Méthodes d'éléments finis d'ordre élevé pour la simulation numérique de la propagation d'ondes

Sébastien Jund

INSTITUT DE RECHERCHE MATHÉMATHIQUE AVANCÉE
Université Louis Pasteur et CNRS (UMR 7501)
7, rue René Descartes
67084 Strasbourg Cedex

Méthodes d'éléments finis d'ordre élevé pour la simulation numérique de la propagation d'ondes.

par

Sébastien JUND

Thèse soutenue le 28 novembre 2007 devant le jury composé de

Patrick CIARLET	Examinateur
Gary COHEN	Rapporteur externe
Philippe HELLUY	Rapporteur interne
Serge PIPERNO	Rapporteur externe
Stéphanie SALMON	Examinateur
Jacques SEGRÉ	Invité
Éric SONNENDRÜCKER	Directeur de thèse

Remerciements

Je tiens à remercier en tout premier lieu mon directeur de thèse Eric Sonnendrücker, pour m'avoir encouragé à me lancer dans cette thèse alors que j'éprouvais moi-même les plus grands doutes quant à mes capacités à mener à bien de tels travaux de recherche. Pour sa disponibilité, ses conseils éclairés et sa gentillesse je lui suis extrêmement reconnaissant.

Merci également à Stéphanie Salmon qui a encadré l'ensemble de mes travaux. Depuis les premiers codes d'éléments finis qu'elle m'a fournis, jusqu'a la relecture de ce mémoire, elle n'a jamais été avare de son temps et je profite de ces quelques lignes pour lui signifier ma gratitude.

Je remercie mes rapporteurs d'avoir accepté cette tâche. Un merci tout particulier à Gary Cohen d'avoir accepté de rapporter sur des travaux qui sont fortement inspirés des siens et de ses collaborateurs, à Serge Piperno pour ses remarques et conseils ayant permis l'amélioration du présent document et à Philippe Helluy pour son enthousiasme relatif à mes travaux et sa sympathie de manière plus générale. Merci également aux autres membres du jury, Patrick Ciarlet et Jacques Segré pour l'intérêt qu'ils ont portés sur ces travaux.

Un très grand merci à Claus-Dieter Munz, Michael Dumbser et Jens Utzmann de l'"Institut für Aerodynamik und Gasdynamik" de Stuttgart pour leur collaboration dans le cadre de la comparaison éléments finis conforme - Galerkin discontinus. Merci pour leur accueil plus que chaleureux pendant mon séjour à Stuttgart dont je garde un très bon souvenir.

Merci à l'agréable Hyam Abboud et au sympathique Hamdi Zorgati pour leur collaboration, durant le CEMRACS'05, aux travaux portant sur la méthode de résolution deux échelles. Ils m'ont permis de reprendre goût à la recherche au moment où j'étais sûrement le plus démotivé.

Merci à mes collègues de bureau, Isabelle Metzmeyer-David dont la gentillesse, la bonne humeur et l'enthousiasme qu'elle peut exprimer en particulier par rapport à l'enseignement sont des raisons plus que suffisantes pour mériter mon admiration et mes remerciements, et Alexandre Mouton que je remercie tout particulièrement pour avoir fait plus que sa part de travail dans le cadre du cours de T.A.N., me déchargeant ainsi d'une part du mien, et de m'avoir ainsi permis de m'adonner plus librement à la rédaction de ce mémoire.

Merci à toute l'équipe EDP de l'IRMA dont les séminaires et les discussions avec les personnes la composant m'ont permis d'élargir ma vision des mathématiques et merci à l'IRMA lui-même pour les conditions de travail plus que favorables dont il nous fait bénéficier. Merci notamment au personnel administratif et technique pour leur efficacité.

Mes derniers remerciements iront à ma famille et mes amis, que je remercie simplement pour tout.

Notations

Sauf mention contraire les notations suivantes valent pour l'ensemble de ce document. Soit Ω un ouvert régulier de $\mathbb{R}^2$ de frontière $\partial\Omega = \Gamma$, $\overrightarrow{n}$ le vecteur unitaire normal sortant de Ω sur Γ et $\overrightarrow{\tau}$ le vecteur unitaire tangent à Ω sur Γ faisant de $(\overrightarrow{n}, \overrightarrow{\tau})$ une base directe du plan. Notons $(x, y) \in \mathbb{R}^2$ un point du plan.
Les opérateurs différentiels sont définis par

- pour $u : \mathbb{R}^2 \to \mathbb{R}$, $\triangle u = \dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2}$,
- pour $u : \mathbb{R}^2 \to \mathbb{R}$, $\overrightarrow{\nabla} u = \begin{pmatrix} \dfrac{\partial u}{\partial x} \\ \dfrac{\partial u}{\partial y} \end{pmatrix}$,
- pour $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} : \mathbb{R}^2 \to \mathbb{R}^2$, $\nabla \times \overrightarrow{u} = \dfrac{\partial u_2}{\partial x} - \dfrac{\partial u_1}{\partial y}$,
- pour $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} : \mathbb{R}^2 \to \mathbb{R}^2$, $\nabla \cdot \overrightarrow{u} = \dfrac{\partial u_1}{\partial x} + \dfrac{\partial u_2}{\partial y}$,
- pour $u : \mathbb{R}^2 \to \mathbb{R}$, $\overrightarrow{\nabla} \times u = \begin{pmatrix} \dfrac{\partial u}{\partial y} \\ -\dfrac{\partial u}{\partial x} \end{pmatrix}$.

Les espaces fonctionnels (voir [2] ou [8]) sont définis par

- $L^2(\Omega) = \left\{ u : \Omega \to \mathbb{R} \text{ mesurable sur } \Omega \mid \displaystyle\int_\Omega u^2 \, dxdy < \infty \right\}$,
- $H^1(\Omega) = \left\{ u \in L^2(\Omega) \mid \nabla u \in (L^2(\Omega))^2 \right\}$,
- $H(rot, \Omega) = \left\{ \overrightarrow{u} \in (L^2(\Omega))^2 \mid \nabla \times \overrightarrow{u} \in L^2(\Omega) \right\}$,
- $H(div, \Omega) = \left\{ \overrightarrow{u} \in (L^2(\Omega))^2 \mid \nabla \cdot \overrightarrow{u} \in L^2(\Omega) \right\}$.

Les espaces $H^1(\Omega)$, $H(rot, \Omega)$ et $H(div, \Omega)$ pouvant être munis des applications respectives trace, trace tangentielle et trace normale (voir [53]) nous définissons

- $H_0^1(\Omega) = \left\{ u \in H^1(\Omega) \mid u_{|\Gamma} = 0 \right\}$,
- $H_0(rot, \Omega) = \left\{ \overrightarrow{u} \in H(rot, \Omega) \mid \overrightarrow{u} \cdot \overrightarrow{\tau}_{|\Gamma} = 0 \right\}$,
- $H_0(div, \Omega) = \left\{ \overrightarrow{u} \in H(div, \Omega) \mid \overrightarrow{u} \cdot \overrightarrow{n}_{|\Gamma} = 0 \right\}$.

Table des matières

Introduction

La modélisation des phénomènes de propagation d'ondes est un problème que l'on retrouve dans de nombreuses applications physiques telles que la propagation d'ondes acoustiques ou électromagnétiques. Malgré plusieurs décennies de recherches actives, la simulation numérique de ces phénomènes reste un problème délicat.

La méthode des différences finies fait partie des méthodes qui se sont imposées de par sa robustesse et sa facilité d'implémentation. Les études de ce type de schémas (voir par exemple [3] ou [31] dans le cadre géophysique, [50] dans le cadre électromagnétique, ou encore [66] dans le cadre de l'acoustique) ont montré l'intérêt non négligeable de l'utilisation de méthodes d'ordre élevé, celles-ci ayant de bien meilleures propriétés de dissipation et de dispersion, ce qui permet une propagation d'onde plus précise à moindre coût, tant du point de vue du temps de calcul que du stockage de données.

Ces méthodes ne sont toutefois bien adaptées qu'à des domaines à géométrie simple (rectangulaires par exemple), et ne s'adaptent que très difficilement à des domaines à géométrie complexe. Pour ce type de domaine, la méthode des éléments finis semble naturellement plus adaptée et plus maniable. Parmi les avantages de la méthode des éléments finis qui nous ont poussé à nous intéresser particulièrement à celle-ci, outre sa souplesse d'utilisation dans le cadre d'un domaine de calcul à géométrie complexe via l'utilisation d'une discrétisation de ce domaine par un maillage non-nécessairement structuré, il faut souligner que son utilisation nous permet naturellement une discrétisation conforme des équations entrant en jeu, c'est-à-dire une résolution des équations dans des espaces de discrétisation inclus dans les espaces continus dans lesquels "vivent" leur solution. Cette notion de conformité est d'une importance toute particulière dans le cadre de la propagation d'ondes électromagnétiques décrite par les équations de Maxwell. En effet, une discrétisation conforme de ces équations nous permet de vérifier automatiquement l'équation de conservation de la charge, et par conséquent la troisième équation constituant les équations de Maxwell dite loi de Gauss. Nous pourrons donc résoudre les équations de Maxwell en ne résolvant que les deux premières lois constituant ces équations, dites lois d'Ampère et de Faraday. Les travaux que l'on a menés sur la résolution des équations de Maxwell dans le cadre de cette thèse sont aussi à intégrer dans le cadre du projet financé par l'Agence National de la Recherche High Order Finite Element Particle-In-Cell Solvers on Unstructured Grids "HOUPIC" (ANR-06-CIS6-013-01). Ce projet vise à développer des méthodes numériques pour la simulation de phénomènes issus de la physique des accélérateurs et des plasmas incluant les plasmas de fusion modélisés par les équations de Vlasov-Maxwell. Dans le

cadre d'une résolution numérique des équations de Vlasov couplées avec les équations de Maxwell, les densités de charge et de courant n'étant connues que numériquement à partir des positions et vitesses des particules chargées, l'équation de conservation de la charge (faisant intervenir ces densités) n'est en général pas automatiquement vérifiée, ce qui peut mener à des solutions non physiques des équations de Vlasov-Maxwell (voir [5]). L'utilisation d'éléments finis conformes pour la résolution des équations de Maxwell apparaît donc tout naturellement dans ce contexte comme une alternative intéressante.

La méthode des éléments finis est à mettre en opposition, par rapport à cette notion de conformité, aux méthodes de type Galerkin discontinus, introduites par W. H. Reed [62] dans le cadre du transport de neutrons en 1973 et qui connaissent un regain d'intérêt depuis les travaux de B. Cockburn et al. [22]-[21] du début des années 90. Parmi les travaux sur les éléments finis de type Galerkin discontinus citons ceux de G. Cohen et al. [24], J. S. Hesthaven et T. Warburton [49], et notamment ceux de C.-D. Munz et M. Dumbser qui ont adapté l'approche ADER de E. F. Toro et V. A. Titarev [69] à la résolution des équations d'Euler (linéarisées ou non) pour obtenir des schémas d'ordre arbitrairement élevé (qu'ils ont testé jusqu'à l'ordre 10) en espace et en temps sur des maillages triangulaires non structurés [35]. C'est dans cet état d'esprit qu'ont été menés les travaux dont les résultats sont consignés dans ce manuscrit : le but de ces travaux est la construction de schémas numériques pour la résolution de phénomènes de propagation d'onde basés sur des discrétisations en espace par éléments finis conformes, ces schémas ayant pour vocation à être d'ordre arbitrairement élevé et aussi efficaces que possible.

Si l'on parle d'efficacité des schémas à discrétisation en espace par éléments finis, c'est qu'il faut être conscient que la méthode des éléments finis à un désavantage majeur : celui de nécessiter l'inversion d'une matrice, dite matrice de masse, à chaque pas de temps (voire plusieurs fois par pas de temps pour les discrétisations en temps d'ordre élevé). L'idée qui vient alors naturellement pour rendre la méthode des éléments finis plus attractive, est donc de rendre cette matrice de masse diagonale. On parle alors de condensation de la matrice de masse (ou, sans ambiguïté possible, de condensation de l'élément fini). La condensation de la matrice de masse est un problème qui n'est résolu que dans un certain nombre de cas particuliers. Considérant les éléments finis adaptés à la résolution de l'équation des ondes scalaire, le cas des éléments finis de Lagrange en une dimension d'espace est un cas que l'on peut résoudre à l'aide de la théorie des polynômes orthogonaux. En effet, la condensation de la matrice de masse issue de l'utilisation d'éléments finis de Lagrange passe par la détermination de formules de quadrature ayant des propriétés bien précises, et il se trouve que les formules de quadrature de Gauss-Lobatto ont toutes les propriétés nécessaires. Le cas des éléments finis en dimensions d'espace supérieures construits par produit tensoriel de ces éléments finis 1D (c'est-à-dire les éléments finis quadrilatéraux en 2D ou hexaédraux en 3D) est alors naturellement réglé. L'utilisation de ces éléments finis n'est toutefois pas adaptée à toute géométrie de domaine : l'efficacité des méthodes d'éléments finis est liée à la "qualité" du maillage, or la génération de "bon" maillages quadrangulaires (en deux dimensions d'espace), c'est-à-dire de maillage dont les éléments sont peu déformés, est très délicate, beaucoup plus en tout cas que la génération de maillages triangulaires. Le cas de la condensation de la matrice de masse issue des éléments finis de Lagrange triangulaires

n'a été résolu que pour les sept premiers ordres (voir les travaux de G. Cohen et al. [26][70] pour la condensation de la matrice de masse issue des éléments finis P_1 à P_3, de W. A. Mulder et al. [14][56] pour la condensation de la matrice de masse issue des éléments finis P_4 et P_5, et plus récemment de F. X. Giraldo et M. A. Taylor [42] pour la condensation de la matrice de masse issue des éléments finis P_6 et P_7). Parmi la classe des éléments finis d'arête que nous allons considérer, seule la condensation de la matrice de masse issue des éléments finis d'arête triangulaires de plus bas degré [51] et des éléments finis d'arête rectangulaires sur maillages cartésiens [23][37] ont été traités. Notons qu'il existe une autre classe d'éléments finis d'arête pour lesquels il est possible de condenser la matrice de masse (plus précisément de remplacer la matrice de masse par une matrice diagonale par bloc, ce qui est numériquement aussi efficace que la condensation décrite plus haut), mais que ceux-ci ne sont pas adaptés à la résolution des équations de Vlasov-Maxwell, dans la mesure où leur utilisation fait apparaître des ondes parasites [36].

Aux problèmes liés à la discrétisation en espace viennent s'ajouter ceux liés à la discrétisation en temps, l'ordre d'un schéma dépendant non seulement de l'ordre de sa discrétisation en espace mais aussi de celui de sa discrétisation en temps. D'un point de vue pratique il est assez difficile de trouver dans la littérature des discrétisations en temps d'ordre élevé (c'est-à-dire au-delà de l'ordre 4). Cela est essentiellement dû au fait que la détermination des paramètres de discrétisation en temps efficaces d'ordre élevé, devient très complexe. Parmi les méthodes très populaires citons notamment les méthodes de Runge-Kutta, dont la classe des méthodes diagonalement implicite [9][43][44] nous intéressera particulièrement et les discrétisation en temps symplectiques [73]. Pour rester dans cet esprit de construction de schémas d'ordre arbitrairement élevé, nous avons développé nos propres discrétisations en temps, dont la montée en ordre, se fait de manière itérative.

Le dernier problème auquel nous allons être confrontés est le suivant : dans un certain nombre d'applications, la simulation d'une propagation d'ondes nécessite une résolution plus précise des équations régissant cette propagation dans certaines régions du domaine de calcul. C'est dans cette optique qu'ont été introduites les méthodes de décomposition de domaine. L'idée de ces méthodes est de décomposer le domaine de calcul en plusieurs sous-domaines, avec ou sans recouvrement suivant les méthodes, et de définir des espaces de discrétisation propres à chacun des sous-domaines. Ainsi il est possible d'ajuster la finesse de l'espace de discrétisation, a priori, en fonction des données du problème (par rapport à la régularité du terme source imposé générant une onde par exemple). Notons que lorsque l'on parle de finesse de l'espace de discrétisation il faut notamment entrevoir la possibilité d'un raffinement local du maillage mais aussi la possibilité d'utiliser localement des espaces d'éléments finis d'ordre plus élevé. On voit alors apparaître la délicate question de la conformité de la méthode au niveau des frontières communes à plusieurs sous-domaines, question qui a motivé le développement de la plupart des méthodes de décompositions de domaine. Parmi les méthodes classiques citons notamment la méthode itérative de Schwarz [54], et plus particulièrement les méthodes dites directes, utilisant des multiplicateurs de Lagrange dont différentes variantes sont proposées par exemple par P. Le Tallec et T. Sassi [68] ou par C. Bernardi, Y. Maday et A. T. Patera [7]. Dans le cadre de la propagation d'ondes électromagnétiques, des méthodes de décomposition de domaine

ont été développées notamment par N. Canouet, L. Fezoui et S. Piperno [12][11] pour des discrétisations en espace par des méthodes de type Galerkin discontinus ou encore par F. Collino, T. Fouquet et P. Joly [29] pour des discrétisations en temps par différences finis.

Dans le premier chapitre de ce manuscrit nous exposons les équations régissant les propagations d'ondes que nous allons considérer, à savoir l'équation des ondes scalaire et les équations de Maxwell. Nous explicitons notamment leur formulation variationnelle et les conditions aux limites qui leur seront respectivement associées. Nous en profitons aussi pour donner certaines contraintes liées à la régularité des espaces de discrétisation associés aux équations de Maxwell, nécessaires au caractère bien posé de ces équations.

Le second chapitre n'a pour vocation que de rappeler brièvement ce qu'est la méthode des éléments finis, notamment la construction des espaces de discrétisation à partir de la définition d'un élément fini, et de mettre en évidence, par une approche intuitive, les mécanismes et contraintes liées à l'utilisation de cette méthode. Nous en profitons pour expliciter un certain nombre d'éléments finis classiques, notamment les éléments finis de Lagrange standards en une et deux dimensions d'espace qui nous intéresseront plus particulièrement par la suite.

Le troisième chapitre résume de manière exhaustive les outils permettant une discrétisation en espace par éléments finis d'ordre arbitrairement élevé, tant dans le cadre de l'équation des ondes scalaire que dans le cadre des équations de Maxwell. Dans le cadre des équations de Maxwell nous considérerons deux types d'éléments finis : les éléments finis d'arête triangulaires et les éléments finis d'arête rectangulaires. S'il est possible d'utiliser ces derniers en toute généralité sur une quadrangulation du domaine de calcul, nous limitons volontairement leur champ d'application dans le cadre d'un maillage cartésien de ce domaine. Cette limitation sera toutefois compensée, comme nous le montrons à la fin de ce chapitre, par le fait que la définition des éléments finis d'arête rectangulaires et triangulaires nous permet d'envisager un couplage conforme de ces éléments finis sur des maillages hybrides (c'est-à-dire cartésiens sauf dans les régions du domaine qui ne le permettent pas, au voisinage des frontières par exemple, où l'on maille par une triangulation).

Dans le quatrième chapitre nous traitons le problème de la condensation de la matrice de masse. Dans un premier temps, en une dimension d'espace pour les éléments finis de Lagrange, nous exposons l'approche qui nous a permis de retrouver les formules de quadrature de Gauss-Lobatto, sur lesquelles sont basée la condensation de la matrice de masse. En deux dimensions d'espace, sur des triangles, cette approche est généralisée par des considérations que l'on a trouvé dans les travaux de N. Tordjman [70], notamment sur les questions de symétrie. Nous reprenons alors ces travaux en généralisant toutefois la description des espaces fonctionnels permettant la réalisation de la condensation de la matrice de masse. Nous décrivons alors un algorithme nous permettant de déterminer les formules de quadrature nécessaires à la réalisation de la condensation de la matrice de masse issue des éléments finis de Lagrange triangulaires. L'algorithme décrit nous a permis non seulement de retrouver les éléments finis de Lagrange condensés $\tilde{P}_1$ à $\tilde{P}_5$, mais aussi de construire un nouvel élément fini condensé $\tilde{P}_6$. Étant donné que la détermination effective

de ces formules de quadrature passe par la résolution de systèmes polynomiaux dont le degré augmente avec l'ordre des éléments finis que l'on cherche à condenser, il ne nous a pas été possible pour le moment d'aller au-delà de cet élément fini condensé $\tilde{P}_6$. Dans de récents travaux, F. X. Giraldo et M. A. Taylor [42] construisent un autre élément fini de type P_6 avec condensation de masse et réussissent même à aller un rang au delà en en déterminant un de type P_7. Nous introduisons ensuite de nouveaux éléments finis qui nous permettent une condensation partielle de la matrice de masse. L'idée est d'orthogonaliser un maximum de fonctions de bases de manière à optimiser le profil de la matrice de masse. Nous terminons alors par le cas de la condensation des éléments finis rectangulaires sur maillage cartésien.

Le cinquième chapitre traite des discrétisations en temps. Nous exposons en premier lieu les discrétisations en temps que nous avons développées. Celles-ci sont explicites et d'ordre arbitrairement élevé. Basées sur une procédure connue sous le nom de procédure de Cauchy-Kowalewski, qui s'apparente à l'approche de l'équation modifiée de Dablain [31] dans l'idée de remplacer les dérivées successives apparaissant dans un développement de Taylor en temps par des dérivées en espace, nous montrons que ces discrétisations ne sont stables que pour certains ordres de discrétisations (dans le cadre de l'équation des ondes scalaire, tout comme celui des équations de Maxwell), et comment les stabiliser lorsqu'elles sont instables. Nous en profitons pour exposer deux autres types de discrétisations en temps qui nous ont paru attrayantes pour la résolution de nos problèmes et auxquelles il nous a paru intéressant de comparer nos discrétisations en temps : les discrétisations en temps symplectiques et les discrétisations en temps de type Runge-Kutta diagonalement implicites.

Dans le sixième chapitre nous faisons une étude comparative des schémas que nous avons développés. Les critères de stabilité, convergence, dissipation, dispersion et rapport coût/précision sont autant de critères qui nous permettront de préférer l'utilisation de l'un ou l'autre des schémas. Si la supériorité des schémas d'ordre élevé, en terme purement de propagation d'onde, sera mise en valeur dans un premier temps, nous mettrons aussi en évidence les limites de la montée en ordre des schémas, dans le cadre de la résolution de l'équation des ondes, sur un cas test plus "physique" dit des tourbillons co-rotatifs.

Le septième chapitre expose une méthode de résolution deux échelles que nous avons développée dans le cadre de la résolution des équations de Maxwell. Nous avons adapté une méthode proposée par R. Glowinski et al. [41] dans le cadre de la résolution d'un laplacien, qui peut s'interpréter comme une méthode de type mortier avec recouvrement total de sous-domaines (voir [6]), dans un premier temps dans le cas stationnaire puis instationnaire.

Chapitre 1

Équations régissant la propagation d'onde

1.1 L'équation d'onde scalaire

Considérons un domaine ouvert $\Omega \subseteq \mathbb{R}^d$. Sauf mention contraire nous ne considérerons que des sous-domaines du plan, c'est-à-dire que nous ne considérerons que le cas particulier où $d = 2$.
Sous les termes d'équation d'onde scalaire (ou plus simplement équation des ondes) nous désignerons l'équation suivante :

$$\partial_t^2 u - \triangle u = f$$

où $u(x,t) : \Omega \times \mathbb{R}^+ \to \mathbb{R}$ est une fonction inconnue que l'on cherche à déterminer et $f(x,t) : \Omega \times \mathbb{R}^+ \to \mathbb{R}$ est une donnée du problème.
La fonction f du second membre de l'équation des ondes sera appelée force imposée ou terme source, et dans le cas où cette fonction est identiquement nulle au cours du temps l'équation des ondes sera dite homogène.
De manière à être bien posée il faut adjoindre à cette équation les conditions initiales

$$u(x,0) = u_0(x) \text{ et } \partial_t u(x,0) = u_1(x) \qquad \forall x \in \Omega,$$

et des conditions de bord (encore appelées conditions aux limites) portant sur la frontière $\Gamma = \partial\Omega$ de Ω dans le cas où celui-ci est strictement inclus dans $\mathbb{R}^d$.
Par condition de bord de Dirichlet nous désignerons la contrainte

$$u(x,t) = g(x,t) \qquad \forall (x,t) \in \Gamma \times \mathbb{R}^+,$$

et par condition de bord de Neumann nous désignerons la contrainte

$$\frac{\partial u}{\partial \overrightarrow{n}}(x,t) = h(x,t) \qquad \forall (x,t) \in \Gamma \times \mathbb{R}^+,$$

où $\overrightarrow{n}$ désigne le vecteur normal unitaire sortant du domaine Ω sur la frontière Γ. Ceci n'a bien entendu de sens que si cette frontière est assez régulière, ce qui est une hypothèse nécessaire au caractère bien posé de l'équation des ondes. Nous désignerons alors Ω comme

un ouvert régulier sans plus de détails.
Nous considérerons aussi deux autres types de conditions aux limites : les conditions aux limites périodiques pour simuler des domaines non bornés, et les conditions aux limites absorbantes, qui s'écrivent

$$\partial_t u(x,t) + \frac{\partial u}{\partial \overrightarrow{n}}(x,t) = 0 \qquad \forall (x,t) \in \Gamma \times \mathbb{R}^+,$$

sensées simuler le caractère sortant du domaine de la propagation d'onde, et ont fait l'objet d'une recherche intensive depuis les premiers travaux de B. Engquist et A. Majda [38] et E.L. Lindmann [52]. Nous ne considérerons que cette condition aux limites absorbante, dite condition aux limites absorbante d'ordre 1, exacte pour les ondes d'incidence normale à Γ.

Formulation variationnelle du problème

Dans l'optique de résoudre l'équation des ondes par une méthode d'éléments finis il nous faut dériver une formulation variationnelle de cette équation.
On se donne donc une fonction $\varphi \in H^1(\Omega)$, on multiplie l'équation des ondes par cette fonction et on intègre sur Ω :

$$\int_\Omega \partial_t^2 u \, \varphi \, dx - \int_\Omega \triangle u \, \varphi \, dx = \int_\Omega f \varphi \, dx.$$

Rappelons alors la formule de Green portant sur le laplacien :

$$\int_\Omega -\triangle u \, \varphi \, dx = \int_\Omega \nabla u \cdot \nabla \varphi \, dx - \int_\Gamma \frac{\partial u}{\partial \overrightarrow{n}} \varphi \, dx,$$

que l'on applique au second terme du premier membre de l'équation pour obtenir

$$\int_\Omega \partial_t^2 u \, \varphi \, dx + \int_\Omega \nabla u \cdot \nabla \varphi \, dx - \int_\Gamma \frac{\partial u}{\partial \overrightarrow{n}} \varphi \, dx = \int_\Omega f \varphi \, dx. \tag{1.1}$$

Les conditions aux limites sont alors directement intégrées à cette formulation variationnelle.

Dans le cas de conditions aux limites de Dirichlet homogènes ($u(x,t) = 0 \ \forall (x,t) \in \Gamma \times \mathbb{R}^+$), il convient de considérer $\varphi \in H_0^1(\Omega)$ et l'équation (1.1) devient

$$\int_\Omega \partial_t^2 u \, \varphi \, dx + \int_\Omega \nabla u \cdot \nabla \varphi \, dx = \int_\Omega f \varphi \, dx,$$

la trace de φ étant identiquement nulle sur Γ. Le traitement d'un point de vue théorique des conditions aux limites de Dirichlet non homogènes est plus délicat et est traité largement dans [53]. Du point de vue du numéricien disons simplement que l'on se ramène au cas homogène par un relèvement de la fonction inconnue.

Dans le cas de conditions aux limites de Neumann, l'équation (1.1) devient

$$\int_\Omega \partial_t^2 u\, \varphi\, dx + \int_\Omega \nabla u \cdot \nabla \varphi\, dx = \int_\Omega f\varphi\, dx + \int_\Gamma h\varphi\, dx.$$

Finalement dans le cas de conditions aux limites absorbantes, l'équation (1.1) devient

$$\int_\Omega \partial_t^2 u\, \varphi\, dx + \int_\Omega \nabla u \cdot \nabla \varphi\, dx + \int_\Gamma \partial_t u\, \varphi\, dx = \int_\Omega f\varphi\, dx.$$

1.2 Les équations de Maxwell

La propagation d'ondes électromagnétiques est modélisée par les équations de Maxwell. Considérons un domaine $\Omega \subset \mathbb{R}^2$. Sous certaines conditions il est possible de découpler les équations de Maxwell en deux jeux d'équations appelés mode transverse électrique, faisant intervenir les champs (E_x, E_y, B_z), et mode transverse magnétique, faisant intervenir les champs (B_x, B_y, E_z). Nous ne considérerons ici que le premier mode, le second pouvant se traiter de manière analogue. Nous considérons donc le mode transverse électrique qui s'écrit :

$$\frac{\partial \overrightarrow{E}}{\partial t} - \overrightarrow{\nabla} \times B = -\overrightarrow{J}, \tag{1.2}$$

$$\frac{\partial B}{\partial t} + \nabla \times \overrightarrow{E} = 0, \tag{1.3}$$

$$\nabla \cdot \overrightarrow{E} = \rho, \tag{1.4}$$

où les composantes sont définies par $\overrightarrow{E} = \begin{pmatrix} E_x \\ E_y \end{pmatrix}$, $B = B_z$ et les opérateurs par $\overrightarrow{\nabla} \times B = \begin{pmatrix} \frac{\partial B}{\partial y} \\ -\frac{\partial B}{\partial x} \end{pmatrix}$ et $\nabla \times \overrightarrow{E} = \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}$. À ces équations il faut bien entendu ajouter des conditions de bord et conditions initiales. Par souci de simplification nous ne considérerons pour le moment que des conditions de bord de type conducteur parfait, ce qui se traduit par $\overrightarrow{E} \times \overrightarrow{n} = 0$, où $\overrightarrow{n}$ désigne le vecteur normal unitaire sortant de Ω sur $\Gamma = \partial\Omega$. Nous supposerons de plus que $\overrightarrow{E} \cdot \overrightarrow{n} = 0$ et $\overrightarrow{J} \cdot \overrightarrow{n} = 0$.
Remarquons dès à présent que l'équation (1.4) du système (dite loi de Gauss) est une conséquence directe de la loi de conservation de la charge :

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \overrightarrow{J} = 0.$$

En effet, en supposant que $\overrightarrow{E}$ et B vérifient la loi d'Ampère (1.2) et en considérant de manière formelle la divergence de cette équation nous obtenons :

$$\frac{\partial}{\partial t} \nabla \cdot \overrightarrow{E} - \underbrace{\nabla \cdot \overrightarrow{\nabla} \times B}_{=0} = -\nabla \cdot \overrightarrow{J}.$$

Si de plus l'équation de conservation de la charge est aussi vérifiée nous obtenons :

$$\frac{\partial}{\partial t} \nabla \cdot \overrightarrow{E} = \frac{\partial \rho}{\partial t}.$$

Il suffit alors que l'équation (1.4) soit vérifiée initialement pour que celle-ci soit automatiquement vérifiée au cours du temps.

Formulation variationnelle du problème

La résolution des équations de Maxwell par la méthode des éléments finis passe par la dérivation d'une formulation variationnelle de ces équations. Soient donc $\overrightarrow{\psi}$, φ et ϕ suffisamment régulières. On multiplie (1.2) par $\overrightarrow{\psi}$, (1.3) par φ et (1.4) par ϕ et on intègre sur Ω :

$$\frac{d}{dt}\int_\Omega \overrightarrow{E}\cdot\overrightarrow{\psi}\,dX - \int_\Omega(\overrightarrow{\nabla}\times B)\cdot\overrightarrow{\psi}\,dX = -\int_\Omega \overrightarrow{J}\cdot\overrightarrow{\psi}\,dX, \tag{1.5}$$

$$\frac{d}{dt}\int_\Omega B\varphi\,dX + \int_\Omega(\nabla\times\overrightarrow{E})\varphi\,dX = 0, \tag{1.6}$$

$$\int_\Omega(\nabla\cdot\overrightarrow{E})\phi\,dX = \int_\Omega \rho\phi\,dX. \tag{1.7}$$

Rappelons alors les formules de Green portant sur le rotationel et sur la divergence :

$$\int_\Omega(\overrightarrow{\nabla}\times G)\cdot\overrightarrow{F}\,dX = \int_\Omega G(\nabla\times\overrightarrow{F})\,dX - \int_\Gamma(G\times\overrightarrow{n})\cdot\overrightarrow{F}\,dS\ ,\ \forall\overrightarrow{F}\in H(rot,\Omega)\text{ et }\forall G\in H^1(\Omega) \tag{1.8}$$

et

$$\int_\Omega(\nabla\cdot\overrightarrow{F})G\,dX = -\int_\Omega\overrightarrow{F}\cdot(\nabla G)\,dX + \int_\Gamma(\overrightarrow{F}\cdot\overrightarrow{n})G\,dS\ ,\ \forall\overrightarrow{F}\in H(div,\Omega)\text{ et }\forall G\in H^1(\Omega). \tag{1.9}$$

En appliquant la formule de Green sur le deuxième terme de l'équation (1.5) nous obtenons :

$$\int_\Omega(\overrightarrow{\nabla}\times B)\cdot\overrightarrow{\psi}\,dX = \int_\Omega B(\nabla\times\overrightarrow{\psi})\,dX - \int_\Gamma(B\times\overrightarrow{n})\cdot\overrightarrow{\psi}\,dS. \tag{1.10}$$

Nous en profitons pour introduire les deux types de conditions aux limites que nous utiliserons (en dehors des conditions limites périodiques) : les conditions limites du type conducteur parfait qui s'écrivent

$$\overrightarrow{E}\times\overrightarrow{n} = 0,$$

et les conditions aux limites absorbantes d'ordre 1 de Silver-Müller (voir [15]) qui s'écrivent

$$(\overrightarrow{E}\times\overrightarrow{n} + B)\times\overrightarrow{n} = 0.$$

L'intégration des conditions limites de conducteur parfait se fait alors de manière naturelle en imposant la contrainte $\overrightarrow{\psi}\times\overrightarrow{n} = 0$ sur Γ, de sorte que le second terme du second membre de l'équation (1.10) s'annule :

$$\begin{aligned}\int_\Omega(\overrightarrow{\nabla}\times B)\cdot\overrightarrow{\psi}\,dX &= \int_\Omega B(\nabla\times\overrightarrow{\psi})\,dX - \int_\Gamma(B\times\overrightarrow{n})\cdot\overrightarrow{\psi}\,dS\\ &= \int_\Omega B(\nabla\times\overrightarrow{\psi})\,dX + \int_\Gamma B(\overrightarrow{\psi}\times\overrightarrow{n})\,dS\\ &= \int_\Omega B(\nabla\times\overrightarrow{\psi})\,dX,\end{aligned} \tag{1.11}$$

ce qui nous transforme l'équation (1.5) en

$$\frac{d}{dt}\int_\Omega \overrightarrow{E}\cdot\overrightarrow{\psi}\;dX - \int_\Omega B(\nabla\times\overrightarrow{\psi})\;dX = -\int_\Omega \overrightarrow{J}\cdot\overrightarrow{\psi}\;dX.$$

L'intégration des conditions limites de Silver-Müller se fait elle aussi à partir de l'équation (1.10) :

$$\begin{aligned}\int_\Omega(\overrightarrow{\nabla}\times B)\cdot\overrightarrow{\psi}\;dX &= \int_\Omega B(\nabla\times\overrightarrow{\psi})\;dX - \int_\Gamma (B\times\overrightarrow{n})\cdot\overrightarrow{\psi}\;dS\\ &= \int_\Omega B(\nabla\times\overrightarrow{\psi})\;dX + \int_\Gamma((\overrightarrow{E}\times\overrightarrow{n})\times\overrightarrow{n})\cdot\overrightarrow{\psi}\;dS\\ &= \int_\Omega B(\nabla\times\overrightarrow{\psi})\;dX - \int_\Gamma(\overrightarrow{E}\times\overrightarrow{n})(\overrightarrow{\psi}\times\overrightarrow{n})\;dS,\end{aligned} \tag{1.12}$$

ce qui nous transforme cette fois l'équation (1.5) en

$$\frac{d}{dt}\int_\Omega \overrightarrow{E}\cdot\overrightarrow{\psi}\;dX - \int_\Omega B(\nabla\times\overrightarrow{\psi})\;dX + \int_\Gamma(\overrightarrow{E}\times\overrightarrow{n})(\overrightarrow{\psi}\times\overrightarrow{n})\;dS = -\int_\Omega \overrightarrow{J}\cdot\overrightarrow{\psi}\;dX.$$

Par la suite et sauf mention contraire nous ne considérerons que les équations de Maxwell avec des conditions aux limites de conducteur parfait.

Pour l'équation (1.7) nous utilisons la formule de Green portant sur la divergence pour obtenir la forme variationnelle de l'équation de Gauss

$$-\int_\Omega \overrightarrow{E}\cdot(\nabla\phi)\;dX = \int_\Omega \rho\phi\;dX.$$

Pour finir dérivons une formulation variationnelle de l'équation de conservation de la charge

$$\frac{d}{dt}\int_\Omega \rho\xi\;dX + \int_\Omega(\nabla\cdot\overrightarrow{J})\xi\;dX = 0,$$

qui devient après utilisation de la formule de Green (1.9)

$$\frac{d}{dt}\int_\Omega \rho\xi\;dX - \int_\Omega \overrightarrow{J}\cdot(\nabla\xi)\;dX = 0.$$

Ne faisant pas de traitement particulier sur l'équation de Faraday (1.6), celle-ci est naturellement bien posée dès que $\varphi\in L^2(\Omega)$.
Nous obtenons donc le problème suivant :
Trouver $\overrightarrow{E}\in H_0(rot,\Omega)$ et $B\in L^2(\Omega)$ tels que $\forall\overrightarrow{\psi}\in H_0(rot,\Omega)$, $\forall\varphi\in L^2(\Omega)$ et $\forall\phi\in H^1(\Omega)$:

$$\frac{d}{dt}\int_\Omega \overrightarrow{E}\cdot\overrightarrow{\psi}\;dX - \int_\Omega B(\nabla\times\overrightarrow{\psi})\;dX = -\int_\Omega \overrightarrow{J}\cdot\overrightarrow{\psi}\;dX, \tag{1.13}$$

$$\frac{d}{dt}\int_\Omega B\varphi\;dX + \int_\Omega(\nabla\times\overrightarrow{E})\varphi\;dX = 0, \tag{1.14}$$

$$-\int_\Omega \overrightarrow{E}\cdot(\nabla\phi)\;dX = \int_\Omega \rho\phi\;dX. \tag{1.15}$$

Il est alors important de remarquer que sous sa forme variationnelle aussi, l'équation de Gauss est automatiquement vérifiée au cours du temps si celle-ci est initialement vérifiée et que l'équation de conservation de la charge

$$\frac{d}{dt}\int_\Omega \rho\xi \, dX - \int_\Omega \overrightarrow{J}\cdot(\nabla\xi)\, dX = 0 \qquad \forall \xi \in H^1(\Omega) \tag{1.16}$$

est vérifiée au cours du temps. En effet, pour tout $\xi \in H^1(\Omega)$ on a $\nabla\times(\nabla\xi) = 0$, ce qui signifie en particulier que $\nabla\xi \in H(rot,\Omega)$ et que l'on peut donc évaluer l'équation (1.13) pour $\overrightarrow{\psi} = \nabla\xi$:

$$\frac{d}{dt}\int_\Omega \overrightarrow{E}\cdot\nabla\xi\, dX - \int_\Omega B(\nabla\times\nabla\xi)\, dX = -\int_\Omega \overrightarrow{J}\cdot\nabla\xi\, dX$$

c'est-à-dire

$$\frac{d}{dt}\int_\Omega \overrightarrow{E}\cdot\nabla\xi\, dX = -\int_\Omega \overrightarrow{J}\cdot(\nabla\xi)\, dX,$$

puis en utilisant l'équation (1.16)

$$\frac{d}{dt}\int_\Omega \overrightarrow{E}\cdot\nabla\xi\, dX + \frac{d}{dt}\int_\Omega \rho\xi\, dX = 0,$$

il suffit alors que l'équation (1.15) soit vérifiée initialement pour que celle-ci soit automatiquement vérifiée au cours du temps.
Il est important de souligner que tout ceci ne tient que parce que la suite

$$H^1(\Omega) \overset{\overrightarrow{\nabla}}{\longrightarrow} H(rot,\Omega) \overset{\nabla\times}{\longrightarrow} L^2(\Omega)$$

est exacte, ce qui signifie en particulier que l'image de $H^1(\Omega)$ par l'opérateur $\overrightarrow{\nabla}$ est non seulement incluse dans $H(rot,\Omega)$ mais aussi dans le noyau de l'opérateur $\nabla\times$. C'est un point qui, s'il est respecté d'un point de vue discret, nous permettra de ne résoudre que les formes variationnelles des équations d'Ampère (1.13) et de Faraday (1.14) pour résoudre les équations de Maxwell. Plus précisément il nous faudra introduire des sous-espaces vectoriels de dimension finie $X \subset H^1(\Omega)$, $W \subset H(rot,\Omega)$ et $V \subset L^2(\Omega)$ vérifiant aussi

$$X \overset{\overrightarrow{\nabla}}{\longrightarrow} W \overset{\nabla\times}{\longrightarrow} V,$$

pour obtenir une approximation de la bonne solution des équations de Maxwell.
Des études mathématiques bien plus complètes des équations de Maxwell ont été menées par exemple dans [55], [32] ou [4].

Remarque 1.2.1. *Pour dériver une formulation variationnelle des équations de Maxwell, nous aurions pu, plutôt que d'intégrer par partie le deuxième terme de l'équation (1.5), intégrer par partie le deuxième terme de l'équation (1.6) en utilisant la formule de Green suivante :*

$$\int_\Omega (\nabla\times\overrightarrow{G})F\, dX = \int_\Omega \overrightarrow{G}\cdot(\overrightarrow{\nabla}\times F)\, dX - \int_\Gamma (\overrightarrow{G}\times\overrightarrow{n})F\, dS \ , \forall \overrightarrow{G} \in H(rot,\Omega) \ et \ \forall F \in H^1(\Omega) \tag{1.17}$$

de manière à obtenir le problème suivant à résoudre pour $\overrightarrow{\psi} \in H_0(rot,\Omega)$, $\varphi \in H^1(\Omega)$, $\phi \in H^1(\Omega)$ *:*

$$\frac{d}{dt}\int_\Omega \overrightarrow{E}\cdot\overrightarrow{\psi}\,dX - \int_\Omega (\overrightarrow{\nabla}\times B)\cdot\overrightarrow{\psi}\,dX = -\int_\Omega \overrightarrow{J}\cdot\overrightarrow{\psi}\,dX, \tag{1.18}$$

$$\frac{d}{dt}\int_\Omega B\varphi\,dX + \int_\Omega \overrightarrow{E}\cdot(\overrightarrow{\nabla}\times\varphi)\,dX = 0, \tag{1.19}$$

$$-\int_\Omega \overrightarrow{E}\cdot(\nabla\phi)\,dX = \int_\Omega \rho\phi\,dX. \tag{1.20}$$

La suite exacte nous permettant de ne résoudre que les équations d'Ampère et Faraday (1.18) et (1.19) pour obtenir la bonne solution des équations de Maxwell (l'équation de Gauss (1.20) étant à nouveau une conséquence de l'équation de conservation de la charge qui est vérifiée) à considérer est alors la suivante :

$$H^1(\Omega) \xrightarrow{\overrightarrow{\nabla}\times} H(div,\Omega) \xrightarrow{\nabla\cdot} L^2(\Omega).$$

Les raisons qui nous ont poussées à ne pas considérer cette approche seront motivées par la suite dans la section 3.2.

Chapitre 2

Introduction à la méthode des éléments finis

2.1 Rappel de la méthode de Ritz-Galerkin

La méthode de Ritz-Galerkin repose sur la formulation variationnelle d'une équation aux dérivées partielles. Pour se fixer les idées nous prendrons pour exemple le problème du laplacien dans un domaine Ω de frontière Γ, avec condition de bord de Dirichlet homogène :
Trouver $u : \Omega \to \mathbb{R}$ tel que

$$\begin{cases} -\triangle u &= f \quad \text{dans } \Omega, \\ u &= 0 \quad \text{sur } \Gamma. \end{cases}$$

Sous forme variationnelle ce problème s'écrit :
Trouver $u \in H_0^1(\Omega)$ tel que

$$\int_\Omega \nabla u \cdot \nabla v \, dx = \int_\Omega f v \, dx \qquad \forall v \in H_0^1(\Omega).$$

Cette formulation peut être vue comme le cas particulier d'un problème du type suivant :
Trouver $u \in V$ tel que

$$a(u, v) = l(v) \qquad \forall v \in V, \tag{2.1}$$

avec $a(\cdot, \cdot)$ une forme bilinéaire symétrique, et $l(\cdot)$ une forme linéaire. Rappelons que le lemme fondamental de Lax-Milgram [17] nous donne l'existence et l'unicité de la solution d'un tel problème sous certaines conditions de régularité sur a et l :

Lemme 2.1.1. *Soit V un espace de Hilbert muni d'une norme $\| \cdot \|_V$. Soit $a(\cdot, \cdot)$ une forme bilinéaire continue et coercive sur $V \times V$, i.e.*

- *Continuité : il existe une constante C telle que pour tout $u, v \in V$*

$$|a(u, v)| \leq C \parallel u \parallel_V \parallel v \parallel_V,$$

- *Coercivité : il existe une constante $\alpha > 0$ telle que pour tout $u \in V$*

$$|a(u, u)| > \alpha \parallel u \parallel_V^2 .$$

Soit $l(\cdot)$ une forme linéaire et continue sur V, i.e.
- *Il existe une constante C telle que pour tout $u \in V$*

$$|l(u)| \leq C \parallel u \parallel_V .$$

Alors il existe un unique $u \in V$ tel que

$$a(u, v) = l(v) \qquad \forall v \in V.$$

La méthode de Ritz-Galerkin consiste alors à chercher une solution approchée u_h du problème (2.1) dans un sous-espace de dimension finie de V. L'approximation u_h de u sera alors d'autant meilleure que l'espace de discrétisation V_h sera proche de l'espace V. Le problème (2.1) se réécrit alors naturellement :
Trouver $u_h \in V_h$ tel que

$$a(u_h, v_h) = l(v_h) \qquad \forall v_h \in V_h, \tag{2.2}$$

où $V_h \subset V$ est un sous-espace vectoriel de V de dimension finie N. Il n'est alors bien entendu pas nécessaire de résoudre le problème (2.2) pour tout $v_h \in V_h$ mais uniquement sur une base $(\Psi_1, \ldots, \Psi_N)$ de V_h. Reste alors à remarquer qu'un élément $u_h \in V_h$ s'écrit $u_h(x) = \sum_{i=1}^{N} u_i \Psi_i(x)$ et d'utiliser la linéarité de l'opérateur a par rapport à sa première composante pour réécrire finalement le problème sous la forme :
Trouver $(u_1, \ldots, u_N) \in \mathbb{R}^N$ tel que

$$\sum_{i=1}^{N} u_i a(\Psi_i, \Psi_j) = l(\Psi_j) \qquad \forall j = 1, \ldots, N. \tag{2.3}$$

Ainsi le problème à résoudre n'est rien d'autre qu'un système linéaire de dimension $N \times N$ que l'on écrit :

$$AU_N = L, \tag{2.4}$$

où $A = (a(\Psi_i, \Psi_j))_{1 \leq i,j \leq N}$, L est le vecteur colonne de composantes $l(\Psi_i)$ et U est le vecteur colonne contenant les coefficients u_i de u_h dans la base $(\Psi_1, \ldots, \Psi_N)$ de V_h.

Il reste maintenant à expliciter la construction des espaces V_h inclus dans V. Dans la pratique c'est à l'aide d'éléments finis que l'on construit ces espaces fonctionnels.

2.2 Définition d'un élément fini

On considère un triplet $(\hat{K}, P, \Sigma)$ où
(i) $\hat{K}$ est un sous-ensemble fermé de $\mathbb{R}^d$ d'intérieur non vide,
(ii) P est un espace vectoriel de dimension finie de fonctions définies sur $\hat{K}$,
(iii) Σ est un ensemble de formes linéaires sur P de cardinal fini N.

Remarque 2.2.1. *Dans la pratique les formes linéaires de Σ sont appelés degrés de liberté de l'élément fini.*

Définition 2.2.2. *On dit que Σ est P-unisolvant si pour tout N-uplet $(\alpha_1, \ldots, \alpha_N)$, il existe un unique élément $p \in P$ tel que $\sigma_i(p) = \alpha_i$ pour $i = 1, \ldots, N$.*

Ce qui nous amène à la définition d'un élément fini :

Définition 2.2.3. *Le triplet $(\hat{K}, P, \Sigma)$ est appelé élément fini de $\mathbb{R}^n$ s'il satisfait (i), (ii) et (iii) et si Σ est P-unisolvant.*

Remarque 2.2.4. *Comme application il est possible de montrer que pour qu'un ensemble Σ de formes linéaires puisse être P-unisolvant il faut que son cardinal soit égal à la dimension de P, i.e.*

$$Card(\Sigma) = Dim(P).$$

Remarque 2.2.5. *Il est bon de remarquer le lien entre l'unisolvance et l'interpolation : dire qu'un ensemble Σ est P-unisolvant, c'est exactement dire que pour tout N-uplet $(\alpha_1, \ldots, \alpha_N)$ il existe un unique interpolé dans P, l'interpolation ce faisant via les formes linéaires de Σ. Cela signifie en particulier que toute fonction de P peut être reconstruite (c'est-à-dire aussi représentée en machine) de manière unique par un N-uplet $(\alpha_1, \ldots, \alpha_N)$.*

Le lemme suivant, dont la preuve peut se trouver dans [17] par exemple, est d'une importance fondamentale dans la mise en oeuvre des éléments finis.

Lemme 2.2.6. *L'ensemble Σ est P-unisolvant si et seulement si les deux propriétés suivantes sont vérifiées :*
- *Il existe N fonctions $p_i \in P$ telles que $\sigma_j(p_i) = \delta_{ij}$,*
- *$Dim(P) = N$.*

La famille $\{p_1, \ldots, p_N\}$ est alors une base de P.

Ainsi si la dimension de P est égale au cardinal de Σ, Σ devient P-unisolvant si et seulement si on est capable d'exhiber une famille $\{\Psi_1, \ldots, \Psi_N\}$ de P vérifiant

$$\sigma_i(\Psi_j) = \delta_{ij} \qquad \forall i, j = 1 \ldots N. \tag{2.5}$$

Toute fonction $p \in P$ sera alors décomposée sur la base $(\Psi_1, \ldots, \Psi_N)$ de la manière suivante :

$$p(x) = \sum_{i=1}^{N} \sigma_i(p)\Psi_i(x). \tag{2.6}$$

Il suffit ensuite de définir un maillage (pour l'instant le terme de maillage désigne une quelconque partition) de Ω par des éléments K_i du même "type" que $\hat{K}$ (c'est-à-dire que chaque K_i est l'image de $\hat{K}$ par une transformation bijective) et de définir l'espace V_h de la manière suivante :

$$V_h = \{v \in V \ / \ v_{|K_i} \in P\}. \tag{2.7}$$

Cette définition de V_h nous assure automatiquement la conformité de la méthode, c'est-à-dire que V_h est inclus dans V. Dans la pratique il est préférable d'exhiber la régularité nécessaire pour vérifier cette condition et de l'expliciter directement dans la définition de

l'espace V_h. Par exemple sur le problème du laplacien, nous avons $V = H_0^1(\Omega)$. Or dans l'espace $H^1(\Omega)$ nous avons le théorème suivant :

Théoreme 2.2.7. *Soit Ω et $(\Omega_i)_{i=1,\dots,N}$ des ouverts de $\mathbb{R}^n$ tels que $\Omega_i \cap \Omega_j = \emptyset$ et $\overline{\Omega} = \cup_{i=1}^N \overline{\Omega}_i$. Soit $(u_1, \dots, u_N) \in \Pi_{i=1}^N H^1(\Omega_i)$ et u la fonction définie par $u_{|\Omega_i} = u_i$. On suppose que $u_{i|\sum_{ij}} = u_{j|\sum_{ij}}$ sur l'interface $\sum_{ij}$ entre les domaines Ω_i et Ω_j pour tout $i, j = 1, \dots, N$. Alors $u \in H^1(\Omega)$.*

Dans le cadre plus restrictif où l'on ne considère pas en toute généralité $(u_1, \dots, u_N) \in \Pi_{i=1}^N H^1(\Omega_i)$ mais $(u_1, \dots, u_N) \in \Pi_{i=1}^N C^0(\Omega_i)$, la condition $u_{i|\sum_{ij}} = u_{j|\sum_{ij}}$ devient aussi nécessaire pour que $u \in H^1(\Omega)$. Ainsi pour des fonctions polynomiales par élément du maillage, c'est à dire si P est un espace vectoriel de fonctions polynomiales définies sur K (nous ne considérerons que ce cas), $u \in H^1(\Omega)$ si et seulement si $u \in C^0(\Omega)$. De sorte que l'espace V_h soit décrit plus précisément par :

$$V_h = \{v \in C^0(\Omega) \ / \ v_{|K_i} \in P\}. \tag{2.8}$$

Le choix de l'élément fini n'est donc pas anodin, puisqu'il doit permettre d'assurer naturellement la condition de continuité à l'interface de deux éléments adjacents de manière à ce que la résolution du système linéaire (2.4) nous donne une solution qui a déjà la régularité nécessaire à la conformité de la méthode. Ceci n'est bien entendu possible que si deux éléments adjacents partagent un nombre suffisant de degrés de liberté. Nous allons préciser cette notion de partage de degrés de liberté, mais avant tout il nous faut terminer d'expliciter la construction de l'espace V_h.

2.3 Construction pratique de l'espace de discrétisation V_h

Une fois que l'on s'est fixé un élément fini $(\hat{K}, P, \Sigma)$ il convient d'expliciter plus en détail la construction de l'espace de discrétisation V_h que l'on a défini jusqu'à présent par :

$$V_h = \{v \in V \ / \ v_{|K_i} \in P\}, \tag{2.9}$$

où les K_i sont les éléments définissant un maillage du domaine Ω.
Il suffit bien entendu d'expliciter une base de V_h pour que celui-ci soit entièrement déterminé. La définition de l'espace V_h dans (2.9) nous laisse penser qu'il faut construire pour chacun des éléments du maillage une base de l'espace P restreint à K_i, ce qui est le cas, mais qui se fait de manière totalement transparente dès lors que l'on voit chaque élément du maillage comme l'image de l'élément $\hat{K}$, dit élément de référence, par une transformation bijective, et que l'on transporte par la même transformation les formes linéaires définies sur l'élément de référence vers chacun des éléments K_i.
De manière plus précise, notons $\{\hat{\sigma}_i\}_{i=1,\dots,N}$ l'ensemble des formes linéaires de Σ et $\{\hat{\Psi}_i\}_{i=1,\dots,N}$ l'ensemble des fonctions de base associées (c'est-à-dire les N fonctions de P déterminées de manière unique par $\hat{\sigma}_i(\hat{\Psi}_j) = \delta_{ij}$). Fixons-nous un élément K_l du maillage et notons Φ_l la transformation bijective transportant l'élément de référence $\hat{K}$ vers l'élément K_l.
Nous définissons alors un nouvel ensemble de formes linéaires $\{\sigma_i^l\}_{i=1,\dots,N}$ par

$$\sigma_i^l(f) = \hat{\sigma}_i(f \circ \Phi_l),$$

c'est-à-dire que nous transportons les formes linéaires définies sur l'élément de référence $\hat{K}$ vers l'élément K_l. À ce nouvel ensemble de formes linéaires il faut associer un nouvel ensemble de fonctions de bases $\{\Psi_i^l\}_{i=1,\dots,N}$ vérifiant $\sigma_i^l(\Psi_j^l) = \delta_{ij}$, de sorte que l'on aura explicité une base de V_h localement sur chaque élément K_l du maillage, ce qui est nécessaire et suffisant à la description de V_h. Un calcul élémentaire nous donne alors

$$\sigma_i^l(\Psi_j^l) = \hat{\sigma}_i(\Psi_j^l \circ \Phi_l) = \delta_{ij}.$$

Or, la définition même de $\hat{\Psi}_j$ et l'unisolvance de Σ sur P, nous dit que ceci n'est possible que si

$$\Psi_j^l \circ \Phi_l = \hat{\Psi}_j,$$

ou de manière totalement équivalente

$$\Psi_j^l = \hat{\Psi}_j \circ \Phi_l^{-1}.$$

Remarque 2.3.1. *Il est important de se rendre compte que la donnée explicite des degrés de liberté, tout comme de l'expression locale sur chaque élément constituant le maillage du domaine Ω des fonctions de base, est indispensable à la projection sur l'espace V_h d'une fonction et à la reconstruction d'une fonction V_h qui n'est connue que par les valeurs prises par l'ensemble des degrés de libertés .*

Remarque 2.3.2. *La souplesse de la méthode des éléments finis vient exactement de ce qui vient d'être écrit; s'il n'est donc qu'une chose à retenir c'est bien celle-ci :* **chaque élément K_l du maillage est vu comme l'image par une transformation bijective Φ_l d'un élément de référence $\hat{K}$. Chaque fonction de base Ψ_j^l de l'espace V_h définie localement sur l'élément K_l comme étant l'unique solution de $\sigma_i^l(\Psi_j^l) = \delta_{ij}$ $(i = 1, \dots, dim(P) = card(\Sigma))$, où $\sigma_i^l(f) = \hat{\sigma}_i(f \circ \Phi_l)$ n'est autre que l'image de la $i^{\text{ième}}$ forme linéaire de Σ transportée sur l'élément K_l et est donnée explicitement par $\Psi_j^l = \hat{\Psi}_j \circ \Phi_l^{-1}$, où $\hat{\Psi}_j$ désigne la $j^{\text{ième}}$ fonction de base définie comme étant l'unique solution de $\hat{\sigma}_i(\hat{\Psi}_j) = \delta_{ij}$ $(i = 1, \dots, dim(P) = card(\Sigma))$.**

Maintenant que l'on a donné un sens précis aux degrés de liberté définissant les fonctions de base de l'espace V_h, il nous est possible de préciser la notion de partage de degrés de liberté assurant la conformité de la méthode. Dans l'ensemble $\{\sigma_i^l(f) = \hat{\sigma}_i(f \circ \Phi_l)\}$ défini pour $i = 1, \dots, card(\Sigma)$ et l parcourant les éléments du maillage, il est tout à fait envisageable de faire correspondre deux formes linéaires, c'est-à-dire que pour certains couples (l_1, l_2) et (i, j) il peut y avoir l'égalité :

$$\sigma_i^{l_1} = \sigma_j^{l_2},$$

de sorte qu'il est possible d'imposer la conformité de l'espace V_h en imposant l'égalité des valeurs prises par ces degrés de liberté partagés par plusieurs éléments (ou de manière optimale en ne considérant qu'un seul degré de liberté dans chaque classe de degrés de liberté partagés) dès lors que ces degrés de liberté portent effectivement sur la contrainte de conformité.
Il faut alors remarquer que la contrainte de conformité porte sur l'espace des traces, aux

interfaces des éléments, des fonctions de P (la notion de trace dont on parle ici prend un sens différent suivant le problème que l'on considère et donc aussi de l'élément fini). En effet sur l'ouverture de chaque élément du maillage les fonctions de V_h sont des fonctions de P, c'est-dire aussi régulières que l'on veut (P étant en général un espace polynomial). Les seuls problèmes qui peuvent apparaître sont donc localisés aux interfaces d'éléments voisins (points, arêtes ou facettes suivant que l'on se place en une, deux ou trois dimensions d'espace). L'idée est alors de maîtriser entièrement cet espace de trace en s'assurant qu'un certain nombre des degrés de liberté définissant l'élément fini, définissent aussi, en particulier un élément fini dans l'espace des traces, et que ces mêmes degrés de liberté sont partagés par des éléments voisins. Cela implique en particulier un recollement conforme des interfaces de deux éléments voisins, c'est pourquoi à partir d'ici et sauf mention contraire nous désignerons par maillages les maillages conformes :

Définition 2.3.3. *On appelle maillage conforme d'un domaine Ω toute partition de ce domaine tel que les frontières d'éléments voisins de cette partition se recouvrent exactement.*

Par exemple en deux dimensions d'espace les éléments K_i doivent former une partition de Ω, tel que l'intersection de deux arêtes de deux éléments distincts est soit vide, soit réduite à un sommet commun aux deux éléments, soit une arête pour chacun des deux éléments.

Remarque 2.3.4. *La condition d'unisolvance d'un sous-ensemble de Σ sur l'espace fonctionnel des traces de fonction de P aux interfaces des éléments peut apparaître comme une vraie difficulté à la construction d'un élément fini ; c'est pourquoi dans la pratique on prend le problème à l'envers : une fois que l'on s'est fixé un espace P, on regarde la nature de la trace d'une fonction de P aux interfaces de $\hat{K}$ et l'on définit un certain nombre de degrés de liberté unisolvants sur cet espace. Ensuite il suffit de chercher un ensemble de formes linéaires unisolvant sur le sous-espace de P constitué des fonctions dont la trace est nulle aux interfaces. Ce dernier espace étant exactement le complémentaire du premier dans P, on a trivialement que la réunion des deux ensembles de formes linéaires est unisolvant sur l'espace P.*

2.4 Quelques exemples d'éléments finis

Cette section a pour objectif non seulement de donner à titre d'exemple quelques éléments finis classiques, mais aussi d'en expliquer la construction. Commençons dans un premier temps par les éléments finis de Lagrange en une dimension d'espace. On appelle éléments finis de Lagrange tout élément fini dont les formes linéaires constituant Σ ne sont que des évaluations de fonctions en des points de $\hat{K}$, c'est-à-dire

$$\sigma_i(p) = p(a_i) \qquad \forall p \in P,$$

où $\{a_i\}_{i=1,\dots,N}$ est un ensemble donné de points de $\hat{K} = [0,1]$. On note généralement P_k l'élément fini de Lagrange d'ordre k, c'est-à-dire l'élément fini dont l'espace fonctionnel P n'est autre que l'espace des polynômes de degré inférieur ou égal à k, que l'on notera $\mathbb{P}_k$. Cet espace étant de dimension $k+1$, l'ensemble Σ est automatiquement constitué

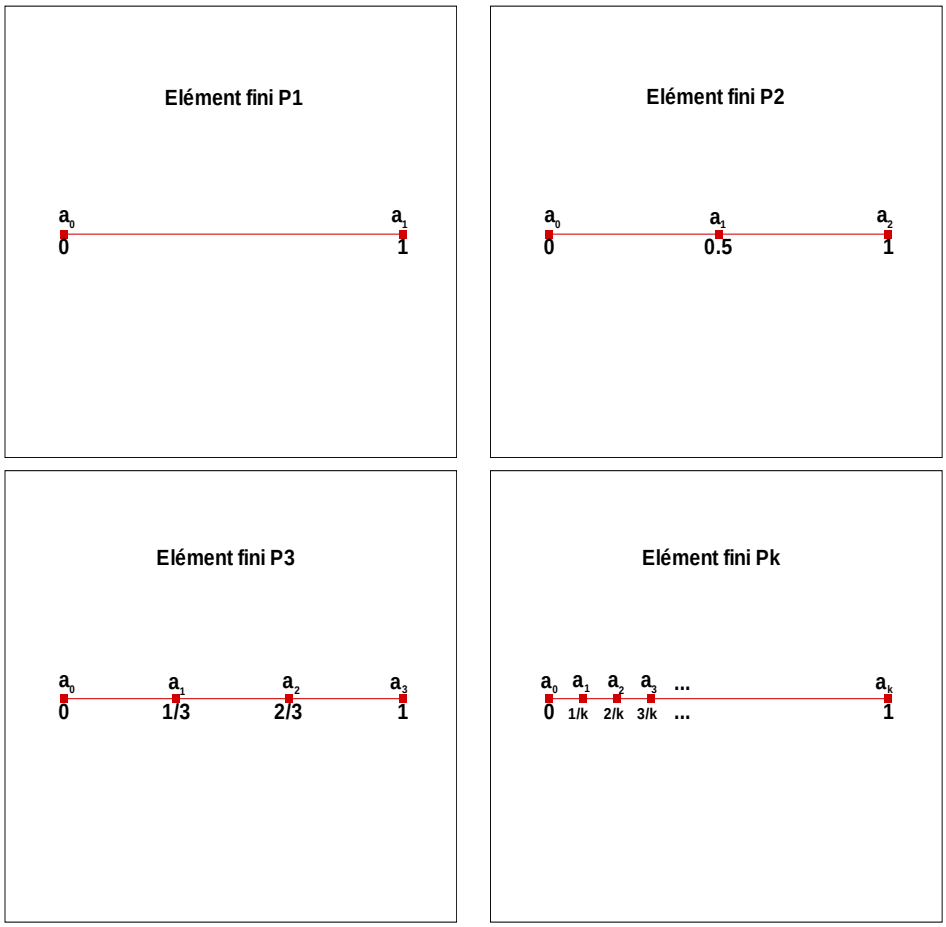

FIG. 2.1 –

de $k+1$ formes linéaires qui sont les évaluations de fonctions de P en $k+1$ points a_i de $\hat{K} = [0,1]$. Ces éléments ayant pour vocation à être conformes dans $H^1(\Omega)$, il est nécessaire et suffisant qu'au moins les extrémités du domaine $\hat{K}$ soit incluses dans l'ensemble des $\{a_i\}$ pour assurer la continuité des fonctions de V_h à l'interface des éléments partitionnant Ω (et par là même l'inclusion $V_h \subset V$). La localisation de ces points pour les éléments finis de Lagrange standards, représentée dans la figure 2.1, est donnée par

$$a_i = \frac{i}{k} \qquad \forall i = 0, \ldots, k.$$

Remarquons que toute autre localisation de $k+1$ points donnera un élément fini du moment que ces points sont distincts (pour l'unisolvance) et que les bords de l'intervalle en font partie (pour la conformité).

Un autre exemple classique d'élément fini en une dimension d'espace est l'exemple des éléments finis d'Hermite. Ceux-ci font non seulement apparaître, en tant que formes linéaires, l'évaluation de fonctions en certains points de $\hat{K}$ mais aussi l'évaluation de dérivées de fonctions en ces points. Par exemple l'élément fini d'Hermite de plus bas degré est défini

comme suit : $\hat{K} = [0,1]$, $P = \mathbb{P}_3$ (l'ensemble des polynômes de degré inférieur ou égal à 3) et $\Sigma = \{\sigma_1, \ldots, \sigma_4\}$ où $\forall p \in P$ on définit

$$\sigma_1(p) = p(0) \qquad \sigma_2(p) = p(1) \qquad \sigma_3(p) = p\prime(0) \qquad \sigma_4(p) = p\prime(1).$$

Cet élément fini a la particularité d'imposer non seulement la continuité de la solution mais aussi sa dérivabilité et la continuité de sa dérivée. En effet, les fonctions de l'espace de discrétisation V_h ainsi défini étant polynomiales à l'intérieur de chaque élément partitionnant le domaine Ω, il suffit d'assurer l'égalité des valeurs des fonctions et de leur nombre dérivé à l'interface de chaque élément partitionnant le domaine Ω, ce qui est garanti par la définition même des degrés de liberté, pour que ces fonctions soient $C^1(\Omega)$.

Remarque 2.4.1. *Bien que pour les éléments finis de Lagrange P_3 et de Hermite de plus bas degré, les espaces polynomiaux P définissant ces éléments finis sont les mêmes, c'est-à-dire $\mathbb{P}_3$; les espaces de discrétisation V_h résultant, pour un domaine Ω et un maillage donné, sont clairement différents.*

Passons maintenant au cas des éléments finis de Lagrange en deux dimensions d'espace. Soit $\hat{K} = [0,1] \times [0,1]$, le carré unité et $P = \mathbb{Q}_k =< x^m y^n \ / \ 0 \leq m, n \leq k >$. $\mathbb{Q}_k$ étant un espace polynomial de dimension $(k+1)^2$, l'ensemble Σ doit nécessairement être composé de $(k+1)^2$ formes linéaires, qui sont des évaluations de polynômes de $\mathbb{Q}_k$ en $(k+1)^2$ points. Il est important de remarquer que la trace (dans ce contexte il ne s'agit que de la restriction) de n'importe quel polynôme de $\mathbb{Q}_k$ à une ligne horizontale ou verticale (c'est-à-dire en fixant la première ou la seconde composante) est un polynôme univarié de degré k. Bien entendu tout polynôme de ce type est déterminé de manière unique par $k+1$ valeurs qu'il prend en $k+1$ points distincts. Ceci signifie qu'en aucune manière une ligne horizontale ou verticale ne doit faire apparaître plus de $k+1$ points auxquels sont associés les degrés de liberté sous peine de perdre l'unisolvance de l'élément fini (la $k + 2^{\text{ième}}$ valeur d'un polynôme étant fixée par les $k+1$ premières, il est impossible que quelles que soit les $(k+1)^2$ valeurs α_i des degrés de liberté, il existe un polynôme de $p \in \mathbb{Q}_k$ tel que $p(a_i) = \alpha_i$). Cette remarque s'appliquant aussi en particulier aux arêtes de K, la condition de conformité nous impose de mettre exactement $k+1$ points par arête (les degrés de liberté associés à des points localisés sur une arête commune à deux éléments voisins étant partagés par ces deux éléments, ceux-ci doivent définir un unique polynôme univarié de degré k). Suivant ces conditions il est possible de construire un bon nombre d'éléments finis. Donnons par exemple l'élément fini standard Q_k : les points auxquels sont associés les formes linéaires sont localisés à l'intersection de deux fois $k+1$ lignes ($k+1$ horizontales et $k+1$ verticales) équiréparties, dont les arêtes du carré $\hat{K}$ (voir figure 2.2).

Il est important de noter que ces éléments finis respectent une contrainte qu'il ne vaut mieux pas contourner d'un point de vue pratique : la répartition barycentrique des points auxquels sont associés les degrés de liberté est symétrique et identique sur chaque arête. Cela signifie que si un point est localisé sur l'arête $[S_1, S_2]$ définie par les sommets S_1 et S_2 de K (voir figure 2.2), par les coordonnées barycentriques $((S_1, \alpha), (S_2, 1-\alpha))$ avec $0 \leq \alpha \leq 1$; alors il est nécessaire de définir $((S_1, 1-\alpha), (S_2, \alpha))$ et aussi ces mêmes points sur les trois autres arêtes $[S_2, S_3]$, $[S_3, S_4]$ et $[S_4, S_1]$. La raison à ceci est la suivante : l'utilisation de la méthode des éléments finis passe par un maillage du domaine de calcul Ω

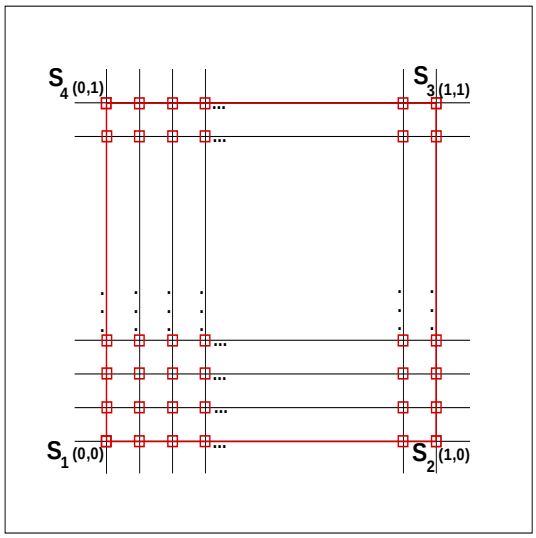

FIG. 2.2 – Localisation des points auxquels sont associés les degrés de liberté de l'élément fini de Lagrange standard Q_k.

par des éléments qui sont des transformations d'un élément de référence (transformation bilinéaire dans le cas présent). En particulier chaque arête du maillage est l'image d'une arête de l'élément de référence. Maintenant, chaque arête étant en général commune à deux éléments voisins, et les degrés de liberté devant être partagés sur ces arêtes (pour respecter la conformité de l'élément fini) ; il faut, soit que la numérotation locale des sommets de chaque élément soit cohérente de manière à ce que chaque arête (orientée) du maillage, commune à deux éléments, soit l'image d'une unique arête de l'élément de référence, qu'elle soit vue comme arête de l'un ou l'autre élément (ce qui est sûrement possible sur un maillage régulier d'un domaine Ω rectangulaire par des éléments rectangulaires, mais qui s'avère impossible en toute généralité), soit que la localisation des points sur les arêtes auxquels on associe les degrés de liberté respecte les symétries décrites précédemment.

Nous illustrons ces propos par l'utilisation d'un élément fini de type Q_2 non standard (en ce sens que l'on a modifié la localisation des points auxquels on associe les degrés de liberté) à l'aide de la figure 2.3 : l'élément de référence (en noir) est transformé en deux éléments voisins du maillage (éléments en rouge et en bleu) de manière à ce que ces deux éléments aient une arête en commun sur laquelle les degrés de libertés sont partagés. Sachant que la restriction à cette arête de toute fonction de l'espace polynomial $\mathbb{Q}_2$ (à partir duquel nous avons construit cet élément fini) est un polynôme univarié de degré deux, nous disposons de trois valeurs prises par les degrés de liberté pour définir de manière unique ce polynôme. Or sur la figure de gauche il est clair qu'il est impossible d'assurer en toute généralité l'égalité deux à deux de trois couples de formes linéaires associées à deux éléments voisins, celles-ci étant l'évaluation de fonctions en des points qui ne sont pas confondus. Ceci entrant en conflit avec la condition de conformité, cet élément fini n'est utilisable dans la pratique que dans les cas très particuliers où le maillage permet une numérotation relative des sommets des éléments le composant cohérente (figure de droite). La force des éléments finis résidant dans sa souplesse d'utilisation sur des maillages non-structurés, tous les éléments finis de

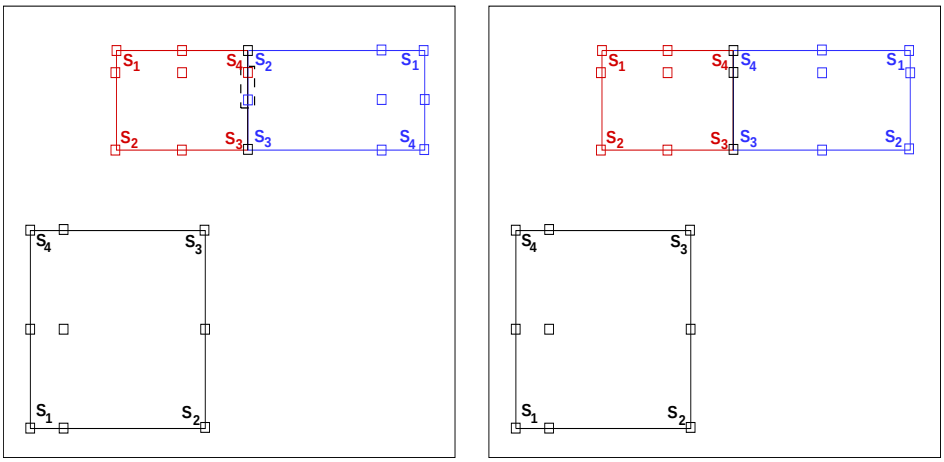

FIG. 2.3 – Localisation de points non symétrique pour un élément fini Q_2 non standard.

Lagrange standards respectent les contraintes de symétrie de localisation des points sur les arêtes auxquels sont associés les degrés de libertés, et la construction d'autres éléments finis de Lagrange passe, de la même façon, par le respect de ces conditions.

Terminons par les éléments finis de Lagrange triangulaires. Pour ces éléments finis en deux dimensions d'espace nous considérons $\hat{K}$ comme étant le triangle délimité par les sommets $(0,0)$, $(1,0)$ et $(0,1)$; $P = \mathbb{P}_k =< x^m y^n \ / \ 0 \leq m,n \leq k$ et $m+n \leq k >$ et Σ un ensemble de $\frac{(k+1)(k+2)}{2}$ formes linéaires qui sont les évaluations de fonctions aux intersections d'une grille régulière composée de $k+1$ lignes horizontales et $k+1$ lignes verticales (voir figure 2.4).

Remarquons que cette fois la restriction de toute fonction de $P = \mathbb{P}_k$ à n'importe quelle droite (et plus seulement aux droites horizontales et verticales) est un polynôme univarié de degré k. Aucun élément de type P_k ne sera unisolvant dès lors que strictement plus de $k+1$ points auxquels sont associés les degrés de liberté seront alignés.

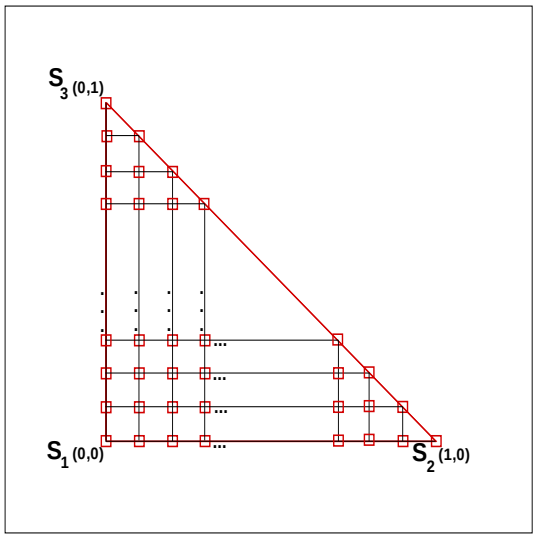

FIG. 2.4 – Localisation des points auxquels sont associés les degrés de liberté de l'élément fini de Lagrange standard P_k.

Chapitre 3

Éléments finis d'ordre arbitrairement élevé

3.1 Éléments finis adaptés à l'équation des ondes

Soit $\Omega \in \mathbb{R}^2$ un ouvert régulier. Nous allons considérer la formulation variationnelle de l'équation des ondes homogène soumise à des conditions aux limites de Dirichlet homogènes. Le problème que l'on se propose de résoudre est donc le suivant :

Trouver $u \in H_0^1(\Omega)$ tel que

$$\frac{d^2}{dt^2}\int_\Omega u\,\varphi\,dx + \int_\Omega \nabla u \cdot \nabla\varphi\,dx = 0 \qquad \forall \varphi \in H_0^1(\Omega).$$

En suivant la méthode de Ritz-Galerkin il nous faut introduire un sous-espace de dimension finie de $H_0^1(\Omega)$. Soit T_h une triangulation du domaine Ω. Notons $\{K_i, i = 1, \ldots, n\}$ l'ensemble des éléments de cette triangulation et

$$V_h = \{\psi \in C^0(\Omega) \ / \ \psi_{|K_i} \in \mathbb{P}_k \ \ \forall i = 1, \ldots, n \text{ et } \psi_{|\Gamma} = 0\}$$

le sous-espace de discrétisation subordonné à T_h associé à l'élément fini de Lagrange $(\hat{K}, \mathbb{P}_k, \Sigma)$ défini à la fin de la section (2.4). Le problème devient alors :
Trouver $u_h \in V_h$ tel que

$$\frac{d^2}{dt^2}\int_\Omega u_h\,v_h\,dx + \int_\Omega \nabla u_h \cdot \nabla v_h\,dx = 0 \qquad \forall v_h \in V_h.$$

3.1.1 Mise en oeuvre des éléments finis de Lagrange sur l'équation des ondes

Désignons par $\{a_i\}$ l'ensemble des points du domaine auxquels on associe les degrés de liberté et $\{\psi_i\}$ l'ensemble des fonctions de base associées. Si U_h désigne le vecteur dont les composantes sont les coordonnées de u_h dans la base $\{\psi_i\}$, alors le problème est équivalent au système d'équations différentielles ordinaires suivant :

$$\frac{d^2}{dt^2} MU_h(t) + KU_h(t) = 0,$$

où les matrices de masse et de raideur, respectivement M et K, sont définies par

$$\begin{cases} M_{ij} = \displaystyle\int_\Omega \psi_i(x)\psi_j(x)dx, \\ K_{ij} = \displaystyle\int_\Omega \nabla\psi_i(x) \cdot \nabla\psi_j(x)dx. \end{cases}$$

Dans la pratique il est préférable d'écrire que ces matrices valent

$$M_{ij} = \sum_{l=1}^{n} \int_{K_l} \psi_i(x,y)\psi_j(x,y)dxdy,$$

et

$$K_{ij} = \sum_{l=1}^{n} \int_{K_l} \nabla\psi_i(x,y) \cdot \nabla\psi_j(x,y)dxdy.$$

Ce petit jeu d'écriture a pour conséquence de ne plus calculer d'intégrale sur le domaine Ω tout entier mais sur chaque élément de la triangulation, sur lequel chaque fonction ψ_i est soit nulle soit entièrement déterminée en fonction d'une des fonctions $\hat{\psi}_{\hat{i}}$ associée à l'élément de référence $\hat{K}$ et de la transformation affine transportant cet élément vers l'élément K_l, de sorte que la matrice de masse M_{ij} est entièrement déterminée à partir d'une matrice de masse dite matrice élémentaire calculée sur l'élément de référence. Plus exactement considérons un élément K quelconque de sommet $S_i, i = 1..3$ ayant pour coordonnées (x_i, y_i) et l'élément de référence $\hat{K}$ qui a pour coordonnées respectives $(0,0)$, $(1,0)$ et $(0,1)$. L'application affine

$$\Phi(x,y) = A \begin{pmatrix} x \\ y \end{pmatrix} + B$$

avec

$$A = \begin{pmatrix} x_2 - x_1 & x_3 - x_1 \\ y_2 - y_1 & y_3 - y_1 \end{pmatrix} \quad , \quad B = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix},$$

transforme $\hat{K}$ en K. On a donc en effectuant le changement de variable $(\hat{x}, \hat{y}) = \Phi^{-1}(x,y)$

$$\begin{aligned} \int_K \psi_i(x,y)\psi_j(x,y)\ dxdy &= \int_K \hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y))\hat{\psi}_{\hat{j}}(\Phi^{-1}(x,y))\ dxdy \\ &= (\det A)\int_{\hat{K}} \hat{\psi}_{\hat{i}}(\hat{x},\hat{y})\hat{\psi}_{\hat{j}}(\hat{x},\hat{y})\ d\hat{x}d\hat{y}. \end{aligned}$$

D'autre part il nous faut calculer

$$\int_K \nabla\psi_i(x,y) \cdot \nabla\psi_j(x,y)\ dxdy = \int_K \nabla(\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y))) \cdot \nabla(\hat{\psi}_{\hat{j}}(\Phi^{-1}(x,y)))\ dxdy. \qquad (3.1)$$

Pour cela nous déterminons

$$\begin{aligned}\nabla(\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y))) &= \begin{pmatrix} \partial_x(\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y))) \\ \partial_y(\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y))) \end{pmatrix} \\ &= \begin{pmatrix} \partial_x\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y))\partial_x\Phi_1^{-1}(x,y) + \partial_y\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y))\partial_x\Phi_2^{-1}(x,y) \\ \partial_x\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y))\partial_y\Phi_1^{-1}(x,y) + \partial_y\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y))\partial_y\Phi_2^{-1}(x,y) \end{pmatrix}\end{aligned}$$

où

$$\Phi^{-1}(x,y) = \begin{pmatrix} \Phi_1^{-1}(x,y) \\ \Phi_2^{-1}(x,y) \end{pmatrix} = \frac{1}{\det A}\begin{pmatrix} (y_3-y_1)(x-x_1) - (x_3-x_1)(y-y_1) \\ -(y_2-y_1)(x-x_1) + (x_2-x_1)(y-y_1) \end{pmatrix},$$

ce qui donne après calculs

$$\begin{aligned}\nabla(\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y))) &= \frac{1}{\det A}\begin{pmatrix} (y_3-y_1)\partial_x\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y)) - (y_2-y_1)\partial_y\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y)) \\ -(x_3-x_1)\partial_x\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y)) + (x_2-x_1)\partial_y\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y)) \end{pmatrix} \\ &= \frac{1}{\det A}({}^tA)^{-1}\nabla\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y)).\end{aligned}$$

En injectant ceci dans (3.1) nous obtenons

$$\begin{aligned}&\int_K \nabla\psi_i(x,y)\cdot\nabla\psi_j(x,y)\,dxdy \\ &\qquad = \frac{1}{(\det A)^2}\int_K ({}^tA)^{-1}\nabla\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y))\cdot({}^tA)^{-1}\nabla\hat{\psi}_{\hat{i}}(\Phi^{-1}(x,y))\,dxdy \\ &\qquad = \frac{1}{\det A}\int_{\hat{K}} ({}^tA)^{-1}\nabla\hat{\psi}_{\hat{i}}(\hat{x},\hat{y})\cdot({}^tA)^{-1}\nabla\hat{\psi}_{\hat{i}}(\hat{x},\hat{y})\,d\hat{x}d\hat{y},\end{aligned}$$

puis finalement en développant

$$\begin{aligned}&\int_K \nabla\psi_i(x,y).\nabla\psi_j(x,y)dxdy = \\ &\qquad \frac{1}{\det A}\{((y_3-y_1)^2+(x_1-x_3)^2)\int_{\hat{K}}\partial_{\hat{x}}\hat{\psi}_{\hat{i}}\partial_{\hat{x}}\hat{\psi}_{\hat{j}}\,d\hat{x}d\hat{y} \\ &\qquad\quad +((y_1-y_2)^2+(x_2-x_1)^2)\int_{\hat{K}}\partial_{\hat{y}}\hat{\psi}_{\hat{i}}\partial_{\hat{y}}\hat{\psi}_{\hat{j}}\,d\hat{x}d\hat{y} \\ &\qquad +((y_3-y_1)(y_1-y_2)+(x_1-x_3)(x_2-x_1))\int_{\hat{K}}\partial_{\hat{x}}\hat{\psi}_{\hat{i}}\partial_{\hat{y}}\hat{\psi}_{\hat{j}} + \partial_{\hat{y}}\hat{\psi}_{\hat{i}}\partial_{\hat{x}}\hat{\psi}_{\hat{j}}\,d\hat{x}d\hat{y}\}.\end{aligned}$$

On remarque alors que le calcul des matrices M et K se fait à partir d'un certain nombre de matrices élémentaires calculées sur l'élément de référence, à savoir

$$\int_{\hat{K}}\hat{\psi}_i(\hat{x},\hat{y})\hat{\psi}_j(\hat{x},\hat{y})d\hat{x}d\hat{y},$$

$$\int_{\hat{K}}\partial_{\hat{x}}\hat{\psi}_i\partial_{\hat{x}}\hat{\psi}_j d\hat{x}d\hat{y},$$

$$\int_{\hat{K}}\partial_{\hat{y}}\hat{\psi}_i\partial_{\hat{y}}\hat{\psi}_j d\hat{x}d\hat{y},$$

$$\int_{\hat{K}}\partial_{\hat{x}}\hat{\psi}_i\partial_{\hat{y}}\hat{\psi}_j + \partial_{\hat{y}}\hat{\psi}_i\partial_{\hat{x}}\hat{\psi}_j d\hat{x}d\hat{y}.$$

3.1.2 Génération automatique des matrices de masse et de raideur

La question que l'on se pose est comment générer automatiquement les matrices qui sont nécessaires à l'assemblage des matrices élémentaires globales. Il suffit alors de remarquer que pour calculer les matrices dont nous avons besoin la seule connaissance des fonctions de base $\{\hat{\psi}_i\}$ sur l'élément de référence $\hat{K}$ nous suffit. Une fois ces fonctions connues, les matrices dont nous avons besoin le seront aussi à l'aide de n'importe quel logiciel de calcul formel (Maple© par exemple [74]). Il s'avère qu'il est possible d'expliciter ces fonctions dans le cas des éléments finis de Lagrange triangulaires standards. Pour cela il nous faut dans un premier temps se fixer une numérotation des points auxquels sont associés les degrés de liberté. Soit $\{S(i) \in \mathbb{R}^2\}_{i=1\ldots3}$ l'ensemble des coordonnées des trois sommets du triangle de référence, auquel on ajoute les points $S(0) = S(3)$ et $S(4) = S(1)$. En notant $\{a_i\}_{i=1\ldots\frac{(k+1)(k+2)}{2}}$ l'ensemble des points auxquels on associe les degrés de liberté, et $\tilde{k}$ la partie entière de $\frac{k}{3}$, nous avons que :

pour m de 0 à $\tilde{k}$,
  pour j de 0 à $k-(3m+1)$,
    pour i de 1 à 3,

le $\xi^{\text{ième}}$ de ces points, en suivant une numérotation qui n'a rien de standard mais qui apparaît naturellement en localisant les points en terme de coordonnées barycentriques (voir figure 3.1), a pour coordonnée :

$$a_\xi = \frac{mS(i-1) + (k-2m-j)S(i) + (j+m)S(i+1)}{k}$$

où l'indice ξ est lui-même une fonction de (k,m,j,i) donnée par

$$\xi(k,m,j,i) = \left(3k - \frac{9(m-1)}{2}\right) m + 3j + i.$$

Les $\frac{(k+1)(k+2)}{2}$ fonctions de base sont alors définies suivant les mêmes modalités par

$$\Psi_\xi(\Lambda_1,\Lambda_2,\Lambda_3) = \frac{\psi_\xi(\Lambda_1,\Lambda_2,\Lambda_3)}{\psi_\xi(\Lambda_1(a_\xi),\Lambda_2(a_\xi),\Lambda_3(a_\xi))},$$

où

$$\psi_\xi(\Lambda_1,\Lambda_2,\Lambda_3) = \prod_{l=0}^{m-1}\prod_{n=1}^{3}\left(\Lambda_n - \frac{l}{k}\right)\prod_{l=m}^{k-2m-(j+1)}\left(\Lambda_i - \frac{l}{k}\right)\prod_{l=m}^{j+m-1}\left(\Lambda_{i+1} - \frac{l}{k}\right)$$

et $\{\Lambda_i\}_{i=1\ldots3}$ désigne l'unique ensemble de polynômes de $\mathbb{P}_1$ vérifiant $\Lambda_i(S(j)) = \delta_{ij}\ \forall i,j = i = 1\ldots3$ (c'est aussi en particulier la base associée à l'élément fini P_1).

Remarque 3.1.1. *Si $k \equiv 0[3]$, il ne faut considérer que m de 0 à $\tilde{k}$ et définir les coordonnées du dernier point, qui n'est autre que le centre de gravité du triangle par*

$$a_{\frac{(k+1)(k+2)}{2}} = \frac{S(1)+S(2)+S(3)}{3},$$

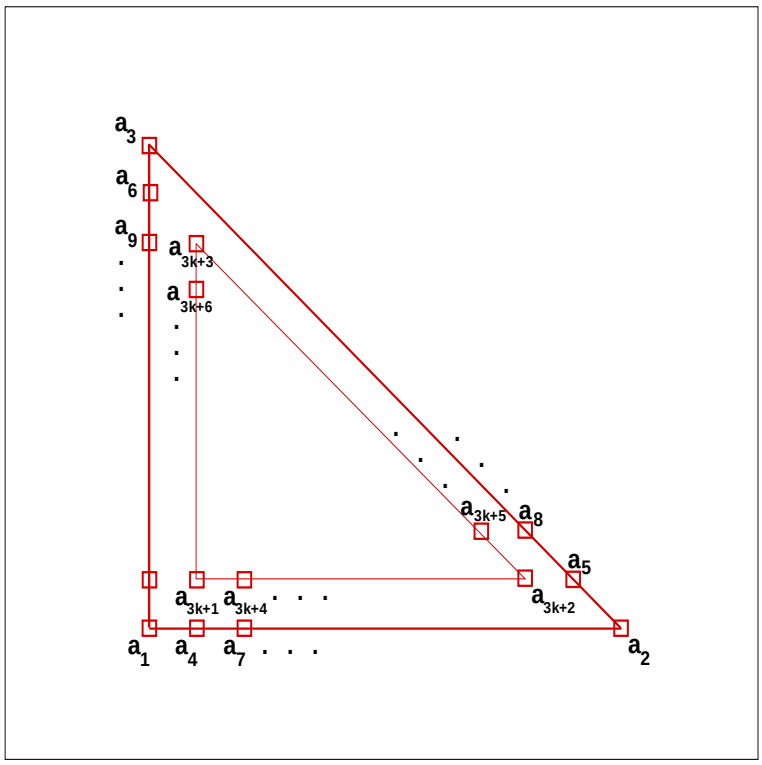

FIG. 3.1 – Numérotation des noeuds de l'élément fini P_k standard.

et la dernière fonction de base qui lui est associée par

$$\Psi_{\frac{(k+1)(k+2)}{2}}(\Lambda_1,\Lambda_2,\Lambda_3)=\frac{\prod_{l=0}^{\tilde{k}-1}\prod_{n=1}^{3}\left(\Lambda_n-\frac{l}{k}\right)}{\left(\prod_{l=1}^{\tilde{k}}\left(\frac{l}{k}\right)\right)^3}.$$

On a alors effectivement explicité les $\frac{(k+1)(k+2)}{2}$ fonctions de base de l'espace $\mathbb{P}_k$ associé à l'élément fini P_k standard. La procédure Maple© générant ces fonctions de base et les matrices élémentaires est donnée en Annexe 7.5.7.
Dans le cas où l'on souhaite utiliser des éléments finis de Lagrange P_k non standards, c'est à dire si l'on souhaite modifier la localisation des points auxquels sont associées les formes linéaires, il faudra résoudre au cas par cas les $\frac{(k+1)(k+2)}{2}$ systèmes linéaires définissant cette base. Ces systèmes, seront bien entendu résolus par une procédure Maple©, qui pour un espace fonctionnel et un jeu de points donnés renvoie les fonctions de base dont les coefficients seront écrits dans un fichier lu par le code de calcul.

3.1.3 Sur l'influence de la localisation des points auxquels sont associées les formes linéaires

Dans cette sous-section nous abordons un problème qui apparaît lorsque l'on veut augmenter l'ordre des discrétisations en espace par éléments finis. Dans la sous-section précédente (3.1.2) nous avons explicité les fonctions de base associées à des degrés de liberté lagrangiens sur des points équirépartis. Typiquement, si une telle répartition de points est viable à des ordres bas, cela peut mener à des phénomènes indésirables une fois que l'on monte en ordre : d'une part la projection sur le sous-espace de discrétisation sera de mauvaise qualité, d'autre part la matrice de masse sera de plus en plus mal conditionnée.

Les problèmes de qualité de projection sont bien connus dans le cadre de l'interpolation de Lagrange (puisque la projection sur un sous-espace d'éléments finis de Lagrange correspond à une interpolation de Lagrange localement sur chaque élément du maillage aux points auxquels sont associés les degrés de liberté) sous les termes de phénomène de Runge : par exemple en une dimension d'espace, approcher une fonction par son interpolée de degré N en $N+1$ points équirépartis, peut se révéler être une très mauvaise idée, c'est-à-dire que l'interpolée sera une mauvaise approximation de la fonction, si les variations de gradient de la fonction sont trop élevées. La figure 3.2 représente la gaussienne centrée normalisée et son interpolée de degré dix sur les onze entiers équirépartis dans $[-5,5]$.

Nous remarquons que l'interpolée a une fâcheuse tendance à osciller aux bords de l'intervalle. Cela est dû au fait que les fonctions de la base lagrangienne sont elles-mêmes fortement oscillantes lorsque les points d'interpolation sont équirépartis. Ce qui nous amène au second problème : la matrice de masse, c'est-à-dire la matrice constituée des intégrales des produits de fonctions de base, sera mal conditionnée.

Pour formaliser tout ceci il est bon d'introduire un certain nombre de notations. Considérons $(\hat{K},\mathbb{P}_k,\Sigma)$ l'élément fini de Lagrange standard, f une fonction continue sur $\hat{K}$ et $\tilde{p}$

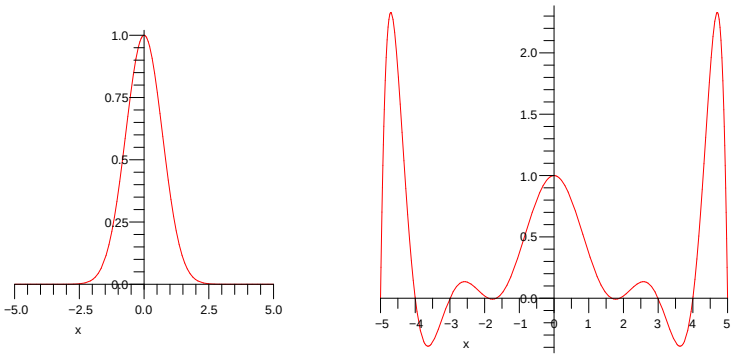

FIG. 3.2 – La gaussienne centrée normalisée et son interpolée sur les onze entier équirépartis dans $[-5, 5]$.

la meilleure approximation de f en norme $L^\infty(\hat{K})$, c'est-à-dire la fonction de $\mathbb{P}_k$ réalisant le minimum

$$\min\{\|f(x) - p(x)\|_\infty \ , p \in \mathbb{P}_k\},$$

qui existe dans la mesure où f est continue sur $\hat{K}$ (voir [33]), même si sa détermination est un problème encore ouvert. Notons Π la projection naturellement associée à l'élément fini sur l'espace polynomial $\mathbb{P}_k$:

$$\Pi(f) = \sum_{i=1}^{N} \sigma_i(f)\varphi_i$$

où N désigne la dimension de $\mathbb{P}_k$ et $\{\varphi_1, \ldots, \varphi_N\}$ la base lagrangienne associée aux formes linéaires de Σ. En définissant finalement

$$\|\Pi\|_\infty = \sup_{f \neq 0} \frac{\|\Pi(f)\|_\infty}{\|f\|_\infty},$$

il est possible de montrer le lemme

Lemme 3.1.2.

$$\|\Pi\|_\infty = \max_{x \in \hat{K}} \sum_{i=1}^{N} |\varphi_i(x)|$$

et le théorème suivant, dû à Lebesgue nous dit que

Théoreme 3.1.3. *Sous l'hypothèse de continuité de f sur $\hat{K}$,*

$$\|f(x) - \Pi(f)(x)\|_\infty \leq (1 + \Delta(\Pi))\|f(x) - \tilde{p}(x)\|_\infty$$

où

$$\Delta(\Pi) = \|\Pi\|_\infty = \max_{x \in \hat{K}} \sum_{i=1}^{N} |\varphi_i(x)|$$

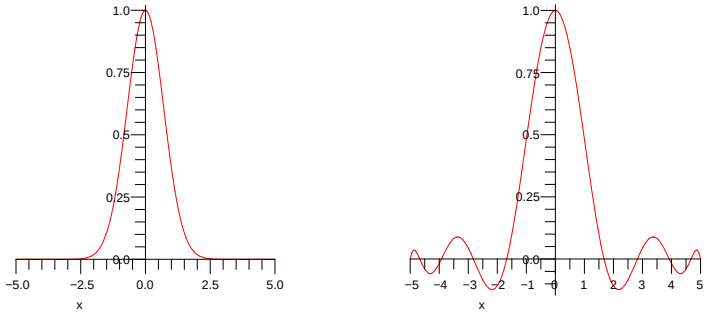

FIG. 3.3 – La gaussienne centrée normalisée et son interpolée sur les onze points de quadrature de la formule de Gauss-Lobatto rapportés à $[-5, 5]$.

est appelée la constante de Lebesgue.

La projection sur le sous-espace d'élément fini est donc d'autant meilleure que la constante de Lebesgue, qui ne dépend que de la localisation des points auxquels sont associées les formes linéaires, est petite.

Pour améliorer l'interpolation il est bon de ne pas utiliser des points équirépartis mais d'utiliser une localisation de ces points aux points de quadrature de formules de quadrature de type Newton-Cotes. Ceci n'est bien évidemment pas dû au hasard puisque les formules de quadrature de type Newton-Cotes sont basées sur l'idée que pour approcher l'intégrale d'une fonction il suffit d'intégrer l'interpolée de Lagrange de cette fonction, du moment que cette interpolée est une bonne approximation de la fonction. La localisation des points de quadrature de ce type de formule est donc optimisée de manière à minimiser l'erreur de projection, ou en d'autres termes de minimiser la constante de Lebesgue associée à ces points. La figure 3.3 représente l'interpolé de degré dix de la même gaussienne, l'interpolation se faisant cette fois aux points de quadrature de la formule de quadrature de Gauss-Lobatto (voir annexe 7.5.7 pour plus de détails sur ces formules de quadrature).

Le problème de la localisation des points auxquels sont associées les formes linéaires, qui est mis en évidence ici en une dimension d'espace, reste vrai en deux dimensions d'espace et est d'autant plus compliqué pour l'interpolation sur des triangles, géométrie pour laquelle nous ne disposons pas d'une théorie aussi aboutie pour la construction de formules de quadrature que pour les segments en une dimension d'espace ou tout produit tensoriel de ce type de domaine en dimension supérieure. Nous renvoyons le lecteur aux travaux de P. Silvester [65], S. Wandzurat et H. Xiao [71] ou plus particulièrement de J. S. Hesthaven [46] dont nous utilisons la localisation de points représentés dans les figures 3.4 à 3.6.

Nous insistons sur le fait que, même si la relocalisation des points n'est quasiment pas visible jusqu'aux éléments finis P_5, celle-ci le deviendra de plus en plus à mesure que l'on

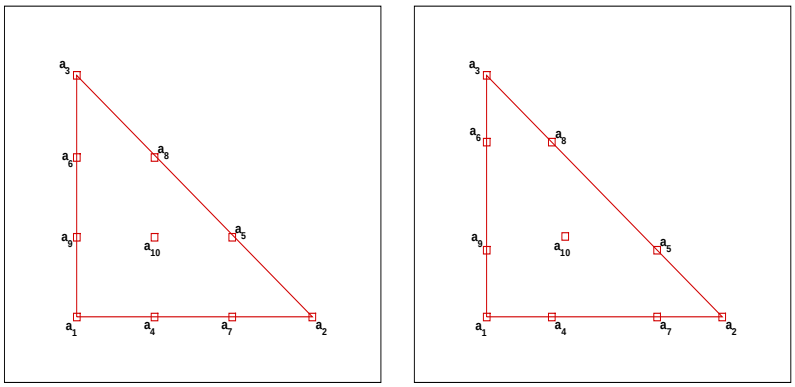

FIG. 3.4 – Relocalisation des points associés à l'élément P_3.

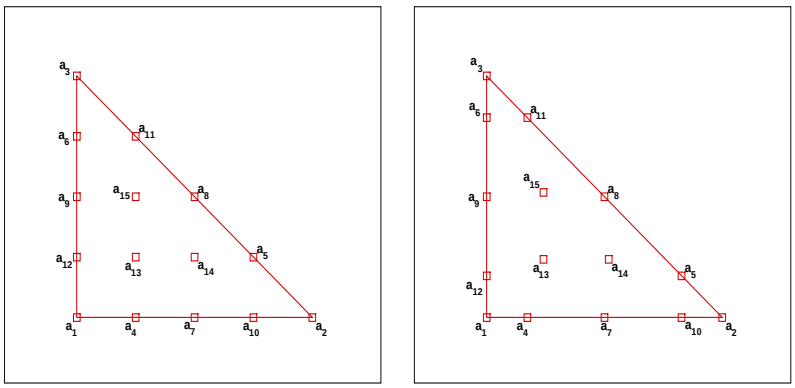

FIG. 3.5 – Relocalisation des points associés à l'élément P_4.

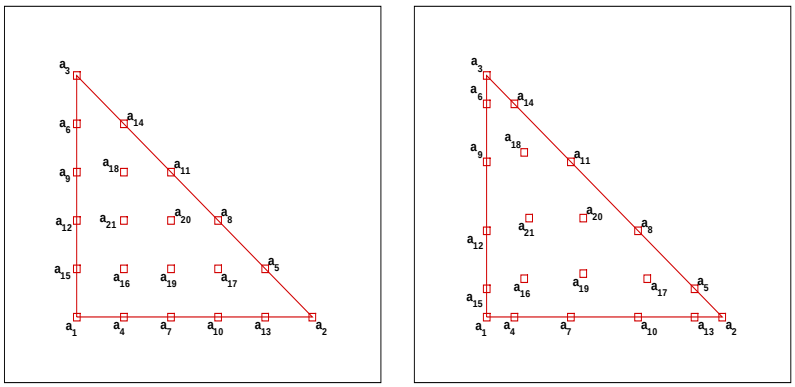

FIG. 3.6 – Relocalisation des points associés à l'élément P_5.

localisation équirépartie	P_3	P_4	P_5	P_6	P_7
conditionnement	67.66	114.41	239.44	548.00	1340.77
constante de Lebesgue	2.27	3.47	5.45	8.7	14.3
localisation optimisée	P_3	P_4	P_5	P_6	P_7
conditionnement	68.73	101.32	134.77	218.73	329.31
constante de Lebesgue	2.12	2.63	3.19	4.06	4.75

TAB. 3.1 – Conditionnement en norme L^2 de la matrice de masse et constante de Lebesgue calculés sur l'élément de référence pour les éléments finis P_3 à P_7.

augmente l'ordre des éléments et qu'à ce titre, le problème de la localisation des points auxquels seront associés les degrés de liberté n'aura plus rien d'anecdotique.

À titre d'exemple, pour illustrer les problèmes de projection et de conditionnement de la matrice de masse issus d'une mauvaise localisation des points auxquels sont associés les degrés de liberté, nous donnons dans le tableau 3.1 la constante de Lebesgue et le conditionnement en norme L^2 de la matrice de masse calculés sur l'élément de référence pour les éléments finis P_3 à P_7 pour la localisation équirépartie et la localisation proposée par J. S. Hesthaven dans [46].

Il faut remarquer non seulement que la relocalisation des points permet de diminuer ces constantes, mais aussi que le rapport entre ces constantes croît fortement (c'est-à-dire que la relocalisation montrera d'autant plus son intérêt que l'ordre des schémas sera élevé).

Remarque 3.1.4. *Dans la pratique nous n'avons pas sensiblement ressenti le gain de la relocalisaton des points auxquels sont associés les degrés de liberté jusqu'aux schémas à discrétisation en espace par éléments finis de Lagrange d'ordre 6. Mais il ne fait aucun doute que pour des discrétisations d'ordre plus élevé la question de la localisation des points influera plus fortement sur l'efficacité des schémas.*

3.2 Éléments finis adaptés à la propagation d'ondes électromagnétiques

Nous avons vu dans la section (1.2) qu'il nous faut construire des espaces X, W et V vérifiant

$$\begin{array}{ccccc} & \overrightarrow{\nabla} & & \nabla\times & \\ H^1(\Omega) & \longrightarrow & H(rot,\Omega) & \longrightarrow & L^2(\Omega) \\ \cup & & \cup & & \cup \\ X & \longrightarrow & W & \longrightarrow & V \end{array} . \tag{3.2}$$

Nous allons supposer pour l'instant de tels espaces construits.
Soit $\{\overrightarrow{\psi}_i\}_{i=1\ldots N}$ une base de $W \subset H(rot,\Omega)$ et $\{\varphi_k\}_{k=1\ldots M}$ une base de $V \subset L^2(\Omega)$. En réécrivant les équations (1.13) et (1.14) notre but est maintenant de résoudre le problème suivant :

Trouver $(\overrightarrow{E}, B) \in W \times V$ vérifiant

$$\left\{ \begin{array}{ll} \dfrac{d}{dt}\displaystyle\int_\Omega \overrightarrow{E}\cdot\overrightarrow{\psi_i}\,dX - \int_\Omega B(\nabla\times\overrightarrow{\psi_i})\,dX &= -\displaystyle\int_\Omega \overrightarrow{J}\cdot\overrightarrow{\psi_i}\,dX, \quad \forall i = 1\dots N, \\ \dfrac{d}{dt}\displaystyle\int_\Omega B\varphi_k\,dX + \int_\Omega (\nabla\times\overrightarrow{E})\varphi_k\,dX &= 0, \qquad\qquad\qquad\qquad \forall k = 1\dots M, \end{array} \right. \tag{3.3}$$

ce qui se traduit, une fois $\overrightarrow{E}$ et B décomposés sur les bases respectives de W et V en $\overrightarrow{E} = \displaystyle\sum_{j=1}^{N} e_j \overrightarrow{\psi}_j$ et $B = \displaystyle\sum_{l=1}^{M} b_l\varphi_l$, par :
Trouver $(e_1,\dots,e_N)$ et $(b_1,\dots,b_M)$ vérifiant

$$\left\{ \begin{array}{ll} \dfrac{d}{dt}\displaystyle\sum_{j=1}^{N} e_j \int_\Omega \overrightarrow{\psi_j}\cdot\overrightarrow{\psi_i}\,dX - \sum_{l=1}^{M} b_l \int_\Omega \varphi_l(\nabla\times\overrightarrow{\psi_i})\,dX &= -\displaystyle\int_\Omega \overrightarrow{J}\cdot\overrightarrow{\psi_i}\,dX, \quad \forall i = 1\dots N, \\ \dfrac{d}{dt}\displaystyle\sum_{l=1}^{M} b_l \int_\Omega \varphi_l\varphi_k\,dX + \sum_{j=1}^{N} e_j \int_\Omega (\nabla\times\overrightarrow{\psi_j})\varphi_k\,dX &= 0, \qquad\qquad \forall k = 1\dots M. \end{array} \right. \tag{3.4}$$

ou encore sous forme matricielle :

$$\left\{ \begin{array}{ll} M_w\dot{E} - KB &= \tilde{J} \\ M_v\dot{B} + {}^tKE &= 0 \end{array} \right. \tag{3.5}$$

avec

$$\begin{aligned} (M_w)_{1\le i,j\le N} &= \int_\Omega \overrightarrow{\psi_j}\cdot\overrightarrow{\psi_i}\,dX, \\ (M_v)_{1\le i,j\le M} &= \int_\Omega \varphi_i\varphi_j\,dX, \\ (K)_{\substack{1\le i\le N \\ 1\le j\le M}} &= \int_\Omega \varphi_j(\nabla\times\overrightarrow{\psi_i})\,dX. \end{aligned}$$

Remarque 3.2.1. *Il est important de remarquer que si $\{\overrightarrow{\psi_i}\}_{i=1,\dots,N}$ est une base d'élément fini, c'est-à-dire une base uniquement déterminée par N formes linéaires $\{\sigma_i^W\}_{i=1,\dots,N}$ et la relation $\sigma_i(\overrightarrow{\psi_i}) = \delta_{ij}$; alors dans la décomposition $\overrightarrow{E} = \sum_{j=1}^{N} e_j\overrightarrow{\psi}_j$ de E sur la base $\{\overrightarrow{\psi_i}\}_{i=1,\dots,N}$, les e_j ne sont autres que les $\sigma_j^W(\overrightarrow{E})$. De la même manière si $\{\varphi_k\}_{k=1,\dots,M}$ est la base associée à M formes linéaires $\{\sigma_k^V\}_{k=1,\dots,M}$, alors dans la décomposition $B = \sum_{l=1}^{M} b_l\varphi_l$ de B sur cette base, on a clairement $b_l = \sigma_l^V(B)\ \ \forall l = 1,\dots,M$.*

Dans la formulation variationnelle de l'équation de Faraday (c'est-à-dire la deuxième équation du système (3.3)), il apparaît un $\nabla\times\overrightarrow{E}$ que l'on décompose naturellement dans le système (3.4) sur les rotationnels de la base $\{\overrightarrow{\psi_i}\}_{i=1,\dots,N}$, c'est-à-dire par $\nabla\times\overrightarrow{E} = \sum_{j=1}^{N} e_j(\nabla\times\overrightarrow{\psi_j})$. Or si les propriétés de suite exacte décrites plus haut sont vérifiées, on a en particulier que $\nabla\times\overrightarrow{E} \in V$, et l'on peut donc décomposer le rotationnel de $\overrightarrow{E}$ sur la

base $\{\varphi_k\}_{k=1,\dots,M}$ de V. Ceci nous donne :

$$\begin{aligned}
\nabla \times \overrightarrow{E} &= \sum_{l=1}^{M} \sigma_l^V (\nabla \times \overrightarrow{E}) \varphi_l \\
&= \sum_{l=1}^{M} \sigma_l^V (\sum_{j=1}^{N} e_j (\nabla \times \overrightarrow{\psi_j})) \varphi_l \\
&= \sum_{l=1}^{M} \sigma_l^V (\sum_{j=1}^{N} \sigma_j^W (\overrightarrow{E}) (\nabla \times \overrightarrow{\psi_j})) \varphi_l \\
&= \sum_{l=1}^{M} \sum_{j=1}^{N} \sigma_l^V (\sigma_j^W (\overrightarrow{E}) (\nabla \times \overrightarrow{\psi_j})) \varphi_l \\
&= \sum_{l=1}^{M} \sum_{j=1}^{N} \sigma_j^W (\overrightarrow{E}) \sigma_l^V (\nabla \times \overrightarrow{\psi_j}) \varphi_l .
\end{aligned}$$

En particulier nous obtenons que $\sigma_l^V(\nabla \times \overrightarrow{E}) = \sum_{j=1}^{N} \sigma_j^W(\overrightarrow{E}) \sigma_l^V(\nabla \times \overrightarrow{\psi_j})$, puis en réécrivant le terme concerné dans l'équation de Faraday (1.14)

$$\begin{aligned}
\int_\Omega (\nabla \times \overrightarrow{E}) \varphi_k \, dX &= \int_\Omega \sum_{l=1}^{M} \sum_{j=1}^{N} \sigma_j^W (\overrightarrow{E}) \sigma_l^V (\nabla \times \overrightarrow{\psi_j}) \varphi_l \varphi_k \, dX \\
&= \sum_{l=1}^{M} \sum_{j=1}^{N} \int_\Omega \varphi_l \varphi_k \, dX \; \sigma_l^V (\nabla \times \overrightarrow{\psi_j}) \; \sigma_j^W (\overrightarrow{E}).
\end{aligned}$$

Cette dernière quantité est alors vue comme la $k^{\text{ième}}$ ligne (pour k de 1 à M) du vecteur

$$M_v R E$$

où R est la matrice définie par

$$(R)_{\substack{1 \le i \le M \\ 1 \le j \le N}} = \sigma_i^V (\nabla \times \overrightarrow{\psi_j}),$$

de sorte que le système (3.5) se réécrive de manière totalement équivalente sous la forme suivante :

$$\left\{ \begin{array}{rl} M_w \dot{E} - KB &= \tilde{J} \\ \dot{B} + RE &= 0 \end{array} \right. . \tag{3.6}$$

L'intérêt d'une telle manipulation tient bien entendu dans le fait que la discrétisation de la forme variationnelle de l'équation de Faraday est maintenant explicite, c'est-à-dire que le calcul de la dérivée de B ne nécessite plus l'inversion de la matrice M_v.

Remarque 3.2.2. *Si nous avions choisi d'utiliser la formulation variationnelle alternative proposée dans la remarque 1.2.1, consistant à intégrer par partie l'équation de Faraday (1.6), plutôt que l'équation d'Ampère (1.5), et d'ainsi obtenir le système suivant :*

$$\left\{ \begin{array}{rll} \dfrac{d}{dt} \displaystyle\int_\Omega \overrightarrow{E} \cdot \overrightarrow{\psi_i} \, dX - \int_\Omega (\overrightarrow{\nabla} \times B) \cdot \overrightarrow{\psi_i} \, dX &= - \displaystyle\int_\Omega \overrightarrow{J} \cdot \overrightarrow{\psi_i} \, dX, & \forall i = 1 \dots N, \\ \dfrac{d}{dt} \displaystyle\int_\Omega B \varphi_k \, dX + \int_\Omega \overrightarrow{E} \cdot (\overrightarrow{\nabla} \times \varphi_k) \, dX &= 0, & \forall k = 1 \dots M, \end{array} \right. \tag{3.7}$$

où $\{\overrightarrow{\psi}_i\}_{i=1...N}$ *est une base de* $W \subset H(rot,\Omega)$ *et* $\{\varphi_k\}_{k=1...M}$ *est une base de* $V \subset H^1(\Omega)$, *nous aurions pu faire la même manipulation pour exprimer cette fois* $\overrightarrow{\nabla} \times B$ *sur la base* $\{\overrightarrow{\psi}_i\}_{i=1...N}$ *pour obtenir un système du type*

$$\left\{ \begin{array}{rl} \dot{E} - \tilde{R}B & = \tilde{J} \\ M_v\dot{B} + \tilde{K}E & = 0 \end{array} \right. , \tag{3.8}$$

avec

$$(\tilde{K})_{\substack{1 \leq i \leq M \\ 1 \leq j \leq N}} = \int_\Omega \overrightarrow{\psi_i} \cdot (\overrightarrow{\nabla} \times \varphi_j)\, dX$$

et

$$(\tilde{R})_{\substack{1 \leq i \leq M \\ 1 \leq j \leq N}} = \sigma_j^W(\overrightarrow{\nabla} \times \varphi_i).$$

En anticipant les problèmes de condensation de masse que nous allons aborder dans le chapitre 4, il est bon de se demander laquelle de ces deux approches alternatives est la plus efficace. En effet il faut remarquer que dans la deuxième approche, l'équation d'Ampère est explicite, c'est-à-dire que la résolution des équations du système (3.8) ne nécessite plus d'inversion de la matrice de masse M_w *issue de l'utilisation d'éléments finis conformes dans* $H(rot,\Omega)$. *L'utilisation d'éléments finis de Lagrange triangulaires nous permettant de condenser la matrice de masse* M_v *(jusqu'aux éléments finis d'ordre 6, tout du moins...), le système (3.8) devient entièrement explicite. Nous avons toutefois préféré la première approche dans la mesure où la résolution des équations de Maxwell en trois dimensions d'espace nous demandera de toute façon une réflexion sur l'optimisation de l'inversion de la matrice de masse issue de l'utilisation d'éléments finis d'arêtes. Cette optimisation a été faite ici par condensation des éléments finis rectangulaires sur maillage cartésien (éléments finis que nous avons développés et dont la généralisation en trois dimensions d'espace est immédiate) et couplage conforme avec des éléments finis triangulaires dans le cadre de domaines à géométrie complexe (voir sous-section 4.5.1 et sous-section 3.2.3).*

3.2.1 Discrétisation conforme dans le cas d'éléments finis d'arête rectangulaires

Il nous faut à présent construire des espaces de discrétisation conformes, c'est-à-dire construire $X \subset H^1(\Omega)$, $W \subset H(rot,\Omega)$ et $V \subset L^2(\Omega)$. Bien entendu, cette construction passe par un maillage du domaine Ω. Nous ne considérons pour l'instant que des domaines rectangulaires que l'on quadrille par des lignes horizontales et verticales équiréparties définissant un maillage régulier de Ω par un ensemble de rectangles $\{K_i\}_{i=1,...,r}$. Nous introduisons alors les espaces fonctionnels suivants :

$$X = \{\xi \in H^1(\Omega) \ | \ \xi_{|K_i} \in \mathbb{Q}_k(K_i)\ ,\ \forall i = 1,\ldots,r\},$$

$$W = \{\overrightarrow{\psi} \in H(rot,\Omega) \ | \ \overrightarrow{\psi}_{|K_i} \in \left(\begin{array}{c} \mathbb{Q}_{k-1,k}(K_i) \\ \mathbb{Q}_{k,k-1}(K_i) \end{array} \right)\ ,\ \forall i = 1,\ldots,r\}$$

où

$$\mathbb{Q}_{m,n} = < x^i y^j \ / \ 0 \leq i \leq m, 0 \leq j \leq n >,$$

$$V = \{\varphi \in L^2(\Omega) \ | \ \varphi_{|K_i} \in \mathbb{Q}_{k-1}(K_i)\ ,\ \forall i = 1,\ldots,r\}.$$

Le choix de ces espaces n'est pas trivial et a été orienté non seulement par notre volonté d'utiliser des éléments finis conformes (d'où la forme très particulière de l'espace W), mais aussi par notre volonté de respecter d'un point de vue discret les propriétés de suite exacte des espaces continus.
Les propriétés suivantes se démontrent de manière immédiate :

$$\overrightarrow{\nabla} X \subset W, \quad \nabla \times (\overrightarrow{\nabla} X) = \{0\}, \quad \nabla \times W \subset V,$$

ce qui suffit pour affirmer que la suite

$$\begin{array}{ccccc} & & \overrightarrow{\nabla} & & \nabla\times \\ X & \longrightarrow & W & \longrightarrow & V, \end{array}$$

est exacte.
S'il est notable qu'une fonction qui est polynomiale par morceaux est déjà dans $L^2(\Omega)$, notre attention est portée sur le fait que tout comme dans $H^1(\Omega)$ il ne suffit pas d'être polynomial par morceaux pour être dans $H(rot, \Omega)$. En effet le théorème suivant, dont la démonstration se trouve dans [55] par exemple, nous dit que :

Théoreme 3.2.3. *(Recollement dans H(rot)) Soit Ω_1 et Ω_2 deux ouverts disjoints de $\mathbb{R}^2$. On note $\Gamma = \overline{\Omega_1} \cap \overline{\Omega_2}$, et $\overrightarrow{\tau}$ un vecteur unitaire tangent à Γ. Soit $\overrightarrow{E_1} \in H(rot, \Omega_1)$ et $\overrightarrow{E_2} \in H(rot, \Omega_2)$. Si $\overrightarrow{E_1}.\overrightarrow{\tau} = \overrightarrow{E_2}.\overrightarrow{\tau}$ sur Γ alors $\overrightarrow{E_1}\chi_{\Omega_1} + \overrightarrow{E_2}\chi_{\Omega_2} \in H(rot, \Omega_1 \cup \Omega_2)$.*

Remarque 3.2.4. *Si le théorème précédent ne nous donne qu'une condition suffisante pour que le recollement de deux fonctions soit conforme dans $H(rot, \Omega)$, celle-ci devient nécessaire dès lors que l'on ne considère plus en toute généralité des fonctions qui sont $H(rot)$ par morceaux mais des fonctions polynomiales par morceaux.*

À nouveau c'est la définition même de l'élément fini qui doit assurer la continuité de la trace tangentielle sur chaque arête du maillage.
Soit donc $(\hat{K}, P, \Sigma)$ où

(i) $\hat{K} = [0,1]^2$ le carré unité,

(ii) $P = \begin{pmatrix} \mathbb{Q}_{k-1,k}(\hat{K}) \\ \mathbb{Q}_{k,k-1}(\hat{K}) \end{pmatrix}$, espace polynomial de dimension $2k(k+1)$,

(iii) Σ est un ensemble de formes linéaires sur P de cardinal fini $2k(k+1)$.

Pour expliciter l'ensemble Σ nous définissons $y_i = \frac{i}{k}, \forall i = 0, \ldots, k$ et nous désignons par Γ_{y_i} le segment horizontal passant par y_i inclus dans K, $\overrightarrow{\tau_{y_i}}$ le vecteur unitaire tangent associé et de manière similaire les x_i, Γ_{x_i} et $\overrightarrow{\tau_{x_i}}$ (voir figure 3.7).

Proposition 3.2.5. *L'ensemble des $2k(k+1)$ degrés de liberté décrits comme suit :*

$$\sigma_{\xi_i}^m(\overrightarrow{p}) = \int_{\Gamma_{\xi_i}} l_m(\bar{\xi})\overrightarrow{p} \cdot \overrightarrow{\tau_{\xi_i}}\, d\Gamma, \quad \begin{array}{l} \xi = x \text{ ou } y \\ i = 0, \ldots, k \\ m = 0, \ldots, k-1 \end{array},$$

où l_m désigne le $m^{ième}$ polynôme de Legendre normalisé sur $[0,1]$, est P-unisolvant et assurent la conformité de l'élément fini dans $H(rot)$.

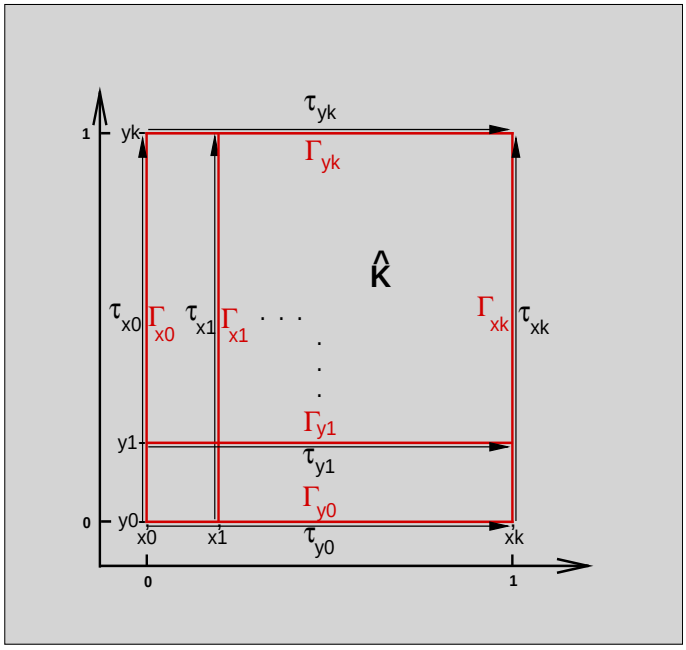

FIG. 3.7 –

Démonstration. Pour la conformité il suffit de se rendre compte que la composante tangentielle d'un élément de P est, sur chacune des quatres arêtes, un polynôme univarié de degré $k-1$.
Soit $\overrightarrow{p} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \in \begin{pmatrix} \mathbb{Q}_{k-1,k}(\hat{K}) \\ \mathbb{Q}_{k,k-1}(\hat{K}) \end{pmatrix}$.
Sur Γ_{y_0} (resp.Γ_{y_k}) la composante tangentielle de p vaut

$$\overrightarrow{p} \cdot \overrightarrow{\tau_{y_0|y=0}} = p_{1|y=0} \text{ (resp. } \overrightarrow{p} \cdot \overrightarrow{\tau_{y_k|y=1}} = p_{1|y=1} \text{)},$$

et sur Γ_{x_0} (resp.Γ_{x_k}) la composante tangentielle de p vaut

$$\overrightarrow{p} \cdot \overrightarrow{\tau_{x_0|x=0}} = p_{2|x=0} \text{ (resp. } \overrightarrow{p} \cdot \overrightarrow{\tau_{x_k|x=1}} = p_{2|x=1} \text{)}.$$

Cela signifie que sur les arêtes horizontales de $\hat{K}$ les composantes tangentielles de $\overrightarrow{p}$ sont des polynômes univariés en x et sur les arêtes horizontales de $\hat{K}$ celles-ci sont des polynômes univariés en y.
Ainsi la composante tangentielle de $\overrightarrow{p}$ est entièrement déterminée sur chacune des arêtes de $\hat{K}$ par k degrés de liberté que l'on choisit comme étant ses k premiers moments, ou, pour être plus précis les $\sigma_{\xi_i}^m$, pour $\xi = x$ ou y, $i = 0$ ou k et m de 0 à $k-1$.
Pour l'unisolvance nous utilisons le lemme 2.2.6.
Le fait d'utiliser des polynômes orthogonaux nous permet de donner la forme explicite de $2k(k+1)$ fonctions

$$\overrightarrow{\psi_{\xi_i}^m} = L_{\xi_i}(\xi) l_m(\bar{\xi}) \begin{pmatrix} \delta_{\xi y} \\ \delta_{\xi x} \end{pmatrix},$$

où

– L_{ξ_i} désigne le $i^{\text{ème}}$ polynôme de la base lagrangienne associée à l'ensemble $\{\xi_i\}_{i=0,\dots,k}$,
– l_m désigne le $m^{\text{ième}}$ polynôme de Legendre normalisé sur Γ_{ξ_i},
– $\overline{\xi} = x$ ou y respectivement si $\xi = y$ ou x,

vérifiant

$$\sigma^m_{\xi_i}(\overrightarrow{\psi^n_{\eta_j}}) = \delta_{(\xi,i,m),(\eta,j,n)},$$

où

$$\delta_{(\xi,i,m),(\eta,j,n)} = \begin{cases} 1 & \text{si } (\xi,i,m) = (\eta,j,n) \\ 0 & \text{sinon} \end{cases}.$$

En effet

$$\begin{aligned} \sigma^m_{\xi_i}(\overrightarrow{\psi^n_{\eta_j}}) &= \int_{\Gamma_{\xi_i}} l_m(\overline{\xi})\overrightarrow{\psi^n_{\eta_j}} \cdot \overrightarrow{\tau_{\xi_i}}\, d\Gamma \\ &= \int_{\Gamma_{\xi_i}} l_m(\overline{\xi}) L_{\eta_j}(\eta) l_n(\overline{\eta}) \begin{pmatrix} \delta_{\eta y} \\ \delta_{\eta x} \end{pmatrix} \cdot \overrightarrow{\tau_{\xi_i}}\, d\Gamma. \end{aligned}$$

Or $\overrightarrow{\tau_{\xi_i}} = \begin{pmatrix} \delta_{\xi y} \\ \delta_{\xi x} \end{pmatrix}$, de sorte que $\begin{pmatrix} \delta_{\eta y} \\ \delta_{\eta x} \end{pmatrix} \cdot \overrightarrow{\tau_{\xi_i}} = \delta_{\xi\eta}$. Ainsi

$$\begin{aligned} \sigma^m_{\xi_i}(\overrightarrow{\psi^n_{\eta_j}}) &= \int_{\Gamma_{\xi_i}} l_m(\overline{\xi}) L_{\eta_j}(\eta) l_n(\overline{\eta}) \begin{pmatrix} \delta_{\eta y} \\ \delta_{\eta x} \end{pmatrix} \cdot \overrightarrow{\tau_{\xi_i}}\, d\Gamma \\ &= \delta_{\xi\eta} \int_{\Gamma_{\xi_i}} l_m(\overline{\xi}) L_{\xi_j}(\xi) l_n(\overline{\xi})\, d\Gamma. \end{aligned}$$

Il faut ensuite remarquer que le polynôme de Lagrange $L_{\xi_j}(\xi)$ est constant sur Γ_{ξ_i} et vaut 1 ou 0 suivant que $i = j$ ou non. D'où

$$\begin{aligned} \sigma^m_{\xi_i}(\overrightarrow{\psi^n_{\eta_j}}) &= \delta_{\xi\eta} \int_{\Gamma_{\xi_i}} l_m(\overline{\xi}) L_{\xi_j}(\xi) l_n(\overline{\xi})\, d\Gamma \\ &= \delta_{\xi\eta}\delta_{ij} \int_{\Gamma_{\xi_i}} l_m(\overline{\xi}) l_n(\overline{\xi})\, d\Gamma. \end{aligned}$$

Le résultat vient alors naturellement du fait que les polynômes de Legendre ont été construits orthogonaux et normalisés :

$$\begin{aligned} \sigma^m_{\xi_i}(\overrightarrow{\psi^n_{\eta_j}}) &= \delta_{\xi\eta}\delta_{ij} \int_{\Gamma_{\xi_i}} l_m(\overline{\xi}) l_n(\overline{\xi})\, d\Gamma \\ &= \delta_{\xi\eta}\delta_{ij}\delta_{mn} \\ &= \delta_{(\xi,i,m),(\eta,j,n)}. \end{aligned}$$

□

Remarque 3.2.6. *Pour les mêmes raisons que celles données dans la sous-section 3.1.3, il n'est pas des plus judicieux de considérer des points x_i et y_i équirépartis : nous venons de montrer en démontrant la proposition 3.2.5 que la base associée aux formes linéaires définissant l'élément fini s'exprime en fonction de la base lagrangienne en une dimension*

Ordre de l'élément fini d'arête localisation équirépartie	1	2	3	4	5
conditionnement	3	9	12.19	19.14	32.53
constante de Lebesgue	1	3.4	8.1	16.75	32.8
Ordre de l'élément fini d'arête localisation optimisée	1	2	3	4	5
conditionnement	3	9	11.07	15	17.23
constante de Lebesgue	1	3.4	7.42	12.4	18.8

TAB. 3.2 – Conditionnement en norme L^2 de la matrice de masse et constante de Lebesgue calculés sur l'élément de référence pour les éléments finis d'arête rectangulaires d'ordre 1 à 5.

d'espace associée aux points x_i ou y_i. Ainsi il convient d'optimiser non-seulement la projection sur le sous-espace de discrétisation mais aussi le conditionnement de la matrice de masse en optimisant la localisation de ces points suivant les points de quadrature des formules de Gauss-Lobatto. Le tableau 3.2 consigne le conditionnement en norme L^2 de la matrice de masse et la constante de Lebesgue calculés sur l'élément de référence pour les éléments finis d'arête rectangulaires d'ordre 1 à 5, pour des points x_i et y_i équirépartis et pour des points x_i et y_i répartis aux points de quadrature des formules de Gauss-Lobatto.
Nous remarquons à nouveau que la relocalisation optimisée des points x_i et y_i permet de diminuer significativement le conditionnement de la matrice de masse et la constante de Lebesgue (même si ces constantes ne sont pas encore dramatiquement élevées jusqu'aux schémas d'ordre 5).

Remarque 3.2.7. *Nous parlerons des degrés de liberté comme étant les moments successifs des composantes tangentielles sur les segments Γ_{ξ_i}, même si l'on n'intègre pas à proprement parler les composantes tangentielles contre la base canonique $\{1, t, \ldots, t^{k-1}\}$.*

Il n'est pas inutile de se rendre compte dès à présent que la première composante p_1 d'un élément $\overrightarrow{p} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \in \begin{pmatrix} \mathbb{Q}_{k-1,k}(\hat{K}) \\ \mathbb{Q}_{k,k-1}(\hat{K}) \end{pmatrix}$ est entièrement déterminée par les $k(k+1)$ degrés de libertés $\sigma^m_{y_i}(P)$, tandis que sa deuxième composante p_2 est entièrement déterminée par les $k(k+1)$ degrés de libertés $\sigma^m_{x_i}(P)$, c'est-à-dire que la description de l'espace $\begin{pmatrix} \mathbb{Q}_{k-1,k}(\hat{K}) \\ \mathbb{Q}_{k,k-1}(\hat{K}) \end{pmatrix}$ se fait de manière découplée par les degrés de liberté associés aux lignes horizontales Γ_{x_i} pour la première composante, et par les degrés de liberté associés aux lignes verticales Γ_{y_i} pour la seconde composante. Cette propriété est d'autant plus remarquable que celle-ci restera vraie pour l'espace $\begin{pmatrix} \mathbb{Q}_{k-1,k}(\tilde{K}) \\ \mathbb{Q}_{k,k-1}(\tilde{K}) \end{pmatrix}$ où $\tilde{K}$ désigne un quelconque élément du maillage dès lors que la transformation bilinéaire transformant $\hat{K}$ en $\tilde{K}$ est en particulier affine, c'est-à-dire que $\tilde{K}$ est un rectangle aux cotés parallèles aux axes (Ox) et (Oy).

Remarque 3.2.8. *Dans le cadre des éléments finis d'arête rectangulaires il n'est pas indis-*

pensable d'exhiber les fonctions de base sur un élément quelconque du maillage en fonction des fonctions de base sur l'élément de référence étant donné que l'on peut aisément définir l'élément $[0, h_x] \times [0, h_y]$ *comme étant l'élément de référence, chacun des éléments du maillage se rapportant à cet élément via une translation de l'origine du plan.*

3.2.2 Discrétisation conforme dans le cas d'éléments finis d'arête triangulaires

Considérons à présent un maillage du domaine Ω par une triangulation $\{T_i\}_{i=1,\dots,r}$. Les espaces fonctionnels $X \subset H^1(\Omega)$, $W \subset H(rot, \Omega)$ et $V \subset L^2(\Omega)$ qui nous permettront une discrétisation conforme des équations de Maxwell et qui nous permettront aussi de conserver les propriétés de suite exacte de ces espaces que nous proposons sont les suivants :

$$X = \{\xi \in H^1(\Omega) \ | \ \xi_{|T_i} \in \mathbb{P}_k(T_i) + \overline{\mathbb{P}}_{k-1}(T_i)xy \ , \ \forall i = 1, \dots, r\},$$

$$W = \{\overrightarrow{\psi} \in H(rot, \Omega) \ | \ \overrightarrow{\psi}_{|T_i} \in \mathbb{P}^2_{k-1}(T_i) + \overline{\mathbb{P}}_{k-1}(T_i) \begin{pmatrix} y \\ -x \end{pmatrix} \ , \ \forall i = 1, \dots, r\},$$

$$V = \{\varphi \in L^2(\Omega) \ | \ \varphi_{|T_i} \in \mathbb{P}_{k-1}(T_i) \ , \ \forall i = 1, \dots, r\},$$

où $\overline{\mathbb{P}}_{k-1}$ désigne l'ensemble des polynômes de degré exactement $k-1$. Les espaces qui doivent être effectivement construits sont à nouveau les espaces W et V. L'espace V étant un espace localement $\mathbb{P}_{k-1}$ par élément du maillage et discontinu (conforme dans $L^2(\Omega)$) dont la construction ne pose pas de problème particulier, nous nous intéressons plus spécifiquement à la construction de l'espace W pour lequel il faut assurer la conformité dans $H(rot, \Omega)$, c'est à dire la continuité de la trace tangentielle des fonctions de W à travers les arêtes de la triangulation. Pour cela nous introduisons l'élément fini $(\hat{T}, P, \Sigma)$ suivant :

(i) $\hat{T}$ le triangle défini par les sommets $(0,0)$, $(1,0)$ et $(0,1)$,

(ii) $P = \mathbb{P}^2_{k-1}(\hat{T}) + \overline{\mathbb{P}}_{k-1}(\hat{T}) \begin{pmatrix} y \\ -x \end{pmatrix}$, espace polynomial de dimension $k(k+2)$,

(iii) Σ est un ensemble de formes linéaires sur P de cardinal fini $k(k+2)$.

Nous décrivons les formes linéaires de Σ de la manière suivante. Notant Γ_i, $i = 1, \dots, 3$, les trois arêtes de $\hat{T}$ et $\overrightarrow{\tau_i}$ un vecteur tangent unitaire associé nous définissons dans un premier temps un jeu de $3k$ formes linéaires sur l'espace des traces tangentielles par

$$\hat{\sigma}_i^m(\overrightarrow{p}) = \int_{\Gamma_i} l_m \overrightarrow{p} \cdot \overrightarrow{\tau_i} \, d\Gamma, \quad \begin{matrix} i = 1, \dots, 3 \\ m = 0, \dots, k-1 \end{matrix} \quad . \tag{3.9}$$

Nous définissons ensuite les $k(k+2) - 3k = k(k-1)$ formes linéaires restantes par

$$\hat{\tilde{\sigma}}_i^m(\overrightarrow{p}) = p_i(X_m), \quad \begin{matrix} i = 1, 2 \\ m = 1, \dots, \frac{(k-1)k}{2} \end{matrix} \quad , \tag{3.10}$$

où p_i, $i = 1, 2$, désigne la première ou la deuxième composante d'une fonction vectorielle $\overrightarrow{p} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$ et $\{X_m\}_{m=1,\dots,\frac{(k-1)k}{2}}$ désigne un jeu de points intérieurs au triangle faisant de l'ensemble $\{\sigma_i(f) := f(X_i)\}_{i=1,\dots,\frac{(k-1)k}{2}}$ un ensemble $\mathbb{P}_{k-2}$-unisolvant.

Proposition 3.2.9. *L'ensemble des $k(k+2)$ formes linéaires décrites dans (3.9) et (3.10) est P-unisolvant et assure la conformité de l'élément fini dans $H(rot,\Omega)$.*

Démonstration. Nous commençons par la conformité de l'élément :
Soit $M_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ et $M_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ deux points du plan. Nous allons regarder la forme de la trace tangentielle sur le segment $[M_1, M_2]$ d'une fonction $\overrightarrow{p}$ de P.
Soit $\overrightarrow{p}(x,y) = \begin{pmatrix} p_1(x,y) + \overline{p}(x,y)y \\ p_2(x,y) - \overline{p}(x,y)x \end{pmatrix}$, avec p_1 et p_2 deux fonctions de $\mathbb{P}_{k-1}$ et $\overline{p} \in \overline{\mathbb{P}}_{k-1}$. Un vecteur tangent au segment $[M_1, M_2]$ est donné par $\overrightarrow{\tau} = \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \end{pmatrix}$ et une paramétrisation de ce segment par $\Gamma(\alpha) = (\xi(\alpha), \eta(\alpha)) = (x_1 + \alpha(x_2 - x_1), y_1 + \alpha(y_2 - y_1))$. La composante tangentielle de $\overrightarrow{p}$ sur $[M_1, M_2]$ est alors donnée par :

$$\begin{aligned}\overrightarrow{p} \cdot \overrightarrow{\tau}(\Gamma(\alpha)) =& (x_2 - x_1)\left\{p_1(\Gamma(\alpha)) + \overline{p}(\Gamma(\alpha))\eta(\alpha)\right\} + \\ & (y_2 - y_1)\left\{p_2(\Gamma(\alpha)) - \overline{p}(\Gamma(\alpha))\xi(\alpha)\right\} \\ =& (x_2 - x_1)p_1(\Gamma(\alpha)) + (y_2 - y_1)p_2(\Gamma(\alpha)) + \\ & \overline{p}(\Gamma(\alpha))\left\{(x_2 - x_1)\eta(\alpha) - (y_2 - y_1)\xi(\alpha)\right\}.\end{aligned}$$

Après simplification il vient

$$(x_2 - x_1)\eta(\alpha) - (y_2 - y_1)\xi(\alpha) = x_2 y_1 - x_1 y_2,$$

d'où

$$\begin{aligned}\overrightarrow{p} \cdot \overrightarrow{\tau}(\Gamma(\alpha)) = & (x_2 - x_1)p_1(\Gamma(\alpha)) \\ & + (y_2 - y_1)p_2(\Gamma(\alpha)) \\ & + (x_2 y_1 - x_1 y_2)\overline{p}(\Gamma(\alpha)).\end{aligned}$$

Cela signifie que la trace tangentielle sur un quelconque segment du plan paramétré par α d'un élément de P est invariablement un polynôme univarié en α de degré $k-1$. En particulier la trace tangentielle sur l'une des arêtes du triangle $\hat{T}$ (ou d'un quelconque triangle du maillage) est donc déterminée de manière unique par ces k premiers moments, ce qui correspond (modulo le fait que l'on intègre pas exactement la trace tangentielle contre la base canonique mais contre une base de polynômes de Legendre) à la définition des formes linéaires que l'on a données dans (3.9).

Pour l'unisolvance nous venons de voir que les $3k$ formes linéaires définies dans (3.9) forment un ensemble unisolvant dans l'espace des traces tangentielles des fonctions de P sur les trois arêtes. Il nous reste donc à voir que les $(k-1)k$ formes linéaires définies dans (3.10) forment un ensemble unisolvant dans le sous-espace généré par les éléments de P à trace tangentielle nulle sur chacune des arêtes. Pour cela il est possible de montrer (en toute généralité cela s'avère compliqué à écrire puisqu'il faut expliciter comment se répercute la contrainte de la trace tangentielle nulle sur les trois arêtes sur les coefficients d'un quelconque polynôme de P) que tout élément de ce sous-espace de P s'écrit comme un polynôme de $\begin{pmatrix} y\mathbb{P}_{k-1} \\ x\mathbb{P}_{k-1} \end{pmatrix}$ tel que la partie en $\overline{\mathbb{P}}_{k-1}$ des deux polynômes de $\mathbb{P}_{k-1}$ est liée à

Ordre de l'élément fini d'arête degrés de liberté intérieurs de Lagrange	1	2	3	4	5
conditionnement	3	13.3	44.0	74.9	226
constante de Lebesgue	2.18	5.95	10.8	16.6	23.5
Ordre de l'élément fini d'arête degrés de liberté intérieurs surfaciques	1	2	3	4	5
conditionnement	3	72.7	831	21213	629180
constante de Lebesgue	2.18	8.5	38.5	215	1260

TAB. 3.3 – Conditionnement en norme L^2 de la matrice de masse et constante de Lebesgue calculés sur l'élément de référence pour les éléments finis d'arête triangulaires d'ordre 1 à 5.

leur partie en $\mathbb{P}_{k-2}$ de sorte que ce sous-espace est en bijection avec $\mathbb{P}^2_{k-2}$. Ainsi l'ensemble des $(k-1)k$ formes linéaires définies dans (3.10), qui est par définition unisolvant sur $\mathbb{P}^2_{k-2}$, l'est aussi sur le sous-espace généré par les éléments de P à trace tangentielle nulle sur chacune des arêtes. □

Remarque 3.2.10. *Quitte à nous répéter, nous soulignons l'importance du choix de la localisation des points X_i, notamment dans le cadre des éléments finis d'arête triangulaires que nous venons de définir : dans la pratique il n'est pas indispensable de définir les degrés de liberté n'assurant pas la conformité de l'éléments comme dans (3.10), c'est-à-dire par des formes linéaires lagrangiennes, tout autre jeu de $k(k-1)$ formes linéaire unisolvant sur le sous-espace de W constitué des fonctions de W à trace tangentielle nulle sur les trois arêtes faisant l'affaire. Nous avions initialement défini ces $k(k-1)$ formes linéaire comme étant des moments surfaciques sur le triangle, sans jamais réussir à générer des fonctions de base dont les oscillations seraient raisonnables à partir des éléments du cinquième ordre. Ceci illustre bien les problèmes qui apparaissent dans la construction de méthodes d'éléments finis d'ordre élevé : il est indispensable de s'assurer que les degrés de liberté sont choisis de manière à optimiser la projection qu'ils définissent sur le sous-espace de discrétisation et, par conséquent, à minimiser les oscillations des fonctions de base qui leur sont associées.*

Le tableau 3.3 liste le conditionnement de la matrice de masse en norme L^2 et la constante de Lebesgue calculés sur le triangle de référence pour les éléments finis d'arête triangulaires que l'on a retenu (c'est-à-dire ceux dont les formes linéaires intérieures sont les formes linéaires lagrangiennes décrites dans (3.10)) et pour les éléments finis d'arête triangulaires dont les formes linéaires intérieures sont des moments surfaciques.
Les résultats parlent d'eux-mêmes : le conditionnement de la matrice de masse et la constante de Lebesgue explosent littéralement avec les degrés de libertés intérieurs surfaciques alors qu'avec les degrés de liberté intérieurs lagrangiens correctement localisés nous arrivons à maîtriser ces constantes.

Nous désignons alors par $\{\overrightarrow{\hat{\psi}_i^m}, \overrightarrow{\hat{\tilde{\psi}}_j^n}\}$ l'ensemble des $k(k+2)$ fonctions de base associées ($3k$ fonctions $\overrightarrow{\hat{\psi}_i^m}$ associées aux formes linéaires $\hat{\sigma}_i^m$ et $k(k-1)$ fonctions $\overrightarrow{\hat{\tilde{\psi}}_i^m}$ associées aux

formes linéaires $\hat{\tilde{\sigma}}_i^m$) et nous cherchons à exhiber les fonctions de base associées aux formes linéaires transportées sur un élément quelconque de la triangulation défini par les sommets (x_1, y_1),(x_2, y_2) et (x_3, y_3).
Rappelons que l'élément de référence est transporté vers cet élément par la transformation affine définie par

$$\overrightarrow{\Phi}\begin{pmatrix} x \\ y \end{pmatrix} = \underbrace{\begin{pmatrix} x_2 - x_1 & x_3 - x_1 \\ y_2 - y_1 & y_3 - y_1 \end{pmatrix}}_{A}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}.$$

Considérons dans un premier temps l'une des $3k$ formes linéaires d'arête $\hat{\sigma}_i^m$

$$\hat{\sigma}_i^m(\overrightarrow{p}) = \int_{\hat{\Gamma}_i} l_m \overrightarrow{p} \cdot \overrightarrow{\tau_i}\, d\Gamma = \int_0^{\|\hat{\Gamma}_i\|} l_m(t) \left(\begin{pmatrix} P_1 \\ P_2 \end{pmatrix} \cdot \begin{pmatrix} \tau_1^i \\ \tau_2^i \end{pmatrix} \right) (\overrightarrow{\hat{\Gamma}}_i(t))\, dt,$$

où $\overrightarrow{\hat{\Gamma}}_i(t) = \begin{pmatrix} \hat{\Gamma}_1^i(t) \\ \hat{\Gamma}_2^i(t) \end{pmatrix}$ est l'unique paramétrisation affine normalisée de $\hat{\Gamma}_i$, de sorte que $\begin{pmatrix} \tau_1^i(\overrightarrow{\hat{\Gamma}}_i(t)) \\ \tau_2^i(\overrightarrow{\hat{\Gamma}}_i(t)) \end{pmatrix} = \begin{pmatrix} \hat{\Gamma_1^i}'(t) \\ \hat{\Gamma_2^i}'(t) \end{pmatrix}$, et est en particulier constant sur $\hat{\Gamma}_i$. Finalement,

$$\hat{\sigma}_i^m(\overrightarrow{p}) = \int_0^{\|\hat{\Gamma}_i\|} l_m(t) \left(\overrightarrow{p}(\overrightarrow{\hat{\Gamma}}_i(t)) \cdot \begin{pmatrix} \hat{\Gamma_1^i}' \\ \hat{\Gamma_2^i}' \end{pmatrix} \right) dt. \tag{3.11}$$

Nous définissons alors la forme linéaire σ_i^m, correspondant à la transposition de la forme linéaire $\hat{\sigma}_i^m$ de l'élément de référence vers l'élément considéré, par

$$\sigma_i^m(\overrightarrow{p}) = \int_0^{\|\hat{\Gamma}_i\|} l_m(t) \left(\begin{pmatrix} P_1 \\ P_2 \end{pmatrix} \cdot \begin{pmatrix} \tau_1^i \\ \tau_2^i \end{pmatrix} \right) (\overrightarrow{\Phi}(\overrightarrow{\hat{\Gamma}}_i(t)))\, dt,$$

avec

$$\begin{aligned}
\begin{pmatrix} \tau_1^i(\overrightarrow{\Phi}(\overrightarrow{\hat{\Gamma}}_i(t))) \\ \tau_2^i(\overrightarrow{\Phi}(\overrightarrow{\hat{\Gamma}}_i(t))) \end{pmatrix} &= \begin{pmatrix} \overrightarrow{\Phi}(\overrightarrow{\hat{\Gamma}}_i(t))_1' \\ \overrightarrow{\Phi}(\overrightarrow{\hat{\Gamma}}_i(t))_2' \end{pmatrix} \\
&= \begin{pmatrix} ((x_2 - x_1)\overrightarrow{\hat{\Gamma}}_1^i(t) + (x_3 - x_1)\overrightarrow{\hat{\Gamma}}_2^i(t) + x_1)' \\ ((y_2 - y_1)\overrightarrow{\hat{\Gamma}}_1^i(t) + (y_3 - y_1)\overrightarrow{\hat{\Gamma}}_2^i(t) + y_1)' \end{pmatrix} \\
&= \begin{pmatrix} (x_2 - x_1)\overrightarrow{\hat{\Gamma}_1^i}' + (x_3 - x_1)\overrightarrow{\hat{\Gamma}_2^i}' \\ (y_2 - y_1)\overrightarrow{\hat{\Gamma}_1^i}' + (y_3 - y_1)\overrightarrow{\hat{\Gamma}_2^i}' \end{pmatrix} \\
&= A \begin{pmatrix} \hat{\Gamma_1^i}' \\ \hat{\Gamma_2^i}' \end{pmatrix}.
\end{aligned}$$

Ainsi

$$\begin{aligned}
\sigma_i^m(\overrightarrow{p}) &= \int_0^{\|\hat{\Gamma}_i\|} l_m(t) \left(\begin{pmatrix} P_1(\overrightarrow{\Phi}(\overrightarrow{\hat{\Gamma}}_i(t))) \\ P_2(\overrightarrow{\Phi}(\overrightarrow{\hat{\Gamma}}_i(t))) \end{pmatrix} \cdot A \begin{pmatrix} \hat{\Gamma_1^i}' \\ \hat{\Gamma_2^i}' \end{pmatrix} \right) dt \\
&= \int_0^{\|\hat{\Gamma}_i\|} l_m(t) \left({}^tA\overrightarrow{p}(\overrightarrow{\Phi}(\overrightarrow{\hat{\Gamma}}_i(t))) \cdot \begin{pmatrix} \hat{\Gamma_1^i}' \\ \hat{\Gamma_2^i}' \end{pmatrix} \right) dt.
\end{aligned} \tag{3.12}$$

Il faut alors bien se garder d'affirmer que l'unisolvance et la définition des $3k$ fonctions de base $\overrightarrow{\hat{\psi}_i^m}$ impliquent dès à présent que

$$ {}^tA\overrightarrow{\psi}_j^m\left(\overrightarrow{\Phi}\begin{pmatrix} x \\ y \end{pmatrix}\right) = \overrightarrow{\hat{\psi}}_j^m\begin{pmatrix} x \\ y \end{pmatrix}, \forall \begin{pmatrix} x \\ y \end{pmatrix} \in T. \tag{3.13} $$

En effet l'identification des équations (3.11) et (3.12) en suivant l'argument que seules les fonctions $\overrightarrow{\hat{\psi}_i^m}$ vérifient la relation $\hat{\sigma}_i^m(\overrightarrow{\hat{\psi}_j^n}) = \delta_{(i,j),(m,n)}$ ne nous permet que de conclure que l'égalité (3.13) ne vaut que sur les arêtes de T (jusqu'à présent nous n'avons transporté que les formes linéaires définissant les traces tangentielles d'une fonction...).
Si l'on veut que l'équation (3.13) soit vraie, c'est-à-dire si l'on veut pouvoir étendre la régularité donnée par l'équation (3.13), qui est vraie sur chacune des arêtes de T, à tout le triangle T, il faut définir convenablement les formes linéaires $\tilde{\sigma}_i^m$, transposées des formes linéaires $\hat{\tilde{\sigma}}_i^m$ de l'élément de référence T vers l'élément considéré.
C'est pourquoi plutôt que définir (de la manière la plus intuitive) les $k(k-1)$ formes linéaires $\tilde{\sigma}_i^m$ par

$$ \tilde{\sigma}_i^m(\overrightarrow{p}) = \hat{\tilde{\sigma}}_i^m(\overrightarrow{p} \circ \overrightarrow{\Phi}) = p_i(\overrightarrow{\Phi}(X_m)), $$

nous définissons ces formes linéaires par

$$ \tilde{\sigma}_i^m(\overrightarrow{p}) = \hat{\tilde{\sigma}}_i^m(({}^tA\overrightarrow{p}) \circ \overrightarrow{\Phi}) = ({}^tA\overrightarrow{p})_i(\overrightarrow{\Phi}(X_m)). \tag{3.14} $$

Les fonctions de bases $\{\overrightarrow{\psi_i^m}, \overrightarrow{\tilde{\psi}_j^n}\}$ définies comme étant les uniques fonctions vérifiant

$$ \begin{cases} \sigma_i^m(\overrightarrow{\psi_j^n}) &= \delta_{(i,j),(m,n)} \\ \sigma_i^m(\overrightarrow{\tilde{\psi}_j^n}) &= 0 \\ \tilde{\sigma}_i^m(\overrightarrow{\psi_j^n}) &= 0 \\ \tilde{\sigma}_i^m(\overrightarrow{\tilde{\psi}_j^n}) &= \delta_{(i,j),(m,n)} \end{cases}, $$

résolvent donc le système explicitement donné par,

$$ \begin{cases} \displaystyle\int_0^{\|\hat{\Gamma}_i\|} l_m(t)\left({}^tA\overrightarrow{\psi_j^n}(\overrightarrow{\Phi}(\overrightarrow{\hat{\Gamma}}_i(t))) \cdot \begin{pmatrix} \hat{\Gamma}_1^{i\,\prime} \\ \hat{\Gamma}_2^{i\,\prime} \end{pmatrix}\right) dt &= \delta_{(i,j),(m,n)} \\ \displaystyle\int_0^{\|\hat{\Gamma}_i\|} l_m(t)\left({}^tA\overrightarrow{\tilde{\psi}_j^n}(\overrightarrow{\Phi}(\overrightarrow{\hat{\Gamma}}_i(t))) \cdot \begin{pmatrix} \hat{\Gamma}_1^{i\,\prime} \\ \hat{\Gamma}_2^{i\,\prime} \end{pmatrix}\right) dt &= 0 \\ ({}^tA\overrightarrow{\psi_j^n})_i(\overrightarrow{\Phi}(X_m)) &= 0 \\ ({}^tA\overrightarrow{\tilde{\psi}_j^n})_i(\overrightarrow{\Phi}(X_m)) &= \delta_{(i,j),(m,n)} \end{cases}. $$

Or par définition même, les fonctions $\overrightarrow{\hat{\psi}_j^n}$ et $\overrightarrow{\hat{\tilde{\psi}}_j^n}$ sont les seules à résoudre le système

$$\left\{\begin{array}{rcl} \displaystyle\int_0^{\|\hat{\Gamma}_i\|} l_m(t)\left(\overrightarrow{\hat{\psi}_j^n}(\overrightarrow{\hat{\Gamma}}_i(t))\cdot\begin{pmatrix}\hat{\Gamma_1^i}'\\ \hat{\Gamma_2^i}'\end{pmatrix}\right)dt &=& \delta_{(i,j),(m,n)}\\ \displaystyle\int_0^{\|\hat{\Gamma}_i\|} l_m(t)\left(\overrightarrow{\hat{\tilde{\psi}}_j^n}(\overrightarrow{\hat{\Gamma}}_i(t))\cdot\begin{pmatrix}\hat{\Gamma_1^i}'\\ \hat{\Gamma_2^i}'\end{pmatrix}\right)dt &=& 0\\ (\overrightarrow{\hat{\psi}_j^n})_i(X_m) &=& 0\\ (\overrightarrow{\hat{\tilde{\psi}}_j^n})_i(X_m) &=& \delta_{(i,j),(m,n)}\end{array}\right. .$$

Par identification il vient donc tout naturellement

$$\begin{array}{rcl} {}^tA\overrightarrow{\psi_j^n}\left(\overrightarrow{\Phi}\begin{pmatrix}x\\ y\end{pmatrix}\right) &=& \overrightarrow{\hat{\psi}_j^n}\begin{pmatrix}x\\ y\end{pmatrix},\\ {}^tA\overrightarrow{\tilde{\psi}_j^n}\left(\overrightarrow{\Phi}\begin{pmatrix}x\\ y\end{pmatrix}\right) &=& \overrightarrow{\hat{\tilde{\psi}}_j^n}\begin{pmatrix}x\\ y\end{pmatrix},\end{array}$$

ou de manière équivalente

$$\begin{array}{rcl} \overrightarrow{\psi_j^n}\circ\overrightarrow{\Phi} &=& {}^tA^{-1}\overrightarrow{\hat{\psi}_j^n},\\ \overrightarrow{\tilde{\psi}_j^n}\circ\overrightarrow{\Phi} &=& {}^tA^{-1}\overrightarrow{\hat{\tilde{\psi}}_j^n}.\end{array}$$

En résumé, en définissant sur un élément quelconque de la triangulation les $3k$ formes linéaires définissant la trace tangentielle d'une fonction de l'espace polynomial comme dans (3.12) et le reste des $k(k-1)$ formes linéaires comme dans (3.14) nous nous assurons que les fonctions de base définies sur cet élément sont localement définies par

$$\overrightarrow{\psi}\circ\overrightarrow{\Phi} = {}^tA^{-1}\overrightarrow{\hat{\psi}},$$

où A désigne la partie linéaire de la transformation affine $\overrightarrow{\Phi}$ transportant le triangle de référence T vers l'élément considéré.

De la même manière que pour les éléments finis de Lagrange pour lesquels il nous a fallu déterminer, pour le calcul de la matrice de raideur, le gradient des fonctions de base, il nous faut à présent déterminer le rotationnel de ces fonctions :

$$\begin{array}{rcl} \nabla\times\overrightarrow{\psi}\begin{pmatrix}x\\ y\end{pmatrix} &=& \partial_x\psi_2\begin{pmatrix}x\\ y\end{pmatrix}-\partial_y\psi_1\begin{pmatrix}x\\ y\end{pmatrix}\\ &=& \partial_x\left\{({}^tA^{-1}\overrightarrow{\hat{\psi}})_2\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\ y\end{pmatrix}\right)\right\}-\partial_y\left\{({}^tA^{-1}\overrightarrow{\hat{\psi}})_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\ y\end{pmatrix}\right)\right\}\\ &=& \partial_x\left\{\frac{1}{\det A}\left(-(x_3-x_1)\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\ y\end{pmatrix}\right)+(x_2-x_1)\hat{\psi}_2\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\ y\end{pmatrix}\right)\right)\right\}-\\ && \partial_y\left\{\frac{1}{\det A}\left((y_3-y_1)\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\ y\end{pmatrix}\right)-(y_2-y_1)\hat{\psi}_2\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\ y\end{pmatrix}\right)\right)\right\},\end{array}$$

avec

$$\begin{array}{rcl} \partial_x\left(\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\ y\end{pmatrix}\right)\right) &=& \partial_x\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\ y\end{pmatrix}\right)\partial_x\Phi_1^{-1}\begin{pmatrix}x\\ y\end{pmatrix}+\\ && \partial_y\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\ y\end{pmatrix}\right)\partial_x\Phi_2^{-1}\begin{pmatrix}x\\ y\end{pmatrix},\end{array}$$

c'est-à-dire

$$\begin{aligned}\partial_x\left(\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)\right) &= \frac{1}{\det A}\left\{(y_3-y_1)\partial_x\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)-\right.\\ &\qquad\left.(y_2-y_1)\partial_y\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)\right\},\end{aligned}$$

et de la même manière

$$\begin{aligned}\partial_x\left(\hat{\psi}_2\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)\right) &= \frac{1}{\det A}\left\{(y_3-y_1)\partial_x\hat{\psi}_2\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)-\right.\\ &\qquad\left.(y_2-y_1)\partial_y\hat{\psi}_2\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)\right\}\\ \partial_y\left(\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)\right) &= \frac{1}{\det A}\left\{-(x_3-x_1)\partial_x\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)+\right.\\ &\qquad\left.(x_2-x_1)\partial_y\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)\right\}\\ \partial_y\left(\hat{\psi}_2\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)\right) &= \frac{1}{\det A}\left\{-(x_3-x_1)\partial_x\hat{\psi}_2\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)+\right.\\ &\qquad\left.(x_2-x_1)\partial_y\hat{\psi}_2\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)\right\}.\end{aligned}$$

Il s'ensuit que

$$\begin{aligned}&\nabla\times\overrightarrow{\psi}\begin{pmatrix}x\\y\end{pmatrix}\\ &=\frac{1}{(\det A)^2}\left\{-(x_3-x_1)\left(\quad(y_3-y_1)\partial_x\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)-(y_2-y_1)\partial_y\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)\right)\right.\\ &\qquad+(x_2-x_1)\left(\quad(y_3-y_1)\partial_x\hat{\psi}_2\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)-(y_2-y_1)\partial_y\hat{\psi}_2\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)\right)\\ &\qquad-(y_3-y_1)\left(-(x_3-x_1)\partial_x\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)+(x_2-x_1)\partial_y\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)\right)\\ &\qquad\left.+(y_2-y_1)\left(-(x_3-x_1)\partial_x\hat{\psi}_2\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)+(x_2-x_1)\partial_y\hat{\psi}_2\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)\right)\right\}\\ &=\frac{1}{(\det A)^2}\left\{\left((x_2-x_1)(y_3-y_1)-(y_2-y_1)(x_3-x_1)\right)\partial_x\hat{\psi}_2\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)+\right.\\ &\qquad\left.\left((x_3-x_1)(y_2-y_1)-(y_3-y_1)(x_2-x_1)\right)\partial_y\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)\right\}\\ &=\frac{1}{(\det A)}\left\{\partial_x\hat{\psi}_2\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)-\partial_y\hat{\psi}_1\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right)\right\}\\ &=\frac{1}{(\det A)}\left(\nabla\times\overrightarrow{\hat{\psi}}\right)\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right).\end{aligned}$$

Finalement,

$$\nabla\times\overrightarrow{\psi}\begin{pmatrix}x\\y\end{pmatrix}=\frac{1}{(\det A)}\left(\nabla\times\overrightarrow{\hat{\psi}}\right)\left(\overrightarrow{\Phi}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right).$$

3.2.3 Couplage conforme des éléments finis d'arête rectangulaires et triangulaires

Une fois que l'on a défini les éléments finis d'arête rectangulaires et triangulaires, la question que l'on s'est posée est la suivante : est-il possible de coupler ces éléments finis sur un maillage hybride ?
Cette question nous est venue naturellement pour plusieurs raisons : l'utilisation de maillages cartésiens n'étant possible que sur un nombre très restreints de domaines de calculs (en particulier les domaines rectangulaires), les éléments finis d'arêtes triangulaires semblent plus adaptés sur des domaine quelconques. En effet s'il est envisageable d'utiliser les éléments finis rectangulaires sur une quadrangulation du domaine, il est reconnu que la génération et l'utilisation dans le cadre des éléments finis d'une quadrangulation d'un domaine est bien plus délicate qu'une triangulation. C'est en particulier pourquoi nous n'avons décrit les éléments finis rectangulaires que dans le cadre d'un maillage cartésien.
L'utilisation d'éléments finis rectangulaires sur un maillage cartésien offre en revanche l'avantage, par rapport à l'utilisation d'éléments finis triangulaires sur un maillage non-structuré, de maîtriser entièrement la déformation du maillage, tous les éléments étant identiques.
C'est pourquoi nous avons envie de tirer un maximum de profit de chacune des méthodes en maillant le domaine de calcul par un maillage cartésien, là où la géométrie de ce domaine le permet (typiquement dans les zones intérieures au domaine), et par une triangulation sur le reste du domaine (typiquement à proximité des frontière du domaine).
Les espaces de discrétisation associés à ce maillage seront à ce moment localement différents suivant que l'on se place sur un élément rectangulaire ou triangulaire, mais les propriétés de suites exactes resteront vérifiées. Ainsi il reste à vérifier la conformité de la méthode, c'est-à-dire l'inclusion des espaces de discrétisation dans les espaces continus, pour que le diagramme (3.2) soit vérifié et que la résolution des équations de Maxwell se résume de nouveau à la résolution des équations (3.4).
La conformité de la méthode se résume dans le cas présent à l'inclusion $W \subset H(rot, \Omega)$. Il suffit donc de s'assurer de la continuité de la trace tangentielle aux interfaces d'éléments voisins. Rappelons alors que cette trace tangentielle est un polynôme univarié de degré $k-1$ entièrement déterminé par ses k premiers moments, que l'on considère un élément fini rectangulaire ou triangulaire. La trace tangentielle d'un élément de W est donc automatiquement continue dès lors que l'on peut imposer l'égalité de ces degrés de liberté, c'est-à-dire dès que le maillage hybride est lui aussi conforme (le recouvrement des arêtes partagées par deux éléments voisins doit de nouveau être exact, que ces éléments soient du même type ou non (voir figure 3.8)).

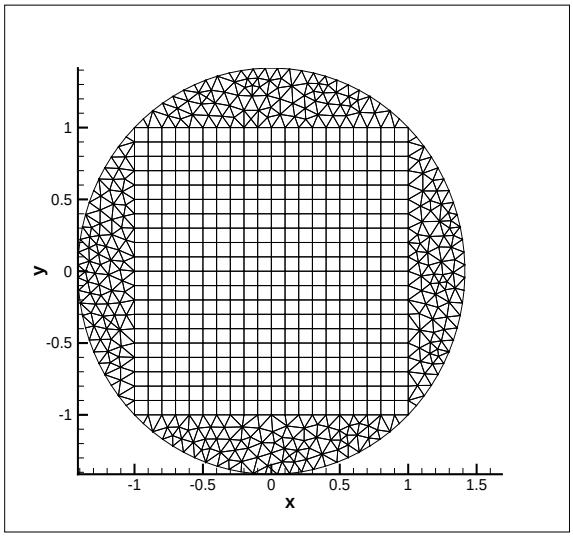

FIG. 3.8 – Un exemple de maillage hybride d'un disque.

Chapitre 4

Condensation de la matrice de masse

Nous avons vu que la semi-discrétisation en espace de l'équation des ondes, tout comme celle des équations de Maxwell, fait apparaître une matrice dite matrice de masse. Nous allons voir au chapitre 5 que suivant la semi-discrétisation en temps, il nous faudra inverser un certain nombre de fois cette matrice (ce nombre étant d'autant plus élevé que l'ordre de la discrétisation en temps est élevé) et ce à chaque itération en temps. L'idéal serait donc que cette matrice soit diagonale de manière à ce que le coût de ces inversions, en terme de temps de calcul, ne soit pas trop élevé. Or, ni la matrice de masse issue de l'utilisation des éléments finis de Lagrange triangulaires standards, ni la matrice de masse issue de l'utilisation des éléments finis d'arête triangulaires ou rectangulaires, ne sont diagonales (quel que soit l'ordre de ces éléments).
La question de la condensation de la matrice de masse (mass-lumping) issue de l'utilisation des éléments finis de Lagrange triangulaires a été étudiée par G.Cohen et al. dans [26] et [70] pour les éléments finis P_1 à P_3 et par W. A. Mulder et al. [14] pour les éléments finis P_4 et P_5 par exemple. Nous reprenons les considérations de symétrie proposées par N. Tordjman [70] et développons un algorithme de construction systématique d'éléments finis permettant la condensation de la matrice de masse. Cet algorithme nous a permis non seulement de retrouver les éléments finis des cinq premiers ordres permettant la condensation de la matrice de masse, mais aussi de déterminer l'espace polynomial et la formule de quadrature nécessaire à la condensation des éléments finis de type P_6. Nous proposons aussi la construction de nouveaux éléments finis dont la matrice de masse sera partiellement condensée : l'idée à mi-chemin entre la méthode des éléments finis conformes et la méthode des éléments finis de Galerkin discontinus consiste à orthogonaliser le plus de fonctions de base possible de manière à creuser le profil de la matrice de masse.
Le problème de la condensation de la matrice de masse issue des éléments finis d'arête que nous allons considérer est un problème encore largement ouvert. À notre connaissance les seuls cas particuliers ayant été résolus sont, le cas des éléments finis d'arête triangulaires de plus bas degré par Y. Haugazeau et P. Lacoste [51], et le cas particulier des éléments finis d'arête rectangulaires sur maillage cartésien par G. Cohen et P. Monk [27][28]. Nous proposons ici de nouveaux éléments finis permettant de condenser la matrice de masse issue des éléments finis d'arête rectangulaires sur maillage cartésien.

4.1 Principe de la condensation de la matrice de masse

L'idée du mass-lumping pour condenser la matrice de masse associée aux éléments finis de Lagrange est assez simple : si on arrivait à déterminer une formule de quadrature assez précise (de manière à ne pas faire diminuer l'ordre de la méthode), il suffirait alors de choisir les points auxquels on associe les degrés de liberté comme étant les points de quadrature pour obtenir une matrice de masse diagonale. Tout le problème réside donc dans la détermination de formules de quadrature : il faut non seulement que celles-ci soient assez précises de manière à ne pas faire diminuer l'ordre de la méthode mais aussi que le jeu de points de quadrature associé respecte un certain nombre de contraintes (liées aux conditions de symétrie, conformité et unisolvance), celui-ci ayant pour vocation à être confondu avec le jeu de points auquel on associe l'ensemble des degrés de liberté.

Plus précisément considérons la matrice de masse M_{ij} associée aux éléments finis de Lagrange :

$$M_{ij} = \int_\Omega \psi_i(x)\psi_j(x)dx.$$

Si maintenant, plutôt que de calculer exactement les intégrales définissant M_{ij}, nous les évaluons approximativement en utilisant, dans chaque triangle, une formule de quadrature ; c'est à dire si nous remplaçons le produit scalaire usuel de L^2 par un produit scalaire discret (ce qui revient exactement à remplacer l'intégrale par une formule de quadrature) défini par :

$$(u, v)_h = \sum_l w_{l,h} u(a_l)v(a_l),$$

où $\{a_l\}$ désigne l'ensemble des points de quadrature sur la réunion des triangles et $\{w_{l,h}\}$ l'ensemble des poids associés ; chacun pourra se persuader que la nouvelle matrice de masse ainsi obtenue sera diagonale dès lors que les degrés de liberté seront imposés comme étant les points de quadrature : en effet, comme chaque fonction de base ψ_i (associée à ces nouveaux degrés de liberté) n'est non nulle qu'en un unique degré de liberté qui lui est propre, nous aurons

$$M_{ij} := (\psi_i, \psi_j)_h = \textstyle\sum_l w_{l,h}\psi_i(a_l)\psi_j(a_l) = 0 \quad \forall i \neq j.$$

Ce qui signifie bien que les termes extra-diagonaux de la matrice $M_{i,j}$ sont nécessairement nuls. Tout le problème du mass-lumping réside donc dans la détermination de formules de quadrature appropriées.

4.2 Le cas 1D

Nous allons nous placer sur le segment $[0, 1]$ et déterminer une formule de quadrature ayant un certain nombre de propriétés. La première question à se poser est la suivante : quel doit être l'ordre de la formule de quadrature de manière à ne pas perdre l'ordre de la méthode ? Nous nous référons aux travaux de P. G. Ciarlet [17], pour répondre à cette question : si l'on veut conserver l'ordre d'une méthode d'éléments finis P_k en utilisant une formule de quadrature pour calculer la matrice de masse, plutôt que de calculer les

intégrales qui la composent de manière exacte, il faut que cette formule de quadrature intègre exactement les polynômes de degré inférieur ou égal à $2k-1$ au moins (ce résultat n'est valable qu'en une dimension d'espace, nous verrons comment celui-ci se généralise en deux dimensions).

Dans ce qui suit nous allons considérer uniquement des formules de quadrature symétriques : si Q désigne le jeu de points de quadrature et si $q \in Q$, alors $1-q \in Q$ et leur poids sont égaux. Remarquons que la plupart des formules de quadrature sont symétriques. Ces formules ont par ailleurs, de part leur forme spécifique, des propriétés intéressantes, notamment celle d'intégrer exactement toute fonction "impaire" (c'est-à-dire toute fonction vérifiant $f(x)+f(1-x)=0$), dont l'intégrale sur $[0,1]$ est bien entendu nulle. Bien que cette notion de symétrie d'une formule de quadrature ne soit pas essentielle d'un point de vue théorique (il serait envisageable de ne pas se restreindre à ce type de formule de quadrature pour atteindre la condensation de la matrice de masse), elle se révèle extrêmement utile sur le plan pratique puisqu'elle va nous permettre de réduire le système dont la solution détermine entièrement la formule de quadrature. Plus précisément, la formule de quadrature devant intégrer un ensemble de polynômes de degré inférieur ou égal à $2k-1$, il suffit de vérifier que cette formule intègre exactement une base de $\mathbb{P}_{2k-1}$. Si dans le choix de cette base deux fonctions sont symétriques sur $[0,1]$ (c'est-à-dire si f et g sont telles que $f(x) = g(1-x)\ \forall x \in [0,1]$), alors il suffit de s'assurer que l'une d'entre elles est intégrée exactement pour que l'autre le soit aussi. Ceci est dans la pratique très aisée via l'utilisation des coordonnées barycentriques :

Lemme 4.2.1. *Soit $\Lambda_1(x) = 1-x$ et $\Lambda_2(x) = x$ les coordonnées barycentriques des extrémités du segment $[0,1]$. Alors l'ensemble*

$$\{\Lambda_1^{2k-1}, \Lambda_1^{2k-2}\Lambda_2, \ldots, \Lambda_2^{2k-1}\}$$

est une base de $\mathbb{P}_{2k-1}$.

Démonstration. L'ensemble $\{\Lambda_1^{2k-1}, \Lambda_1^{2k-2}\Lambda_2, \ldots, \Lambda_2^{2k-1}\}$ est constitué de $2k$ fonctions dans l'espace $\mathbb{P}_{2k-1}$ de dimension $2k$. Il suffit donc de vérifier que ces fonctions sont libres pour que celles-ci forment une base de $\mathbb{P}_{2k-1}$. Soit $(a_0, \ldots, a_{2k-1}) \in \mathbb{R}^{2k}$ tel que

$$P = \sum_{i=0}^{2k-1} a_i \Lambda_1^{2k-1-i}\Lambda_2^i = 0.$$

Montrons que $a_0 = \cdots = a_{2k-1} = 0$. Étant donné que

$$P(0) = \sum_{i=0}^{2k-1} a_i \Lambda_1^{2k-1-i}(0)\Lambda_2^i(0) = a_0,$$

et

$$P(1) = \sum_{i=0}^{2k-1} a_i \Lambda_1^{2k-1-i}(1)\Lambda_2^i(1) = a_{2k-1},$$

nous avons directement que $a_0 = a_{2k-1} = 0$. Alors

$$P = \sum_{i=1}^{2k-2} a_i \Lambda_1^{2k-1-i}\Lambda_2^i = \Lambda_1\Lambda_2 \sum_{i=1}^{2k-2} a_i \Lambda_1^{2k-2-i}\Lambda_2^{i-1} = 0.$$

Comme $\Lambda_1\Lambda_2 \neq 0$, nous avons que

$$\tilde{P} = \sum_{i=1}^{2k-2} a_i \Lambda_1^{2k-2-i}\Lambda_2^{i-1} = 0.$$

Il suffit alors d'évaluer $\tilde{P}$ en 0 et 1 pour vérifier que $a_1 = a_{2k-2} = 0$ et ainsi de suite pour montrer que tous les coefficients des termes extrêmes de la somme sont nuls pour conclure que $a_0 = \cdots = a_{2k-1} = 0$. □

En remarquant que $\Lambda_1^{2k-1-i}\Lambda_2^i$ et $\Lambda_1^i\Lambda_2^{2k-1-i}$ pour tout $i = 0, \ldots, 2k$ sont deux fonctions symétriques :

$$\Lambda_1^{2k-1-i}(x)\Lambda_2^i(x) = (1-x)^{2k-1-i}x^i = x^i(1-x)^{2k-1-i} = \Lambda_1^i(1-x)\Lambda_2^{2k-1-i}(1-x),$$

il suffit de vérifier que la formule de quadrature intègre exactement les k premières fonctions de base décrites dans le lemme précédent pour que celle-ci soit donc exacte sur $\mathbb{P}_{2k-1}$. La forme générique d'une formule de quadrature symétrique sur $[0,1]$ pouvant être donnée par

$$I(f) = \sum_i w_i(f(x_i) + f(1-x_i))$$

nous disposons d'un certain nombre de degrés de liberté associés à la formule de quadrature, c'est-à dire un certain nombre de poids (les w_i) et de points (les x_i). Pour éviter toute ambiguïté entre les degrés de liberté associés à l'élément fini (c'est-à-dire des formes linéaires de Σ) et les degrés de liberté associés à la formule de quadrature, nous parlerons de ces derniers plutôt en termes de paramètres de la formule de quadrature.

Sachant que l'on veut transformer l'élément fini P_k standard en un nouvel élément fini que l'on notera $\tilde{P}_k$, et que celui-ci fait apparaître $k+1$ points auxquels sont associés les formes linéaires, il nous faut construire une formule de quadrature symétrique utilisant $k+1$ points et intégrant exactement les polynômes de degré $2k-1$. Regardons de plus près combien de paramètres nous donnent ces $k+1$ points. Pour vérifier la condition de conformité il est indispensable de définir les points 0 et 1 comme points de quadrature auxquels on associe un unique poids comme paramètre. Il reste alors $k-1$ points à répartir de manière symétrique dans l'intérieur de l'intervalle $[0,1]$: si k est impair on divise ces points en $\frac{k-1}{2}$ couples de points symétriques, x_i et $1-x_i$, qui engendrent chacun deux paramètres (un poids w_i et un paramètre de localisation x_i) ; si k est pair nous sommes obligés de définir le point $\frac{1}{2}$ comme point de quadrature qui nous donne un poids en tant que paramètre puis de diviser les $k-2$ points restant en $\frac{k-2}{2}$ couples de points symétriques engendrant chacun deux degrés de liberté. Que k soit pair ou impair nous disposons donc de k paramètres (en terme de poids ou de paramètres de localisation) pour déterminer une formule de quadrature symétrique exacte dans $\mathbb{P}_{2k-1}$, c'est-à-dire intégrant exactement les k fonctions de base définies plus haut.

Le système engendré par ces k égalités n'étant pas linéaire mais polynomial (il est en fait linéaire sur les poids et polynomial sur les paramètres de localisation), il n'est pas possible d'affirmer l'existence d'une solution à ce système, et donc l'existence d'une formule de quadrature de ce type permettant la condensation de masse. Remarquons toutefois qu'il est en général impossible qu'il y ait unicité de la solution dans la mesure où si $\{w_i, x_i\}$ est

une solution, on peut par exemple remplacer n'importe quel x_i par $1 - x_i$ pour avoir une autre solution, ou bien intervertir les couples $(x_i, 1 - x_i)$ par un autre couple $(x_j, 1 - x_j)$ avec leur poids (une bonne question à se poser serait de savoir s'il y a unicité modulo ces permutations, ce qui semble être le cas). Ces systèmes sont résolus à l'aide du logiciel de calcul formel MAPLE© (de manière exacte tant que celui-ci est assez robuste pour nous retourner une solution puis de manière numérique). Nous donnons dans les tableaux suivant le système à résoudre, sa (ou plutôt une) solution et la formule de quadrature qui en découle pour les éléments finis $\tilde{P}_1$ à $\tilde{P}_5$.

élément fini	$\tilde{P}_1$
forme générique de la formule de quadrature	$I(f) = w_s(f(0) + f(1))$
système	$w_s = \frac{1}{2}$
solution	$w_s = \frac{1}{2}$

élément fini	$\tilde{P}_2$
forme générique de la formule de quadrature	$I(f) = w_s(f(0) + f(1)) + w_m f(\frac{1}{2})$
système	$\begin{cases} w_s + \frac{1}{8}w_m = \frac{1}{4} \\ \frac{1}{8}w_m = \frac{1}{12} \end{cases}$
solution	$\begin{cases} w_s = \frac{1}{6} \\ w_m = \frac{2}{3} \end{cases}$

élément fini	$\tilde{P}_3$
forme générique de la formule de quadrature	$I(f) = w_s(f(0) + f(1)) + w_\alpha(f(\alpha) + f(1-\alpha))$
système	$\begin{cases} w_s + w_\alpha((1-\alpha)^5 + \alpha^5) = \frac{1}{6} \\ w_\alpha((1-\alpha)^4\alpha + \alpha^4(1-\alpha)) = \frac{1}{30} \\ w_\alpha((1-\alpha)^3\alpha^2 + \alpha^3(1-\alpha)^2) = \frac{1}{60} \end{cases}$
solution	$\begin{cases} w_s = \frac{1}{12} \\ w_\alpha = \frac{5}{12} \\ \alpha = \frac{1}{2} - \frac{1}{10}\sqrt{5} \end{cases}$

élément fini	$\tilde{P}_4$
forme générique de la formule de quadrature	$\begin{aligned} I(f) \quad &= w_s(f(0)+f(1)) \\ &+w_\alpha(f(\alpha)+f(1-\alpha)) \\ &+w_m f(\frac{1}{2}) \end{aligned}$
système	$\begin{cases} w_s + w_\alpha((1-\alpha)^7+\alpha^7) + \frac{1}{128}w_m = \frac{1}{8} \\ w_\alpha((1-\alpha)^6\alpha + \alpha^6(1-\alpha)) + \frac{1}{128}w_m = \frac{1}{56} \\ w_\alpha((1-\alpha)^5\alpha^2 + \alpha^5(1-\alpha)^2) + \frac{1}{128}w_m = \frac{1}{168} \\ w_\alpha((1-\alpha)^4\alpha^3 + \alpha^4(1-\alpha)^3) + \frac{1}{128}w_m = \frac{1}{280} \end{cases}$
solution	$\begin{cases} w_s = \frac{1}{20} \\ w_\alpha = \frac{49}{180} \\ w_m = \frac{16}{45} \\ \alpha = \frac{1}{2} - \frac{1}{14}\sqrt{21} \end{cases}$

élément fini	$\tilde{P}_5$
forme générique de la formule de quadrature	$\begin{aligned} I(f) \quad &= w_s(f(0)+f(1)) \\ &+w_{\alpha_1}(f(\alpha_1)+f(1-\alpha_1)) \\ &+w_{\alpha_2}(f(\alpha_2)+f(1-\alpha_2)) \end{aligned}$
système	$\begin{cases} w_s + w_{\alpha_1}((1-\alpha_1)^9+\alpha_1^9) + w_{\alpha_2}((1-\alpha_2)^9+\alpha_2^9) = \frac{1}{10} \\ w_{\alpha_1}((1-\alpha_1)^8\alpha_1 + \alpha_1^8(1-\alpha_1)) + w_{\alpha_2}((1-\alpha 2)^8\alpha_2 + \alpha_2^8(1-\alpha_2)) = \frac{1}{90} \\ w_{\alpha_1}((1-\alpha_1)^7\alpha_1^2 + \alpha_1^7(1-\alpha_1)^2) + w_{\alpha_2}((1-\alpha_2)^7\alpha_2^2 + \alpha_2^7(1-\alpha_2)^2) = \frac{1}{360} \\ w_{\alpha_1}((1-\alpha_1)^6\alpha_1^3 + \alpha_1^6(1-\alpha_1)^3) + w_{\alpha_2}((1-\alpha_2)^6\alpha_2^3 + \alpha_2^6(1-\alpha_2)^3) = \frac{1}{840} \\ w_{\alpha_1}((1-\alpha_1)^5\alpha_1^4 + \alpha_1^5(1-\alpha_1)^4) + w_{\alpha_2}((1-\alpha_2)^5\alpha_2^4 + \alpha_2^5(1-\alpha_2)^4) = \frac{1}{1260} \end{cases}$
solution	$\begin{cases} w_s = \frac{1}{30} \\ w_{\alpha_1} = \frac{7}{30} - \frac{1}{60}\sqrt{7} \\ w_{\alpha_2} = \frac{7}{30} + \frac{1}{60}\sqrt{7} \\ \alpha_1 = \frac{1}{2} - \frac{1}{42}\sqrt{147+42\sqrt{7}} \\ \alpha_2 = \frac{1}{2} - \frac{1}{42}\sqrt{147-42\sqrt{7}} \end{cases}$

Remarquons que les formules de quadrature que l'on a ainsi déterminées corespondent aux formules de quadrature de Gauss-Lobatto rapportés de l'intervalle $[-1,1]$ à l'intervalle $[0,1]$. Nous renvoyons le lecteur à l'annexe 7.5.7 pour plus de détails sur ces formules.

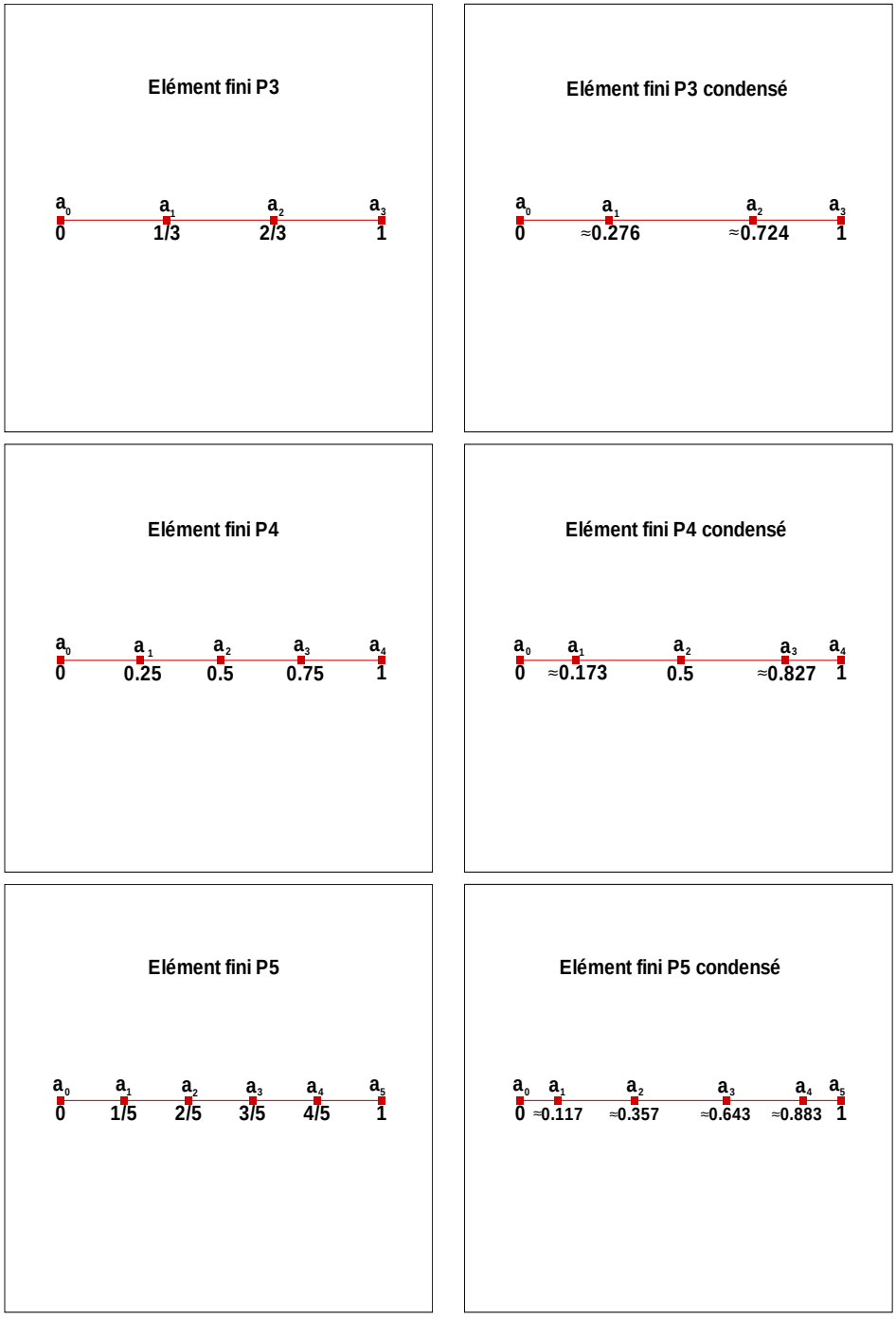

FIG. 4.1 – Passage des éléments finis P_k standards aux éléments finis $\tilde{P}_k$.

4.3 Condensation de la matrice de masse issue des éléments finis de Lagrange triangulaires

Le problème de la condensation de la matrice de masse issue des éléments finis de Lagrange en une dimension d'espace peut donc être résolu à l'aide de la théorie des polynômes orthogonaux qui nous permet de construire des formules de quadrature appropriées. Si nous avons choisi de ne pas considérer cet aspect, mais de construire "à la main" des formules de quadrature ayant de bonnes propriétés c'est pour se donner l'idée d'un algorithme que l'on pourrait généraliser pour la détermination de formules de quadrature en deux dimensions d'espace sur un triangle. En effet la construction de formules de quadrature sur un triangle n'a pas de fondements théoriques aussi aboutis que la construction de formules de quadrature sur un segment.
L'idée essentielle de la méthode est d'enrichir l'espace polynomial P définissant l'élément fini de manière à disposer de plus de formes linéaires dans Σ, c'est-à-dire de plus de degrés de liberté (et donc aussi de plus de points auxquels ceux-ci sont associés), afin de construire "à la main" une formule de quadrature adaptée. Plus précisément, en voulant modifier $\mathbb{P}_k$ (l'espace polynomial des éléments finis de Lagrange d'ordre k) en $\tilde{\mathbb{P}}_k$ de la manière suivante :

$$\mathbb{P}_k \subseteq \tilde{\mathbb{P}}_k \subset \mathbb{P}_{k'} \quad k \leq k',$$

il est possible de montrer que nous ne perdons pas en précision si la formule de quadrature est exacte à l'ordre $k + k' - 2$ (voir [39],[40] ou [17]). D'autres conditions, plus restrictives, viennent se rajouter à celle-ci :

- l'espace $\tilde{\mathbb{P}}_k$ doit être aussi petit que possible tout en contenant $\mathbb{P}_k$.
- le jeu de points de quadrature doit être $\tilde{\mathbb{P}}_k$-unisolvant.
- le nombre de points de quadrature sur les arêtes doit être suffisant pour assurer la conformité de l'élément fini.
- les poids associés à la formule de quadrature doivent être strictement positifs.

Les trois premières conditions sont intuitives : si la première n'est pas nécessaire d'un point de vue mathématique, elle vise simplement à minimiser le nombre de degrés de liberté global de la méthode, elle incite entre autre à choisir k' aussi petit que possible ; la seconde quant à elle, est nécessaire de part le fait que les degrés de liberté seront exactement les points de quadrature, et la troisième s'explique par le même argument. La quatrième condition est liée à un critère de stabilité de la méthode : des poids nuls impliqueraient des termes nuls sur la diagonale de la matrice de masse qui ne serait donc plus inversible et des poids négatifs entraîneraient un schéma inconditionnellement instable quelle que soit la semi-discrétisation en temps choisie (voir [70]).

4.3.1 Quelques remarques préliminaires sur les formules de quadrature symétriques dans un triangle

Dans la suite de cette seconde partie, nous allons discuter de la construction d'un point de vue pratique de formules de quadrature dans un triangle. Plus particulièrement, nous allons montrer comment il est possible de réduire de manière significative le système dont les inconnues déterminent les paramètres de la formule de quadrature (en terme de poids et de

localisation des points). Vouloir déterminer une formule de quadrature c'est tout d'abord se donner un ordre de précision, c'est-à-dire par exemple le degré des polynômes qui seront intégrés de manière exacte. Il est généralement reconnu qu'il est préférable d'exiger plus que simplement le fait d'intégrer de manière exacte des polynômes jusqu'à un certain degré. Par exemple en une dimension d'espace sur le segment $[-1,1]$, sachant que toute fonction impaire est d'intégrale nulle il est raisonnable de vouloir que l'évaluation d'une formule de quadrature sur une telle fonction retourne elle aussi zéro. Cela est naturellement réalisé dès lors que les points de quadrature sont symétriquement localisés dans l'intervalle $[-1,1]$ et que les deux poids associés à deux points symétriques sont égaux. Une telle formule de quadrature est dite symétrique. Une autre propriété remarquable des formules de quadrature symétriques est que les approximations des intégrales de fonctions symétriques (c'est-à-dire de fonctions vérifiant $g(x) = f(-x)\ \forall x \in [-1,1]$) sont égales. C'est exactement cette notion de symétrie qu'on l'on va généraliser pour les formules de quadrature définies sur un triangle.
Nous voulons déterminer une formule de quadrature qui intègre de manière exacte l'espace polynomial $\mathbb{P}_k$. Nous introduisons pour cela un certain nombre de notations, à savoir que $x = (x_1, x_2)$ désigne une variable de $\mathbb{R}^2$ et K un triangle du plan de sommets S_1, S_2 et S_3. Les coordonnées barycentriques par rapport à ces trois sommets seront données respectivement par $\Lambda_1(x)$, $\Lambda_2(x)$ et $\Lambda_3(x)$, c'est-à-dire que $\Lambda_i(x)$ est l'unique polynôme de $\mathbb{P}_1$ vérifiant $\Lambda_i(S_j) = \delta_{ij}$ (voir [61]). Le résultat fondamental est alors le suivant :

Lemme 4.3.1. *L'ensemble $B(\mathbb{P}_k)$ défini par*

$$B(\mathbb{P}_k) = \{\Lambda_1^l(x)\Lambda_2^m(x)\Lambda_3^n(x); 0 \leq l,m,n \leq k; l+m+n = k\}$$

est une base de $\mathbb{P}_k$.

Démonstration. Nous ne donnons ici que les idées de la démonstration.
Dans un premier temps nous montrons que le cardinal de $B(\mathbb{P}_k)$ est égal à la dimension de l'espace $\mathbb{P}_k$, c'est-à-dire $\frac{(k+1)(k+2)}{2}$ en dénombrant l'ensemble $\{0 \leq l,m,n \leq k; l+m+n = k\}$.
Il suffit ensuite de montrer que la famille d'éléments de $B(\mathbb{P}_k)$ est libre. On se donne alors une combinaison linéaire nulle de la famille.
Parmi les éléments de $B(\mathbb{P}_k)$ il n'y en a que $3k$ que l'on ne peut pas factoriser par $\Lambda_1(x)\Lambda_2(x)\Lambda_3(x)$, les autres étant identiquement nuls sur les arêtes du triangle K, et parmi ceux-là seuls $k+1$ éléments ne peuvent se factoriser par $\Lambda_3(x)$, les autres étant identiquement nuls sur l'arête du triangle K délimitée par les sommets $S1$ et $S2$. Nous sommes alors exactement dans les conditions du lemme 4.2.1 en considérant les restrictions à cette arête de ces $k+1$ éléments, qui sont donc une base de $\mathbb{P}_k([S1,S2])$ ce qui nous permet de conclure que ce sont les coefficients de ces éléments, dans la combinaison linéaire, qui sont nuls. On fait de même pour les deux autres arêtes.
Il ne reste alors dans la combinaison linéaire que les éléments de $B(\mathbb{P}_k)$ qui se factorisent par $\Lambda_1(x)\Lambda_2(x)\Lambda_3(x)$, facteur que l'on peut simplifier, celui-ci n'étant pas identiquement nul.
Il suffit ensuite de recommencer l'opération, tant qu'il reste dans la combinaison linéaire des polynômes se factorisant par $\Lambda_1(x)\Lambda_2(x)\Lambda_3(x)$, pour montrer que tous les coefficients de la combinaison linéaire sont nuls. □

Cela implique en particulier que toute fonction donnée $f \in \mathbb{P}_k$ peut être exprimée comme une unique combinaison linéaire de fonctions de $B(\mathbb{P}_k)$, ou plus précisément qu'il existe une unique fonction $\hat{f}$ telle que :

$$f(x) = \hat{f}(\Lambda_1(x), \Lambda_2(x), \Lambda_3(x)) \quad \forall x \in \mathbb{R}^2.$$

Considérons alors $\mathbb{S}_3$ le groupe des permutations sur $\{1, 2, 3\}$. Remarquant que

$$\int_K \Lambda_i(x)\, dx = \int_K \Lambda_{\sigma(i)}(x)\, dx \quad \forall i = 1, 2, 3 \text{ et } \forall \sigma \in \mathbb{S}_3,$$

il devient évident que $\forall \sigma \in \mathbb{S}_3, (l, m, n) \in \mathbb{N}^3$,

$$\int_K \Lambda_1^l(x)\Lambda_2^m(x)\Lambda_3^n(x)\, dx = \int_K \Lambda_{\sigma(1)}^l(x)\Lambda_{\sigma(2)}^m(x)\Lambda_{\sigma(3)}^n(x)\, dx,$$

de sorte que $\forall \sigma \in \mathbb{S}_3$,

$$\int_K f(x)\, dx = \int_K \hat{f}(\Lambda_1(x), \Lambda_2(x), \Lambda_3(x))\, dx = \int_K \hat{f}(\Lambda_{\sigma(1)}(x), \Lambda_{\sigma(2)}(x), \Lambda_{\sigma(3)}(x))\, dx. \tag{4.1}$$

Nous demanderons donc à notre formule de quadrature de vérifier cette dernière propriété de symétrie par rapport au groupe des permutations à trois éléments.
Il est bien connu que tout point $(x_1, x_2) \in K$ peut être localisé en terme de coordonnées barycentriques $(\lambda_1, \lambda_2, \lambda_3)$ par rapport aux trois sommets de K et que cette localisation devient unique dès lors que l'on impose la contrainte $\lambda_1 + \lambda_2 + \lambda_3 = 1$. Nous dirons alors que deux points $(x_1, x_2) \in K$ et $(\hat{x}_1, \hat{x}_2) \in K$ de coordonnées barycentriques respectives $(\lambda_1, \lambda_2, \lambda_3)$ et $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3)$ sont symétriques s'il existe $\sigma \in \mathbb{S}_3$ tel que

$$(\lambda_1, \lambda_2, \lambda_3) = (\hat{\lambda}_{\sigma(1)}, \hat{\lambda}_{\sigma(2)}, \hat{\lambda}_{\sigma(3)}).$$

Il devient alors évident que toute formule de quadrature I_K définie par un jeu de points symétriques Q, c'est-à-dire un jeu de points vérifiant

$$(\lambda_1, \lambda_2, \lambda_3) \in Q \Rightarrow (\lambda_{\sigma(1)}, \lambda_{\sigma(2)}, \lambda_{\sigma(3)}) \in Q \quad \forall \sigma \in \mathbb{S}_3,$$

et des poids associés vérifiant

$$w(\lambda_1, \lambda_2, \lambda_3) = w(\lambda_{\sigma(1)}, \lambda_{\sigma(2)}, \lambda_{\sigma(3)}) \quad \forall \sigma \in \mathbb{S}^3,$$

où bien entendu $w(\lambda_1, \lambda_2, \lambda_3)$ désigne le poids associé au point de coordonnées barycentriques $(\lambda_1, \lambda_2, \lambda_3)$, vérifie la version discrète de la propriété de symétrie (4.1) :

$$I_K(\hat{f}(\Lambda_1(x), \Lambda_2(x), \Lambda_3(x))) = I_K(\hat{f}(\Lambda_{\sigma(1)}(x), \Lambda_{\sigma(2)}(x), \Lambda_{\sigma(3)}(x))) \quad \forall \sigma \in \mathbb{S}_3.$$

En effet, le seul point qu'il faut vérifier est que

$$I_K(\Lambda_i(x)) = I_K(\Lambda_{\sigma(i)}(x)) \quad \forall \sigma \in \mathbb{S}_3, i = 1, 2, 3,$$

ce qui est trivial et qui implique immédiatement que

$$I_K(\Lambda_1^l(x)\Lambda_2^m(x)\Lambda_3^n(x)) = I_K(\Lambda_{\sigma(1)}^l(x)\Lambda_{\sigma(2)}^m(x)\Lambda_{\sigma(3)}^n(x)) \quad \forall \sigma \in \mathbb{S}_3, (l, m, n) \in \mathbb{N}^3.$$

Une telle formule de quadrature sera alors dite symétrique.
Si l'on veut maintenant assurer qu'une formule de quadrature symétrique intègre de manière exacte toute fonction polynomiale de $\mathbb{P}_k$ il suffit d'assurer que celle-ci intègre de manière exacte l'ensemble des fonctions
$\Lambda_1^l(x)\Lambda_2^m(x)\Lambda_3^n(x) \in B(\mathbb{P}_k)$. Or sachant que

$$\begin{aligned} I_K(\Lambda_1^l(x)\Lambda_2^m(x)\Lambda_3^n(x)) =& I_K(\Lambda_{\sigma(1)}^l(x)\Lambda_{\sigma(2)}^m(x)\Lambda_{\sigma(3)}^n(x)) \\ & \forall \sigma \in \mathbb{S}_3, \Lambda_1^l(x)\Lambda_2^m(x)\Lambda_3^n(x) \in B(\mathbb{P}_k), \end{aligned}$$

il est évident que l'on a pas besoin d'assurer l'intégration exacte des $\frac{(k+1)(k+2)}{2}$ fonctions de base de $B(\mathbb{P}_k)$ mais seulement pour un représentant de chaque classe d'équivalence de $B(\mathbb{P}_k)/\sim$, où la relation d'équivalence $\sim$ est définie par :

$$\begin{gathered} \Lambda_1^l(x)\Lambda_2^m(x)\Lambda_3^n(x) \sim \Lambda_1^{\hat{l}}(x)\Lambda_2^{\hat{m}}(x)\Lambda_3^{\hat{n}}(x) \\ \Longleftrightarrow \\ \exists \sigma \in \mathbb{S}_3 \ / \ \Lambda_1^l(x)\Lambda_2^m(x)\Lambda_3^n(x) = \Lambda_{\sigma(1)}^{\hat{l}}(x)\Lambda_{\sigma(2)}^{\hat{m}}(x)\Lambda_{\sigma(3)}^{\hat{n}}(x). \end{gathered}$$

Pour illustrer ces propos nous donnons ici en exemple les classes d'équivalence $C_i^k, i = 1 \dots \tilde{M}_k$ de $B(\mathbb{P}_k)$ par espace polynomial $\mathbb{P}_k$ pour les cinq premières valeurs de k, où $\tilde{M}_k$ désigne alors naturellement le nombre de classes d'équivalence de $B(\mathbb{P}_k)$ et N_k désigne le nombre de degrés de liberté associé à l'élément fini P_k (qui correspond aussi à la dimension de l'espace polynomial $\mathbb{P}_k$ qui lui est associé) :

(i) $k=1, \quad N_1=3, \quad \tilde{M}_1=1, \quad C_1^1 = \{\Lambda_1, \Lambda_2, \Lambda_3\}$

(ii) $k=2, \quad N_2=6, \quad \tilde{M}_2=2, \quad \begin{cases} C_1^2 = \{\Lambda_1^2, \Lambda_2^2, \Lambda_3^2\} \\ C_2^2 = \{\Lambda_1\Lambda_2, \Lambda_1\Lambda_3, \Lambda_2\Lambda_3\} \end{cases}$

(iii) $k=3, \quad N_3=10, \quad \tilde{M}_3=3, \quad \begin{cases} C_1^3 = \{\Lambda_1^3, \Lambda_2^3, \Lambda_3^3\} \\ C_2^3 = \{\Lambda_1^2\Lambda_2, \Lambda_1^2\Lambda_3, \Lambda_2^2\Lambda_1, \Lambda_2^2\Lambda_3, \Lambda_3^2\Lambda_1, \Lambda_3^2\Lambda_2\} \\ C_3^3 = \{\Lambda_1\Lambda_2\Lambda_3\} \end{cases}$

(iv) $k=4, \quad N_4=15, \quad \tilde{M}_4=4, \quad \begin{cases} C_1^4 = \{\Lambda_1^4, \Lambda_2^4, \Lambda_3^4\} \\ C_2^4 = \{\Lambda_1^3\Lambda_2, \Lambda_1^3\Lambda_3, \Lambda_2^3\Lambda_1, \Lambda_2^3\Lambda_3, \Lambda_3^3\Lambda_1, \Lambda_3^3\Lambda_2\} \\ C_3^4 = \{\Lambda_1^2\Lambda_2^2, \Lambda_1^2\Lambda_3^2, \Lambda_2^2\Lambda_3^2, \} \\ C_4^4 = \{\Lambda_1^2\Lambda_2\Lambda_3, \Lambda_2^2\Lambda_1\Lambda_3, \Lambda_3^2\Lambda_1\Lambda_2\} \end{cases}$

(v) $k=5, \quad N_5=21, \quad \tilde{M}_5=5, \quad \begin{cases} C_1^5 = \{\Lambda_1^5, \Lambda_2^5, \Lambda_3^5\} \\ C_2^5 = \{\Lambda_1^4\Lambda_2, \Lambda_1^4\Lambda_3, \Lambda_2^4\Lambda_1, \Lambda_2^4\Lambda_3, \Lambda_3^4\Lambda_1, \Lambda_3^4\Lambda_2\} \\ C_3^5 = \{\Lambda_1^3\Lambda_2^2, \Lambda_1^3\Lambda_3^2, \Lambda_2^3\Lambda_1^2, \Lambda_2^3\Lambda_3^2, \Lambda_3^3\Lambda_1^2, \Lambda_3^3\Lambda_2^2, \} \\ C_4^5 = \{\Lambda_1^3\Lambda_2\Lambda_3, \Lambda_2^3\Lambda_1\Lambda_3, \Lambda_3^3\Lambda_1\Lambda_2\} \\ C_5^5 = \{\Lambda_1^2\Lambda_2^2\Lambda_3, \Lambda_1^2\Lambda_3^2\Lambda_2, \Lambda_2^2\Lambda_3^2\Lambda_1\} \end{cases}$

Remarque 4.3.2. *Il serait faux de croire qu'en toute généralité $\tilde{M}_k = k$ puisqu'en particulier $\tilde{M}_6 = 7$.*

4.3.2 Construction pratique des éléments finis condensés

Le premier point dont il nous faut discuter dans cette partie est le suivant : quelles sont les différentes classes d'équivalence que la contrainte de symétrie nous permet de considérer,

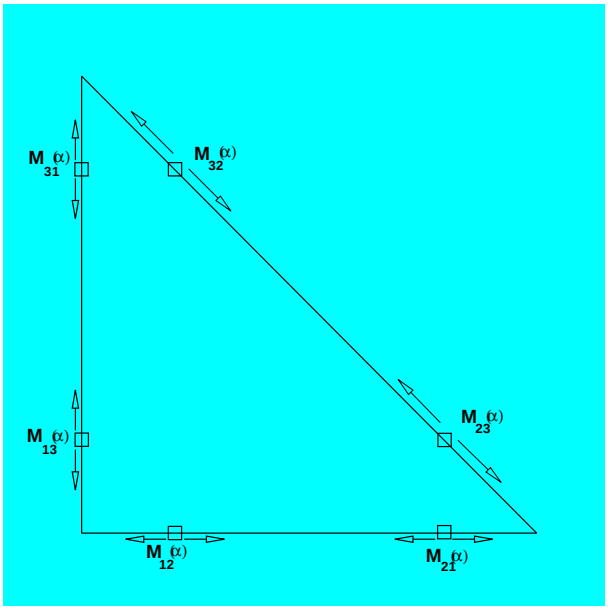

FIG. 4.2 – Le jeu de six points localisés sur les arêtes.

et le nombre de paramètres que leur utilisation implique ?
Les classes d'équivalence composées de points localisés sur les arêtes ne sont que de trois types :

- le premier type est bien entendu la classe composée par les **trois sommets** S_i. Le jeu de points composant cette classe peut être entièrement déterminé par un unique paramètre w_s qui correspond au poids associé aux trois points,
- le second type est la classe composée par les **trois points milieux des trois arêtes** M_i, auquel nous associons un poids w_m,
- le dernier type est composé par un jeu de **six points** M_{ij} localisés par un paramètre α. Plus précisément $M_{ij}(\alpha)$ **est le barycentre des deux sommets** S_i **and** S_j **respectivement pondérés par** α **et** $1-\alpha$**, où** $\alpha \in]0, \frac{1}{2}[$. Dans ce cas la classe d'équivalence est donc déterminée de manière unique par deux paramètres, le poids w_α associé aux six points et le paramètre de localisation α(voir figure 4.2).

Donnons maintenant les trois types de classe d'équivalence de points localisés à l'intérieur du triangle :

- la classe composée par un unique point, le **barycentre** du triangle, pondéré par w_g,
- la classe composée par **trois points** G_i **localisés sur les médianes** par un paramètre β ; **où** $G_i(\beta)$ **est le barycentre des trois sommets,** S_i **pondéré par** β **et les deux autres pondérés par** $\frac{1-\beta}{2}$, où $\beta \in]0,1[$ et $\beta \neq \frac{1}{3}$ de manière à ce que cette classe d'équivalence ne dégénère pas en l'unique barycentre du triangle. Cette classe d'équivalence est donc déterminée par deux paramètres : un poids w_β et un paramètre de localisation β (voir figure 4.3),
- la dernière classe composée par **six points** G_{ij} localisés cette fois par **deux paramètres** ω_1 et ω_2 dans $]0,1[$ où $\omega_1 \neq \omega_2$ (de manière à ne pas dégénérer en

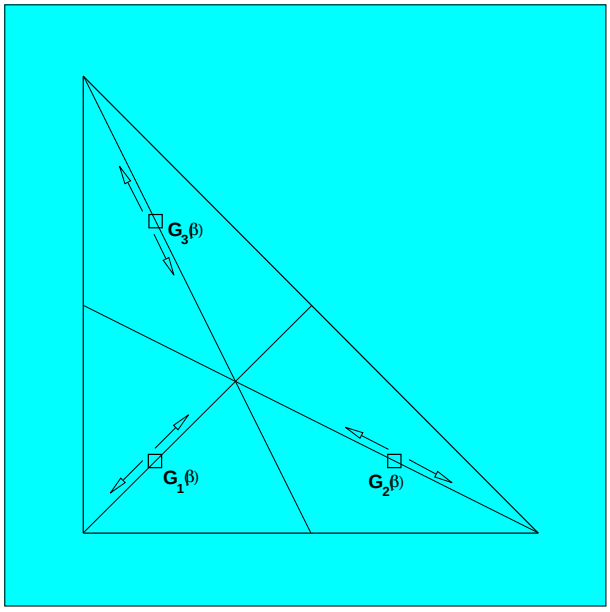

FIG. 4.3 – Le jeu de trois points symétriques intérieurs au triangle.

l'une des deux dernières classes mentionnées). Les points $G_{ij}(\omega_1, \omega_2)$ **sont les barycentres des trois sommets** S_i, S_j **et** S_k **respectivement pondérés par** ω_1, ω_2 **et** $1 - \omega_1 - \omega 2$. Cette classe d'équivalence sera donc entièrement définie par trois paramètres : un poids w_ω, et deux paramètres de localisation ω_1 et ω_2 (voir figure 4.4).

Dans l'optique de condenser la matrice de masse des éléments finis de Lagrange standards P_k, il nous faut considérer un espace $\tilde{\mathbb{P}}_k$ tel que $\mathbb{P}_k \subseteq \tilde{\mathbb{P}}_k \subset \mathbb{P}_{k'}$. Une fois cet espace fixé, le nombre de points que l'on doit utiliser pour la formule de quadrature à associer est uniquement déterminé par la dimension de cet espace polynomial. Il nous est donc possible de scinder ces points de quadrature en un certain nombre de classes d'équivalence et ainsi d'exhiber formellement la formule de quadrature associée et le nombre de paramètres dont celle-ci nous permet de disposer. Mais jusqu'à maintenant nous ne savons toujours pas quels espaces $\tilde{\mathbb{P}}_k$ sont potentiellement convenables à la condensation de masse. C'est en fait la condition de conformité qui va déterminer leur forme générale. Le fait même qu'il nous faut respecter cette condition nous amène à ne considérer que des espaces $\tilde{\mathbb{P}}_k$ pour lesquels l'ensemble des formes linéaires a une chance d'être unisolvant lorsque ces formes linéaires sont associées à un jeu de points faisant apparaître exactement autant de points sur les arêtes que ceux associés à l'élément fini P_k, c'est-à dire $3k$ (dont $k-1$ sur chaque arête, plus les 3 sommets), de manière à, d'une part assurer la continuité des éléments à travers les arêtes du maillage, d'autre part assurer de ne pas contenir entièrement $\mathbb{P}_{k+1}$. Ceci nous amène à une détermination plus précise des espaces polynomiaux à considérer : si nous voulons conserver l'unisolvance de l'élément fini il nous faut enrichir l'espace $\mathbb{P}_k$ par un espace de polynômes de degré supérieur à k qui s'annulent identiquement sur les

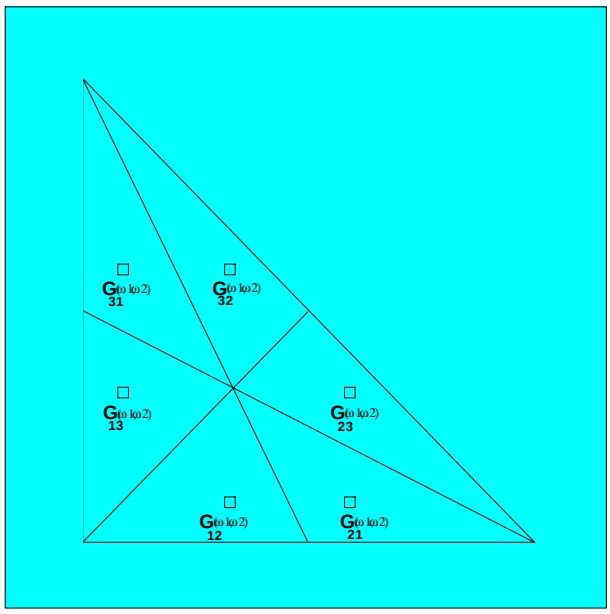

FIG. 4.4 – Le jeu de six points symétriques intérieurs au triangle.

arêtes du triangle. Ce nouvel espace aura alors une forme du type :

$$\tilde{\mathbb{P}}_k = \mathbb{P}_k + b\mathbb{P}_{\tilde{k}},$$

où b est la fonction dite fonction bulle $\Lambda_1\Lambda_2\Lambda_3$ qui s'annule identiquement sur les arêtes du triangle. Soulignons que cette expression n'a de sens que pour $k \geq 0$, $\tilde{k} \geq 0$ et $\tilde{k} \geq k-2$: d'une part il faut bien entendu que les espaces considérés soient des espaces de polynômes (d'où les conditions $k \geq 0$ et $\tilde{k} \geq 0$), d'autre part pour que l'espace $\mathbb{P}_k$ soit effectivement enrichi il faut que $\tilde{k} \geq k-2$ (b étant un polynôme de degré 3), sinon nous ajouterions à $\mathbb{P}_k$ un espace qui est déjà inclus dans $\mathbb{P}_k$. À ce moment nous aurons

$$\mathbb{P}_k \subset \tilde{\mathbb{P}}_k \subset \mathbb{P}_{\tilde{k}+3}.$$

Une autre manière d'enrichir l'espace polynomial serait par exemple de considérer

$$\tilde{\mathbb{P}}_k = \mathbb{P}_k + b^2\mathbb{P}_{\tilde{k}},$$

ce qui n'a de sens que pour $k \geq 0$, $\tilde{k} \geq 0$ et $\tilde{k} \geq k-5$ (b^2 étant cette fois ci un polynôme de degré 6). Ce nouvel espace polynomial vérifie alors

$$\mathbb{P}_k \subset \tilde{\mathbb{P}}_k \subset \mathbb{P}_{\tilde{k}+6}.$$

Il est assez aisé de généraliser cette construction en faisant attention à ce que les espaces ainsi définis aient bien un sens (les deux seules conditions à vérifier étant que les espaces définis soient bien des espaces polynomiaux, et à veiller à effectivement enrichir les espaces). Par exemple tout espace du type

$$\tilde{\mathbb{P}}_k = \mathbb{P}_k + b(\mathbb{P}_{\tilde{k}} + b\mathbb{P}_{\tilde{\tilde{k}}}),$$

ou

$$\tilde{\mathbb{P}}_k = \mathbb{P}_k + b(\mathbb{P}_{\tilde{k}} + b^2\mathbb{P}_{\tilde{\tilde{k}}}),$$

ou

$$\tilde{\mathbb{P}}_k = \mathbb{P}_k + b^2(\mathbb{P}_{\tilde{k}} + b\mathbb{P}_{\tilde{\tilde{k}}}),$$

est potentiellement un bon candidat du moment qu'il a un sens.

Nous pouvons maintenant donner une approche pratique pour condenser les éléments finis P_k :

- dans un premier temps nous choisissons un espace polynomial $\tilde{\mathbb{P}}_k$ entre $\mathbb{P}_k$ et $\mathbb{P}_{k'}$,
- nous déterminons ensuite le nombre de paramètres qui doivent apparaître formellement dans la formule de quadrature (en termes de poids et de paramètres de localisation) de manière à ce que celle-ci soit d'un ordre assez élevé,
- nous construisons formellement une formule de quadrature qui utilise un jeu de points de quadrature symétrique faisant apparaître ce nombre de paramètres,
- nous déterminons finalement les paramètres de la formule de quadrature en résolvant un système polynomial.

Il n'y a que deux types de problèmes qui peuvent apparaître et nous empêcher de condenser les éléments P_k :

- l'espace polynomial que l'on considère ne nous permet pas de disposer d'assez de paramètres pour la formule de quadrature, c'est-à-dire que le jeu de classes d'équivalence constitué par les points de quadrature (qui sont au nombre de la dimension de l'espace polynomial) ne fait pas apparaître assez de poids et de paramètres de localisation pour que le système polynomial ait une chance d'avoir une solution,
- le système polynomial n'admet pas de solution convenable.

Pour ces deux problèmes nous serons obligés de recommencer en considérant, soit un autre espace polynomial entre $\mathbb{P}_k$ et $\mathbb{P}_{k'}$, puisqu'il peut y en avoir plusieurs, soit en considérant un k' plus grand.

Le choix initial de l'espace polynomial $\tilde{\mathbb{P}}_k$ est bien entendu $\tilde{\mathbb{P}}_k = \mathbb{P}_k$. Pour ce choix il nous faut déterminer une formule de quadrature exacte pour les polynômes de degré inférieur ou égal à $2k-2$. S'il apparaît que ceci est impossible, il faut alors enrichir $\tilde{\mathbb{P}}_k$. Nous exhibons alors tous les espaces $\tilde{\mathbb{P}}_k$ du type décrit précédemment, tels que

$$\mathbb{P}_k \subset \tilde{\mathbb{P}}_k \subset \mathbb{P}_{k+1},$$

et essayons pour chacun (en commençant par celui de plus petite dimension) de déterminer une formule de quadrature exacte pour les polynômes de degré inférieur ou égal à $k + (k+1) - 2$. Si ceci n'est possible pour aucun de ces espaces, nous continuons cette procédure en considérant un k' plus grand que $k+1$ et tous les espaces $\tilde{\mathbb{P}}_k$ tels que

$$\mathbb{P}_k \subset \tilde{\mathbb{P}}_k \subset \mathbb{P}_{k'}.$$

Rappelons que le nombre de points de quadrature qu'il faut utiliser correspond à la dimension de l'espace considéré, et que le nombre de paramètres qui doivent apparaître dans

la formule de quadrature est déterminé en calculant le nombre $\tilde{M}$ de classes d'équivalence dans $B(\mathbb{P}_{k+k'-2})$.
Une fois connu le nombre de paramètres dont nous avons besoin, nous pouvons voir s'il est possible de construire une formule de quadrature qui fait apparaître ce nombre de paramètres et utilisant le nombre donné de points de quadrature. Si cela est effectivement possible, nous disposons formellement d'une formule de quadrature, et il ne nous reste plus qu'à en déterminer les paramètres en résolvant le système polynomial composé par les égalités entre l'évaluation de la formule de quadrature et l'intégration exacte d'un représentant de chacune des classes d'équivalence de $B(\mathbb{P}_{k+k'-2})$.

Remarque 4.3.3. *Il ne faut pas croire que la construction de la formule de quadrature dépend d'un quelconque choix, la forme générique de celle-ci étant fixée par l'espace $\tilde{\mathbb{P}}_k$. En effet les points de quadrature ayant pour vocation à être les points auxquels on associe les degrés de liberté, leur localisation doit permettre à ces formes linéaires d'être $\tilde{\mathbb{P}}_k$-unisolvant. Ce que nous avons passé sous silence jusqu'à présent c'est que la donnée d'un espace $\tilde{\mathbb{P}}_k$ du type décrit précédemment nous permet de déterminer exactement quels jeux de points symétriques (et en quel nombre) feront de l'ensemble des formes linéaires un ensemble $\tilde{\mathbb{P}}_k$-unisolvant. C'est pourquoi une fois fixé $\tilde{\mathbb{P}}_k$, nous parlerons de "la" formule de quadrature. Dans la pratique c'est quelque chose qui se fait intuitivement : par exemple en considérant l'espace $\mathbb{P}_2 + b\mathbb{P}_2$ on a naturellement envie de localiser les douze points de quadrature (cet espace polynomial étant de dimension égale à douze) de manière symétrique comme dans le premier diagramme de la figure (4.5), de sorte que la formule de quadrature prenne la forme générique suivante*

$$\begin{aligned} I_K^{app}(f) =& mes(K)\{w_s(f(S_1)+f(S_2)+f(S_3))+ \\ & w_m(f(M_1)+f(M_2)+f(M_3))+ \\ & w_{\beta_1}(f(G_1(\beta_1))+f(G_2(\beta_1))+f(G_3(\beta_1)))+ \\ & w_{\beta_2}(f(H_1(\beta_2))+f(H_2(\beta_2))+f(H_3(\beta_2)))\}, \end{aligned}$$

et nous permet de disposer de 6 paramètres (4 poids et 2 paramètres de localisation). Or nous pourrions très bien imaginer répartir les points de manière symétrique plutôt suivant le deuxième schéma de la figure (4.5), ce qui génère une formule de quadrature à 5 paramètres (3 poids et 2 paramètres de localisation) de la forme

$$\begin{aligned} I_K^{app}(f) =& mes(K)\{w_s(f(S_1)+f(S_2)+f(S_3))+ \\ & w_m(f(M_1)+f(M_2)+f(M_3))+ \\ & w_\omega \sum_{j=1}^{6} f(G_j(\omega_1,\omega_2))\}. \end{aligned}$$

Cette deuxième construction est du point de vue des contraintes de symétrie, du nombre de points (global et sur les arêtes), tout aussi légitime que la première, mais un tel jeu de point ne peut faire de l'ensemble des formes linéaires qui lui sont associés un ensemble

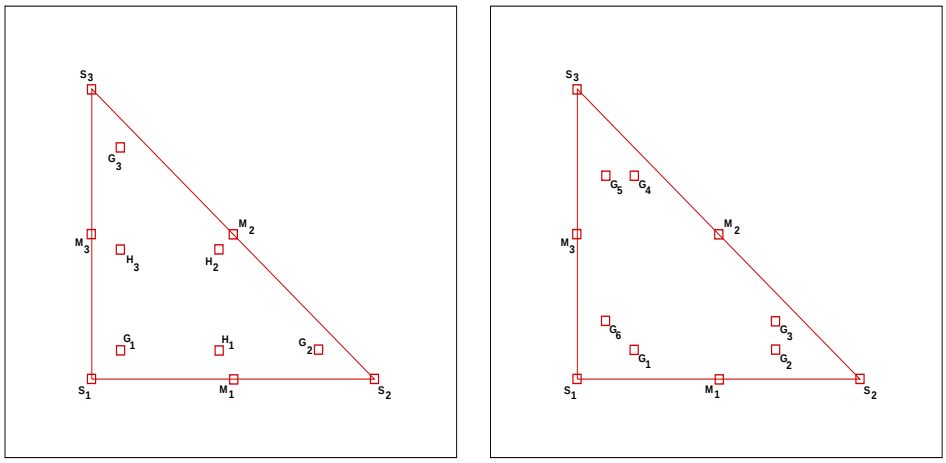

FIG. 4.5 – Les deux possibilités de répartir les douze points auxquels seront associés les degrés de liberté de l'élément fini basé sur l'espace polynomial $\mathbb{P}_2 + b\mathbb{P}_2$, de manière symétrique, et respectant la contrainte de conformité (issue de $\mathbb{P}_2$).

$\mathbb{P}_2 + b\mathbb{P}_2$-*unisolvant. Pour voir ceci on peut par exemple exhiber une base de* $\mathbb{P}_2 + b\mathbb{P}_2$ *:*

$$\left\{ \begin{array}{l} \Lambda_1^2, \Lambda_2^2, \Lambda_3^2, \\ \Lambda_1\Lambda_2, \Lambda_1\Lambda_3, \Lambda_2\Lambda_3, \\ \Lambda_1^3\Lambda_2\Lambda_3, \Lambda_1\Lambda_2^3\Lambda_3, \Lambda_1\Lambda_2\Lambda_3^3, \\ \Lambda_1^2\Lambda_2^2\Lambda_3, \Lambda_1^2\Lambda_2\Lambda_3^2, \Lambda_1\Lambda_2^2\Lambda_3^2 \end{array} \right\}$$

et suivre une argumentation similaire à celle de la démonstration du lemme 4.3.1.

Nous allons à présent appliquer cet algorithme pour condenser les éléments finis de Lagrange triangulaires. Les éléments finis condensés que l'on va déterminer vont coïncider exactement avec les éléments finis déjà connus et décrits par N. Tordjman [70] pour la condensation des éléments finis P_1 à P_3 et par W. A. Mulder [14] pour la condensation des éléments finis P_4 et P_5.

4.3.3 L'exemple de P_1

Il semble naturel de considérer $k' = 1$ et de considérer la formule des trapèzes comme formule de quadrature, i.e :

$$I_K^{app}(f) = \frac{mes(K)}{3} \left\{ f(S_1) + f(S_2) + f(S_3) \right\},$$

où S_1, S_2 et S_3 désignent les trois sommets du triangle K. Il s'avère que toutes les conditions sont vérifiées.

4.3.4 L'exemple de P_2

Nous allons dans un premier temps considérer $k' = 2$. Il nous faut alors déterminer, comme nous l'avons vu précédemment, une formule de quadrature qui intègre exactement les polynômes de degré inférieur ou égal à deux (2+2-2), et pour cela elle doit faire apparaître deux paramètres (poids ou localisations des points de quadrature), le nombre de classes d'équivalence dans $B(\mathbb{P}_2)$ étant de deux. Dans la mesure où il n'y a pas d'autre choix possible, les points de quadrature sont définis comme étant les degrés de liberté usuels associés à P_2 (voir partie gauche de la figure 4.6). La formule de quadrature est alors nécessairement de la forme

$$I_K^{app}(f) = mes(K)\left\{w_s(f(S_1) + f(S_2) + f(S_3)) + w_m(f(M_1) + f(M_2) + f(M_3))\right\},$$

et fait alors effectivement apparaître les deux paramètres qui sont les deux poids w_s et w_m. Il apparaît, après résolution, que la seule solution est

$$w_s = 0 \quad et \quad w_m = \tfrac{1}{3}.$$

Malheureusement, comme les poids associés aux sommets du triangle sont nuls, les fonctions de base correspondantes ont une norme discrète nulle, ce qui implique que certains termes diagonaux de la matrice de masse sont nuls, et donc sa non-inversibilité.

Nous considérons alors $k' = 3$, c'est-à-dire que nous devons déterminer les espaces $\tilde{\mathbb{P}}_2$ convenables (dans le sens que nous avons décrit précédemment) vérifiant

$$\mathbb{P}_2 \subset \tilde{\mathbb{P}}_2 \subset \mathbb{P}_3.$$

Le seul espace convenable de ce type est

$$\tilde{\mathbb{P}}_2 = \mathbb{P}_2 + b\mathbb{P}_0.$$

Nous cherchons alors à déterminer une formule de quadrature exacte pour les polynômes de degré inférieur ou égal à trois (2+3-2). Le nombre de classes d'équivalence dans $B(\mathbb{P}_3)$ étant de trois, la formule de quadrature doit faire apparaître trois paramètres.
Le seul jeu de points convenable dans $\mathbb{P}_2 + b\mathbb{P}_0$ est représenté dans la partie droite de la figure 4.6 et la formule de quadrature associée

$$I_K^{app}(f) = mes(K)\left\{w_s\sum_{j=1}^{3} f(S_j) + w_m\sum_{j=1}^{3} f(M_j) + w_g f(G)\right\},$$

nous fait bien bénéficier des trois paramètres nécessaires (les trois poids w_s, w_m et w_g). Il s'avère qu'il existe une unique solution au problème qui de plus satisfait la condition de positivité et de précision :

$$w_s = \tfrac{1}{20}, \quad w_m = \tfrac{2}{15}, \quad w_g = \tfrac{9}{20}.$$

Comme de plus le nombre de points de quadrature est optimal (nous n'en avons rajouté qu'un seul par rapport à une formule qui, nous l'avons montré, ne peut vérifier les conditions requises au mass-lumping), et que les autres conditions sont clairement vérifiées, nous avons effectivement obtenu le résultat cherché.

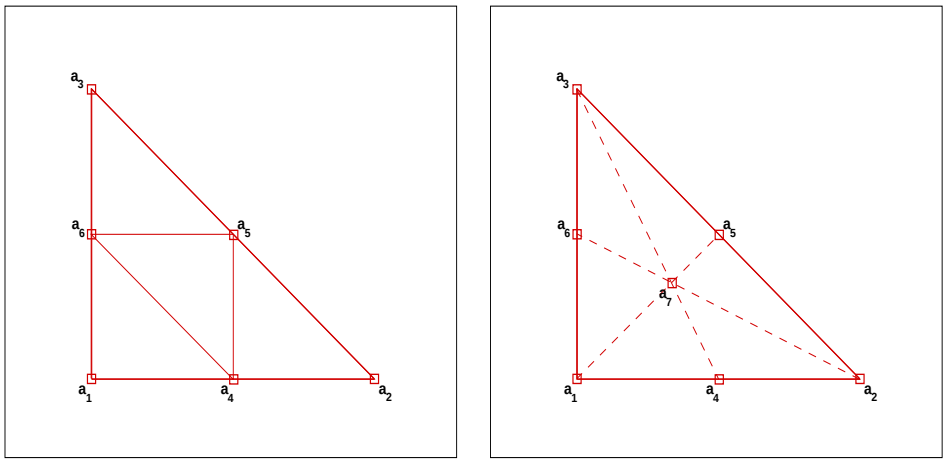

FIG. 4.6 – Localisation des noeuds de l'élément fini P_2 standard et P_2 condensé.

4.3.5 L'exemple de P_3

À nouveau, nous considérons $k' = 3$ et nous cherchons une formule de quadrature exacte pour les polynômes de degré inférieur ou égal à quatre (3+3-2). Comme nous l'avons vu, le nombre de classes d'équivalence dans $B(\mathbb{P}_4)$ est de quatre, ce qui signifie qu'il nous faut déterminer une formule de quadrature à quatre degrés de liberté. La seule combinaison de jeux de points symétriques faisant de l'ensemble des formes linéaires un ensemble P_3-unisolvant tout en respectant les contraintes liées à la conformité de l'élément fini est donné par :

- les sommets $\{S_1,S_2,S_3\}$,
- les six points situés sur les côtés
  $\{M_{12}(\alpha), M_{21}(\alpha), M_{13}(\alpha), M_{31}(\alpha), M_{23}(\alpha), M_{32}(\alpha)\}$,
- le barycentre du triangle $\{G\}$,

où α désigne un paramètre réel entre 0 et 1 et $M_{ij}(\alpha)$ est le barycentre de Si et Sj de poids respectifs α et $1-\alpha$. Nous rappelons que la localisation habituelle des degrés de liberté de P_3 correspond à $\alpha = \frac{1}{3}$ (voir partie gauche de la figure 4.7). La formule de quadrature associée est de la forme

$$I_K^{app}(f) = mes(K)\left\{ w_s \sum_{j=1}^{3} f(S_j) + w_\alpha \sum_{\substack{i,j=1 \\ i \neq j}}^{3} f(M_{ij}(\alpha)) + w_g f(G) \right\},$$

et fait bien apparaître quatre paramètres, α, w_s, w_α et w_g. La résolution du système composé des égalités entre l'évaluation par la formule de quadrature et l'intégration exacte d'un représentant de chaque classe d'équivalence de $B(\mathbb{P}_4)$ nous renvoie une unique solution :

$$\alpha = \tfrac{3-\sqrt{3}}{6}, \quad w_s = -\tfrac{1}{60}, \quad w_\alpha = \tfrac{1}{10}, \quad w_g = \tfrac{9}{20}.$$

La formule de quadrature correspondante a la fâcheuse propriété d'avoir l'un de ses poids, celui associé aux sommets, strictement négatif. Ceci a pour conséquence de rendre le schéma instable, et ceci quelle que soit la semi-discrétisation en temps utilisée (voir [70]).

On essaie alors de construire un espace $\tilde{\mathbb{P}}_3$ plus grand en considérant $k' = 4$. Le seul espace $\tilde{\mathbb{P}}_3$ convenable vérifiant $\mathbb{P}_3 \subset \tilde{\mathbb{P}}_3 \subset \mathbb{P}_4$ est

$$\tilde{\mathbb{P}}_3 = \mathbb{P}_3 + b\mathbb{P}_1.$$

Il nous faut déterminer une formule de quadrature qui intègre exactement les polynômes de degré inférieur ou égal à cinq (3+4-2). Pour cela il nous faut déterminer une formule de quadrature à cinq paramètres (ce qui correspond au nombre de classes d'équivalence dans $B(\mathbb{P}_5)$). Le jeu de points de quadrature à considérer représenté par la partie droite de la figure 4.7 est le suivant :

- les sommets $\{S_1,S_2,S_3\}$,
- les six points situés sur les côtés
  $\{M_{12}(\alpha), M_{21}(\alpha), M_{13}(\alpha), M_{31}(\alpha), M_{23}(\alpha), M_{32}(\alpha)\}$,
- les trois points intérieurs au triangle $\{G_1(\beta), G_2(\beta), G_3(\beta), \}$,

où $G_i(\beta)$ est le barycentre des points S_i de poids β, S_j de poids $\frac{1-\beta}{2}(j \neq i)$ et $M_{ij}(\alpha)$ est le barycentre de S_i et S_j de poids respectifs α et $1-\alpha$.
La formule de quadrature associée, qui est de la forme :

$$I_K^{app}(f) = mes(K) \left\{ w_s \sum_{j=1}^{3} f(S_j) + w_\alpha \sum_{\substack{i,j=1 \\ i \neq j}}^{3} f(M_{ij}(\alpha)) + w_\beta \sum_{j=1}^{3} f(G_j(\beta)) \right\},$$

nous fait bien bénéficier de cinq paramètres.
Nous allons maintenant donner en détail la détermination des degrés de liberté de cette formule de quadrature.
Une résolution brutale du système constitué des cinq équations issues des égalités entre l'évaluation par la formule de quadrature et l'intégration exacte d'un représentant de chacune des cinq classes d'équivalence de $\tilde{\mathbb{P}}_5$ s'avère ici (et pour la première fois) infructueuse (rappelons que l'on a à faire ici à un système de cinq équations polynomiales de degré cinq). Il nous faut alors choisir à l'intuition les fonctions polynomiales de degré inférieur ou égal à cinq pour lesquelles l'égalité entre la formule de quadrature et l'intégrale fait apparaître le moins de paramètres simultanément de manière à les déterminer de proche en proche.
On considère dans un premier temps la fonction :

$$\Lambda_1\Lambda_2\Lambda_3(\Lambda_1 - \frac{1-\beta}{2})(\Lambda_1 - \beta)$$

qui nous donne la première égalité :

$$\frac{-1}{5040} + \frac{\beta}{360} - \frac{\beta^2}{240} = 0$$

dont on détermine les deux solutions :

$$\tfrac{1}{3} - \tfrac{2\sqrt{7}}{21} \quad , \quad \tfrac{1}{3} + \tfrac{2\sqrt{7}}{21}.$$

Nous faisons alors le choix arbitraire de poser :

$$\beta = \frac{1}{3} + \frac{2\sqrt{7}}{21}.$$

Si nous avions choisi l'autre solution de l'équation nous n'aurions pas trouvé de formule de quadrature convenable dans la mesure où nécessairement α aurait été complexe.
On considère ensuite la fonction :

$$\Lambda_1\Lambda_2\Lambda_3$$

qui nous donne la seconde égalité :

$$\frac{w_\beta}{6174}(7+2\sqrt{7})(-7+\sqrt{7})^2 = \frac{1}{120}$$

d'où nécessairement :

$$w_\beta = \frac{147}{40(14+\sqrt{7})} = \frac{21\sqrt{7}}{40(2\sqrt{7}+1)}.$$

En considérant la fonction :

$$\Lambda_1(\Lambda_1-\alpha)(\Lambda_1-(1-\alpha))(1-\Lambda_1)$$

il vient :

$$\frac{-1}{60} + \frac{\alpha}{12} - \frac{\alpha^2}{12} = \frac{-40+189\alpha-189\alpha^2+\sqrt{7}}{180(14+\sqrt{7})}$$

dont les solutions sont :

$$\tfrac{-15\sqrt{7}-21+\sqrt{168+174\sqrt{7}}}{2(-15\sqrt{7}-21)} \quad , \quad \tfrac{-15\sqrt{7}-21-\sqrt{168+174\sqrt{7}}}{2(-15\sqrt{7}-21)}.$$

À ce niveau le choix importe peu dans la mesure où ces deux solutions pour localiser les points sur les arêtes sont symétriques par rapport à $\frac{1}{2}$ dans l'intervalle $[0,1]$ et génèrent donc le même jeu de six points.
On choisit alors :

$$\alpha = \frac{-15\sqrt{7}-21+\sqrt{168+174\sqrt{7}}}{2(-15\sqrt{7}-21)}.$$

Pour déterminer w_α nous considérons la fonction :

$$\Lambda_1(1-\Lambda_1)$$

qui nous renvoie l'égalité :

$$\frac{7(6920w_\alpha+1960w_\alpha\sqrt{7}+1008+315\sqrt{7})}{30(\sqrt{7}+7)^2(14+\sqrt{7})} = \frac{1}{12}$$

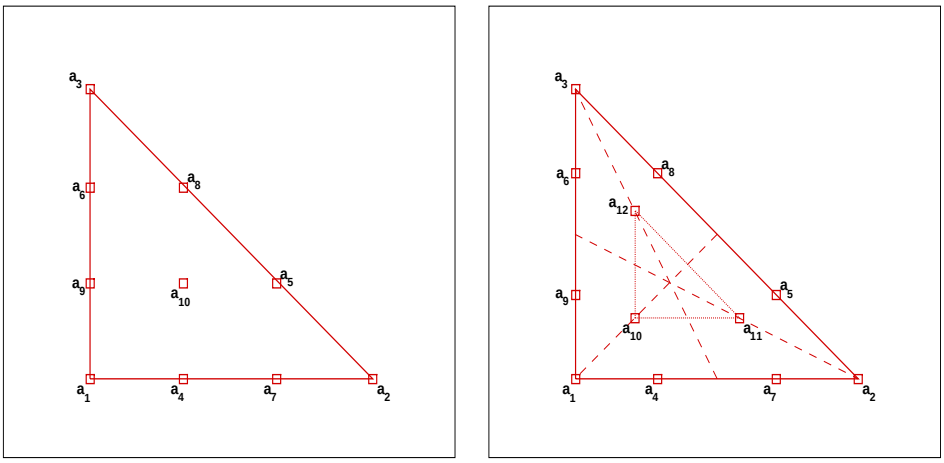

FIG. 4.7 – Localisation des noeuds de l'élément fini P_3 standard et P_3 condensé.

dont la solution est :

$$w_\alpha = \frac{287 + 115\sqrt{7}}{40(173 + 49\sqrt{7})}.$$

Il reste alors à déterminer w_s par exemple à l'aide de la fonction Λ_1 :

$$\frac{110600w_s + 34360w_s\sqrt{7} + 35077 + 10997\sqrt{7}}{80(173 + 49\sqrt{7})(14 + \sqrt{7})} = \frac{1}{6}$$

d'où

$$w_s = \frac{5369 + 1369\sqrt{7}}{120(2765 + 859\sqrt{7})} = \frac{1369 + 767\sqrt{7}}{120(859 + 395\sqrt{7})}.$$

On a donc démontré qu'il existe une unique solution acceptable qui intègre $\mathbb{P}_5$ exactement :

$$\beta = \tfrac{1}{3} + \tfrac{2\sqrt{7}}{21} \simeq 0.5853$$
$$\alpha = \tfrac{-15\sqrt{7}-21+\sqrt{168+174\sqrt{7}}}{2(-15\sqrt{7}-21)} \simeq 0.2935$$

et les poids strictement positifs associés :

$$w_s = \tfrac{1369+767\sqrt{7}}{120(859+395\sqrt{7})} \simeq 0.0148$$
$$w_\alpha = \tfrac{287+115\sqrt{7}}{40(173+49\sqrt{7})} \simeq 0.0488$$
$$w_\beta = \tfrac{21\sqrt{7}}{40(2\sqrt{7}+1)} \simeq 0.02208$$

4.3.6 L'exemple de P_4

À nouveau nous considérons a priori l'espace le plus petit possible tout en contenant P_4, c'est a dire $k' = 4$, et donc

$$\tilde{\mathbb{P}}_4 = \mathbb{P}_4$$

Il nous faut alors considérer une formule de quadrature qui intègre exactement les polynômes de degré inférieur ou égal à six (4+4-2), et pour cela, celle-ci doit faire apparaître sept degrés de liberté.
En effet le nombre de classe d'équivalence $\tilde{M}_6$ dans $B(P_6)$ n'est cette fois plus égal au degré des polynômes considérés (six dans le cas présent) mais est égal à sept :

$$k=6,\quad \tilde{M}_6=7,\quad \begin{cases} C_1=\{\Lambda_1^6,\Lambda_2^6,\Lambda_3^6\}\\ C_2=\{\Lambda_1^5\Lambda_2,\Lambda_1^5\Lambda_3,\Lambda_2^5\Lambda_1,\Lambda_2^5\Lambda_3,\Lambda_3^5\Lambda_1,\Lambda_3^5\Lambda_2\}\\ C_3=\{\Lambda_1^4\Lambda_2^2,\Lambda_1^4\Lambda_3^2,\Lambda_2^4\Lambda_1^2,\Lambda_2^4\Lambda_3^2,\Lambda_3^4\Lambda_1^2,\Lambda_3^4\Lambda_2^2\}\\ C_4=\{\Lambda_1^4\Lambda_2\Lambda_3,\Lambda_2^4\Lambda_1\Lambda_3,\Lambda_3^4\Lambda_1\Lambda_2\}\\ C_5=\{\Lambda_1^3\Lambda_2^3,\Lambda_1^3\Lambda_3^3,\Lambda_2^3\Lambda_3^3\}\\ C_6=\{\Lambda_1^3\Lambda_2^2\Lambda_3,\Lambda_1^3\Lambda_3^2\Lambda_2,\Lambda_2^3\Lambda_1^2\Lambda_3,\Lambda_2^3\Lambda_3^2\Lambda_1,\Lambda_3^3\Lambda_1^2\Lambda_2,\Lambda_3^3\Lambda_2^2\Lambda_1\}\\ C_7=\{\Lambda_1^2\Lambda_2^2\Lambda_3^2\} \end{cases}$$

Malheureusement la formule de quadrature que l'on doit définir sur le jeu de points faisant de l'ensemble des formes linéaires qui lui sont associés un ensemble $\mathbb{P}_4$-unisolvant ne nous permet de disposer que de six paramètres :

- 4 poids w_s, w_l,w_α et w_β associés respectivement aux 3 sommets, 3 milieux des arêtes, 6 autres points sur les arêtes et les 3 points intérieurs au triangle,
- 2 paramètres de localisation α et β associés respectivement aux 6 points sur les arêtes et aux 3 points intérieurs au triangle.

Nous allons donc considérer $k'=5$. Il nous faut déterminer une formule de quadrature exacte pour les polynômes de degré inférieur ou égal à sept (4+5-2). Pour cela il nous faut alors considérer une formule de quadrature à huit degrés de liberté. En effet le nombre de classe d'équivalence $\tilde{M}_7$ dans $B(\mathbb{P}_7)$ est cette fois de huit :

$$k=7,\quad \tilde{M}_7=8,\quad \begin{cases} C_1=\{\Lambda_1^7,\Lambda_2^7,\Lambda_3^7\}\\ C_2=\{\Lambda_1^6\Lambda_2,\Lambda_1^6\Lambda_3,\Lambda_2^6\Lambda_1,\Lambda_2^6\Lambda_3,\Lambda_3^6\Lambda_1,\Lambda_3^6\Lambda_2\}\\ C_3=\{\Lambda_1^5\Lambda_2^2,\Lambda_1^5\Lambda_3^2,\Lambda_2^5\Lambda_1^2,\Lambda_2^5\Lambda_3^2,\Lambda_3^5\Lambda_1^2,\Lambda_3^5\Lambda_2^2,\}\\ C_4=\{\Lambda_1^5\Lambda_2\Lambda_3,\Lambda_2^5\Lambda_1\Lambda_3,\Lambda_3^5\Lambda_1\Lambda_2\}\\ C_5=\{\Lambda_1^4\Lambda_2^3,\Lambda_1^4\Lambda_3^3,\Lambda_2^4\Lambda_1^3,\Lambda_2^4\Lambda_3^3,\Lambda_3^4\Lambda_1^3,\Lambda_3^4\Lambda_2^3\}\\ C_6=\{\Lambda_1^4\Lambda_2^2\Lambda_3,\Lambda_1^4\Lambda_3^2\Lambda_2,\Lambda_2^4\Lambda_1^2\Lambda_3,\Lambda_2^4\Lambda_3^2\Lambda_1,\Lambda_3^4\Lambda_1^2\Lambda_2,\Lambda_3^4\Lambda_2^2\Lambda_1\}\\ C_7=\{\Lambda_1^3\Lambda_2^3\Lambda_3,\Lambda_2^3\Lambda_3^3\Lambda_1,\Lambda_3^3\Lambda_1^3\Lambda_2\}\\ C_8=\{\Lambda_1^3\Lambda_2^2\Lambda_3^2,\Lambda_2^3\Lambda_3^2\Lambda_1^2,\Lambda_3^3\Lambda_1^2\Lambda_2^2\} \end{cases}$$

Remarquons qu'il n'y a de nouveau qu'une seule possibilité de construire un espace $\tilde{P}_4$ dans le sens décrit précédemment et vérifiant $P_4\subset\tilde{P}_4\subset P_5$:

$$\tilde{P}_4=P_4+bP_2$$

À cet espace polynomial nous associons le jeu de points de quadrature suivant :

- les sommets $\{S_1,S_2,S_3\}$
- les milieux des arêtes $\{M_1,M_2,M_3\}$
- les six points situés sur les côtés
  $\{M_{12}(\alpha),M_{21}(\alpha),M_{13}(\alpha),M_{31}(\alpha),M_{23}(\alpha),M_{32}(\alpha)\}$

– les six points intérieurs au triangle
$\{G_1(\beta), G_2(\beta), G_3(\beta), GM_1(\Omega), GM_2(\Omega), GM_3(\Omega)\}$

où $G_i(\beta)$ est le barycentre des points S_i de poids β, S_j de poids $\frac{1-\beta}{2}(j \neq i)$, $GM_i(\Omega)$ est le barycentre des points S_i de poids Ω, S_j de poids $\frac{1-\Omega}{2}(j \neq i)$ et $M_{ij}(\alpha)$ est le barycentre de S_i et S_j de poids respectifs α et $1-\alpha$ (voir partie droite de la figure 4.8).

On cherche alors une formule de quadrature à huit paramètres de la forme :

$$I_K^{app}(f) = mes(K) \left\{ \begin{array}{l} w_s \sum_{j=1}^3 f(S_j) + w_m \sum_{j=1}^3 f(M_j) + w_\alpha \sum_{\substack{i,\,j=1 \\ i \neq j}}^3 f(M_{ij}(\alpha)) \\ + w_\beta \sum_{j=1}^3 f(G_j(\beta)) + w_\Omega \sum_{j=1}^3 f(GM_j(\Omega)) \end{array} \right\}$$

Donnons à nouveau le cheminement complet menant à la détermination de ces huit paramètres de manière à bien mettre en évidence la complexité grandissante du problème. La première fonction test que nous considérons est :

$$\Lambda_1\Lambda_2\Lambda_3(\Lambda_1 - \frac{1-\Omega}{2})(\Lambda_1 - \Omega)(\Lambda_1 - \frac{1-\beta}{2})(\Lambda_1 - \beta)$$

et l'égalité entre la formule de quadrature et l'intégrale est :

$$\frac{-\beta}{10080} - \frac{\Omega}{10080} + \frac{\Omega^2}{10080} + \frac{\Omega\beta}{1008} - \frac{\Omega^2\beta}{720} + \frac{\beta^2}{10080} + \frac{\Omega^2\beta^2}{480} - \frac{\Omega\beta^2}{720} + \frac{1}{30240} = 0.$$

Nous résolvons cette équation par rapport à Ω ce qui nous renvoie deux solutions en fonction de β :

$$\begin{array}{rcl} \Omega & = & \dfrac{-30\beta + 3 + 42\beta^2 + \sqrt{360\beta^2 + 24\beta - 1260\beta^3 - 3 + 1008\beta^4}}{2(63\beta^2 + 3 - 42\beta)}, \\ \text{ou} & & \\ \Omega & = & \dfrac{-30\beta + 3 + 42\beta^2 - \sqrt{360\beta^2 + 24\beta - 1260\beta^3 - 3 + 1008\beta^4}}{2(63\beta^2 + 3 - 42\beta)}. \end{array}$$

À nouveau nous choisissons arbitrairement l'une des solutions, par exemple la seconde (quitte à choisir l'autre solution si celle-ci ne s'avère pas convenable) ; la fonction

$$\Lambda_1\Lambda_2\Lambda_3(\Lambda_1 - \frac{1-\Omega}{2})(\Lambda_1 - \Omega)(\Lambda_2 - \frac{1-\beta}{2})(\Lambda_2 - \beta)$$

nous amène une seconde égalité entre Ω et β :

$$\frac{42\beta^3 - 66\beta^2 + 26\beta + (2\beta - 1)\sqrt{-3 + 360\beta^2 + 24\beta + 1008\beta^4 - 1260\beta^3} - 1}{120960(-14\beta + 1 + 21\beta^2)} = 0$$

dont les solutions sont

$$\frac{4 - \sqrt{7}}{9} \quad , \quad \frac{4 + \sqrt{7}}{9}.$$

Nous choisissons alors

$$\beta = \frac{4+\sqrt{7}}{9},$$

ce qui nous détermine automatiquement Ω :

$$\Omega = \frac{-1+2\sqrt{7}}{10+7\sqrt{7}}.$$

Remarque 4.3.4. *Si nous avions choisi précédemment l'autre solution pour Ω nous aurions alors nécessairement une solution négative pour Ω ou β suivant le choix de la solution ci-dessus et si nous avions choisi l'autre solution ci-dessus (tout en ayant choisi la solution convenable pour Ω) nous aurions alors $\Omega = \beta$ ce qui est bien entendu à exclure. Nous sommes donc à un point où nous avons une unique solution potentiellement convenable pour Ω et β.*

En considérant ensuite la fonction :

$$\Lambda_1\Lambda_2\Lambda_3,$$

il vient :

$$\frac{201382w_\beta + 85369w_\beta\sqrt{7} + 453438w_\Omega + 175689w_\Omega\sqrt{7}}{972(10+7\sqrt{7})^3} = \frac{1}{120},$$

que l'on résout par rapport à w_Ω :

$$w_\Omega = -\frac{-914490 - 364581\sqrt{7} + 2013820w_\beta + 853690w_\beta\sqrt{7}}{7290(622+241\sqrt{7})}.$$

Puis la fonction :

$$\Lambda_1\Lambda_2\Lambda_3(\Lambda_1 - \frac{1}{2})(\Lambda_2 - \frac{1}{2})(\Lambda_3 - \frac{1}{2}),$$

d'où l'équation :

$$(4+\sqrt{7})(-1+2\sqrt{7})((440081188534400 + 166356161968760\sqrt{7})w_\beta$$
$$-11102086102347 - 4200265555434\sqrt{7}) = \frac{56687040(622+241\sqrt{7})(10+7\sqrt{7})^6}{20160}.$$

Il vient alors :

$$w_\beta = \frac{19683(4491766196\sqrt{7} + 11884797941)}{140(6242276236090\sqrt{7} + 16516136607649)},$$

et automatiquement :

$$w_\Omega = \frac{156107549422310943\sqrt{7} + 4130142353460756370}{1260(6242276236090\sqrt{7} + 16516136607649)(622+241\sqrt{7})}.$$

Il nous reste alors à déterminer les paramètres w_α, α, w_s et w_m (nous n'écrirons plus les équations issues des égalités entre l'intégration exacte et la quadrature, celles-ci devenant

encore plus longues que précédemment).
En utilisant la fonction :

$$\Lambda_1(\Lambda_1 - \frac{1-\beta}{2})(\Lambda_1 - \frac{1}{2})(\Lambda_1 - \beta)(1 - \Lambda_1)$$

nous déterminons w_α en fonction de α :

$$w_\alpha = -\frac{1}{840(5\alpha + 4\alpha^3 - 1 - 8\alpha^2)\alpha},$$

puis la fonction :

$$(\Lambda_1 - \frac{1}{2})(\Lambda_2 - \frac{1}{2})(\Lambda_3 - \frac{1}{2})$$

nous permet de déterminer w_s en fonction de α :

$$w_s = \frac{-26\alpha + 26\alpha^2 + 3}{1260(-1+\alpha)\alpha}.$$

Finalement la fonction :

$$\Lambda_1(\Lambda_1 - \frac{1}{2})(\Lambda_1 - \alpha)(\Lambda_2 - \alpha)(\Lambda_1 - \frac{1-\alpha}{2})(\Lambda_2 - \frac{1-\alpha}{2})$$

nous permet de déterminer α parmi deux solutions symétriques par rapport à $\frac{1}{2}$ dans $[0, 1]$:

$$\alpha = \frac{3-\sqrt{3}}{6}.$$

Il vient alors :

$$w_s = \frac{2}{315} \quad , \quad w_\alpha = \frac{3}{140}.$$

Reste alors à déterminer w_m, par exemple à l'aide de la fonction Λ_1 :

$$w_m = \frac{8}{315}.$$

On a donc démontré qu'il existe une unique formule de quadrature de la forme considérée qui intègre exactement $\mathbb{P}_7$, ses huit paramètres sont :

$$\Omega = \frac{-1+2\sqrt{7}}{10+7\sqrt{7}} \simeq 0.1504$$
$$\beta = \frac{4+\sqrt{7}}{9} \simeq 0.7384$$
$$\alpha = \frac{3-\sqrt{3}}{6} \simeq 0.2113$$

et les poids strictement positifs associés :

$$w_\Omega = \frac{156107549422310943\sqrt{7} + 4130142353460756370}{1260(6242276236090\sqrt{7} + 16516136607649)(622 + 241\sqrt{7})} \simeq 0.1575$$
$$w_\beta = \frac{19683(4491766196\sqrt{7} + 11884797941)}{140(6242276236090\sqrt{7} + 16516136607649)} \simeq 0.1011$$
$$w_\alpha = \frac{3}{140} \simeq 0.02142$$
$$w_s = \frac{2}{315} \simeq 0.006349$$
$$w_m = \frac{8}{315} \simeq 0.02539$$

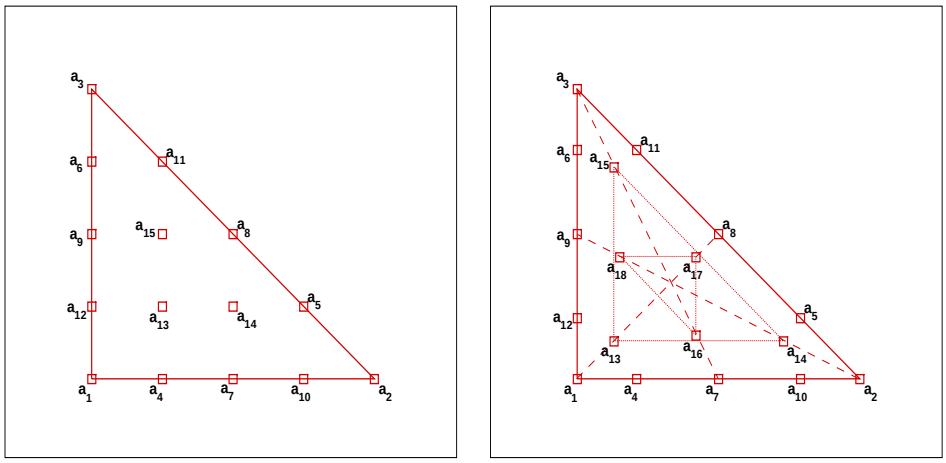

FIG. 4.8 – Localisation des noeuds de l'élément fini P_4 standard et P_4 condensé.

4.3.7 L'exemple de P_5

Nous allons considérer a priori l'espace le plus petit possible qui contienne $\mathbb{P}_5$, c'est-à-dire $k' = 5$ et donc

$$\tilde{\mathbb{P}}_5 = \mathbb{P}_5$$

Il nous faut alors construire une formule de quadrature à vingt-et-un points qui intègre exactement les polynômes de degré inférieur ou égal à huit (5+5-2), et pour cela, celle-ci doit faire apparaître dix degrés de liberté, le nombre de classe d'équivalence $\tilde{M}_8$ dans $B(\mathbb{P}_8)$:

$$k = 8, \quad N_8 = 28, \quad \tilde{M}_8 = 10, \quad \begin{cases} C_1 = \{\Lambda_1^8, \Lambda_2^8, \Lambda_3^8\} \\ C_2 = \{\Lambda_1^7\Lambda_2, \Lambda_1^7\Lambda_3, \Lambda_2^7\Lambda_1, \Lambda_2^7\Lambda_3, \Lambda_3^7\Lambda_1, \Lambda_3^7\Lambda_2\} \\ C_3 = \{\Lambda_1^6\Lambda_2^2, \Lambda_1^6\Lambda_3^2, \Lambda_2^6\Lambda_1^2, \Lambda_2^6\Lambda_3^2, \Lambda_3^6\Lambda_1^2, \Lambda_3^6\Lambda_2^2\} \\ C_4 = \{\Lambda_1^6\Lambda_2\Lambda_3, \Lambda_2^6\Lambda_1\Lambda_3, \Lambda_3^6\Lambda_1\Lambda_2\} \\ C_5 = \{\Lambda_1^5\Lambda_2^3, \Lambda_1^5\Lambda_3^3, \Lambda_2^5\Lambda_1^3, \Lambda_2^5\Lambda_3^3, \Lambda_3^5\Lambda_1^3, \Lambda_3^5\Lambda_2^3\} \\ C_6 = \{\Lambda_1^5\Lambda_2^2\Lambda_3, \Lambda_1^5\Lambda_3^2\Lambda_2, \Lambda_2^5\Lambda_1^2\Lambda_3, \Lambda_2^5\Lambda_3^2\Lambda_1, \Lambda_3^5\Lambda_1^2\Lambda_2, \Lambda_3^5\Lambda_2^2\Lambda_1\} \\ C_7 = \{\Lambda_1^4\Lambda_2^4, \Lambda_1^4\Lambda_3^4, \Lambda_2^4\Lambda_3^4\} \\ C_8 = \{\Lambda_1^4\Lambda_2^3\Lambda_3, \Lambda_1^4\Lambda_3^3\Lambda_2, \Lambda_2^4\Lambda_1^3\Lambda_3, \Lambda_2^4\Lambda_3^3\Lambda_1, \Lambda_3^4\Lambda_1^3\Lambda_2, \Lambda_3^4\Lambda_2^3\Lambda_1\} \\ C_9 = \{\Lambda_1^4\Lambda_2^2\Lambda_3^2, \Lambda_2^4\Lambda_1^2\Lambda_3^2, \Lambda_3^4\Lambda_1^2\Lambda_2^2\} \\ C_{10} = \{\Lambda_1^3\Lambda_2^3\Lambda_3^2, \Lambda_1^3\Lambda_3^3\Lambda_2^2, \Lambda_2^3\Lambda_3^3\Lambda_1^2\} \end{cases}$$

Or la formule de quadrature que l'on doit définir sur le jeu de points faisant de l'ensemble des formes linéaires qui lui sont associés un ensemble $\mathbb{P}_5$-unisolvant ne nous permet de disposer au maximum que de neuf degrés de liberté :

- 5 poids w_s, $w_{\alpha 1}$, $w_{\alpha 2}$, $w_{\beta 1}$, $w_{\beta 2}$ associés respectivement aux 3 sommets, les 2 fois 6 points situés sur les arêtes et les 2 fois 3 points intérieurs au triangle,
- 4 paramètres de localisation $\alpha 1$, $\alpha 2$, $\beta 1$ et $\beta 2$ associés respectivement aux 2 fois 6 points sur les arêtes et aux 2 fois 3 points intérieurs au triangle.

Nous allons donc considérer $k' = 6$. Il nous faut alors enrichir notre espace polynomial

de manière à ce que

$$\mathbb{P}_5 \subset \tilde{\mathbb{P}}_5 \subset \mathbb{P}_6.$$

Chacun peut se convaincre que les seuls espaces $\tilde{\mathbb{P}_5}$ convenables (dans le sens décrit précédemment) sont les suivants :

$$\tilde{\mathbb{P}}_5 = \mathbb{P}_5 + b\mathbb{P}_3$$

et

$$\tilde{\mathbb{P}}_5 = \mathbb{P}_5 + b^2\mathbb{P}_0.$$

Une fois choisi l'un de ces deux espaces il nous faut déterminer une formule de quadrature exacte pour les polynômes de degré inférieur ou égal à neuf (5+6-2). Pour cela il nous faut alors considérer une formule de quadrature à douze paramètres :

$$k = 9, \quad \tilde{M}_9 = 12, \quad \begin{cases} C_1 = \{\Lambda_1^9, \Lambda_2^9, \Lambda_3^9\} \\ C_2 = \{\Lambda_1^8\Lambda_2, \lambda_1^8\Lambda_3, \Lambda_2^8\Lambda_1, \Lambda_2^8\Lambda_3, \Lambda_3^8\Lambda_1, \Lambda_3^8\Lambda_2\} \\ C_3 = \{\Lambda_1^7\Lambda_2^2, \Lambda_1^7\Lambda_3^2, \Lambda_2^7\Lambda_1^2, \Lambda_2^7\Lambda_3^2, \Lambda_3^7\Lambda_1^2, \Lambda_3^7\Lambda_2^2, \} \\ C_4 = \{\Lambda_1^7\Lambda_2\Lambda_3, \Lambda_2^7\Lambda_1\Lambda_3, \Lambda_3^7\Lambda_1\Lambda_2\} \\ C_5 = \{\Lambda_1^6\Lambda_2^3, \Lambda_1^6\Lambda_3^3, \Lambda_2^6\Lambda_1^3, \Lambda_2^6\Lambda_3^3, \Lambda_3^6\Lambda_1^3, \Lambda_3^6\Lambda_2^3\} \\ C_6 = \{\Lambda_1^6\Lambda_2^2\Lambda_3, \Lambda_1^6\Lambda_3^2\Lambda_2, \Lambda_2^6\Lambda_1^2\Lambda_3, \Lambda_2^6\Lambda_3^2\Lambda_1, \Lambda_3^6\Lambda_1^2\Lambda_2, \Lambda_3^6\Lambda_2^2\Lambda_1\} \\ C_7 = \{\Lambda_1^5\Lambda_2^4, \Lambda_1^5\Lambda_3^4, \Lambda_2^5\Lambda_1^4, \Lambda_2^5\Lambda_3^4, \Lambda_3^5\Lambda_1^4, \Lambda_3^5\Lambda_2^4\} \\ C_8 = \{\Lambda_1^5\Lambda_2^3\Lambda_3, \Lambda_1^5\Lambda_3^3\Lambda_2, \Lambda_2^5\Lambda_1^3\Lambda_3, \Lambda_2^5\Lambda_3^3\Lambda_1, \Lambda_3^5\Lambda_1^3\Lambda_2, \Lambda_3^5\Lambda_2^3\Lambda_1\} \\ C_9 = \{\Lambda_1^5\Lambda_2^2\Lambda_3^2, \Lambda_2^5\Lambda_1^2\Lambda_3^2, \Lambda_3^5\Lambda_1^2\Lambda_2^2\} \\ C_{10} = \{\Lambda_1^4\Lambda_2^4\Lambda_3, \Lambda_1^4\Lambda_3^4\Lambda_2, \Lambda_2^4\Lambda_3^4\Lambda_1\} \\ C_{11} = \{\Lambda_1^4\Lambda_2^3\Lambda_3^2, \Lambda_1^4\Lambda_3^3\Lambda_2^2, \Lambda_2^4\Lambda_1^3\Lambda_3^2, \Lambda_2^4\Lambda_3^3\Lambda_1^2, \Lambda_3^4\Lambda_1^3\Lambda_2^2, \Lambda_3^4\Lambda_2^3\Lambda_1^2\} \\ C_{12} = \{\Lambda_1^3\Lambda_2^3\Lambda_3^3\} \end{cases}$$

La dimension de l'espace $\tilde{\mathbb{P}}_5 = \mathbb{P}_5 + b\mathbb{P}_3$ est de vingt-cinq. Il nous faut donc construire une formule de quadrature sur vingt-cinq points, sachant que quinze d'entre eux sont localisés sur les arêtes. La seule manière de définir un jeu de vingt cinq-points faisant de l'ensemble des formes linéaires qui lui sont associés un ensemble $\mathbb{P}_5 + b\mathbb{P}_3$-unisolvant est la suivante et ne nous permet de disposer que de onze paramètres :

- 6 poids w_s, $w_{\alpha 1}$, $w_{\alpha 2}$, $w_{\beta 1}$, w_ω et w_g associés respectivement aux 3 sommets, aux 2 fois 6 points situés sur les arêtes, aux 3 points intérieurs au triangle situé sur les médianes, aux 6 autres points intérieurs et enfin au barycentre du triangle,
- 5 paramètres de localisation $\alpha 1$, $\alpha 2$,$\beta 1$, $\omega 1$ et $\omega 2$ associés respectivement au 2 fois 6 points sur les arêtes, aux 3 points intérieurs au triangle situés sur les trois médianes et aux 6 autres points intérieurs au triangle.

La dimension de l'espace $\tilde{\mathbb{P}}_5 = \mathbb{P}_5 + b^2\mathbb{P}_0$ est de vingt-deux. Chacun se rendra compte que les vingt-deux points dont on dispose (dont quinze sont nécessairement sur les arêtes) ne nous donnent que dix paramètres (six poids et quatre paramètres de localisation) ce qui est moins que les douze nécessaires .

Nous considérons alors $k' = 7$. Les seuls espaces convenables entre P_5 et P_7 sont les suivants :

$$\tilde{\mathbb{P}}_5 = \mathbb{P}_5 + b^2\mathbb{P}_1,$$

$$\tilde{\mathbb{P}}_5 = \mathbb{P}_5 + b(\mathbb{P}_3 + b\mathbb{P}_1),$$

$$\tilde{\mathbb{P}}_5 = \mathbb{P}_5 + b\mathbb{P}_4.$$

Il nous faut construire une formule de quadrature intégrant de manière exacte les polynômes de degré inférieur ou égal à dix (5+7-2) faisant apparaître quatorze paramètres :

$$k = 10, \quad \tilde{M}_{10} = 14, \quad \begin{cases} C_1 = \{\Lambda_1^{10}, \Lambda_2^{10}, \Lambda_3^{10}\} \\ C_2 = \{\Lambda_1^9\Lambda_2, \Lambda_1^9\Lambda_3, \Lambda_2^9\Lambda_1, \Lambda_2^9\Lambda_3, \Lambda_3^9\Lambda_1, \Lambda_3^9\Lambda_2\} \\ C_3 = \{\Lambda_1^8\Lambda_2^2, \Lambda_1^8\Lambda_3^2, \Lambda_2^8\Lambda_1^2, \Lambda_2^8\Lambda_3^2, \Lambda_3^8\Lambda_1^2, \Lambda_3^8\Lambda_2^2, \} \\ C_4 = \{\Lambda_1^8\Lambda_2\Lambda_3, \Lambda_2^8\Lambda_1\Lambda_3, \Lambda_3^8\Lambda_1\Lambda_2\} \\ C_5 = \{\Lambda_1^7\Lambda_2^3, \Lambda_1^7\Lambda_3^3, \Lambda_2^7\Lambda_1^3, \Lambda_2^7\Lambda_3^3, \Lambda_3^7\Lambda_1^3, \Lambda_3^7\Lambda_2^3\} \\ C_6 = \{\Lambda_1^7\Lambda_2^2\Lambda_3, \Lambda_1^7\Lambda_3^2\Lambda_2, \Lambda_2^7\Lambda_1^2\Lambda_3, \Lambda_2^7\Lambda_3^2\Lambda_1, \Lambda_3^7\Lambda_1^2\Lambda_2, \Lambda_3^7\Lambda_2^2\Lambda_1\} \\ C_7 = \{\Lambda_1^6\Lambda_2^4, \Lambda_1^6\Lambda_3^4, \Lambda_2^6\Lambda_1^4, \Lambda_2^6\Lambda_3^4, \Lambda_3^6\Lambda_1^4, \Lambda_3^6\Lambda_2^4\} \\ C_8 = \{\Lambda_1^6\Lambda_2^3\Lambda_3, \Lambda_1^6\Lambda_3^3\Lambda_2, \Lambda_2^6\Lambda_1^3\Lambda_3, \Lambda_2^6\Lambda_3^3\Lambda_1, \Lambda_3^6\Lambda_1^3\Lambda_2, \Lambda_3^6\Lambda_2^3\Lambda_1\} \\ C_9 = \{\Lambda_1^6\Lambda_2^2\Lambda_3^2, \Lambda_2^6\Lambda_1^2\Lambda_3^2, \Lambda_3^6\Lambda_1^2\Lambda_2^2\} \\ C_{10} = \{\Lambda_1^5\Lambda_2^5, \Lambda_1^5\Lambda_3^5, \Lambda_2^5\Lambda_3^5\} \\ C_{11} = \{\Lambda_1^5\Lambda_2^4\Lambda_3, \Lambda_1^5\Lambda_3^4\Lambda_2, \Lambda_2^5\Lambda_1^4\Lambda_3, \Lambda_2^5\Lambda_3^4\Lambda_1, \Lambda_3^5\Lambda_1^4\Lambda_2, \Lambda_3^5\Lambda_2^4\Lambda_1\} \\ C_{12} = \{\Lambda_1^5\Lambda_2^3\Lambda_3^2, \Lambda_1^5\Lambda_3^3\Lambda_2^2, \Lambda_2^5\Lambda_1^3\Lambda_3^2, \Lambda_2^5\Lambda_3^3\Lambda_1^2, \Lambda_3^5\Lambda_1^3\Lambda_2^2, \Lambda_3^5\Lambda_2^3\Lambda_1^2\} \\ C_{13} = \{\Lambda_1^4\Lambda_2^4\Lambda_3^2, \Lambda_1^4\Lambda_3^4\Lambda_2^2, \Lambda_2^4\Lambda_3^4\Lambda_1^2\} \\ C_{14} = \{\Lambda_1^4\Lambda_2^3\Lambda_3^3, \Lambda_2^4\Lambda_3^3\Lambda_1^3, \Lambda_3^4\Lambda_1^3\Lambda_2^3\} \end{cases}$$

En considérant $\tilde{\mathbb{P}}_5 = \mathbb{P}_5 + b^2\mathbb{P}_1$ nous ne pouvons avoir que onze paramètres (six poids et cinq paramètres de localisation) tandis qu'avec $\tilde{\mathbb{P}}_5 = \mathbb{P}_5 + b(\mathbb{P}_3 + b\mathbb{P}_1)$ nous ne pouvons en avoir que douze (six poids et six paramètres de localisation).
Nous considérons alors le dernier espace convenable entre $\mathbb{P}_5$ et $\mathbb{P}_7$:

$$\tilde{\mathbb{P}}_5 = \mathbb{P}_5 + b\mathbb{P}_4$$

Cette fois il nous faut utiliser un jeu de trente points de quadrature. La seule manière de définir un jeu de trente points faisant de l'ensemble des formes linéaires qui lui sont associés un ensemble $\mathbb{P}_5 + b\mathbb{P}_4$-unisolvant (tout en respectant les contraintes liées à la conformité) est la suivante et nous donne les quatorze degrés de liberté dont nous avons besoin (voir partie droite de la figure 4.9) :

- sept poids w_s, w_{α_1}, w_{α_2}, w_{β_1}, w_{β_2}, w_{β_3} et w_ω associés respectivement aux trois sommets, deux jeux de six points sur les arêtes, trois jeux de trois points sur les médianes et un jeu de six points intérieurs au triangle,
- sept paramètres de localisation α_1, α_2, β_1, β_2, β_3, ω_1 et ω_2 associés respectivement aux deux jeux de six points sur les arêtes, trois jeux de trois points sur les médianes et aux six points intérieurs au triangle.

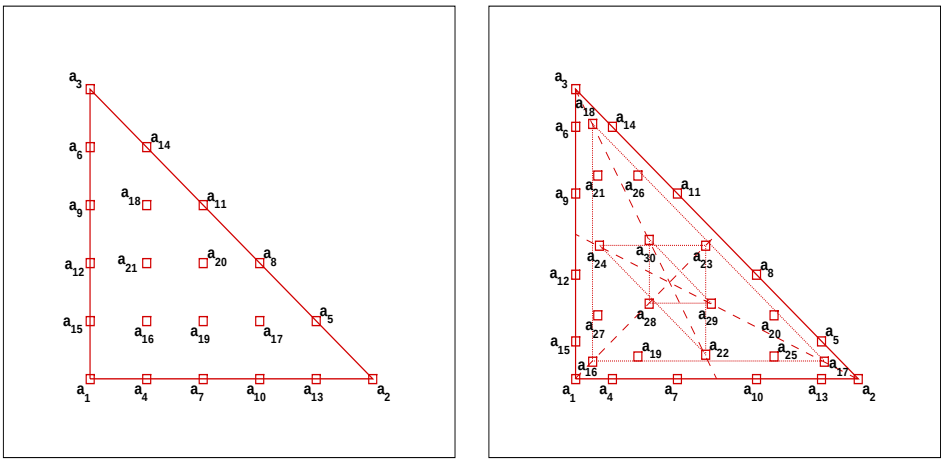

FIG. 4.9 – Localisation des noeuds de l'élément fini P_5 standard et P_5 condensé.

La formule de quadrature associée sera de la forme :

$$
\begin{aligned}
I_K^{app}(f) = mes(K)\{ & w_s \sum_{j=1}^{3} f(S_j) + w_{\alpha_1} \sum_{\substack{i,j=1 \\ i \neq j}}^{3} f(M_{ij}(\alpha_1)) \\
& + w_{\alpha_2} \sum_{\substack{i,j=1 \\ i \neq j}}^{3} f(M_{ij}(\alpha_2)) + w_{\beta_1} \sum_{j=1}^{3} f(G_j(\beta_1)) \\
& + w_{\beta_2} \sum_{j=1}^{3} f(G_j(\beta_2)) + w_{\beta_3} \sum_{j=1}^{3} f(G_j(\beta_3)) \\
& + w_{\omega} \sum_{j=1}^{6} f(G_j(\omega_1, \omega_2))\}.
\end{aligned}
$$

Nous résolvons numériquement ce système à l'aide de MAPLE© et cette fois il existe une solution convenable :

$$
\begin{aligned}
\alpha_1 &\simeq 0.13226458163271398535388822004364735894320922145235 \\
\alpha_2 &\simeq 0.36329807415368604570550633618418105322598405901322 \\
\beta_1 &\simeq 0.88494463117717978867836496446913619911878691250281 \\
\beta_2 &\simeq 0.08432632384167779612299356515186339305516402888832 \\
\beta_3 &\simeq 0.48628178547608184787221833703559663865211105673340 \\
\omega_1 &\simeq 0.22100121875989000797812820146484191542882927860498 \\
\omega_2 &\simeq 0.07819258362551702199888597846982582812764202045477
\end{aligned}
$$

et les poids strictement positifs :

$$\begin{aligned}
w_s &\simeq 0.00141884794135849195859201314216756690368214898897 29\\
w_{\alpha_1} &\simeq 0.00696115728097842131688535505330660662678389159796 1\\
w_{\alpha_2} &\simeq 0.01238113000735325822823625385145974077706964686367 6\\
w_{\beta_1} &\simeq 0.02325227091923514227899684743112030085978283029949 8\\
w_{\beta_2} &\simeq 0.06906086075456558705677702806082523413852311846006 2\\
w_{\beta_3} &\simeq 0.09180247526152571475403829582713371371109543840295 4\\
w_{\omega} &\simeq 0.05455715193999251909734296553127691145627136012928 5
\end{aligned}$$

4.3.8 L'exemple de P_6

En considérant l'espace polynomial $\mathbb{P}_6$ lui-même, nous ne pouvons disposer que de douze paramètres sur les quatorze requis : il faut intégrer les polynômes de degré inférieur ou égal à dix.

Nous déterminons les espaces entre $\mathbb{P}_6$ et $\mathbb{P}_7$:

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b^2\mathbb{P}_1,$$

et

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b\mathbb{P}_4.$$

Ces espaces nous renvoient respectivement treize et quinze paramètres sur les seize nécessaires à l'intégration exacte des polynômes de degré inférieur ou égal à onze.

Nous continuons en considérant les espaces entre $\mathbb{P}_6$ et $\mathbb{P}_8$:

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b^2\mathbb{P}_2,$$

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b(\mathbb{P}_4 + b\mathbb{P}_2),$$

et

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b\mathbb{P}_5.$$

Il nous faut cette fois dix neuf paramètres pour intégrer de manière exacte les polynômes de degré inférieur ou égal à douze, mais nous ne pouvons en disposer respectivement que de quinze, dix-sept et dix-huit.

Nous listons alors les espaces entre $\mathbb{P}_6$ et $\mathbb{P}_9$:

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b^2(\mathbb{P}_1 + b\mathbb{P}_0),$$

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b^2(\mathbb{P}_2 + b\mathbb{P}_0),$$

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b^2\mathbb{P}_3,$$

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b(\mathbb{P}_4 + b^2\mathbb{P}_0),$$

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b(\mathbb{P}_4 + b(\mathbb{P}_2 + b\mathbb{P}_0)),$$

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b(\mathbb{P}_4 + b\mathbb{P}_3),$$

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b(\mathbb{P}_5 + b^2\mathbb{P}_0),$$

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b(\mathbb{P}_5 + b\mathbb{P}_3),$$

et

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b\mathbb{P}_6.$$

De ces neuf espaces, aucun ne nous permet de disposer exactement des vingt-et-un paramètres nécessaires à l'intégration exacte des polynômes de degré inférieur ou égal à treize, mais l'un d'entre eux nous permet de disposer d'un paramètre en plus des vingt-et-un nécessaires :

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b\mathbb{P}_6.$$

Le problème est que MAPLE© n'est capable de résoudre numériquement que des systèmes carrés (c'est-à-dire des systèmes ayant autant d'équations que d'inconnues). Nous avons toutefois essayé de lier deux paramètres de différentes manières, par exemple en imposant que l'un des poids soit la différence entre un et la somme des autres poids (la somme des poids, pondérée par l'aire du triangle de référence, étant automatiquement égale à un dès que la formule de quadrature intègre de manière exacte les constantes, nous ne figeons rien de manière arbitraire en faisant cela)...malheureusement sans succès.

En continuant à considérer des espaces encore plus grands nous trouvons :

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b(\mathbb{P}_4 + b\mathbb{P}_4),$$

et

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b\mathbb{P}_7,$$

entre $\mathbb{P}_6$ et $\mathbb{P}_{10}$, qui nous donne respectivement vingt-cinq et vingt-six paramètres sur les vingt-quatre nécessaires et finalement

$$\tilde{\mathbb{P}}_6 = \mathbb{P}_6 + b(\mathbb{P}_6 + b(\mathbb{P}_4 + b\mathbb{P}_2))$$

entre $\mathbb{P}_6$ et $\mathbb{P}_{11}$ qui est le premier espace polynomial qui nous permet de disposer d'exactement autant de paramètres que nécessaires à l'intégration exacte des polynômes de degré inférieur ou égal à quinze, soit vingt-sept dont :

- treize poids w_s, w_m, w_{α_1}, w_{α_2}, w_{β_1}, w_{β_2}, w_{β_3}, w_{β_4}, w_{β_5}, w_{β_6}, w_ω, w_γ et w_η associés respectivement aux trois sommets, trois milieux des arêtes, deux jeux de six points sur les arêtes, six jeux de trois points sur les médianes et trois jeux de six points intérieurs au triangle,
- quatorze paramètres de localisation α_1, α_2, β_1, β_2, β_3, β_4, β_5, β_6, ω_1, ω_2, γ_1, γ_2, η_1 et η_2 associés respectivement aux deux jeux de six points sur les arêtes, six jeux de trois points sur les médianes et aux trois jeux de six points intérieurs au triangle.

La formule de quadrature prend la forme générique suivante :

$$\begin{aligned}
I_K^{app}(f) &= mes(K)\{w_s \sum_{j=1}^{3} f(S_j) + w_{\alpha_1} \sum_{\substack{i,j=1 \\ i\neq j}}^{3} f(M_{ij}(\alpha_1)) \\
&+ w_{\alpha_2} \sum_{\substack{i,j=1 \\ i\neq j}}^{3} f(M_{ij}(\alpha_2)) + w_m \sum_{j=1}^{3} f(M_j) \\
&+ w_{\beta_1} \sum_{j=1}^{3} f(G_j(\beta_1)) + w_\omega \sum_{j=1}^{6} f(G_j(\omega_1,\omega_2)) \\
&+ w_\gamma \sum_{j=1}^{6} f(G_j(\gamma_1,\gamma_2)) + w_{\beta_2} \sum_{j=1}^{3} f(G_j(\beta_2)) \\
&+ w_{\beta_3} \sum_{j=1}^{3} f(G_j(\beta_3)) + w_\eta \sum_{j=1}^{6} f(G_j(\eta_1,\eta_2)) \\
&+ w_{\beta_4} \sum_{j=1}^{3} f(G_j(\beta_4)) + w_{\beta_5} \sum_{j=1}^{3} f(G_j(\beta_5)) \\
&+ w_{\beta_6} \sum_{j=1}^{3} f(G_j(\beta_6))\}.
\end{aligned}$$

La résolution (numérique) du système polynomial constitué des égalités entre l'évaluation de cette formule de quadrature et l'intégration exacte d'un représentant de chacune des vingt-sept classes d'équivalence de $B(\mathbb{P}_{15})$ n'a été réussie qu'avec une bonne initialisation de la localisation des points, ce qui nous a demandé un peu d'intuition et pas mal de chance. La solution est alors donnée par :

$$\begin{aligned}
\alpha_1 &\simeq 0.048328814080479057424558067859816742717695705184333 \\
\alpha_2 &\simeq .156889155584739310316668383523551556210834974318 18 \\
\beta_1 &\simeq .903816639580845208205785171145054509967881631956 05 \\
\beta_2 &\simeq 0.047898621838323026959279690164038119179373445721984 \\
\beta_3 &\simeq .716636980609917842476639778085854223663471286266 01 \\
\beta_4 &\simeq .160483242810439247600579591085466729265970295168 29 \\
\beta_5 &\simeq .527174106166824908938278227507049766213809149185 33 \\
\beta_6 &\simeq .290327693520982168611368591032989565609526399756 82 \\
\omega_1 &\simeq .155281907716302352530682633073166485264172185803 20 \\
\omega_2 &\simeq 0.043808855299202926165197147854605590935259610826989 \\
\gamma_1 &\simeq 0.310873920620226987440019945161334282012877012853 04 \\
\gamma_2 &\simeq 0.018669969955364815357041953565838632145399456409655 \\
\eta_1 &\simeq .290949955416471355427114590719986865682349802512 89 \\
\eta_2 &\simeq 0.097955775850824782170700589609487451292547438792040
\end{aligned}$$

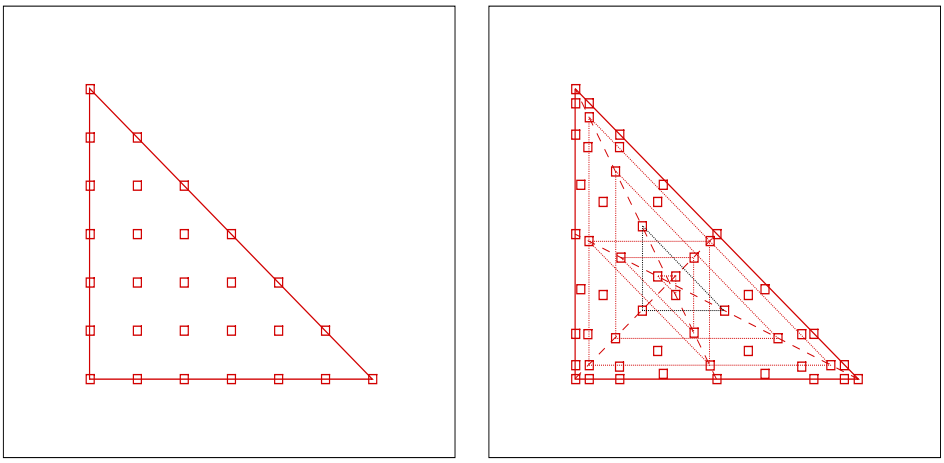

FIG. 4.10 – Localisation des 28 noeuds de l'élément fini P_6 standard et des 54 noeuds du P_6 condensé.

et les poids strictement positifs :

$$
\begin{aligned}
w_s &\simeq 0.00035670532998628816007141134943849466217587254778624\\
w_m &\simeq 0.00495913645821493157849702712501582493644169346082 38\\
w_{\alpha_1} &\simeq 0.00213103187981581819185069995423427542911774009188 19\\
w_{\alpha_2} &\simeq 0.00330090466471432719336754321733421567498004200694 32\\
w_{\beta_1} &\simeq 0.01263009150243493058100502579701472673592187609094 0\\
w_{\beta_2} &\simeq 0.02828054422581060256709947087336084060746969803746 1\\
w_{\beta_3} &\simeq 0.02910355497490404211786397724202240316256735432975 8\\
w_{\beta_4} &\simeq 0.04349000218458691879150442589926265606245259291696 4\\
w_{\beta_5} &\simeq 0.04426968803761943676516328443776891570676200185471 2\\
w_{\beta_6} &\simeq 0.02050041511701979284852466615471517168522030217894 2\\
w_{\omega} &\simeq 0.01904580637999886473147430024616613731741091889276 9\\
w_{\gamma} &\simeq 0.01637365078767301495952534989936201045053333891784 3\\
w_{\eta} &\simeq 0.03402020403917616988558412891027051101511893104853 5
\end{aligned}
$$

La partie droite de la figure 4.10 représente la localisation des noeuds de l'élément fini condensé $\check{P}_6$.

4.4 De nouveaux éléments finis conformes dans $H^1(\Omega)$ partiellement condensés

Comme nous l'avons vu, la condensation de la matrice de masse issue de l'utilisation des élément finis de Lagrange triangulaires, devient complexe dès lors que l'on cherche à monter en ordre. Cela est principalement dû au fait qu'il est impossible de résoudre, en toute généralité, les systèmes polynomiaux dont les solutions déterminent une formule de quadrature permettant cette condensation. Le but recherché de la condensation de masse étant, s'il est besoin de le rappeler, d'obtenir une matrice de masse diagonale, la

question que l'on se pose maintenant est la suivante : s'il paraît probable que la construction d'éléments finis décrite dans les sections précédentes aboutisse à la condensation d'éléments finis d'ordre plus élevés, mais que d'un point de vue pratique il nous est impossible de déterminer les formules de quadrature appropriées, n'est-il pas possible de construire de nouveaux éléments finis qui engendreront des matrices de masse assez proches de matrices diagonales ?
L'idée d'une telle approche est licite dans la mesure où, du point de vue d'un numéricien, il ne faut pas voir une matrice comme étant diagonale ou non, mais plutôt comme ayant un profil optimal ou non.
On peut définir le profil d'une matrice symétrique de dimension $n \times n$ de la manière suivante (cette définition n'a rien de standard et nous sert uniquement à illustrer nos propos) : c'est la donnée d'un n-uplet dont le $i^{\text{ème}}$ élément correspond à la distance du premier élément non nul de la $i^{\text{ème}}$ ligne à l'élément diagonal. Par exemple une matrice diagonale aurait un profil égal à $(0, \ldots, 0)$ et une matrice pleine (dont aucun élément n'est nul) aurait un profil égal à $(0, 1, \ldots, n-1)$. Dans la pratique, il n'est nécessaire de stocker les éléments d'une matrice que sur son profil, les autres étant nuls. Le résultat qui nous intéresse est le suivant : toute matrice symétrique définie positive M admet une unique factorisation de Cholesky LL^t (ceci est un résultat classique d'algèbre linéaire) dont la matrice L à même profil que M (c'est un résultat que l'on vérifie aisément à la main). L'inversion de la matrice M passant par une descente-remontée sur L et L^t, il devient clair que le coût de celle-ci sera d'autant amoindri que l'on optimise le profil de L, c'est à dire celui de M.
L'idée de ce qui suit part du fait suivant : rendre la matrice de masse diagonale, c'est exactement orthogonaliser les fonctions de bases :

$$\int_T \phi_i \phi_j \, dX = \delta_{ij}.$$

C'est ce que l'on a fait jusqu'à présent de manière approchée, via l'utilisation de formules de quadrature :

$$\int_T \phi_i \phi_j \, dX \simeq \sum_k w_k \phi_i(X_k) \phi_j(X_k) = \delta_{ij}.$$

Pourquoi alors ne pas chercher à définir un ensemble de degrés de liberté $\{\sigma_1, \ldots, \sigma_N\}$, unisolvant sur un certain espace polynomial de dimension N, dont la base associée, c'est-à-dire l'unique ensemble de fonctions $\{\phi_1, \ldots, \phi_N\}$ vérifiant

$$\sigma_i(\phi_j) = \delta_{ij} \qquad \forall i, j = 1, \ldots, N$$

serait orthogonale.
La solution qui semble toute trouvée est la suivante : une fois déterminée une base orthogonale $\{\phi_1, \ldots, \phi_{\frac{(k+1)(k+2)}{2}}\}$ de l'espace $\mathbb{P}_k$, par exemple en utilisant le procédé de Gram-Schmidt sur la base canonique, nous pouvons définir un ensemble de formes linéaires par :

$$\sigma_i(P) = \int_T \phi_i P \, dX, \qquad \forall i = 1, \ldots, \frac{(k+1)(k+2)}{2}. \tag{4.2}$$

L'orthogonalité des ϕ_i impliquant que

$$\sigma_i(\phi_j) = \int_T \phi_i \phi_j \, dX = \delta_{ij},$$

l'ensemble de fonctions $\{\phi_1, \ldots, \phi_{\frac{(k+1)(k+2)}{2}}\}$ est par définition la base associée aux formes linéaires définies par (4.2).

Cette solution nous donne bien un élément fini, mais celui-ci n'est toutefois pas adapté au problème : comment assurer la conformité de cet élément dans H^1 ?
La restriction d'une fonction de $\mathbb{P}_k$ à n'importe quelle arête de T étant un polynôme univarié de degré k, il faudrait se fixer $k+1$ degrés de liberté comme étant des degrés de liberté partagés entre un élément et son voisin et pouvoir assurer que $k+1$ valeurs fixées prises par ces degrés de liberté déterminent de manière unique la restriction d'une fonction de $\mathbb{P}_k$ à cette arête, et ceci pour chacune des trois arêtes de T. Dans le cas où ceci serait possible, le recollement des arêtes étant aléatoire (dans le sens où il n'est pas possible d'assurer que, dans la numérotation relative propre à chaque élément issue de la transformation affine transformant l'élément de référence en chaque élément, toutes les arêtes du maillage ont le même numéro qu'elles soient vues en tant qu'arête d'un élément ou d'un de ses voisins), il faudrait en plus que $k+1$ valeurs fixées déterminent un unique polynôme, que ces valeurs soient associées aux $k+1$ premiers degrés de liberté déterminant le polynôme sur la première arête, aux $k+1$ seconds degrés de liberté déterminant le polynôme sur la seconde arête ou bien aux $k+1$ troisièmes degrés de liberté déterminant le polynôme sur la troisième arête.
Pour éclaircir la situation, il faut dans un premier temps revenir à la définition de la décomposition d'une fonction $P \in \mathbb{P}_k$ sur une base d'élément fini :

$$P(X) = \sum_{i=1}^{\frac{(k+1)(k+2)}{2}} \sigma_i(P)\phi_i(X). \tag{4.3}$$

Notons $\sigma_n(P) = \int_T \phi_n P \, dX = a_n$ la valeur du $n^{\text{ième}}$ degré de liberté. La définition (4.3) nous permet de remarquer la chose suivante : modifier la valeur prise pas le $n^{\text{ième}}$ degré de liberté ne modifie la fonction P que sur le support de ϕ_n. Cette constatation nous permet d'affirmer que si l'on veut que $k+1$ degrés de liberté définissent, de manière unique et indépendante du reste des degrés de liberté, la restriction d'une fonction de $\mathbb{P}_k$ à une arête, il faut et il suffit que les fonctions associées à ces degrés de liberté aient un support contenant cette arête alors que les supports des fonctions associées au reste des degrés de liberté ne doivent pas contenir cette arête. Et c'est bien là que réside tout le problème, car si construire une base de $\mathbb{P}_k$ orthogonale sur un triangle n'a rien de compliqué, gérer le support des fonctions qui composent cette base paraît très improbable à réaliser, d'autant que ces fonctions doivent respecter les conditions de symétrie décrites ci-dessus.
S'il est impossible pour nous de dire, en toute généralité, si oui ou non une telle base de polynômes orthogonaux existe, et encore moins d'en envisager la construction systématique par un algorithme à définir, nous pouvons toutefois utiliser ces considérations pour orthogonaliser un maximum de fonctions de base de manière à optimiser le profil de la matrice de masse.
L'idée est alors de décomposer l'espace $\mathbb{P}_k$ en somme directe de deux espaces, l'un qui définira un espace de fonctions dont le support ne contient pas les arêtes du triangle, c'est à dire le sous-espace de $\mathbb{P}_k$ constitué des fonctions à trace nulle sur les arêtes, l'autre étant par définition le supplémentaire de cet espace dans $\mathbb{P}_k$. Cette décomposition est explicite

et peut être écrite sous la forme (pour $k \geq 3$) :

$$\mathbb{P}_k = b\mathbb{P}_{k-3} \oplus \mathbb{H}$$

Ceci nous permettra :
- de définir dans un premier temps une base orthogonale $\{\phi_i\}_{i=1,\ldots,\frac{(k-2)(k-1)}{2}}$ de $b\mathbb{P}_{k-3}$ et d'associer à chacune de ces fonctions de base la forme linéaire $\sigma_i(P) = \int_T \phi_i P \, dX$,
- de construire ensuite les $3k$ fonctions de base $\{\phi_i\}_{i=\frac{(k-2)(k-1)}{2}+1,\ldots,\frac{(k+1)(k+2)}{2}}$ de l'espace $\mathbb{H}$ comme étant non seulement lagrangiennes sur un jeu de $3k$ points $\{X_i\}$ symétriques sur les arêtes assurant la conformité de type $\mathbb{P}_k$ de l'élément fini (les points et les formes linéaires qui leur sont associées étant numérotés de $\frac{(k-2)(k-1)}{2}+1$ à $\frac{(k+1)(k+2)}{2}$) mais aussi orthogonales aux $\frac{(k-2)(k-1)}{2}$ fonctions ϕ_i pour $i = 1, \ldots, \frac{(k-2)(k-1)}{2}$.

Ainsi

$$\sigma_i(\phi_j) = \int_T \phi_i \phi_j \, dX = \delta_{ij}, \qquad \begin{array}{rcl} \forall i & = & 1, \ldots, \frac{(k-2)(k-1)}{2} \\ \forall j & = & 1, \ldots, \frac{(k+1)(k+2)}{2} \end{array},$$

par orthogonalité de l'ensemble des fonctions de bases contre les $\frac{(k-2)(k-1)}{2}$ premières fonctions de base que l'on a définies,

$$\sigma_i(\phi_j) = \phi_j(X_i) = \delta_{ij}(=0), \qquad \begin{array}{rcl} \forall i & = & \frac{(k-2)(k-1)}{2}+1, \ldots, \frac{(k+1)(k+2)}{2} \\ \forall j & = & 1, \ldots, \frac{(k-2)(k-1)}{2} \end{array},$$

les $\frac{(k-2)(k-1)}{2}$ premières fonctions de bases étant par définition identiquement nulles sur les arêtes du triangle et donc en particulier aux points X_i, et

$$\sigma_i(\phi_j) = \phi_j(X_i) = \delta_{ij}, \qquad \begin{array}{rcl} \forall i & = & \frac{(k-2)(k-1)}{2}+1, \ldots, \frac{(k+1)(k+2)}{2} \\ \forall j & = & \frac{(k-2)(k-1)}{2}+1, \ldots, \frac{(k+1)(k+2)}{2} \end{array},$$

les $3k$ dernières fonctions de base étant par définition lagrangiennes aux points X_i.
L'ensemble $\{\phi_i\}_{i=1,\ldots,\frac{(k+1)(k+2)}{2}}$ définie alors bien la base d'élément fini associée aux $\frac{(k+1)(k+2)}{2}$ formes linéaires que l'on a définies (dans la mesure où $\sigma_i(\phi_j) = \delta_{ij}, \forall i, j = 1, \ldots, \frac{(k+1)(k+2)}{2}$).
La trace sur l'une des arêtes du triangle d'une quelconque fonction de $\mathbb{P}_k$ étant entièrement définie par $k+1$ des formes linéaires lagrangiennes, la conformité de l'élément peut être assurée. Nous avons donc bien défini un élément fini conforme dans H^1, mais pour cet élément, seules les $3k$ fonctions de base associées aux formes linéaires lagrangiennes ne sont pas orthogonales entre elles.

Nous illustrons ceci sur l'exemple de $\mathbb{P}_4$: le lemme 4.3.1 nous dit qu'une base de $\mathbb{P}_4$ peut être donné par l'ensemble de fonctions

$$\begin{array}{l} \{\Lambda_1^4, \Lambda_2^4, \Lambda_3^4, \\ \Lambda_1^3\Lambda_2, \Lambda_1^3\Lambda_3, \Lambda_2^3\Lambda_1, \Lambda_2^3\Lambda_3, \Lambda_3^3\Lambda_1, \Lambda_3^3\Lambda_2, \\ \Lambda_1^2\Lambda_2^2, \Lambda_1^2\Lambda_3^2, \Lambda_2^2\Lambda_3^2, \\ \Lambda_1^2\Lambda_2\Lambda_3, \Lambda_2^2\Lambda_1\Lambda_3, \Lambda_3^2\Lambda_1\Lambda_2\} \end{array}$$

Nous pouvons alors expliciter la décomposition

$$\mathbb{P}_4 = b\mathbb{P}_1 \oplus \mathbb{H}$$

en donnant une base de $b\mathbb{P}_1$ et de H par :

$$b\mathbb{P}_1 =< \Lambda_1^2\Lambda_2\Lambda_3, \Lambda_2^2\Lambda_1\Lambda_3, \Lambda_3^2\Lambda_1\Lambda_2 >,$$

$$\mathbb{H} =< \Lambda_1^4, \Lambda_2^4, \Lambda_3^4, \Lambda_1^3\Lambda_2, \Lambda_1^3\Lambda_3, \Lambda_2^3\Lambda_1, \Lambda_2^3\Lambda_3, \Lambda_3^3\Lambda_1, \Lambda_3^3\Lambda_2, \Lambda_1^2\Lambda_2^2, \Lambda_1^2\Lambda_3^2, \Lambda_2^2\Lambda_3^2 > .$$

Nous utilisons le procédé de Gram-Schmidt pour déterminer une base orthogonale $\{\phi_1, \phi_2, \phi_3\}$ de $b\mathbb{P}_1$, où

$$\begin{aligned}
\phi_1 &:= \Lambda_1^2\Lambda_2\Lambda_3 \\
\phi_2 &:= \Lambda_2^2\Lambda_1\Lambda_3 - \tfrac{3}{4}\Lambda_1^2\Lambda_2\Lambda_3 \\
\phi_3 &:= \Lambda_3^2\Lambda_1\Lambda_2 - \tfrac{3}{7}\Lambda_2^2\Lambda_1\Lambda_3 - \tfrac{3}{7}\Lambda_1^2\Lambda_2\Lambda_3
\end{aligned}$$

Il faut ensuite se fixer un jeu de points $\{X_i\}_{i=4,\ldots,15}$, symétrique sur les arêtes (le nombre de points étant fixé par la dimension de l'espace $\mathbb{H}$, ici 12, sachant qu'il faut 5 points par arête symétriquement localisés et de manière identique sur les trois arêtes pour la conformité) : par exemple le jeu de points standard de l'élément fini P_4 (il n'y a finalement aucune raison d'en changer). Il suffit alors de résoudre un système linéaire pour chacune des fonctions de base restant à déterminer, ce système étant formé des équations issues des égalités :

$$\sigma_i(\phi_j) = \delta_{ij},$$

où $\{\phi_j\}_{i=4,\ldots,15}$ désigne un ensemble de polynômes quelconques de $\mathbb{P}_4$ pris sous leur forme générique, et

$$\begin{aligned}
&\sigma_i(P) = \textstyle\int_T \phi_i P \, dX \text{ pour } i = 1, \ldots, 3; \\
&\sigma_i(P) = P(X_i) \text{ pour } i = 4, \ldots, 15.
\end{aligned}$$

Remarque 4.4.1. *Il peut être bon de se demander comment évoluent le conditionnement de la matrice de masse et la constante de Lebesgue avec la condensation des éléments finis. À titre de comparaison nous reprenons dans le tableau 4.1 le tableau 3.1 en y ajoutant les constantes de Lebesgue et le conditionnement de la matrice de masse en norme L^2 pour les éléments finis condensés et partiellement condensés décrits dans les sections 4.3 et 4.4.*

Il faut remarquer que les constantes de Lebesgue associées aux éléments finis condensés sont du même ordre que les constantes de Lebesgue associées aux éléments finis standards avec une localisation équirépartie des points, alors que le conditionnement de la matrice de masse est bien meilleur pour les éléments finis condensés que le conditionnement de la matrice de masse des éléments finis standards (que la localisation des points soit optimisée ou non). D'autre part nous remarquons que le conditionnement de la matrice de masse et la constante de Lebesgue sont nettement plus élevés lorsque l'on utilise la condensation partielle des éléments finis : cela est dû au fait que le procédé de Gram-Schmidt engendre de mauvaises bases orthonormées du sous-espace de l'espace polynomial associé à l'élément fini constitué des polynômes à trace nulle sur les arêtes du triangle, dans le sens où les fonctions de ces bases sont fortement oscillantes. Si, comme nous allons le voir dans la section 6 pour les éléments finis de lagrange partiellement condensés P_3 à P_5, cela ne pénalise pas encore les schémas, il faut toutefois rester conscient qu'il sera nécessaire de construire des bases

Éléments finis standards localisation équirépartie	P_3	P_4	P_5	P_6	P_7
conditionnement	67.66	114.41	239.44	548.00	1340.77
constante de Lebesgue	2.27	3.47	5.45	8.7	14.3
Éléments finis standards localisation optimisée	P_3	P_4	P_5	P_6	P_7
conditionnement	68.73	101.32	134.77	218.73	329.31
constante de Lebesgue	2.12	2.63	3.19	4.06	4.75
Éléments finis condensés	$\tilde{P}_3$	$\tilde{P}_4$	$\tilde{P}_5$	$\tilde{P}_6$	$\tilde{P}_7$
conditionnement	14.84	24.81	64.70	124.10	
constante de Lebesgue	2.22	3.75	5.24	9.65	
Éléments finis partiellement condensés	$\tilde{P}_3$	$\tilde{P}_4$	$\tilde{P}_5$	$\tilde{P}_6$	$\tilde{P}_7$
conditionnement	299.25	660	1295.28	2301.12	
constante de Lebesgue	3.48	6.9	11.5	17.5	

TAB. 4.1 – Conditionnement en norme L^2 de la matrice de masse et constante de Lebesgue calculés sur l'élément de référence pour les éléments finis P_3 à P_7.

orthonormées de manière plus réfléchie pour les ordres plus élevés (de la même manière, finalement, qu'il nous faut réfléchir à la localisation des points pour les éléments finis de Lagrange standards : le problème de la "bonne" définition des degrés de libertés générant une "bonne" base reste le même pour la condensation partielle des éléments finis).

4.5 Condensation des éléments finis d'arête

Dans la classe des éléments finis d'arête que nous allons considérer, le problème de la condensation de la matrice de masse est un problème encore très ouvert. À notre connaissance les seuls cas particuliers ayant été résolus sont, le cas des éléments finis d'arête triangulaires de plus bas degré par Y. Haugazeau et P. Lacoste [51] mais sous des contraintes très restrictives, la stabilité de la méthode étant fortement dépendante de la qualité du maillage (aucun angle d'aucun des triangles du maillage ne devant être obtus), et le cas particulier des éléments finis d'arête rectangulaires sur maillage cartésien par G. Cohen et P. Monk [27][28] qui ramènent le problème de la condensation des éléments finis d'arête rectangulaires au problème de la condensation des éléments finis de Lagrange en une dimension d'espace en définissant les degrés de liberté par des formes linéaires Lagrangiennes. Nous allons voir que la définition que l'on a fait des éléments finis d'arête rectangulaires dans la sous-section 3.2.1 nous permet à nous aussi de condenser la matrice de masse sur des maillages cartésiens, mais en utilisant partiellement l'orthogonalité des polynômes de Legendre (le reste de la condensation étant pour nous aussi ramené au problème de la condensation des éléments finis de Lagrange en une dimension d'espace).

4.5.1 Condensation des éléments finis d'arête rectangulaires sur maillage régulier

Nous reconsidérons ici les éléments finis d'arête rectangulaires définis dans la sous-section 3.2.1. Les fonctions de bases associées à ces éléments finis, que l'on a déjà déterminées dans la démonstration de la proposition 3.2.5, ayant une forme bien particulière, nous permettent d'envisager une condensation de la matrice de masse associée à ces éléments. En effet il suffit d'expliciter ce que vaut

$$\int_K \overrightarrow{\psi^m_{\xi_i}}.\overrightarrow{\psi^n_{\eta_j}}\ dX =$$
$$\int_K \left(L_{\xi_i}(\xi) l_m(\overline{\xi}) \begin{pmatrix} \delta_{\xi y} \\ \delta_{\xi x} \end{pmatrix} \right) \cdot \left(L_{\eta_j}(\eta) l_n(\overline{\eta}) \begin{pmatrix} \delta_{\eta y} \\ \delta_{\eta x} \end{pmatrix} \right)\ dX =$$
$$\delta_{\xi\eta} \int_0^1 L_{\xi_i}(\xi) L_{\xi_j}(\xi)\ d\xi \int_0^1 l_m(\overline{\xi}) l_n(\overline{\xi})\ d\overline{\xi} =$$
$$\delta_{\xi\eta}\delta_{mn} \int_0^1 L_{\xi_i}(\xi) L_{\xi_j}(\xi)\ d\xi,$$

pour se rendre compte que le seul facteur gênant est une intégrale unidimensionnelle d'un produit de fonctions de base de Lagrange, ce qui signifie que le problème de la condensation de la matrice de masse issue des éléments finis vectoriels rectangulaires, conformes dans $H(rot, \Omega)$, se résume au problème de condensation de la matrice de masse issue des éléments finis de Lagrange en une dimension d'espace, problème que l'on a déjà résolu : il suffit de localiser les $k+1$ points $\{\xi_i\}_{i=0,\ldots,k}$ non plus de manière équirépartie mais aux points de quadrature de la formule de quadrature à $k+1$ points de Gauss-Lobatto de sorte que :

$$\int_K \overrightarrow{\psi^m_{\xi_i}}.\overrightarrow{\psi^n_{\eta_j}}\ dX = \delta_{\xi\eta}\delta_{mn} \int_0^1 L_{\xi_i}(\xi) L_{\xi_j}(\xi)\ d\xi \simeq \omega_i \delta_{ij} \delta_{\xi\eta} \delta_{mn},$$

où $\{\omega_i\}_{i=0,\ldots,k}$ désigne l'ensemble des poids associés aux points de quadrature. Ceci signifie bien entendu que la matrice de masse peut être remplacée par une matrice diagonale.

Remarque 4.5.1. *Il est indispensable de se rendre compte que cette condensation de la matrice de masse n'est possible que sur des maillages réguliers. En effet, nous venons de voir qu'il est possible de condenser la matrice de masse élémentaire, c'est-à-dire la matrice de masse calculée sur l'élément de référence $K = [0,1]^2$. Cela n'est plus suffisant pour impliquer la condensation de la matrice de masse en toute généralité : pour cela il faudrait que toute fonction de base "physique", c'est-à- dire les fonctions de base qui sont déterminés comme les transformations des fonctions de bases sur l'élément de référence, ait la forme bien particulière d'un produit de fonctions de Legendre contre une fonction de Lagrange portant sur l'une ou l'autre composante, ce qui n'est pas le cas étant donné que la transformation transportant l'élément de référence vers un élément du maillage n'est en général plus affine mais bilinéaire, ce qui implique que les fonctions de bases ne seront plus données en terme de produit de fonctions de Legendre contre des fonctions de Lagrange évaluées en x ou y, mais qu'elles seront évaluées en une combinaison linéaire de ces quantités, et que donc il ne sera plus possible de découpler l'intégrale sur l'élément en produit d'intégrales unidimensionnelles suivant chacune des deux directions x et y.*

Ordre de l'élément fini d'arête localisation équirépartie	1	2	3	4	5
conditionnement	3	9	12.19	19.14	32.53
constante de Lebesgue	1	3.4	8.1	16.75	32.8
Ordre de l'élément fini d'arête localisation optimisée	1	2	3	4	5
conditionnement	3	9	11.07	15	17.23
constante de Lebesgue	1	3.4	7.42	12.4	18.8
Ordre de l'élément fini d'arête condensé	1	2	3	4	5
conditionnement	1	4	5	7.11	8.32
constante de Lebesgue	1	3.4	7.42	12.4	18.8

TAB. 4.2 – Conditionnement en norme L^2 de la matrice de masse et constante de Lebesgue calculés sur l'élément de référence pour les éléments finis d'arête rectangulaires d'ordre 1 à 5.

Remarque 4.5.2. *Nous reprenons dans le tableau 4.2 le tableau 3.2 en y ajoutant le conditionnement en norme L^2 de la matrice de masse et la constante de Lebesgue calculés sur l'élément de référence pour les éléments finis d'arête condensés.*
La localisation des points x_i et y_i étant la même pour les éléments finis d'arête rectangulaires avec localisation des points optimisée et les éléments finis d'arête rectangulaires condensés (localisation aux points des formules de quadrature de Gauss-Lobatto), les constantes de Lebsgue sont naturellement les mêmes. En revanche nous remarquons que la matrice de masse condensée est bien mieux conditionnée.

4.5.2 Cas particulier de la condensation des éléments finis d'arête de premier ordre : Schéma de Yee

Nous allons nous intéresser au cas particulier de la condensation de l'élément fini d'arête sur un quadrangle d'ordre le plus bas pour montrer que cette discrétisation en espace couplée à une discrétisation en temps symplectique d'ordre 2 (nous renvoyons le lecteur à la section 5.1.5 pour se rendre compte que ce schéma n'est autre qu'un schéma en temps du type "saute-mouton") se confond avec la méthode classique de résolution des équations de Maxwell par un schéma de Yee [72]. Cela signifie en particulier que les schémas d'éléments finis avec condensation de la matrice de masse couplés avec des schémas en temps symplectiques peuvent être vus comme une généralisation aux ordres plus élevés du schéma de Yee couramment utilisé pour la résolution des équations de Maxwell.
Nous reconsidérons notre problème initial écrit de la manière suivante :

$$\left\{\begin{array}{rcl} \dfrac{\partial E_x}{\partial t} - \dfrac{\partial B_z}{\partial y} & = & -J_x \\ \dfrac{\partial E_y}{\partial t} + \dfrac{\partial B_z}{\partial x} & = & -J_y \\ \dfrac{\partial B_z}{\partial t} + \dfrac{\partial E_y}{\partial x} - \dfrac{\partial E_x}{\partial y} & = & 0 \end{array}\right.$$

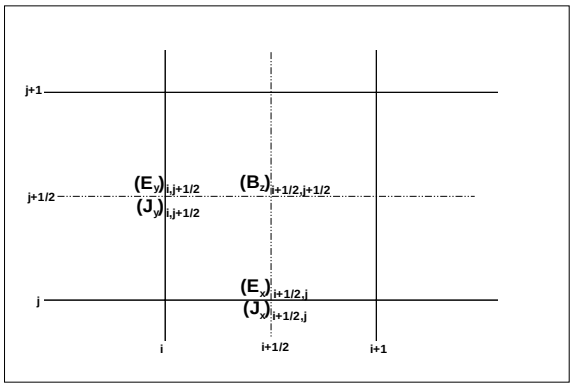

FIG. 4.11 – Localisation des champs discrets par la discrétisation de Yee

La figure 4.11 représente un élément du maillage et la localisation des champs discrets associés à la discrétisation par le schéma de Yee, que l'on écrit de la manière suivante :

$$\left\{\begin{array}{rcl} \frac{(E_x)^{n+1}_{i+\frac{1}{2},j}-(E_x)^{n}_{i+\frac{1}{2},j}}{\Delta t}-\frac{(B_z)^{n+\frac{1}{2}}_{i+\frac{1}{2},j+\frac{1}{2}}-(B_z)^{n+\frac{1}{2}}_{i+\frac{1}{2},j-\frac{1}{2}}}{\Delta y} &=& -(J_x)^{n+\frac{1}{2}}_{i+\frac{1}{2},j} \\ \frac{(E_y)^{n+1}_{i,j+\frac{1}{2}}-(E_y)^{n}_{i,j+\frac{1}{2}}}{\Delta t}+\frac{(B_z)^{n+\frac{1}{2}}_{i+\frac{1}{2},j+\frac{1}{2}}-(B_z)^{n+\frac{1}{2}}_{i-\frac{1}{2},j+\frac{1}{2}}}{\Delta x} &=& -(J_y)^{n+\frac{1}{2}}_{i,j+\frac{1}{2}} \\ \frac{(B_z)^{n+\frac{1}{2}}_{i+\frac{1}{2},j+\frac{1}{2}}-(B_z)^{n-\frac{1}{2}}_{i+\frac{1}{2},j+\frac{1}{2}}}{\Delta t}+\frac{(E_y)^{n}_{i+1,j+\frac{1}{2}}-(E_y)^{n}_{i,j+\frac{1}{2}}}{\Delta x}-\frac{(E_x)^{n}_{i+\frac{1}{2},j+1}-(E_x)^{n}_{i+\frac{1}{2},j}}{\Delta y} &=& 0 \end{array}\right. \tag{4.4}$$

où bien entendu $(E_x)^n_{i+\frac{1}{2},j}$ désigne une approximation de la valeur de la première composante du champ électrique au point $(i+\frac{1}{2},j)$ pris au temps $t^n = n\Delta t$, $(E_y)^n_{i,j+\frac{1}{2}}$ désigne une approximation de la valeur de la deuxième composante du champ électrique au point $(i,j+\frac{1}{2})$ pris au temps t^n, les champs $(J_x)^n_{i+\frac{1}{2},j}$ et $(J_y)^n_{i,j+\frac{1}{2}}$ étant interprétés de la même manière, et finalement $(B_z)^{n+\frac{1}{2}}_{i+\frac{1}{2},j+\frac{1}{2}}$ désigne une approximation de la valeur du champ magnétique au point $(i+\frac{1}{2},j+\frac{1}{2})$ pris au temps $t^{n+\frac{1}{2}} = (n+\frac{1}{2})\Delta t$.
Considérons alors les éléments finis de plus bas degré. Les 4 fonctions de bases définissant localement sur l'élément de référence $[0,\Delta x]\times[0,\Delta y]$ l'espace W sont données, dans une numérotation locale des arêtes auxquelles sont associés les degrés de liberté qui les définissent (figure 4.12), par

$$\psi_1 = \begin{pmatrix} (1-\frac{y}{\Delta y})\frac{1}{\sqrt{\Delta x}} \\ 0 \end{pmatrix}, \psi_2 = \begin{pmatrix} \frac{y}{\Delta y}\frac{1}{\sqrt{\Delta x}} \\ 0 \end{pmatrix},$$
$$\psi_3 = \begin{pmatrix} 0 \\ (1-\frac{x}{\Delta x})\frac{1}{\sqrt{\Delta y}} \end{pmatrix}, \psi_4 = \begin{pmatrix} 0 \\ \frac{x}{\Delta x}\frac{1}{\sqrt{\Delta y}} \end{pmatrix},$$

et bien entendu l'unique fonction de base définissant localement sur ce même élément l'espace V est donné par $\varphi_1 = 1$.

Il convient alors de réécrire le système (3.4) (que l'on réécrit dans un premier temps

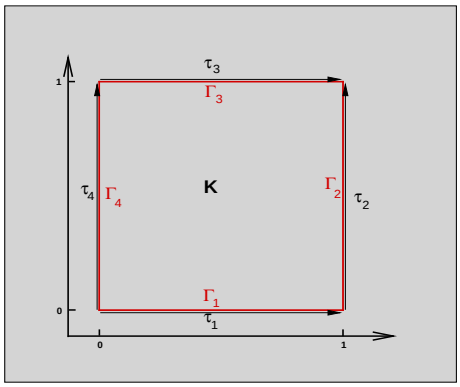

FIG. 4.12 – Numérotation locale des arêtes de l'élément de référence

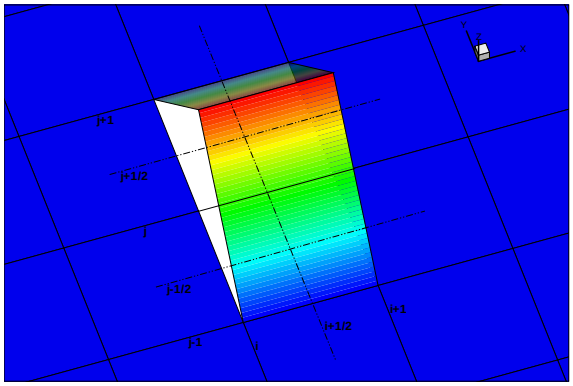

FIG. 4.13 – Première composante de $\psi_{i+\frac{1}{2},j}$ (la seconde étant identiquement nulle).

ci-dessous en renommant les indices pour éviter les conflits)

$$\left\{\begin{array}{ll} \dfrac{d}{dt}\displaystyle\sum_k e_k \int_\Omega \overrightarrow{\psi_k}\cdot\overrightarrow{\psi_n}\,dX - \sum_l b_l \int_\Omega \varphi_l(\nabla\times\overrightarrow{\psi_n})\,dX & = -\displaystyle\int_\Omega \overrightarrow{J}\cdot\overrightarrow{\psi_n}\,dX, \\ \dfrac{d}{dt}\displaystyle\sum_l b_l \int_\Omega \varphi_l\varphi_m\,dX + \sum_k e_k \int_\Omega (\nabla\times\overrightarrow{\psi_k})\varphi_m\,dX & = 0, \end{array}\right. \tag{4.5}$$

dans une numérotation globale des arêtes et des centres des éléments similaire à celle utilisée pour expliciter le schéma de Yee. Plus précisément $\overrightarrow{\psi}_{i+\frac{1}{2},j}$ désignera la fonction de base de l'espace W associée à l'unique degré de liberté portant sur l'arête horizontale ayant pour centre les coordonnées $(i+\frac{1}{2},j)$ (voir figure 4.13), $\overrightarrow{\psi}_{i,j+\frac{1}{2}}$ désignera la fonction de base de l'espace W associée à l'unique degré de liberté portant sur l'arête verticale ayant pour centre les coordonnées $(i,j+\frac{1}{2})$ (voir figure 4.14), et enfin $\varphi_{i+\frac{1}{2},j+\frac{1}{2}}$ désignera la fonction de base de l'espace V associée à l'unique degré de liberté portant sur l'élément $K_{i+\frac{1}{2},j+\frac{1}{2}}$, c'est-à-dire l'élément ayant pour centre les coordonnées $(i+\frac{1}{2},j+\frac{1}{2})$.

Considérons dans un premier temps une arête horizontale d'indice $n=(i+\frac{1}{2},j)$. Remar-

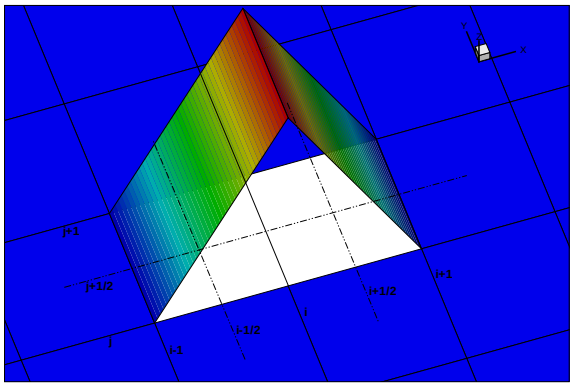

FIG. 4.14 – Seconde composante de $\psi_{i,j+\frac{1}{2}}$ (la première étant identiquement nulle).

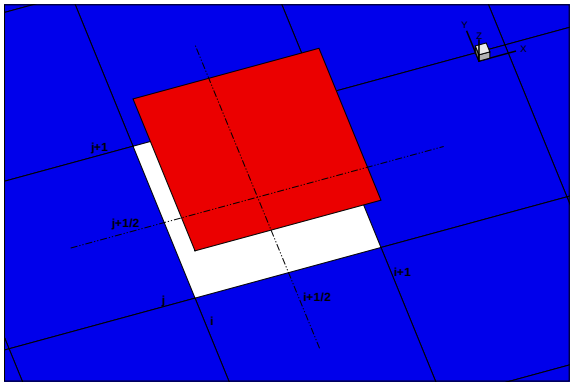

FIG. 4.15 – $\varphi_{i+\frac{1}{2},j+\frac{1}{2}}$.

quons que les seuls termes non-nuls de la combinaison linéaire

$$\sum_k e_k \int_\Omega \overrightarrow{\psi_k} \cdot \overrightarrow{\psi}_{i+\frac{1}{2},j}\, dX$$

sont les termes d'indices $k = (i+\frac{1}{2}, j-1)$, $k = (i+\frac{1}{2}, j)$ et $k = (i+\frac{1}{2}, j+1)$. En effet, pour les indices portant sur les arêtes verticales, les fonctions de base sont identiquement nulles sur leur première composante d'où la nullité du produit scalaire, tandis que pour les fonctions associées aux indices portant sur les arêtes horizontales, seules celles énumérées n'ont pas un support disjoint de celui de $\overrightarrow{\psi}_{i+\frac{1}{2},j}$. Si de plus nous faisons le choix de condenser la matrice de masse, alors le seul indice k apportant une contribution à cette somme est $k = (i+\frac{1}{2}, j)$ et celle-ci est alors explicitement approchée par

$$\sum_k e_k \int_\Omega \overrightarrow{\psi_k} \cdot \overrightarrow{\psi}_{i+\frac{1}{2},j}\, dX \simeq e_{i+\frac{1}{2},j}\Delta y,$$

dont $e_{i+\frac{1}{2},j}\frac{\Delta y}{2}$ pour la contribution de la quadrature de $\overrightarrow{\psi}_{i+\frac{1}{2},j} \cdot \overrightarrow{\psi}_{i+\frac{1}{2},j}$ sur l'élément $K_{i+\frac{1}{2},j-\frac{1}{2}}$ et $e_{i+\frac{1}{2},j}\frac{\Delta y}{2}$ pour la contribution de la quadrature de $\overrightarrow{\psi}_{i+\frac{1}{2},j} \cdot \overrightarrow{\psi}_{i+\frac{1}{2},j}$ sur l'élément $K_{i+\frac{1}{2},j+\frac{1}{2}}$.
De la même manière, les seuls termes non nuls dans la somme

$$\sum_l b_l \int_\Omega \varphi_l (\nabla \times \overrightarrow{\psi_n})\, dX,$$

sont ceux d'indice $l = (i+\frac{1}{2}, j-\frac{1}{2})$ et $l = (i+\frac{1}{2}, j+\frac{1}{2})$, et cette somme devient alors

$$\sum_l b_l \int_\Omega \varphi_l (\nabla \times \overrightarrow{\psi_n})\, dX = -b_{i+\frac{1}{2},j-\frac{1}{2}}\sqrt{\Delta x} + b_{i+\frac{1}{2},j+\frac{1}{2}}\sqrt{\Delta x}.$$

En choisissant de calculer le second membre de l'équation en projetant $\overrightarrow{J}$ sur l'espace d'éléments finis et en approchant les intégrales de produits de fonctions de base via la technique de condensation de masse il vient naturellement

$$-\int_\Omega \overrightarrow{J} \cdot \overrightarrow{\psi_n}\, dX \simeq -J_{i+\frac{1}{2},j}\Delta y.$$

De sorte que l'approximation de la première composante du champ électrique, qui, rappelons-le, est entièrement déterminée par les valeurs des degrés de libertés associés aux arêtes horizontales du maillage sera donnée comme la solution du système différentiel

$$\frac{d}{dt}e_{i+\frac{1}{2},j}\Delta y - (-b_{i+\frac{1}{2},j-\frac{1}{2}}\sqrt{\Delta x} + b_{i+\frac{1}{2},j+\frac{1}{2}}\sqrt{\Delta x}) = -J_{i+\frac{1}{2},j}\Delta y$$

ou de manière totalement équivalente

$$\frac{d}{dt}\frac{e_{i+\frac{1}{2},j}}{\sqrt{\Delta x}} - \frac{b_{i+\frac{1}{2},j+\frac{1}{2}} - b_{i+\frac{1}{2},j-\frac{1}{2}}}{\Delta y} = -\frac{J_{i+\frac{1}{2},j}}{\sqrt{\Delta x}}.$$

Si maintenant nous considérons une arête verticale d'indice $n = (i, j + \frac{1}{2})$, une argumentation similaire explicite la première équation du système (4.5) par

$$\frac{d}{dt}\frac{e_{i,j+\frac{1}{2}}}{\sqrt{\Delta y}} + \frac{b_{i+\frac{1}{2},j+\frac{1}{2}} - b_{i-\frac{1}{2},j+\frac{1}{2}}}{\Delta x} = -\frac{J_{i,j+\frac{1}{2}}}{\sqrt{\Delta y}}.$$

Pour finir nous explicitons la seconde équation de (4.5) en $m = (i+\frac{1}{2}, j+\frac{1}{2})$ qui devient

$$\frac{d}{dt}b_{i+\frac{1}{2},j+\frac{1}{2}} + \frac{\frac{e_{i+1,j+\frac{1}{2}}}{\sqrt{\Delta y}} - \frac{e_{i,j+\frac{1}{2}}}{\sqrt{\Delta y}}}{\Delta x} - \frac{\frac{e_{i+\frac{1}{2},j+1}}{\sqrt{\Delta x}} - \frac{e_{i+\frac{1}{2},j}}{\sqrt{\Delta x}}}{\Delta y} = 0.$$

Le système (4.5) se réécrit donc sous la forme

$$\left\{\begin{array}{rcl} \frac{d}{dt}\frac{e_{i+\frac{1}{2},j}}{\sqrt{\Delta x}} - \frac{b_{i+\frac{1}{2},j+\frac{1}{2}} - b_{i+\frac{1}{2},j-\frac{1}{2}}}{\Delta y} & = & -\frac{J_{i+\frac{1}{2},j}}{\sqrt{\Delta x}} \\ \frac{d}{dt}\frac{e_{i,j+\frac{1}{2}}}{\sqrt{\Delta y}} + \frac{b_{i+\frac{1}{2},j+\frac{1}{2}} - b_{i-\frac{1}{2},j+\frac{1}{2}}}{\Delta x} & = & -\frac{J_{i,j+\frac{1}{2}}}{\sqrt{\Delta y}} \\ \frac{d}{dt}b_{i+\frac{1}{2},j+\frac{1}{2}} + \frac{\frac{e_{i+1,j+\frac{1}{2}}}{\sqrt{\Delta y}} - \frac{e_{i,j+\frac{1}{2}}}{\sqrt{\Delta y}}}{\Delta x} - \frac{\frac{e_{i+\frac{1}{2},j+1}}{\sqrt{\Delta x}} - \frac{e_{i+\frac{1}{2},j}}{\sqrt{\Delta x}}}{\Delta y} & = & 0 \end{array}\right. . \tag{4.6}$$

Il nous reste alors à introduire la discrétisation symplectique d'ordre deux en temps, plus connue sous le terme de schéma "saute-mouton". Les deux premières équations du système (4.5.2) font apparaître des dérivées en temps portant sur le champ électrique et nécessitent la connaissance du champ magnétique (et du courant), tandis que la troisième équation de (4.5.2) fait apparaître une dérivée en temps portant sur le champ magnétique et nécessite la connaissance du champ électrique. Il est donc naturel de résoudre les deux premières équations de ce système à un temps donné (chaque demi pas de temps, par exemple) et de décaler la résolution de la troisième équation d'un demi pas de temps (chaque pas de temps entier, donc...). Explicitement cela donne après discrétisation des dérivées temporelles par

$$\frac{d}{dt}\frac{e^{n+\frac{1}{2}}_{i+\frac{1}{2},j}}{\sqrt{\Delta x}} \simeq \frac{\frac{e^{n+1}_{i+\frac{1}{2},j}}{\sqrt{\Delta x}} - \frac{e^{n}_{i+\frac{1}{2},j}}{\sqrt{\Delta x}}}{\Delta t}$$

$$\frac{d}{dt}\frac{e^{n+\frac{1}{2}}_{i,j+\frac{1}{2}}}{\sqrt{\Delta y}} \simeq \frac{\frac{e^{n+1}_{i,j+\frac{1}{2}}}{\sqrt{\Delta y}} - \frac{e^{n}_{i,j+\frac{1}{2}}}{\sqrt{\Delta y}}}{\Delta t}$$

$$\frac{d}{dt}b^{n}_{i+\frac{1}{2},j+\frac{1}{2}} \simeq \frac{b^{n+\frac{1}{2}}_{i+\frac{1}{2},j+\frac{1}{2}} - b^{n-\frac{1}{2}}_{i+\frac{1}{2},j+\frac{1}{2}}}{\Delta t}$$

le système suivant

$$\left\{\begin{array}{rcl} \dfrac{\frac{e^{n+1}_{i+\frac{1}{2},j}}{\sqrt{\Delta x}}-\frac{e^{n}_{i+\frac{1}{2},j}}{\sqrt{\Delta x}}}{\Delta t}-\dfrac{b^{n+\frac{1}{2}}_{i+\frac{1}{2},j+\frac{1}{2}}-b^{n+\frac{1}{2}}_{i+\frac{1}{2},j-\frac{1}{2}}}{\Delta y} & = & -\dfrac{J^{n+\frac{1}{2}}_{i+\frac{1}{2},j}}{\sqrt{\Delta x}} \\ \dfrac{\frac{e^{n+1}_{i,j+\frac{1}{2}}}{\sqrt{\Delta y}}-\frac{e^{n}_{i,j+\frac{1}{2}}}{\sqrt{\Delta y}}}{\Delta t}+\dfrac{b^{n+\frac{1}{2}}_{i+\frac{1}{2},j+\frac{1}{2}}-b^{n+\frac{1}{2}}_{i-\frac{1}{2},j+\frac{1}{2}}}{\Delta x} & = & -\dfrac{J^{n+\frac{1}{2}}_{i,j+\frac{1}{2}}}{\sqrt{\Delta y}} \\ \dfrac{b^{n+\frac{1}{2}}_{i+\frac{1}{2},j+\frac{1}{2}}-b^{n-\frac{1}{2}}_{i+\frac{1}{2},j+\frac{1}{2}}}{\Delta t}+\dfrac{\frac{e^{n}_{i+1,j+\frac{1}{2}}}{\sqrt{\Delta y}}-\frac{e^{n}_{i,j+\frac{1}{2}}}{\sqrt{\Delta y}}}{\Delta x}-\dfrac{\frac{e^{n}_{i+\frac{1}{2},j+1}}{\sqrt{\Delta x}}-\frac{e^{n}_{i+\frac{1}{2},j}}{\sqrt{\Delta x}}}{\Delta y} & = & 0 \end{array}\right. . \tag{4.7}$$

On reconnaît alors ici le système (4.4) obtenu par la discrétisation de la méthode de Yee, modulo certaines normalisations. En effet si pour le champ magnétique, les $b_{i+\frac{1}{2},j+\frac{1}{2}}$ correspondent exactement aux valeurs de ce champ aux milieux des éléments (rappelons que nous avons fait le choix de résoudre le champ magnétique en utilisant des éléments finis de Lagrange...), c'est-à-dire que l'on a exactement $b_{i+\frac{1}{2},j+\frac{1}{2}} = (B_z)_{i+\frac{1}{2},j+\frac{1}{2}}$, cela n'est plus exactement le cas pour le champ électrique. Celui-ci étant calculé à l'aide des éléments finis d'arêtes, les $e_{i+\frac{1}{2},j}$ et $e_{i,j+\frac{1}{2}}$ ne sont plus respectivement les valeurs de la première composante au point $(i+\frac{1}{2},j)$ et de la deuxième composante au point $(i+\frac{1}{2},j+\frac{1}{2})$ mais les valeurs des degrés de libertés associés aux arêtes ayant pour milieux ces points. C'est-à-dire que $e_{i+\frac{1}{2},j}$ correspond au premier moment (normalisé par le premier polynôme de Legendre) de la composante tangentielle du champ électrique sur l'arête ayant pour milieu $(i+\frac{1}{2},j)$, i.e. de la première composante du champ électrique, l'arête étant horizontale ; et $e_{i,j+\frac{1}{2}}$ correspond au premier moment de la composante tangentielle du champ électrique sur l'arête ayant pour milieu $(i,j+\frac{1}{2})$, i.e. de la deuxième composante du champ électrique, l'arête étant verticale. De sorte que $\frac{e_{i+\frac{1}{2},j}}{\sqrt{\Delta x}}$ et $(E_x)_{i,j+\frac{1}{2}}$ ne sont pas à proprement parler égaux mais représentent deux approximations conformes de la même quantité, à savoir la valeur moyenne de la première composante du champ électrique sur l'arête ayant pour milieu $(i+\frac{1}{2},j)$; et pareillement $\frac{e_{i,j+\frac{1}{2}}}{\sqrt{\Delta y}}$ et $(E_y)_{i,j+\frac{1}{2}}$ représentent deux approximations conformes de la deuxième composante du champ électrique sur l'arête ayant pour milieu $(i,j+\frac{1}{2})$.

Chapitre 5

Discrétisations en temps

Étant donné que l'ordre global d'un schéma dépend non seulement de l'ordre de la semi-discrétisation en espace mais aussi de l'ordre de la semi-discrétisation en temps, développer une discrétisation d'ordre élevé en espace n'a qu'un intérêt très limité si l'on n'en fait pas de même pour la discrétisation en temps. En effet, s'il est possible (et c'est un phénomène que nous avons vu apparaître à l'utilisation) que l'ordre numérique de convergence d'un schéma soit l'ordre de sa discrétisation en espace alors que l'ordre de sa discrétisation en temps est moins élevée, cela s'explique par le fait que la restriction sur le rapport entre le pas de temps et le pas d'espace est parfois si faible que l'asymptotique de convergence en temps n'est jamais atteint, et que donc l'erreur reste dominée par l'erreur de la discrétisation en espace. Cette super-convergence numérique reste toutefois un comportement marginal des schémas et l'on ne peut affirmer en toute généralité qu'un schéma est d'un certain ordre que si sa discrétisation en espace et en temps sont de ce même ordre.
Toujours dans l'optique de développer des schémas d'ordre arbitrairement élevé nous avons développé nos propres discrétisations en temps. Celles-ci sont explicites et basées sur une procédure connue sous le nom de Cauchy-Kowalewski [45], que l'on a dû stabiliser. Ces discrétisations en temps ont, comme nous allons le voir, comme principal défaut de faire apparaître les dérivées en temps successives de la force imposée, à calculer à chaque pas de temps. Celles-ci seront approchées par différences divisées (à un ordre assez élevé de manière à ne pas faire descendre l'ordre de la discrétisation en temps), même si cette force est connue de manière analytique. Remarquons toutefois que, la force imposée n'étant en général pas connue analytiquement, mais uniquement en chaque pas de temps, il n'y a pas une différence fondamentale entre la détermination de dérivées en temps successives de la force imposée par différences divisées et la détermination de la force imposée en des pas de temps intermédiaires par interpolation, qu'il faut réaliser pour les discrétisations classiques du type Runge-Kutta par exemple.
Nous avons par ailleurs cherché dans la littérature, parmi les discrétisations en temps d'ordre élevé, lesquelles nous paraissaient adaptées et en avons retenu deux : les discrétisations en temps symplectiques (voir [13]) et les discrétisations en temps de type Runge-Kutta diagonalement implicites (voir [43], [44] ou [9]).
Deux aspects des discrétisations en temps symplectiques nous ont paru particulièrement attrayants. Le premier est que ce type de schéma, développé pour l'intégration numérique de systèmes hamiltoniens, permet de conserver de manière exacte l'énergie totale d'un tel

système. Cela implique que l'utilisation de ce type de schéma nous permettra une propagation d'onde non dissipative (c'est à dire sans perte d'amplitude). Notons que l'utilisation de schémas symplectiques a déjà été abordée dans le cadre de la simulation de propagation d'ondes électromagnétiques par discrétisation en temps par une méthode de type Galerkin discontinus par S. Piperno par exemple dans [59]. Le deuxième est que la détermination et l'implémentation de ces discrétisations en temps s'est révélée être très aisée via l'algorithme proposé par R. Rieben [63] pour les quatres premiers ordres, puis par composition de ces discrétisations (par elles-mêmes) pour atteindre tous les ordres pairs plus élevés [73].
Notre intérêt s'est aussi porté sur les discrétisations de type Runge-Kutta diagonalement implicites. Étant implicites, elles nous permettent d'utiliser nos schémas sans restriction théorique sur le pas de temps. De plus leur forme très particulière nous permet une résolution bien plus efficace que les discrétisations de type Runge-Kutta implicites ne le permettent en général.
Il y a toutefois pour ces deux types de discrétisations des inconvénients majeurs (c'est en particulier ce qui nous a poussé à développer nos propres discrétisations en temps) : pour les discrétisations en temps symplectiques, H. Yoshida nous montre dans [73] qu'il est possible de construire, en composant par lui-même un schéma d'ordre $(2n)$, un schéma d'ordre $(2n+2)$. Or ceci implique une croissance exponentielle du nombre de pas de temps intermédiaires. Pour les discrétisations en temps diagonalement implicites, l'inconvénient est que leur détermination, comme toute discrétisation en temps du type Runge-Kutta d'ordre élevé, est d'une extrême complexité.

5.1 Discrétisation explicite

5.1.1 Discrétisation d'ordre arbitrairement élevé : procédure Cauchy-Kowalewski

La construction des discrétisations en temps que nous présentons ici s'inspire des travaux de M. Dumbser et C.-D. Munz [35] [34] et se fonde sur une procédure connue sous le nom de procédure Cauchy-Kowalewski [45] qui consiste à remplacer les dérivées en temps apparaissant dans le développement de Taylor de la solution entre deux pas de temps successifs par les dérivées en espace via la définition même du système différentiel que l'on cherche à résoudre.
Si cette procédure génère des discrétisations en temps adaptées dans le cadre de schémas à discrétisation en espace du type Galerkin discontinus à flux décentré amont, nous allons voir que celles-ci ne sont stables que pour certains ordres, et proposons une méthode de stabilisation pour celles qui ne le sont pas.

Sachant que les problèmes que nous considérerons pourront se réécrire de cette manière, nous ne considérerons ici que les systèmes différentiels ordinaires du premier ordre de la forme suivante :

$$\frac{dU}{dt} = AU, \tag{5.1}$$

où U est le vecteur à composantes dépendantes du temps inconnues et A une matrice à

coefficients constants (sans plus de précision pour l'instant).
Introduisons une discrétisation de l'axe temporel par $t_n = n\Delta_t$, où le pas de temps Δ_t est un réel positif fixé. Supposons que l'on connaisse la solution de ce système linéaire à un temps t_n donné. Nous proposons de déterminer la solution au temps t_{n+1} en écrivant un développement de Taylor entre les temps t_n et t_{n+1} :

$$U(t_{n+1}) = U(t_n) + \Delta_t \frac{dU(t_n)}{dt} + \frac{{\Delta_t}^2}{2!}\frac{d^2U(t_n)}{dt^2} + \cdots + \frac{{\Delta_t}^p}{p!}\frac{d^pU(t_n)}{dt^p} + O({\Delta_t}^{p+1}).$$

L'équation (5.1) impliquant naturellement que $\dfrac{d^iU}{dt^i} = A^iU$, $\forall i = 1, \ldots, p$; nous obtenons

$$U(t_{n+1}) = U(t_n) + \Delta_t AU(t_n) + \frac{{\Delta_t}^2}{2!}A^2U(t_n) + \cdots + \frac{{\Delta_t}^p}{p!}A^pU(t_n) + O({\Delta_t}^{p+1}).$$

Suivant cette approche nous sommes donc amenés à considérer les approximations successives U_n de $U(t_n)$ données par le schéma :

$$U_{n+1} = (I + \Delta_t A + \frac{{\Delta_t}^2}{2!}A^2 + \cdots + \frac{{\Delta_t}^p}{p!}A^p)U_n. \tag{5.2}$$

Nous introduisons alors la matrice

$$\mathbb{A} = I + \Delta_t A + \frac{{\Delta_t}^2}{2!}A^2 + \cdots + \frac{{\Delta_t}^p}{p!}A^p$$

appelée matrice d'amplification. La méthode numérique décrite par

$$U_{n+1} = \mathbb{A}U_n$$

sera dite stable si

$$\|\mathbb{A}\| \leq 1.$$

On peut montrer (voir [16]) que pour toute matrice $\mathbb{A}$,

$$\rho(\mathbb{A}) \leq \|\mathbb{A}\|,$$

où $\rho(\mathbb{A})$ désigne le rayon spectral de $\mathbb{A}$, c'est-à-dire la plus grande des valeurs propres de $\mathbb{A}$ en module, et que l'égalité de ces quantités se produit en particulier lorsque la matrice $\mathbb{A}$ est symétrique ou anti-symétrique (ou semblable à une telle matrice) et la norme considérée est la norme subordonnée à la norme vectorielle euclidienne.
Supposons alors que la matrice A soit elle-même symétrique (resp. anti-symétrique), A est alors diagonalisable et toutes ses valeurs propres sont réelles (resp. imaginaires pures). On se donne alors une valeur propre λ de A. Il est aisé de se rendre compte que $1+\Delta_t\lambda+\cdots+\frac{{\Delta_t}^p}{p!}\lambda$ est une valeur propre de $\mathbb{A}$ (et que toutes les valeurs propres de $\mathbb{A}$ sont décrites de manière similaire en fonction des valeurs propres de A). Pour étudier la stabilité de la discrétisation en temps il suffit donc de trouver la zone du plan complexe pour laquelle la fonction polynomiale d'une variable complexe $\mu = \Delta_t\lambda$ donnée par $R(\mu) = 1 + \mu + \cdots + \frac{\mu^p}{p!}$, est de module inférieur à 1.
Nous traçons alors à l'aide du logiciel de calcul formel MAPLE© ces zones de stabilité (figure 5.1).

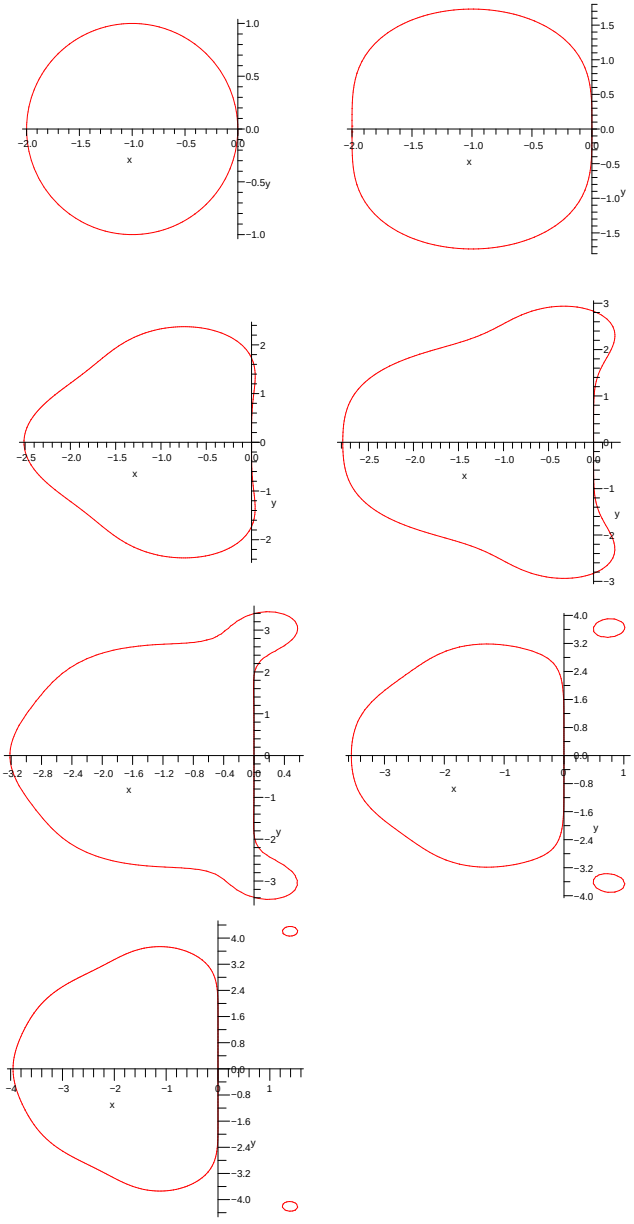

FIG. 5.1 – zones de stabilité pour les discrétisations en temps d'ordre 1 à 7.

5.1.2 Application à l'équation des ondes

Considérons la forme semi-discrétisée de l'équation des ondes

$$M\frac{d^2}{dt^2}U + KU = 0.$$

Il nous faut dans un premier temps réécrire ce système sous la forme d'un système différentiel du premier ordre

$$\left\{ \begin{array}{rl} M\frac{d}{dt}V + KU & = 0 \\ \frac{d}{dt}U - V & = 0 \end{array} \right. \tag{5.3}$$

ou encore sous forme matricielle

$$\frac{d}{dt}\begin{pmatrix} V \\ U \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & -M^{-1}K \\ I & 0 \end{pmatrix}}_{A}\begin{pmatrix} V \\ U \end{pmatrix}.$$

Proposition 5.1.1. *Si λ désigne une valeur propre non nulle de A associée au vecteur propre $\begin{pmatrix} V \\ U \end{pmatrix}$, alors $-\lambda^2$ est valeur propre de $M^{-1}K$ associée au vecteur propre U, et λ est alors imaginaire pure. Réciproquement si μ désigne une valeur propre non nulle de $M^{-1}K$ associée au vecteur propre U, alors μ est réel, strictement positif et $\lambda = \pm i\sqrt{\mu}$ est valeur propre de A associée au vecteur propre $\begin{pmatrix} \pm i\sqrt{\mu}U \\ U \end{pmatrix}$.*

Démonstration. Soit $\lambda \neq 0$ une valeur propre de A associée au vecteur propre $\begin{pmatrix} V \\ U \end{pmatrix}$. Alors

$$-M^{-1}KU = \lambda V, \tag{5.4}$$
$$V = \lambda U. \tag{5.5}$$

λ étant différent de 0, il vient de l'équation (5.5) que $U \neq 0$ et $V \neq 0$ ($\begin{pmatrix} V \\ U \end{pmatrix}$étant un vecteur propre, il est nécessairement non nul). En combinant (5.4) et (5.5) on a que $M^{-1}KU = -\lambda^2 U$ de sorte que $-\lambda^2$ est valeur propre de $M^{-1}K$ associée à U. De cette dernière égalité il vient naturellement

$$KU = -\lambda^2 MU,$$

puis

$${}^t\overline{U}KU = -\lambda^2\, {}^t\overline{U}MU.$$

La matrice M étant symétrique et définie positive, ${}^t\overline{U}MU$ est non seulement réel, mais de plus ${}^t\overline{U}MU > 0$. La matrice symétrique K n'est quant-à-elle que semi-définie positive : il suffit de remarquer que ${}^t\overline{V}KV$ n'est autre que la norme du gradient de la fonction de composantes V dans l'espace de discrétisation, de sorte que ${}^t\overline{V}KV$ est un réel positif, et

que pour toute fonction constante non nulle de composante $V \neq 0$ nous avons ${}^t\overline{V}KV = 0$. Il vient alors

$$\underbrace{{}^t\overline{U}KU}_{\geq 0} = -\lambda^2 \underbrace{{}^t\overline{U}MU}_{>0}.$$

Ainsi $-\lambda^2$ est en particulier réel, $-\lambda^2 > 0$ et λ est naturellement imaginaire pur.
Considérons maintenant μ une valeur propre non nulle de $M^{-1}K$ associée au vecteur propre U. Alors $M^{-1}KU = \mu U$. La même manipulation nous permet d'affirmer que μ est réel et strictement positif. Posons $V = \pm i\sqrt{\mu}U$ de sorte que

$$U = \pm\frac{1}{i\sqrt{\mu}}V = \mp\frac{i}{\sqrt{\mu}}V.$$

Alors

$$M^{-1}KU = \mu U = \mp i\frac{\mu}{\sqrt{\mu}}V = \mp i\sqrt{\mu}V$$

et

$$V = \pm i\sqrt{\mu}U$$

ou encore

$$A\begin{pmatrix} V \\ U \end{pmatrix} = \pm i\sqrt{\mu}\begin{pmatrix} V \\ U \end{pmatrix}.$$

Finalement $\pm i\sqrt{\mu}$ est valeur propre de A associée au vecteur propre $\begin{pmatrix} V \\ U \end{pmatrix} = \begin{pmatrix} \pm i\sqrt{\mu}U \\ U \end{pmatrix}$.
□

Remarque 5.1.2. *La proposition précédente nous dit en particulier que* $\rho(A) = \sqrt{\rho(M^{-1}K)}$.

Maintenant que l'on sait que les valeurs propres de A sont imaginaires pures il convient de s'intéresser à l'intersection de chaque zone de stabilité par discrétisation en temps avec l'axe des nombres complexes imaginaires purs. Les graphiques des zones de stabilité associés aux discrétisations en temps d'ordre 1 et 2 nous disent que l'intersection de la zone de stabilité avec l'axe des nombres complexes imaginaires purs se résume au singleton $\{0\}$, ce qui signifie que ces schémas sont inconditionnellement instables sauf pour $\Delta_t = 0$, et n'ont donc plus aucun intérêt.
Sur le graphique associé à la discrétisation en temps d'ordre 3 cette intersection que l'on détermine à l'aide de MAPLE© est l'intervalle $[-i\sqrt{3}, i\sqrt{3}]$, pour l'ordre 4 nous obtenons l'intervalle $[-2i\sqrt{2}, 2i\sqrt{2}]$ et pour l'ordre 7 l'intervalle $[-1.764421325i, 1.764421325i]$. Cela signifie que ces schémas seront stables dès que $\Delta_t\rho(A) \leq \sqrt{3}$ pour le schéma d'ordre 3, $\Delta_t\rho(A) \leq 2\sqrt{2}$ pour le schéma d'ordre 4 et $\Delta_t\rho(A) \leq 1.764421325$ pour le schéma d'ordre 7.
Pour les discrétisations d'ordre 5 et 6 il convient de regarder de plus près le comportement de la zone de stabilité autour de l'axe imaginaire (figure 5.2).
S'il devient clair que la discrétisation d'ordre 6 est inconditionnellement instable dès que $\Delta_t > 0$, nous déterminons l'intersection de la zone de stabilité associée à la discrétisation d'ordre 5 avec l'axe imaginaire qui vaut $\left[-\frac{i}{2}\sqrt{30+2\sqrt{65}}, -\frac{i}{2}\sqrt{30-2\sqrt{65}}\right] \cup$

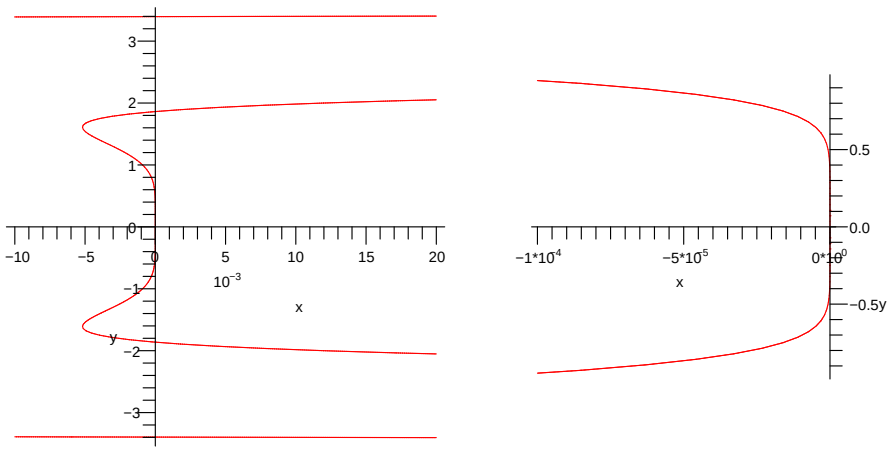

FIG. 5.2 – Zoom aux abords de l'axe imaginaire des zones de stabilité pour les discrétisations en temps d'ordre 5 et 6.

$\{0\} \cup \left[\frac{i}{2}\sqrt{30-2\sqrt{65}}, \frac{i}{2}\sqrt{30+2\sqrt{65}}\right]$. Cette discrétisation est donc marginalement stable et n'est pas utilisable dans la pratique en toute généralité (elle l'est dans le cas où l'amplitude des modules des valeurs propres n'est pas trop importante pour pouvoir déterminer des bornes inférieures et supérieures sur Δ_t de manière à ce que $R(\Delta_t\lambda)$ soit un élément de l'ensemble décrit ci-dessus quelle que soit la valeur propre λ de A).

D'un point de vue pratique il serait très peu judicieux de résoudre brutalement le système (5.2) en calculant la matrice $\mathbb{A}$, celle-ci étant donnée en fonction des matrices M et K qui sont symétriques et définies positives, ce qui nous permet des stockages (et une inversion pour le cas de M) optimisés de ces matrices. Chacun pourra se convaincre que l'on résout ce système de manière tout à fait équivalente en écrivant :

$$\begin{aligned} U^{n+1} &= U^n + \Delta t\dot{U}^n + \tfrac{\Delta t^2}{2!}\ddot{U}^n + \ldots \\ V^{n+1} &= V^n + \Delta t\dot{V}^n + \tfrac{\Delta t^2}{2!}\ddot{V}^n + \ldots \end{aligned}$$

puis en substituant dans le développement de U^{n+1} les dérivées successives de U^n par celles de V^n grâce à la deuxième relation de ce système, elles-mêmes évaluées à l'aide de la première relation du système :

$$\begin{aligned} U^{n+1} &= U^n + \Delta t V^n + \tfrac{\Delta t^2}{2!}\dot{V}^n + \ldots \\ V^{n+1} &= V^n + \Delta t \underbrace{\dot{V}^n}_{-M^{-1}KU^n} + \tfrac{\Delta t^2}{2!}\underbrace{\ddot{V}^n}_{-M^{-1}KV^n} + \ldots \end{aligned}$$

Cette procédure se généralise alors très intuitivement pour l'implémentation de schémas en temps résolvant, par exemple, l'équation des ondes avec second membre dépendant du

temps, soumise à des conditions aux limites absorbantes :

$$\begin{cases} \partial_t^2 u - \triangle u = f(x,t) & (x,t) \in \Omega \times \mathbb{R}^+ \\ u(x,0) = u_0(x) & x \in \Omega \\ \partial_t u(x,0) = u_1(x) & x \in \Omega \\ \partial_t u(x,t) + \partial_{\overrightarrow{n}} u(x,t) = 0 & (x,t) \in \Gamma = \partial\Omega \times \mathbb{R}^+ \end{cases}$$

où $\overrightarrow{n}$ désigne le vecteur normal sortant au domaine Ω.
Une fois la semi-discrétisation effectuée le problème devient :

$$\begin{cases} M\dot{V} + M_\Gamma V + KU &= MF \\ \dot{U} - V &= 0 \end{cases}$$

où M_Γ désigne la matrice de composantes

$$(M_\Gamma)_{ij} = \int_\Gamma \psi_i(x,y)\psi_j(x,y)dxdy,$$

et F le vecteur de composantes $f_i = f(x_i,t)$.
Nous effectuons alors comme précédemment la semi-discrétisation en temps pour obtenir le problème suivant :

$$\begin{array}{l} U^{n+1} = U^n + \Delta t V^n + \frac{\Delta t^2}{2!}\dot{V}^n + ... \\ V^{n+1} = V^n + \Delta t \underbrace{\dot{V}^n}_{F^n - M^{-1}(M_\Gamma V^n + KU^n)} + \frac{\Delta t^2}{2!} \underbrace{\ddot{V}^n}_{F^n - M^{-1}(M_\Gamma \dot{V}^n + KV^n)} + \dots \end{array}$$

ou finalement

$$\begin{array}{rcl} U^{n+1} = U^n & + & \Delta t V^n + \frac{\Delta t^2}{2!}\dot{V}^n + ... \\ V^{n+1} = V^n & + & \Delta t(F^n - M^{-1}(M_\Gamma V^n + KU^n)) \\ & + & \frac{\Delta t^2}{2!}(\dot{F}^n - M^{-1}(M_\Gamma(F^n - M^{-1}(M_\Gamma V^n + KU^n)) + KV^n)) \\ & + & \dots \end{array}$$

Remarque 5.1.3. *Cette dernière expression est très révélatrice du coût des méthodes d'ordre élevé : celles-ci font non seulement intervenir un nombre grandissant d'inversion de matrices de masse à chaque pas de temps, mais aussi, pour cette discrétisation en temps, de dérivées successives du second membre.*

5.1.3 Application aux équations de Maxwell

Nous reconsidérons à présent le système semi-discrétisé des équations de Maxwell

$$\begin{cases} M_w \frac{d}{dt} E - KB &= 0 \\ M_v \frac{d}{dt} B + {}^t KE &= 0 \end{cases}, \tag{5.6}$$

que l'on réécrit sous forme matricielle

$$\frac{d}{dt}\begin{pmatrix} E \\ B \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & M_w^{-1}K \\ -M_v^{-1}\,{}^t K & 0 \end{pmatrix}}_{A} \begin{pmatrix} E \\ B \end{pmatrix}.$$

Proposition 5.1.4. *Si λ désigne une valeur propre non nulle de A associée au vecteur propre $\begin{pmatrix} E \\ B \end{pmatrix}$, alors $-\lambda^2$ est valeur propre de $M_v^{-1}\,{}^tKM_w^{-1}K$ et $M_w^{-1}KM_v^{-1}\,{}^tK$ associée respectivement aux vecteurs propres B et E, et λ est alors imaginaire pure. Réciproquement si μ désigne une valeur propre non nulle de $M_v^{-1}\,{}^tKM_w^{-1}K$, alors μ est aussi valeur propre de $M_w^{-1}KM_v^{-1}\,{}^tK$ (et réciproquement), μ est réel, strictement positif et $\lambda = \pm i\sqrt{\mu}$ est valeur propre de A.*

Démonstration. Soit $\lambda \neq 0$ une valeur propre de A associée au vecteur propre $\begin{pmatrix} E \\ B \end{pmatrix}$. Alors

$$M_w^{-1}KB = \lambda E, \tag{5.7}$$
$$-M_v^{-1}\,{}^tKE = \lambda B. \tag{5.8}$$

Notons que nécessairement E et B sont tous les deux non nuls. En effet en supposant $E = 0$ alors $\lambda B = 0$, ce qui n'est possible que si $B = 0$, λ étant non nul. De sorte que $\begin{pmatrix} E \\ B \end{pmatrix} = 0$ ne puisse plus être un vecteur propre de A. Ainsi $E \neq 0$ et en conséquence immédiate de (5.7) $B \neq 0$.
En multipliant (5.7) par $M_v^{-1}\,{}^tK$ et en utilisant (5.8) nous obtenons

$$M_v^{-1}\,{}^tKM_w^{-1}KB = \lambda M_v^{-1}\,{}^tKE = -\lambda^2 B,$$

de sorte que $-\lambda^2$ est valeur propre de $M_v^{-1}\,{}^tKM_w^{-1}K$ associée au vecteur propre B.
De manière similaire nous montrons que

$$M_w^{-1}KM_v^{-1}\,{}^tKE = -\lambda M_w^{-1}KB = -\lambda^2 E,$$

c'est-à-dire que $-\lambda^2$ est valeur propre de $M_w^{-1}KM_v^{-1}\,{}^tK$ associée au vecteur propre E.
Considérons alors une valeur propre μ de $M_v^{-1}\,{}^tKM_w^{-1}K$ associée au vecteur propre B. Alors

$${}^tKM_w^{-1}KB = \mu M_v B,$$

et en multipliant les deux membres de l'équation par ${}^t\overline{B}$

$${}^t(\overline{KB})M_w^{-1}KB = \mu\,{}^t\overline{B}M_vB.$$

Les matrices M_w^{-1} et M_v étant symétriques définies positives, on a que $\mu \geq 0$, de sorte qu'en particulier $-\lambda^2 > 0$ et donc que λ est imaginaire pur. Ce résultat se retrouve bien évidemment en considérant les valeurs propres de $M_w^{-1}KM_v^{-1}\,{}^tK$.
Réciproquement considérons $\mu \neq 0$ une valeur propre de $M_v^{-1}\,{}^tKM_w^{-1}K$ associée au vecteur propre B. On a alors

$$M_v^{-1}\,{}^tKM_w^{-1}KB = \mu B,$$

d'où

$${}^t(\overline{KB})M_w^{-1}KB = \mu\,{}^t\overline{B}M_vB,$$

et μ est alors un réel strictement positif.
Posant

$$E = \pm\frac{1}{i\sqrt{\mu}} M_w^{-1} K B = \mp\frac{i}{\sqrt{\mu}} M_w^{-1} K B,$$

E est non nul (μ et B étant non nuls), et cette équation ce réécrit

$$\pm i\sqrt{\mu} M_v^{-1\ t} K E = \mu B,$$

ou encore

$$M_v^{-1\ t} K E = \mp i\sqrt{\mu} B.$$

Ainsi $M_w^{-1} K B = \pm i\sqrt{\mu} E$, et $-M_v^{-1\ t} K E = \pm i\sqrt{\mu} B$ de sorte que $\pm i\sqrt{\mu}$ est valeur propre de A. De cette dernière équation on déduit aussi que

$$M_v^{-1\ t} K E = (\pm)^2 \mu K^{-1} M_w E$$

c'est-à-dire que

$$M_w^{-1} K M_v^{-1\ t} K E = \mu E.$$

Finalement μ est aussi une valeur propre de $M_w^{-1} K M_v^{-1\ t} K$. □

Remarque 5.1.5. *La proposition précédente nous dit que les valeurs propres de A sont à nouveau imaginaires pures. Les zones de stabilité décrites dans le cadre de l'équation des ondes restent donc valables en considérant cette fois le rayon spectral de la matrice $A = \begin{pmatrix} 0 & M_w^{-1} K \\ -M_v^{-1\ t} K & 0 \end{pmatrix}$ qui n'est autre que la racine du rayon spectral de $M_w^{-1} K M_v^{-1\ t} K$ ou de $M_v^{-1\ t} K M_w^{-1} K$.*

5.1.4 Stabilisation de ces discrétisations en temps

Nous venons de voir que les discrétisations en temps d'ordre 1, 2, 5 et 6 sont instables quel que soit $\Delta_t > 0$, que ce soit dans le cadre de la résolution de l'équation des ondes ou de la résolution des équations de Maxwell. Nous cherchons alors à stabiliser ces méthodes de la manière suivante : en rajoutant un terme d'ordre $p+1$ dans le développement de Taylor définissant la discrétisation en temps, on ne modifie pas l'ordre de cette discrétisation, mais peut-être est-il possible de modifier la zone de stabilité de manière à y inclure un voisinage de 0 de l'axe imaginaire.

Plus précisément pour la discrétisation d'ordre 1 nous proposons de chercher ξ de sorte que l'intersection de la zone de stabilité avec l'axe imaginaire ne se résume plus à 0 mais contienne un intervalle le plus grand possible lui-même contenant 0, où la zone de stabilité est définie comme étant l'ensemble des μ tel que

$$R(\mu) = 1 + \mu + \xi\frac{\mu^2}{2!}$$

est de module inférieur à 1. Il s'avère qu'il est effectivement possible de stabiliser la discrétisation en temps d'ordre 1 par cette méthode. Une dichotomie nous donne que le ξ

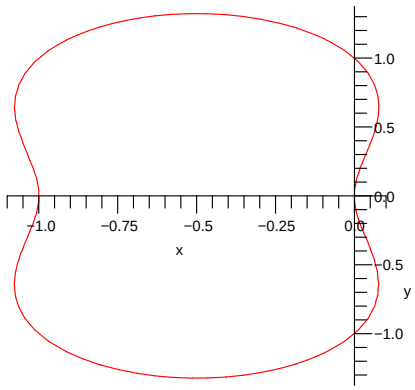

FIG. 5.3 – Stabilisation de la discrétisation en temps d'ordre 1.

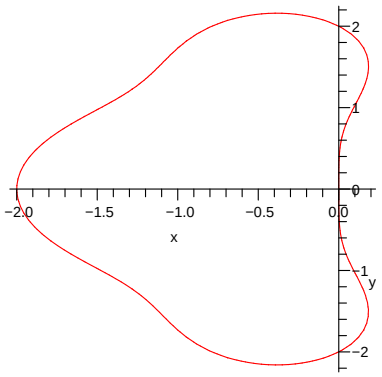

FIG. 5.4 – Stabilisation de la discrétisation en temps d'ordre 2.

optimal est atteint en $\xi = 2$. La figure 5.3 représente la zone de stabilité correspondante à cette valeur de ξ et l'on détermine la zone de stabilité sur l'axe imaginaire qui vaut $[-i, i]$. La condition de stabilité est alors donné par $\Delta_t \rho(A) \leq 1$.

Pour la discrétisation en temps d'ordre 2 le même procédé appliqué à la fonction

$$R(\mu) = 1 + \mu + \frac{\mu^2}{2!} + \xi \frac{\mu^3}{3!}$$

nous donne un ξ optimal valant $\xi = \frac{3}{2}$, la zone de stabilité est alors donnée par $[-2i, 2i]$ (figure 5.4) et la contrainte de stabilité devient $\Delta_t \rho(A) \leq 2$.

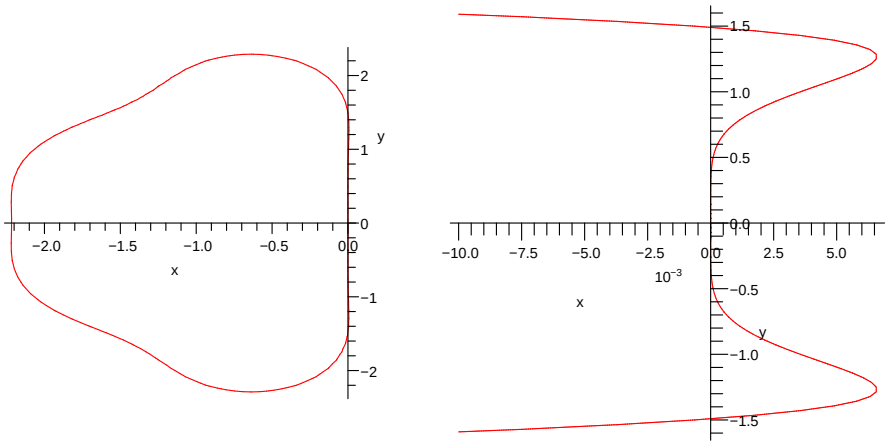

FIG. 5.5 – Stabilisation de la discrétisation en temps d'ordre 5 : zoom au voisinage de l'axe imaginaire.

Pour la discrétisation en temps d'ordre 5 nous considérons la fonction

$$R(\mu) = 1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \frac{\mu^4}{4!} + \frac{\mu^5}{5!} + \xi\frac{\mu^6}{6!}.$$

Nous déterminons alors une approximation par dichotomie du ξ optimal $\simeq 6.15746160$, la zone de stabilité qui en résulte est alors donnée par $[-1.491320186i, 1.491320186i]$ (figure 5.5) et la contrainte de stabilité devient $\Delta_t\rho(A) \leq 1.491320186$.

Pour finir nous déterminons la stabilisation de la discrétisation en temps d'ordre 6 en considérant

$$R(\mu) = 1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \frac{\mu^4}{4!} + \frac{\mu^5}{5!} + \frac{\mu^6}{6!} + \xi\frac{\mu^7}{7!}.$$

Le ξ optimal est alors atteint aux environs de $\xi \simeq 2.505288240$, la zone de stabilité est donnée par $[-2.751711543i, 2.751711543i]$ (figure 5.6) et donc la contrainte de stabilité devient $\Delta_t\rho(A) \leq 2.751711543$.

5.1.5 Discrétisation symplectique

Nous allons décrire à présent les discrétisations en temps symplectiques, par exemple sur le système d'équation différentielle ordinaire que l'on obtient après semi-discrétisation en espace de la formulation variationnelle de l'équation des ondes :

$$\left\{ \begin{array}{rl} M\dot{V} + KU & = 0 \\ \dot{U} - V & = 0 \end{array} \right. ,$$

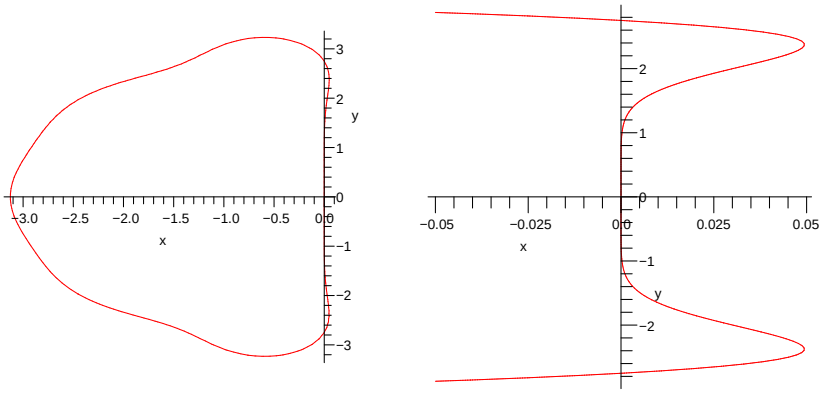

FIG. 5.6 – Stabilisation de la discrétisation en temps d'ordre 6 : zoom au voisinage de l'axe imaginaire.

ou sous forme matricielle

$$\begin{aligned}\frac{d}{dt}\begin{pmatrix} V \\ U \end{pmatrix} &= \underbrace{\begin{pmatrix} 0 & -M^{-1}K \\ I & 0 \end{pmatrix}}_{A}\begin{pmatrix} V \\ U \end{pmatrix} \\ &= \underbrace{\begin{pmatrix} 0 & 0 \\ I & 0 \end{pmatrix}}_{\mathbb{T}}\begin{pmatrix} V \\ 0 \end{pmatrix} + \underbrace{\begin{pmatrix} 0 & -M^{-1}K \\ 0 & 0 \end{pmatrix}}_{\mathbb{V}}\begin{pmatrix} 0 \\ U \end{pmatrix}.\end{aligned}$$

Après avoir introduit une discrétisation de l'axe temporel par $t^n = n\Delta t$, où Δt désigne le pas de temps, la solution du problème au temps t^{n+1} en fonction de la solution au temps t^n peut être donnée de manière formelle par :

$$\begin{aligned}\begin{pmatrix} V^{n+1} \\ U^{n+1} \end{pmatrix} &= \exp(\Delta t A)\begin{pmatrix} V^n \\ U^n \end{pmatrix} \\ &= \exp(\Delta t(\mathbb{T}+\mathbb{V}))\begin{pmatrix} V^n \\ U^n \end{pmatrix}.\end{aligned}$$

Supposons à présent déterminé un ensemble de réels $\{(a_i, b_i), i = 1 \dots p\}$, tel que

$$\exp(\Delta t(\mathbb{T}+\mathbb{V})) = \exp(a_1\Delta t\mathbb{T})\exp(b_1\Delta t\mathbb{V}) \times \cdots \times \exp(a_p\Delta t\mathbb{T})\exp(b_p\Delta t\mathbb{V}) + \mathcal{O}(\Delta t^m)$$

alors par définition

$$\exp(a_1\Delta t\mathbb{T})\exp(b_1\Delta t\mathbb{V}) \times \cdots \times \exp(a_p\Delta t\mathbb{T})\exp(b_p\Delta t\mathbb{V})\begin{pmatrix} V^n \\ U^n \end{pmatrix} \tag{5.9}$$

est une approximation d'ordre m de $\begin{pmatrix} V^{n+1} \\ U^{n+1} \end{pmatrix}$. Ainsi il est possible de construire des discrétisations en temps symplectiques en déterminant les réels $\{(a_i, b_i), i = 1 \dots p\}$ par identification des termes du développement du produit

$$\exp(a_1\Delta t\mathbb{T})\exp(b_1\Delta t\mathbb{V}) \times \cdots \times \exp(a_p\Delta t\mathbb{T})\exp(b_p\Delta t\mathbb{V})$$

$p = 1$	
$a_1 = 1$	$b_1 = 1$
$p = 2$	
$a_1 = \frac{1}{2}$	$b_1 = 0$
$a_2 = \frac{1}{2}$	$b_2 = 1$
$p = 3$	
$a_1 = \frac{2}{3}$	$b_1 = \frac{7}{24}$
$a_2 = -\frac{2}{3}$	$b_2 = \frac{3}{4}$
$a_3 = 1$	$b_3 = -\frac{1}{24}$
$p = 4$	
$a_1 = \frac{2+2^{\frac{1}{3}}+2^{-\frac{1}{3}}}{6}$	$b_1 = 0$
$a_2 = \frac{1-2^{\frac{1}{3}}-2^{-\frac{1}{3}}}{6}$	$b_2 = \frac{1}{2-2^{\frac{1}{3}}}$
$a_3 = \frac{1-2^{\frac{1}{3}}-2^{-\frac{1}{3}}}{6}$	$b_3 = \frac{1}{1-2^{\frac{2}{3}}}$
$a_4 = \frac{2+2^{\frac{1}{3}}+2^{-\frac{1}{3}}}{6}$	$b_4 = \frac{1}{2-2^{\frac{1}{3}}}$

TAB. 5.1 – Coefficients des discrétisations en temps symplectiques

en puissance de Δt avec les termes du développement de $\exp(\Delta t(\mathbb{T}+\mathbb{V}))$.

Le schéma défini par (5.9) peut être explicité par l'algorithme décrit par R. Rieben dans [63], et que nous adaptons ici au système d'équation obtenu par semi-discrétisation en espace de l'équation des ondes non-homogène :

On initialise les champs de calcul $U_{in} = U^n$ et $V_{in} = V^n$, où U^n et V^n désignent la solution du problème au temps t^n et l'on itère pour i de 1 à p :

$$\begin{array}{lcl} t_i & = & t^n + \sum_{k=1}^{i-1} a_k \Delta t \\ V_{out} & = & V_{in} + b_i \Delta t M^{-1}(F(t_i) - K U_{in}) \\ U_{out} & = & U_{in} + a_i \Delta t V_{out} \\ V_{in} & \leftarrow & V_{out} \\ U_{in} & \leftarrow & U_{out} \end{array}$$

où p, le nombre de pas intermédiaires, est égal à l'ordre de la méthode pour les quatres premiers ordres, et les coefficients $\{(a_i, b_i), i = 1 \dots p\}$ sont donnés dans le tableau 5.1 pour ces mêmes ordres.

Pour les méthodes d'ordre plus élevé, H. Yoshida propose dans [73] de composer les méthodes d'ordre pair $(2m)$ par elle-même pour obtenir une méthode d'ordre $(2m+2)$ de la manière suivante : en notant S_{2m} l'opérateur de passage de la solution du temps t^n au temps t^{n+1}, c'est-à-dire

$$\begin{pmatrix} V^{n+1} \\ U^{n+1} \end{pmatrix} = S_{2m}(\Delta t) \begin{pmatrix} V^n \\ U^n \end{pmatrix},$$

une méthode d'ordre $(2m+2)$ peut être donnée par la composition

$$S_{2m+2}(\Delta t) = S_{2m}(\alpha\Delta t)S_{2m}(\beta\Delta t)S_{2m}(\alpha\Delta t),$$

avec $\alpha = \dfrac{1}{2-2^{\frac{1}{2m+1}}}$ et $\beta = -\dfrac{2^{\frac{1}{2m+1}}}{2-2^{\frac{1}{2m+1}}}$ (c'est un résultat qui se retrouve à nouveau par identification des développements en puissances de Δt des expressions formelles des opérateurs). On pourra remarquer que la discrétisation d'ordre 4 n'est autre que la composée de la discrétisation d'ordre 2 composée par elle-même.
Nous avons tenté de déterminer, sans succès, une discrétisation en temps symplectique d'ordre 5 par la résolution directe décrite précédemment qui nous a pourtant permis de retrouver les discrétisations d'ordre 1 à 4.

5.2 Discrétisation implicite

Pour finir nous allons décrire l'adaptation des méthodes de Runge-Kutta diagonalement implicites de nouveau sur le système d'équations différentielles ordinaires que l'on obtient après semi-discrétisation en espace de la formulation variationnelle de l'équation des ondes :

$$\left\{ \begin{array}{rl} M\dot{V} + KU & = F \\ \dot{U} - V & = 0 \end{array} \right. .$$

Une méthode de Runge-Kutta diagonalement implicite est la donnée de deux vecteurs $B = (b_1, \ldots, b_s)$ et $T = (t_1, \ldots, t_s)$, et de la matrice carrée $A = (a_{ij})_{1 \leq i,j \leq s}$, triangulaire inférieure dont les termes diagonaux sont non nuls. Après avoir introduit une discrétisation de l'axe temporel par $t^n = n\Delta t$, où Δt désigne le pas de temps, la solution au temps t^{n+1} est alors définie par :

$$\begin{pmatrix} V^{n+1} \\ U^{n+1} \end{pmatrix} = \begin{pmatrix} V^n \\ U^n \end{pmatrix} + \Delta t \sum_{i=1}^{s} b_i \begin{pmatrix} V_i \\ U_i \end{pmatrix},$$

où U_i et V_i sont les solutions du problème suivant :

$$\begin{pmatrix} \begin{pmatrix} V_1 \\ U_1 \end{pmatrix} \\ \vdots \\ \begin{pmatrix} V_s \\ U_s \end{pmatrix} \end{pmatrix} = \begin{pmatrix} N\begin{pmatrix} V^n \\ U^n \end{pmatrix} \\ \vdots \\ N\begin{pmatrix} V^n \\ U^n \end{pmatrix} \end{pmatrix} + \Delta t \begin{pmatrix} a_{11}N & & \\ \vdots & \ddots & \\ a_{s1}N & \cdots & a_{ss}N \end{pmatrix} \begin{pmatrix} \begin{pmatrix} V_1 \\ U_1 \end{pmatrix} \\ \vdots \\ \begin{pmatrix} V_s \\ U_s \end{pmatrix} \end{pmatrix} + \begin{pmatrix} \begin{pmatrix} F(t_1) \\ 0 \end{pmatrix} \\ \vdots \\ \begin{pmatrix} F(t_s) \\ 0 \end{pmatrix} \end{pmatrix},$$

où N désigne la matrice

$$N = \begin{pmatrix} 0 & -M^{-1}K \\ Id & 0 \end{pmatrix}.$$

Remarquons alors que le fait que la matrice A soit triangulaire inférieure (c'est-à-dire le fait de ne considérer que des méthodes de Runge-Kutta diagonalement implicites) nous

ordre 2, $s = 1$

$\frac{1}{2}$	$\frac{1}{2}$
	1

ordre 3, $s = 2$

γ	γ	0
$1-\gamma$	$1-2\gamma$	γ
	$\frac{1}{2}$	$\frac{1}{2}$

, $\gamma = \frac{3+\sqrt{3}}{6}$

ordre 4, $s = 3$

γ	γ	0	0
$\frac{1}{2}$	$\frac{1}{2}-\gamma$	γ	0
$1-\gamma$	2γ	$1-4\gamma$	γ
	δ	$1-2\delta$	δ

, $\gamma = \frac{1}{\sqrt{3}}\cos(\frac{\pi}{18}) + \frac{1}{2}$, $\delta = \frac{1}{6(2\gamma-1)^2}$

ordre 5, $s = 5$

$\frac{6-\sqrt{6}}{10}$	$\frac{6-\sqrt{6}}{10}$	0	0	0	0
$\frac{6+9\sqrt{6}}{35}$	$\frac{-6+5\sqrt{6}}{14}$	$\frac{6-\sqrt{6}}{10}$	0	0	0
1	$\frac{888+607\sqrt{6}}{2850}$	$\frac{126-161\sqrt{6}}{1425}$	$\frac{6-\sqrt{6}}{10}$	0	0
$\frac{4-\sqrt{6}}{10}$	$\frac{3153-3082\sqrt{6}}{14250}$	$\frac{3213+1148\sqrt{6}}{28500}$	$\frac{-267+88\sqrt{6}}{500}$	$\frac{6-\sqrt{6}}{10}$	0
$\frac{4+\sqrt{6}}{10}$	$\frac{-32583+14638\sqrt{6}}{71250}$	$\frac{-17199+364\sqrt{6}}{142500}$	$\frac{1329-544\sqrt{6}}{2500}$	$\frac{-96+131\sqrt{6}}{625}$	$\frac{6-\sqrt{6}}{10}$
	0	0	$\frac{1}{9}$	$\frac{16-\sqrt{6}}{36}$	$\frac{16+\sqrt{6}}{36}$

TAB. 5.2 – Coefficients des discrétisations en temps implicites

permet de résoudre ce problème en s étapes en résolvant à chaque étape un problème dont la forme générique s'écrit :

$$\left\{ \begin{array}{rcl} (M + (a_{ii}\Delta t)^2 K)U_i &=& M(V^n + \Delta t \sum_{j=1}^{i-1} a_{ij} V_j) \\ &+& a_{ii}\Delta t(MF(t_i) - K(U^n + \Delta t \sum_{j=1}^{i-1} a_{ij} U_j)) \\ V_i &=& \frac{1}{a_{ii}\Delta t}(U_i - (V^n + \sum_{j=1}^{i-1} a_{ij}\Delta t V_j)) \end{array} \right.$$

Les paramètres de ces discrétisations en temps d'ordre 2 à 5 (qui sont les seuls que l'on ait trouvé dans la littérature) sont ceux décrits par E. Hairer dans [43] pour les ordres 2 et 3, dans [44] pour l'ordre 4, et déterminés par G. J. Cooper et A. Sayfy dans [30] pour l'ordre 5, et que l'on donne dans le tableau 5.2 sous la forme $\begin{array}{c|c} \mathrm{T} & \mathrm{A} \\ \hline & {}^t\mathrm{B} \end{array}$.

Remarquons que si cette discrétisation est implicite et que donc la stabilité du schéma est automatiquement assurée quel que soit le nombre CFL, ce n'est cette fois plus la matrice de masse qu'il faut inverser mais une combinaison linéaire des matrices de masse et de raideur, de sorte que l'on perd tout l'intérêt de la condensation de la matrice de masse.

Chapitre 6

Efficacité des schémas

Nous allons maintenant tester numériquement l'efficacité des schémas que nous avons développés : c'est-à-dire déterminer les conditions de stabilité, étudier la dissipation et la dispersion de ceux-ci. Il nous faut notamment vérifier si les ordres de convergence sont cohérents et vérifier que l'utilisation du mass-lumping n'entraîne pas une diminution de l'ordre des schémas. Les schémas ont été effectivement implémentés de manière à résoudre l'équation des ondes avec second membre dépendant du temps et conditions de bord périodiques, absorbantes, de Neumann homogènes ou de Dirichlet dépendant du temps. Nous renvoyons le lecteur à l'annexe 7.5.7 pour les détails de l'imposition de ces dernières, les seules à poser des problèmes techniques. Les schémas adaptés à la résolution des équations de Maxwell ont, quant à eux, été implémentés avec des conditions aux limites de type conducteur parfait, périodiques, et de Silver-Müller.

6.1 Schémas adaptés à la résolution de l'équation des ondes

6.1.1 Stabilité et ordre de convergence

Nous allons considérer dans un premier temps les discrétisations explicites d'ordre arbitrairement élevé (stabilisées) que nous avons développées, couplées aux éléments finis de Lagrange standards (avec localisation optimisée des points), condensés et partiellement condensés.
Pour se fixer les idées nous utiliserons dans un premier temps des maillages structurés dont nous donnons un exemple dans la figure 6.1.

Il nous faut tout d'abord déterminer les limites de stabilité sur chacun des schémas : pour cela nous propageons une onde plane à travers le domaine périodique $\Omega = [-5, 5] \times [-2.5, 2.5]$ (voir figure 6.2), et nous déterminons le plus grand nombre CFL (rapport du pas de temps sur le pas d'espace) permettant la propagation de cette onde sur 100 périodes. Plus précisément nous résolvons le système suivant :

$$\left\{ \begin{array}{rcll} \partial_t^2 u - \triangle u & = & 0 & (x,y,t) \in \Omega \times [0,100] \\ u(x,y,0) & = & 0.005 \sin((k\pi x)) & (x,y) \in \Omega \\ \partial_t u(x,y,0) & = & -0.005\, \omega\, \cos((k\pi x)) & (x,y) \in \Omega \end{array} \right. ,$$

avec $k = 2$, où k et ω sont reliés par la relation $w = k\pi$, et dont la solution est explicitement

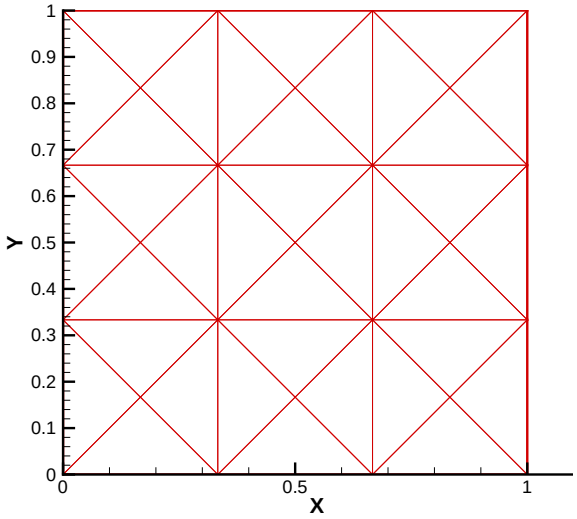

FIG. 6.1 – Maillage structuré de $3 \times 3 \times 4$ soit 36 éléments

Ordre théorique de la discrétisation	2	3	4	5	6	7
Lagrange standard	0.47	0.19	0.19	0.072	0.097	0.055
Lagrange condensé	0.94	0.26	0.24	0.079	0.058	0.0052
Lagrange partiellement condensé	XX	XX	0.19	0.072	0.097	

TAB. 6.1 – Nombres CFL optimaux pour les discrétisations en temps explicites d'ordre arbitrairement élevé.

donnée par

$$u(x, y, t) = 0.005 \sin((k\pi x) - \omega t).$$

La figure 6.3 nous donne deux pick-points (évolution temporelle du signal en un point particulier du domaine), l'un à nombre CFL optimal (l'onde est alors propagée) et l'autre à un nombre CFL trop élevé (l'approximation explose).

Les nombres CFL obtenus sont listés dans le tableau 6.1. Rappelons que la condensation partielle des éléments finis de Lagrange P_k n'a de sens que lorsque le sous-espace $\mathbb{P}_{k-3}$ n'est pas dégénéré, c'est-à-dire à partir des éléments finis P_3. Il n'existe donc pas d'éléments finis partiellement condensés P_1 et P_2 (en tout cas pas dans le sens où nous avons défini cette condensation partielle), d'où la notation XX dans le tableau 6.1. D'autre part, n'ayant à ce jour pas encore réglé le problème de la construction d'une "bonne" base orthogonale pour la condensation partielle des éléments finis (cf remarque 4.4.1), nous ne pouvons utiliser cette condensation partielle au-delà des schémas d'ordre 6.

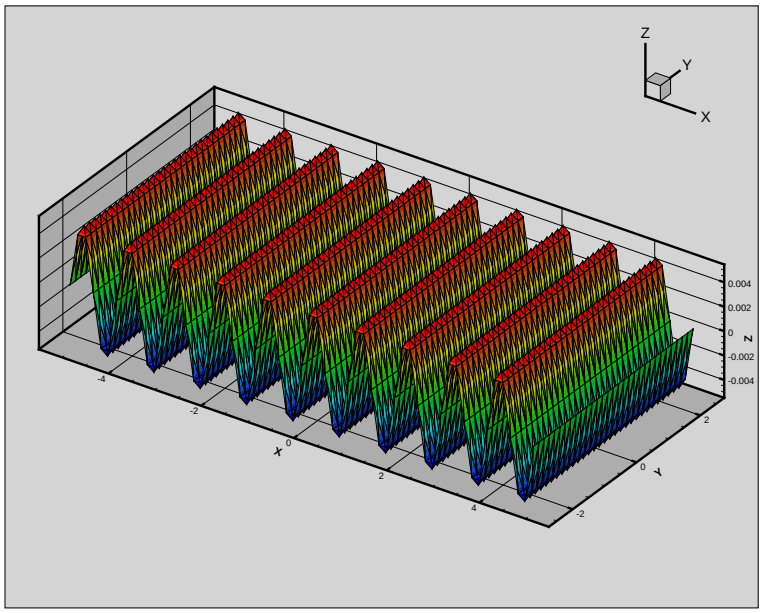

FIG. 6.2 – Propagation d'une onde plane

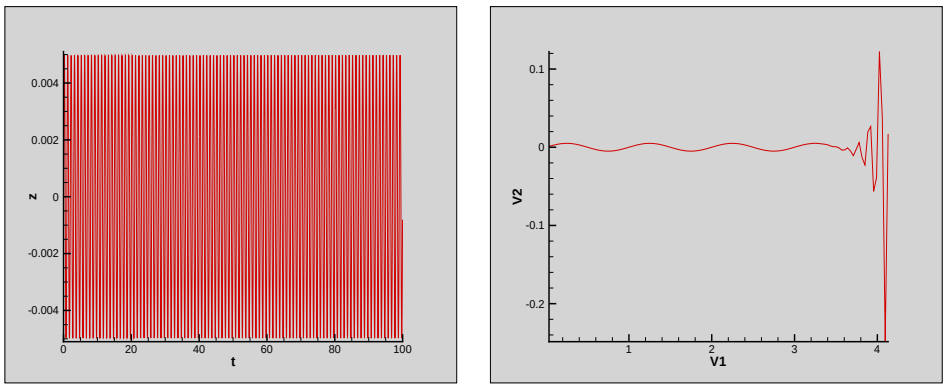

FIG. 6.3 – Pick-points d'un schéma à nombre CFL optimal (à gauche) et trop élevé (à droite).

Ordre théorique de la discrétisation	2	3	4	5	6
Lagrange standard	0.47	0.28	0.11	0.077	0.056
Lagrange condensé	0.94	0.38	0.13	0.084	0.033
Lagrange partiellement condensé	XX	XX	0.11	0.077	0.056

TAB. 6.2 – Nombres CFL optimaux pour les discrétisations en temps explicites symplectiques.

Il faut remarquer que les nombres CFL diminuent de manière assez alarmante lorsque l'on augmente l'ordre des schémas. Cela est essentiellement dû au fait que l'on considère comme pas d'espace l'aire du plus petit triangle du maillage, ce qui ne tient pas du tout compte, par exemple pour les éléments finis de Lagrange, ni du nombre, ni de la proximité des points auxquels on associe les degrés de liberté. Dans la pratique, il serait préférable de définir le pas d'espace comme étant la plus petite distance entre deux points auxquels on associe les degrés de liberté, ceci n'ayant de sens que pour les éléments finis de Lagrange. Remarquons d'autre part que la condensation partielle des éléments finis ne modifie en rien les limites de stabilité par rapport à l'utilisation des éléments finis standards. En revanche l'utilisation d'éléments finis (totalement) condensés nous permet d'utiliser des nombres CFL significativement plus élevés que ceux déterminés pour les éléments finis standards d'ordre bas, mais le rapport de ces nombres diminue avec l'augmentation de l'ordre (de 0.94 par rapport à 0.47 pour les éléments d'ordre 2, à 0.079 par rapport à 0.072 pour les éléments d'ordre 5), et va jusqu'à s'inverser pour les éléments d'ordre 6 et 7 (le nombre CFL devenant même plus de dix fois inférieur pour les éléments finis condensés d'ordre 7 par rapport à celui des éléments finis standards).

Nous déterminons les limites de stabilité de la même manière pour les différentes discrétisations en espace couplées cette fois-ci aux discrétisations symplectiques en temps. Les résultats sont listés dans le tableau 6.2. Nous avons fait le choix de ne pas tester les schémas à discrétisation en temps symplectique au-delà de l'ordre 6. En effet cela nous aurait fait utiliser une discrétisation en temps symplectique d'ordre 8 (rappelons qu'au-delà de l'ordre 4, seules les discrétisations symplectiques d'ordre pair sont connues), qui est la composée par elle-même de la discrétisation en temps symplectique d'ordre 6, ce qui implique un nombre de pas de temps intermédiaires de 256, et qui paraît très peu compétitif.

L'évolution des nombres CFL optimaux pour les schémas à semi-discrétisation en temps symplectique est en tous points similaire à celle des nombres CFL optimaux pour les schémas à semi-discrétisation en temps d'ordre arbitrairement élevé.
Il faut toutefois remarquer que l'utilisation des schémas symplectiques en temps semble plus avantageux (sur le seul critère d'un nombre CFL optimal plus élevé) pour les discrétisations en temps d'ordre impair (nettement pour les schémas d'ordre 3, et légèrement pour les schémas d'ordre 5) tandis que pour les ordres pairs ce sont les schémas à discrétisation en temps d'ordre arbitrairement élevé (sauf pour les schémas d'ordre 2 pour lesquels les contraintes sur le nombre CFL sont strictement identiques).

Ordre théorique de la discrétisation	2	3	4	5	6	7
Lagrange standard	1.99	3.07	4.15	4.93	5.63	6.96
Lagrange condensé	1.97	3.00	3.87	4.86	5.95	6.81
Lagrange partiellement condensé	XX	XX	4.14	4.95	5.56	

TAB. 6.3 – Ordres numériques de convergence pour les discrétisations en temps explicites d'ordre arbitrairement élevé.

Ordre théorique de la discrétisation	2	3	4	5	6
Lagrange standard	1.98	3.00	4.13	4.89	5.75
Lagrange condensé	2.00	3.06	3.88	4.76	5.92
Lagrange partiellement condensé	XX	XX	4.13	4.91	5.65

TAB. 6.4 – Ordres numériques de convergence pour les discrétisations en temps explicites symplectiques.

Nous allons maintenant vérifier que les ordres de convergence numériques correspondent bien aux ordres théoriques. Pour cela nous traçons les courbes de régression données par le logarithme de l'erreur (en norme L^2) en fonction du logarithme du pas d'espace, pour chacun des schémas numériques considérés, en raffinant le maillage. Pour obtenir une convergence plus propre, nous faisons cette étude au bout d'une seule période du signal, et en ne propageant que deux ondes sur la longueur du domaine (c'est-à-dire $k = \frac{2}{5}$) de manière à atteindre le comportement asymptotique des schémas plus facilement. Ces courbes de régression sont données par les figures 6.4, pour les discrétisations en temps explicites d'ordre arbitrairement élevé, et 6.5 pour les discrétisations en temps explicites symplectiques. Les ordres de convergence qui en sont déduits sont donnés dans les tableaux 6.3 et 6.4.

Chacun se rendra compte que, même s'il existe des écarts entre les ordres de convergence numérique et les ordres théoriques des schémas, les résultats restent toutefois très cohérents.

Dans la mesure où l'on perd tout l'intérêt de la condensation (même partielle) de la matrice de masse dès que l'on couple la discrétisation en espace avec une discrétisation en temps implicite, nous ne considérerons ces discrétisations en temps que lorsque nous utiliserons des éléments finis de Lagrange standards. Les discrétisations en temps implicites n'étant pas soumises à des restrictions sur le nombre CFL, nous allons tracer les courbes de régression pour un demi, un et deux fois le nombre CFL optimal que l'on a déterminé pour les discrétisations par éléments finis de Lagrange standards en espace couplées aux discrétisations explicites d'ordre arbitrairement élevé.
Nous déterminons parallèlement pour ces discrétisations en temps implicites le nombre CFL pour lequel la courbe de régression est la plus basse possible. C'est bien entendu pour ce nombre CFL que l'erreur de la solution numérique par rapport à la solution exacte est minimisée. Lorsque la courbe de régression semble insensible sur une large plage de nombres CFL (ce qui est le cas pour les discrétisations en temps implicites d'ordre élevé) nous déterminons le plus grand des nombres CFL pour lequel la courbe de régression reste

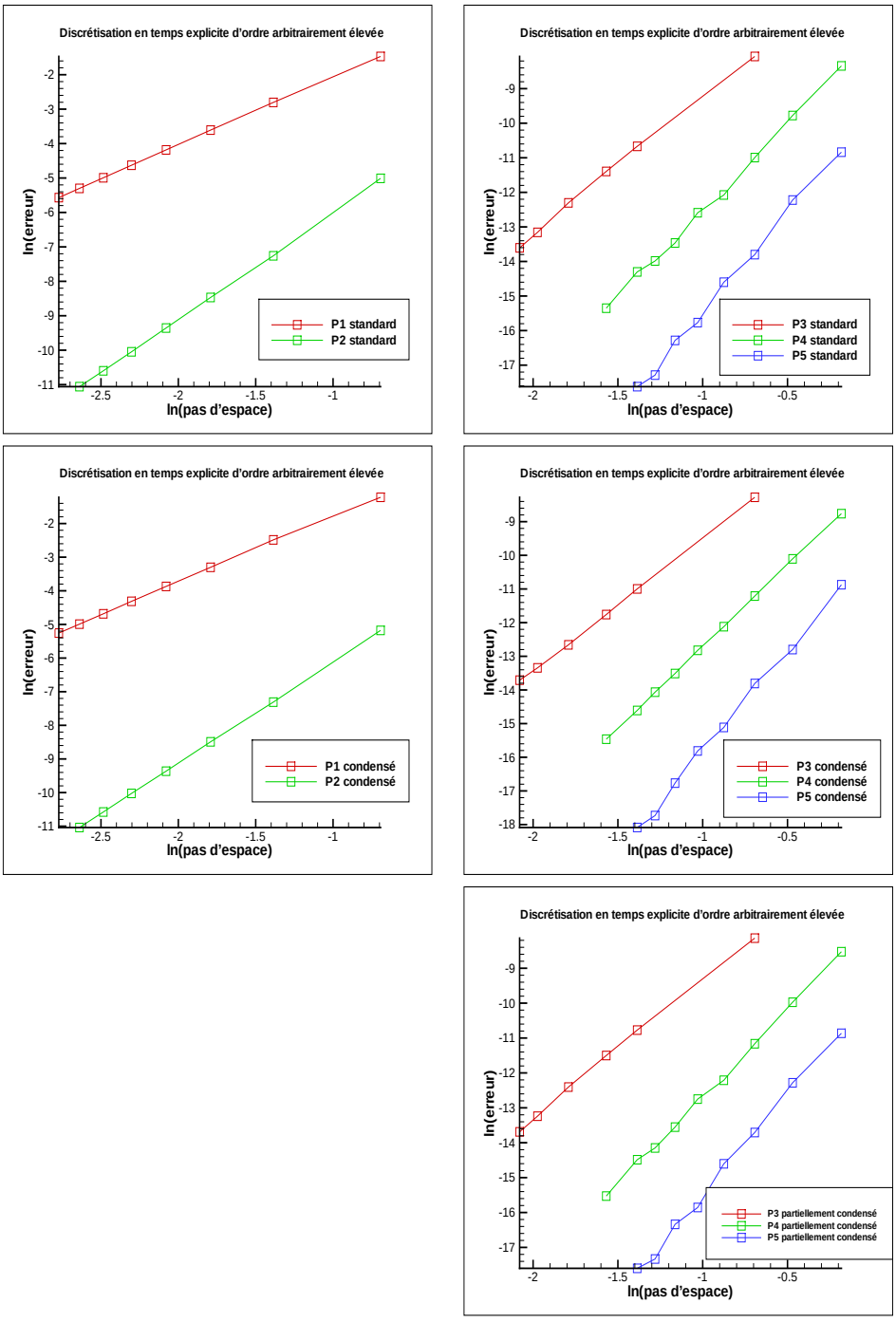

FIG. 6.4 – Convergence des schémas à discrétisation en temps explicite d'ordre arbitrairement élevé.

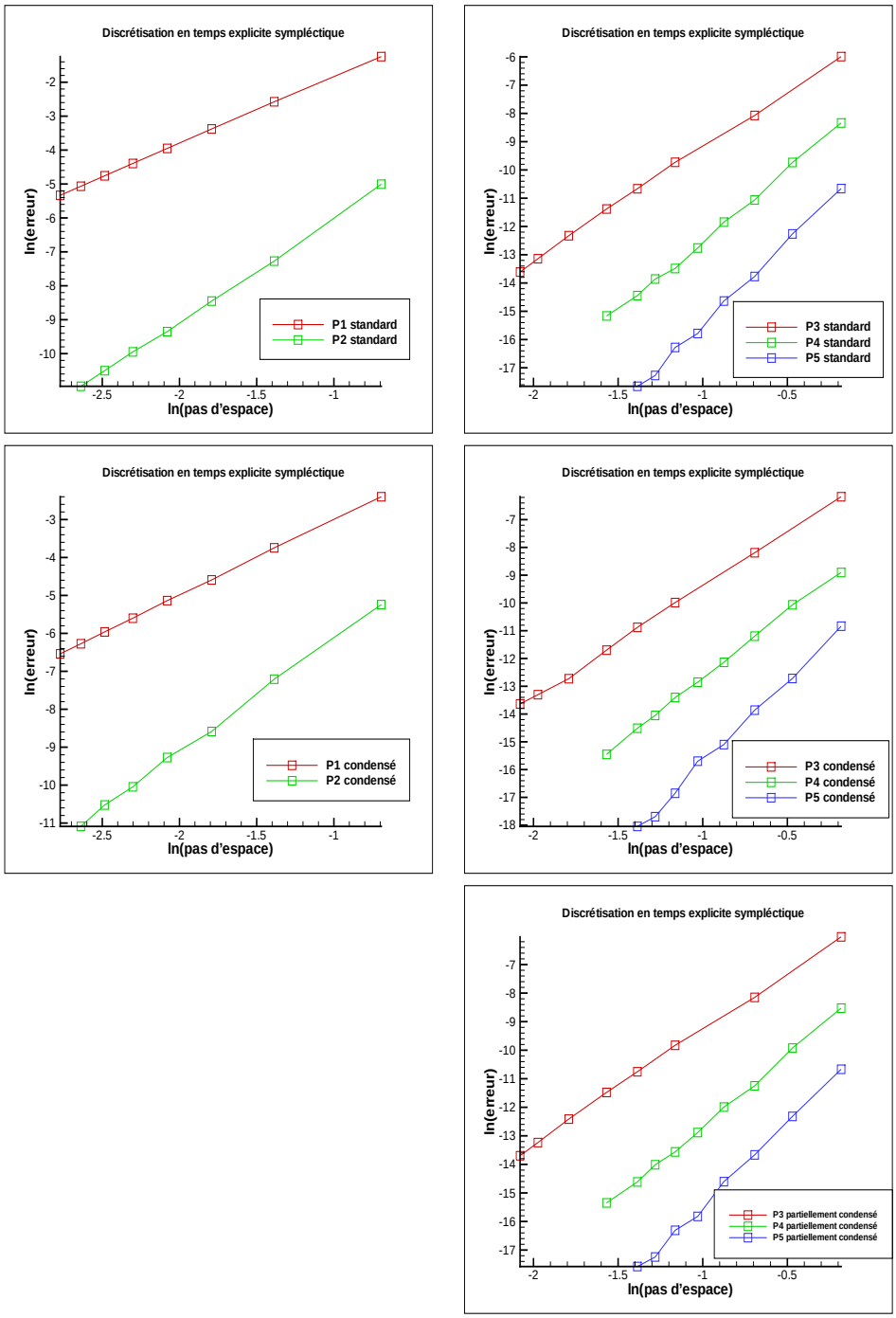

FIG. 6.5 – Convergence des schémas à discrétisation en temps explicite symplectique.

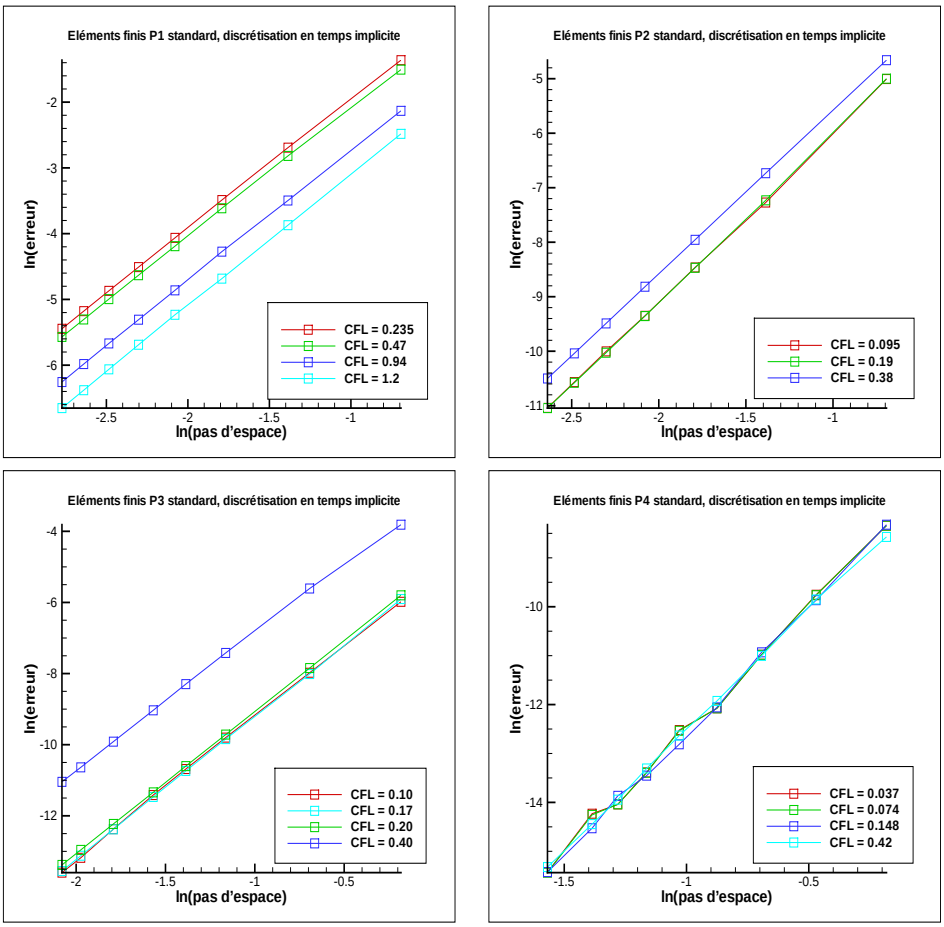

FIG. 6.6 – Convergence des schémas à discrétisation en temps implicite.

sensiblement au plus bas.
Ces courbes de régression sont données dans les figures 6.6, les nombres CFL minimisant l'erreur et les ordres numériques de convergence (calculés pour ces nombres CFL) sont listés respectivement dans les tableaux 6.5 et 6.6. Rappelons que nous n'avons pas trouvé dans la littérature de discrétisation en temps diagonalement implicite d'ordre supérieur à cinq. C'est pourquoi nous marquons à nouveau d'un XX les données faisant référence à un schéma que l'on n'a pas réellement implémenté.

Remarque 6.1.1. *Les nombres CFL minimisant l'erreur au bout d'une période de la propagation d'une onde plane, qui ne sont donnés qu'à titre indicatif, ne sont pas forcément les plus judicieux à utiliser pour simuler la propagation d'onde en temps long, et il serait faux de croire qu'ils sont optimaux (dans le sens où ce serait les plus grands nombres CFL minimisant l'erreur quel que soit le cas test considéré). En effet le fait de se fixer une période du signal comme temps final pour déterminer l'ordre de convergence des schémas nous permet d'atteindre le comportement asymptotique des schémas plus facilement, mais nous fait négliger les mauvaises (ou bonnes) propriétés de dissipation et dispersion de ces*

Ordre théorique de la discrétisation	2	3	4	5	6
Lagrange standard	1.2	0.19	0.17	0.42	XX

TAB. 6.5 – Nombres CFL minimisant l'erreur au bout d'une période de propagation d'une onde plane pour les discrétisations en temps implicites.

Ordre théorique de la discrétisation	2	3	4	5	6
Lagrange standard	2.02	3.07	4.04	4.91	XX

TAB. 6.6 – Ordres numériques de convergence pour les discrétisations en temps implicites.

schémas. Les nombres CFL minimisant l'erreur au bout d'une période du signal ne sont donc pas forcément ceux minimisant l'erreur au bout d'une centaine de périodes du signal.

6.1.2 Quantification de la dissipation et de la dispersion numérique

Nous allons à présent nous intéresser aux propriétés de dissipation et de dispersion numérique de nos schémas. Un schéma numérique est dit dissipatif lorsque l'amplitude d'une onde plane qui est propagée n'est pas conservée. Un schéma est dit dispersif lorsque une onde plane qui est propagée l'est à une mauvaise vitesse.

Pour quantifier la dissipation numérique de nos schémas nous allons donc propager une onde plane sur le domaine $\Omega = [0,1] \times [0,1]$ et déterminer le raffinement du maillage nécessaire à ce que la dissipation numérique soit inférieure à 1% au bout de 1000 périodes (théoriques) du signal. Il nous a bien entendu fallu adapter le nombre d'onde de cette onde plane, c'est-à-dire le nombre de périodes du signal sur la largeur du domaine Ω à un temps fixé, suivant la dissipation des schémas considérés : pour les schémas les moins dissipatifs, qui ont une dissipation déjà inférieure à la tolérance que l'on s'est fixée sur le maillage le plus grossier que l'on puisse considérer (maillage $1 \times 1 \times 4$ éléments), nous avons augmenté le nombre d'onde. Pour comparer la dissipation des différents schémas il nous faut un bon indicateur : typiquement en une dimension d'espace nous aurions considéré le nombre de degrés de liberté par longueur d'onde. C'est essentiellement pour cela que nous considérons un domaine carré : l'onde plane se propageant suivant une seule dimension d'espace (l'axe des abscisses), et les degrés de libertés étant uniformément répartis sur le domaine (le maillage étant périodique il n'y a pas de zone à plus forte densité de degrés de liberté), nous estimerons le nombre moyen de degrés de liberté par longueur d'onde par la racine carrée du nombre de degrés de liberté sur le domaine, divisée par le nombre d'onde.

Les résultats de cette étude de dissipation sont consignés dans le tableau 6.7 pour les schémas à discrétisation en temps explicite d'ordre arbitrairement élevé.

La première constatation que l'on fait à la vue des résultats est la diminution spectaculaire du nombre de degrés de liberté par longueur d'onde qui accompagne l'augmentation

Éléments finis de Lagrange standards					
Ordre théorique du schéma	2	3	4	5	6
Nombre d'onde	1	2	4	6	6
Nombre d'éléments du maillage	16384	5184	676	324	196
Nombre de degrés de liberté	8321	10513	3121	2665	2521
Nombre de degrés de liberté par longueur d'onde	91	51	14	9	8
Éléments finis de Lagrange condensés					
Ordre théorique du schéma	2	3	4	5	6
Nombre d'onde	1	2	4	6	6
Nombre d'éléments du maillage	64516	9604	1156	484	196
Nombre de degrés de liberté	32513	29009	7617	5413	4285
Nombre de degrés de liberté par longueur d'onde	180	85	22	12	11
Éléments finis de Lagrange partiellement condensés					
Ordre théorique du schéma	2	3	4	5	6
Nombre d'onde	XX	XX	4	6	6
Nombre d'éléments du maillage	XX	XX	676	324	196
Nombre de degrés de liberté	XX	XX	3121	2665	2521
Nombre de degrés de liberté par longueur d'onde	XX	XX	14	9	8

TAB. 6.7 – Raffinement du maillage nécessaire à une dissipation inférieure à 1% au bout de 1000 périodes pour les discrétisations en temps explicites d'ordre arbitrairement élevé.

Éléments finis de Lagrange standard					
Ordre théorique du schéma	2	3	4	5	6
Nombre d'onde	non-dissipatif	1	2	3	XX
Nombre d'éléments du maillage	non-dissipatif	2116	676	900	XX
Nombre de degrés de liberté	non-dissipatif	4325	3121	7321	XX
Nombre de degrés de liberté par longueur d'onde	non-dissipatif	66	28	29	XX

TAB. 6.8 – Raffinement du maillage nécessaire à une dissipation inférieure à 1% au bout de 1000 périodes pour les discrétisations en temps implicites.

de l'ordre des schémas numériques quel que soit le type de discrétisation en temps. Remarquons aussi que si l'utilisation de la condensation partielle des éléments finis ne modifie en rien la dissipation des schémas par rapport aux éléments finis standards, il n'en est pas de même pour l'utilisation de la condensation (totale) des éléments finis. En effet pour les 91 degrés de liberté nécessaires aux éléments finis standards d'ordre 2 pour une dissipation inférieure à 1% au bout de 1000 périodes, il en faut le double pour les éléments finis condensés du même ordre. Toutefois le rapport de ces nombres de degrés de liberté tend à diminuer avec l'augmentation en ordre des schémas.
Nous avons testé numériquement que ces mauvaises propriétés de dissipations peuvent être effacées en ne prenant pas les nombres CFL optimaux pour les éléments finis condensés mais en les diminuant aux valeurs des nombres CFL optimaux pour les éléments finis standards : cela signifie exactement que la condensation des éléments finis nous permet d'utiliser des nombres CFL plus élevés mais que cela implique de plus mauvaises propriétés de dissipation.

Pour les schémas numériques à discrétisation en temps symplectique, qui sont par définition non-dissipatifs, nous avons vérifié avec succès qu'il n'y a aucune perte sur l'amplitude de l'onde propagée.

Pour les schémas numériques à discrétisation en temps implicite, nous avons dans un premier temps fait cette étude pour les nombre CFL minimisant l'erreur au bout d'une période de la propagation de l'onde. Les résultats sont consignés dans le tableau 6.8.
Nous avons constaté que la discrétisation en temps implicite d'ordre 2 est non-dissipative. Si les 66 degrés de liberté par longueur d'onde nécessaires à une dissipation inférieure à la tolérance que l'on s'est fixée pour la discrétisation implicite d'ordre 3 sont tout à fait comparables aux 51 degrés de liberté par longueur d'onde nécessaires à la discrétisation en temps d'ordre arbitrairement élevé du même ordre (couplée, bien entendu, avec la même discrétisation en espace), il n'en va plus de même avec les discrétisations d'ordre plus élevées : pour les schémas d'ordre 4 et 5 il faut respectivement 28 degrés de liberté par longueur d'onde contre les 14 (soit le double) et 29 contre les 9 (soit plus du triple) degrés de liberté par longueur d'onde nécessaires pour les discrétisations en temps d'ordre arbitrairement élevé. Rappelons que les nombres CFL utilisés pour ces discrétisations en temps sont ceux du tableau 6.5 et n'ont rien d'optimal en temps long : sans doutes ceux-ci

Éléments finis de Lagrange standards					
Ordre théorique du schéma	2	3	4	5	6
Nombre d'onde	non-dissipatif	2	4	6	XX
Nombre d'éléments du maillage	non-dissipatif	5184	676	324	XX
Nombre de degrés de liberté	non-dissipatif	10513	3121	2665	XX
Nombre de degrés de liberté par longueur d'onde	non-dissipatif	51	14	9	XX
Nombre CFL	non-dissipatif	0.14	0.084	0.12	XX

TAB. 6.9 – Nombres CFL optimaux nécessaires à une dissipation inférieure à 1% au bout de 1000 périodes pour les discrétisations en temps implicites en comparaison aux discrétisations en temps explicites d'ordre arbitrairement élevé .

sont bien trop élevés (notamment pour la discrétisation implicite d'ordre 5 pour laquelle on utilise un nombre CFL de 0.42 à comparer au nombre CFL de 0.072 pour la discrétisation en temps d'ordre arbitrairement élevé d'ordre 5) pour assurer une dissipation du même ordre que les schémas à discrétisation en temps d'ordre arbitrairement élevé.
C'est pourquoi nous nous sommes aussi intéressés, pour les schémas à discrétisation en temps implicite, aux nombres CFL les plus grands possibles permettant une dissipation inférieure à 1% au bout de 1000 périodes dans les mêmes conditions que les schémas à discrétisation en temps explicite d'ordre arbitrairement élevé dont la dissipation est elle-même inférieure à cette tolérance. Ces nombres CFL sont listés dans le tableau 6.9.
Il y a, à ce moment déjà, pas mal de questions à se poser par rapport à l'efficacité des schémas à discrétisation en temps implicite. Rappelons que si notre intérêt s'est porté sur ce type de discrétisation c'est pour nous débarrasser de la borne sur le nombre CFL assurant la stabilité de la discrétisation en temps. Or pour assurer une dissipation du même ordre que les discrétisations en temps d'ordre arbitrairement élevé nous voyons à présent que l'on ne peut pas augmenter le nombre CFL à volonté : il faut même diminuer ce nombre de 0.19 à 0.14 pour le schéma d'ordre 3 et de 0.19 à 0.084 pour le schéma d'ordre 4.

Nous allons à présent étudier de la même manière la dispersion numérique de nos schémas. Nous allons déterminer le raffinement nécessaire à ce que la dispersion numérique de nos schémas soit inférieure à une tolérance que l'on se fixe à 0.1%, c'est-à-dire à ce que le nombre de périodes effectivement propagées soit compris entre 999 et 1001 pour 1000 périodes théoriques. Nous avons expérimenté que les schémas numériques d'ordre élevé, notamment ceux utilisant une discrétisation spatiale par éléments finis de Lagrange condensés, sont extrêmement peu dispersifs quelle que soit la discrétisation temporelle utilisée, c'est pourquoi nous abaissons cette tolérance jusqu'à 0.001% pour ces schémas. Dans la pratique nous ne propageons pas le signal sur 100000 périodes théoriques mais nous déterminons le raffinement du maillage nécessaire à ce que la dispersion soit inférieure à un centième de période au bout de 1000 périodes théoriques.
Les résultats pour les schémas à discrétisation en espace par éléments finis de Lagrange standards, condensés et partiellement condensés sont donnés dans le tableau 6.10 lorsqu'ils sont couplés aux discrétisations en temps explicites d'ordre arbitrairement élevé et dans le

Éléments finis de Lagrange standards					
Ordre théorique du schéma	2	3	4	5	6
Dispersion	0.1%	0.1%	0.01%	0.01%	0.001%
Nombre d'onde	1	3	4	4	5
Nombre d'éléments du maillage	3844	784	576	196	256
Nombre de degrés de liberté	1985	1625	3613	1625	3281
Nombre de degrés de liberté par longueur d'onde	45	13	13	10	11
Éléments finis de Lagrange condensés					
Ordre théorique du schéma	2	3	4	5	6
Dispersion	0.1%	0.001%	0.001%	0.001%	0.001%
Nombre d'onde	1	2	4	6	8
Nombre d'éléments du maillage	5184	1156	900	484	576
Nombre de degrés de liberté	2665	3537	5941	5413	12505
Nombre de degrés de liberté par longueur d'onde	52	30	19	12	14
Éléments finis de Lagrange partiellement condensés					
Ordre théorique du schéma	2	3	4	5	6
Dispersion	XX	XX	0.01%	0.01%	0.001%
Nombre d'onde	XX	XX	4	4	5
Nombre d'éléments du maillage	XX	XX	576	196	256
Nombre de degrés de liberté	XX	XX	3613	1625	3281
Nombre de degrés de liberté par longueur d'onde	XX	XX	13	10	11

TAB. 6.10 – Raffinement du maillage nécessaire à une dispersion inférieure à 0.1%, 0.01% ou 0.001% pour les discrétisations en temps explicites d'ordre arbitrairement élevé.

tableau 6.11 lorsqu'ils sont couplés aux discrétisations en temps explicites symplectiques.

De nouveau l'augmentation de l'ordre des schémas s'accompagne d'une diminution du nombre de degrés de liberté par longueur d'onde nécessaires à une tolérance de dispersion donnée. Par exemple pour les schémas à discrétisation en espace par éléments finis de Lagrange standards couplés aux discrétisations en temps symplectiques, s'il faut 50 degrés de liberté par longueur d'onde au schéma d'ordre 2 pour une dispersion inférieure à 1 pour 1000, il n'en faut que 11 au schéma d'ordre 6 pour une dispersion inférieure à 1 pour 100000. Seuls les schémas à discrétisation en espace par éléments finis de Lagrange condensés d'ordre 6 ne vont pas dans ce sens : que ce soit avec une discrétisation d'ordre arbitrairement élevé ou symplectique il faut 14 degrés de liberté par longueur d'onde, soit 2 de plus que les 12 nécessaires aux schémas d'ordre 5.
Remarquons aussi qu'à nouveau l'utilisation d'éléments finis partiellement condensés, par rapport à l'utilisation des éléments finis standards, ne modifie en rien les propriétés de dispersion des schémas. En revanche l'utilisation d'éléments finis (totalement) condensés entraîne des propriétés de dispersion bien plus avantageuses par rapport à l'utilisation

Éléments finis de Lagrange standards					
Ordre théorique du schéma	2	3	4	5	6
Dispersion	0.1%	0.1%	0.01%	0.01%	0.001%
Nombre d'onde	1	2	3	4	5
Nombre d'éléments du maillage	4900	324	324	196	256
Nombre de degrés de liberté	2521	685	1513	1625	3281
Nombre de degrés de liberté par longueur d'onde	50	13	13	10	11
Éléments finis de Lagrange condensés					
Ordre théorique du schéma	2	3	4	5	6
Dispersion	0.01%	0.001%	0.001%	0.001%	0.001%
Nombre d'onde	1	2	4	6	8
Nombre d'éléments du maillage	7396	1024	400	484	576
Nombre de degrés de liberté	3785	3137	2661	5413	12505
Nombre de degrés de liberté par longueur d'onde	62	28	13	12	14
Éléments finis de Lagrange partiellement condensés					
Ordre théorique du schéma	2	3	4	5	6
Dispersion	XX	XX	0.01%	0.01%	0.001%
Nombre d'onde	XX	XX	3	4	5
Nombre d'éléments du maillage	XX	XX	324	196	256
Nombre de degrés de liberté	XX	XX	1513	1625	3281
Nombre de degrés de liberté par longueur d'onde	XX	XX	13	10	11

TAB. 6.11 – Raffinement du maillage nécessaire à une dispersion inférieure à 0.1%, 0.01% ou 0.001% pour les discrétisations en temps explicites symplectiques.

des éléments finis standards : par exemple, pour le schéma d'ordre 4 à discrétisation en temps symplectique, s'il faut 13 degrés de liberté par longueur d'onde aux éléments finis de Lagrange standards pour une dispersion inférieure à 1 pour 10000, il en faut exactement autant aux éléments finis de Lagrange condensés pour une dispersion cette fois-ci inférieure à 1 pour 100000, soit dix fois inférieure. De la même manière pour le schéma d'ordre 5 à discrétisation en temps d'ordre arbitrairement élevé, s'il faut 10 degrés de liberté par longueur d'onde aux éléments finis de Lagrange standards pour une dispersion inférieure à 1 pour 10000, il n'en faut que 12 aux éléments finis de Lagrange condensés pour une dispersion 10 fois inférieure (nous avons bien entendu vérifié qu'avec 12 degrés de liberté par longueur d'onde, les éléments finis de Lagrange standards sont encore bien loin de cette précision de 1 pour 100000). De nouveau les éléments finis condensés d'ordre 6 semblent avoir un comportement qui ne va pas dans ce sens : que ce soit pour le schéma à discrétisation en temps d'ordre arbitrairement élevé ou symplectique, l'utilisation de la discrétisation en espace par éléments finis de Lagrange condensés nécessite 14 degrés de liberté par longueur d'onde pour une dispersion de 1 pour 100000 plutôt que les 11 nécessaires aux éléments finis de Lagrange standards.
Il reste alors à comparer les propriétés de dispersion des schémas à discrétisation en temps d'ordre arbitrairement élevé aux schémas à discrétisation en temps symplectique. Celles-ci sont tout à fait du même ordre, voire même identiques dans la plupart des cas, sauf pour les éléments finis de Lagrange condensés d'ordre 4, qui, s'ils sont couplés à la discrétisation symplectique du même ordre nécessite 13 degrés de liberté par longueur d'onde pour une dispersion de 1 pour 100000 contre les 19 nécessaires à la discrétisation en temps d'ordre arbitrairement élevé ; et surtout pour les éléments finis de Lagrange condensés d'ordre 2, pour lesquels 62 degrés de liberté par longueur d'onde suffisent à une dispersion inférieure à 1 pour 10000 avec la discrétisation en temps symplectique, alors qu'il faut déjà 52 degrés de liberté par longueur d'onde à la discrétisation en temps d'ordre arbitrairement élevé pour une dispersion de 1 pour 1000, soit dix fois supérieure. Rappelons que le schéma à discrétisation en espace par élément finis de Lagrange condensés couplée à la discrétisation en temps d'ordre arbitrairement élevé d'ordre 2 était déjà le plus dissipatif de nos schémas, il se trouve maintenant que c'est aussi le plus dispersif.

Les schémas à discrétisation en temps implicite se sont révélés avoir des propriétés tout à fait remarquables : il semble être possible d'ajuster la vitesse de propagation d'une onde plane (ou plutôt de l'approximation numérique que l'on en fait), et donc de minimiser la dispersion numérique de ces schémas de manière à la rendre aussi petite que l'on veut, en ajustant le nombre CFL. Dans la pratique cela n'a d'intérêt que pour la propagation d'une onde plane, c'est-à-dire lorsque la simulation numérique ne fait intervenir qu'une seule longueur d'onde. C'est pourquoi nous choisissons de ne pas déterminer le nombre CFL minimisant la dispersion numérique de manière très précise, mais uniquement sur deux chiffres significatifs, et de quantifier la dispersion du schéma pour ce nombre CFL. Le tableau 6.12 liste ces nombres CFL pour les différents ordres de schémas et pour différents nombres de degrés de liberté par longueur d'onde. Il est alors bon de déterminer les bonnes ou mauvaises propriétés de dissipation des schémas utilisés dans ces conditions (c'est-à-dire de se demander si les nombres CFL rendant les schémas optimaux en terme de dispersion, font de ces schémas des schémas raisonnablement dissipatifs), c'est pourquoi nous listons conjointement dans ce même tableau la proportion de signal dis-

sipé au bout de 1000 périodes théoriques. À titre de comparaison nous avons aussi listé dans le tableau 6.13 la dispersion des schémas à discrétisation en temps implicite dans les conditions décrites par le tableau 6.9, c'est-à-dire dans les conditions rendant ces schémas aussi peu dissipatifs que les schémas à discrétisation en temps d'ordre arbitrairement élevé.

La première constatation est que pour les schémas d'ordre 2 et 3, le nombre CFL permettant à l'onde de se propager à la bonne vitesse ne dépend pas du nombre de degrés de liberté par longueur d'onde : entre 1.1 et 1.2 pour le schéma d'ordre 2 et 0.46 pour le schéma d'ordre 3. Ceci est une très bonne propriété puisque sur des cas tests plus réalistes il n'y a pas qu'une seule onde à propager mais bien une superposition de plusieurs ondes qui ont toutes leur propre longueur d'onde. La discrétisation en temps implicite d'ordre 2 étant par ailleurs non-dissipative, cela nous laisse préjuger d'un schéma du second ordre des plus efficaces. Malheureusement pour la discrétisation en temps d'ordre 3, le nombre CFL de 0.46 rend le schéma extrêmement dissipatif : même à 41 degrés de liberté par longueur d'onde, c'est plus de 30% du signal qui est dissipé au bout de 1000 périodes de propagation. Pour ce schéma il devient donc clair que le nombre CFL optimisant la dispersion est bien trop élevé pour assurer une dissipation raisonnable : pour un nombre CFL de 0.14 et 51 degrés de liberté par longueur d'onde la dissipation est inférieure à 1% au bout de 1000 périodes, alors que la dispersion n'est que de 0.00042%. Il semble donc être préférable dans ce cas de ne pas minimiser la dispersion du schéma en choisissant un nombre CFL de 0.46 mais de diminuer ce nombre de manière à en minimiser la dissipation. Considérant le schéma d'ordre 4, non seulement le nombre CFL permettant à l'onde de se propager à la bonne vitesse dépend fortement du nombre de degrés de liberté par longueur d'onde, mais en plus le signal est entièrement dissipé quel que soit ce nombre de degrés de liberté. Ces nombres CFL sont à nouveau bien trop élevés par rapport aux nombres CFL de 0.084 rendant la dissipation de ce schéma inférieure à 1% au bout de 1000 périodes pour uniquement 14 degrés de liberté par longueur d'onde. D'autant que la dispersion de 0.0063% dans ces conditions, reste tout à fait du même ordre que les 0.0024% ou 0.0035% de dispersion que l'on obtient pour les nombre CFL optimisant la dispersion de ce schéma pour respectivement 9 et 17 degrés de liberté par longueur d'onde. Remarquons finalement que pour le schéma d'ordre 5, si pour les 9 degrés de liberté par longueur d'onde il faut utiliser un nombre CFL de 0.39 pour minimiser la dispersion, nous avons vu qu'il faut utiliser un nombre CFL de 0.12 pour que la dissipation soit inférieure à 1% au bout de 1000 périodes, c'est pourquoi la quasi-totalité du signal (89.2%) est dissipée. En revanche la dispersion de 0.023% pour ce nombre CFL de 0.12 reste raisonnablement faible, même si elle n'atteint pas les 0.0012% que nous obtenons pour le nombre CFL de 0.39.

Pour résumer ce qui vient d'être dit sur les schémas à discrétisation en temps implicite, il faut retenir que s'il est possible d'ajuster le nombre CFL suivant le nombre de degrés de liberté par longueur d'onde pour minimiser la dispersion numérique des schémas, cet ajustement ne semble pas être des plus judicieux dans la mesure où l'influence du nombre CFL sur la dispersion des schémas ne semble que très faible par rapport à son influence sur la dissipation de ces schémas. L'utilisation des discrétisations en temps implicites va donc s'avérer très délicate (sauf pour le schéma d'ordre 2) : il va s'agir de faire un compromis entre notre volonté d'utiliser un nombre CFL élevé pour atteindre le temps final du calcul plus

Schéma à discrétisation en temps implicite d'ordre 2					
Nombre d'onde	2	1	1	1	1
Nombre d'éléments du maillage	576	256	900	3600	10000
Nombre de degrés de liberté	313	145	481	1861	5101
Nombre de degrés de liberté par longueur d'onde	9	12	21	43	71
Nombre CFL	1.1	1.1	1.2	1.2	1.2
Dispersion du schéma	0.27%	0.13%	0.042%	0.0099%	0.0035%
proportion de signal dissipé	0%	0%	0%	0%	0%
Schéma à discrétisation en temps implicite d'ordre 3					
Nombre d'onde	3	3	3	3	
Nombre d'éléments du maillage	400	1764	3600	7396	
Nombre de degrés de liberté	841	3613	7321	14965	
Nombre de degrés de liberté par longueur d'onde	10	20	29	41	
Nombre CFL	0.46	0.46	0.46	0.46	
Dispersion du schéma	0.011%	0.00019%	0.000025%	0.00001%	
proportion de signal dissipé	100%	99.1%	81%	30.6%	
Schéma à discrétisation en temps implicite d'ordre 4					
Nombre d'onde	3	2	2		
Nombre d'éléments du maillage	144	256	484		
Nombre de degrés de liberté	685	1201	2245		
Nombre de degrés de liberté par longueur d'onde	9	17	24		
Nombre CFL	0.44	0.88	1.2		
Dispersion du schéma	0.0024%	0.0035%	0.0045%		
proportion de signal dissipé	100%	100%	100%		
Schéma à discrétisation en temps implicite d'ordre 5					
Nombre d'onde	4	3	2		
Nombre d'éléments du maillage	144	484	484		
Nombre de degrés de liberté	1201	3961	3961		
Nombre de degrés de liberté par longueur d'onde	9	21	31		
Nombre CFL	0.39	0.29	0.25		
Dispersion du schéma	0.0012%	<0.00001%	<0.00001%		
proportion de signal dissipé	89.2%	0.61%	0.052%		

TAB. 6.12 – Nombres CFL minimisant la dispersion des schémas à discrétisation en temps implicite en fonction du nombre de degrés de liberté par longueur d'onde et dissipation inhérente à leur utilisation.

Éléments finis de Lagrange standard			
Ordre théorique du schéma	3	4	5
Nombre d'onde	2	4	6
Nombre d'éléments du maillage	5184	676	324
Nombre de degrés de liberté	10513	3121	2665
Nombre de degrés de liberté par longueur d'onde	51	14	9
Nombre CFL	0.14	0.084	0.12
Dispersion du schéma	0.00042%	0.0063%	0.023%

TAB. 6.13 – Dispersion des schémas à discrétisation en temps implicite dans les conditions les rendant aussi peu dissipatifs que les schémas à discrétisation explicite d'ordre arbitrairement élevé.

rapidement, et la nécessité d'ajuster ce nombre CFL de manière à minimiser conjointement la dissipation et la dispersion des schémas.

6.1.3 Rapport coût/précision

Nous allons à présent nous intéresser au rapport coût/précision des différents schémas de la manière suivante : nous allons déterminer le nombre de degrés de liberté (abrégé par NTDDL pour Nombre Total de Degrés De Liberté) et le temps de calcul nécessaire pour atteindre une précision donnée à un temps final que l'on se fixe pour chacun des schémas numériques. Dans la pratique nous avons propagé une onde plane (de longueur d'onde 1) sur le domaine $\Omega = [0,1] \times [0,1]$ sur dix périodes et raffiné le maillage jusqu'à atteindre une erreur (en norme $L^2(\Omega)$) inférieure à $e^{-5.60}$, $e^{-11.20}$ puis $e^{-15.20}$ et consigné les résultats dans les tableaux 6.14 à 6.16. Dans ces tableaux, et dans les commentaires les accompagnant, nous ferons référence aux discrétisations en temps d'ordre arbitrairement élevé simplement sous les termes de discrétisation en temps explicite (bien que les discrétisation en temps symplectiques soient elles aussi explicites).

La première constatation que l'on peut faire est qu'il nous a été impossible d'atteindre l'erreur de $e^{-11.20}$ avec aucun des schémas d'ordre 2, ni l'erreur de $e^{-15.20}$ avec aucun des schémas d'ordre 3, le coût en terme de stockage mémoire devenant bien trop élevé.
Parmi les schémas d'ordre 2, le plus coûteux est le P_1 standard-explicite, qui nécessite $1069.s$ et 31501 degrés de liberté pour atteindre l'erreur de $e^{-5.60}$, alors qu'en utilisant le P_1 condensé-explicite, même s'il faut plus de degrés de liberté (le schéma étant plus dissipatif et plus dispersif), il ne faut plus que $24.s$ (la matrice de masse étant diagonale et le nombre CFL pouvant être doublé) pour atteindre la même précision. Nous avions déjà souligné le fait que le schéma d'ordre 2 à discrétisation en temps implicite serait sûrement très efficace, notamment à cause de sa non-dissipativité : il ne lui faut que $2.s$ et 3445 degrés de liberté pour atteindre l'erreur de $e^{-5.60}$. Si c'est pour ce dernier schéma que l'on a besoin du moins de degrés de liberté parmi les schémas d'ordre 2, c'est toutefois pour le P_1 condensé-symplectique que l'on a besoin de moins de temps, soit 1.s, même si le nombre de degrés de liberté est bien supérieur. Pour tous les schémas d'ordre plus élevé,

ln(erreur) < -5.60					
Disc. en espace	Disc. en temps	nombre CFL	ln(erreur)	NTDDL	temps CPU
P_2 standard	explicite	0.19	-5.85	841	1.s
P_2 standard	symplectique	0.28	-5.87	841	<1.s
P_2 standard	implicite	0.19	-5.81	841	1.s
P_2 condensé	explicite	0.26	-5.70	801	<1.s
P_2 condensé	symplectique	0.38	-5.59	457	<1.s
P_1 standard	explicite	0.47	-5.60	31501	1069.s
P_1 standard	implicite	1.2	-5.60	3445	2.s
P_1 condensé	explicite	0.94	-5.60	44105	24.s
P_1 condensé	symplectique	0.94	-5.61	6845	1.s

TAB. 6.14 – Coût du calcul (en temps et nombre de degrés de liberté) pour atteindre une erreur (en norme $L^2(\Omega)$) inférieure à $e^{-5.60}$

cette erreur est atteinte quasi instantanément et pour un nombre de degrés de liberté bien inférieur.
Pour les schémas à discrétisation en temps implicite d'ordre 3, 4 et 5, nous avons déterminé les nombres CFL minimisant les temps de calcul. Ces nombres CFL sont de 0.16 pour le schéma d'ordre 3, de 0.11 pour le schéma d'ordre 4 et de 0.32 ou 0.27 pour le schéma d'ordre 5 suivant que l'on veuille atteindre une précision de respectivement $e^{-11.20}$ ou $e^{-15.20}$. Remarquons qu'il n'y a que pour le schéma d'ordre 5 que ce nombre est significativement plus élevé que celui pour les schémas du même ordre à discrétisation en temps explicite ou symplectique, et que même dans ce cas ce n'est pas le schéma minimisant le temps de calcul, les schémas à discrétisation en espace condensée faisant systématiquement mieux. Le fait de ne plus avoir de borne sur le nombre CFL s'avère donc être un gain purement théorique et inexploitable dans la pratique.
De manière générale, il faut se rendre compte que l'utilisation de schémas d'ordre plus élevé nous permet d'atteindre une précision donnée avec un nombre bien inférieur de degrés de liberté mais pas forcément plus rapidement : par exemple, si pour le schéma P_2 condensé-symplectique il ne faut que $7.s$ pour atteindre une erreur de $e^{-11.20}$, il en faut tout de même $18.s$ au P_3 standard-symplectique.
Si les discrétisations en espace par éléments finis de Lagrange condensés $\tilde{P}_1$ à $\tilde{P}_5$ montrent leur efficacité en terme de temps CPU, malgré le surcoût en degrés de liberté nécessaires pour compenser une dissipation accrue lorsqu'ils sont couplés avec une discrétisation explicite en temps, il n'en va plus de même pour les éléments finis de Lagrange condensés $\tilde{P}_6$: cela est essentiellement dû à la restriction beaucoup plus forte sur le nombre CFL qui n'arrive plus à être compensée ni par le fait que la matrice de masse est diagonale, ni par le fait que l'espace polynomial associé aux éléments finis condensés $\tilde{P}_6$ est plus fin que celui associé aux éléments finis standards P_6. Par rapport aux schémas à discrétisation en espace partiellement condensée, notons que si le gain en temps n'est pas aussi spectaculaire, celui-ci tend à augmenter avec l'ordre du schéma. Cela est naturel dans la mesure où le nombre de fonctions de base à orthogonaliser par rapport au nombre de fonctions de base de l'espace polynomial définissant l'élément fini augmente avec l'ordre de cet élément, ce

ln(erreur) < -11.20					
Disc. en espace	Disc. en temps	nombre CFL	ln(erreur)	NTDDL	temps CPU
P_5 standard	explicite	0.097	-12.39	841	3.s
P_5 standard	symplectique	0.056	-12.29	841	12.s
P_5 condensé	explicite	0.058	-12.71	1417	2.s
P_5 condensé	symplectique	0.033	-12.43	1417	9.s
P_5 part. condensé	explicite	0.097	-12.40	841	2.s
P_5 part. condensé	symplectique	0.056	-12.36	841	8.s
P_4 standard	explicite	0.072	-11.72	1201	6.s
P_4 standard	symplectique	0.077	-11.69	1201	15.s
P_4 standard	implicite	0.42	-11.41	2113	5.s
P_4 standard	implicite	0.32	-11.26	1201	2.s
P_4 standard	implicite	0.21	-11.93	1201	3.s
P_4 condensé	explicite	0.079	-12.03	1633	1.s
P_4 condensé	symplectique	0.084	-11.97	1633	3.s
P_4 part. condensé	explicite	0.072	-11.90	1201	5.s
P_4 part. condensé	symplectique	0.077	-11.85	1201	11.s
P_3 standard	explicite	0.19	-11.59	2665	11.s
P_3 standard	symplectique	0.11	-11.66	2665	18.s
P_3 standard	implicite	0.17	-11.21	4705	40.s
P_3 standard	implicite	0.11	-11.28	2665	18.s
P_3 standard	implicite	0.085	-11.59	2665	25.s
P_3 condensé	explicite	0.24	-11.35	3817	2.s
P_3 condensé	symplectique	0.13	-11.21	3213	2.s
P_3 part. condensé	explicite	0.19	-11.22	2245	5.s
P_3 part. condensé	symplectique	0.11	-11.71	2665	15.s
P_2 standard	explicite	0.19	-11.20	14965	427.s
P_2 standard	symplectique	0.28	-11.22	13613	229.s
P_2 standard	implicite	0.19	-11.21	18625	404.s
P_2 standard	implicite	0.16	-11.20	15665	353.s
P_2 standard	implicite	0.13	-11.20	14281	371.s
P_2 condensé	explicite	0.26	-11.21	29009	22.s
P_2 condensé	symplectique	0.38	-11.21	15697	7.s

TAB. 6.15 – Coût du calcul (en temps et nombre de degrés de liberté) pour atteindre une erreur (en norme $L^2(\Omega)$) inférieure à $e^{-11.20}$

ln(erreur) < -15.20					
Disc. en espace	Disc. en temps	nombre CFL	ln(erreur)	NTDDL	temps CPU
P_6 standard	explicite	0.055	-15.43	1201	8.s
P_6 condensé	explicite	0.0052	-15.79	2865	57.s
P_5 standard	explicite	0.097	-15.68	2521	31.s
P_5 standard	symplectique	0.056	-15.75	2521	139.s
P_5 condensé	explicite	0.058	-15.96	4285	10.s
P_5 condensé	symplectique	0.033	-16.13	4285	46.s
P_5 part. condensé	explicite	0.097	-15.67	2521	16.s
P_5 part. condensé	symplectique	0.056	-15.67	2521	73.s
P_4 standard	explicite	0.072	-15.32	4705	148.s
P_4 standard	symplectique	0.077	-15.79	5513	494.s
P_4 standard	implicite	0.42	-15.19	9385	94.s
P_4 standard	implicite	0.27	-15.20	4705	37.s
P_4 standard	implicite	0.21	-15.42	4705	49.s
P_4 standard	implicite	0.10	-15.38	4705	100.s
P_4 condensé	explicite	0.079	-15.50	6433	11.s
P_4 condensé	symplectique	0.084	-15.51	6433	29.s
P_4 part. condensé	explicite	0.072	-15.46	4705	111.s
P_4 part. condensé	symplectique	0.077	-15.28	4705	257.s
P_3 standard	explicite	0.19	-15.22	16381	521.s
P_3 standard	symplectique	0.11	-15.19	16381	923.s
P_3 condensé	explicite	0.24	-15.27	26817	26.s
P_3 condensé	symplectique	0.13	-15.39	25173	46.s
P_3 part. condensé	explicite	0.19	-15.28	16381	491.s
P_3 part. condensé	symplectique	0.11	-15.26	16381	884.s

TAB. 6.16 – Coût du calcul (en temps et nombre de degrés de liberté) pour atteindre une erreur (en norme $L^2(\Omega)$) inférieure à $e^{-15.20}$

qui optimise d'autant le profil de la matrice de masse.
Il reste alors à se demander laquelle des discrétisations en temps est la plus efficace. Cela semble dépendre fortement de l'ordre du schéma. Pour le schéma d'ordre 2, nous avons tendance à donner un avantage à la discrétisation en temps implicite : même si le schéma P_1 standard-implicite est légèrement plus lent que le P_1 condensé-symplectique, il a le mérite de nécessiter bien moins de degrés de liberté. Pour le schéma d'ordre 3, c'est la discrétisation en temps symplectique qui s'impose nettement. À partir des schémas d'ordre 4 ce sont les schémas en temps explicites qui semblent être les plus avantageux. En regardant de plus près il suffit de se rendre compte qu'à partir de cet ordre, les schémas à discrétisation en temps explicite sont si peu dissipatifs qu'il n'est pas nécessaire d'augmenter le nombre de degrés de liberté pour atteindre les même précisions que les schémas à discrétisation en temps symplectiques. Ainsi pour les schémas d'ordre 4, la discrétisation en temps explicite est plus avantageuse essentiellement grâce au fait que l'on peut utiliser un nombre CFL plus élevé par rapport à la discrétisation en temps symplectique (quelle que soit la discrétisation en espace) et pour les schémas d'ordre 5 et 6 ceci vient essentiellement du fait que, ne disposant pas de discrétisation en temps symplectique d'ordre 5, le schéma d'ordre 5 est utilisé avec une discrétisation en temps d'ordre 6 qui n'est autre que la composée de la discrétisation en temps d'ordre 4 par elle-même. Ainsi le nombre de pas intermédiaires d'une itération en temps est largement supérieur pour les schémas d'ordre 5 et 6 à discrétisation en temps symplectique par rapport à celui des schémas d'ordre 5 et 6 à discrétisation en temps explicite.

6.1.4 Cas test des tourbillons co-rotatifs

Nous allons à présent nous intéresser à un cas test plus physique, à savoir le cas test des tourbillons co-rotatifs, pour mettre en évidence les limites de l'efficacité de la montée en ordre des éléments finis. En effet jusqu'à présent nous n'avons testé l'efficacité de nos schémas que sur la propagation d'une onde plane. Ceci est tout à fait légitime dans la mesure où toute onde se propageant n'est qu'une superposition d'ondes planes. Nous avons alors mis en évidence le gain que l'on peut avoir en augmentant l'ordre des schémas numériques, en terme de dissipation et dispersion, ce qui s'est traduit par le fait de pouvoir atteindre des précisions très supérieures avec un nombre bien inférieur de degrés de liberté. Il serait toutefois faux de croire que ceci reste vrai quel que soit le cas test que l'on considère. La figure 6.7 représente le terme source initial, qui évolue dans le temps par rotation autour de l'origine du domaine, et génère les tourbillons co-rotatifs dont la solution exacte et numérique au temps $t = 150$ sont représentées dans la figure 6.8.

Nous avons testé nos schémas d'ordre 2 à 6, à discrétisation en espace par éléments finis standards couplés avec une discrétisation en temps d'ordre arbitrairement élevé, en se fixant un nombre de degrés de liberté d'approximativement 15400 sur le domaine pour chaque schéma.
Le nombre exact de degrés de liberté pour chacun des 5 schémas est donné dans le tableau (6.17) et les pick-points au point de coordonnées $(50, 0)$ sont représentés dans la figure 6.9.

Nous remarquons alors qu'il n'y a guère que le schéma d'ordre 6 (et d'ordre 5 dans

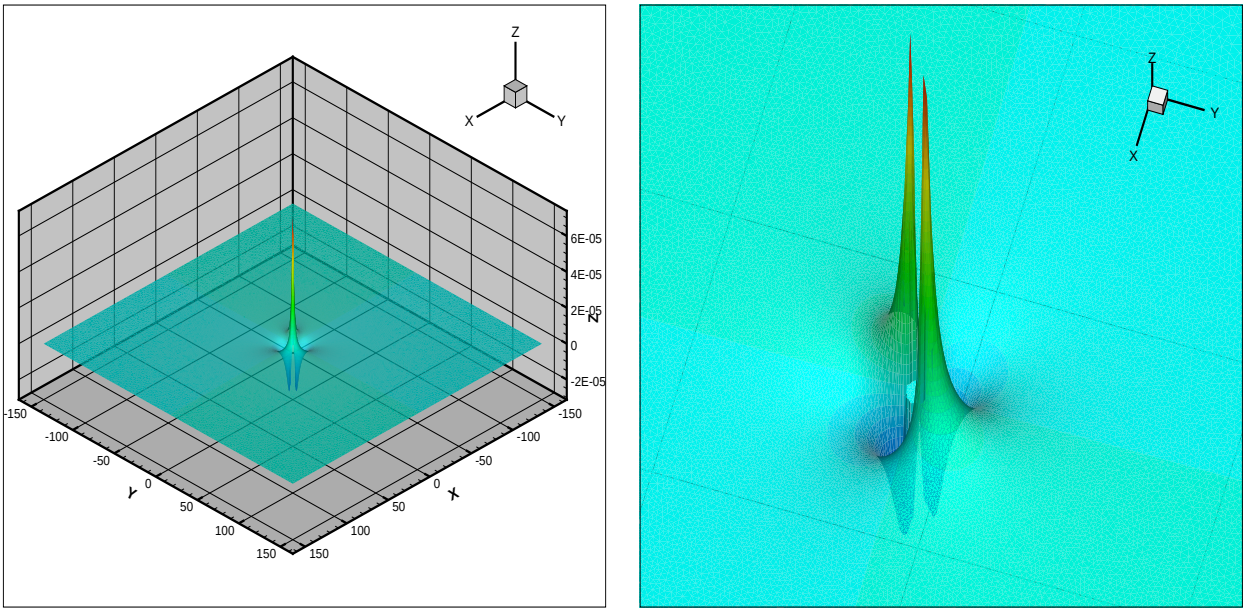

FIG. 6.7 – Force imposée générant les tourbillons co-rotatifs.

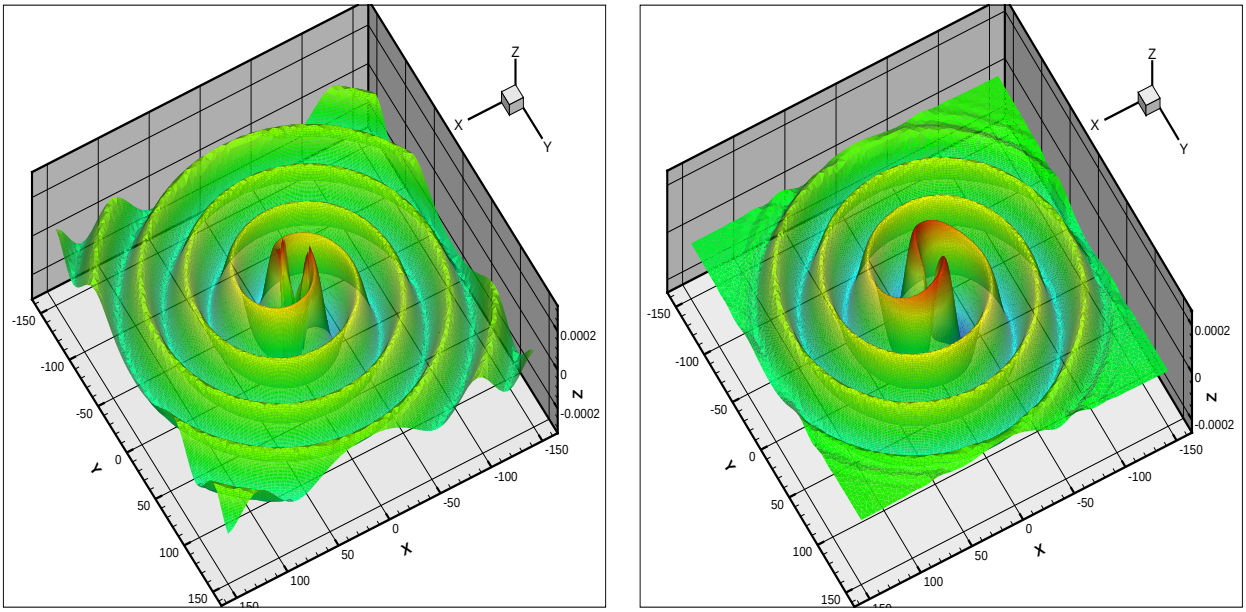

FIG. 6.8 – Solution exacte et numérique au temps t=150.

Ordre théorique du schéma	2	3	4	5	6
nombre de degrés de liberté	15425	15401	15313	15449	15461

TAB. 6.17 – Nombre de degrés de liberté utilisés par schéma pour le cas test des tourbillons co-rotatifs.

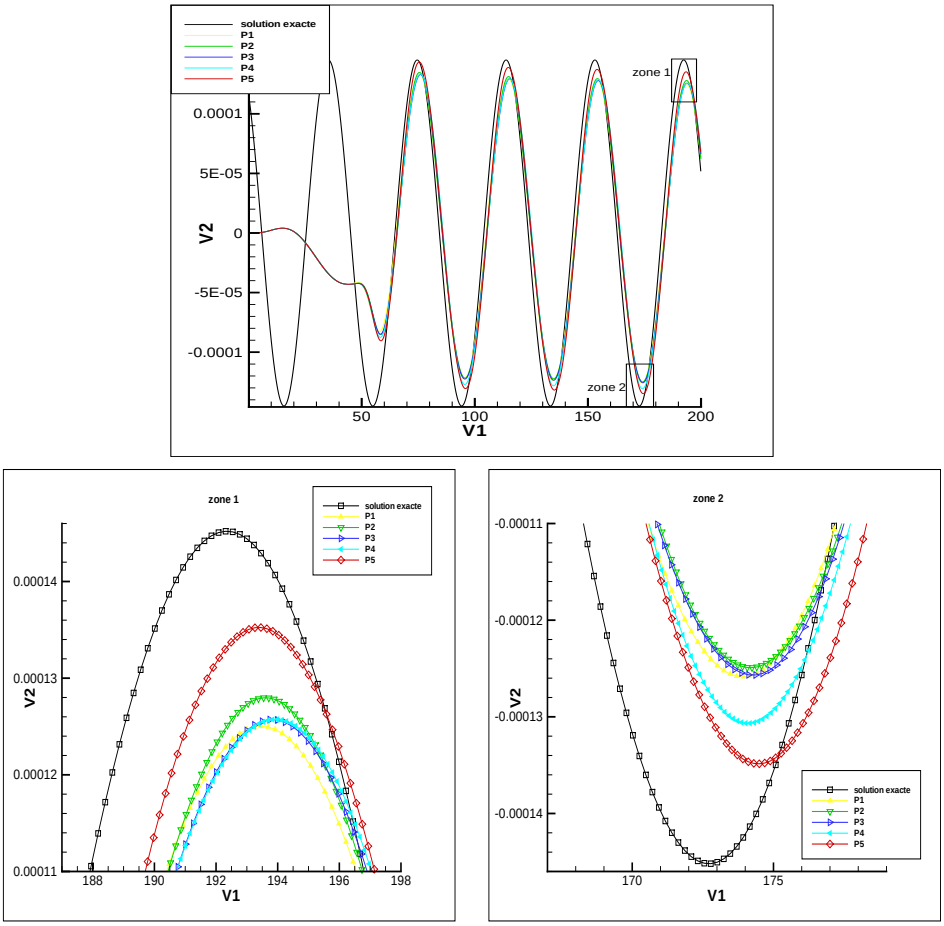

FIG. 6.9 – Pick-point au point de coordonnée $(50, 0)$ de la solution exacte et des solutions numériques.

une moindre mesure) qui semble s'approcher de la solution exacte, mais l'amélioration de la solution est bien loin de ce que l'on aurait pu espérer à la vue de l'amélioration de la propagation d'une onde plane qui accompagne la montée en ordre des schémas. Ceci s'explique de la manière suivante : en général une onde qui se propage est générée par un terme source. L'onde générée peut alors mieux se propager avec des schémas d'ordre élevé, mais si celle-ci est mal générée, la solution numérique n'a pas de raison d'être plus proche de la solution exacte. Le problème est ici clairement un problème de projection de la force imposée sur l'espace de discrétisation. À un nombre de degrés de liberté fixé, le fait d'approcher une force imposée par des polynômes de degré plus élevé n'améliore l'approximation que si cette force est assez régulière. Or en regardant la figure 6.7 nous remarquons que la force imposée générant les tourbillons co-rotatifs admet de très fortes variations de gradient. Pour remédier à ce problème il n'y pas d'autre choix que d'améliorer la projection de la force imposée, par exemple en raffinant le maillage dans la zone où cette force admet les plus fortes variations de gradient. Notons que c'est essentiellement pour ce type d'application que nous avons développé la méthode deux échelles, exposée dans le chapitre 7 de ce document, dans le cadre de la résolution des équations de Maxwell, mais pourrait très bien être adaptée à l'équation des ondes.

6.2 Schémas adaptés à la résolution des équations de Maxwell

Pour tester l'efficacité des schémas numériques adaptés à la résolution des équations de Maxwell nous allons à nouveau nous focaliser sur la propagation d'une onde plane (nous préciserons plus loin ce que l'on entend par onde plane dans le cadre des équations de Maxwell).
Nous allons tout d'abord déterminer les limites de stabilité des différents schémas, puis vérifier que les ordres numériques de convergence sont cohérents avec les ordres théoriques. Nous n'allons cette fois plus faire une étude aussi exhaustive de la dissipation et de la dispersion des différents schémas, l'étude du rapport coût/précision étant déjà des plus significatives sur les bonnes ou mauvaises propriétés conjointes de dissipation et dispersion des schémas.

6.2.1 Éléments finis rectangulaires sur maillage structuré

Nous considérons dans cette sous-section les discrétisations en espace par éléments finis d'arête rectangulaires définis dans la sous-section 3.2.1 et leur versions condensées définies dans la sous-section 4.5.1.
Dans un premier temps nous déterminons les limites de stabilité en propageant une onde plane à travers la diagonale du domaine $\Omega = [0,1] \times [0,1]$. Cela signifie que l'on résout le

Ordre théorique de la discrétisation	1	2	3	4	5
Discrétisation en temps d'ordre arbitraire	0.20	0.18	0.093	0.10	0.038
Discrétisation en temps symplectique	0.40	0.18	0.13	0.057	0.041

TAB. 6.18 – Nombres CFL optimaux pour les discrétisations en espace standards couplés aux discrétisations d'ordre arbitrairement élevé et symplectiques en temps.

Ordre théorique de la discrétisation	1	2	3	4	5
Discrétisation en temps d'ordre arbitraire	0.35	0.28	0.14	0.14	0.053
Discrétisation en temps symplectique	0.70	0.28	0.20	0.082	0.056

TAB. 6.19 – Nombres CFL optimaux pour les discrétisations condensées en espace couplés aux discrétisations d'ordre arbitrairement élevé et symplectiques en temps.

problème suivant :

$$\left\{\begin{array}{rcl} \dfrac{\partial \overrightarrow{E}}{\partial t} - \overrightarrow{\nabla} \times B & = & 0 \\ \dfrac{\partial B}{\partial t} + \nabla \times \overrightarrow{E} & = & 0 \\ \overrightarrow{E}(x,y,0) & = & \begin{pmatrix} \frac{\alpha w}{2k} \sin(k(x-y)) \\ \frac{\alpha w}{2k} \sin(k(x-y)) \end{pmatrix} \\ B(x,y,0) & = & \alpha \sin(k(x-y)) \end{array}\right. \tag{6.1}$$

avec $w = \sqrt{2}k$, et dont la solution est donnée par

$$\begin{array}{rcl} \overrightarrow{E}(x,y,t) & = & \begin{pmatrix} \frac{\alpha w}{2k} \sin(k(x-y) - wt) \\ \frac{\alpha w}{2k} \sin(k(x-y) - wt) \end{pmatrix}, \\ B(x,y,t) & = & \alpha \sin(k(x-y) - wt). \end{array}$$

Si l'on a choisi de propager l'onde plane suivant la diagonale du domaine Ω et non plus suivant l'un ou l'autre des axes ce n'est pas tant pour déterminer les limites de stabilité de nos schémas que pour en déterminer les ordres numérique de convergence : propager l'onde plane suivant l'un des axes reviendrait à privilégier la direction dans laquelle l'espace de discrétisation polynomial est de degré le plus élevé, et donc s'exposer à une super convergence.

Les nombres CFL optimaux, c'est-à-dire les nombres CFL les plus élevés nous permettant de propager l'onde plane au-delà d'une centaine de périodes sans que celle-ci n'explose sont donnés par type de discrétisation en temps dans le tableau 6.18 pour les discrétisation en espace par éléments finis d'arête standards, et dans le tableau 6.19 pour les discrétisation en espace par éléments finis d'arête condensés.

Ordre théorique de la discrétisation	1	2	3	4	5
Discrétisation en temps d'ordre arbitraire	0.94	1.97	2.99	3.99	4.99
Discrétisation en temps symplectique	1.01	2.00	2.99	3.99	4.99

TAB. 6.20 – Ordre numérique de convergence pour les schémas à discrétisations en espace standards.

Ordre théorique de la discrétisation	1	2	3	4	5
Discrétisation en temps d'ordre arbitraire	0.94	1.99	2.99	3.99	4.99
Discrétisation en temps symplectique	1.00	2.00	2.99	3.99	4.99

TAB. 6.21 – Ordre numérique de convergence pour les schémas à discrétisations condensées en espace.

Les ordres de convergence sont déterminés sur le même cas test au bout d'une période (théorique) de propagation et listés dans le tableau 6.20 pour les schémas à discrétisation en espace par éléments finis d'arête standards et dans le tableau 6.21 pour les schémas à discrétisation en espace par éléments finis d'arête condensés. Nous ne donnons ici que les ordres de convergence du champ magnétique, les erreurs étant considérées en norme $L^2(\Omega)$, en ayant vérifié que ceux du champ électrique ne diffèrent que très peu de ceux-ci et uniquement pour les schémas d'ordre 1 et 2.

Nous allons donc passer directement au rapport coût/précision des différents schémas en déterminant le nombre total de degrés de liberté pour le champ électrique et pour le champ magnétique (abrévié respectivement NTDDL_E et NTDDL_B) et le temps de calcul nécessaire pour atteindre une erreur (sur le champ magnétique en norme $L^2(\Omega)$) inférieure à un seuil que l'on se fixe, au bout de 10 périodes (théoriques) de propagation de l'onde plane. Les résultats sont consignés dans les tableaux 6.22 à 6.25 pour des seuils d'erreur allant de $e^{-6.60}$ à $e^{-20.50}$. Dans ces tableaux et dans les remarques les accompagnant nous désignerons de nouveau par discrétisations en temps explicites les discrétisation en temps explicites d'ordre arbitrairement élevé bien que les discrétisations en temps symplectiques soient elles aussi explicites.

Nous faisons alors remarquer à nouveau la supériorité, en terme de nombre de degrés de liberté des schémas d'ordre élevé : par exemple il ne faut que 4900 degrés de liberté aux schémas d'ordre 5 pour atteindre une erreur inférieure à $e^{-20.50}$ sur le champ magnétique, alors qu'il en faut au minimum 19600 aux schémas d'ordre 4. Mais cela ne signifie pas forcement une diminution du temps de calcul : le schéma à discrétisation condensée d'ordre 4 en espace couplée avec la discrétisation explicite en temps ne nécessite que 134 secondes pour atteindre une erreur inférieure à $e^{-20.50}$ alors qu'il en faut 195 au schéma à discrétisation en espace standard d'ordre 5 couplée avec la discrétisation explicite en temps. Nous remarquons à nouveau le gain résultant de l'utilisation d'éléments finis d'arête condensés par rapport aux éléments finis d'arête standards, et ceci pour tous les ordres.
La comparaison entre les discrétisations en temps est toutefois sensiblement dépendante de

ln(erreur) < -6.60					
Disc. en espace	Disc. en temps	nombre CFL	ln(erreur_E) ln(erreur_B)	NTDDL_E NTDDL_B	temps CPU
Standard d'ordre 1	explicite	0.20	-6.95 -6.60	26912 13456	375.s
Standard d'ordre 1	symplectique	0.40	-7.03 -6.66	800 400	<1.s
Condensée d'ordre 1	explicite	0.35	-6.94 -6.60	81608 40804	50.s
Condensée d'ordre 1	symplectique	0.70	-7.11 -6.63	162 81	<1.s
Standard d'ordre 2	explicite	0.18	-7.15 -6.80	288 144	<1.s
Standard d'ordre 2	symplectique	0.18	-6.96 -6.62	200 100	<1.s
Condensée d'ordre 2	explicite	0.28	-6.96 -6.61	648 324	<1.s
Condensée d'ordre 2	symplectique	0.28	-7.13 -6.78	392 196	<1.s

TAB. 6.22 – Coût du calcul (en temps et nombre de degrés de liberté) pour atteindre une erreur (sur le champ magnétique en norme $L^2(\Omega)$) inférieure à $e^{-6.60}$

ln(erreur) < -11.50					
Disc. en espace	Disc. en temps	nombre CFL	ln(erreur_E) ln(erreur_B)	NTDDL_E NTDDL_B	temps CPU
Standard d'ordre 2	explicite	0.18	-11.86 -11.50	38088 19044	621.s
Standard d'ordre 2	symplectique	0.18	-11.88 -11.51	20000 10000	123.s
Condensée d'ordre 2	explicite	0.28	-11.85 -11.50	38088 19044	98.s
Condensée d'ordre 2	symplectique	0.28	-11.86 -11.51	46208 23104	26.s
Standard d'ordre 3	explicite	0.093	-12.08 -11.62	1458 729	1.s
Standard d'ordre 3	symplectique	0.13	-12.47 -11.78	1152 576	1.s
Condensée d'ordre 3	explicite	0.14	-12.05 -11.70	3042 1521	<1.s
Condensée d'ordre 3	symplectique	0.20	-11.47 -11.78	1152 576	<1.s

TAB. 6.23 – Coût du calcul (en temps et nombre de degrés de liberté) pour atteindre une erreur (sur le champ magnétique en norme $L^2(\Omega)$) inférieure à $e^{-11.50}$

ln(erreur) < -16.75					
Disc. en espace	Disc. en temps	nombre CFL	ln(erreur_E) ln(erreur_B)	NTDDL_E NTDDL_B	temps CPU
Standard d'ordre 3	explicite	0.093	-17.21 -16.76	45000 22500	1098.s
Standard d'ordre 3	symplectique	0.13	-17.45 -16.75	31752 15876	432.s
Condensée d'ordre 3	explicite	0.14	-17.10 -16.74	88200 44100	179.s
Condensée d'ordre 3	symplectique	0.20	-17.45 -16.75	31752 15876	26.s
Standard d'ordre 4	explicite	0.10	-17.34 -16.80	6272 3136	25.s
Standard d'ordre 4	symplectique	0.057	-17.43 -16.86	6272 3136	45.s
Condensée d'ordre 4	explicite	0.14	-17.29 -16.92	10368 5184	5.s
Condensée d'ordre 4	symplectique	0.082	-17.13 -16.75	9248 4624	9.s

TAB. 6.24 – Coût du calcul (en temps et nombre de degrés de liberté) pour atteindre une erreur (sur le champ magnétique en norme $L^2(\Omega)$) inférieure à $e^{-16.75}$

ln(erreur) < -20.50					
Disc. en espace	Disc. en temps	nombre CFL	ln(erreur_E) ln(erreur_B)	NTDDL_E NTDDL_B	temps CPU
Standard d'ordre 4	explicite	0.10	-21.12 -20.58	41472 20736	833.s
Standard d'ordre 4	symplectique	0.057	-21.09 -20.52	39200 19600	1303.s
Condensée d'ordre 4	explicite	0.14	-20.96 -20.58	61952 30976	134.s
Condensée d'ordre 4	symplectique	0.082	-20.93 -20.55	61952 30976	229.s
Standard d'ordre 5	explicite	0.038	-21.51 -20.82	9800 4900	195.s
Standard d'ordre 5	symplectique	0.041	-21.51 -20.82	9800 4900	360.s
Condensée d'ordre 5	explicite	0.053	-21.50 -20.81	9800 4900	24.s
Condensée d'ordre 5	symplectique	0.056	-21.51 -20.82	9800 4900	43.s

TAB. 6.25 – Coût du calcul (en temps et nombre de degrés de liberté) pour atteindre une erreur (sur le champ magnétique en norme $L^2(\Omega)$) inférieure à $e^{-20.50}$

l'ordre du schéma : si pour les schémas d'ordre 1, 2 et 3 ce sont les schémas à discrétisation en temps symplectiques qui sont nettement plus avantageux en terme de coût en nombre de degrés de liberté et en temps de calcul, pour les schémas d'ordre 4 et 5 ce n'est plus la cas. Les schémas à discrétisation en temps explicite deviennent à partir de l'ordre 4 si peu dissipatifs et dispersifs qu'il n'y a plus vraiment d'intérêt à conserver de manière exacte l'amplitude de l'onde plane via l'utilisation de discrétisations en temps symplectiques. Si pour les schémas d'ordre 4 le nombre de degrés de liberté nécessaire pour atteindre une erreur inférieure à $e^{-20.50}$ est sensiblement le même, que l'on utilise une discrétisation en temps explicite ou symplectique, il faut remarquer que la restriction sur le nombre CFL est plus sévère pour la discrétisation en temps symplectique, d'où le surcoût en temps de calcul. Pour les schémas d'ordre 5, la différence de temps de calcul entre l'utilisation de la discrétisation en temps explicite et symplectique, ne vient ni d'une restriction sur le nombre CFL plus sévère (cette restriction étant même moins forte pour la discrétisation en temps symplectique), ni d'un surcoût en nombre de degrés de liberté, mais de nouveau du fait qu'au-delà de l'ordre 3, seules les discrétisations symplectiques d'ordre pair (2n) sont connues, et ceci par composition de discrétisation symplectiques d'ordre inférieur (2n-2), ce qui fait considérablement augmenter le coût d'une itération en temps.

6.2.2 Éléments finis triangulaires

Nous considérons à présent dans cette sous-section les éléments finis d'arête triangulaires définis dans la sous-section 3.2.2.
Pour les éléments finis d'arête triangulaires nous allons simplement déterminer les restrictions sur le nombre CFL assurant la stabilité des schémas et vérifier les ordres de convergences, l'étude du rapport coût/précision nous ayant montré que le comportement des éléments finis d'arête triangulaires est en tout point similaire aux schémas à discrétisation en espace par éléments finis d'arête rectangulaires standards.
Nous propageons donc une onde plane, cette fois suivant l'axe des abscisses, c'est-à-dire que nous résolvons le problème suivant :

$$\begin{cases} \dfrac{\partial \overrightarrow{E}}{\partial t} - \overrightarrow{\nabla} \times B &= 0 \\ \dfrac{\partial B}{\partial t} + \nabla \times \overrightarrow{E} &= 0 \\ \overrightarrow{E}(x,y,0) &= \begin{pmatrix} 0 \\ \alpha \sin(kx) \end{pmatrix} \\ B(x,y,0) &= \alpha \sin(kx) \end{cases} \tag{6.2}$$

dont la solution est donnée par

$$\begin{aligned} \overrightarrow{E}(x,y,t) &= \begin{pmatrix} 0 \\ \alpha \sin(k(x-t)) \end{pmatrix}, \\ B(x,y,t) &= \alpha \sin(k(x-t)), \end{aligned}$$

Ordre théorique de la discrétisation	1	2	3	4	5
Discrétisation en temps d'ordre arbitraire	0.23	0.24	0.13	0.14	0.056
Discrétisation en temps symplectique	0.47	0.24	0.19	0.082	0.060

TAB. 6.26 – Nombres CFL optimaux pour les discrétisations en temps d'ordre arbitrairement élevé et symplectiques.

Ordre théorique de la discrétisation	1	2	3	4	5
Discrétisation en temps d'ordre arbitraire	0.93	1.99	2.99	3.99	5.34
Discrétisation en temps symplectique	1.00	2.00	2.99	4.05	5.02

TAB. 6.27 – Ordre numérique de convergence.

et où l'on prend soin d'adapter k au domaine de calcul considéré de manière à rester cohérent avec l'utilisation des conditions limites périodiques.

Sur le domaine de calcul $[-5,5] \times [-2.5,2.5]$ avec $k = 2\pi$, nous avons déterminé les limites de stabilité de nos schémas et consigné les résultats dans le tableau 6.26 pour les discrétisations en temps d'ordre arbitrairement élevé et les discrétisations en temps symplectiques.

Les ordres de convergences sont déterminés pour ces nombres CFL optimaux en propageant une onde plane de longueur d'onde 1 suivant l'axe des abscisses sur le domaine $[0,1] \times [0,1]$ pendant une seule période. Ces ordres sont consignés dans le tableau 6.27.

6.2.3 Couplage conforme des éléments finis rectangulaires et triangulaires

Nous allons à présent tester le couplage entre les éléments finis d'arête rectangulaires condensés et les éléments finis triangulaires. Pour ce faire nous allons propager une impulsion initiale dérivant d'une gaussienne sur un domaine circulaire avec la condition aux limites absorbante de Silver-Müller, une première fois sur un maillage hybride, puis sur une triangulation de ce domaine. Les maillages que l'on s'est fixés sont donnés dans la figure 6.10 et génèrent un nombre sensiblement égal, à un ordre de schéma fixé, de degrés de liberté, ces nombres étant donnés dans le tableau 6.28.

Les figures 6.11 et 6.12 représentent les projections des impulsions initiales sur les espaces de discrétisation associés à l'utilisation des schémas de plus bas degré respectivement pour les éléments finis d'arête couplés sur le maillage hybride et pour les éléments finis d'arête triangulaires sur la triangulation.

Nous nous intéressons dans un premier temps à l'évolution temporelle de la première

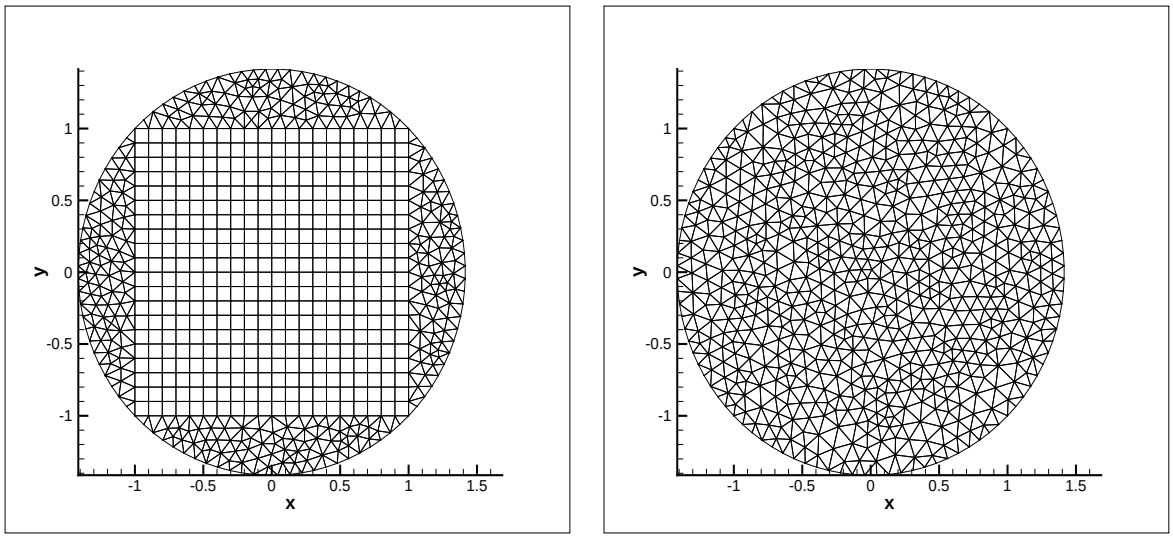

FIG. 6.10 – Maillage hybride et triangulation du domaine.

Ordre théorique de la discrétisation	1	2	3	4	5
Nombre de DDL pour le champ électrique :					
pour le maillage hybride	1795	6450	13965	24340	37575
pour la triangulation	2015	6650	13905	23780	36275
Nombre de DDL pour le champ magnétique :					
pour le maillage hybride	1030	3490	7380	12700	19450
pour la triangulation	1310	3930	7860	13100	19650

TAB. 6.28 – Nombre de degrés de liberté par ordre de schéma générés par les éléments finis d'arête couplés sur le maillage hybride et par les éléments finis d'arête triangulaires sur la triangulation.

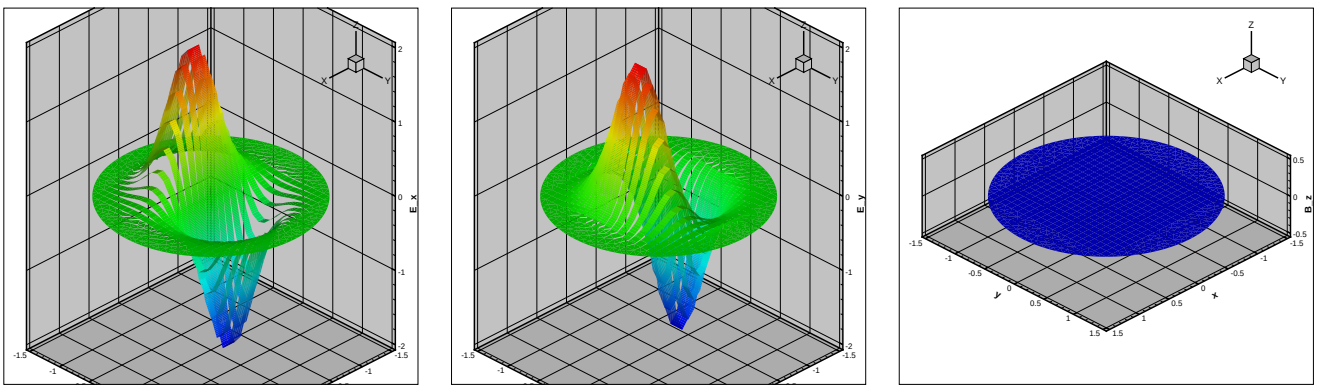

FIG. 6.11 – Impulsion initiale avec les éléments finis d'arête de plus bas degré sur le maillage hybride.

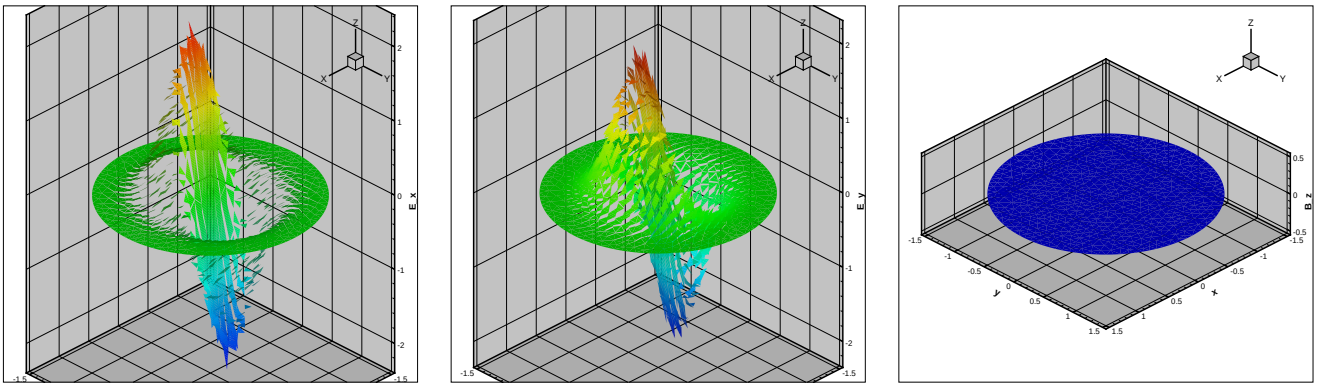

FIG. 6.12 – Impulsion initiale avec les éléments finis d'arête de plus bas degré sur le maillage non-structuré.

composante du champ électrique au point $(0.5, 0.5)$, c'est-à-dire en un point intérieur à la zone maillée par les éléments rectangulaires pour le maillage hybride, d'une part pour la solution numérique issue des éléments finis couplés (figure 6.13), d'autre part pour la solution numérique issue des éléments finis triangulaires (figure 6.14).

Remarquons que le comportement des schémas à discrétisation en espace par éléments finis d'arête couplés est sensiblement identique au comportement des schémas à discrétisation en espace par éléments finis d'arête triangulaire, à savoir que pour les schémas d'ordre 1, 2 et 3 les solutions sont bien distinctes même si globalement leur profil sont identiques, alors qu'a partir des schémas d'ordre 3 il devient impossible de les distinguer. Nous n'avons pas superposé à un ordre de schéma fixé la solution issue du schéma à discrétisation en espace par éléments finis d'arête couplés et la solution issue du schéma à discrétisation en espace par éléments finis triangulaires parce que dans la pratique celles-ci sont à nouveau indiscernables à partir des schémas du troisième ordre.
Nous avons aussi vérifié que l'évolution temporelle du signal au point $(1.3, 0)$, c'est-à-dire cette fois en un point extérieur à la zone maillée par les éléments rectangulaire pour le maillage hybride, nous permettait de tirer les mêmes conclusions qu'au point $(0.5, 0.5)$.

Remarque 6.2.1. *Si le fait que les schémas à discrétisation en espace par éléments finis d'arête couplés permettent "visiblement" de propager l'impulsion initiale correctement n'est pas suffisant pour valider la méthode, nous avons aussi vérifié avec succès sur la propagation d'une onde plane que les ordres de convergence de ces schémas sur un maillage hybride sont cohérents.*

Rappelons que l'intérêt d'utiliser des schémas à discrétisation en espace par éléments finis d'arête couplés est de permettre une condensation partielle de la matrice de masse. Le tableau 6.29 nous donne les temps de calcul des schémas à discrétisation en espace par éléments finis d'arête couplés et des schémas à discrétisation en espace par éléments finis d'arête triangulaires par ordre de schéma.

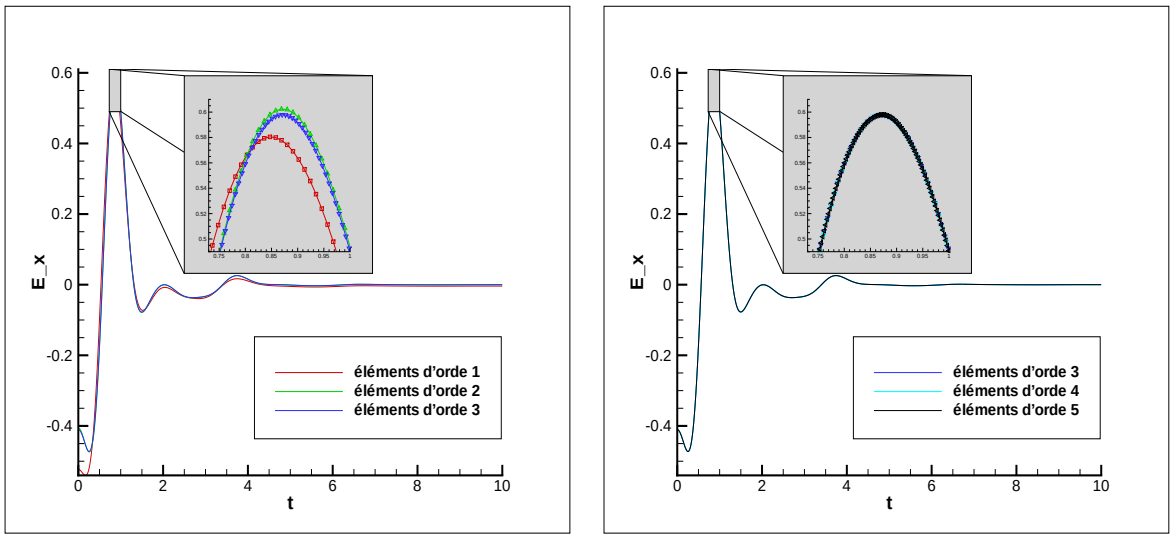

FIG. 6.13 – Pick point de la première composante du champ électrique au point de coordonnée $(0.5, 0.5)$ pour les éléments finis couplés.

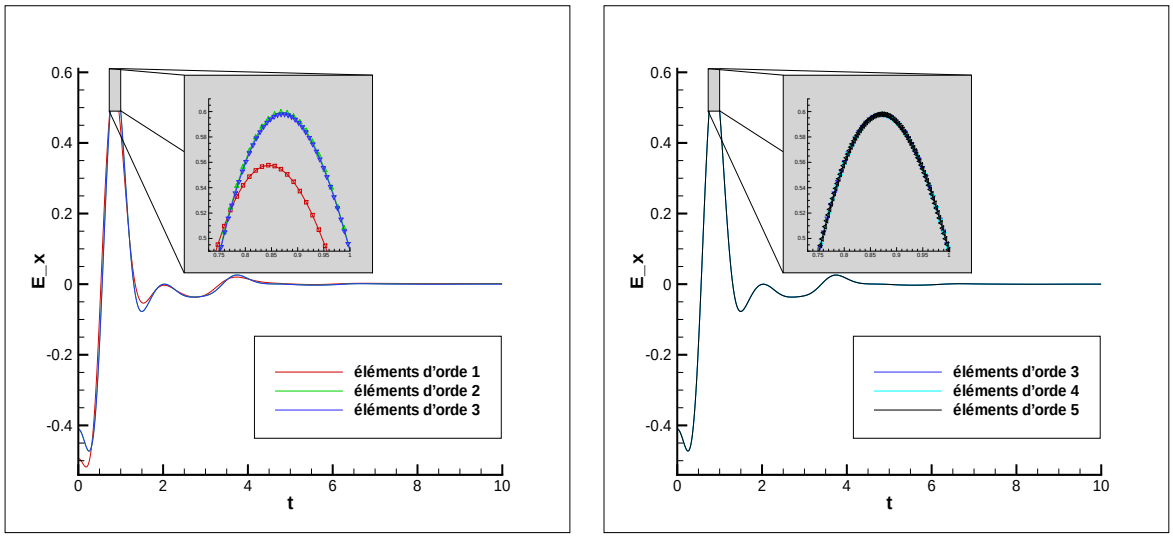

FIG. 6.14 – Pick point de la première composante du champ électrique au point de coordonnée $(0.5, 0.5)$ pour les éléments finis d'arête triangulaires.

Ordre théorique de la discrétisation	1	2	3	4	5
Temp CPU pour les éléments finis couplés	<1.s	5.s	32.s	88.s	653.s
Temp CPU pour les éléments finis triangulaires	<1.s	7.s	43.s	116.s	827.s

TAB. 6.29 – Temps de calcul comparés des schémas à discrétisation en espace par éléments finis d'arête couplés et triangulaires.

Nous remarquons alors une nette diminution des temps de calcul liée à l'utilisation des éléments finis couplés. Ce gain est tout à fait comparable à celui que l'on a en utilisant des éléments finis de Lagrange partiellement condensés plutôt que les éléments finis de Lagrange standards pour la résolution de l'équation des ondes. Remarquons que le gain pourrait être encore nettement amélioré en optimisant la zone maillée par les éléments rectangulaires, c'est-à-dire en ajustant au mieux la zone discrétisée par le maillage cartésien dans le domaine Ω.

6.2.4 Comparaison éléments finis conformes-Galerkin discontinus

Nous présentons à présent les résultats issus de la comparaison entre la résolution des équations de Maxwell par nos schémas d'éléments finis conformes et le résolution des équations de Maxwell par une méthode de type Galerkin discontinus. Dans la pratique le code utilisant une discrétisation en espace de type Galerkin discontinus est un code développé à l'"Institut für Aerodynamik und Gasdynamik" de Stuttgart par l'équipe de C.-D Munz dont on peut trouver une description dans [35] ou [34], résolvant les équations de l'acoustique, mais on peut montrer qu'il est algébriquement équivalent de résoudre les équations de Maxwell en utilisant des éléments finis conformes dans H(rot) et de résoudre les équations de l'acoustique en utilisant des éléments finis conformes dans H(div).
Nous reprenons le cas test décrit dans la sous-section 6.2.2 de la propagation d'une onde plane suivant la direction (Ox) sur le domaine $\Omega = [-5, 5] \times [2.5, 2.5]$ à travers des bords périodiques. Le tableau 6.30 consigne les résultats obtenus après 100 périodes théoriques de la propagation d'une onde plane de nombre d'onde égal à dix sur le maillage représenté dans la figure 6.15 pour les schémas d'ordre 2 à 6. Les erreurs sont données en norme $L^2(\Omega)$ sur la composante magnétique et sur les deux composantes du champ électrique. Les notations FE et DG font références respectivement aux résultats issus de la méthode d'éléments finis conforme (Finite Element) et issus de la méthode Galerkin discontinus (Discontinuous Galerkin), NTDDL_E et NTDDL_B désignent le nombre total de degrés de libertés respectivement pour le champ électrique et magnétique.

Il faut remarquer que les résultats sont conformes à l'intuition que l'on peut avoir des avantages et inconvénients des deux méthodes. En terme de norme d'erreur le gain lié à la conformité de la méthode des éléments finis est indéniable. Par exemple sur le champ magnétique, si l'erreur est identique pour les deux méthodes pour les schémas du second ordre (2.50E-2 et 2.49E-2), pour les schémas d'ordre 6 cette erreur devient quasiment dix fois plus petite pour les schémas à discrétisation en espace par éléments finis conforme (1.17E-5) que pour les schémas à discrétisation en espace du type Galerkin discontinus (1.12E-4), alors que le nombre de degrés de liberté est globalement moindre. Remarquons que si le gain sur la composante E_y du champ électrique, c'est-à-dire la composante sur laquelle l'onde est vraiment propagée, est encore plus net, il l'est moins sur la composante E_x, composante sur laquelle la solution exacte est nulle, et même en défaveur des éléments finis conformes pour les schémas d'ordre 2 et 3. Pour expliquer ceci il faut souligner que lorsque l'on utilise des éléments finis d'arête, les deux composantes du champ électrique sont liés par la contrainte de conformité alors qu'avec la méthode Galerkin discontinue les composantes du champ électrique, bien que liées par les équations de Maxwell, sont définies

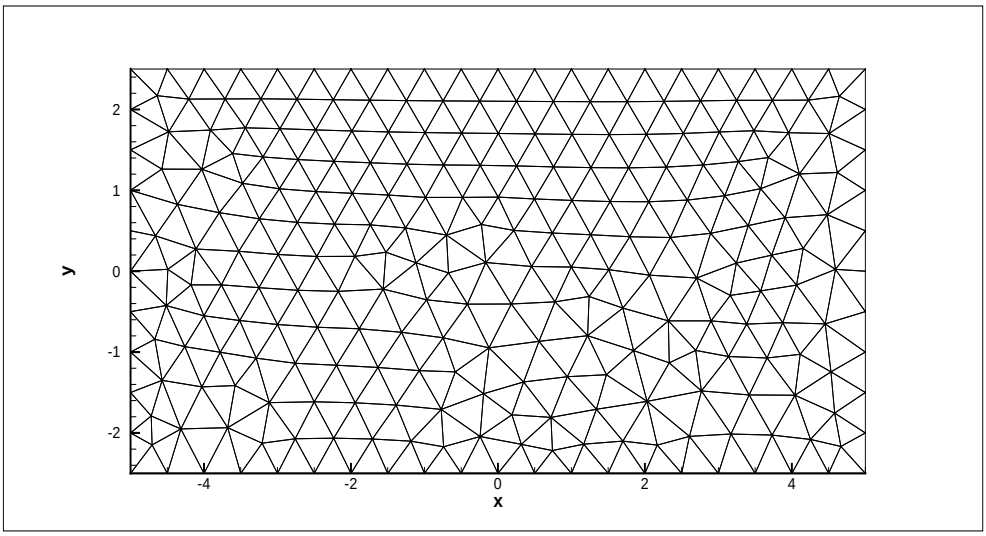

FIG. 6.15 – Maillage utilisé pour la comparaison éléments finis conformes-Galerkin discontinus.

Ordre théorique du schéma		2	3	4	5	6
NTDDL_E	FE	2290	4809	8244	12595	17862
	DG	2748	5496	9160	13740	19236
erreur E_x	FE	1.10E-3	1.12E-3	2.15E-4	3.10E-5	7.44E-6
	DG	4.06E-4	8.06E-5	4.44E-4	1.22E-4	1.19E-5
erreur E_y	FE	2.50E-2	1.51E-2	2.97E-4	4.00E-5	7.51E-6
	DG	2.49E-2	2.50E-2	1.27E-3	1.02E-3	1.13E-4
NTDDL_B	FE	1374	2748	4580	6870	9618
	DG	1374	2748	4580	6870	9618
erreur B	FE	2.50E-2	1.53E-2	3.61E-4	3.77E-5	1.17E-5
	DG	2.49E-2	2.50E-2	1.29E-2	1.05E-3	1.12E-4
pas de temps	FE	7.11E-2	3.82E-2	4.31E-2	1.64E-2	2.31E-2
	DG	2.87E-2	1.72E-2	1.23E-2	9.56E-3	7.83E-3
temps CPU	FE	4.s	31.s	88.s	668.s	951.s
	DG	19.s	47.s	121.s	247.s	478.s

TAB. 6.30 – Comparaison de l'efficacité et du coût de schémas d'éléments finis conforme et Galerkin discontinus.

indépendamment l'une de l'autre. Par exemple la projection, sur l'espace de discrétisation issu de l'utilisation d'éléments finis conformes, de la condition initiale du cas test de la propagation de l'onde plane que l'on considère, portant sur le champ électrique, n'est pas identiquement nulle sur sa première composante alors que dans le cadre de l'utilisation de la méthode de Galerkin discontinue elle l'est. Si l'espace de discrétisation n'est pas assez fin, la contrainte de conformité peut donc se révéler être un facteur limitant dans la résolution des équations de Maxwell.
Au niveau des temps de calculs il faut toutefois donner un net avantage aux discrétisations en espace de type Galerkin discontinues. Ces méthodes sont bien plus efficaces dès que l'espace de discrétisation devient grand, soit par raffinement du maillage, soit comme nous le vérifions ici par augmentation de l'ordre de la discrétisation. En effet les méthodes de type Galerkin discontinues ne nécessitent pas l'inversion d'une matrice de masse globale mais d'un grand nombre de matrices de masses locales (ce nombre étant proportionnel au nombre d'éléments du maillage), ce qui se révèle plus efficace. Il faut toutefois relativiser ceci en précisant que pour cette comparaison nous avons utilisé les éléments finis d'arête triangulaires que l'on ne sait pas condenser et qu'il nous serait possible d'accélérer la résolution via l'utilisation d'éléments finis d'arête rectangulaires condensés (ce qui n'est pas le but de cette comparaison).
Ce qui n'apparaît pas ici, et qu'il faut aussi mettre en faveur des méthodes de type Galerkin discontinues, c'est que, dans la mesure où la propagation d'onde est simulée par un calcul de flux à travers les arêtes du maillage, il n'y a besoin que de stocker des matrices locales, ces dernières pouvant même être calculées à la volée à partir des matrices déterminées sur un éléments de référence. Le coût en terme de stockage est donc très largement moindre pour ces méthodes que pour les méthodes d'éléments finis conformes.

Chapitre 7

Résolution deux échelles des équations de Maxwell

Dans un certain nombre d'applications la simulation d'une propagation d'ondes nécessite une résolution précise des équations régissant cette propagation dans certaines régions du domaine. C'est le cas notamment lorsqu'une source générant une onde est localisée dans une région notablement plus petite que le domaine de calcul. Si la source oscille à une fréquence élevée, l'onde générée aura une longueur d'onde courte, tandis que si la fréquence d'oscillation de la source est faible, l'onde générée aura une longueur plus élevée. C'est pourquoi si la source oscille à une faible fréquence il est possible d'utiliser un maillage plus grossier pour la simulation que si la source oscillait à une fréquence plus élevée tout en gardant une précision comparable. Il faut toutefois nuancer ce principe dans la mesure où si l'on utilise un maillage trop grossier, la source elle-même sera mal capturée (typiquement cela signifie que la projection de la source sur l'espace de discrétisation sera une mauvaise approximation de la source), alors que l'onde générée aurait été assez lisse (nous emploierons ce terme pour désigner des fonctions ayant de faibles variations de gradients) pour être convenablement approchée sur l'espace de discrétisation (c.f. le cas test des tourbillons co-rotatifs, section 6.1.4). Pour ce type d'application il est donc préférable d'utiliser un maillage fin au voisinage de la source et un maillage plus grossier dans le reste du domaine.
Notre objectif étant de développer un algorithme deux échelles pour la résolution des équations de Maxwell, nous avons choisi d'étendre la méthode développée par Glowinski et associés [41] pour la résolution d'un laplacien à l'aide d'éléments finis de Lagrange, à l'équation de Maxwell du second ordre à l'aide d'éléments finis d'arête.

7.1 Problème continu : le cas stationnaire

Introduisons dans un premier temps un certain nombre de notations. Soit $\Omega \subset \mathbb{R}^2$ un ouvert borné de $\mathbb{R}^2$ de frontière $\Gamma = \partial\Omega$ suffisamment régulière et soit $\omega \subset \Omega$ un second ouvert de frontière γ. Notons $\overrightarrow{n}$ le vecteur normal unitaire sortant de ω sur γ.

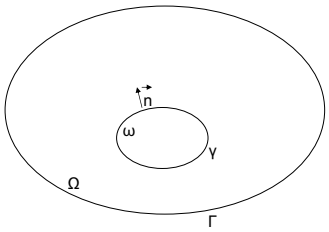

FIG. 7.1 – Domaine.

Dans ce qui suit nous considérerons le problème suivant : trouver $U \in H(rot, \Omega)$ tel que

$$\begin{cases} U + \nabla \times \nabla \times U = f \text{ in } \Omega, \\ U \times \overrightarrow{n} = 0 \text{ on } \Gamma. \end{cases} \tag{7.1}$$

où $f \in (L^2(\Omega))^2$ peut être décomposée en somme de deux fonctions $f_1, f_2 \in (L^2(\Omega))^2$,

$$f = f_1 + f_2$$

f_2 ayant un support inclus dans ω,

$$supp(f_2) \subset \omega.$$

Il faut bien entendu voir f_2 comme étant une source fine et f_1 comme une source de fond assez lisse. En gardant bien à l'esprit que $supp(f_2) \subset \omega \subset \Omega$ et que ω et un domaine nettement plus petit que Ω, nous introduisons les deux problèmes auxiliaires :
Trouver $V \in H(rot, \Omega)$ tel que

$$\begin{cases} V + \nabla \times \nabla \times V &= f_1 & \text{dans } \Omega \backslash \overline{\omega} \cup \omega, \\ V \times \overrightarrow{n} &= 0 & \text{sur } \Gamma, \\ \left[V\right] &= 0 & \text{sur } \gamma, \\ \left[(\nabla \times V) \times \overrightarrow{n}\right] &= -\lambda & \text{sur } \gamma \end{cases} \tag{7.2}$$

et trouver $W \in H(rot, \Omega)$ tel que

$$\begin{cases} W + \nabla \times \nabla \times W &= f_1 + f_2 & \text{dans } \omega, \\ W \times \overrightarrow{n} &= V \times \overrightarrow{n} & \text{sur } \gamma \end{cases} \tag{7.3}$$

où, $[\psi] = \psi^+ - \psi^-$ désigne le saut de ψ à travers γ, ψ^+ et ψ^- sont respectivement les restrictions de ψ sur $\Omega \backslash \overline{\omega}$ et ω. Remarquons que la solution W du problème (7.3) dépend de la solution du problème (7.2), elle-même dépendant de l'indéterminée λ. Nous cherchons alors à déterminer λ de manière à ce que la solution de (7.1) soit donnée par

$$U = V^+ . \chi(\Omega \backslash \overline{\omega}) + W . \chi(\omega),$$

où V et W sont les solutions respectives de (7.2) et (7.3). Cela signifie que l'on sera capable de définir la solution de notre problème initial comme étant la solution de (7.2) dans $\Omega \backslash \overline{\omega}$,

qui a pour vocation à être calculée sur un maillage grossier (sachant que la solution sera lisse dans cette région), et comme étant la solution de (7.3) dans ω, qui a pour vocation à être calculée sur un maillage fin (sachant que la source f_2 nécessite une résolution plus précise).
Sachant que $supp(f_2) \subset \omega$, et à la vue de (7.2) et (7.3), il suffit de remarquer que

Proposition 7.1.1. *Si λ est choisi tel que*

$$\lambda = (\nabla \times V^-) \times \overrightarrow{n} - (\nabla \times W) \times \overrightarrow{n}, \tag{7.4}$$

alors la solution U du problème initial (7.1) *est donnée par*

$$U = V^+ \chi(\Omega \backslash \overline{\omega}) + W \chi(\omega) \tag{7.5}$$

Démonstration. Soit $U \in H_0(rot, \Omega)$ vérifiant (7.5). Alors $\forall \phi \in H_0(rot, \Omega)$,

$$\begin{aligned} \int_\Omega U \cdot \phi \, dX + \int_\Omega (\nabla \times U)(\nabla \times \phi) \, dX = \\ \int_\Omega U \cdot \phi \, dX + \int_{\Omega \backslash \overline{\omega}} (\nabla \times V^+)(\nabla \times \phi) \, dX + \int_\omega (\nabla \times W)(\nabla \times \phi) \, dX \end{aligned} \tag{7.6}$$

avec

$$\int_{\Omega \backslash \overline{\omega}} (\nabla \times V^+)(\nabla \times \phi) \, dX = \int_{\Omega \backslash \overline{\omega}} (\nabla \times \nabla \times V^+) \cdot \phi \, dX + \int_{\Gamma \cup \gamma} (\nabla \times V^+ \times \overrightarrow{m}) \cdot \phi \, d\sigma \tag{7.7}$$

et

$$\int_\omega (\nabla \times W)(\nabla \times \phi) \, dX = \int_\omega (\nabla \times \nabla \times W) \cdot \phi \, dX + \int_\gamma (\nabla \times W \times \overrightarrow{n}) \cdot \phi \, d\sigma \tag{7.8}$$

où $\overrightarrow{m}$ est le vecteur normal unitaire sortant sur $\partial(\Omega \backslash \overline{\omega})$. Remarquons en particulier que $\overrightarrow{m} = -\overrightarrow{n}$ sur γ.
Alors en combinant (7.7) et (7.8) l'équation (7.6) devient :

$$\begin{aligned} \int_\Omega U \cdot \phi \, dX + \int_\Omega (\nabla \times U)(\nabla \times \phi) \, dX = \int_{\Omega \backslash \overline{\omega}} f_1 \cdot \phi \, dX + \int_\omega (f_1 + f_2) \cdot \phi \, dX \\ + \int_\gamma \Big\{ ((\nabla \times W) \times \overrightarrow{n}) - ((\nabla \times V^+) \times \overrightarrow{n}) \Big\} \cdot \phi \, d\sigma. \end{aligned} \tag{7.9}$$

En utilisant (7.2) et (7.4) nous obtenons

$$\lambda := (\nabla \times V^-) \times \overrightarrow{n} - (\nabla \times V^+) \times \overrightarrow{n} = (\nabla \times V^-) \times \overrightarrow{n} - (\nabla \times W) \times \overrightarrow{n},$$

c'est-à-dire

$$\int_\gamma \{ ((\nabla \times W) \times \overrightarrow{n}) - ((\nabla \times V^+) \times \overrightarrow{n}) \} \cdot \phi \, d\sigma = 0.$$

Finalement,

$$\int_\Omega U \cdot \phi \, dX + \int_\Omega (\nabla \times U)(\nabla \times \phi) \, dX = \int_\Omega (f_1 + f_2) \cdot \phi \, dX.$$

□

Ainsi, en définissant l'opérateur

$$T\lambda = ((\nabla \times W) \times \overrightarrow{n}) - ((\nabla \times V^+) \times \overrightarrow{n}),$$

il suffit de trouver λ tel que $T\lambda = 0$.
Le résultat fondamental pour la résolution de notre problème est le suivant

Lemme 7.1.2. *L'opérateur T est explicitement donné par*

$$T\lambda = \lambda + ((\nabla \times \overline{W}) \times \overrightarrow{n}) - ((\nabla \times \overline{V}^+) \times \overrightarrow{n}) \tag{7.10}$$

où $\overline{V}$ et $\overline{W}$ sont les solutions respectives de

$$\begin{cases} \overline{V} + \nabla \times \nabla \times \overline{V} = f_1 \text{ in } \Omega, \\ \overline{V} \times \overrightarrow{n} = 0 \text{ on } \Gamma, \end{cases} \tag{7.11}$$

et

$$\begin{cases} \overline{W} + \nabla \times \nabla \times \overline{W} = f_1 + f_2 \text{ in } \omega, \\ \overline{W} \times \overrightarrow{n} = \overline{V} \times \overrightarrow{n} \text{ on } \gamma. \end{cases} \tag{7.12}$$

Démonstration. Introduisons $\widetilde{W} = W - V^-$ et $\widetilde{\overline{W}} = \overline{W} - \overline{V}^-$ et notons que ces deux quantités vérifient toutes les deux

$$\begin{cases} E + \nabla \times \nabla \times E = f_2 \text{ in } \omega, \\ E \times \overrightarrow{n} = 0 \text{ on } \gamma. \end{cases} \tag{7.13}$$

L'unicité de la solution de ce problème nous donne $(\nabla \times \widetilde{W}) \times \overrightarrow{n} = (\nabla \times \widetilde{\overline{W}}) \times \overrightarrow{n}$ ce qui signifie que

$$(\nabla \times W) \times \overrightarrow{n} - (\nabla \times V^-) \times \overrightarrow{n} = (\nabla \times \overline{W}) \times \overrightarrow{n} - (\nabla \times \overline{V}^-) \times \overrightarrow{n}.$$

Nous obtenons alors le résultat en additionnant et soustrayant dans chacun des membres de l'équation ci-dessus $(\nabla \times V^+) \times \overrightarrow{n}$ et en remarquant que

$$(\nabla \times \overline{V}^+) \times \overrightarrow{n} - (\nabla \times \overline{V}^-) \times \overrightarrow{n} = 0,$$

$$(\nabla \times V^+) \times \overrightarrow{n} - (\nabla \times V^-) \times \overrightarrow{n} = -\lambda,$$

et

$$((\nabla \times W) \times \overrightarrow{n}) - ((\nabla \times V^+) \times \overrightarrow{n}) = T\lambda.$$

□

Ainsi l'équation $T\lambda = 0$ est trivialement équivalente à

$$\lambda = ((\nabla \times \overline{V}^{+}) \times \overrightarrow{n}) - ((\nabla \times \overline{W}) \times \overrightarrow{n}).$$

Bien entendu, cela n'est vrai que si l'on considère les problèmes continus (7.11) et (7.12). Il est certain que de résoudre les problèmes discrétisés correspondants et de définir

$$\lambda_H = ((\nabla \times \overline{V}_H^{+}) \times \overrightarrow{n}) - ((\nabla \times \overline{W}_h) \times \overrightarrow{n}) \tag{7.14}$$

introduira une erreur que l'on devra corriger.
Toutefois, l'algorithme de résolution deux échelles que l'on va utiliser apparaît déjà clairement : Résoudre les problèmes discrétisés issus de (7.11) et (7.12) nous permet de définir $\lambda_H = ((\nabla \times \overline{V}_H^{+}) \times \overrightarrow{n}) - ((\nabla \times \overline{W}_h) \times \overrightarrow{n})$, puis de résoudre les problèmes discrétisés issus de (7.2) et (7.3), et finalement de définir une approximation de la solution de notre problème initial (7.1) comme dans (7.5).

7.2 Problème discret

Nous allons naturellement utiliser les éléments finis d'arête que nous avons développés. Une fois que l'on s'est fixé les deux maillages, à savoir un maillage grossier sur le domaine Ω et un maillage plus fin sur le domaine ω, nous définissons deux espaces de discrétisation d'éléments finis P_H et P_h respectivement associés au maillage grossier et au maillage fin. En supposant que la frontière γ corresponde à la réunion d'un certain nombre d'arêtes du maillage grossier, nous définissons aussi un espace de discrétisation d'éléments finis Δ_H qui correspond à l'espace des traces tangentielles sur γ des fonctions de l'espace P_H. Remarquons que les degrés de liberté définissant de manière unique toute fonction de cet espace correspondent aux degrés de liberté d'arête définissant les fonctions de P_H associées aux arêtes définissant γ. C'est dans cet espace Δ_H que l'on doit chercher λ.

L'idée est alors de définir un opérateur $T_H : \Delta_H \to \Delta_H$ qui soit une approximation appropriée de $T\lambda$ et de déterminer λ_H tel que $T_H\lambda_H = 0$ de manière à corriger l'approximation λ_H définie par (7.14). Ainsi, nous définissons une approximation de la solution U du problème (7.1) par

$$U_{Hh} = V_H^{+}\chi(\Omega\backslash\overline{\omega}) + W_h\chi(\omega) \tag{7.15}$$

où V_H^{+} est la restriction sur $\Omega\backslash\overline{\omega}$ de la solution de l'équation suivante

$$\int_\Omega V_H \cdot \phi \, dX + \int_\Omega (\nabla \times V_H)(\nabla \times \phi) \, dX = \int_\Omega f_1 \cdot \phi \, dX + \int_\gamma \lambda_H \cdot \phi \, d\sigma \quad \forall \phi \in P_H, \tag{7.16}$$

et W_h est la solution de

$$\int_\omega W_h \cdot \phi \, dX + \int_\omega (\nabla \times W_h)(\nabla \times \phi) \, dX = \int_\omega (f_1 + f_2) \cdot \phi \, dX \quad \forall \phi \in P_h \tag{7.17}$$

avec la condition de bord $W_h \times \overrightarrow{n} = V_H \times \overrightarrow{n}$ sur γ.

Il faut alors se rendre compte que si la trace tangentielle $((\nabla \times V_H^{+}) \times \overrightarrow{n})$ sur γ est par définition dans Δ_H, ce n'est pas le cas de $((\nabla \times W_h) \times \overrightarrow{n})$. Cela signifie que l'on ne peut

pas définir directement $T_H = ((\nabla \times W_h) \times \overrightarrow{n}) - ((\nabla \times V_H^+) \times \overrightarrow{n})$.

La façon la plus naturelle de construire une approximation conforme de $T\lambda$ est alors de trouver $\delta_H \in \Delta_H$ tel que

$$\int_\gamma \delta_H \cdot \Phi \, d\sigma = \int_\gamma ((\nabla \times W_h) \times \overrightarrow{n}) - ((\nabla \times V_H^+) \times \overrightarrow{n}) \cdot \Phi \, d\sigma \quad \forall \Phi \in \Delta_H, \tag{7.18}$$

et de définir $T_H \lambda_H = \delta_H$.
Une méthode plus élaborée de définir T_H est la suivante : plutôt que d'intégrer directement

$$\int_\gamma ((\nabla \times W_h) \times \overrightarrow{n}) \cdot \Phi \, d\sigma$$

et

$$\int_\gamma ((\nabla \times V_H^+) \times \overrightarrow{n}) \cdot \Phi \, d\sigma$$

dans l'équation (7.18), ce qui est une approximation extrêmement grossière étant donné que l'on n'utilise que l'information sur la frontière γ, il est préférable d'utiliser le fait que si

$$V + \nabla \times \nabla \times V = f_1 \text{ in } \Omega, \tag{7.19}$$

alors

$$\begin{aligned} &\int_{\Omega \setminus \overline{\omega}} V_H^+ \cdot \phi \, dX + \int_{\Omega \setminus \overline{\omega}} (\nabla \times V_H^+)(\nabla \times \phi) \, dX \\ &\qquad = \int_{\Omega \setminus \overline{\omega}} f_1 \cdot \phi \, dX - \int_\gamma ((\nabla \times V_H^+) \times \overrightarrow{n}) \cdot \phi \, d\sigma \quad \forall \phi \in P_H, \end{aligned} \tag{7.20}$$

puis en combinant cette équation avec (7.16) nous obtenons

$$\begin{aligned} &\int_\gamma ((\nabla \times V_H^+) \times \overrightarrow{n}) \cdot \Phi \, d\sigma = \\ &\qquad \int_\omega V_H^- \cdot \tilde{\Phi} \, dX + \int_\omega (\nabla \times V_H^-)(\nabla \times \tilde{\Phi}) \, dX - \int_\omega f_1 \cdot \tilde{\Phi} \, dX - \int_\gamma \lambda_H \cdot \Phi \, d\sigma, \end{aligned} \tag{7.21}$$

quel que soit Φ dans Δ_H, où $\tilde{\Phi}$ désigne un prolongement quelconque de Φ dans P_H.
De la même manière nous obtenons une autre expression de

$$\int_\gamma ((\nabla \times W_h) \times \overrightarrow{n}) \cdot \Phi \, d\sigma = \int_\omega W_h \cdot \tilde{\Phi} \, dX + \int_\omega (\nabla \times W_h)(\nabla \times \tilde{\Phi}) \, dX - \int_\omega (f_1 + f_2) \cdot \tilde{\Phi} \, dX. \tag{7.22}$$

Nous proposons finalement de calculer $T_H \lambda_H = \delta_H$, où δ_H vérifie : $\forall \Phi \in \Delta_H$,

$$\begin{aligned} \int_\gamma \delta_H . \Phi \, d\sigma = \int_\gamma \lambda_H \cdot \Phi \, d\sigma \quad &+ \left(\int_\omega W_h \cdot \tilde{\Phi} \, dX + \int_\omega (\nabla \times W_h)(\nabla \times \tilde{\Phi}) \, dX - \int_\omega (f_1 + f_2) \cdot \tilde{\Phi} \, dX \right) \\ &- \left(\int_\omega V_H^- \cdot \tilde{\Phi} \, dX + \int_\omega (\nabla \times V_H^-)(\nabla \times \tilde{\Phi}) \, dX - \int_\omega f_1 \cdot \tilde{\Phi} \, dX \right). \end{aligned} \tag{7.23}$$

7.3 Algorithme de résolution

Nous avons déjà évoqué à la fin de la section 7.1 le fait que le passage des équations continues aux équations discrètes introduit certaines erreurs qu'il nous faut corriger. Plus précisément cela signifie que définir λ_H comme dans (7.14) n'implique pas que $T_H\lambda_H = 0$, même si cette approximation de λ est conforme. C'est pourquoi nous choisissons d'appliquer un point fixe à l'opérateur $(I - T_H)$ de la manière suivante :

1. Définir $\lambda_H = 0$ pour initialisation.

2. Résoudre
$$\begin{cases} V_H + \nabla \times \nabla \times V_H = f_1 \text{ dans } \Omega, \\ V_H \times \overrightarrow{n} = 0 \text{ sur } \Gamma, \\ \left[V_H \times \overrightarrow{n}\right] = 0 \text{ sur } \gamma, \\ \left[(\nabla \times V_H) \times \overrightarrow{n}\right] = -\lambda_H \text{ sur } \gamma \end{cases} \tag{7.24}$$
et de cette solution récupérer $V_H \times \overrightarrow{n}$ sur γ.

3. Résoudre
$$\begin{cases} W_h + \nabla \times \nabla \times W_h = f_1 + f_2 \text{ dans } \omega, \\ W_h \times \overrightarrow{n} = V_H \times \overrightarrow{n} \text{ sur } \gamma. \end{cases} \tag{7.25}$$

4. Définir $T_H\lambda_H = \delta_H$ en utilisant l'équation (7.23).

5. Finalement définir $\lambda_H = \lambda_H - T_H\lambda_H$.

6. Retourner à la seconde étape de l'algorithme.

Il faut remarquer qu'à la première itération de l'algorithme, à $\lambda_H = 0$, résoudre (7.24) et (7.25), définir $T_H\lambda_H = \delta_H$ et redéfinir $\lambda_H = \lambda_H - T_H\lambda_H$ correspond exactement à ce que l'on à dit a la fin de la section 7.1, c'est-à-dire résoudre la version discrétisée de (7.11) et (7.12)et définir λ_H comme dans (7.14), de sorte que définir U_{Hh} par (7.15) avec V_H et W_h calculés à la seconde itération est déjà une bonne approximation de la solution de notre problème initial. Les itérations suivantes ne serviront qu'à corriger λ_H de manière à forcer la nullité de $T_H\lambda_H$ et donc d'améliorer l'approximation.

7.4 Application au problème dépendant du temps

Nous allons maintenant considérer le problème suivant :

$$\begin{cases} \partial_t^2 U + \nabla \times \nabla \times U = f \text{ in } \Omega \times [0,T] \\ U \times \overrightarrow{n} = 0 \text{ on } \Gamma \times [0,T] \end{cases} \tag{7.26}$$

où T désigne un temps final que l'on se fixe, et f est la somme de deux fonctions $f_1, f_2 \in (L^2(\Omega)\times[0,T])^2$ avec $supp(f_2(.,t)) \subset \omega$ pour tout temps $t \in [0,T]$. Nous introduisons alors

les deux problèmes auxiliaires :

$$\begin{cases} \partial_t^2 V + \nabla \times \nabla \times V = f_1 \text{ dans } \Omega \times [0,T] \\ V \times \overrightarrow{n} = 0 \text{ sur } \Gamma \times [0,T] \\ \left[V \times \overrightarrow{n}\right] = 0 \text{ sur } \gamma \times [0,T] \\ \left[(\nabla \times V) \times \overrightarrow{n}\right] = -\lambda \text{ sur } \gamma \times [0,T] \end{cases} \tag{7.27}$$

et

$$\begin{cases} \partial_t^2 W - \nabla \times \nabla \times W = f_1 + f_2 \text{ dans } \omega \times [0,T] \\ W \times \overrightarrow{n} = V \times \overrightarrow{n} \text{ sur } \gamma \times [0,T] \end{cases} \tag{7.28}$$

Après avoir fixé un pas de temps Δ_t, introduit la discrétisation de l'axe temporel $t^n = n\Delta_t$, et discrétisé la dérivée temporelle seconde par le schéma centré standard d'ordre deux, ces deux problèmes deviennent :

$$\begin{cases} V^{n+1} + \Delta_t^2 \nabla \times \nabla \times V^{n+1} = \Delta_t^2 f_1^{n+1} + 2V^n - V^{n-1} \text{ dans } \Omega, \\ V^{n+1} \times \overrightarrow{n} = 0 \text{ sur } \Gamma, \\ \left[V^{n+1} \times \overrightarrow{n}\right] = 0 \text{ sur } \gamma, \\ \left[(\nabla \times V^{n+1}) \times \overrightarrow{n}\right] = -\lambda^{n+1} \text{ sur } \gamma \end{cases} \tag{7.29}$$

et

$$\begin{cases} W^{n+1} + \Delta_t^2 \nabla \times \nabla \times W^{n+1} = \Delta_t^2 (f_1^{n+1} + f_2^{n+1}) + 2W^n - W^{n-1} \text{ dans } \omega, \\ W^{n+1} \times \overrightarrow{n} = V^{n+1} \times \overrightarrow{n} \text{ on } \gamma \end{cases} \tag{7.30}$$

que l'on doit résoudre à chaque pas de temps $n\Delta_t$. En définissant

$$\tilde{f}_1 = \Delta_t^2 f_1^{n+1} + 2V^n - V^{n-1},$$

et

$$\tilde{f}_1 + \tilde{f}_2 = \Delta_t^2 (f_1^{n+1} + f_2^{n+1}) + 2W^n - W^{n-1},$$

nous obtenons automatiquement

$$\tilde{f}_2 = \Delta_t^2 f_2^{n+1} + 2(W^n - V^n) - (W^{n-1} - V^{n-1}).$$

Le plus **important** est alors d'interpréter les termes sources $\tilde{f}_1$ et $\tilde{f}_2$. En effet, $\tilde{f}_1$ et $\tilde{f}_2$ font apparaître les termes V^n et V^{n-1} qui doivent être vus comme les solutions du problème initial (7.26) sur le maillage grossier aux pas de temps respectifs $n\Delta_t$ et $(n-1)\Delta_t$, et **pas** comme les solutions du problème (7.29) : rappelons que le problème (7.29) ne donne une approximation de la solution du problème initial que dans $\Omega \backslash \overline{\omega}$ et que cette solution est donnée par W^n (solution du problème (7.30)) dans ω, de sorte que les termes V^n et V^{n-1} dans $\tilde{f}_1$ et $\tilde{f}_2$ doivent être définis comme V^n et V^{n-1} (solutions du problème (7.29) au temps $n\Delta_t$ et $(n-1)\Delta_t$) dans $\Omega \backslash \overline{\omega}$ et comme la projection sur l'espace d'éléments finis associé au maillage grossier de W^n et W^{n-1} (solutions du problème (7.30) au temps $n\Delta_t$ et $(n-1)\Delta_t$) dans ω.
Nous pouvons alors résoudre les problèmes (7.29) et (7.30) à chaque pas de temps en utilisant la méthode deux échelles décrite précédemment.

7.5 Simulations numériques

Dans cette section nous allons effectuer une série de simulations numériques pour mettre en évidence l'efficacité de la méthode deux échelles.
Dans tout ce qui suit, nous appellerons solution numérique la solution de notre problème initial issue de l'utilisation de notre méthode deux échelles, et solution de référence la solution de notre problème initial calculée sur une extension du maillage fin sur tout le domaine.
Dans la mesure où l'on ne s'intéresse qu'à la perte de précision que peut entraîner l'utilisation de notre méthode deux échelles par rapport à une résolution directe sur un maillage uniformément raffiné, nous ne considérerons que l'erreur relative entre ces deux solutions, que nous déterminerons en norme $L^2(\Omega)$ et $H(rot, \Omega)$.
Ces erreurs seront d'autre part scindées en erreurs dite "intérieure" et "extérieure" ce qui correspondra respectivement aux erreurs dans ω et $\Omega\backslash\overline{\omega}$.

Les deux premiers cas tests ont été choisis de manière à ce que la solution de notre problème initial soit donné dans $\Omega\backslash\overline{\omega}$ par la restriction à $\Omega\backslash\overline{\omega}$ de la solution de (7.2) (c'est à dire de manière à ce que la solution de notre problème initial soit donnée à $\lambda = 0$). Pour le troisième cas test la solution de notre problème initial dans $\Omega\backslash\overline{\omega}$ est cette fois vraiment dépendante de λ, et le quatrième cas test est identique au troisième à ceci près que l'on considère un domaine ω légèrement plus grand que le support de f_2 (alors que dans le troisième cas test $\omega = supp(f_2)$). Nous testons notre méthode deux échelles sur le problème dépendant du temps dans les cas tests cinq et six, en propageant une impulsion initiale dérivant d'une gaussienne dans le cas test cinq, et en générant une onde via l'imposition d'une force oscillante dans le sixième cas test.
Pour les quatre premiers cas test, c'est-à-dire dans le cas stationnaire, nous n'utiliserons, que des éléments finis d'arête de plus bas degré, tandis que nous profiterons des deux cas test instationnaires pour comparer la méthode deux échelles avec des éléments finis d'arête d'ordre un à trois.

7.5.1 Cas test 1

Pour notre premier cas test nous considérons $\Omega = [-\frac{\pi}{2}, \frac{\pi}{2}] \times [-\frac{\pi}{2}, \frac{\pi}{2}], \omega = [-\frac{\pi}{6}, \frac{\pi}{6}] \times [-\frac{\pi}{6}, \frac{\pi}{6}]$, $f_1 = \begin{pmatrix} 2\cos(y) \\ 2\cos(x) \end{pmatrix}$ comme force de fond et f_2 identiquement nulle.
La solution exacte de ce problème est donnée par $U = \begin{pmatrix} \cos(y) \\ \cos(x) \end{pmatrix}$.
Remarquons que la composante tangentielle de la solution sur chacune des arêtes du maillage est constante, de sorte que ce cas test est parfaitement adapté pour une résolution par éléments finis d'arête de plus bas degré.
Notons aussi que dans la mesure où $f_2 = 0$, le seul intérêt de ce cas test est de vérifier que l'on a déjà une bonne approximation de la solution à la première itération de l'algorithme, et que cette solution reste quasiment stable au cours des itérations ultérieures (la solution devrait être légèrement modifiée de par le fait que la résolution dans le domaine ω, qui fait apparaître $f_1 + f_2$ en terme source, est meilleure grâce au raffinement du maillage).

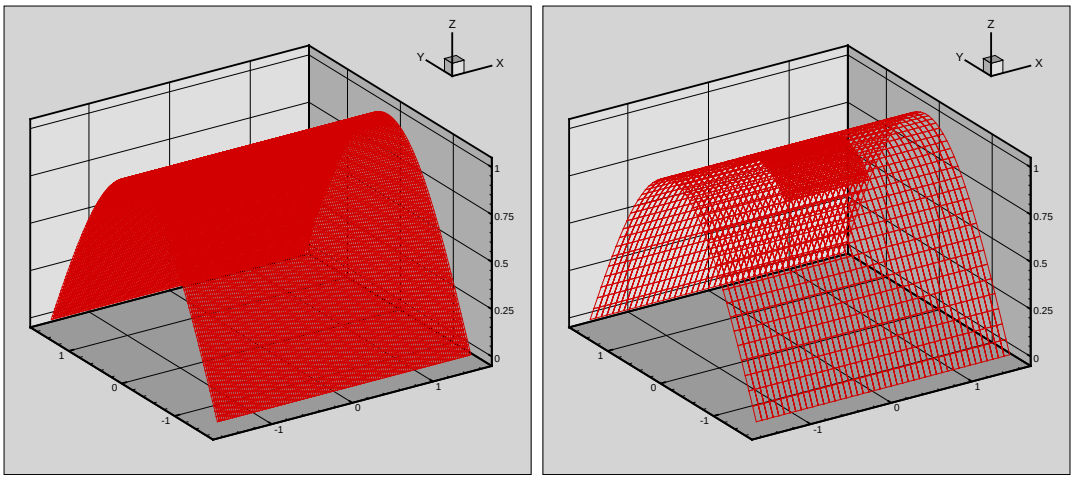

FIG. 7.2 – Première composante de la solution de référence et de la solution numérique après une itération pour le premier cas test.

Itération	1	2	5	10
Erreur L^2	0.127792E-01	0.129802E-01	0.129983E-01	0.129983E-01
Erreur L^2 extérieure	0.141880E-01	0.144605E-01	0.144782E-01	0.144782E-01
Erreur L^2 intérieure	0.376103E-02	0.300233E-02	0.305109E-02	0.305110E-02
Erreur $H(rot)$	0.599090E-01	0.598985E-01	0.598982E-01	0.598982E-01
Erreur $H(rot)$ extérieure	0.635364E-01	0.635261E-01	0.635258E-01	0.635258E-01
Erreur $H(rot)$ intérieure	0.106011E-02	0.451137E-03	0.433064E-03	0.433062E-03

TAB. 7.1 – Erreurs sur un maillage de 9×9 éléments pour le premier cas test.

Nous donnons dans la figure 7.2 la première composante de la solution de référence et de la solution numérique après une itération. Les tableaux 7.1 à 7.3 donnent les erreurs après 1, 2, 5 et 10 itérations pour des maillages grossiers respectivement de 9×9, 27×27 et 45×45 éléments et des maillages fin trois fois plus raffinés dans chacun des cas.

7.5.2 Cas test 2

Pour ce deuxième cas test nous avons voulu construire un problème dont la solution dans $\Omega \backslash \overline{\omega}$ ne dépend pas de la solution du problème (7.3) dans ω. Pour cela il faut considérer une force imposée f_2 localisée dans ω dont toutes les dérivées successives s'annulent (numériquement tout du moins) sur γ, sans toutefois que la force f_2 soit identiquement nulle.
Plus précisément, nous considérons $\Omega = [-12, 12] \times [-12, 12]$, $\omega = [-4, 4] \times [-4, 4]$, $f_1 = \begin{pmatrix} (\frac{\pi^2}{576} + 1)\cos(\frac{\pi y}{24}) \\ (\frac{\pi^2}{576} + 1)\cos(\frac{\pi x}{24}) \end{pmatrix}$ comme force de fond et $f_2 = \begin{pmatrix} \eta e^{\frac{-(x^2+y^2)}{\epsilon}}(1 + \frac{2}{\epsilon} - 4\frac{(y^2 - xy)}{\epsilon^2}) \\ \eta e^{\frac{-(x^2+y^2)}{\epsilon}}(1 + \frac{2}{\epsilon} - 4\frac{(x^2 - xy)}{\epsilon^2}) \end{pmatrix}$

Itération	1	2	5	10
Erreur L^2	0.142712E-02	0.144246E-02	0.144295E-02	0.144295E-02
Erreur L^2 extérieure	0.158463E-02	0.161307E-02	0.161352E-02	0.161352E-02
Erreur L^2 intérieure	0.417327E-03	0.184581E-03	0.187740E-03	0.187741E-03
Erreur $H(rot)$	0.199886E-01	0.199878E-01	0.199878E-01	0.199878E-01
Erreur $H(rot)$ extérieure	0.212009E-01	0.212000E-01	0.212000E-01	0.212000E-01
Erreur $H(rot)$ intérieure	0.118266E-03	0.178047E-04	0.183512E-04	0.183514E-04

TAB. 7.2 – Erreurs sur un maillage de 27×27 éléments pour le premier cas test.

Itération	1	2	5	10
Erreur L^2	0.513971E-03	0.519377E-03	0.519473E-03	0.519473E-03
Erreur L^2 extérieure	0.570703E-03	0.580997E-03	0.581085E-03	0.581085E-03
Erreur L^2 intérieure	0.150221E-03	0.595088E-04	0.602927E-04	0.602931E-04
Erreur $H(rot)$	0.119941E-01	0.119939E-01	0.119939E-01	0.119939E-01
Erreur $H(rot)$ extérieure	0.127216E-01	0.127214E-01	0.127214E-01	0.127214E-01
Erreur $H(rot)$ intérieure	0.425890E-04	0.753758E-05	0.769877E-05	0.769884E-05

TAB. 7.3 – Erreurs sur un maillage de 45×45 éléments pour le premier cas test.

comme force locale dans ω avec $\epsilon = 0.5$ et $\eta = 10$.
La figure 7.3 représente la première composante de la solution de référence et de la solution numérique après une itération calculées sur un maillage grossier de 45×45 et un maillage fin trois fois plus raffiné (ce qui signifie en particulier que la solution de référence est calculée sur un maillage de 135×135 éléments). Les tableaux 7.4 à 7.6 listent les erreurs calculées dans les mêmes conditions que pour le premier cas test.

7.5.3 Cas test 3

Pour ce cas test nous considérons $\Omega = [-\frac{\pi}{2}, \frac{\pi}{2}] \times [-\frac{\pi}{2}, \frac{\pi}{2}]$, $\omega = [-\frac{\pi}{6}, \frac{\pi}{6}] \times [-\frac{\pi}{6}, \frac{\pi}{6}]$, $f_1 = \begin{pmatrix} 2\cos(y) \\ 2\cos(x) \end{pmatrix}$ comme force de fond et $f_2 = 10\cos(9x)\cos(9y)\chi(\omega)\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Itération	1	2	5	10
Erreur L^2	0.823368E-02	0.854288E-02	0.859216E-02	0.859221E-02
Erreur L^2 extérieure	0.101939E-01	0.105637E-01	0.106163E-01	0.106164E-01
Erreur L^2 intérieure	0.132116E-02	0.154363E-02	0.165281E-02	0.165288E-02
Erreur $H(rot)$	0.946069E-02	0.947588E-02	0.948614E-02	0.948615E-02
Erreur $H(rot)$ extérieure	0.123393E-01	0.123511E-01	0.123566E-01	0.123566E-01
Erreur $H(rot)$ intérieure	0.760773E-03	0.927762E-03	0.106761E-02	0.106779E-02

TAB. 7.4 – Erreurs sur un maillage de 9×9 éléments pour le second cas test.

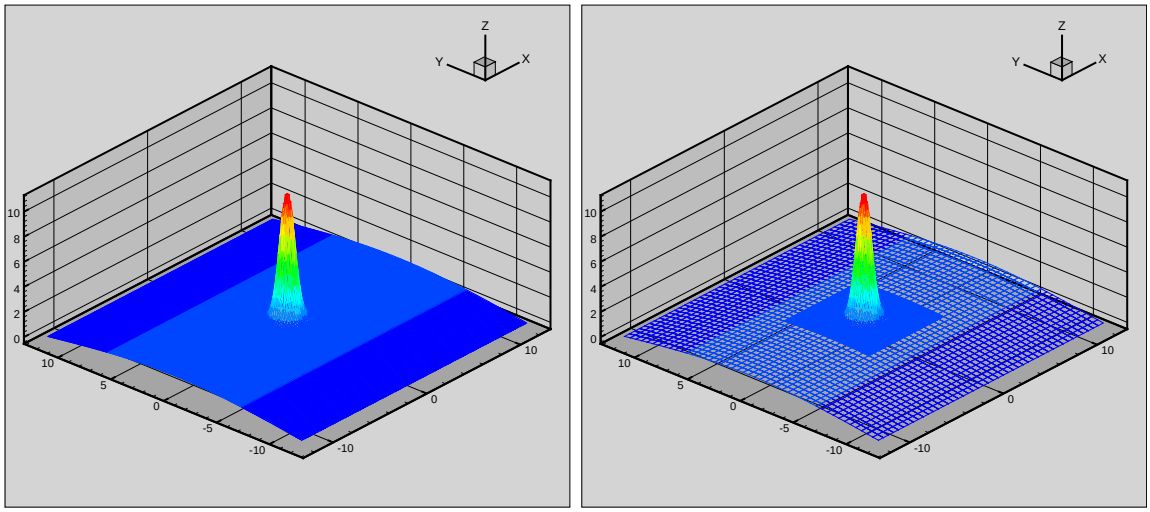

FIG. 7.3 – Première composante de la solution de référence et de la solution numérique après une itération pour le second cas test.

Itération	1	2	5	10
Erreur L^2	0.869482E-03	0.874768E-03	0.875010E-03	0.875010E-03
Erreur L^2 extérieure	0.113507E-02	0.114716E-02	0.114750E-02	0.114750E-02
Erreur L^2 intérieure	0.137202E-03	0.508224E-04	0.503533E-04	0.503526E-04
Erreur $H(rot)$	0.270281E-02	0.270232E-02	0.270232E-02	0.270232E-02
Erreur $H(rot)$ extérieure	0.410376E-02	0.410342E-02	0.410342E-02	0.410342E-02
Erreur $H(rot)$ intérieure	0.589900E-04	0.308610E-04	0.313108E-04	0.313108E-04

TAB. 7.5 – Erreurs sur un maillage de 27×27 éléments pour le second cas test.

Itération	1	2	5	10
Erreur L^2	0.312617E-03	0.312908E-03	0.312931E-03	0.312931E-03
Erreur L^2 extérieure	0.408765E-03	0.411261E-03	0.411291E-03	0.411291E-03
Erreur L^2 intérieure	0.499898E-04	0.110458E-04	0.110195E-04	0.110195E-04
Erreur $H(rot)$	0.160328E-02	0.160314E-02	0.160314E-02	0.160314E-02
Erreur $H(rot)$ extérieure	0.246170E-02	0.246159E-02	0.246159E-02	0.246159E-02
Erreur $H(rot)$ intérieure	0.210766E-04	0.672441E-05	0.677168E-05	0.677168E-05

TAB. 7.6 – Erreurs sur un maillage de 45×45 éléments pour le second cas test.

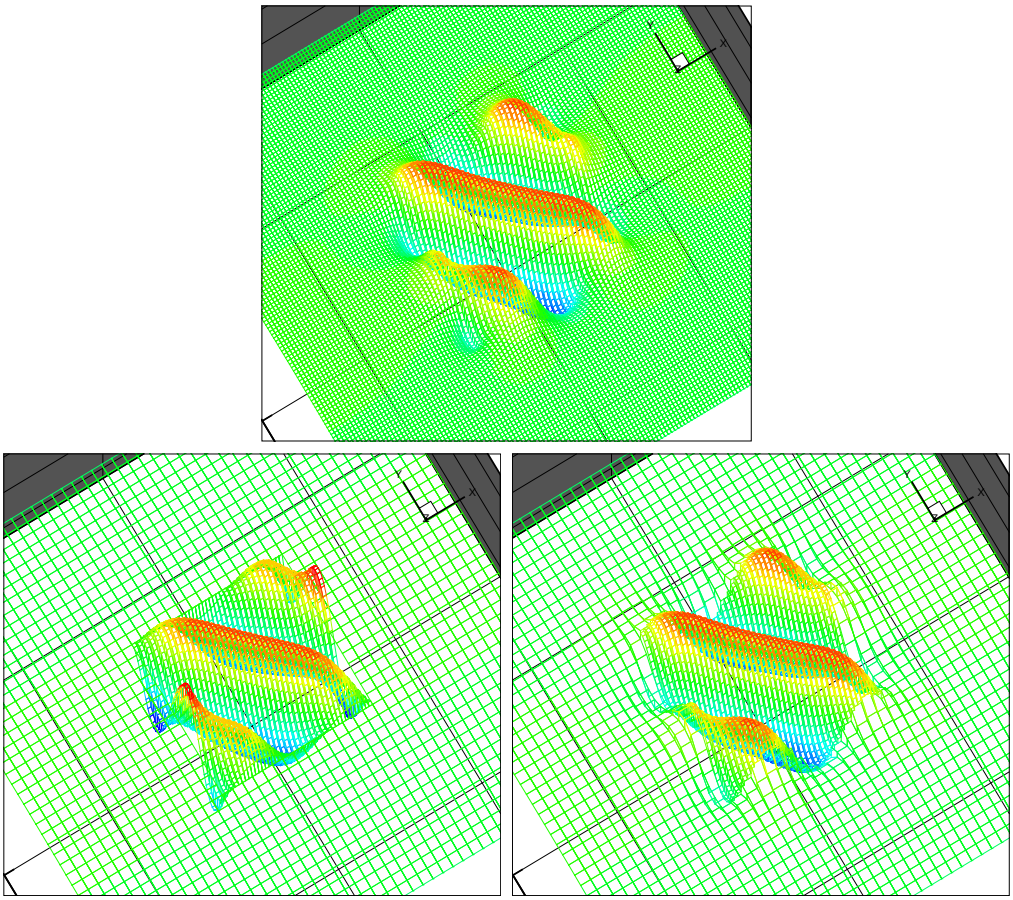

FIG. 7.4 – Première composante de la solution de référence, de la solution numérique après une itération et de la solution numérique après deux itérations pour le troisième cas test.

Remarquons que, bien que la force f_2 s'annule identiquement sur γ, la solution de notre problème initial dans $\Omega\backslash\overline{\omega}$ dépend cette fois de la solution du problème (7.3) dans ω, ce qui signifie que la solution du problème découplé (7.2)-(7.3) ne coincide plus, trivialement lorsque $\lambda = 0$, avec la solution de notre problème initial (7.1).
La figure 7.4 représente la première composante de la solution de référence, de la solution numérique après une itération et de la solution numérique après deux itérations.
Remarquons que ces figures sont cohérentes avec ce que nous attendions de l'algorithme, à savoir que ce que l'on calcule à la première itération n'est pas, en général, une approximation de la solution mais une donnée qui nous permet de calculer λ_H, approximation de λ. C'est seulement ensuite, à la seconde itération, que l'algorithme nous retourne une approximation de notre problème initial.
Les tableaux 7.7 à 7.9 listent les erreurs calculées dans les mêmes conditions que pour le premier cas test.

Iteration	1	2	5	10
Erreur L^2	0.398497E+00	0.281039E+00	0.278372E+00	0.278371E+00
Erreur L^2 extérieure	0.493352E+00	0.379490E+00	0.387446E+00	0.387477E+00
Erreur L^2 intérieure	0.347533E+00	0.223060E+00	0.211780E+00	0.211754E+00
Erreur $H(rot)$	0.562216E-01	0.454014E-01	0.450850E-01	0.450844E-01
Erreur $H(rot)$ extérieure	0.751863E-01	0.636354E-01	0.635357E-01	0.635356E-01
Erreur $H(rot)$ intérieure	0.302904E-01	0.164788E-01	0.151184E-01	0.151155E-01

TAB. 7.7 – Erreurs sur un maillage de 9×9 éléments pour le troisième cas test.

Iteration	1	2	5	10
Erreur L^2	0.387811E+00	0.740680E-01	0.731386E-01	0.731386E-01
Erreur L^2 extérieure	0.508526E+00	0.136386E+00	0.136992E+00	0.136993E+00
Erreur L^2 intérieure	0.329459E+00	0.203193E-01	0.126115E-01	0.126112E-01
Erreur $H(rot)$	0.409944E-01	0.147756E-01	0.147125E-01	0.147125E-01
Erreur $H(rot)$ extérieure	0.515924E-01	0.222988E-01	0.222920E-01	0.222920E-01
Erreur $H(rot)$ intérieure	0.304299E-01	0.195685E-02	0.877946E-03	0.877902E-03

TAB. 7.8 – Erreurs sur un maillage de 27×27 éléments pour le troisième cas test.

Iteration	1	2	5	10
Erreur L^2	0.386945E+00	0.435838E-01	0.433808E-01	0.433808E-01
Erreur L^2 extérieure	0.509749E+00	0.821687E-01	0.822958E-01	0.822958E-01
Erreur L^2 intérieure	0.328072E+00	0.668444E-02	0.350526E-02	0.350523E-02
Erreur $H(rot)$	0.396788E-01	0.886878E-02	0.885411E-02	0.885411E-02
Erreur $H(rot)$ extérieure	0.492684E-01	0.134717E-01	0.134723E-01	0.134723E-01
Erreur $H(rot)$ intérieure	0.304401E-01	0.746564E-03	0.296818E-03	0.296817E-03

TAB. 7.9 – Erreurs sur un maillage de 45×45 éléments pour le troisième cas test.

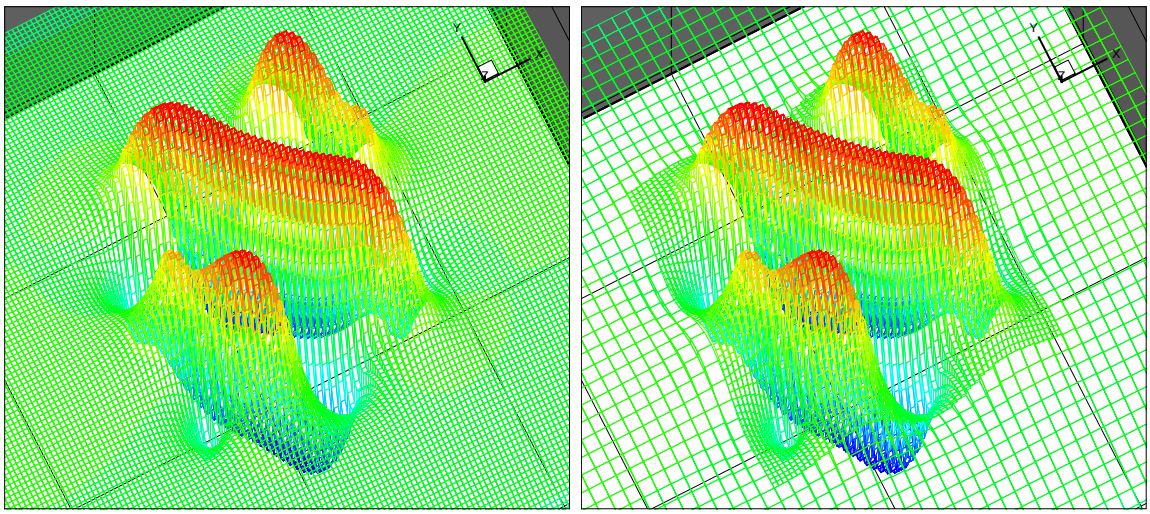

FIG. 7.5 – Première composante de la solution de référence et de la solution numérique après deux itérations pour le quatrième cas test.

Iteration	1	2	5	10
Erreur L^2	0.881475E-01	0.743515E-02	0.741946E-02	0.741946E-02
Erreur L^2 extérieure	0.145889E+00	0.182483E-01	0.182695E-01	0.182695E-01
Erreur L^2 intérieure	0.714757E-01	0.768414E-03	0.398887E-03	0.398886E-03
Erreur $H(rot)$	0.250256E-01	0.695290E-02	0.695237E-02	0.695237E-02
Erreur $H(rot)$ extérieure	0.371278E-01	0.120720E-01	0.120716E-01	0.120716E-01
Erreur $H(rot)$ intérieure	0.159048E-01	0.815904E-04	0.357121E-04	0.357123E-04

TAB. 7.10 – Erreurs sur un maillage de 45×45 éléments pour le quatrième cas test.

7.5.4 Cas test 4

Pour ce quatrième cas test nous considérons $\Omega = [-\frac{\pi}{2}, \frac{\pi}{2}] \times [-\frac{\pi}{2}, \frac{\pi}{2}]$, $\omega = [-\frac{7\pi}{30}, \frac{7\pi}{30}] \times [-\frac{7\pi}{30}, \frac{7\pi}{30}]$, $f_1 = \begin{pmatrix} 2\cos(y) \\ 2\cos(x) \end{pmatrix}$ comme force de fond et $f_2 = 10\cos(9x)\cos(9y)\chi([-\frac{\pi}{6}, \frac{\pi}{6}] \times [-\frac{\pi}{6}, \frac{\pi}{6}]) \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.
La seule différence entre le troisième et le quatrième cas test est que l'on considère cette fois-ci ω légèrement plus grand que $supp(f_2)$. Nous commenterons l'intérêt d'un telle manipulation par la suite dans la sous-section 7.5.7.
Nous donnons dans la figure 7.5 la première composante de la solution de référence et de la solution numérique après deux itérations et dans le tableau 7.10 les erreurs calculées sur un maillage grossier de 45×45 éléments et un maillage fin trois plus raffiné.

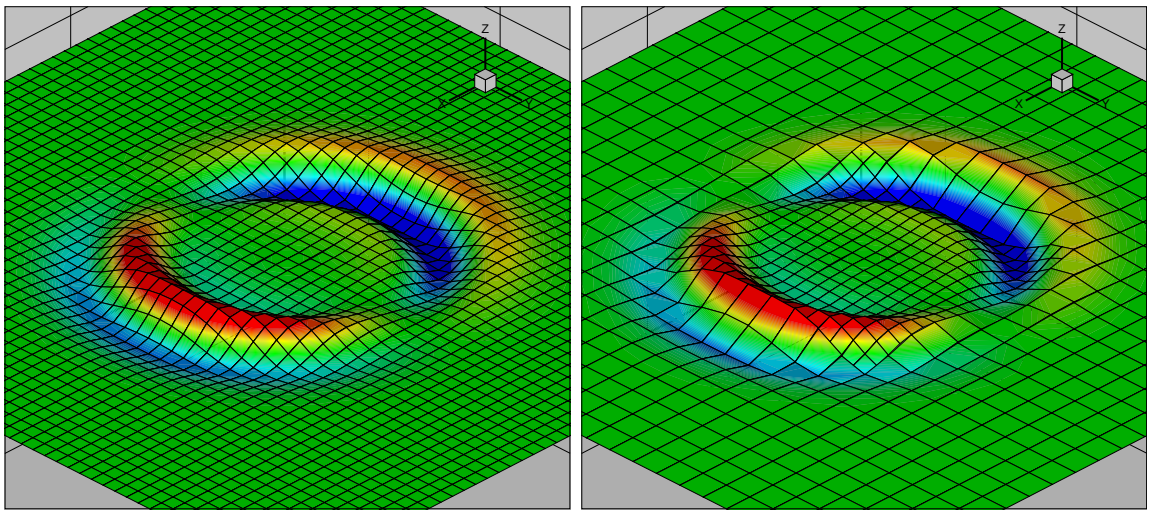

FIG. 7.6 – Seconde composante de la solution de référence et de la solution numérique après quarante-cinq itérations en temps pour le premier cas test dépendant du temps.

7.5.5 Cas test 5

Pour ce premier cas test sur le problème dépendant du temps nous considérons $\Omega = [-12, 12] \times [-12, 12]$, $\omega = [-4, 4] \times [-4, 4]$, $f_1 = f_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Nous imposons alors l'impulsion initiale $U(x, y, 0) = \begin{pmatrix} -2\eta\frac{y}{\epsilon}e^{(-\frac{x^2+y^2}{\epsilon})} \\ 2\eta\frac{x}{\epsilon}e^{(-\frac{x^2+y^2}{\epsilon})} \end{pmatrix}$ et $\partial_t U(x, y, 0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, où $\eta = 10$ et $\epsilon = 0.5$.
La figure 7.6 représente la seconde composante de la solution de référence et de la solution numérique après quarante-cinq itérations en temps (soit au temps $t = 6.25$) et le tableau 7.11 liste les erreurs calculées sur un maillage grossier de 27×27 éléments et un maillage fin deux fois plus raffiné aux temps $t = 4$, $t = 8$ et $t = 12$.

7.5.6 Cas test 6

Pour ce second cas test sur le problème dépendant du temps nous considérons cette fois $\Omega = [-12, 12]\times[-12, 12], \omega = [-4, 4]\times[-4, 4]$, $f_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, et $f_2 = \begin{pmatrix} -2\eta\frac{y}{\epsilon}e^{(-\frac{x^2+y^2}{\epsilon})}\sin(t) \\ 2\eta\frac{x}{\epsilon}e^{(-\frac{x^2+y^2}{\epsilon})}\sin(t) \end{pmatrix}$, où $\eta = 10$ et $\epsilon = 0.5$ et nous imposons les conditions initiales $U(x, y, 0) = \partial_t U(x, y, 0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.
La figure 7.7 représente la seconde composante de la solution de référence et de la solution numérique au temps $t = 12$, et nous donnons dans le tableau 7.12 les erreurs calculées sur un maillage grossier de 27×27 éléments et un maillage fin deux fois plus raffiné aux temps $t = 4$, $t = 8$ et $t = 12$.

Éléments finis d'arête du premier ordre			
Time t	4	8	12
Erreur L^2	0.581116E+00	0.719711E+00	0.635122E+00
Erreur L^2 extérieure	0.531007E+00	0.715346E+00	0.634705E+00
Erreur L^2 intérieure	0.236067E+00	0.791401E-01	0.229959E-01
Erreur $H(rot)$	0.189652E+01	0.181267E+01	0.135066E+01
Erreur $H(rot)$ extérieure	0.183644E+01	0.180503E+01	0.135005E+01
Erreur $H(rot)$ intérieure	0.473557E+00	0.166193E+00	0.403889E-01
Éléments finis d'arête du second ordre			
Time t	4	8	12
Erreur L^2	0.924122E-01	0.860510E-01	0.432997E-01
Erreur L^2 extérieure	0.886897E-01	0.855219E-01	0.432661E-01
Erreur L^2 intérieure	0.259643E-01	0.952856E-02	0.170669E-02
Erreur $H(rot)$	0.420503E+00	0.318012E+00	0.156138E+00
Erreur $H(rot)$ extérieure	0.418192E+00	0.316847E+00	0.156076E+00
Erreur $H(rot)$ intérieure	0.440241E-01	0.272051E-01	0.439168E-02
Éléments finis d'arête du troisième ordre			
Time t	4	8	12
Erreur L^2	0.139498E-01	0.103088E-01	0.449157E-02
Erreur L^2 extérieure	0.137516E-01	0.102902E-01	0.449112E-02
Erreur L^2 intérieure	0.234338E-02	0.618670E-03	0.640346E-04
Erreur $H(rot)$	0.141093E+00	0.606411E-01	0.250898E-01
Erreur $H(rot)$ extérieure	0.141044E+00	0.606128E-01	0.250890E-01
Erreur $H(rot)$ intérieure	0.369926E-02	0.185147E-02	0.199354E-03

TAB. 7.11 – Erreurs sur un maillage de 27×27 éléments pour le premier cas test dépendant du temps.

Éléments finis d'arête du premier ordre			
Time t	4	8	12
Erreur L^2	0.223674E+00	0.779898E+00	0.105697E+01
Erreur L^2 extérieure	0.199908E+00	0.753085E+00	0.104881E+01
Erreur L^2 intérieure	0.100334E+00	0.202743E+00	0.131048E+00
Erreur $H(rot)$	0.481584E+00	0.149183E+01	0.218860E+01
Erreur $H(rot)$ extérieure	0.469498E+00	0.147570E+01	0.218318E+01
Erreur $H(rot)$ intérieure	0.107218E+00	0.218813E+00	0.153950E+00
Éléments finis d'arête du second ordre			
Time t	4	8	12
Erreur L^2	0.267213E-01	0.686831E-01	0.874091E-01
Erreur L^2 extérieure	0.255068E-01	0.680402E-01	0.867683E-01
Erreur L^2 intérieure	0.796434E-02	0.937600E-02	0.105641E-01
Erreur $H(rot)$	0.575619E-01	0.195273E+00	0.228699E+00
Erreur $H(rot)$ extérieure	0.569467E-01	0.195035E+00	0.228447E+00
Erreur $H(rot)$ intérieure	0.839323E-02	0.964150E-02	0.107185E-01
Éléments finis d'arête du troisième ordre			
Time t	4	8	12
Erreur L^2	0.275997E-02	0.615208E-02	0.775471E-02
Erreur L^2 extérieure	0.269564E-02	0.611029E-02	0.771919E-02
Erreur L^2 intérieure	0.592397E-03	0.715845E-03	0.741333E-03
Erreur $H(rot)$	0.160975E-01	0.203000E-01	0.235309E-01
Erreur $H(rot)$ extérieure	0.160859E-01	0.202872E-01	0.235192E-01
Erreur $H(rot)$ intérieure	0.610490E-03	0.720756E-03	0.742916E-03

TAB. 7.12 – Erreurs sur un maillage de 27×27 éléments pour le second cas test dépendant du temps.

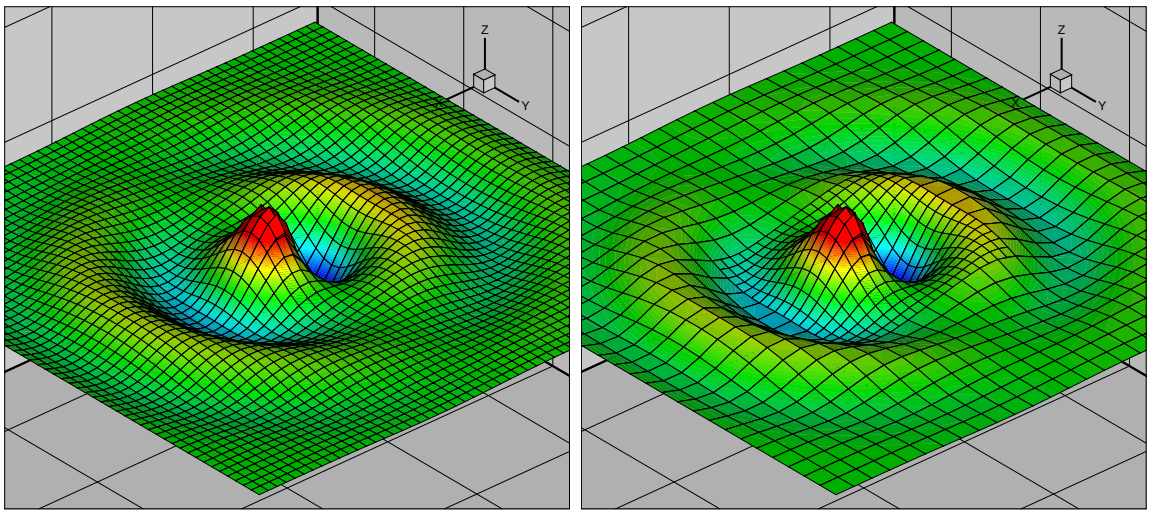

FIG. 7.7 – Seconde composante de la solution de référence et de la solution numérique après quarante-cinq itérations en temps pour le second cas test dépendant du temps.

7.5.7 Interprétation

Considérons dans un premier temps les résultats des cas tests stationnaires. Remarquons pour commencer que l'erreur globale (c'est-à-dire sur l'ensemble du domaine) se stabilise après un très petit nombre d'itérations. Cela s'explique naturellement par le fait que la deuxième itération de l'algorithme est déjà censée nous retourner une approximation de notre problème initial, les itérations successives ne faisant qu'améliorer cette solution en forçant la nullité de l'opérateur $T_H\lambda_H$.
En regardant de plus près nous pouvons préciser un peu les choses : il faut remarquer que la plus grande partie de l'erreur vient de ce que l'on a appelé l'erreur extérieure, c'est-à-dire l'erreur restreinte au domaine $\Omega\backslash\overline{\omega}$. Dans les cas tests qui nous intéressent vraiment, c'est-à-dire ceux pour lesquels la force f_2 admet de fortes variations de gradients, les erreurs intérieures et extérieures sont même à des échelles très différentes. Cela signifie que l'erreur supplémentaire que l'on introduit par l'utilisation de notre méthode deux échelles par rapport à une résolution directe sur un maillage uniformément raffiné, vient essentiellement du fait que le maillage grossier de $\Omega\backslash\overline{\omega}$ ne permet pas une représentation de la solution aussi fine que le maillage raffiné (ce dont il fallait de toute façon se douter).
Au regard des troisième et quatrième cas tests nous remarquons qu'il n'est pas seulement nécessaire d'assurer que l'espace de discrétisation est assez fin pour représenter le terme source (ce que l'on fait en définissant l'espace P_h dans ω) mais aussi que l'espace de discrétisation est assez fin pour représenter la solution sur l'ensemble du domaine Ω, sachant que l'incidence d'un terme source à fortes variations de gradient à support inclus dans ω portera aussi en dehors de ce support. Ce fait est mis en évidence en considérant un domaine ω légèrement plus grand que le support de f_2 dans le quatrième cas test, et en comparant les résultats de ce cas test à ceux du troisième pour voir de combien la résolution est améliorée.
Considérant les cas tests dépendants du temps, c'est-à-dire les cinquième et sixième cas

tests, signalons que la simulation semble visuellement bien se passer. En terme d'erreur remarquons que l'augmentation de l'ordre des schémas s'accompagne d'une diminution des normes d'erreur pour les deux cas tests. Cela signifie que l'erreur que l'on introduit par l'utilisation de notre méthode deux échelles, plutôt qu'une simulation directe sur un maillage uniformément raffiné, diminue avec l'augmentation de la capacité des schémas à capturer le terme source et à propager l'onde générée. Si l'erreur intérieure diminue fortement au cours du temps pour le cinquième cas test, c'est essentiellement parce que l'impulsion initiale sort du domaine ω. Par ailleurs on peut se rendre compte que cette diminution est d'autant plus nette que l'ordre du schéma est élevé : on vérifie donc là numériquement que la réflexion des ondes sortant du ω non-résolues sur l'espace de discrétisation $\Omega\backslash\overline{\omega}$ est d'autant amoindrie que cet espace est capable de représenter ces ondes.

Faute de temps nous n'avons pas poussé plus loin nos recherches sur l'efficacité de cette méthode deux échelles. Donnons par exemple quelques pistes qu'il aurait été intéressant de suivre :

- une étude approfondie du rapport entre le gain, en terme de temps de calcul et coût de stockage, et la perte de résolution liée à l'utilisation de la méthode,
- l'utilisation d'éléments finis d'ordres différents sur les maillages fin et grossier (un ordre plus bas sur le maillage fin, c'est-à-dire là où les fortes variations de gradient du terme source nécessite un raffinement du maillage plutôt que l'utilisation d'une approximation d'ordre élevé, et un ordre plus élevé sur le maillage grossier , c'est-à-dire là où l'onde propagée est plus lisse),
- l'adaptation de la méthode à l'équation des ondes et son implémentation en vue de la tester sur le cas test des tourbillons co-rotatifs.

Conclusions et perspectives

Le but de ces travaux de recherche a été de développer et d'implémenter des méthodes d'éléments finis conformes d'ordre élevé pour la simulation de propagation d'ondes.

Dans un premier temps, après avoir décrit les outils nécessaires pour les discrétisations en espace par éléments finis d'un ordre quelconque donné, c'est-à-dire la connaissance localement sur un élément de référence des degrés de liberté, des fonctions de base et des matrices de références, et s'être rendu compte que la génération d'un code de calcul d'ordre arbitrairement élevé (en espace tout du moins) ne nécessitait qu'un effort raisonnable d'implémentation (qui se résume dans la pratique à automatiser la gestion, pour un ordre quelconque d'élément fini, de tableaux de voisinages donnant la connection entre degrés de libertés), nous nous somme posés la question de l'efficacité de ces méthodes. Il en est ressorti que l'utilisation de méthodes d'éléments finis d'ordre élevé nécessite une réflexion sur la localisation (ou plutôt la "nature" puisque le terme de localisation n'a de sens que pour les éléments finis de Lagrange) des degrés de liberté. En effet nous avons remarqué dans le cadre des éléments finis d'arête triangulaires qu'un mauvais choix des formes linéaires définissant l'élément fini menait à des projections sur le sous-espace de discrétisation et des conditionnements de la matrice de masse très mauvais (il nous est même arrivés de perdre la convergence pour les discrétisations par éléments finis d'arête triangulaires du cinquième ordre). Si ce phénomène n'est pas apparu de manière aussi flagrante pour les discrétisations par éléments finis de Lagrange (bien qu'il n'y ait pas de doute sur le fait qu'il apparaîtra pour des éléments finis d'ordre plus élevé que ceux que l'on a testés), nous avons remarqué qu'il était possible d'augmenter l'efficacité des schémas simplement en relocalisant de manière réfléchie les points auxquels sont associés les degrés de liberté, plutôt que de les équirépartir.

Parallèlement à ces questions d'efficacité en terme d'optimisation d'erreur, nous nous sommes intéressés à l'efficacité en terme de temps de calcul des schémas. C'est dans cette optique que nous avons décrit un algorithme de construction d'éléments finis de Lagrange dont il est possible de condenser la matrice de masse, c'est-à-dire de remplacer la matrice de masse par une matrice diagonale, et donc d'en optimiser l'inversion. Si théoriquement cet algorithme nous laisse espérer la construction d'éléments finis de Lagrange condensés d'ordre quelconque, celui-ci ne nous a permis que de retrouver les versions condensées des éléments finis de Lagrange P_1 à P_5, et de déterminer une version condensée de l'éléments finis P_6. En effet la détermination des formules de quadratures nécessaires à la condensation de masse nécessite la résolution de systèmes polynomiaux dont le degré augmente

avec l'ordre des éléments. Ainsi, même s'il est possible d'exhiber un bon espace fonctionnel et formellement une bonne formule de quadrature qui nous permettraient d'envisager la condensation de la matrice de masse issue des éléments finis de Lagrange d'ordre quelconque, il nous est impossible, en toute généralité, de résoudre le système polynomial dont la solution détermine entièrement la formule de quadrature. Pour les méthodes d'ordre plus élevé nous avons toutefois développé des éléments finis partiellement condensés en redéfinissant les degrés de libertés de Lagrange n'intervenant pas dans la contrainte de conformité par des intégrales contre les fonctions d'une base orthogonale du sous-espace fonctionnel, de l'espace polynomial définissant l'élément fini, constitué des polynômes à trace nulle sur les trois arêtes. Ceci nous a permis de "creuser" le profil de la matrice de masse en orthogonalisant (de manière exacte cette fois, et non plus via l'utilisation d'une formule de quadrature) l'ensemble des fonctions de base, sauf les fonctions associées aux degrés de liberté imposant la conformité de l'élément fini entre elles, ce qui est dommageable dans la mesure où ce sont les interactions avec ces fonctions de base qui remplissent le profil de la matrice de masse de manière incontrôlée : si l'on parvenait à orthogonaliser ces fonctions de base entre elles et avec les fonctions de base associées au reste des degrés de liberté, les seules interactions entre fonctions de base se feraient localement sur chaque élément, et non plus aussi entre éléments voisins, de sorte que la matrice de masse serait diagonale par bloc (la taille de chaque bloc correspondant à la dimension de l'espace P_{k-3} pour les éléments finis P_k). Mais c'est bien en cela que réside tout le problème : comment orthogonaliser les fonctions des bases (ou plutôt définir les degrés de liberté qui les génèrent) tout en maîtrisant leur trace sur les arêtes pour assurer la conformité de l'élément ?

De la même manière notre intérêt s'est porté sur la condensation des éléments finis d'arête. Nous avons construit des éléments finis d'arête rectangulaires d'ordre quelconque de manière à pouvoir condenser la matrice de masse associée sur des maillage cartésiens. Si ceci peut paraître aller à l'encontre de la philosophie des éléments finis dont l'un des avantages majeurs est de permettre de traiter des domaines à géométrie complexe, nous avons montré que la construction des éléments finis d'arête triangulaires que l'on a fait permet un couplage conforme de ces deux types d'éléments finis sur un maillage hybride. Tant que l'on ne saura pas condenser les éléments finis d'arête triangulaires, il nous sera donc possible, par cette technique, d'optimiser le profil de la matrice de masse et donc d'en optimiser l'inversion.

À coté de ces travaux portant sur la discrétisation en espace par des méthodes d'éléments finis, il nous a fallu construire des discrétisations en temps d'ordre (arbitrairement) élevé adaptées à la résolution des systèmes différentiels ordinaires issus de la semi-discrétisation en espace des équations d'ondes considérées. Nous avons montré que si la procédure de Cauchy-Kowalewski, qui consiste à remplacer les dérivées temporelles successives de la solution du système dans le développement de Taylor en temps de cette solution entre deux pas de temps successifs par les dérivées spatiales via la définition même du système différentiel que l'on cherche à résoudre, n'est pas adaptée à nos problèmes semi-discrétisés dans la mesure où elle mène, suivant l'ordre de la discrétisation, à des schémas inconditionnellement instables, il était possible de les stabiliser en rajoutant un terme d'ordre plus élevé dans ledit développement de Taylor. Nous avons comparé ces discrétisations en temps à des

discrétisations en temps symplectiques et diagonalement implicites. Il en est ressorti que, si les discrétisations en temps diagonalement implicites nous permettent théoriquement d'utiliser des nombres CFL plus élevés, cela n'a aucun intérêt pratique dans la mesure où les schémas deviennent extrêmement dissipatifs. De plus ce type de discrétisation implique l'inversion, non plus de la matrice de masse, mais d'une combinaison linéaire de la matrice de masse et de raideur, de sorte que l'on perd aussi le bénéfice de la condensation de masse. Si les essais numériques ont montré que pour les schémas d'ordre bas il y a un véritable intérêt à l'utilisation de discrétisations en temps symplectiques, la dissipation des discrétisations en temps que nous avons développées étant trop importante, ils ont aussi montré que pour les discrétisations d'ordre élevé, l'utilisation de discrétisations en temps symplectiques n'est plus indispensable, la dissipation des discrétisations en temps que nous avons développées devenant négligeable. D'autant que les discrétisations en temps symplectique d'ordre élevé ne sont connues que par composition par elles-mêmes de discrétisations d'ordre plus bas, ce qui signifie une croissance beaucoup plus forte du nombre de pas de temps intermédiaires que pour les discrétisations en temps que nous avons développées.

Les travaux que l'on a mené au cours de cette thèse vont se prolonger naturellement dans le cadre du projet "HOUPIC" (High Order Finite Element Particle-In-Cell Solvers on Unstructured Grids) financé par l'Agence National de la Recherche. Il nous faudra non seulement étendre la construction que l'on a fait des éléments finis d'arête en trois dimensions d'espace dans le but de résoudre les équations de Maxwell 3D, mais optimiser l'inversion de la matrice de masse. La piste que l'on a suivie dans cette thèse, qui consiste à optimiser le profil de la matrice de masse en couplant des éléments finis d'arête rectangulaires condensés sur maillage cartésiens avec des éléments finis triangulaires nous paraît être une solution généralisable en trois dimensions d'espace. La généralisation de la condensation des éléments finis d'arête rectangulaires aux éléments finis de facette cubiques étant immédiate, le seul vrai problème réside dans le couplage des éléments finis de facette cubiques avec des éléments finis de facette tétraèdriques. La solution que l'on envisage est de construire des éléments finis de facette pyramidaux et d'utiliser ces éléments finis dans une zone tampon nous permettant un couplage conforme des trois types d'éléments finis considérés.

Quelques rappels sur les formules de quadrature de Gauss-Lobatto

Rappelons dans un premier temps la définition des polynômes de Legendre : ce sont les polynômes orthogonaux pour le produit scalaire

$$< P, Q >= \int_{-1}^{1} P(x)Q(x)\, dx$$

que l'on récupère par le procédé d'orthogonalisation de Gram-Schmidt à partir de la base canonique $\{1, x, x^2, x^3, \ldots\}$ et normalisés de sorte que leur évaluation en 1 vaut 1. Les 7 premiers polynômes de Legendre sont alors donnés par :

$$\begin{array}{l} P_0(x) = 1 \\ P_1(x) = x \\ P_2(x) = \frac{1}{2}(3x^2 - 1) \\ P_3(x) = \frac{1}{2}(5x^3 - 3x) \\ P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3) \\ P_5(x) = \frac{1}{8}(63x^5 - 70x^3 + 15x) \\ P_6(x) = \frac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5). \end{array}$$

Il est toutefois plus pratique de définir les polynômes de Legendre par la relation de récurrence :

$$\left\{ \begin{array}{rcl} P_0(x) & = & 1 \\ P_1(x) & = & x \\ (n+1)P_n(x) & = & (2n+1)xP_n(x) - nP_{n-1}(x) \ , \forall n \geq 1 \end{array} \right.$$

Les formules de quadrature de Gauss-Lobatto sont définies sur l'intervalle $[-1, 1]$. Elles ont la particularité de faire apparaître comme points de quadrature les extrémités de ce segment. Considérant la formule de quadrature de Gauss-Lobatto à N points $x_1 = -1$, x_2, $\ldots$, x_{N-1}, $x_N = 1$; la localisation des $N-2$ points intérieurs à $[-1, 1]$ est donnée par les zéros de la dérivée du $N-1^{\text{ième}}$ polynôme de Legendre P'_{N-1}. Ces zéros sont listés dans le tableau suivant.

Polynôme	Racines
P_2'	0
P_3'	$-\frac{1}{5}\sqrt{5}, \frac{1}{5}\sqrt{5}$
P_4'	$-\frac{1}{7}\sqrt{21}, 0, \frac{1}{7}\sqrt{21}$
P_5'	$-\frac{1}{21}\sqrt{147+42\sqrt{7}}, -\frac{1}{21}\sqrt{147-42\sqrt{7}}$ $\frac{1}{21}\sqrt{147-42\sqrt{7}}, \frac{1}{21}\sqrt{147+42\sqrt{7}}$
P_6'	$-\frac{1}{33}\sqrt{495+66\sqrt{15}}, -\frac{1}{33}\sqrt{495-66\sqrt{15}}, 0$ $\frac{1}{33}\sqrt{495-66\sqrt{15}}, \frac{1}{33}\sqrt{495+66\sqrt{15}}$

Les poids associés à ces points sont respectivement donnés par

$$w_1 = w_N = \frac{2}{N(N-1)}$$

et

$$w_i = \frac{2}{N(N-1)P_{N-1}(x_i)^2} \qquad i = 2, \ldots, N-1.$$

Une telle formule est d'ordre $2N-3$, cela signifie que tout les polynômes de degré inférieur ou égal à $2N-3$ seront intégrés exactement sur $[-1,1]$. Remarquons que les formules de quadrature de Gauss-Lobatto sont optimales en ce sens qu'aucune autre formule de quadrature à N points faisant apparaître les extrémités du segment ne peut être d'ordre supérieur (au sens large).

Sur l'imposition de conditions aux limites de Dirichlet

Notre problème modèle est le suivant :
Trouver u : $\Omega \to \mathbb{R}$ tel que

$$\left\{ \begin{array}{rll} -\triangle u & = f & (x,t) \in \Omega \\ u & = 0 & x \in \Gamma = \partial\Omega \end{array} \right.$$

qui devient sous forme variationnelle :
Trouver $u \in H_0^1(\Omega)$ tel que

$$\int_\Omega \nabla u \cdot \nabla \phi dx = \int_\Omega f\phi dx \quad \forall \phi \in H_0^1(\Omega)$$

Dans la pratique, pour la résolution de ce problème par éléments finis, une fois fixé un maillage de Ω, nous sommes tentés de définir un espace d'éléments finis V inclus dans $H_0^1(\Omega)$, de résoudre le problème intérieur à Ω et d'apposer simplement les valeurs nulles du bord à la solution. Mais ceci ne fonctionne pas, et pour cause, le support de f n'est pas nécessairement inclus dans Ω. En ne considérant qu'un espace $V \subset H_0^1(\Omega)$ on perd totalement les valeurs de f sur Γ, plus exactement, on décompose f, qui n'est pas nécessairement nulle sur Γ, dans une base d'éléments finis incluse dans $H_0^1(\Omega)$. Il y a donc là un décalage évident entre le nombre d'inconnues, qui sont les degrés de liberté intérieurs à Ω (valeurs nodales de u pour les éléments finis de Lagrange), et les données, évaluation des degrés de libertés associés à f sur $\overline{\Omega}$. Pour bien comprendre le problème que l'on cherche à résoudre, le plus simple est alors de se fixer les m premiers degrés de liberté, parmi les N au total, comme étant les degrés de liberté associés aux m noeuds du bord et d'écrire explicitement le système à résoudre. Cherchant $u = \sum_{j=1}^N u_j\phi_j$, et connaissant $f = \sum_{j=1}^N f_j\phi_j$, le problème devient :
Trouver $(u_i)_{i=1\ldots N}$ tel que :

$$\left\{ \begin{array}{rll} u_i & = 0 & \forall i = 1 \ldots m \\ \displaystyle\sum_{j=m+1}^N u_j \int_\Omega \nabla\phi_i \cdot \nabla\phi_j dX & = \displaystyle\sum_{j=1}^N f_j \int_\Omega \phi_i\phi_j dX & \forall i = m+1 \ldots N \end{array} \right. ,$$

ou encore sous forme matricielle :

$$\begin{pmatrix} Id_m & 0 \\ 0 & K_{ii} \end{pmatrix} \begin{pmatrix} \overline{u}_b \\ \overline{u}_i \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ M_{ib} & M_{ii} \end{pmatrix} \begin{pmatrix} \overline{f}_b \\ \overline{f}_i \end{pmatrix},$$

où

$$\overline{u}_b = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}, \quad \overline{u}_i = \begin{pmatrix} u_{m+1} \\ \vdots \\ u_N \end{pmatrix}, \quad \overline{f}_b = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}, \quad \overline{f}_i = \begin{pmatrix} f_{m+1} \\ \vdots \\ f_N \end{pmatrix},$$

Id_m désigne la matrice identité de dimension m,

$$K_{ii} = \begin{pmatrix} \int_\Omega \nabla\phi_{m+1} \cdot \nabla\phi_{m+1} dX & \dots & \int_\Omega \nabla\phi_{m+1} \cdot \nabla\phi_N dX \\ \vdots & \ddots & \vdots \\ \int_\Omega \nabla\phi_N \cdot \nabla\phi_{m+1} dX & \dots & \int_\Omega \nabla\phi_N \cdot \nabla\phi_N dX \end{pmatrix},$$

$$M_{ib} = \begin{pmatrix} \int_\Omega \phi_{m+1}\phi_1 dX & \dots & \int_\Omega \phi_{m+1}\phi_m dX \\ \vdots & \ddots & \vdots \\ \int_\Omega \phi_N\phi_1 dX & \dots & \int_\Omega \phi_N\phi_m dX \end{pmatrix},$$

et

$$M_{ii} = \begin{pmatrix} \int_\Omega \phi_{m+1}\phi_{m+1} dX & \dots & \int_\Omega \phi_{m+1}\phi_N dX \\ \vdots & \ddots & \vdots \\ \int_\Omega \phi_N\phi_{m+1} dX & \dots & \int_\Omega \phi_N\phi_N dX \end{pmatrix}.$$

Reste alors à résoudre ce système soit de manière optimisée, c'est-à-dire en générant effectivement les matrices qui y apparaissent, soit de manière détournée : les matrices qui sont générées par les codes d'éléments finis sont en général les matrices de masse $(M_{ij})_{i,j=1\dots N} = (\int_\Omega \phi_i\phi_j dX)_{i,j=1\dots N}$ et de raideur $(K_{ij})_{i,j=1\dots N} = (\int_\Omega \nabla\phi_i \cdot \nabla\phi_j dX)_{i,j=1\dots N}$, ces matrices étant symétriques, les codes optimisés ne stockent qu'une des deux parties triangulaires de celle-ci ; or la matrice apparaissant au second membre n'est pas symétrique, donc plutôt que de repenser spécifiquement la gestion de cette matrice (et de son produit par un vecteur), il peut sembler préférable de calculer le second membre à partir de la matrice (M_{ij}) puis d'imposer la nullité des degrés de liberté associés aux noeuds du bord ...

Passons maintenant à un problème légèrement plus général :
Trouver u : $\Omega \to \mathbb{R}$ tel que

$$\begin{cases} -\triangle u &= f & (x,t) \in \Omega \\ u &= g & x \in \Gamma = \partial\Omega \end{cases}$$

D'un point de vue théorique ce problème se résout par un relèvement de u, c'est-à-dire par un changement de fonction inconnue, en définissant : $v = u - \tilde{g}$ où $\tilde{g}$ désigne toute fonction dont la trace sur Γ vaut g. Le problème devient alors :
Trouver v : $\Omega \to \mathbb{R}$ tel que

$$\begin{cases} -\triangle v &= f + \triangle\tilde{g} & (x,t) \in \Omega \\ v &= 0 & x \in \Gamma = \partial\Omega \end{cases},$$

ou encore sous forme variationnelle :
Trouver $v \in H_0^1(\Omega)$ tel que

$$\int_\Omega \nabla v.\nabla\phi dx = \int_\Omega f\phi dx - \int_\Omega \nabla\tilde{g}\cdot\nabla\phi dx \quad \forall\phi \in H_0^1(\Omega)\ .$$

Dans la pratique nous définissons bien entendu $\tilde{g}$ comme étant le prolongement de g nul sur les degrés de liberté associés aux noeuds intérieurs à Ω. De nouveau en faisant bien attention au fait que ni f ni $\tilde{g}$ n'ont de raison d'être nuls sur Γ, nous obtenons, en gardant les mêmes conventions de notation, le système suivant à résoudre :

$$\begin{pmatrix} Id_m & 0 \\ 0 & K_{ii} \end{pmatrix}\begin{pmatrix} \overline{v}_b \\ \overline{v}_i \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ M_{ib} & M_{ii} \end{pmatrix}\begin{pmatrix} \overline{f}_b \\ \overline{f}_i \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ K_{ib} & K_{ii} \end{pmatrix}\begin{pmatrix} \overline{g}_b \\ 0 \end{pmatrix}.$$

À nouveau si la matrice apparaissant au premier membre s'avère aisée à construire, à partir de la matrice K, et surtout à manipuler (en terme d'inversion) puisque celle-ci reste symétrique, il peut être plus judicieux de travailler avec les matrice M et K pour le second membre et d'imposer la nullité des degrés de liberté associés aux noeuds du bord. Remarquons qu'écrit sous cette forme, le système nous permet de voir que la solution u ne dépend effectivement pas du prolongement de g dans Ω choisi. Pour s'en persuader il suffit de le réécrire en terme de u et de considérer un prolongement de g non nécessairement nul :

$$\begin{pmatrix} Id_m & 0 \\ 0 & K_{ii} \end{pmatrix}\begin{pmatrix} \overline{u}_b - \overline{g}_b \\ \overline{u}_i - \overline{g}_i \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ M_{ib} & M_{ii} \end{pmatrix}\begin{pmatrix} \overline{f}_b \\ \overline{f}_i \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ K_{ib} & K_{ii} \end{pmatrix}\begin{pmatrix} \overline{g}_b \\ \overline{g}_i \end{pmatrix}$$

pour voir que les seuls termes faisant apparaître $\overline{g}_i$ se simplifient. Le système peut alors se réécrire sous la forme suivante :

$$\begin{pmatrix} Id_m & 0 \\ 0 & K_{ii} \end{pmatrix}\begin{pmatrix} \overline{u}_b \\ \overline{u}_i \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ M_{ib} & M_{ii} \end{pmatrix}\begin{pmatrix} \overline{f}_b \\ \overline{f}_i \end{pmatrix} - \begin{pmatrix} -Id_m & 0 \\ K_{ib} & K_{ii} \end{pmatrix}\begin{pmatrix} \overline{g}_b \\ 0 \end{pmatrix}$$

et c'est ce dernier que l'on résout dans la pratique, en utilisant les matrices M et K pour calculer le second membre puis en imposant, non plus la nullité des degrés de liberté associés aux noeuds du bord, mais l'évaluation sur g de ces degrés de liberté.

Regardons maintenant comment se traduisent ces considérations sur la résolution de l'équation d'onde. Considérons le problème suivant :
Trouver u : $\Omega \times \mathbb{R}^+ \to \mathbb{R}$ tel que

$$\begin{cases} \partial_t^2 u - \triangle u = f & (x,t) \in \Omega\times\mathbb{R}^+ \\ u(x,t) = g(x,t) & (x,t) \in \Gamma = \partial\Omega\times\mathbb{R}^+ \\ + \text{ Conditions initiales} \end{cases}$$

Nous faisons un relèvement de u en définissant $v = u - g$, le problème devient :
Trouver v : $\Omega \times \mathbb{R}^+ \to \mathbb{R}$ tel que

$$\begin{cases} \partial_t^2 v - \triangle v = f - \partial_t^2\tilde{g} + \triangle\tilde{g} & (x,t) \in \Omega\times\mathbb{R}^+ \\ v(x,t) = 0 & (x,t) \in \Gamma = \partial\Omega\times\mathbb{R}^+ \\ + \text{ Conditions initiales} \end{cases}$$

puis sous sa forme variationnelle :
Trouver $v(x,.) : \mathbb{R}^+ \to H_0^1(\Omega)$ tel que

$$\frac{d}{dt^2}\int_\Omega v\phi dx + \int_\Omega \nabla v \cdot \nabla\phi dx = \int_\Omega f\phi dx - \frac{d}{dt^2}\int_\Omega g\phi dx - \int_\Omega \nabla g \cdot \nabla\phi dx \quad \forall \phi \in H_0^1(\Omega) \ .$$

Une fois fixés un maillage de Ω et un espace d'éléments finis $V \subset H^1(\Omega)$ (et non strictement inclus dans $H_0^1(\Omega)$ puisque ni f ni g ne sont en général nuls sur Γ) nous réécrivons dans un premier temps le problème sous la forme du système d'équation qu'il faut effectivement résoudre après décomposition des fonctions (inconnues et données) sur une base de V :

$$\left\{\begin{array}{rcll} v_i & = & 0 & ,\forall i = 1 \dots m \\ \displaystyle\sum_{j=m+1}^{N} \frac{dv_j}{dt^2}\int_\Omega \phi_i\phi_j dX & + & \displaystyle\sum_{j=m+1}^{N} v_j \int_\Omega \nabla\phi_i \cdot \nabla\phi_j dX = & \\ \displaystyle\sum_{j=1}^{N} f_j \int_\Omega \phi_i\phi_j dX - \sum_{j=1}^{N} \frac{dg_j}{dt^2}\int_\Omega \phi_i\phi_j dX & - & \displaystyle\sum_{j=1}^{N} g_j \int_\Omega \nabla\phi_i \cdot \nabla\phi_j dX & ,\forall i = m+1 \dots N, \end{array}\right.$$

puis sous forme matricielle

$$\begin{array}{rcl} \begin{pmatrix} 0 & 0 \\ 0 & M_{ii} \end{pmatrix}\begin{pmatrix} \ddot{\overline{v}}_b \\ \ddot{\overline{v}}_i \end{pmatrix} + \begin{pmatrix} Id_m & 0 \\ 0 & K_{ii} \end{pmatrix}\begin{pmatrix} \overline{v}_b \\ \overline{v}_i \end{pmatrix} & = & \\ \begin{pmatrix} 0 & 0 \\ M_{ib} & M_{ii} \end{pmatrix}\begin{pmatrix} \overline{f}_b \\ \overline{f}_i \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ M_{ib} & M_{ii} \end{pmatrix}\begin{pmatrix} \ddot{\overline{g}}_b \\ 0 \end{pmatrix} & - & \begin{pmatrix} 0 & 0 \\ K_{ib} & K_{ii} \end{pmatrix}\begin{pmatrix} \overline{g}_b \\ 0 \end{pmatrix} \end{array}$$

que l'on réécrit en terme de u, sachant que $\overline{g}_i$ est imposé comme étant nul au cours du temps, $\ddot{\overline{g}}_i$ l'est automatiquement aussi, de même que $\ddot{\overline{v}}_b$ par définition, il reste

$$\begin{array}{rcl} \begin{pmatrix} 0 & 0 \\ 0 & M_{ii} \end{pmatrix}\begin{pmatrix} 0 \\ \ddot{\overline{u}}_i \end{pmatrix} + \begin{pmatrix} Id_m & 0 \\ 0 & K_{ii} \end{pmatrix}\begin{pmatrix} \overline{u}_b - \overline{g}_b \\ \overline{u}_i \end{pmatrix} & = & \\ \begin{pmatrix} 0 & 0 \\ M_{ib} & M_{ii} \end{pmatrix}\begin{pmatrix} \overline{f}_b \\ \overline{f}_i \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ M_{ib} & M_{ii} \end{pmatrix}\begin{pmatrix} \ddot{\overline{g}}_b \\ 0 \end{pmatrix} & - & \begin{pmatrix} 0 & 0 \\ K_{ib} & K_{ii} \end{pmatrix}\begin{pmatrix} \overline{g}_b \\ 0 \end{pmatrix}, \end{array}$$

pour finir après simplification :

$$\begin{array}{rcl} \begin{pmatrix} 0 & 0 \\ 0 & M_{ii} \end{pmatrix}\begin{pmatrix} 0 \\ \ddot{\overline{u}}_i \end{pmatrix} + \begin{pmatrix} Id_m & 0 \\ 0 & K_{ii} \end{pmatrix}\begin{pmatrix} \overline{u}_b \\ \overline{u}_i \end{pmatrix} & = & \\ \begin{pmatrix} 0 & 0 \\ M_{ib} & M_{ii} \end{pmatrix}\begin{pmatrix} \overline{f}_b \\ \overline{f}_i \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ M_{ib} & M_{ii} \end{pmatrix}\begin{pmatrix} \ddot{\overline{g}}_b \\ 0 \end{pmatrix} & - & \begin{pmatrix} -Id_m & 0 \\ K_{ib} & K_{ii} \end{pmatrix}\begin{pmatrix} \overline{g}_b \\ 0 \end{pmatrix}. \end{array}$$

Routine MAPLE© pour la génération des fonctions de base et des matrices

```
> Basis_and_matrices_calculation:=proc(k)

  local k1,m,j,i,l:
  global lambda,a,A,IJXX,IJYY,IJXYYX:

  A:=matrix((k+1)*(k+2)/2,(k+1)*(k+2)/2):
  IJXX:=matrix((k+1)*(k+2)/2,(k+1)*(k+2)/2):
  IJYY:=matrix((k+1)*(k+2)/2,(k+1)*(k+2)/2):
  IJXYYX:=matrix((k+1)*(k+2)/2,(k+1)*(k+2)/2):

  lambda[1]:=(x,y)->1-x-y:
  lambda[2]:=(x,y)->x:
  lambda[3]:=(x,y)->y:
  lambda[4]:=lambda[1]:

  if k=1 then

    a[1]:=lambda[1]:
    a[2]:=lambda[2]:
    a[3]:=lambda[3]:

  fi:

  if k>1 then

    k1:=iquo(k,3):
    if modp(k,3)=0 then

      for m from 0 to (k1-1) do
      for j from 0 to (k-(3*m+1)) do
      for i from 1 to 3 do
```

```
a[(3*k-9*(m-1)/2)*m+3*j+i]:=
             product(lambda[1]-l/k,l=0..m-1)*
             product(lambda[2]-l/k,l=0..m-1)*
             product(lambda[3]-l/k,l=0..m-1)*
             product(lambda[i  ]-l/k,l=m..k-2*m-(j+1))*
             product(lambda[i+1]-l/k,l=m..j+m-1      )*
             1/(product(l/k,l=1..m)*
             product(l/k,l=      j+1..    m+j)*
             product(l/k,l=k-3*m-j+1..k-2*m-j)*
             product(l/k,l=        1..k-3*m-j)*
             product(l/k,l=        1..      j)):

     od:od:od:

a[(k+1)*(k+2)/2]:=product(lambda[1]-l/k,l=0..k1-1)*
                  product(lambda[2]-l/k,l=0..k1-1)*
                  product(lambda[3]-l/k,l=0..k1-1)*
                  1/(product(l/k,l=1..k1))^3:

   else

     for m from 0 to k1 do
     for j from 0 to (k-(3*m+1)) do
     for i from 1 to 3 do

a[(3*k-9*(m-1)/2)*m+3*j+i]:=
             product(lambda[1]-l/k,l=0..m-1)*
             product(lambda[2]-l/k,l=0..m-1)*
             product(lambda[3]-l/k,l=0..m-1)*
             product(lambda[i  ]-l/k,l=m..k-2*m-(j+1))*
             product(lambda[i+1]-l/k,l=m..      j+m-1)*
             1/(product(l/k,l=1..m)*
             product(l/k,l=      j+1..    m+j)*
             product(l/k,l=k-3*m-j+1..k-2*m-j)*
             product(l/k,l=        1..k-3*m-j)*
             product(l/k,l=        1..      j)):

     od:od:od:

   fi:
 fi:

 for i from 1 to (k+1)*(k+2)/2 do
   for j from 1 to (k+1)*(k+2)/2 do
```

```
A[i,j]:=int(int(a[i](x,y)*a[j](x,y),y=0..1-x),x=0..1);
IJXX[i,j]:=int(int(diff(a[i](x,y),x)*diff(a[j](x,y),x),y=0..1-x),x=0..1):
IJYY[i,j]:=int(int(diff(a[i](x,y),y)*diff(a[j](x,y),y),y=0..1-x),x=0..1):
IJXYYX[i,j]:=int(int(
                    (diff(a[i](x,y),x)*diff(a[j](x,y),y))
                   +(diff(a[i](x,y),y)*diff(a[j](x,y),x))
                     ,y=0..1-x),x=0..1):

   od:
 od:

 end proc:
```

Bibliographie

[1] Y. Achdou, Y. Maday,
The mortar element method with overlapping subdomains,
SIAM J. Numer. Anal., Vol. 40, No. 2, pp. 601-628, 2002.

[2] R. A. Adams, J. Fournier,
Sobolev Spaces (2nd edn.),
Academic Press, 2003.

[3] R. M. Alford, K. R. Kelly, D. M. Boore,
Accuracy of finite difference modeling of the acoustic wave equation,
Geophysics, Vol. 39, Issue 6, pp. 834-842, 1974.

[4] F. Assous, P. Ciarlet Jr.,
Modèles et méthodes pour les équations de Maxwell,
Rapport de Recherche ENSTA 347, 2001.

[5] R. Barthelmé,
Le problème de conservation de la charge dans le couplage des équations de Vlasov et de Maxwell,
Thèse, Université Louis Pasteur, Strasbourg, 2005.

[6] F. B. Belgacem,
The Mortar finite element method with Lagrange multipliers,
Numer. Math., Vol. 84, pp. 173-197, 1999.

[7] C. Bernardi , Y. Maday , A. T. Patera,
A New Non Conforming Approach to Domain Decomposition : The Mortar Element Method,
Collège de France Seminar, Pitman, H. Brezis, J.-L. Lions, 1990.

[8] H. Brezis,
Analyse fonctionnelle. Théorie et applications,
-Paris : Masson, 1983.

[9] J. C. Butcher,
Numerical methods for ordinary differential equations,
-Chichester : John Wiley, 2003.

[10] J. Candy and W. Rozmus,
A symplectic integration algorithm for seperable Hamiltonian functions,
J. Comput. Phys., Vol. 92, pp. 230-256, 1991.

[11] N. Canouet,
Schémas multi-échelles pour la résolution numérique des équations de Maxwell,
Thèse, École Nationale des Ponts et Chaussées, 2003.

[12] N. Canouet, L. Fezoui, S. Piperno,
Discontinuous Galerkin Time-Domain solution of Maxwell's equations on locally-refined nonconforming Cartesian grids,
COMPEL, vol. 24, No. 4, pp. 1381-1401, 2005.

[13] P. J. Channell, and C. Scovel,
Symplectic integration of Hamiltonian systems,
Nonlinearity 3, pp. 231-259, 1990.

[14] M. J. S. Chin-Joe-Kong, W. A. Mulder, M. Van Veldhuizen,
Higher-order triangular and tetrahedral finite elements with mass lumping for solving the wave equation,
Journal of Engineering Mathematics, Vol. 35, No.4, pp. 405-426, 1999.

[15] P. Ciarlet Jr.,
Augmented formulations for solving Maxwell equations,
Comput. Methods Appl. Mech. Eng., Vol. 194, No. 2-5, 559-586, 2005.

[16] P. G. Ciarlet,
Introduction à l'analyse numérique matricielle et à l'optimisation,
Masson, Paris, 1982.

[17] P. G. Ciarlet,
The Finite Element Method for Elliptic Problems,
North-Holland, Amsterdam, 1978.

[18] B. Cockburn, S. Hou, C.-W. Shu,
The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws IV : the multidimensional case,
Mathematics of Computation, Vol. 54, pp. 545-581, 1990.

[19] B. Cockburn, S.-Y. Lin, C.-W. Shu,
TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws III : one dimensional systems,
J. Comput. Phys., Vol. 84, pp. 90-113, 1989.

[20] B. Cockburn, C.-W. Shu,
The Runge-Kutta local projection P1-discontinuous Galerkin finite element method for scalar conservation laws,
Mathematical Modelling and Numerical Analysis, Vol. 25, pp. 337-361, 1991.

[21] B. Cockburn, C.-W. Shu,
The Runge-Kutta discontinuous Galerkin method for conservation laws V : multidimensional systems,
J. Comput. Phys., Vol. 141, pp. 199-224, 1998.

[22] B. Cockburn, C.-W. Shu,
TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws II : general framework,
Mathematics of Computation, Vol. 52, pp. 411-435, 1989.

[23] G. Cohen,
Higher-Order Numerical Methods for Transient Wave Equation,
Springer-Verlag, 2001.

[24] G. Cohen, X. Ferrieres et S. Pernet,
Discontinuous Galerkin methods for Maxwell's equations in the time domain,
Comptes Rendus Physique, Vol. 7, Issue 5, pp. 494-500, 2006.

[25] G. Cohen, P. Joly, J. E. Roberts et N. Tordjman,
Construction and analysis of higher order finite elements with mass lumping for the wave equation,
In : R. Kleinman, T. Angell, D. Colton, F. Santosa and I.stakgold (eds), Proc. of the 2nd International Conference on Mathematical and Nuerical Aspect of Wave Propagation. Philadelphia : SIAM, pp. 152-160, 1993.

[26] G. Cohen, P. Joly, J. E. Roberts et N. Tordjman,
Higher order triangular finite element with mass lumping for the wave equation,
SIAM J. Numer. Anal., Vol. 38, No. 6, pp. 2047-2078, 2001.

[27] G. Cohen, P. Monk
Efficient Edge Finite Element Schemes in Computational Electromagnetism,
Proc. of the 3rd Conf. on Mathematical and Numerical Aspects of Wave Propagation Phenomena, SIAM, april 1995.

[28] G. Cohen, P. Monk
Gauss point mass lumping schemes for Maxwell's equations,
NMPDE Journal, Vol. 14, No. 1, pp. 63-88, 1998.

[29] F. Collino, T. Fouquet, P. Joly,
Conservative space-time mesh refinement methods for the FDTD solution of Maxwell's equations,
J. Comput. Phys., Vol. 211, Issue 1, pp. 9-35 , 2006.

[30] G. J. Cooper, A. Sayfy,
Semiexplicit A-stable Runge-Kutta methods,
Mathematics of Computation, Vol. 33, pp. 541-556, 1979.

[31] M. A. Dablain,
The application of high order differencing for the scalar wave equation,
Geophysics, Vol. 51, Issue 1, pp. 54-66, 1986.

[32] R. Dautray, J.-L. Lions,
Analyse mathématique et calcul numérique pour les sciences et les techniques,
Vol. 8, Masson, Paris, 1988.

[33] P. J. Davis,
Interpolation and Approximation,
Dover Publications, New York, 1975.

[34] M. Dumbser, C.-D. Munz,
ADER Discontinuous Galerkin Schemes for Aeroacoustics,
C. R. Mecanique, Vol. 333, Issue 9, pp. 683-687, 2005.

[35] M. Dumbser, C.-D. Munz,
Arbitrary High Order Discontinuous Galerkin Schemes,

Numerical methods for hyperbolic and kinetic problems, eds. S. Cordier, T. Goudon, M. Gutnic, E. Sonnendrucker, IRMA series in mathematics and theoretical physics, European Mathematical Society, 2005.

[36] M. Duruflé,
Intégration numérique et éléments finis d'ordre élevé appliqués aux équations de Maxwell en régime harmonique,
Thèse, Université Paris-Dauphine, 2006.

[37] A. Elmkies, P. Joly,
Eléments finis et condensation de masse pour les équations de Maxwell : le cas 2D,
Rapport de Recherche INRIA, No. 3035, 1996.

[38] B. Engquist, A. Majda,
Absorbing boundary conditions for the numerical simulation of waves,
Mathematics of Computation, Vol. 31, No. 139, pp. 629-651, 1977.

[39] G. J. Fix,
Effects of quadrature errors in the finite element approximation of steady state, eigen value and parabolic problems,
Mathematical Foundation of the Finite Element Methode with Application to Partial Differential Equations, A.K Aziz, ed., Academic Press, New York and London, pp. 525-556, 1972.

[40] I. Fried, D. S. Malkus,
Finite element mass matrix lumping by numerical integration with no convergence rate loss,
Int. J. Solids Struc., Vol. 11, Issue 4, pp. 461-466, 1975.

[41] R. Glowinski, J. He, J. Rappaz & J. Wagner,
A multi-domain method for solving numerically multi-scale elliptic problems,
C. R. Acad. Sci. Paris, Ser. I 338, 2004.

[42] F. X. Giraldo, M. A. Taylor,
A diagonal-mass-matrix triangular-spectral-element method based on cubature points,
Journal of Engineering Mathematics, Vol. 56, No.3, pp. 307-322, 2006.

[43] E. Hairer, S. P. Noersett, G. Wanner,
Solving ordinary differential equations I,
-Berlin : Springer, 1987

[44] E. Hairer, G. Wanner,
Solving ordinary differential equations II,
-Berlin, Heidelberg, New York : Springer, 1991.

[45] A. Harten, B. Engquist, S. Osher, S. Chakravarthy,
Uniformly high order essentially non-oscillatory schemes, III,
J. Comput. Phys., Vol. 71,pp. 231-303, 1987.

[46] J. S. Hesthaven,
From Electrostatics to Almost Optimal Nodal Sets for Polynomial Interpolation in a Simplex,
SIAM J. Nuner. Anal., Vol. 35, No. 2, pp. 655-676, 1998.

[47] J. S. Hesthaven, D. Gottlieb
Stable spectral methods for conservation laws on triangles with unstructured grids,
Comput. Methods Appl. Mech. Engrg., Vol. 175, pp. 361-381, 1999.

[48] J. S. Hesthaven, T. Warburton,
Nodal High-Order Methods on Unstructured Grids, I. Time-Domain Solution of Maxwell's Equations,
J. Comput. Phys., Vol. 181, pp. 186-221, 2002.

[49] J. S. Hesthaven, T. Warburton
High-order nodal discontinuous Galerkin methods for the Maxwell eigenvalue problem,
Philos. Transact. A Math. Phys. Eng. Sci., Vol. 362, pp.493-524, 2004.

[50] H. M. Jurgens, D. W. Zingg,
Numerical Solution of the Time-Domain Maxwell Equations Using High-Accuracy Finite-Difference Methods,
SIAM J. Sci. Comput., Vol. 22, Issue 5, pp. 1675-1696, 2000.

[51] P. Lacoste,
Les éléments finis des équations de Maxwell dans le code PALAS. Eléments finis nouveaux pour le cadre axisymétrique. La condensation des matrices masses,
Thèse, Université de Bordeaux I, 1994.

[52] E. L. Lindmann,
Free-space boundary conditions for the time dependant wave equation,
J. Comput. Phys., Vol. 18, pp. 66-78, 1975.

[53] J.-L. Lions, E. Magenes,
Problème aux limites non homogènes et applications,
Vol. 1, Dunod, 1968.

[54] P. L. Lions,
On the Schwarz alternating method III : a variant for nonoverlapping subdomains,
Third International Symposium on Domain Decomposition Methods for Partial Differential Equations, T. F. Chan, R. Glowinski, J. Periaux, and O. B. Widlund (eds.), SIAM, pp. 202-223, 1989.

[55] P. Monk
Finite Element Methods for Maxwell's Equations,
Oxford University Press, 2003.

[56] W. A. Mulder,
Higher order mass-lumped finite elements for the wave equation,
J. Comput. Acoust., Vol. 9, No.2, pp. 671-680, 2001

[57] J. C. Nédélec,
Mixed finite elements in $\mathbb{R}^3$,
Numer. Math., Vol. 35, pp. 315-341, 1980.

[58] J. C. Nédélec,
A new family of mixed finite elements in $\mathbb{R}^3$,
Numer. Math., Vol. 50, pp. 57-81, 1986.

[59] S. Piperno,
Symplectic local time-stepping in non-dissipative DGTD methods applied to wave pro-

pagation problems,
Mathematical Modelling and Numerical Analysis, Vol. 40 No. 5, pp. 815-841, 2006.

[60] A. Quarteroni, A. Valli,
Domain Decomposition Methods for Partial Differential Equations,
Oxford University Press, Oxford, 1999.

[61] P.A Raviart et J.M Thomas,
Introduction à l'analyse numérique des équations aux dérivées partielles,
Masson, Paris, 1983.

[62] W. H. Reed, T. R. Hill
Triangular mesh methods for neutron transport equation,
Tech. Report LA-UR-73-479, Los Alamos Scientific Laboratory, 1973.

[63] R. Rieben, D. White, and G. Rodrigue,
High Order Symplectic Integration Methods for Finite Element Solutions to Time Dependent Maxwell Equations,
IEEE Transactions on Antennas and Propagation, Vol. 52, No. 8, pp. 2190-2195, 2004.

[64] D. Ruth,
A canonical integration technique,
IEEE Trans. Nucl. Sci., Vol. NS-30, pp. 2669-2671, 1983.

[65] P. Silvester,
Symmetric quadrature formulae for simplexes,
Mathematics of Computation, Vol. 24, pp. 95-100, 1970.

[66] B. Strand,
Simulations of Acoustic Wave Phenomena Using High-Order Finite Difference Approximations,
SIAM J. Sci. Comput., Vol. 20, Issue 5, pp. 1585-1604, 1999.

[67] A. H. Stroud,
Approximate Calculation of Multiple Integrals,
Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1971.

[68] P. Le Tallec, T. Sassi,
Domain decomposition with nonmatching grids : augmented Lagrangian approach,
Math. Comp., Vol. 64, pp. 1367-1396, 1995.

[69] V. A. Titarev, E. F. Toro,
ADER : Arbitrary High Order Godunov Approach,
Journal of Scientific Computing, Vol. 17, pp. 609-618, 2002.

[70] N. Tordjman,
Eléments finis d'ordre élevés avec condensation de masse pour l'équation des ondes,
Thèse, Université Paris IX Dauphine, Paris, 1995.

[71] S. Wandzurat, H. Xiao,
Symmetric quadrature rules on a triangle,
Comput. and Math. with Applications, Vol.45, Issue 12, pp. 1829-1840, 2003.

[72] K. S. Yee,
Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media,
IEEE Trans. Ant. Propagat., Vol. 14, 302, 1966.

[73] H. Yoshida,
Construction of higher order symplectic integrators,
Physics Letters A, Vol. 150, Issues 5-7, pp. 262-268, 12 November 1990.

[74] *Maple*©,
http ://www.maplesoft.com