



# Intégration numérique et éléments finis d'ordre élevé appliqués aux équations de Maxwell en régime harmonique

Marc Duruflé

## ► To cite this version:

Marc Duruflé. Intégration numérique et éléments finis d'ordre élevé appliqués aux équations de Maxwell en régime harmonique. Modélisation et simulation. ENSTA ParisTech, 2006. Français.  
tel-00068590

HAL Id: tel-00068590

<https://pastel.archives-ouvertes.fr/tel-00068590>

Submitted on 12 May 2006

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

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

année : 2006

N° attribué par la bibliothèque



# THÈSE

présentée à

**UNIVERSITÉ PARIS DAUPHINE**

pour obtenir le titre de  
DOCTEUR EN SCIENCES

Spécialité

Mathématiques appliquées

soutenue par

**Marc DURUFLE**

le 7 février 2006

Titre

**Intégration numérique et  
éléments finis d'ordre élevé  
appliqués aux équations de Maxwell  
en régime harmonique.**

Directeur de thèse : Gary Cohen

Jury

Rapporteurs : M. **Martin Costabel**  
M. **Marcus Grote**

Suffragants : M. **David Levadoux**  
M. **Peter Monk**  
M. **Gabriel Turinici**



“l’Université n’entend donner aucune approbation, ni improbation aux opinions émises dans les thèses : ces opinions doivent être considérées comme propres à leurs auteurs.”



# Table des matières

<b>Introduction</b>	<b>9</b>
Problématique . . . . .	9
Méthodes classiques utilisées . . . . .	9
Eléments finis d'arête . . . . .	10
Autres méthodes éléments finis . . . . .	10
Solution proposée et plan de la thèse . . . . .	11
<b>I Equation de Helmholtz</b>	<b>13</b>
<b>1 Résolution de l'équation de Helmholtz par éléments finis spectraux</b>	<b>15</b>
1.1 Formulation variationnelle classique et éléments finis spectraux . . . . .	17
1.1.1 Cas modèle traité . . . . .	17
1.1.2 Formulation variationnelle . . . . .	18
1.1.3 Expression des matrices en 2-D et 3-D, utilisation de la condensation de masse . . . . .	21
1.2 Éléments finis courbes . . . . .	24
1.2.1 Cas 2-D . . . . .	24
1.2.2 Cas 3-D . . . . .	26
1.3 Précision de la méthode . . . . .	29
1.3.1 Cas 2-D . . . . .	29
1.3.2 Cas 3-D . . . . .	35
1.4 Application aux filtres optiques, intérêt de l'ordre élevé . . . . .	38
1.4.1 Faisceaux gaussiens . . . . .	38
1.4.2 Propriétés du filtre optique . . . . .	39
1.4.3 Apologie de l'ordre élevé . . . . .	46
1.5 Conclusion . . . . .	48
<b>2 Algorithmes itératifs de résolution</b>	<b>49</b>
2.1 Formulation mixte . . . . .	50
2.1.1 Formulation variationnelle, propriétés des matrices . . . . .	50
2.1.2 Intérêt de la factorisation . . . . .	52
2.2 Résolution directe . . . . .	55
2.3 Résolution itérative . . . . .	58
2.3.1 Cas 2-D . . . . .	58
2.3.2 Cas 3-D . . . . .	61
2.4 Préconditionnement . . . . .	63
2.4.1 Préconditionnement par l'équation de Helmholtz avec amortissement . . . . .	64
2.4.2 Décomposition en sous-domaines . . . . .	70

2.4.3	Solveur itératif ? . . . . .	73
2.5	Conclusion . . . . .	73
<b>3</b>	<b>Comparaison hexaèdres / tétraèdres</b>	<b>75</b>
3.1	Analyse de dispersion . . . . .	76
3.2	Cas académique de la sphère . . . . .	84
3.2.1	Coût du produit matrice-vecteur . . . . .	84
3.2.2	Comparaison sur une sphère parfaitement conductrice . . . . .	85
3.2.3	Comparaison sur une sphère diélectrique . . . . .	88
3.3	Résultats numériques sur des cas plus complexes . . . . .	88
3.3.1	Cavité cobra . . . . .	89
3.3.2	Cone-sphère revêtu . . . . .	91
3.4	Conclusion . . . . .	92
<b>II</b>	<b>Equations de Maxwell 2-D et 3-D</b>	<b>95</b>
<b>4</b>	<b>Seconde famille de Nédélec sur les quadrillatères</b>	<b>97</b>
4.1	Formulation mixte des équations de Maxwell . . . . .	98
4.1.1	Formulation variationnelle standard . . . . .	98
4.1.2	Expression des matrices élémentaires . . . . .	100
4.1.3	Une formulation variationnelle mixte possible . . . . .	101
4.1.4	Intérêt de la formulation variationnelle mixte . . . . .	102
4.2	Pollution de la solution : modes parasites . . . . .	105
4.2.1	Etude de convergence . . . . .	105
4.2.2	Etude de la distribution des valeurs propres . . . . .	109
4.3	Conclusion . . . . .	114
<b>5</b>	<b>Première famille sur les quadrillatères/hexaèdres</b>	<b>115</b>
5.1	Formulation variationnelle et espace d'approximation . . . . .	116
5.2	Expression des matrices élémentaires . . . . .	116
5.2.1	Cas 2-D . . . . .	117
5.2.2	Cas 3-D . . . . .	118
5.3	Précision de la méthode . . . . .	123
5.4	Algorithme rapide du produit matrice-vecteur . . . . .	128
5.4.1	Factorisation discrète . . . . .	128
5.4.2	Formulation mixte . . . . .	129
5.4.3	Produit $\hat{R}E$ et $\hat{C}E$ en 2-D . . . . .	130
5.4.4	Produit $\hat{R}E$ et $\hat{C}E$ en 3-D . . . . .	131
5.4.5	Complexité du produit matrice-vecteur . . . . .	134
5.5	Calcul de modes propres . . . . .	136
5.5.1	Maillage régulier . . . . .	137
5.5.2	Maillage non-régulier . . . . .	137
5.6	Préconditionnement du système linéaire . . . . .	137
5.6.1	Préconditionnement par un sous-maillage $Q_1$ . . . . .	139
5.6.2	Préconditionnement à l'aide d'une factorisation incomplète . . . . .	141
5.6.3	Préconditionnement utilisant la décomposition de Helmholtz . . . . .	142
5.6.4	Multigrille . . . . .	144
5.7	Conclusion . . . . .	145

<b>6 Méthode de Galerkin discontinue sur les quadrilatères/hexaèdres</b>	<b>147</b>
6.1 Description de la formulation Galerkin discontinue . . . . .	148
6.1.1 Formulation variationnelle . . . . .	148
6.1.2 Expression des matrices . . . . .	149
6.1.3 Termes de pénalisation . . . . .	153
6.1.4 Calculs de complexité . . . . .	153
6.1.5 Conditions aux limites . . . . .	154
6.1.6 Résolution du système linéaire . . . . .	155
6.2 Présence de modes parasites ? . . . . .	155
6.2.1 Etude de convergence . . . . .	155
6.2.2 Etude de valeurs propres . . . . .	161
6.3 Conclusion . . . . .	169
<b>7 Comparaison hexaèdres / tétraèdres pour les équations de Maxwell 3-D</b>	<b>171</b>
7.1 Analyse de dispersion . . . . .	172
7.1.1 Cas 2-D . . . . .	172
7.1.2 Cas 3-D . . . . .	175
7.2 Cas académique de la sphère . . . . .	176
7.2.1 Coût du produit matrice-vecteur . . . . .	176
7.2.2 Sphère parfaitement conductrice . . . . .	177
7.2.3 Sphère diélectrique . . . . .	178
7.3 Cavité cobra . . . . .	179
7.4 Conclusion . . . . .	182
<b>III Equations de Maxwell en domaine axisymétrique</b>	<b>183</b>
<b>8 Résolution des équations de Maxwell axi-symétriques par éléments finis d'arête</b>	<b>185</b>
8.1 Description de la méthode éléments finis . . . . .	186
8.1.1 Choix de la formulation variationnelle . . . . .	186
8.1.2 Calcul de la matrice éléments finis . . . . .	192
8.2 Précision de la méthode . . . . .	198
8.2.1 Cas de la sphère parfaitement conductrice . . . . .	198
8.2.2 Cas du cone-sphère . . . . .	198
8.2.3 Cas du cylindre . . . . .	201
8.3 Conclusion . . . . .	202
<b>9 Équations intégrales d'ordre élevé pour les équations de Maxwell sur des domaines à symétrie de révolution</b>	<b>203</b>
9.1 Obtention de la formulation variationnelle . . . . .	204
9.1.1 Notations . . . . .	204
9.1.2 Formulation EFIE . . . . .	205
9.1.3 Formulation MFIE . . . . .	209
9.1.4 Formulation CFIE . . . . .	210
9.2 Méthode d'intégration . . . . .	211
9.2.1 Méthode d'intégration pour la partie régulière . . . . .	211
9.2.2 Règles simples d'intégration dans le cas régulier . . . . .	213
9.2.3 Calcul des intégrales singulières . . . . .	215
9.3 Précision de la méthode . . . . .	221
9.3.1 Cas de la sphère parfaitement conductrice . . . . .	221

9.3.2	Cas du cylindre parfaitement conducteur . . . . .	221
9.3.3	Cas du cône-sphère parfaitement conducteur . . . . .	222
9.4	Couplage avec les éléments finis . . . . .	226
9.4.1	Définition des opérateurs . . . . .	226
9.4.2	Cas de la sphère revêtue par un matériau diélectrique . . . . .	227
9.4.3	Cas du cylindre revêtu par du diélectrique . . . . .	228
9.4.4	Cas du cone-sphère revêtue par du diélectrique . . . . .	229
9.5	Conclusion . . . . .	230
	<b>Conclusion</b>	<b>233</b>
<b>A</b>	<b>Solutions analytiques des problèmes de diffraction d'ondes planes par une sphère</b>	<b>235</b>
A.1	Notations des fonctions spéciales . . . . .	236
A.2	Solutions analytiques pour l'équation de Helmholtz . . . . .	237
A.2.1	Diffraction par un disque . . . . .	237
A.2.2	Diffraction par une sphère . . . . .	239
A.3	Solutions analytiques pour les équations de Maxwell . . . . .	240
A.3.1	Développement d'un champ en harmoniques sphériques . . . . .	240
A.3.2	Potentiels de Debye pour les équations de Maxwell . . . . .	241
A.3.3	Traces des champs tangentiels associés aux potentiels de Debye . . . . .	243
A.3.4	Décomposition d'une onde plane dans les potentiels de Debye . . . . .	244
A.3.5	Diffraction par une sphère parfaitement conductrice . . . . .	245
A.3.6	Expression de E et H dans tout l'espace . . . . .	247
A.3.7	Diffraction d'une onde plane par une sphère diélectrique . . . . .	249
A.3.8	Prise en compte de la condition de Silver-Müller . . . . .	251
<b>B</b>	<b>Condition transparente pour les équations de Maxwell en régime harmonique utilisant une formule de représentation intégrale</b>	<b>253</b>
B.1	Problème Modèle . . . . .	254
B.2	Description de la condition transparente . . . . .	255
B.3	Formulation variationnelle . . . . .	257
B.4	Calcul du produit matrice-vecteur . . . . .	259
B.4.1	Cas 2-D . . . . .	259
B.4.2	Cas 3-D . . . . .	260
B.5	Equations de Maxwell 3-D . . . . .	261
<b>C</b>	<b>Factorisation discrète et intégration exacte</b>	<b>265</b>
C.1	Cas des éléments finis $H^1$ . . . . .	266
C.2	Cas des éléments finis de Nédélec de la première famille . . . . .	268
C.3	Cas de la formulation Galerkin discontinue . . . . .	269

# Introduction

## Problématique

Ce mémoire a pour objet la résolution des équations de Maxwell en régime fréquentiel, à l'aide de méthodes éléments finis. Ces travaux ont été menés, en majeure partie au sein du projet POEMS (Propagation des Ondes : Etude Mathématique et Simulation) de l'INRIA (Institut National de Recherche en Informatique et Automatique), spécialisé dans l'étude mathématique des problèmes de propagation d'ondes. Ces travaux ont également été menés en collaboration avec l'Office National d'Etudes et de Recherche en Aéronautique (ONERA), et plus particulièrement avec le Département Electromagnétisme et Radar (DEMR). L'application recherchée est la détermination des signatures radar de diverses cibles (avions, missiles ...). Notre motivation pour cette étude vient du fait que la simulation numérique est un outil puissant et flexible. De plus l'utilisation de méthodes d'ordre élevé est intéressante à deux titres. En premier lieu, les méthodes d'ordre élevé permettent de traiter des cas haute-fréquence avec une très grande précision. En second lieu, elles permettent de calculer avec fidélité la signature radar de cibles dites "furtives", dont le signal de retour est très faible.

Nous allons maintenant faire un bref rappel des méthodes les plus pratiquées afin de mieux cerner les enjeux de cette thèse. La littérature étant très vaste sur ce sujet, nous ne serons pas exhaustifs.

## Méthodes "classiques" utilisées

Les méthodes les plus populaires pour résoudre ce type de problème, sont sans conteste les équations intégrales. Dans le cas d'un objet parfaitement conducteur placé dans le vide, on peut ramener la résolution 3-D des équations de Maxwell à une résolution 2-D sur la surface de l'objet. Une présentation générale de ces méthodes est disponible dans le livre [Colton et Kress, 1983]. Elles permettent de traiter également des objets uniformément diélectriques, mais le cas parfaitement conducteur est le plus pratiqué. Bien qu'on ait gagné une dimension en espace, on a perdu le caractère local des équations. Les systèmes linéaires sont pleins, ce qui limite la taille des cas de calculs. Pour lever cette contrainte, des méthodes multipoles rapides ont été mises au point, nous renvoyons le lecteur à [Coifman *et al.*, 1993], [Song. *et al.*, 1997]. Ces méthodes fournissent un produit matrice-vecteur rapide pour un coût de stockage faible. A l'aide de ces techniques, les cas traités peuvent être de très grande taille (plusieurs millions de ddl).

Sur le cas d'un objet parfaitement conducteur ou homogène (diélectrique), ces techniques sont extrêmement performantes. Toutefois, les cas réalistes sont parfois plus complexes, les objets ont souvent des hétérogénéités locales, difficiles à traiter en équations intégrales pures. Une solution attrayante est de résoudre les parties hétérogènes de l'objet par une méthode éléments finis, et coupler les éléments finis avec les équations intégrales, pour résoudre la totalité du problème. Une analyse de cette stratégie est menée dans [Simon, 2003], en utilisant des éléments finis d'arête de plus bas ordre et des équations intégrale de plus bas ordre.

Autant les équations intégrales de plus bas ordre se révèlent précises car l'erreur de dispersion est extrêmement faible, autant les éléments finis d'arête 3-D souffrent d'un manque de précision, car ils sont beaucoup plus dispersifs. Un des buts de cette thèse est d'exhiber les défauts des éléments finis de bas ordre, et de montrer que les éléments finis d'ordre élevé sont une réponse convaincante afin d'obtenir une méthode précise.

## Eléments finis d'arête

On a vu croître **un très vif intérêt porté aux méthodes d'ordre élevé ces dix dernières années**. Cela se traduit par une abondance de travaux publiés récemment sur la question. On peut dire, pour simplifier, que si les éléments finis d'arêtes de plus bas degré étaient d'actualité dans les années 1980-1990, cela ne semble plus le cas actuellement. Avant de s'aventurer dans la jungle des éléments finis d'arête, il nous a semblé important de signaler deux ouvrages de référence. Tout d'abord le livre de Peter Monk qui est paru en 2002, [Monk, 2002]. P. Monk est le grand spécialiste mondial des méthodes numériques pour le calcul des solutions de Maxwell en régime harmonique. Son livre, quoique technique est abordable et permet, rien qu'à la lecture des introductions de chaque chapitre de se faire une idée des enjeux sur la question. Le deuxième ouvrage que nous voudrions citer est plus tourné vers la mise en œuvre des éléments finis d'arête, et complète ainsi le livre de Monk. Il s'agit de la monographie de Jin, [Jin, 1993] qu'il a publié en 1993.

Le premier article de référence est celui de Jean-Claude Nédélec qui introduit la première famille de Nédélec en 1980 dans [Nédélec, 1980], puis la seconde famille en 1986 dans [Nédélec, 1986]. Notons que Nédélec ne donne pas de fonctions de base dans ses articles et que leur détermination à tout ordre a fait l'objet de travaux spécifiques. Pour  $p = 0$ , les fonctions de base de la première famille avaient été déjà proposées par Whitney mais dans un contexte tout à fait différent de celui de l'approximation par éléments finis [Bossavit, 1998]. Il est à noter que, chaque fois que l'on choisit un type d'élément différent (triangle, quadrangle en 2-D, tétraèdre, hexaèdre, pyramide, prisme en 3-D), le travail de recherche d'une "bonne base" est à mener : il n'y a pas unicité de la base et le choix de la meilleure est l'objet de controverses animées.

Ainsi, pour générer les espaces de Nédélec, on peut choisir des bases "génératrices" différentes. Ce choix dépendra des propriétés qu'on veut obtenir sur la base. Dans les diverses bases présentées, les fonctions de bases sont identiques à l'ordre zéro, mais pas aux ordres supérieurs.

Deux approches s'opposent dans la littérature, : les hexaèdres ou les tétraèdres. Les tétraèdres sont les plus populaires, notamment ces dernières années, car les meilleurs non-structurés tétraédriques sont extrêmement courants. Parmi la multitude de papiers sur la manière de monter en ordre sur les tétraèdres, nous citerons [Graglia *et al.*, 1997], [Webb, 1999], [Ainsworth et Coyle, 2003] et [Demkowicz, 2000]. La montée en ordre sur les hexaèdres est plus simple, elle est détaillée dans [Cohen, 2002].

## Autres méthodes éléments finis

Des auteurs proposent une alternative aux sacro-saints éléments finis d'arête. Nous avons noté deux approches concurrentielles. La première approche est la réhabilitation des éléments finis nodaux pour les équations de Maxwell. Longtemps, ces éléments finis ont été boudés par les numériciens car ils généraient des ondes parasites qui polluaient la solution. De nombreux auteurs ont proposé une technique de régularisation , nous renvoyons le lecteur au papier [Costabel et Dauge, 2002].

La seconde approche est l'utilisation de méthodes de Galerkin discontinues. Ces méthodes ont connu un regain d'intérêt exceptionnel ces dernières années. Des travaux complets sur les équations de Maxwell, utilisant les tétraèdres, ont été menés à l'université de Brown, [Hesthaven et Warburton, 2002], [Hesthaven et Warburton, 2004], [Olson et Hesthaven, 2004]. Pour les hexaèdres, nous sommes partis de la thèse de [Pernet, 2004].

## Solution proposée et plan de la thèse

L'utilisation de méthodes d'ordre élevé se heurte souvent au problème d'un temps de calcul prohibitif et d'un stockage important par rapport aux méthodes de plus bas ordre. Cet inconvénient a été levé par G. Cohen et P. Monk pour les éléments finis d'arête de la seconde famille de Nédélec sur les hexaèdres [Cohen et Monk, 1999]. Toutefois, cette approche s'est révélée inadaptée, car on obtient de nombreuses ondes parasites. Un remède étudié dans [Pernet, 2004] est d'adopter une formulation Galerkin discontinue, mais cette solution n'est pas très heureuse en régime fréquentiel. Nous proposons d'utiliser les hexaèdres de la première famille de Nédélec, pour lesquels nous avons mis au point un algorithme de produit matrice-vecteur rapide. Cet algorithme s'appuie sur la tensorisation des fonctions de bases et permet de gagner à la fois en stockage et en temps de calcul. On met en œuvre des préconditionneurs classiques pour montrer l'intérêt de monter en ordre sur ce type de discréétisation.

Nous proposons également une méthode de discréétisation, basée sur la première famille, pour résoudre les équations de Maxwell sur des domaines axi-symétriques. Pour traiter des domaines non-bornés, on utilise un couplage éléments finis-équations intégrales axi-symétriques. Des résultats numériques montrent que l'approche choisie donne entièrement satisfaction, et que l'ordre élevé apporte un réel gain en nombre de degrés de liberté et en temps de calcul.

Cette thèse est divisée en trois parties et neuf chapitres.

Dans la première partie, nous étudions la résolution de l'équation de Helmholtz par éléments finis spectraux. Les méthodes d'éléments finis mixtes quadrilatéraux ou hexaédriques avec condensation de masse (ou éléments finis mixtes spectraux) ont été utilisées avec succès pour le régime transitoire [Fauqueux, 2003]. Pour de telles équations, elles conduisent à des algorithmes peu coûteux en temps de calcul et en stockage. Nous montrons qu'en régime harmonique, on conserve ces bonnes propriétés à condition d'utiliser des algorithmes itératifs pour l'inversion des systèmes linéaires. Nous montrons qu'il est possible de construire des préconditionneurs efficaces lorsqu'on monte en ordre. Nous proposons également une comparaison tétraèdres/hexaèdres sur des cas complexes.

Dans la seconde partie, nous nous intéressons aux équations de Maxwell 2-D et 3-D. Nous montrons les inconvénients lorsque l'on utilise la seconde famille de Nédélec sur les hexaèdres ou une méthode Galerkin discontinue. Nous détaillons les hexaèdres de la première famille, notamment la manière dont on obtient l'algorithme rapide pour effectuer le produit matrice-vecteur. Nous utilisons des préconditionneurs classiques sur ce type de discréétisation. Nous montrons, que déjà sur des cas de moyenne taille, la montée en ordre présente un intérêt non-négligeable.

La troisième partie est dédiée au cas axisymétrique. Dans ce cas, les équations de Maxwell se réduisent à une résolution d'une suite de problèmes 2-D indépendants, chaque problème étant relié à un mode de la décomposition de Fourier. Nous proposons une formulation mixte utilisant à la fois les éléments finis nodaux et les éléments finis d'arête pour résoudre chaque problème

2-D. Afin de prendre en compte la condition de Sommerfeld de manière exacte, nous couplons les éléments finis avec des équations intégrales d'ordre élevé. L'intégration des singularités est discutée. Nous montrons sur des cas complexes le gain apporté par la montée en ordre.

# Première partie

## Equation de Helmholtz



# Chapitre 1

## Résolution de l'équation de Helmholtz par éléments finis spectraux

*Ce chapitre introduit les éléments finis spectraux utilisés pour la résolution de l'équation de Helmholtz ainsi que leurs bonnes propriétés. La première section s'attache à décrire la discréttisation choisie et le calcul rapide de la matrice éléments finis qui en découle. La deuxième section décrit les éléments courbes qu'on utilise en 2-D et en 3-D. La troisième section aborde la précision de la méthode sur des cas d'objets lisses et avec des coins. La dernière section montre tout l'intérêt d'utiliser des éléments finis d'ordre élevé sur un cas 2-D réaliste.*

### Sommaire

---

<b>1.1</b>	<b>Formulation variationnelle classique et éléments finis spectraux</b>	<b>17</b>
1.1.1	Cas modèle traité . . . . .	17
1.1.2	Formulation variationnelle . . . . .	18
1.1.3	Expression des matrices en 2-D et 3-D, utilisation de la condensation de masse . . . . .	21
<b>1.2</b>	<b>Éléments finis courbes</b>	<b>24</b>
1.2.1	Cas 2-D . . . . .	24
1.2.2	Cas 3-D . . . . .	26
<b>1.3</b>	<b>Précision de la méthode</b>	<b>29</b>
1.3.1	Cas 2-D . . . . .	29
1.3.2	Cas 3-D . . . . .	35
<b>1.4</b>	<b>Application aux filtres optiques, intérêt de l'ordre élevé</b>	<b>38</b>
1.4.1	Faisceaux gaussiens . . . . .	38
1.4.2	Propriétés du filtre optique . . . . .	39
1.4.3	Apologie de l'ordre élevé . . . . .	46
<b>1.5</b>	<b>Conclusion</b>	<b>48</b>

---

Afin d'approfondir ce chapitre, on pourra consulter

- COHEN, G. (2002). *Higher-order numerical methods for transient wave equations.* Springer Verlag,
- CIARLET, P. (1978). *The finite element method for elliptic problems.* North-Holland,
- ZIENKIEWICZ, O. et TAYLOR, R. (1989). *The finite element method.* McGraw-Hill,
- SOLIN, P., SEGETH, K. et DOLEZEL, I. (2003). *Higher-order finite elements methods.* Studies in Advanced Mathematics, Chapman and Hall

## 1.1 Formulation variationnelle classique et éléments finis spéciaux

### 1.1.1 Cas modèle traité

On cherche à résoudre l'équation de Helmholtz dans le cas hétérogène

$$-\rho \omega^2 u - \operatorname{div}(\mu \nabla u) = f \quad (1.1)$$

où  $\rho$  et  $\mu$  sont deux coefficients qui peuvent être constants. C'est ce qu'on appellera le cas homogène. Le cas hétérogène a lieu quand les coefficients  $\rho$  et  $\mu$  dépendent de la position. On a :

$$\frac{\rho}{\mu} = \frac{1}{c^2}$$

où la constante  $c$  est la vitesse de l'onde. Pour les équations de Maxwell,  $\rho$  s'identifie avec  $\varepsilon$  et  $\mu$  avec  $1/\mu$  dans le cas transverse électrique. Dans le cas homogène, l'équation de Helmholtz s'écrit classiquement :

$$-k^2 u - \Delta u = f \quad (1.2)$$

avec la relation de dispersion :

$$k = \frac{\omega}{c}$$

On résoudra l'équation de Helmholtz dans un domaine borné  $\Omega$ , on choisira des conditions aux limites de type Dirichlet

$$u = u_d$$

de type Neumann

$$\frac{\partial u}{\partial n} = u_n$$

ou une condition absorbante d'ordre 1

$$\frac{\partial u}{\partial n} - iku = 0$$

On considère un maillage conforme en quadrillatères/hexaèdres du domaine  $\Omega$  :

$$\Omega = \bigcup_{i=1}^{N_e} K_i$$

$$\forall i \neq j \quad K_i \cap K_j = \emptyset$$

Dans la figure 1.1, on montre un exemple de maillage quadrilatéral.

On considère ensuite la transformation  $F_j$ , qui va transformer le carré unité  $\hat{K} = [0, 1]^2$  en un quadrilatère quelconque  $K_j$ . Cette transformation n'est pas affine sur des maillages quelconques. Soit  $K_j$ , le quadrilatère de sommets  $A_1, A_2, A_3, A_4$ .  $F_j$  s'écrit :

$$F_j = \sum_{\ell=1}^4 A_\ell \hat{\varphi}_\ell$$

avec

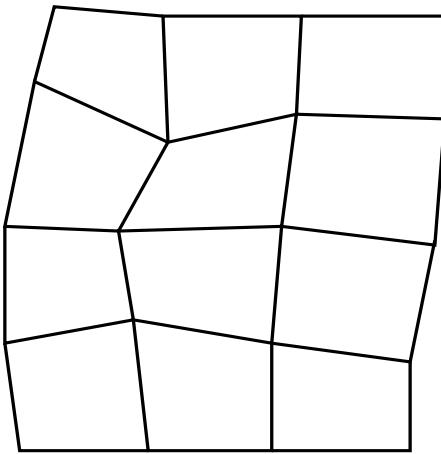


FIG. 1.1 – Maillage quadrilatéral :  $\mathcal{M}_h = \bigcup_{j=1}^{N_e} K_j$ ,

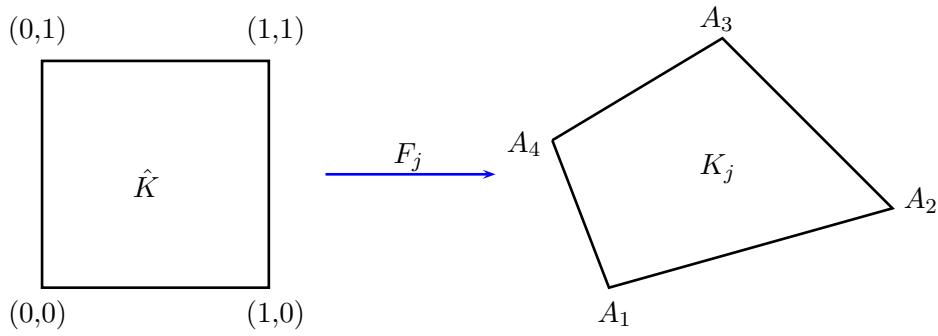


FIG. 1.2 – La transformation  $\vec{F}_j$

$$\hat{\varphi}_1 = (1 - \hat{x}_1)(1 - \hat{x}_2)$$

$$\hat{\varphi}_2 = \hat{x}_1(1 - \hat{x}_2)$$

$$\hat{\varphi}_3 = \hat{x}_1\hat{x}_2$$

$$\hat{\varphi}_4 = (1 - \hat{x}_1)\hat{x}_2$$

On notera  $DF_j$  la matrice jacobienne de la transformation  $F_j$ , et  $J_j$  le déterminant de  $DF_i$ .

### 1.1.2 Formulation variationnelle

On écrit la formulation variationnelle “classique” de l’équation de Helmholtz :

$$-\omega^2 \int_{\Omega} \rho u v + \int_{\Omega} \mu \nabla u \cdot \nabla v = \int_{\Omega} f v \quad \forall v \in U_h$$

Les fonctions  $u$  et  $v$  sont prises dans l'espace d'approximation :

$$U_h = \{v \in H^1(\Omega) \mid v|_{K_i} \circ F_i \in Q_r\}$$

On choisit ensuite la position des degrés de liberté sur le carré unité, en prenant les points de Gauss-Lobatto en 2-D. Pour une méthode d'ordre  $r$ , on aura  $(r+1)^2$  degrés de liberté sur le

carré unité. Sur la figure 1.3, on voit l’agencement de ces points pour  $r=5$ . C’est le choix de ces points d’interpolation qui donnent le qualificatif de spectral à la méthode éléments finis. En effet, les points de Gauss-Lobatto conduisent à une convergence exponentielle, quand on monte en ordre, de l’erreur  $L^2$  entre une fonction et son interpolée. Les points réguliers n’ont pas cette propriété. Ils ont même la fâcheuse tendance à donner une interpolée qui oscille de plus en plus lorsqu’on monte en ordre, ce qui est connu comme le phénomène de Runge. Cette propriété est montrée sur la figure 1.4. On note les points de Gauss-Lobatto :

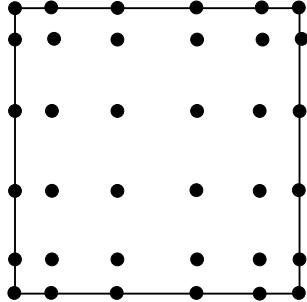


FIG. 1.3 – Les 36 points de Gauss-Lobatto pour des éléments spectraux  $Q_5$  en 2D

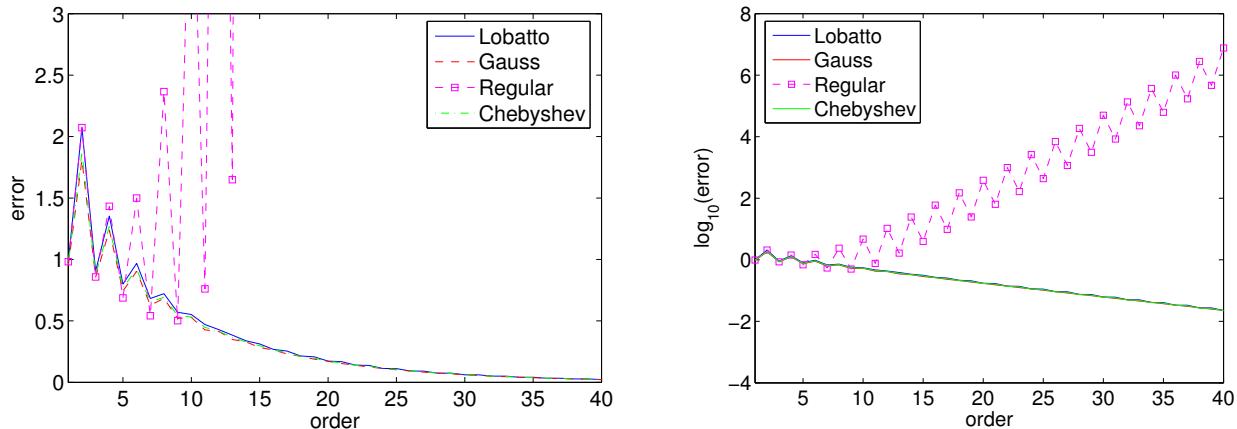


FIG. 1.4 – Erreur  $L^2$  entre la fonction  $1/(1+x^2)$  et son interpolée sur l’intervalle  $[-5, 5]$ . À gauche, erreur en échelle linéaire, à droite échelle logarithmique

$$\hat{\xi}_{k,l} = (\hat{\xi}_k, \hat{\xi}_l) \quad k = 1..r+1 \quad l = 1..r+1$$

Chaque degré de liberté sur  $\hat{K}$  est associé à une fonction de base, qui est un simple polynôme d’interpolation de Lagrange. La fonction de base associée au degré de liberté  $(k, l)$  est un polynôme d’ordre  $r$  en  $\hat{x}_1$  et  $\hat{x}_2$ , qui vaut 1 au point  $\hat{\xi}_{k,l}$  et 0 sur tous les autres points d’interpolation. On note cette fonction de base  $\hat{\varphi}_{k,l}$ , elle vaut :

$$\hat{\varphi}_{k,l}(\hat{x}_1, \hat{x}_2) = \prod_{m=1, m \neq k}^{r+1} \frac{\hat{x}_1 - \hat{\xi}_k}{\hat{\xi}_m - \hat{\xi}_k} \prod_{m=1, m \neq l}^{r+1} \frac{\hat{x}_2 - \hat{\xi}_l}{\hat{\xi}_m - \hat{\xi}_l}$$

Elle vérifie bien

$$\hat{\varphi}_{k,l}(\hat{\xi}_p, \hat{\xi}_q) = \delta_{k,p} \delta_{l,q}$$

où  $\delta$  est le symbole de Kronecker. Pour les fonctions de base définies sur  $\hat{K}$ , on adoptera indifféremment la notation avec un couple d'indices  $(i_1, i_2)$  ou avec un seul indice  $i$ . Ainsi, on peut aussi écrire :

$$\hat{\varphi}_i(\hat{\xi}_j) = \delta_{i,j}$$

Les fonctions de base sur le quadrilatère  $K_e$  vérifient par définition :

$$\hat{\varphi}_i = \varphi_i \circ F_e$$

d'où

$$\begin{aligned}\varphi_i &= \hat{\varphi}_i \circ F_e^{-1} \\ \nabla \varphi_i &= D F_e^{-t} \hat{\nabla} \hat{\varphi}_i \circ F_e^{-1}\end{aligned}$$

$\hat{\nabla}$  étant le gradient par rapport à  $\hat{x}_1, \hat{x}_2$ . Les fonctions de base  $\hat{\varphi}_i$  sur le carré unité forment une base de l'espace  $Q_r$ .  $Q_r$  est l'espace des polynômes de degré inférieur ou égal à  $r$  en chaque variable sur le carré unité  $\hat{K}$ .

$$Q_r = \left\{ \sum_{i,j=0}^r a_{i,j} \hat{x}_1^i \hat{x}_2^j \text{ avec } a_{i,j} \in \mathbb{R} \ \forall (i,j) \right\}$$

Sur chaque maille, la solution numérique est combinaison linéaire des fonctions de base  $\varphi_i$ . Si on note  $(u_i)_{1 \leq i \leq (r+1)^2}$ , les composantes de  $u$  dans cette base, on a

$$u|_{K_e}(x, y) = \sum_{i=1}^{(r+1)^2} u_i \varphi_i(x)$$

En injectant cette expression, dans la formulation variationnelle et en prenant  $v = \varphi_j$ , on obtient le système discret suivant :

$$-\omega^2 M_h U_h + K_h U_h = F_h$$

$U_h$  est le vecteur des composantes de  $u$ . Le second membre s'écrit :

$$(F_h)_i = \int_{\Omega} f \varphi_i = \sum_{e \in \text{supp}(\varphi_i)} \int_{\hat{K}} J_e f(x) \hat{\varphi}_i(\hat{x}_1, \hat{x}_2) d\hat{x}_1 d\hat{x}_2$$

On pose dans toute la suite de l'exposé comme convention :

$$J_e > 0$$

Pour les quadrillatères/hexaèdres pour lesquels le jacobien est négatif, on permute au préalable la numérotation des sommets afin de se ramener à un jacobien positif. Si le jacobien change de signe, c'est que le quadrilatère est dégénéré. La matrice de masse  $M_h$  est de terme générique :

$$(M_h)_{i,j} = \int_{\Omega} \rho \varphi_i \varphi_j = \sum_{e \in \text{supp}(\varphi_i) \cap \text{supp}(\varphi_j)} \int_{\hat{K}} \rho J_e \hat{\varphi}_i(\hat{x}_1, \hat{x}_2) \hat{\varphi}_j(\hat{x}_1, \hat{x}_2) d\hat{x}_1 d\hat{x}_2$$

La matrice de rigidité  $K_h$  a pour expression :

$$(K_h)_{i,j} = \int_{\Omega} \mu \nabla \varphi_i \nabla \varphi_j = \sum_{e \in \text{supp}(\varphi_i) \cap \text{supp}(\varphi_j)} \int_{\hat{K}} \mu J_e D F_e^{-1} D F_e^{-t} \hat{\nabla} \hat{\varphi}_i(\hat{x}_1, \hat{x}_2) \cdot \hat{\nabla} \hat{\varphi}_j(\hat{x}_1, \hat{x}_2) d\hat{x}_1 d\hat{x}_2$$

### 1.1.3 Expression des matrices en 2-D et 3-D, utilisation de la condensation de masse

Pour évaluer les intégrales sur  $\hat{K}$ , on utilise les formules d'intégration de Gauss-Lobatto, exactes pour les polynômes de degré inférieur ou égal à  $2r - 1$ . Or, le produit  $\hat{\varphi}_i \hat{\varphi}_j$  est un polynôme de degré  $2r$ . On a également la présence de  $J_e$ , qui est un polynôme de degré supérieur à 1, sur des maillages quelconques. Plus précisément,  $J_e \in P_1$  en 2-D (quadrilatère) et  $J_e \in P_4 \cap Q_2$  en 3-D (hexaèdre). Par conséquent, l'intégration utilisée est approchée, mais cette approximation n'affecte pas la précision de la méthode, ce qui est vérifié soit en faisant une estimation d'erreur [Ciarlet, 1978], ou une analyse de dispersion [Fauqueux, 2003]. Le premier auteur fait la démonstration dans le cas de tétraèdres, alors que le second auteur fait l'analyse de dispersion sur des maillages hexaédriques réguliers. Nous avons observé une perte de précision due à la condensation de masse, sur des maillages hexaédriques fortement déformés. Ce phénomène sera étudié plus en détail dans le chapitre 3. L'utilisation des formules de quadrature de Gauss-Lobatto conduit à la condensation de masse, i.e la matrice de masse est une matrice diagonale. En effet, on a :

$$(M_h)_{i,j} \approx \sum_{e \in \text{supp}(\varphi_i) \cap \text{supp}(\varphi_j)} \sum_{k=1}^{(r+1)^d} \rho \omega_k J_e \hat{\varphi}_i(\hat{\xi}_k) \hat{\varphi}_j(\hat{\xi}_k)$$

$\hat{\xi}_k$  est le point d'intégration (2-D ou 3-D),  $\omega_k$  le poids d'intégration. On a, par construction des fonctions de base

$$\hat{\varphi}_i(\hat{\xi}_k) = \delta_{i,k}$$

D'où

$$(M_h)_{i,j} \approx \sum_{e \in \text{supp}(\varphi_i) \cap \text{supp}(\varphi_j)} \omega_i \rho(\hat{\xi}_i) J_e(\hat{\xi}_i) \delta_{i,j}$$

ce qui démontre que  $M_h$  est diagonale.

Pour ce qui est de la matrice de rigidité, les expressions sont plus simples lorsqu'on utilise les points de Gauss-Lobatto pour intégrer de manière approchée les intégrales. Nous allons détailler ces expressions et montrer qu'elles mènent à un calcul plus rapide de la matrice de rigidité.

#### Cas 2-D

On s'intéresse dans cette partie au calcul de la matrice de rigidité élémentaire :

$$(K_h)_{i,j} = \int_{\hat{K}} \mu J_e D F_e^{-1} D F_e^{-t} \hat{\nabla} \hat{\varphi}_i(\hat{x}_1, \hat{x}_2) \hat{\nabla} \hat{\varphi}_j(\hat{x}_1, \hat{x}_2) d\hat{x}_1 d\hat{x}_2$$

avec  $i, j$  les numéros locaux des degrés de liberté. On note la matrice :

$$B_e(\hat{x}_1, \hat{x}_2) = \mu J_e D F_e^{-1} D F_e^{-t}$$

Cette matrice est symétrique, elle contient l'information sur la géométrie.

$$\begin{aligned} (K_h)_{i,j} &= \sum_{m,n=1}^{r+1} \omega_{m,n} B_{11}(\hat{\xi}_m, \hat{\xi}_n) \hat{\varphi}'_{i_1}(\hat{\xi}_m) \hat{\varphi}_{i_2}(\hat{\xi}_n) \hat{\varphi}'_{j_1}(\hat{\xi}_m) \hat{\varphi}_{j_2}(\hat{\xi}_n) \\ &\quad + \omega_{m,n} B_{21}(\hat{\xi}_m, \hat{\xi}_n) \hat{\varphi}_{i_1}(\hat{\xi}_m) \hat{\varphi}'_{i_2}(\hat{\xi}_n) \hat{\varphi}'_{j_1}(\hat{\xi}_m) \hat{\varphi}_{j_2}(\hat{\xi}_n) \\ &\quad + \omega_{m,n} B_{21}(\hat{\xi}_m, \hat{\xi}_n) \hat{\varphi}'_{i_1}(\hat{\xi}_m) \hat{\varphi}_{i_2}(\hat{\xi}_n) \hat{\varphi}_{j_1}(\hat{\xi}_m) \hat{\varphi}'_{j_2}(\hat{\xi}_n) \\ &\quad + \omega_{m,n} B_{22}(\hat{\xi}_m, \hat{\xi}_n) \hat{\varphi}_{i_1}(\hat{\xi}_m) \hat{\varphi}'_{i_2}(\hat{\xi}_n) \hat{\varphi}_{j_1}(\hat{\xi}_m) \hat{\varphi}'_{j_2}(\hat{\xi}_n) \end{aligned}$$

Après simplification, on trouve

$$\begin{aligned}
(K_h)_{i,j} &= \sum_{m=1}^{r+1} \omega_{m,i_2} B_{11}(\hat{\xi}_m, \hat{\xi}_{i_2}) \hat{\varphi}'_{i_1}(\hat{\xi}_m) \hat{\varphi}'_{j_1}(\hat{\xi}_m) \delta_{i_2, j_2} \\
&\quad + \omega_{i_1, j_2} B_{21}(\hat{\xi}_{i_1}, \hat{\xi}_{j_2}) \hat{\varphi}'_{i_2}(\hat{\xi}_{j_2}) \hat{\varphi}'_{j_1}(\hat{\xi}_{i_1}) \\
&\quad + \omega_{i_2, j_1} B_{21}(\hat{\xi}_{i_2}, \hat{\xi}_{j_1}) \hat{\varphi}'_{i_1}(\hat{\xi}_{j_1}) \hat{\varphi}'_{j_2}(\hat{\xi}_{i_2}) \\
&\quad + \sum_{n=1}^{r+1} \omega_{i_1, n} B_{22}(\hat{\xi}_{i_1}, \hat{\xi}_n) \hat{\varphi}'_{i_2}(\hat{\xi}_n) \hat{\varphi}'_{j_2}(\hat{\xi}_n) \delta_{i_1, j_1}
\end{aligned}$$

Si on avait choisi d'intégrer plus précisément, par exemple en utilisant les points de Gauss au lieu des points de Gauss-Lobatto, on n'aurait pas eu de simplifications. Le calcul de la matrice de rigidité aurait été de complexité  $O(r^6)$  (boucle sur  $i_1, j_1, i_2, j_2, m, n$ ). Lorsqu'on utilise les points de Gauss-Lobatto, le calcul de la matrice de rigidité est de complexité  $O(r^4)$ . De fait, on a une boucle sur  $i_1, j_1, i_2, j_2$  pour les termes croisés contenant  $B_{21}$ . Pour les termes diagonaux impliquant  $B_{11}, B_{22}$ , on a une boucle sur respectivement  $i_1, j_1, i_2, m$  et  $i_1, i_2, j_2, n$ . Le calcul de la matrice est donc plus rapide et il est comparable au coût de calcul de la matrice pour des éléments triangulaires  $P_r$  en  $O(r^4)$ . Cette complexité n'est exacte que pour des éléments triangulaires droits, pour lesquels on peut utiliser des matrices élémentaires précalculées. En présence de triangles courbes, la matrice jacobienne n'est pas constante par élément, on ne peut pas utiliser ces matrices élémentaires. Le coût devient alors plus important en  $O(r^6)$ , car nous avons une sommation sur tous les points d'intégration :

$$(K_h)_{i,j} = \sum_m \omega_m B(\hat{\xi}_m) \hat{\nabla} \hat{\varphi}_i(\hat{\xi}_m) \cdot \hat{\nabla} \hat{\varphi}_j(\hat{\xi}_m)$$

On devra utiliser  $O(r^2)$  points d'intégration pour évaluer chaque interaction élémentaire. Nous n'avons pas cet inconvénient sur les quadrangles où le coût est toujours en  $O(r^4)$ , à cause des simplifications exhibées.

### Cas 3-D

Dans le cas 3-D, les calculs sont très proches, et les conclusions identiques. En utilisant les points de Gauss-Lobatto pour intégrer, on aboutit à une complexité en  $O(r^5)$  au lieu de  $O(r^9)$  si on utilise des points différents. Le gain est appréciable, ce qui est corroboré par les expériences numériques. On explicite les calculs, mais le lecteur n'est pas obligé de les lire ! On utilise toujours la matrice

$$B_e(\hat{x}_1, \hat{x}_2, \hat{x}_3) = \mu J_e D F_e^{-1} D F_e^{-t}$$

$$\begin{aligned}
(K_h)_{i,j} = & \sum_{l,m,n=1}^{r+1} \left( \omega_{l,m,n} B_{11}(\hat{\xi}_l, \hat{\xi}_m, \hat{\xi}_n) \hat{\varphi}'_{i_1}(\hat{\xi}_l) \hat{\varphi}_{i_2}(\hat{\xi}_m) \hat{\varphi}_{i_3}(\hat{\xi}_n) \hat{\varphi}'_{j_1}(\hat{\xi}_l) \hat{\varphi}_{j_2}(\hat{\xi}_m) \hat{\varphi}_{j_3}(\hat{\xi}_n) \right. \\
& + \omega_{l,m,n} B_{22}(\hat{\xi}_l, \hat{\xi}_m, \hat{\xi}_n) \hat{\varphi}_{i_1}(\hat{\xi}_l) \hat{\varphi}'_{i_2}(\hat{\xi}_m) \hat{\varphi}_{i_3}(\hat{\xi}_n) \hat{\varphi}_{j_1}(\hat{\xi}_l) \hat{\varphi}'_{j_2}(\hat{\xi}_m) \hat{\varphi}_{j_3}(\hat{\xi}_n) \\
& + \omega_{l,m,n} B_{33}(\hat{\xi}_l, \hat{\xi}_m, \hat{\xi}_n) \hat{\varphi}_{i_1}(\hat{\xi}_l) \hat{\varphi}_{i_2}(\hat{\xi}_m) \hat{\varphi}'_{i_3}(\hat{\xi}_n) \hat{\varphi}_{j_1}(\hat{\xi}_l) \hat{\varphi}_{j_2}(\hat{\xi}_m) \hat{\varphi}'_{j_3}(\hat{\xi}_n) \\
& + \omega_{l,m,n} B_{21}(\hat{\xi}_l, \hat{\xi}_m, \hat{\xi}_n) \hat{\varphi}'_{i_1}(\hat{\xi}_l) \hat{\varphi}_{i_2}(\hat{\xi}_m) \hat{\varphi}_{i_3}(\hat{\xi}_n) \hat{\varphi}_{j_1}(\hat{\xi}_l) \hat{\varphi}'_{j_2}(\hat{\xi}_m) \hat{\varphi}_{j_3}(\hat{\xi}_n) \\
& + \omega_{l,m,n} B_{21}(\hat{\xi}_l, \hat{\xi}_m, \hat{\xi}_n) \hat{\varphi}_{i_1}(\hat{\xi}_l) \hat{\varphi}'_{i_2}(\hat{\xi}_m) \hat{\varphi}_{i_3}(\hat{\xi}_n) \hat{\varphi}'_{j_1}(\hat{\xi}_l) \hat{\varphi}_{j_2}(\hat{\xi}_m) \hat{\varphi}_{j_3}(\hat{\xi}_n) \\
& + \omega_{l,m,n} B_{31}(\hat{\xi}_l, \hat{\xi}_m, \hat{\xi}_n) \hat{\varphi}'_{i_1}(\hat{\xi}_l) \hat{\varphi}_{i_2}(\hat{\xi}_m) \hat{\varphi}_{i_3}(\hat{\xi}_n) \hat{\varphi}_{j_1}(\hat{\xi}_l) \hat{\varphi}_{j_2}(\hat{\xi}_m) \hat{\varphi}'_{j_3}(\hat{\xi}_n) \\
& + \omega_{l,m,n} B_{31}(\hat{\xi}_l, \hat{\xi}_m, \hat{\xi}_n) \hat{\varphi}_{i_1}(\hat{\xi}_l) \hat{\varphi}_{i_2}(\hat{\xi}_m) \hat{\varphi}'_{i_3}(\hat{\xi}_n) \hat{\varphi}'_{j_1}(\hat{\xi}_l) \hat{\varphi}_{j_2}(\hat{\xi}_m) \hat{\varphi}_{j_3}(\hat{\xi}_n) \\
& + \omega_{l,m,n} B_{32}(\hat{\xi}_l, \hat{\xi}_m, \hat{\xi}_n) \hat{\varphi}_{i_1}(\hat{\xi}_l) \hat{\varphi}'_{i_2}(\hat{\xi}_m) \hat{\varphi}_{i_3}(\hat{\xi}_n) \hat{\varphi}_{j_1}(\hat{\xi}_l) \hat{\varphi}_{j_2}(\hat{\xi}_m) \hat{\varphi}'_{j_3}(\hat{\xi}_n) \\
& \left. + \omega_{l,m,n} B_{32}(\hat{\xi}_l, \hat{\xi}_m, \hat{\xi}_n) \hat{\varphi}_{i_1}(\hat{\xi}_l) \hat{\varphi}_{i_2}(\hat{\xi}_m) \hat{\varphi}'_{i_3}(\hat{\xi}_n) \hat{\varphi}_{j_1}(\hat{\xi}_l) \hat{\varphi}'_{j_2}(\hat{\xi}_m) \hat{\varphi}_{j_3}(\hat{\xi}_n) \right)
\end{aligned}$$

Après simplification, on trouve

$$\begin{aligned}
(K_h)_{i,j} = & \sum_{l=1}^{r+1} \omega_{l,i_2,i_3} B_{11}(\hat{\xi}_l, \hat{\xi}_{i_2}, \hat{\xi}_{i_3}) \hat{\varphi}'_{i_1}(\hat{\xi}_l) \hat{\varphi}'_{j_1}(\hat{\xi}_l) \delta_{i_2,j_2} \delta_{i_3,j_3} \\
& + \sum_{m=1}^{r+1} \omega_{i_1,m,i_3} B_{22}(\hat{\xi}_{i_1}, \hat{\xi}_m, \hat{\xi}_{i_3}) \hat{\varphi}'_{i_2}(\hat{\xi}_m) \hat{\varphi}'_{j_2}(\hat{\xi}_m) \delta_{i_1,j_1} \delta_{i_3,j_3} \\
& + \sum_{n=1}^{r+1} \omega_{i_1,i_2,n} B_{33}(\hat{\xi}_{i_1}, \hat{\xi}_{i_2}, \hat{\xi}_n) \hat{\varphi}'_{i_3}(\hat{\xi}_n) \hat{\varphi}'_{j_3}(\hat{\xi}_n) \delta_{i_1,j_1} \delta_{i_2,j_2} \\
& + \omega_{j_1,i_2,i_3} B_{21}(\hat{\xi}_{j_1}, \hat{\xi}_{i_2}, \hat{\xi}_{i_3}) \hat{\varphi}'_{i_1}(\hat{\xi}_{j_1}) \hat{\varphi}'_{j_2}(\hat{\xi}_{i_2}) \delta_{i_3,j_3} \\
& + \omega_{i_1,j_2,i_3} B_{21}(\hat{\xi}_{i_1}, \hat{\xi}_{j_2}, \hat{\xi}_{i_3}) \hat{\varphi}'_{j_1}(\hat{\xi}_{i_1}) \hat{\varphi}'_{i_2}(\hat{\xi}_{j_2}) \delta_{i_3,j_3} \\
& + \omega_{j_1,i_2,i_3} B_{31}(\hat{\xi}_{j_1}, \hat{\xi}_{i_2}, \hat{\xi}_{i_3}) \hat{\varphi}'_{i_1}(\hat{\xi}_{j_1}) \hat{\varphi}'_{j_3}(\hat{\xi}_{i_3}) \delta_{i_2,j_2} \\
& + \omega_{i_1,i_2,j_3} B_{31}(\hat{\xi}_{i_1}, \hat{\xi}_{i_2}, \hat{\xi}_{j_3}) \hat{\varphi}'_{j_1}(\hat{\xi}_{i_1}) \hat{\varphi}'_{i_3}(\hat{\xi}_{j_3}) \delta_{i_2,j_2} \\
& + \omega_{i_1,j_2,i_3} B_{32}(\hat{\xi}_{i_1}, \hat{\xi}_{j_2}, \hat{\xi}_{i_3}) \hat{\varphi}'_{i_2}(\hat{\xi}_{j_2}) \hat{\varphi}'_{j_3}(\hat{\xi}_{i_3}) \delta_{i_1,j_1} \\
& + \omega_{i_1,i_2,j_3} B_{32}(\hat{\xi}_{i_1}, \hat{\xi}_{i_2}, \hat{\xi}_{j_3}) \hat{\varphi}'_{j_2}(\hat{\xi}_{i_2}) \hat{\varphi}'_{i_3}(\hat{\xi}_{j_3}) \delta_{i_1,j_1}
\end{aligned}$$

Le calcul de la matrice est donc plus rapide, de complexité  $O(r^5)$ , et il est même plus rapide que dans le cas d'éléments tétraédriques  $P_r$  en  $O(r^6)$ . Cette particularité est due au fait que seuls les degrés de liberté placés localement dans le même plan (parallèle à  $Oxy$ ,  $Oxz$  ou  $Oyz$ ) ont une interaction non-nulle. De même qu'en 2-D, on garde cette complexité pour les éléments courbes alors que pour les tétraèdres, la complexité est alors en  $O(r^9)$ .

## 1.2 Éléments finis courbes

On s'est inspiré de l'ouvrage [Solin *et al.*, 2003], qui donne les formules pour les quadrilatères, triangles, hexaèdres et tétraèdres.

### 1.2.1 Cas 2-D

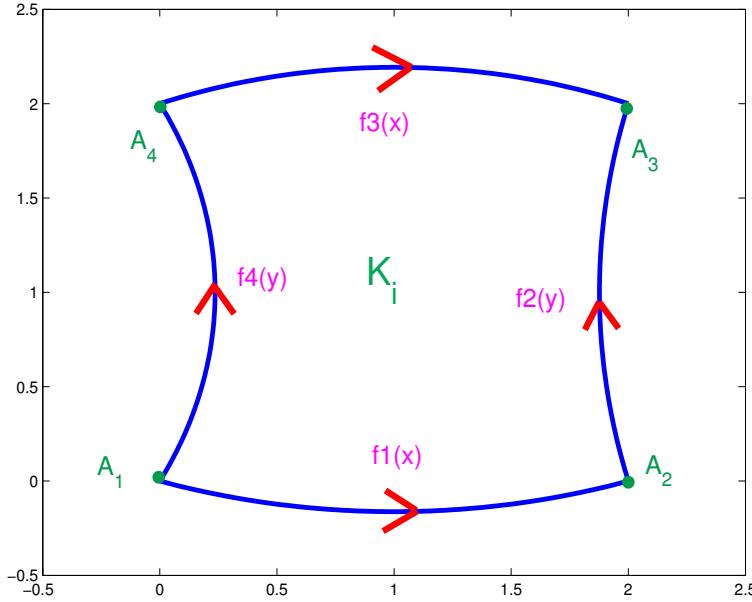


FIG. 1.5 – Quadrilatère courbe

On suppose connaitre la paramétrisation des quatre arêtes du maillage par les fonctions \$f\_1, f\_2, f\_3, f\_4\$ (pour les notations, cf. figure 1.5). \$f\_1\$ est une fonction qui va de l'intervalle \$[0, 1]\$ au plan \$\mathbb{R}^2\$. Elle vérifie des conditions de compatibilité avec les coins du quadrilatère :

$$f_1(0) = A_1 \quad f_1(1) = A_2$$

Sur les quadrilatères, on considère la transformation de Gordon-Hall [Gordon et Hall, 1973] :

$$\begin{aligned} \tilde{F}_i(\hat{x}_1, \hat{x}_2) &= \hat{x}_1 f_2(\hat{x}_2) + (1 - \hat{x}_1) f_4(\hat{x}_2) + \hat{x}_2 f_3(\hat{x}_1) + (1 - \hat{x}_2) f_1(\hat{x}_1) \\ &\quad - \left[ (1 - \hat{x}_1)(1 - \hat{x}_2) A_1 + \hat{x}_1(1 - \hat{x}_2) A_2 + \hat{x}_1 \hat{x}_2 A_3 + (1 - \hat{x}_1) \hat{x}_2 A_4 \right] \end{aligned} \quad (1.3)$$

On choisit de projeter les points de Gauss-Lobatto du carré unité \$\hat{K}\$ sur le quadrilatère courbe par cette transformation (cf. figure 1.6)

$$P_k = \tilde{F}_i(\hat{\xi}_{k1}, \hat{\xi}_{k2}) \quad \forall k = (k_1, k_2) \quad 1 \leq k_1, k_2 \leq r + 1$$

On utilise ensuite une interpolation lagrangienne comme transformation finale \$F\_i\$ entre le carré unité et le quadrilatère courbe.

$$F_i(\hat{x}_1, \hat{x}_2) = \sum_{k_1, k_2=1}^{r+1} P_{k_1, k_2} \hat{\varphi}_{k_1}(\hat{x}_1) \hat{\varphi}_{k_2}(\hat{x}_2) \quad (1.4)$$

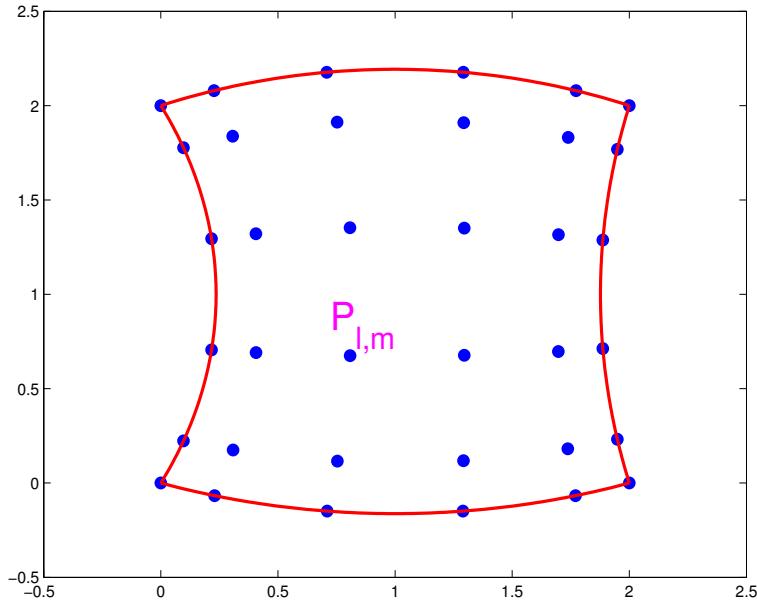


FIG. 1.6 – Projection des points de Gauss-Lobatto par la transformation de Gordon-Hall

où  $\hat{\varphi}_{k1}(\hat{x}_1) = \prod_{i=1, i \neq k1}^{r+1} \frac{\hat{x}_1 - \hat{\xi}_{k1}}{\hat{\xi}_i - \hat{\xi}_{k1}}$  est la fonction de Lagrange associée au point  $\hat{\xi}_{k1}$ . On a souvent besoin de la matrice jacobienne  $DF_i$ , on calcule les dérivées :

$$\begin{aligned} \frac{\partial F_i}{\partial \hat{x}_1}(\hat{x}_1, \hat{x}_2) &= \sum_{k1, k2=1}^{r+1} \frac{d\hat{\varphi}_{k1}}{d\hat{x}_1}(\hat{x}_1) \hat{\varphi}_{k2}(\hat{x}_2) P_{k1, k2} \\ \frac{\partial F_i}{\partial \hat{x}_2}(\hat{x}_1, \hat{x}_2) &= \sum_{k1, k2=1}^{r+1} \hat{\varphi}_{k1}(\hat{x}_1) \frac{d\hat{\varphi}_{k2}}{d\hat{x}_2}(\hat{x}_2) P_{k1, k2} \end{aligned} \quad (1.5)$$

Pour éviter de calculer les dérivées des fonctions de base, on va le faire une seule fois en calculant les valeurs

$$\frac{d\hat{\varphi}_i}{dx}(\hat{\xi}_j) \quad i, j = 1, (r+1)$$

Ce sont  $(r+1)^2$  valeurs à calculer une seule fois. On en déduit l'expression de  $DF_i$  sur les points de Gauss-Lobatto

$$\begin{aligned} \frac{\partial F_i}{\partial \hat{x}_1}(\hat{\xi}_{j1}, \hat{\xi}_{j2}) &= \sum_{k1=1}^{r+1} \frac{d\hat{\varphi}_{k1}}{d\hat{x}_1}(\hat{\xi}_{j1}) P_{k1, j2} \\ \frac{\partial F_i}{\partial \hat{x}_2}(\hat{\xi}_{j1}, \hat{\xi}_{j2}) &= \sum_{k2=1}^{r+1} \frac{d\hat{\varphi}_{k2}}{d\hat{x}_2}(\hat{\xi}_{j2}) P_{j1, k2} \quad j1, j2 = 1..(r+1) \end{aligned} \quad (1.6)$$

la dérivée d'une fonction de base se calcule à partir de ses valeurs sur les points de Gauss-Lobatto par simple interpolation :

$$\frac{d\hat{\varphi}_i}{d\hat{x}_1}(\hat{x}_1) = \sum_{j=1}^{r+1} \frac{d\hat{\varphi}_i}{d\hat{x}_1}(\hat{\xi}_j) \hat{\varphi}_j(\hat{x}_1) \quad (1.7)$$

On en déduit  $DF_i$  en tout point :

$$\begin{aligned}\frac{\partial F_i}{\partial \hat{x}_1}(\hat{x}_1, \hat{x}_2) &= \sum_{k1, k2, j=1}^{r+1} \frac{d\hat{\varphi}_{k1}}{d\hat{x}_1}(\hat{\xi}_j) \hat{\varphi}_j(\hat{x}_1) \hat{\varphi}_{k2}(\hat{x}_2) P_{k1, k2} \\ \frac{\partial F_i}{\partial \hat{x}_2}(\hat{x}_1, \hat{x}_2) &= \sum_{k1, k2, j=1}^{r+1} \hat{\varphi}_{k1}(\hat{x}_1) \frac{d\hat{\varphi}_{k2}}{d\hat{x}_2}(\hat{\xi}_j) \hat{\varphi}_j(\hat{x}_2) P_{k1, k2}\end{aligned}\quad (1.8)$$

En pratique, on ne doit connaître les matrices jacobienes aux seuls points de Gauss-Lobatto, car on utilise les formules d'intégration de Gauss-Lobatto pour calculer la matrice et le second membre.

### 1.2.2 Cas 3-D

Transformation générale  $F_i$

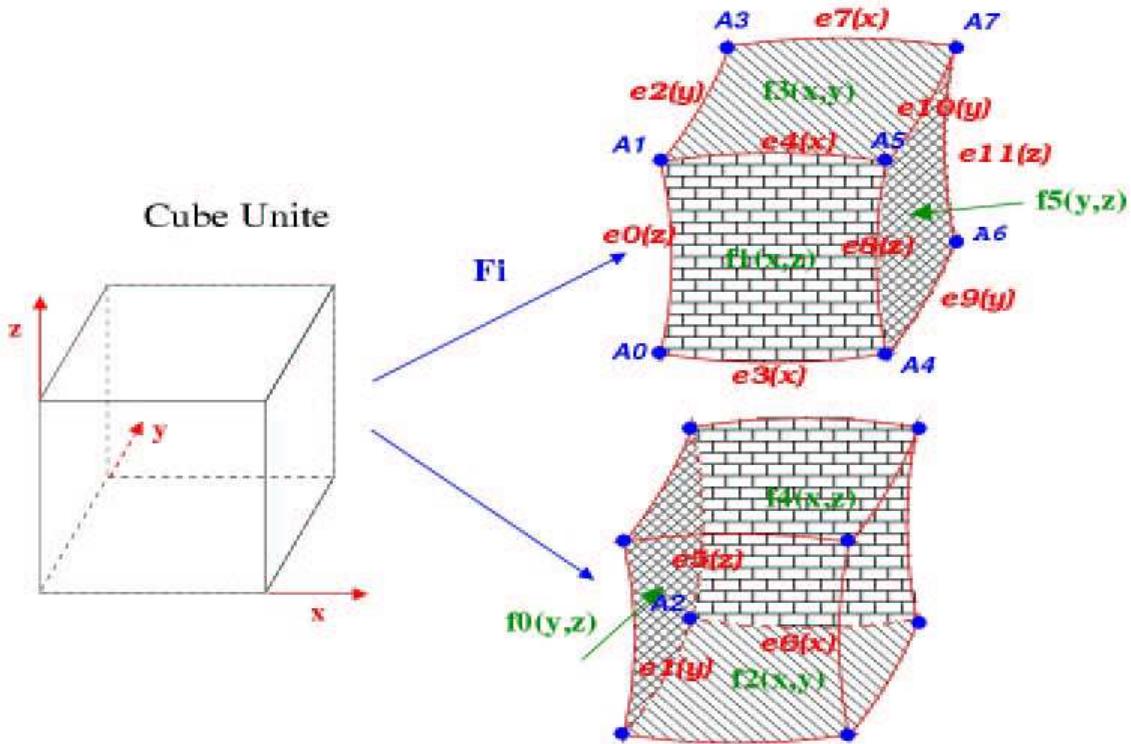


FIG. 1.7 – Hexaèdre Courbe

On suppose disposer de la paramétrisation des arêtes et des faces d'un hexaèdre (pour les notations, cf. figure 1.7). On considère, de manière similaire au cas 2-D, la transformation pour

passer d'un cube unité en un hexaèdre :

$$\begin{aligned}
\tilde{F}_i(\hat{x}_1, \hat{x}_2, \hat{x}_3) = & (1 - \hat{x}_1)(1 - \hat{x}_2)(1 - \hat{x}_3) A_0 + (1 - \hat{x}_1)(1 - \hat{x}_2)\hat{x}_3 A_1 + (1 - \hat{x}_1)\hat{x}_2(1 - \hat{x}_3) A_2 \\
& + (1 - \hat{x}_1)\hat{x}_2\hat{x}_3 A_3 + \hat{x}_1(1 - \hat{x}_2)(1 - \hat{x}_3) A_4 + \hat{x}_1(1 - \hat{x}_2)\hat{x}_3 A_5 \\
& + \hat{x}_1\hat{x}_2(1 - \hat{x}_3) A_6 + \hat{x}_1\hat{x}_2\hat{x}_3 A_7 \\
& - (1 - \hat{x}_1)(1 - \hat{x}_2) e0(\hat{x}_3) - (1 - \hat{x}_1)(1 - \hat{x}_3) e1(\hat{x}_2) - (1 - \hat{x}_1)\hat{x}_3 e2(\hat{x}_2) \\
& - (1 - \hat{x}_2)(1 - \hat{x}_3) e3(\hat{x}_1) - (1 - \hat{x}_2)\hat{x}_3 e4(\hat{x}_1) - (1 - \hat{x}_1)\hat{x}_2 e5(\hat{x}_3) \\
& - \hat{x}_2(1 - \hat{x}_3) e6(\hat{x}_1) - \hat{x}_2\hat{x}_3 e7(\hat{x}_1) - \hat{x}_1(1 - \hat{x}_2) e8(\hat{x}_3) \\
& - \hat{x}_1(1 - \hat{x}_3) e9(\hat{x}_2) - \hat{x}_1\hat{x}_3 e10(\hat{x}_2) - \hat{x}_1\hat{x}_2 e11(\hat{x}_3) \\
& + (1 - \hat{x}_1) f0(\hat{x}_2, \hat{x}_3) + \hat{x}_1 f5(\hat{x}_2, \hat{x}_3) + (1 - \hat{x}_2) f1(\hat{x}_1, \hat{x}_3) \\
& + \hat{x}_2 f4(\hat{x}_1, \hat{x}_3) + (1 - \hat{x}_3) f2(\hat{x}_1, \hat{x}_2) + \hat{x}_3 f3(\hat{x}_1, \hat{x}_2)
\end{aligned} \tag{1.9}$$

Cette transformation assure que :

$$\hat{x}_1 = 0 \implies \tilde{F}_i(\hat{x}_1, \hat{x}_2, \hat{x}_3) = f0(\hat{x}_2, \hat{x}_3)$$

On a la même propriété pour les autres faces. La vérification de cette propriété est aisée :

$$\begin{aligned}
\tilde{F}_i(0, \hat{x}_2, \hat{x}_3) = & (1 - \hat{x}_2)(1 - \hat{x}_3) A_0 + (1 - \hat{x}_2)\hat{x}_3 A_1 + \hat{x}_2(1 - \hat{x}_3) A_2 + \hat{x}_2\hat{x}_3 A_3 \\
& - (1 - \hat{x}_2)(1 - \hat{x}_3) e3(0) - (1 - \hat{x}_2)\hat{x}_3 e4(0) - \hat{x}_2(1 - \hat{x}_3) e6(0) - \hat{x}_2\hat{x}_3 e7(0) \\
& - (1 - \hat{x}_2) e0(\hat{x}_3) - (1 - \hat{x}_3) e1(\hat{x}_2) - \hat{x}_3 e2(\hat{x}_2) - \hat{x}_2 e5(\hat{x}_3) \\
& + (1 - \hat{x}_2) f1(0, \hat{x}_3) + \hat{x}_2 f4(0, \hat{x}_3) + (1 - \hat{x}_3) f2(0, \hat{x}_2) + \hat{x}_3 f3(0, \hat{x}_2) \\
& + f0(\hat{x}_2, \hat{x}_3)
\end{aligned} \tag{1.10}$$

Or, nous connaissons les extrémités des arêtes

$$e3(0) = A_0 \quad e4(0) = A_1 \quad e6(0) = A_2 \quad e7(0) = A_3 \tag{1.11}$$

Nous avons ainsi les deux premières lignes de l'équation (1.10) qui se suppriment. Nous connaissons aussi les bords des faces :

$$f1(0, \hat{x}_3) = e0(\hat{x}_3) \quad f4(0, \hat{x}_3) = e5(\hat{x}_3) \quad f2(0, \hat{x}_2) = e1(\hat{x}_2) \quad f3(0, \hat{x}_2) = e2(\hat{x}_2) \tag{1.12}$$

On a bien  $\tilde{F}_i(0, \hat{x}_2, \hat{x}_3) = f0(\hat{x}_2, \hat{x}_3)$ .

Une fois cette transformation établie, l'idée est semblable à ce qu'on a fait en 2-D : on projette les points de Gauss-Lobatto du cube unité sur l'hexaèdre courbe paramétrisé par les équations des six faces.

$$P_k = \tilde{F}_i(\hat{\xi}_{k1}, \hat{\xi}_{k2}, \hat{\xi}_{k3}) \quad \forall k = (k_1, k_2, k_3) \quad 1 \leq k_1, k_2, k_3 \leq r+1$$

$r+1$  est le nombre de points de Gauss-Lobatto.

On utilise ensuite une interpolation lagrangienne comme transformation finale  $F_i$  entre le cube unité et l'hexaèdre courbe :

$$F_i(\hat{x}_1, \hat{x}_2) = \sum_{k_1, k_2, k_3=1}^{r+1} P_{k_1, k_2, k_3} \hat{\varphi}_{k_1}(\hat{x}_1) \hat{\varphi}_{k_2}(\hat{x}_2) \hat{\varphi}_{k_3}(\hat{x}_3) \tag{1.13}$$

On a souvent besoin de la matrice jacobienne  $DF_i$ , on calcule les dérivées :

$$\begin{aligned}\frac{\partial F_i}{\partial \hat{x}_1}(\hat{x}_1, \hat{x}_2, \hat{x}_3) &= \sum_{k_1, k_2, k_3=1}^{r+1} \frac{d\hat{\varphi}_{k_1}}{d\hat{x}_1}(\hat{x}_1) \hat{\varphi}_{k_2}(\hat{x}_2) \hat{\varphi}_{k_3}(\hat{x}_3) P_{k_1, k_2, k_3} \\ \frac{\partial F_i}{\partial \hat{x}_2}(\hat{x}_1, \hat{x}_2, \hat{x}_3) &= \sum_{k_1, k_2, k_3=1}^{r+1} \hat{\varphi}_{k_1}(\hat{x}_1) \frac{d\hat{\varphi}_{k_2}}{d\hat{x}_2}(\hat{x}_2) \hat{\varphi}_{k_3}(\hat{x}_3) P_{k_1, k_2, k_3} \\ \frac{\partial F_i}{\partial \hat{x}_3}(\hat{x}_1, \hat{x}_2, \hat{x}_3) &= \sum_{k_1, k_2, k_3=1}^{r+1} \hat{\varphi}_{k_1}(\hat{x}_1) \hat{\varphi}_{k_2}(\hat{x}_2) \frac{d\hat{\varphi}_{k_3}}{d\hat{x}_3}(\hat{x}_3) P_{k_1, k_2, k_3}\end{aligned}\tag{1.14}$$

Pour éviter de calculer les dérivées des fonctions de base, on ne calcule que les valeurs :

$$\frac{d\hat{\varphi}_i}{d\hat{x}_1}(\hat{\xi}_j) \quad i, j = 1, (r+1)$$

Ce sont  $(r+1)^2$  valeurs à calculer une seule fois. On en déduit l'expression de  $DF_i$  sur les points de Gauss-Lobatto

$$\begin{aligned}\frac{\partial F_i}{\partial \hat{x}_1}(\hat{\xi}_{j_1}, \hat{\xi}_{j_2}, \hat{\xi}_{j_3}) &= \sum_{k_1=1}^{r+1} \frac{d\hat{\varphi}_{k_1}}{d\hat{x}_1}(\hat{\xi}_{j_1}) P_{k_1, j_2, j_3} \\ \frac{\partial F_i}{\partial \hat{x}_2}(\hat{\xi}_{j_1}, \hat{\xi}_{j_2}, \hat{\xi}_{j_3}) &= \sum_{k_2=1}^{r+1} \frac{d\hat{\varphi}_{k_2}}{d\hat{x}_2}(\hat{\xi}_{j_2}) P_{j_1, k_2, j_3} \\ \frac{\partial F_i}{\partial \hat{x}_3}(\hat{\xi}_{j_1}, \hat{\xi}_{j_2}, \hat{\xi}_{j_3}) &= \sum_{k_3=1}^{r+1} \frac{d\hat{\varphi}_{k_3}}{d\hat{x}_3}(\hat{\xi}_{j_3}) P_{j_1, j_2, k_3} \quad j_1, j_2, j_3 = 1..(r+1)\end{aligned}\tag{1.15}$$

La dérivée d'une fonction de base se calcule à partir de ses valeurs sur les points de Gauss-Lobatto par simple interpolation :

$$\frac{d\hat{\varphi}_i}{d\hat{x}_1}(\hat{x}_1) = \sum_{j=1}^{r+1} \frac{d\hat{\varphi}_i}{d\hat{x}_1}(\hat{\xi}_j) \hat{\varphi}_j(\hat{x}_1) \tag{1.16}$$

On en déduit  $DF_i$  en tout point :

$$\begin{aligned}\frac{\partial F_i}{\partial \hat{x}_1}(\hat{x}_1, \hat{x}_2, \hat{x}_3) &= \sum_{k_1, k_2, k_3, j=1}^{r+1} \frac{d\hat{\varphi}_{k_1}}{d\hat{x}_1}(\hat{\xi}_j) \hat{\varphi}_j(\hat{x}_1) \hat{\varphi}_{k_2}(\hat{x}_2) \hat{\varphi}_{k_3}(\hat{x}_3) P_{k_1, k_2, k_3} \\ \frac{\partial F_i}{\partial \hat{x}_2}(\hat{x}_1, \hat{x}_2, \hat{x}_3) &= \sum_{k_1, k_2, k_3, j=1}^{r+1} \hat{\varphi}_{k_1}(\hat{x}_1) \frac{d\hat{\varphi}_{k_2}}{d\hat{x}_2}(\hat{\xi}_j) \hat{\varphi}_j(\hat{x}_2) \hat{\varphi}_{k_3}(\hat{x}_3) P_{k_1, k_2, k_3} \\ \frac{\partial F_i}{\partial \hat{x}_3}(\hat{x}_1, \hat{x}_2, \hat{x}_3) &= \sum_{k_1, k_2, k_3, j=1}^{r+1} \hat{\varphi}_{k_1}(\hat{x}_1) \hat{\varphi}_{k_2}(\hat{x}_2) \frac{d\hat{\varphi}_{k_3}}{d\hat{x}_3}(\hat{\xi}_j) \hat{\varphi}_j(\hat{x}_3) A_{k_1, k_2, k_3}\end{aligned}\tag{1.17}$$

Là aussi, il ne sera nécessaire en pratique de connaître  $DF_i$  uniquement aux points de Gauss-Lobatto.

## 1.3 Précision de la méthode

Nous présentons ici uniquement des résultats numériques qui valident la méthode éléments finis utilisée, et qui suggèrent des ordres de convergence. Pour des estimations d'erreurs théoriques, on pourra se référer à [Arnold *et al.*, 2000] et [Grob, 2006]. Sur des cas académiques, on compare les différents ordres d'approximation afin de déterminer s'il est intéressant de monter en ordre sur des géométries lisses, des géométries présentant des arêtes vives ou des coins.

### 1.3.1 Cas 2-D

#### Diffraction par un disque

Nous vérifions la convergence des éléments finis spectraux sur le cas académique de la diffraction par un disque. Le problème modèle s'écrit :

$$\left\{ \begin{array}{l} \text{Trouver } u \in H^1(\Omega) \\ -k^2 u - \Delta u = 0 \quad \in \Omega \\ u = -u^{inc} = -\exp(ikx) \quad \text{pour } r = a \\ \frac{\partial u}{\partial n} - iku = 0 \quad \text{pour } r = b \end{array} \right. \quad (1.18)$$

$k$  est le nombre d'onde, il est égal à  $\omega$ , car on a choisi une vitesse de propagation de 1. Le domaine de calcul  $\Omega$  est la couronne comprise entre les deux cercles de rayon  $a = 1$  et  $b = 2$ . La solution analytique de ce problème est représentée sur la figure 1.8. L'expression des solutions analytiques pour la diffraction d'une sphère ou d'un disque, est rappelée succinctement en annexe A. On peut voir sur les figures 1.9 et 1.10 l'évolution de l'erreur entre la solution

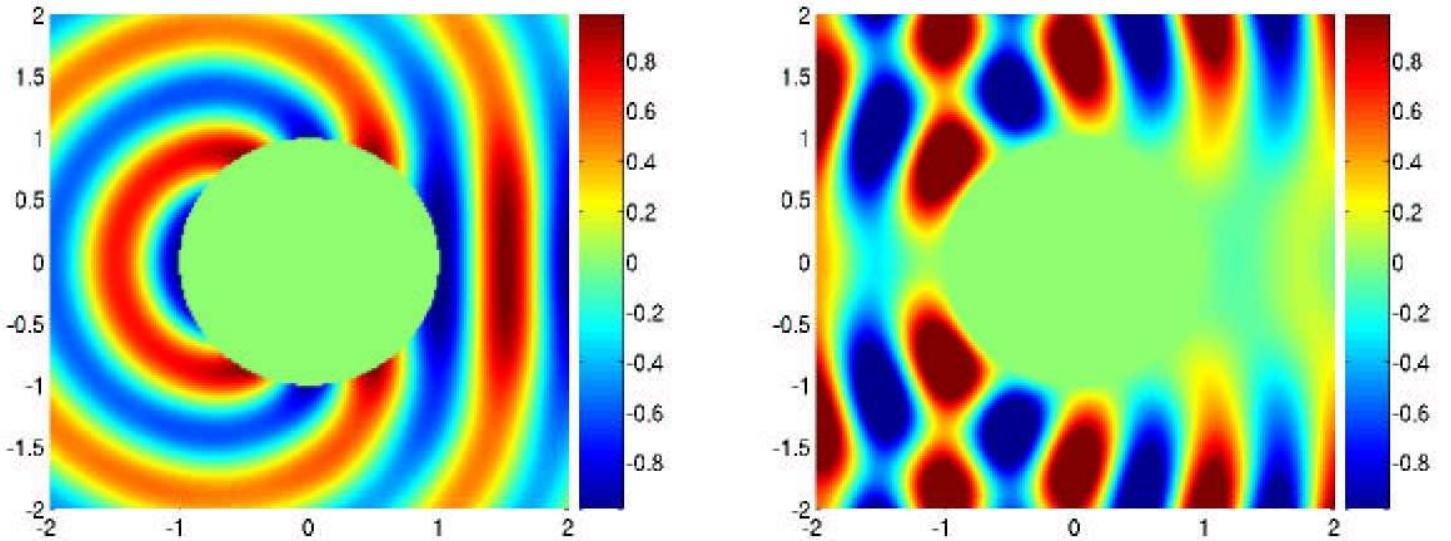


FIG. 1.8 – A gauche, partie réelle du champ diffracté, à droite, partie réelle du champ total. Comme on a une expression analytique de la solution, on peut la calculer sur tout le carré  $[-2, 2]$ .

numérique et la solution analytique respectivement en fonction du pas de maillage et en fonction du nombre de degrés de liberté. Pour cette dernière courbe, on utilise la grandeur  $h/r$  en lieu et place du nombre de degrés de liberté (ddl) effectif,  $r$  étant l'ordre d'approximation. En effet,

sur un maillage régulier la grandeur  $h/r$  est reliée au nombre de ddl, par une relation du type :

$$\left(\frac{h}{r}\right)^2 = \frac{C}{N_{ddl}}$$

où  $C$  est une constante indépendante de  $r$ . On notera que l'ordre de convergence est optimal

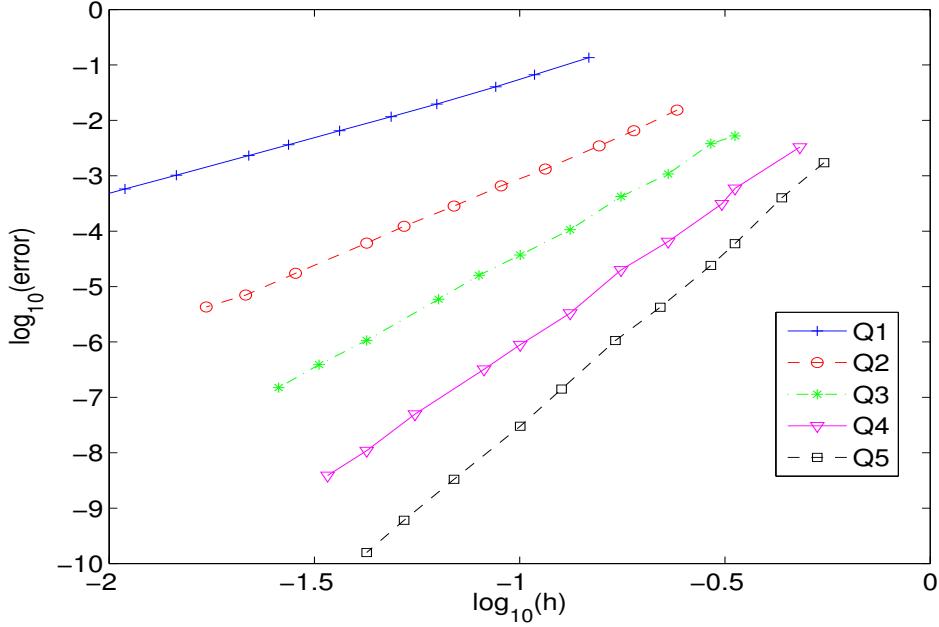


FIG. 1.9 – Evolution de l'erreur  $L^2$  en fonction du pas de maillage, échelle log-log. Cas du disque sur des maillages pseudo-réguliers.  $\log_{10}(h) = -1$  correspond à  $h = \frac{\lambda}{10}$ .

en  $h^{r+1}$  et que, pour des pas de maillages utilisés en pratique, l'ordre élevé apporte un gain de précision substantiel pour un même nombre de degrés de liberté. On remarque que l'ordre de convergence en norme  $H^1$  est en  $h^r$  sur la figure 1.11. Pour finir sur le cas du disque, nous utilisons des maillages non-structurés, obtenus par découpage de maillages triangulaires (cf. figure 2.8). Chaque triangle est découpé en trois quadrilatères. Sur la figure 1.12, on a représenté l'erreur  $L^2$  en fonction du pas de maillage. On mesure des pentes de 2.09, 3.03 et 4.03 pour respectivement  $Q_1$ ,  $Q_2$  et  $Q_3$ . Ces résultats tendent à nous faire penser qu'on garde l'ordre de convergence optimal  $O(h^{r+1})$ , sur des maillages non-structurés et en utilisant la condensation de masse.

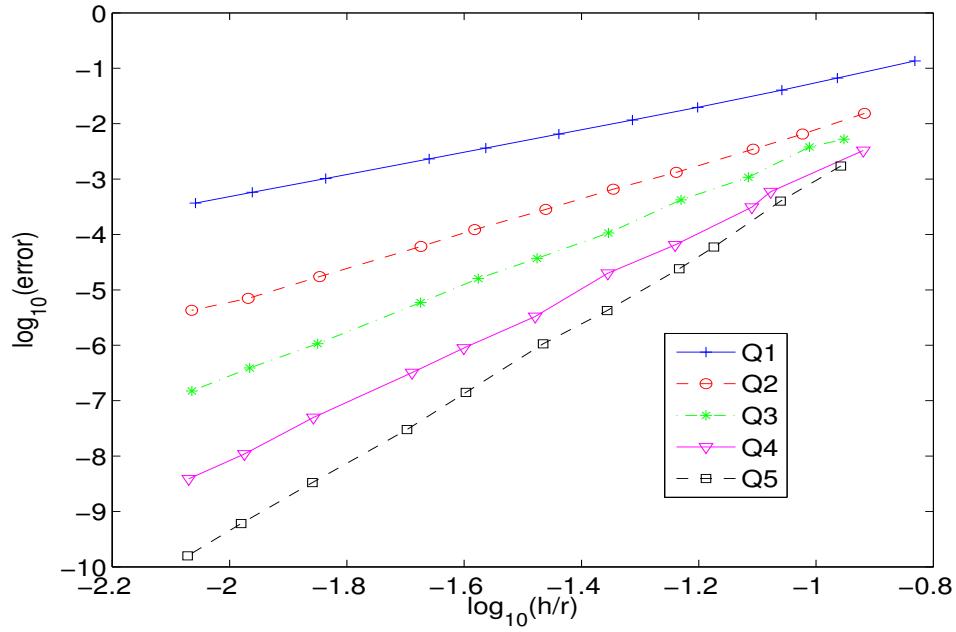


FIG. 1.10 – Evolution de l'erreur  $L^2$  en fonction de  $\frac{h}{r}$ , échelle log-log. Cas du disque sur des maillages pseudo-réguliers.

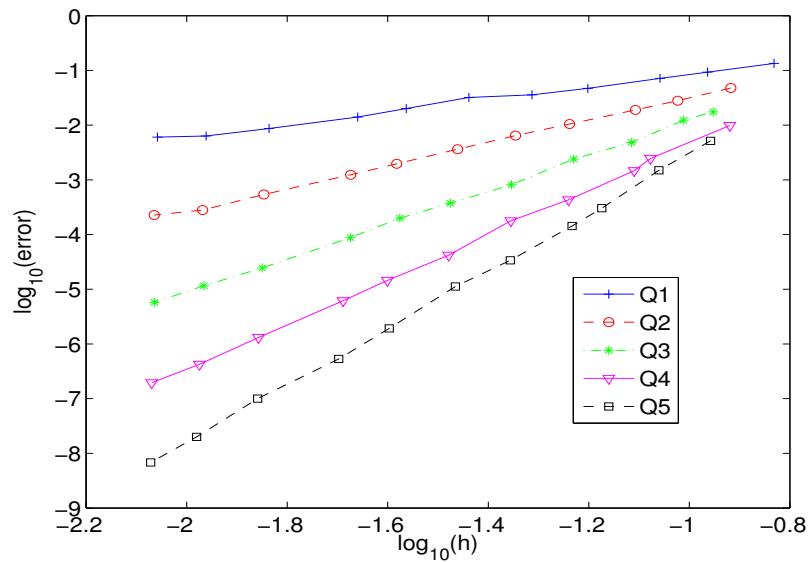


FIG. 1.11 – Evolution de l'erreur  $H^1$  en fonction de  $\frac{h}{r}$ , échelle log-log. Cas du disque sur des maillages pseudo-réguliers.

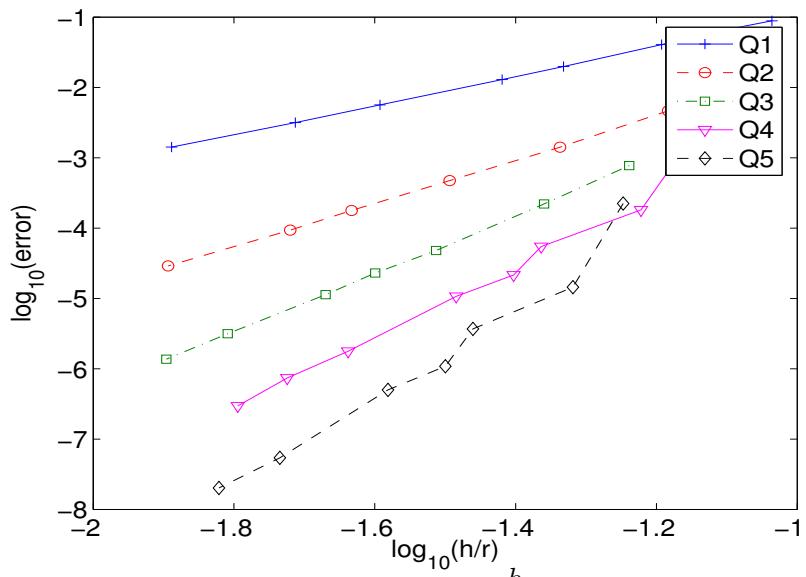


FIG. 1.12 – Evolution de l'erreur  $L^2$  en fonction de  $\frac{h}{r}$ , échelle log-log. Cas du disque sur des maillages triangulaires découpés.

## Diffracton par un carré

On étudie la diffracton d'une onde plane par un carré (voir figure 1.13). Dans le cas d'obs-

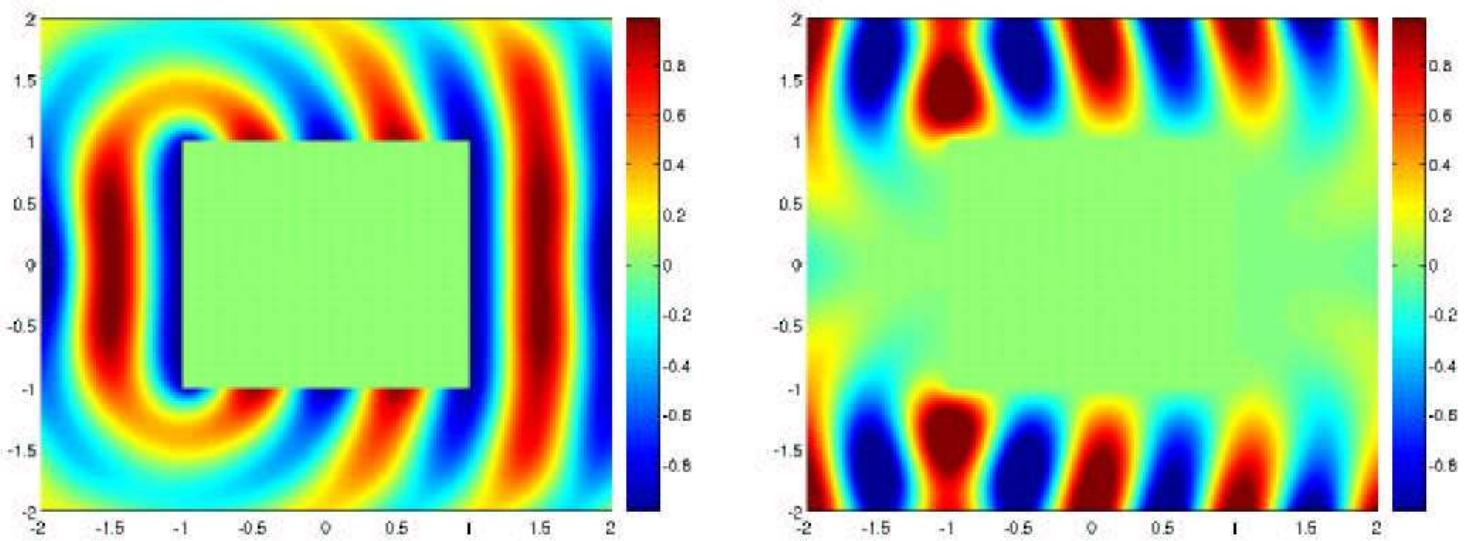


FIG. 1.13 – A gauche, partie réelle du champ diffracté, à droite, partie réelle du champ total

tacles à frontière  $C^1$  comme pour le disque, la convergence de la méthode éléments finis pour un ordre élevé est très rapide. Par conséquent, il est avantageux de monter en ordre. Pour des géométries présentant des coins, la convergence est plus laborieuse comme nous le constatons sur les figures 1.14 et 1.15. L'erreur faite sur la solution numérique est mesurée à partir d'une solution de référence calculée avec une approximation  $Q_4$  et avec un pas de maillage cinq fois plus petit. On observe un ordre de convergence en  $h^{\frac{4}{3}}$  quel que soit l'ordre, et pour la norme  $L^2$

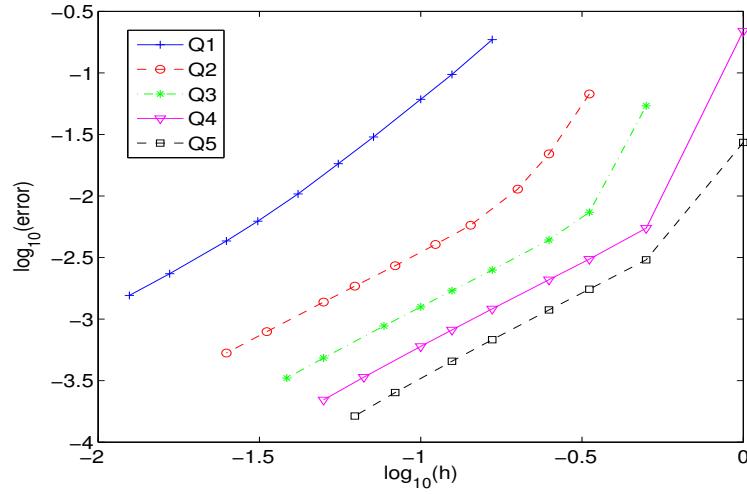


FIG. 1.14 – Evolution de l'erreur  $L^2$  en fonction de  $h$ , échelle log-log. Cas du carré sur des maillages réguliers.

ou  $H^1$ . Ce phénomène est confirmé par la théorie, qui prévoit un ordre de convergence de  $\frac{2\pi}{\alpha} \cdot \alpha$

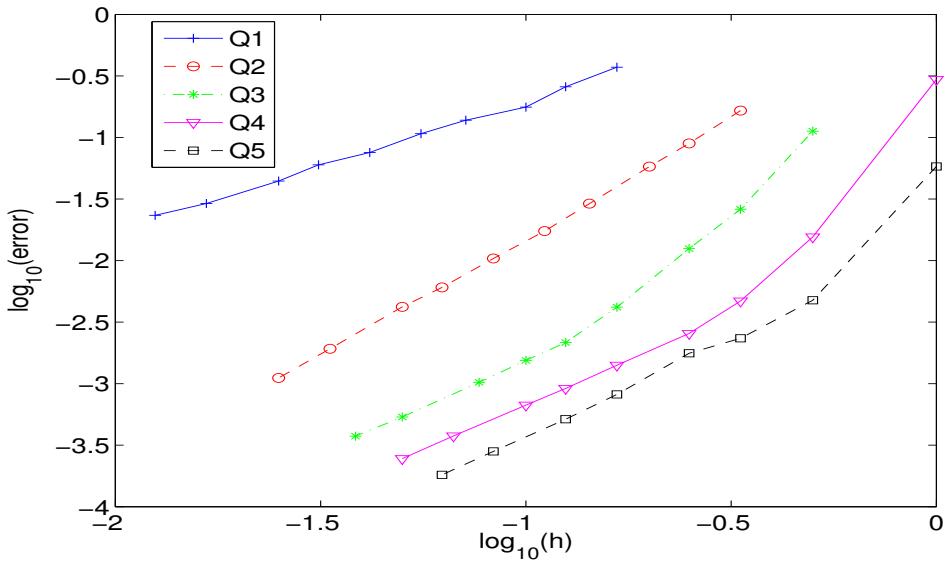


FIG. 1.15 – Evolution de l'erreur  $H^1$  en fonction de  $h$ , échelle log-log. Cas du carré sur des maillages réguliers.

est l'angle convexe que forme le coin, pour le carré  $\alpha = \frac{3\pi}{2}$ , ce qui donne bien un ordre de  $\frac{4}{3}$ . On se pose la question suivante : “L'ordre élevé est-il plus précis à nombre de ddl constant ?”. La réponse est oui, comme l'illustre la figure 1.16. Néanmoins, l'avantage de monter en ordre

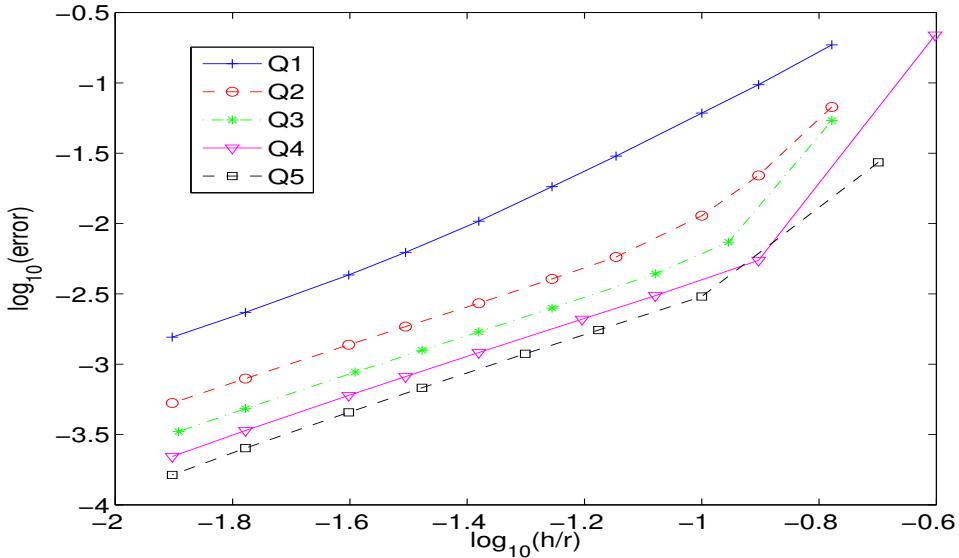


FIG. 1.16 – Evolution de l'erreur  $L^2$  en fonction de  $\frac{h}{r}$ , échelle log-log. Cas du carré sur des maillages réguliers.

(au delà de  $Q_3$ ) semble mince...

### 1.3.2 Cas 3-D

#### Diffraction par une sphère

Nous vérifions la convergence des éléments finis spectraux sur le cas académique de la diffraction par une sphère. Le problème modèle s'écrit :

$$\left\{ \begin{array}{l} \text{Trouver } u \in H^1(\Omega) \\ -k^2 u - \Delta u = 0 \quad \in \Omega \\ u = -u^{inc} = -\exp(ikx) \quad \text{pour } r = a \\ \frac{\partial u}{\partial n} - iku = 0 \quad \text{pour } r = b \end{array} \right. \quad (1.19)$$

Le domaine de calcul  $\Omega$  est la couronne comprise entre les deux sphères de rayon  $a = 1$  et  $b = 1.5$ . La solution analytique de ce problème est représentée sur la figure 1.17. On peut

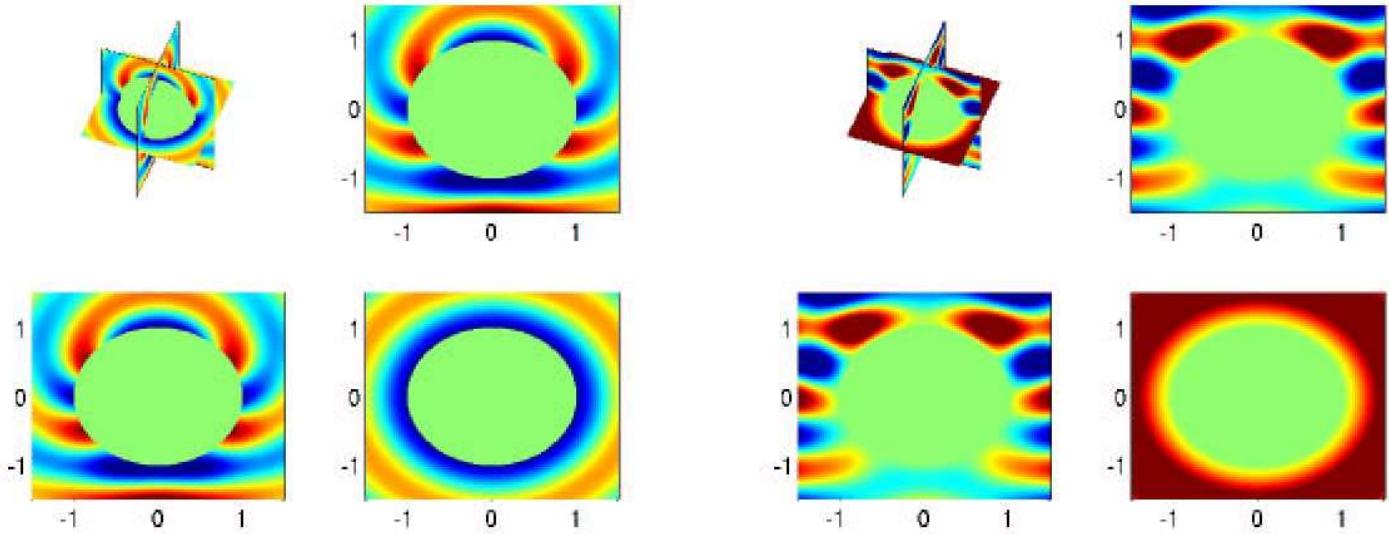


FIG. 1.17 – A gauche, partie réelle du champ diffracté, à droite, partie réelle du champ total

voir sur les figures 1.18 et 1.19 l'évolution de l'erreur entre la solution numérique et la solution analytique respectivement en fonction du pas de maillage ou en fonction du nombre de degrés de liberté. On notera que l'ordre de convergence est bien de  $h^{r+1}$  et que pour des pas de maillages utilisés en pratique, l'ordre élevé apporte un gain de précision substantiel pour un même nombre de degrés de liberté. Sur des maillages non-réguliers, il est difficile d'obtenir de jolies droites et d'estimer l'ordre de convergence. Des calculs sur l'erreur de dispersion, effectués au chapitre 3, nous mènent à penser que la condensation de masse nous fait perdre un ordre de convergence sur des maillages tétraédriques découpés.

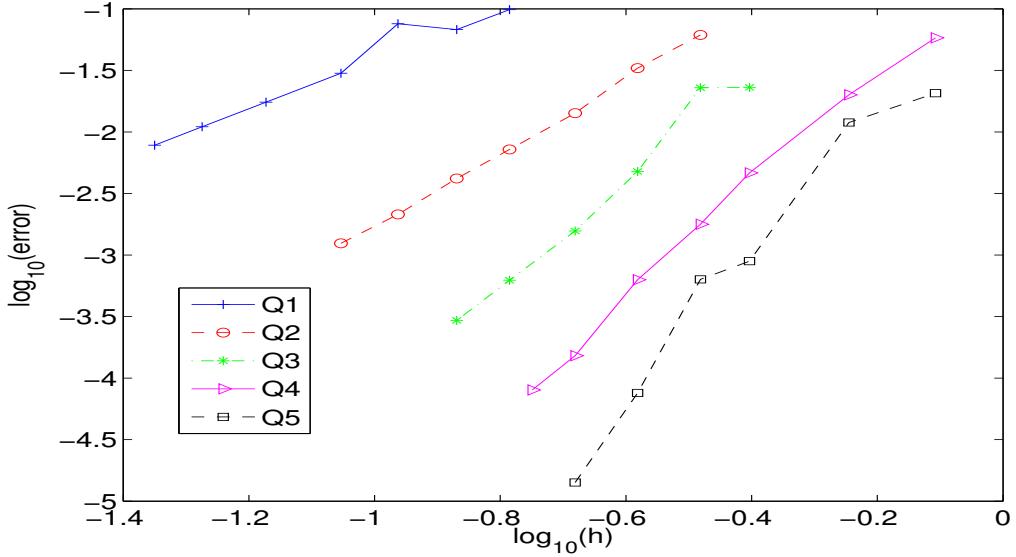


FIG. 1.18 – Evolution de l’erreur  $L^2$  en fonction du pas de maillage, échelle log-log. Cas de la sphère sur des maillages pseudo-réguliers.

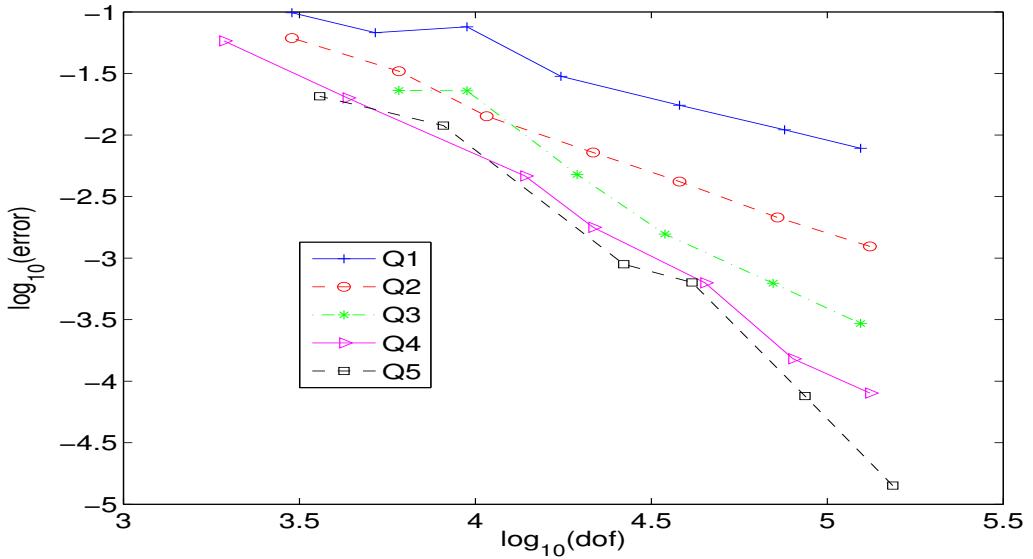


FIG. 1.19 – Evolution de l’erreur  $L^2$  en fonction du nombre de ddl, échelle log-log. Cas de la sphère sur des maillages pseudo-réguliers. En abscisse, on mesure le logarithme du nombre de degrés de liberté (numérique), sans passer par la grandeur  $h/r$ .

## Diffractioп par un parallépipède

Comme dans le cas 2-D, on s'intéresse à la diffraction par un cube, qui présente à la fois des arêtes vives et des coins. On représente la solution numérique sur la figure 1.20. On fait

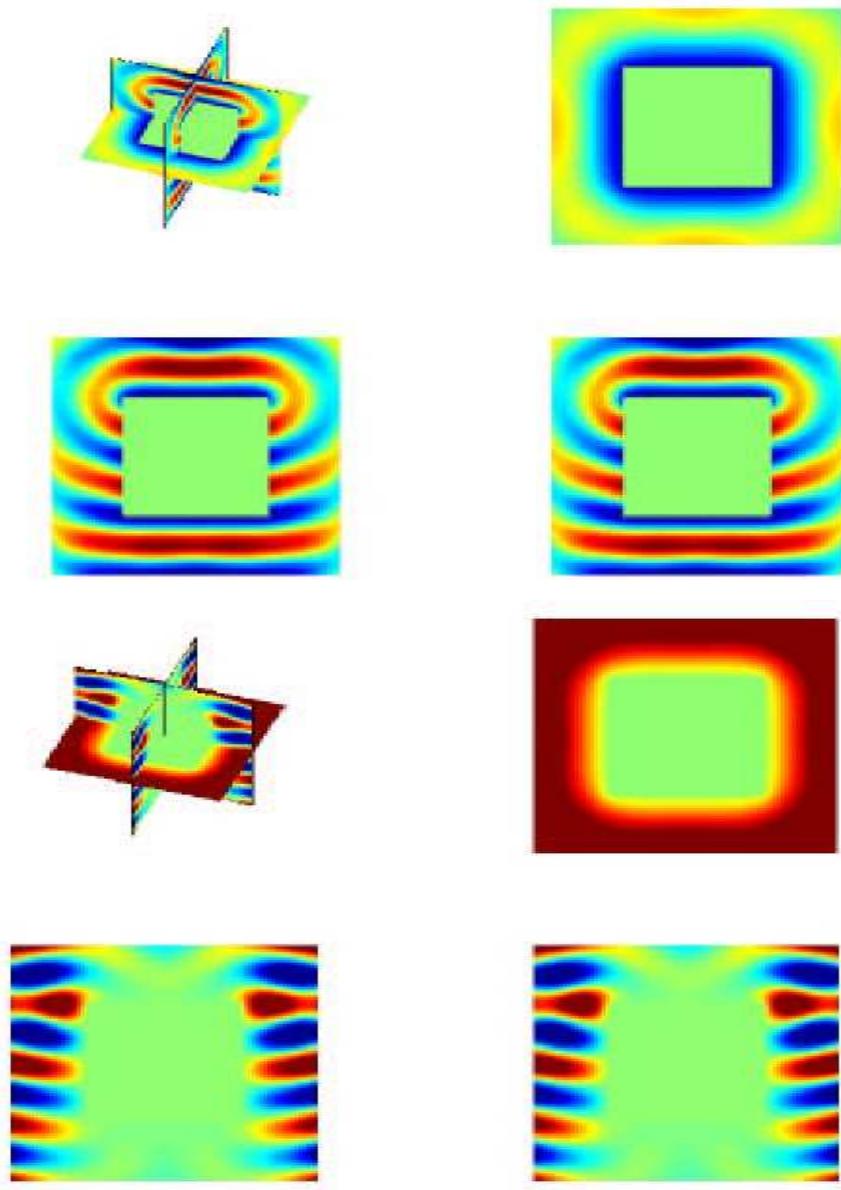


FIG. 1.20 – En haut, partie réelle du champ diffracté, en bas, partie réelle du champ total

une étude de convergence en prenant trois maillages de pas  $h$ ,  $\frac{h}{2}$  et  $\frac{h}{4}$  avec du  $Q_4$ . On note les solutions obtenues sur chaque maillage  $u_1$ ,  $u_2$ ,  $u_3$ . L'ordre de convergence se calcule par la formule :

$$r = \frac{\log(\|u_2 - u_1\|) - \log(\|u_3 - u_2\|)}{\log(2)}$$

On trouve numériquement  $r = 1.38$ . On peut supposer que l'ordre de convergence théorique est identique à celui qu'on avait en 2-D, à savoir  $\frac{4}{3}$ .

## 1.4 Application aux filtres optiques, intérêt de l'ordre élevé

Dans ce paragraphe, nous étudions un dispositif optique appelé filtre à hyper-fréquence de Fabry-Perrot. Ce cas nous a été fourni par la défunte société ATMEL. Nous rappelons, dans un premier temps, le type d'onde incidente utilisée : les faisceaux gaussiens. Dans un second temps, nous présenterons les principales caractéristiques du dispositif. Nous terminerons par une apologie de l'ordre élevé sur ce cas.

### 1.4.1 Faisceaux gaussiens

Les faisceaux gaussiens sont des solutions approchées de l'équation de Helmholtz. Ils sont utilisés pour simuler des ondes se propageant selon un axe privilégié. Ils sont construits de la manière suivante.

Si  $Oz$  est l'axe privilégié et  $Ox$  l'axe perpendiculaire, on recherche des solutions de

$$k^2 u + \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial z^2} = 0$$

sous la forme

$$u(x, z) = v(x, z)e^{ikz}$$

Un calcul simple montre que  $v$  vérifie

$$\frac{\partial^2 v}{\partial z^2} + \left( 2ik \frac{\partial v}{\partial z} + \frac{\partial^2 v}{\partial x^2} \right) = 0$$

Si  $v(x, z)$  varie lentement en fonction de  $z$ , on peut négliger la variation seconde de  $v(x, z)$  devant le terme entre parenthèses, il reste

$$2ik \frac{\partial v}{\partial z} + \frac{\partial^2 v}{\partial x^2} = 0 \quad (1.20)$$

soit, si l'on revient à  $u$

$$ik \frac{\partial u}{\partial z} + k^2 u - \frac{1}{2} \frac{\partial^2 u}{\partial x^2} = 0.$$

Cette équation est connue sous le nom d'approximation parabolique de l'équation de Helmholtz. Cette équation possède également des solutions ondes planes de la forme  $e^{i(k_x x + k_z z)}$  avec

$$k_z = ik \left( 1 - \frac{k_x^2}{2k^2} \right)$$

Cette relation de dispersion est à comparer avec la relation de dispersion classique pour laquelle

$$k_z = \pm ik \sqrt{1 - \frac{k_x^2}{k^2}}$$

On voit ainsi que l'équation parabolique ne rend compte que des ondes se dirigeant dans une seule direction (il n'y a qu'un seul  $k_z$  pour l'équation parabolique contre deux pour l'équation de Helmholtz) et que le nombre d'onde en  $z$  est approché à l'ordre 2 car

$$\sqrt{1 - x^2} = 1 - \frac{x^2}{2} + O(x^4)$$

Maintenant si l'on revient à (1.20), on peut vérifier que

$$v(x, z) = e^{-\frac{1}{2} \log(1+i\frac{z}{z_0}) - \frac{k_x^2}{z_0(1+i\frac{z}{z_0})}}$$

en est, pour tout  $z_0$ , une solution particulière ; la fonction

$$u(x, z) = e^{-\frac{1}{2} \log(1+i\frac{z}{z_0}) - \frac{kx^2}{z_0(1+i\frac{z}{z_0})} + ikz}$$

est appelé faisceau gaussien. En  $z = 0$ , la trace du faisceau est

$$u(x, 0) = e^{-\frac{1}{2} \frac{kx^2}{z_0}} = e^{-\frac{x^2}{w^2}}$$

où

$$w = \sqrt{\frac{2z_0}{k}}$$

est le col du faisceau. Le faisceau est dit large lorsque  $w$  vaut plusieurs longueurs d'onde ( $w$  est appelé aussi waist du faisceau)

$$w = N_\lambda \lambda$$

avec, pour fixer les idées,  $N_\lambda = 3, 6, 10, \dots$ . On a alors

$$z_0 = \pi N_\lambda^2 \lambda$$

Maintenant, si l'on calcule l'erreur sur l'équation d'Helmholtz, adimensionalisée par  $k^{-2}$ ,

$$\varepsilon(x, z) = u + \frac{1}{k^2} \Delta u$$

On trouve

$$\varepsilon(x, z) = \frac{-e^{ik_z z}}{4w^4 k^4 (1+i\frac{z}{z_0})^5} \frac{d^4 e^{-t^2}}{dt^4} (t = t(x, z)), \quad t(x, z) = \frac{x}{w \sqrt{1+i\frac{z}{z_0}}},$$

et on a donc

$$\sup_{(x, z), z \geq 0} |\varepsilon(x, z)| = \frac{C_0}{4w^4 k^4}, \quad \text{avec } C_0 \simeq 12$$

ce qui est petit, puisque

$$w^4 k^4 = (2\pi N_\lambda)^4;$$

l'approximation est donc d'autant meilleure que le col est grand.  $w$  sera appelé indifféremment largeur du faisceau gaussien, ou "waist".

### 1.4.2 Propriétés du filtre optique

#### Description générale du dispositif

Le dispositif étudié est un empilement de couches de diélectrique (InP) séparées par des couches d'air, comme le montre la figure 1.21. On note la longueur d'onde nominale du dispositif  $\lambda_0$ . L'écart entre deux couches d'InP est de  $0.25 \lambda_0$ , excepté pour la cavité centrale, dont la largeur est de  $\lambda_0$ . La largeur des couches d'InP est de  $\frac{5\lambda_0}{4n}$  où  $n$  est l'indice de réfraction du milieu. On prendra pour toutes les expériences numériques :

$$\lambda = 1.55 \mu m$$

$$n = 3.155 \quad \rho = n^2$$

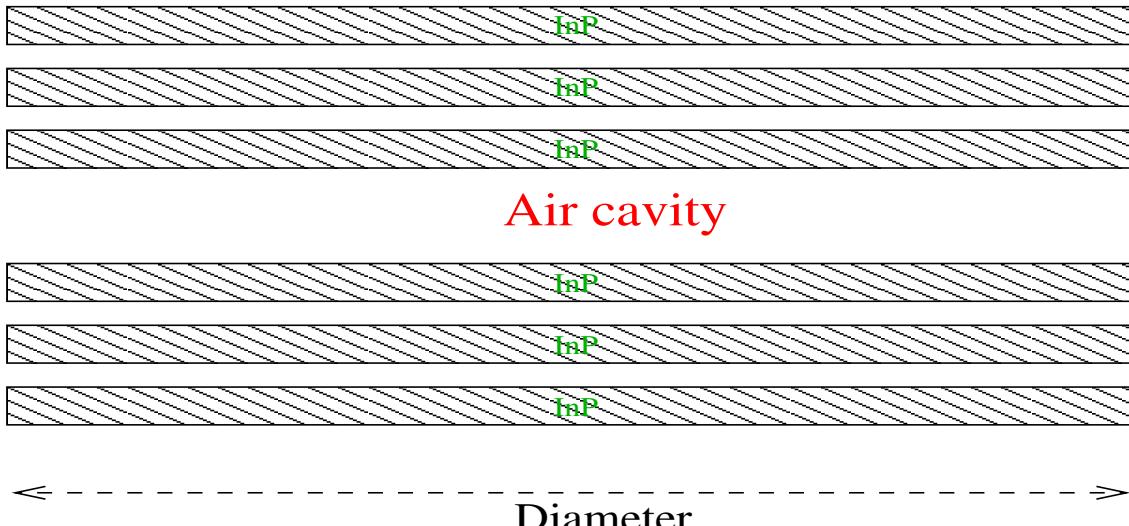


FIG. 1.21 – Dispositif Atmel

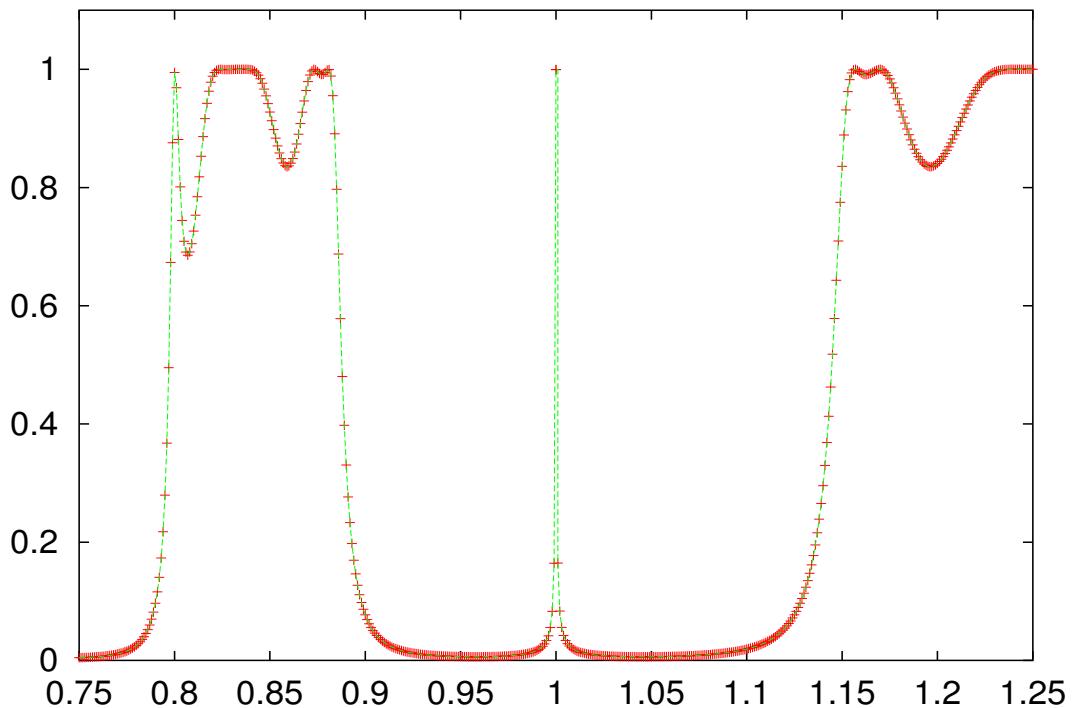


FIG. 1.22 – Coefficient de transmission du dispositif Atmel en fonction de la fréquence.

Toutes les grandeurs sont adimensionnalisées, les **unités** sur les axes sont des **longueurs d'onde** dans le vide. De même, on adimensionnalise la fréquence par rapport à la fréquence  $f_0 = \frac{c_0}{\lambda_0}$ . On parle alors de fréquence relative  $F = \frac{f}{f_0}$ . Une première étude par ondes planes nous fournit le coefficient de transmission à la sortie du dispositif suivant la fréquence relative. Sur la figure 1.22, on peut constater que le dispositif réalise bien une fonction de filtre passe-bande très sélectif. La fréquence  $F = 1$  s'apparente à une fréquence de résonance, pour laquelle le coefficient de

transmission est de 1. Dès qu'on s'écarte légèrement de cette fréquence, le coefficient retombe à 0.

## Maillage utilisé

On s'intéresse maintenant à la simulation numérique. Le premier obstacle qui s'est présenté a été la réalisation du maillage. La plupart des meilleurs réalisent des maillages en triangles. On peut néanmoins découper les triangles en trois et les quadrangles en quatre, ce qui nous permet de mailler n'importe quel domaine en quadrangles. Malheureusement, la méthode éléments finis s'avère plus efficace sur des maillages réguliers. On a donc choisi de développer un outil de maillage spécifique à un ensemble d'empilements, comme le dispositif Atmel. Le maillage “optimal” (pour  $Q_5$ ) trouvé lors des simulations est présenté sur la figure 1.23.

La structure particulière du dispositif privilégie certains ordres notamment  $Q_5$  et  $Q_7$ . On ne maille que la moitié du domaine et on reconstitue le reste de la solution par symétrie... Sur l'axe de symétrie, on impose une condition de Neumann, on s'astreint à ne traiter que le cas de faisceaux gaussiens à incidence nulle. Dans le cas de faisceaux gaussiens quelconques, il faudrait décomposer la source en partie symétrique et antisymétrique. Pour la partie symétrique, on imposerait une condition de Neumann, et pour la partie antisymétrique une condition de Dirichlet. On utilise des PML (Perfectly Matched Layers) au lieu de la condition absorbante

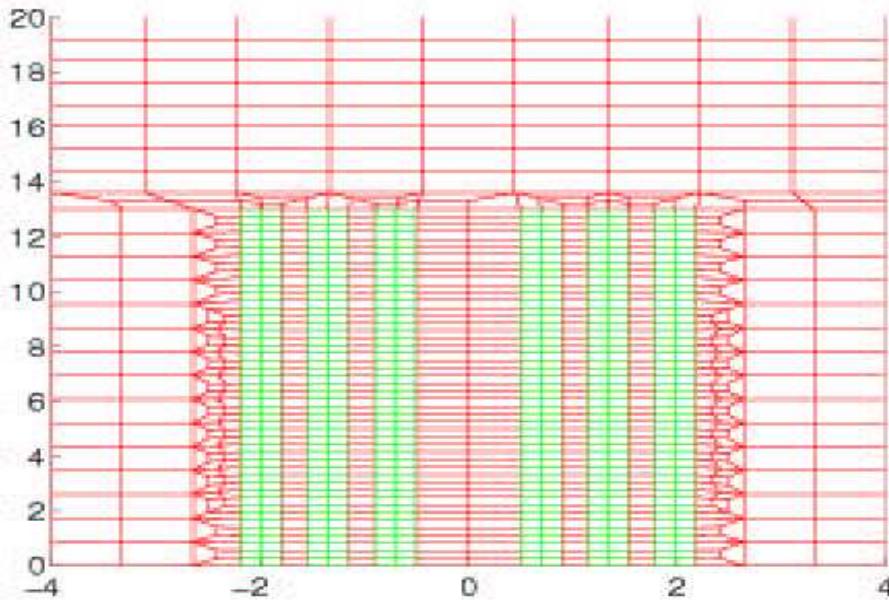


FIG. 1.23 – Maillage du domaine  $x \geq 0$  utilisé avec les éléments  $Q_5$ . Les diélectriques sont en vert. Les unités sont comptées en  $\lambda_0 = 1.55\mu m$  : les longueurs physiques sont obtenues en multipliant par  $1.55\mu m$ .

d'ordre 1, afin de simuler de manière plus fine un domaine non-borné. On ne décrira pas les PML dans cette thèse ; on renvoie le lecteur à la description de G. Cohen dans son livre [Cohen, 2002]. Les couches PML sont rajoutées sur tout le pourtour du domaine de calcul, après la construction du maillage. C'est pour cette raison, qu'on ne les voit pas sur les figures. L'utilisation de PML est particulièrement adapté pour ce type de géométrie “cartésienne”. Pour d'autres géométries, on préférera bien souvent l'introduction d'une condition transparente. Cette dernière est décrite en annexe B.

## Résultats avec des bords droits

La figure 1.24 montre la solution obtenue pour une fréquence à 80% de la fréquence de résonance. On retrouve le cas 1-D, avec un coefficient de transmission égal à 1 pour une fréquence de 0.8. C'est évidemment une fréquence hors "régime de fonctionnement" du filtre optique. Comme l'a montré le diagramme 1-D, la plage de fonctionnement du dispositif est entre  $F = 0.95$  et  $F = 1.05$ .

On prend maintenant des couches de diamètre  $112\lambda$  soit  $173.6\mu\text{m}$ . La solution obtenue est présentée dans la figure 1.25. On remarque que le champ s'étale quasiment dans tout le dispositif, mais il n'atteint pas les bords de manière sensible. Le dispositif étant suffisamment grand, les bord n'interviennent donc pas. On va maintenant étudier l'effet des bords, en prenant un dispositif plus petit. On prend un dispositif de diamètre  $26\lambda$ , soit  $40.2\mu\text{m}$ . On observe un décalage de la fréquence de résonance, qui est de  $1+10^{-4}$  au lieu de 1, dans le cas 1-D. Ce décalage est probablement dû à l'approximation utilisée pour l'implémentation d'un faisceau gaussien et à la présence de bords. A cette fréquence, nous avons tracé la solution sur la figure 1.27. On a effectué un balayage en fréquence de la solution. On observe des comportements semblables au cas 1-D, avec l'apparition de bosses pour des fréquences supérieures à 1, relativement à la fréquence de résonance du système. On représente sur la figure 1.26 quelques solutions à des fréquences proches de la fréquence de résonance. On observe des "rebonds", la courbe de transmission n'est plus une gaussienne. On a également des oscillations, qui sont dues à la présence de bords. Pour des fréquences inférieures à 1, on a une gaussienne, dont la hauteur diminue quand on s'éloigne de la fréquence. On n'observe pas de rebonds.

## Résultats en prenant des bords courbes

On perturbe le système en courbant légèrement les 2 couches diélectriques de part et d'autre de la cavité. La déformation crée alors une cavité optique concave qui se referme légèrement sur elle-même.

On effectue un balayage en fréquence afin de déterminer la fréquence de "résonance", pour

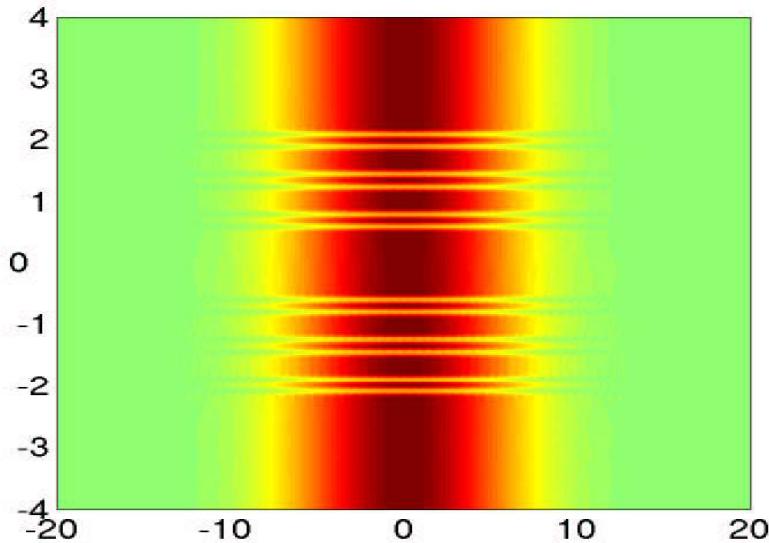


FIG. 1.24 – Module du champ total pour une fréquence relative de 0.8 (ce qui correspond à augmenter la longueur d'onde de  $\frac{1}{0.8}$  par rapport à  $\lambda_0$ ) et un waist de  $w = 4.96\mu\text{m}$

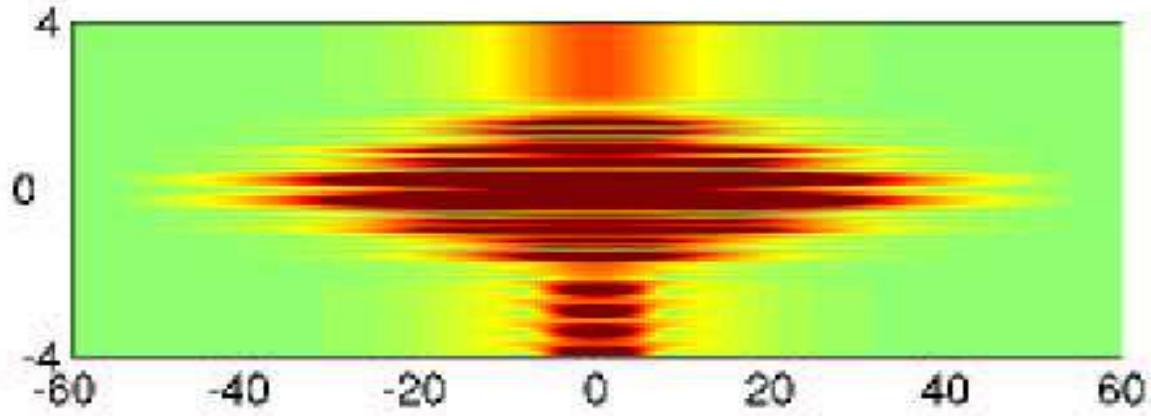


FIG. 1.25 – Diamètre de  $173.6\mu\text{m}$ , fréquence relative de 1 (la longueur d’onde est  $1.55\mu\text{m}$ ), waist  $w=9.92\mu\text{m}$ . Les unités sont comptées en  $\lambda_0 = 1.55\mu\text{m}$  : les longueurs physiques sont obtenues en multipliant par  $1.55\mu\text{m}$ .

laquelle le coefficient de transmission est maximal, et le coefficient de réflexion proche de zéro. On trace la solution obtenue à cette fréquence de résonance sur la figure 1.28.

On constate que le faisceau reste confiné dans le dispositif, alors que dans le cas de couches planes, le faisceau s’élargissait dans tout le dispositif. Le faisceau en sortie est donc plus étroit que dans le cas de couches planes. Les bords n’interviennent quasiment pas.

On représente sur la figure 1.29 le module du champ total suivant la direction d’atténuation du faisceau (direction perpendiculaire à la direction de propagation). On n’a pas d’oscillations, comme dans le cas de couches planes. Lorsqu’on effectue un balayage en fréquence, la gaussienne

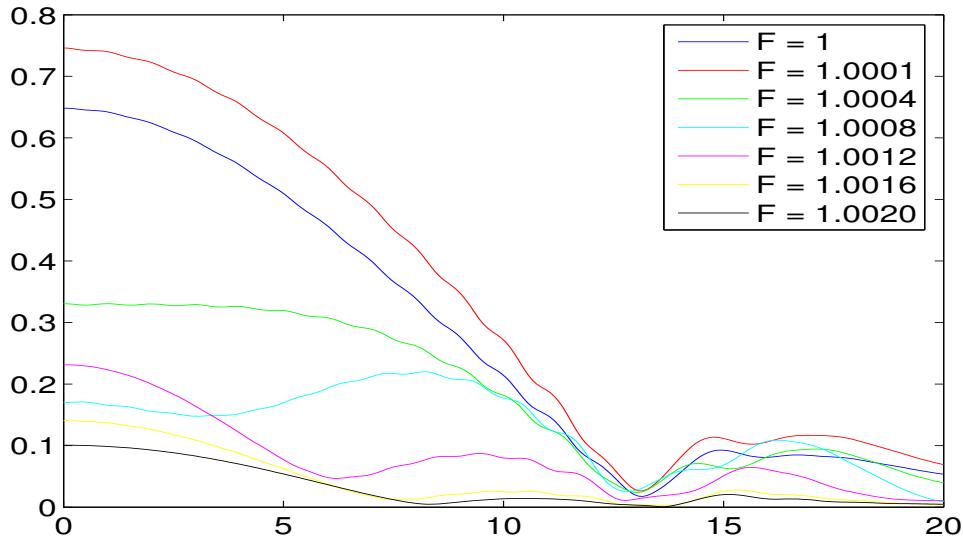


FIG. 1.26 – Module du champ transmis  $|u(x, z = Z)|$ ,  $x > 0$  pour un dispositif de diamètre  $D=40.2\mu\text{m}$ , et un waist  $w = 9.92 \mu\text{m}$ . Le  $Z$  correspond à 4 longueurs d’onde en aval du filtre. Axe horizontal : axe parallèle au barreau  $x > 0$ . Les différentes courbes correspondent à un balayage en fréquence ou encore un décalage en longueur d’onde variant de  $-3.094\text{ nm}$  à 0.

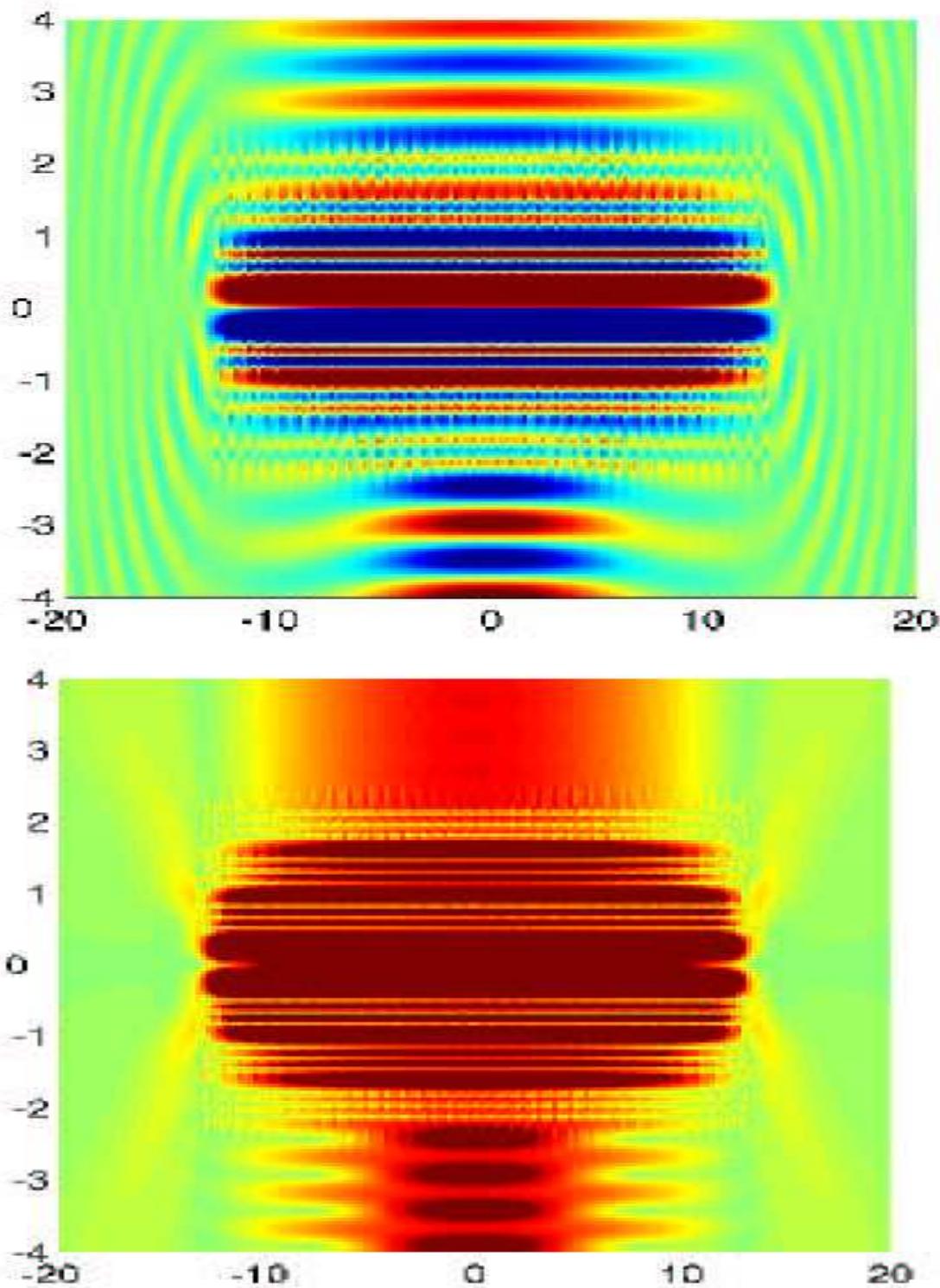


FIG. 1.27 – Solution pour un diamètre  $D=40.2\mu\text{m}$ , une fréquence relative  $F$  de 1.0001 (décalage de la longueur d'onde de  $-0.1548452\text{ nm}$ ), waist  $w = 9.92\mu\text{m}$ . En haut, partie réelle du champ total, en bas, le module. Les unités sont comptées en  $\lambda_0 = 1.55\mu\text{m}$  : les longueurs physiques sont obtenues en multipliant par  $1.55\mu\text{m}$ .

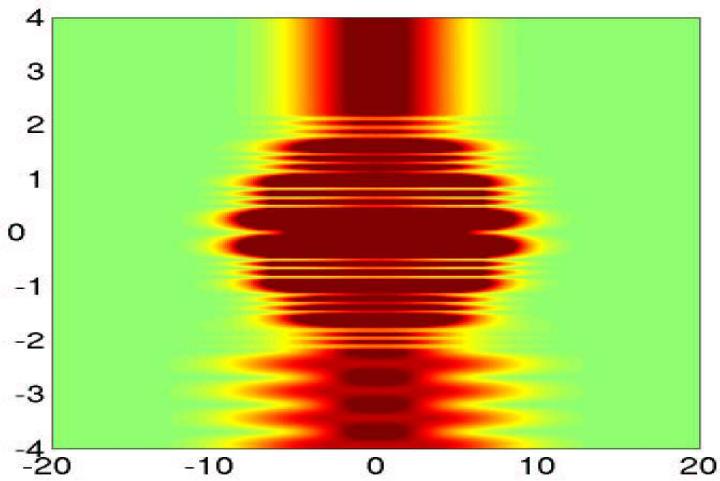


FIG. 1.28 – Solution pour un diamètre  $D=40.2\mu\text{m}$ , une fréquence relative  $F$  de 1.000732 (longueur d'onde de  $1.548866\mu\text{ m}$ ), waist  $w = 9.92\mu\text{m}$ . La flèche est de  $20\text{ nm}$  au centre de la cavité. Cas des bords courbes.

en entrée reste à peu près une gaussienne en sortie, les “rebonds” sont moins marqués que dans le cas de couches planes.

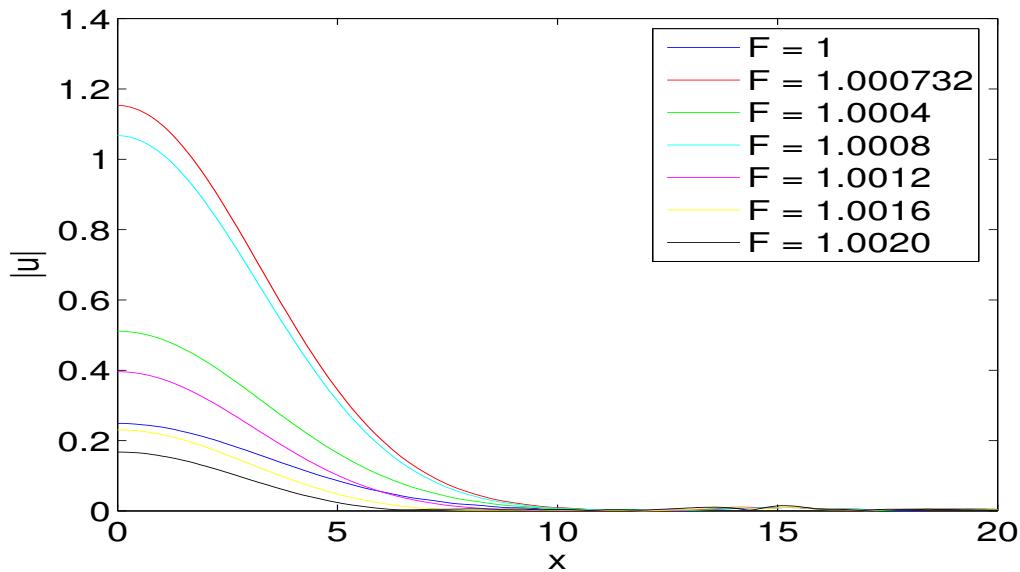


FIG. 1.29 – Module du champ transmis  $D=40.2\mu\text{m}$ ,  $F$  de 1.0 à 1.0020  $w=9.92\mu\text{m}$ . Cas des bords courbes.

### 1.4.3 Apologie de l'ordre élevé

Nous nous concentrons uniquement sur le dernier cas (bords courbes avec une diamètre de  $40.2\mu\text{m}$ , une fréquence relative  $F$  de 1.000732 et un waist  $w = 9.92\mu\text{m}$ ). Nous commençons par faire la simulation numérique avec du  $Q_2$ , 10 points par longueur d'onde (dans le diélectrique et dans l'air). On obtient les résultats des figures 1.30 et 1.31

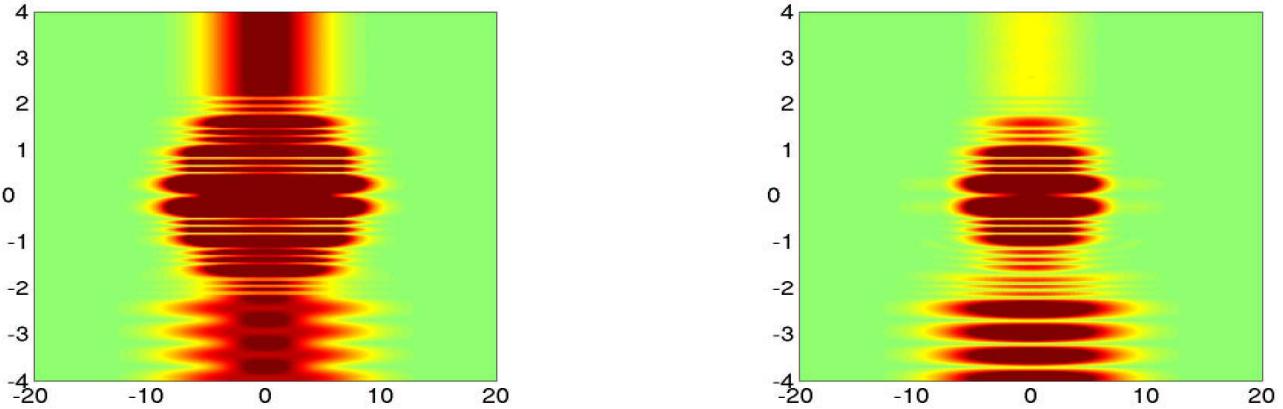


FIG. 1.30 – Solution pour un diamètre  $D=40.2\mu\text{m}$ , une fréquence relative  $F$  de 1.000732 (longueur d'onde de  $1.548866\mu\text{ m}$ ), waist  $w = 9.92\mu\text{m}$ . A gauche, solution de référence ; à droite solution pour  $Q_2$  avec 10 points par longueur d'onde

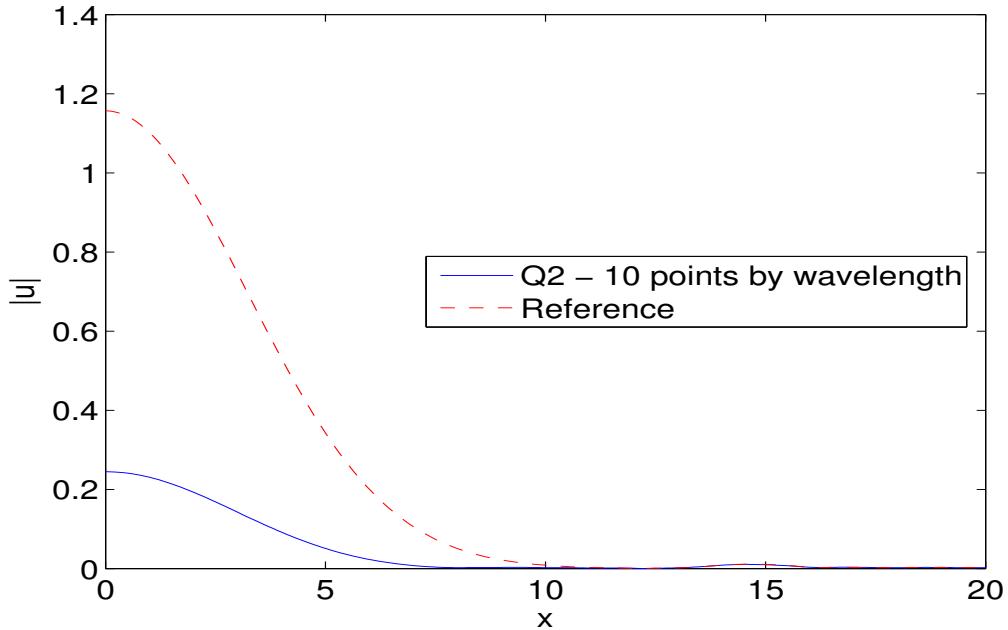


FIG. 1.31 – Faisceau gaussien à la sortie du dispositif, pour la solution de référence et la solution numérique  $Q_2$  avec 10 points par longueur d'onde

Oh malheur, la solution  $Q_2$  part dans le décor ! Si on mesure l'erreur à la sortie du dispositif, on trouve 100 % d'erreur. Ce phénomène peut s'expliquer de manière simple. La dispersion des

éléments  $Q_2$  déplace légèrement la fréquence de résonance, de ce fait on tombe à côté de la fréquence de résonance, et le filtre ne laisse passer qu'un signal faible. De façon plus précise, on peut tracer le coefficient de transmission en fonction de la fréquence pour la solution de référence et pour la solution  $Q_2$  avec 10 points par longueur d'onde. On voit nettement le décalage de la fréquence de résonance sur la figure 1.32, la fréquence de résonance numérique du maillage  $Q_2$  est égale à  $F = 0.99989$  au lieu de  $f = 1.00073$  pour la solution de référence.

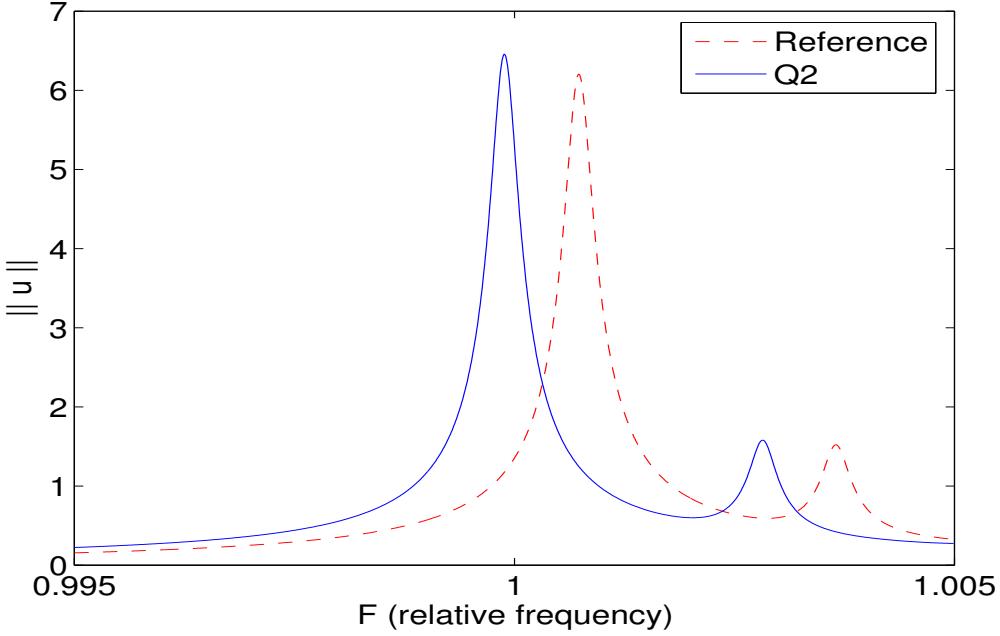


FIG. 1.32 – Coefficient de transmission en fonction de la fréquence, pour la solution de référence et la solution numérique  $Q_2$  avec 10 points par longueur d'onde

Pour la fréquence de résonance, on se fixe un objectif d'erreur à 10 %, on affiche dans le tableau 1.1 le nombre de ddl nécessaire pour atteindre cette erreur, ainsi que la taille de la matrice LU (dans le cas d'une résolution directe). Il apparaît évident que sur ce cas, il est vain

Ordre	Nombre ddl	Stockage LU
2	453000	598Mo
3	69800	94Mo
4	52000	78Mo
5	33200	58Mo
6	47700	93Mo
7	42200	96Mo

TAB. 1.1 – Nombre de ddl nécessaires pour atteindre moins de 10 % d'erreur

et inutile de tenter de faire du  $Q_1$  ou du  $Q_2$ . L'ordre le plus approprié semble être le  $Q_5$ , pour des questions de respect de la géométrie. Si on utilise un ordre supérieur, on se retrouve avec des mailles petites pour respecter la géométrie (typiquement une maille dans les interstices d'air qui font un quart de longueur d'onde ...). On a donc par conséquent plus de degrés de liberté. Le maillage utilisé pour  $Q_5$  est sur la figure 1.23 (hors PML). On ne met que deux mailles dans la cavité et deux mailles sur chaque diélectrique, suivant l'axe de propagation.

## 1.5 Conclusion

Dans ce chapitre, nous avons décrit la méthode des éléments finis nodaux sur le cas particulier de l'équation de Helmholtz. Nous choisissons les points de Gauss-Lobatto comme points d'interpolation et points d'intégration, ce qui permet de réaliser la condensation de masse. De plus, nous avons montré que ce choix menait à un calcul rapide de la matrice de rigidité. Nous avons également introduit la technique utilisée pour prendre en compte de manière fine la géométrie, technique dite des "éléments courbes isoparamétriques".

Les résultats numériques tendent à prouver qu'en 2-D, les approximations faites ne détériorent pas l'ordre de convergence. Sur des géométries lisses, on semble obtenir une méthode qui converge en  $O(h^{r+1})$  en norme  $L^2$  et en  $O(h^r)$  en norme  $H^1$ . Sur des géométries singulières, l'ordre de convergence est le même quel que soit l'ordre d'approximation. Néanmoins, La montée en ordre permet d'obtenir une solution plus précise pour un même nombre de degrés de liberté.

Nous conclurons par la robustesse des méthodes d'ordre élevé, qui sur des cas difficiles comme le cas Atmel, donnent une solution précise, alors que les méthodes d'ordre un ou deux nécessitent d'utiliser un nombre de degrés de liberté très important. Les méthodes d'ordre élevé sont flexibles, car on peut utiliser un ordre différent si on veut changer la fréquence (plutôt que de refaire le maillage!). On peut aussi calculer la solution pour  $Q_4$  et  $Q_5$  par exemple, et on aura une estimation de l'erreur commise.

## Chapitre 2

# Algorithmes itératifs de résolution

*L'objectif de ce chapitre est de développer des méthodes itératives efficaces pour la résolution du système linéaire issu de la discrétisation éléments finis de l'équation de Helmholtz . La première section apporte le premier élément de réponse, on y propose un algorithme rapide pour effectuer le produit matrice vecteur. Cet algorithme s'accompagne d'un gain de stockage appréciable lorsqu'on monte en ordre. La deuxième section montre qu'un solveur direct est très largement suffisant pour le cas 2-D et qu'il est nécessaire de se tourner vers un solveur itératif en 3-D. La troisième section expose les performances des algorithmes itératifs basés sur les espaces de Krylov. On fera le choix d'algorithmes optimaux, le BICGCR ou le COCG. La troisième section aborde le problème du préconditionnement, on compare divers préconditionneurs : la factorisation incomplète, le multigrille et la décomposition en sous-domaines. Les deux premiers sont tous deux efficaces si on ajoute de l'amortissement.*

## Sommaire

---

<b>2.1</b>	<b>Formulation mixte</b>	<b>50</b>
2.1.1	Formulation variationnelle, propriétés des matrices	50
2.1.2	Intérêt de la factorisation	52
<b>2.2</b>	<b>Résolution directe</b>	<b>55</b>
<b>2.3</b>	<b>Résolution itérative</b>	<b>58</b>
2.3.1	Cas 2-D	58
2.3.2	Cas 3-D	61
<b>2.4</b>	<b>Préconditionnement</b>	<b>63</b>
2.4.1	Préconditionnement par l'équation de Helmholtz avec amortissement	64
2.4.2	Décomposition en sous-domaines	70
2.4.3	Solveur itératif?	73
<b>2.5</b>	<b>Conclusion</b>	<b>73</b>

---

Afin d'approfondir ce chapitre, on pourra consulter

- FAUQUEUX, S. (2003). *Eléments finis mixtes spectraux et couches absorbantes parfaitement adaptées pour la propagation d'ondes élastiques en régime transitoire*. Thèse de doctorat, Université de Paris IX Dauphine,
- HACKBUSCH, W. (1994). *Iterative solution of large sparse systems of equations*. Springer Verlag,
- HACKBUSCH, W. (1985). *Multigrid methods and applications*. Springer-Verlag,
- SAAD, Y. (1996). *Iterative methods for sparse linear systems*. Series in Computer Science

## 2.1 Formulation mixte

L'utilisation d'une formulation mixte pour l'équation des ondes a été originalement introduite par [Cohen et Fauqueux, 2000]. La transposition au régime harmonique est immédiate, nous la détaillons ici pour exhiber les bonnes propriétés d'une telle approche.

### 2.1.1 Formulation variationnelle, propriétés des matrices

L'idée de base est d'introduire une inconnue intermédiaire  $\mathbf{v}$  afin de ne conserver que des opérateurs différentiels d'ordre 1. On considère le système :

$$\begin{aligned} -\omega^2 \rho u - \operatorname{div}(\mathbf{v}) &= f \\ \frac{1}{\mu} \mathbf{v} - \nabla u &= 0 \end{aligned}$$

On établit la formulation variationnelle en faisant l'intégration par parties sur la première équation :

$$\begin{aligned} -\omega^2 \int_{\Omega} \rho u \varphi + \int_{\Omega} \mathbf{v} \cdot \nabla \varphi &= \int_{\Omega} f \varphi \\ \int_{\Omega} \frac{1}{\mu} \mathbf{v} \cdot \psi - \int_{\Omega} \nabla u \cdot \psi &= 0 \end{aligned}$$

$u$  est choisi dans le même espace d'approximation que dans la formulation standard :

$$U_h = \{v \in H^1(\Omega) \mid v|_{K_i} \circ F_i \in Q_r\}$$

$\mathbf{v}$  est choisi dans l'espace d'approximation suivant :

$$V_h = \{\mathbf{v} \in (L^2(\Omega))^2 \mid J_i D F_i^{-1} \mathbf{v} \circ F_i \in (Q_r)^2\}$$

Par conséquent, les fonctions de base vérifient :

$$\begin{aligned} \varphi_i \circ F_i(\hat{x}) &= \hat{\varphi}_i(\hat{x}) \\ \psi_i \circ F_i(\hat{x}) &= \frac{1}{J_i} D F_i \hat{\psi}_i(\hat{x}) \end{aligned}$$

Les fonctions de base vectorielles sont construites à partir des fonctions de base scalaire (associées aux points de Gauss-Lobatto comme dans la formulation standard) :

$$\text{En 2-D} \quad \hat{\psi}_i = \begin{vmatrix} \hat{\varphi}_i \\ 0 \end{vmatrix} \quad \text{ou} \quad \begin{vmatrix} 0 \\ \hat{\varphi}_i \end{vmatrix}$$

On aboutit au système linéaire suivant :

$$\begin{aligned} -\omega^2 M_h U_h + R_h V_h &= F_h \\ B_h V_h - R_h^t U_h &= 0 \end{aligned}$$

La matrice  $M_h$  est identique à celle du chapitre 1, elle est diagonale, lorsqu'on utilise les points de Gauss-Lobatto. Les matrices élémentaires de  $B_h$  et  $R_h$  sont égales à :

$$\begin{aligned} (B_h)_{k,l} &= \int_{K_i} \psi_k \cdot \psi_l \\ &= \int_{\hat{K}} \frac{1}{\mu J_i} D F_i^t D F_i \hat{\psi}_k \cdot \hat{\psi}_l \end{aligned}$$

$$\begin{aligned}
(R_h)_{k,l} &= \int_{K_i} \psi_l \cdot \nabla \varphi_k \\
&= \int_{\hat{K}} J_i \frac{1}{J_i} DF_i \hat{\psi}_l \cdot DF_i^{-t} \hat{\nabla} \hat{\varphi}_k \\
&= \int_{\hat{K}} \hat{\psi}_l \cdot \hat{\nabla} \hat{\varphi}_k
\end{aligned}$$

Le choix de l'espace d'approximation pour  $\mathbf{v}$  n'est pas anodin, il a été choisi de telle sorte que la matrice élémentaire de  $R_h$  ne dépende pas de la géométrie. La matrice élémentaire de  $R_h$  est identique quel que soit l'élément :  $R_h = \hat{R}$ . Le coût de stockage de la matrice globale  $R_h$  est donc quasi-nul, car on ne stocke que la matrice élémentaire  $\hat{R}$ . De plus cette matrice élémentaire  $\hat{R}$  est creuse, du fait de la tensorisation des fonctions de base. En effet, choisissons :

$$\hat{\varphi}_k(\hat{x}, \hat{y}) = \hat{\varphi}_{k_1}(\hat{x})\hat{\varphi}_{k_2}(\hat{y}) \text{ et } \hat{\psi}_l = \hat{\varphi}_{l_1}(\hat{x})\hat{\varphi}_{l_2}(\hat{y}) \mathbf{e}_1$$

En intégrant à l'aide des points de Gauss-Lobatto :

$$\begin{aligned}
(R_h)_{k,l} &= \sum_{m,n=1}^{r+1} \omega_{m,n} \hat{\varphi}'_{k_1}(\hat{\xi}_m) \hat{\varphi}_{k_2}(\hat{\xi}_n) \hat{\varphi}_{l_1}(\hat{\xi}_m) \hat{\varphi}_{l_2}(\hat{\xi}_n) \\
&= \omega_{l_1,l_2} \hat{\varphi}'_{k_1}(\hat{\xi}_{l_1}) \delta_{k_2,l_2}
\end{aligned}$$

De manière analogue, pour le degré de liberté vectoriel orienté suivant  $\mathbf{e}_2$ , on obtient :

$$(R_h)_{k,l} = \omega_{l_1,l_2} \hat{\varphi}'_{k_2}(\hat{\xi}_{l_2}) \delta_{k_1,l_1}$$

La présence de  $\delta_{k_1,l_1}$  démontre que la matrice élémentaire est creuse. Sur la figure 2.1, on peut voir les degrés de liberté vectoriels qui donnent une interaction non-nulle avec le degré de liberté scalaire symbolisé par le point bleu marine. En 2-D, chaque ligne de la matrice  $\hat{R}$  contient

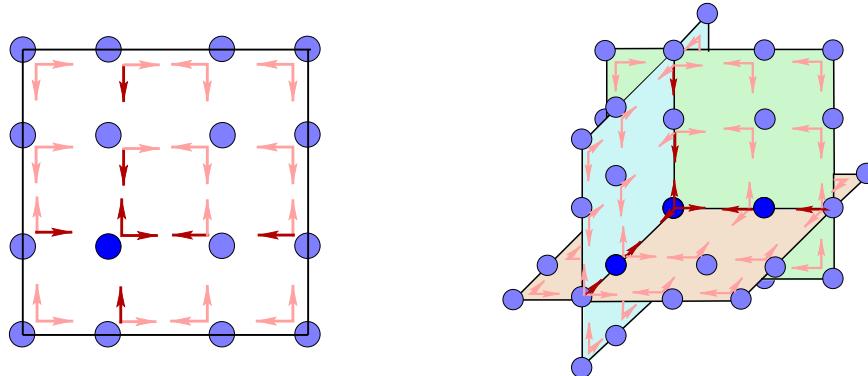


FIG. 2.1 – Interactions entre degrés de liberté scalaires et degrés de liberté vectoriels.

$2(r+1)$  valeurs non-nulles, alors que le nombre de colonnes est de  $2(r+1)^2$ . En 3-D, chaque ligne de la matrice  $\hat{R}$  contient  $3(r+1)$  valeurs non-nulles, alors que le nombre de colonnes est de  $3(r+1)^3$ . Le caractère creux de la matrice est diablement plus important en 3-D !

Intéressons-nous finalement aux propriétés de la matrice  $B_h$ . Prenons :

$$\psi_k = \hat{\varphi}_k \mathbf{e}_s \quad s = 1 \text{ ou } 2 \quad \psi_l = \hat{\varphi}_l \mathbf{e}_t \quad t = 1 \text{ ou } 2$$

En utilisant les points de Gauss-Lobatto pour intégrer  $B_h$ , on obtient alors :

$$(B_h)_{k,l} = \frac{\omega_k}{\mu J_i} (DF_i^t DF_i)_{s,t} \delta_{k,l}$$

La matrice élémentaire de  $B_h$  est par conséquent une matrice diagonale par blocs 2x2, chaque bloc s'appliquant sur les deux inconnues vectorielles associées au même point de Gauss-Lobatto. L'espace d'approximation pour  $\mathbf{v}$  étant discontinu, la matrice global  $B_h$  est également diagonale par blocs 2x2 (blocs 3x3 en 3-D).

$M_h$ , on le rappelle au bon souvenir du lecteur, est diagonale. On a par conséquent un coût de stockage extrêmement faible : une matrice diagonale et une matrice diagonale par blocs 2x2. De plus, on peut éliminer l'inconnue vectorielle en résolvant directement le système :

$$(-\omega^2 D_h + R_h B_h^{-1} R_h^t) U_h = F_h$$

En pratique, on calcule directement l'inverse de  $B_h$  :

$$(B_h^{-1})_{k,l} = \frac{\mu J_i}{\omega_k} (DF_i^{-1} DF_i^{-t})_{s,t} \delta_{k,l}$$

Et on utilise l'algorithme de produit matrice-vecteur suivant (cas 3-D) :

Produit matrice-vecteur  $Y = R_h B_h^{-1} R_h^t U$

$Y = 0$

Pour  $e = 1$ , nombre d'hexaèdres du maillage

Pour  $i = 1, (r+1)^3$   
 $U_{local}(i) = U_{global}(lg(e,i))$

Fin Pour

// produit matrice-vecteur creux standard

$V_{local} = R_h * U_{local}$

// produit par la matrice diagonale par blocs  $B_h^{-1}$

Pour  $i = 1, (r+1)^3$   
 Pour  $j = 1, 3$   
 $V(j) = V_{local}(3*(i-1) + j)$

$B_h^{-1} V = B_h^{-1}(e,i) * V$

Pour  $j = 1, 3$   
 $V_{local}(3*(i-1) + j) = B_h^{-1} V(j)$

Fin Pour

// produit matrice-vecteur creux standard

$U_{local} = R_h^t * V_{local}$

Pour  $i = 1, (r+1)^3$

$Y(lg(e,i)) += U_{local}(i)$

Fin Pour

Fin boucle éléments

$lg(e,i)$  est le numéro global du degré de liberté  $i$  de l'élément  $e$ .

## 2.1.2 Intérêt de la factorisation

Par complément de Schur sur le système (2.1.1), on peut éliminer l'inconnue vectorielle :

$$(-\omega^2 M_h + R_h B_h^{-1} R_h^t) U_h = F_h$$

Or, S. Fauqueux a démontré que la matrice de rigidité  $K_h$  de la formulation standard était égale à :

$$K_h = R_h B_h^{-1} R_h^t$$

Cette égalité est vraie si on choisit les points de Gauss-Lobatto pour intégrer les matrices. La formulation mixte est donc strictement équivalente à la formulation standard, son seul et unique avantage est de fournir une factorisation de la matrice de rigidité. Cette factorisation a l'intérêt majeur de fournir un produit matrice-vecteur peu coûteux en stockage (matrice diagonale et diagonale par blocs à stocker) et en temps de calcul (matrice de rigidité élémentaire creuse). On peut effectuer des calculs de complexité pour illustrer ces deux propriétés.

## Cas 2-D

On note  $r$  l'ordre d'approximation,  $N_e$  le nombre d'éléments. Sur un maillage régulier, on a globalement  $(r^2 N_e + O(1))$  degrés de liberté pour  $u$ . L'inconnue vectorielle  $\mathbf{v}$  étant discontinue, on a  $2(r+1)^2 N_e$  degrés de liberté pour  $\mathbf{v}$ .

Pour la formulation mixte, la matrice diagonale coûte  $r^2 N_e$  en stockage. La matrice diagonale par blocs  $2 \times 2 B_h$  est symétrique, son stockage est de  $3(r+1)^2 N_e$ , le tableau de correspondance numérotation locale/numérotation globale  $lg(e,i)$  coûte  $0.5(r+1)^2$ .

Pour la formulation standard, on doit stocker la matrice globale  $A_h$ . Comptabilisons le nombre d'éléments non-nuls de cette matrice. On a  $(2r+1)^2$  coefficients non-nuls pour chaque ddl associé à un sommet du maillage,  $(r+1)(2r+1)$  coefficients non-nuls pour chaque ddl associé à une arête du maillage, et  $(r+1)^2$  coefficients non-nuls pour chaque ddl interne. En tout, on dénombre  $((2r+1)^2 + 2(r+1)(2r+1)(r-1) + (r+1)^2(r-1)^2) N_e$  coefficients non-nuls dans la matrice. On utilise la propriété de symétrie de la matrice, en divisant par 2 ce nombre et en rajoutant la moitié du coût d'une matrice diagonale.

Stockage utilisé pour la formulation mixte 2-D :  $(r^2 + 3.5(r+1)^2) N_e$

Stockage utilisé pour la formulation standard 2-D :  $(0.5r^4 + 2r^3 + 2.5r^2) N_e$

Pour le coût de calcul, on considère que l'addition, la soustraction ou la multiplication ont un coût de 1. Pour la formulation standard, on aura une multiplication et une addition, pour chaque entrée de la matrice (la symétrie n'intervient pas). Pour la formulation mixte, on a  $2(r+1)^3$  termes non-nuls dans la matrice  $\hat{R}$ . Comme la factorisation fait intervenir deux multiplications avec  $R_h$  et  $R_h^t$ , on aura un coût de  $8(r+1)^3 N_e$ . A ce coût, il faut rajouter le coût de la multiplication par  $M_h$  et  $B_h^{-1}$ .

Opérations pour la formulation mixte 2-D :  $(2r^2 + 6(r+1)^2 + 8(r+1)^3) N_e$

Opérations pour la formulation standard 2-D :  $(2r^4 + 8r^3 + 8r^2) N_e$

Asymptotiquement pour  $r$  grand, on trouve un coût en temps de calcul en  $O(r^4)$  pour la formulation standard contre un coût en  $O(r^3)$  pour la formulation mixte. On divise ces coûts par le nombre de degrés de liberté ( $r^2 N_e$ ), afin de comparer à nombre de degrés de liberté constant. On trouve les résultats de la figure 2.2. Sur cette figure, on voit que la formulation mixte devient plus rapide dès  $Q_4$ , elle est moins onéreuse en stockage dès  $Q_2$ . Le surcoût de calcul lorsqu'on monte en ordre est relativement faible, il est loin d'être rédhibitoire. En outre, si on utilise un ordre élevé, on a besoin de moins de degrés de liberté pour obtenir la même précision sur la solution. Des comparaisons plus "justes" seront effectuées par la suite.

## Cas 3-D

On a  $r^3 N_e$  degrés de liberté pour  $u$  et  $3(r+1)^3 N_e$  ddl pour  $\mathbf{v}$ . Pour la formulation mixte, on stocke  $r^3 N_e$  coefficients pour  $M_h$  et  $6.5(r+1)^3 N_e$  coefficients pour  $B_h$ .

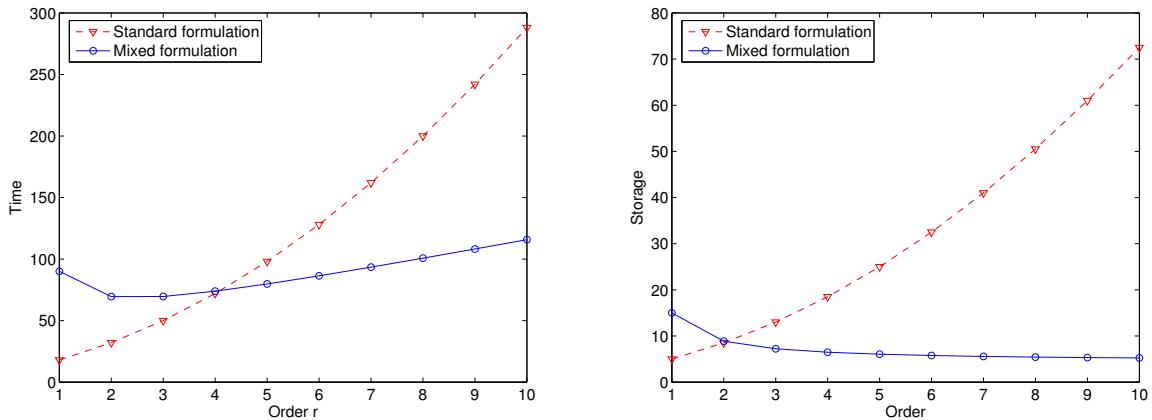


FIG. 2.2 – A gauche temps de calcul en fonction de l’ordre d’approximation, à droite stockage

Pour la formulation standard, on a  $12r^2 + 6r + 1$  termes non-nuls pour chaque ddl associé à un sommet du maillage,  $4r(r+1) + (r+1) + 4r^2$  termes pour chaque ddl associé à une arête,  $(r+1)^2 + 2r(r+1) + 2r^2$  termes pour chaque ddl associé à une face et  $3(r+1)^2 - 3r - 2$  termes pour chaque ddl interne.

Stockage utilisé pour la formulation mixte 3-D :  $(r^3 + 6.5(r+1)^3)N_e$

Stockage utilisé pour la formulation standard 3-D :  $(1.5r^5 + 4.5r^4 + 4r^3)N_e$

Opérations pour la formulation mixte 3-D :  $(2r^3 + 15(r+1)^3 + 12(r+1)^4)N_e$

Opérations pour la formulation standard 3-D :  $(6r^5 + 18r^4 + 14r^3)N_e$

Asymptotiquement pour  $r$  grand, on trouve un coût en temps de calcul en  $O(r^5)$  pour la formulation standard contre un coût en  $O(r^4)$  pour la formulation mixte. On divise ces coûts par le nombre de degrés de liberté ( $r^3 N_e$ ), afin de comparer à nombre de degrés de liberté constant. On trouve les résultats de la figure 2.3 Sur cette figure, on voit que la formulation

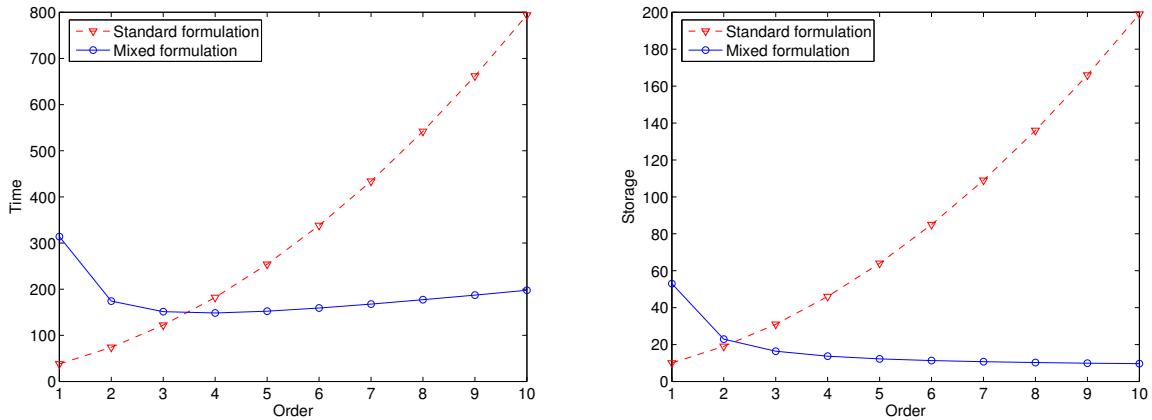


FIG. 2.3 – A gauche temps de calcul, à droite stockage

mixte devient plus rapide dès  $Q_4$ , elle est moins onéreuse en stockage dès  $Q_3$ . Pour la formulation standard, on a utilisé la propriété fondamentale que la matrice élémentaire est creuse. En effet, lorsqu’on utilise les points de Gauss-Lobatto, seuls les degrés de liberté placés localement dans

le même plan (parallèle à Oxy, Oxz ou Oyz) ont une interaction non-nulle. Cette propriété a été mise en évidence dans le chapitre 1. Cela donne lieu à un coût asymptotique en  $O(r^5)$  au lieu du coût “classique” en  $O(r^6)$  si on avait utilisé des points d’intégration de Gauss. Dans le chapitre 3, on donnera des points de comparaison avec les tétraèdres.

On notera que la formulation mixte fait apparaître un ordre “optimal”, qui donne un temps de calcul minimal pour un nombre de degrés de liberté constant. En 2-D, l’ordre optimal théorique est  $Q_2$ , en 3-D c’est  $Q_4$ . En conclusion, on utilisera la factorisation  $K_h = R_h B_h^{-1} R_h^t$  pour calculer le produit matrice-vecteur si  $\mathbf{r} \geq \mathbf{3}$ . Pour des ordres inférieurs, on assemblera la matrice classiquement comme explicité dans le chapitre 1.

Au niveau du stockage, la formulation mixte permet de stocker à peu près 8 vecteurs en lieu et place de la matrice. Si les coefficients physiques  $\rho$  et  $\mu$  sont réels, on ne stocke que 4 vecteurs. En vérité, ce sont les vecteurs nécessaires à l’algorithme itératif de résolution, qui coûtent le plus cher en stockage !

## Intégration exacte

Le point clé pour obtenir cette factorisation est l’utilisation de formules de quadrature approchées (ie formules de Gauss-Lobatto). Peut-on obtenir une factorisation similaire si on utilise une formule de quadrature plus précise (formules de Gauss) ? La réponse est oui, plus précisément on obtient la factorisation suivante :

$$K_h = C_h R_h B_h^{-1} R_h^t C_h^t$$

avec les notations :

$$(C_h)_{j,k} = \hat{\varphi}_j(\hat{\xi}_k^G)$$

$$(R_h)_{j,k} = \hat{\nabla} \hat{\varphi}_j(\hat{\xi}_k^G)$$

La matrice  $B_h$  est similaire à la matrice  $B_h$  des sections précédentes sauf qu’elle est exprimée aux points de Gauss. Lorsqu’on utilise les points de Gauss, on perd la condensation de masse, la matrice de masse devient alors factorisable en :

$$M_h = C_h D_h C_h^t$$

avec la matrice diagonale :

$$(D_h)_{j,j} = \rho \omega_j J_e(\hat{\xi}_j)$$

Une démonstration de ces factorisations est présentée en annexe C.

Néanmoins, l’utilisation de formules de quadrature de Gauss, fournit un algorithme de calcul deux fois plus lent, il reste donc plus intéressant d’utiliser les formules approchées de Gauss-Lobatto.

## 2.2 Résolution directe

Avant d’aborder la résolution itérative, il n’est pas inutile de rappeler brièvement les fondamentaux d’une résolution directe. C’est en exhibant les avantages et les inconvénients d’une méthode directe, qu’on pourra conclure sur la nécessité ou non d’une résolution itérative. On s’intéresse au système linéaire :

$$(-\omega^2 M_h + R_h B_h^{-1} R_h^t) U_h = F_h$$

Une résolution directe consistera à calculer la matrice complexe symétrique

$$A_h = -\omega^2 M_h + R_h B_h^{-1} R_h^t$$

et à la factoriser sous la forme :

$$A_h = L D L^t$$

Cette factorisation est connue sous le nom de factorisation de Crout. On doit expliciter la matrice, ce qui revient à calculer la matrice éléments finis de la formulation standard ! On est alors confronté au problème d'avoir une matrice très volumineuse quand on monte en ordre. La factorisation sera encore plus volumineuse, rendant prohibitive l'utilisation de méthodes directes pour des cas de grande taille. Pour illustrer notre propos, nous donnons les tailles des matrices éléments finis, et la place mémoire prise par la factorisation  $L D L^t$  dans les tableaux 2.1 et 2.2. Les cas tests sont respectivement la diffraction par un disque et par une sphère. On impose une condition de Neumann sur l'objet, et une condition absorbante sur la frontière extérieure du domaine de calcul. Afin de comparer ce qui est comparable, on se place à nombre de degrés de libertés constant.

Nombre ddl	11 000				
Ordre d'approximation	1	2	3	5	7
Taille matrice $A_h$	1 Mo	1 Mo	2 Mo	5 Mo	8 Mo
Taille factorisation	11 Mo	11 Mo	12 Mo	17 Mo	25 Mo
Avec condensation	11 Mo	10 Mo	9 Mo	7 Mo	6 Mo
Nombre ddl	100 000				
Ordre d'approximation	1	2	3	5	7
Taille matrice $A_h$	9 Mo	15 Mo	24 Mo	46 Mo	77 Mo
Taille factorisation	117 Mo	126 Mo	134 Mo	171 Mo	244 Mo
Avec condensation	117 Mo	109 Mo	100 Mo	88 Mo	83 Mo
Nombre ddl	1 000 000				
Ordre d'approximation	1	2	3	5	7
Taille matrice $A_h$	96 Mo	163 Mo	252 Mo	481 Mo	773 Mo
Taille factorisation	1 472 Mo	1 617 Mo	1 673 Mo	2 045 Mo	2 700 Mo
Avec condensation	1 472 Mo	1 440 Mo	1 349 Mo	1 187 Mo	1 088 Mo

TAB. 2.1 – Taille nécessaire en fonction de l'ordre d'approximation, cas 2-D

On utilise MUMPS comme solveur direct [Amestoy *et al.*, 2003], qui est basé sur une technique multifrontale [Duff et Reid, 1983] et très performant [Amestoy *et al.*, 2000]. Dans le tableau 2.1, on trouve une complexité en  $O(N)$  pour la matrice  $A_h$  et une complexité quasi-linéaire en  $O(N^{1.1})$  pour la factorisation ( $N$  étant le nombre de degrés de liberté). En 3-D, les complexités sont respectivement  $O(N)$  et  $O(N^2)$ , ce qui traduit bien que les matrices éléments finis 3-D requièrent rapidement une mémoire très grande. En 2-D, avec 2 Go de mémoire, on peut passer des cas à un million de degrés de liberté. Il nous apparaît, en 2-D, qu'un solveur itératif a peu de chances d'être compétitif face à un solveur direct, sauf pour de très gros cas. En revanche, le tableau 2.2 montre qu'un solveur direct est particulièrement inefficace en 3-D. En effet, avec 2 Go de mémoire, on est limité à des cas de moins de 150 000 ddl. Or, des cas réalistes 3-D comprennent plusieurs millions de ddl. Il est nécessaire de se tourner vers un solveur itératif !

Une technique intéressante est la condensation statique, cela consiste à éliminer les degrés de liberté intérieurs par un complément de Schur. On décompose la matrice élémentaire sous la

Nombre ddl	110 000				
Ordre d'approximation	1	2	3	5	7
Taille matrice $A_h$	19 Mo	36 Mo	59 Mo	138 Mo	219 Mo
Taille factorisation	1 485 Mo	2 094 Mo	1 505 Mo	2 018 Mo	2 234 Mo
Avec condensation	1 485 Mo	1 361 Mo	1 476 Mo	1 741 Mo	1 605 Mo
Nombre ddl	400 000				
Ordre d'approximation	1	2	3	5	7
Taille matrice $A_h$	186 Mo	343 Mo	559 Mo	464 Mo	836 Mo
Taille factorisation	37 Go	32 Go	42 Go	-	-
Avec condensation	37 Go	33 Go	28 Go	-	-

TAB. 2.2 – Taille nécessaire en fonction de l'ordre d'approximation, cas 3-D

forme :

$$A = \begin{pmatrix} A_{\text{bord,bord}} & A_{\text{bord,int}} \\ A_{\text{int,bord}} & A_{\text{int,int}} \end{pmatrix}$$

où l'indice “bord” se rapporte aux degrés de liberté sur la frontière de l'élément (en 2-D, ddl associés aux sommets et aux arêtes). L'indice “int” se rapporte aux degrés de liberté internes de l'élément (cf. figure 2.4).

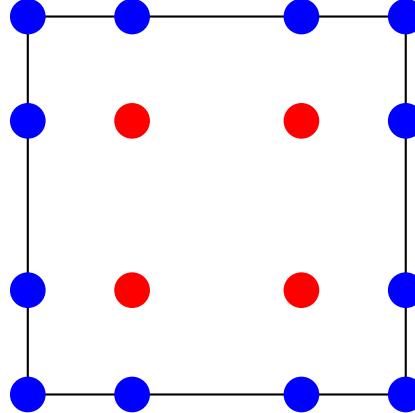


FIG. 2.4 – En bleu, ddl du bord, en rouge ddl internes de l'élément

On a le système :

$$A_{\text{bord,bord}} U_{\text{bord}} + A_{\text{bord,int}} U_{\text{int}} = F_{\text{bord}}$$

$$A_{\text{int,bord}} U_{\text{bord}} + A_{\text{int,int}} U_{\text{int}} = F_{\text{int}}$$

soit en éliminant  $U_{\text{int}}$  :

$$\begin{aligned} U_{\text{int}} &= (A_{\text{int,int}})^{-1} (F_{\text{int}} - A_{\text{int,bord}} U_{\text{bord}}) \\ (A_{\text{bord,bord}} - A_{\text{bord,int}} (A_{\text{int,int}})^{-1} A_{\text{int,bord}}) U_{\text{bord}} &= F_{\text{bord}} - A_{\text{bord,int}} (A_{\text{int,int}})^{-1} F_{\text{int}} \end{aligned}$$

En pratique, on remplace la matrice élémentaire  $A$ , par le complément de schur :

$$A_{\text{bord,bord}} - A_{\text{bord,int}} (A_{\text{int,int}})^{-1} A_{\text{int,bord}}$$

L'inverse  $A_{\text{int,int}}$  est calculé élément par élément lors de l'assemblage, mais on ne le stocke pas. Par conséquent, on est amené à recalculer ces inverses si on veut reconstituer la solution

sur tous les degrés de liberté du maillage. Il faut également modifier le terme source dans le cas d'une source "volumique". On voit dans le tableau 2.1, que la condensation statique permet de diminuer la mémoire requise par un solveur direct, lorsqu'on monte en ordre. Cette diminution paraît moindre en 3-D. En effet, la condensation statique a pour effet de coupler tous les degrés de liberté du bord, on perd malheureusement l'orthogonalité partielle qu'on avait (sans condensation, un ddl n'interagit qu'avec les ddls situés sur un même plan parallèle à Oxy, Oxz ou Oyz) ! Ainsi en 2-D, la condensation statique nous fait passer d'une complexité en  $O(r^4)$  à une complexité en  $O(r^2)$ , tandis qu'en 3-D on passe seulement d'une complexité en  $O(r^5)$  à une complexité en  $O(r^4)$ .

Pour conclure cette sous-section, un solveur direct a d'excellentes propriétés en 2-D, permettant de traiter de gros cas. On peut même économiser de la mémoire quand on monte en ordre, en effectuant une condensation statique. En 3-D, un solveur direct requiert beaucoup trop de mémoire - et donc de temps de calcul -, il est nécessaire d'avoir recours à un solveur itératif.

## 2.3 Résolution itérative

Les méthodes itératives les plus utilisées dans le cadre des équations de Maxwell, sont spécialement dévolues aux systèmes complexes symétriques. C'est notamment COCG (Conjugate Orthogonal Conjugate Gradient) et BICGCR (BiConjugate Gradient Conjugate Residual method), elles sont détaillées dans [Clemens et Weiland, 2002]. Des méthodes plus classiques sont également utilisées comme GMRES (Generalized Minimal RESidual, cf. [Saad, 1996] [Saad et Schultz, 1986]), BICGSTAB (BiConjugate Gradient STABilized cf. [der Vorst, 1992]) ou SQMR (Symmetric Quasi-Minimal Residual [Freund et Nachtigal, 1991]). GMRES et BICGSTAB ont l'avantage d'être conçus pour des matrices non-symétriques. Tous les autres algorithmes utilisent fortement la propriété de symétrie. Par ailleurs, COCG et SQMR sont les équivalents de BICG et QMR. Lorsque la matrice est symétrique, on peut simplifier l'algorithme du BICG et QMR pour obtenir respectivement COCG et SQMR. Le COCG est équivalent au gradient conjugué pour des matrices réelles. Le paramètre de restart pour GMRES est pris égal à 20. Malheureusement, ces méthodes convergent d'autant plus lentement que le conditionnement de la matrice se détériore. Pour accélérer la convergence de ces méthodes, on "préconditionne" la matrice par un inverse "approché". Dans cette section, on n'utilisera pas de préconditionneur, la question du préconditionnement sera abordée dans la section suivante.

### 2.3.1 Cas 2-D

#### Différentes méthodes de Krylov

Nous testons ces différentes méthodes itératives sur le cas du disque parfaitement conducteur de diamètre 20 longueurs d'onde. On a ainsi un cas relativement haute-fréquence, on utilise une approximation  $Q_5$  avec 10 points par longueur d'onde. Le nombre de ddl est égal à 20 000 environ. On fixe un critère d'arrêt de  $\varepsilon = 10^{-6}$ , on obtient les résultats de la figure 2.5 et du tableau 2.3. On observe que BICGCR et COCG sont les algorithmes les plus rapides avec un résidu fortement oscillant. On peut rajouter un "post-traitement" afin de lisser ce résidu en utilisant le MRS (Minimum Residual Smoothing). Pour mieux connaître ce type de technique, le lecteur pourra lire [Zhou et Walker, 1994] et [Gutknecht et Rozloznik, 2001]. En pratique, le lissage n'améliore pas grandement le nombre d'itérations nécessaires, on a choisi de ne pas l'utiliser. QMR donne des résultats proches avec un résidu évoluant par paliers. BICGSTAB a tendance à stagner, il n'est pas intéressant à utiliser sans préconditionnement. GMRES converge correctement, mais pas dans le cas diélectrique (cf. figure 2.6). Sans préconditionneur, il semble que la méthode la plus rapide sur ce cas test soit le BICGCR. Le COCG a cependant l'avantage

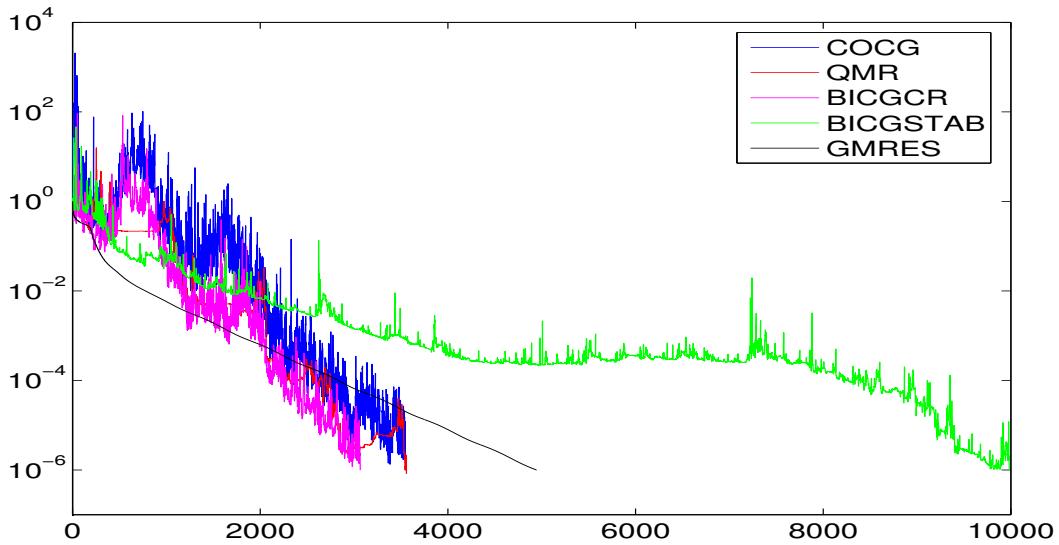


FIG. 2.5 – Evolution du résidu en fonction du nombre d’itérations, échelle logarithmique. Cas du disque parfaitement conducteur.

Méthode	Nombre d’itérations	Temps
COCG	3 547	53 s
QMR	3 558	58 s
BICGCR	3 068	49 s
BICGSTAB	9 982	149 s
GMRES	4 495	78 s

TAB. 2.3 – Nombre d’itérations et temps suivant l’algorithme itératif

d’être plus simple et d’avoir un stockage réduit (4 vecteurs). En l’absence de mention, la méthode itérative utilisée est le COCG.

Nous testons maintenant ces méthodes de Krylov sur la diffraction d’un disque diélectrique ( $\rho = 4 \mu = 1$ ). Le nombre de longueurs d’onde dans le diélectrique est d’une vingtaine environ. On obtient les résultats de la figure 2.6. On notera que sur le cas hétérogène, GMRES(20) et BICGSTAB stagnent très fortement, ce qui prohibe leur utilisation dans ce cas. Nous avons noté la même difficulté pour les problèmes de cavité. Nous verrons ultérieurement que même en présence d’un préconditionneur, ces méthodes sont souvent moins efficaces que le COCG, BICGCR ou QMR.

### Influence de la fréquence

On s’intéresse au nombre d’itérations nécessaires lorsqu’on monte en fréquence en adaptant le pas de maillage afin d’avoir dix points par longueur d’onde. On peut soit raffiner le maillage, soit monter en ordre. On remarque sur la figure 2.7, que le nombre d’itérations augmente dans les deux cas, l’augmentation étant plus prononcée quand on monte en ordre. La détérioration de la convergence est loin d’être problématique, il est probable que si on avait choisi des points réguliers au lieu des points de Gauss-Lobatto, on aurait eu des résultats bien pires. Il semble que la complexité du nombre d’itérations soit en  $O(k)$  (relation affine entre le nombre d’itérations et le nombre d’onde). En multipliant par la complexité du produit matrice-vecteur en  $O(k^2)$ ,

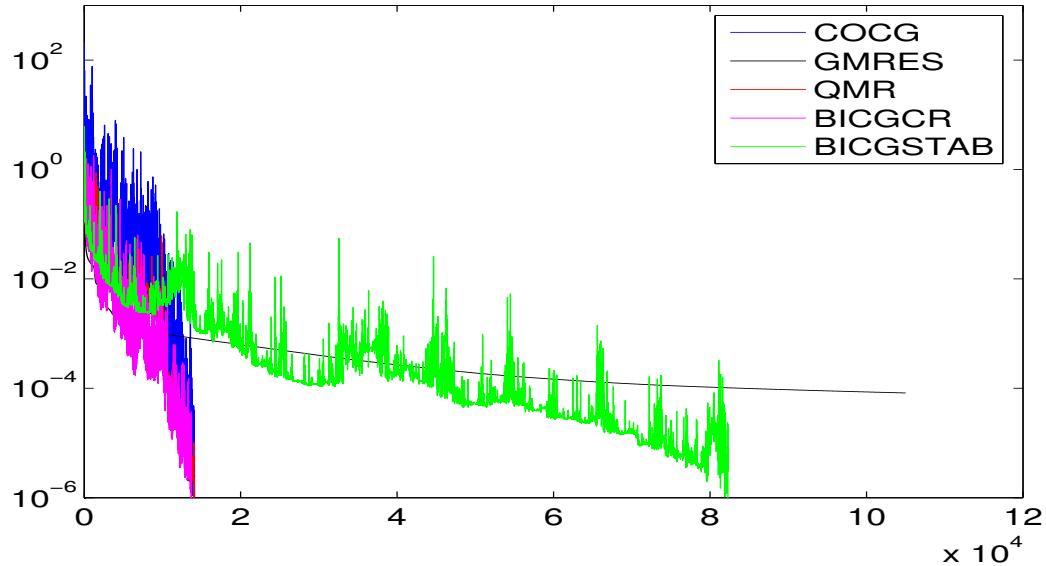


FIG. 2.6 – Evolution du résidu en fonction du nombre d’itérations, échelle logarithmique. Cas du disque diélectrique.

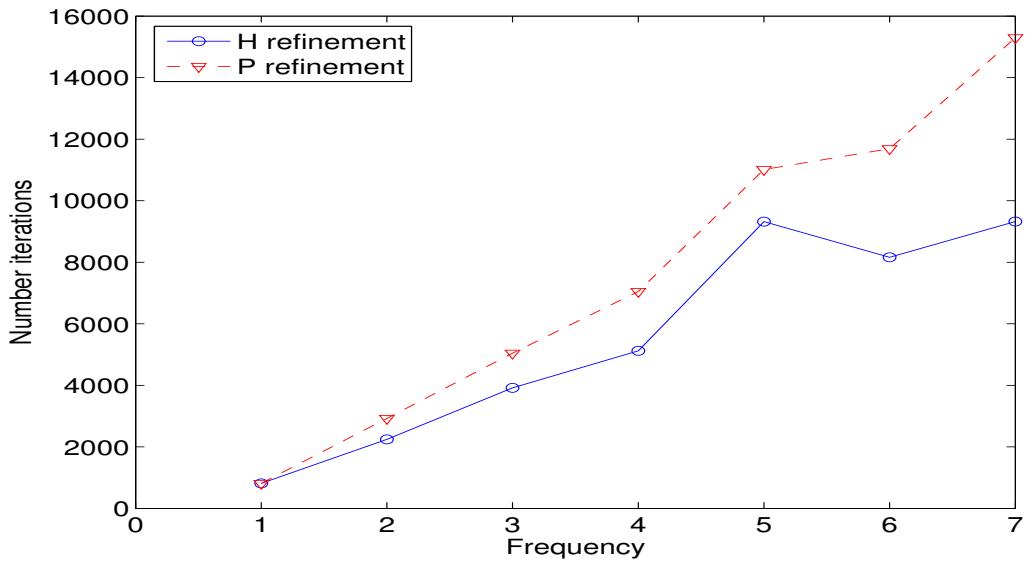


FIG. 2.7 – Nombre d’itérations en fonction de la fréquence (le pas de maillage s’adapte). Cas du disque parfaitement conducteur.

on trouve que la méthode éléments finis sans préconditionnement est de complexité  $O(k^3)$  en 2-D.

### Influence du maillage

Une autre interrogation concerne la robustesse d’une méthode itérative suivant le maillage utilisé. On considère trois maillages différents (cf. figure 2.8) avec 8 points par longueur d’onde en  $Q_5$ . Le nombre de degrés de liberté est le même sur les trois maillages (environ 20 000). On

obtient les résultats du tableau 2.4

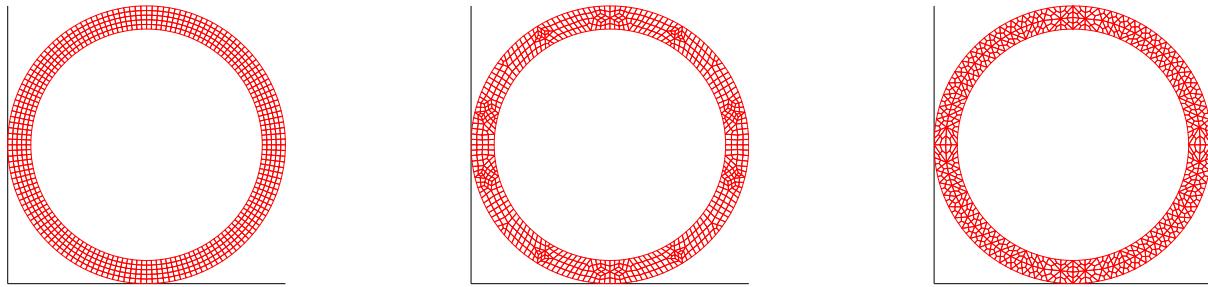


FIG. 2.8 – Maillage régulier à gauche, maillage quadrilatéral au milieu, maillage “triangles découpés” à droite

Maillage	Nombre d’itérations	Temps
régulier	3 809	58 s
quadrilatéral	5 058	74 s
triangles découpés	5 870	92 s

TAB. 2.4 – Performance du COCG avec différents maillages. Cas du disque parfaitement conducteur.

Sans surprise, le conditionnement de la matrice se détériore si on raffine le maillage, si on utilise de l’ordre élevé, si on monte en fréquence et si on utilise des maillages de qualité médiocre !

### 2.3.2 Cas 3-D

#### Différentes méthodes de Krylov

On fait le même test sur le cas de la diffraction par une sphère de diamètre  $20\lambda$ . Le maillage  $Q_5$  contient 1 130 000 degrés de liberté (8 points par longueur d’onde), on obtient les résultats de la figure 2.9 et du tableau 2.5. Les résultats sont identiques à ce qu’on a obtenu en 2-D !

Méthode	Nombre d’itérations	Temps
COCG	7 717	3 h 3mn
QMR	8 226	3 h 40mn
BICGCR	7 021	3 h
BICGSTAB	19 741	7 h 44mn
GMRES	8 742	4 h 8mn

TAB. 2.5 – Nombre d’itérations et temps suivant l’algorithme itératif. Cas de la sphère parfaitement conductrice

On notera néanmoins que le ratio  $\frac{\text{Nombre d’itérations}}{\text{Nombre ddl}}$  est plus faible qu’en 2-D car pour un cas de même taille, on a beaucoup plus de degrés de liberté alors que le nombre d’itérations a doublé (pour le disque, on comptait 4 000 itérations, contre 8 000 environ pour la sphère). Ce comportement va à l’opposé de ce qu’on avait pour un solveur direct, en effet quand on passe du 2-D au 3-D, le temps de calcul nécessaire pour une résolution directe devient plus important pour un même nombre de ddl. Pour un solveur itératif, c’est l’inverse, pour un même nombre

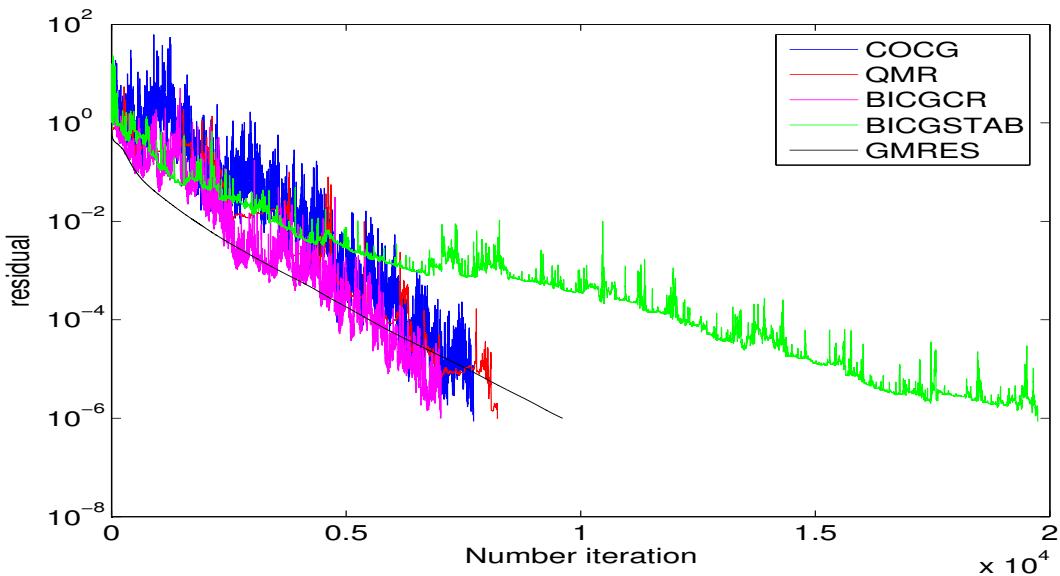


FIG. 2.9 – Evolution du résidu en fonction du nombre d’itérations, échelle logarithmique. Cas de la sphère parfaitement conductrice.

de ddl, le temps de calcul sera plus faible en 3-D qu’en 2-D. On ne montre pas les résultats pour la diffraction d’une sphère diélectrique car ils sont semblables à ce qu’on a observé en 2-D. De manière générale, les cas 2-D et 3-D ont de fortes similitudes.

### Influence de la fréquence

Comme en 2-D, on peut constater que le nombre d’itérations croît linéairement en fonction de la fréquence (cf. figure 2.10). La complexité d’un calcul 3-D en utilisant un solveur itératif sans préconditionnement est donc en  $O(k^4)$ .

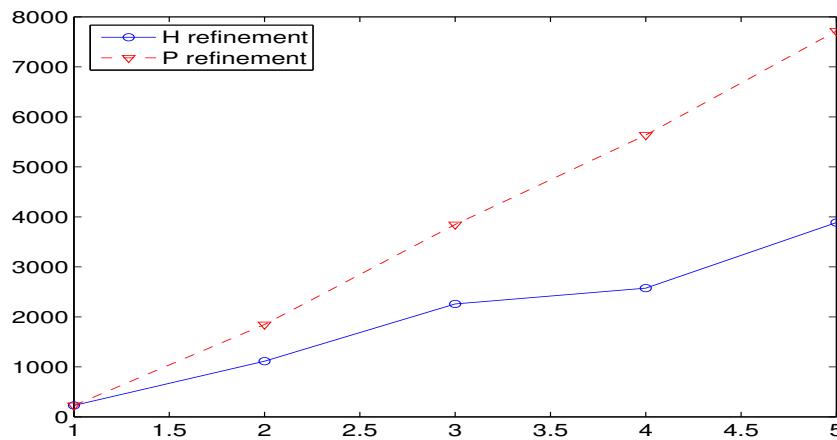


FIG. 2.10 – Nombre d’itérations lorsqu’on raffine en maillage ou en ordre, la fréquence est augmentée dans la même mesure. Cas de la sphère parfaitement conductrice.

## 2.4 Préconditionnement

Les matrices qu'on obtient sont mal conditionnées, certaines valeurs propres peuvent être proches de zéro (fréquence proche d'une résonance comme dans le cas Atmel), certaines valeurs propres peuvent être très grandes sur des maillages de qualité médiocre. Il est nécessaire de préconditionner le système linéaire, ce qui revient à remplacer le système linéaire :

$$A_h U_h = F_h$$

par le système :

$$M_h A_h U_h = M_h F_h$$

On peut également préconditionner à droite et résoudre :

$$A_h M_h Y_h = F_h \quad U_h = M_h Y_h$$

$M_h$  est le préconditionneur, c'est un opérateur linéaire de telle sorte que :

$$\text{conditionnement}(A_h M_h) \leq \text{conditionnement}(A_h)$$

Si on utilise des algorithmes symétriques (comme le COCG), il est nécessaire de disposer de préconditionneurs symétriques. On n'utilise pas les techniques précédentes car le produit  $A_h M_h$  n'est pas symétrique. Le préconditionneur s'applique en modifiant le produit scalaire :

$$\langle x, y \rangle = (x, M_h y)$$

En pratique, il n'est pas obligatoire de construire la matrice  $M_h$  de manière effective, on n'a besoin que d'avoir un algorithme de calcul de l'application du préconditionneur sur un vecteur (produit matrice-vecteur) :

$$Y = M_h X$$

De nombreux travaux ont été menés pour trouver des préconditionneurs efficaces pour l'équation de Helmholtz. Une approche séduisante est l'utilisation de la fft pour la résolution approchée de cette équation sur des domaines tensoriels, c'est une technique exploitée par [Elman et O'Leary, 1998], [Otto et Larsson, 2000], [Heikkola *et al.*, 2003a] et [Heikkola *et al.*, 2003b]. On peut citer dans le même ordre d'idées les travaux originaux de [Gander et Nataf, 2001], qui proposent d'utiliser la factorisation analytique de l'opérateur de Helmholtz en deux opérateurs d'ordre un pour réaliser une factorisation incomplète analytique. Afin d'exploiter les idées sous-jacentes à ces différents travaux, il est nécessaire de coupler les éléments finis avec des différences finies pour exploiter la structure tensorielle des opérateurs. Ce couplage ne nous semble pas facile à réaliser, nous avons choisi d'ignorer cette technique.

Une autre approche est d'utiliser l'équation de Helmholtz avec amortissement, soit en utilisant une factorisation incomplète, soit en utilisant un algorithme multigrille. Nous avons trouvé cette approche satisfaisante car simple à réaliser, et pouvant s'adapter à n'importe quelle discrétisation. Le lecteur pourra lire [Bayliss *et al.*, 1983], [Vuik *et al.*, 2003] et [Erlangga, 2002] pour approfondir le sujet.

On peut également utiliser la décomposition en sous-domaines sur ce type d'équation. Cette approche est surtout intéressante à utiliser pour paralléliser la résolution. Pour une implémentation en séquentiel, le gain en mémoire réalisé grâce à cette technique n'est pas des plus formidables. Parmi les nombreux articles sur le sujet, on pourra lire [Collino *et al.*, 1988], [Benamou et Despres, 1996], [Larsson, 1999], [Toselli, 1998], [Gander *et al.*, 2002].

Une autre alternative est de faire du multigrille directement sans adjoindre de l'amortissement. On distingue le multigrille géométrique, qui utilise des maillages plus grossiers pour

approcher la solution, et le multigrille algébrique, qui construit la matrice du maillage grossier uniquement à partir de la matrice du maillage fin. Le principal défaut du multigrille géométrique est que le maillage grossier doit être suffisamment fin (quatre points par longueur d'ondes est un bon critère), le multigrille devient alors du 2-grilles, cette approche reste quand même intéressante, [Elman *et al.*, 2001]. On n'a pas regardé en détail le multigrille algébrique pour donner un avis pertinent, mais cette technique semble a priori intéressante [Vanek *et al.*, 1998a], [Vanek *et al.*, 1998b], [Vanek *et al.*, 1997].

#### 2.4.1 Préconditionnement par l'équation de Helmholtz avec amortissement

On considère l'équation de Helmholtz amortie :

$$-k^2(\alpha + i\beta)u - \Delta u = 0 \quad (2.1)$$

Il y aura absorption si  $\beta > 0$ , ce signe est à mettre en relation avec la convention qu'on a choisie pour la condition absorbante :

$$\frac{\partial u}{\partial n} - ik u = 0$$

On utilise les mêmes éléments finis que pour l'équation de Helmholtz pour construire  $M_h^{-1}$  :

$$(M_h^{-1})_{i,j} = -k^2(\alpha + i\beta) \int_{\Omega} \varphi_i \varphi_j dx + \int_{\Omega} \nabla \varphi_i \cdot \nabla \varphi_j dx$$

Appliquer le préconditionneur  $M$  à un vecteur  $X$ , revient à résoudre le système linéaire :

$$M_h^{-1} X = Y$$

L'idée d'un tel préconditionneur vient de Bayliss, Goldstein and Turkel [Bayliss *et al.*, 1983] qui ont proposé de préconditionner l'équation de Helmholtz par le laplacien ( $\alpha = 0$   $\beta = 0$ ), Laird a ensuite proposé de prendre  $\alpha = -1$   $\beta = 0$ . Plus récemment, Y.A. Erlangga et al [Vuik *et al.*, 2003] ont étudié les qualités de ce préconditionneur pour  $\alpha \leq 0$  et  $\beta > 0$ . En choisissant ces deux conditions, on s'assure de la coercivité de la formulation variationnelle issue de l'équation (2.1) (ce qui n'est pas le cas pour l'équation de Helmholtz). La matrice  $M_h^{-1}$  est définie positive, on peut alors approcher  $M_h$  par des préconditionneurs efficaces du laplacien, comme la factorisation incomplète ou du multigrille. Y.A. Erlangga a montré que le choix  $(\alpha, \beta) = (0, 1)$  était optimal sous la condition  $\alpha \leq 0$ . Néanmoins il a observé numériquement que le choix  $(\alpha, \beta) = (1, 0.5)$  donnait les meilleurs résultats. Nous garderons le plus souvent ce choix de paramètres. La première approche conduit à un stockage important, le coût du préconditionneur est tel que le gain en temps de calcul et en stockage réalisé sur la matrice à l'aide de la formulation mixte est négligeable. Le seul intérêt de cette approche est de gagner en stockage par rapport à un solveur direct. On mettra en évidence les gains réalisés par la suite. En revanche, la seconde approche est plus satisfaisante comme le montre [Erlangga *et al.*, 2004], elle permet d'avoir un stockage extrêmement réduit en utilisant le produit matrice-vecteur présenté précédemment.

#### Factorisation incomplète

On résout le système linéaire

$$M_h^{-1} X = Y$$

par une méthode de factorisation incomplète : ILUT( $\varepsilon$ ), détaillée dans [Saad, 1996]. Cela revient à faire une factorisation LU de la matrice en éliminant les coefficients de module inférieur au seuil

$\varepsilon$ . Le désavantage de cette approche est de conduire à un stockage relativement important. On exploite la symétrie de la matrice en ne stockant que la partie supérieure de la factorisation. Le préconditionneur est alors symétrique, toutefois nous montrons quelques résultats numériques utilisant GMRES(20).

Dans un premier temps, on met en évidence l'importance de l'amortissement. En effet, le tableau 2.6 montre qu'en l'absence d'amortissement, la factorisation incomplète “décroche” rapidement, et peut même être plus coûteuse que la résolution sans préconditionneur.

seuil $\varepsilon$	1e-3	3e-3	1e-2	2e-2	4e-2
$\alpha = 1 \quad \beta = 0$	13 / 28 Mo	37 / 27 Mo	$\infty$ / 25 Mo	$\infty$ / 22 Mo	$\infty$ / 16 Mo
$\alpha = 1 \quad \beta = 0.5$	117 / 26 Mo	118 / 23 Mo	133 / 19 Mo	160 / 15 Mo	254 / 10 Mo
$\alpha = 1 \quad \beta = 1.0$	215 / 23 Mo	216 / 20 Mo	224 / 16 Mo	239 / 13 Mo	295 / 9 Mo

TAB. 2.6 – Nombre d’itérations en préconditionnant par une factorisation incomplète. Le cas test est la diffraction par un disque parfaitement conducteur de rayon 10. On adopte le format “x / y Mo”, où x est le nombre d’itérations et y la taille mémoire utilisée par la factorisation incomplète.

On voit qu’en rajoutant de l’amortissement ( $\beta = 0.5$ ), on évite cet écueil, de plus la factorisation incomplète nécessite alors moins d’espace mémoire. Sur cet exemple, le choix  $\beta = 0.5$  est meilleur que le choix plus classique  $\beta = 1$ . On obtient des gains appréciables en coût de stockage par rapport à un solveur direct. On divise par 2 à 3 fois le stockage par rapport à une factorisation classique.

Dans un second temps, on s’intéresse à l’influence de l’ordre d’approximation, à nombre de degrés de liberté constant, cf. tableau 2.7. Le cas test est la diffraction d’une sphère parfaitement conductrice ( $a = 4$ ,  $b = 5$ ), avec 105 000 degrés de liberté.

seuil $\varepsilon$	1e-3	3e-3	1e-2	2e-2	4e-2
Q1	57 / 117 Mo	58 / 75 Mo	71 / 40 Mo	108 / 27 Mo	195 / 18 Mo
Q2	57 / 283 Mo	60 / 155 Mo	89 / 68 Mo	187 / 43 Mo	316 / 27 Mo
Q3	57 / 306 Mo	60 / 185 Mo	88 / 90 Mo	183 / 57 Mo	343 / 37 Mo
Q4	57 / 333 Mo	60 / 210 Mo	80 / 110 Mo	150 / 73 Mo	306 / 48 Mo
Q5	57 / 389 Mo	60 / 233 Mo	95 / 115 Mo	184 / 78 Mo	338 / 52 Mo

TAB. 2.7 – Nombre d’itérations et taille de la factorisation incomplète. Diffraction par une sphère.

L’ordre d’approximation n’affecte que la taille de la matrice LU, le nombre d’itérations ne varie pas beaucoup. En 3-D, le gain de stockage à l’aide de la factorisation incomplète est plus substantiel, allant parfois jusqu’à un facteur de 10!. Au vu de ces expériences, il semble raisonnable de choisir  $\varepsilon = 10^{-2}$ , voire légèrement supérieur si le stockage est trop important.

Une parade pour diminuer le stockage pour des ordres d’approximation élevés, consiste à utiliser un sous-maillage  $Q_1$  du domaine de calcul. Ce sous-maillage  $Q_1$  est généré en subdivisant le maillage initial sur les points de Gauss-Lobatto, comme le montre la figure 2.11. Les points du sous-maillage coincident exactement avec les degrés de liberté du maillage initial. On calcule alors la factorisation incomplète sur la matrice éléments finis  $Q_1$  de ce sous-maillage. On a par conséquent quel que soit l’ordre d’approximation le même stockage, le nombre d’itérations est légèrement plus élevé comme le montre le tableau 2.8.

Globalement le temps de calcul est sensiblement le même car chaque itération coûte moins

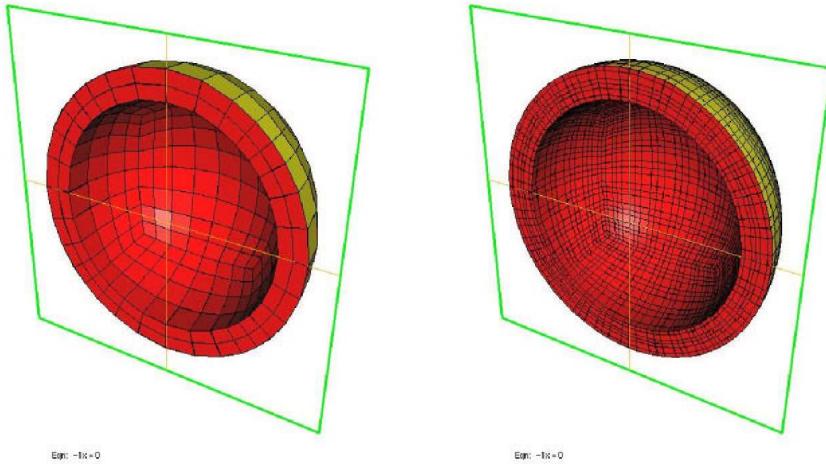


FIG. 2.11 – A gauche, maillage  $Q_3$ , à droite sous-maillage  $Q_1$

seuil	1e-3	3e-3	1e-2	2e-2	4e-2
$Q_2$	80(57)	81(60)	101(89)	167(187)	326(316)
$Q_3$	86(57)	87(60)	102(98)	171(183)	335(343)
$Q_4$	89(57)	91(60)	104(80)	170(150)	316(306)
$Q_5$	91(57)	94(60)	118(95)	208(184)	396(338)

TAB. 2.8 – Nombre d’itérations et taille de la factorisation incomplète. Diffraction d’une sphère. Utilisation du sous-maillage  $Q_1$ .

cher (factorisation moins volumineuse), mais on a plus d’itérations. Le point positif de cette technique est bien d’avoir un stockage moins important.

## Multigrille

On peut encore gagner en stockage en utilisant un algorithme multigrille au lieu d’une factorisation incomplète. Une analyse complète et détaillée de cette méthode utilisant des schémas différences finies est présentée dans le rapport de Y.A. Erlangga. Nous décrivons brièvement l’itération multigrille, et plus en détail les opérateurs de restriction et de prolongement, spécifiques à la méthode éléments finis qu’on utilise.

**Itération multigrille** On utilise le même maillage  $\Omega_h$ , on ne fait varier que l’ordre d’approximation noté  $p$ . L’algorithme classique s’écrit :

```

Multigrille( $A_p$ , p, x, b)
  Pré-lissage  $x = S_p^{\nu_1}(x, b)$ 
  Calcul résidu  $r = b - A_p x$ 
  Restriction  $r_c = R_p r$ 
  Si  $p/2 \leq r_m$ 
    Résolution directe  $x_c = A_{p/2}^{-1} r_c$ 
  sinon
     $x_c = 0$ 
    Pour i = 1,  $\gamma$ 
      Multigrille( $A_{p/2}$ , p/2,  $x_c$ ,  $r_c$ )
    Fin Pour
  Fin Si
  Prolongement  $r = P_p x_c$ 
   $x = x + r$ 
  Post-lissage  $x = S_p^{\nu_2}(x, b)$ 
Fin Multigrille

```

On a choisi les notations suivantes :

- $A_p$ , matrice élément fini pour l’ordre d’approximation  $p$
- $R_p$ , opérateur de restriction de l’ordre d’approximation  $p$  vers  $p/2$
- $P_p$ , opérateur de prolongement de l’ordre d’approximation  $p/2$  vers  $p$
- $r_m$ , ordre minimal pour lequel on fait une résolution directe
- $\gamma = 1$  correspond à un V-cycle,  $\gamma = 2$  correspond à un W-cycle
- $S_p$  lisseur
- $\nu_1$  nombre d’itérations de pré-lissage
- $\nu_2$  nombre d’itérations de post-lissage

Pour le lisseur, on utilise du Jacobi relaxé avec  $\omega = 0.5$  :

$$S_p(x, b) = x + \omega(b - A_p x)$$

On a choisi de mettre une itération de post-lissage et une itération de pré-lissage. On a ainsi un préconditionneur symétrique (les opérateurs de prolongement et de restriction sont transposés l’un de l’autre, comme on va le voir dans le paragraphe suivant). On choisit un W-cycle ( $\gamma = 2$ ) pour des questions de performance.

**Opérateurs de prolongement et restriction** On note  $\psi_i^c$  les fonctions de base du maillage grossier (ordre d’approximation de  $p/2$ ), et  $\varphi_i^f$  les fonctions de base du maillage fin (ordre d’approximation  $p$ ). Le maillage est le même, seul l’ordre change. Cependant, par abus de notation de notation, nous parlons de maillage fin et de maillage grossier.

L’opérateur de prolongement est défini par :

$$P_{i,j} = \psi_j^c(\xi_i^f)$$

où les points  $\xi_i^f$  désigne la position du degré de liberté  $i$  du maillage fin. On prend de manière classique l’opérateur de restriction suivant :

$$R = P^t$$

Avec ces définitions, on peut espérer obtenir un préconditionneur symétrique.

A partir de là, deux possibilités s'offrent à nous : soit on assemble la matrice  $P$ , et on utilise un produit matrice vecteur creux classique pour évaluer  $Y = P X$ ; soit on fait un produit matrice-vecteur qui utilise la tensorisation des fonctions de base. En utilisant le dernier procédé, qu'on appellera “procédé boîte noire”, on aura une complexité en  $O(r^4)$  contre  $O(r^6)$  pour le premier procédé (cas 3-D).

Explicitons donc le procédé boîte noire, qui s'appuie sur la tensorisation des fonctions de base :

$$\hat{\psi}_j^c(\hat{\xi}_i^f) = \hat{\psi}_{j_1}^c(\hat{\xi}_{i_1}^f) \hat{\psi}_{j_2}^c(\hat{\xi}_{i_2}^f) \hat{\psi}_{j_3}^c(\hat{\xi}_{i_3}^f)$$

Le produit matrice vecteur  $Y = P X$  s'écrit localement comme une triple somme :

$$Y_{i_1,i_2,i_3} = \sum_{j_1,j_2,j_3} \hat{\psi}_{j_1}^c(\hat{\xi}_{i_1}^f) \hat{\psi}_{j_2}^c(\hat{\xi}_{i_2}^f) \hat{\psi}_{j_3}^c(\hat{\xi}_{i_3}^f) X_{j_1,j_2,j_3}$$

On décompose la triple somme en trois sommes simples, et on obtient l'algorithme souhaité !

$$w1_{j_1,j_2,i_3} = \sum_{j_3} \hat{\psi}_{j_3}^c(\hat{\xi}_{i_3}^f) X_{j_1,j_2,j_3}$$

$$w2_{j_1,i_2,i_3} = \sum_{j_2} \hat{\psi}_{j_2}^c(\hat{\xi}_{i_2}^f) w1_{j_1,j_2,i_3}$$

$$w3_{i_1,i_2,i_3} = \sum_{j_1} \hat{\psi}_{j_1}^c(\hat{\xi}_{i_1}^f) w2_{j_1,i_2,i_3}$$

$$Y_{i_1,i_2,i_3} = w3_{i_1,i_2,i_3}$$

Il n'est donc nécessaire de ne stocker que les coefficients  $\hat{\psi}_{j_1}^c(\hat{\xi}_{i_1}^f)$ , indépendants de la géométrie. Le coût de stockage est nul par rapport au premier procédé qui exigeait de stocker toute la matrice  $P$ . On note  $p_c = p/2$  et  $p_f = p$ . On effectue des calculs de complexité de temps de calcul produit matrice-vecteur des deux procédés :

Opérations pour P assemblée (2-D) :  $[(p_c - 1)(p_f + 1) + (2p_f + 1)]^2$

Opérations pour P assemblée (3-D) :  $[(p_c - 1)(p_f + 1) + (2p_f + 1)]^3$

Opérations pour P boîte noire (2-D) :  $2(p_f + 1)^2(p_c + 1) + 2(p_f + 1)(p_c + 1)^2 + (p_f + 1)^2$

Opérations pour P boîte noire (3-D) :  $2(p_f + 1)^3(p_c + 1) + 2(p_f + 1)^2(p_c + 1)^2 + 2(p_f + 1)(p_c + 1)^3 + (p_f + 1)^3$

Les graphes représentant ces complexités sont sur la figure 2.12.

Pour les ordres impairs, on a pris l'arrondi supérieur de  $p_c = p/2$ . En effet, le passage de  $Q_3$  à  $Q_1$  ou  $Q_5$  à  $Q_2$  est trop brutal, et donne des résultats assez médiocres. En 2-D, on utilisera le procédé classique pour  $p \leq 3$  et le procédé boîte noire pour des ordres supérieurs. En 3-D, le procédé boîte noire est toujours le plus performant.

**Résultats préliminaires sur la sphère** On considère la même sphère que dans le cas de la factorisation incomplète. On utilise du COCG préconditionné par une itération multigrille. Pour le maillage grossier  $Q_1$ , on utilise la factorisation incomplète ILUT(3e-3) pour la résolution du système linéaire.

Le gain en nombre d'itérations est appréciable(cf. tableau 2.9), le gain en temps de calcul n'est pas négligeable notamment pour des ordres d'approximation élevés. Le nombre d'itérations est néanmoins largement supérieur au nombre d'itérations qu'on pouvait espérer (NDLR : 60 itérations). Pour expliquer cette difficulté nous regardons le nombre d'itérations en fonction du pas de maillage pour  $Q_2$  (cf. tableau 2.10) :

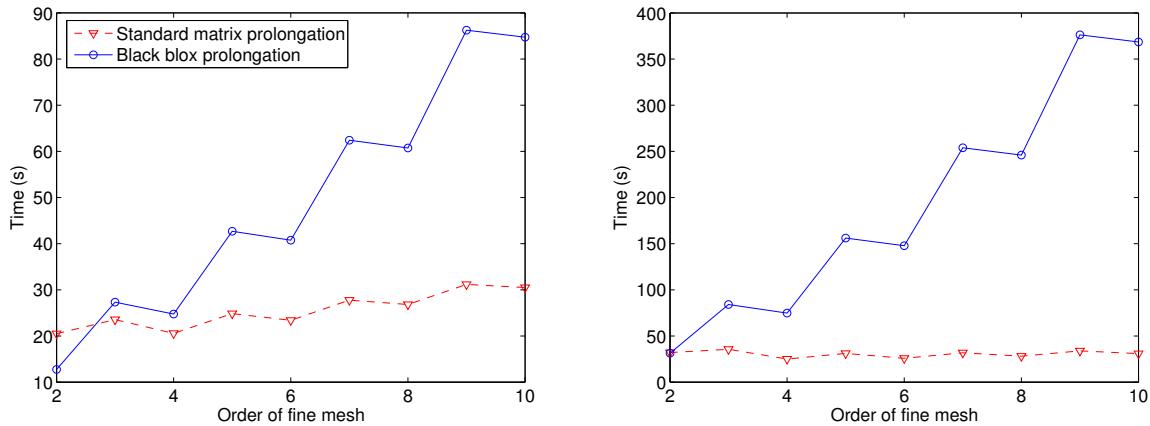


FIG. 2.12 – Temps de calcul pour l’opérateur de prolongement/restriction. A gauche 2-D, à droite 3-D

Ordre	$Q_2$	$Q_4$	$Q_8$
Nombre itérations	211	151	134
Temps	79s	72s	64s
Nombre itérations sans préconditionneur	1 971	2 235	2 988
Temps sans préconditionneur	169s	264s	346s

TAB. 2.9 – Performances du préconditionneur multigrille sur la sphère parfaitement conductrice (condition de Neumann)

Pas de maillage	1.0	0.5	0.25	0.125
Nombre itérations	228	245	211	80

TAB. 2.10 – Nombre d’itérations du COCG préconditionné par une itération multigrille, pour différents pas de maillage

Le pas de maillage  $h = 0.25$  correspond à huit points par longueur d’onde. Lorsqu’on raffine le maillage, la grille grossière est mieux adaptée, on a donc moins d’itérations. Une difficulté est le manque de précision évident de  $Q_1$ . De fait, il est souvent plus adéquat de poser comme ordre minimal :

$$r_m = 2$$

Mais cela impose d’utiliser au minimum du  $Q_4$ .

**Une alternative a priori séduisante...** On peut également avoir l’idée d’utiliser le sous-maillage  $Q_1$  (cf. figure 2.8) pour préconditionner  $Q_k$  par une itération multigrille sur le sous-maillage  $Q_1$ . En faisant cette opération, on stocke la matrice  $Q_1$  (qui peut être relativement volumineuse), mais on a un produit matrice-vecteur plus rapide que le produit matrice vecteur avec de l’ordre élevé. Le coût du lisseur est moindre, et a priori les opérateurs de restriction et de prolongement sont peu coûteux. Pour ces derniers, on utilise des opérateurs classiques pour les fonctions de base d’ordre 1, cf. [Hackbusch, 1985]. Sur le même cas que précédemment on compare les temps obtenus avec les deux techniques sur le tableau 2.11, ainsi que l’espace mémoire requis. On utilise cette fois un solveur direct pour le maillage grossier et du BICGCR, ce qui donne légèrement moins d’itérations que dans le paragraphe précédent. On peut voir ainsi que pour du  $Q_8$ , le BICGCR préconditionné par du multigrille classique a besoin de 129

Ordre	$Q_2$	$Q_4$	$Q_8$
Nombre itérations	193 / 212	144 / 184	129 / 186
Temps	74s / 62s	64s / 59s	58s / 51s
Mémoire	78 Mo / 105 Mo	35 Mo / 65 Mo	24 Mo / 51 Mo

TAB. 2.11 – Nombre d’itérations du BICGCR préconditionné par du multigrille. A gauche, statistiques pour le multigrille classique, à droite pour le multigrille utilisant le sous-maillage  $Q_1$

itérations alors qu’il a besoin de 186 itérations lorsqu’on utilise du multigrille sur le sous-maillage  $Q_1$ . Néanmoins, on gagne légèrement en temps de calcul, car chaque itération est beaucoup moins coûteuse. En revanche, au niveau de l’espace mémoire, on a besoin de 51 Mo contre 24 Mo pour le multigrille classique. Dans cet espace mémoire on ne fait pas figurer le coût des vecteurs d’itérations (20 Mo). L’espace mémoire est un facteur important pour la résolution de cas 3-D de grandes tailles, on n’utilisera donc pas cette alternative.

### Défaut des préconditionneurs basés sur l’équation de Helmholtz amortie

Les préconditionneurs basés sur l’équation de Helmholtz amortie paraissent souffrir d’une limitation importante. Ils ont l’avantage non-négligeable d’avoir une complexité indépendante du pas de maillage, de la déformation du maillage et de l’ordre d’approximation. Mais leur complexité dépend toujours fortement de la fréquence ! En outre, sur le cas hétérogène, le nombre d’itérations est assez important. Nous illustrons ces propriétés sur le cas de la diffraction d’un disque diélectrique ( $\rho = 4 \mu = 1$ ), le même cas qu’au début du chapitre. Sur le tableau 2.12, on note le nombre d’itérations nécessaires lorsqu’on préconditionne par l’équation de Helmholtz amortie (factorisation complète !).

Fréquence	0.25	0.5	1.0	2.0	3.0
Nombre itérations BICGCR	68	203	600	1800	2700

TAB. 2.12 – Nombre d’itérations pour différentes fréquences. BICGCR préconditionné par l’équation de Helmholtz amortie  $\alpha = 1.0 \quad \beta = 0.5$ . Une fréquence de 1 correspond à la diffraction d’un disque diélectrique de diamètre 20 longueurs d’onde (dans le diélectrique).

### 2.4.2 Décomposition en sous-domaines

L’idée est de résoudre des problèmes locaux sur des domaines de petite taille. L’avantage escompté de cette technique est de gagner en stockage si on utilise un solveur direct pour résoudre les systèmes linéaires dans chaque sous-domaine. En effet, on a vu que la complexité d’un solveur direct était à peu près en  $O(N^2)$  en 3-D. En morcellant le domaine, on peut espérer atteindre une complexité en  $O(N)$ .

On utilise une méthode de Schwarz additive sans recouvrement. On note la partition en sous-domaines :

$$\Omega = \bigcup_{i=1}^{N_s} \Omega_i$$

Sur chaque sous-domaine  $\Omega_i$ , on résout le problème aux limites :

$$\begin{aligned} -\rho \omega^2 u - \Delta u &= 0 \text{ dans } \Omega_i \\ \frac{\partial u}{\partial n} - ik u &= 0 \quad \text{sur } \partial\Omega_i \cap \partial\Omega_j \\ + \text{ conditions aux limites} &\quad \text{sur } \partial\Omega_i \cap \partial\Omega \end{aligned}$$

Sur les interfaces entre sous-domaines, on impose une condition absorbante d'ordre 1. On s'assure ainsi que chaque problème dans un sous-domaine est bien posé. On note  $A_i$  la matrice éléments finis du sous-domaine  $\Omega_i$ . Le préconditionneur s'écrit alors :

$$M^{-1} = \sum_{i=1}^{N_s} P_i A_i^{-1} P_i^t$$

$P_i$  est l'opérateur de prolongement du sous-domaine  $\Omega_i$  vers le domaine  $\Omega$ . On prend comme opérateur :

$$(P_i)_{j,k} = \frac{\int_{\Omega_i} \varphi_j \varphi_k^i}{\int_{\Omega} \varphi_j \varphi_j}$$

où  $\varphi_j$  est une fonction de base du domaine  $\Omega$  et  $\varphi_k^i$  une fonction de base du domaine  $\Omega_i$ . Par condensation de masse,  $P_i$  est une matrice diagonale. Sur tous les degrés de liberté du domaine  $\Omega_i$ , on a :

$$(P_i)_{j,j} = \frac{1}{\text{Nombre sous-domaines partageant le ddl } \text{glob}(j)}$$

où  $\text{glob}(j)$  est le numéro global (dans le domaine  $\Omega$ ) du degré de liberté  $j$  du domaine  $\Omega_i$ . En pratique, l'opérateur de prolongement ne fait que renommer les inconnues du domaine  $\Omega_i$  vers le domaine  $\Omega$ , sauf sur la frontière du domaine  $\Omega_i$ , où il faut en plus pondérer la valeur des inconnues. Cette pondération réalise une moyenne de  $u$  pour les degrés de liberté partagés par plusieurs sous-domaines. On obtient ainsi une propriété de conservation de la masse. En effet, si on choisit

$$u_i = 1 \quad \text{sur tous les sous-domaines}$$

On a alors :

$$u = \sum_{i=1}^{N_s} P_i u_i = 1 \quad \text{sur } \Omega$$

Pour découper le maillage  $\Omega$  en petits morceaux, on utilise un découpage en boîtes (cf. figure 2.13).

Sur le cas de la diffraction par un disque diélectrique, nous obtenons les résultats du tableau 2.13. Globalement, on a besoin de moins d'itérations qu'avec un préconditionneur utili-

Fréquence	0.25	0.5	1.0	2.0	3.0
2x2 sous-domaines	51	101	169	300	322
4x4 sous-domaines	132	311	500	922	1002
8x8 sous-domaines	251	651	1125	1982	2088

TAB. 2.13 – Nombre d'itérations du BICGCR préconditionné par la méthode de Schwarz additive.

sant l'équation de Helmholtz amortie. C'est d'autant plus vrai que le nombre de sous-domaines

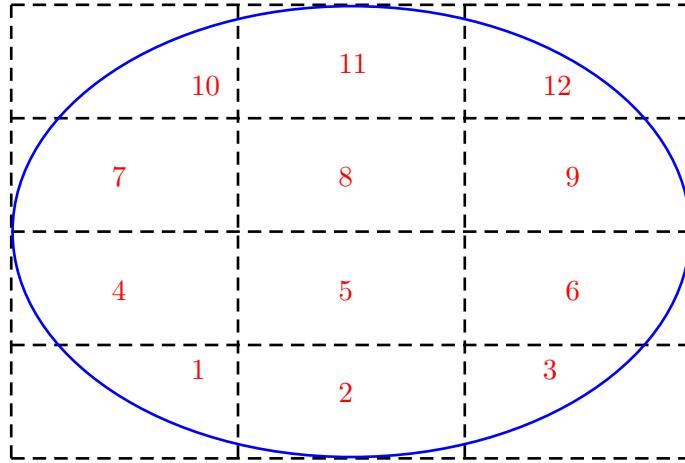


FIG. 2.13 – Découpage en boîtes d'un domaine de calcul. Chaque numéro est associé à un sous-domaine, intersection du petit rectangle avec le domaine de calcul.

n'est pas trop important. La dépendance vis-à-vis de la fréquence semble moins forte. Lorsqu'on choisit une fréquence 3 contre 1, on passe de 169 itérations à 322, alors qu'avec l'équation de Helmholtz amortie, on passait de 600 itérations à 2700 !

Sur le tableau 2.14, on met en évidence le gain en stockage, sur le cas 3-D de la diffraction d'une sphère. On utilise du  $Q_4$ , et pour chaque sous-domaine, on factorise la matrice associée au sous-mailage  $Q_1$ . Comme on l'a vu antérieurement, on peut éventuellement effectuer une factorisation incomplète du moment qu'on choisit un seuil assez petit ( $\varepsilon = 1e - 3$  donne des résultats corrects). Le cas avec 1x1x1 domaine est en vérité le préconditionnement par la matrice associée au sous maillage  $Q_1$ , c'est le cas d'un seul sous-domaine... Dans ce cas, on retrouve les limitations de la factorisation incomplète lorsqu'on ne met pas d'amortissement. En effet, le nombre d'itérations est de 202 contre 45 itérations si on utilise un solveur direct.

Décomposition en	1x1x1	2x2x2	4x4x4	8x8x8
Nombre d'itérations avec solveur direct	45	66	137	224
Nombre d'itérations avec factorisation incomplète	202	124	142	224
Mémoire	366 Mo	207 Mo	101 Mo	64 Mo

TAB. 2.14 – Nombre d'itérations du BICGCR préconditionné par la méthode de Schwarz additive. Pour factoriser les matrices éléments finis de chaque sous-domaine, on utilise soit un solveur direct, soit une factorisation incomplète. Cas de la sphère.

Cet effet est moins crucial lorsqu'on décompose en sous-domaines, bien qu'il soit encore présent. Sur ce cas test, la factorisation incomplète avec amortissement est moins gourmande en mémoire (37 Mo contre 64 Mo) et en temps de calcul. Sur le cas hétérogène, il sera parfois plus intéressant de faire de la décomposition de domaines.

Pour améliorer les performances lorsqu'il y a beaucoup de sous-domaines, on peut faire du recouvrement et utiliser une correction sur maillage grossier, [Cai *et al.*, 1998]. Nous excluons cette technique sur un code séquentiel. En effet le recouvrement induit des matrices bien plus importantes (notamment lorsqu'on utilise du  $Q_5$  par exemple), où est le gain par rapport à un solveur direct ?

### 2.4.3 Solveur itératif?

On a présenté des résultats numériques utilisant parfois du COCG, parfois du GMRES(20), parfois du BICGCR. On peut se demander quel solveur itératif est le plus judicieux, une fois le système préconditionné. Nous mettons en compétition plusieurs algorithmes : BICGCR, GMRES(20), QCGS, TFQMR et BICGSTAB. QCGS est une variante “lisse” du CGS (Conjugate Gradient Squared), il est décrit dans [Tong, 1992]. BICGCR s’applique exclusivement aux matrices complexes symétriques tandis que les autres algorithmes s’appliquent à des matrices quelconques. De plus, ils ne nécessitent pas de connaître la transposée de la matrice, contrairement au BICG et QMR. Le cas test est la diffraction d’une sphère diélectrique (cf. chapitre 3 pour les détails du cas test). On fait varier la fréquence et on utilise un préconditionneur multigrille. Les résultats sont consignés sur le tableau 2.15. Sur ce cas test, on ne modifie pas le

Fréquence	BICGCR	QCGS	GMRES(20)	TFQMR	BICGSTAB
0.125	34	42	<b>30</b>	40	40
0.25	<b>72</b>	94	99	100	123
0.5	<b>145</b>	182	288	185	215
1.0	<b>383</b>	544	719	577	975

TAB. 2.15 – Nombre d’itérations en fonction de la fréquence et pour divers solveurs

maillage, le maillage est “adapté” à la fréquence 1, il est trop fin pour les autres fréquences. On voit que pour une fréquence peu élevée, GMRES(20) demande très peu d’itérations. Lorsqu’on augmente la fréquence, il devient beaucoup moins intéressant que le BICGCR. Ces résultats soulignent bien que la prise en compte de la symétrie de la matrice mènent à un gain substantiel du nombre d’itérations. Si on est amené à manipuler des préconditionneurs non-symétriques, il semble plus judicieux d’utiliser QCGS ou TFQMR. Notons enfin, une complexité linéaire du nombre d’itérations en fonction de la fréquence. Lorsqu’on multiplie par deux la fréquence, on multiplie également par deux le nombre d’itérations. Dans la suite, on utilisera plutôt le BICGCR.

Comme on l’a signalé auparavant, une itération de chaque algorithme représente un produit matrice-vecteur et une application du préconditionneur. Typiquement QCGS est donné dans la littérature avec 2 produits matrice vecteur et 2 applications du préconditionneur par itération. Nous multiplions en conséquence par deux le nombre d’itérations pour coller à notre règle. Or, le coût du produit matrice-vecteur et du préconditionneur est majoritaire (il représente plus de 90 % du coût total) par rapport aux additions et produit scalaires calculés dans les algorithmes itératifs. Le temps de calcul est donc proportionnel au nombre d’itérations quel que soit l’algorithme itératif utilisé.

## 2.5 Conclusion

Dans ce chapitre, nous avons décrit un algorithme rapide pour effectuer le produit matrice-vecteur, pour la discréétisation choisie dans le chapitre 1. On a comparé la complexité de l’algorithme standard avec cet algorithme rapide. On a trouvé qu’il était intéressant pour des ordres d’approximation supérieurs à 3, que ce soit en 2-D ou en 3-D.

Nous avons ensuite observé qu’un solveur direct était satisfaisant pour l’ensemble des problèmes rencontrés en 2-D. En 3-D, le solveur direct est trop gourmand en mémoire, il faut avoir recours à un solveur itératif. En l’absence de préconditionneur, le solveur itératif est très coûteux car le nombre d’itérations augmente lorsqu’on raffine le maillage, ou si on augmente la fréquence ou lorsqu’on a des maillages de qualité médiocre.

Parmi les préconditionneurs proposés dans la littérature, notre choix s'est arrêté sur la factorisation incomplète, le multigrille et la décomposition en sous-domaines. Pour les deux premiers, on doit ajouter de l'amortissement à l'équation de Helmholtz, afin qu'ils soient stables. La factorisation incomplète limite toutefois le nombre de degrés de liberté qu'on peut traiter, car le coût en mémoire de ce préconditionneur est très important. Le multigrille n'a pas cet inconvénient, le principal défaut du multigrille est la relative "faiblesse" du lisseur utilisé (Jacobi par points).

La décomposition en sous-domaines est très coûteuse en espace mémoire, elle n'est intéressante à utiliser qu'avec une résolution parallèle. En effet, lorsqu'il y a peu de sous-domaines, le nombre d'itérations est assez réduit.

## Chapitre 3

# Comparaison hexaèdres / tétraèdres

*Nous nous proposons dans ce chapitre d'établir des comparaisons entre les éléments finis hexaédriques qu'on a présentés et des éléments finis tétraédriques utilisés classiquement. Nous faisons en premier lieu une étude de dispersion sur des maillages périodiques. Nous ferons une comparaison sur le cas académique de la sphère. Finalement, nous traiterons des cas réalistes 3-D, utilisant les préconditionneurs présentés dans le chapitre 2.*

### Sommaire

---

<b>3.1</b>	<b>Analyse de dispersion</b>	<b>76</b>
<b>3.2</b>	<b>Cas académique de la sphère</b>	<b>84</b>
3.2.1	Coût du produit matrice-vecteur	84
3.2.2	Comparaison sur une sphère parfaitement conductrice	85
3.2.3	Comparaison sur une sphère diélectrique	88
<b>3.3</b>	<b>Résultats numériques sur des cas plus complexes</b>	<b>88</b>
3.3.1	Cavité cobra	89
3.3.2	Cone-sphère revêtu	91
<b>3.4</b>	<b>Conclusion</b>	<b>92</b>

---

### 3.1 Analyse de dispersion

On considère un milieu homogène infini maillé par un motif élémentaire, le motif le plus simple étant un cube de taille  $h$ . Pour simplifier, on choisit  $\rho = 1$   $\mu = 1$ , on a ainsi une vitesse de propagation égale à 1. L'analyse de dispersion consiste à étudier des solutions ondes planes du schéma discret.

$$U(x) = U_0 \exp(i\vec{k} \cdot x)$$

Les solutions non-triviales du schéma discret fournissent alors une relation entre la pulsation  $\omega$  et le nombre d'onde  $k = |\vec{k}|$ . Cette relation est connue sous le nom de relation de dispersion discrète. Elle approche la relation de dispersion continue :

$$\omega^2 = k^2$$

#### Cas 2-D

Prenons par exemple le cas d'un triangle  $P2$ . On considère le maillage de la figure 3.1. Les quatre premiers degrés de liberté de ce maillage sont indépendants. Tous les autres degrés de liberté satisfont une relation de périodicité. Typiquement, nous avons les relations suivantes :

$$u_5 = u_1 \exp(ik_x h)$$

$$u_6 = u_3 \exp(ik_x h)$$

$$u_7 = u_1 \exp(i\frac{3}{2}k_x h + i\frac{\sqrt{3}}{2}k_y h)$$

On considère les quatre premières lignes de la matrice éléments finis associée à ce maillage. Pour toutes les colonnes  $j > 4$ , on utilise la relation de périodicité que satisfait l'inconnue  $u_j$ . On élimine ainsi toutes les inconnues  $u_j$ ,  $j > 4$ . On obtient, un problèmes aux valeurs propres (4x4) de la forme :

$$-\omega^2 h^2 \tilde{D}_h U_h + \tilde{K}_h U_h = 0$$

Une valeur propre correspond à la relation de dispersion continue, elle satisfait un développement du type :

$$\omega^2 h^2 = k^2 h^2 + c(kh)^{p+2} + o((kh)^{p+2})$$

$p$  est appelé l'ordre de l'erreur de dispersion. La constante  $c$  dépend de la direction du vecteur d'onde. En 2-D, on a ainsi une dépendance suivant l'angle  $\theta$  de l'onde plane. Les autres valeurs propres correspondent à des ondes "parasites", leur vitesse de propagation tend vers l'infini, quand le pas de maillage  $h$  tend vers 0.

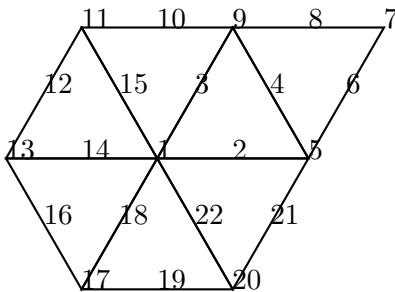


FIG. 3.1 – Maillage périodique P2

On introduit les variables :

$$\xi_1 = k h \cos \theta$$

$$\xi_2 = k h \sin \theta$$

Les éléments finis quadrilatéraux **sans** condensation de masse (en utilisant les formules de Gauss exactes pour  $2r+1$ ) vérifient les relations discrètes suivantes (nous ne mettons que les termes prépondérants du développement) :

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 + \frac{\xi_1^4}{12} + \frac{\xi_2^4}{12} \quad \text{pour } r = 1$$

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 + \frac{\xi_1^6}{720} + \frac{\xi_2^6}{720} \quad \text{pour } r = 2$$

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 + \frac{\xi_1^8}{100800} + \frac{\xi_2^8}{100800} \quad \text{pour } r = 3$$

Les relations discrètes sont établies pour tous les ordres d'approximation dans [Ainsworth, 2004a]. Dans la thèse de S. Fauqueux, les calculs sont détaillés pour les éléments quadrilatéraux **avec** condensation de masse, nous faisons figurer ici les résultats :

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 - \frac{\xi_1^4}{12} - \frac{\xi_2^4}{12} \quad \text{pour } r = 1$$

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 - \frac{\xi_1^6}{1440} - \frac{\xi_2^6}{1440} \quad \text{pour } r = 2$$

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 - \frac{\xi_1^8}{302400} - \frac{\xi_2^8}{302400} \quad \text{pour } r = 3$$

Comme on peut le constater, la condensation de masse sur les quadrilatères/hexaèdres diminue la dispersion par  $r$ . Une opinion commune est que la condensation de masse est une méthode de sous-intégration, car on n'intègre pas exactement la formulation variationnelle. On pourrait penser que l'approximation ainsi faite fasse perdre de la précision. Il n'en est rien, on gagne même de la précision, du moins en ce qui concerne l'erreur de dispersion. On remarquera que la condensation de masse inverse le signe de l'erreur de dispersion. C'est une propriété qu'on retrouve également pour les triangles/tétraèdres [Elmkies, 1998]. On peut moyenner les deux méthodes (avec condensation et sans condensation) :

$$A_h = \alpha A_h^{\text{avec condensation}} + (1 - \alpha) A_h^{\text{sans condensation}}$$

On choisit :

$$\alpha = \frac{r}{r+1}$$

On obtient les relations discrètes suivantes :

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 - \frac{\xi_1^6}{240} - \frac{\xi_1^4 \xi_2^2}{144} - \frac{\xi_1^2 \xi_2^4}{144} - \frac{\xi_2^6}{240} \quad \text{pour } r = 1$$

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 - \frac{\xi_1^8}{37800} - \frac{\xi_2^8}{37800} \quad \text{pour } r = 2$$

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 - \frac{\xi_1^{10}}{15876000} - \frac{\xi_2^{10}}{15876000} \quad \text{pour } r = 3$$

On remarque qu'on gagne ainsi deux ordres dans l'erreur de dispersion. Cette technique tient plus du gadget que de l'efficacité, car ce gain n'est vérifié que sur des maillages réguliers.

On calcule des relations similaires pour les triangles avec le motif périodique de la figure 3.1

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 + \frac{\xi_1^4}{16} + \frac{\xi_1^2 \xi_2^2}{8} + \frac{\xi_2^4}{16} \quad \text{pour } r = 1$$

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 + \frac{\xi_1^6}{1280} + \frac{\xi_1^4 \xi_2^2}{640} + \frac{11 \xi_1^2 \xi_2^4}{3840} + \frac{\xi_2^6}{1440} \quad \text{pour } r = 2$$

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 + \frac{19 \xi_1^8}{5017600} + \frac{23 \xi_1^6 \xi_2^2}{2257920} + \frac{1427 \xi_1^4 \xi_2^4}{67737600} + \frac{1819 \xi_1^2 \xi_2^6}{101606400} + \frac{263 \xi_2^8}{81285120} \quad \text{pour } r = 3$$

Il est difficile d'interpréter de telles relations, et comme je ne suis pas un fanatique des courbes de dispersion, j'ai préféré introduire une constante de dispersion " $L^2$ " :

$$\text{constante}L^2 = \left( \frac{1}{2\pi} \int_0^{2\pi} c(\theta)^2 d\theta \right)^{1/2}$$

où  $c(\theta)$  est la constante qui apparaît dans le développement limité de la relation de dispersion :

$$\frac{\omega^2}{k^2} = 1 + c(\theta)(k h)^p + o((k h)^p)$$

On moyenne ainsi la dispersion pour tous les angles d'incidence de l'onde plane. Nous voulons aussi comparer la dispersion entre les différents ordres d'approximation. Pour ce faire, on introduit une variable adimensionnelle :

$$K = \frac{6 k h}{2\pi \tilde{r}}$$

Où  $\tilde{r}$  est défini en fonction du nombre de degrés indépendants du maillage :

$$\tilde{r} = \sqrt{\frac{\text{Nombre de degrés de liberté indépendants}}{\text{Aire du motif élémentaire}}}$$

La valeur  $K = 1$  correspond ainsi à un maillage contenant exactement 6 degrés de liberté par longueur d'onde, quel que soit l'ordre d'approximation. Afin d'avoir une constante  $L^2$  raisonnable (pas minuscule quand on monte en ordre), on utilisera plutôt cette définition de  $c(\theta)$  :

$$\frac{\omega^2}{k^2} = 1 + c(\theta)K^p + o(K^p)$$

Ainsi  $\text{constante}L^2$  représentera directement l'erreur de dispersion moyenne quand on utilise un maillage avec 6 points par longueur d'onde !

Les résultats sont notés dans les tableaux 3.2 et 3.1. L'ordre de l'erreur de dispersion est égal à  $2r$  pour les éléments finis quadrilatéraux et triangulaires.

Ordre d'approximation	$Q_1$	$Q_2$	$Q_3$	$Q_4$	$Q_5$	$Q_6$	$Q_7$
Sans condensation de masse	6.76e-2	1.80e-2	5.98e-3	2.22e-3	8.74e-4	3.60e-4	1.52e-4
Avec condensation de masse	6.76e-2	8.98e-3	1.99e-3	5.54e-4	1.74e-4	6.00e-5	2.18e-5

TAB. 3.1 – Constante  $L^2$  de l'erreur de dispersion pour les éléments finis quadrilatéraux

On voit ainsi qu'avec 6 points par longueur d'onde, il est intéressant de monter en ordre quel que soit l'élément fini considéré. On notera que ce sont les éléments quadrilatéraux avec condensation de masse, qui donnent la dispersion la plus petite. Ils sont talonnés de près par les éléments finis triangulaires sur un maillage parfaitement régulier. La qualité du maillage

Ordre d'approximation	P1	P2	P3	P4	P5	P6	P7
Triangles droits	1.02e-1	3.26e-2	1.17e-2	6.14e-3	3.42e-3	2.31e-3	1.57e-3
Triangles équilatéraux	6.33e-2	1.45e-2	3.81e-3	1.28e-3	4.43e-4	1.74e-4	6.967e-5

TAB. 3.2 – Constante  $L^2$  de l'erreur de dispersion pour les éléments finis triangulaires

triangulaire a une certaine influence, pour tous les ordres d'approximation. En tenant compte de l'ordre de l'erreur de dispersion, on peut calculer le nombre de points par longueur d'onde nécessaire pour le maillage “triangles droits”, afin d'obtenir la même erreur de dispersion que le maillage “triangles équilatéraux”. Ainsi pour P1 on a besoin de 7.6 points par longueur d'onde, pour P7 on a besoin de 7.5 points par longueur d'onde, contre 6 points par longueur d'onde pour le maillage triangles équilatéraux.

De la même manière, on peut quantifier l'apport de la condensation de masse sur les éléments finis quadrilatéraux. Les éléments sans condensation de masse nécessitent 7.13 points par longueur d'onde pour  $Q_1$ , et 6.89 points par longueur d'onde pour  $Q_7$ ! Cet écart peut sembler minime, mais en 3-D, ça se traduit par une augmentation de 50% du nombre de degrés de liberté en  $Q_7$ ! Nous nous sommes également intéressés à savoir si cet avantage de condenser la masse était conservé sur des maillages déformés. Nous avons pris le maillage de la figure 3.2, et nous avons évalué les constantes de dispersion  $L^2$  (cf. tableau 3.3). L'ordre de l'erreur de dispersion est toujours égal à  $2r$  avec ou sans condensation de masse.

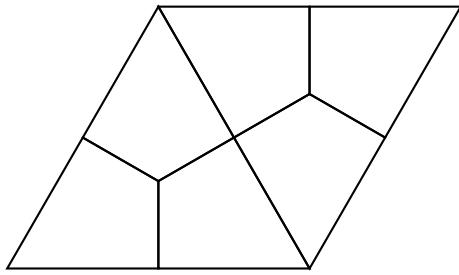


FIG. 3.2 – Motif élémentaire de maillage quadrilatéral déformé

Ordre d'approximation	$Q_1$	$Q_2$	$Q_3$	$Q_4$	$Q_5$	$Q_6$
Sans condensation de masse	7.74e-2	2.57e-2	1.00e-2	5.07e-3	2.77e-3	1.67e-3
Avec condensation de masse	1.19e-1	8.29e-3	1.91e-3	3.39e-4	1.06e-4	6.93e-5

TAB. 3.3 – Constante  $L^2$  de l'erreur de dispersion sur un maillage fortement déformé

On voit que la condensation de masse détériore les résultats pour  $Q_1$ , mais les améliore pour les autres ordres d'approximation. L'amélioration semble même plus importante que dans le cas régulier. Les constantes obtenues sont à un poil près les mêmes que dans le cas du maillage régulier. On a remarqué que les triangles “droits” découpés (au lieu des triangles équilatéraux) donnaient des constantes bien plus élevées.

### Cas 3-D

Pour ce qui est des hexaèdres réguliers, les conclusions sont identiques au cas 2-D, du fait de la tensorisation du maillage. La condensation de masse fournit toujours un gain de précision. La constante  $L^2$  de l'erreur de dispersion, fait intervenir une intégrale mettant en jeu deux angles

$\theta, \phi :$

$$\text{constante} L^2 = \left( \frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi (c(\theta, \varphi))^2 \sin \theta d\theta d\varphi \right)^{1/2}$$

La grandeur  $\tilde{r}$  est alors égale à :

$$\tilde{r} = \sqrt[3]{\frac{\text{Nombre de degrés de liberté indépendants}}{\text{Volume du motif élémentaire}}}$$

Nous utilisons toujours la variable  $K$  :

$$K = \frac{6k h}{2\pi \tilde{r}}$$

Nous faisons une étude de dispersion, dans le cas du maillage de la figure 3.3, dont on découpe chaque tétraèdre en 4 hexaèdres. Du fait de la complexité des calculs, il est difficile de calculer la constante  $L^2$  pour des ordres supérieurs ou égaux à 3 sur ce maillage. Nous avons choisi de n'évaluer cette constante de l'erreur de dispersion que pour une seule direction du vecteur d'onde, la direction  $(1, 0, 0)$ . On a mis dans le tableau 3.4, le terme principal de l'erreur de dispersion. Comme en 2-D, les éléments avec condensation de masse dispersent plus pour  $Q_1$ , mais ont une constante plus petite que les éléments sans condensation de masse pour les ordres supérieurs. On notera que la condensation de masse fait perdre deux ordres de dispersion pour  $r \geq 2$ , l'ordre de l'erreur de dispersion est de  $2(r-1)$  au lieu de  $2r$ , comme on pouvait s'attendre. Ce n'est pas seulement la déformation du maillage qui est la cause de cette perte de précision, mais également la structure du maillage. En effet, si on prend le maillage hexaédrique de la figure 3.4, on a un ordre  $2r$  pour l'erreur de dispersion !

Ordre	$Q_1$	$Q_2$	$Q_3$	$Q_4$
Sans condensation de masse	$0.108 K^2$	$0.0447 K^4$	$0.0215 K^6$	$0.0144 K^8$
Avec condensation de masse	$0.227 K^2$	$0.00898 K^2$	$0.00392 K^4$	$0.00229 K^6$

TAB. 3.4 – Erreur de dispersion pour  $k = (1, 0, 0)$

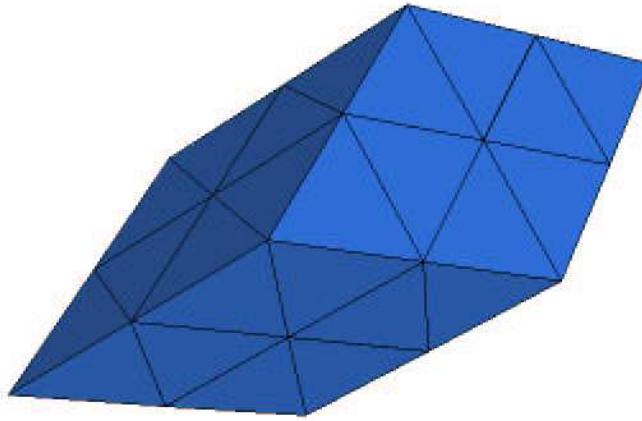


FIG. 3.3 – Maillage tétraédrique dit “régulier”

Nous comparons les erreurs de dispersion des éléments finis hexaédriques sur maillage régulier avec les éléments finis tétraédriques, sur deux maillages (cf. fig 3.5, 3.3). Le second maillage est dit “régulier”, il n'en est rien. Pour cause, il est impossible de remplir l'espace 3-D

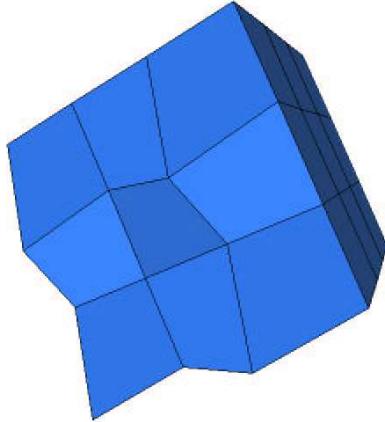


FIG. 3.4 – Maillage hexaédrique légèrement modifié

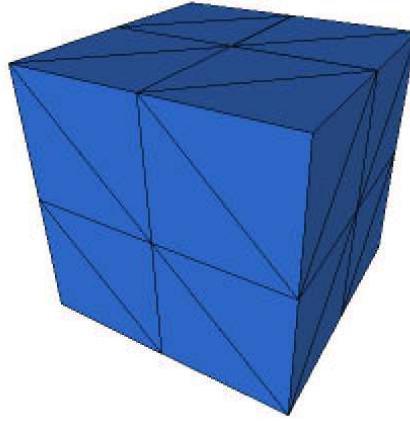


FIG. 3.5 – Maillage “tétraèdres droits”

avec des tétraèdres réguliers car l’angle formé par deux faces est égal à 70.53 degrés, qui ne divise pas 360... Le second maillage est néanmoins plus régulier que le premier (le ratio plus grande arête/ plus petite arête est moins élevé). Le tableau comparatif 3.5 montre nettement que les éléments finis hexaédriques sur maillage régulier sont moins dispersifs. L’ordre de l’erreur de dispersion est pour tous ces configurations de  $2r$ .

Ordre	1	2	3	4	5
Tétraèdres droits	0.168	5.51e-2	2.05e-2	1.27e-2	8.41e-3
Tétraèdres “réguliers”	0.1082	2.85e-2	8.59e-3	4.24e-3	2.30e-3
Hexaèdres réguliers	0.0653	8.27e-3	1.81e-3	5e-4	1.57e-4

TAB. 3.5 – Constante  $L^2$  de l’erreur de dispersion pour les tétraèdres et hexaèdres (avec condensation de masse)

On veut illustrer les effets de la condensation de masse sur le problème “concret” de recherche

de valeurs propres :

$$\left\{ \begin{array}{l} \text{trouver } (\omega, u) \in \mathbb{R} \times H^1(\Omega) \quad u \neq 0 \quad \text{tel que} \\ -\omega^2 u - \Delta u = 0 \quad \in \Omega \\ u = 0 \quad \in \partial\Omega \end{array} \right.$$

On connaît les modes propres du cube  $[0, 1]^3$  avec la condition de Dirichlet :

$$\omega_{m,n} = \pi \sqrt{l^2 + m^2 + n^2} \quad u = \sin(\pi l x) \sin(\pi m x) \sin(\pi n x) \quad l, m, n \in \mathbb{N}^*$$

On étudie la convergence de la valeur propre (1,1,1) et du mode propre associé sur les figures 3.6, 3.7 et 3.8, pour un maillage hexaédrique régulier. On voit sur la première figure, que les éléments avec condensation de masse donnent une valeur propre plus précise que les éléments sans condensation de masse. En revanche, c'est le contraire pour les vecteurs propres, qui sont mieux approchés avec les éléments sans condensation de masse. Les valeurs propres ont une convergence en  $O(h^{2r})$  tandis que les vecteurs propres convergent en  $O(h^r)$  en norme  $H^1$  et en  $O(h^{r+1})$  en norme  $L^2$ . Lorsqu'on prend un maillage non-régulier, les valeurs propres ont une convergence en  $O(h^{2r-2})$  pour  $r \geq 2$  si on utilise de la condensation de masse, ce qu'on peut voir sur la figure 3.9. Sur ce type de maillage, les éléments  $Q_1$  avec condensation de masse semblent donner des valeurs propres plus précises que les éléments  $Q_1$  sans condensation de masse. Notons enfin que pour obtenir 1% d'erreur sur la valeur propre à l'aide d'une approximation  $Q_1$ , on a besoin de 5 fois plus de degrés de liberté en maillage non-régulier, ce qui est assez désappointant... On effectuera d'autres comparaisons entre maillage régulier/non-régulier dans la section suivante.

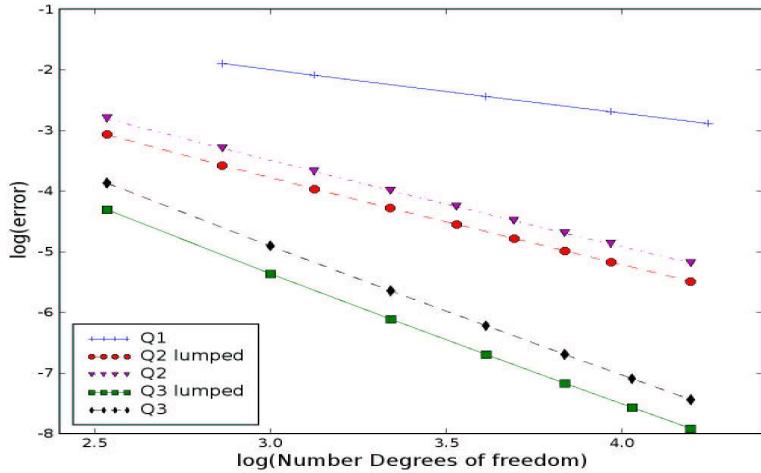


FIG. 3.6 – Convergence de la valeur propre numérique vers la valeur propre analytique en fonction du nombre de ddl. Echelle log-log, maillages réguliers.

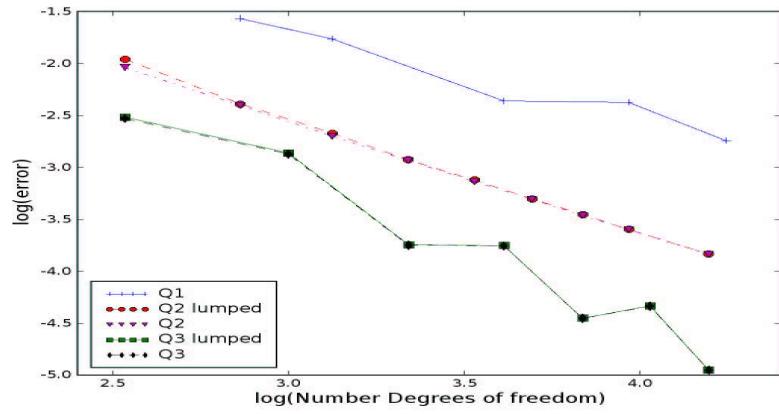


FIG. 3.7 – Convergence du mode numérique vers le mode analytique en fonction du nombre de ddl. Echelle log-log, norme  $L^2$ , maillages réguliers.

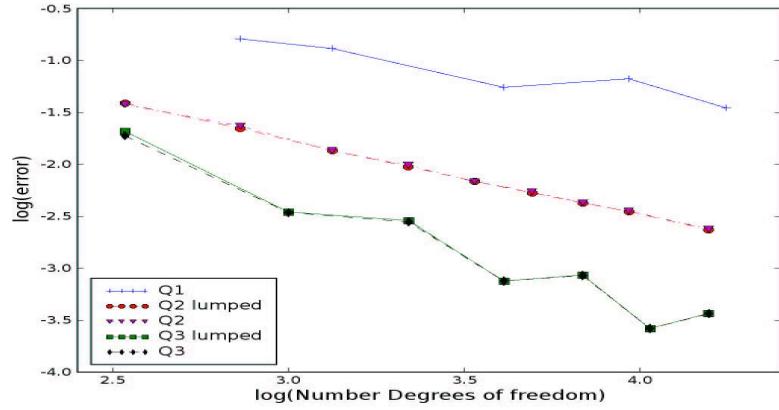


FIG. 3.8 – Convergence du mode numérique vers le mode analytique en fonction du nombre de ddl. Echelle log-log, norme  $H^1$ , maillages réguliers

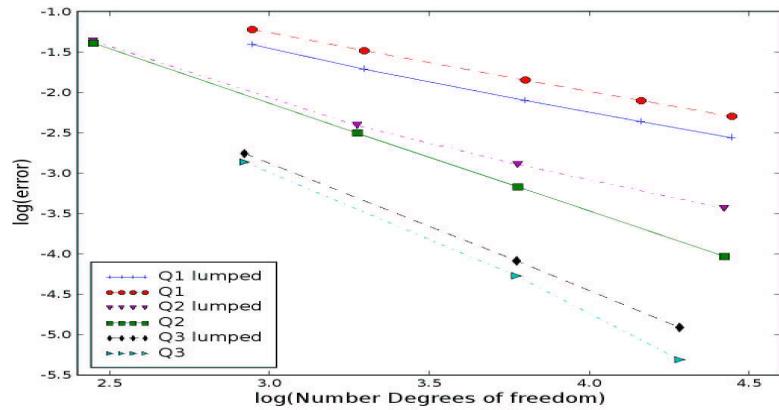


FIG. 3.9 – Convergence de la valeur propre numérique vers la valeur propre analytique en fonction du nombre de ddl. Echelle log-log, maillage “tétraèdres découpés”.

## 3.2 Cas académique de la sphère

En préambule, nous précisons que dans la suite du chapitre, nous utiliserons exclusivement BICGCR comme algorithme itératif de résolution. Comme on l'a mentionné dans le chapitre 2, c'est l'algorithme le plus robuste, aussi robuste que QMR, mais plus efficace que ce dernier. Nous ne montrons aucun résultat numérique réalisé avec du  $P_1$  ou du  $Q_1$ . Pour cause, ces deux éléments sont bien trop dispersifs, et sur des cas de grande taille, il faut souvent mettre 50 points par longueur d'onde pour obtenir un résultat potable. Ils ont également le désavantage de nécessiter des maillages très volumineux, qui prennent tout l'espace mémoire ! On exclut leur utilisation d'emblée.

### 3.2.1 Coût du produit matrice-vecteur

Sur maillage tétraédrique “régulier”, on effectue des calculs de complexité du coût produit matrice-vecteur pour les éléments tétraédriques. On dénombre les interactions suivantes de la matrice éléments finis :

- Nombre d'interactions par sommet :

$$15 + 50(r-1) + 60(r-1)(r-2)/2 + 24(r-1)(r-2)(r-3)/6$$

- Nombre d'interactions par ddl d'une arête :

$$8 + 18(r-1) + 12(r-1)(r-2)/2 + 6(r-1)(r-2)(r-3)/6$$

- Nombre d'interactions par ddl d'une face :

$$5 + 8(r-1) + 6(r-1)(r-2)/2 + 2(r-1)(r-2)(r-3)/6$$

- Nombre d'interactions par ddl intérieur :

$$(r+1)(r+2)(r+3)/6$$

Pour chaque interaction de la matrice, on compte deux opérations (une multiplication et une addition), on obtient alors les complexités suivantes :

$$\text{Opérations : } 1/3 r^6 + 4r^5 + 28/3 r^4 - 2r^3 + 122r^2 - 174r + 70$$

$$\text{Stockage utilisé : Nombre opérations}/4 + 0.5r^3$$

Pour le stockage, on a pris en compte la symétrie (on divise par deux le nombre d'interactions et on rajoute la moitié de la diagonale). Lorsqu'on divise par le nombre de degrés de liberté, afin de comparer à nombre de ddl constant, on obtient la figure 3.10.

On a également validé ce calcul théorique par des expériences numériques sur le cas de la

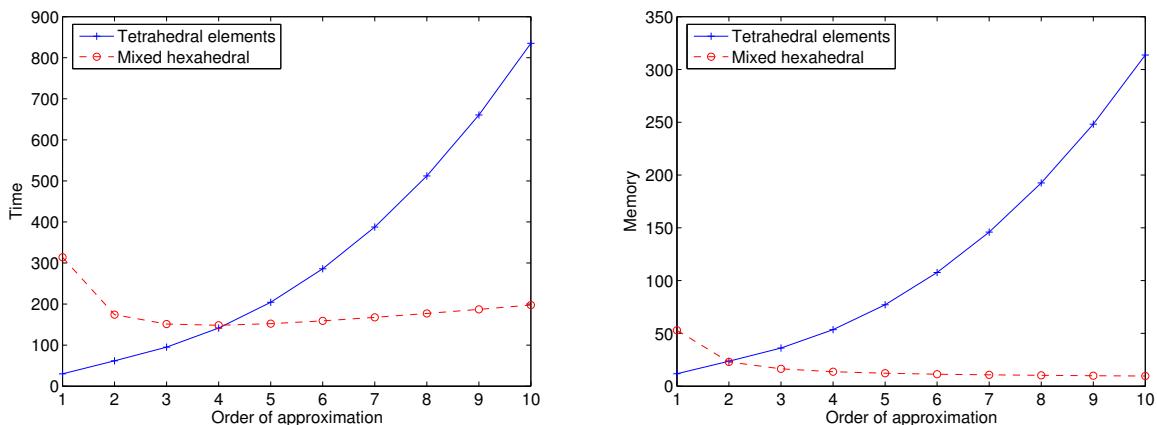


FIG. 3.10 – A gauche, temps de calcul pour les tétraèdres et pour les hexaèdres (formulation mixte). A droite, stockage pour les deux éléments.

sphère, cf. figure 3.11. Les résultats théoriques concordent bien avec l'expérience numérique.

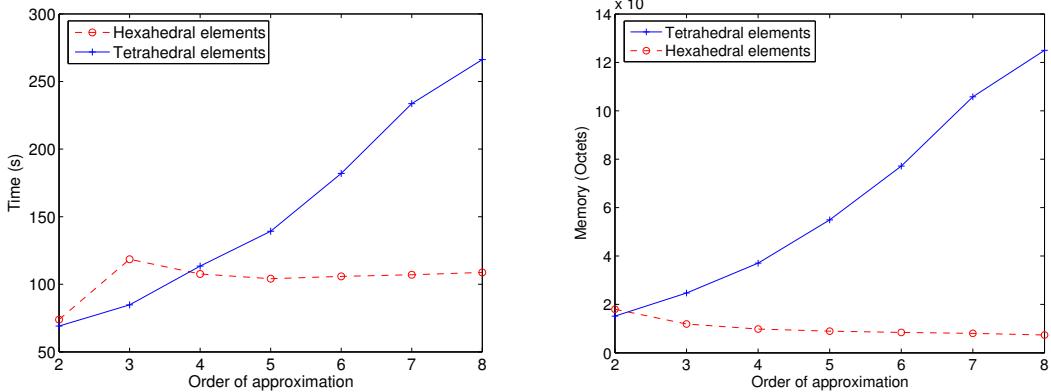


FIG. 3.11 – A gauche, temps de calcul pour les tétraèdres et pour les hexaèdres (formulation mixte). A droite, stockage pour les deux éléments. Le temps de calcul est mesuré sur 1000 itérations de COCG pour un cas de 100 000 ddl (le maillage est différent pour chaque ordre, afin d'avoir le même nombre de ddl). Pour  $Q_2$ , on utilise la formulation standard

Les tétraèdres sont plus rapides pour  $P_1$ ,  $P_2$  et  $P_3$ , ils sont plus lents pour les ordres supérieurs. Au niveau de l'espace mémoire, la formulation mixte apporte un gain dès l'ordre 2. En pratique, on n'ira jamais au-delà de  $P_4$ , car alors l'espace mémoire requis est trop important. Pour  $Q_1$  et  $Q_2$ , on utilisera exclusivement la formulation standard, pour les ordres supérieurs on utilisera la formulation mixte. On remarquera que numériquement  $Q_5$  est plus rapide que  $Q_4$ , alors que la théorie donne  $Q_4$  gagnant. Nous n'avons pas d'explications de cette différence minime...

### 3.2.2 Comparaison sur une sphère parfaitement conductrice

On considère une sphère de rayon  $a = 4$ , la condition absorbante est placée sur une sphère extérieure de rayon  $b = 6$ . On impose une condition de Dirichlet sur la sphère intérieure. On a représenté le champ diffracté solution sur la figure 3.12.

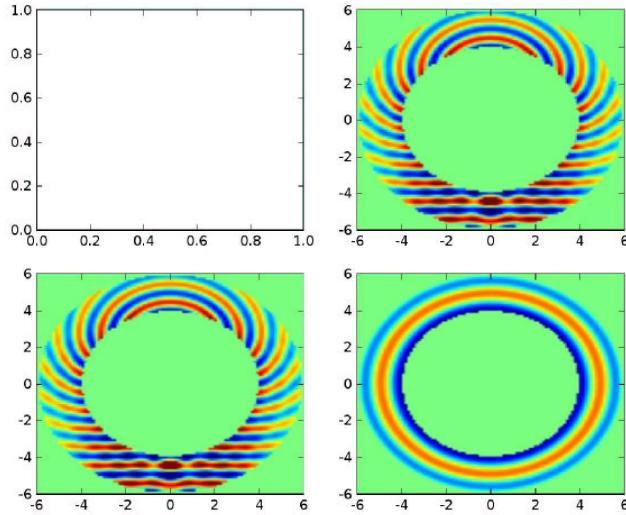


FIG. 3.12 – Partie réelle du champ diffracté par une sphère parfaitement conductrice de diamètre 8 longueurs d'onde

Pour mailler le domaine en tétraèdres, on utilise gmsh, un mailleur tétraédrique gratuit. On découpe ensuite chaque tétraèdres en 4 hexaèdres, ce qui donne le maillage de la figure 3.13.

On compare dans un premier temps la convergence des tétraèdres classiques à la convergence des hexaèdres - avec condensation de masse - sur des maillages tétraédriques découpés. La convergence de ces éléments est montrée sur la figure 3.14. En abscisse, on fait figurer, le

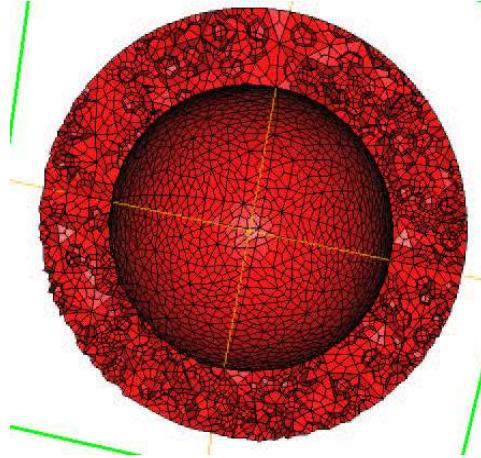


FIG. 3.13 – Exemple de maillage hexaédrique utilisé

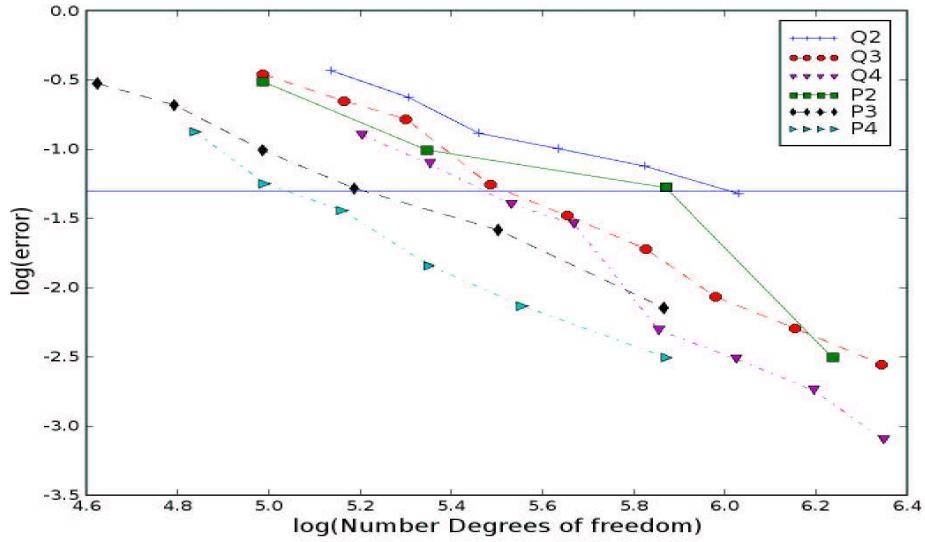


FIG. 3.14 – Evolution de l'erreur  $L^2$  en fonction du nombre de degrés de liberté. Échelle log-log, cas de la sphère parfaitement conductrice sur des maillages tétraédriques découpés

nombre de degrés de libertés en échelle logarithmique. En ordonnées, on représente l'erreur relative par rapport à la solution analytique, en échelle logarithmique aussi. On voit que plus on monte en ordre, plus la convergence est rapide, et on a besoin de moins de degrés de liberté pour obtenir une solution précise. On a tracé une ligne horizontale en trait plein, qui symbolise les 5% d'erreur. On peut alors collecter le nombre de degrés de liberté nécessaires pour avoir une erreur relative d'environ 5%, ce qui est fait sur la figure 3.15. Sur cette figure, on affiche le nombre de degrés de liberté nécessaires pour atteindre cette erreur, pour trois types d'éléments :

- Éléments tétraédriques classiques
- Éléments hexaédriques sur maillage “pseudo-régulier”
- Éléments hexaédriques sur maillage “tétraèdres découpés”

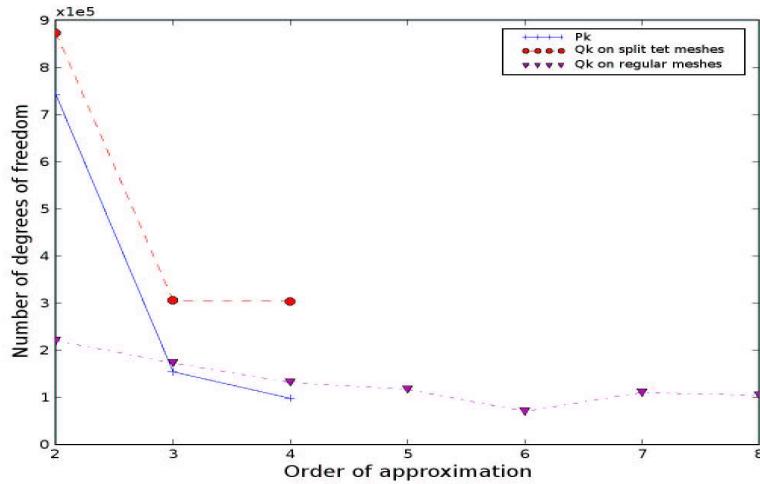


FIG. 3.15 – Nombre de degrés de liberté nécessaire pour atteindre une erreur de 5% pour diverses approximations.

On peut observer qu'on divise par 7 le nombre de degrés de liberté quand on utilise du  $P_4$  au lieu du  $P_2$ . La technique des tétraèdres découpés est moins performante que les purs tétraèdres, on a besoin de 300 000 degrés de liberté pour  $Q_4$  contre moins de 100 000 pour du  $P_4$ . Même en utilisant des hexaèdres réguliers, il faut utiliser au minimum du  $Q_6$  pour nécessiter moins de degrés de liberté. Le lecteur se pose sûrement la question suivante : “Pourquoi on peut pas utiliser du  $Q_7$  sur des tétraèdres découpés ?”. La raison est purement technique en vérité. En effet, lorsqu'on demande un maillage très grossier en tétraèdres, le meilleur fournit des tétraèdres assez plats. Par malheur, lorsqu'on courbe ces tétraèdres, il peut arriver qu'une arête intérieure du tétraèdre traverse la face courbe du tétraèdre !! Inutile de préciser qu'on dégrade complètement la solution, le conditionnement de la matrice ... C'est relativement aisément détecter, car sur ces éléments dégénérés, le jacobien change de signe. Un exemple d'élément dégénéré est illustré sur la figure 3.16. Pour obtenir cette figure, on a découpé le tétraèdre suivant la carte locale  $F_i$ . Il faut se rendre compte que pour obtenir 100 000 degrés de liberté en découplant des

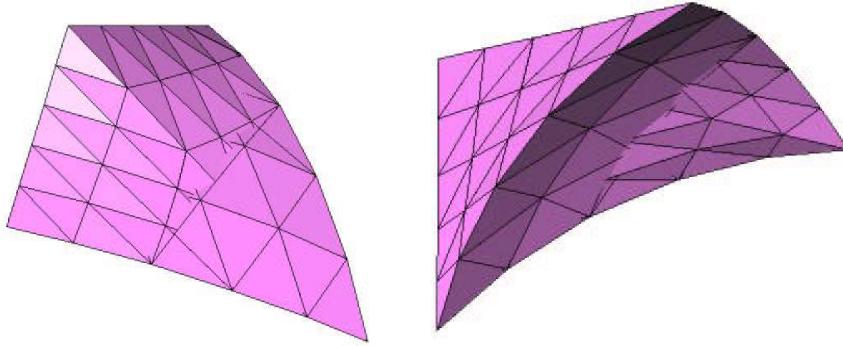


FIG. 3.16 – Tétraèdre dégénéré, vue de dessus, et vue de dessous

tétraèdres avec du  $Q_7$ , il faut environ 60 tétraèdres. C'est un vrai défi de mailler ce domaine avec aussi peu de tétraèdres en ayant en outre l'exigence que le maillage soit sympathique... Pour le maillage régulier hexaédrique, il se pose deux problèmes. Le premier c'est que si on met 8 points par longueur d'onde sur la sphère intérieure, on a 5 points par longueur d'onde sur la sphère extérieure, à cause du ratio 2/3 entre les deux rayons ! Le second problème, c'est qu'il est difficile d'obtenir pile 5% d'erreur pour  $Q_7$  et  $Q_8$ , un coup le maillage est trop grossier, un coup il est bien trop fin. Sur le tableau 3.6, on a mis les temps de résolution avec ou sans préconditionnement pour ces divers éléments. Quand on met  $Q_2$  régulier, c'est  $Q_2$  utilisé avec un maillage quasi-régulier de la sphère. Lorsqu'on met  $Q_2$  déstructuré, c'est  $Q_2$  utilisé avec un maillage tétraédrique découpé. Le préconditionneur multigrille et la factorisation incomplète ILUT(1e-2) sont deux préconditionneurs basés sur l'équation de Helmholtz amortie, ils sont détaillés dans le chapitre 2. Comme on peut le voir sur ce tableau, la factorisation incomplète

Élément fini	Nombre ddl	Temps sans préconditionneur	Multigrille	ILUT(1e-2)
$Q_2$ régulier	220 000	244s	157s	42s
$Q_4$ régulier	132 000	242s	120s	28s
$Q_6$ régulier	<b>70 200</b>	<b>150s</b>	<b>80s</b>	<b>15s</b>
$Q_2$ déstructuré	874 000	3090s	703s	876s
$Q_4$ déstructuré	304 000	1298s	305s	99s
$P_2$	743 000	1672s	250s	653s
$P_4$	98 000	195s	105s	25s

TAB. 3.6 – Performances sur la sphère parfaitement conductrice

donne les résultats les plus intéressants, sauf sur des cas exigeant une place mémoire trop importante. On peut conseiller l'utilisation de ce préconditionneur pour des cas de moins de 500 000 ddl, au-delà seul le préconditionneur multigrille n'est pas coûteux en espace mémoire. La décomposition en sous-domaines est une technique trop coûteuse en espace mémoire.

### 3.2.3 Comparaison sur une sphère diélectrique

Le problème modèle est la diffraction d'une sphère diélectrique :

$$\rho = 4, \mu = 1 \quad \text{pour } r \geq a$$

$$\rho = 1, \mu = 1 \quad \text{pour } r > a$$

Le rayon de la sphère diélectrique est pris égal à 2. Une condition absorbante d'ordre 1 est imposée sur une sphère extérieure de rayon 3. Le champ diffracté obtenu est représenté sur la figure 3.17. Le maillage hexaédrique régulier utilisé est affiché sur la figure 3.18. On notera qu'avec ce maillage, on réalise presque naturellement l'adaptation du pas de maillage à la longueur d'onde ! En revanche, un meilleur tétraédrique ne réalisera pas cette condition, ce qui donne des performances meilleures pour les hexaèdres sur ce type de maillage, comme on peut le voir sur le tableau 3.7. Mais la technique des tétraèdres découpés est moins performante encore une fois ! Pour P2, il est nécessaire d'avoir au moins un million de ddl, on n'a pas jugé utile de mettre les stats du P2...

## 3.3 Résultats numériques sur des cas plus complexes

On n'a pas la prétention de traiter des cas réalistes, mais de traiter d'autres géométries que la sphère, pour lesquelles on ne connaît pas de solution analytique. On traite le cas d'une

Élément fini	Nombre ddl	Temps sans préconditionneur	Multigrille	ILUT(1e-2)
$Q_2$ régulier	220 000	1 391s	535s	324s
$Q_4$ régulier	85 000	<b>708s</b>	185s	91s
$Q_6$ régulier	<b>78 000</b>	787s	<b>165s</b>	<b>77s</b>
$Q_4$ déstructuré	243 000	5795s	729s	534s
$P_4$	180 000	1597s	695s	363s

TAB. 3.7 – Performances sur la sphère diélectrique

cavité cobra et d'un cone-sphère revêtu. Dans tous les cas, on utilise une condition absorbante d'ordre 1. Dans l'annexe B, on a introduit une condition transparente, qui s'obtient en itérant la résolution du problème éléments finis avec la condition absorbante d'ordre 1. Si on sait résoudre rapidement ce dernier problème, on saura résoudre rapidement le problème avec une condition transparente.

### 3.3.1 Cavité cobra

C'est le cas d'une cavité (cf. figure 3.19), dont on a ouvert l'extrémité droite, en imposant une condition absorbante d'ordre 1. Sur les autres parois de la cavité, on impose une condition de Neumann. Les dimensions de cette cavité sont environ 20 longueurs d'onde sur 4 longueurs d'onde. Le champ diffracté solution est représenté sur la figure 3.20. Cette solution de référence a été obtenue sur un maillage hexaédrique, en prenant une approximation  $Q_8$  avec un million de ddl. Comme dans le cas de la sphère, on détermine le nombre de ddl nécessaires pour atteindre 5 % d'erreur avec une approximation  $P_4$ ,  $Q_4$ ... On récapitule les résultats obtenus sur la cavité cobra dans le tableau 3.8. On utilise plutôt un algorithme 2-grille, l'algorithme multigrille étant moins performant sur ce cas-là. En pratique, au vu de l'espace mémoire requis par la factorisation incomplète, il nous semble inutile d'aller au-delà de 3-grilles. En effet, en prenant trois grilles, la taille mémoire utilisée pour factoriser la matrice de la grille grossière est négligeable devant la mémoire nécessaire pour stocker la matrice de la grille fine ainsi que les vecteurs d'itération.

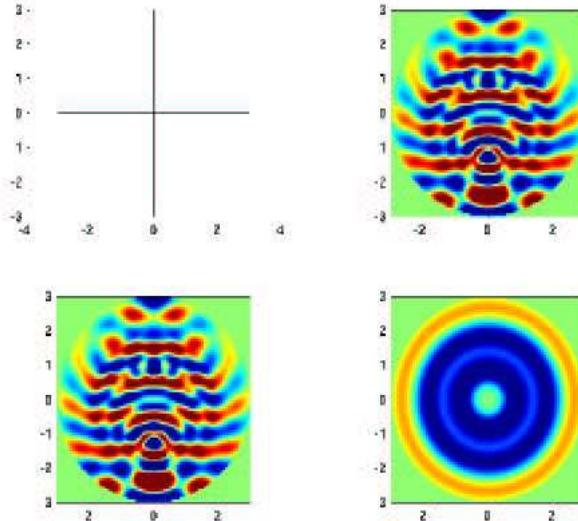


FIG. 3.17 – Partie réelle du champ diffracté par une sphère diélectrique

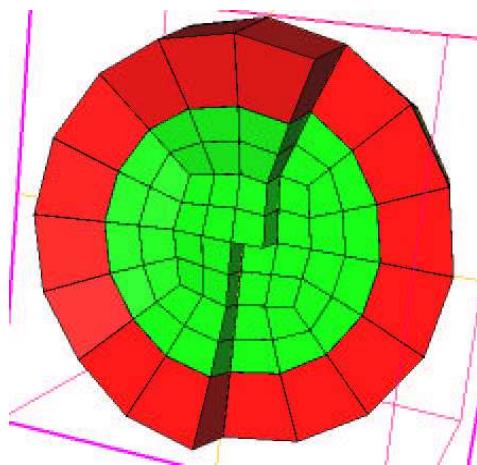


FIG. 3.18 – Maillage hexaédrique quasi-régulier d'une sphère diélectrique

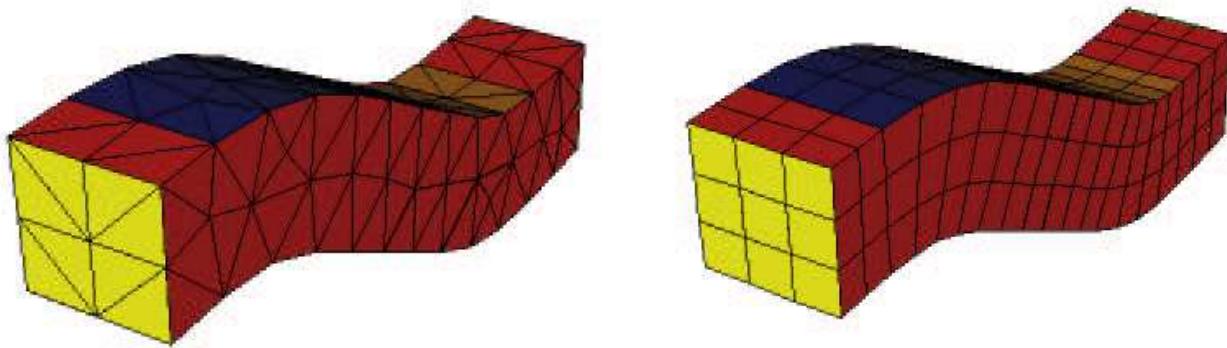


FIG. 3.19 – Maillage de la cavité en tétraèdres à gauche, en hexaèdres à droite

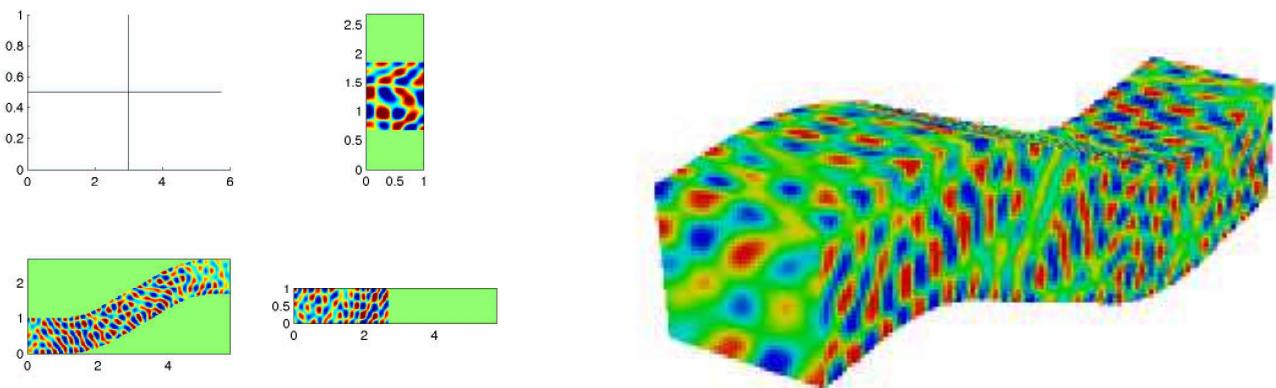


FIG. 3.20 – Partie réelle du champ diffracté par la cavité cobra. A gauche, coupe à l'intérieur de la cavité, à droite valeur sur la surface.

Ce sont la matrice et les vecteurs qui coûtent chers en espace mémoire, limitant les cas 3-D à 5 millions de ddl (2 Go de ram). La formulation mixte donne à cette occasion entièrement satisfaction, car une approximation P4 nécessite beaucoup plus de stockage.

Sur ce cas, on arrive à produire un maillage tétraédrique suffisamment grossier (cf. fi-

Élément fini	Nombre ddl	Temps sans préconditionneur	2-grille	ILUT(1e-2)
$Q_4$ régulier	330 000	29 863s	3 589s	6 920s
$Q_6$ régulier	185 000	17 272s	1 916s	1 564s
$Q_8$ régulier	<b>95 600</b>	<b>9 860s</b>	<b>1 082s</b>	<b>1 021s</b>
$Q_4$ déstructuré	567 400	NC	7 009s	17 947s
$Q_6$ déstructuré	466 700	NC	6 821s	13 766s
$P_4$	358 900	NC	14 016s	8 036s

TAB. 3.8 – Performances sur la cavité cobra

gure 3.19) pour utiliser  $Q_6$  sur des tétraèdres découpés. Néanmoins, le nombre de degrés de liberté est au minimum doublé lorsqu'on passe d'un maillage structuré à un maillage non-structuré.  $P_4$  nécessite moins de ddl que  $Q_4$  et  $Q_6$  non-structurés, mais il est plus lent à l'utilisation avec un préconditionneur 2-grilles. Enfin, tous les feux sont au vert pour  $Q_8$  structuré. En effet sur ce cas, l'erreur de dispersion joue un rôle beaucoup plus important que sur les autres cas rencontrés. Les éléments ayant une erreur de dispersion très faible sont avantageux. C'est pour cette raison, qu'il est inutile de vouloir faire du  $P_2$  ou du  $Q_2$  sur ce cas, ils ont besoin d'un nombre trop important de ddl.

### 3.3.2 Cone-sphère revêtu

On traite le cas du cone-sphère de la figure 3.21. On prend des indices complexes pour la

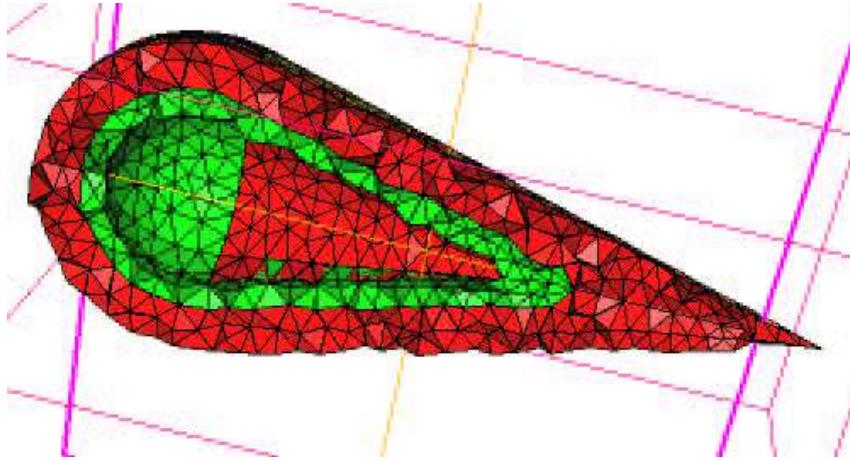


FIG. 3.21 – Maillage du cone-sphère en tétraèdres. En vert, revêtement, en rouge le domaine extérieur (vide).

partie diélectrique :

$$\rho = 3 + 0.5i \quad \mu = 0.5 - 0.5i$$

La partie imaginaire (positive pour  $\rho$  et négative pour  $\mu$ ) correspond à un amortissement. Comme on peut le voir sur la figure 3.22, la solution s'atténue fortement au fur et à mesure qu'elle pénètre dans la partie diélectrique. L'amortissement introduit rend ce cas extrêmement bien conditionné, comme en témoignent les résultats du tableau 3.9. Le rayon de la sphère est de 2, contre une épaisseur de 0.8 du revêtement. A cause de cette épaisseur relativement faible, il nous est impossible de créer des maillages tétraédriques non-dégénérés avec un pas de maillage strictement supérieur à 0.51. Avec ce pas de maillage, les éléments droits fournissent une erreur de 5 %, quel que soit l'ordre d'approximation. Par conséquent, on ne peut espérer prendre un

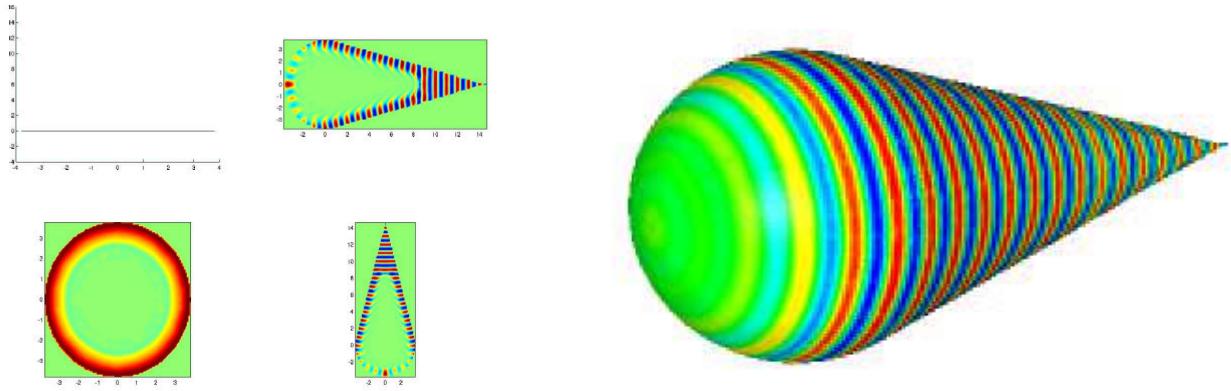


FIG. 3.22 – Partie réelle du champ total pour un cone sphère. A gauche, coupe intérieure, à droite valeur du champ sur la surface extérieure.

pas de maillage plus grand, utiliser des éléments droits et obtenir une erreur de 5 % ! C'est ce qui est reflété dans le tableau, la géométrie nous impose le maillage, il est plus intéressant dans ce cas d'utiliser du  $Q_2$  non-structuré que du  $Q_4$  non-structuré... On a quasiment le même souci pour les tétraèdres (en moins problématique). P4 donne avec ce maillage 1.7% d'erreur si on utilise les éléments courbes. Mais on ne peut pas prendre un pas de maillage plus grand pour obtenir une erreur de 5%, car on récupère des tétraèdres dégénérés. Sur ce type de cas, P2

Élément fini	Nombre ddl	Temps sans préconditionneur	Multigrille	ILUT(1e-2)
$Q_2$ déstructuré	494 000	1 787s	274s	370s
$Q_4$ déstructuré	3 838 000	42 200s	1 426s	-
$P_2$	178 000	<b>193s</b>	<b>24s</b>	<b>21s</b>
$P_4$	<b>166 000</b>	516s	107s	27s

TAB. 3.9 – Performances sur le cone-sphère revêtu

l'emporte pour les raisons invoquées.

### 3.4 Conclusion

Dans ce chapitre, nous avons montré qu'en 2-D, la condensation de masse diminuait l'erreur de dispersion, en maillage régulier et non-régulier. La comparaison avec les triangles se révèle favorables aux quadrillatères. En 3-D, la condensation de masse donne une erreur de dispersion en  $O(h^{2r-2})$  au lieu de  $O(h^{2r})$ , sur des maillages tétraédriques découpés. Toutefois, sur le cas régulier, Les hexaèdres sont moins dispersifs que les tétraèdres.

Des comparaisons, tant au niveau théorique que numérique, ont montré que le produit matrice vecteur des hexaèdres était plus rapide que pour les tétraèdres à partir de l'ordre 4. En ce qui concerne la mémoire, on observe un gain dès l'ordre 2.

L'ensemble des résultats numériques montrent que les hexaèdres sont plus performants que les tétraèdres avec ou sans préconditionneur, pourvu qu'on dispose d'un maillage hexaédrique "régulier". La technique des tétraèdres découpés donne des résultats assez décevants. Sur certains cas favorables comme la cavité cobra (grande zone à mailler), elle est relativement performante par rapport aux tétraèdres. Elle permet d'étudier des cas de grande taille car les

hexaèdres sont beaucoup moins coûteux en mémoire. Sur des cas défavorables, comme le cone-sphère (couche de petite épaisseur), il est préférable d' utiliser les tétraèdres. Les tétraèdres sont difficiles à courber, ce qui donne souvent lieu à des tétraèdres dégénérés lorsque le maillage est trop grossier. Cette contrainte pénalise d'autant plus la technique des tétraèdres découpés. Il nous semble que les hexaèdres sont plus faciles à courber, il y a un intérêt non-négligeable à disposer de meilleurs hexaédriques non-structurés.



## **Deuxième partie**

### **Equations de Maxwell 2-D et 3-D**



# Chapitre 4

## Seconde famille de Nédélec sur les quadrillatères

Ce chapitre présente la discréétisation des équations de Maxwell 2-D par les éléments finis quadrillatéraux utilisant la seconde famille de Nédélec comme espace d'approximation. La première section montre comment on aboutit à un produit matrice-vecteur rapide en réalisant la condensation de masse. La seconde section met en valeur les modes parasites qui polluent la solution numérique.

### Sommaire

---

<b>4.1</b>	<b>Formulation mixte des équations de Maxwell . . . . .</b>	<b>98</b>
4.1.1	Formulation variationnelle standard . . . . .	98
4.1.2	Expression des matrices élémentaires . . . . .	100
4.1.3	Une formulation variationnelle mixte possible . . . . .	101
4.1.4	Intérêt de la formulation variationnelle mixte . . . . .	102
<b>4.2</b>	<b>Pollution de la solution : modes parasites . . . . .</b>	<b>105</b>
4.2.1	Etude de convergence . . . . .	105
4.2.2	Etude de la distribution des valeurs propres . . . . .	109
<b>4.3</b>	<b>Conclusion . . . . .</b>	<b>114</b>

---

Afin d'approfondir ce chapitre, on pourra consulter

- COHEN, G. (2002). *Higher-order numerical methods for transient wave equations*. Springer Verlag,
- BOSSAVIT, A. (1998). *Computational electromagnetism*. Academic Press (Boston),
- MONK, P. (2002). *Finite elements methods for Maxwell's equations*. Oxford Science Publication, 2002,
- JIN, J. (1993). *The finite element method in electromagnetics*. John Wiley and Sons Inc,
- NÉDÉLEC, J. C. (1980). Mixed finite elements in  $\mathbb{R}^3$ . *Numer. Math.*, 35(3):315–341,
- NÉDÉLEC, J. C. (1986). A new family of mixed finite elements in  $\mathbb{R}^3$ . *Numer. Math.*, 51(1):57–81

Pourquoi commençons-nous par la seconde famille, avant la première ? La raison est purement historique. En effet, seule la seconde famille de Nédélec sur les quadrilatères conduit à la condensation de masse. On peut étendre de manière simple la formulation mixte pour les éléments  $H^1$  vers une formulation mixte pour les éléments H-rot en utilisant la seconde famille de Nédélec. Cette extension est détaillée dans [Pernet, 2004]. Il nous est paru donc plus logique de privilégier la seconde famille. La suite nous a montré l'étendue de notre erreur ! On rappellera brièvement comment on réalise cette extension, et les bonnes propriétés qui en découlent. On conclura sur un inconvénient majeur de ce choix de discrétisation, à savoir la présence d'ondes parasites.

## 4.1 Formulation mixte des équations de Maxwell

### 4.1.1 Formulation variationnelle standard

On suppose comme dans la première partie une dépendance harmonique du champ électrique et magnétique :

$$\mathbf{E}(x, t) = \mathbf{E}(x) e^{-i\omega t}$$

$$H(x, t) = H(x) e^{-i\omega t}$$

On se place dans le cas transverse magnétique, avec  $\mathbf{E}$  vecteur du plan, et  $H$  scalaire ( on a  $\mathbf{H} = H\mathbf{z}$  ). On définit un rotationnel scalaire et un rotationnel vectoriel :

$$\text{rot } \mathbf{v} = \frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}$$

$$\mathbf{rot } u = \left( \frac{\partial u}{\partial y}, -\frac{\partial u}{\partial x} \right)$$

Le champ électrique  $\mathbf{E}(x)$  et le champ magnétique sont solutions des équations :

$$\begin{cases} -i\omega \varepsilon \mathbf{E} - \mathbf{rot } H = \mathbf{f} \\ -i\omega \mu H + \text{rot } \mathbf{E} = 0 \end{cases}$$

On effectue un changement de variable, afin de se ramener à un système plus sympathique.

$$\varepsilon = \varepsilon_0 \varepsilon_r$$

$$\mu = \mu_0 \mu_r$$

$$k_0^2 = \frac{\omega^2}{c_0^2} = \omega^2 \mu_0 \varepsilon_0$$

$$\bar{\mathbf{E}} = \frac{\mathbf{E}}{\varepsilon_0}$$

$$\bar{H} = -i\omega \mu_0 H$$

On obtient ainsi le système :

$$\begin{cases} -k_0^2 \varepsilon_r \bar{\mathbf{E}} - \mathbf{rot } \bar{H} = \mathbf{f} \\ \mu_r \bar{H} + \text{rot } \bar{\mathbf{E}} = 0 \end{cases}$$

Par abus de notation, on omettra la barre sur  $\bar{\mathbf{E}}$  et  $\bar{H}$ , on utilisera  $\omega$  alors qu'il s'agit de  $k_0$ . On parlera également de  $\varepsilon$ ,  $\mu$ , alors qu'on devrait dire  $\varepsilon_r$  et  $\mu_r$ . Avec ces abus de notations, on a le système suivant :

$$\begin{cases} -\omega^2 \varepsilon \mathbf{E} - \mathbf{rot} H = \mathbf{f} \\ \mu H + \mathbf{rot} \mathbf{E} = 0 \end{cases}$$

Le milieu environnant est le vide, où :

$$\varepsilon = 1 \quad \mu = 1$$

On peut éventuellement éliminer  $H$  et obtenir une équation en  $\mathbf{E}$  :

$$-\omega^2 \varepsilon \mathbf{E} + \mathbf{rot} \left( \frac{1}{\mu} \mathbf{rot} \mathbf{E} \right) = \mathbf{f}$$

On établit la formulation variationnelle, dite formulation standard, des équations de Maxwell 2-D :

$$-\omega^2 \int_{\Omega} \varepsilon \mathbf{E} \cdot \boldsymbol{\varphi} + \int_{\Omega} \frac{1}{\mu} \mathbf{rot} \mathbf{E} \mathbf{rot} \boldsymbol{\varphi} + \int_{\partial\Omega} \frac{1}{\mu} \mathbf{rot} \mathbf{E} \boldsymbol{\varphi} \times \mathbf{n} = - \int_{\Omega} \mathbf{f} \cdot \boldsymbol{\varphi} \quad (4.1)$$

On choisit  $\mathbf{E}, \boldsymbol{\varphi}$  dans l'espace :

$$H(\text{rot}, \Omega) = \{ \mathbf{u} \in (L^2(\Omega))^2 \text{ tel que } \mathbf{rot} \mathbf{u} \in L^2(\Omega) \}$$

Le terme de bord est égal à :

$$\int_{\partial\Omega} \frac{1}{\mu} \mathbf{rot} \mathbf{E} \boldsymbol{\varphi} \times \mathbf{n} = - \int_{\partial\Omega} H \boldsymbol{\varphi} \times \mathbf{n}$$

Si on impose une condition de conducteur parfait,

$$\mathbf{E} \times \mathbf{n} = 0$$

le terme de bord devient nul, car on impose  $\boldsymbol{\varphi} \times \mathbf{n} = 0$  dans l'espace de discréétisation. Si on impose une condition de Silver-Müller :

$$H - i\omega \sqrt{\frac{\varepsilon}{\mu}} \mathbf{E} \times \mathbf{n} = 0$$

on obtient le terme de bord suivant :

$$-i\omega \int_{\partial\Omega} \sqrt{\frac{\varepsilon}{\mu}} \mathbf{E} \times \mathbf{n} \boldsymbol{\varphi} \times \mathbf{n}$$

Le facteur  $-i\omega$  apparaît à cause du changement de variable qu'on a introduit.

On choisit comme espace de discréétisation :

$$V_h = \{ \mathbf{u} \in H(\text{rot}, \Omega) \text{ tel que } DF_i^t \mathbf{u} \circ F_i \in Q_r^2 \}$$

où  $r$  est l'ordre d'approximation. Pour passer d'une fonction définie sur le carré unité  $\hat{K}$  vers une fonction définie sur un quadrilatère, on utilise donc la formule :

$$\boldsymbol{\varphi}_j \circ F_i(\hat{x}) = DF_i^{-t}(\hat{x}) \hat{\boldsymbol{\varphi}}_j(\hat{x})$$

La transformation  $DF_i^{-t}$  est dite H-rot conforme. Si une fonction de base  $\hat{\boldsymbol{\varphi}}$  définie sur  $\hat{K}$  appartient à  $H(\text{rot}, \hat{K}_1 \cup \hat{K}_2)$ , alors la fonction de base  $\boldsymbol{\varphi}$  construite à l'aide de cette transformation appartiendra à  $H(\text{rot}, K_i \cup K_j)$ . La démonstration de cette propriété revient à démontrer que la transformation  $DF_i^{-t}$  conserve la continuité de la trace tangentielle. Cette démonstration est faite dans l'annexe de [Cohen, 2002].

Les degrés de libertés sont pris, comme dans le cas scalaire, aux points de Gauss-Lobatto. On exige que les degrés de liberté tangentiels soient continus d'un élément à un autre. Cette continuité est vérifiée en attribuant un numéro global identique pour un degré de liberté tangentiel partagé par deux éléments, cf. figure 4.1.

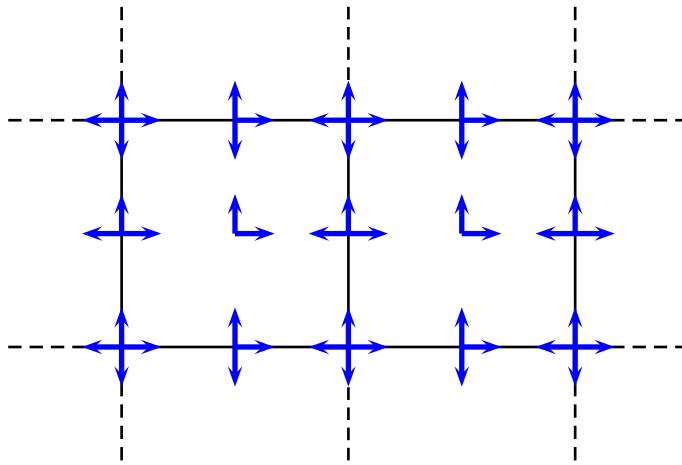


FIG. 4.1 – Degrés de liberté de la seconde famille sur les quadrilatères pour  $r = 2$

#### 4.1.2 Expression des matrices élémentaires

La matrice de masse élémentaire a pour terme générique :

$$(M_h)_{j,k} = \int_{K_i} \varepsilon \varphi_j \cdot \varphi_k dx$$

Après changement de variables, on obtient :

$$(M_h)_{j,k} = \int_{\hat{K}} J_i DF_i^{-1} \varepsilon DF_i^{-t} \hat{\varphi}_j \cdot \hat{\varphi}_k d\hat{x}$$

On obtient la condensation de masse en utilisant les points de Gauss-Lobatto comme points d'intégration :

$$(M_h)_{j,k} = \omega_j (J_i DF_i^{-1} \varepsilon DF_i^{-t})(\hat{\xi}_j) \delta_{j,k}$$

La matrice de masse élémentaire est diagonale par blocs 2x2. Lorsqu'on assemble cette matrice, on peut avoir des blocs de plus grande taille, du fait de la continuité des degrés de liberté tangentiels, chaque bloc étant associé à un point de quadrature du maillage. Comme on peut le voir sur la figure 4.1, on a des blocs 4x4 sur les sommets de ce maillage, des blocs 3x3 sur les points de quadrature placés sur les arêtes, des blocs 2x2 sur les points de quadrature intérieurs.

La matrice de rigidité élémentaire a pour terme générique :

$$(K_h)_{j,k} = \int_{K_i} \frac{1}{\mu} \operatorname{rot} \varphi_j \operatorname{rot} \varphi_k dx$$

Après changement de variables, on obtient :

$$(K_h)_{j,k} = \int_{\hat{K}} \frac{1}{\mu J_i} \operatorname{rot} \hat{\varphi}_j \operatorname{rot} \hat{\varphi}_k d\hat{x}$$

On distingue deux types de degrés de liberté :

$$\hat{\varphi}_{j_1,j_2}^1 = \hat{\varphi}_{j_1}(x_1) \hat{\varphi}_{j_2}(x_2) \mathbf{e}_1$$

$$\hat{\varphi}_{j_1,j_2}^2 = \hat{\varphi}_{j_1}(x_1) \hat{\varphi}_{j_2}(x_2) \mathbf{e}_2$$

En utilisant les points de Gauss-Lobatto, on obtient les expressions suivantes :

$$\begin{aligned}
\text{Interaction } (j_1, j_2, 1) \text{ et } (k_1, k_2, 1) & : \sum_{n=1}^{r+1} \frac{1}{\mu J_i(\hat{\xi}_{j_1}, \hat{\xi}_n)} \omega_{j_1, n} \hat{\varphi}'_{j_2}(\hat{\xi}_n) \hat{\varphi}'_{k_2}(\hat{\xi}_n) \delta_{j_1, k_1} \\
\text{Interaction } (j_1, j_2, 2) \text{ et } (k_1, k_2, 2) & : \sum_{n=1}^{r+1} \frac{1}{\mu J_i(\hat{\xi}_n, \hat{\xi}_{j_2})} \omega_{j_1, n} \hat{\varphi}'_{j_1}(\hat{\xi}_n) \hat{\varphi}'_{k_1}(\hat{\xi}_n) \delta_{j_2, k_2} \\
\text{Interaction } (j_1, j_2, 1) \text{ et } (k_1, k_2, 2) & : \frac{1}{\mu J_i(\hat{\xi}_{j_1}, \hat{\xi}_{k_2})} \omega_{j_1, k_2} \hat{\varphi}'_{j_2}(\hat{\xi}_{k_2}) \hat{\varphi}'_{k_1}(\hat{\xi}_{j_1})
\end{aligned}$$

On a une matrice de rigidité élémentairement creuse, mais le nombre d'éléments non-nuls tend asymptotiquement vers 50 % pour  $r$  tendant vers l'infini.

On a une propriété similaire pour le cas 3-D, les degrés de liberté orientés suivant  $\mathbf{e}_s$  ont une interaction non-nulle, si et seulement si

$$j_s = k_s$$

Les degrés de liberté orientés suivant  $\mathbf{e}_s$  ont une interaction non-nulle avec les degrés de liberté orientés suivant  $\mathbf{e}_t$  ( $s \neq t$ ) si et seulement si

$$j_1 = k_1 \text{ ou } j_2 = k_2 \text{ ou } j_3 = k_3$$

On retrouve la propriété énoncée dans le cas scalaire, les degrés de liberté n'interagissent qu'avec ceux situés sur un même plan. La matrice de rigidité est élémentairement creuse, bien plus creuse en 3-D qu'en 2-D.

#### 4.1.3 Une formulation variationnelle mixte possible

Elle s'obtient en conservant l'inconnue  $H$ . On effectue une intégration par parties, pour ne garder que le terme rotationnel s'appliquant sur des inconnues vectorielles :

$$\begin{aligned}
-\omega^2 \int_{\Omega} \varepsilon \mathbf{E} \cdot \boldsymbol{\varphi} - \int_{\Omega} H \operatorname{rot}(\boldsymbol{\varphi}) - \int_{\partial\Omega} H \boldsymbol{\varphi} \times \mathbf{n} &= \int_{\Omega} \mathbf{f} \cdot \boldsymbol{\varphi} \\
\int_{\Omega} \mu H \psi + \int_{\Omega} \operatorname{rot} \mathbf{E} \psi &= 0
\end{aligned}$$

Le terme de bord est identique à celui obtenu pour la formulation variationnelle standard (on a mis  $H$  au lieu de  $\operatorname{rot} \mathbf{E}$  ...). Il est donc traité de la même manière, on ne revient pas dessus. On choisit  $\mathbf{E}$  dans le même espace d'approximation discret,  $H$  est choisi dans l'espace d'approximation discret suivant :

$$W_h = \{u \in L^2(\Omega) \text{ tel que } u \circ F_i \in Q_r\}$$

Les degrés de liberté pour  $H$  sont pris aux points de Gauss-Lobatto, les fonctions de base sont identiques à celles introduites dans le cadre de l'équation de Helmholtz.

Après discréttisation, on obtient le système discret suivant :

$$\begin{cases} -\omega^2 M_h E_h - R_h H = F_h \\ D_h H_h + R_h^t E_h = 0 \end{cases}$$

avec

$M_h$  comme définie dans la section précédente

$$(D_h)_{i,j} = \int_{\Omega} \mu \psi_i \psi_j$$

$$(R_h)_{i,j} = \int_{\Omega} \text{rot}(\varphi_i) \psi_j$$

On élimine l'inconnue  $H_h$  :

$$-\omega_2 M_h E_h + R_h D_h^{-1} R_h^t E_h = F_h$$

On a ainsi remplacé la matrice  $K_h$  par  $R_h D_h^{-1} R_h^t$ . Il est possible de démontrer, si on choisit les points de Gauss-Lobatto comme points d'intégration, que :

$$K_h = R_h D_h^{-1} R_h^t$$

Cette démonstration est faite dans [Pernet, 2004].

#### 4.1.4 Intérêt de la formulation variationnelle mixte

Les avantages sont du même acabit que pour l'équation de Helmholtz. La matrice élémentaire est indépendante de la géométrie :

$$(R_h)_{j,k} = \int_{\hat{K}} \hat{\text{rot}}(\hat{\varphi}_j) \hat{\psi}_k d\hat{x}$$

Elle ne nécessite aucun stockage, de plus elle est élémentairement creuse. A chaque degré de liberté vectoriel, on ne compte que  $r+1$  degrés de liberté scalaires qui interagissent. Les matrices de masse  $M_h$  et  $D_h$  sont respectivement diagonale par blocs et diagonale.

On a par conséquent un coût du produit matrice vecteur en  $O(r^3)$  en 2-D et  $O(r^4)$  en 3-D, contre un coût en  $O(r^4)$  et  $O(r^5)$  si on utilise la formulation variationnelle standard. On fait un calcul de complexité comme dans le cas de l'équation de Helmholtz, sur un maillage régulier.

Nombre d'opérations formulation standard 2-D :  $(4r^4 + 20r^3 + 42r^2 + 42r + 16) N_e$

Nombre d'opérations formulation mixte 2-D :  $(8r^3 + 25r^2 + 22r + 5) N_e$

Nombre d'opérations formulation standard 3-D :  $(42r^5 + 162r^4 + 156r^3 - 6r^2 - 36r + 6) N_e$

Nombre d'opérations formulation mixte 3-D :  $(24r^4 + 132r^3 + 228r^2 + 162r + 42) N_e$

Pour faire ce calcul, on a considéré que la matrice issue de la formulation standard était assemblée ainsi que la matrice de masse  $M_h$  de la formulation mixte. Il est relativement aisés de retrouver le terme prépondérant de ces développements.

Pour la formulation standard 2-D, chaque degré de liberté vectoriel (il y en a  $2(r+1)^2$ ) interagit avec  $(r+1)^2$  degrés de liberté vectoriels, on a donc principalement  $2r^4$  interactions. Chaque interaction donne lieu à deux opérations (une multiplication et une addition), on retrouve ainsi le terme prépondérant  $4r^4$ .

Pour la formulation mixte 2-D, la matrice de rigidité élémentaire compte  $r+1$  interactions pour chaque degré de liberté vectoriel, soit principalement  $2r^3$  termes non-nuls dans cette matrice. La factorisation fait apparaître deux produits impliquant cette matrice, on a bien un terme prépondérant en  $8r^3$ .

Pour la formulation mixte 3-D, la matrice de rigidité élémentaire comptabilise  $2(r+1)$  interactions pour chaque degré de liberté vectoriel, on a donc un terme prépondérant en  $3r^3$ .

$2r * 4 = 24r^4$ . Notons au passage le caractère particulièrement creux de la matrice de rigidité élémentaire de la formulation mixte. Comparativement aux nombres d'éléments de la matrice, elle est plus creuse que la matrice obtenue pour l'équation de Helmholtz. Le taux de remplissage de la matrice est égal à :

$$\frac{\text{Nombre éléments non-nuls}}{\text{Nombre éléments}} = \frac{2}{3(r+1)^2}$$

alors que pour Helmholtz, il est égal à

$$\frac{\text{Nombre éléments non-nuls}}{\text{Nombre éléments}} = \frac{1}{(r+1)^2}$$

Pour le nombre d'interactions de la formulation standard, il suffit de diviser par 4 le nombre d'opérations. Pour la formulation mixte, on ne stocke que des matrices de masse. En 2-D, on stocke quatre coefficients par point de quadrature (trois pour une matrice bloc 2x2 symétrique et un coefficient pour la matrice diagonale). En 3-D, on stocke douze coefficients par point de quadrature (six pour chaque matrice bloc 3x3 symétrique). Aux matrices de masse, on rajoute la numérotation du maillage, on obtient :

Stockage formulation mixte 2-D :  $5(r+1)^2 N_e$

Stockage formulation mixte 3-D :  $13.5(r+1)^3 N_e$

On veut comparer à nombre de degrés de liberté égal, il est donc nécessaire d'estimer le nombre de degrés de liberté :

Nombre de ddl en 2-D :  $2(r+1)^2 - 2(r+1)$

Nombre de ddl en 3-D :  $3(r+1)^3 - 9(r+1) - 3(2(r-1)^2 + 4(r-1))$

Sur les figures 4.2 et 4.3, on a mis les temps de calculs (fictifs) en fonction de l'ordre d'approximation, ainsi que l'espace mémoire requis. On trouve une différence notable par rapport à Helmholtz, la formulation mixte est plus efficace que la formulation standard pour tous les ordres d'approximation, on ne conserve donc que la formulation mixte. Ce comportement se

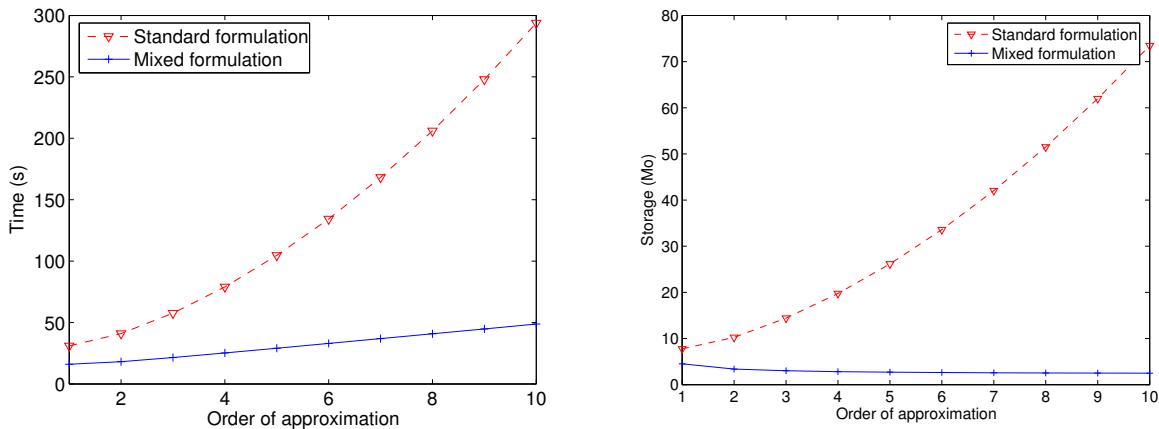


FIG. 4.2 – A gauche temps de calcul en fonction de l'ordre d'approximation, à droite stockage. Nombre ddl constant.

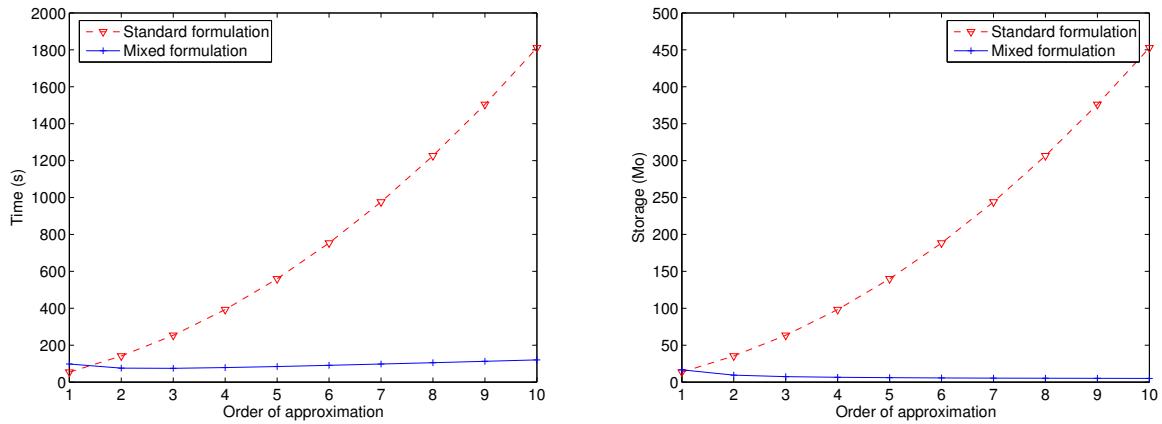


FIG. 4.3 – A gauche temps de calcul en fonction de l’ordre d’approximation, à droite stockage. Nombre ddl constant.

comprend relativement aisément. En 2-D, on a finalement peu de degrés de liberté qui sont en commun avec deux éléments. L’assemblage de la matrice nous fait gagner moins d’interactions que dans le cas de Helmholtz. En 3-D, ce qui prime plus, c’est le caractère très creux de la matrice de rigidité, qui donne un avantage significatif à la formulation mixte. L’ordre “optimal”, nécessitant le moins de calculs est le  $Q_3$  en 3-D.

## 4.2 Pollution de la solution : modes parasites

Dans cette section, nous nous limitons au cas 2-D, car nous n'avons pas traité le cas 3-D. En effet, au vu des inconvénients rencontrés en 2-D, il nous est paru inutile de tester la méthode en 3-D. Des résultats numériques montrant la déficience d'une telle approche en 3-D sont dans la thèse [Pernet, 2004]. Dans un premier temps, nous nous proposons de montrer des résultats de convergence en maillage régulier et en maillage “triangles découpés”. Dans un second temps, nous mettrons en évidence les modes propres parasites inhérents à la discrétisation choisie. Pour une étude théorique, on pourra lire [Boffi *et al.*, 1999], [Caorsi *et al.*, 1999] et [M. Costabel, 2003].

### 4.2.1 Etude de convergence

#### Comportement sur maillages réguliers

Nous considérons le cas d'un disque :

$$\left\{ \begin{array}{l} -\omega^2 \mathbf{E} + \operatorname{rot} \operatorname{rot} \mathbf{E} = 0 \\ \mathbf{E} \times \mathbf{n} = -\mathbf{E}^{\text{inc}} \times \mathbf{n} \text{ sur } C_a \\ \operatorname{rot} \mathbf{E} - i\omega \mathbf{E} \times \mathbf{n} = 0 \text{ sur } C_b \end{array} \right.$$

Le champ incident  $(E^{\text{inc}}, H^{\text{inc}})$  vérifie les équations de Maxwell avec :

$$H^{\text{inc}} = \exp(-i\mathbf{k} \cdot \mathbf{x})$$

La frontière intérieure  $C_a$  est un cercle de rayon  $a = 10$ , la frontière extérieure un cercle de rayon  $b = 12$ . La solution de ce problème est représentée sur la figure 4.4 (figure de droite). On peut voir sur cette figure que la condition de Silver-Müller pollue pas mal la zone d'ombre ! Des résultats numériques mettant en jeu une condition transparente, plutôt que la condition de Silver-Müller, sont montrés un peu plus loin. On peut ainsi effectuer une étude de convergence

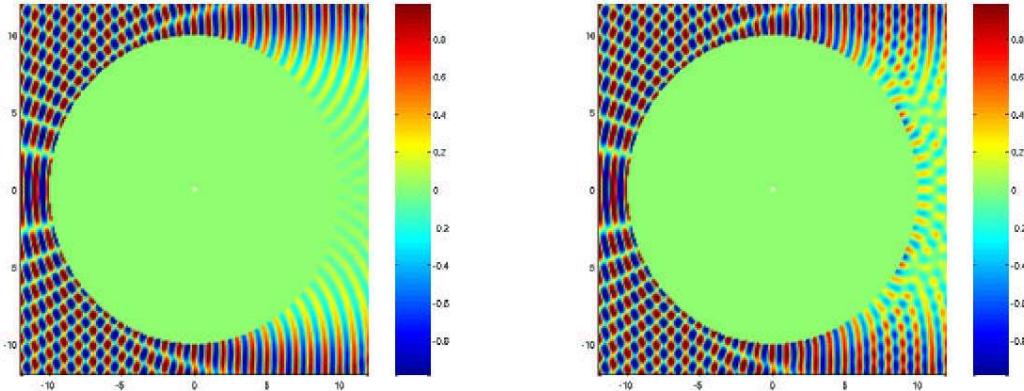


FIG. 4.4 – Partie réelle du champ total pour un disque de rayon 10. A gauche, solution analytique avec une condition transparente, à droite solution analytique avec une condition de Silver-Müller.

sur ce cas académique. Cette étude est réalisée sur la figure 4.5. On utilise la norme  $H\text{-rot}$ , car en pratique on calcule le champ scalaire  $H$ . Le calcul de l'erreur se fait entre ce champ scalaire numérique et la solution analytique  $H^{\text{analytique}}$ . Cette solution vérifie l'équation de Helmholtz avec une condition de Neumann inhomogène. On observe une convergence en  $O(h^r)$

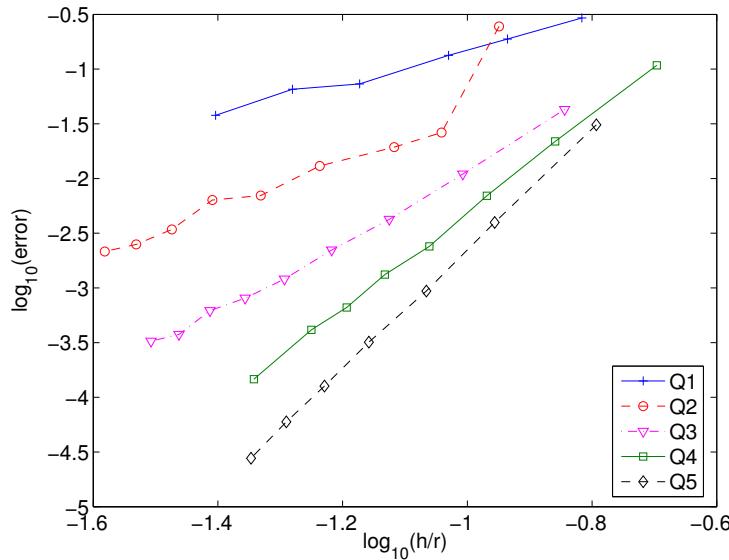


FIG. 4.5 – Evolution de l'erreur H-rot entre solution numérique et la solution analytique en fonction de  $h/r$ , où  $h$  est le pas de maillage,  $r$  l'ordre d'approximation. Echelle log-log, maillages réguliers sur le cas du disque parfaitement conducteur.

en maillage régulier. On a vraisemblablement un ordre de convergence optimal, lorsqu'on utilise de la condensation de masse et des éléments finis courbes. En abscisse de la figure 4.5, on a mis  $h/r$  et non  $h$ , afin de comparer les différents ordres d'approximations. On voit que sur ce cas, il est très intéressant de monter en ordre.

Sur des maillages quasi-réguliers de la sphère lorsqu'on utilise des sources de type “onde plane”, on ne rencontre pas de difficultés. Ce n'est pas le cas d'une source ponctuelle. Prenons pour illustrer notre propos, un domaine carré :

$$\left\{ \begin{array}{l} -\omega^2 \mathbf{E} + \mathbf{rot} \mathbf{rot} \mathbf{E} = \mathbf{f} \text{ sur } \Omega \\ \mathbf{E} \times \mathbf{n} = 0 \text{ sur } \Gamma \\ \mathbf{f}(x) = \mathbf{rot} \left( \frac{1}{r_0^2} \exp \left( -7 \frac{r^2}{r_0^2} \right) \right) \end{array} \right.$$

$\Omega$  est le carré  $[-1, 1]^2$ ,  $\Gamma$  le bord du carré,  $r_0$  le rayon de distribution de la gaussienne. on choisit le jeu de paramètres :

$$\omega = 2.02\pi \quad r_0 = 0.4$$

On obtient les solutions de la figure 4.6, lorsqu'on prend un maillage régulier avec 15 points par longueur d'onde. On voit que la solution est perturbée par des parasites et ce quel que soit l'ordre d'approximation utilisé, excepté pour  $Q_5$  sur cet exemple. Il faut noter que sur cet exemple, nous avons fait exprès de surmailler en mettant 15 points par longueur d'onde. C'est pour montrer que même sur des maillages fins et réguliers, il est possible d'observer des ondes parasites. Une manière de les enlever est de prendre un maillage encore plus fin... En dépit de ces ondes parasites, on constate la convergence de la méthode, comme le montre le tableau 4.1. La convergence est-elle en  $O(h^r)$ ? Elle est pour le moins irrégulière, il n'est pas possible de conclure. La solution de référence est calculée à l'aide de la première famille (cf. chapitre 5).

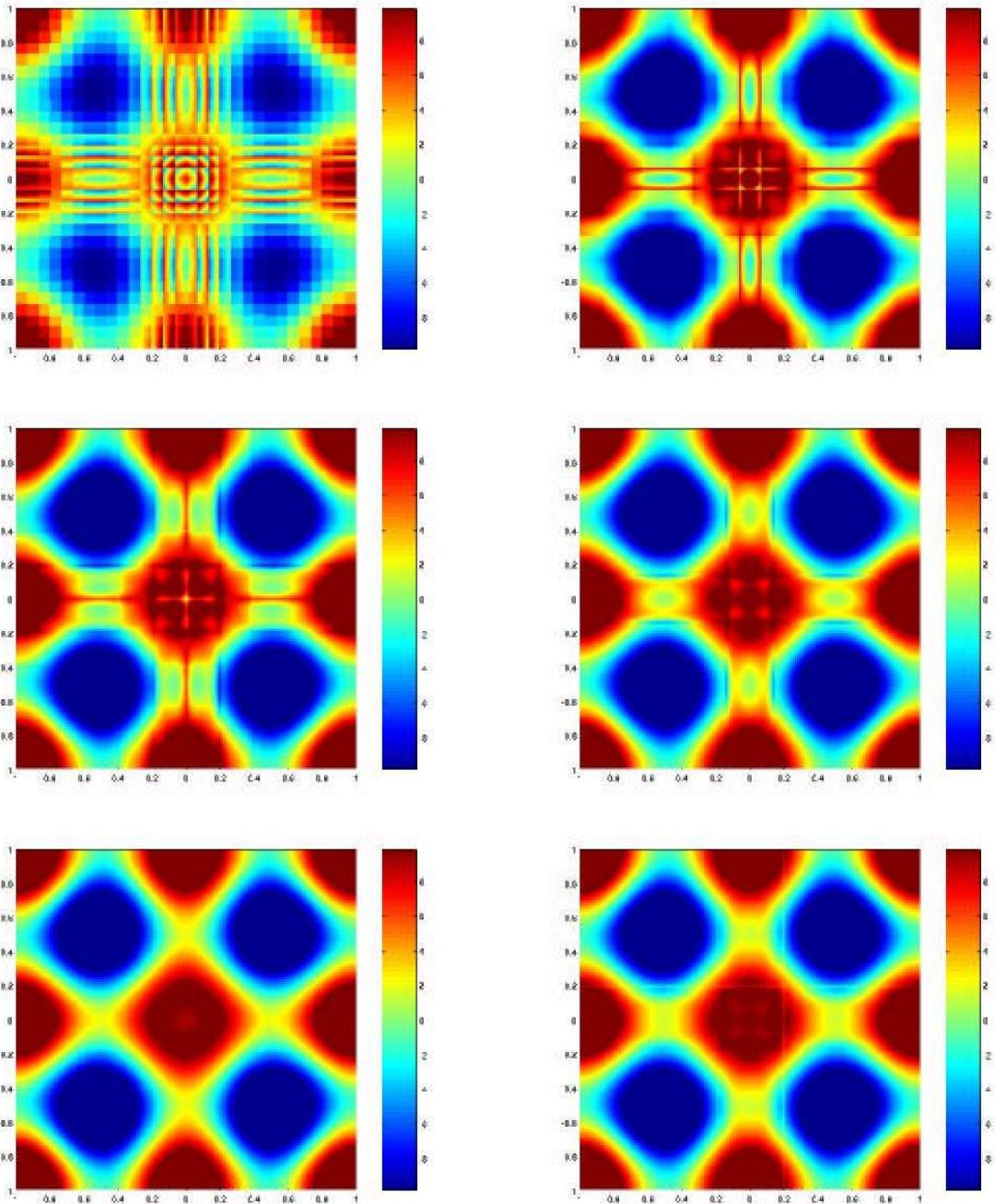


FIG. 4.6 – Solution numérique  $H$  sur maillage régulier, avec un point source. De gauche à droite, et de haut en bas, on utilise une approximation  $Q_1$ ,  $Q_2$ ,  $Q_3$ ,  $Q_4$ ,  $Q_5$  et  $Q_6$ .

### Comportement sur maillages non-réguliers

Sur des maillages réguliers, on rencontre quasiment tout le temps des parasites sur la solution, à moins de prendre un maillage très fin (30 points par longueur d'onde). Nous considérons la diffraction par un carré diélectrique (cf. figure 4.7). Le maillage utilisé est obtenu en découplant les triangles en trois quadrilatères. La solution numérique obtenue est sur la figure 4.8. On voit nettement que les ondes parasites épousent la forme du maillage et sont localisées. Elles

Pas de maillage $h$	Erreur	Ordre de convergence
0.667	0.664	-
0.333	0.0893	2.89
0.167	4.87e-4	7.52
0.0833	4.84e-4	0.01
0.0417	3.91e-5	3.62
0.0208	1.27e-5	1.62

TAB. 4.1 – Evolution de l'erreur pour une approximation  $Q_4$ , sur le carré avec un point source

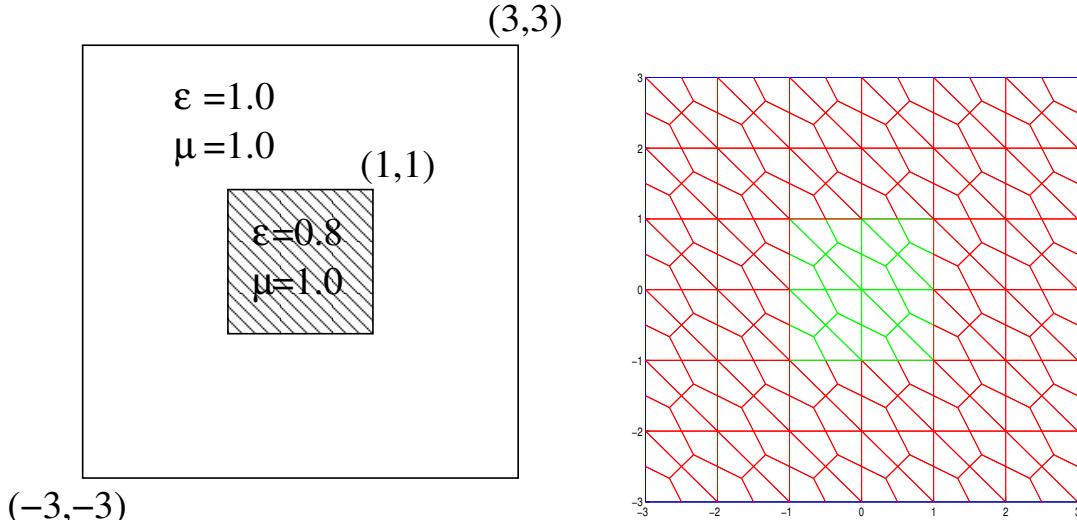


FIG. 4.7 – Etude de la diffraction d'un carré diélectrique. A droite, maillage utilisé pour les simulations.

s'ajoutent à la solution physique, sans la perturber. On a également testé si la condensation de

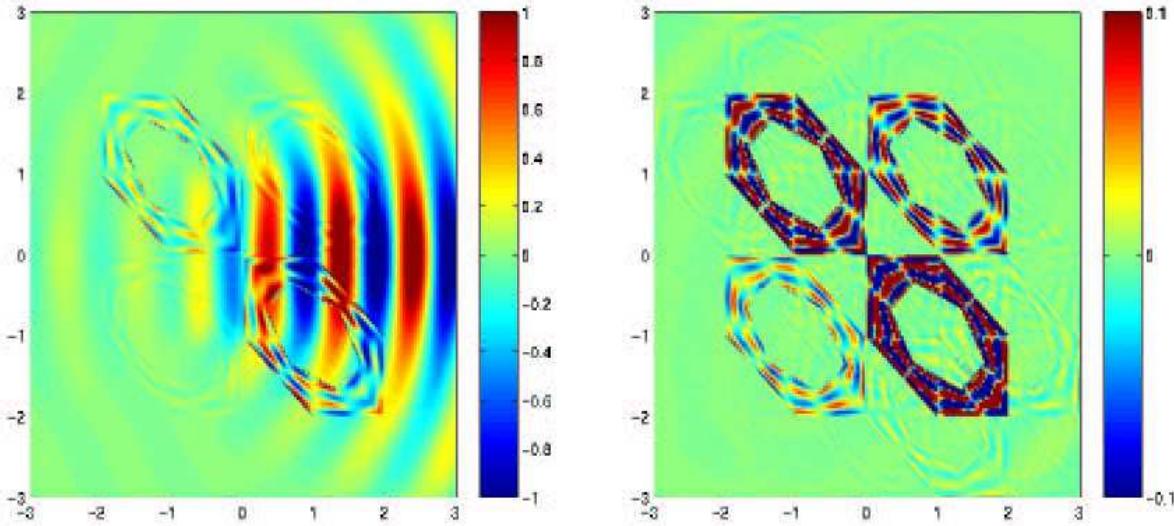


FIG. 4.8 – A gauche, partie réelle du champ diffracté  $H$  avec des éléments finis d'arête  $Q_5$  avec condensation de masse. A droite, différence avec la solution de référence (zoom de 10).

masse ne pouvait pas être responsable de ce phénomène. Il n'en est rien, on obtient également

des ondes parasites lorsqu'on ne condense pas la masse. Néanmoins on a observé que les ondes parasites étaient totalement différentes si on utilisait les formules de Gauss-Lobatto ou les formules de Gauss pour calculer les intégrales.

De manière générale, l'apparition de ces ondes parasites est assez imprévisible, on a beau raffiner le maillage, on n'est pas à l'abri de ces parasites, particulièrement en maillage non-régulier. On illustre cette propriété sur la figure 4.9. Le cas est la diffraction d'un disque parfaitement

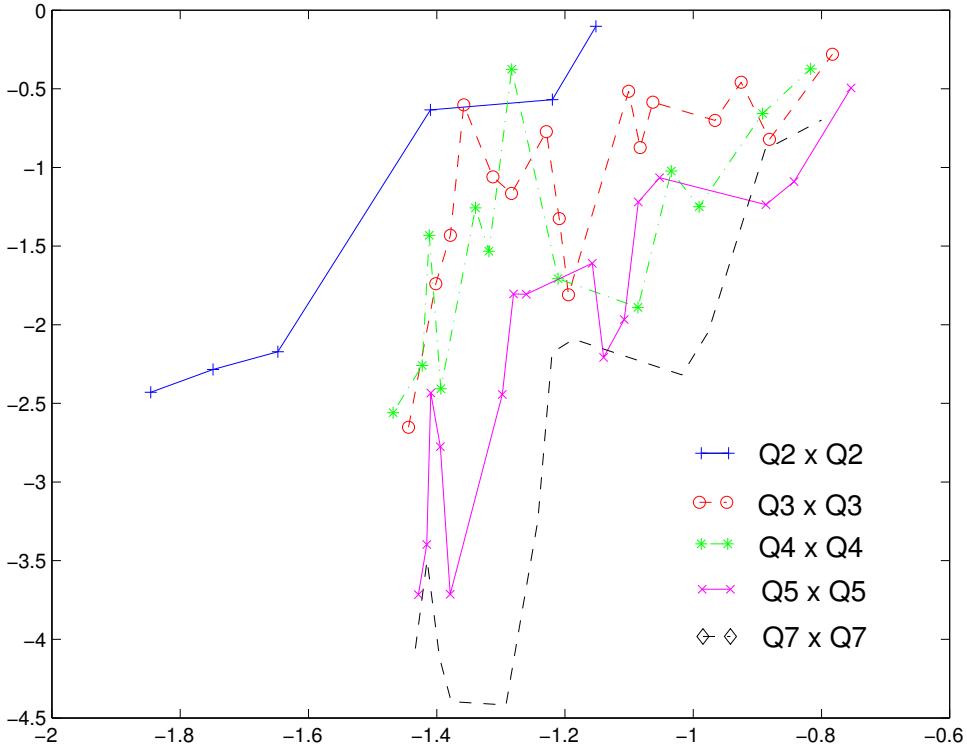


FIG. 4.9 – Evolution de l'erreur  $H_{\text{rot}}$  entre la solution numérique et la solution analytique, sur des maillages “triangles découpés. Echelle log-log, utilisation de la condition transparente. Cas du disque parfaitement conducteur.

conducteur, de mêmes caractéristiques que la section précédente, la solution analytique est sur la figure 4.4 (figure de gauche). On utilise cette fois une condition transparente au lieu de la condition de Silver-Müller, et des maillages de mauvaise qualité (triangles découpés). On observe une convergence très erratique de la solution numérique, mais visiblement la méthode est néanmoins consistante. Il peut arriver que les ondes parasites soient localisées à un endroit précis du maillage et de très forte amplitude. Elles donnent une erreur très importante, alors que sur le reste du maillage, l'erreur est minime.

#### 4.2.2 Etude de la distribution des valeurs propres

Une manière différente d'aborder ce problèmes de modes parasites, est de faire une étude de valeurs propres et de les exhiber plus clairement. Le problème aux valeurs propres s'écrit :

$$\left\{ \begin{array}{l} \text{trouver } (\omega, \mathbf{u}) \in \mathbb{R} \times H(\text{rot}, \Omega) \quad \omega \neq 0 \quad \mathbf{u} \neq \mathbf{0} \quad \text{tel que} \\ -\omega^2 \mathbf{u} + \text{rot rot} \mathbf{u} = 0 \quad \in \Omega \\ \mathbf{u} \times \mathbf{n} = 0 \quad \in \partial\Omega \end{array} \right.$$

Le domaine  $\Omega$  considéré est le carré  $[-1, 1]^2$ , les valeurs propres analytiques sont égales à :

$$\omega_{m,n} = \frac{\pi \sqrt{m^2 + n^2}}{2} \quad m \geq 0 \quad n \geq 0 \quad (m, n) \neq (0, 0)$$

Les modes propres associés sont égaux à :

$$H_{m,n}(x, y) = \cos\left(\frac{\pi m(1+x)}{2}\right) \cos\left(\frac{\pi n(1+y)}{2}\right)$$

### Modes propres en maillage régulier

On calcule les modes propres avec Arpack [Lehoucq *et al.*, 1996]. Cette librairie permet de calculer rapidement des valeurs propres et modes propres d'une matrice creuse à l'aide de méthodes itératives. Le mode de résolution le plus intéressant est le mode de Cayley, qui permet d'éliminer la valeur propre nulle et de sélectionner un nombre limité de modes autour d'une fréquence centrale. On commence par un maillage  $Q_1$  avec 20 cellules dans chaque direction. On affiche sur la figure 4.10, la distribution de valeurs propres. Pour ce cas-là, on a calculé toutes les valeurs propres (1 600 degrés de liberté), afin de ne pas sous-estimer la multiplicité. On ne tient pas en compte des valeurs propres nulles. On voit qu'on a les bonnes valeurs

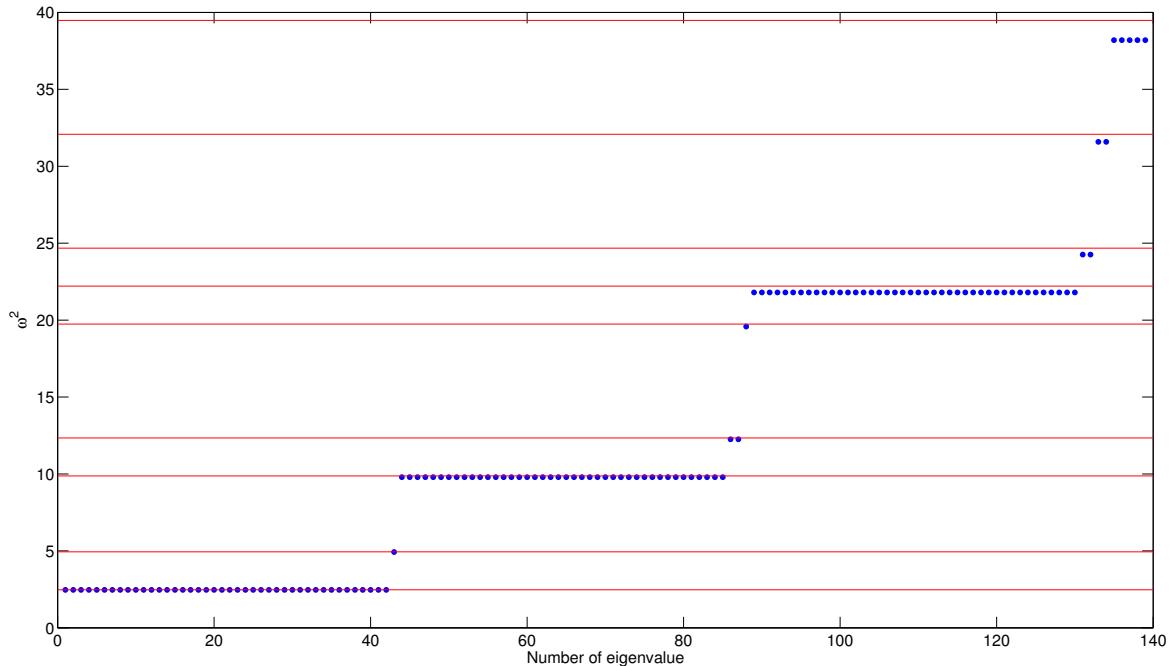


FIG. 4.10 – Distribution des valeurs propres sur un maillage  $20 \times 20$  avec une approximation  $Q_1$ . Les traits rouges symbolisent les valeurs propres analytiques. Les points bleus représentent les valeurs propres numériques. Les décalages sont dûs à l'erreur de dispersion importante en  $Q_1$ .

propres, mais la multiplicité est incorrecte. De fait, certaines valeurs propres correspondent à des modes propres non-physiques. On affiche sur la figure 4.11 quelques modes propres associés à la première valeur propre. On contemple ainsi les modes propres parasites ! La première valeur propre est de multiplicité théorique 2 (le mode  $(1,0)$  et le mode  $(0,1)$ ) alors que numériquement on trouve une multiplicité 42 ! Le spectre de l'opérateur est complètement pollué, il est difficile

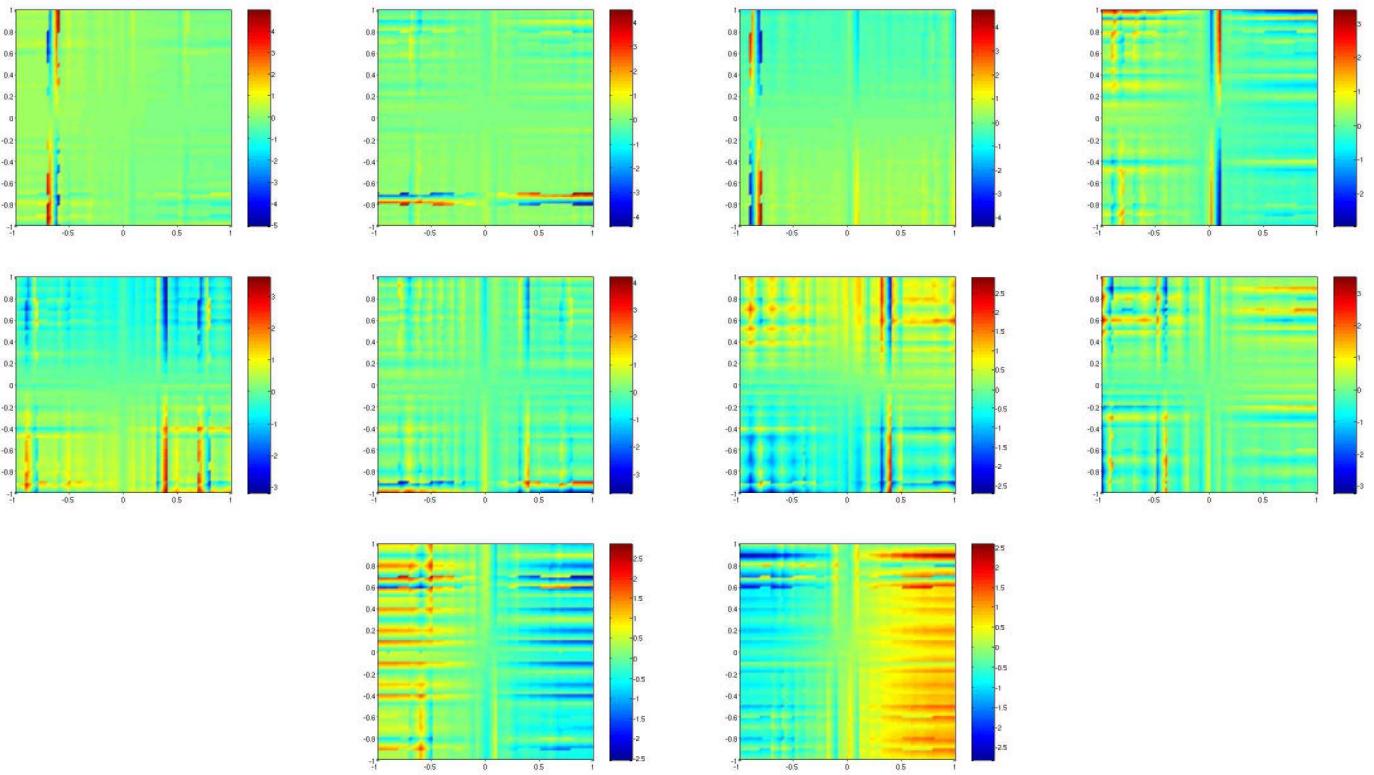


FIG. 4.11 – Quelques modes propres associés à la première valeur propre (1,0). On utilise une approximation  $Q_1$  sur un maillage 20x20.

d'isoler les modes propres physiques. Les deux modes physiques ne sont pas identifiables, ce qui est facile à expliquer. Les modes propres forment une base orthogonale de l'espace des vecteurs propres associés à la première valeur propre. Cette base n'est pas unique, on peut faire des combinaisons linéaires entre ces modes propres pour obtenir une autre base orthogonale. On pourrait donc retrouver les modes propres physiques si on savait, quelle combinaison linéaire de tous les modes, permettrait de les obtenir. Lorsqu'on raffine le maillage, les modes physiques convergent vers leur valeur théorique, tandis que les modes parasites varient suivant le maillage. Mais comme les deux types de modes sont indissociables car ils ont la même valeur propre, il nous est impossible de vérifier la convergence vers le mode physique. Au niveau des valeurs propres, le fait de raffiner le maillage ne rejette pas les valeurs propres parasites vers l'infini. On a remarqué que la multiplicité des valeurs propres augmente lorsqu'on raffine le maillage !

On peut faire la même analyse pour une approximation  $Q_4$  avec 5 cellules dans chaque direction, On met les modes propres parasites sur la figure 4.12. Les conclusions sont identiques à ce qu'on a observé en  $Q_1$ . On observe également le même phénomène lorsqu'on prend le cas limite où le maillage est constitué d'un seul élément avec une approximation  $Q_8$ , cf. figure 4.13. Sur tous les cas numériques traités, les valeurs propres simples ne semblent pas être parasitées.

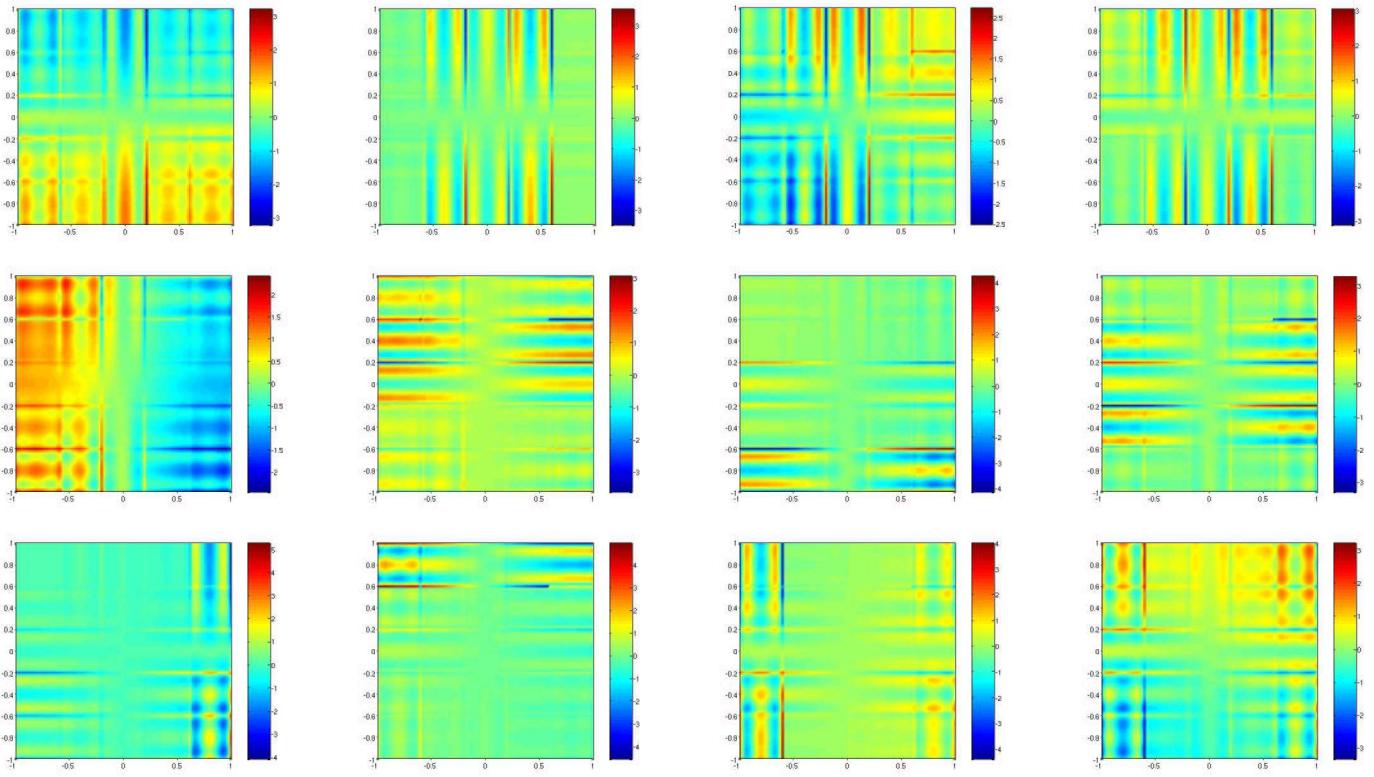


FIG. 4.12 – Quelques modes propres associés à la première valeur propre  $(1,0)$ . On utilise une approximation  $Q_4$  sur un maillage  $5 \times 5$ .

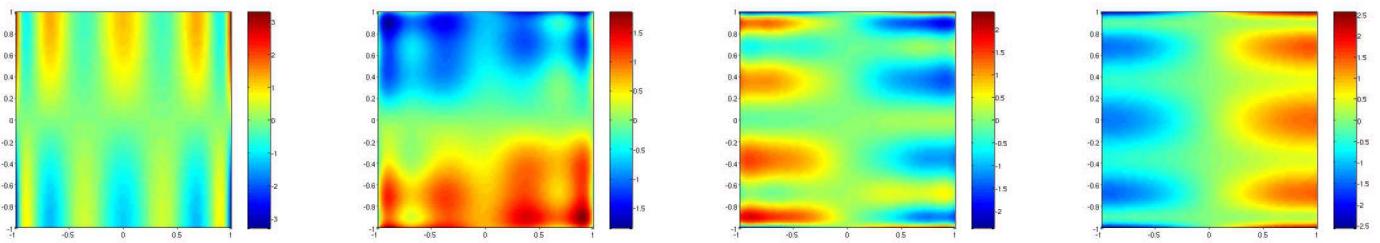


FIG. 4.13 – Tous les modes propres associés à la première valeur propre  $(1,0)$ . On utilise une approximation  $Q_8$  sur un seul élément.

## Modes propres en maillage non-régulier

Nous faisons la même étude sur un petit maillage non régulier (cf. figure 4.14). On utilise une approximation  $Q_5$ . Sur ce type de maillage, les valeurs propres physiques ont la bonne mul-

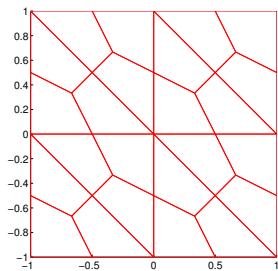


FIG. 4.14 – Maillage utilisé pour le calcul des modes propres

tiplicité, mais on a des valeurs propres parasites. On a synthétisé la différence maillage régulier/

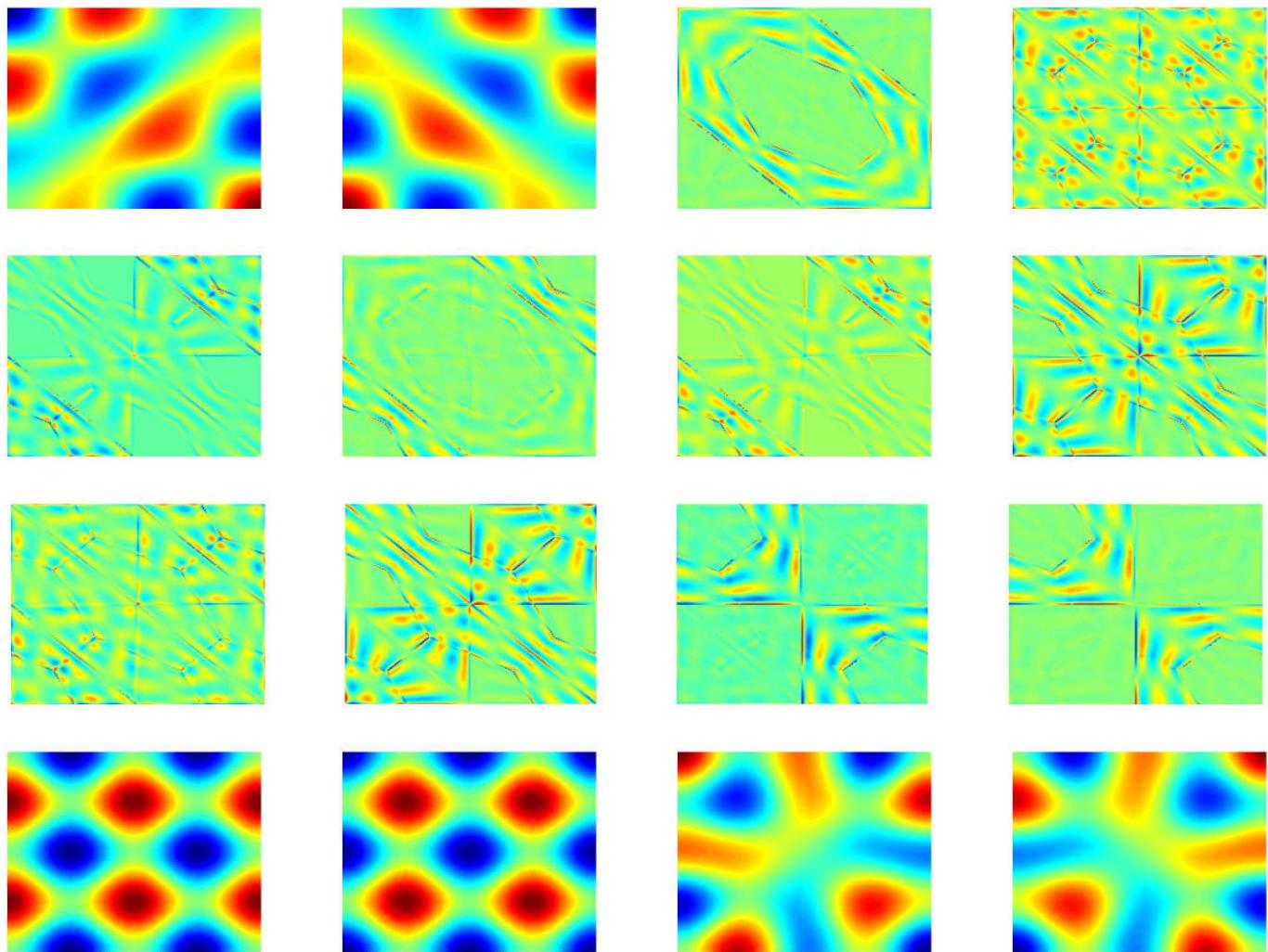


FIG. 4.15 – Quelques modes physiques entourés de modes parasites !

maillage non-régulier sur la figure 4.16. On a sélectionné une bande de spectre correspondant à la finesse du maillage. L'utilisation de maillages non-réguliers a un avantage, celui de séparer

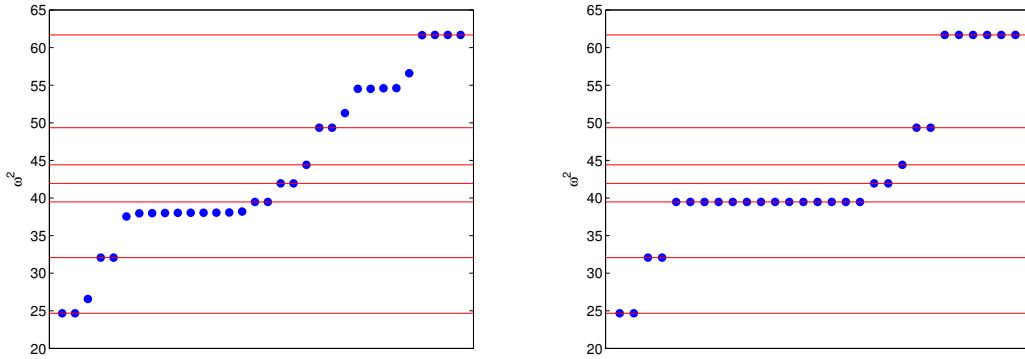


FIG. 4.16 – A gauche, distribution des valeurs propres sur maillage “triangles découpés” ; à droite, distribution sur maillage régulier. Les traits horizontaux rouges symbolisent les valeurs propres analytiques. Les points bleus sont les valeurs propres numériques. On utilise une approximation  $Q_5$ , l'erreur de dispersion n'est pas visible comme en  $Q_1$ .

les modes propres physiques des modes propres parasites ! Mais probablement que l'apparition de valeurs propres parasites est la cause des solutions beaucoup plus perturbées lorsqu'on est en maillage non-régulier.

### 4.3 Conclusion

Dans ce chapitre, nous avons décrit comment l'utilisation de la seconde famille de Nédélec sur les quadrilatères menait à un gain en temps de calcul et en stockage. Des études de complexité ont montré que ce gain était réalisé à partir de l'ordre 1.

Nous avons ensuite constaté que cette discréttisation donne des résultats numériques décevants, à cause de l'apparition d'ondes “parasites”. Ces parasites sont très pénalisants lorsque l'on utilise des maillages triangulaires découpés.

Une étude numérique des valeurs propres et vecteurs propres, exhibe plus clairement les modes propres parasites. Sur des maillages réguliers, les modes propres parasites ont la même valeur propre que les modes propres physiques. Sur maillage non-régulier, les valeurs propres diffèrent.

La présence de valeurs propres parasites détériore le conditionnement de la matrice, même en utilisant des préconditionneurs. Toutes ces raisons nous ont définitivement décidés à exclure l'utilisation de ces éléments, même sur des maillages réguliers.

# Chapitre 5

## Première famille sur les quadrilatères/hexaèdres

*Une alternative séduisante de la seconde famille est la première famille, dont l'espace local de polynômes est légèrement différent. Nous verrons qu'avec cet espace d'approximation, on obtient une méthode spectralement correcte. L'inconvénient majeur est la perte de la condensation de masse. Néanmoins, il est possible de trouver un algorithme rapide pour réaliser un produit matrice-vecteur. Dans la dernière section, nous verrons comment on peut préconditionner le système linéaire issu d'une telle discrétisation.*

### Sommaire

---

<b>5.1</b>	<b>Formulation variationnelle et espace d'approximation</b>	<b>116</b>
<b>5.2</b>	<b>Expression des matrices élémentaires</b>	<b>116</b>
5.2.1	Cas 2-D	117
5.2.2	Cas 3-D	118
<b>5.3</b>	<b>Précision de la méthode</b>	<b>123</b>
<b>5.4</b>	<b>Algorithme rapide du produit matrice-vecteur</b>	<b>128</b>
5.4.1	Factorisation discrète	128
5.4.2	Formulation mixte	129
5.4.3	Produit $\hat{R}E$ et $\hat{C}E$ en 2-D	130
5.4.4	Produit $\hat{R}E$ et $\hat{C}E$ en 3-D	131
5.4.5	Complexité du produit matrice-vecteur	134
<b>5.5</b>	<b>Calcul de modes propres</b>	<b>136</b>
5.5.1	Maillage régulier	137
5.5.2	Maillage non-régulier	137
<b>5.6</b>	<b>Préconditionnement du système linéaire</b>	<b>137</b>
5.6.1	Préconditionnement par un sous-maillage $Q_1$	139
5.6.2	Préconditionnement à l'aide d'une factorisation incomplète	141
5.6.3	Préconditionnement utilisant la décomposition de Helmholtz	142
5.6.4	Multigrille	144
<b>5.7</b>	<b>Conclusion</b>	<b>145</b>

---

## 5.1 Formulation variationnelle et espace d'approximation

On choisit d'adopter dans cette section, des notations 3-D. On part de la formulation variationnelle d'ordre deux, obtenue dans le chapitre précédent :

$$-\omega^2 \int_{\Omega} \varepsilon \mathbf{E} \cdot \boldsymbol{\varphi} + \int_{\Omega} \frac{1}{\mu} \operatorname{rot} \mathbf{E} \operatorname{rot} \boldsymbol{\varphi} + \int_{\partial\Omega} \frac{1}{\mu} \operatorname{rot} \mathbf{E} \boldsymbol{\varphi} \times \mathbf{n} = - \int_{\Omega} \mathbf{f} \cdot \boldsymbol{\varphi} \quad (5.1)$$

On choisit  $\mathbf{E}, \boldsymbol{\varphi}$  dans l'espace :

$$H(\operatorname{rot}, \Omega) = \{\mathbf{u} \in (L^2(\Omega))^3 \mid \operatorname{rot} \mathbf{u} \in (L^2(\Omega))^3\}$$

L'espace d'approximation est la première famille de Nédélec sur les hexaèdres :

$$V_h = \{\mathbf{u} \in H(\operatorname{rot}, \Omega) \text{ tel que } DF_i^t \mathbf{u} \circ F_i \in Q_{r-1,r,r} \times Q_{r,r-1,r} \times Q_{r,r,r-1}\}$$

où  $r$  est l'ordre d'approximation. Les fonctions de bases locales sont égales à (cf. [Cohen et Monk, 1998]) :

$$\hat{\boldsymbol{\varphi}}_{(i,1)} = \hat{\varphi}_{i_1}^G(\hat{x}_1) \hat{\varphi}_{i_2}^{GL}(\hat{x}_2) \hat{\varphi}_{i_3}^{GL}(\hat{x}_3) \mathbf{e}_1 \quad i_1 \in [1, r] \quad i_2 \in [1, r+1] \quad i_3 \in [1, r+1]$$

$$\hat{\boldsymbol{\varphi}}_{(i,2)} = \hat{\varphi}_{i_1}^{GL}(\hat{x}_1) \hat{\varphi}_{i_2}^G(\hat{x}_2) \hat{\varphi}_{i_3}^{GL}(\hat{x}_3) \mathbf{e}_2 \quad i_1 \in [1, r+1] \quad i_2 \in [1, r] \quad i_3 \in [1, r+1]$$

$$\hat{\boldsymbol{\varphi}}_{(i,3)} = \hat{\varphi}_{i_1}^{GL}(\hat{x}_1) \hat{\varphi}_{i_2}^{GL}(\hat{x}_2) \hat{\varphi}_{i_3}^G(\hat{x}_3) \mathbf{e}_3 \quad i_1 \in [1, r+1] \quad i_2 \in [1, r+1] \quad i_3 \in [1, r]$$

On note les points de Gauss :

$$\hat{\xi}_i^G \quad i \in [1, r]$$

et les points de Gauss-Lobatto :

$$\hat{\xi}_i^{GL} \quad i \in [1, r+1]$$

les fonctions de base lagrangiennes associées aux points de Gauss :

$$\text{Soit } i \in [1, r], \quad \hat{\varphi}_i^G \in P_{r-1} \quad \text{tel que} \quad \forall j \in [1, r] \quad \hat{\varphi}_i^G(\hat{\xi}_j^G) = \delta_{i,j}$$

les fonctions de base lagrangiennes associées aux points de Gauss-Lobatto :

$$\text{Soit } i \in [1, r+1], \quad \hat{\varphi}_i^{GL} \in P_r \quad \text{tel que} \quad \forall j \in [1, r+1] \quad \hat{\varphi}_i^{GL}(\hat{\xi}_j^{GL}) = \delta_{i,j}$$

Les degrés de liberté obtenus sont affichés sur la figure 5.1. On introduit la matrice de masse :

$$(M_h)_{j,k} = \int_{\Omega} \varepsilon \boldsymbol{\varphi}_j \cdot \boldsymbol{\varphi}_k$$

et la matrice de rigidité :

$$(K_h)_{j,k} = \int_{\Omega} \frac{1}{\mu} \operatorname{rot} \boldsymbol{\varphi}_j \cdot \operatorname{rot} \boldsymbol{\varphi}_k$$

## 5.2 Expression des matrices élémentaires

On effectue le changement de variable pour se ramener à l'élément de référence :

$$(M_h)_{j,k} = \int_{\hat{K}} \varepsilon J_i DF_i^{-1} DF_i^{-t} \hat{\boldsymbol{\varphi}}_j \cdot \hat{\boldsymbol{\varphi}}_k$$

$$(K_h)_{j,k} = \int_{\hat{K}} \frac{1}{J_i \mu} DF_i^t DF_i \operatorname{rot} \hat{\boldsymbol{\varphi}}_j \cdot \operatorname{rot} \hat{\boldsymbol{\varphi}}_k$$

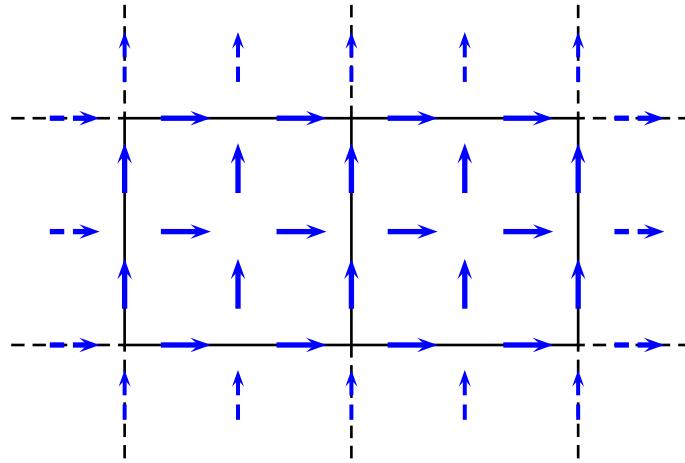


FIG. 5.1 – Degrés de liberté de la première famille sur les quadrilatères pour  $r = 2$

$DF_i$  n'est pas une matrice constante dans le cas d'hexaèdres droits quelconques, une première solution est d'utiliser brutallement  $(r + 1)^3$  points de Gauss (formule exacte pour  $(Q_{2r+1})^3$ ) pour intégrer les deux matrices. Cette solution ne nous convient pas car en 3-D, on aura une complexité du calcul de la matrice en  $O(r^9)$  où  $r$  est l'ordre d'approximation, ce qui est assez prohibitif si on veut faire du  $Q_6$  ou du  $Q_7$ . On choisit de sous-intégrer ces deux matrices, on évalue la matrice de masse avec  $(r + 1)^3$  points de Gauss-Lobatto et la matrice de rigidité avec  $r^3$  points de Gauss. Dans les deux cas, on utilise une formule exacte pour  $(Q_{2r-1})^3$ . On aboutira à une complexité en  $O(r^4)$  en 2-D et  $O(r^7)$  en 3-D.

### 5.2.1 Cas 2-D

#### Matrice de masse

On introduit la matrice 2x2 :

$$B = \varepsilon J_i DF_i^{-1} DF_i^{-t}$$

On évalue cette matrice sur tous les points de Gauss-Lobatto. On a alors :

$$(M_h)_{(j,1),(k,1)} = \sum_{m,n} (B_{11})(\hat{\xi}_m^{GL}, \hat{\xi}_n^{GL}) \omega_{m,n}^{GL} \hat{\varphi}_{j_1}^G(\hat{\xi}_m^{GL}) \hat{\varphi}_{j_2}^G(\hat{\xi}_n^{GL}) \hat{\varphi}_{k_1}^G(\hat{\xi}_m^{GL}) \hat{\varphi}_{k_2}^G(\hat{\xi}_n^{GL})$$

Soit après simplification :

$$(M_h)_{(j,1),(k,1)} = \delta_{j_2,k_2} \sum_{m=1}^{r+1} (B_{11})(\hat{\xi}_m^{GL}, \hat{\xi}_{j_2}^{GL}) \omega_{m,j_2}^{GL} \hat{\varphi}_{j_1}^G(\hat{\xi}_m^{GL}) \hat{\varphi}_{k_1}^G(\hat{\xi}_m^{GL})$$

De façon analogue, on trouve les autres termes de la matrice :

$$(M_h)_{(j,2),(k,2)} = \delta_{j_1,k_1} \sum_{m=1}^{r+1} (B_{22})(\hat{\xi}_{j_1}^{GL}, \hat{\xi}_m^{GL}) \omega_{j_1,m}^{GL} \hat{\varphi}_{j_2}^G(\hat{\xi}_m^{GL}) \hat{\varphi}_{k_2}^G(\hat{\xi}_m^{GL})$$

$$(M_h)_{(j,1),(k,2)} = (B_{21})(\hat{\xi}_{k_1}^{GL}, \hat{\xi}_{j_2}^{GL}) \omega_{k_1,j_2}^{GL} \hat{\varphi}_{j_1}^G(\hat{\xi}_{k_1}^{GL}) \hat{\varphi}_{k_2}^G(\hat{\xi}_{j_2}^{GL})$$

On a bien une complexité en  $O(r^4)$  pour le calcul de la matrice de masse. Contrairement à la première famille, la matrice de masse n'est pas diagonale. Toutefois sur maillage orthogonal, il est

possible de la condenser en utilisant des points biens choisis (points de Gauss dans une direction et Gauss-Lobatto dans l'autre direction). Nous n'utiliserons pas cette propriété, exploitée en régime temporel dans [Cohen et Monk, 1998]. Si on applique des éléments finis, c'est d'abord pour pouvoir traiter des géométries complexes et donc des maillages hexaédriques quelconques.

### Matrice de rigidité

Dans le cas 2-D, le rotationnel est orienté suivant  $\mathbf{e}_z$ , et la matrice jacobienne  $DF_i$  doit être interprétée comme :

$$DF_i = \begin{pmatrix} DF_i^{2D} & 0 \\ 0 & 1 \end{pmatrix}$$

La matrice de rigidité élémentaire se simplifie :

$$(K_h)_{j,k} = \int_{\hat{K}} \frac{1}{J_i \mu} \operatorname{rot} \hat{\varphi}_j \operatorname{rot} \hat{\varphi}_k$$

On note le scalaire :

$$A = \frac{1}{J_i \mu}$$

On évalue cette quantité sur tous les points de Gauss, on a alors :

$$(K_h)_{(j,1),(k,1)} = \sum_{m,n} A(\hat{\xi}_m^G, \hat{\xi}_n^G) \omega_{m,n}^G \hat{\varphi}_{j_1}^G(\hat{\xi}_m^G) \hat{\varphi}'_{j_2}^{GL}(\hat{\xi}_n^G) \hat{\varphi}_{k_1}^G(\hat{\xi}_m^G) \hat{\varphi}'_{k_2}^{GL}(\hat{\xi}_n^G)$$

Soit après simplification :

$$(K_h)_{(j,1),(k,1)} = \delta_{j_1, k_1} \sum_{m=1}^r A(\hat{\xi}_{j_1}^G, \hat{\xi}_m^G) \omega_{j_1, m}^G \hat{\varphi}'_{j_2}^{GL}(\hat{\xi}_m^G) \hat{\varphi}'_{k_2}^{GL}(\hat{\xi}_m^G)$$

De façon analogue :

$$(K_h)_{(j,2),(k,2)} = \delta_{j_2, k_2} \sum_{m=1}^r A(\hat{\xi}_m^G, \hat{\xi}_{j_2}^G) \omega_{m, j_2}^G \hat{\varphi}'_{j_1}^{GL}(\hat{\xi}_m^G) \hat{\varphi}'_{k_1}^{GL}(\hat{\xi}_m^G)$$

$$(K_h)_{(j,1),(k,2)} = A(\hat{\xi}_{j_1}^G, \hat{\xi}_{k_2}^G) \omega_{j_1, k_2}^G \hat{\varphi}'_{j_2}^{GL}(\hat{\xi}_{k_2}^G) \hat{\varphi}'_{k_1}^{GL}(\hat{\xi}_{j_1}^G)$$

Là aussi, la complexité du calcul de la matrice de rigidité est en  $O(r^4)$ .

### 5.2.2 Cas 3-D

On explicite ici le calcul de la matrice en utilisant les points de Gauss-Lobatto pour la matrice de masse **et** la matrice de rigidité.

#### Matrice de masse

On introduit la matrice 3x3 :

$$B = \varepsilon J_i DF_i^{-1} DF_i^{-t}$$

On évalue cette matrice sur tous les points de Gauss-Lobatto. On a alors :

$$(M_h)_{(j,1),(k,1)} = \sum_{m_1, m_2, m_3} (B_{11})(\hat{\xi}_m^{GL}) \omega_{m_1, m_2, m_3}^{GL} \hat{\varphi}_{j_1}^G(\hat{\xi}_{m_1}^{GL}) \hat{\varphi}_{j_2}^{GL}(\hat{\xi}_{m_2}^{GL}) \hat{\varphi}_{j_3}^{GL}(\hat{\xi}_{m_3}^{GL}) \hat{\varphi}_{k_1}^G(\hat{\xi}_{m_1}^{GL}) \hat{\varphi}_{k_2}^{GL}(\hat{\xi}_{m_2}^{GL}) \hat{\varphi}_{k_3}^{GL}(\hat{\xi}_{m_3}^{GL})$$

Soit après simplification :

$$(M_h)_{(j,1),(k,1)} = \delta_{j_2,k_2} \delta_{j_3,k_3} \sum_{m=1}^{r+1} (B_{11})(\hat{\xi}_{m,j_2,j_3}^{GL}) \omega_{m,j_2,j_3}^{GL} \hat{\varphi}_{j_1}^G(\hat{\xi}_m^{GL}) \hat{\varphi}_{k_1}^G(\hat{\xi}_m^{GL})$$

De manière analogue :

$$(M_h)_{(j,2),(k,2)} = \delta_{j_1,k_1} \delta_{j_3,k_3} \sum_{m=1}^{r+1} (B_{22})(\hat{\xi}_{j_1,m,j_3}^{GL}) \omega_{j_1,m,j_3}^{GL} \hat{\varphi}_{j_2}^G(\hat{\xi}_m^{GL}) \hat{\varphi}_{k_2}^G(\hat{\xi}_m^{GL})$$

$$(M_h)_{(j,3),(k,3)} = \delta_{j_1,k_1} \delta_{j_2,k_2} \sum_{m=1}^{r+1} (B_{33})(\hat{\xi}_{j_1,j_2,m}^{GL}) \omega_{j_1,j_2,m}^{GL} \hat{\varphi}_{j_3}^G(\hat{\xi}_m^{GL}) \hat{\varphi}_{k_3}^G(\hat{\xi}_m^{GL})$$

$$(M_h)_{(j,1),(k,2)} = \delta_{j_3,k_3} (B_{21})(\hat{\xi}_{k_1,j_2,j_3}^{GL}) \omega_{k_1,j_2,j_3}^{GL} \hat{\varphi}_{j_1}^G(\hat{\xi}_{k_1}^{GL}) \hat{\varphi}_{j_2}^G(\hat{\xi}_{j_3}^{GL})$$

$$(M_h)_{(j,1),(k,3)} = \delta_{j_2,k_2} (B_{31})(\hat{\xi}_{k_1,j_2,j_3}^{GL}) \omega_{k_1,j_2,j_3}^{GL} \hat{\varphi}_{j_1}^G(\hat{\xi}_{k_1}^{GL}) \hat{\varphi}_{k_3}^G(\hat{\xi}_{j_3}^{GL})$$

$$(M_h)_{(j,2),(k,3)} = \delta_{j_1,k_1} (B_{32})(\hat{\xi}_{j_1,k_2,j_3}^{GL}) \omega_{j_1,k_2,j_3}^{GL} \hat{\varphi}_{j_2}^G(\hat{\xi}_{k_2}^{GL}) \hat{\varphi}_{k_3}^G(\hat{\xi}_{j_3}^{GL})$$

La complexité de calcul de la matrice de masse est donc en  $O(r^5)$ , elle n'est pas diagonale.

### Matrice de rigidité

Le calcul de la matrice de rigidité est un poil plus complexe. On introduit la matrice 3x3 :

$$A = \varepsilon \frac{1}{J_i \mu} DF_i^t DF_i$$

On calcule le rotationnel des fonctions de base :

$$\text{rot} \hat{\varphi}_{(j,1)} = \begin{vmatrix} 0 \\ \hat{\varphi}_{j_1}^G \hat{\varphi}_{j_2}^{GL} \hat{\varphi}_{j_3}'^{GL} \\ -\hat{\varphi}_{j_1}^G \hat{\varphi}_{j_2}'^{GL} \hat{\varphi}_{j_3}^{GL} \end{vmatrix} \quad \text{rot} \hat{\varphi}_{(j,2)} = \begin{vmatrix} -\hat{\varphi}_{j_1}^{GL} \hat{\varphi}_{j_2}^G \hat{\varphi}_{j_3}'^{GL} \\ 0 \\ \hat{\varphi}_{j_1}'^{GL} \hat{\varphi}_{j_2}^G \hat{\varphi}_{j_3}^{GL} \end{vmatrix} \quad \text{rot} \hat{\varphi}_{(j,3)} = \begin{vmatrix} \hat{\varphi}_{j_1}^{GL} \hat{\varphi}_{j_2}'^{GL} \hat{\varphi}_{j_3}^G \\ -\hat{\varphi}_{j_1}'^{GL} \hat{\varphi}_{j_2}^{GL} \hat{\varphi}_{j_3}^G \\ 0 \end{vmatrix}$$

On applique la matrice A à ces rotationnels :

$$A \text{rot} \hat{\varphi}_{(j,1)} = \begin{vmatrix} A_{21} \hat{\varphi}_{j_1}^G \hat{\varphi}_{j_2}^{GL} \hat{\varphi}_{j_3}'^{GL} - A_{32} \hat{\varphi}_{j_1}^G \hat{\varphi}_{j_2}'^{GL} \hat{\varphi}_{j_3}^{GL} \\ A_{22} \hat{\varphi}_{j_1}^G \hat{\varphi}_{j_2}^{GL} \hat{\varphi}_{j_3}'^{GL} - A_{32} \hat{\varphi}_{j_1}^G \hat{\varphi}_{j_2}'^{GL} \hat{\varphi}_{j_3}^{GL} \\ A_{32} \hat{\varphi}_{j_1}^G \hat{\varphi}_{j_2}^{GL} \hat{\varphi}_{j_3}'^{GL} - A_{33} \hat{\varphi}_{j_1}^G \hat{\varphi}_{j_2}'^{GL} \hat{\varphi}_{j_3}^{GL} \end{vmatrix}$$

$$A \text{rot} \hat{\varphi}_{(j,2)} = \begin{vmatrix} -A_{11} \hat{\varphi}_{j_1}^{GL} \hat{\varphi}_{j_2}^G \hat{\varphi}_{j_3}'^{GL} + A_{31} \hat{\varphi}_{j_1}'^{GL} \hat{\varphi}_{j_2}^G \hat{\varphi}_{j_3}^{GL} \\ -A_{21} \hat{\varphi}_{j_1}^{GL} \hat{\varphi}_{j_2}^G \hat{\varphi}_{j_3}'^{GL} + A_{32} \hat{\varphi}_{j_1}'^{GL} \hat{\varphi}_{j_2}^G \hat{\varphi}_{j_3}^{GL} \\ -A_{31} \hat{\varphi}_{j_1}^{GL} \hat{\varphi}_{j_2}^G \hat{\varphi}_{j_3}'^{GL} + A_{33} \hat{\varphi}_{j_1}'^{GL} \hat{\varphi}_{j_2}^G \hat{\varphi}_{j_3}^{GL} \end{vmatrix}$$

$$A \text{rot} \hat{\varphi}_{(j,3)} = \begin{vmatrix} A_{11} \hat{\varphi}_{j_1}^{GL} \hat{\varphi}_{j_2}'^{GL} \hat{\varphi}_{j_3}^G - A_{21} \hat{\varphi}_{j_1}'^{GL} \hat{\varphi}_{j_2}'^{GL} \hat{\varphi}_{j_3}^G \\ A_{21} \hat{\varphi}_{j_1}^{GL} \hat{\varphi}_{j_2}'^{GL} \hat{\varphi}_{j_3}^G - A_{22} \hat{\varphi}_{j_1}'^{GL} \hat{\varphi}_{j_2}'^{GL} \hat{\varphi}_{j_3}^G \\ A_{31} \hat{\varphi}_{j_1}^{GL} \hat{\varphi}_{j_2}'^{GL} \hat{\varphi}_{j_3}^G - A_{32} \hat{\varphi}_{j_1}'^{GL} \hat{\varphi}_{j_2}'^{GL} \hat{\varphi}_{j_3}^G \end{vmatrix}$$

Il suffit ensuite de faire le produit scalaire avec le rotationnel de chaque fonction de base  $\varphi_{(k,1)}$ ,  $\varphi_{(k,2)}$  et  $\varphi_{(k,3)}$ . Sur chaque terme obtenu, on fait les simplifications qui en découlent du fait que :

$$\hat{\varphi}_i^{GL}(\hat{\xi}_j^{GL}) = \delta_{i,j}$$

On détaille les calculs pour l'interaction  $(j, 1)$  avec  $(k, 1)$ , on donne l'expression finale pour les autres interactions. L'interaction de  $(j, 1)$  avec  $(k, 1)$ , vaut donc :

$$\begin{aligned} (K_h)_{(j,1),(k,1)} &= \int_{\hat{K}} A_{22} \hat{\varphi}_{j_1}^G \hat{\varphi}_{j_2}^{GL} \hat{\varphi}_{j_3}'^{GL} \hat{\varphi}_{k_1}^G \hat{\varphi}_{k_2}^{GL} \hat{\varphi}_{k_3}'^{GL} \\ &\quad - \int_{\hat{K}} A_{32} \hat{\varphi}_{j_1}^G \hat{\varphi}_{j_2}^{GL} \hat{\varphi}_{j_3}'^{GL} \hat{\varphi}_{k_1}^G \hat{\varphi}_{k_2}'^{GL} \hat{\varphi}_{k_3}^{GL} \\ &\quad - \int_{\hat{K}} A_{32} \hat{\varphi}_{j_1}^G \hat{\varphi}_{j_2}'^{GL} \hat{\varphi}_{j_3}^{GL} \hat{\varphi}_{k_1}^G \hat{\varphi}_{k_2}^{GL} \hat{\varphi}_{k_3}'^{GL} \\ &\quad + \int_{\hat{K}} A_{33} \hat{\varphi}_{j_1}^G \hat{\varphi}_{j_2}'^{GL} \hat{\varphi}_{j_3}^{GL} \hat{\varphi}_{k_1}^G \hat{\varphi}_{k_2}^{GL} \hat{\varphi}_{k_3}'^{GL} \end{aligned}$$

Le premier terme donne après intégration numérique

$$\sum_{m_1, m_2, m_3} A_{22}(\hat{\xi}_m^{GL}) \omega_m^{GL} \hat{\varphi}_{j_1}^G(\hat{\xi}_{m_1}^{GL}) \hat{\varphi}_{j_2}^{GL}(\hat{\xi}_{m_2}^{GL}) \hat{\varphi}_{j_3}'^{GL}(\hat{\xi}_{m_3}^{GL}) \hat{\varphi}_{k_1}^G(\hat{\xi}_{m_1}^{GL}) \hat{\varphi}_{k_2}^{GL}(\hat{\xi}_{m_2}^{GL}) \hat{\varphi}_{k_3}'^{GL}(\hat{\xi}_{m_3}^{GL})$$

On obtient la simplification suivante :

$$j_2 = m_2 = k_2$$

On élimine ainsi deux indices sur 9 indices initiaux. On sait dès maintenant qu'on va obtenir une complexité en  $O(r^7)$ . Peut-on faire mieux ? Oui, on peut faire mieux en utilisant sur ce terme les points d'intégration :

$$(\hat{\xi}_i^G, \hat{\xi}_j^{GL}, \hat{\xi}_k^{GL})$$

On éliminerait ainsi 4 indices au lieu de 2 et on obtient une complexité en  $O(r^5)$ . C'est un cas relativement favorable car les fonctions de base de Gauss sont sur le même indice  $m_1$ . Dans un cas moins favorable (interaction croisée  $(j, 1)$   $(k, 2)$ ), on a ce type de terme :

$$- \sum_{m_1, m_2, m_3} A_{21}(\hat{\xi}_m^{GL}) \omega_m^{GL} \hat{\varphi}_{j_1}^G(\hat{\xi}_{m_1}^{GL}) \hat{\varphi}_{j_2}^{GL}(\hat{\xi}_{m_2}^{GL}) \hat{\varphi}_{j_3}'^{GL}(\hat{\xi}_{m_3}^{GL}) \hat{\varphi}_{k_1}^G(\hat{\xi}_{m_1}^{GL}) \hat{\varphi}_{k_2}^{GL}(\hat{\xi}_{m_2}^{GL}) \hat{\varphi}_{k_3}'^{GL}(\hat{\xi}_{m_3}^{GL})$$

Sur le premier indice  $m_1$ , le choix d'un point de Gauss ou d'un point de Gauss-Lobatto donne une seule simplification. Sur le deuxième indice  $m_2$ , même cas de figure, même punition. Sur le dernier indice  $m_3$ , qu'on mette du Gauss ou du Gauss-Lobatto, on est cuits à cause des deux dérivées conjointes. On a donc, quelles que soient les combinaisons de points qu'on choisit, deux simplifications, et donc forcément une complexité en  $O(r^7)$ .

Par conséquent, il est plus intelligent de choisir des points de Gauss-Lobatto pour tous les indices. Primo, ce choix aboutit à une simplicité d'implémentation, ce qui est agréable. Deuxio, le produit matrice-vecteur rapide, qui se déduit de ces calculs, comporte un nombre minimal de termes lorsqu'on utilise les points de Gauss-Lobatto. Ce point sera développé ultérieurement. Tertio, l'utilisation des points de Gauss-Lobatto ne conduit pas à l'apparition de modes parasites alors que l'utilisation des points de Gauss est néfaste sur des maillages tétraédriques découverts

Revenons à notre premier terme de l'interaction entre  $(j, 1)$  et  $(k, 1)$ . Après simplification, il vaut :

$$\delta_{j_2, k_2} \sum_{m,n} A_{22}(\hat{\xi}_m^{GL}, \hat{\xi}_{j_2}^{GL}, \hat{\xi}_n^{GL}) \omega_{m,j_2,n}^{GL} \hat{\varphi}_{j_1}^G(\hat{\xi}_m^{GL}) \hat{\varphi}_{j_3}^{'GL}(\hat{\xi}_n^{GL}) \hat{\varphi}_{k_1}^G(\hat{\xi}_m^{GL}) \hat{\varphi}_{k_3}^{'GL}(\hat{\xi}_n^{GL})$$

On a trois autres termes :

$$- \sum_m A_{32}(\hat{\xi}_m^{GL}, \hat{\xi}_{k_2}^{GL}, \hat{\xi}_{j_3}^{GL}) \omega_{m,k_2,j_3}^{GL} \hat{\varphi}_{j_1}^G(\hat{\xi}_m^{GL}) \hat{\varphi}_{j_2}^{'GL}(\hat{\xi}_{k_2}^{GL}) \hat{\varphi}_{k_1}^G(\hat{\xi}_m^{GL}) \hat{\varphi}_{k_3}^{'GL}(\hat{\xi}_{j_3}^{GL})$$

$$- \sum_{m,n} A_{32}(\hat{\xi}_m^{GL}, \hat{\xi}_{j_2}^{GL}, \hat{\xi}_{k_3}^{GL}) \omega_{m,j_2,k_3}^{GL} \hat{\varphi}_{j_1}^G(\hat{\xi}_m^{GL}) \hat{\varphi}_{j_3}^{'GL}(\hat{\xi}_{k_3}^{GL}) \hat{\varphi}_{k_1}^G(\hat{\xi}_m^{GL}) \hat{\varphi}_{k_2}^{'GL}(\hat{\xi}_{j_2}^{GL})$$

$$\delta_{j_3, k_3} \sum_{m,n} A_{33}(\hat{\xi}_m^{GL}, \hat{\xi}_n^{GL}, \hat{\xi}_{j_3}^{GL}) \omega_{m,n,j_3}^{GL} \hat{\varphi}_{j_1}^G(\hat{\xi}_m^{GL}) \hat{\varphi}_{j_2}^{'GL}(\hat{\xi}_n^{GL}) \hat{\varphi}_{k_1}^G(\hat{\xi}_m^{GL}) \hat{\varphi}_{k_2}^{'GL}(\hat{\xi}_n^{GL})$$

On adopte les notations suivantes :

$$\text{G\_GL}(i, j) = \hat{\varphi}_i^G(\hat{\xi}_j^{GL})$$

$$\text{dGL\_GL}(i, j) = \hat{\varphi}_i^{'GL}(\hat{\xi}_j^{GL})$$

$$\bar{a}22(i, j, k) = A_{22}(\hat{\xi}_i^{GL}, \hat{\xi}_j^{GL}, \hat{\xi}_k^{GL}) \omega_{i,j,k}^{GL}$$

Le premier terme de la matrice de rigidité est alors égal à :

$$\begin{aligned} (K_h)_{(j,1),(k,1)} &= \delta_{j_2, k_2} \sum_{m,n=1}^{r+1} \bar{a}22(m, j_2, n) \text{G\_GL}(j_1, m) \text{dGL\_GL}(j_3, n) \text{G\_GL}(k_1, m) \text{dGL\_GL}(k_3, n) \\ &\quad - \sum_{m=1}^{r+1} \bar{a}32(m, k_2, j_3) \text{G\_GL}(j_1, m) \text{dGL\_GL}(j_2, k_2) \text{G\_GL}(k_1, m) \text{dGL\_GL}(k_3, j_3) \\ &\quad - \sum_{m=1}^{r+1} \bar{a}32(m, j_2, k_3) \text{G\_GL}(j_1, m) \text{dGL\_GL}(j_3, k_3) \text{G\_GL}(k_1, m) \text{dGL\_GL}(k_2, j_2) \\ &\quad + \delta_{j_3, k_3} \sum_{m,n=1}^{r+1} \bar{a}33(m, n, j_3) \text{G\_GL}(j_1, m) \text{dGL\_GL}(j_2, n) \text{G\_GL}(k_1, m) \text{dGL\_GL}(k_2, n) \end{aligned}$$

Les autres termes de la matrice de rigidité sont égaux à :

$$\begin{aligned} (K_h)_{(j,2),(k,2)} &= \delta_{j_1, k_1} \sum_{m,n=1}^{r+1} \bar{a}11(j_1, m, n) \text{G\_GL}(j_2, m) \text{dGL\_GL}(j_3, n) \text{G\_GL}(k_2, m) \text{dGL\_GL}(k_3, n) \\ &\quad - \sum_{m=1}^{r+1} \bar{a}31(k_1, m, j_3) \text{G\_GL}(j_2, m) \text{dGL\_GL}(j_1, k_1) \text{G\_GL}(k_2, m) \text{dGL\_GL}(k_3, j_3) \\ &\quad - \sum_{m=1}^{r+1} \bar{a}31(j_1, m, k_3) \text{G\_GL}(j_2, m) \text{dGL\_GL}(j_3, k_3) \text{G\_GL}(k_2, m) \text{dGL\_GL}(k_1, j_1) \\ &\quad + \delta_{j_3, k_3} \sum_{m,n=1}^{r+1} \bar{a}33(m, j_2, n) \text{G\_GL}(j_2, n) \text{dGL\_GL}(j_1, m) \text{G\_GL}(k_2, n) \text{dGL\_GL}(k_1, m) \end{aligned}$$

$$\begin{aligned}
(K_h)_{(j,3),(k,3)} &= \delta_{j_1,k_1} \sum_{m,n=1}^{r+1} \bar{a}11(j_1, m, n) G\_GL(j_3, n) dGL\_GL(j_2, m) G\_GL(k_3, n) dGL\_GL(k_2, m) \\
&\quad - \sum_{m=1}^{r+1} \bar{a}21(k_1, j_2, m) G\_GL(j_3, m) dGL\_GL(j_1, k_1) G\_GL(k_3, m) dGL\_GL(k_2, j_2) \\
&\quad - \sum_{m=1}^{r+1} \bar{a}21(j_1, k_2, m) G\_GL(j_3, m) dGL\_GL(j_2, k_2) G\_GL(k_3, m) dGL\_GL(k_1, j_1) \\
&\quad + \delta_{j_2,k_2} \sum_{m,n=1}^{r+1} \bar{a}22(m, j_2, n) G\_GL(j_3, n) dGL\_GL(j_1, m) G\_GL(k_3, n) dGL\_GL(k_1, m) \\
(K_h)_{(j,1),(k,2)} &= - \sum_{m=1}^{r+1} \bar{a}21(k_1, j_2, m) G\_GL(j_1, k_1) dGL\_GL(j_3, m) G\_GL(k_2, j_2) dGL\_GL(k_3, m) \\
&\quad + \sum_{m=1}^{r+1} \bar{a}31(k_1, m, j_3) G\_GL(j_1, k_1) dGL\_GL(j_2, m) G\_GL(k_2, m) dGL\_GL(k_3, j_3) \\
&\quad + \sum_{m=1}^{r+1} \bar{a}32(m, j_2, k_3) G\_GL(j_1, m) dGL\_GL(j_3, k_3) G\_GL(k_2, j_2) dGL\_GL(k_1, m) \\
&\quad - \delta_{j_3,k_3} \sum_{m,n=1}^{r+1} \bar{a}33(m, n, j_3) G\_GL(j_1, m) dGL\_GL(j_2, n) G\_GL(k_2, n) dGL\_GL(k_1, m) \\
(K_h)_{(j,1),(k,3)} &= + \sum_{m=1}^{r+1} \bar{a}21(k_1, j_2, m) G\_GL(j_1, k_1) dGL\_GL(j_3, m) G\_GL(k_3, m) dGL\_GL(k_2, j_2) \\
&\quad - \sum_{m=1}^{r+1} \bar{a}31(k_1, m, j_3) G\_GL(j_1, k_1) dGL\_GL(j_2, m) G\_GL(k_3, j_3) dGL\_GL(k_2, m) \\
&\quad - \delta_{j_2,k_2} \sum_{m,n=1}^{r+1} \bar{a}22(m, j_2, n) G\_GL(j_1, m) dGL\_GL(j_3, n) G\_GL(k_3, n) dGL\_GL(k_1, m) \\
&\quad + \sum_{m=1}^{r+1} \bar{a}32(m, k_2, j_3) G\_GL(j_1, m) dGL\_GL(j_2, k_2) G\_GL(k_3, j_3) dGL\_GL(k_1, m) \\
(K_h)_{(j,2),(k,3)} &= - \delta_{j_1,k_1} \sum_{m,n=1}^{r+1} \bar{a}11(j_1, m, n) G\_GL(j_2, m) dGL\_GL(j_3, n) G\_GL(k_3, n) dGL\_GL(k_2, m) \\
&\quad + \sum_{m=1}^{r+1} \bar{a}31(k_1, m, j_3) G\_GL(j_2, m) dGL\_GL(j_1, k_1) G\_GL(k_3, j_3) dGL\_GL(k_2, m) \\
&\quad + \sum_{m=1}^{r+1} \bar{a}21(j_1, k_2, m) G\_GL(j_2, k_2) dGL\_GL(j_3, m) G\_GL(k_3, m) dGL\_GL(k_1, j_1) \\
&\quad - \sum_{m=1}^{r+1} \bar{a}32(m, k_2, j_3) G\_GL(j_2, k_2) dGL\_GL(j_1, m) G\_GL(k_3, j_3) dGL\_GL(k_1, m)
\end{aligned}$$

Les autres termes de la matrice se déduisent de ceux-ci par symétrie. Une fois cette matrice élémentaire calculée, on rappelle au lecteur qu'il faut prendre en compte les signes des degrés

de liberté locaux par rapport aux degrés de liberté globaux, pour assurer la continuité de la composante tangentielle des inconnues. Typiquement, lorsque le sens de parcours local d'une arête est opposé au sens de parcours global de l'arête, il faut multiplier par -1 la ligne et la colonne de la matrice élémentaire, pour chaque degré de liberté associé à l'arête en question.

Au final, les expressions exhibées nous permettent d'obtenir un calcul de la matrice  $-\omega^2 M_h + K_h$  en  $O(r^7)$ , ce qui est plus intéressant que la complexité en  $O(r^9)$  si on ne fait pas de sous-intégration dans les hexaèdres, ou sur des éléments tétraédriques courbes.

### 5.3 Précision de la méthode

#### Cas 2-D

On étudie la convergence de la solution numérique vers la solution analytique pour un disque parfaitement conducteur de rayon 1 (cf. figure 5.2) Sur maillage régulier, on obtient l'évolution

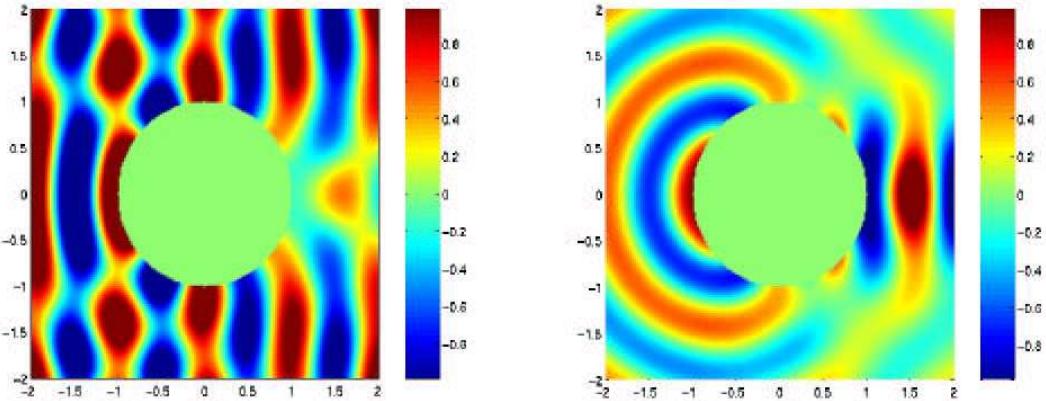


FIG. 5.2 – A gauche, partie réelle du champ total pour un disque de rayon 1. à droite partie réelle du champ diffracté.

de l'erreur de la figure 5.3. On met comme d'habitude en abscisse la quantité  $h/r$ , proportionnelle au nombre de ddl. On voit que  $Q_5$  nécessite moins de ddl que les autres ordres d'approximation pour atteindre une erreur de 0.1%. On fait la même étude sur des maillages non-structurés (cf. figure 5.4). On notera que les éléments d'ordre 1 semblent ne pas converger, l'erreur reste constante. Sur la figure 5.5, on a représenté la solution pour le maillage le plus fin. On voit que les oscillations de la solution sont captées correctement, mais qu'on a une erreur d'interpolation très importante. Les ordres de convergence trouvés sont récapitulés dans le tableau 5.1. On

Ordre d'approximation	1	2	3	4	5
Ordre de convergence, maillage régulier	1.03	2.01	2.97	4.02	4.97
Ordre de convergence, maillage non-structuré	0.06	1.08	2.06	3.04	4.02

TAB. 5.1 – Ordres de convergence mesurés sur les quadrilatères de la première famille

notera que la méthode converge en  $O(h^r)$  en norme H-rot pour des maillages réguliers et en  $O(h^{r-1})$  pour des maillages non-structurés. On a pensé d'abord que cette perte de précision était due à la sous-intégration. Malheureusement, cette conjecture s'est avérée fausse, on obtient les mêmes ordres de convergence lorsqu'on calcule de manière exacte les intégrales. Probablement

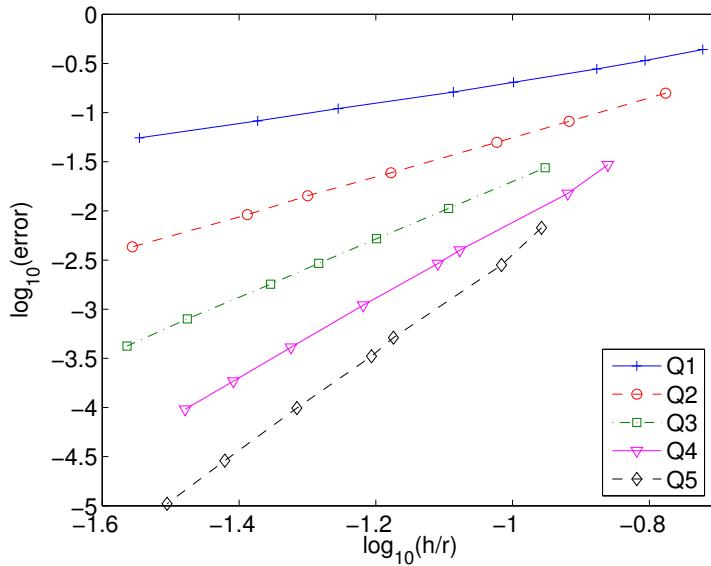


FIG. 5.3 – Evolution de l'erreur H-rot entre la solution numérique et la solution analytique en fonction de  $h/r$ , où  $h$  est le pas de maillage,  $r$  l'ordre d'approximation. Echelle log-log. Cas du disque sur des maillages réguliers.

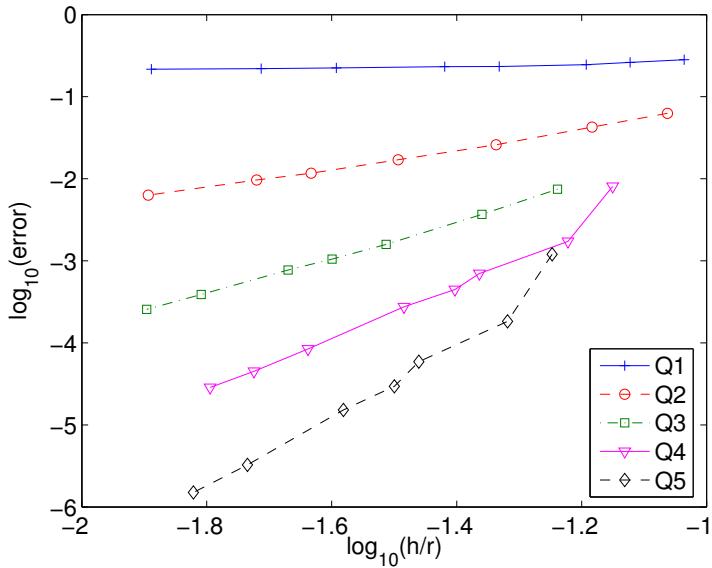


FIG. 5.4 – Evolution de l'erreur H-rot entre la solution numérique et la solution analytique en fonction de  $h/r$ , où  $h$  est le pas de maillage,  $r$  l'ordre d'approximation. Echelle log-log. cas du disque sur des maillages triangulaires découpés.

que cette perte de précision est due à la transformation  $DF_i^{-t}$  qu'on doit appliquer aux fonctions de base.

Comme dans le cas de l'équation de Helmholtz, on s'est posé la question :

“ La présence d'une singularité désavantage-t-elle les ordres d'approximations élevés ? ”. La réponse est non, son illustration est dans le chapitre suivant (cf. figure 6.8). Bien que l'ordre de convergence soit le même pour tous les ordres d'approximation ( $O(h^{4/3})$  pour le carré), la

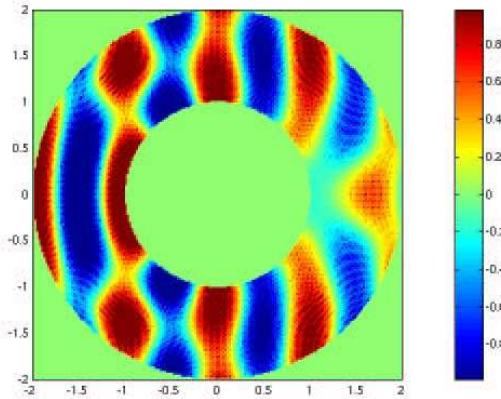


FIG. 5.5 – Partie réelle du champ total pour un disque de rayon 1. Maillage très fin de triangles découpés. On utilise les éléments finis quadrilatéraux de la première famille à l'ordre 1.

constante est bien plus faible pour les ordres élevés, qui ont donc besoin de moins de ddI pour une précision donnée.

### Cas 3-D

On s'intéresse au cas d'une sphère parfaitement conductrice (cf. figure 5.6), on utilise une condition de Silver-Müller sur la frontière extérieure. Le champ électrique  $\mathbf{E}$  diffracté est solution de :

$$-\omega^2 \mathbf{E} + \operatorname{rot} \operatorname{rot} \mathbf{E} = 0 \quad \text{pour } 1 \leq r \leq 2$$

$$\mathbf{E} \times \mathbf{n} = -\mathbf{E}^{\text{inc}} \times \mathbf{n} \quad \text{pour } r = 1$$

$$\operatorname{rot} \mathbf{E} \times \mathbf{n} - i\omega \mathbf{n} \times (\mathbf{E} \times \mathbf{n}) = 0 \quad \text{pour } r = 2$$

$$\mathbf{E}^{\text{inc}} = \exp(-i\omega z) \mathbf{e}_x$$

Le champ électrique total vaut :

$$\mathbf{E}^{\text{tot}} = \mathbf{E}^{\text{inc}} + \mathbf{E}$$

On obtient la courbe de convergence de la figure 5.7. Il semblerait qu'on ait comme en 2-D, une convergence en  $O(h^r)$ , pour la norme H-rot.

On peut également se poser la question de la précision de la méthode sur des maillages tétraédriques découpés. On obtient les courbes de convergence de la figure 5.8. Comme en 2-D, on trouve une convergence en  $h^{r-1}$ , l'ordre 1 n'est pas consistant. Une analyse de dispersion sur maillage non-régulier faite au chapitre 7 tend à démontrer ce résultat.

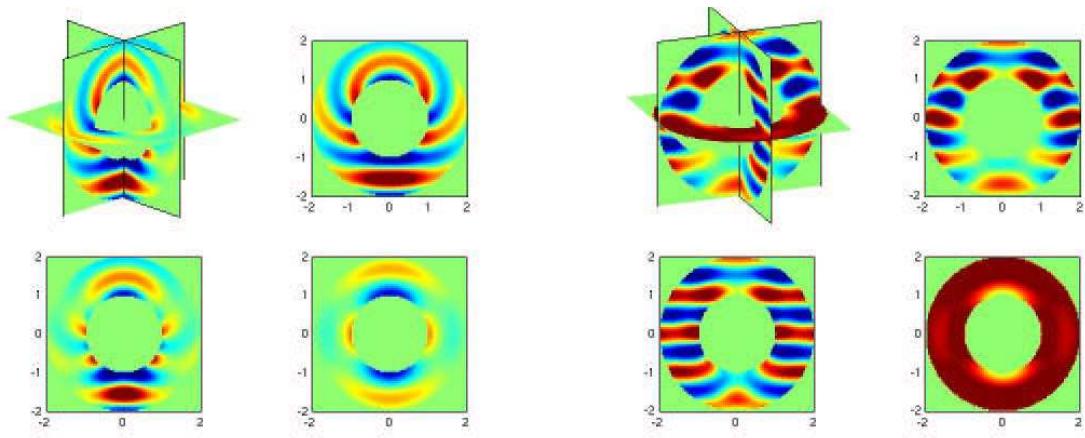


FIG. 5.6 – Solution analytique de la diffraction par une sphère. A gauche, partie réelle de la composante suivant  $x$  du champ diffracté  $E$ . A droite, champ total.

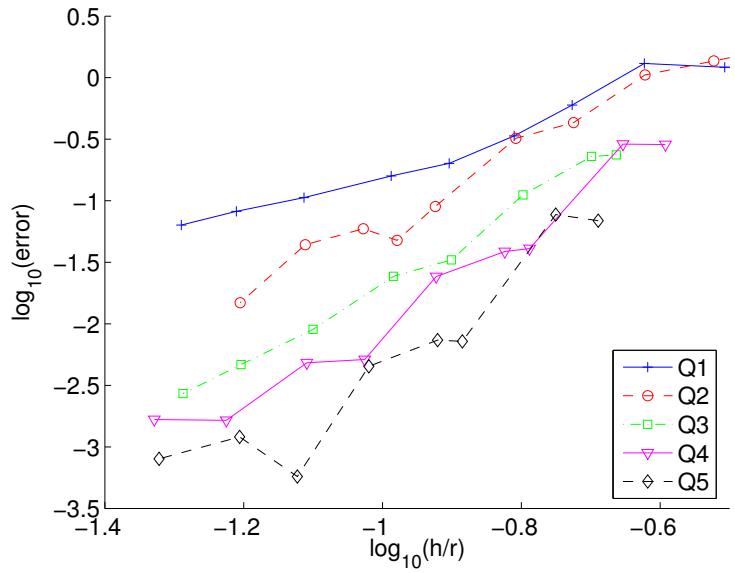


FIG. 5.7 – Evolution de l'erreur H-rot entre la solution numérique et la solution analytique en fonction de  $h/r$ , où  $h$  est le pas de maillage,  $r$  l'ordre d'approximation. Echelle log-log. Cas de la sphère sur des maillages réguliers.

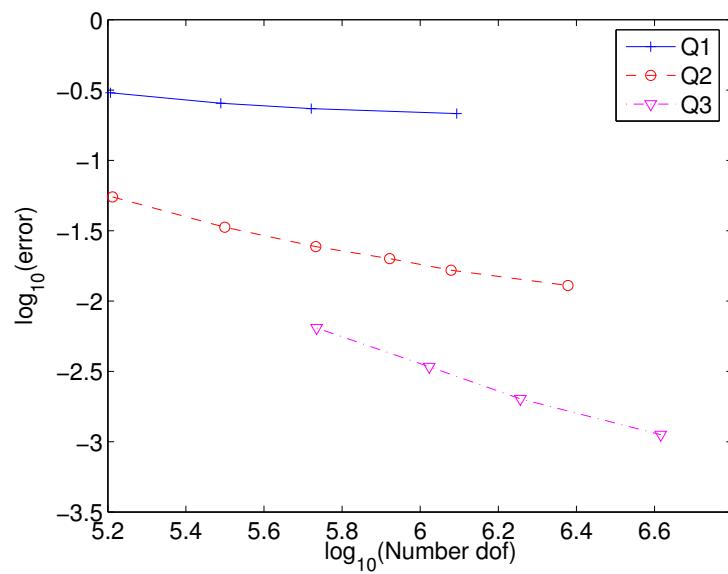


FIG. 5.8 – Evolution de l’erreur H-rot entre la solution numérique et la solution analytique en fonction du nombre de degrés de liberté (échelle log-log). Cas de la sphère avec des maillages tétraédriques découpés

## 5.4 Algorithme rapide du produit matrice-vecteur

### 5.4.1 Factorisation discrète

On s'intéresse au produit matrice vecteur :

$$Y_h = -\omega^2 M_h E_h + K_h E_h$$

On traite uniquement l'hexaèdre  $K_i$ , le vecteur global  $Y_h$  est obtenu par assemblage. Sur l'hexaèdre  $K_i$ , le produit matrice vecteur s'écrit :

$$Y_j = -\omega^2 \sum_{k,m} \omega_m^{GL} B(\hat{\xi}_m^{GL}) \hat{\varphi}_k(\hat{\xi}_m^{GL}) \cdot \hat{\varphi}_j(\hat{\xi}_m^{GL}) E_k + \sum_{k,m} \omega_m^G A(\hat{\xi}_m^G) \text{rot}(\hat{\varphi}_j)(\hat{\xi}_m^G) \cdot \text{rot}(\hat{\varphi}_j)(\hat{\xi}_m^G) E_k$$

en rappelant les expressions des matrices dépendant de la géométrie :

$$B(\hat{x}) = \varepsilon J_i D F_i^{-1} D F_i^{-t}$$

$$A(\hat{x}) = \varepsilon \frac{1}{J_i \mu} D F_i^t D F_i$$

On sépare la double somme en deux sommes simples. La première somme fera en quelque sorte l'évaluation de  $E$  aux points de Gauss-Lobatto et de  $\text{rot}(E)$  aux points de Gauss. La seconde somme fera l'intégration contre les fonctions tests, et leurs rotationnels. On introduit deux vecteurs intermédiaires :

$$\tilde{E}_j = \omega_j^{GL} B(\hat{\xi}_j^{GL}) \sum_k \hat{\varphi}_k(\hat{\xi}_j^{GL}) E_k$$

$$\tilde{H}_j = \omega_j^G A(\hat{\xi}_j^G) \sum_k \text{rot}(\hat{\varphi}_k)(\hat{\xi}_j^G) E_k$$

On notera que  $\tilde{E}_j$  est un vecteur à trois composantes en 3-D. Nous introduisons les matrices diagonales :

$$\bar{B}_{j,k} = B(\hat{\xi}_j^{GL}) \delta_{j,k}$$

$$\bar{A}_{j,k} = A(\hat{\xi}_j^G) \delta_{j,k}$$

Ainsi que les matrices :

$$\hat{C}_{j,k} = \hat{\varphi}_k(\hat{\xi}_j^{GL})$$

$$\hat{R}_{j,k} = \text{rot} \hat{\varphi}_k(\hat{\xi}_j^G)$$

On a donc par construction :

$$\tilde{E} = \bar{B} \hat{C} E$$

$$\tilde{H} = \bar{A} \hat{R} E$$

Le vecteur produit  $Y_h$  est obtenu par la somme :

$$Y_h = -\omega^2 \sum_m \tilde{E}_m \cdot \hat{\varphi}_j(\hat{\xi}_m^{GL}) + \sum_m \tilde{H}_m \cdot \text{rot} \hat{\varphi}_j(\hat{\xi}_m^{GL})$$

On retrouve les matrices  $\hat{C}$  et  $\hat{R}$ , mais transposées ! On peut écrire cette somme sous la forme matricielle :

$$Y = (-\omega^2 \hat{C}^t \bar{B} \hat{C} + \hat{R}^t \bar{A} \hat{R}) E$$

On a ainsi trouvé une factorisation de la matrice de masse et de rigidité. On notera que  $\hat{C}$  et  $\hat{R}$  sont élémentairement creuses et indépendantes de la géométrie. L'information sur la géométrie est stockée dans les matrices diagonales par blocs  $\bar{A}$  et  $\bar{B}$ . En 3-D, on préférera l'utilisation des points de Gauss-Lobatto pour évaluer les intégrales.

### 5.4.2 Formulation mixte

On a ainsi mis en place tous les ingrédients nécessaires à un produit matrice vecteur peu coûteux en stockage et en temps de calcul. La factorisation n'a pas été obtenue par une formulation mixte comme dans les chapitres précédents. Qu'à celà ne tienne, faisons en une! On part du système d'équations :

$$\begin{cases} -\omega^2 \tilde{E} + \text{rot}(H) = 0 \\ \mu H - \text{rot}(E) = 0 \\ \frac{1}{\varepsilon} \tilde{E} = E \end{cases}$$

On établit la formulation variationnelle, en faisant l'intégration par parties sur la première équation :

$$\begin{cases} -\omega^2 \int_{\Omega} \tilde{E} \varphi + \int_{\Omega} H \text{rot} \varphi = 0 \\ \int_{\Omega} \mu H \psi - \int_{\Omega} \text{rot}(E) \psi = 0 \\ \int_{\Omega} \frac{1}{\varepsilon} \tilde{E} \lambda = \int_{\Omega} E \lambda \end{cases}$$

$E$  et sa fonction test associée  $\varphi$ , sont pris dans le même espace  $V_h$  de la formulation standard.  $H$  et sa fonction test associée  $\psi$ , sont pris dans l'espace :

$$W_h = \{u \in (L^2(\Omega))^3 \text{ tel que } DF_i^t u \circ F_i \in Q_{r-1}^3\}$$

On utilise les points de Gauss pour discréteriser cet espace. En 3-D, on préférera l'utilisation des points de Gauss-Lobatto et l'espace d'approximation suivant :

$$W_h = \{u \in (L^2(\Omega))^3 \text{ tel que } DF_i^t u \circ F_i \in Q_r^3\}$$

$\tilde{E}$  et sa fonction test associée  $\lambda$ , sont pris dans l'espace :

$$P_h = \{u \in (L^2(\Omega))^3 \text{ tel que } J_i DF_i^{-1} u \circ F_i \in Q_r^3\}$$

On utilise les points de Gauss-Lobatto pour discréteriser cette espace. Après changement de variables, les matrices élémentaires en jeu sont :

$$\begin{aligned} C'_{j,k} &= \int_{K_i} \lambda_k \cdot \varphi_j = \int_{\hat{K}} \hat{\lambda}_k \cdot \hat{\varphi}_j \\ R'_{j,k} &= \int_{K_i} \psi_k \cdot \text{rot} \varphi_j = \int_{\hat{K}} \hat{\psi}_k \cdot \text{rot} \hat{\varphi}_j \\ A'_{j,k} &= \int_{K_i} \mu \psi_k \cdot \psi_j = \int_{\hat{K}} \mu J_i DF_i^{-1} DF_i^{-t} \hat{\psi}_k \cdot \hat{\psi}_j \\ B'_{j,k} &= \int_{K_i} \lambda_k \cdot \lambda_j = \int_{\hat{K}} \frac{1}{\varepsilon J_i} DF_i^t DF_i \hat{\lambda}_k \cdot \hat{\lambda}_j \end{aligned}$$

Le système linéaire à résoudre est :

$$\begin{cases} -\omega^2 C' \tilde{E} + R' H = 0 \\ A' H - (R')^t H = 0 \\ B' \tilde{E} = (C')^t E \end{cases}$$

On élimine les inconnues intermédiaires  $\tilde{E}$  et  $H$  :

$$(-\omega^2 C'(B')^{-1}(C')^t + R'(A')^{-1}(R')^t) E = 0$$

On obtient la même factorisation que dans le paragraphe précédent. La matrice  $C'$  est égale à la matrice précédemment nommée  $\hat{C}^t$  aux poids de Gauss-Lobatto près. La matrice  $R'$  est égale à la matrice précédemment nommée  $\hat{R}^t$  aux poids de Gauss près. Les matrices  $(A')^{-1}$  et  $(B')^{-1}$  sont égales aux matrices précédemment nommées  $\bar{A}$  et  $\bar{B}$  aux poids d'intégration près. On aboutit ainsi à la même factorisation en utilisant une formulation mixte ou en décomposant le produit matrice vecteur au niveau discret. Néanmoins, il nous semble que la décomposition au niveau discret possède l'avantage d'être une démarche plus "systématique" que l'établissement de la formulation mixte.

### 5.4.3 Produit $\hat{R}E$ et $\hat{C}E$ en 2-D

Explicitons la matrice élémentaire  $\hat{C}$  en 2-D :

$$\hat{C}_{(j,1),(k,1)} = \varphi_{k_1}^G(\hat{\xi}_{j_1}^{GL}) \hat{\varphi}_{k_2}^{GL}(\hat{\xi}_{j_2}^{GL})$$

$(j, 1)$  désigne comme d'habitude une fonction de base orientée suivant  $\mathbf{e}_x$ . Après simplification :

$$\hat{C}_{(j,1),(k,1)} = \varphi_{k_1}^G(\hat{\xi}_{j_1}^{GL}) \delta_{j_2, k_2}$$

Par analogie :

$$\hat{C}_{(j,2),(k,2)} = \varphi_{k_2}^G(\hat{\xi}_{j_2}^{GL}) \delta_{j_1, k_1}$$

Les interactions croisées sont nulles :

$$\hat{C}_{(j,1),(k,2)} = \hat{C}_{(j,2),(k,1)} = 0$$

La matrice élémentaire est creuse, on compte  $r$  termes pour chaque ligne de la matrice, soit au total  $2r(r+1)^2$  éléments non-nuls dans la matrice. On utilise un produit matrice vecteur creux (cf. chapitre 1) pour effectuer le produit matrice vecteur au niveau élémentaire.

Explicitons la matrice élémentaire  $\hat{R}$  en 2-D :

$$\hat{R}_{(j,1),(k,1)} = -\varphi_{k_1}^G(\hat{\xi}_{j_1}^G) \hat{\varphi}_{k_2}^{GL}(\hat{\xi}_{j_2}^G)$$

soit après simplification :

$$\hat{R}_{(j,1),(k,1)} = -\hat{\varphi}_{k_2}^{GL}(\hat{\xi}_{j_2}^G) \delta_{j_1, k_1}$$

De même, on a :

$$\hat{R}_{(j,1),(k,2)} = \hat{\varphi}_{k_1}^{GL}(\hat{\xi}_{j_1}^G) \delta_{j_2, k_2}$$

Ce sont les seules interactions, car en 2-D, l'inconnue  $H$  est scalaire. La matrice élémentaire est creuse, on compte  $2(r+1)$  termes pour chaque ligne de la matrice, soit  $2(r+1)r^2$  éléments non-nuls dans la matrice. On remarquera qu'en 2-D, le coût du produit matrice vecteur est réparti de manière quasi-égale entre la contribution de la matrice de masse et celle de la matrice de rigidité. La contribution de la matrice de masse est même légèrement supérieure.

#### 5.4.4 Produit $\hat{R}E$ et $\hat{C}E$ en 3-D

On adopte les notations suivantes :

$$E = \sum_k \hat{\varphi}_{(k,1)} E_k^1 + \sum_k \hat{\varphi}_{(k,2)} E_k^2 + \sum_k \hat{\varphi}_{(k,3)} E_k^3$$

$$Y = -\omega^2 \hat{C}^t \bar{B} \hat{C} E + \hat{R}^t \bar{A} \hat{R} E$$

On décompose également le vecteur produit :

$$Y = \sum_k \hat{\varphi}_{(k,1)} Y_k^1 + \sum_k \hat{\varphi}_{(k,2)} Y_k^2 + \sum_k \hat{\varphi}_{(k,3)} Y_k^3$$

Explicitons la matrice élémentaire  $\hat{C}$  en 3-D :

$$\hat{C}_{(m,1),(k,1)} = \hat{\varphi}_{k_1}^G(\hat{\xi}_{m_1}^{GL}) \hat{\varphi}_{k_2}^{GL}(\hat{\xi}_{m_2}^{GL}) \hat{\varphi}_{k_3}^{GL}(\hat{\xi}_{m_3}^{GL})$$

Après simplification :

$$\hat{C}_{(m,1),(k,1)} = \hat{\varphi}_{k_1}^G(\hat{\xi}_{m_1}^{GL}) \delta_{m_2, k_2} \delta_{m_3, k_3}$$

Par analogie :

$$\hat{C}_{(m,2),(k,2)} = \hat{\varphi}_{k_2}^G(\hat{\xi}_{m_2}^{GL}) \delta_{m_1, k_1} \delta_{m_3, k_3}$$

$$\hat{C}_{(m,3),(k,3)} = \hat{\varphi}_{k_3}^G(\hat{\xi}_{m_3}^{GL}) \delta_{m_1, k_1} \delta_{m_2, k_2}$$

Les interactions croisées sont nulles :

$$\hat{C}_{(m,1),(k,2)} = \hat{C}_{(m,2),(k,1)} = \hat{C}_{(m,1),(k,3)} = \hat{C}_{(m,3),(k,1)} = \hat{C}_{(m,2),(k,3)} = \hat{C}_{(m,3),(k,2)} = 0$$

La matrice élémentaire est creuse, on compte  $r$  termes pour chaque ligne de la matrice, soit au total  $3r(r+1)^3$  éléments non-nuls dans la matrice. Le produit  $\hat{C}E$  revient à calculer l'interpolation de  $E$  aux points de Gauss-Lobatto :

$$\begin{aligned} w1(m_1, m_2, m_3) &= \sum_{k_1=1}^r G\_GL(k_1, m_1) E_{k_1, m_2, m_3}^1 \\ w2(m_1, m_2, m_3) &= \sum_{k_2=1}^r G\_GL(k_2, m_2) E_{m_1, k_2, m_3}^2 \\ w3(m_1, m_2, m_3) &= \sum_{k_3=1}^r G\_GL(k_3, m_3) E_{m_1, m_2, k_3}^3 \end{aligned}$$

Explicitons la matrice élémentaire  $\hat{R}$  en 3-D, en utilisant les points de Gauss-Lobatto comme points d'intégration :

$$\hat{R}_{(m,2),(k,1)} = \hat{\varphi}_{k_1}^G(\hat{\xi}_{m_1}^{GL}) \hat{\varphi}_{k_3}^{GL}(\hat{\xi}_{m_3}^{GL}) \delta_{m_2, k_2}$$

$$\hat{R}_{(m,3),(k,1)} = -\hat{\varphi}_{k_1}^G(\hat{\xi}_{m_1}^{GL}) \hat{\varphi}_{k_2}^{GL}(\hat{\xi}_{m_2}^{GL}) \delta_{m_3, k_3}$$

$$\hat{R}_{(m,1),(k,2)} = -\hat{\varphi}_{k_2}^G(\hat{\xi}_{m_2}^{GL}) \hat{\varphi}_{k_3}^{GL}(\hat{\xi}_{m_3}^{GL}) \delta_{m_1, k_1}$$

$$\hat{R}_{(m,3),(k,2)} = \hat{\varphi}_{k_2}^G(\hat{\xi}_{m_2}^{GL}) \hat{\varphi}_{k_1}^{GL}(\hat{\xi}_{m_1}^{GL}) \delta_{m_3, k_3}$$

$$\hat{R}_{(m,1),(k,3)} = \hat{\varphi}_{k_3}^G(\hat{\xi}_{m_3}^{GL}) \hat{\varphi}_{k_2}^{GL}(\hat{\xi}_{m_2}^{GL}) \delta_{m_1, k_1}$$

$$\hat{R}_{(m,2),(k,3)} = -\hat{\varphi}_{k_3}^G(\hat{\xi}_{m_3}^{GL}) \hat{\varphi}_{k_1}^{GL}(\hat{\xi}_{m_1}^{GL}) \delta_{m_2, k_2}$$

Les interactions “diagonales” sont nulles :

$$\hat{R}_{(m,1),(k,1)} = \hat{R}_{(m,2),(k,2)} = \hat{R}_{(m,3),(k,3)} = 0$$

On n'est pas très heureux, car on a un seul symbole de kronecker, ce qui donnera une complexité en  $O(r^5)$ , si on utilise un produit matrice-vecteur creux  $\hat{R}E$  classique. Nous ne calculons pas le produit  $\hat{R}E$  de manière standard, nous allons séparer la double somme de chaque interaction élémentaire en simples sommes. On décompose  $\hat{R}$  en six composantes :

$$\hat{R} = \hat{R}_{12} + \hat{R}_{13} + \hat{R}_{21} + \hat{R}_{23} + \hat{R}_{31} + \hat{R}_{32}$$

Le produit matrice vecteur  $\hat{R}_{12}E$  s'écrit :

$$(\hat{R}_{21}E)_{m_1,m_2,m_3} = \sum_{k_1,k_3} \text{G\_GL}(k_1, m_1) \text{dGL\_GL}(k_3, m_3) E_{k_1,m_2,k_3}^1$$

On décompose cette double somme en deux étapes :

$$\begin{aligned} w_{m_1,m_2,m_3}^{21} &= \sum_{k_1=1}^r \text{G\_GL}(k_1, m_1) E_{k_1,m_2,m_3}^1 \\ (\hat{R}_{21}E)_{m_1,m_2,m_3} &= \sum_{k_3=1}^{r+1} \text{dGL\_GL}(k_3, m_3) w_{m_1,m_2,k_3}^{21} \end{aligned}$$

La première étape est de complexité  $2(r+1)^3 r$  alors que la seconde est de complexité  $2(r+1)^4$ . On trouve bien la complexité optimale en  $O(r^4)$ , on met toutes les étapes de calcul pour les autres composantes de  $\hat{R}$  :

$$\begin{aligned} w_{m_1,m_2,m_3}^{31} &= \sum_{k_1=1}^r \text{G\_GL}(k_1, m_1) E_{k_1,m_2,m_3}^1 \\ (\hat{R}_{31}E)_{m_1,m_2,m_3} &= - \sum_{k_2=1}^{r+1} \text{dGL\_GL}(k_2, m_2) w_{m_1,k_2,m_3}^{31} \\ w_{m_1,m_2,m_3}^{12} &= \sum_{k_2=1}^r \text{G\_GL}(k_2, m_2) E_{m_1,k_2,m_3}^2 \\ (\hat{R}_{12}E)_{m_1,m_2,m_3} &= - \sum_{k_3=1}^{r+1} \text{dGL\_GL}(k_3, m_3) w_{m_1,m_2,k_3}^{12} \\ w_{m_1,m_2,m_3}^{32} &= \sum_{k_2=1}^r \text{G\_GL}(k_2, m_2) E_{m_1,k_2,m_3}^2 \\ (\hat{R}_{32}E)_{m_1,m_2,m_3} &= \sum_{k_1=1}^{r+1} \text{dGL\_GL}(k_1, m_1) w_{k_1,m_2,m_3}^{32} \\ w_{m_1,m_2,m_3}^{13} &= \sum_{k_3=1}^r \text{G\_GL}(k_3, m_3) E_{m_1,m_2,k_3}^3 \\ (\hat{R}_{13}E)_{m_1,m_2,m_3} &= \sum_{k_2=1}^{r+1} \text{dGL\_GL}(k_2, m_2) w_{m_1,k_2,m_3}^{13} \end{aligned}$$

$$w_{m_1, m_2, m_3}^{23} = \sum_{k_3=1}^r \text{G\_GL}(k_3, m_3) E_{m_1, m_2, k_3}^3$$

$$(\hat{R}_{23} E)_{m_1, m_2, m_3} = - \sum_{k_1=1}^{r+1} \text{dGL\_GL}(k_1, m_1) w_{k_1, m_2, m_3}^{23}$$

Les sommes finales faites sont :

$$Z_{m_1, m_2, m_3}^1 = (\hat{R}_{12} E)_{m_1, m_2, m_3} + (\hat{R}_{13} E)_{m_1, m_2, m_3}$$

$$Z_{m_1, m_2, m_3}^2 = (\hat{R}_{21} E)_{m_1, m_2, m_3} + (\hat{R}_{23} E)_{m_1, m_2, m_3}$$

$$Z_{m_1, m_2, m_3}^3 = (\hat{R}_{31} E)_{m_1, m_2, m_3} + (\hat{R}_{32} E)_{m_1, m_2, m_3}$$

Le lecteur aura remarqué que les valeurs intermédiaires ont déjà été calculées lors du produit avec la matrice  $\hat{C}$ . On a de fait :

$$w1(m_1, m_2, m_3) = w_{m_1, m_2, m_3}^{21} = w_{m_1, m_2, m_3}^{31}$$

$$w2(m_1, m_2, m_3) = w_{m_1, m_2, m_3}^{12} = w_{m_1, m_2, m_3}^{32}$$

$$w3(m_1, m_2, m_3) = w_{m_1, m_2, m_3}^{13} = w_{m_1, m_2, m_3}^{23}$$

On fait la même démarche pour le produit avec  $\hat{R}^t$ . On peut là aussi regrouper des calculs avec  $\hat{C}^t$ .

La raison est que  $\hat{R}$  peut être factorisée sous la forme :

$$\hat{R} = \hat{S} \hat{C}$$

La matrice  $\hat{S}$  est la matrice de rigidité qu'on avait introduite pour la seconde famille. Les termes non-nuls de  $\hat{S}$  sont :

$$\hat{S}_{(m,2),(k,1)} = \hat{\varphi}_{k_3}^{'GL}(\hat{\xi}_{m_3}^{GL}) \delta_{m_1, k_1} \delta_{m_2, k_2}$$

$$\hat{S}_{(m,3),(k,1)} = -\hat{\varphi}_{k_2}^{'GL}(\hat{\xi}_{m_2}^{GL}) \delta_{m_1, k_1} \delta_{m_3, k_3}$$

$$\hat{S}_{(m,1),(k,2)} = -\hat{\varphi}_{k_3}^{'GL}(\hat{\xi}_{m_3}^{GL}) \delta_{m_1, k_1} \delta_{m_2, k_2}$$

$$\hat{S}_{(m,3),(k,2)} = \hat{\varphi}_{k_1}^{'GL}(\hat{\xi}_{m_1}^{GL}) \delta_{m_2, k_2} \delta_{m_3, k_3}$$

$$\hat{S}_{(m,1),(k,3)} = \hat{\varphi}_{k_2}^{'GL}(\hat{\xi}_{m_2}^{GL}) \delta_{m_1, k_1} \delta_{m_3, k_3}$$

$$\hat{S}_{(m,2),(k,3)} = -\hat{\varphi}_{k_1}^{'GL}(\hat{\xi}_{m_1}^{GL}) \delta_{m_2, k_2} \delta_{m_3, k_3}$$

Pour résumer, on dispose de l'algorithme suivant pour effectuer le produit matrice-vecteur  $\textcolor{red}{Y} = -(\textcolor{green}{\omega}^2 \textcolor{brown}{M}_h + \textcolor{blue}{K}_h) \textcolor{red}{E}$  :

$$\text{Egl} = \hat{C} \textcolor{red}{E}$$

$$\text{Hgl} = \hat{S} \text{Egl}$$

$$H = \bar{A} \text{Hgl}$$

$$\text{Ystiff} = \hat{S}^t H$$

$$\text{Ygl} = -\omega^2 \bar{B} \text{Egl} + \text{Ystiff}$$

$$\textcolor{red}{Y} = \hat{C}^t \text{Ygl}$$

#### 5.4.5 Complexité du produit matrice-vecteur

Dans cette sous-section, on effectue des calculs de complexité au niveau théorique, afin d'évaluer si le produit matrice-vecteur, obtenu à l'aide d'une factorisation discrète, est plus rapide que le produit matrice-vecteur standard. Pour ce dernier, on stocke la matrice globale, et on fait un produit matrice-vecteur creux standard.

#### Cas 2-D

Pour la formulation standard, on compte le nombre d'éléments non nuls de la matrice globale sur un maillage régulier. Ainsi, chaque degré de liberté associé à une arête interagit avec  $(2r+1)r + 2r(r+1)$  degrés de liberté. Chaque degré de liberté associé à l'intérieur d'un élément interagit avec tous les degrés de liberté de l'élément soit  $2r(r+1)$  ddl. Sur maillage régulier, on a  $2rN_e$  degrés de libertés associés aux arêtes et  $2r(r-1)N_e$  degrés de libertés intérieurs.  $N_e$  représente nombre d'éléments du maillage.

Nombre ddl maillage :  $2r^2 N_e$

Nombre d'opérations formulation standard :  $(4r^4 + 8r^3 + 2r^2)N_e$

Stockage formulation standard :  $(r^4 + 2r^3 + 1.5r^2)N_e$

Pour la factorisation discrète, on comptabilise  $2r(r+1)^2$  éléments non-nuls dans la matrice  $\hat{C}$ . On prend en compte deux additions et deux multiplications pour chaque élément de cette matrice (produit avec  $\hat{C}$  et sa transposée), soit quatre opérations. On a également 4 opérations pour chaque élément non-nul de la matrice  $\hat{R}$  ( $2r^2(r+1)$  éléments non-nuls). Pour les matrices dépendant de la géométrie  $\bar{A}$  et  $\bar{B}$ , on compte six opérations pour chaque point de Gauss-Lobatto - multiplication par une matrice 2x2 - et une opération pour chaque point de Gauss - multiplication par un scalaire. On ne stocke que ces matrices, soit trois coefficients par point de Gauss-Lobatto et un coefficient par point de Gauss.

Nombre d'opérations factorisation discrète :  $(16r^3 + 31r^2 + 20r + 6)N_e$

Stockage factorisation discrète :  $(3(r+1)^2 + r^2)N_e$

La factorisation discrète est plus efficace à partir de  $Q_4$ , ce qu'on peut voir sur la figure 5.9.

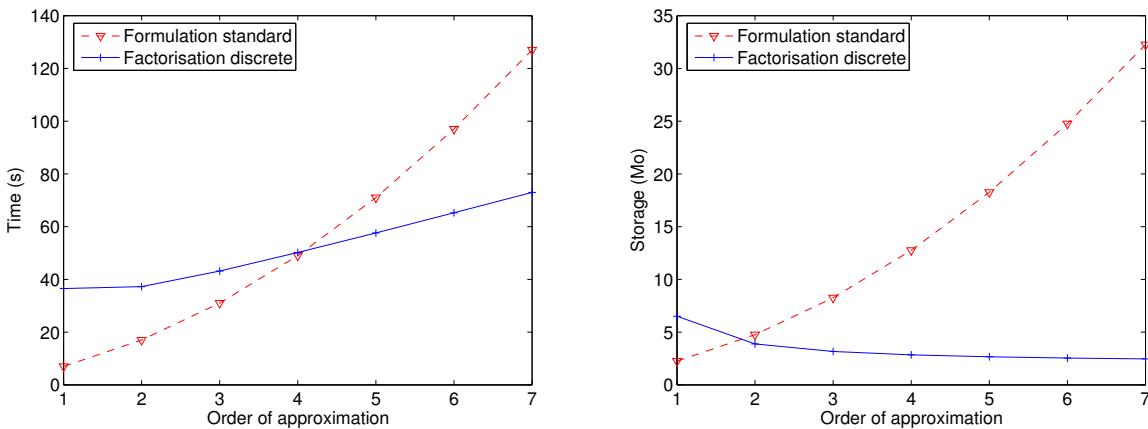


FIG. 5.9 – A gauche temps de calcul en fonction de l'ordre d'approximation, à droite stockage. Première famille sur les quadrilatères. Nombre ddl constant.

### Cas 3-D

Pour la formulation standard, on a pour chaque degré de liberté le nombre d'interactions suivant :

$$\text{Interactions avec un ddl associé à une arête : } 12r^3 + 14r^2 + 6r$$

$$\text{Interactions avec un ddl associé à une face : } 6r^3 + 10r^2 + 4r$$

$$\text{Interactions avec un ddl interne : } 3r^3 + 6r^2 + 3r$$

Sur un maillage hexaédrique régulier, on a la répartition suivante :

$$\text{Nombre ddl associés aux arêtes : } 3rN_e$$

$$\text{Nombre ddl associés aux faces : } (6r^2 - 6r)N_e$$

$$\text{Nombre ddl internes : } (3r^3 - 6r^2 + 3r)N_e$$

On obtient alors :

$$\text{Nombre ddl maillage : } 3r^3N_e$$

$$\text{Nombre d'opérations formulation standard : } (18r^6 + 72r^5 + 84r^4 + 12r^3 + 6r^2)N_e$$

$$\text{Stockage formulation standard : } (4.5r^6 + 18r^5 + 21r^4 + 4.5r^3 + 1.5r^2)N_e$$

Pour la factorisation discrète, les produits avec la matrice  $\hat{C}$  et sa transposée demandent  $12r(r+1)^3$  opérations. Les produits avec la matrice  $\hat{S}$  et sa transposée demandent  $24(r+1)^4$  opérations. Les produits avec les matrices dépendant de la géométrie demandent  $15(r+1)^3 + 15(r+1)^3$  opérations. Au niveau du stockage, on stocke deux matrices  $3 \times 3$  symétriques pour chaque point de Gauss-Lobatto, soit 12 coefficients. Au final, on obtient les complexités suivantes :

$$\text{Nombre opérations factorisation discrète : } (36r^4 + 162r^3 + 270r^2 + 198r + 54)N_e$$

$$\text{Stockage factorisation discrète : } (12(r+1)^3)N_e$$

Asymptotiquement, le produit matrice vecteur coûtera 50% plus cher avec la première famille par rapport à la seconde famille, car le terme prédominant est  $36r^4$ , alors qu'il est de  $24r^4$  pour la seconde famille. Les résultats sont résumés sur la figure 5.10. La factorisation discrète est

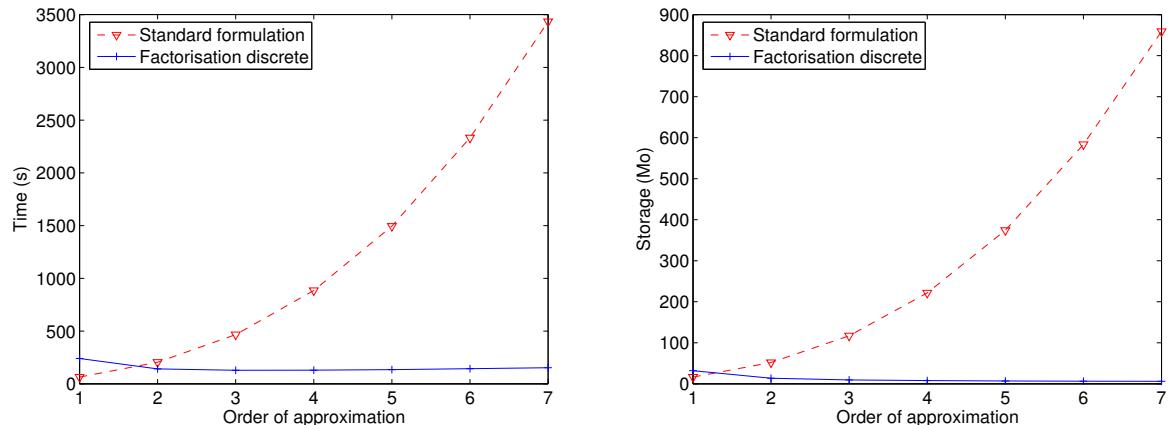


FIG. 5.10 – A gauche temps de calcul en fonction de l'ordre d'approximation, à droite stockage. Première famille sur les hexaèdres. Nombre ddl constant.

plus efficace à partir de  $Q_2$ . On a un coût maîtrisé quand on monte en ordre, alors que le coût explose lorsqu'on utilise une formulation standard. Des expériences numériques confirment cette étude, comme on peut le voir sur le tableau 5.2 L'espace mémoire requis par la factorisation discrète est dérisoire pour  $r$  assez grand. Asymptotiquement, on a besoin de 12 coefficients par point de quadrature, soit 4 coefficients par degré de liberté. Le stockage de la matrice se réduit donc à 4 vecteurs dans le cas général et 2 vecteurs si les indices physiques  $\varepsilon, \mu$  sont réels. Il faut

Ordre d'approximation	1	2	3	4	5	6	7
Temps de calcul formulation standard	<b>55s</b>	<b>127s</b>	224s	380s	631	939s	1380s
Temps de calcul factorisation discrète	244s	128s	<b>106s</b>	<b>97s</b>	<b>96s</b>	<b>98s</b>	<b>103s</b>
Stockage formulation standard	18 Mo	50 Mo	105 Mo	187 Mo	308 Mo	473 Mo	681 Mo
Stockage factorisation discrète	23 Mo	9.9 Mo	6.9 Mo	5.7 Mo	5.0 Mo	4.6 Mo	4.3 Mo

TAB. 5.2 – Comparaison entre la formulation standard et la formulation discrète pour effectuer 1000 itérations du COCG pour un nombre de ddl constant égal à 100 000. Cas des hexaèdres.

comparer ce coût aux 6 vecteurs nécessaires au COCG, et 8 vecteurs pour le BICGCR... On voit clairement que ce sont les vecteurs nécessaires à l'algorithme itératif, qui coûtent le plus. Un autre point positif est que le coût de stockage de la matrice est largement inférieur - d'un rapport 4 environ - au coût de la matrice éléments finis d'ordre 1. En revanche, le temps de calcul est plus élevé, il est environ égal au double. Ce rapport est vrai dans le cas d'indices physiques réels. Dans le cas où ces indices sont complexes, le temps de calcul du produit matrice-vecteur de la matrice éléments finis d'ordre élevé sera du même ordre de grandeur que pour la matrice éléments finis d'ordre 1.

### Intégration exacte

Comme pour le cas de Helmholtz, on peut se poser la question d'un choix de points d'intégration plus précis ( $k + 1$  points de Gauss), pour évaluer les intégrales. La factorisation obtenue avec ces points, est similaire :

$$M_h = \hat{C} B_h \hat{C}^t$$

$$R_h = \hat{C} \hat{S} A_h \hat{S}^t \hat{C}^t$$

avec  $A_h$  et  $B_h$  définies aux points de Gauss, au lieu de Gauss-Lobatto, et :

$$\hat{C}_{j,k} = \hat{\varphi}_j(\hat{\xi}_k^G)$$

$$\hat{S}_{j,k} = \hat{\text{curl}}\hat{\varphi}_j(\hat{\xi}_k^G)$$

Une démonstration de ces factorisations est fournie en annexe C.

Le coût du produit matrice-vecteur sera alors de  $60r^4 + o(r^4)$  au lieu de  $36r^4 + o(r^4)$ . Cependant, l'utilisation de formules d'intégration exacte ou approchée donne le même ordre de convergence. De plus, l'erreur de dispersion est en  $O(h^{2r-2})$  sur des maillages quelconques, comme on le montrera dans le chapitre 7. Il est donc plus avantageux d'évaluer de manière approchée les intégrales, afin de garder un produit matrice-vecteur optimal.

## 5.5 Calcul de modes propres

En 2-D, on obtient tous les modes physiques et aucun mode parasite, contrairement à la seconde famille. Par souci de concision, on ne présentera donc que le cas 3-D. Le problème est la recherche de modes propres dans une cavité cubique :

$$\left\{ \begin{array}{l} \text{trouver } (\omega, \mathbf{E}) \in \mathbb{R} \times H(\text{rot}, \Omega) \quad \mathbf{E} \neq \mathbf{0} \quad \text{tel que} \\ -\omega^2 \mathbf{E} + \text{rot rot } \mathbf{E} = 0 \quad \in \Omega = [-1, 1]^3 \\ \mathbf{E} \times \mathbf{n} = 0 \quad \in \partial\Omega \end{array} \right.$$

Les valeurs propres analytiques d'un cube de côté  $L$  sont connues :

$$\omega_{k,m,n}^2 = \frac{\pi^2 (k^2 + m^2 + n^2)}{L^2} \quad k \geq 0 \quad m > 0 \quad n > 0$$

### 5.5.1 Maillage régulier

On fait le calcul sur un maillage  $3 \times 3 \times 3$  avec une approximation d'ordre 5, on trouve les modes physiques et uniquement les modes physiques. On en affiche quelques-uns sur la figure 5.11. On

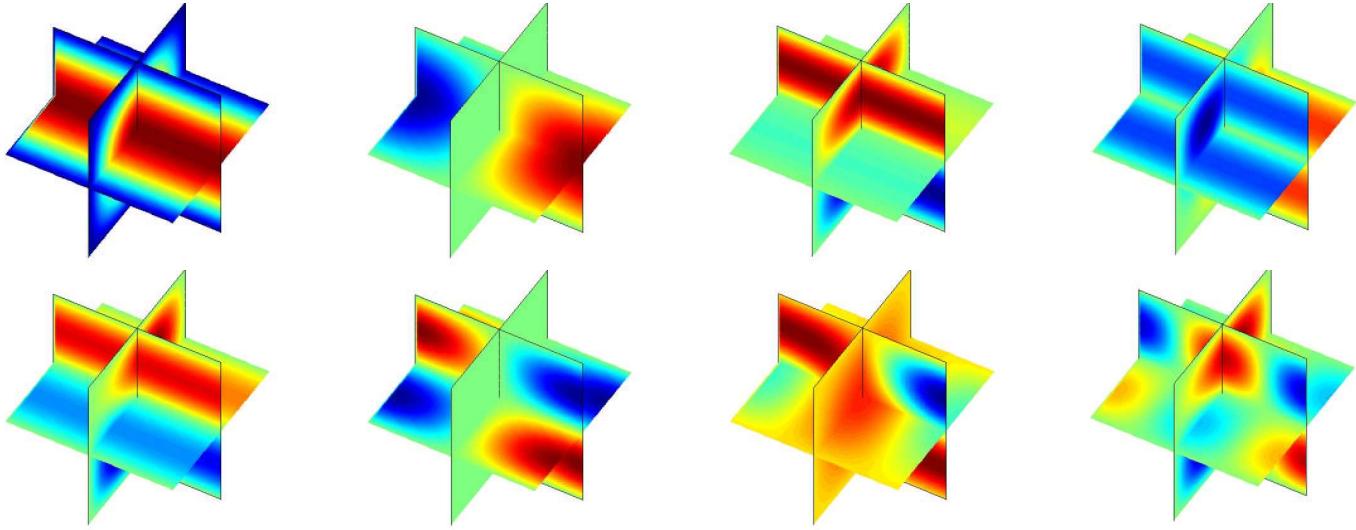


FIG. 5.11 – Quelques modes propres sur un maillage régulier  $3 \times 3 \times 3$  avec du  $Q_5$  (11 500 ddl). Sous-intégration avec des points de Gauss.

voit que sur ce cas, la sous-intégration avec les points de Gauss n'apporte pas de parasites.

### 5.5.2 Maillage non-régulier

Sur un maillage tétraédrique découpé, il en va autrement. Sur la figure 5.12, on a disposé certains modes propres trouvés, un certain nombre est constitué de modes parasites. Le nombre de modes parasites est de plus en plus important lorsqu'on diminue l'ordre d'approximation,  $Q_1$  étant le plus touché. De plus, les valeurs propres parasites s'éloignent assez "lentement" lorsqu'on raffine le maillage, et de fait elles polluent la solution pour des pas de maillages utilisés en pratique. Il est donc préférable d'utiliser les points de Gauss-Lobatto (on sous-intègre toujours la matrice de rigidité!). Comme on peut le voir sur la figure 5.13, l'utilisation des points de Gauss-Lobatto pour intégrer la matrice de rigidité enlève totalement les modes parasites, et ce pour tous les ordres d'approximation. Pour synthétiser la différence entre l'intégration avec points de Gauss et points de Gauss-Lobatto, on affiche le spectre obtenu par les deux intégrations sur la figure 5.14.

## 5.6 Préconditionnement du système linéaire

Comme dans le cas de l'équation de Helmholtz, la matrice éléments finis obtenue est très mal conditionnée, le conditionnement se détériore lorsqu'on raffine le maillage, lorsqu'on augmente la fréquence et lorsqu'on utilise des maillages de qualité médiocre. On a également une difficulté supplémentaire car le noyau de l'opérateur rot-rot est de dimension infinie. De fait, le conditionnement de la matrice sera également mauvais lorsqu'on choisira des fréquences proches de zéro. Dans le présent exposé, on privilégiera le cas haute-fréquence par rapport au cas basse-fréquence, qui nécessite des techniques spécifiques. On aura souvent un domaine de calcul de plus de quatre longueurs d'ondes de diamètre et un maillage en espace autour des huit points par longueur d'onde.

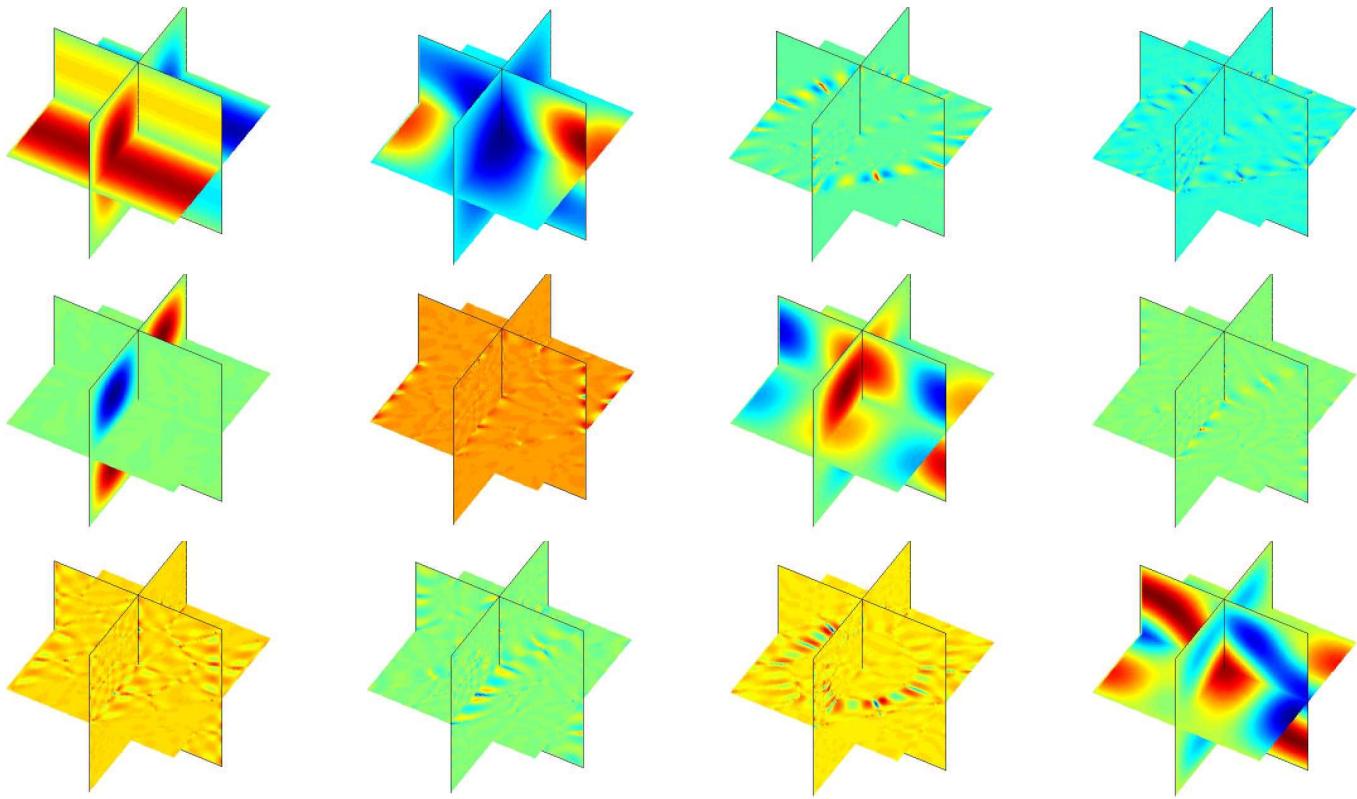


FIG. 5.12 – Quelques modes propres sur un maillage tétraèdrique découpé avec du  $Q_4$  (40 000 dd). Sous-intégration avec des points de Gauss.

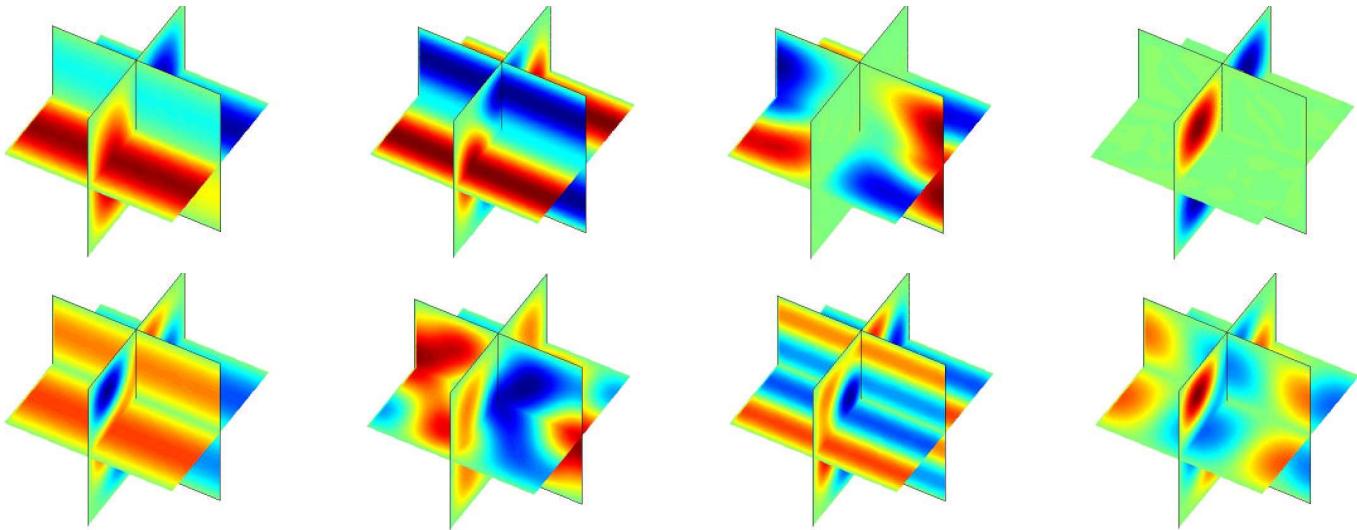


FIG. 5.13 – Quelques modes propres sur un maillage tétraèdrique découpé avec du  $Q_4$  (40 000 dd). Sous-intégration avec des points de Gauss-Lobatto.

Pour pallier au mauvais conditionnement de la matrice, il est nécessaire de préconditionner le système linéaire. On présente ici plusieurs préconditionneurs, dont la factorisation incomplète et le multigrille. Nous n'utiliserons pas de décomposition en sous-domaines, car cette technique n'est pas très intéressante à utiliser dans un code séquentiel (cf. chapitre 2). Pour les lecteurs

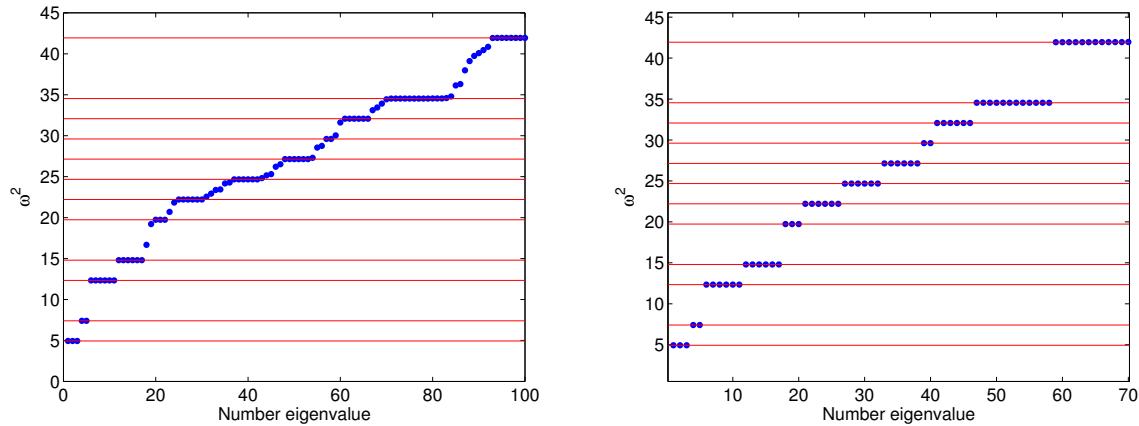


FIG. 5.14 – Distribution des valeurs propres en utilisant les points de Gauss à gauche, et les points de Gauss-Lobatto à droite. Les lignes horizontales symbolisent les valeurs propres analytiques.

qui veulent faire du parallèle, on peut citer [Rappetti et Toselli, 2002] ou [Toselli, 2000].

### 5.6.1 Préconditionnement par un sous-maillage $Q_1$

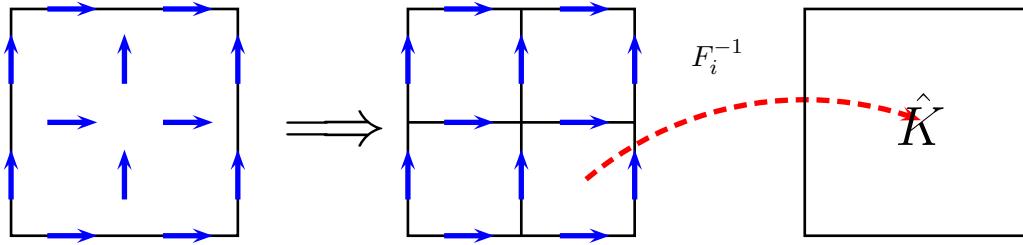


FIG. 5.15 – Sous-maillage  $Q_1$  d'une cellule élémentaire  $Q_2$ . transformation  $F_i$  pour passer du carré de référence vers un petit carré du sous-maillage.

L'idée est similaire à ce qu'on a utilisé pour Helmholtz. On découpe le maillage initial suivant les points de Gauss-Lobatto. On obtient ainsi une correspondance directe entre les degrés de liberté du sous maillage  $Q_1$  et les degrés de liberté initiaux. Une illustration 2-D du procédé est affichée sur la figure 5.15. Contrairement au cas scalaire, il est cependant nécessaire de définir un opérateur de projection, à cause de la transformation  $DF_i^{-t}$ . On note les fonctions de base d'ordre élevé  $\varphi_{(i,s)}$ , et les fonctions de base d'ordre 1  $\psi_{(j,t)}$ . L'indice  $i$  représente la position du degré de liberté tandis que l'indice  $s$  représente son orientation (1 2 ou 3). L'opérateur de projection est une simple injection :

$$P_{i,j} = \psi_{(j,t)}(\xi_i) \cdot \mathbf{e}_s$$

Soit après changement de variables :

$$P_{i,j} = (DF_i^{-t})_{s,t} \hat{\psi}_j(\hat{\xi}_{(i,s)})$$

Or la fonction de base d'ordre 1 a pour composante tangentielle 1 sur l'arête à laquelle elle est associée, et 0 sur les autres arêtes. On a par conséquent :

$$P_{i,j} = (DF_i^{-t})_{s,t} \delta_{i,j}$$

La transformation  $DF_i$  transforme le cube de référence  $\hat{K}$  en un petit pavé, issu du découpage du cube de référence sur les points de Gauss-Lobatto (voir figure 5.15). Les deux sommets opposés du pavé sont :

$$(\hat{\xi}_{i_1}^{GL}, \hat{\xi}_{i_2}^{GL}, \hat{\xi}_{i_3}^{GL}) \quad \text{et} \quad (\hat{\xi}_{i_1+1}^{GL}, \hat{\xi}_{i_2+1}^{GL}, \hat{\xi}_{i_3+1}^{GL})$$

On en déduit l'expression de  $DF_i$ , qui est donc diagonale :

$$DF_i = \begin{pmatrix} \hat{\xi}_{i_1+1}^{GL} - \hat{\xi}_{i_1}^{GL} & 0 & 0 \\ 0 & \hat{\xi}_{i_2+1}^{GL} - \hat{\xi}_{i_2}^{GL} & 0 \\ 0 & 0 & \hat{\xi}_{i_3+1}^{GL} - \hat{\xi}_{i_3}^{GL} \end{pmatrix}$$

En définitive, la matrice de projection est diagonale, du moment que la numérotation des ddl du sous-maillage  $Q_1$  coïncide avec la numérotation des ddl du maillage original  $Q_k$ . Si ce n'est pas le cas, il faut en plus introduire une matrice de permutation. Le préconditionneur vaut donc :

$$M_h = P_h A_h^{-1} P_h^T$$

où  $A_h$  est la matrice éléments finis du sous-maillage  $Q_1$ . On a bien un préconditionneur symétrique. On utilise un solveur direct pour résoudre les systèmes linéaires  $A_h X = B$ .

On s'intéresse dans un premier temps à la diffraction d'une sphère parfaitement conductrice (cf. figure 5.16). On prend un maillage avec 8/10 points par longueur d'onde pour une fréquence de 1.0, et on note le nombre d'itérations suivant la fréquence, sur le tableau 5.3. On voit ici

Ordre	$F = 0.125$	$F = 0.25$	$F = 0.5$	$F = 1.0$	$F = 1.5$
$Q_2(110\ 000\text{ddl})$	$NC$	49	19	16	49
$Q_4(92\ 000\text{ddl})$	$NC$	$NC$	42	30	123
$Q_6(72\ 000\text{ddl})$	$NC$	$NC$	71	47	159

TAB. 5.3 – Nombre d'itérations du BICGCR pour une sphère parfaitement conductrice, pour différentes fréquences et avec un même maillage. Le préconditionneur utilisé est la matrice éléments finis d'ordre 1.

que ce préconditionneur n'est pas adapté au cas basse-fréquence. En effet lorsque le maillage est beaucoup trop fin vis-à-vis de la fréquence considérée, l'algorithme préconditionné ne converge pas au bout de 1000 itérations. C'est un préconditionneur qui donne des résultats très satisfaisants lorsqu'on est dans le régime de fonctionnement (fréquence de 1). Sans surprise, on a une détérioration lorsqu'on monte en ordre.

Dans un second temps, on étudie le préconditionneur pour le cas d'un point source dans une cavité cubique  $[-2, 2]^3$ , on fait évoluer la fréquence EN adaptant le maillage pour avoir toujours 10 points par longueur d'onde, pas plus, pas moins. On obtient les résultats du tableau 5.4. On a choisi le cas de la cavité car c'est un cas plus difficile que la sphère parfaitement conductrice. On peut ainsi clairement observer que le nombre d'itérations croît linéairement en fonction de la fréquence. La difficulté du cas masque les différences entre les ordres d'approximation, alors qu'on les voyait sur la sphère parfaitement conductrice.

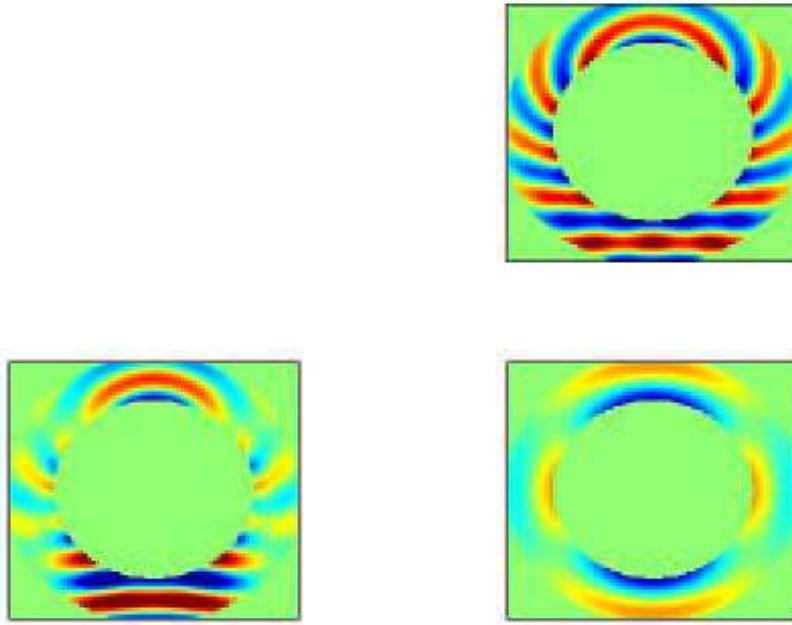


FIG. 5.16 – Partie réelle du champ diffracté par une sphère parfaitement conductrice.

Ordre	$F = 0.25$	$F = 0.5$	$F = 0.8$
$Q_2$	30	74	124
$Q_4$	27	72	131

TAB. 5.4 – Nombre d’itérations du BICGCR pour une cavité cubique, pour différentes fréquences, avec un maillage adapté. Le préconditionneur utilisé est la matrice éléments finis d’ordre 1.

### 5.6.2 Préconditionnement à l’aide d’une factorisation incomplète

Au lieu d’utiliser un solveur direct pour résoudre le système linéaire de la matrice éléments finis d’ordre 1, on propose dans cette sous-section d’utiliser une factorisation incomplète. Pour que cette technique marche, il est nécessaire (cf. chapitre 2), de rajouter de l’amortissement, et donc de calculer la matrice d’ordre 1,  $A_h$  avec un indice physique :

$$\tilde{\varepsilon} = \varepsilon(\alpha + i\beta)$$

Dans le chapitre 2, on avait vu que le choix  $(\alpha, \beta) = (1, 0.5)$  était satisfaisant, ainsi que le choix d’un seuil de 0.01. Sur le tableau 5.5, on effectue quelques tests afin de déterminer le bon choix de paramètres. On voit que la factorisation incomplète échoue quasiment tout le temps si on ne met pas d’amortissement. Même un seuil de  $1e - 3$  est insuffisant ! Quand on ajoute de l’amortissement, la factorisation incomplète donne de bons résultats et mène à des gains importants de stockage, allant jusqu’à diviser par 10 le coût mémoire. Le choix entre  $\beta = 0.5$  et  $\beta = 1$  est cornélien, il faut choisir entre être rapide ou être peu cher en mémoire. Comme ce dernier point nous semble relativement important, on choisit de prendre  $\alpha = 1$   $\beta = 1$

Seuil	$1e - 4$	$1e - 3$	0.01	0.05	0.08	0.1
$\alpha = 1 \beta = 0$	$30/370 Mo$	$\infty/350 Mo$	$\infty/340 Mo$	$\infty/326 Mo$	$\infty/318 Mo$	$\infty/314 Mo$
$\alpha = 1 \beta = 0.5$	$55/299 Mo$	$55/242 Mo$	$55/149 Mo$	$82/74 Mo$	$116/55 Mo$	$145/47 Mo$
$\alpha = 1 \beta = 1$	$97/244 Mo$	$97/197 Mo$	$99/108 Mo$	$110/53 Mo$	$133/40 Mo$	$155/34 Mo$

TAB. 5.5 – Nombre d’itérations du BICGCR pour une sphère parfaitement conductrice, pour différents jeux de paramètres. Le préconditionneur utilisé est la factorisation incomplète sur la matrice éléments finis d’ordre 1.

et un seuil de 5e-2. On remarquera qu’il est nécessaire de prendre un seuil bien plus important que pour l’équation de Helmholtz (1e-2) afin de gagner suffisamment en espace mémoire par rapport à un solveur direct.

### 5.6.3 Préconditionnement utilisant la décomposition de Helmholtz

On propose dans cette sous-section d’utiliser le préconditionneur mis en avant dans [Perrussel et al., 2004]. Ce préconditionneur est détaillé pour le cas des éléments tétraédriques de la première famille de Nédélec de plus bas ordre. L’extension aux hexaèdres est immédiate, mais l’extension pour des ordres d’approximation élevés semble délicate. Le préconditionneur utilise la décomposition de Helmholtz :

$$E_h = \nabla \phi_h + \tilde{E}_h$$

avec  $\phi_h$  potentiel scalaire exprimé dans la base des éléments finis nœuds et  $\tilde{E}_h$  qui appartient à l’orthogonal du noyau du rotationnel. On construit ainsi un opérateur  $N$  de projection de l’espace des éléments finis nœuds vers l’espace des éléments finis d’arête. Pour les éléments d’ordre 1, il s’écrit

$$N_{i,j} = \begin{cases} -1 & \text{si le point } j \text{ est la première extrémité de l’arête } i \\ 1 & \text{si le point } j \text{ est la seconde extrémité de l’arête } i \\ 0 & \text{sinon} \end{cases}$$

On assemble également la matrice éléments finis nœuds :

$$A_\phi = N^t A_h N$$

Le préconditionneur  $x_n = M r_n$  s’écrit alors :

$$x_n = 0 \quad x_\phi = 0$$

$$r_\phi = N^t (r_n - A_h x_n)$$

Descente de Gauss-Seidel sur le système  $A_\phi x_\phi = r_\phi$

$$x_n = x_n + N x_\phi$$

Descente et remontée de Gauss-Seidel sur le système  $A x_n = r_n$

$$x_\phi = 0$$

$$r_\phi = N^t (r_n - A_h x_n)$$

Remontée de Gauss-Seidel sur le système  $A_\phi x_\phi = r_\phi$

$$x_n = x_n + N x_\phi$$

L’auteur alterne les descentes et remontées de Gauss-Seidel afin d’avoir un préconditionneur

symétrique. On se contente ici d'utiliser ce préconditionneur sur le sous-maillage  $Q_1$ . Les cas considérés dans [Perrussel *et al.*, 2004] sont plutôt des cas basses-fréquences, où le préconditionneur semble particulièrement efficace. On a disposé sur le tableau 5.6, le nombre d'itérations selon la fréquence pour un pas de maillage fixé, adapté pour la fréquence 1. Pour des fréquences faibles, le gain en nombre d'itérations est impressionnant. Néanmoins, ce préconditionneur souffre d'in-

Ordre	$F = 0.125$	$F = 0.25$	$F = 0.5$	$F = 0.75$	$F = 1$
Pas de préconditionneur	1262	1157	1253	1373	1250
Helmholtz ( $\alpha = 1.0 \quad \beta = 0$ )	86	148	272	517	2750
Helmholtz ( $\alpha = 1.0 \quad \beta = 1$ )	117	140	197	252	361

TAB. 5.6 – Nombre d'itérations du BICGCR préconditionné en utilisant la décomposition de Helmholtz, en ajoutant ou non de l'amortissement. L'élément fini utilisé est  $Q_1$  avec 110 000 ddl environ.

stabilités pour des fréquences assez élevées. Typiquement, si je prends huit points par longueur d'onde sur un maillage  $Q_1$  pour la sphère parfaitement conductrice et une fréquence de 1, l'algorithme itératif ne converge pas (il met 2750 itérations avec 10 points par longueur d'onde). Lorsqu'on regarde de plus près l'effet du préconditionneur sur ce maillage, on s'aperçoit que la norme du vecteur après application du préconditionneur est multiplié par  $10^{12}$  ! La précision machine est de  $10^{-16}$  environ, insuffisante pour que ce type d'instabilité soit contrôlée par l'algorithme itératif. Afin de stabiliser ce préconditionneur, il est indispensable de rajouter de l'amortissement, comme dans le cas de la factorisation incomplète. Comme on peut voir sur le tableau 5.6, l'amortissement fait gagner en nombre d'itérations surtout pour des fréquences élevées. Pour le cas basse-fréquence, le nombre d'itérations est légèrement supérieur, sans qu'il y ait matière à crier au drame.

Lorsqu'on utilise ce préconditionneur pour des méthodes d'ordre élevé (en passant par le sous-maillage), on perd ses bonnes propriétés de convergence pour le cas basse-fréquence. Mais comme on peut le voir sur le tableau 5.7, son efficacité est correcte dans le cas haute-fréquence. Il faut comparer les 361 itérations pour  $Q_1$  (fréquence 1) et les 406, 456, et 518 itérations pour respectivement  $Q_2$ ,  $Q_4$  et  $Q_6$ . De plus, quand on monte en ordre, le coût du préconditionneur reste constant car on passe par une matrice d'ordre 1. Le ratio coût préconditionneur/coût produit matrice vecteur diminue en conséquence. Théoriquement, le préconditionneur coûte moins de trois produits matrice-vecteur pour l'ordre un, ce ratio sera plus proche de 1.5 pour des ordres d'approximation élevés. Mais il faut prendre en compte qu'on a été obligé de rajouter une partie imaginaire à  $\varepsilon$ , on a donc un coût deux fois plus élevé si les indices physiques sont réels.

Ordre	$F = 0.25$	$F = 0.5$	$F = 1.0$	$F = 1.5$
$Q_2(110\,000\text{ddl})$	399	270	406	592
$Q_4(92\,000\text{ddl})$	$NC$	438	456	$NC$
$Q_6(72\,000\text{ddl})$	$NC$	589	518	$NC$

TAB. 5.7 – Nombre d'itérations du BICGCR pour une sphère parfaitement conductrice, pour différentes fréquences et avec un même maillage. Le préconditionneur utilisé est la décomposition de Helmholtz sur la matrice éléments finis d'ordre 1.

Si on veut garder un bon comportement en basse fréquence, il est nécessaire d'effectuer la décomposition de Helmholtz directement sur les fonctions de base d'ordre élevé. Cette solution

est difficile à mettre en oeuvre, et surtout le gain espéré est quasi-nul, parce que le coût du préconditionneur sera très important car le préconditionneur utilise du Gauss-Seidel, pour lequel il est indispensable de stocker toute la matrice ! Comme on l'a vu précédemment, le coût de stockage et de temps de calcul devient assez vite prohibitif lorsqu'on monte en ordre.

#### 5.6.4 Multigrille

On utilise un algorithme multigrille (cf. chapitre 2), comme préconditionneur. Cette approche a été abordée par quelques auteurs, dont [Hiptmair, 1998], [Beck et Hiptmair, 1999], [Gopalakrishnan *et al.*, 2004], [Perrussel, 2005]. La discréttisation choisie est pour tous ces auteurs des éléments finis tétraédriques ou hexaédriques de plus bas degré. De plus, quand ils font des hexaèdres, ce sont en fait des cubes : forcément les hexaèdres quelconques de plus bas degré ne sont pas consistants ! Les premiers utilisent plutôt du multigrille géométrique tandis que le dernier fait du multigrille algébrique. Nous nous limiterons ici au multigrille géométrique. Nous utilisons, comme dans les autres préconditionneurs, un sous-maillage  $Q_1$  intermédiaire, sur lequel on applique une itération multigrille.

Pour l'opérateur de prolongement, on utilise l'injection classique, à savoir :

$$P_{(i,s),(j,t)} = \psi_{(j,t)}^c(\xi_i^f) \cdot \mathbf{e}_s$$

où on a noté  $\psi_{(j,t)}^c$  les fonctions de base du maillage grossier, et  $\xi_i^f$  les points associés aux degrés de liberté du maillage fin. L'opérateur de restriction est choisi égal au transposé de l'opérateur de prolongement. On choisit également un opérateur de post-lissage transposé de l'opérateur de prélassage. On a ainsi un préconditionneur symétrique. Les opérateurs de lissage sont basés sur la décomposition de Helmholtz, comme dans le cas de la section précédente. L'opérateur de prélassage s'écrit :

$$\begin{aligned} x_\phi &= 0 \\ r_\phi &= N^t(r_n - A_h x_n) \\ \text{Relaxation } R_\phi \text{ sur le système } A_\phi x_\phi &= r_\phi \\ x_n &= x_n + N x_\phi \\ \text{Relaxation } R \text{ sur le système } A x_n &= r_n \end{aligned}$$

L'opérateur de post-lissage est le transposé :

$$\begin{aligned} \text{Relaxation } R^t \text{ sur le système } A x_n &= r_n \\ x_\phi &= 0 \\ r_\phi &= N^t(r_n - A_h x_n) \\ \text{Relaxation } R_\phi^t \text{ sur le système } A_\phi x_\phi &= r_\phi \\ x_n &= x_n + N x_\phi \end{aligned}$$

Pour les étapes de relaxation, on pourra utiliser du Gauss-Seidel comme dans la section précédente ou du Jacobi. L'avantage principal de ce dernier est de pouvoir espérer l'utiliser avec des fonctions de base d'ordre élevé de manière efficace, plutôt que de passer par le sous-maillage  $Q_1$ . On n'a pas eu le temps de réaliser cette idée, nous comparons les deux lisseurs sur les tableaux 5.8 et 5.9. On garde l'amortissement  $\alpha = 1$   $\beta = 1$ , pour les raisons précédemment énoncées (stabilité, robustesse). Sur la grille grossière on utilise une factorisation incomplète au lieu d'un solveur direct. Pour  $Q_2$ , on fait du 2-grilles, pour  $Q_4$  du 3-grilles et pour  $Q_6$  du 4-grilles. Dans ce dernier cas, le préconditionneur ne semble pas très efficace ... Lorsqu'on utilise du Jacobi au lieu de Gauss-Seidel, on est plus robuste sur les cas hautes-fréquences, le BICGCR converge toujours. Dans [Gopalakrishnan *et al.*, 2004], il apparaît nécessaire que le pas de maillage du

Ordre	$F = 0.25$	$F = 0.5$	$F = 1.0$	$F = 1.5$
$Q_2(110\ 000\mathrm{ddl})$	87	56	162	560
$Q_4(92\ 000\mathrm{ddl})$	641	80	519	$NC$
$Q_6(72\ 000\mathrm{ddl})$	$NC$	243	630	$NC$

TAB. 5.8 – Nombre d’itérations du BICGCR pour une sphère parfaitement conductrice, pour différentes fréquences et avec un même maillage. Le préconditionneur utilisé est une itération multigrille sur la matrice éléments finis d’ordre 1. Relaxation de Gauss-Seidel.

Ordre	$F = 0.25$	$F = 0.5$	$F = 1.0$	$F = 1.5$
$Q_2(110\ 000\mathrm{ddl})$	62	63	188	575
$Q_4(92\ 000\mathrm{ddl})$	549	95	294	784
$Q_6(72\ 000\mathrm{ddl})$	$NC$	946	540	878

TAB. 5.9 – Nombre d’itérations du BICGCR pour une sphère parfaitement conductrice, pour différentes fréquences et avec un même maillage. Le préconditionneur utilisé est une itération multigrille sur la matrice éléments finis d’ordre 1. Relaxation de Jacobi ( $\omega = 0.5$ ).

maillage grossier soit au moins de 4 points par longueur d’onde (il n’ajoute pas d’amortissement néanmoins). Or pour la fréquence 1.5, le maillage grossier ne respecte pas cette contrainte pour  $Q_2$ ,  $Q_4$  ou  $Q_6$ . Cette contrainte est respectée pour  $Q_2$  avec une fréquence inférieure à 1, pour  $Q_4$  avec une fréquence inférieure à 0.5. Ceci explique en partie la perte d’efficacité du préconditionneur sur certains cas. L’autre facteur en jeu est que pour les fréquences trop basses, le passage par le sous-maillage  $Q_1$  n’est pas robuste. Globalement, l’ajout d’amortissement a tendance à stabiliser les cas qui autrement ne convergeraient pas du tout.

On peut choisir d’autres valeurs pour  $\alpha$  et  $\beta$  plus appropriées, notamment  $\alpha \neq 1$ . L’indice physique du maillage grossier de niveau  $k$  (le maillage grossier de niveau 0 est le maillage fin) est :

$$\tilde{\varepsilon} = \varepsilon (\alpha^{k+1} + \beta \times (\alpha)^k)$$

On diminue ainsi le nombre d’onde lorsqu’on passe sur des maillages de plus en plus grossiers. Nous regardons l’évolution du nombre d’itérations en fonction de  $\alpha$ , en choisissant  $\beta = \frac{\alpha}{2}$  sur la figure 5.17, en choisissant  $\beta = \frac{\alpha}{2}$  sur la figure 5.18. On utilise l’algorithme de Jacobi comme algorithme de relaxation.

## 5.7 Conclusion

Dans ce chapitre, nous avons montré qu’il était possible de réaliser un produit matrice-vecteur efficace sur les quadrilatères/hexaèdres de la première famille. En 2-D, on peut se satisfaire des points de Gauss-Lobatto et des points de Gauss pour intégrer respectivement la matrice de masse et de rigidité. En 3-D, ce choix donne lieu à des modes parasites, il est préférable d’utiliser les points de Gauss-Lobatto pour intégrer les deux matrices. Avec ce choix, nous avons une méthode vierge de modes parasites.

En outre, nous avons présenté quelques préconditionneurs efficaces pour ce type de discréttisation. La plupart de ces préconditionneurs sont basés sur les équations de Maxwell avec amortissement, afin d’obtenir des algorithmes stables. Dans le chapitre sept, nous montrerons des expériences

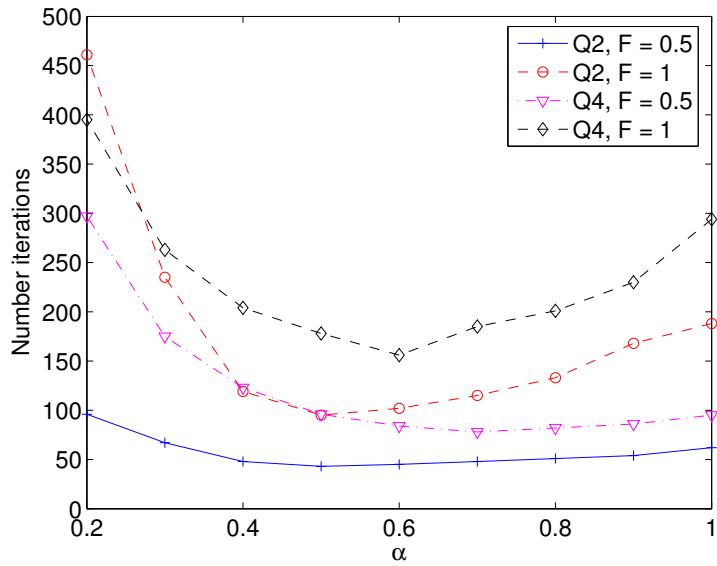


FIG. 5.17 – Nombre d’itérations du BICGCR préconditionné par une itération multigrille, en fonction de  $\alpha$ . On choisit  $\beta = \alpha$ .

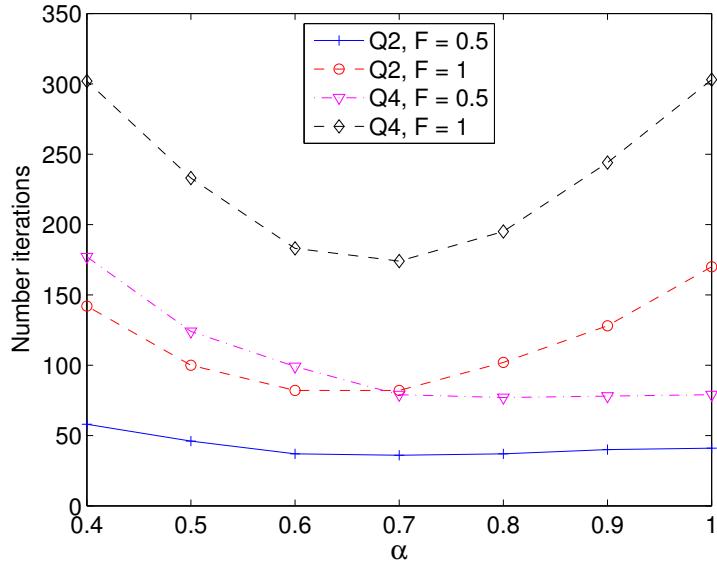


FIG. 5.18 – Nombre d’itérations du BICGCR préconditionné par une itération multigrille, en fonction de  $\alpha$ . On choisit  $\beta = \frac{\alpha}{2}$

numériques sur des cas plus complexes.

# Chapitre 6

## Méthode de Galerkin discontinue sur les quadrilatères/hexaèdres

*Face aux difficultés rencontrées par la seconde famille, on a tout d'abord pensé à utiliser une formulation Galerkin discontinue, utilisant toujours le même espace d'approximation local  $Q_r^d$ . On garde ainsi la condensation de masse, la matrice de rigidité creuse... Une telle approche est très satisfaisante dans le domaine temporel, elle n'est pas très heureuse dans le régime fréquentiel, où le nombre de degrés de liberté est une quantité capitale lors de la résolution du système linéaire. Dans la première section, on explicite la discréétisation choisie et le produit matrice-vecteur rapide qui en découle. Dans la deuxième section, on montre qu'il est nécessaire d'ajouter un terme de pénalisation afin d'obtenir une méthode spectralement correcte.*

### Sommaire

---

<b>6.1 Description de la formulation Galerkin discontinue . . . . .</b>	<b>148</b>
6.1.1 Formulation variationnelle . . . . .	148
6.1.2 Expression des matrices . . . . .	149
6.1.3 Termes de pénalisation . . . . .	153
6.1.4 Calculs de complexité . . . . .	153
6.1.5 Conditions aux limites . . . . .	154
6.1.6 Résolution du système linéaire . . . . .	155
<b>6.2 Présence de modes parasites ? . . . . .</b>	<b>155</b>
6.2.1 Etude de convergence . . . . .	155
6.2.2 Etude de valeurs propres . . . . .	161
<b>6.3 Conclusion . . . . .</b>	<b>169</b>

---

## 6.1 Description de la formulation Galerkin discontinue

On s'intéresse à une méthode Galerkin discontinue sur les quadrilatères hexaèdres, notre principale source d'inspiration est [Pernet, 2004]. Pour les tétraèdres, le lecteur pourra consulter [Hesthaven et Warburton, 2002], et pour une présentation plus abstraite, [Arnold *et al.*, 2002].

### 6.1.1 Formulation variationnelle

On choisit dans cette section des notations plutôt 3-D, on montrera des résultats numériques 2-D et 3-D. On part du système en  $\mathbf{E}$  et  $\mathbf{H}$  des équations de Maxwell :

$$\begin{cases} -\omega \varepsilon \mathbf{E} - \operatorname{rot} \mathbf{H} = \mathbf{f} \\ -\omega \mu \mathbf{H} - \operatorname{rot} \mathbf{E} = 0 \end{cases}$$

On n'utilise pas tout à fait le même changement de variable que précédemment. On a choisi

$$\bar{\mathbf{H}} = -i\mu_0 \mathbf{H}$$

afin d'avoir une symétrie dans le rôle de  $\mathbf{E}$  et  $\mathbf{H}$ . La formulation Galerkin discontinue s'écrit (cf. [Pernet, 2004]) :

$$-\omega \int_{K_i} \varepsilon \mathbf{E} \cdot \varphi - \int_{K_i} \mathbf{H} \cdot \operatorname{rot} \varphi - \int_{\partial K_i} \{\mathbf{H}\} \cdot \varphi \times \mathbf{n} = \int_{K_i} \mathbf{f} \cdot \varphi \quad (6.1)$$

$$-\omega \int_{K_i} \mu \mathbf{H} \cdot \varphi - \int_{K_i} \operatorname{rot} \mathbf{E} \cdot \varphi - \frac{1}{2} \int_{\partial K_i} [\mathbf{E}] \times \mathbf{n} \cdot \varphi = 0 \quad (6.2)$$

(6.3)

On s'est placé sur un hexaèdre  $K_i$ . Si on considère la frontière commune avec un autre hexaèdre

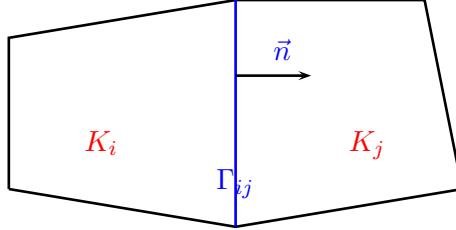


FIG. 6.1 – Interface entre deux mailles élémentaires  $K_i$  et  $K_j$

$K_j$  (cf. figure 6.1), on a choisi les conventions :

$$\{\mathbf{H}\} = \frac{1}{2}(\mathbf{H}_i + \mathbf{H}_j)$$

$$[\mathbf{E}] = (\mathbf{E}_i - \mathbf{E}_j)$$

$\mathbf{n}$  sortante de  $K_i$  vers  $K_j$

On a choisi de faire une intégration par parties sur la première équation et faire apparaître ainsi la moyenne plutôt que le saut de  $\mathbf{H}$ . Ce choix répond au désir de montrer que la formulation variationnelle est symétrique. En effet, les matrices de rigidité sont bien transposées l'une de

l'autre. Il nous reste à montrer que les deux termes de flux sont symétriques. La contribution intérieure des termes de flux est trivialement symétrique :

$$-\frac{1}{2} \int_{\partial K_i} \mathbf{H}_i \cdot \boldsymbol{\varphi}_i \times \mathbf{n}_{i,j} \quad \text{et} \quad -\frac{1}{2} \int_{\partial K_i} \mathbf{E}_i \times \mathbf{n}_{i,j} \cdot \boldsymbol{\varphi}_i$$

En ce qui concerne la contribution extérieure, elle est également symétrique, mais c'est moins immédiat. A priori, on a des termes de signe opposé :

$$-\frac{1}{2} \int_{\partial K_i} \mathbf{H}_j \cdot \boldsymbol{\varphi}_i \times \mathbf{n}_{i,j} \quad \text{et} \quad +\frac{1}{2} \int_{\partial K_i} \mathbf{E}_j \times \mathbf{n}_{i,j} \cdot \boldsymbol{\varphi}_i$$

Mais ces deux termes ne correspondent pas à des termes croisés, en fait il faut chercher les termes transposés sur l'hexaèdre adjacent  $K_j$ . Sur ce dernier, on trouve les contributions extérieures suivantes :

$$+\frac{1}{2} \int_{\partial K_i} \mathbf{H}_i \cdot \boldsymbol{\varphi}_j \times \mathbf{n}_{i,j} \quad \text{et} \quad -\frac{1}{2} \int_{\partial K_i} \mathbf{E}_i \times \mathbf{n}_{i,j} \cdot \boldsymbol{\varphi}_j$$

On vérifie bien la symétrie, le terme en  $\mathbf{H}_j \cdot \boldsymbol{\varphi}_i \times \mathbf{n}$  a pour transposé  $\mathbf{E}_i \times \mathbf{n} \cdot \boldsymbol{\varphi}_j$ , les deux sont égaux (facteur  $-1/2$ ).

L'espace d'approximation pour les deux inconnues est le même :

$$V_h = \{\mathbf{u} \in (L^2(\Omega))^3 \text{ tel que } DF_i^t \mathbf{u} \circ F_i \in Q_r^3\}$$

On a choisi localement le même espace de polynômes que dans le chapitre précédent. La présence de  $DF_i^t$  permet d'obtenir des matrices de rigidité et de sauts indépendantes de la géométrie. On obtient ainsi un produit matrice-vecteur rapide, qu'on explicite dans la suite de ce chapitre. Toutefois, si on omet cette transformation, il est toujours possible d'obtenir un produit matrice-vecteur rapide. La démonstration de ce point est fournie en annexe C. Néanmoins, on obtient un algorithme de calcul légèrement plus lent, c'est pour cette raison qu'on n'a préféré garder la transformation  $DF_i^t$ .

Cette formulation Galerkin discontinue est souvent appelé LDG (Local Discontinuous Galerkin) dans la littérature. C'est par opposition à des formulations du second ordre, a priori plus adapté à une résolution du second ordre. Une présentation des formulations du second ordre - dont IIPG, NIPG et SIPG - est faite dans [Riviere *et al.*, 1999]. La dernière formulation SIPG (Symmetric Interior Penalty Galerkin) aboutit à un système linéaire symétrique, ce qui est primordial en régime harmonique. Elle est utilisée par [Houston *et al.*, 2005] et [Olson et Hesthaven, 2004].

### 6.1.2 Expression des matrices

Pour les éléments  $H^1$  ou  $H(\text{rot})$ , il était nécessaire de prendre les points de Gauss Lobatto pour réaliser à la fois la condensation de masse et respecter la continuité de l'espace d'approximation. Avec la formulation Galerkin discontinue, on a tout un éventail de points qui réalisent la condensation de masse. Nous considérerons deux alternatives :

- Points de Gauss-Lobatto
- Points de Gauss

L'avantage des point de Gauss-Lobatto est d'avoir des termes de flux peu coûteux. L'avantage des points de Gauss est d'avoir une intégration plus précise, et de donner une erreur de dispersion plus faible [Pernet, 2004].

Le système linéaire discret de la formulation Galerkin discontinue s'écrit :

$$\begin{cases} -\omega B_h^1 E_h - R_h H_h - S_h H_h = F_h \\ -\omega B_h^2 H_h - R_h^t E_h - S_h^t E_h = 0 \end{cases}$$

On note les matrices de masse :

$$(B_h^1)_{j,k} = \sum_i \int_{K_i} \varepsilon \boldsymbol{\varphi}_j \cdot \boldsymbol{\varphi}_k$$

$$(B_h^2)_{j,k} = \sum_i \int_{K_i} \mu \boldsymbol{\varphi}_j \cdot \boldsymbol{\varphi}_k$$

la matrice de rigidité, à laquelle on rajoute la contribution intérieure des flux :

$$(R_h)_{j,k} = \sum_i \int_{K_i} \operatorname{rot} \boldsymbol{\varphi}_j \cdot \boldsymbol{\varphi}_k + \frac{1}{2} \int_{\partial K_i} \boldsymbol{\varphi}_k \cdot \boldsymbol{\varphi}_j \times \mathbf{n}$$

la matrice de saut non-locale :

$$(S_h)_{m,n} = \frac{1}{2} \sum_{i,j \text{ voisins}} \int_{\partial K_i \cap \partial K_j} \boldsymbol{\varphi}_n|_{K_j} \cdot \boldsymbol{\varphi}_m|_{K_i} \times \mathbf{n}_{i,j}$$

Explicitons ces différentes matrices, après changement de variable sur  $\hat{K}$ . On adopte la notation :

$$\hat{\boldsymbol{\varphi}}_{(j,s)} = \hat{\varphi}_j \mathbf{e}_s = \hat{\varphi}_{j_1}(\hat{x}_1) \hat{\varphi}_{j_2}(\hat{x}_2) \hat{\varphi}_{j_3}(\hat{x}_3) \mathbf{e}_s$$

Les matrices de masse élémentaires sont identiques à celles du chapitre précédent :

$$(B_h^1)_{(j,s),(k,t)} = \omega_j (J_i D F_i^{-1} \varepsilon D F_i^{-t}) (\hat{\xi}_j) \mathbf{e}_s \cdot \mathbf{e}_t \delta_{j,k}$$

$$(B_h^2)_{(j,s),(k,t)} = \omega_j (J_i D F_i^{-1} \mu D F_i^{-t}) (\hat{\xi}_j) \mathbf{e}_s \cdot \mathbf{e}_t \delta_{j,k}$$

Elles sont diagonales par bloc 3x3. La matrice de rigidité élémentaire est indépendante de la géométrie :

$$(R_h)_{j,k} = \int_{\hat{K}} \hat{\operatorname{rot}} \hat{\boldsymbol{\varphi}}_j \cdot \hat{\boldsymbol{\varphi}}_k + \frac{1}{2} \int_{\partial \hat{K}} \hat{\boldsymbol{\varphi}}_k \cdot \hat{\boldsymbol{\varphi}}_j \times \hat{\mathbf{n}}$$

On utilise les deux égalités :

$$\boldsymbol{\varphi}_k \circ F_i = D F_i^{-t} \hat{\boldsymbol{\varphi}}_k \quad \operatorname{rot} (\boldsymbol{\varphi}_j \circ F_i) = \frac{1}{J_i} D F_i \hat{\operatorname{rot}} \hat{\boldsymbol{\varphi}}_j$$

Le terme surfacique peut s'écrire comme une différence de deux intégrales volumiques indépendantes de la géométrie, il est donc indépendant de la géométrie. La matrice de saut élémentaire est indépendante de la géométrie :

$$(S_h)_{m,n} = \frac{1}{2} \int_{\partial \hat{K}_1 \cap \partial \hat{K}_2} \hat{\boldsymbol{\varphi}}_n|_{\hat{K}_1} \cdot \hat{\boldsymbol{\varphi}}_m|_{\hat{K}_2} \times \hat{\mathbf{n}}_{1,2}$$

où  $\hat{K}_1$  et  $\hat{K}_2$  sont deux cubes unités ayant une face commune.

### Matrice de rigidité élémentaire

Détaillons un peu plus les interactions de la matrice de rigidité élémentaire. On choisit deux fonctions de base orientées suivant  $\mathbf{e}_1$  et  $\mathbf{e}_2$  :

$$\hat{\boldsymbol{\varphi}}_{(j,1)} = \hat{\varphi}_{j_1}(\hat{x}_1) \hat{\varphi}_{j_2}(\hat{x}_1) \hat{\varphi}_{j_3}(\hat{x}_3) \mathbf{e}_1$$

$$\boldsymbol{\varphi}_{(k,2)} = \hat{\varphi}_{k_1}(\hat{x}_1) \hat{\varphi}_{k_2}(\hat{x}_2) \hat{\varphi}_{k_3}(\hat{x}_3) \mathbf{e}_2$$

Le rotationnel de la première fonction de base vaut :

$$\text{rot} \varphi_j = \begin{vmatrix} 0 \\ \hat{\varphi}_{j_1} \hat{\varphi}_{j_2} \hat{\varphi}'_{j_3} \\ -\hat{\varphi}_{j_1} \hat{\varphi}'_{j_2} \hat{\varphi}_{j_3} \end{vmatrix}$$

La partie volumique de la matrice de rigidité vaut :

$$\sum_m \omega_m \hat{\varphi}_{j_1}(\hat{\xi}_{m_1}) \hat{\varphi}_{j_2}(\hat{\xi}_{m_2}) \hat{\varphi}'_{j_3}(\hat{\xi}_{m_3}) \hat{\varphi}_{k_1}(\hat{\xi}_{m_1}) \hat{\varphi}_{k_2}(\hat{\xi}_{m_2}) \hat{\varphi}_{k_3}(\hat{\xi}_{m_3})$$

En utilisant la relation  $\hat{\varphi}_i(\hat{\xi}_j) = \delta_{i,j}$ , on trouve :

$$\omega_{k_1, k_2, k_3} \hat{\varphi}'_{j_3}(\hat{\xi}_{k_3}) \delta_{j_1, k_1} \delta_{j_2, k_2}$$

La partie surfacique de la matrice de rigidité se réduit à deux intégrales sur les faces opposées  $\hat{x}_3 = 0$  et  $\hat{x}_3 = 1$  :

$$\begin{aligned} & -\frac{1}{2} \int_{[0,1]^2} \hat{\varphi}_{j_1}(\hat{x}_1) \hat{\varphi}_{j_2}(\hat{x}_2) \hat{\varphi}_{j_3}(0) \hat{\varphi}_{k_1}(\hat{x}_1) \hat{\varphi}_{k_2}(\hat{x}_2) \hat{\varphi}_{k_3}(0) \text{Det}(\mathbf{e}_y, \mathbf{e}_x, \mathbf{e}_z) d\hat{x}_1 d\hat{x}_2 \\ & + \frac{1}{2} \int_{[0,1]^2} \hat{\varphi}_{j_1}(\hat{x}_1) \hat{\varphi}_{j_2}(\hat{x}_2) \hat{\varphi}_{j_3}(1) \hat{\varphi}_{k_1}(\hat{x}_1) \hat{\varphi}_{k_2}(\hat{x}_2) \hat{\varphi}_{k_3}(1) \text{Det}(\mathbf{e}_y, \mathbf{e}_x, \mathbf{e}_z) d\hat{x}_1 d\hat{x}_2 \end{aligned}$$

Sur les autres faces, la contribution est nulle, à cause du déterminant. En utilisant l'intégration numérique adéquate sur les faces, on trouve la quantité :

$$\frac{1}{2} \omega_{k_1, k_2} (\hat{\varphi}_{j_3}(0) \hat{\varphi}_{k_3}(0) - \hat{\varphi}_{j_3}(1) \hat{\varphi}_{k_3}(1)) \delta_{j_1, k_1} \delta_{j_2, k_2}$$

Finalement, on recense les interactions suivantes pour la matrice de rigidité :

$$\begin{aligned} (R_h)_{(j,1),(k,2)} &= \left[ \omega_{k_3} \hat{\varphi}'_{j_3}(\hat{\xi}_{k_3}) + \frac{1}{2} (\hat{\varphi}_{j_3}(0) \hat{\varphi}_{k_3}(0) - \hat{\varphi}_{j_3}(1) \hat{\varphi}_{k_3}(1)) \right] \omega_{k_1} \omega_{k_2} \delta_{j_1, k_1} \delta_{j_2, k_2} \\ (R_h)_{(j,1),(k,3)} &= - \left[ \omega_{k_2} \hat{\varphi}'_{j_2}(\hat{\xi}_{k_2}) + \frac{1}{2} (\hat{\varphi}_{j_2}(0) \hat{\varphi}_{k_2}(0) - \hat{\varphi}_{j_2}(1) \hat{\varphi}_{k_2}(1)) \right] \omega_{k_1} \omega_{k_3} \delta_{j_1, k_1} \delta_{j_3, k_3} \\ (R_h)_{(j,2),(k,3)} &= \left[ \omega_{k_1} \hat{\varphi}'_{j_1}(\hat{\xi}_{k_1}) + \frac{1}{2} (\hat{\varphi}_{j_1}(0) \hat{\varphi}_{k_1}(0) - \hat{\varphi}_{j_1}(1) \hat{\varphi}_{k_1}(1)) \right] \omega_{k_2} \omega_{k_3} \delta_{j_2, k_2} \delta_{j_3, k_3} \end{aligned}$$

Les autres interactions sont de signes opposés :

$$(R_h)_{(j,2),(k,1)} = -(R_h)_{(j,1),(k,2)}$$

$$(R_h)_{(j,3),(k,1)} = -(R_h)_{(j,1),(k,3)}$$

$$(R_h)_{(j,3),(k,2)} = -(R_h)_{(j,2),(k,3)}$$

On remarquera que le terme surfacique et le terme volumique donnent la même structure creuse de la matrice. Chaque ligne de la matrice élémentaire de rigidité  $R_h$  contient  $2(r+1)$  éléments non-nuls. Mais, on dispose de la propriété suivante (valable pour les points de Gauss et de Gauss-Lobatto) :

$$\omega_k \hat{\varphi}'_k(\hat{\xi}_k) + \frac{1}{2} (\hat{\varphi}_k(0) \hat{\varphi}_k(0) - \hat{\varphi}_k(1) \hat{\varphi}_k(1)) = 0 \quad \forall k$$

Cette propriété nous permet d'affirmer qu'on a en fait  $2r$  éléments non-nuls sur chaque ligne de la matrice. On a choisi de prendre cette définition de la matrice de rigidité (avec la contribution des termes de flux intérieurs), afin d'avoir ce nombre optimal d'éléments non-nuls. Globalement, on aura un algorithme de produit matrice-vecteur optimal. La matrice de rigidité contient donc finalement  $6r(r+1)^3$  entrées non-nulles.

## Termes de flux non-locaux

On considère la face  $\hat{x}_1 = 0$  partagée entre deux hexaèdres de référence  $\hat{K}_1$  et  $\hat{K}_2$ , la fonction test  $\hat{\varphi}_{(j,2)}$  est à support dans  $\hat{K}_1$ , alors que la fonction de base  $\hat{\varphi}_{(k,3)}$  est à support dans  $\hat{K}_2$ . Le terme de flux est alors égal à :

$$\frac{1}{2} \int_{[0,1]^2} \hat{\varphi}_{j_1}(0) \hat{\varphi}_{j_2}(\hat{x}_2) \hat{\varphi}_{j_3}(\hat{x}_3) \hat{\varphi}_{k_1}(1) \hat{\varphi}_{k_2}(\hat{x}_2) \hat{\varphi}_{k_3}(\hat{x}_3) \text{Det}(\mathbf{e}_z, \mathbf{e}_y, \mathbf{e}_x) d\hat{x}_2 d\hat{x}_3$$

Ce qui donne après intégration numérique :

$$-\frac{1}{2} \omega_{j_2} \omega_{j_3} \hat{\varphi}_{j_1}(0) \hat{\varphi}_{k_1}(1) \delta_{j_2, k_2} \delta_{j_3, k_3}$$

On peut mener des calculs similaires sur les autres faces.

Le lecteur aura remarqué que lorsqu'on utilise des points de Gauss-Lobatto :

$$\hat{\varphi}_j(0) = 0 \quad \forall j \in ]0, r+1[$$

La matrice de saut non-locale a par conséquent une interaction non-nulle pour chaque degré de liberté tangentiel de chaque face. Lorsqu'on utilise les points de Gauss, ce n'est plus vrai. Tous les degrés de liberté de  $\hat{K}_1$  orientés suivant  $\mathbf{e}_y$  ou  $\mathbf{e}_z$  vont interagir avec tous les degrés de liberté orientés de  $\hat{K}_2$  suivant  $\mathbf{e}_y$  ou  $\mathbf{e}_z$ , placés sur le même axe parallèle à  $\mathbf{Ox}$ . Ainsi un degré de liberté orienté suivant  $\mathbf{e}_y$  interagira avec :

- $r+1$  degrés de liberté orientés suivant  $\mathbf{e}_z$  de  $\hat{K}_2$
- $r+1$  ddl orientés suivant  $\mathbf{e}_z$  de l'hexaèdre adjacent à la face  $x = 1$
- $r+1$  ddl orientés suivant  $\mathbf{e}_x$  de l'hexaèdre adjacent à la face  $z = 0$
- $r+1$  ddl orientés suivant  $\mathbf{e}_x$  de l'hexaèdre adjacent à la face  $z = 1$ .

Le nombre d'éléments non-nuls de chaque ligne de  $S_h$  devient alors égal à  $4(r+1)$  (interaction avec 4 hexaèdres). On aurait à première vue un produit matrice-vecteur  $S_h X$  deux fois plus coûteux que le produit  $R_h X$  lorsqu'on utilise les points de Gauss ! Heureusement, on dispose d'une seconde vue, et il est assez clair qu'on peut écrire un produit matrice-vecteur rapide pour  $S_h X$ .

En effet, ce produit s'écrit pour la face  $x = 0$  et les ddl orientés suivant  $\mathbf{e}_y$  :

$$(S_h X)_{(j,2)} = -\frac{1}{2} \omega_{j_2} \omega_{j_3} \hat{\varphi}_{j_1}(0) \sum_{k_1} \hat{\varphi}_{k_1}(1) X_{[(k_1, j_2, j_3), 3]}$$

On remarque que la somme ne fait pas intervenir  $j_1$ , on évalue donc dans un premier temps l'extrapolation de  $X$  au point  $(0, \hat{\xi}_{j_2}, \hat{\xi}_{j_3})$  :

$$v_{j_2, j_3} = \sum_{k_1} \hat{\varphi}_{k_1}(1) X_{[(k_1, j_2, j_3), 3]}$$

Et dans un second temps, on calcule le vecteur produit :

$$(S_h X)_{(j,2)} = -\frac{1}{2} \omega_{j_2} \omega_{j_3} \hat{\varphi}_{j_1}(0) v_{j_2, j_3}$$

La première étape, comme la seconde, nécessite  $2(r+1)^3$  opérations (on considère qu'on assemble le produit, on a donc une multiplication et une addition). Chaque face de l'hexaèdre  $\hat{K}_1$  demande  $8(r+1)^3$  opérations. Le produit matrice vecteur  $S_h X$  est donc de complexité en  $O(r^3)$  alors que le produit matrice vecteur  $R_h X$  est en  $O(r^4)$ , donc le coût de calcul des flux devient négligeable si on utilise un ordre d'approximation suffisamment élevé.

### 6.1.3 Termes de pénalisation

De manière classique, les mathématiciens introduisent dans la formulation Galerkin discontinue, des termes de pénalisation, qui leur permettent de montrer des estimations d'erreur optimale. On peut ainsi ajouter à l'équation (6.1) :

$$-i\omega \alpha \int_{\partial K_i} [\mathbf{E} \times \mathbf{n}] \cdot \boldsymbol{\varphi} \times \mathbf{n} dx$$

Et un terme similaire pour l'équation (6.2).

$$-i\omega \delta \int_{\partial K_i} [\mathbf{H} \times \mathbf{n}] \cdot \boldsymbol{\varphi} \times \mathbf{n} dx$$

Les coefficients  $\alpha$  et  $\delta$  doivent être positifs pour correspondre à un amortissement (ce signe est lié à la convention  $-i\omega t$  qu'on a choisi). Dans la suite, on prendra toujours :

$$\delta = 0 \quad \alpha = 0.5$$

En effet, en l'absence d'amortissement sur  $H$ , il est aisément de faire un complément de Schur pour éliminer l'inconnue  $H$ . La valeur 0.5 est classiquement utilisée dans les méthodes Galerkin discontinues. Il existe d'autres valeurs plus "optimales", ce n'est pas notre préoccupation.

Après changement de variables, le terme de pénalisation devient égal à :

$$P_h E = -i\omega \int_{\hat{K}} \frac{1}{ds} DF_i^t DF_i [\hat{\mathbf{E}} \times \hat{\mathbf{n}}] \cdot \hat{\boldsymbol{\varphi}} \times \hat{\mathbf{n}}$$

On choisit comme d'habitude de ne pas expliciter la matrice associée à ce terme. On préfère utiliser un produit matrice vecteur rapide. La première étape est le calcul de  $\hat{\mathbf{E}}$  aux points d'intégration de chaque face (extrapolation de  $E$ ). Cette phase a déjà été réalisée lors du calcul des sauts. La seconde étape est l'application de la transformation géométrique  $\frac{1}{ds} DF_i^t DF_i$ . On calcule cette matrice  $3 \times 3$  sur tous les points de quadrature de chaque face. Pour chaque point de quadrature on doit multiplier cette matrice par la valeur extrapolée de  $\hat{\mathbf{E}} \times \hat{\mathbf{n}}$ . Comme on ne considère que les valeurs tangentielles du champ électrique, on n'a besoin en pratique que d'une sous-matrice  $2 \times 2$ . La troisième et dernière étape est l'intégration contre les fonctions de base. Cette étape peut être regroupée avec le calcul des termes de flux. Pour chaque point de quadrature, on a 2 soustractions (différence entre les composantes tangentielles de  $E_i$  et  $E_j$ ), 4 multiplications et 2 additions pour la multiplication avec la matrice  $2 \times 2$  et deux additions pour rajouter la contribution du terme de pénalisation aux termes de flux. Au total, on a besoin de 10 opérations par point de quadrature.

### 6.1.4 Calculs de complexité

Afin de mieux fixer les idées, il est utile de faire un calcul de complexité afin de comparer les méthodes de Galerkin discontinue (points de Gauss ou Gauss-Lobatto) avec la seconde famille de Nédélec. On récapitule les coûts de calcul en utilisant les points de Gauss-Lobatto :

Coût  $B_h^1 X$  et  $B_h^2 X$  :  $30(r+1)^3 N_e$

Coût  $R_h X$  et  $R_h^t X$  :  $24r(r+1)^3 N_e$

Coût  $S_h X$  et  $S_h^t X$  :  $48(r+1)^2 N_e$

Pour les points de Gauss, ce dernier coût est plus élevé :

Coût  $S_h X$  et  $S_h^t X$  :  $96(r+1)^3 N_e$

On rajoute le coût du terme de pénalisation : Coût  $P_h X$  :  $60(r+1)^2 N_e$

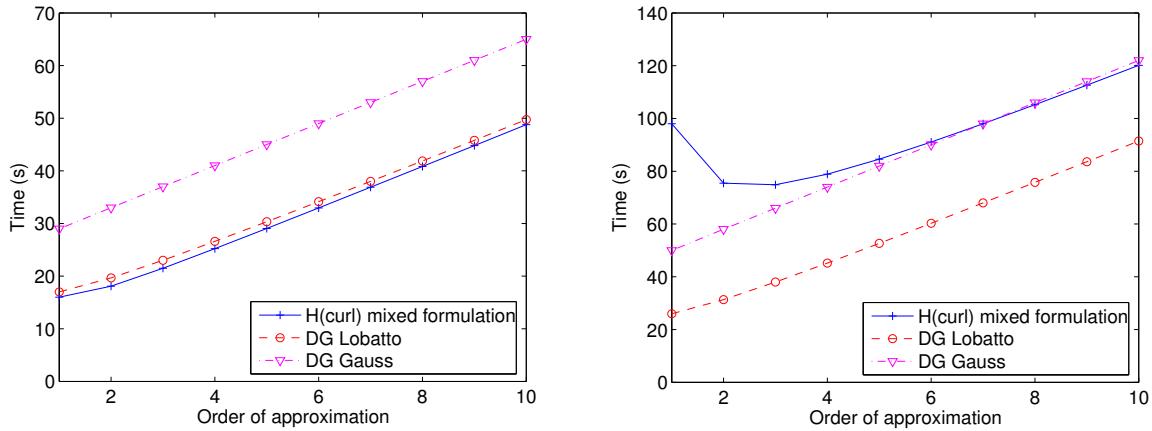


FIG. 6.2 – A gauche temps de calcul en fonction de l’ordre d’approximation pour le cas 2-D, à droite cas 3-D. Nombre de ddl constant.

Le nombre de degrés de liberté est égal à  $3(r+1)^3 N_e$ . Sur la figure 6.2, on a représenté le temps de calcul en fonction de l’ordre d’approximation, lorsqu’on compare à nombre de degrés de liberté constant. Logiquement, on trouve une complexité linéaire pour la formulation Galerkin discontinue, car on n’a plus de continuité de certains degrés de liberté comme dans la formulation  $H^1$  ou  $H(\text{rot})$ . Si on prend un même pas de maillage, le nombre de degrés de liberté en Galerkin discontinue est bien plus important en 3-D qu’en 2-D par rapport aux nombres de degrés de liberté en formulation  $H(\text{rot})$ . Ce point explique, qu’à nombre de ddl constant, la formulation Galerkin discontinue donne un produit matrice vecteur plus rapide que la formulation  $H(\text{rot})$  en 3-D. Si on utilise un ordre bas, les points de Gauss fournissent une solution plus précise que les points de Gauss-Lobatto, mais le coût du produit matrice-vecteur est bien plus élevé. Si on utilise un ordre élevé, le coût est proche, mais le gain en précision n’est pas terrible. Le choix est cornélien, seule l’expérience nous permet de trancher et de préférer les points de Gauss, il est toujours bon d’avoir le moins de degrés de liberté possible. Le stockage est comparable à ce qu’on a pour la première famille, lorsque l’on utilise la factorisation discrète pour cette dernière.

### 6.1.5 Conditions aux limites

Nous rappelons ici comment on prend en compte les conditions aux limites, pour une description plus complète, le lecteur pourra se référer à [Piperno et Fezoui, 2003]. On se place sur un hexaèdre  $K_i$  dont une des faces  $\Gamma$  est un bord du domaine de calcul. On n’a pas d’hexaèdre de l’autre côté de cette face, il faut donc donner une autre définition des flux numériques :

$$\begin{aligned} & -\frac{1}{2} \int_{\Gamma} (\mathbf{H}_i + \mathbf{H}_j) \cdot \boldsymbol{\varphi} \times \mathbf{n} d\Gamma \\ & -\frac{1}{2} \int_{\Gamma} (\mathbf{E}_i - \mathbf{E}_j) \times \mathbf{n} \cdot \boldsymbol{\varphi} d\Gamma \end{aligned}$$

Les valeurs  $\mathbf{E}_j$  et  $\mathbf{H}_j$  ne sont pas définies car il n’existe pas d’hexaèdre  $K_j$ . On exploite alors la condition aux limites sur  $\Gamma$  pour donner une définition à ces deux valeurs.

Condition de Dirichlet  $\mathbf{E} \times \mathbf{n} = \mathbf{f}$  :

$$\mathbf{E}_j \times \mathbf{n} = -\mathbf{E}_i \times \mathbf{n} + 2\mathbf{f} \quad \mathbf{H}_j = \mathbf{H}_i$$

Condition de Neumann  $\mathbf{n} \times (\mathbf{H} \times \mathbf{n}) = \mathbf{f}$  :

$$\mathbf{E}_j = \mathbf{E}_i \quad \mathbf{H}_j = -\mathbf{H}_i + 2\mathbf{f}$$

Condition de Silver-Müller  $\mathbf{n} \times (\mathbf{H} \times \mathbf{n}) = i\sqrt{\frac{\epsilon}{\mu}}\mathbf{E} \times \mathbf{n}$  :

$$\mathbf{E}_j \times \mathbf{n} = -i\sqrt{\frac{\mu}{\epsilon}}\mathbf{n} \times (\mathbf{H}_i \times \mathbf{n}) \quad \mathbf{H}_j = i\sqrt{\frac{\epsilon}{\mu}}\mathbf{E}_i \times \mathbf{n}$$

### 6.1.6 Résolution du système linéaire

On peut choisir de garder les deux inconnues  $E$  et  $H$  ou faire un complément de schur pour ne conserver que  $E$ . Le complément fournit alors le système :

$$-\omega^2 B_h^1 E_h + (R_h + S_h)(B_h^2)^{-1}(R_h^t + S_h^t)E_h = \omega F_h$$

Sur le tableau 6.1, on donne les tailles des matrices LU nécessaire à une résolution directe, suivant qu'on choisit ou non d'éliminer une inconnue. Ces simulations ont été réalisées sur des

Ordre d'approximation	1	2	3
DG Lobatto sans complément	1 687 Mo	1 774 Mo	2 647 Mo
DG Gauss sans complément	5 709 Mo	11 623 Mo	28 100 Mo
DG Lobatto	1 688 Mo	2 010 Mo	2 520 Mo
DG Gauss	5 642 Mo	11 398 Mo	24 622 Mo
Première famille	496 Mo	544 Mo	768 Mo

TAB. 6.1 – Taille mémoire requise par une résolution directe, cas 3-D. On compare à nombre de ddl constant (80 000).

maillages déformés. On compare avec la première famille sur les hexaèdres (cf. chapitre 5). On remarquera que le complément de Schur ne nous fait pas gagner en stockage. Les points de Gauss donnent des matrices LU nécessitant un stockage environ 10 fois plus important qu'en utilisant les points de Gauss-Lobatto. Les points de Gauss-Lobatto fournissent également un stockage très important, ce qui exclut l'utilisation d'un solveur direct sur les méthodes Galerkin discontinues 3-D.

En ce qui concerne le solveur itératif, il est intéressant de considérer plutôt l'équation fournie par le complément de Schur. Cela permet d'avoir des vecteurs d'itération deux fois plus petits, et donc de gagner un facteur 2 en stockage. On utilisera comme pour l'équation de Helmholtz, du BICGCR au besoin préconditionné par une factorisation incomplète sur l'équation amortie, ou par une décomposition en sous-domaines.

Pour trouver les modes propres et valeurs propres, on passera par le complément de Schur, car il nous ramène à étudier les valeurs propres d'une matrice symétrique positive. Si on avait gardé les deux inconnues  $E$  et  $H$ , on aurait eu une matrice symétrique mais indéfinie (de fait lorsque  $\omega$  est valeur propre,  $-\omega$  est également valeur propre).

## 6.2 Présence de modes parasites ?

### 6.2.1 Etude de convergence

#### Cas 2-D

**Cas du disque** On s'intéresse dans un premier temps au cas du disque parfaitement conducteur. On reprend le cas test du cinquième chapitre (cf. figure 5.2), on utilise une condition de Silver-Müller sur la frontière extérieure. On trace les courbes d'erreur en fonction du pas de maillage sur la figure 6.3. L'erreur est calculée sur  $H$ , qui est le rotationnel de  $\mathbf{E}$  à une constante

près. Pour la seconde famille, on calculait également l'erreur sur  $H$ , mais comme  $H$  n'était pas une inconnue principale de la formulation variationnelle, on devait évaluer le rotationnel de  $\mathbf{E}$ . L'avantage de la formulation Galerkin discontinue est que  $H$  est une inconnue du problème et les deux inconnues  $\mathbf{E}$  et  $H$  jouent un rôle similaire. Les erreurs commises sur  $E$  ou  $H$  sont équivalentes en 3-D, en 2-D c'est moins clair.

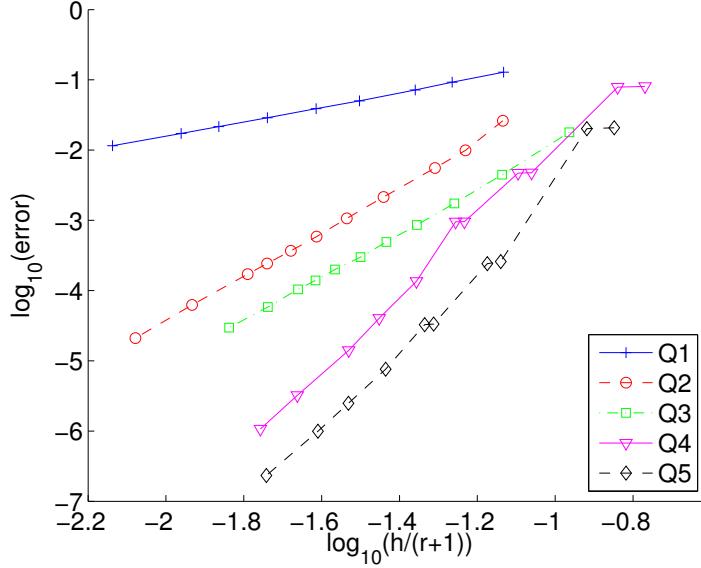


FIG. 6.3 – Evolution de l'erreur sur  $H$  entre la solution numérique et la solution analytique en fonction de  $h/(r + 1)$ , où  $h$  est le pas de maillage,  $r$  l'ordre d'approximation. Echelle log-log. Cas du disque sur des maillages réguliers.

On choisit de prendre en abscisse la variables  $h/(r + 1)$ , cela permet de comparer de façon équitable les ordres d'approximation. On voit que sur ce petit cas, à nombre de degrés de liberté fixé, on gagne en précision lorsqu'on monte en ordre. Au niveau des ordres de convergence, on mesure une convergence en  $O(h^{1.02})$  pour  $Q_1$ , une convergence en  $O(h^{4.3})$  pour  $Q_2$ , une convergence en  $O(h^{3.10})$  pour  $Q_3$ , une convergence en  $O(h^{5.88})$  pour  $Q_4$  et une convergence en  $O(h^{5.06})$  pour  $Q_5$ . Au vu de ces mesures, on peut conjecturer que la méthode de Galerkin discontinue converge en  $O(h^{r+2})$  si l'ordre est pair et en  $O(h^r)$  si l'ordre est impair. Cette distinction ordre pair/ordre impair a été signalée via une étude numérique 1-D [Piperno, 2003]. On effectue la même démarche sur des maillages triangulaires découpés, qui nous posaient des problèmes dans le chapitre précédent : On n'observe pas de convergence erratique comme dans le cas de la seconde famille de Nédélec sur les quadrangles. Contrairement au cas régulier, on obtient une convergence en  $O(h^{r+1})$  (on mesure une pente de 1.95 pour  $Q_1$  et de 3.02 pour  $Q_2$ ). Il n'y a plus de distinction ordre pair/ordre impair. On compare la précision obtenue avec les points de Gauss, par rapport aux points de Gauss-Lobatto sur la figure 6.5. On voit que l'utilisation des points de Gauss-Lobatto donne le même ordre de convergence sur des maillages triangulaires découpés, mais la constante est bien plus élevée pour les points de Gauss-Lobatto.

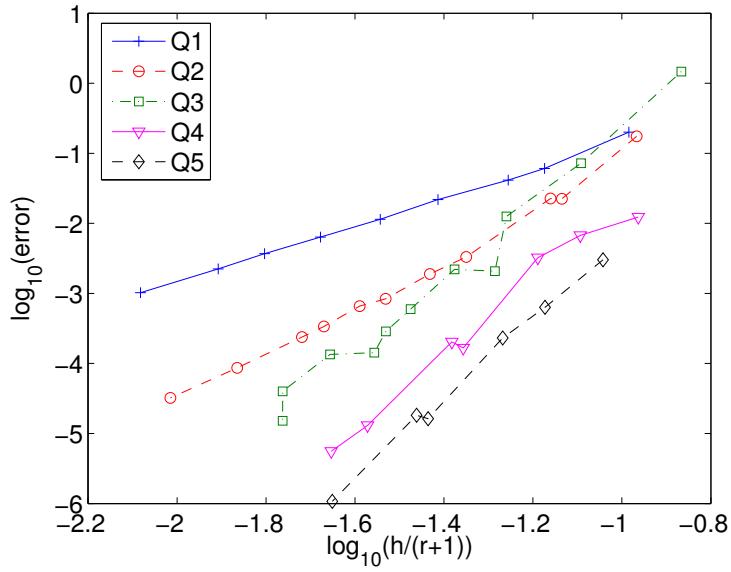


FIG. 6.4 – Evolution de l’erreur sur  $H$  entre la solution numérique et la solution analytique en fonction de  $h/(r + 1)$ , où  $h$  est le pas de maillage,  $r$  l’ordre d’approximation. Echelle log-log. Cas du disque sur des maillages triangulaires découpés.

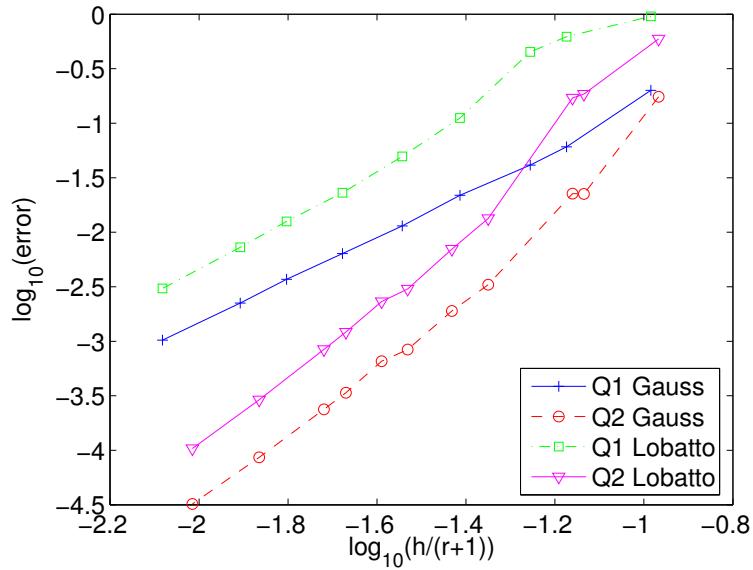


FIG. 6.5 – Evolution de l’erreur sur  $H$  entre la solution numérique et la solution analytique en fonction de  $h/(r + 1)$ , où  $h$  est le pas de maillage,  $r$  l’ordre d’approximation. Echelle log-log. Cas du disque sur des maillages triangulaires découpés.

**Cas du carré** On considère un carré parfaitement conducteur 6.6. La solution de référence

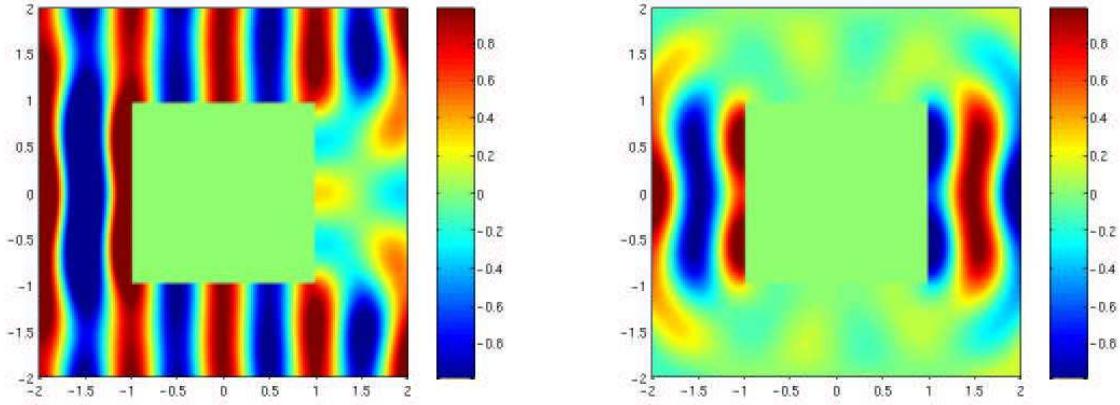


FIG. 6.6 – A gauche, partie réelle du champ total pour carré de côté 2. à droite partie réelle du champ diffracté.

est calculée sur un maillage fin avec de l'ordre élevé, on garantit une erreur inférieure à  $10^{-4}$  sur la solution de référence. On trace les courbes de convergence pour la méthode Galerkin discontinue sur la figure 6.7. On voit sur cette figure, qu'on obtient une convergence en  $O(h^{4/3})$ ,

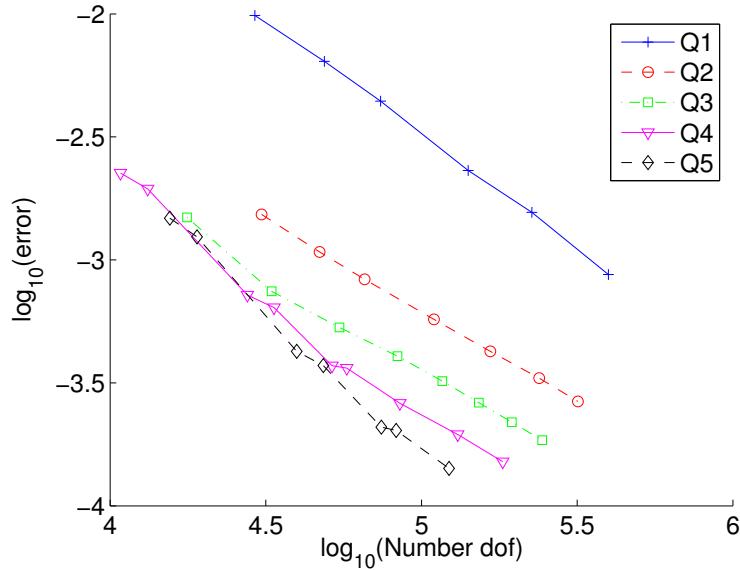


FIG. 6.7 – Evolution de l'erreur sur  $H$  entre la solution numérique et la solution de référence en fonction de  $h/(r + 1)$ , où  $h$  est le pas de maillage,  $r$  l'ordre d'approximation. Echelle log-log. Cas du carré sur des maillages triangulaires découpés.

mais que la constante est bien plus petite quand on monte ordre. Il reste donc intéressant de monter en ordre en présence de coins. Sur la figure 6.8, on compare les erreurs commises par la formulation Galerkin discontinue, la seconde famille et la première famille en fonction du nombre de ddl, ce pour divers ordres d'approximation. On voit que la formulation Galerkin discontinue est compétitive par rapport à la première famille. La seconde famille donne toujours une convergence en dents de scie.

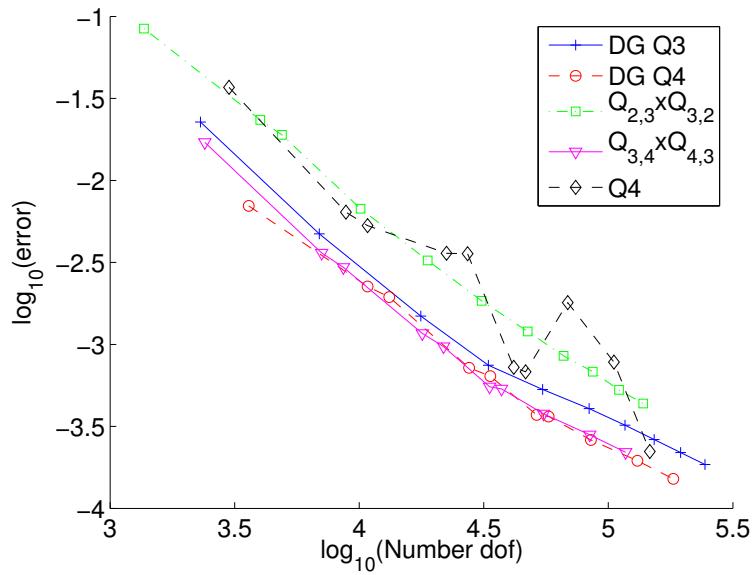


FIG. 6.8 – Evolution de l'erreur sur  $H$  entre la solution numérique et la solution de référence en fonction du nombre de ddl. Echelle log-log. Cas du carré sur des maillages triangles découpés, plusieurs éléments finis utilisés.

Dans l'ensemble des expériences numériques menées, on n'a jamais observé d'ondes parasites, même sur des maillages relativement grossiers (6/7 points par longueur d'onde). La méthode semble très robuste comparativement à la seconde famille.

### Cas 3-D

**Maillages réguliers** On s'intéresse au cas d'une sphère parfaitement conductrice (cf. figure 5.6), on utilise une condition de Silver-Müller sur la frontière extérieure. On introduit l'erreur H-rot (qui est en fait une erreur  $L^2$ , car on dispose de  $H$ ) :

$$\text{error} = \frac{\left( \|\mathbf{E}^{\text{numérique}} - \mathbf{E}^{\text{analytique}}\|_{L^2(\Omega)}^2 + \|\mathbf{H}^{\text{numérique}} - \mathbf{H}^{\text{analytique}}\|_{L^2(\Omega)}^2 \right)^{1/2}}{\left( \|\mathbf{E}^{\text{analytique}}\|_{L^2(\Omega)}^2 + \|\mathbf{H}^{\text{analytique}}\|_{L^2(\Omega)}^2 \right)^{1/2}}$$

On trace les courbes d'erreur en fonction du nombre de degrés de liberté sur la figure 6.9. Le nombre de degrés de liberté est le nombre d'inconnues utilisées pour  $E$  uniquement. Le pas de maillage  $h$  est considéré proportionnel à l'inverse de la racine cubique du nombre de ddl. Au niveau des ordres de convergence, on trouve un ordre de 1.08, 3.42, 3.30, 2.65 et 5.28 pour

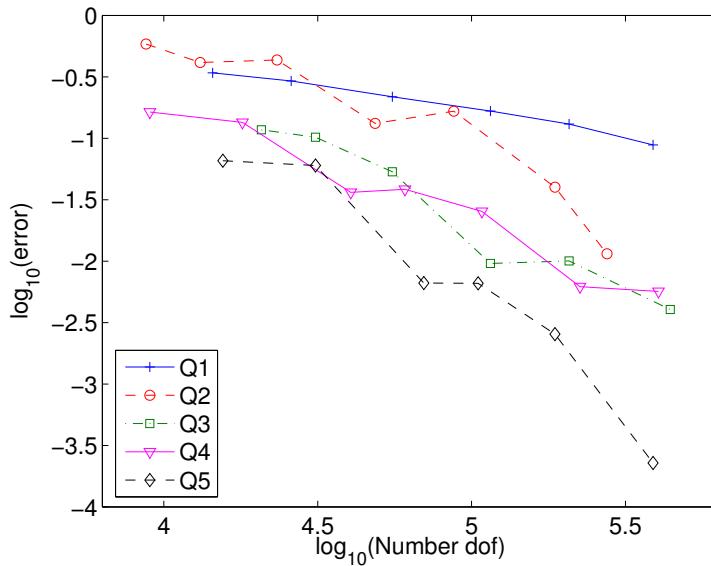


FIG. 6.9 – Evolution de l'erreur H-rot entre la solution numérique et la solution analytique en fonction de  $h/(r+1)$ , où  $h$  est le pas de maillage,  $r$  l'ordre d'approximation. Echelle log-log.

respectivement  $Q_1$ ,  $Q_2$ ,  $Q_3$ ,  $Q_4$  et  $Q_5$ . Pour  $Q_1$ , il semble qu'on converge effectivement en  $O(h)$ , on retrouve le cas 2-D sur des maillages réguliers. Pour les autres ordres, on ne peut pas conclure, la convergence étant irrégulière.

**Maillages non-structurés** On met en valeur dans ce paragraphe l'importance de pénaliser la formulation Galerkin discontinue sur des maillages non-structurés. Le premier cas test est la diffraction par une sphère diélectrique d'indices  $\varepsilon = 4$   $\mu = 1$ . Les solutions numériques de ce problème sont affichées sur la figure 6.10. Un cas plus difficile est le cas d'un cube, à l'intérieur duquel on a placé une source dipolaire :

$$-\omega^2 \mathbf{E} + \mathbf{rot} \mathbf{rot} \mathbf{E} = \mathbf{f} \quad \text{sur } \Omega = [-1, 1]^3$$

$$\mathbf{E} \times \mathbf{n} = 0 \quad \text{sur } \partial\Omega$$

$$\mathbf{f} = \frac{1}{r_0^2} \exp\left(-\frac{7\pi r^2}{r_0^2}\right) \mathbf{e}_x$$

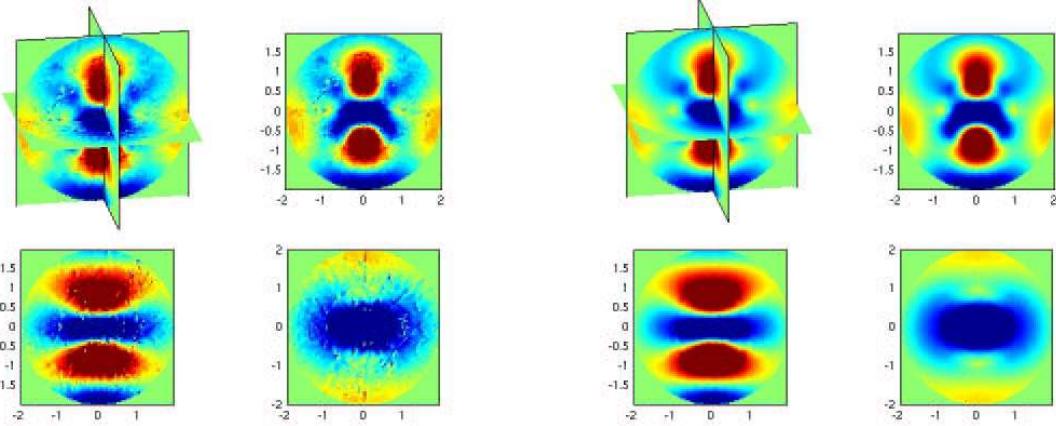


FIG. 6.10 – Champ diffracté  $E_x$  pour la formulation Galerkin discontinue (points de Gauss). À gauche sans pénalisation et à droite avec pénalisation  $\alpha = 0.5$ . On utilise le même maillage et une approximation  $Q_3$ .

$r_0$  est le rayon de la gaussienne, égal à 0.6. Sur un maillage tétraédrique découpé (cf. figure 3.5), on obtient les solutions de la figure 6.11. On voit que le terme de pénalisation permet d'obtenir une solution relativement propre, alors que la formulation Galerkin discontinue sans pénalisation donne une solution fortement perturbée. Dans la suite, on exhibera les parasites à l'aide d'une

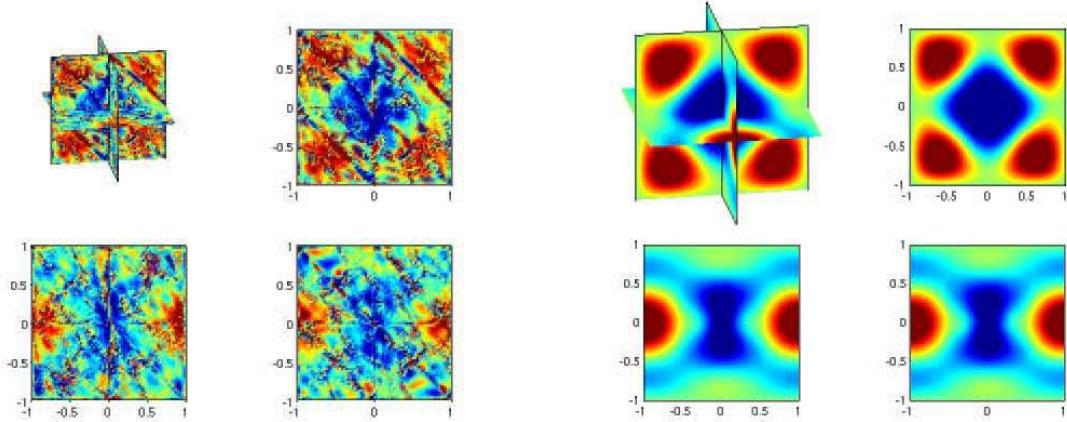


FIG. 6.11 – Champ  $E_x$  pour la formulation Galerkin discontinue (points de Gauss). À gauche sans pénalisation et à droite avec pénalisation  $\alpha = 0.5$ . On utilise le même maillage et une approximation  $Q_4$ .

étude de valeurs propres.

### 6.2.2 Etude de valeurs propres

Une étude numérique pour des triangles/tétraèdres est réalisée dans [Hesthaven et Warburton, 2004]. Pour l'analyse théorique, on renvoie le lecteur à [Buffa et Perugia, 2005]. Nous nous limitons dans cette partie uniquement au cas de quadrilatères/hexaèdres sur des maillages réguliers et non-structurés. Au premier abord, il nous a semblé que la formulation Galerkin discontinue était exempte d'ondes parasites. On va constater dans cette sous-section qu'il n'en

est rien. On distingue la formulation Galerkin discontinue avec points de Gauss et points de Gauss-Lobatto.

## Cas 2-D, points de Gauss-Lobatto

**Modes propres en maillage régulier** Comme dans le cas de la seconde famille, on n'a pas de valeurs propres parasites, mais une multiplicité incorrecte. Les modes propres physiques sont indissociables des modes propres parasites comme on le voit sur la figure 6.12. Lorsqu'on raffine

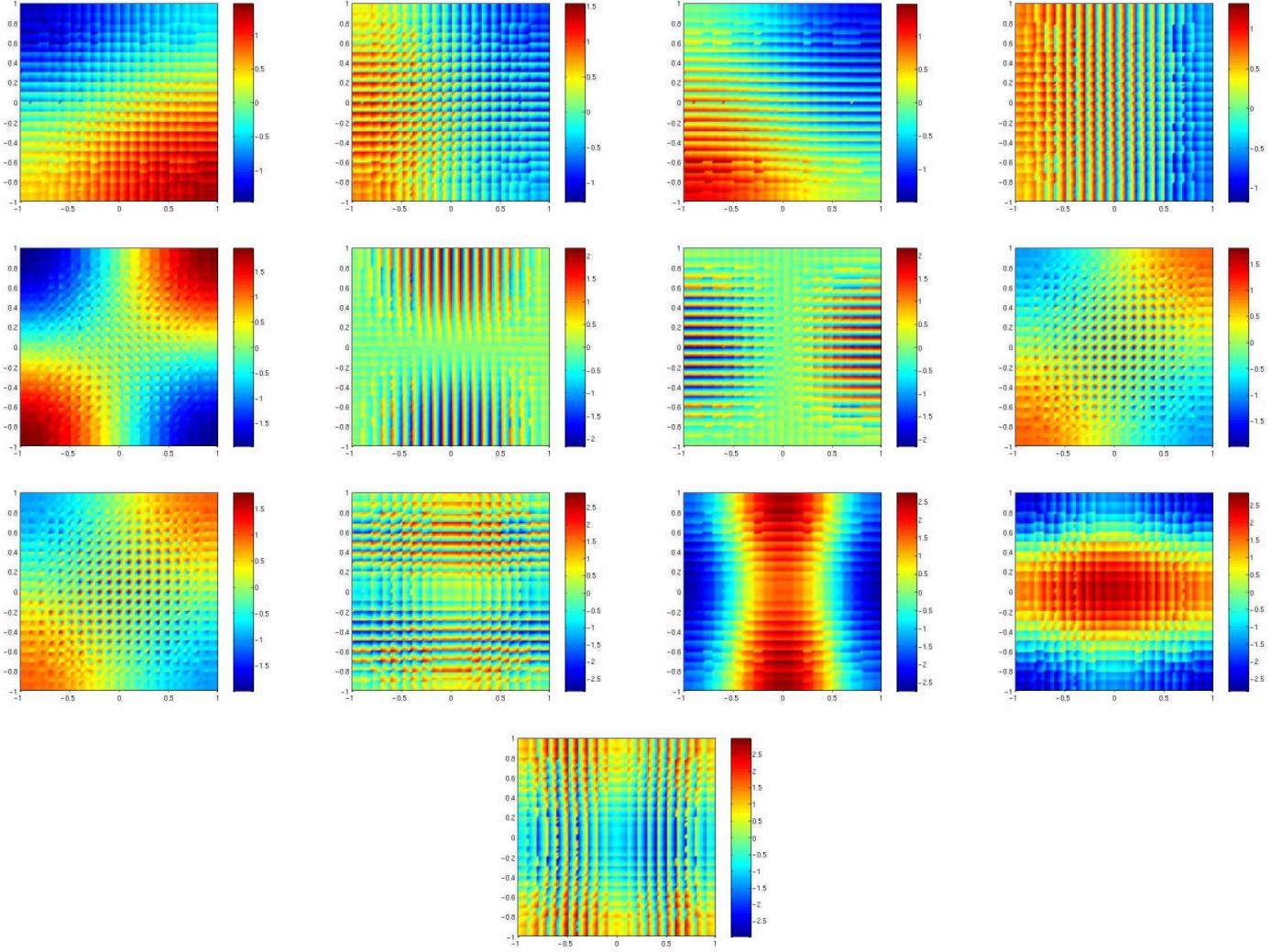


FIG. 6.12 – 13 premiers modes propres pour un maillage  $20 \times 20$  avec une approximation  $Q_1$ . Utilisation des points de Gauss-Lobatto.

le maillage, la multiplicité des valeurs propres n'est pas modifiée, on n'a pas de prolifération incontrôlée des modes parasites comme dans le cas de la seconde famille. Les modes parasites sont toujours à la même place, mais leur forme varie quand on raffine le maillage.

**Modes propres en maillage non-régulier** L'utilisation de maillages non-réguliers est très bénéfique, on recense beaucoup moins de modes propres parasites que dans le cas régulier, nous en exhibons quelques uns sur la figure 6.13, sur le maillage 4.14 avec une approximation

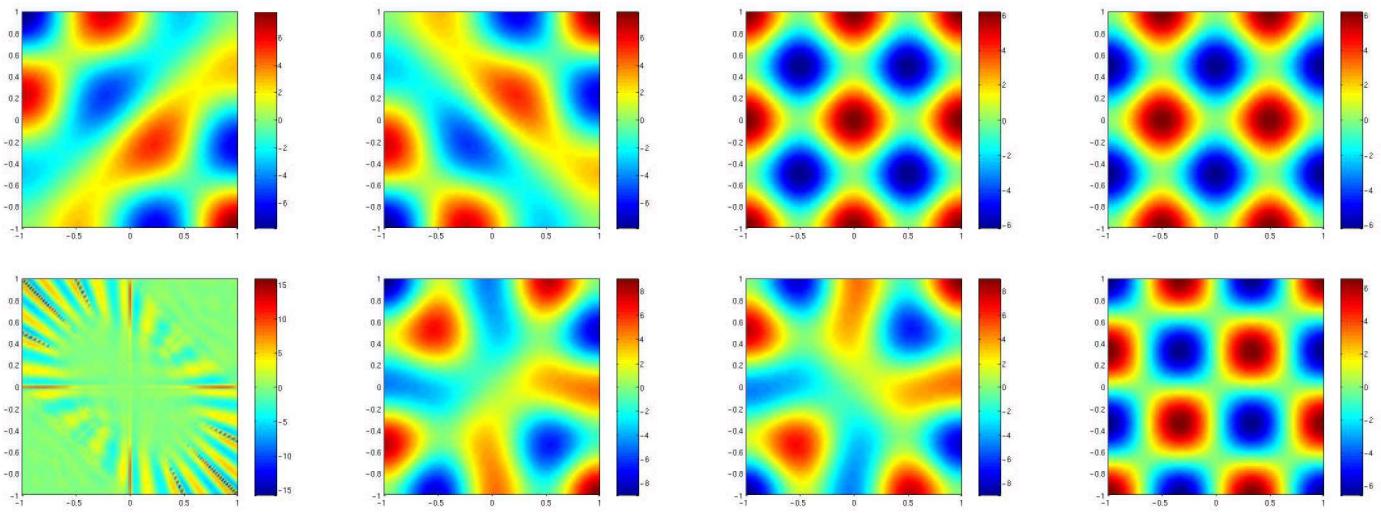


FIG. 6.13 – Quelques modes propres pour un maillage triangulaire découpé (approximation  $Q_5$ ). Utilisation des points de Gauss-Lobatto

$Q_5$ . Sur ce maillage, le mode parasite affiché est le premier qu'on rencontre dans le spectre ! Lorsqu'on utilise un maillage triangulaire plus fin qu'on découpe, les valeurs propres parasites sont repoussées plus loin dans le spectre. Il est possible que sur des maillages non-réguliers, la méthode Galerkin discontinue soit spectralement correcte, alors qu'elle ne l'est pas sur des maillages réguliers.

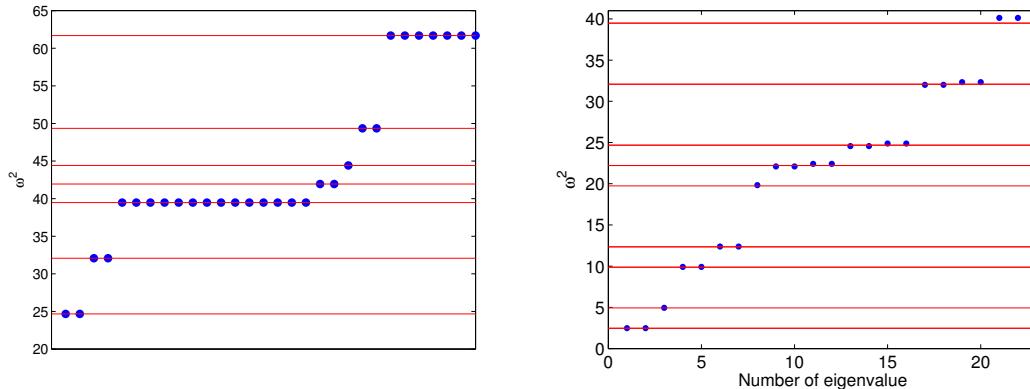


FIG. 6.14 – A gauche distribution des valeurs propres pour la seconde famille, à droite distribution pour Galerkin discontinue avec points de Gauss. Les traits horizontaux rouges symbolisent les valeurs propres analytiques. Les points bleus sont les valeurs propres numériques. On utilise une approximation  $Q_1$  sur un maillage 20x20 pour les deux.

### Cas 2-D, points de Gauss

**Maillage régulier** On observe une différence notable avec les points de Gauss-Lobatto. On obtient des valeurs propres parasites, les valeurs propres physiques ont une multiplicité correcte. La conséquence immédiate est que les modes propres physiques sont dissociés des modes propres parasites (voir figure 6.15). On synthétise la différence de distribution des valeurs propres entre la seconde famille et Galerkin discontinue sur la figure 6.14. Cette figure confirme, d'une part,

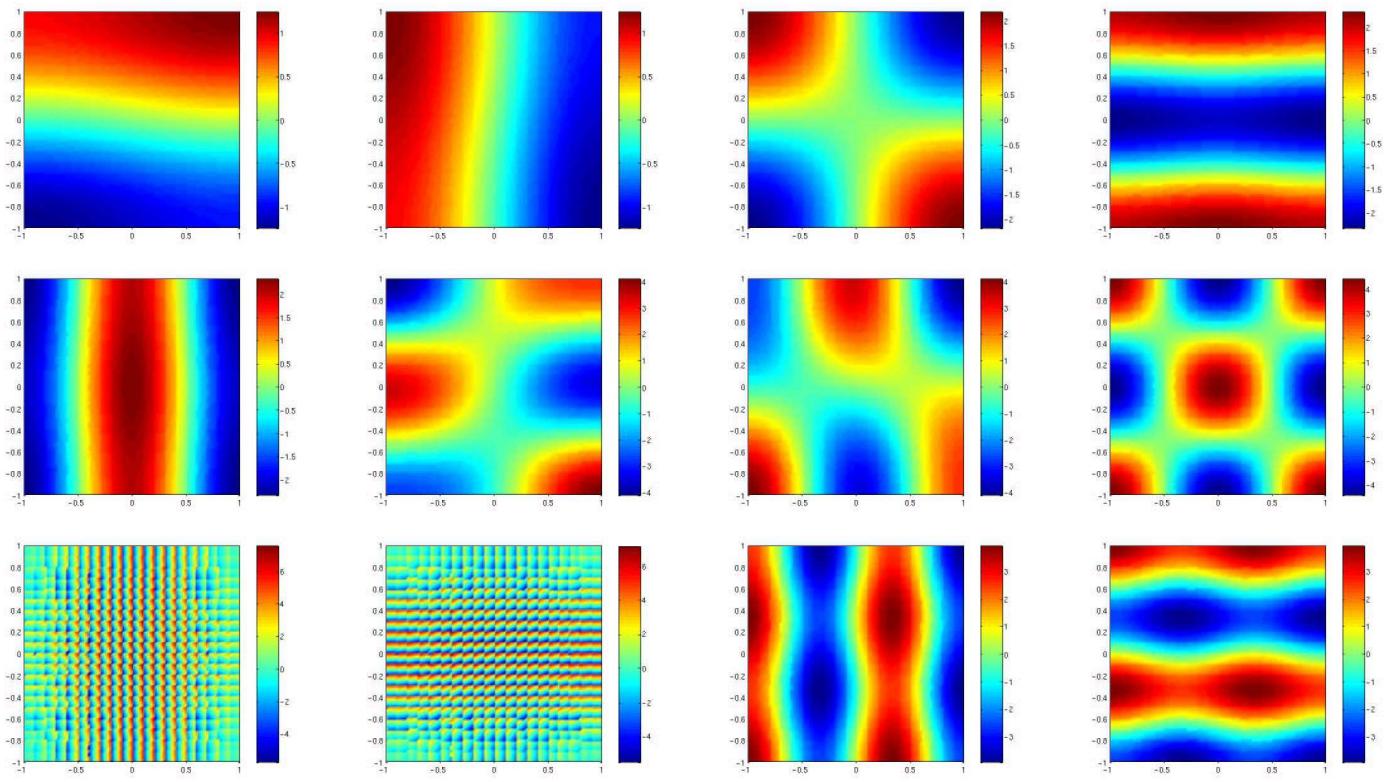


FIG. 6.15 – 13 premiers modes propres pour un maillage  $20 \times 20$  avec une approximation  $Q_1$ . Utilisation des points de Gauss.

la prolifération des modes propres parasites pour la seconde famille et, d'autre part, la non-prolifération pour Galerkin discontinu.

**Maillage non-régulier** Il n'y a pas de différence notable avec les points de Gauss-Lobatto, on affiche sur la figure 6.16 les modes propres trouvés avec le premier mode parasite trouvé sur ce maillage. D'autres modes parasites existent, mais bien plus loin dans le spectre. Le maillage non-régulier est aussi très bénéfique lorsqu'on utilise les points de Gauss.

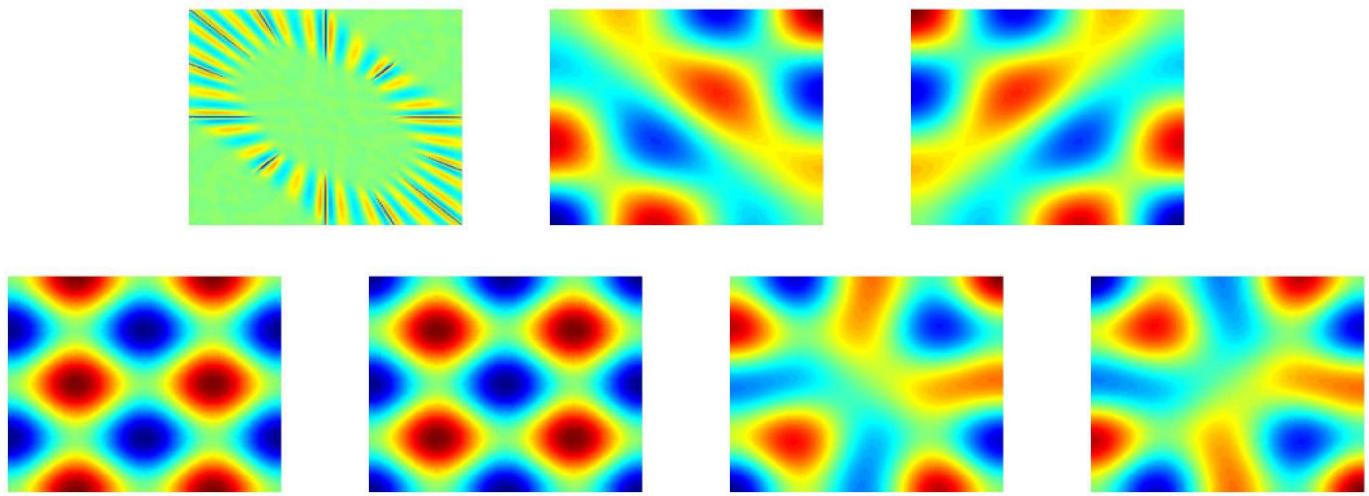


FIG. 6.16 – Quelques modes propres pour un maillage triangulaire découpé (approximation  $Q_5$ ). Utilisation des points de Gauss

### Cas 3-D

**Maillage régulier** On n'a pas noté de différence fondamentale entre les points de Gauss et les points de Gauss-Lobatto. En effet lorsqu'on utilise un maillage régulier, on a des valeurs propres parasites et des valeurs propres de multiplicité incorrecte. Les valeurs propres trouvées correspondent à des valeurs propres du type :

$$\omega_{k,m,n}^2 = \frac{\pi^2 (k^2 + m^2 + n^2)}{L^2} \quad k > 0 \ m \geq 0 \ n \geq 0$$

Ainsi certaines valeurs propres sont parasites comme on peut le voir sur la figure 6.17. Du fait de

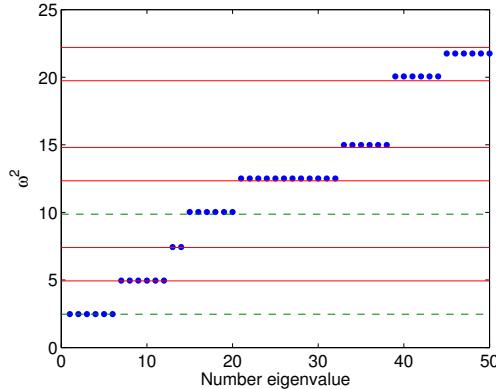


FIG. 6.17 – Distribution des valeurs propres pour Galerkin discontinu avec points de Gauss. Les traits horizontaux rouges symbolisent les valeurs propres analytiques. Les traits horizontaux verts symbolisent les valeurs propres en choisissant  $(k, m, n) = (1, 0, 0)$  ou  $(2, 0, 0)$ . Les points bleus sont les valeurs propres numériques. On utilise une approximation  $Q_1$  sur un maillage 10x10x10 (24 000 ddl).

la multiplicité incorrecte, il est impossible de dissocier les modes propres physiques des modes propres parasites, on exhibe quelques modes parasites sur la figure 6.18. Le mode  $(0,1,1)$  est de multiplicité physique 3, alors qu'on trouve numériquement six modes associés. Le mode  $(1,1,1)$  de multiplicité physique 2, n'est pas parasité. Néanmoins, nous avons comme en 2-D une non-prolifération des modes parasites. Lorsqu'on raffine le maillage, le nombre de modes associés à une valeur propre est constant.

**Maillage non-structuré** Le maillage considéré dans cette section est celui de la figure 3.5 découpé en hexaèdres. On utilise une approximation  $Q_4$ . Un tel maillage est adapté à la fréquence 1 (10 points par longueur d'onde). On s'attend à observer le même apport “bénéfique” qu'on avait en 2-D. Bien au contraire, l'effet est très maléfique ! Sur la figure 6.19, on a représenté la distribution des valeurs propres obtenues. On a beaucoup de valeurs propres parasites, qui polluent le spectre entier, avec un espacement régulier. Nous avons observé une prolifération des modes parasites lorsqu'on raffinait le maillage. Plus on raffine le maillage, plus on a de valeurs propres parasites. Lorsqu'on monte en ordre en gardant le même maillage, le nombre de valeurs propres parasites diminue. On a représenté sur la figure 6.20 les modes parasites et les modes physiques, qu'on peut dissocier en maillage non-régulier. Néanmoins, il faut partir à la pêche pour les trouver, générer 100 modes, pour espérer en trouver quelques uns de physiques ... Les premiers modes physiques ( mode  $(0,1,1)$  en haut à droite, mode  $(1,1,1)$  en-dessous) sont très bien approchés.

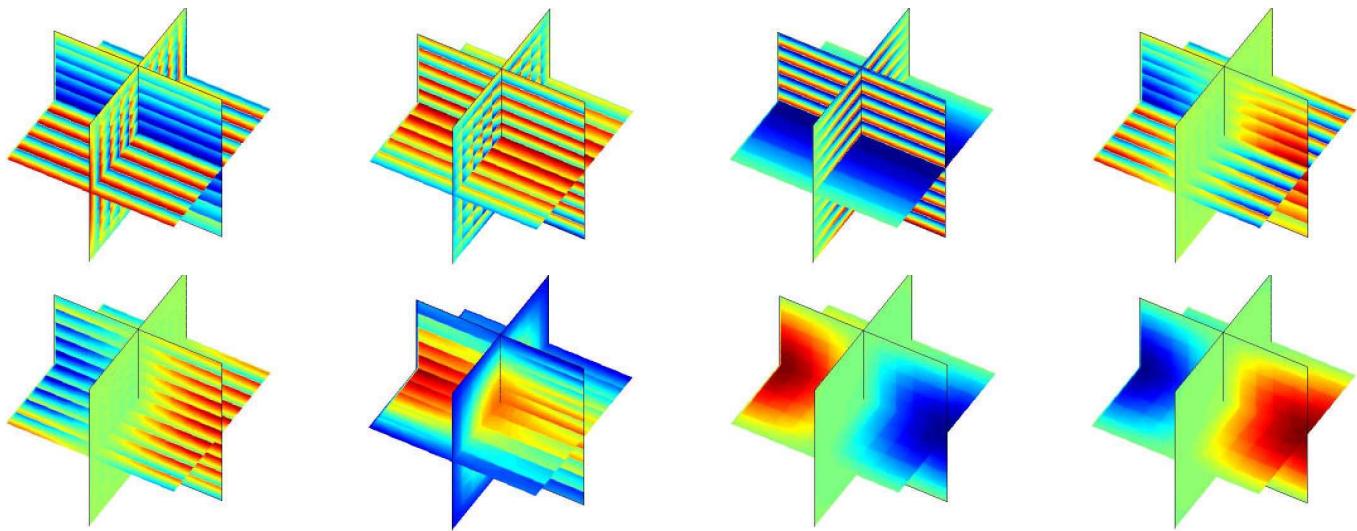


FIG. 6.18 – Quelques modes propres sur un maillage régulier  $10 \times 10 \times 10$  (24 000 ddl). Utilisation des points de Gauss

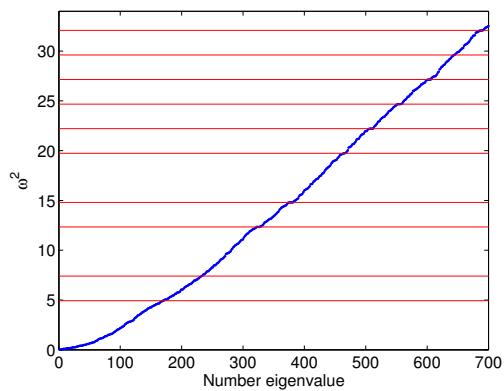


FIG. 6.19 – Distribution des valeurs propres pour Galerkin discontinu avec points de Gauss-Lobatto. Les traits horizontaux rouges symbolisent les valeurs propres analytiques. Les points bleus sont les valeurs propres numériques. On utilise une approximation  $Q_4$  sur un maillage non-structuré.

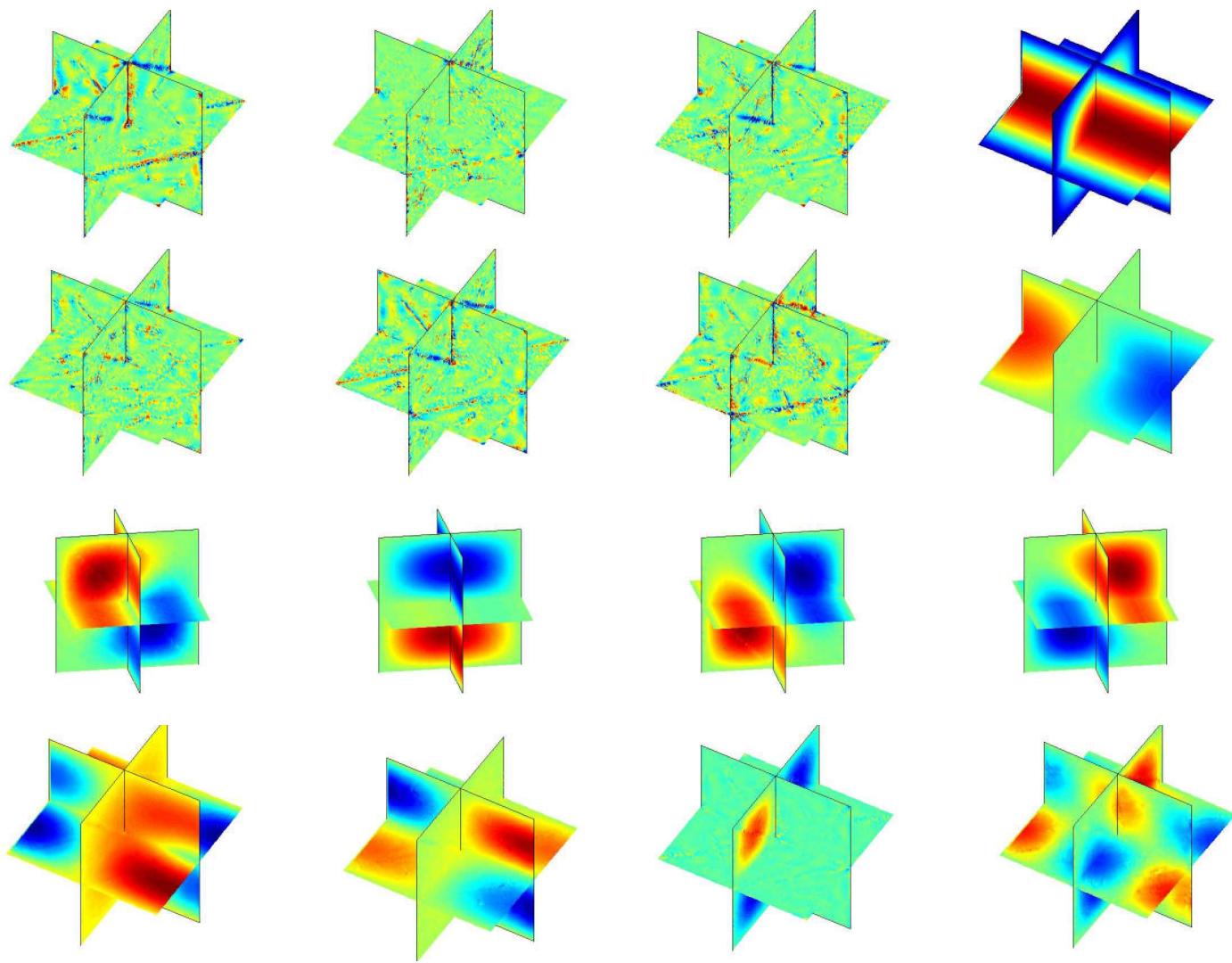


FIG. 6.20 – Quelques modes propres sur un maillage non-structuré (72 000 ddl). Utilisation des points de Gauss-Lobatto. On représente principalement les modes propres physiques.

**Influence de la pénalisation** Au niveau discret le problème aux valeurs propres écrit :

$$-\omega (B_h^1 + i\alpha P_h) E + (R_h + S_h) H = 0$$

$$(R_h + S_h)^t E - \omega (B_h^2 + i\delta P_h) H = 0$$

On ne peut pas éliminer  $H$  si  $\delta \neq 0$ , mais on le prend égal à zéro ! On élimine donc  $H$  et on cherche les valeurs propres d'une matrice complexe symétrique. Les valeurs propres sont donc complexes. On considère le même maillage tétraèdrique découpé de la figure précédente, on utilise une approximation  $Q_3$ . L'ajout du terme de pénalisation rejette les modes parasites dans le plan complexe avec une partie imaginaire assez élevée, comme on peut le voir sur la figure 6.21. Seuls les modes physiques gardent une partie imaginaire proche de zéro. Sur la figure 6.22, on

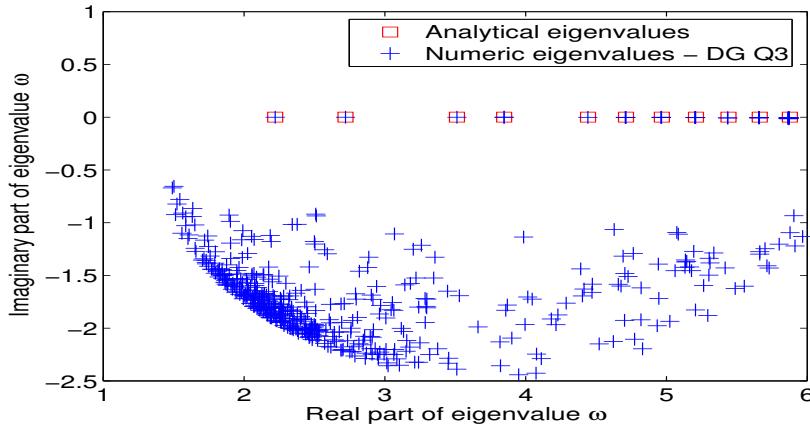


FIG. 6.21 – Distribution des valeurs propres pour Galerkin discontinu avec points de Gauss-Lobatto et terme de pénalisation  $\alpha = 0.5$ . On utilise une approximation  $Q_3$  sur un maillage non-structuré.

a disposé les modes propres physiques pour des fréquences croissantes. Sur cette figure, on a choisi de visualiser le module du champ électrique plutôt que la partie réelle de la composante suivant  $x$ . Dans nos expériences numériques (en maillage régulier ou non-structuré), nous n'avons jamais trouvé de mode parasite qui n'était pas rejeté dans le plan complexe. De plus, la partie imaginaire des modes physiques est très petite, d'autant plus que le maillage est fin (forcément les valeurs propres convergent vers les valeurs propres analytiques réelles). Probablement, que la formulation Galerkin discontinu est spectralement correcte pour les hexaèdres de la seconde famille.

### 6.3 Conclusion

Dans ce chapitre, nous avons décrit une méthode de Galerkin discontinu sur les hexaèdres, utilisant les mêmes espaces locaux que la seconde famille de Nédélec. Les matrices de rigidité obtenues sont creuses et indépendantes de la géométrie. Les matrices de masse sont diagonales par bloc 3x3. On obtient ainsi un produit matrice-vecteur rapide et peu coûteux en stockage.

ELorsqu'on utilise une formulation Galerkin discontinu, on ne rencontre pas de difficultés dans le cas 2-D, où le nombre de modes parasites est restreint et leur impact sur la qualité de la solution est négligeable. Lorsqu'on passe au cas 3-D, le nombre de parasites est bien plus important, ce qui oblige à rajouter à la formulation variationnelle un terme de pénalisation.

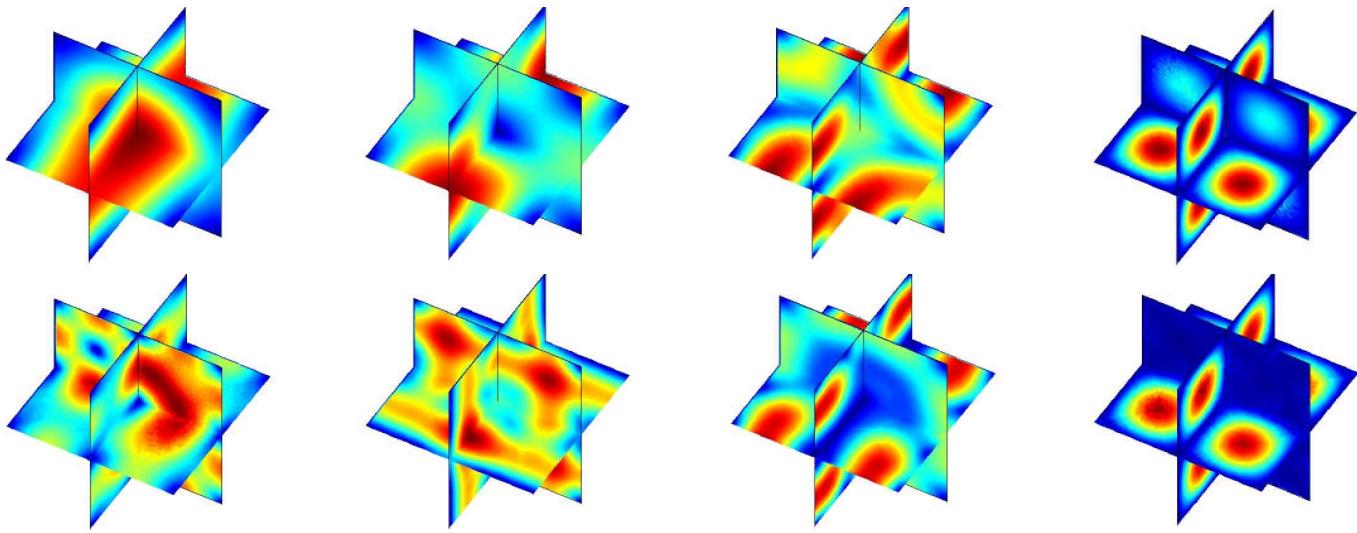


FIG. 6.22 – Quelques modes propres sur un maillage non-structuré (39 000 ddl). Utilisation des points de Gauss-Lobatto avec terme de pénalisation. On n'observe que des modes physiques. Module du champ électrique.

Grâce à ce terme de pénalisation, on obtient une méthode robuste, gardant un stockage très faible et un produit matrice-vecteur rapide.

L'inconvénient principal reste le nombre de degrés de liberté très important, par rapport à la première famille.

# Chapitre 7

## Comparaison hexaèdres / tétraèdres pour les équations de Maxwell 3-D

*Nous nous proposons dans ce chapitre d'établir des comparaisons entre les éléments finis hexaédriques de la première famille et les éléments finis tétraédriques de la première famille. Nous faisons en premier lieu une étude de dispersion sur des maillages périodiques. Nous ferons une comparaison sur le cas académique de la sphère. Finalement, nous traiterons des cas complexes 3-D, utilisant les préconditionneurs présentés dans le chapitre .*

### Sommaire

---

<b>7.1</b>	<b>Analyse de dispersion</b>	<b>172</b>
7.1.1	Cas 2-D	172
7.1.2	Cas 3-D	175
<b>7.2</b>	<b>Cas académique de la sphère</b>	<b>176</b>
7.2.1	Coût du produit matrice-vecteur	176
7.2.2	Sphère parfaitement conductrice	177
7.2.3	Sphère diélectrique	178
<b>7.3</b>	<b>Cavité cobra</b>	<b>179</b>
<b>7.4</b>	<b>Conclusion</b>	<b>182</b>

---

## 7.1 Analyse de dispersion

Le procédé d'obtention des erreurs de dispersion est identique à celui utilisé dans le chapitre trois pour l'équation de Helmholtz. On ne revient pas dessus. En outre, on adopte une définition similaire pour évaluer la constante  $L^2$  de l'erreur de dispersion. La seule différence est dans la définition de  $\tilde{r}$ . Comme on a des degrés de liberté vectoriels, il est nécessaire de choisir la définition suivante :

$$\tilde{r} = \sqrt{\frac{\text{Nombre de degrés de liberté indépendants / 2}}{\text{Aire du motif élémentaire}}} \quad \text{en 2-D}$$

$$\tilde{r} = \sqrt[3]{\frac{\text{Nombre de degrés de liberté indépendants / 3}}{\text{Volume du motif élémentaire}}} \quad \text{en 3-D}$$

On rappelle qu'on utilise la variable  $K$  :

$$K = \frac{6k h}{2\pi \tilde{r}}$$

### 7.1.1 Cas 2-D

#### Eléments finis quadrilatéraux

Pour les quadrilatères, on obtient les mêmes erreurs de dispersion que pour le cas scalaire. Une démonstration d'une telle propriété est faite par [Cohen et Monk, 1998], [Ainsworth, 2004b], la démonstration est également réalisée dans le cas de la seconde famille. Il est intéressant de noter que la première et la seconde famille donnent la même erreur de dispersion, alors que la première famille nécessite moins de degrés de liberté. On rappelle l'erreur moyenne de dispersion, sur les éléments finis quadrilatéraux (cf. chapitre 3), sur le tableau 7.1.

Ordre d'approximation	$Q_1$	$Q_2$	$Q_3$	$Q_4$	$Q_5$
Intégration exacte	$6.76e - 2 K^2$	$1.80e - 2 K^4$	$5.98e - 3 K^6$	$2.22e - 3 K^8$	$8.74e - 4 K^{10}$
Intégration approchée	$6.76e - 2 K^2$	$8.98e - 3 K^4$	$1.99e - 3 K^6$	$5.54e - 4 K^8$	$1.74e - 4 K^{10}$

TAB. 7.1 – Constante  $L^2$  de l'erreur de dispersion pour les éléments finis quadrilatéraux

On effectue les mêmes calculs dans le cas de maillages triangulaires découpés, on s'attend dans ce cas à avoir une différence par rapport au cas scalaire, car les maillages ne sont pas créés par tensorisation d'un maillage 1-D. Les erreurs de dispersion trouvées dans ce cas, sont visibles sur le tableau 7.2. Sur des maillages fortement modifiés, l'ordre de l'erreur de dispersion

Ordre d'approximation	$Q_1$	$Q_2$	$Q_3$	$Q_4$	$Q_5$
Intégration exacte	$4.54e - 2 K^0$	$6.29e - 3 K^2$	$1.88e - 3 K^4$	$7.5e - 4 K^6$	$3.58e - 4 K^8$
Intégration approchée	$9.89e - 2 K^2$	$9.16e - 3 K^4$	$2.39e - 3 K^6$	$7.46e - 4 K^8$	$2.99e - 4 K^{10}$

TAB. 7.2 – Constante  $L^2$  de l'erreur de dispersion pour les éléments finis quadrilatéraux sur des maillages triangulaires découpés

est de  $2r$  lorsqu'on utilise une intégration numérique approchée. En revanche, on obtient bien un ordre de  $2r - 2$ , si on utilise une intégration exacte. On observe ainsi une non-consistance de  $Q_1$  sur des maillages triangulaires équilatéraux découpés, ce qu'on avait subodoré dans le chapitre 5 à l'aide d'une étude numérique de convergence. Le terme prépondérant de la relation de dispersion s'écrit pour le maillage étudié :

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 + \frac{\xi_1^2}{22} + \frac{\xi_2^2}{22}$$

On a bien un biais de  $\frac{1}{22}$  de la valeur propre physique. L'intégration numérique approchée, bizarrement, fait gagner de la précision. La relation de dispersion discrète pour  $Q_1$  avec une intégration numérique approchée s'écrit :

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 - \frac{5\xi_1^4}{96} - \frac{5\xi_1^2\xi_2^2}{48} - \frac{5\xi_2^4}{96}$$

Pour ce développement, on a considéré que  $h$  était égal à la demi-longueur du triangle d'origine (qu'on découpe). La constante de dispersion (environ 0.104) est plus grande que la constante sur maillage régulier (environ 0.0833).

### Eléments finis triangulaires

Lorsqu'on utilise les éléments finis triangulaires de la première famille sur des triangles équilatéraux, on obtient les relations de dispersion discrètes suivantes :

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 + \frac{\xi_1^6}{3840} - \frac{\xi_1^4\xi_2^2}{256} + \frac{\xi_1^2\xi_2^4}{768} - \frac{7\xi_2^6}{11520} \quad \text{pour } r = 1$$

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 + \frac{13\xi_1^6}{86400} + \frac{71\xi_1^4\xi_2^2}{57600} - \frac{\xi_1^2\xi_2^4}{14400} + \frac{41\xi_2^6}{172800} \quad \text{pour } r = 2$$

Pour  $r = 1$ , une justification détaillée est disponible dans [Monk et Parrott, 1994]. On s'aperçoit qu'on obtient un ordre 4 pour l'erreur de dispersion pour les éléments triangulaires de plus bas ordre (au lieu d'un ordre 2 pour les quadrangles). De plus les constantes sont très petites, de fait ces éléments dispersent très peu lorsqu'on a que des triangles équilatéraux. Malheureusement, ils sont très sensibles à la déformation du maillage. En effet, on obtient les relations de dispersion suivantes pour des triangles rectangles :

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 - \frac{\xi_1^4}{36} - \frac{\xi_1^3\xi_2}{18} + \frac{\xi_1^2\xi_2^2}{9} - \frac{\xi_1\xi_2^3}{18} - \frac{\xi_1^4}{36} \quad \text{pour } r = 1$$

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 + \frac{\xi_1^6}{16200} - \frac{13\xi_1^5\xi_2}{10800} + \frac{13\xi_1^4\xi_2^2}{5400} - \frac{13\xi_1^3\xi_2^3}{5400} + \frac{13\xi_1^2\xi_2^4}{5400} - \frac{13\xi_1\xi_2^5}{10800} + \frac{\xi_2^6}{16200} \quad \text{pour } r = 2$$

Ainsi pour  $R_1$ , on obtient un ordre 2 de dispersion et un ordre 4 pour  $R_2$ . On synthétise la différence triangles rectangles/ triangles équilatéraux sur le tableau 7.3. Seul  $R_1$  est très

Ordre d'approximation	R1	R2	R3	R4	R5
Triangles rectangles	$2.80e - 2K^2$	$1.74e - 2K^4$	$9.85e - 3K^6$	$7.34e - 3K^8$	$4.60e - 3K^{10}$
Triangles équilatéraux	$7.23e - 4K^4$	$7.87e - 3K^4$	$2.97e - 3K^6$	$1.41e - 3K^8$	$5.35e - 4K^{10}$

TAB. 7.3 – Constante  $L^2$  de l'erreur de dispersion pour les éléments finis triangles de la première famille sur des maillages triangulaires réguliers

sensible à la déformation du maillage, pour les ordres supérieurs, on obtient un ordre de  $2r$ , et les constantes sont du même ordre de grandeur, avec un avantage net pour les triangles équilatéraux.

En ce qui concerne la seconde famille, on obtient les relations suivantes pour les triangles équilatéraux :

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 + \frac{\xi_1^4}{24} + \frac{\xi_1^2\xi_2^2}{12} + \frac{\xi_2^4}{24} \quad \text{pour } r = 1$$

$$\omega^2 h^2 = \xi_1^2 + \xi_2^2 + \frac{3\xi_1^6}{6400} + \frac{9\xi_1^4\xi_2^2}{6400} + \frac{9\xi_1^2\xi_2^4}{6400} + \frac{3\xi_2^6}{6400} \quad \text{pour } r = 2$$

Bien que la seconde famille soit un “enrichissement” de la première famille ( $R_1 \subset (P_1)^2$ ,  $R_2 \subset (P_2)^2$ ), les triangles de la seconde famille sont plus dispersifs que ceux de la première famille, notamment pour l’ordre 1 et 2. Pour les ordres plus élevés, la différence est minime. On donne les constantes de l’erreur de dispersion pour les triangles de la seconde famille dans le tableau 7.4. Le

Ordre d’approximation	P1	P2	P3	P4	P5
Triangles rectangles	$1.94e - 1 K^2$	$6.54e - 2 K^4$	$3.82e - 2 K^6$	$2.07e - 2 K^8$	$1.48e - 2 K^{10}$
Triangles équilatéraux	$1.58e - 1 K^2$	$3.68e - 2 K^4$	$1.37e - 2 K^6$	$4.48e - 3 K^8$	$1.89e - 3 K^{10}$

TAB. 7.4 – Constante  $L^2$  de l’erreur de dispersion pour les éléments finis triangulaires de la seconde famille sur des maillages triangulaires réguliers

drame a lieu pour  $P1$  qui donne une erreur de dispersion d’ordre 2 sur des maillages équilatéraux alors qu’on a un ordre 4 pour  $R1$ . Ainsi  $P1$  nécessite deux fois plus de degrés de liberté que  $R1$  et détruit les bonnes propriétés de cet élément. C’est bien à cause de ce phénomène que les triangles de la première famille de plus bas ordre sont très populaires en électromagnétisme.

On illustre sur un cas concret la sensibilité de  $R1$  vis-à-vis du maillage. Le problème modèle est la diffraction par un point source dans une cavité carrée (cf. chapitre 4). La cavité est un carré  $[-5, 5]^2$ , la fréquence est prise égale à 1.01, le rayon de distribution de la gaussienne 0.6. On considère un maillage triangles rectangles et un maillage non-structuré ; on compare avec la solution de référence sur la figure 7.1. Les maillages utilisés pour  $R1$  contiennent eux 10 points par longueur d’onde. On voit que les triangles équilatéraux donnent une solution correcte (même si l’erreur  $L^2$  est de 20%), alors que les triangles rectangles donnent 100% d’erreur à cause de la dispersion numérique.

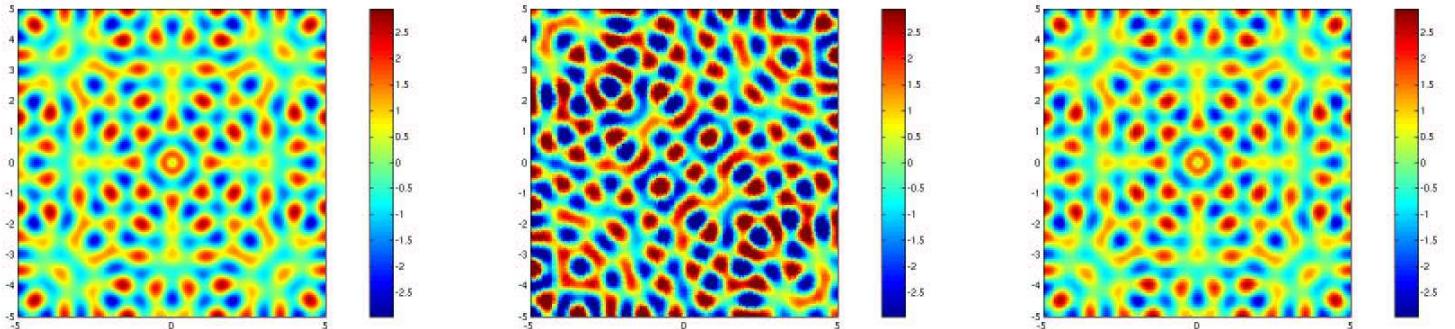


FIG. 7.1 – Point source dans une cavité. A gauche, solution numérique de référence, au milieu solution numérique avec un maillage de triangles rectangles, à droite avec un maillage de triangles quasi-équilatéraux. Triangles de la première famille, maillés avec dix points par longueur d’onde.

Sur ce même cas test, on illustre la dispersion pour  $Q_1$  sur des maillages réguliers. Sur la figure 7.2, on voit que la solution obtenue à l’aide d’une intégration approchée possède les mêmes noeuds et ventres que la solution de référence, alors que l’intégration exacte fournit une solution fausse à 100%. Les maillages contiennent plus de 60 points par longueur d’onde. On notera que l’erreur en dispersion en  $O(K^{2r})$  sur des maillages triangulaires découpés en quadrilatères n’est obtenue que si on utilise  $(r+1)^2$  points de Gauss-Lobatto pour la matrice de masse et  $r^2$  points de Gauss pour la matrice de rigidité, comme on l’a présenté au chapitre 5. Si on met des points de Gauss-Lobatto pour intégrer la matrice de rigidité, on retombe sur une convergence en  $O(K^{2r-2})$ .

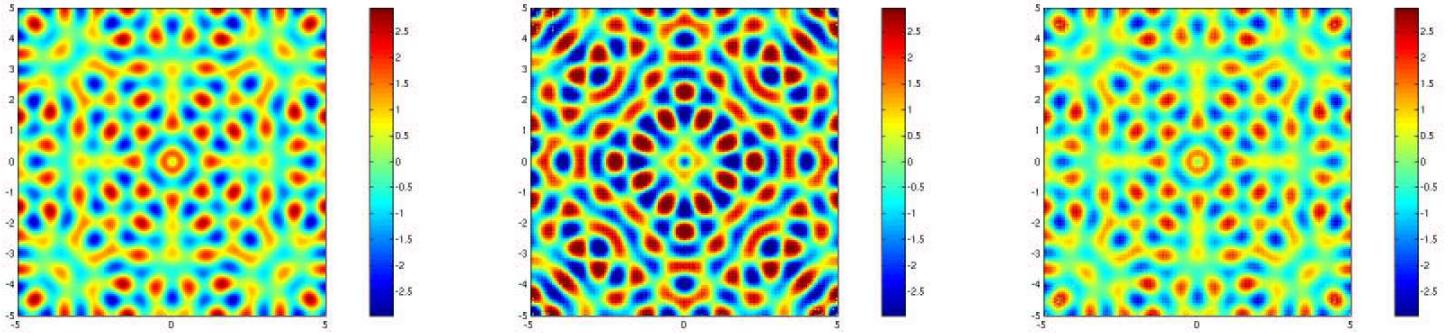


FIG. 7.2 – Point source dans une cavité. A gauche, solution numérique de référence, au milieu solution numérique avec une intégration exacte, à droite avec une intégration approchée. Quadrilatères de la première famille, maillés avec 60 points par longueur d’onde.

### 7.1.2 Cas 3-D

#### Eléments finis hexaédriques

La dispersion des hexaèdres de la première famille sur maillage régulier est identique au cas 2-D (et donc à Helmholtz). L’intégration numérique approchée donne une dispersion plus faible. Dans le cas des tétraèdres découpés, les calculs sont relativement onéreux, on se contente de les faire pour l’ordre 1 et 2. Le maillage utilisé est celui de la figure 3.3. On trouve les constantes du tableau 7.5. La déformation du maillage nous fait perdre deux ordres, on obtient une erreur de

Ordre	$Q_1$	$Q_2$
Intégration exacte	$3.71e - 2 K^0$	$5.54e - 3 K^2$
Intégration approchée	$3.28e - 2 K^0$	$2.94e - 3 K^2$

TAB. 7.5 – Constante  $L^2$  de l’erreur de dispersion sur un maillage tétraèdrique découpé.

dispersion en  $O(K^{2r-2})$ , que ce soit en évaluant de manière exacte les intégrales ou de manière approchée en utilisant les points de Gauss-Lobatto. C’est surtout pénalisant pour  $Q_1$ , qui en conséquence n’est pas consistant, que ce soit en terme d’erreur d’interpolation ou d’erreur de dispersion. Les constantes sont sensiblement les mêmes suivant la méthode d’intégration choisie. On notera que cette perte d’ordre est vraie pour un maillage hexaédrique légèrement modifié (par exemple celui de la figure 3.4). C’est donc bien la non-linéarité de la transformation  $F_i$  qui est la cause de cette perte d’ordre.

#### Eléments finis tétraédriques

On ne considère que les tétraèdres de la première famille. Malheureusement en 3-D, il est impossible de générer des maillages tétraédriques réguliers pour mailler l’espace. C’est par ailleurs un sujet ouvert que de trouver un motif élémentaire pour mailler l’espace de telle sorte que les tétraèdres soient le plus “réguliers” possible, comme le montre l’article [Eppstein *et al.*, 2004]. Par conséquent, nous nous contentons des deux motifs élémentaires du chapitre 3, cf. figures 3.3 et 3.5. Nous obtenons les constantes de dispersion du tableau 7.6. Nous observons comme en 2-D, une sensibilité importante vis-à-vis du maillage pour  $R_1$ . Lorsqu’on utilise un maillage plus “régulier”, on divise par 5 l’erreur de dispersion. Cela signifie qu’il faut prendre un maillage 2.2 fois plus fin si on utilise des tétraèdres droits et qu’on veut obtenir une erreur de dispersion semblable aux tétraèdres dits réguliers.

Ordre	1	2	3
$R_k$ sur des tétraèdres droits	$3.7e - 1 K^2$	$4.99e - 2 K^4$	$2.96e - 2 K^6$
$R_k$ sur des tétraèdres “réguliers”	$6.76e - 2 K^2$	$8.4e - 3 K^4$	$5.6e - 3 K^6$
$Q_k$ sur des hexaèdres réguliers	$6.53e - 2 K^2$	$8.27e - 3 K^4$	$1.81e - 3 K^6$

TAB. 7.6 – Constante  $L^2$  de l’erreur de dispersion pour les éléments tétraédriques de la première famille. Comparaison avec les hexaèdres de la première famille sur maillage régulier.

Pour conclure, il semble plus judicieux de choisir  $R_1$  et  $R_2$  plutôt que  $Q_1$  et  $Q_2$  car ils donnent une dispersion équivalente à ces derniers sur des maillages “réguliers”. Ils ont l’avantage considérable de donner un ordre optimal pour l’erreur de dispersion sur des maillages non-structurés alors que les hexaèdres souffrent d’une perte de précision dans ce cas. Pour des ordres supérieurs, il est préférable d’utiliser  $Q_k$ , car la perte de précision est moins criante et le gain en stockage et en temps de calcul compense largement cette perte de précision.

## 7.2 Cas académique de la sphère

Nous faisons des comparaisons sur le cas académique de la sphère parfaitement conductrice et de la sphère diélectrique. Nous calculons sur ces objets le champ lointain, appelé aussi S.E.R (Section Équivalente Radar). La SER est obtenue en évaluant l’intégrale suivante pour différents vecteurs unitaires  $\mathbf{u}$  :

$$\boldsymbol{\sigma}(\mathbf{u}) = \frac{k^2}{4\pi} \int_{\Sigma} e^{ik\mathbf{u}\cdot\mathbf{OM}} \left[ \mathbf{u} \times (\mathbf{n} \times \mathbf{H}) + (u \otimes u - I)(\mathbf{E} \times \mathbf{n}) \right] dM$$

Pour la justification de cette formule, le lecteur pourra se référer à [Monk, 2002].  $\mathbf{u}$  est donc la direction pour laquelle on désire connaître le champ lointain  $\boldsymbol{\sigma}$ . Le plus souvent, on impose au vecteur directeur d’appartenir à un plan, et on fait varier l’angle d’incidence  $\theta$ . Le champ lointain est un vecteur complexe, dont on calcule la grandeur :

$$\text{SER} = 10 \log_{10}(\|\boldsymbol{\sigma}\|^2)$$

C’est cette grandeur qu’on affichera quand on parlera de SER sur les figures (ou RCS = Radar Cross Section, pour la traduction anglaise). Afin de traiter des domaines bornés, on utilise la condition transparente présentée en annexe B. On utilise un critère d’arrêt de  $10^{-6}$  pour le solveur de la matrice éléments finis, et un critère d’arrêt de  $10^{-3}$  pour le solveur de la condition transparente. Pour calculer le champ lointain, on fait une intégration numérique standard. On a alors besoin d’évaluer  $H$  en des points de quadrature. Pour ce faire, on calcule le rotationnel de  $E$  au niveau discret. Comme on l’a signalé,  $E$  est assez bien évalué alors que son rotationnel l’est moins bien. Il est probable qu’on aurait de meilleurs résultats si on effectuait une évaluation “variationnelle” de  $H$  comme décrit dans [Monk et Parrott, 2001]. De même, la condition transparente nécessite l’évaluation de  $H$ .

Nous utilisons les éléments finis tétraédriques de la première famille de Nédélec, et comme discréttisation de cette espace, les éléments finis proposés par Graglia *et al.* [1997]. D’autres éléments finis tétraédriques d’ordre élevé ont été proposés, mais surtout pour la seconde famille dont [Webb, 1999], [Ainsworth et Coyle, 2003], [Demkowicz, 2000].

### 7.2.1 Coût du produit matrice-vecteur

On fait ici une comparaison purement numérique entre le coût d’un produit matrice vecteur avec les tétraèdres et avec les hexaèdres. On obtient la figure 7.3 sur un cas de 200 000 degrés

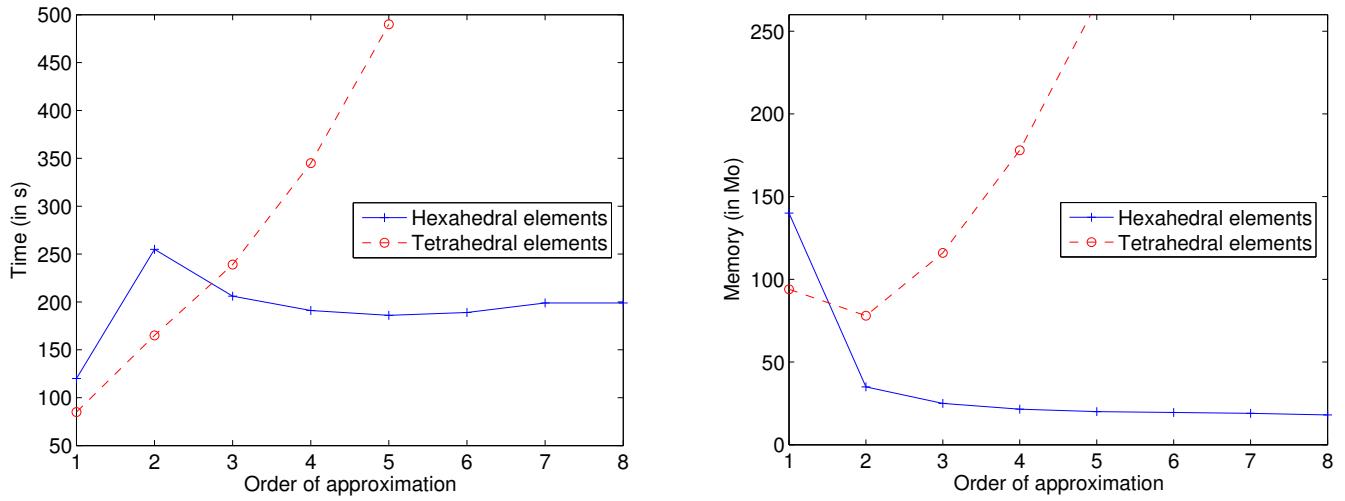


FIG. 7.3 – A gauche, temps pour 1 000 itérations de COCG sur un cas de 200 000 ddl. A droite, stockage requis avant le début des itérations.

de liberté. On voit que les hexaèdres sont plus rapides que les tétraèdres dès l'ordre 3. Pour le stockage, les hexaèdres sont moins coûteux dès l'ordre 2. On a choisi ici de faire figurer le stockage nécessaire pour construire la matrice, le maillage ... Ce stockage comprend donc toutes les variables créées excepté les vecteurs d'itérations. On voit notamment que les méthodes d'ordre un sont pénalisées à cause du maillage qui coûte très cher. En effet, on stocke toutes les informations sur les faces, arêtes, sommets et les informations de connectivité entre ces entités. Pour chaque élément du maillage, on connaît la liste des sommets, des arêtes, des degrés de liberté, des faces. Pour chaque face, on connaît les deux éléments adjacents, la liste des arêtes de la face, la liste des sommets. Toute cette connectivité a son utilité lorsqu'on utilise de l'ordre élevé, c'est un peu superflu pour les méthodes d'ordre 1. A mon sens, faire de l'ordre 1 est difficile dans le sens où il faut ne stocker que les tableaux dont on a strictement besoin, pour gagner de la mémoire. L'avantage des méthodes d'ordre élevé qu'on propose est de s'affranchir de cette contrainte et de conduire à un faible stockage sans trop d'efforts.

A cause du coût de stockage trop important pour les tétraèdres d'ordre élevé, et d'une convergence trop lente des tétraèdres d'ordre bas, nous ne présentons pas de résultats numériques sur les tétraèdres.

## 7.2.2 Sphère parfaitement conductrice

On étudie la diffraction par une sphère parfaitement conductrice de rayon 4 (voir figure 7.4). On se fixe comme objectif d'atteindre une erreur maximale de 0.5 dB sur la SER. Pour illustrer ce seuil, on affiche la SER analytique et la SER avec 0.43 dB d'erreur obtenue pour du  $Q_4$ , sur la figure 7.5. On obtient les résultats du tableau 7.7. Sur ce cas-là,  $Q_8$  est optimal, car on a alors besoin de ne mettre qu'une seule maille entre la sphère intérieure (condition de Dirichlet) et la sphère extérieure (condition transparente).  $Q_2$  nécessite plus d'un million de degrés de liberté alors qu'on est en maillage régulier. Sur des tétraèdres découpés, c'est inutile d'espérer que  $Q_2$  convergera vers une solution assez précise !

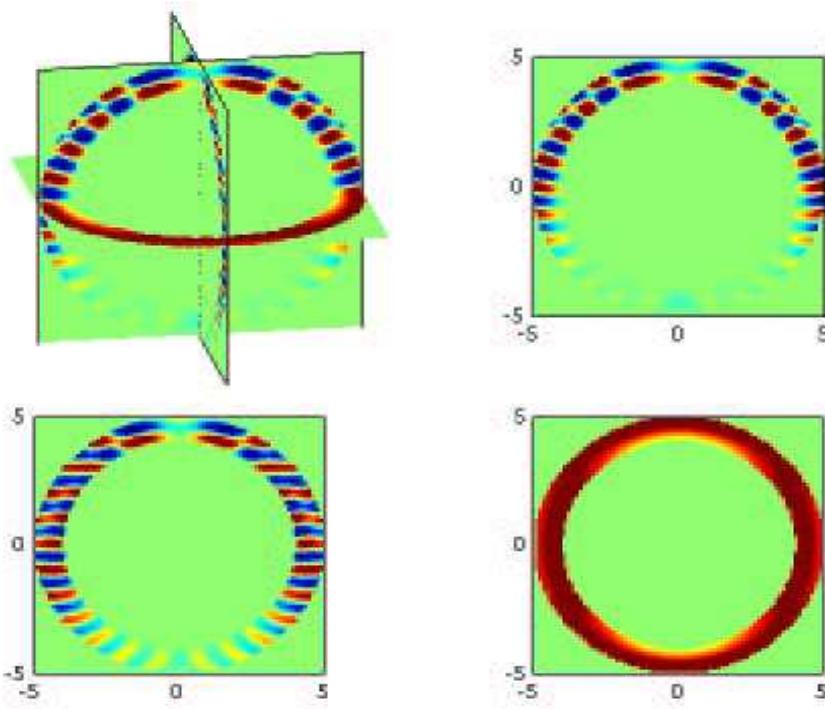


FIG. 7.4 – Partie réelle du champ total pour une sphère parfaitement conductrice. Composante suivant x du champ électrique

Ordre	Nombre ddl	Erreur	BICGCR	ILUT(0.05)	Helmholtz	2-grille
$Q_2$	1 015 000	0.74 dB	4 740 s 3 660 (186 Mo)	- -	5 186 s 801 (587 Mo)	1 556 s 133 (938 Mo)
$Q_4$	250 000	0.54 dB	852 s 3 457 (30 Mo)	146 s 213 (240 Mo)	706 s 561 (126 Mo)	376 s 202 (180 Mo)
$Q_6$	295 000	0.2 dB	2 484 s 8 404 (33 Mo)	218 s 276 (268 Mo)	2 796 s 1 870 (149 Mo)	1 382 s 741 (207 Mo)
$Q_8$	160 000	0.28 dB	864 s 5 041 (21 Mo)	109 s 262 (133 Mo)	550 s 793 (81 Mo)	543 s 622 (109 Mo)

TAB. 7.7 – Performances sur la sphère parfaitement conductrice. Sur les lignes du bas, on fait figurer le nombre d’itérations ainsi que la place mémoire utilisée lors de la simulation.

### 7.2.3 Sphère diélectrique

On étudie la diffraction par une sphère diélectrique de rayon 2 (voir figure 7.6), avec pour indices :

$$\varepsilon = 3.5 \quad \mu = 1$$

On se fixe comme objectif d’atteindre une erreur maximale de 0.5 dB sur la SER. Pour illustrer ce seuil, on affiche la SER analytique et la SER avec 0.3 dB d’erreur obtenue pour du  $Q_4$ , sur la figure 7.7. Sur ce cas test,  $Q_8$  est encore bien adapté, tout comme  $Q_4$ .

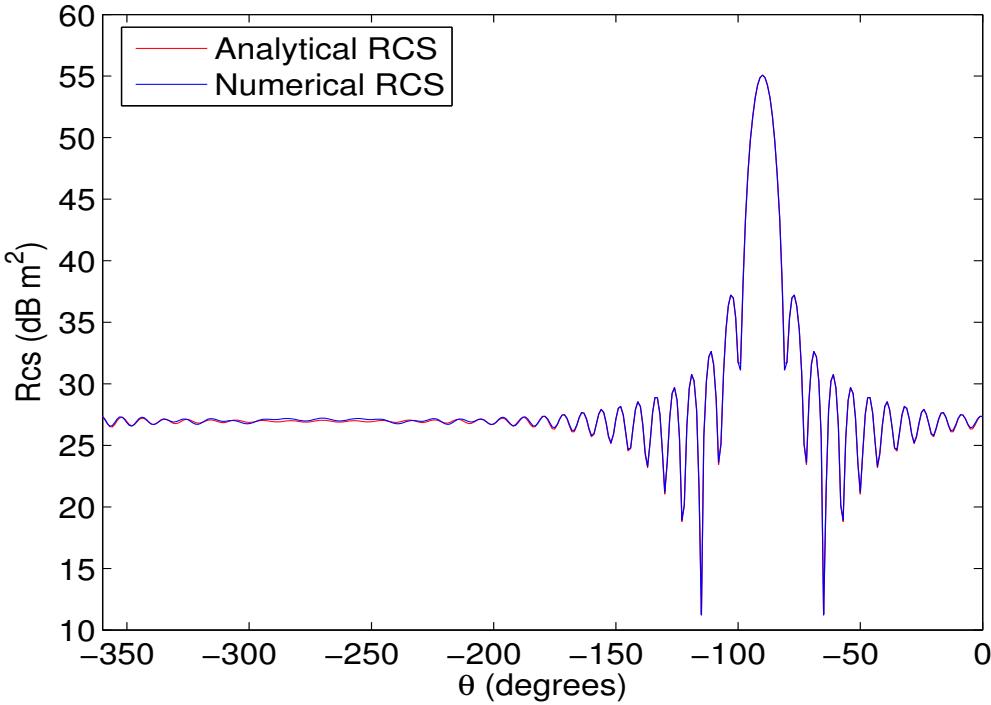


FIG. 7.5 – Section Equivalente Radar pour une sphère parfaitement conductrice. Le maillage  $Q_4$  utilisé est maillé avec 8 points par longueur d'onde (250 000 ddl environ).

Ordre	Nombre ddl	Erreur	BICGCR	ILUT(0.05)	Helmholtz	2-grille
$Q_2$	940 000	0.95 dB	19 486 s 15 227 (171 Mo)	- -	17 970 s 3 603 (574 Mo)	4 344 s 628 (947 Mo)
$Q_4$	88 000	0.30 dB	894 s 9 886 (10 Mo)	189 s 718 (99 Mo)	2 305 s 5 646 (47 Mo)	488 s 813 (67 Mo)
$Q_6$	230 000	0.18 dB	4 401 s 18 800 (24 Mo)	1 035 s 1 455 (271 Mo)	3 787 s 3 479 (123 Mo)	1 095 s 794 (180 Mo)
$Q_8$	88 000	0.03 dB	1 484 s 15 300 (10 Mo)	307 s 1 200 (90 Mo)	5 260 s 12 700 (47 Mo)	952 s 1 800 (66 Mo)

TAB. 7.8 – Performances sur la sphère diélectrique

### 7.3 Cavité cobra

On étudie la diffraction par une cavité cobra (cf. figure 7.8). On se fixe comme objectif d'atteindre une erreur maximale de 0.5 dB sur la SER. La SER de référence est affichée sur la figure 7.9. Sur ce cas test,  $Q_6$  est mieux adapté que  $Q_4$  et  $Q_8$  (on n'a pas communiqué les

Ordre	Nombre ddl	Erreur	BICGCR	ILUT(0.05)	Helmholtz	2-grille
$Q_4$	412 000	0.45 dB	14 039 s 34 800 (47 Mo)	2 247 s 1 900 (391 Mo)	9 371 s 5 800 (184 Mo)	9 294 s 4 300 (260 Mo)
$Q_6$	187 000	0.4 dB	12 096 s 31 500 (22 Mo)	846 s 1 700 (161 Mo)	4 821 s 6 400 (87 Mo)	10 063 s 10 500 (130 Mo)

TAB. 7.9 – Performances sur la cavité cobra

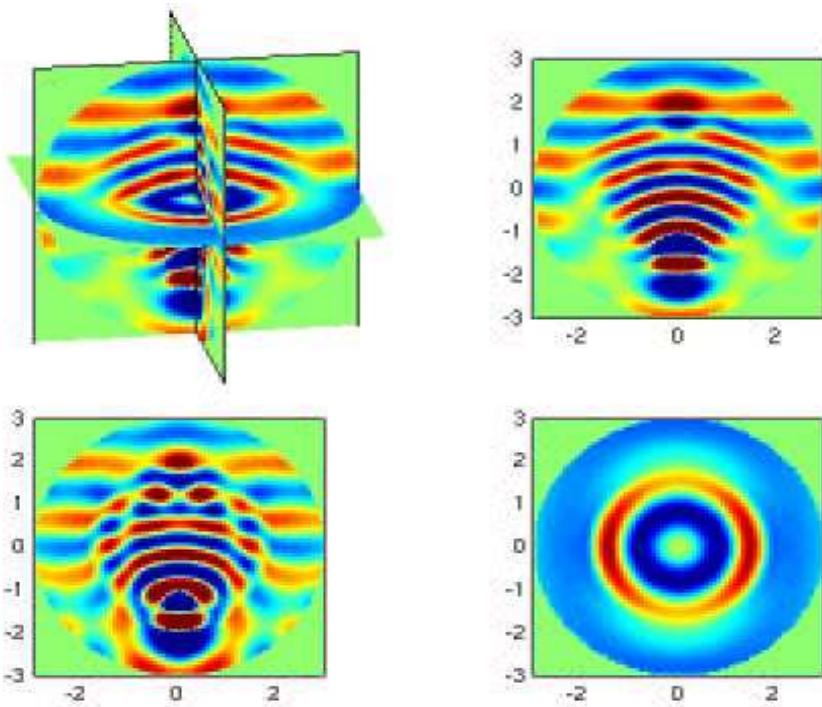


FIG. 7.6 – Partie réelle du champ total pour une sphère diélectrique. Composante suivant  $y$  du champ magnétique

statistiques de ce dernier).

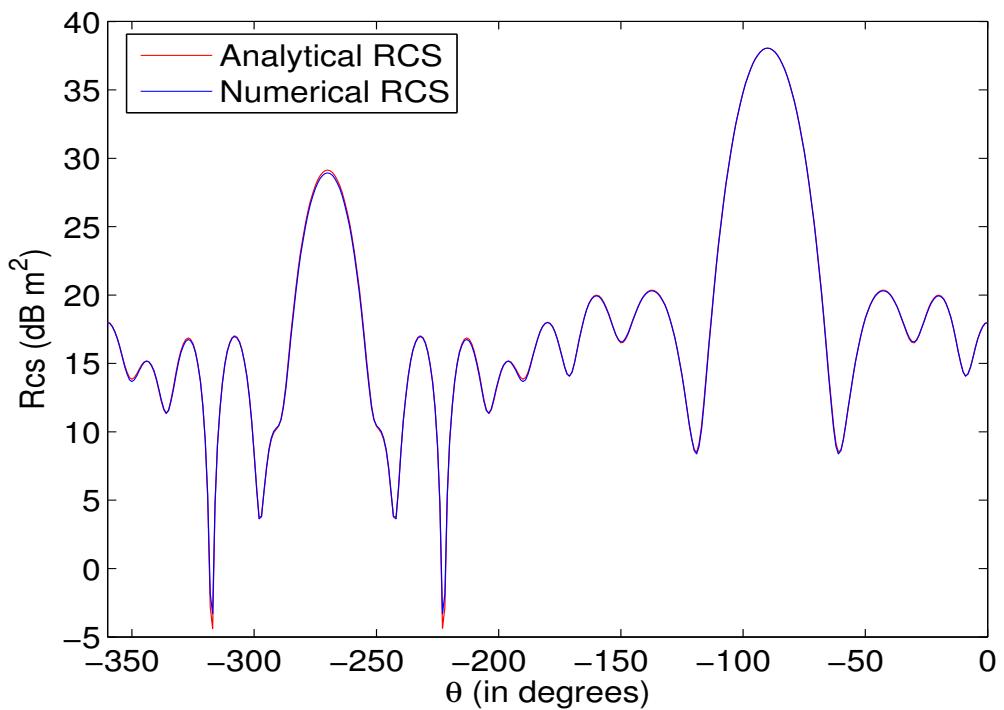


FIG. 7.7 – Section Equivalente Radar pour une sphère diélectrique. Le maillage  $Q_4$  utilisé est maillé avec 8 points par longueur d’onde (90 000 ddl environ).

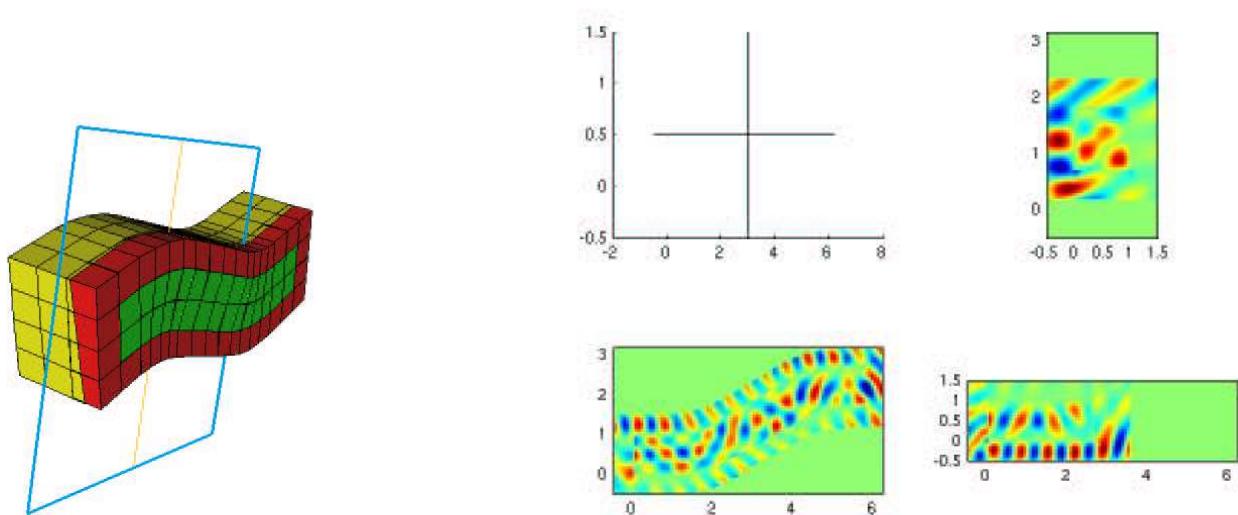


FIG. 7.8 – A gauche, maillage de la cavité cobra, à droite, partie réelle du champ total pour une cavité cobra. Composante suivant x du champ électrique.

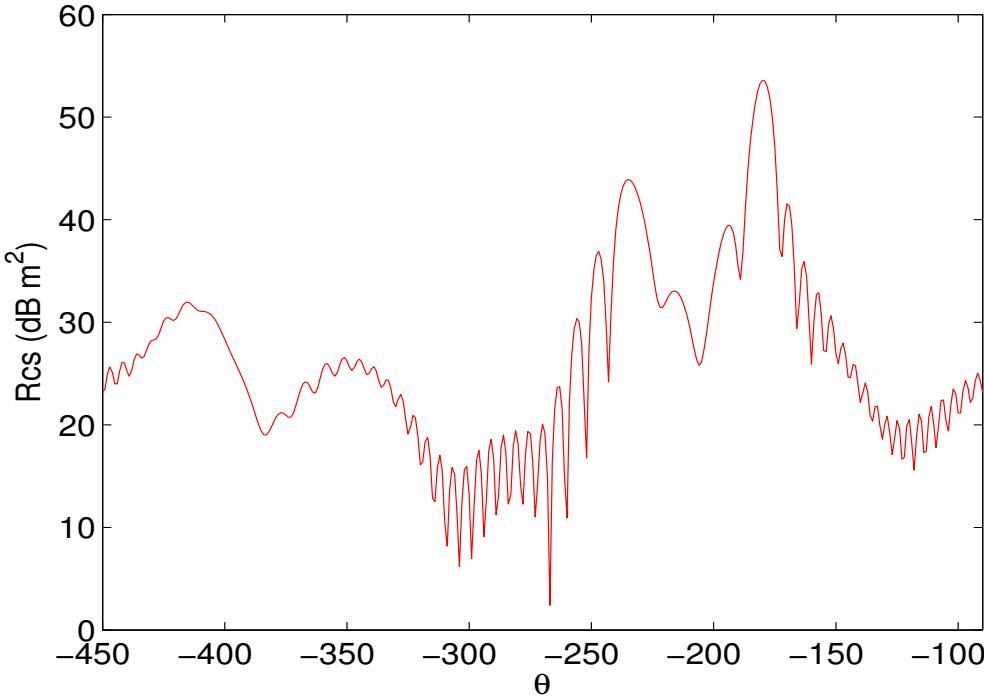


FIG. 7.9 – Section Equivalente Radar pour une cavité cobra.

## 7.4 Conclusion

Dans ce chapitre, nous avons fait une étude de dispersion sur des maillages non-réguliers. En 2-D, les triangles de la première famille de plus bas ordre ont la remarquable propriété d'avoir une erreur de dispersion en  $O(h^4)$  sur des maillages triangulaires équilatéraux, alors qu'elle est en  $O(h^2)$  sur des maillages “triangles rectangles”. Lorsqu'on monte en ordre sur les triangles de la première famille, on n'a pas de gain d'ordre, mais des constantes plus petites pour les triangles équilatéraux. Les triangles de la seconde famille sont plus dispersifs que la première famille. En ce qui concerne les quadrilatères de la première famille, on obtient une erreur de dispersion en  $O(h^{2r-2})$  lorsque l'on calcule de manière exacte les intégrales, et une erreur en  $O(h^{2r})$  lorsqu'elles sont calculées de manière approchées. De plus, les constantes sont plus plus petites que pour les triangles de la première famille, pour des ordres d'approximation élevés. En 3-D, on a toujours une erreur de dispersion en  $O(h^{2r-2})$  sur les hexaèdres de la première famille, que ce soit avec des intégrales approchées ou exactes, sur des maillages quelconques.

Nous avons également comparé les tétraèdres et les hexaèdres au niveau du produit matrice-vecteur. A nombre de degrés de liberté égal, les hexaèdres sont plus rapides que les tétraèdres dès l'ordre 3. On obtient un gain en stockage à partir de l'ordre 2. Les tétraèdres demandent un stockage beaucoup trop important lorsqu'on veut monter en ordre.

Sur l'ensemble des cas-tests présentés, le préconditionneur le plus efficace est la factorisation incomplète devant le multigrille. Ce dernier est peu efficace sur des ordres très élevés comme  $Q_6$  et  $Q_8$ , il donne des résultats très satisfaisants pour  $Q_2$ , un peu moins satisfaisants pour  $Q_4$  et décevants pour les ordres supérieurs. Il faudrait probablement éviter le passage par le sous-maillage  $Q_1$ , pour obtenir un algorithme plus robuste.

## Troisième partie

# Equations de Maxwell en domaine axisymétrique



## Chapitre 8

# Résolution des équations de Maxwell axi-symétriques par éléments finis d'arête

*Nous nous intéressons ici à la résolution des équations de Maxwell pour des domaines présentant une symétrie de révolution. La source - en général une onde plane - ne présente pas cette symétrie. On décompose la source et la solution en séries de Fourier suivant l'angle de révolution  $\theta$ . On aboutit à des problèmes 2-D indépendants, qu'on peut alors résoudre par une méthode éléments finis. Dans la première section, nous décrivons la discrétisation choisie, ses avantages par rapport aux autres choix possibles. Dans la seconde section, nous nous attachons à montrer via des exemples numériques que ce choix est satisfaisant.*

### Sommaire

---

<b>8.1</b>	<b>Description de la méthode éléments finis</b>	<b>186</b>
8.1.1	Choix de la formulation variationnelle	186
8.1.2	Calcul de la matrice éléments finis	192
<b>8.2</b>	<b>Précision de la méthode</b>	<b>198</b>
8.2.1	Cas de la sphère parfaitement conductrice	198
8.2.2	Cas du cone-sphère	198
8.2.3	Cas du cylindre	201
<b>8.3</b>	<b>Conclusion</b>	<b>202</b>

---

## 8.1 Description de la méthode éléments finis

### 8.1.1 Choix de la formulation variationnelle

#### Equations de Maxwell en coordonnées cylindriques

Le changement des coordonnées cartésiennes (x,y,z) vers les coordonnées cylindriques ( $r, \theta, z$ ) s'écrit :

$$\begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta \\ z &= z \end{aligned}$$

La base cylindrique ( $\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{z}}$ ) s'exprime :

$$\hat{\mathbf{r}} = \begin{vmatrix} \cos \theta \\ \sin \theta \\ 0 \end{vmatrix}, \quad \hat{\boldsymbol{\theta}} = \begin{vmatrix} -\sin \theta \\ \cos \theta \\ 0 \end{vmatrix}, \quad \hat{\mathbf{z}} = \begin{vmatrix} 0 \\ 0 \\ 1 \end{vmatrix}$$

Le rotationnel d'un champ de vecteurs  $\mathbf{E}$  devient dans les coordonnées cylindriques :

$$\text{rot}(E) = \begin{vmatrix} \frac{1}{r} \left( \frac{\partial E_z}{\partial \theta} - \frac{\partial (rE_\theta)}{\partial z} \right) \\ \frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r} \\ \frac{1}{r} \left( \frac{\partial (rE_\theta)}{\partial r} - \frac{\partial E_r}{\partial \theta} \right) \end{vmatrix}$$

On considère les équations de Maxwell :

$$-i\omega \varepsilon \mathbf{E} - \text{rot} \mathbf{H} = 0$$

$$-i\omega \mu \mathbf{H} + \text{rot} \mathbf{E} = 0$$

On décompose le champ électrique et le champ magnétique en séries de Fourier :

$$\mathbf{E} = \sum_{m=-\infty}^{+\infty} \begin{vmatrix} E_{r,m} \\ E_{\theta,m} \\ E_{z,m} \end{vmatrix} e^{-im\theta}, \quad \mathbf{H} = \sum_{m=-\infty}^{+\infty} \begin{vmatrix} H_{r,m} \\ H_{\theta,m} \\ H_{z,m} \end{vmatrix} e^{-im\theta}$$

m sera appelé numéro de mode. Par simple propriété d'orthogonalité , chaque mode ( $\mathbf{E}_m, \mathbf{H}_m$ ) est solution d'un problème indépendant 2-D :

$$\begin{cases} -i\omega \varepsilon r E_r = -im H_z - \frac{\partial(rH_\theta)}{\partial z} \\ -i\omega \varepsilon E_\theta = \frac{\partial H_r}{\partial z} - \frac{\partial H_z}{\partial r} \\ -i\omega \varepsilon r E_z = \frac{\partial(rH_\theta)}{\partial r} + im H_r \end{cases} \quad \begin{cases} -i\omega \mu r H_r = \frac{\partial(rE_\theta)}{\partial z} + im E_z \\ -i\omega \mu H_\theta = \frac{\partial E_z}{\partial r} - \frac{\partial E_r}{\partial z} \\ -i\omega \mu r H_z = -im E_r - \frac{\partial(rE_\theta)}{\partial r} \end{cases}$$

On introduit les notations suivantes :

$$\mathbf{rot}(u(r, z)) = \begin{vmatrix} \frac{\partial u}{\partial z} \\ -\frac{\partial u}{\partial r} \end{vmatrix} \quad \text{rot}(\mathbf{v}) = \frac{\partial v_z}{\partial r} - \frac{\partial v_r}{\partial z} \quad \tilde{\mathbf{v}} = \begin{vmatrix} v_z \\ -v_r \end{vmatrix}$$

On note les nouvelles inconnues :

$$\mathbf{E} = \begin{vmatrix} E_r \\ E_z \end{vmatrix} \quad \mathbf{H} = \omega \begin{vmatrix} H_r \\ H_z \end{vmatrix} \quad \bar{E}_\theta = i E_\theta \quad \bar{H}_\theta = i \omega H_\theta$$

On obtient le système d'équations suivant :

$$-\omega^2 \varepsilon \mathbf{E} + \frac{m}{r} \tilde{\mathbf{H}} - \frac{1}{r} \text{rot}(r H_\theta) = 0 \quad (8.1)$$

$$\mu \mathbf{H} + \frac{m}{r} \tilde{\mathbf{E}} - \frac{1}{r} \text{rot}(r E_\theta) = 0 \quad (8.2)$$

$$-\omega^2 \varepsilon E_\theta + \text{rot} \mathbf{H} = 0 \quad (8.3)$$

$$\mu H_\theta + \text{rot} \mathbf{E} = 0 \quad (8.4)$$

On a  $\mathbf{E}, \mathbf{H} \in H(\text{rot}, \Omega)$  et  $E_\theta, H_\theta \in H^1(\Omega)$ .

On peut se poser la question du choix d'inconnues principales à choisir. On expose en premier lieu en quoi le choix de  $E_\theta$  et  $H_\theta$  mène à une impasse. Dans un second temps, on fait le choix plus "classique" de  $\mathbf{E}$  et  $E_\theta$ . Ce choix aboutit à l'évaluation d'intégrales divergentes en  $1/r$ , nous montrons comment on peut détourner cette difficulté apparente.

### Pourquoi ne pas prendre $E_\theta$ et $H_\theta$ ?

Un choix attractif est de ne choisir que  $E_\theta$  et  $H_\theta$ . On aboutit alors au système d'équations suivantes :

$$\begin{cases} -\omega^2 \varepsilon E_\theta - \text{rot} [\zeta(r) (m r \nabla(r H_\theta) - \omega^2 \varepsilon r \mathbf{rot}(r E_\theta))] = 0 \\ -\mu r H_\theta + \text{rot} [\zeta(r) (m \nabla(r E_\theta) + \mu r \mathbf{rot}(r H_\theta))] = 0 \end{cases} \quad (8.5)$$

en notant :

$$\zeta(r) = \frac{1}{\omega^2 \varepsilon \mu r^2 - m^2}$$

La formulation variationnelle s'obtient par intégration par parties sur les deux équations :

$$\begin{cases} -\omega^2 \int_{\Omega} \varepsilon r E_\theta \psi + \omega^2 \int_{\Omega} \varepsilon r \zeta(r) \mathbf{rot}(r E_\theta) \cdot \mathbf{rot}(r \psi) - m \int_{\Omega} \zeta(r) \mathbf{rot}(r \psi) \cdot \nabla(r H_\theta) = 0 \\ - \int_{\Omega} \mu r H_\theta \lambda + \int_{\Omega} \mu r \zeta(r) \mathbf{rot}(r H_\theta) \cdot \mathbf{rot}(r \lambda) + m \int_{\Omega} \zeta(r) \nabla(r E_\theta) \cdot \mathbf{rot}(r \lambda) = 0 \end{cases} \quad (8.6)$$

Les variables  $E_\theta$  et  $H_\theta$  sont dans  $H^1(\Omega)$ , ainsi que les fonctions tests  $\psi$  et  $\lambda$ . Pour avoir une formulation variationnelle bien définie, il est nécessaire de rajouter une condition sur l'axe  $r = \frac{m}{k}$  :

$$m \nabla(r E_\theta) + \mu r \mathbf{rot}(r H_\theta) = 0$$

En pratique, il est difficile d'imposer cette condition aux limites.

## Choix plus classique de $\mathbf{E}$ et $E_\theta$

L'équation (8.2) nous fournit les deux relations :

$$\mu \tilde{\mathbf{H}} = \frac{1}{r} (m \mathbf{E} - \nabla(r E_\theta))$$

$$\text{rot}(H) = m \text{div}\left(\frac{\mathbf{E}}{\mu r}\right) + \text{rot}\left(\frac{1}{\mu r} \text{rot}(r E_\theta)\right)$$

Ces deux relations et l'équation (8.4) nous permet d'éliminer les inconnues  $\mathbf{H}$  et  $\mathbf{H}_\theta$  :

$$(-\omega^2 \varepsilon r + \frac{m^2}{\mu r}) \mathbf{E} - \frac{m}{\mu r} \nabla(r E_\theta) + \text{rot}\left(\frac{r}{\mu} \text{rot} E\right) = 0 \quad (8.7)$$

$$-\omega^2 \varepsilon r E_\theta + m r \text{div}\left(\frac{\mathbf{E}}{\mu r}\right) + r \text{rot}\left(\frac{1}{\mu r} \text{rot}(r E_\theta)\right) = 0 \quad (8.8)$$

On effectue la formulation variationnelle de ce système afin d'obtenir une formulation variationnelle symétrique :

$$-\omega^2 \int_{\Omega} \varepsilon r \mathbf{E} \cdot \boldsymbol{\varphi} + m \int_{\Omega} \frac{1}{\mu r} (m \mathbf{E} - \nabla(r E_\theta)) \cdot \boldsymbol{\varphi} + \int_{\Omega} \frac{r}{\mu} \text{rot} \mathbf{E} \text{rot} \boldsymbol{\varphi} = 0 \quad (8.9)$$

$$-\omega^2 \int_{\Omega} \varepsilon r E_\theta \psi - \int_{\Omega} \frac{1}{\mu r} (m \mathbf{E} - \nabla(r E_\theta)) \cdot \nabla(r \psi) = 0 \quad (8.10)$$

Afin de mettre en évidence la symétrie, on peut ajouter les deux équations et obtenir le problème suivant :

$$\begin{cases} \text{Trouver } (\mathbf{E}, E_\theta) \in \text{H(rot, } \Omega) \times H^1(\Omega) \text{ tel que} \\ -\omega^2 \int_{\Omega} \varepsilon r (\mathbf{E} \cdot \boldsymbol{\varphi} + E_\theta \psi) dr dz + \int_{\Omega} \frac{1}{\mu r} (m \mathbf{E} - \nabla(r E_\theta)) \cdot (m \boldsymbol{\varphi} - \nabla(r \psi)) dr dz \\ + \int_{\Omega} \frac{r}{\mu} \text{rot}(\mathbf{E}) \text{rot}(\boldsymbol{\varphi}) dr dz = 0 \quad \forall (\boldsymbol{\varphi}, \psi) \in \text{H(rot, } \Omega) \times H^1(\Omega) \end{cases}$$

Le lecteur aura noté la présence de  $1/r$  sur une des intégrales, rendant nécessaire l'ajout d'une condition aux limites sur l'axe. A cette formulation variationnelle, il faut rajouter la condition aux limites :

$$(m \mathbf{E} - \nabla(r E_\theta)) = 0 \quad \text{pour } r = 0$$

Cette condition est évidemment vérifiée par les solutions du problème continu car :

$$(m \mathbf{E} - \nabla(r E_\theta)) = \mu r \tilde{\mathbf{H}}$$

En projetant cette condition sur les deux composantes, on obtient les deux conditions :

$$m E_r - E_\theta = 0 \quad (8.11)$$

$$m E_z = 0 \quad (8.12)$$

Il est difficile d'imposer ces conditions dans l'espace discret. En effet, lorsqu'on adopte une discréttisation par éléments finis d'arête, on ne sait traiter que des conditions aux limites portant sur  $\mathbf{E} \times n$ . Par conséquent, les inconnues discrètes ne vérifieront pas a priori les conditions (8.11) et (8.12). La formulation variationnelle fait donc intervenir des intégrales divergentes. Le papier [Hiptmair et Ledger, 2005] propose d'évaluer ces intégrales avec  $(3r + 1)^2$  points de quadrature sans avoir besoin d'un traitement supplémentaire. Quelque part, on peut

interpréter cette technique de surintégration comme une technique de pénalisation. On met de manière artificielle des coefficients élevés dans la matrice afin de forcer le système discret à vérifier les conditions aux limites voulues. Cette justification est heuristique, les résultats numériques montrent que cette technique de surintégration fonctionne correctement et qu'il n'est pas nécessaire de prendre autant de points d'intégration, une intégration normale avec  $(r + 1)^2$  points de Gauss suffit (voir figure 8.2). Le cas test choisi est la diffraction par une sphère diélectrique d'indices  $\epsilon = 1.0$   $\mu = 3.5$  (cf. figure 8.1). La méthode ne paraît pas

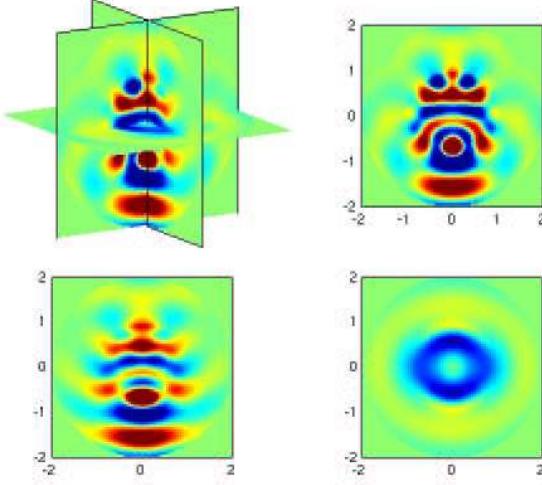


FIG. 8.1 – Partie réelle du champ diffracté par une sphère diélectrique. Composante x du champ électrique.

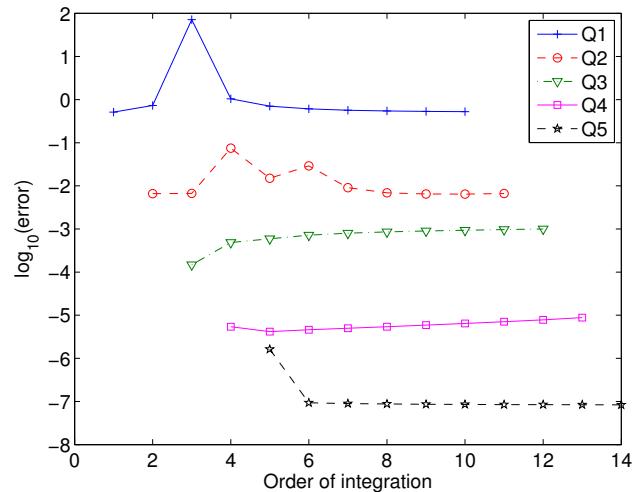


FIG. 8.2 – Evolution de l'erreur entre la solution numérique et la solution analytique pour divers ordres d'approximations. Utilisation du même maillage triangles découpés pour tous les ordres.

très robuste à cause de cette difficulté sur l'axe. Cette difficulté peut être levée en utilisant des éléments finis nodaux au lieu des éléments finis d'arête, la condition aux limites est alors simple à traiter. Néanmoins, il faut un traitement spécifique des singularités géométriques, on renvoie le lecteur aux travaux de [Assous *et al.*, 2003].

## Modification de la formulation variationnelle

Nous proposons dans cette sous-sous-section une approche intéressante pour lever la difficulté des intégrales en  $1/r$ . L'idée originale vient de [Lacoste, 2000], elle consiste à faire le changement de variable :

$$\mathbf{U} = \frac{(m\mathbf{E} - \nabla(rE_\theta))}{r}$$

et de manière similaire pour la fonction-test :

$$\mathbf{V} = \frac{(m\varphi - \nabla(r\psi))}{r}$$

L'auteur considère cette idée comme anecdotique, il la met en remarque. Il préfère revenir à la formulation variationnelle de départ et construire des éléments finis qui respecteront de manière naturelle les conditions aux limites. Si on adopte sa démarche, on ne peut pas utiliser des éléments finis classiques, on est amené à devoir modifier les fonctions de base. C'est pour cette raison qu'on préfère utiliser le changement de variable et ainsi garder les éléments finis standard  $H - rot$  et  $H^1$ . Nous obtenons la formulation variationnelle suivante :

$$\left\{ \begin{array}{l} \text{Trouver } (\mathbf{U}, E_\theta) \in H(\text{rot}, \Omega) \times H^1(\Omega) \text{ tel que} \\ -\omega^2 \int_{\Omega} \varepsilon r (r\mathbf{U} + \nabla(rE_\theta)) \cdot (r\mathbf{V} + \nabla(r\psi)) - \omega^2 m^2 \int_{\Omega} \varepsilon r E_\theta \psi \\ + m^2 \int_{\Omega} \frac{r}{\mu} \mathbf{U} \cdot \mathbf{V} + \int_{\Omega} \frac{r}{\mu} \text{rot}(rU) \text{rot}(rV) = 0 \quad \forall (\mathbf{V}, \psi) \in H(\text{rot}, \Omega) \times H^1(\Omega) \end{array} \right.$$

On s'est servi de la relation :

$$\text{rot}(r\mathbf{U}) = m \text{rot} \mathbf{E}$$

Dans cette formulation variationnelle, on a bien ce qu'on recherche, à savoir la symétrie et aucune singularité. De plus, la solution  $\mathbf{U}$  satisfait la relation :

$$\mathbf{U} = \mu \tilde{\mathbf{H}}$$

Cette propriété tend à montrer que le bon choix de variables n'était ni  $E_\theta$ ,  $H_\theta$ , ni  $\mathbf{E}$ ,  $E_\theta$ , mais plutôt  $\mathbf{H}$ ,  $E_\theta$ . Pour conclure sur l'intérêt de cette formulation variationnelle, nous comparons les erreurs  $L^2$  pour les solutions obtenues à partir de la formulation variationnelle en  $E, E_\theta$  (avec des intégrales divergentes) avec les solutions obtenues par la formulation variationnelle en  $U, E_\theta$ . Cette comparaison est faite sur la figure 8.3. On voit que cette formulation variationnelle donne des résultats plus précis, quel que soit l'ordre d'approximation.

## Formulation mixte

Une autre possibilité pour traiter la singularité sur l'axe est d'utiliser une formulation mixte sur tous les éléments proches de l'axe. On reprend le système (8.1) en  $\mathbf{E}, \mathbf{H}, E_\theta, H_\theta$ . On effectue les intégrations par parties sur les équations (8.1) et (8.3), afin de choisir  $\mathbf{H}, H_\theta$  dans  $L^2$ .

$$\left\{ \begin{array}{l} -\omega^2 \int_{\Omega} \varepsilon r \mathbf{E} \cdot \varphi + m \int_{\Omega} \varphi \times \mathbf{H} - \int_{\Omega} r H_\theta \text{rot} \varphi - \int_{\Gamma} r H_\theta \varphi \times \mathbf{n} = 0 \\ -\omega^2 \int_{\Omega} \varepsilon r E_\theta \psi + \int_{\Omega} \mathbf{H} \cdot \text{rot}(r\psi) - \int_{\Gamma} r \psi \mathbf{H} \times \mathbf{n} = 0 \\ -\int_{\Omega} \mu r \mathbf{H} \cdot \phi + m \int_{\Omega} \mathbf{E} \times \phi + \int_{\Omega} \text{rot}(rE_\theta) \cdot \phi = 0 \\ -\int_{\Omega} \mu r H_\theta \lambda - \int_{\Omega} r \lambda \text{rot}(\mathbf{E}) = 0 \end{array} \right.$$

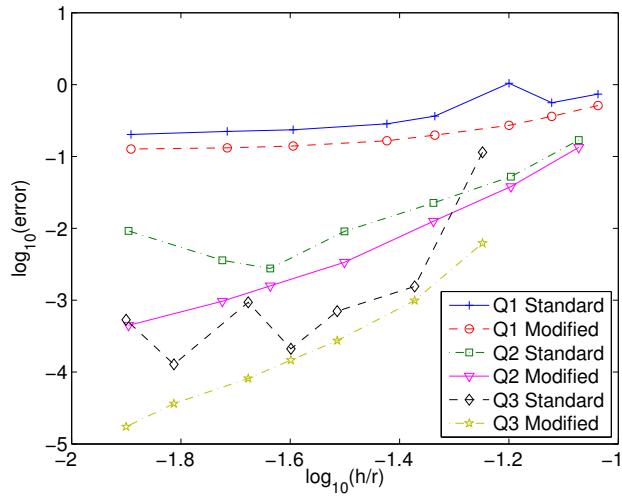


FIG. 8.3 – Evolution de l'erreur  $L^2$  pour la FV  $E, E_\theta$  et la FV  $U, E_\theta$ . Courbes en échelles log-log. Maillages triangles découpés et intégration exacte ( $(r+1)^2$  points de Gauss)

Les espaces d'approximation sont :

$$\begin{aligned}\mathbf{E} \in V_h &= \{\varphi \in H(\text{rot}, \Omega) \text{ tel que } DF_i^t \varphi \circ F_i \in Q_{r-1,r} \times Q_{r,r-1}\} \\ E_\theta \in W_h &= \{\psi \in H^1(\Omega) \text{ tel que } \psi \circ F_i \in Q_{r,r}\} \\ \mathbf{H} \in T_h &= \{\phi \in (L^2(\Omega))^2 \text{ tel que } DF_i^t \varphi \circ F_i \in Q_{r,r} \times Q_{r,r}\} \\ H_\theta \in U_h &= \{\lambda \in L^2(\Omega) \text{ tel que } \lambda \circ F_i \in Q_{r-1,r-1}\}\end{aligned}$$

$V_h$  est la première famille de Nédélec sur les quadrangles, tandis que  $W_h$  est l'espace classique des éléments finis nodaux sur les quadrangles. Sur le même cas test que précédemment, on peut comparer cette méthode à la dernière technique. On voit donc sur la figure 8.4 que cette

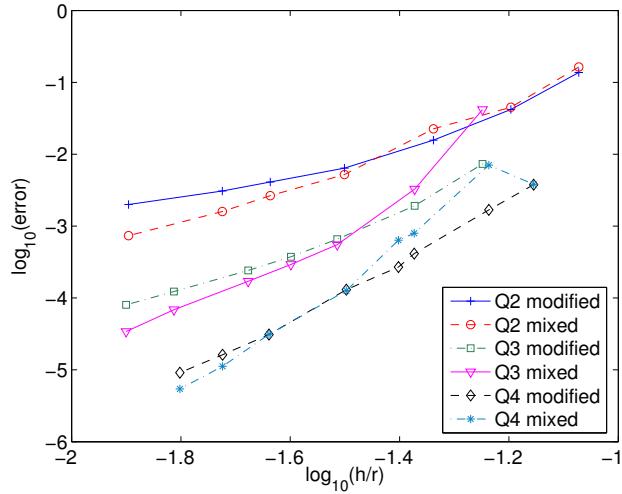


FIG. 8.4 – Evolution de l'erreur  $L^2$  pour la FV mixte  $E, E_\theta$  et la FV  $U, E_\theta$ . Courbes en échelles log-log, maillages triangles découpés et intégration exacte.

méthode se comporte bien. C'est pourquoi on a retenu cette technique par la suite. On a fait

figurer dans la formulation variationnelle les termes de bord ; on explicitera ultérieurement le devenir de ces termes de bord suivant les conditions aux limites choisies.

### 8.1.2 Calcul de la matrice éléments finis

#### Expression des matrices élémentaires

Le choix des espaces d'approximation a été donné précédemment. On rappelle en premier lieu l'expression des fonctions de base pour les 4 espaces d'approximation. Les fonctions de base de  $V_h$  sur le carré unité  $\hat{K}$  sont les suivantes :

$$\hat{\varphi}_{i,j}^1(\hat{x}, \hat{y}) = \hat{\varphi}_i^G(\hat{x}) \hat{\varphi}_j^{GL}(\hat{y}) \mathbf{e_x} \quad 1 \leq i \leq r \quad 1 \leq j \leq r+1$$

$$\hat{\varphi}_{i,j}^2(\hat{x}, \hat{y}) = \hat{\varphi}_i^{GL}(\hat{x}) \hat{\varphi}_j^G(\hat{y}) \mathbf{e_y} \quad 1 \leq i \leq r+1 \quad 1 \leq j \leq r$$

Les fonctions de base pour  $W_h, T_h$  et  $U_h$  :

$$\hat{\psi}_{i,j}(\hat{x}, \hat{y}) = \hat{\varphi}_i^{GL}(\hat{x}) \hat{\varphi}_j^{GL}(\hat{y}) \quad 1 \leq i \leq r+1 \quad 1 \leq j \leq r+1$$

$$\hat{\varphi}_{i,j}^l(\hat{x}, \hat{y}) = \hat{\varphi}_i^{GL}(\hat{x}) \hat{\varphi}_j^{GL}(\hat{y}) \mathbf{e_l} \quad l = 1..2 \quad 1 \leq i \leq r+1 \quad 1 \leq j \leq r+1$$

$$\hat{\lambda}_{i,j}(\hat{x}, \hat{y}) = \hat{\varphi}_i^G(\hat{x}) \hat{\varphi}_j^G(\hat{y}) \quad 1 \leq i \leq r \quad 1 \leq j \leq r$$

où  $\hat{\varphi}_i^G, \hat{\varphi}_i^{GL}$  sont respectivement les fonctions de base associées aux points de Gauss, et aux points de Gauss-Lobatto. On a par définition des espaces d'approximation les définitions suivantes :

$$\varphi(x, y) = DF_i^{-t} \hat{\varphi}(\hat{x}, \hat{y}) \quad \psi(x, y) = \hat{\psi}(\hat{x}, \hat{y})$$

$$\phi(x, y) = DF_i^{-t} \hat{\phi}(\hat{x}, \hat{y}) \quad \lambda(x, y) = \hat{\lambda}(\hat{x}, \hat{y})$$

Après changement de variables les matrices s'écrivent alors :

$$(B_h^1)_{j,k} = -\omega^2 \int_{\hat{K}} \varepsilon r J_i DF_i^{-1} DF_i^{-t} \hat{\varphi}_j \cdot \hat{\varphi}_k$$

$$(B_h^2)_{j,k} = -\omega^2 \int_{\hat{K}} \varepsilon r J_i \hat{\psi}_j \hat{\psi}_k$$

$$(D_h^1)_{j,k} = - \int_{\hat{K}} \mu r J_i DF_i^{-1} DF_i^{-t} \hat{\phi}_j \cdot \hat{\phi}_k$$

$$(D_h^2)_{j,k} = - \int_{\hat{K}} \mu r J_i \hat{\lambda}_j \hat{\lambda}_k$$

$$(C_h)_{j,k} = m \int_{\hat{K}} \hat{\varphi}_j \times \hat{\phi}_k$$

$$(R_h^1)_{j,k} = - \int_{\hat{K}} r \hat{\lambda}_k \hat{\text{rot}} \varphi_j$$

$$(R_h^2)_{j,k} = + \int_{\hat{K}} \hat{\phi}_k \cdot \hat{\text{rot}}(r \hat{\psi}_k)$$

La formulation variationnelle aboutit au système linéaire suivant :

$$\begin{pmatrix} B_h^1 & 0 & C_h & R_h^1 \\ 0 & B_h^2 & R_h^2 & 0 \\ C_h^t & (R_h^2)^t & D_h^1 & 0 \\ (R_h^1)^t & 0 & 0 & D_h^2 \end{pmatrix} \begin{pmatrix} E \\ E_\theta \\ H \\ H_\theta \end{pmatrix} = \begin{pmatrix} F_E \\ F_{E_\theta} \\ F_H \\ F_{H_\theta} \end{pmatrix}$$

On précisera ultérieurement comment on obtient les termes source. Comme les inconnues  $\mathbf{H}$  et  $H_\theta$  sont discontinues d'un élément à un autre, on peut les éliminer à l'aide d'un complément de Schur, et ne garder que les inconnues principales  $\mathbf{E}$  et  $E_\theta$ . On obtient alors le système linéaire suivant :

$$\begin{aligned} [B_h^1 - C_h(D_h^1)^{-1}C_h^t - R_h^1(D_h^2)^{-1}(R_h^1)^t]E \\ - C_h(D_h^1)^{-1}(R_h^2)^t E_\theta &= F_E - C_h(D_h^1)^{-1}F_H - R_h^1(D_h^2)^{-1}F_{H_\theta} \\ [B_h^2 - R_h^2(D_h^1)^{-1}(R_h^2)^t]E_\theta - R_h^2(D_h^1)^{-1}C_h^t E &= F_{E_\theta} - R_h^2(D_h^1)^{-1}F_H \end{aligned}$$

Au lieu de faire l'élimination des inconnues  $H$  et  $H_\theta$  sur le problème continu, on le fait au niveau discret. On peut faire le lien de chaque terme matriciel avec un terme de la formulation standard en  $E, E_\theta$ . Ainsi le terme  $-C_h(D_h^1)^{-1}C_h^t$  correspond au terme suivant de la formulation standard :

$$m^2 \int_{\Omega} \frac{1}{\mu r} \mathbf{E} \cdot \boldsymbol{\varphi} dr dz$$

Le terme  $-R_h^1(D_h^2)^{-1}(R_h^1)^t$  correspond au rot-rot :

$$\int_{\Omega} \frac{r}{\mu} \text{rot}(\mathbf{E}) \text{rot}(\boldsymbol{\varphi}) dr dz$$

Le terme de couplage  $-C_h(D_h^1)^{-1}(R_h^2)^t$  est associé à :

$$-m \int_{\Omega} \frac{1}{\mu r} \mathbf{E} \cdot \nabla(r \psi)$$

On démontrera ultérieurement que si on choisit des points d'intégration adéquats, on obtient une équivalence stricte entre la formulation standard et la formulation mixte.

Les matrices  $D_h^1$  et  $D_h^2$  sont définies négatives, elles sont toujours inversibles. On doit cependant ne pas utiliser les points de Gauss-Lobatto comme points d'intégration pour les éléments proches de l'axe.

### Traitements des conditions aux limites

On s'intéresse dans ce paragraphe à la prise en compte des conditions aux limites de type Dirichlet et de Silver-Müller. On considère une condition de Dirichlet inhomogène :

$$\mathbf{E} \times \mathbf{n} = f \quad E_\theta = g$$

On la traite de manière classique car nos inconnues principales sont  $\mathbf{E}$  et  $E_\theta$ .

La condition absorbante d'ordre 1 souvent appelée condition de Silver-Müller s'écrit :

$$E^{3D} \times n + n \times H^{3D} \times n = 0$$

On décompose le champ électrique sur ses composantes :

$$E^{3D} = E_r \hat{\mathbf{r}} + E_z \hat{\mathbf{z}} + E_\theta \hat{\boldsymbol{\theta}}$$

La normale  $n$  a pour composante :

$$\mathbf{n} = n_r \hat{\mathbf{r}} + n_z \hat{\mathbf{z}}$$

Le produit vectoriel est donc égal à :

$$\mathbf{E}^{3D} \times \mathbf{n} = E_\theta n_z \hat{\mathbf{r}} - E_\theta n_r \hat{\mathbf{z}} - (E_r n_z - E_z n_r) \hat{\boldsymbol{\theta}}$$

On reconnaît la quantité :

$$\mathbf{E} \times \mathbf{n} = (E_r n_z - E_z n_r)$$

De même, le champ magnétique tangentiel est égal à :

$$\mathbf{n} \times \mathbf{H}^{3D} \times \mathbf{n} = n_z \mathbf{H} \times \mathbf{n} \hat{\mathbf{r}} - n_r \mathbf{H} \times \mathbf{n} \hat{\mathbf{z}} + H_\theta \hat{\boldsymbol{\theta}}$$

La condition absorbante d'ordre 1 s'écrit alors :

$$E_\theta + \mathbf{H} \times \mathbf{n} = 0$$

$$-\mathbf{E} \times \mathbf{n} + H_\theta = 0$$

ce qu'on réécrit en exploitant le changement de variables présenté en début de chapitre ;

$$-i\omega E_\theta + \mathbf{H} \times \mathbf{n} = 0$$

$$-i\omega \mathbf{E} \times \mathbf{n} + H_\theta = 0$$

On rappelle que la formulation variationnelle fait apparaître les termes de bord :

$$- \int_{\Gamma} r H_\theta \boldsymbol{\varphi} \times \mathbf{n} ds$$

$$- \int_{\Gamma} r \psi \mathbf{H} \times \mathbf{n} ds$$

Ces termes deviennent donc égaux à :

$$-i\omega \int_{\Gamma} r \mathbf{E} \times \mathbf{n} \boldsymbol{\varphi} \times \mathbf{n} ds$$

$$-i\omega \int_{\Gamma} r E_\theta \psi ds$$

On retrouve des termes analogues à ce qu'on avait pour les équations de Maxwell de 2-D pour  $\mathbf{E}$  et pour l'équation de Helmholtz pour  $E_\theta$ . Le poids d'intégration est différent, on note la présence de  $r$ . Le coefficient  $-i\omega$  est lié à la convention  $-i\omega t$  qu'on a choisie tout au long de l'exposé.

## Expression des termes sources

On considère la décomposition d'une onde plane :

$$E^{\text{incident}} = \sum_{m=-\infty}^{+\infty} E_{r,m}^{\text{inc}} \hat{\mathbf{r}} + E_{z,m}^{\text{inc}} \hat{\mathbf{z}} + E_{\theta,m}^{\text{inc}} \hat{\boldsymbol{\theta}}$$

Par souci de légèreté des notations, on introduit :

$$\mathbf{E}^{\text{inc}} = \begin{vmatrix} E_{r,m}^{\text{inc}} \\ E_{z,m}^{\text{inc}} \end{vmatrix} \quad E_{\theta}^{\text{inc}} = E_{\theta,m}^{\text{inc}}$$

On utilise des notations similaires pour le champ magnétique incident, les termes sources de la formulation variationnelle s'écrivent alors :

$$\begin{aligned} (F_E)_i &= \omega^2 \int_{\Omega} (\varepsilon - \varepsilon_0) r \mathbf{E}^{\text{inc}} \cdot \boldsymbol{\varphi}_i \\ (F_{E_{\theta}})_i &= \omega^2 \int_{\Omega} (\varepsilon - \varepsilon_0) r E_{\theta}^{\text{inc}} \psi_i \\ (F_H)_i &= \int_{\Omega} (\mu - \mu_0) r \mathbf{H}^{\text{inc}} \cdot \boldsymbol{\phi}_i \\ (F_{H_{\theta}})_i &= \int_{\Omega} (\mu - \mu_0) r H_{\theta}^{\text{inc}} \lambda_i \end{aligned}$$

Nous explicitons maintenant la décomposition d'une onde planes en modes. On considère une onde plane classique dans la base cartésienne ( $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$ ) :

$$\mathbf{E}^{\text{inc}} = \begin{vmatrix} E_x^0 \\ E_y^0 \\ E_z^0 \end{vmatrix} e^{i\mathbf{k} \cdot \mathbf{x}}$$

Le champ électrique incident s'écrit dans la base cylindrique ( $\mathbf{e}_r, \mathbf{e}_{\theta}, \mathbf{e}_z$ )

$$\mathbf{E}^{\text{inc}} = \begin{vmatrix} \cos \theta E_x^0 + \sin \theta E_y^0 \\ \cos \theta E_y^0 - \sin \theta E_x^0 \\ E_z^0 \end{vmatrix} e^{i\mathbf{k} \cdot \mathbf{x}}$$

On veut décomposer ce champ sous la forme

$$E^{\text{inc}} = \sum_{m=-\infty}^{+\infty} \begin{vmatrix} E_{r,m} \\ E_{\theta,m} \\ E_{z,m} \end{vmatrix} e^{-im\theta}$$

Du fait de l'orthogonalité des modes,  $E_{r,m}$  est égal à

$$E_{r,m} = \frac{1}{2\pi} \int_0^{2\pi} E_r^{\text{inc}} e^{im\theta}$$

$$\mathbf{k} \cdot \mathbf{x} = k_x r \cos \theta + k_y r \sin \theta + k_z z$$

On introduit un angle d'incidence  $\theta_0$  de l'onde plane, tel que :

$$k_x \cos \theta + k_y \sin \theta = k_{\perp} \cos(\theta - \theta_0)$$

$k_{\perp}$  et  $\theta_0$  sont définis par les relations

$$k_{\perp} = \sqrt{k_x^2 + k_y^2} \quad \cos \theta_0 = \frac{k_x}{\sqrt{k_x^2 + k_y^2}}$$

On a donc

$$E_{r,m} = e^{ik_z z} \frac{1}{2\pi} \int_0^{2\pi} (\cos \theta E_x^0 + \sin \theta E_y^0) e^{ik_{\perp} r \cos(\theta - \theta_0)} e^{im\theta} d\theta$$

On utilise le développement de Jacobi-Anger :

$$e^{ik_{\perp} r \cos(\theta - \theta_0)} = \sum_{n=-\infty}^{+\infty} i^n J_n(k_{\perp} r) e^{in(\theta - \theta_0)} = J_0(k_{\perp} r) + 2 \sum_{n=1}^{+\infty} i^n J_n(k_{\perp} r) \cos n(\theta - \theta_0)$$

On obtient finalement l'expression suivante :

$$E_{r,m} = \frac{1}{2} e^{ik_z z} \left[ i^{m-1} J_{m-1}(k_{\perp} r) e^{i(m-1)\theta_0} (E_x^0 + iE_y^0) + i^{m+1} J_{m+1}(k_{\perp} r) e^{i(m+1)\theta_0} (E_x^0 - iE_y^0) \right]$$

Pour obtenir  $E_{\theta,m}$ , il suffit de remplacer  $E_x^0$  par  $E_y^0$  et  $E_y^0$  par  $-E_x^0$ .

$$E_{\theta,m} = \frac{1}{2} e^{ik_z z} \left[ i^{m-1} J_{m-1}(k_{\perp} r) e^{i(m-1)\theta_0} (E_y^0 - iE_x^0) + i^{m+1} J_{m+1}(k_{\perp} r) e^{i(m+1)\theta_0} (E_y^0 + iE_x^0) \right]$$

$E_{z,m}$  est égal à

$$E_{z,m} = i^m J_m(k_{\perp} r) E_z^0 e^{ik_z z} e^{im\theta_0}$$

On notera que pour une incidence axiale ( $k_{\perp} = 0$ ), l'onde plane se décompose sur les deux modes -1 et +1 :

$$E_{r,-1} = \frac{1}{2} (E_x^0 - iE_y^0)$$

$$E_{r,+1} = \frac{1}{2} (E_x^0 + iE_y^0)$$

$$E_{\theta,-1} = \frac{1}{2} (E_y^0 + iE_x^0)$$

$$E_{\theta,+1} = \frac{1}{2} (E_y^0 - iE_x^0)$$

$$E_{z,-1} = E_{z,+1} = 0$$

### Equivalence formulation mixte - formulation standard

Sur les éléments proches de l'axe, il est nécessaire d'utiliser une intégration avec des points de Gauss ; en pratique on prend  $(r+1)^2$  points de Gauss. Sur les éléments qui ne touchent pas l'axe, rien ne nous empêche de prendre une intégration approchée. Le but recherché est d'obtenir un calcul rapide de la matrice éléments finis, par exemple une complexité en  $O(r^4)$  où  $r$  est l'ordre d'approximation. Une intégration exacte nous fournit un calcul relativement lent, de complexité  $O(r^6)$  rendant rédhibitoire l'utilisation de  $Q_8$ .

On choisit donc des points d'intégration de Gauss-Lobatto pour évaluer les matrices de masse  $D_h^1, D_h^2, B_h^1, B_h^3$ , la matrice de rigidité de l'inconnue scalaire  $R_h^2$  et la matrice de couplage  $C_h$ .

On choisit les points de Gauss pour évaluer la matrice de rigidité vectorielle  $R_h^1$ . Comme l'on a vu au chapitre 7, ce choix donne une erreur de dispersion optimale sur les équations de Maxwell 2-D. De plus, ce choix aboutit à une équivalence entre la formulation mixte et la formulation standard, si on utilise les points de Gauss-Lobatto pour toutes les matrices élémentaires excepté la matrice de rigidité vectorielle où on prend les points de Gauss. On s'attache dans cette sous-sous-section à démontrer de manière relativement simple cette équivalence.

On commence par le terme de masse vectoriel de la formulation standard :

$$\int_{\hat{K}} \left( -\omega^2 \varepsilon r + \frac{m^2}{\mu r} \right) J_i DF_i^{-1} DF_i^{-t} \hat{\varphi}_j \cdot \hat{\varphi}_k \mathbf{e}_t \times \mathbf{e}_s$$

On veut prouver que ce terme est égal à  $B_h^1 - C_h (D_h^1)^{-1} C_h^t$  si on utilise les points d'intégration de Gauss-Lobatto. La partie  $B_h^1$  est identique sur les deux formulations mixte et standard. Intéressons nous à la partie  $-C_h (D_h^1)^{-1} C_h^t$ . Le terme générique de  $C_h^t$  est égal à :

$$(C_h^t)_{(j,s),(k,t)} = m \omega_j^{GL} \hat{\varphi}_k(\hat{\xi}_j^{GL}) \mathbf{e}_t \times \mathbf{e}_s$$

La matrice de masse  $D_h^1$  est diagonale par blocs :

$$(D_h^1)_{(j,s),(k,t)}^{-1} = -\frac{1}{\omega_j^{GL} \mu r J_i} DF_i^t DF_i \mathbf{e}_t \cdot \mathbf{e}_s \delta_{j,k}$$

Le produit s'écrit alors :

$$-(C_h (D_h^1)^{-1} C_h^t)_{(j,s),(k,t)} = m^2 \sum_{n,p,q} \frac{\omega_n^{GL}}{(\mu r J_i)(\hat{\xi}_n^{GL})} \hat{\varphi}_j(\xi_n^{GL}) (DF_i^t DF_i)(\hat{\xi}_n)_{p,q} \mathbf{e}_s \times \mathbf{e}_p \mathbf{e}_t \times \mathbf{e}_q \hat{\varphi}_k(\xi_n^{GL})$$

Ce qu'on peut réécrire sous la forme

$$(-C_h (D_h^1)^{-1} C_h^t)_{(j,s),(k,t)} = m^2 \sum_n \frac{\omega_n^{GL}}{(\mu r)(\hat{\xi}_n^{GL})} \hat{\varphi}_j(\xi_n^{GL}) \hat{\varphi}_k(\xi_n^{GL}) A(\xi_n^{GL})_{s,t}$$

La matrice intermédiaire  $A$  est égale à :

$$A_{s,t} = \sum_{p,q} \frac{1}{J_i} (DF_i^t DF_i) \mathbf{e}_s \times \mathbf{e}_p \mathbf{e}_t \times \mathbf{e}_q$$

En faisant cette double sommation sur  $p$  et  $q$ , on calcule en pratique la comatrice de  $DF_i^t DF_i$ , qui est donc égale à  $\det(DF_i^t DF_i)(DF_i^t DF_i)^{-1} = J_i^2 DF_i^{-1} DF_i^{-t}$ . Au final, on trouve donc :

$$-(C_h (D_h^1)^{-1} C_h^t)_{(j,s),(k,t)} = m^2 \sum_n \frac{\omega_n^{GL}}{(\mu r)(\hat{\xi}_n^{GL})} \hat{\varphi}_j(\xi_n^{GL}) \hat{\varphi}_k(\xi_n^{GL}) J_i DF_i^{-1} DF_i^{-t} \mathbf{e}_s \cdot \mathbf{e}_t$$

On a bien l'équivalence annoncée sur ce terme. On remarquera que le calcul de ce terme est de complexité  $O(r^4)$ , car ce terme est une matrice de masse qu'on a rencontrée dans le chapitre 5.

On regarde maintenant le terme de rigidité :

$$\int_{\hat{K}} \frac{r}{\mu J_i} \text{rot}(\hat{\varphi}_j \mathbf{e}_s) \text{rot}(\hat{\varphi}_j \mathbf{e}_t)$$

On veut prouver qu'il est égal au terme :

$$- R_h^1 (D_h^2)^{-1} (R_h^1)^t$$

On utilise les points de Gauss pour intégrer toutes ces matrices.  $D_h^2$  est diagonale :

$$(D_h^2)_{j,k}^{-1} = -\frac{1}{\mu r J_i \omega_k^G} \delta_{j,k}$$

$R_h^1$  est d'expression relativement simple :

$$(R_h^1)_{j,k} = -\omega_k^G \operatorname{rot}(\hat{\varphi}_j \mathbf{e}_s)(\hat{\xi}_k^G) \delta_{j,k}$$

Lorsqu'on multiplie, on fait apparaître la sommation sur les points de quadrature :

$$-(R_h^1 (D_h^2)^{-1} (R_h^1)^t)_{j,k} = \sum_n \frac{\omega_n^G}{(\mu r J_i)(\hat{\xi}_n^G)} \operatorname{rot}(\hat{\varphi}_j \mathbf{e}_s)(\hat{\xi}_n^G) \operatorname{rot}(\hat{\varphi}_j \mathbf{e}_s)(\hat{\xi}_n^G)$$

On retrouve bien l'expression annoncée, avec une sommation sur les points d'intégration de Gauss. Là aussi, un terme analogue a déjà été analysé au chapitre 5, on a une complexité en  $O(r^4)$  sur ce terme.

On laisse le soin au lecteur de montrer l'équivalence pour les autres termes de la formulation variationnelle. L'équivalence pour la matrice de rigidité scalaire a été signalée au chapitre 1. La complexité est bien en  $O(r^4)$ . Pour le terme de couplage, la démonstration est un peu plus ardue. Le terme de couplage de la formulation standard s'écrit :

$$m \int_K \frac{J_i}{\mu r} D F_i^{-t} \hat{\varphi}_i \cdot \nabla(r \psi_j)$$

On utilise des points de Gauss-Lobatto pour intégrer cette matrice. On aura donc un indice de sommation en moins car les fonctions de base  $\hat{\varphi}$  ont un facteur  $\hat{\varphi}_{j_2}^{GL}$  sur une variable. On aura également un autre indice de sommation en moins car le gradient des fonctions  $\psi$  ont également un facteur  $\hat{\varphi}_{j_2}^{GL}$  sur une variable. La complexité du calcul de la matrice élémentaire de couplage est en  $O(r^4)$ .

## 8.2 Précision de la méthode

### 8.2.1 Cas de la sphère parfaitement conductrice

On étudie la précision de la méthode sur le cas académique de la sphère parfaitement conductrice (voir figure 5.6), avec une condition de Silver-Müller sur la frontière extérieure. On obtient les résultats de la figure 8.5 pour des maillages réguliers. On mesure des pentes de 1.04, 2.02, 2.99, 4.07, 5.10 pour respectivement  $Q_1$ ,  $Q_2$ ,  $Q_3$ ,  $Q_4$  et  $Q_5$ . On a, semble-t-il, une convergence de la méthode en  $O(h^r)$ . Sur des maillages “triangles découpés”, on retrouvera vraisemblablement au mieux une convergence en  $O(h^{r-1})$  du fait de l'utilisation de la première famille de Nédélec sur les quadrangles, qui donnait cet ordre de convergence sur les équations de Maxwell 2-D (cf. chapitre 5).

### 8.2.2 Cas du cone-sphère

On s'intéresse maintenant au cas intermédiaire du cône-sphère, dont la géométrie possède une singularité uniquement sur un point de l'axe. La géométrie est donc “faiblement singulière” puisque la surface 3-D est singulière uniquement en un point. La solution de ce problème est affichée sur la figure 8.6. On calcule une solution de référence avec du  $Q_8$  sur un maillage d'un million de ddl. On obtient les courbes de convergence de la figure 8.7 pour des maillages réguliers. On utilise des maillages quadrangulaires non-structurés (cf. figure 8.8), qui ne sont pas obte-

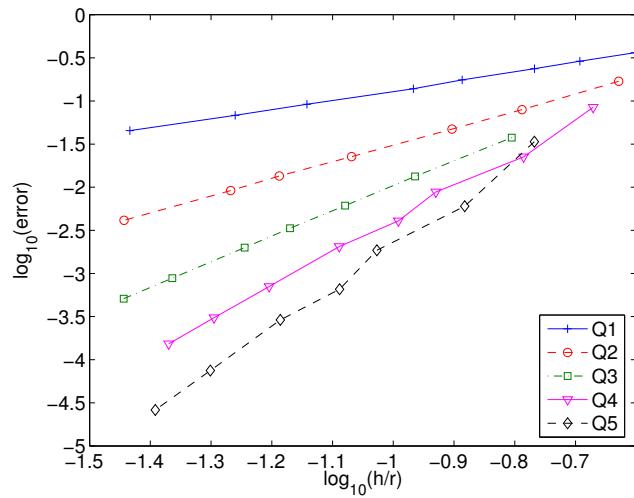


FIG. 8.5 – Evolution de l'erreur H-rot entre la solution numérique et la solution analytique en échelle log-log. Cas de la sphère parfaitement conductrice sur des maillages réguliers.

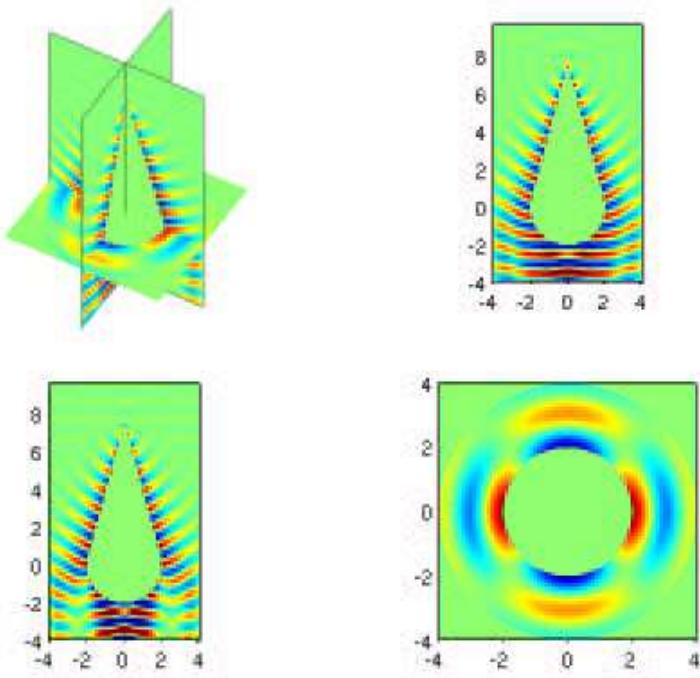


FIG. 8.6 – Partie réelle du champ diffracté par une onde plane pour un cone-sphère parfaitement conducteur. L'onde plane est axiale, elle vient par le haut.

nus en découplant des triangles. Ces maillages sont obtenus en recombinant des triangles entre eux pour obtenir des quadrangles. Le maillage hybride quadrangles/triangles est redécoupé afin de n'obtenir que des quadrangles. Ces maillages sont généralement plus sympathiques que les triangles découpés, on a besoin de moins de degrés de liberté pour obtenir la même précision. Toutefois, comme le montre la figure 8.5, il est difficile de mesurer un ordre de convergence.

Nous préférons donc faire une étude de convergence en subdivisant le maillage 8.8. On obtient le tableau 8.1.  $h = 1.0$  correspond au maillage initial,  $h = 0.5$  correspond à ce maillage deux fois plus fin, etc ... La première information que donne ce tableau est que cette faible singularité ne perturbe pas l'ordre de convergence de la méthode, lorsqu'on évalue l'erreur sur un domaine excluant la singularité. La seconde information est que lorsqu'on utilise des maillages

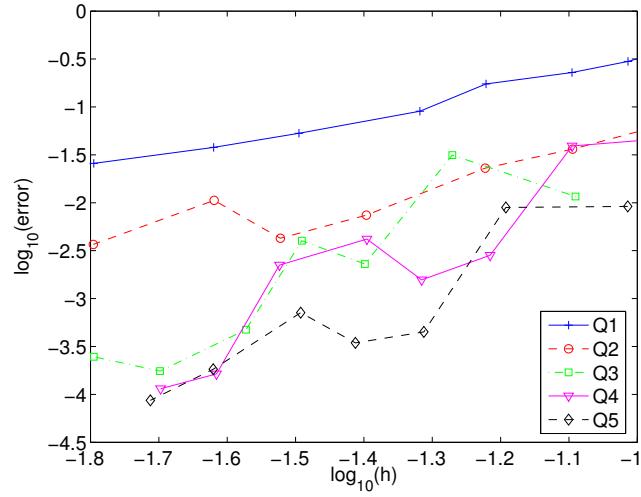


FIG. 8.7 – Evolution de l'erreur H-rot entre la solution numérique et la solution de référence en échelle log-log. Cas du cone-sphère parfaitement conducteur sur des maillages non-structurés. L'erreur est calculée en omettant la région  $r < 0.5$  qui contient la singularité.

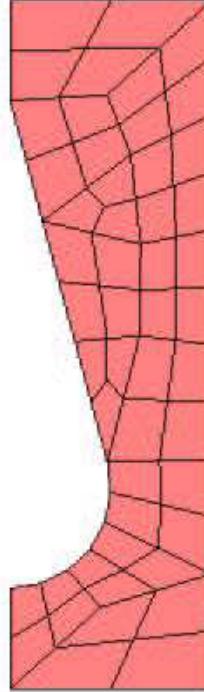


FIG. 8.8 – Maillage utilisé pour mesurer l'ordre de convergence

	$h = 1.0$	$h = 0.5$	$h = 0.25$	$h = 0.125$	$h = 0.0625$	$h = 0.03125$	$h = 0.015625$
Erreur avec $Q_1$	-	-	1.08	0.46	0.14	0.051	0.023
Ordre de convergence	-	-	-	1.23	1.74	1.42	1.15
Erreur avec $Q_2$	-	1.14	0.074	0.038	4.6e-3	9.54e-4	-
Ordre de convergence	-	-	3.94	0.96	3.04	2.27	-
Erreur avec $Q_4$	0.34	0.027	0.0024	7.99e-4	2.53e-5	-	-
Ordre de convergence	-	3.65	3.47	1.6	4.98	-	-

TAB. 8.1 – Erreur H-rot entre la solution numérique et la solution de référence pour différents pas de maillage. Cas d'un maillage dont chaque segment est subdivisé en  $2^k$  sous-intervalles.

quadrangulaires obtenus par subdivision, il semble qu'on obtienne un ordre de convergence en  $O(h^r)$ . Cette observation peut se comprendre car la subdivision d'un maillage va faire tendre les matrices jacobiniennes vers des matrices jacobiniennes diagonales. Les termes extra-diagonaux de  $DF_i$  vont tendre vers 0 lorsque le pas de maillage tend vers 0. Cette convergence est "globale", localement les quadrangles qui sont à la jonction des sous-domaines ne vérifient pas cette propriété. En subdivisant un maillage initial, on va se rapprocher du comportement d'un maillage régulier.

### 8.2.3 Cas du cylindre

On traite maintenant un cas avec une singularité plus étendue que le cas précédent. On considère la diffraction d'une onde plane par un cylindre parfaitement conducteur (cf. figure 8.9). On peut cette fois, utiliser des maillages parfaitement réguliers, et obtenir les courbes de conver-

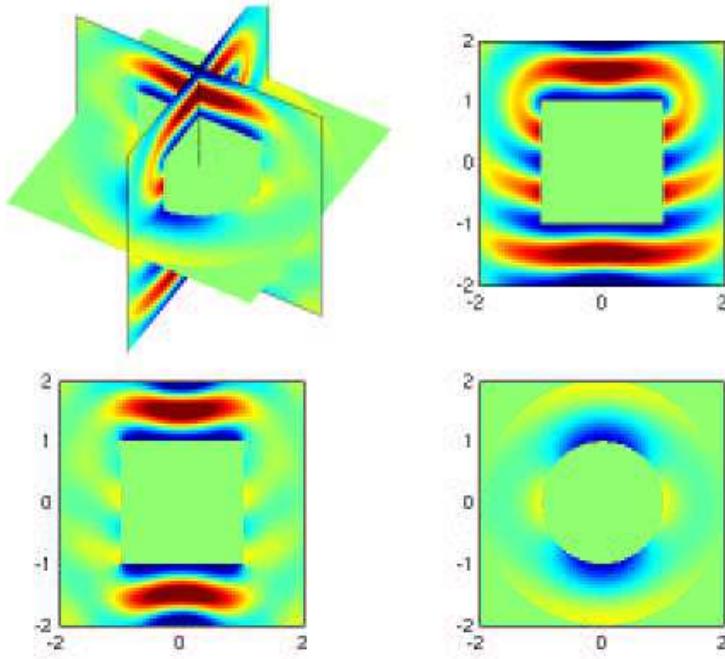


FIG. 8.9 – Partie réelle du champ diffracté par un cylindre parfaitement conducteur

gence sympathiques de la figure 8.10. La solution de référence est calculée sur un maillage  $Q_8$  d'un million de ddl avec un raffinement local sur les deux coins du maillage. On retrouve ce qu'on a déjà signalé pour l'équation de Helmholtz 2-D. On a une convergence en  $O(h^{1.33})$  quel que soit l'ordre d'approximation, mais les constantes sont de plus en plus faibles lorsqu'on monte

en ordre. A précision fixée, on a besoin de moins de degrés de liberté en  $Q_8$  qu'en  $Q_2$ , si on utilise un maillage régulier uniforme. Pour capter le bon ordre de convergence, on fait une étude sur le tableau 8.2, en subdivisant un maillage initial. Ce tableau est instructif, car sur les dia-

$h$	1.0	0.5	0.25	0.125	0.0625	0.03125	0.015625	0.0078125
Erreur avec $Q_1$	-	-	0.67	0.215	0.0918	0.0428	0.0211	0.0109
Ordre de convergence	-	-	-	1.63	1.22	1.10	1.02	0.95
Erreur avec $Q_2$	-	0.65	0.0708	0.0176	4.99e-3	1.55e-3	5.23e-4	-
Ordre de convergence	-	-	3.20	2.01	1.82	1.69	1.57	-
Erreur avec $Q_4$	0.217	0.0156	3.67e-3	1.36e-3	5.09e-4	2.0e-4	-	-
Ordre de convergence	-	3.8	2.09	1.43	1.42	1.34	-	-

TAB. 8.2 – Erreur H-rot entre la solution numérique et la solution de référence pour différents pas de maillage. Cas d'un maillage dont chaque segment est subdivisé en  $2^k$  sous-intervalles.

gonales, on a l'erreur commise pour un nombre de degrés de liberté constant, pour  $Q_1$ ,  $Q_2$  et  $Q_4$ . Ainsi, lorsque l'on met huit points par longueur d'onde ( $h = 0.125$  pour  $Q_1$ ), on obtient une erreur de 21 % en  $Q_1$ , de 7% en  $Q_2$  et de 1.5 % en  $Q_4$ . Sur toutes les diagonales, on a bien une décroissance de l'erreur.

### 8.3 Conclusion

Dans ce chapitre, nous avons présenté une méthode de discréétisation des équations de Maxwell sur des domaines axi-symétriques. Cette méthode de discréétisation utilise des éléments finis quadrilatéraux H-rot pour  $\mathbf{E}$  et  $H^1$  pour  $E_\theta$ . Afin d'éviter la présence d'intégrales avec un poids en  $1/r$ , nous proposons une formulation mixte. Nous avons validé cette approche sur le cas de la sphère, du cône-sphère et du cylindre.

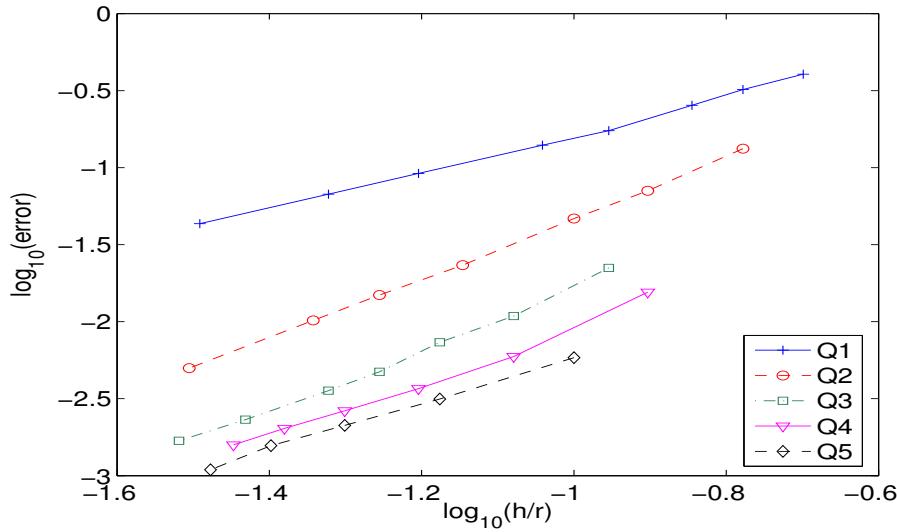


FIG. 8.10 – Evolution de l'erreur H-rot entre la solution numérique et la solution de référence en échelle log-log. En abscisse , on fait figurer  $h/r$  afin de pouvoir comparer les différents ordres d'approximation. Cas du cylindre parfaitement conducteur sur des maillages réguliers. On calcule l'erreur sur un domaine excluant tout le pourtour de l'objet.

## Chapitre 9

# Équations intégrales d'ordre élevé pour les équations de Maxwell sur des domaines à symétrie de révolution

*Nous montrons dans ce chapitre comment on établit une formulation intégrale lorsque le domaine de calcul présente une symétrie de révolution. Comme dans le cas volumique, on aboutit à une succession de problèmes 2-D indépendants. La principale difficulté réside dans le calcul des intégrales singulières. Nous proposons plusieurs approches concurrentielles pour lever cette difficulté. Finalement, nous discutons du couplage avec les éléments finis, et l'intérêt de monter en ordre.*

### Sommaire

---

<b>9.1</b>	<b>Obtention de la formulation variationnelle</b>	<b>204</b>
9.1.1	Notations	204
9.1.2	Formulation EFIE	205
9.1.3	Formulation MFIE	209
9.1.4	Formulation CFIE	210
<b>9.2</b>	<b>Méthode d'intégration</b>	<b>211</b>
9.2.1	Méthode d'intégration pour la partie régulière	211
9.2.2	Règles simples d'intégration dans le cas régulier	213
9.2.3	Calcul des intégrales singulières	215
<b>9.3</b>	<b>Précision de la méthode</b>	<b>221</b>
9.3.1	Cas de la sphère parfaitement conductrice	221
9.3.2	Cas du cylindre parfaitement conducteur	221
9.3.3	Cas du cône-sphère parfaitement conducteur	222
<b>9.4</b>	<b>Couplage avec les éléments finis</b>	<b>226</b>
9.4.1	Définition des opérateurs	226
9.4.2	Cas de la sphère revêtue par un matériau diélectrique	227
9.4.3	Cas du cylindre revêtu par du diélectrique	228
9.4.4	Cas du cone-sphère revêtu par du diélectrique	229
<b>9.5</b>	<b>Conclusion</b>	<b>230</b>

---

## 9.1 Obtention de la formulation variationnelle

Notre principale source d'inspiration est [Volpert et Levadoux, 2001]. Dans ce rapport, l'auteur présente une discréétisation d'ordre 1 pour une formulation intégrale couplée avec des éléments finis d'ordre 1. Notre apport est d'étendre cette approche à l'ordre élevé, en utilisant des formulations variationnelles légèrement différentes. L'obtention de la formulation s'effectue en considérant une formulation variationnelle 3-D, et en injectant des fonctions de base spécifiques. On obtient alors une suite de formulations variationnelles 2-D indépendantes. Les techniques d'intégration de singularités sont néanmoins spécifiques à l'axi-symétrique. Elles sont différentes de ce qu'on peut rencontrer en 2-D ou en 3-D, bien qu'on puisse faire quelques similitudes avec le 3-D.

### 9.1.1 Notations

On considère une courbe du plan  $\Gamma$  paramétrée par l'abscisse curviligne  $s$ . Un point  $M$  appartenant à  $\Gamma$  a pour coordonnées :

$$M(s) = (r(s), z(s))$$

La surface générée par révolution sera notée  $\Sigma$ . Un point de cette surface a pour coordonnées :

$$M(s, \theta) = (r \cos \theta, r \sin \theta, z)$$

L'élément surfacique vaut alors

$$d\sigma = r ds d\theta$$

On note  $\mathbf{t}$ , le vecteur tangent unitaire à  $\Gamma$  :

$$\mathbf{t} = (t_x, t_z) = \frac{1}{\sqrt{(\frac{\partial r}{\partial s})^2 + (\frac{\partial z}{\partial s})^2}} \left( \frac{\partial r}{\partial s}, \frac{\partial z}{\partial s} \right)$$

Par la suite, afin de simplifier les calculs, nous supposons que  $s$  est un paramétrage de la courbe qui vérifie :

$$\sqrt{(\frac{\partial r}{\partial s})^2 + (\frac{\partial z}{\partial s})^2} = 1$$

$s$  correspond dans ce cas à ce qu'on appelle l'abscisse curviligne. On introduit la base ortho-normée directe  $(\mathbf{t}, \mathbf{n}, \mathbf{b})$  telle que

$$\mathbf{t} = \begin{vmatrix} t_x \cos \theta \\ t_x \sin \theta \\ t_z \end{vmatrix} \quad \mathbf{n} = \begin{vmatrix} t_z \cos \theta \\ t_z \sin \theta \\ -t_x \end{vmatrix} \quad \mathbf{b} = \begin{vmatrix} -\sin \theta \\ \cos \theta \\ 0 \end{vmatrix}$$

Les vecteurs  $(\mathbf{t}, \mathbf{b})$  sont tangents à la surface  $\Sigma$ .  $\mathbf{n}$  est la normale unitaire, extérieure à  $\Sigma$ , si on suppose que la frontière  $\Gamma$  est parcourue dans le sens direct.

On considère un second point :

$$M' = (r' \cos \theta', r' \sin \theta', z')$$

La distance entre le point  $M$  et le point  $M'$ , vaut :

$$R^2 = ||\mathbf{M}\mathbf{M}'||^2 = (r' \cos \theta' - r \cos \theta)^2 + (r' \sin \theta' - r \sin \theta)^2 + (z' - z)^2$$

$$\begin{aligned} R^2 &= r^2 + r'^2 - 2r r' \cos(\theta - \theta') + (z - z')^2 \\ R &= \sqrt{(r' - r)^2 + (z' - z)^2 + 2rr' (1 - \cos \varphi)} \end{aligned}$$

où on a introduit l'angle  $\varphi$  :

$$\varphi = \theta - \theta'$$

Le vecteur  $\mathbf{MM}'$  a pour composantes dans le système  $\mathbf{t}', \mathbf{n}', \mathbf{b}'$  :

$$\begin{aligned} MM'_{t'} &= r' t'_x - r t'_x \cos \varphi + (z' - z) t'_z \\ MM'_{b'} &= -r \sin \varphi \\ MM'_{n'} &= r' t'_z - r t'_z \cos \varphi - (z' - z) t'_x \end{aligned}$$

Les produits scalaires entre le système  $\mathbf{t}', \mathbf{n}', \mathbf{b}'$  et le système  $\mathbf{t}, \mathbf{n}, \mathbf{b}$  valent :

$$\begin{aligned} \mathbf{t}' \cdot \mathbf{t} &= t'_x t_x \cos \varphi + t'_z t_z & \mathbf{t}' \cdot \mathbf{b} &= -t'_x \sin \varphi & \mathbf{t}' \cdot \mathbf{n} &= -t'_z t_x + t'_x t_z \cos \varphi \\ \mathbf{b}' \cdot \mathbf{t} &= t_x \sin \varphi & \mathbf{b}' \cdot \mathbf{b} &= \cos \varphi & \mathbf{b}' \cdot \mathbf{n} &= t_z \sin \varphi \\ \mathbf{n}' \cdot \mathbf{t} &= -t'_x t_z + t'_z t_x \cos \varphi & \mathbf{n}' \cdot \mathbf{b} &= -t'_z \sin \varphi & \mathbf{n}' \cdot \mathbf{n} &= t'_z t_z \cos \varphi + t'_x t_x \end{aligned}$$

### 9.1.2 Formulation EFIE

On note les courants, en gardant  $\mathbf{n}$  normale extérieure à  $\Sigma$

$$\mathbf{J} = \mathbf{n} \times \mathbf{H} \quad \mathbf{K} = \mathbf{E} \times \mathbf{n}$$

On suppose que l'objet est parfaitement conducteur, on a alors  $\mathbf{K} = 0$ . On ne garde que l'inconnue  $\mathbf{J}$ , la formulation variationnelle 3-D de l'EFIE s'écrit :

$$-i \int_{\Sigma} \int_{\Sigma} k G(x, y) \left( \mathbf{J}(y) \cdot \bar{\mathbf{J}}^t(x) - \frac{1}{k^2} \operatorname{div}_{\sigma}(\mathbf{J}(y)) \operatorname{div}_{\sigma}(\bar{\mathbf{J}}^t(x)) \right) d\sigma(y) d\sigma(x) = \int_{\Sigma} \mathbf{E}^{\text{inc}}(x) \cdot \bar{\mathbf{J}}^t(x) d\sigma(x)$$

$\mathbf{J}^t$  est la fonction test, elle est prise dans le même espace fonctionnel que  $\mathbf{J}$  : l'espace  $H(\operatorname{div}, \Gamma)$ . Souvent, cette formulation variationnelle est écrite en omettant le conjugué sur la fonction test ( $\mathbf{J}^t(x)$  au lieu de  $\bar{\mathbf{J}}^t(x)$ ). En effet, les deux notations sont équivalentes lorsqu'on prend des fonctions de base réelles, ce qui est le cas en 3-D. Dans notre cas, on prend des fonctions de base complexes, on doit donc ne pas omettre le conjugué.  $\mathbf{J}$  est tangentiel à la surface  $\Sigma$ . C'est pour cette raison qu'on choisit d'exprimer ses projections sur les deux vecteurs tangentiels  $\mathbf{t}$  et  $\mathbf{b}$ .

On ne discrétise ainsi que deux inconnues scalaires  $J_t$  et  $J_b$ .

On choisit des fonctions de base de la forme :

$$\begin{aligned} J_t(s, \theta) &= \varphi(s) e^{-im\theta} \\ J_b(s, \theta) &= \psi(s) e^{-im\theta} \end{aligned}$$

La condition  $\mathbf{J} \in H(\operatorname{div}, \Sigma)$  se réinterprète comme :

$$\varphi \in H^1(\Gamma) \quad \psi \in L^2(\Gamma)$$

Un choix possible de fonctions de base est de prendre la trace des fonctions de base volumiques issues de la formulation  $H - rot$  :

$$\hat{\varphi}_i(\hat{s}) = \hat{\varphi}_i^{GL}(\hat{s}) \quad i = 1..(r+1)$$

$$\hat{\varphi}_i(\hat{s}) = \hat{\varphi}_i^G(\hat{s}) \quad i = 1..r$$

Une autre choix est de prendre :

$$\varphi \in H^1(\Gamma) \quad \psi \in H^1(\Gamma)$$

On choisit alors les mêmes fonctions de base

$$\hat{\varphi}_i(\hat{s}) = \hat{\psi}_i(\hat{s}) = \hat{\varphi}_i^{GL}(\hat{s})$$

Dans la suite, nous garderons la notation avec deux fonctions de base différentes. Au niveau des résultats numériques, on a fait le choix de prendre les mêmes fonctions de base. Les deux inconnues  $J_t$  et  $J_b$  sont continues et appartiennent à  $H^1$ . Nous verrons ultérieurement en quoi ce choix est judicieux pour réaliser le couplage avec les éléments finis.

On va maintenant remarquer qu'on aboutit à des problèmes indépendants pour chaque mode  $m$ . En effet supposons qu'on ait décomposé  $\mathbf{J}$  sous la forme :

$$\mathbf{J}(x') = \sum_{m'=-\infty}^{+\infty} \mathbf{J}_{m'}(s') e^{-im'\theta'}$$

On choisit comme fonctions tests

$$\mathbf{J}^t(x) = \mathbf{J}_m^t(s) e^{-im\theta}$$

On note :

$$a(\mathbf{u}(x'), \mathbf{v}(x)) = -ik G(x, x') [\mathbf{u}(x') \cdot \bar{\mathbf{v}}(x) - \frac{1}{k^2} \operatorname{div}_\sigma \mathbf{u}(x') \operatorname{div}_\sigma \bar{\mathbf{v}}(x)]$$

on constate alors que

$$a(\mathbf{J}(x'), \mathbf{J}^t(x)) = \sum_{m'=-\infty}^{+\infty} h_{m'}(s, s', \varphi) e^{im\varphi} e^{-i(m'-m)\theta}$$

On a séparé la dépendance en  $\varphi$  et la dépendance en  $\theta$ . On a utilisé le fait que le noyau de Green ne dépend que de la distance  $R$ , dont l'expression ne fait intervenir que  $\varphi$ . De même, les produits scalaires  $\mathbf{t}' \cdot \mathbf{t}$ ,  $\mathbf{t}' \cdot \mathbf{b}$ ... ne dépendent que de  $\varphi$ . Lorsqu'on effectue la double intégration, on a :

$$\int_{\Sigma} \int_{\Sigma} a(\mathbf{J}(x'), \mathbf{J}^t(x)) dx' dx = \int_0^{2\pi} \int_0^{2\pi} \int_{\Gamma} \int_{\Gamma} a(\mathbf{J}(x'), \mathbf{J}^t(x)) r r' ds ds' d\theta' d\theta$$

On fait le changement de variables :

$$\varphi = \theta - \theta'$$

On a alors :

$$\int_{\Sigma} \int_{\Sigma} a(\mathbf{J}(x'), \mathbf{J}^t(x)) dx' dx = \int_0^{2\pi} \int_{\theta}^{\theta+2\pi} \int_{\Gamma} \int_{\Gamma} a(\mathbf{J}(x'), \mathbf{J}^t(x)) r r' ds ds' d\varphi d\theta$$

Par  $2\pi$ -périodicité de  $h_{m'}$  par rapport à  $\varphi$ , on a :

$$\int_{\Sigma} \int_{\Sigma} a(\mathbf{J}(x'), \mathbf{J}^t(x)) dx' dx = \sum_{m'=-\infty}^{+\infty} \int_0^{2\pi} \left( \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} h_{m'}(s, s', \varphi) e^{im\varphi} r r' ds ds' d\varphi \right) e^{-i(m'-m)\theta} d\theta$$

Si on choisit  $m \neq m'$ , on aura une intégrale en  $\theta$  nulle. On en déduit que :

$$\int_{\Sigma} \int_{\Sigma} a(\mathbf{J}_{m'}(x') e^{-im'\theta}, \mathbf{J}_{\mathbf{m}}^{\mathbf{t}}(x) e^{-im\theta}) dx dx' = 0 \quad \forall m' \neq m$$

On a bien des problèmes indépendants.

Dans la suite, on divisera la formulation variationnelle par  $2\pi$  pour avoir :

$$\frac{1}{2\pi} \int_{\Sigma} \int_{\Sigma} a(\mathbf{J}(x'), \mathbf{J}^{\mathbf{t}}(x)) dx' dx = \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} h_m(s, s', \varphi) e^{im\varphi} r r' ds ds' d\varphi$$

Le second membre s'écrit :

$$\frac{1}{2\pi} \int_0^{2\pi} \int_{\Gamma} \mathbf{E}^{\text{inc}}(x) \cdot \bar{\mathbf{J}}_{\mathbf{m}}(s) e^{im\theta} r ds d\theta = \int_{\Gamma} \mathbf{E}_{\mathbf{m}}^{\text{inc}}(s) \cdot \bar{\mathbf{J}}_{\mathbf{m}}(s) r ds$$

$\mathbf{E}_{\mathbf{m}}^{\text{inc}}$  est la quantité qu'on a introduite dans la section précédente. C'est la composante associée au mode  $\mathbf{m}$ , du champ électrique incident.

Soit une fonction scalaire  $u$ , exprimons le gradient dans le système  $\mathbf{t}, \mathbf{b}, \mathbf{n}$

$$u = \frac{\partial u}{\partial r} \hat{\mathbf{r}} + \frac{1}{r} \frac{\partial u}{\partial \theta} \hat{\boldsymbol{\theta}} + \frac{\partial u}{\partial z} \hat{\mathbf{z}} = (t_x \frac{\partial u}{\partial r} + t_z \frac{\partial u}{\partial z}) \mathbf{t} + \frac{1}{r} \frac{\partial u}{\partial \theta} \mathbf{b} + (-t_z \frac{\partial u}{\partial r} + t_x \frac{\partial u}{\partial z}) \mathbf{n}$$

Le gradient surfacique est par définition la composante tangentielle du gradient, soit :

$$\nabla_{\sigma} u = \frac{\partial u}{\partial s} \mathbf{t} + \frac{1}{r} \frac{\partial u}{\partial \theta} \mathbf{b}$$

On cherche l'expression de la divergence surfacique en effectuant une intégration par parties :

$$\begin{aligned} \int_0^{2\pi} \int_{\Gamma} \nabla_{\sigma} u \cdot \mathbf{v} r ds d\theta &= \int_0^{2\pi} \int_{\Gamma} (r \frac{\partial u}{\partial s} v_t) + \frac{\partial u}{\partial \theta} v_b ds d\theta \\ &= - \int_0^{2\pi} \int_{\Gamma} u \left( r \frac{\partial v_t}{\partial s} + \frac{\partial r}{\partial s} v_t + \frac{\partial v_b}{\partial \theta} \right) ds d\theta \end{aligned}$$

Comme  $t_x = \frac{\partial r}{\partial s}$ , on en déduit l'expression de la divergence surfacique :

$$\text{div}_{\sigma} \mathbf{v} = \frac{\partial v_t}{\partial s} + \frac{t_x}{r} v_t + \frac{1}{r} \frac{\partial v_b}{\partial \theta}$$

On applique cette expression sur les fonctions de base de  $\mathbf{J}$  :

$$\text{div}_{\sigma}(J_t(x') \mathbf{t}') = \left( \frac{\partial \varphi_i}{\partial s'}(s') + \frac{t'_x}{r'} \varphi_i(s') \right) e^{-im\theta'}$$

$$\text{div}_{\sigma}(J_b(x') \mathbf{b}') = -\frac{i}{r'} \psi_i(s') e^{-im\theta'}$$

Explicitons dans un premier temps la matrice

$$S_{ij} = -\frac{i}{k} \int_{\Sigma} \int_{\Sigma} G(x, x') \text{div}_{\sigma} \mathbf{w}_j(x') \text{div}_{\sigma} \bar{\mathbf{w}}_i(x) dx' dx$$

$\mathbf{w}_i$  étant une fonction de base de  $\mathbf{J}$ . On distingue quatre cas :

$$\text{Cas 1 : } \mathbf{w}_i = \varphi_i(s) e^{-im\theta} \mathbf{t} \quad \mathbf{w}_j = \varphi_j(s') e^{-im\theta'} \mathbf{t}'$$

$$S_{ij} = -\frac{i}{k} \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} G(x, x') \left( \frac{\partial \varphi_j}{\partial s'} + \frac{t'_x}{r'} \varphi_j(s') \right) \left( \frac{\partial \varphi_i}{\partial s} + \frac{t_x}{r} \varphi_i(s) \right) e^{im\varphi} r r' ds' ds d\varphi$$

$$\text{Cas 2 : } \mathbf{w}_i = \varphi_i(s) e^{-im\theta} \mathbf{t} \quad \mathbf{w}_j = \psi_j(s') e^{-im\theta'} \mathbf{b}'$$

$$S_{ij} = -\frac{m}{k} \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} G(x, x') \frac{\psi_j(s')}{r'} \left( \frac{\partial \varphi_i}{\partial s} + \frac{t_x}{r} \varphi_i(s) \right) e^{im\varphi} r r' ds' ds d\varphi$$

$$\text{Cas 3 : } \mathbf{w}_i = \psi_i(s) e^{-im\theta} \mathbf{b} \quad \mathbf{w}_j = \varphi_j(s') e^{-im\theta'} \mathbf{t}'$$

$$S_{ij} = \frac{m}{k} \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} G(x, x') \left( \frac{\partial \varphi_j}{\partial s'} + \frac{t'_x}{r'} \varphi_j(s') \right) \frac{\psi_i(s)}{r} e^{im\varphi} r r' ds' ds d\varphi$$

$$\text{Cas 4 : } \mathbf{w}_i = \psi_i(s) e^{-im\theta} \mathbf{b} \quad \mathbf{w}_j = \psi_j(s') e^{-im\theta'} \mathbf{b}'$$

$$S_{ij} = -\frac{im^2}{k} \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} G(x, x') \frac{\psi_j(s')}{r'} \frac{\psi_i(s)}{r} e^{im\varphi} r r' ds' ds d\varphi$$

Explicitons dans un second temps la matrice

$$B_{ij} = -ik \int_{\Sigma} \int_{\Sigma} G(x, x') \mathbf{w}_j(x') \cdot \bar{\mathbf{w}}_i(x) dx' dx$$

$$\text{Cas 1 : } \mathbf{w}_i = \varphi_i(s) e^{-im\theta} \mathbf{t} \quad \mathbf{w}_j = \varphi_j(s') e^{-im\theta'} \mathbf{t}'$$

$$B_{ij} = -ik \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} G(x, x') \varphi_j(s') \varphi_i(s) (t_x t'_x \cos \varphi + t_z t'_z) e^{im\varphi} r r' ds' ds d\varphi$$

$$\text{Cas 2 : } \mathbf{w}_i = \varphi_i(s) e^{-im\theta} \mathbf{t} \quad \mathbf{w}_j = \psi_j(s') e^{-im\theta'} \mathbf{b}'$$

$$B_{ij} = -ik \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} G(x, x') \varphi_j(s') \psi_i(s) t_x \sin \varphi e^{im\varphi} r r' ds' ds d\varphi$$

$$\text{Cas 3 : } \mathbf{w}_i = \psi_i(s) e^{-im\theta} \mathbf{b} \quad \mathbf{w}_j = \varphi_j(s') e^{-im\theta'} \mathbf{t}'$$

$$B_{ij} = ik \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} G(x, x') \psi_j(s') \varphi_i(s) t'_x \sin \varphi e^{im\varphi} r r' ds' ds d\varphi$$

$$\text{Cas 4 : } \mathbf{w}_i = \psi_i(s) e^{-im\theta} \mathbf{b} \quad \mathbf{w}_j = \psi_j(s') e^{-im\theta'} \mathbf{b}'$$

$$B_{ij} = -ik \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} G(x, x') \psi_j(s') \psi_i(s) \cos \varphi e^{im\varphi} r r' ds' ds d\varphi$$

Le second membre s'écrit :

$$F_i^{EFIE} = \int_{\Gamma} \mathbf{E}_{\mathbf{m}}^{\text{inc}}(s) \cdot \bar{\mathbf{w}}_i(s) r ds$$

$$\text{Cas 1 : } \mathbf{w}_i = \varphi_i(s) e^{-im\theta} \mathbf{t}$$

$$F_i^{EFIE} = \int_{\Gamma} (t_x E_{r,m}^{\text{inc}} + t_z E_{z,m}^{\text{inc}}) \varphi_i(s) r ds$$

$$\text{Cas 2 : } \mathbf{w}_i = \psi_i(s) e^{-im\theta} \mathbf{b}$$

$$F_i^{EFIE} = \int_{\Gamma} E_{\theta,m}^{\text{inc}} \psi_i(s) r ds$$

La formulation EFIE conduit à la résolution du système linéaire :

$$(B - S)J = F^{EFIE}$$

On notera que la matrice peut être rendue symétrique en multipliant par -1 les équations portant sur  $J_b$ , il faut faire la même opération sur les composantes correspondantes du second membre. Nous n'utiliserons pas cette propriété de symétrie, car nous préférerons l'utilisation de la CFIE.

### 9.1.3 Formulation MFIE

La formulation variationnelle 3-D de la MFIE s'écrit :

$$\frac{1}{2} \int_{\Sigma} \mathbf{J}(x) \cdot \bar{\mathbf{J}}^t(x) d\sigma(x) + \int_{\Sigma} \int_{\Sigma} (\nabla_y \mathbf{G}(x, y) \times \mathbf{J}(y)) \cdot (\bar{\mathbf{J}}^t(x) \times n(x)) d\sigma(y) d\sigma(x) = \int_{\Sigma} n(x) \times \mathbf{H}^{\text{inc}}(x) \cdot \bar{\mathbf{J}}^t(x) d\sigma(x)$$

$$\mathbf{J}(x), \mathbf{J}^t(x) \in H(\text{div}, \Sigma)$$

$\mathbf{J}^t$  est la fonction test, elle est prise dans le même espace fonctionnel que  $\mathbf{J}$  : l'espace  $H(\text{div}, \Gamma)$ . A priori, la MFIE est consistante si on choisit comme espace fonctionnel  $L^2(\Gamma)$ , mais il paraît plus judicieux de choisir le même espace d'approximation que pour l'EFIE, afin de pouvoir utiliser la CFIE. Explicitons la matrice

$$I_{ij} = \int_{\Sigma} \mathbf{w}_i(x) \cdot \bar{\mathbf{w}}_j(x) dx$$

$$\text{Cas 1 : } \mathbf{w}_i = \varphi_i(s) e^{-im\theta} \mathbf{t} \quad \mathbf{w}_j = \varphi_j(s) e^{-im\theta} \mathbf{t}$$

$$I_{ij} = \int_{\Gamma} \varphi_i(s) \varphi_j(s) r ds$$

$$\text{Cas 4 : } \mathbf{w}_i = \psi_i(s) e^{-im\theta} \mathbf{b} \quad \mathbf{w}_j = \psi_j(s) e^{-im\theta} \mathbf{b}$$

$$I_{ij} = \int_{\Gamma} \psi_i(s) \psi_j(s) r ds$$

Dans le cas 2 et 3 (interactions croisées), les termes de la matrice sont nuls. La matrice  $I$  est une simple matrice de masse.

Explicitons la matrice :

$$Q_{ij}^{\times} = \int_{\Sigma} \int_{\Sigma} \nabla'_x \mathbf{G}(x, x') \times \mathbf{w}_j(x') \cdot (\bar{\mathbf{w}}_i(x) \times n(x)) d\sigma(y) d\sigma(x)$$

$$\text{Cas 1 : } \mathbf{w}_i = \varphi_i(s) e^{-im\theta} \mathbf{t} \quad \mathbf{w}_j = \varphi_j(s') e^{-im\theta'} \mathbf{t}'$$

$$Q_{ij}^{\times} = \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} \nabla'_x \mathbf{G}(x, x') \times \mathbf{t}' \cdot \mathbf{b} \varphi_j(s') \varphi_i(s) e^{im\varphi} r r' ds' ds d\varphi$$

Or, on a :

$$\nabla'_x \mathbf{G}(x, x') = \left( \frac{ik}{R} - \frac{1}{R^2} \right) G(x, x') \mathbf{M} \mathbf{M}'$$

En utilisant

$$\mathbf{M} \mathbf{M}' \times \mathbf{t}' = \mathbf{M} \mathbf{M}' \cdot \mathbf{b}' \mathbf{n}' - \mathbf{M} \mathbf{M}' \cdot \mathbf{n}' \mathbf{b}'$$

On trouve l'expression :

$$Q_{ij}^{\times} = \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} \left( \frac{ik}{R} - \frac{1}{R^2} \right) G(x, x') \left( r t'_z - \cos \varphi (r' t'_z - (z' - z) t'_x) \right) e^{im\varphi} \varphi_j(s') \varphi_i(s) r r' ds' ds d\varphi$$

$$\text{Cas 2 : } \mathbf{w}_i = \varphi_i(s) e^{-im\theta} \mathbf{t} \quad \mathbf{w}_j = \psi_j(s') e^{-im\theta'} \mathbf{b}'$$

$$Q_{ij}^{\times} = \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} \nabla'_x \mathbf{G}(x, x') \times \mathbf{b}' \cdot \mathbf{b} \psi_j(s') \varphi_i(s) r r' ds' ds d\varphi$$

$$Q_{ij}^{\times} = \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} \left( \frac{ik}{R} - \frac{1}{R^2} \right) G(x, x') \sin \varphi(z' - z) e^{im\varphi} \psi_j(s') \varphi_i(s) r r' ds' ds d\varphi$$

Cas 3 :  $\mathbf{w}_i = \psi_i(s) e^{-im\theta} \mathbf{b}$      $\mathbf{w}_j = \varphi_j(s') e^{-im\theta'} \mathbf{t}'$

$$Q_{ij}^{\times} = - \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} \nabla'_{\mathbf{x}} \mathbf{G}(x, x') \times \mathbf{t}' \cdot \mathbf{t} \varphi_j(s') \psi_i(s) r r' ds' ds d\varphi$$

$$Q_{ij}^{\times} = - \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} \left( \frac{ik}{R} - \frac{1}{R^2} \right) G(x, x') \sin \varphi(-r't'_z t_x + rt'_x t_z + (z' - z)t'_x t_x) e^{im\varphi} \varphi_j(s') \psi_i(s) r r' ds' ds d\varphi$$

Cas 4 :  $\mathbf{w}_i = \psi_i(s) e^{-im\theta} \mathbf{b}$      $\mathbf{w}_j = \psi_j(s') e^{-im\theta'} \mathbf{b}'$

$$Q_{ij}^{\times} = - \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} \nabla'_{\mathbf{x}} \mathbf{G}(x, x') \times \mathbf{b}' \cdot \mathbf{t} \psi_j(s') \psi_i(s) r r' ds' ds d\varphi$$

$$Q_{ij}^{\times} = - \int_{-\pi}^{\pi} \int_{\Gamma} \int_{\Gamma} \left( \frac{ik}{R} - \frac{1}{R^2} \right) G(x, x') \left( r't_z - \cos \varphi(r t_z + (z' - z) t_x) \right) e^{im\varphi} \psi_j(s') \psi_i(s) r r' ds' ds d\varphi$$

Le second membre s'écrit :

$$F_i^{MFIE} = \int_{\Gamma} \mathbf{J}_{\mathbf{m}}^{\text{inc}}(s) \cdot \bar{\mathbf{w}}_i(s) r ds$$

Cas 1 :  $\mathbf{w}_i = \varphi_i(s) e^{-im\theta} \mathbf{t}$

$$F_i^{MFIE} = \int_{\Gamma} H_{\theta, m}^{inc} \varphi_i(s) r ds$$

Cas 2 :  $\mathbf{w}_i = \psi_i(s) e^{-im\theta} \mathbf{b}$

$$F_i^{MFIE} = - \int_{\Gamma} (t_x H_{r, m}^{inc} + t_z H_{z, m}^{inc}) \psi_i(s) r ds$$

La formulation MFIE conduit à la résolution du système linéaire :

$$\left( \frac{1}{2} I + Q^{\times} \right) J = Z^{MFIE}$$

La matrice ne peut être rendue symétrique.

#### 9.1.4 Formulation CFIE

Elle s'obtient par simple combinaison linéaire de l'EFIE et la MFIE. On obtient le système linéaire suivant :

$$\left[ \alpha(B - S) + (1 - \alpha) \left( \frac{1}{2} I + Q^{\times} \right) \right] J = \alpha F^{EFIE} + (1 - \alpha) F^{MFIE}$$

L'EFIE et la MFIE sont mal posées pour un certain nombre de fréquences, qui correspondent aux modes propres de l'intérieur de l'objet. La CFIE est toujours bien posée pour des fréquences réelles. Dans toute la suite de l'exposé, on prendra toujours  $\alpha = 0.5$ .

## 9.2 Méthode d'intégration

### 9.2.1 Méthode d'intégration pour la partie régulière

#### Première intégration en $\varphi$

Nous introduisons les variables intermédiaires suivantes :

$$\begin{aligned} G1^{EFIE}(s, s') &= -i \int_{-\pi}^{\pi} G(x, x') e^{im\varphi} d\varphi = -2i \int_0^{\pi} G(x, x') \cos(m\varphi) d\varphi \\ G\cos^{EFIE}(s, s') &= -i \int_{-\pi}^{\pi} G(x, x') \cos \varphi e^{im\varphi} d\varphi = -2i \int_0^{\pi} G(x, x') \cos \varphi \cos(m\varphi) d\varphi \\ G\sin^{EFIE}(s, s') &= -i \int_{-\pi}^{\pi} G(x, x') \sin \varphi e^{im\varphi} d\varphi = 2 \int_0^{\pi} G(x, x') \sin \varphi \sin(m\varphi) d\varphi \end{aligned}$$

L'idée est de calculer d'abord ces variables pour éliminer l'intégration en  $\varphi$ . On définit des variables similaires pour la MFIE :

$$\begin{aligned} G1^{MFIE}(s, s') &= 2 \int_0^{\pi} \left( \frac{ik}{|x - x'|} - \frac{1}{|x - x'|^2} \right) G(x, x') \cos(m\varphi) d\varphi \\ G\cos^{MFIE}(s, s') &= 2 \int_0^{\pi} \left( \frac{ik}{|x - x'|} - \frac{1}{|x - x'|^2} \right) G(x, x') \cos \varphi \cos(m\varphi) d\varphi \\ G\sin^{MFIE}(s, s') &= 2i \int_0^{\pi} \left( \frac{ik}{|x - x'|} - \frac{1}{|x - x'|^2} \right) G(x, x') \sin \varphi \sin(m\varphi) d\varphi \end{aligned}$$

Il ne reste alors plus que la double intégrale sur  $\Gamma$ . Les variables  $G1$  et  $G\cos$  ne sont pas définies pour  $s = s'$ , car on a alors une intégrale divergente.

#### Intégrales régulières pour l'EFIE

On doit maintenant évaluer des intégrales de la forme :

$$I = \int_{s_1}^{s_2} \int_{s'_1}^{s'_2} g(s, s') ds' ds \quad ,$$

qu'on intègre en deux étapes :

$$h(s) = \int_{s'_1}^{s'_2} g(s') ds'$$

Cette intégrale sera appelée **intégrale intérieure**. L'autre étape est :

$$I = \int_{s_1}^{s_2} h(s) ds$$

Cette intégrale sera appelée **intégrale extérieure**. On note :

$$\begin{aligned} A1(s, s') &= \left[ k(t_x t'_x G\cos^{EFIE} + t_z t'_z G1^{EFIE}) rr' - \frac{1}{k} t_x t'_x G1^{EFIE} \right] ds' ds \\ A2(s, s') &= -\frac{t_x r'}{k} G1^{EFIE} ds \\ A3(s, s') &= -\frac{t'_x r}{k} G1^{EFIE} ds' \end{aligned}$$

$$\begin{aligned}
A4(s, s') &= -\frac{rr'}{k} G1^{\text{EFIE}} \\
A5(s, s') &= (k \text{Gcos}^{\text{EFIE}}_{rr'} - \frac{m^2}{k} G1^{\text{EFIE}}) ds ds' \\
A6(s, s') &= (k \text{Gsin}^{\text{EFIE}}_{t_x rr'} + \frac{im}{k} t_x G1^{\text{EFIE}}) ds ds' \\
A7(s, s') &= \frac{im}{k} G1^{\text{EFIE}}_r r ds' \\
A8(s, s') &= -(k \text{Gsin}^{\text{EFIE}}_{t'_x rr'} + \frac{im}{k} t'_x G1^{\text{EFIE}}) ds ds' \\
A9(s, s') &= -\frac{im}{k} G1^{\text{EFIE}}_r r' ds
\end{aligned}$$

Ces variables apparaissent lorsqu'on fait un changement de variables afin de ramener l'arc  $[s_1, s_2]$ , et l'arc  $[s'_1, s'_2]$  au segment unité  $[0, 1]$ . Explicitons la partie régulière de l'opérateur (B-S) sur les quatre cas précédemment rencontrés :

Cas 1 : interaction (t,t')

$$\begin{aligned}
(B - S)_{ij} &= \sum_{p,q=1}^{k+1} \left[ A1(\xi_p, \xi'_q) \hat{\varphi}_i(\hat{\xi}_p) \hat{\varphi}_j(\hat{\xi}'_q) + A2(\xi_p, \xi'_q) \hat{\varphi}_i(\hat{\xi}_p) \frac{\partial \hat{\varphi}_j}{\partial \hat{s}}(\hat{\xi}'_q) \right. \\
&\quad \left. + A3(\xi_p, \xi'_q) \frac{\partial \hat{\varphi}_i}{\partial \hat{s}}(\hat{\xi}_p) \hat{\varphi}_j(\hat{\xi}'_q) + A4(\xi_p, \xi'_q) \frac{\partial \hat{\varphi}_i}{\partial \hat{s}}(\hat{\xi}_p) \frac{\partial \hat{\varphi}_j}{\partial \hat{s}}(\hat{\xi}'_q) \right] \omega_p \omega_q
\end{aligned}$$

Cas 2 : interaction (t,b')

$$(B - S)_{ij} = \sum_{p,q=1}^{k+1} \left( A6(\xi_p, \xi'_q) \hat{\varphi}_i(\hat{\xi}_p) \hat{\psi}_j(\hat{\xi}'_q) + A7(\xi_p, \xi'_q) \frac{\partial \hat{\varphi}_i}{\partial \hat{s}}(\hat{\xi}_p) \hat{\psi}_j(\hat{\xi}'_q) \right) \omega_p \omega_q$$

Cas 3 : interaction (b,t')

$$(B - S)_{ij} = \sum_{p,q=1}^{k+1} \left( A8(\xi_p, \xi'_q) \hat{\psi}_i(\hat{\xi}_p) \hat{\varphi}_j(\hat{\xi}'_q) + A9(\xi_p, \xi'_q) \hat{\psi}_i(\hat{\xi}_p) \frac{\partial \hat{\varphi}_j}{\partial \hat{s}}(\hat{\xi}'_q) \right) \omega_p \omega_q$$

Cas 4 : interaction (b, b')

$$(B - S)_{ij} = \sum_{p,q=1}^{k+1} A5(\xi_p, \xi'_q) \hat{\psi}_i(\hat{\xi}_p) \hat{\psi}_j(\hat{\xi}'_q) \omega_p \omega_q$$

On a ici utilisé une intégration avec points de Gauss sur l'intégrale extérieure et l'intégrale intérieure.  $\xi_p$  et  $\omega_p$  désignent respectivement le  $p$ -ième point de Gauss et le  $p$ -ième poids de Gauss.

### Intégrales régulières pour la MFIE

De manière similaire, on fait apparaître les termes suivants pour le calcul de la MFIE :

$$F11 = \left[ \text{Gsin}^{\text{MFIE}}(r' t'_z t_x - r t'_x t_z - (z' - z) t'_x t_x) \right] r r' ds' ds$$

$$F_{12} = - \left[ G1^{\text{MFIE}} r' t_z - G\cos^{\text{MFIE}}(r t_z + (z' - z) t_x) \right] r r' ds' ds$$

$$F_{21} = \left[ G1^{\text{MFIE}} r t'_z - G\cos^{\text{MFIE}}(r' t'_z - (z' - z) t'_x) \right] r r' ds' ds$$

$$F_{22} = \left[ G\sin^{\text{MFIE}}(z' - z) \right] r r' ds' ds$$

On exprime le terme générique de la matrice Q en fonction de ces termes :

Cas 1 : interaction ( $\mathbf{t}, \mathbf{t}'$ )

$$Q_{ij} = \sum_{p,q=1}^{k+1} \left( F_{21}(\xi_p, \xi'_q) \hat{\varphi}_i(\hat{\xi}_p) \hat{\varphi}_j(\hat{\xi}'_q) \right) \omega_p \omega_q$$

Cas 2 : interaction ( $\mathbf{t}, \mathbf{b}'$ )

$$Q_{ij} = \sum_{p,q=1}^{k+1} \left( F_{22}(\xi_p, \xi'_q) \hat{\varphi}_i(\hat{\xi}_p) \hat{\psi}_j(\hat{\xi}'_q) \right) \omega_p \omega_q$$

Cas 1 : interaction ( $\mathbf{b}, \mathbf{t}'$ )

$$Q_{ij} = - \sum_{p,q=1}^{k+1} \left( F_{11}(\xi_p, \xi'_q) \hat{\psi}_i(\hat{\xi}_p) \hat{\varphi}_j(\hat{\xi}'_q) \right) \omega_p \omega_q$$

Cas 1 : interaction ( $\mathbf{b}, \mathbf{b}'$ )

$$Q_{ij} = - \sum_{p,q=1}^{k+1} \left( F_{12}(\xi_p, \xi'_q) \hat{\psi}_i(\hat{\xi}_p) \hat{\psi}_j(\hat{\xi}'_q) \right) \omega_p \omega_q$$

### 9.2.2 Règles simples d'intégration dans le cas régulier

On a séparé la triple intégrale sur  $(s, s', \varphi)$ , en une première intégration sur  $\varphi$  avec les variables G1, Gcos et Gsin. L'intégration sur  $\varphi$  s'effectue en prenant  $N_\phi + 1$  points de Gauss sur l'intervalle  $[0, \pi]$ . L'intégration sur  $s$  et  $s'$  s'effectue en utilisant  $k + 1$  points de Gauss sur l'intervalle  $[s_1, s_2]$  et l'intervalle  $[s'_1, s'_2]$ . La première interrogation est sur le choix de l'entier  $k$ . Pour répondre à cette question, nous utilisons une intégration très précise sur  $\varphi$ , et nous calculons l'erreur maximale faite sur le coefficient de la matrice en fonction de  $k$ . Cette étude est faite sur le tableau 9.1. On observe une convergence exponentielle en fonction du nombre de

Nombre de points de Gauss - 1	1	2	3	4	5
Erreur $L^\infty$	1.37e-2	1.98e-3	1.3e-5	4.7e-8	2.1e-10

FIG. 9.1 – Erreur  $L^\infty$  sur la matrice en fonction du nombre de points d'intégration de  $s, s'$

points. C'est normal car on intègre une fonction infiniment régulière et analytique. Au vu des valeurs, il semble judicieux de choisir  $k = 2$  sur cet exemple, qui est l'ordre d'approximation (P2). Dans la suite, on prendra  $k$  égal à l'ordre d'approximation utilisé. Il nous reste maintenant à déterminer la règle d'intégration pour  $\varphi$ . La fonction-type à intégrer vaut :

$$\int_0^\pi \frac{e^{ikR}}{R} \cos(m\varphi) d\varphi \quad \text{où} \quad R = \sqrt{(r' - r)^2 + (z' - z)^2 + 2r'r'(1 - \cos\varphi)}$$

On a un produit de deux fonctions oscillantes, qui donnent donc une sinusoïde avec une phase, qui est au pire la somme des deux phases. On doit donc considérer la phase suivante :

$$k R + m \varphi$$

La variable  $R$  peut varier sur l'intervalle  $[0, 2r_{\max}]$ , le cas le plus défavorable se produit lorsque  $s = s'$ ,  $\varphi = \pi$ . La distance entre les deux points d'intégration  $x$  et  $x'$  est dans ce cas de  $2r$ . Pour ce qui est de la variable  $\varphi$ , elle varie de 0 à  $\pi$ . La phase maximale est donc de :

$$2k r_{\max} + m\pi$$

On choisit d'utiliser 4 points d'intégration par longueur d'onde pour évaluer les fonctions oscillantes, on exige l'ordre d'intégration suivant :

$$N_\phi = 2(m + 4fx_{\max})$$

où  $f$  désigne la fréquence,  $f = \frac{k}{2\pi}$ . On a deux termes en concurrence ; le premier est dû à l'oscillation du noyau de Green et le second est dû à l'oscillation des fonctions de base. Ce dernier effet n'est pas négligeable, une première idée est de choisir  $N_\phi$  indépendant du numéro du mode. Sur la figure 9.2, on compare les deux approches (choisir  $N_\phi$  constant ou dépendant du mode). On voit que si on choisit  $N_\phi$  constant, l'erreur commise est exponentiellement croissante suivant  $m$ . Lorsqu'on adapte  $N_\phi$  au numéro du mode, on observe une erreur légèrement décroissante lorsque  $m$  augmente. Sur la figure, ce n'est pas visible, car on a atteint les limites de la précision machine. La séparation de l'intégrale en  $\varphi$  des autres variables d'intégration permet d'obtenir

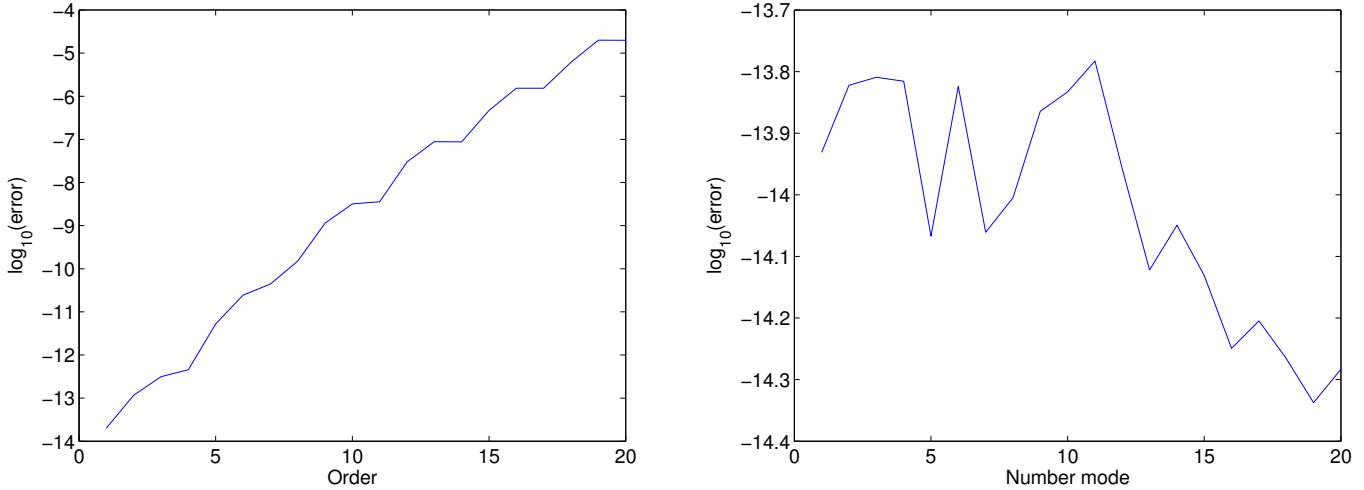


FIG. 9.2 – Evolution du logarithme de l'erreur  $L^\infty$  sur la matrice suivant le numéro de mode. On ne considère ici que la partie régulière de la matrice, le nombre de points d'intégration sur  $s, s'$  est fixé. A gauche, on fixe le nombre de points d'intégration sur  $\varphi$ , à droite on l'adapte au numéro du mode.

un coût qui est de complexité :

$$C(k+1)^2 N_e^2 N_\phi$$

La constante  $C$  est indépendante de l'ordre d'approximation  $k$  sous les hypothèses :

$$N_e, N_\phi \text{ assez grand}$$

$N_e$  représente le nombre de segments du maillage. On fait le quotient de ce coût par le carré du nombre de degrés de liberté, pour comparer les différents ordres d'approximation :

$$\frac{\text{Coût calcul partie régulière}}{(\text{Nombre ddl})^2} = C \frac{(k+1)^2}{k^2} N_\phi$$

Comme on l'a vu, le nombre de points d'intégration  $N_\phi$  et la constante C sont indépendants de  $k$ . On peut donc tracer la complexité de ce calcul en fonction de l'ordre d'approximation sur la figure 9.3. On constate qu'il est avantageux de monter en ordre car la partie régulière est plus

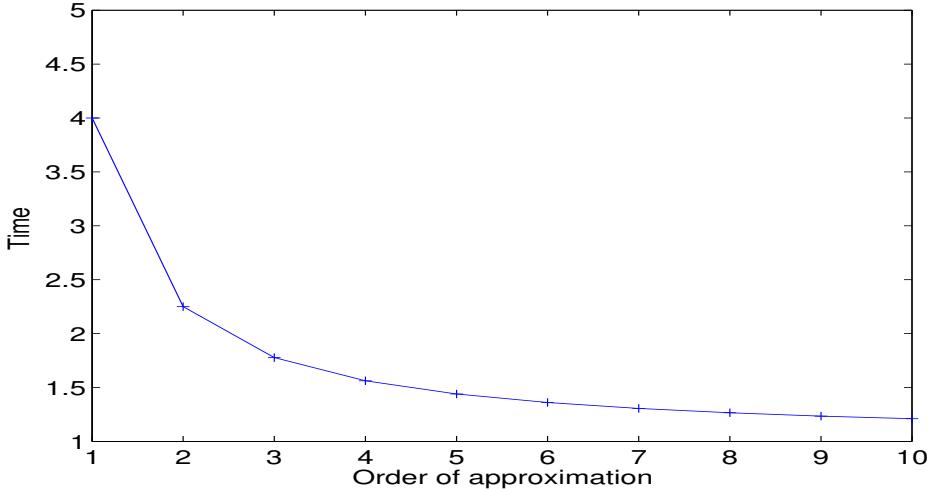


FIG. 9.3 – Temps de calcul théorique en fonction de l'ordre d'approximation, pour un nombre de ddl fixé. Partie régulière uniquement.

rapide à évaluer.

### 9.2.3 Calcul des intégrales singulières

Trois cas distincts sont à traiter, le cas où la double intégrale s'effectue sur le même élément (éléments confondus), le cas où la double intégrale s'effectue sur deux éléments proches (éléments joints) et le cas où la double intégrale s'effectue sur le même élément proche de l'axe (éléments axiaux). Par proche, on veut signifier qu'une des extrémités du segment appartient à l'axe. La problématique d'intégration de singularités est largement abordée dans la littérature [Sauter et Krapp, 1996], [Sauter et Lage, 2000], [Singh et Tanaka, 1999], [Schwab et Wendland, 1992].

#### Éléments confondus

On choisit d'intégrer une partie des intégrales singulières par la technique d'intégration régulière. Pour  $\varphi \in [\varphi_1, \pi]$ , on utilise la technique d'intégration régulière. On est confronté à un choix cornélien. Soit on prend  $\varphi_1$  assez élevé, on a ainsi une partie régulière facile à intégrer. En contrepartie, on aura plus de problèmes sur la partie singulière car la fonction à intégrer sur  $[0, \varphi_1]$  variera beaucoup sur cet intervalle. Soit on prend  $\varphi_1$  petit, la partie régulière sera plus délicate à intégrer, mais la partie singulière n'aura pas beaucoup d'oscillations. On choisit de privilégier ce dernier point, en choisissant  $\varphi_1$  de telle sorte que l'intervalle  $[0, \varphi_1]$  soit inférieur à une demi-période de l'oscillation de la fonction à intégrer. On prend en conséquence :

$$\varphi_1 = \frac{\pi}{m + 4 f x_{max}}$$

Sur cette partie régulière, il est nécessaire de la surintégrer, on choisit en conséquence  $2k + 1$  points de Gauss pour l'intégrale extérieure (en  $s$ ), et  $2k + 1$  points de Gauss pour l'intégrale intérieure (en  $s'$ ).

On s'intéresse maintenant à l'intégration sur l'intervalle  $[0, \varphi_1]$ . On cherche à estimer numériquement le nombre de points d'intégration "satisfaisant" pour évaluer l'intégrale extérieure. On obtient les résultats de la figure 9.4. On a calculé l'erreur pour 20 modes, et on observe que le maximum

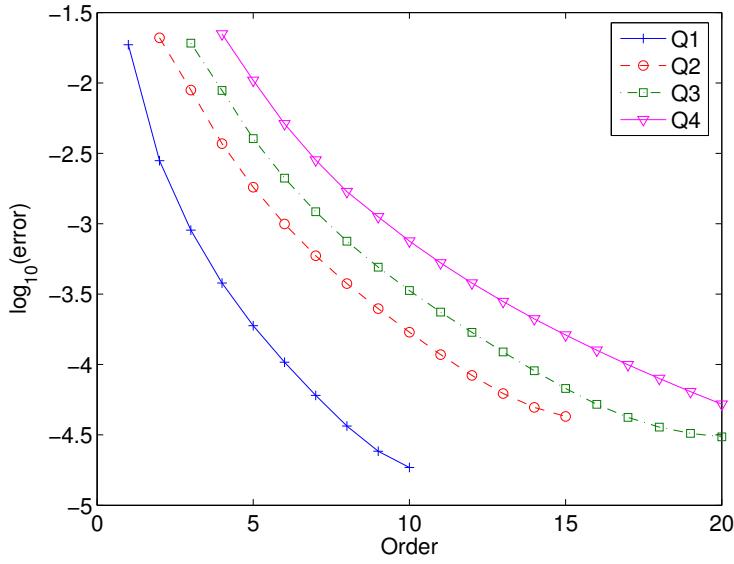


FIG. 9.4 – Evolution de l'erreur  $L^\infty$  sur l'intégrale extérieure en fonction de  $k$  ( $k + 1$  points de quadrature) pour divers ordres d'approximation

est atteint pour le mode 1. La convergence est rapide, mais il faut néanmoins prendre  $2k + 1$  points de Gauss d'intégration pour avoir une erreur faible.

En ce qui concerne l'intégrale intérieure, on doit intégrer une fonction singulière au point  $(\xi_p, 0)$  sur le rectangle  $[0, 1] \times [0, \varphi]$ . Plusieurs possibilités s'offrent à nous. Une technique relativement simple d'intégration de fonctions singulières est la technique utilisant la transformation de Duffy [Duffy, 1982] [Schwab et Wendland, 1992]. On découpe notre domaine d'intégration en triangles (cf. figure 9.5). Sur chaque triangle, on applique la transformation de Duffy, pour passer du cube unité  $[0, 1]^2$  vers le triangle de sommets  $S_0$ ,  $S_1$  et  $S_2$  :

$$F_T(\hat{x}, \hat{y}) = (1 - \hat{x} - (1 - \hat{x})\hat{y})S_0 + \hat{x}S_1 + (1 - \hat{x})\hat{y}S_2$$

On a choisi le point  $S_1$  comme point singulier, les points d'intégration vont s'accumuler sur ce point. On effectue le changement de variables :

$$\int_T f(x, y) dx dy = \int_{\hat{T}} f(x, y) |Det(DF_T)| d\hat{x} d\hat{y}$$

On utilise des points de Gauss sur le carré unité. On peut réinterpréter ce changement de variables comme une formule de quadrature sur le triangle :

$$\int_T f(x, y) dx dy = \sum_m \omega_m f(\xi_m^1, \xi_m^2)$$

avec les points définis par :

$$(\xi_m^1, \xi_m^2) = F_T(\hat{\xi}_m^1, \hat{\xi}_m^2)$$

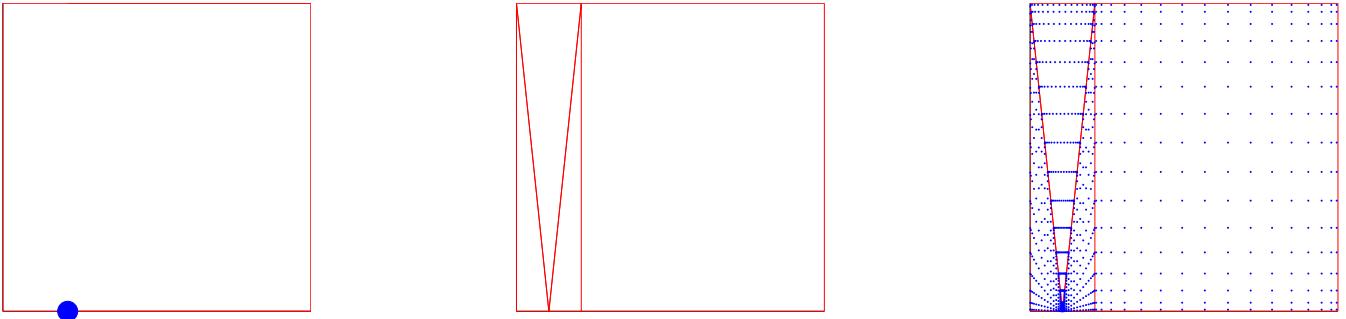


FIG. 9.5 – Position des points d'intégration après application de la transformation de Duffy. A gauche, domaine d'intégration (le point bleu symbolise la localisation de la singularité), au milieu maillage du domaine d'intégration pour s'adapter à la singularité, et à droite position des points d'intégration.

et les poids définis par :

$$\omega_m = \hat{\omega}_m^1 \hat{\omega}_m^2 |\text{Det}(DF_T)|(\hat{\xi}_m^1, \hat{\xi}_m^2)$$

Cette règle d'intégration est connue sous le nom de formules de Gauss-Radau.

Une autre possibilité est de passer en coordonnées polaires, l'origine étant le point  $(\xi_p, 0)$ . On utilise alors le même découpage en triangles que pour la transformation de Duffy. On réinterprète le passage en coordonnées polaires comme des formules de quadrature sur chaque triangle.

Une autre possibilité, pas très bonne, est de prendre les points de Gauss sur les trois rectangles de la figure 9.6.

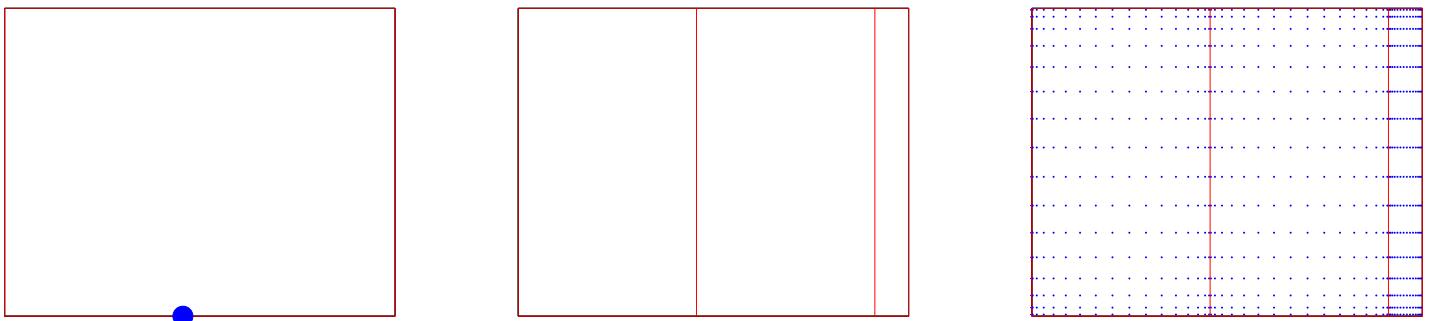


FIG. 9.6 – Position des points d'intégration de Gauss. A gauche, domaine d'intégration (le point bleu symbolise la localisation de la singularité), au milieu maillage du domaine d'intégration pour s'adapter à la singularité, et à droite position des points d'intégration.

Une dernière possibilité, est de garder la séparation des variables en  $s$ ,  $s'$  et  $\varphi$ . Sur l'intégrale en  $s$ , on utilise les points de Gauss  $\xi_p$  classiques. Sur l'intégrale en  $s'$ , on utilise les points de Gauss en leur faisant subir une transformation pour accumuler les points sur le point singulier  $x_0 = \xi_p$ . On prend les points et poids suivants pour  $s'$  sur l'intervalle  $[0, \xi_p]$  :

$$\xi_q = x_0 (1 - \hat{\xi}_q^2) \quad \omega_q = 2 \hat{\xi}_q x_0 \hat{\omega}_q$$

De même, on considère les points suivants sur l'intervalle  $[\xi_p, 2\xi_p]$  :

$$\xi_q = x_0 + \hat{\xi}_q^2 x_0 \quad \omega_q = 2 \hat{\xi}_q x_0 \hat{\omega}_q$$

L'intégrale en  $\varphi$  sur l'intervalle  $[0, \varphi_1]$  est évaluée avec des points de Gauss s'accumulant en 0 :

$$\xi_q = \varphi_1 \hat{\xi}_q^2 \quad \omega_q = 2 \hat{\xi}_q \varphi_1 \hat{\omega}_q$$

L'avantage de cette technique est d'être moins coûteuse que les autres techniques qui ne réalisent pas la séparation de variables, notamment pour des ordres d'approximation élevés. On appellera cette technique "Gauss-squared", car on utilise le carré des points de Gauss, de telle sorte que les points d'intégration s'accumulent près de la singularité (cf. figure 9.7).

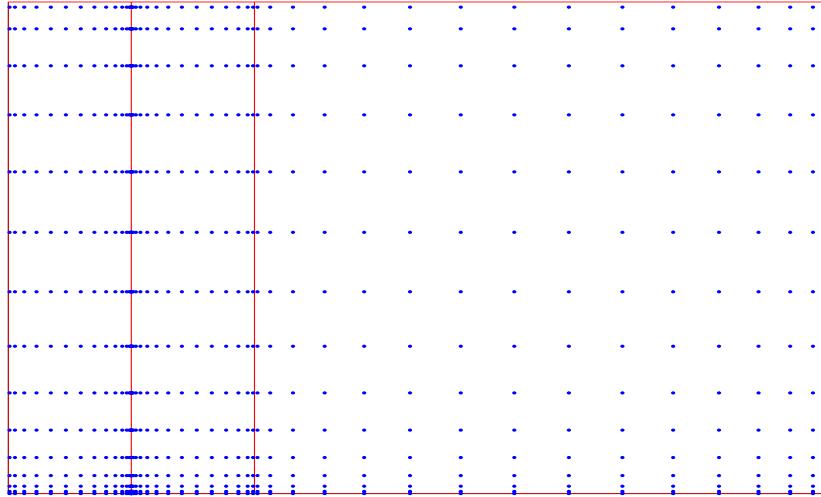


FIG. 9.7 – Distribution des points d'intégration pour la technique "Gauss-squared".

On cherche à valider ces différentes techniques d'intégration pour l'intégrale intérieure. On obtient les résultats de la figure 9.8. Sur ce cas, la transformation de Duffy converge plus rapidement pour des ordres d'intégration raisonnables. Toutefois, la technique "Gauss-squared" donne des résultats proches, et elle est plus efficace lorsqu'on monte en ordre, à cause de la séparation des variables  $s, s', \varphi$ . On choisit d'utiliser cette technique avec  $2k + 1$  points de quadrature pour l'intégrale intérieure et extérieure dans le cas confondu.

### Éléments joints

C'est le cas où les deux segments d'intégration sont adjacents. Au lieu d'avoir une ligne singulière  $s = s'$ , on a un seul point singulier  $s = s' = 0$ . A priori, on peut utiliser des points de Gauss pour évaluer l'intégrale intérieure et extérieure. Comme ces points ne passent pas par les extrémités 0 et 1, les deux intégrales sont régulières. Toutefois, il faut utiliser un ordre d'intégration suffisamment élevé pour obtenir une erreur faible sur les coefficients de la matrice. Sur la figure 9.9, on a calculé l'erreur  $L^\infty$  pour les deux intégrales en fonction du nombre de points de quadrature. Contrairement à ce qu'on pouvait imaginer, c'est l'intégrale extérieure qui est évaluée la moins précisément lorsqu'on utilise le même ordre d'intégration. On utilise le même nombre de points de quadrature que dans le cas des éléments confondus ( $2k + 1$ ).

### Éléments axiaux

Pour tous les éléments proches de l'axe, en plus de la singularité  $s = s'$ , on a une singularité pour  $s = s' = 0$  et pour tous les angles  $\varphi \in [0, \pi]$ . On ne peut donc évaluer une partie régulière

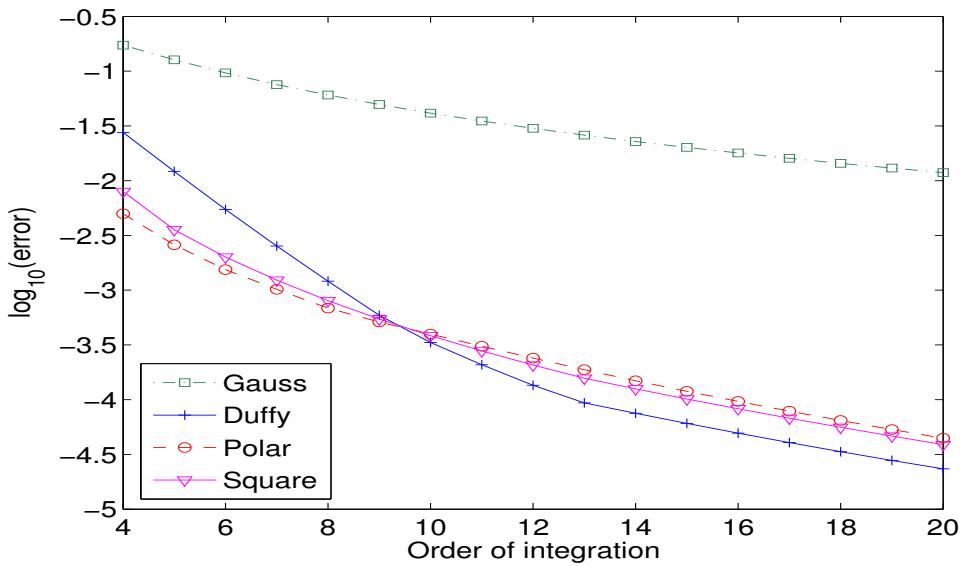


FIG. 9.8 – Evolution de l'erreur  $L^\infty$  sur la matrice, lorsque l'on évalue l'intégrale intérieure avec un certain nombre de points de quadrature, avec diverses techniques d'intégration.

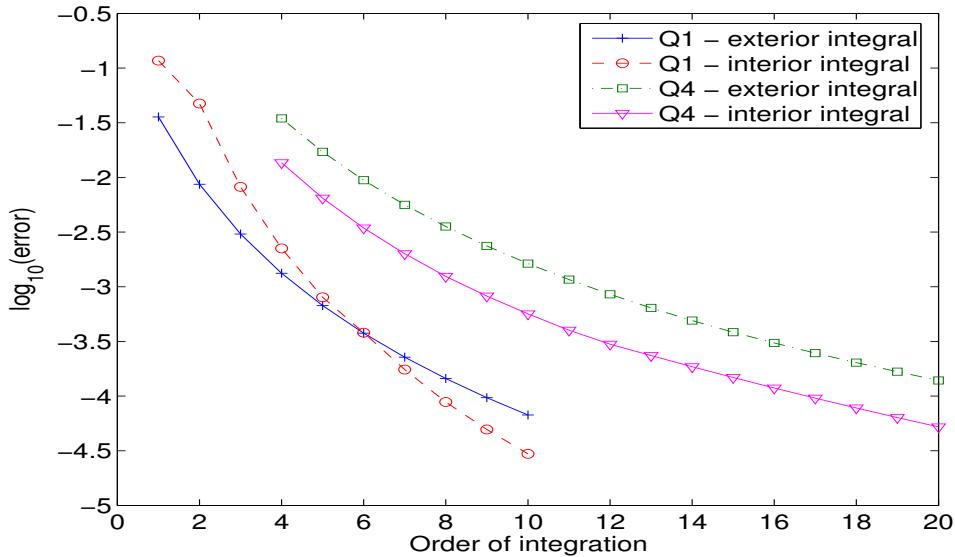


FIG. 9.9 – Evolution de l'erreur  $L^\infty$  sur la matrice, lorsque l'on évalue les intégrales faisant intervenir des segments adjacents, avec un certain nombre de points de quadrature.

comme pour les autres éléments. On fait le choix d'utiliser des points de Gauss s'accumulant en 0 pour l'intégrale extérieure en  $s$  :

$$\xi_q = \hat{\xi}_q^2 \quad \omega_q = 2\hat{\xi}_q \hat{\omega}_q$$

On élimine ainsi les singularités du type  $\frac{1}{\sqrt{s}}$  (a priori, les singularités sont logarithmiques en  $s$ ). Pour l'intégrale intérieure en  $(s', \varphi)$ , on utilise la même règle d'intégration que pour les éléments

confondus (Gauss-squared). Le nombre de points d'intégration est pris égal à :

$$N_\phi = \max(3k + 1, 2m + 1)$$

On choisit de faire évoluer le nombre de points d'intégration en fonction du numéro de mode.

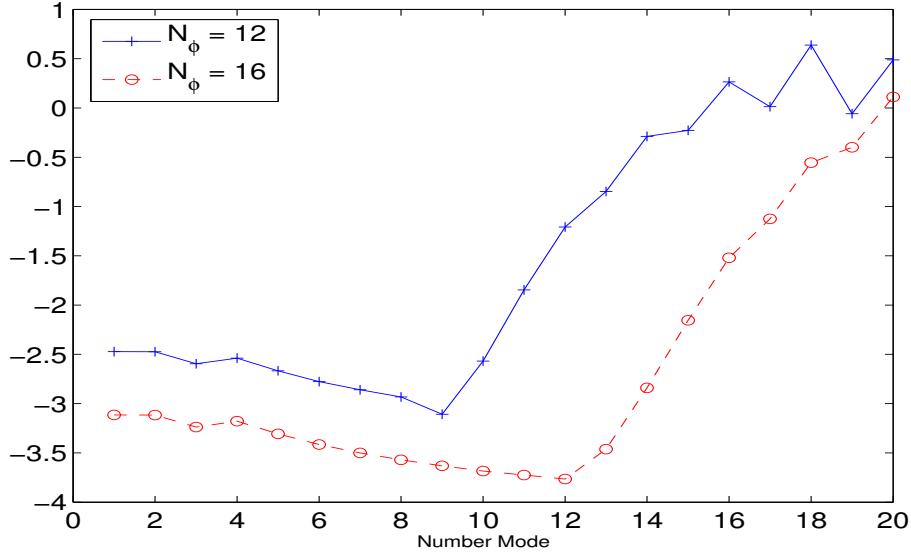


FIG. 9.10 – Evolution de l'erreur  $L^\infty$  sur la matrice, lorsque l'on évalue les intégrales des éléments axiaux, en fonction du numéro de mode. Le nombre de points de l'intégrale intérieure est fixé.

Si on n'adapte pas le nombre de points au numéro du mode, on obtient une erreur exponentiellement croissante sur les interactions de l'axe (cf. figure 9.10). On calcule les erreurs commises

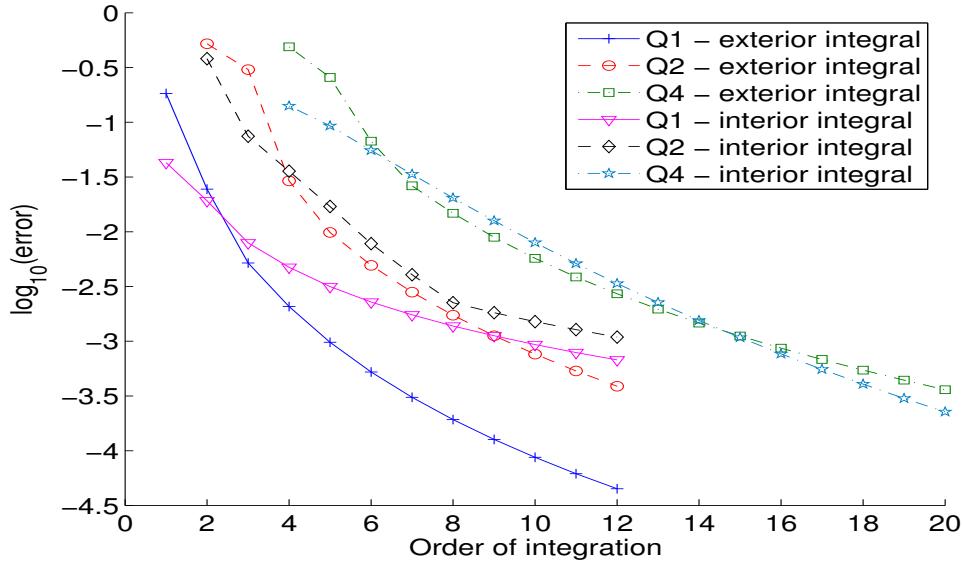


FIG. 9.11 – Evolution de l'erreur  $L^\infty$  sur la matrice, lorsque l'on évalue les intégrales des éléments axiaux. On fait varier le nombre de points d'intégration de l'intégrale extérieure et intérieure. Mode 1.

sur la matrice en fonction du nombre de points d'intégration, lorsqu'on s'intéresse au mode 1. On décide de prendre  $3k + 1$  points d'intégration pour évaluer l'intégrale extérieure.

### 9.3 Précision de la méthode

Dans cette section, on valide la méthode numérique en étudiant la convergence de celle-ci sur des cas tests relativement simples.

#### 9.3.1 Cas de la sphère parfaitement conductrice

On considère la diffraction par une sphère de rayon 5. On affiche les courants  $J_t$  sur un arc de sphère ( $\theta = 0$ ) sur la figure 9.12. Il est également possible de les visualiser sur toute la surface de l'objet (cf. figure 9.13). Nous calculons l'erreur  $L^2$  entre la solution numérique

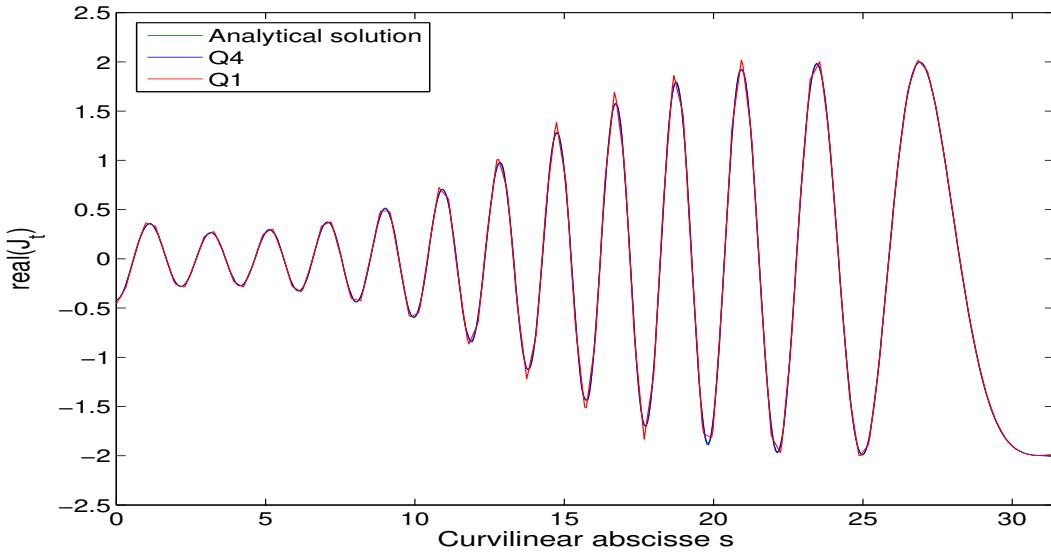


FIG. 9.12 – Partie réelle du courant  $J_t$  en fonction de l'abscisse curviligne  $s$ . La solution  $P_1$  colle très bien à la solution analytique, bien qu'elle soit en “dents de scie”.

et la solution analytique en fonction du pas de maillage, sur la figure 9.14. On peut voir que l'intégration approchée est la cause d'un “plateau” dans la convergence. Au delà d'un certain pas de maillage, il est inutile de raffiner le maillage pour obtenir une solution plus précise. Le raffinement de maillage a presque l'effet inverse, l'erreur augmente de plus en plus. Si on désire obtenir une solution plus précise, il faut prendre plus de points d'intégration, notamment sur les intégrales singulières. Cet effet plateau intervient assez tardivement sur des maillages contenant plus de dix points par longueur d'onde (-1 en abscisse). Sur ce cas simple, on compare l'ordre 1 avec l'ordre 5, en exigeant une précision de 5%. Pour  $Q_1$ , il est nécessaire d'avoir un maillage de 170 degrés de liberté contre 128 degrés de liberté pour  $Q_5$ . Sur ce cas académique, on voit qu'il est intéressant de monter en ordre.

#### 9.3.2 Cas du cylindre parfaitement conducteur

On considère la diffraction par un cylindre avec peu de longueurs d'onde (cf. figure 9.15) On obtient les courbes de convergence de la figure 9.16. La solution de référence a été calculée sur

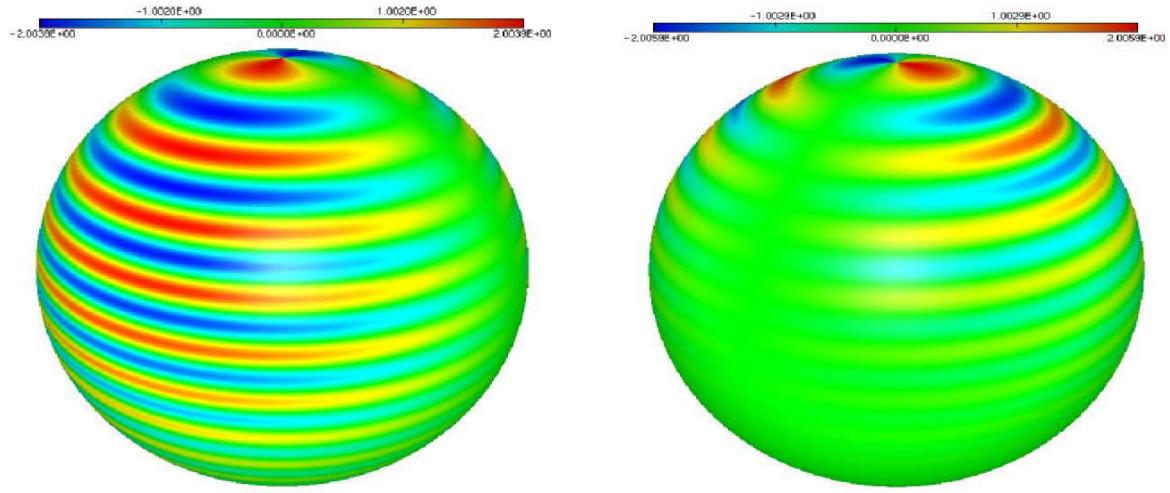


FIG. 9.13 – A gauche, partie réelle de  $J_t$ , à droite, partie réelle de  $J_b$

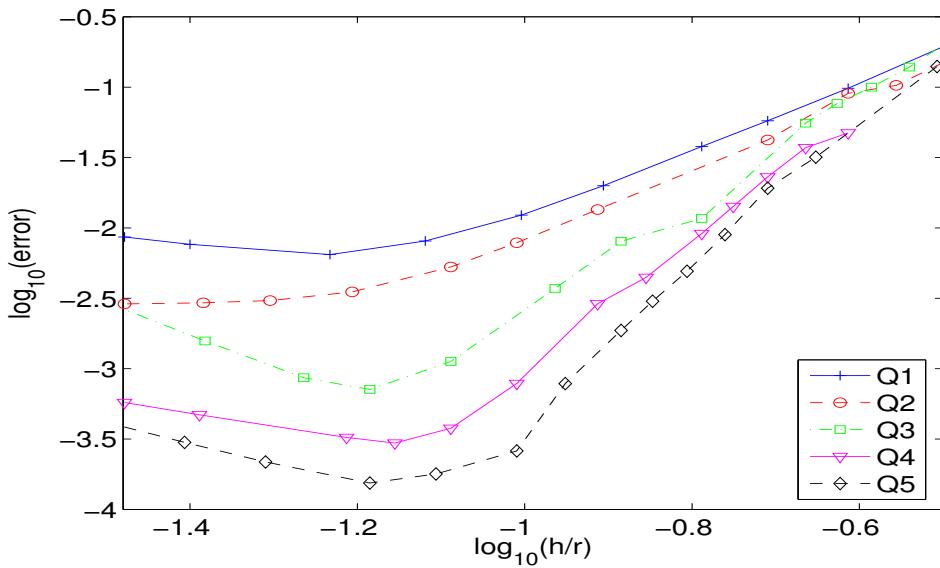


FIG. 9.14 – Evolution de l'erreur en fonction de  $h/r$  où  $r$  est l'ordre d'approximation. Cas de la sphère parfaitement conductrice de rayon 5.

un maillage  $Q_{15}$  avec 240 degrés de liberté. On voit que la présence d'un coin dans le maillage perturbe l'ordre de convergence de la méthode. Néanmoins, il reste toujours très intéressant de monter en ordre. Pour atteindre une erreur relative de 5%,  $Q_1$  a besoin de 100 degrés de liberté, alors que  $Q_4$  n'a besoin que de 82 degrés de liberté.

### 9.3.3 Cas du cône-sphère parfaitement conducteur

On s'intéresse à un cône parfaitement conducteur pour lequel on dispose d'une référence (JINA 90). On donne ainsi une validation externe du code de calcul. La géométrie testée est le noyau conducteur du cône-sphère de la figure 9.17. On calcule sur cette géométrie la SER monostatique pour une fréquence physique de 2.92 Ghz, en faisant varier l'angle d'incidence,

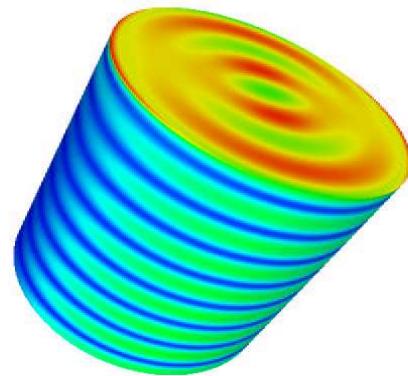


FIG. 9.15 – module de la partie réelle du courant magnétique ( $|Re(J)|$ ) sur la surface du cylindre. L’onde plane est axiale.

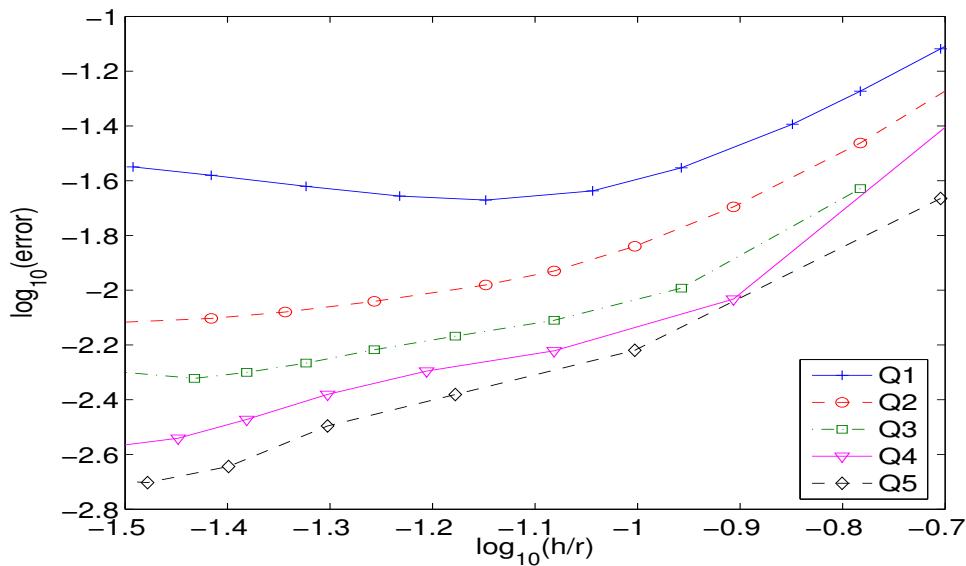


FIG. 9.16 – Evolution de l’erreur  $L^2$  en fonction de  $h/r$  où  $r$  est l’ordre d’approximation. Cas du cylindre parfaitement conducteur de rayon 2 et de hauteur 4.

pour la polarisation HH et la polarisation VV. On obtient une bonne concordance avec la référence du JINA, comme le montrent les figures 9.18 et 9.19.

Pour calculer cette SER, on s'est imposé un nombre  $M$  de modes tel que :

$$\forall p \geq M \quad \| E_p^{\text{inc}} \|_{L^2(\Gamma)} \leq 10^{-6} \max_m \| E_m^{\text{inc}} \|_{L^2(\Gamma)}$$

La solution est ainsi calculée sur 16 modes. Sur le tableau 9.1, on compte le nombre de degrés de liberté nécessaire pour atteindre une erreur inférieure à 0.5 dB. Pour  $Q_1$ , on a besoin de prendre 10 points par longueur d'onde alors que  $Q_3$  ne nécessite que 5 points par longueur d'onde, on gagne un facteur deux sur ce cas simple. Pour les ordres  $Q_4$  et  $Q_5$ , on est contraint par la géométrie.

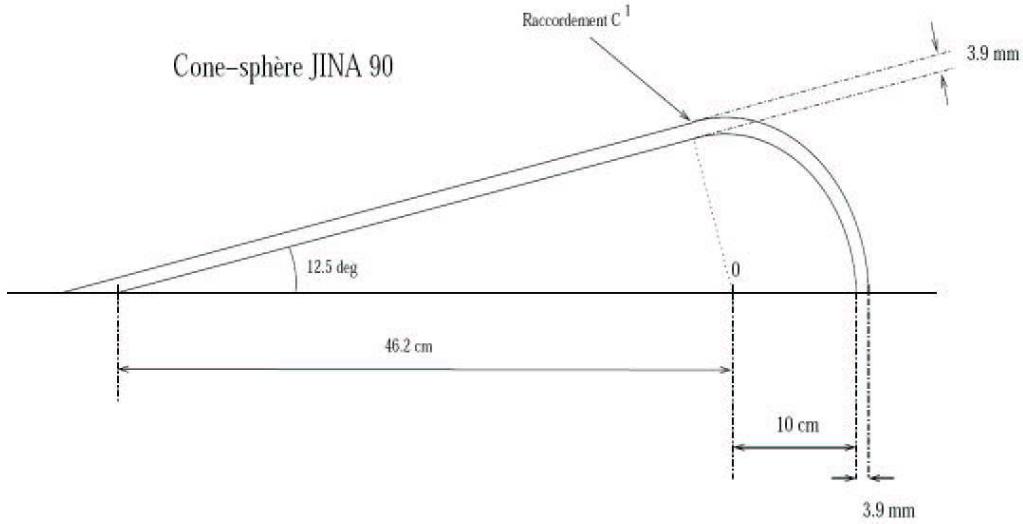


FIG. 9.17 – Paramètres géométriques du cône-sphère.

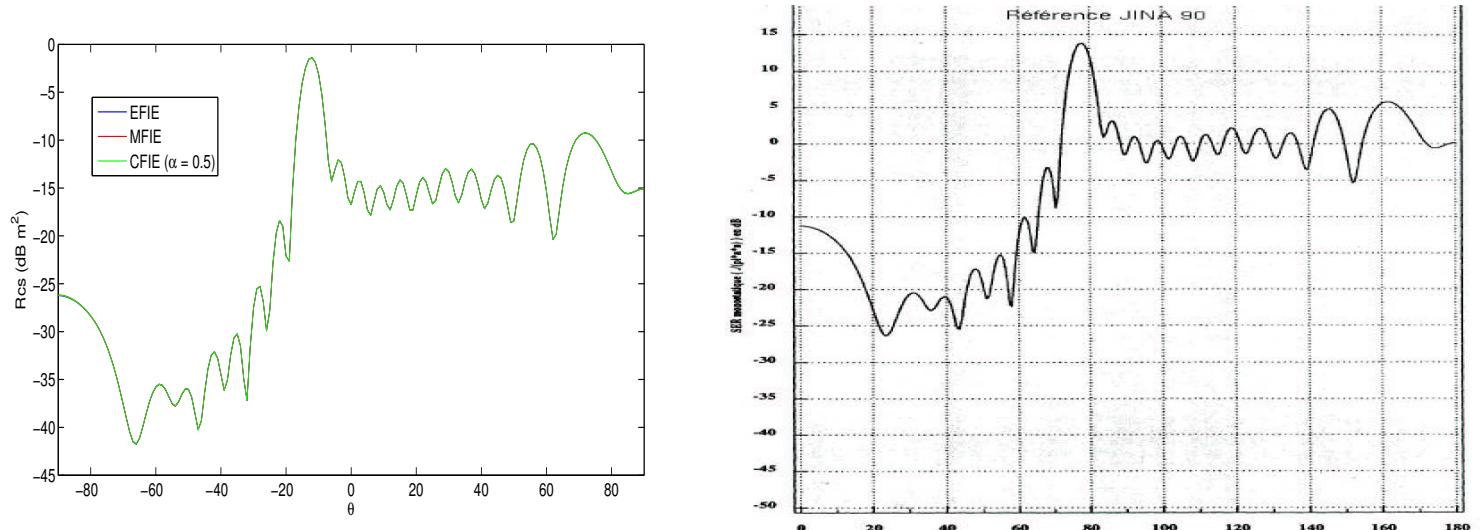


FIG. 9.18 – Ser pour un cône-sphère parfaitement conducteur. A gauche, résultat fourni par notre code, à droite référence du JINA. Polarisation HH

Ordre	$Q_1$	$Q_2$	$Q_3$	$Q_4$	$Q_5$
Ddl	120	66	62	74	72

TAB. 9.1 – Nombre de degrés de liberté nécessaires pour obtenir une SER précise à 0.5 dB (en norme  $L^\infty$ ).

Afin de mieux différencier les différents ordres d'approximation, on multiplie la fréquence par trois, le nombre de modes est pour sa part multiplié par deux (33 au lieu de 16). On obtient la SER monostatique de la figure 9.20. On cherche là aussi pour une précision donnée (toujours 0.5 dB), le nombre de degrés de liberté nécessaire. Les résultats sont synthétisés dans le tableau 9.2. Sur ce cas, on voit un peu plus nettement l'avantage de  $Q_5$  par rapport à  $Q_2$ , mais ce n'est

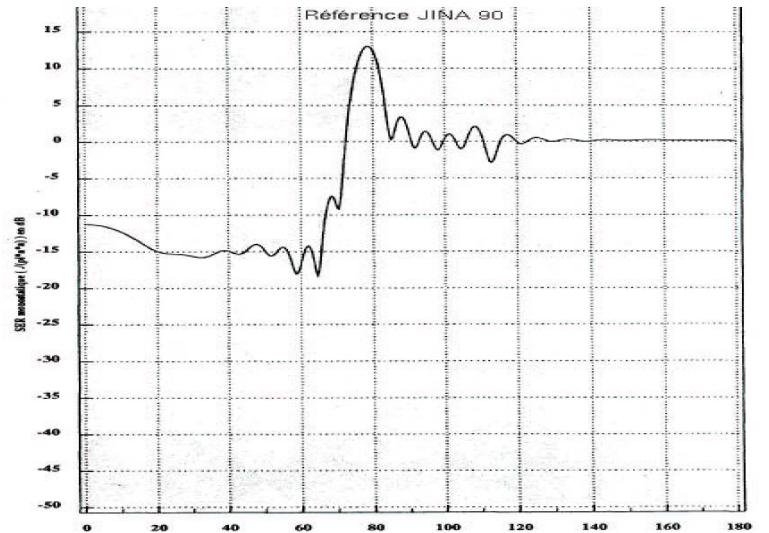
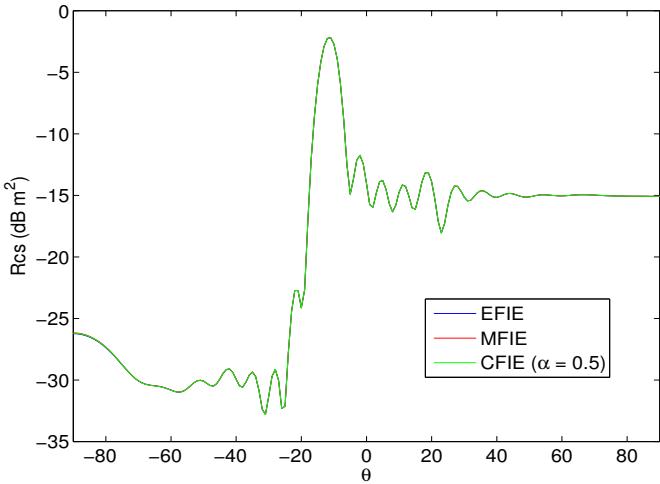


FIG. 9.19 – Ser pour un cône-sphère parfaitement conducteur. A gauche, résultat fourni par notre code, à droite référence du JINA. Polarisation VV

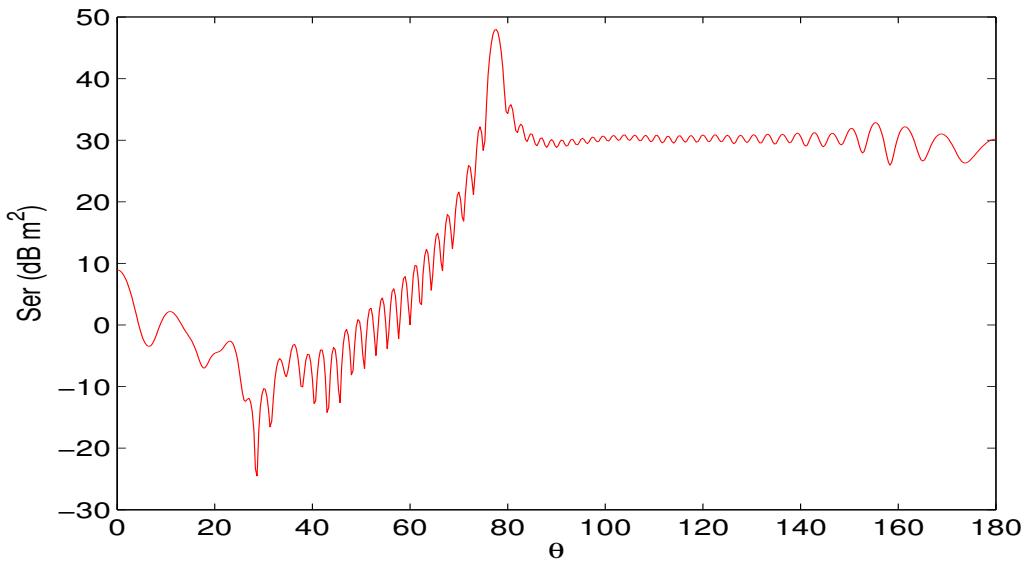


FIG. 9.20 – Ser pour un cône-sphère parfaitement conducteur, fréquence de 8.76 Ghz. Polarisation HH

Ordre	$Q_1$	$Q_2$	$Q_3$	$Q_4$	$Q_5$
Ddl	488	270	266	274	250

TAB. 9.2 – Nombre de degrés de liberté nécessaires pour obtenir une SER précise à 0.5 dB (en norme  $L^\infty$ ).

pas encore très flagrant. Si on veut prendre en défaut  $Q_2$ , soit il faut demander une précision plus grande, soit il faut trouver un cas difficile à résoudre, par exemple un objet présentant une pseudo-cavité.

## 9.4 Couplage avec les éléments finis

### 9.4.1 Définition des opérateurs

Lorsqu'on effectue le couplage avec les éléments finis, on considère les formules de représentation intégrales suivantes [Levillain, 1991], [Simon, 2003] :

$$\begin{aligned} \int_{\Sigma} \mathbf{E}^{inc} \cdot \mathbf{J}^t &= -ik \int_{\Sigma} \int_{\Sigma} G(x, x') \mathbf{J}(x') \cdot \mathbf{J}^t(x) dx dx' + \frac{i}{k} \int_{\Sigma} \int_{\Sigma} G(x, x') \operatorname{div}_{\Sigma}(\mathbf{J}) \operatorname{div}_{\Sigma}(\mathbf{J}^t) dx' dx \\ &\quad + \frac{1}{2} \int_{\Sigma} (\mathbf{n} \times \mathbf{K}) \cdot \mathbf{J}^t(x) dx + \int_{\Sigma} \int_{\Sigma} \mathbf{K}(x') \times \nabla'_x G(x, x') \cdot \mathbf{J}^t(x) dx' dx \\ \int_{\Sigma} \mathbf{H}^{inc} \cdot (\mathbf{J}^t(x) \times \mathbf{n}(x)) dx &= \frac{1}{2} \int_{\Sigma} \mathbf{J}(x) \cdot \mathbf{J}^t(x) dx - \int_{\Sigma} \int_{\Sigma} \mathbf{J}(x') \times \nabla'_x G(x, x') \cdot (\mathbf{J}^t(x) \times \mathbf{n}) dx' dx \\ &\quad - ik \int_{\Sigma} \int_{\Sigma} G(x, x') \mathbf{K}(x') \cdot (\mathbf{J}^t(x) \times \mathbf{n}(x)) dx dx' + \frac{i}{k} \int_{\Sigma} \int_{\Sigma} G(x, x') \operatorname{div}_{\Sigma}(\mathbf{K}(x')) \operatorname{div}_{\Sigma}(\mathbf{J}^t(x) \times \mathbf{n}) dx' dx \end{aligned}$$

La première équation est équivalente à l'EFIE, si on prend  $\mathbf{K} = 0$ , ce qui correspond à la condition de conducteur parfait. La seconde équation est équivalente à la MFIE, si on prend  $\mathbf{K} = 0$ . Notons :

$$\begin{aligned} Z^{EFIE} \mathbf{J} &= -ik \int_{\Sigma} \int_{\Sigma} G(x, x') \mathbf{J} \cdot \mathbf{J}^t dx' dx + \frac{i}{k} \int_{\Sigma} \int_{\Sigma} G(x, x') \operatorname{div}_{\Sigma}(\mathbf{J}(x')) \operatorname{div}_{\Sigma}(\mathbf{J}^t(x)) dx' dx \\ Z^{MFIE} \mathbf{J} &= \frac{1}{2} \int_{\Sigma} \mathbf{J}(x) \cdot \mathbf{J}^t(x) dx + \int_{\Sigma} \int_{\Sigma} (\nabla'_x G(x, x') \times \mathbf{J}(x')) \cdot (\mathbf{J}^t(x) \times \mathbf{n}(x)) dx' dx \end{aligned}$$

Il apparaît, en plus des opérateurs EFIE et MFIE, deux opérateurs :

$$\begin{aligned} Z^{EFIE \text{ croisée}} \mathbf{J} &= -ik \int_{\Sigma} \int_{\Sigma} G(x, x') \mathbf{J} \cdot (\mathbf{J}^t \times \mathbf{n}) dx' dx + \frac{i}{k} \int_{\Sigma} \int_{\Sigma} G(x, x') \operatorname{div}_{\Sigma}(\mathbf{J}(x')) \operatorname{div}_{\Sigma}(\mathbf{J}^t(x) \times \mathbf{n}) dx' dx \\ Z^{MFIE \text{ croisée}} \mathbf{J} &= \frac{1}{2} \int_{\Sigma} (\mathbf{n} \times \mathbf{J}) \cdot \mathbf{J}^t(x) dx + \int_{\Sigma} \int_{\Sigma} \mathbf{J}(x') \times \nabla'_x G(x, x') \cdot \mathbf{J}^t(x) dx' dx \end{aligned}$$

Ces deux opérateurs sont similaires à  $Z^{EFIE}$  et  $Z^{MFIE}$ , il suffit de remplacer la fonction test  $\mathbf{J}^t$  par  $\mathbf{J}^t \times \mathbf{n}$ . Pour une résolution axisymétrique, cette opération revient à intervertir les inconnues  $J_t$  et  $J_b$  avec un signe moins sur l'une d'entre elles. Matriciellement, ça se traduit par l'opération suivante :

$$\begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix} \implies \begin{pmatrix} Z_{21} & Z_{22} \\ -Z_{11} & -Z_{12} \end{pmatrix}$$

On réécrit les deux équations à l'aide des opérateurs définis précédemment :

$$\begin{aligned} F^{EFIE} &= Z^{MFIE \text{ croisée}} K + Z^{EFIE} J \\ F^{MFIE} &= Z^{MFIE} J + Z^{EFIE \text{ croisée}} K \end{aligned} \tag{9.1}$$

Les inconnues  $\mathbf{J}$  sont discrétisées de la même manière que dans le cas parfaitement conducteur. Les inconnues  $\mathbf{K}$  sont imposées par la discréttisation volumique.  $K_t \in H^1(\Gamma)$  est discrétisée par les points de Gauss-Lobatto.  $K_b \in L^2(\Gamma)$  est discrétisée par les points de Gauss. L'opérateur  $Z^{EFIE \text{ croisée}}$  est parfaitement défini car on utilise des fonctions tests appartenant à  $H^1(\Gamma)$ . On a fait le choix de prendre les inconnues  $J_t$   $J_b$  et leurs fonctions tests dans  $H^1$ , justement pour avoir un couplage avec les éléments finis, qui ne pose pas de problème au niveau des espaces fonctionnels.

#### 9.4.2 Cas de la sphère revêtue par un matériau diélectrique

On valide le couplage avec les éléments finis du chapitre 8, sur le cas académique d'une sphère revêtue. Le noyau conducteur est de rayon 4, l'épaisseur du revêtement est de 1. Les indices physiques de ce dernier sont égaux à :

$$\varepsilon = 3 \quad \mu = 2$$

Le courant magnétique solution de ce problème est affiché sur la figure 9.21. On obtient la

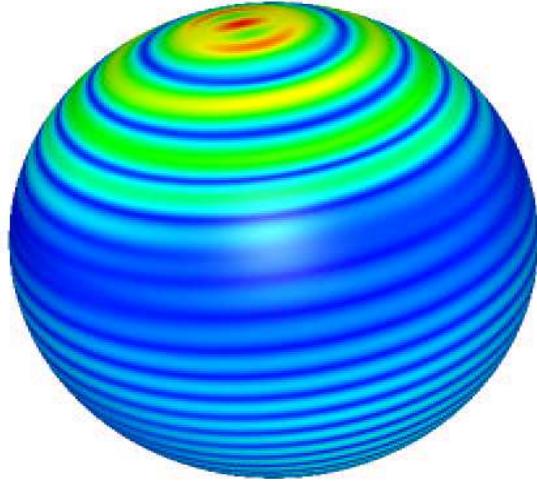


FIG. 9.21 – module de la partie réelle du courant magnétique ( $|Re(J)|$ ) sur la surface de la sphère diélectrique.

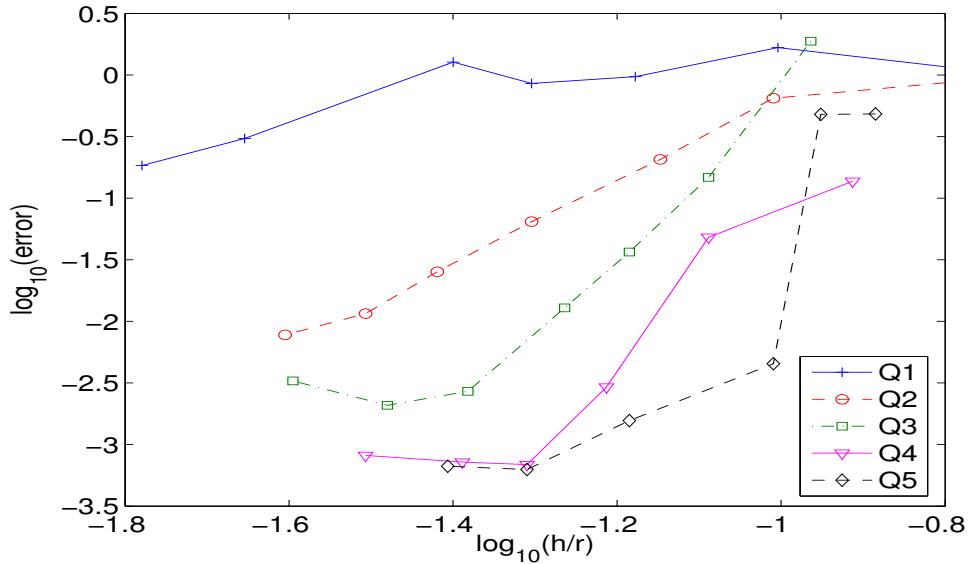


FIG. 9.22 – Evolution de l'erreur  $L^2$  en fonction de  $h/r$  où  $r$  est l'ordre d'approximation. Cas de la sphère de rayon 4, revêtue d'une couche d'épaisseur 1.

convergence de la figure 9.22. Ce cas est intéressant, car la longueur d'onde du diélectrique est

environ 2.5 fois plus petite que la longueur d'onde dans le vide. On voit qu'il est nécessaire de mettre suffisamment de points par longueur d'onde dans le diélectrique. Pour  $Q_5$ , on doit par exemple avoir un maillage avec 10 points par longueur d'onde dans le vide (soit 4 points dans le diélectrique) pour avoir une erreur inférieure à 5 %. Les éléments finis  $Q_1$  sont très dispersifs par rapport aux équations intégrales, ils nous obligent à discréteriser avec 60 points par longueur d'onde dans le vide (25 dans le diélectrique), pour obtenir une erreur inférieure à 20%. Quant au nombre de ddl surfaciques,  $Q_5$  demande 320 degrés de libertés pour atteindre une erreur inférieure à 5 %, alors qu'il en faut 826 pour  $Q_2$ .

Cet exemple nous montre également l'intérêt potentiel de faire des maillages volumiques qui ne s'adaptent pas au maillage de frontière utilisé pour les équations intégrales. Notamment lorsqu'on utilise des méthodes d'ordre 1, le maillage de frontière n'a pas besoin d'être extrêmement fin alors que le maillage volumique doit être fortement surmaillé.

#### 9.4.3 Cas du cylindre revêtu par du diélectrique

On choisit dans cette section d'étudier la diffraction par un cylindre revêtu d'une couche diélectrique. On choisit pour ce cas des indices pas trop méchants :

$$\varepsilon = 2 \quad \mu = 1$$

Le courant magnétique solution de ce problème est représenté sur la figure 9.23. Grace à ce

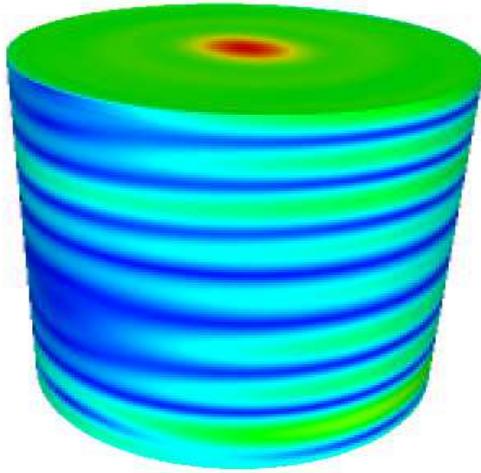


FIG. 9.23 – module de la partie réelle du courant magnétique ( $|Re(J)|$ ) sur la surface du cylindre diélectrique.

choix d'indices, on obtient des courbes de convergence plus propres que dans la sous-section précédente. Le lecteur pourra se délecter de ces courbes sur la figure 9.24.  $Q_1$  donne une convergence d'ordre 2 (on mesure une pente de 1.92), la singularité doit perturber l'ordre de convergence de  $Q_1$  mais pour des pas de maillage plus petits. Pour  $Q_2$ ,  $Q_3$ ,  $Q_4$  et  $Q_5$ , la méthode convergen avec le même ordre, car la géométrie présente une singularité de type arête. La montée en ordre reste avantageuse. Pour atteindre une erreur inférieure à 5 %, il faut 220 degrés de liberté en  $Q_2$  contre 160 degrés de liberté en  $Q_5$ .

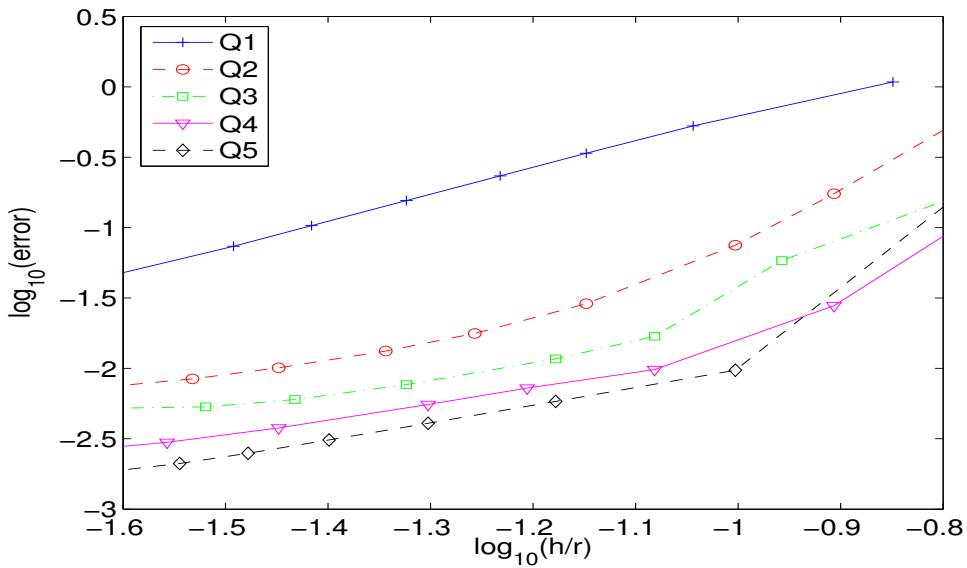


FIG. 9.24 – Evolution de l'erreur  $L^2$  en fonction de  $h/r$  où  $r$  est l'ordre d'approximation. Cas du cylindre, revêtu d'une couche d'épaisseur 1.

#### 9.4.4 Cas du cone-sphère revêtu par du diélectrique

Nous reprenons le cas du JINA de la figure 9.17, afin de valider la méthode sur un cas différent de la sphère. Les indices du revêtement sont pris égaux à :

$$\varepsilon = 15 + 1.8i \quad \mu = 1.7 + 1.7i$$

On calcule sur cette géométrie la SER monostatique pour la même fréquence que dans le cas conducteur (2.92 Ghz). On obtient une bonne concordance avec la référence du JINA (cf. figures 9.25 et 9.26). Si on se fixe d'atteindre une erreur inférieure à 0.5 dB, on obtient le nombre

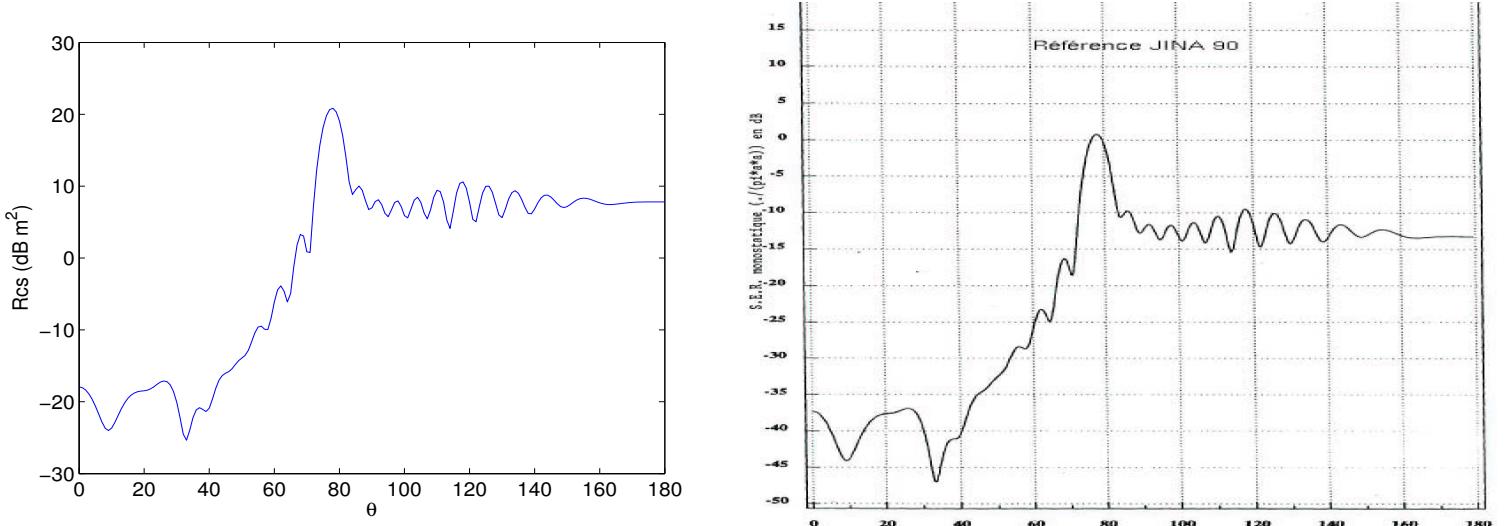


FIG. 9.25 – SER pour un cône-sphère revêtu. A gauche, résultat fourni par notre code, à droite référence du JINA. Polarisation HH

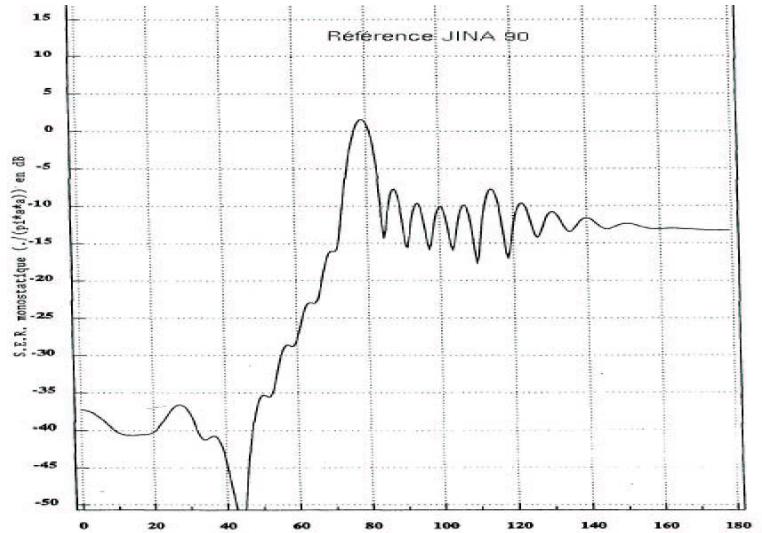
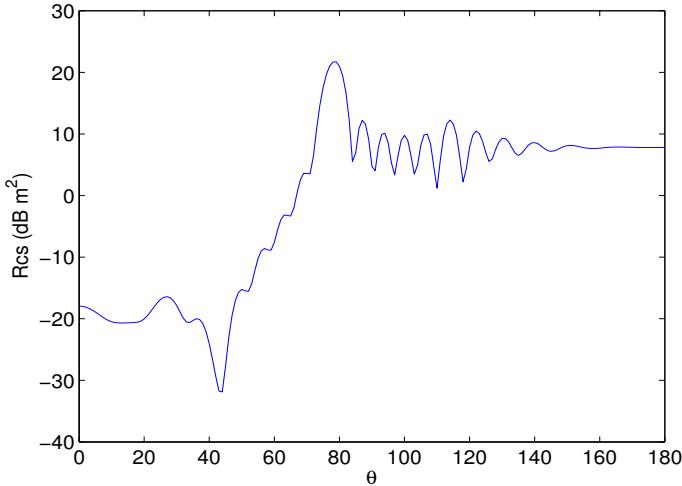


FIG. 9.26 – Ser pour un cône-sphère revêtu. A gauche, résultat fourni par notre code, à droite référence du JINA. Polarisation VV

Ordre	$Q_1$	$Q_2$	$Q_3$	$Q_4$	$Q_5$
ddl	634	258	182	170	162

TAB. 9.3 – Nombre de ddl surfaciques nécessaires pour atteindre 0.5 dB d’erreur, sur le cas du cône-sphère revêtu. On ne compte ici que les ddl pour le courant  $J$ . Il faut multiplier ce nombre par 2, car on a également besoin de  $K$  dans la résolution.

de degrés minimal donné par le tableau 9.3. Pour  $Q_1$ , on a obtenu une erreur de 1.5 dB pour un maillage comprenant 634 degrés de liberté surfaciques. Il nous a semblé inutile de mailler plus finement pour obtenir l’objectif fixé. Comme on l’a déjà signalé,  $Q_1$  converge lentement car les éléments finis volumiques ont de mauvaises propriétés de convergence. Sur cet exemple, on voit qu’il est intéressant de faire au moins du  $Q_3$ .

## 9.5 Conclusion

Dans ce chapitre, nous avons décrit les équations intégrales sur des domaines à symétrie de révolution, pour un objet parfaitement conducteur. Nous avons choisi d’utiliser une formulation CFIE, qui présente l’avantage de ne pas présenter de fréquences de résonance. Nous avons montré comment on menait les calculs de manière efficace lorsque l’on montait en ordre. Le point-clé est une première intégration sur la variable  $\varphi$ , cette intégration est commune à tous les ordres d’intégration, c’est elle qui en pratique domine le temps de calcul des matrices.

Pour évaluer les intégrales singulières, nous avons comparé diverses approches, et fixé notre choix sur une approche utilisant une technique “Gauss Squared”. Nous avons établi des règles de quadrature afin d’évaluer correctement tous les termes de la matrice. L’intégration numérique des singularités se traduit par le fait qu’au-delà d’un certain pas de maillage, la précision obtenue sur la solution stagne. Nous avons ainsi pu vérifier qu’en utilisant nos règles d’intégration, cette stagnation n’apparaissait que pour des pas de maillage trop fins, qui ne sont pas utilisés en pratique.

Nous avons comparé les différents ordres d’approximation sur le cas du cône-sphère parfaitement conducteur. Sur ce cas, on a montré qu’il était intéressant d’utiliser de l’ordre 2, plutôt que de l’ordre 1. Sur ce cas, l’utilisation d’ordres plus élevés ne fait pas gagner beaucoup de

degrés de liberté par rapport à  $Q_2$ .

Finalement, nous avons traité le cas d'objets avec revêtement, en couplant les éléments finis volumiques avec les équations intégrales. Sur le cas du cone-sphère revêtu, nous avons montré que la montée en ordre était très intéressante, car les éléments finis volumiques constituaient le facteur pénalisant de la méthode.



# Conclusion

## Bilan

L'objectif de la thèse était de construire une méthode numérique permettant de résoudre les équations de Maxwell dans des milieux hétérogènes, en régime fréquentiel par une méthode précise et rapide. Le premier point est assuré par l'utilisation de méthodes d'ordre élevé. Le second point est assuré par la mise au point de produits matrice-vecteur rapides et l'utilisation de préconditionneurs efficaces.

Dans la première partie, nous avons développé ces techniques sur le cas simplifié de l'équation de Helmholtz 2-D et 3-D. Nous avons montré qu'en 2-D, un solveur direct était suffisant pour la plupart des problèmes rencontrés. En 3-D, on a recours à un solveur itératif, le BICGCR préconditionné, soit par une factorisation incomplète avec amortissement, soit par une itération multigrille.

En ce qui concerne les équations de Maxwell, nous avons montré que l'utilisation de la seconde famille de Nédélec sur les hexaèdres était néfaste à cause d'ondes parasites. La formulation Galerkin discontinue, adaptée au régime temporel, n'est pas très avantageuse en régime fréquentiel. Nous avons donc privilégié la première famille de Nédélec sur les hexaèdres. On a décrit comment on pouvait réaliser un produit matrice vecteur rapide en utilisant cette discrétisation. Le solveur itératif préconisé est le BICGCR préconditionné par une factorisation incomplète avec amortissement. Cette factorisation incomplète est calculée sur un maillage de bas ordre, obtenu en subdivisant le maillage initial. On obtient ainsi un stockage pas très élevé, mais ce stockage reste pénalisant pour les problèmes de grande taille. Toutefois, il est bien plus raisonnable que celui requis par les tétraèdres d'ordre élevé.

Finalement, nous avons étudié la discrétisation des équations de Maxwell sur des domaines axisymétriques. On utilise une méthode éléments finis couplée avec une équation intégrale sur la frontière. Pour la méthode éléments finis, nous avons fait le choix d'une formulation mixte, qui évite les problèmes de singularité sur l'axe. Pour la méthode intégrale, nous avons utilisé des éléments finis de frontière d'ordre élevé. Nous avons ainsi obtenu une méthode qui nécessite moins de degrés de liberté lorsqu'on monte en ordre, pour atteindre une précision donnée.

## Perspectives

Pour l'équation de Helmholtz, nous pensons que les préconditionneurs utilisés sont efficaces, mais ils peuvent être améliorés. Notamment, certaines voies, comme le multigrille algébrique, ou l'utilisation de la FFT sur des domaines tensorisés, gagneraient à être étudiées pour les éléments finis hexaédriques d'ordre élevé.

Pour les équations de Maxwell 3-D, nous avons aussi des problèmes d'efficacité des préconditionneurs. Un autre point à étudier serait le couplage éléments finis - équations intégrales. Une première approche est d'utiliser une discrétisation des équations intégrales, d'ordre élevé. La difficulté est en pratique l'intégration des singularités. Pour les méthodes d'ordre un, les intégrales sin-

gulières sont évaluées analytiquement. Lorsqu'on utilise de l'ordre élevé, cette technique n'est pas praticable, il faut avoir recours à des intégrations numériques. Le couplage avec les éléments finis serait du même type que celui proposé en axisymétrique.

La deuxième approche serait de coupler des éléments finis d'ordre élevé avec les équations intégrales d'ordre un. On pourrait ainsi avoir une non-conformité entre le maillage de surface et le maillage de volume.

En ce qui concerne, les équations de Maxwell axi-symétriques, nous pensons qu'une piste à explorer serait l'utilisation de maillages non-conformes entre le maillage surfacique utilisé par les équations intégrales et le maillage volumique éléments finis. On pourrait ainsi gagner des degrés de liberté pour représenter la solution sur la surface de l'objet, sans avoir à rajouter une couche intermédiaire dans le maillage volumique. Une autre perspective prometteuse, est de coupler une résolution axi-symétrique avec une résolution 3-D. On pourrait ainsi étudier la diffraction d'objets globalement axi-symétriques, comportant des défauts 3-D.

## Annexe A

# Solutions analytiques des problèmes de diffraction d'ondes planes par une sphère

*Nous rappelons dans cette annexe les expressions utilisées pour calculer les solutions analytiques des problèmes de diffraction par un disque en 2-D, et une sphère en 3-D. La première section rappelle les notations utilisées pour les fonctions spéciales. La deuxième section donne les expression pour l'équation de Helmholtz, avec condition de Sommerfeld exacte ou condition absorbante d'ordre 1. La troisième et dernière section construit les solutions pour les équations de Maxwell. On pourra utiliser indifféremment une condition de Silver-Müller à distance infinie, ou à distance finie.*

### Sommaire

---

<b>A.1</b>	<b>Notations des fonctions spéciales</b>	<b>236</b>
<b>A.2</b>	<b>Solutions analytiques pour l'équation de Helmholtz</b>	<b>237</b>
A.2.1	Diffraction par un disque	237
A.2.2	Diffraction par une sphère	239
<b>A.3</b>	<b>Solutions analytiques pour les équations de Maxwell</b>	<b>240</b>
A.3.1	Développement d'un champ en harmoniques sphériques	240
A.3.2	Potentiels de Debye pour les équations de Maxwell	241
A.3.3	Traces des champs tangentiels associés aux potentiels de Debye	243
A.3.4	Décomposition d'une onde plane dans les potentiels de Debye	244
A.3.5	Diffraction par une sphère parfaitement conductrice	245
A.3.6	Expression de E et H dans tout l'espace	247
A.3.7	Diffraction d'une onde plane par une sphère diélectrique	249
A.3.8	Prise en compte de la condition de Silver-Müller	251

---

## A.1 Notations des fonctions spéciales

On introduit les fonctions spéciales, qui sont à la base de tous les développements qui suivront. Pour une meilleure compréhension, nous renvoyons lecteur à l'ouvrage [Lebedev *et al.*, 1972].

- $J_{n+\frac{1}{2}}(x)$ , la fonction de Bessel de première espèce d'ordre  $n + \frac{1}{2}$ .
- $j_n(x)$ , la fonction de Bessel sphérique de première espèce d'ordre  $n$ .
- $\psi_n(x)$ , la fonction de Riccati-Bessel de première espèce d'ordre  $n$ .
- $Y_{n+\frac{1}{2}}(x)$ , la fonction de Bessel de deuxième espèce d'ordre  $n + \frac{1}{2}$ .
- $y_n(x)$ , la fonction de Bessel sphérique de deuxième espèce d'ordre  $n$ .
- $\zeta_n(x)$ , la fonction de Riccati-Bessel de deuxième espèce d'ordre  $n$ .
- $H_{n+\frac{1}{2}}^{(1)}(x)$ , la fonction de Hankel de première espèce d'ordre  $n + \frac{1}{2}$ .
- $h_n^{(1)}(x)$ , la fonction de Hankel sphérique de première espèce d'ordre  $n$ .
- $\xi_n^{(1)}(x)$ , la fonction de Hankel de première espèce d'ordre  $n$ .
- $H_{n+\frac{1}{2}}^{(2)}(x)$ , la fonction de Hankel de deuxième espèce d'ordre  $n + \frac{1}{2}$ .
- $h_n^{(2)}(x)$ , la fonction de Hankel sphérique de deuxième espèce d'ordre  $n$ .
- $\xi_n^{(2)}(x)$ , la fonction de Hankel de deuxième espèce d'ordre  $n$ .

Ces fonctions vérifient les relations mutuelles :

$$\begin{cases} j_n(x) = \sqrt{\frac{\pi}{2x}} J_{n+\frac{1}{2}}(x), & \psi_n(x) = x j_n(x), \\ y_n(x) = \sqrt{\frac{\pi}{2x}} Y_{n+\frac{1}{2}}(x), & \zeta_n(x) = x y_n(x), \\ h_n^{(1)}(x) = \sqrt{\frac{\pi}{2x}} H_{n+\frac{1}{2}}^{(1)}(x) & \xi_n^{(1)}(x) = x h_n^{(1)}(x). \end{cases}$$

et d'autres relations classiques :

$$\begin{cases} J_{\ell-1}(x) + J_{\ell+1}(x) = \frac{2\ell}{x} J_\ell(x), & Y_{\ell-1}(x) + Y_{\ell+1}(x) = \frac{2\ell}{x} Y_\ell(x) \\ J_{\ell-1}(x) - J_{\ell+1}(x) = 2J'_\ell(x), & Y_{\ell-1}(x) - Y_{\ell+1}(x) = 2Y'_\ell(x), \end{cases}$$

Nous en déduisons que :

$$\begin{cases} \psi_{n-1}(x) + \psi_{n+1}(x) = \frac{2n+1}{x} \psi_n(x) \\ \xi_{n-1}^{(1)}(x) + \xi_{n+1}^{(1)}(x) = \frac{2n+1}{x} \xi_n^{(1)}(x) \\ \psi'_n(x) = \frac{x}{2n+1} ((n+1)j_{n-1}(x) - nj_{n+1}(x)) \\ \xi_n^{(1)'}(x) = \frac{x}{2n+1} \left( (n+1)h_{n-1}^{(1)}(x) - nh_{n+1}^{(1)}(x) \right). \end{cases}$$

Nous rappelons à tout hasard les identités du wronskien :

$$\psi_n(x) \zeta'_n(x) - \psi'_n(x) \zeta_n(x) = 1, \quad (\text{A.1})$$

$$\psi_n(x) \xi_n^{(1)'}(x) - \psi'_n(x) \xi_n^{(1)}(x) = i, \quad (\text{A.2})$$

Ces deux relations viennent de la propriété bien connue :

$$J_\ell(x) Y'_\ell(x) - J'_\ell(x) Y_\ell(x) = \frac{2}{\pi x}.$$

## A.2 Solutions analytiques pour l'équation de Helmholtz

### A.2.1 Diffraction par un disque

Lorsque la géométrie est indépendante de  $\theta$ , on peut décomposer toute solution de l'équation de Helmholtz sous la forme :

$$u = \sum_{n=-\infty}^{\infty} i^n (\alpha_n H_n^{(1)}(kr) + \beta_n H_n^{(2)}(kr)) e^{in\theta}$$

$\alpha_n$  et  $\beta_n$  sont les coefficients de la décomposition modale de  $u$ .

L'onde plane peut se développer en fonctions de Bessel (développement de Jacobi-Anger) :

$$u^{\text{incident}} = e^{ikr \cos \theta} = \sum_{n=-\infty}^{+\infty} i^n J_n(kr) e^{in\theta} \quad (\text{A.3})$$

On développe le champ total en rajoutant à l'onde plane les fonctions de Hankel de première espèce (qui vérifient la condition de Sommerfeld) :

$$u(r, \theta) = \sum_{n=-\infty}^{+\infty} i^n [J_n(kr) + \alpha_n H_n^{(1)}(kr)] e^{in\theta} \quad (\text{A.4})$$

Ainsi, le champ diffracté vérifie la condition de Sommerfeld.

#### Condition de Dirichlet

Le cas modèle de la diffraction par un disque de rayon  $a$  s'écrit :

$$\begin{cases} -k^2 u - \Delta u = 0 & \in \Omega \\ u = 0 & \text{pour } r = a \\ & +\text{Condition de Sommerfeld sur } u - u^{\text{incident}} \end{cases} \quad (\text{A.5})$$

Le champ total doit vérifier la condition aux limites pour  $r = a$  :

$$u(a) = 0$$

ce qui donne l'équation suivante :

$$J_n(ka) + \alpha_n H_n^{(1)}(ka) = 0$$

On en déduit l'expression de  $\alpha_n$  :

$$\alpha_n = -\frac{J_n(ka)}{H_n^{(1)}(ka)}$$

Le champ total  $u$  est recomposé en utilisant la formule (A.4).

#### Condition de Neumann

Lorsqu'on impose la condition de Neumann, on obtient l'équation :

$$J'_n(ka) + \alpha_n H_n^{(1)'}(ka) = 0$$

On en déduit les coefficients  $\alpha_n$  :

$$\alpha_n = -\frac{J'_n(ka)}{H_n^{(1)'}(ka)}$$

## Disque diélectrique

Le cas modèle de la diffraction par un disque diélectrique, de rayon  $a$  et d'indices  $\rho, \mu$ , s'écrit :

$$\left\{ \begin{array}{l} -\frac{\rho}{\mu} k^2 u - \Delta u = 0 \quad r \leq a \\ -k^2 u - \Delta u = 0 \quad r \geq a \end{array} \right. \quad (\text{A.6})$$

+Condition de Sommerfeld sur  $u - u^{\text{incident}}$

On note le nombre d'onde du disque intérieur  $k_i$ , il vaut :

$$k_i = k \sqrt{\frac{\rho}{\mu}}$$

La solution intérieure est développée en fonctions de Bessel (avec pour nombre d'onde  $k_i$ )

$$u(r, \theta) = \sum_{n=-\infty}^{+\infty} i^n \beta_n J_n(k_i r) e^{in\theta} \quad r \leq a \quad (\text{A.7})$$

Les conditions de transmission sont

$$u(r, \theta) \text{ et } \mu \frac{\partial u}{\partial r}(r, \theta) \text{ continues en } r = a$$

C'est bien  $\mu \frac{\partial u}{\partial r}$  qui est continue, car l'équation de Helmholtz dans tout le domaine est

$$-\rho k^2 u - \operatorname{div}(\mu \nabla u) = 0$$

On obtient alors le système 2x2

$$\begin{aligned} J_n(ka) + \alpha_n H_n^{(1)}(ka) &= \beta_n J_n(k_i a) \\ k J'_n(ka) + \alpha_n k H_n^{(1)'}(ka) &= \mu \beta_n k_i J'_n(k_i a) \end{aligned}$$

soit

$$\left\{ \begin{array}{l} -\alpha_n H_n^{(1)}(ka) + \beta_n J_n(k_i a) = J_n(ka) \\ -\alpha_n k H_n^{(1)'}(ka) + \mu \beta_n k_i J'_n(k_i a) = k J'_n(ka) \end{array} \right.$$

Les solutions sont :

$$\alpha_n = \frac{\mu k_i J'_n(k_i a) J_n(ka) - k J_n(k_i a) J'_n(ka)}{k H_n^{(1)'}(ka) J_n(k_i a) - \mu k_i J'_n(k_i a) H_n^{(1)}(ka)} \quad (\text{A.8})$$

$$\beta_n = \frac{k H_n^{(1)'}(ka) J_n(ka) - k H_n^{(1)}(ka) J'_n(ka)}{k H_n^{(1)'}(ka) J_n(k_i a) - \mu k_i J'_n(k_i a) H_n^{(1)}(ka)} \quad (\text{A.9})$$

### Prise en compte de la condition absorbante d'ordre 1

Dans nos expériences numériques, on a souvent approché la condition exacte de Sommerfeld, par une condition absorbante d'ordre sur un cercle de rayon  $b$  :

$$\frac{\partial(u - u^{\text{incident}})}{\partial r} - i k (u - u^{\text{incident}}) = 0 \quad \text{pour } r = b$$

On remplace le développement (A.4), par celui-ci :

$$u(r, \theta) = \sum_{n=-\infty}^{+\infty} i^n \left[ J_n(kr) + \alpha_n H_n^{(1)}(kr) + \delta_n H_n^{(2)}(kr) \right] e^{in\theta} \quad (\text{A.10})$$

La condition aux limites pour le disque de rayon  $b$  nous fournit l'équation :

$$\alpha_n [k H_n^{(1)'}(kb) - ik H_n^{(1)}(kb)] + \delta_n [k H_n^{(2)'}(kb) - ik H_n^{(2)}(kb)]$$

On en déduit l'expression de  $\delta_n$  :

$$\delta_n = -\frac{H_n^{(1)'}(kb) - i H_n^{(1)}(kb)}{H_n^{(2)'}(kb) - i H_n^{(2)}(kb)} \alpha_n$$

On peut réécrire  $u$  sous une forme plus "agréable" :

$$u(r, \theta) = \sum_{n=-\infty}^{+\infty} i^n \left[ J_n(kr) + \alpha_n \tilde{H}_n^{(1)}(kr) \right] e^{in\theta} \quad (\text{A.11})$$

où  $\tilde{H}_n^{(1)}$  est une fonction de Hankel de première espèce perturbée :

$$\tilde{H}_n^{(1)}(kr) = H_n^{(1)}(kr) - \left( \frac{H_n^{(1)'}(kb) - i H_n^{(1)}(kb)}{H_n^{(2)'}(kb) - i H_n^{(2)}(kb)} \right) H_n^{(2)}(kr)$$

Toutes les expressions calculées pour une condition de Sommerfeld exacte sont donc valables, en remplaçant  $H_n^{(1)}$  par  $\tilde{H}_n^{(1)}$ . De cette manière, la prise en compte de la condition absorbante d'ordre 1 est relativement simple à mettre en œuvre.

### A.2.2 Diffraction par une sphère

On se place en coordonnées sphériques

$$\begin{aligned} x &= r \sin \theta \cos \varphi \\ y &= r \sin \theta \sin \varphi \\ z &= r \cos \theta \end{aligned}$$

Le développement de Jacobi-Anger s'écrit

$$e^{ikr \cos \theta} = \sum_{n=0}^{+\infty} i^n (2n+1) j_n(kr) P_n(\cos \theta) \quad (\text{A.12})$$

avec  $j_n(kr)$  fonction de Bessel sphérique de première espèce. Comme l'onde incidente est invariante suivant  $\varphi$ , la solution est également invariante par rapport à cette variable. Le champ total est recherché sous la forme :

$$u(r, \theta, \phi) = \sum_{n=0}^{+\infty} i^n (2n+1) \left[ j_n(kr) + \alpha_n h_n^{(1)}(kr) \right] P_n(\cos \theta) \quad r \geq a \quad (\text{A.13})$$

Ce développement est très proche de celui qu'on a obtenu en 2-D. Pour passer du 2-D au 3-D, il suffit donc de remplacer  $H_n^{(1)}$ ,  $J_n$ ,  $H_n^{(2)}$  par leurs équivalents sphériques  $h_n^{(1)}$ ,  $j_n$ ,  $h_n^{(2)}$ .

## Condition de Dirichlet

$$\alpha_n = -\frac{j_n(k a)}{h_n^{(1)}(k a)}$$

## Condition de Neumann

$$\alpha_n = -\frac{j'_n(k a)}{h_n^{(1)'}(k a)}$$

## Sphère diélectrique

$$\alpha_n = \frac{\mu k_i j'_n(k_i a) j_n(k a) - k j_n(k_i a) j'_n(k a)}{k h_n^{(1)'}(k a) j_n(k_i a) - \mu k_i j'_n(k_i a) h_n^{(1)}(k a)} \quad (\text{A.14})$$

$$\beta_n = \frac{k h_n^{(1)'}(k a) j_n(k a) - k h_n^{(1)}(k a) j'_n(k a)}{k h_n^{(1)'}(k a) j_n(k_i a) - \mu k_i j'_n(k_i a) h_n^{(1)}(k a)} \quad (\text{A.15})$$

## Prise en compte de la condition de Sommerfeld

On remplace  $h_n^{(1)}$  par  $\tilde{h}_n^{(1)}$  :

$$\tilde{h}_n^{(1)}(k r) = h_n^{(1)}(k r) - \left( \frac{h_n^{(1)'}(k b) - i h_n(1)(k b)}{h_n^{(2)'}(k b) - i h_n(2)(k b)} \right) h_n^{(2)}(k r)$$

## A.3 Solutions analytiques pour les équations de Maxwell

Les équations de Maxwell sont différentes de l'équation de Helmholtz qu'à partir de la dimension 3. On ne donnera ici que les expressions des solutions analytiques sur des sphères.

### A.3.1 Développement d'un champ en harmoniques sphériques

Soit  $S_a^2$  une sphère de rayon  $a$ . Dans la suite, nous noterons  $\hat{r}$  la normale unitaire sortante de la sphère. Nous notons  $TL^2(S_a^2)$  l'ensemble des champs tangentiels de carré intégrable :

$$TL^2(S_a^2) = \{J(a\hat{r}) \in L^2(S_a^2)^3, J(a\hat{r}).\hat{r} = 0\}.$$

Cet espace vectoriel peut être générée par un ensemble de fonctions de bases : les harmoniques sphériques. Nous détaillons leur construction ci-après.

Si  $P_n(x)$  est le polynôme de Legendre d'ordre  $n$ ,  $P_n^m(x)$  est la fonction de Legendre associée :

$$P_n^m(x) = (1 - x^2)^{\frac{m}{2}} \frac{d^m P_n(x)}{dx^m}, m \geq 0,$$

Nous définissons :

$$\begin{cases} \tilde{Y}_n^m(\hat{r}) = P_n^{|m|}(\cos \theta) e^{im\varphi}, & n \geq 0, |m| \leq n \\ d_{n,m} = \frac{1}{4\pi} \frac{(n - |m|)! (2n + 1)}{(n + |m|)! n(n + 1)}, & \end{cases} \quad (\text{A.16})$$

où  $(\theta, \varphi)$  sont les angles usuels en coordonnées sphériques  $\tilde{Y}_n^m(\hat{r})$  est connue comme une harmonique sphérique. Nous introduisons :

$$\begin{cases} u_{n,m}^{(+)}(a\hat{r}) = d_{n,m}^{1/2} \vec{\nabla}_{S_a^2} \tilde{Y}_n^m(\hat{r}) \\ u_{n,m}^{(-)}(a\hat{r}) = d_{n,m}^{1/2} \hat{r} \wedge \vec{\nabla}_{S_a^2} \tilde{Y}_n^m(\hat{r}), \end{cases} \quad (\text{A.17})$$

où

$$\begin{cases} \vec{\nabla}_{S_a^2} f(r, \theta, \varphi) = \frac{1}{r} \left( \frac{\partial f(r, \theta, \varphi)}{\partial \theta} \hat{\theta} + \frac{1}{\sin \theta} \frac{\partial f(r, \theta, \varphi)}{\partial \varphi} \hat{\varphi} \right) \\ \hat{r} \wedge \vec{\nabla}_{S_a^2} f(r, \theta, \varphi) = \frac{1}{r} \left( \frac{\partial f(r, \theta, \varphi)}{\partial \theta} \hat{\varphi} - \frac{1}{\sin \theta} \frac{\partial f(r, \theta, \varphi)}{\partial \varphi} \hat{\theta} \right), \end{cases} \quad (\text{A.18})$$

et

$$\hat{r} = \begin{bmatrix} \sin \theta \cos \varphi \\ \sin \theta \sin \varphi \\ \cos \theta \end{bmatrix}, \quad \hat{\theta} = \begin{bmatrix} \cos \theta \cos \varphi \\ \cos \theta \sin \varphi \\ -\sin \theta \end{bmatrix}, \quad \hat{\varphi} = \begin{bmatrix} -\sin \varphi \\ \cos \varphi \\ 0 \end{bmatrix}.$$

Tout champ dans l'espace vectoriel  $L^2(S_a^2)$  peut être décomposé sous la forme

$$J(a\hat{r}) = \sum_{n=1}^{\infty} \sum_{m=-n}^{+n} \sum_{\varepsilon=\pm} J_{mn}^{(\varepsilon)} u_{n,m}^{(\varepsilon)}(a\hat{r}). \quad (\text{A.19})$$

Nous avons

$$\|J(a\hat{r})\|_{TL^2(S_a^2)}^2 = \sum_{n=1}^{\infty} \sum_{m=-n}^{+n} \sum_{\varepsilon=\pm} |J_{mn}^{(\varepsilon)}|^2. \quad (\text{A.20})$$

On peut introduire une base orthormée :

$$\begin{cases} u_{n,m}^{(\varepsilon,c)}(a\hat{r}) = \frac{1}{\sqrt{2}} (u_{n,m}^{(\varepsilon)}(a\hat{r}) + u_{n,-m}^{(\varepsilon)}(a\hat{r})) \\ u_{n,m}^{(\varepsilon,s)}(a\hat{r}) = \frac{-i}{\sqrt{2}} (u_{n,m}^{(\varepsilon)}(a\hat{r}) - u_{n,-m}^{(\varepsilon)}(a\hat{r})) \end{cases} \quad (\text{A.21})$$

Nous avons également :

$$J(a\hat{r}) = \sum_{n=1}^{\infty} \sum_{m=0}^{+n} \sum_{\varepsilon=\pm} J_{mn}^{(\varepsilon,c)} u_{n,m}^{(\varepsilon,c)}(a\hat{r}) + \sum_{n=1}^{\infty} \sum_{m=1}^{+n} \sum_{\varepsilon=\pm} J_{mn}^{(\varepsilon,s)} u_{n,m}^{(\varepsilon,s)}(a\hat{r}) \quad (\text{A.22})$$

et :

$$\|J(a\hat{r})\|_{TL^2(S_a^2)}^2 = \sum_{n=1}^{\infty} \sum_{m=0}^n \sum_{\varepsilon=\pm} |J_{mn}^{(\varepsilon,c)}|^2 + \sum_{n=1}^{\infty} \sum_{m=1}^n \sum_{\varepsilon=\pm} |J_{mn}^{(\varepsilon,s)}|^2. \quad (\text{A.23})$$

### A.3.2 Potentiels de Debye pour les équations de Maxwell

Soit  $w(x)$  une solution de l'équation de Helmholtz :

$$k^2 w(x) + \Delta w(x) = 0.$$

A partir de ce potentiel  $w(x)$ , il est possible de construire une solution des équations de Maxwell. Nous commençons par l'identité triviale :

$$\operatorname{curl} \left( \vec{x}(k^2 w(x) + \Delta w(x)) + \vec{\nabla} w(x) \right) = 0,$$

( $\operatorname{curl} \vec{\nabla} w = 0$  est évident)

$$\operatorname{curl} \left( k^2 \vec{x}w(x) + \vec{\Delta}(\vec{x}w(x)) \right) = 0,$$

Comme  $\vec{\Delta} = \vec{\nabla} \operatorname{div} - \operatorname{curl} \operatorname{curl}$ ,

$$\operatorname{curl} \left( k^2 \vec{x}w(x) - \operatorname{curl} \operatorname{curl}(\vec{x}w(x)) \right) = 0. \quad (\text{A.24})$$

Soit

$$E(x) = \operatorname{curl}(\vec{x}w(x)), \quad H(x) = -\frac{i}{k Z_0} \operatorname{curl} \operatorname{curl}(\vec{x}w(x)),$$

Nous avons par construction

$$\operatorname{curl} E(x) - ikZ_0 H(x) = 0, \quad (\text{A.25})$$

L'équation (A.24) fournit immédiatement

$$\operatorname{curl} H(x) + ikZ_0^{-1} E(x) = 0, \quad (\text{A.26})$$

Nous pouvons faire le lien avec les équations de Maxwell standard :

$$\begin{aligned} -i\omega \epsilon_r E(x) - \operatorname{curl} H(x) &= 0 \\ -i\omega \mu_r H(x) + \operatorname{curl} E(x) &= 0 \end{aligned} \quad (\text{A.27})$$

On en déduit les relations vérifiées par  $Z_0$  et  $k$

$$\frac{ik}{Z_0} = i\omega \epsilon_r$$

$$ikZ_0 = i\omega \mu_r$$

Nous avons l'expression suivante de  $Z_0$   $k$  en fonction de  $\epsilon_r$  et  $\mu_r$

$$\begin{aligned} Z_0 &= \sqrt{\frac{\mu_r}{\epsilon_r}} \\ k &= \sqrt{\mu_r \epsilon_r} \omega \end{aligned} \quad (\text{A.28})$$

$Z_0$  est souvent appelée l'impédance du matériau alors que  $k$  est le nombre d'onde.

Soit  $v(x)$  une autre solution de l'équation de Helmholtz. En échangeant le rôle de  $E$  et  $H$ , on peut montrer que :

$$E(x) = \frac{i}{k} \operatorname{curl} \operatorname{curl}(\vec{x}v(x)), \quad H(x) = \frac{1}{Z_0} \operatorname{curl}(\vec{x}v(x)),$$

est aussi une solution des deux équations (A.25) et (A.26). En ajoutant les deux champs électromagnétiques, on obtient que :

$$\begin{cases} E(x) = \operatorname{curl}(\vec{x}w(x)) + \frac{i}{k} \operatorname{curl} \operatorname{curl}(\vec{x}v(x)), \\ H(x) = -\frac{i}{kZ_0} \operatorname{curl} \operatorname{curl}(\vec{x}w(x)) + \frac{1}{Z_0} \operatorname{curl}(\vec{x}v(x)), \end{cases} \quad (\text{A.29})$$

est une solution particulière des équations de Maxwell (A.25)-(A.26). Les fonctions  $v(x)$  et  $w(x)$  sont les potentiels de Debye associés au champ électromagnétique. Ces potentiels sont des outils

puissants pour résoudre des problèmes de diffractions dans des géométries sphériques. Nous pouvons décomposer ces potentiels sur les harmoniques sphériques

$$\begin{bmatrix} v(x) \\ w(x) \end{bmatrix} = \begin{bmatrix} v(r, \theta, \varphi) \\ w(r, \theta, \varphi) \end{bmatrix} == \sum_{n=1}^{\infty} \sum_{m=-n}^n \begin{bmatrix} v_{n,m} \\ w_{n,m} \end{bmatrix} \chi_n(kr) \tilde{Y}_{n,m}(\theta, \varphi), \quad (\text{A.30})$$

où  $v_{n,m}$  et  $w_{n,m}$  sont une suite de nombres complexes, et  $\chi_n(kr)$  tient lieu pour une des trois fonctions sphériques  $j_n(kr)$ ,  $y_n(kr)$  ou  $h_n^{(1)}(kr)$  (de telle sorte que  $\chi_n(kr) \tilde{Y}_{n,m}(\theta, \varphi)$  vérifie l'équation de Helmholtz). Résoudre un problème de diffraction sur une géométrie sphérique revient à trouver les expressions analytiques pour les coefficients  $v_{n,m}$  et  $w_{n,m}$ .

### A.3.3 Traces des champs tangentiels associés aux potentiels de Debye

Soit  $(E(x), H(x))$  un champ électromagnétique associé aux potentiels  $v(x)$  et  $w(x)$ . Nous assumons que  $v(x)$  et  $w(x)$  sont définis à l'extérieur ou à l'intérieur du domaine délimité par  $S_a^2$ , la sphère de rayon  $a$ . Nous voulons obtenir l'expression des champs tangentiels :

$$E_t(a\hat{r}) = \lim_{x \rightarrow a\hat{r}} \hat{r} \wedge (E(x) \wedge \hat{r}), \quad H_t(a\hat{r}) = \lim_{x \rightarrow a\hat{r}} \hat{r} \wedge (H(x) \wedge \hat{r}).$$

Nous utilisons les coordonnées sphériques  $(r, \theta, \varphi)$ , et les vecteurs unitaires associés  $\hat{r}, \hat{\theta}, \hat{\varphi}$ . Les expressions du rotationnel et du rot-rot d'un champ dirigé dans la direction radiale sont données par :

$$\operatorname{curl}(v(x)\vec{x}) = \operatorname{curl}(rv(r, \theta, \varphi)\hat{r}) = \frac{1}{r} \left( \frac{1}{\sin \theta} \frac{\partial rv(r, \theta, \varphi)}{\partial \varphi} \hat{\theta} - \frac{\partial rv(r, \theta, \varphi)}{\partial \theta} \hat{\varphi} \right)$$

et

$$\operatorname{curl} \operatorname{curl}(v(x)\vec{x}) = \operatorname{curl} \operatorname{curl}(rv(r, \theta, \varphi)\hat{r})$$

$$\begin{aligned} &= -\frac{1}{r^2} \left( \frac{\partial^2 rv(r, \theta, \varphi)}{\partial \theta^2} + \frac{1}{\tan \theta} \frac{\partial rv(r, \theta, \varphi)}{\partial \theta} + \frac{1}{\sin^2 \theta} \frac{\partial^2 rv(r, \theta, \varphi)}{\partial \varphi^2} \right) \hat{r} \\ &\quad + \frac{1}{r} \frac{\partial^2 rv(r, \theta, \varphi)}{\partial r \partial \theta} \hat{\theta} + \frac{1}{r \sin \theta} \frac{\partial^2 rv(r, \theta, \varphi)}{\partial r \partial \varphi} \hat{\varphi} \end{aligned}$$

En particulier, nous voyons que

$$\left\{ \begin{array}{l} \operatorname{curl}(v(x)\vec{x}) \cdot \hat{r} = 0 \\ \operatorname{curl} \operatorname{curl}(v(x)\vec{x}) \cdot \hat{r} = \\ -\frac{1}{r} \left( \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial v(r, \theta, \varphi)}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 v(r, \theta, \varphi)}{\partial \varphi^2} \right) \end{array} \right. \quad (\text{A.31})$$

Dans le même temps, la définition (A.18) permet d'écrire

$$\left\{ \begin{array}{l} \lim_{r \rightarrow a} \hat{r} \wedge (\operatorname{curl}(rv(r, \theta, \varphi)\hat{r}) \wedge \hat{r}) = -\hat{r} \wedge \vec{\nabla}_{S_a^2} av(a, \theta, \varphi) \\ \lim_{r \rightarrow a} \hat{r} \wedge (\operatorname{curl} \operatorname{curl}(rv(r, \theta, \varphi)\hat{r}) \wedge \hat{r}) = \left( \vec{\nabla}_{S_r^2} \frac{\partial rv(r, \theta, \varphi)}{\partial r} \right)_{r=a} \end{array} \right.$$

En particulier, si  $v(r, \theta, \varphi) = \chi(kr)Y(\theta, \varphi)$

$$\left\{ \begin{array}{l} \lim_{r \rightarrow a} \hat{r} \wedge (\operatorname{curl}(rv(r, \theta, \varphi)\hat{r}) \wedge \hat{r}) = -\frac{1}{k} (t\chi(t))_{t=ka} \hat{r} \wedge \vec{\nabla}_{S_a^2} Y(\theta, \varphi) \\ \lim_{r \rightarrow a} \hat{r} \wedge (\operatorname{curl} \operatorname{curl}(rv(r, \theta, \varphi)\hat{r}) \wedge \hat{r}) = \left( \frac{d}{dt} (t\chi(t)) \right)_{t=ka} \vec{\nabla}_{S_r^2} Y(\theta, \varphi) \end{array} \right. \quad (\text{A.32})$$

De ces résultats, nous en déduisons que ;

$$\begin{cases} E_t(a\hat{r}) &= -\hat{r} \wedge \vec{\nabla}_{S_a^2} aw(a, \theta, \varphi) + \frac{i}{k} \left( \vec{\nabla}_{S_r^2} \frac{\partial rv(r, \theta, \varphi)}{\partial r} \right)_{r=a}, \\ H_t(a\hat{r}) &= -\frac{i}{kZ_0} \left( \vec{\nabla}_{S_r^2} \frac{\partial rw(r, \theta, \varphi)}{\partial r} \right)_{r=a} - \frac{1}{Z_0} \hat{r} \wedge \vec{\nabla}_{S_a^2} av(a, \theta, \varphi), \end{cases}$$

tandis que, si  $v$  et  $w$  sont définis par le développement :

$$\begin{cases} E_t(a\hat{r}) &= -\frac{1}{k} \sum_{n=1}^{\infty} \sum_{m=-n}^n w_{n,m} (t\chi_n(t))_{t=ka} \hat{r} \wedge \vec{\nabla}_{S_a^2} \tilde{Y}_{n,m}(\theta, \varphi) \\ &\quad + \frac{1}{k} \sum_{n=1}^{\infty} \sum_{m=-n}^n v_{n,m} i(t\chi_n(t))'_{t=ka} \vec{\nabla}_{S_a^2} \tilde{Y}_{n,m}(\theta, \varphi) \\ H_t(a\hat{r}) &= -\frac{1}{kZ_0} \sum_{n=1}^{\infty} \sum_{m=-n}^n w_{n,m} i(t\chi_n(t))'_{t=ka} \vec{\nabla}_{S_a^2} \tilde{Y}_{n,m}(\theta, \varphi) \\ &\quad - \frac{1}{kZ_0} \sum_{n=1}^{\infty} \sum_{m=-n}^n v_{n,m} (t\chi_n(t))_{t=ka} \hat{r} \wedge \vec{\nabla}_{S_a^2} \tilde{Y}_{n,m}(\theta, \varphi) \end{cases} \quad (\text{A.33})$$

Notons que puisque  $\chi_n(t)$  est une fonction de Bessel sphérique ou de Hankel.  $t\chi_n(t)$  est une fonction de Riccati-Bessel ou de Ricatti-Hankel.

#### A.3.4 Décomposition d'une onde plane dans les potentiels de Debye

Nous considérons une onde plane incident se propageant dans la direction  $\hat{z}$  ( $\hat{x}$ ,  $\hat{y}$ ,  $\hat{z}$  sont les vecteurs unitaires usuels en coordonnées cartésiennes)

$$E^{inc}(x) = \hat{x} e^{-ik\hat{z} \cdot x}, \quad Z_0 H^{inc}(x) = -\hat{y} e^{-ik\hat{z} \cdot x}. \quad (\text{A.34})$$

Pour obtenir les potentiels de Debye associés à ce champ, nous commençons, par regarder la décomposition de la composante radiale :

$$\begin{bmatrix} E^{inc}(x) \cdot \hat{r} \\ -Z_0 H^{inc}(x) \cdot \hat{r} \end{bmatrix} = \begin{bmatrix} \sin \theta \cos \varphi \\ \sin \theta \sin \varphi \end{bmatrix} e^{-ikr \cos \theta}$$

En différentiant le développement de Jacobi-Anger :

$$e^{-ikr \cos \theta} = \sum_{n=0}^{\infty} (-i)^n (2n+1) j_n(kr) P_n(\cos \theta),$$

suivant  $\theta$ , on obtient :

$$ikr \sin \theta e^{-ikr \cos \theta} = - \sum_{n=1}^{\infty} (-i)^n (2n+1) j_n(kr) \sin \theta P'_n(\cos \theta),$$

et puisque  $\sin \theta P'_n(\cos \theta) = P_n^1(\cos \theta)$ , nous voyons que

$$\begin{bmatrix} E^{inc}(x) \cdot \hat{r} \\ -Z_0 H^{inc}(x) \cdot \hat{r} \end{bmatrix} = \frac{i}{kr} \sum_{n=1}^{\infty} (-i)^n (2n+1) j_n(kr) \begin{bmatrix} P_n^1(\cos \theta) \cos \varphi \\ P_n^1(\cos \theta) \sin \varphi \end{bmatrix}$$

Soit  $w^{inc}(x)$  et  $v^{inc}(x)$  deux potentiels de Debye associés au champ électromagnétique  $(E^{inc}(x), H^{inc}(x))$

$$\begin{cases} E^{inc}(x) &= \operatorname{curl}(\vec{x}w^{inc}(x)) + \frac{i}{k} \operatorname{curl} \operatorname{curl}(\vec{x}v^{inc}(x)), \\ H^{inc}(x) &= -\frac{i}{kZ_0} \operatorname{curl} \operatorname{curl}(\vec{x}w^{inc}(x)) + \frac{1}{Z_0} \operatorname{curl}(\vec{x}v^{inc}(x)). \end{cases} \quad (\text{A.35})$$

Les équations (A.31) fournissent les relations :

$$\begin{bmatrix} E^{inc}(x) \cdot \hat{r} \\ -Z_0 H^{inc}(x) \cdot \hat{r} \end{bmatrix} = -\frac{i}{kr} \left( \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} \right) \begin{bmatrix} v^{inc}(x) \\ w^{inc}(x) \end{bmatrix}$$

Nous remarquons que puisque  $rP_n^1(\cos \theta) \cos \varphi$  et  $rP_n^1(\cos \theta) \sin \varphi$  sont harmoniques,

$$\begin{cases} \left( \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} \right) \begin{bmatrix} P_n^1(\cos \theta) \cos \varphi \\ P_n^1(\cos \theta) \sin \varphi \end{bmatrix} = \\ -n(n+1) \begin{bmatrix} P_n^1(\cos \theta) \cos \varphi \\ P_n^1(\cos \theta) \sin \varphi \end{bmatrix} \end{cases}$$

Par une simple comparaison, nous obtenons la décomposition souhaitée

$$\begin{bmatrix} v^{inc}(r, \theta, \varphi) \\ w^{inc}(r, \theta, \varphi) \end{bmatrix} = \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} j_n(kr) P_n^1(\cos \theta) \begin{bmatrix} \cos \varphi \\ \sin \varphi \end{bmatrix} \quad (\text{A.36})$$

De cette relation, nous en déduisons aisément la trace tangentielle du champ électromagnétique sur la sphère  $S_a^2$

$$\begin{cases} E_t^{inc}(a\hat{r}) = -\frac{1}{k} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} \psi_n(ka) \hat{r} \wedge \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \sin \varphi \\ \quad + \frac{1}{k} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} i\psi'_n(ka) \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \cos \varphi \\ H_t^{inc}(a\hat{r}) = -\frac{1}{kZ_0} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} i\psi'_n(ka) \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \sin \varphi \\ \quad - \frac{1}{kZ_0} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} \psi_n(ka) \hat{r} \wedge \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \cos \varphi \end{cases} \quad (\text{A.37})$$

### A.3.5 Diffraction par une sphère parfaitement conductrice

#### Formulation du problème

Nous considérons une onde plane :

$$E^{inc}(x) = \hat{x} e^{-ik\hat{z} \cdot x}, \quad Z_0 H^{inc}(x) = -\hat{y} e^{-ik\hat{z} \cdot x}. \quad (\text{A.38})$$

Nous cherchons la solution au problème de diffraction par une sphère parfaitement conductrice :

$$\begin{cases} \operatorname{curl} E - ikZ_0 H = 0, & \text{in } D_a^+ \\ \operatorname{curl} H + ikZ_0^{-1} E = 0, & \text{in } D_a^+, \end{cases} \quad (\text{A.39})$$

avec la condition de Silver-Müller à l'infini

$$\lim_{|x| \rightarrow \infty} |x| \left( Z_0(H - H^{inc}) \wedge \frac{x}{|x|} - (E - E^{inc}) \right) = 0, \quad (\text{A.40})$$

and la condition aux limites (la normale  $n(x)$  est  $\hat{r}$  pour une sphère)

$$E_t(x) + \beta(n(x) \wedge Z_0 H(x)) = 0. \quad (\text{A.41})$$

Le coefficient  $\beta$  est complexe avec  $\Re e \beta \leq 0$  pour assurer que le problème est bien posé. Le cas  $\beta = 0$  correspond à une sphère parfaitement conductrice et  $\beta = -1$  est associé à une condition de Silver-Müller posée à distance finie.

### Solution exprimée à l'aide des potentiels de Debye

La solution peut être décomposée à l'aide des potentiels de Debye

$$E(x) = E^{inc}(x) + E^d(x), \quad H(x) = H^{inc}(x) + H^d(x)$$

avec

$$\begin{cases} E^d(x) &= \operatorname{curl}(\vec{x} w^d(x)) + \frac{i}{k} \operatorname{curl} \operatorname{curl}(\vec{x} v^d(x)), \\ H^d(x) &= -\frac{i}{k Z_0} \operatorname{curl} \operatorname{curl}(\vec{x} w^d(x)) + \frac{1}{Z_0} \operatorname{curl}(\vec{x} v^d(x)). \end{cases} \quad (\text{A.42})$$

La décomposition !(A.36) de l'onde plane nous mène à chercher les potentiels du champ diffracté sous la forme :

$$\begin{bmatrix} v^d(r, \theta, \varphi) \\ w^d(r, \theta, \varphi) \end{bmatrix} = \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} h_n^{(1)}(kr) P_n^1(\cos \theta) \begin{bmatrix} \alpha_n \cos \varphi \\ \beta_n \sin \varphi \end{bmatrix} \quad (\text{A.43})$$

où  $\alpha_n$  et  $\beta_n$  sont des coefficients inconnus déterminés par la condition aux limites. Notons que le choix de la fonction de Hankel de première espèce a été retenu pour assurer la condition de Sommerfeld.

Les expressions des composantes tangentielles d'un champ électromagnétique associé aux potentiels de Debye sont données par les relations (A.33). Nous les utilisons ensemble avec (A.36) et (A.43) pour obtenir les traces tangentielles

$$\begin{aligned} E_t(a\hat{r}) &= \frac{1}{k} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} i(\psi'_n(ka) + \alpha_n \xi_n^{(1)'}(ka)) \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \cos \varphi \\ &\quad - \frac{1}{k} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} (\psi_n(ka) + \beta_n \xi_n^{(1)}(ka)) \hat{r} \wedge \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \sin \varphi \\ H_t(a\hat{r}) &= \frac{-1}{k Z_0} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} i(\psi'_n(ka) + \beta_n \xi_n^{(1)'}(ka)) \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \sin \varphi \\ &\quad - \frac{1}{k Z_0} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} (\psi_n(ka) + \alpha_n \xi_n^{(1)}(ka)) \hat{r} \wedge \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \cos \varphi \end{aligned}$$

A cause de l'orthogonalité mutuelle des termes du développement, la condition aux limites (A.41) est vérifiée dès que :

$$\beta_n = -\frac{\psi_n(ka) + i\beta\psi'_n(ka)}{\xi_n^{(1)}(ka) + i\beta\xi_n^{(1)'}(ka)}, \quad \alpha_n = -\frac{\beta\psi_n(ka) + i\psi'_n(ka)}{\beta\xi_n^{(1)}(ka) + i\xi_n^{(1)'}(ka)}.$$

le problème est complètement résolu.

### A.3.6 Expression de E et H dans tout l'espace

Rappelons le rotationnel d'un vecteur en coordonnées sphériques

$$\operatorname{curl} u = \begin{vmatrix} \frac{1}{r^2 \sin \theta} (r \cos \theta u_\phi + r \sin \theta \frac{\partial u_\phi}{\partial \theta} - r \frac{\partial u_\theta}{\partial \phi}) \\ \frac{1}{r \sin \theta} (\frac{\partial u_r}{\partial \phi} - \sin \theta u_\phi - r \sin \theta \frac{\partial u_\phi}{\partial r}) \\ \frac{1}{r} (u_\theta + r \frac{\partial u_\theta}{\partial r} - \frac{\partial u_r}{\partial \theta}) \end{vmatrix}$$

Ainsi, nous avons

$$\operatorname{curl}(v \vec{x}) = \begin{vmatrix} 0 \\ \frac{1}{\sin \theta} \frac{\partial v}{\partial \phi} \\ -\frac{\partial v}{\partial \theta} \end{vmatrix}$$

et

$$\operatorname{curl}(\operatorname{curl}(v \vec{x})) = \begin{vmatrix} -\frac{1}{r} (\frac{1}{\tan \theta} \frac{\partial v}{\partial \theta} + \frac{\partial^2 v}{\partial \theta^2} + \frac{1}{\sin^2 \theta} \frac{\partial^2 v}{\partial \phi^2}) \\ \frac{1}{r} (\frac{\partial v}{\partial \theta} + r \frac{\partial^2 v}{\partial \theta \partial r}) \\ \frac{1}{r \sin \theta} (\frac{\partial v}{\partial \phi} + r \frac{\partial^2 v}{\partial r \partial \phi}) \end{vmatrix}$$

On peut en déduire que

$$\operatorname{curl}(v^d(r, \theta, \phi) \vec{x}) = \begin{vmatrix} 0 \\ -\sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} h_n^{(1)}(kr) \alpha_n P_n^1(\cos \theta) \frac{\sin \phi}{\sin \theta} \\ -\sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} h_n^{(1)}(kr) \alpha_n \frac{\partial}{\partial \theta} (P_n^1(\cos \theta)) \cos \phi \end{vmatrix} \quad (\text{A.44})$$

$$\operatorname{curl}(w^d(r, \theta, \phi) \vec{x}) = \begin{vmatrix} 0 \\ +\sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} h_n^{(1)}(kr) \beta_n P_n^1(\cos \theta) \frac{\cos \phi}{\sin \theta} \\ -\sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} h_n^{(1)}(kr) \beta_n \frac{\partial}{\partial \theta} (P_n^1(\cos \theta)) \sin \phi \end{vmatrix} \quad (\text{A.45})$$

De la même manière

$$\text{curl}(\text{curl}(v^d(r, \theta, \phi) \vec{x})) = \begin{cases} \frac{1}{r} \sum_{n=1}^{\infty} (-i)^n (2n+1) h_n^{(1)}(kr) \alpha_n P_n^1(\cos \theta) \cos \phi \\ \frac{1}{r} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} \xi_n^{(1)'}(kr) \alpha_n \frac{\partial}{\partial \theta} (P_n^1(\cos \theta)) \cos \phi \\ -\frac{1}{r \sin \theta} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} \xi_n^{(1)'}(kr) \alpha_n P_n^1(\cos \theta) \sin \phi \end{cases} \quad (\text{A.46})$$

$$\text{curl}(\text{curl}(w^d(r, \theta, \phi))) \vec{x} = \begin{cases} \frac{1}{r} \sum_{n=1}^{\infty} (-i)^n (2n+1) h_n^{(1)}(kr) \beta_n P_n^1(\cos \theta) \sin \phi \\ + \frac{1}{r} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} \xi_n^{(1)'}(kr) \beta_n \frac{\partial}{\partial \theta} (P_n^1(\cos \theta)) \sin \phi \\ + \frac{1}{r \sin \theta} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} \xi_n^{(1)'}(kr) \beta_n P_n^1(\cos \theta) \cos \phi \end{cases} \quad (\text{A.47})$$

Finallement, nous avons l'expression :

$$E^d = \begin{cases} \frac{1}{kr} \sum_{n=1}^{\infty} (-i)^n (2n+1) i h_n^1(kr) \alpha_n P_n^1(\cos \theta) \cos \phi \\ \frac{1}{kr} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} \left( \alpha_n i \xi_n^{(1)'}(kr) \frac{\partial}{\partial \theta} (P_n^1(\cos \theta)) + \frac{\beta_n}{\sin \theta} \xi_n^{(1)}(kr) P_n^1(\cos \theta) \right) \cos \phi \\ -\frac{1}{kr} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} \left( \frac{\alpha_n}{\sin \theta} i \xi_n^{(1)'}(kr) P_n^1(\cos \theta) + \beta_n \frac{\partial}{\partial \theta} (P_n^1(\cos \theta)) \xi_n^{(1)}(kr) \right) \sin \phi \end{cases} \quad (\text{A.48})$$

Elle peut être réécrite :

$$\begin{aligned} E^d &= \frac{i}{kr} \sum_{n=1}^{\infty} (-i)^n (2n+1) h_n^{(1)}(kr) \alpha_n P_n^1(\cos \theta) \cos \phi \hat{r} \\ &\quad + \frac{1}{k} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} i \alpha_n \xi_n^{(1)'}(kr) \nabla_{S_r} (P_n^1(\cos \theta) \cos \phi) \\ &\quad - \frac{1}{k} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} \beta_n \xi_n^{(1)}(kr) \hat{r} \times \nabla_{S_r} (P_n^1(\cos \theta) \sin \phi) \end{aligned} \quad (\text{A.49})$$

Le champ magnétique diffracté est égal à

$$\begin{aligned}
H^d &= -\frac{i}{Z_0 kr} \sum_{n=1}^{\infty} (-i)^n (2n+1) h_n^{(1)}(kr) \beta_n P_n^1(\cos \theta) \sin \phi \hat{r} \\
&\quad - \frac{1}{k Z_0} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} i \beta_n \xi_n^{(1)'}(kr) \nabla_{S_r}(P_n^1(\cos \theta) \sin \phi) \\
&\quad - \frac{1}{k Z_0} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} \alpha_n \xi_n^{(1)}(kr) \hat{r} \times \nabla_{S_r}(P_n^1(\cos \theta) \cos \phi)
\end{aligned} \tag{A.50}$$

### A.3.7 Diffraction d'une onde plane par une sphère diélectrique

Maintenant étudions la diffraction d'une sphère diélectrique de rayon  $a$  et d'indices  $\epsilon_r, \mu_r$ . OA l'extérieur de la sphère, nous avons un domaine homogène infini où  $\epsilon_r = 1$  and  $\mu_r = 1$ , et  $Z_0 = 1$ . A l'intérieur, nous avons  $\epsilon_r \neq 1 \quad \mu_r \neq 1$  and  $Z_0 = \sqrt{\frac{\mu_r}{\epsilon_r}}$ . A partir de maintenant,  $Z_0$  remplacera  $\sqrt{\frac{\mu_r}{\epsilon_r}}$  et le domaine extérieur aura une impédance de 1

La solution à l'extérieur de la sphère, peut être développée sous la forme :

$$\begin{aligned}
v^d &= \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} h_n^{(1)}(kr) P_n^1(\cos \theta) \alpha_n \cos \phi \\
w^d &= \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} h_n^{(1)}(kr) P_n^1(\cos \theta) \beta_n \sin \phi
\end{aligned} \tag{A.51}$$

A l'intérieur de la sphère, la fonction de Bessel sphérique  $j_n$  sera utilisé pour le champ total

$$\begin{aligned}
v^i &= \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} j_n(k_i r) P_n^1(\cos \theta) \gamma_n \cos \phi \\
w^i &= \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} j_n(k_i r) P_n^1(\cos \theta) \delta_n \sin \phi
\end{aligned} \tag{A.52}$$

$k_i$  est le nombre d'onde de la sphère diélectrique, il est égal à :  $k_i = k \sqrt{\epsilon_r \mu_r}$

Rappelons les expression des traces tangentiellles de  $E$  et  $H$  à l'extérieur de la sphère quand  $r$  tends vers  $a$ . Ces traces sont les traces du champ total :

$$\begin{aligned}
E_t(a\hat{r}) &= \frac{1}{k} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} i(\psi'_n(ka) + \alpha_n \xi_n^{(1)'}(ka)) \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \cos \varphi \\
&\quad - \frac{1}{k} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} (\psi_n(ka) + \beta_n \xi_n^{(1)}(ka)) \hat{r} \times \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \sin \varphi \\
H_t(a\hat{r}) &= \frac{-1}{k} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} i(\psi'_n(ka) + \beta_n \xi_n^{(1)'}(ka)) \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \sin \varphi \\
&\quad - \frac{1}{k} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} (\psi_n(ka) + \alpha_n \xi_n^{(1)}(ka)) \hat{r} \times \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \cos \varphi
\end{aligned}$$

Explicitons maintenant les traces tangentielles du champ total  $E$  et  $H$  à l'intérieur de la sphère quand  $r$  tends vers  $a$ .

$$\begin{aligned} E_t(a\hat{r}) &= \frac{1}{k_i} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} i \gamma_n \psi'_n(k_i a) \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \cos \varphi \\ &\quad - \frac{1}{k_i} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} \delta_n \psi_n(k_i a) \hat{r} \times \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \sin \varphi \\ H_t(a\hat{r}) &= \frac{-1}{k_i Z_0} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} i \delta_n \psi'_n(k_i a) \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \sin \varphi \\ &\quad - \frac{1}{k_i Z_0} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} \gamma_n \psi_n(k_i a) \hat{r} \times \vec{\nabla}_{S_a^2} P_n^1(\cos \theta) \cos \varphi \end{aligned}$$

The traces tangentielles de  $E$  et  $H$  sont continues, donc les coefficients  $\alpha_n, \beta_n, \gamma_n$  and  $\delta_n$  vérifient les équations

$$\left\{ \begin{array}{lcl} \frac{i}{k} (\psi'_n(ka) + \alpha_n \xi_n^{(1)'}(ka)) & = & \frac{i}{k_i} \gamma_n \psi'_n(k_i a) \\ \frac{1}{k} (\psi_n(ka) + \beta_n \xi_n^{(1)}(ka)) & = & \frac{1}{k_i} \delta_n \psi_n(k_i a) \\ \frac{i}{k} (\psi'_n(ka) + \beta_n \xi_n^{(1)'}(ka)) & = & \frac{i}{k_i Z_0} \delta_n \psi'_n(k_i a) \\ \frac{1}{k} (\psi_n(ka) + \alpha_n \xi_n^{(1)}(ka)) & = & \frac{1}{k_i Z_0} \gamma_n \psi_n(k_i a) \end{array} \right. \quad (A.53)$$

Ainsi nous avons un système 2x2 en  $\alpha_n$  and  $\gamma_n$

$$\begin{aligned} -k_i \alpha_n \xi_n^{(1)'}(ka) + k \gamma_n \psi'_n(k_i a) &= k_i \psi'_n(ka) \\ -k_i Z_0 \alpha_n \xi_n^{(1)}(ka) + k \gamma_n \psi_n(k_i a) &= k_i Z_0 \psi_n(ka) \end{aligned} \quad (A.54)$$

Multiplions la première équation par  $\psi_n(k_i a)$  et soustrayons la à la seconde équation multipliée par  $\psi'_n(k_i a)$  Ainsi, nous obtenons :

$$\alpha_n (k_i Z_0 \psi'_n(k_i a) \xi_n^{(1)}(ka) - k_i \xi_n^{(1)'}(ka) \psi_n(k_i a)) = k_i \psi'_n(ka) \psi_n(k_i a) - k_i Z_0 \psi_n(ka) \psi'_n(k_i a) \quad (A.55)$$

$\alpha_n$  est égale à :

$$\alpha_n = \frac{\psi'_n(k a) \psi_n(k_i a) - Z_0 \psi_n(k a) \psi'_n(k_i a)}{Z_0 \psi'_n(k_i a) \xi_n^{(1)}(ka) - \xi_n^{(1)'}(ka) \psi_n(k_i a)} \quad (A.56)$$

Nous avons aussi deux équations en  $\beta_n$  et  $\delta_n$

$$\begin{aligned} -k_i \beta_n \xi_n^{(1)}(ka) + k \delta_n \psi_n(k_i a) &= k_i \psi_n(ka) \\ -k_i Z_0 \beta_n \xi_n^{(1)'}(ka) + k \delta_n \psi'_n(k_i a) &= k_i Z_0 \psi'_n(ka) \end{aligned} \quad (A.57)$$

Multiplions la première équation par  $\psi'_n(k_i a)$  et soustrayons la à la deuxième équation multipliée par  $\psi_n(k_i a)$ . Nous avons :

$$\beta_n (k_i Z_0 \xi_n^{(1)'}(ka) \psi_n(k_i a) - k_i \psi'_n(k_i a) \xi_n^{(1)}(ka)) = k_i \psi_n(ka) \psi'_n(k_i a) - k_i Z_0 \psi'_n(ka) \psi_n(k_i a) \quad (A.58)$$

$\beta_n$  est égal à

$$\beta_n = \frac{\psi_n(ka) \psi'_n(k_i a) - Z_0 \psi'_n(ka) \psi_n(k_i a)}{Z_0 \xi_n^{(1)'}(ka) \psi_n(k_i a) - \psi'_n(k_i a) \xi_n^{(1)}(ka)} \quad (\text{A.59})$$

De la même manière, on peut calculer  $\gamma_n$  et  $\delta_n$

$$\begin{aligned} \gamma_n &= \frac{k_i}{k} \frac{Z_0 \xi_n^{(1)}(ka) \psi'_n(ka) - Z_0 \xi_n^{(1)'}(ka) \psi_n(ka)}{Z_0 \psi'_n(k_i a) \xi_n^{(1)}(ka) - \psi_n(k_i a) \xi_n^{(1)'}(ka)} \\ \delta_n &= \frac{k_i}{k} \frac{Z_0 \xi_n^{(1)'}(ka) \psi_n(ka) - Z_0 \xi_n^{(1)}(ka) \psi'_n(ka)}{Z_0 \xi_n^{(1)'}(ka) \psi_n(k_i a) - \xi_n^{(1)}(ka) \psi'_n(k_i a)} \end{aligned} \quad (\text{A.60})$$

Rappelons les coefficients  $\alpha_n$  et  $\beta_n$

$$\begin{aligned} \alpha_n &= \frac{\psi'_n(k a) \psi_n(k_i a) - Z_0 \psi_n(ka) \psi'_n(k_i a)}{Z_0 \psi'_n(k_i a) \xi_n^{(1)}(ka) - \xi_n^{(1)'}(ka) \psi_n(k_i a)} \\ \beta_n &= \frac{\psi_n(ka) \psi'_n(k_i a) - Z_0 \psi'_n(ka) \psi_n(k_i a)}{Z_0 \xi_n^{(1)'}(ka) \psi_n(k_i a) - \psi'_n(k_i a) \xi_n^{(1)}(ka)} \end{aligned} \quad (\text{A.61})$$

### A.3.8 Prise en compte de la condition de Silver-Müller

De même que pour l'équation de Helmholtz, il suffit de remplacer la fonction de Riccati-Hankel  $\xi_n^{(1)}$  par  $\tilde{\xi}_n^{(1)}$  :

$$\tilde{\xi}_n^{(1)}(k r) = \xi_n^{(1)}(k r) - \left( \frac{\xi_n^{(1)'}(k b) - i \xi_n(1)(k b)}{\xi_n^{(2)'}(k b) - i \xi_n(2)(k b)} \right) \xi_n^{(2)}(k r)$$



## Annexe B

# Condition transparente pour les équations de Maxwell en régime harmonique utilisant une formule de représentation intégrale

*Nous rappelons dans cette annexe, une méthode pour calculer un problème de diffraction en domaine non-borné. Cette méthode, qu'on appellera “condition transparente” utilise une représentation intégrale de la solution. Cette approche a été introduite par [Hazard et Lenoir, 1996] et également par [J. Liu, 2001].*

### Sommaire

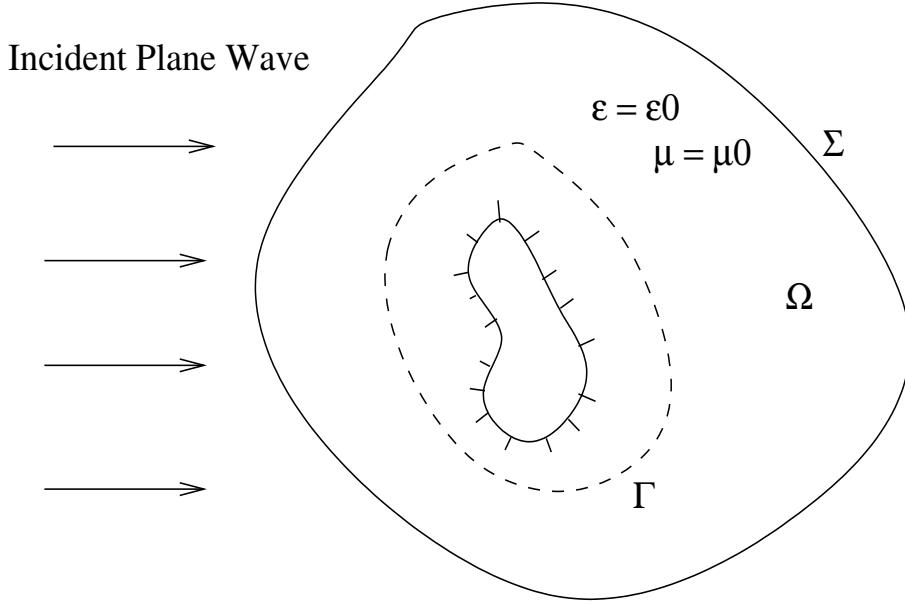
---

B.1	Problème Modèle . . . . .	254
B.2	Description de la condition transparente . . . . .	255
B.3	Formulation variationnelle . . . . .	257
B.4	Calcul du produit matrice-vecteur . . . . .	259
B.4.1	Cas 2-D . . . . .	259
B.4.2	Cas 3-D . . . . .	260
B.5	Équations de Maxwell 3-D . . . . .	261

---

## B.1 Problème Modèle

On étudie la diffraction d'une onde incidente par un obstacle



Dans un premier temps, on étudie l'équation de Helmholtz en domaine non borné. On note  $\mathcal{O}$  l'intérieur de l'obstacle,  $\partial\mathcal{O}$  le bord de l'obstacle. On décompose la solution  $u$  en

$$u = u_d + u_{inc}$$

$u_{inc}$  est une onde incidente vérifiant l'équation de Helmholtz, typiquement une onde plane.  $u_d$ , le champ diffracté est solution de

$$\begin{cases} -k^2 u - \Delta u &= 0 \quad \text{dans } \mathbb{R}^2 \setminus \mathcal{O} \\ \lim_{r \rightarrow +\infty} \sqrt{r} \left( \frac{\partial u}{\partial n} - iku \right) &= 0 \\ + \text{condition aux limites sur } \partial\mathcal{O} \end{cases} \quad (\text{B.1})$$

La seconde ligne correspond à la condition de Sommerfeld, on impose au champ diffracté de satisfaire à cette condition. Suivant la nature de l'obstacle (diélectrique, conducteur parfait ...), on adaptera la condition aux limites sur  $\partial\mathcal{O}$ , la frontière de l'obstacle. Pour simplifier notre discours, on prendra une condition de Neumann homogène sur  $\partial\mathcal{O}$ , ce qui donne l'équation sur  $u_d$

$$\begin{cases} -k^2 u - \Delta u &= 0 \quad \text{dans } \mathbb{R}^2 \setminus \mathcal{O} \\ \lim_{r \rightarrow +\infty} \sqrt{r} \left( \frac{\partial u}{\partial n} - iku \right) &= 0 \\ \frac{\partial u}{\partial n} &= -\frac{\partial u_{inc}}{\partial n} \quad \text{sur } \partial\mathcal{O} \end{cases} \quad (\text{B.2})$$

Malheureusement, les ordinateurs ont une mémoire finie et on ne peut faire des calculs en domaine fini, on doit utiliser des domaines bornés. On a alors plusieurs choix pour approximer la

condition de Sommerfeld, les conditions absorbantes, PML... Une condition absorbante revient à calculer la solution sur un domaine borné  $\Omega$  et imposer une condition aux limites sur la frontière extérieure  $\Sigma$ . Le choix le plus simple de condition aux limites est la condition absorbante d'ordre un. On résout alors

$$\begin{cases} -k^2 u - \Delta u = f & \text{dans } \Omega \\ \frac{\partial u}{\partial n} - iku = 0 & \text{sur } \Sigma \\ \frac{\partial u}{\partial n} = -\frac{\partial u_{inc}}{\partial n} & \text{sur } \partial\Omega \end{cases} \quad (\text{B.3})$$

On va maintenant décrire une condition transparente, qui utilise la condition absorbante d'ordre un. Puis, on discutera de la discréétisation par éléments finis de l'équation de Helmholtz avec cette condition transparente.

## B.2 Description de la condition transparente

Dans la figure 1, on a fait figurer le domaine  $\Omega$ , la frontière  $\Sigma$  et l'obstacle. On a également rajouté une frontière interne au domaine  $\Gamma$  et on a supposé que le milieu était homogène entre la  $\Gamma$  et  $\Sigma$ . A l'intérieur de  $\Gamma$ , on peut résoudre une équation de Helmholtz plus générale

$$-\rho(x) u(x) - \operatorname{div}(\mu(x) \operatorname{grad}(u(x))) = f \quad (\text{B.4})$$

$\rho$  et  $\mu$  sont des coefficients positifs qui peuvent dépendre de  $x$ . Comme le milieu est homogène à l'extérieur de  $\Gamma$ , on peut calculer  $u$  dans tout l'espace extérieur à  $\Gamma$  grâce à la formule de représentation

$$u(x) = \int_{\Gamma} \frac{\partial \phi(x, y)}{\partial n(y)} u(y) - \phi(x, y) \frac{\partial u(y)}{\partial n(y)} dy \quad (\text{B.5})$$

pour tout  $x$  à l'extérieur de  $\Gamma$ , notamment sur  $\Sigma$ .  $\phi(x, y)$  est le noyau de Green de l'équation de Helmholtz en milieu homogène.

$$\begin{cases} \phi(x, y) = \frac{i}{4} H_0^{(1)}(k|x - y|) & \text{en 2-D} \\ \phi(x, y) = \frac{\exp(ik|x - y|)}{4\pi|x - y|} & \text{en 3-D} \end{cases} \quad (\text{B.6})$$

Une première idée serait alors d'utiliser cette formule de représentation comme condition aux limites sur  $\Sigma$ .  $u_d$  est alors solution de

$$\begin{cases} -k^2 u(x) - \Delta u(x) = 0 & x \in \Omega \\ u(x) = \int_{\Gamma} \frac{\partial \phi(x, y)}{\partial n(y)} u(y) - \phi(x, y) \frac{\partial u(y)}{\partial n(y)} dy & x \in \Sigma \\ \frac{\partial u}{\partial n}(x) = -\frac{\partial u_{inc}}{\partial n} & x \in \partial\Omega \end{cases} \quad (\text{B.7})$$

On appelle la condition sur  $\Sigma$  condition transparente, car la solution en domaine non-borné vérifie cette condition, ce n'est donc pas une approximation. D'autres auteurs préfèrent employer

l'expression “condition exacte”. Notre objectif est de dériver une formulation variationnelle à partir de ce système. Un premier désavantage de cette condition est qu'elle s'écrit comme un opérateur Dirichlet-to-Neumann

$$u = D(u, \frac{\partial u}{\partial n})$$

Cette écriture est difficile à insérer dans la formulation variationnelle, car lorsqu'on fait l'intégration par parties, on a

$$-\int_{\Omega} k^2 u(x)v(x)dx + \int_{\Omega} \nabla u \cdot \nabla v - \int_{\Sigma} \frac{\partial u}{\partial n} v ds = \int_{\Omega} f v \quad (\text{B.8})$$

On intègre la condition aux limites relative à  $\Sigma$  via le terme de bord sur  $\Sigma$ . Une condition de Dirichlet est traitée en l'imposant sur l'espace d'approximation. On préfère traiter des conditions qui s'écrivent sous la forme d'un opérateur Neumann-to-Dirichlet

$$\frac{\partial u}{\partial n} = N(u)$$

Afin d'obtenir cette forme, on dérive la formule de représentation intégrale

$$\frac{\partial u(x)}{\partial n(x)} = \int_{\Gamma} \frac{\partial^2 \phi(x, y)}{\partial n(y) \partial n(x)} u(y) - \frac{\partial \phi(x, y)}{\partial n(x)} \frac{\partial u(y)}{\partial n(y)} dy \quad (\text{B.9})$$

On considère alors  $u_d$  solution de

$$\left\{ \begin{array}{l} -k^2 u(x) - \Delta u(x) = 0 \quad x \in \Omega \\ \frac{\partial u(x)}{\partial n(x)} = \int_{\Gamma} \frac{\partial^2 \phi(x, y)}{\partial n(y) \partial n(x)} u(y) - \frac{\partial \phi(x, y)}{\partial n(x)} \frac{\partial u(y)}{\partial n(y)} dy \quad x \in \Sigma \\ \frac{\partial u}{\partial n}(x) = -\frac{\partial u_{inc}}{\partial n} \quad x \in \partial\Omega \end{array} \right. \quad (\text{B.10})$$

On peut aussi faire des combinaisons linéaires des deux conditions, et obtenir ainsi une condition absorbante “modifiée”.

$u_d$  est solution de

$$\left\{ \begin{array}{l} -k^2 u(x) - \Delta u(x) = 0 \quad x \in \Omega \\ \frac{\partial u}{\partial n}(x) - iku(x) = \int_{\Gamma} \frac{\partial^2 \phi(x, y)}{\partial n(y) \partial n(x)} u(y) - \frac{\partial \phi(x, y)}{\partial n(x)} \frac{\partial u(y)}{\partial n(y)} dy - ik \int_{\Gamma} \frac{\partial \phi(x, y)}{\partial n(y)} u(y) - \phi(x, y) \frac{\partial u(y)}{\partial n(y)} dy \\ \frac{\partial u}{\partial n}(x) = 0 \quad x \in \partial\Omega \end{array} \right. \quad (\text{B.11})$$

Par la suite, c'est cette condition qu'on appellera condition transparente. On essaiera également de justifier qu'on privilégie celle-ci par rapport à la condition “Neumann”. On va maintenant écrire la formulation variationnelle.

### B.3 Formulation variationnelle

On prend une fonction  $v$  qu'on multiplie à l'équation de Helmholtz, et on intègre sur tout le domaine. On fait une intégration par parties sur le terme en laplacien. On obtient alors

$$-\int_{\Omega} k^2 u(x)v(x)dx + \int_{\Omega} \nabla u \cdot \nabla v - \int_{\Sigma} \frac{\partial u}{\partial n} v ds - \int_{\partial\Omega} \frac{\partial u}{\partial n} v ds = \int_{\Omega} f v \quad (\text{B.12})$$

On injecte les conditions aux limites sur  $\Sigma$  et  $\partial\Omega$

$$\begin{aligned} & -\int_{\Omega} k^2 u(x)v(x)dx + \int_{\Omega} \nabla u \cdot \nabla v - ik \int_{\Sigma} uv ds - \int_{\Sigma} \left[ \int_{\Gamma} \frac{\partial^2 \phi(x,y)}{\partial n(y) \partial n(x)} u(y) - \frac{\partial \phi(x,y)}{\partial n(x)} \frac{\partial u(y)}{\partial n(y)} dy \right. \\ & \quad \left. - ik \int_{\Gamma} \frac{\partial \phi(x,y)}{\partial n(y)} u(y) - \phi(x,y) \frac{\partial u(y)}{\partial n(y)} dy \right] v ds = \int_{\Omega} f v - \int_{\partial\Omega} \frac{\partial u_{inc}}{\partial n} v dx \end{aligned}$$

On choisit  $u$  et  $v$  dans  $H^1(\Omega)$ , bien que ça puisse être problématique car  $\frac{\partial u}{\partial n}$  intervient dans les intégrales sur  $\Gamma$ . Toutes les intégrales dans le premier membre contribueront dans l'expression de la matrice après discréétisation, tandis que les autres intégrales constitueront le second membre. On choisit un espace d'approximation discret  $V_h$ . On note alors

$$(\varphi_i)_{1 \leq i \leq N}$$

les  $N$  fonctions de base de  $V_h$ .

La matrice “éléments finis” est de terme générique

$$A_{i,j}^c = -\int_{\Omega} k^2 \varphi_i(x) \varphi_j(x) dx + \int_{\Omega} \nabla \varphi_i \cdot \nabla \varphi_j - ik \int_{\Sigma} \varphi_i \varphi_j ds \quad (\text{B.13})$$

C'est la matrice qu'on obtiendrait si on mettait la seule condition absorbante d'ordre 1. Elle est particulièrement creuse, seuls les degrés de liberté partageant un même élément interagissent. On la note  $A^c$ , c comme creuse !

La matrice “équation intégrale” est de terme générique

$$A_{i,j}^p = \int_{\Sigma} \left[ \int_{\Gamma} \frac{\partial^2 \phi(x,y)}{\partial n(y) \partial n(x)} \varphi_j(y) - \frac{\partial \phi(x,y)}{\partial n(x)} \frac{\partial \varphi_j(y)}{\partial n(y)} dy - ik \int_{\Gamma} \frac{\partial \phi(x,y)}{\partial n(y)} \varphi_j(y) - \phi(x,y) \frac{\partial \varphi_j(y)}{\partial n(y)} dy \right] \varphi_i(x) ds \quad (\text{B.14})$$

Cette matrice contient une sous-matrice pleine, traduisant le caractère non local de la condition transparente. Si on décompose le vecteur  $\mathbf{U}$

$$\mathbf{U} = [U_i, U_g, U_s]$$

$U_s$  les degrés de liberté associés aux fonctions de base qui ne s'annulent pas sur  $\Sigma$

$U_g$  les degrés de liberté associés aux fonctions de base qui ne s'annulent pas sur  $\Gamma$  ainsi que leur gradient

$U_i$  tous les autres degrés de liberté

On a alors l'écriture bloc de  $A^p$

$$A_p = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & A^f & 0 \end{pmatrix}$$

$A^p$  s'applique à une “inconnue sur  $\Gamma$ ” et renvoie une “inconnue sur  $\Sigma$ ”. La sous-matrice  $A^f$  est pleine. Le second membre s’écrit

$$F_i = \int_{\Omega} f \varphi_i - \int_{\partial\Omega} \frac{\partial u_{inc}}{\partial n} \varphi_i ds \quad (\text{B.15})$$

$U$  est alors solution du système linéaire

$$(A^c - A^p)U = F \quad (\text{B.16})$$

$A^c$  a de très bonnes propriétés, notamment elle est très creuse, donc peu coûteuse en stockage. Elle est symétrique, ce qui est important car la factorisation  $LDL^t$  est moins coûteuse en stockage et en temps de calcul qu’une factorisation  $LU$ . Un système linéaire symétrique est également plus facile à résoudre par des méthodes itératives. Notamment la méthode COCG, qui est une version du gradient conjugué adapté aux matrices complexes symétriques, et semble particulièrement efficace sur ce type de matrices.

En revanche,  $A^p$  a de mauvaise propriétés, elle a une sous-matrice pleine de taille importante, elle est non-symétrique. C’est pour ces raisons, qu’on choisit de traiter cette matrice de manière itérative. On ne stocke pas cette matrice, mais on dispose d’une procédure qui effectue le produit matrice vecteur. On explicitera dans la suite comment on réalise cette dernière opération sans avoir à stocker la matrice.

On suppose avoir une procédure de résolution du système linéaire creux

$$A^c X = B$$

Par exemple on peut faire au préalable une factorisation  $LDL^t$  de la matrice, et résoudre alors des systèmes linéaires triangulaires.

On s’intéresse alors à la résolution par une méthode itérative de l’équation

$$U - (A^c)^{-1} A^p U = (A^c)^{-1} F = G \quad (\text{B.17})$$

$G$  est la solution de Helmholtz avec la condition d’ordre 1 homogène. Une première idée est d’appliquer l’algorithme de Jacobi. On prend  $U^0 = G$ , qui est déjà une bonne approximation de la solution exacte.

```

    Choisir  $U^0 = G$ 
    Pour  $i = 1 \dots N$ 
         $U^{n+1} = G + (A^c)^{-1} A^p U^n$ 
        Faire tant que  $\|A^p U^{n+1} - A^p U^n\| > \epsilon$ 

```

Le résidu de (B.16) est égal à

$$r = A^c U^{n+1} - A^p U^{n+1} - F$$

soit en utilisant l’équation

$$A^c U^{n+1} = F + A^p U^n$$

$$r = F + A^p U^n - A^p U^{n+1} - F$$

finalement

$$r = A^p U^n - A^p U^{n+1}$$

On peut aussi appliquer des méthodes itératives plus sophistiquées comme GMRES(m) sur le système linéaire

$$(I - (A^c)^{-1} A^p)U = G \quad (\text{B.18})$$

L'avantage de l'utilisation de telles méthodes est le gain de robustesse, car la méthode de Jacobi peut ne pas converger, on l'a constaté numériquement et analytiquement dans le cas du disque.

## B.4 Calcul du produit matrice-vecteur

### B.4.1 Cas 2-D

On note  $H_n^{(1)}$  la fonction de Hankel de première espèce d'ordre n.

Le noyau de Green s'écrit

$$\phi(x, y) = \frac{i}{4} H_0^{(1)}(k|x - y|)$$

Or

$$H_0^{(1)'}(x) = -H_1^{(1)}(x)$$

Et

$$\nabla_y |x - y| = \frac{(y - x)}{|x - y|}$$

d'où

$$\frac{\partial \phi(x, y)}{\partial n(y)} = \frac{ik}{4} H_1^{(1)}(k|x - y|) \frac{(x - y) \cdot n(y)}{|x - y|} \quad (\text{B.19})$$

On sait que

$$H_1^{(1)'}(x) = H_0^{(1)}(x) - \frac{1}{x} H_1^{(1)}(x)$$

On en déduit que

$$\begin{aligned} \frac{\partial^2 \phi(x, y)}{\partial n(x) \partial n(y)} &= \frac{ik^2}{4} H_0^{(1)}(k|x - y|) \frac{(x - y) \cdot n(y)}{|x - y|} \frac{(x - y) \cdot n(x)}{|x - y|} \\ &\quad - \frac{ik}{4} H_1^{(1)}(k|x - y|) \frac{(x - y) \cdot n(y)(x - y) \cdot n(x)}{|x - y|^3} \\ &\quad + \frac{ik}{4} H_1^{(1)}(k|x - y|) \frac{n(x) \cdot n(y)}{|x - y|} \\ &\quad - \frac{ik}{4} H_1^{(1)}(k|x - y|) \frac{(x - y) \cdot n(y)(x - y) \cdot n(x)}{|x - y|^3} \end{aligned}$$

Finalement

$$\begin{aligned} \frac{\partial^2 \phi(x, y)}{\partial n(x) \partial n(y)} &= \frac{i}{4} \frac{(x-y) \cdot n(y)(x-y) \cdot n(x)}{|x-y|^2} \left( k^2 H_0^{(1)}(k|x-y|) - 2k \frac{H_1^{(1)}(k|x-y|)}{|x-y|} \right) \\ &+ \frac{ik}{4} H_1^{(1)}(k|x-y|) \frac{n(x) \cdot n(y)}{|x-y|} H_0^{(1)}(k|x-y|) \end{aligned} \quad (\text{B.20})$$

On sépare le produit matrice vecteur  $Y = ApU$  en trois étapes

Etape 1 On calcule  $\frac{\partial u}{\partial n}$  aux points de quadrature de la frontière intérieure  $\Gamma$   
Sur chaque segment la composant, on choisit les points de Gauss.

Etape 2 On calcule  $g(x) = \int_{\Gamma} \frac{\partial \phi(x, y)}{\partial n(y)} u(y) - \phi(x, y) \frac{\partial u(y)}{\partial n(y)} dy$

$$\text{et } h(x) = \int_{\Gamma} \frac{\partial^2 \phi(x, y)}{\partial n(y) \partial n(x)} u(y) - \frac{\partial \phi(x, y)}{\partial n(x)} \frac{\partial u(y)}{\partial n(y)} dy \quad x \in \Sigma$$

aux points de quadrature sur la frontière extérieure  $\Sigma$

Etape 3 On calcule  $\int_{\Sigma} [h(x) - ik g(x)] \varphi_i ds$   
pour toute fonction de base ne s'annulant pas sur  $\Sigma$   
Ce coefficient est alors la i-ème composante du vecteur  $Y$

#### B.4.2 Cas 3-D

L'algorithme de calcul est identique en 3-D, ce qui est modifié c'est la fonction de green. On explicite dans cette section, ses dérivées comme on l'a fait pour le cas 2-D.

$$\phi(x, y) = \frac{\exp(ik|x-y|)}{4\pi|x-y|} \quad (\text{B.21})$$

$$\frac{\partial \phi(x, y)}{\partial n(y)} = \frac{(x-y) \cdot n(y)}{4\pi|x-y|^3} \exp(ik|x-y|) - ik \frac{(x-y) \cdot n(y)}{4\pi|x-y|^2} \exp(ik|x-y|) \quad (\text{B.22})$$

En dérivant, on obtient

$$\begin{aligned} \frac{\partial^2 \phi(x, y)}{\partial n(x) \partial n(y)} &= \frac{n(x) \cdot n(y)}{4\pi|x-y|^3} \exp(ik|x-y|) - 3 \frac{(x-y) \cdot n(y)(x-y) \cdot n(x)}{4\pi|x-y|^5} \exp(ik|x-y|) + \\ &\frac{ik(x-y) \cdot n(y)(x-y) \cdot n(x)}{4\pi|x-y|^4} \exp(ik|x-y|) - \frac{ikn(x) \cdot n(y)}{4\pi|x-y|^2} \exp(ik|x-y|) \\ &+ \frac{k^2(x-y) \cdot n(y)(x-y) \cdot n(x)}{4\pi|x-y|^3} + \frac{2ik(x-y) \cdot n(y)(x-y) \cdot n(x)}{4\pi|x-y|^4} \exp(ik|x-y|) \end{aligned}$$

En regroupant, on trouve

$$\frac{\partial^2 \phi(x, y)}{\partial n(x) \partial n(y)} = \frac{\exp(ik|x-y|)}{4\pi|x-y|} \left[ n(x) \cdot n(y) \left( \frac{1}{|x-y|^2} - \frac{ik}{|x-y|} \right) + \frac{(x-y) \cdot n(y)(x-y) \cdot n(x)}{|x-y|^2} \left( k^2 + \frac{3ik}{|x-y|} - \frac{3}{|x-y|^2} \right) \right] \quad (\text{B.23})$$

## B.5 Equations de Maxwell 3-D

On étudie la diffraction d'un objet métallique parfaitement conducteur de frontière  $\Gamma$ . On borne le domaine de calcul  $\Omega$  par une frontière extérieure  $\Sigma$ , sur laquelle on met une condition de Silver-Müller. L'objet métallique est plongé dans un milieu hétérogène de permittivité diélectrique  $\epsilon$ , et de perméabilité magnétique  $\mu$ . On suppose, qu'à l'infini, le milieu est homogène, d'indices  $\epsilon_0$  et  $\mu_0$ . Le champ diffracté est alors solution de

$$\begin{aligned} -i\omega \epsilon_r E(x) - \vec{\text{rot}} H(x) &= i\omega (\epsilon - \epsilon_0) E^i \quad x \in \Omega \\ -i\omega \mu(x) H(x) + \text{rot} E(x) &= i\omega (\mu - \mu_0) H^i \quad x \in \Omega \\ n \times E(x) &= -n \times E^i(x) \quad x \in \Gamma \\ \text{rot}(E) \times n &= \frac{ik}{\mu} (n \times E) \times n \quad x \in \Sigma \end{aligned} \quad (\text{B.24})$$

$\vec{E}$  et  $H$  sont respectivement le champ électrique et le champ magnétique. La dernière condition est la condition de Silver-Müller. On modifie cette condition de la même manière qu'on l'a fait pour l'équation de Helmholtz.

$$\text{rot}(E) \times n = \frac{ik}{\mu} (n \times E) \times n + \text{rot}(E^{pot}) \times n - \frac{ik}{\mu} (n \times E^{pot}) \times n \quad (\text{B.25})$$

Soit en faisant intervenir  $H^{pot}$  :

$$\text{rot}(E) \times n = \frac{ik}{\mu} (n \times E) \times n + i\omega \mu H^{pot} \times n - \frac{ik}{\mu} (n \times E^{pot}) \times n \quad (\text{B.26})$$

$H^{pot}$  et  $E^{pot}$  sont les champs magnétiques et électriques sur  $\Sigma$  calculés à l'aide d'une formule de représentation intégrale utilisant  $E$  et  $H$  sur  $\Gamma$ . La formule de représentation intégrale de type Stratton-Chu s'écrit :

$$\begin{aligned} E^{pot}(x) &= \int_{\Gamma} ik G(x, y) (n \times H)(y) dy + \int_{\Gamma} (n \times E)(y) \times \nabla_y \phi(x, y) dy \\ H^{pot}(x) &= - \int_{\Gamma} ik G(x, y) (n \times E)(y) dy + \int_{\Gamma} (n \times H)(y) \times \nabla_y \phi(x, y) dy \end{aligned} \quad (\text{B.27})$$

Le noyau de Green classique en 3D :

$$\phi(x, y) = \frac{e^{ik|x-y|}}{4\pi|x-y|}$$

La fonction de Green dyadique :

$$G(x, y) = \phi(x, y) I + \frac{1}{k^2} \nabla_y \nabla_y \phi(x, y)$$

Nous avons donc besoin de calculer le gradient du noyau de Green, ainsi que la matrice hessienne :

$$\frac{d\phi(x, y)}{dy_m} = \left\{ \frac{(x_m - y_m)}{4\pi|x-y|^3} - \frac{ik(x_m - y_m)}{4\pi|x-y|^2} \right\} e^{ik|x-y|} \quad (\text{B.28})$$

$$\begin{aligned}
\frac{d^2\phi}{dy_m \partial y_l} &= \left[ -\frac{\delta_{m,l}}{4\pi|x-y|^3} + \frac{3(x_m - y_m)(x_l - y_l)}{4\pi|x-y|^5} + \frac{ik\delta_{m,l}}{4\pi|x-y|^2} - \frac{2ik(x_m - y_m)(x_l - y_l)}{4\pi|x-y|^4} \right. \\
&\quad \left. - \frac{ik(x_m - y_m)(x_l - y_l)}{4\pi|x-y|^4} - k^2 \frac{(x_m - y_m)(x_l - y_l)}{4\pi|x-y|^3} \right] e^{ik|x-y|} \\
&= \left[ \delta_{m,l} (ik - \frac{1}{|x-y|}) + \frac{(x_m - y_m)(x_l - y_l)}{|x-y|} (-k^2 - \frac{3ik}{|x-y|} + \frac{3}{|x-y|^2}) \right] \frac{e^{ik|x-y|}}{4\pi|x-y|^2}
\end{aligned}$$

La formulation variationnelle de l'équation du second ordre en  $E$  s'écrit :

$$-\int_{\Omega} \omega^2 \epsilon_r E \varphi dx + \int_{\Omega} \frac{1}{\mu_r} \text{rot}(E) \text{rot}(\varphi) - \int_{\Sigma} \text{rot}E \times n \cdot \varphi ds = \int_{\Omega} f \cdot \varphi \quad (\text{B.29})$$

Le terme de bord est égal à

$$\begin{aligned}
-\int_{\Sigma} \text{rot}E \times n \cdot \varphi ds &= -ik \int_{\Sigma} \left[ (E \times n) \cdot (\varphi \times n) + H^{pot} \times n \cdot \varphi - (E^{pot} \times n) \cdot (\varphi \times n) \right] \\
&= -ik \int_{\Sigma} (E \times n) \cdot (\varphi \times n) + ik \int_{\Sigma} \left[ -(H^{pot} \times n) \times n + (E^{pot} \times n) \right] \cdot (\varphi \times n) \quad (\text{B.30})
\end{aligned}$$

Sur la figure B.1, on a affiché la SER obtenue avec la condition de Silver-Müller, la condition transparente et la SER analytique. On utilise des éléments finis d'arête de la première famille  $Q4 - Q3$ . Sur la figure B.2, on fait la même expérience pour un rayon plus petit, toujours avec  $Q4 - Q3$ .

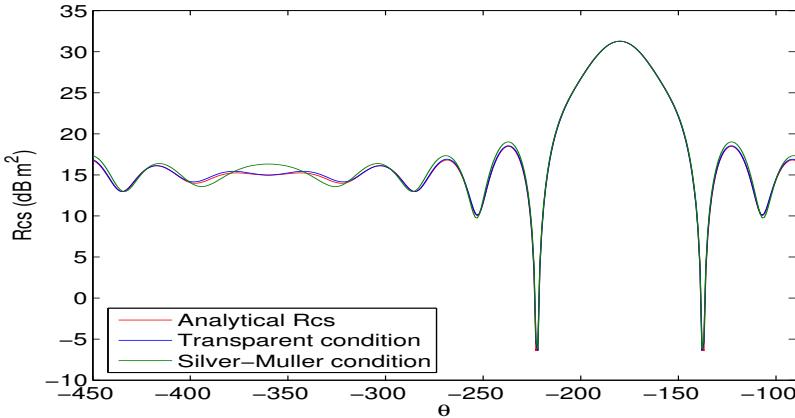


FIG. B.1 – SER d'une sphère parfaitement conductrice de rayon 1.  $k = 2\pi$ , la frontière extérieure du domaine de calcul est placée sur une sphère de rayon 2. En rouge, SER avec la condition transparente. En bleu, SER obtenue analytiquement. En vert, SER obtenue avec la condition de Silver-Müller. Le même maillage est utilisé, on compte 22 000 degrés de liberté.

On valide également le cas diélectrique sur la figure B.3

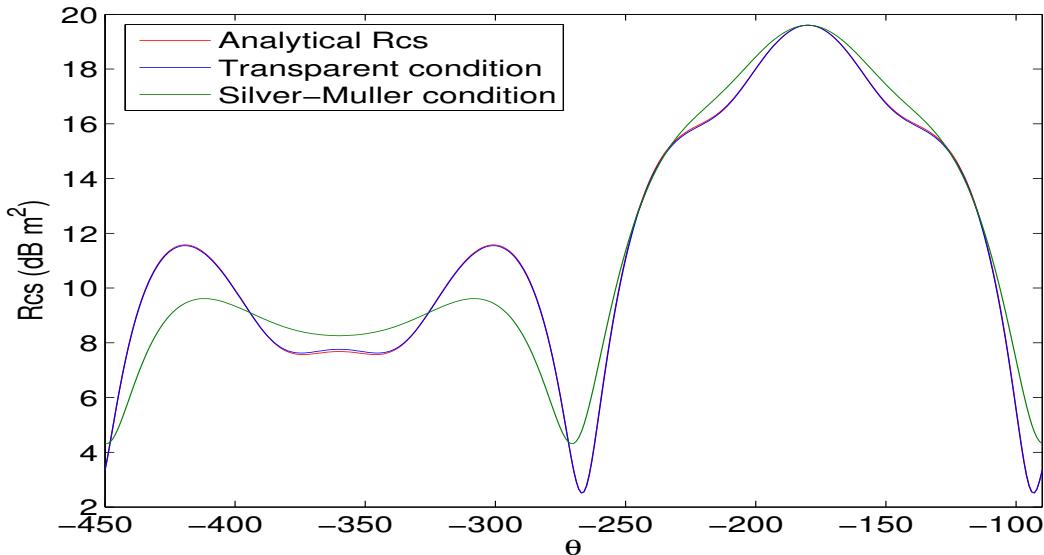


FIG. B.2 – SER d'une sphère parfaitement conductrice de rayon  $0.5 \cdot k = 2\pi$ , la frontière extérieure du domaine de calcul est placée sur une sphère de rayon  $0.6$ . En rouge, SER avec la condition transparente. En bleu, SER obtenue analytiquement. En vert, SER obtenue avec la condition de Silver-Müller. Le même maillage est utilisé, on compte 5 200 degrés de liberté.

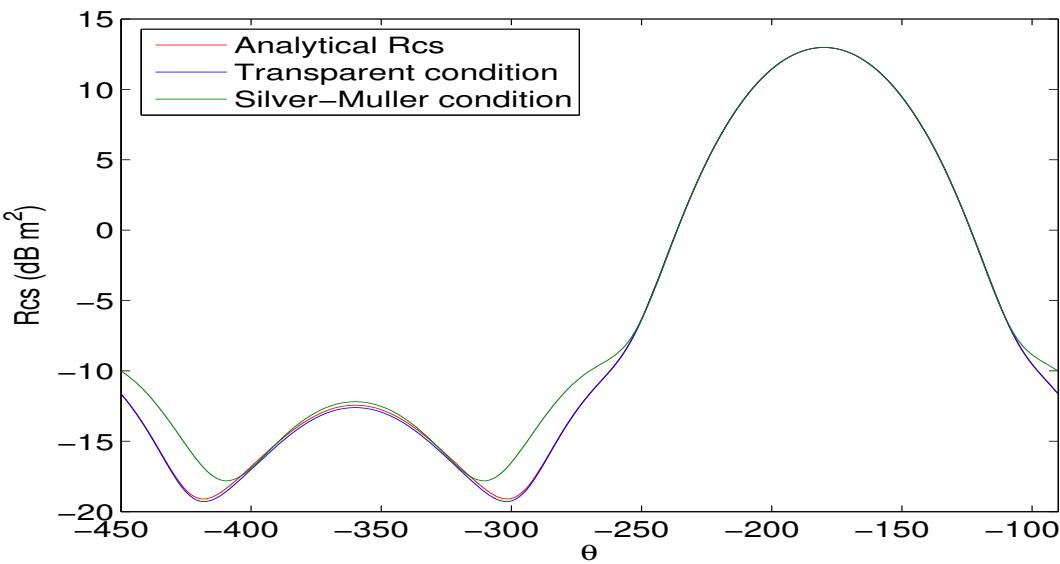


FIG. B.3 – SER d'une sphère diélectrique de rayon  $1.0 \cdot k = \pi \quad \epsilon_r = 1.5$ , la frontière extérieure du domaine de calcul est placée sur une sphère de rayon  $1.5$ . En rouge, SER avec la condition transparente. En bleu, SER obtenue analytiquement. Le maillage utilisé compte 11 000 degrés de liberté.



## Annexe C

# Factorisation discrète et intégration exacte

*Nous rappelons dans cette annexe comment on obtient une factorisation de la matrice de rigidité dans le cas des éléments finis nodaux, lorsqu'on utilise une intégration presque exacte. On effectue la même démarche sur les éléments finis d'arête et pour la formulation Galerkin discontinue. Les algorithmes obtenus ont une complexité en  $O(r^4)$ , comme dans le cas d'une intégration approchée mais les constantes sont légèrement plus élevées.*

### Sommaire

---

C.1	Cas des éléments finis $H^1$ . . . . .	266
C.2	Cas des éléments finis de Nédélec de la première famille . . . . .	268
C.3	Cas de la formulation Galerkin discontinue . . . . .	269

---

Les développements donnés dans cette annexe sont très formels, on ne précisera que rarement l'espace de définition des variables et opérateurs qu'on introduit. Ces développements servent d'un point de vue pratique, pour obtenir des produits matrice-vecteur efficaces. C'est pour cette raison qu'il nous a semblé plus important de garder les idées bases, plutôt que de produire un raisonnement rigoureux avec des notations complexes. On adopte une notation avec des indices  $i, j, k, m$ , qui sont des indices 3-D, on note  $i_1, i_2, i_3$  les trois composantes de l'indice. Les indices  $t, p, q, s$  vont de 1 à 3, ils désignent les composantes des vecteurs. Souvent, on omettra ces indices de composante, ils seront considérés implicites à cause de la structure des objets mis en jeu (matrices 3x3, vecteur à trois composantes).  $r$  désigne l'ordre d'approximation.

## C.1 Cas des éléments finis $H^1$

Rappelons l'expression de la matrice de rigidité élémentaire sur un hexaèdre  $K_e$  :

$$(K_h)_{i,j} = \int_{\hat{K}} J_e DF_e^{-1} \mu DF_e^{-t} \hat{\nabla} \hat{\varphi}_i \cdot \hat{\nabla} \hat{\varphi}_j$$

On utilise les mêmes notations que dans le chapitre 1. Après intégration en utilisant  $(r+1)^3$  points de Gauss (les points sont notés  $\hat{\xi}_k^G$  et les poids  $\omega_k$ ), on obtient :

$$(K_h)_{i,j} = \sum_{k=1}^{(r+1)^3} \omega_k (J_e DF_e^{-1} \mu DF_e^{-t})(\hat{\xi}_k^G) \hat{\nabla} \hat{\varphi}_i(\hat{\xi}_k^G) \cdot \hat{\nabla} \hat{\varphi}_j(\hat{\xi}_k^G)$$

On introduit la matrice  $B_h$  diagonale par blocs 3x3 :

$$(B_h)_{j,k} = \omega_k (J_e DF_e^{-1} \mu DF_e^{-t})(\hat{\xi}_k^G) \delta_{j,k}$$

chaque bloc  $(B_h)_{k,k}$  est une matrice 3x3 symétrique. On introduit la matrice  $\hat{S}$  :

$$\hat{S}_{j,k} = \hat{\nabla} \hat{\varphi}_j(\hat{\xi}_k^G)$$

On sépare le produit matrice vecteur  $K_h U$  en trois étapes :

1.  $V_k = \sum_{j=1}^{(r+1)^3} \hat{\nabla} \hat{\varphi}_j(\hat{\xi}_k^G) U_j$
2.  $W_k = (B_h)_{k,k} V_k$
3.  $Y_i = \sum_{k=1}^{(r+1)^3} \hat{\nabla} \hat{\varphi}_i(\hat{\xi}_k^G) W_k$

On a exhibé la factorisation suivante :

$$K_h = \hat{S} B_h \hat{S}^t$$

Cette factorisation permet d'isoler la géométrie dans la matrice diagonale par blocs  $B_h$ . Elle est donc cruciale lorsqu'on veut construire des algorithmes peu coûteux en stockage.

Cependant, la matrice  $\hat{S}$  est pleine, car les fonctions de base  $\hat{\varphi}_i$  utilisent les points de Gauss-Lobatto, alors que les points de quadrature  $\hat{\xi}_k^G$  sont les points de Gauss. Pour pallier à cet inconvénient, nous utilisons la factorisation suivante de la matrice  $\hat{S}$  :

$$\hat{S} = \hat{C} \hat{R}$$

où

$$\begin{aligned}\hat{C}_{j,k} &= \hat{\varphi}_j(\hat{\xi}_k^G) \\ \hat{R}_{j,k} &= \hat{\nabla} \hat{\varphi}_j^G(\hat{\xi}_k^G)\end{aligned}$$

Cette factorisation vient de l'identité :

$$\hat{\varphi}_j(\hat{x}) = \sum_m \hat{\varphi}_i(\xi_m^G) \hat{\varphi}_m^G(\hat{x})$$

Cette identité est immédiate du fait de l'unicité de l'écriture polynomiale, les fonctions de base  $\hat{\varphi}_i$  et  $\hat{\varphi}_i^G$  constituant une base de l'espace des polynômes  $Q_k$ . En dérivant cette égalité, on obtient alors :

$$\hat{\nabla} \hat{\varphi}_j(\hat{\xi}_k^G) = \sum \hat{\varphi}_i(\xi_m^G) \hat{\nabla} \hat{\varphi}_m^G(\hat{\xi}_k^G)$$

On reconnaît bien le produit matrice vecteur recherché  $\hat{C} \hat{R}$ .

La matrice  $\hat{R}$  est creuse pour la même raison invoquée dans le chapitre 2. La matrice  $\hat{C}$  est pleine, mais du fait de la tensorisation des degrés de liberté, on peut séparer la triple somme en trois sommes simples :

$$\hat{C}U = \sum_{i_1, i_2, i_3=1}^{r+1} u_{k_1, k_2, k_3} \hat{\varphi}_{i_1}(\hat{\xi}_{k_1}^G) \hat{\varphi}_{i_2}(\hat{\xi}_{k_2}^G) \hat{\varphi}_{i_3}(\hat{\xi}_{k_3}^G)$$

est scindé en :

$$\begin{aligned}w1_{k_1, k_2, i_3} &= \sum_{k_3} \hat{\varphi}_{i_3}(\hat{\xi}_{k_3}^G) u_{k_1, k_2, k_3} \\ w2_{k_1, i_2, i_3} &= \sum_{k_2} \hat{\varphi}_{i_2}(\hat{\xi}_{k_2}^G) w1_{k_1, k_2, i_3} \\ w3_{i_1, i_2, i_3} &= \sum_{k_1} \hat{\varphi}_{i_1}(\hat{\xi}_{k_1}^G) w2_{k_1, i_2, i_3}\end{aligned}$$

De manière sous-jacente, on a une factorisation :

$$\hat{C} = \hat{C}_1 \hat{C}_2 \hat{C}_3$$

avec  $\hat{C}_1, \hat{C}_2, \hat{C}_3$  des matrices creuses.

La matrice de masse élémentaire s'écrit :

$$(M_h)_{i,j} = \int_{\hat{K}} J_e \rho \hat{\varphi}_i \hat{\varphi}_j$$

après intégration numérique, elle vaut :

$$(M_h)_{i,j} = \sum_{k=1}^{(r+1)^3} \omega_k (J_e \rho)(\hat{\xi}_k^G) \hat{\varphi}_i(\hat{\xi}_k^G) \hat{\varphi}_j(\hat{\xi}_k^G)$$

De la même manière que pour la matrice de rigidité, on trouve la factorisation suivante :

$$M_h = \hat{C} D_h \hat{C}^t$$

avec la matrice diagonale  $D_h$  :

$$(D_h)_{j,k} = \omega_k (J_e \rho)(\hat{\xi}_k^G) \delta_{j,k}$$

Evaluons maintenant la complexité. Le coût d'un produit matrice vecteur avec  $\hat{C}$  est identique au coût avec  $\hat{R}$ , car on a trois sommes simples à évaluer. Le produit matrice vecteur  $(-\omega^2 D_h + K_h) U_h$  coûte deux fois plus cher, lorsqu'on fait de l'intégration exacte, par rapport à l'intégration approchée.

## C.2 Cas des éléments finis de Nédélec de la première famille

La démonstration est très similaire au cas  $H^1$ . On part de l'expression de la matrice de rigidité élémentaire :

$$(K_h)_{i,j} = \int_{\hat{K}} J_e DF_e^t \mu^{-1} DF_e \hat{\nabla} \times \hat{\varphi}_i \cdot \hat{\nabla} \times \hat{\varphi}_j$$

Après intégration en utilisant  $(r+1)^3$  points de Gauss, on obtient :

$$(K_h)_{i,j} = \sum_k \omega_k (J_e DF_e^t \mu^{-1} DF_e)(\hat{\xi}_k^G) \hat{\nabla} \times \hat{\varphi}_i(\hat{\xi}_k^G) \cdot \hat{\nabla} \times \hat{\varphi}_j(\hat{\xi}_k^G)$$

On introduit la matrice  $B_h$  diagonale par blocs :

$$(A_h)_{j,k} = \omega_k (J_e DF_e^t \mu^{-1} DF_e)(\hat{\xi}_k^G) \delta_{j,k}$$

chaque bloc  $(A_h)_{k,k}$  est une matrice 3x3 symétrique. On introduit la matrice  $\hat{S}$  :

$$\hat{S}_{j,k} = \hat{\nabla} \times \hat{\varphi}_j(\hat{\xi}_k^G)$$

On sépare le produit matrice vecteur  $K_h U$  en trois étapes :

$$1. V_k = \sum_j \hat{\nabla} \times \hat{\varphi}_j(\hat{\xi}_k^G) U_j$$

$$2. W_k = (A_h)_{k,k} V_k$$

$$3. Y_i = \sum_k \hat{\nabla} \times \hat{\varphi}_i(\hat{\xi}_k^G) W_k$$

On a exhibé la factorisation suivante :

$$K_h = \hat{S} B_h \hat{S}^t$$

Cette factorisation permet d'isoler la géométrie dans la matrice diagonale par blocs  $B_h$ . Elle est donc cruciale lorsqu'on veut construire des algorithmes peu coûteux en stockage.

La matrice  $\hat{S}$  est pleine, mais on peut obtenir la factorisation suivante :

$$\hat{S} = \hat{C} \hat{R}$$

où

$$\hat{C}_{j,k} = \hat{\varphi}_j(\hat{\xi}_k^G)$$

$$\hat{R}_{j,k} = \hat{\nabla} \times \hat{\varphi}_j^G(\hat{\xi}_k^G)$$

Cette factorisation vient de l'identité :

$$\hat{\varphi}_j(\hat{x}) = \sum_m \hat{\varphi}_i(\hat{\xi}_m^G) \hat{\varphi}_m^G(\hat{x})$$

Cette identité est vérifiée car l'espace engendré par les fonctions de base  $\hat{\varphi}_j$  est  $Q_{r-1,r,r} \times Q_{r,r-1,r} \times Q_{r,r,r-1}$ . Cet espace est inclus dans  $Q_r^3$ , espace engendré par les fonctions de base  $\hat{\varphi}_m^G$ . En prenant le rotationnel de cette identité, on obtient la factorisation souhaitée.

La matrice  $\hat{R}$  est creuse, le produit matrice vecteur avec  $\hat{R}$  et sa transposée donne une complexité de  $24(r+1)^4$ . Explicitons le produit matrice vecteur  $\hat{C} U$ , pour les fonctions de base orientées suivant  $e_1$  :

$$\hat{C}_1 U = \sum_{k_1, k_2, k_3} u_{k_1, k_2, k_3} \hat{\varphi}_{i_1}^{GA}(\hat{\xi}_{k_1}^G) \hat{\varphi}_{i_2}^{GL}(\hat{\xi}_{k_2}^G) \hat{\varphi}_{i_3}^{GL}(\hat{\xi}_{k_3}^G)$$

où  $GA$  désigne les  $r^3$  points de Gauss ( $G$  désigne toujours les  $(r+1)^3$  points de Gauss). On peut séparer cette triple en trois simples sommes comme on l'a fait dans le cas scalaire. On aura une factorisation du type :

$$\hat{C}_1 = \hat{C}_{11}\hat{C}_{12}\hat{C}_{13}$$

On aura des factorisations semblables pour  $\hat{C}_2$  (fonctions de base orientées suivant  $e_2$ ) et  $\hat{C}_3$ . Les matrices  $\hat{C}_{ij}$  auront comme bonne propriété d'être creuses.

La matrice de masse élémentaire s'écrit :

$$(M_h)_{i,j} = \int_{\hat{K}} J_e DF_e^{-1} \varepsilon DF_e^{-t} \hat{\varphi}_i \hat{\varphi}_j$$

après intégration numérique, elle vaut :

$$(M_h)_{i,j} = \sum_k \omega_k (J_e DF_e^{-1} \varepsilon DF_e^{-t})(\hat{\xi}_k^G) \hat{\varphi}_i(\hat{\xi}_k^G) \hat{\varphi}_j(\hat{\xi}_k^G)$$

De la même manière que pour la matrice de rigidité, on trouve la factorisation suivante :

$$M_h = \hat{C} B_h \hat{C}^t$$

avec la matrice diagonale par blocs 3x3,  $B_h$  :

$$(B_h)_{j,k} = \omega_k (J_e DF_e^{-1} \varepsilon DF_e^{-t})(\hat{\xi}_k^G) \delta_{j,k}$$

Evaluons la complexité du produit matrice vecteur

$$(-\omega^2 M_h + K_h) U_h = [\hat{C} (-\omega^2 B_h + \hat{R} A_h \hat{R}^t) \hat{C}^t] U_h$$

Le cout de  $\hat{R}$  et  $\hat{R}^t$  est de  $24(r+1)^4$ .  $\hat{C}_1 V$  nécessite trois sommes, avec respectivement  $r$  termes,  $r+1$  et  $r+1$  termes, soit une complexité de  $(6r+4)(r+1)^3$ .  $\hat{C}_2$  et  $\hat{C}_3$  demandent chacune ce même nombre d'opérations. Le cout de  $\hat{C}$  et  $\hat{C}^t$  est donc de  $(36r+24)(r+1)^3$ .

La complexité est donc principalement de  $60r^4$  contre  $36r^4$  lorsqu'on fait de l'intégration approchée.

### C.3 Cas de la formulation Galerkin discontinue

Nous nous intéressons au cas d'une formulation Galerkin discontinue (voir chapitre 6). On choisit dans cette partie de prendre comme espace local :

$$V_h = \{\mathbf{u} \in (L^2(\Omega))^3 \text{ tel que } \mathbf{u} \circ F_e \in Q_r^3\}$$

L'avantage de ce choix est d'obtenir des matrices de masse diagonales, par exemple la matrice relative au champ électrique :

$$(M_h)_{j,k}^1 = \varepsilon J_e(\hat{\xi}_k) \omega_k \delta_{j,k}$$

De plus, si on orthonormalise les fonctions de base avec le facteur  $(J_e(\hat{\xi}_k) \omega_k)^{-1/2}$ , on obtient des matrices de masse diagonales et constantes par élément, car elles ne contiennent que des indices physiques  $\varepsilon, \mu, \dots$

Cependant, les matrices de rigidité dépendent alors de la géométrie :

$$(R_h)_{(i,r),(j,s)} = \int_{K_e} \nabla \times (\hat{\varphi}_j e_s) \cdot (\hat{\varphi}_i e_r)$$

Nous allons montrer qu'on peut trouver une factorisation, du type :

$$R_h = B_h \hat{S}$$

avec  $B_h$  matrice diagonale par blocs 3x3, et  $\hat{S}$  une matrice indépendante de la géométrie.

Pour démontrer cette factorisation, nous nous plaçons dans un cadre plus général, avec un opérateur :  $A_1 \frac{\partial}{\partial x_1} + A_2 \frac{\partial}{\partial x_2} + A_3 \frac{\partial}{\partial x_3}$  Dans le cas du rotationnel  $\nabla \times$ , nous avons trivialement :

$$A_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \quad A_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \quad A_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

On a :

$$\nabla U_1 = DF_e^{-t} \hat{\nabla} \hat{U}_1$$

on note la matrice 3x3  $b^k$  :

$$b^k = \omega_k (J_e DF_e^{-t})(\hat{\xi}_k^G)$$

On a donc :

$$\omega_k J_e \frac{\partial \varphi_i}{\partial x_p}(\hat{\xi}_k^G) = \sum_q b_{p,q}^k \frac{\partial \hat{\varphi}_i}{\partial \hat{x}_q}(\hat{\xi}_k^G)$$

La matrice de rigidité s'écrit :

$$(R_h)_{(i,t),(j,s)} = \int_{K_e} \sum_p A_p \frac{\partial \varphi_j e_s}{\partial x_p} \cdot (\varphi_i e_t)$$

soit

$$(R_h)_{(i,t),(j,s)} = \int_{K_e} \sum_p (A_p)_{t,s} \frac{\partial \varphi_j}{\partial x_p} \varphi_i$$

Le produit matrice vecteur  $R_h U$  s'écrit après changement de variables :

$$(R_h U)_{i,t} = \sum_{k,p,q,j,s} (A_p)_{t,s} b_{p,q}^k \frac{\partial \hat{\varphi}_j}{\partial \hat{x}_q}(\hat{\xi}_k^G) \hat{\varphi}_i(\hat{\xi}_k^G) u_{j,s}$$

Les fonctions de base  $\hat{\varphi}_i$  étant associées aux points de Gauss, on obtient  $i = k$ .

$$(R_h U)_{i,t} = \sum_{p,q,j,s} (A_p)_{t,s} b_{p,q}^i \frac{\partial \hat{\varphi}_j}{\partial \hat{x}_q}(\hat{\xi}_i^G) u_{j,s}$$

On va séparer cette intégrale en deux étapes :

$$(v_i)_{(q,s)} = \sum_j \frac{\partial \hat{\varphi}_j}{\partial \hat{x}_q}(\hat{\xi}_i^G) u_{j,s}$$

$$(R_h U)_{i,t} = \sum_{p,q,s} (A_p)_{t,s} b_{p,q}^i (v_i)_{(q,s)}$$

La première étape calcule les dérivées des trois composantes de  $u$  selon les trois variables d'espaces  $\hat{x}_1, \hat{x}_2$  et  $\hat{x}_3$ . La deuxième applique les transformations géométriques pour passer à un élément quelconque du maillage. On a ainsi exhibé la factorisation annoncée.

Explicitons maintenant les matrices de sauts :

$$S_{h(i,t),(j,s)} = \int_{\partial \hat{K}} ds_e \sum_p A_p \hat{\varphi}_j e_s n_p \cdot \varphi_i e_t$$

Soit après intégration sur les points de Gauss de la frontière  $\hat{\zeta}_m^G$ , le produit matrice-vecteur s'écrit :

$$(S_h U) = \sum_{m,p,j,s} \omega_m (n_p ds_e) (\hat{\zeta}_m^G) (A_p)_{t,s} \hat{\varphi}_j(\hat{\zeta}_m^G) \hat{\varphi}_i(\hat{\zeta}_m^G) u_{j,s}$$

On va décomposer cette sommation en trois étapes :

1.  $v_{m,s} = \sum_j \hat{\varphi}_j(\hat{\zeta}_m^G) u_{j,s}$
2.  $w_{m,t} = \sum_{p,s} \omega_m (n_p ds_e) (\hat{\zeta}_m^G) (A_p)_{t,s} v_{m,s}$
3.  $y_{i,t} = \sum_m \hat{\varphi}_i(\hat{\zeta}_m^G) w_{m,t}$

Ce qu'on peut formellement réinterpréter comme une factorisation :

$$S_h = \hat{C} A_h \hat{C}^t$$

La première étape sert à extrapoler la valeur des fonctions de base intérieures aux points de Gauss de la frontière. Du fait de la tensorisation des fonctions de base, la matrice  $\hat{C}$  sera creuse. Plus exactement elle comportera  $3(r+1)^4$  éléments non-nuls. La seconde étape applique les transformations géométriques. Et la troisième étape réalise l'intégration contre les fonctions tests.

Evaluons maintenant la complexité du produit matrice-vecteur avec  $R_h$ . Le calcul  $\hat{S}U$  nécessite  $18(r+1)^4$  opérations car on doit évaluer 9 dérivées, et chaque dérivée demande  $r+1$  multiplications et additions. Le calcul  $B_h V$  nécessite  $36(r+1)^3$  opérations, car les matrices  $A_p$  ont en tout 6 éléments non-nuls.

coût en temps de calcul de  $R_h U$  et  $R_h^t U$  :  $36(r+1)^4 + 72(r+1)^3$

Il faut mettre en comparaison ces complexités avec ce qu'on avait obtenu avec l'utilisation de la transformation  $DF_i^t$  :

coût en temps de calcul de  $R_h U$  et  $R_h^t U$  :  $24r(r+1)^3$

On voit que l'utilisation de cette transformation est avantageuse, car elle conserve le rotationnel quand on passe d'un élément quelconque vers l'élément de référence. On n'a donc pas besoin

d'évaluer les dérivées  $\frac{\partial \hat{U}_1}{\partial \hat{x}_1}, \frac{\partial \hat{U}_2}{\partial \hat{x}_2}, \frac{\partial \hat{U}_3}{\partial \hat{x}_3}$ , ce qui explique la constante 36 au lieu de 24. En ce qui concerne la complexité du produit matrice-vecteur avec  $S_h$  et  $S_h^t$ , on trouve :

coût en temps de calcul de  $S_h U$  et  $S_h^t U$  :  $144(r+1)^3 + 144(r+1)^2$

On trouve là aussi un coût environ 50% supérieur par rapport au cas avec la transformation  $DF_i^t$ . La raison est que cette transformation permet de conserver les normales. Par conséquent, en chaque point de quadrature d'une surface, on a seulement les deux degrés de liberté tangentiels qui vont intervenir pour évaluer les termes de saut, alors qu'en l'absence de cette transformation, les trois degrés de liberté situés autour de ce point vont interagir.

Au niveau du stockage, on obtient un algorithme attractif car on a besoin de ne stocker que les matrices  $DF_i^{-t}$  en chaque point de quadrature, et des matrices 3x3 en chaque point de quadrature de la frontière, frontière qui est partagée par deux hexaèdres. Le stockage est donc de  $9(r+1)^3 + 27(r+1)^2$  pour chaque élément du maillage.

Cet algorithme est intéressant car il permet de traiter les équations de Maxwell dans un cadre plus général des systèmes hyperboliques linéaires (via les matrices  $A_p$ ), mais il est environ 50 % plus coûteux que l'algorithme utilisant la transformation  $DF_i^t$ .



# Bibliographie

- AINSWORTH, M. (2004a). Discrete dispersion relation for hp-version finite element approximation at high wave number. *SIAM Journal on Numerical Analysis*, 42(2):553–575.
- AINSWORTH, M. (2004b). Dispersive properties of high order Nedelec/edge element approximation of the time-harmonic Maxwell equations. *Phil. Trans. Roy. Soc. Series A*, 362:471–493.
- AINSWORTH, M. et COYLE, J. (2003). Hierarchic finite element bases on unstructured tetrahedral meshes. *Computer Methods Applied Mechanical Engineering*, 58(14):2,103–2,130.
- AMESTOY, P., DUFF, I. S., L'EXCELLENT, J.-Y. et KOSTER, J. (2003). Multifrontal Massively Parallel Solver (MUMPS version 4.3), user's guide. Rapport technique, a determiner.
- AMESTOY, P. R., DUFF, I. S., L'EXCELLENT, J.-Y. et LI, X. S. (2000). Analysis, tuning and comparison of two general sparse solvers for distributed memory computers. Rapport technique, a determiner.
- ARNOLD, D. N., BOFFI, D. et FALK, R. S. (2000). Approximation by quadrilateral finite elements. *Math. Computation.*, 71(239):909–922.
- ARNOLD, D. N., BREZZI, F., COCKBURN, B. et MARINI, L. D. (2002). Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM Journal on Numerical Analysis*, 39(5):1749–1779.
- ASSOUS, F., JR, P. C., LABRUNIE, S. et SEGRÉ, J. (2003). Numerical solution to the time-dependent Maxwell equations in axisymmetric singular domains : the singular complement method. *Journal of Computational Physics*, 191:147–176.
- BAYLISS, A., GOLDSTEIN, C. et TURKEL, E. (1983). An iterative method for Helmholtz equation. *Journal of Computational Physics*, 49:443–457.
- BECK, R. et HIPTMAIR, R. (1999). Multilevel solution of the time-harmonic Maxwell's equations based on edge element. *International Journal of Numerical Methods in Engineering*, 45:901–920.
- BENAMOU, J. D. et DESPRES, B. (1996). A domain decomposition method for the Helmholtz equation and related optimal control problems. Rapport de Recherche INRIA 2791, INRIA.
- BOFFI, D., FERNANDES, P., GASTALDI, L. et PERUGIA, I. (1999). Computational models of electromagnetic resonators : analysis of edge element approximation. *SIAM Journal on Numerical Analysis*, 36:1264–1290.
- BOSSAVIT, A. (1998). *Computational electromagnetism*. Academic Press (Boston).
- BUFFA, A. et PERUGIA, I. (2005). Discontinuous Galerkin approximation of the Maxwell eigenproblem. Rapport technique, Technical Report, IMATI-CNR Pavia, Italy.

- CAI, X.-C., CASARIN, M. A., ELLIOT, J. F. W. et WIDLUND., O. B. (1998). Overlapping Schwarz algorithms for solving Helmholtz's equation. *Domain Decomposition Methods 10, AMS Contemporary Mathematics*, 218:391–399.
- CAORSI, S., FERNANDES, P. et RAFFETTO, M. (1999). On the convergence of Galerkin finite element approximations of electromagnetic eigenproblems. *SIAM Journal on Numerical Analysis*, 38:627–649.
- CIARLET, P. (1978). *The finite element method for elliptic problems*. North-Holland.
- CLEMENS, M. et WEILAND, T. (2002). Iterative methods for the solution of very large complex symmetric linear systems of equations in electrodynamics. Fachbereich 18 elektrische nachrichtentechnik, Technische Hochschule Darmstadt.
- COHEN, G. (2002). *Higher-order numerical methods for transient wave equations*. Springer Verlag.
- COHEN, G. et FAUQUEUX, S. (2000). Mixed finite elements with mass-lumping for the transient wave equation. *Journal of Computational Acoustics*, 8:171–188.
- COHEN, G. et MONK, P. (1998). Gauss point mass lumping schemes for Maxwell's equations. *NMPDE Journal*, 14(1):63–88.
- COHEN, G. et MONK, P. (1999). Mur-Nédélec finite element schemes for Maxwell's equations. *Computer Methods Applied Mechanical Engineering*, 169(3-4):197–217.
- COIFMAN, R., ROKHLIN, V. et WANDZURA, S. (1993). The fast multipole method for the wave equation : a pedestrian prescription. *IEEE Antennas and Propagation*, 35:14–19.
- COLLINO, F., GHANEMI, S. et JOLY, P. (1988). Domain decomposition methods for harmonic wave propagation, a general presentation. Rapport de Recherche INRIA 3473, INRIA.
- COLTON, D. et KRESS, P. (1983). *Integral equations methods in scattering*. Wiley and Sons (New York).
- COSTABEL, M. et DAUGE, M. (2002). Weighted regularization of Maxwell equations in polyhedral domains : a rehabilitation of nodal finite element. *Numer. Math.*, 93:239–277.
- DEMKOWICZ, L. (2000). Edge finite elements of variable order for maxwell's equations - a discussion. Rapport technique, Texas institute for Computational and Applied Mathematics.
- der VORST, H. V. (1992). Bi-cgstab : a fast and smoothly converging variant of BiCG for the solution of non-symmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 13:631–644.
- DUFF, I. S. et REID, J. K. (1983). The multifrontal solution of indefinite sparse symmetric linear equations. *ACM Transaction on Mathematical Software*, 9(3):302–325.
- DUFFY, M. (1982). Quadrature over a pyramid or cube of integrands with a singularity at a vertex. *SIAM J. Numer. Anal.*, 19:1260–1262.
- ELMAN, H., ERNST, O. G. et O'LEARY, D. P. (2001). A multigrid method enhanced by Krylov subspace iteration for discrete Helmholtz equations. *SIAM J. Sci. Comput.*, 23(4):1291–1315.
- ELMAN, H. C. et O'LEARY, D. P. (1998). Efficient iterative solution of the three-dimensional Helmholtz equation. *jcp*, 142:163–181.

ELMKIES, A. (1998). *Sur les éléments finis d'arête pour la résolution des équations de Maxwell en milieu anisotrope et pour des maillages quelconques*. Thèse de doctorat, Université de Paris XI Orsay.

EPPSTEIN, D., SULLIVAN, J. M. et UNGOR, A. (2004). Tiling space and slabs with acute tetrahedra. *Computational Geometry Theory And Applications*, 27:237.

ERLANGGA, Y. A. (2002). Some numerical aspects for solving sparse large linear systems derived from the Helmholtz equation. Report of delft university technology, Delft University Technology.

ERLANGGA, Y. A., VUIK, C. et OSTERLEE, C. (2004). A novel multigrid based preconditioner for heterogeneous Helmholtz problems. Report of delft university technology, Delft University Technology.

FAUQUEUX, S. (2003). *Eléments finis mixtes spectraux et couches absorbantes parfaitement adaptées pour la propagation d'ondes élastiques en régime transitoire*. Thèse de doctorat, Université de Paris IX Dauphine.

FREUND, R. W. et NACHTIGAL, N. M. (1991). A quasi-minimal residual method for non-hermitian linear systems. *Numer. Math.*, 60:315–339.

GANDER, M., MAGOULÈS, F. et NATAF, F. (2002). Optimized Schwarz methods without overlap for the Helmholtz equation. *SIAM J. Sci. Comput.*, 24(1):38–60.

GANDER, M. et NATAF, F. (2001). Ailu for Helmholtz problems : a new preconditioner based on the analytic parabolic factorization. *Journal of Computational Physics*, 9(4):1499–1509.

GOPALAKRISHNAN, J., PASCIAK, J. E. et DEMKOWICZ, L. F. (2004). Analysis of a multigrid algorithm for time-harmonic Maxwell equations. *SIAM Journal on Numerical Analysis*, 42: 90–108.

GORDON, W. et HALL, C. (1973). Transfinite element methods : blending functions interpolation over arbitrary element domains. *Numer. Math.*, 21:109–129.

GRAGLIA, R. D., WILTON, D. R. et PETERSON, A. F. (1997). Higher order interpolatory vector bases for computational electromagnetics. *IEEE Transactions on Antennas and Propagation*, 45(3):329–342.

GROB, P. (2006). *Couplage éléments finis spectraux d'ordre élevé - potentiels retardés. Application à la vibro-acoustique instationnaire*. Thèse de doctorat, Université de Paris IX Dauphine.

GUTKNECHT, M. H. et ROZLOZNIK, M. (2001). Residual smoothing techniques : do they improve the limiting accuracy of iterative solvers ? *BIT*, 41:86–114.

HACKBUSCH, W. (1985). *Multigrid methods and applications*. Springer-Verlag.

HACKBUSCH, W. (1994). *Iterative solution of large sparse systems of equations*. Springer Verlag.

HAZARD, C. et LENOIR, M. (1996). On the solution of time-harmonic scattering problems for Maxwell's equations. *SIAM Journal on Numerical Analysis*, 27:1597–1630.

HEIKKOLA, E., ROSSI, T. et TOIVANEN, J. (2003a). Fast direct solution of the Helmholtz equation with a perfectly matched layer/an absorbing boundary condition. *International Journal for Numerical Methods in Engineering*, 57(14):2007–2025.

- HEIKKOLA, E., ROSSI, T. et TOIVANEN, J. (2003b). A parallel fictitious domain method for the three-dimensional Helmholtz equation. *SIAM J. Sci. Comput.*, 24(5):1567–1588.
- HESTHAVEN, J. et WARBURTON, T. (2002). High order unstructured grid methods for time-domain electromagnetics. Rapport technique, Division of Applied Mathematics, Brown University.
- HESTHAVEN, J. et WARBURTON, T. (2004). High order discontinuous Galerkin methods for the Maxwell eigenvalue problem. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 362(1816):493–524.
- HIPTMAIR, R. (1998). Multigrid method for Maxwell's equations. *SIAM Journal on Numerical Analysis*, 36:204–225.
- HIPTMAIR, R. et LEDGER, P. D. (2005). Computation of resonant modes for axisymmetric Maxwell cavities using  $hp$ -version edge finite elements. *Internat. J. Numer. Methods Engrg.*, 62(12):1652–1676.
- HOUSTON, P., PERUGIA, I., SCHNEEBELI, A. et SCHOETZAU, D. (2005). Interior penalty method for the indefinite time-harmonic Maxwell equations. *Numer. Math.*, 100:485–518.
- J. LIU, J. J. (2001). A novel hybridization of higher order finite element and boundary integral methods for electromagnetic scattering and radiation problems. *IEEE Trans Antennas Propagat.*, 49:1794–1806.
- JIN, J. (1993). *The finite element method in electromagnetics*. John Wiley and Sons Inc.
- LACOSTE, P. (2000). Solution of Maxwell equation in axisymmetric geometry by Fourier series decomposition and by use of  $H(\text{rot})$  conforming finite element. *Numer. Math.*, 84(4):577–609.
- LARSSON, E. (1999). A domain decomposition for the Helmholtz equation in a multilayer domain. *SIAM J. Sci. Comput.*, 20(5):1713–1731.
- LEBEDEV, N., LEBEDEV, N. et SILVERMAN, R. (1972). *Special functions and their applications*. Dover Publications, New York.
- LEHOUCQ, R. B., SORENSEN, D. et YANG, C. (1996). Arpack user's guide : solution of large scale eigenvalues problems by implicitly restarted Arnoldi methods. Rapport technique, Available at <http://www.netlib.org>.
- LEVILLAIN, V. (1991). *Couplage éléments finis-équations intégrales pour la résolution des équations de Maxwell en milieu hétérogène*. Thèse de doctorat, Ecole Polytechnique.
- M. COSTABEL, M. (2003). Computation of resonance frequencies for Maxwell equations in non smooth domains. volume 31.
- MONK, P. (2002). *Finite elements methods for Maxwell's equations*. Oxford Science Publication, 2002.
- MONK, P. et PARROTT, K. (1994). A dispersion analysis of finite element methods for Maxwell's equations. *SIAM J. Sci. Comput.*, 15(4):916–937.
- MONK, P. et PARROTT, K. (2001). Phase-accuracy comparisons and improved far-field estimates for 3-d edge elements on tetrahedral meshes. *Journal of Computational Physics*, 170(2):614–641.

- NÉDÉLEC, J. C. (1980). Mixed finite elements in  $\mathbb{R}^3$ . *Numer. Math.*, 35(3):315–341.
- NÉDÉLEC, J. C. (1986). A new family of mixed finite elements in  $\mathbb{R}^3$ . *Numer. Math.*, 51(1):57–81.
- OLSON, L. et HESTHAVEN, J. (2004). Preconditioning high-order discontinuous Galerkin discretizations of the Helmholtz equation. In *a determiner*.
- OTTO, K. et LARSSON, E. (2000). Iterative solution of the Helmholtz equation by a second order method. *SIAM, Journal of Matrix. Anal. Applied*, 21(1):209–229.
- PERNET, S. (2004). *Etude de méthodes d'ordre élevé pour résoudre les équations de Maxwell dans le domaine temporel. Application à la détection et à la compatibilité électromagnétique*. Thèse de doctorat, Université de Paris IX Dauphine.
- PERRUSSEL, R. (2005). *Méthodes multiniveau algébriques pour les éléments finis d'arête. Application à l'électromagnétisme*. Thèse de doctorat, Ecole Centrale de Lyon.
- PERRUSSEL, R., NICOLAS, L. et MUSY, F. (2004). An efficient preconditioner for linear systems issued from the finite-element method for scattering problems. *IEEE Trans. Mag.*, 40:1080–1083.
- PIPERNO, S. (2003). Schémas en éléments finis discontinus localement raffinés en espace et en temps pour les équations de Maxwell 1-d. Rapport de Recherche INRIA 4986, INRIA.
- PIPERNO, S. et FEZOUI, L. (2003). A centered discontinuous Galerkin finite volume scheme for the 3-d heterogeneous Maxwell equations on 3-d unstructured meshes. Rapport de Recherche INRIA 4733, INRIA.
- RAPPETTI, F. et TOSELLI, A. (2002). A feti preconditioner for two dimensionnal edge element approximations of Maxwell's equations on non-matching grids. *SIAM J. Sci. Comput.*, 23(1):92–108.
- RIVIERE, B., WHEELER, M. F. et GIRAUT, V. (1999). Improved energy estimates for interior penalty constrained and discontinuous Galerkin methods for elliptic problems, part i. *Computational Geosciences*, 3:337–360.
- SAAD, Y. (1996). *Iterative methods for sparse linear systems*. Series in Computer Science.
- SAAD, Y. et SCHULTZ, M. H. (1986). Gmres : A generalized minimal residual algorithm for solving nonsymmetric linear systems. *Journal of Scientific Computing*, 16(7):856–869.
- SAUTER, S. et KRAAPP, A. (1996). On the effect of numerical integration in the Galerkin boundary element method. *Numer. Math.*, 74(3):337–360.
- SAUTER, S. et LAGE, C. (2000). Transformation of hypersingular integrals and black-box cubature. *Math. Comp.*, 70(233):223–250.
- SCHWAB, C. et WENDLAND, W. (1992). On the numerical cubatures of singular surface integrals in boundary element methods. *Numer. Math.*, 62:343–369.
- SIMON, J. (2003). *Extension des méthodes multipôles rapides : résolution pour des seconds membres multiples et application aux objets diélectriques*. Thèse de doctorat, Université de Versailles Saint-Quentin-en-Yvelines.
- SINGH, K. et TANAKA, M. (1999). Analytical evaluation of weakly singular integrals for Helmholtz equation. *Transactions of JSSES*.

SOLIN, P., SEGETH, K. et DOLEZEL, I. (2003). *Higher-order finite elements methods*. Studies in Advanced Mathematics, Chapman and Hall.

SONG., J., LU, C., et CHEW, W. C. (1997). Multilevel fast multipole algorithm for electromagnetic scattering. *IEEE Antennas and Propagation*, 45:1488–1493.

TONG, C. H. (1992). A comparative study of preconditioned Lanczos methods for non-symmetric linear systems. Sandia reports, no sand91-8240b, livermore, Sandia National Laboratories.

TOSELLI, A. (1998). Some results on overlapping Schwarz methods for the Helmholtz equation employing perfectly matched layers. Rapport technique, Technical Report 765, Courant Institute, New York University.

TOSELLI, A. (2000). Overlapping Schwarz methods for Maxwell's equations in three dimensions. *Numer. Math.*, 86:733–752.

VANEK, P., BREZINA, M. et MANDEL, J. (1998a). Convergence of algebraic multigrid based on smoothed aggregation. Rapport technique, UCD/CCM Report 110.

VANEK, P., MANDEL, J. et BREZINA, M. (1997). Solving a two-dimensionnal Helmholtz problem using algebraic multigrid. Rapport technique, UCD/CCM Report 110.

VANEK, P., MANDEL, J. et BREZINA, M. (1998b). Two-level algebraic multigrid for the Helmholtz problem. *Contemporary Mathematics*, 18:349–356.

VOLPERT, D. et LEVADOUX, D. (2001). Expertise ser et code axisymétrique pour objets de révolution. Rapport technique, ONERA, département DEMR.

VUIK, C., ERLANGGA, Y. A. et OSTERLEE, C. (2003). Shifted Laplace preconditioners for the Helmholtz equation. Report of delft university technology, Delft University Technology.

WEBB, J. P. (1999). Hierachal vector basis functions of arbitrary order for triangular and tetrahedral finite elements. *IEEE Transactions on Antennas and Propagation*, 47(8):1244–1253.

ZHOU, L. et WALKER, H. F. (1994). Residual smoothing techniques for iterative methods. *SIAM J. Sci. Comput.*, 15:297–312.

ZIENKIEWICZ, O. et TAYLOR, R. (1989). *The finite element method*. McGraw-Hill.

Vu : le Président  
M.....

Vu : les suffragants  
MM.....

Vu et permis d'imprimer :  
le Vice-président du Conseil Scientifique chargé de la Recherche de l'Université de PARIS IX  
DAUPHINE

## Résumé

Dans cette thèse, nous nous intéressons à la résolution des équations de Maxwell en régime fréquentiel, afin de calculer précisément la signature radar de cibles diverses. Pour avoir une grande précision nécessaire pour des expériences de grande taille, nous utilisons des méthodes d'ordre élevé.

Dans le cas scalaire, les éléments finis spectraux hexaédriques avec condensation de masse, permettent d'obtenir un produit matrice vecteur rapide et peu coûteux en stockage. Dans le cas vectoriel, les hexaèdres de la première famille ne réalisent pas la condensation de masse, mais on peut écrire un algorithme rapide de produit matrice-vecteur. Des résultats numériques 3-D montrent la performance de l'algorithme proposé.

Nous traitons également le cas où la géométrie présente une symétrie de révolution. On est alors ramenés à une succession de problèmes 2-D indépendants. Nous proposons une méthode éléments finis d'ordre élevé couplée à des équations intégrales d'ordre élevé.

**Mots clés :** équations de Maxwell, éléments finis d'ordre élevé, méthode de Galerkin discontinue, éléments finis d'arête, axisymétrique, équation de Helmholtz, équations intégrales

## Abstract

In this thesis, time-harmonic Maxwell's equations are our main interest, in order to compute accurately the radar cross section of electromagnetic targets. To have a good precision in large-scale experiments, higher-order finite element methods are used.

In the scalar case, hexahedral spectral finite element, with mass lumping, provide a fast matrix-vector product with a low storage. In the vectorial case, Nedelec's first family hexahedral element doesn't lead to mass lumping, but a fast matrix-vector can be found. Numerical results in 3-D show the efficiency of this algorithm.

The case, where the geometry is invariant under rotation, is treated. This symmetry leads to solve a sequence of 2-D independent problems. A higher-order finite element method is proposed. This method is coupled with higher-order boundary element method.

**Key words :** Maxwell's equations, higher-order finite element methods, Discontinuous Galerkin methods, edge finite-element, axisymmetric, Helmholtz equation, boundary element method