x_3)\, \hat{\varphi}_{\mathbf{j}}(x_1,0,x_3)\, dx_1 dx_3 \\ K^N_{3,p}(\mathbf{i},\mathbf{j}) &= \int_{[0,1]^2} \frac{h}{2} \left(\hat{\varphi}_{\mathbf{i}}(x_1,x_2,1)\, \nabla\hat{\varphi}_{\mathbf{j}}(x_1,x_2,0) + \hat{\varphi}_{\mathbf{j}}(x_1,x_2,0)\, \nabla\hat{\varphi}_{\mathbf{i}}(x_1,x_2,1)\right) \nu_N \\ &\quad - h^2 \hat{\varphi}_{\mathbf{i}}(x_1,x_2,1)\, \hat{\varphi}_{\mathbf{j}}(x_1,x_2,0)\, dx_1 dx_2 \end{aligned}$$

$$\begin{aligned} K^W_{3,p}((i_1,i_2,i_3),(j_1,j_2,j_3)) &= K^E_{3,p}((j_1,i_2,i_3),(i_1,j_2,j_3)) \\ K^B_{3,p}((i_1,i_2,i_3),(j_1,j_2,j_3)) &= K^F_{3,p}((i_1,j_2,i_3),(j_1,i_2,j_3)) \\ K^S_{3,p}((i_1,i_2,i_3),(j_1,j_2,j_3)) &= K^N_{3,p}((i_1,i_2,j_3),(j_1,j_2,i_3)) \end{aligned}$$

Alors, en multipliant l'équation (2.24) par l'inverse de la matrice de masse $M_{3,p}$, on obtient

$$\begin{aligned} \delta^n U_{I_1,I_2,I_3} = \quad & N_{3,p} U_{I_1,I_2,I_3} + N^E_{3,p} U_{I_1+1,I_2,I_3} + N^W_{3,p} U_{I_1-1,I_2,I_3} \\ & + N^F_{3,p} U_{I_1,I_2+1,I_3} + N^B_{3,p} U_{I_1,I_2-1,I_3} + N^N_{3,p} U_{I_1,I_2,I_3+1} \\ & + N^S_{3,p} U_{I_1,I_2,I_3-1} \end{aligned} \tag{2.25}$$

où $N_{3,p} = M^{-1}_{3,p} K_{3,p}$ et $N^C_{3,p} = M^{-1}_{3,p} K^C_{3,p}$.

A présent, nous allons nous intéresser à la façon de réécrire les matrices $N_{3,p}$ et $N^C_{3,p}$ par rapport aux matrices que nous avons obtenues dans le cas mono dimensionnel.

2.3.1 Du cas 3D au cas 1D

Les coefficients de $M_{3,p}$, $K_{3,p}$, $K^C_{3,p}$, $N_{3,p}$ et $N^C_{3,p}$ peuvent être déduits des coefficients de $M_{1,p}$, $K_{1,p}$, $K^W_{1,p}$, $N_{1,p}$ et $N^W_{1,p}$ grâce au théorème suivant.

Théorème 2.3.1. *Pour tout* $\mathbf{m} = (m_k)_{k=1,\dots,3} \in \{1,\dots,p+1\}^3$, $\mathbf{n} = (n_k)_{k=1,\dots,3} \in \{1,\dots,p+1\}^3$ *et* $C = \{E, W, N, S, B, F\}$, *nous avons*

$$
\begin{aligned}
&1.\quad M_{3,p}(\mathbf{m},\mathbf{n}) = \prod_{i=1}^{3} M_{1,p}(m_i,n_i)\,,\\
&2.\quad K_{3,p}(\mathbf{m},\mathbf{n}) = \sum_{i=1}^{3}\left(K_{1,p}(m_i,n_i)\prod_{k=1,\,k\neq i}^{3} M_{1,p}(m_k,n_k)\right),\\
&3.\quad K^C_{3,p}(\mathbf{m},\mathbf{n}) = K^W_{1,p}(m_{p_C},n_{p_C})\prod_{k=1,\,k\neq p_C}^{3} M_{1,p}(m_k,n_k)\,,\\
&4.\quad N_{3,p}(\mathbf{m},\mathbf{n}) = \sum_{p=1}^{3} N_{1,p}(m_p,n_p)\prod_{k=1,\,k\neq p}^{3}\delta_{m_k,n_k},\\
&5.\quad N^C_{3,p}(\mathbf{m},\mathbf{n}) = N^W_{1,p}(m_{p_C},n_{p_C})\prod_{k=1,\,k\neq p_C}^{3}\delta_{m_k,n_k},
\end{aligned}
\tag{2.26}
$$

où $p_C = \begin{cases} 1 & \text{si } C \in \{E,W\}\,,\\ 2 & \text{si } C \in \{N,S\}\,, \\ 3 & \text{si } C \in \{B,F\}\,, \end{cases}$ *et* $N_{1,p} = M^{-1}_{1,p}K_{1,p}$ *et* $N^W_{1,p} = M^{-1}_{1,p}K^W_{1,p}$.

La preuve de ce théorème est donnée en annexe 2.C.

2.3.2 Conséquences sur l'analyse de stabilité

Appliquons tout d'abord une transformée de Fourier dans les trois directions d'espace à (2.25) pour obtenir, pour $\boldsymbol{\beta} \in [-\pi,\pi]^3$,

$$
\delta^n \tilde{U}_{\beta_1,\beta_2,\beta_3} = \mathbf{N}_\beta \tilde{U}_{\beta_1,\beta_2,\beta_3} \tag{2.27}
$$

où la matrice $\mathbf{N}_{\boldsymbol{\beta}}$ est définie par

$$
\begin{aligned}
\mathbf{N}_{\boldsymbol{\beta}}(\mathbf{m},\mathbf{n}) \;=& \sum_{j=1}^{3}\Bigg[N_{1,p}(m_j,n_j)\prod_{q=1,\,q\neq j}^{3}\delta_{m_q,n_q} + e^{i\beta_j}N^W_{1,p}(m_j,n_j)\prod_{q=1,\,q\neq j}^{3}\delta_{m_q,n_q}\\
&+e^{-i\beta_j}N^W_{1,p}(n_j,m_j)\prod_{q=1,\,q\neq j}^{3}\delta_{m_q,n_q}\Bigg]
\end{aligned}
$$

ce qui peut être réécrit comme

$$
\begin{aligned}
\mathbf{N}_{\boldsymbol{\beta}}(\mathbf{m},\mathbf{n}) \;&= \sum_{j=1}^{3}\left(\left(N_{1,p}(m_j,n_j) + e^{i\beta_j}N^W_{1,p}(m_j,n_j) + e^{-i\beta_j}N^W_{1,p}(n_j,m_j)\right)\prod_{q=1,\,q\neq j}^{3}\delta_{m_q,n_q}\right)\\
&= \sum_{j=1}^{3} N_{\beta_j}(m_j,n_j)\prod_{q=1,\,q\neq j}^{3}\delta_{m_q,n_q}.
\end{aligned}
$$

En utilisant l'analyse de stabilité comme dans la section 2.2.1, la stabilité du schéma est assurée si et seulement si

$$\lambda_{\min,3} \geq 0 \qquad \text{et} \qquad \lambda_{\max,3} \leq \frac{4}{\Delta t^2}$$

où $\lambda_{\min,3} = \min\limits_{\boldsymbol{\beta}\in[-\pi,\pi]^3} (\min \Lambda(\mathbf{N}_{\boldsymbol{\beta}}))$, $\lambda_{\max,3} = \max\limits_{\boldsymbol{\beta}\in[-\pi,\pi]^3} (\max \Lambda(\mathbf{N}_{\boldsymbol{\beta}}))$ et $\Lambda(\mathbf{N}_{\boldsymbol{\beta}})$ est l'ensemble des valeurs propres de $\mathbf{N}_{\boldsymbol{\beta}}$.

Pour calculer ces valeurs, nous utilisons le lemme suivant.

Lemme 2.3.2. *Soit $\left(\lambda_i^{\boldsymbol{\beta}}\right)_{i=1,\dots,p+1}$ les valeurs propres de $N_{\boldsymbol{\beta}}$ et $\left(v_i^{\boldsymbol{\beta}}\right)_{i=1,\dots,p+1}$ les vecteurs propres associés. Alors, les valeurs propres de $\mathbf{N}_{\boldsymbol{\beta}}$ sont données par*

$$\lambda_{\mathbf{i}}^{\boldsymbol{\beta}} = \lambda_{i_1,i_2,i_3}^{\beta_1,\beta_2,\beta_3} = \sum_{k=1}^{3} \lambda_{i_k}^{\beta_k}$$

et le vecteur propre associé à la valeur propre $\lambda_{\mathbf{i}}^{\boldsymbol{\beta}}$ est défini par

$$v_{\mathbf{i}}^{\boldsymbol{\beta}}(\mathbf{m}) = v_{i_1,i_2,i_3}^{\beta_1,\beta_2,\beta_3}(m_1,m_2,m_3) = \prod_{k=1}^{3} v_{i_k}^{\beta_k}(m_k) \tag{2.28}$$

où $\mathbf{m} = (m_1,m_2,m_3)$ et $v_{i_k}^{\beta_k}(m_k)$ est la m_k-ième composante du vecteur $v_{i_k}^{\beta_k}$.

Démonstration. Soit $v_{\mathbf{i}}^{\boldsymbol{\beta}}$ défini par (2.28). Alors

$$\begin{aligned}
\left(N_{\boldsymbol{\beta}} v_{\mathbf{i}}^{\boldsymbol{\beta}}\right)(\mathbf{m}) &= \sum_{n_1,\dots,n_3=1}^{p+1} \left(\sum_{q=1}^{3} N_{\beta_q}(m_q,n_q) \prod_{k=1,\,k\neq q}^{3} \delta_{m_k,n_k} \prod_{k=1}^{3} v_{i_k}^{\beta_k}(n_k) \right) \\
&= \sum_{q=1}^{3} \left(\left(\sum_{n_q=1}^{p+1} N_{\beta_q}(m_q,n_q)\, v_{i_q}^{\beta_q}(n_q) \right) \prod_{k=1,\,k\neq q}^{3} v_{i_k}^{\beta_k}(m_k) \right) \\
&= \sum_{q=1}^{3} \left(\lambda_{i_q}^{\beta_q} v_{i_q}^{\beta_q}(m_q) \prod_{k=1,\,k\neq q}^{3} v_{i_k}^{\beta_k}(m_k) \right) \\
&= \left(\sum_{q=1}^{3} \lambda_{i_q}^{\beta_q} \right) v_{\mathbf{i}}^{\boldsymbol{\beta}}(\mathbf{m})
\end{aligned}$$

□

Il est alors clair que

$$\lambda_{\min,3} = 3\lambda_{\min} \qquad \text{et} \qquad \lambda_{\max,3} = 3\lambda_{\max}.$$

Le schéma (2.28) est donc stable si et seulement si

$$\lambda_{\min} \geq 0 \qquad \text{et} \qquad \frac{c\Delta t}{h} \leq \frac{1}{\sqrt{3}} \sqrt{\frac{2}{\lambda_{\max}}}.$$

La première condition est équivalente à la condition (2.2) tandis que la seconde implique (2.3) et (2.4).

2.3.3 Extension aux maillages rectangulaires ou parallélépipèdiques

Comme nous l'avons précisé lors de l'analyse du théorème 2.1.1, le cas des maillages carrés ou cubiques ne nous permet pas de déterminer le choix le plus approprié de ξ_F. En effet, les rayons des cercles inscrits et circonscrits sont alors proportionnels. Dans le cas de maillages rectangulaires ou parallélépipèdiques, nous avons en 2D :

$$2\rho_{\text{ins}} = \min(h_x, h_y) \quad \text{et} \quad 2\rho_{\text{circ}} = \sqrt{h_x^2 + h_y^2}$$

et en 3D :

$$2\rho_{\text{ins}} = \min(h_x, h_y, h_z) \quad \text{et} \quad 2\rho_{\text{circ}} = \sqrt{h_x^2 + h_y^2 + h_y^2}.$$

Ici, h_x, h_y et h_z représentent respectivement la longueur des côtés des éléments dans les directions x, y et z.
On peut étendre le théorème 2.1.1 aux cas de maillages rectangulaires ou parallélépipèdiques pour montrer qu'une condition nécessaire de stabilité est, en 2D :

$$\gamma \geq \frac{p(p+1)}{2\min(h_x, h_y)}$$

et en 3D :

$$\gamma \geq \frac{p(p+1)}{2\min(h_x, h_y, h_z)}.$$

Nous ne démontrons pas ce résultat ici, mais c'est une extension triviale du théorème 2.1.1. Il indique que le choix de ξ_F faisant intervenir le rayon de la sphère inscrite (ou du cercle en 2D) est le plus appropriée.
La preuve peut aussi être étendue pour obtenir une condition CFL, mais son expression est compliquée et n'apporte pas plus de précisions.

2.4 Résultats numériques

Dans cette section, nous représentons tout d'abord le comportement de la condition CFL par rapport à α et nous montrons que l'ensemble $V_{p,\alpha}$ est vide pour la plupart des valeurs de α (section 2.4.1). Cela illustre le fait que le théorème 2.1.1 fournit une condition qui est en fait nécessaire et suffisante pour la plupart des valeurs de α. Nous comparons ensuite la condition CFL analytique en domaine infini avec la condition CFL calculée numériquement sur des maillages finis pour illustrer le résultat du théorème 2.1.3 (section 2.4.2).

2.4.1 Comportement de la condition CFL par rapport à α

Dans les Fig. 2.2, 2.4, 2.6, 2.8 et 2.10, nous avons représenté les fonctions $C_{1,p}$ (ligne bleue avec losanges), $C_{2,p}(\alpha)$ (ligne rouge avec cercles) et $C_{3,p}(\alpha)$ (ligne noire) respectivement pour $p = 1, 2, 3, 4$ et 5. La condition nécessaire de stabilité est obtenue pour α donné en choisissant le minimum de ces trois courbes.
La fonction $C_{3,p}(\alpha)$ modifie seulement la condition CFL dans un voisinage restreint de $\alpha_{1,p}$. Le comportement est confirmé par les Fig. 2.3, 2.5, 2.7, 2.9 et 2.11 qui représentent un zoom autour de $\alpha_{1,p}$. On illustre ici encore le fait que le théorème 2.1.1 fournit une condition nécessaire et suffisante à l'exception d'un intervalle très étroit autour de $\alpha_{1,p}$. De plus, la condition CFL reste constante pour α allant de $\dfrac{p(p+1)}{2}$ à une valeur proche de $\alpha_{1,p}$, ce qui signifie qu'il n'est pas nécessaire de choisir $\alpha = \dfrac{p(p+1)}{2}$ pour optimiser le pas de temps.

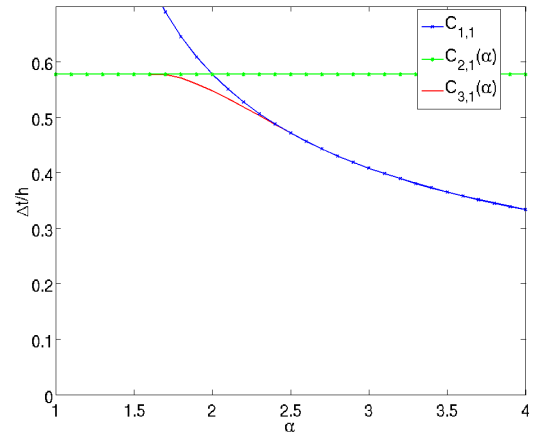

FIGURE 2.2 – Les 3 conditions de stabilité pour $p = 1$

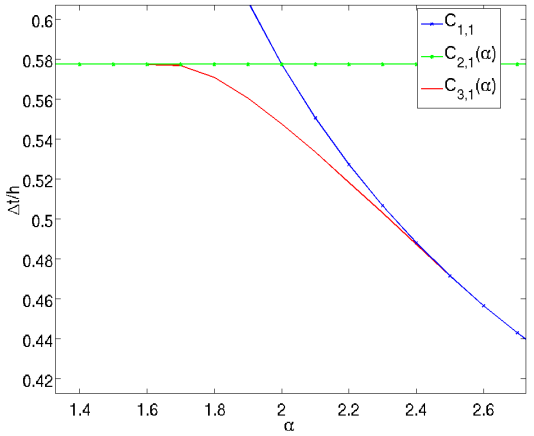

FIGURE 2.3 – Zoom sur les 3 conditions de stabilité pour $p = 1$

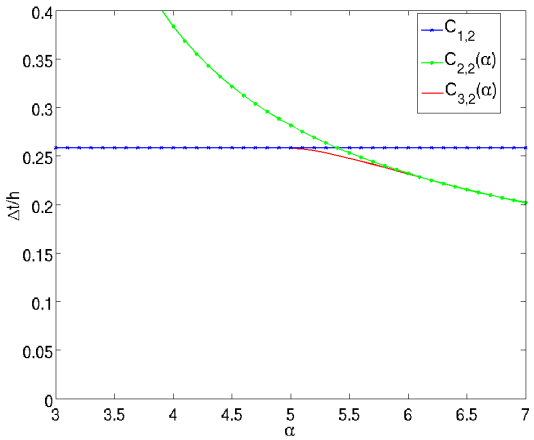

FIGURE 2.4 – Les 3 conditions de stabilité pour $p = 2$

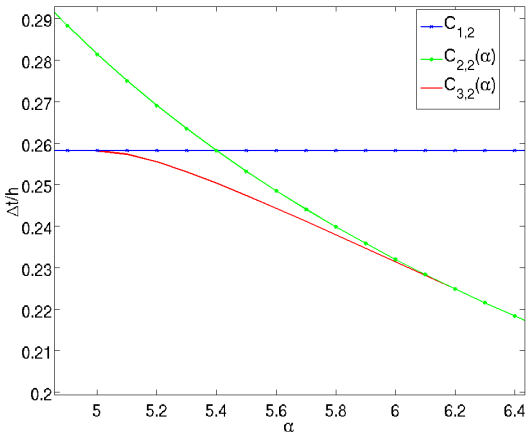

FIGURE 2.5 – Zoom sur les 3 conditions de stabilité pour $p = 2$

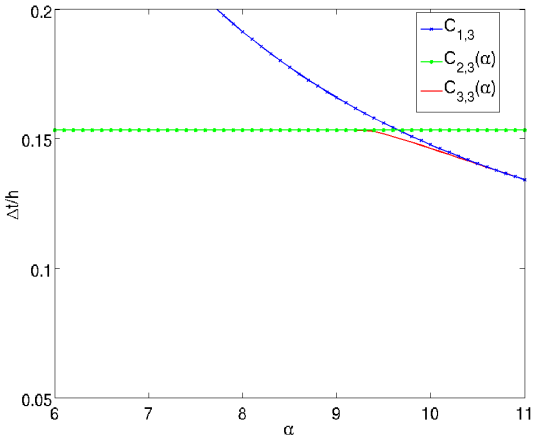

FIGURE 2.6 – Les 3 conditions de stabilité pour $p = 3$

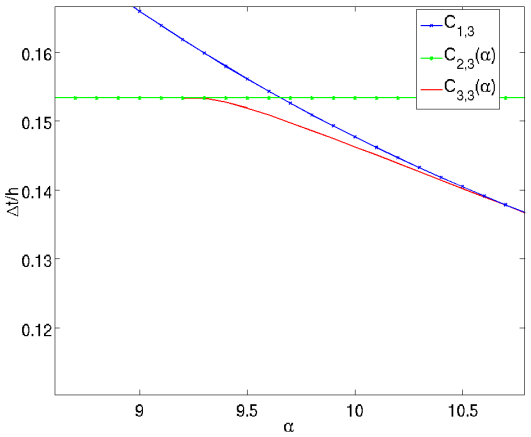

FIGURE 2.7 – Zoom sur les 3 conditions de stabilité pour $p = 3$

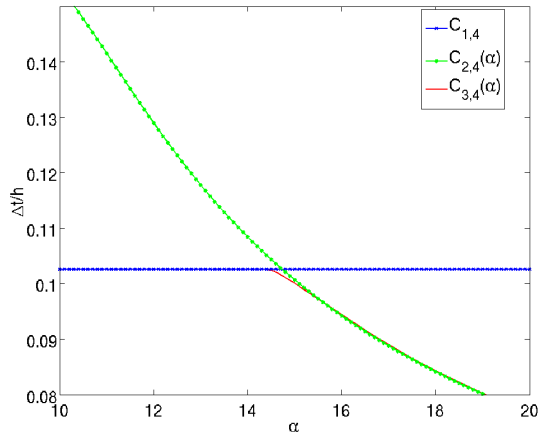

FIGURE 2.8 – Les 3 conditions de stabilité pour $p = 4$

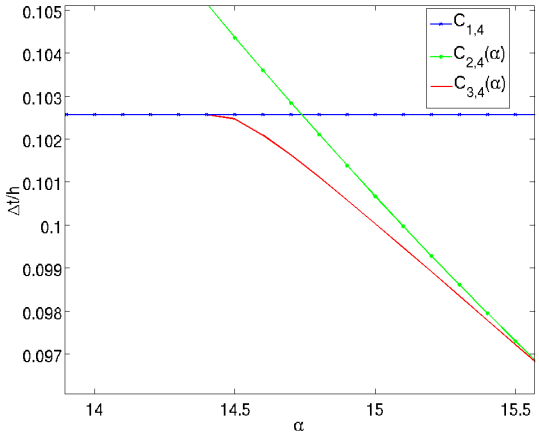

FIGURE 2.9 – Zoom sur les 3 conditions de stabilité pour $p = 4$

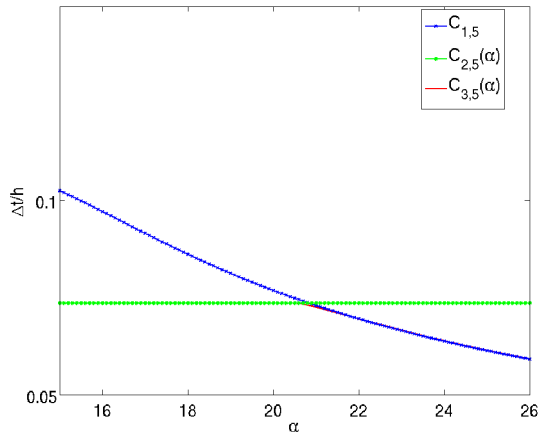

FIGURE 2.10 – Les 3 conditions de stabilité pour $p = 5$

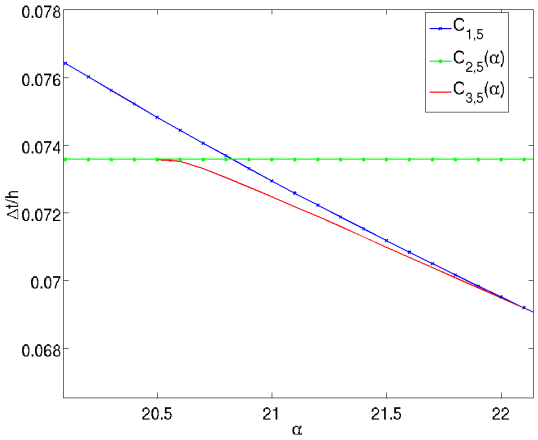

FIGURE 2.11 – Zoom sur les 3 conditions de stabilité pour $p = 5$

2.4.2 Confrontation à des résultats numériques

Dans cette section, nous confrontons les résultats que nous avons obtenus précédemment à des expériences numériques. Les conditions CFL 2D et 3D étant proportionnelles aux conditions CFL 1D, on se limite ici à des expériences en dimension un. Nous considérons la simulation de la propagation des ondes dans un domaine homogène 1D $\Omega = [0, 10]$ avec une vitesse $c = (\mu/\rho)^{1/2} = 1\ \mathrm{ms}^{-1}$. Nous imposons aussi des conditions de bord de Dirichlet aux deux extrémités du domaine et la longueur du pas d'espace est $h = 0.1$.
Nous calculons numériquement la plus grande valeur propre $\lambda_{\max}$ de la matrice $M^{-1}K$ et nous en déduisons la condition CFL numérique du schéma en utilisant la formule $\dfrac{c\Delta t}{h} \leq \dfrac{2}{\sqrt{\lambda_{\max}}}$.
Dans les Fig. 2.12, 2.13, 2.14, 2.15 et 2.16 nous comparons la condition CFL analytique (ligne rouge) donnée par le théorème 2.1.3 à la condition CFL numérique (triangles), respectivement pour $p = 1, 2, 3, 4$ et 5. Toutes les figures montrent une très bonne concordance entre les CFL analytique et numérique, ce qui illustre le fait que les conditions que nous avons calculées sont en pratique nécessaires et suffisantes.

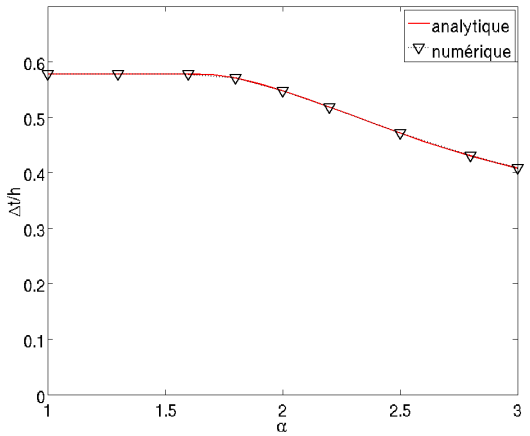

FIGURE 2.12 – Comparaison numérique en P^1

2.5 Conclusion

Dans ce chapitre, nous avons proposé des conditions nécessaires de stabilité L^2 pour la méthode IPDG utilisant des maillages réguliers. Nous avons donc confirmé par la preuve la conjecture de Ainsworth, Monk and Muniz jusqu'à $p = 5$. De plus, nous avons observé que la condition CFL est constante par rapport à α sur un segment $\left[\dfrac{p\,(p+1)}{2}, \tilde{\alpha}\right]$ et décroît comme $\alpha^{-1/2}$ pour $\alpha > \tilde{\alpha}$. Cela signifie qu'il n'est pas nécessaire de choisir α trop proche de $\dfrac{p\,(p+1)}{2}$ pour améliorer la condition CFL. Finalement, nous avons observé à la section 2.3.3 qu'un bon choix pour ξ_F serait de considérer le rayon du cercle (ou de la sphère) inscrite. Cela sera confirmé par une analyse sur des maillages triangulaires dans le chapitre 3.

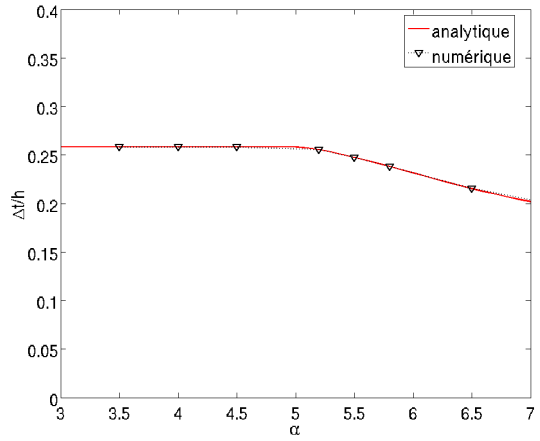

FIGURE 2.13 – Comparaison numérique en P^2

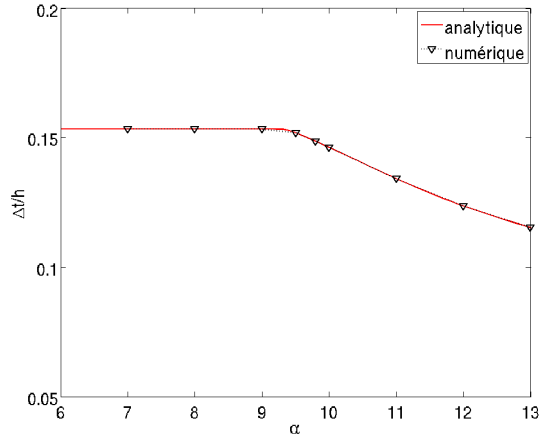

FIGURE 2.14 – Comparaison numérique en P^3

2.A Expression du polynôme q_α

Dans cette section, nous présentons les expressions du polynôme q_α, introduit dans la section 2.2.3, pour des degrés polynomiaux égaux à 1, 2, 4 et 5.

- Dans le cas de fonctions de base discontinues de degré 1, nous pouvons facilement obtenir le polynôme caractéristique suivant, associé à la matrice N_β

$$q_\alpha(\beta, \lambda) = \lambda^2 + c_1(\alpha, \beta)\lambda + c_0(\alpha, \beta) \tag{2.29}$$

avec

$$\begin{cases} c_1(\alpha, \beta) = & \frac{4}{h^2}((3-\alpha)\cos(\beta) - 2\alpha) \\ c_0(\alpha, \beta) = & \frac{12}{h^4}\left(\cos^2(\beta) - 2\alpha\cos(\beta) + 2\alpha - 1\right). \end{cases}$$

- Dans le cas de fonctions de base discontinues de degré 2, le polynôme caractéristique associé à la matrice N_β est

$$q_\alpha(\beta, \lambda) = -\lambda^3 + c_2(\alpha, \beta)\lambda^2 + c_1(\alpha, \beta)\lambda + c_0(\alpha, \beta) \tag{2.30}$$

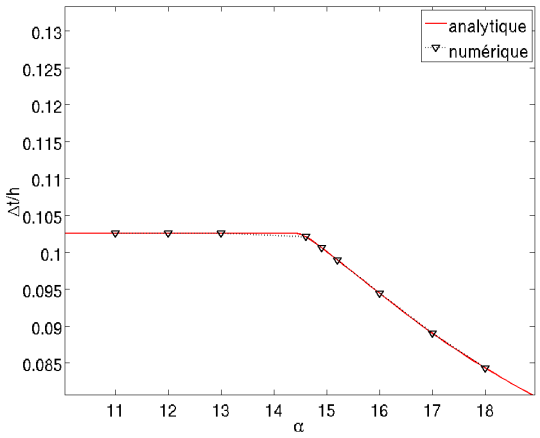

FIGURE 2.15 – Comparaison numérique en P^4

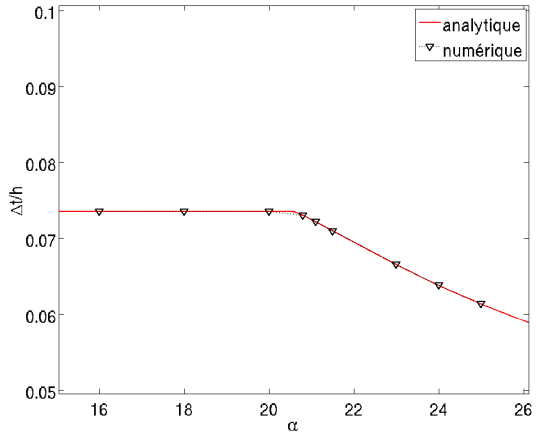

FIGURE 2.16 – Comparaison numérique en P^5

avec

$$\begin{cases} c_2(\alpha,\beta) = & -\dfrac{6}{h^2}\left((\alpha-8)\cos(\beta)-3\alpha\right) \\ c_1(\alpha,\beta) = & \dfrac{12}{h^4}\left(-6\cos^2(\beta)-2(15+4\alpha)\cos(\beta)+4(24-13\alpha)\right) \\ c_0(\alpha,\beta) = & -\dfrac{1440}{h^6}\left(\cos^2(\beta)+(\alpha-3)\cos(\beta)+2-\alpha\right). \end{cases}$$

- Dans le cas de fonctions de base discontinues de degré 4, le polynôme caractéristique associé à la matrice N_β est

$$q_\alpha(\beta,\lambda) = -\lambda^5 + \sum_{i=0}^{4} c_i(\alpha,\beta)\,\lambda^i \tag{2.31}$$

avec

$$\begin{cases} c_4(\alpha,\beta) = & -\dfrac{10}{h^2}\left((\alpha-24)\cos(\beta)-5\alpha\right) \\ c_3(\alpha,\beta) = & -\dfrac{120}{h^4}\left(5\cos^2(\beta)+(4\alpha+287)\cos(\beta)+10(15\alpha-88)\alpha\right) \\ c_2(\alpha,\beta) = & -\dfrac{10080}{h^6}\left(5\cos^2(\beta)+(3\alpha-305)\cos(\beta)+990-133\alpha\right) \\ c_1(\alpha,\beta) = & -\dfrac{201600}{h^8}\left(15\cos^2(\beta)+(8\alpha+165)\cos(\beta)+2(59\alpha-468)\right) \\ c_0(\alpha,\beta) = & -\dfrac{50803200}{h^{10}}\left(2\cos^2(\beta)+(\alpha-10)\cos(\beta)+8-\alpha\right). \end{cases}$$

- Dans le cas de fonctions de base discontinues de degré 5, le polynôme caractéristique associé à la matrice N_β est

$$q_\alpha(\lambda,\beta) = \lambda^6 + \sum_{i=0}^{5} c_i(\alpha,\beta)\lambda^i \tag{2.32}$$

avec

$$\begin{cases} c_5(\alpha,\beta) = & -\dfrac{12}{h^2}\left((\alpha-35)\cos(\beta)+6\alpha\right) \\ c_4(\alpha,\beta) = & -\dfrac{420}{h^4}\left(-3\cos^2(\beta)+(2\alpha+336)\cos(\beta)+1155-134\alpha\right) \\ c_3(\alpha,\beta) = & -\dfrac{40320}{h^6}\left(-4\cos^2(\beta)+(2\alpha-702)\cos(\beta)+256\alpha-2849\alpha\right) \\ c_2(\alpha,\beta) = & -\dfrac{1814400}{h^8}\left(-9\cos^2(\beta)+(4\alpha770)\cos(\beta)+9343-326\alpha\right) \\ c_1(\alpha,\beta) = & -\dfrac{101606400}{h^{10}}\left(-12\cos^2(\beta)+(5\alpha-303)\cos(\beta)+94\alpha-1170\right) \\ c_0(\alpha,\beta) = & -\dfrac{10059033600}{h^{12}}\left(-5\cos^2(\beta)+(2\alpha-20)\cos(\beta)+25-2\alpha\right). \end{cases}$$

2.B Définition de $Q_{p,\alpha}$ et $\tilde{Q}_{p,\alpha}$

Nous présentons ici les expressions du polynôme $Q_{p,\alpha}$ et de la fonction rationnelle $\tilde{Q}_{p,\alpha}$ pour $1 \leq p \leq 5$.

On rappelle que la fonction rationnelle $\tilde{Q}_{p,\alpha}$ est obtenue en cherchant $(\beta_0, \lambda(\beta_0))$ tel que $\dfrac{\partial q_\alpha}{\partial \beta}(\beta_0, \lambda(\beta_0)) = 0$. A partir, de ce calcul, on va pouvoir déterminer $\cos(\beta_0)$ que l'on notera $\tilde{Q}_{p,\alpha}$. Le polynôme $Q_{p,\alpha}$ est ensuite déterminé en injectant cette expression de $\cos(\beta_0)$ dans le polynôme caractéristique q_α.

- Pour les polynômes de degré 1, $\tilde{Q}_{1,\alpha}$ est défini par

$$\tilde{Q}_{1,\alpha}(\lambda) = \frac{h^2\lambda}{2}\left(\frac{\alpha}{3}-1\right)+\alpha.$$

Nous avons de plus

$$Q_{1,\alpha}(\lambda) = \sum_{i=0}^{2} \lambda^i h^{2i} \tilde{d}_i(\alpha)$$

avec

$$\begin{cases} \tilde{d}_0(\alpha) = & 36\left(\alpha^2 - 2\alpha + 1\right), \\ \tilde{d}_1(\alpha) = & 12\left(\alpha^2 - \alpha\right), \\ \tilde{d}_2(\alpha) = & \alpha^2 - 6\alpha + 6, \end{cases}$$

- Dans le cas $p = 2$, la définition de $\tilde{Q}_{p,\alpha}$ est

$$\tilde{Q}_{2,\alpha}(\lambda) = -\frac{(\alpha-1)h^4\lambda^2 + 4(15+4\alpha)h^2\lambda + 240(\alpha-3)}{24(h^2\lambda+20)}.$$

Le polynôme $Q_{2,\alpha}$ est tel que

$$Q_{2,\alpha}(\lambda) = \sum_{i=0}^{4} \lambda^i h^{2i} \tilde{d}_i(\alpha)$$

avec

$$\begin{cases} \tilde{d}_0(\alpha) = & 57600\left(\alpha^2 - 2\alpha + 1\right), \\ \tilde{d}_1(\alpha) = & 1920\left(4\alpha^2 - 43\alpha + 39\right), \\ \tilde{d}_2(\alpha) = & 16\left(46\alpha^2 - 342\alpha + 1521\right), \\ \tilde{d}_3(\alpha) = & 8\left(4\alpha^2 + \alpha - 140\right), \\ \tilde{d}_4(\alpha) = & \alpha^2 - 16\alpha + 56. \end{cases}$$

- Pour $p = 3$, nous avons $\tilde{Q}_{p,\alpha}$ défini par

$$\tilde{Q}_{3,\alpha}(\lambda) = \frac{(\alpha-15)h^6\lambda^3 + 30(23+\alpha)h^4\lambda^2 + 360(3\alpha-65)h^2\lambda + 25200(\alpha-3)}{60(h^4\lambda^2 + 48h^2\lambda + 1260)}$$

et

$$Q_{p,\alpha}(\lambda) = \sum_{i=0}^{6} \lambda^i h^{2i} \tilde{d}_i(\alpha)$$

avec

$$\begin{cases} \tilde{d}_0(\alpha) = & 635040000\left(\alpha^2 - 12\alpha + 36\right), \\ \tilde{d}_1(\alpha) = & 3628800\left(15\alpha^2 + 70\alpha - 96\right), \\ \tilde{d}_2(\alpha) = & 86400\left(31\alpha^2 - 447\alpha + 5316\right), \\ \tilde{d}_3(\alpha) = & 14400\left(8\alpha^2 - 135\alpha - 1728\right), \\ \tilde{d}_4(\alpha) = & 180\left(17\alpha^2 - 442\alpha + 7740\right), \\ \tilde{d}_5(\alpha) = & 60\left(\alpha^2 + 16\alpha - 357\right), \\ \tilde{d}_6(\alpha) = & \alpha^2 - 30\alpha + 210. \end{cases}$$

- Pour les polynômes de degré 4, $\tilde{Q}_{4,\alpha}$ est défini par

$$\tilde{Q}_{4,\alpha}(\lambda) = -\frac{\tilde{B}_{4,\alpha}(\lambda)}{120(169344 + h^6\lambda^3 + 84h^4\lambda^2 + 5040h^2\lambda)}$$

avec

$$\begin{aligned}\tilde{B}_{4,\alpha}(\lambda) = \quad & \lambda^4 h^8(\alpha-24)+12\lambda^3 h^6(4\alpha+287)+1008\lambda^2 h^4(3\alpha-305)\\ & +20160\lambda h^2(8\alpha-165)+5080320(\alpha-1)\end{aligned}$$

De plus, nous avons

$$Q_{4,\alpha}(\lambda)=\sum_{i=0}^{8}\lambda^i h^{2i}\tilde{d}_i(\alpha).$$

avec

$$\begin{cases}\tilde{d}_0(\alpha) = & 25809651302400(\alpha-6)^2,\\ \tilde{d}_1(\alpha) = & 204838502400\left(8\alpha^2-357\alpha+1854\right),\\ \tilde{d}_2(\alpha) = & 81285120\left(698\alpha^2+3882\alpha+292185\right),\\ \tilde{d}_3(\alpha) = & 203212800\left(72\alpha^2-13791\alpha-328\right),\\ \tilde{d}_4(\alpha) = & 48384\left(719\alpha^2-12750\alpha+2419275\right),\\ \tilde{d}_5(\alpha) = & 8064\left(76\alpha^2-972\alpha-286209\right),\\ \tilde{d}_6(\alpha) = & 144\left(58\alpha^2-5282\alpha+201609\right),\\ \tilde{d}_7(\alpha) = & 24\left(4\alpha^2+241\alpha-6972\right),\\ \tilde{d}_8(\alpha) = & \alpha^2-48\alpha+55.\end{cases}$$

- Pour $p=5$, $\tilde{Q}_{p,\alpha}$ est tel que

$$\tilde{Q}_{5,\alpha}(\lambda)=\frac{\tilde{B}_{5,\alpha}(\lambda)}{210\left(39916800+\lambda^4h^8+128\lambda^3h^6+12960\lambda^2h^4+967680\lambda h^2\right)}$$

avec

$$\begin{aligned}\tilde{B}_{5,\alpha}(\lambda) = \quad & \lambda^5 h^{10}(\alpha-35)+70\lambda^4 h^8(\alpha+168)+6720\lambda^3 h^6(\alpha-351)\\ & +302400\lambda^2 h^4(2\alpha+385)+8467200\lambda h^2(5\alpha-303)+1676505600(\alpha-1)\end{aligned}$$

et, on a

$$Q_{5,\alpha}(\lambda)=\sum_{i=0}^{10}\lambda^i h^{2i}\tilde{d}_i(\alpha)$$

avec

$$\begin{cases} \tilde{d}_0(\alpha) = & 2810671026831360000\,(\alpha-15)^2\,, \\ \tilde{d}_1(\alpha) = & 28390616432640000\,(5\alpha+168)\,(\alpha-15)\,, \\ \tilde{d}_2(\alpha) = & 10241925120000\left(373\alpha^2-35067\alpha+855423\right), \\ \tilde{d}_3(\alpha) = & 3072577536000\left(24\alpha^2-895\alpha-159240\right), \\ \tilde{d}_4(\alpha) = & 1016064000\left(1151\alpha^2-11360\alpha+24379995\right), \\ \tilde{d}_5(\alpha) = & 67737600\left(257\alpha^2-3570\alpha-9096540\right), \\ \tilde{d}_6(\alpha) = & 2822400\left(76\alpha^2-2417\alpha+2988895\right), \\ \tilde{d}_7(\alpha) = & 67200\left(32\alpha^2+2383\alpha-974820\right), \\ \tilde{d}_8(\alpha) = & 140\left(131\alpha^2-37062\alpha+2285235\right), \\ \tilde{d}_9(\alpha) = & 140\left(\alpha^2+151\alpha-5912\right), \\ \tilde{d}_{10}(\alpha) = & \alpha^2-70\alpha+1190. \end{cases}$$

2.C Preuve du théorème 2.3.1

Cette section est consacrée à la preuve du théorème suivant

Théorème 4.1 *Pour tout* $\mathbf{m} = (m_k)_{k=1,\dots,3} \in \{1,\dots,p+1\}^3$ *et* $\mathbf{n} = (n_k)_{k=1,\dots,3} \in \{1,\dots,p+1\}^3$*, nous avons*

$$\begin{aligned} &1.\quad M_{3,p}(\mathbf{m},\mathbf{n}) = \prod_{i=1}^{3} M_{1,p}(m_i,n_i)\,, \\ &2.\quad K_{3,p}(\mathbf{m},\mathbf{n}) = \sum_{i=1}^{3}\left(K_{1,p}(m_i,n_i)\prod_{k=1,\,k\neq i}^{3} M_{1,p}(m_k,n_k)\right) \\ &3.\quad K^C_{3,p}(\mathbf{m},\mathbf{n}) = K^W_{1,p}(m_{p_C},n_{p_C})\prod_{k=1,\,k\neq p_C}^{3} M_{1,p}(m_k,n_k) \qquad (2.33) \\ &4.\quad N_{3,p}(\mathbf{m},\mathbf{n}) = \sum_{p=1}^{3} N_{1,p}(m_p,n_p)\prod_{k=1,\,k\neq p}^{3}\delta_{m_k,n_k} \\ &5.\quad N^C_{3,p}(\mathbf{m},\mathbf{n}) = N^W_{1,p}(m_{p_C},n_{p_C})\prod_{k=1,\,k\neq p_C}^{3}\delta_{m_k,n_k} \end{aligned}$$

où $p_C = \begin{cases} 1 & \text{si } C\in\{E,W\}\,, \\ 2 & \text{si } C\in\{N,S\}\,, \\ 3 & \text{si } C\in\{B,F\}\,, \end{cases}$ *et* $N_{1,p} = M_{1,p}^{-1}K_{1,p}$ *et* $N^C_{1,p} = M_{1,p}^{-1}K^C_{1,p}$.

Démonstration.

- *Preuve de 1.*
En considérant les notations et les résultats de la section 2.3, nous avons

$$\begin{aligned} M_{3,p}(\mathbf{m},\mathbf{n}) &= h^3 \int_{\hat{K}} \hat{\varphi}_{\mathbf{m}} \hat{\varphi}_{\mathbf{n}} d\mathbf{x} \\ &= h^3 \int_{\hat{K}} \prod_{i=1}^{3} \hat{\varphi}_{m_i}(x_i) \prod_{i=1}^{3} \hat{\varphi}_{n_i}(x_i) \prod_{i=1}^{3} dx_i \\ &= \prod_{i=1}^{3} h \int_{[0,1]} \hat{\varphi}_{m_i}(x_i) \hat{\varphi}_{n_i}(x_i) dx_i \\ &= \prod_{i=1}^{3} M_{1,p}(m_i,n_i) \end{aligned}$$

- *Preuve de 2.*
Tout d'abord, nous avons le lemme suivant pour le terme en volume.
Lemme 2.C.1. *Pour tout* $\mathbf{m} = (m_k)_{k=1,\dots,3} \in \{1,\dots,p+1\}^3$ *et* $\mathbf{n} = (n_k)_{k=1,\dots,3} \in \{1,\dots,p+1\}^3$*, nous avons*

$$h \int_{\hat{K}} \nabla\hat{\varphi}_{\mathbf{m}} \cdot \nabla\hat{\varphi}_{\mathbf{n}} d\mathbf{x} = \sum_{i=1}^{3} \left(\frac{1}{h} \int_{[0,1]} \frac{\partial\hat{\varphi}_{m_i}}{\partial x_i}(x_i) \frac{\partial\hat{\varphi}_{n_i}}{\partial x_i}(x_i) dx_i \prod_{k=1,\, k\neq i}^{3} M_{1,p}(m_k,n_k) \right).$$

Démonstration. Nous savons que $\forall \mathbf{m} \in \{1,\dots,p+1\}^d$

$$\hat{\varphi}_{\mathbf{m}}(\mathbf{x}) = \prod_{k=1}^{3} \hat{\varphi}_{m_p}(x_p)$$

ce qui implique que

$$\frac{\partial\hat{\varphi}_{\mathbf{m}}}{\partial x_k}(\mathbf{x}) = \frac{\partial\hat{\varphi}_{m_k}}{\partial x_k}(x_k) \prod_{i=1,\, i\neq k}^{3} \hat{\varphi}_{m_i}(x_i).$$

Alors, en utilisant le même raisonnement que précédemment,

$$\begin{aligned} h \int_{\hat{K}} \nabla\hat{\varphi}_{\mathbf{m}} \cdot \nabla\hat{\varphi}_{\mathbf{n}} d\mathbf{x} &= \sum_{i=1}^{3} \left(\frac{1}{h} \int_{[0,1]} \frac{\partial\hat{\varphi}_{m_i}}{\partial x_i}(x_i) \frac{\partial\hat{\varphi}_{n_i}}{\partial x_i}(x_i) dx_i \prod_{k=1,\, k\neq i}^{3} h \int_{[0,1]} \hat{\varphi}_{m_k}(x_k) \hat{\varphi}_{n_k}(x_k) dx_k \right) \\ &= \sum_{i=1}^{3} \left(\frac{1}{h} \int_{[0,1]} \frac{\partial\hat{\varphi}_{m_i}}{\partial x_i}(x_i) \frac{\partial\hat{\varphi}_{n_i}}{\partial x_i}(x_i) dx_i \prod_{k=1,\, k\neq i}^{3} M_{1,p}(m_k,n_k) \right) \end{aligned}$$

ce qui achève la preuve. □

Maintenant, nous avons à considérer les termes surfaciques.
Remarquons tout d'abord que, sur toutes les faces Γ^C,

$$\nabla\hat{\varphi}_{\mathbf{m}} \cdot \nu|_{\Gamma^C} = \frac{\partial\hat{\varphi}_{m_{p_C}}}{\partial x_{p_C}}(x_{p_C}) \nu_{1,C} \prod_{k=1,\, k\neq p_C}^{3} \hat{\varphi}_{m_k}(x_k) \tag{2.34}$$

où $\nu_{1,C}$ est le vecteur normal unitaire extérieur dans le cas mono dimensionnel défini par

$$\nu_{1,C} = \begin{cases} 1 & \text{si } C \in \{E, N, F\}, \\ -1 & \text{si } C \in \{W, S, B\} \end{cases}$$

et x_{p_C} est défini par

$$x_{p_C} = \begin{cases} 1 & \text{si } C \in \{E, N, F\}, \\ 0 & \text{si } C \in \{W, S, B\}. \end{cases}$$

Alors, nous pouvons énoncer le lemme suivant.

Lemme 2.C.2. *Pour tout $C \in \{E, W, N, S, F, B\}$, nous avons $\forall \mathbf{m}, \mathbf{n} \in \{1, \ldots, p+1\}^d$*

$$h \int_{\Gamma^C} \hat{\varphi}_{\mathbf{m}} \left(\nabla \hat{\varphi}_{\mathbf{n}} \cdot \nu\right) d\sigma = \frac{1}{h} \hat{\varphi}_{m_{p_C}}(x_{p_C}) \frac{\partial \hat{\varphi}_{n_{p_C}}}{\partial x_{p_C}}(x_{p_C}) \nu_{1,C} \prod_{k=1,\, k \neq p_C}^{3} M_{1,p}(m_k, n_k)$$

et

$$h^2 \int_{\Gamma^C} \gamma \hat{\varphi}_{\mathbf{m}} \hat{\varphi}_{\mathbf{n}} d\sigma = \gamma \hat{\varphi}_{m_{p_C}}(x_{p_C}) \hat{\varphi}_{n_{p_C}}(x_{p_C}) \prod_{k=1,\, k \neq p_C}^{3} M_{1,p}(m_k, n_k).$$

Démonstration. Tout d'abord, en utilisant (2.34), on a

$$\begin{aligned} h \int_{\Gamma^C} \hat{\varphi}_{\mathbf{m}} \left(\nabla \hat{\varphi}_{\mathbf{n}} \cdot \nu\right) d\sigma &= h \int_{\Gamma^C} \left(\prod_{k=1}^{3} \hat{\varphi}_{m_k}(x_k) \right) \left(\frac{\partial \hat{\varphi}_{n_{p_C}}}{\partial x_{p_C}}(x_{p_C}) \nu_{1,C} \prod_{k=1,\, k \neq p_C}^{3} \hat{\varphi}_{n_k}(x_k) \right) \prod_{k=1, k \neq p_C}^{3} dx_k \\ &= \frac{1}{h} \hat{\varphi}_{m_{p_C}}(x_{p_C}) \frac{\partial \hat{\varphi}_{n_{p_C}}}{\partial x_{p_C}}(x_{p_C}) \nu_{1,C} \prod_{k=1,\, k \neq p_C}^{3} h^2 \int_{[0,1]} \hat{\varphi}_{m_k}(x_k) \hat{\varphi}_{n_k}(x_k) dx_k \end{aligned}$$

qui peut être réécrit comme

$$\int_{\Gamma^C} \hat{\varphi}_{\mathbf{m}} \left(\nabla \hat{\varphi}_{\mathbf{n}} \cdot \nu\right) d\sigma = \hat{\varphi}_{m_{p_C}}(x_{p_C}) \frac{\partial \hat{\varphi}_{n_{p_C}}}{\partial x_{p_C}}(x_{p_C}) \nu_{1,C} \prod_{k=1,\, k \neq p_C}^{3} M_{1,p}(m_k, n_k).$$

De la même manière, pour le terme de pénalisation, nous avons :

$$\begin{aligned} h^2 \int_{\Gamma^C} \gamma \hat{\varphi}_{\mathbf{m}} \hat{\varphi}_{\mathbf{n}} d\sigma &= \hat{\varphi}_{m_{p_C}}(x_{p_C}) \hat{\varphi}_{n_{p_C}}(x_{p_C}) h^2 \int_{\Gamma^C} \gamma \prod_{k=1,\, k \neq p_C}^{3} \hat{\varphi}_{m_k}(x_k) \prod_{k=1,\, k \neq p_C}^{3} \hat{\varphi}_{n_k}(x_k) \prod_{k=1,\, k \neq p_C}^{3} d\sigma_k \\ &= \gamma \hat{\varphi}_{m_{p_C}}(x_{p_C}) \hat{\varphi}_{n_{p_C}}(x_{p_C}) \prod_{k=1,\, k \neq p_C}^{3} h^2 \int_{[0,1]} \hat{\varphi}_{m_k}(x_k) \hat{\varphi}_{n_k}(x_k) dx_k \end{aligned}$$

ce qui implique clairement que

$$\int_{\Gamma^C} \gamma \hat{\varphi}_{\mathbf{m}} \hat{\varphi}_{\mathbf{n}} d\sigma = \gamma \hat{\varphi}_{m_{p_C}}(x_{p_C}) \hat{\varphi}_{n_{p_C}}(x_{p_C}) \prod_{k=1,\, k \neq p_C}^{3} M_{1,p}(m_k, n_k)$$

ce qui achève la preuve. □

Finalement, en utilisant les deux lemmes, nous obtenons

$$K_{3,p}(\mathbf{m},\mathbf{n}) = \sum_{i=1}^{3}\left(K_{1,p}(m_i,n_i)\prod_{k=1,\,k\neq i}^{3}M_{1,p}(m_k,n_k)\right). \tag{2.35}$$

- *Preuve de 3.*
  Pour réécrire les termes $K^C_{3,p}(\mathbf{m},\mathbf{n})$ pour tout $\mathbf{m},\ \mathbf{n}$, nous utilisons un raisonnement similaire à celui pour $K_{3,p}$.
- *Preuve de 4.*
  Pour prouver 4. et 5., nous avons besoin du lemme suivant.
  Lemme 2.C.3. *Soit* $\mathbf{m} = (m_p)_{p=1,\dots,3}$ *et* $\mathbf{n} = (n_p)_{p=1,\dots,3}$.
  Nous avons

$$M^{-1}_{3,p}(\mathbf{m},\mathbf{n}) = \prod_{k=1}^{3}M^{-1}_{1,p}(m_k,n_k).$$

Démonstration. Soit A la matrice définie $\forall \mathbf{m},\mathbf{n}\in\{1,\dots,p+1\}^d$ par

$$A(\mathbf{m},\mathbf{n}) = \prod_{k=1}^{3}M^{-1}_{1,p}(m_k,n_k).$$

Nous avons

$$\begin{aligned}(AM)(\mathbf{m},\mathbf{n}) &= \sum_{l_1,\dots,l_3=1}^{p+1}\left(\prod_{k=1}^{3}M^{-1}_{1,p}(m_k,l_k)\prod_{k=1}^{3}M_{1,p}(l_k,n_k)\right)\\ &= \sum_{l_1,l_2=1}^{p+1}\left(\prod_{k=1}^{2}M^{-1}_{1,p}(m_k,l_k)\prod_{k=1}^{2}M_{1,p}(l_k,n_k)\sum_{l_3=1}^{3}M^{-1}_{1,p}(m_3,l_3)M_{1,p}(l_3,n_3)\right)\\ &= \prod_{k=1}^{3}\left(\sum_{l_k=1}^{p+1}M^{-1}_{1,p}(m_k,l_k)M_{1,p}(l_k,n_k)\right).\end{aligned}$$

Mais,

$$\sum_{l_k=1}^{p+1}M^{-1}_{1,p}(m_k,l_k)M_{1,p}(l_k,n_k) = \left(M^{-1}_{1,p}M_{1,p}\right)(m_k,n_k) = \delta_{m_k,n_k}$$

ainsi

$$(AM)(\mathbf{m},\mathbf{n}) = \prod_{k=1}^{3}\delta_{m_k,n_k} = I(\mathbf{m},\mathbf{n})$$

où I est la matrice identité, ce qui termine la preuve. □

Considérons maintenant la matrice $N_{3,p} = M^{-1}_{3,p}K_{3,p}$. Premièrement, nous réécrivons $K_{3,p}$ comme

$$K_{3,p} = \sum_{q=1}^{3}T_q,$$

avec $T_q(\mathbf{m},\mathbf{n}) = K_{1,p}(m_q,n_q)\prod_{k=1,\,k\neq q}^{3} M_{1,p}(m_k,n_k)$.

Alors, $M_{3,p}^{-1}K_{3,p} = \sum_{q=1}^{3} M_{3,p}^{-1}T_q$ et, en utilisant le lemme 2.C.3

$$\begin{aligned}
\left(M_{3,p}^{-1}T_1\right)(\mathbf{m},\mathbf{n}) &= \sum_{l_1,\ldots,l_3=1}^{p+1}\left(\prod_{k=1}^{3} M_{1,p}^{-1}(m_k,l_k)\,K_{1,p}(l_1,n_1)\prod_{k=2}^{3} M_{1,p}(l_k,n_k)\right)\\
&= \sum_{l_1,\ldots,l_3=1}^{p+1}\Big(M_{1,p}^{-1}(m_1,l_1)\,K_{1,p}(l_1,n_1)\,M_{1,p}^{-1}(m_2,l_2)\\
&\qquad \times M_{1,p}(l_2,n_2)\,M_{1,p}^{-1}(m_3,l_3)\,M_{1,p}(l_3,n_3)\Big)\\
&= \sum_{l_1=1}^{p+1} M_{1,p}^{-1}(m_1,l_1)\,K_{1,p}(l_1,n_1)\prod_{k=2}^{3}\left(\sum_{l_k=1}^{p+1} M_{1,p}^{-1}(m_k,l_k)\,M_{1,p}(l_k,n_k)\right).
\end{aligned}$$

Alors,

$$\left(M_{3,p}^{-1}T_1\right)(\mathbf{m},\mathbf{n}) = N_{1,p}(m_1,n_1)\,\delta_{m_2,n_2}\delta_{m_3,n_3}.$$

En effectuant les mêmes calculs pour T_2 et T_3, on obtient

$$N_{3,p}(\mathbf{m},\mathbf{n}) = \sum_{p=1}^{3} N_{1,p}(m_p,n_p)\prod_{k=1,\,k\neq p}^{3}\delta_{m_k,n_k}.$$

- *Preuve de 5.*
  Nous appliquons la technique utilisée pour prouver 4. pour montrer que

$$N_{3,p}^{C}(\mathbf{m},\mathbf{n}) = N_{1,p}^{W}(m_{p_C},n_{p_C})\prod_{k=1,\,k\neq p_C}^{3}\delta_{m_k,n_k}$$

□

2.D Preuve du lemme 2.2.1

Nous définissons les fonctions symétriques des racines :

$$\begin{cases}
\sigma_0^n = \prod_{i=1}^{n}\lambda_i & \\
\sigma_p^n = \sum_{i_1=1}^{n}\left[\sum_{i_2>i_1}^{n}\cdots\sum_{i_{p+1}>i_p}^{n}\ \prod_{\substack{j=1\\ j\neq i_p,\ \forall p\in\{1,\ldots,p+1\}}}^{n}\lambda_k\right] & \text{si } 1\leq p\leq n-2\\
\sigma_{n-1}^n = \sum_{i=1}^{n}\lambda_i &
\end{cases}$$

où λ_i, $i = 1, \ldots, n$ représente les racines du polynôme P.
Nous pouvons remarquer que $c_i = (-1)^i \sigma_i^n$ et nous devons alors montrer que

$$\lambda_i \geq 0 \Longleftrightarrow \sigma_i^n \geq 0, \ \forall i$$

– Il est clair que si toutes les racines λ_i sont positives, alors pour tout i,

$$\sigma_i^n \geq 0.$$

– A présent, supposons que pour tout $i \in \{0, \ldots, n-1\}$, $\sigma_i^n \geq 0$.
En raisonnant par l'absurde, nous supposons qu'il existe au moins une racine strictement négative, que nous notons λ.
Dans un premier temps, supposons que n est pair. On a

$$P(\lambda) = \lambda^n + \sum_{i=0}^{n-1} (-1)^{i+1} \sigma_{n-1-i} \lambda^{n-1-i} = 0.$$

Si i est pair alors $(-1)^{i+1} < 0$ et $\lambda^{n-1+i} \leq 0$ et dans le cas où i est impair on a $(-1)^{i+1} > 0$ et $\lambda^{n-1+i} \geq 0$. Dans les deux cas chaque terme de la somme est postif ce qui implique que λ est une racine multiple d'ordre n nulle.
Le raisonnement est identique pour le cas n impair et on arrive à la conclusion que tous les termes de la somme sont négatifs et que λ est une racine multiple d'ordre n nulle.
On a ainsi obtenu une contradiction avec l'hypothèse initiale donc toutes les racines λ_i sont positives.

Chapitre 3

Étude numérique de la stabilité de la méthode IPDG sur des maillages triangulaires

Dans le chapitre précédent, nous avons réalisé une étude de stabilité de la méthode IPDG sur des maillages réguliers. Cette étude nous a permis d'obtenir des conditions de stabilité L^2 confirmant la conjecture de Ainsworth, Monk and Muniz (cf. [2]) jusqu'à un degré polynomial égal à 5. Puis nous avons étudié le comportement de la condition CFL en fonction du coefficient de pénalisation de la méthode IPDG. Les résultats obtenus pour des maillages rectangulaires ou parallélépipédiques montrent qu'il serait préférable d'utiliser, dans le choix de ce paramètre, le rayon du cercle (en 2D) ou de la sphère (en 3D) inscrit dans chaque élément du maillage (rectangle ou parallélépipède).
Le but de ce chapitre est d'étudier quel est le meilleur choix du paramètre ξ_F en se basant sur des maillages triangulaires périodiques uniformes. Dans la section 3.1, nous comparerons les deux choix de ξ_F et nous montrerons que le rayon du cercle inscrit est plus adapté que le rayon du cercle circonscrit. Nous remarquerons que ce choix n'est pas optimal et nous l'améliorerons à la section 3.2 en considérant des paramètres géomètriques supplémentaires. Enfin, on s'attachera dans la section 3.3 à analyser l'influence des différents choix de ξ_F sur le comportement de la condition CFL en considérant, cette fois-ci, des maillages non-uniformes.

3.1 Comparaison des différents choix de ξ_F

Les remarques du chapitre précédent (cf. section 2.3.3) suggèrent que ξ_F devrait être le rayon du cercle inscrit (en 2D) et celui de la sphère inscrite (en 3D). Cependant, les résultats ont été obtenus pour des maillages réguliers avec des mailles rectangulaires ou parallélépipédiques. Le but de cette section est d'étendre l'analyse au cas de maillages triangulaires. Comme nous l'avons vu au chapitre 2, ξ_F devrait être choisi de telle sorte que α soit indépendant du maillage, ou le moins dépendant du maillage possible. Pour déterminer quel est le choix de ξ_F le plus approprié, nous proposons de considérer des maillages réguliers construits à partir d'un triangle de référence selon une approche que nous préciserons plus loin. Soit K_0 le triangle de sommets $A = (0, 0)$, $B = (1, 0)$, $C = (x_1, y_1)$ (cf. Fig. 3.1), avec $y_1 > 0$. La coordonnée x_1 détermine la nature de K_0 : triangle rectangle pour $x_1 = 0$, isocèle pour $x_1 = 0.5$ et quelconque pour toute autre valeur. Nous définissons aussi K_0^*, le triangle de sommets C, B, $D = (1 + x_1, y_1)$. Comme $K_0 \cup K_0^*$ est un parallélogramme, nous pouvons facilement construire un maillage périodique uniforme par

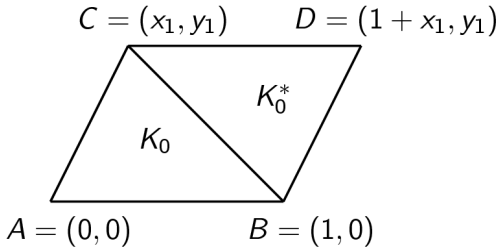

FIGURE 3.1 – K_0 et K_0^*

translation de $K_0 \cup K_0^*$ le long des axes $(1, 0)$ et (x_1, y_1) (cf. Fig. 3.2).
Pour mener l'étude, nous avons tout d'abord considéré des maillages infinis et, en utilisant une

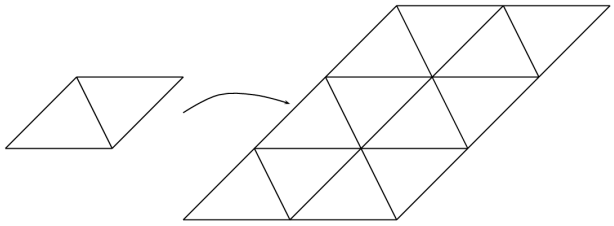

FIGURE 3.2 – Construction d'un maillage périodique

transformée de Fourier suivant les directions $(1, 0)$ et (x_1, y_1), nous avons effectué une analyse similaire à celle du chapitre 2. Nous n'avons cependant pas été capables de calculer une condition CFL et nous n'avons pu exprimer qu'une condition nécessaire de stabilité sur le paramètre γ_1. De plus, nous n'avons pas réussi à exprimer ce paramètre comme une fonction de (x_1, y_1) ou d'une quelconque grandeur géométrique caractéristique de K_0. Pour illustrer ces propos, nous donnons dans l'annexe 3.A le détail des calculs pour des triangles rectangles et pour des triangles isocèles. Nous avons alors décidé de calculer le paramètre de stabilisation minimal numériquement au lieu de calculer son expression analytique. Pour cela, nous avons considéré des maillages finis composés de trente triangles dans les deux directions $(1, 0)$ et (x_1, y_1) avec une source ponctuelle en espace placée au centre du maillage en $x_0 = (0, 0.5)$. L'expression de la source en temps est donnée par

$$f(t) = 2\lambda \left(\lambda (t - t_0)^2 - 1\right) e^{-\lambda (t-t_0)^2},$$

avec $\lambda = \pi^2 f_0^2$, $f_0 = 5$ et $t_0 = 1/f_0$.
Le paramètre minimal, $\gamma_{\min}$, que nous définissons comme le plus petit paramètre stabilisant le schéma IPDG, a été calculé par dichotomie en utilisant l'algorithme 1. Le principe de cet algorithme consiste à résoudre l'équation des ondes pour un paramètre donné puis à diminuer ce paramètre si le schéma est stable ou à l'augmenter sinon.

Pour déterminer si le schéma explose ou pas, nous avons calculé à chaque pas de temps l'énergie

$$E^{n+\frac{1}{2}} = \left(M \frac{U^{n+1} - U^n}{\Delta t}, \frac{U^{n+1} - U^n}{\Delta t}\right) + \left(KU^n, U^{n+1}\right).$$

Si elle n'explose pendant 10000 itérations, *i.e.* si $E^{n+\frac{1}{2}}$ ne dépasse pas $10e16$, le schéma est considéré stable, autrement il est considéré instable.

Algorithme 1

1: $\Delta t = 4.3\ 10^{-4}$, $\gamma_1 = 1$ et $\gamma_2 = 5$
2: $\gamma = \dfrac{\gamma_1 + \gamma_2}{2}$
3: On calcule la solution avec Δt et γ après 10000 itérations
4: **si** explosion **alors**
5: $\quad \gamma_1 = \gamma$
6: **sinon**
7: $\quad \gamma_2 = \gamma$
8: **finsi**
9: **si** $|\gamma - \dfrac{\gamma_1 + \gamma_2}{2}| < 10^{-5}$ **alors**
10: $\quad \gamma_{\min} = \gamma_2$
11: **sinon**
12: $\quad$ Retour en 2.
13: **finsi**

Le pas de temps Δt a été choisi suffisamment petit de telle sorte que seul le paramètre de pénalisation puisse induire des instabilités. En rappelant que l'on cherche à exprimer $\gamma_{\min}$ sous la forme

$$\gamma_{\min} = \frac{\alpha_{\min}}{\xi_F}$$

avec $\alpha_{\min}$ indépendant de la géométrie du maillage, le choix de ξ_F le plus approprié devrait être tel que $\gamma_{\min}\xi_F$ est indépendant de la géométrie du maillage et donc de y_1 et de x_1. Dans un premier temps, on considère les trois configurations

- configuration 1 : K_0 est un triangle rectangle, $x_1 = 0$,
- configuration 2 : K_0 est un triangle isocèle, $x_1 = 0.5$,
- configuration 3 : K_0 est un triangle quelconque, $x_1 = 0.75$

et nous calculons $\gamma_{\min}$ pour différentes valeurs de y_1 en utilisant des éléments P^1. Sur la Fig. 3.3 (resp. 3.4 et 3.5), nous représentons les valeurs de $\gamma_{\min}\rho_{\text{ins}}$ (courbe noire) et $\gamma_{\min}\rho_{\text{circ}}$ (courbe rouge en pointillés) comme des fonctions de y_1 pour $x_1 = 0$ (resp. $x_1 = 0.5$ et $x_1 = 0.75$). Il est clair que $\gamma_{\min}\rho_{\text{circ}}$ dépend fortement de y_1 et par conséquent de la géométrie du triangle de référence, alors que $\gamma_{\min}\rho_{\text{ins}}$ semble être beaucoup plus indépendant de y_1 et x_1. De plus, $\gamma_{\min}\rho_{\text{circ}}$ tend vers l'infini quand y_1 tend vers l'infini alors que $\gamma_{\min}\rho_{\text{ins}}$ reste borné.

Nous obtenons des résultats similaires pour des éléments P^2 (cf. Fig. 3.6-3.8) et pour des éléments P^3 (cf. Fig. 3.9-3.11). Par conséquent, nous concluons que $\xi_F = \rho_{\text{ins}}$ est un choix plus approprié que $\xi_F = \rho_{\text{circ}}$. Cependant, il est clair sur toutes les figures (en particulier sur les Fig. 3.9-3.11) que $\gamma_{\min}\rho_{\text{ins}}$ dépend légérement de y_1. Cela signifie que le choix de ξ_F peut être amélioré en prenant en compte des paramètres supplémentaires.

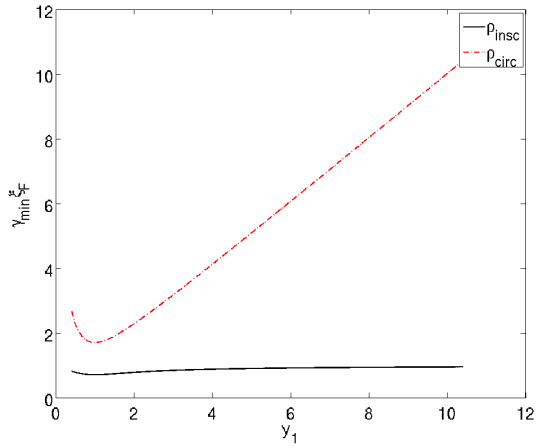

FIGURE 3.3 – $\gamma_{\min}\xi_F$ pour $x_1 = 0$ avec des éléments P^1

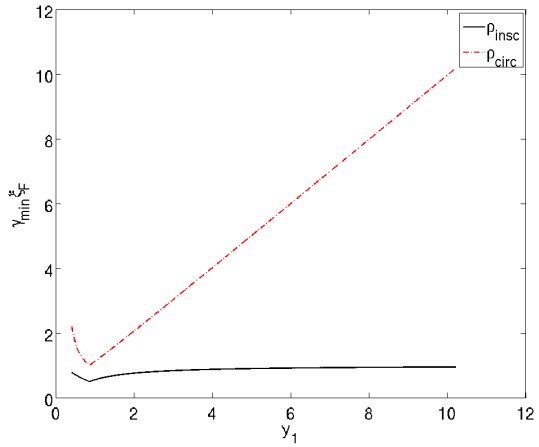

FIGURE 3.4 – $\gamma_{\min}\xi_F$ pour $x_1 = 0.5$ avec des éléments P^1

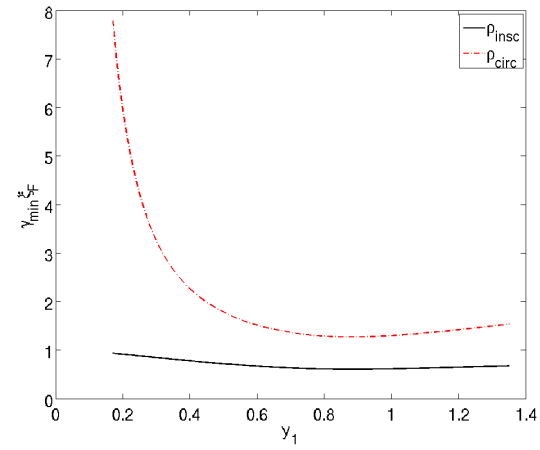

FIGURE 3.5 – $\gamma_{\min}\xi_F$ pour $x_1 = 0.75$ avec des éléments P^1

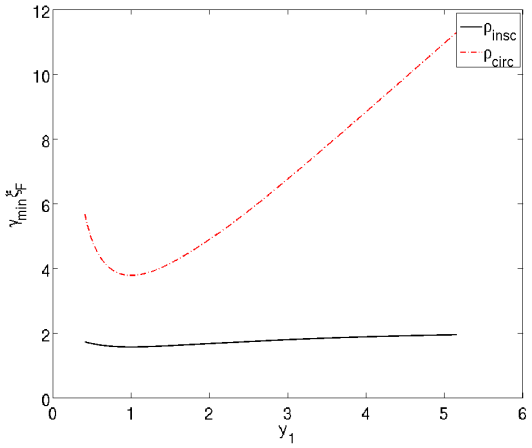

FIGURE 3.6 – $\gamma_{\min}\xi_F$ pour $x_1 = 0$ avec des éléments P^2

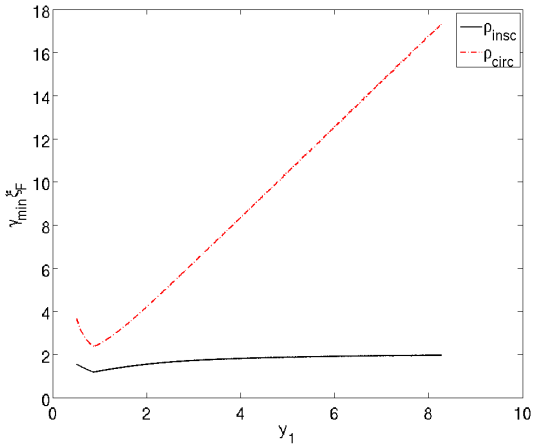

FIGURE 3.7 – $\gamma_{\min}\xi_F$ pour $x_1 = 0.5$ avec des éléments P^2

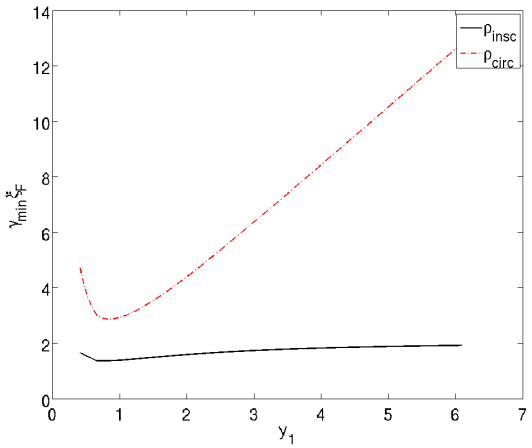

FIGURE 3.8 – $\gamma_{\min}\xi_F$ pour $x_1 = 0.75$ avec des éléments P^2

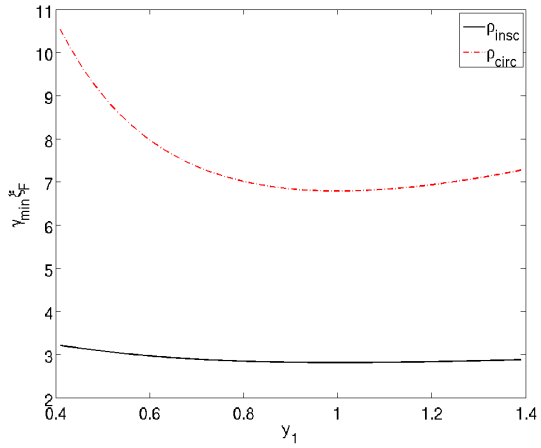

FIGURE 3.9 – $\gamma_{\min}\xi_F$ pour $x_1 = 0$ avec des éléments P^3

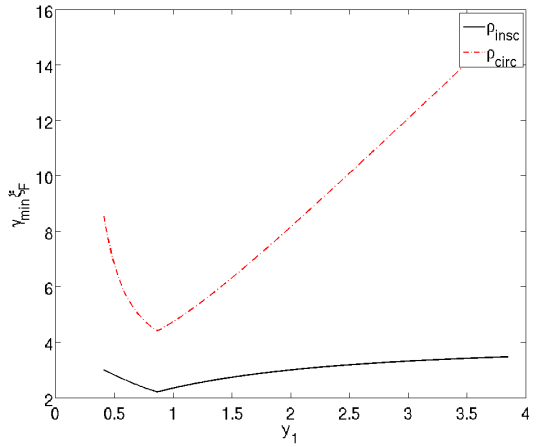

FIGURE 3.10 – $\gamma_{\min}\xi_F$ pour $x_1 = 0.5$ avec des éléments P^3

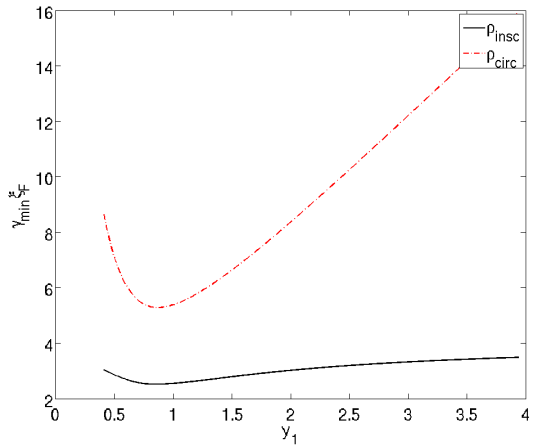

FIGURE 3.11 – $\gamma_{\min}\xi_F$ pour $x_1 = 0.75$ avec des éléments P^3

3.2 Amélioration du choix de ξ_F

Comme nous l'avons dit précédemment, le produit $\gamma_{\min} \times \rho_{\text{ins}}$ n'est pas constant et dépend légèrement de x_1 et y_1. Nous avons tout d'abord tenté d'exprimer ce produit comme une fonction de x_1 et y_1, mais nous n'avons pas trouvé de relation satisfaisante. Nous avons alors essayé de trouver un autre paramètre géométrique qui pourrait mieux décrire la variation de $\gamma_{\min}\rho_{\text{ins}}$. Il est apparu que l'angle minimum de K_0, que nous noterons $\theta_{\min}$, est le plus approprié. Nous représentons dans un premier temps sur la Fig. 3.12 $\gamma_{\min}\rho_{\text{ins}}$ comme une fonction de $\theta_{\min}$ pour des éléments P^1 et pour $x_1 = 0$ (courbe noire), $x_1 = 0.5$ (courbe en points rouges) et $x_1 = 0.75$ (courbe bleue en pointillés).
Même si la variation de $\gamma_{\min}\rho_{\text{ins}}$ n'est pas aussi importante que la variation de $\gamma_{\min}\rho_{\text{circ}}$, elle est de l'ordre de 100% (de 0.5 à 1), ce qui reste élevé.

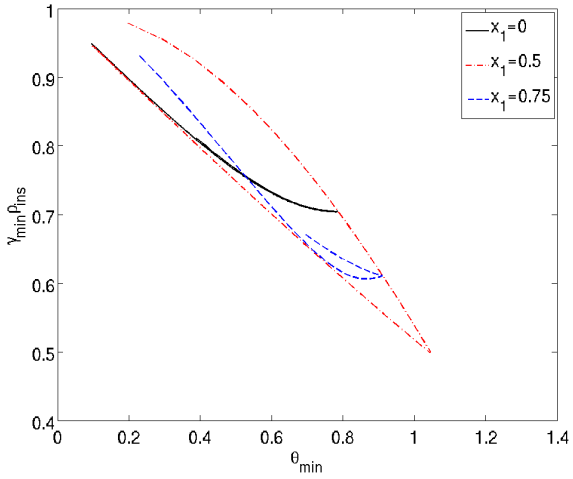

FIGURE 3.12 – $\gamma_{\min}\rho_{\text{ins}}$ comme fonction de $\theta_{\min}$ en P^1

Remarquons que chaque graphe est en fait composé de deux branches (pour le triangle rectangle, ces deux branches sont superposées). Pour expliquer ce phénomène, nous supposons, sans perte de généralité, que $x_1 \leq 0.5$. L'angle minimum est l'angle $\widehat{ABC}$ (cf. premier triangle de la Fig. 3.13) pour de petits y_1 et il s'agit de l'angle $\widehat{ACB}$ pour de grands y_1 (cf. deuxième triangle de la Fig. 3.13). Le point critique, joignant les deux branches, est obtenu quand y_1 est tel que $\widehat{ACB} = \widehat{ABC}$ (cf. troisième triangle de la Fig. 3.13). Pour les triangles rectangles ($x_1 = 0$), ce point correspond à $y_1 = 1$, *i.e.* K_0 est un triangle rectangle isocèle.
Concentrons nous à présent sur le cas des triangles isocèles. Il est clair que le point critique correspond au cas où K_0 est un triangle équilatéral et $y_1 = \dfrac{\sqrt{3}}{2}$. La branche supérieure est obtenue pour $y_1 < \dfrac{\sqrt{3}}{2}$. Elle correspond aux triangles isocèles dont l'angle relatif à la base est plus petit que $\dfrac{\pi}{3}$. La branche inférieure est obtenue quand $y_1 > \dfrac{\sqrt{3}}{2}$. Elle correspond aux triangles isocèles dont l'angle relatif à la base est plus grand que $\dfrac{\pi}{3}$. La branche supérieure semble être un maximum de $\gamma_{\min}\xi_F$ pour toutes les configurations alors que la branche inférieure semble être un minimum.

Pour confirmer ce point, nous considérons cinq autres configurations qui sont représentées sur

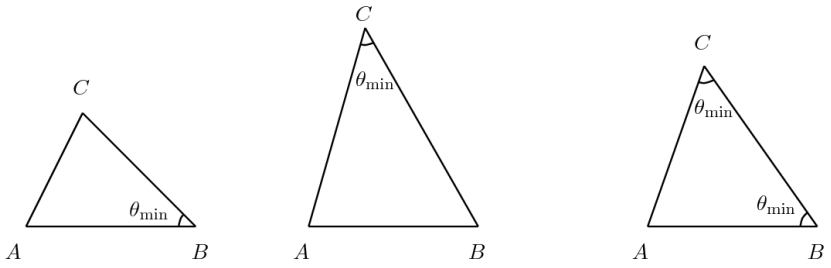

FIGURE 3.13 – Angle minimum en fonction de la nature du triangle

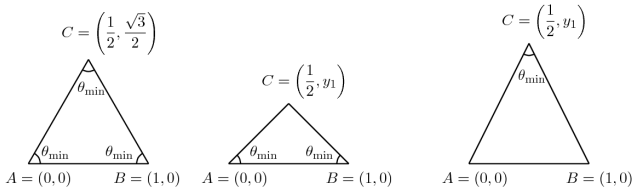

FIGURE 3.14 – Triangles isocèles pour $y_1 = \frac{\sqrt{3}}{2}$, $y_1 < \frac{\sqrt{3}}{2}$ et $y_1 > \frac{\sqrt{3}}{2}$

les Fig. 3.15 et 3.16.

- $x_1 = 0.66$ et représentée en bleu avec tirets et pointillés,
- $x_1 = 0.12$ et représentée en magenta avec tirets,
- $x_1 = 1.5$ et représentée en vert avec tirets,
- $x_1 = 1.3$ et représentée en cyan,
- $x_1 = 1.1$ et représentée en orange avec tirets et pointillés.

Notons que les trois dernières configurations correspondent à des triangles obtusangles ($x_1 > 1$) ce qui témoigne que notre analyse n'est pas restreinte au cas de triangles aigus. Il est clair sur les

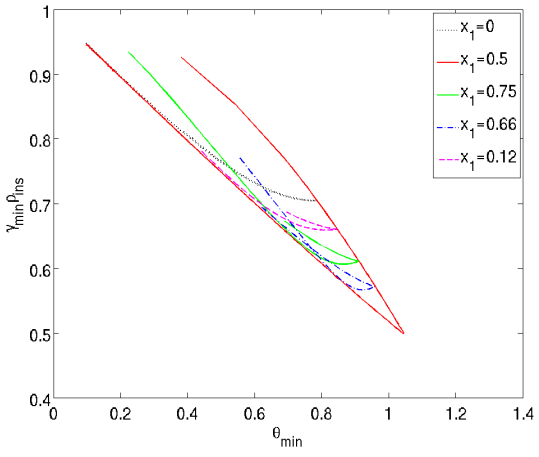

FIGURE 3.15 – $\xi_F = \rho_{ins}$ pour huit configurations en P^1 ($x_1 < 1$)

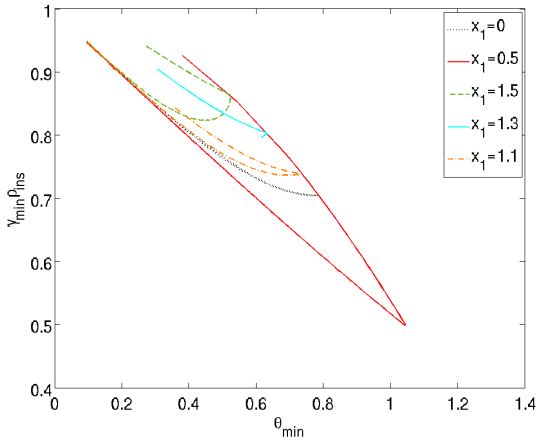

FIGURE 3.16 – $\xi_F = \rho_{ins}$ pour huit configurations en P^1 ($x_1 > 1$)

Fig. 3.15 et 3.16 que, quelle que soit la configuration, la branche supérieure de la courbe isocèle

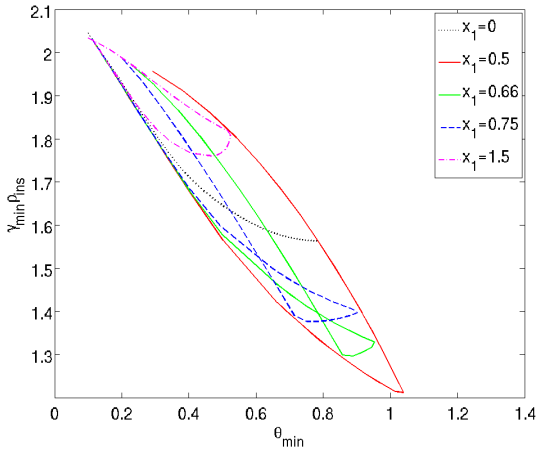

FIGURE 3.17 – $\xi_F = \rho_{ins}$ pour cinq configurations en P^2

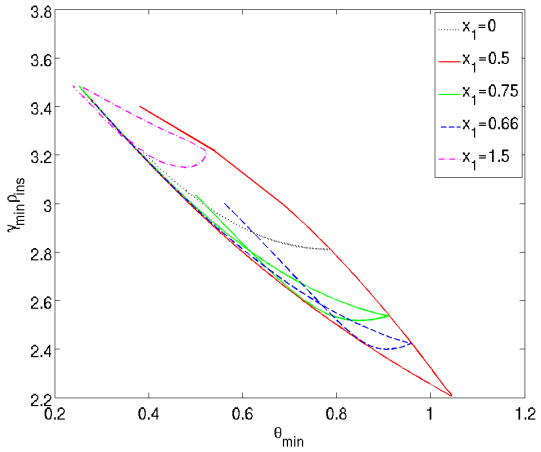

FIGURE 3.18 – $\xi_F = \rho_{ins}$ pour cinq configurations en P^3

est une borne supérieure de $\gamma_{\min}\rho_{\text{ins}}$ alors que la branche inférieure en est une borne inférieure. Nous avons effectué des expériences similaires pour des éléments P^2 (cf. Fig. 3.17) et P^3 (cf. Fig. 3.18) et nous avons obtenu la même conclusion.
Nous proposons donc maintenant de déterminer l'expression de ces deux branches comme des fonctions de $\theta_{\min}$ pour améliorer le choix de ξ_F. Cela sera l'objet des deux sous-sections suivantes, alors que la troisième sera dédiée à l'étude d'un troisième paramètre, l'angle maximal du triangle de référence.

3.2.1 Majoration de $\gamma_{\min}\rho_{\text{ins}}$

Pour approcher la branche supérieure de la courbe isocèle, nous avons essayé diverses approximations polynomiales mais les fonctions trigonométriques semblent être les plus adaptées. En effet, la fonction

$$f_{1,1} \ : \ \theta_{\min} \longmapsto f_{1,1}\left(\theta_{\min}\right) = \cos\left(\theta_{\min}\right),$$

fournit une approximation précise de cette branche. Sur la Fig. 3.19, nous représentons $f_{1,1}$ (courbe rouge) et la branche supérieure (courbe noire avec +), les deux courbes sont superposées. Sur la Fig. 3.20, nous traçons l'erreur relative entre les deux fonctions qui est plus petite que $3\,{}^{0}/_{00}$. Puisque $\gamma_{\min}$ est calculé numériquement avec trois chiffres de précision, $f_{1,1}$ est une approximation très proche de $\gamma_{\min}\rho_{\text{ins}}$ pour les triangles isocèles avec $y_1 < \dfrac{\sqrt{3}}{3}$.
$f_{1,1}$ est également une borne supérieure de $\gamma_{\min}\rho_{\text{ins}}$ pour toutes les configurations que nous avons

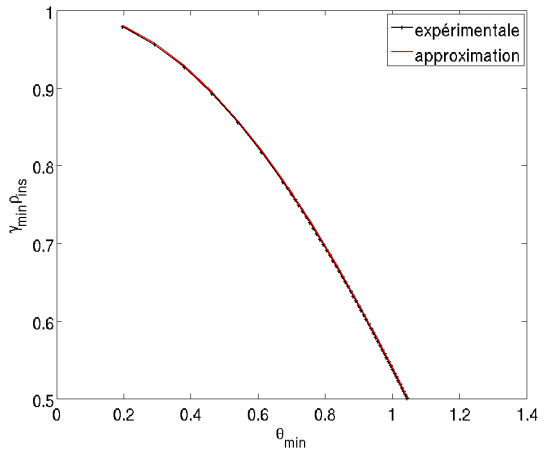

FIGURE 3.19 – Comparaison entre la branche supérieure et son approximation

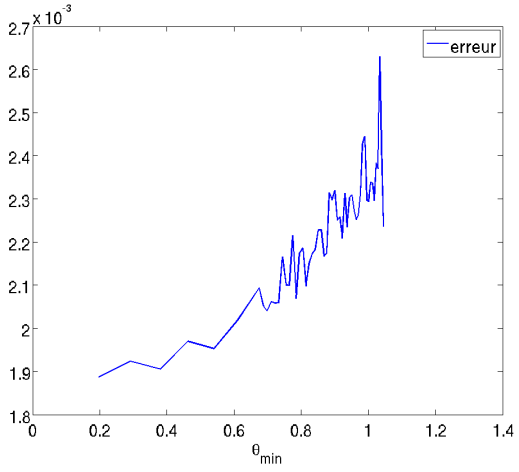

FIGURE 3.20 – Erreur relative entre la branche supérieure et son approximation

considérées. On a donc $\gamma_{\min}\rho_{\text{ins}} = f_{1,1}\left(\theta_{\min}\right)$ pour les triangles isocèles dont l'angle minimal est l'angle opposé à la base et $\gamma_{\min}\rho_{\text{ins}} \leq f_{1,1}\left(\theta_{\min}\right)$ pour tous les autres triangles. Ainsi,

$$\gamma_{\min} \leq \frac{f_{1,1}\left(\theta_{\min}\right)}{\rho_{\text{ins}}} \leq \frac{1}{\rho_{\text{ins}}}.$$

L'approximation $\frac{f_{1,1}(\theta_{\min})}{\rho_{\text{ins}}}$ est donc plus précise que l'approximation $\frac{1}{\rho_{\text{ins}}}$ et nous proposons donc d'utiliser

$$\xi_F = \frac{\rho_{\text{ins}}}{f_{1,1}(\theta_{\min})}. \tag{3.1}$$

Sur les Fig. 3.21 et 3.22, nous représentons $\frac{\gamma_{\min}\rho_{\text{ins}}}{f_{1,1}}$ pour les huit configurations que nous avons considérées. Cette quantité varie de 0.85 à 1 (18%) et $\frac{\rho_{\text{ins}}}{f_{1,1}}$ est alors un meilleur choix de ξ_F que ρ_{ins}, qui menait à des variations de 100%.

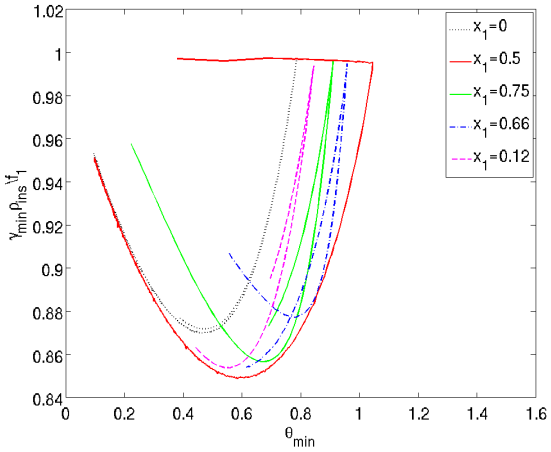

FIGURE 3.21 – $\xi_F = \rho_{ins}/f_{1,1}$ pour huit configurations en P^1 ($x_1 < 1$)

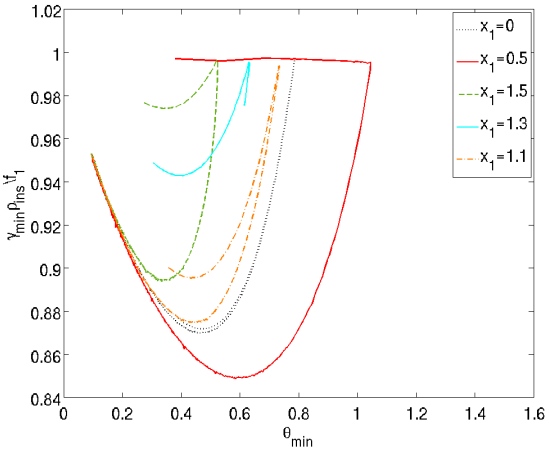

FIGURE 3.22 – $\xi_F = \rho_{ins}/f_{1,1}$ pour huit configurations en P^1 ($x_1 > 1$)

Pour des polynômes P^2 et P^3, nous avons trouvé, en utilisant la méthode des moindres carrés, qu'une bonne approximation de la branche supérieure est respectivement

$$f_{1,2} \ : \ \theta_{\min} \longmapsto f_{1,2}(\theta_{\min}) = 0.13 + 1.85\cos(\theta_{\min}) + 0.16\sin(\theta_{\min}),$$

et

$$f_{1,3} = : \ \theta_{\min} \longmapsto f_{1,3}(\theta_{\min}) = 0.29 + 3.21\cos(\theta_{\min}) + 0.36\sin(\theta_{\min}),$$

Sur la Fig, 3.23 (resp. Fig. 3.25), nous représentons $f_{1,2}$ (resp. $f_{1,3}$) et la branche supérieure pour de polynômes P^2 (resp. P^3). Cette fois encore, les courbes sont bien superposées. Sur la Fig. 3.24 (resp. Fig. 3.26), nous traçons l'erreur relative entre les deux fonctions pour des polynômes P^2 (resp. P^3).

Sur la Fig. 3.27 (resp. Fig. 3.28), nous représentons $\frac{\gamma_{\min}\rho_{\text{ins}}}{f_{1,2}}$ (resp. $\frac{\gamma_{\min}\rho_{\text{ins}}}{f_{1,3}}$) pour les huit configurations avec des polynômes P^2 (resp. P^3). Cette quantité varie de 0.84 à 1, *i.e.* 19% (resp. de 0.89 à 1, *i.e.* 12%). Ce choix est déjà satisfaisant mais il peut être amélioré en calculant l'expression de la branche inférieure obtenue avec la courbe isocèle puis en prenant en compte un troisième paramètre, l'angle maximum du triangle de référence.

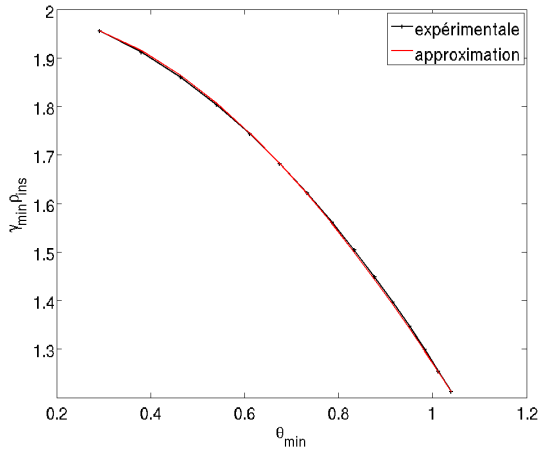

FIGURE 3.23 – Comparaison entre la branche supérieure et son approximation en P^2

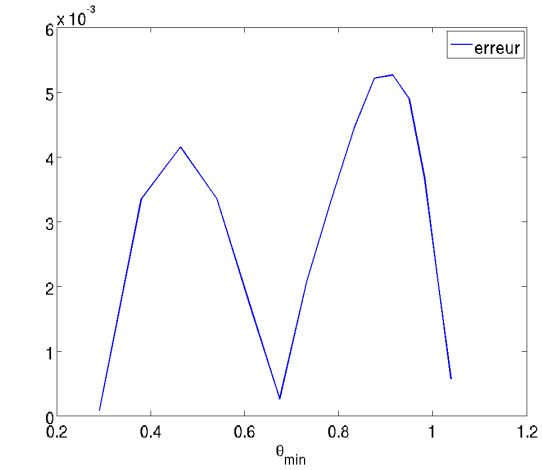

FIGURE 3.24 – Erreur relative entre la branche supérieure et son approximation en P^2

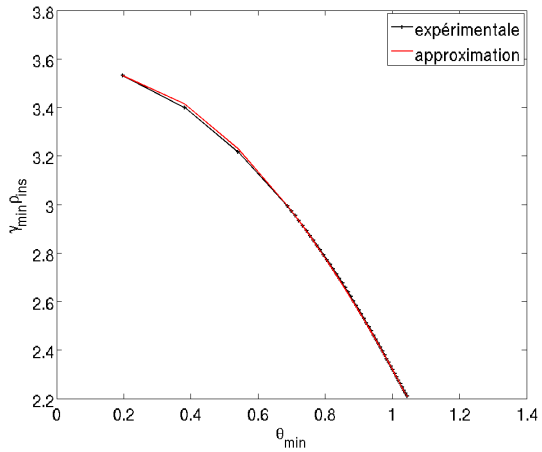

FIGURE 3.25 – Comparaison entre la branche supérieure et son approximation en P^3

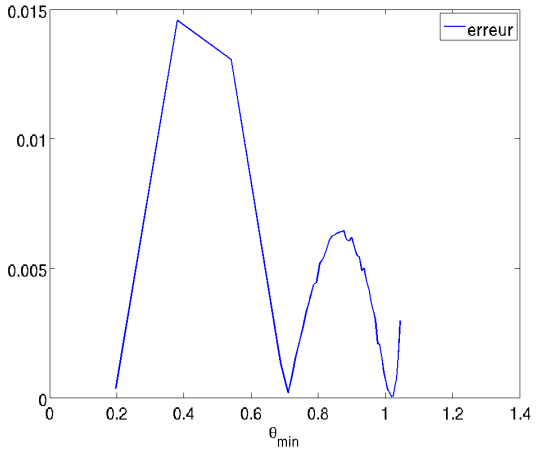

FIGURE 3.26 – Erreur relative entre la branche supérieure et son approximation en P^3

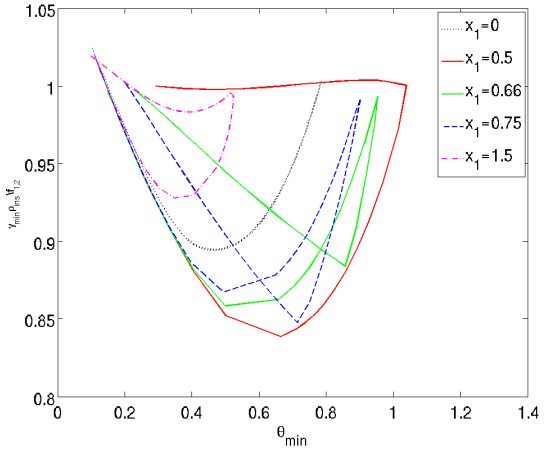

FIGURE 3.27 – $\xi_F = \rho_{ins}/f_{1,2}$ pour les huit configurations en P^2

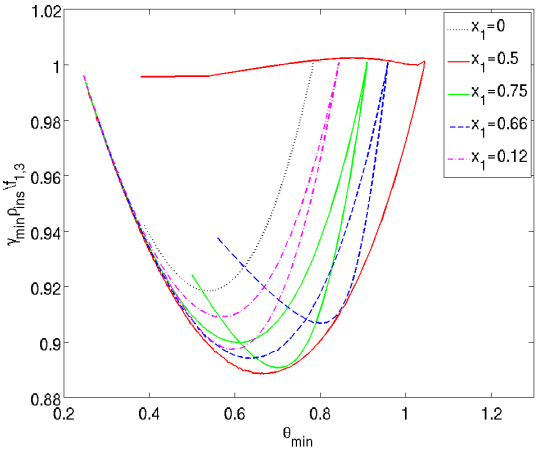

FIGURE 3.28 – $\xi_F = \rho_{ins}/f_{1,3}$ pour les huit configurations en P^3

3.2.2 Minoration de $\gamma_{\min}\rho_{\text{ins}}$

Comme pour la majoration de $\gamma_{\min}\rho_{\text{ins}}$, nous avons d'abord essayé d'approcher la branche inférieure de la courbe isocèle des Fig. 3.15 et 3.16 par des fonctions polynomiales de $\theta_{\min}$, mais nous nous sommes rendus compte que les fonctions trigonomètriques étaient les plus appropriées. Nous avons trouvé qu'une bonne approximation de cette branche inférieure est donnée par

$$f_{2,1} \ : \ (\theta_{\min}) \longmapsto f_{2,1}\left(\theta_{\min}\right) = 1 - \sin\left(\frac{\theta_{\min}}{2}\right).$$

Sur la Fig. 3.29, nous représentons $f_{2,1}$ (courbe rouge) et la branche inférieure (courbe noire avec +). Les deux courbes sont superposées. Sur la Fig. 3.30, nous traçons l'erreur relative entre les deux fonctions. Cette erreur est inférieure à 1%, de telle sorte que $f_{2,1}$ est une bonne approximation de la branche inférieure. Pour des polynômes P^2 et P^3, nous avons trouvé, en utilisant la méthode

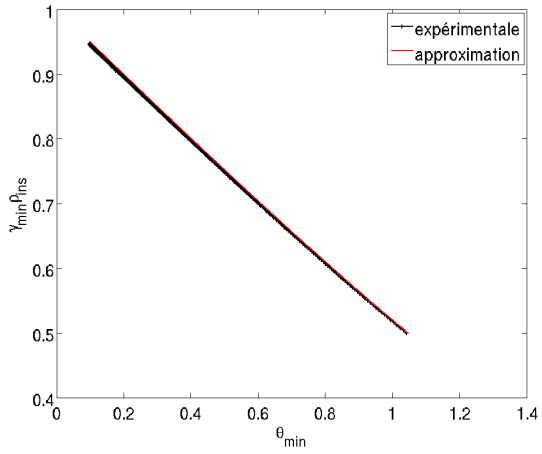

FIGURE 3.29 – Comparaison entre la branche inférieure et son approximation

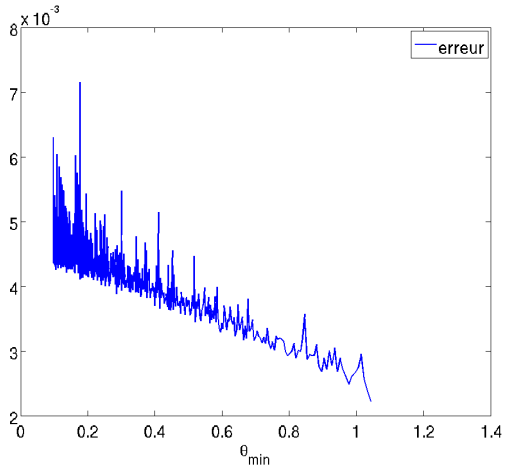

FIGURE 3.30 – Erreur relative entre la branche inférieure et son approximation

des moindres carrés, qu'une bonne approximation de la branche inférieure est respectivement

$$f_{2,2} \ : \ \theta_{\min} \longmapsto f_{2,2}\left(\theta_{\min}\right) = 2.68 - 0.49\cos\left(\theta_{\min}\right) - 1.41\sin\left(\theta_{\min}\right),$$

et

$$f_{2,3} =: \theta_{\min} \longmapsto f_{2,3}(\theta_{\min}) = 4.03 - 0.05\cos(\theta_{\min}) - 2.08\sin(\theta_{\min}),$$

Sur la Fig, 3.31 (resp. Fig. 3.33), nous représentons $f_{2,2}$ (resp. $f_{2,3}$) et la branche inférieure pour des polynômes P^2 (resp. P^3). Ici aussi, les courbes sont superposées. Sur la Fig. 3.32 (resp. Fig. 3.34), nous traçons l'erreur relative entre les deux fonctions pour des polynômes P^2 (resp. P^3).

Nous n'avons pas représenté $\frac{\gamma_{\min}\rho_{\text{ins}}}{f_{2,i}}$ car $\frac{f_{2,i}}{\rho_{\text{ins}}}$ ne fournit qu'une minoration du paramètre optimal

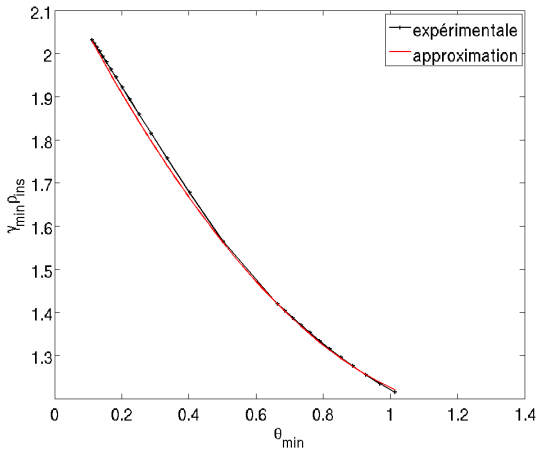

FIGURE 3.31 – Comparaison entre la branche inférieure et son approximation en P^2

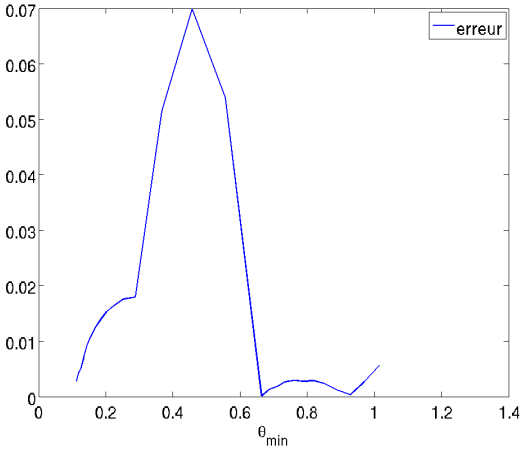

FIGURE 3.32 – Erreur relative entre la branche inférieure et son approximation en P^2

et ne peut donc garantir la stabilité. Cependant, nous allons voir à la section suivante que cette quantité, couplée à l'angle maximal du triangle, permet d'obtenir une approximation très précise de $\gamma_{\min}$.

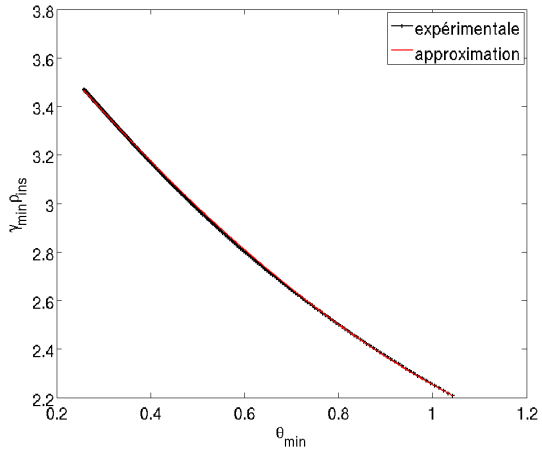

FIGURE 3.33 – Comparaison entre la branche inférieure et son approximation en P^3

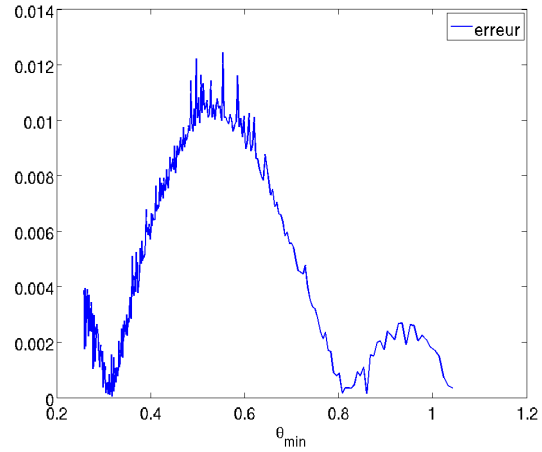

FIGURE 3.34 – Erreur relative entre la branche inférieure et son approximation en P^3

3.2.3 Influence de l'angle maximal

Dans cette section, nous étudions l'influence de l'angle maximal de K_0, $\theta_{\max}$ sur $\gamma_{\min}$. En effet, jusqu'à présent nous n'avons considéré que deux paramètres ρ_{ins} et $\theta_{\min}$, pour représenter $\gamma_{\min}$, alors qu'un triangle est complétement caractérisé par trois paramètres.
Soit $\theta_{\min} \in \left[0; \frac{\pi}{3}\right]$. Nous notons θ le troisième angle de K_0. De manière évidente,

$$\theta_{\min} \leq \theta \leq \theta_{\max} \text{ et } \theta_{\max} + \theta + \theta_{\min} = \pi.$$

Alors, pour un $\theta_{\min}$ donné, nous avons

$$\frac{\pi - \theta_{\min}}{2} \leq \theta_{\max} \leq \pi - 2\theta_{\min}.$$

La borne inférieure est atteinte quand $\theta = \theta_{max}$, *i.e.* quand K_0 est un triangle isocèle tel que l'angle relatif à la base est plus grand que $\frac{\pi}{3}$ (cf. Fig 3.35). Dans cette configuration, nous savons, d'après la section 3.2.2, que $\gamma_{\min}\rho_{\text{ins}} \simeq f_{2,i}$. La borne supérieure est atteinte quand $\theta = \theta_{\min}$, *i.e.*

quand K_0 est un triangle isocèle avec un angle relatif à la base plus petit que $\frac{\pi}{3}$(cf. Fig 3.36). Dans cette configuration, nous savons d'après la section 3.2.1, que $\gamma_{\min}\rho_{\text{ins}} \simeq f_{1,i}$.
Puisque nous savons également que

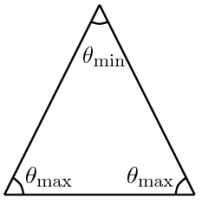

FIGURE 3.35 – Triangle isocèle où $\theta = \theta_{\max}$

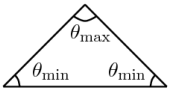

FIGURE 3.36 – Triangle isocèle où $\theta = \theta_{\min}$

$$f_{2,i}\left(\theta_{\min}\right) \leq \gamma_{\min}\rho_{\text{ins}} \leq f_{1,i}\left(\theta_{\min}\right)$$

nous cherchons une approximation de $\gamma_{\min}\rho_{\text{ins}}$ sous la forme

$$F\left(\theta_{\max}, \theta_{\min}\right) = f_{1,i}\left(\theta_{\min}\right) + G\left(\theta_{\max}\right)\left(f_{2,i}\left(\theta_{\min}\right) - f_{1,i}\left(\theta_{\min}\right)\right).$$

où G est tel que

$$\begin{cases} G\left(\frac{\pi-\theta_{\min}}{2}\right) = 1, \\ G\left(\pi - 2\theta_{\min}\right) = 0, \\ 0 \leq G\left(\theta_{\max}\right) \leq 1, \text{ pour } \frac{\pi-\theta_{\min}}{2} \leq \theta_{\max} \leq \pi - 2\theta_{\min}. \end{cases}$$

Ici aussi, nous avons d'abord tenté d'exprimer G comme une fonction polynomiale de $\theta_{\max}$ mais l'expression la plus appropriée semble être

$$G = \cos\left(\Theta\right) \text{ avec } \Theta = \frac{\pi}{2}\left(\frac{2\theta_{\max} - \pi + \theta_{\min}}{\pi - 3\theta_{\min}}\right).$$

Sur les Fig. 3.37 et 3.38, nous représentons $\gamma_{\min}\rho_{\text{ins}}$ et son approximation par $F\left(\theta_{\min}, \theta_{\max}\right)$ pour les huit configurations décrites précédemment et pour des fonctions de base P^1.
Pour toutes les configurations, les courbes sont parfaitement superposées. Sur la Fig. 3.39, nous représentons l'erreur relative entre $\gamma_{\min}\rho_{\text{ins}}$ et son approximation par F. Par souci de clarté, nous nous restreignons aux configurations $x_1 = 0.5$; $x_1 = 0.12$; $x_1 = 1.5$ et $x_1 = 0$. Néanmoins les résultats sont similaires pour les trois autres configurations. On conclut donc qu'une bonne approximation de $\gamma_{\min}$ pourrait être $\dfrac{F\left(\theta_{\max}, \theta_{\min}\right)}{\rho_{ins}}$ et nous proposons d'utiliser

$$\xi_F = \frac{\rho_{\text{ins}}}{F\left(\theta_{\max}, \theta_{\min}\right)}.$$

Sur la Fig. 3.40, nous traçons $\dfrac{\gamma_{\min}\rho_{\text{ins}}}{F}$ pour les huit configurations que nous avons considérées. Cette quantité varie de 0.99 à 1 (1%) et $\dfrac{\rho_{\text{ins}}}{F}$ est alors un meilleur choix de ξ_F que ρ_{ins} ou $\dfrac{\rho_{\text{ins}}}{f_{1,1}}$.
Sur la Fig. 3.41 (resp. Fig. 3.42), nous représentons $\dfrac{\gamma_{\min}\rho_{\text{ins}}}{F}$ pour les cinq configurations que nous

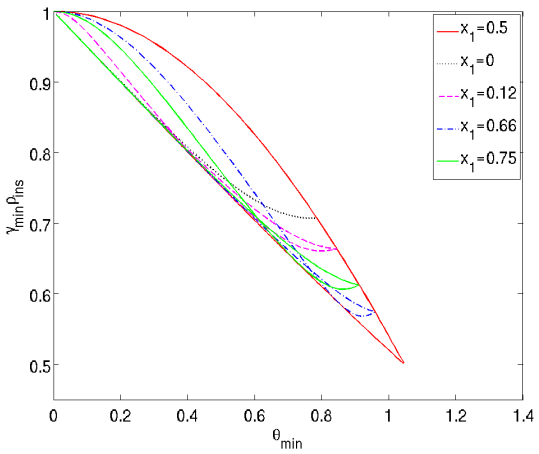

FIGURE 3.37 – Comparaison entre $\gamma_{\min}\rho_{\text{ins}}$ et $F(\theta_{\min}, \theta_{\max})$ en P^1 ($x_1 < 1$)

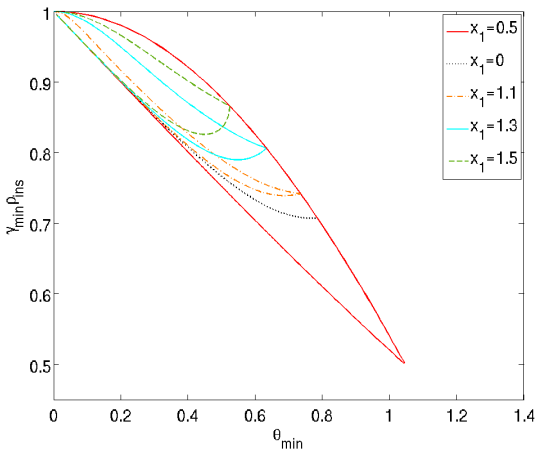

FIGURE 3.38 – Comparaison entre $\gamma_{\min}\rho_{\text{ins}}$ et $F(\theta_{\min}, \theta_{\max})$ en P^1 ($x_1 > 1$)

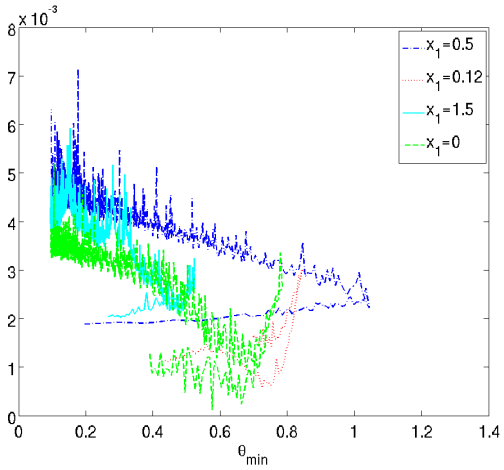

FIGURE 3.39 – Erreur entre $\gamma_{\min}\rho_{\text{ins}}$ et $F(\theta_{\min}, \theta_{\max})$

avons considérées en P^2 (resp. en P^3). Cette quantité varie de 0.995 à 1.04 (3%) (resp. de 0.995 à 1.008 (1%)) et ici aussi, pour des polynômes P^2 et P^3, $\frac{\rho_{\text{ins}}}{F}$ est un meilleur choix de ξ_F que ρ_{ins} ou $\frac{\rho_{\text{ins}}}{f_{1,1}}$.
Ce dernier choix nous permet donc de calculer un paramètre de pénalisation indépendant du maillage. Ainsi, il n'est plus nécessaire de réajuster α pour chaque maillage, soit de choisir α très élevé et donc de sur-pénaliser la forme bilinéaire. Il nous reste maintenant à étudier l'influence de ces différents choix sur la condition CFL pour des maillages non-uniformes.

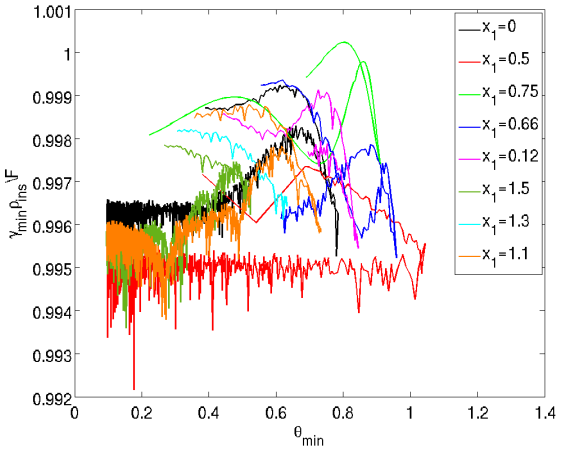

FIGURE 3.40 – $\xi_F = \rho_{ins}/F$ pour les huit configurations

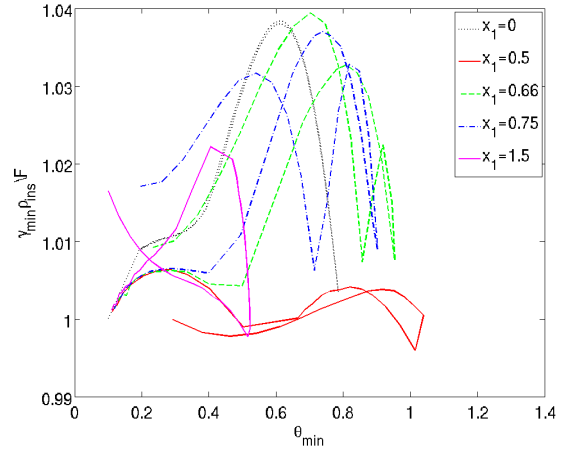

FIGURE 3.41 – $\xi_F = \rho_{ins}/F$ pour les huit configurations en P^2

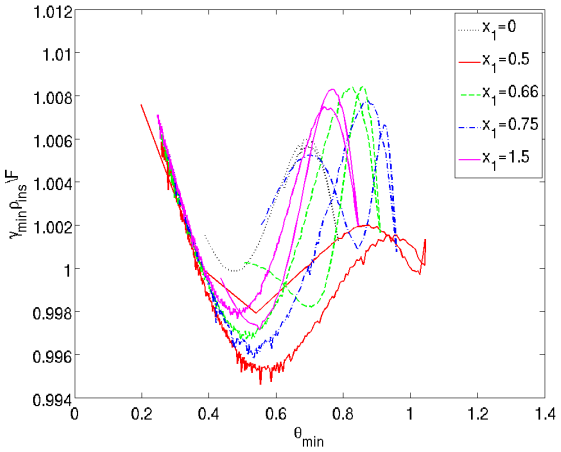

FIGURE 3.42 – $\xi_F = \rho_{ins}/F$ pour les huit configurations en P^3

3.3 Etude de la condition CFL

Au cours de ce chapitre, nous avons considéré quatre choix possibles de ξ_F : ρ_{circ}, ρ_{ins}, $\rho_{\text{ins}}/f_{1,1}$ et ρ_{ins}/F et nous avons conclu que le dernier était le plus adapté. Nous avons également vu au chapitre précédent que le choix du coefficient de pénalisation pouvait influencer fortement la condition de stabilité du schéma saute-moutons. Il est donc naturel de se demander comment les quatre choix de ξ_F peuvent modifier la condition CFL. Dans cette section, nous nous sommes intéréssés à la discrétisation du carré $[0,1]^2$ par un maillage triangulaire non-uniforme (cf. Fig. 3.43). Nous avons considéré une source ponctuelle en espace placée en $(0.5, 0.5)$. L'expression de la source en temps est la même que celle donnée dans la section 3.1.

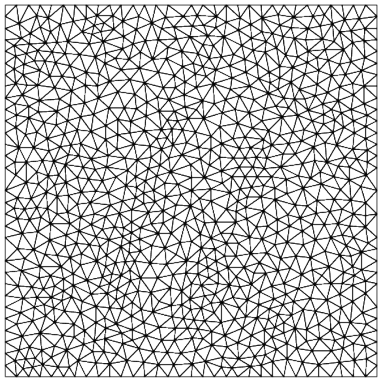

FIGURE 3.43 – Maillage du carré unité $[0,1]^2$

Pour chacun des quatre choix de ξ_F, nous avons étudié numériquement la dépendance entre la condition CFL et le paramètre de pénalisation γ_1. Le calcul de la condition CFL pour γ_1 fixé est réalisé par dichotomie en utilisant l'algorithme 2. Le principe de cet algorithme est similaire à celui de l'algorithme 1 : on résout l'équation des ondes pour un pas de temps donné puis on diminue le pas de temps en cas d'explosion et on l'augmente sinon. Nous présentons sur la figure

Algorithme 2

1: $\Delta t_1 = 6.9\ 10^{-3}$ et $\Delta t_2 = 0.17$
2: $\Delta t = \dfrac{\Delta t_1 + \Delta t_2}{2}$
3: Nous calculons la solution avec Δt
4: **si** explosion **alors**
5: $\quad \Delta t_2 = \Delta t$
6: **sinon**
7: $\quad \Delta t_1 = \Delta t$
8: **finsi**
9: **si** $|\Delta t - \dfrac{\Delta t_1 + \Delta t_2}{2}| < 10^{-5}\Delta t$ **alors**
10: $\quad \Delta_{\min} = \Delta t_1^2$
11: **sinon**
12: $\quad$ Retour en 2.
13: **finsi**

3.44 les résultats obtenus pour des éléments finis P^1 et pour :

– $\xi_F = \rho_{circ}$: ligne noire pleine,
– $\xi_F = \rho_{ins}$: ligne rouge avec tirets et pointillés,
– $\xi_F = \dfrac{\rho_{ins}}{f_{1,1}}$: ligne bleue avec tirets,
– $\xi_F = \dfrac{\rho_{ins}}{F}$:ligne magenta avec pointillés.

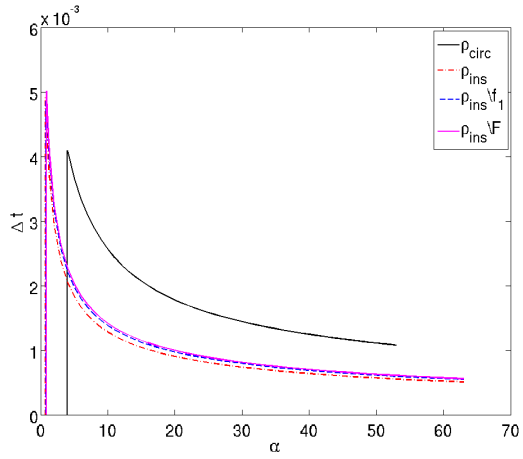

FIGURE 3.44 – Condition CFL pour quatre choix de ξ_F en P^1

Remarquons tout d'abord que, comme nous pouvions nous y attendre, le paramètre optimal $\alpha_{\min}$ dépend du choix de ξ_F. Il est également clair sur la figure 3.44 que les trois choix ρ_{ins}, $\rho_{\text{ins}}/f_{1,1}$ et ρ_{ins}/F permettent d'utiliser un pas de temps plus élevé $\left(5\ 10^{-3}\right)$ que le choix ρ_{circ} $\left(4.1\ 10^{-3}\right)$, soit un gain de 22% en temps de calcul.
Il n'y a cependant pas de différence notable entre les trois choix et nous préconisons par conséquent de choisir $\xi_F = \rho_{\text{ins}}$, ce qui est le plus simple à mettre en oeuvre.
Sur la figure 3.45 (resp. 3.46), nous représentons l'évolution de la condition CFL en fonction de γ_1 pour des éléments P^2 (resp. P^3) et pour :
– $\xi_F = \rho_{circ}$: ligne noire pleine,
– $\xi_F = \rho_{ins}$: ligne rouge avec tirets et pointillés.

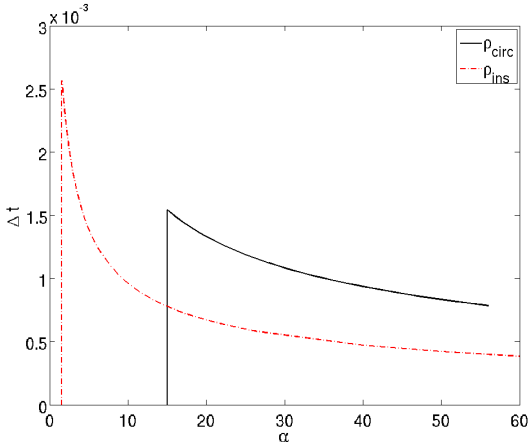

FIGURE 3.45 – Condition CFL pour deux choix de ξ_F en P^2

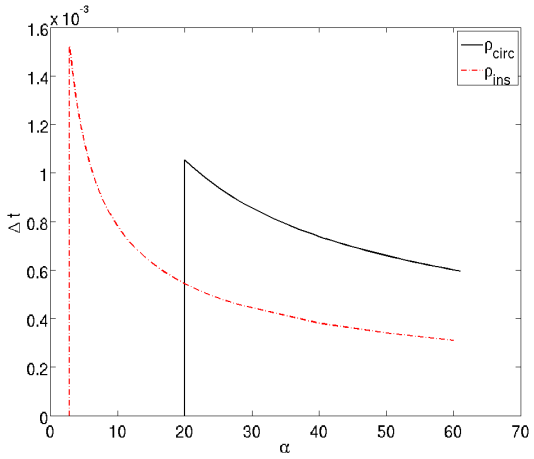

FIGURE 3.46 – Condition CFL pour deux choix de ξ_F en P^3

On remarque qu'on peut utiliser un pas de temps de $2.56\ 10^{-3}$ (resp. $1.53\ 10^{-3}$) en considérant le rayon du cercle inscrit et de $1.55\ 10^{-3}$ (resp. $1.05\ 10^{-3}$) en considérant le rayon du cercle circonscrit, ce qui représente un gain de 65% (resp. 46%) en temps de calcul.
Nous avons également réalisé des expériences en dimension 3 en discrétisant le cube $[0,1]^3$ par un maillage tétraèdrique non uniforme. Pour des raisons évidentes de coût de calcul, nous n'avons pas pu mener une étude aussi précise qu'en dimension 2 mais nous avons obtenu des gains de 20% pour des éléments P^1, de 25% pour des éléments P^2 et de 33% pour des éléments P^3 en utilisant le rayon de la sphère inscrite plutôt que celui de la sphère circonscrite.

3.4 Conclusion

Dans ce chapitre, nous avons comparé numériquement les différents choix du paramètre ξ_F que l'on peut trouver dans la littérature. Comme nous l'avions entrevu au chapitre 2, le choix le plus judicieux semble être de considérer le rayon du cercle inscrit. Cela permet en effet d'obtenir une valeur beaucoup plus constante du paramètre de pénalisation qu'avec les autres choix. Néanmoins, ce résultat n'était clairement pas suffisant puisque le paramètre de pénalisation variait encore fortement (de l'ordre de 100%). C'est pourquoi nous avons proposé plusieurs choix un peu plus sophistiqués amenant à une variation du paramètre de pénalisation de l'ordre du pourcent. Ainsi, il n'est plus nécessaire d'ajuster le paramètre de pénalisation en fonction de l'expérience pour assurer la stabilité de la méthode IPDG quelle que soit la nature du maillage. De plus, on a également pu constater que ce choix influe fortement sur la condition CFL. Effectivement, les résultats de la section 3.3 témoignent d'un gain oscillant entre 22% et 65% en 2D et entre 20% et 33% en 3D sur le pas de temps, c'est-à-dire que grâce à ce choix de ξ_F, on peut diviser le nombre d'itérations nécessaire pour atteindre un même temps final par une constante comprise entre 1.22 et 1.65.

3.A Analyse de stabilité

Dans cette section, nous allons présenter les résultats analytiques que nous avons obtenus pour différents maillages triangulaires. Afin d'appliquer une analyse de stabilité similaire à celle faite dans le chapitre 2, nous n'allons pas simplement considérer pour élément de base un triangle mais

plutôt le parallélogramme $E_{I,J}$ formé par $K_{I,J} \cup K^*_{I,J}$. Nous représentons $E_{I,J}$ ainsi que ses voisins sur la figure 3.47 dans le cas de triangles équilatéraux.

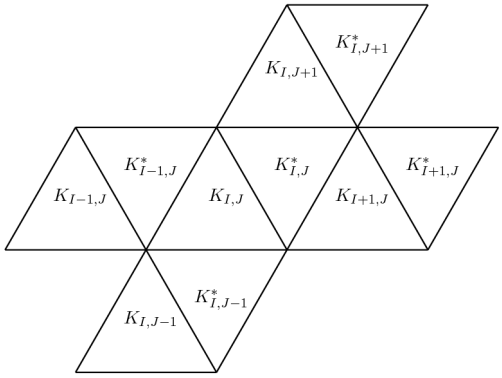

FIGURE 3.47 – Voisins d'une maille $E_{I,J} = K_{I,J} \cup K^*_{I,J}$

Si l'on considère l'équation des ondes totalement discrétisée sur un élément E_0 du maillage, nous avons pour tout $E_0 \in \mathcal{T}_h$

$$M_{2,p}\frac{U^{n+1}_{I,J} - 2U^n_{I,J} + U^{n-1}_{I,J}}{\Delta t^2} + K_{2,p}U^n_{I,J} + \left(K^W_{2,p}\right)^T U^n_{I-1,J} + K^W_{2,p}U^n_{I+1,J} + \left(K^N_{2,p}\right)^T U^n_{I,J-1} + K^N_{2,p}U^n_{I,J+1} = 0 \tag{3.2}$$

où $U_{I,J}$ correspond au vecteur d'inconnues U restreint à l'élément $E_{I,J}$, $M_{2,p}$ (resp. $K_{2,p}$) est la matrice de masse (resp. matrice de raideur) relative à l'élément $E_{I,J}$, $K^W_{2,p}$ (resp. $K^N_{2,p}$) est la matrice de raideur relative à l'interface entre les éléments $E_{I,J}$ et $E_{I+1,J}$ (resp. $E_{I,J}$ et $E_{I,J+1}$) en considérant des polynômes de degré p. Toutes ces matrices sont de taille $(2N, 2N)$ où N est le nombre de fonctions de base considérées car une maille $E_{I,J}$ est composée de deux éléments triangulaires.
Ainsi, en appliquant la transformée de Fourier discrète à (3.2) suivant les directions $(1, 0)$ et (x_1, y_1), on obtient, $\forall \beta_1, \beta_2 \in [-\pi, \pi]$

$$M_{2,p}\frac{\hat{U}^{n+1} - 2\hat{U}^n + \hat{U}^{n-1}}{\Delta t^2} + K_{\beta_1,\beta_2}\hat{U}^n = 0 \tag{3.3}$$

où $K_{\beta_1,\beta_2} = K_{2,p} + K^W_{2,p}e^{i\beta_1} + \left(K^W_{2,p}\right)^T e^{-i\beta_1} + K^N_{2,p}e^{i\beta_2} + \left(K^N_{2,p}\right)^T e^{-i\beta_2}$.
De la même manière que dans le chapitre 2, la stabilité du schéma est assurée si et seulement si

$$0 \leq \lambda \leq \frac{4}{\Delta t^2}$$

où $\lambda \in \Lambda(\beta_1, \beta_2)$, $\Lambda(\beta_1, \beta_2)$ représentent l'ensemble des valeurs propres de $N_{\beta_1,\beta_2} := M^{-1}_{2,p}K_{\beta_1,\beta_2}$.
Une condition nécessaire et suffisante de stabilité du schéma est donc

$$\lambda_{\text{min}} \geq 0 \qquad \text{et} \qquad \Delta t \leq \frac{2}{\sqrt{\lambda_{\text{max}}}}$$

avec $\lambda_{\text{min}} = \min\limits_{\beta_1,\beta_2 \in [-\pi,\pi]} [\min(\Lambda(\beta_1, \beta_2))]$ et $\lambda_{\text{max}} = \max\limits_{\beta_1,\beta_2 \in [-\pi,\pi]} [\max(\Lambda(\beta_1, \beta_2))]$.

3.A.1 Maillage issu de triangles équilatéraux

Dans cette sous-section, on s'intéresse à un maillage généré à partir de triangles équilatéraux *i.e.* construit à partir du triangle de référence K_0 défini par $x_1 = 1/2$ et $y_1 = \sqrt{3}/2$, comme sur la figure 3.47.
A titre indicatif, on donne les expressions des différentes matrices intervenant dans (3.3). La matrice de masse est donnée par

$$M_{2,p} = \begin{pmatrix} M^1_{2,p} & 0 \\ 0 & M^1_{2,p} \end{pmatrix} \qquad \text{où} \qquad M^1_{2,p} = \frac{1}{24}\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

Les matrices de raideurs $K^W_{2,p}$ et $K^N_{2,p}$ sont les suivantes

$$K^W_{2,p} = \begin{pmatrix} 0 & 0 \\ K^{W,1}_{2,p} & 0 \end{pmatrix} \qquad \text{où} \qquad K^{W,1}_{2,p} = \frac{1}{6}\begin{pmatrix} \sqrt{3}-\alpha & -\sqrt{3} & \sqrt{3}-2\alpha \\ -\sqrt{3} & 0 & -\sqrt{3} \\ \sqrt{3}-2\alpha & -\sqrt{3} & \sqrt{3}-\alpha \end{pmatrix}$$

et

$$K^N_{2,p} = \begin{pmatrix} 0 & 0 \\ K^{N,1}_{2,p} & 0 \end{pmatrix} \qquad \text{où} \qquad K^{N,1}_{2,p} = \frac{1}{6}\begin{pmatrix} \sqrt{3}-\alpha & \sqrt{3}-2\alpha & -\sqrt{3} \\ \sqrt{3}-2\alpha & \sqrt{3}-\alpha & -\sqrt{3} \\ -\sqrt{3} & -\sqrt{3} & 0 \end{pmatrix}.$$

Enfin la matrice $K_{2,p}$, qui est pleine, se formule de la façon suivante

$$K_{2,p} = \begin{pmatrix} K^1_{2,p} & K^2_{2,p} \\ K^2_{2,p} & K^1_{2,p} \end{pmatrix}$$

avec

$$K^1_{2,p} = \frac{\alpha}{6}\begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix}, \qquad K^2_{2,p} = \frac{1}{6}\begin{pmatrix} 0 & -\sqrt{3} & -\sqrt{3} \\ -\sqrt{3} & \sqrt{3}-\alpha & \sqrt{3}-2\alpha \\ -\sqrt{3} & \sqrt{3}-2\alpha & \sqrt{3}-\alpha \end{pmatrix}.$$

L'étude des valeurs propres de la matrice N_{β_1,β_2} s'avérant particulièrement compliquée pour tout β_1 et β_2, nous nous sommes intéressés aux cas $\beta_1 = \beta_2 = 0$ et $\beta_1 = \beta_2 = \pi$ pour obtenir une condition nécessaire de stabilité.

– Dans le cas où $\beta_1 = \beta_2 = 0$, le polynôme caractéristique $p_\alpha(\lambda, \beta_1, \beta_2)$ de la matrice N_{β_1,β_2} est donné par

$$p_\alpha(\lambda, 0, 0) = \lambda^6 + \sum_{i=0}^{5} c_i(\alpha, 0, 0)\,\lambda^i$$

où les coefficients $c_i(\alpha, 0, 0)$ sont définis par

$$\begin{cases} c_5(\alpha,0,0) = & -60\alpha, \\ c_4(\alpha,0,0) = & 288\left(5\alpha^2 - 3\right), \\ c_3(\alpha,0,0) = & 3456\alpha\left(9 - 5\alpha^2\right), \\ c_2(\alpha,0,0) = & 20736\left(9 - 18\alpha^2 + 5\alpha^4\right), \\ c_1(\alpha,0,0) = & 248832\alpha\left(-9 + 6\alpha^2 - \alpha^5\right), \\ c_0(\alpha,0,0) = & 0. \end{cases}$$

Nous allons utiliser le lemme 2.2.1 afin d'étudier la positivité des valeurs propres de la matrice $N_{0,0}$. Nous présentons brièvement les résultats obtenus pour chaque coefficient $c_i(\alpha,0,0)$, la technique employée étant la même que celle utilisée dans le chapitre 2.

- $-c_5(\alpha,0,0) \geq 0 \Leftrightarrow \alpha \geq 0$,
- $c_4(\alpha,0,0) \geq 0 \Leftrightarrow \alpha \geq \dfrac{\sqrt{15}}{5}$,
- $-c_3(\alpha,0,0) \geq 0 \Leftrightarrow \alpha \geq \dfrac{3\sqrt{5}}{5}$,
- $c_2(\alpha,0,0) \geq 0 \Leftrightarrow \alpha \geq \sqrt{3}$,
- $-c_1(\alpha,0,0) \geq 0 \Leftrightarrow \alpha \geq \sqrt{3}$.

Ainsi, il résulte de ce cas que les valeurs propres de la matrice $N_{0,0}$ sont positives si et seulement si

$$\alpha \geq \sqrt{3} \simeq 1.73. \tag{3.4}$$

– Intéressons nous à présent au cas où $\beta_1 = \beta_2 = \pi$. Le polynôme caractéristique de la matrice N_{β_1,β_2} est ainsi donné par

$$p_\alpha(\lambda,\pi,\pi) = \lambda^6 + \sum_{i=0}^{5} c_i(\alpha,\pi,\pi)\,\lambda^i$$

où les coefficients $c_i(\alpha,\pi,\pi)$ sont définis par

$$\begin{cases}
c_5(\alpha,\pi,\pi) = & -60\alpha, \\
c_4(\alpha,\pi,\pi) = & 8\left(160\alpha^2 + 72\sqrt{3}\alpha - 183\right), \\
c_3(\alpha,\pi,\pi) = & -288\alpha\left(40\alpha^2 + 76\sqrt{3}\alpha - 191\right), \\
c_2(\alpha,\pi,\pi) = & 48\left(848\alpha^4 + 4944\sqrt{3}\alpha^3 - 6088\alpha^2 - 9648\sqrt{3}\alpha + 10731\right), \\
c_1(\alpha,\pi,\pi) = & -576\alpha\left(80\alpha^4 + 1216\sqrt{3}\alpha^3 + 5712\alpha^2 - 13608\sqrt{3}\alpha + 15387\right), \\
c_0(\alpha,\pi,\pi) = & 4608\left(120\sqrt{3}\alpha^5 + 1676\alpha^4 - 810\sqrt{3}\alpha^3 - 11753\alpha^2\right. \\
& \left. +11052\sqrt{3}\alpha - 7776\right).
\end{cases}$$

L'étude des signes de chaque coefficient, que nous ne détaillons pas ici, nous montre que les valeurs propres de la matrice $N_{\pi,\pi}$ sont positives si et seulement si

$$\alpha \geq \frac{9\sqrt{3}}{10} \simeq 1.56. \tag{3.5}$$

– Enfin, nous considérons le cas $\beta_1 = 0$, $\beta_2 = \pi$, qui est identique au cas $\beta_1 = \pi$, $\beta_2 = 0$. Avec un tel choix de β_1 et β_2, la matrice $N_{0,\pi}$ admet pour polynôme caractéristique

$$p_\alpha(\lambda,0,\pi) = \lambda^6 + \sum_{i=0}^{5} c_i(\alpha,0,\pi)\,\lambda^i$$

avec pour coefficients $c_i\,(\alpha,0,\pi)$:

$$\begin{cases} c_5\,(\alpha,\pi,\pi) = & -60\alpha, \\ c_4\,(\alpha,\pi,\pi) = & 1280\alpha^2+384\sqrt{3}\alpha-672, \\ c_3\,(\alpha,\pi,\pi) = & -11520\alpha^3-14592\sqrt{3}\alpha^2+25344\alpha, \\ c_2\,(\alpha,\pi,\pi) = & 40704\alpha^4+158208\sqrt{3}\alpha^3-141312\alpha^2-142848\sqrt{3}\alpha+112896, \\ c_1\,(\alpha,\pi,\pi) = & -46080\alpha^5-466944\sqrt{3}\alpha^4-1419264\alpha^3+2433024\sqrt{3}\alpha^2-1963008\alpha, \\ c_0\,(\alpha,\pi,\pi) = & 368640\sqrt{3}\alpha^5+3391488\alpha^4-1327104\sqrt{3}\alpha^3-11206656\alpha^2 \\ & +7741440\sqrt{3}\alpha-3981312. \end{cases}$$

L'étude de ces différents coefficients nous amène à la conclusion que les valeurs propres de $N_{0,\pi}$ sont positives si et seulement si

$$\alpha \geq \frac{3\sqrt{3}}{5} \simeq 1.04. \tag{3.6}$$

Ainsi en regroupant les conditions (3.4) à (3.5), on obtient donc qu'il faut que $\alpha \geq \sqrt{3} \simeq 1.73$ pour que le schéma soit stable.

Remarque 3.A.1. *Les expériences numériques que nous avons effectuées sur de tels maillages corroborent bien ce résultat puisque numériquement, nous avons stabilité lorsque* $\alpha \geq 1.74$*. Il semblerait donc que l'étude du cas* $\beta_1=\beta_2=0$ *soit suffisant pour déterminer la condition de stabilité du schéma.*

3.A.2 Maillage issu de triangles rectangles

Dans cette sous-section, on considère pour triangle de référence le triangle K_0 avec $x_1=0$ et $y_1=1$, ce qui donne un maillage constitué de triangles rectangles isocèles.
Dans une telle configuration, la matrice de masse est identique à celle présentée en sous-section 3.A.1 et les matrices de raideurs $K^W_{2,p}$ et $K^N_{2,p}$ sont données par

$$K^W_{2,p}=\begin{pmatrix} 0 & 0 \\ K^{W,1}_{2,p} & 0 \end{pmatrix} \qquad \text{où} \qquad K^{W,1}_{2,p}=\frac{1}{12}\begin{pmatrix} -2\alpha+6 & -3 & -4\alpha+3 \\ -3 & 0 & -3 \\ -4\alpha+3 & -3 & -2\alpha \end{pmatrix}$$

et

$$K^N_{2,p}=\begin{pmatrix} 0 & 0 \\ K^{N,1}_{2,p} & 0 \end{pmatrix} \qquad \text{où} \qquad K^{N,1}_{2,p}=\frac{1}{12}\begin{pmatrix} -2\alpha+6 & -4\alpha+3 & -3 \\ -4\alpha+3 & -2\alpha & -3 \\ -3 & -3 & 0 \end{pmatrix}.$$

La matrice $K_{2,p}$ est, quant à elle, de la forme

$$K_{2,p}=\begin{pmatrix} K^1_{2,p} & K^2_{2,p} \\ K^2_{2,p} & K^1_{2,p} \end{pmatrix}$$

avec

$$K^1_{2,p}=\frac{\alpha}{6}\begin{pmatrix} 4 & 1 & 1 \\ 1 & 2\left(1+\sqrt{2}\right) & \sqrt{2} \\ 1 & \sqrt{2} & 2\left(1+\sqrt{2}\right) \end{pmatrix}, \qquad K^2_{2,p}=\frac{1}{12}\begin{pmatrix} 0 & -6 & -6 \\ -12 & -2\alpha\sqrt{2}+6 & -4\alpha\sqrt{2}+6 \\ -12 & -4\alpha\sqrt{2}+6 & -2\alpha\sqrt{2}+6 \end{pmatrix}.$$

Etudions à présent la positivité des valeurs propres de la matrice N_{β_1,β_2} dans les mêmes cas que dans la sous-section 3.A.1.

– Supposons que $\beta_1 = \beta_2 = 0$. Dans ce cas, le polynôme caractéristique de la matrice $N_{0,0}$ est donné par

$$p_\alpha(\lambda, 0, 0) = \lambda^6 + \sum_{i=0}^{5} c_i(\alpha, 0, 0)\,\lambda^i$$

où les coefficients $c_i(\alpha, 0, 0)$ sont définis par

$$\begin{cases} c_5(\alpha,0,0) = & -20\left(2+\sqrt{2}\right)\alpha, \\ c_4(\alpha,0,0) = & 768\left(1+\sqrt{2}\right)\alpha^2 + 192\left(\sqrt{2}-1\right)\alpha - 1440, \\ c_3(\alpha,0,0) = & -10368\left(1+\sqrt{2}\right)\alpha^3 + 1152\left(3-4\sqrt{2}\right)\alpha^2 + 17280\left(\sqrt{2}+2\right)\alpha, \\ c_2(\alpha,0,0) = & 6912\left(13+8\sqrt{2}\right)\alpha^4 + 27648\sqrt{2}\alpha^3 - 138240\left(2\sqrt{2}+3\right)\alpha^2 \\ & +82944\left(\sqrt{2}-1\right)\alpha + 186624, \\ c_1(\alpha,0,0) = & -82944\left(\sqrt{2}+4\right)\alpha^5 - 165888\alpha^4 + 829440\left(3+\sqrt{2}\right)\alpha^3 \\ & -497664\alpha^2 - 746496\left(\sqrt{2}+2\right)\alpha, \\ c_0(\alpha,0,0) = & 0. \end{cases}$$

En étudiant le signe de chaque coefficient $c_i(\alpha, 0, 0)$, on obtient que les valeurs propres de la matrice $N_{0,0}$ sont positives si et seulement si

$$\alpha \geq 1 + \sqrt{2} \simeq 2.41. \tag{3.7}$$

– Si $\beta_1 = \beta_2 = \pi$, alors $N_{\pi,\pi}$ admet pour polynôme caractéristique

$$p_\alpha(\lambda, \pi, \pi) = \lambda^6 + \sum_{i=0}^{5} c_i(\alpha, \pi, \pi)\,\lambda^i$$

où les coefficients $c_i(\alpha, 0, 0)$ sont définis par

$$\begin{cases} c_5(\alpha,\pi,\pi) = & -20\left(2+\sqrt{2}\right)\alpha \\ c_4(\alpha,\pi,\pi) = & 32\left(24+19\sqrt{2}\right)\alpha^2 + 384\left(\sqrt{2}+1\right)\alpha - 864 \\ c_3(\alpha,\pi,\pi) = & -384\left(19+16\sqrt{2}\right)\alpha^3 - 1152\left(15+14\sqrt{2}\right)\alpha^2 + 2304\left(4\sqrt{2}+11\right)\alpha \\ c_2(\alpha,\pi,\pi) = & 768\left(37+30\sqrt{2}\right)\alpha^4 + 4608\left(41\sqrt{2}+50\right)\alpha^3 - 4608\left(17\sqrt{2}+36\right)\alpha^2 \\ & -13824\left(11\sqrt{2}+14\right)\alpha + 186624 \\ c_1(\alpha,\pi,\pi) = & -9216\left(3\sqrt{2}+4\right)\alpha^5 - 18432\left(47+36\sqrt{2}\right)\alpha^4 - 55296\left(23+15\sqrt{2}\right)\alpha^3 \\ & +165888\left(23+20\sqrt{2}\right)\alpha^2 - 82944\left(11\sqrt{2}+34\right)\alpha \\ c_0(\alpha,\pi,\pi) = & 221184\left(3\sqrt{2}+4\right)\alpha^5 + 884736\left(5+3\sqrt{2}\right)\alpha^4 - 1327104\left(1+3\sqrt{2}\right)\alpha^3 \\ & -1327104\left(10\sqrt{2}+9\right)\alpha^2 + 1990656\left(7\sqrt{2}+10\right)\alpha - 11943936 \end{cases}$$

L'étude du signe de chaque coefficient $c_i(\alpha, \pi, \pi)$ nous permet d'affirmer que les valeurs propres de la matrice $N_{0,0}$ sont positives si et seulement si

$$\alpha \geq \frac{1}{2}\left(\sqrt{9+12\sqrt{2}}-3\right) \simeq 1.05. \tag{3.8}$$

– Enfin, intéressons-nous au cas $\beta_1 = 0$ et $\beta_2 = \pi$. Dans ce cas, la matrice $N_{0,\pi}$ a pour polynôme caractéristique

$$p_\alpha(\lambda, 0, \pi) = \lambda^6 + \sum_{i=0}^{5} c_i(\alpha, 0, \pi)\,\lambda^i$$

où les coefficients $c_i(\alpha, 0, \pi)$ sont définis par

$$\begin{cases}
c_5(\alpha,0,\pi) = & -20\left(2+\sqrt{2}\right)\alpha \\
c_4(\alpha,0,\pi) = & 688\left(1+\sqrt{2}\right)\alpha^2 + 96\left(3\sqrt{2}+5\right)\alpha - 1152 \\
c_3(\alpha,0,\pi) = & -384\left(19+18\sqrt{2}\right)\alpha^3 - 384\left(40+41\sqrt{2}\right)\alpha^2 + 1152\left(11\sqrt{2}+26\right)\alpha \\
c_2(\alpha,0,\pi) = & 768\left(57+23\sqrt{2}\right)\alpha^4 + 2304\left(81\sqrt{2}+80\right)\alpha^3 - 2304\left(57\sqrt{2}+74\right)\alpha^2 \\
& -6912\left(13\sqrt{2}+40\right)\alpha + 165888 \\
c_1(\alpha,0,\pi) = & -9216\left(\sqrt{2}+8\right)\alpha^5 - 313344\left(3+\sqrt{2}\right)\alpha^4 - 55296\left(9+20\sqrt{2}\right)\alpha^3 \\
& +55296\left(63+52\sqrt{2}\right)\alpha^2 - 165888\left(6\sqrt{2}+11\right)\alpha \\
c_0(\alpha,0,\pi) = & 110592\left(\sqrt{2}+8\right)\alpha^5 + 110592\left(45+11\sqrt{2}\right)\alpha^4 - 331776\left(24+5\sqrt{2}\right)\alpha^3 \\
& -331776\left(8\sqrt{2}+15\right)\alpha^2 + 1990656\left(\sqrt{2}+5\right)\alpha - 2985984
\end{cases}$$

En étudiant le signe des cinq expressions ci-dessus, on établit que $N_{0,\pi}$ est positive si et seulement si

$$\alpha \geq \frac{1}{62}\left(17+25\sqrt{2}+\sqrt{1911-638\sqrt{2}}\right) \simeq 1.36. \tag{3.9}$$

Si l'on regroupe les conditions (3.7) à (3.9), il vient qu'une condition nécessaire pour que les valeurs propres de N_{β_1,β_2} soient positives est que

$$\alpha \geq 1+\sqrt{2}.$$

Remarque 3.A.2. *Les expériences numériques que nous avons effectuées sur de tels maillages confirment bien ce résultat puisque numériquement, nous avons stabilité lorsque $\alpha \geq 2.42$. A nouveau, la condition de stabilité peut être calculée grâce au cas $\beta_1 = \beta_2 = 0$.*

3.A.3 Maillage issu de triangles quelconques

Dans cette sous-section, on considère pour triangle de référence le triangle K_0 avec $x_1 = 3/4$ et $y_1 = 1$, ce qui donne un maillage constitué de triangles quelconques.

Ici aussi, la matrice de masse est identique à celle présentée en sous-section 3.A.1 et les matrices de raideurs $K_{2,p}^W$ et $K_{2,p}^N$ sont données par

$$K_{2,p}^W = \begin{pmatrix} 0 & 0 \\ K_{2,p}^{W,1} & 0 \end{pmatrix} \qquad \text{où} \qquad K_{2,p}^{W,1} = \begin{pmatrix} -\frac{5}{24}\alpha + \frac{13}{32} & -\frac{25}{64} & -\frac{5}{12}\alpha + \frac{25}{64} \\ -\frac{25}{64} & 0 & -\frac{25}{64} \\ -\frac{5}{12}\alpha + \frac{25}{64} & -\frac{25}{64} & -\frac{5}{23}\alpha + \frac{3}{8} \end{pmatrix}$$

et

$$K_{2,p}^N = \begin{pmatrix} 0 & 0 \\ K_{2,p}^{N,1} & 0 \end{pmatrix} \qquad \text{où} \qquad K_{2,p}^{N,1} = \frac{1}{24}\begin{pmatrix} -4\alpha + 3 & -8\alpha + 6 & -6 \\ -8\alpha + 6 & -4\alpha + 9 & -6 \\ -6 & -6 & 0 \end{pmatrix}.$$

La matrice $K_{2,p}$ est, quant à elle, de la forme

$$K_{2,p} = \begin{pmatrix} K_{2,p}^1 & K_{2,p}^2 \\ K_{2,p}^2 & K_{2,p}^1 \end{pmatrix}$$

avec

$$K_{2,p}^1 = \frac{\alpha}{24}\begin{pmatrix} 18 & 4 & 5 \\ 4 & 2\left(4+\sqrt{17}\right) & \sqrt{17} \\ 5 & \sqrt{17} & 2\left(5+\sqrt{17}\right) \end{pmatrix}$$

et

$$K_{2,p}^2 = \begin{pmatrix} 0 & -\frac{17}{64} & -\frac{17}{64} \\ -\frac{17}{64} & -\frac{1}{24}\alpha\sqrt{17} + \frac{13}{32} & -\frac{1}{12}\alpha\sqrt{17} + \frac{17}{64} \\ -\frac{17}{64} & -\frac{1}{12}\alpha\sqrt{17} + \frac{17}{64} & -\frac{1}{24}\alpha\sqrt{17} + \frac{1}{8} \end{pmatrix}.$$

Etudions à présent la positivité des valeurs propres de la matrice N_{β_1,β_2} dans les mêmes cas que dans la sous-section 3.A.1.

– Supposons que $\beta_1 = \beta_2 = 0$. Dans ce cas, le polynôme caractéristique de la matrice $N_{0,0}$ est donné par

$$p_\alpha(\lambda, 0, 0) = \lambda^6 + \sum_{i=0}^{5} c_i(\alpha, 0, 0)\,\lambda^i$$

et les coefficients $c_i\left(\alpha,0,0\right)$ sont définis par

$$\begin{cases} c_5\left(\alpha,0,0\right) = & -5\left(9+\sqrt{17}\right)\alpha, \\ c_4\left(\alpha,0,0\right) = & 36\left(23+6\sqrt{17}\right)\alpha^2+\frac{3}{2}\left(45-\sqrt{17}\right)\alpha-\frac{4113}{4}, \\ c_3\left(\alpha,0,0\right) = & -36\left(333+121\sqrt{17}\right)\alpha^3+\frac{9}{2}\left(59-63\sqrt{17}\right)\alpha^2+\frac{12339}{4}\left(\sqrt{17}+9\right)\alpha, \\ c_2\left(\alpha,0,0\right) = & 243\left(273+80\sqrt{17}\right)\alpha^4+\frac{81}{2}\left(311\sqrt{17}-1005\right)\alpha^3 \\ & -\frac{27}{8}\left(17343\sqrt{17}+85951\right)\alpha^2+648\left(45-7\sqrt{17}\right)\alpha+186624, \\ c_1\left(\alpha,0,0\right) = & -1620\left(20\sqrt{17}+153\right)\alpha^5+810\left(425-117\sqrt{17}\right)\alpha^4 \\ & +\frac{81}{2}\left(32949+5933\sqrt{17}\right)\alpha^3+1944\left(99\sqrt{17}-485\right)\alpha^2 \\ & -186624\left(\sqrt{17}+9\right)\alpha, \\ c_0\left(\alpha,0,0\right) = & 0. \end{cases}$$

En étudiant le signe de chaque coefficient $c_i\left(\alpha,0,0\right)$, on obtient que les valeurs propres de la matrice $N_{0,0}$ sont positives si et seulement si

$$\alpha \geq \frac{1}{8}\left(9+\sqrt{17}+\frac{3}{85}\sqrt{37570-7310\sqrt{17}}\right) \simeq 2.02. \tag{3.10}$$

– Si $\beta_1=\beta_2=\pi$, alors $N_{\pi,\pi}$ admet pour polynôme caractéristique

$$p_\alpha\left(\lambda,\pi,\pi\right)=\lambda^6+\sum_{i=0}^{5}c_i\left(\alpha,\pi,\pi\right)\lambda^i$$

et les coefficients $c_i(\alpha,0,0)$ sont définis par

$$\begin{cases} c_5(\alpha,\pi,\pi) = & -5\left(9+\sqrt{17}\right)\alpha, \\ c_4(\alpha,\pi,\pi) = & 9\left(92+19\sqrt{17}\right)\alpha^2+3\left(29\sqrt{17}+135\right)\alpha-\dfrac{6561}{8}, \\ c_3(\alpha,\pi,\pi) = & -45\left(7155+1935\sqrt{17}\right)\alpha^3-27\left(\dfrac{1137}{2}+145\sqrt{17}\right)\alpha^2 \\ & +\dfrac{9}{8}\left(2135\sqrt{17}+20799\right)\alpha, \\ c_2(\alpha,\pi,\pi) = & 3\left(8411+2700\sqrt{17}\right)\alpha^4+\dfrac{9}{4}\left(21073\sqrt{17}+76689\right)\alpha^3 \\ & -\dfrac{27}{4}\left(3609\sqrt{17}+17339\right)\alpha^2-\dfrac{27}{16}\left(20689\sqrt{17}+95139\right)\alpha+\dfrac{32673537}{256}, \\ c_1(\alpha,\pi,\pi) = & -540\left(20\sqrt{17}+51\right)\alpha^5-9\left(\dfrac{111979}{2}+17280\sqrt{17}\right)\alpha^4 \\ & -\dfrac{27}{16}\left(622791+112795\sqrt{17}\right)\alpha^3 s+\dfrac{81}{8}\left(\dfrac{983053}{4}+74013\sqrt{17}\right)\alpha^2 \\ & -\dfrac{81}{256}\left(250073\sqrt{17}+6564321\right)\alpha, \\ c_0(\alpha,\pi,\pi) = & 6885\left(20\sqrt{17}+51\right)\alpha^5+\dfrac{459}{4}\left(\dfrac{80971}{4}+5805\sqrt{17}\right)\alpha^4 \\ & +\dfrac{1377}{64}\left(8343-59041\sqrt{17}\right)\alpha^3-\dfrac{4131}{64}\left(\dfrac{123873}{4}\sqrt{17}+42871\right)\alpha^2 \\ & +\dfrac{4131}{4}\left(1057\sqrt{17}+8271\right)\alpha-3370896. \end{cases}$$

L'étude du signe de chaque coefficient $c_i(\alpha,\pi,\pi)$ nous permet d'affirmer que les valeurs propres de la matrice $N_{0,0}$ sont positives si et seulement si

$$\alpha \geq \frac{3311}{1976}-\frac{160}{741}\sqrt{17}+\frac{1}{5928}\sqrt{80990249-18599424\sqrt{17}} \simeq 1.14. \tag{3.11}$$

– Enfin, intéressons-nous au cas $\beta_1 = 0$ et $\beta_2 = \pi$. Dans ce cas, la matrice $N_{0,\pi}$ a pour polynôme caractéristique

$$p_\alpha(\lambda,0,\pi) = \lambda^6+\sum_{i=0}^{5} c_i(\alpha,0,\pi)\,\lambda^i$$

où les coefficients $c_i(\alpha, 0, \pi)$ sont définis par

$$\begin{cases} c_5(\alpha,\pi,\pi) = & -5\left(9+\sqrt{17}\right)\alpha, \\ c_4(\alpha,\pi,\pi) = & 28\left(26+7\sqrt{17}\right)\alpha^2 + \frac{27}{4}\left(7\sqrt{17}+139\right)\alpha - \frac{29817}{16}, \\ c_3(\alpha,\pi,\pi) = & -15\left(401+149\sqrt{17}\right)\alpha^3 - \frac{9}{4}\left(11135+2333\sqrt{17}\right)\alpha^2 \\ & +\frac{9}{16}\left(11199\sqrt{17}+87551\right)\alpha, \\ c_2(\alpha,\pi,\pi) = & 3\left(10941+2140\sqrt{17}\right)\alpha^4 + \frac{27}{4}\left(12527\sqrt{17}+29131\right)\alpha^3 \\ & -\frac{9}{32}\left(277789\sqrt{17}+618877\right)\alpha^2 - 324\left(174\sqrt{17}+2275\right)\alpha + 557280, \\ c_1(\alpha,\pi,\pi) = & -540\left(20\sqrt{17}+51\right)\alpha^5 - 9\left(\frac{111979}{2}+17280\sqrt{17}\right)\alpha^4 \\ & -\frac{27}{16}\left(622791+112795\sqrt{17}\right)\alpha^3 + \frac{81}{8}\left(\frac{983053}{4}+74013\sqrt{17}\right)\alpha^2 \\ & -\frac{81}{256}\left(250073\sqrt{17}+6564321\right)\alpha, \\ c_0(\alpha,\pi,\pi) = & 3240\left(20\sqrt{17}+323\right)\alpha^5 + 297\left(14607+5395\sqrt{17}\right)\alpha^4 \\ & -486\left(27323+931\sqrt{17}\right)\alpha^3 - 648\left(13123\sqrt{17}+24291\right)\alpha^2 \\ & +3732480\left(2\sqrt{17}+13\right)\alpha - 26873856. \end{cases}$$

Grâce à l'étude du signe de ces six coefficients, il vient que $N_{0,\pi}$ est positive si et seulement si

$$\alpha \geq \frac{1}{34422}\left(19183+5011\sqrt{17}+\sqrt{1191400986-184462342\sqrt{17}}\right) \simeq 1.76. \quad (3.12)$$

Si l'on regroupe les conditions (3.10) à (3.12), il vient qu'une condition nécessaire afin d'assurer que les valeurs propres de N_{β_1,β_2} sont positives est que

$$\alpha \geq \frac{1}{8}\left(9+\sqrt{17}+\frac{3}{85}\sqrt{37570-7310\sqrt{17}}\right).$$

Remarque 3.A.3. *Les expériences numériques que nous avons effectuées sur de tels maillages confirment bien ce résultat puisque numériquement, nous avons stabilité lorsque* $\alpha \geq 2.02$. *Dans cette configuration également, c'est le cas* $\beta_1 = \beta_2 = 0$ *qui permet de conclure quant à la stabilité du schéma.*

Chapitre 4

Analyse de stabilité du Δ^2-schéma

Dans les deux chapitres précédents, nous avons déterminé des conditions nécessaires de stabilité pour la méthode IPDG classique appliquée à l'équation des ondes. A présent, nous nous intéressons à l'analyse de stabilité du Δ^2-schéma que nous avons proposé dans le chapitre 1. Plus précisément, il s'agit de déterminer comment choisir les différents paramètres de pénalisation intervenant dans le Δ^2-schéma. Nous nous intéressons également à la dépendance de la condition CFL vis à vis de ces paramètres. En particulier, on peut se demander si le paramètre de pénalisation γ_1 propre à l'opérateur harmonique suffit seul à assurer la stabilité du schéma sous une certaine condition CFL. On peut également chercher à définir l'influence du paramètre de pénalisation $\gamma_{2,1}$ de l'opérateur biharmonique sur la stabilité du schéma ainsi que sur le comportement de la condition CFL. Ainsi, dans la suite, nous allons utiliser une méthodologie similaire à celle employée dans le chapitre 2 afin d'étudier la stabilité du Δ^2-schéma. Dans la section 4.2, nous proposerons une analyse de stabilité dans le cas où tous les paramètres de pénalisation (liés aux opérateurs harmonique et biharmonique) sont fixés d'après les résultats numériques du chapitre 1, c'est-à-dire le cas où $\gamma_{1,1} = {}^{8}/_{h}$ et $\gamma_{2,1} = \gamma_{2,2} = 0$ et celui où $\gamma_{1,1} = {}^{8}/_{h}$, $\gamma_{2,1} = {}^{10}/_{h}$ et $\gamma_{2,2} = 0$. Le lecteur intéressé par les conditions CFL obtenus avec ces deux jeux de paramètres peut sauter les démonstrations dans une première lecture et aller directement à la conclusion. Nous avons aussi calculé une condition nécessaire de stabilité en fonction de γ_1 que nous présentons dans l'annexe 4.A. Ces résultats ont été obtenus pour $\gamma_{2,1} = \gamma_{2,2} = 0$ puis pour $\gamma_{2,1} = 10$ et $\gamma_{2,2} = 0$. Enfin, des expériences numériques illustrerons les résultats présentés.

4.1 Préliminaires

Comme dans le chapitre 2, nous considérons le domaine $\Omega = \mathbb{R}$ et une vitesse de propagation constante $c^2 = {}^{\mu}/_{\rho} = 1$ mais l'analyse peut être étendue à des valeurs arbitraires de c en utilisant le changement de variable $\Delta t' = c\Delta t$. On rappelle que le Δ^2-schéma est de la forme

$$M\frac{U^{n+1} - 2U^n + U^{n-1}}{\Delta t^2} + K^* U^n = 0, \tag{4.1}$$

avec $K^* = K_1 - \dfrac{\Delta t^2}{12}K_2$ et K_1 et K_2 sont les matrices de raideur relatives respectivement aux opérateurs harmonique et biharmonique.
On rappelle la notation introduite au chapitre 2 :

$$\gamma_1 = \frac{\alpha_1}{h}$$

et on pose, par analogie :

$$\gamma_{2,1} = \frac{\alpha_{2,1}}{h} \qquad \text{et} \qquad \gamma_{2,2} = \frac{\alpha_{2,2}}{h}.$$

Nous avons vu au chapitre 1 que ce schéma conservait l'énergie

$$E^{n+\frac{1}{2}} = \left(\left(M - \frac{\Delta t^2}{4}K^*\right)\frac{U^{n+1}-U^n}{\Delta t}, \frac{U^{n+1}-U^n}{\Delta t}\right) + \left(K^*\frac{U^{n+1}+U^n}{2}, \frac{U^{n+1}+U^n}{2}\right)$$

et que sa stabilité était assurée par la positivité des matrices $M - \dfrac{\Delta t^2}{4}K^*$ et K^*. Le théorème suivant garantit l'existence d'une condition de stabilité pour $\alpha_1 > 6$.

Théorème 4.1.1. *Si $\alpha_1 > 6$, alors le Δ^2-schéma est stable sous une condition CFL, c'est-à-dire qu'il existe $\Delta t_0 > 0$ tel que K^* et $M - \dfrac{\Delta t^2}{4}K^*$ soient positives pour tout $\Delta t < \Delta t_0$.*

Démonstration. Remarquons tout d'abord que K_1 et K_2 admettent toutes deux une valeur propre nulle associée au vecteur constant. De plus, si $\alpha_1 > 6$, toutes les valeurs propres de K_1 sont positives. Comme K_2 est symétrique, toutes ses valeurs propres sont réelles. Nous notons $\kappa_1^{\min}$ la plus petite valeur propre non nulle de K_1 et $\kappa_2^{\max}$ la plus grande valeur propre de K_2.
D'après les remarques précédentes $K^* = K_1 - \dfrac{\Delta t^2}{12}K_2$ est positive si $\kappa_1^{\min} - \dfrac{\Delta t^2}{12}\kappa_2^{\max} \geq 0$.
Cette condition est toujours vraie si $\kappa_2^{\max} \leq 0$. Si $\kappa_2^{\max} > 0$, K^* est positive si

$$\Delta t \leq \Delta t_1 = 2\sqrt{\frac{3\kappa_1^{\min}}{\kappa_2^{\max}}}.$$

On note maintenant $\mu_1^{\min}$ la plus petite valeur propre de M, $\kappa_1^{\max}$ la plus grande valeur propre de K_1 et $\kappa_2^{\min}$ la plus petite valeur propre de K_2. On vérifie facilement que $\mu_1^{\min} > 0$ (M est définie positive).
Pour que $M - \dfrac{\Delta t^2}{4}K^*$ soit positive, il suffit que

$$\mu_1^{\min} - \frac{\Delta t^2}{4}\kappa_1^{\max} + \frac{\Delta t^4}{48}\kappa_2^{\min} \geq 0. \tag{4.2}$$

Comme $\mu_1^{\min} > 0$, il existe $\Delta t_2 > 0$ tel que cette inégalité soit vraie pour tout $\Delta t \in [0; \Delta t_2[$.
En effet, en notant X_1 et X_2 les racines de $P(X) = \mu_{\min} - \dfrac{X}{4}\kappa_1^{\max} + \dfrac{X^2}{48}\kappa_2^{\min}$.
Si ces deux racines sont complexes ou négatives, l'inégalité (4.2) est toujours vérifiée. En effet, dans ce cas $P(X)$ ne s'annule pas pour $X \geq 0$ et reste donc du signe de $P(0) = \mu_{\min}$.
Si une seule racine est positive, par exemple X_1, alors $P(X)$ est du signe de $P(0) = \mu_{\min}$ pour $0 \leq X \leq X_1$ et l'inégalité (4.2) est vraie pour tout $\Delta t \leq \sqrt{X_1}$.
Si les deux racines sont positives, on suppose sans perte de généralité que $X_1 \leq X_2$ et on montre à nouveau que l'inégalité (4.2) est vraie pour tout $\Delta t \leq \sqrt{X_1}$. Finalement, le schéma est stable pour tout $\Delta t \leq \Delta t_0 = \min(\Delta t_1, \Delta t_2)$. □

Remarque 4.1.2. *Ce théorème montre que les paramètres de pénalisation associés à l'opérateur biharmonique ne sont pas utiles à la stabilité du schéma. En effet, nous n'avons pas besoin de la positivité de K_2 pour démontrer le théorème. Néanmoins, nous verrons par la suite qu'un choix judicieux de ces paramètres permet d'optimiser la condition CFL.*

La complexité des systèmes à résoudre ne nous a pas permis d'expliciter la dépendance de Δt_0 par rapport à α_1, $\alpha_{2,1}$ et $\alpha_{2,2}$ et nous nous sommes limités au calcul de bornes supérieures de Δt_0 pour certaines valeurs particulières des paramètres de pénalisation. Pour cela, en suivant la technique développée au chapitre 2, nous considérons la restriction de l'équation (4.1) à un élément $J \in \mathcal{T}_h$ du maillage :

$$M_{1,p}\frac{U_J^{n+1}-2U_J^n+U_J^{n-1}}{\Delta t^2}+\left(K_p^{*,E}\right)^T U_{J-1}^n+K_p^* U_J^n+K_p^{*,E}U_{J+1}^n=0 \tag{4.3}$$

avec $K_p^*=K_{1,p}-\dfrac{\Delta t^2}{12}K_{2,p}$ et $K_p^{*,E}=K_{1,p}^E-\dfrac{\Delta t^2}{12}K_{2,p}^E$.
$M_{1,p}$, $K_{1,p}$ et $K_{1,p}^E$ sont les matrices de masse et de raideur en dimension un déjà définies dans le chapitre 2. $K_{2,p}$ et $K_{2,p}^E$ correspondent aux matrices de raideurs issue de l'opérateur biharmonique en dimension un en considérant des polynômes de degré p :

$$\begin{aligned}
K_{2,p}(i,j) &= \frac{1}{h^3}\int_{[0,1]}\frac{\partial^2\hat\varphi_i}{\partial\hat x^2}(\hat x)\frac{\partial^2\hat\varphi_j}{\partial\hat x^2}(\hat x)\,d\hat x-\frac{1}{2h^3}\frac{\partial^2\hat\varphi_i}{\partial\hat x^2}(1)\frac{\partial\hat\varphi_j}{\partial\hat x}(1)-\frac{1}{2h^3}\frac{\partial^2\hat\varphi_j}{\partial\hat x^2}(1)\frac{\partial\hat\varphi_i}{\partial\hat x}(1)\\
&\quad+\frac{1}{2h^3}\frac{\partial^2\hat\varphi_i}{\partial\hat x^2}(0)\frac{\partial\hat\varphi_j}{\partial\hat x}(0)+\frac{1}{2h^3}\frac{\partial^2\hat\varphi_j}{\partial\hat x^2}(0)\frac{\partial\hat\varphi_i}{\partial\hat x}(0)+\frac{1}{2h^3}\frac{\partial^3\hat\varphi_i}{\partial\hat x^3}(1)\hat\varphi_j(1)\\
&\quad+\frac{1}{2h^3}\frac{\partial^3\hat\varphi_j}{\partial\hat x^3}(1)\hat\varphi_i(1)-\frac{1}{2h^3}\frac{\partial^3\hat\varphi_i}{\partial\hat x^3}(0)\hat\varphi_j(1)-\frac{1}{2h^3}\frac{\partial^3\hat\varphi_j}{\partial\hat x^3}(0)\hat\varphi_i(0)\\
&\quad+\frac{\alpha_{2,1}}{h^3}\frac{\partial\hat\varphi_i}{\partial\hat x}(1)\frac{\partial\hat\varphi_j}{\partial\hat x}(1)+\frac{\alpha_{2,1}}{h^3}\frac{\partial\hat\varphi_i}{\partial\hat x}(0)\frac{\partial\hat\varphi_j}{\partial\hat x}(0)+\frac{\alpha_{2,2}}{h^3}\hat\varphi_i(1)\hat\varphi_j(1)\\
&\quad+\frac{\alpha_{2,2}}{h^3}\hat\varphi_i(0)\hat\varphi_j(0),\\
K_{2,p}^E(i,j) &= \frac{1}{2h^3}\frac{\partial^2\hat\varphi_i}{\partial\hat x^2}(1)\frac{\partial\hat\varphi_j}{\partial\hat x}(0)-\frac{1}{2h^3}\frac{\partial^2\hat\varphi_j}{\partial\hat x^2}(0)\frac{\partial\hat\varphi_j}{\partial\hat x}(1)-\frac{1}{2h^3}\frac{\partial^3\hat\varphi_i}{\partial\hat x^3}(0)\hat\varphi_j(1)\\
&\quad+\frac{1}{2h^3}\frac{\partial^3\hat\varphi_j}{\partial\hat x^3}(1)\hat\varphi_i(0)-\frac{\alpha_{2,1}}{h^3}\frac{\partial\hat\varphi_i}{\partial\hat x}(1)\frac{\partial\hat\varphi_j}{\partial\hat x}(0)-\frac{\alpha_{2,2}}{h^3}\hat\varphi_i(1)\hat\varphi_j(0)
\end{aligned} \tag{4.4}$$

où $\{\hat\varphi_i\}_{i=1,\dots,p+1}$ sont les fonctions de base discontinues classiques de Lagrange sur l'élément de référence $[0,1]$.
En appliquant à (4.3) la transformée de Fourier discrète définie au chapitre 2, nous obtenons, $\forall\beta\in[-\pi,\pi]$

$$M_{1,p}\frac{\tilde U_J^{n+1}(\beta)-2\tilde U_J^n(\beta)+\tilde U_J^{n-1}(\beta)}{\Delta t^2}+K_\beta^*\tilde U_J^n(\beta)=0 \tag{4.5}$$

où $\beta=hk$ et $K_\beta^*=\left(K_p^{*,E}\right)^T e^{-i\beta}+K_p^*+K_p^{*,E}e^{i\beta}$.
La stabilité L^2 de (4.5) pour tout $\beta\in[-\pi,\pi]$, est équivalente à la stabilité L^2 de (4.1), grâce aux égalités de Parseval. Or si l'on applique une analyse de stabilité classique, comme nous l'avons fait au chapitre 1 dans la preuve du théorème 1.2.8, la stabilité L^2 du schéma est assurée si et seulement si les matrices $M_{1,p}-\dfrac{\Delta t^2}{12}K_\beta^*$ et K_β^* sont positives.

Nous devons donc vérifier que toutes les valeurs propres de ces deux matrices sont bien positives. Néanmoins, contrairement à ce que nous avons fait au chapitre 2, il n'y a plus une seule inconnue mais quatre : les trois coefficients de pénalisation (α_1, $\alpha_{2,1}$ et $\alpha_{2,2}$) ainsi que le pas de temps Δt. Nous n'allons donc pas être en mesure de trouver une condition de stabilité sans fixer certains de ces paramètres. Par conséquent, nous allons considérer plusieurs cas qui nous semblent être caractéristiques, en fixant les valeurs de certains paramètres de pénalisation :

1. On pose $\alpha_1 = 8$: α_1 est ainsi dans le palier déterminé au chapitre 2, et on fixe les deux autres coefficients de pénalisation à zéro afin de vérifier si l'on peut assurer la stabilité du schéma en se contentant du coefficient de pénalisation provenant de l'opérateur laplacien,
2. On prend $\alpha_1 = 8$, $\alpha_{2,1} = 10$ et $\alpha_{2,2} = 0$, la valeur de $\alpha_{2,1}$ qui nous a semblé judicieuse par rapport aux expériences numériques que nous avons effectuées et présentées dans le chapitre 1,
3. On fixe $\alpha_{2,1}$ et $\alpha_{2,2}$ à 0 et on étudie la stabilité du schéma en fonction de α_1,
4. On fixe $\alpha_{2,1} = 10$ et $\alpha_{2,2} = 0$ et on étudie la stabilité du schéma en fonction de α_1.

Les deux premiers points feront l'objet de la section 4.2 et les deux suivants de la section 4.A.

Remarque 4.1.3. *Ici, nous ne présenterons que le cas $p = 3$, les cas $p < 3$ n'ayant pas d'intérêt puisque pour de tels polynômes, la forme bilinéaire $a_{2,h}$ associée à l'opérateur bilaplacien est nulle.*

Dans les sections suivantes, nous cherchons Δt_0 tel que les valeurs propres de matrices K_β^* et $M - \dfrac{\Delta t^2}{12} K_\beta^*$ soient positives pour tout $\beta \in [-\pi, \pi]$ et pour tout $\Delta t \leq \Delta t_0$. Pour cela, nous devons considérer les polynômes caractéristiques de K_β^* et $M - \dfrac{\Delta t^2}{12} K_\beta^*$ notés respectivement $q_{1,\Delta t}$ et $q_{2,\Delta t}$ qui vérifient, pour $j \in \{1, 2\}$

$$q_{j,\Delta t}(\beta, \lambda) = (-1)^{p+1} \lambda^{p+1} + \sum_{i=0}^{3} c_{j,i}(\Delta t, \beta) \lambda^i.$$

D'après le lemme 2.2.1, montrer la positivité des racines de ces deux polynômes revient à montrer que pour $j \in \{1, 2\}$, $i \in \{0, \ldots, 3\}$ et pour tout $\beta \in [-\pi, \pi]$,

$$(-1)^i c_{j,i}(\Delta t, \beta) \geq 0.$$

Malheureusement, ces coefficients sont bien plus compliqués que ceux obtenus dans le chapitre 2 et on ne peut faire une analyse aussi fine. Cependant, nous avons vu dans la section 2.2.3, que les informations obtenues en considérant les cas $\beta = 0$ et $\beta = \pi$ donnent quasiment l'essentiel du comportement de la condition CFL hormis pour un petit intervalle de valeurs que peut prendre le coefficient de pénalisation. Par conséquent, nous allons effectuer l'analyse dans les cas où $\beta = 0$ et $\beta = \pi$ et, par la suite, nous analyserons la pertinence de ces résultats grâce aux résultats numériques que nous présenterons.

4.2 Une condition CFL pour α_1 fixé

Dans cette section, nous étudions la stabilité du schéma (4.1) en fixant les trois paramètres de pénalisation pour $p = 3$.
Le premier théorème nous fournit une condition nécessaire de stabilité dans le cas $\alpha_1 = 8$, $\alpha_{2,1} = \alpha_{2,2} = 0$.

Théorème 4.2.1. *Si $\alpha_1 = 8$ et $\alpha_{2,1} = \alpha_{2,2} = 0$, alors*

$$\Delta t_0 \leq h\sqrt{X_0} \simeq 0.145h$$

où X_0 est la plus petite racine positive du polynôme

$$9X^7 - \frac{63}{4}X^6 + \frac{93}{8}X^5 - \frac{501}{140}X^4 + \frac{39}{280}X^3 + \frac{41}{200}X^2 - \frac{39}{3500}X + \frac{1}{7000}.$$

Nous avons ensuite considéré le cas $\alpha_1 = 8$, $\alpha_{2,2} = 0$ et $\alpha_{2,1} = 10$ qui nous a paru le plus optimal lors des tests numériques, ce qui nous a permis de prouver le théorème suivant.

Théorème 4.2.2. *Si $\alpha_1 = 8$, $\alpha_{2,1} = 10$ et $\alpha_{2,2} = 0$, alors*

$$\Delta t_0 \leq h\sqrt{X_0} \simeq 0.1822h$$

où X_0 est la première racine positive du polynôme

$$21X^7 - \frac{171}{4}X^6 + \frac{2139}{56}X^5 - \frac{35}{2}X^4 + \frac{1167}{280}X^3 + \frac{457}{140}X^2 - \frac{39}{3500}X + \frac{1}{7000}.$$

Remarque 4.2.3.

- *Nous n'avons obtenu que des conditions nécessaires. Néanmoins les tests numériques montrent qu'en pratique les conditions sont suffisantes.*
- *En comparant les bornes obtenues dans les théorèmes 4.2.1 et 4.2.2, on constate que celle du théorème 4.2.2 est moins restrictive, ce qui correspond aux résultats numériques du chapitre 1.*
- *Nous rappelons que d'après le théorème 2.1.1, le schéma saute-moutons est stable dans un certain intervalle de valeur de α_1 si $\Delta t \leq \sqrt{\frac{2}{45+\sqrt{1605}}}h \simeq 0.153h$. Si $\alpha_1 = 8$, $\alpha_{2,1} = 10$ et $\alpha_{2,2} = 0$, on a donc une condition CFL moins restrictive avec le Δ^2-schéma qu'avec le schéma saute-moutons, ce qui est observable sur les tests numériques du chapitre 1.*

4.2.1 Preuve du théorème 4.2.1

Nous posons ici $\alpha_1 = 8$ et $\alpha_{2,1} = \alpha_{2,2} = 0$. Dans la première sous-section, nous étudions la positivité des valeurs propres de la matrice K_β^* tandis que la deuxième sous-section est consacrée à l'étude de la matrice $M - \frac{\Delta t^2}{4}K_\beta^*$.

4.2.1.1 Positivité de K_β^*

Nous nous limitons à l'étude de la positivité de K_0^* et K_π^*, ce qui explique que nous n'obtenons qu'une condition nécessaire de stabilité.
Supposons tout d'abord que $\beta = 0$. Dans de telles conditions, le polynôme caractéristique de K_0^* s'écrit

$$q_{1,\Delta t}(0,\lambda) = \lambda^4 + \sum_{i=0}^{3} c_{1,i}(\Delta t, 0)\,\lambda^i$$

avec

$$\begin{cases} c_{1,3}(\Delta t,0) = & \frac{207}{2}\frac{\Delta t^2}{h^2} - 32, \\ c_{1,2}(\Delta t,0) = & -\frac{14823}{16}\frac{\Delta t^4}{h^4} - 2214\frac{\Delta t^2}{h^2} + \frac{19467}{80}, \\ c_{1,1}(\Delta t,0) = & \frac{19683}{16}\frac{\Delta t^6}{h^6} + \frac{98415}{8}\frac{\Delta t^4}{h^4} + \frac{846369}{80}\frac{\Delta t^2}{h^2} - \frac{19683}{40}, \\ c_{1,0}(\Delta t,0) = & 0. \end{cases}$$

Rappelons que nous cherchons une condition sur Δt pour que $(-1)^i c_{1,i}(\Delta t, 0) \geq 0$, $\forall i \in \{0, \ldots, 3\}$.

- Etudions tout d'abord la condition sur $c_{1,3}$. On a,

$$-c_{1,3}(\Delta t, 0) \geq 0 \quad \Leftrightarrow \quad 32 - \frac{207}{2}\frac{\Delta t^2}{h^2} \geq 0$$

Par conséquent, $-c_{1,3}(\Delta t, 0) \geq 0$, si et seulement si

$$\frac{\Delta t^2}{h^2} \leq \frac{64}{207} \simeq 0.31 \tag{4.6}$$

- Considérons à présent la condition $c_{1,2}(\Delta t, 0) \geq 0$ qui se réécrit, en posant $X = \dfrac{\Delta t^2}{h^2}$,

$$f(X) := -\frac{14823}{16}X^2 - 2214X + \frac{19467}{80} \geq 0. \tag{4.7}$$

Ce polynôme du second degré admet deux racines :

$$\begin{cases} X_1 = -\dfrac{656}{549} - \dfrac{1}{2745}\sqrt{12737545}, \\ X_2 = -\dfrac{656}{549} + \dfrac{1}{2745}\sqrt{12737545}. \end{cases}$$

Le coefficient de plus haut degré de f étant négatif, f est positif si $X \in [X_1, X_2]$. Comme $X_1 < 0$, il s'en suit que $c_{1,2}(\Delta t, 0) \geq 0$, si et seulement si

$$\frac{\Delta t^2}{h^2} \leq -\frac{656}{549} + \frac{1}{2745}\sqrt{12737545} \simeq 0.11. \tag{4.8}$$

- Maintenant, étudions le signe de $c_{1,1}(\Delta t, 0)$.
La condition $-c_{1,1}(\Delta t, 0) \geq 0$ est équivalente à

$$f(X) := \frac{19683}{16}X^3 + \frac{98415}{8}X^2 + \frac{846369}{80}X - \frac{19683}{40} \leq 0$$

avec $X = \dfrac{\Delta t^2}{h^2}$.
Le polynôme f admet les trois racines suivantes :

$$\begin{cases} X_1 = -1, \\ X_2 = -\dfrac{9}{2} - \dfrac{1}{10}\sqrt{2065}, \\ X_3 = -\dfrac{9}{2} + \dfrac{1}{10}\sqrt{2065}. \end{cases}$$

Seule X_3 est positive et le coefficient de plus haut degré de f est positif donc, $-c_{1,1}(\Delta t, 0) \geq 0$ si et seulement si

$$\frac{\Delta t^2}{h^2} \leq -\frac{9}{2} + \frac{1}{10}\sqrt{2065} \simeq 0.044. \tag{4.9}$$

- Finalement, la positivité de $c_{1,0}(\Delta t, 0)$ est triviale puisque $c_{1,0}(\Delta t, 0) = 0$.

En résumé, une condition nécessaire et suffisante de positivité de K_β^* pour $\beta = 0$ est

$$\frac{\Delta t^2}{h^2} \leq -\frac{9}{2} + \frac{1}{10}\sqrt{2065} \simeq 0.044.$$

A présent, nous allons supposer que $\beta = \pi$. Dans ce cas, les coefficients du polynôme caractéristique $q_{1,\Delta t}(\pi, \lambda)$ sont

$$\begin{cases} c_{1,3}(\Delta t, \pi) = & -\frac{207}{2}\frac{\Delta t^2}{h^2} - 36, \\ c_{1,2}(\Delta t, \pi) = & -\frac{14823}{16}\frac{\Delta t^4}{h^4} + 1494\frac{\Delta t^2}{h^2} + \frac{30091}{80}, \\ c_{1,1}(\Delta t, \pi) = & -\frac{19683}{16}\frac{\Delta t^6}{h^6} + \frac{77517}{8}\frac{\Delta t^4}{h^4} - \frac{382347}{80}\frac{\Delta t^2}{h^2} - \frac{10017}{8}, \\ c_{1,0}(\Delta t, \pi) = & \frac{19683}{2}\frac{\Delta t^6}{h^6} - \frac{255879}{16}\frac{\Delta t^4}{h^4} + \frac{334611}{80}\frac{\Delta t^2}{h^2} + \frac{19683}{16}. \end{cases}$$

- La condition $-c_{1,3}(\Delta t, 0) \geq 0$ est trivialement vérifiée.
- En posant $X = \dfrac{\Delta t^2}{h^2}$, la condition sur $c_{1,2}$ nous donne

$$f(X) := -\frac{14823}{16}X^2 + 1494X + \frac{30091}{80} \geq 0 \tag{4.10}$$

qui admet deux racines :

$$\begin{cases} X_1 = \dfrac{1328}{1647} - \dfrac{1}{8235}\sqrt{71622865}, \\ X_2 = \dfrac{1328}{1647} + \dfrac{1}{8235}\sqrt{71622865}. \end{cases}$$

Le coefficient de plus haut degré de $f(X)$ est négatif. Donc, f est positif si $X \in [X_1, X_2]$. Or, $X_1 < 0$, par conséquent $c_{1,2}(\Delta t, \pi) \geq 0$, si et seulement si

$$\frac{\Delta t^2}{h^2} \leq \frac{1328}{1647} + \frac{1}{8235}\sqrt{71622865} \simeq 1.83. \tag{4.11}$$

- Maintenant, étudions le signe de $c_{1,1}(\Delta t, \pi)$.
  En posant $X = \dfrac{\Delta t^2}{h^2}$, la condition $-c_{1,1}(\Delta t, \pi) \geq 0$ est équivalente à,

$$f(X) := -\frac{19683}{16}X^3 + \frac{77517}{8}X^2 - \frac{382347}{80}X - \frac{10017}{8} \leq 0.$$

En utilisant Maple, on obtient les trois racines suivantes :

$$\begin{cases} X_1 = -2a\cos\left(b - \dfrac{\pi}{3}\right) + \dfrac{638}{243} \simeq -0.188, \\ X_2 = -2a\cos\left(b + \dfrac{\pi}{3}\right) + \dfrac{638}{243} \simeq 0.738, \\ X_3 = 2a\cos(b) + \dfrac{638}{243} \simeq 7.327, \end{cases}$$

où $a = \dfrac{1}{1215}\sqrt{5}\sqrt{1652873}$, $b = \dfrac{1}{3}\arctan\left(\dfrac{27\sqrt{3}\sqrt{5}}{4480261580}\sqrt{229118584411451}\right)$.

Le coefficient de plus haut degré de f étant négatif, la positivité de $-c_{1,1}(\Delta t, \pi)$ est équivalente à $\dfrac{\Delta t^2}{h^2} \in [0; X_2] \cup [X_3; +\infty[$.

Il est clair que la condition $\dfrac{\Delta t^2}{h^2} \geq X_3 \simeq 7.327$ est incompatible avec la condition (4.11). Nous considérons donc uniquement le cas

$$\frac{\Delta t^2}{h^2} \leq X_2 \simeq 0.738. \tag{4.12}$$

- Enfin, la positivité de $c_{1,0}(\Delta t, \pi)$ est équivalente à

$$f(X) := \frac{19683}{2}X^3 - \frac{255879}{16}X^2 + \frac{334611}{80}X + \frac{19683}{16} \geq 0$$

avec $X = \dfrac{\Delta t^2}{h^2}$.

Le polynôme f admet les trois racines suivantes :

$$\begin{cases} X_1 = \dfrac{1}{2} - \dfrac{3}{10}\sqrt{5}, \\ X_2 = \dfrac{5}{8}, \\ X_3 = \dfrac{1}{2} + \dfrac{3}{10}\sqrt{5}. \end{cases}$$

La première racine est négative, les deux autres positives. Le coefficient de plus haut degré de f étant positif, la positivité de $c_{1,0}(\Delta t, \pi)$ est équivalente à $\dfrac{\Delta t^2}{h^2} \in [0; X_2] \cup [X_3; +\infty[$.

La condition $\dfrac{\Delta t^2}{h^2} \geq X_3 \simeq 1.17$ est incompatible avec (4.12) et nous nous ne retenons donc que la condition

$$\frac{\Delta t^2}{h^2} \leq X_2 \simeq 0.625. \tag{4.13}$$

Une condition nécessaire et suffisante de positivité de K^*_β pour $\beta = \pi$ est donc

$$\frac{\Delta t^2}{h^2} \leq X_2 \simeq 0.625.$$

et la condition la plus restrictive est celle obtenue pour $\beta = 0$

$$\frac{\Delta t^2}{h^2} \leq -\frac{9}{2} + \frac{1}{10}\sqrt{2065} \simeq 0.044$$

Une condition nécessaire de positivité de K^*_β est donc

$$\frac{\Delta t}{h} \leq \sqrt{-\frac{9}{2} + \frac{1}{10}\sqrt{2065}} \simeq 0.210.$$

Cette condition est uniquement nécessaire car nous n'avons pas étudié la positivité de K^*_β pour tout β de $[-\pi; \pi]$. Nous avons déterminé numériquement la condition garantissant la positivité de K^*.

Algorithme 3

$\Delta t_1 = 0$, $\Delta t_2 = 1$ et $\epsilon = 10^{-8}$
$\Delta t = \dfrac{\Delta t_1 + \Delta t_2}{2}$
$K^* = K_1 - \dfrac{\Delta t^2}{12} K_2$
Calcul de $\lambda_{\min}$, la plus petite des valeurs propres de K^*
tantque $|\Delta t_1 - \Delta t_2| > \epsilon$ ou $\lambda_{\min} < 0$ **faire**
  si $\lambda_{\min} < 0$ **alors**
    $\Delta t_2 = \Delta t$
  sinon
    $\Delta t_1 = \Delta t$
  finsi
  $\Delta t = \dfrac{\Delta t_1 + \Delta t_2}{2}$
  $K^* = K_1 - \dfrac{\Delta t^2}{12} K_2$
fin tantque
$\Delta t = \min(\Delta t_1, \Delta t_2)$

Pour cela nous avons repris le code 1D présenté au chapitre 1 et nous avons utilisé l'algorithme 3 qui permet de déterminer ce pas de temps par dichotomie.

Nous avons obtenu la condition numérique

$$\frac{\Delta t}{h} \leq 0.211.$$

qui semble montrer que l'étude mathématique restreinte aux cas $\beta = 0$ et $\beta = \pi$ soit suffisante.

4.2.1.2 Positivité de $M - \dfrac{\Delta t^2}{4} K_\beta^*$

On va à présent s'intéresser à la positivité des valeurs propres de la matrice $M - \dfrac{\Delta t^2}{12} K_\beta^*$. Comme dans le cas précédent, nous nous sommes limités aux cas $\beta = 0$ et $\beta = \pi$.
Dans un premier temps, nous allons supposer que $\beta = 0$. Avec un tel choix, les coefficients de $q_{2,\Delta t}(0, \lambda)$ sont

$$\begin{cases}
c_{2,3}(\Delta t, 0) = & -\dfrac{207}{8}\dfrac{\Delta t^4}{h^4} + 8\dfrac{\Delta t^2}{h^2} - \dfrac{97}{105}, \\
c_{2,2}(\Delta t, 0) = & -\dfrac{14823}{256}\dfrac{\Delta t^8}{h^8} - \dfrac{1107}{8}\dfrac{\Delta t^6}{h^6} + \dfrac{277533}{8960}\dfrac{\Delta t^4}{h^4} - \dfrac{10541}{2240}\dfrac{\Delta t^2}{h^2} + \dfrac{8797}{33600}, \\
c_{2,1}(\Delta t, 0) = & -\dfrac{19683}{1024}\dfrac{\Delta t^{12}}{h^{12}} - \dfrac{98415}{512}\dfrac{\Delta t^{10}}{h^{10}} - \dfrac{716121}{5120}\dfrac{\Delta t^8}{h^8} + \dfrac{173853}{2560}\dfrac{\Delta t^6}{h^6} - \dfrac{776673}{89600}\dfrac{\Delta t^4}{h^4} \\
& + \dfrac{837}{1120}\dfrac{\Delta t^2}{h^2} - \dfrac{93}{4000}, \\
c_{2,0}(\Delta t, 0) = & \dfrac{19683}{4096}\dfrac{\Delta t^{12}}{h^{12}} + \dfrac{98415}{2048}\dfrac{\Delta t^{10}}{h^{10}} + \dfrac{793881}{20480}\dfrac{\Delta t^8}{h^8} - \dfrac{137781}{20480}\dfrac{\Delta t^6}{h^6} + \dfrac{422091}{716800}\dfrac{\Delta t^4}{h^4} \\
& - \dfrac{2187}{71680}\dfrac{\Delta t^2}{h^2} + \dfrac{2187}{3584000}.
\end{cases}$$

Maintenant, étudions le signe de chacun de ces coefficients. Ces derniers étant des polynômes de plus haut degré que dans la sous-section précédente, nous avons utilisé le logiciel Maple afin de déterminer leurs racines.

- Vérifions que $-c_{2,3}(\Delta t, 0) \geq 0$. Cela revient à vérifier l'inéquation du second degré
$$\frac{207}{8}X^2 - 8X + \frac{97}{105} \geq 0$$
où $X = \dfrac{\Delta t^2}{h^2}$. Comme le discriminant de cette équation est strictement négatif, on conclut qu'elle n'admet pas de racine réelle et que la condition est toujours vérifiée.
- Intéressons nous à la condition sur $c_{2,2}$. On veut établir sous quelle condition
$$-\frac{14823}{256}\frac{\Delta t^8}{h^8} - \frac{1107}{8}\frac{\Delta t^6}{h^6} + \frac{277533}{8960}\frac{\Delta t^4}{h^4} - \frac{10541}{2240}\frac{\Delta t^2}{h^2} + \frac{8797}{33600} \geq 0.$$
On effectue le changement de variable $X = \dfrac{\Delta t^2}{h^2}$, et, d'après Maple, ce polynôme admet deux racines complexes conjuguées et deux racines réelles $X_1 \simeq -2.6072$ et $X_2 \simeq 0.0841$. Le coefficient correspondant au terme de plus haut degré est positif.On en déduit donc que $c_{2,2}$ est positif entre X_1 et X_2. Par conséquent, la positivité de $c_{2,2}(\Delta t, 0)$ est équivalente à
$$\frac{\Delta t^2}{h^2} \leq 0.0841. \tag{4.14}$$
- De la même façon que précédemment, pour la condition sur $c_{2,1}$, Maple nous donne deux racines complexes conjuguées ainsi que les quatre racines réelles
$$\begin{cases} X_1 = -9.1632, & X_2 = -1.1997, \\ X_3 = 0.0500, & X_4 = 0.2292. \end{cases}$$
Le terme de plus haut degré du polynôme $c_{2,1}$ étant négatif, on peut donc en conclure que $-c_{2,1}(\Delta t, 0)$ est positif si et seulement si $\frac{\Delta t^2}{h^2} \in [0; X_3] \cup [X_4; +\infty]$.
La condition $\dfrac{\Delta t^2}{h^2} \geq X_4$ étant incompatible avec la condition (4.14), on se limitera à la condition
$$\frac{\Delta t^2}{h^2} \leq 0.05 \tag{4.15}$$
- Enfin, la condition sur $c_{2,0}$ est équivalente à
$$\frac{19683}{4096}X^6 + \frac{98415}{2048}X^5 + \frac{793881}{20480}X^4 - \frac{137781}{20480}X^3 + \frac{422091}{716800}X^2 - \frac{2187}{71680}X + \frac{2187}{3584000} \geq 0$$
en posant $X = \dfrac{\Delta t^2}{h^2}$. Les solutions de cette équation nous sont données par Maple, deux racines complexes conjuguées et quatre racines réelles
$$\begin{cases} X_1 = -9.0961, & X_2 = -1.0627, \\ X_3 = 0.0411, & X_4 = 0.0627. \end{cases}$$
Par conséquent, $c_{2,0}(\Delta t, 0)$ est positif si et seulement si $\frac{\Delta t^2}{h^2} \in [0; X_3] \cup [X_4; +\infty]$ et on considérera uniquement
$$\frac{\Delta t^2}{h^2} \leq 0.0411 \tag{4.16}$$

En résumé, une condition nécessaire et suffisante de positivité de $M - \dfrac{\Delta t^2}{4} K_\beta^*$ pour $\beta = 0$ est

$$\frac{\Delta t^2}{h^2} \leq 0.0411.$$

A présent, on suppose que $\beta = \pi$. Les coefficients de $q_{2,\Delta t}(\pi, \lambda)$ sont

$$\begin{cases} c_{2,3}(\Delta t, \pi) = & \dfrac{207}{8}\dfrac{\Delta t^4}{h^4} + 9\dfrac{\Delta t^2}{h^2} - \dfrac{97}{105}, \\ c_{2,2}(\Delta t, \pi) = & -\dfrac{14823}{256}\dfrac{\Delta t^8}{h^8} + \dfrac{747}{8}\dfrac{\Delta t^6}{h^6} + \dfrac{69373}{8960}\dfrac{\Delta t^4}{h^4} - \dfrac{40207}{6720}\dfrac{\Delta t^2}{h^2} + \dfrac{8797}{33600}, \\ c_{2,1}(\Delta t, \pi) = & \dfrac{19683}{1024}\dfrac{\Delta t^{12}}{h^{12}} - \dfrac{77517}{512}\dfrac{\Delta t^{10}}{h^{10}} + \dfrac{102519}{1024}\dfrac{\Delta t^8}{h^8} - \dfrac{60039}{2560}\dfrac{\Delta t^6}{h^6} - \dfrac{3969}{512}\dfrac{\Delta t^4}{h^4} \\ & + \dfrac{621}{560}\dfrac{\Delta t^2}{h^2} - \dfrac{93}{4000}, \\ c_{2,0}(\Delta t, \pi) = & \dfrac{19683}{512}\dfrac{\Delta t^{14}}{h^{14}} - \dfrac{137781}{2048}\dfrac{\Delta t^{12}}{h^{12}} + \dfrac{203391}{4096}\dfrac{\Delta t^{10}}{h^{10}} - \dfrac{1095687}{71680}\dfrac{\Delta t^8}{h^8} + \dfrac{85293}{143360}\dfrac{\Delta t^6}{h^6} \\ & + \dfrac{89667}{102400}\dfrac{\Delta t^4}{h^4} - \dfrac{85293}{1792000}\dfrac{\Delta t^2}{h^2} + \dfrac{2187}{3584000}. \end{cases}$$

Regardons le signe de chacun de ces coefficients.

- La condition $-c_{2,3}(\Delta t, \pi) \geq 0$ est équivalente à

  $$-\frac{207}{8}X^2 - 9X + \frac{97}{105} \geq 0$$

  en posant $X = \dfrac{\Delta t^2}{h^2}$. Cette équation admet deux racines réelles $X_1 = -\dfrac{4}{23} - \dfrac{2}{7245}\sqrt{865410}$ et $X_2 = -\dfrac{4}{23} + \dfrac{2}{7245}\sqrt{865410}$. Comme le coefficient devant le terme de plus haut degré est négatif, le polynôme $-c_{2,3}(\Delta t, \pi)$ est positif si et seulement si

  $$\frac{\Delta t^2}{h^2} \leq -\frac{4}{23} + \frac{2}{7245}\sqrt{865410} \simeq 0.0829. \tag{4.17}$$

- La positivité de $c_{2,2}(\Delta t, \pi)$ revient à étudier le signe de la fonction

  $$f : X \longmapsto -\frac{14823}{256}X^4 + \frac{747}{8}X^3 + \frac{69373}{8960}X^2 - \frac{40207}{6720}X + \frac{8797}{33600}$$

  qui admet, d'après Maple, quatre racines réelles

  $$\begin{cases} X_1 = -0.2879, & X_2 = 0.0485, \\ X_3 = 0.1953, & X_4 = 1.6567. \end{cases}$$

  Le terme de plus haut degré du polynôme étant négatif, on peut donc en conclure que $c_{2,2}(\Delta t, \pi)$ est positif si et seulement si $\dfrac{\Delta t^2}{h^2} \in [0; X_2] \cup [X_3; X_4]$. La condition $\dfrac{\Delta t^2}{h^2} \geq X_3$ est incompatible avec (4.17). On se limitera donc à la condition

  $$\frac{\Delta t^2}{h^2} \leq 0.0485. \tag{4.18}$$

- La condition $-c_{2,1}(\Delta t, \pi) \geq 0$ revient à étudier

$$-\frac{19683}{1024}X^6 + \frac{77517}{512}X^5 - \frac{102519}{1024}X^4 + \frac{60039}{2560}X^3 + \frac{3969}{512}X^2 - \frac{621}{560}X + \frac{93}{4000} \geq 0.$$

D'après Maple, ce polynôme admet deux racines complexes conjuguées et quatre racines réelles

$$\begin{cases} X_1 = -0.2336, & X_2 = 0.0260, \\ X_3 = 0.0936, & X_4 = 7.1754. \end{cases}$$

Le coefficient de plus haut degré est négatif donc $-c_{2,1}(\Delta t, \pi)$ est positif si et seulement si $\frac{\Delta t^2}{h^2} \in [0; X_2] \cup [X_3; X_4]$. Or (4.18) est incompatible avec la condition $\frac{\Delta t^2}{h^2} \leq X_3$. On se limite donc à

$$\frac{\Delta t^2}{h^2} \leq 0.0260. \tag{4.19}$$

- Enfin, la positivité de $c_{2,0}(\Delta t, \pi)$ nous conduit à étudier, en posant $X = \frac{\Delta t^2}{h^2}$ le signe de

$$\begin{aligned} f(X) = \ & \frac{19683}{512}X^7 - \frac{137781}{2048}X^6 + \frac{203391}{4096}X^5 - \frac{1095687}{71680}X^4 + \frac{85293}{143360}X^3 \\ & + \frac{89667}{102400}X^2 - \frac{85293}{1792000}X + \frac{2187}{3584000}. \end{aligned}$$

En utilisant Maple, cette fonction admet quatre racines complexes conjuguées deux à deux et trois racines réelles

$$\begin{cases} X_1 = -0.1926, \\ X_2 = 0.0210, \\ X_3 = 0.0332. \end{cases}$$

La première racine est négative et les deux autre positives. Le coefficient de plus haut degré est positif et le plus haut degré est impair. La positivité de $c_{2,0}(\Delta t, \pi)$ est équivalente à $f(X) \geq 0$ et donc à $\frac{\Delta t^2}{h^2} \in [0; X_2] \cup [X_3; +\infty]$. La condition $\frac{\Delta t^2}{h^2} \geq X_3$ est incompatible avec (4.19) donc nous nous limitons à la condition

$$0 \leq \frac{\Delta t^2}{h^2} \leq X_2 = 0.0210. \tag{4.20}$$

Une condition nécessaire et suffisante de positivité de $M - \frac{\Delta t^2}{4}K^*_\beta$ pour $\beta = \pi$ est donc

$$\frac{\Delta t^2}{h^2} \leq 0.0210$$

qui est plus restrictive que la condition obtenue pour $\beta = 0$.

Une condition nécessaire de positivité de $M - \frac{\Delta t^2}{4}K^*_\beta$ est donc, pour $\beta = 0$ et $\beta = \pi$:

$$\frac{\Delta t}{h} \leq \sqrt{0.0210} \simeq 0.1449.$$

Comme précédemment, cette condition est uniquement nécessaire. C'est pourquoi nous avons également étudié numériquement la positivité de $M - \dfrac{\Delta t^2}{4} K^*$. Pour cela, nous avons couplé le code 1D avec l'algorithme 4 pour déterminer par dichotomie le plus petit pas de temps garantissant la positivité de cette matrice.

Algorithme 4

$\Delta t_1 = 0$, $\Delta t_2 = 1$ et $\epsilon = 10^{-8}$
$\Delta t = \dfrac{\Delta t_1 + \Delta t_2}{2}$
$A = M - \dfrac{\Delta t^2}{4}\left(K_1 - \dfrac{\Delta t^2}{12} K_2\right)$
Calcul de $\lambda_{\min}$, la plus petite des valeurs propres de A
tantque $|\Delta t_1 - \Delta t_2| > \epsilon$ ou $\lambda_{\min} < 0$ **faire**
  si $\lambda_{\min} < 0$ **alors**
    $\Delta t_2 = \Delta t$
  sinon
    $\Delta t_1 = \Delta t$
  finsi
  $\Delta t = \dfrac{\Delta t_1 + \Delta t_2}{2}$
  $A = M - \dfrac{\Delta t^2}{4}\left(K_1 - \dfrac{\Delta t^2}{12} K_2\right)$
fin tantque
$\Delta t = \min(\Delta t_1, \Delta t_2)$

Nous avons obtenu la condition numérique

$$\frac{\Delta t}{h} \leq 0.1459.$$

qui semble à nouveau montrer que l'étude mathématique restreinte aux cas $\beta = 0$ et $\beta = \pi$ soit suffisante.

Si l'on regroupe les résultats des sous-sections 4.2.1.1 et 4.2.1.2, il vient que la condition CFL la plus restrictive que l'on obtient est

$$\frac{\Delta t}{h} \leq 0.1449.$$

Remarque 4.2.4. *En pratique, dans toutes les expériences que nous avons menées, on a pu vérifier qu'il s'agissait bien de la bonne condition CFL puisque dès qu'on dépasse de peu cette valeur critique, nous perdons la stabilité du schéma.*

4.2.2 Preuve du théorème 4.2.2

Dans cette section, nous étudions la positivité des deux matrices K^*_β et $M - \dfrac{\Delta t^2}{4} K^*_\beta$ en fixant cette fois le paramètre de pénalisation $\alpha_{2,1}$ à 10. On rappelle que nous avons utilisé cette valeur dans les tests 1D et elle nous semble optimale. On conserve ici $\alpha_1 = 8$ et $\alpha_{2,2} = 0$. Afin d'alléger la présentation des résultats, nous ne donnerons que les résultats provenant de chaque condition sans développer les calculs, comme on a pu le faire dans la sous-section précédente.

4.2.2.1 Positivité de K_β^*

En fixant $\beta = 0$, le polynôme $q_{1,\Delta t}(0,\lambda)$ a pour coefficient

$$\begin{cases} c_{1,3}(\Delta t, 0) = & 171\dfrac{\Delta t^2}{h^2} - 32, \\ c_{1,2}(\Delta t, 0) = & \dfrac{104247}{16}\dfrac{\Delta t^4}{h^4} - \dfrac{31347}{8}\dfrac{\Delta t^2}{h^2} + \dfrac{19467}{80}, \\ c_{1,1}(\Delta t, 0) = & -\dfrac{177147}{16}\dfrac{\Delta t^6}{h^6} - 98415\dfrac{\Delta t^4}{h^4} + \dfrac{1240029}{80}\dfrac{\Delta t^2}{h^2} - \dfrac{19683}{40} \\ c_{1,0}(\Delta t, 0) = & 0. \end{cases}$$

La condition sur $c_{1,0}(\Delta t, 0)$ est clairement vérifiée. L'étude des autres conditions donne les résultats suivants

- $-c_{1,3}(\Delta t, 0) \geq 0 \Leftrightarrow \dfrac{\Delta t^2}{h^2} \leq \dfrac{32}{171} \simeq 0.187,$

- $c_{1,2}(\Delta t, 0) \geq 0 \Leftrightarrow \dfrac{\Delta t^2}{h^2} \in [0, X_1] \cup [X_2, +\infty]$

  où $X_1 = \dfrac{43}{143} - \dfrac{4}{6435}\sqrt{137355} \simeq 0.070$ et $X_2 = \dfrac{43}{143} + \dfrac{4}{6435}\sqrt{137355} \simeq 0.531,$

- $-c_{1,1}(\Delta t, 0) \geq 0 \Leftrightarrow \dfrac{\Delta t^2}{h^2} \in [0, X_1] \cup [X_2, +\infty]$

  où $X_1 = -\dfrac{9}{2} + \dfrac{1}{10}\sqrt{2065} \simeq 0.044$ et $X_2 = \dfrac{1}{9}.$

Quant au polynôme $q_{1,\Delta t}(\pi, \lambda)$, ses coefficients sont donnés par

$$\begin{cases} c_{1,3}(\Delta t, \pi) = & \dfrac{812}{3}\dfrac{\Delta t^2}{h^2} - 36, \\ c_{1,2}(\Delta t, \pi) = & \dfrac{35307}{16}\dfrac{\Delta t^4}{h^4} - \dfrac{105115}{24}\dfrac{\Delta t^2}{h^2} + \dfrac{30091}{80}, \\ c_{1,1}(\Delta t, \pi) = & \dfrac{45927}{16}\dfrac{\Delta t^6}{h^6} - \dfrac{95157}{4}\dfrac{\Delta t^4}{h^4} + \dfrac{1429803}{80}\dfrac{\Delta t^2}{h^2} - \dfrac{10017}{8} \\ c_{1,0}(\Delta t, \pi) = & -\dfrac{45927}{2}\dfrac{\Delta t^6}{h^6} + \dfrac{702027}{16}\dfrac{\Delta t^4}{h^4} - \dfrac{1633689}{80}\dfrac{\Delta t^2}{h^2} + \dfrac{19683}{16}. \end{cases}$$

L'étude du signe de ces quatre coefficients nous donne les résultats suivants :

- $c_{1,3}(\Delta t, \pi) \geq 0 \Leftrightarrow \dfrac{\Delta t^2}{h^2} \leq \dfrac{27}{203} \simeq 0.133,$

- $c_{1,2}(\Delta t, \pi) \geq 0 \Leftrightarrow \dfrac{\Delta t^2}{h^2} \in [0, X_1] \cup [X_2, +\infty]$

  où $\begin{cases} X_1 = \dfrac{105115}{105921} - \dfrac{2}{529605}\sqrt{57105012115} \simeq 0.089, \\ X_2 = \dfrac{105115}{105921} + \dfrac{2}{529605}\sqrt{57105012115} \simeq 0.090, \end{cases}$

- $-c_{1,1}(\Delta t, \pi) \geq 0 \Leftrightarrow \dfrac{\Delta t^2}{h^2} \in [0, X_1] \cup [X_2, X_3]$.

  Ici, on ne précisera que la forme de X_1 afin de ne pas alourdir inutilement la présentation. On a ainsi, $X_1 = -2a\cos\left(b - \dfrac{\pi}{3}\right) + \dfrac{42292}{15309} \simeq 0.078$,

  avec $a = \dfrac{\sqrt{5}\sqrt{6510971417}}{76545}$ et $b = \dfrac{1}{3}\arctan\left(\dfrac{30618\sqrt{3}\sqrt{5}}{1139228280271105}\sqrt{5849041382980548989}\right)$

- $c_{1,0}(\Delta t, \pi) \geq 0 \Leftrightarrow \dfrac{\Delta t^2}{h^2} \in [0, X_1] \cup [X_2, X_3]$

  où $X_1 = \dfrac{9}{14} - \dfrac{1}{70}\sqrt{1605} \simeq 0.0705$, $X_2 = \dfrac{5}{8}$ et $X_3 = \dfrac{9}{14} + \dfrac{1}{70}\sqrt{1605} \simeq 1.215$.

En regroupant les différents résultats obtenus ci-dessus, on obtient la condition nécessaire suivante de positivité de la matrice K_β^*

$$\frac{\Delta t}{h} \leq \sqrt{-\frac{9}{2} + \frac{1}{10}\sqrt{2065}} \simeq 0.209.$$

Afin d'analyser ce résultat, nous avons utilisé l'algorithme 3 introduit précédemment mais en fixant cette fois-ci la valeur du coefficient de pénalisation $\alpha_{2,1}$ à 10. Nous avons obtenu numériquement la condition

$$\frac{\Delta t}{h} \leq 0.211.$$

4.2.2.2 Positivité de $M - \dfrac{\Delta t^2}{4}K_\beta^*$

Les coefficients du polynôme $q_{2,\Delta t}(0, \lambda)$ sont donnés par

$$\begin{cases}
c_{2,3}(\Delta t, 0) = & -\dfrac{171}{4}\dfrac{\Delta t^4}{h^4} + 8\dfrac{\Delta t^2}{h^2} - \dfrac{97}{105}, \\
c_{2,2}(\Delta t, 0) = & \dfrac{104247}{256}\dfrac{\Delta t^8}{h^8} - \dfrac{31347}{128}\dfrac{\Delta t^6}{h^6} + \dfrac{390753}{8960}\dfrac{\Delta t^4}{h^4} - \dfrac{10541}{2240}\dfrac{\Delta t^2}{h^2} + \dfrac{8797}{33600}, \\
c_{2,1}(\Delta t, 0) = & \dfrac{177147}{1024}\dfrac{\Delta t^{12}}{h^{12}} + \dfrac{98415}{64}\dfrac{\Delta t^{10}}{h^{10}} - \dfrac{2164401}{5120}\dfrac{\Delta t^8}{h^8} + \dfrac{274833}{2560}\dfrac{\Delta t^6}{h^6} - \dfrac{249507}{22400}\dfrac{\Delta t^4}{h^4} \\
& + \dfrac{837}{1120}\dfrac{\Delta t^2}{h^2} - \dfrac{93}{4000}, \\
c_{2,0}(\Delta t, 0) = & -\dfrac{177147}{4096}\dfrac{\Delta t^{12}}{h^{12}} - \dfrac{98415}{256}\dfrac{\Delta t^{10}}{h^{10}} + \dfrac{1646811}{20480}\dfrac{\Delta t^8}{h^8} - \dfrac{203391}{20480}\dfrac{\Delta t^6}{h^6} + \dfrac{487701}{716800}\dfrac{\Delta t^4}{h^4} \\
& - \dfrac{2187}{71680}\dfrac{\Delta t^2}{h^2} + \dfrac{2187}{3584000}.
\end{cases}$$

La condition sur $c_{2,3}(\Delta t, 0)$ est toujours vérifiée et l'étude des trois autres coefficients donne

- $c_{2,2}(\Delta t, 0) \geq 0 \Leftrightarrow \dfrac{\Delta t^2}{h^2} \in [0, X_1] \cup [X_2, +\infty]$ où $X_1 \simeq 0.1130$ et $X_2 \simeq 0.3932$,

- $-c_{2,1}(\Delta t, 0) \geq 0 \Leftrightarrow \dfrac{\Delta t^2}{h^2} \leq 0.0587,$

- $c_{2,0}(\Delta t, 0) \geq 0 \Leftrightarrow \dfrac{\Delta t^2}{h^2} \leq 0.0411.$

Les coefficients du polynôme $q_{2,\Delta t}(\pi, \lambda)$ sont donnés par

$$\begin{cases} c_{2,3}(\Delta t, \pi) = & -\dfrac{203}{3}\dfrac{\Delta t^4}{h^4} + 9\dfrac{\Delta t^2}{h^2} - \dfrac{97}{105}, \\ c_{2,2}(\Delta t, \pi) = & \dfrac{35307}{256}\dfrac{\Delta t^8}{h^8} - \dfrac{105115}{384}\dfrac{\Delta t^6}{h^6} + \dfrac{1772459}{26880}\dfrac{\Delta t^4}{h^4} - \dfrac{40207}{6720}\dfrac{\Delta t^2}{h^2} + \dfrac{8797}{33600}, \\ c_{2,1}(\Delta t, \pi) = & -\dfrac{45927}{1024}\dfrac{\Delta t^{12}}{h^{12}} + \dfrac{95157}{256}\dfrac{\Delta t^{10}}{h^{10}} - \dfrac{2440989}{7168}\dfrac{\Delta t^8}{h^8} + \dfrac{2618541}{17920}\dfrac{\Delta t^6}{h^6} - \dfrac{163257}{8960}\dfrac{\Delta t^4}{h^4} \\ & + \dfrac{621}{560}\dfrac{\Delta t^2}{h^2} - \dfrac{93}{4000}, \\ c_{2,0}(\Delta t, \pi) = & -\dfrac{45927}{512}\dfrac{\Delta t^{14}}{h^{14}} + \dfrac{373977}{2048}\dfrac{\Delta t^{12}}{h^{12}} - \dfrac{4677993}{28672}\dfrac{\Delta t^{10}}{h^{10}} + \dfrac{76545}{1024}\dfrac{\Delta t^8}{h^8} - \dfrac{2552229}{143360}\dfrac{\Delta t^6}{h^6} \\ & + \dfrac{999459}{716800}\dfrac{\Delta t^4}{h^4} - \dfrac{85293}{1792000}\dfrac{\Delta t^2}{h^2} + \dfrac{2187}{3584000}. \end{cases}$$

Ici aussi, $-c_{2,3}(\Delta t, \pi) \geq 0$ et l'étude du signe des trois autres coefficients nous donne

- $c_{2,2}(\Delta t, \pi) \geq 0 \Leftrightarrow \dfrac{\Delta t^2}{h^2} \in [0, X_1] \cup [X_2, +\infty]$ où $X_1 \simeq 0.1478$ et $X_2 \simeq 1.7213,$

- $-c_{2,1}(\Delta t, \pi) \geq 0 \Leftrightarrow \dfrac{\Delta t^2}{h^2} \in [0, X_1] \cup [X_2, +\infty]$ où $X_1 \simeq 0.0383$ et $X_2 \simeq 7.3088,$

- $c_{2,0}(\Delta t, \pi) \geq 0 \Leftrightarrow \dfrac{\Delta t^2}{h^2} \leq 0.0332.$

En rassemblant les différentes conditions obtenues et en ne gardant que la plus restrictive, il vient qu'une condition nécessaire de positivité de $M - \dfrac{\Delta t^2}{12} K_\beta^*$ s'écrit

$$\frac{\Delta t}{h} \leq 0.1822.$$

Comme précédemment, nous avons utilisé l'algorithme 4 afin de pouvoir comparer ce résultat. On obtient numériquement que la matrice $M - \dfrac{\Delta t^2}{12} K_\beta^*$ est positive si

$$\frac{\Delta t}{h} \leq 0.1821.$$

Finalement, la condition de stabilité du schéma s'écrit

$$\frac{\Delta t}{h} \leq 0.1821.$$

4.3 Conclusion

Dans ce chapitre, nous avons proposé des conditions nécessaires de stabilité pour le Δ^2-schéma introduit au chapitre 1. En particulier, nous avons montré que pour $\alpha_1 = 8$, $\alpha_{2,1} = \alpha_{2,2} = 0$ une condition nécessaire de stabilité est

$$\frac{\Delta t}{h} \leq 0.145$$

et pour $\alpha_1 = 8$, $\alpha_{2,1} = 10$ et $\alpha_{2,2} = 0$:

$$\frac{\Delta t}{h} \leq 0.1822$$

Nous rappelons que la condition CFL du schéma saute-moutons est $\dfrac{\Delta t}{h} \leq 0.153$.
Les résultats numériques indiquent que ces conditions sont quasi suffisantes en pratique. On constate donc qu'il n'est pas nécessaire de pénaliser l'opérateur biharmonique pour avoir un schéma stable. Cependant, sans ce paramètre de pénalisation, la condition CFL est inférieure à celle du schéma saute-moutons. La pénalisation de l'opérateur biharmonique nous permet donc d'optimiser la condition CFL.
Nous avons également effectué des calculs pour établir la dépendance de la condition CFL par rapport au paramètre de pénalisation α_1 de l'opérateur harmonique. Ces calculs, qui sont présentés en annexe, montrent que la condition CFL décroît quand α_1 augmente.

4.A Une condition CFL dépendant de α_1

Nous avons vu dans la section précédente qu'on pouvait trouver une estimation relativement satisfaisante de la condition CFL lorsque l'on fixe les trois coefficients de pénalisation. Une configuration, qui serait intéressante ici, est de considérer le terme de pénalisation relatif à l'opérateur harmonique comme étant une inconnue du problème. Ainsi, on espère pouvoir déterminer une condition CFL dépendant du choix de ce paramètre de pénalisation. Comme on a pu le faire dans la section précédente, nous allons, dans un premier temps, fixer les deux coefficients de pénalisation liés à l'opérateur biharmonique à zéro puis nous considérons $\alpha_{2,1} = 10$.
Nous proposons les deux conjectures suivantes.

Conjecture 4.A.1. *Si $\alpha_{2,1} = \alpha_{2,2} = 0$, une condition nécessaire de stabilité du schéma (4.1) est donnée par*

$$\frac{\Delta t}{h} \leq \min\left(\sqrt{f_1(\alpha_1)}, \sqrt{X_{1,1}(\alpha_1)}, \sqrt{X_{2,1}(\alpha_1)}\right)$$

où $X_{1,1}(\alpha_1)$ est la plus petite racine positive du polynôme

$$\begin{aligned} p_1(X) \;:=\; & \frac{9}{4}X^6 + \frac{9}{2}\left(1+\frac{\alpha_1}{2}\right)X^5 - 3\left(\frac{5}{4}+\frac{3}{5}\alpha_1\right)X^4 + \frac{3}{70}\left(\frac{61}{2}-13\alpha_1\right)X^3 \\ & + \frac{1}{140}\left(-43+\frac{51}{5}\alpha_1\right)X^2 + \frac{1}{350}(3-\alpha_1)X + \frac{1}{3500} \end{aligned}$$

$X_{2,1}(\alpha_1)$ est celle du polynôme

$$\begin{aligned} p_2(X) \;:=\; & \frac{9}{2}X^7 + 9(1-\alpha_1)X^6 + \frac{9}{2}(-27+15\alpha_1)X^5 + \frac{3}{5}\left(61-\frac{11}{70}\alpha_1\right)X^4 + \frac{3}{35}\left(-\frac{3}{2}+\alpha_1\right)X^3 \\ & + \frac{1}{350}(-169+57\alpha_1)X^2 - \frac{3}{175}\left(1+\frac{1}{5}\alpha_1\right)X + \frac{1}{1750} \end{aligned}$$

et $f_1(\alpha_1) = -\dfrac{\alpha_1+1}{2} + \dfrac{1}{10}\sqrt{-95+70\alpha_1+25\alpha_1^2}$.

Conjecture 4.A.2. *Si* $\alpha_{2,1} = 10$ *et* $\alpha_{2,2} = 0$*, une condition nécessaire de stabilité du schéma (4.1) est donnée par*

$$\frac{\Delta t}{h} \leq \min\left(\sqrt{f_2}, \sqrt{f_3(\alpha_1)}, \sqrt{X_{3,1}(\alpha_1)}, \sqrt{X_{4,1}(\alpha_1)}\right)$$

où $X_{3,1}(\alpha_1)$ *est la plus petite racine positive du polynôme*

$$\begin{aligned} p_3(X) \;:=\; & -\frac{81}{8}X^6 - 9\left(1+\frac{9}{8}\alpha_1\right)X^5 + \frac{3}{4}\left(-\frac{17}{2}+\frac{21}{5}\alpha_1\right)X^4 + \frac{3}{140}\left(\frac{151}{2}-23\alpha_1\right)X^3 \\ & + \frac{1}{280}\left(-37+\frac{51}{5}\alpha_1\right)X^2 + \frac{1}{700}(3-\alpha_1)X + \frac{1}{7000} \end{aligned}$$

et $X_{4,1}(\alpha_1)$ *est celle du polynôme*

$$\begin{aligned} p_4(X) \;:=\; & \frac{21}{16}\alpha_1 X^7 + 3\left(-\frac{7}{8}+\alpha_1\right)X^6 + \frac{3}{16}\left(25-\frac{111}{7}\alpha_1\right)X^5 + \frac{1}{140}\left(-437+\frac{831}{4}\alpha_1\right)X^4 \\ & + \frac{3}{280}\left(\frac{139}{2}-33\alpha_1\right)X^3 + \frac{1}{2800}(1+57\alpha_1)X^2 - \frac{3}{1400}\left(1+\frac{1}{5}\alpha_1\right)\frac{\Delta t^2}{h^2} + \frac{1}{14000}. \end{aligned}$$

$f_2 = \dfrac{9}{14} - \dfrac{1}{70}\sqrt{1605}$

et $f_3(\alpha_1) = -\dfrac{1}{2}\left(1+\alpha_1 - \dfrac{1}{5}\sqrt{-95+70\alpha_1+25\alpha_1^2}\right)$.

On a représenté sur la figure 4.1 les conditions obtenues d'après la conjecture 4.A.1 (en rouge avec tirets et pointillés) et la conjecture 4.A.2 (en bleu) et on les a comparées à la condition CFL du schéma saute-moutons (en vert avec tirets). On constate que le fait de considérer un coefficient

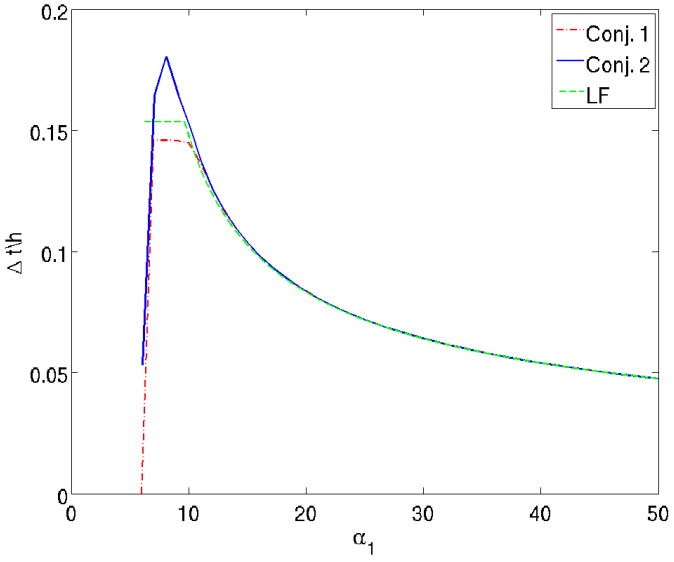

FIGURE 4.1 – Comparaison des conditions issues des conjectures 4.A.1 et 4.A.2.

de pénalisation adéquat pour l'opérateur biharmonique (conjecture 4.A.2) permet d'avoir une condition CFL moins restrictive que pour le schéma saute-moutons et amène donc moins de coûts de

calcul. Néanmoins, un choix peu judicieux de ce coefficient ($\alpha_{2,1} = 0$, conjecture 4.A.1) amène à une condition CFL moins avantageuse que pour le schéma saute-moutons et pénalise donc la rapidité de l'algorithme.

Nous n'énonçons ici que des conjectures car nous ne pouvons justifier rigoureusement que ces conditions sont bien nécessaires. La difficulté vient de l'étude des valeurs propres de $M - \dfrac{\Delta t^2}{4} K_\beta^*$ (voir section 4.A.1.2). Néanmoins, les résultats numériques que nous avons obtenus (voir sections 4.A.1.3 et 4.A.2.3) montrent que ces conditions semblent être réellement nécessaires. De plus, ces résultats indiquent qu'elles sont quasi suffisantes.

4.A.1 Sans pénalisation pour l'opérateur biharmonique

Nous considérons ici $\alpha_{2,1} = \alpha_{2,2} = 0$ et nous allons à nouveau étudier la positivité des matrices K_β^* et $M - \dfrac{\Delta t^2}{4} K_\beta^*$ en considérant leurs polynômes caractéristiques respectifs.

4.A.1.1 Etude de la positivité de K^*

En étudiant la positivité de K_β^* pour $\beta = 0$ et $\beta = \pi$, nous obtenons le lemme suivant.

Lemme 4.A.3. *Si $\alpha_{2,1} = \alpha_{2,2} = 0$, une condition nécessaire de positivité de la matrice K^* est*

$$\frac{\Delta t}{h} \leq f_1(\alpha_1) = \sqrt{-\frac{\alpha_1 + 1}{2} + \frac{1}{10}\sqrt{-95 + 70\alpha_1 + 25\alpha_1^2}} \text{ et } \alpha_1 \geq 6.$$

Démonstration. En utilisant Maple, on vérifie que, pour $\beta = 0$, le coefficient $c_{1,1}(\Delta t, 0)$ de $q_1(\Delta t, 0)$ s'écrit

$$\begin{aligned} c_{1,1}(\Delta t, 0) = \; & \frac{19683}{16}\frac{\Delta t^{12}}{h^{12}} + \frac{19683}{8}\left(1 + \frac{\alpha_1}{2}\right)\frac{\Delta t^8}{h^8} + \left(\frac{216513}{80} + \frac{19683}{20}\alpha_1\right)\frac{\Delta t^4}{h^4} \\ & + \frac{59049}{40} - \frac{19683}{80}\alpha_1. \end{aligned}$$

On vérifie que ce polynôme de degré 3 en $\dfrac{\Delta t^4}{h^4}$ admet deux racines négatives et une positive qui s'écrit

$$f_1(\alpha_1) = -\frac{\alpha_1 + 1}{2} + \frac{1}{10}\sqrt{-95 + 70\alpha_1 + 25\alpha_1^2}.$$

Une condition nécessaire de positivité de la matrice K^* est $c_{1,1}(\Delta t,) \leq 0$, c'est-à-dire

$$\frac{\Delta t}{h} \leq \sqrt{-\frac{\alpha_1 + 1}{2} + \frac{1}{10}\sqrt{-95 + 70\alpha_1 + 25\alpha_1^2}} \text{ et } \alpha_1 \geq 6.$$

Cette condition ne peut être vérifiée que pour $\alpha_1 \geq 6$, de plus petites valeurs conduisant à une condition du type $\Delta t^2 \leq 0$. On retrouve ici la minoration de α_1 que l'on avait déterminée au chapitre 2.

Il est à noter que $c_{1,0}(\Delta t, 0)$ est toujours positif et n'apporte donc aucune précision quant à la condition la plus restrictive. Les coefficients $c_{1,i}(\Delta t, 0)$, $i \neq 1$, conduisent à des conditions moins restrictives que nous ne détaillons donc pas ici. Pour illustrer ce point, nous représentons ces 3

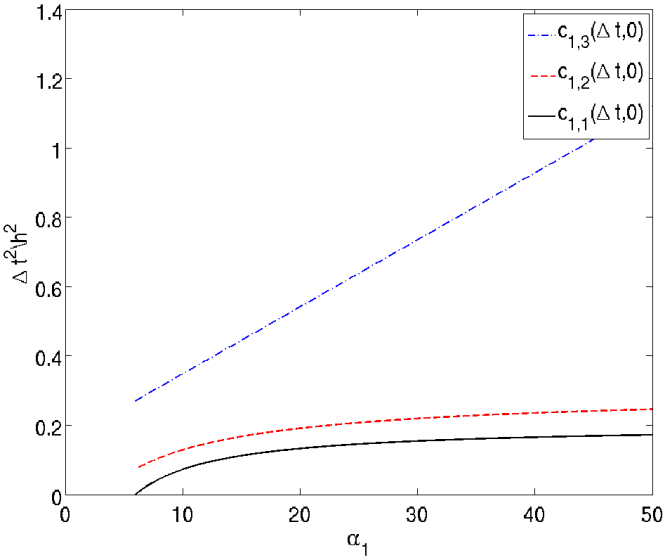

FIGURE 4.2 – Comparaison des différentes conditions issues de l'étude des coefficients $c_{1,i}(\Delta t, 0)$.

conditions en fonction de α_1 sur un intervalle $[6, 50]$ sur la figure 4.2. Ce choix d'intervalle se justifie par le fait que pour que $c_{1,1}(\Delta t, 0)$ soit négatif, une condition nécessaire et suffisante est que $\alpha_1 \geq 6$. Nous avons représenté en bleu (resp. en rouge et en noir) la condition issue de l'étude de $c_{1,3}(\Delta t, 0)$ (resp. de $c_{1,2}(\Delta t, 0)$ et de $c_{1,1}(\Delta t, 0)$) et il est clair que la condition la plus restrictive sur le pas de temps est celle provenant de $c_{1,1}(\Delta t, 0)$.
Les coefficients $c_{1,i}(\Delta t, \pi)$ fournissent également des conditions moins restrictives. $c_{1,3}(\Delta t, \pi)$ conduit à une condition toujours vérifiée et nous représentons sur la figure 4.3 en bleu (resp. en rouge, en vert et en noir) la condition issue de l'étude de $c_{1,1}(\Delta t, 0)$ (resp. de $c_{1,2}(\Delta t, \pi)$, de $c_{1,1}(\Delta t, \pi)$ et de $c_{1,0}(\Delta t, \pi)$). pour différentes valeurs de α_1. Là encore, il est clair que la condition la plus restrictive sur le pas de temps est celle provenant de $c_{1,1}(\Delta t, 0)$.

□

4.A.1.2 Etude de la positivité de $M - \frac{\Delta t^2}{4} K^*$

En étudiant la positivité de $M - \frac{\Delta t^2}{4} K^*_\beta$ pour $\beta = 0$ et $\beta = \pi$, nous avons établi la conjecture suivante.

Conjecture 4.A.4. *Si $\alpha_{2,1} = \alpha_{2,2} = 0$, pour avoir positivité de la matrice $M - \frac{\Delta t^2}{12} K^*$, il faut que*

$$\frac{\Delta t}{h} \leq \min\left(X_{1,1}(\alpha_1), X_{2,1}(\alpha_1)\right)$$

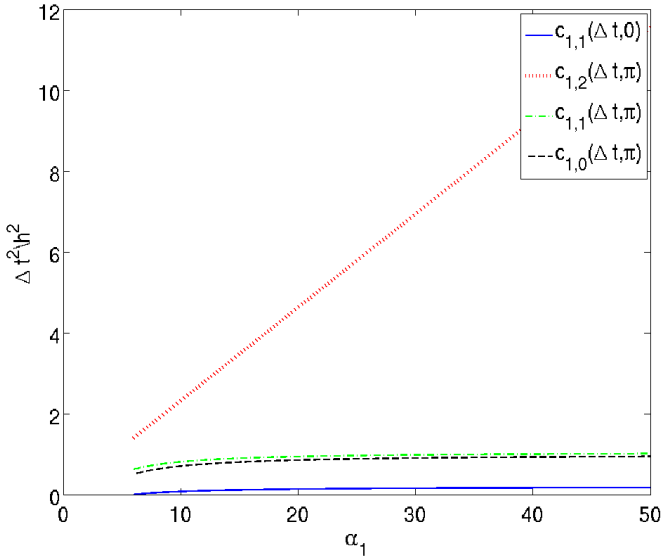

FIGURE 4.3 – Comparaison des différentes conditions issues de l'étude des coefficients $c_{1,i}(\Delta t, \pi)$.

où $X_{1,1}(\alpha_1)$ *est la plus petite racine positive du polynôme*

$$\begin{aligned} p_1(X) \;:=\; & \frac{9}{4}X^6 + \frac{9}{2}\left(1+\frac{\alpha_1}{2}\right)X^5 - 3\left(\frac{5}{4}+\frac{3}{5}\alpha_1\right)X^4 + \frac{3}{70}\left(\frac{61}{2}-13\alpha_1\right)X^3 \\ & + \frac{1}{140}\left(-43+\frac{51}{5}\alpha_1\right)X^2 + \frac{1}{350}(3-\alpha_1)X + \frac{1}{3500} \end{aligned}$$

et $X_{2,1}(\alpha_1)$ *est celle du polynôme*

$$\begin{aligned} p_2(X) \;:=\; & \frac{9}{2}X^7 + 9(1-\alpha_1)X^6 + \frac{9}{2}(-27+15\alpha_1)X^5 + \frac{3}{5}\left(61-\frac{11}{70}\alpha_1\right)X^4 + \frac{3}{35}\left(-\frac{3}{2}+\alpha_1\right)X^3 \\ & + \frac{1}{350}(-169+57\alpha_1)X^2 - \frac{3}{175}\left(1+\frac{1}{5}\alpha_1\right)X + \frac{1}{1750}. \end{aligned}$$

Pour arriver à cette conjecture, nous considérons le coefficient $c_{2,0}(\Delta t, 0)$ de $q_{2,\Delta t}(0,\lambda)$

$$\begin{aligned} c_{2,0}(\Delta t, 0) = \; & \frac{2187}{1024}\left[\frac{9}{4}\frac{\Delta t^{12}}{h^{12}} + \frac{9}{2}\left(1+\frac{\alpha_1}{2}\right)\frac{\Delta t^{10}}{h^{10}} - 3\left(\frac{5}{4}+\frac{3}{5}\alpha_1\right)\frac{\Delta t^8}{h^8}\right. \\ & + \frac{3}{70}\left(\frac{61}{2}-13\alpha_1\right)\frac{\Delta t^6}{h^6} + \frac{1}{140}\left(-43+\frac{51}{5}\alpha_1\right)\frac{\Delta t^4}{h^4} \\ & \left. + \frac{1}{350}(3-\alpha_1)\frac{\Delta t^2}{h^2} + \frac{1}{3500}\right] = p_1\left(\frac{\Delta t^2}{h^2}\right). \end{aligned}$$

Pour étudier complétement le signe de ce coefficient, nous devons calculer toutes les racines positives de p_1. Nous supposons qu'il y en a N positives (éventuellement multiples) notées $(X_{1,i})_{i=1,\dots,N}$ et classées par ordre croissant.
Comme $c_{2,0}(0,0) \geq 0$ et que $\lim_{\Delta t \to +\infty} c_{2,0}(\Delta t, 0) = +\infty$, N est pair de sorte que la condition

pour que $c_{2,0}(0,0) \geq 0$ s'écrit

$$\frac{\Delta t^2}{h^2} \in [0; X_{1,1}] \cup [X_{1,2}; X_{1,3}] \cup \ldots \cup [X_{1,N}; +\infty] .$$

Nous ferons ici la conjecture que les conditions du type $\frac{\Delta t^2}{h^2} \in [X_{1,i}; X_{1,i+1}]$ ou $\frac{\Delta t^2}{h^2} \in [X_{1,N}; +\infty]$ sont incompatibles avec les conditions fournies par les autres coefficients, comme c'était le cas à la section précédente.
A partir de cette conjecture, nous déduisons qu'une condition nécessaire de positivité de $M - \frac{\Delta t^2}{4} K^*$ est

$$\frac{\Delta t}{h} \leq \sqrt{X_{1,1}(\alpha_1)}$$

où $X_{1,1}(\alpha_1)$ est la plus petite racine de p_1.
Afin d'illustrer que ce coefficient est le plus restrictif, nous représentons sur la figure 4.4 (en vert avec tirets) la condition analytique issue de l'étude de $c_{2,3}(\Delta t, 0)$ et en noir avec croix (resp. rouge avec tirets et pointillés et bleu) la courbe obtenue numériquement en fixant les valeurs de α_1 pour le coefficient $c_{2,2}(\Delta t, 0)$ (resp. $c_{2,1}(\Delta t, 0)$ et $c_{2,0}(\Delta t, 0)$).

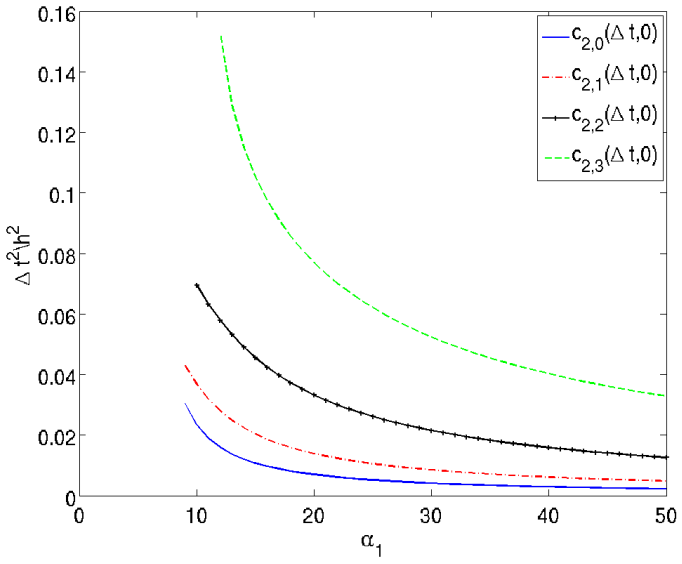

FIGURE 4.4 – Comparaison des différentes conditions issues de l'étude des coefficients $c_{2,i}(\Delta t, 0)$.

Remarque 4.A.5. *On notera que toutes les courbes présentées sur la figure 4.4 ne donnent pas de conditions pour les premières valeurs de α_1. Cela est tout simplement dû au fait que dans ces cas, le coefficient $c_{2,i}(\Delta t, 0)$ n'admet que des racines complexes et a donc le signe souhaité.*

Dans le cas $\beta = \pi$, la condition la plus restrictive est celle correspondant au coefficient

$c_{2,0}(\Delta t, \pi)$

$$c_{2,0}(\Delta t,\pi) = \frac{2187}{2048}\left[\frac{9}{2}\frac{\Delta t^{14}}{h^{14}} + 9(1-\alpha_1)\frac{\Delta t^{12}}{h^{12}} + \frac{9}{2}(-27+15\alpha_1)\frac{\Delta t^{10}}{h^{10}} + \frac{3}{5}\left(61-\frac{11}{70}\alpha_1\right)\frac{\Delta t^8}{h^8}\right.$$
$$+\frac{3}{35}\left(-\frac{3}{2}+\alpha_1\right)\frac{\Delta t^6}{h^6} + \frac{1}{350}(-169+57\alpha_1)\frac{\Delta t^4}{h^4}$$
$$\left.-\frac{3}{175}\left(1+\frac{1}{5}\alpha_1\right)\frac{\Delta t^2}{h^2} + \frac{1}{1750}\right] = p_2\left(\frac{\Delta t^2}{h^2}\right).$$

Comme pour le cas $\beta = 0$, nous conjecturons qu'une condition nécessaire de positivité de $M - \frac{\Delta t^2}{4}K^*$ est

$$\frac{\Delta t}{h} \leq \sqrt{X_{2,1}(\alpha_1)}$$

où $X_{2,1}(\alpha_1)$ est la plus petite racine positive de p_2.
Sur la figure 4.5, nous avons représenté en noir avec croix (resp. en vert avec tirets, en rouge avec pointillés et en bleu) la condition sur $c_{2,3}(\Delta t, \pi)$ (resp. $c_{2,2}(\Delta t, \pi)$, $c_{2,1}(\Delta t, \pi)$ et $c_{2,0}(\Delta t, \pi)$) et nous constatons que la condition la plus restrictive est celle correspondant au coefficient $c_{2,0}(\Delta t, \pi)$.

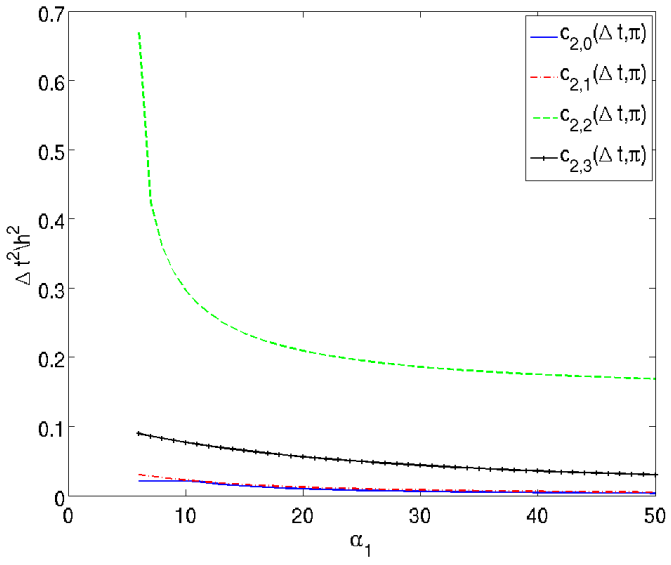

FIGURE 4.5 – Comparaison des différentes conditions issues de l'étude des coefficients $c_{2,i}(\Delta t, \pi)$.

Ainsi, une condition nécessaire de positivité de $M - \frac{\Delta t^2}{4}K^*$ est

$$\frac{\Delta t}{h} \leq \min\left(\sqrt{X_{1,1}(\alpha_1)}, \sqrt{X_{2,1}(\alpha_1)}\right).$$

On compare les conditions obtenues sur $c_{2,0}(\Delta t, 0)$ et sur $c_{2,0}(\Delta t, \pi)$ sur la figure 4.6. On représente en bleu la condition issue de $c_{2,0}(\Delta t, 0)$ et en rouge avec pointillés celle issue de $c_{2,0}(\Delta t, \pi)$.

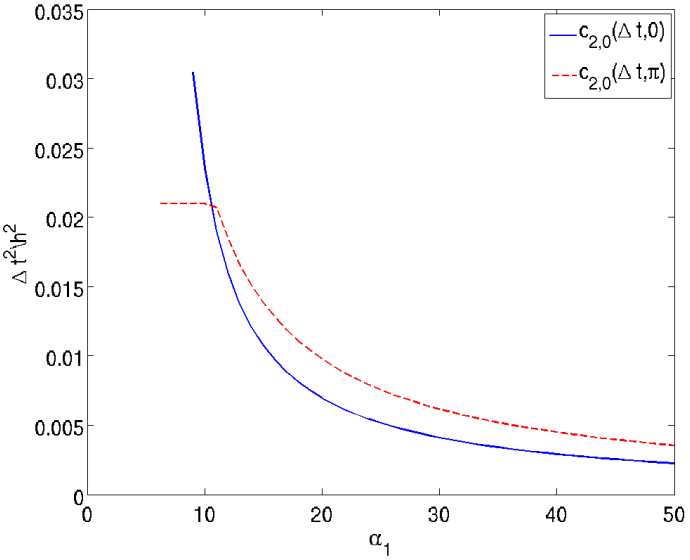

FIGURE 4.6 – Comparaison des conditions $c_{2,0}\left(\Delta t, 0\right)$ et $c_{2,0}\left(\Delta t, \pi\right)$.

4.A.1.3 Comparaison avec des expériences numériques

Afin d'analyser ces résultats, nous allons utiliser l'algorithme 3 en faisant varier α_1. Nous obtiendrons ainsi les valeurs optimales du pas de temps Δt assurant la positivité de la matrice K^* en fonction du paramètre de pénalisation α_1. On a représenté sur la figure 4.7 la condition nécessaire de positivité de K^* issue du lemme 4.A.3 en bleu et les résultats obtenus grâce à l'algorithme 3 avec des croix rouges. Il est clair que l'on dispose d'une bonne condition de positivité de K^*.

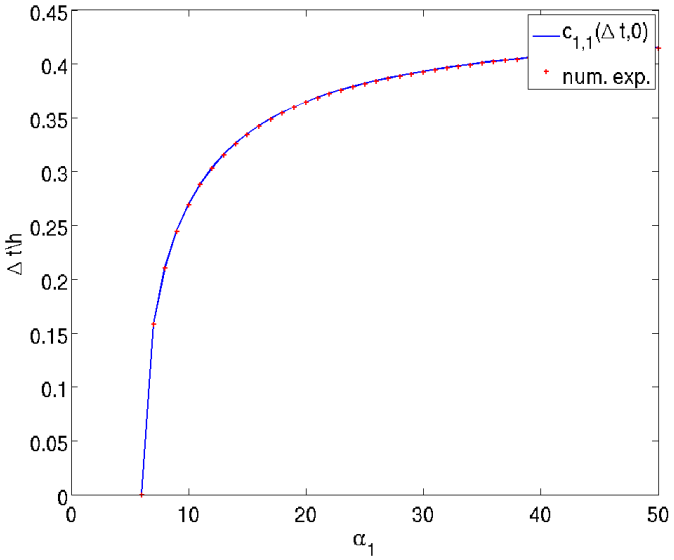

FIGURE 4.7 – Comparaison entre la condition nécessaire de positivité de K^* et la solution numérique

Nous allons maintenant utiliser l'algorithme 4 en faisant varier α_1 afin d'obtenir les valeurs

optimales du pas de temps Δt assurant la positivité de la matrice $M - \frac{\Delta t^2}{4} K^*$ en fonction du paramètre de pénalisation α_1. On a représenté sur la figure 4.8 la condition nécessaire de positivité de $M - \frac{\Delta t^2}{4} K^*$ issue de la conjecture 4.A.4 en bleu et les résultats obtenus grâce à l'algorithme 4 avec des croix rouges.

Enfin, si l'on compare les deux conditions obtenues, d'après les études sur K^*_β et sur $M - \frac{\Delta t^2}{4} K^*_\beta$,

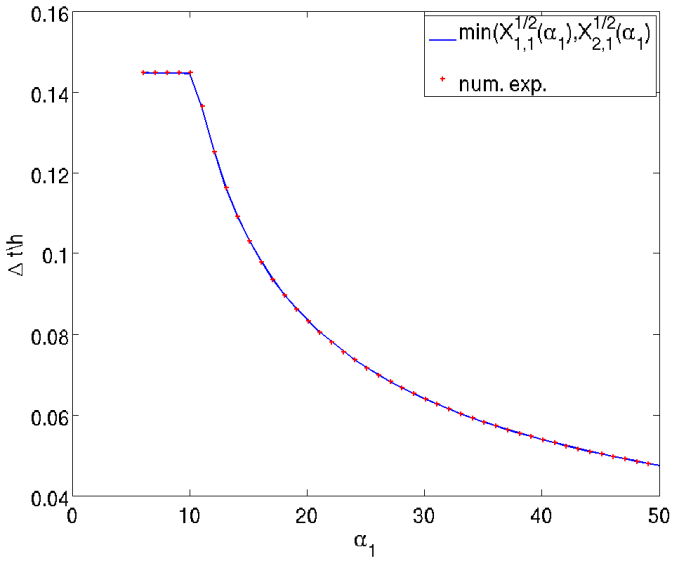

FIGURE 4.8 – Comparaison entre la condition nécessaire de positivité de $M - \frac{\Delta t^2}{4} K^*$ et la solution numérique

avec les résultats obtenus numériquement, on obtient la figure 4.9. On a représenté en bleu avec tirets la condition issue de l'étude de K^*_β, en vert celle issue de l'étude de $M - \frac{\Delta t^2}{4} K^*_\beta$ et enfin avec des croix rouges, la solution numérique obtenue grâce à l'algorithme 5. On constate sur cette figure que l'on a obtenu une condition nécessaire de stabilité qui donne une bonne approximation de la condition nécessaire et suffisante de stabilité.

Algorithme 5

$\Delta t_1 = 0$, $\Delta t_2 = 1$ et $\epsilon = 10^{-8}$
$\Delta t = \dfrac{\Delta t_1 + \Delta t_2}{2}$
$A = K_1 - \dfrac{\Delta t^2}{12} K_2$ et $B = M - \dfrac{\Delta t^2}{4} A$
Calcul de $\lambda_{\min,A}$ et $\lambda_{\min,B}$, la plus petite des valeurs propres de A et de B
tantque $|\Delta t_1 - \Delta t_2| > \epsilon$ ou $\lambda_{\min,A} < 0$ ou $\lambda_{\min,B} < 0$ **faire**
  si $\lambda_{\min,A} < 0$ ou $\lambda_{\min,B} < 0$ **alors**
    $\Delta t_2 = \Delta t$
  sinon
    $\Delta t_1 = \Delta t$
  finsi
  $\Delta t = \dfrac{\Delta t_1 + \Delta t_2}{2}$
  $A = K_1 - \dfrac{\Delta t^2}{12} K_2$ et $B = M - \dfrac{\Delta t^2}{4} A$
fin tantque
$\Delta t = \min(\Delta t_1, \Delta t_2)$

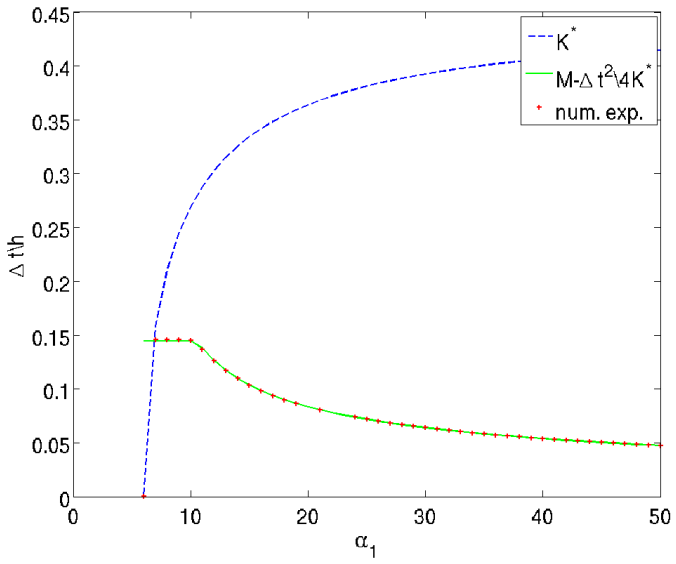

FIGURE 4.9 – Comparaison entre les conditions numériques et théoriques.

4.A.2 Avec pénalisation sur l'opérateur biharmonique

Cette fois, nous considérons $\alpha_{2,1} = 10$ et $\alpha_{2,2} = 0$ et nous allons à nouveau étudier la positivité des matrices K^*_β et $M - \dfrac{\Delta t^2}{4} K^*_\beta$ en considérant leurs polynômes caractéristiques respectifs.

4.A.2.1 Etude de la positivité de K^*

En étudiant la positivité de K^*_β pour $\beta = 0$ et $\beta = \pi$, nous obtenons le lemme suivant.

Lemme 4.A.6. *Si $\alpha_{2,1} = 10$ et $\alpha_{2,2} = 0$, une condition nécessaire de positivité de la matrice K^**

est

$$\frac{\Delta t}{h} \leq \min\left(\sqrt{\frac{9}{14}-\frac{1}{70}\sqrt{1605}}, \sqrt{-\frac{1}{2}\left(1+\alpha_1-\frac{1}{5}\sqrt{-95+70\alpha_1+25\alpha_1^2}\right)}\right).$$

Démonstration. En utilisant Maple, on vérifie que, pour $\beta = 0$, le coefficient $c_{1,1}(\Delta t, 0)$ de $q_1(\Delta t, 0)$ s'écrit

$$\begin{aligned} c_{1,1}(\Delta t,0) = & -\frac{177147}{16}\frac{\Delta t^6}{h^6} - \frac{19683}{2}\left(1+\frac{9}{8}\alpha_1\right)\frac{\Delta t^4}{h^4} + \frac{137781}{40}\left(\alpha_1-\frac{7}{2}\right)\frac{\Delta t^2}{h^2} \\ & + \frac{19683}{40}(3-\alpha_1). \end{aligned}$$

L'étude de la positivité de ce coefficient, que nous ne détaillons pas ici, nous mène à la condition nécessaire

$$\frac{\Delta t^2}{h^2} \leq -\frac{1}{2}\left(1+\alpha_1-\frac{1}{5}\sqrt{25\alpha_1^2+70\alpha_1-95}\right) = f_3(\alpha_1)$$

Pour $\beta = \pi$, le coefficient $c_{1,0}(\Delta t, \pi)$ s'écrit

$$c_{1,0}(\Delta t,\pi) = -\frac{45927}{16}\frac{\Delta t^6}{h^6} + 6561\left(\alpha_1-\frac{21}{16}\right)\frac{\Delta t^4}{h^4} + 19683\left(\frac{9}{16}\frac{\alpha_1}{5}\right)\frac{\Delta t^2}{h^2} + \frac{19683}{80}(\alpha_1-3).$$

L'étude des racines de ce polynôme en $\dfrac{\Delta t^2}{h^2}$ montre qu'une condition nécessaire de stabilité de K^* s'écrit

$$\frac{\Delta t^2}{h^2} \leq \frac{1}{14}\left(9-\frac{1}{5}\sqrt{1605}\right) = f_2$$

Une condition nécessaire de positivité de la matrice K^* est donc

$$\frac{\Delta t}{h} \leq \min\left(f_2, f_3(\alpha_1)\right)$$

c'est-à-dire

$$\frac{\Delta t}{h} \leq \min\left(\sqrt{\frac{1}{14}\left(9-\frac{1}{5}\sqrt{1605}\right)}, \sqrt{-\frac{1}{2}\left(1+\alpha_1-\frac{1}{5}\sqrt{-95+70\alpha_1+25\alpha_1^2}\right)}\right). \quad (4.21)$$

□

4.A.2.2 Etude de la positivité de $M - \dfrac{\Delta t^2}{4}K^*$

En étudiant la positivité de $M - \dfrac{\Delta t^2}{4}K^*_\beta$ pour $\beta = 0$ et $\beta = \pi$, nous avons établi la conjecture suivante

Conjecture 4.A.7. *Si $\alpha_{2,1} = 10$ et $\alpha_{2,2} = 0$, pour avoir positivité de la matrice $M - \dfrac{\Delta t^2}{12}K^*$, $\dfrac{\Delta t}{h}$ doit être plus petit que $\min\left(X_{3,1}(\alpha_1), X_{4,1}(\alpha_1)\right)$ où $X_{3,1}(\alpha_1)$ est la plus petite racine positive du polynôme*

$$\begin{aligned} p_3(X) := & -\frac{81}{8}X^6 - 9\left(1+\frac{9}{8}\alpha_1\right)X^5 + \frac{3}{4}\left(-\frac{17}{2}+\frac{21}{5}\alpha_1\right)X^4 + \frac{3}{140}\left(\frac{151}{2}-23\alpha_1\right)X^3 \\ & + \frac{1}{280}\left(-37+\frac{51}{5}\alpha_1\right)X^2 + \frac{1}{700}(3-\alpha_1)X + \frac{1}{7000} \end{aligned}$$

et $X_{4,1}(\alpha_1)$ *est celle du polynôme*

$$\begin{aligned} p_4(X) &:= \frac{21}{16}\alpha_1 X^7 + 3\left(-\frac{7}{8}+\alpha_1\right)X^6 + \frac{3}{16}\left(25-\frac{111}{7}\alpha_1\right)X^5 + \frac{1}{140}\left(-437+\frac{831}{4}\alpha_1\right)X^4 \\ &\quad + \frac{3}{280}\left(\frac{139}{2}-33\alpha_1\right)X^3 + \frac{1}{2800}\left(1+57\alpha_1\right)X^2 - \frac{3}{1400}\left(1+\frac{1}{5}\alpha_1\right)\frac{\Delta t^2}{h^2} + \frac{1}{14000}. \end{aligned}$$

Pour établir cette conjecture, nous considérons uniquement le cas $\beta = 0$. Dans ce cas, le coefficient $c_{2,0}(\Delta t, 0)$ de $q_{2,\Delta t}(0,\lambda)$ s'écrit

$$\begin{aligned} c_{2,0}(\Delta t, 0) &= \frac{2187}{512}\left[-\frac{81}{8}\frac{\Delta t^{12}}{h^{12}} - 9\left(1+\frac{9}{8}\alpha_1\right)\frac{\Delta t^{10}}{h^{10}} + \frac{3}{4}\left(-\frac{17}{2}+\frac{21}{5}\alpha_1\right)\frac{\Delta t^8}{h^8} + \right. \\ &\quad \left. \frac{3}{140}\left(\frac{151}{2}-23\alpha_1\right)\frac{\Delta t^6}{h^6} + \frac{1}{280}\left(-37+\frac{51}{5}\alpha_1\right)\frac{\Delta t^4}{h^4} + \frac{1}{700}\left(3-\alpha_1\right)\frac{\Delta t^2}{h^2} + \frac{1}{7000}\right] \\ &= p_3\left(\frac{\Delta t^2}{h^2}\right). \end{aligned}$$

En utilisant à nouveau la conjecture de la sous-section 4.A.1.2, nous déduisons qu'une condition nécessaire de positivité de $M - \frac{\Delta t^2}{4}K^*$ est

$$\frac{\Delta t}{h} \leq \sqrt{X_{3,1}(\alpha_1)}$$

où $X_{3,1}(\alpha_1)$ est la plus petite racine du polynôme $p_3(X)$.

Dans le cas $\beta = \pi$, le coefficient $c_{2,0}(\Delta t, \pi)$ s'écrit

$$\begin{aligned} c_{2,0}(\Delta t, \pi) = \frac{2187}{256}&\left[\frac{21}{16}\alpha_1\frac{\Delta t^{14}}{h^{14}} + 3\left(-\frac{7}{8}+\alpha_1\right)\frac{\Delta t^{12}}{h^{12}} + \frac{3}{16}\left(25-\frac{111}{7}\alpha_1\right)\frac{\Delta t^{10}}{h^{10}}\right. \\ &+ \frac{1}{140}\left(-437+\frac{831}{4}\alpha_1\right)\frac{\Delta t^8}{h^8} + \frac{3}{280}\left(\frac{139}{2}-33\alpha_1\right)\frac{\Delta t^6}{h^6} + \frac{1}{2800}\left(1+57\alpha_1\right)\frac{\Delta t^4}{h^4} \\ &\left. - \frac{3}{1400}\left(1+\frac{1}{5}\alpha_1\right)\frac{\Delta t^2}{h^2} + \frac{1}{14000}\right] = p_4\left(\frac{\Delta t^2}{h^2}\right) \end{aligned}$$

Ainsi, une condition nécessaire de positivité de $M - \frac{\Delta t^2}{4}K^*$ est

$$\frac{\Delta t}{h} \leq \min\left(\sqrt{X_{3,1}(\alpha_1)}, \sqrt{X_{4,1}(\alpha_1)}\right)$$

où $X_{5,1}(\alpha_1)$ est la plus petite racine du polynôme $c_{2,0}(\Delta t, \pi)$. On compare les conditions obtenues sur $c_{2,0}(\Delta t, 0)$ et sur $c_{2,0}(\Delta t, \pi)$ sur la figure 4.10. On représente en bleu la condition issue de $c_{2,0}(\Delta t, 0)$ et en rouge avec pointillés celle issue de $c_{2,0}(\Delta t, \pi)$. On constate que les deux conditions interviennent dans le calcul de la condition la plus restrictive.

Enfin, on représente sur la figure 4.11 les conditions obtenues d'après les études de K^*_β et de $M - \frac{\Delta t^2}{4}K^*_\beta$. Nous avons tracé en bleu la condition provenant de l'étude K^*_β et en rouge avec tirets celle provenant de l'étude de $M - \frac{\Delta t^2}{4}K^*_\beta$. On remarque que pour α_1 petit, la condition sur

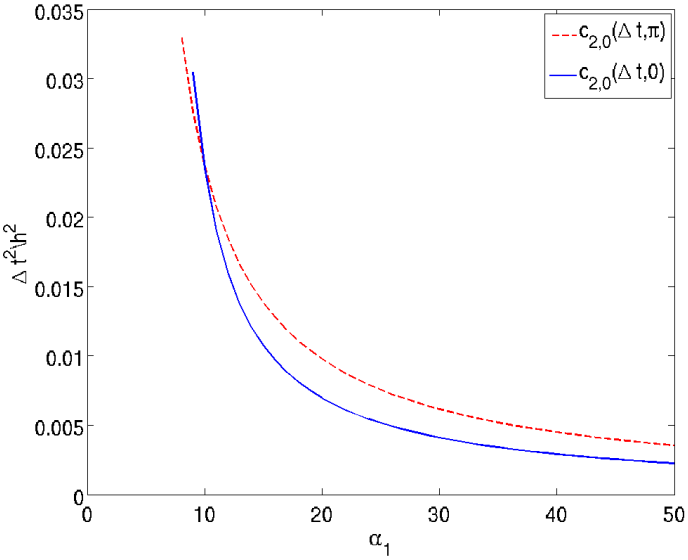

FIGURE 4.10 – Comparaison des conditions $c_{2,0}\left(\Delta t, 0\right)$ et $c_{2,0}\left(\Delta t, \pi\right)$.

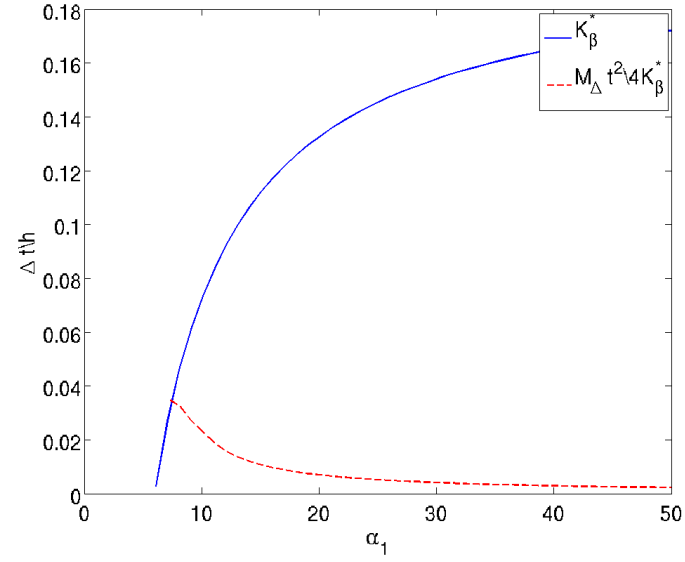

FIGURE 4.11 – Comparaison des conditions sur K_β^* et $M - \dfrac{\Delta t^2}{4} K_\beta^*$.

K_β^* est la plus restrictive alors que pour α_1 grand, c'est la condition sur $M - \dfrac{\Delta t^2}{4} K_\beta^*$ qui intervient.

Ainsi une condition nécessaire nécessaire de stabilité du schéma (4.1) est

$$\frac{\Delta t}{h} \leq \min\left(\sqrt{f_2}, \sqrt{f_3\left(\alpha_1\right)}, \sqrt{X_{3,1}\left(\alpha_1\right)}, \sqrt{X_{4,1}\left(\alpha_1\right)}\right) \tag{4.22}$$

4.A.2.3 Comparaisons avec des expériences numériques

Sur la figure 4.12, nous comparons la condition (4.22) (avec des croix rouges) à la condition de stabilité numérique (en bleu) obtenue grâce à l'algorithme 5. La condition (4.22) est donc une bonne approximation de la condition nécessaire et suffisante de stabilité.

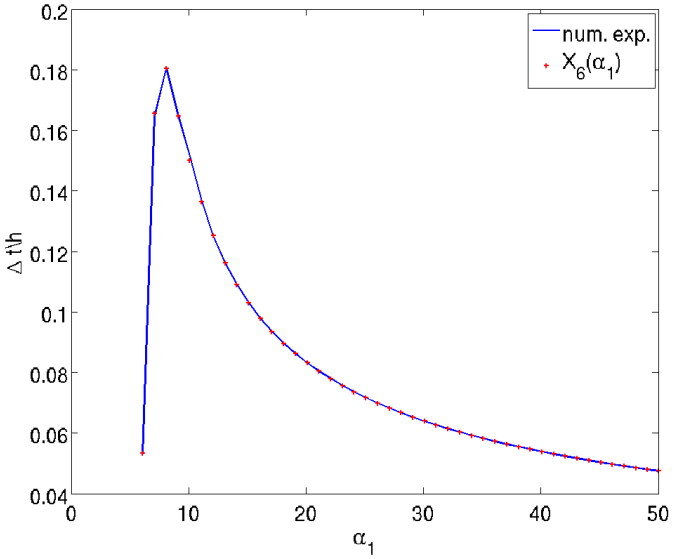

FIGURE 4.12 – Comparaison entre les conditions numérique et théorique.

Chapitre 5

Adaptativité en temps et en espace

Dans de nombreux cas pratiques, on peut trouver judicieux de combiner des approximations d'ordres différents en divers endroits du domaine étudié, que ce soit en temps ou en espace. C'est par exemple le cas lorsqu'on modélise la propagation d'ondes sismiques dans des régions montagneuses. La forte topographie doit alors être discrétisée par un maillage très fin pour prendre en compte les effets du relief. Par contre, loin de la topographie on peut utiliser un maillage beaucoup plus grossier. On doit donc faire des calculs sur un maillage beaucoup plus fin dans une zone que dans une autre et ainsi, si l'on veut obtenir une précision à peu près homogène sur l'ensemble du domaine, on peut penser utiliser des degrés polynomiaux ou des ordres en temps différents. Par exemple, on pourrait considérer une approximation par des polynômes de degré un dans la zone fine et des polynômes de degré trois dans la zone grossière. L'adaptation en espace et en temps est une approche très intéressante pour les problèmes d'ondes (cf. par exemple [6, 39, 52]) car elle permet de réduire sensiblement les coûts de calcul pour une précision souhaitée. L'intérêt de ce chapitre est donc d'étudier les propriétés d'adaptativité du Δ^p-schéma que nous avons introduit dans le chapitre 1. Nous montrons tout d'abord dans la section 5.1 que les schémas p-harmonique permettent de faire très facilement de l'adaptativité en espace et en temps. Dans la section 5.2, nous nous intéresserons à l'adaptativité en dimension un des schémas biharmonique et triharmonique et nous la comparons à celle de l'équation modifiée. Enfin dans la section 5.3, nous illustrerons l'utilité de la méthode par des calculs 2D. Finalement, nous nous intéresserons à la prise en compte de conditions aux limites absorbantes (CLA). Il n'est pas trivial de coupler ces conditions avec des schémas d'ordre élevé en temps. Une alternative consiste alors à considérer des schémas d'ordre élevé à l'intérieur du domaine de calcul et des schémas d'ordre deux au voisinage de la CLA. L'adaptativité nous permet de combiner ces deux schémas.

5.1 Adaptativité du Δ^p-schéma

Dans cette section, nous allons nous intéresser aux propriétés d'adaptativité des Δ^p-schémas. Tout d'abord, rappelons que lorsque l'on s'intéresse au schéma d'ordre quatre, on est amené à résoudre le problème variationnel suivant :

$$\left\{\begin{array}{l} \text{Trouver } u_h^{n+1} \in V_h \text{ tel que,} \qquad \forall v \in V_h, \\ \displaystyle\sum_{K\in\mathcal{T}_h} \int_K \frac{1}{\mu} \frac{u_h^{n+1} - 2u_h^n + u_h^{n-1}}{\Delta t^2} v = \quad -a_{1h}\left(u_h^n, v\right) + \frac{\Delta t^2}{12} a_{2h}\left(u_h^n, v\right). \end{array}\right. \tag{5.1}$$

où a_{1h} et a_{2h} sont les deux formes bilinéaires introduites au chapitre 1 associées aux opérateurs d'ordre deux et quatre.

Il est à remarquer que la forme bilinéaire $a_{2,h}$ ne contient que des termes faisant intervenir des opérateurs différentiels d'ordre deux au moins, hormis pour les termes de pénalisation. Par conséquent, il est clair que si l'on considère des fonctions de base de degré strictement inférieur à deux, alors la forme bilinéaire $a_{2,h}$ se réduit aux seuls termes de pénalisation. Ces derniers n'ont plus de raison d'être dans ce cas, puisqu'ils sont là pour assurer la coercivité de la forme $a_{2,h}$ et compenser l'erreur introduite en symétrisant la forme.
Ainsi, dans toute zone d'un domaine où l'on considère une approximation par des polynômes de degré un, le problème (5.1) est équivalent à

$$\left\{ \begin{array}{l} \text{Trouver } u_h^{n+1} \in V_h \text{ tel que,} \qquad \forall v \in V_h, \\ \displaystyle\sum_{K \in \mathcal{T}_h} \int_K \frac{1}{\mu} \frac{u_h^{n+1} - 2u_h^n + u_h^{n-1}}{\Delta t^2} v = \quad -a_{1h}\left(u_h^n, v\right). \end{array} \right. \tag{5.2}$$

Nous précisons également que le résultat reste vrai pour les schémas d'ordre $2p$. En effet, pour p fixé, en utilisant la même technique, le Δ^p-schéma fait intervenir p formes bilinéaires $a_{i,h}$, $1 \leq i \leq p$ et chaque forme $a_{i,h}$ ne contient que des opérateurs différentiels d'ordre au moins égal à i. Ainsi, si l'on considère des fonctions de base de degré q sur tout le domaine d'étude, les formes bilinéaires $a_{i,h}$ sont non nulles si et seulement si $q \geq i$.
Par conséquent, on obtient la propriété suivante

Remarque 5.1.1. *En utilisant les Δ^p-schémas, il est suffisant d'adapter l'ordre polynomial des fonctions de base pour obtenir un schéma adaptatif à la fois en temps et en espace.*

5.2 Résultats numériques en dimension un

Dans cette section, nous allons présenter des expériences en dimension un d'espace pour étudier l'adaptativité des différentes méthodes présentées dans ce manuscrit. Dans toutes les expériences que nous allons mener, le domaine sera un segment de longueur 8 avec une inclusion en son centre de longueur L qui sera maillée finement, la partie ne figurant pas dans cette inclusion sera elle maillée beaucoup plus grossièrement. Il est à noter que nous encadrerons cette zone fine par deux zones de transition permettant de passer progressivement de celle-ci à la zone grossière. Ce choix se justifie en partie par le fait qu'en dimension supérieure (2D et 3D), il paraît peu judicieux de coller directement des mailles très fines à des mailles très grossières. En effet, une telle configuration conduirait à considérer des maillages non conformes ou alors des mailles extrêmement déformées, ce qui peut poser des problèmes avec les méthodes par éléments finis.
Dans la suite, nous utiliserons les notations suivantes :

- h_f désigne le pas de maillage dans l'inclusion ;
- h_g désigne le pas de maillage dans le domaine privé de l'inclusion ainsi que des zones de transition ;
- h désigne le pas de maillage dans la zone de transition, ce dernier variant de h_f à h_g.

Le domaine ainsi que les notations sont illustrés dans les Figures 5.1 et 5.2. De plus, nous imposons également des conditions de bord périodiques aux deux extrémités du domaine, le terme source est supposé nul et les données initiales sont

$$u_0(x) = \left\{ \begin{array}{ll} (x - x_0)\, e^{-\left(\frac{2\pi (x - x_0)}{r_0}\right)^2} & \text{si } |x - x_0| \leq r_0 \\ 0 & \text{sinon,} \end{array} \right.$$

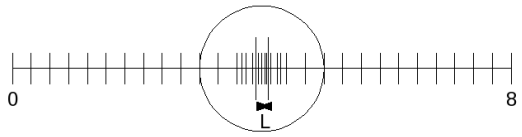

FIGURE 5.1 – Domaine de calcul

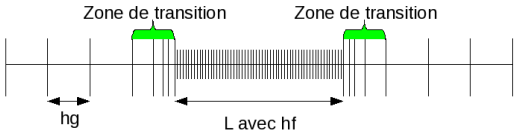

FIGURE 5.2 – Zoom sur le centre du domaine de calcul

et

$$u_1(x) = \begin{cases} \left(8\left(\frac{(x-x_0)\pi}{r_0}\right)^2 - 1\right) e^{-\left(\frac{2\pi(x-x_0)}{r_0}\right)^2} & \text{si } |x-x_0| \leq r_0 \\ 0 & \text{sinon.} \end{cases}$$

Par conséquent, la solution exacte est donnée par :

$$u^{\text{ex}}(x,t) = \sum_{i=0}^{+\infty} u_0(x+8i-t).$$

Dans la suite, nous fixons $x_0 = 2$ et $r_0 = 2$ et nous considérons deux types de configurations. Dans la première, l'inclusion est de longueur $L = 0.2$ maillée avec un pas $h_f = 1/160$ alors que dans la deuxième elle est de longueur $L = 0.2$ et maillée avec un pas $h_f = 1/320$. Dans les deux cas, h_f reste constant et nous étudions le comportement de l'erreur en fonction de h_g.

5.2.1 Maillages non uniformes

Dans cette sous-section, nous pratiquons le même type de tests que ceux effectués dans [6], ce qui nous permet dans un même temps de valider et de comparer certains de nos résultats. Ainsi, nous allons comparer, pour un temps final de simulation long (T=80s) et un temps court (T=8s), l'erreur relative $L^2([0,T],\Omega)$ obtenue pour différents pas d'espace h_g allant de 0.2 à 0.025. Dans un premier temps, nous comparerons ces résultats à ceux obtenus avec les schémas d'ordre 4 (MES-4 et Δ^2-schéma) pour des polynômes de degré 3 puis on s'intéressera aux schémas d'ordre 6 (MES-6 et Δ^3-schéma) en utilisant des éléments P^5. Il faut également préciser que nous utilisons pour coefficient de pénalisation $\gamma_1 = 8$.
Les erreurs relatives, obtenues avec un pas $h_f = 1/160$ dans la zone fine, sont données dans les tableaux 5.1 et 5.2 et sont représentées en échelle logarithmique sur les figures 5.3 et 5.4.

Comme nous pouvons le constater, les erreurs obtenues pour les deux méthodes sont extrêmement proches, ce que nous avions déjà observé au chapitre 1. Sur les figures 5.3 et 5.4, on peut aisément remarquer que pour un temps final $T = 8s$ (ligne rouge avec tirets), les deux méthodes sont d'ordre 4 et que pour un temps final plus long de $T = 80s$ (ligne verte avec tirets et pointillés), l'ordre garanti est au moins égal à 4. Néanmoins, si l'on compare ces résultats avec ceux de [6], on remarque que nous avons les mêmes résultats pour un temps final d'expérience $T = 8s$ mais que ce n'est pas exactement le cas pour $T = 80s$. En effet, dans [6] on obtient une erreur

h_g	$T = 8$s	$T = 80$s
0.2	$1.9317\ 10^{-3}$	$8.6354\ 10^{-3}$
0.1	$1.2058\ 10^{-4}$	$3.8398\ 10^{-4}$
0.05	$7.1728\ 10^{-6}$	$1.3366\ 10^{-5}$
0.025	$4.3326\ 10^{-7}$	$4.7794\ 10^{-7}$

TABLE 5.1 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour le MES-4 ($L = 0.2$ et $h_f = 1/160$).

h_g	$T = 8$s	$T = 80$s
0.2	$1.9333\ 10^{-3}$	$8.6585\ 10^{-3}$
0.1	$1.2058\ 10^{-4}$	$3.8198\ 10^{-4}$
0.05	$7.1706\ 10^{-6}$	$1.3285\ 10^{-5}$
0.025	$4.3006\ 10^{-7}$	$4.7746\ 10^{-7}$

TABLE 5.2 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour le Δ^2-schéma ($L = 0.2$ et $h_f = 1/160$).

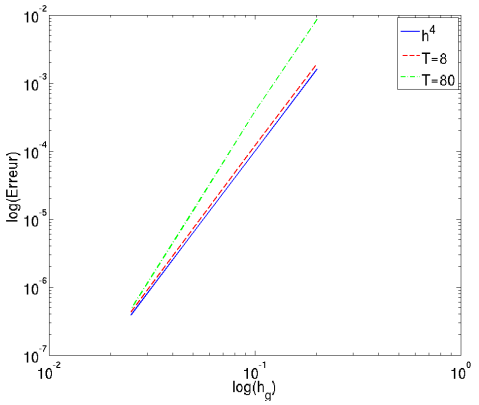

FIGURE 5.3 – Courbes de convergence pour le schéma MES-4 ($L = 0.2$ et $h_f = 1/160$).

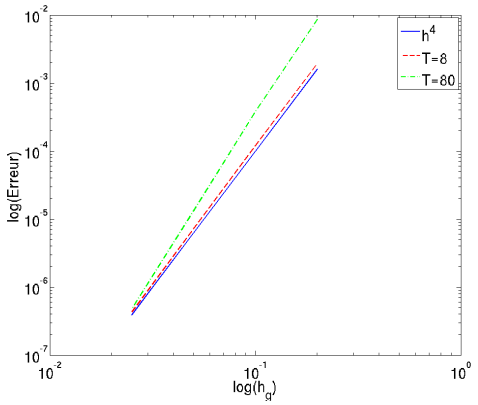

FIGURE 5.4 – Courbes de convergence pour le Δ^2-schéma ($L = 0.2$ et $h_f = 1/160$).

plus élevée pour $h_g = 0.2$. La différence avec nos résultats provient du fait que nous avons choisi une valeur différente pour le coefficient de pénalisation. En effet, nous avons pris $\alpha = 8$ alors que dans [6] le choix s'est porté sur $\alpha = 7$. Nous avons donc refait les mêmes tests avec $\alpha = 7$ et nous obtenons bien les mêmes résultats. On remarque ainsi que la valeur du coefficient de pénalisation influe sur l'ordre de convergence de la méthode.

On propose le même type d'expériences pour un pas $h_f = 1/320$ dans la zone fine afin d'analyser l'importance du choix du pas d'espace dans la zone fine sur l'erreur relative. On présente les erreurs relatives obtenues avec un tel pas d'espace dans le Tableau 5.3 (resp. dans le Tableau 5.4) pour l'équation modifiée (resp. le Δ^2-schéma). Les résultats étant très proches, nous nous contentons donc de présenter les courbes de convergence pour le Δ^2-schéma sur la figure 5.5.

Les courbes de la figure 5.5 donnent le même type de résultats que ceux obtenus pour $h_f = 1/160$. En effet, pour un temps final $T = 8s$ (ligne rouge avec tirets), on a bien une convergence à l'ordre 4 tandis que pour $T = 80s$ (ligne verte avec tirets et pointillés), l'ordre de convergence est au moins égal à 4.

Pour analyser ces résultats, rappelons que l'erreur du schéma peut se mettre sous la forme $E = C_1 h_g^4 + C_2 h_f^4$. Comme nous ne faisons varier que h_g, le fait que l'erreur converge à l'ordre quatre montre que dans tous les cas que nous avons étudiés, le terme $C_1 h_g^4$ (l'erreur dans les mailles grossières) est beaucoup plus fort que le terme $C_2 h_f^4$ (l'erreur dans les mailles fines). Dans le cas contraire, l'erreur aurait convergé vers $C_2 h_f^4$.

h_g	$T = 8$s	$T = 80$s
0.2	$1.8824\ 10^{-3}$	$8.2809\ 10^{-3}$
0.1	$1.2200\ 10^{-4}$	$3.9479\ 10^{-4}$
0.05	$7.2048\ 10^{-6}$	$1.3523\ 10^{-5}$
0.025	$4.3500\ 10^{-7}$	$4.8655\ 10^{-7}$

TABLE 5.3 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour le MES-4 ($L = 0.1$ et $h_f = 1/320$).

h_g	$T = 8$s	$T = 80$s
0.2	$1.8828\ 10^{-3}$	$8.2865\ 10^{-3}$
0.1	$1.2200\ 10^{-4}$	$3.9428\ 10^{-4}$
0.05	$7.2042\ 10^{-6}$	$1.3501\ 10^{-5}$
0.025	$4.3420\ 10^{-7}$	$4.8551\ 10^{-7}$

TABLE 5.4 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour le Δ^2-schéma ($L = 0.1$ et $h_f = 1/320$).

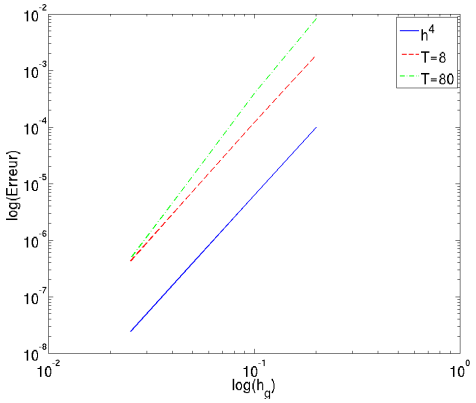

FIGURE 5.5 – Courbes de convergence pour le Δ^2-schéma ($L = 0.1$ et $h_f = 1/320$).

Pour gagner en temps de calcul, nous pourrions donc perdre en précision dans la zone fine sans pénaliser la qualité de la solution. Une première idée serait de considérer des mailles plus larges, mais dans la pratique, la taille des mailles est souvent contrainte par la configuration du domaine. L'autre solution, comme nous l'avons évoqué en introduction, est d'utiliser un ordre d'approximation plus faible dans la zone fine. On diminue ainsi le nombre de degrés de liberté et le coût de calcul global. En utilisant par exemple des éléments P^1 au lieu d'éléments P^3, on divise par deux le nombre d'inconnues dans la zone fine. De plus, la condition CFL pour des éléments P^1 est moins contraignante que celle associée à des éléments P^3. On pourra donc utiliser un pas de temps plus élevé.

5.2.2 Adaptation de l'ordre en temps et en espace

Dans cette sous-section, nous étudions la possibilité d'utiliser des éléments P^1 dans la zone fine. Il est à noter que la méthode de Galerkine discontinue se prête tout à fait à ce genre d'expérience grâce aux discontinuités des fonctions de base. En effet, on peut facilement affecter à tout élément du maillage un certain ordre et à son voisin un ordre différent. Il faut juste prendre garde au choix du paramètre de pénalisation entre deux éléments n'étant pas du même ordre. Le terme de pénalisation venant assurer la continuité du flux à l'interface entre deux éléments, il semble logique d'utiliser le coefficient de pénalisation le plus fort entre ces deux éléments. Ainsi, lorsque l'on rencontre, par exemple, une interface entre un élément d'ordre deux et un élément d'ordre quatre, on utilise le coefficient de pénalisation correspondant à un élément d'ordre 4, à savoir $\gamma_1 = 8$. En ce qui concerne la pénalisation de l'opérateur biharmonique, nous avons considéré

le paramètre qui nous avait semblé optimal numériquement dans le cas de maillage régulier (cf. chapitre 1) à savoir $\gamma_{2,1} = 10$.
Dans les expériences qui sont présentées par la suite, nous utiliserons donc des éléments P^1 dans la zone fine avec le schéma saute-moutons et des éléments P^3 partout ailleurs (zone grossière et zone de transition) ce que nous noterons dans les légendes MES-4-LF et Δ^2-schéma-LF (LF étant l'acronyme de "Leap-Frog scheme" i.e. schéma saute-moutons). Les erreurs obtenues à $T = 8s$ et $80s$ avec un pas d'espace fin $h_f = 1/160$ et h_g variant de 0.2 à 0.025 avec le MES-4 (resp. Δ^2-schéma) sont présentées dans le tableau 5.5 (resp. tableau 5.6). La figure 5.6 représente les erreurs des deux schémas à $T = 8s$ en fonction du pas d'espace en échelle logarithmique (les résultats à $T = 80s$ conduisent à des courbes similaires).

h_g	$T = 8$s	$T = 80$s
0.2	$1.9351\ 10^{-3}$	$8.6716\ 10^{-3}$
0.1	$1.7514\ 10^{-4}$	$1.0313\ 10^{-3}$
0.05	$1.0629\ 10^{-4}$	$7.2900\ 10^{-4}$
0.025	$1.0545\ 10^{-4}$	$7.2159\ 10^{-4}$

TABLE 5.5 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour le MES-4 ($L = 0.2$ et $h_f = 1/160$).

h_g	$T = 8$s	$T = 80$s
0.2	$2.0069\ 10^{-3}$	$9.4069\ 10^{-3}$
0.1	$5.6599\ 10^{-4}$	$3.8479\ 10^{-3}$
0.05	$5.5002\ 10^{-4}$	$3.7935\ 10^{-3}$
0.025	$5.4989\ 10^{-4}$	$3.7924\ 10^{-3}$

TABLE 5.6 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour le Δ^2-schéma ($L = 0.2$ et $h_f = 1/160$) avec $\gamma_{2,1} = 10$.

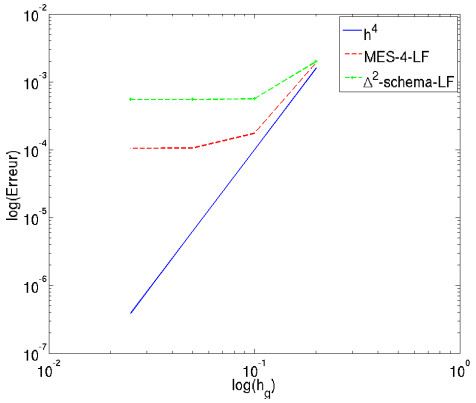

FIGURE 5.6 – Courbes de convergence pour le schéma MES-4 et le Δ^2-schéma ($L = 0.2$ et $h_f = 1/160$).

Dans un premier temps, on peut constater sur ces tableaux et figure que la technique de l'équation modifiée donne de meilleurs résultats que le Δ^2-schéma.
Si l'on s'intéresse aux résultats de la figure 5.6, on constate, de la même façon que dans [6], que pour h_g grand, les erreurs obtenues en utilisant des éléments du même ordre et celles obtenues avec des ordres différents sont extrêmement proches voire identiques. Lorsque le pas d'espace dans la zone grossière h_g diminue, l'erreur tend vers une constante ce qui signifie que l'erreur effectuée dans la zone fine prédomine sur l'erreur effectuée dans la zone grossière. Ce comportement

s'explique par le fait que l'erreur est cette fois de la forme

$$E = C_1 h_g^4 + C_2 h_f^2.$$

Comme nous ne faisons varier que h_g, le terme $C_1 h_g^4$ devient négligeable devant $C_2 h_f^2$ pour h_g assez petit et l'erreur ne décroît plus. Remarquons que ce phénomène arrive beaucoup plus tard pour le MES-4 que pour le Δ^2-schéma, ce qui semble montrer que le MES-4 est plus propice à l'adaptativité que le Δ^2-schéma. Afin d'étudier l'influence du pas d'espace dans la zone fine, on présente le même type de résultats pour $h_g = 1/320$ dans les tableaux 5.7 et 5.8. On représente sur la figure 5.7 (resp. figure 5.8), en échelle logarithmique l'erreur obtenue pour $T = 8s$ avec la MES (resp. Δ^2-schéma) que l'on comparera aux courbes obtenues sans adaptation de l'ordre ni en espace ni en temps.

h_g	$T = 8$s	$T = 80$s
0.2	$1.8824\ 10^{-3}$	$8.2807\ 10^{-3}$
0.1	$1.2554\ 10^{-4}$	$4.6554\ 10^{-4}$
0.05	$1.7494\ 10^{-5}$	$1.1134\ 10^{-4}$
0.025	$1.5443\ 10^{-5}$	$1.0433\ 10^{-4}$

TABLE 5.7 – Erreur relative $L^2([0,T],\Omega)$ pour le MES-4 ($L = 0.1$ et $h_f = 1/320$).

h_g	$T = 8$s	$T = 80$s
0.2	$1.8912\ 10^{-3}$	$8.3655\ 10^{-3}$
0.1	$2.2733\ 10^{-4}$	$1.3670\ 10^{-3}$
0.05	$1.9232\ 10^{-4}$	$1.3125\ 10^{-3}$
0.025	$1.7906\ 10^{-4}$	$1.3124\ 10^{-3}$

TABLE 5.8 – Erreur relative $L^2([0,T],\Omega)$ pour le Δ^2-schéma ($L = 0.1$ et $h_f = 1/320$) avec $\gamma_{2,1} = 10$.

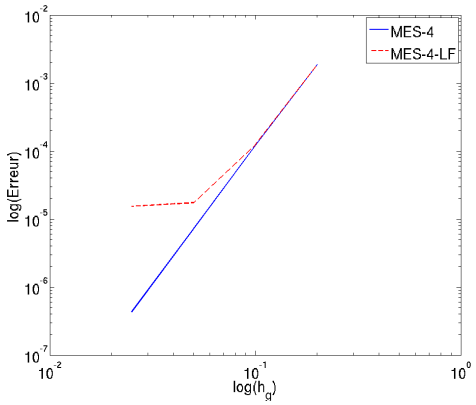

FIGURE 5.7 – Courbes de convergence pour le schéma MES-4 ($L = 0.1$ et $h_f = 1/320$).

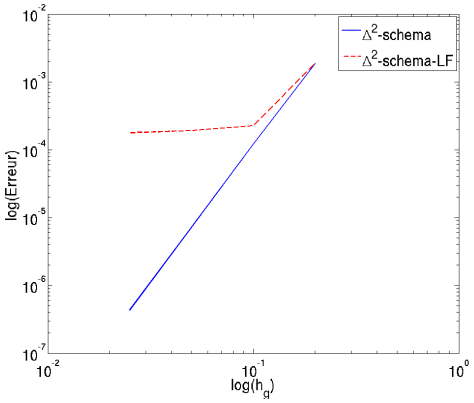

FIGURE 5.8 – Courbes de convergence pour le Δ^2-schéma ($L = 0.1$ et $h_f = 1/320$) avec $\gamma_{2,1} = 10$.

Lorsqu'on s'intéresse à la convergence du Δ^2-schéma, on constate que l'ordre de convergence n'est même pas assuré à l'ordre minimal qu'est l'ordre deux. En effet, comparons par exemple les erreurs relatives des cas $h_g = 0.1$ du tableau 5.5 et $h_g = 0.05$ du tableau 5.3. Le maillage utilisé dans le deuxième cas est deux fois plus fin que dans le premier cas (dans la zone grossière et dans la zone fine). Le rapport entre les deux erreurs est d'environ 10, ce qui correspond bien à un ordre de convergence d'une méthode mêlant schéma d'ordre 2 et schéma d'ordre 4. Par contre, si on

effectue le même rapport (grâce aux tableaux 5.6 et 5.8) dans le cas du Δ^2-schéma, on obtient un rapport de 2.94, ce qui n'est pas du tout satisfaisant.
Nous avons donc décidé de refaire cette série d'expériences en considérant le Δ^2-schéma sans paramètre de pénalisation sur l'opérateur biharmonique ($\gamma_{2,1} = 0$). On rappelle que dans le chapitre 4, nous avions étudié ce cas et constaté que la seule pénalisation de l'opérateur harmonique suffit à assurer la stabilité du schéma. De plus, dans le cas présent, la condition CFL ne sera pas affectée par ce choix dans la mesure où celle-ci est dictée par la zone fine dans laquelle on considère le schéma saute-moutons avec des éléments P^1. Les résultats sont présentés dans les tableaux 5.9 et 5.10 respectivement pour $h_f = 1/160$ et $h_f = 1/320$. Les figures 5.9 et 5.10 permettent de comparer la nouvelle courbe de convergence à celle du MES-4, respectivement pour $h_f = 1/160$ et $h_f = 1/320$. On remarque que dans ce cas, les erreurs sont du même ordre. Ce résultat montre à nouveau l'importance du choix des paramètres de pénalisation.

h_g	$T = 8$s	$T = 80$s
0.2	$1.9364\ 10^{-3}$	$8.6862\ 10^{-3}$
0.1	$1.7517\ 10^{-4}$	$1.0302\ 10^{-3}$
0.05	$1.0628\ 10^{-5}$	$7.2847\ 10^{-4}$
0.025	$1.0544\ 10^{-5}$	$7.2105\ 10^{-4}$

TABLE 5.9 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour le Δ^2-schéma ($L = 0.2$ et $h_f = 1/160$) avec $\gamma_{2,1} = 0$.

h_g	$T = 8$s	$T = 80$s
0.2	$1.8827\ 10^{-3}$	$8.2845\ 10^{-3}$
0.1	$1.2557\ 10^{-4}$	$4.6529\ 10^{-4}$
0.05	$1.7498\ 10^{-5}$	$1.1125\ 10^{-4}$
0.025	$1.5446\ 10^{-5}$	$1.0424\ 10^{-4}$

TABLE 5.10 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour le Δ^2-schéma ($L = 0.1$ et $h_f = 1/320$) avec $\gamma_{2,1} = 0$.

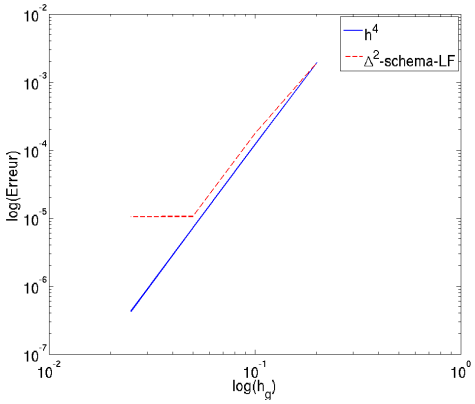

FIGURE 5.9 – Courbes de convergence pour le Δ^2-schéma ($L = 0.2$ et $h_f = 1/160$) avec $\gamma_{2,1} = 0$.

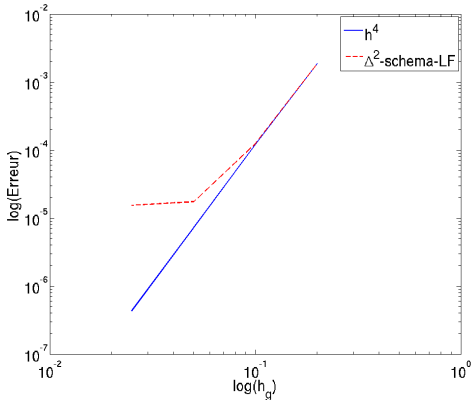

FIGURE 5.10 – Courbes de convergence pour le Δ^2-schéma ($L = 0.1$ et $h_f = 1/320$) avec $\gamma_{2,1} = 0$.

Les résultats présentés dans les tableux 5.9 et 5.10 sont à présent très proches de ceux obtenus dans les tableaux 5.5 et 5.3 avec l'équation modifiée. Si on s'intéresse au même rapport d'erreur que précédemment, on obtient aussi un rapport de 10, ce qui est bien plus acceptable. Nous allons maintenant nous intéresser au cas du Δ^3-schéma.

5.2.3 Le cas du Δ^3-schéma

Dans cette sous-section, nous allons nous intéresser au cas de l'adaptativité du Δ^3-schéma. Nous allons effectuer le même type d'expérience que dans la sous-section 5.2.2.

5.2.3.1 Adaptation schéma d'ordre 6 - schéma d'ordre 2

Nous considérons des éléments P^1 dans la zone fine avec le schéma saute-moutons et des éléments P^5 partout ailleurs (zone grossière et zone de transition) avec le Δ^3-schéma (resp. le MES-6) ce que nous noterons dans les légendes Δ^3-LF-schéma et MES-6-LF. Les erreurs obtenues à $T = 8s$ avec un pas d'espace fin $h_f = 1/160$ et h_g variant de 0.4 à 0.05 avec le MES-6-LF et le Δ^3-LF-schéma sont présentées dans le tableau 5.11. Dans ce tableau, le cas 1 correspond au cas où l'on pénalise les différents opérateurs avec les valeurs que nous avions déterminées numériquement au chapitre 1, à savoir $\gamma_1 = \gamma_{2,1} = \gamma_{3,1} = 20$ avec des éléments P^5. Le cas 2 correspond au cas où l'on ne pénalise ni l'opérateur biharmonique ni l'opérateur triharmonique. Sur la figure 5.11, nous avons représenté les erreurs obtenues avec le Δ^3-LF-schéma dans les cas 1 (en vert avec tirets et croix) et 2 (en rouge avec tirets) en fonction du pas d'espace en échelle logarithmique. Nous précisons que nous n'avons pas représenté les résultats de l'équation modifiée puisque ceux-ci sont extrêmement proches du Δ^3-LF-schéma dans le cas 2.

h_g	MES-6-4	Δ^3-LF-schéma : cas 1	Δ^3-LF-schéma : cas 2
0.4	$5.7331\ 10^{-3}$	$5.7416\ 10^{-3}$	$5.7330\ 10^{-3}$
0.2	$1.0566\ 10^{-4}$	$3.9161\ 10^{-4}$	$1.0517\ 10^{-4}$
0.1	$1.0448\ 10^{-4}$	$3.9073\ 10^{-4}$	$1.0395\ 10^{-4}$
0.05	$1.0444\ 10^{-4}$	$3.6208\ 10^{-4}$	$1.0394\ 10^{-4}$

TABLE 5.11 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour $L = 0.2$ et $h_f = 1/160$.

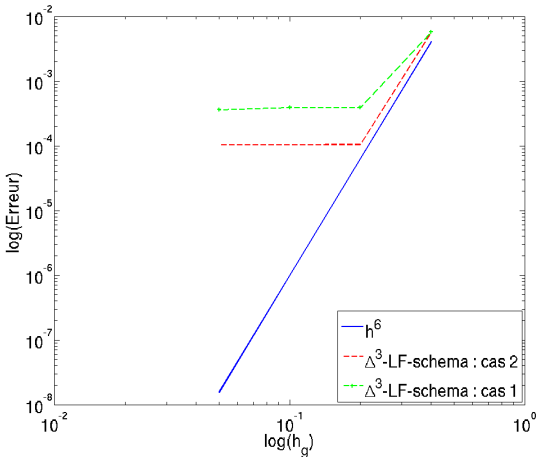

FIGURE 5.11 – Courbes de convergence pour le Δ^3-LF-schéma avec $L = 0.2$ et $h_f = 1/160$.

Les conclusions qu'on peut tirer sont sensiblement les mêmes que celles de la sous-section précédente : le fait de ne pas considérer de pénalisations sur les opérateurs biharmonique et tri-

harmonique permet d'obtenir des résultats identiques à ceux du schéma MES-6-LF ; et lorsque le pas d'espace dans la zone grossière h_g diminue, l'erreur tend vers une constante. Là encore, l'erreur effectuée dans la zone fine prédomine sur l'erreur effectuée dans la zone grossière. En effet, l'erreur est cette fois de la forme

$$E = C_1 h_g^6 + C_2 h_f^2.$$

Comme nous ne faisons varier que h_g, le terme $C_1 h_g^6$ devient rapidement négligeable devant $C_2 h_f^2$ pour h_g assez petit et l'erreur ne décroît plus.
Ainsi, comme précédemment, nous allons analyser l'influence du pas d'espace dans la zone fine. On reprend donc les mêmes expériences mais en fixant cette fois le pas dans la zone fine à $h_f = 1/320$. On représente dans le tableau 5.12 les résultats obtenus pour le MES-6-LF et pour le Δ^3-LF-schéma dans les cas 1 et 2. Ici, on constate à nouveau que les résultats obtenus avec le Δ^3-LF-schéma dans le cas 2 sont très proches de ceux correspondant au MES-6-LF alors que lorsque l'on pénalise les opérateurs biharmonique et triharmonique, les erreurs relatives sont bien moins bonnes.

h_g	MES-6-4	Δ^3-LF-schéma : cas 1	Δ^3-LF-schéma : cas 2
0.4	$5.7320\ 10^{-3}$	$5.7343\ 10^{-3}$	$5.7320\ 10^{-3}$
0.2	$2.1178\ 10^{-5}$	$1.5823\ 10^{-4}$	$2.1127\ 10^{-5}$
0.1	$1.5302\ 10^{-5}$	$1.4540\ 10^{-4}$	$1.5232\ 10^{-5}$
0.05	$1.5297\ 10^{-5}$	$1.4648\ 10^{-4}$	$1.5227\ 10^{-5}$

TABLE 5.12 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour $L = 0.1$ et $h_f = 1/320$.

On a représenté en échelle logarithmique les résultats du Δ^3-LF-schéma dans le cas 1 (en vert avec tirets et croix) et dans le cas 2 (en rouge avec tirets) sur la figure 5.12. La courbe correspondant au cas du MES-6-2 n'a pas été représentée sur cette figure puisque les valeurs sont extrêmement proches de celles du Δ^3-LF-schéma dans le cas 2. On constate ici aussi que lorsqu'on pénalise les opérateurs biharmonique et triharmonique, l'ordre de convergence n'est même pas assuré à l'ordre minimal qui est l'ordre deux. Dans le cas contraire, on obtient une convergence comparable à celles des schémas MES.

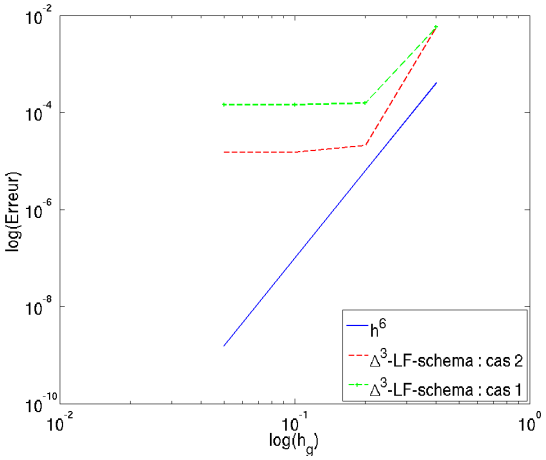

FIGURE 5.12 – Courbes de convergence pour le Δ^3-LF-schéma avec $L = 0.2$ et $h_f = 1/160$.

5.2.3.2 Adaptation schéma d'ordre 6 - schéma d'ordre 4

A présent, nous allons considérer des éléments P^3 dans la zone fine avec le Δ^2-schéma (resp. le MES-4) et des éléments P^5 partout ailleurs (zone grossière et zone de transition) avec le Δ^3-schéma (resp. le MES-6) ce qui sera noté dans les légendes Δ^3-Δ^2-schéma et MES-6-4. Les erreurs obtenues à $T = 8s$ avec un pas d'espace fin $h_f = 1/160$ et h_g variant de 0.4 à 0.05 avec le MES-6-4 et le Δ^3-Δ^2-schéma sont présentées dans le tableau 5.13. Comme précédemment, le cas 1 correspond au cas où l'on pénalise les différents opérateurs avec les valeurs que nous avions déterminées numériquement au chapitre 1, à savoir $\gamma_1 = \gamma_{2,1} = \gamma_{3,1} = 20$ avec les éléments P^5 et $\gamma_1 = 8$ et $\gamma_{2,1} = 10$ avec les éléments P^3. Le cas 2 est toujours le cas où l'on ne pénalise ni l'opérateur biharmonique ni l'opérateur triharmonique. Sur la figure 5.13, nous avons représenté les erreurs obtenues avec le Δ^3-Δ^2-schéma dans les cas 1 (en vert avec tirets et croix) et 2 (en rouge avec tirets) en fonction du pas d'espace en échelle logarithmique. Ici aussi, les résultats obtenus en utilisant l'équation modifiée sont très proches de ceux obtenus avec le Δ^3-Δ^2-schéma dans le cas 2 ce qui explique que nous ne représentons que les résultats issus du cas 2.

h_g	MES-6-4	Δ^3-Δ^2-schéma : cas 1	Δ^3-Δ^2-schéma : cas 2
0.4	$5.7318\ 10^{-3}$	$5.7363\ 10^{-3}$	$5.7318\ 10^{-3}$
0.2	$1.4684\ 10^{-5}$	$2.4262\ 10^{-4}$	$1.4683\ 10^{-5}$
0.1	$2.7331\ 10^{-7}$	$2.4167\ 10^{-4}$	$2.7353\ 10^{-7}$
0.05	$4.5607\ 10^{-9}$	$2.1197\ 10^{-4}$	$4.5635\ 10^{-9}$

TABLE 5.13 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour $L = 0.2$ et $h_f = 1/160$.

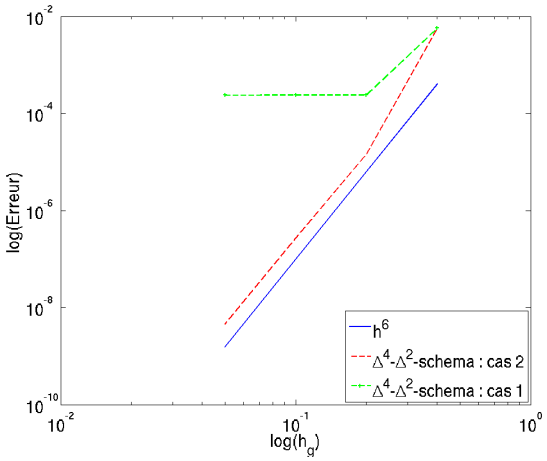

FIGURE 5.13 – Courbes de convergence pour le Δ^3-Δ^2-schéma avec $L = 0.2$ et $h_f = 1/160$.

On obtient des résultats identiques à ceux du schéma MES-6-LF en ne considérant pas de pénalisations sur les opérateurs biharmonique et triharmonique. Contrairement au cas précédent, on constate sur la figure 5.13 que l'on obtient une convergence à l'ordre 6. L'erreur est cette fois de la forme

$$E = C_1 h_g^6 + C_2 h_f^4.$$

Pour obtenir le même type de résultat que précédemment, c'est-à-dire une erreur tendant vers une constante, il faut que le terme $C_1 h_g^6$ devienne négligeable devant $C_2 h_f^4$, ce qui arriverait si l'on considérait h_g encore plus petit.

On reprend à présent les mêmes expériences mais en fixant cette fois le pas dans la zone fine à $h_f = 1/320$. On représente dans le tableau 5.14 les résultats obtenus pour le MES-6-Δ^2 et pour les Δ^3-Δ^2-schémas dans les cas 1 et 2. Ici, on constate à nouveau que les résultats obtenus avec le Δ^3-Δ^2-schéma dans le cas 2 sont très proches de ceux correspondant au MES-6-LF alors que lorsqu'on pénalise les opérateurs biharmonique et triharmonique, les erreurs relatives sont bien moins bonnes. Les résultats présentés dans ce tableau sont similaires à ceux du tableau 5.13, voire moins bons, ce qui peut s'expliquer par le fait que les erreurs de troncature dans la zone fine l'emportent sur l'erreur globale.

h_g	MES-6-4	Δ^3-Δ^2-schéma : cas 1	Δ^3-Δ^2-schéma : cas 2
0.4	$5.7318\ 10^{-3}$	$5.7330\ 10^{-3}$	$5.7318\ 10^{-3}$
0.2	$1.4478\ 10^{-5}$	$1.1349\ 10^{-4}$	$1.4479\ 10^{-5}$
0.1	$2.7538\ 10^{-7}$	$1.0595\ 10^{-4}$	$2.7543\ 10^{-7}$
0.05	$4.5770\ 10^{-9}$	$1.0113\ 10^{-4}$	$4.5790\ 10^{-9}$

TABLE 5.14 – Erreur relative $L^2\left([0,T],\Omega\right)$ pour $L = 0.1$ et $h_f = 1/320$.

On a représenté en échelle logarithmique les résultats du Δ^3-Δ^2-schéma dans le cas 1 (en vert avec tirets et croix) et dans le cas 2 (en rouge avec tirets) sur la figure 5.14. La courbe correspondant au cas du MES-6-4 n'a pas été représentée sur cette figure puisque les valeurs sont extrêmement proches de celles du Δ^3-Δ^2-schéma dans le cas 2. Les conclusions que l'on peut tirer de cette figure sont exactement les mêmes que celles de la figure 5.13.

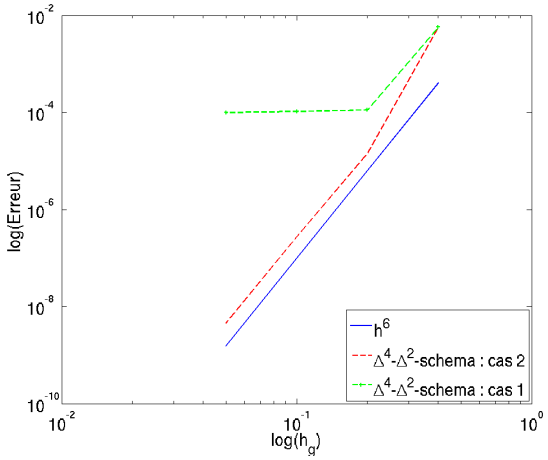

FIGURE 5.14 – Courbes de convergence pour le Δ^3-Δ^2-schéma avec $L = 0.2$ et $h_f = 1/160$.

Nous allons maintenant nous intéresser à des expériences en dimension deux.

5.3 Résultats numériques en dimension deux

Dans cette section, nous allons nous intéresser à des cas en dimension deux d'espace qui peuvent nous permettre de tester l'adaptativité de la méthode sur des cas simples. Dans toutes les expériences que nous allons présenter, nous considérons des données initiales nulles et une source qui est une dérivée seconde de Gaussienne en temps et un point source en espace

$$f = \delta_{x_0} 2\lambda \left(\lambda (t-t_0)^2 - 1\right) e^{-\lambda (t-t_0)^2}, \tag{5.3}$$

avec $\lambda = \pi^2 f_0^2$, $f_0 = 5$ et $t_0 = 1/f_0$ et x_0 désigne le point d'application de cette source.
Le premier cas auquel on s'intéressera dans la sous-section 5.3.1 sera celui d'une fine zone maillée par des éléments P^1 venant s'insérer entre deux couches beaucoup plus importantes maillées par des éléments P^3 beaucoup plus grossiers. Dans la sous-section 5.3.2, on s'intéressera à un domaine bicouche dont l'interface n'est plus plane mais sinusoidale et on étudiera enfin, dans la sous-section 5.3.3, le cas d'un domaine avec un fort rétrécissement qui sera beaucoup plus caractéristique des problèmes pratiques que l'on peut rencontrer.

5.3.1 Fine couche

Dans cette sous-section, on va s'intéresser à un domaine présentant une fine couche entre deux couches beaucoup plus épaisses comme présenté dans la figure 5.15. Le domaine est le carré $[-2,2]^2$ où la zone fine est $\Omega_3 = [-2,2] \times [-0.02, 0.02]$ et la partie supérieure (resp. inférieure) sera notée Ω_1 (resp. Ω_2). Le but de cette expérience est de considérer une zone assez fine qui ne contient dans la largeur qu'une seule maille afin d'illustrer qu'il n'y a aucun problème à avoir des changements d'ordre même sur une zone très peu étendue.

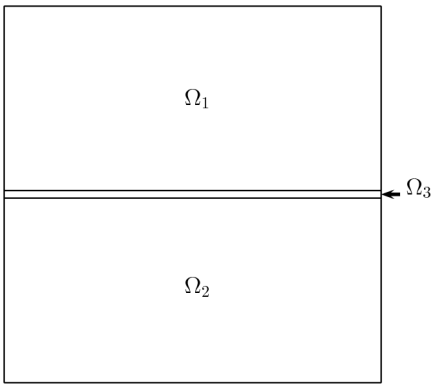

FIGURE 5.15 – Domaine avec une fine couche Ω_3.

Le maillage que nous avons utilisé est présenté sur la figure 5.16 et pour plus de clarté, nous avons effectué un zoom sur la zone fine pour montrer qu'elle ne contient effectivement que très peu de mailles dans sa largeur (figure 5.17).

Par la suite, sur un tel domaine, deux choses peuvent être intéressantes. La première est de considérer une vitesse constante sur tout le domaine afin de s'assurer que l'adaptation de l'ordre en espace ne provoque aucun effet parasite au passage de cette zone. La deuxième est de considérer une même vitesse dans les zones Ω_1 et Ω_2 et une vitesse différente dans la zone Ω_3 pour vérifier que des variations de caractéristiques physiques peuvent être prises en compte numériquement même dans des zones très fines. Ainsi, dans la première expérience, nous allons considérer

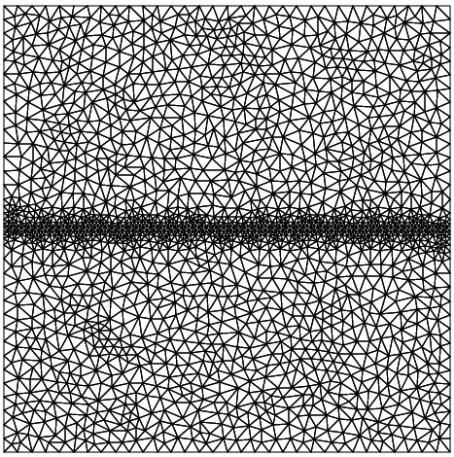

FIGURE 5.16 – Maillage du domaine avec fine couche.

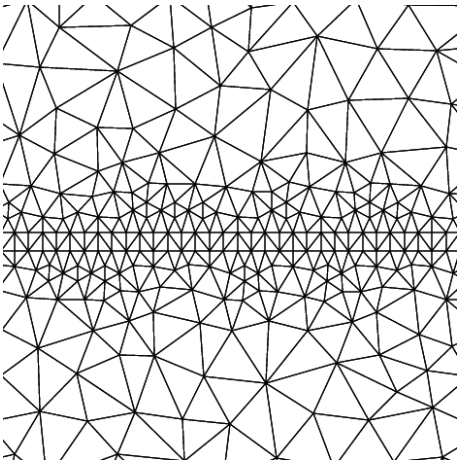

FIGURE 5.17 – Zoom sur le maillage de la fine couche.

$\mu = \rho = 2$ dans l'intégralité du domaine et nous utiliserons le Δ^2-schéma sur Ω_1 et Ω_2 avec des polynômes de degré trois alors que dans le sous-domaine Ω_2, on utilisera le schéma saute-moutons avec des éléments P^1. La source est placée en $x_0 = (0, 1)$ et un instantané de la simulation est donné sur la figure 5.18. On remarque qu'aucun effet numérique indésirable n'apparaît lorsque

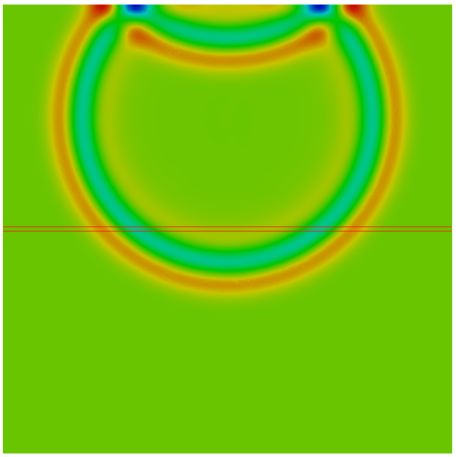

FIGURE 5.18 – Simulation sur le domaine avec zone fine

l'onde incidente traverse le domaine Ω_3. Afin de confirmer cette impression, nous allons utiliser la même méthode qu'au chapitre 1, c'est-à-dire une méthode de Cagniard-De Hoop, qui donne la solution analytique en un point du domaine au cours du temps. Nous avons placé un récepteur au point $(0, -1)$, c'est-à-dire de l'autre côté de la fine couche par rapport à la source et nous avons ainsi pu comparer notre solution approchée à la solution analytique en ce point. Les résultats obtenus sont présentés sur la figure 5.19 où nous avons tracé en bleu la solution approchée et en rouge la solution analytique en fonction du temps. Les résultats étant très proches, nous avons tracé le résidu entre ces deux solutions sur la figure 5.20 et on peut constater que la solution approchée est une bonne approximation de la solution exacte. Ainsi, on peut en conclure que le fait de traiter une zone du domaine avec des ordres différents, que ce soit en temps ou en espace, ne joue ni sur l'allure générale ni sur le comportement de la solution.

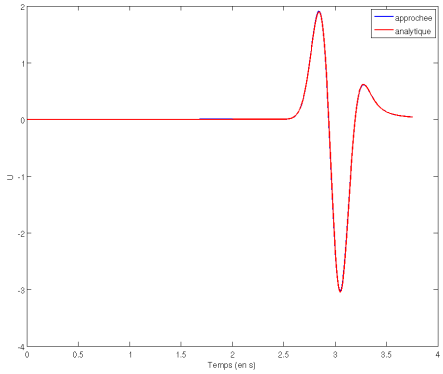

FIGURE 5.19 – Sismogrammes sur le domaine avec zone fine

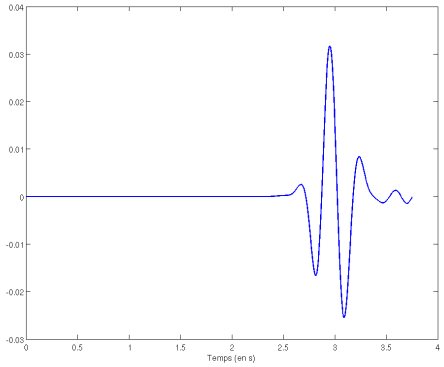

FIGURE 5.20 – Résidu entre la solution approchée et la solution analytique.

La deuxième expérience reprend la même configuration que celle de la première. La seule différence provient du fait que l'on considère des caractéristiques physiques différentes dans le domaine Ω_3 : $\mu = 8$ et $\rho = 4$. On présente un instantané de la propagation sur la figure 5.21. Il n'est pas évident sur une zone aussi mince de voir le changement de vitesse. Néanmoins, la réflexion issue de l'interface est bien caractéristique d'une telle variation. Ces résultats tendent donc à confirmer qu'il n'y a pas de problème majeur à considérer des variations d'ordre, que ce soit en espace ou en temps, dans des zones de tailles plus ou moins importantes.

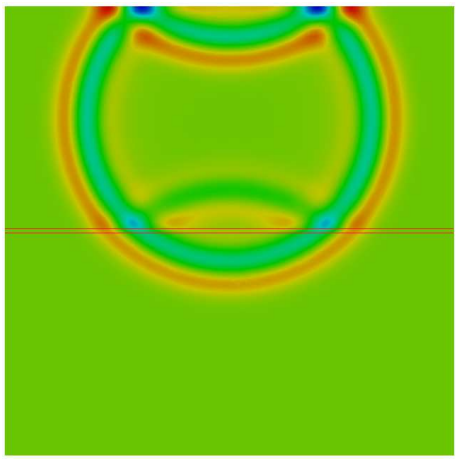

FIGURE 5.21 – Simulation sur le domaine avec zone fine et contraste de vitesses.

5.3.2 Interface sinusoidale

Dans cette partie, on va s'intéresser au cas d'une interface sinusoïdale qui nécessite d'être maillée finement par rapport au reste du domaine. Le domaine est un carré $\Omega = [-1, 1]^2$ et l'interface est définie par la fonction qui à $x \in [-1, 1]$ associe $\dfrac{\sin(5\pi x)}{10}$. Le nombre d'oscillations est ainsi assez conséquent par rapport à la taille du domaine comme le montre la figure 5.22. La fontière sinusoidale sépare Ω en deux sous-domaines, Ω_1 au dessus de la frontière et Ω_2 en dessous. Il

est clair que l'on doit mailler très finement l'interface pour prendre en comptre sa géométrie (cf. figure 5.23). Or, si nous maillons finement cette zone, il n'est pas nécessaire de considérer une approximation polynomiale d'aussi haut degré que dans le reste du domaine de calcul.

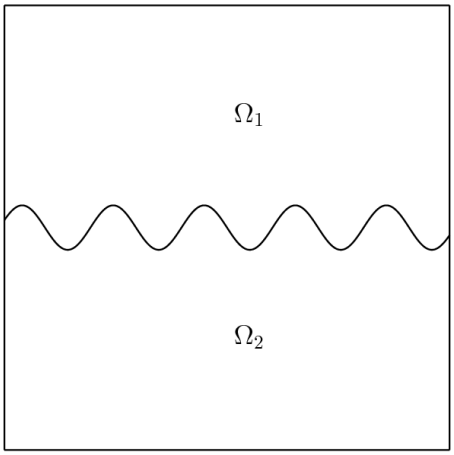

FIGURE 5.22 – Domaine avec interface sinusoidale.

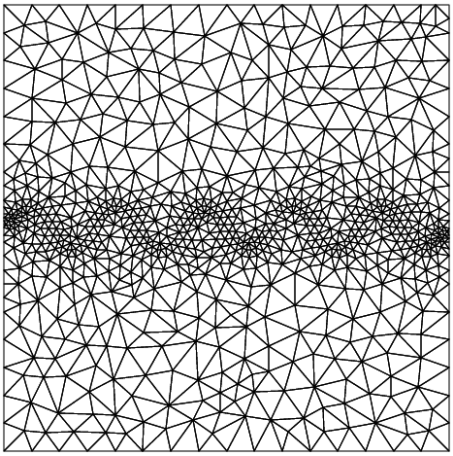

FIGURE 5.23 – Maillage du domaine avec interface sinusoidale.

Dans les expériences numériques que nous avons menées, nous avons imposé $\mu = \rho = 2$ dans Ω_1 et $\mu = 8$, $\rho = 4$ dans la partie Ω_2. Contrairement aux cas précédents où la zone fine était clairement déterminée, ici, nous ne savons pas *a priori* où utiliser le schéma d'ordre deux et des éléments P^1. Un critère discriminant est donc nécessaire pour effectuer un tel choix et nous proposons d'utiliser l'aire des mailles de notre triangulation. Nous considérons le schéma saute-moutons et des éléments P^1 si l'aire d'une maille est plus petite que l'aire totale du domaine divisé par le nombre de mailles du maillage (c'est-à-dire l'aire moyenne). Sur le cas que l'on présente ici, on a généré un maillage comportant 1564 éléments. Ainsi toutes les mailles ayant une aire inférieure à $4/1564 = 2.558\ 10^{-3}$ sont automatiquement traitées avec des éléments P^1 et le schéma saute-moutons. Ces mailles sont représentées en bleu sur la figure 5.24. Le résultat

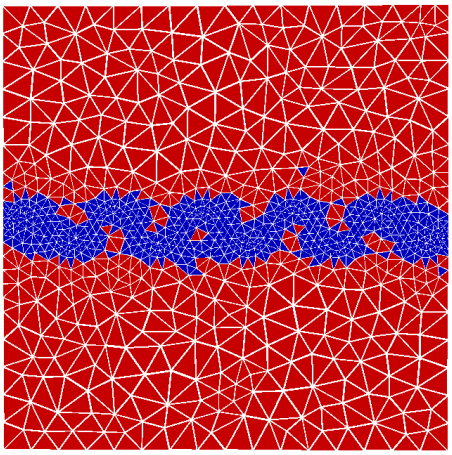

FIGURE 5.24 – Géométrie du rétrécissement

obtenu est présenté sur la figure 5.25 aux instants $t = 1.064\ 10^{-2}$s, $1.127\ 10^{-2}$s, $1.315\ 10^{-2}$s, et $1.628\ 10^{-2}$s.

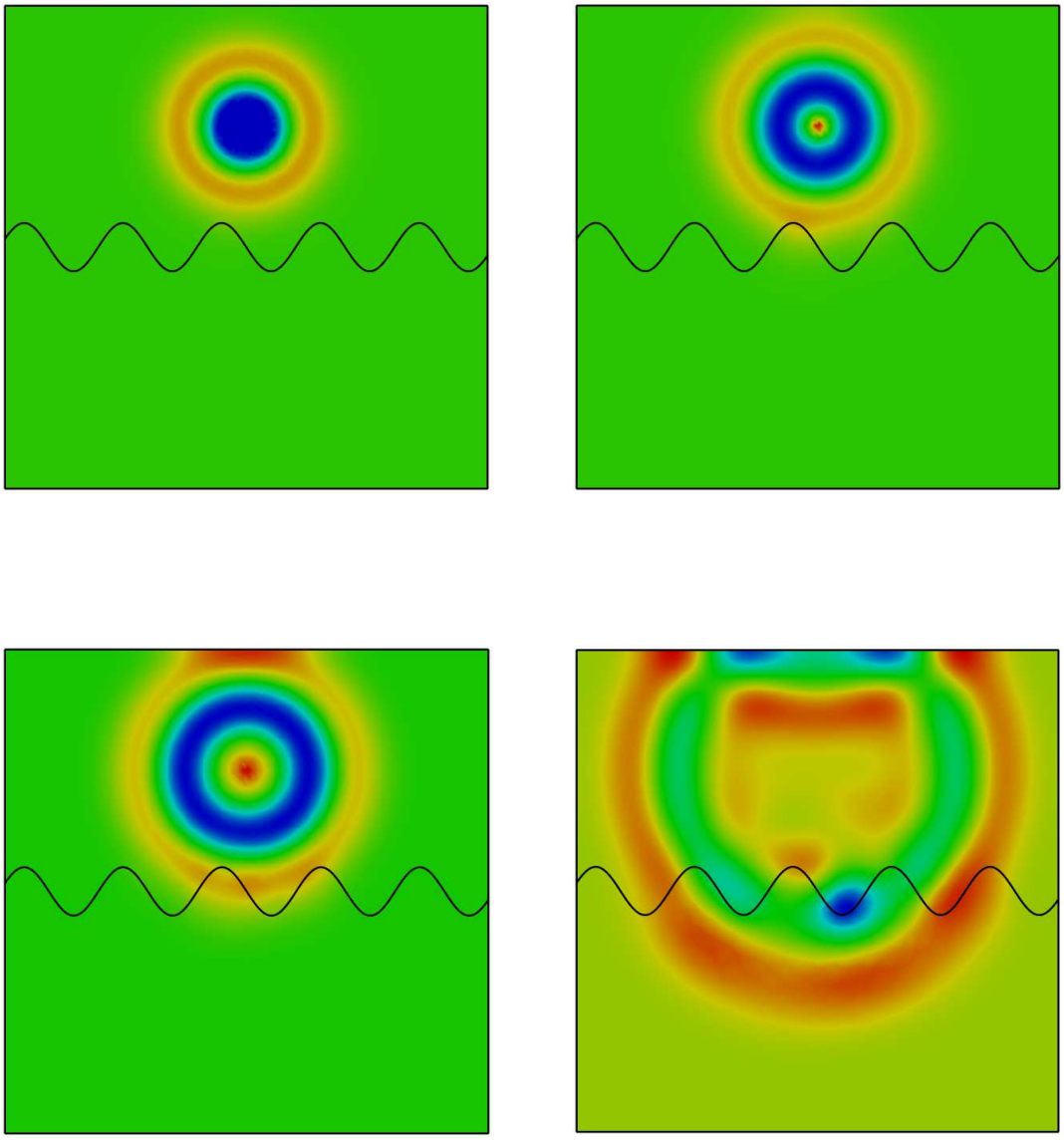

FIGURE 5.25 – Simulation sur le domaine avec interface sinusoidale

5.3.3 Cas d'un rétrécissement

Dans cette sous-section, nous allons nous intéresser au domaine présenté à la figure 5.26.

Ce domaine est formé de deux rectangles $\Omega_1 = [0, 10] \times [7, 10]$ et $\Omega_2 = [0, 10] \times [0, 3]$ reliés l'un à l'autre par un autre domaine $\Omega_3 = [4.9, 5.1] \times [3, 7]$ beaucoup plus étroit. Dans cette expérience, nous considérons une vitesse constante égale à 1. Le rapport entre mailles fines et grossières étant de l'ordre de cinq, il est clair que nous n'avons pas besoin de mailler aussi finement les domaines Ω_1 et Ω_2 que Ω_3 (cf. Fig. 5.27 et 5.28).

De plus, pour atteindre un même niveau d'approximation, nous ne devons pas utiliser les mêmes degrés polynomiaux partout. C'est pourquoi nous utiliserons des éléments P^3 dans les domaines Ω_1 et Ω_2 maillés grossièrement et des éléments P^1 dans le sous-domaine Ω_3 maillé finement. Avec une telle expérience, l'intérêt d'avoir une méthode permettant de considérer facilement des éléments d'ordres différents, d'un élément à l'autre, est indéniable. De plus, en ce qui concerne l'ordre temporel, on considérera le Δ^2-schéma dans les zones Ω_1 et Ω_2 alors que dans la zone

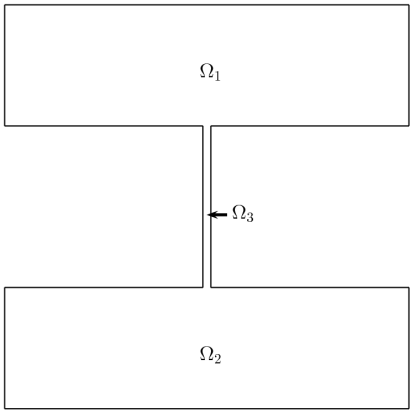

FIGURE 5.26 – Géométrie du rétrécissement

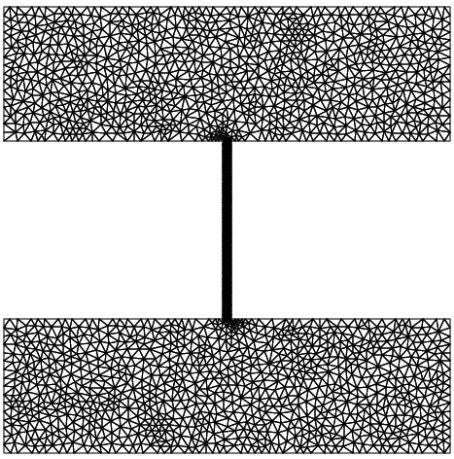

FIGURE 5.27 – Maillage du domaine avec rétrécissement

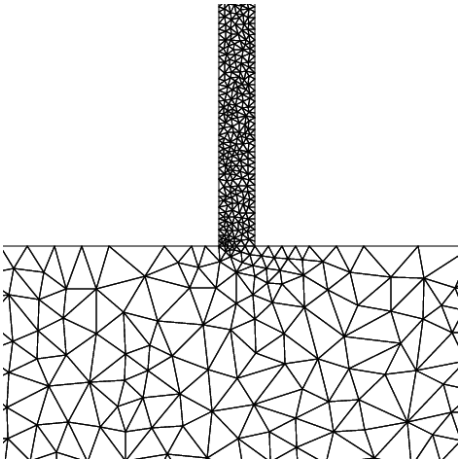

FIGURE 5.28 – Zoom sur le maillage du rétrécissement

Ω_3, nous considérons le schéma saute-moutons. On placera la source définie en (5.3) au point $x_0 = (5, 8)$. En considérant une telle configuration, on obtient les résultats présentés sur la figure 5.29 aux instants $t = 1.97\ 10^{-2}s$, $5.25\ 10^{-2}s$, $6.75\ 10^{-2}s$, et $9.84\ 10^{-2}s$.

Comme à la section précédente, de nombreuses réflections apparaissent aux bords du domaine. Il serait donc intéressant d'utiliser des conditions aux limites absorbantes sur les bords supérieurs et inférieurs du domaine. Néanmoins, il n'est pas évident de développer des conditions aux limites absorbantes pour des schémas d'ordre quatre en temps donc en particulier pour les Δ^p-schémas. Une alternative intéressante serait de considérer des éléments de bas degré sur les bords du domaine pour pouvoir utiliser une condition absorbante classique. Comme à la sous-section 5.3.1, le but est ici de considérer cette zone la plus fine possible *i.e.* de la taille d'une maille grossière.
Sur le domaine que nous avons considéré, cela revient à considérer le maillage représenté sur la figure 5.30. La figure 5.31 est un zoom de la partie supérieure de ce maillage. Nous avons imposé dans ces deux zones fines le schéma saute-moutons avec des éléments P^1 et nous avons utilisé sur les bords supérieur et inférieur du domaine une condition aux limites absorbantes classique

$$\partial_t u + \partial_n u = 0.$$

Les résultats numériques sont reportés sur la figure 5.32.

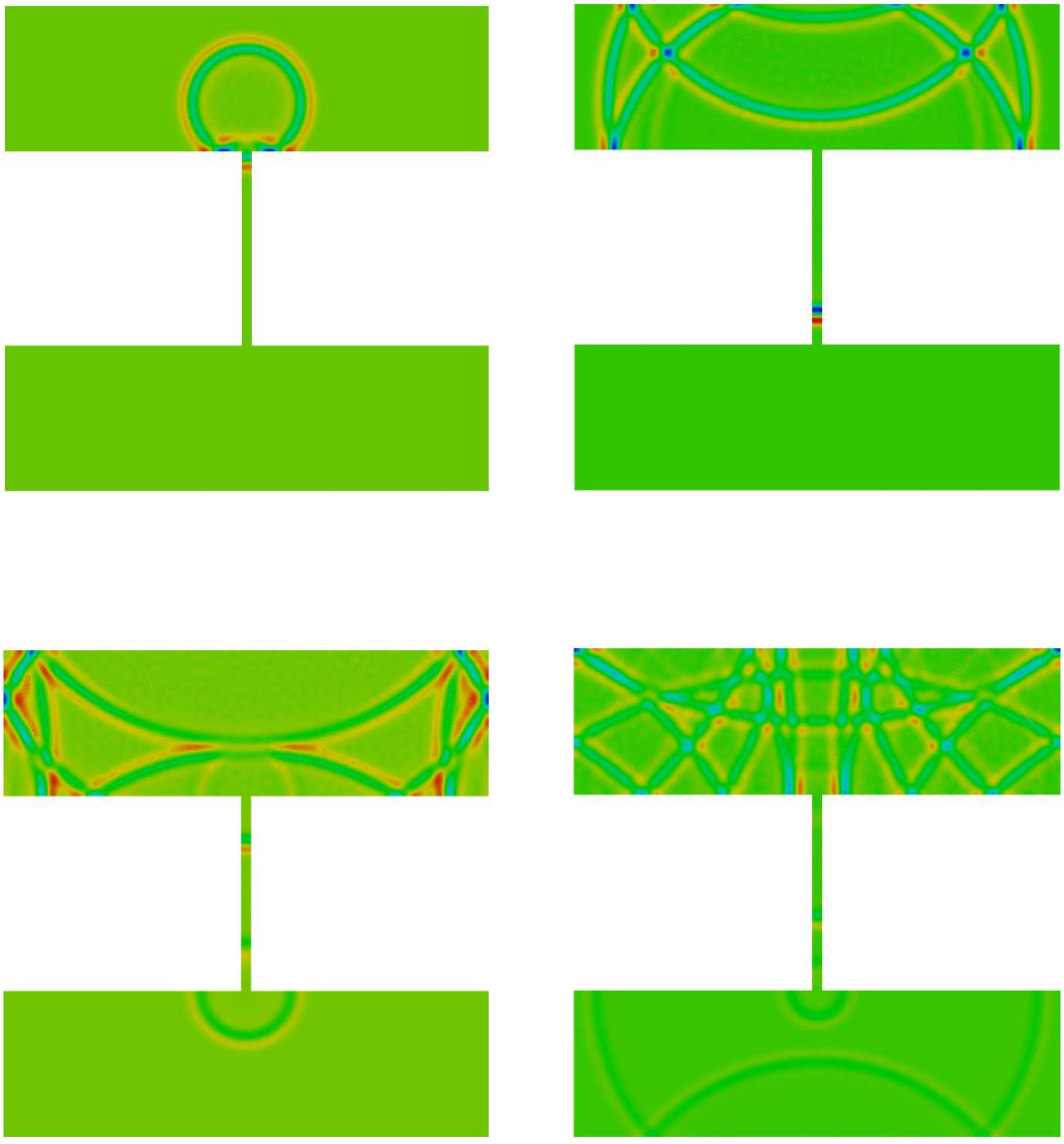

FIGURE 5.29 – Simulation sur le domaine avec rétrécissement

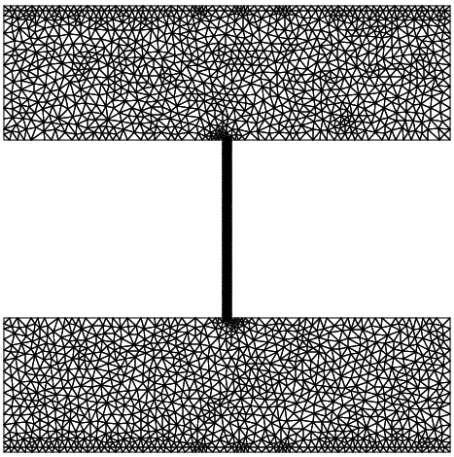

FIGURE 5.30 – Maillage du domaine avec rétrécissement et zones fines.

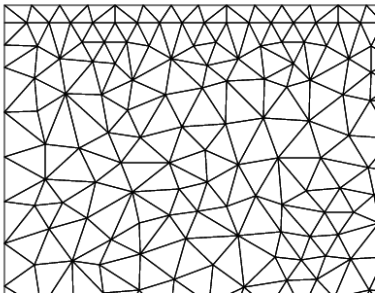

FIGURE 5.31 – Zoom de la couche fine dans la parite supérieure.

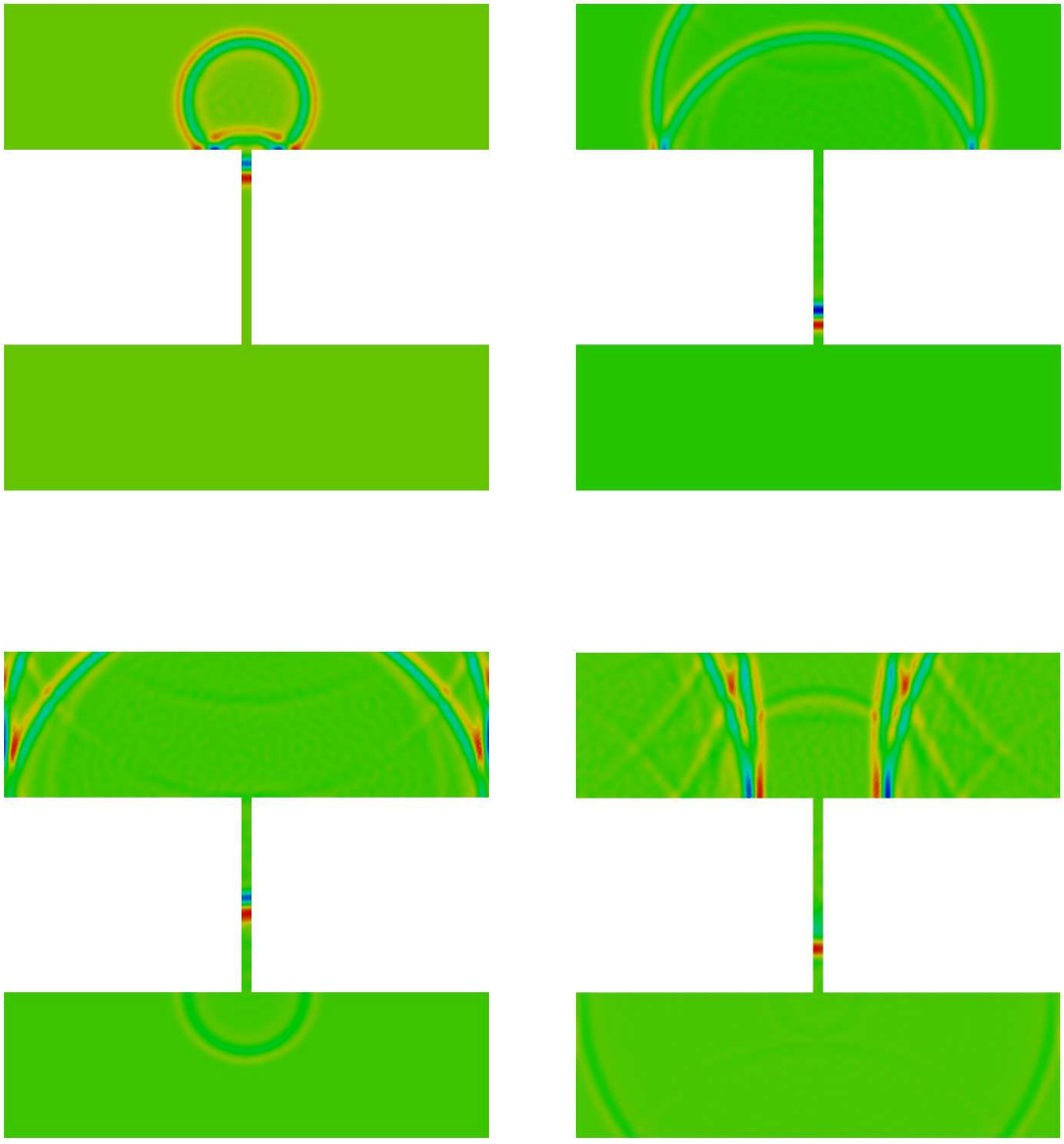

FIGURE 5.32 – Simulation sur le domaine avec rétrécissement et conditions aux limites absorbantes

Le résultat de la simulation illustre que l'onde est absorbée aux bords supérieur et inférieur du domaine. Néanmoins, afin de nous assurer du bon comportement de la condition aux limites absorbantes, nous avons étudié le comportement de l'énergie associée au schéma. On rappelle que pour le Δ^2-schéma, cette énergie est définie comme suit (cf. chapitre 1)

$$E^{n+\frac{1}{2}} = \left(M \frac{U^{n+1} - U^n}{\Delta t}, \frac{U^{n+1} - U^n}{\Delta t} \right) + \left(K^* U^n, U^{n+1} \right)$$

où $K^* = K_1 - \dfrac{\Delta t^2}{12} K_2$.
On a représenté sur la figure 5.33 cette énergie calculée en fonction du temps d'expérience. On constate que l'énergie décroît dès que la source a fini d'émettre, ce qui est caractéristique du bon fonctionnement de la condition aux limites absorbantes. En effet, en appliquant une telle condition, le schéma devient dissipatif et l'énergie doit donc décroitre. Elle ne tend pas vers 0 car nous n'imposons pas des conditions absorbantes sur tous les bords du domaine. De plus, une partie de l'énergie est piégée dans le sous-domaine Ω_3.

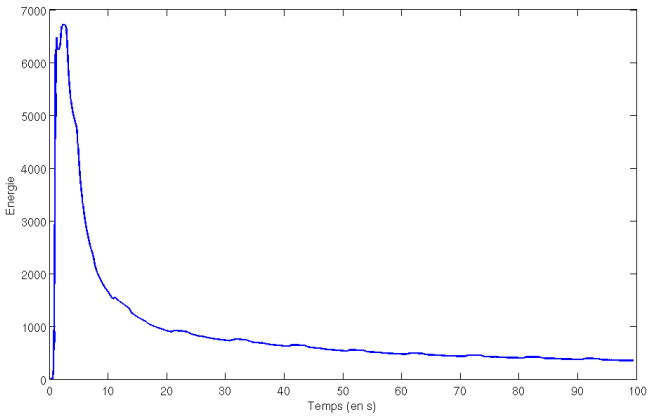

FIGURE 5.33 – Décroissance d'une énergie

5.4 Conclusion

Dans ce chapitre, nous avons vu que les schémas que nous avons proposés présentent des propriétés tout à fait intéressantes vis-à-vis de l'adaptativité en temps et en espace. En effet, on a observé qu'il suffit d'adapter l'ordre des fonctions de base pour adapter l'ordre en temps de la méthode numérique. De plus, on a vérifié que l'on pouvait contourner certaines difficultés pour prendre en compte des conditions aux limites absorbantes aux bords des domaines de calcul. Néanmoins, cela n'est qu'une étape puisqu'il serait très intéressant de développer des conditions aux limites absorbantes spécifiques à nos schémas et ainsi éviter de devoir utiliser au bord du domaine une zone où l'on ne considère que le schéma d'ordre deux en temps. De plus, nous n'avons considéré ici que la p adaptativité (*i.e.* l'adaptativité en ordre) en temps et il faudrait maintenant considérer la Δt adaptativité, c'est-à-dire des techniques de pas de temps local. Ces techniques (cf. par exemple [25, 32, 18]) reposent sur l'utilisation de différents pas de temps en divers endroits du domaine. En effet, la condition CFL est contrainte par la plus petite maille du domaine et il est donc intéressant d'utiliser des pas de temps différents pour pouvoir effectuer moins de calculs dans des zones où les mailles sont beaucoup plus grosses.

Bibliographie

[1] C. Agut and J. Diaz. High-order schemes combining the modified equation approach and discontinuous galerkin approximations for the wave equation. *Communication In Computationnal Physics*, 2011.

[2] M. Ainsworth, P. Monk, and W. Muniz. Dispersive and dissipative properties of discontinuous galerkin finite element methods for the second-order wave equation. *Journal of Scientific Computing*, 27, 2006.

[3] L. Anné, P. Joly, and Q. H. Tran. Construction and analysis of higher order finite difference schemes for the 1d wave equation. *Computational Geosciences*, 4 :207–249, 2000.

[4] D. N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAM J. Numer. Anal.*, 19, No. 4 :742–760, 1982.

[5] D. N. Arnold, F. Brezzi, B. Cockburn, and L.D. Marini. Unified analysis of disconitnuous galerkin methods for elliptic problems. *SIAM J. Numer. Anal.*, 39 :1749–1779, 2002.

[6] C. Baldassari. Modélisation et simulation numérique pour la migration terrestre par équation d'ondes. *PhD Thesis*, 2009.

[7] C. Baldassari, H. Barucq, H. Calandra, B. Denel, and J. Diaz. High-order discontinuous galerkin method for the reverse time migration. *Communications in Computational Physics*, 2010.

[8] F. Bassi and S. Rebay. A high-order accurate discontinuous finite element method for the numerical solution of the compressible navier-stokes equations. *J. Comput. Phys.*, 131 :267–279, 1997.

[9] G. Benitez Alvarez, A.F. Dourado Loula, E.G. Dutra do Carmo, and A. Alves Rochinha. A discontinuous finite element formulation for helmholtz equation. *Comput. Methods. Appl. Mech. Engrg.*, 195 :4018–4035, 2006.

[10] E. Bossy. Caractérisation ultrasonore de l'os par propagation de l'onde latérale : modèle de propagation, transfert de technologie et application à l'ostéoporose. *PhD Thesis*, 2003.

[11] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer, Berlin, 2002.

[12] H. Brezis. Analyse fonctionnelle, théorie et applications. *Dunod*, 1999.

[13] L. Cagniard. Réflexion et réfraction des ondes sismiques progressives. *Gauthier-Villard*, 1939.

[14] L. Cagniard. Reflection and refraction of progressive seismic waves. *McGraw-Hill*, 1962, traduit de [13].

[15] P. G. Ciarlet. The finite element method for elliptic problems. *North-Holland Publishing Co., Amsterdam, Studies in Mathematics and its Applications*, 4, 1978.

[16] F. Clearbout, J. Toward a unified theory of reflector imaging. *Geophysics*, 36 :467–481, 1971.

[17] G. Cohen, P. Joly, and N. Tordjman. Construction and analysis of higher order finite elements with mass lumping for the wave equation. *Second International Conference on Mathematical and Numerical Aspects of Wave Propagation (Newark, DE)*, SIAM, Philadelphia, PA :152–160, 1993.

[18] Gary Cohen, Xavier Ferrieres, and Sébastien Pernet. A spatial high-order hexahedral discontinuous galerkin method to solve maxwell's equations in time domain. *J. Comput. Phys.*, 217 :340–363, 2006.

[19] S. Cohen and P. Joly. Construction analysis of fourth-order finite difference schemes for the acoustic wave equation in nonhomogeneous media. *SIAM J. Numer. Anal.*, 33 :1266–1302, 1996.

[20] S. Cohen, P. Joly, J.E. Roberts, and N. Tordjman. Higher-order triangular finite elements with mass-lumping for the wave equation. *SIAM J. Numer. Anal.*, Vol. 44, 6 :2408–2431, 2006.

[21] S. Cohen, P. Joly, and N. Tordjman. Higher-order finite elements with mass-lumping for the 1d wave equation. *Finite Elements in Analysis and Design*, Vol. 16, Issues 3-4 :329–336, 1994.

[22] M. A. Dablain. The application of high order differencing for the scalar wave equation. *Geophysics*, pages 51 :54–56, 1, 1986.

[23] J. D. De Basabe and M. K. Sen. New developments in the finite-element method for seismic modeling. *Leading Edge*, 28 :562–567, 2009.

[24] A. T. De Hoop. The surface line source problem. *Appl. Sci. Res.*, B 8 :349–356, 1959.

[25] J. Diaz and M. Grote. Energy conserving explicit local time-stepping for second-order wave equations. *SIAM Journal on Scientific Computing*, 31 (3) :1985–2014, 2009.

[26] L. C. Evans. *Partial Differential Equations,Graduate Studies in Mathematics*, volume 19. American Mathematical Society, 1998.

[27] S. Fauqueux. Eléments finis mixtes spectraux et couches absorbantes parfaitement adaptées pour la propagation d'ondes élastiques en régime transitoire. *PhD Thesis*, 2003.

[28] X. Feng and H. Wu. *hp*-discontinuous galerkin methods for the helmholtz equation with large wave number. *SIAM J. Numer. Anal.*, 47, No. 4, 2009.

[29] E.H. Georgoulis and P. Houston. Discontinuous galerkin methods for the biharmonic problem. *IMA Journal of Numerical Analysis*, 2008.

[30] J.-C. Gilbert and P. Joly. Higher order time stepping for second order hyperbolic problems and optimal cfl conditions. *Numerical Analysis and Scientific Computing for PDE's and their Challenging Applications*, 2006.

[31] J.L. Goncalves, P.R.B. Devloo, and S.M. Gomes. Goal-oriented error estimation for the discontinuous galerkin method applied to the biharmonic equation. *Numerical Mathematics and Advanced Applications 2009*, 2 :369–376, 2010.

[32] M. Grote and T. Mitkova. Explicit local time stepping methods for maxwell's equations. *J. Comp. Appl. Math.*, 234 :3283–3302, 2010.

[33] M. J. Grote, A. Schneebeli, and D. Schötzau. Discontinuous galerkin finite element method for the wave equation. *SIAM J. on Numerical Analysis*, 44 :2408–2431, 2006.

[34] M. J. Grote and D. Schötzau. Convergence analysis of a fully discrete dicontinuous galerkin method for the wave equation. *Preprint No. 2008-04*, 2008.

[35] B. Gustafsson and E. Mossberg. Time compact high-order difference methods for wave propagation. *SIAM J. Sci. Comput.*, 26 (1) :259–271, 2004.

[36] B. Gustafsson and P. Wahlund. Time compact high-order difference methods for wave propagation, 2d. *Journal of Scientific Computing*, 25 (1/2), 2005.

[37] T.J.R. Hughes. The finite element method. *Prentice Hall Inc., Englewood Cliffs, NJ. Linear static and dynamic finite element analysis, With the collaboration of Robert M. Ferencz and Arthur M. Raefsky*, 1987.

[38] P. Joly and J. Rodríguez. Optimized higher order time discretization of second order hyperbolic problems : Construction and numerical study. *J. of Computational and Applied Mathematics*, 234 :p. 1953–1961, 2010.

[39] M. Käser and A. Iske. ADER Schemes on Adaptive Triangular Meshes for Scalar Conservation Laws. *Journal of Computational Physics*, 205 :486–508, 2005.

[40] D. Komatitsch and J. Tromp. Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophys, J. Int.*, 139 :806–822, 1999.

[41] D. Komatitsch and J.P. Vilotte. The spectral-element method : an efficient tool to simulate the seismic response of 2d and 3d geological structures. *Bulletin of the Seismological Society of America*, 88(2) :368–392, 1998.

[42] D. Komatitsch, J.P. Vilotte, R. Vai, J.M. Castillo-Covarrubias, and Sanchez-Sesma F.J. The spectral element method for elastic wave equations : application to 2d and 3d seismic problems. *International Journal for numerical methods in engineering*, 45 :1139–1164, 1999.

[43] I. Mozolevski and E. Suli. *hp*-version interior penalty dgfems for the biharmonic equation. Technical report, Oxford University Computing Laboratory, 2004.

[44] A. Patera. A spectral element method for fluid dynamics : laminar flow in a channel expansion. *J. Comput. Phys.*, 54 :468–488, 1984.

[45] S. Prudhomme, f. Pascal, J. T. Oden, and Romkes A. Review of a priori error estimation for discontinuous galerkin methods. *TICAM REPORT 00-27, Texas Institute for Computational and Applied Mathematics, The University of Texas at Austin*, 2000.

[46] W.H. Reed and T.R. Hill. Triangular mesh methods for the neutron transport equation. *Tech. Report LA-UR-73-479, Los Alamos Scientific Laboratory, Los Alamos, NM*, 1973.

[47] E. Robein. Seismic imaging, a review of the techniques, their principles, merits and limitations. *EAGE, Education Tour Series*, 2010.

[48] C. Schwab. *p*- and *hp*- finite element methods. theory and applications to solid and fluid mechanics. *Oxford University Press, Oxford*, 1998.

[49] K. Shahbazi. An explicit expression for the penalty parameter of the interior penalty method. *J. of Computational Physics*, 205 :401–407, 2005.

[50] G. R. Shubin and J. B. Bell. A modified equation approach to constructing fourth-order methods for acoustic wave propagation. *SIAM J. Sci. Statist. Comput.*, 8 :135–151, 1987.

[51] E. Süli and I. Mozolevski. *hp*-version interior penalty dgfems for the biharmonic equation. *Computer Methods in Applied Mechanics and Engineering*, Vol. 196, No. 13-16 :p. 1851–1863, 2007.

[52] E. Toro, M. Käser, M. Dumbser, and C. Castro. ADER Shock-Capturing Methods and Geophysical Applications. In *Proceedings of the 25th International Symposium on Shock Waves - ISSW25*, Bangalore, 2005.

[53] T. Warburton and J. S. Hesthaven. On the constants in *hp*-finite element trace inverse inequalities. *Comput. Methods Appl. Mech. Engrg.*, 192 :2765–2773, 2003.

[54] R.F. Warming and B.J. Hyett. The modified equation approachto the stability and accuracy analysis of finite difference methods. *Journal of Computational Physics*, 14 :159–179, 1974.