THÈSE de DOCTORAT de l'UNIVERSITÉ PARIS IX DAUPHINE

Spécialité :

MATHÉMATIQUES

présentée

par Olivier Cessenat
pour obtenir le grade de DOCTEUR de l'UNIVERSITÉ PARIS IX DAUPHINE

Sujet de la thèse :
APPLICATION D'UNE NOUVELLE FORMULATION VARIATIONNELLE AUX ÉQUATIONS D'ONDES HARMONIQUES.
PROBLÈMES DE HELMHOLTZ 2D ET DE MAXWELL 3D.

soutenue le 13 Décembre 1996 devant le jury composé de :

M. A. De La Bourdonnaye	Rapporteur
M. J. C. Nédélec	Rapporteur
M. C. Bardos	Examinateur
M. D. Bouche	Examinateur
M. G. Chavent	Examinateur
M. B. Després	Examinateur
M. P. Joly	Examinateur
M. M. Lenoir	Examinateur

A Louis-Robert et Etienne Pelletier.

A Elie Etzer Ben Yehuda.

A mes parents.

I cannot forecast to you the action of Russia. It is a riddle wrapped in a mystery inside an enigma[1].

We shall not flag or fail. We shall fight in France, we shall fight on the seas and oceans, we shall fight with growing confidence and growing strength in the air, we shall defend our island, whatever the cost may be, we shall fight on the beaches, we shall fight on the landing grounds, we shall fight in the fields and in the streets, we shall fight in the hills ; we shall never surrender[2].

Beware, for the time may be short. A shadow has fallen across the scenes so lately lighted by the Allied victory. Nobody knows what Soviet Russia and its Communist international organization intend to do in the immediate future. From Stettin in the Baltic to Trieste in the Adriatic an Iron Curtain has descended across the Continent[3].

Sir Winston Leonard Spencer Churchill.

[1] *Broadcast,* 1939.
[2] *June* 1940.
[3] *Speech,* 1946.

Je remercie Mrs J.C. Nédelec et A. De La Bourdonnaye d'avoir accepté d'être rapporteurs. Messieurs C. Bardos, D. Bouche, G. Chavent et M. Lenoir d'avoir accepté de faire partie du jury.

Je remercie Pr. Patrick Joly de m'avoir pris comme l'un de ses étudiants. Patrick Joly m'a régulièrement suivi et a constamment amélioré la rigueur de mes démonstrations et la clarté de leur exposé. Je lui dois la clarification de beaucoup de notions mathématiques que je n'avais pas bien comprises. J'ai admiré sa capacité à rédiger de façon parfaite en une seule fois... et de me comprendre avant que j'ouvre la bouche. Je lui sais gré pour sa gentillesse, sa modestie, l'aide gracieuse qu'il nous a apportée à Bruno et moi pour l'écriture de l'article. Mieux qu'un long discours, il est plus simple de dire que je souhaite à tout le monde d'avoir Patrick comme professeur.

Je tiens à remercier tout particulièrement Bruno Després qui m'a encadré au sein du CEA et qui a été l'initiateur de cette thèse dont il a posé les fondements. Il a fait preuve d'une attention constante à mon égard durant toute la durée de cette thèse, ne se désintéressant jamais de ce que je faisais, en étant toujours disponible à n'importe quel moment alors qu'il était très occupé. Bruno a toujours été là dans les moments difficiles, m'aura toujours bien conseillé. J'insiste sur sa bonne humeur constante et sa gentillesse qui motivent mon admiration et ma reconnaissance sincères envers lui.

Je remercie Daniel Bouche qui, pendant mon service, m'a mis en contact avec feu le groupe électromagnétisme de Limeil, ainsi que ceux qui m'y ont accueilli : R. Le Martret (qui m'a conduit avec sa Ferrari de nombreuses fois), B. Scheurer, R. Sentis et D. Verwaerde. Je remercie ceux qui ont relu mes écrits pour le compte du service et du département, en particulier B. Scheurer, G. Meurant et D. Bouche.

Je tiens à remercier tous ceux qui m'ont aidé dont la liste est semi exhaustive : Christophe Le Potier (avec une gentillesse infinie), Roland Le Martret (avec une patience infinie), Guilhem Chevalier (Guilhem est un saint qui s'ignore), Pierre Bonnemason (notamment sur les logiciels SER, aussi pour son humour constant), Nicolas L'Hullier (mon spécialiste du C), Isabelle Bertron (ma spécialiste en tout), Emmanuelle Bonneaux (qui m'a relu un nombre infini de fois, initié à Psyche et beaucoup d'autres logiciels, bref mon envoyée spéciale nouveautés informatiques et ma correspondante permanente auprès de la langue française), Olivier Lafitte (pour ses nombreuses relectures), Jean François Clouet (pour son ouverture d'esprit), Christian Quine (pour sa bonne humeur), Bruno Bodin (pour notre conception commune de l'informatique), Julien Pascual (mon spécialiste du C-shell), Bruno Stupfel (pour son humour coriace), Michel Cessenat (pour son livre et ses conseils), Eric Sonnendrücker (mon spécialiste de latex), Eric Puertolas, Emmanuel Frénod (mon ami ennemi du mal), Françoise Angrand (pour son profil (NACA)), Pierre Henri Maire (pour ses nombreuses invitations à dîner et sa charcuterie de Luxeuil)... et Robert North, citoyen de sa majesté, pour sa revue shakespearienne de l'article.

Je remercie aussi tous les anonymes qui m'auront aidé sans le savoir ou sans que je le sache. Entre autres, les mécènes du *free-ware*, latex, maple, et les non mécènes fabricants de logiciels (gnuplot, ideas, islanddraw), de langages ou de machines, notamment la société CRAY R. Inc. Je remercie l'équipe “Psyche”, Y. Demur pour Ideas.

Enfin, *last but not least*, mes parents dont le soutien moral et financier m'aura permis de commencer et de poursuivre cette thèse sans soucis. J'en profite pour, au nom de leurs quatre enfants, les remercier de nous avoir tous permis de faire des études, au sacrifice de leurs propres loisirs.

Abstract

A new technique to solve Elliptic Linear PDEs has been introduced in [27]. It is called Ultra Weak Variational Formulation. This work is devoted to an evaluation of the potentialities of this technique for solving waves problems, in particular the 2-dimensional Helmholtz and the 3-dimensional Maxwell problems. These two problems have wide industrial applications.

For Helmholtz, let us recall the applications to the mine prospection, the sub-marine navigation, the surveillance of structures in nuclear plants or in building, archeological research, more generally to any problem involving detection of cracks or cavities. Furthermore, the simplified scalar Helmholtz problem in a bounded two-dimensional non dissipative constant medium is of pedagogic interest. The study of this problem is presented in a first part following the same style of presentation as the classical one of the Finite Element Method, even though they are definitely conceptually different methods. The first chapter is committed to the variational formulation and to the continuous problem. The second chapter is to define the discretization process using a Galerkin procedure. The third chapter actually studies the efficiency of the technique from the order of convergence point of view. This is achieved using theoretical proofs and a series of numerical experiments. In particular, it is observed and proven that the order of convergence is lower bounded by a linear function of the number of degrees of freedom. Applications to scattering problems are presented in a fourth chapter. At last, an additional chapter presents the generalization of the variational formulation to the Helmholtz problem with varying coefficients.

The second part of this work deals with harmonic electromagnetic problems in a bounded three dimensional complex medium. Amidst many various fields of industrial applications, let us recall this problem interests communications problems, free in space or guided, and also radar detection problems of great military importance (playing a major role during the "battle of Britain") as well as air traffic control. This part, slightly less complete than the first one for its theoretical analysis of the method, shows a large variety of numerical tests and comparisons with classical methods. Nevertheless, the same logic is followed all along this part, underlining the new difficulties brought about by the three dimensional feature of the unknown and of the domain as well as by the new conditions that are the Gauss relations. An important non obvious result deals with the order of convergence of the method in free space with no volumic source of energy. It evolves as a square root of the number of degrees of freedom, which, when dealing with a three dimensional problem, is a good result.

Keywords : Variational formulation, ultra weak, Helmholtz, Maxwell, harmonic waves.

Résumé

Une nouvelle technique de résolution des Equations aux Dérivées Partielles Elliptiques Linéaires a été introduite dans [27]. Nous appellerons cette technique "Formulation Variationnelle Ultra-Faible" (UWVF). L'objet de ce travail est l'étude des potentialités de cette méthode pour les problèmes d'onde, en particulier, les problèmes de Helmholtz et de Maxwell. Ces deux problèmes ont de nombreuses applications industrielles.

Citons pour Helmholtz les problèmes de prospection minière, de navigation sous-marine, de contrôle des structures dans les centrales nucléaires ou pour le bâtiment, de recherche archéologique, de façon générale pour tous les problèmes de détection de fissures ou de cavités. De plus, le cas simplifié de Helmholtz scalaire dans un milieu borné bi-dimensionnel à caractéristiques constantes non dissipatives est intéressant pour ses vertus pédagogiques. Cette étude fait l'objet d'une première partie dont la présentation suit celle de la Méthode des Eléments Finis (FEM). En effet, si les deux méthodes sont conceptuellement différentes, elles sont proches dans la mise en œuvre pratique. Le premier chapitre est dédié à l'étude du problème continu et de la formulation variationnelle. Le deuxième chapitre concerne la technique de discrétisation de type Galerkin. Le troisième chapitre étudie l'efficacité de la méthode du point de vue de l'ordre de convergence. Ceci est réalisé par des démonstrations mathématiques ainsi qu'une large série d'expériences numériques. En particulier, il est observé et prouvé que l'ordre de convergence est minoré par une loi linéaire en fonction du nombre de degrés de liberté. Des applications à des problèmes de diffraction sont présentées dans un quatrième chapitre. Enfin, un dernier chapitre généralise la formulation variationnelle au cas de coefficients variables.

La deuxième partie de ce travail étudie les problèmes harmoniques d'électromagnétisme en domaine borné dans des milieux aux caractéristiques scalaires complexes. Entre autres applications industrielles, citons les problèmes de communications, libres dans l'espace, ou canalisées par un guide d'onde, et les problèmes de détection radar aux nombreuses applications militaires (jouant un rôle majeur pendant "la bataille d'Angleterre") et pour l'aviation civile. Cette partie, légèrement moins complète que la première en ce qui concerne les analyses théoriques de la méthode, présente un grand nombre de simulations numériques et de comparaisons aux méthodes classiques. Néanmoins, l'analyse du problème suit la même démarche logique en présentant les difficultés supplémentaires induites par le caractère tridimensionnel des inconnues et du domaine ainsi que par les conditions supplémentaires que sont les relations de Gauss ou relations de divergence. Un résultat important non trivial concerne l'ordre de convergence de la méthode dans le vide pour un problème sans source volumique. L'ordre de la méthode évolue en racine carrée du nombre de degrés de liberté, ce qui, pour un problème tridimensionnel, constitue un bon résultat.

Mots-clefs : Formulation variationnelle, ultra-faible, Helmholtz, Maxwell, ondes harmoniques.

Table des matières

Introduction.

Ce travail traite de l'application d'une nouvelle formulation variationnelle ultra-faible aux problèmes d'ondes harmoniques. Cette formulation fut proposée pour la première fois par Bruno Després dans [25]. La difficulté essentielle dans la résolution des problèmes d'ondes harmoniques réside dans le fait que la solution est oscillante, gouvernée par la longueur d'onde λ.

Les méthodes classiques actuelles de l'analyse numérique utilisées pour résoudre les problèmes d'ondes harmoniques sont nombreuses. Ces méthodes sont pour la plupart des méthodes générales de résolution de problèmes de physique, nous ne citerons que leurs spécificités dans le cas des problèmes harmoniques de propagation d'ondes.

Les Méthodes des Eléments Finis (FEM) ([54], [17]), des Volumes Finis ([43]) et des Différences Finies ([60]) sont capables de résoudre un problème volumique, en particulier dans des milieux non homogènes. Par exemple, en électromagnétisme, la méthode des éléments finis non conformes ou éléments finis de Nédelec ([48], [49], [47]) prend bien en compte les relations de continuité tangentielle. L'avantage numérique de ces méthodes est de mener à un système linéaire dont la matrice est creuse. En revanche, la matrice du système linéaire de ces techniques n'est généralement pas hermitienne et la mise en œuvre d'un algorithme linéaire itératif convergent d'inversion du système linéaire discret est parfois difficile (selon la répartition des valeurs propres [51]). Par exemple, la méthode du gradient conjugué [18] requiert souvent l'utilisation préalable d'un pré-conditionneur. Néanmoins, l'emploi de méthodes itératives est général et performant, citons notamment GMRes [56] et Bi-CGStab [58]. Une étude comparative extensive a d'ailleurs été menée [22] à la fois sur Helmholtz et sur Maxwell. En revanche, l'inconvénient principal est d'exiger le maillage d'un volume, ce qui est un lourd handicap de par l'augmentation exponentielle du stockage informatique en fonction de la dimension d'espace. Pour la résolution de problèmes dans un milieu infini (comme c'est le cas dans les problèmes de scattering), ces méthodes exigent de tronquer le domaine de calcul. Cela consiste à établir des conditions aux limites absorbantes. La mise en place de conditions aux limites d'ordre élevé performantes est coûteuse et n'a connu que des progrès récents ([5], [19]).

Les Equations Intégrales ([50], [20], [8]) sont bien adaptées aux problèmes de scattering sur des obstacles dans des milieux infinis puisqu'elles tiennent compte exactement de la condition de radiation. Les équations intégrales sont largement utilisées dans le domaine numérique, ne citons que [8]. De plus elles permettent la réduction d'un problème de dimension N à un problème de dimension $N-1$ puisque les calculs sont restreints à la frontière de l'obstacle. Néanmoins ceci a l'inconvénient de restreindre le champ d'application de la méthode à des obstacles dont le comportement peut se modéliser par un opérateur surfacique, par exemple par une condition de Léontovich (ou impédance). Ce type de conditions est utilisable dans un champ d'application limité ([30], [59]). Un inconvénient supplémentaire sur le plan numérique est que la discrétisation conduit à une matrice pleine inversée par des méthodes directes. La mise en œuvre d'algorithmes itératifs est assez difficile, mais il en existe, comme la méthode multi-pôles rapide.

Les avantages respectifs des méthodes d'éléments finis et d'équations intégrales peuvent être rassemblés en couplant ces deux méthodes ([35], [38] et [39]). La méthode couplée est affranchie des problèmes de conditions aux limites absorbantes approchées et peut traiter des objets revêtus d'une couche de diélectrique. Il faut néanmoins une condition aux limites de couplage adaptée aux deux méthodes. Dans ce cadre, les éléments finis mixtes hybrides, permettent d'écrire des conditions de raccord adéquates [12] et l'utilisation de fonctions tests discontinues [52]. Ce type d'éléments est aussi adapté aux méthodes de décomposition de domaine. Le système linéaire discret peut être résolu par un algorithme itératif, comme l'ont fait une équipe de l'Onera, [3] et [2] : leur système linéaire est creux.

La méthode des éléments finis estime l'erreur entre la solution exacte et la solution numérique approchée. Dans la méthode des éléments finis P_k, l'espace de discrétisation est constitué de polynômes, de degré k. L'interpolation à l'aide de polynômes permet de faire une estimation optimale de l'erreur. Pour un problème de Laplacien homogène en bidimensionnel, l'erreur est majorée par une fonction exponentielle de la racine carrée de k le degré de la méthode [17]. En pratique, augmenter le degré k est long à implémenter numériquement, mais ne pose pas de problème théorique.

Ces méthodes, appliquées à la résolution de problèmes d'ondes en fréquence, sont basées sur une formulation variationnelle dont l'opérateur se décompose en la somme d'un opérateur coercif et d'une perturbation compacte de l'opérateur coercif [54]. Leur convergence est conditionnée par le rapport entre le pas de discrétisation h et la longueur d'onde λ par des relations de la forme

$$h \approx \lambda/N$$

où N a une valeur empirique. Par exemple $N = 5$ pour une méthode intégrale sur un problème bidimensionnel, $N = 10$ pour une méthode d'éléments finis toujours dans un domaine bidimensionnel.

Citons d'autres méthodes générales de résolution de problèmes elliptiques ou de discrétisation numérique qui ne sont pas forcément basées sur une formulation variationnelle de la forme d'un opérateur coercif et d'une perturbation compacte.

La méthode d'approximations spectrales [6] est une méthode d'ordre élevé. Cette méthode est basée sur une formulation variationnelle et utilise des polynômes d'ordre N élevé pour l'espace test. L'approximation spectrale est optimale et le degré des polynômes joue le rôle de $1/h$ dans les éléments finis. La méthode spectrale basée sur des formules de collocation est utilisée sur le plan numérique avec de bons résultats pour les problèmes elliptiques. Cette méthode nous semble intéressante et éventuellement pourrait être utilisée dans le cadre de notre formulation variationnelle.

Les méthodes multi-grilles [34], dont le principe est connu depuis le début des années 60, sont des méthodes de résolution numérique effectivement utilisées depuis quelques années. Ce sont des méthodes itératives de résolution basées sur des discrétisations du même problème sur des grilles de tailles différentes. Cette technique est particulièrement efficace lorsque la grille la plus grossière est de taille très réduite, ce qui semble de prime abord particulièrement difficile à réaliser pour des problèmes d'ondes en fréquence.

Les méthodes de décomposition de domaine ([24],[45]) sont des méthodes récentes, développées simultanément à l'apparition des calculateurs à architecture parallèle. Ces méthodes, en pleine extension [32], [4], promettent d'utiliser au mieux les possibilités offertes par les calculateurs de prochaine génération [26]. Elles présentent l'avantage de ne pas être liées à une condition entre la longueur d'onde et la taille de la décomposition du domaine. De plus, parmi toutes les méthodes de résolution numérique, elles proposent des preuves de convergence de la résolution du système linéaire discret ([29], [13]). Nous verrons que la formulation ultra-faible est basée sur des idées de décomposition de domaine.

Pour un résumé pédagogique plus étendu de toutes ces méthodes, nous renvoyons à [36].

Les méthodes asymptotiques sont aujourd'hui les seules réellement capables de résoudre les problèmes d'ondes à haute fréquence ([10], [9]). Leur difficulté est de justifier, sur le plan théorique, les développements effectués ([40]) : ces analyses reposent sur l'analyse micro-locale et sont extrêmement techniques. De plus, la validité de l'application des méthodes asymptotiques à moyenne fréquence est un problème ouvert.

Nous présentons ici une approche nouvelle de résolution des problèmes d'ondes harmoniques. Cette méthode doit son origine à des techniques de décomposition de domaine [24]. Notons que c'est la première fois que cette formulation est utilisée pour la résolution numérique d'équations aux dérivées partielles. Les éléments fondamentaux de la nouvelle formulation sont :

- l'utilisation de fonctions de base discontinues aux interfaces. Ceci explique la terminologie “ultra-faible”. Notons que c'est aussi le cas dans la méthode des éléments finis discontinus (non standards).

– Le problème discret est inconditionnellement bien posé. Il n'y a pas de relation liant la stabilité et l'inversibilité du système linéaire au paramètre de discrétisation.
– Le système linéaire est résolu à l'aide de l'algorithme itératif de Richardson. On montre la convergence de l'algorithme à l'aide d'arguments simples.
– L'ordre de convergence est élevé. Pour un problème bidimensionnel, l'ordre est minoré par une fonction linéaire du nombre de fonctions de base par élément, pour un problème tridimensionnel par une fonction en racine carrée du nombre de fonctions de base par élément. Ceci n'est pas le cas dans la méthode des éléments finis où l'ordre de convergence est en racine carrée du nombre de degrés de liberté pour un problème bidimensionnel [17] et en racine cubique pour un problème tridimensionnel. Ceci signifie que la méthode ultra-faible est asymptotiquement d'ordre plus élevé et nécessite alors un stockage informatique plus faible que la méthode des éléments finis.

L'idée de base de la formulation variationnelle ultra-faible est de réaliser une partition, ou maillage, du domaine de travail. Nous définissons un cadre fonctionnel de fonctions d'énergie finie sur les faces de la partition. Un nouveau problème est posé sur ces faces : la première caractéristique importante de la formulation est de mener à un problème dont les inconnues sont définies sur des interfaces. L'inconnue de la formulation ultra-faible est obtenue à partir des traces tangentielles de la solution au problème harmonique. Plus précisément, c'est une combinaison linéaire des traces tangentielles (en domaine tridimensionnel, des traces normales en bidimensionnel) et de l'application de l'opérateur de Calderon à ces traces. Il est ensuite possible de remonter à la solution du problème harmonique sur ces interfaces. L'utilisation de fonctions de base solutions du problème dual permet d'obtenir un ordre de convergence élevé ; nous ne montrons pas ce résultat de façon générale, mais il nous semble intuitif : les fonctions de base contiennent intrinsèquement l'information concernant la fréquence, au contraire par exemple d'une méthode d'éléments finis où la fréquence n'apparaît pas dans les fonctions de base. Le problème résultant est un système linéaire injectif qui se décompose en la somme de l'opérateur identité (du cadre fonctionnel de la formulation) et d'un opérateur de norme inférieure à l'unité.

Cette formulation est générale, valable pour toute équation linéaire dont le symbole principal est elliptique (cf [27]). De plus, c'est toujours une formulation exacte du problème continu, même si elle est liée à un maillage du domaine. La discrétisation du problème apparaît comme une deuxième étape.

Une caractéristique spécifique aux problèmes d'ondes harmoniques est l'utilisation de fonctions de base issues d'ondes planes ce qui permet un calcul analytique du système linéaire discret. Ceci présente l'intérêt essentiel de permettre l'introduction de connaissances sur la physique du problème dans l'espace des fonctions tests. Ainsi, l'utilisation d'informations sur le comportement asymptotique de la solution est possible [9]. Notons qu'une méthode de discrétisation mettant en œuvre des ondes planes est aussi utilisée pour la discrétisation haute fréquence des équations intégrales [37].

L'utilisation d'un maillage rend la mise en œuvre pratique de notre méthode semblable à celle de la méthode des Eléments Finis, mais sans obligation de respecter une condition entre la longueur d'onde et la taille du maillage.

Cette étude est divisée en deux parties : le problème de Helmholtz en bidimensionnel, puis le problème de Maxwell tridimensionnel dans un milieu diélectrique isotrope.

Première partie

Le problème de Helmholtz bidimensionnel.

Présentation de la première partie.

Cette première partie traite de l'application à un problème modèle de propagation d'ondes harmoniques, le problème de Helmholtz bidimensionnel (ou, en abrégé, $2D$). Ce travail reprend pour beaucoup la note CEA [15] et l'article accepté par *SIAM Numerical Analysis* [16]. Ce problème simplifié est étudié en détail car il permet de comprendre nombre de points difficiles de la problématique générale à la fois de la formulation variationnelle et des équations d'ondes harmoniques. C'est pourquoi nous insisterons sur les difficultés inhérentes à l'étude de tous les problèmes d'ondes harmoniques et sur les caractéristiques générales de la formulation.

Le problème de Helmholtz $2D$ découle de l'équation des ondes après avoir supprimé la dépendance en temps. Dans un domaine Ω borné de frontière Γ assez régulière pour définir la normale presque partout, le problème harmonique se formalise, en normalisant la célérité de l'onde à 1 par

$$\text{(I.0.1)} \qquad \begin{cases} -\Delta u - \omega^2 u = f & \text{dans } \Omega \\ (\partial_\nu + i\omega)u = Q(-\partial_\nu + i\omega)u + g & \text{sur } \Gamma \\ |Q| < 1, \ Q \in \mathbb{C}. \end{cases}$$

La dérivée normale extérieure est notée ∂_ν, la pulsation est notée ω. Les conditions aux limites sont données par l'opérateur Q que l'on supposera constant dans les preuves théoriques et de module strictement inférieur à l'unité. Néanmoins, dans les applications, notons qu'il est possible de poser $Q = -1$ (condition aux limites de Dirichlet) sur une partie de la frontière. De même, il est possible de poser $Q = 1$ (condition aux limites de Neumann) sur une partie de la frontière. L'essentiel est que le problème (I.0.1) soit bien posé et que la solution soit assez régulière (nous préciserons ultérieurement).

Notons qu'une condition aux limites absorbante est donnée par $Q = 0$: c'est le condition aux limites d'ordre le plus bas. Cette condition est connue pour ne pas être très performante, mais nous ne discuterons pas ce point qui ne fait pas l'objet de notre exposé.

Rappelons que les difficultés principales de la résolution des problèmes d'ondes harmoniques résident dans la nature non coercive du problème et dans le fait que la solution est oscillante, gouvernée par la longueur d'onde λ liée à la pulsation ω par

$$\lambda = \frac{2\pi}{\omega} \ .$$

Les méthodes classiques de l'analyse numérique utilisées pour résoudre les problèmes d'ondes harmoniques présentent l'inconvénient d'être basées sur la décomposition en un système de la forme

$$\text{(I.0.2)} \qquad (F - K)u = b$$

où K est une perturbation compacte de l'opérateur coercif F (cf section I.1.1.2). L'injectivité de la matrice du système linéaire du problème discret est conditionnée par le pas de discrétisation (proposition 5). Par exemple, dans la méthode des Eléments Finis (en abrégé : "FEM"), pour un problème bidimensionnel, un pas de discrétisation h vérifiant

$$\text{(I.0.3)} \qquad h \approx \lambda/10$$

est demandé sur le plan pratique pour assurer la stabilité du problème. La méthode par équations intégrales est aussi conditionnée par une telle relation, avec un maillage surfacique de l'objet en $\lambda/5$.

Au contraire des méthodes classiques, la stabilité de la nouvelle formulation n'est en aucun cas conditionnée par une relation entre la longueur d'onde et le pas de discrétisation. De plus, nous verrons que l'ordre de convergence de la méthode est une fonction linéaire du nombre de fonctions de base par élément. Ceci n'est pas le cas dans la méthode des éléments finis où l'ordre de convergence est en racine carrée du nombre de degrés de liberté du problème [17]. Ceci signifie que pour le même taux de précision la méthode issue de la formulation variationnelle ultra-faible (en abrégé : "UWVF") nécessite un stockage informatique plus faible que la FEM.

L'idée de base de la formulation variationnelle ultra-faible est de considérer une nouvelle inconnue ainsi qu'un nouveau problème continu à partir desquels on pourra remonter au problème initial et à sa solution (I.0.1). Pour cela, il convient de réaliser une partition de Ω, c'est-à-dire

$$\left| \begin{array}{l} \overline{\Omega} = \cup \overline{\Omega_k}, \ \Omega_k \cap \Omega_j = \emptyset \text{ for } k \neq j \\ \Gamma_k = \overline{\Omega_k} \cap \Gamma \\ \Sigma_{kj} = \overline{\Omega_k} \cap \overline{\Omega_j} \text{ orienté de } \Omega_k \text{ vers } \Omega_j \\ \partial\Omega_k = (\cup_j \Sigma_{kj}) \cup \Gamma_k \ . \end{array} \right. \tag{I.0.4}$$

En pratique la partition est un maillage du domaine Ω. Les ensembles Ω_k sont les éléments. Nous noterons Σ_{kj} les arêtes des interfaces et Γ_k les arêtes libres (figure 1).

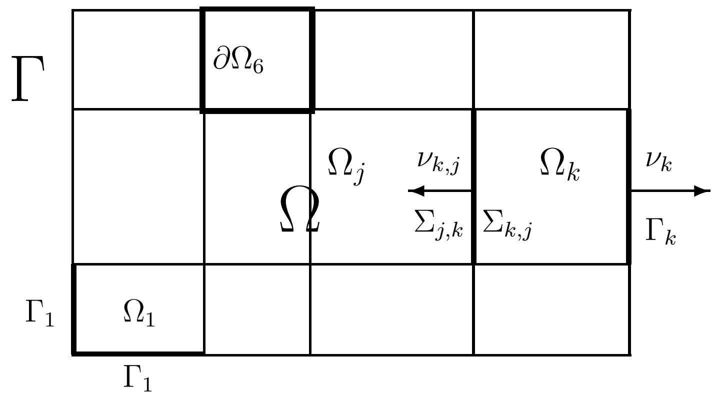


FIG. 1 – Partition d'un domaine Ω en éléments Ω_k

Nous définissons l'espace fonctionnel V de la formulation UWVF comme étant

$$V = \prod_{k=1}^{K} L^2(\partial\Omega_k) \tag{I.0.5}$$

muni du produit scalaire naturel

$$(x, y) = \sum_k \int_{\partial\Omega_k} x_{|\partial\Omega_k} \overline{y_{|\partial\Omega_k}} \tag{I.0.6}$$

qui définit la norme $||.||_V$ et la norme induite d'un opérateur $A \in V$ par :

$$||A|| = \sup_{x \neq 0} \frac{||Ax||_V}{||x||_V} \ . \tag{I.0.7}$$

Remarque 1 Le cadre fonctionnel V dépend du maillage, mais ce n'est pas un espace de discrétisation de dimension finie. Nous devrions nous y référer en notant l'indice K du maillage. Nous garderons néanmoins la notation V, adoptée par [25] et [15] pour ne pas surcharger ce document.

Notons ∂_{ν_k} la dérivée normale extérieure $\partial\Omega_k$. La valeur de l'inconnue notée x de la formulation UWVF est définie à partir de u la solution de (I.0.1) comme étant :

$$x_{|\partial\Omega_k} = (-\partial_{\nu_k} + i\omega) u_{|\partial\Omega_k} \ , \tag{I.0.8}$$

en faisant l'hypothèse de régularité

(I.0.9) $$(-\partial_{\nu_k} + i\omega)u_{|\partial\Omega_k} \in V \ .$$

Cette hypothèse est valable dès que la solution u au problème (I.0.1) est suffisamment régulière. Ainsi, la première caractéristique importante de la formulation UWVF est de mener à un problème dont les inconnues sont définies sur les interfaces. Le problème résultant est un système linéaire de la forme

(I.0.10) $$\begin{cases} \text{pour } b \in V, \text{ trouver } x \in V \\ (I - A)x = b \end{cases}$$

où A est un opérateur linéaire dans V vérifiant $||A|| \leq 1$ et $(I - A)$ inversible. Soulignons que cette formulation est générale à toute équation aux dérivées partielles (EDP) linéaire elliptique (cf [27]).

La présentation de la formulation est organisée dans cette partie en cinq chapitres dont nous donnons ci-dessous les points essentiels.

– La section I.1 présente la nouvelle formulation. Sous l'hypothèse de régularité : $\forall k \ \dfrac{\partial u}{\partial \nu_k} \in L^2(\partial\Omega_k)$, nous définissons x et V par (I.0.8) et (I.0.5). Nous montrons (théorème 3) que, moyennant un opérateur de relèvement, la nouvelle formulation variationnelle est équivalente au problème de Helmholtz (I.0.1). Ceci montre que la formulation UWVF est bien posée. L'espace test de la formulation est constitué de fonctions z, dont les restrictions z_k à $\partial\Omega_k$ sont construites à partir de fonctions e, données par

  (I.0.11) $$\begin{cases} (-\Delta - \omega^2)e_k = 0 & \text{dans } \Omega_k \\ (-\partial_\nu + i\omega)e_k = z_k & \text{sur } \partial\Omega_k \ . \end{cases}$$

  Ces équations sont les équations adjointes homogènes (i.e. $f = 0$ dans Ω) des équations du problème de Helmholtz (I.0.1). On montre alors que la formulation a la propriété suivante.

  Proposition 1 *Il existe un opérateur A dans $\mathcal{L}(V)$ et b dans V tels que $(I - A)x = b$ et*

  (I.0.12) $$\begin{cases} ||A|| \leq 1 & \textit{(proposition 6)} \\ (I - A) \textit{ est injectif} & \textit{(proposition 7).} \end{cases}$$

– La section I.2 présente le procédé de discrétisation par une méthode de Galerkin. La solution approchée x_h est cherchée dans un sous-espace V_h de V, V_h étant de dimension finie. Nous montrons comment définir et calculer une solution approchée u_h de u la solution du problème de départ (I.0.1) à partir de x_h. L'espace de Galerkin V_h est construit à l'aide de p fonctions de base $z_{kl,l=1..p}$ sur tout élément Ω_k (avec k variant de 1 à K) qui sont des solutions du problème homogène dual (I.0.11) et de support inclus dans Ω_k. Le problème discret s'écrit formellement

  (I.0.13) $$(I - P_h A)x_h = P_h b$$

  et mène au système linéaire de dimension finie pK :

  (I.0.14) $$(D - C)X = b \ ,$$

  où $X = (x_{kl})_{kl} \in \mathbb{C}^{pK}$ et $(x_h)_{|\partial\Omega_k} = \sum_{l=1}^{p} x_{kl} z_{kl}$. Nous affirmons alors

  Proposition 2 *Le système (I.0.14) est inversible pour tout paramètre de discrétisation h (théorème 5). La matrice $(D - C)$ du système linéaire est creuse.*

  L'utilisation d'ondes planes pour la construction des fonctions de base permet un calcul par une formule analytique des coefficients des matrices D et C. De par le fait que $||A|| \leq 1$, on montre qu'il existe un algorithme itératif convergent pour résoudre le système linéaire discret. Remarquons que l'on commande la qualité de l'approximation à l'aide de deux paramètres : le paramètre de taille du maillage (qui définit l'espace de la formulation continue V) et le nombre de fonctions de base.

– La section I.3 est dédiée à l'étude de l'ordre de convergence de la méthode. Cette section est divisée en deux parties. Une partie numérique montre le comportement "linéaire" [1] de l'ordre de convergence en fonction du nombre de fonctions de base par élément. Une partie théorique mène aux deux résultats essentiels que sont les deux propositions suivantes.

[1] En fait, la loi est linéaire en fonction de $[p/2]$, la partie entière du nombre de fonctions de base divisé par deux.

Proposition 3 *Soit u une solution du problème (I.0.1) homogène ($f = 0$), solution assez régulière pour que x défini par (I.0.8) vérifie $x \in V$ (I.0.5). Soit x_h une solution de (I.0.13). Notons $[\alpha]$ la partie entière de α. Nous supposons que u est de classe $C^{[(p+1)/2]}(\Omega)$ où $p \geq 3$ est le nombre de fonctions de base par élément. Nous supposons aussi que les fonctions de base sont des ondes planes aux directions fixées, identiques pour tous les éléments. Alors,*

$$||x - x_h||_{L^2(\Gamma)} \leq Ch^{[(p-1)/2]-1/2}||u||_{C^{[(p+1)/2]}(\Omega)} \tag{I.0.15}$$

L'ordre de convergence, défini comme l'exposant de h est ainsi $[(p-1)/2] - 1/2$. C'est une fonction "linéaire" de p. Par exemple $p = 3$ ou $p = 4$ donnent $h^{1/2}$.

A l'aide d'une technique de dualité on montre l'existence d'une loi de convergence linéaire dans le cas non homogène ($f \neq 0$) dans un espace de Sobolev d'exposant négatif.

Proposition 4 *Soit u une solution du problème général (I.0.1), solution assez régulière pour que x défini par (I.0.8) vérifie $x \in V$ (I.0.5). Soit x_h une solution de (I.0.13). Nous supposons que u est de classe $C^2(\Omega)$ et $p \geq 3$ est le nombre de fonctions de base par élément, les fonctions de base étant des ondes planes aux directions fixées, identiques pour tous les éléments, alors,*

$$\forall s > [(p-1)/2] - 1/2, \ ||x - x_h||_{H^{-s}(\Gamma)} \leq Ch^{[(p-1)/2]}||u||_{C^2(\Omega)} \ . \tag{I.0.16}$$

L'ordre de convergence est ainsi $[(p-1)/2]$. C'est une fonction linéaire de p. Par exemple $p = 3$ ou $p = 4$ donnent h^1.

- La section I.4.2 présente diverses simulations numériques : deux cas tests où l'on calculera les valeurs de u, deux cas tests de scattering avec des calculs de Section Efficace Radar (SER).
- Nous présentons la formulation variationnelle sur un problème de Helmholtz à coefficients variables section I.5.

Chapitre I.1

Présentation de la Formulation ultra-faible.

I.1.1 Rappels sur le problème modèle dans le vide.

Nous utiliserons dans toute cette étude la convention de notation suivante

Notation 1

1. *Le domaine Ω est un ouvert borné de $\mathbb{R}^2$.*
2. *Le bord Γ est la frontière "assez régulière" de Ω. La régularité de Γ sera en général Lipschitz continue par morceaux, nous supposerons Γ plus régulière en fonction des résultats à obtenir.*
3. *La dérivée normale est notée ∂_ν (c'est le produit scalaire de la normale extérieure avec le gradient).*
4. *La pulsation est notée ω. Par adimensionnement du vecteur d'onde $\vec{k}$ en prenant $c = 1$ la célérité de l'onde, ω sera aussi le nombre d'onde ($\omega \in \mathbb{R}$).*
5. *Le second membre de l'équation dans le domaine Ω est noté f. La fonction source volumique f est dans un espace fonctionnel donné, précisé ultérieurement.*
6. *La condition sur le bord Γ est donnée par une fonction notée g.*
7. *L'opérateur de bord Q est une fonction réelle qui relie les traces d'énergie entrante et sortante sur Γ.*

I.1.1.1 Problématique.

Le problème de Helmholtz bidimensionnel est le suivant. Pour Ω un domaine borné de $\mathbb{R}^2$ de frontière Γ assez régulière, pour f une fonction d'énergie finie dans Ω, pour g d'énergie finie sur le bord Γ et pour Q un opérateur de bord de norme inférieure à 1, trouver u dans Ω vérifiant, à la pulsation $\omega \in \mathbb{C}$, $\omega \neq 0$,

$$(-\Delta - \omega^2)u = f$$

et sur le bord Γ

$$(\partial_\nu + i\omega)u = Q(-\partial_\nu + i\omega)u + g\ .$$

I.1.1.2 Résultats classiques.

Rappelons les résultats classiques suivants.

Théorème 1 (Résultat d'existence et d'unicité) *Soit Ω un ouvert borné de frontière Γ de classe C^1. Soit $f \in L^2(\Omega)$ et $g \in H^{1/2}(\Gamma)$. On pose $\zeta = \dfrac{1-Q}{1+Q}$ et on suppose Q constant, $|Q| < 1$ (alors*

$\Re(\zeta) > 0$). *Alors, il existe un unique* $u \in H^1(\Omega)$ *tel que*

$$\text{(I.1.1)} \qquad \begin{cases} \forall v \in H^1(\Omega) \\ \displaystyle\int_\Omega \nabla u \, \overline{\nabla v} - \omega^2 \int_\Omega u \, \overline{v} + i\omega\zeta \int_\Gamma u \, \overline{v} = \int_\Omega f \, \overline{v} + \frac{1}{1+Q} \int_\Gamma g \, \overline{v} \, . \end{cases}$$

Ce problème est la formulation variationnelle de : pour $f \in L^2(\Omega)$ *et* $g \in H^{1/2}(\Gamma)$, *trouver* $u \in H^1(\Omega)$ *unique tel que*

$$\text{(I.1.2)} \qquad \begin{cases} (-\Delta - \omega^2)u = f & \textit{dans } \Omega \\ (\partial_\nu + i\omega)u = Q(-\partial_\nu + i\omega)u + g & \textit{sur } \Gamma \, . \end{cases}$$

Preuve. On se rapporte à [24]. □

Citons un résultat supplémentaire.

Théorème 2 (Résultat de régularité) *Soit* Ω *un ouvert borné de* $\mathbb{R}^n$ *de frontière* Γ. *On suppose que* Γ *est une variété indéfiniment différentiable de dimension* $n-1$ *et que* Ω *est localement d'un seul côté de* Γ. *Soit* $s \in \mathbb{R}, s \geq 2$ *et* $f \in H^{s-2}(\Omega)$ *et* $g \in H^{s-3/2}(\Gamma)$. *Alors il existe un unique* $u \in H^s(\Omega)$ *solution de :*

$$\text{(I.1.3)} \qquad \begin{cases} (-\Delta - \omega^2)u = f & \textit{dans } \Omega \\ (\partial_\nu + i\omega)u = g & \textit{sur } \Gamma \, . \end{cases}$$

De plus, on a l'estimation suivante

$$\text{(I.1.4)} \qquad ||u||_{H^s(\Omega)} \leq C \left\{ ||f||_{H^{s-2}(\Omega)} + ||g||_{H^{s-3/2}(\Gamma)} \right\} \, .$$

Enfin, si $f = 0$ *l'estimation (I.1.4) reste vraie pour tout* s *réel.*

Preuve. On se rapporte à [44] p.202 remarque 7.2. □

Proposition 5 (Un résultat classique d'erreur numérique) *On considère un maillage régulier ([17]) définissant un paramètre de taille de la discrétisation* h *"assez petit". Ce maillage permet la discrétisation par la méthode des éléments finis* P_1 *de la formulation variationnelle (I.1.1). On suppose que la solution* u *du problème est* $H^2(\Omega)$. *Alors, il existe* C_1 *et* C_2 *deux constantes strictement positives indépendantes de* h *telles que*

$$\text{(I.1.5)} \qquad ||u - u_h||_{L^2(\Omega)} \leq \frac{C_1 h^2}{1 - \omega^2 h^2 C_2} ||u||_{H^2(\Omega)} \, .$$

Preuve. On se rapporte à [25], [17] ou [36]. Nous ne présentons pas tous les éléments de la preuve qui font appel à beaucoup trop de notions sur les éléments finis. Nous supposons connus l'espace d'éléments finis $\mathcal{W}_h$ et l'opérateur d'interpolation Π_h.

Pour une triangulation régulière, on sait estimer l'erreur d'interpolation en norme H^1 par rapport à la norme H^2 de la solution du problème variationnel (I.1.1) ([17], corollaire 7 de [36]) :

$$\exists C_a > 0 / \; ||u - \Pi_h u||_{H^1(\Omega)} \leq C_a h ||u||_{H^2(\Omega)} \, .$$

D'autre part, on majore la norme H^2 de u par la norme L^2 de f grâce à (I.1.4) (pour $s = 2$). On déduit aisément du lemme d'Aubin-Nitsche ([17] p. 137 théorème 3.2.4 avec $V = L^2$ et $H = H^1$) qu'il existe $C > 0$ tel que

$$||u - u_h||_{L^2(\Omega)} \leq Ch||u - u_h||_{H^1(\Omega)} \, .$$

La formulation variationnelle (I.1.1) s'écrit sous la forme

$$\text{(I.1.6)} \qquad \begin{cases} \forall v \in H^1(\Omega) \\ a(u,v) - \omega^2 b(u,v) = l(v) \end{cases}$$

où a est coercif sur H^1 et b est une forme sesquilinéaire continue :

$$\begin{cases} a(u,v) = \displaystyle\int_\Omega \nabla u \, \overline{\nabla v} + \int_\Omega u \, \overline{v} + i\omega\zeta \int_\Gamma u \, \overline{v} \\ b(u,v) = -\dfrac{(1+\omega^2)}{\omega^2} \displaystyle\int_\Omega u \, \overline{v} \\ l(v) = \displaystyle\int_\Omega f \, \overline{v} + \int_\Gamma g \, \overline{v} \; . \end{cases}$$

Alors, pour tout $v = v_h \in \mathcal{W}_h \subset H^1(\Omega)$ on a

$$a(u-u_h, v_h) - \omega^2 b(u-u_h, v_h) = 0 \; .$$

En particulier, pour $v_h = (\Pi_h u - u) + (u - u_h) \in \mathcal{W}_h$, on a

$$a(u-u_h, u-u_h) - \omega^2 b(u-u_h, u-u_h) = a(u-u_h, u-\Pi_h u) - \omega^2 b(u-u_h, u-\Pi_h u) \; .$$

La coercivité de a et les continuités de a et b, qui s'expriment par,

$$\left| \begin{array}{l} \exists \alpha \in \mathbb{R} - \{0\} / \; \forall v \in H^1(\Omega), \;\; a(v,v) \geq \alpha ||v||^2_{H^1(\Omega)} \\ \exists M \in \mathbb{R} - \{0\} / \; \forall (u,v) \in (H^1(\Omega))^2, \;\; a(u,v) \leq M ||u||_{H^1(\Omega)} ||v||_{H^1(\Omega)} \\ \exists C_b \in \mathbb{R} - \{0\} / \; \forall (u,v) \in (L^2(\Omega))^2, \;\; |b(u,v)| \leq C_b ||u||_{L^2(\Omega)} ||v||_{L^2(\Omega)} \; , \end{array} \right.$$

impliquent

$$\begin{aligned} \alpha ||u-u_h||^2_{H^1(\Omega)} - \omega^2 C_b ||u-u_h||^2_{L^2(\Omega)} \leq & M ||u-u_h||_{H^1(\Omega)} ||u-\Pi_h u||_{H^1(\Omega)} \\ & + \omega^2 C_b ||u-u_h||_{L^2(\Omega)} ||u-\Pi_h u||_{L^2(\Omega)} \; . \end{aligned}$$

Or, pour un maillage régulier, il existe C_e et C_d ([36], lemme 13 relation 45) tels que

$$||u-\Pi_h u||_{L^2(\Omega)} \leq C_e h^2 ||u||_{H^2(\Omega)} \text{ et } ||u-\Pi_h u||_{H^1(\Omega)} \leq C_d h ||u||_{H^2(\Omega)} \; .$$

Alors,

$$(\alpha - C_b C^2 \omega^2 h^2) ||u-u_h||^2_{H^1(\Omega)} \leq ||u-u_h||_{H^1(\Omega)} (M C_d h + \omega^2 C_b C h C_e h^2) ||u||_{H^2(\Omega)}$$

ce qui montre que pour h assez petit, en posant $C_1 > M C_d / \alpha$ et $C_2 = C_b C^2 / \alpha$, on a la relation (I.1.5). □

La relation (I.1.5) établit la condition de stabilité $C_2 \omega^2 h^2 \leq 1$, condition qui se réécrit

$$\frac{\lambda}{h} \geq 2\pi \sqrt{C_2} \; . \qquad \text{(I.1.7)}$$

Il est connu empiriquement que si $\dfrac{\lambda}{h} \geq 10$, alors la condition de stabilité (I.1.7) est vérifiée et la consistance des calculs est suffisante pour obtenir des résultats précis.

I.1.2 Etude de la formulation variationnelle.

I.1.2.1 La formulation variationnelle.

Le théorème 3 spécifie l'équivalence entre le problème originel et notre formulation dont les inconnues sont définies sur les arêtes $\partial\Omega_k$. Cette méthode apparaît comme une nouvelle formulation variationnelle (I.1.25) puis (I.1.27) par le théorème 4. Les propriétés du système (I.1.27) sont résumées par les propositions 6 et 7.

Théorème 3 *Soit $u \in H^1(\Omega)$ une solution du problème de Helmholtz (I.0.1) vérifiant l'hypothèse de régularité $\partial_{\nu_k} u \in L^2(\partial\Omega_k)$ pour tout k. Alors x défini par $x_{|\partial\Omega_k} = x_k$ avec $x_k = ((-\partial_{\nu_k} + i\omega)u_{|\Omega_k})_{|\partial\Omega_k}$ vérifie :*

$$\begin{aligned}
&\left(\sum_k \int_{\partial\Omega_k} x_k\, \overline{(-\partial_{\nu_k} + i\omega)e_k}\right) \\
&- \left(\sum_{k,j} \int_{\Sigma_{kj}} x_j\, \overline{(\partial_{\nu_k} + i\omega)e_k} + \sum_k \int_{\Gamma_k} Q x_k\, \overline{(\partial_{\nu_k} + i\omega)e_k}\right) \\
&= -2i\omega \sum_k \int_{\Omega_k} f\, \overline{e_k} + \sum_k \int_{\Gamma_k} g\, \overline{(\partial_{\nu_k} + i\omega)e_k}\ .
\end{aligned} \tag{I.1.8}$$

pour tout

$$e \in H,\ \ e = (e_k)_{k=1\ldots K},\ \ H = \prod_{k=1}^{K} H_k \tag{I.1.9}$$

avec

$$H_k = \left\{ v_k \in H^1(\Omega_k) \,\middle|\, \begin{array}{l} (-\Delta - \omega^2)v_k = 0 \ \ dans\ \Omega_k \\ (-\partial_{\nu_k} + i\omega)(v_k)_{|\partial\Omega_k} \in L^2(\partial\Omega_k) \end{array} \right\}\ . \tag{I.1.10}$$

Réciproquement, si x est solution de (I.1.8) alors la fonction u définie par

$$\left| \begin{array}{l} u_{|\Omega_k} = u_k\ , u_k \in H^1(\Omega_k) \\ (-\Delta - \omega^2)u_k = f_{|\Omega_k} \\ (-\partial_\nu + i\omega)u_k = x_k\ . \end{array} \right. \tag{I.1.11}$$

est la solution unique du problème de Helmholtz dans Ω (I.0.1).

Preuve. Par hypothèse, $u \in H^1(\Omega)$, donc la trace de u sur toute surface régulière $\partial\Omega_k$ de Ω est $u_{|\partial\Omega_k} \in H^{1/2}(\partial\Omega_k)$. On suppose aussi que $\partial_\nu u \in V$. Cela nous permet d'écrire

$$\begin{aligned}
&\int_{\partial\Omega_k} (-\partial_{\nu_k} + i\omega)u\, \overline{(-\partial_{\nu_k} + i\omega)e_k} \\
&= \int_{\partial\Omega_k} (+\partial_{\nu_k} + i\omega)u\, \overline{(+\partial_{\nu_k} + i\omega)e_k} - 2i\omega \int_{\partial\Omega_k} \left(u\,(\overline{\partial_{\nu_k} e_k}) - (\partial_{\nu_k} u)\,\overline{e_k}\right)
\end{aligned} \tag{I.1.12}$$

en développant le produit puis en regroupant. Rappelons que, d'après (I.0.1) et (I.1.10), on a les équations volumiques

$$\begin{cases} a)(-\Delta - \omega^2)u = f & \text{dans } \Omega\ , \\ b)(-\Delta - \omega^2)\,\overline{e}_k = 0 & \text{dans } \Omega_k\ . \end{cases} \tag{I.1.13}$$

Par des intégrations par partie dans (I.1.13), contre $\overline{e}_k$ dans $a)$ et contre u dans $b)$, on obtient

$$\begin{cases} \displaystyle\int_{\Omega_k} \nabla u\, \overline{\nabla e_k} - \omega^2 u\, \overline{e}_k = \int_{\partial\Omega_k} u(\overline{\partial_{\nu_k} e_k}) \\ \displaystyle\int_{\Omega_k} \nabla u\, \overline{\nabla e_k} - \omega^2 u\, \overline{e}_k - \int_{\Omega_k} f\, \overline{e}_k = \int_{\partial\Omega_k} (\partial_{\nu_k} u)\, \overline{e}_k\ . \end{cases} \tag{I.1.14}$$

En remplaçant les expressions de $\int_{\partial\Omega_k} u(\overline{\partial_{\nu_k} e_k})$ et $\int_{\partial\Omega_k} (\partial_{\nu_k} u)\,\overline{e}_k$ données par (I.1.14) dans la relation (I.1.12), on a

$$\begin{aligned}
&\int_{\partial\Omega_k} (-\partial_{\nu_k} + i\omega)u\, \overline{(-\partial_{\nu_k} + i\omega)e_k} \\
&\qquad - \int_{\partial\Omega_k} (+\partial_{\nu_k} + i\omega)u\, \overline{(+\partial_{\nu_k} + i\omega)e_k} = -2i\omega \int_{\Omega_k} f\, \overline{e}_k\ .
\end{aligned} \tag{I.1.15}$$

Remarquons que la continuité de u sur Σ_{kj} et la condition au bord de (I.0.1) s'écrivent :

$$\text{(I.1.16)} \qquad \begin{cases} (+\partial_{\nu_k} + i\omega)u_{|\Sigma_{kj}} = (-\partial_{\nu_j} + i\omega)u_{|\Sigma_{jk}} \\ (+\partial_{\nu_k} + i\omega)u_{|\Gamma_k} = Q(-\partial_{\nu_k} + i\omega)u_{|\Gamma_k} + g \ . \end{cases}$$

On peut alors sommer sur k l'indice d'élément Ω_k dans (I.1.15), puis remplacer la somme sur Ω_k dans le terme $\int_{\partial\Omega_k}(+\partial_{\nu_k} + i\omega)u\,\overline{(+\partial_{\nu_k} + i\omega)e_k}$ par une somme sur Γ_k et Σ_{kj}. L'utilisation des relations (I.1.16) donne alors directement l'équation (I.1.8).

Réciproquement, soit x une solution de (I.1.8) et u défini par (I.1.11) pour toutes ses restrictions à Ω_k. Par hypothèse sur u et e, on a (I.1.13) qui conduit à (I.1.15). En sommant sur tous les éléments, on a

$$\text{(I.1.17)} \qquad \begin{aligned} &\sum_k \int_{\partial\Omega_k} x_k\,\overline{(-\partial_{\nu_k} + i\omega)e_k} \\ &\qquad - \sum_k \int_{\partial\Omega_k} (+\partial_{\nu_k} + i\omega)u\,\overline{(+\partial_{\nu_k} + i\omega)e_k} = -2i\omega \sum_k \int_{\Omega_k} f\,\overline{e}_k \ . \end{aligned}$$

Comme x vérifie (I.1.8) et en combinant avec (I.1.17), on a

$$\text{(I.1.18)} \qquad \begin{cases} \forall\ (+\partial_{\nu_k} + i\omega)e_k \in V,\ e_k \in H_k \\ \displaystyle\sum_{k,j} \int_{\Sigma_{kj}} (+\partial_{\nu_k} + i\omega)u\,\overline{(+\partial_{\nu_k} + i\omega)e_k} + \sum_k \int_{\Gamma_k} (+\partial_{\nu_k} + i\omega)u\,\overline{(+\partial_{\nu_k} + i\omega)e_k} = \\ \qquad\qquad \displaystyle\sum_{k,j} \int_{\Sigma_{kj}} x_j\,\overline{(\partial_{\nu_k} + i\omega)e_k} + \sum_k \int_{\Gamma_k} (Qx_k + g)\ \overline{(\partial_{\nu_k} + i\omega)e_k} \ . \end{cases}$$

Cela donne (I.1.16). Une fonction, dont les restrictions à Ω_k sont $H^1(\Omega_k)$ et vérifient (I.1.11), et qui vérifie les relations de continuité (I.1.16) est la solution du problème de Helmholtz - une telle fonction est admissible et c'est la seule d'après le théorème d'existence et d'unicité 1. □

Le théorème 4 résume la formulation (I.1.8) du théorème 3 après introduction des opérateurs E, F, Π et A.

Définition 1 *Soient les applications de relèvement E et E_f*

$$\text{(I.1.19)} \qquad E_f = \begin{cases} V \to H = \displaystyle\prod_{k=1}^{K} H_k \\ z \mapsto e = (e_k),\ e_k = e_{|\Omega_k} \end{cases}$$

où e_k est l'unique solution de

$$\begin{cases} (-\partial_{\nu_k} + i\omega)e_k = z_{|\partial\Omega_k} & \text{sur } \partial\Omega_k \\ (-\Delta - \omega^2)e_k = f_{|\Omega_k} & \text{dans } \Omega_k \ , \end{cases}$$

et

$$\text{(I.1.20)} \qquad E = E_0 \ .$$

Remarque 2 L'application E est un opérateur linéaire alors que E_f pour $f \neq 0$ est un opérateur affine.

Définition 2 *Soit l'opérateur $F \in \mathcal{L}(V)$ défini par :*

$$\text{(I.1.21)} \qquad F = \begin{cases} V \to V \\ z \mapsto Fz = ((\partial_{\nu_k} + i\omega)(E(z)_{|\Omega_k})_{|\partial\Omega_k})_k \ . \end{cases}$$

liant la trace sortante $(-\partial_{\nu_k} + i\omega)e_k$ à la trace entrante $(\partial_{\nu_k} + i\omega)e_k$.

Remarque 3 Ces définitions ont un sens d'après le théorème 1. Ceci implique que $(E(z)_{|\Omega_k})_{|\partial\Omega_k}$ existe et est unique dans H. De $Fz = -z + 2i\omega(E(z)_{|\Omega_k})_{|\partial\Omega_k}$, on obtient $Fz \in \prod_k(L^2(\partial\Omega_k)) = V$. Remarquons que

$$\text{(I.1.22)} \qquad \prod_{k=1}^{K} H_k \text{ est isomorphe à } Y = \{(-\partial_{\nu_k} + i\omega)(v_k)_{|\partial\Omega_k}, v_k \in H_k, k = 1 \dots K\} \ .$$

Définition 3 *Soit Q une fonction définie sur le bord Γ à valeur complexe vérifiant $|Q| \leq 1$. Nous définissons l'opérateur linéaire Π par :*

$$\text{(I.1.23)} \qquad \Pi \in \mathcal{L}(V) \begin{cases} \Pi z_{|\Sigma_{kj}} = z_{|\Sigma_{jk}} \\ \Pi z_{|\Gamma_k} = Q z_{|\Gamma_k} \end{cases} .$$

Lemme 1 *L'opérateur linéaire Π ci-dessus vérifie évidemment $||\Pi||_V \leq 1$ pour $|Q| \leq 1$ sur Γ.*

Définition 4 *Soit $F^* \in \mathcal{L}(V)$ (V est identifié à son espace dual) l'adjoint de F. Soit $A \in \mathcal{L}(V)$ défini par*

$$\text{(I.1.24)} \qquad A = F^*\Pi \ .$$

Dans l'équation (I.1.8) nous reconnaissons dans le premier terme intégral le produit scalaire dans V, dans le second terme nous reconnaissons à gauche l'opérateur Π appliqué à x, à droite l'opérateur F appliqué à la fonction de base y. Nous sommes donc en mesure de formuler le théorème d'existence et d'unicité de la formulation variationnelle ultra-faible sous l'hypothèse de régularité $x \in V$:

Théorème 4

a) *Le problème (I.1.8) est équivalent à*

$$\text{(I.1.25)} \qquad \begin{cases} \textit{Trouver } x \in V \textit{ tel que } \forall y \in V \\ (x, y)_V - (\Pi x, Fy)_V = (b, y)_V \end{cases}$$

où le second membre $b \in V$ est défini, via le théorème de représentation de Riesz, par :

$$\text{(I.1.26)} \qquad \forall y \in V \quad (b, y)_V = -2i\omega \sum_k \int_{\Omega_k} f\, \overline{E(y)}_{\Omega_k} + \sum_k \int_{\Gamma_k} g\, \overline{F(y)}_{\Gamma_k} \ .$$

b) *Si u est solution du problème de Helmholtz (I.0.1) vérifiant l'hypothèse de régularité (I.0.9), alors $x = (-\partial_{\nu_k} + i\omega)u$ est solution dans V de (I.1.25).*

c) *Réciproquement, si x est solution de (I.1.25) alors $u = E_f(x)$ est l'unique solution de (I.0.1). Le problème (I.1.25) est équivalent à :*

$$\text{(I.1.27)} \qquad \begin{cases} \textit{Pour } b \in V, \textit{ trouver } x \in V \\ \boxed{(I - A)x = b} \end{cases}$$

Preuve. Soit $x = (-\partial_{\nu_k} + i\omega)u$, et $y \in V$ donnés. De y, on définit e par $E(y) = e$ (d'après le théorème d'existence et d'unicité 1 pour $f = 0$). Alors $y = (-\partial_{\nu_k} + i\omega)e_k$. Par l'égalité (I.1.8) et les définitions de F (I.1.21) et Π (I.1.23) on obtient (I.1.26). Puisque (I.1.26) est valable pour tout $y \in V$ (E est défini sur V), nous pouvons affirmer (I.1.27). La réciproque est assurée par le théorème 1. □

I.1.2.2 Propriétés de la formulation.

L'opérateur A vérifie les propriétés essentielles suivantes :

Proposition 6 *la norme de A vérifie $||A|| \leq 1$,*

Proposition 7 *l'opérateur $(I - A)$ est injectif.*

Les preuves sont aisées à l'aide du lemme suivant :

Lemme 2 *L'opérateur F est une isométrie dans V.*

Preuve. Cette propriété d'isométrie a déjà été étudiée [33] dans un tout autre contexte. Soit y vérifiant $e = E(y)$ (comme dans (I.1.19)). Alors :

$$(Fy, Fy) = \int_{\partial\Omega_k} |(\frac{\partial}{\partial\nu_k} + i\omega)e_k|^2 = \int_{\partial\Omega_k} |\frac{\partial}{\partial\nu_k}e_k|^2 + \omega^2|e_k|^2 - 2\omega\Im\left(\int_{\partial\Omega_k} \frac{\partial}{\partial\nu_k}e_k.\overline{e_k}\right) .$$

En intégrant par parties dans $\int_{\Omega_k}(-\Delta - \omega^2)e_k\overline{e}_k = 0$, on a :

$$\int_{\Omega_k} |\nabla e_k|^2 - \omega^2|e_k|^2 = \int_{\partial\Omega_k} \frac{\partial}{\partial\nu_k}e_k.\overline{e_k} \in \mathbb{R},$$

d'où,

$$\int_{\partial\Omega_k} |(\frac{\partial}{\partial\nu_k} + i\omega)e_k|^2 = \int_{\partial\Omega_k} |\frac{\partial}{\partial\nu_k}e_k|^2 + \omega^2|e_k|^2 = \int_{\partial\Omega_k} |(-\frac{\partial}{\partial\nu_k} + i\omega)e_k|^2 .$$

Ceci implique que

$$\|Fy\|^2 = \sum_k \int_{\partial\Omega_k} |(\frac{\partial}{\partial\nu_k} + i\omega)e_k|^2 = \sum_k \int_{\partial\Omega_k} |(-\frac{\partial}{\partial\nu_k} + i\omega)e_k|^2 = \|y\|^2 .$$

Comme ceci est vrai pour tout y de V, on a $F^*F = I$. L'adjoint F^* est la fonction qui à $(\partial_{\nu_k} + i\omega)e_k$ associe $(-\partial_{\nu_k} + i\omega)e_k$. C'est donc une isométrie injective et $FF^* = I$. □

Preuve. (de la proposition 6). L'opérateur Π défini par (I.1.23) vérifie $\|\Pi\| \leq 1$ (lemme 1). Ceci, combiné avec le lemme 11, conduit immédiatement à $||A|| \leq 1$. □

Preuve. (de la proposition 7). Cette preuve peut être vue comme une simple conséquence du théorème (4) qui spécifie l'équivalence entre le problème ultra-faible et le problème de Helmholtz (sous l'hypothèse de régularité de la dérivée normale). La preuve peut aussi être faite explicitement comme suit. Nous supposons l'existence de x tel que $x = Ax$ et nous vérifions $x = 0$. Soit $w = E(x)$ (E défini par (I.1.19)), en d'autres termes $x_{|\partial\Omega_k} = (-\partial_{\nu_k} + i\omega)w_{|\Omega_k}$ avec

$$(-\Delta - \omega^2)w_{|\Omega_k} = 0 . \tag{I.1.28}$$

L'équation $x = Ax$ multipliée à gauche par l'opérateur F devient $FF^*\Pi x = Fx$. Ceci, en utilisant $FF^* = I$ puisque F est une isométrie bijective (lemme 2), se transcrit sur w :

$$(+\partial_{\nu_k} + i\omega)w_{|\Sigma_{kj}} = (-\partial_{\nu_j} + i\omega)w_{|\Sigma_{jk}} \tag{I.1.29}$$

$$(+\partial_{\nu_k} + i\omega)w_{|\Gamma_k} = Q(-\partial_{\nu_k} + i\omega)w_{|\Gamma_k} . \tag{I.1.30}$$

Les équations (I.1.28) et (I.1.29) signifient que w est une solution du problème de Helmholtz. L'équation (I.1.28) signifie que le second membre f est nul sur Ω. L'équation (I.1.30) signifie que g est nul sur Γ. Sous les hypothèses d'existence et d'unicité du théorème 1, nous savons que $w = 0$. Ainsi, nous avons bien $x = 0$ ce qui montre par la contraposée l'injectivité de $(I - A)$. □

Chapitre I.2

Discrétisation du problème.

I.2.1 Approximation de Galerkin.

I.2.1.1 Existence et unicité de la solution approchée.

Dans cette partie, on s'intéresse à la discrétisation du problème variationnel (I.1.25). On utilise une procédure de type Galerkin. On introduit un espace de dimension finie $V_h \subset V$. On obtient la formulation :

$$\text{trouver } x_h \in V_h, \text{ tel que} \tag{I.2.1}$$
$$\begin{cases} \forall y_h \in V_h \\ (x_h, y_h)_V - (\Pi x_h, F y_h)_V = (b, y_h)_V \end{cases}$$

Théorème 5 *Le problème (I.2.1) a une solution unique.*

Preuve. Soit $P_h : V \to V_h$ l'opérateur de projection orthogonale de V dans V_h. L'équation (I.2.1) signifie

$$(I - P_h A)x_h = P_h b \ . \tag{I.2.2}$$

Prouvons seulement l'unicité de (I.2.2) puisque V_h étant un espace de dimension finie, l'unicité est équivalente à l'existence. L'unicité signifie que si $P_h b = 0$ dans la formulation (I.2.2) alors $x_h = 0$. En utilisant le fait que P_h est un projecteur orthogonal, $(I - P_h A)x_h = 0$ est équivalent à :

$$||x_h||^2 = ||P_h A x_h||^2 = ||Ax_h||^2 - ||(I - P_h)Ax_h||^2 \ . \tag{I.2.3}$$

Comme $||A|| \leq 1$ (proposition 6) nous avons $||x_h||^2 \leq ||x_h||^2 - ||(I - P_h)Ax_h||^2$ donc $(I - P_h)Ax_h = 0$. Ceci, avec $(I - P_h A)x_h = 0$ entraîne $(I - A)x_h = 0$. Comme $(I - A)$ est injectif (proposition 7), nous obtenons finalement $x_h = 0$. □

I.2.1.2 Remarque sur les notations V et V_h.

La notation V_h est issue de la terminologie des éléments finis (cf [17]) et des hypothèses d'uniforme régularité H1, H2 et H3 faites sur le maillage.

Hypothèse 1 *Le bord $\partial\Omega_k$ est Lipschitz-continu,*

Hypothèse 2 *On définit un "maillage régulier" comme dans [17]. On note h_k le diamètre de Ω_k et ρ_k le maximum des diamètres des sphères inscrites dans Ω_k. On suppose que le maillage est non dégénéré et que $\rho_k \neq 0$. On peut donc définir*

$$\sigma_k = \frac{h_k}{\rho_k} \ , \tag{I.2.4}$$

et on suppose qu'il existe σ tel que

$$\forall k, \ h_k \leq \sigma \rho_k \ . \tag{I.2.5}$$

On définit alors le paramètre de raffinement du maillage h par

(I.2.6)
$$h = \max_k h_k \ .$$

Hypothèse 3 *Σ_{kj} est une arête commune de Ω_k et de Ω_j.*

Dans la méthode des éléments finis, la donnée du degré des polynômes (les polynômes sont les fonctions de base des éléments finis) et du maillage définit entièrement l'espace d'approximation V_h. Dans notre méthode, la donnée du maillage fait partie de l'étude du problème continu, préliminaire à l'étude du problème discret. La donnée du maillage construit l'espace continu V, mais nous avons encore entière liberté sur le choix des fonctions de base. Nous noterons V_h le sous espace de V construit à l'aide d'un nombre fini de fonctions de base. La notation en indice h exprime donc à la fois la dimension finie et, pour un maillage uniformément régulier, la taille caractéristique du maillage.

I.2.1.3 Construction de Galerkin de l'espace V_h.

Dans chaque élément Ω_k, nous considérons un nombre fini de fonctions e_{kl} solutions indépendantes de l'équation de Helmholtz homogène dans l'élément Ω_k. Un ensemble particulier de fonctions e_{kl} est celui des fonctions de support contenu dans Ω_k :

(I.2.7)
$$\begin{cases} (e_{kl})_{|\Omega_j} = 0 \text{ si } k \neq j \\ (-\Delta - \omega^2)(e_{kl})_{|\Omega_k} = 0 \ . \end{cases}$$

L'utilisation de fonctions à support compact, comme cela est habituel dans la méthode des Eléments Finis, mène à un système dont la matrice est creuse. Remarquons que les fonctions e_{kl} ne sont pas continues sur Ω mais qu'elles sont C^∞ sur l'intérieur strict de Ω_k.

L'indice k (variant de 1 à K le nombre total de mailles) désigne le numéro de la maille dans laquelle e_{kl} n'est pas identiquement nulle. Le deuxième indice l correspond au numéro d'une fonction de base pour la maille Ω_k : c'est une numérotation locale à la maille Ω_k. Pour simplifier, nous prendrons un nombre constant, noté p, de fonctions de base pour toute maille Ω_k. Les fonctions z_{kl} sont définies de façon univoque par $z_{kl} = (-\partial_{\nu_k} + i\omega)e_{kl}$, e_{kl} étant une solution du système (I.2.7). Nous avons donc :

(I.2.8)
$$\begin{cases} (z_{kl})_{|\partial\Omega_j} = 0 \text{ si } k \neq j \\ (z_{kl})_{|\partial\Omega_k} = (-\partial_{\nu_k} + i\omega)e_{kl} \ . \end{cases}$$

On vérifie que $\{z_{kl}\}_{k,l}$ est une base de V_h sous les hypothèses du lemme suivant :

Lemme 3 *La famille de fonctions $\{z_{kl}\}_{1\leq l\leq p}$ définies par (I.2.8) est libre dans V si et seulement si la famille de fonctions $\{e_{kl}\}_{1\leq l\leq p}$ est libre dans H_k.*

Preuve. C'est une conséquence du théorème (4). Plus précisément, si la famille $\{e_{kl}\}_{1\leq l\leq p}$ n'est pas libre, il est évident que les fonctions $\{z_{kl}\}_{1\leq l\leq p}$ ne seront pas libres. Réciproquement, montrons que si les fonctions e_{kl} sont libres, alors les fonctions z_{kl} le sont aussi, ou, en d'autres termes, supposons que les fonctions $\{z_{kl}\}_{1\leq l\leq p}$ sont liées et montrons qu'alors, les fonctions e_{kl} le sont aussi. Montrons qu'alors les fonctions $\{e_{kl}\}_{1\leq l\leq p}$ sont liées. L'hypothèse signifie qu'il existe $y_h \in V$ et $w_h \in H$ (définition (I.1.9)) et $Y = (y_{kl})_{k,l} \in \mathbb{C}^{pK}$ tels que

(I.2.9)
$$\begin{cases} y_h = \displaystyle\sum_{l=1}^{p} y_{kl}\, z_{kl} \\ w_h = E(y_h) \ , \end{cases}$$

avec

(I.2.10)
$$||y_h||_V = 0 \ .$$

De façon équivalente à (I.2.10), on suppose

(I.2.11)
$$\sum_k \int_{\partial\Omega_k} |(-\partial_{\nu_k} + i\omega)w_h|^2 = 0 \ .$$

Alors, $w_h = E(y_h)$ vérifie, pour tout élément Ω_k

$$\text{(I.2.12)} \qquad \begin{cases} (-\Delta - \omega^2)w_h = 0 \text{ dans } \Omega_k \\ (-\partial_{\nu_k} + i\omega)w_h = 0 \text{ sur } \partial\Omega_k \ , \end{cases}$$

l'équation (I.2.12) implique que ce qui montre que $w_h = 0$ sur $L^2(\Omega_k)$ pour tout élément Ω_k (d'après le théorème 1 d'existence et d'unicité). Par définition de w_h (I.2.9), on a donc

$$\text{(I.2.13)} \qquad \forall k, \forall \vec{x} \in \Omega_k \ \sum_{l=1}^{p} y_{kl} e_{kl} = 0 \ .$$

qui montre que les fonctions e_{kl} sont liées. □

On construit finalement l'espace discret V_h comme l'espace vectoriel engendré par les fonctions z_{kl}, autrement dit :

$$\text{(I.2.14)} \qquad V_h = \text{ Vect } (z_{kl})_{1 \leq k \leq K}^{1 \leq l \leq p} \ .$$

On cherche donc la solution approchée x_h sous la forme d'une combinaison linéaire des fonctions de base : la solution approchée x_h est entièrement définie par la donnée des pK coefficients complexes x_{kl} définissant $X = (x_{kl})_{1 \leq k \leq K}^{1 \leq l \leq p}$ tels que

$$\text{(I.2.15)} \qquad (x_h)_{|\partial\Omega_k} = \sum_{l=1}^{p} x_{kl} \, z_{kl} \ ,$$

ou exprimée en fonction des e_{kl} :

$$\text{(I.2.16)} \qquad (x_h)_{|\partial\Omega_k} = \sum_{l=1}^{p} x_{kl} (-\partial_{\nu_k} + i\omega) e_{kl} \ .$$

I.2.1.4 Construction des opérateurs discrets.

Nous construisons les termes correspondant à la formulation (I.1.27). La formulation discrétisée (I.2.1) de (I.1.25) conduit au système (I.2.17), forme discrète de (I.1.27).

$$\text{(I.2.17)} \qquad \begin{cases} \text{Trouver } X \in \mathbb{C}^{pK} \\ (D - C)X = b \end{cases}$$

La matrice notée D est la matrice du produit scalaire dans V_h. La matrice notée C est la matrice de la forme bilinéaire $(\Pi x_h, F y_h)$. Le second membre, par abus de langage, est toujours noté b.

i) Les coefficients de la matrice D définis par $D_{k,j}^{l,m} = (z_{jm}, z_{kl})_V$ sont :

$$\text{(I.2.18)} \qquad D_{k,j}^{l,m} = \int_{\partial\Omega_k} (-\partial_{\nu_k} + i\omega) e_{jm} \, \overline{(-\partial_{\nu_k} + i\omega) e_{kl}} \ .$$

Remarque 4 Pour $j \neq k$ les supports de e_{jm} et e_{kl} sont disjoints, donc $D_{k,j}^{l,m} = 0$.

ii) Les coefficients de la matrice C sont, soit une intégrale sur les interfaces Σ_{kj} et définis par $C_{k,j}^{l,m} = (\Pi z_{jm}, F z_{kl})_V$, soit une intégrale sur une face de bord Γ_k et définis par $C_k^{l,m} = (\Pi z_{km}, F z_{kl})_V$.

$$\text{(I.2.19)} \qquad C_{k,j}^{l,m} = \int_{\Sigma_{kj}} (+\partial_{\nu_k} + i\omega) e_{jm} \, \overline{(+\partial_{\nu_k} + i\omega) e_{kl}}$$

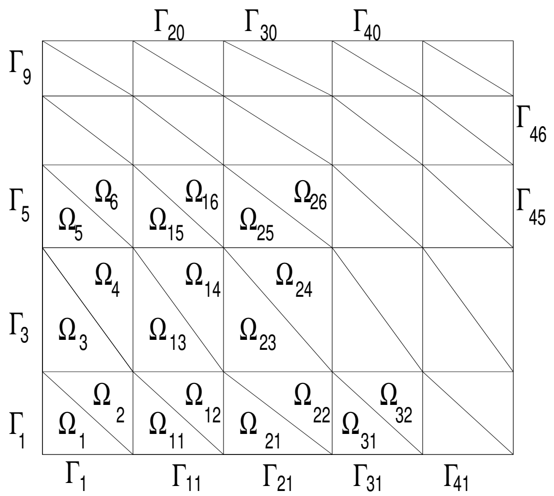
$$\text{(I.2.20)} \qquad C_k^{l,m} = \int_{\Gamma_k} Q_k \, (-\partial_{\nu_k} + i\omega) e_{jm} \, \overline{(+\partial_{\nu_k} + i\omega) e_{kl}}$$

iii) Le second membre b est défini par $b_{k,l} = (b, z_{kl})_V$ (le second membre de la formulation discrète et le second membre du problème variationnel sont abusivement notés de la même façon) soit

$$b_{k,l} = -2i\omega \int_{\Omega_k} f\,\overline{(e_{kl})} + \int_{\Gamma_k} g\,\overline{(+\partial_{\nu_k} + i\omega)e_{kl}}\ . \tag{I.2.21}$$

Nous considérons la figure I.2.1 de maillage d'un rectangle en triangles. Nous présentons la forme (essentiellement creuse) des matrices D et C.

FIG. I.2.1 – Un exemple de maillage d'un rectangle en triangles.



i) La matrice D est diagonale par blocs : $D_{kj} = 0$ pour $k \neq j$:

$$D = \left| \begin{array}{cccccc} D_1 & 0 & \dots & & & \\ 0 & \ddots & 0 & & & \\ \vdots & 0 & D_k & 0 & & \\ & & 0 & \ddots & 0 & \\ & & & 0 & D_K & \end{array} \right| \text{où } (D_k)_{(l,m)} = D_{kk}^{lm}\ .$$

ii) La matrice C est non nulle sur des blocs diagonaux correspondant aux éléments Ω_k ayant une face sur le bord extérieur Γ_k (le terme $C_{k,k}$ est nul si $\Gamma_k = \emptyset$) :

$$\text{Diag}\ (C) = \begin{bmatrix} C_{1,1} & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & C_{3,3} & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & C_{5,5} \\ \vdots & \vdots & \vdots & \vdots & 0 \end{bmatrix}.$$

iii) Le terme non diagonal de couplage de C est nul au bord. Nous donnons la forme générale de ce terme, et les blocs précis pour les couplages des 5 premiers éléments du maillage :

$$\begin{bmatrix} 0 & C_{1,2} & 0 & 0 & 0 & 0 & 0\,0\,0\,0 & 0 & 0 & 0 & \dots \\ C_{2,1} & 0 & C_{2,3} & 0 & 0 & 0 & 0\,0\,0\,0 & C_{2,11} & 0 & 0 & \dots \\ 0 & C_{3,2} & 0 & C_{3,4} & 0 & 0 & 0\,0\,0\,0 & 0 & 0 & 0 & \dots \\ 0 & 0 & C_{4,3} & 0 & C_{4,5} & 0 & 0\,0\,0\,0 & 0 & 0 & C_{2,13} & \dots \\ 0 & 0 & 0 & C_{5,4} & 0 & C_{5,4} & 0\,0\,0\,0 & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \\ \dots & C_{k,j_1} & \dots & \dots & \dots & C_{k,j_2} & \dots & \dots & & C_{k,j_3} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \end{bmatrix}$$

où (j_1, j_2, j_3) sont les indices des trois voisins d'un élément Ω_k de la triangulation.

I.2.1.5 Lien avec le problème initial : construction de u_h à partir de x_h.

Nous décrivons comment définir et calculer une approximation de u (la solution du problème de Helmholtz), notée u_h à partir de x_h solution de (I.2.1). Il y a deux techniques indépendantes d'approximation de u.

i) Reconstruction de u_h sur $\partial\Omega_k$

En utilisant $u_{|\Sigma_{kj}} = u_{|\Sigma_{jk}}$ et $+\partial_{\nu_j} u_{|\Sigma_{jk}} = -\partial_{\nu_k} u_{|\Sigma_{kj}}$, nous avons sur Σ_{kj} :

$$\begin{aligned}(I+\Pi)x &= (-\partial_{\nu_k} + i\omega)u_{|\Sigma_{kj}} + (-\partial_{\nu_j} + i\omega)u_{|\Sigma_{jk}} \\ &= (-\partial_{\nu_k} + i\omega)u_{|\Sigma_{kj}} + (+\partial_{\nu_k} + i\omega)u_{|\Sigma_{kj}} \\ &= 2i\omega u\ .\end{aligned}$$

En utilisant $(\partial_{\nu_k} + i\omega)u_{|\Gamma_k} = Q(-\partial_{\nu_k} + i\omega)u_{|\Gamma_k} + g$, nous avons sur Γ_k :

$$\begin{aligned}(I+\Pi)x + g &= (-\partial_{\nu_k} + i\omega)u_{|\Gamma_k} + Q(-\partial_{\nu_k} + i\omega)u_{|\Gamma_k} + g \\ &= (-\partial_{\nu_k} + i\omega)u_{|\Gamma_k} + (+\partial_{\nu_k} + i\omega)u_{|\Gamma_k} \\ &= 2i\omega u\ .\end{aligned}$$

Ainsi, la trace de u sur V est liée à x par

$$\text{(I.2.22)} \qquad \begin{cases} u = \dfrac{1}{2i\omega}[(I+\Pi)x] & \text{sur } \Sigma_{kj} \\ u = \dfrac{1}{2i\omega}[(I+\Pi)x + g] & \text{sur } \Gamma_k\ . \end{cases}$$

Il est donc naturel de définir u_h sur les arêtes du maillage par

$$\text{(I.2.23)} \qquad \begin{cases} u_h = \dfrac{1}{2i\omega}[(I+\Pi)x_h] & \text{sur } \partial\Omega_k \\ u_h = \dfrac{1}{2i\omega}[(I+\Pi)x_h + g] & \text{sur } \Gamma_k\ . \end{cases}$$

Pratiquement, l'équation (I.2.23) signifie que u_h est calculé

1. sur Γ_k par

$$\text{(I.2.24)} \qquad 2i\omega(u_h)_{|\Gamma_k} = g + (1 + Q_k)\sum_l x_{kl}(-\partial_{\nu_k} + i\omega)e_{kl}$$

2. et sur Σ_{kj} par

$$\text{(I.2.25)} \qquad 2i\omega(u_h)_{|\Sigma_{kj}} = \sum_{l(k)} x_{kl}(-\partial_{\nu_k} + i\omega)e_{kl} + \sum_{l(j)} x_{jl}(-\partial_{\nu_j} + i\omega)e_{jl}\ .$$

ii) Reconstruction de u_h dans Ω.

Nous savons qu'il est théoriquement possible d'inverser l'opérateur E_f (I.1.19) sur chacune de ses restrictions à Ω_k pour tous les éléments du maillage. En pratique, le fait d'inverser

$$\begin{cases} -(\Delta+\omega^2)u = f & \text{dans } \Omega_k \\ (-\partial_\nu + i\omega)u = x & \text{sur } \partial\Omega_k \ , \end{cases}$$

ou sous la forme discrète

$$\begin{cases} -(\Delta+\omega^2)u_h = f & \text{dans } \Omega_k \\ (-\partial_\nu + i\omega)u_h = x_h & \text{sur } \partial\Omega_k \ . \end{cases} \tag{I.2.26}$$

est *a priori* équivalent à résoudre le problème de Helmholtz originel. Néanmoins, il peut être intéressant de résoudre beaucoup de problèmes dans des domaines restreints plutôt qu'un seul dans un grand domaine. Ceci est l'idée initiale des techniques de décomposition de domaine [24]. Dans le cas particulier où $f = 0$ sur Ω_k, l'équation (I.2.26) est simple à résoudre : la linéarité de l'opérateur de relèvement E (I.1.20) implique que

$$(u_h)_{|\Omega_k} = \sum_{l=1}^{p} x_{kl}\, e_{kl} \ . \tag{I.2.27}$$

est solution du problème (I.2.26) dans tout Ω_k. Attirons l'attention du lecteur sur le fait que la formule (I.2.27) peut être utilisée sur les arêtes du maillage à la place de (I.2.24) et (I.2.25) si nécessaire.

Remarque 5 Dans le cas d'un problème où f n'est pas identiquement nulle sur l'élément Ω_k il est toujours possible de résoudre (I.2.26) en utilisant d'autres méthodes, comme la méthode des Eléments Finis (FEM), ou d'utiliser à nouveau la méthode ultra-faible sur les sous-domaines Ω_k. La formulation ultra-faible est donc adaptée à la fois aux méthodes de décomposition de domaine et aux méthodes multi-grilles.

Remarque 6 Sur le plan pratique, les logiciels de représentation graphique requièrent habituellement les valeurs aux nœuds du maillage. Ceci se fait par la formule (I.2.23). Une autre possibilité consiste à continuer à utiliser la formule (I.2.27) qui est aussi une approximation linéaire de u sur Ω_k connaissant u sur $\partial\Omega_k$. Dans les deux cas, nous approchons la solution à l'ordre 1.

I.2.2 Mise en œuvre d'un espace d'approximation particulier.

I.2.2.1 Choix de V_h.

Nous avons choisi de construire notre espace d'approximation à partir :

i) D'une partition de Ω en éléments triangulaires réguliers. La terminologie "éléments réguliers" signifie que le maillage vérifie les hypothèses H1, H2 et H3. L'intérêt est bien sûr que cette partition se réalise aisément à l'aide d'un mailleur.

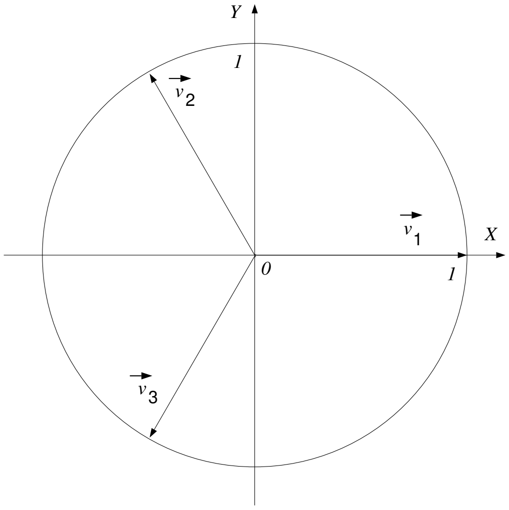
ii) De fonctions de base construites à l'aide d'ondes planes sur Ω_k de la forme

$$\begin{cases} e_{kl} = e^{(i\omega \vec{v}_{kl}.\vec{x})} \\ \vec{v}_{kl}\,\vec{v}_{kl} = 1, \quad \text{et } l \neq m \Rightarrow \vec{v}_{kl} \neq \vec{v}_{km} \ . \end{cases} \tag{I.2.28}$$

iii) En outre, lors des simulations numériques, les vecteurs d'onde de ces ondes planes ont été choisis réels équirépartis dans le plan :

$$\forall k \in \{1 \dots K\},\ \vec{v}_{kl,\ l=1\dots p} = \begin{pmatrix} \cos(2\pi \dfrac{l-1}{p}) \\ \sin(2\pi \dfrac{l-1}{p}) \end{pmatrix} \ , \tag{I.2.29}$$

et identiques sur tous les éléments[1]. Cela donne par exemple pour trois et sept fonctions de base par élément, la disposition de la figure I.2.2.



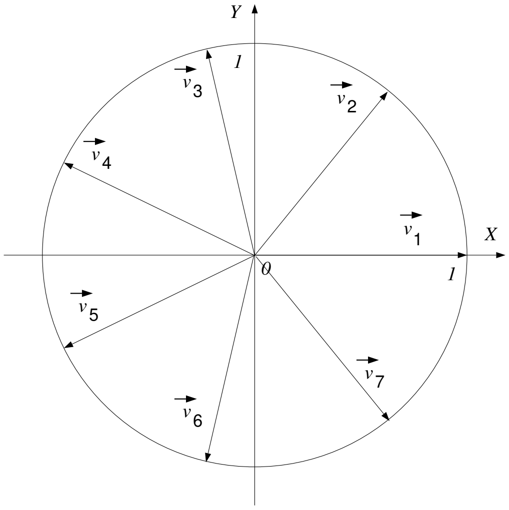


FIG. I.2.2 – Répartition géométrique des directions des fonctions de base.

Lemme 4 *Les fonctions* $\{z_{kl}\}_{1\leq l\leq p}$ *définies par* $z_{kl} = (-\partial_{\nu_k} + i\omega)e_{kl}$ *avec* e_{kl} *vérifiant (I.2.28) forment une base de* V_h.

Preuve. Le lemme 3 stipule l'équivalence entre l'indépendance linéaire des fonctions z_{kl} dans V_h et celle des fonctions e_{kl} dans H (voir (I.1.9)). La condition d'indépendance des ondes planes est une propriété bien connue des exponentielles sur $\mathbb{R}^2$. Cette propriété est facile à montrer en supposant l'existence de coefficients non nuls α_l liant les fonctions e_{kl} par

$$\sum_{l=1}^{p} \alpha_l e_{kl} = 0\ . \tag{I.2.30}$$

Nous proposons alors deux preuves.

- La dérivation d'une exponentielle par rapport à l'une des coordonnées donne cette même exponentielle multipliée par une constante qui ne dépend que du vecteur d'onde de l'onde plane. Appliquant l'opérateur différentiel (D_x+iD_y) (où le symbole D dénote l'opérateur de dérivation) p fois à (I.2.30), on obtient un système linéaire dont la matrice a un déterminant de Vandermonde qui est non nul si et seulement si tous les vecteurs $\vec{v}_{km}$ sont différents.
- On peut appliquer l'opérateur de Fourier-Plancherel à (I.2.30) en prolongeant la fonction sur Ω à $\mathbb{R}^2$ entier. On a alors
$$\sum_{l=1}^{p} \alpha_l \delta_{-\vec{v}_{k,l}} = 0\ ,$$
qui n'est possible, les points $-\vec{v}_{k,l}$ étant distincts, que si, pour tout l de 1 à p, on a : $\alpha_l = 0$.

□

Remarque 7 (Choix du maillage) Le choix d'un maillage triangulaire n'est absolument pas obligatoire. Nous aurions aussi pu prendre un maillage quadrangulaire ou mixte. C'est seulement pour la simplicité de la mise en œuvre informatique que ce choix a été fixé ; un autre choix n'implique rien de fondamentalement différent. De même, c'est pour la simplicité de la mise en œuvre informatique que nous avons choisi un nombre constant de fonctions de base.

[1]Les directions peuvent partir de la direction $(1,0)$ comme c'est le cas pour la fonction $\vec{v}_{k,l=0}$ dans la formule (I.2.29) ou de $\vec{v}_0$ dans les problèmes de "scattering" (cf section I.4.2) où $\vec{v}_0$ est la direction de l'onde plane incidente.

Remarque 8 (Choix des fonctions de base) Le choix de fonctions de base issues d'ondes planes est justifié par le fait que le calcul des termes des matrices est analytique.

Le choix de directions équiréparties pour les vecteurs d'onde des ondes planes est le choix le plus simple quand on ne connaît pas la forme de la solution. Par exemple, si l'on étudie la diffraction sur un dièdre infini d'une onde plane incidente, il est clair qu'introduire l'onde incidente et l'onde diffractée dans l'espace des fonctions de base donnera un excellent résultat. L'utilisation d'informations sur le comportement asymptotique pour le choix des fonctions de base est une voie judicieuse pour améliorer les études de diffraction. Pour effectuer une telle étude nous conseillons de se reporter à [9] qui étudie en détail avec une approche à la fois physique et mathématique les comportements asymptotiques de la propagation d'onde.

Un autre choix de directions des ondes planes est de choisir les normales aux triangles. Nous avons comparé sur des cas tests un tel choix avec celui où les fonctions de base sont équiréparties et nous avons constaté un meilleur résultat dans le second cas (surtout lorsque le maillage n'est pas très régulier). Nous avons enfin deux justifications supplémentaires de l'emploi a priori de directions équiréparties.

1. Le coût de détermination des directions est mineur et permet d'avoir des fonctions de base égales sur tous les éléments. Ceci permet d'économiser le stockage des fonctions de base et le stockage de $K_{int} \times p^2$ termes matriciels où K_{int} est le nombre d'éléments intérieurs (éléments n'ayant pas d'arête libre) et où p est le nombre de fonctions de base par élément. Ceci est issu de propriétés particulières liant les matrices D et C lorsque les fonctions de base sont les mêmes sur tous les éléments.
2. Nous montrons que le choix de fonctions de bases équiréparties maximise le déterminant des matrices D_k quand le paramètre de raffinement du maillage tend vers 0 et quand on prend $p = 3$ fonctions de base par élément. Ceci permet alors une meilleure majoration du conditionnement. On observe que l'optimisation du déterminant s'effectue en deux temps, la première étape permettant de séparer les facteurs géométriques de l'élément des facteurs dépendant des directions des fonctions de base. Ceci montre que le choix des normales aux arêtes du maillage dans ce cas n'est pas le meilleur.

I.2.2.2 Construction du système linéaire.

Le lecteur peut se reporter à l'annexe (III.A p. 172) pour les formules de calcul des termes du système linéaire construit dans l'espace V_h.

I.2.3 Solution du système matriciel.

Le système (I.2.17) est résolu en deux étapes :

1. Inversion de D par une méthode directe, ce qui mène au système

$$(I - D^{-1}C) = D^{-1}b\ .$$

2. Résolution finale grâce à l'algorithme itératif de Richardson, qui s'interprète en l'occurrence comme la version sous-relaxée discrète du développement en série de Neumann de $(I-A)$. Cet algorithme est connu dans le cas général des matrices non symétriques [51] mais semble particulièrement efficace pour le système linéaire présenté ici où les valeurs propres de la matrice $M = D^{-1}C$ sont dans le disque complexe unité privé de la valeur 1. L'utilisation de l'algorithme itératif de Richardson (I.2.37) est issue des idées de décomposition de domaine (cf [24]). L'algorithme est analogue aux méthodes de Jacobi sous-relaxées.

Notons que la technique d'inversion proposée est naturelle dans les techniques de décomposition de domaine, mais ce n'est évidemment pas la seule possible. On aurait pu étudier l'emploi d'autres algorithmes itératifs comme GMRes [56] ou Bi-CGStab [58] ou même QMR [31] qui sont des méthodes efficaces et robustes de résolution de systèmes linéaires. A ce propos, citons la thèse de L. Crouzet [22] qui étudie en détail ces méthodes pour la résolution des problèmes de Helmholtz et Maxwell à l'aide d'une formulation éléments finis. Nous ne discuterons pas de ces méthodes qui ne sont pas l'objet de notre exposé.

I.2.3.1 Inversibilité de D, propriété de $M = D^{-1}C$.

Lemme 5 *La matrice D définie par (I.2.18), est hermitienne positive, définie positive si et seulement si les fonctions de base e_{kl} sont linéairement indépendantes dans Ω. Dans le cas où les fonctions de base e_{kl} sont des ondes planes à support dans Ω_k et de direction $\vec{v}_{kl}$, la condition d'indépendance linéaire est assurée si et seulement si les vecteurs d'onde $\vec{v}_{kl}$ sont tous distincts, soit $\forall k, (\forall (i,j), \vec{v}_{ki} \neq \vec{v}_{kj})$.*

Preuve. La matrice D représente le produit scalaire dans V_h, elle est donc hermitienne positive. Le lemme 4 montre que D est définie dès que les directions de propagation des ondes planes de tout élément Ω_k diffèrent. □

On pose alors :

(I.2.31)
$$\begin{cases} b' = D^{-1}b \\ M = D^{-1}C \ . \end{cases}$$

Lemme 6 *Les valeurs propres de M définie par (I.2.31) sont situées dans le disque complexe de rayon 1, la valeur 1 exclue.*

Preuve. Nous notons $\lambda \in \mathbb{C}$ une valeur propre de M et $Y \in \mathbb{C}^N$ son vecteur propre associé en notant y_{kl} les coefficients du vecteur $Y = (y_{kl})_{k,l}$.

i) Montrons tout d'abord que $|\lambda| \leq 1$. L'égalité $MY = \lambda Y$ et la définition (I.2.31) donnant M impliquent :

(I.2.32)
$$(CY, Y) = \lambda(DY, Y) \ .$$

Soit $y_h \in V_h$ la fonction dont les coefficients sont y_{kl} : $(y_h)_{|\partial\Omega_k} = \sum_l y_{kl} z_{kl}$. Par définition de C et D, nous avons :

(I.2.33)
$$\begin{cases} (CY, Y) = (Ay_h, y_h)_V \\ (DY, Y) = (y_h, y_h)_V \ . \end{cases}$$

Donc, à l'aide de (I.2.32) et (I.2.33) nous obtenons

(I.2.34)
$$\lambda = \frac{(Ay_h, y_h)_V}{(y_h, y_h)_V} \ .$$

Comme $y_h \in V_h \subset V$ et $||A|| \leq 1$, l'équation (I.2.34) entraîne $|\lambda| \leq 1$.

ii) Montrons maintenant que la valeur $\lambda = 1$ est exclue. Supposons que $\lambda = 1$. Alors,

$$(Ay_h, y_h)_V = (y_h, y_h)_V \ ,$$

or

$$\begin{aligned} ||(I - A)y_h||^2 &= ||y_h||^2 + ||Ay_h||^2 - 2\Re e \ (Ay_h, y_h)_V \\ &\leq ||y_h||^2[1 + 1 - 2] = 0 \ , \end{aligned}$$

donc $(I - A)y_h = 0$. L'injectivité de $I - A$ implique $y_h = 0$, ainsi Y est nul. Ceci est contradictoire avec le fait que Y est vecteur propre.

□

Corollaire 1 *Soit $B = (1 - \beta)I + \beta M$ (M définie par (I.2.31)) avec $\beta \in]0; 1[$. Les valeurs propres de la matrice B sont situées à l'intérieur du disque complexe unitaire ouvert (la valeur 1 est donc exclue). Notons $\sigma(B)$ le rayon spectral de B. Comme la matrice B est de dimension finie, il existe $\zeta_0 \in]0; 1[$ tel que :*

(I.2.35)
$$\sigma(B) \leq \zeta_0 \ .$$

Soient η et ε des réels strictement positifs avec $\eta \leq 1 - \varepsilon < 1$. Alors, pour toute suite réelle $\beta_n \in]\eta; 1 - \varepsilon[$ et $B_n = (1 - \beta_n)I + \beta_n M$, il existe $\zeta \in]0; 1[$ tel que :

(I.2.36)
$$\sup_{n \in \mathbb{N}} \sigma(B_n) \leq \zeta \ .$$

On note (figure I.2.3) $\lambda(M)$ une valeur propre de M et $\lambda(B)$ une valeur propre de B. Le cercle de rayon 1 centré sur l'origine délimite le spectre ponctuel $\sigma_p(M)$ de M (ensemble des valeurs propres en dimension finie). L'ensemble hachuré, noté $\sigma_p(B)$ contient les valeurs propres de B. L'introduction du correcteur $\beta \in]0;1[$ tend globalement à rapprocher de l'origine les valeurs propres de M et permet donc d'améliorer la convergence sur le plan numérique.

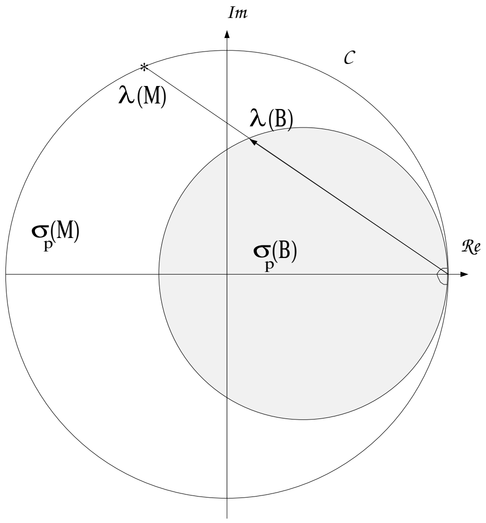


FIG. I.2.3 – Valeurs propres de M et B pour $\beta = 0,6$.

I.2.3.2 Construction d'un algorithme itératif.

Le système (I.2.17) se met sous la forme (I.2.37) où M est défini par (I.2.31) :

$$(I - M)X = b' \tag{I.2.37}$$

Nous terminons la résolution du système (I.2.37) par un développement en série de Neumann de $(I - M)$. Si un tel développement ne converge pas forcément pour tout opérateur continu de norme inférieure à 1, il converge pour l'opérateur discret associé : ceci sera montré par le lemme 7. Notons que cette preuve est aussi effectuée par [51] dans le cadre plus général des matrices non hermitiennes sous des conditions générales sur le spectre de la matrice. Nous proposons deux versions algorithmiques de schéma itératif, pour une sous-relaxation constante ou non.

i) Soit $\beta \in]0,5;1[$ fixe, l'algorithme est :

$$\begin{cases} X_1 = \beta b' \\ X_{n+1} = \beta b' + [(1-\beta)I + \beta M]X_n \ . \end{cases} \tag{I.2.38}$$

ii) Soit $\beta_n \in]0,5;1-\varepsilon[,\ \varepsilon > 0$ suite de nombres aléatoires, l'algorithme est :

$$\beta_n \in]0,5;1-\varepsilon[,\ \varepsilon > 0 \begin{cases} X_1 = \beta_1 b' \\ X_{n+1} = \beta_n b' + [(1-\beta_n)I + \beta_n M]X_n \ . \end{cases} \tag{I.2.39}$$

Remarque 9 C'est le second algorithme (I.2.31) que nous avons le plus utilisé. Nous avons pris les coefficients β_n répartis selon une loi aléatoire uniforme entre 0,5 et 0,99. On constate en effet qu'avec environ 10 ou 20% d'itérations en moins, on obtient une précision égale à celle obtenue par le premier algorithme (I.2.38).

Remarque 10 Le lecteur notera que l'algorithme (I.2.39) utilise des coefficients de relaxation non cycliques. Il serait intéressant d'étudier d'autres choix des termes de relaxation β_n de façon à diminuer encore le nombre d'itérations [51]. Nous n'avons pas mené cette étude qui serait à faire pour optimiser le coût calcul de la méthode.

Remarque 11 La génération de nombres aléatoires β_n est faite d'après une méthode proposée par Knuth dont l'avantage est d'être de période quasi infinie, et d'avoir une faible corrélation séquentielle (cf [53], Ch 7, p.196 fonction *RAN1*). L'avantage de la programmation de cette fonction est sa portabilité d'un système informatique à un autre.

Lemme 7 *Les algorithmes (I.2.38) et (I.2.39) convergent vers X, la solution de $(D - C)X = b$.*

Preuve. La convergence de (I.2.38) est donnée dans [57]. Cette preuve peut être vue comme un cas particulier de la convergence de (I.2.39) qui est démontrée ici. Soit N $(= pK)$ la taille de D et C. Définissons Δ et J comme la décomposition de Jordan de la matrice M : la matrice Δ est diagonale et J est son bloc de Jordan associé (c'est une matrice nilpotente). Soit P la matrice de passage de la transformation linéaire de Jordan, alors $M = P(\Delta + J)P^{-1}$. Définissons :

$$\begin{aligned} M_q &= (1 - \beta_q)I + \beta_q M \\ \Delta_q &= (1 - \beta_q)I + \beta_q \Delta \\ T_q &= \Delta_q + \beta_q J \\ D_q^{\alpha_q} &= \begin{cases} \beta_q I \text{ si } \alpha_q = 1 \\ \Delta_q \text{ si } \alpha_q = 0 \end{cases} \end{aligned} \tag{I.2.40}$$

Pour toute matrice B, soit $||B||$ la norme définie par :

$$||B|| = \sup_{Y \in \mathbb{C}^N} \frac{|BY|}{|Y|}, \text{ avec } |Y| = ||Y||_{L^2(\mathbb{C}^N)}. \tag{I.2.41}$$

i) Calcul explicite de X_n.

Remarquons que, pour tout entier n, X (la solution de (I.2.37)) vérifie :

$$(I - M_n)X = \beta_n b' \ . \tag{I.2.42}$$

Soit $Y_n = X_n - X$, alors :

$$\begin{aligned} Y_{n+1} &= [(1 - \beta_n)I + \beta_n M]X_n - X + \beta_n b' \\ &= X_n - X - \beta_n[I - M]X_n + \beta_n(I - M)X \\ &= [(1 - \beta_n)I + \beta_n M](X_n - X) \\ &= M_n Y_n \ . \end{aligned}$$

Alors, à l'aide de $M_q = P\left[(1 - \beta_q)I + \beta_q(\Delta + J)\right]P^{-1} = PT_qP^{-1}$, nous avons :

$$Y_{n+1} = P\left(\prod_{q=1}^{n} T_q\right)P^{-1}Y_1 \ . \tag{I.2.43}$$

Remarque 12 Remarquons que pour tout $q = 1 \dots n$, la matrice T_q est un polynôme en J et Δ qui commutent. Ainsi, le symbole $\prod$, qui représente le produit dans un anneau commutatif est valable dans l'anneau commutatif des polynômes en J et Δ.

ii) Prouvons que $\prod_{q=1}^{n} T_q$ tend vers zéro lorsque n tend vers l'infini.

Nous utiliserons les assertions suivantes.

– D'après la remarque 12 et en factorisant J nous avons :

$$\prod_{q=1}^{n} T_q = \sum_{(\alpha_1 \ldots \alpha_n) \in \{0,1\}^n} \prod_{q=1}^{n} D_q^{\alpha_q} J^{(\alpha_1 + \ldots + \alpha_n)} \ . \tag{I.2.44}$$

– En multipliant à gauche par J on effectue un décalage à droite des colonnes. C'est pourquoi, nous avons la propriété de nilpotence :

$$(q \geq N) \Rightarrow (J^q = 0) \ . \tag{I.2.45}$$

– Notons $\sigma(T)$ le rayon spectral de T. Par définition de $D_q^{\alpha_q}$, nous avons d'une part pour $\alpha_q = 1$, $\sigma(D_q^{\alpha_q}) \leq |\beta_q| < 1 - \varepsilon$ et d'autre part pour $\alpha_q = 0$ $\sigma(D_q^{\alpha_q}) \leq |\Delta_q| < \zeta$ en appliquant le corollaire 1 avec la formule (I.2.36). Alors, nous pouvons majorer le rayon spectral de $D_q^{\alpha_q}$ par $\lambda = \max(\zeta, 1 - \varepsilon)$ qui est inférieur à 1. Comme $D_q^{\alpha_q}$ est diagonal, nous pouvons affirmer :

$$||D_q^{\alpha_q}|| \leq \lambda < 1 \ . \tag{I.2.46}$$

Les relations (I.2.44) et (I.2.45) donnent

$$\prod_{q=1}^{n} T_q = \sum_{\substack{(\alpha_1 \ldots \alpha_n) \in \{0,1\}^n \\ \sum_{l=1}^{n} \alpha_l < N}} \prod_{q=1}^{n} D_q^{\alpha_q} J^{(\alpha_1 + \ldots + \alpha_n)} \ . \tag{I.2.47}$$

En utilisant les équations (I.2.46), (I.2.47) et le fait que la norme du produit de deux matrices est inférieure au produit des normes de ces mêmes matrices, on a

$$\| \prod_{q=1}^{n} T_q \| \leq \sum_{\substack{(\alpha_1 \ldots \alpha_n) \in \{0,1\}^n \\ \sum_{l=1}^{n} \alpha_l < N}} \| J^{(\alpha_1 + \ldots + \alpha_n)} \| \lambda^n \ . \tag{I.2.48}$$

En réordonnant la somme ci-dessus, on a

$$\sum_{\substack{(\alpha_1 \ldots \alpha_n) \in \{0,1\}^n \\ \sum_{l=1}^{n} \alpha_l < N}} \| J^{(\alpha_1 + \ldots + \alpha_n)} \| \lambda^n = \sum_{q=0}^{N-1} \sum_{\sum_{l=1}^{n} \alpha_l = q} ||J||^q \lambda^n \tag{I.2.49}$$

Le nombre d'occurrences de $\|J\|^q$ dans (I.2.49) est le coefficient du binôme de Pascal C_n^q (par la définition du coefficient du binôme comme le cardinal de l'ensemble des n nombres logiques dont la somme est q). C'est pourquoi,

$$|| \prod_{q=1}^{n} T_q || \leq \sum_{q=0}^{N-1} C_n^q \lambda^n ||J||^q \ . \tag{I.2.50}$$

On déduit de

$$(\lambda < 1, q \leq N \in \mathbb{N}) \rightarrow (C_n^q ||J||^q \lambda^n \leq ||J||^N n^N \lambda^n$$

que la sommation (I.2.50) est bornée supérieurement par $N||J||^N n^N \lambda^n$, donc tend vers zéro quand n tend vers l'infini, i.e. :

$$\lim_{n \to \infty} \left\| \prod_{q=1}^{n} T_q \right\| = 0 \ . \tag{I.2.51}$$

De (I.2.43) et (I.2.51) on déduit $\lim_{n \to \infty} Y_n = 0$. De plus, $X_n = X + Y_n$ donne finalement

$$\lim_{n \to \infty} X_n = X \ .$$

□

Chapitre I.3

Analyse de la méthode.

Ce chapitre est dédié à l'étude de la précision de la formulation variationnelle. Il est organisé comme suit.

1. Rappels concernant l'ordre de convergence dans la Méthode des Eléments Finis (section I.3.1) et extension à notre méthode.
2. Série de tests numériques pour découvrir une loi de convergence (section I.3.2).
3. Majorations menant aux résultats de convergences théoriques, corollaires 4 et 5 (section I.3.3).
4. Etude du conditionnement par des simulations informatiques et démonstration de la loi linéaire (section I.3.4, théorème 8).

I.3.1 Notion d'ordre de convergence et rappels sur la méthode des éléments finis.

Soit un domaine bidimensionnel Ω borné. La théorie classique (cf [17]) des éléments finis montre que si l'on discrétise le problème

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ u = 0 & \text{sur } \Gamma \end{cases}$$

à l'aide d'éléments finis P_k, on obtient l'estimation[1] suivante, si $u \in H^{k+1}(\Omega)$ avec $k+1 \geq m$ (cf [54]), alors

$$||u - u_h||_{H^m(\Omega)} \leq Ch^{k+1-m}||u||_{H^{k+1}(\Omega)} \ .$$

Dans l'inégalité ci-dessus, h est le paramètre de raffinement du maillage (que l'on suppose régulier, vérifiant les hypothèses H1, H2 et H3 section I.2.1.2). L'exposant de h est ce que l'on appelle l'ordre de la méthode. Par exemple, pour $k = 1$, on a

$$||u - u_h||_{L^2(\Omega)} \leq Ch^2||u||_{H^2(\Omega)} \ .$$

Pour $k = 2$, on a

$$||u - u_h||_{L^2(\Omega)} \leq Ch^3||u||_{H^3(\Omega)} \ .$$

Comme dans la méthode des Eléments Finis (FEM) nous dirons que la méthode est d'ordre $n \in \mathbb{N}$ si, pour toute solution régulière, il existe une constante positive C telle que, pour une norme sur le bord Γ,

$$||x - x_h||_\Gamma \leq Ch^n$$

où x_h est défini par (I.2.15) et x par (I.0.8). Nous avons réalisé des tests numériques afin d'étudier l'ordre de convergence dans d'autres normes telles que $||x - x_h||_V$, $||u - u_h||_\Gamma$ et $||u - u_h||_\Omega$ section I.3.2. Les

[1] Cette estimation, obtenue à l'aide de fonctions polynômiales est optimale.

estimations théoriques effectuées dans ce chapitre porteront sur des estimations L^2 ou H^{-s} avec $s > 1/2$: nous supposerons donc que Γ est Lipschitz continue par morceaux.

L'ordre de convergence permet d'estimer asymptotiquement l'erreur en fonction de k et h. Nous sommes d'autre part en mesure de comparer les évolutions de la place mémoire et du temps de calcul entre la FEM et la formulation UWVF.

Prenons un exemple précis. Soit $\mathcal{C}$ un carré de base unitaire maillé par une triangulation structurée. Le maillage est constitué de $2n^2$ éléments. Dans cette section, on définit le paramètre de raffinement du maillage h par $h = 1/n$: c'est la longueur des côtés des triangles dans les deux directions d'axes[1]. Il y a $(n+1)^2$ nœuds, $4n$ arêtes libres, $3n^2 - 2n$ interfaces, donc, au total, $3n^2 + 2n$ arêtes. La figure I.3.1 donne le maillage de $\mathcal{C}$.

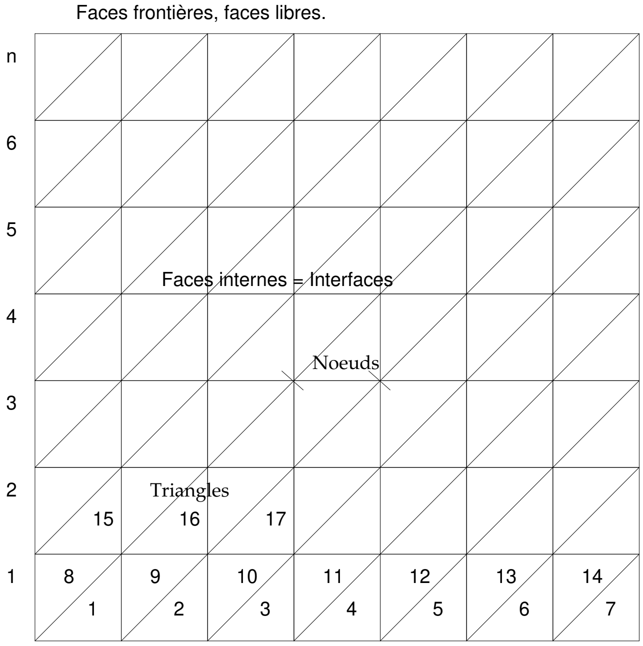


FIG. I.3.1 – Maillage de $\mathcal{C}$.

Le nombre de degrés de liberté dans la FEM par des éléments P_1, c'est-à-dire par des fonctions centrées aux nœuds, est exactement le nombre de nœuds $(n+1)^2$ si l'espace d'approximation est $H^1(\mathcal{C})$, ou le nombre de nœuds internes $(n-1)^2$ si l'espace d'approximation est $H_0^1(\mathcal{C})$ (lorsque l'on traite d'un problème de Dirichlet). Considérant des éléments P_2, on rajoute les fonctions dont les supports sont localisés sur les milieux des arêtes. Ces fonctions sont au nombre de $3n^2 + 2n$ pour $H^1(\mathcal{C})$ et $3n^2 - 2n$ moins le nombre d'interfaces ayant un nœud à la frontière $(8n - 6)$ pour $H_0^1(\mathcal{C})$, totalisant $4n^2 + 4n + 1$ et $4n^2 - 12n + 7$ respectivement.

Dans un contexte plus général, si l'on prend des éléments finis P_k sur la triangulation ci-dessus, le nombre de degrés de liberté de l'espace d'approximation V_h de $H^1(\mathcal{C})$ est $(n+1)^2$ le nombre de nœuds, plus $(k-1)$ fois le nombre d'arêtes $(3n^2+2n)$ plus le nombre d'éléments du maillage $(2n^2)$ fois la dimension de P_k qui est $\dfrac{(k+2)(k+1)}{2}$ (cf [54]) moins $3k$ qui est le nombre de nœuds des arêtes d'un triangle. On a donc $dim(V_h) = (n+1)^2 + (k-1)(3n^2+2n) + 2n^2(\dfrac{(k+2)(k+1)}{2} - 3k)$ soit $dim(V_h) = n^2k^2 + 2nk + 1$. Comptons maintenant le nombre d'éléments non nuls de la matrice du système linéaire en considérant la matrice ligne par ligne. Les éléments non nuls pour une ligne numéro q proviennent des couplages entre la fonction numéro q de V_h et ses fonctions voisines, c'est-à-dire les fonctions ayant un support d'intersection non vide avec le support de la q-ième fonction de base. Pour des éléments P_1, ceci est équivalent à compter le nombre de nœuds voisins, qui est de 6 pour un nœud interne, ce qui est le cas

[1]Cette définition est légèrement différente de la définition (I.3.5) qui donne $h = \sqrt{2}/2n$.

pour la plupart des nœuds. Pour des éléments P_2, ceci revient, pour chaque triangle, à compter les nœuds voisins et les nœuds milieux de chaque segment jusqu'à trois points. Si l'on considère une fonction à l'intérieur du domaine (ce qui représente la plupart des fonctions), on dénombre 36 nœuds voisins. Pour P_3, on a 90 fonctions à coupler. Ceci augmente avec k comme $2(2k)^2 - 2 + (2k-2)(2k-1) = 12k^2 - 6k$. Le stockage est équivalent (il est en réalité légèrement inférieur à cause des nœuds proches de la frontière) à $(12k^2 - 6k)(n^2k^2 + 2nk + 1) = 12\,n^2k^4 + 24\,nk^3 - 6\,n^2k^3 + 12\,k^2 - 12\,nk^2 - 6\,k$. Pour des Eléments Finis P_1, en considérant un grand nombre n, le stockage se comporte en $6n^2$. Pour des EF P_2, en $144n^2$, et enfin pour des EF P_3 en $810n^2$. La FEM a donc les caractéristiques suivantes

Tab. I.3.1 – FEM

éléments	nœuds	interfaces	arêtes libres	arêtes
$2n^2$	$(n+1)^2$	$3n^2 - 2n$	$4n$	$3n^2 + 2n$

Type d'éléments finis	P_1	P_2	P_3	P_k
Degrés de Liberté des EF	$(n+1)^2$	$4n^2 + 4n + 1$	$9n^2 + 6n + 1$	$n^2k^2 + 2nk + 1$
Erreur	n^{-2}	n^{-3}	n^{-4}	$n^{-(k+1)}$
Couplage de fonctions de base	6	36	90	$12k^2 - 6k$
Stockage	$6(n+1)^2$	$36(4n^2 + 4n + 1)$	$\approx 810n^2$	$\approx 6n^2k^3(2k-1)$

Dans la méthode UWVF, le nombre de degrés de liberté est le nombre de fonctions de base par élément (noté p) fois le nombre d'éléments $2n^2$ (si l'on en prend un nombre constant de fonctions de base par élément). Le stockage est donné par le couplage hermitien des fonctions de base sur chaque élément $n^2p(p+1)$, les couplages sur les $3n^2 - 2n$ interfaces du maillage, i.e. $(3n^2 - 2n)p^2$, et enfin par le couplage non hermitien des fonctions de base aux arêtes libres, ce qui rajoute $2np(p-1)$ termes. Ceci totalise $2n^2((p+1)(p)/2 + 3p^2) - 2np \approx 7n^2p^2$ unités de stockage. Nous voyons section I.3.3.5 par (I.3.67) que l'erreur sur $||u - u_h||_{L^2(\Gamma)}$ se comporte en $h^{[(p-1)/2]-1/2}||u||_{C^{[(p+1)]/2}(\Omega)}$.

Les comparaisons asymptotiques de la taille mémoire et de l'ordre de convergence entre la méthode UWVF et la méthode FEM sont résumées dans le tableau (I.3.2)[1].

Tab. I.3.2 – Comparaison UWVF-FEM

Méthode	FEM (k)	UWVF (p)
Ordre de l'erreur	$(k+1)$	$-1/2 + [(p-1)/2]$
Stockage	$12n^2k^4$	$7n^2p^2$
Degrés de Liberté	n^2k^2	$2n^2p$

Remarque 13 Les estimations pour la FEM sont données pour des normes sur le domaine entier Ω, alors que pour la méthode UWVF nous ne donnons d'estimation que sur la frontière Γ. C'est pourquoi les ordres en norme $L^2(\Omega)$ et en norme $L^2(\Gamma)$ diffèrent de 1/2. D'autre part, rappelons que l'ordre donné pour la méthode UWVF dans le tableau (I.3.2) est l'ordre théorique démontré dans ce chapitre : cet ordre n'est pas optimal, l'ordre numérique observé est la fonction

$$+1/2 + [(p-1)/2] \ .$$

Cette loi numérique montre l'intérêt de la méthode pour p petit.

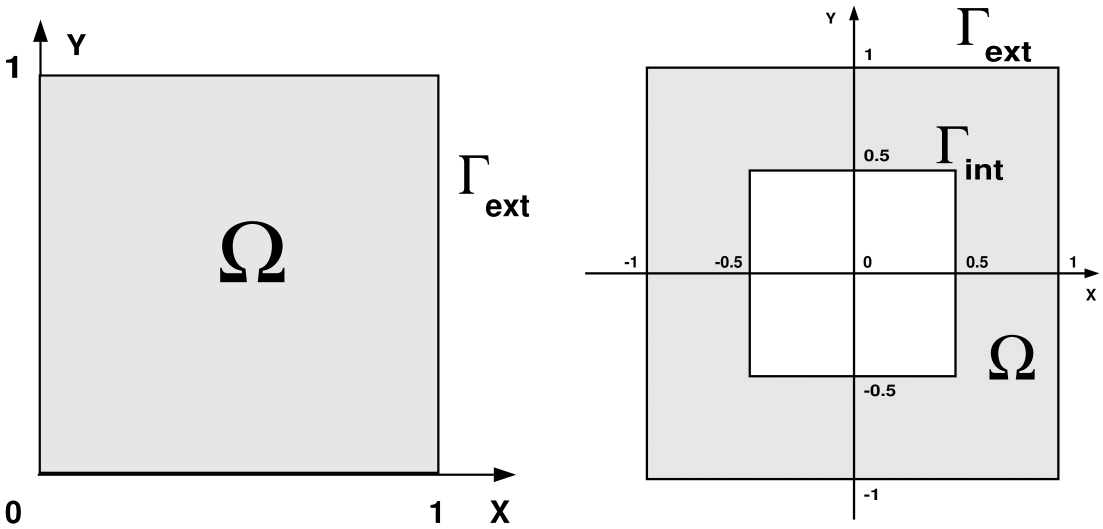
Remarque 14 Le tableau (I.3.2) met en lumière le fait que le même taux de précision peut être atteint en utilisant la méthode UWVF à la place de la FEM pour un stockage informatique réduit. Il en est de même du temps de calcul. Ceci provient du fait que la méthode UWVF est une méthode d'ordre asymptotiquement plus élevé que la FEM au sens où l'ordre de convergence est en racine carrée du stockage dans la méthode UWVF au lieu de la racine quatrième du stockage pour la FEM.

[1]Comme $p \geq 3$, l'ordre de la méthode est supérieur à 1/2.

I.3.2 Etude numérique de l'ordre de convergence.

Nous nous sommes intéressés aux deux domaines Ω de la figure I.3.2

FIG. I.3.2 – Types de domaines Ω.



a) Domaine connexe b) Domaine non simplement connexe

Afin de pouvoir estimer aisément la différence entre la solution exacte u du problème de Helmholtz et la solution approchée, nous nous sommes donné une solution exacte simple sous la forme d'une onde plane de direction complexe $\vec{v}_0$. En notant $\vec{x} = (x_1, x_2)$ la position dans le plan, la solution de référence est donnée par[1]

$$u = e^{i\omega\vec{v}_0.\vec{x}} \text{ avec } \vec{v}_0 = (v_0^1, v_0^2) \ . \tag{I.3.1}$$

Ce type de solution, selon les valeurs de $\vec{v}_0$ donne lieu à deux problèmes dont on calcule simplement le second membre (avec des conditions aux limites précisées aux paragraphes I.3.2.3 et I.3.2.4) :

i) pour $(v_0^1)^2 + (v_0^2)^2 = 1$, on a $(-\Delta - \omega^2)u = 0$, c'est-à-dire $f = 0$ (problème de Helmholtz homogène),

ii) pour $(v_0^1)^2 + (v_0^2)^2 \neq 1$, on a $(-\Delta - \omega^2)u \neq 0$, c'est-à-dire $f \neq 0$ (problème de Helmholtz non homogène).

I.3.2.1 Calcul numérique de l'erreur.

On calcule l'erreur entre la solution exacte x de (I.0.8) et la solution approchée x_h définie par (I.2.16) : $x_h = \sum_{l=1}^{L} x_{kl}\, z_{kl}$. Dans le cas homogène, on calculera aussi l'erreur entre la solution exacte u et la solution approchée u_h calculée par (I.2.27) : $u_h = \sum_{l=1}^{L} x_{kl}\, e_{kl}$. Dans le cas non homogène, on peut calculer u_h par (I.2.23).

Nous montrerons un calcul effectué à l'aide de la formule (I.2.27) (valable dans le cas homogène) pour montrer les limitations de cette formule.

Explicitons deux calculs d'erreur relative :

i) Calcul de $||x - x_h||_V$.

Par définition du produit scalaire sur V, on a :

$$||x - x_h||_V^2 = \sum_k \int_{\partial\Omega_k} |x - \sum_{l=1}^{n} x_{kl}\, z_{kl}|^2 \ . \tag{I.3.2}$$

[1]Noter que $\vec{v}_0$ n'est pas nécessairement dans l'espace des directions de propagation des fonctions de base.

En développant le carré du module de (I.3.2) on intervertit le signe somme sur l et le signe intégral sur $\partial\Omega_k$, on obtient l'égalité (I.3.3).

$$\text{(I.3.3)} \qquad \begin{aligned} ||x - x_h||_V^2 = & \sum_k \int_{\partial\Omega_k} |x|^2 - 2Re\left[\sum_{l=1}^{n}\sum_k \int_{\partial\Omega_k} x_{kl}\, z_{kl}\overline{x}\right] \\ & + \sum_{l=1}^{n}\sum_{m=1}^{n}\sum_k \int_{\partial\Omega_k} x_{kl}\, z_{kl}\overline{x_{km}\, z_{km}} \ . \end{aligned}$$

Le premier terme de (I.3.3) est $||x||_V^2$. Le deuxième terme de (I.3.3) se calcule presque comme ${}^\top bX$. La différence est que l'on somme sur $\partial\Omega_k$ et non seulement sur Γ_k et que l'on prend $-\partial_{\nu_k}$ au lieu de ∂_{ν_k}. On reconnaît que le troisième terme de l'égalité (I.3.3) est exactement ${}^\top XDX$.

ii) Dans le cas homogène, calcul de $||u - u_h||_{L^2(\Omega)}^2$.

On effectue une inversion du signe somme sur l et du signe intégral sur Ω_k comme en i). On a alors

$$\text{(I.3.4)} \qquad \begin{aligned} ||u - u_h||_{L^2(\Omega)}^2 = & \sum_k \int_{\Omega_k} |u|^2 - 2Re\left[\sum_{l=1}^{n}\sum_k \int_{\Omega_k} x_{kl}\, u_{kl}\overline{u}\right] \\ & + \sum_{l=1}^{n}\sum_{m=1}^{n}\sum_k \int_{\Omega_k} x_{kl}\, u_{kl}\overline{x_{km}\, u_{km}} \ . \end{aligned}$$

Le premier terme de (I.3.4) est $||u||_{L^2(\Omega)}^2$. Le deuxième terme de (I.3.4) se calcule comme le deuxième terme de (I.3.3), à la différence que l'on effectue des intégrales sur des surfaces au lieu d'arêtes. On reconnaît que le troisième terme de l'égalité (I.3.4) est exactement ${}^\top XD'X$ où D' est la matrice du produit scalaire des fonctions (e_{kl}, e_{km}) sur $\oplus_k L^2(\Omega_k)$.

Conclusion : Le calcul numérique de l'erreur se fait donc sans difficulté supplémentaire par rapport à l'assemblage de la matrice D, et toujours à l'aide d'une formule d'intégration exacte.

I.3.2.2 Une simulation informatique de calcul d'erreur.

La figure I.3.3 indique la valeur de l'erreur relative sur u en norme $L^2(\Omega)$ en fonction de $1/h$ pour différents nombres de fonctions de base p par élément. Rappelons que h est le raffinement du maillage que l'on définit par :

$$\text{(I.3.5)} \qquad h = \sqrt{\frac{\text{Surface de } \Omega}{\text{Nombre total d'éléments de la triangulation}}} \ .$$

Les paramètres du cas sont dans le tableau I.3.3. La procédure adoptée pour calculer l'ordre de convergence consiste à mesurer la pente négative maximale des courbes (en échelle logarithmique). On constate que le faisceau de courbes de la figure I.3.3 est composé de droites approximativement parallèles pour p impair et pour $p+1$, droites de pentes négatives dont la valeur absolue augmente avec la valeur de p.

Remarque 15 Nous devons éliminer les points qui donnent une pente positive. Ces points apparaissent lorsque la discrétisation devient trop fine (en espace quand le paramètre de taille du maillage h tend vers zéro, ou quand p le nombre de fonctions de base par élément tend vers l'infini). Ceci provient du lien entre l'ordre et le conditionnement des matrices blocs D_k (D est diagonale par blocs). En effet, on montre que le conditionnement des matrices D_k est supérieur à $h_k^{-2[p/2]+2}$ pour $p \geq 4$, (théorème 8 p. 56), estimation corroborée par les tests numériques (figure I.3.8). Les matrices D_k^{-1} sont alors mal calculées, ce qui nous prive du gain en précision dû au grand nombre de fonctions de base. Noter que ceci n'est pas en pratique un handicap pour l'utilisation de la méthode, ce genre de problème ne survenant que lorsque la discrétisation est excessivement fine. En revanche, on a constaté numériquement (figure I.3.8) et prouvé (annexe III.E.1) que, pour trois fonctions de base par élément, le conditionnement des matrices D_k ne dépend pas de h.

Remarque 16 Nous avons constaté empiriquement que le nombre d'itérations de l'algorithme de Richardson augmente quand le conditionnement de la matrice hermitienne D augmente. A titre purement indicatif et expérimental, indiquons que 200 itérations suffisent pour résoudre un problème pour lequel la matrice D est bien conditionnée (voir par exemple la figure I.4.17), 600 dans le cas d'un problème pour lequel la matrice D est mal conditionnée (et qui mène à des normes d'erreur relative de l'ordre de la racine carrée de la précision machine) : c'est le nombre maximal d'itérations effectuées pour les calculs de la figure I.3.3. On peut donc conclure que,

- dans le cadre de notre formulation, la technique de résolution du système linéaire par la méthode de Richardson est parfaitement robuste ; ceci provient de notre connaissance du spectre de la matrice $D^{-1}C$.
- La méthode ultra-faible, et l'algorithme de résolution du système discret, ne sont pas, en pratique, sujets aux problèmes de conditionnement. Au contraire, les problèmes de conditionnement des systèmes linéaires issus de la FEM ou des équations intégrales obligent souvent l'emploi d'un préconditionneur, ce qui est une opération coûteuse. Le nombre d'itérations des algorithmes de résolution de ces méthodes augmente aussi avec le conditionnement, et pour une précision finale de calcul beaucoup plus faible.
- La méthode est remarquablement précise, par exemple jusqu'à une précision de 10^{-6} sur la norme relative $||u - u_h||_{L^2(\Omega)}$ dans le cas de la figure I.3.3.

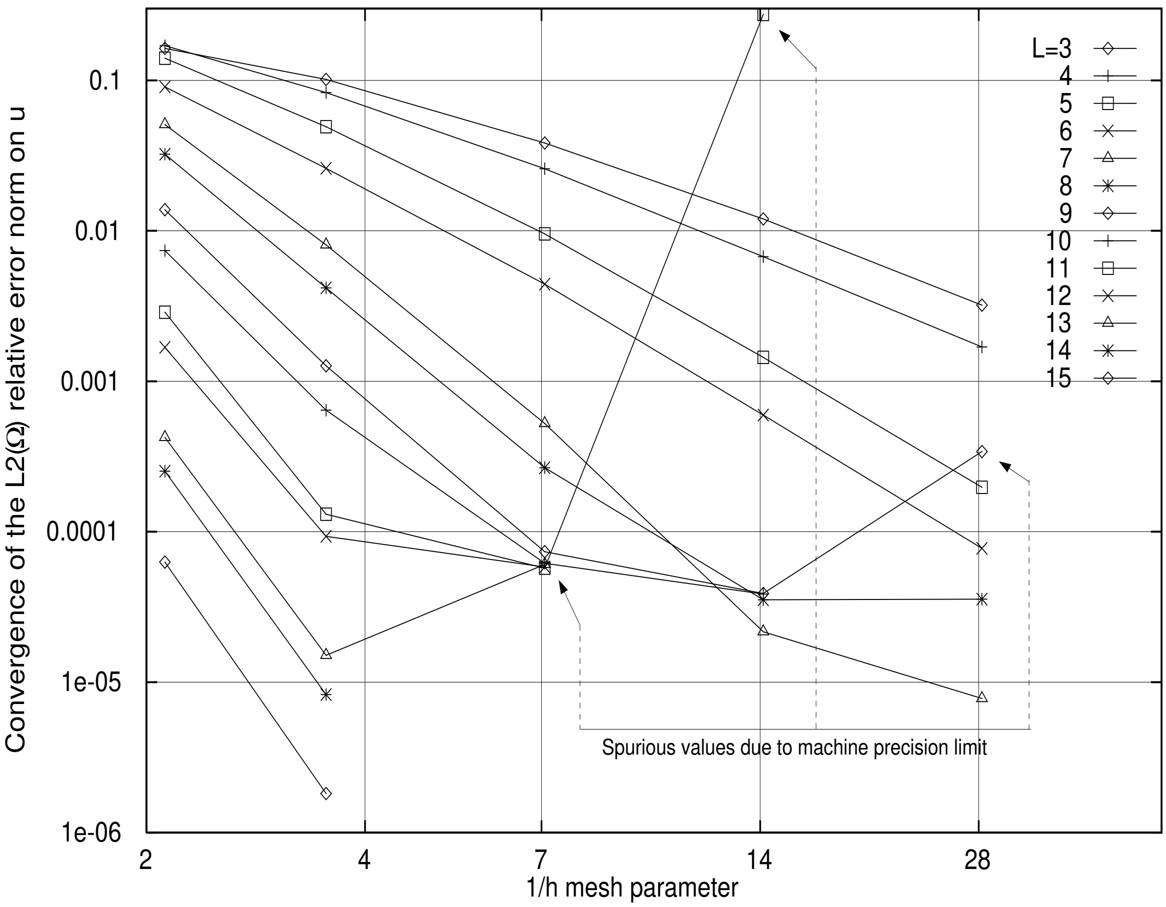


FIG. I.3.3 – Faisceau de courbes d'erreur en fonction de $1/h$. Le nombre de fonctions de base par élément varie de 3 à 15 (de haut en bas).

I.3.2.3 Calculs d'ordre dans le cas homogène : $f = 0$.

La condition aux limites g est :

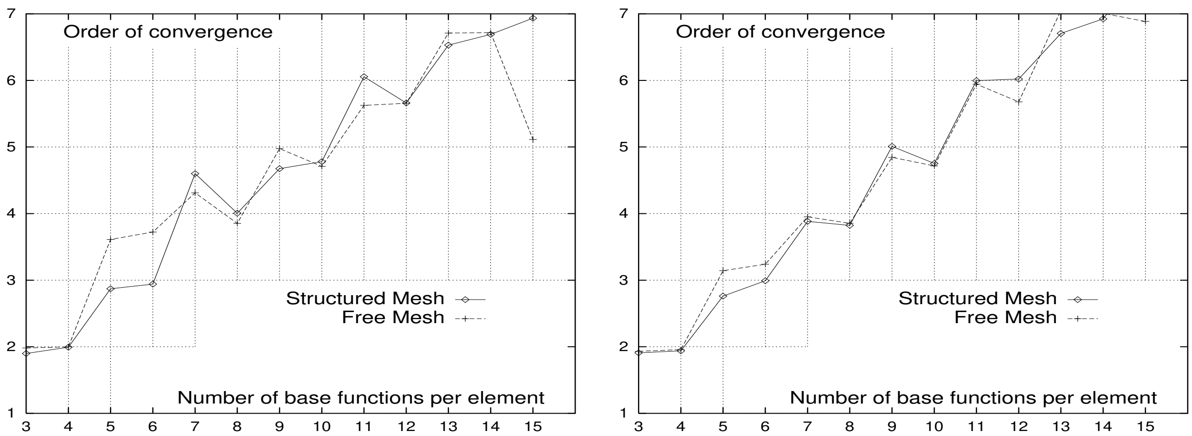
$$g = ((1 + Q)\partial_\nu + (1 - Q)i\omega)e^{(i\omega\vec{v}_0.\vec{x})} \ , \tag{I.3.6}$$

avec $\vec{v}_0 = (v_0^1, v_0^2)$ vérifiant $(v_0^1)^2 + (v_0^2)^2 = 1$. La solution exacte est $e^{(i\omega\vec{v}_0.\vec{x})}$. Les autres données du problème sont dans le tableau I.3.3 : la fréquence de calcul est de 600 MHz ($\omega = 4\pi$).

TAB. I.3.3 – Paramètres du cas homogène

Variables	**Valeur**
f	0
Q	0.1
$\vec{v}_0$	$(1.009946454058, i \times 0.1413925035682)$
ω	12.56637061436 ($\lambda = 0,5$ m)

Nous évaluons l'évolution de l'ordre de convergence pour un maillage structuré et un maillage non structuré en fonction du nombre de fonctions de base. La norme est la norme d'erreur relative sur u -Figures I.3.4 et I.3.5.



Norme $L^2(\Omega)$ Norme $L^2(\Gamma)$

FIG. I.3.4 – Cas $f = 0$, ordres sur u, approché par (I.2.27).

Pour $p = 3$ fonctions de base par élément, on observe, figure I.3.4, que

$$\begin{aligned} ||u - u_h||_{L^2(\Omega)} &\leq Ch^2 \\ ||(u - u_h)_{|\Gamma}||_{L^2(\Gamma)} &\leq Ch^2 \ . \end{aligned} \tag{I.3.7}$$

L'estimation dans $L^2(\Omega)$ semble optimale au sens que trois fonctions de base permettent d'approcher les termes d'ordre zéro et le gradient de la solution de façon à ce que u_h approche u en $O(h^2)$. L'estimation dans $L^2(\Gamma)$ est meilleure que celle que l'on attendrait d'après les théorèmes de trace (perte de 1/2 sur l'ordre par rapport à l'estimation dans $L^2(\Omega)$).

Remarque 17 L'approximation (I.2.27) de u est meilleure que celle donnée par (I.2.23) : on observe un décalage de 1/2 de l'ordre de convergence entre les deux types d'approximation : la formule (I.2.23) donne la même loi d'ordre que celles des courbes I.3.5.

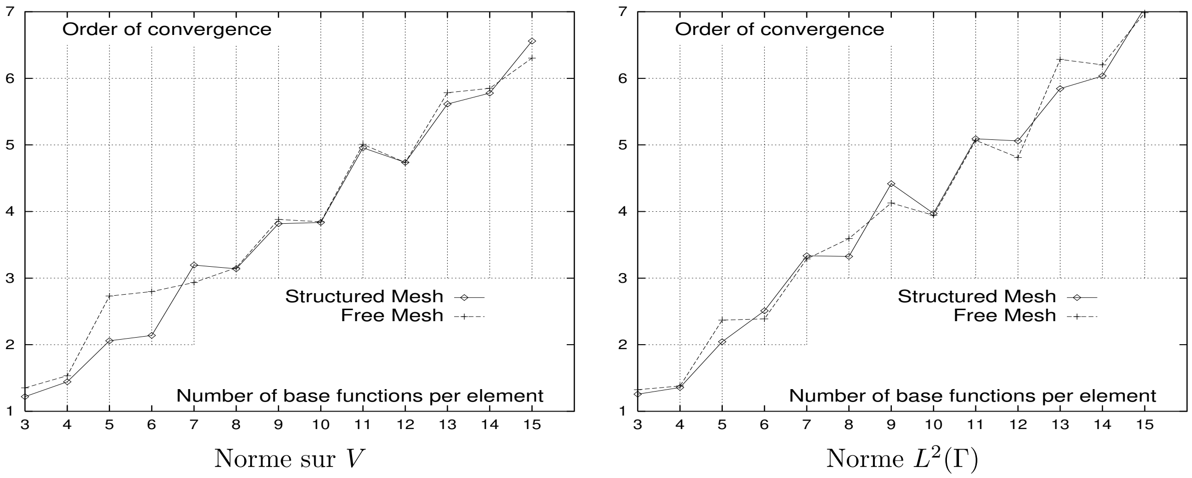


FIG. I.3.5 – Cas $f = 0$, ordres sur x.

I.3.2.4 Calculs d'ordre dans le cas non homogène : $f \neq 0$.

Par rapport au cas précédent de la section I.3.2.3, on choisit $\vec{v}_0 = (v_0^1, v_0^2)$ vérifiant

$$(v_0^1)^2 + (v_0^2)^2 = 1 + \frac{\mu}{\omega^2},\ \mu \in \mathbb{C},\ \mu \neq 0\ ,$$

de façon à ce que $e^{(i\omega\vec{v}_0.\vec{x})}$ soit la solution du problème (I.0.1) avec

$$\begin{cases} f = \mu e^{(i\omega\vec{v}_0.\vec{x})} \\ g = ((1+Q)\partial_\nu + (1-Q)i\omega)e^{(i\omega\vec{v}_0.\vec{x})}\ . \end{cases} \tag{I.3.8}$$

Le tableau I.3.4 complète ces données de façon à définir le cas test étudié.

TAB. I.3.4 – Paramètres du cas non homogène

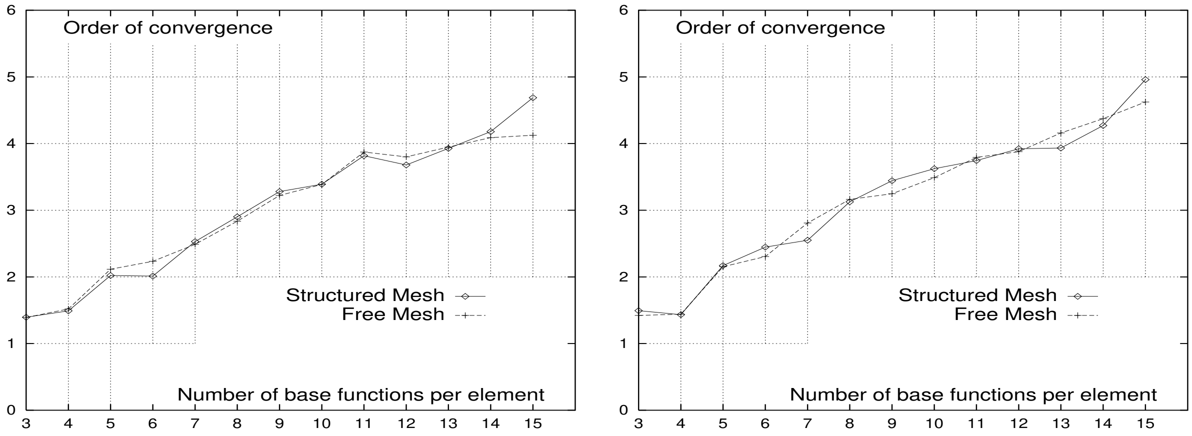
Variables	**Valeur**
Q	0.1
$\vec{v}_0$	$(1.4282799726, i \times 0.199959196164)$
ω	12.56637061436
μ	157.9136704174

La figure I.3.6 montre l'évolution de l'ordre de l'erreur sur la norme de x.

Remarque 18 D'après le lemme 10 qui donne l'estimation (I.3.36), on sait que les estimations de u sur Γ restent bonnes à condition de bien calculer u_h sur Γ par (I.2.23) section I.2.1.5. Ainsi, l'ordre de l'erreur relative sur u_h en norme $L^2(\Gamma)$ ne peut pas être inférieur à celui donnée par la figure I.3.6. En pratique, la loi est la même (nous ne montrons pas ces simulations qui n'apportent pas grand chose par rapport aux courbes I.3.6). En revanche, regardons ce qui se passe si l'on essaie d'approcher u par u_h, combinaison linéaire des fonctions e_{kl}, situation adaptée au cas homogène (équation (I.2.27), section I.2.1.5). On constate que les normes

$$\frac{||u - u_h||_{L^2(\Omega)}}{||u||_{L^2(\Omega)}} \text{ et } \frac{||u - u_h||_{L^2(\Gamma)}}{||u||_{L^2(\Gamma)}}$$

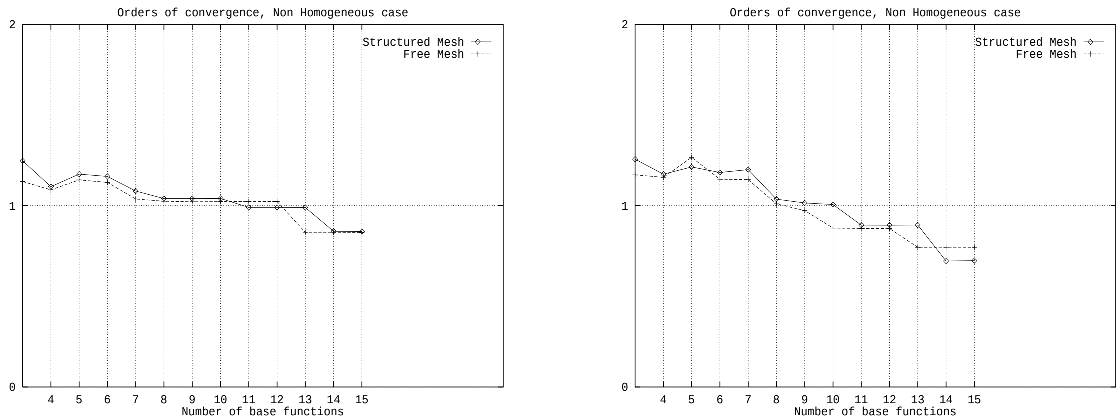
de la figure I.3.7 n'évoluent pas en fonction de p, le nombre de fonctions de base par élément. Elles suivent une loi de convergence en h^1, le paramètre de raffinement du maillage.



Norme sur V Norme $L^2(\Gamma)$

FIG. I.3.6 – Cas $f \neq 0$, ordres sur x.

Cet ordre de convergence constant était attendu, on essaie en effet d'approcher une solution d'un problème non homogène par une solution homogène (noter la lente dégradation du résultat, due aux problèmes de conditionnement). Jusqu'à présent, l'approximation de u dans Ω est plutôt mauvaise. Il faut garder à l'esprit que, sur le bord Γ, nous devons absolument prendre la formule d'approximation (I.2.23) et non celle qui est plus aisée à calculer numériquement (I.2.27).



Norme $L^2(\Omega)$ Norme $L^2(\Gamma)$

FIG. I.3.7 – Cas $f \neq 0$, ordres sur u, approché par (I.2.27).

I.3.2.5 Bilan : lois approchées de convergence numérique.

D'après les figures (I.3.4), (I.3.5), (I.3.6) et (I.3.7), il semble que nous pouvons dégager des lois d'ordre de convergence. Nous séparons ces lois en deux cas, le cas $f = 0$ et le cas $f \neq 0$. Pour chacune des normes étudiées, on aura donc deux lois, une pour le problème homogène et une autre pour le problème non homogène.

1. Les normes $\dfrac{||x - x_h||_V}{||x||_V}$ et $\dfrac{||x - x_h||_{L^2(\Gamma)}}{||x||_{L^2(\Gamma)}}$ suivent les lois du tableau I.3.5.
2. La norme $\dfrac{||u - u_h||_{L^2(\Gamma)}}{||u||_{L^2(\Gamma)}}$ suit la loi du tableau I.3.5 lorsque la méthode de calcul utilise la formule d'approximation générale (I.2.23).

3. Les normes $\frac{||u-u_h||_{L^2(\Omega)}}{||u||_{L^2(\Omega)}}$ et $\frac{||u-u_h||_{L^2(\Gamma)}}{||u||_{L^2(\Gamma)}}$ suivent les lois du tableau I.3.6, à l'aide de la formule d'approximation (I.2.27) qui ne devrait être utilisée que dans le cas homogène.

TAB. I.3.5 – Bilan : ordres de convergence en fonction de p

p nb. de f.b. par élt		3 4	5 6	7 8	9 10	11 12	13 14
cas	Non homogène	1.5	2	2.5	3	3.5	4
	Homogène	1.5	2.5	3.5	4.5	5.5	?

TAB. I.3.6 – Bilan : ordres de convergence en fonction de p

p nb. de f.b. par élt		3 4	5 6	7 8	9 10	11 12	13 14
cas	Non homogène	1	1	1	1	1	1
	Homogène	2	3	4	5	6	?

Remarque 19 On observe une croissance linéaire de l'ordre de convergence en fonction du nombre de fonctions de base p par élément. En première approximation, la pente de la loi de convergence du problème non homogène est deux fois plus faible que celle du problème homogène.

Remarque 20 Nous avons des résultats numériques plus complets dans l'étude du problème non homogène que dans celle du problème homogène (d'où les points d'interrogation dans les tableaux I.3.5 et I.3.6). En effet, alors que le conditionnement ne dépend pas du second membre f du problème (I.0.1), l'ordre de convergence du problème non homogène augmente à un taux plus faible que celui du problème homogène, d'environ la moitié. La perte de précision numérique due au conditionnement importe plus dans un calcul très précis (ce qui est le cas du problème homogène) que dans un calcul moins précis ((ce qui est le cas du problème non homogène). Ces constatations sont expliquées théoriquement sections I.3.3.5 et I.3.4 avec les corollaires 4 et 5 et le théorème 8.

Remarque 21 Il peut être utile de relire la remarque 18 pour comprendre les différences entre le cas non homogène du tableau I.3.5 et du tableau I.3.6. Ces deux tableaux donnent l'ordre de convergence de la trace de u sur Γ. Dans le cas homogène, on obtient une amélioration d'environ 1/2 de la norme de u sur Γ par les formules (I.2.27) à la place de l'approximation (I.2.23).

I.3.2.6 Conclusion de l'étude numérique de l'ordre.

Cette étude numérique de l'ordre de convergence du problème de Helmholtz sans coefficient a clairement montré l'existence de lois linéaires en fonction de la partie entière de la moitié du nombre de fonctions de base par élément. Cette propriété sera étudiée sur le plan théorique dans la section suivante.

I.3.3 Etude théorique de l'ordre de convergence.

Le but de cette section est d'expliquer sur le plan théorique les résultats observés numériquement dans la section précédente. La raison pour laquelle l'ordre de convergence du problème homogène est majoré par une loi linéaire du nombre de fonctions de base par élément est enracinée dans l'utilisation de fonctions de base solutions du problème dual homogène. Une technique de dualité, équivalente au lemme d'Aubin-Nitsche, étend cette loi aux problèmes non homogènes.

Pour simplifier l'étude, nous allons nous intéresser aux normes sur le bord Γ.

L'étude est découpée en plusieurs parties.

1. Nous établissons l'estimation qui relie l'erreur $(x-x_h)$ à l'erreur d'interpolation $||(I-P_h)x||$ (section I.3.3.1, lemme 8) :
$$||(I-A)(x-x_h)|| \leq 2||(I-P_h)x|| \ ,$$
où P_h dénote le projecteur orthogonal sur l'espace de discrétisation V_h.
2. Par une technique de dualité, nous montrons (section I.3.3.2, théorème 6) une estimation en normes de Sobolev à exposants négatifs (supposant une régularité plus faible que L^2) sur Γ : pour tout $s > 1/2$,
$$||x-x_h||_{H^{-s}(\Gamma)} \leq 2||(I-P_h)x|| \sup_{\psi \in H^s(\Gamma)} \frac{||(I-P_h)y(\psi)||}{||\psi||_{H^s(\Gamma)}}$$
où $y(\psi)$ est donné par la solution du problème de Helmholtz homogène.
3. Avec l'hypothèse $|Q| \leq \delta < 1$, on obtient des majorations de l'erreur à la frontière (lemme 9) en norme $L^2(\Omega)$
$$||x-x_h||_{L^2(\Gamma)} \leq \frac{2}{\sqrt{1-\delta^2}}||(I-P_h)x|| \ .$$
4. Nous estimons l'erreur $u-u_h$ au bord à partir de l'erreur $x-x_h$ (lemme 10, section I.3.3.3.2) dans tout espace de Sobolev pour s positif ou nul
$$||u-u_h||_{H^{-s}(\Gamma)} \leq \frac{1+\delta}{2\omega}||x-x_h||_{H^{-s}(\Gamma)} \ .$$
5. Dans le cas particulier du problème de Helmholtz homogène ($f = 0$ dans (I.0.1)) et u de classe $C^{[(p+1)]/2}$, nous estimons le terme d'interpolation $||(I-P_h)x||_V$ (section I.3.3.4, théorème 7). Pour p fonctions de base par élément construites à partir d'ondes planes, et pour un maillage régulier, on a
$$||(I-P_h)x||_V \leq Ch^{[(p-1)]/2-1/2}||u||_{C^{[(p+1)]/2}(\Omega)} \ .$$
6. A l'aide du théorème 7 nous obtenons les résultats finaux concernant les estimations d'ordres de convergence en des normes sur la frontière (section I.3.3.5). Dans le cas spécifiquement homogène, nous avons des estimations en norme $L^2(\Gamma)$ (corollaire 4, lemme 9). Une estimation plus générale (sans hypothèse sur f) complète ce résultat (corollaire 5). Cette estimation est valable dans un espace de Sobolev à exposant négatif et est meilleure que celle dans l'espace $L^2(\Gamma)$ (de fonctions plus régulières). Le lemme 10 stipule que les estimations sont du même ordre sur $u-u_h$ que sur $x-x_h$.

Cette étude ne concerne que des normes sur le bord. Nous ne faisons pas d'estimation intérieure. De telles estimations peuvent faire l'objet d'une étude ultérieure par une technique de dualité appropriée.

I.3.3.1 Estimation du résidu.

Lemme 8 *Soit $x \in V$ la solution de (I.1.27) et $x_h \in V_h$ la solution de (I.2.1). Soit P_h le projecteur orthogonal sur V_h. Nous avons :*

(I.3.9)
$$(I-A)(x-x_h) \in V_h^{\perp}$$

(I.3.10)
$$\boxed{||(I-A)(x-x_h)|| \leq 2||(I-P_h)x||} \ .$$

Preuve. Pour cette preuve, le lecteur peut se reporter à [25]. Nous en donnons ici une légère variante sans différence fondamentale.

Appliquant P_h aux égalités

(I.3.11)
$$\begin{aligned} &(I-A)x = b \\ &(I-P_hA)x_h = b_h = P_hb \end{aligned}$$

nous avons

(I.3.12)
$$P_h(x-x_h) = P_hA(x-x_h)$$

qui est équivalent à (I.3.9). D'après l'équation (I.3.12), nous avons

$$\begin{aligned} x - x_h &= x - P_h x + P_h x - x_h \\ &= (I - P_h)x + P_h A(x - x_h) \\ &= (I - P_h)x - (I - P_h)A(x - x_h) + A(x - x_h) \ , \end{aligned} \tag{I.3.13}$$

d'où,

$$(I - A)(x - x_h) = (I - P_h)x - (I - P_h)A(x - x_h) \ . \tag{I.3.14}$$

Il reste maintenant à estimer le terme de droite de l'équation ci-dessus (I.3.14). Puisque P_h est un projecteur orthogonal, nous avons

$$||(I - P_h)A(x - x_h)||^2 = ||A(x - x_h)||^2 - ||P_h A(x - x_h)||^2 \ . \tag{I.3.15}$$

En utilisant (I.3.12) dans le terme de droite de (I.3.15), on obtient

$$||(I - P_h)A(x - x_h)||^2 = ||A(x - x_h)||^2 - ||P_h(x - x_h)||^2 \ , \tag{I.3.16}$$

et puisque $||A|| \leq 1$:

$$||(I - P_h)A(x - x_h)||^2 \leq ||(x - x_h)||^2 - ||P_h(x - x_h)||^2 \ . \tag{I.3.17}$$

Toujours puisque P_h est un projecteur orthogonal, nous avons

$$||(I - P_h)A(x - x_h)||^2 \leq ||(I - P_h)(x - x_h)||^2 \ , \tag{I.3.18}$$

et comme $(I - P_h)x_h = 0$

$$||(I - P_h)A(x - x_h)||^2 \leq ||(I - P_h)x||^2 \ . \tag{I.3.19}$$

Des équations (I.3.14) et (I.3.19), on écrit finalement :

$$||(I - A)(x - x_h)|| \leq 2||(I - P_h)x|| \ . \tag{I.3.20}$$

□

I.3.3.2 Une estimation au bord en norme de Sobolev négatif.

Nous réalisons cette estimation par une technique de dualité. C'est l'équivalent du lemme d'Aubin-Nitsche. Le résultat présenté sera utile dans l'étude du cas $f \neq 0$.

Théorème 6 *Soit Ω un domaine borné de $\mathbb{R}^2$ de frontière Γ assez régulière, nous prendrons C^∞ pour simplifier. Supposons Q (l'opérateur de bord du problème de Helmholtz (I.0.1)) une fonction à valeur complexe et $|Q| \leq 1$. Considérons $x \in V$ la solution de (I.1.27) et $x_h \in V_h$ la solution de (I.2.1). Soit P_h le projecteur orthogonal sur l'espace V_h. Soit $s > 1/2$ et $\psi \in H^s(\Gamma)$. Nous définissons w par :*

$$\begin{cases} (-\Delta - \omega^2)w = 0 & \text{dans } \Omega \\ (-\partial_\nu + i\omega)w = \overline{Q}(+\partial_\nu + i\omega)w + \psi & \text{sur } \Gamma \ . \end{cases} \tag{I.3.21}$$

et $y(\psi) \in V$ par

$$y(\psi)_{|\partial\Omega_k} = (-\partial_{\nu_k} + i\omega)w \quad \text{sur } \partial\Omega_k \ .$$

Alors, on a $\forall s > 1/2$

$$\boxed{||x - x_h||_{H^{-s}(\Gamma)} \leq 2||(I - P_h)x|| \sup_{\psi \in H^s(\Gamma)} \frac{||(I - P_h)y(\psi)||}{||\psi||_{H^s(\Gamma)}}} \tag{I.3.22}$$

Preuve.

i) Montrons que $y(\psi) \in V$ et $w \in C^m(\overline{\Omega})$ avec $m = [s + 1/2]$.

– D'après les résultats de régularité, (cf le théorème 2, section I.1.1.2, ou [44]), on sait que si $g \in H^s(\Gamma)$ (et $s \geq 1/2$) alors $w \in H^{s+3/2}(\Omega)$.
– On utilise les inégalités de Sobolev dans le cas de la dimension $N = 2$ et pour les fonctions de carré intégrable (cf [11], corollaire IX.13). On pose $m = [s+3/2-2/2] \in \mathbb{N}-\{0\}$ et $\theta = s+1/2-k \in]0,1[$. Alors, on a injection continue de $H^{s+3/2}(\Omega)$ dans $C^m(\overline{\Omega})$ et $m \geq 1$ puisque $s > 1/2$. Par exemple, pour $s = 1/2 + \theta$, on a injection continue de $H^{2+\theta}$ dans $C^1(\overline{\Omega})$.
– Des deux points précédents, on a $w \in C^m(\overline{\Omega})$ avec $m \geq 1$, ce qui implique que son gradient est continu sur $\overline{\Omega}$. Comme Ω est borné, $y(\psi) \in V$ pour tout $s > 1/2$.

On a donc $m = [s + 1/2]$ où $[\alpha]$ désigne la partie entière de α.

ii) Vérifions que $y(\psi)$ est tel que $(I - A^*)y(\psi) = (\psi)_{|\Gamma}$.

Notons A^* l'adjoint de A (A défini en (I.1.24) section I.1.2.1). En effet, w vérifie

$$\begin{cases} (\partial_{\nu_j} + i\omega)w_j = (-\partial_{\nu_k} + i\omega)w_k & \text{sur tout } \Sigma_{kj} \\ (-\partial_{\nu_k} + i\omega)w_k = \overline{Q}(+\partial_\nu + i\omega)w + \psi & \text{sur tout } \Gamma_k \ , \end{cases}$$

donc, pour tout $z \in V$, on a

$$\sum_k \int_{\partial\Omega_k} (-\partial_{\nu_k} + i\omega)w_k \overline{z_{|\partial\Omega_k}} - \sum_{kj} \int_{\Sigma_{kj}} (\partial_{\nu_j} + i\omega)w_j \overline{z_{|\Sigma_{kj}}} - \sum_k \int_{\Gamma_k} \overline{Q_k}(+\partial_{\nu_k} + i\omega)w_k \overline{z_{|\Gamma_k}}$$
$$= \sum_k \int_{\Gamma_k} \psi \overline{z_{|\Gamma_k}} \ .$$

Autrement dit,

$$\forall z \in V, \ (y(\psi), z)_V - (Fy(\psi), \Pi z)_V = (\psi, z)_{L^2(\Gamma)} \ . \tag{I.3.23}$$

iii) Prouvons l'inégalité $|(x - x_h, \psi)_\Gamma| \leq 2||(I - P_h)x|| \ ||(I - P_h)y(\psi)||$.

En effet, d'après (I.3.21), et comme $P_h(I - A)(x - x_h) = 0$ (lemme 10), on a :

$$\begin{aligned} (x - x_h, \psi)_\Gamma &= (x - x_h, (I - A^*)y(\psi))_V \\ &= ((I - A)(x - x_h), y(\psi))_V \\ &= ((I - A)(x - x_h), (I - P_h)y(\psi))_V \ . \end{aligned} \tag{I.3.24}$$

De (I.3.24) et de l'équation (I.3.10) du lemme 10, on a, par l'inégalité de Cauchy-Schwarz :

$$|(x - x_h, \psi)_\Gamma| \leq 2||(I - P_h)x|| \ ||(I - P_h)y(\psi)|| \ . \tag{I.3.25}$$

iv) Concluons à l'aide de (I.3.12) et

$$||x - x_h||_{H^{-s}(\Gamma)} = \sup_{\psi \in H^s(\Gamma)} \frac{(x - x_h, \psi)_{L^2(\Gamma)}}{||\psi||_{H^s(\Gamma)}} \ . \tag{I.3.26}$$

□

Remarque 22 L'intérêt de (I.3.22) vient de ce que $\sup_{\psi \in H^s(\Gamma)} \frac{||(I - P_h)y(\psi)||}{||\psi||_{H^s(\Gamma)}}$ tend vers zéro avec h comme l'indique le théorème (7). L'estimation obtenue est meilleure dans un espace de Sobolev négatif $H^{-s}(\Gamma)$ que dans l'espace classique $L^2(\Gamma)$ de mesure d'énergie.

Remarque 23 Ce genre de majoration dans l'espace de Sobolev $H^{-s}(\Gamma)$ avec $s > 1/2$ garde un intérêt pratique. En effet, l'amplitude de diffusion, importante pour les applications (entre autres le calcul de la Section Efficace Radar (aussi appelée Surface Equivalente Radar), section I.4.2), notée $a(\theta)$, est obtenue à l'aide de (I.3.27).

$$a(\theta) = \int_{\Gamma_a} e^{i\omega\vec{x}\vec{e}_\theta}(-i\omega\vec{\nu}\,\vec{e}_\theta u + \partial_\nu u) \tag{I.3.27}$$

On prend $\Gamma_a = \Gamma_{int}$ la frontière de l'objet diffractant, et on utilise la majoration (I.3.36) du lemme 9. On majore alors $|a(\theta) - a_h(\theta)|$ par $||x - x_h||_{H^{-s}(\Gamma)}$ grâce à l'inégalité de Cauchy-Schwarz.

I.3.3.3 Estimations d'énergie sur la frontière.

I.3.3.3.1 Estimation sur la frontière de $x - x_h$.

Nous étudions ici la norme du résidu $||x-x_h||_{L^2(\Gamma)}$ par rapport à l'erreur d'interpolation $||(I-P_h)x||$. Ceci est l'équivalent du lemme de Céa.

Lemme 9 *Soit Q (l'opérateur de bord du problème de Helmholtz (I.0.1)) constant tel que $|Q| \leq \delta < 1$. Soit $x \in V$ la solution de (I.1.27) et $x_h \in V_h$ la solution de (I.2.1). Soit P_h le projecteur orthogonal sur l'espace V_h. Nous avons*

$$\boxed{||x - x_h||_{L^2(\Gamma)} \leq \frac{2}{\sqrt{1-\delta^2}}||(I - P_h)x||_V \ .} \tag{I.3.28}$$

Preuve. Posons $\varepsilon_h = (x-x_h)$ et $e_h = u-u_h$. D'après la définition de A, l'inégalité de Cauchy-Schwarz et le fait que F est une isométrie, nous avons

$$\begin{aligned} ((I-A)\varepsilon_h, \varepsilon_h)_V &= ||\varepsilon_h||^2 - (\Pi\varepsilon_h, F\varepsilon_h)_V \\ &\geq ||\varepsilon_h||^2 - ||\Pi\varepsilon_h|| \; ||F\varepsilon_h|| \\ &\geq ||\varepsilon_h||^2 \left(1 - \frac{||\Pi\varepsilon_h||}{||\varepsilon_h||}\right) \ . \end{aligned} \tag{I.3.29}$$

Par définition de Π

$$\begin{aligned} ||\Pi\varepsilon_h||^2 &= + \sum_k \int_{\Gamma_k} |Q|^2 \; |(-\partial_{\nu_k} + i\omega)e_h|^2 + \sum_{kj} \int_{\Sigma_{kj}} |(-\partial_{\nu_k} + i\omega)e_h|^2 \\ &= + \sum_k \int_{\Gamma_k} (|Q|^2 - 1) \; |(-\partial_{\nu_k} + i\omega)e_h|^2 + \sum_k \int_{\partial\Omega_k} |(-\partial_{\nu_k} + i\omega)e_h|^2 \ , \end{aligned} \tag{I.3.30}$$

soit, en définissant $||\varepsilon_h||_\Gamma^2 = \int_\Gamma |\varepsilon_h|^2$, nous obtenons $||\Pi\varepsilon_h||^2 \leq ||\varepsilon_h||^2 - (1-|\delta|^2)||\varepsilon_h||_\Gamma^2$. D'où,

$$||\Pi\varepsilon_h|| \leq ||\varepsilon_h|| \left(1 - \frac{1-|\delta|^2}{2} \frac{||\varepsilon_h||_\Gamma^2}{||\varepsilon_h||^2}\right) \ . \tag{I.3.31}$$

Des inégalités (I.3.31) et (I.3.29), nous avons

$$|((I-A)\varepsilon_h, \varepsilon_h)_V| \geq ||\varepsilon_h||^2 \left[1 - (1 - \frac{1-|\delta|^2}{2} \frac{||\varepsilon_h||_\Gamma^2}{||\varepsilon_h||^2})\right] \ , \tag{I.3.32}$$

qui donne la majoration sur $||\varepsilon_h||_\Gamma$

$$|((I-A)\varepsilon_h, \varepsilon_h)_V| \geq \frac{1-|\delta|^2}{2} ||\varepsilon_h||_\Gamma^2 \ . \tag{I.3.33}$$

D'après le lemme 8 et l'inégalité de Cauchy-Schwarz, nous pouvons écrire

$$|((I-A)\varepsilon_h, \varepsilon_h)|_V = |((I-A)\varepsilon_h, (I-P_h)x)|_V \leq 2||(I-P_h)x||^2 \ . \tag{I.3.34}$$

Ainsi, de (I.3.33) et (I.3.34), on a :

$$||(I-P_h)x|| \geq \frac{\sqrt{1-\delta^2}}{2} ||\varepsilon_h||_\Gamma \ . \tag{I.3.35}$$

□

I.3.3.3.2 Estimation sur la frontière de $u - u_h$.

Lemme 10 *Soit Q (l'opérateur de bord du problème de Helmholtz (I.0.1)) constant tel que $|Q| \leq \delta < 1$. Soit $x \in V$ la solution de (I.1.27) et $x_h \in V_h$ la solution de (I.2.1), u la solution de (I.2.22) et u_h défini par (I.2.23). Alors, pour tout réel positif s*

$$\boxed{||u - u_h||_{H^{-s}(\Gamma)} \leq \frac{1+\delta}{2\omega}||x - x_h||_{H^{-s}(\Gamma)}\ .} \tag{I.3.36}$$

Preuve. Ceci est trivial d'après les relations (I.2.23) et (I.2.22), section I.2.1.5 :

$$\begin{cases} u = \dfrac{1}{2i\omega}[(I + \Pi)x + g] & \text{sur } \Gamma_k \\ u_h = \dfrac{1}{2i\omega}[(I + \Pi)x_h + g] & \text{sur } \Gamma \end{cases}$$

et l'hypothèse $|Q| \leq \delta$. □

Remarque 24 Nous observons numériquement que la majoration (I.3.36) du lemme 10 qui utilise l'approximation (I.2.27) pour u_h n'est pas optimale dans le cadre d'un problème homogène (la figure I.3.4 montre un gain d'environ un demi sur l'ordre de l'erreur sur u à la frontière Γ par rapport à l'ordre de l'erreur sur x de la figure I.3.5).

Nous pensons qu'une analyse plus fine devra étudier directement l'opérateur de relèvement linéaire E (I.1.20) et obtenir une majoration directe de l'erreur $u - u_h$ par rapport à l'erreur d'interpolation $(I - P_h^{'})u$ où $P_h^{'}$ sera l'opérateur de projection orthogonale sur les fonctions de base $e_{k,l}$ permettant d'approcher u_h par $(u_h)_{\Omega_k} = \sum_{l=1}^{p} x_{k,l}e_{k,l}$.

Corollaire 2 *Sous les hypothèses du lemme (9), il vient de (I.3.36) et (I.3.28) la majoration suivante en norme $L^2(\Gamma)$ de l'erreur sur u par rapport à l'erreur d'interpolation $||(I - P_h)x||_V$:*

$$||u - u_h||_{L^2(\Gamma)} \leq \frac{1}{\omega}\sqrt{\frac{1+\delta}{1-\delta}}||(I - P_h)x||_V\ . \tag{I.3.37}$$

Corollaire 3 *Sous les hypothèses du théorème (6), il vient de (I.3.36) et (I.3.22) la majoration suivante en norme de Sobolev négatif $H^{-s}(\Gamma)$ pour tout $s > 1/2$ de l'erreur sur u :*

$$||u - u_h||_{H^{-s}(\Gamma)} \leq \frac{1+\delta}{\omega}||(I - P_h)x|| \sup_{\psi \in H^s(\Gamma)} \frac{||(I - P_h)y(\psi)||}{||\psi||_{H^s(\Gamma)}}\ . \tag{I.3.38}$$

I.3.3.4 Etude de l'erreur d'interpolation.

Nous avons : $\forall x_a \in V_h$, $||(I - P_h)x|| \leq ||x - x_a||_V$. La clef de la majoration est de trouver une fonction particulière x_a pour laquelle nous avons une "bonne" estimation.

Théorème 7 *Soit u une solution d'un problème de Helmholtz homogène. Nous supposons que u est de classe C^{n+1} avec $n \geq 1$. Soit $x \in V$ vérifiant :*

$$x_{|\partial\Omega_k} = (-\partial\nu_k + i\omega)u_{|\partial\Omega_k}\ .$$

Nous supposons que le maillage $(\Omega_k)_{k=1\ldots K}$ vérifie les hypothèses d'uniforme régularité (H1, H2 et H3 de la section I.2.1.2). L'espace d'approximation V_h est construit de $p = 2n + 1$ fonctions z_{kl} par élément Ω_k telles que $z_{kl} = (-\partial_{\nu_k} + i\omega)e_{kl}$ et $(e_{kl})_{l=1\ldots p}$ soit une famille libre d'ondes planes. Pour les besoins de simplicité de la preuve sur le plan technique, nous supposons que les directions de ces ondes planes sont fixées une fois pour toutes (et par exemple toutes égales d'un élément à un autre). Alors, en notant $[\alpha]$ la partie entière de α, on a

$$\begin{cases} \exists C > 0 \text{ dépendant de } n \text{ et des données du problème (I.0.1)} \\ ||(I - P_h)x||_V \leq Ch^{n-1/2}||u||_{C^{n+1}(\Omega)},\ \text{avec } n = [\dfrac{p-1}{2}]\ . \end{cases} \tag{I.3.39}$$

Remarque 25 La constante C dans (I.3.39) dépend de ω. Dans la preuve du théorème 7, nous verrons que l'on peut faire apparaître que $C = \omega^{n+1} \times C^{'}$, mais $C^{'}$ dépend encore de ω de façon non triviale.

Preuve.

a) Etape 1. Supposons qu'il existe une fonction u_a dans l'espace vectoriel engendré par les ondes planes e_{kl} telle que, sur chaque élément Ω_k, u_a approche u à l'ordre n. Nous omettons l'indice k de l'élément Ω_k. Plus précisément, si l'on note $\vec{x} = (x_1, x_2)$ un point de Ω_k dans un système de coordonnées centré sur le barycentre de cet élément, nous supposons qu'il existe C_1 constante positive telle que

$$|u(\vec{x}) - u_a(\vec{x})| \leq C_1 h^{n+1} ||u||_{C^{n+1}(\Omega)} \tag{I.3.40}$$

et

$$|\nabla u(\vec{x}) - \nabla u_a(\vec{x}))| \leq C_1 h^{n} ||u||_{C^{n+1}(\Omega)} \tag{I.3.41}$$

où C_1 dépend de ω et Ω seulement (pas de h ou u) avec u_a vérifiant

$$u_a = \sum_{l=1}^{p} x_l^n e_l \ . \tag{I.3.42}$$

Comme u_a est supposé au moins $C^{n+1}(\Omega)$ on peut définir x_a par

$$x_a = (-\partial_\nu + i\omega) u_a \ . \tag{I.3.43}$$

Par définition de $||x - x_a||_{L^2(\partial\Omega_k)}$, on a :

$$\begin{aligned} ||x - x_a||^2_{L^2(\partial\Omega_k)} &= \int_{\partial\Omega_k} |(-\partial_{\nu_k} + i\omega)(u - u_a)|^2 \\ &\leq 2 \int_{\partial\Omega_k} \left\{ |\left(\nabla u(\vec{x}) - \nabla u_a(\vec{x})\right) .\vec{\nu}|^2 + \omega^2 |u(\vec{x}) - u_a(\vec{x})|^2 \right\} \\ &\leq 2 C_1^2 h^{2n} \left(\int_{\partial\Omega_k} (1 + \omega^2 h^2) \right) ||u||^2_{C^{n+1}(\Omega)} \end{aligned}$$

Par définition, pour h assez petit, l'intégrale sur $\partial\Omega_k$ peut être majorée comme suit ($C^{'} > 0$) :

$$\int_{\partial\Omega_k} \left(1 + \omega^2 h^2\right) \leq C^{'} h \ .$$

C'est pourquoi, en posant $C^2 = 2 C_1^2 C^{'} \omega^2$, on a

$$||x - x_a||_{L^2(\partial\Omega_k)} \leq C h^{n+1/2} ||u||_{C^{n+1}(\Omega)} \ . \tag{I.3.44}$$

Sachant que, pour un maillage régulier de taille h, le nombre total d'éléments est majoré par C/h^2, on en déduit :

$$||x - x_a||_V \leq C h^{n-1/2} ||u||_{C^{n+1}(\Omega)} \ . \tag{I.3.45}$$

La constante C dans (I.3.45) ne dépend pas de k puisque les directions des ondes planes sont fixes.

b) Etape 2. Nous cherchons la condition d'existence de u_a vérifiant la condition (I.3.40), omettant l'indice k pour simplifier les notations.

Comme u est de classe C^{n+1}, le développement de Taylor de u (en un point de Ω_k noté (x_1, x_2) dans un système de coordonnées centré sur le barycentre de l'élément considéré) à l'ordre n existe au sens suivant

$$\left| u(\vec{x}) - \sum_{m=0}^{n} \sum_{q_1+q_2=m}^{q_1\geq 0, q_2 \geq 0} B_{q_1,q_2} x_1^{q_1} x_2^{q_2} \right| \leq C h^{n+1} ||u||_{C^{n+1}(\Omega)} \ . \tag{I.3.46}$$

Les fonctions de base e_l sont C^∞ dans l'intérieur de Ω_k (et même analytiques). Soit e_l^n le développement de Taylor tronqué à l'ordre n de e_l, i.e. $|e_l - e_l^n| \leq C_{l,n} h^{n+1} ||e_l||_{C^{n+1}(\Omega)}$. Alors

$$e_l^n(\vec{x}) = \sum_{m=0}^{n} \sum_{q_1+q_2=m}^{q_1\geq 0, q_2 \geq 0} M^l_{q_1,q_2} x_1^{q_1} x_2^{q_2} \ . \tag{I.3.47}$$

la dérivée comme étant $lim_{h\to 0}\dfrac{e(x+h)-e(x)}{h}$) De plus, nous avons (puisque e_l est analytique),

$$|\Delta e_l(\vec{x}) - \Delta e_l^n(\vec{x})| \leq C_{l,n} h^{n-1} ||e_l||_{C^{n+1}(\Omega)} \ . \tag{I.3.48}$$

L'existence de u_a tel que (I.3.40) est vérifié est équivalente à l'existence de p coefficients complexes x_l^n tels que

$$\begin{cases} B_{q_1,q_2} - \displaystyle\sum_{l=1}^{p} x_l^n M_{q_1,q_2}^l = 0 \\ \forall (q_1,q_2) \in \mathbb{N}^2, \ q_1 \geq 0, \ q_2 \geq 0, \ 0 \leq q_2 + q_1 \leq n \ . \end{cases} \tag{I.3.49}$$

Le système d'équations linéaires (I.3.49) peut être synthétisé sous la forme matricielle

$$\begin{cases} \text{Trouver } X^n \in \mathbb{C}^{2n+1} \text{ tel que} \\ M_n X^n = B_n \end{cases} \tag{I.3.50}$$

La matrice M_n a $p = 2n+1$ colonnes et $(n+2)(n+1)/2$ lignes. Les termes de (I.3.50) sont de la forme

$$M_n = \begin{bmatrix} M_{0,0}^1 & M_{0,0}^2 & \dots & M_{0,0}^p \\ M_{1,0}^1 & M_{1,0}^2 & \dots & M_{1,0}^p \\ M_{0,1}^1 & M_{0,1}^2 & \dots & M_{0,1}^p \\ M_{2,0}^1 & M_{2,0}^2 & \dots & M_{2,0}^p \\ M_{1,1}^1 & M_{1,1}^2 & \dots & M_{1,1}^p \\ M_{0,2}^1 & M_{0,2}^2 & \dots & M_{0,2}^p \\ \vdots & \vdots & \vdots & \vdots \\ M_{n,0}^1 & M_{n,0}^2 & \dots & M_{n,0}^p \\ \vdots & \vdots & \vdots & \vdots \\ M_{0,n}^1 & M_{0,n}^2 & \dots & M_{0,n}^p \end{bmatrix}, \quad B_n = \begin{bmatrix} B_{0,0} \\ B_{1,0} \\ B_{0,1} \\ B_{2,0} \\ B_{1,1} \\ B_{0,2} \\ \vdots \\ B_{n,0} \\ \vdots \\ B_{0,n} \end{bmatrix} \quad \text{et } X_n = \begin{bmatrix} x_1^n \\ x_2^n \\ x_3^n \\ x_4^n \\ x_5^n \\ x_6^n \\ \vdots \\ \vdots \\ \vdots \\ x_p^n \end{bmatrix}$$

La matrice M_n dépend uniquement des directions des ondes planes. Le second membre B_n dépend uniquement des données du problème (I.0.1). Nous montrons que ce système admet une solution.

i) Montrons que l'image de M_n est incluse dans un sous-espace K de dimension $2n+1$. Montrons aussi que le terme de droite B_n dans le système linéaire (I.3.50) est un élément de K. Ceci est résumé par

$$\begin{cases} \text{Im}(M_n) \subset K \\ B_n \in K \\ dim(K) = 2n+1 \ . \end{cases}$$

Calculons Δe_l^n, en développant dans un premier temps $\dfrac{\partial^2 e_l^n}{\partial^2 x_1}$:

$$\begin{aligned} \frac{\partial^2 e_l^n}{\partial^2 x_1}(\vec{x}) &= \frac{\partial^2}{\partial^2 x_1} \sum_{m=0}^{n} \sum_{q_1+q_2=m}^{q_1\geq 0, q_2 \geq 0} M_{q_1,q_2}^l x_1^{q_1} x_2^{q_2} \\ &= \sum_{m=0}^{n} \sum_{q_1+q_2=m}^{q_1\geq 0, q_2 \geq 0} M_{q_1,q_2}^l q_1(q_1-1) x_1^{q_1-2} x_2^{q_2} \\ &= \sum_{m=0}^{n} \sum_{q_1+q_2=m}^{q_1\geq 0, q_2 \geq 0} M_{q_1+2,q_2}^l (q_1+1)(q_1+2) x_1^{q_1} x_2^{q_2} \ . \end{aligned} \tag{I.3.51}$$

De $-(\Delta + \omega^2)e_l = 0$, relation (I.3.48) et en remplaçant $\Delta e_l = \dfrac{\partial^2 e_l}{\partial^2 x_1} + \dfrac{\partial^2 e_l}{\partial^2 x_2}$ par son développement (obtenu à l'aide de (I.3.51) puis en substituant x_1 par x_2 dans (I.3.51)), on obtient les relations (I.3.52) entre les coefficients de Taylor de e_l.

$$\text{(I.3.52)} \qquad \begin{cases} \forall l \in \{1..p\} \\ \forall (q_1, q_2) \in \mathbb{N}^2,\ q_1 \geq 0,\ q_2 \geq 0,\ 0 \leq q_1 + q_2 \leq n-2 \\ (q_1+1)(q_1+2)M^l_{q_1+2,q_2} + (q_2+1)(q_2+2)M^l_{q_1,q_2+2} = \omega^2 M^l_{q_1,q_2}\ . \end{cases}$$

Comme u aussi vérifie l'équation de Helmholtz homogène, nous avons :

$$\text{(I.3.53)} \qquad \begin{cases} \forall (q_1, q_2) \in \mathbb{N}^2,\ q_1 \geq 0,\ q_2 \geq 0,\ 0 \leq q_1 + q_2 \leq n-2 \\ (q_1+1)(q_1+2)B_{q_1+2,q_2} + (q_2+1)(q_2+2)B_{q_1,q_2+2} = \omega^2 B_{q_1,q_2}\ . \end{cases}$$

Soit K l'espace vectoriel défini par

$$\text{(I.3.54)} \qquad \begin{aligned} K = \{(B_{q_1,q_2}) \in \mathbb{C}^{\frac{n(n+1)}{2}}, q_1 \geq 0,\ q_2 \geq 0,\ q_1+q_2 \leq n \text{ avec} \\ \forall (q_1, q_2) \in \mathbb{N}^2,\ q_1 \geq 0,\ q_2 \geq 0,\ 0 \leq q_1 + q_2 \leq n-2 \\ (q_1+1)(q_1+2)B_{q_1+2,q_2} + (q_2+1)(q_2+2)B_{q_1,q_2+2} = \omega^2 B_{q_1,q_2}\}\ . \end{aligned}$$

Il est clair que $B_n \in K$ et $Im(M_n) \subset K$. Les vecteurs colonnes de M_n appartiennent à K, donc l'espace vectoriel engendré par les colonnes de M_n est inclus dans K. Montrons que $dim(K) \leq 2n+1$. L'espace vectoriel K est défini par $\frac{n(n-1)}{2}$ relations que l'on écrit sous la forme :

$$N(B_{q_1,q_2}) = 0$$

avec N une $(\frac{n(n-1)}{2}, \frac{n(n+1)}{2})$ matrice dont les $\frac{n(n-1)}{2}$ premières colonnes sont

$$\begin{bmatrix} -\omega^2 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & 0 & 0 & \dots \\ 0 & -\omega^2 & 0 & 0 & 0 & 0 & 6 & 0 & 6 & 0 & \dots \\ 0 & 0 & -\omega^2 & 0 & 0 & 0 & 0 & 6 & 0 & 6 & \dots \\ 0 & 0 & 0 & -\omega^2 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & -\omega^2 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & -\omega^2 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & -\omega^2 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\omega^2 & 0 & 0 & \dots \\ \vdots & & & & & & & & \ddots & & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\omega^2 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\omega^2 \end{bmatrix}$$

formant une matrice triangulaire supérieure qui est évidemment inversible. Les $\frac{n(n-1)}{2}$ relations qui définissent K sont indépendantes. Ceci signifie que la dimension de K est égale à $\frac{n(n+1)}{2} - \frac{n(n-1)}{2}$, qui est $2n+1$.

ii) Montrons que M_n a un rang supérieur à $2n+1$.

La partie précédente de la preuve utilisait seulement le fait que les fonctions e_l sont des solutions exactes de l'équation de Helmholtz homogène. Cette partie utilise la forme exacte des fonctions, c'est-à-dire le fait que ce sont des ondes planes, et nécessite, pour des raisons techniques, que ces fonctions soient fixées dès le départ. Nous extrayons de la matrice M_n ou d'une transformée linéaire de M_n une matrice carrée inversible de rang $2n+1$.

Notons G le barycentre de l'élément Ω_k. La matrice M_n est définie par les coefficients $M^l_{q,s} = \frac{(i\omega)^{q+s}}{q!s!} \frac{\partial^{q+s}}{\partial^q_{x_1} \partial^s_{x_2}} e_l(G)$. Soit S_n la matrice définie par $S_{n+q+1,l} = (\partial_{x_1} + i\partial_{x_2})^q e_l(G)$ pour $q = 0$ à $q = n$ et $S_{n-q+1,l} = (\partial_{x_1} - i\partial_{x_2})^q e_l(G)$ pour $q = 0$ à $q = n$. Cette matrice est une transformée linéaire de M_n

puisque les opérateurs de dérivation mixtes utilisés pour la définir sont des combinaisons linéaires des opérateurs de dérivation qui donnent la matrice M_n. Plus précisément,

$$\text{(I.3.55)} \qquad 0 \leq q \leq n \Rightarrow \begin{cases} S_{n-q+1,l} = \dfrac{q!}{(i\omega)^q} \displaystyle\sum_{s=0}^{q} (-i)^s M^l_{s-q,s} \\ S_{n+q+1,l} = \dfrac{q!}{(i\omega)^q} \displaystyle\sum_{s=0}^{q} i^s M^l_{s-q,s} \end{cases}$$

Les relations ci-dessus montrent qu'il existe une matrice P_n de taille $(\dfrac{n(n+1)}{2} \times 2n+1)$ qui transforme M_n en S_n : $S_n = P_n\ M_n$. La matrice transformée P_n dépend de ω et n seulement. Elle ne dépend pas de u. Les coefficients de P_n sont $\dfrac{q!}{(i\omega)^q} i^s$ et $\dfrac{q!}{(i\omega)^q}(-i)^s$. Montrons que M_n est de rang supérieur à $2n+1$.

La matrice S_n est une matrice carrée de taille $2n+1$. Pour e_l une onde plane, soit

$$\text{(I.3.56)} \qquad e_l = e^{i\omega(u_l x_1 + v_l x_2)},$$

nous avons

$$\text{(I.3.57)} \qquad \begin{cases} (\partial_x + i\partial_y)^q e_l = (u_l + i v_l)^q\, e_l = z_l^q \\ (\partial_x - i\partial_y)^q e_l = (u_l - i v_l)^q\, e_l = z_l^{-q} \end{cases}$$

considérant que $(u_l - iv_l) = 1/(u_l + iv_l)$ puisque $u_l^2 + v_l^2 = 1$ (e_l est une solution du problème de Helmholtz homogène (I.0.1) et que $e_l = 1$ au barycentre. La matrice construite S_n est ainsi

$$\text{(I.3.58)} \qquad S_n = \begin{bmatrix} z_1^{-n} & \dots & z_p^{-n} \\ \vdots & \dots & \vdots \\ z_1^{-1} & \dots & z_p^{-1} \\ 1 & \dots & 1 \\ z_1 & \dots & z_p \\ \vdots & \dots & \vdots \\ z_1^n & \dots & z_p^n \end{bmatrix} .$$

La matrice S_q^n a un déterminant de Vandermonde. Il vaut :

$$\text{(I.3.59)} \qquad \prod_{i=1}^{n} z_i^{-n} \prod_{i<j} (z_i - z_j) .$$

Le produit ci-dessus (I.3.59) est non nul si et seulement si $\forall (i,j) z_i \neq z_j$. Pour récapituler, S_q^n est une matrice carrée de dimensions $2n+1$ lignes, $p = 2n+1$ colonnes et est inversible. Nous avons donc : $rang(M_n) \geq rang(S_q^n) = 2n+1$.

iii) Récapitulons les points précédents i) et ii). Le point i) montre l'inclusion de l'image de M_n dans K, un espace vectoriel de dimension $2n+1$. Le point ii) nous apprend que la dimension de $\mathrm{Im}(M_n)$ est au moins $2n+1$. Donc $Im(M_n) = K$. De plus, le second membre B_n est un élément de K. Nous pouvons résumer :

$$\left.\begin{matrix} K = \mathrm{Im}\ (M_n) \\ B_n \in K \end{matrix}\right\} \Rightarrow B_n \in \mathrm{Im}\ (M_n) .$$

Ceci prouve que le système (I.3.50) a une unique solution et que u_a (défini par (I.3.42)) existe. De l'approximation de Taylor de $u - u_a$, nous avons

$$\text{(I.3.60)} \qquad \begin{aligned} |u - u_a| \leq & Ch^{n+1} \sup ||u - u_a||_{C^{n+1}(\Omega)} \\ \leq & Ch^{n+1} \left(||u||_{C^{n+1}(\Omega)} + ||u_a||_{C^{n+1}(\Omega)} \right) , \end{aligned}$$

et d'après (I.3.42)

$$\text{(I.3.61)} \qquad ||u_a||_{C^{n+1}(\Omega)} \leq \sum_{l=1}^{p} |x_l^n| ||e_l||_{C^{n+1}(\Omega)} .$$

De $S_n X_n = P_n M_n X_n = P_n B_n$ et S_n est une matrice donnée, inversible et de dimension finie, on obtient

$$X^n = (S_n)^{-1} P_n B_n \ .$$

Puisque M_n dépend seulement des fonctions de base données et a donc sa norme bornée supérieurement uniformément par $\sup_{l=1}^{p} ||e_l||_{C^{n+1}(\Omega)}$, et puisque les coefficients B_n sont bornés supérieurement par la $C^{n+1}(\Omega)$ norme de u, il existe une constante positive C_2 telle que pour $l = 1$ à $p = 2n+1$ nous avons

(I.3.62) $$|x_l^n| \leq C_2 ||u||_{C^{n+1}(\Omega)} \ .$$

De (I.3.61) et (I.3.62) dans (I.3.60), nous avons en posant $C_1 = C(pC_2 + 1)$

(I.3.63) $$|u - u_a| \leq C_1 h^{n+1} ||u||_{C^{n+1}(\Omega)} \ ,$$

Nous vérifions donc l'hypothèse (I.3.40) de l'étape 1. Vérifions maintenant l'hypothèse (I.3.41). De l'approximation de Taylor de $u - u_a$, nous avons :

(I.3.64) $$|\nabla u(\vec{x}) - \nabla u_a(\vec{x}))| \leq C h^n ||u - u_a||_{C^{n+1}(\Omega)} + |\nabla u_n(\vec{x}) - \sum_{l=1}^{2n+1} x_l^n \nabla e_l^n(\vec{x}))|$$

où u_n est le polynôme de Taylor à l'ordre n de u comme dans (I.3.46). Par construction, on a donc montré que l'on a

$$u_n(\vec{x}) - \sum_{l=1}^{2n+1} x_l^n e_l^n(\vec{x})) = 0 \text{ et } \nabla u_n(\vec{x}) - \sum_{l=1}^{2n+1} x_l^n \nabla e_l^n(\vec{x})) = 0 \ .$$

Nous avons donc, comme dans (I.3.63)

(I.3.65) $$|\nabla u(\vec{x}) - \nabla u_a(\vec{x}))| \leq C_3 h^n ||u||_{C^{n+1}(\Omega)}$$

qui est l'hypothèse (I.3.41) de l'étape 1. □

Remarque 26 On peut calculer exactement le déterminant de M_n pour 3 fonctions de base. On remarquera que l'on obtient le déterminant de S_q^n et qu'il est maximal pour des fonctions de base équiréparties (annexe III.E.1).

Remarque 27 La preuve du théorème (7) exige l'utilisation d'ondes planes pour extraire un sous-système libre de taille $(2n+1)$ de M_n. Il devrait être possible de généraliser ce résultat à d'autres types de fonctions solutions du problème de Helmholtz homogène sur une maille, à condition qu'elles soient assez régulières, et sous d'autres conditions qu'il faudra exhiber.

Remarque 28 Ce résultat pourrait être établi dans des espaces de Sobolev en estimant directement le reste intégral du développement de Taylor. Nous pensons qu'une estimation directe dans les espaces de Sobolev devrait mener à un résultat plus optimal.

I.3.3.5 Bilan : majorations de l'ordre de convergence.

Voici nos principaux résultats concernant l'ordre de convergence. Ces résultats sont de simples conséquences du théorème 7.

Remarque 29 Dans la démonstration du théorème 7, le fait que u est solution d'un problème homogène est essentiel. Cependant, dans le cas $n = 1$, le résultat est encore vrai pour u solution d'un problème de Helmholtz non homogène. En effet, on a $p = 2n+1 = 3 = (n+2)(n+1)/2$ fonctions de base. L'espace vectoriel K (I.3.54) est de dimension 3 car $\frac{n(n-1)}{2} = 0$. Pour tout $p \geq 3$, x défini par (I.0.3) et u la solution du problème de Helmholtz homogène ou non homogène (I.0.1), on a :

(I.3.66) $$||(I - P_h)x||_V \leq C h^{1/2} ||u||_{C^2(\Omega)} \ .$$

D'autre part, d'après le théorème 7 et le lemme 9, on a le

Corollaire 4 *Supposons que les fonctions de base vérifient les hypothèses du théorème 7. Soit u une solution de (I.0.1) homogène ($f = 0$), x la solution de (I.1.27) et $x_h \in V_h$ la solution de (I.2.1) pour $p \geq 3$ fonctions de base par élément. Soit $[\alpha]$ la partie entière de α. Nous supposons que u est de classe $C^{[(p+1)/2]}(\Omega)$, alors :*

$$\text{(I.3.67)} \qquad \begin{cases} ||x - x_h||_{L^2(\Gamma)} \leq Ch^{[(p-1)/2]-1/2}||u||_{C^{[(p+1)/2]}(\Omega)} \\ ||u - u_h||_{L^2(\Gamma)} \leq C^{'} h^{[(p-1)/2]-1/2}||u||_{C^{[(p+1)/2]}(\Omega)} \end{cases} ,$$

Par exemple $p = 3$ ou $p = 4$ donne $h^{1/2}$.

Ces estimations ne sont pas optimales puisque pour $p = 3$ nous avons observé numériquement $h^{3/2}$. Pour u approché par (I.2.27), on a même obtenu une loi en h^2.

Corollaire 5 *Supposons que les fonctions de base vérifient les hypothèses du théorème 7. Soit u une solution de (I.0.1) non homogène (f n'est pas nécessairement identiquement nulle sur $L^2(\Omega)$), x défini par (I.1.27) et $x_h \in V_h$ la solution de (I.2.1) avec $p \geq 3$ fonctions de base par élément. Soit $[\alpha]$ la partie entière de α. Nous supposons que u est de classe $C^2(\Omega)$, alors :*

$$\text{(I.3.68)} \qquad \forall s > [(p-1)/2] - 1/2 \begin{cases} ||x - x_h||_{H^{-s}(\Gamma)} \leq Ch^{[(p-1)/2]}||u||_{C^2(\Omega)} \\ ||u - u_h||_{H^{-s}(\Gamma)} \leq C^{'} h^{[(p-1)/2]}||u||_{C^2(\Omega)} \end{cases}$$

Par exemple $p = 3$ ou $p = 4$ donne h^1. D'après la relation (I.3.36), nous obtenons la même loi de convergence sur $||u - u_h||_{H^{-s}(\Gamma)}$.

Preuve. Rappelons la majoration par dualité (I.3.22) section I.3.3.2, théorème 6, on a $\forall s > 1/2$:

$$\text{(I.3.69)} \qquad ||x - x_h||_{H^{-s}(\Gamma)} \leq 2||(I - P_h)x||_V \sup_{||\psi||_{H^s(\Gamma)}=1} ||(I - P_h)y(\psi)||_V \ .$$

Pour $p \geq 3$ et $u \in C^1(\Omega)$, on utilise la majoration (I.3.66) de la remarque 29 :

$$\text{(I.3.70)} \qquad ||(I - P_h)x||_V \leq Ch^{1/2}||u||_{C^2(\Omega)} \ .$$

De plus, w (donné par (I.3.21)), la solution du problème de Helmholtz avec $f = 0$, $Q = 0$ et $g = \psi$, est $C^{[s+1/2]}(\Omega)$: ceci est assuré par le point **i)** de la preuve du théorème 6. Donc, pour tout $s > [(p-1)/2]-1/2$, on peut appliquer la majoration (I.3.39) du théorème 7 :

$$\text{(I.3.71)} \qquad \sup_{||\psi||_{H^s(\Gamma)}=1} ||(I - P_h)y(\psi)||_V \leq Ch^{[(p-1)/2]-1/2} \ .$$

Finalement, d'après (I.3.71) et (I.3.70), on a la majoration (I.3.67). □

I.3.4 Conditionnement de la matrice de produit scalaire.

Nous définissons le conditionnement $K(D)$ de la matrice de produit scalaire D comme le rapport de la plus grande valeur propre $\lambda_{\max}(D)$ et de la plus petite $\lambda_{\min}(D)$. Le conditionnement dépend des paramètres p, le nombre de fonctions de base par élément du maillage, et h, le pas du maillage. Nous allons montrer (théorème 8) qu'il existe une constante C indépendante de p et h telle que l'on peut minorer le conditionnement par

$$\text{(I.3.72)} \qquad K(D) \geq Ch^{-q(p)}$$

où $q(p)$, que l'on appellera l'ordre du conditionnement, est une fonction uniquement de p. Nous nous sommes intéressés au conditionnement des sous-matrices blocs D_k de la matrice D du système linéaire car l'algorithme de résolution (section I.2.3) exige l'inversion de ces matrices. Rappelons que nous avons utilisé une méthode de Cholesky, sensible aux problèmes de conditionnement. Lors de l'étude numérique de l'ordre de convergence de la méthode, nous avons rencontré ce type de problème lorsque la discrétisation effectuée était trop raffinée. Nous présentons ensuite quelques résultats numériques pour lesquels l'évolution du conditionnement semble bien suivre la loi théorique de la minoration $q(p) = 2[p/2] - 2$.

I.3.4.1 Majoration théorique du conditionnement.

Dans l'annexe III.E.1, on montrera que pour $p = 3$ fonctions de base par élément, on peut majorer le conditionnement de D_k indépendamment de h. On montrera également que l'équirépartition des ondes planes maximise le déterminant de D_k indépendamment de la géométrie de Ω_k. Nous montrons dans l'annexe III.E.2 que ce résultat se généralise au problème scalaire de Helmholtz dans l'espace $\mathbb{R}^3$.

Dans le cas $p \geq 4$, nous avons de plus l'estimation suivante.

Théorème 8 *Supposons que les p fonctions de base ($p \geq 4$) vérifient les hypothèses du théorème 7. Nous avons une croissance linéaire (I.3.73) du conditionnement de D_k en fonction de p. Soit h_k le diamètre de Ω_k et $[\alpha]$ la partie entière de α. Il existe C positif tel que :*

$$\frac{\lambda_{max}}{\lambda_{min}} \geq C h_k^{-2[p/2]+2} \tag{I.3.73}$$

soit

$$q(p) \geq 2[p/2] - 2$$

Par exemple $p = 4$ ou $p = 5$ donne h_k^{-2} et $q(p) = 2$.

Preuve. Soient λ_{min} la plus petite valeur propre de D_k et λ_{max} la plus grande. Nous avons :

$$\begin{cases} \forall Y \in \mathbb{C}^p \\ \lambda_{min} \leq \dfrac{Y^\top D_k Y}{||Y||^2} \leq \lambda_{max} \ . \end{cases} \tag{I.3.74}$$

i) Estimation de la plus petite valeur propre. Rappelons que les fonctions de base z_l dérivent des ondes planes e_l par la relation $z_l = (-\partial_\nu + i\omega)e_l$. Considérons Y tel que $Y^\top = \left[X', -1\right]$ où X' est le vecteur de $\mathbb{C}^{p-1}$ qui approxime la fonction $(-\partial_\nu + i\omega)e_p$ à l'aide des $(p-1)$ fonctions $(z_1, \ldots, z_{p-1})$. Nous sommes dans les conditions d'application du théorème 7 avec $p - 1$ fonctions de base et $u = e_p$ puisque les ondes planes sont solutions du problème de Helmholtz homogène. Nous prenons $p - 1 = 2n + 1$. L'inégalité (I.3.44) du théorème 7 stipule :

$$Y^\top D_k Y = \left(||z_p - \sum_{l=1}^{p-1} x_l' z_l||_{L^2(\partial\Omega_k)} \right)^2 \leq C(h_k^{n+1/2})^2 \ .$$

Ceci implique, avec $||Y||^2 = 1 + \sum_{l=1}^{p-1} x_l'^2 \geq 1$ et (I.3.74) que :

$$\lambda_{min} \leq C h_k^{2n+1} \ .$$

ii) Estimation de la plus grande valeur propre. Considérons Y tel que $Y^\top = [0_{p-1}, -1]$. En substituant dans l'équation (I.3.74), on obtient :

$$\begin{aligned} \lambda_{max} \geq Y^\top D_k Y &= \omega^2 \sum_{i=1}^{3} L_i (1 - \vec{\nu}_i . \vec{e}_p)^2 \\ &\geq \omega^2 \left[(L_1 + L_2 + L_3) - 2.\vec{e}_p \sum_{i=1}^{3} L_i \vec{\nu}_i \right] \\ &= \omega^2 (L_1 + L_2 + L_3) \geq \omega^2 \sup(L_1, L_2, L_3) \geq \omega^2 h_k \end{aligned}$$

car $L_1\vec{\nu}_1 + L_2\vec{\nu}_2 + L_3\vec{\nu}_3 = 0$ - les normales ν_n sont des rotations d'angle $\pi/2$ de $\dfrac{\vec{x}_{n+1} - \vec{x}_n}{|\vec{x}_{n+1} - \vec{x}_n|}$ de même que $L_1\vec{\nu}_1 + L_2\vec{\nu}_2 + L_3\vec{\nu}_3$ est une rotation d'angle $\pi/2$ de $(\vec{x}_2 - \vec{x}_1) + (\vec{x}_3 - \vec{x}_2) + (\vec{x}_1 - \vec{x}_3)$ qui est le vecteur nul.

iii) Conclusion. Des deux précédents points i) et ii), nous obtenons une minoration du comportement asymptotique pour h tendant vers zéro du conditionnement (d'après les hypothèses $n = [p/2] - 1$ et $p \geq 4$) :

$$K(D) \leq C h_k^{1-(2n+1)} = C h_k^{-2[p/2]+2} \ ,$$

et donc un ordre $q(p)$ tel que

$$q(p) \geq 2[p/2] - 2 \ .$$

$\square$

Corollaire 6 *Sous les mêmes hypothèses et les mêmes notations que celles des théorèmes 8 et 7, l'ordre du conditionnement de la matrice hermitienne D du système linéaire est minoré par la fonction*

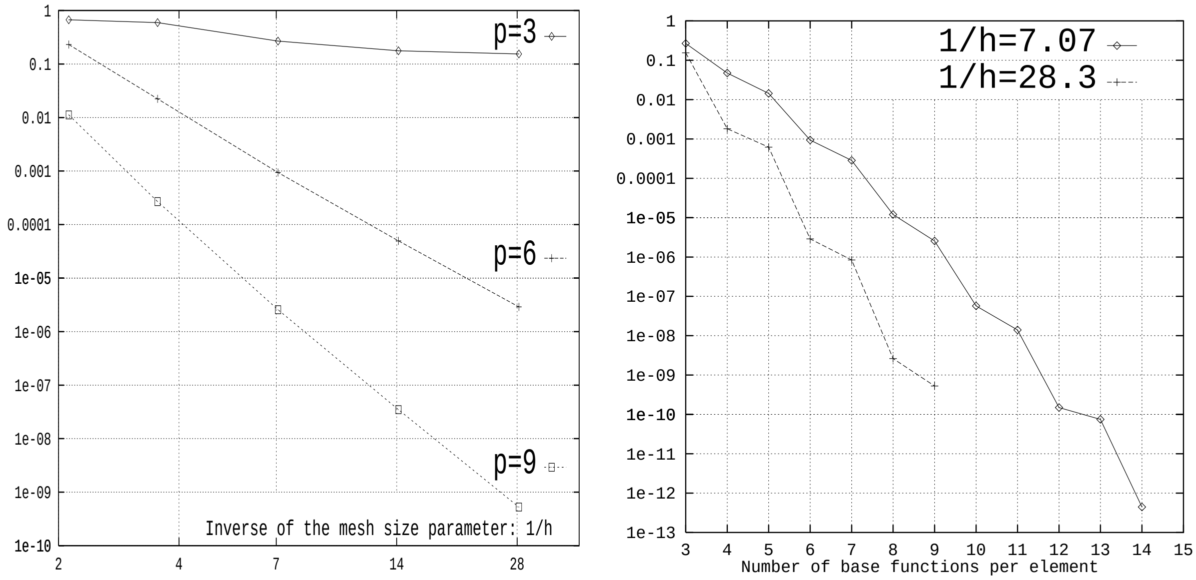
$$2[p/2] - 2$$

lorsque h tend vers zéro et pour p fonctions de base par élément, $p \geq 4$.

Preuve. Ce résultat est trivial sous les hypothèses de régularité du maillage et d'uniformité des fonctions de base. Il suffit d'effectuer des majorations uniformes sur $\lambda_{\max}(D)$ et $\lambda_{\min}(D)$. $\square$

I.3.4.2 Evolution numérique du conditionnement.

Proposons un exemple numérique (figure I.3.8) d'évolution du conditionnement (de D) en fonction du nombre de fonctions de base, ou du paramètre h du maillage, tous les autres paramètres étant fixés (tableau I.3.4) section I.3.2.4.



fonction de $1/h$ — fonction de p

FIG. I.3.8 – Inverse du conditionnement, cas limites.

On notera que

1. Les problèmes de conditionnement n'apparaissent pas pour $p = 3$ fonctions de base.
2. Le logarithme du conditionnement évolue linéairement en fonction de $2p$ et de $1/h$.
3. Pour deux fois plus de fonctions de base, le conditionnement est deux fois plus élevé pour le même maillage.
4. Pour $2p$ et $2p + 1$ fonctions de base, le conditionnement n'est que légèrement augmenté, l'ordre semble identique.

5. La loi d'évolution du conditionnement est proche de la fonction

$$q(p) = 2[p/2] - 2$$

En effet, pour $p = 6$ (courbe I.3.8 de gauche), pour $h = 1/7$ le conditionnement est de 1 000 et est multiplié par 100 pour $h \approx 1/21$. Ceci nous donne approximativement $q = 2 \times \log(3)$ soit 4.2, proche de 4. De même, pour $p = 9$ sur la même figure le conditionnement est de 10 000 pour $h = 1/4$ et est multiplié par 10 000 pour $h \approx 1/18$. Ceci signifie $q \approx 6.1$, proche de 6.

I.3.4.3 Conclusion sur le problème de l'inversion de la matrice D.

Les problèmes de conditionnement n'apparaissent que lors d'une sur-discrétisation du problème, ce qui est somme toute logique. Nous n'avons pas étudié le conditionnement global de la matrice $D - C$, mais nous avons observé numériquement une corrélation entre le conditionnement de la matrice D et le nombre d'itérations à effectuer pour résoudre le système linéaire.

Chapitre I.4

Résultats numériques.

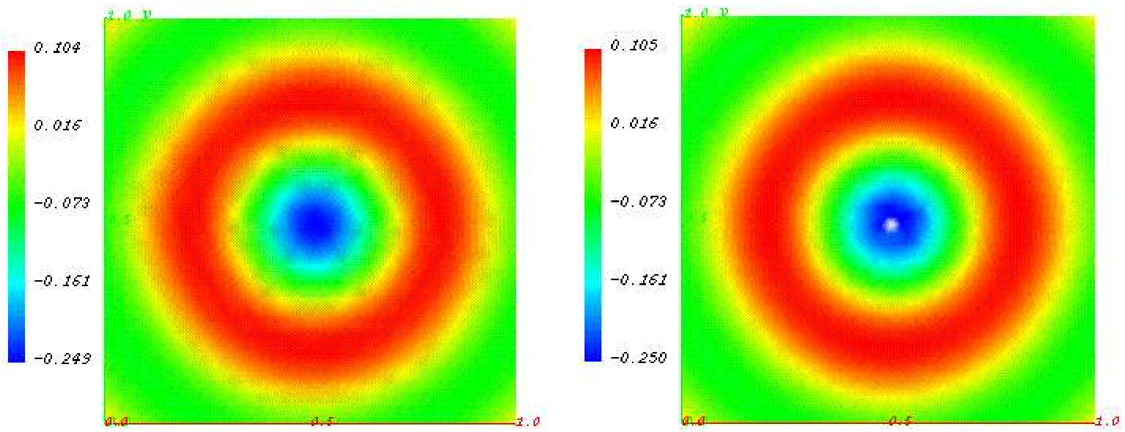
Nous étudions différents problèmes et observons les valeurs des champs ou de la SER. Nous observons ensuite un facteur important du coût informatique en temps de calcul de la méthode : la vitesse de convergence de l'algorithme itératif. Cette présentation de cas n'est pas exhaustive, nous en avons déjà proposé à Oxford (Roland Le Martret [41] et Bruno Després [28]) : résonateur de Helmholtz et ogive. Bruno Després a aussi exposé une carte des champs autour du résonateur de Helmholtz lors de sa remise du *prix Cisi 1996*.

I.4.1 Observation des valeurs des champs.

Nous proposons deux cas types simples :
- le cas d'un Dirac à l'intérieur du domaine Ω,
- les problèmes de Dirichlet et Neumann homogènes sur un carré.

I.4.1.1 Cas d'une source d'énergie ponctuelle.

Cet exemple teste le comportement de notre formulation vis-à-vis de conditions très singulières sur le second membre f.



Visualisation aux nœuds, visualisation sur une grille fine.

FIG. I.4.1 – Partie imaginaire de u.

Nous représentons, figure I.4.1, la partie imaginaire de la solution u au problème de Helmholtz avec $f = \delta$ (la fonction Dirac) et des conditions aux limites absorbantes d'ordre zéro ($Q = 0, g = 0$) sur le bord Γ (nous donnons les caractéristiques précises de ce cas dans le tableau I.4.1).

Nous donnons deux représentations graphiques de la partie imaginaire de u.

TAB. I.4.1 – Domaine connexe, Dirac

Variables	**Valeurs**
(Q, g)	$(0, 0)$
$f = \delta$ en	$(0.512443, 0.4926793)$
ω	4π
λ	0.5
$1/h$	29.899832775
(h, p)	$(\frac{\lambda}{15}, 5)$

i) Représentation de u aux nœuds du maillage utilisé pour le calcul.

On calcule u_h à l'aide de la formule (I.2.23 p. 27) et en effectuant une moyenne des valeurs de $(u_h)_{|\Sigma_{kj}}$ et $(u_h)_{|\Gamma_k}$ à chaque nœud.

ii) Représentation de u sur une grille plus fine.

On calcule u_h par la formule (I.2.27 p. 28) sur une grille plus fine que celle du maillage. Le pas de cette grille de représentation graphique est de $\lambda/30$ (2 fois plus fin que le maillage de calcul). Ce calcul est justifié dans toutes les mailles du domaine où $f_{|\Omega_k} = 0$ sauf celle dans laquelle se trouve le Dirac.

On obtient par ii) une représentation plus fine pour toutes les mailles sauf une. Le cas i) est alors préférable sur cette maille. Ceci est une bonne illustration de l'intérêt de la formule (I.2.27) et de ses limitations.

On constate que, au point où se trouve le Dirac, la partie imaginaire de la solution approchée vaut approximativement $-1/4$. Cette valeur est celle de la partie imaginaire de la fonction de Hänkel, solution du problème posé dans tout l'espace (avec les conditions de radiation de Sommerfeld).

I.4.1.2 Deux problèmes de diffraction, Dirichlet et Neumann homogènes.

Cet exemple s'intéresse à un problème physique de diffraction par réflexion sur un objet simple avec une frontière de Condition aux Limites Absorbante assez loin de l'objet pour ne pas entraîner de diffraction visible au bord. Nous avons utilisé un maillage non structuré de pas $\lambda/7.5$. Les caractéristiques du cas sont données dans le tableau I.4.2. Nous étudions des problèmes de Neumann ou Dirichlet homogènes

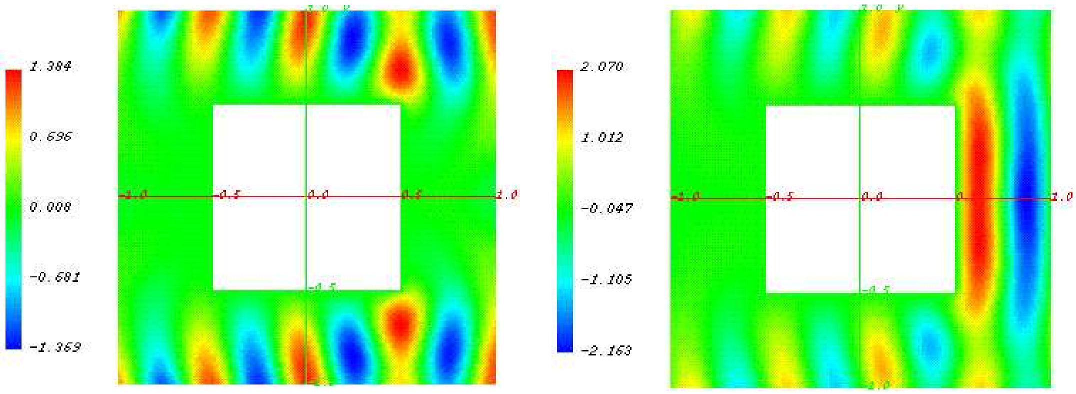
TAB. I.4.2 – Diffraction sur un carré

f	0
$\vec{v}_0$	$(-1, 0)$
ω	12.56637061436
λ	0.5
$1/h$	15.231546212
p	5
Q_{ext}	0.0

	Champ total	Champ diffracté
g_{int}	0.0	cf équation (I.3.6)
g_{ext}	cf équation (I.3.6)	0.0
	Dirichlet	Neumann
Q_{int}	-1	$+1$

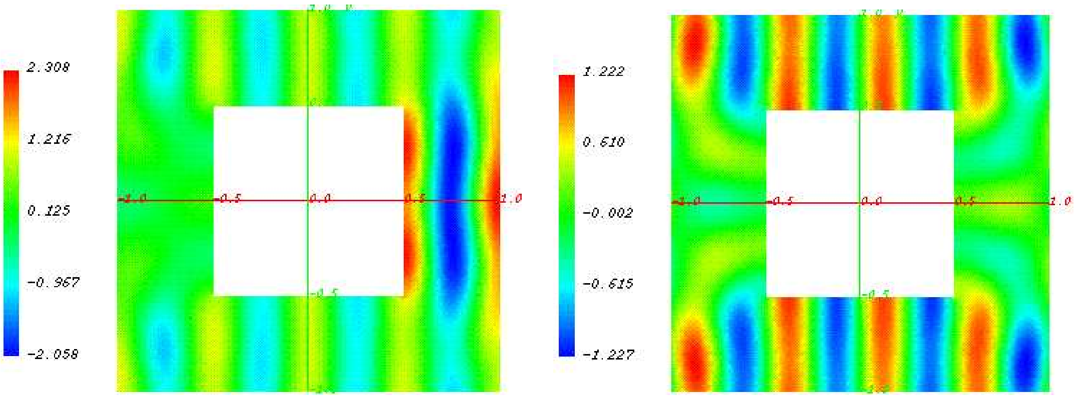
($f = 0$) en champ total ou diffracté. Les figures I.4.2, I.4.3, I.4.4, I.4.5, I.4.6 et I.4.7 sont les cartes des valeurs de u en partie réelle, imaginaire ou en module.

Nous donnons une seule représentation graphique de u à l'aide d'une projection de u_h sur une grille plus fine que celle du maillage. Le pas de la grille de représentation graphique est de $\lambda/25$ (3.3 fois plus fin que le maillage de calcul et juste suffisant pour donner une image assez nette). On calcule naturellement u_h par la formule (I.2.27) section I.2.1.3, puisque l'on traite du problème de Helmholtz homogène.



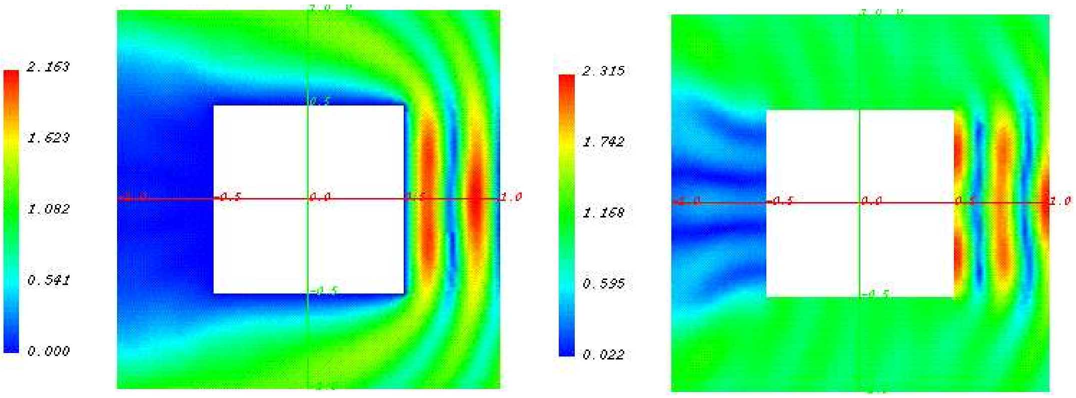
Partie réelle, partie imaginaire.

Fig. I.4.2 – Dirichlet, champ total.



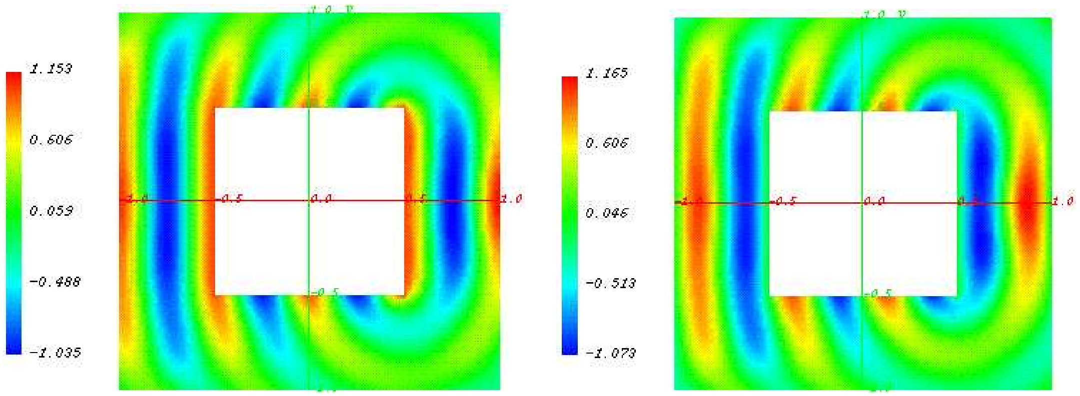
Partie réelle, partie imaginaire.

Fig. I.4.3 – Neumann, champ total.



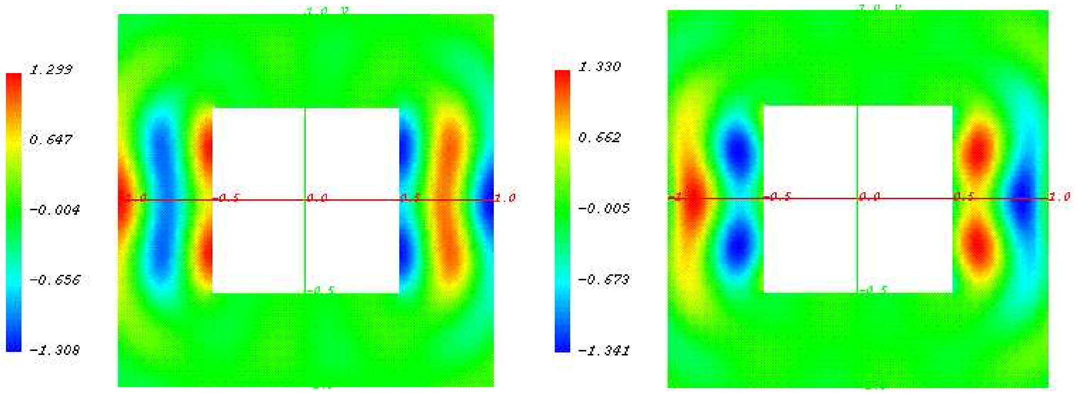
Dirichlet, Neumann.

Fig. I.4.4 – Module, champ total.



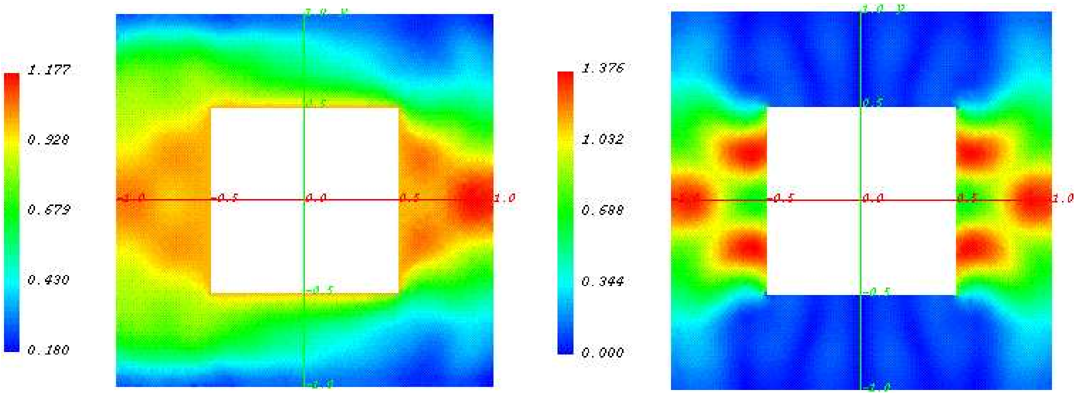
Partie réelle, partie imaginaire.

FIG. I.4.5 – Dirichlet, champ diffracté.



Partie réelle, partie imaginaire.

FIG. I.4.6 – Neumann, champ diffracté.



Dirichlet, Neumann.

FIG. I.4.7 – Module, champ diffracté.

I.4.2 Application à des problèmes de *scattering*.

Le *scattering* est le terme technique anglais de la diffraction d'une onde plane sur un objet placé dans un domaine sans source d'énergie, ce qui, pour le problème de Helmholtz bidimensionnel, se modélise par l'équation

$$\begin{cases} (-\Delta - \omega^2)u = 0 & \text{dans } \Omega \\ (\partial_\nu + i\omega)u = 0 & \text{sur } \Gamma_{ext} \\ u = -u_{inc} & \text{sur } \Gamma_{int}\ , \end{cases} \tag{I.4.1}$$

où $\Gamma = \Gamma_{int} \cup \Gamma_{ext}$, les frontières intérieure et extérieure de Ω, la frontière intérieure Γ_{int} étant la frontière de l'objet diffractant étudié (comme, par exemple, c'est le cas figures I.3.2 p. 38 b) ou I.4.8). La notation u_{inc} porte sur l'onde incidente.

Remarque 30 Le problème (I.4.1) est donné pour $|Q| = 1$ sur Γ_{int}. Ce cas n'est couvert par la théorie présentée que si Γ_{ext} est de mesure non nulle.

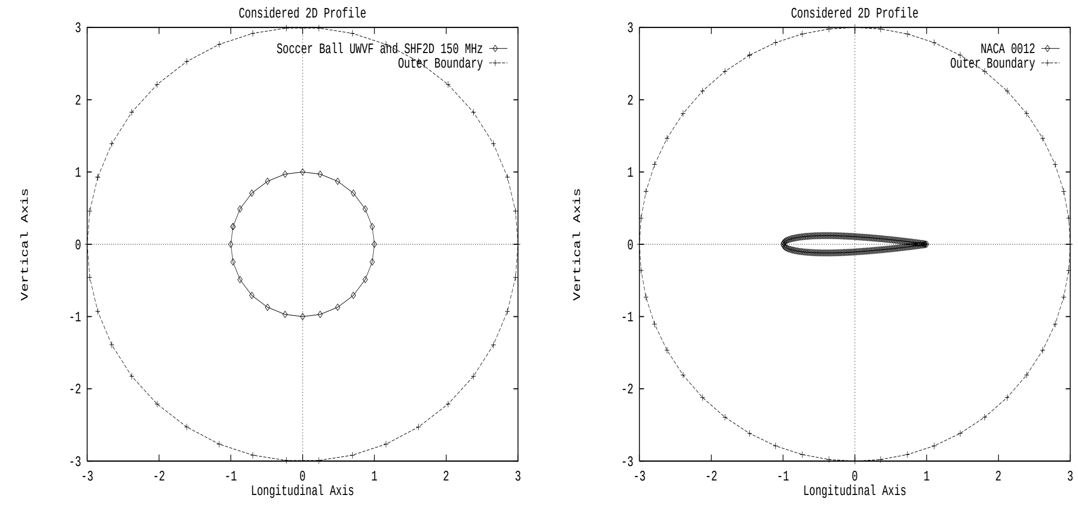
I.4.2.1 Présentation des tests.

Nous considérons ici deux problèmes de *scattering* : un disque très grossièrement maillé en 24 segments, que nous avons appelé "ballon de football", et un profil NACA, le NACA 0012, profil symétrique dont la génératrice est donnée par l'équation

$$\begin{cases} y = \dfrac{t}{2}\left(0.2969\sqrt{x} - 0.126 * x - 0.3516 * x^2 + 0.2843 * x^3 - 0.1015 * x^4\right) \\ t = 0,12 \end{cases} \tag{I.4.2}$$

pour $x \in [0, 1]$.

Ces deux profils sont représentés figure I.4.8 avec leurs domaines de calcul Ω. Le tableau I.4.3 p. 64



Ballon de football NACA 0012

FIG. I.4.8 – Domaines avec maillage de frontière

résume les caractéristiques des deux cas tests. On note L la longueur de l'aile ou le diamètre du ballon. Nos deux objets sont tels que $L = 2$. On définit le paramètre de raffinement du maillage h par (I.3.5 p. 39).

TAB. I.4.3 – Caractéristiques des maillages

Caractéristiques		
f	0	
Paramètres de raffinement	Valeur	
p : nb. de fonctions de base par élt.	5	
NACA 0012	pas du maillage	
	L/h Zone 1	100
	L/h Zone 2	20
	L/h Zone 3	5
Ballon de football	pas du maillage	
	L/h	7.4
fréquence	150 MHz	1500 MHz
L/λ	1	10
Conditions aux Limites	Dirichlet champ diffracté, mode TM	Neumann champ diffracté, mode TE
g_{int}	cf l'équation (I.3.6)	cf l'équation (I.3.6)
Q_{int}	-1	$+1$
g_{ext}	0.0	0.0
Q_{ext}	0.0	0.0
Onde incidente	bord de fuite	bord d'attaque
$\vec{v}_0$	$(-1,0)$	$(+1,0)$

I.4.2.2 Notion de Section Efficace Radar.

Le lecteur peut se reporter à [55] pour une définition de la SER. Le calcul de la SER est pour nous un moyen de comparer notre code à un code existant et fiable, un code d'équations intégrales (IE) appelé SHF2D et développé au CEA-CEL-V par P. Bonnemason, B. Stupfel [8]. Ce code simule des Conditions aux Limites de Neumann et de Dirichlet à l'aide d'un matériau parfaitement isolant ou conducteur, c'est-à-dire d'impédance nulle ou infinie. Nous utilisons la formule (I.3.27 p. 47) dans laquelle $\vec{e}_\theta = (\cos\theta, \sin\theta)$, θ étant l'angle bistatique d'observation de l'onde incidente $\vec{v}_0$:

$$a(\theta) = \frac{1}{\sqrt{\omega}} \int_{\Gamma_{int}} e^{i\omega\vec{x}\vec{e}_\theta}(-i\omega\vec{\nu}\,\vec{e}_\theta u_h + \partial_\nu u_h)\ . \tag{I.4.3}$$

En utilisant la relation vérifiée par g (I.3.6), la définition (I.2.16) de x_h, et la condition aux limites (I.2.24) définissant u_h à l'aide de x_h, nous avons :

$$\begin{cases} (x_h)_{|\Gamma_k} = i\omega \sum_l x_{kl}(1-\vec{\nu}_k\,\vec{v}_{kl})e^{i\omega\vec{v}_{kl}\,\vec{x}} \\ (g)_{|\Gamma_k} = i\omega((1+Q_k)\vec{\nu}_k\,\vec{v}_0 + (1-Q_k))e^{i\omega\vec{v}_0\,\vec{x}} \end{cases} \qquad \begin{cases} (-\partial_{\nu_k} + i\omega)(u_h)_{|\Gamma_k} = (x_h)_{|\Gamma_k} \\ (+\partial_{\nu_k} + i\omega)(u_h)_{|\Gamma_k} = g + Q_k(x_h)_{|\Gamma_k}. \end{cases}$$

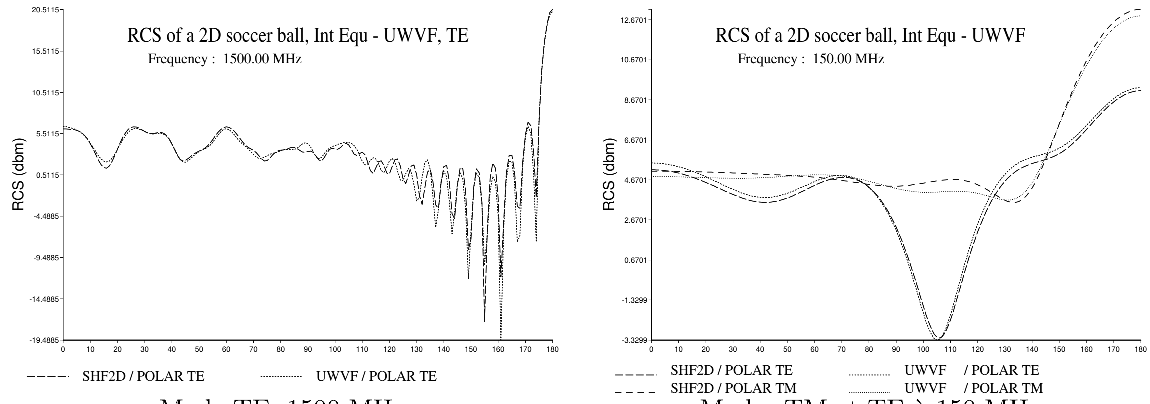
Nous calculons le terme intégral de l'équation (I.4.3) sur Γ_k :

$$\begin{cases} 2i\omega(u_h)_{|\Gamma_k} = g + (1+Q_k)(x_h)_{|\Gamma_k} \quad \text{et} \quad \partial_{\nu_k}(u_h)_{|\Gamma_k} = g + (Q_k-1)(x_h)_{|\Gamma_k} \\ 2(-i\omega\vec{\nu}_k\,\vec{e}_\theta u_h + \partial_{\nu_k}u_h) = (-\vec{\nu}_k\,\vec{e}_\theta(g + (1+Q_k)(x_h)_{|\Gamma_k}) + (g + (Q_k-1)(x_h)_{|\Gamma_k}) \end{cases} \tag{I.4.4}$$

Ceci remplacé dans l'équation (I.4.3) permet un calcul exact de l'amplitude de diffusion. Selon la polarisation, nous multiplions $a(\theta)$ de l'équation (I.4.3) par -1 dans le mode de polarisation TM (qui correspond à prendre des conditions aux limites de Dirichlet sur Γ_{int}), et par i dans le mode de polarisation TE (qui correspond à prendre des conditions aux limites de Neumann sur Γ_{int}). Par définition, la SER est $20\log(|a(\theta)|^2)$. Notons qu'en bidimensionnel cette quantité est en toute rigueur appelée LER et se mesure en $dB.m$.

I.4.2.3 Diffraction sur un ballon de football.

Le ballon de football est maillé dans un domaine entouré d'un cercle de diamètre trois fois le diamètre du ballon. Le nombre d'éléments du maillage est 346. Le nombre d'arêtes libres extérieures (frontière absorbante artificielle) est 38, le nombre d'arêtes libres intérieures est 24. Nous obtenons figure I.4.9 les calculs de SER. Soulignons que notre pas de discrétisation du ballon de football à 1500 MHz est extrêmement grossier : le paramètre de raffinement du maillage est de l'ordre de la longueur d'onde. Pour comparer avec le code SHF2D, nous avons dû multiplier le nombre d'éléments autour du profil par un facteur 5 environ, de façon à vérifier la loi empirique de discrétisation des Equations Intégrales $h \approx \lambda/5$. Dans un cas contraire, le code d'équations intégrales ne converge pas, même vers une solution éloignée.



Mode TE, 1500 MHz. Modes TM et TE à 150 MHz.

FIG. I.4.9 – Ballon de football SER, UWVF-IE.

I.4.2.4 Diffraction sur un profil NACA.

Le domaine de calcul est maillé en trois parties.

- Domaine 1. Autour du profil, ressemble à une couche limite, épaisse d'environ 3% de la longueur de l'aile. Le paramètre de taille du maillage représente 1% de la longueur de l'aile.
- Domaine 2. Ce domaine entoure la couche limite (domaine 1) dans un domaine limité, à peu près de même surface que le profil. Le paramètre de taille du maillage représente 5% de la longueur de l'aile. Les domaines de calcul 1 et 2 sont les domaines dans lesquels nous représentons la carte des iso-valeurs des champs, figures I.4.10 à I.4.13.
- Domaine 3. C'est le domaine principal en ce qui concerne la surface. Limité par le domaine précédent et le cercle de diamètre trois fois la longueur du profil. Le paramètre de taille du maillage représente 20% de la longueur de l'aile. Soulignons que ce maillage est extrêmement grossier puisque à 1500 MHz il correspond à $h \approx 2\lambda$, qui est très loin de la loi empirique (I.0.3) entre le pas de discrétisation et la fréquence.

Nous obtenons 2976 éléments et 1615 nœuds. Le nombre d'arêtes libres extérieures (frontière absorbante artificielle) est de 48, le nombre d'arêtes libres intérieures est de 202. Pour l'aile NACA, nous donnons une représentation du champ u autour du profil. Ceci est réalisé à l'aide d'une grille plus fine que celle du maillage de discrétisation. Nous calculons u_h à l'aide de la formule (I.2.27) section I.2.1.5. Le maillage utilisé pour la représentation du champ est constitué des domaines 1 et 2 décrits ci-dessus, mais avec un paramètre de taille du maillage uniforme qui vaut 1% de la longueur de l'aile. Nous obtenons 3572 éléments et 2010 nœuds. Le nombre d'arêtes libres extérieures (frontière graphique artificielle) est de 246, le nombre d'arêtes libres intérieures est de 202. Les valeurs des parties réelles ou imaginaires du champ diffracté à 1 500 MHz sont données par les courbes de niveau figures I.4.10, I.4.11, I.4.12 et I.4.13. L'onde incidente aborde le profil en incidence rasante, soit par le bord de fuite (figures I.4.10 et I.4.11), soit par le bord d'attaque (figures I.4.12 et I.4.13).

Fig. I.4.10 – Bord de fuite, Dirichlet TM, $\Im(u_D)$.

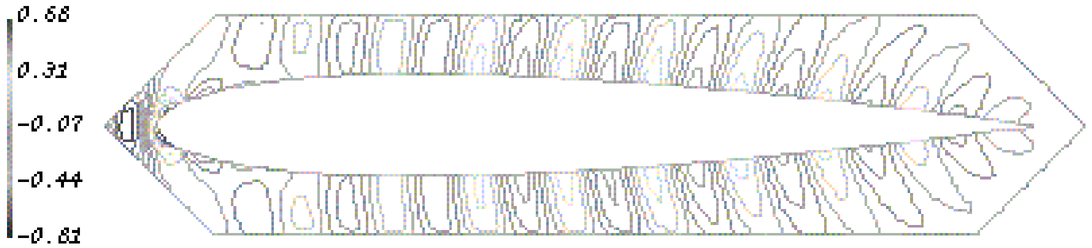


Fig. I.4.11 – Bord de fuite, Neumann TE, $\Im(u_D)$.

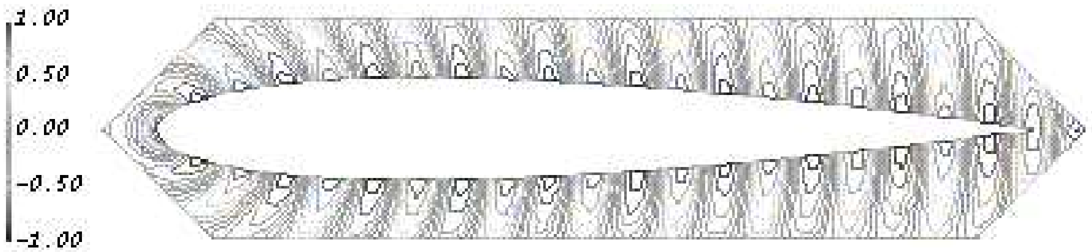


Fig. I.4.12 – Bord d'attaque, Dirichlet TM, $\Re(u_D)$.

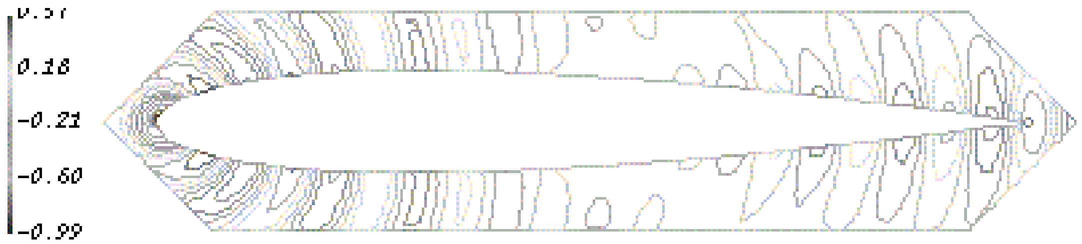
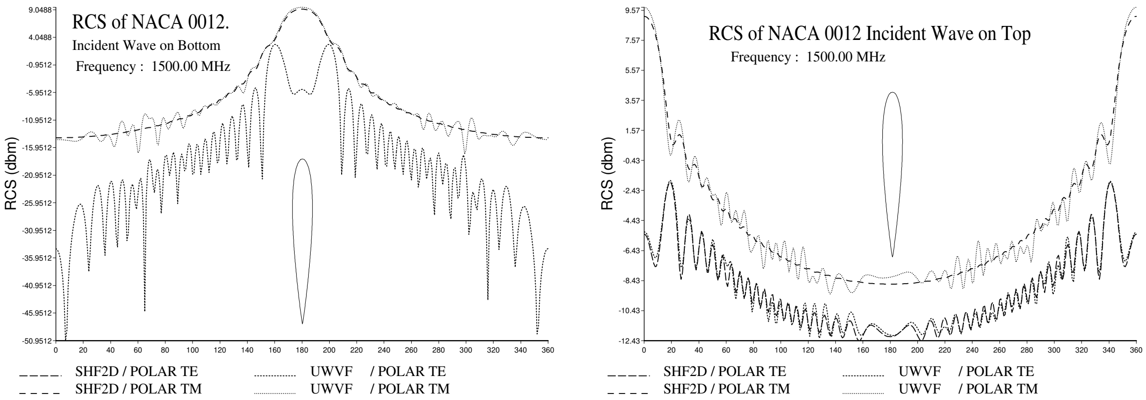


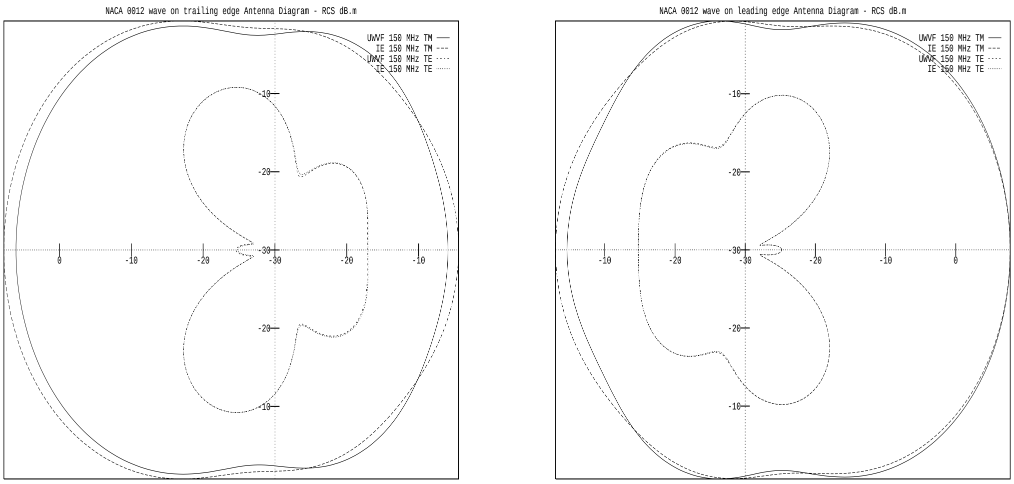
Fig. I.4.13 – Bord d'attaque, Neumann TE, $\Re(u_D)$.

Nous obtenons, figures I.4.14 et I.4.15, les calculs de SER comparés à ceux du code SHF2D en balayage bistatique. La figure I.4.15 présente la comparaison sous la forme d'un diagramme d'antenne, c'est-à-dire en coordonnées polaires : ce genre de représentation permet de voir facilement l'énergie diffusée selon la direction d'observation. Le lecteur notera que, à 1500 MHz, nos calculs donnent de légères oscillations autour des valeurs calculées par SHF2D dans le mode TM : ceci provient d'une contribution de la CLA, contribution parasite faible, qui est plus apparente dans le mode TM.



Bord de fuite, bord d'attaque.

Fig. I.4.14 – NACA 0012 SER, UWVF-IE comparaison à 1500 MHz.



Bord de fuite, bord d'attaque.

Fig. I.4.15 – Diagramme d'antenne NACA 0012 SER, UWVF-IE comparaison à 150 MHz.

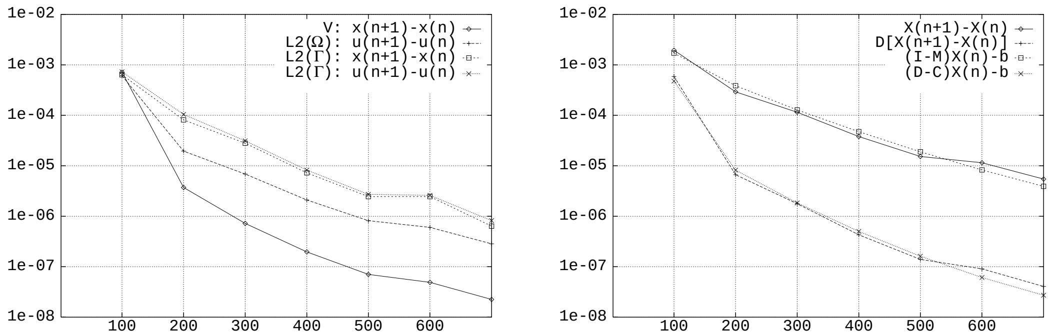
I.4.3 Vitesse de convergence de l'algorithme itératif.

Nous observons la vitesse de convergence de l'algorithme itératif sur le cas test de la propagation libre dans un cube décrit section I.3.2 avec les données du tableau I.4.4. Cette simulation est motivée par la forte dépendance du temps total de calcul du code par rapport au nombre d'itérations effectuées par l'algorithme itératif (I.2.31) de résolution du système linéaire (cf annexe III.C.4).

Tab. I.4.4 – Etude de la convergence de l'algorithme

Variable	Valeur
f	0
g	cf équation (I.3.6)
Q	0
$\vec{v}_0$	$(1.009946454058, i \times 0.1413925035682)$
ω	4.188786015996
p	5
discrétisation	$\frac{\lambda}{5}$

Nous étudions (figures I.4.16 et I.4.17) les évolutions de différents critères en fonction du nombre d'itérations effectuées.



(α) Normes d'espaces fonctionnels. (β) Normes de vecteurs.

Fig. I.4.16 – Evolution de normes entre deux itérations

1. Normes relatives sur Γ et Ω dans les espaces V et L^2 de la différence entre les solutions numériques respectivement x_h et u_h aux itérations n et $n+1$, figure I.4.16 (α).
2. Normes relatives dans $L^2(\mathbb{C}^{pK})$ des vecteurs solutions X aux itérations n et $n+1$, de DX, de $(I-M)X-b$ et de $(D-C)X-b$, figure I.4.16 (β).
3. Mêmes mesures qu'au point 1 mais sur $(x-x_h)$ et $(u-u_h)$, figure I.4.17.

Nous constatons qu'il y a un nombre optimal d'itérations à effectuer pour atteindre la précision maximale possible pour une discrétisation donnée. En l'occurrence, après 200 itérations la solution approchée n'évolue plus guère par rapport à la solution exacte. Nous observons aussi qu'en l'absence de connaissance de la solution exacte il serait difficile de savoir quand stopper les itérations.

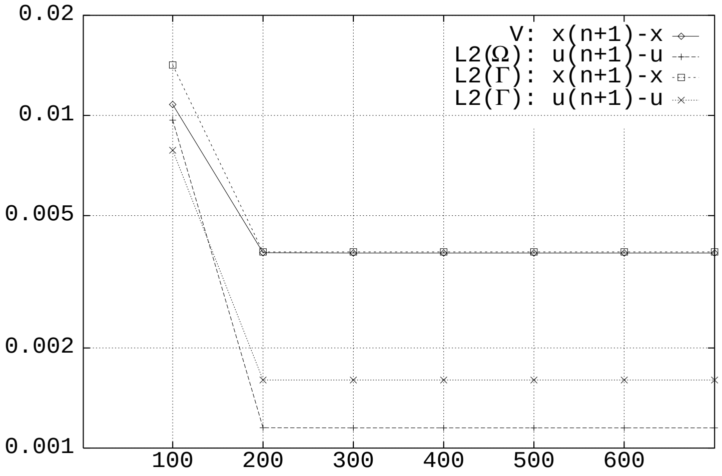


FIG. I.4.17 – Evolution de normes d'erreur en fonction du nombre d'itérations.

Chapitre I.5

Extension au cas des coefficients variables.

Le but de ce chapitre n'est pas d'étendre tous les résultats obtenus sur le problème de Helmholtz à coefficients constants mais de montrer comment la généralisation peut s'effectuer. Nous montrerons donc essentiellement la formulation variationnelle et ses propriétés.

I.5.1 Présentation du problème et formulation.

I.5.1.1 Cadre du problème.

Nous étudions la généralisation du problème modèle (I.0.1) au cas de la propagation acoustique plan dans un milieu borné Ω, de frontière Γ lipschitzienne, constitué de matériaux aux caractéristiques constantes par morceaux qui définissent le tenseur bidimensionnel μ et la fonction scalaire ρ. On modélise ce phénomène par

$$\text{(I.5.1)} \qquad \begin{cases} -\nabla.[\mu\,\nabla\, u] - \omega^2 \rho u = f & \text{dans } \Omega \\ (+\nu^\top \mu\,\nabla\, u + i\omega\sigma u) = Q(-\nu^\top \mu\,\nabla\, u + i\omega\sigma u) + g & \text{sur } \Gamma \end{cases}$$

où σ est une fonction essentiellement bornée et Q un opérateur réel de norme strictement inférieure à 1. Les hypothèses physiques sur les fonctions μ, ρ et σ sont qu'elles sont supérieurement et inférieurement essentiellement bornées, soit, (au sens matriciel pour μ),

$$\text{(I.5.2)} \qquad \begin{cases} \Re(\mu) \geq \mu_0 \ \ \text{une matrice réelle définie positive,} \\ \Re\rho \geq \rho_0 > 0, \\ \Re\sigma \geq \sigma_0 > 0\ . \end{cases}$$

D'autre part, on suppose que le milieu est dissipatif, soit

$$\text{(I.5.3)} \qquad \begin{cases} \Im(\mu) \geq 0 \\ \Im\rho \leq 0\ . \end{cases}$$

Remarquons que pour $\rho = 1$, $\sigma = 1$ et $\mu = I_2$ la matrice identité du plan, le problème (I.5.1) est le problème de Helmholtz sans coefficient traité dans cette première partie. En effet, l'opérateur $\nu^\top \mu\,\nabla$ vérifie alors

$$\nu^\top \mu\,\nabla = \frac{\partial}{\partial\nu}\ .$$

La formulation variationnelle classique est (en posant $\zeta = \sigma\dfrac{1-Q}{1+Q}$ avec $\Re(\zeta) > 0$)

$$\text{(I.5.4)} \qquad \begin{cases} \forall v \in H^1(\Omega) \\ \displaystyle\int_\Omega \nabla\, u^\top \mu\,\nabla\, \overline{v} - \omega^2 \int_\Omega \rho u\, \overline{v} + i\omega \int_\Gamma \zeta u\, \overline{v} = \int_\Omega f\, \overline{v} + \int_\Gamma g\, \overline{v}\ . \end{cases}$$

L'hypothèse $\Re(\mu) \geq \mu_0$ définie positive signifie que la forme

$$(u,v) \in [H^1(\Omega)]^2 \mapsto \int_\Omega \nabla u^\top \mu \nabla \overline{v} + u\,\overline{v} \tag{I.5.5}$$

est coercive sur $H^1(\Omega)$. On montre ensuite que l'opérateur

$$(u,v) \in [H^1(\Omega)]^2 \mapsto -(1+\omega^2)\int_\Omega \rho u\,\overline{v} - \int_\Omega u\,\overline{v} + i\omega \int_\Gamma \zeta u\,\overline{v} \tag{I.5.6}$$

est une perturbation compacte de l'opérateur coercif par l'injection compacte de $H^1(\Omega)$ dans $L^2(\Omega)$ pour Ω borné ([23]) et par l'injection continue de $H^{1/2}(\Gamma)$ dans $L^2(\Gamma)$. Cela nous place dans le cadre de l'alternative de Fredholm. de conclure à l'existence et l'unicité du problème continu.

On utilise alors le fait que $\Re\sigma \geq \sigma_0 > 0$ et $\Im\mu \geq 0$ et $\Im\rho \leq 0$: ceci permet de montrer l'unicité par le théorème de prolongement d'Holmgren.

Par la suite, nous supposerons que σ est réel.

I.5.1.2 Nouvelle formulation variationnelle.

Théorème 9 *Soit u une solution du problème de Helmholtz (I.5.1) qui vérifie l'hypothèse de régularité $-{\nu_k}^\top \mu \nabla u \in L^2(\partial\Omega_k)$ pour tout k. Alors, pour tout $e_k \in L^2(\Omega_k)$ tel que, pour σ un réel positif quelconque,*

$$\begin{cases} -\nabla.[\mu^* \nabla e_k] - \omega^2 \overline{\rho} e_k = 0 \ \textit{dans}\ \Omega_k \\ (-\nu^\top \mu \nabla + i\omega\sigma)(e_k)_{|\partial\Omega_k} \in L^2(\partial\Omega_k) \ , \end{cases} \tag{I.5.7}$$

on a :

$$\begin{aligned} &\sum_k \int_{\partial\Omega_k} (-{\nu_k}^\top \mu \nabla u + i\omega\sigma u)\,\overline{(-{\nu_k}^\top \mu^* \nabla e_k + i\omega\sigma e_k)} \\ &- \sum_{k,j} \int_{\Sigma_{kj}} (-{\nu_j}^\top \mu \nabla u + i\omega\sigma u)\,\overline{(+{\nu_k}^\top \mu^* \nabla e_k + i\omega\sigma e_k)} \\ &- \sum_k \int_{\Gamma_k} Q(-\nu^\top \mu \nabla u + i\omega\sigma u)\,\overline{(+\nu^\top \mu^* \nabla e_k + i\omega\sigma e_k)} \\ &= -2i\omega\sigma \sum_k \int_{\Omega_k} f\,\overline{e_k} + \sum_k \int_{\Gamma_k} g\,\overline{(+\nu^\top \mu^* \nabla e_k + i\omega\sigma e_k)} \ . \end{aligned} \tag{I.5.8}$$

Preuve. En développant le premier terme dans l'égalité (I.5.8), nous avons :

$$\begin{aligned} &\int_{\partial\Omega_k} (-{\nu_k}^\top \mu \nabla u + i\omega\sigma u)\,\overline{(-{\nu_k}^\top \mu^* \nabla e_k + i\omega\sigma e_k)} \\ &\quad = \int_{\partial\Omega_k} (+{\nu_k}^\top \mu \nabla u + i\omega\sigma u)\,\overline{(+{\nu_k}^\top \mu^* \nabla e_k + i\omega\sigma e_k)} \\ &\qquad - 2i\omega \int_{\partial\Omega_k} \left(\sigma u\,\overline{({\nu_k}^\top \mu^* \nabla e_k)} - ({\nu_k}^\top \mu \nabla u)\,\overline{\sigma e_k}\right) \ . \end{aligned} \tag{I.5.9}$$

Rappelons que, d'après (I.5.1) et (I.5.7),

$$\begin{cases} -\nabla.[\mu^* \nabla e_k] - \omega^2 \overline{\rho} e_k = 0 & \text{dans } \Omega_k \\ -\nabla.[\mu \nabla u] - \omega^2 \rho u = f & \text{dans } \Omega \ , \end{cases} \tag{I.5.10}$$

ainsi, par intégrations par parties dans (I.5.10),

$$\begin{cases} \displaystyle\int_{\Omega_k} (\nabla \overline{u})^\top \mu^* \nabla e_k - \omega^2 \overline{\rho u} e_k = \int_{\partial\Omega_k} \overline{u}({\nu_k}^\top \mu^* \nabla e_k) \\ \displaystyle\int_{\Omega_k} (\nabla \overline{e_k})^\top \mu \nabla u - \omega^2 \rho u \overline{e_k} = \int_{\partial\Omega_k} ({\nu_k}^\top \mu \nabla u)\overline{e_k} + \int_{\Omega_k} f\,\overline{e_k} \ . \end{cases} \tag{I.5.11}$$

Le terme $(\nabla \overline{u})^\top \mu^* e_k$ est scalaire donc égal à son transposé. On conjugue tous les termes dans la première équation de (I.5.11) :

$$\left\{ \begin{aligned} &\int_{\Omega_k} (\nabla \overline{e_k})^\top \mu \nabla u - \omega^2 \rho u \overline{e_k} = \int_{\partial\Omega_k} u \overline{(\nu_k{}^\top \mu^* \nabla e_k)} \\ &\int_{\Omega_k} (\nabla \overline{e_k})^\top \mu \nabla u - \omega^2 \rho u \overline{e_k} = \int_{\partial\Omega_k} (\nu_k{}^\top \mu \nabla u) \overline{e_k} + \int_{\Omega_k} f\, \overline{e_k} \ . \end{aligned} \right. \tag{I.5.12}$$

De (I.5.9) et (I.5.12) nous avons

$$\begin{aligned} &\int_{\partial\Omega_k} (-\nu_k{}^\top \mu \nabla u + i\omega\sigma u)\, \overline{(-\nu_k{}^\top \mu^* \nabla e_k + i\omega\sigma e_k)} \\ &- \int_{\partial\Omega_k} (+\nu_k{}^\top \mu \nabla u + i\omega\sigma u)\, \overline{(+\nu_k{}^\top \mu^* \nabla e_k + i\omega\sigma e_k)} = 2i\omega\sigma \int_{\Omega_k} f\, \overline{e}_k \ . \end{aligned} \tag{I.5.13}$$

La continuité de u sur Σ_{kj} et la condition aux limites du système (I.5.1) s'écrivent :

$$\left\{ \begin{aligned} &(\nu_k{}^\top \mu \nabla u + i\omega\sigma u)_{|\Sigma_{kj}} = (-\nu_j{}^\top \mu \nabla + i\omega\sigma) u_{|\Sigma_{jk}} \\ &(+\nu^\top \mu \nabla u + i\omega\sigma u) = Q(-\nu^\top \mu \nabla u + i\omega\sigma u) + g \ . \end{aligned} \right. \tag{I.5.14}$$

Dans le deuxième terme de l'équation (I.5.13) nous remplaçons la somme sur $\partial\Omega_k$ par une somme sur Γ_k et Σ_{kj}. Alors, à l'aide des relations (I.5.14), nous obtenons immédiatement l'équation (I.5.8). □

Comme pour le problème sans coefficient, nous définissons les opérateurs E, F, Π et A.

Définition 5

$$\begin{aligned} &E = \left\{ \begin{aligned} &V \to L^2(\Omega) \\ &y \mapsto e = (e_k),\ e_k = e_{|\Omega_k} \end{aligned} \right. \\ &\textit{avec} \left\{ \begin{aligned} &y_{|\partial\Omega_k} = (-\nu^\top \mu \nabla e_k + i\omega\sigma e_k) && \textit{sur } \partial\Omega_k \\ &-\nabla.[\mu^* \nabla e_k] - \omega^2 \overline{\rho} e_k = 0 && \textit{dans } \Omega_k \ . \end{aligned} \right. \end{aligned} \tag{I.5.15}$$

$$F = \left\{ \begin{aligned} &V \to V \\ &z = (+\nu^\top \mu^* \nabla + i\omega\sigma) e_k \mapsto Fz = (-\nu^\top \mu^* \nabla + i\omega\sigma) e_k \\ &Fz = (-\nu^\top \mu^* \nabla + i\omega\sigma) e_k (E(z))_{|\partial\Omega_k})_k \ . \end{aligned} \right. \tag{I.5.16}$$

Considérant une fonction complexe Q définie sur Γ vérifiant $|Q| \leq 1$, définissons l'opérateur Π par :

$$\Pi = \left\{ \begin{aligned} &V \to V \\ &z_{|\Sigma_{kj}} \mapsto z_{|\Sigma_{jk}} \\ &z_{|\Gamma_k} \mapsto Q z_{|\Gamma_k} \ . \end{aligned} \right. \tag{I.5.17}$$

Notons F^* l'adjoint de F. Définissons l'opérateur A de V dans V par

$$A = F^* \Pi \ . \tag{I.5.18}$$

On a de la même façon que pour le problème sans coefficient le théorème d'existence et d'unicité suivant. Les arguments des preuves sont identiques.

Théorème 10

a) Trouver x solution de (I.1.8) est équivalent à

$$\left\{ \begin{aligned} &\textit{Trouver } x \in V \textit{ tel que } \forall y \in V \\ &(x, y)_V - (\Pi x, F y)_V = (b, y)_V \ . \end{aligned} \right. \tag{I.5.19}$$

où le second membre $b \in V$ est défini, via le théorème de représentation de Riesz, par :

$$\forall y \in V \quad (b, y)_V = -2i\omega\sigma \sum_k \int_{\Omega_k} f\, \overline{E(y)}_{\Omega_k} + \sum_k \int_{\Gamma_k} g\, \overline{F(y)}_{\Gamma_k} \ . \tag{I.5.20}$$

b) Si u est solution du problème de Helmholtz (I.5.1) alors $x = (-\nu_k{}^\top \mu \nabla u + i\omega\sigma u)$ est solution de (I.5.19) sous l'hypothèse de régularité $x \in V$.

c) Réciproquement, si x est solution de (I.5.19) alors $u = E_f(x)$ est l'unique solution de (I.5.1). Le problème (I.5.19) est équivalent à :

$$\begin{cases} \text{trouver } x \in V \\ \boxed{(I - A)x = b} \end{cases} \tag{I.5.21}$$

I.5.1.3 Propriétés de la formulation.

L'opérateur A vérifie les propriétés essentielles suivantes :

Proposition 8 *La norme induite de A vérifie $||A|| \leq 1$.*

Proposition 9 *L'opérateur $(I - A)$ est injectif.*

Les preuves sont aisées à l'aide du lemme suivant :

Lemme 11 *L'opérateur F est une contraction.*

Preuve. Soit $e = E(y)$ (comme dans la définition (5)). Alors :

$$\begin{aligned} (Fy, Fy) &= \int_{\partial\Omega_k} |(\nu_k{}^\top \mu^* \nabla e_k + i\omega\sigma e_k)|^2 \\ &= \int_{\partial\Omega_k} |\nu_k{}^\top \mu^* \nabla e_k|^2 + \omega^2 |\sigma|^2 |e_k|^2 - 2\omega \Im \left(\int_{\partial\Omega_k} \sigma \nu_k{}^\top \mu^* \nabla e_k . \overline{\sigma e_k} \right) . \end{aligned} \tag{I.5.22}$$

$$\begin{aligned} (y, y) &= \int_{\partial\Omega_k} |(-\nu_k{}^\top \mu^* \nabla e_k + i\omega\sigma e_k)|^2 \\ &= \int_{\partial\Omega_k} |\nu_k{}^\top \mu^* \nabla e_k|^2 + \omega^2 |\sigma|^2 |e_k|^2 + 2\omega \Im \left(\int_{\partial\Omega_k} \sigma \overline{\nu_k{}^\top \mu^* \nabla e_k} \sigma e_k \right) . \end{aligned} \tag{I.5.23}$$

$$(y, y) - (Fy, Fy) = 2\omega \Im \left(\int_{\partial\Omega_k} \sigma \overline{\nu_k{}^\top \mu^* \nabla e_k} \sigma e_k - \int_{\partial\Omega_k} \sigma \nu_k{}^\top \mu^* \nabla e_k . \overline{\sigma e_k} \right) . \tag{I.5.24}$$

$$(y, y) - (Fy, Fy) = -4\omega \Im \left(\int_{\partial\Omega_k} \sigma \nu_k{}^\top \mu^* \nabla e_k . \overline{\sigma e_k} \right) . \tag{I.5.25}$$

Intégrons par parties dans

$$\int_{\Omega_k} (-\nabla . [\mu^* \nabla e_k] - \omega^2 \overline{\rho} e_k) \overline{\sigma e_k} = 0 \tag{I.5.26}$$

on a alors :

$$\int_{\Omega_k} \overline{\sigma} (\nabla \overline{e_k})^\top \mu^* \nabla e_k - \omega^2 \overline{\sigma\rho} |e_k|^2 = \int_{\partial\Omega_k} \overline{\sigma} \nu_k{}^\top \mu^* \nabla e_k \overline{\sigma e_k}, \tag{I.5.27}$$

et

$$(y, y) - (Fy, Fy) = -4\omega \Im \int_{\Omega_k} \overline{\sigma} (\nabla \overline{e_k})^\top \mu^* \nabla e_k - \omega^2 \overline{\sigma\rho} |e_k|^2 \tag{I.5.28}$$

Le terme

$$4\omega \Im \int_{\Omega_k} \sigma \nabla e_k{}^\top \mu \nabla \overline{e_k} - \omega^2 \sigma\rho |e_k|^2 \tag{I.5.29}$$

est positif. En effet, nous avons supposé σ réel positif non nul et

$$\begin{cases} \Im\rho \leq 0 \\ \Im\mu \geq 0 \ . \end{cases} \tag{I.5.30}$$

Ceci implique que

$$\|F\|^2 = \sup_{y\in V, y\neq 0} \sqrt{\frac{(Fy,Fy)}{(y,y)}} \leq 1 \tag{I.5.31}$$

□

Preuve. (de la proposition 8). L'opérateur Π défini par (I.5.17) vérifie $\|\Pi\| \leq 1$. Ceci, combiné avec le lemme 11 donne $||A|| \leq 1$. □

Preuve. (de la proposition 9). Nous supposons l'existence de x tel que $x = Ax$ et nous vérifions $x = 0$. Soit $w = E(x)$ (E défini par (5)), en d'autres termes $x_{|\partial\Omega_k} = (-\nu^\top \mu \nabla w + i\omega\sigma w)_{|\Omega_k}$ avec

$$-\nabla.[\mu^* \nabla w] - \omega^2 \overline{\rho} w = 0 \text{ dans } \Omega_k \ . \tag{I.5.32}$$

Soit G l'opérateur défini par :

$$G = \begin{cases} V \to V \\ z = (+\nu^\top \mu^* \nabla + i\omega\sigma)w \mapsto Gz = (-\nu^\top \mu^* \nabla + i\omega\sigma)w \ . \end{cases} \tag{I.5.33}$$

L'équation $x = Ax$ multipliée à gauche par l'opérateur G devient $GF^*\Pi x = Gx$, à l'aide de $GF^* = I$, on a $Gx = \Pi x$. Ceci se réécrit sur w :

$$(\nu_k{}^\top \mu \nabla + i\omega\sigma) w_{|\Sigma_{kj}} = (-\nu_j{}^\top \mu \nabla + i\omega\sigma) w_{|\Sigma_{jk}} \tag{I.5.34}$$

$$(\nu_k{}^\top \mu \nabla + i\omega\sigma) w_{|\Gamma_k} = Q(-\nu_k{}^\top \mu \nabla + i\omega\sigma) w_{|\Gamma_k} \ . \tag{I.5.35}$$

Les équations (I.5.32) et (I.5.35) signifient que w est solution d'un problème de Helmholtz (I.5.1). L'équation (I.5.32) implique que le second membre f est nul sur Ω. L'équation (I.5.35) implique que g est nul sur Γ. D'après le théorème d'existence et d'unicité, on a $w = 0$. Ainsi, nous avons $x = 0$. □

I.5.2 Approximation de Galerkin.

On effectue la même procédure de Galerkin qu'en I.2.1.1 p. 23. On a donc le même théorème d'existence et d'unicité du problème variationnel dans V_h (I.2.1) et l'on définit toujours les matrices du problème linéaire (I.5.36).

$$\begin{cases} \text{Trouver } X \in \mathbb{C}^{pK} \\ (D - C)X = b \end{cases} \tag{I.5.36}$$

La matrice D est la matrice du produit scalaire dans V_h, C est la matrice de la forme bilinéaire $(\Pi x_h, Fy_h)$, le second membre, par abus de langage, est toujours noté b. Nous présentons comment calculer les termes du système linéaire I.5.36 puis comment calculer une approchée discrète de la solution du problème I.5.21. L'implémentation numérique est proposée par le choix d'un espace V_h qui permet une inversibilité inconditionnelle du système I.5.36.

I.5.2.1 Construction des opérateurs discrétisés.

i) Les coefficients de la matrice D définis par $D_{k,j}^{l,m} = (z_{jm}, z_{kl})_V$ sont donnés par

$$D_{k,j}^{l,m} = \int_{\partial\Omega_k} (-\nu_k{}^\top \mu \nabla + i\omega\sigma) e_{jm} \overline{(-\nu_k{}^\top \mu \nabla + i\omega\sigma) e_{kl}} \ . \tag{I.5.37}$$

On a donc $j \neq k \Rightarrow D_{k,j}^{l,m} = 0$.

ii) Les coefficients de la matrice C définis par $C_{k,j}^{l,m} = (\Pi z_{jm}, F z_{kl})_V$ sont :

$$\begin{cases} C_{k,j\neq k}^{l,m} = \displaystyle\int_{\Sigma_{kj}} (+\nu_k{}^\top \mu \nabla + i\omega\sigma) e_{jm} \overline{(+\nu_k{}^\top \mu \nabla + i\omega\sigma) e_{kl}} \\ C_{k,k}^{l,m} = \displaystyle\int_{\Gamma_k} Q\,(-\nu_k{}^\top \mu \nabla + i\omega\sigma) e_{jm} \overline{(+\nu_k{}^\top \mu \nabla + i\omega\sigma) e_{kl}} \end{cases} \tag{I.5.38}$$

iii) Le second membre b est construit par $b_{k,l} = (b, z_{kl})_V$:

$$b_{k,l} = -2i\omega\sigma \int_{\Omega_k} f\,\overline{(e_{kl})} + \int_{\Gamma_k} g\,\overline{(+\nu_k{}^\top \mu\,\nabla + i\omega\sigma) e_{kl}}\ . \tag{I.5.39}$$

I.5.2.2 Définition et construction de u_h à partir de x_h.

Nous décrivons comment définir et calculer une approximation de u (la solution du problème de Helmholtz), notée u_h à partir de x_h solution de (I.5.36). Il y a deux techniques indépendantes d'approximation de u :

i) Reconstruction de u_h sur $\partial\Omega_k$.

Comme dans le cas sans coefficient, on montre aisément que la trace de u sur V est liée à x par

$$\begin{cases} u = \dfrac{1}{2i\omega\sigma}[(I+\Pi)x] & \text{sur } \Sigma_{kj} \\ u = \dfrac{1}{2i\omega\sigma}[(I+\Pi)x + g] & \text{sur } \Gamma_k\ , \end{cases} \tag{I.5.40}$$

et l'on définit u_h sur les arêtes du maillage par

$$\begin{cases} u_h = \dfrac{1}{2i\omega\sigma}[(I+\Pi)x_h] & \text{sur } \partial\Omega_k \\ u_h = \dfrac{1}{2i\omega\sigma}[(I+\Pi)x_h + g] & \text{sur } \Gamma_k\ . \end{cases} \tag{I.5.41}$$

Pratiquement, l'équation (I.5.41) signifie que :

1. nous construisons u_h sur Γ par

$$2i\omega\sigma (u_h)_{|\Gamma_k} = g + (1+Q_k)\sum_l x_{kl}(-\nu_k{}^\top \mu\,\nabla + i\omega\sigma) e_{kl}\ , \tag{I.5.42}$$

2. et nous construisons u_h sur Σ_{kj} par

$$2i\omega\sigma (u_h)_{|\Sigma_{kj}} = \sum_{l(k)} x_{kl}(-\nu_k{}^\top \mu\,\nabla + i\omega\sigma) e_{kl} + \sum_{l(j)} x_{jl}(-\nu_j{}^\top \mu\,\nabla + i\omega\sigma) e_{jl}\ . \tag{I.5.43}$$

ii) Reconstruction de u_h dans Ω.

Comme pour le problème de Helmholtz sans coefficient, lorsque $f = 0$ et μ et ρ réels, nous pouvons reconstruire u_h par

$$(u_h)_{|\Omega_k} = \sum_{l=1}^{p} x_{kl}\, e_{kl}\ . \tag{I.5.44}$$

En effet, dans ce cas, l'opérateur de relèvement, E, est linéaire.

I.5.2.3 Choix de l'espace d'approximation et résolution du système linéaire.

L'introduction des coefficients dans le problème de Helmholtz bidimensionnel nous amène à définir des solutions de $-\nabla.[\mu^* \nabla e_k] - \omega^2 \overline{\rho} e_k = 0$ sous la forme d'ondes planes :

$$\begin{cases} \text{pour } ||\vec{v}_{kl}|| = 1\ , \\ e_{kl} = e^{(i\omega \vec{v_{k,l}}^\top \sqrt{\frac{\rho}{\mu}}\vec{x})}\ . \end{cases}$$

En effet,

$$\nabla e_k = i\omega \vec{v_{k,l}}^\top \sqrt{\frac{\rho}{\mu}} e^{(i\omega \vec{v_{k,l}}^\top \sqrt{\frac{\rho}{\mu}}\vec{x})}\ ,$$

d'où

$$\mu^* \nabla e_k = \sqrt{\mu^*}\sqrt{\mu^*} \nabla e_k = i\omega\sqrt{\rho}\sqrt{\mu^*}\sqrt{\mu^*}(\sqrt{\mu}^\top)^{-1} \vec{v_{k,l}} e^{(i\omega \vec{v_{k,l}}^\top \sqrt{\frac{\rho}{\mu}}\vec{x})} = i\omega\sqrt{\rho}\sqrt{\mu^*}\vec{v_{k,l}} e^{(i\omega \vec{v_{k,l}}^\top \sqrt{\frac{\rho}{\mu}}\vec{x})}$$

et finalement

$$\nabla.[\mu^* \nabla e_k] = -\omega^2 \rho \vec{v_{k,l}}\sqrt{\mu^*}(\sqrt{\mu}^\top)^{-1} \vec{v_{k,l}} e^{(i\omega \vec{v_{k,l}}^\top \sqrt{\frac{\rho}{\mu}}\vec{x})} = -\omega^2 \rho \vec{v_{k,l}} \vec{v_{k,l}} e^{(i\omega \vec{v_{k,l}}^\top \sqrt{\frac{\rho}{\mu}}\vec{x})}.$$

Remarque 31 Nous sommes toujours libres de choisir les directions de propagation $v_{k,l}$ des ondes planes comme en I.2.29 équiréparties dans le plan. De même, il est possible d'utiliser des informations d'études asymptotiques du comportement de la solution et d'introduire les ondes de pression et de cisaillement adéquates.

Le système (I.5.36) est résolu comme dans le cas sans coefficient. L'inversibilité de la matrice D est assurée pour des ondes planes si et seulement si les vecteurs d'onde sont tous distincts $\forall k, (\forall (i,j), \vec{v}_{ki} \neq \vec{v}_{kj})$ puisque la matrice $\sqrt{\frac{\rho}{\mu}}$ est définie positive.

I.5.3 Conclusion de l'étude du problème à coefficients variables.

L'introduction de coefficients dans l'équation de Helmholtz bidimensionnelle complique l'analyse de la méthode. Nous ne généraliserons pas les résultats. Il est clair néanmoins que l'analyse est généralisable sous des hypothèses restrictives sur ces coefficients.

La mise en œuvre numérique que nous proposons est en revanche aisée à effectuer et permet de traiter numériquement de nombreuses applications.

Conclusions et perspectives tirées de l'étude du problème de Helmholtz.

Sur le cas modèle du problème de Helmholtz bidimensionnel, nous avons montré que les propriétés essentielles de notre méthode sont les suivantes.

1. La mise en œuvre est semblable à celle d'une méthode d'éléments finis. Ceci permet d'appliquer la méthode à des milieux autres que le vide.
2. Pour un maillage donné, l'ordre de l'erreur est une fonction linéaire de la taille du système linéaire alors que dans le cas des éléments finis, l'ordre est en racine carrée de ce nombre.
3. Pour un problème continu dont la solution unique est assez régulière, la formulation ultra-faible discrète est toujours inversible, quel que soit le rapport entre le pas de la discrétisation et la longueur d'onde. Sur plusieurs exemples numériques, nous avons observé une bonne consistance des résultats pour une discrétisation de taille h de l'ordre de la longueur d'onde λ.
4. La méthode est remarquablement précise. Ceci provient du calcul analytique du système matriciel et de la robustesse de l'algorithme de Richardson utilisé dans le cadre de la formulation ultra-faible.

Cette première étude nous a donc encouragé à étudier l'utilisation de la formulation pour un autre problème d'ondes en fréquence, le problème de Maxwell harmonique tridimensionnel dans un milieu à caractéristiques réelles ou complexes. Cette étude fait l'objet de la deuxième partie de ce travail.

Deuxième partie

Le problème de Maxwell tridimensionnel.

Présentation de la deuxième partie.

Cette partie présente l'application de la formulation variationnelle ultra-faible au problème harmonique de Maxwell dans un domaine tridimensionnel borné Ω, dans le vide ou en présence d'un diélectrique de permittivité ε et de perméabilité μ par rapport au vide. Comme lors de l'étude du problème de Helmholtz, la frontière Γ du domaine sera supposée assez régulière, au moins lipschitzienne. Le problème de Maxwell en régime harmonique se modélise par :

Problème 1 **:** *trouver* $(\mathbf{E}, \mathbf{H})$ *tel que, dans* Ω*, pour des termes sources* $\mathbf{m}$ *et* $\mathbf{j}$ *de divergence nulle,*

$$\begin{cases} \nabla\wedge \mathbf{E} - i\omega\mu\mathbf{H} = -\mathbf{m} \\ \nabla\wedge \mathbf{H} + i\omega\varepsilon\mathbf{E} = \mathbf{j} \\ \nabla.(\varepsilon\mathbf{E}) = 0 \\ \nabla.(\mu\mathbf{H}) = 0 \end{cases} \tag{II.6.1}$$

et sur la frontière Γ*,*

$$-|\sqrt{\varepsilon}|\mathbf{E}\wedge\nu + |\sqrt{\mu}|\,(\mathbf{H}\wedge\nu)\wedge\nu = Q\left(|\sqrt{\varepsilon}|\mathbf{E}\wedge\nu + |\sqrt{\mu}|\,(\mathbf{H}\wedge\nu)\wedge\nu\right) + g\ . \tag{II.6.2}$$

▭

Le problème de Maxwell harmonique entre dans le cadre général d'application de la formulation variationnelle ultra-faible. Nous allons introduire le formalisme variationnel ultra-faible à l'aide d'un espace fonctionnel hilbertien L^2 sur les traces tangentielles de la solution sur le bord d'une partition du domaine Ω en K sous-domaines Ω_k, espace que nous notons V. Dans le vide, l'inconnue de notre formulation est

$$\mathcal{X} = \mathbf{E}\wedge\nu + (\mathbf{H}\wedge\nu)\wedge\nu$$

sur toutes les interfaces et faces frontières de la partition. La mise en place de la formulation et l'équivalence avec le problème de Maxwell de départ sont globalement présentées de la même façon que dans la première partie. Sous des hypothèses de régularité de la solution $(\mathbf{E}, \mathbf{H})$, nous montrons que le problème se traduit, comme pour le problème de Helmholtz, par la formulation

$$\begin{cases} \text{Trouver } \mathcal{X}\in V \text{ tel que} \\ (I-A)\mathcal{X} = b\ , \end{cases}$$

où $b \in V$ et l'opérateur $I - A$ a les mêmes propriétés qu'en première partie.

L'espace des fonctions tests est l'espace des fonctions $\mathcal{Y} \in V$ données par

$$\mathcal{Y} = \mathbf{E}^{'}\wedge\nu + \left(\mathbf{H}^{'}\wedge\nu\right)\wedge\nu$$

où les fonctions $(\mathbf{E}^{'}, \mathbf{H}^{'})$ sont solutions de

$$\text{(II.6.3)} \qquad \left\{ \begin{array}{l} \nabla\wedge \mathbf{E}^{'} - i\omega\overline{\mu}\mathbf{H}^{'} = 0 \\ \nabla\wedge \mathbf{H}^{'} + i\omega\overline{\varepsilon}\mathbf{E}^{'} = 0 \end{array} \right.$$

sur les éléments Ω_k, où l'on suppose ε et μ constants. On suppose que $(\mathbf{E}^{'}, \mathbf{H}^{'})$ sont d'énergie finie et que leur traces tangentielles sont L^2. Nous proposons une écriture équivalente de (II.6.3) dans le vide, où $\varepsilon = \mu = 1$, en introduisant des fonctions à polarisations "circulaires" $\mathbf{F}$ et $\mathbf{G}$ qui vérifient les équations découplés

$$\text{(II.6.4)} \qquad \left\{ \begin{array}{l} \nabla\wedge \mathbf{F} = -\omega\mathbf{F} \\ \nabla\wedge \mathbf{G} = +\omega\mathbf{G} \end{array} \right.$$

et qui sont reliées à $(\mathbf{E}^{'}, \mathbf{H}^{'})$ par

$$\text{(II.6.5)} \qquad \left\{ \begin{array}{l} \mathbf{F} = \mathbf{E}^{'} + i\mathbf{H}^{'} \\ \mathbf{G} = \mathbf{E}^{'} - i\mathbf{H}^{'} \ . \end{array} \right.$$

La discrétisation s'effectue par une procédure de Galerkin. Comme pour le problème de Helmholtz, nous choisissons des ondes planes, aux directions de propagation notées $V_{k,l}$. L'indice k indique que l'onde plane a son support dans l'élément Ω_k, l'indice l, qui varie de 1 à p le nombre de directions choisies par élément, est un indice local à l'élément Ω_k. Le nombre p, pour des raisons de mise en œuvre informatique, est constant sur tous les éléments. Le caractère tridimensionnel du problème impose de munir les ondes planes d'un vecteur polarisation constant. Les relations de Gauss (ou de divergence) imposent l'orthogonalité des directions de propagation et des polarisations. Le découplage des équations de Maxwell dans le vide (II.6.4) nous a permis de considérer des fonctions de base orthogonales dans le vide pour le produit scalaire dans V, issues de p couples d'ondes planes aux polarisations complexes conjuguées dans le vide. Ces ondes planes, en notant $\mathbf{X}$ le vecteur position, sont (sur un élément Ω_k) de la forme[1],

$$\left| \begin{array}{ll} \mathbf{E}^{'}{}_{k,l} & = \sqrt{\overline{\mu}_k}\left(\mathbf{E}^0_{k,l} + i\mathbf{E}^0_{k,l} \wedge V_{k,l}\right) e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,l}\,.\,\mathbf{X})} \\ \mathbf{E}^{'}{}_{k,l+p} & = \sqrt{\overline{\mu}_k}\left(\mathbf{E}^0_{k,l} - i\mathbf{E}^0_{k,l} \wedge V_{k,l}\right) e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,l}\,.\,\mathbf{X})} \end{array} \right. ,$$

où pour les indices k et l fixés, la famille

$$\left(\mathbf{E}^0_{k,l}, V_{k,l}, \mathbf{E}^0_{k,l} \wedge V_{k,l}\right)$$

est une base orthonormée directe de $\mathbb{R}^3$.

L'orthogonalité dans V des fonctions de base issues de ces ondes planes permet, dans le vide, de réduire (presque) de moitié la taille mémoire nécessaire au stockage du problème matriciel

$$\left\{ \begin{array}{l} \text{trouver } \mathcal{X} \in \mathbb{C}^{2pK} \text{ tel que} \\ (D - C)\mathcal{X} = b \ , \end{array} \right.$$

où $b \in \mathbb{C}^{2pK}$. La matrice D du produit scalaire dans V_h est inversible sous la condition que les directions de propagation soient toutes distinctes deux à deux. Le fait que Ω est tridimensionnel complique la preuve par rapport à la preuve équivalente de la première partie. Le système linéaire est inversé par le même algorithme itératif. La matrice du système linéaire a la même propriété d'injectivité que celle issue de la discrétisation du problème de Helmholtz.

La similitude profonde, à la fois théorique sur les opérateurs du problème continu et du système linéaire discret, et pratique sur la mise en œuvre de l'espace de discrétisation constitué d'ondes planes,

[1] La racine carrée complexe est définie par la racine de partie réelle positive.

nous a poussé à examiner les problèmes d'estimations d'erreur, estimations étudiées en détails dans la première partie. Par exemple, l'estimation du résidu par l'erreur d'interpolation est une conséquence pure des propriétés de la formulation, communes aux deux parties. Sans chercher à simplement répéter les extrapolations logiques des lois concernant Helmholtz bidimensionnel à des lois pour Maxwell tridimensionnel, nous avons néanmoins mis en avant le problème central, à savoir l'estimation de l'erreur d'interpolation. Cette estimation montre qu'il existe un choix adéquat de $p = (N+1)(N+3)$ directions de propagation des ondes planes tel que l'erreur d'interpolation soit majorée par

$$||(I - P_h)\mathcal{X}||_V \leq Ch^{N+1/2} \ ,$$

où C est une constante strictement positive dépendant des données du problème harmonique et du choix des fonctions de base, mais indépendante de h le paramètre de taille du maillage. Cette propriété essentielle permet de montrer que la formulation ultra-faible appliquée au problème de Maxwell harmonique est asymptotiquement (en fonction de p) d'ordre plus élevé que la méthode des éléments finis P_k (en fonction de k). La loi d'ordre de convergence n'est plus linéaire comme en bidimensionnel, mais en racine carrée. Dans la méthode des éléments finis, l'ordre de convergence est en racine cubique de k.

L'intérêt d'un code Maxwell harmonique nous a poussé à nous concentrer sur l'étude des résultats numériques. Nous avons comparé le code $\mathcal{L}ior$ issu de la formulation ultra-faible à d'autres codes de résolution, notamment des codes éléments finis et volumes finis. Nous avons considéré particulièrement les aspects qui motivaient initialement cette étude dans le cadre du CEA : la viabilité de la méthode en terme de place mémoire, le coût informatique global (temps de calcul), la précision et la difficulté de la mise en œuvre des calculs. En somme, il s'agit de montrer en quoi cette formulation est intéressante alors qu'il existe déjà un grand nombre de méthodes déjà largement validées et utilisées.

Chapitre II.7

Construction de la formulation variationnelle.

II.7.1 Rappels sur le problème de Maxwell.

II.7.1.1 Problème physique.

Nous traitons du problème de Maxwell stationnaire dans un domaine tridimensionnel Ω décrit par la permittivité ε et la perméabilité μ, fonctions de la fréquence f (ou de la pulsation $\omega = 2\pi f$) et de la position notée $\mathbf{X}$. Les ondes électromagnétiques décrites par les champs électrique $\mathbf{E}$ et magnétique $\mathbf{H}$ sont éventuellement perturbées par des termes sources volumiques d'énergie finie, la densité de courant électrique $\mathbf{j}$ et la densité de charge magnétique $\mathbf{m}$. Nous formulons en outre les hypothèses suivantes :

1. le milieu de propagation est dissipatif, ce qui, dans la convention qui définit la transformée de Fourier d'une fonction f par $F(k) = \int_{-\infty}^{\infty} e^{+i\omega t} f(t) dt$, s'exprime par

$$\begin{aligned} &\forall \mathbf{X} \in \Omega, \\ &\Im(\varepsilon(\mathbf{X})) \geq 0 \\ &\Im(\mu(\mathbf{X})) \geq 0 \ , \end{aligned} \tag{II.7.1}$$

2. les fonctions ε et μ sont inférieurement bornées : il existe $\varepsilon_0^{'} > 0$ et $\mu_0^{'} > 0$ tels que

$$\begin{aligned} &\forall \mathbf{X} \in \Omega, \\ &\Re(\varepsilon(\mathbf{X})) \geq \varepsilon_0^{'} \\ &\Re(\mu(\mathbf{X})) \geq \mu_0^{'} \ . \end{aligned} \tag{II.7.2}$$

Dans le vide, la permittivité ε_0 et la perméabilité μ_0 définissent la célérité ou vitesse de propagation c des ondes électromagnétiques par

$$\varepsilon_0 \mu_0 = \frac{1}{c^2} \ . \tag{II.7.3}$$

Nous pouvons donc définir la permittivité et la perméabilité du milieu relativement au vide, ε_r et μ_r, par

$$\begin{cases} \varepsilon_r = \dfrac{\varepsilon}{\varepsilon_0} \\ \mu_r = \dfrac{\mu}{\mu_0} \ . \end{cases} \tag{II.7.4}$$

Par la suite, nous considérerons toujours les permittivité et perméabilité relatives que nous noterons dorénavant ε et μ puisqu'il n'y aura pas risque de confusion. Les caractéristiques du vide seront alors naturellement $\varepsilon = 1$ et $\mu = 1$.

Sous ces hypothèses, l'onde électromagnétique $(\mathbf{E},\mathbf{H})$ vérifie les équations de Maxwell harmoniques à la pulsation ω dans le domaine Ω

$$\begin{cases} \nabla\wedge\mathbf{E} - i\omega\mu\mathbf{H} = -\mathbf{m} & \text{Maxwell-Faraday,} \\ \nabla\wedge\mathbf{H} + i\omega\varepsilon\mathbf{E} = \mathbf{j} & \text{Maxwell-Ampère,} \\ \nabla.(\varepsilon\mathbf{E}) = 0 & \text{Gauss électrique,} \\ \nabla.(\mu\mathbf{H}) = 0 & \text{Gauss magnétique,} \end{cases} \tag{II.7.5}$$

où les termes sources volumiques $\mathbf{m}$ et $\mathbf{j}$ vérifient

$$\begin{cases} \nabla.(\mathbf{m}) = 0 \\ \nabla.(\mathbf{j}) = 0 \end{cases}$$

Notons que physiquement, la quantité de charge magnétique $\mathbf{m}$ n'existe pas mais nous la gardons pour la symétrie des équations. Nous pouvons définir les quantités suivantes qui sont intrinsèques au milieu de propagation :

1. le nombre d'onde k,
$$k = \sqrt{\omega^2\varepsilon_0\mu_0\varepsilon\mu}\ , \tag{II.7.6}$$
2. la longueur d'onde λ,
$$\lambda = \frac{2\pi}{|\Re(k)|}\ , \tag{II.7.7}$$
3. l'indice du milieu n,
$$n = \frac{ck}{\omega} =^{+}_{-} \sqrt{\frac{\varepsilon\mu}{\varepsilon_0\mu_0}}\ , \tag{II.7.8}$$
4. l'impédance relative du milieu Z (ou impédance intrinsèque [55]), de partie réelle positive,
$$Z = \frac{\omega\mu}{k} = \frac{k}{\omega\varepsilon} = \sqrt{\frac{\mu}{\varepsilon}}\ . \tag{II.7.9}$$

Rappelons que dans un domaine infini le champ électromagnétique $(\mathbf{E},\mathbf{H})$ doit vérifier la condition de radiation de Silver-Müller, condition qui s'écrit pour une position $\vec{r}$ à l'infini (dans un domaine complémentaire d'un borné) dans le vide où $\varepsilon = \mu = 1$ (d'après II.7.4) :

$$\begin{aligned} \lim_{r\to\infty} r\left(\mathbf{H}\wedge\frac{\vec{r}}{r} - \mathbf{E}\right) &= 0 \\ \lim_{r\to\infty} r\left(\mathbf{E}\wedge\frac{\vec{r}}{r} + \mathbf{H}\right) &= 0\ . \end{aligned} \tag{II.7.10}$$

Le champ à l'infini de la solution du problème (II.7.5) est donné, au premier ordre en r, par

$$\begin{aligned} \mathbf{E} &= \frac{e^{i\vec{k}.\vec{r}}}{r}\vec{a} + O(\frac{1}{r^2}) \\ \mathbf{H} &= \frac{e^{i\vec{k}.\vec{r}}}{r}\vec{b} + O(\frac{1}{r^2}) \end{aligned}$$

où a et b ne dépendent pas de r. On a donc, pour $\vec{e}_r = \frac{\vec{r}}{r}$,

$$\begin{aligned} \mathbf{H}\wedge\vec{e}_r &= \mathbf{E} + O(\frac{1}{r^2}) \\ \mathbf{E}\wedge\vec{e}_r &= -\mathbf{H} + O(\frac{1}{r^2})\ . \end{aligned}$$

Cette condition, écrite sur une sphère de rayon r en un point où le vecteur $\vec{e_r}$ est la normale ν, donne de façon équivalente

$$\left|\begin{array}{l}\mathbf{H}.\nu = O(\dfrac{1}{r^2}) \approx 0\\ \mathbf{E}.\nu = O(\dfrac{1}{r^2}) \approx 0\\ -\mathbf{E}\wedge\nu + (\mathbf{H}\wedge\nu)\wedge\nu = O(\dfrac{1}{r^2}) \approx 0\ .\end{array}\right.$$

Notons que cette condition aux limites n'est pas très performante pour approcher le problème infini en domaine borné. Nous ne discuterons pas ce point qui ne fait pas l'objet de notre exposé.

En domaine borné, nous pouvons écrire cette condition aux limites sur une frontière sphérique éloignée du centre du domaine (cf figure II.7.1) de façon à obtenir une condition aux limites absorbante d'ordre le plus faible ([8] ou [30]). Nous généralisons cette condition à tout domaine borné, assez régulier pour pouvoir définir la normale sortante ν sur son bord extérieur Γ_{ext} (que l'on suppose placé dans le vide, comme c'est le cas figure II.7.1), en présence d'un terme de courant de surface g. On obtient la condition aux limites

(II.7.11) $$-\mathbf{E}\wedge\nu + (\mathbf{H}\wedge\nu)\wedge\nu = g\ .$$

Nous modélisons la présence d'un obstacle opaque (figure II.7.1) dans le domaine Ω par la condition sur la trace tangentielle du champ électrique sur le bord intérieur Γ_{int},

(II.7.12) $$\mathbf{E}\wedge\nu = g\ ,$$

appelée condition de conducteur parfait ou condition sur un objet métallique. Notons aussi la condition aux limites retenue pour les équations intégrales qui modélise la présence d'une couche de matériau de faible épaisseur et de fort indice n (II.7.8) sur un objet métallique,

(II.7.13) $$-\mathbf{E}\wedge\nu + Z\,(\mathbf{H}\wedge\nu)\wedge\nu = g\ ,$$

où $Z \in \mathbb{C}$, ($\Re Z \geq 0$), est l'impédance de la couche ou l'impédance intrinsèque définie par (II.7.9). Cette condition est appelée condition de Léontovich ou condition d'impédance. En pratique, le terme "fort indice" signifie que l'on doit avoir $n \geq 10$ [59].

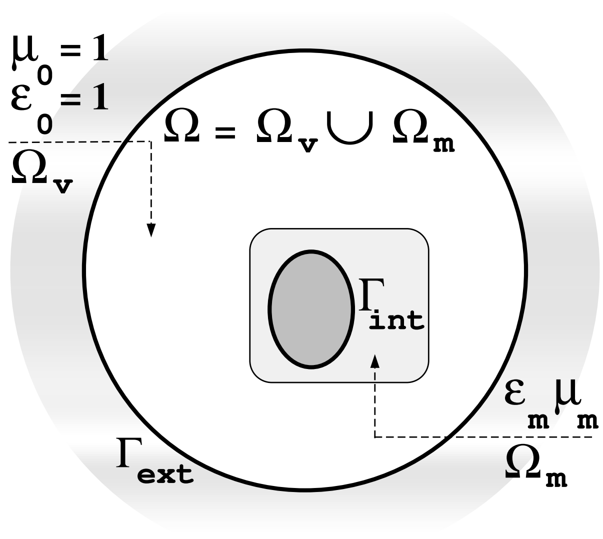


Fig. II.7.1 – Problème type en domaine borné.

Toutes ces conditions aux limites peuvent se résumer, en introduisant un opérateur de bord noté Q (qui pourra être une simple fonction du bord Γ à valeur complexe, ou être constant), de norme strictement inférieure à 1, et à l'aide d'une fonction g terme source d'énergie finie au bord du domaine, par

$$-|\sqrt{\varepsilon}|\mathbf{E}\wedge\nu+|\sqrt{\mu}|\,(\mathbf{H}\wedge\nu)\wedge\nu=Q\left(|\sqrt{\varepsilon}|\mathbf{E}\wedge\nu+|\sqrt{\mu}|\,(\mathbf{H}\wedge\nu)\wedge\nu\right)+g\ . \tag{II.7.14}$$

où les coefficients $|\sqrt{\varepsilon}|$ et $|\sqrt{\mu}|$ sont réels et permettent d'adimensionner les quantités à sommer. Nous retrouvons la condition aux limites de conducteur parfait (II.7.12) pour $Q=1$, la condition aux limites absorbante pour $Q=0$ dans le vide (II.7.11) ou même sur une frontière dans un milieu non absorbant où $|\sqrt{\varepsilon}|=\sqrt{\varepsilon}$ et de même pour μ, la condition d'impédance intrinsèque (II.7.13) pour $Q=0$ dans un matériau non absorbant, une relation d'impédance quelconque pour

$$Z=\frac{1-Q}{1+Q}\frac{|\sqrt{\mu}|}{|\sqrt{\varepsilon}|}$$

(la transformée de Cayley de Q) dans un matériau absorbant.

Résoudre le problème de Maxwell harmonique en domaine borné signifie donc trouver le couple $(\mathbf{E},\mathbf{H})$ vérifiant à la fois II.7.5 dans Ω et II.7.14 sur $\Gamma=\partial\Omega$. Notons que dans le vide, les lois II.7.5 deviennent, puisque $\varepsilon=1$ et $\mu=1$,

$$\begin{cases}\nabla\wedge\mathbf{E}-i\omega\mathbf{H}=0\\ \nabla\wedge\mathbf{H}+i\omega\mathbf{E}=\mathbf{j}\end{cases} \tag{II.7.15}$$

pour un terme source $\mathbf{j}$ de divergence nulle dans Ω (et $\mathbf{m}=0$, ce qui est une hypothèse physique), et

$$\begin{cases}\nabla.\,\mathbf{E}=0\\ \nabla.\,\mathbf{H}=0\ .\end{cases} \tag{II.7.16}$$

Nous constatons que la relation de Maxwell-Faraday implique la relation de Gauss magnétique. Il en est de même de la relation de Gauss électrique en l'absence de courant et de charges extérieures. Nous reformulons donc le problème de Maxwell dans l'espace des fonctions de divergence nulle dans le vide ($\varepsilon=\mu=1$) par

$$\begin{cases}\nabla\wedge\mathbf{E}-i\omega\mathbf{H}=-\mathbf{m} & \text{dans } \Omega\\ \nabla\wedge\mathbf{H}+i\omega\mathbf{E}=\mathbf{j} & \text{dans } \Omega\\ -\mathbf{E}\wedge\nu+(\mathbf{H}\wedge\nu)\wedge\nu=Q\left(\mathbf{E}\wedge\nu+(\mathbf{H}\wedge\nu)\wedge\nu\right)+g & \text{sur } \Gamma\end{cases} \tag{II.7.17}$$

où les termes sources $\mathbf{m}$ et $\mathbf{j}$ sont aussi de divergence nulle. En présence de matériau, nous considérerons le cas décrit par la figure type II.7.1 où les coefficients ε et μ sont constants par morceaux

$$\begin{cases}\nabla\wedge\mathbf{E}-i\omega\mu\mathbf{H}=-\mathbf{m} & \text{dans } \Omega\\ \nabla\wedge\mathbf{H}+i\omega\varepsilon\mathbf{E}=\mathbf{j} & \text{dans } \Omega\\ -|\sqrt{\varepsilon}|\mathbf{E}\wedge\nu+|\sqrt{\mu}|\,(\mathbf{H}\wedge\nu)\wedge\nu=Q\left(|\sqrt{\varepsilon}|\mathbf{E}\wedge\nu+|\sqrt{\mu}|\,(\mathbf{H}\wedge\nu)\wedge\nu\right)+g & \text{sur } \Gamma\ .\end{cases} \tag{II.7.18}$$

La situation de la figure II.7.1 nous permet de rester dans les espaces de fonctions de divergence nulle par morceaux avec les relations de saut (II.7.19), ou relations de continuité des traces tangentielles des champs, aux interfaces de discontinuité des perméabilité et permittivité entre les deux domaines Ω_m et Ω_v. Nous supposons que l'interface de discontinuité est assez régulière pour y définir une normale ν.

$$\begin{cases}\mathbf{H}_{|\Omega_m}\wedge\nu=\mathbf{H}_{|\Omega_v}\wedge\nu \text{ sur } \Omega_m\cap\Omega_v\\ \mathbf{E}_{|\Omega_m}\wedge\nu=\mathbf{E}_{|\Omega_v}\wedge\nu \text{ sur } \Omega_m\cap\Omega_v\ .\end{cases} \tag{II.7.19}$$

II.7.1.2 Cadre mathématique usuel et résultats d'existence et d'unicité.

Nous définissons l'espace des fonctions d'énergie finie $\mathrm{H}(\mathrm{rot},\Omega)$:

$$\mathrm{H}(\mathrm{rot},\Omega)=\left\{\mathbf{F}\in(L^2(\Omega))^3;\ \nabla\wedge\mathbf{F}\in(L^2(\Omega))^3\right\}\ . \tag{II.7.20}$$

Les fonctions $\mathbf{E}$ et $\mathbf{H}$ vérifiant (II.7.17) où ε et μ sont constants sur Ω seront naturellement, comme les termes sources $\mathbf{m}$ et $\mathbf{j}$, dans l'espace des fonctions de divergence nulle :

$$\text{(II.7.21)} \qquad \mathrm{H}(\mathrm{div}_0, \Omega) = \left\{ \mathbf{F} \in (L^2(\Omega))^3;\ \nabla . \mathbf{F} = 0 \right\} \ .$$

Le cadre fonctionnel de résolution des équations de Maxwell dans le vide est constitué des espaces suivants.

- Les termes sources volumiques $\mathbf{m}$ et $\mathbf{j}$ sont des fonctions de

$$\text{(II.7.22)} \qquad \mathrm{H}(\mathrm{div}_0, \Omega) \ .$$

- Le terme de bord g sur Γ appartient à

$$\text{(II.7.23)} \qquad L_t^2(\Gamma) = \left\{ \mathbf{F}; \mathbf{F}_{|\Gamma} \wedge \nu \in (L^2(\Gamma))^3;\ \mathbf{F}_{|\Gamma} . \nu = 0 \ \right\} \ .$$

- Le cadre fonctionnel adéquat pour les fonctions $\mathbf{E}$ et $\mathbf{H}$ pour un problème dans un milieu de caractéristiques réelles et constantes est

$$\text{(II.7.24)} \qquad \tilde{\mathcal{H}}(\Omega, \Gamma) = \left\{ \mathbf{F} \in (\mathrm{H}(\mathrm{rot}, \Omega) \cap \mathrm{H}(\mathrm{div}_0, \Omega));\ \mathbf{F}_{|\Gamma} \wedge \nu \in L_t^2(\Gamma) \right\} \ .$$

  Pour étudier les problèmes dans des milieux comme ceux de la figure II.7.1 formés de deux sous-domaines Ω_v et Ω_m d'interface assez régulière et où les permittivité et perméabilité sont égales à 1 sur Ω_v et constantes vérifiant les conditions (II.7.1) et (II.7.2) sur Ω_m, nous définissons $\mathcal{H}(\Omega, \Gamma)$ par

$$\text{(II.7.25)} \ \mathcal{H}(\Omega, \Gamma) = \left\{ \mathbf{F}; \mathbf{F}_{|\Omega_m} \in \tilde{\mathcal{H}}(\Omega_m, \partial\Omega_m),\ \mathbf{F}_{|\Omega_v} \in \tilde{\mathcal{H}}(\Omega_v, \partial\Omega_v);\ \mathbf{F}_{|\Omega_m} \wedge \nu = \mathbf{F}_{|\Omega_v} \wedge \nu \right\}$$

  Notons que si $\mathbf{F} \in \mathrm{H}(\mathrm{rot}, \Omega)$ alors $\mathbf{F}_{|\Gamma} \wedge \nu \in \mathrm{H}^{-1/2}(\mathrm{div}, \Gamma)$ par le théorème de trace continue surjective ([14]).

Donnons maintenant un résultat d'existence et d'unicité pour le problème (1) dans le cas de coefficients constants :

Problème 2 **:** *trouver* $(\mathbf{E}, \mathbf{H}) \in (\mathrm{H}(\mathrm{rot}, \Omega) \cap \mathrm{H}(\mathrm{div}_0, \Omega))^2$ *tel que*

- *dans* Ω,

$$\text{(II.7.26)} \qquad \begin{cases} \nabla \wedge \mathbf{E} - i\omega\mu\mathbf{H} = -\mathbf{m} \\ \nabla \wedge \mathbf{H} + i\omega\varepsilon\mathbf{E} = \mathbf{j} \end{cases}$$

- *sur* Γ,

$$\text{(II.7.27)} \ -|\sqrt{\varepsilon}|(\mathbf{E} \wedge \nu) + |\sqrt{\mu}|((\mathbf{H} \wedge \nu) \wedge \nu) = Q \quad (|\sqrt{\varepsilon}|(\mathbf{E} \wedge \nu) + |\sqrt{\mu}|((\mathbf{H} \wedge \nu) \wedge \nu)) + g$$

▭

Théorème 11 (Existence et unicité à coefficients constants.) *Soit Ω un domaine de $\mathbb{R}^3$ borné de frontière Γ assez régulière (par exemple C^1). Soit $(\mathbf{m}, \mathbf{j}) \in (\mathrm{H}(\mathrm{div}_0, \Omega))^2$ et $g \in \mathrm{H}^{-1/2}(\mathrm{div}, \Gamma)$. Les fonctions ε et μ sont des constantes qui vérifient les hypothèses (II.7.1) et (II.7.2). On suppose $||Q||_\Gamma < 1$. Alors, le problème 2 admet une solution unique. Si de plus $g \in L_t^2(\Gamma)$ et $Q \neq 1$ ou -1, (par exemple $Q = 0$) alors $(\mathbf{E}, \mathbf{H}) \in (\tilde{\mathcal{H}}(\Omega, \Gamma))^2$ [24].*

Remarque 32 Si $\Omega \in C^{1,1}$ et si $g \in \mathrm{H}_t^{1/2}(\Gamma)$ on a la solution dans $H^1(\Omega)$.

Citons les résultats de Mitrea [46] qui montre un résultat de régularité L^2 sur la frontière pour le problème de Maxwell sans terme source (i.e. $(\mathbf{m}, \mathbf{j}) = (0, 0)$) avec diffraction sur un obstacle ($Q = -1$ sur $\partial\Omega$) pour une fonction de bord $g \in L_t^2(\partial\Omega)$ de divergence surfacique $L^2(\partial\Omega)$ dans le cas très général où la frontière $\partial\Omega$ est lipschitzienne.

Nous définissons le problème homogène adjoint suivant, dans l'hypothèse où ε et μ sont constants sur Ω (domaine qui sera, en pratique, une maille Ω_k où l'hypothèse ε et μ constants sera naturellement vérifiée) :

Problème 3 **:** *trouver* $(\mathbf{E}, \mathbf{H}) \in (\mathrm{H}(\mathrm{rot}, \Omega) \cap \mathrm{H}(\mathrm{div}_0, \Omega))^2$ *tel que*

– *dans* Ω,

$$(\text{II.7.28}) \qquad \left\{\begin{array}{l} \nabla\wedge \mathbf{E} - i\omega\overline{\mu}\mathbf{H} = 0 \\ \nabla\wedge \mathbf{H} + i\omega\overline{\varepsilon}\mathbf{E} = 0 \end{array}\right.$$

– *et sur* $\partial\Omega$,

$$(\text{II.7.29}) \qquad -|\sqrt{\varepsilon}|(\mathbf{E}\wedge\nu) + |\sqrt{\mu}|((\mathbf{H}\wedge\nu)\wedge\nu) = g$$

où $g \in L^2_t(\partial\Omega)$. ▭

Corollaire 7 *Sous les mêmes hypothèses que celles du théorème 11, le problème 3 admet une solution unique dans* $(\mathrm{H}(\mathrm{rot},\Omega)\cap\mathrm{H}(\mathrm{div}_0,\Omega))^2$*. Si* $g \in L^2_t(\partial\Omega)$ *et* $Q \neq 1$ *ou* -1*, alors* $(\mathbf{E},\mathbf{H}) \in (\tilde{\mathcal{H}}(\Omega,\Gamma))^2$.

Donnons une généralisation du théorème d'existence au cas des coefficients variables. Pour la preuve nous renvoyons le lecteur à [1]. Le lecteur constatera que le résultat est valable pour des perméabilité et permittivité complexes ou tenseurs symétriques bornés définis positifs.

Théorème 12 (Extension aux coefficients constants par morceaux) *Sous les hypothèses du théorème 11 à la variante près que les coefficients* ε *et* μ *sont constants sur deux parties disjointes* Ω_v *et* Ω_m *de* Ω *dont l'interface est assez régulière (figure II.7.1), et pour* $g \in \mathrm{H}^{-1/2}(\mathrm{div},\Gamma)$*, on a l'existence et l'unicité de la solution du problème,*

$$(\text{II.7.30})\left\{\begin{array}{ll} \nabla\wedge \mathbf{E} - i\omega\mu\mathbf{H} = -\mathbf{m} & \textit{dans } \Omega \\ \nabla\wedge \mathbf{H} + i\omega\varepsilon\mathbf{E} = \mathbf{j} & \textit{dans } \Omega \\ -|\sqrt{\varepsilon}|(\mathbf{E}\wedge\nu) + |\sqrt{\mu}|((\mathbf{H}\wedge\nu)\wedge\nu) = Q \quad (|\sqrt{\varepsilon}|(\mathbf{E}\wedge\nu) + |\sqrt{\mu}|((\mathbf{H}\wedge\nu)\wedge\nu)) + g & (\Gamma)\ , \end{array}\right.$$

avec les relations (II.7.19) dans l'espace des fonctions $\mathrm{H}(\mathrm{rot},\Omega)$*. Si* $g \in L^2_t(\Gamma)$ *et* $Q \neq 1$ *ou* -1 *les traces de* $\mathbf{E}$ *et* $\mathbf{H}$ *sur* Γ_m *et* Γ_v *seront* $L^2_t(\Gamma_m)$ *et* $L^2_t(\Gamma_v)$*. Par abus de langage, cet espace sera noté* $\mathcal{H}(\Omega,\Gamma)$ *(II.7.25).*

Remarque 33 On déduit du fait que $(\mathbf{E},\mathbf{H}) \in (\mathrm{H}(\mathrm{rot},\Omega))^2$ et que $(\mathbf{m}_{|\Omega_v},\mathbf{j}_{|\Omega_v}) \in (\mathrm{H}(\mathrm{div}_0,\Omega_v))^2$ que $(\mathbf{E}_{|\Omega_v},\mathbf{H}_{|\Omega_v}) \in (\mathrm{H}(\mathrm{div}_0,\Omega_v))^2$ et de même sur Ω_m.

Nous n'effectuons pas la preuve du théorème 11 qui n'est pas l'objet essentiel de notre exposé. Le point crucial de la preuve effectuée dans [24] dans le cas de la propagation libre dans le vide est la mise sous la forme variationnelle classique d'un opérateur coercif et d'une perturbation dont on montre la compacité par les théorèmes d'injection continue de $\mathcal{H}(\Omega,\Gamma)$ dans $(H^{1/2}(\Omega))^3$ et l'injection compacte de $(H^{1/2}(\Omega))^3$ dans $(L^2(\Omega))^3$ qui nous mène à l'alternative de Fredholm. On montre l'unicité à l'aide de la condition de divergence comprise dans l'espace fonctionnel de travail $H(\mathrm{div}_0,\Omega)$. Notons que l'hypothèse "frontière Γ assez régulière" peut signifier lipschitzienne [21], cadre qui nous sera utile dans nos applications. Le théorème 12 ([1]) est essentiel aussi pour toute application numérique dans un cadre industriel où les caractéristiques du milieu ne peuvent être constantes. Nous montrons ici seulement la façon d'obtenir la formulation variationnelle dans le cadre du théorème 11 et du corollaire 7.

Lemme 12 *Le problème 2 s'écrit sous la forme variationnelle sur* $\mathbf{E}$ *dans* $\mathcal{H}(\Omega,\Gamma)$ *:*

$$(\text{II.7.31}) \qquad \left\{\begin{array}{l} \displaystyle\int_\Omega \frac{1}{\mu}(\nabla\wedge\mathbf{E})(\overline{\nabla\wedge\mathbf{E}'}) - i\omega\int_\Gamma \zeta((\mathbf{E}\wedge\nu)\wedge\nu)(\overline{(\mathbf{E}'\wedge\nu)\wedge\nu}) - \omega^2\int_\Omega \varepsilon\mathbf{E}\overline{\mathbf{E}'} \\ \displaystyle = \int_\Omega f'\overline{\mathbf{E}'} + \int_\Gamma g'\,\overline{((\mathbf{E}'\wedge\nu)\wedge\nu)} \end{array}\right.$$

pour tout $\mathbf{E}' \in \mathcal{H}(\Omega,\Gamma)$ *en posant*

$$(\text{II.7.32}) \qquad \left\{\begin{array}{l} f' = +i\omega\mathbf{j} - \nabla\wedge(\dfrac{\mathbf{m}}{\mu}) \\ g' = \dfrac{i\omega}{1-Q}\dfrac{1}{|\sqrt{\mu}|}g\wedge\nu + \dfrac{\mathbf{m}\wedge\nu}{\mu} \end{array}\right.$$

avec $\zeta = \dfrac{1+Q}{1-Q}\dfrac{|\sqrt{\varepsilon}|}{|\sqrt{\mu}|}$.

Preuve. De la première relation de l'équation (II.7.26) on a

$$\mathbf{H} = \frac{\nabla\wedge\mathbf{E} + \mathbf{m}}{i\omega\mu}\ , \tag{II.7.33}$$

ce qui permet de supprimer le champ $\mathbf{H}$ dans les relations (II.7.26) et (II.7.27). On obtient les deux relations, dans Ω,

$$\nabla\wedge(\frac{1}{i\omega\mu}\nabla\wedge\mathbf{E}) + i\omega\varepsilon\mathbf{E} = +\mathbf{j} - \nabla\wedge(\frac{\mathbf{m}}{i\omega\mu}) \tag{II.7.34}$$

et sur Γ,

$$\frac{1}{i\omega\mu}((\nabla\wedge\mathbf{E})\wedge\nu)\wedge\nu = \frac{1+Q}{1-Q}\frac{|\sqrt{\varepsilon}|}{|\sqrt{\mu}|}(\mathbf{E}\wedge\nu) + \frac{1}{1-Q}\frac{1}{|\sqrt{\mu}|}g - \frac{1}{i\omega\mu}(\mathbf{m}\wedge\nu)\wedge\nu\ . \tag{II.7.35}$$

Rappelons la formule d'intégration par partie (formule de Green pour des champs $\mathrm{H}(\mathrm{rot},\Omega)$) pour des champs assez réguliers :

$$\int_\Omega (\nabla\wedge\mathbf{F})\mathbf{G} - (\nabla\wedge\mathbf{G})\mathbf{F} = \int_\Gamma (\mathbf{F}\wedge\nu)((\mathbf{G}\wedge\nu)\wedge\nu) \tag{II.7.36}$$

d'où en remplaçant dans (II.7.36) la fonction $\mathbf{F}$ par $\frac{1}{i\omega\mu}\nabla\wedge\mathbf{E}$ et la fonction $\mathbf{G}$ par $\overline{\mathbf{E}'}$, on a

$$\int_\Omega (\nabla\wedge(\frac{1}{i\omega\mu}\nabla\wedge\mathbf{E}))\overline{\mathbf{E}'} = \int_\Omega \frac{1}{i\omega\mu}(\nabla\wedge\mathbf{E})(\overline{\nabla\wedge\mathbf{E}'}) + \int_\Gamma \frac{1}{i\omega\mu}((\nabla\wedge\mathbf{E})\wedge\nu)(\overline{(\mathbf{E}'\wedge\nu)\wedge\nu}) \tag{II.7.37}$$

De plus, remarquons que

$$(\nabla\wedge\mathbf{E})\wedge\nu = -(((\nabla\wedge\mathbf{E})\wedge\nu)\wedge\nu)\wedge\nu\ , \tag{II.7.38}$$

donc, en intégrant la formule (II.7.34) sur Ω contre une fonction test $\mathbf{E}^{'}$ on déduit de (II.7.37) puis (II.7.38) que

$$\begin{aligned}&\int_\Omega \frac{1}{i\omega\mu}(\nabla\wedge\mathbf{E})(\overline{\nabla\wedge\mathbf{E}'}) - \int_\Gamma \frac{1}{i\omega\mu}(((\nabla\wedge\mathbf{E})\wedge\nu)\wedge\nu)\wedge\nu(\overline{(\mathbf{E}'\wedge\nu)\wedge\nu})\\&+ i\omega\int_\Omega \varepsilon\mathbf{E}\overline{\mathbf{E}'} = \int_\Omega \mathbf{j}\overline{\mathbf{E}'} - \int_\Omega \nabla\wedge(\frac{\mathbf{m}}{i\omega\mu})\overline{\mathbf{E}'}\ .\end{aligned} \tag{II.7.39}$$

Puis, en multipliant (II.7.35) à droite par $\wedge\nu$, on a

$$\begin{aligned}&\frac{1}{i\omega\mu}(((\nabla\wedge\mathbf{E})\wedge\nu)\wedge\nu)\wedge\nu\\&= \frac{1+Q}{1-Q}\frac{|\sqrt{\varepsilon}|}{|\sqrt{\mu}|}((\mathbf{E}\wedge\nu)\wedge\nu) + \frac{1}{1-Q}\frac{1}{|\sqrt{\mu}|}g\wedge\nu + \frac{1}{i\omega\mu}\mathbf{m}\wedge\nu\ .\end{aligned} \tag{II.7.40}$$

A l'aide de (II.7.39) et (II.7.40), on tire (rappelons que l'on pose $\zeta = \frac{1+Q}{1-Q}\frac{|\sqrt{\varepsilon}|}{|\sqrt{\mu}|}$)

$$\begin{aligned}&\int_\Omega \frac{1}{i\omega\mu}(\nabla\wedge\mathbf{E})(\overline{\nabla\wedge\mathbf{E}'}) - \int_\Gamma \zeta((\mathbf{E}\wedge\nu)\wedge\nu)(\overline{(\mathbf{E}'\wedge\nu)\wedge\nu}) + i\omega\int_\Omega \varepsilon\mathbf{E}\overline{\mathbf{E}'}\\&= \int_\Omega \mathbf{j}\overline{\mathbf{E}'} - \int_\Omega \nabla\wedge(\frac{\mathbf{m}}{i\omega\mu})\overline{\mathbf{E}'} + \int_\Gamma\left(\frac{1}{1-Q}\frac{1}{|\sqrt{\mu}|}g\wedge\nu + \frac{\mathbf{m}\wedge\nu}{i\omega\mu}\right)(\overline{(\mathbf{E}'\wedge\nu)\wedge\nu})\end{aligned} \tag{II.7.41}$$

Le problème se réduit alors à l'étude de la forme sesquilinéaire

$$\int_\Omega \frac{1}{\mu}(\nabla\wedge\mathbf{E})(\overline{\nabla\wedge\mathbf{E}'}) - i\omega\int_\Gamma \zeta((\mathbf{E}\wedge\nu)\wedge\nu)(\overline{(\mathbf{E}'\wedge\nu)\wedge\nu}) - \omega^2\int_\Omega \varepsilon\mathbf{E}\overline{\mathbf{E}'}\ . \tag{II.7.42}$$

ou, sous une autre forme,

$$\int_\Omega \frac{i\overline{\mu}}{|\mu|^2}(\nabla\wedge\mathbf{E})(\overline{\nabla\wedge\mathbf{E}'}) - \omega\int_\Gamma \zeta((\mathbf{E}\wedge\nu)\wedge\nu)(\overline{(\mathbf{E}'\wedge\nu)\wedge\nu}) - \omega^2\int_\Omega i\varepsilon\mathbf{E}\overline{\mathbf{E}'}\ . \tag{II.7.43}$$

Remarquons que si ε et μ sont dissipatifs constants dans Ω et si $Q \in]-1,1[$ (ce qui assure $\zeta \geq 0$) alors la forme sesquilinéaire (II.7.43) est coercive ce qui permet de conclure à l'aide du théorème de Lax-Milgram seul.
$\square$

Lemme 13 *Le problème 3 s'écrit sous la forme variationnelle sur* $\mathbf{E}$ *dans* $\mathcal{H}(\Omega,\Gamma)$ *:*

$$(II.7.44)\qquad \begin{cases} \displaystyle\int_\Omega \frac{1}{\overline{\mu}}(\nabla\wedge\mathbf{E})(\overline{\nabla\wedge\mathbf{E}'}) + i\omega\int_\Gamma \frac{|\sqrt{\varepsilon}|}{|\sqrt{\mu}|}((\mathbf{E}\wedge\nu)\wedge\nu)(\overline{(\mathbf{E}'\wedge\nu)\wedge\nu}) - \omega^2\int_\Omega \overline{\varepsilon}\mathbf{E}\overline{\mathbf{E}'} \\ \displaystyle = i\omega\int_\Gamma \frac{1}{|\sqrt{\mu}|}(g\wedge\nu)(\overline{(\mathbf{E}'\wedge\nu)\wedge\nu}) \end{cases}$$

pour tout $\mathbf{E}^{'} \in \mathcal{H}(\Omega,\Gamma)$.

Preuve. On a

$$(II.7.45)\qquad \begin{cases} \mathbf{H} = \dfrac{1}{i\omega\overline{\mu}}\nabla\wedge\mathbf{E} \\ \nabla\wedge(\dfrac{1}{\overline{\mu}}\nabla\wedge\mathbf{E}) - \omega^2\overline{\varepsilon}\mathbf{E} = 0 \\ +\dfrac{1}{i\omega\overline{\mu}}(((\nabla\wedge\mathbf{E})\wedge\nu)\wedge\nu) = \dfrac{1}{|\sqrt{\mu}|}g - \dfrac{|\sqrt{\varepsilon}|}{|\sqrt{\mu}|}(\mathbf{E}\wedge\nu)\ . \end{cases}$$

d'où

$$(II.7.46)\qquad \begin{aligned} &\int_\Omega \frac{1}{\overline{\mu}}(\nabla\wedge\mathbf{E})(\overline{\nabla\wedge\mathbf{E}'}) + i\omega\int_\Gamma \frac{|\sqrt{\varepsilon}|}{|\sqrt{\mu}|}((\mathbf{E}\wedge\nu)\wedge\nu)(\overline{(\mathbf{E}'\wedge\nu)\wedge\nu}) - \omega^2\int_\Omega \overline{\varepsilon}\mathbf{E}\overline{\mathbf{E}'} \\ &= i\omega\int_\Gamma \frac{1}{|\sqrt{\mu}|}g\wedge\nu(\overline{(\mathbf{E}'\wedge\nu)\wedge\nu}) \end{aligned}$$

qui est le problème adjoint homogène (i.e. $(\mathbf{m},\mathbf{j}) = (0,0)$) de (II.7.31). $\square$

II.7.1.3 Découplage des équations de Maxwell.

On considère les équations de Maxwell sur le couple $(\mathbf{E},\mathbf{H})$ dans Ω,

$$(II.7.47)\qquad \begin{cases} \nabla\wedge\mathbf{E} - i\omega\mu\mathbf{H} = -\mathbf{m} \\ \nabla\wedge\mathbf{H} + i\omega\varepsilon\mathbf{E} = \mathbf{j}\ . \end{cases}$$

Posons

$$(II.7.48)\qquad \begin{cases} \mathbf{G} = \sqrt{\varepsilon}\mathbf{E} + i\sqrt{\mu}\mathbf{H} \\ \mathbf{F} = \sqrt{\varepsilon}\mathbf{E} - i\sqrt{\mu}\mathbf{H}\ . \end{cases}$$

Un calcul élémentaire montre que le problème (II.7.47) est équivalent à

$$(II.7.49)\qquad \begin{cases} \nabla\wedge\mathbf{G} - \omega\sqrt{\varepsilon\mu}\mathbf{G} = -\sqrt{\varepsilon}\mathbf{m} + i\sqrt{\mu}\mathbf{j} \\ \nabla\wedge\mathbf{F} + \omega\sqrt{\varepsilon\mu}\mathbf{F} = -\sqrt{\varepsilon}\mathbf{m} - i\sqrt{\mu}\mathbf{j} \end{cases}$$

avec

$$(II.7.50)\qquad \begin{cases} \mathbf{E} = \dfrac{(\mathbf{G}+\mathbf{F})}{2\sqrt{\varepsilon}} \\ \mathbf{H} = \dfrac{(\mathbf{G}-\mathbf{F})}{2i\sqrt{\mu}}\ . \end{cases}$$

Autrement dit, les champs $\mathbf{G}$ et $\mathbf{F}$ sont solutions d'équations différentielles découplées.

Dans un matériau réel, la condition aux limites absorbante est

$$(II.7.51)\qquad -\sqrt{\varepsilon}(\mathbf{E}\wedge\nu) + \sqrt{\mu}((\mathbf{H}\wedge\nu)\wedge\nu) = g\ .$$

Substituons le couple $(\mathbf{G}, \mathbf{F})$ au couple $(\mathbf{E}, \mathbf{H})$. On obtient, sur Γ,

$$-(\mathbf{G} \wedge \nu + \mathbf{F} \wedge \nu) - i\left((\mathbf{G} \wedge \nu) \wedge \nu - (\mathbf{F} \wedge \nu) \wedge \nu\right) = 2g \ . \tag{II.7.52}$$

On écrit les composantes de $\mathbf{G}$ et $\mathbf{F}$ dans un repère orthonormé direct dont le troisième axe est ν,

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \mathbf{G}_3 \end{pmatrix} \text{ et } \mathbf{F} = \begin{pmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \\ \mathbf{F}_3 \end{pmatrix} \tag{II.7.53}$$

La relation (II.7.52) devient

$$\begin{cases} -(\mathbf{G}_2 + \mathbf{F}_2) + i(\mathbf{G}_1 - \mathbf{F}_1) = 2g & \text{sur } \Gamma \\ -(\mathbf{G}_1 + \mathbf{F}_1) - i(\mathbf{G}_2 - \mathbf{F}_2) = -2g & \text{sur } \Gamma \end{cases} \tag{II.7.54}$$

En multipliant la deuxième équation par i et en sommant avec la première, puis en multipliant la première par i et en sommant avec la deuxième, on a les équations découplées

$$\begin{cases} -i\mathbf{F}_1 - \mathbf{F}_2 = (1-i)g \\ \mathbf{G}_1 + i\mathbf{G}_2 = (i+1)g \end{cases} \tag{II.7.55}$$

Pour résumer, nous avons à résoudre les systèmes découplés

$$\begin{cases} \nabla\wedge \mathbf{G} - \omega\sqrt{\varepsilon\mu}\mathbf{G} = \sqrt{\varepsilon}\mathbf{m} + i\sqrt{\mu}\mathbf{j} & \text{dans } \Omega \\ \mathbf{G}_1 + i\mathbf{G}_2 = (i+1)g & \text{sur } \Gamma \end{cases} \tag{II.7.56}$$

et

$$\begin{cases} \nabla\wedge \mathbf{F} + \omega\sqrt{\varepsilon\mu}\mathbf{F} = \sqrt{\varepsilon}\mathbf{m} - i\sqrt{\mu}\mathbf{j} & \text{dans } \Omega \\ -i\mathbf{F}_1 - \mathbf{F}_2 = (1-i)g & \text{sur } \Gamma \end{cases} \tag{II.7.57}$$

II.7.2 La formulation ultra-faible et ses propriétés.

Le but de cette section est de présenter la formulation variationnelle ultra-faible appliquée aux problèmes d'électromagnétisme, de démontrer l'existence et l'unicité de la solution et l'équivalence entre notre formulation et la formulation classique. Le plan de travail est donc le suivant.

1. Obtention de la formulation en montrant qu'une solution du problème de Maxwell vérifie la formulation ultra-faible. C'est l'objet du théorème 13.
2. Inversement, on montre l'existence d'une solution de la formulation ultra-faible et que l'on obtient les champs électrique et magnétique par un opérateur de relèvement (théorème 14).
3. Les deux points précédents montrent l'existence et l'unicité du problème ultra-faible et l'équivalence avec la formulation classique. Nous définissons les opérateurs de notre formulation et nous étudions leurs propriétés. Le résultat fondamental (section II.7.2.6) est la décomposition de l'opérateur de la formulation en l'identité moins un opérateur de norme inférieure à 1.

II.7.2.1 Partition du domaine tridimensionnel et notations.

Comme pour le problème de Helmholtz bidimensionnel, nous réalisons une partition du domaine tridimensionnel Ω.

Notation 2 (Rappels et notations des faces de bord Σ_{kk}) *Soit Ω un domaine borné de $\mathbb{R}^3$ partitionné en K domaines Ω_k lipschitziens. Pour un élément Ω_k donné, nous considérons un de ses voisins Ω_j. L'interface entre Ω_k et Ω_j est notée Σ_{kj} comme pour Helmholtz. Considérons un élément Ω_k qui possède une frontière libre, c'est-à-dire sur laquelle il n'y a pas de voisin Ω_j. Nous avons noté Γ_k une telle frontière dans l'étude du problème de Helmholtz. Dans l'étude du problème de Maxwell avec matériaux,* *nous verrons qu'une telle notation n'est pas très judicieuse. En effet, d'une part elle risque d'introduire une confusion avec les notations classiques définissant Γ_k comme étant le bord de la maille Ω_k, d'autre part cette notation crée une asymétrie entre les types de frontières alors qu'elles jouent le même rôle lorsque l'on s'intéresse simplement au bord de Ω_k. C'est pourquoi nous appellerons dorénavant une telle frontière Σ_{kk}, ce qui permet d'écrire que*

(II.7.58) $$\partial\Omega_k = \bigcup_{j(k),\ \Omega_j \text{ voisin de } \Omega_k,\ \text{ou } j=k \text{ sur une face frontière}} \Sigma_{kj}$$

où $j(k)$ décrit l'ensemble des voisins de Ω_k qui sont des éléments de la partition de Ω, plus la frontière libre éventuelle Σ_{kk}. Par la suite, lorsqu'il n'y aura pas de risque de confusion, $j(k)$ sera noté j. La figure II.7.2 résume ces notations : il s'agit d'un exemple de partition de domaine sous la forme d'un maillage en éléments hexaédriques dont on effectue une coupe dans le plan de la feuille.

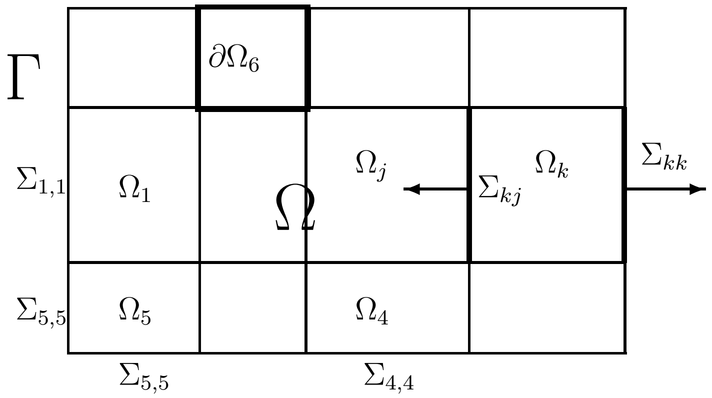


FIG. II.7.2 – Illustration de la définition de Σ_{kk}.

Pour un élément Ω_k quelconque nous notons :
- ν_k une normale à l'élément Ω_k,
- $\nu_{k,j}$ une normale à l'interface $\Sigma_{k,j}$.

Définissons maintenant les fonctions ε_k, ε_{kk} et ε_{kj} introduites dans la formulation variationnelle à partir de la perméabilité ε. Nous définirons de la même façon les fonctions μ_k, μ_{kk} et μ_{kj}.

1. La fonction ε_k est la restriction de ε à l'élément Ω_k,

   (II.7.59) $$\varepsilon_k = \varepsilon_{|\Omega_k} .$$

2. La fonction réelle ε_{kj} est la valeur absolue de la moyenne géométrique à l'interface Σ_{kj} de la fonction ε sur Ω_k et sur Ω_j,

   (II.7.60) $$\varepsilon_{kj} = |\sqrt{\varepsilon_k \varepsilon_j}| .$$

3. Sur une face frontière Σ_{kk}, la fonction réelle ε_{kk} vaut d'après (II.7.60) :

   (II.7.61) $$\varepsilon_{kk} = |\varepsilon_k| .$$

   Ne pas confondre ε_{kk} et ε_k, ces fonctions sont égales sur Σ_{kk} si et seulement si ε_k est réelle (positive par hypothèse sur ε, cf (II.7.1)).

Remarque 34 Notons l'intérêt du choix de la moyenne géométrique plutôt que de la moyenne arithmétique pour la définition des fonctions approchées ε_{kk} et μ_{kk} dans le cas d'une discontinuité des permittivité et perméabilité entre Ω_k et Ω_j :

(II.7.62) $$\begin{cases} |\varepsilon_k \mu_k| = 1 \\ |\varepsilon_j \mu_j| = 1 \end{cases} \Longrightarrow \varepsilon_{kj}\mu_{kj} = 1 .$$

D'autre part, si ε_k est réelle, alors

$$\frac{\varepsilon_{kj}}{\sqrt{\varepsilon_k}} = |\sqrt{\varepsilon_j}| \tag{II.7.63}$$

ne dépend plus de k. Enfin, les fonctions réelles ε_{kj} et μ_{kj} vérifient

$$\left| \begin{array}{l} \varepsilon_{kj} = \varepsilon_{jk} \\ \mu_{kj} = \mu_{jk} \end{array} \right. \tag{II.7.64}$$

Comme pour le problème de Helmholtz, définissons l'espace fonctionnel V de la formulation ultra-faible. Notons toujours que cet espace dépend du maillage, mais que ce n'est pas un espace de discrétisation de dimension finie.

Définition 6 *L'espace de travail V est l'espace de Hilbert produit des espaces de Hilbert $L^2_t()$ définis en II.7.23 :*

$$V = \prod_{k=1}^{K} L^2_t(\partial\Omega_k) \ , \tag{II.7.65}$$

ou, de façon équivalente, par,

$$V = \prod_{k=1}^{K} \prod_{j(k)} L^2_t(\Sigma_{kj}) \ .$$

L'espace de travail est muni du produit scalaire naturel

$$(\mathcal{X}, \mathcal{Y}) = \sum_k \int_{\partial\Omega_k} \mathcal{X}_{|\partial\Omega_k} \overline{\mathcal{Y}_{|\partial\Omega_k}} \tag{II.7.66}$$

qui définit la norme $||.||_V$ et la norme induite d'un opérateur $A \in V$ par :

$$||A|| = \sup_{x \neq 0} \frac{||Ax||_V}{||x||_V} \ . \tag{II.7.67}$$

II.7.2.2 Relation vérifiée par la solution des équations de Maxwell.

Dans toute cette étude du problème de Maxwell (1 p. 79), nous ferons l'hypothèse H4 de régularité des champs $(\mathbf{E}, \mathbf{H})$.

Hypothèse 4 *On utilise la partition du domaine Ω de sorte que ε et μ sont constants sur tout élément Ω_k. Soit $(\mathbf{E}, \mathbf{H})$ une solution du problème de Maxwell (1 p. 79) à coefficients constants par morceaux qui vérifie l'hypothèse de régularité $(\mathbf{E}, \mathbf{H}) \in (\mathcal{H}(\Omega, \Gamma))^2$. L'espace $\mathcal{H}(\Omega, \Gamma)$ est donc l'espace des fonctions $\tilde{\mathcal{H}}$ (II.7.24) sur tous les sous-domaines de Ω où ε et μ sont constants et des fonctions vérifiant les relations de continuité des traces tangentielles des champs (II.7.19).*

Sous l'hypothèse H4 ci-dessus, nous pouvons définir une nouvelle inconnue notée $\mathcal{X}$.

Définition 7 *On définit $\mathcal{X}$ par ses restrictions $\mathcal{X}_k = (\mathcal{X})_{|\partial\Omega_k}$ sur $\partial\Omega_k$ et ses restrictions sur $\Sigma_{k,j}$ ($j = j(k)$ désigne l'indice d'un voisin Ω_j à Ω_k) :*

$$(\mathcal{X}_k)_{|\Sigma_{k,j}} = +\sqrt{\varepsilon_{kj}}(\mathbf{E}_k \wedge \nu_k) + \sqrt{\mu_{kj}}((\mathbf{H}_k \wedge \nu_k) \wedge \nu_k) \ ,$$

où $\sqrt{\varepsilon_{kj}}$ est défini par (II.7.60). Sous l'hypothèse 4, on a bien $\mathcal{X}_k \in L^2_t(\partial\Omega_k)$ et $\mathcal{X} \in V$.

Notation 3 *On indice par k les quantités relatives à un élément Ω_k. En particulier, nous notons $\mathcal{X}_k$ les restrictions de $\mathcal{X}$ à $\partial\Omega_k = \bigcup_{j(k)} \Sigma_{kj}$ vues de Ω_k vers Ω_j. Nous notons aussi $\mathbf{E}_k = (\mathbf{E})_{|\Omega_k}$.*

Théorème 13 *On suppose que l'hypothèse 4 est vérifiée. Alors, pour tout* $(\mathbf{E}^{'},\mathbf{H}^{'})$ *dont les restrictions* $(\mathbf{E}^{'}{}_k,\mathbf{H}^{'}{}_k) \in (\tilde{\mathcal{H}}(\Omega_k,\partial\Omega_k))^2$ *vérifient*

$$\text{(II.7.68)} \qquad \begin{cases} \nabla\wedge \mathbf{E}^{'}{}_k - i\omega\overline{\mu}_k\mathbf{H}^{'}{}_k = 0 \ \textit{dans}\ \Omega_k \\ \nabla\wedge \mathbf{H}^{'}{}_k + i\omega\overline{\varepsilon}_k\mathbf{E}^{'}{}_k = 0 \ \textit{dans}\ \Omega_k \\ \sqrt{\varepsilon_{kj}}(\mathbf{E}^{'}{}_k\wedge\nu_k) + \sqrt{\mu_{kj}}((\mathbf{H}^{'}{}_k\wedge\nu_k)\wedge\nu_k) \in L^2_t(\partial\Omega_k)\ , \end{cases}$$

on a,

$$\text{(II.7.69)} \quad \begin{aligned} &\sum_{k,j}\int_{\Sigma_{kj}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}\mathcal{X}_k\ \left(\overline{\sqrt{\varepsilon_{kj}}\mathbf{E}'{}_k\wedge\nu_k + \sqrt{\mu_{kj}}((\mathbf{H}'\wedge\nu)\wedge\nu)}\right)\\ &-\sum_{k,j\neq k}\int_{\Sigma_{kj}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}\mathcal{X}_j\ \left(\overline{-\sqrt{\varepsilon_{kj}}\mathbf{E}'{}_k\wedge\nu_k + \sqrt{\mu_{kj}}((\mathbf{H}'{}_k\wedge\nu_k)\wedge\nu_k)}\right)\\ &-\sum_{k}\int_{\Sigma_{kk}} \frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}}Q_k\mathcal{X}_k\ \left(\overline{-\sqrt{\varepsilon_{kk}}\mathbf{E}'{}_k\wedge\nu_k + \sqrt{\mu_{kk}}((\mathbf{H}'{}_k\wedge\nu_k)\wedge\nu_k)}\right)\\ &= -2\sum_k\int_{\Omega_k}\mathbf{m}\overline{\mathbf{H}'{}_k} + \mathbf{j}\overline{\mathbf{E}'{}_k} + \sum_k\int_{\Sigma_{kk}}\frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}}g\ \left(\overline{-\sqrt{\varepsilon_{kk}}\mathbf{E}'{}_k\wedge\nu_k + \sqrt{\mu_{kk}}((\mathbf{H}'{}_k\wedge\nu_k)\wedge\nu_k)}\right)\ . \end{aligned}$$

Remarque 35 La relation (II.7.69) s'écrit aussi sous la forme

$$\text{(II.7.70)} \quad \begin{aligned} &\sum_{k,j}\int_{\Sigma_{kj}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}(\sqrt{\varepsilon_{kj}}\mathbf{E}_k\wedge\nu_k + \sqrt{\mu_{kj}}((\mathbf{H}_k\wedge\nu_k)\wedge\nu_k))\ \left(\overline{\sqrt{\varepsilon_{kj}}\mathbf{E}'{}_k\wedge\nu_k + \sqrt{\mu_{kj}}((\mathbf{H}'{}_k\wedge\nu_k)\wedge\nu_k)}\right)\\ &-\sum_{k,j\neq k}\int_{\Sigma_{kj}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}(-\sqrt{\varepsilon_{kj}}\mathbf{E}_j\wedge\nu_k + \sqrt{\mu_{kj}}((\mathbf{H}_j\wedge\nu_k)\wedge\nu_k))\ \left(\overline{-\sqrt{\varepsilon_{kj}}\mathbf{E}'{}_k\wedge\nu_k + \sqrt{\mu_{kj}}((\mathbf{H}'{}_k\wedge\nu_k)\wedge\nu_k)}\right)\\ &-\sum_{k}\int_{\Sigma_{kk}} \frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}}t(\sqrt{\varepsilon_{kk}}\mathbf{E}_k\wedge\nu_k + \sqrt{\mu_{kk}}((\mathbf{H}_k\wedge\nu_k)\wedge\nu_k))\ \left(\overline{-\sqrt{\varepsilon_{kk}}\mathbf{E}'{}_k\wedge\nu_k + \sqrt{\mu_{kk}}((\mathbf{H}'{}_k\wedge\nu_k)\wedge\nu_k)}\right)\\ &= -2\sum_k\int_{\Omega_k}\mathbf{m}\overline{\mathbf{H}'{}_k} + \mathbf{j}\overline{\mathbf{E}'{}_k} + \sum_k\int_{\Sigma_{kk}}\frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}}g\ \left(\overline{-\sqrt{\varepsilon_{kk}}\mathbf{E}'{}_k\wedge\nu_k + \sqrt{\mu_{kk}}((\mathbf{H}'{}_k\wedge\nu_k)\wedge\nu_k)}\right)\ . \end{aligned}$$

Preuve. En développant le premier terme sous l'intégrale dans l'égalité (II.7.70), nous avons :

$$\text{(II.7.71)} \quad \begin{aligned} &\frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}(\sqrt{\varepsilon_{kj}}\mathbf{E}_k\wedge\nu_k + \sqrt{\mu_{kj}}((\mathbf{H}_k\wedge\nu_k)\wedge\nu_k))\ \left(\overline{\sqrt{\varepsilon_{kj}}\mathbf{E}'{}_k\wedge\nu_k + \sqrt{\mu_{kj}}((\mathbf{H}'{}_k\wedge\nu_k)\wedge\nu_k)}\right)\\ =\ &\frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}(-\sqrt{\varepsilon_{kj}}\mathbf{E}_k\wedge\nu_k + \sqrt{\mu_{kj}}((\mathbf{H}_k\wedge\nu_k)\wedge\nu_k))\ \left(\overline{-\sqrt{\varepsilon_{kj}}\mathbf{E}'{}_k\wedge\nu_k + \sqrt{\mu_{kj}}((\mathbf{H}'{}_k\wedge\nu_k)\wedge\nu_k)}\right)\\ +\ &2\frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}\left(\sqrt{\varepsilon_{kj}}(\mathbf{E}_k\wedge\nu_k)\,\overline{\sqrt{\mu_{kj}}((\mathbf{H}'{}_k\wedge\nu_k)\wedge\nu_k)} + (\overline{\sqrt{\varepsilon_{kj}}\mathbf{E}'{}_k\wedge\nu_k})\,\sqrt{\mu_{kj}}((\mathbf{H}_k\wedge\nu_k)\wedge\nu_k)\right). \end{aligned}$$

On remarque que

$$\text{(II.7.72)} \quad \left| \begin{aligned} &\frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}\left(\sqrt{\varepsilon_{kj}}(\mathbf{E}_k\wedge\nu_k)\,\overline{\sqrt{\mu_{kj}}((\mathbf{H}'{}_k\wedge\nu_k)\wedge\nu_k)}\right) = \left((\mathbf{E}_k\wedge\nu_k)\,\overline{((\mathbf{H}'{}_k\wedge\nu_k)\wedge\nu_k)}\right)\\ &\frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}\left(+(\overline{\sqrt{\varepsilon_{kj}}\mathbf{E}'{}_k\wedge\nu_k})\,\sqrt{\mu_{kj}}((\mathbf{H}_k\wedge\nu_k)\wedge\nu_k)\right) = \left((\overline{\mathbf{E}'{}_k\wedge\nu_k})\,((\mathbf{H}_k\wedge\nu_k)\wedge\nu_k)\right) \end{aligned}\right.$$

Rappelons la formule classique d'intégration par parties pour des champs de vecteurs de $\mathbb{C}^3$ dans $\mathcal{D}(\overline{\Omega})^3$ ou $\mathrm{H}(\mathrm{rot},\Omega)$ (formule de Green) :

$$\text{(II.7.73)} \qquad \int_{\Omega_k}(\nabla\wedge\mathbf{F})\mathbf{G} - (\nabla\wedge\mathbf{G})\mathbf{F} = \int_{\partial\Omega_k}(\mathbf{F}\wedge\nu)((\mathbf{G}\wedge\nu)\wedge\nu)$$

On a alors

$$\text{(II.7.74)} \qquad \int_{\partial\Omega_k}(\mathbf{E}_k\wedge\nu_k)\,(\overline{((\mathbf{H}'{}_k\wedge\nu_k)\wedge\nu_k)}) = \int_{\Omega_k}(\nabla\wedge\mathbf{E}_k)\overline{\mathbf{H}'{}_k} - (\nabla\wedge\overline{\mathbf{H}'{}_k})\mathbf{E}_k\ ,$$

et

$$\text{(II.7.75)} \qquad \int_{\partial\Omega_k}(\overline{\mathbf{E}'{}_k\wedge\nu_k})\,(((\mathbf{H}_k\wedge\nu_k)\wedge\nu_k)) = \int_{\Omega_k}(\nabla\wedge\overline{\mathbf{E}'{}_k})\mathbf{H}_k - (\nabla\wedge\mathbf{H}_k)\overline{\mathbf{E}'{}_k}\ .$$

Rappelons les équations (II.6.1) et (II.7.68) :

$$\text{(II.7.76)} \qquad \begin{cases} h)\, \nabla\wedge \mathbf{E} - i\omega\mu\mathbf{H} = -\mathbf{m} & \text{dans } \Omega \\ e)\, \nabla\wedge \mathbf{H} + i\omega\varepsilon\mathbf{E} = +\mathbf{j} & \text{dans } \Omega \\ \overline{h})\, \nabla\wedge \overline{\mathbf{E}'_k} + i\omega\mu\overline{\mathbf{H}'_k} = 0 & \text{dans } \Omega_k \\ \overline{e})\, \nabla\wedge \overline{\mathbf{H}'_k} - i\omega\varepsilon\overline{\mathbf{E}'_k} = 0 & \text{dans } \Omega_k \ . \end{cases}$$

Alors, en intégrant les équations (II.7.76) sur Ω_k contre $\overline{\mathbf{H}'_k}$ pour l'équation (II.7.76 h), contre $\mathbf{H}_k$ pour l'équation (II.7.76 $\overline{h}$), contre $\overline{\mathbf{E}'_k}$ pour l'équation (II.7.76 e), et enfin contre $\mathbf{E}_k$ pour l'équation (II.7.76 $\overline{e}$), on a

$$\text{(II.7.77)} \qquad \begin{cases} \displaystyle\int_{\Omega_k} (\nabla\wedge \mathbf{E}_k)\overline{\mathbf{H}'_k} - i\omega \int_{\Omega_k} \mu_k \mathbf{H}_k \overline{\mathbf{H}'_k} = -\int_{\Omega_k} \mathbf{m}\overline{\mathbf{H}'_k} \\ \displaystyle\int_{\Omega_k} (\nabla\wedge \overline{\mathbf{E}'_k})\mathbf{H}_k + i\omega \int_{\Omega_k} \mu_k \mathbf{H}_k \overline{\mathbf{H}'_k} = 0 \end{cases}$$

et

$$\text{(II.7.78)} \qquad \begin{cases} \displaystyle\int_{\Omega_k} (\nabla\wedge \mathbf{H}_k)\overline{\mathbf{E}'_k} + i\omega \int_{\Omega_k} \varepsilon_k \mathbf{E}_k \overline{\mathbf{E}'_k} = \int_{\Omega_k} \mathbf{j}\overline{\mathbf{E}'_k} \\ \displaystyle\int_{\Omega_k} (\nabla\wedge \overline{\mathbf{H}'_k})\mathbf{E}_k - i\omega \int_{\Omega_k} \varepsilon_k \mathbf{E}_k \overline{\mathbf{E}'_k} = 0 \ . \end{cases}$$

D'après (II.7.74) et (II.7.75), on a aussi

$$\text{(II.7.79)} \qquad \begin{aligned} &\int_{\partial\Omega_k} \mathbf{E}_k \wedge \nu_k \left(\overline{(\mathbf{H}'_k \wedge \nu_k) \wedge \nu_k}\right) + \left(\overline{\mathbf{E}'_k \wedge \nu_k}\right) \left((\mathbf{H}_k \wedge \nu_k) \wedge \nu_k\right) \\ &= \int_{\Omega_k} (\nabla\wedge \mathbf{E}_k)\overline{\mathbf{H}'_k} - (\nabla\wedge \overline{\mathbf{H}'_k})\mathbf{E}_k + (\nabla\wedge \overline{\mathbf{E}'_k})\mathbf{H}_k - (\nabla\wedge \mathbf{H}_k)\overline{\mathbf{E}'_k} \\ &= -\int_{\Omega_k} \mathbf{m}\overline{\mathbf{H}'_k} + \mathbf{j}\overline{\mathbf{E}'_k} \ . \end{aligned}$$

Finalement, de (II.7.79) combiné avec (II.7.71) et (II.7.72) on a, en sommant sur toutes les mailles,

$$\text{(II.7.80)} \qquad \begin{aligned} &\sum_{k,j} \int_{\Sigma_{kj}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}} \left(\sqrt{\varepsilon_{kj}}\mathbf{E}_k \wedge \nu_k + \sqrt{\mu_{kj}}((\mathbf{H}_k \wedge \nu_k) \wedge \nu_k)\right) \left(\overline{\sqrt{\varepsilon_{kj}}\mathbf{E}'_k \wedge \nu_k + \sqrt{\mu_{kj}}((\mathbf{H}'_k \wedge \nu_k) \wedge \nu_k)}\right) \\ &= \sum_{k,j} \int_{\Sigma_{kj}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}} \left(-\sqrt{\varepsilon_{kj}}\mathbf{E}_k \wedge \nu_k + \sqrt{\mu_{kj}}((\mathbf{H}_k \wedge \nu_k) \wedge \nu_k)\right) \left(\overline{-\sqrt{\varepsilon_{kj}}\mathbf{E}'_k \wedge \nu_k + \sqrt{\mu_{kj}}((\mathbf{H}'_k \wedge \nu_k) \wedge \nu_k)}\right) \\ &\quad - 2\sum_k \int_{\Omega_k} \mathbf{m}\overline{\mathbf{H}'_k} + \mathbf{j}\overline{\mathbf{E}'_k} \ . \end{aligned}$$

Les relations de continuité aux interfaces Σ_{kj} s'écrivent

$$\begin{aligned} -(\mathbf{E}_k \wedge \nu_k)_{|\Sigma_{kj}} &= (\mathbf{E}_j \wedge \nu_j)_{|\Sigma_{jk}} \\ (((\mathbf{H}_k \wedge \nu_k) \wedge \nu_k))_{|\Sigma_{kj}} &= (((\mathbf{H}_j \wedge \nu_j) \wedge \nu_j))_{|\Sigma_{jk}} \end{aligned}$$

et comme par définition $\varepsilon_{kj} = \varepsilon_{jk}$ et $\mu_{kj} = \mu_{jk}$ (II.7.64) on obtient la relation de compatibilité aux interfaces. Avec la condition aux limites sur la frontière libre on peut écrire

$$\text{(II.7.81)} \qquad \begin{cases} (-\sqrt{\varepsilon_{kj}}\mathbf{E}_k \wedge \nu_k + \sqrt{\mu_{kj}}((\mathbf{H}_k \wedge \nu_k) \wedge \nu_k))_{|\Sigma_{kj}} = (+\sqrt{\varepsilon_{jk}}\mathbf{E}_j \wedge \nu_j + \sqrt{\mu_{jk}}((\mathbf{H}_j \wedge \nu_j) \wedge \nu_j))_{|\Sigma_{jk}} \\ (-\sqrt{\varepsilon_{kk}}\mathbf{E}_k \wedge \nu_k + \sqrt{\mu_{kk}}((\mathbf{H}_k \wedge \nu_k) \wedge \nu_k)) = Q(+\sqrt{\varepsilon_{kk}}\mathbf{E}_k \wedge \nu_k + \sqrt{\mu_{kk}}((\mathbf{H}_k \wedge \nu_k) \wedge \nu_k)) + g \ . \end{cases}$$

Dans le deuxième terme de l'équation (II.7.80) nous replaçons la somme sur $\Sigma_{kj},\ \forall j$ par une somme sur Σ_{kk} et sur Σ_{kj} pour $j \neq k$. Alors, à l'aide des relations (II.7.81), nous obtenons immédiatement l'équation (II.7.69). En remarquant que $\nu_k = -\nu_j$, on obtient les relations

$$\text{(II.7.82)} \qquad \begin{cases} (-\sqrt{\varepsilon_{kj}}\mathbf{E}_k \wedge \nu_k + \sqrt{\mu_{kj}}((\mathbf{H}_k \wedge \nu_k) \wedge \nu_k))_{|\Sigma_{kj}} = (-\sqrt{\varepsilon_{kj}}\mathbf{E}_j \wedge \nu_k + \sqrt{\mu_{kj}}\,(\mathbf{H}_j \wedge \nu_k) \wedge \nu_k)_{|\Sigma_{jk}} \\ (-\sqrt{\varepsilon_{kk}}\mathbf{E}_k \wedge \nu_k + \sqrt{\mu_{kk}}((\mathbf{H}_k \wedge \nu_k) \wedge \nu_k)) = Q(+\sqrt{\varepsilon_{kk}}\mathbf{E}_k \wedge \nu_k + \sqrt{\mu_{kk}}((\mathbf{H}_k \wedge \nu_k) \wedge \nu_k)) + g \ , \end{cases}$$

qui donnent la relation (II.7.70). □

II.7.2.3 Réciproque, mise en place de la formulation ultra-faible.

Le théorème 14 est la réciproque du théorème 13 assurant l'équivalence, par un relèvement, entre le problème de Maxwell initial (1 p. 79) et la formulation ultra-faible correspondante (II.7.69).

Définition 8 *On définit $\mathcal{D}(\Omega, K)$ comme l'espace des fonctions*

$$\prod_{k=1}^{K} \tilde{\mathcal{H}}(\Omega_k, \partial\Omega_k) \tag{II.7.83}$$

qui vérifient les relations de continuité des traces tangentielles aux interfaces $\Sigma_{k,j\neq k}$ (II.7.19).

Théorème 14 *On suppose que l'hypothèse H4 est vérifiée. Nous supposons que $\mathcal{X}$ vérifie (II.7.69), c'est-à-dire,*

$$\begin{aligned}
&\sum_{k,j}\int_{\Sigma_{kj}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}\mathcal{X}_k \left(\overline{\sqrt{\varepsilon_{kj}}\mathbf{E}'_k \wedge \nu_k + \sqrt{\mu_{kj}}((\mathbf{H}'_k \wedge \nu_k)\wedge \nu_k)}\right)\\
&-\sum_{k,j\neq k}\int_{\Sigma_{kj}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}\mathcal{X}_j \left(\overline{-\sqrt{\varepsilon_{kj}}\mathbf{E}'_k \wedge \nu_k + \sqrt{\mu_{kj}}((\mathbf{H}'_k \wedge \nu_k)\wedge \nu_k)}\right)\\
&-\sum_{k}\int_{\Sigma_{kk}} \frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}}Q_k\mathcal{X}_k \left(\overline{-\sqrt{\varepsilon_{kk}}\mathbf{E}'_k \wedge \nu_k + \sqrt{\mu_{kk}}((\mathbf{H}'_k \wedge \nu_k)\wedge \nu_k)}\right)\\
&= -2\sum_k\int_{\Omega_k} \mathbf{m}\overline{\mathbf{H}'_k} + \mathbf{j}\overline{\mathbf{E}'_k} + \sum_k\int_{\Sigma_{kk}} \frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}} g \left(\overline{-\sqrt{\varepsilon_{kk}}\mathbf{E}'_k \wedge \nu_k + \sqrt{\mu_{kk}}((\mathbf{H}'_k \wedge \nu_k)\wedge \nu_k)}\right)
\end{aligned} \tag{II.7.84}$$

pour tout $(\mathbf{E}', \mathbf{H}')$ dont les restrictions à Ω_k notées $(\mathbf{E}'_k, \mathbf{H}'_k) \in \tilde{\mathcal{H}}(\Omega_k, \partial\Omega_k)$ vérifient

$$\begin{cases}
\nabla\wedge \mathbf{E}'_k - i\omega\overline{\mu}\mathbf{H}'_k = 0 & \textit{dans } \Omega_k\\
\nabla\wedge \mathbf{H}'_k + i\omega\overline{\varepsilon}\mathbf{E}'_k = 0 & \textit{dans } \Omega_k\\
\sqrt{\varepsilon_{kj}}(\mathbf{E}'_k \wedge \nu_k) + \sqrt{\mu_{kj}}((\mathbf{H}'_k \wedge \nu_k)\wedge \nu_k) & \in L^2_t(\partial\Omega_k)\ .
\end{cases} \tag{II.7.85}$$

Alors la fonction $(\mathbf{E}, \mathbf{H})$ définie par

$$\begin{cases}
\mathbf{E}_{|\Omega_k} = \mathbf{E}_k\\
\mathbf{H}_{|\Omega_k} = \mathbf{H}_k\\
\nabla\wedge \mathbf{E}_k - i\omega\mu_k\mathbf{H}_k = -\mathbf{m}_{|\Omega_k}\\
\nabla\wedge \mathbf{H}_k + i\omega\varepsilon_k\mathbf{E}_k = +\mathbf{j}_{|\Omega_k}\\
+\sqrt{\varepsilon_{kj}}(\mathbf{E}_k \wedge \nu_k) + \sqrt{\mu_{kj}}((\mathbf{H}_k \wedge \nu_k)\wedge \nu_k) = \mathcal{X}_k\ .
\end{cases} \tag{II.7.86}$$

est solution du problème de Maxwell (1 p. 79) dans Ω entier et est une fonction de $\mathcal{D}(\Omega, K)$.

Preuve. D'après le théorème 11, les restrictions $(\mathbf{E}_k, \mathbf{H}_k)$ de $(\mathbf{E}, \mathbf{H})$ à Ω_k existent dans $\tilde{\mathcal{H}}(\Omega_k, \partial\Omega_k)$ pour tout $k = 1, K$ et d'après le corollaire 7 $(\mathbf{E}'_k, \mathbf{H}'_k)$ existent aussi dans $\tilde{\mathcal{H}}(\Omega_k, \partial\Omega_k)$ pour tout $k = 1, K$. Donc, comme dans le théorème 13, les hypothèses (II.7.85) et (II.7.86) soit (II.7.76) que l'on réécrit ici :

$$\begin{cases}
\nabla\wedge \mathbf{E}_k - i\omega\mu_k\mathbf{H}_k = -\mathbf{m}_{|\Omega_k}\\
\nabla\wedge \mathbf{H}_k + i\omega\varepsilon_k\mathbf{E}_k = +\mathbf{j}_{|\Omega_k}\\
\nabla\wedge \mathbf{E}'_k - i\omega\overline{\mu}\mathbf{H}'_k = 0 \text{ dans } \Omega_k\\
\nabla\wedge \mathbf{H}'_k + i\omega\overline{\varepsilon}\mathbf{E}'_k = 0 \text{ dans } \Omega_k
\end{cases} \tag{II.7.87}$$

mènent à l'égalité (II.7.80). En substituant $\mathcal{X}_k = +\sqrt{\varepsilon_{kj}}(\mathbf{E}_k \wedge \nu_k) + \sqrt{\mu_{kj}}((\mathbf{H}_k \wedge \nu_k)\wedge \nu_k)$ dans (II.7.80), on a

$$\begin{aligned}
&\sum_{k,j}\int_{\Sigma_{kj}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}\mathcal{X}_k \left(\overline{\sqrt{\varepsilon_{kj}}\mathbf{E}'_k \wedge \nu_k + \sqrt{\mu_{kj}}((\mathbf{H}'_k \wedge \nu_k)\wedge \nu_k)}\right)\\
&= \sum_{k,j\neq k}\int_{\Sigma_{kj}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}\left(-\sqrt{\varepsilon_{kj}}\mathbf{E}_k \wedge \nu_k + \sqrt{\mu_{kj}}((\mathbf{H}_k \wedge \nu_k)\wedge \nu_k)\right) \left(\overline{-\sqrt{\varepsilon_{kj}}\mathbf{E}'_k \wedge \nu_k + \sqrt{\mu_{kj}}((\mathbf{H}'_k \wedge \nu_k)\wedge \nu_k)}\right)\\
&+\sum_{k}\int_{\Sigma_{kk}} \frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}}\left(-\sqrt{\varepsilon_{kk}}\mathbf{E}_k \wedge \nu_k + \sqrt{\mu_{kk}}((\mathbf{H}_k \wedge \nu_k)\wedge \nu_k)\right) \left(\overline{-\sqrt{\varepsilon_{kk}}\mathbf{E}'_k \wedge \nu_k + \sqrt{\mu_{kk}}((\mathbf{H}'_k \wedge \nu_k)\wedge \nu_k)}\right)\\
&-2\int_{\Omega_k}\left(\mathbf{m}\overline{\mathbf{H}'_k} + \mathbf{j}\overline{\mathbf{E}'_k}\right)\ .
\end{aligned} \tag{II.7.88}$$

Or l'hypothèse (II.7.84) est

$$
\begin{aligned}
&\sum_{k,j}\int_{\Sigma_{kj}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}\mathcal{X}_k \left(\overline{\sqrt{\varepsilon_{kj}}\mathbf{E}'_k\wedge\nu_k+\sqrt{\mu_{kj}}((\mathbf{H}'\wedge\nu)\wedge\nu)}\right)\\
&=\sum_{k,j\neq k}\int_{\Sigma_{kj}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}\mathcal{X}_j \left(\overline{-\sqrt{\varepsilon_{kj}}\mathbf{E}'_k\wedge\nu_k+\sqrt{\mu_{kj}}((\mathbf{H}'_k\wedge\nu_k)\wedge\nu_k)}\right)\\
&+\sum_{k}\int_{\Sigma_{kk}} \frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}}Q_k\mathcal{X}_k \left(\overline{-\sqrt{\varepsilon_{kk}}\mathbf{E}'_k\wedge\nu_k+\sqrt{\mu_{kk}}((\mathbf{H}'_k\wedge\nu_k)\wedge\nu_k)}\right)\\
&-2\sum_{k}\int_{\Omega_k} \mathbf{m}\overline{\mathbf{H}'_k}+\mathbf{j}\overline{\mathbf{E}'_k}\\
&+\sum_{k}\int_{\Sigma_{kk}} \frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}}g \left(\overline{-\sqrt{\varepsilon_{kk}}\mathbf{E}'_k\wedge\nu_k+\sqrt{\mu_{kk}}((\mathbf{H}'_k\wedge\nu_k)\wedge\nu_k)}\right)
\end{aligned}
\tag{II.7.89}
$$

et en faisant la différence avec (II.7.88) on a, en mettant à gauche les sommes sur les interfaces et à droite les sommes sur les bords, après simplification :

$$
\begin{aligned}
&\sum_{k,j\neq k}\int_{\Sigma_{kj}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}\mathcal{X}_j \left(\overline{-\sqrt{\varepsilon_{kj}}\mathbf{E}'_k\wedge\nu_k+\sqrt{\mu_{kj}}((\mathbf{H}'_k\wedge\nu_k)\wedge\nu_k)}\right)\\
&-\sum_{k,j\neq k}\int_{\Sigma_{kj}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}\left(-\sqrt{\varepsilon_{kj}}\mathbf{E}_k\wedge\nu_k+\sqrt{\mu_{kj}}((\mathbf{H}_k\wedge\nu_k)\wedge\nu_k)\right) \left(\overline{-\sqrt{\varepsilon_{kj}}\mathbf{E}'_k\wedge\nu_k+\sqrt{\mu_{kj}}((\mathbf{H}'_k\wedge\nu_k)\wedge\nu_k)}\right)\\
=&\sum_{k}\int_{\Sigma_{kk}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}g \left(\overline{-\sqrt{\varepsilon_{kj}}\mathbf{E}'_k\wedge\nu_k+\sqrt{\mu_{kj}}((\mathbf{H}'_k\wedge\nu_k)\wedge\nu_k)}\right)\\
&+\sum_{k}\int_{\Sigma_{kk}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}\left(-\sqrt{\varepsilon_{kj}}\mathbf{E}_k\wedge\nu_k+\sqrt{\mu_{kj}}((\mathbf{H}_k\wedge\nu_k)\wedge\nu_k)\right) \left(\overline{-\sqrt{\varepsilon_{kj}}\mathbf{E}'_k\wedge\nu_k+\sqrt{\mu_{kj}}((\mathbf{H}'_k\wedge\nu_k)\wedge\nu_k)}\right)\\
&-\sum_{k}\int_{\Sigma_{kk}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}}t\mathcal{X}_k \left(\overline{-\sqrt{\varepsilon_{kj}}\mathbf{E}'_k\wedge\nu_k+\sqrt{\mu_{kj}}((\mathbf{H}'_k\wedge\nu_k)\wedge\nu_k)}\right) .
\end{aligned}
\tag{II.7.90}
$$

Ceci étant vrai pour toute fonction $(\mathbf{E}^{'},\mathbf{H}^{'})$ on a les relations de continuité (II.7.81) que l'on réécrit

$$
\begin{cases}
(-\sqrt{\varepsilon_{kj}}\mathbf{E}_k\wedge\nu_k+\sqrt{\mu_{kj}}((\mathbf{H}_k\wedge\nu_k)\wedge\nu_k))_{|\Sigma_{kj}} = (+\sqrt{\varepsilon_{jk}}\mathbf{E}_j\wedge\nu_j+\sqrt{\mu_{jk}}((\mathbf{H}_j\wedge\nu_j)\wedge\nu_j))_{|\Sigma_{jk}}\\
(-\sqrt{\varepsilon_{kk}}\mathbf{E}_k\wedge\nu_k+\sqrt{\mu_{kk}}((\mathbf{H}_k\wedge\nu_k)\wedge\nu_k)) = Q(+\sqrt{\varepsilon_{kk}}\mathbf{E}_k\wedge\nu_k+\sqrt{\mu_{kk}}((\mathbf{H}_k\wedge\nu_k)\wedge\nu_k))+g\ .
\end{cases}
\tag{II.7.91}
$$

Des relations (II.7.86) et (II.7.91) on élimine les relations aux interfaces et l'on obtient finalement que la fonction $(\mathbf{E},\mathbf{H})$ vérifie

$$
\begin{cases}
\text{dans } \Omega\\
\nabla\wedge\mathbf{E}-i\omega\mu\mathbf{H}=-\mathbf{m}\\
\nabla\wedge\mathbf{H}+i\omega\varepsilon\mathbf{E}=+\mathbf{j}\\
\text{et sur } \Gamma\\
-\sqrt{\varepsilon_{kk}}\mathbf{E}\wedge\nu+\sqrt{\mu_{kk}}\,(\mathbf{H}\wedge\nu)\wedge\nu = Q(+\sqrt{\varepsilon_{kk}}\mathbf{E}\wedge\nu+\sqrt{\mu_{kk}}\,(\mathbf{H}\wedge\nu)\wedge\nu)+g\ .
\end{cases}
\tag{II.7.92}
$$

Le théorème 12 assure l'existence et l'unicité de $(\mathbf{E},\mathbf{H})$ vérifiant (II.7.86) dans l'espace $\mathrm{H}(\mathrm{rot},\Omega)$ et $\mathrm{H}(\mathrm{div},\Omega_i)$ sur tous les ensembles Ω_i où les coefficients ε et μ sont constants. □

II.7.2.4 Définition des opérateurs formels.

Nous définissons pour Maxwell les opérateurs équivalents á ceux définis lors de l'étude du problème de Helmholtz. Nous supposons toujours les coefficients ε et μ constants sur les éléments à frontière lipschitzienne de la partition de Ω effectuée section II.7.2.1.

Définition 9 *D'après le théorème 14 nous pouvons définir $E_{\mathbf{j},\mathbf{m}}$ l'application de relèvement du problème de Maxwell par*

$$
E_{\mathbf{j},\mathbf{m}} = \begin{cases} V \to (\mathcal{D}(\Omega,K))^2\\ \mathcal{X}\mapsto(\mathbf{E},\mathbf{H})\end{cases}
\tag{II.7.93}
$$

qui à $\mathcal{X}$ solution de la formulation variationnelle II.7.84 fait correspondre $(\mathbf{E},\mathbf{H})$ *solution du problème (1 p. 79).*

Définition 10 *D'après le corollaire 7 nous pouvons définir E^* l'opérateur de relèvement pour les fonctions tests de la formulation variationnelle ultra-faible :*

$$E^* = \begin{cases} V \to (\mathcal{D}(\Omega,K))^2 \\ \mathcal{Y} \mapsto (\mathbf{E}',\mathbf{H}') = ((\mathbf{E}'{}_k,\mathbf{H}'{}_k))_{k=1\ldots K}, \ (\mathbf{E}'{}_k,\mathbf{H}'{}_k) = (\mathbf{E}',\mathbf{H}')_{|\Omega_k} \end{cases} \tag{II.7.94}$$

où $(\mathbf{E}'{}_k,\mathbf{H}'{}_k)$ est l'unique solution dans la maille Ω_k de

$$\begin{cases} \nabla\wedge \mathbf{E}'{}_k - i\omega\overline{\mu}\mathbf{H}'{}_k = 0 \ \textit{dans}\ \Omega_k \\ \nabla\wedge \mathbf{H}'{}_k + i\omega\overline{\varepsilon}\mathbf{E}'{}_k = 0 \ \textit{dans}\ \Omega_k \\ \sqrt{\varepsilon_{kj}}(\mathbf{E}'{}_k\wedge\nu_k) + \sqrt{\mu_{kj}}((\mathbf{H}'{}_k\wedge\nu_k)\wedge\nu_k) = \mathcal{Y}_{|\Sigma_{kj}} \qquad \textit{sur}\ \partial\Omega_k = \bigcup_{j(k)}\Sigma_{kj}\ . \end{cases}$$

Notons que l'opérateur E^* est linéaire et que cette définition a un sens d'après le corollaire 7.

Notation 4 *Nous noterons $E^1_{\mathbf{j},\mathbf{m}}$ l'application $E_{\mathbf{j},\mathbf{m}}$ dont l'image est restreinte à son premier champ* $\mathbf{E}$. *En d'autres termes, avec les notations de la définition 9 nous avons*

$$E^1_{\mathbf{j},\mathbf{m}}(\mathcal{X}) = \mathbf{E}$$
$$E^2_{\mathbf{j},\mathbf{m}}(\mathcal{X}) = \mathbf{H}$$

et de même pour l'opérateur linéaire E^*_1 avec les notations de la définition 10 :

$$E^*_1(\mathcal{Y}) = \mathbf{E}'$$
$$E^*_2(\mathcal{Y}) = \mathbf{H}'$$

Définition 11 *Soit $\mathcal{Y} = \sqrt{\varepsilon_{kj}}(\mathbf{E}'\wedge\nu) + \sqrt{\mu_{kj}}((\mathbf{H}'\wedge\nu)\wedge\nu)$. On définit l'opérateur F par l'application qui transforme la trace sortante $\sqrt{\varepsilon_{kj}}(\mathbf{E}'\wedge\nu) + \sqrt{\mu_{kj}}((\mathbf{H}'\wedge\nu)\wedge\nu)$ en la trace entrante $-\sqrt{\varepsilon_{kj}}(\mathbf{E}'\wedge\nu) + \sqrt{\mu_{kj}}((\mathbf{H}'\wedge\nu)\wedge\nu)$:*

$$F = \begin{cases} V \to V \\ \mathcal{Y} \mapsto F\mathcal{Y} = -\sqrt{\varepsilon_{kj}}(\mathbf{E}'\wedge\nu) + \sqrt{\mu_{kj}}((\mathbf{H}'\wedge\nu)\wedge\nu)\ . \end{cases} \tag{II.7.95}$$

Remarque 36 On a $(F\mathcal{Y})_{|\Sigma_{k,j}} = -(\mathcal{Y})_{|\Sigma_{k,j}} + 2\sqrt{\varepsilon_{kj}}((E^*_1(\mathcal{Y}))_{|\Sigma_{k,j}}\wedge\nu_{k,j}$ où $(\mathcal{Y})_{|\Sigma_{k,j}} \in L^2_t(\Sigma_{kj})$ par hypothèse et $(E^*_1(\mathcal{Y}))_{|\Omega_k} \in \mathcal{H}(\Omega_k,\partial\Omega_k)$ d'après le corollaire 7 p. 87. Ceci montre que $(F\mathcal{Y})_{|\Sigma_{k,j}} \in L^2_t(\Sigma_{kj})$ et donc que $F\mathcal{Y} \in V$.

Définition 12 *Définissons l'opérateur Π par*

$$\Pi = \begin{cases} V \to V \\ \mathcal{Y}_{|\Sigma_{kj}} \mapsto \mathcal{Y}_{|\Sigma_{jk}} \\ \mathcal{Y}_{|\Gamma_k} \mapsto Q\mathcal{Y}_{|\Gamma_k}\ . \end{cases} \tag{II.7.96}$$

La fonction complexe Q, définie sur Γ, vérifiant $|Q| \leq 1$, est donnée dans la condition de bord (II.6.2). Notons que $||\Pi|| \leq 1$ puisque l'on suppose $|Q| \leq 1$.

Définition 13 *Notons F^* l'adjoint de F. Définissons l'opérateur A de V dans V par*

$$A = F^*\Pi\ . \tag{II.7.97}$$

II.7.2.5 Forme synthétique.

Dans l'équation (II.7.84) nous reconnaissons dans le premier terme intégral le produit scalaire dans V, dans le second terme nous reconnaissons à gauche l'opérateur Π appliqué à $\mathcal{X}$, à droite l'opérateur F appliqué à la fonction de base $\mathcal{Y}$. Nous sommes donc en mesure d'énoncer le théorème 15.

Théorème 15 (Equivalence des formulations variationnelles classique et ultra-faible)
a) *Le problème (II.7.84) est équivalent à*

$$\text{(II.7.98)} \qquad \begin{cases} \forall \mathcal{Y} \in V \ \text{ trouver } \mathcal{X} \in V \ \text{ tel que} \\ (\mathcal{X}, \mathcal{Y})_V - (\Pi\mathcal{X}, F\mathcal{Y})_V = (b, \mathcal{Y})_V \end{cases}$$

où le second membre $b \in V$ est défini, via le théorème de représentation de Riesz, par :

$$\forall \mathcal{Y} \in V$$

$$\text{(II.7.99)} \qquad \begin{aligned}(b, \mathcal{Y})_V = & -2\sum_k \int_{\Omega_k} \mathbf{m}\overline{\mathbf{H}'_k} + \mathbf{j}\overline{\mathbf{E}'_k} \\ & + \sum_k \int_{\Gamma_k} \frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}} g\left(\overline{-\sqrt{\varepsilon_{kk}}\mathbf{E}'_k \wedge \nu_k + \sqrt{\mu_{kk}}((\mathbf{H}'_k \wedge \nu_k) \wedge \nu_k)}\right) \ .\end{aligned}$$

b) *Si $(\mathbf{E}, \mathbf{H})$ est solution de (1 p. 79) assez régulière (dans $\mathcal{D}(\Omega, K)$ (Définition 8)) alors $\mathcal{X} = (\mathcal{X}_{k,j})_{k=1\ldots K, j=j(k)}$ avec $\mathcal{X}_{k,j} = \mathcal{X}_{|\Sigma_{k,j}} = \sqrt{\varepsilon_{kj}}(\mathbf{E}_k \wedge \nu_k) + \sqrt{\mu_{kj}}((\mathbf{H}_k \wedge \nu_k) \wedge \nu_k)$ est solution de (II.7.98).*

c) *Réciproquement, si $\mathcal{X}$ est solution de (II.7.98) alors $(\mathbf{E}, \mathbf{H}) = E_{\mathbf{j},\mathbf{m}}(\mathcal{X})$ est l'unique solution de (1 p. 79). Le problème (II.7.98) est équivalent à :*

$$\text{(II.7.100)} \qquad \begin{cases} \text{Pour } b \in V, \text{ vérifiant (II.7.99), trouver } \mathcal{X} \in V \\ \boxed{(I - A)\mathcal{X} = b} \end{cases}$$

Preuve. Soit $\mathcal{X} = \sqrt{\varepsilon_{kj}}(\mathbf{E} \wedge \nu) + \sqrt{\mu_{kj}}((\mathbf{H} \wedge \nu) \wedge \nu)$ et $\mathcal{Y} \in V$ donné. De $\mathcal{Y}$, on définit $(\mathbf{E}', \mathbf{H}')$ par $E^*(\mathcal{Y}) = (\mathbf{E}', \mathbf{H}')$ (d'après le théorème 11 assurant l'existence et l'unicité). Alors $\mathcal{Y} = +\sqrt{\varepsilon_{kj}}(\mathbf{E}' \wedge \nu) + \sqrt{\mu_{kj}}((\mathbf{H}' \wedge \nu) \wedge \nu)$. En remplaçant dans l'égalité (II.7.84) par les définitions de F (définition 11, équation (II.7.95)) et Π (définition 12, équation (II.7.96)) on obtient (II.7.98). Comme (II.7.98) est valable pour tout $\mathcal{Y} \in V$ (E est défini sur V), on peut affirmer (II.7.100). □

II.7.2.6 Propriétés des opérateurs.

L'opérateur de la formulation variationnelle classique se décompose en la somme d'un opérateur coercif et d'une perturbation compacte. Nous montrons ici que l'opérateur $I - A$ de notre formulation (II.7.100) est injectif et que A est de norme inférieure à 1 (propositions 10 et 11). Cette propriété mathématique constitue un point fort essentiel de la formulation ultra-faible.

Lemme 14 *L'opérateur F est de norme inférieure à 1 et $||F^*|| = ||F||$.*

Preuve. Soit $E^*(\mathcal{Y}) = (\mathbf{E}', \mathbf{H}')$. Alors :

$$\text{(II.7.101)} \qquad \begin{aligned}(F\mathcal{Y}, F\mathcal{Y}) &= \int_{\partial\Omega_k} \frac{1}{\varepsilon_{kj}\mu_{kj}} \left|-\sqrt{\varepsilon_{kj}}(\mathbf{E}'_k \wedge \nu_k) + \sqrt{\mu_{kj}}((\mathbf{H}'_k \wedge \nu_k) \wedge \nu_k)\right|^2 \\ &= \int_{\partial\Omega_k} \frac{1}{\varepsilon_{kj}\mu_{kj}} \left[\varepsilon_{kj}|(\mathbf{E}'_k \wedge \nu_k)|^2 + \mu_{kj}|((\mathbf{H}'_k \wedge \nu_k) \wedge \nu_k)|^2\right] \\ &\quad - 4\Re\left(\int_{\partial\Omega_k} (\mathbf{E}'_k \wedge \nu_k).\overline{((\mathbf{H}'_k \wedge \nu_k) \wedge \nu_k)}\right) \ .\end{aligned}$$

Or

$$\text{(II.7.102)} \qquad \int_{\partial\Omega_k} (\mathbf{E}'_k \wedge \nu_k).\overline{((\mathbf{H}'_k \wedge \nu_k) \wedge \nu_k)_k} = \int_{\Omega_k} (\nabla \wedge \mathbf{E}'_k)\overline{\mathbf{H}'_k} - (\nabla \wedge \overline{\mathbf{H}'_k})\mathbf{E}'_k \ .$$

En intégrant par parties dans

$$\text{(II.7.103)} \qquad \begin{cases} \nabla\wedge \mathbf{E}'_k - i\omega\overline{\mu}\mathbf{H}'_k = 0 & \text{dans } \Omega_k \\ \nabla\wedge \mathbf{H}'_k + i\omega\overline{\varepsilon}\mathbf{E}'_k = 0 & \text{dans } \Omega_k \end{cases}$$

nous avons :

$$\text{(II.7.104)} \qquad \begin{cases} \displaystyle\int_{\Omega_k} (\nabla\wedge \mathbf{E}'_k)\overline{\mathbf{H}'_k} - i\omega \int_{\Omega_k} \overline{\mu}\mathbf{H}'_k\overline{\mathbf{H}'_k} = 0 \\ \displaystyle\int_{\Omega_k} (\nabla\wedge \mathbf{H}'_k)\overline{\mathbf{E}'_k} + i\omega \int_{\Omega_k} \overline{\varepsilon}\mathbf{E}'_k\overline{\mathbf{E}'_k} = 0 \end{cases}$$

d'où, de (II.7.102) et (II.7.104) on obtient en conjuguant la deuxième équation de (II.7.104) :

$$\text{(II.7.105)} \qquad \int_{\Omega_k} (\nabla\wedge \mathbf{E}'_k)\overline{\mathbf{H}'_k} - (\nabla\wedge \overline{\mathbf{H}'_k})\mathbf{E}'_k = i\omega \left(\int_{\Omega_k} \overline{\mu}|\mathbf{H}'_k|^2 - \varepsilon|\mathbf{E}'_k|^2 \right)$$

et en remarquant que $\Im(\overline{\varepsilon}_k) = -\Im(\varepsilon_k)$ et $\Re(i\alpha) = -\Im(\alpha)$

$$\text{(II.7.106)} \qquad \begin{aligned} (\mathcal{Y}, \mathcal{Y}) - (F\mathcal{Y}, F\mathcal{Y}) &= 4\Re \int_{\partial\Omega_k} (\mathbf{E}'_k \wedge \nu_k).\overline{((\mathbf{H}'_k \wedge \nu_k) \wedge \nu_k)} \\ &= -4\Im\omega \int_{\Omega_k} \overline{\mu}_k\|\mathbf{H}'_k\|^2 + \overline{\varepsilon}_k\|\mathbf{E}'_k\|^2 \end{aligned}$$

et finalement

$$\text{(II.7.107)} \qquad (\mathcal{Y}, \mathcal{Y}) - (F\mathcal{Y}, F\mathcal{Y}) = 4\Im\omega \int_{\Omega_k} \mu_k\|\mathbf{H}'_k\|^2 + \varepsilon_k\|\mathbf{E}'_k\|^2$$

D'après l'hypothèse (II.7.1) qui stipule $\Im(\mu) \geq 0$ et $\Im(\varepsilon) \geq 0$, l'équation (II.7.107) implique

$$\|F\|^2 = \sup_{\mathcal{Y}\in V, \mathcal{Y}\neq 0} \sqrt{\frac{(F\mathcal{Y}, F\mathcal{Y})}{(\mathcal{Y}, \mathcal{Y})}} \leq 1 \ .$$

On a montré que F est borné sur V (V est un Hilbert, définition 6 p. 92) un espace de Banach [11]. Ceci montre que $||F^*|| = ||F|| \leq 1$. □

Remarque 37 Dans le cas de matériaux non dispersifs, (le vide par exemple), on a $\Im(\varepsilon) = 0$ et $\Im(\mu) = 0$. Ceci entraîne que F est une isométrie comme pour le problème de Helmholtz sans coefficient.

Proposition 10 *L'opérateur $(I - A)$ est injectif.*

Preuve. Ceci est une conséquence du théorème 15. Si $b = 0$ dans (II.7.100), alors on peut prendre $\mathbf{m} = \mathbf{j} = 0$ sur Ω et $g = 0$ sur Γ. Le théorème d'unicité du problème de Maxwell montre que $(\mathbf{E}, \mathbf{H}) = (0, 0)$ puis $\mathcal{X} = 0$ par linéarité. □

Proposition 11 *La norme induite de A vérifie $||A|| \leq 1$.*

Preuve. L'opérateur Π défini par (II.7.96) vérifie $\|\Pi\| \leq 1$. Ceci, combiné avec le lemme 14 (F est de norme inférieure à 1) donne $||A|| \leq 1$. □

Chapitre II.8

Discrétisation du problème de Maxwell.

Le chapitre précédent a introduit exactement le même formalisme que celui introduit partie I pour le problème de Helmholtz. La même procédure de discrétisation de Galerkin est employée. Nous introduisons un espace de dimension finie V_h inclus dans V. A l'aide des propositions 10 ($(I - A)$ est injectif) et 11 ($||A|| \leq 1$) nous montrons le théorème d'existence et d'unicité de la formulation discrète, équivalent pour Maxwell du théorème 5 p. 23 de la partie I.

Théorème 16 *Le problème*

$$\text{(II.8.1)} \qquad \begin{aligned} &\textit{Trouver } \mathcal{X}_h \in V_h \subset V, \ \textit{tel que} \\ &\begin{cases} \forall \mathcal{Y}_h \in V_h \\ (\mathcal{X}_h, \mathcal{Y}_h)_V - (\Pi \mathcal{X}_h, F\mathcal{Y}_h)_V = (b, \mathcal{Y}_h)_V \end{cases} \end{aligned}$$

a une solution unique.

La procédure de discrétisation de Galerkin définit l'espace d'approximation $V_h \subset V$ par l'introduction d'un nombre fini I de fonctions de base $\mathcal{Z}_i$ où i est l'indice d'une numérotation J que nous construirons dans la section II.8.1.1. La formulation discrète (II.8.1) de (II.7.98) conduit au système matriciel (II.8.2), version discrète de (II.7.100), où l'on garde abusivement mais sans risque de confusion la notation b pour le second membre.

$$\text{(II.8.2)} \qquad \begin{cases} \text{Trouver } X = ([\mathcal{X}_i]_{i \in J}) \in \mathbb{C}^I \text{ tel que} \\ (D - C)X = b \end{cases}$$

La solution approchée $\mathcal{X}_h$ est entièrement définie par la donnée des I coefficients complexes $\mathcal{X}_i$ par

$$\text{(II.8.3)} \qquad \mathcal{X}_h = \sum_{i \in J} \mathcal{X}_i \mathcal{Z}_i \ .$$

Notons que la matrice D est la matrice du produit scalaire dans V_h (opérateur discret de I), la matrice C est la matrice de la forme bilinéaire $(\Pi \mathcal{Z}_i, F\mathcal{Z}_{i'})$ (opérateur discret de A).

II.8.1 Approximation de type Galerkin.

II.8.1.1 Construction de l'espace d'approximation V_h.

Nous effectuons les mêmes hypothèses sur le maillage 3-D que sur le maillage 2-D de Helmholtz (hypothèses H1, H2 et H3 p. 24). Avec les notations de la section II.7.2.1, définissons les fonctions de base $\mathcal{Z}$ et leur numérotation k, l de l'espace $V_h \subset V$.

Définition 14 (Base de l'espace de discrétisation V_h.) *On considère la partition de Ω en K éléments Ω_k et pour chaque élément un nombre L constant de fonctions $(\mathbf{E}'_{kl}, \mathbf{H}'_{kl})$ linéairement indépendantes entre elles dans $(\mathcal{H}(\Omega_k, \partial\Omega_k))^2$ et vérifiant*

$$\text{(II.8.4)} \qquad \begin{cases} ((\mathbf{E}'_{kl}, \mathbf{H}'_{kl}))_{|\Omega_j} = (0,0) \text{ si } k \neq j \\ \nabla\wedge \mathbf{E}'_{kl} - i\omega\overline{\mu}\mathbf{H}'_{kl} = 0 \text{ dans } \Omega_k \\ \nabla\wedge \mathbf{H}'_{kl} + i\omega\overline{\varepsilon}\mathbf{E}'_{kl} = 0 \text{ dans } \Omega_k \ . \end{cases}$$

On définit alors les LK fonctions de base $\mathcal{Z}_{kl} \in V$ de l'espace de discrétisation V_h par

$$\text{(II.8.5)} \qquad \begin{cases} (\mathcal{Z}_{kl})_{|\partial\Omega_j} = 0 \text{ si } k \neq j \\ (\mathcal{Z}_{kl})_{|\partial\Omega_k} = \sqrt{\varepsilon_{kj}}\mathbf{E}'_{k,l} \wedge \nu_k + \sqrt{\mu_{kj}}(\mathbf{H}'_{k,l} \wedge \nu_k) \wedge \nu_k \ . \end{cases}$$

Vérifions que les LK fonctions $\mathcal{Z}_{kl}$ forment une base de V_h.

Lemme 15 *Les fonctions $\{\mathcal{Z}_{kl}\}_{1\leq l\leq L}$ définies par (II.8.5) sont linéairement indépendantes dans V si et seulement si les fonctions $\{\mathbf{E}'_{kl}\}_{1\leq l\leq L}$ définies par (II.8.4) sont linéairement indépendantes dans $\mathcal{H}(\Omega_k, \partial\Omega_k)$.*

Preuve. Si la famille $\{\mathbf{E}'_{kl}\}_{1\leq l\leq L}$ n'est pas libre sur $\mathcal{H}(\Omega_k, \partial\Omega_k)$, alors la famille $\{\mathbf{H}'_{kl}\}_{1\leq l\leq L}$ n'est pas libre puisque $i\omega\overline{\mu}\mathbf{H}'_{kl} = \nabla\wedge \mathbf{E}'_{kl}$ et $\mathbf{E}'_{kl} \in H(\mathrm{rot}, \Omega_k)$. Ceci assure que la famille $\{\mathcal{Z}_{kl}\}_{1\leq l\leq L}$ n'est pas libre. La réciproque est une conséquence de la linéarité de l'opérateur de relèvement E_1^* (p. 97). □

Notons toujours que, comme dans la méthode des éléments finis, la procédure de discrétisation (II.8.4) à l'aide de l'espace V_h construit par l'espace des fonctions $\mathcal{Z}_{kl}$ (II.8.5) a l'avantage de mener à un système où la matrice est creuse.

II.8.1.2 Construction des opérateurs du système linéaire.

Nous explicitons les termes D, C et b du système (II.8.2).

i) Les coefficients de la matrice D définis par $D^{l,m}_{k,j} = \delta_{kj}(\mathcal{Z}_{jm}, \mathcal{Z}_{kl})_V$ sont :

$$\text{(II.8.6)} \qquad \begin{aligned} D^{l,m}_{k,k} = \sum_{j(k)} \int_{\Sigma_{k,j}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}} &\left(\sqrt{\varepsilon_{kj}}\mathbf{E}'_{k,m} \wedge \nu_k + \sqrt{\mu_{kj}}(\mathbf{H}'_{k,m} \wedge \nu_k) \wedge \nu_k\right) \\ &\left(\overline{+\sqrt{\varepsilon_{kj}}\mathbf{E}'_{k,l} \wedge \nu_k + \sqrt{\mu_{kj}}(\mathbf{H}'_{k,l} \wedge \nu_k) \wedge \nu_k}\right) \ . \end{aligned}$$

ii) Les coefficients de la matrice C définis par $C^{l,m}_{k,j} = (\Pi\mathcal{Z}_{jm}, F\mathcal{Z}_{kl})_V$ sont :

$$\text{(II.8.7)} \qquad \begin{aligned} C^{l,m}_{k,j\neq k} = \int_{\Sigma_{kj}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}} &\left(-\sqrt{\varepsilon_{kj}}\mathbf{E}'_{j,m} \wedge \nu_k + \sqrt{\mu_{kj}}(\mathbf{H}'_{j,m} \wedge \nu_k) \wedge \nu_k\right) \\ &\left(\overline{-\sqrt{\varepsilon_{kj}}\mathbf{E}'_{k,l} \wedge \nu_k + \sqrt{\mu_{kj}}(\mathbf{H}'_{k,l} \wedge \nu_k) \wedge \nu_k}\right) \end{aligned}$$

$$\text{(II.8.8)} \qquad \begin{aligned} C^{l,m}_{k,k} = \int_{\Sigma_{k,k}} \frac{Q_k}{\sqrt{\varepsilon_{kk}\mu_{kk}}} &\left(+\sqrt{\varepsilon_{kk}}\mathbf{E}'_{k,m} \wedge \nu_k + \sqrt{\mu_{kk}}(\mathbf{H}'_{k,m} \wedge \nu_k) \wedge \nu_k\right) \\ &\left(\overline{-\sqrt{\varepsilon_{kk}}\mathbf{E}'_{k,l} \wedge \nu_k + \sqrt{\mu_{kk}}(\mathbf{H}'_{k,l} \wedge \nu_k) \wedge \nu_k}\right) \end{aligned}$$

iii) Le second membre b est construit par $b_{k,l} = (b, \mathcal{Z}_{kl})_V$ (notation abusive sans risque de confusion) :

$$\text{(II.8.9)} \qquad \begin{aligned} b_{k,l} = &-2\int_{\Omega_k} \mathbf{m}\,\overline{(\mathbf{H}'_{kl})} + \mathbf{j}\,\overline{(\mathbf{E}'_{kl})} \\ &+ \int_{\Sigma_{k,k}} g \left(\overline{-\sqrt{\varepsilon_{kk}}\mathbf{E}'_{k,l} \wedge \nu_k + \sqrt{\mu_{kk}}(\mathbf{H}'_{k,l} \wedge \nu_k) \wedge \nu_k}\right) \ . \end{aligned}$$

Remarque 38 Le terme $C^{l,m}_{k,k}$ est nul pour une face $\Sigma_{k,k}$ sur la frontière extérieure où l'on pose $Q = 0$ (condition aux limites absorbante).

II.8.1.3 Résolution du système linéaire.

La solution approchée $\mathcal{X}_h$ est entièrement définie par

$$(\mathcal{X}_h)_{\partial\Omega_k} = \sum_{l=1}^{L} \mathcal{X}_{kl}\mathcal{Z}_{kl} \ . \tag{II.8.10}$$

où les LK coefficients complexes $\mathcal{X}_{kl}$ définissent le vecteur X solution du système linéaire discret

$$(D - C)X = b \ . \tag{II.8.11}$$

Le formalisme présenté est strictement identique à celui de la résolution de la formulation discrète du problème de Helmholtz dans le vide. Le système est résolu par les mêmes étapes que nous rappelons.

1. Inversion de D menant au système dans $\mathbb{C}^{LK}$:

   $$(I - D^{-1}C)X = D^{-1}b \ . \tag{II.8.12}$$

   Nous inversons D par la méthode directe de Cholesky, très performante sur les matrices hermitiennes de taille réduite, comme c'est le cas des blocs D_k.
2. Calcul de $M = D^{-1}C$ et $b^{'} = D^{-1}b$ que l'on note encore b, donnant le système dans $\mathbb{C}^{LK}$:

   $$(I - M)X = b \ . \tag{II.8.13}$$

3. Résolution finale de II.8.13 grâce à l'algorithme itératif de Richardson présenté section I.2.3.2 par (I.2.39) :

   $$\beta_n \in]0,5;1[\begin{cases} X_1 = \beta_1 b \\ X_{n+1} = \beta_n b + [(1-\beta_n)I + \beta_n M]X_n \ . \end{cases} \tag{II.8.14}$$

La première étape est assurée par le lemme suivant :

Lemme 16 *La matrice D construite à l'aide des fonctions de base $\mathcal{Z}_{kl}$ (définition 14) est hermitienne définie strictement positive. Elle est donc inversible.*

Preuve. La matrice D est hermitienne définie positive puisque c'est la matrice du produit scalaire dans l'espace de dimension finie V_h. L'inversibilité est assurée par le lemme 15 qui montre que la famille $\{\mathcal{Z}_{kl}\}_{1\leq l\leq L}$ forme une base de V_h. □

II.8.1.4 Construction des traces tangentielles des champs.

Nous décrivons comment définir et calculer une approximation des traces tangentielles du champ électromagnétique $(\mathbf{E},\mathbf{H})$ solution du problème de Maxwell stationnaire (1 p. 79) à partir de la connaissance de $\mathcal{X}_h$. Plus précisément

$$\begin{array}{l}(\mathbf{E}_h)_{|\partial\Omega_k} \wedge \nu_k \\ \big((\mathbf{H}_h)_{|\partial\Omega_k} \wedge \nu_k\big) \wedge \nu_k \ .\end{array}$$

Dans cette section, $\Sigma_{k,j}$ sera l'interface entre Ω_k et Ω_j, avec $k \neq j$, de façon à faire la distinction entre les interfaces du maillage et les faces de bord qui seront toujours notées $\Sigma_{k,k}$.

La continuité aux interfaces de la solution du problème de Maxwell $(\mathbf{E},\mathbf{H})$ nous donne les relations

$$\begin{array}{l}(\mathbf{E}_k \wedge \nu_k)_{|\Sigma_{kj}} = -(\mathbf{E}_j \wedge \nu_j)_{|\Sigma_{jk}} \\ ((\mathbf{H}_k \wedge \nu_k)\wedge \nu_k)_{|\Sigma_{kj}} = ((\mathbf{H}_j \wedge \nu_j)\wedge \nu_j)_{|\Sigma_{jk}}\end{array} \tag{II.8.15}$$

Or,

$$\mathcal{X} = (\sqrt{\varepsilon_{kj}}\mathbf{E}_k \wedge \nu_k + \sqrt{\mu_{kj}}(\mathbf{H}_k \wedge \nu_k)\wedge \nu_k)$$

et, par définition, sur Σ_{kj} (pour $j \neq k$),

$$\Pi\mathcal{X} = (\sqrt{\varepsilon_{kj}}\mathbf{E}_j \wedge \nu_j + \sqrt{\mu_{kj}}(\mathbf{H}_j \wedge \nu_j)\wedge \nu_j) \ .$$

On a donc immédiatement

$$(I+\Pi)\mathcal{X} = 2\sqrt{\mu_{kj}}((\mathbf{H}_k \wedge \nu_k) \wedge \nu_k)\ .$$

De plus sur Σ_{kk} la condition aux limites s'écrit

$$\Pi\mathcal{X} + g = (-\sqrt{\varepsilon_{kk}}\mathbf{E}_k \wedge \nu_k + \sqrt{\mu_{kk}}(\mathbf{H}_k \wedge \nu_k) \wedge \nu_k)_{|\Sigma_{kk}}$$

conduisant à

$$(I+\Pi)\mathcal{X} + g = 2\sqrt{\mu_{kk}}((\mathbf{H}_k \wedge \nu_k) \wedge \nu_k)\ .$$

Ainsi, la trace de $(\mathbf{H}_k \wedge \nu_k) \wedge \nu_k$ sur V est liée à $\mathcal{X}$ par

$$\text{(II.8.16)} \qquad \begin{cases} (\mathbf{H}_k \wedge \nu_k) \wedge \nu_k = \dfrac{1}{2\sqrt{\mu_{kj}}}[(I+\Pi)\mathcal{X}] & \text{sur } \Sigma_{kj} \\ (\mathbf{H}_k \wedge \nu_k) \wedge \nu_k = \dfrac{1}{2\sqrt{\mu_{kk}}}[(I+\Pi)\mathcal{X} + g] & \text{sur } \Sigma_{kk} \end{cases}$$

et de même la trace $\mathbf{E}_k \wedge \nu_k$ sur V est liée à $\mathcal{X}$ par

$$\text{(II.8.17)} \qquad \begin{cases} \mathbf{E}_k \wedge \nu_k = \dfrac{1}{2\sqrt{\varepsilon_{kj}}}[(I-\Pi)\mathcal{X}] & \text{sur } \Sigma_{kj} \\ \mathbf{E}_k \wedge \nu_k = \dfrac{1}{2\sqrt{\varepsilon_{kk}}}[(I-\Pi)\mathcal{X} - g] & \text{sur } \Sigma_{kk}\ . \end{cases}$$

Il est donc naturel de définir $(\mathbf{H}_k^h \wedge \nu_k) \wedge \nu_k$ sur les arêtes du maillage par

$$\text{(II.8.18)} \qquad \begin{cases} (\mathbf{H}_k^h \wedge \nu_k) \wedge \nu_k = \dfrac{1}{2\sqrt{\mu_{kj}}}[(I-\Pi)\mathcal{X}_h] & \text{sur } \Sigma_{kj} \\ (\mathbf{H}_k^h \wedge \nu_k) \wedge \nu_k = \dfrac{1}{2\sqrt{\mu_{kk}}}[(I-\Pi)\mathcal{X}_h - g] & \text{sur } \Sigma_{kk}\ . \end{cases}$$

et $\mathbf{E}_k^h \wedge \nu_k$ par

$$\text{(II.8.19)} \qquad \begin{cases} \mathbf{E}_k^h \wedge \nu_k = \dfrac{1}{2\sqrt{\varepsilon_{kj}}}[(\Pi - I)\mathcal{X}_h] & \text{sur } \Sigma_{kj} \\ \mathbf{E}_k^h \wedge \nu_k = \dfrac{1}{2\sqrt{\varepsilon_{kk}}}[(\Pi - I)\mathcal{X}_h + g] & \text{sur } \Sigma_{kk}\ . \end{cases}$$

Nous étudions annexe III.B.2 différentes techniques de calcul des traces des champs approchées $\mathbf{E}_h$ et $\mathbf{H}_h$ sur le maillage ainsi que le problème de la reconstruction dans Ω tout entier, problème dont la difficulté dépend de la linéarité de l'opérateur de relèvement E^* (cf (II.7.94)).

II.8.2 Un choix particulier de l'espace V_h.

Nous particularisons la construction générale de la formulation variationnelle ultra-faible proposée section II.8.1. Nous construisons l'espace V_h à l'aide d'ondes planes. Cette idée est naturelle, puisque ce type de fonctions a déjà été testé lors de l'étude du problème de Helmholtz. La difficulté du choix d'ondes planes réside dans la nature tridimensionnelle du problème. Une onde plane $\mathbf{E}(\mathbf{X})$ est donc définie par deux vecteurs, la polarisation $\mathbf{E}_0$ et la direction de propagation $\mathbf{k}_0$ telles que

$$\mathbf{E}(\mathbf{X}) = \mathbf{E}_0 e^{i\omega(\mathbf{k}_0\,.\,\mathbf{X})}$$

où les relations de Gauss imposent

$$(\mathbf{k}_0\,.\,\mathbf{E}_0) = 0\ .$$

Une idée naturelle serait de considérer, pour une direction de propagation donnée, deux ondes planes aux polarisations réelles, orthogonales entre elles et orthogonales à la direction de propagation de façon à vérifier la condition de divergence.

L'étude du découplage des équations de Maxwell (section II.7.1.3) nous a suggéré d'utiliser, pour une direction de propagation donnée, deux polarisations complexes conjuguées combinaisons de $\mathbf{E}_0$ et de $\mathbf{E}_0 \wedge \mathbf{k}_0$.

Nous verrons qu'un tel choix permet une réduction appréciable de la place mémoire nécessaire lors de l'implémentation informatique.

II.8.2.1 Construction de solutions du problème dual adjoint.

Définissons d'abord les fonctions $\mathbf{E}'_{kl}$ et $\mathbf{H}'_{kl}$ (II.8.4) qui d'après la définition 14 donnent les fonctions de base $\mathcal{Z}_{kl}$ par (II.8.5). Pour cela introduisons les ondes planes de polarisations complexes que nous appellerons "fonctions de type $\mathbf{F}$ ou de type $\mathbf{G}$".

Définition 15 *Sur chaque maille Ω_k on considère p directions de propagation réelles normées $V_{k,l'}$, l'indice l' variant de 1 à p, telles que*

(II.8.20) $$\forall (l,m) \in [1,p]^2,\ l \neq m \Longrightarrow V_{k,l} \neq V_{k,m}\ .$$

Pour une direction de propagation $V_{k,l'}$ donnée, on choisit un vecteur polarisation réel et normé $\mathbf{E}^0_{k,l'}$, quelconque dans le plan orthogonal à la direction de propagation $V_{k,l'}$. Ce vecteur réel $\mathbf{E}^0_{k,l'}$ définit les polarisation complexes $\mathbf{F}_{k,l'}$ et $\mathbf{G}_{k,l'}$ par

(II.8.21) $$\begin{aligned} \mathbf{F}_{k,l'} &= \left(\mathbf{E}^0_{k,l'} + i\mathbf{E}^0_{k,l'} \wedge V_{k,l'}\right) \\ \mathbf{G}_{k,l'} &= \left(\mathbf{E}^0_{k,l'} - i\mathbf{E}^0_{k,l'} \wedge V_{k,l'}\right)\ . \end{aligned}$$

On définit alors deux fonctions $\mathbf{E}^{\mathbf{F}}_{k,l'}$ et $\mathbf{E}^{\mathbf{G}}_{k,l'}$:

1. *la fonction $\mathbf{E}^{\mathbf{F}}_{k,l'}$ par*

   (II.8.22) $$\mathbf{E}^{\mathbf{F}}_{k,l'} = \sqrt{\overline{\mu}_k}\mathbf{F}_{k,l'}e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}\left(V_{k,l'}\,.\mathbf{x}\right)}\ ,$$

2. *la fonction $\mathbf{E}^{\mathbf{G}}_{k,l'}$ par*

   (II.8.23) $$\mathbf{E}^{\mathbf{G}}_{k,l'} = \sqrt{\overline{\mu}_k}\mathbf{G}_{k,l'}e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}\left(V_{k,l'}\,.\mathbf{x}\right)}\ .$$

Les fonctions $\mathbf{E}^{\mathbf{F}}_{k,l'}$ et $\mathbf{E}^{\mathbf{G}}_{k,l'}$ sont étendues à tout le domaine Ω par la fonction nulle

(II.8.24) $$j \neq k \Longrightarrow \begin{cases} \mathbf{E}^{\mathbf{F}}_{j,l'} = 0 \\ \mathbf{E}^{\mathbf{G}}_{j,l'} = 0\ . \end{cases}$$

Définition 16 *Pour $V_{k,l'}$ et $\mathbf{E}^0_{k,l'}$ donnés, on définit la fonction $\mathbf{H}^{\mathbf{F}}_{k,l'}$ par*

(II.8.25) $$\mathbf{H}^{\mathbf{F}}_{k,l'} = +i\sqrt{\overline{\varepsilon}_k}\mathbf{F}_{k,l'}e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}\left(V_{k,l'}\,.\mathbf{x}\right)}$$

et la fonction $\mathbf{H}^{\mathbf{G}}_{k,l'}$ par

(II.8.26) $$\mathbf{H}^{\mathbf{G}}_{k,l'} = -i\sqrt{\overline{\varepsilon}_k}\mathbf{G}_{k,l'}e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}\left(V_{k,l'}\,.\mathbf{x}\right)}$$

Un calcul élémentaire montre que l'on a la proposition 12 suivante.

Proposition 12 *Les couples $(\mathbf{E}^{\mathbf{F}}_{k,l'}, \mathbf{H}^{\mathbf{F}}_{k,l'})$ et $(\mathbf{E}^{\mathbf{G}}_{k,l'}, \mathbf{H}^{\mathbf{G}}_{k,l'})$ vérifient les équations de Maxwell adjointes sans source dans Ω_k, soit*

$$\begin{aligned} \nabla\wedge\mathbf{E}' - i\omega\overline{\mu}_k\mathbf{H}' &= 0 \\ \nabla\wedge\mathbf{H}' + i\omega\overline{\varepsilon}_k\mathbf{E}' &= 0\ . \end{aligned}$$

Ces couples vérifient les relations de la définition de V_h (14) et permettent donc de définir l'espace de discrétisation V_h.

Proposition 13 *Les conditions de la définition 15, soit,*

(II.8.27) $$\begin{cases} \mathbf{E}^0_{k,l'}\mathbf{E}^0_{k,l'} = 1 \\ V_{k,l'}V_{k,l'} = 1 \\ \mathbf{E}^0_{k,l'}V_{k,l'} = 0 \\ (\mathbf{E}^0_{k,l'}, V_{k,l'}) \in \mathbb{R}^2\ , \end{cases}$$

impliquent

(II.8.28) $$\mathbf{E}^{\mathbf{F}}_{k,l'}\overline{\mathbf{E}^{\mathbf{G}}_{k,l'}} = 0 \ .$$

Cela signifie que, pour tout k de 1 à K et tout $l^{'}$ de 1 à p, les fonctions $\mathbf{E}^{\mathbf{F}}_{k,l'}$ et $\mathbf{E}^{\mathbf{G}}_{k,l'}$ sont orthogonales dans $\mathbb{C}^3$ pour tout $\mathbf{X}$ donné dans Ω_k donc dans $\mathcal{H}(\Omega_k, \partial\Omega_k)$.

Preuve. Le terme $(\mathbf{E}^0_{k,l'}\overline{\mathbf{E}^0_{k,l'} \wedge V_{k,l'}})$ est nul car c'est le produit mixte de trois vecteurs réels dont deux sont égaux. En outre, la famille $(\mathbf{E}^0_{k,l'}, V_{k,l'}, \mathbf{E}^0_{k,l'} \wedge V_{k,l'})$ forme une base orthonormée directe de $\mathbb{R}^3$ donc

$$|\mathbf{E}^0_{k,l'} \wedge V_{k,l'}|^2 = 1 \ .$$

On calcule alors

$$\mathbf{E}^{\mathbf{F}}_{k,l'}\overline{\mathbf{E}^{\mathbf{G}}_{k,l'}} = 0 \ .$$

On a

$$\begin{aligned}
\mathbf{E}^{\mathbf{F}}_{k,l'}\overline{\mathbf{E}^{\mathbf{G}}_{k,l'}} &= \left(\mathbf{E}^0_{k,l'} + i\mathbf{E}^0_{k,l'} \wedge V_{k,l'}\right)\left(\mathbf{E}^0_{k,l'} + i\mathbf{E}^0_{k,l'} \wedge V_{k,l'}\right) \\
&= (\mathbf{E}^0_{k,l'})^2 - (\mathbf{E}^0_{k,l'} \wedge V_{k,l'})^2 + 2i\mathbf{E}^0_{k,l'}\left(\mathbf{E}^0_{k,l'} \wedge V_{k,l'}\right) \\
&= 1 - 1 + 2i \times 0
\end{aligned}$$

□

Lemme 17 *La famille $(\mathbf{E}^{\mathbf{F}}_{k,l}, \mathbf{E}^{\mathbf{G}}_{k,l})_{l=1,p}$ de cardinal $L = 2p$ est libre dans Ω_k.*

Preuve. Supposons qu'il existe une famille de complexes $(\alpha_l, \alpha_{l+p})_{l=1,p}$ de cardinal $2p$ telle que si l'on définit la fonction vectorielle P par

$$P(\mathbf{X}) = \sum_{l=1}^{p} \alpha_l \mathbf{E}^{\mathbf{F}}_{k,l}(\mathbf{X}) + \alpha_{l+p}\mathbf{E}^{\mathbf{G}}_{k,l}(\mathbf{X})$$

on ait

$$\begin{cases} \forall \mathbf{X} \in \Omega_k \\ P(\mathbf{X}) = 0 \ . \end{cases}$$

On définit les trois fonctions $P_n \in C^\infty(\Omega_k)$ pour $n = 1$ à 3 telles que

$$P(\mathbf{X}) = (P_1(\mathbf{X}), P_2(\mathbf{X}), P_3(\mathbf{X})) \ .$$

On peut appliquer l'opérateur de dérivation autant de fois que souhaité aux fonctions P_n. Par exemple, pour une dérivation, on obtient

$$\frac{\partial P_n}{\partial \mathbf{X}_n} = i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}\sum_{l=1}^{p}(V_{k,l})_n \times \left(\alpha_l(\mathbf{E}^{\mathbf{F}}_{k,l})_n + \alpha_{l+p}(\mathbf{E}^{\mathbf{G}}_{k,l})_n\right) \ .$$

On définit la matrice carrée M_n de taille $p \times p$ et le vecteur $T_n(\mathbf{X})$ de taille p par

$$\left| \begin{array}{l} M^n_{l,j} = \left(i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,j})_n\right)^l \\ T^j_n(\mathbf{X}) = (\alpha_j \mathbf{E}^{\mathbf{F}}_{k,j} + \alpha_{j+p}\mathbf{E}^{\mathbf{G}}_{k,j})_n \end{array} \right.$$

en remarquant que M_n ne dépend pas de $\mathbf{X}$. On a donc

$$\begin{cases} \forall \mathbf{X} \in \Omega_k \\ \displaystyle\sum_{j=1}^{p} M^n_{l,j} T^j_n(\mathbf{X}) = 0 \end{cases}$$

Le déterminant de la matrice M_n est un Vandermonde ce qui prouve que M_n est inversible [57]. On a alors

$$\forall \mathbf{X} \in \Omega_k,\ \forall 1 \leq j \leq p,\ \forall 1 \leq n \leq 3,\ T_n^j(\mathbf{X}) = 0$$

Cette relation signifie que les polarisations de deux ondes planes de même direction de propagation sont liées dans Ω_k, soit

$$\forall \mathbf{X} \in \Omega_k,\ \forall l \leq p \ \ \alpha_l \mathbf{E}^{\mathbf{F}}_{k,l}(\mathbf{X}) + \alpha_{l+p}\mathbf{E}^{\mathbf{G}}_{k,l}(\mathbf{X}) = 0\ ,$$

ce qui implique $(\alpha_l, \alpha_{l+p}) = (0,0)$ pour tout l de 1 à p puisque les polarisations sont orthogonales d'après (II.8.28). □

II.8.2.2 Construction d'un espace d'approximation particulier.

Nous avons alors choisi de construire notre espace d'approximation V_h à partir

i) d'une partition de Ω en K éléments réguliers Ω_k soit tétraédriques soit hexaédriques. Notons que ce maillage définit l'espace fonctionnel "continu" V de la formulation variationnelle ultra-faible (II.7.84). La terminologie "éléments réguliers" signifie que le maillage vérifie les hypothèses H1, H2 et H3 ([17]).

ii) Sur chaque maille Ω_k on considère les p directions de propagation $V_{k,l'}$ vérifiant (II.8.20), l'indice l' variant de 1 à p, et p vecteurs polarisation réels et normés $\mathbf{E}^0_{k,l'}$ orthogonaux aux directions de propagation respectives. Ces vecteurs définissent les p fonctions de type $\mathbf{F}$, $(\mathbf{E}^{\mathbf{F}}_{k,l'}, \mathbf{H}^{\mathbf{F}}_{k,l'})$ ((II.8.22) et (II.8.25)) et les p fonctions associées de type $\mathbf{G}$, $(\mathbf{E}^{\mathbf{G}}_{k,l'}, \mathbf{E}^{\mathbf{G}}_{k,l'})$ ((II.8.23) et (II.8.26)).

Les fonctions de base $\mathcal{Z}_{kl}$ sont construites pour l variant de 1 à $L = 2p$ par les relations (II.8.5) à l'aide des fonctions $(\mathbf{E}'_{kl}, \mathbf{H}'_{kl}) = (\mathbf{E}^{\mathbf{F}}_{k,l}, \mathbf{H}^{\mathbf{F}}_{k,l})$ pour $1 \leq l \leq p$ et des fonctions $(\mathbf{E}'_{kl}, \mathbf{H}'_{kl}) = (\mathbf{E}^{\mathbf{G}}_{k,l-p}, \mathbf{H}^{\mathbf{G}}_{k,l-p})$ pour $p+1 \leq l \leq 2p$. On a donc

$$\text{(II.8.29)} \quad \begin{cases} (\mathcal{Z}_{kl})_{|\partial\Omega_j} = 0 \text{ si } k \neq j \\ (\mathcal{Z}_{kl})_{|\Sigma_{k,j}} = \left(\sqrt{\varepsilon_{kj}}\sqrt{\overline{\mu}_k}\mathbf{F}_{k,l} \wedge \nu_k + i\sqrt{\mu_{kj}}\sqrt{\overline{\varepsilon}_k}\,(\mathbf{F}_{k,l} \wedge \nu_k) \wedge \nu_k\right) e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,l}\,.\,\mathbf{X})} \\ \qquad \text{pour } 1 \leq l \leq p\ . \\ (\mathcal{Z}_{k,l})_{|\Sigma_{k,j}} = \left(\sqrt{\varepsilon_{kj}}\sqrt{\overline{\mu}_k}\mathbf{G}_{k,l-p} \wedge \nu_k - i\sqrt{\mu_{kj}}\sqrt{\overline{\varepsilon}_k}\,(\mathbf{G}_{k,l-p} \wedge \nu_k) \wedge \nu_k\right) e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,l-p}\,.\,\mathbf{X})} \\ \qquad \text{pour } 2p \geq l > p\ . \end{cases}$$

Lemme 18 *Les fonctions de base choisies sont bien indépendantes. En effet, d'après le lemme 17 la famille de fonctions analytiques sur Ω_k, $(\mathbf{E}^{\mathbf{F}}_{k,l}, \mathbf{E}^{\mathbf{G}}_{k,l})_{l=1,p}$, de cardinal $2p$ est linéairement indépendante sur Ω_k donc sur $\mathcal{H}(\Omega_k, \partial\Omega_k)$. Le lemme 15 montre alors que la famille $(\mathcal{Z}_{kl})_{l=1,2p}$ est libre. Nous sommes donc dans le cadre des hypothèses du lemme 16 qui assure l'inversibilité de la matrice D du système linéaire discret (II.8.2).*

Proposition 14 *La donnée des p vecteurs réels $\mathbf{E}^0_{k,l'}$ et $V_{k,l'}$ définit complètement les fonctions de base $\mathcal{Z}_{kl}$ et donc la matrice hermitienne de produit scalaire D_k par*

$$D_k^{l,m} = (\mathcal{Z}_{km}, \mathcal{Z}_{kl})_V\ .$$

Montrons que le déterminant de D_k ne dépend pas du choix des vecteurs $\mathbf{E}^0_{k,l'}$ à partir d'une direction de propagation $V_{k,l'}$ fixée.

Preuve. Dans la définition (15), le choix des polarisation $\mathbf{E}^0_{k,l'}$ n'est pas unique pour une direction de propagation $V_{k,l'}$ fixée. En effet, considérons le vecteur $\mathbf{E}^1_{k,l'}$ réel unitaire orthogonal à la direction de propagation $V_{k,l'}$ tel que

$$\mathbf{E}^1_{k,l'} = \sin\theta_{l'}\,\mathbf{E}^0_{k,l'} + \cos\theta_{l'}\,V_{k,l'} \wedge \mathbf{E}^0_{k,l'}\ .$$

Le vecteur $\mathbf{E}^1_{k,l'}$ définit

$$\mathcal{Z}_{kl}(\mathbf{E}^1_{k,l}) = e^{i\theta_l}\mathcal{Z}_{kl}(\mathbf{E}^0_{k,l})$$

pour $1 \leq l \leq p$ et

$$\mathcal{Z}_{kl}(\mathbf{E}^1_{k,l-p}) = e^{-i\theta_{l-p}} \mathcal{Z}_{kl}(\mathbf{E}^0_{k,l-p})$$

pour $p+1 \leq l \leq 2p$.

La matrice N_k, définie à l'aide des polarisation réelles $\mathbf{E}^1_{k,l'}$ et des directions de propagations $V_{k,l'}$, est obtenue à partir de la matrice D_k, donnée par $(\mathbf{E}^0_{k,l'}, V_{k,l'})_{1 \leq l' \leq p}$, par les relations

$$N_k^{l,m} = e^{i\theta_l} D_k^{l,m} \overline{e^{i\theta_m}}$$

en définissant $\theta_l = -\theta_{l-p}$ pour $p+1 \leq l \leq p$. On vérifie par un calcul élémentaire que $N_k = \Theta D_k \overline{\Theta}$ où $\Theta_{l,m} = \delta_{l,m} e^{i\theta_m}$. La matrice Θ est de déterminant 1. □

Remarque 39 (Sur le choix du maillage) Comme pour le choix de l'espace pour le problème de Helmholtz, la partition proposée se réalise simplement à l'aide d'un mailleur. Notons que les éléments tétraédriques sont mieux adaptés que les hexaèdres pour mailler un objet quelconque. Nous proposons de fixer le type des éléments du maillage pour faciliter la mise en œuvre informatique sur un ordinateur à architecture vectorielle. Notons en outre que l'utilisation d'éléments tétraédriques minimise les couplages entre éléments voisins, ce qui réduit à la fois la place mémoire informatique et le temps de calcul. Pour K éléments et L fonctions de base par élément, nous estimons sommairement que la taille mémoire nécessaire est proportionnelle à

$$K(L^2)N + K(\frac{L(L+1)}{2}) \tag{II.8.30}$$

où N est le nombre de faces par élément, N vaut 4 pour des tétraèdres, 5 pour des pentaèdres, 6 pour des hexaèdres. L'assemblage de la matrice est deux fois plus long sur une face à quatre sommets par rapport à une face à trois sommets. En outre une itération de l'algorithme itératif de résolution du système matriciel (I.2.39) demande 50% de calculs en plus pour un maillage en hexaèdres que pour un maillage en tétraèdres (pour le même nombre d'éléments). Le nombre de degrés de libertés reste inchangé, égal à $2pK$.

Remarque 40 (Sur le choix des fonctions de base) Comme pour le problème de Helmholtz, le choix de fonctions de base issues d'ondes planes se justifie par le fait que les termes des matrices du système linéaire se calculent par des formules analytiques. Fixer le nombre de fonctions de base facilite la mise en œuvre informatique sur un ordinateur à architecture vectorielle. L'utilisation de fonctions de bases identiques sur tous les éléments permet une amélioration du temps d'assemblage informatique des matrices du système linéaire, et éventuellement une réduction du stockage. De plus, nous exposons dans la section II.8.2.3 l'intérêt spécifique aux équations de Maxwell de choisir des fonctions de base issues de fonctions aux polarisations complexes conjuguées. Ce choix particulier permet de réduire au mieux la taille mémoire du système linéaire. Nous montrons, section II.8.2.3, que pour un problème dans le vide, la place mémoire du système linéaire est proportionnelle à

$$2K(p^2)N + K\,(p(p+1)) \tag{II.8.31}$$

au lieu de (II.8.30) pour $L = 2p$, soit approximativement une division par deux de la place mémoire.

II.8.2.3 Particularité avantageuse de l'espace d'approximation.

La section II.7.1.3 montre le découplage des équations de Maxwell dans Ω et de la condition aux limites absorbante. C'est ce découplage qui a inspiré l'utilisation de polarisations complexes conjuguées présentée dans la section II.8.2.2. Nous allons montrer que si le domaine Ω est constitué d'un milieu à caractéristiques ε et μ réelles et constantes, alors presque la moitié des termes de la matrice du système linéaire sont nuls. Plus précisément,

Lemme 19 *Soit une face Σ_{kj} telle que ε et μ sont réels et constants entre Ω_k et Ω_j.*

– *Si $j = k$, alors*

$$\begin{cases} \forall 1 \leq l,m \leq p \\ C^{l,m}_{k,k} = C^{l+p,m+p}_{k,k} = 0\ . \end{cases} \tag{II.8.32}$$

– *Si* $j \neq k$, *alors*

$$\text{(II.8.33)} \qquad \begin{cases} \forall 1 \leq l, m \leq p \\ C_{k,j}^{l,m+p} = C_{k,j}^{l+p,m} = 0 \ . \end{cases}$$

Soit un élément Ω_k *tel que* ε *et* μ *sont réels constants entre* Ω_k *et tous ses voisins, alors*

–

$$\text{(II.8.34)} \qquad \begin{cases} \forall 1 \leq l, m \leq p \\ D_{k,k}^{l,m+p} = D_{k,k}^{l+p,m} = 0 \ . \end{cases}$$

Preuve. [De (II.8.34)] D'après (II.8.6) on a

$$\text{(II.8.35)} \qquad D_{k,k}^{l,m+p} = (\mathcal{Z}_{k,m+p}, \mathcal{Z}_{k,l})_V \ .$$

En utilisant les fonctions $\mathbf{F}_{k,l}$ et $\mathbf{G}_{k,m}$ (cf (II.8.21)), on sait, d'après (II.8.29), que l'on a

$$\text{(II.8.36)} \qquad \begin{aligned} \mathcal{Z}_{k,l} &= \left(\sqrt{\varepsilon_{kj}}\sqrt{\overline{\mu}_k}\mathbf{F}_{k,l} \wedge \nu_k + i\sqrt{\mu_{kj}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{F}_{k,l} \wedge \nu_k\right) \wedge \nu_k\right) e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,l} \, . \, \mathbf{X})} \\ \mathcal{Z}_{k,m+p} &= \left(\sqrt{\varepsilon_{kj}}\sqrt{\overline{\mu}_k}\mathbf{G}_{k,m} \wedge \nu_k - i\sqrt{\mu_{kj}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{G}_{k,m} \wedge \nu_k\right) \wedge \nu_k\right) e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,m} \, . \, \mathbf{X})} \ , \end{aligned}$$

et, dans le cas de matériaux réels et identiques sur Ω_k et Ω_j on a, pour tout $j = j(k)$

$$\text{(II.8.37)} \qquad \begin{cases} \varepsilon_{kj} = (\varepsilon_k)_{|\Sigma_{k,j}} \\ \mu_{kj} = (\mu_k)_{|\Sigma_{k,j}} \\ \overline{\varepsilon}_k = \varepsilon_k \\ \overline{\mu}_k = \mu_k \ , \end{cases}$$

ce qui, dans (II.8.36), donne

$$\text{(II.8.38)} \qquad \begin{aligned} \mathcal{Z}_{k,l} &= \sqrt{\varepsilon_k\mu_k}\left(\mathbf{F}_{k,l} \wedge \nu_k + i\left(\mathbf{F}_{k,l} \wedge \nu_k\right) \wedge \nu_k\right) e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,l} \, . \, \mathbf{X})} \\ \mathcal{Z}_{k,m+p} &= \sqrt{\varepsilon_k\mu_k}\left(\mathbf{G}_{k,m} \wedge \nu_k - i\left(\mathbf{G}_{k,m} \wedge \nu_k\right) \wedge \nu_k\right) e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,m} \, . \, \mathbf{X})} \ . \end{aligned}$$

Dans une base orthonormée de $\mathbb{R}^3$ dont le troisième vecteur est la normale ν_k, on a

$$\text{(II.8.39)} \qquad \begin{aligned} \mathcal{Z}_{k,l} &= \sqrt{\varepsilon_k\mu_k}\left(\begin{bmatrix} \mathbf{F}_{k,l}^2 \\ -\mathbf{F}_{k,l}^1 \\ 0 \end{bmatrix} + i\begin{bmatrix} -\mathbf{F}_{k,l}^1 \\ -\mathbf{F}_{k,l}^2 \\ 0 \end{bmatrix}\right) e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,l} \, . \, \mathbf{X})} \\ \mathcal{Z}_{k,m+p} &= \sqrt{\varepsilon_k\mu_k}\left(\begin{bmatrix} \mathbf{G}_{k,m}^2 \\ -\mathbf{G}_{k,m}^1 \\ 0 \end{bmatrix} - i\begin{bmatrix} -\mathbf{G}_{k,m}^1 \\ -\mathbf{G}_{k,m}^2 \\ 0 \end{bmatrix}\right) e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,m} \, . \, \mathbf{X})} \end{aligned}$$

que l'on simplifie en

$$\text{(II.8.40)} \qquad \begin{aligned} \mathcal{Z}_{k,l} &= \sqrt{\varepsilon_k\mu_k}\begin{bmatrix} \mathbf{F}_{k,l}^2 - i\mathbf{F}_{k,l}^1 \\ -\mathbf{F}_{k,l}^1 - i\mathbf{F}_{k,l}^2 \\ 0 \end{bmatrix} e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,l} \, . \, \mathbf{X})} \\ &= \sqrt{\varepsilon_k\mu_k}\left(\mathbf{F}_{k,l}^1 + i\mathbf{F}_{k,l}^2\right)\begin{bmatrix} -i \\ -1 \\ 0 \end{bmatrix} e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,l} \, . \, \mathbf{X})} \\ \mathcal{Z}_{k,m+p} &= \sqrt{\varepsilon_k\mu_k}\begin{bmatrix} \mathbf{G}_{k,m}^2 + i\mathbf{G}_{k,m}^1 \\ -\mathbf{G}_{k,m}^1 + i\mathbf{G}_{k,m}^2 \\ 0 \end{bmatrix} e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,m} \, . \, \mathbf{X})} \\ &= \sqrt{\varepsilon_k\mu_k}\left(\mathbf{G}_{k,m}^1 - i\mathbf{G}_{k,m}^2\right)\begin{bmatrix} i \\ -1 \\ 0 \end{bmatrix} e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,m} \, . \, \mathbf{X})} \ . \end{aligned}$$

Il est clair que le produit scalaire dans $\mathbb{C}^3$

$$
(\begin{bmatrix} i \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} -i \\ -1 \\ 0 \end{bmatrix})_{\mathbb{C}^3}
$$

est nul. Ceci montre que la contribution de l'interface $\Sigma_{k,j}$ au calcul de D est nul. S'il en est de même sur les autres faces de Ω_k, alors D est nul. □

Preuve. [De (II.8.33)] D'après la section II.8.1.2, on a

$$
C^{l,m+p}_{k,j\neq k} = (\Pi\mathcal{Z}_{j,m+p}, F\mathcal{Z}_{k,l})_V \ . \tag{II.8.41}
$$

En utilisant les fonctions $\mathbf{F}_{k,l}$ et $\mathbf{G}_{j,m}$ (cf (II.8.21)), on calcule sans difficulté d'après (II.8.8) que

$$
\begin{aligned}
F\mathcal{Z}_{k,l} &= \left(-\sqrt{\varepsilon_{kk}}\sqrt{\overline{\mu}_k}\mathbf{F}_{k,l}\wedge\nu_k + i\sqrt{\mu_{kk}}\sqrt{\overline{\varepsilon}_k}\,(\mathbf{F}_{k,l}\wedge\nu_k)\wedge\nu_k\right)e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,l}\,.\,\mathbf{X})} \\
\Pi\mathcal{Z}_{j,m+p} &= \left(+\sqrt{\varepsilon_{jj}}\sqrt{\overline{\mu}_j}\mathbf{G}_{j,m}\wedge\nu_j - i\sqrt{\mu_{jj}}\sqrt{\overline{\varepsilon}_j}\,(\mathbf{G}_{j,m}\wedge\nu_j)\wedge\nu_j\right)e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{j,m}\,.\,\mathbf{X})}
\end{aligned} \tag{II.8.42}
$$

et, dans le cas de matériaux réels et identiques sur Ω_k et Ω_j (avec $j\neq k$), les relations (II.8.37) induisent

$$
\begin{aligned}
F\mathcal{Z}_{k,l} &= \sqrt{\varepsilon_k\mu_k}\,(-\mathbf{F}_{k,l}\wedge\nu_k + i\,(\mathbf{F}_{k,l}\wedge\nu_k)\wedge\nu_k)\,e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,l}\,.\,\mathbf{X})} \\
\Pi\mathcal{Z}_{j,m+p} &= \sqrt{\varepsilon_k\mu_k}\,(-\mathbf{G}_{j,m}\wedge\nu_k - i\,(\mathbf{G}_{j,m}\wedge\nu_k)\wedge\nu_k)\,e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{j,m}\,.\,\mathbf{X})} \ .
\end{aligned} \tag{II.8.43}
$$

Dans une base orthonormée de $\mathbb{R}^3$ dont le troisième vecteur est la normale ν_k, on a

$$
\begin{aligned}
F\mathcal{Z}_{k,l} &= \sqrt{\varepsilon_k\mu_k}\begin{bmatrix} -\mathbf{F}^2_{k,l} - i\mathbf{F}^1_{k,l} \\ \mathbf{F}^1_{k,l} - i\mathbf{F}^2_{k,l} \\ 0 \end{bmatrix} e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,l}\,.\,\mathbf{X})} \\
&= \sqrt{\varepsilon_k\mu_k}\left(\mathbf{F}^1_{k,l} - i\mathbf{F}^2_{k,l}\right)\begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix} e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,l}\,.\,\mathbf{X})} \\
\Pi\mathcal{Z}_{j,m+p} &= \sqrt{\varepsilon_k\mu_k}\begin{bmatrix} -\mathbf{G}^2_{j,m} + i\mathbf{G}^1_{j,m} \\ \mathbf{G}^1_{j,m} + i\mathbf{G}^2_{j,m} \\ 0 \end{bmatrix} e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{j,m}\,.\,\mathbf{X})} \\
&= \sqrt{\varepsilon_k\mu_k}\left(\mathbf{G}^1_{j,m} + i\mathbf{G}^2_{j,m}\right)\begin{bmatrix} i \\ 1 \\ 0 \end{bmatrix} e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{j,m}\,.\,\mathbf{X})} \ .
\end{aligned} \tag{II.8.44}
$$

Le produit scalaire dans $\mathbb{C}^3$

$$
(\begin{bmatrix} i \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix})_{\mathbb{C}^3}
$$

est nul. Ceci montre, d'après (II.8.41) que

$$
C^{l,m+p}_{k,j\neq k} = (\Pi\mathcal{Z}_{j,m+p}, F\mathcal{Z}_{k,l})_V = 0 \ . \tag{II.8.45}
$$

De la même manière, on montre que

$$
C^{l+p,m}_{k,j\neq k} = 0 \ .
$$

□

Preuve. [De (II.8.32)] De (II.8.8), on a la relation

$$
C^{l,m}_{k,k} = (\Pi\mathcal{Z}_{k,m}, F\mathcal{Z}_{k,l})_V
$$

et de même pour les indices $l+p$ et $m+p$. Dans le cas d'un matériau réel sur Σ_{kk} on a

$$\text{(II.8.46)} \qquad \begin{cases} \overline{\varepsilon}_k = \varepsilon_k \\ \overline{\mu}_k = \mu_k \ . \end{cases}$$

En combinant les résultats précédents (II.8.40) et (II.8.44) dans une base dont la troisième composante est la normale ν_k à $\Sigma_{k,k}$, on obtient le résultat suivant :

$$\text{(II.8.47)} \qquad \begin{aligned} F\mathcal{Z}_{k,l} &= \sqrt{\varepsilon_k\mu_k}\left(\mathbf{F}^1_{k,l} - i\mathbf{F}^2_{k,l}\right)\begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix} e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,l}\,.\,\mathbf{X})} \\ F\mathcal{Z}_{k,l+p} &= \sqrt{\varepsilon_k\mu_k}\left(\mathbf{G}^1_{k,l} + i\mathbf{G}^2_{k,l}\right)\begin{bmatrix} i \\ 1 \\ 0 \end{bmatrix} e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,l}\,.\,\mathbf{X})} \\ \Pi\mathcal{Z}_{k,m}&\sqrt{\varepsilon_k\mu_k}\left(\mathbf{G}^1_{k,m} + i\mathbf{G}^2_{k,m}\right)\begin{bmatrix} -i \\ -1 \\ 0 \end{bmatrix} e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,m}\,.\,\mathbf{X})} \\ \Pi\mathcal{Z}_{k,m+p}&\sqrt{\varepsilon_k\mu_k}\left(\mathbf{G}^1_{k,m} - i\mathbf{G}^2_{k,m}\right)\begin{bmatrix} i \\ -1 \\ 0 \end{bmatrix} e^{i\omega\sqrt{\overline{\varepsilon}_k\overline{\mu}_k}(V_{k,m}\,.\,\mathbf{X})} \ . \end{aligned}$$

Nous calculons alors

$$(\begin{bmatrix} -i \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix})_{\mathbb{C}^3} = 0, \quad (\begin{bmatrix} i \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} i \\ 1 \\ 0 \end{bmatrix})_{\mathbb{C}^3} = 0, \quad (\begin{bmatrix} i \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix})_{\mathbb{C}^3} = -2 \ ,$$

ce qui montre (II.8.32), mais, qu'en revanche, $C_k^{l,m+p}$ et $C_k^{l+p,m}$ ne sont pas forcément nuls. □

II.8.3 Un choix important, le choix d'une base de V_h.

Pour une maille Ω_k donnée nous cherchons à construire les p directions de propagation $V_{k,l}$ (avec $1 \leq l \leq p$) et les p polarisations associées $\mathbf{E}^0_{k,l}$ pour les fonctions $\mathbf{E}^{\mathbf{F}}_{k,l}$ et $\mathbf{E}^{\mathbf{G}}_{k,l}$ définies par (II.8.22) et (II.8.23), définition 15 p. 104.

Le choix des directions $V_{k,l}$ et des polarisations $\mathbf{E}^0_{k,l}$ est en 3-D beaucoup plus laborieux qu'en 2-D. En effet, la notion d'équirépartition des directions des ondes planes ne se fait plus simplement. La notion de maillage régulier (en éléments identiques par rotations) n'existe que pour un nombre fini d'éléments et de nœuds. Nous ne présentons pas d'algorithme de construction optimal, mais nous montrons comment nous choisissons ces directions dans les cas qui nous intéressent.

Nous présentons d'abord comment les polarisations sont choisies à partir des directions. Nous montrons ensuite comment choisir les directions de propagation sur un élément donné, le nombre de directions p étant fixé. Enfin, nous montrons comment choisir les directions pour tous les éléments en choisissant une répartition aléatoire d'un élément à un autre.

II.8.3.1 Choix des polarisations.

Les polarisations sont des vecteurs unitaires, orthogonaux aux directions. Nous déterminons la polarisation $\mathbf{E}^0_{k,l}$ par (où p_m est la précision machine) l'algorithme suivant :

$$\left|\begin{array}{l}
\text{si } (V_{k,l})_1^2 + (V_{k,l})_2^2 \geq 2/3 - 20 * p_m,\ \text{alors } \mathbf{E}^0_{k,l} = \dfrac{1}{\sqrt{(V_{k,l})_1^2 + (V_{k,l})_2^2}} \begin{bmatrix} -(V_{k,l})_2 \\ (V_{k,l})_1 \\ 0 \end{bmatrix},\ \text{sinon,} \\
\text{si } (V_{k,l})_3^2 + (V_{k,l})_2^2 \geq 2/3 - 20 * p_m,\ \text{alors } \mathbf{E}^0_{k,l} = \dfrac{1}{\sqrt{(V_{k,l})_3^2 + (V_{k,l})_2^2}} \begin{bmatrix} 0 \\ -(V_{k,l})_3 \\ (V_{k,l})_2 \end{bmatrix},\ \text{sinon,} \\
\text{puisque } (V_{k,l})_1^2 + (V_{k,l})_3^2 \geq 2/3 - 20 * p_m, \mathbf{E}^0_{k,l} = \dfrac{1}{\sqrt{(V_{k,l})_1^2 + (V_{k,l})_3^2}} \begin{bmatrix} -(V_{k,l})_3 \\ 0 \\ (V_{k,l})_1 \end{bmatrix}.
\end{array}\right.$$

Nous représentons figure II.8.2 quelques polarisations dans les cas de 20 et 44 fonctions de base. Nous voyons clairement que la répartition des polarisations n'est pas homogène entre les trois plans (xOy), (yOz) et (xOz) dans lesquels les polarisations se trouvent nécessairement. Une répartition homogène oblige à choisir des coefficients différents dans l'algorithme conditionnel ci-dessus. Dans le cas où l'on fait varier les directions de propagation de façon aléatoire d'un élément à un autre, nous voyons apparaître l'ensemble des polarisations figure II.8.5.

II.8.3.2 Choix des directions de propagation sur la sphère unité.

II.8.3.2.1 Choix des directions de propagation de référence.

Nous donnons des directions de référence des fonctions de base qui serviront dans le choix final des directions. Nous avons choisi de prendre en priorité des directions équiréparties, sinon le découpage S_n de la sphère, sinon des directions qui nous semblent assez bien découper la sphère, et enfin nous proposons un algorithme général de choix des directions, algorithme qu'il serait intéressant d'améliorer à l'avenir.

i) Equiréparties. Dans certains cas, nous avons pu choisir des directions équiréparties dans l'espace. Pour trois fonctions de base les directions forment un triangle équilatéral dans le plan équatorial. Pour quatre, un tétraèdre régulier. Pour six, deux tétraèdres réguliers assemblés par une face. Pour huit, un cube. Pour douze on a pris les nœuds de l'icosaèdre.

ii) Découpage S_n de la sphère en $n(n+2)$ directions. Ce découpage est bien connu des neutroniciens. Il consiste à effectuer un découpage de l'espace des phases (ϕ, μ) où la quantité μ est liée à l'angle θ par $\mu = \cos\theta$. Les angles (θ, ϕ) donnent les coordonnées d'un point de la sphère unité dans un repère cartésien par $(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. L'espace des phases (ϕ, μ) est donc le quadrilatère $[0, \pi] \times [-1, +1]$. L'avantage de cette répartition est qu'elle découpe l'espace des phases en quadrangles d'aires égales. Les neutroniciens l'utilisent pour des propriétés de quadratures des moments du vecteur courant $\vec{\mathbf{\Omega}}$ défini par

$$(\sqrt{1-\mu^2}\cos\phi, \sqrt{1-\mu^2}\sin\phi, \mu)\ .$$

Ce découpage nous a semblé assez bien "équiréparti" au sens qu'il produit une partition quadrangulaire de la sphère unité qui est assez régulière. Nous renvoyons aux ouvrages de neutronique qui utilisent ce découpage pour des explications plus précises, nous ne donnons ici que le tableau (II.8.1) de construction des directions $V_{m,l}$ de la demi-sphère unité. L'indice m correspond à une latitude, l'indice l à une longitude. La figure (II.8.1) représente le découpage pour $n = 16$ (image Guillaume Pottier). Les directions sont proches des barycentres des quadrangles.

iii) Construction "à la main". Pour des nombres différents de fonctions de base, nous avons pris dans certains cas des directions construites à la main. Ainsi, pour cinq fonctions de base, nous avons pris les trois fonctions équiréparties du cas ci-dessus, puis nous avons rajouté les deux pôles. Pour sept fonctions nous avons pris celles du tétraèdre régulier plus celles du plan équatorial de façon à

TAB. II.8.1 – Récapitulatif du découpage S_n.

$\boxed{\mu}$	$1 \leq m \leq \frac{n}{2}$	$\frac{n}{2} < m \leq n$
μ_m	$\left(\frac{4m^2}{n(n+2)} - 1\right)\sqrt{\frac{n(n+2)}{n(n+2)-2}}$	$\left(-\frac{4(n+1-m)^2}{n(n+2)} + 1\right)\sqrt{\frac{n(n+2)}{n(n+2)-2}}$
$\boxed{\phi}$	$1 \leq l \leq 2m$	$1 \leq l \leq 2(n+1-m)$
$\phi_{m,l}$	$\pi\left(1 - \frac{2l-1}{4m}\right)$	$\pi\left(1 - \frac{2l-1}{4(n+1-m)}\right)$

FIG. II.8.1 – Découpage $S_{n=16}$ de la sphère unité.

former, vu de dessus, un point à l'origine et deux triangles obtenus l'un de l'autre par homothétie et symétrie par rapport à l'origine. Pour neuf fonctions nous avons pris les trois directions du plan équatorial auxquelles nous avons rajouté trois fonctions dans un plan parallèle à une latitude de 45 degrés dans un repère terrestre. Ces directions forment un triangle équilatéral dans leur plan et leur projection sur le plan équatorial est un triangle aux côtés parallèles aux côtés du triangle du plan équatorial, mais par les faces les plus éloignées. Les trois dernières directions sont symétriques par rapport au plan équatorial. Enfin pour vingt fonctions nous avons pris les nœuds du dodécahèdre sachant qu'ils ne forment pas un maillage équiréparti de la sphère.

iv) Génération automatique non équirépartie. Pour d'autres nombres de fonctions de base nous construisons les directions par un algorithme qui, asymptotiquement, donne des directions équiréparties. Nous prenons le pôle Nord et le pôle Sud puis un nombre N pour l'instant inconnu de fonctions de base aux directions équiréparties dans le plan équatorial. A une latitude d'angle ϕ égal à $(N/4)*4$ (division entière) fois l'angle entre deux directions consécutives du plan équatorial nous prenons un nombre $[N*\cos(\phi)]$ (rappelons que $[\alpha]$ est le symbole partie entière de α) de directions équiréparties dans leur plan. Cet algorithme permet une anisotropie polaire faible par rapport à un simple maillage quadrangulaire en latitude-longitude, pôles exclus. En même temps que nous construisons les directions de l'hémisphère Nord nous construisons les directions symétriques de l'hémisphère Sud. Nous ne tombons pas forcément sur exactement le bon nombre de fonctions de base. Nous calculons d'abord le nombre N donnant le plus petit nombre de directions construites par l'algorithme supérieur au nombre de fonctions de bases demandé. Nous stoppons ensuite le procédé lorsque le nombre demandé est atteint ce qui peut créer une légère anisotropie supplémentaire aux pôles. Par exemple pour 17 fonctions de base demandées l'algorithme est prêt à en constituer 20. Pour 40 l'algorithme serait pleinement efficace pour 44. A la place de 90 fonctions il serait mieux d'en demander 95, pour 200 ce serait 203 et 800 ce serait 814. La figure II.8.2 pour 300 directions est typique : en coupe équatoriale nous observons un "trou" aux deux pôles dans les répartitions des directions. C'est le problème de la couche d'ozone...

II.8.3.2.2 Exemples.

Nous appelons vue géostationnaire une projection sur un plan dans lequel les pôles (de la terre) sont des diamètres de la projection. Nous appelons vue "soleil d'été" une projection en perspective selon le vecteur donné par ($\theta = \pi/4$, $\phi = \pi/4$) et en doublant les distances sur l'axe z). L'appellation est justifiée par le fait que l'on aperçoit le pôle Nord en angle presque rasant. Les figures (II.8.2) représentent les vecteurs directeurs et les polarisations associées pour certains nombres de fonctions de base demandées. En se déplaçant de gauche à droite puis de haut en bas, les deux premières vues sont dans le plan équatorial, la troisième est une vue géostationnaire, toutes les autres sont des vues "soleil d'été".

II.8.3.2.3 Choix des directions à partir des directions de référence.

Les directions des fonctions de base sont prises de la façon suivante.

i) Constantes. Tous les éléments ont les mêmes directions de propagation.

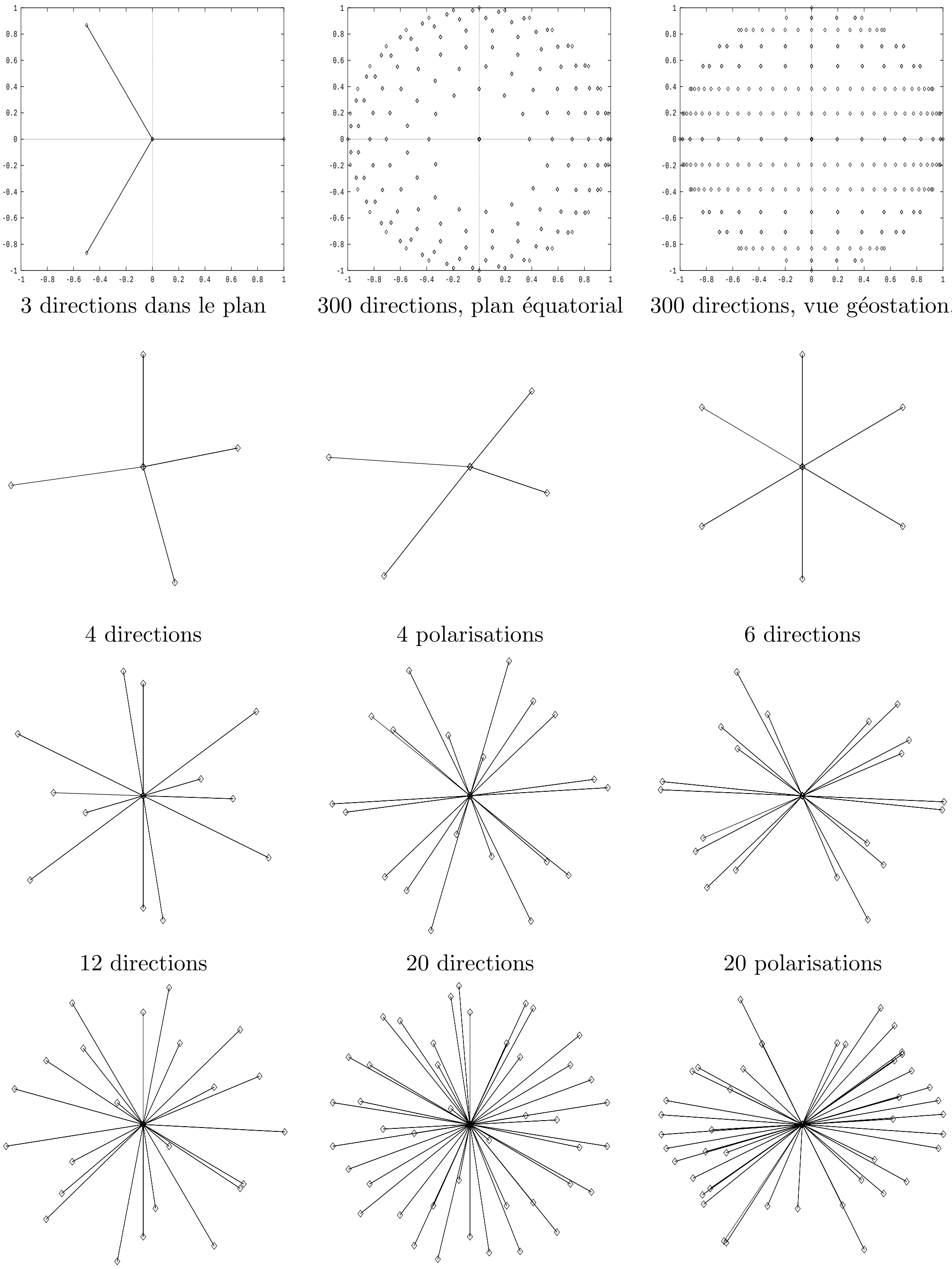
ii) Variantes d'un élément à l'autre. A partir des positions de référence des ondes planes on effectue dans un repère terrestre une rotation en latitude d'angle ϕ entre $-\pi/2$ et $+\pi/2$ puis on effectue une rotation en longitude d'angle θ entre $-\pi$ et $+\pi$. La première étape consiste à multiplier les coordonnées du premier vecteur de la base canonique par la matrice

$$\begin{bmatrix} \cos(\phi) & 0 & \sin(\phi) \\ 0 & 1 & 0 \\ -\sin(\phi) & 0 & \cos(\phi) \end{bmatrix},$$

la deuxième étape consiste à multiplier par la matrice

$$\begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Les lois de répartition des variables aléatoires ϕ et θ peuvent être des deux formes suivantes :



3 directions dans le plan — 300 directions, plan équatorial — 300 directions, vue géostation.

4 directions — 4 polarisations — 6 directions

12 directions — 20 directions — 20 polarisations

23 directions — 44 directions — 44 polarisations

FIG. II.8.2 – Directions de propagation et polarisations de référence, vues Soleil d'été.

1. Loi uniforme en ϕ : cela introduit une anisotropie polaire sur la loi de répartition de la fonction de base de coordonnées

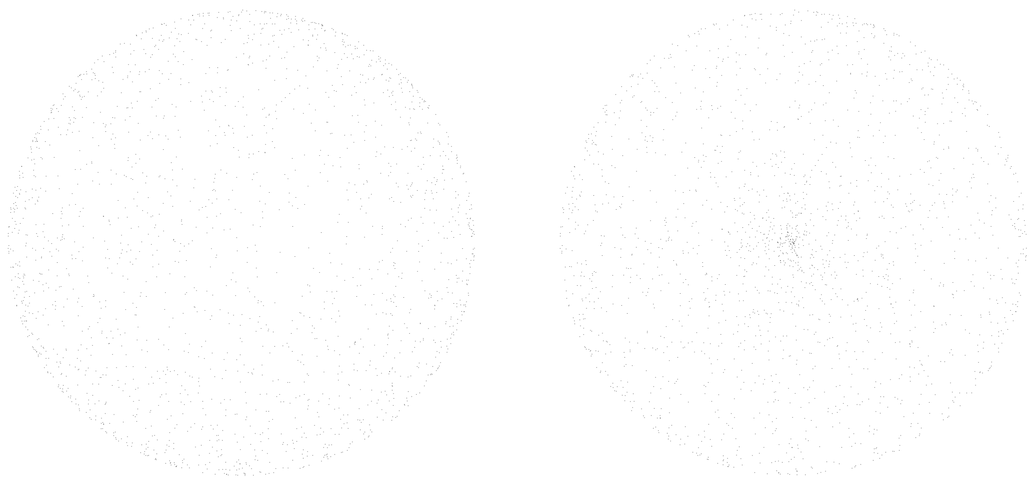
$$\begin{pmatrix} \cos(\theta)\cos(\phi) & -\sin(\theta)\cos(\phi) & -\sin(\phi) \end{pmatrix}$$

   Se reporter à la figure II.8.3 *b*) donnant une idée de cette anisotropie polaire parasite. La loi de répartition, en $1/\cos(\phi)$, est singulière aux pôles.

2. Loi de probabilité sur la variable aléatoire de latitude ϕ en $\cos(\phi)$: Cela supprime l'anisotropie polaire sur la loi de répartition de la fonction de base citée ci-dessus. On constate, sur la figure II.8.3 *a*), la répartition uniforme sur la sphère de cette fonction de base. C'est cette loi que nous avons prise. Le lecteur vérifiera que les changements de variables aléatoires C^1 difféomorphismes,

$$\begin{cases} \alpha \mapsto \phi = -\arcsin(1-2\alpha) \\ [0,1] \to [-\pi/2, \pi/2] \end{cases} \text{ et } \begin{cases} \beta \mapsto \theta = \pi(1-2\beta) \\ [0,1] \to [-\pi, \pi] \end{cases}$$

   appliqués à des paramètres α et β suivant des lois de répartition uniformes sur le segment $[0,1]$, donnent bien une loi uniforme sur θ et une loi de probabilité sur ϕ en $\cos(\phi)$, donc une loi de répartition uniforme sur la sphère.



a) b)

FIG. II.8.3 – Lois de répartition selon le choix des variables aléatoires

II.8.3.2.4 Conclusion : choix pratique des directions et polarisations.

Les directions des fonctions de base sont pour conclure transformées par l'application linéaire dont la matrice othonormale directe est

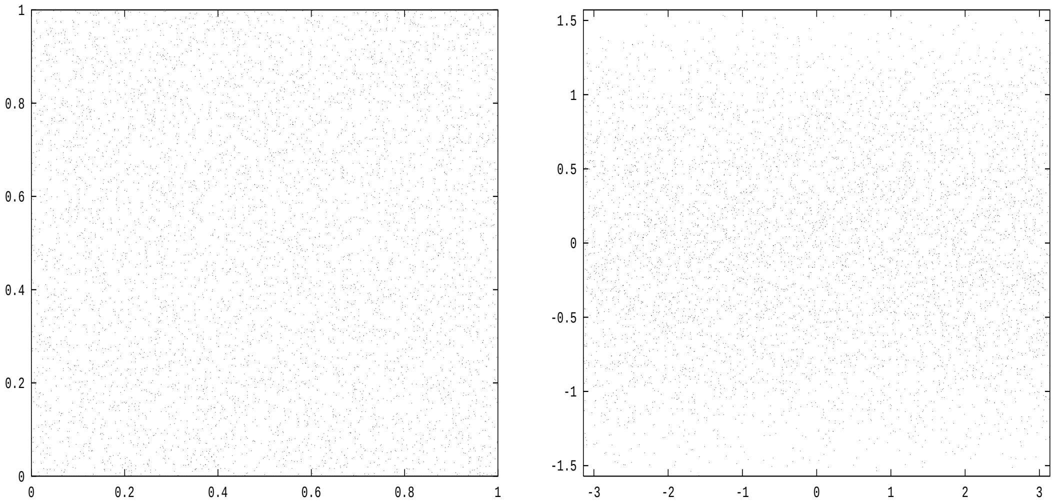
$$\begin{bmatrix} \cos(\theta)\cos(\phi) & \sin(\theta)\cos(\psi)+\cos(\theta)\sin(\phi)\sin(\psi) & -\sin(\theta)\sin(\psi)+\cos(\theta)\sin(\phi)\cos(\psi) \\ -\sin(\theta)\cos(\phi) & \cos(\theta)\cos(\psi)-\sin(\theta)\sin(\phi)\sin(\psi) & -\cos(\theta)\sin(\psi)-\sin(\theta)\sin(\phi)\cos(\psi) \\ -\sin(\phi) & \cos(\phi)\sin(\psi) & \cos(\phi)\cos(\psi) \end{bmatrix}$$

pour des variables aléatoires ϕ, θ et ψ données par les changements de variables

$$\begin{cases} \alpha \mapsto \phi = -\arcsin(1-2\alpha) \\ [0,1] \to [-\pi/2, \pi/2] \end{cases} \quad \begin{cases} \beta \mapsto \theta = \pi(1-2\beta) \\ [0,1] \to [-\pi, \pi] \end{cases} \begin{cases} \gamma \mapsto \psi = \pi(1-2\gamma) \\ [0,1] \to [-\pi, \pi] \end{cases}$$

où α, β et γ sont des variables aléatoires indépendantes de répartition uniforme sur $[0, 1]$.

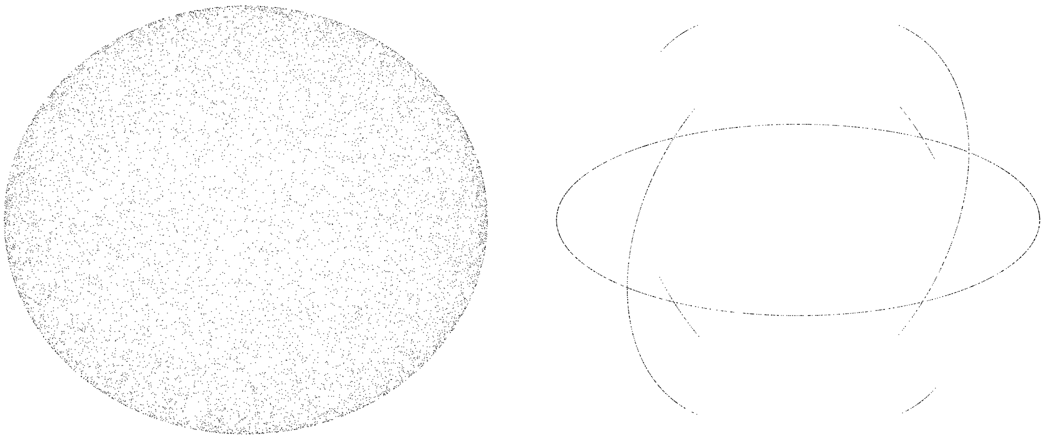
Cela nous donne une équirépartition des directions des fonctions de base pour un grand nombre d'éléments en gardant les angles entre les directions de référence. Nous appliquons ce procédé au cas de quatre fonctions de base par élément (figures II.8.5) dont la position de référence des fonctions de base est le tétraèdre régulier.



Lois en $\beta(\alpha)$ Lois en $\phi(\theta)$

FIG. II.8.4 – Lois de répartition des paramètres.

Nous effectuons 2000 tirs des variables aléatoires α et β sur $[0, 1]$ qui donnent les variables ϕ (entre $-\pi/2$ et $+\pi/2$) et θ (entre $-\pi$ et $+\pi$) indépendamment l'un de l'autre comme le montre la loi de répartition de la figure II.8.4. Remarquons que la densité de points semble bien être proportionnelle à $\cos(\phi)$.



Quatre directions (tétraèdre régulier), Polarisations correspondantes.

FIG. II.8.5 – Loi de répartition pour quatre fonctions de base (Vue Soleil d'été)

Chapitre II.9

Analyse de la méthode sur le problème de Maxwell tridimensionnel.

Dans l'étude du problème de Helmholtz nous avons d'abord montré des simulations numériques qui présentaient des lois d'ordre de convergence. Nous avons ensuite vérifié certaines de ces lois à l'aide d'une majoration fondamentale, la majoration de l'erreur d'interpolation $(I - P_h)X$. Nous avons réussi à majorer des normes d'erreur sur le bord Γ du domaine par l'erreur d'interpolation dans l'espace des fonctions d'énergie finie pour le cas où il n'y a pas de source d'énergie volumique dans le domaine Ω. Par un raisonnement par dualité nous avons aussi effectué une majoration d'erreur sur le bord dans un espace H^{-s} avec $s > 1/2$. Nous avons constaté que nos majorations n'étaient pas optimales, et en outre les simulations numériques montraient qu'il était certainement possible de faire des estimations volumiques d'énergie.

Le but de ce chapitre n'est pas de recommencer exactement la même analyse. En effet, d'après l'étude du problème de Helmholtz et d'après les similitudes avec le problème de Maxwell, il nous semble clair que des lois semblables existent pour le problème de Maxwell. Nous n'effectuons donc pas de simulations numériques pour vérifier ces lois. Tous les points théoriques étudiés dans la première partie peuvent être reconduits, avec quelques variantes, pour le problème de Maxwell.

Par exemple, le résultat d'estimation du résidu pour le problème de Helmholtz (section I.3.3.1) est toujours valable, puisqu'aucun des arguments cités dans la preuve n'est spécifique au problème de Helmholtz, mais général à tout problème qui s'écrit sous la forme variationnelle ultra-faible.

Lemme 20 *Soit $\mathcal{X} \in V$ la solution de (II.7.100) et $\mathcal{X}_h \in V_h$ la solution de (II.8.1). Soit P_h le projecteur orthogonal sur V_h. Nous avons :*

(II.9.1) $$(I - A)(\mathcal{X} - \mathcal{X}_h) \in V_h^{\perp}$$

(II.9.2) $$\boxed{||(I - A)(\mathcal{X} - \mathcal{X}_h)|| \leq 2||(I - P_h)\mathcal{X}||}\ .$$

En revanche, l'estimation de l'erreur d'interpolation est beaucoup plus compliquée. C'est pourquoi, l'analyse de la méthode est menée selon le plan suivant qui met en avant

- la difficulté essentielle nouvelle : l'étude de l'erreur d'interpolation. L'analyse diffère radicalement de l'analyse effectuée dans la première partie. Cette étude est menée dans la première section (II.9.1) de ce chapitre.
- Comme pour le problème de Helmholtz, nous majorons l'erreur sur le bord dans le cas où il n'y a pas de source volumique pour le problème donné. Nous en déduisons les lois d'ordre de convergence qui sont proposées dans la section (II.9.2).
- Enfin, nous étudions, section (II.9.3), le conditionnement du problème avec une analyse non triviale proche de celle effectuée en première partie.

II.9.1 Etude de l'erreur d'interpolation.

Cette étude essentielle dans l'analyse de la méthode ne peut reprendre les points expliqués pour le problème de Helmholtz bidimensionnel dans le vide. Les difficultés majeures proviennent du caractère vectoriel des équations et de la prise en compte des relations de divergence. De plus, les résultats obtenus pour le problème de Helmholtz ne se généralisent pas simplement au problème de Maxwell tridimensionnel, le choix des fonctions de base n'est plus aussi simple. Nous verrons que choisir des directions de propagation toutes distinctes ne suffit plus pour contrôler l'erreur d'interpolation.

Nous considérons le problème de Maxwell dans le vide sans terme source, soit le problème sur l'inconnue $(\mathbf{E}, \mathbf{H})$ vérifiant les équations de Maxwell-Faraday et Maxwell-Gauss ci dessous :

$$\left\{\begin{array}{l} \nabla\wedge\mathbf{E} - i\omega\mathbf{H} = 0 \\ \nabla\wedge\mathbf{H} + i\omega\mathbf{E} = 0 \ . \end{array}\right. \tag{II.9.3}$$

Ces relations impliquent dans le vide

$$\left\{\begin{array}{l} \nabla.\,\mathbf{E} = 0 \\ \nabla.\,\mathbf{H} = 0 \ . \end{array}\right. \tag{II.9.4}$$

En posant,

$$\left\{\begin{array}{l} \mathbf{F} = \dfrac{1}{2}\,(\mathbf{E} + i\mathbf{H}) \\ \mathbf{G} = \dfrac{1}{2i}\,(\mathbf{E} - i\mathbf{H}) \ , \end{array}\right. \tag{II.9.5}$$

nous obtenons le problème équivalent sur $(\mathbf{F}, \mathbf{G})$ devant vérifier les équations de Maxwell (ou conditions de rotationnel)

$$\left\{\begin{array}{l} \nabla\wedge\mathbf{F} - \omega\mathbf{F} = 0 \\ \nabla\wedge\mathbf{G} + \omega\mathbf{G} = 0 \end{array}\right. \tag{II.9.6}$$

et les relations de Gauss (ou conditions de divergence)

$$\left\{\begin{array}{l} \nabla.\,\mathbf{F} = 0 \\ \nabla.\,\mathbf{G} = 0 \ . \end{array}\right. \tag{II.9.7}$$

Nous supposons que la fonction $\mathbf{F}$ solution admet un développement de Taylor uniforme à l'ordre N sur Ω de telle sorte qu'il existe

$$3 \times \frac{(N+3)(N+2)(N+1)}{6} \tag{II.9.8}$$

coefficients complexes $\mathbf{F}_i^{q_1,q_2,q_3}$ définissant la fonction polynômiale vectorielle F_a par

$$\mathbf{F}_a = (\sum_{n=0}^{N} \sum_{q_1+q_2+q_3=n}^{q_i\in\mathbb{N}} \mathbf{F}_i^{q_1,q_2,q_3} \left(\prod_{j=1}^{3} x_j^{q_j}\right))_{i=1,2,3}$$

et tels que pour tout élément Ω_k de Ω il existe une constante C_0 strictement positive donnant la majoration

$$\left\{\begin{array}{l} \forall(\mathbf{X}, \mathbf{X}^{'}) \in (\Omega_k)^2 \\ |\mathbf{F}(\mathbf{X}) - \mathbf{F}_a(\mathbf{X}^{'})|_{C^{N+1}(\Omega)} \leq C_0(\Omega_k, \omega) h^{N+1} |\mathbf{F}|_{C^{N+1}(\Omega)} \ . \end{array}\right. \tag{II.9.9}$$

Dans l'hypothèse (II.9.9), h est toujours le paramètre de taille du maillage que l'on suppose régulier, vérifiant les hypothèses H1, H2 et H3 p. 24. Ceci assure que dans Ω_k, on a $|\mathbf{X} - \mathbf{X}^{'}| \leq h$. On suppose que les dérivées de $\mathbf{F}$ admettent aussi un développement de Taylor uniforme obtenu par dérivation de celui de la fonction $\mathbf{F}$, soit par dérivation de la fonction polynômiale $\mathbf{F}_a$.

Nous allons chercher la condition d'existence de p fonctions $\mathbf{F}_l$ vérifiant les équations de Maxwell et les relations de Gauss

$$\left\{\begin{array}{l} \nabla\wedge\mathbf{F}_l = \omega\mathbf{F}_l \\ \nabla.\,\mathbf{F}_l = 0 \end{array}\right. \tag{II.9.10}$$

et admettant un développement de Taylor à l'ordre N sur un élément Ω_k de Ω, c'est-à-dire telles que l'on puisse définir les fonctions polynômiales vectorielles $\mathbf{F}_l^a$ par les coefficients complexes $\mathbf{F}_{l,i}^{q_1,q_2,q_3}$

(II.9.11) $$\mathbf{F}_l^a = (\sum_{n=0}^{N} \sum_{q_1+q_2+q_3=n}^{q_i \in \mathbb{N}} \mathbf{F}_{l,i}^{q_1,q_2,q_3} \left(\prod_{j=1}^{3} x_j^{q_j} \right))_{i=1,2,3}$$

tels que

(II.9.12) $$|\mathbf{F}_l - \mathbf{F}_l^a|_{C^{N+1}(\Omega_k)} \leq C_l(\Omega_k, \omega) h^{N+1} |\mathbf{F}_l|_{C^{N+1}(\Omega_k)}$$

et qu'il existe p coefficients complexes bornés α_l assurant

(II.9.13) $$\mathbf{F}_a - \sum_{l=1}^{p} \alpha_l \mathbf{F}_l^a = 0 \ .$$

Si les fonctions $\mathbf{F}_l$ existent, alors on pourra majorer

$$|\mathbf{F} - \sum_{l=1}^{p} \alpha_l \mathbf{F}_l|_{C^{N+1}(\Omega)} \leq C h^{N+1}$$

avec

$$C = \left(C_0(\Omega_k, \omega) |\mathbf{F}|_{C^{N+1}(\Omega)} + \sum_{l=1}^{p} C_l(\Omega_k, \omega) |\alpha_l| |\mathbf{F}_l|_{C^{N+1}(\Omega_k)} \right) \ .$$

En conjuguant toutes les relations écrites jusqu'à présent, on aura les mêmes relations avec les coefficients $\overline{\alpha_l}$ sur les fonctions $\mathbf{G}_l$ par rapport à la solution $\mathbf{G}$. Alors, en définissant $\mathbf{E}_a$ par

$$\begin{array}{c} \mathbf{E}_a = \sum_{l=1}^{p} \mathbf{E}_l \\ \mathbf{E}_l = \mathbf{F}_l + i\mathbf{G}_l \end{array}$$

et $\mathbf{H}_a$ par

$$\begin{array}{c} \mathbf{H}_a = \sum_{l=1}^{p} \mathbf{H}_l \\ i\mathbf{H}_l = \mathbf{F}_l - i\mathbf{G}_l \end{array}$$

on obtient que

(II.9.14) $$||(\mathbf{E}, \mathbf{H}) - (\mathbf{E}_a, \mathbf{H}_a)||^2_{C^{N+1}(\Omega_k)} \leq 4C(h^{N+1})^2.$$

Alors, en posant

$$\mathcal{X}_a = \mathbf{E}_a \wedge \nu + (\mathbf{H}_a \wedge \nu) \wedge \nu$$

on aura, en intégrant sur $\partial\Omega_k$ dont la mesure est de l'ordre de h^2,

(II.9.15) $$||\mathcal{X} - \mathcal{X}_a||^2_{L^2(\partial\Omega_k)} \leq 8Ch^{2N+2}h^2$$

puis en sommant sur les

$$K = O\left(\frac{1}{h^3}\right)$$

éléments, on a,

$$\exists C \geq 0 / \ ||\mathcal{X} - \mathcal{X}_a||^2_V \leq Ch^{2N+2}h^{-1},$$

soit finalement

(II.9.16) $$\begin{cases} \exists C \geq 0 \\ ||\mathcal{X} - \mathcal{X}_a||_V \leq Ch^{N+1/2} \ . \end{cases}$$

On a donc

(II.9.17) $$||(I - P_h)\mathcal{X}||_V \leq Ch^{N+1/2} \ .$$

La difficulté de notre travail réside maintenant dans le fait de montrer qu'il existe p coefficients α_l vérifiant (II.9.13), soit

(II.9.18) $$\mathbf{F}_a - \sum_{l=1}^{p} \alpha_l \mathbf{F}_l^a = 0$$

pour des fonctions $\mathbf{F}_l$ bien choisies. Le système (II.9.18) sur les fonctions polynômiales $\mathbf{F}_l^a$ et $\mathbf{F}_a$ s'écrit de façon équivalente sous la forme d'un système linéaire sur les coefficients de ces fonctions polynômiales, soit

$$\text{(II.9.19)} \qquad \begin{cases} \forall (q_1, q_2, q_3) \in \mathbb{N}^3 \text{ tel que } q_1 + q_2 + q_3 \leq N \\ \mathbf{F}_i^{q_1,q_2,q_3} - \displaystyle\sum_{l=1}^{p} \alpha_l \mathbf{F}_{l,i}^{q_1,q_2,q_3} = 0 \ . \end{cases}$$

Le système linéaire (II.9.19) écrit sous forme matricielle est équivalent au problème (II.9.20) dans l'espace vectoriel $\mathbb{C}^{\frac{(N+3)(N+2)(N+1)}{2}}$.

$$\text{(II.9.20)} \qquad \begin{cases} \text{Trouver } A = (\alpha_l)_{l=1,p} \text{ tel que} \\ MA = [\mathbf{F}] \end{cases}$$

Dans (II.9.20) $[\mathbf{F}] \in \mathbb{C}^{\frac{(N+3)(N+2)(N+1)}{2}}$ est un vecteur dont le terme générique est $[\mathbf{F}]_m = \mathbf{F}_i^{q_1,q_2,q_3}$. La matrice M est constituée de p colonnes d'indice l et $3 \times \frac{(N+3)(N+2)(N+1)}{6}$ lignes d'indice m et le terme générique de M est $M_{m,l} = \mathbf{F}_{l,i}^{q_1,q_2,q_3}$. L'indice de ligne m est une fonction des indices (i, q_1, q_2, q_3). Nous avons toute liberté de choisir une bijection Q liant (i, q_1, q_2, q_3) à m de telle sorte que l'on ait

$$\text{(II.9.21)} \qquad M_{m,l} = \mathbf{F}_{l,i}^{q_1,q_2,q_3}, \ m = Q(i, q_1, q_2, q_3)$$

et

$$\text{(II.9.22)} \qquad [\mathbf{F}]_m = \mathbf{F}_i^{q_1,q_2,q_3}, \ m = Q(i, q_1, q_2, q_3) \ .$$

Le problème (II.9.20) se ramène à l'étude de l'image et du noyau de M. Nous allons suivre la démarche adoptée pour l'étude du problème de Helmholtz.

1. Montrer que le noyau de M est de dimension supérieure à $\dfrac{(N+3)(N+1)(N)}{2}$. Ceci s'effectue en analysant les relations linéaires données par les conditions de divergence et de rotationnel (ou conditions de Gauss et de Maxwell).
   (a) Etudier le noyau de M dans le cas simplifié $N = 1$ (il n'y a pas de relation à l'ordre $N = 0$).
   (b) A l'aide d'une construction explicite technique de la fonction Q qui ordonne les indices (i, q_1, q_2, q_3) dans $\mathbb{N}$, nous montrons la généralisation pour tout N.

   Ce point est conceptuellement aisé à montrer même si la preuve, obligeant à expliciter une loi de construction Q d'un vecteur à partir des indices (i, q_1, q_2, q_3), est très lourde.
2. Montrer qu'il existe un sous espace libre de M de dimension $(N+3)(N+1)$ sous certaines conditions sur le choix des directions de propagation des ondes planes.
   (a) Nous montrons que dans le cas $N = 0$ la matrice M est inversible si et seulement si les $p = 3$ directions de propagation des ondes planes qui construisent l'espace V_h sont distinctes 2 à 2.
   (b) Dans le cas $N = 1$ nous montrons que cette propriété ne se généralise pas.
   (c) La généralisation s'effectue pour tout N sous des hypothèses supplémentaires sur les directions des ondes planes. Plus précisément, nous montrons qu'il existe $p = (N+1)(N+3)$ ondes planes telles que la matrice M soit inversible.

   La démonstration est très longue et très technique.

II.9.1.1 Etude du noyau de la matrice M des coefficients de Taylor des fonctions de base.

II.9.1.1.1 Etude du noyau de M pour $N = 1$.

Commençons à caractériser le noyau de M pour $N = 1$. Pour $N = 1$, la fonction polynômiale vectorielle approchée $\mathbf{F}_a$ est donnée par 12 (ce nombre est donné par la relation (II.9.8) pour $N = 1$) coefficients $\mathbf{F}_i^{q_1,q_2,q_3}$ tels que les trois composantes $(\mathbf{F}_1^a, \mathbf{F}_2^a, \mathbf{F}_3^a)$ de $\mathbf{F}_a$ vérifient

$$\text{(II.9.23)} \qquad \mathbf{F}_i^a = \mathbf{F}_i^{0,0,0} + x_1 \mathbf{F}_i^{1,0,0} + x_2 \mathbf{F}_i^{0,1,0} + x_3 \mathbf{F}_i^{0,0,1} \ .$$

On déduit de $\nabla\wedge \mathbf{F} = \omega\mathbf{F}$ trois relations liant les termes d'ordre zéro aux termes du premier ordre,

$$\text{(II.9.24)} \qquad \begin{cases} \mathbf{F}_3^{0,1,0} - \mathbf{F}_2^{0,0,1} = \omega\mathbf{F}_1^{0,0,0} \\ \mathbf{F}_1^{0,0,1} - \mathbf{F}_3^{1,0,0} = \omega\mathbf{F}_2^{0,0,0} \\ \mathbf{F}_2^{1,0,0} - \mathbf{F}_1^{0,1,0} = \omega\mathbf{F}_3^{0,0,0} \ , \end{cases}$$

et de $\nabla.\mathbf{F} = 0$ une relation sur les termes du premier ordre,

$$\text{(II.9.25)} \qquad \mathbf{F}_1^{1,0,0} + \mathbf{F}_2^{0,1,0} + \mathbf{F}_3^{0,0,1} = 0 \ .$$

Nous proposons d'ordonner les indices (i, q_1, q_2, q_3) de façon à ce que le vecteur $[\mathbf{F}]$ s'écrive sous la forme ci-dessous. Remarquons que cette écriture définit une fonction Q pour le cas $N = 1$ (par (II.9.22)).

On considère alors la matrice P décrite ci-dessous. On a noté $\delta = -1$ pour une raison technique de mise en page. Les quatre premières lignes de la matrice P sont la réécriture matricielle (dans l'ordre) des quatre relations (II.9.24) et (II.9.25). Ceci se traduit par $P[F] = [F_r]$ où les quatre premières lignes de $[F_r]$ sont nulles. La matrice P est triangulaire supérieure inversible.

$$P = \begin{bmatrix} \omega & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \delta & 0 & 1 & 0 \\ 0 & \omega & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \delta & 0 & 0 \\ 0 & 0 & \omega & 0 & \delta & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} [F] = \begin{bmatrix} \mathbf{F}_1^{0,0,0} \\ \mathbf{F}_2^{0,0,0} \\ \mathbf{F}_3^{0,0,0} \\ \mathbf{F}_1^{1,0,0} \\ \mathbf{F}_1^{0,1,0} \\ \mathbf{F}_1^{0,0,1} \\ \mathbf{F}_2^{1,0,0} \\ \mathbf{F}_2^{0,1,0} \\ \mathbf{F}_2^{0,0,1} \\ \mathbf{F}_3^{1,0,0} \\ \mathbf{F}_3^{0,1,0} \\ \mathbf{F}_3^{0,0,1} \end{bmatrix} [F_r] = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \mathbf{F}_1^{0,1,0} \\ \mathbf{F}_1^{0,0,1} \\ \mathbf{F}_2^{1,0,0} \\ \mathbf{F}_2^{0,1,0} \\ \mathbf{F}_2^{0,0,1} \\ \mathbf{F}_3^{1,0,0} \\ \mathbf{F}_3^{0,1,0} \\ \mathbf{F}_3^{0,0,1} \end{bmatrix}$$

La matrice M, dont le terme général est donné par (II.9.21), est maintenant ordonnée puisque Q est construit en même temps que le vecteur $[\mathbf{F}_a]$ (II.9.22) donné ci-dessus. Nous écrivons la matrice M en donnant sa colonne d'indice l, puis la colonne d'indice l de la matrice PM :

$$[M]_l = \begin{bmatrix} \mathbf{F}_{l,1}^{0,0,0} \\ \mathbf{F}_{l,2}^{0,0,0} \\ \mathbf{F}_{l,3}^{0,0,0} \\ \mathbf{F}_{l,1}^{1,0,0} \\ \mathbf{F}_{l,1}^{0,1,0} \\ \mathbf{F}_{l,1}^{0,0,1} \\ \mathbf{F}_{l,2}^{1,0,0} \\ \mathbf{F}_{l,2}^{0,1,0} \\ \mathbf{F}_{l,2}^{0,0,1} \\ \mathbf{F}_{l,3}^{1,0,0} \\ \mathbf{F}_{l,3}^{0,1,0} \\ \mathbf{F}_{l,3}^{0,0,1} \end{bmatrix} \quad [PM]_l = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \mathbf{F}_{l,1}^{0,1,0} \\ \mathbf{F}_{l,1}^{0,0,1} \\ \mathbf{F}_{l,2}^{1,0,0} \\ \mathbf{F}_{l,2}^{0,1,0} \\ \mathbf{F}_{l,2}^{0,0,1} \\ \mathbf{F}_{l,3}^{1,0,0} \\ \mathbf{F}_{l,3}^{0,1,0} \\ \mathbf{F}_{l,3}^{0,0,1} \end{bmatrix}$$

Les quatre premières lignes de la matrice PM sont nulles puisque les fonctions $\mathbf{F}_l$ vérifient les relations (II.9.10) et donc les trois conditions de la forme (II.9.24) pour les conditions de rotationnel (ou équations de Maxwell) et la condition de la forme (II.9.25) pour la condition de divergence (ou relation de Gauss).

Pour conclure, nous avons montré que

$$dim(Ker(M)) = dim(Ker(PM)) \geq 4 \ ,$$

ou, de façon équivalente puisque PM a $12-4=8$ lignes non nulles et p colonnes, que

$$rang(M) = rang(PM) \leq \min(8,p) \ .$$

Remarquons que pour $N=1$, on a $1 = \frac{(N+1)N}{2}$ relation de Gauss, $3 = \frac{(N+2)(N+1)N}{2}$ relations de Maxwell et que le rang de M est inférieur à $8 = (N+1)(N+3)$.

II.9.1.1.2 Relations de Gauss obtenues à partir des relations de Maxwell pour $N \geq 2$.

Dans le cas général, pour tout N, on déduit de $\nabla \wedge \mathbf{F} = \omega \mathbf{F}$ les $3 \times \dfrac{(N+2)(N+1)N}{6}$ relations

$$\text{(II.9.26)} \qquad \begin{cases} \forall n \leq N-1 \\ \forall (q_1,q_2,q_3) \in \mathbb{N}^3 / q_1+q_2+q_3 = n \\ (q_2+1)\mathbf{F}_3^{q_1,q_2+1,q_3} - (q_3+1)\mathbf{F}_2^{q_1,q_2,q_3+1} = \omega \mathbf{F}_1^{q_1,q_2,q_3} \\ (q_3+1)\mathbf{F}_1^{q_1,q_2,q_3+1} - (q_1+1)\mathbf{F}_3^{q_1+1,q_2,q_3} = \omega \mathbf{F}_2^{q_1,q_2,q_3} \\ (q_1+1)\mathbf{F}_2^{q_1+1,q_2,q_3} - (q_2+1)\mathbf{F}_1^{q_1,q_2+1,q_3} = \omega \mathbf{F}_3^{q_1,q_2,q_3} \end{cases}$$

et de $\nabla . \mathbf{F} = 0$ les $\dfrac{(N+2)(N+1)N}{6}$ relations

$$\text{(II.9.27)} \qquad \begin{cases} \forall n \leq N-1 \\ \forall (q_1,q_2,q_3) \in \mathbb{N}^3 / q_1+q_2+q_3 = n \\ (q_1+1)\mathbf{F}_1^{q_1+1,q_2,q_3} + (q_2+1)\mathbf{F}_2^{q_1,q_2+1,q_3} + (q_3+1)\mathbf{F}_3^{q_1,q_2,q_3+1} = 0 \ . \end{cases}$$

Remarquons que pour $N \geq 2$ et $n \leq N-2$, les premières relations du rotationnel s'écrivent aussi

$$\begin{cases} \forall n \leq N-2 \\ \forall (q_1,q_2,q_3) \in \mathbb{N}^3 / q_1+q_2+q_3 = n \\ \omega(q_1+1)\mathbf{F}_1^{q_1+1,q_2,q_3} = (q_1+1)(q_2+1)\mathbf{F}_3^{q_1+1,q_2+1,q_3} - (q_1+1)(q_3+1)\mathbf{F}_2^{q_1+1,q_2,q_3+1} \\ \omega(q_2+1)\mathbf{F}_2^{q_1,q_2+1,q_3} = (q_2+1)(q_3+1)\mathbf{F}_1^{q_1,q_2+1,q_3+1} - (q_2+1)(q_1+1)\mathbf{F}_3^{q_1+1,q_2+1,q_3} \\ \omega(q_3+1)\mathbf{F}_3^{q_1,q_2,q_3+1} = (q_3+1)(q_1+1)\mathbf{F}_2^{q_1+1,q_2,q_3+1} - (q_3+1)(q_2+1)\mathbf{F}_1^{q_1,q_2+1,q_3+1} \end{cases}$$

qui, en faisant la somme des trois équations ci-dessus pour tout $q_1+q_2+q_3 = n \leq N-2$, impliquent naturellement les $\dfrac{(N+1)N(N-1)}{6}$ relations

$$\text{(II.9.28)} \qquad \begin{cases} \forall n \leq N-2 \\ \forall (q_1,q_2,q_3) \in \mathbb{N}^3 / q_1+q_2+q_3 = n \\ (q_1+1)\mathbf{F}_1^{q_1+1,q_2,q_3} + (q_2+1)\mathbf{F}_2^{q_1,q_2+1,q_3} + (q_3+1)\mathbf{F}_3^{q_1,q_2,q_3+1} = 0 \ . \end{cases}$$

II.9.1.1.3 Etude du cas général, construction d'un noyau de la matrice M.

Nous allons montrer l'indépendance des $3 \times \dfrac{(N+2)(N+1)N}{6}$ relations

$$\text{(II.9.29)} \qquad \begin{cases} \forall n \leq N-1 \\ \forall (q_1,q_2,q_3) \in \mathbb{N}^3 / q_1+q_2+q_3 = n \\ (q_2+1)\mathbf{F}_3^{q_1,q_2+1,q_3} - (q_3+1)\mathbf{F}_2^{q_1,q_2,q_3+1} = \omega \mathbf{F}_1^{q_1,q_2,q_3} \\ (q_3+1)\mathbf{F}_1^{q_1,q_2,q_3+1} - (q_1+1)\mathbf{F}_3^{q_1+1,q_2,q_3} = \omega \mathbf{F}_2^{q_1,q_2,q_3} \\ (q_1+1)\mathbf{F}_2^{q_1+1,q_2,q_3} - (q_2+1)\mathbf{F}_1^{q_1,q_2+1,q_3} = \omega \mathbf{F}_3^{q_1,q_2,q_3} \end{cases}$$

et des $\dfrac{(N+1)N}{2}$ relations

$$\text{(II.9.30)} \qquad \begin{cases} \forall (q_1, q_2, q_3) \in \mathbb{N}^3 / q_1 + q_2 + q_3 = N - 1 \\ (q_1+1)\mathbf{F}_1^{q_1+1,q_2,q_3} + (q_2+1)\mathbf{F}_2^{q_1,q_2+1,q_3} + (q_3+1)\mathbf{F}_3^{q_1,q_2,q_3+1} = 0 \ . \end{cases}$$

Remarquons que si l'on ordonne le vecteur $\mathbf{F}_i^{q_1,q_2,q_3}$ par blocs croissants en fonction de $q_1+q_2+q_3$, puis en sous-blocs croissants en fonction de i, puis encore en sous-sous-blocs décroissants en fonction de q_1, nous pourrons faire apparaître les relations (II.9.29) et (II.9.30) sous une forme triangulaire supérieure comme dans le cas $N = 1$ à l'aide d'une matrice P.

Notons qu'une fonction Q vérifiant ce rangement n'est pas unique, puisqu'on a toute liberté d'ordonner les derniers blocs en fonction de q_2 ou en fonction de q_3. Nous avons choisi de ranger ces derniers blocs de façon décroissante en fonction de q_2 (c'est-à-dire de façon croissante avec q_3 puisque le bloc à ranger est donné pour q_1 et $q_1 + q_2 + q_3$ fixé).

La donnée de la fonction Q nous permet d'expliciter entièrement la matrice M, dont le terme $M_{m,l}$ (où m est l'indice de ligne et l de colonne) est donné par (II.9.21) avec $m = Q(i, q_1, q_2, q_3)$. On vérifie que la fonction Q définie ci-dessous dans la Définition 17 correspond bien à un tel rangement.

Définition 17 (Fonction bijective Q)

1. *Le premier terme est* $Q(1, 0, 0, 0) = 1$.
2. *Pour tout* $n \in \mathbb{N}$ *tel que* $1 \leq n \leq N$, *on a la relation de récurrence*

$$Q(1, n, 0, 0) = Q(1, n-1, 0, 0) + 3 * C_{n+1}^2$$

   croissante en fonction de n.
3. *Pour tout* $n \in \mathbb{N}^3$ *tel que* $n \leq N$, *et pour* $2 \leq i \leq 3$, *on a la relation de récurrence*

$$Q(i, n, 0, 0) = Q(i-1, n, 0, 0) + C_{n+1}^2$$

   croissante en fonction de i.
4. *Pour tout* $(q_1, q_2) \in \mathbb{N}^2$ *tel que* $q_1 + q_2 \leq N$, $1 \leq q_1$ *et* $q_2 \leq N-1$, *et pour* $1 \leq i \leq 3$, *on a la relation de récurrence*

$$Q(i, q_1 - 1, q_2 + 1, 0) = Q(i, q_1, q_2, 0) + (q_2 + 1)$$

   décroissante en fonction de q_1.
5. *Pour tout* $(q_1, q_2, q_3) \in \mathbb{N}^3$ *tel que* $q_1 + q_2 + q_3 \leq N$, $1 \leq q_2$ *et* $q_3 \leq N - 1$, *et pour* $1 \leq i \leq 3$, *on a la relation de récurrence*

$$Q(i, q_1, q_2 - 1, q_3 + 1) = Q(i, q_1, q_2, q_3) + 1$$

   décroissante en fonction de q_2, *croissante en fonction de* q_3.

Remarque 41 On déduit de la définition de Q les relations de récurrence supplémentaires ci-dessous.
- Pour tout $(q_1, q_2, q_3) \in \mathbb{N}^3$ tel que $q_1 + q_2 + q_3 \leq N$, $i \geq 1$, et pour $2 \leq i \leq 3$, on a la relation de récurrence

$$Q(i, q_1, q_2, q_3) = Q(i-1, q_1, q_2, q_3) + C_{n+1}^2$$

  croissante en fonction de i.
- Pour tout $(q_1, q_2, q_3) \in \mathbb{N}^3$ tel que $q_1 + q_2 + q_3 \leq N$, $1 \leq q_1$ et $q_2 \leq N-1$, et pour $1 \leq i \leq 3$, on a la relation de récurrence

$$Q(i, q_1 - 1, q_2 + 1, q_3) = Q(i, q_1, q_2, q_3) + (q_2 + q_3 + 1)$$

  décroissante en fonction de q_1.

Remarque 42 On déduit de la définition de Q les relations d'ordre ci-dessous.

– Pour tout $(q_1, q_2, q_3) \in \mathbb{N}^3$, $(s_1, s_2, s_3) \in \mathbb{N}^3$ et tout $1 \le i, j \le 3$, on a la relation d'ordre

$$q_1 + q_2 + q_3 \ge s_1 + s_2 + s_3 \Rightarrow Q(i, q_1, q_2, q_3) \ge Q(j, s_1, s_2, s_3) \ . \tag{II.9.31}$$

– Pour tout $(q_1, q_2, q_3) \in \mathbb{N}^3$ et $(s_1, s_2, s_3) \in \mathbb{N}^3$ tels que $q_1 + q_2 + q_3 = s_1 + s_2 + s_3$, alors on a la relation d'ordre

$$i \ge j \Rightarrow Q(i, q_1, q_2, q_3) \ge Q(j, s_1, s_2, s_3) \ . \tag{II.9.32}$$

– Pour tout $(q_1, q_2, q_3) \in \mathbb{N}^3$ et $(s_1, s_2, s_3) \in \mathbb{N}^3$ tels que $q_1 + q_2 + q_3 = s_1 + s_2 + s_3$, et pour $1 \le i \le 3$, on a la relation d'ordre

$$q_1 \ge s_1 \Rightarrow Q(i, q_1, q_2, q_3) \le Q(i, s_1, s_2, s_3) \ . \tag{II.9.33}$$

On définit alors la matrice P comme dans le cas $N = 1$.

1. La matrice P est non nulle sur la diagonale. La diagonale est ordonnée en quatre blocs croissants. Le bloc $a)$ est constitué de $3 \times \dfrac{(N+2)(N+1)(N)}{6}$ termes qui correspondent aux équations du rotationnel. Le bloc $b)$ est constitué de $\dfrac{(N+1)N}{2}$ termes qui correspondent aux équations de divergence. Les blocs $c)$ et $d)$ sont constitués des $(N+1)(N+3)$ termes manquant pour compléter la matrice P de façon à ce qu'aucun terme de la diagonale ne soit nul. Le bloc $c)$ est constitué de $2\dfrac{(N+1)N}{2}$ termes et le bloc $d)$ de $3(N+1) = 3 \times \left(\dfrac{(N+2)(N+1)}{2} - \dfrac{(N+1)N}{2}\right)$ termes. Nous présentons la diagonale de P sous deux formes,

$$\begin{cases} a) \ q_1 + q_2 + q_3 \le N - 1, \ k = Q(i, q_1, q_2, q_3) \Rightarrow P(k,k) = \omega \\ b) \ q_1 + q_2 + q_3 = N - 1, \ k = Q(1, q_1 + 1, q_2, q_3) \Rightarrow P(k,k) = (q_1 + 1) \\ c) \ q_1 + q_2 + q_3 = N - 1, i \ne 1, \ k = Q(i, q_1 + 1, q_2, q_3) \Rightarrow P(k,k) = 1 \\ d) \ q_1 + q_2 + q_3 = N, \ k = Q(i, 0, q_2, q_3) \Rightarrow P(k,k) = 1 \end{cases}$$

   et, pour $k = Q(i, q_1, q_2, q_3)$, toujours sous la forme croissante (d'après la remarque 42),

$$\begin{cases} a) \ q_1 + q_2 + q_3 \le N - 1 \Rightarrow P(k,k) = \omega \\ b) \ q_1 + q_2 + q_3 = N, i = 1, q_1 \ge 1, \Rightarrow P(k,k) = (q_1) \\ c) \ q_1 + q_2 + q_3 = N, i \ne 1, q_1 \ge 1, \Rightarrow P(k,k) = 1 \\ d) \ q_1 + q_2 + q_3 = N, q_1 = 0, \Rightarrow P(k,k) = 1 \ . \end{cases}$$

   On vérifie que les termes des blocs ci-dessus sont bien ordonnés de façon croissante d'après la remarque 42. En particulier, les termes du bloc $c)$ sont rangés avant les termes du bloc $d)$ d'après la relation (II.9.33).

2. La matrice P est non nulle sur une partie supérieure correspondant aux équations du rotationnel. On note

$$j \equiv i \iff j - 1 = i - 1 \text{ modulo } 3$$

   ce qui nous permet de prendre, comme représentants de la classe d'équivalence définie par "modulo 3", non pas $\{0, 1, 2\}$ mais $\{1, 2, 3\}$. Alors, pour $1 \le i \le 3$ et pour tout $(q_1, q_2, q_3) \in \mathbb{N}^3$ tel que $q_1 + q_2 + q_3 \le N - 1$, on a

$$P(k,m) = -(q_b + 1) \text{ avec } \begin{cases} k = Q(i, q_1, q_2, q_3) \\ m = Q(j, q_1 + \delta_{1,b}, q_2 + \delta_{2,b}, q_3 + \delta_{3,b}) \\ j \equiv i + 1 \text{ et } b \equiv i + 2 \end{cases}$$

   et

$$P(k,m) = +(q_b + 1) \text{ avec } \begin{cases} k = Q(i, q_1, q_2, q_3) \\ m = Q(j, q_1 + \delta_{1,b}, q_2 + \delta_{2,b}, q_3 + \delta_{3,b}) \\ b \equiv i + 1 \text{ et } j \equiv i + 2 \end{cases}$$

   où l'on vérifie que $m \ge k$ puisque les termes sont rangés en fonction croissante de $q_1 + q_2 + q_3$ (cf (II.9.31)).

3. La matrice P est non nulle sur une partie supérieure correspondant aux équations de divergence. Pour tout $1 < j \leq 3$ et pour tout $(q_1, q_2, q_3) \in \mathbb{N}^3$ tel que $q_1 + q_2 + q_3 = N - 1$, on a

$$P(k,m) = (q_j + 1) \text{ avec } \begin{cases} k = Q(1, q_1 + 1, q_2, q_3) \\ m = Q(j, q_1, q_2 + \delta_{2,j}, q_3 + \delta_{3,j}) \end{cases}$$

où l'on vérifie que $m \geq k$. En effet, pour $q_1 + q_2 + q_3$ fixé, on range les termes en fonction de i, et on a ici $j > i$ (cf (II.9.32)).

4. Les autres termes de la matrice P sont tous nuls.

La matrice P est triangulaire supérieure de diagonale non nulle : son déterminant est égal à

$$\left(\omega^{\frac{(N+2)(N+1)(N)}{2}}\right) \times \left(\prod_{q_1=0}^{N-1} (q_1 + 1)^{N-q_1}\right)$$

et est donc non nul si et seulement si $\omega \neq 0$. P est donc inversible.

Le produit $P[\mathbf{F}]$ est le vecteur dont les

$$3 \times \frac{(N+2)(N+1)(N)}{6} + \frac{(N+1)N}{2}$$

premières lignes sont nulles.

Il en est de même de la matrice dont la l-ième colonne est $[\mathbf{F}_l]$ puisque les fonctions de base vérifient les relations sur le rotationnel et la divergence.

On a donc

$$\begin{cases} \forall N \in \mathbb{N} \\ dim(Ker(M)) = dim(Ker(PM)) \geq 3 \times \dfrac{(N+2)(N+1)(N)}{6} + \dfrac{(N+1)N}{2} \end{cases}.$$

De façon équivalente, puisque la matrice PM a

$$3\frac{(N+3)(N+2)(N+1)}{6} - 3\frac{(N+2)(N+1)(N)}{6} - \frac{(N+1)N}{2} = (N+1)(N+3)$$

lignes non nulles et p colonnes, on a le lemme suivant.

Lemme 21 *Le rang de la matrice M, noté $rang(M)$, vérifie*

$$\begin{cases} \forall N \in \mathbb{N} \\ rang(M) \leq \min\left[(N+1)(N+3), p\right] \end{cases}.$$

II.9.1.2 Etude de l'image de la matrice M.

Montrons que l'on peut extraire de M un sous espace libre de dimension $(N+1)(N+3)$ sous certaines conditions sur le choix des fonctions de base.

Pour cela, nous avons besoin d'utiliser le fait que les fonctions $\mathbf{F}_l$ sont des ondes planes, c'est-à-dire des fonctions de la forme

(II.9.34)
$$\mathbf{F}_l = \vec{\mathbf{E}}^l e^{j\omega V^l \mathbf{X}} \ (j^2 = -1)$$

avec

(II.9.35)
$$V^l = \begin{bmatrix} \cos\theta_l \cos\phi_l \\ \cos\theta_l \sin\phi_l \\ \sin\theta_l \end{bmatrix} \ \mathbf{E}^l = \begin{bmatrix} \sin\theta_l \cos\phi_l + i \sin\phi_l \\ \sin\theta_l \sin\phi_l - i\cos\phi_l \\ -\cos\theta_l \ . \end{bmatrix}$$

dont le développement donne, pour $n = q_1 + q_2 + q_3$,

(II.9.36)
$$\mathbf{F}_{l,i}^{q_1,q_2,q_3} = \mathbf{E}_i^l (j\omega)^n \prod_{s=1}^{3} \frac{(V_s^l)^{q_s}}{q_s!}$$

Remarque 43 Dans un repère en coordonnées sphériques $(e_r, e_\theta, e_\phi)(\theta_l, \phi_l)$ orthonormé direct, on constate que

$$\text{(II.9.37)} \qquad \begin{cases} V^l = e_r(\theta_l, \phi_l) \\ \mathbf{E}^l = -(e_\theta - ie_\phi)(\theta_l, \phi_l) \end{cases}$$

et surtout que

$$\text{(II.9.38)} \qquad V^l \wedge \mathbf{E}^l = -i\mathbf{E}^l$$

II.9.1.2.1 Etude de la dimension de M pour $N = 0$.

Nous cherchons à caractériser le rang de la matrice M dans le cas $N = 0$ et à établir la condition sur le choix des fonctions de base pour que le système (II.9.20) admette une solution.

Pour $N = 0$ et pour $p = 3$ fonctions $\mathbf{F}_l$ la matrice M est la matrice carrée des polarisations : $M_{m,l} = \mathbf{F}_{l,m}$, soit d'après (II.9.36) la matrice $M_{m,l} = \mathbf{E}^l_m$. Dans une base orthonormée directe bien choisie (cf annexe (III.E.3)), cette matrice est de la forme

$$\text{(II.9.39)} \qquad M = \begin{bmatrix} \sin(\theta)\cos(\phi) + i\sin(\phi) & i\cos(\zeta) & i \\ \sin(\theta)\sin(\phi) - i\cos(\phi) & 1 & 1 \\ -\cos(\theta) & -i\sin(\zeta) & 0 \end{bmatrix}$$

où les angles (ζ, θ, ϕ) définissent les trois vecteurs polarisations $\mathbf{E}^1, \mathbf{E}^2, \mathbf{E}^3$:

$$\text{(II.9.40)} \qquad \begin{bmatrix} \cos(\theta)\cos(\phi) \\ \cos(\theta)\sin(\phi) \\ \sin(\theta) \end{bmatrix} \begin{bmatrix} \sin(\zeta) \\ 0 \\ \cos(\zeta) \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Le déterminant de (II.9.39) est nul si et seulement si

$$\text{(II.9.41)} \qquad \begin{cases} \sin\zeta \sin\phi\,(\sin\theta - 1) = 0 \\ \sin\zeta\sin\theta\cos\phi - \sin\zeta\cos\phi - \cos\zeta\cos\theta + \cos\theta = 0\ . \end{cases}$$

Nous allons montrer que M est inversible si et seulement les trois directions de propagation V^1, V^2, V^3 (II.9.35) sont distinctes deux à deux. Pour cela, étudions la contraposée, c'est-à-dire les conditions pour que les relations (II.9.41) soient vérifiées simultanément.

1. Si $\sin\zeta = 0$ alors si $\cos\zeta = 1$ les vecteurs 2 et 3 de directions de propagation sont égaux. Si $\cos\zeta = -1$ alors $\cos\theta = 0$ qui implique que les vecteurs 1 et 2 ou 3 sont égaux.
2. Si $\sin\theta = 1$ alors $\cos\theta = 0$ et les vecteurs 1 et 3 de directions de propagation sont égaux.
3. Si $\sin\phi = 0$ alors si $\cos\phi = 1$, on a

   $$\cos\theta - \sin\zeta = \cos(\theta + \zeta)$$

   et si $\cos\phi = -1$,

   $$\cos\theta + \sin\zeta = \cos(\theta - \zeta)$$

   La fonction $f(\theta, \zeta) = \cos(\theta - \zeta) - \cos\theta - \sin\zeta$ est une fonction de signe constant. On pose $\psi = \pi/2 - \zeta$, alors

   $$\begin{aligned} \sin\zeta &= \sin(\pi/2 - \psi) = \cos\psi \\ \cos(\theta - \zeta) &= \cos(-\pi/2 + \theta + \psi) = \sin(\theta + \psi) \\ f(\theta, \zeta) &= \sin(\theta + \psi) - \cos\theta - \cos\psi \end{aligned}$$

   En posant $t = \tan(\theta/2)$ et $q = \tan(\psi/2)$, on a

   $$f(\theta, \zeta) = 2\frac{(q-1)(t-1)(qt-1)}{(1+q^2)(1+t^2)}$$

   Le cas $q = 1$ signifie $\sin\theta = 1$ qui est exclu dans la situation précédente. De même pour $t = 1$. Le cas $qt = 1$ signifie $\cos\frac{(\theta + \pi/2 - \zeta)}{2} = 0$ soit $\cos(\theta - \zeta) = 0$ et puisque $f(\theta, \zeta) = 0$ on a $\cos\theta = -\sin\zeta$ et $\sin\zeta(\sin\theta - \cos\zeta) = 0$. Le cas $\sin\zeta = 0$ est déjà traité, le cas $\sin\theta - \cos\zeta = 0$ impliquerait que les vecteurs 1 et 2 soient égaux.

Le résumé de l'étude de l'image de M pour $N=0$ est donné dans la

Proposition 15 *Dans le cas $N=0$, la matrice M constituée de*

$$3\frac{(N+3)(N+2)(N+1)}{6}=(N+1)(N+3)=3$$

lignes et de

$$p=3$$

colonnes est inversible si et seulement si les trois directions de propagations des ondes planes choisies sont distinctes deux à deux.

II.9.1.2.2 Un contre-exemple inattendu dans le cas $N=1$.

Dans le cas $N=1$ et $p=8$ fonctions $\mathbf{F}_l$, la matrice PM vaut, après un ré-ordonnancement des lignes,

$$\begin{bmatrix} \mathbf{E}_3^l V_3^l \\ \mathbf{E}_3^l V_2^l \\ \mathbf{E}_3^l V_1^l \\ \mathbf{E}_2^l V_3^l \\ \mathbf{E}_2^l V_2^l \\ \mathbf{E}_2^l V_1^l \\ \mathbf{E}_1^l V_3^l \\ \mathbf{E}_1^l V_2^l \end{bmatrix} \tag{II.9.42}$$

pour V^l et $\mathbf{E}^l$ donnés par (II.9.35).

La configuration équirépartie sur la sphère unité, soit pour m de 1 à 4, $\theta_{2m+1}=\pi/4$, $\theta_{2m+2}=-\pi/4$ et $\phi_{2m+1}=\phi_{2m+2}=\pi/4+(m-1)*\pi/2$, nous donne un déterminant non nul de valeur 1. En revanche, la configuration équirépartie dans le plan (x,z), soit, $\theta_1=-\pi/2$, $\theta_2=\pi/2$ puis pour m de 2 à 4, $\phi_{2m+1}=0$, $\phi_{2m+2}=\pi$ et $\theta_{2m+1}=\theta_{2m+2}=(m-3)*\pi/4$, nous donne un déterminant nul (la matrice a notamment trois lignes nulles, celles qui correspondent à V_2 puisque $\forall l, \sin\phi_l=0$). Ce résultat est valable pour tout développement à un ordre supérieur à 1.

II.9.1.2.3 Conclusion, pas de généralisation de la condition obtenue pour Helmholtz.

La condition pour l'existence d'ondes planes approchant le développement limité à l'ordre $N\geq 1$ de la solution n'est donc plus seulement que les ondes planes soient de directions de propagation distinctes comme pour le problème de Helmholtz bidimensionnel.

II.9.1.2.4 Etude de l'existence d'un sous-espace libre de M de taille $(N+1)(N+3)$.

Dans le cas général, la matrice PM étudiée est

$$\begin{cases} q_1+q_2+q_3=N, i\neq 1, q_1\geq 1, \Rightarrow \mathbf{F}_{l,i\neq 1}^{q_1+1,q_2,q_3}=\mathbf{E}_i^l(j\omega)^N\dfrac{(V_1^l)^{q_1}}{(q_1)!}\dfrac{(V_2^l)^{q_2}}{q_2!}\dfrac{(V_3^l)^{q_3}}{q_3!} \\ q_1+q_2+q_3=N, q_1=0, \Rightarrow \mathbf{F}_{l,i}^{q_1,q_2,q_3}=\mathbf{E}_i^l(j\omega)^N\dfrac{(V_2^l)^{q_2}}{q_2!}\dfrac{(V_3^l)^{q_3}}{q_3!} \end{cases} \tag{II.9.43}$$

de taille

$$2\times\frac{(N+1)N}{2}+3\times\left(\frac{(N+2)(N+1)}{2}-\frac{(N+1)N}{2}\right)=(N+1)N+3\times(N+1)$$

En exprimant $\mathbf{E}_i^l$ et V_i^l en fonctions de θ_l et ϕ_l (d'après les relations II.9.35), on a le système défini

1. pour $q_1+q_2+q_3=N-1$, par

$$\begin{cases} (\sin\theta_l\sin\phi_l-i\cos\phi_l)\,(\cos\theta_l)^{q_1+1+q_2}(\cos\phi_l)^{q_1+1}(\sin\phi_l)^{q_2}(\sin\theta_l)^{q_3} \\ -(\cos\theta_l)\,(\cos\theta_l)^{q_1+1+q_2}(\cos\phi_l)^{q_1+1}(\sin\phi_l)^{q_2}(\sin\theta_l)^{q_3} \end{cases} \tag{II.9.44}$$

2. et pour $q_2 + q_3 = N$, par

$$\text{(II.9.45)} \qquad \begin{cases} (\sin\theta_l \cos\phi_l + i \sin\phi_l)\,(\cos\theta_l)^{q_2}(\sin\phi_l)^{q_2}(\sin\theta_l)^{q_3} \\ (\sin\theta_l \sin\phi_l - i\cos\phi_l)\,(\cos\theta_l)^{q_2}(\sin\phi_l)^{q_2}(\sin\theta_l)^{q_3} \\ -\,(\cos\theta_l)\,(\cos\theta_l)^{q_2}(\sin\phi_l)^{q_2}(\sin\theta_l)^{q_3}\ . \end{cases}$$

Nous allons démontrer l'existence de $p = (N+1)(N+3)$ couples (θ_l, ϕ_l) tels que la matrice PM, dont les lignes sont données par (II.9.44) et (II.9.45), soit inversible.

Pour cela, il suffit de montrer les deux conjectures suivantes.

1. Le déterminant d'une matrice dont le terme générique est $f_i(x_j)$ où les fonctions f_i sont linéairement indépendantes n'est pas nul partout. Si les fonctions f_i sont continues et libres sur tout intervalle, l'ensemble des $(x_1, \ldots, x_n)$ annulant le déterminant est de mesure nulle. Ceci fait l'objet du lemme 22.
2. Les fonctions construisant la matrice sont linéairement indépendantes sur (θ, ϕ) dans tout sous domaine de mesure non nulle de $[0 \ldots \pi] \times [-\pi \ldots \pi]$. Ceci fait l'objet du lemme 23.

Lemme 22 *Soit $(f_i)_{i=1\ldots I}$ une famille de fonctions continues et linéairement indépendantes de Ω un ouvert de mesure non nulle de $\mathbb{R}^k$, $k \in \mathbb{N}, k \neq 0$ dans $\mathbb{C}$. Alors, il existe une famille $(\mathbf{X}_j)_{j=1\ldots I} \in \Omega^I$ telle que la matrice $M = (M_{i,j})_{1\leq i,j\leq I}$ définie par $M_{i,j} = f_i(\mathbf{X}_j)$ soit inversible. L'ensemble $(\mathbf{X}_j)_{j=1\ldots I} \in \Omega^I$ tel que $det(M) \neq 0$ est de mesure non nulle. Si la famille est libre sur tout compact de Ω, alors l'ensemble $(\mathbf{X}_j)_{j=1\ldots I} \in \Omega^I$ tel que $det(M) = 0$ est de mesure nulle.*

Preuve. La preuve est effectuée par récurrence et par l'absurde. Supposons donc que : $\forall(\mathbf{X}_j)_{j=1\ldots I} \in \Omega^I$, le déterminant de M est nul. Le déterminant est une application multilinéaire alternée donc, pour $(\mathbf{X}_j)$ fixé de 2 à I, on a, en développant le déterminant par rapport à la première colonne,

$$\begin{cases} \forall \mathbf{X}_1 \in \Omega \\ \displaystyle\sum_{i=1}^{I} \lambda_i((\mathbf{X}_j)_{j=2\ldots I}) f_i(\mathbf{X}_1) = 0 \end{cases}$$

où $\lambda_i((\mathbf{X}_j)_{j=2\ldots I})$ est le déterminant mineur. Comme les fonctions f_i sont linéairement indépendantes sur Ω, on a nécessairement $\lambda_i((\mathbf{X}_j)_{j=2\ldots I}) = 0$ qui exprime la nullité de tous les déterminants mineurs par rapport à la première colonne de la matrice M. Alors, par récurrence, on a : $\forall i, f_i = 0$ sur Ω, ce qui est impossible. Enfin, remarquons que le déterminant est une application continue donc non nulle sur un voisinage dans Ω^I d'un point de déterminant non nul (résultat classique des applications continues). □

Nous avons montré l'indépendance linéaire des fonctions de (θ_l, ϕ_l) que sont les fonctions (II.9.44) et (II.9.45) dans les cas particuliers $N = 1$ et $N = 2$. Le but était d'en "intuiter" un raisonnement général. Ces preuves sont extrêmement calculatoires et consistent à développer les expressions en $\sin^2$ en fonction de $1 - \cos^2$ puis par des multiplications par des matrices inversibles à faire apparaître que les fonctions utilisées sont des combinaisons linéaires de fonctions libres utilisées une seule fois dans tous les termes. Ces preuves sont longues à faire : pour $N = 1$ on obtient 8 fonctions, pour $N = 2$ on obtient 15 fonctions. Nous ne donnons pas ces preuves : trop explicites, elles n'ont pas fait se dégager un raisonnement valable pour tout N.

Après une longue recherche, essayant des raisonnements par récurrence ou par l'absurde ou recherchant les points communs avec Helmholtz bi ou tridimensionnel, nous avons réussi à montrer le lemme 23 par les trois étapes que sont les sous-lemmes 24, 25 et 26.

Lemme 23 *A l'aide des fonctions*

$$\text{(II.9.46)} \qquad \begin{aligned} u &= \sin\theta\cos\phi + i\sin\phi \\ v &= \sin\theta\sin\phi - i\cos\phi \\ w &= -\cos\theta \end{aligned}$$

on définit

1. les $2 \times \dfrac{(N+1)(N+2)}{2}$ *fonctions données par*

$$\text{(II.9.47)} \qquad \begin{aligned} g_1(q_1,q_2,q_3) &= v(\cos\theta)^{q_1+q_2}(\cos\phi)^{q_1}(\sin\phi)^{q_2}(\sin\theta)^{q_3} \\ g_2(q_1,q_2,q_3) &= w(\cos\theta)^{q_1+q_2}(\cos\phi)^{q_1}(\sin\phi)^{q_2}(\sin\theta)^{q_3} \end{aligned}$$

pour $q_1+q_2+q_3 = N$,

2. puis les $(N+1)$ *fonctions données par*

$$\text{(II.9.48)} \qquad g_3(q_1=0,q_2,q_3) = u(\cos\theta)^{q_1+q_2}(\cos\phi)^{q_1}(\sin\phi)^{q_2}(\sin\theta)^{q_3}$$

pour $q_2+q_3 = N$.

Alors, les $(N+1)(N+3)$ *fonctions de* $(\theta,\phi) \in [0\ldots\pi]\times[-\pi\ldots\pi]$ *définies par (II.9.47) et (II.9.48) sont linéairement indépendantes (sur tout ouvert de mesure non nulle de* $[0\ldots\pi]\times[-\pi\ldots\pi]$*).*

Remarque 44 Remarquons qu'il est essentiel de ne pas rajouter la famille $g_3(q_1,q_2,q_3)$ pour $q_1 \neq 0$. En effet, on déduit de

$$u\cos\theta\cos\phi + v\cos\theta\sin\phi + w\sin\theta = 0$$

que

$$\text{(II.9.49)} \qquad \left| \begin{array}{l} \forall (q_1,q_2,q_3)\in\mathbb{N}^3;\ q_1+q_2+q_3 = N-1 \\ g_2(q_1,q_2,q_3+1) + g_1(q_1,q_2+1,q_3) + g_3(q_1+1,q_2,q_3) = 0\ . \end{array} \right.$$

Lemme 24 *La famille décrite par (II.9.47) et (II.9.48) est une sous famille de la famille décrite par (II.9.47) pour* $N+1$*. L'espace vectoriel engendré par la famille (II.9.47) est le même que celui engendré par la famille*

$$\text{(II.9.50)} \qquad \begin{cases} \forall q_1+q_2+q_3 = N \\ f_1(q_1,q_2,q_3) = u(\cos\theta)^{q_1+q_2}(\cos\phi)^{q_1}(\sin\phi)^{q_2}(\sin\theta)^{q_3} \\ f_2(q_1,q_2,q_3) = v(\cos\theta)^{q_1+q_2}(\cos\phi)^{q_1}(\sin\phi)^{q_2}(\sin\theta)^{q_3} \end{cases}$$

avec

$$\begin{aligned} u &= \sin\theta\cos\phi + i\sin\phi \\ v &= \sin\theta\sin\phi - i\cos\phi\ . \end{aligned}$$

Preuve. Supposons l'existence de $(N+1)(N+3)$ coefficients $\alpha_{1,2}^{(q_2,q_3)}, \alpha_3^{q_3}$ pour $q_2+q_3 \le N$ tels que

$$\text{(II.9.51)} \qquad \begin{aligned} 0 = &\sum_{q_2,q_3} \alpha_1^{q_2,q_3} g_1(q_1,q_2,q_3) \\ &+ \sum_{q_2,q_3} \alpha_2^{q_2,q_3} g_2(q_1,q_2,q_3) \\ &+ \sum_{q_3} \alpha_3^{q_3} g_3(q_1=0,q_2,q_3) \end{aligned}$$

On peut multiplier l'égalité (II.9.51) par $\cos\theta\cos\phi$ et on utilise la relation (II.9.49). On a alors pour $q_1+q_2+q_3 = N$,

$$\begin{aligned} 0 = &\sum_{q_2,q_3} \alpha_1^{q_2,q_3} g_1(q_1+1,q_2,q_3) \\ &+ \sum_{q_2,q_3} \alpha_2^{q_2,q_3} g_2(q_1+1,q_2,q_3) \\ &- \sum_{q_3} \alpha_3^{q_3} \left(g_1(0,q_2,q_3+1) + g_2(0,q_2+1,q_3)\right) \end{aligned}$$

soit

$$\begin{aligned} 0 = &\sum_{q_2,q_3} \alpha_1^{q_2,q_3} g_1(q_1+1,q_2,q_3) - \sum_{q_3} \alpha_3^{q_3} g_1(0,q_2,q_3+1) \\ &+ \sum_{q_2,q_3} \alpha_2^{q_2,q_3} g_2(q_1+1,q_2,q_3) - \sum_{q_3} \alpha_3^{q_3} g_2(0,q_2+1,q_3) \end{aligned}$$

qui est une combinaison linéaire extraite de la famille (II.9.47) pour $N+1$ (ce qui signifie pour $q_1+q_2+q_3 = N+1$) de la famille

$$g_1(q_1,q_2,q_3) = v(\cos\theta)^{q_1+q_2}(\cos\phi)^{q_1}(\sin\phi)^{q_2}(\sin\theta)^{q_3}$$
(II.9.52)
$$g_2(q_1,q_2,q_3) = w(\cos\theta)^{q_1+q_2}(\cos\phi)^{q_1}(\sin\phi)^{q_2}(\sin\theta)^{q_3}\ .$$

Les permutations entre u, v et w sont permises par la relation (II.9.49). Ceci montre la relation (II.9.50) du lemme. □

Lemme 25 *La famille f_1 décrite par (II.9.50) est libre. Il en est de même de la famille f_2 (et de la famille g_2).*

Preuve. Supposons qu'il existe $\dfrac{(N+1)(N+2)}{2}$ coefficients γ^{q_2,q_3} tels que

(II.9.53)
$$\sum_{0\le q_2+q_3\le N} \gamma^{q_2,q_3} f_1(q_1,q_2,q_3) = 0\ .$$

En posant,

(II.9.54)
$$p(q_1,q_2,q_3) = (\cos\theta)^{q_1+q_2}(\cos\phi)^{q_1}(\sin\phi)^{q_2}(\sin\theta)^{q_3}\ ,$$

la relation (II.9.53) est équivalente, après division par $u \neq 0$, à

(II.9.55)
$$\sum_{0\le q_2+q_3\le N} \gamma^{q_2,q_3} p(q_1,q_2,q_3) = 0\ .$$

On effectue le changement de variable

(II.9.56)
$$\begin{cases} x = \tan(\dfrac{\theta}{2}) \\ y = \tan(\dfrac{\phi}{2})\ , \end{cases}$$

et on exprime $p(q_1,q_2,q_3)$ en fonction de (x,y) :

$$\begin{aligned} p(q_1,q_2,q_3) =& (\frac{1-x^2}{1+x^2})^{q_1+q_2}(\frac{1-y^2}{1+y^2})^{q_1}(\frac{2y}{1+y^2})^{q_2}(\frac{2x}{1+x^2})^{q_3} \\ =& \frac{(2x)^{q_3}(1-x^2)^{q_1+q_2}(2y)^{q_2}(1-y^2)^{q_1}}{(1+x^2)^{q_1+q_2+q_3}(1+y^2)^{q_1+q_2}} \\ =& \frac{P(q_1,q_2,q_3)}{(1+x^2)^{q_1+q_2+q_3}(1+y^2)^{q_1+q_2+q_3}} \end{aligned}$$

avec

(II.9.57)
$$P(q_1,q_2,q_3) = (2x)^{q_3}(1-x^2)^{q_1+q_2}(2y)^{q_2}(1-y^2)^{q_1}(1+y^2)^{q_3}\ .$$

En remarquant que $q_1+q_2+q_3 = N$, la relation (II.9.55) est équivalente, après multiplication par $(1+x^2)^N(1+y^2)^N$, à

(II.9.58)
$$\sum_{0\le q_2+q_3\le N} \gamma^{q_2,q_3} P(q_1,q_2,q_3) = 0\ .$$

On remarque que $P(q_1,q_2,q_3)$ est de degré

$$x^{2N-q_3}y^{2N-q_2}$$

donc deux polynômes $P(q_1,q_2,q_3)$ et $P(s_1,s_2,s_3)$ ne sont du même degré que s'ils sont égaux. La famille f_1 est donc libre. Il en est évidemment de même pour les deux autres familles. □

Preuve. **[du lemme 23]** Pour achever de montrer l'indépendance des fonctions (f_1,f_2) définies par (II.9.50) il suffit de montrer le lemme 26 qui montre que l'espace vectoriel $[f_1]\times[f_2]$ défini par

$$[f_1]\times[f_2] = \big\{\{f_1(q_1,q_2,q_3), f_2(q_1,q_2,q_3)\}\ ,\ \forall (q_1,q_2,q_3)\in\mathbb{N}^3\ /\ q_1+q_2+q_3 = N\big\}$$

peut aussi se construire par un algorithme classique de construction d'une base (non orthonormée pour le produit scalaire canonique des fonctions de $[0\dots\pi]\times[-\pi\dots\pi]$) qui consiste à construire l'espace vectoriel $[f_2]$ de dimension $C^2_{N+2} = \frac{(N+2)(N+1)}{2}$,

$$[f_2] = \left\{ f_2(q_1, q_2, q_3),\ \forall (q_1, q_2, q_3) \in \mathbb{N}^3\ /\ q_1 + q_2 + q_3 = N \right\}$$

puis à rajouter une à une les C^2_{N+2} fonctions $f_1(q_1, q_2, q_3)$. Cette construction peut s'effectuer à l'aide de la fonction bijective $R(q_1, q_2, q_3)$ définie à l'aide de la fonction Q (cf définition 17 p. 123) par

$$R(q_1, q_2, q_3) = Q(1, q_1, q_2, q_3) - Q(1, N, 0, 0) + 1 \in [1, C^2_{N+2}]\ .$$

En ordonnant les triplets (q_1, q_2, q_3), la fonction R permet de construire l'espace vectoriel $< f_1 >_{(q_1,q_2,q_3)}$ par

$$< f_1 >_{(q_1,q_2,q_3)} = \{ f_1(s_1, s_2, s_3),\ s_1 + s_2 + s_3 = N,\ R(s_1, s_2, s_3) < R(q_1, q_2, q_3) \}$$

et l'on aura

$$f_1(q_1, q_2, q_3) \notin \left([f_2] \times < f_1 >_{(q_1,q_2,q_3)} \right)\ .$$

□

Lemme 26 *Aucune fonction $f_1(r_1, r_2, r_3)$ pour $r_1 + r_2 + r_3 = N$ n'est dans l'espace vectoriel $[f_2]$.*

Preuve. Montrons qu'il est impossible de trouver (r_1, r_2, r_3) et des complexes β^{q_2,q_3} tels que

$$f_1(r_1, r_2, r_3) = \sum_{0 \leq q_2, q_3 \leq N} \beta^{q_2,q_3} f_2(q_1, q_2, q_3),$$

ou, en d'autres termes, montrons qu'une telle relation mène à une incohérence. Cette relation se traduit, avec (u, v, w) définis par (II.9.46) et p par (II.9.54), par

$$up(r_1, r_2, r_3) = v \sum_{0 \leq q_2, q_3 \leq N} \beta^{q_2,q_3} p(q_1, q_2, q_3)\ . \tag{II.9.59}$$

La relation (II.9.59) exprimée en fonction de (x, y) (II.9.56) est équivalente à

$$UP(r_1, r_2, r_3) = V \sum_{0 \leq q_2, q_3 \leq N} \beta^{q_2,q_3} P(q_1, q_2, q_3) \tag{II.9.60}$$

où P est défini par (II.9.57) et (U, V, W) par

$$\begin{aligned} U =& (2x)(1 - y^2) + i(1 + x^2)(2y) = 2ix^2y - 2xy^2 + 2x - 2iy \\ V =& (2x)(2y) - i(1 + x^2)(1 - y^2) = ix^2y^2 - ix^2 + iy^2 + 4xy - i \\ W =& -(1 - x^2)(1 + y^2) = x^2y^2 + x^2 - y^2 - 1 \end{aligned} \tag{II.9.61}$$

Le terme de plus haut degré d_1 de $UP(r_1, r_2, r_3)$ est la somme

$$d_1 = (-1)^{-r_2} 2^{r_3+r_2} x^{2N-r_3} y^{2N-r_2} \left(-2xy^2 + 2ix^2y \right)$$

Les termes d_2 de plus haut degré pour f_2 sont de la forme

$$d_2 = (-1)^{-q_2} 2^{q_3+q_2} x^{2N-q_3} y^{2N-q_2} \left(ix^2y^2 \right)$$

L'égalité des termes de plus haut degré impose donc l'existence de deux couples (q_1, q_2, q_3) et (s_1, s_2, s_3) tels que

$$\begin{array}{lll} q_1 = r_1 + 1 & q_2 = r_2 & q_3 = r_3 - 1 \\ s_1 = r_1 + 1 & s_2 = r_2 - 1 & s_3 = r_3 \end{array}$$

ce qui impose une condition sur (r_1, r_2, r_3). On identifie les termes de plus haut degré et l'on a

$$\left(-2xy^2 + 2ix^2y \right) = \frac{1}{2} \left(-\beta^{r_2-1,r_3}(ixy^2) + \beta^{r_2,r_3-1}(ix^2y) \right)$$

soit $\beta^{r_2-1,r_3} = -4i$ et $\beta^{r_2,r_3-1} = 4$. Remarquons en outre que U est impaire et que V est paire. Nous pouvons écrire que si la relation (II.9.60) est vraie alors

$$\begin{aligned} UP(r_1,r_2,r_3) = &-4iVP(r_1+1,r_2-1,r_3)+4VP(r_1+1,r_2,r_3-1) \\ &+\sum_{q_2,q_3}\beta^{r_2+q_2,r_3+q_3}VP(r_1-q_2-q_3,r_2+q_2,r_3+q_3) \end{aligned} \tag{II.9.62}$$

où $(q_2,q_3)\in\mathbb{N}^2$, $q_2+q_3\leq N-r_1$ et q_2+q_3 est impaire. L'analyse des termes suivants ne nous a pas permis d'exprimer simplement une incohérence. Nous allons voir que l'incohérence est montrée à l'aide de la proposition 16 qui peut être vérifiée par un calcul élémentaire. La signification géométrique de la proposition 16 sera montrée après la fin de la preuve de ce lemme.

Proposition 16 *Avec (U,V,W) définis par (II.9.61), et (A,B,C) donnés par*

$$\begin{aligned} A &= (1-x^2)(1-y^2) \\ B &= 2y(1-x^2) \\ C &= (2x)(1+y^2) \end{aligned}$$

on a

$$(A^2+C^2)U = (BA+iC(1+x^2)(1+y^2))V \ . \tag{II.9.63}$$

Multiplions la relation (II.9.62) par (A^2+C^2). Alors, en divisant par $V\neq 0$, on a

$$\begin{aligned} &P(r_1+1,r_2+1,r_3)+i(1+x^2)(1+y^2)P(r_1,r_2,r_3+1) = \\ &-4i(P(r_1+3,r_2-1,r_3)+P(r_1+1,r_2-1,r_3+2)) \\ &+4(P(r_1+3,r_2,r_3-1)+P(r_1+1,r_2,r_3+1))+S \end{aligned}$$

où S est un polynôme de degré inférieur à $x^{2N-r_3+4}y^{2N-r_2+4}$. L'analyse des degrés des termes est effectuée dans le tableau (II.9.64). Dans ce tableau on donne les degrés des polynômes de la colonne de gauche. Le degré en x (resp. y) d'un polynôme est donné par la somme de d_x, colonne du milieu, (resp. d_y, colonne de droite) et de $2r_1+2r_2+r_3$ (resp. $2r_1+r_2+2r_3$).

(II.9.64)

Polynôme	d_x	d_y
$P(r_1+1,r_2+1,r_3)$	4	3
$P(r_1,r_2,r_3+1)$	1	2
$P(r_1+3,r_2-1,r_3)$	4	5
$P(r_1+1,r_2-1,r_3+2)$	2	5
$P(r_1+3,r_2,r_3-1)$	5	4
$P(r_1+1,r_2,r_3+1)$	3	4
S	4	4

On observe que le polynôme de plus haut degré en x est $P(r_1+3,r_2,r_3-1)$. Le coefficient de ce polynôme doit donc être nul, soit $4=0$. □

Preuve. **[de la proposition 16].** Avec

$$\begin{aligned} a &= \cos\theta\cos\phi \\ b &= \cos\theta\sin\phi \\ c &= \sin\theta \end{aligned}$$

on sait, pour des raisons géométriques expliquées dans la remarque 43, que pour (u,v,w) définis par (II.9.46), on a (II.9.38) :

$$(a,b,c)\wedge(u,v,w) = -i(u,v,w) \ ,$$

ce qui se traduit par

$$\begin{aligned} bw-cv &= -iu \\ cu-aw &= -iv \\ av-bu &= -iw \ . \end{aligned}$$

On a alors $w = iav-ibu$ et $u = ib(iav-ibu)-icv$ soit $u=-bav+b^2u-icv$ d'où $(1-b^2)u = -(ba+ic)v$. On déduit de $a^2+b^2+c^2=1$ que

$$(a^2+c^2)u = (ba+ic)v \ .$$

Il suffit alors de multiplier cette relation par $(1+x^2)^3(1+y^2)^3$ pour obtenir (II.9.63). □

II.9.1.3 Bilan de l'étude de l'erreur d'interpolation.

Nous pouvons maintenant rassembler les résultats obtenus dans cette section par le théorème 17.

Théorème 17 (Majoration de l'erreur d'interpolation) *Soit* $(\mathbf{E}, \mathbf{H})$ *une solution du problème de Maxwell (II.9.3) dans le vide sans source, soit* $\varepsilon = \mu = 1$ *et* $(\mathbf{m}, \mathbf{j}) = (0, 0)$ *dans* Ω. *Nous supposons* $(\mathbf{E}, \mathbf{H}) \in (C^{N+1}(\Omega))^2$ *et* $N \geq 0$. *Soit* $\mathcal{X}$ *la solution du problème variationnel avec* $\mathcal{X} = \mathbf{E} \wedge \nu + (\mathbf{H} \wedge \nu) \wedge \nu$. *Nous supposons que le maillage de* Ω *vérifie les hypothèses d'uniforme régularité. Alors, il existe* $p = (N+1)(N+3)$ *directions de propagation définissant les* $2pK$ *fonctions de base de la formulation discrète et une constante positive* C *dépendant de* N, *des fonctions de bases et des données du problème, telles que*

$$||(I - P_h)\mathcal{X}||_V \leq Ch^{N+1/2} \ . \tag{II.9.65}$$

Par exemple, pour $p = 3$ *directions, l'ordre de l'erreur d'interpolation est* $1/2$.

Remarque 45 Pour N paire, le nombre p de directions données par $p = (N + 1)(N + 3)$ correspond au nombre de directions donné par le découpage S_N de la sphère unité (section II.8.3.2.1).

Remarque 46 Dans la démonstration du théorème 17, le fait que $(\mathbf{m}, \mathbf{j}) = (0, 0)$ dans Ω est essentiel pour $N \geq 1$. En revanche, pour $N = 0$ on a vu que le système linéaire (II.9.20) admet une solution : les conditions de divergence (II.9.7) et de rotationnel (II.9.6) ne sont pas utilisées. On aura donc $\forall p \geq 3$, pour $(\mathbf{E}, \mathbf{H}) \in (C^1(\Omega))^2$ solution du problème de Maxwell (II.9.3) dans le vide,

$$||(I - P_h)\mathcal{X}||_V \leq Ch^{1/2} \ . \tag{II.9.66}$$

II.9.2 Etude de l'ordre de convergence de la méthode.

Le but de cette section est d'aboutir à des majorations d'erreur L^2 du bord sur la quantité $\mathcal{X} - \mathcal{X}_h$. Nous utiliserons le résultat de majoration de l'erreur d'interpolation obtenu dans la section (II.9.1) et les extensions au problème de Maxwell des majorations des quantités de bord par rapport à l'erreur d'interpolation. Notons que nous n'effectuons pas l'estimation par dualité (I.3.3.2) pour le problème de Maxwell. Cette extension est néanmoins envisageable et pourrait s'appuyer sur les résultats de [46] étendus aux problèmes de régularité.

II.9.2.1 Une estimation énergétique d'erreur au bord.

Nous étudions ici la norme du résidu $||\mathcal{X} - \mathcal{X}_h||_{L^2(\Gamma)}$ par rapport à l'erreur d'interpolation $||(I - P_h)\mathcal{X}||$. La preuve est copiée sur celle du lemme 9, la différence étant que l'on n'utilise pas que F est une isométrie mais une application de norme inférieure à 1.

Lemme 27 *Supposons que* Q *l'opérateur de bord du problème de Maxwell est constant et* $|Q| \leq \delta < 1$. *Considérons* $\mathcal{X} \in V$ *la solution de (II.7.84) et* $\mathcal{X}_h \in V_h$ *la solution de (II.8.1). Soit* P_h *le projecteur orthogonal sur l'espace* V_h. *Nous avons (II.9.67).*

$$\boxed{||\mathcal{X} - \mathcal{X}_h||_{L^2(\Gamma)} \leq \frac{2}{\sqrt{1 - \delta^2}} ||(I - P_h)\mathcal{X}||_V} \tag{II.9.67}$$

Preuve. Posons $\eta_h = (\mathcal{X} - \mathcal{X}_h)$. En utilisant la définition de A, l'inégalité de Cauchy-Schwarz et le fait que F est une contraction, nous avons

$$\begin{aligned} ((I - A)\eta_h, \eta_h)_V &= ||\eta_h||^2 - (\Pi\eta_h, F\eta_h)_V \\ &\geq ||\eta_h||^2 - ||\Pi\eta_h|| \ ||F\eta_h|| \\ &\geq ||\eta_h||^2 \left(1 - \frac{||\Pi\eta_h||}{||\eta_h||}\right) \ . \end{aligned} \tag{II.9.68}$$

Par définition de Π

$$||\Pi\eta_h||^2 = +\sum_k \int_{\Gamma_k} |Q|^2\ |\mathcal{X}_k - \mathcal{X}_h^k + \sum_{kj} \int_{\Sigma_{jk}} |\mathcal{X}_j - \mathcal{X}_h^j|^2\ ,$$

la somme sur j et k est réordonnée en une somme sur k et j :

$$||\Pi\eta_h||^2 = +\sum_k \int_{\Gamma_k} |Q|^2\ |\mathcal{X}_k - \mathcal{X}_h^k|^2 + \sum_{kj} \int_{\Sigma_{kj}} |\mathcal{X}_k - \mathcal{X}_h^k|^2\ ,$$

soit, en définissant $||\eta_h||_\Gamma^2 = \int_\Gamma |\eta_h|^2$, nous obtenons $||\Pi\eta_h||^2 \leq ||\eta_h||^2 - (1-|\delta|^2)||\eta_h||_\Gamma^2$. Alors,

$$||\Pi\eta_h|| \leq ||\eta_h|| \left(1 - \frac{1-|\delta|^2}{2} \frac{||\eta_h||_\Gamma^2}{||\eta_h||^2}\right)\ . \tag{II.9.69}$$

Des inégalités (II.9.69) et (II.9.68) nous avons

$$|((I-A)\eta_h, \eta_h)_V| \geq ||\eta_h||^2 \left[1 - (1 - \frac{1-|\delta|^2}{2} \frac{||\eta_h||_\Gamma^2}{||\eta_h||^2})\right]\ , \tag{II.9.70}$$

d'où, la majoration (II.9.71) sur $||\eta_h||_\Gamma$

$$|((I-A)\eta_h, \eta_h)_V| \geq \frac{1-|\delta|^2}{2} ||\eta_h||_\Gamma^2\ . \tag{II.9.71}$$

D'après le lemme 20 qui donne $(I - P_h)\mathcal{X} = \mathcal{X} - \mathcal{X}_h$ et l'inégalité de Cauchy-Schwarz, nous avons

$$|((I-A)\eta_h, \eta_h)|_V = |((I-A)\eta_h, (I-P_h)\mathcal{X})|_V \leq 2||(I-P_h)\mathcal{X}||^2\ . \tag{II.9.72}$$

Ainsi, de (II.9.71) et (II.9.72), nous avons :

$$||(I-P_h)\mathcal{X}|| \geq \frac{\sqrt{1-\delta^2}}{2} ||\eta_h||_\Gamma\ . \tag{II.9.73}$$

□

Notons que le lemme 27 qui relie l'erreur sur la quantité $\mathcal{X}$ sur le bord à l'erreur d'interpolation demande que Q soit strictement inférieur à 1. Cette hypothèse est demandée par la démonstration qui effectue un raisonnement uniforme sur tout le bord. Il est possible de raffiner ce raisonnement et d'obtenir un même type de majoration avec des hypothèses moins restrictives sur Q. Par exemple, sous les mêmes hypothèse que celles du lemme 27 mais avec $Q = 1$ sur Γ_1 et $Q = 0$ sur Γ_2 où Γ_1 et Γ_2 sont deux parties non vides de Γ, on aura une estimation d'erreur sur le bord Γ_2.

$$\boxed{||\mathcal{X} - \mathcal{X}_h||_{L^2(\Gamma_2)} \leq ||(I-P_h)\mathcal{X}||_V} \tag{II.9.74}$$

II.9.2.2 Estimation énergétique au bord sur les traces tangentielles.

Lemme 28 *Supposons que Q l'opérateur de bord du problème de Maxwell est constant et $|Q| \leq \delta < 1$. Considérons $\mathcal{X} \in V$ la solution de (II.7.84) et $\mathcal{X}_h \in V_h$ la solution de (II.8.1). Soit $\mathbf{E}$ la solution de (II.8.17), et $\mathbf{E}_h$ définie par (II.8.19) qui définissent d'une part $\mathbf{E} \wedge \nu$, $\mathbf{E}_h \wedge \nu$ et d'autre part $(\mathbf{H} \wedge \nu) \wedge \nu$, $(\mathbf{H}_h \wedge \nu) \wedge \nu$. Nous avons (II.9.75).*

$$\begin{aligned} ||\mathbf{E} \wedge \nu - \mathbf{E}_h \wedge \nu||_{L^2(\Gamma)} &\leq \frac{1+\delta}{2 \min\limits_\Gamma \sqrt{\varepsilon_{kk}}} ||\mathcal{X} - \mathcal{X}_h||_{L^2(\Gamma)} \\ ||(\mathbf{H} \wedge \nu) \wedge \nu - (\mathbf{H}_h \wedge \nu) \wedge \nu||_{L^2(\Gamma)} &\leq \frac{1+\delta}{2 \min\limits_\Gamma \sqrt{\mu_{kk}}} ||\mathcal{X} - \mathcal{X}_h||_{L^2(\Gamma)} \end{aligned} \tag{II.9.75}$$

Preuve. Rappelons les équations (II.8.16) et (II.8.18), section II.8.1.4 :

$$\begin{cases} (\mathbf{H} \wedge \nu) \wedge \nu = \dfrac{1}{2\sqrt{\mu_{kk}}}[(I+\Pi)\mathcal{X} + g] & \text{sur } \Gamma_k \\ (\mathbf{H}_h \wedge \nu) \wedge \nu = \dfrac{1}{2\sqrt{\mu_{kk}}}[(I+\Pi)\mathcal{X}_h + g] & \text{sur } \Gamma_k \ . \end{cases}$$

Ainsi,

$$||(\mathbf{H} \wedge \nu) \wedge \nu - (\mathbf{H}_h \wedge \nu) \wedge \nu||_{L^2(\Gamma)} = \frac{1}{2\sqrt{\mu_{kk}}}||(I+\Pi)(\mathcal{X} - \mathcal{X}_h)||_{L^2(\Gamma)}$$

d'où, à l'aide de la définition de Π et de l'hypothèse $|Q| \leq \delta$, nous avons la deuxième relation de (II.9.75). Il en est de même pour les traces tangentielles du champ électrique $\mathbf{E}$ sur Γ_k d'après les relations (II.8.17) et (II.8.19) :

$$\begin{cases} \mathbf{E} \wedge \nu = \dfrac{1}{2\sqrt{\varepsilon_{kk}}}[(I-\Pi)\mathcal{X} - g] \\ \mathbf{E}_h \wedge \nu = \dfrac{1}{2\sqrt{\varepsilon_{kk}}}[(I-\Pi)\mathcal{X}_h - g] \ . \end{cases}$$

□

II.9.2.3 Estimations de l'ordre de convergence.

D'après le théorème 17 et les lemmes 27 et 28, on a le corollaire 8 de majoration de l'erreur au bord dans le cas $(\mathbf{m}, \mathbf{j}) = (0,0)$ dans Ω.

Corollaire 8 *Soit* $(\mathbf{E}, \mathbf{H})$ *une solution du problème de Maxwell (II.9.3) dans le vide sans source, soit* $\varepsilon = \mu = 1$ *et* $(\mathbf{m}, \mathbf{j}) = (0,0)$ *dans* Ω. *Nous supposons* $(\mathbf{E}, \mathbf{H}) \in (C^{N+1}(\Omega))^2$ *et* $N \geq 0$. *Soit* $\mathcal{X}$ *la solution du problème variationnel avec* $\mathcal{X} = \mathbf{E} \wedge \nu + (\mathbf{H} \wedge \nu) \wedge \nu$. *Nous supposons que le maillage de* Ω *vérifie les hypothèses d'uniforme régularité.*

Alors, il existe $p = (N+1)(N+3)$ *directions de propagation définissant les* $2pK$ *fonctions de base de la formulation discrète et deux constantes positives* C *et* C' *dépendant de* N, *des fonctions de bases et des données du problème, telles que*

(II.9.76)
$$\begin{aligned} &||\mathcal{X} - \mathcal{X}_h||_{L^2(\Gamma)} \leq C h^{N+1/2} \\ &||\mathbf{E} \wedge \nu - \mathbf{E} \wedge \nu_h||_{L^2(\Gamma)} \leq C' h^{N+1/2} \end{aligned}$$

Dans le cas $(\mathbf{m}, \mathbf{j}) \neq (0,0)$ *dans* Ω *on a toujours le résultat (II.9.76) pour* $N = 0$ *et* $p = 3$ *directions de propagation.*

Remarquons que $N = [\sqrt{p+1}] - 2$ où $[\alpha]$ désigne la partie entière de α.

II.9.3 Etude du conditionnement.

Supposons que les $p = (N+1)(N+3)$ directions de propagation des fonctions de base ($p \geq 4$) vérifient les hypothèses du théorème 17. Supposons aussi que l'élément Ω_k et tous ses voisins sont dans le vide. Nous avons vu que le conditionnement de la matrice D_k de couplage hermitien des fonctions de base sur l'élément Ω_k ne dépend pas du choix des polarisations (proposition 14). En outre, nous savons que pour trois directions distinctes le conditionnement de D_k est non nul (annexe III.E.3).

Sous ces hypothèses, nous avons en plus le résultat énoncé dans le théorème suivant.

Théorème 18 *Nous supposons que le milieu présent dans l'élément* Ω_k *et dans tous ses voisins est occupé par le vide. Soient* $p = (N+1)(N+3)+1$ *directions de propagation* ($p \geq 4$) *vérifiant les hypothèses du théorème 17. Soit* h_k *le diamètre de* Ω_k *et* $[\alpha]$ *la partie entière de* α. *Il existe* C *positif tel que le conditionnement* $K(D_k)$ *de* D_k *est minoré par la loi*

(II.9.77)
$$K(D_k) = \frac{\lambda_{\max}}{\lambda_{\min}} \geq C h_k^{-(2[\sqrt{p}]-2)} \ .$$

Par exemple, on peut choisir p directions, $p \in [4, 8]$, telles que le conditionnement de la matrice D_k soit supérieur à h_k^{-1}.

Preuve. La matrice D_k est hermitienne donc $K(D_k) = \frac{\lambda_{\max}}{\lambda_{\min}}$ avec λ_{min} la plus petite valeur propre de D_k et λ_{max} la plus grande. Par définition des valeurs propres, on a

$$\begin{cases} \forall \mathcal{Y} \in \mathbb{C}^{p+1} \\ \lambda_{\min} \leq \dfrac{\mathcal{Y}^\top D_k \mathcal{Y}}{||Y||^2} \leq \lambda_{\max} \ . \end{cases} \tag{II.9.78}$$

Nous allons majorer $\lambda_{\min}$ et minorer $\lambda_{\max}$ ce qui assurera une minoration du conditionnement. On sait que, dans le vide, la matrice D_k est formée de deux blocs carrés diagonaux de taille p. Nous allons étudier les valeurs propres de chacun de ces blocs.

Majoration de $\lambda_{\min}$. Rappelons que les fonctions de base $\mathcal{Z}_{k,l}$ dérivent des ondes planes solutions d'un problème de Maxwell dans le vide et sans terme source.

On peut donc appliquer le théorème 17 pour $p-1$ directions de propagation avec $\mathcal{Y}^\top = [\mathcal{X}_\mathbf{F}, -1]$ où $\mathcal{X}_\mathbf{F}$ est le $\mathbb{C}^{p-1}$ vecteur qui approche la dernière onde plane par une relation de la forme (II.9.18). On a donc, d'après (II.9.15), l'existence d'une constante C telle que

$$\mathcal{Y}^\top D_k \mathcal{Y} = ||\mathcal{X}_{k,p} - \sum_{l=1}^{p-1} \mathcal{X}_\mathbf{F}^l \, \mathcal{X}_{k,l}||^2_{L^2(\partial\Omega_k)} \leq C h_k^{2N+4} \ .$$

Ceci implique, à l'aide de $||\mathcal{Y}||^2 = 1 + \sum_{l=1}^p |\mathcal{X}_\mathbf{F}^l|^2 \geq 1$ et (II.9.78) que :

$$\lambda_{\min} \leq C h_k^{2N+4} \ .$$

Minoration de $\lambda_{\max}$. De même, considérons $\mathcal{Y}$ tel que $\mathcal{Y}^\top = [0, \ldots, 0, -1]$. On émet l'hypothèse, pour simplifier la preuve, que le maillage est constitué de tétraèdres. L'élément Ω_k a donc 4 faces de surfaces S_i. Alors, en appliquant l'équation (II.9.78), nous obtenons

$$\lambda_{\max} \geq \sum_{i=1}^4 S_i \times |\mathbf{F}_{k,p} \wedge \nu_i + i\,(\mathbf{F}_{k,p} \wedge \nu_i) \wedge \nu_i|^2$$

On sait que ce terme est non nul et proportionnel à h_k^2 d'après l'étude de l'annexe III.E.3 qui montre le découplage du déterminant de la matrice limite D_k/h pour trois directions de propagation.

On déduit des deux points précédents que

$$K(D_k) \geq C h_k^{2-(2N+4)} \leq C h_k^{-(2N+2)} \ .$$

En outre, $p = (N+1)(N+3) + 1$ donc $p = (N+2)^2$ soit $N = [\sqrt{p}] - 2$. □

Chapitre II.10

Résultats numériques pour le problème de Maxwell.

Choix d'un nom pour notre programme Maxwell $3-D$. Afin de ne pas systématiquement parler de "notre code", nous avons décidé de lui donner un nom. L'histoire de ce nom remonte à une conversation téléphonique avec mon père le jour où je lui ai dit que je pensais avoir fini la partie programmation de ma thèse et surtout que le code était "débuggé". Il m'a suggéré de lui donner un nom grec. J'ai tout de suite été enthousiasmé par l'idée et ai pensé l'appeler "Ariane" qui est le très joli prénom d'une de mes sœurs et qui évoque la mythologie grecque dont je suis friand. Malheureusement pour moi, ce nom est déjà utilisé par le consortium européen de construction spatiale. De plus, le lancement de la cinquième fusée de la gamme avait été réalisé avec échec quelques jours auparavant. J'ai alors pensé à Sophie, mon autre sœur à prénom antique. La supériorité de Sophie est aussi celle de Scholl et du frère Hans, héroïques opposants au régime nazi pendant la guerre. Mais j'ai pensé que la signification de Sophie avait certainement déjà attiré beaucoup de gens. L'attrait d'un nom grec provient de la base culturelle de notre civilisation, de l'admiration portée à l'antiquité depuis la renaissance par les artistes, de la beauté et du sens de la responsabilité inspirés par l'Illiade, de la puissance de la logique de l'idéologie rationaliste qui tourne les scientifiques vers Socrate en plus de l'évocation naturelle au Pays merveilleux d'Hélène. Il est injuste néanmoins de toujours faire la part belle à cette grande civilisation. Nous oublions toujours les racines celtes dont nos traditions sont issues, nous oublions surtout l'influence majeure de la civilisation et du peuple du Livre. C'est pourquoi j'ai tout naturellement pensé à un nom hébraïque, et pour un code qui traite de la propagation d'ondes et qui est destiné à augmenter les fréquences de résolution des problèmes d'électromagnétisme, il me semble que le nom de $\mathcal{L}ior$ qui signifie "lumière" est bien choisi. C'est de surcroît un prénom féminin ou masculin, et je salue tous les $\mathcal{L}ior$.

But du chapitre.

Le but de ce chapitre est de montrer que le code $\mathcal{L}ior$ satisfait les critères suivants.

1. Obtenir des résultats aussi précis que d'autres méthodes classiques de discrétisation.
2. Travailler avec des objets éventuellement revêtus d'une couche de matériau.
3. Prendre en compte des milieux absorbants (caractéristiques électromagnétiques complexes).
4. Fonctionner avec des maillages très grossiers, par exemple pour des paramètres de discrétisation h de l'ordre de la longueur d'onde λ.
5. Obtenir des résultats comparables (en terme de précision) à une méthode d'éléments finis avec une place mémoire et un temps de calcul du même ordre, mais en utilisant un maillage plus grossier.
6. Travailler indifféremment sur différents types de maillage, en l'occurrence sur des tétraèdres comme sur des hexaèdres, moyennant des coûts d'utilisation différents.

Pour cela, nous avons divisé notre étude en plusieurs sections en observant les points forts et faibles du code $\mathcal{L}ior$ de façon à indiquer son mode optimal d'utilisation.

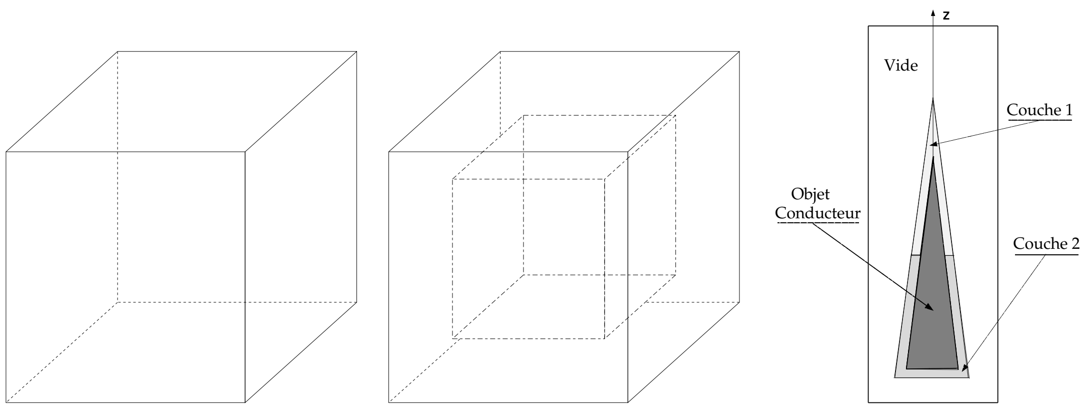
1. Une section de validation du code $\mathcal{L}ior$ dont le but est de nous convaincre des points 1, 2 et 3 cités précédemment et de l'absence d'erreurs de programmation dans le code $\mathcal{L}ior$.

2. Une section dont le but est de montrer les avantages pratiques du code $\mathcal{L}ior$ par les points 4 et 5. Nous comparerons $\mathcal{L}ior$ à d'autres codes.
3. Une petite étude sur des maillages en tétraèdres ou en hexaèdres. Nous savons que l'utilisation d'hexaèdres est plus coûteuse en place mémoire et en temps de calcul, mais nous ne savions pas "a priori" si ces difficultés n'étaient pas compensées par d'autres caractéristiques numériques.

II.10.1 Tests de validation du code.

II.10.1.1 Présentation des tests.

Les simulations effectuées ont essentiellement porté sur trois types de géométries figurées en II.10.1.
- Le domaine représenté par la figure II.10.1 a) est un cube centré sur l'origine de longueur L.
- Le domaine représenté par la figure II.10.1 b) est un cube centré sur l'origine et d'arête de longueur a placé dans un autre cube plus gros d'arête de longueur b. Le domaine de calcul dans le cas b) est donc le volume compris entre les deux cubes. Nous avons utilisé les cas $a = 30$ centimètres et $b = 50$ centimètres puis $a = 50$ centimètres et $b = 70$ centimètres. Notons que dans les deux cas la frontière artificielle est très proche de l'objet.
- Afin de réaliser des simulations qui sont susceptibles d'être proches de problèmes industriels, nous avons considéré le cas du cône dans un cylindre décrit dans la note technique de Christophe Le Potier [42]. Le cône parfaitement conducteur (impénétrable aux ondes électromagnétiques) d'axe de symétrie de révolution z mesure $1,3$ m et le point le plus haut pour z positif est à $0,6$ m. La base du cône est un disque de rayon $0,18$ m. La frontière bornant le domaine est un cylindre de $2,3$ m de hauteur et de rayon $0,4$ m, le point le plus élevé du cylindre pour z positif est à $1,4$ m. Le cône est éventuellement revêtu de deux couches de matériau. Ces couches sont de 5 centimètres d'épaisseur, l'une est à la pointe du cône pour $z > 0$, l'autre est au culot pour $z \leq 0$[1].



a) Propagation libre, b) Diffraction sur un cube, c) Coupe du cône revêtu.

FIG. II.10.1 – Exemples de domaines Ω utilisés.

Nous utiliserons dans cette section les notions de "scattering", calcul en "champ total" ou en "champ diffracté", de conditions de conducteur parfait ou d'absorbant parfait. Ces notions ont déjà été vues pour l'étude du problème de Helmholtz, nous les clarifions pour le problème de Maxwell.

Définition 18 ("scattering" et propagation libre) *Dans toutes les simulations effectuées, nous considérons le problème de "scattering" ou propagation d'une onde plane (que nous appelons onde plane incidente) dans un milieu sans source d'énergie. Le mot "scattering" signifie diffraction lorsque l'onde plane incidente rencontre un objet impénétrable aux ondes électromagnétiques. Lorsque le milieu de propagation a des permittivités et perméabilités variables, le terme "scattering" signifie diffusion : nous n'employons*

[1]La hauteur du cône avec la couche de matériau est donc de $1,715$ m et le point le plus bas sur l'axe z est à -75 centimètres, le rayon de la base du cône revêtu de matériau est de $23,7$ centimètres.

pas ce terme ambigü que l'on risque de confondre avec d'autres phénomènes physiques. En l'absence de diffraction et de diffusion, nous parlerons de propagation libre dans un milieu homogène.

Définition 19 (champ total) *On appelle "calcul en champ total" le problème de Maxwell dans un milieu sans source d'énergie volumique dont la condition aux limites est vérifiée par une onde plane sur la frontière extérieure $\Gamma_{ext} \subset \partial\Omega$ (évidemment non vide). La condition aux limites sur la frontière intérieure Γ_{int} de l'objet diffractant (éventuellement vide) est vérifiée par la solution nulle. Le problème se met sous la forme*

$$\text{(II.10.1)} \qquad \begin{cases} \nabla\wedge \mathbf{E} - i\omega\mu\mathbf{H} = 0 \ \textit{sur}\ \Omega \\ \nabla\wedge \mathbf{H} + i\omega\varepsilon\mathbf{E} = 0 \ \textit{sur}\ \Omega \\ -(1+Q)\sqrt{\varepsilon_{kk}}\mathbf{E}\wedge\nu + (1-Q)\sqrt{\mu_{kk}}\,(\mathbf{H}\wedge\nu)\wedge\nu = g \ \textit{sur}\ \partial\Omega \\ g_{|\Gamma_{ext}} = \left(-(1+Q)\sqrt{\varepsilon_{kk}}\mathbf{E}_0\wedge\nu - (1-Q)\sqrt{\mu_{kk}}\sqrt{\dfrac{\varepsilon}{\mu}}\,(\mathbf{E}_0\wedge\nu)\wedge\nu\right)e^{i\omega\sqrt{\varepsilon\mu}(\mathbf{k}_0\mathbf{X})} \\ g_{|\Gamma_{int}} = 0, \quad Q_{|\Gamma_{ext}} = 0 \end{cases}$$

où l'on a

$$\text{(II.10.2)} \qquad \begin{cases} (\mathbf{k}_0, \mathbf{E}_0) \in (\mathbb{R}^3)^2 \\ \mathbf{k}_0\mathbf{E}_0 = 0 \\ |\mathbf{k}_0| = |\mathbf{E}_0| = 1\ . \end{cases}$$

Notons que lorsque $\Gamma_{int} = \emptyset$ et ε et μ les permittivité et perméabilité sont constantes dans tout le domaine Ω, le couple $(\mathbf{E}, \mathbf{H})$ donné par

$$\text{(II.10.3)} \qquad \begin{cases} \mathbf{E} = \mathbf{E}_0 e^{i\omega\sqrt{\varepsilon\mu}(\mathbf{k}_0\mathbf{X})} \\ \mathbf{H} = -\sqrt{\dfrac{\varepsilon}{\mu}}\mathbf{E}_0\wedge\mathbf{k}_0 e^{i\omega\sqrt{\varepsilon\mu}(\mathbf{k}_0\mathbf{X})} \end{cases}$$

est solution du problème II.10.1. C'est ce que nous avons appelé le problème de propagation libre dans un milieu homogène (définition 18). Noter que la solution (II.10.3) est analytique et s'intègre facilement sur des surfaces planes.

Définition 20 (champ diffracté) *On note $(\tilde{\mathbf{E}}, \tilde{\mathbf{H}})$ la solution du problème (II.10.1). Soit*

$$(\mathbf{E}, \mathbf{H}) = (\tilde{\mathbf{E}}, \tilde{\mathbf{H}}) - (\mathbf{E}_0, \mathbf{H}_0)$$

où $(\mathbf{E}_0, \mathbf{H}_0)$ est le champ incident. On dit que $(\mathbf{E}, \mathbf{H})$ est le champ diffracté. Dans le cas où les caractéristiques du domaine Ω sont celles du vide partout $(\mathbf{E}, \mathbf{H})$ est la solution du système

$$\text{(II.10.4)} \qquad \begin{cases} \nabla\wedge \mathbf{E} - i\omega\mathbf{H} = 0 \ \textit{sur}\ \Omega \\ \nabla\wedge \mathbf{H} + i\omega\mathbf{E} = 0 \ \textit{sur}\ \Omega \\ -(1+Q)\mathbf{E}\wedge\nu + (1-Q)\,(\mathbf{H}\wedge\nu)\wedge\nu = g \ \textit{sur}\ \partial\Omega \\ g_{|\Gamma_{int}} = -\left(-(1+Q)\mathbf{E}_0\wedge\nu - (1-Q)\,(\mathbf{E}_0\wedge\nu)\wedge\nu\right)e^{i\omega(\mathbf{k}_0\mathbf{X})} \\ g_{|\Gamma_{ext}} = 0, \quad Q_{|\Gamma_{ext}} = 0 \end{cases}$$

avec $(\mathbf{k}_0, \mathbf{E}_0)$ vérifiant les conditions (II.10.2). Cela correspond à supposer que l'onde plane incidente est issue de la frontière intérieure $\Gamma_{int} \subset \partial\Omega$ ($\Gamma_{int} \neq \emptyset$).

Définition 21 (condition d'absorbant parfait) *La condition d'absorbant parfait correspond à prendre $Q = 0$ sur la frontière de l'objet Γ_{int} dans la condition aux limites (II.7.11), ou de façon équivalente par la condition (II.7.13) pour une impédance Z égale à 1. On a alors*

$$\text{(II.10.5)} \qquad \mathbf{E}\wedge\nu = (\mathbf{H}\wedge\nu)\wedge\nu \ \textit{sur}\ \Gamma_{int}\ .$$

Définition 22 (condition de conducteur parfait) *La condition de conducteur parfait correspond à prendre $Q = 1$ sur la frontière de l'objet Γ_{int} dans la condition aux limites (II.7.11). On obtient*

$$\mathbf{E} \wedge \nu = 0 \ \textit{sur} \ \Gamma_{int} \ . \tag{II.10.6}$$

La direction incidente $\mathbf{k}_0$ du problème de "scattering" est définie dans un repère sphérique par des angles θ et ϕ qui seront par la suite toujours indiqués dans la convention $e^{-i\vec{k}\mathbf{X}}$ en degrés. Les conventions de polarisation, modes TM ou TE, sont aussi indiquées dans la convention $e^{-i\vec{k}\mathbf{X}}$. Toutes ces conventions sont clairement expliquées dans [7] par J.L. Bonnefoy. En revanche, les vecteurs directions de propagation et polarisations sont donnés dans la convention utilisée dans le code $\mathcal{L}ior$ soit $e^{+i\vec{k}\mathbf{X}}$.

Dans tout ce chapitre on utilise deux types d'approximations des champs, types ¡¡ G ¿¿ et ¡¡ H ¿¿.

Définition 23 (Approximation ¡¡ G ¿¿ des courants totaux) *C'est l'approximation naturelle des traces tangentielles utilisant la quantité approchée $\mathcal{X}_h$ des deux côtés d'une interface. Cette approximation est issue des calculs (III.B.62) et (III.B.64) des traces tangentielles de $\mathbf{H}_h$ section III.B.2.2 p. 184. Le courant total J_h, pour un calcul en champ total, est donné par les formules ci-dessous :*
sur une face de bord $\Sigma_{k,k}$,

$$(J_h)_{|\Sigma_{k,k}} = -\frac{1}{2\sqrt{\mu_{kk}}}(g + (1+Q_k)\sum_{l=1}^{2p} x_{kl}\mathcal{Z}_{k,l}) \wedge \nu_{k,k} \ ,$$

sur une face intérieure $\Sigma_{k,j\neq k}$,

$$(J_h)_{|\Sigma_{k,j}} = -\frac{1}{2\sqrt{\mu_{kj}}}(\sum_{l=1}^{2p} \mathcal{X}_{k,l}\mathcal{Z}_{k,l} + \sum_{m=1}^{2p} \mathcal{X}_{j,m}\mathcal{Z}_{j,m}) \wedge \nu_{k,k} \ .$$

Définition 24 (Approximation ¡¡ H ¿¿ des courants totaux) *C'est l'approximation valable seulement pour des matériaux à caractéristiques électromagnétiques réelles et un problème de Maxwell homogène. Cette approximation est issue de la formule (III.B.60) des traces tangentielles de $\mathbf{H}_h$ section III.B.2.1 p. 184. Le courant total J_h, pour un calcul en champ total, est donné par la formule*

$$(J_h)_{|\Sigma_{k,j}} = i\sqrt{\varepsilon_k}\left(\sum_{l=1}^{p} \mathcal{X}_{k,l}\left(\mathbf{E}^0_{k,l} + i\mathbf{E}^0_{k,l} \wedge V_{k,l}\right) + \mathcal{X}_{k,l+p}\left(\mathbf{E}^0_{k,l} - i\mathbf{E}^0_{k,l} \wedge V_{k,l}\right)\right) \wedge \nu_{k,j}.$$

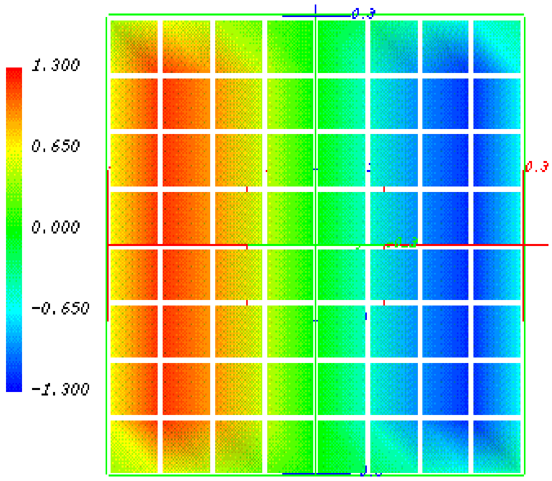
II.10.1.2 Propagation libre en milieu homogène.

Nous nous intéressons à la géométrie de la figure II.10.1 *a*) où $\Gamma_{int} = \emptyset$ pour $L = 60$ centimètres. Les caractéristiques électromagnétiques du domaine de propagation Ω sont constantes et différentes de celles du vide. La solution exacte est donc donnée par la relation (II.10.3). Nous pouvons donc comparer la solution calculée par le code $\mathcal{L}ior$ à la solution analytique. Si les sections II.10.3 et II.10.2 étudient une norme sur le bord, nous nous contentons dans cette section d'effectuer des comparaisons qualitatives visuelles sur les valeurs du courant total sur la frontière extérieure. Cela permet d'observer la différence entre un matériau absorbant, section II.10.1.2.2, et un matériau non absorbant, section II.10.1.2.1. Les caractéristiques du maillage en éléments hexaédriques et de l'onde incidente sont données par le tableau II.10.1 où K est le nombre d'éléments du maillage.

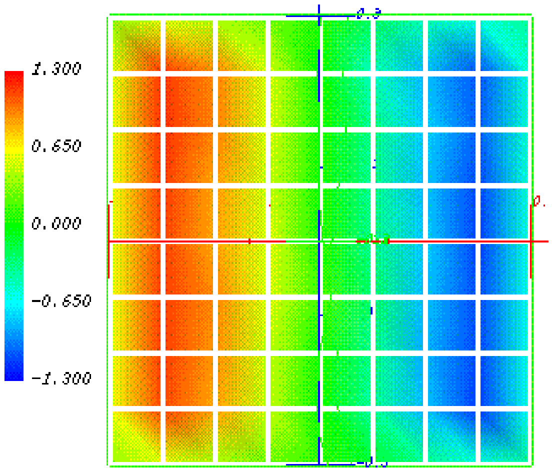
TAB. II.10.1 – Propagation libre en milieu homogène.

(θ, ϕ)	$(90, 0)$	polarisation	TE
$\mathbf{k}_0$	$(-1, 0, 0)$	$\mathbf{E}_0$	$(0, -1, 0)$
K	512	h	0.075

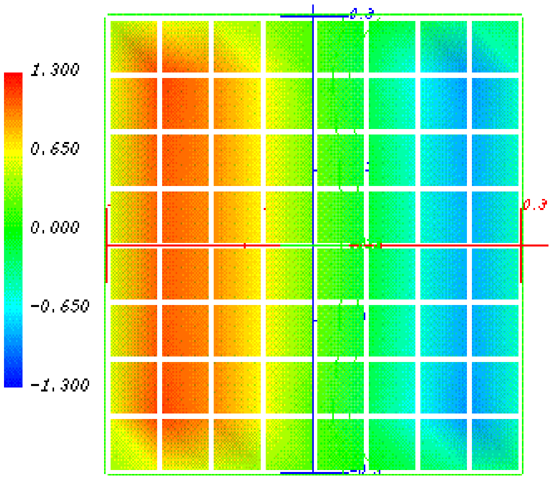
L'onde incidente frappe donc la face $x = L$ et se propage vers les positions décroissantes sur l'axe x.



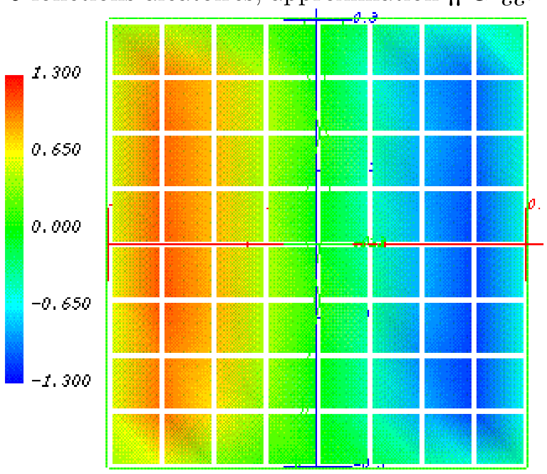
Solution exacte.



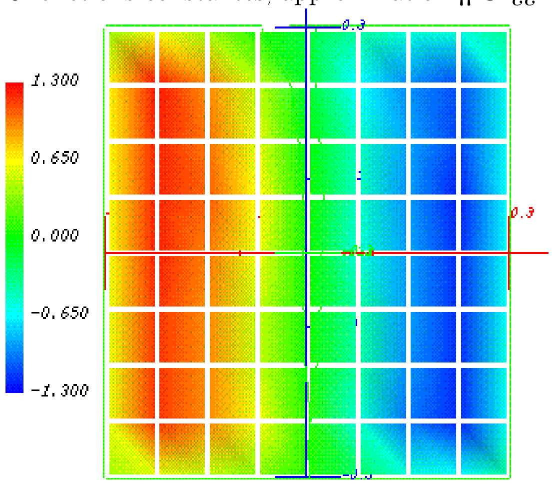
5 fonctions aléatoires, approximation ¡¡ G ¿¿.



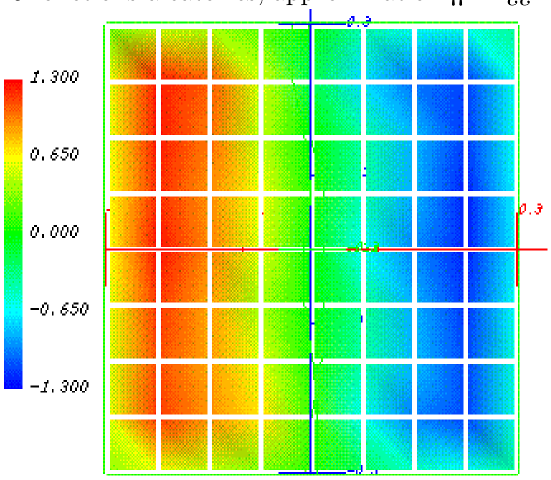
5 fonctions constantes, approximation ¡¡ G ¿¿.



5 fonctions aléatoires, approximation ¡¡ H ¿¿.



10 fonctions aléatoires, approximation ¡¡ G ¿¿.



10 fonctions aléatoires, approximation ¡¡ H ¿¿.

FIG. II.10.2 – Propagation libre en milieu homogène, milieu non absorbant, courant total $\Im(J_x)$.

II.10.1.2.1 Milieu non absorbant.

Nous étudions le cas où les permittivité et perméabilité dans le domaine Ω sont constantes et réelles. Le milieu de propagation est donc non absorbant. Le tableau II.10.2 indique les caractéristiques relatives du milieu (par rapport au vide) et la fréquence de travail, ce qui, en complément du tableau (II.10.1), achève de définir le problème (II.10.1) et sa solution $(\mathbf{E}, \mathbf{H})$ donnée par (II.10.3).

Tab. II.10.2 – Propagation libre en milieu homogène, milieu non absorbant.

f (MHz)	80	(ε, μ)	$(3.22, 2.03)$
λ (vide)	3.75 m	λ (réel)	1.45 m
λ/h (vide)	50	λ/h (réel)	19

Nous représentons, figure II.10.2, la partie imaginaire du courant total $J_t = \mathbf{H} \wedge \nu$ sur la frontière du domaine par une vue dans le plan (x, z) éclairée par l'axe y sens positif (on voit donc la face $y = -L$). On observe donc l'onde plane incidente se propager le long de la frontière extérieure de droite à gauche dans le sens décroissant sur l'axe x. Les simulations ont été effectuées avec 5 ou 10 directions de propagation par élément, constantes ou aléatoires d'un élément à l'autre. Le mode d'approximation des valeurs du courant total varie, c'est soit le mode ¡¡ G ¿¿ (définition 23 p. 140) soit le mode ¡¡ H ¿¿ (définition 24 p. 140).

L'étude visuelle qualitative des valeurs du courant total sur la frontière du domaine Ω des figures II.10.2 nous permet de proposer les observations et les conclusions suivantes.

1. Le type d'approximation ¡¡ G ¿¿ des courants semble de meilleure qualité que le mode d'approximation ¡¡ H ¿¿. Notamment, l'approximation ¡¡ G ¿¿ avec 5 directions est meilleure que l'approximation ¡¡ H ¿¿ avec 10 directions.
2. Les simulations avec des directions de propagation qui varient aléatoirement d'un élément à l'autre à partir de la même position de référence sont meilleures globalement que les simulations avec des directions constantes sur tous les éléments.
3. Nous observons néanmoins une plus grande précision sur la partie gauche de la face $y = -L$ dans le cas de directions constantes. Nous constatons ceci à l'allure du front d'onde, bien droit parallèle à l'axe z (orthogonal à la direction de propagation) avec des directions constantes, plus flou avec des directions aléatoires.
4. Nous déduisons des deux observations précédentes que l'utilisation de directions aléatoires peut améliorer les problèmes d'anisotropie globale de la discrétisation basée sur des fonctions aux directions fixes pour un élément, mais peut au contraire engendrer des problèmes de diffusion numérique localement.
5. L'augmentation du nombre de directions de propagation (entre 5 et 10) améliore globalement la représentation du courant, mais les phénomènes de diffusion du front d'onde augmentent, particulièrement sur la partie droite de la face. Pour 10 directions de propagation aléatoires entre les éléments nous observons une zone floue vers $x = 12$ centimètres et $z = 22$ centimètres dans l'approximation ¡¡ G ¿¿, quatre zones floues dans l'approximation ¡¡ H ¿¿. Nous ne sommes pas en mesure d'expliquer ces phénomènes. Ils sont certainement renforcés par le logiciel de représentation graphique. En effet, nous calculons le courant sur les barycentres des mailles et le logiciel interpole ces quantités pour représenter les courants sur les nœuds. L'interpolation peut augmenter l'effet visuel de flou.

II.10.1.2.2 Milieu absorbant.

Nous remplaçons les caractéristiques du milieu et de la fréquence de travail du problème précédent (section II.10.1.2) par celles du tableau II.10.3. Le milieu a des permittivité et perméabilité complexes, il est donc absorbant. Le tableau II.10.3 nous donne des informations numériques sur l'inversion et le conditionnement maximal des blocs D_k (construits par (II.8.6)) de la matrice D, sur la qualité de son inversion en norme L_∞, sur l'erreur maximale des coefficients de $D^{-1}D$ par rapport à la matrice identité

I, sur le nombre d'itérations effectuées par l'algorithme de résolution finale du système, enfin sur la norme relative dans $L^2(\mathbb{C}^{2pK})$ de la différence du vecteur $(\mathcal{X}^n_{k,l})_{k=1\ldots K, l=1\ldots 2p}$ des coefficients de la solution à l'itération n et du vecteur à l'itération précédente pour la dernière itération du calcul.

TAB. II.10.3 – Propagation libre en milieu homogène, milieu absorbant.

f (MHz)	500	(ε, μ)	$(1 + 0.5 * j, 1)$
λ (vide)	0.63 m	λ (réel)	0.6 m
λ/h (vide)	8	λ/h (réel)	7, 6
Nombre d'itérations	120	$\lvert\mathcal{X}^n - \mathcal{X}^{n-1}\rvert_{L^2(\mathbb{C}^{2pK})}$	$3, 7 * 10^{-4}$
Conditionnement	124	$\lvert D^{-1}D - I\rvert_{L_\infty}$	$2, 7 * 10^{-11}$

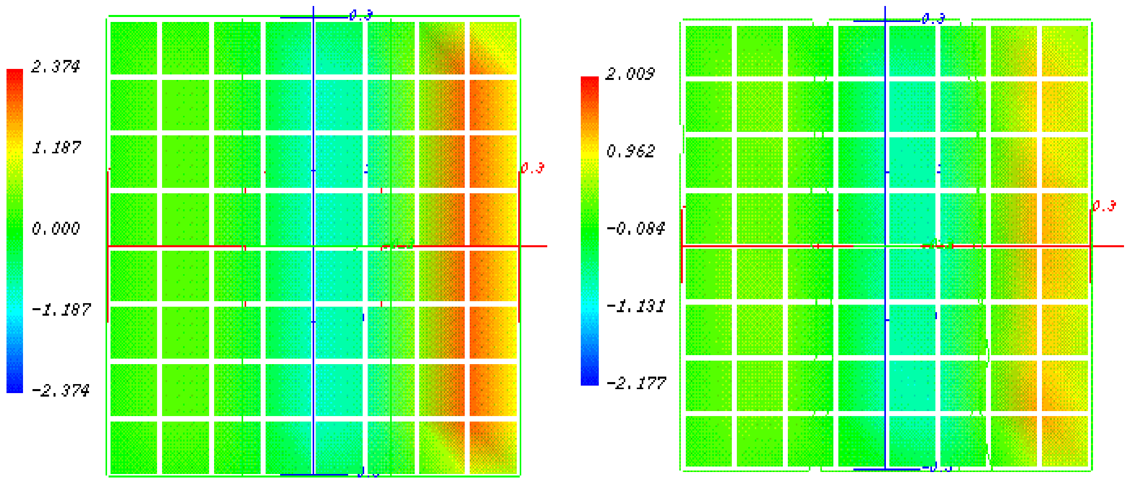
On n'utilise plus que le type d'approximations des champs appelé ¡¡ G ¿¿ (définition 23 p. 140)) de par la présence de matériau absorbant (l'autre type d'approximation n'est plus valable). Nous représentons, figure II.10.3, la partie imaginaire du courant total $J_t = \mathbf{H} \wedge \nu$ par une vue dans le plan (x, z) éclairée dans l'axe y sens positif.

La figure II.10.3 p. 143 montre la décroissance de l'onde incidente le long de la direction de propagation due à la présence d'un matériau absorbant. La qualité du calcul effectué par $\mathcal{L}ior$ nous semble satisfaisante.

II.10.1.2.3 Conclusion de l'étude des courants dans un milieu homogène sans diffraction.

Des valeurs du courant total sur la frontière du domaine Ω des sections II.10.1.2.1 et II.10.1.2.2 nous proposons les conclusions suivantes :

1. le type d'approximation ¡¡ G ¿¿ (définition 23) est plus général que le type d'approximation ¡¡ H ¿¿ (définition 24),
2. la qualité de l'approximation nous semble suffisante.
3. l'utilisation de fonctions de base aléatoires est préférable globalement, mais cette propriété ne nous semble pas générale pour tout problème,
4. le code $\mathcal{L}ior$ est capable de calculer la solution d'un problème avec de fortes absorptions. L'utilisation de fonctions de base issues du problème adjoint du problème de Maxwell permet de discrétiser de façon correcte le problème de Maxwell. Notons que cette propriété théorique n'était pas évidente sur le plan numérique.



Solution exacte. 5 fonctions aléatoires.

FIG. II.10.3 – Propagation libre en milieu homogène, milieu absorbant, courant total $\Re(J_x)$.

II.10.1.3 Notion de SER.

On définit la SER rétro-diffusée dans la direction v par

$$SER = 10 * \log_{10}\left(\frac{\omega^2}{32\pi}|a|^2\right) \tag{II.10.7}$$

où l'amplitude de diffusion a est donnée par

$$|a|^2 = |a^h_{\mathbf{F}}|^2 + |a^h_{\mathbf{G}}|^2 \tag{II.10.8}$$

pour

$$a^h_{\mathbf{F}} = \int_\Gamma ((\mathbf{H}_{\mathbf{F}_v} \wedge \nu) \wedge \nu) . (\mathbf{E}_h \wedge \nu) - (\mathbf{E}_{\mathbf{F}_v} \wedge \nu) . ((\mathbf{H}_h \wedge \nu) \wedge \nu) \tag{II.10.9}$$

(et de même pour $a^h_{\mathbf{G}}$ en remplaçant $(\mathbf{E}_{\mathbf{F}_v}, \mathbf{H}_{\mathbf{F}_v})$ par $(\mathbf{E}_{\mathbf{G}_v}, \mathbf{H}_{\mathbf{G}_v})$) avec

$$\begin{cases} \mathbf{E}_{\mathbf{F}_v} = (e^v_\theta - ie^v_\phi)e^{i\omega\vec{y}e^v_r} \\ \mathbf{E}_{\mathbf{G}_v} = (e^v_\theta + ie^v_\phi)e^{i\omega\vec{y}e^v_r} \end{cases} \quad \text{et} \quad \begin{cases} \mathbf{H}_{\mathbf{F}_v} = i\mathbf{E}_{\mathbf{F}_v} \\ \mathbf{H}_{\mathbf{G}_v} = -i\mathbf{E}_{\mathbf{G}_v} \end{cases}, \tag{II.10.10}$$

où le vecteur $v = e^v_r$ définit les polarisations e^v_θ et e^v_ϕ dans un repère sphérique (cf [7]).

La frontière Γ de calcul de la SER est n'importe quelle surface incluant l'objet diffractant et placée dans le vide.

Remarque 47 Le fait que Γ soit dans le vide indique que l'on peut

- calculer la SER sur la frontière artificielle car elle est toujours placée dans le vide par hypothèse,
- calculer la SER sur deux frontières à savoir la frontière de l'objet décrite par les faces $\Sigma_{k,k}$ là où il n'est pas recouvert de matériau, et l'interface vide/matériau. Cette interface est décrite par les faces $\Sigma_{k,j}$ telle que Ω_k soit dans le vide et Ω_j possède un matériau. Ceci exclut l'éventualité d'un nuage dans le domaine de calcul.

Pour des raisons de précision numérique il est traditionnel d'effectuer le calcul de la SER sur la surface dans le vide la plus proche possible de l'objet, ce qui correspond à la deuxième situation.

On montre que le calcul de la SER, comme le calcul d'erreur par rapport à une solution exacte qui serait une onde plane, est de la même complexité que le calcul des termes matriciels pour la contribution du bord Γ du domaine Ω. Nous ne développerons pas ces calculs. Notons simplement que, pour,

$$\begin{aligned} \mathbf{F}_{-v} &= (e^v_\theta + ie^v_\phi)e^{-i\omega\vec{y}e^v_r} \\ \mathbf{G}_{-v} &= -(e^v_\theta - ie^v_\phi)e^{-i\omega\vec{y}e^v_r} \end{aligned}, \tag{II.10.11}$$

la contribution des faces de bord pour l'amplitude de diffusion est donnée par les formules

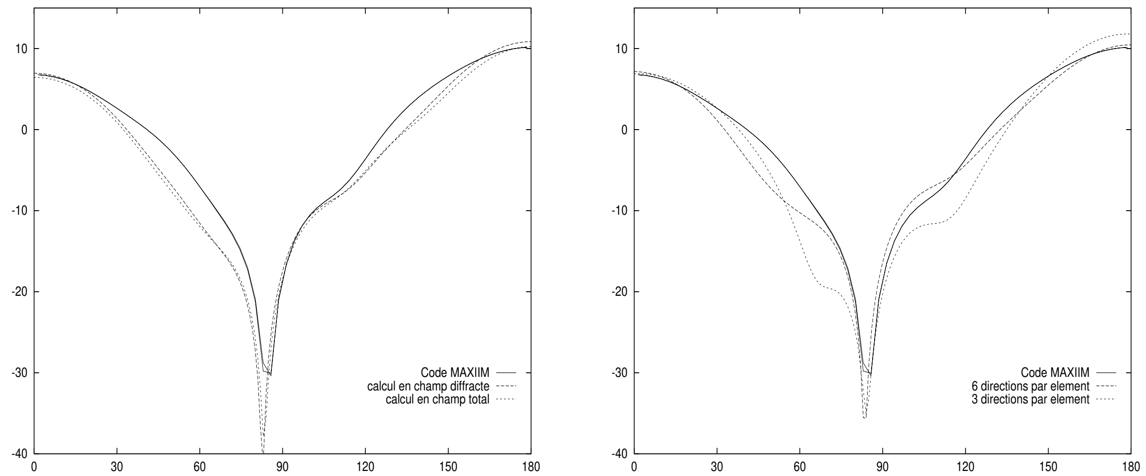
$$\begin{aligned} 2a^h_{\mathbf{F}} = &\sum_{m=1}^{p} \sum_{\Sigma_{k,k}\subset\Gamma} \mathcal{X}^{\mathbf{F}}_{k,m} \int_{\Sigma_{k,k}} \overline{(+\mathbf{F}_{-v} \wedge \nu_k + i\,(\mathbf{F}_{-v} \wedge \nu_k) \wedge \nu_k)}.\,(\mathbf{F}_{k,m} \wedge \nu_k + i\,(\mathbf{F}_{k,m} \wedge \nu_k) \wedge \nu_k) \\ &- \sum_{m=1}^{p} \sum_{\Sigma_{k,k}\subset\Gamma} \mathcal{X}^{\mathbf{G}}_{k,m} \int_{\Sigma_{k,k}} Q_k \overline{(-\mathbf{F}_{-v} \wedge \nu_k + i\,(\mathbf{F}_{-v} \wedge \nu_k) \wedge \nu_k)}.\,(+\mathbf{G}_{k,m} \wedge \nu_k - i\,(\mathbf{G}_{k,m} \wedge \nu_k) \wedge \nu_k) \\ &- \sum_{\Sigma_{k,k}\subset\Gamma} \int_{\Sigma_{k,k}} \overline{(-\mathbf{F}_{-v} \wedge \nu_k + i\,(\mathbf{F}_{-v} \wedge \nu_k) \wedge \nu_k)}.\,(g) \end{aligned}$$

et

$$\begin{aligned} 2a^h_{\mathbf{G}} = &\sum_{m=1}^{p} \sum_{\Sigma_{k,k}\subset\Gamma} \mathcal{X}^{\mathbf{G}}_{k,m} \int_{\Sigma_{k,k}} \overline{(+\mathbf{G}_{-v} \wedge \nu_k - i\,(\mathbf{G}_{-v} \wedge \nu_k) \wedge \nu_k)}.\,(\mathbf{G}_{k,m} \wedge \nu_k - i\,(\mathbf{G}_{k,m} \wedge \nu_k) \wedge \nu_k) \\ &- \sum_{m=1}^{p} \sum_{\Sigma_{k,k}\subset\Gamma} \mathcal{X}^{\mathbf{G}}_{k,m} \int_{\Sigma_{k,k}} Q_k \overline{(-\mathbf{G}_{-v} \wedge \nu_k - i\,(\mathbf{G}_{-v} \wedge \nu_k) \wedge \nu_k)}.\,(+\mathbf{F}_{k,m} \wedge \nu_k + i\,(\mathbf{F}_{k,m} \wedge \nu_k) \wedge \nu_k) \\ &- \sum_{\Sigma_{k,k}\subset\Gamma} \int_{\Sigma_{k,k}} \overline{(-\mathbf{G}_{-v} \wedge \nu_k - i\,(\mathbf{G}_{-v} \wedge \nu_k) \wedge \nu_k)}.\,(g) \ . \end{aligned}$$

Un calcul similaire montre que la contribution des interfaces internes entre des éléments Ω_k dans le vide et Ω_j quelconque, s'obtient aussi comme des calculs de termes du système linéaire par des couplages des fonctions de base avec les fonctions $\mathbf{F}_{-v}$ et $\mathbf{G}_{-v}$ qui ne dépendent que de la direction d'observation.

II.10.1.4 Diffraction dans le vide sur un cube.



Champ diffracté ou total, 6 directions aléatoires. 3 ou 6 directions constantes en champ diffracté.

FIG. II.10.4 – Comparaison $\mathcal{M}axiim/\mathcal{L}ior$ en conducteur parfait.

Nous considérons le problème de “scattering” (diffraction d’une onde plane incidente sur un objet) dans le vide du tableau II.10.4 avec la géométrie de la figure II.10.1 *b*) avec $a = 30$ cm et $b = 50$ cm.

TAB. II.10.4 – Diffraction dans le vide sur un cube.

(θ, ϕ)	$(0, 0)$	polarisation	TM
$\mathbf{k}_0$	$(0, 0, -1)$	$\mathbf{E}_0$	$(0, -1, 0)$
p	3 ou 6	K	5711
$(h, a, b)(cm)$	$(5.2, 30, 60)$	f (MHz)	500
λ	0.6 m	λ/h	11.6

Nous ne connaissons pas la solution exacte. Nous comparons nos calculs de Section Efficace Radar (cf section II.10.1.3) à un code de référence, le code $\mathcal{M}axiim$, appelé ainsi car il utilise une méthode de résolution des équations de Maxwell par une équation intégrale et un algorithme itératif (MAXwell Integral Iterative Method). Ce code est l’œuvre de Bruno Després.

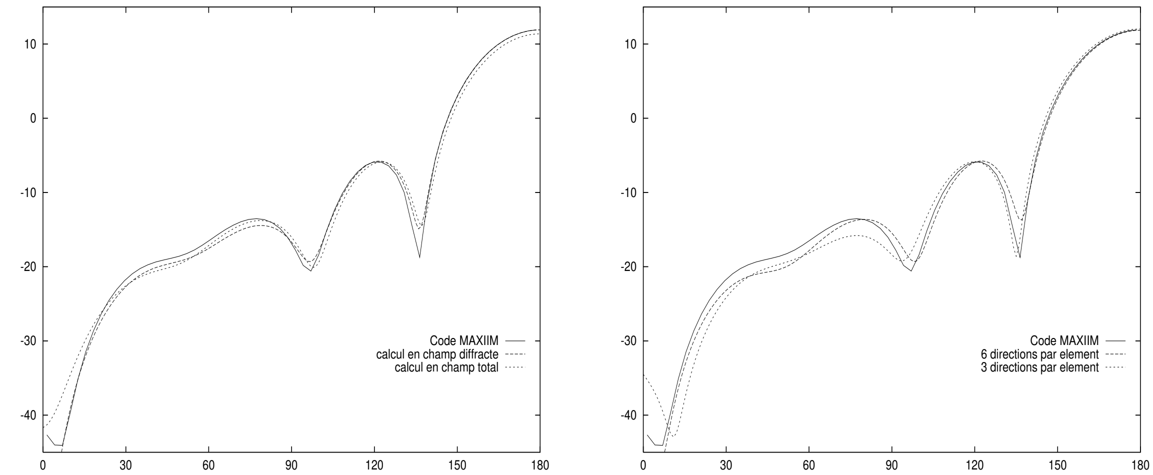
Nous calculons la SER en balayage bistatique de 0 à 180 degrés pour la fréquence et l’incidence données. Nous comparons nos résultats à ceux du code $\mathcal{M}axiim$ pour deux types de conditions aux limites sur le cube :

a) une condition de conducteur parfait sur l’objet (II.10.6), figures II.10.4, II.10.6 et II.10.7,

b) une condition d’absorbant parfait sur l’objet (II.10.5), figures II.10.5, II.10.8 et II.10.9.

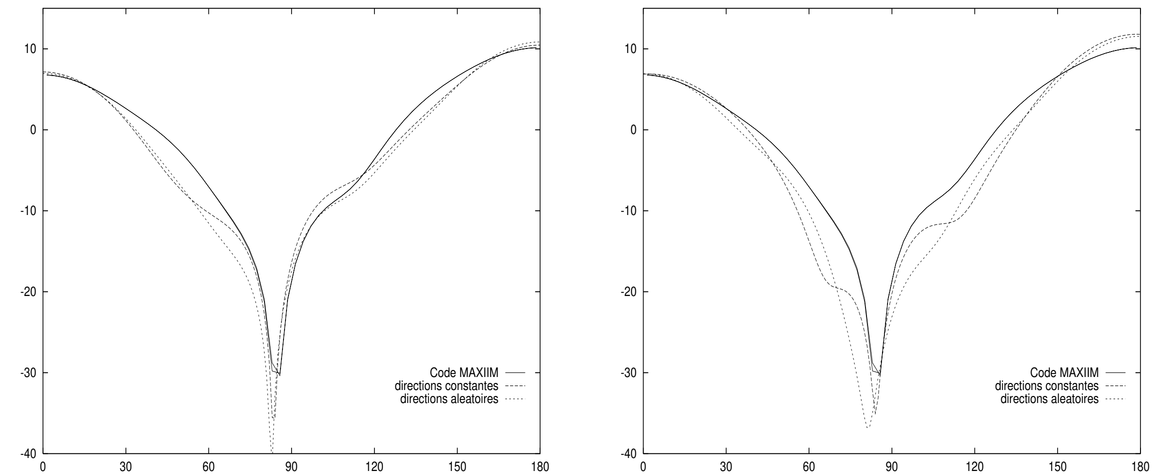
Nous comparons successivement plusieurs facteurs importants du code $\mathcal{L}ior$.

1. Il faut vérifier la convergence des calculs effectués par $\mathcal{L}ior$ vers une solution unique. Pour cela, on augmente le nombre de directions de propagation des fonctions de base de façon à obtenir une courbe limite de SER. En outre, il faut vérifier que, pour les courbes limites obtenues, les calculs effectués en champ total ou en champ diffracté (cf section II.10.1.1) donnent des résultats identiques. Se reporter aux figures II.10.4 et II.10.5
2. Nous regardons la dépendance du code $\mathcal{L}ior$ par rapport au choix des fonctions de base entre les éléments du maillage en les choisissant soit constantes soit aléatoires (figures II.10.6 et II.10.8).
3. Nous étudions s’il est plus intéressant d’effectuer un calcul en champ diffracté ou en champ total lorsque la discrétisation n’est pas optimale (figures II.10.5 et II.10.9).



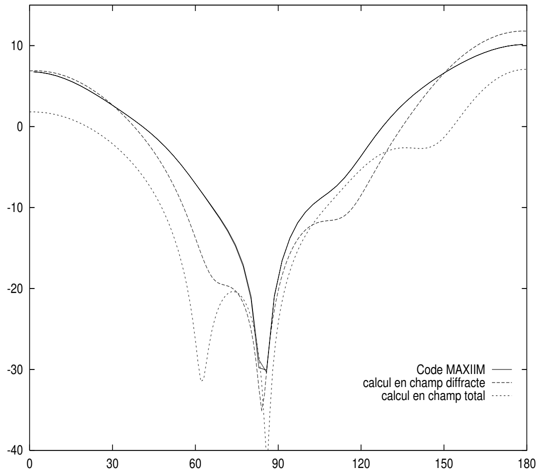
Champ diffracté ou total, 6 directions aléatoires. 3 ou 6 directions constantes en champ diffracté.

FIG. II.10.5 – Comparaison $\mathcal{M}axiim/\mathcal{L}ior$ en absorbant parfait.

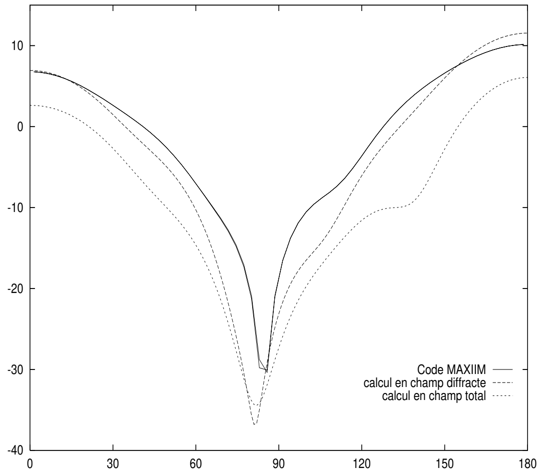


Champ diffracté, 6 directions. Champ diffracté, 3 directions.

FIG. II.10.6 – Directions aléatoires ou constantes, conducteur parfait.

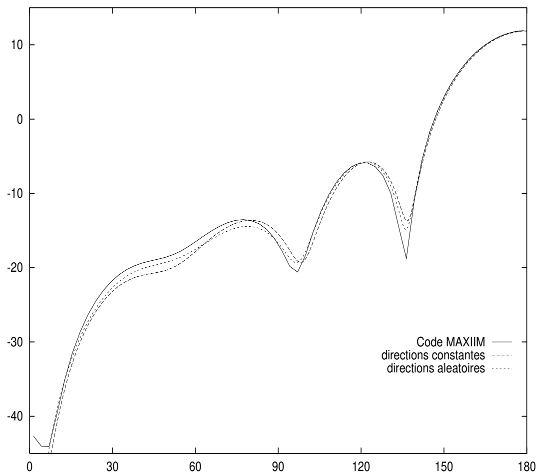


3 directions aléatoires.

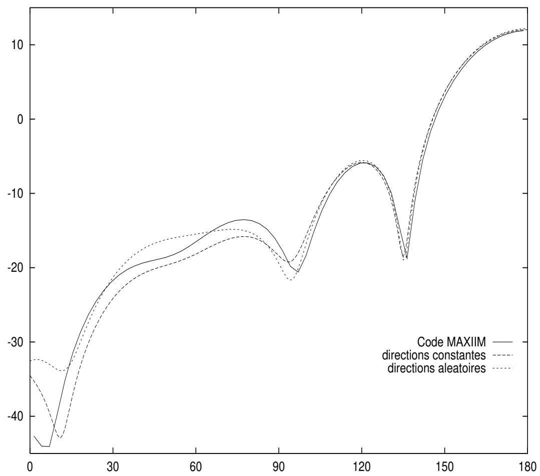


3 directions constantes.

FIG. II.10.7 – Calculs en champ total ou diffracté, conducteur parfait.



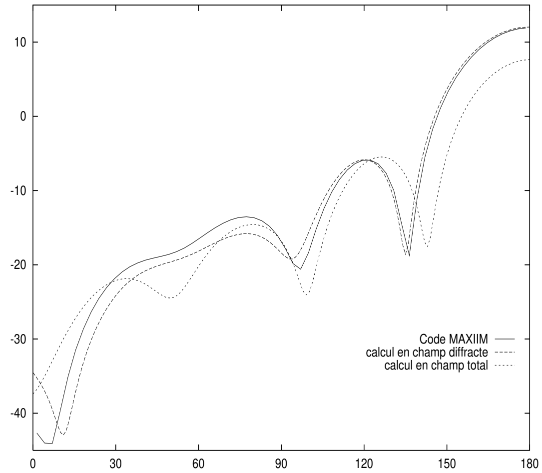
Champ diffracté, 6 directions.



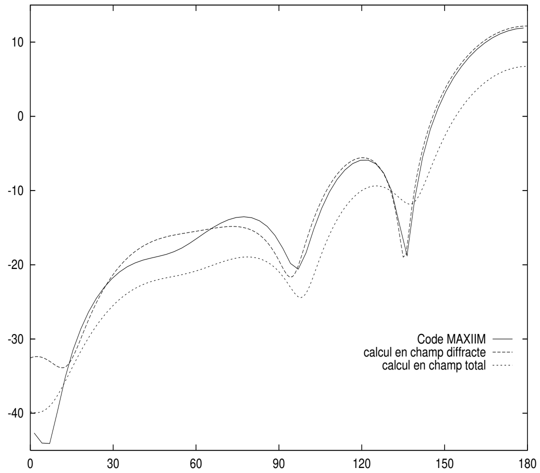
Champ diffracté, 3 directions.

FIG. II.10.8 – Directions aléatoires ou constantes, absorbant parfait.

Nous constatons l'influence de la Condition aux Limites Absorbante sur les figures II.10.4 et II.10.5 où le code $\mathcal{L}ior$ atteint une discrétisation suffisante (notamment le calcul en champ total ou en champ diffracté donne le même résultat pour des fonctions tirées aléatoirement ou constantes). Ceci n'est guère surprenant puisque la surface sur laquelle on impose la CLA est très proche de l'objet. Nous constatons aussi que le choix des directions par un tir aléatoire ou non n'est pas essentiel. Enfin nous constatons que le calcul effectué en champ diffracté est plus précis comme c'était le cas pour le problème de Helmholtz bidimensionnel. Il semble qu'approcher la solution en champ total revient grossièrement à calculer la solution en champ diffracté et à approcher l'onde incidente. Nous avions déjà observé ce phénomène lors de l'étude du problème de Helmholtz dans le vide.



3 directions aléatoires.



3 directions constantes.

FIG. II.10.9 – Calculs en champ total ou diffracté, absorbant parfait.

II.10.1.5 Diffraction sur un cône.

Nous nous intéressons à la géométrie de la figure II.10.1 c) (cf [42]) du cône déjà étudié par Christophe Le Potier pour le problème du tableau II.10.5. Le nombre total de tétraèdres du maillage est K_{total}, le nombre d'éléments dans lequel se trouve éventuellement un matériau est K_{couche}, répartis en K_1 éléments sur la pointe et K_2 éléments sur le culot. Cette géométrie est non triviale et commence à être plus

TAB. II.10.5 – Diffraction sur le cône.

(θ, ϕ)	$(0,0)$	polarisation	TM	h (cm)	$3,97$
$\mathbf{k}_0$	$(0,0,-1)$	$\mathbf{E}_0$	$(-1,0,0)$	p	3 ou 6
K_{total}	17684	K_{couche}	1536	f	300 MHz
K_1	612	K_2	924	λ	1.0 m

intéressante pour les applications. En effet, elle possède une pointe de diffraction dans l'axe d'éclairement, un cercle de discontinuité C^1 au culot (raccord entre le cône et le disque délimitant la surface de l'objet). De plus, la géométrie est axi-symétrique, ce qui nous permet de comparer nos calculs à ceux du code axi-symétrique équations intégrales / éléments finis couplés, code SHFC développé par Roland Le Martret et Bruno Stupfel. Le nombre de nœuds utilisés dans le maillage bidimensionnel axi-symétrique pour SHFC est 1112, le nombre de quadrangles est 1927.

Nous avons considéré les trois cas suivants :

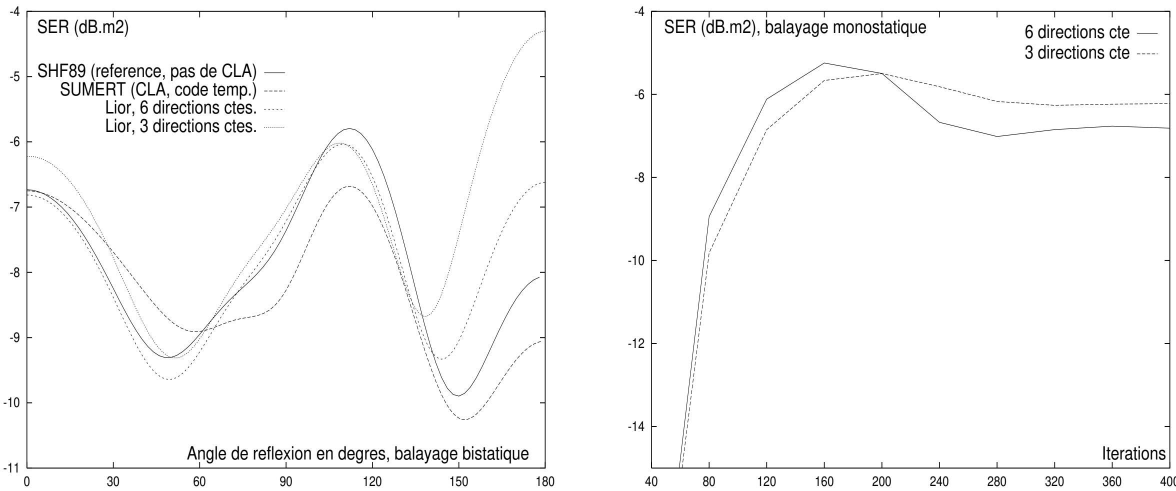
1. le cône parfaitement conducteur est nu,
2. le cône parfaitement conducteur est revêtu de deux matériaux non absorbants,

3. le cône parfaitement conducteur est revêtu de deux matériaux absorbants.

Dans les deux premiers cas nous comparons aussi au code SUMERT développé par Christophe Le Potier et Roland Le Martret. SUMERT est un code de résolution des équations de Maxwell temporelles par une méthode de volumes finis. On obtient les résultats en fréquence par une transformée de Fourier. L'intérêt de la comparaison avec le code SUMERT est que nous avons disposé du même maillage volumique. Dans le premier cas nous comparons aussi à un code haute fréquence, SHF89 développé par Pierre Bonnemason et Bruno Stupfel.

II.10.1.5.1 Conducteur parfait dans le vide.

Ce cas correspond dans la figure II.10.1 *c*) à enlever les couches de matériaux 1 et 2 (ou à prendre le matériau fictif dont les caractéristiques sont celles du vide). Les caractéristiques du cas sont données dans le tableau II.10.5.



Comparaisons à des codes de référence. Convergence : fonction du nombre d'itérations.

FIG. II.10.10 – Cône dans le vide.

Nous représentons, figure II.10.10, les SER comparées entre SUMERT, SHF89 et $\mathcal{L}ior$, pour une convergence suffisante (pour 6 directions de propagation constantes par élément). Nous étudions aussi sommairement le coût de $\mathcal{L}ior$ en observant l'évolution de la SER monostatique en fonction du nombre d'itérations effectuées. Il est clair qu'optimiser le nombre d'itérations de l'algorithme de résolution du système linéaire en fonction de la précision globale du code sera un point important d'un développement ultérieur de $\mathcal{L}ior$. Nous débutons cette réflexion dans la section II.10.1.5.4.

II.10.1.5.2 Conducteur revêtu de matériaux diélectriques.

Le cône contenu dans un cylindre de la section II.10.1.5.1 est revêtu d'une couche de matériau non absorbant comme dans l'étude menée par Christophe Le Potier dans [42]. Rappelons que la couche de matériau est d'épaisseur 5 centimètres, constante par morceau, à la pointe pour z positif de perméabilité relative $1,5$ (permittivité relative 2), au culot pour z négatif de perméabilité relative 2 (permittivité relative $1,5$).

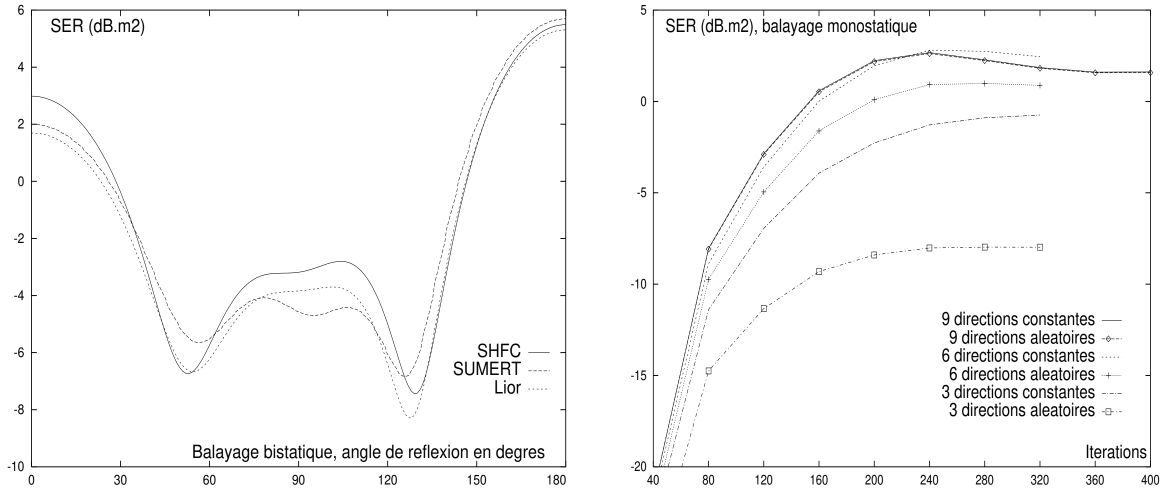
Le balayage effectué est celui du tableau II.10.5 avec les longueurs d'ondes dans le matériau et dans le vide calculées dans le tableau II.10.6.

Nous comparons figure II.10.11 les niveaux de SER pour le balayage bistatique donné à ceux obtenus par les codes SHFC et SUMERT. Nous constatons que pour un nombre suffisant de directions de propagation, le code $\mathcal{L}ior$ est assez précis, plus que SUMERT. Nous étudions la convergence de l'algorithme itératif pour différentes directions de propagation pour le calcul de la SER monostatique.

On constate que, plus la précision finale du calcul est élevée, plus grand est le nombre d'itérations nécessaires pour obtenir une SER parfaitement stable. Nous remarquons aussi que le choix de directions

TAB. II.10.6 – Diffraction sur le cône avec une couche de matériau réel.

(ε, μ) (pointe)	$(2, 1.5)$	(ε, μ) (culot)	$(1.5, 2)$
λ (vide)	1 m	λ/h (vide)	$25, 2$
λ (matériau)	$0,588$m	λ/h (matériau)	$14, 5$



Comparaisons pour 6 directions aléatoires. SER monostatique.

FIG. II.10.11 – Cône avec matériau non absorbant.

aléatoires n'est pas judicieux ici. En effet, on obtient une précision plus grande avec moins de fonctions de base en les choisissant constantes, ceci est frappant en comparant les courbes pour 6 directions constantes et 7 directions aléatoires. Nous vérifions que pour suffisamment de fonctions de base, en l'occurrence 9 directions de propagation par élément, les tirs aléatoires ou constants des directions donnent le même résultat.

Nous vérifions que le nombre de fonctions de base choisi est suffisant figure II.10.12. En augmentant le nombre de directions de propagation, on constate que la courbe de SER dans ce balayage bistatique n'évolue plus guère entre 9 et 7 directions par élément. En outre que pour 9 directions il est indifférent de choisir un tir aléatoire ou constant des directions d'un élément à l'autre. En revanche nous vérifions comme sur la figure de droite II.10.11 qu'il est préférable d'employer toujours les mêmes directions de propagation pour tous les éléments lorsque le nombre de fonctions choisies reste faible.

Nous représentons, figure II.10.13 p. 152, le module du courant total $J_t = \mathbf{H} \wedge \nu$, aux nœuds du maillage pour le calcul par SUMERT, et aux barycentres des mailles pour $\mathcal{L}ior$ sur l'interface vide / matériau. Les trois figures du haut sont dans le plan $(-y, z)$ et celles du bas dans le plan $(-y, -x)$. Les plans entre les figures du haut et du bas correspondent. La normale est définie dans le sens : vide vers matériau. Le mode de représentation (aux nœuds ou aux barycentres des mailles) augmente l'effet visuel de différence entre $\mathcal{L}ior$ et SUMERT. Notons que SUMERT n'est pas forcément la bonne référence, mais plutôt l'approximation ¡¡ H ¿¿.

II.10.1.5.3 Conducteur revêtu d'un matériau absorbant.

Nous remplaçons les matériaux aux caractéristiques réelles et non absorbantes de la section II.10.1.5.2 par des matériaux absorbants dont les valeurs sont consignées dans le tableau II.10.7.

La longueur d'onde et le rapport au paramètre de raffinement du maillage sont données dans le vide et dans le matériau. Le balayage effectué est toujours celui du tableau II.10.5.

Nous comparons, figure II.10.14, les balayages effectués par le code SHFC et le code $\mathcal{L}ior$, pour un balayage bistatique puis pour un balayage fréquentiel en monostatique sur une petite plage de fréquences. Ces calculs sont effectués pour 6 directions de propagation par élément, toujours les mêmes d'un élément

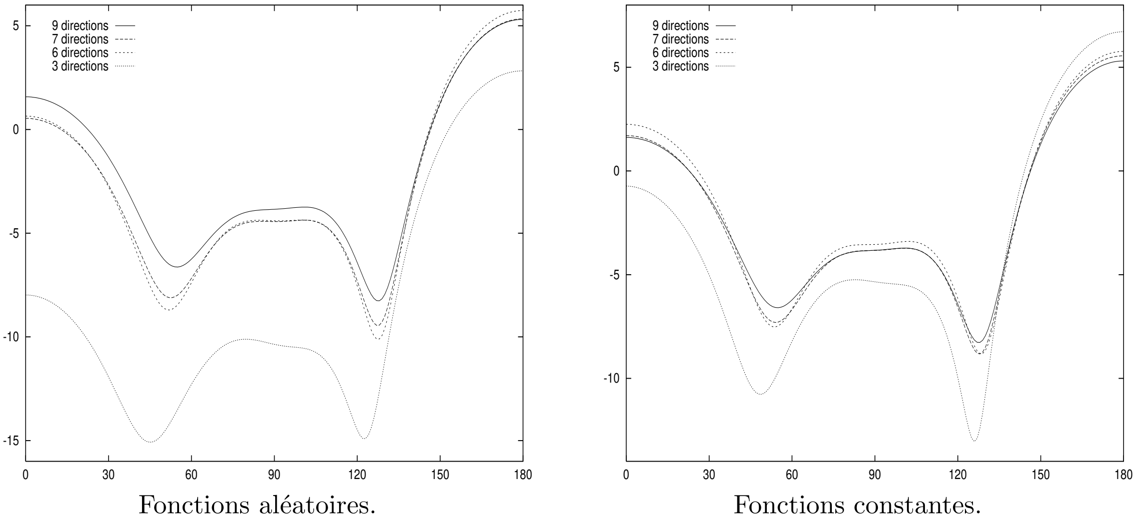


FIG. II.10.12 – Cône avec matériau non absorbant, balayage bistatique.

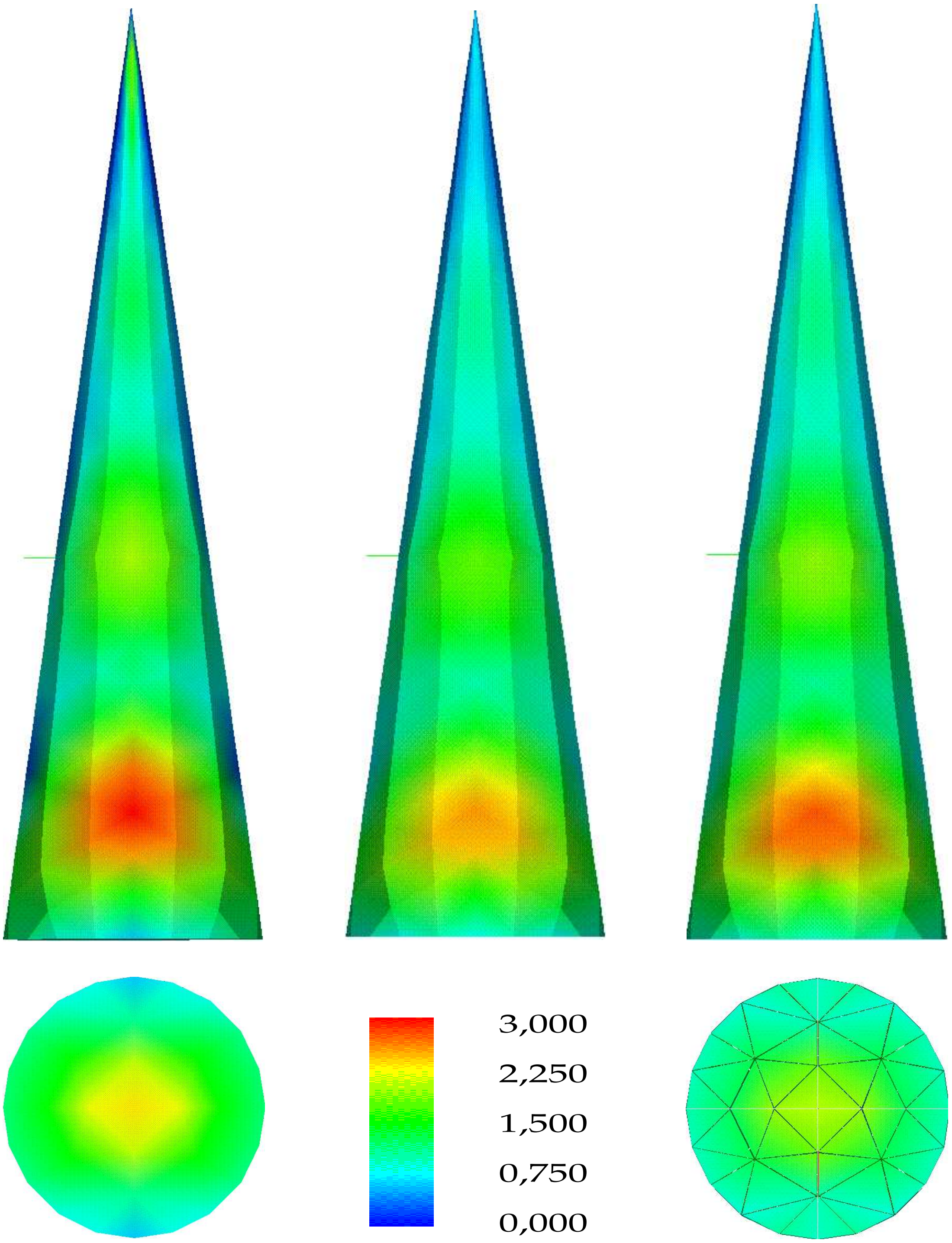
TAB. II.10.7 – Diffraction sur le cône avec une couche de matériau complexe.

(ε, μ) (pointe)	$(2 + 0.5j, 1.5)$	(ε, μ) (culot)	$(1.5 + 0.5j, 2)$
λ (vide)	1 m	λ/h (vide)	$25,2$
λ (matériau)	$0,562$ m	λ/h (matériau)	$14,2$

à l'autre. La différence dans les deux cas est inférieure à 1 $dB.m^2$ de SER.

Nous représentons, figure II.10.15 p. 154, les valeurs des modules des courants aux barycentres des faces de l'interface vide matériau, de l'objet, et enfin de la frontière artificielle pour mesurer l'influence de la CLA sur notre méthode. Nous pouvons considérer que la CLA est positionnée suffisamment loin puisque l'on observe des courants proches de ceux donnés par l'onde incidente seule (se propageant dans le vide).

La figure II.10.16 étudie l'évolution de la SER du balayage fréquentiel effectué par le code $\mathcal{L}ior$ en une fois de 300 à 340 MHz par pas de 10 MHz en fonction du nombre d'itérations effectuées pour la résolution du système linéaire. En effet, nous voyons, section II.10.1.5.4, que le temps de calcul de $\mathcal{L}ior$ est fortement conditionné par le nombre d'itérations effectuées. Lors d'un balayage fréquentiel ou angulaire, on peut, lorsque le premier calcul en fréquence ou en angle est effectué, repartir de la valeur calculée de la solution pour l'étape suivante. Lorsque la variation par rapport aux données est petite, la variation de la solution est généralement faible aussi. Nous avons imposé à $\mathcal{L}ior$ d'effectuer systématiquement 400 itérations pour toutes les fréquences, la figure II.10.16 permet d'observer l'évolution de la SER bistatique. Il appartiendra à l'utilisateur de décider du nombre minimal d'itérations à effectuer en fonction de la précision qui lui convient. Notons que si la discrétisation est faible il n'est pas nécessaire de pousser l'algorithme itératif vers une précision absolue, de même si la frontière bornant le domaine est trop proche de l'objet. Le critère d'arrêt de l'algorithme peut s'obtenir ici d'après le calcul de référence effectué à l'aide de SHFC (cf figure II.10.14). En l'absence de solution "exacte" de référence, on peut aussi cherchant la solution approchée la meilleure en effectuant plusieurs simulations, comme cela nous semble clair section II.10.1.5.2, figure II.10.12 lors de l'étude du cône recouvert de matériaux au caractéristiques réelles. Il est alors théoriquement possible de diminuer fortement le temps de calcul du code $\mathcal{L}ior$ alors que dans SHFC, le temps de calcul reste constant par fréquence, en l'occurrence de 9 secondes.

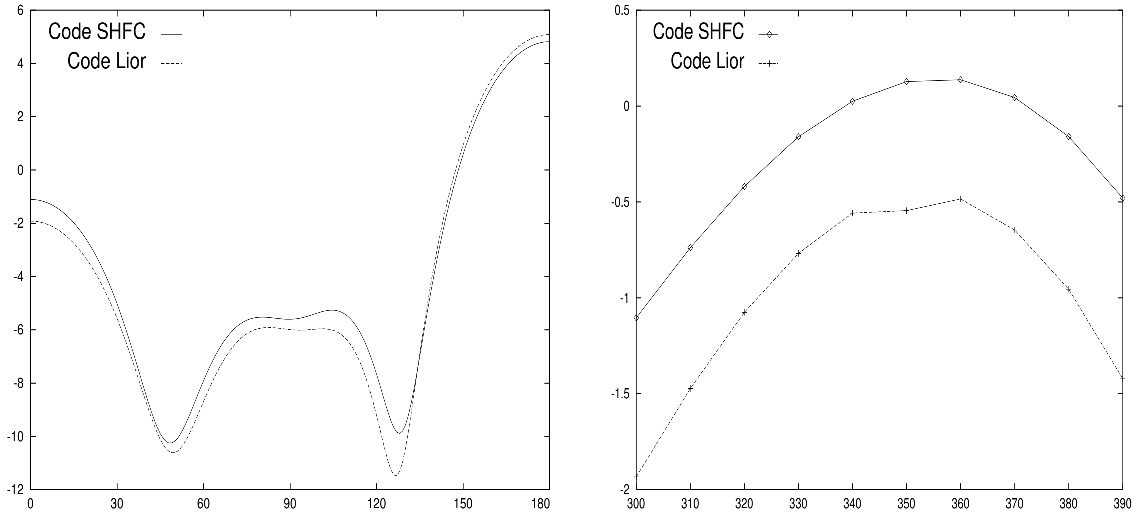


Code SUMERT.

6 fonctions constantes, approximation ¡¡ G ¿¿.

6 fonctions constantes, approximation ¡¡ H ¿¿.

FIG. II.10.13 – Matériau réel sur le cône, modules des courants, échelle 1/12.



Balayage bistatique — Balayage fréquentiel.

FIG. II.10.14 – Cône avec matériau absorbant, comparaison avec SHFC.

II.10.1.5.4 Etude du coût.

Les trois sections précédentes (II.10.1.5.1, II.10.1.5.2 et II.10.1.5.3) ont abordé la question du coût du code $\mathcal{L}ior$ en terme qualitatif de temps de calcul. Nous étudions ici

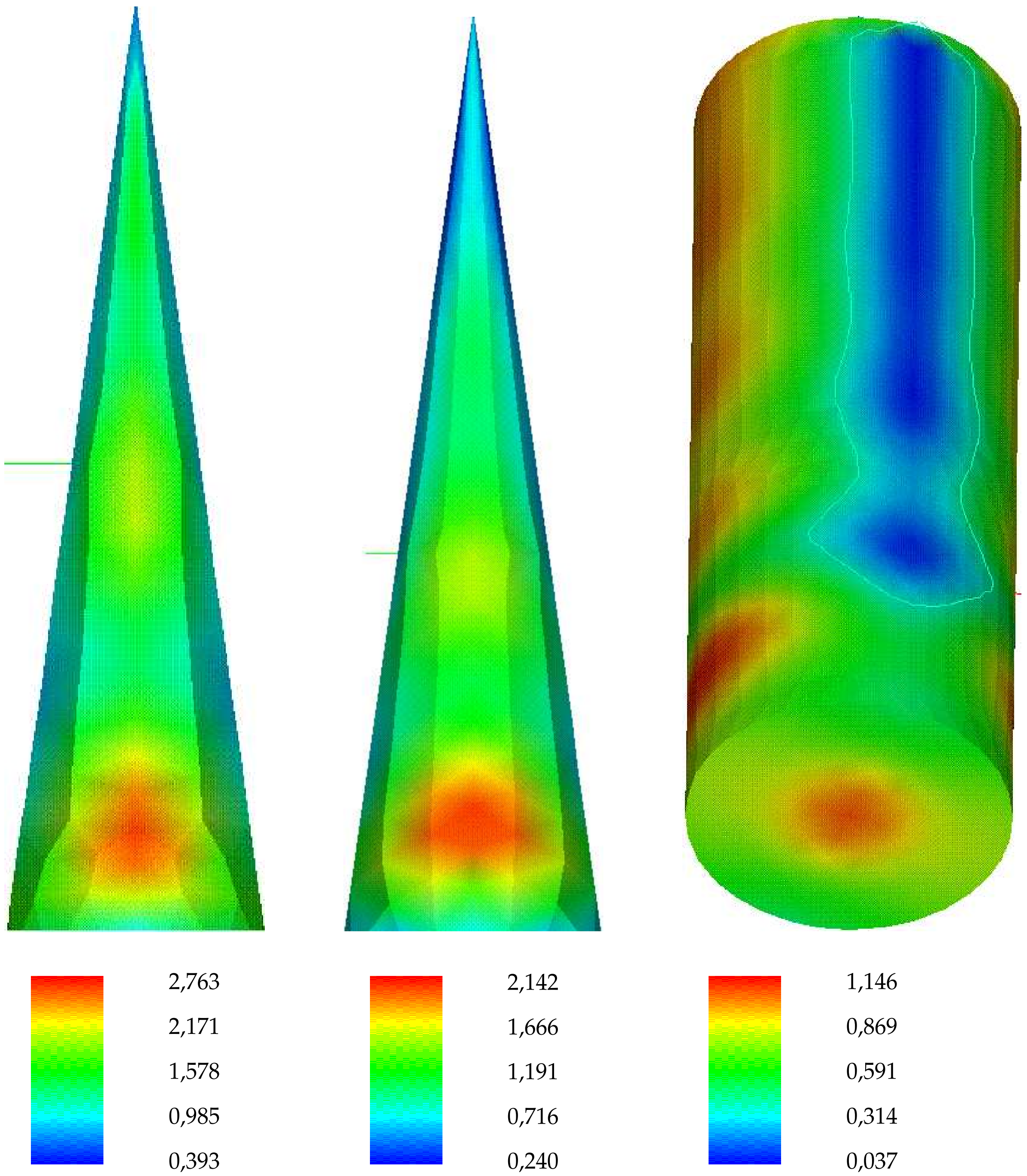
a) le nombre d'itérations à effectuer pour un critère de convergence fort ou faible, pour un balayage en fréquence ou angulaire. Pour chaque fréquence (ou chaque angle d'incidence) l'algorithme itératif converge selon le même critère d'évolution de la SER au cours des itérations. Notez la différence avec la situation de la figure II.10.16 où le nombre d'itérations est constant.

b) le temps de calcul en fonction du nombre d'itérations, du nombre de fonctions de base, du type de la couche de matériau (vide, matériau absorbant ou non), du choix aléatoire ou constant des directions de propagation entre les éléments. Nous cherchons les droites de coût (t_T) du code $\mathcal{L}ior$ en fonction du nombre d'itérations (N_i) effectuées par l'algorithme de résolution du système linéaire (une itération est effectuée en un temps t_1) pour une incidence et une fréquence de calcul données, à l'exception des post-traitements (ces tâches sont effectuées en un temps t_0). Ces droites sont d'équation

$$t_T = t_0 + N_i \times t_1 \ .$$

Pour ce faire, nous effectuons les simulations suivantes :

1. La figure II.10.17 de gauche donne le nombre d'itérations effectuées par l'algorithme itératif de résolution du système linéaire pour un critère fort de convergence de l'algorithme itératif pour lequel on obtient une courbe très proche de celle de la figure II.10.14 à droite (point II.10.1.5.4 ci-dessus). La simulation est effectuée sur le cône revêtu de matériau absorbant pour 6 fonctions de base constantes et le balayage fréquentiel de la section II.10.1.5.3.
2. La figure II.10.18 de gauche étudie le nombre d'itérations du balayage fréquentiel pour un critère faible de convergence de l'algorithme (point II.10.1.5.4) pour les mêmes données que la figure II.10.17 de gauche.
3. La figure II.10.18 de droite nous donne une idée de la précision du code $\mathcal{L}ior$ pour le critère faible de convergence par comparaison à la courbe obtenue par SHFC (point II.10.1.5.4) pour les mêmes données que la figure II.10.17 de gauche. Cette courbe est à comparer à la courbe II.10.14.
4. La figure II.10.17 de droite donne les droites de coût du code $\mathcal{L}ior$ du point II.10.1.5.4 en fonction du nombre de directions de propagation. Les directions de la simulation sont aléatoires d'un élément à l'autre, la géométrie est celle du cône revêtu d'un matériau non absorbant.

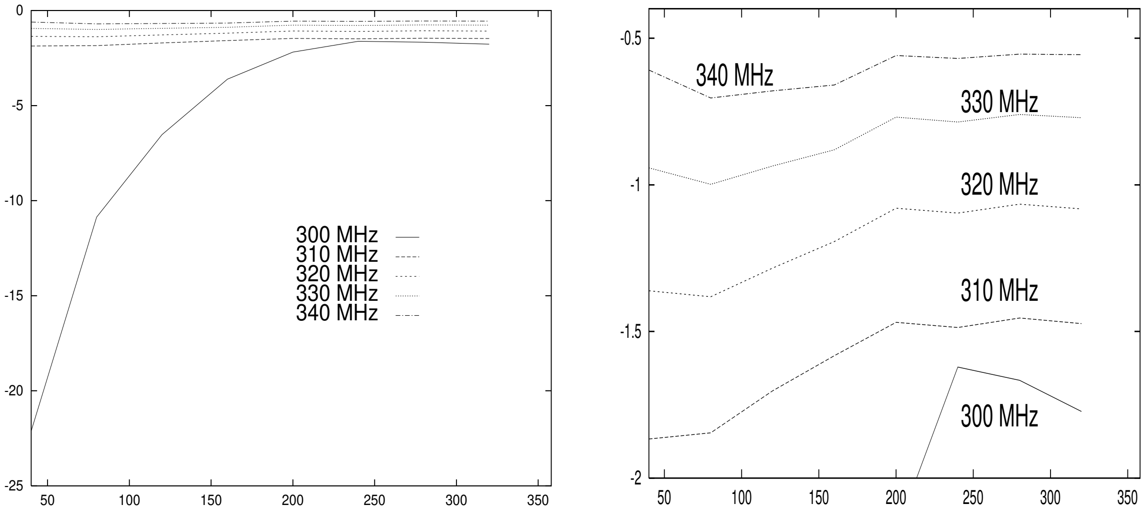


Frontière de l'objet.
Echelle 1/9.

Interface vide matériau.
Echelle 1/12.

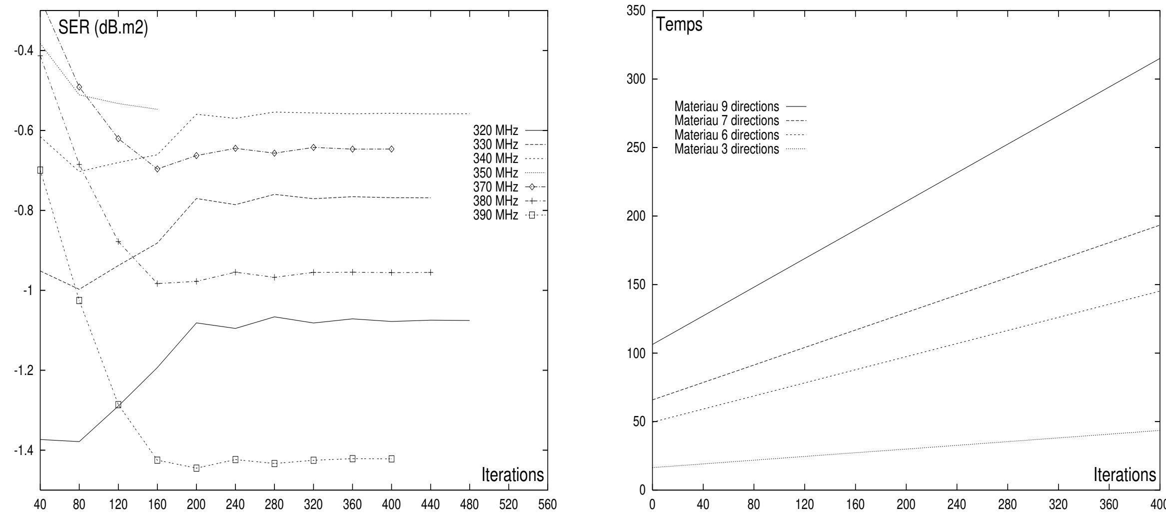
Frontière extérieure.
Echelle 1/21.

FIG. II.10.15 – Module du courant total avec un matériau complexe.



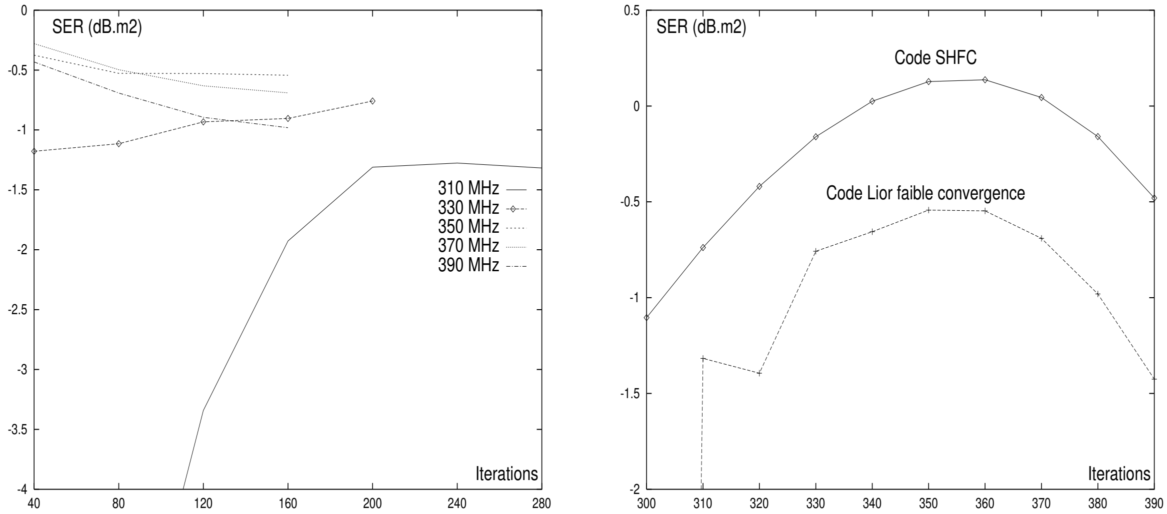
Plage complète de SER. Zoom sur la zone de convergence.

Fig. II.10.16 – Evolution de la SER en fonction des itérations pour un balayage fréquentiel



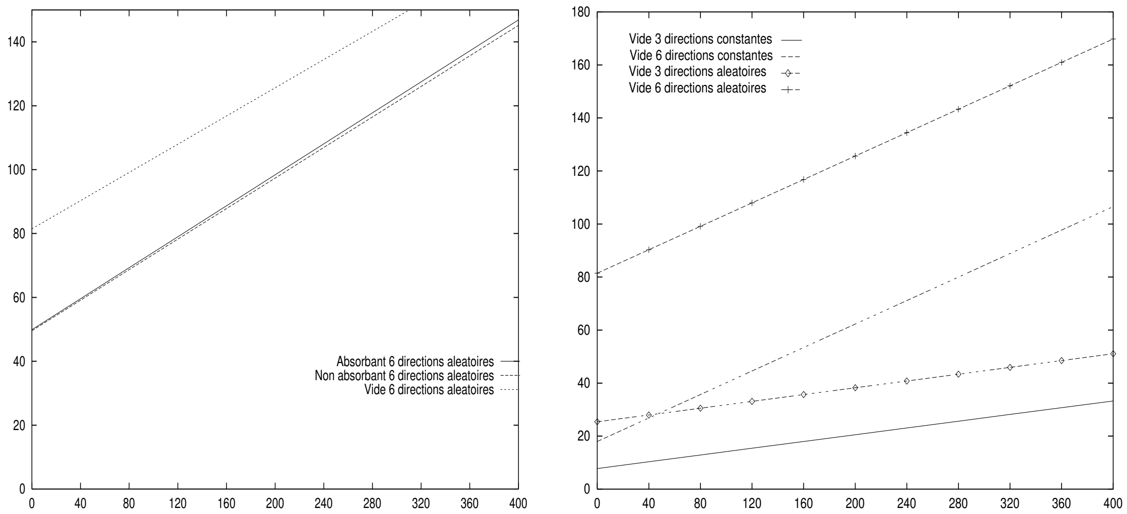
Itérations à partir d'une solution proche. Droite de coût selon le nombre de directions.

Fig. II.10.17 – Cône, étude du rapport coût / précision en fonction du nombre d'itérations.



Itérations à partir d'une solution éloignée. Balayage fréquentiel avec peu d'itérations.

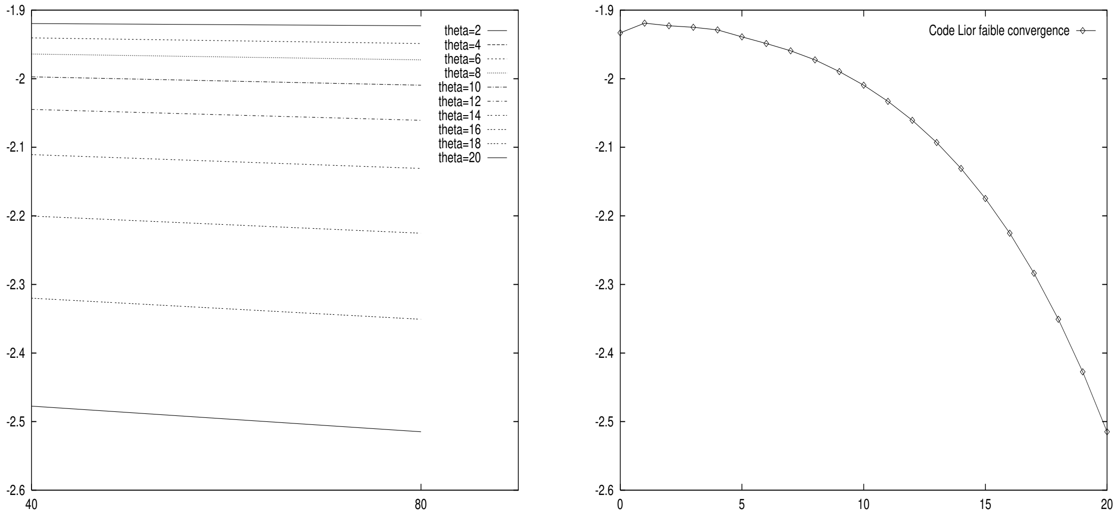
Fig. II.10.18 – Cône, étude du coût d'un balayage fréquentiel pour peu d'itérations effectuées.



Fonction du matériau, vide ou non. Type de directions, aléatoires ou constantes.

Fig. II.10.19 – Cône, droites de coût selon le type de fonctions et les caractéristiques du milieu.

5. La figure II.10.19 de gauche donne les droites de coût du code $\mathcal{L}ior$ du point II.10.1.5.4 en fonction des caractéristiques des couches recouvrant le cône : vide (comme section II.10.1.5.1), matériau absorbant (comme section II.10.1.5.3) ou non (comme section II.10.1.5.2 où les permittivité et perméabilité sont réelles). Les 6 directions de propagation sont aléatoires d'un élément à l'autre.
6. La figure II.10.19 de droite donne les droites de coût du code $\mathcal{L}ior$ du point II.10.1.5.4 en fonction du choix de directions de propagation constantes ou aléatoires entre les éléments. La simulation est effectuée sur le cône revêtu de vide comme dans la section II.10.1.5.1.



Itérations à partir d'une solution éloignée. Balayage angulaire avec peu d'itérations.

Fig. II.10.20 – Cône, SER d'un balayage angulaire pour peu d'itérations effectuées.

7. La figure II.10.20 de gauche étudie le nombre d'itérations du balayage angulaire pour un critère faible de convergence de l'algorithme (point II.10.1.5.4) avec les mêmes données que celles de la figure II.10.17 de gauche.
8. La figure II.10.20 de droite nous donne une idée de la précision du code $\mathcal{L}ior$ pour le critère faible de convergence du balayage angulaire. En effet, nous avons observé lors du balayage fréquentiel (figure II.10.18 de droite) que les difficultés de convergence se traduisaient par des discontinuités dans la courbe de SER entre les fréquences, particulièrement pour celles où le nombre d'itérations effectuées était trop faible. De plus, nous avons constaté sur la figure II.10.18 de gauche que l'évolution de la SER bistatique en fonction du nombre d'itérations n'était pas très stable. Nous observons au contraire dans les courbes II.10.20 une évolution régulière de la SER bistatique du balayage angulaire en angle θ d'incidence (dans le plan (x, z)).

Nous avons cherché à compléter l'étude de coût du point II.10.1.5.4 par une étude par tâche effectuée par le code $\mathcal{L}ior$ et non plus globalement pour tout le code. Cette étude nous permettra de mieux comparer les coûts de chaque tâche du code, et non plus seulement la coût de l'algorithme itératif. Ceci nous permettra de prévoir les temps de calcul selon le balayage effectué et les caractéristiques de la discrétisation pour une géométrie donnée. Nous présentons les types de tâches effectuées dans le tableau II.10.8 qui affecte un numéro à chacune des tâches essentielles du programme. Les coûts de ces tâches sont consignés dans les figures II.10.21. La figure II.10.21 de gauche évalue les coûts des différentes tâches en fonction du nombre de fonctions de base, la figure II.10.21 de droite compare, pour 6 directions de propagation, les coûts à la fois en fonction du type de matériau (vide ou non) et en fonction du choix aléatoire ou constant des directions entre les éléments.

Remarque 48 Dans le tableau II.10.8 la tâche numéro 5 consiste essentiellement à assembler le second membre sur les portions de faces non traitées, la tâche 4 ayant déjà calculé le second membre pour les trois premiers sommets des faces quadrangulaires. La tâche 4 calcule aussi le produit $D^{-1}b$ et les caractéristiques de l'onde incidente. Nous constatons que le temps de calcul du second membre est négligeable par rapport au calcul du simple produit matrice vecteur $D^{-1}b$.

Tab. II.10.8 – Cône, étude du coût par tâche.

Numéro	Tâche
1	Place mémoire en Méga-Words.
2	Lecture des données, initialisation, calcul des normales, divers.
3	Construction des polarisations et directions de propagation.
4	Temps hors itérations de 100 balayages angulaires pour des tétraèdres.
5	Idem, coût supplémentaire au coût de la tâche 4 pour des hexaèdres (pour 100 balayages angulaires).
6	Temps de 2 assemblages des matrices du système linéaire pour des tétraèdres.
7	Coût supplémentaire à celui de la tâche 6 pour des hexaèdres (pour 2 assemblages des matrices).
8	Coût supplémentaire à ceux des tâches 6 et 7 pour effectuer 2 balayages en fréquence.
9	Temps du post-traitement pour une fréquence et une incidence
10	Coût supplémentaire à celui de la tâche 9 pour des hexaèdres.
11	Coût de 100 itérations de l'algorithme itératif.
12	Coût total d'un calcul à une fréquence et une incidence sans itération et sans post-traitement
13	Idem avec 100 itérations de l'algorithme itératif.

Nous constatons les points suivants.

1. Imposer un critère de convergence élevé à l'algorithme itératif pour des balayages en fréquence ne permet pas de réduire fortement les temps de calcul du code $\mathcal{L}ior$.
2. Imposer un critère de convergence faible à l'algorithme itératif pour des balayages en fréquence dégrade la précision du code $\mathcal{L}ior$.
3. Imposer un critère de convergence faible à l'algorithme itératif pour des balayages angulaires ne dégrade pas la précision du code $\mathcal{L}ior$ et réduit considérablement le temps de calcul.
4. L'augmentation du nombre de directions de propagation est néfaste au temps de calcul qui est proportionnel au carré du nombre de fonctions de base dans la plupart des étapes du code $\mathcal{L}ior$ en dehors des étapes de post-traitement de calcul de la SER ou des valeurs aux nœuds.
5. L'utilisation de matériaux réels ou complexes est indifférent sur le temps de calcul, en revanche un calcul sur un élément dans un matériau implique un coût double par rapport à un calcul dans le vide.
6. Le choix de directions qui diffèrent d'un élément à l'autre, plus coûteux pour l'assemblage de la matrice du système linéaire, ne change rien à l'évolution du coût en fonction du nombre d'itérations.

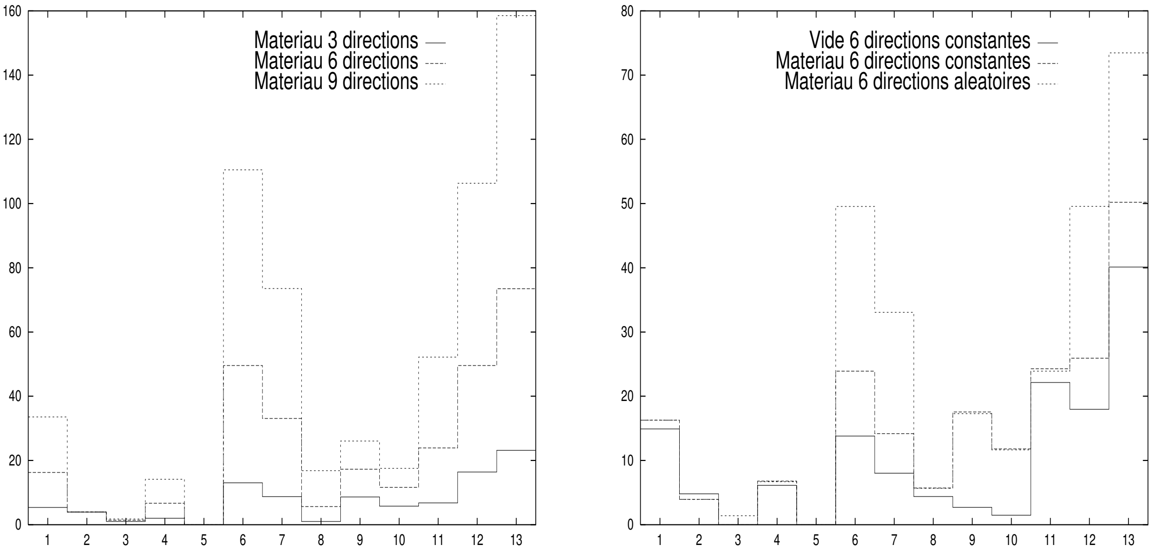
II.10.2 Utilisation optimale.

Deux simulations montrent que l'on peut

1. utiliser un maillage très grossier (section II.10.2.1),
2. avoir un coût en place mémoire et en temps de calcul équivalent à celui d'une méthode d'éléments finis pour un maillage plus grossier (section II.10.2.2).

II.10.2.1 Un cas hors de portée des méthodes classiques.

On s'intéresse au cas modèle du cube II.10.1 a) dans le vide avec $L = 60$ centimètres. On utilise un maillage structuré en 8 hexaèdres à la fréquence de 1000 MHz. Nous effectuons deux calculs, pour deux incidences différentes : $\theta = 19,2$ et $\phi = 11,4$, puis $\theta = 39,2$ et $\phi = 31,4$. Le tableau II.10.9) indique les caractéristiques du cas traité pour la première incidence. Nous choisissons,



En fonction du nombre de directions aléatoires sur le cône avec matériau.

Selon le type de fonctions, directions aléatoires ou constantes et pour le cône revêtu ou non.

FIG. II.10.21 – Cône, étude du coût par tâche.

- soit, des directions constantes sur les 8 éléments,
- soit, de prendre des fonctions de base différentes sur les 8 éléments hexaédriques de façon à ne pas privilégier une direction particulière pour l'ensemble du problème discret,

Dans le deuxième cas, on utilise toujours une loi de répartition aléatoire uniforme des directions de propagation à partir d'une situation de référence qui dépend du nombre de fonctions choisies (cf section II.8.3).

TAB. II.10.9 – Simulation avec des hexaèdres réguliers d'arête la longueur d'onde.

$\mathbf{k}_0$	$(-0,322\,;\,-0.065\,;\,-0.944)$	polarisation	TM
$\mathbf{E}_0$	$(-0.925\,;\,-0.187\,;\,0.329)$	(θ,ϕ)	$(19,2;11,4)$
p	de 3 à 70 (aléatoires)	K	8
f	1 GHz	λ	$0,3$ m
h	$0,3$ m	λ/h	1

Le maillage est extrêmement grossier puisque les arêtes des cubes du maillage ont comme longueur la longueur d'onde du problème. Un tel problème n'est pas abordable pour une méthode classique d'éléments finis. Rappelons, que dans la méthode des éléments finis, on calcule h comme le plus grand des diamètres des éléments. Comme le diamètre d'un cube est $\sqrt{3}$ fois son arête, la méthode des éléments finis, sur ce calcul, indiquerait $\lambda = \sqrt{3}\,.\,h$, et non $\lambda = h$. Nous n'avons pas inclus d'objet diffractant de façon à connaître la solution analytique exacte du problème, mais les résultats seraient semblables dans le cas par exemple d'un cube diffractant. La figure II.10.22 donne, pour les deux incidences et les deux choix des directions, l'évolution, en fonction du nombre de directions de propagation par élément, de la norme $|\mathcal{X} - \mathcal{X}_h|_{L^2(\Gamma)}$ entre la solution approchée calculée par $\mathcal{L}ior$ et la solution exacte $\mathcal{X} = \mathbf{E} \wedge \nu + (\mathbf{H} \wedge \nu) \wedge \nu$ où $\mathbf{E}$ et $\mathbf{H}$ sont donnés par (II.10.3). L'annexe III.B.3 explique ce calcul.

Dans le cas d'une diffraction sur un objet, la SER est liée à la norme $|\mathcal{X}-\mathcal{X}_h|_{L^2(\Gamma)}$ (cf section II.10.1.3). Par exemple, une erreur de 10 % sur la norme de l'erreur équivaut à une erreur de 1 $dB.m^2$ sur la SER par $20 * \log_{10}(1 - 0,1) \approx 1$. Nous constatons que

1. Le code $\mathcal{L}ior$ converge toujours, quel que soit le nombre de directions.

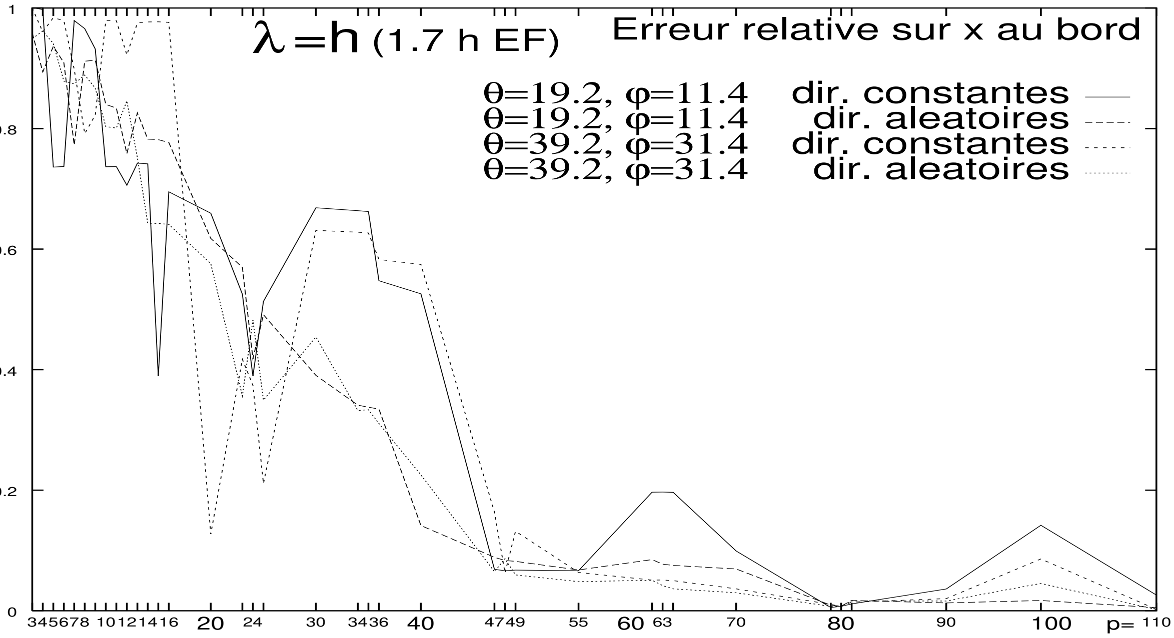


FIG. II.10.22 – Evolution de l'erreur de bord pour des grosses mailles

2. Le nombre d'itérations nécessaires pour résoudre le système linéaire est très faible, la précision du calcul étant de toute façon assez faible.
3. Il faut au moins 45 directions par élément pour obtenir une erreur inférieure à 10 %, au moins 35 pour une erreur inférieure à 30 % (qui donnerait approximativement 3 $dB.m^2$ d'écart par rapport à la solution exacte.
4. Le choix des directions est important pour obtenir une précision suffisante : ceci se voit à l'allure non monotone de la courbe II.10.22. Les résultats sont globalement meilleurs pour des directions de propagation plus équiréparties (cf section II.8.3). Ainsi, le calcul avec 12 directions est meilleur qu'avec 14 moins bien réparties sur la sphère unité. On observe surtout que le choix des directions aléatoires d'un élément à un autre, donne, en général, une erreur plus faible que le choix de directions toutes identiques. Ceci peut s'expliquer par l'anisotropie plus forte induite par le choix de directions constantes sur tous les éléments du maillage, alors que le choix de directions variables corrige globalement ce phénomène.
5. Pour p directions par élément, le stockage est proportionnel à $p(p+1)+12p^2$ et au nombre d'éléments du maillage (pour des hexaèdres). Pour 70 directions cela donne 63770 (unités de stockage) alors que pour 7 on a 644 soit 99 fois moins. Par comparaison, une méthode d'éléments finis P_1 pour un maillage en $\lambda/20$, utilise $(20\sqrt{3})^3$ fois plus d'éléments, soit 42000 (et donc 333 000 éléments au total). Dans une méthode d'éléments finis d'arête le stockage est proportionnel à $\dfrac{N(N+1)}{2}$ où N est le nombre d'arêtes de l'élément. Sur des hexaèdres, le stockage sera donc proportionnel à 78 fois le nombre d'éléments. Dans une telle situation, le stockage pour la méthode éléments finis serait de 26 millions de termes dans la matrice du système linéaire, soit un stockage 50 fois supérieur à celui atteint par $\mathcal{L}ior$ pour 70 fonctions de base par élément. Dans le cas de l'utilisation de 35 directions de propagation pour $\mathcal{L}ior$ et d'un maillage en $\lambda/10$ pour une méthode éléments finis, le stockage demandé par $\mathcal{L}ior$ sera 25 fois inférieur à celui demandé par la méthode des éléments finis.

II.10.2.2 Intérêt par rapport à une méthode d'éléments finis.

Nous effectuons une simulation sur un cube parfaitement conducteur placé dans le vide comme dans la figure II.10.1 *b*). L'arête du cube diffractant est de longueur $a = 50$ cm. Le domaine de maillage est centré sur le centre de ce cube et contenu dans un cube d'arête de longueur $b = 70$ cm. Nous effectuons une simulation à l'aide de trois maillages en tétraèdres dont le nombre d'éléments est $K = N = 10\,464$, environ $8N$ ($K = 83\,712$) et environ $N/8$ ($K = 1\,308$). Les caractéristiques précises du cas étudié sont données dans le tableau II.10.4. Le code $\mathcal{L}ior$ est utilisé avec des fonctions de base constantes d'un élément à un autre, et opère en champ total.

Tab. II.10.10 – Une comparaison avec les éléments finis.

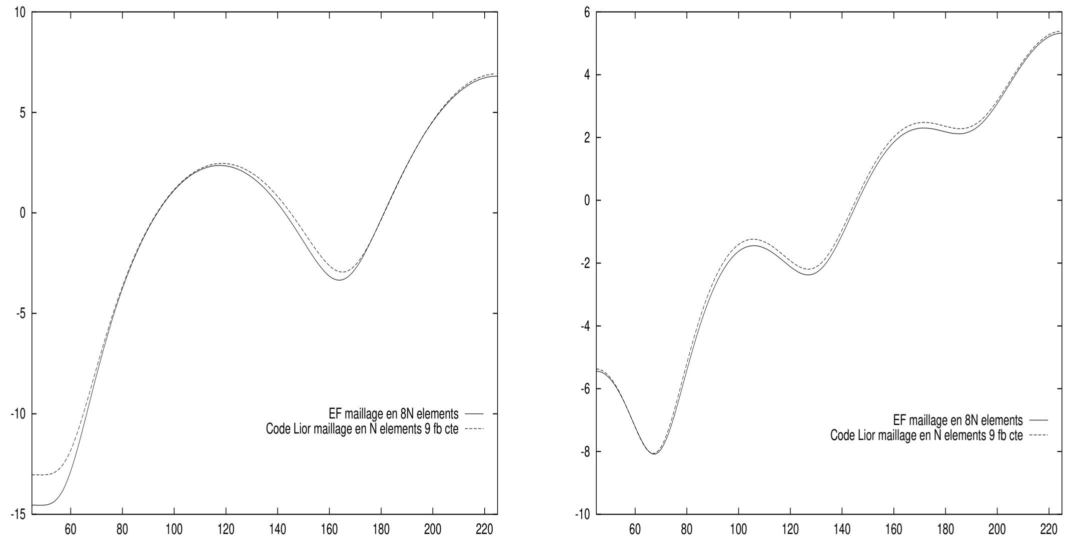
(θ, ϕ)	$(45, 0)$	$\mathbf{k}_0$	$-(1/\sqrt{2}, 0, 1/\sqrt{2})$
$\mathbf{E}_0$	$(0, -1, 0)$	p	3, 6, 9 ou 12
f (MHz)	400	K	1 308, 10 464 ou 83 712
λ	0.75 m	λ/h	13, 6 ; 27, 3 ou 54, 5

Nous comparons le code $\mathcal{L}ior$ à un code éléments finis standard écrit par Bruno Stupfel. Ce code est utilisé pour la même condition aux limites absorbante. Notons que la technique de résolution du système linéaire pour la méthode d'éléments finis utilisée est un algorithme itératif de gradient conjugué par produit scalaire non Hermitien. Nous allons comparer simultanément

– les temps de calculs,
– la place mémoire,
– la précision du calcul.

Notons que les temps de calcul sont dans les deux cas assujettis au nombre d'itérations des algorithmes itératifs respectifs des deux méthodes de résolution du système linéaire. Les temps de calcul ne doivent être pris qu'à titre purement qualitatif.

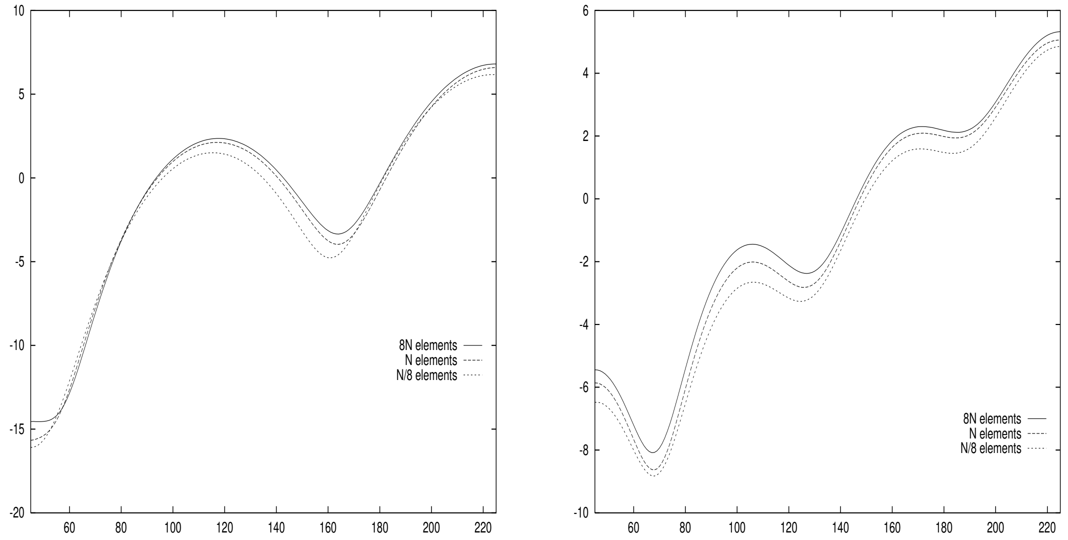
Notons que l'on ne connaît pas la solution exacte de ce problème de "scattering" (II.10.1). La précision obtenue par les simulations numériques est donc seulement estimée par rapport à des courbes de référence qui sont des approximations de la solution exacte. Ces courbes sont obtenues pour des discrétisations élevées du problème. En pratique cela correspond à une courbe "limite" lorsque la discrétisation augmente, mais évidemment la limite n'est pas atteinte. Nous représentons, figure II.10.23, les valeurs de la SER en balayage bistatique de 45 à 225 degrés pour ces courbes de référence, pour $\mathcal{L}ior$ comme pour le code Eléments Finis dans les deux polarisations.



mode TE   mode TM.

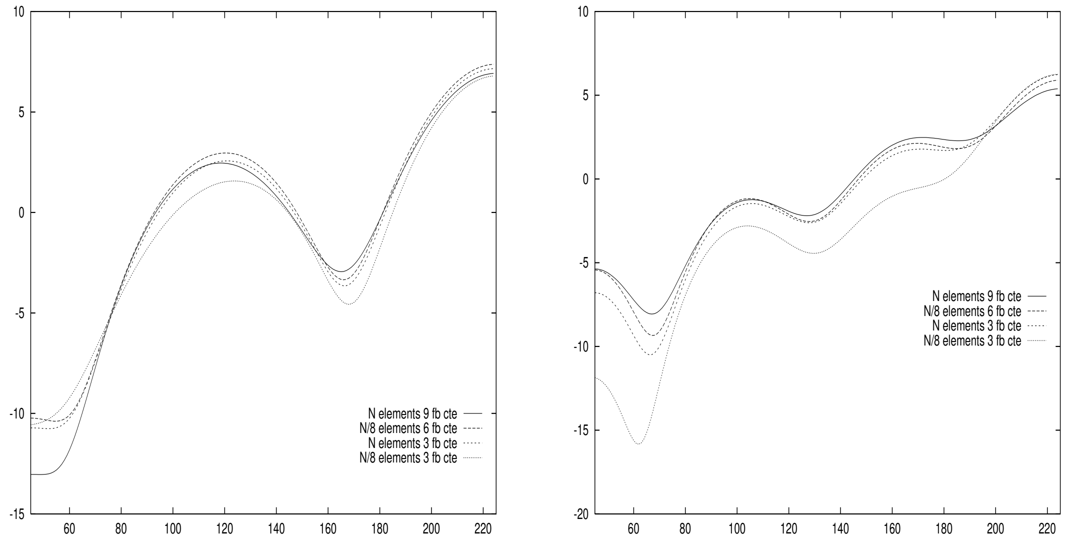
Fig. II.10.23 – SER de "référence" comparée à un code Eléments Finis.

La figure II.10.24 étudie la convergence de la méthode des éléments finis par rapport au paramètre de raffinement du maillage, les figures II.10.25 et II.10.26 la convergence du code $\mathcal{L}ior$ par rapport au nombre de fonctions de base et au paramètre de raffinement du maillage. La figure II.10.25 représente les



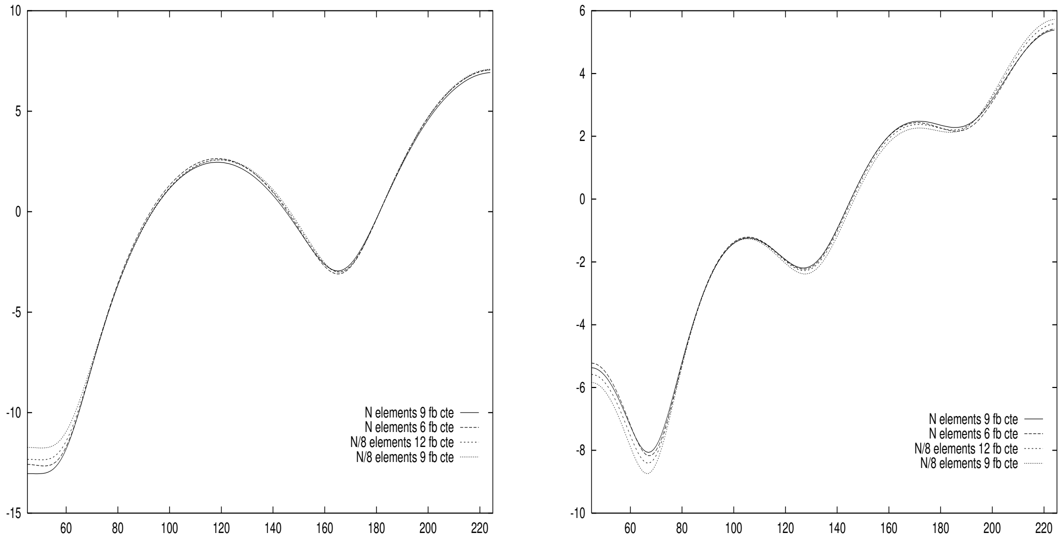
mode TE mode TM.

FIG. II.10.24 – Etude de convergence des Eléments Finis P1, maillage variable.



mode TE mode TM.

FIG. II.10.25 – Etude de convergence du code $\mathcal{L}ior$: discrétisation insuffisante.



mode TE — mode TM.

FIG. II.10.26 – Etude de convergence du code $\mathcal{L}ior$: discrétisation “suffisante”.

cas pour lesquels nous considérons que la discrétisation est trop grossière (par comparaison à la courbe de référence pour N éléments et 9 directions), la figure II.10.26 les cas où le code $\mathcal{L}ior$ donne des résultats corrects et comparables à ceux de la courbe de référence II.10.23.

Nous avons testé les deux codes par rapport à la discrétisation. Le code Eléments Finis utilise des éléments P_1, nous n'avons pas fait varier le degré mais uniquement le nombre d'éléments du maillage. Pour $\mathcal{L}ior$ nous avons fait varier le nombre de fonctions de base par élément. Nous n'avons pas utilisé le maillage le plus raffiné car notre but est de montrer que l'on peut grossièrement atteindre la même précision de calcul pour une place mémoire et un temps de calcul du même ordre mais pour moins d'éléments que par la méthode d'éléments finis. Nous laissons le lecteur libre d'accepter ce résultat au vu des courbes II.10.24 et II.10.26. Nous constatons en revanche que notre méthode, même si elle calcule toujours une solution, est moins précise pour une discrétisation trop lâche (figure II.10.25).

Les places mémoires sont

1. Pour les éléments finis d'arête P_1 avec des tétraèdres, la taille du système linéaire est en $21 \times N$. En effet, la matrice est composée de blocs hermitiens dont la taille est le nombre d'arêtes dans un tétraèdre, soit 6 (on a $21 = \frac{6*(6+1)}{2}$). La matrice est composée de termes réels ou complexes, mais pour plus de commodité, elle est stockée uniformément en complexes.
2. Pour $\mathcal{L}ior$ la taille mémoire pour des tétraèdres dans le vide est proportionnelle à $2 \times \frac{p(p+1)}{2} + 4 \times p^2$ (elle double dans un matériau). Tous les termes sont complexes sauf les $2p$ coefficients de la diagonale de la matrice hermitienne.

TAB. II.10.11 – Evolutions comparées des temps CPU et tailles mémoires.

Discrétisation, méthode	Temps CPU (s)	Taille mémoire relative
EF maillage $8N$	400	168
EF maillage N	35	21
$\mathcal{L}ior$ maillage $N/8$, 12 fb	$23,7$	163
$\mathcal{L}ior$ maillage $N/8$, 9 fb	$13,1$	92
$\mathcal{L}ior$ maillage $N/8$, 6 fb	$8,5$	41
$\mathcal{L}ior$ maillage N, 9 fb	105	738
$\mathcal{L}ior$ maillage N, 6 fb	60	330

Nous résumons les temps de calcul et les tailles mémoires des cas les plus intéressants des deux méthodes dans le tableau II.10.11. On voit qu'il est intéressant pour le code $\mathcal{L}ior$ de prendre un nombre élevé de fonctions de bases par élément et un maillage réduit. On réussit alors dans le cas de 12 fonctions de base par élément avec le maillage en $N/8$ éléments à obtenir une précision équivalente à celle du maillage en $8N$ éléments pour la méthode d'éléments finis.

II.10.3 Maillages tétraédriques ou hexaédriques.

Nous comparons les différences pratiques d'utilisation d'un maillage en tétraèdres et en hexaèdres. En effet, un maillage hexaédrique est, pour un nombre d'éléments fixé, plus coûteux en place mémoire et en temps de calcul, que ce soit pour l'assemblage des termes du système linéaire ou lors de l'algorithme itératif de résolution du système, qu'un maillage tétraédrique. Nous regardons si la précision atteinte est la même, pour un nombre d'éléments fixé et si le temps de calcul est le même.

Pour cela, nous calculons la norme $|\mathcal{X} - \mathcal{X}_h|_{L^2(\Gamma)}$ (entre la solution approchée calculée par $\mathcal{L}ior$ et la solution exacte $\mathcal{X} = \mathbf{E} \wedge \nu + (\mathbf{H} \wedge \nu) \wedge \nu$ où $\mathbf{E}$ et $\mathbf{H}$ sont donnés par (II.10.3)) sur le cas modèle du cube II.10.1 a) avec $L = 60$ centimètres dans le vide. La solution de ce problème est une onde plane dont les caractéristiques sont données par le tableau II.10.12.

TAB. II.10.12 – Maillages tétraédriques ou hexaédriques, caractéristiques de l'onde plane solution.

(θ, ϕ)	$(19,2; 11,4)$	$\mathbf{k}_0$	$(-0,322\,; -0.065\,; -0.944)$
polarisation	TM	$\mathbf{E}_0$	$(-0.925\,; -0.187\,; 0.329)$

Nous utiliserons trois maillages dont les caractéristiques sont données par le tableau II.10.13 où F est le nombre de faces sur la frontière du domaine.

TAB. II.10.13 – Maillages tétraédriques ou hexaédriques, caractéristiques des trois maillages.

Caractéristiques	Maillage M_1	Maillage M_2	Maillage M_3
Type du maillage	tétraèdre	hexaèdre	tétraèdre
Nombre K d'éléments	465	512	550
Nombre F de faces	192	384	228
Rapport F/K en %	$41,2$	$75,0$	$41,45$
Raffinement h en cm	$7,74$	$7,50$	$7,32$

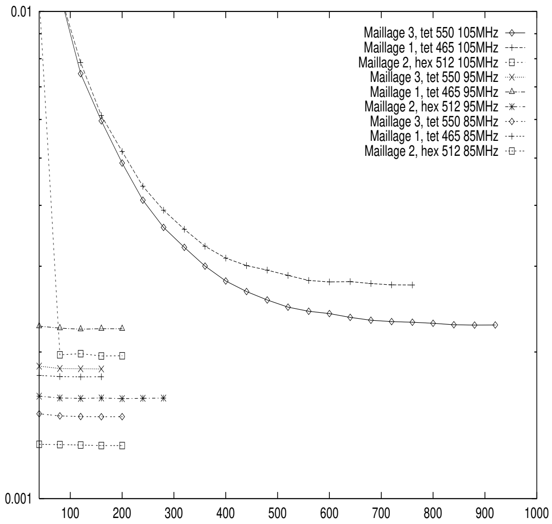
II.10.3.1 Etude d'un cas limite sur-discrétisé.

Nous avons effectué des simulations pour trois très basses fréquences par rapport au paramètre h de raffinement du maillage. Les fréquences de calculs sont 105, 95 ou 85 MHz, les longueurs d'onde sont donc respectivement de 2.857, 3.158 et 3.529 mètres.

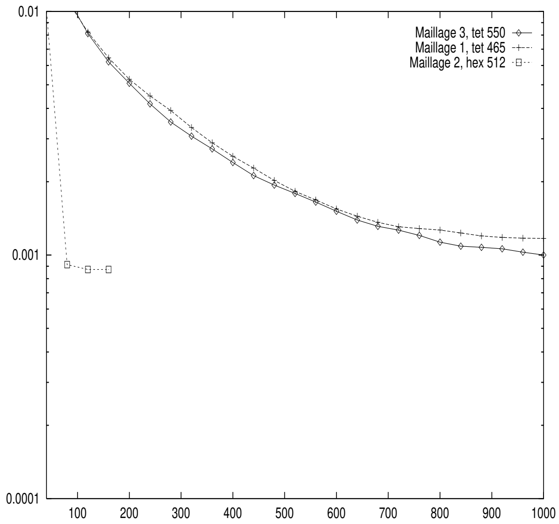
La figure II.10.27 représente l'évolution de la précision de la norme L^2 relative $|\mathcal{X} - \mathcal{X}_h|_{L^2(\Gamma)}$ en fonction du nombre d'itérations effectuées pour les trois maillages. Notons que la figure II.10.27 a) avec 11 directions de propagation aléatoires entre les éléments est issue d'un calcul effectué en une seule fois, de 105 MHz à 85 MHz par pas de -10 MHz, ce qui explique que le nombre d'itérations effectuées est plus grand pour la première fréquence. La figure II.10.27 b) ne concerne que la plus basse fréquence de 85 MHz et est obtenue pour 13 directions de propagation aléatoires entre les éléments.

Les rapports entre la longueur d'onde et le raffinement du maillage sont données dans les tableaux II.10.14 et II.10.15 pour respectivement 105 et 85 MHz pour les trois maillages.

Nous donnons quelques informations pour la simulation avec 11 directions de propagation par élément (choisies aléatoirement entre les éléments) sur le conditionnement maximal des sous matrices D_k (cf



a) Balayage fréquentiel pour 11 directions.



b) Calcul à 85 MHz avec 13 directions.

FIG. II.10.27 – Nombre d'itérations et précision selon les types de maillages, cas limite sur-discrétisé.

TAB. II.10.14 – Maillages tétraédriques ou hexaédriques, simulation à 105 MHz.

Quantité	Maillage M_1	Maillage M_2	Maillage M_3
λ/h	37	38	39
Conditionnement	$1.1E11$	$3.2E10$	$8.2E10$
$\|D^{-1}D - I\|_{L_\infty}$	$1.4E-5$	$1.7E-5$	$4.2E-5$

section II.8.1.2) de D et la qualité de l'inversion de la matrice globale D (chapitre II.8, problème (II.8.2)), comme nous l'avions déjà indiqué dans la section II.10.1.2.2, en norme L_∞ sur l'erreur maximale des coefficients de $D^{-1}D$ par rapport à la matrice identité I.

TAB. II.10.15 – Maillages tétraédriques ou hexaédriques, simulation à 85 MHz.

Quantité	Maillage M_1	Maillage M_2	Maillage M_3
λ/h	46	47	48
Conditionnement	$3,2E10$	$9,0E9$	$2,3E10$
$\lvert D^{-1}D - I\rvert_{L_\infty}$	$3,6E-6$	$5,8E-6$	$8,6E-6$

Nous constatons sur ce cas précis les points suivants.

1. Le nombre d'itérations nécessaires pour les trois fréquences 105, 95 et 85 MHz est inférieur pour le maillage hexaédrique que pour les deux autres maillages tétraédriques. Le coût supplémentaire en temps de calcul induit par le maillage en hexaèdre est donc compensé par la plus grande rapidité de résolution du système linéaire.
2. La précision finale est meilleure avec le maillage en hexaèdre. Le coût supplémentaire en place mémoire induit par le maillage en hexaèdre peut donc être compensé par l'utilisation d'un maillage moins fin, donc moins compliqué à construire.
3. Le maillage en hexaèdre est plus robuste par rapport au conditionnement des matrices hermitiennes de produit scalaire D_k, trop élevé comme on le constate dans le tableau II.10.14. Cela permet d'obtenir des résultats plus précis en sur-discrétisant.

II.10.3.2 Etude exhaustive sur un cas de propagation dans le vide.

Il est faux de croire que le nombre d'itérations nécessaires pour inverser le système linéaire dans le code $\mathcal{L}ior$ est systématiquement inférieur pour un maillage hexaédrique que pour un maillage tétraédrique (à nombre égal d'éléments).

Après un grand nombre de simulations (environ 80) nous avons observé que ceci est vrai lorsque la fréquence étudiée est trop basse par rapport à la discrétisation du problème, ce qui correspond à une sur-discrétisation. Cette situation fait apparaître des problèmes numériques de conditionnement comme pour toute méthode numérique de discrétisation qui tend à se rapprocher de l'opérateur continu.

Dans le cas général le choix des hexaèdres n'apporte rien par rapport au choix des tétraèdres comme le montrent les quelques exemples de la figure II.10.28 p. 167 où l'on représente l'évolution de la précision de la norme L^2 relative $|\mathcal{X} - \mathcal{X}_h|_{L^2(\Gamma)}$ en fonction du nombre d'itérations. Ces calculs sont effectués à différentes fréquences que nous n'indiquons pas ; ces courbes ne sont là qu'à titre indicatif.

En particulier, dans les situations réalistes où le problème n'est pas sur-discrétisé, nous constatons qu'à nombre égal d'éléments, les maillages en hexaèdres, par rapport aux maillages en tétraèdres,

1. occupent une place mémoire presque 50 % supérieure,
2. demandent autant d'itérations pour la résolution itérative du système linéaire,
3. prennent environ $2,5$ fois plus de temps pour l'assemblage des termes du système linéaire quand leurs faces sont planes, 3 fois quand ce n'est pas le cas,
4. demandent un temps total de calcul majoré d'au moins 50 %.

II.10.4 Conclusion de l'étude du programme $\mathcal{L}ior$.

L'étude quasi extensive du code $\mathcal{L}ior$ menée dans cette section nous semble permettre d'affirmer les conclusions suivantes.

1. Le nombre de simulations effectuées, sur des cas très variés, et le nombre de comparaisons avec d'autres codes nous semblent suffisants pour affirmer que le code $\mathcal{L}ior$ est fiable.

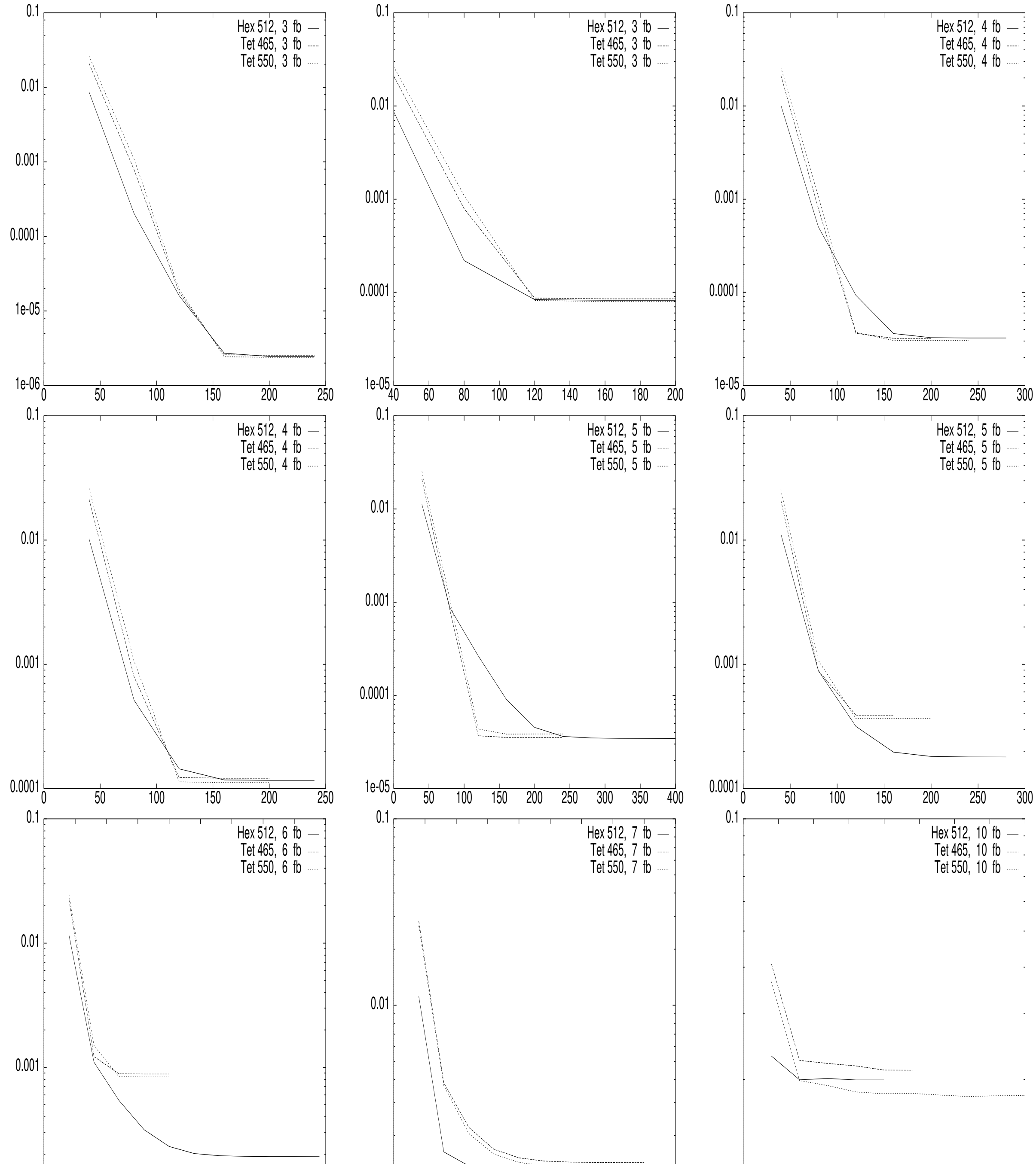


Fig. II.10.28 – Comparaison du nombre d'itérations et de la précision selon le type de maillage, les directions sont aléatoires, fréquences variables.

2. Le code $\mathcal{L}ior$ tire parti de toutes les situations particulières du problème et de la discrétisation, notamment
   (a) dans le vide, la place mémoire est minimisée. Dans le cas du cône revêtu, le gain en place mémoire par rapport au cas du cône parfaitement conducteur est d'environ 50% sur tous les éléments de la couche.
   (b) Du fait précédent, tous les calculs effectués sur un élément dans le vide sont réduits. Notamment, le nombre d'opérations effectuées par l'algorithme itératif est réduit de moitié.
   (c) Sur un élément dans le vide strictement intérieur à Ω et dont tous les voisins sont dans le vide, et lorsque les fonctions de base sont toutes égales, le temps d'assemblage des matrices est minimisé. On utilise la relation (III.B.46) qui permet d'optimiser le calcul de D à partir du calcul de C.
   (d) Dans une situation identique à la précédente mais sous la condition moins restrictive que la face $\Sigma_{k,j}$ est dans le vide et interne, on utilise la relation (III.B.47). Ceci permet d'optimiser le calcul de la matrice C.
   (e) La programmation de $\mathcal{L}ior$ pour des hexaèdres est effectuée en rajoutant les termes intégraux sur le deuxième triangle des faces quadrangulaires. Cette technique de programmation a permis d'étendre le code au cas d'hexaèdres par un programme d'intelligence artificielle (triviale) de génération automatique de programme. Le but de l'extension de $\mathcal{L}ior$ était d'effectuer des simulations numériques visant à comparer les avantages respectifs des deux types d'éléments de maillage, tétraèdres ou hexaèdres. Ce but a pu être atteint, mais l'algorithme de programmation ne permet pas d'étudier des maillages en hexaèdres à faces non planes.
3. Le code $\mathcal{L}ior$ nous semble parfaitement optimisé sur machine à “architecture vectorielle”. Nous vérifierons cela dans l'annexe III.C qui présente les performances de $\mathcal{L}ior$ en terme de “Méga-Flops”.

Le cahier des charges est rempli : code Maxwell 3-D. En plus, le code $\mathcal{L}ior$

1. travaille indifféremment dans le vide ou en milieu absorbant isotrope,
2. effectue les post-traitements de calcul des valeurs du courant total et de calcul de la SER,
3. présente des caractéristiques d'un code utilisable en “boîte noire”. Cela se traduit par
   (a) des contrôles de fiabilité effectués par le code lui-même automatiquement : vérification du conditionnement de la matrice D, vérification de l'inversion de la matrice, vérification de la résolution finale du système linéaire.
   (b) Vérification de la cohérence des données spécifiées par l'utilisateur, en l'absence de données caractéristiques de la discrétisation par la formulation variationnelle (nombre de directions, type de ces directions (aléatoires ou constantes), critère d'arrêt de l'algorithme itératif), mise en place de valeurs par défaut qui dépendent des caractéristiques du problème.

Enfin, il nous semble que l'intérêt de $\mathcal{L}ior$ par rapport à une méthode d'éléments finis est de

1. pouvoir travailler sur des maillages grossiers où la taille mémoire et le temps de calcul du code sont inférieurs pour une précision égale. Dans les cas où l'utilisation d'un maillage grossier est rendue difficile par les données géométriques du problème (objet à surface très irrégulière, couche de matériaux à caractéristiques très variables), les performances relatives de $\mathcal{L}ior$ baissent par rapport aux méthodes classiques.
2. Dans le cas où l'utilisation d'un maillage grossier est possible, le code $\mathcal{L}ior$ permet d'étudier des problèmes à des fréquences hors de porté par une méthode d'éléments finis, au moins à cause de la réalisation du maillage. Il est aujourd'hui difficile de mailler un objet avec plus d'un million de tétraèdres.
3. Dans le cas général, $\mathcal{L}ior$ permet d'effectuer des balayages en fréquence à moindre coût ingénieur (utilisation d'un seul maillage) et d'optimiser les temps de calcul puisqu'il suffit de modifier le nombre p de directions par élément pour monter en fréquence. Cette possibilité offre en outre l'avantage de vérifier un calcul par un autre pour plus de fonctions de base.

Enfin, une amélioration sensible du code serait de le coupler avec une méthode intégrale ou avec le code $\mathcal{M}axiim$ de façon à s'affranchir des problèmes de conditions aux limites absorbantes approchées.

Synthèse de l'étude du problème de Maxwell.

Notre étude du problème de Maxwell harmonique tridimensionnel nous amène aux conclusions suivantes.

1. Nous rencontrons les mêmes avantages théoriques et pratiques pour la résolution des équations de Maxwell que pour celles de Helmholtz : mise en œuvre pratique, ordre de convergence élevé, inversibilité inconditionnelle du système linéaire (section I.5.3).
2. La formulation ultra-faible reste valable dans un milieu absorbant.
3. Les simulations numériques nous ont montré que la formulation ultra-faible est une alternative intéressante aux autres méthodes de discrétisation comme les éléments finis, les différences finies ou les volumes finis. Nous pouvons notamment utiliser des maillages grossiers et réduire la place mémoire en conservant une précision suffisante.

Conclusion et perspectives.

Cette étude a montré la viabilité théorique et numérique de notre méthode pour la résolution des problèmes de Helmholtz ou de Maxwell, en deux ou trois dimensions, avec les avantages cités dans les conclusions des deux parties de ce travail, pages 77 et 169. Notons que cette étude pourra être poursuivie par les points suivants concernant la résolution des problèmes d'ondes harmoniques.

Analyse de la méthode.

- Etendre les lois d'ordre d'erreur sur le bord au domaine entier.
- Compléter les tests numériques de résolution à l'aide de grosses mailles, et effectuer une étude théorique. Ceci permettrait, au moins empiriquement, selon les données du problème et la taille des mailles, de déterminer le nombre de fonctions de base à utiliser pour une précision demandée. Cette étude n'a été que partiellement réalisée.
- Etudier les problèmes de dispersion numérique.

Choix de l'espace de discrétisation.

- Initier l'utilisation d'informations sur le comportement asymptotique de la solution ([9]) pour discrétiser le problème avec les fonctions de base adéquates.
- En l'absence d'informations sur le comportement asymptotique de la solution pour le problème de Maxwell, rechercher des algorithmes efficaces de choix des directions de propagation et trouver un critère d'équirépartition pour la méthode.
- Utiliser d'autres fonctions de base que des ondes planes.
- Utiliser un maillage en mailles à faces paraboliques, ou d'ordres plus élevés : ceci permettra de mieux suivre une frontière courbe et donc d'augmenter la taille des éléments.

Résolution du système linéaire.

- Améliorer la vitesse de convergence de l'algorithme de résolution du système linéaire en étudiant le choix optimal des coefficients de relaxation.
- Diminuer le nombre d'itérations effectuées par l'algorithme de résolution du système linéaire en construisant au préalable une solution proche de la solution discrète.
- Optimiser le nombre d'itérations effectuées en fonction de la précision globale du cas traité, précision naturellement limitée par l'espace de discrétisation.
- Comparer aux techniques itératives classiques : GMRes, Bi-CGStab, QMR...

Mise en place d'une méthode "exacte".

- Utiliser des conditions aux limites d'ordre élevé.
- Etudier un couplage avec une méthode intégrale affranchie des problèmes de longueur d'onde : ceci permettrait d'entrer véritablement en compétition avec les méthodes hybrides éléments finis, équations intégrales.

Notons que la formulation proposée est également applicable à une classe très large d'équations aux dérivées partielles (cf [27]), ce qui pourrait aussi faire l'objet d'un travail ultérieur.

Troisième partie

Annexes

Annexe III.A

Mise en œuvre informatique de l'espace V_h choisi pour Helmholtz.

Nous calculons analytiquement les termes des matrices et du second membre de la formulation discrète (I.1.27) du problème modèle de Helmholtz bidimensionnel sans coefficient. L'espace de discrétisation est l'espace V_h construit à l'aide des ondes planes section I.2.2.1. Nous donnons d'abord les notations nécessaires à cette section.

i) Notations relatives à la géométrie (sommets, arêtes, normales, longueurs).

1. Soient (x_1^k, x_2^k, x_3^k) les positions des trois sommets du triangle Ω_k.
2. Soient (x_1^{kj}, x_2^{kj}) les extrémités de l'arête Σ_{kj} commune aux éléments Ω_k et Ω_j.
3. Soient $(\nu_1^k, \nu_2^k, \nu_3^k)$ les trois normales sortantes du triangle Ω_k, respectivement des arêtes (x_1^k, x_2^k), (x_2^k, x_3^k) et (x_3^k, x_1^k).
4. Soit (ν_{kj}) la normale sortante du triangle Ω_k vers le triangle Ω_j. C'est donc aussi la normale à l'arête (x_1^{kj}, x_2^{kj}).
5. Soient (L_1^k, L_2^k, L_3^k) les longueurs des trois arêtes du triangle Ω_k : $|x_1^k - x_2^k|$, $|x_2^k - x_3^k|$, $|x_3^k - x_1^k|$.
6. Soit L_{kj} la longueur de l'arête Σ_{kj} : $|x_1^{kj} - x_2^{kj}|$.

ii) Notations relatives aux termes exponentiels des fonctions de base.

1. Pour $n = 1$ à 3, $h_n^k = \omega(\mathbf{v}_{km} - \mathbf{v}_{kl}) \dfrac{(\vec{x}_{n+1}^k - \vec{x}_n^k)}{2}$ avec $\vec{x}_4^k = \vec{x}_1^k$.
2. Pour $n = 1$ à 3, $Z_n^k = e^{(i\omega(\mathbf{v}_{km} - \mathbf{v}_{kl})\, \vec{x}_n^k)}$.
3. Soit $h_{kj} = \omega(\mathbf{v}_{jm} - \mathbf{v}_{kl}) \dfrac{(\vec{x}_2^{kj} - \vec{x}_1^{kj})}{2}$.
4. Soit $Z_{kj} = e^{(i\omega(\mathbf{v}_{jm} - \mathbf{v}_{kl})\, \vec{x}_1^k)}$.

Les matrices D (I.2.18) et C (I.2.19 et I.2.20) et le second membre b (I.2.21) donnés section I.2.1.4 sont calculés à l'aide des ondes planes définissant l'espace de discrétisation V_h.

i) Le calcul de D (I.2.18) pour des ondes planes fait apparaître un terme constant à multiplier par l'intégrale d'une onde plane sur un segment. A l'aide de l'annexe III.D.1.1 qui donne l'intégrale sur un segment de la fonction $e^{i\mathbf{kX}}$, on calcule facilement que les termes non nuls de D, notés $D_k^{l,m}$, sont explicités par

$$D_k^{l,m} = \omega^2 \sum_{n=1}^{3} L_n^k Z_n^k (1 - \vec{\nu}_n^k\, \mathbf{v}_{km})(1 - \vec{\nu}_n^k\, \mathbf{v}_{kl}) \frac{\sin h_n^k}{h_n^k} e^{ih_n^k} \ . \tag{III.A.1}$$

Le terme $D_k^{l,m}$ est la somme des contributions des arêtes de l'élément Ω_k pour les fonctions de base de directions $\mathbf{v}_{km}$ et $\mathbf{v}_{kl}$.

ii) La matrice C, constituée de termes de couplage entre deux éléments voisins (I.2.19) et de termes diagonaux de couplage au bord du domaine (I.2.20), est calculée de la même façon que D. On obtient donc après calculs (qui utilisent toujours l'intégrale sur un segment de la fonction $e^{i\mathbf{kX}}$ de l'annexe III.D.1.1) les deux formules de calcul des termes non nuls de C (III.A.2) et (III.A.3) ci-dessous.

1. Dans le cas où Ω_k est voisin de Ω_j, la contribution $C_{k,j}^{l,m}$ de l'interface Σ_{kj} pour les fonctions de base de directions $\mathbf{v}_{jm}$ et $\mathbf{v}_{kl}$ est donnée par

$$\text{(III.A.2)} \qquad C_{k,j}^{l,m} = \omega^2 L_{kj} Z_{kj} (1 + \vec{\nu}_{kj}.\mathbf{v}_{jm})(1 + \vec{\nu}_{kj}.\mathbf{v}_{kl}) e^{ih_{kj}} \frac{\sin h_{kj}}{h_{kj}} \ .$$

2. Dans le cas où Γ_k est non vide, la contribution $C_{k,k}^{l,m}$ du bord Γ_k pour les fonctions de base sur Ω_k de directions $\mathbf{v}_{km}$ et $\mathbf{v}_{kl}$ est donnée par

$$\text{(III.A.3)} \qquad C_{k,k}^{l,m} = \sum_{n/[x_n^k, x_{n+1}^k] \in \Gamma_k} Q_k \omega^2 L_n^k Z_n^k (1 - \vec{\nu}_n^k.\mathbf{v}_{km})(1 + \vec{\nu}_n^k.\mathbf{v}_{kl}) e^{ih_n^k} \frac{\sin h_n^k}{h_n^k}$$

où l'on utilise l'hypothèse $Q_k = Q_{|\Gamma_k}$ constant sur Γ_k.

iii) Le second membre, donné par l'équation (I.2.21), est calculé pour des fonctions de base issues d'ondes planes par

$$\text{(III.A.4)} \qquad b_{k,l} = -2i\omega \int_{\Omega_k} f.\overline{e^{(i\omega\mathbf{v}_{kl}.\vec{x})}} + \int_{\Gamma_k} g.\overline{(+\partial_{\nu_k} + i\omega) e^{(i\omega\mathbf{v}_{kl}.\vec{x})}} \ .$$

Nous explicitons le calcul analytique de (III.A.4) dans les cas suivants de valeurs des fonctions f et g.

1. La contribution d'un terme source localisé en un point x_0 de Ω, soit pour $f = \delta_{x_0}$, est donnée par

$$\text{(III.A.5)} \qquad b_{k,l} = \left| \begin{array}{ll} -2i\omega e^{-i\omega(\mathbf{v}_{kl}.\vec{x_0})} & \text{si } x_0 \in \Omega_k \\ 0 & \text{sinon.} \end{array} \right.$$

2. La contribution d'une source surfacique sur un bord $B = \Gamma \cap \partial\Omega_k$ correspondant à l'excitation par une onde plane incidente de direction $\mathbf{v}_0$, modélisée par

$$g_{|\Gamma_k \subset B} = [(1 + Q_k)(\partial_{\nu_k}) + i\omega(1 - Q_k)] e^{i\omega(\mathbf{v}_0.\vec{x})}$$

où l'on suppose $Q_k = Q_{|\Gamma_k}$ constant, est

$$\text{(III.A.6)} \qquad \begin{array}{l} b_{k,l} = \displaystyle\sum_{n/[x_n^k, x_{n+1}^k] \in \Gamma_k} \omega^2 L_n Z_n^k \xi (1 + \vec{\nu}_n \, \mathbf{v}_{kl}) e^{ih_n^k} \frac{\sin h_n^k}{h_n^k} \\ \text{avec } \xi = (1 + Q_k)\vec{\nu}_n \mathbf{v}_0 + 1 - Q_k \ . \end{array}$$

3. On calcule la contribution d'une source volumique $f = \mu\omega^2 e^{i\omega(\mathbf{v}_0.\vec{x})}$. Sur un triangle de surface S et de barycentre $\vec{G}$, le calcul de l'intégrale de l'onde plane $e^{i\mathbf{kX}}$ est effectué annexe III.D.1.2. Alors, pour $\mathbf{k} = \omega(\mathbf{v}_0 - \mathbf{v}_{kl})$, la contribution de la source volumique f est, pour un triangle (Ω_k),

$$\text{(III.A.7)} \qquad \begin{array}{l} b_{k,l} = (-4i\omega^3 \mu S).(e^{i\mathbf{k}\vec{G}} e^{-\frac{2}{3}i(\alpha+\beta)} \dfrac{e^{i\alpha}\dfrac{\sin\alpha}{\alpha} - e^{i\beta}\dfrac{\sin\beta}{\beta}}{2i(\alpha - \beta)}) \\ \text{avec } \left\{ \begin{array}{l} \alpha = \omega(\mathbf{v}_0 - \mathbf{v}_{kl}) \, \dfrac{(\vec{x}_1^k - \vec{x}_2^k)}{2} \\ \beta = \omega(\mathbf{v}_0 - \mathbf{v}_{kl}) \, \dfrac{(\vec{x}_3^k - \vec{x}_2^k)}{2} \ . \end{array} \right. \end{array}$$

La mise en œuvre informatique du calcul des termes explicités ci-dessus n'est pas directe. En effet, les intégrations des ondes planes sur un segment ou sur un triangle donnent des fonctions dont le domaine de définition pose problème.

1. Les termes matriciels D et C ne sont pas définis pour $\alpha = h_n^k$ ou $\alpha = h_{kj}$ nuls. En effet, la fonction $I_1(\alpha)$, intégrale sur un segment $[x_1, x_2]$ de $e^{i\mathbf{kX}}$, vaut

$$I_1(\alpha) = Le^{i\mathbf{k}\vec{G}} \frac{\sin \alpha}{\alpha} ,$$

   où $\vec{G}$ est la position du barycentre du segment et L sa longueur, α est donné par $\alpha = \mathbf{k}\frac{(x_1 - x_2)}{2}$. Cette fonction, qui permet de calculer les termes matriciels, est définie sur $\mathbb{R} - \{0\}$. Elle se prolonge mathématiquement par continuité en 0, mais cette notion est ignorée par les calculateurs. Nous expliquons annexe III.D.2.1 comment calculer la fonction $I_1(\alpha)$ sur $\mathbb{R}$ ou $\mathbb{C}$ sur tout calculateur de façon à obtenir la meilleure précision possible. Le lecteur intéressé par les problèmes d'optimisation sur "super-calculateurs" consultera l'annexe III.D.3.1 pour l'implémentation sur machine à architecture vectorielle.
2. Le calcul du second membre dans le cas d'une source volumique de la forme $f = \mu\omega^2 e^{i\omega(\mathbf{v}_0.\vec{x})}$ n'est pas défini sur les droites $\alpha = 0$, $\beta = 0$ et $\alpha = \beta$ dans l'équation (III.A.7). Comme pour l'intégrale sur un segment $I_1(\alpha)$, nous expliquons, annexe III.D.2.2, comment calculer par ordinateur la fonction intégrale sur un triangle $[x_1, x_2, x_3]$ de surface S de $e^{i\mathbf{kX}}$, soit, pour $\vec{G}$ la position du barycentre du triangle et $\mathbf{k} = \omega(\mathbf{v}_0 - \mathbf{v}_{kl})$,

$$I_2(\alpha, \beta) = 2Se^{i\mathbf{k}\vec{G}} e^{-\frac{2}{3}i(\alpha+\beta)} \left(\frac{e^{i\alpha}\dfrac{\sin \alpha}{\alpha} - e^{i\beta}\dfrac{\sin \beta}{\beta}}{2i(\alpha - \beta)}\right) ,$$

   et comment optimiser le calcul sur machine vectorielle annexe III.D.3.2.

Le lecteur trouvera comment nous avons optimisé le problème suivant :
- garder une excellente précision de calcul,
- effectuer le calcul le plus rapidement possible.

Annexe III.B

Mise en œuvre informatique de l'espace V_h choisi pour Maxwell.

Ce chapitre a pour but d'expliciter entièrement la mise en œuvre pratique de la formulation discrète, autant pour la construction numérique analytique du système linéaire que pour la reconstruction éventuelle des champs électrique et magnétique $\mathbf{E}$ et $\mathbf{H}$. Dans tout ce chapitre, afin de simplifier les notations, nous considérons que $\Sigma_{k,j}$ est une face plane triangulaire.

III.B.1 Construction du système linéaire.

III.B.1.1 Introduction, notations.

Nous considérons les contributions d'une interface $\Sigma_{k,j}$ ou d'une face de bord $\Sigma_{k,j=k}$ sans effectuer l'éventuelle sommation sur toutes les faces $\Sigma_{k,j}$ concernées (par exemple dans le calcul des termes de la matrice D ou du second membre). Nous laissons au lecteur le soin de calculer les contributions d'une face quadrangulaire et d'effectuer les sommations adéquates. Nous considérons les quantités qui dépendent de la géométrie de l'interface $\Sigma_{k,j}$, des fonctions de base et du milieu présent de part et d'autre de l'interface.

– Les trois sommets de l'interface $\Sigma_{k,j}$ sont notés

$$(\mathbf{X}^1_{k,j}, \mathbf{X}^2_{k,j}, \mathbf{X}^3_{k,j}) \ . \tag{III.B.1}$$

Pour faciliter un jeu d'écriture, on utilisera $\mathbf{X}^4_{k,j}$ désignant $\mathbf{X}^1_{k,j}$, et ainsi de suite.

– Nous notons $S_{k,j}$ la surface de la face plane $\Sigma_{k,j}$. Ce terme est calculé par

$$S_{k,j} = \frac{1}{2}(\mathbf{X}^1_{k,j}, \mathbf{X}^2_{k,j}, \mathbf{X}^3_{k,j})_{mixte} \tag{III.B.2}$$

– Pour $1 \leq l, m \leq p$, indices des directions de propagation dans Ω_k ou Ω_j, on note

$$\mathbf{v}^{l,m}_{k,j} = \frac{\omega}{2}(\sqrt{\overline{\varepsilon}_j \overline{\mu}_j} V_{j,m} - \sqrt{\varepsilon_k \mu_k} V_{k,l}) \ . \tag{III.B.3}$$

– Sur les trois arêtes définissant Σ_{kj} et pour les directions de propagation définissant $\mathbf{v}^{l,m}_{k,j}$ (III.B.3), on pose

$$h^n_{k,j,l,m} = \mathbf{v}^{l,m}_{k,j} \, . \, (\mathbf{X}^{n+1}_{k,j} - \mathbf{X}^n_{k,j}) \tag{III.B.4}$$

et

$$Z^n_{k,j,l,m} = e^{2i\mathbf{v}^{l,m}_{k,j} \, . \, \mathbf{X}^n_{k,j}} \ . \tag{III.B.5}$$

– Nous notons $f^{l,m}_{k,j}$ le terme intégral sur $\Sigma_{k,j}$ de l'onde plane définie par son vecteur d'onde $2\mathbf{v}^{l,m}_{k,j}$:

$$f^{l,m}_{k,j} = \int_{\Sigma_{k,j}} e^{2i\mathbf{v}^{l,m}_{k,j} \, . \, \mathbf{X}} \, d\mathbf{X} \ . \tag{III.B.6}$$

Nous noterons abusivement $f_{k,k}^{l,m}$ l'intégrale sur $\Sigma_{k,j}$ (et non sur $\Sigma_{k,k}$)

$$f_{k,k}^{l,m} = \int_{\Sigma_{k,j}} e^{2i\mathbf{v}_{k,k}^{l,m} \cdot \mathbf{X}} \, d\,\mathbf{X} \;. \tag{III.B.7}$$

Cette intégrale dépend de la géométrie de Ω_j mais ne dépend que des fonctions de base de Ω_k : en toute rigueur, il faudrait introduire un troisième indice et noter $f_{k,k,j}^{l,m}$.

Nous calculons annexe III.D.1.2 l'intégrale de $e^{i\mathbf{kX}}$ sur une face plane triangulaire. D'après la relation (III.D.5) pour $\mathbf{k} = 2i\mathbf{v}_{k,j}^{l,m}$, on calcule analytiquement $f_{k,j}^{l,m}$ par

$$f_{k,j}^{l,m} = 2S_{k,j} Z_{k,j,l,m}^{n+1} \frac{e^{ih_{k,j,l,m}^{n+1}} \dfrac{\sin h_{k,j,l,m}^{n+1}}{h_{k,j,l,m}^{n+1}} - e^{-ih_{k,j,l,m}^{n}} \dfrac{\sin h_{k,j,l,m}^{n}}{h_{k,j,l,m}^{n}}}{2i(h_{k,j,l,m}^{n+1} + h_{k,j,l,m}^{n})} \tag{III.B.8}$$

indépendamment de n, l'indice des sommets $\mathbf{X}_{k,j}^{n}$ de $\Sigma_{k,j}$. A l'aide du barycentre $\vec{G}$ de l'interface $\Sigma_{k,j}$, on a aussi

$$f_{k,j}^{l,m} = 2S_{k,j} e^{2i\mathbf{v}_{k,j}^{l,m}\vec{G}} e^{-\frac{2}{3}i\left(h_{k,j,l,m}^{n+1} - h_{k,j,l,m}^{n}\right)} \frac{e^{ih_{k,j,l,m}^{n+1}} \dfrac{\sin h_{k,j,l,m}^{n+1}}{h_{k,j,l,m}^{n+1}} - e^{-ih_{k,j,l,m}^{n}} \dfrac{\sin h_{k,j,l,m}^{n}}{h_{k,j,l,m}^{n}}}{2i(h_{k,j,l,m}^{n+1} + h_{k,j,l,m}^{n})} \tag{III.B.9}$$

ou, à l'aide de $\mathbf{v}_{k,j}^{l,m}$ (III.B.3),

$$f_{k,j}^{l,m} = 2S_{k,j} \frac{e^{i\mathbf{v}_{k,j}^{l,m}(\mathbf{X}_{k,j}^{n+2} + \mathbf{X}_{k,j}^{n+1})} \dfrac{\sin \mathbf{v}_{k,j}^{l,m}(\mathbf{X}_{k,j}^{n+2} - \mathbf{X}_{k,j}^{n+1})}{\mathbf{v}_{k,j}^{l,m}(\mathbf{X}_{k,j}^{n+2} - \mathbf{X}_{k,j}^{n+1})} - e^{i\mathbf{v}_{k,j}^{l,m}(\mathbf{X}_{k,j}^{n} + \mathbf{X}_{k,j}^{n+1})} \dfrac{\sin \mathbf{v}_{k,j}^{l,m}(\mathbf{X}_{k,j}^{n} - \mathbf{X}_{k,j}^{n+1})}{\mathbf{v}_{k,j}^{l,m}(\mathbf{X}_{k,j}^{n} - \mathbf{X}_{k,j}^{n+1})}}{2i\mathbf{v}_{k,j}^{l,m}(\mathbf{X}_{k,j}^{n+2} - \mathbf{X}_{k,j}^{n}))} \tag{III.B.10}$$

L'implémentation informatique de cette formule et son implémentation sur une machine à architecture vectorielle sont présentées dans les annexes III.D.2.2 et III.D.3.2.

Les formules donnant l'intégrale $f_{k,k}^{l,m}$ s'obtiennent des formules ci-dessus en remplaçant j par k partout où apparaissent les fonctions de base, mais pas dans les termes intrinsèques à la géométrie.

Enfin, rappelons quelques définitions ou notations qui servent dans ce chapitre.

– Les $L = 2p$ fonctions de base $\mathcal{Z}_{kl}$ de l'élément Ω_k sont définies, pour $1 \le l \le p$, par

$$\begin{cases} \mathcal{Z}_{k,l} = \mathcal{Z}_{k,l}^{0} e^{i\omega\sqrt{\varepsilon_k \mu_k}(V_{k,l} \,.\, \mathbf{X})} \\ \mathcal{Z}_{k,l+p} = \mathcal{Z}_{k,l+p}^{0} e^{i\omega\sqrt{\varepsilon_k \mu_k}(V_{k,l} \,.\, \mathbf{X})} \end{cases} \tag{III.B.11}$$

avec

$$\begin{cases} \mathcal{Z}_{kl}^{0} = \left(\sqrt{\varepsilon_{kj}}\sqrt{\mu_k}\mathbf{F}_{k,l} \wedge \nu_k + i\sqrt{\mu_{kj}}\sqrt{\varepsilon_k}\,(\mathbf{F}_{k,l} \wedge \nu_k) \wedge \nu_k\right) \\ \mathcal{Z}_{k,l+p}^{0} = \left(\sqrt{\varepsilon_{kj}}\sqrt{\mu_k}\mathbf{G}_{k,l} \wedge \nu_k - i\sqrt{\mu_{kj}}\sqrt{\varepsilon_k}\,(\mathbf{G}_{k,l} \wedge \nu_k) \wedge \nu_k\right) \end{cases} \tag{III.B.12}$$

où les fonctions $\mathbf{F}_{k,l}$ et $\mathbf{G}_{k,l}$ sont de la forme

$$\begin{cases} \mathbf{F}_{k,l} = \left(\mathbf{E}_{k,l}^{0} + i\mathbf{E}_{k,l}^{0} \wedge V_{k,l}\right) \\ \mathbf{G}_{k,l} = \left(\mathbf{E}_{k,l}^{0} - i\mathbf{E}_{k,l}^{0} \wedge V_{k,l}\right) \;. \end{cases} \tag{III.B.13}$$

– Remarquons que les fonctions de base $\mathcal{Z}_{kl}$ pour $1 \le l \le p$ sont issues des fonctions $\mathbf{F}_{k,l}$. De même les fonctions de base $\mathcal{Z}_{kl}$ pour $p+1 \le l \le 2p$ sont issues des fonctions $\mathbf{G}_{k,l-p}$. C'est pourquoi nous emploierons le terme de "fonction de type $\mathbf{F}$" (resp. "type $\mathbf{G}$") fonction que nous noterons $\mathcal{Z}_{kl}^{\mathbf{F}}$ (resp. $\mathcal{Z}_{kl}^{\mathbf{G}}$). Les fonctions $\mathcal{Z}_{kl}^{\mathbf{F}}$ et $\mathcal{Z}_{kl}^{\mathbf{G}}$ sont définies par

$$\begin{aligned} &1 \le l \le p \Rightarrow \mathcal{Z}_{k,l}^{\mathbf{F}} = \mathcal{Z}_{k,l} \\ &2p \ge l > p \Rightarrow \mathcal{Z}_{k,l-p}^{\mathbf{G}} = \mathcal{Z}_{k,l} \;. \end{aligned} \tag{III.B.14}$$

– Le vecteur $\mathbf{E}^0_{k,l}$ est réel unitaire et orthogonal à $V_{k,l}$ lui aussi réel unitaire.
– La fonction réelle ε_{kj} (resp. μ_{kj}) est définie comme la moyenne géométrique de la valeur absolue de la permittivité ε (resp. perméabilité μ) sur Ω_k et sur Ω_j :

$$\text{(III.B.15)} \qquad \begin{aligned} \varepsilon_{kj} &= \sqrt{|\varepsilon_k||\varepsilon_j|} \\ \mu_{kj} &= \sqrt{|\mu_k||\mu_j|} \ . \end{aligned}$$

Nous utiliserons aussi, pour $1 \leq l \leq p$, la notation

$$\text{(III.B.16)} \qquad \begin{cases} \mathcal{Z}^1_{kl} = \Big(-\sqrt{\varepsilon_{kj}}\sqrt{\mu_k}\mathbf{F}_{k,l} \wedge \nu_k + i\sqrt{\mu_{kj}}\sqrt{\varepsilon_k}\,(\mathbf{F}_{k,l} \wedge \nu_k) \wedge \nu_k\Big) \\ \mathcal{Z}^1_{k,l+p} = \Big(-\sqrt{\varepsilon_{kj}}\sqrt{\mu_k}\mathbf{G}_{k,l} \wedge \nu_k - i\sqrt{\mu_{kj}}\sqrt{\varepsilon_k}\,(\mathbf{G}_{k,l} \wedge \nu_k) \wedge \nu_k\Big) \ . \end{cases}$$

Remarque 49 D'après (III.B.13), on a

$$\mathbf{G}_{k,l} = \overline{\mathbf{F}_{k,l}} \ ,$$

d'où

$$\begin{aligned} \mathcal{Z}^0_{k,l+p} &= \Big(\sqrt{\varepsilon_{kj}}\sqrt{\mu_k}\mathbf{G}_{k,l} \wedge \nu_k - i\sqrt{\mu_{kj}}\sqrt{\varepsilon_k}\,(\mathbf{G}_{k,l} \wedge \nu_k) \wedge \nu_k\Big) \\ &= \Big(\overline{\sqrt{\varepsilon_{kj}}\sqrt{\mu_k}\mathbf{F}_{k,l} \wedge \nu_k + i\sqrt{\mu_{kj}}\sqrt{\varepsilon_k}\,(\mathbf{F}_{k,l} \wedge \nu_k) \wedge \nu_k}\Big) \ . \end{aligned}$$

Si ε et μ sont réels sur Ω_k alors

$$\text{(III.B.17)} \qquad \begin{cases} \mathcal{Z}^0_{k,l+p} = \overline{\mathcal{Z}^0_{k,l}} \\ \mathcal{Z}^1_{k,l+p} = \overline{\mathcal{Z}^1_{k,l}} \ . \end{cases}$$

III.B.1.2 Forme des matrices.

La construction de Galerkin de l'espace de discrétisation implique que la matrice $D - C$ du système linéaire est essentiellement creuse. Nous avons montré, section II.8.2.3, la nullité, dans le vide, de certains des termes des matrices. Nous illustrons ces résultats par l'exemple de la configuration de la figure (III.B.1) qui est un maillage d'une pyramide à base quadrangulaire Ω en 4 éléments tétraédriques. La figure est une projection du maillage sur le plan de base de la pyramide. Dans cette figure, 2 éléments sont dans le vide où $\varepsilon = \mu = 1$. L'existence d'éléments dans le vide entraîne la nullité de certains termes des matrices (lemme 19 p. 107). C'est pourquoi nous avons repéré les différents types d'interfaces $\Sigma_{k,j}$ selon les cas suivants.

i) Cas général où $\Sigma_{k,j}$ est une face entre deux éléments Ω_k et Ω_j, Ω_k possédant du matériau, Ω_j quelconque.

ii) Cas simplifié où les deux éléments sont placés dans le vide.

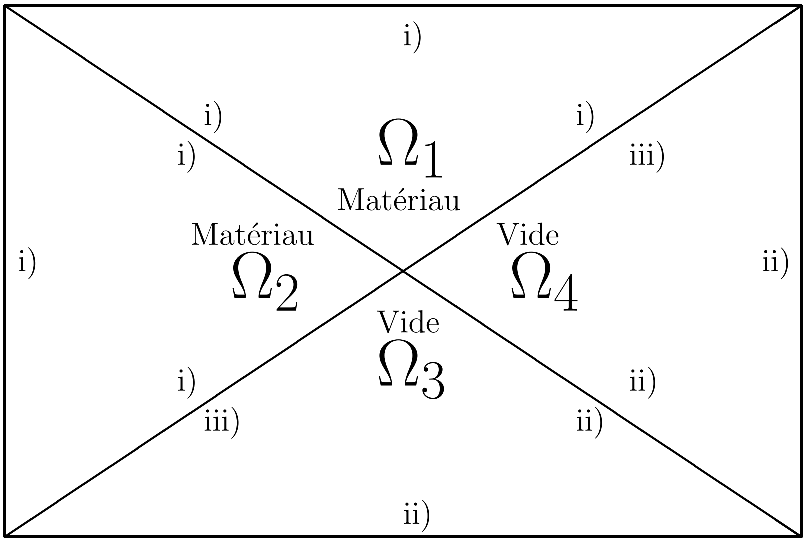
iii) Cas d'une interface ($j \neq k$) telle que Ω_k soit dans le vide et Ω_j possède un matériau.

En plus de la réduction de la taille mémoire (remarque 40 p. 107), ces différents cas ont une importance pour l'optimisation du calcul numérique (par exemple en utilisant les propriétés de la section III.B.1.3.3).

La matrice D est nulle partout sauf sur des blocs diagonaux de taille $2p \times 2p$. Nous notons $D^{\mathbf{F},\mathbf{F}}$ les termes d'indices $1 \leq l, m \leq p$ correspondant aux produits scalaires des fonctions de type $\mathbf{F}$, $D^{\mathbf{F},\mathbf{G}}$ pour les indices $l \leq p$ et $m \geq p+1$ et $D^{\mathbf{G},\mathbf{G}}$ pour $p+1 \leq l, m \leq 2p$. La forme de la matrice hermitienne D est la suivante :

$$D = \begin{bmatrix} \begin{matrix} D_1^{\mathbf{F},\mathbf{F}} & D_1^{\mathbf{F},\mathbf{G}} \\ - & D_1^{\mathbf{G},\mathbf{G}} \end{matrix} & 0 & 0 & 0 \\ - & \begin{matrix} D_2^{\mathbf{F},\mathbf{F}} & D_2^{\mathbf{F},\mathbf{G}} \\ - & D_2^{\mathbf{G},\mathbf{G}} \end{matrix} & 0 & 0 \\ - & - & \begin{matrix} D_3^{\mathbf{F},\mathbf{F}} & D_{3,2}^{\mathbf{F},\mathbf{G}} \\ - & D_3^{\mathbf{G},\mathbf{G}} \end{matrix} & 0 \\ - & - & - & \begin{matrix} D_4^{\mathbf{F},\mathbf{F}} & D_{4,1}^{\mathbf{F},\mathbf{G}} \\ - & D_4^{\mathbf{G},\mathbf{G}} \end{matrix} \end{bmatrix}$$

FIG. III.B.1 – Exemple de configuration des éléments.



où la notation $D_{3,2}^{\mathbf{F},\mathbf{G}}$ (resp. $D_{4,1}^{\mathbf{F},\mathbf{G}}$) est la contribution de la face $\Sigma_{4,1}$ entre Ω_4 et Ω_1 (resp. de la face $\Sigma_{3,2}$ entre Ω_3 et Ω_2) au calcul de $D_3^{\mathbf{F},\mathbf{G}}$ (resp. $D_4^{\mathbf{F},\mathbf{G}}$). On remarque que $D_3^{\mathbf{F},\mathbf{G}} = D_{3,2}^{\mathbf{F},\mathbf{G}}$ (resp. $D_4^{\mathbf{F},\mathbf{G}} = D_{4,1}^{\mathbf{F},\mathbf{G}}$) puisque les contributions des autres faces sont nulles.

De même, la matrice C est de la forme essentiellement creuse suivante

$$C = \begin{bmatrix}
C_{1,1}^{\mathbf{F},\mathbf{F}} & C_{1,1}^{\mathbf{F},\mathbf{G}} & C_{1,2}^{\mathbf{F},\mathbf{F}} & C_{1,2}^{\mathbf{F},\mathbf{G}} & & & C_{1,4}^{\mathbf{F},\mathbf{F}} & C_{1,4}^{\mathbf{F},\mathbf{G}} \\
C_{1,1}^{\mathbf{G},\mathbf{F}} & C_{1,1}^{\mathbf{G},\mathbf{G}} & C_{1,2}^{\mathbf{G},\mathbf{F}} & C_{1,2}^{\mathbf{G},\mathbf{G}} & \multicolumn{2}{c}{0} & C_{1,4}^{\mathbf{G},\mathbf{F}} & C_{1,4}^{\mathbf{G},\mathbf{G}} \\
C_{2,1}^{\mathbf{F},\mathbf{F}} & C_{2,1}^{\mathbf{F},\mathbf{G}} & C_{2,2}^{\mathbf{F},\mathbf{F}} & C_{2,2}^{\mathbf{F},\mathbf{G}} & C_{2,3}^{\mathbf{F},\mathbf{F}} & C_{2,3}^{\mathbf{F},\mathbf{G}} & \multicolumn{2}{c}{0} & \\
C_{2,1}^{\mathbf{G},\mathbf{F}} & C_{2,1}^{\mathbf{G},\mathbf{G}} & C_{2,2}^{\mathbf{G},\mathbf{F}} & C_{2,2}^{\mathbf{G},\mathbf{G}} & C_{2,3}^{\mathbf{G},\mathbf{F}} & C_{2,3}^{\mathbf{G},\mathbf{G}} & & \\
 & & C_{3,2}^{\mathbf{F},\mathbf{F}} & C_{3,2}^{\mathbf{F},\mathbf{G}} & 0 & C_{3,3}^{\mathbf{F},\mathbf{G}} & C_{3,4}^{\mathbf{F},\mathbf{F}} & 0 \\
\multicolumn{2}{c}{0} & C_{3,2}^{\mathbf{G},\mathbf{F}} & C_{3,2}^{\mathbf{G},\mathbf{G}} & C_{3,3}^{\mathbf{G},\mathbf{F}} & 0 & 0 & C_{3,4}^{\mathbf{G},\mathbf{G}} \\
C_{4,1}^{\mathbf{F},\mathbf{F}} & C_{4,1}^{\mathbf{F},\mathbf{G}} & \multicolumn{2}{c}{0} & C_{4,3}^{\mathbf{F},\mathbf{F}} & 0 & 0 & C_{4,4}^{\mathbf{F},\mathbf{G}} \\
C_{4,1}^{\mathbf{G},\mathbf{F}} & C_{4,1}^{\mathbf{G},\mathbf{G}} & & & 0 & C_{4,3}^{\mathbf{G},\mathbf{G}} & C_{4,4}^{\mathbf{G},\mathbf{F}} & 0
\end{bmatrix}$$

On remarque que certains sous-blocs de couplage de fonctions de types différents (couplage des fonctions de type $\mathbf{F}$ avec des fonctions de type $\mathbf{G}$ ou l'inverse) sont nuls. C'est le cas des sous-blocs $C_{3,4}$ et $C_{4,3}$, couplages sur une interface dans le vide puisque ces deux éléments sont dans le vide. On remarque que certains sous-blocs de C sont nuls pour des termes de couplage de fonctions de même type (types $\mathbf{F}$ ou $\mathbf{G}$). C'est le cas des blocs $C_{3,3}$ et $C_{4,4}$, couplages sur des faces de bord dans le vide (puisque ces deux éléments sont dans le vide).

III.B.1.3 Assemblage des matrices.

Nous calculons par des formules analytiques la contribution de $\Sigma_{k,j}$ pour les matrices D (II.8.6) et C (II.8.7 pour $j \neq k$ et II.8.8 pour $j = k$). Nous écrirons ces termes en fonction des vecteurs définis en (III.B.13).

III.B.1.3.1 Calcul de la matrice D.

Nous notons $D^{k,j}_{l,m}$ la contribution de la face $\Sigma_{k,j}$ pour le calcul de $D^{l,m}_k$ dans (II.8.6). On a alors, pour $1 \leq l, m \leq L = 2p$, à l'aide des notations (III.B.3) et (III.B.12),

$$D^{k,j}_{l,m} = \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}} \mathcal{Z}^0_{k,m} \overline{\mathcal{Z}}^0_{k,l} \int_{\Sigma_{k,j}} e^{2i\mathbf{v}^{l,m}_{k,k} \cdot \mathbf{X}} \, d\mathbf{X} \tag{III.B.18}$$

soit, par définition de $f^{l,m}_{k,k}$ (III.B.7),

$$D^{k,j}_{l,m} = \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}} \mathcal{Z}^0_{k,m} \overline{\mathcal{Z}}^0_{k,l} f^{l,m}_{k,k} \; . \tag{III.B.19}$$

Nous savons calculer analytiquement le terme $f^{l,m}_{k,k}$ par la formule (III.B.8). Il reste à calculer la matrice hermitienne D^0 donnée par

$$D^0_{l,m} = \mathcal{Z}^0_{k,m} \overline{\mathcal{Z}}^0_{k,l} \tag{III.B.20}$$

pour $1 \leq l, m \leq 2p$. Cette matrice est constituée de quatre blocs selon les valeurs des indices par rapport à p. Dans les formules énumérées ci-dessous, les indices l et m sont tels que $1 \leq l, m \leq p$.

1. Le bloc supérieur gauche des fonctions de type $\mathbf{F}$, termes $D^0_{l,m}$, se calcule par

$$\begin{aligned} D^0_{l,m} = &\left(\overline{\sqrt{\varepsilon_{kj}}\sqrt{\overline{\mu}_k}\mathbf{F}_{k,l} \wedge \nu_k + i\sqrt{\mu_{kj}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{F}_{k,l} \wedge \nu_k\right) \wedge \nu_k}\right) \\ &\left(\sqrt{\varepsilon_{kj}}\sqrt{\overline{\mu}_k}\mathbf{F}_{k,m} \wedge \nu_k + i\sqrt{\mu_{kj}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{F}_{k,m} \wedge \nu_k\right) \wedge \nu_k\right) \; . \end{aligned} \tag{III.B.21}$$

2. Le bloc supérieur droit des produits scalaires de types $(\mathbf{F}, \mathbf{G})$, termes $D^0_{l,m+p}$, se calcule par

$$\begin{aligned} D^0_{l,m+p} = &\left(\overline{\sqrt{\varepsilon_{kj}}\sqrt{\overline{\mu}_k}\mathbf{F}_{k,l} \wedge \nu_k + i\sqrt{\mu_{kj}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{F}_{k,l} \wedge \nu_k\right) \wedge \nu_k}\right) \\ &\left(\sqrt{\varepsilon_{kj}}\sqrt{\overline{\mu}_k}\mathbf{G}_{k,m} \wedge \nu_k - i\sqrt{\mu_{kj}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{G}_{k,m} \wedge \nu_k\right) \wedge \nu_k\right) \; , \end{aligned} \tag{III.B.22}$$

   ou, d'après la remarque 49, par

$$\begin{aligned} \overline{D^0_{l,m+p}} = &\left(\sqrt{\varepsilon_{kj}}\sqrt{\overline{\mu}_k}\mathbf{F}_{k,l} \wedge \nu_k + i\sqrt{\mu_{kj}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{F}_{k,l} \wedge \nu_k\right) \wedge \nu_k\right) \\ &\left(\sqrt{\varepsilon_{kj}}\sqrt{\mu_k}\mathbf{F}_{k,m} \wedge \nu_k + i\sqrt{\mu_{kj}}\sqrt{\varepsilon_k}\left(\mathbf{F}_{k,m} \wedge \nu_k\right) \wedge \nu_k\right) \; . \end{aligned} \tag{III.B.23}$$

   Si $\varepsilon = \mu = 1$ sur Ω_k et Ω_j alors

$$D^0_{l,m+p} = 0 \; . \tag{III.B.24}$$

3. Le bloc inférieur droit des fonctions de type $\mathbf{G}$, termes $D^0_{l+p,m+p}$, se calcule par

$$\begin{aligned} D^0_{l+p,m+p} = &\left(\overline{\sqrt{\varepsilon_{kj}}\sqrt{\overline{\mu}_k}\mathbf{G}_{k,l} \wedge \nu_k - i\sqrt{\mu_{kj}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{G}_{k,l} \wedge \nu_k\right) \wedge \nu_k}\right) \\ &\left(\sqrt{\varepsilon_{kj}}\sqrt{\overline{\mu}_k}\mathbf{G}_{k,m} \wedge \nu_k - i\sqrt{\mu_{kj}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{G}_{k,m} \wedge \nu_k\right) \wedge \nu_k\right) \; . \end{aligned} \tag{III.B.25}$$

   D'après la remarque 49, on peut exprimer $D_{l+p,m+p}$ en fonction de $\mathbf{F}_{k,l}$ et $\mathbf{F}_{k,m}$:

$$\begin{aligned} D^0_{l+p,m+p} = &\left(\sqrt{\varepsilon_{kj}}\sqrt{\mu_k}\mathbf{F}_{k,l} \wedge \nu_k + i\sqrt{\mu_{kj}}\sqrt{\varepsilon_k}\left(\mathbf{F}_{k,l} \wedge \nu_k\right) \wedge \nu_k\right) \\ &\left(\overline{\sqrt{\varepsilon_{kj}}\sqrt{\mu_k}\mathbf{F}_{k,m} \wedge \nu_k + i\sqrt{\mu_{kj}}\sqrt{\varepsilon_k}\left(\mathbf{F}_{k,m} \wedge \nu_k\right) \wedge \nu_k}\right) \end{aligned} \tag{III.B.26}$$

   ce qui montre que si ε et μ sont réels sur Ω_k alors

$$D^0_{l+p,m+p} = \overline{D^0_{l,m}} \; .$$

III.B.1.3.2 Calcul de la matrice C.

i) Rappelons que la contribution de l'interface interne $\Sigma_{k,j\neq k}$ dans la matrice de couplage non hermitien C (II.8.7) est donnée par

$$C_{l,m}^{k,j} = \int_{\Sigma_{k,j}} \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}} \mathcal{Z}_{j,m} \overline{F \mathcal{Z}_{k,l}} d\,\mathbf{X} \tag{III.B.27}$$

On a alors, pour $1 \leq l, m \leq L = 2p$, à l'aide des notations (III.B.3), (III.B.12) et (III.B.16),

$$C_{l,m}^{k,j} = \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}} \mathcal{Z}_{j,m}^{0} \overline{\mathcal{Z}}_{k,l}^{1} \int_{\Sigma_{k,j}} e^{2i\mathbf{v}_{k,j}^{l,m}\cdot\mathbf{X}} d\,\mathbf{X} \tag{III.B.28}$$

soit, par définition de $f_{k,j}^{l,m}$ (III.B.6),

$$C_{l,m}^{k,j} = \frac{1}{\sqrt{\varepsilon_{kj}\mu_{kj}}} \mathcal{Z}_{j,m}^{0} \overline{\mathcal{Z}}_{k,l}^{1} f_{k,j}^{l,m} \ . \tag{III.B.29}$$

Nous savons calculer analytiquement le terme $f_{k,j}^{l,m}$ par la formule (III.B.8). Il reste à calculer la matrice C^0 donnée par

$$C_{l,m}^{0} = \mathcal{Z}_{j,m}^{0} \overline{\mathcal{Z}}_{k,l}^{1}$$

pour $1 \leq l, m \leq 2p$. Cette matrice est constituée de quatre blocs selon les valeurs des indices par rapport à p. Nous calculons C^0 par les formules qui suivent en faisant varier les indices l et m de 1 à p.

1. Le bloc supérieur gauche des fonctions de type $\mathbf{F}$, termes $C_{l,m}^{0}$, se calcule par

$$\begin{aligned} C_{l,m}^{0} = & \overline{\left(-\sqrt{\varepsilon_{kj}}\sqrt{\mu_k}\mathbf{F}_{k,l}\wedge\nu_k + i\sqrt{\mu_{kj}}\sqrt{\varepsilon_k}\left(\mathbf{F}_{k,l}\wedge\nu_k\right)\wedge\nu_k\right)} \\ & \left(-\sqrt{\varepsilon_{kj}}\sqrt{\mu_j}\mathbf{F}_{j,m}\wedge\nu_k + i\sqrt{\mu_{kj}}\sqrt{\varepsilon_j}\left(\mathbf{F}_{j,m}\wedge\nu_k\right)\wedge\nu_k\right) \ . \end{aligned} \tag{III.B.30}$$

2. Le bloc supérieur droit des couplages des fonctions de type $\mathbf{F}$ avec les fonctions de type $\mathbf{G}$, termes $C_{l,m+p}^{0}$, se calcule par

$$\begin{aligned} C_{l,m+p}^{0} = & \overline{\left(-\sqrt{\varepsilon_{kj}}\sqrt{\mu_k}\mathbf{F}_{k,l}\wedge\nu_k + i\sqrt{\mu_{kj}}\sqrt{\varepsilon_k}\left(\mathbf{F}_{k,l}\wedge\nu_k\right)\wedge\nu_k\right)} \\ & \left(-\sqrt{\varepsilon_{kj}}\sqrt{\mu_j}\mathbf{G}_{j,m}\wedge\nu_k - i\sqrt{\mu_{kj}}\sqrt{\overline{\varepsilon}_j}\left(\mathbf{G}_{j,m}\wedge\nu_k\right)\wedge\nu_k\right) \ , \end{aligned} \tag{III.B.31}$$

   et, lorsque $\varepsilon = \mu = 1$ sur Ω_k et Ω_j, on a

$$C_{l,m+p}^{0} = 0 \ . \tag{III.B.32}$$

3. Le bloc inférieur gauche des couplages des fonctions de type $\mathbf{G}$ avec les fonctions de type $\mathbf{F}$, termes $C_{l+p,m}^{0}$, se calcule par

$$\begin{aligned} C_{l+p,m}^{0} = & \left(-\sqrt{\varepsilon_{kj}}\sqrt{\mu_k}\mathbf{F}_{j,m}\wedge\nu_k + i\sqrt{\mu_{kj}}\sqrt{\varepsilon_k}\left(\mathbf{F}_{j,m}\wedge\nu_k\right)\wedge\nu_k\right) \\ & \left(-\sqrt{\varepsilon_{kj}}\sqrt{\mu_j}\mathbf{F}_{k,l}\wedge\nu_k + i\sqrt{\mu_{kj}}\sqrt{\varepsilon_j}\left(\mathbf{F}_{k,l}\wedge\nu_k\right)\wedge\nu_k\right) \ , \end{aligned} \tag{III.B.33}$$

   et, lorsque $\varepsilon = \mu = 1$ sur Ω_k et Ω_j, on a

$$C_{l,m+p}^{0} = 0 \ . \tag{III.B.34}$$

4. Le bloc inférieur droit des fonctions de type $\mathbf{G}$, termes $C_{l+p,m+p}^{0}$, se calcule par

$$\begin{aligned} C_{l+p,m+p}^{0} = & \overline{\left(-\sqrt{\varepsilon_{kj}}\sqrt{\mu_k}\mathbf{G}_{k,l}\wedge\nu_k - i\sqrt{\mu_{kj}}\sqrt{\varepsilon_k}\left(\mathbf{G}_{k,l}\wedge\nu_k\right)\wedge\nu_k\right)} \\ & \left(-\sqrt{\varepsilon_{kj}}\sqrt{\mu_j}\mathbf{G}_{j,m}\wedge\nu_k - i\sqrt{\mu_{kj}}\sqrt{\overline{\varepsilon}_j}\left(\mathbf{G}_{j,m}\wedge\nu_k\right)\wedge\nu_k\right) \ . \end{aligned} \tag{III.B.35}$$

D'après la remarque 49 on peut exprimer $C_{l+p,m+p}$ en fonction de $\mathbf{F}_{k,l}$ et $\mathbf{F}_{j,m}$:

$$\text{(III.B.36)}\qquad \begin{aligned} C^0_{l+p,m+p} = &\left(-\sqrt{\varepsilon_{kj}}\sqrt{\mu_k}\mathbf{F}_{k,l}\wedge\nu_k + i\sqrt{\mu_{kj}}\sqrt{\varepsilon_k}\left(\mathbf{F}_{k,l}\wedge\nu_k\right)\wedge\nu_k\right) \\ &\left(\overline{-\sqrt{\varepsilon_{kj}}\sqrt{\mu_j}\mathbf{F}_{j,m}\wedge\nu_k + i\sqrt{\mu_{kj}}\sqrt{\varepsilon_j}\left(\mathbf{F}_{j,m}\wedge\nu_k\right)\wedge\nu_k}\right)\ , \end{aligned}$$

ce qui montre que, si ε et μ sont réels sur Ω_k et Ω_j, alors

$$C^0_{l+p,m+p} = \overline{C^0_{l,m}}\ .$$

ii) La contribution de la face de bord $\Sigma_{k,k}$ dans la matrice de couplage non hermitien C (II.8.8) est

$$\text{(III.B.37)}\qquad C^{k,k}_{l,m} = \int_{\Sigma_{k,k}} Q_k \frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}}\mathcal{Z}_{k,m}\overline{F\mathcal{Z}}_{k,l}d\,\mathbf{X}\ .$$

Alors, pour $1 \le l, m \le L = 2p$, à l'aide des notations (III.B.3), (III.B.12) et (III.B.16), on déduit que

$$\text{(III.B.38)}\qquad C^{k,k}_{l,m} = Q_k\frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}}\mathcal{Z}^0_{k,m}\overline{\mathcal{Z}}^1_{k,l}\int_{\Sigma_{k,k}} e^{2i\mathbf{v}^{l,m}_{k,k}\cdot\mathbf{X}}\,d\,\mathbf{X}\ ,$$

où l'on reconnaît le terme $f^{l,m}_{k,k}$ (III.B.7), que nous savons calculer analytiquement par la formule (III.B.8). Il reste à calculer le bloc C^0 donné par

$$C^0_{l,m} = \mathcal{Z}^0_{k,m}\overline{\mathcal{Z}}^1_{k,l}$$

pour $1 \le l, m \le 2p$. Pour $1 \le l, m \le p$, nous calculons C^0 par les formules ci-dessous.

1. Le bloc supérieur gauche des fonctions de type $\mathbf{F}$, termes $C^0_{l,m}$, se calcule par

$$\text{(III.B.39)}\qquad \begin{aligned} C^0_{l,m} = &\left(\overline{-\sqrt{\varepsilon_{kk}}\sqrt{\overline{\mu}_k}\mathbf{F}_{k,l}\wedge\nu_k + i\sqrt{\mu_{kk}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{F}_{k,l}\wedge\nu_k\right)\wedge\nu_k}\right) \\ &\left(+\sqrt{\varepsilon_{kk}}\sqrt{\overline{\mu}_k}\mathbf{F}_{k,m}\wedge\nu_k + i\sqrt{\mu_{kk}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{F}_{k,m}\wedge\nu_k\right)\wedge\nu_k\right)\ , \end{aligned}$$

   et, lorsque $\varepsilon = \mu = 1$ sur Ω_k, on a

$$\text{(III.B.40)}\qquad C_{l,m} = 0\ .$$

2. Le bloc supérieur droit des couplages des fonctions de type $\mathbf{F}$ avec les fonctions de type $\mathbf{G}$, termes $C^0_{l,m+p}$, se calcule par

$$\text{(III.B.41)}\qquad \begin{aligned} C^0_{l,m+p} = &\left(\overline{-\sqrt{\varepsilon_{kk}}\sqrt{\overline{\mu}_k}\mathbf{F}_{k,l}\wedge\nu_k + i\sqrt{\mu_{kk}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{F}_{k,l}\wedge\nu_k\right)\wedge\nu_k}\right) \\ &\left(+\sqrt{\varepsilon_{kk}}\sqrt{\overline{\mu}_k}\mathbf{G}_{k,m}\wedge\nu_k - i\sqrt{\mu_{kk}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{G}_{k,m}\wedge\nu_k\right)\wedge\nu_k\right)\ . \end{aligned}$$

   D'après la remarque 49, on peut exprimer $C^0_{l,m+p}$ en fonction de $\mathbf{F}_{k,l}$ et $\mathbf{F}_{k,m}$. On a en effet

$$\text{(III.B.42)}\qquad \begin{aligned} \overline{C^0_{l,m+p}} = &\left(-\sqrt{\varepsilon_{kk}}\sqrt{\overline{\mu}_k}\mathbf{F}_{k,l}\wedge\nu_k + i\sqrt{\mu_{kk}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{F}_{k,l}\wedge\nu_k\right)\wedge\nu_k\right) \\ &\left(+\sqrt{\varepsilon_{kk}}\sqrt{\mu_k}\mathbf{F}_{k,m}\wedge\nu_k + i\sqrt{\mu_{kk}}\sqrt{\varepsilon_k}\left(\mathbf{F}_{k,m}\wedge\nu_k\right)\wedge\nu_k\right)\ . \end{aligned}$$

3. Le bloc inférieur gauche des couplages des fonctions de type $\mathbf{G}$ avec les fonctions de type $\mathbf{F}$, termes $C^0_{l+p,m}$, se calcule par

$$\text{(III.B.43)}\qquad \begin{aligned} C^0_{l+p,m} = &\left(-\sqrt{\varepsilon_{kk}}\sqrt{\overline{\mu}_k}\mathbf{F}_{k,m}\wedge\nu_k + i\sqrt{\mu_{kk}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{F}_{k,m}\wedge\nu_k\right)\wedge\nu_k\right) \\ &\left(+\sqrt{\varepsilon_{kk}}\sqrt{\mu_k}\mathbf{F}_{k,l}\wedge\nu_k + i\sqrt{\mu_{kk}}\sqrt{\varepsilon_k}\left(\mathbf{F}_{k,l}\wedge\nu_k\right)\wedge\nu_k\right)\ , \end{aligned}$$

   ce qui montre que, si ε et μ sont réels sur Ω_k, alors

$$C^0_{l+p,m} = \overline{C^0_{l,m+p}}\ .$$

4. Le bloc inférieur droit des fonctions de type $\mathbf{G}$, termes $C^0_{l+p,m+p}$, se calcule par

$$C^0_{l+p,m+p} = \overline{\left(-\sqrt{\varepsilon_{kk}}\sqrt{\overline{\mu}_k}\mathbf{G}_{k,l} \wedge \nu_k - i\sqrt{\mu_{kk}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{G}_{k,l} \wedge \nu_k\right) \wedge \nu_k\right)} \left(+\sqrt{\varepsilon_{kk}}\sqrt{\overline{\mu}_k}\mathbf{G}_{k,m} \wedge \nu_k - i\sqrt{\mu_{kk}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{G}_{k,m} \wedge \nu_k\right) \wedge \nu_k\right) . \tag{III.B.44}$$

D'après la remarque 49, on peut exprimer $C^0_{l+p,m+p}$ en fonction de $\mathbf{F}_{k,l}$ et $\mathbf{F}_{k,m}$:

$$C^0_{l+p,m+p} = \left(-\sqrt{\varepsilon_{kk}}\sqrt{\mu_k}\mathbf{F}_{k,l} \wedge \nu_k + i\sqrt{\mu_{kk}}\sqrt{\varepsilon_k}\left(\mathbf{F}_{k,l} \wedge \nu_k\right) \wedge \nu_k\right) \overline{\left(+\sqrt{\varepsilon_{kk}}\sqrt{\mu_k}\mathbf{F}_{k,m} \wedge \nu_k + i\sqrt{\mu_{kk}}\sqrt{\varepsilon_k}\left(\mathbf{F}_{k,m} \wedge \nu_k\right) \wedge \nu_k\right)} , \tag{III.B.45}$$

et, lorsque $\varepsilon = \mu = 1$ sur Ω_k, on a

$$C^0_{l+p,m+p} = 0 .$$

III.B.1.3.3 Une propriété supplémentaire des matrices.

Pour un élément Ω_k strictement intérieur au domaine Ω (c'est-à-dire sans face incluse dans $\partial\Omega$) tel que le milieu présent dans l'élément Ω_k et dans tous ses voisins soit le vide, si l'on prend des fonctions de base identiques sur Ω_k et sur tous ses voisins, alors

$$D^{l,m}_k = \sum_{j(k)} C^{l,m}_{j,k} . \tag{III.B.46}$$

Toujours dans le cas de fonctions de base identiques sur Ω_k et $\Omega_j \neq \Omega_k$ et où ces deux éléments sont placés dans le vide, alors

$$C^{l,m}_{k,j} = \overline{C^{m,l}_{k,j}} . \tag{III.B.47}$$

III.B.1.4 Assemblage du second membre, cas particuliers.

Nous calculons les $2pK$ termes du second membre b (II.8.2) dans des cas particuliers où les termes sources du problème de Maxwell permettent un calcul analytique. Pour une résolution des équations de Maxwell avec des termes sources plus compliqués, il faudrait assembler les termes du second membre de la formulation discrète grâce à une intégration numérique.

Nous utiliserons les notations de la section (III.B.1.1) en définissant des termes d'indice 0 par

$$\mathbf{v}^{l,0}_{k,j} = \frac{\omega}{2}(\sqrt{\overline{\varepsilon}_j\overline{\mu}_j}\mathbf{k}_0 - \sqrt{\varepsilon_k\mu_k}V_{k,l}) . \tag{III.B.48}$$

et de même pour $h^n_{k,j,l,0}$ (III.B.4), $Z^n_{k,j,l,0}$ (III.B.5) et $f^{l,0}_{k,j}$ (III.B.6).

On suppose que Q est une fonction constante pour une face frontière donnée $\Sigma_{k,k}$. Pour $1 \leq l \leq p$, on a d'après (II.8.9),

$$b_{k,l} = -2\int_{\Omega_k} \left(\mathbf{j}\sqrt{\mu_k}\overline{\mathbf{F}_{k,l}} - i\mathbf{m}\sqrt{\varepsilon_k}\overline{\mathbf{F}_{k,l}}\right) e^{-i\omega\left(\sqrt{\varepsilon_k\mu_k}V_{k,m}\,.\,\mathbf{X}\right)} + \int_{\Sigma_{k,k}} \frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}} g.\overline{F\mathcal{Z}_{k,l}}$$

et

$$b_{k,l+p} = -2\int_{\Omega_k} \left(\mathbf{j}\sqrt{\mu_k}\overline{\mathbf{G}_{k,l}} + i\mathbf{m}\sqrt{\varepsilon_k}\overline{\mathbf{G}_{k,l}}\right) e^{-i\omega\left(\left(\sqrt{\varepsilon_k\mu_k}V_{k,m}\right)\,.\,\mathbf{X}\right)} + \int_{\Sigma_{k,k}} \frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}} g.\overline{F\mathcal{Z}_{k,l+p}}$$

ou

$$\begin{aligned} b_{k,l} = &-2\int_{\Omega_k} \left(\sqrt{\mu_k}\mathbf{j}\overline{\mathbf{F}_{k,l}} - i\sqrt{\varepsilon_k}\mathbf{m}\overline{\mathbf{F}_{k,l}}\right) e^{-i\omega\left(\sqrt{\varepsilon_k\mu_k}V_{k,m}\,.\,\mathbf{X}\right)} \\ &+ \int_{\Sigma_{k,k}} \frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}} g.\overline{\left(-\sqrt{\varepsilon_{kk}}\sqrt{\overline{\mu}_k}\mathbf{F}_{k,l} \wedge \nu_k + i\sqrt{\mu_{kk}}\sqrt{\overline{\varepsilon}_k}\left(\mathbf{F}_{k,l} \wedge \nu_k\right) \wedge \nu_k\right)} e^{-i\omega\left(\sqrt{\varepsilon_k\mu_k}V_{k,l}\,.\,\mathbf{X}\right)} . \end{aligned} \tag{III.B.49}$$

et

$$\begin{aligned} b_{k,l+p} = &-2\int_{\Omega_k} \left(\sqrt{\mu_k}\mathbf{j}\mathbf{F}_{k,l} + i\sqrt{\varepsilon_k}\mathbf{m}\mathbf{F}_{k,l}\right) e^{-i\omega\left(\sqrt{\varepsilon_k\mu_k}V_{k,m}\,.\,\mathbf{X}\right)} \\ &+ \int_{\Sigma_{k,k}} \frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}} g.\left(-\sqrt{\varepsilon_{kk}}\sqrt{\mu_k}\mathbf{F}_{k,l} \wedge \nu_k + i\sqrt{\mu_{kk}}\sqrt{\varepsilon_k}\left(\mathbf{F}_{k,l} \wedge \nu_k\right) \wedge \nu_k\right) e^{-i\omega\left(\sqrt{\varepsilon_k\mu_k}V_{k,l}\,.\,\mathbf{X}\right)} . \end{aligned} \tag{III.B.50}$$

Nous allons expliciter (III.B.49) et (III.B.50) dans les cas suivants.

i) Calculons le second membre b pour une condition aux limites donnée par un terme de bord g tel que

$$g_{|\Sigma_{kk}} = -(1+Q_k)\sqrt{\varepsilon_{kk}}(\mathbf{E}\wedge\nu_{k,k}) + (1-Q_k)\sqrt{\mu_{kk}}((\mathbf{H}\wedge\nu_{k,k})\wedge\nu_{k,k})$$

où le couple $(\mathbf{E},\mathbf{H})$ est donné par

$$\begin{aligned}\mathbf{E} &= \mathbf{E}_0 e^{i\omega\left((\sqrt{\varepsilon\mu}\mathbf{k}_0)\,.\,\mathbf{X}\right)}\\ \mathbf{H} &= -\sqrt{\frac{\varepsilon}{\mu}}\mathbf{E}_0\wedge\mathbf{k}_0 e^{i\omega\left((\sqrt{\varepsilon\mu}\mathbf{k}_0)\,.\,\mathbf{X}\right)}\ .\end{aligned}$$

On pose

(III.B.51) $$g_k^0 = -(1+Q_k)\sqrt{\varepsilon_{kk}}\mathbf{E}_0\wedge\nu_k - (1-Q_k)\sqrt{\mu_{kk}}\sqrt{\frac{\varepsilon_k}{\mu_k}}\left(\mathbf{E}_0\wedge\mathbf{k}_0\wedge\nu_k\right)\wedge\nu_k$$

Alors,

(III.B.52) $$b_{k,l} = +\frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}}g_k^0.\overline{\left(-\sqrt{\varepsilon_{kk}}\sqrt{\mu_k}\mathbf{F}_{k,l}\wedge\nu_k + i\sqrt{\mu_{kk}}\sqrt{\varepsilon_k}\left(\mathbf{F}_{k,l}\wedge\nu_k\right)\wedge\nu_k\right)}f_{k,j}^{l,0}$$

et

(III.B.53) $$b_{k,l+p} = +\frac{1}{\sqrt{\varepsilon_{kk}\mu_{kk}}}g_k^0.\left(-\sqrt{\varepsilon_{kk}}\sqrt{\mu_k}\mathbf{F}_{k,l}\wedge\nu_k + i\sqrt{\mu_{kk}}\sqrt{\varepsilon_k}\left(\mathbf{F}_{k,l}\wedge\nu_k\right)\wedge\nu_k\right)f_{k,j}^{l,0}\ .$$

Notons que la complexité de ce calcul est équivalente à celle du calcul des termes matriciels (section III.B.1.3).

ii) Pour $(\mathbf{m},\mathbf{j}) = (0,\vec{V}_0\delta_{\mathbf{X}_0})$, nous avons :

(III.B.54) $$b_{k,l} = -2\sqrt{\mu_k}\vec{V}_0\overline{\mathbf{F}_{k,l}}e^{-i\omega\left((\sqrt{\varepsilon_k\mu_k}V_{k,m})\,.\,\mathbf{X}_0\right)}$$

et

(III.B.55) $$b_{k,l+p} = -2\sqrt{\mu_k}\vec{V}_0\mathbf{F}_{k,l}e^{-i\omega\left((\sqrt{\varepsilon_k\mu_k}V_{k,m})\,.\,\mathbf{X}_0\right)}$$

puisque $\overline{\mathbf{G}_{k,l}} = \mathbf{F}_{k,l}$.

iii) Calculons le second membre b pour des termes sources $(\mathbf{m},\mathbf{j})$ donnés par

(III.B.56) $$\begin{cases}\mathbf{m} = 0\\ \mathbf{j} = i\omega\varepsilon\left(\mathbf{E}_0\,.\,\mathbf{k}_0\right)\mathbf{k}_0 e^{i\omega\left((\sqrt{\varepsilon\mu}\mathbf{k}_0)\,.\,\mathbf{X}\right)}\end{cases}$$

où $(\mathbf{E}_0,\mathbf{k}_0)\in(\mathbb{R}^3)^2$ sont tels que $|\mathbf{E}_0| = |\mathbf{k}_0| = 1$ et $(\mathbf{E}_0\,.\,\mathbf{k}_0)\neq 0$. Notons que le champ $(\mathbf{E},\mathbf{H})$ donné par

(III.B.57) $$\begin{cases}\mathbf{E} = \mathbf{E}_0 e^{i\omega\left((\sqrt{\varepsilon_k\mu_k}\mathbf{k}_0)\,.\,\mathbf{X}\right)}\\ \mathbf{H} = -\sqrt{\frac{\varepsilon_k}{\mu_k}}\mathbf{E}_0\wedge\mathbf{k}_0 e^{i\omega\left((\sqrt{\varepsilon_k\mu_k}\mathbf{k}_0)\,.\,\mathbf{X}\right)}\end{cases}$$

est solution des équations de Maxwell dans Ω. Notons en outre que $(\mathbf{E},\mathbf{H})$ ne fait pas partie de l'espace des fonctions de base puisque $(\mathbf{E}_0\,.\,\mathbf{k}_0)\neq 0$.

La contribution volumique de ce terme source $(\mathbf{m},\mathbf{j})$ non nul est

(III.B.58) $$\left|\begin{aligned}b_{k,l} &= -\,2i\omega\varepsilon_k\sqrt{\mu_k}\left(\mathbf{E}_0\,.\,\mathbf{k}_0\right)\left(\overline{\mathbf{F}_{k,l}}\,.\,\mathbf{k}_0\right)\int_{\Omega_k}e^{i\omega\left(\sqrt{\varepsilon_k\mu_k}(\mathbf{k}_0-V_{k,m})\,.\,\mathbf{X}\right)}\\ b_{k,l+p} &= -\,2i\omega\varepsilon_k\sqrt{\mu_k}\left(\mathbf{E}_0\,.\,\mathbf{k}_0\right)\left(\mathbf{F}_{k,l}\,.\,\mathbf{k}_0\right)\int_{\Omega_k}e^{i\omega\left(\sqrt{\varepsilon_k\mu_k}(\mathbf{k}_0-V_{k,m})\,.\,\mathbf{X}\right)}\ .\end{aligned}\right.$$

Il reste à calculer l'intégrale $3D$,

$$I_3 = \int_{\Omega_k}e^{i\omega\left(\sqrt{\varepsilon_k\mu_k}(\mathbf{k}_0-V_{k,m})\,.\,\mathbf{X}\right)}d\,\mathbf{X}\ ,$$

donnée annexe III.D.1.3 en posant $\mathbf{k} = \sqrt{\varepsilon_k\mu_k}(\mathbf{k}_0 - V_{k,m})$.

III.B.2 Reconstruction des champs électrique et magnétique.

La solution $\mathcal{X}$ du problème variationnel (II.7.84) est approchée par une fonction $\mathcal{X}_h$ à l'aide des coefficients complexes $\mathcal{X}_{k,l}$ solutions du système linéaire (II.8.12) et des fonctions de base $\mathcal{Z}_{k,l}$ définies en (II.8.29) :

$$(\mathcal{X}_h)_{\partial\Omega_k} = \sum_{l=1}^{L=2p} \mathcal{X}_{k,l}\mathcal{Z}_{k,l}$$

Montrons comment nous pouvons reconstruire les fonctions approchées $\mathbf{E}_h$ et $\mathbf{H}_h$ des solutions $\mathbf{E}$ et $\mathbf{H}$ du problème initial de Maxwell (1 p. 79) à l'aide des fonctions $(\mathbf{E}^{\mathbf{F}}_{k,l}, \mathbf{H}^{\mathbf{F}}_{k,l})$ ((II.8.22) et (II.8.25)) pour $1 \le l \le p$ et $(\mathbf{E}^{\mathbf{G}}_{k,l-p}, \mathbf{H}^{\mathbf{G}}_{k,l-p})$ ((II.8.23) et (II.8.26)) pour $p+1 \le l \le L = 2p$. Cette section s'attache à construire ces solutions approchées dans trois situations :

i) reconstruction dans tout le domaine Ω,

ii) reconstruction des traces tangentielles des champs sur le maillage,

iii) reconstruction des traces sur le maillage.

III.B.2.1 Reconstruction des champs électrique et magnétique dans l'espace.

Nous avons

$$\begin{cases} \nabla\wedge\mathbf{E} - i\omega\mu\mathbf{H} = -\mathbf{m} & \text{dans } \Omega_k \\ \nabla\wedge\mathbf{H} + i\omega\varepsilon\mathbf{E} = +\mathbf{j} & \text{dans } \Omega_k \\ (\sqrt{\varepsilon_{kj}}\mathbf{E}\wedge\nu + \sqrt{\mu_{kj}}(\mathbf{H}\wedge\nu)\wedge\nu) = \mathcal{X} & \text{sur } \partial\Omega_k\ , \end{cases}$$

il semble logique de suggérer,

$$\begin{cases} \nabla\wedge\mathbf{E}_h - i\omega\mu\mathbf{H}_h = -\mathbf{m} & \text{dans } \Omega_k \\ \nabla\wedge\mathbf{H}_h + i\omega\varepsilon\mathbf{E}_h = +\mathbf{j} & \text{dans } \Omega_k \\ (\sqrt{\varepsilon_{kj}}\mathbf{E}_h\wedge\nu + \sqrt{\mu_{kj}}(\mathbf{H}_h\wedge\nu)\wedge\nu) = \mathcal{X}_h & \text{sur } \partial\Omega_k\ . \end{cases} \tag{III.B.59}$$

La résolution de (III.B.59) est *a priori* équivalente à la résolution du problème de Maxwell originel. Néanmoins, il peut être intéressant de résoudre beaucoup de problèmes dans des domaines restreints plutôt qu'un seul dans un grand domaine. Ceci est l'idée initiale des techniques de décomposition de domaines [24]. Néanmoins, dans le cas particulier où $(\mathbf{m}, \mathbf{j}) = (0,0)$ sur Ω_k, on a, grâce à la linéarité de l'opérateur de relèvement E^* (II.7.94),

$$\begin{aligned} (\mathbf{E}_h)_{|\Omega_k} &= \sqrt{\mu_k}\sum_{l=1}^{p} \mathcal{X}_{k,l}\left(\mathbf{E}^0_{k,l} + i\mathbf{E}^0_{k,l}\wedge V_{k,l}\right) + \mathcal{X}_{k,l+p}\left(\mathbf{E}^0_{k,l} - i\mathbf{E}^0_{k,l}\wedge V_{k,l}\right) \\ (\mathbf{H}_h)_{|\Omega_k} &= i\sqrt{\varepsilon_k}\sum_{l=1}^{p} \mathcal{X}_{k,l}\left(\mathbf{E}^0_{k,l} + i\mathbf{E}^0_{k,l}\wedge V_{k,l}\right) + \mathcal{X}_{k,l+p}\left(\mathbf{E}^0_{k,l} - i\mathbf{E}^0_{k,l}\wedge V_{k,l}\right)\ . \end{aligned} \tag{III.B.60}$$

La formule (III.B.60) est la formule qui permet le calcul du courant total par le mode d'approximation ¡¡ H ¿¿, définition 24 du chapitre II.10.

III.B.2.2 Calcul des traces tangentielles.

Nous avons déjà construit les champs électrique et magnétique sur le plan théorique section II.8.1.4. En pratique, les relations obtenues donnent les quatre formules suivantes.

III.B.2.2.1 Calcul des traces tangentielles intérieures.

Sur $\Sigma_{k,j\neq k}$, on a

$$(\mathbf{E}_h)_{|\Sigma_{k,j}}\wedge\nu_{k,j} = \frac{1}{2\sqrt{\varepsilon_{kj}}}\left(\sum_{l=1}^{2p}\mathcal{X}_{k,l}\mathcal{Z}_{k,l} - \sum_{m=1}^{2p}\mathcal{X}_{j,m}\mathcal{Z}_{j,m}\right) \tag{III.B.61}$$

et

(III.B.62) $$\big((\mathbf{H}_h)_{|\Sigma_{k,j}} \wedge \nu_{k,j}\big) \wedge \nu_{k,j} = \frac{1}{2\sqrt{\mu_{kj}}}(\sum_{l=1}^{2p} \mathcal{X}_{k,l}\mathcal{Z}_{k,l} + \sum_{m=1}^{2p} \mathcal{X}_{j,m}\mathcal{Z}_{j,m}) \ .$$

III.B.2.2.2 Calcul des traces tangentielles au bord.

Sur $\Sigma_{k,j\neq k}$, on a, d'après (II.8.19),

(III.B.63) $$(\mathbf{E}_h)_{|\Sigma_{k,k}} \wedge \nu_{k,k} = \frac{1}{2\sqrt{\varepsilon_{kk}}}(-g + (1 - Q_k)\sum_{l=1}^{2p} \mathcal{X}_{kl}\mathcal{Z}_{k,l})$$

et

(III.B.64) $$\big((\mathbf{H}_h)_{|\Sigma_{k,k}} \wedge \nu_{k,k}\big) \wedge \nu_{k,k} = \frac{1}{2\sqrt{\mu_{kk}}}(g + (1 + Q_k)\sum_{l=1}^{2p} \mathcal{X}_{kl}\mathcal{Z}_{k,l})$$

Avec, par exemple, g donné par (III.B.51) :

$$g_{|\Sigma_{kk}} = g_k^0 e^{i\omega\big((\sqrt{\varepsilon_k\mu_k}\mathbf{k}_0)\,.\,\mathbf{X}\big)}$$
$$g_k^0 = -(1 + Q_k)\sqrt{\varepsilon_{kk}}\mathbf{E}_0 \wedge \nu_{k,k} - (1 - Q_k)\sqrt{\mu_{kk}}\sqrt{\frac{\varepsilon_k}{\mu_k}}\,(\mathbf{E}_0 \wedge \mathbf{k}_0 \wedge \nu_{k,k}) \wedge \nu_{k,k} \ .$$

Les formules (III.B.62) et (III.B.64) sont les formules permettant le calcul du courant total par le mode d'approximation ¡¡ G ¿¿, définition 23 du chapitre II.10.

III.B.2.3 Calcul des traces sur le maillage.

III.B.2.3.1 Calcul des traces normales par dérivation des traces tangentielles sur les faces du maillage.

En supposant que $\mathbf{E}_h$ et $\mathbf{H}_h$ vérifient les équations (III.B.59 p. 184), on a

(III.B.65) $$\begin{cases} \mathbf{H}_h = \dfrac{\nabla\wedge \mathbf{E}_h + \mathbf{m}}{i\omega\mu} \\ \mathbf{E}_h = -\dfrac{\nabla\wedge \mathbf{H}_h - \mathbf{j}}{i\omega\varepsilon} \ . \end{cases}$$

Dans un repère orthonormé direct local à une face Σ_{kj} orientée par la normale, on aura

(III.B.66) $$\begin{cases} (\mathbf{H}_h \,.\, \nu_{k,j}) = \dfrac{\dfrac{\partial \mathbf{E}_y^h}{\partial x} - \dfrac{\partial \mathbf{E}_x^h}{\partial y} + \mathbf{m}\nu_{k,j}}{i\omega\mu_k} \\ (\mathbf{E}_h \,.\, \nu_{k,j}) = \dfrac{+\mathbf{j}\nu_{k,j} - \dfrac{\partial \mathbf{H}_y^h}{\partial x} + \dfrac{\partial \mathbf{H}_x^h}{\partial y}}{i\omega\varepsilon_k} \ . \end{cases}$$

Or, pour tout vecteur $\mathbf{E}$ on a $(\mathbf{E}\wedge\nu)_x = +\mathbf{E}_y$ et $(\mathbf{E}\wedge\nu)_y = -\mathbf{E}_x$, donc dans un repère orthonormé direct orthogonal à la normale sortante $\nu_{k,j}$, on a

$$\frac{\partial \mathbf{E}_y^h}{\partial x} - \frac{\partial \mathbf{E}_x^h}{\partial y} = +(\frac{\partial(\mathbf{E}_h \wedge \nu_{k,j})_x}{\partial x} - \frac{\partial(\mathbf{E}_h \wedge \nu_{k,j})_y}{\partial y}) = \nabla\cdot_{\Sigma_{kj}} (\mathbf{E}_h \wedge \nu_{k,j})$$

puisque la composante de $\mathbf{E}_h \wedge \nu_{k,j}$ selon le troisième axe est nulle dans le repère choisi. D'autre part, remarquons que la divergence ne dépend pas du repère orthonormé direct choisi ; on peut donc écrire :

(III.B.67) $$(\mathbf{H}_h \,.\, \nu_{k,j}) = \frac{\nabla\cdot_{\Sigma_{kj}} (\mathbf{E}_h \wedge \nu_{k,j}) + \mathbf{m}\nu_{k,j}}{i\omega\mu_k}\nu_{k,j}$$

et, comme $(\mathbf{H}_h \wedge \nu_{k,j}) \wedge \nu_{k,j} \wedge \nu_{k,j} = -\mathbf{H}_h \wedge \nu_{k,j}$, on a aussi

$$(\mathbf{E}_h \,.\, \nu_{k,j}) = \frac{+\mathbf{j}\nu_{k,j} - \nabla.{}_{\Sigma_{kj}} \left(((\mathbf{H}_h \wedge \nu_{k,j}) \wedge \nu_{k,j}) \wedge \nu_{k,j}\right)}{i\omega\varepsilon_k}\nu_{k,j} \;. \tag{III.B.68}$$

Remarque 50 Une telle approximation exige le calcul d'une fonction dérivée, ce qui fait baisser l'ordre d'approximation par rapport au calcul des traces tangentielles.

Remarque 51 Pour tout vecteur $\mathbf{E}$, on a $\mathbf{E} = (\mathbf{E}\,.\,\nu)\,\nu - \mathbf{E}\wedge\nu$, ce qui montre que l'on peut reconstruire les champs $\mathbf{E}$ et $\mathbf{H}$ sur les arêtes du maillage. On a

$$\mathbf{H}_h = \frac{\nabla.{}_{\Sigma_{kj}} (\mathbf{E}_h \wedge \nu_{k,j}) + (\mathbf{m}\,.\,\nu_{k,j})}{i\omega\mu_k}\nu_{k,j} - ((\mathbf{H}_h \wedge \nu_{k,j}) \wedge \nu_{k,j}) \tag{III.B.69}$$

et

$$\mathbf{E}_h = \frac{+(\mathbf{j}\,.\,\nu_{k,j}) - \nabla.{}_{\Sigma_{kj}} \left(((\mathbf{H}_h \wedge \nu_{k,j}) \wedge \nu_{k,j}) \wedge \nu_{k,j}\right)}{i\omega\varepsilon_k}\nu_{k,j} - \mathbf{E}_h \wedge \nu_{k,j} \;. \tag{III.B.70}$$

III.B.2.3.2 Un calcul possible des champs électrique et magnétique sur les arêtes du maillage.

Nous présentons des formules de calcul des traces des champs sur les arêtes du maillage, formules qui présentent l'avantage d'être facilement calculables informatiquement et qui sont une alternative au calcul précédent (formules (III.B.67) et (III.B.68)) qui utilisait l'opérateur de divergence surfacique.

En effet, on peut approcher les traces de $\mathbf{E}$ et $\mathbf{H}$ sur une arête quelconque du maillage par une combinaison linéaire des traces tangentielles sur les faces contenant cette arête.

Supposons connues les traces tangentielles suivantes sur les faces Σ_{k,j_2} et Σ_{k,j_2} :

$$\begin{array}{ll} (\mathbf{E}_h)_{|\Sigma_{k,j_1}} \wedge \nu_{k,j_1} & (\mathbf{E}_h)_{|\Sigma_{k,j_2}} \wedge \nu_{k,j_2} \\ \left((\mathbf{H}_h)_{|\Sigma_{k,j_1}} \wedge \nu_{k,j_1}\right) \wedge \nu_{k,j_1} & \left((\mathbf{H}_h)_{|\Sigma_{k,j_2}} \wedge \nu_{k,j_2}\right) \wedge \nu_{k,j_2} \;. \end{array}$$

En supposant la continuité de $\mathbf{E}_h$ entre Σ_{k,j_2} et Σ_{k,j_1}, on peut calculer

$$\begin{array}{c} \left((\mathbf{E}_h)_{|\Sigma_{k,j_1}} \wedge \nu_{k,j_1}\right) \wedge \nu_{k,j_1} \\ ((\mathbf{E}_h)_{|\Sigma_{k,j_1}} \wedge \nu_{k,j_2}) \wedge \nu_{k,j_2} \;, \end{array}$$

or, on remarque que l'on a trivialement,

$$\begin{aligned} &(\mathbf{E}_h \wedge \nu_{k,j_2}) \wedge \nu_{k,j_1} = (\mathbf{E}_h\, \nu_{k,j_1})\nu_{k,j_2} - (\nu_{k,j_1}\, \nu_{k,j_2})\mathbf{E}_h \\ &\mathbf{E}_h = (\mathbf{E}_h\, \nu_{k,j_1})\nu_{k,j_1} - (\mathbf{E}_h \wedge \nu_{k,j_1}) \wedge \nu_{k,j_1} \;, \end{aligned}$$

donc, en remplaçant dans la première équation ci-dessus $\mathbf{E}_h$ par sa valeur donnée dans la deuxième, on a

$$(\mathbf{E}_h \wedge \nu_{k,j_2}) \wedge \nu_{k,j_1} = (\mathbf{E}_h\, \nu_{k,j_1})(\nu_{k,j_2} - (\nu_{k,j_1}\, \nu_{k,j_2})\nu_{k,j_1}) + (\nu_{k,j_1}\, \nu_{k,j_2})\,(\mathbf{E}_h \wedge \nu_{k,j_1}) \wedge \nu_{k,j_1} \;,$$

et en projetant sur ν_{k,j_2} :

$$(\mathbf{E}_h\, \nu_{k,j_1}) = \nu_{k,j_2} \frac{((\mathbf{E}_h \wedge \nu_{k,j_2}) \wedge \nu_{k,j_1} - (\nu_{k,j_1}\,.\,\nu_{k,j_2})\,(\mathbf{E}_h \wedge \nu_{k,j_1}) \wedge \nu_{k,j_1})}{1 - (\nu_{k,j_1}\,.\,\nu_{k,j_2})^2} \;.$$

Finalement, on peut calculer $(\mathbf{E}_h)_{|\Sigma_{k,j_1}\cap\Sigma_{k,j_2}}$ par

$$\mathbf{E}_h = \frac{(\nu_{k,j_1}\,.\,\nu_{k,j_2})\,((\mathbf{E}_h \wedge \nu_{k,j_2}) \wedge \nu_{k,j_1}) - ((\mathbf{E}_h \wedge \nu_{k,j_1}) \wedge \nu_{k,j_1})}{1 - (\nu_{k,j_1}\,.\,\nu_{k,j_2})^2} \;, \tag{III.B.71}$$

ou par

$$\mathbf{E}_h = \frac{\mathbf{E}_h \wedge ((\nu_{k,j_1} \wedge \nu_{k,j_2}) \wedge \nu_{k,j_2}) \wedge \nu_{k,j_1}}{|\nu_{k,j_1} \wedge \nu_{k,j_2}|^2} \;. \tag{III.B.72}$$

Remarque 52 L'intérêt de ce calcul est de ne pas effectuer une dérivation supplémentaire pour le calcul des traces normales. Remarquons que ce calcul est toujours possible étant donné que $\nu_{k,j_1} = \nu_{k,j_2}$ correspond à un tétraèdre qui aurait deux faces contigües dans le même plan, alors qu'un tétraèdre est toujours convexe. Avec d'autres éléments, il faudrait imposer leur convexité pour que la formule (III.B.71) soit toujours applicable. En termes de précision, éviter la dérivation est une bonne chose, mais nous remplaçons ce calcul par un calcul de trace aux bords des faces, ce qui nous fait aussi perdre en termes d'ordre de convergence.

Remarque 53 On peut effectuer le calcul des valeurs aux nœuds par la somme

$$\begin{aligned}3\mathbf{E}_h =&\frac{(\mathbf{E}_h)_{|\Sigma_{k,j_1}} \wedge ((\nu_{k,j_1} \wedge \nu_{k,j_2}) \wedge \nu_{k,j_2}) \wedge \nu_{k,j_1}}{|\nu_{k,j_1} \wedge \nu_{k,j_2}|^2} \\ &+ \frac{(\mathbf{E}_h)_{|\Sigma_{k,j_2}} \wedge ((\nu_{k,j_2} \wedge \nu_{k,j_3}) \wedge \nu_{k,j_3}) \wedge \nu_{k,j_2}}{|\nu_{k,j_2} \wedge \nu_{k,j_3}|^2} \\ &+ \frac{(\mathbf{E}_h)_{|\Sigma_{k,j_3}} \wedge ((\nu_{k,j_3} \wedge \nu_{k,j_1}) \wedge \nu_{k,j_1}) \wedge \nu_{k,j_3}}{|\nu_{k,j_3} \wedge \nu_{k,j_1}|^2}\end{aligned} \tag{III.B.73}$$

où j_1, j_2 et j_3 sont les indices des trois éléments voisins de Ω_k et ayant le nœud considéré - pour un nœud interne au maillage.

III.B.3 Calcul d'erreur pour le code Maxwell tridimensionnel.

Le but de cette section est de montrer que le calcul d'erreur dans le cas où la solution exacte est une onde plane présente la même difficulté qu'un assemblage de matrice.

Nous présentons ici le calcul de la norme d'erreur de bord

$$||(\mathcal{X}_h - \mathcal{X})||_{L^2(\Gamma=\partial\Omega)}$$

dans le cas où la quantité exacte $\mathcal{X}$ est connue sous la forme

$$\mathcal{X} = +\mathbf{E}_0 \wedge \nu_k + (\mathbf{H}_0 \wedge \nu_k) \wedge \nu_k \ .$$

La quantité approchée est calculée par

$$\mathcal{X}_h = \sum_{l=1}^{p} \mathcal{X}^{\mathbf{F}}_{k,l} \mathcal{Z}^{\mathbf{F}}_{k,l} + \mathcal{X}^{\mathbf{G}}_{k,l} \mathcal{Z}^{\mathbf{G}}_{k,l} \ .$$

On calcule donc la norme de l'erreur par

$$||\mathcal{X}_h - \mathcal{X}||^2_{L^2(\Gamma)} = \sum_k \int_{\Sigma_{k,k}} |\mathcal{X}_h|^2 + \sum_k \int_{\Sigma_{k,k}} |\mathcal{X}|^2 - 2\sum_k \int_{\Sigma_{k,k}} \Re(\mathcal{X}_h \overline{\mathcal{X}})$$

On pose

$$\begin{aligned}T_1^k &= \int_{\Sigma_{k,k}} |\sum_{l=1}^{p} \mathcal{X}^{\mathbf{F}}_{k,l} \mathcal{Z}^{\mathbf{F}}_{k,l} + \mathcal{X}^{\mathbf{G}}_{k,l} \mathcal{Z}^{\mathbf{G}}_{k,l}|^2 \\ T_2^k &= \Re \int_{\Sigma_{k,k}} \sum_{l=1}^{p} \left(\mathcal{X}^{\mathbf{F}}_{k,l} \mathcal{Z}^{\mathbf{F}}_{k,l} + \mathcal{X}^{\mathbf{G}}_{k,l} \mathcal{Z}^{\mathbf{G}}_{k,l}\right) \overline{\mathcal{X}} \\ T_3^k &= \int_{\Sigma_{k,k}} |\mathcal{X}|^2\end{aligned}$$

Calcul de T_1^k. Dans le vide, on calcule que

$$\begin{aligned}T_1^k =& \sum_{l=1}^{p} |\mathcal{X}^{\mathbf{F}}_{k,l}|^2 (D_k^{l,l}) + |\mathcal{X}^{\mathbf{G}}_{k,l}|^2 (D_k^{l,l}) \\ &+ 2\Re \sum_{m>l}^{p} \mathcal{X}^{\mathbf{F}}_{k,m} (D_k^{l,m}) \overline{\mathcal{X}^{\mathbf{F}}_{k,l}} + \mathcal{X}^{\mathbf{G}}_{k,m} (D_k^{l,m}) \overline{\mathcal{X}^{\mathbf{G}}_{k,l}}\end{aligned}$$

Calcul de T_2^k. On remarque que ce terme est calculé de la même façon que le second membre du couplage des fonctions de base avec la fonction de bord g. On note abusivement, (puisque $\mathcal{X}$ n'est pas égal à g, la condition sur le bord extérieur d'un calcul de diffraction en champ total)

$$b_k^l = \int_{\Sigma_{k,k}} \mathcal{X}\overline{\mathcal{Z}_{k,l}} \ ,$$

et l'on a

$$T_2^k = \Re \sum_{l=1}^{p} \mathcal{X}_{k,l}^{\mathbf{F}} (b_k^l)_{\mathbf{F}} + \mathcal{X}_{k,l}^{\mathbf{G}} (b_k^l)_{\mathbf{G}} \ .$$

Annexe III.C

Performances des codes Helmholtz et Maxwell.

Nous proposons d'étudier les performances des programmes Helmholtz bidimensionnel sans coefficient et Maxwell tridimensionnel avec des caractéristiques scalaires complexes ε et μ.

Nous avons retenu deux indices de performance :

- le taux de "vectorisation", exprimé en Méga-Flops, dont nous expliquerons la signification,
- la lisibilité des programmes. En effet, il n'est pas difficile de créer un code pour une application spécifique définie au début. En revanche, il est plus difficile d'écrire un code lisible par n'importe qui comme doit l'être un rapport technique. De plus, lors de modifications futures éventuelles, par exemple pour répondre à un nouveau cahier des charges, il est nécessaire que le code soit souple, ou que les difficultés apparaissent clairement. Ainsi, certaines "ruses" de programmation, appréciables pour une application donnée, peuvent donner lieu à des erreurs inattendues lorsque ces ruses ne sont pas clairement commentées de façon à être comprises immédiatement par un autre programmeur. Nous allons expliquer brièvement comment nous avons programmé systématiquement toutes les étapes du code, et expliquer l'utilisation d'un outil de visualisation nouveau des programmes fortran en LaTeX, outil que nous avons créé.

III.C.1 Définition de l'indice de performance.

L'efficacité numérique de la méthode et la qualité de la programmation sur un ordinateur à architecture vectorielle (se reporter à la section III.D.3) se mesurent, entre autres, en termes de Méga-Flops (MFlops). Un MFlop est un million d'opérations en virgule flottante, multiplications ou additions, pour des nombres réels en simple précision. Ce nombre tient aussi compte des opérations de conversion de types. Notons que les mesures ont été réalisées sur CRAY YMP qui est un ordinateur à architecture vectorielle dont la simple précision s'effectue sur 64 bits (et non 32 comme c'est généralement le cas sur les ordinateurs actuels). Pour donner une idée de la performance que l'on peut atteindre en pratique, voici un exemple.

```
program test
parameter (N=10000)
real a(N),b(N),c(N),d(N)
```

Boucle 1. Initialisation.

```
do i=1,N
  a(i)=real(i)
  b(i)=real(i)
  c(i)=real(i)
end do
```

Boucle 2. Calcul.

```
do i=1,N
```

```
    d(i)=a(i)+b(i)*c(i)
  end do

  end
```

La performance mesurée pour un processeur YMP donne le tableau III.C.1 suivant.

TAB. III.C.1 – Performances du programme test.

Boucle	**Temps (μs)**	**MFlops**
Initialisation	210	50
Calcul	122	165

III.C.2 Visualisation d'un programme fortran par LaTeX.

Le programme ¡¡ test ¿¿ est composé du texte suivant.

```
      program test
      parameter (N=10000)
      real a(N),b(N),c(N),d(N)
c \begin{description}
c \item[Boucle 1.] Initialisation.
      do i=1,N
        a(i)=real(i)
        b(i)=real(i)
        c(i)=real(i)
      end do
c \item[Boucle 2.] Calcul.
      do i=1,N
        d(i)=a(i)+b(i)*c(i)
      end do
c \end{description}
      end
```

Nous avons utilisé un programme de mise en forme du fortran en TEX, écrit par Van Jacobson (Lawrence Berkeley Laboratory) en 1985, qui permet notamment de mettre en gras les mots-clefs du fortran 77. Nous avons couplé ce premier "logiciel" à un script en C-Shell et à un programme en C pour visualiser des programmes en LaTeX, avec des commentaires codés[1].

Ceci permet de réaliser (enfin !) des listings écrits en français, structurés naturellement comme une note technique de code, avec des explications qui peuvent faire appel à des formules, des schémas, des tableaux et surtout inclure d'autres documents.

Les variables utilisées dans plusieurs sections du programme sont expliquées dans des fichiers séparés que le programme de visualisation recherche et inclut dans le document. A la deuxième apparition de la même variable, on ne donne plus que la référence de la définition.

Ceci assure la cohérence des définitions - à la différence de commentaires écrits en ligne dans les fichiers - de par l'unicité de la définition.

Ceci permet de travailler dans des fichiers de taille raisonnable mais largement commentés sans être encombré par ces commentaires.

On peut aussi demander la création d'un index des variables définies. On effectue automatiquement la bibliographie.

[1] Je tiens à remercier tout particulièrement Guilhem Chevalier et Nicolas L'Hullier sans qui cette adaptation n'aurait pu se faire.

Ce programme de visualisation nous semble particulièrement adapté à l'impression de listings de codes de très grosse taille et à la relecture sur écran de routines pour amélioration ou correction, le temps de visualisation étant donné par le temps de la compilation par le processeur LaTeX. Ainsi, alors que le nombre total de lignes écrites dans les fichiers fortran du code $\mathcal{L}ior$ est 43 225 pour 20 744 lignes d'instructions fortran, le nombre total de lignes de commentaires est 33 808 lorsque tout le code est visualisé en même temps (puisque beaucoup de commentaires sont effectués dans des fichiers séparés des fichiers fortran). Les fichiers fortran visualisés séparément, le nombre total de commentaires est constitué de 65 824 lignes : la visualisation globale du code s'effectue sans redite des mêmes commentaires (au contraire de la plupart des codes). On obtient une moyenne de 61% de lignes de commentaires pour tout le programme et 76% de lignes de commentaires en moyenne en visualisant les fichiers fortran un à un.

III.C.3 Liste des tâches effectuées par les deux programmes, Helmholtz et Maxwell.

Les deux programmes effectuent les tâches numerotées dans le tableau III.C.2.

Tab. III.C.2 – Liste des tâches.

Numéro	Tâche
1	Calcul des normales.
2	Lecture des données, initialisation, divers.
3	Construction des polarisations et directions de propagation.
4	Assemblage du second membre.
5	Assemblage des matrices du système linéaire.
6	Inversion de la matrice D.
7	Calcul de $D^{-1}C$ et $D^{-1}b$.
8	Calculs de post-traitements.
9	Ecritures des sorties.
10	Itération de l'algorithme itératif.

III.C.4 Code Helmholtz bidimensionnel dans le vide.

Sur le cas de la section III.C.4, on donne les temps de calcul des parties essentielles du programme. Nous remarquerons que le code bidimensionnel n'est pas parfaitement optimisé. La performance globale du programme est de l'ordre de 110 MFlops[1]. Le temps de calcul du code dépend donc par ordre décroissant

Tab. III.C.3 – Performances du code Helmholtz $2D$.

Tâche	Temps (s)	MFlops
10	4.15	143.7
2	1.15	0.1
9	0.331	0.2
5	0.43	125.1
6	0.186	15.6
7	0.0564	123.7
8	0.0316	98.9

des procédures suivantes.

[1]Le code expérimental Helmholtz bidimensionnel dans le vide effectue aussi toutes sortes de calculs de normes et de vérifications qui nuisent à son efficacité globale.

i) L'algorithme itératif (I.2.31) qui consiste à effectuer des produits matrice-vecteur. En pratique, un nombre d'itérations de 600 s'est révélé suffisant.

ii) Calculer les matrices D et C.

iii) Inverser les matrices D_k (de petite taille, inférieure, en pratique, à 15×15).

III.C.5 Code Maxwell tridimensionnel avec ou sans matériau.

En moyenne, le code $\mathcal{L}ior$ effectue 150 MFlops sur tout le programme. Si l'on élimine les entrées sorties, il fait 170 MFlops : le code $\mathcal{L}ior$ est donc à notre avis très bien optimisé pour un ordinateur à architecture vectorielle.

Nous donnons (tableau III.C.4) une estimation du nombre de MFlops atteints en moyenne par le programme sur un cas réaliste. Les temps permettent d'estimer les importances relatives des différentes tâches. Remarquons que certaines tâches sont effectuées en plusieurs parties, nous n'avons pas détaillé ces sous-actions.

Pour une compréhension plus détaillée de l'optimisation du code $\mathcal{L}ior$, le lecteur peut se reporter au chapitre (II.10) qui étudie (entre autres) plus précisément les performances du code selon les caractéristiques de V_h l'espace de discrétisation. Les performances atteintes dépassent souvent, toujours en

TAB. III.C.4 – Performances du code $\mathcal{L}ior$.

Tâche	**Temps (s)**	**MFlops**
10	$11,6$	170
2	$1,2$	0.5
5	$1,0$	$170 - 190$
9	$0,5$	1.2
3	$0,31$	80
6	$0,13$	160
7	$0,13$	$160 - 170$
8	$0,02$	150
1	$0,007$	150

termes de Méga-Flops, celles de l'algorithme produit vecteur-vecteur du programme test section III.C.1. Par exemple, l'algorithme itératif de résolution du système linéaire effectue des produits vecteur-vecteur mais simultanément sur les termes issus des fonctions de base de type **F** et des fonctions de type **G**.

Le code $\mathcal{L}ior$ est programmé de manière à ce qu'il soit portable sur toute machine. Il a été programmé en respectant les règles de base de toute programmation, à savoir,

1. absence de "ruses" non portables entre différentes machines ou systèmes d'exploitation,
2. utilisation d'algorithmes astucieux pour l'optimisation sur machine vectorielle largement commentés et expliqués (en respectant la règle précédente),
3. séparation nette des algorithmes valables dans certains cas particuliers des algorithmes généraux,
4. cohérence parfaite des commentaires de par l'utilisation du programme de visualisation (cf section III.C.2),
5. construction des commentaires du programme de façon à ce que la compréhension de tout le programme puisse être effectuée à la simple lecture de ces commentaires, chaque tâche (ou routine) élémentaire du programme est expliquée. Il n'y a nul besoin de lire les commentaires d'une autre routine pour comprendre la routine en cours, sauf indication et référence claire dans des cas où l'adaptation est triviale.

Notons que la méthode est aussi intrinsèquement adaptée à la programmation parallèle.

Annexe III.D

Calcul des termes intégraux du système linéaire.

Dans ce chapitre nous nous intéressons au calcul de l'intégrale de la fonction

$$e^{i\mathbf{k}\mathbf{X}} \tag{III.D.1}$$

sur un segment $[x_1, x_2]$, un triangle $[x_1, x_2, x_3]$ et un tétraèdre $[x_1, x_2, x_3, x_4]$. Les fonctions intégrales sont respectivement I_1, I_2 et I_3. Le calcul de cette intégrale est en effet nécessaire pour formuler les termes du système linéaire, puisque les fonctions de base du problème variationnel sont des ondes planes, que ce soit pour le problème de Helmholtz ou pour le problème de Maxwell. Le calcul simple des formules est effectué dans la section III.D.1. Les trois fonctions I_1, I_2 et I_3 ne sont pas définies sur respectivement $\mathbb{C}$, $\mathbb{C}^2$ et $\mathbb{C}^3$ ce qui interdit d'inclure directement les formules analytiques de calcul dans un programme (pour ordinateur). En revanche, comme ces fonctions se prolongent par des fonctions continues (sur respectivement $\mathbb{C}$, $\mathbb{C}^2$ et $\mathbb{C}^3$), il est possible d'implémenter le calcul à l'aide de tests judicieux pour maximiser la précision du résultat. Ce travail, qui consiste notamment à effectuer les développements limités adéquats, est présenté section III.D.2. Le lecteur averti sait que cette recette de programmation limite les performances d'une machine à architecture vectorielle. Nous présentons section III.D.3 comment opérer la vectorisation à moindre coût. Ce travail consiste à utiliser certaines fonctions intrinsèques dont la vectorisation est complète sur le calculateur.

III.D.1 Formules analytiques des intégrales d'ondes planes.

Cette section calcule simplement les formules d'intégration des ondes planes sur un simplexe de $\mathbb{R}$ (section III.D.1.1), $\mathbb{R}^2$ (section III.D.1.2) et $\mathbb{R}^3$ (section III.D.1.3). Les calculs présentés utilisent les notions élémentaires d'intégration d'une fonction analytique, leur difficulté réside dans la longueur des expressions.

III.D.1.1 Formules d'intégration sur un segment : calcul de I_1.

Nous calculons l'intégrale de (III.D.1) sur le segment $[x_1, x_2]$, soit

$$I_1 = \int_{[x_1,x_2]} e^{i\mathbf{k}\mathbf{X}}\, d\,\mathbf{X} \tag{III.D.2}$$

On pose $\alpha = \mathbf{k}\,\frac{(\vec{x}_2-\vec{x}_1)}{2}$ et on note L la longueur du segment $[x_1, x_2]$ soit $L = |x_2 - x_1|$. On effectue le changement de variable

$$I_1 = L\int_0^1 e^{i\mathbf{k}(x_1+\theta(x_2-x_1))}\, d\theta\ ,$$

et l'intégrale I_1 est donnée par la fonction de α

$$I_1(\alpha) = \left| \begin{array}{ll} Le^{i\mathbf{k}x_1} & \text{si } \alpha = 0 \ , \\ Le^{i\mathbf{k}x_1}\dfrac{e^{2i\alpha}-1}{2i\alpha} & \text{sinon.} \end{array} \right.$$

On remarque que la fonction

$$\frac{e^{2i\alpha}-1}{2i\alpha} = e^{i\alpha}\frac{e^{i\alpha}-e^{-i\alpha}}{2i\alpha} = e^{i\alpha}\frac{\sin\alpha}{\alpha}$$

tend vers 1 lorsque α tend vers zéro. Donc, en prolongeant par continuité, on a finalement, $\forall \alpha \in \mathbb{C}$,

(III.D.3)
$$I_1(\alpha) = Le^{i\mathbf{k}\vec{G}}\frac{\sin\alpha}{\alpha}$$

où $\vec{G}$ est le barycentre du segment $[x_1, x_2]$.

III.D.1.2 Formules d'intégration sur un triangle : calcul de I_2.

Nous calculons I_2 l'intégrale de (III.D.1) sur le triangle $[x_1, x_2, x_3]$ de surface S et de barycentre $\vec{G}$:

(III.D.4)
$$I_2 = \int_{[x_1,x_2,x_3]} e^{i\mathbf{k}\mathbf{X}}\, d\,\mathbf{X} \ .$$

On définit les trois paramètres α, β et γ par $\alpha = \mathbf{k}\dfrac{(\vec{x}_3-\vec{x}_2)}{2}$, $\beta = \mathbf{k}\dfrac{(\vec{x}_1-\vec{x}_2)}{2}$ et $\gamma = \mathbf{k}\dfrac{(\vec{x}_3-\vec{x}_1)}{2}$. Remarquons que l'on a $\gamma = \alpha - \beta$.

On suppose que les trois paramètres α, β et γ sont non nuls et on effectue le changement de variable affine

$$x = x_1 + \theta(x_2 - x_1) + (1-\theta)\phi(x_3 - x_1) \ ,$$

dans (III.D.4). On calcule alors I_2, avec $Z_i = e^{(i\mathbf{k}\,\vec{x}_i)}$:

$$\begin{aligned}
I_2 &= \int\int_{[0,1]\times[0,1]} 2S(1-\theta)e^{i\mathbf{k}(\vec{x_1}+\theta(\vec{x_2}-\vec{x_1})+(1-\theta)\phi(\vec{x_3}-\vec{x_1}))}\, d\theta d\phi \\
&= 2SZ_1\int_{[0,1]}(1-\theta)e^{-2i\theta\beta}\frac{e^{+2i(1-\theta)\gamma}-1}{2i(1-\theta)\gamma}\, d\theta \\
&= 2SZ_1\int_{[0,1]}\frac{e^{2i\gamma}.e^{-2i\theta(\beta+\gamma)}-e^{-2i\theta\beta}}{2i\gamma}\, d\theta \\
&= 2SZ_1\frac{e^{2i\gamma}.\dfrac{e^{-2i\alpha}-1}{-2i\alpha}-\dfrac{e^{-2i\beta}-1}{-2i\beta}}{2i\gamma} \ .
\end{aligned}$$

Après simplification, on obtient

$$\begin{aligned}
I_2 &= 2SZ_1e^{i\gamma}\frac{e^{i\gamma}.e^{-i\alpha}\dfrac{e^{-i\alpha}-e^{i\alpha}}{-2i\alpha}-e^{-i\gamma}e^{-i\beta}\dfrac{e^{-i\beta}-e^{i\beta}}{-2i\beta}}{2i\gamma} \\
I_2 &= 2SZ_1e^{i\gamma}\frac{e^{-i\beta}\dfrac{\sin\alpha}{\alpha}-e^{-i\alpha}\dfrac{\sin\beta}{\beta}}{2i\gamma} = 2SZ_2\frac{e^{i\alpha}\dfrac{\sin\alpha}{\alpha}-e^{i\beta}\dfrac{\sin\beta}{\beta}}{2i(\alpha-\beta)} \ .
\end{aligned}$$

Finalement, l'intégrale I_2 apparaît comme la fonction de α et β suivante

(III.D.5)
$$I_2(\alpha,\beta) = 2Se^{i\mathbf{k}\vec{G}}e^{-\frac{2}{3}i(\alpha+\beta)}\left(\frac{e^{i\alpha}\dfrac{\sin\alpha}{\alpha}-e^{i\beta}\dfrac{\sin\beta}{\beta}}{2i(\alpha-\beta)}\right) \ ,$$

fonction que l'on sait continue en $(\alpha,\beta) \in \mathbb{C}^2$.

III.D.1.3 Formules de l'intégrale sur un tétraèdre : calcul de I_3.

Nous calculons I_3 l'intégrale de (III.D.1) sur le tétraèdre $[x_1, x_2, x_3, x_4]$ de volume V et de barycentre $\vec{G}$:

(III.D.6) $$I_3 = \int_{[x_1,x_2,x_3,x_4]} e^{i\mathbf{k}\mathbf{X}}\, d\,\mathbf{X}\ .$$

On pose, pour $n-1 = 0 \ldots 3$ modulo 4

$$Z_n = e^{i\mathbf{k}\vec{x}_n}$$

et

$$v_n = \frac{\mathbf{k}\vec{x}_n}{2}\ .$$

On effectue le changement de variable affine

$$\vec{x} = \vec{x}_1 + \theta(\vec{x}_2 - \vec{x}_1) + (1-\theta)\left(\phi(\vec{x}_3 - \vec{x}_1) + (1-\phi)\zeta(\vec{x}_4 - \vec{x}_1)\right)$$

dans (III.D.6), ce qui donne

$$I_3 = 6Ve^{i\mathbf{k}\vec{x}_1} \int_{[0,1]} (1-\theta)e^{i\theta\mathbf{k}(\vec{x}_2-\vec{x}_1)}\, d\theta \int\int_{[0,1]^2} (1-\theta)(1-\phi)e^{i(1-\theta)\mathbf{k}(\phi(\vec{x}_3-\vec{x}_1)+(1-\phi)\zeta(\vec{x}_4-\vec{x}_1))}\, d\zeta\, d\phi$$

$$I_3 = 6Ve^{i\mathbf{k}\vec{x}_1} \int_{[0,1]} (1-\theta)e^{2i\theta(v_2-v_1)} \frac{e^{-2i(1-\theta)(v_1-v_4)}.\dfrac{e^{-2i(1-\theta)(v_4-v_3)}-1}{-2i(v_4-v_3)} - \dfrac{e^{2i(1-\theta)(v_3-v_1)}-1}{2i(v_3-v_1)}}{-2i(v_1-v_4)}\, d\theta$$

$$I_3 = 6VZ_1e^{2i(v_2-v_1)} \int_{[0,1]} (1-\theta) \frac{\dfrac{e^{-2i(1-\theta)(v_2-v_3)} - e^{-2i(1-\theta)(v_2-v_4)}}{-2i(v_4-v_3)} - \dfrac{e^{2i(1-\theta)(v_3-v_2)} - e^{2i(1-\theta)(v_1-v_2)}}{2i(v_3-v_1)}}{-2i(v_1-v_4)}\, d\theta$$

$$I_3 = 6VZ_1e^{2i(v_2-v_1)} \frac{\dfrac{\dfrac{e^{-2i(v_2-v_3)}-1}{-2i(v_2-v_3)} - \dfrac{e^{-2i(v_2-v_4)}-1}{-2i(v_2-v_4)}}{-2i(v_4-v_3)} - \dfrac{\dfrac{e^{2i(v_3-v_2)}-1}{2i(v_3-v_2)} - \dfrac{e^{2i(v_1-v_2)}-1}{2i(v_1-v_2)}}{2i(v_3-v_1)}}{-2i(v_1-v_4)}\ .$$

Après simplification, on a

$$I_3 = 6VZ_2 \frac{\dfrac{e^{i(v_4-v_2)}\dfrac{\sin(v_4-v_2)}{(v_4-v_2)} - e^{i(v_3-v_2)}\dfrac{\sin(v_3-v_2)}{(v_3-v_2)}}{2i(v_4-v_3)} - \dfrac{e^{i(v_3-v_2)}\dfrac{\sin(v_3-v_2)}{(v_3-v_2)} - e^{i(v_1-v_2)}\dfrac{\sin(v_1-v_2)}{(v_1-v_2)}}{2i(v_3-v_1)}}{2i(v_4-v_1)}\ ,$$

ou

$$I_3 = 6V \frac{\dfrac{e^{i(v_4+v_2)}\dfrac{\sin(v_4-v_2)}{(v_4-v_2)} - e^{i(v_3+v_2)}\dfrac{\sin(v_3-v_2)}{(v_3-v_2)}}{2i(v_4-v_3)} - \dfrac{e^{i(v_3+v_2)}\dfrac{\sin(v_3-v_2)}{(v_3-v_2)} - e^{i(v_1+v_2)}\dfrac{\sin(v_1-v_2)}{(v_1-v_2)}}{2i(v_3-v_1)}}{2i(v_4-v_1)}\ .$$

En posant $\alpha_n = \mathbf{k}\dfrac{(\vec{x}_{n-1} - \vec{x}_n)}{2}$, $\beta_n = \mathbf{k}\dfrac{(\vec{x}_{n+1} - \vec{x}_n)}{2}$ et $\gamma_n = \mathbf{k}\dfrac{(\vec{x}_{n+2} - \vec{x}_n)}{2}$, pour tout $n-1$ de 0 à 3 modulo 4, on a

$$I_3 = 6VZ_2\frac{I(\gamma_2, \beta_2) - I(\beta_2, \alpha_2)}{2i(\gamma_2 - \alpha_2)}$$

avec

(III.D.7) $$I(\lambda, \mu) = \frac{e^{i\lambda}\dfrac{\sin\lambda}{\lambda} - e^{i\mu}\dfrac{\sin\mu}{\mu}}{2i(\lambda-\mu)}\ .$$

Finalement, par symétrie, l'intégrale I_3 apparaît comme la fonction de α_n, β_n et γ_n suivante

(III.D.8) $$I_3(\alpha_n, \beta_n, \gamma_n) = 6Ve^{i\mathbf{k}\vec{G}}e^{-\frac{1}{2}i(\alpha_n+\beta_n+\gamma_n)}\frac{I(\gamma_n, \beta_n) - I(\beta_n, \alpha_n)}{2i(\gamma_n - \alpha_n)}\ ,$$

fonction que l'on sait continue en $(\alpha_n, \beta_n, \gamma_n) \in \mathbb{C}^3$.

III.D.2 Algorithmes conditionnels de programmation.

Les intégrales $I_1(\alpha)$ (III.D.3), $I_2(\alpha,\beta)$ (III.D.5) et $I_3(\alpha,\beta,\gamma)$ (III.D.8) calculées section III.D.1 sont des fonctions qui ont un sens sur $\mathbb{C}$, $\mathbb{C}^2$ et $\mathbb{C}^3$ par prolongement du domaine de définition grâce à leur continuité. Mathématiquement, ces fonctions ne posent pas de problème. En revanche, sur le plan informatique, étendre le domaine de définition d'une fonction n'est pas aussi simple. Le programmeur est obligé de dissocier la fonction sur son domaine de définition et les valeurs limites prises autour des points qui sortent du domaine. Pour bien comprendre tout l'intérêt numérique de cette section, référons-nous tout de suite à l'exemple de la sous-section III.D.2.1 du calcul de I_1.

III.D.2.1 Programmation de I_1.

Rappelons que I_1 est donné par

$$I_1(\alpha) = Le^{i\mathbf{k}\vec{G}}\frac{\sin\alpha}{\alpha} \tag{III.D.9}$$

Cette fonction dont le domaine de définition est $\mathbb{C} - \{0\}$ admet une limite pour α tendant vers 0,

$$Le^{i\mathbf{k}\vec{G}} ,$$

ce qui permet de donner un sens à $I_1(0)$. En revanche, un ordinateur, quel qu'il soit, ne peut calculer $I_1(0)$ par la formule (III.D.9) puisque la division par zéro est interdite. La notion de limite, ou de prolongement par continuité, est une notion formelle que le calculateur ignore. En revanche, pour tous les autres décimaux (qui sont en nombre fini) connus par la machine, on pourra calculer la fonction I_1. En effet, il existe un nombre, que l'on note ϵ, qui est le plus petit réel (en fait décimal) strictement positif de la machine. Pour ce nombre, et tous les nombres dont la valeur absolue est supérieure, le calculateur peut effectuer la division. La fonction $\sin\alpha$ est approchée par son développement limité à un ordre qui dépend du calculateur. Par exemple, $\sin\alpha$ sera approché par la fonction polynômiale

$$\sin\alpha \approx \alpha - \frac{\alpha^3}{6} .$$

Pour $\alpha = \epsilon$, le nombre $\epsilon^3/6$ n'existe pas dans la machine puisqu'il est compris entre ϵ et 0 (pour $\epsilon \leq \sqrt{6}$, ce qui est évidemment le cas, le contraire ne se conçoit pas). Pour ϵ assez petit, le calculateur remplacera $\epsilon^3/6$ par sa valeur approchée la plus proche, soit 0. On aura donc[1]

$$\sin\epsilon = \epsilon .$$

La division de ϵ par ϵ rend alors 1 le nombre entier naturel représenté par le décimal $1,00\ldots0$ ou en binaire par la mantisse $0\ldots01$ et l'exposant $0\ldots0$ (le nombre de zéros dépend de la précision du calculateur, ou mode de représentation des réels). Nous en tirons les conclusions suivantes :

– l'erreur effectuée par le calculateur sur la fonction sinus cardinale est donc de l'ordre de $\alpha^2/6$ donc inférieure à α et non quantifiable,
– il est inutile de calculer $I_1(\epsilon)$ par la formule (III.D.9) puisque $I_1(\epsilon) = I_1(0)$ pour le calculateur.

Cherchons à prolonger notre raisonnement sur le calcul de $I_1(\epsilon)$ à des nombres plus grands. La question est de savoir à partir de quel nombre ϵ_1 la calculateur fera la différence entre 1 et le sinus cardinal de ϵ_1. En informatique, on définit un nombre, que nous noterons p_m, appelé la "précision machine" qui est le plus grand réel positif tel que le test "le réel 1 plus la précision machine est égal au réel 1" soit vrai. On peut aussi le définir par le plus petit réel positif tel que le test "le réel 1 plus la précision machine est égal au réel 1" soit faux. La différence entre ces deux définitions est ϵ et évidemment (pour tout calculateur à exposant) ϵ est très petit devant p_m. Nous cherchons donc ϵ_1 le plus grand réel positif tel que

$$|I_1(\epsilon_1) - Le^{i\mathbf{k}\vec{G}}| \leq p_m \times Le^{i\mathbf{k}\vec{G}} .$$

[1] Si le calculateur approche $\sin\alpha$ à un ordre supérieur à celui de l'exemple, si ϵ est assez petit (ce qui en général va de pair), on aura toujours $\sin\epsilon = \epsilon$.

Toujours dans le cas où la fonction sinus est connue par son développement à l'ordre 4 (ou 5), ϵ_1 vérifie

$$\epsilon_1^2/6 = p_m \ .$$

Lorsque le calculateur approche la fonction sinus à un ordre plus élevé et que la précision machine est assez petite devant 1, on peut toujours approcher ϵ_1 par $\sqrt{6p_m}$. Par exemple, à l'ordre 6, on aura

$$\epsilon_1^2/6 + \epsilon_1^4/120 = p_m \ ,$$

approximativement vérifié par $\epsilon_1 = \sqrt{6p_m}$ à $0,3p_m^2$ (près), donc à $0,3p_m$ par rapport à p_m (en erreur relative). Cette différence (par définition de p_m) n'est pas représentée par la machine (donne le même nombre ϵ_1). Nous proposons donc l'algorithme de calcul suivant :

- définition de $\epsilon_1 = \sqrt{6p_m}$,
- pour tout $\alpha \in \mathbb{C}$,
  - si $|\alpha| \leq \epsilon_1$, alors $I_1(\alpha) = Le^{i\mathbf{k}\vec{G}}$,
  - sinon $I_1(\alpha) = Le^{i\mathbf{k}\vec{G}}\frac{\sin\alpha}{\alpha}$.

La fonction $I_1(\alpha)$ est ainsi calculée à la précision $p_m \times Le^{i\mathbf{k}\vec{G}}$ en un nombre minimal d'opérations. L'erreur relative sur I_1 est donc de l'ordre de la précision machine, précision maximale inhérente au calculateur.

III.D.2.2 Programmation de I_2.

La programmation de I_2 est beaucoup plus compliquée que celle de I_1, beaucoup de situations possibles sur les paramètres de calcul interviennent. Nous allons présenter au lecteur plusieurs algorithmes possibles de calcul et pour l'un deux un résumé de la marche suivie pour minimiser l'erreur.

Rappelons que I_2 est donné par

$$I_2(\alpha,\beta) = 2Se^{i\mathbf{k}\vec{G}}e^{-\frac{2}{3}i(\alpha+\beta)}\left(\frac{e^{i\alpha}\frac{\sin\alpha}{\alpha} - e^{i\beta}\frac{\sin\beta}{\beta}}{2i(\alpha-\beta)}\right) \tag{III.D.10}$$

Le domaine de définition de $I_2(\alpha,\beta)$ est $\mathbb{C}^2$ privé des droites $\alpha = 0$, $\beta = 0$ et $\alpha = \beta$. On prolonge cette fonction par continuité sur $\mathbb{C}^2$: la fonction prolongée est analytique sur $\mathbb{C}^2$. Comme nous l'avons expliqué pour le calcul de I_1 (III.D.9), il n'est pas possible de calculer I_2 directement sur le plan complexe puisque la division par zéro est interdite sur un calculateur. Nous allons suivre la même démarche que pour le calcul de I_1. On remarque que si l'on pose

$$f(\alpha) = e^{i\alpha}\frac{\sin\alpha}{\alpha}$$

on calcule I_2 par

$$I_2(\alpha,\beta) = 2Se^{i\mathbf{k}\vec{G}}e^{-\frac{2}{3}i(\alpha+\beta)}\left(\frac{f(\alpha)-f(\beta)}{2i(\alpha-\beta)}\right) \ .$$

Nous savons maintenant calculer la fonction $f(\alpha)$ sur $\mathbb{C}$ puisque c'est le même problème que celui du calcul de $I_1(\alpha)$. Ceci implique que nous sommes capables de calculer $I_2(\alpha,\beta)$ sur les axes $\alpha = 0$ et $\beta = 0$. Le problème nouveau de notre étude est donc de calculer la fonction I_2 dans un voisinage de la droite $\alpha = \beta$ à la meilleure précision possible et avec le moins d'opérations possible. Le critère supplémentaire à considérer est donc $\gamma = \alpha - \beta$ proche de zéro ou pas, et dans quel sens. En effet, on a,

$$I_2 = \tilde{I}_2(\alpha,\gamma) = 2Se^{i\mathbf{k}\vec{G}}e^{-\frac{2}{3}i(2\alpha-\gamma)}\left(\frac{f(\alpha)-f(\alpha-\gamma)}{2i(\gamma)}\right) \ .$$

Lorsque $\gamma \to 0$, la fonction $\tilde{I}_2(\alpha,\gamma)$ tend vers

$$-iSe^{i\mathbf{k}\vec{G}}e^{-\frac{4}{3}i\alpha}f'(\alpha)$$

au premier ordre, avec

$$f'(\alpha) = ie^{i\alpha}\frac{\sin\alpha}{\alpha} + e^{i\alpha}\left(\frac{\alpha\cos\alpha - \sin\alpha}{\alpha^2}\right) \ .$$

La fonction dérivée $f'(\alpha)$ admet une limite en zéro. Nous pouvons donc proposer l'algorithme suivant.

- Si $|\gamma| \leq \epsilon_2$, un réel positif à définir que l'on suppose supérieur à ϵ_1, alors
  - si $|\alpha| \leq \epsilon_3$, un autre réel positif à définir, alors on sait que $f'(\alpha) = i$ et l'on a directement $I_2 = Se^{i\mathbf{k}\vec{G}}$,
  - sinon, on calcule

$$I_2 = -iSe^{i\mathbf{k}\vec{G}}e^{-\frac{4}{3}i\alpha}ie^{i\alpha}\frac{\sin\alpha}{\alpha} + e^{i\alpha}\left(\frac{\alpha\cos\alpha - \sin\alpha}{\alpha^2}\right) .$$

- Si $|\gamma| > \epsilon_2$, alors
  - si $|\alpha| \leq \epsilon_1$, (d'où $\beta > \epsilon_1$) alors

$$I_2 = 2Se^{i\mathbf{k}\vec{G}}e^{-\frac{2}{3}i\beta}(\frac{1 - f(\beta)}{2i(\alpha - \beta)}) ,$$

  - de même si l'on inverse les rôles de α et β,
  - sinon, on calcule I_2 par la formule (III.D.10).

Nous laissons au lecteur le soin de déterminer ϵ_2 et ϵ_3 de façon à minimiser l'erreur globale du calcul. Pour cela, il faudra effectuer des estimations pointues des restes des fonctions approchant I_2.

Nous avons préféré utiliser un autre algorithme qui exploite les propriétés particulières de la fonction $I_2(\alpha, \beta)$ issues de considérations géométriques. En effet, rappelons que l'on a défini $\alpha = \mathbf{k}\frac{(\vec{x}_3 - \vec{x}_2)}{2}$, $\beta = \mathbf{k}\frac{(\vec{x}_1 - \vec{x}_2)}{2}$ et $\gamma = \mathbf{k}\frac{(\vec{x}_3 - \vec{x}_1)}{2}$. Si on remplace α par $-\beta$ et β par γ, on sait que

$$I_2 = I_2(\beta, -\gamma) = 2Se^{i\mathbf{k}\vec{G}}e^{-\frac{2}{3}i(\gamma-\beta)}\frac{e^{i\gamma}\frac{\sin\gamma}{\gamma} - e^{-i\beta}\frac{\sin\beta}{\beta}}{2i(\beta + \gamma)}) .$$

La fonction $I_2(\beta, -\gamma)$ est une alternative au calcul de I_2 donné par la formule (III.D.10).

Nous pouvons donc développer l'algorithme plus simple suivant.

- Si $|\gamma| \leq \epsilon_2$, un réel positif à définir que l'on suppose supérieur à ϵ_1, alors nous avons le choix suivant.
  - Si $|\alpha| \geq \epsilon_2$, alors,
    - si $|\beta| \geq \epsilon_1$, alors,

$$I_2 = I_2(\beta, -\gamma) = 2Se^{i\mathbf{k}\vec{G}}e^{-\frac{2}{3}i(\gamma-\beta)}\frac{e^{i\gamma}\frac{\sin\gamma}{\gamma} - e^{-i\beta}\frac{\sin\beta}{\beta}}{2i\alpha}) ,$$

    - sinon,

$$I_2 = 2Se^{i\mathbf{k}\vec{G}}e^{-\frac{2}{3}i(\gamma)}\frac{e^{i\gamma}\frac{\sin\gamma}{\gamma} - 1}{2i\alpha}) .$$

  - Si $|\alpha| < \epsilon_2$, alors on effectue un développement de Taylor par rapport à α et β, et l'on peut calculer I_2 par

$$I_2 = 2Se^{i\mathbf{k}\vec{G}}e^{-\frac{2}{3}i(\gamma)}T_n(\alpha, \beta)$$

  où $T_n(\alpha, \beta)$ est le polynôme approchant le développement limité de la différence des deux sinus cardinaux de α et β à l'ordre $n-2$ par rapport aux deux variables : n est l'ordre du développement de Taylor à effectuer sur la fonction I_2. Par exemple,

$$\begin{aligned} T_7(\alpha, \beta) =& \frac{1}{2} + i\frac{(\alpha + \beta)}{3} - \frac{(\alpha^2 + +\beta^2)}{6} \\ & - i\frac{(\alpha^3 + (\alpha^2\beta) + (\alpha\beta^2) + \beta^3)}{15} \\ & + \frac{(\alpha^4 + (\alpha^3\beta) + (\alpha^2\beta^2) + (\alpha\beta^3) + \beta^4)}{45} . \end{aligned} \tag{III.D.11}$$

Notons que si l'on veut éviter le calcul et la multiplication par $e^{-\frac{2}{3}i(\gamma)}$, on peut aussi calculer I_2 par

$$I_2 = 2Se^{i\mathbf{k}\vec{G}} D_n(\alpha,\beta)$$

où $D_n(\alpha,\beta)$ est le polynôme approchant le développement limité de I_2 à l'ordre n par rapport aux deux variables α et β. Par exemple,

$$\begin{aligned} D_7(\alpha,\beta) =& \frac{1}{2} + \frac{\beta\alpha - \alpha^2 - \beta^2}{18} + i\frac{3\beta^2\alpha + 3\beta\alpha^2 - 2\alpha^3 - 2\beta^3}{405} \\ &+ \frac{\beta^4 + \alpha^4 - 2\beta^3\alpha - 2\beta\alpha^3 + 3\beta^2\alpha^2}{405} \end{aligned} \tag{III.D.12}$$

– Si $|\gamma| > \epsilon_2$, alors,
  – si $|\alpha| \leq \epsilon_1$, (d'où $\beta \geq \epsilon_2 - \epsilon_1$) alors

$$I_2 = 2Se^{i\mathbf{k}\vec{G}} e^{-\frac{2}{3}i\beta} \left(\frac{1 - e^{i\beta}\dfrac{\sin\beta}{\beta}}{2i(\alpha-\beta)} \right) ,$$

  – de même si l'on inverse les rôles de α et β,
  – sinon, on calcule I_2 par la formule (III.D.10).

Il nous reste à choisir ϵ_2 de façon à optimiser la précision du calcul de I_2 puis à minimiser le nombre d'opérations à effectuer. Il s'agit d'étudier les précisions des calculs de I_2 par les différentes formules présentées, puis d'obtenir les mêmes précisions sur tous les cas limites afin que la majoration globale de l'erreur sur le calcul de I_2 soit aussi faible que possible. Sous l'hypothèse $\epsilon_1 \leq \epsilon_2/2$, les différentes formules de calcul utilisent les trois fonctions suivantes :

1. la fonction $h(x,y)$

$$h(x,y) = \frac{e^{ix}\dfrac{\sin x}{x} - e^{iy}\dfrac{\sin y}{y}}{2i(x-y)} \tag{III.D.13}$$

définie sur $|x-y| > \epsilon_2$ et $|x| > \epsilon_1$ et $|y| > \epsilon_1$,

2. la fonction $g(x,y)$

$$g(x,y) = \frac{e^{ix}\dfrac{\sin x}{x} - 1}{2i(x-y)} \tag{III.D.14}$$

définie sur $|x-y| > \epsilon_2$ et $|y| \leq \epsilon_1$,

3. la fonction $d(x,y)$

$$\begin{aligned} d(x,y) =& \frac{1}{2} + i\frac{(x+y)}{3} - \frac{(x^2 + +y^2)}{6} \\ &- i\frac{(x^3 + (x^2y) + (xy^2) + y^3)}{15} \\ &+ \frac{(x^4 + (x^3y) + (x^2y^2) + (xy^3) + y^4)}{45} \end{aligned} \tag{III.D.15}$$

définie sur $|x-y| \leq \epsilon_2$ et $|y| \leq \epsilon_2$.

La fonction $d(x,y)$ permet d'approcher I_2 en effectuant une erreur de l'ordre du premier terme oublié dans la série de Taylor, terme de la forme

$$\frac{2i}{315}\left(x^5 + x^4y + x^3y^2 + x^2y^3 + xy^4 + y^5\right)$$

dont le module est maximal pour $y = \epsilon_2$ et $x = 2\epsilon_2$, et de valeur $\frac{2}{5}\epsilon_2^5$. Pour ϵ_2 assez petit, nous ferons l'hypothèse que l'erreur est donc $\frac{2}{5}\epsilon_2^5$. Par exemple, pour $\epsilon_2 = 1/10$, le terme suivant est 50 fois inférieur à ce terme.

Etudions comment se comporte la fonction $g(x,y)$ (III.D.14) autour du point limite $x=\epsilon_2$ et $y=0$. On a, puisque ϵ_2 est petit,

$$g(\epsilon_2,0)\approx\frac{1+i\epsilon_2-2/3\epsilon_2^2-1}{2i(\epsilon_2)}\ .$$

On constate que cette fonction a bien une limite informatique pour ϵ_2 petit non nul et que cette limite est $1/2$. En revanche, autour du point $x=-i\epsilon_2$ et $y=0$, on a

$$g(-i\epsilon_2,0)\approx\frac{1+\epsilon_2+2/3\epsilon_2^2-1}{2(\epsilon_2)}\ .$$

Pour ϵ_2 assez petit, l'ordinateur calcule $1-1$ qui est nul, soit une erreur de l'ordre de $1/2$, ce qui est inacceptable. L'erreur sur cette fonction de ϵ_2 est

$$p_m\times\frac{1}{\epsilon_2}\ .$$

La même analyse peut être effectuée sur la fonction $h(x,y)$ différence de deux fonctions évaluées à la précision machine et dont l'erreur sur la différence revient à étudier la fonction $g(x,y)$. Nous choisissons donc ϵ_2 de façon à avoir

$$\frac{2}{5}\epsilon_2^5=p_m\times\frac{1}{\epsilon_2}$$

soit,

$$\epsilon_2=\left(\frac{5}{2}p_m\right)^{1/6}$$

où l'on vérifie bien l'hypothèse $\epsilon_1<\epsilon_2/2$ à condition d'avoir $p_m\le 1/10$. L'erreur relative sur le calcul de I_2 est alors majorée par

$$p_m^{5/6}$$

qui est proche de la précision machine.

L'erreur effectuée sur le calcul de I_2 est donc faible, largement inférieure à toute méthode de calcul par intégration numérique (sur des points de Gauss par exemple).

III.D.2.3 Programmation de I_3.

Rappelons que I_3 (III.D.8) est donné par

$$I_3(\alpha,\beta,\gamma)=6Ve^{i\mathbf{k}\vec{G}}e^{-\frac{1}{2}i(\alpha+\beta+\gamma)}\frac{I(\beta,\gamma)-I(\beta,\alpha)}{2i(\gamma-\alpha)}\qquad\text{(III.D.16)}$$

en définissant les fonctions I et f par

$$\begin{cases}I(\lambda,\mu)=\dfrac{f(\lambda)-f(\mu)}{2i(\lambda-\mu)}\\[2mm] f(\lambda)=e^{i\lambda}\dfrac{\sin\lambda}{\lambda}\end{cases}\qquad\text{(III.D.17)}$$

ou en définissant $Z_n=e^{i\mathbf{k}\mathbf{X}_n}$ où $\mathbf{X}_n$ est la position du n-ième sommet du tétraèdre,

$$I_3=6VZ_n\frac{I(\beta,\gamma)-I(\beta,\alpha)}{2i(\gamma-\alpha)}\ .$$

La programmation de cette intégrale volumique est très compliquée. Nous donnons l'algorithme de calcul sans explication, cet algorithme étant dans la ligne logique des algorithmes présentés dans les sections III.D.2.1 et III.D.2.2.

Nous laissons au lecteur le soin de calculer ϵ_1 et ϵ_2 de façon à maximiser la précision des calculs quels que soient les paramètres de la fonction I_3.

1. Pour $|\alpha-\gamma| \geq \epsilon_2$:
   (a) Pour $|\beta-\alpha|$ et $|\beta-\gamma| \geq \epsilon_2$:
      i. Pour $|\alpha| \leq \epsilon_1$, on calcule alors I_3 en remplaçant $f(\alpha)$ par 1.
      ii. Pour $|\beta| \leq \epsilon_1$, on calcule alors I_3 en remplaçant $f(\beta)$ par 1.
      iii. Pour $|\gamma| \leq \epsilon_1$, on calcule alors I_3 en remplaçant $f(\gamma)$ par 1.
      iv. Pour $(\alpha,\beta,\gamma) \geq \epsilon_2$, on calcule I_3 simplement par la formule III.D.16.
   (b) Pour $|\beta-\alpha| \leq \epsilon_2$ ou (exclusif) $|\beta-\gamma| \leq \epsilon_2$, comme le problème est symétrique, on suppose : $|\beta-\alpha| \leq \epsilon_2$. On doit donc remplacer $I(\alpha,\beta)$ par un développement :
      i. Pour $|\alpha| \leq \epsilon_2$ et (ou) $|\beta| \leq \epsilon_2$, on calcule alors $I(\alpha,\beta)$ par son développement limité :

$$\text{(III.D.18)} \qquad \begin{cases} \dfrac{1}{2} + i\dfrac{\alpha+\beta}{3} - \dfrac{\alpha^2+\beta^2+\alpha\,\beta}{6} \\ -\,i\dfrac{\alpha^3+\beta^3+\alpha^2\beta+\alpha\,\beta^2}{15} \\ \dfrac{\alpha^4+\beta^4+\alpha^3\beta+\alpha^2\beta^2+\alpha\,\beta^3}{45} + i\mathrm{O}(\alpha,\beta)^5 \end{cases}$$

      ii. Pour $|\alpha|$ et $|\beta| \geq \epsilon_2$, on calcule alors $I(\alpha,\beta)$ par :

$$\text{(III.D.19)} \qquad \begin{aligned} Z_2 I(\alpha,\beta) &= Z_1 I(\beta-\alpha,-\alpha) \\ &= Z_1 I(\delta,-\alpha) \end{aligned}$$

      où le dénominateur de $I(\delta,-\alpha)$ est forcément assez grand. On remplace $f(\delta)$ par :

$$e^{i\delta}\frac{\sin\delta}{\delta} \approx e^{i\delta}(1-\frac{1}{6}\delta^2+\frac{1}{120}\delta^4)+\mathrm{O}(\delta)^6$$

2. Pour $|\alpha-\gamma| \leq \epsilon_2$:
   (a) Pour $|\alpha|$ et $|\gamma| \geq \epsilon_2$, on a

$$I_3 = VZ_3\frac{I(\beta_3,\gamma_3)-I(\beta_3,\alpha_3)}{2i(\gamma_3-\alpha_3)}\ ,$$

   avec $\alpha_3-\gamma_3=-\alpha$. On se ramène alors à l'étude précédente avec $|\alpha_3-\gamma_3| \geq \epsilon_2$.
   (b) Pour $|\alpha| \leq \epsilon_2$ et (ou) $|\gamma| \leq \epsilon_2$:
      i. Pour $|\beta| \leq \epsilon_2$, on effectue un développement limité complet de I_3 :

$$\text{(III.D.20)} \qquad \begin{cases} \dfrac{1}{6} + i\dfrac{\alpha+\beta+\gamma}{12} - \dfrac{\alpha^2+\beta^2+\gamma^2+\alpha\,\beta+\alpha\,\gamma+\beta\,\gamma}{30} \\ -\dfrac{i}{90}.(\alpha^3+\beta^3+\gamma^3+\alpha^2\beta+\alpha^2\gamma+\alpha\,\beta^2 \\ \quad +\alpha\,\beta\,\gamma+\alpha\,\gamma^2+\beta^2\gamma+\beta\,\gamma^2) \\ +\dfrac{1}{315}.(\alpha^4+\beta^4+\gamma^4+\alpha^3\beta+\alpha^3\gamma+\alpha^2\beta\,\gamma \\ \quad +\alpha^2\beta^2+\alpha^2\gamma^2+\alpha\,\beta^3+\alpha\,\beta^2\,\gamma \qquad + i\mathrm{O}(\alpha,\beta,\gamma)^5 \\ \quad +\alpha\,\beta\,\gamma^2+\alpha\,\gamma^3+\beta^3\gamma+\beta^2\gamma^2+\beta\,\gamma^3) \end{cases}$$

      ii. Pour $|\beta| \geq \epsilon_2$, on a

$$I_3 = VZ_4\frac{I(\beta_4,\gamma_4)-I(\beta_4,\alpha_4)}{2i(\gamma_4-\alpha_4)}$$

      avec $\alpha_4-\gamma_4=-\beta$. On se ramène alors à l'étude précédente avec $|\alpha_4-\gamma_4| \geq \epsilon_2$.

III.D.3 Recettes d'implémentation sur machine vectorielle.

La section III.D.2 a montré comment programmer les formules analytiques des intégrales I_1, I_2 et I_3 de la section III.D.1. La technique utilisée était simple, il s'agissait de tester les paramètres intervenant dans les calculs et selon leurs valeurs d'utiliser telle ou telle formule. Nous avons vu que nous pouvions ainsi maximiser la précision du calcul et minimiser le nombre d'opérations à effectuer, et ce pour n'importe quel calculateur.

Un des progrès, encore récent, de l'informatique scientifique est né d'une architecture particulière des processeurs, architecture que l'on dit "vectorielle". La "vectorisation" consiste à effectuer simultanément des opérations élémentaires menant à un calcul. Une image peut être celle d'une chaîne de construction automobile. Si l'on demande à un opérateur de poser le phare gauche si l'amortisseur arrière droit est déjà monté, ou le feu de brouillard arrière dans le cas contraire, son travail sera considérablement ralenti. Il devra faire le tour de la voiture, effectuer un contrôle, éventuellement changer d'outil avec l'opérateur à qui l'on aura demandé l'inverse. Pire, s'il doit installer la boîte à fusibles avant le tableau de bord... De même les performances d'un super ordinateur à architecture vectorielle sont fortement diminuées dans le cas de tests ou d'indirections. Le but de cette section est donc de montrer comment les tests peuvent être supprimés dans les calculs de la section III.D.2. Nous montrerons comment l'utilisation de fonctions indicatrices permet d'effectuer ce travail[1].

Notons que les points importants d'une bonne vectorisation sont aussi les suivants :
- effectuer le maximum d'opérations simultanément, ce qui permet d'augmenter les MFlops (III.C.1),
- minimiser les sauts d'indices dans les appels,
- effectuer un minimum d'opérations supplémentaires par rapport à l'algorithme optimal initial,
- ne pas prendre trop de temps au concepteur,
- rester lisible, éventuellement à l'aide d'explications très poussées,
- ne pas utiliser de ruses non portables.

III.D.3.1 Vectorisation du calcul de I_1.

Rappelons l'algorithme optimal de programmation de I_1 :
- définition ou calcul (portabilité de l'opération) de la précision machine, p_m,
- définition de $\epsilon_1 = \sqrt{6p_m}$,
- pour tout $\alpha \in \mathbb{C}$,
  - si $|\alpha| \leq \epsilon_1$, alors $I_1(\alpha) = Le^{i\mathbf{k}\vec{G}}$,
  - sinon $I_1(\alpha) = Le^{i\mathbf{k}\vec{G}}\frac{\sin\alpha}{\alpha}$.

Cet algorithme, à partir de $\forall\alpha \in \mathbb{C}$, peut se transformer en
- calcul de la fonction indicatrice de $|\alpha| \leq \epsilon_1$ notée $\mathcal{I}_{[|\alpha|\leq\epsilon_1]}$ qui vaut 1 si $|\alpha| \leq \epsilon_1$, qui vaut 0 sinon,
- calcul direct de I_1 par la formule analytique

$$I_1(\alpha) = Le^{i\mathbf{k}\vec{G}}e^{i\alpha}\left((1-\mathcal{I}_{[|\alpha|\leq\epsilon_1]}) \times \frac{\sin\alpha}{(1-\mathcal{I}_{[|\alpha|\leq\epsilon_1]}) \times \alpha + \mathcal{I}_{[|\alpha|\leq\epsilon_1]}} + \mathcal{I}_{[|\alpha|\leq\epsilon_1]}\right) \tag{III.D.21}$$

sans rien changer à la précision du calcul, de l'ordre de p_m. Sur calculateur Cray, nous proposons de calculer la fonction indicatrice à l'aide de la formule

$$\mathcal{I}_{[|\alpha|\leq\epsilon_1]} = \max(0, nint(sign(1., \epsilon_1 - |\alpha|))) \tag{III.D.22}$$

qui n'utilise que des fonctions intrinsèques au calculateur, fonctions dont la vectorisation est "totale".

Remarque 54 On peut aussi calculer $\mathcal{I}_{[|\alpha|\leq\epsilon_1]}$ par

$$\mathcal{I}_{[|\alpha|\leq\epsilon_1]} = nint(0.5 + sign(0.5, |\alpha| - \epsilon_1)) \tag{III.D.23}$$

et aussi la fonction $I_1(\alpha)$ par

$$I_1(\alpha) = Le^{i\mathbf{k}\vec{G}}e^{i\alpha}\left((1-\mathcal{I}_{[|\alpha|\leq\epsilon_1]}) \times \left(\overline{\alpha}\frac{\sin\alpha}{\max(|\alpha|^2, \epsilon_1^2)}\right) + \mathcal{I}_{[|\alpha|\leq\epsilon_1]}\right) . \tag{III.D.24}$$

[1] En suivant notre allégorie automobile, la vectorisation consiste à systématiquement poser les deux feux le plus vite possible, en affectant à un contrôleur la tâche de reposer un feu mal mis.

III.D.3.2 Vectorisation du calcul de I_2.

La vectorisation devient compliquée, mais est encore abordable. Pour cela, il faut calculer tous les termes définis section III.D.2.2 de façon à ce qu'ils soient définis sur $\mathbb{C}^2$.

On définit toujours les fonctions caractéristiques par rapport aux quantités ϵ_1 et ϵ_2.

$$\mathcal{I}_{[|\alpha|\geq\epsilon]} = \max(0, nint(sign(1., |\alpha| - \epsilon))) \tag{III.D.25}$$

On utilise aussi la fonction $f(\alpha)$ définie et calculée sur $\mathbb{C}$ par

$$f(\alpha) = e^{i\alpha}\left(\mathcal{I}_{[|\alpha|\geq\epsilon_1]} \times \frac{\sin\alpha}{\mathcal{I}_{[|\alpha|\geq\epsilon_1])} \times \alpha + 1 - \mathcal{I}_{[|\alpha|\geq\epsilon_1]}} + 1 - \mathcal{I}_{[|\alpha|\geq\epsilon_1]}\right) . \tag{III.D.26}$$

On calcule alors T_1 et T_2 ainsi que le développement limité $D_7(\alpha, \beta)$ (III.D.12),

$$\begin{aligned} T_1 &= -iSe^{i\mathbf{k}\vec{G}}e^{-\frac{2}{3}i(\alpha+\beta)}\frac{f(\alpha) - f(\beta)}{\mathcal{I}_{[|\gamma|\geq\epsilon_2]} \times \gamma + 1 - \mathcal{I}_{[|\gamma|\geq\epsilon_2]}} \\ T_2 &= -iSe^{i\mathbf{k}\vec{G}}e^{-\frac{2}{3}i(\gamma-\beta)}\frac{f(\gamma) - f(-\beta)}{\mathcal{I}_{[|\beta|\geq\epsilon_2]} \times \alpha + 1 - \mathcal{I}_{[|\beta|\geq\epsilon_2]}} \end{aligned} \tag{III.D.27}$$

et l'on a finalement

$$\begin{aligned} I_2(T) =& T_1\mathcal{I}_{[|\gamma|\geq\epsilon_2]} + T_2\mathcal{I}_{[|\beta|\geq\epsilon_2]}(1 - \mathcal{I}_{[|\gamma|\geq\epsilon_2]}) \\ &+ 2Se^{i\mathbf{k}\vec{G}}D_7(\alpha, \beta)(1 - \mathcal{I}_{[|\gamma|\geq\epsilon_2]})(1 - \mathcal{I}_{[|\beta|\geq\epsilon_2]}) . \end{aligned} \tag{III.D.28}$$

III.D.3.3 Vectorisation du calcul de I_3.

Nous ne présenterons pas la vectorisation du calcul de I_3 puisque ce calcul ne nous a pas servi dans les codes. Notons que la vectorisation de I_3 peut s'opérer

- soit de façon analogue aux vectorisations de I_1 et I_2, augmentant énormément le nombre d'opérations inutiles effectuées par rapport à une programmation conditionnelle,
- soit d'une façon beaucoup plus compliquée à l'aide de tableaux intermédiaires de stockage des indirections. Brièvement, il s'agirait de détecter, pour toutes les valeurs des paramètres α, β et γ, dans quel sous-cas ils se placent (dans III.D.2.3) puis de les regrouper. Par exemple le cas numéro n (définissant $I_3(n)$) serait le cas numéro m du calcul de I_3 par le développement limité (III.D.20). On calculerait alors les M cas qui donnent I_3 à l'aide du développement limité, puis on rangerait ces valeurs dans le tableau initial[2]. Le coût de la vectorisation serait alors faible. Nous considérons que cet algorithme est trop technique pour être présenté ici.

[2] Ceci consiste à séparer la chaîne en deux avec un contrôleur de l'amortisseur arrière droit, à affecter deux opérateurs pour poser les feux, puis à installer un robot pour rassembler les deux chaînes.

Annexe III.E

Déterminant de la matrice D du système linéaire quand h tend vers 0.

L'objectif de ce chapitre est de montrer que les déterminants des matrices hermitiennes $(D_k)/h$ sont non nuls lorsque h le paramètre de discrétisation du maillage tend vers zéro et dés que le nombre de fonctions de base par élément est le nombre minimal, soit

- 3 fonctions pour Helmholtz scalaire bidimensionnel,
- 4 fonctions pour Helmholtz scalaire tridimensionnel,
- 6 fonctions pour Maxwell tridimensionnel.

On montre que le déterminant se découple en un produit d'un terme dépendant uniquement de la géométrie et d'un terme dépendant uniquement du choix des directions de propagation des ondes planes. Pour les problèmes scalaires de Helmholtz discrétisés sur des simplexes, ce résultat est obtenu explicitement en donnant la valeur du terme dépendant de la géométrie. Pour le problème de Maxwell, nous ne calculons pas ce terme, mais nous montrons que ce résultat est obtenu quelle que soit la forme (non dégénérée) des éléments. Cette preuve est aussi applicable aux problèmes de Helmholtz.

Puis, on maximise ce déterminant par rapport au choix des directions de propagation des ondes planes : on montre que des directions équiréparties maximisent le déterminant.

Cela nous permet de majorer le conditionnement des matrices limites D/h lorsque h tend vers zéro. Ceci montre que l'inversion numérique de la matrice D (par une méthode directe, de Cholesky par exemple) est toujours aisée, même si le conditionnement global du système, peut être mauvais : nous n'étudions pas le conditionnement global de la matrice $(D - C)$.

Cette annexe nous permet donc d'initier une réflexion sur le choix des directions des fonctions de base. Une étude plus générale du choix de directions équiréparties (et en quel sens pour les problèmes tridimensionnels) serait une extension intéressante (mais à notre avis difficile) de ce travail.

Les preuves présentées sont très techniques.

Le problème de Helmholtz tridimensionnel utilise des preuves valables dans $\mathbb{R}^n$: la difficulté des preuves dans $\mathbb{R}^n$ avec n quelconque et dans $\mathbb{R}^3$ spécifiquement est du même ordre. Le problème de Helmholtz bidimensionnel utilise des preuves plus simples, dont les arguments sont valables dans $\mathbb{R}^2$ et non généralisables pour tout n.

III.E.1 Problème de Helmholtz bidimensionnel.

Pour le triangle numéroté par l'indice k, on définit les quantités suivantes :

1. les trois longueurs notées L_1, L_2, L_3,
2. les trois normales notées $\vec{\nu}_1$, $\vec{\nu}_2$ et $\vec{\nu}_3$,
3. le périmètre noté P,
4. la surface notée S.

On s'intéresse au cas asymptotique $h \to 0$ et $p = 3$ pour un problème bidimensionnel.

Théorème 19
On introduit la matrice limite D^r définie par

$$D^r = \lim_{h\to 0} \frac{D}{h\omega^2} \ , \tag{III.E.1}$$

ainsi que les quantités réduites indexées $(P^r, S^r, L_1^r, L_2^r, L_3^r)$ qui sont obtenues des précédentes en divisant par $h\omega^2$ sauf S^r où l'on divise par $(h\omega^2)^2$. Alors :

1. *La matrice limite D^r existe et est inversible.*
2. *Le déterminant de D^r se calcule par le produit de deux quantités, l'une faisant intervenir la géométrie du triangle, l'autre la répartition des fonctions de base. Des ondes planes équiréparties maximisent le déterminant de D^r et il vaut alors*

$$\boxed{detD^r = 27.\frac{(P^r S^r)^2}{L_1^r L_2^r L_3^r}} \tag{III.E.2}$$

3. *La majoration du conditionnement à partir du déterminant est maximale pour des ondes planes équiréparties et on majore le conditionnement de D par*

$$\frac{\lambda_{max}}{\lambda_{min}} = \frac{\lambda^r_{max}}{\lambda^r_{min}} \leq \frac{12}{\pi}(4\sigma)^4 \ . \tag{III.E.3}$$

dont le terme majorant $\sigma = h/\rho$ (I.2.4) est minimal dans le cas d'un triangle équilatéral.

La preuve de ce théorème fait l'objet de toute cette section. La preuve est longue, technique et calculatoire. Elle est effectuée en plusieurs étapes et suit l'ordre de présentation des résultats effectuée théorème 19 :

a) simplification du déterminant de la matrice D (où l'on ne garde que les termes d'ordre 1 en h) en un produit dépendant de l'unité d'échelle du triangle h et un terme indépendant de h,

b) découplage du déterminant (toujours de la matrice D où l'on n'a gardé que les termes du premier ordre en h) entre les termes géométriques et les fonctions de base,

c) optimisation du déterminant en fonction du choix des directions de propagation des ondes planes,

d) récapitulatif des points précédents et fin du calcul du déterminant (ceci montre le point 2 du théorème 19),

e) majoration du conditionnement (point 3 du théorème 19).

Le lemme suivant donne la forme de la matrice réduite D^r.

Lemme 29 *Les termes de la matrice D_k (calculés chapitre III.A) sont équivalents au premier ordre à la matrice, encore notée D (par abus) dans toute cette section, dont les termes $D^{l,m}$ pour $(l,m) \in \{1,2,3\}^2$ sont*[1]

$$D^{l,m} = \omega^2 \sum_{n=1}^{3} L_n(1 - \vec{\nu}_n\,\vec{v}_m)(1 - \vec{\nu}_n.\vec{v}_l) \ . \tag{III.E.4}$$

Les termes de la matrice réduite sont donc

$$D^r_{l,m} = \sum_{n=1}^{3} L^r_n(1 - \vec{\nu}_n\,\vec{v}_m)(1 - \vec{\nu}_n.\vec{v}_l) \ . \tag{III.E.5}$$

[1]En toute rigueur, avec les notations de ce chapitre, $D_k^{l,m} = D^{l,m} + O(h^2)$.

Preuve. Rappelons que les termes de la matrice D (annexe III.A) sont donnés à l'aide des directions de propagation $\vec{v}_{kl}$ et $\vec{v}_{jm}$ par

$$\begin{aligned} D_k^{l,m} &= \sum_{n=1}^{3} \int_{\Sigma_{k,j(k,n)}} (1+\vec{\nu}.\vec{v}_{jm}) e^{i\omega \vec{v}_{jm}.\vec{x}} \overline{(1+\vec{\nu}.\vec{v}_{kl}) e^{i\omega \vec{v}_{kl}.\vec{x}}}\, d\sigma \\ &= \omega^2 \sum_{n=1}^{3} L_n (1+\vec{\nu}.\vec{v}_{jm})(1+\vec{\nu}.\vec{v}_{kl}) \int_{\Sigma_{k,j(k,n)}} e^{(i\omega(\vec{v}_{jm}-\vec{v}_{kl}).\vec{x}} \\ &= \omega^2 \sum_{n=1}^{3} L_n (1+\vec{\nu}.\vec{v}_{jm})(1+\vec{\nu}.\vec{v}_{kl}) Z_n \frac{\sin h_n}{h_n} \end{aligned} \tag{III.E.6}$$

avec les notations de l'annexe (III.A). On a clairement :

$$\begin{cases} h_n = \omega(\vec{v}_m - \vec{v}_l).\dfrac{(\vec{x}_{n+1} - \vec{x}_n)}{2} \underset{h\to 0}{\to} 0 \\ Z_n = e^{(i\omega(\vec{v}_m - \vec{v}_l).\vec{x}_n)} \underset{h\to 0}{\to} 1 \\ e^{ih_n} \dfrac{\sin h_n}{h_n} \underset{h\to 0}{\to} 1 \end{cases}$$

□

On pose :

$$\vec{w}_n = \begin{pmatrix} 1 - \vec{\nu}_n.\vec{v}_1 \\ 1 - \vec{\nu}_n.\vec{v}_2 \\ 1 - \vec{\nu}_n.\vec{v}_3 \end{pmatrix} , \tag{III.E.7}$$

et D sera la matrice définie par

$$\begin{aligned} D &= \omega^2 (L_1 D_1 + L_2 D_2 + L_3 D_3) \\ D_n &= [\vec{w}_n \otimes \vec{w}_n] . \end{aligned} \tag{III.E.8}$$

Remarque 55 En toute rigueur, on a

$$D_k = \omega^2 (L_1 D_1 + L_2 D_2 + L_3 D_3) + O(h^2) .$$

On pose en outre, dans toute cette section,

$$\xi = (\vec{w}_1, \vec{w}_2, \vec{w}_3)_{mixte} . \tag{III.E.9}$$

a) Simplification du déterminant de D.

Lemme 30 *La matrice D est inversible si et seulement si les trois vecteurs $\vec{w}_n$ sont indépendants dans $\mathbb{R}^3$:*

$$\boxed{detD = \omega^6 L_1 L_2 L_3.(\vec{w}_1, \vec{w}_2, \vec{w}_3)^2_{mixte}}$$

Preuve. On définit les trois vecteurs $\vec{e}_n = \vec{w}_{n+1} \wedge \vec{w}_{n+2}$ pour $n - 1 = 0$ à 2 modulo 3. On note E la matrice dont les colonnes sont les trois vecteurs $\vec{e}_n$, on a

$$\begin{aligned} \det E &= (\vec{e}_1, \vec{e}_2, \vec{e}_3)_{mixte} \\ &= (\vec{w}_2 \wedge \vec{w}_3, \vec{w}_3 \wedge \vec{w}_1, \vec{w}_1 \wedge \vec{w}_2)_{mixte} \\ &= (\vec{w}_2 \wedge \vec{w}_3) \wedge (\vec{w}_3 \wedge \vec{w}_1).(\vec{w}_1 \wedge \vec{w}_2) \\ &= (\vec{w}_1, \vec{w}_2, (\vec{w}_2 \wedge \vec{w}_3) \wedge (\vec{w}_3 \wedge \vec{w}_1))_{mixte} \end{aligned}$$

Or

$$\begin{aligned} (\vec{w}_2 \wedge \vec{w}_3) \wedge (\vec{w}_3 \wedge \vec{w}_1) &= (\vec{w}_3.(\vec{w}_3 \wedge \vec{w}_1))\vec{w}_2 - (\vec{w}_2.(\vec{w}_3 \wedge \vec{w}_1))\vec{w}_3 \\ &= (\vec{w}_2, \vec{w}_3, \vec{w}_1)_{mixte}\vec{w}_3 - (\vec{w}_3, \vec{w}_3, \vec{w}_1)_{mixte}\vec{w}_2 \\ &= (\vec{w}_1, \vec{w}_2, \vec{w}_3)_{mixte}\vec{w}_3 \end{aligned}$$

Donc

$$\begin{aligned}\det E &= (\vec{w}_1, \vec{w}_2, (\vec{w}_1, \vec{w}_2, \vec{w}_3)_{mixte}\vec{w}_3)_{mixte}\\ &= (\vec{w}_1, \vec{w}_2, \vec{w}_3)^2_{mixte}\end{aligned}$$

Finalement

$$\det\ E = \xi^2 \text{ avec } \xi = (\vec{w}_1, \vec{w}_2, \vec{w}_3)_{mixte}$$

Or, $D_n = [\vec{w}_n \otimes \vec{w}_n]$ donc

$$\begin{aligned}D_n\vec{e}_k &= (\vec{w}_n\ \vec{e}_k)\vec{w}_n = (\vec{w}_n\ \vec{w}_{n+2} \wedge \vec{w}_{n+3})\vec{w}_n\\ &= (\vec{w}_n, \vec{w}_{k+1}, \vec{w}_{k+2})\vec{w}_n\ .\end{aligned}$$

Donc :

$$D_n\vec{e}_k = \begin{cases}\xi\vec{w}_n \text{ si } n = k\\ 0 \text{ si } n \neq k\end{cases}$$

ce qui entraîne que $D\vec{e}_k = \omega^2 L_k\xi\vec{w}_k$. On calcule alors le déterminant :

$$\det(De) = (\omega^2\xi)^3 L_1L_2L_3\xi = \det D.\det E = \det D.\xi^2\ .$$

Donc

$$\det D = \omega^6 L_1L_2L_3.(\vec{w}_1, \vec{w}_2, \vec{w}_3)^2_{mixte}$$

□

b) Découplage du déterminant entre les termes géométriques et les fonctions de base.

Lemme 31 $L_1\vec{\nu}_1 + L_2\vec{\nu}_2 + L_3\vec{\nu}_3 = 0$.

Preuve. On introduit R la rotation d'angle $\pi/2$ dans le plan. Alors :

(III.E.10)
$$\begin{aligned}L_1\vec{\nu}_1 &= R(\vec{x}_2 - \vec{x}_3)\\ L_2\vec{\nu}_2 &= R(\vec{x}_3 - \vec{x}_1)\\ L_3\vec{\nu}_3 &= R(\vec{x}_1 - \vec{x}_2)\end{aligned}$$

(III.E.11)
$$L_1\vec{\nu}_1 + L_2\vec{\nu}_2 + L_3\vec{\nu}_3 = R(\vec{x}_2 - \vec{x}_3 + \vec{x}_3 - \vec{x}_1 + \vec{x}_1 - \vec{x}_2) = \vec{0}$$

□

Lemme 32 *Pour* $\xi = (\vec{w}_1, \vec{w}_2, \vec{w}_3)_{mixte}$ *(cf (III.E.9)) et* $P = (L_1 + L_2 + L_3)$*, on a :*

(III.E.12)
$$\xi = \frac{P}{L_3}(\vec{\nu}_1, \vec{\nu}_2)\begin{vmatrix} v_1^1 & v_1^2 & 1\\ v_2^1 & v_2^2 & 1\\ v_3^1 & v_3^2 & 1\end{vmatrix}\ .$$

Preuve. Par définition,

$$\xi = \begin{vmatrix}(1 - \vec{\nu}_1.\vec{v}_1) & (1 - \vec{\nu}_2.\vec{v}_1) & (1 - \vec{\nu}_3.\vec{v}_1)\\ (1 - \vec{\nu}_1.\vec{v}_2) & (1 - \vec{\nu}_2.\vec{v}_2) & (1 - \vec{\nu}_3.\vec{v}_2)\\ (1 - \vec{\nu}_1.\vec{v}_3) & (1 - \vec{\nu}_2.\vec{v}_3) & (1 - \vec{\nu}_3.\vec{v}_3)\end{vmatrix}$$

Elimination de ν_3. On multiplie chaque colonne par les longueurs respectives L_n, puis on additionne les deux premières colonnes à la troisième. On utilise le lemme 31 : $L_1\vec{\nu}_1 + L_2\vec{\nu}_2 + L_3\vec{\nu}_3 = 0$. Cela donne :

$$\begin{aligned}\xi &= \frac{1}{L_1L_2L_3}\begin{vmatrix}L_1(1 - \vec{\nu}_1.\vec{v}_1) & L_2(1 - \vec{\nu}_2.\vec{v}_1) & L_3(1 - \vec{\nu}_3.\vec{v}_1)\\ L_1(1 - \vec{\nu}_1.\vec{v}_2) & L_2(1 - \vec{\nu}_2.\vec{v}_2) & L_3(1 - \vec{\nu}_3.\vec{v}_2)\\ L_1(1 - \vec{\nu}_1.\vec{v}_3) & L_2(1 - \vec{\nu}_2.\vec{v}_3) & L_3(1 - \vec{\nu}_3.\vec{v}_3)\end{vmatrix}\\ &= \frac{1}{L_1L_2L_3}\begin{vmatrix}L_1(1 - \vec{\nu}_1.\vec{v}_1) & L_2(1 - \vec{\nu}_2.\vec{v}_1) & P - (L_1\vec{\nu}_1.\vec{v}_1 + L_2\vec{\nu}_2.\vec{v}_1 + L_3\vec{\nu}_3.\vec{v}_1)\\ L_1(1 - \vec{\nu}_1.\vec{v}_2) & L_2(1 - \vec{\nu}_2.\vec{v}_2) & P - (L_1\vec{\nu}_1.\vec{v}_2 + L_2\vec{\nu}_2.\vec{v}_2 + L_3\vec{\nu}_3.\vec{v}_2)\\ L_1(1 - \vec{\nu}_1.\vec{v}_3) & L_2(1 - \vec{\nu}_2.\vec{v}_3) & P - (L_1\vec{\nu}_1.\vec{v}_3 + L_2\vec{\nu}_2.\vec{v}_3 + L_3\vec{\nu}_3.\vec{v}_3)\end{vmatrix}\\ &= \frac{P}{L_3}\begin{vmatrix}(1 - \vec{\nu}_1.\vec{v}_1) & (1 - \vec{\nu}_2.\vec{v}_1) & 1\\ (1 - \vec{\nu}_1.\vec{v}_2) & (1 - \vec{\nu}_2.\vec{v}_2) & 1\\ (1 - \vec{\nu}_1.\vec{v}_3) & (1 - \vec{\nu}_2.\vec{v}_3) & 1\end{vmatrix}\end{aligned}$$

On soustrait alors la dernière colonne aux deux premières et inverse les signes des deux premières colonnes :

(III.E.13)
$$\xi = \frac{P}{L_3} \begin{vmatrix} \vec{\nu}_1.\vec{v}_1 & \vec{\nu}_2.\vec{v}_1 & 1 \\ \vec{\nu}_1.\vec{v}_2 & \vec{\nu}_2.\vec{v}_2 & 1 \\ \vec{\nu}_1.\vec{v}_3 & \vec{\nu}_2.\vec{v}_3 & 1 \end{vmatrix}$$

Séparation des variables géométriques des variables de discrétisation. Remarquons que :

(III.E.14)
$$\begin{bmatrix} \vec{\nu}_1.\vec{v}_1 & \vec{\nu}_2.\vec{v}_1 & 1 \\ \vec{\nu}_1.\vec{v}_2 & \vec{\nu}_2.\vec{v}_2 & 1 \\ \vec{\nu}_1.\vec{v}_3 & \vec{\nu}_2.\vec{v}_3 & 1 \end{bmatrix} = \begin{bmatrix} v_1^1 & v_1^2 & 1 \\ v_2^1 & v_2^2 & 1 \\ v_3^1 & v_3^2 & 1 \end{bmatrix} \begin{bmatrix} \vec{\nu}_1^1 & \vec{\nu}_2^1 & 0 \\ \vec{\nu}_1^2 & \vec{\nu}_2^2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Et :

(III.E.15)
$$\begin{vmatrix} \vec{\nu}_1^1 & \vec{\nu}_2^1 & 0 \\ \vec{\nu}_1^2 & \vec{\nu}_2^2 & 0 \\ 0 & 0 & 1 \end{vmatrix} = (\vec{\nu}_1, \vec{\nu}_2) \ .$$

De (III.E.13) puis (III.E.14) et (III.E.15) on tire (III.E.12). □

c) Optimisation du déterminant.

Le calcul du produit mixte de trois vecteurs donne le volume du simplexe déterminé par ces trois vecteurs : nous avons pensé que le déterminant est de module maximal pour des vecteurs v_n équirépartis dans le plan. C'est ce que nous allons montrer dans le lemme suivant.

Lemme 33 *On montre que $|\xi|$ est maximal pour des vecteurs v_n équirépartis dans le plan.*

Preuve. On pose, pour $(\alpha, \beta) \in]-\pi, \pi]^2$ et $\alpha \neq \beta$

$$\begin{array}{ll} v_1^1 = \cos\alpha & v_1^2 = \sin\alpha \\ v_2^1 = \cos\beta & v_2^2 = \sin\beta \\ v_3^1 = 1 & v_3^2 = 0 \end{array}$$

Et $h(\alpha, \beta)$ tel que $\xi = \frac{P}{L_3}(\vec{\nu}_1, \vec{\nu}_2)h(\alpha, \beta)$. On a alors

(III.E.16)
$$\begin{aligned} h(\alpha, \beta) &= \begin{vmatrix} \cos\alpha & \sin\alpha & 1 \\ \cos\beta & \sin\beta & 1 \\ 1 & 0 & 1 \end{vmatrix} \\ &= \sin\alpha - \sin\beta + (\cos\alpha\sin\beta - \sin\alpha\cos\beta) \end{aligned}$$

soit

(III.E.17)
$$h(\alpha, \beta) = \sin\alpha - \sin\beta - \sin(\alpha - \beta) \ .$$

La fonction $h(\alpha, \beta)$ est périodique et continue sur $\mathbb{R}^2$: ses extrema sont donnés pour $dh(\alpha, \beta) = 0$. Rappelons que $(\alpha, \beta) \in]-\pi, \pi]^2$ et $\alpha \neq \beta$:

$$\begin{aligned} &\left\{ \begin{array}{l} \dfrac{\partial h(\alpha,\beta)}{\partial \alpha} = 0 \\ \dfrac{\partial h(\alpha,\beta)}{\partial \beta} = 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{r} \cos\alpha - \cos(\alpha - \beta) = 0 \\ -\cos\beta + \cos(\alpha - \beta) = 0 \end{array} \right. \\ \Leftrightarrow &\left\{ \begin{array}{l} \cos\alpha = \cos\beta \\ \cos\alpha - \cos(\alpha - \beta) = 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} \beta = -\alpha \\ \cos\alpha - \cos 2\alpha = 0 \end{array} \right. \\ \Leftrightarrow &\left\{ \begin{array}{l} \beta = -\alpha \\ \cos\alpha = 1 \text{ ou } -1/2 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} \alpha = +2\pi/3 \text{ et } \beta = -2\pi/3 \\ \alpha = -2\pi/3 \text{ et } \beta = +2\pi/3 \end{array} \right. \end{aligned}$$

puisque l'on exclut $\cos\alpha = 1$, choisissant par hypothèse $\vec{v}_1 \neq \vec{v}_3$. On vérifiera que la hessienne est définie négative et que les points trouvés sont bien des maxima équivalents.

Ceci prouve que le meilleur choix des fonctions de base, en vue du conditionnement de la matrice, est celui où les directions des vecteurs d'onde sont équiréparties dans le plan, et :

$$|h(\alpha,\beta)| = \frac{3\sqrt{3}}{2}\ . \tag{III.E.18}$$

□

d) Récapitulatif et fin du calcul du déterminant.

Lemme 34 *$L_1L_2(\vec{\nu}_1,\vec{\nu}_2)_{mixte}$ = 2 fois l'aire S du triangle.*

Preuve. Le raisonnement est identique à celui de la preuve du lemme 31. On introduit R la rotation d'angle $\pi/2$ dans le plan. Alors, par définition des normales (III.E.10), on a :

$$\begin{aligned} L_1\vec{\nu}_1 &= R(\vec{x}_2-\vec{x}_3)\\ L_2\vec{\nu}_2 &= R(\vec{x}_3-\vec{x}_1) \end{aligned}$$

L'aire S du triangle est définie par

$$2S = |(\vec{x}_2-\vec{x}_3,\vec{x}_3-\vec{x}_1)_{mixte}| \tag{III.E.19}$$

□

Preuve. Nous effectuons la preuve des deux premiers points du théorème 19.

D'après les lemmes successifs 30, 32, 34 et l'équation (III.E.2)

$$\begin{aligned} \det D = \omega^6 L_1L_2L_3.(\vec{w}_1,\vec{w}_2,\vec{w}_3)^2_{mixte} &= \omega^6 L_1L_2L_3.(\frac{P}{L_3}(\vec{\nu}_1,\vec{\nu}_2)\begin{vmatrix} v_1^1 & v_1^2 & 1\\ v_2^1 & v_2^2 & 1\\ v_3^1 & v_3^2 & 1\end{vmatrix})^2\\ &= \omega^6 L_1L_2L_3.(\frac{P}{L_3}(\vec{\nu}_1,\vec{\nu}_2)h(\alpha,\beta))^2 = \omega^6 L_1L_2L_3.(\frac{P}{L_3}(\vec{\nu}_1,\vec{\nu}_2))^2\frac{27}{4}\\ &= \omega^6.\frac{P^2}{L_1L_2L_3}(L_1L_2(\vec{\nu}_1,\vec{\nu}_2))^2\frac{27}{4} = \omega^6.\frac{P^2}{L_1L_2L_3}(4S^2)\frac{27}{4}\\ &= \omega^6.\frac{27P^2S^2}{L_1L_2L_3} \end{aligned}$$

Soit

$$\boxed{\det D = \omega^6.\frac{27P^2}{L_1L_2L_3}S^2}$$

□

e) Majoration du conditionnement. Nous allons montrer le dernier point du théorème 19.

Lemme 35 *Des ondes planes équiréparties assurent la majoration suivante du conditionnement des matrices D et D^r :*

$$\boxed{\frac{\lambda_{max}}{\lambda_{min}} \le \frac{12}{\pi}(4\sigma)^4} \tag{III.E.20}$$

où $\sigma = h/\rho$ est défini en (I.2.4). Cette majoration est obtenue pour un élément régulier, c'est-à-dire vérifiant les hypothèses de régularité H1, H2 et H3 de la section I.2.1.2. Dans le cas d'un triangle équilatéral, nous maximisons le terme majorant σ.

Preuve. On note λ_{min} et λ_{max} les valeurs propres minimale et maximale de la matrice D (toutes deux positives car D_k est définie positive). On définit le conditionnement par :

$$K(D) = \frac{\lambda_{max}}{\lambda_{min}} \tag{III.E.21}$$

1. Majoration des caractéristiques géométriques de Ω_k par h. On suppose que le maillage vérifie les hypothèses d'uniforme régularité. On définit h le diamètre de l'élément Ω_k considéré et ρ le diamètre du plus grand cercle de Ω_k. On suppose l'hypothèse 2 section I.2.1.2 :

$$\exists \sigma > 0 \text{ tel que } h \leq \sigma\rho \ .$$

Alors, il est clair que :
– Les faces du triangle ont des arêtes de longueur inférieure au diamètre du triangle :

$$L_n \leq h \tag{III.E.22}$$

– La surface du triangle est supérieure à l'aire d'un cercle contenu dans ce triangle :

$$S \geq \frac{\pi}{4}\rho^2 \geq \frac{\pi h^2}{4\sigma^2} \tag{III.E.23}$$

– Le périmètre total du triangle est supérieur à la circonférence d'un cercle contenu dans ce triangle (inégalité de droite dans (III.E.24)). D'après (III.E.22), on a trivialement l'inégalité de gauche.

$$3h \geq P \geq \pi\rho \geq \frac{\pi h}{\sigma} \tag{III.E.24}$$

2. Minoration du déterminant de D :

$$\mathrm{det}D = 27\omega^6 \frac{(PS)^2}{L_1 L_2 L_3} \tag{III.E.25}$$

On utilise les inégalités (III.E.24) de droite pour P, (III.E.23) pour S et (III.E.22) pour L_n. On a alors :

$$\mathrm{det}D \geq \frac{27\pi^4}{16}\frac{\omega^6 h^6}{h^3\sigma^6} \tag{III.E.26}$$

qui se simplifie en :

$$\mathrm{det}D \geq \frac{1}{3}\left(\frac{3\pi}{2}\right)^4 \frac{h^3\omega^6}{\sigma^6} \ . \tag{III.E.27}$$

3. Majoration de λ_{max}. On remarque que D a tous ses coefficients bornés par $4\omega^2(L_1 + L_2 + L_3)$, ce qui implique évidemment que :

$$\lambda_{max} \leq 12\omega^2 P \leq 36\omega^2 h \tag{III.E.28}$$

4. Minoration de λ_{min}. On remarque que :

$$\lambda_{min}\lambda_{max}^2 \geq \mathrm{det}D \tag{III.E.29}$$

donc :

$$\lambda_{min} \geq \frac{\mathrm{det}D}{\lambda_{max}^2} \ . \tag{III.E.30}$$

On remplace dans (III.E.30) la majoration de λ_{max} par (III.E.28) et la minoration de $\mathrm{det}D$ par (III.E.27) :

$$\lambda_{min} \geq \frac{1}{3}\left(\frac{3\pi}{12}\right)^4 \frac{h^3\omega^6}{\sigma^6\omega^4 h^2} \tag{III.E.31}$$

ou :

$$\lambda_{min} \geq \frac{1}{3}\left(\frac{\pi}{4}\right)^4 \frac{h\omega^2}{\sigma^6} \tag{III.E.32}$$

5. Minoration du conditionnement. De la définition du conditionnement (III.E.21) et de (III.E.30), puis en utilisant la valeur de $\mathrm{det}D$, on a :

$$K(D) = \frac{\lambda_{max}}{\lambda_{min}} \leq \frac{\lambda_{max}^3}{\mathrm{det}D} \leq (12\omega^2 P)^3 \frac{L_1 L_2 L_3}{27\omega^6 (PS)^2} \tag{III.E.33}$$

qui se simplifie d'abord en :

(III.E.34) $$K(D) \leq 4^3 P \frac{L_1 L_2 L_3}{S^2} \ .$$

On utilise les inégalités (III.E.24) de gauche pour P, (III.E.23) pour S et (III.E.22) pour L_n. On a alors :

(III.E.35) $$K(D) \leq 4^3 \pi 3 h h^3 \frac{16\sigma^4}{\pi^2 h^4}$$

et :

(III.E.36) $$K(D) \leq 4^5 \frac{3}{\pi} \sigma^4 \ .$$

La quantité σ est évidemment minimale lorsque, la surface du triangle étant fixée, le triangle est équilatéral. Notre majoration est optimale dans le cas d'un triangle équilatéral, ce que l'on comprend bien intuitivement. □

f) Remarques.

Remarque 56 Ce résultat est généralisable au problème de Helmholtz scalaire dans un espace tridimensionnel avec un tétraèdre et 4 fonctions de base (cf Annexe III.E.2). En revanche, on ne peut pas le généraliser au cas des n-polygones avec n fonctions de base $\vec{v}_n$. On peut en donner un contre-exemple pour $n = 4$. Prenons $(\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4)$ un carré, $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (1, -1, 1, -1)$. Les vecteurs w_n sont alors liés.

Remarque 57 Il est trivial de remarquer que la matrice limite D^r $(h \to 0)$ avec quatre fonctions de base ou plus $(p \geq 4)$ n'est pas inversible. En effet, la nouvelle matrice D^r obtenue s'écrira toujours sous la forme (en gardant toujours le même sens à la notation en indice r) :

$$\begin{aligned} D^r &= \omega^2 (L_1^r D_1 + L_2^r D_2 + L_3^r D_3) \\ D_n &= [\vec{w}_n \otimes \vec{w}_n] \ , \end{aligned}$$

avec

$$\vec{w}_n = \begin{pmatrix} 1 - \vec{\nu}_n . \vec{v}_1 \\ 1 - \vec{\nu}_n . \vec{v}_2 \\ 1 - \vec{\nu}_n . \vec{v}_3 \\ 1 - \vec{\nu}_n . \vec{v}_4 \end{pmatrix}$$

Le rang de la matrice D^r est au plus de trois, pour une matrice de taille $p \times p$ et $p \geq 4$.

III.E.2 Problème de Helmholtz tridimensionnel.

On s'intéresse au cas asymptotique $h \to 0$ et $p = 4$ pour le problème de Helmholtz tridimensionnel dans le vide.

Théorème 20

On introduit la matrice limite D^r définie par

(III.E.37) $$D^r = \lim_{h \to 0} \frac{D}{h\omega^2} \ ,$$

ainsi que les quantités réduites indexées $(S^r, V^r, S_1^r, S_2^r, S_3^r, S_4^r)$ qui sont obtenues des précédentes en divisant par $(h\omega^2)^2$ sauf V^r où l'on divise par $(h\omega^2)^3$. Alors :

1. *la matrice D^r est inversible,*
2. *des ondes planes équiréparties maximisent le déterminant de D^r et il vaut alors*

(III.E.38) $$\boxed{det D^r = 12\omega^8 \frac{S_r^2 V_r^4}{S_1^r S_2^r S_3^r S_4^r}} \ ,$$

3. *la majoration du conditionnement à partir du déterminant est maximale pour des ondes planes équiréparties et on majore le conditionnement de D par*

$$\frac{\lambda_{max}}{\lambda_{min}} = \frac{\lambda^r_{max}}{\lambda^r_{min}} \leq \frac{4}{3}\left(\frac{48}{\pi}\right)^4 \sigma^{12} \ , \tag{III.E.39}$$

dont le terme majorant $\sigma = h/\rho$ (cf (I.2.4)) est minimal dans le cas d'un tétraèdre régulier.

Comme pour le problème de Helmholtz bidimensionnel, la preuve est très longue. Elle suit néanmoins les mêmes étapes, plus techniques et plus générales encore :

a) simplification du déterminant de D (par abus de langage, il s'agit en fait de la matrice des termes d'ordre 1 en h de la matrice D) de façon à séparer les termes dépendant de h des autres termes,

b) découplage du déterminant en la géométrie et les fonctions de base,

c) optimisation du déterminant en fonction du choix des directions de propagation des ondes planes scalaires dans $\mathbb{R}^3$,

d) récapitulatif et fin du calcul du déterminant (prouve les points 1 et 2 du théorème 20)

e) majoration du conditionnement (prouve le point 3 du théorème 20).

III.E.2.0.1 Notations.

On considère le tétraèdre (numéroté par l'indice k que l'on omet pour ne pas alourdir les notations) défini par

1. les sommets $\vec{x}_1$, $\vec{x}_2$, $\vec{x}_3$ et $\vec{x}_4$.
2. les arêtes $\vec{x}_2 - \vec{x}_1$, $\vec{x}_3 - \vec{x}_1$, $\vec{x}_4 - \vec{x}_1$ qui forment un trièdre orienté positivement
3. les faces numérotées $(1, 2, 3, 4)$ correspondant respectivement aux triangles $(\vec{x}_2, \vec{x}_3, \vec{x}_4)$, $(\vec{x}_1, \vec{x}_3, \vec{x}_4)$, $(\vec{x}_1, \vec{x}_2, \vec{x}_4)$ et $(\vec{x}_1, \vec{x}_2, \vec{x}_3)$,
4. les normales externes (notées $\vec{\nu}_1$, $\vec{\nu}_2$, $\vec{\nu}_3$ et $\vec{\nu}_4$) aux faces $(1, 2, 3, 4)$ sont définies par :

$$\begin{aligned} \vec{\nu}_1 &= \frac{(\vec{x}_3-\vec{x}_2)\wedge(\vec{x}_4-\vec{x}_2)}{|(\vec{x}_3-\vec{x}_2)\wedge(\vec{x}_4-\vec{x}_2)|} & \vec{\nu}_2 &= -\frac{(\vec{x}_3-\vec{x}_1)\wedge(\vec{x}_4-\vec{x}_1)}{|(\vec{x}_3-\vec{x}_1)\wedge(\vec{x}_4-\vec{x}_1)|} \\ \vec{\nu}_3 &= \frac{(\vec{x}_2-\vec{x}_1)\wedge(\vec{x}_4-\vec{x}_1)}{|(\vec{x}_2-\vec{x}_1)\wedge(\vec{x}_4-\vec{x}_1)|} & \vec{\nu}_4 &= -\frac{(\vec{x}_2-\vec{x}_1)\wedge(\vec{x}_3-\vec{x}_1)}{|(\vec{x}_2-\vec{x}_1)\wedge(\vec{x}_3-\vec{x}_1)|} \end{aligned} \tag{III.E.40}$$

ou

$$\begin{aligned} \vec{\nu}_1 &= \frac{(\vec{x}_3-\vec{x}_2)\wedge(\vec{x}_4-\vec{x}_2)}{|(\vec{x}_3-\vec{x}_2)\wedge(\vec{x}_4-\vec{x}_2)|} & \vec{\nu}_2 &= -\frac{(\vec{x}_4-\vec{x}_3)\wedge(\vec{x}_1-\vec{x}_3)}{|(\vec{x}_3-\vec{x}_1)\wedge(\vec{x}_4-\vec{x}_1)|} \\ \vec{\nu}_3 &= \frac{(\vec{x}_1-\vec{x}_4)\wedge(\vec{x}_2-\vec{x}_4)}{|(\vec{x}_2-\vec{x}_1)\wedge(\vec{x}_4-\vec{x}_1)|} & \vec{\nu}_4 &= -\frac{(\vec{x}_2-\vec{x}_1)\wedge(\vec{x}_3-\vec{x}_1)}{|(\vec{x}_2-\vec{x}_1)\wedge(\vec{x}_3-\vec{x}_1)|} \end{aligned} \tag{III.E.41}$$

5. les quatre surfaces notées S_1, S_2, S_3, S_4, la plus grande de ces quatre surfaces sera simplement notée S,
6. le volume noté V.

III.E.2.0.2 Remarque sur la géométrie d'un tétraèdre.

Lemme 36 *Soient $\vec{\nu}_1, \vec{\nu}_2, \vec{\nu}_3, \vec{\nu}_4$ les normales externes aux faces du tétraèdre. On a :*

$$S_1\vec{\nu}_1 + S_2\vec{\nu}_2 + S_3\vec{\nu}_3 + S_4\vec{\nu}_4 = 0 \tag{III.E.42}$$

Preuve. Ce lemme peut se montrer par le calcul à l'aide des définitions des normales (section III.E.2.0.1). Il se montre plus simplement en remarquant que

$$S_1\vec{\nu}_1 + S_2\vec{\nu}_2 + S_3\vec{\nu}_3 + S_4\vec{\nu}_4 = \int_{\partial\Omega} \nu(\mathbf{X})d\,\mathbf{X} = \int_{\Omega} \nu(\mathbf{X})\,\nabla\, 1 = 0\ .$$

Cette preuve s'applique évidemment aussi au lemme 31. □

III.E.2.0.3 Lemmes techniques.

Lemme 37 *Soient N vecteurs $w_n \in \mathbb{C}^N$ et*

(III.E.43)
$$D = \frac{1}{\alpha}\sum_{n=1}^{N}\lambda_n[\overline{w_n}{w_n}^{\top}]$$

alors,

(III.E.44)
$$detD = \frac{|\xi|^2}{\alpha^N}\left(\prod_{n=1}^{N}\lambda_n\right)$$

avec, dans toute cette section,
(III.E.45)
$$\xi = ((w_n)_{n=1\ldots N})_{mixte}\ .$$

Preuve.

1. On construit une base de vecteurs e_m $m = 1..N$ comme suit :
$$(e_m.w_n) = \xi\delta_{n.m}$$
où $\delta_{n.m}$ est le symbole de Kronecker :
$$\delta_{n.m} = \left| \begin{array}{l} 1 \text{ si } n = m \\ 0 \text{ si } n \neq m \end{array}\right.$$
On définit la matrice $[e]$ par la matrice dont les colonnes sont les vecteurs e_n et de même pour la matrice $[w]$. On note I_N la matrice identité de $\mathbb{C}^N$. Alors
$$[w]^{\top}[e] = (w_n.e_m)_{\substack{n=1\ldots N\\ m=1\ldots N}} = \xi I_N$$
Donc :
$$\xi^N = \det([w]^{\top}[e]) = \det[w]\ \det[e] = \xi\det[e]$$
Finalement :
(III.E.46)
$$\det\ e = \xi^{N-1}$$
2. Alors :
$$\begin{aligned}[\overline{w_n}{w_n}^{\top}]e_m &= (w_n.e_m)\overline{w_n}\\ &= \xi\overline{w_n}\delta_{n,m}\end{aligned}$$
ce qui entraîne que :
(III.E.47)
$$De_m = \frac{\lambda_m\xi}{\alpha}\overline{w_m}\ ,$$
l'image par D de la base e_m est l'ancienne base w_n de $\mathbb{C}^N$. De (III.E.47) et (III.E.46), on tire :
$$\det([D][e]) = \frac{1}{\alpha^N}\lambda_1.\ldots.\lambda_N\xi^N\overline{\xi} = \det D.\det e = \det D.\xi^{N-1}$$
Donc :
(III.E.48)
$$\det D = \frac{|\xi|^2}{\alpha^N}\left(\prod_{n=1}^{N}\lambda_n\right)$$

□

Remarque 58 Si

(III.E.49)
$$D = \frac{1}{\alpha}\sum_{n=1}^{N}\lambda_n[w_n{w_n}^{*}]$$

alors

(III.E.50)
$$\det D = \frac{|\xi|^2}{\alpha^N}\left(\prod_{n=1}^{N}\lambda_n\right)$$

Nous pouvons maintenant commencer la preuve du théorème 20.

Proposition 17 *Les termes de la matrice limite* D_k^r *sont donnés par les termes* $D^{l,m}$ *(nous omettrons l'indice* k *de l'élément) pour* $(l,m) \in \{1,2,3,4\}^2$ *tels que*

$$D^{l,m} \approx \frac{\omega^2}{2} \sum_{n-1}^{4} S_n (1 - \vec{\nu}_n \, \vec{v}_m)(1 - \vec{\nu}_n . \vec{v}_l) \ , \tag{III.E.51}$$

ou, directement, par les termes $D_r^{l,m}$ *tels que*

$$D_r^{l,m} = \frac{1}{2} \sum_{n-1}^{4} S_n^r (1 - \vec{\nu}_n \, \vec{v}_m)(1 - \vec{\nu}_n . \vec{v}_l) \ . \tag{III.E.52}$$

Dans toute la suite, le terme $D^{l,m}$ *sera considéré égal à la formule (III.E.51), puisque* $D^{l,m}/h$ *a une limite.*

On pose :

$$\vec{w}_n = \begin{pmatrix} 1 - \vec{\nu}_n . \vec{v}_1 \\ 1 - \vec{\nu}_n . \vec{v}_2 \\ 1 - \vec{\nu}_n . \vec{v}_3 \\ 1 - \vec{\nu}_n . \vec{v}_4 \end{pmatrix} \tag{III.E.53}$$

alors :

$$\begin{aligned} D &= \frac{\omega^2}{2}(S_1 D_1 + S_2 D_2 + S_3 D_3 + S_4 D_4) \\ D_n &= [\vec{w}_n \otimes \vec{w}_n] \end{aligned} \tag{III.E.54}$$

On pose en outre dans toute cette section :

$$\xi = (\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4)_{mixte} \tag{III.E.55}$$

a) Simplification du déterminant de D.

Lemme 38 *On a :*

$$detD = \frac{\omega^8}{16} S_1 S_2 S_3 S_4 . (\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4)^2_{mixte} \tag{III.E.56}$$

Preuve. Il suffit d'utiliser le lemme 37 avec $\lambda_n = S_n$ et $\alpha = \omega^2/2$ et $n = 4$. □

b) Découplage du déterminant en la géométrie et les fonctions de base.

Lemme 39 *On note toujours* $\xi = (\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4)_{mixte}$*, comme en (III.E.55). Alors :*

$$\xi = -\frac{S}{S_4}(\vec{\nu}_1, \vec{\nu}_2, \vec{\nu}_3) \begin{vmatrix} v_1^1 & v_1^2 & v_1^3 & 1 \\ v_2^1 & v_2^2 & v_2^3 & 1 \\ v_3^1 & v_3^2 & v_3^3 & 1 \\ v_4^1 & v_4^2 & v_4^3 & 1 \end{vmatrix} \tag{III.E.57}$$

Preuve. Par définition

$$\xi = \begin{vmatrix} (1 - \vec{\nu}_1 . \vec{v}_1) & (1 - \vec{\nu}_2 . \vec{v}_1) & (1 - \vec{\nu}_3 . \vec{v}_1) & (1 - \vec{\nu}_4 . \vec{v}_1) \\ (1 - \vec{\nu}_1 . \vec{v}_2) & (1 - \vec{\nu}_2 . \vec{v}_2) & (1 - \vec{\nu}_3 . \vec{v}_2) & (1 - \vec{\nu}_4 . \vec{v}_2) \\ (1 - \vec{\nu}_1 . \vec{v}_3) & (1 - \vec{\nu}_2 . \vec{v}_3) & (1 - \vec{\nu}_3 . \vec{v}_3) & (1 - \vec{\nu}_4 . \vec{v}_3) \\ (1 - \vec{\nu}_1 . \vec{v}_4) & (1 - \vec{\nu}_2 . \vec{v}_4) & (1 - \vec{\nu}_3 . \vec{v}_4) & (1 - \vec{\nu}_4 . \vec{v}_4) \end{vmatrix}$$

1. On multiplie chaque colonne par les surfaces S_n, puis on additionne les trois premières colonnes à la quatrième. On termine en notant $S = (S_1 + S_2 + S_3 + S_4)$ et en utilisant la relation (III.E.42) du lemme 36 : $S_1\vec{\nu}_1 + S_2\vec{\nu}_2 + S_3\vec{\nu}_3 + S_4\vec{\nu}_4 = 0$. Cela donne :

$$\xi = \frac{1}{S_1S_2S_3S_4}\begin{vmatrix} S_1(1-\vec{\nu}_1.\vec{v}_1) & S_2(1-\vec{\nu}_2.\vec{v}_1) & S_3(1-\vec{\nu}_3.\vec{v}_1) & S_4(1-\vec{\nu}_4.\vec{v}_1) \\ S_1(1-\vec{\nu}_1.\vec{v}_2) & S_2(1-\vec{\nu}_2.\vec{v}_2) & S_3(1-\vec{\nu}_3.\vec{v}_2) & S_4(1-\vec{\nu}_4.\vec{v}_2) \\ S_1(1-\vec{\nu}_1.\vec{v}_3) & S_2(1-\vec{\nu}_2.\vec{v}_3) & S_3(1-\vec{\nu}_3.\vec{v}_3) & S_4(1-\vec{\nu}_4.\vec{v}_3) \\ S_1(1-\vec{\nu}_1.\vec{v}_4) & S_2(1-\vec{\nu}_2.\vec{v}_4) & S_3(1-\vec{\nu}_3.\vec{v}_4) & S_4(1-\vec{\nu}_4.\vec{v}_4) \end{vmatrix}$$
$$= \frac{S}{S_4}\begin{vmatrix} (1-\vec{\nu}_1.\vec{v}_1) & (1-\vec{\nu}_2.\vec{v}_1) & (1-\vec{\nu}_3.\vec{v}_1) & 1 \\ (1-\vec{\nu}_1.\vec{v}_2) & (1-\vec{\nu}_2.\vec{v}_2) & (1-\vec{\nu}_3.\vec{v}_2) & 1 \\ (1-\vec{\nu}_1.\vec{v}_3) & (1-\vec{\nu}_2.\vec{v}_3) & (1-\vec{\nu}_3.\vec{v}_3) & 1 \\ (1-\vec{\nu}_1.\vec{v}_4) & (1-\vec{\nu}_2.\vec{v}_4) & (1-\vec{\nu}_3.\vec{v}_4) & 1 \end{vmatrix}$$

On soustrait alors la dernière colonne aux trois premières et inverse les signes des trois premières colonnes :

(III.E.58)
$$\xi = -\frac{S}{S_4}\begin{vmatrix} \vec{\nu}_1.\vec{v}_1 & \vec{\nu}_2.\vec{v}_1 & \vec{\nu}_3.\vec{v}_1 & 1 \\ \vec{\nu}_1.\vec{v}_2 & \vec{\nu}_2.\vec{v}_2 & \vec{\nu}_3.\vec{v}_2 & 1 \\ \vec{\nu}_1.\vec{v}_3 & \vec{\nu}_2.\vec{v}_3 & \vec{\nu}_3.\vec{v}_3 & 1 \\ \vec{\nu}_1.\vec{v}_4 & \vec{\nu}_2.\vec{v}_4 & \vec{\nu}_3.\vec{v}_4 & 1 \end{vmatrix}$$

2. On remarque :

(III.E.59)
$$\begin{bmatrix} \vec{\nu}_1.\vec{v}_1 & \vec{\nu}_2.\vec{v}_1 & \vec{\nu}_3.\vec{v}_1 & 1 \\ \vec{\nu}_1.\vec{v}_2 & \vec{\nu}_2.\vec{v}_2 & \vec{\nu}_3.\vec{v}_2 & 1 \\ \vec{\nu}_1.\vec{v}_3 & \vec{\nu}_2.\vec{v}_3 & \vec{\nu}_3.\vec{v}_3 & 1 \\ \vec{\nu}_1.\vec{v}_4 & \vec{\nu}_2.\vec{v}_4 & \vec{\nu}_3.\vec{v}_4 & 1 \end{bmatrix} = \begin{bmatrix} v_1^1 & v_1^2 & v_1^3 & 1 \\ v_2^1 & v_2^2 & v_2^3 & 1 \\ v_3^1 & v_3^2 & v_3^3 & 1 \\ v_4^1 & v_4^2 & v_4^3 & 1 \end{bmatrix}\begin{bmatrix} \nu_1^1 & \nu_2^1 & \nu_3^1 & 0 \\ \nu_1^2 & \nu_2^2 & \nu_3^2 & 0 \\ \nu_1^3 & \nu_2^3 & \nu_3^3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Et :

(III.E.60)
$$(\vec{\nu}_1, \vec{\nu}_2, \vec{\nu}_3) = \begin{vmatrix} \nu_1^1 & \nu_2^1 & \nu_3^1 & 0 \\ \nu_1^2 & \nu_2^2 & \nu_3^2 & 0 \\ \nu_1^3 & \nu_2^3 & \nu_3^3 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

De (III.E.58) puis (III.E.59) et (III.E.60), on tire (III.E.57). □

c) Optimisation du déterminant.

Lemme 40 *On pose :*

(III.E.61)
$$h(\theta, \phi) = \begin{vmatrix} v_1^1 & v_1^2 & v_1^3 & 1 \\ v_2^1 & v_2^2 & v_2^3 & 1 \\ v_3^1 & v_3^2 & v_3^3 & 1 \\ v_4^1 & v_4^2 & v_4^3 & 1 \end{vmatrix}$$

Alors $|h(\theta, \phi)|$ est maximal pour des vecteurs $\vec{v}_n$ formant un tétraèdre régulier. On a :

(III.E.62)
$$\boxed{|h(\theta, \phi)| = \frac{16\sqrt{3}}{9}}$$

Preuve.

i) Le produit mixte de quatre vecteurs calcule le volume du simplexe déterminé par ces quatre vecteurs. Il est géométriquement intuitif de penser que le déterminant ci-dessus est de module maximal pour des vecteurs $\vec{v}_n$ équirépartis dans l'espace, c'est-à-dire formant un tédraèdre régulier.

Vérifions cette intuition :

Dans un repère de type terrestre, on note θ la longitude et ϕ la latitude. On choisit $v_1^2 = 0$ et $\phi \in [-\pi/2, +\pi/2]$. On pose :

$$\begin{array}{lll} v_1^1 = \cos\phi_1 & v_1^2 = 0 & v_1^3 = \sin\phi_1 \\ v_2^1 = \cos\phi_2\cos\theta_2 & v_1^2 = \cos\phi_2\sin\theta_2 & v_2^3 = \sin\phi_2 \\ v_3^1 = \cos\phi_3\cos\theta_3 & v_1^2 = \cos\phi_3\sin\theta_3 & v_3^3 = \sin\phi_3 \\ v_4^1 = 0 & v_4^2 = 0 & v_4^3 = 1 \end{array}$$

On effectue le changement de variable : $\psi_n = \phi_n - \pi/2$. Alors :

$$\begin{array}{lll} v_1^1 = -\sin\psi_1 & v_1^2 = 0 & v_1^3 = \cos\psi_1 \\ v_2^1 = -\sin\psi_2\cos\theta_2 & v_1^2 = -\sin\psi_2\sin\theta_2 & v_2^3 = \cos\psi_2 \\ v_3^1 = -\sin\psi_3\cos\theta_3 & v_1^2 = -\sin\psi_3\sin\theta_3 & v_3^3 = \cos\psi_3 \\ v_4^1 = 0 & v_4^2 = 0 & v_4^3 = 1 \end{array}$$

Ce changement de variable, introduit dans l'expression (III.E.61) de $h(\theta, \psi)$, donne

$$h(\theta,\psi) = \begin{vmatrix} -\sin\psi_1 & 0 & \cos\psi_1 & 1 \\ -\sin\psi_2\cos\theta_2 & -\sin\psi_2\sin\theta_2 & \cos\psi_2 & 1 \\ -\sin\psi_3\cos\theta_3 & -\sin\psi_3\sin\theta_3 & \cos\psi_3 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix}$$

qui se simplifie en

$$h(\theta,\psi) = -\begin{vmatrix} \sin\psi_1 & 0 & 1-\cos\psi_1 \\ \sin\psi_2\cos\theta_2 & \sin\psi_2\sin\theta_2 & 1-\cos\psi_2 \\ \sin\psi_3\cos\theta_3 & \sin\psi_3\sin\theta_3 & 1-\cos\psi_3 \end{vmatrix} . \tag{III.E.63}$$

Puisque la fonction $h(\theta, \psi)$ est périodique et continue sur $\mathbb{R}^5$, $|\xi|$ admet un extremum pour $dh(\theta, \psi) = 0$. Le lecteur vérifiera que la hessienne des points trouvés ci-dessous est bien définie négative, ou alors acceptera l'unicité du maximum pour des raisons géométriques (unicité des maxima équivalents). La section III.E.1 a montré que le problème dégénéré $2D$ était maximal pour des fonctions équiréparties. Nous faisons ces hypothèses dans le lemme 41.

ii) Lemme 41 *Si $\psi 1 = \psi 2 = \psi 3 = \psi$, alors $|h(\theta, \psi)|$ est maximal pour :*

$$\begin{cases} \theta_2 = 2\pi/3 \\ \theta_3 = -2\pi/3 \end{cases} . \tag{III.E.64}$$

Preuve.

$$h(\theta,\psi) = \sin^2\psi(\cos\psi - 1)\begin{vmatrix} 1 & 0 & 1 \\ \cos\theta_2 & \sin\theta_2 & 1 \\ \cos\theta_3 & \sin\theta_3 & 1 \end{vmatrix} \tag{III.E.65}$$

$$\begin{aligned} h(\theta,\psi) &= \sin\theta_2 - \sin\theta_3 + \cos\theta_2\sin\theta_3 - \cos\theta_3\sin\theta_2 \\ &= \sin\theta_2 - \sin\theta_3 - \sin(\theta_2 - \theta_3) \end{aligned}$$

On reconnaît l'expression III.E.17. On sait que cette expression est maximale pour :

$$\begin{cases} \theta_2 = 2\pi/3 \\ \theta_3 = -2\pi/3 \end{cases} \text{ ou } \begin{cases} \theta_2 = -2\pi/3 \\ \theta_3 = +2\pi/3 \end{cases}$$

Retenons :

$$\begin{cases} \theta_2 = 2\pi/3 \\ \theta_3 = -2\pi/3 \end{cases} \tag{III.E.66}$$

□

iii) Alors, si on élimine l'hypothèse $\vec{v}_1 = \vec{v}_2 ...$ (qui implique $\cos\psi \neq 1$ et $\sin\psi \neq 0$), on trouve :

$$\begin{aligned} \frac{\partial h(\theta,\psi)}{\partial\psi} = 0 &\Leftrightarrow \frac{\partial(1-\cos^2\psi)(\cos\psi - 1)}{\partial\psi} = 0 \\ &\Leftrightarrow \frac{\partial \cos\psi - 1 - \cos^3\psi + \cos^2\psi}{\partial\psi} = 0 \\ &\Leftrightarrow -\sin\psi + 3\cos^2\psi\sin\psi - 2\cos\psi\sin\psi = 0 \\ &\Leftrightarrow \sin\psi(3\cos^2\psi - 2\cos\psi - 1) = 0 \\ &\Leftrightarrow 3\cos^2\psi - 2\cos\psi - 1 = 0 \\ &\Leftrightarrow \cos\psi = -1/3 \\ &\Leftrightarrow \psi = -109,47 \text{ degrés} \end{aligned}$$

de par la condition : $\phi \in [-\pi/2, +\pi/2]$ soit $\psi \in [-\pi, 0]$. Sachant (pour des raisons géométriques triviales) que $|h(\theta, \phi)|$ admet un extremum non trivial et un seul, nous avons bien trouvé la solution de ce problème d'optimisation. Ceci prouve que le meilleur choix des fonctions de base, en vue du conditionnement de la matrice, est celui où les directions des vecteurs d'onde sont équiréparties dans l'espace.

De plus, d'après l'équation III.E.65 :

$$|h(\theta, \phi)| = |(-\frac{1}{3} - 1)\frac{8}{9}\frac{3\sqrt{3}}{2}|$$

qui donne la relation (III.E.62).

□

Lemme 42 *Soit V le volume du tétraèdre. Alors :*

$$S_1 S_2 S_3 (\vec{\nu}_1, \vec{\nu}_2, \vec{\nu}_3)_{mixte} = \frac{36}{8} \times V^2 \tag{III.E.67}$$

Preuve.

La relation (III.E.42) du lemme 36 implique

$$S_1 S_2 S_3 (\vec{\nu}_1, \vec{\nu}_2, \vec{\nu}_3)_m = -S_2 S_3 S_4 (\vec{\nu}_2, \vec{\nu}_3, \vec{\nu}_4)_m \ . \tag{III.E.68}$$

On note $\vec{u}_n = (\vec{x}_n - \vec{x}_1)$ pour n de 1 à 4. Comme le produit mixte est indépendant de la base choisie, on décide de prendre comme base affine de calcul de l'espace la base suivante : $(\vec{x}_1,\ \vec{u}_2,\ \vec{u}_3,\ \vec{u}_4)$. Dans cette base, d'après les définitions de $\vec{\nu}_n$ (III.E.40 p. 212) et S_n (section III.E.2.0.1), on a :

$$\begin{cases} S_2\vec{\nu}_2 = -1/2\vec{u}_3 \wedge \vec{u}_4 \\ S_3\vec{\nu}_3 = 1/2\vec{u}_2 \wedge \vec{u}_4 \\ S_4\vec{\nu}_4 = -1/2\vec{u}_2 \wedge \vec{u}_3 \ . \end{cases}$$

D'où :

$$S_2 S_3 S_4 (\vec{\nu}_2, \vec{\nu}_3, \vec{\nu}_4)_{mixte} = 1/8(\vec{u}_3 \wedge \vec{u}_4, \vec{u}_2 \wedge \vec{u}_4, \vec{u}_2 \wedge \vec{u}_3)_{mixte} \ . \tag{III.E.69}$$

En remarquant que le volume du tétraèdre est défini par $(\vec{u}_2, \vec{u}_3, \vec{u}_4)_m = 6V$, on a :

$$8S_2S_3S_4 \begin{bmatrix} \nu_2^1 & \nu_2^2 & \nu_2^3 \\ \nu_3^1 & \nu_3^2 & \nu_3^3 \\ \nu_4^1 & \nu_4^2 & \nu_4^3 \end{bmatrix} \begin{bmatrix} u_2^1 & u_3^1 & u_4^1 \\ u_2^2 & u_3^2 & u_4^2 \\ u_2^3 & u_3^3 & u_4^3 \end{bmatrix} = \begin{bmatrix} 6V & 0 & 0 \\ 0 & -6V & 0 \\ 0 & 0 & 6V \end{bmatrix} \tag{III.E.70}$$

Finalement,

$$S_2 S_3 S_4 (\vec{\nu}_2, \vec{\nu}_3, \vec{\nu}_4)_{mixte} = -\frac{36}{8} V^2 \tag{III.E.71}$$

De (III.E.68) et (III.E.71), on obtient (III.E.67). □

d) Récapitulatif et fin du calcul du déterminant.

Théorème 21 *Des ondes planes équiréparties assurent le meilleur conditionnement de la matrice D : Le déterminant de la matrice limite D est maximal pour un choix d'ondes planes aux vecteurs d'onde équirépartis dans le plan. Il vaut alors :*

$$\boxed{detD = 12\omega^8 \frac{S^2V^4}{S_1S_2S_3S_4}} \tag{III.E.72}$$

où S et V sont respectivement la surface et le volume du tétraèdre, (S_1, S_2, S_3, S_4) les quatre surfaces des faces.

Preuve. On note toujours $\xi = (\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4)_{mixte}$, comme en (III.E.55). La relation (III.E.56) du lemme 38 est :

$$\det D = \frac{\omega^8}{16} S_1 S_2 S_3 S_4 \xi^2,$$

la relation (III.E.57) du lemme 39 est (avec la définition (III.E.61) de $h(\theta, \phi)$) :

$$\xi = \frac{S}{S_4} (\vec{\nu}_1, \vec{\nu}_2, \vec{\nu}_3) h(\theta, \phi)$$

D'après le lemme 40, $|h(\theta, \phi)|$ est maximal pour des vecteurs $\vec{v}_n$ formant un tétraèdre régulier, et l'on a alors d'après (III.E.62) :

$$|h(\theta, \phi)| = \frac{16\sqrt{3}}{9}$$

Donc :

(III.E.73)
$$\begin{aligned} \det D &= \frac{\omega^8}{16} S_1 S_2 S_3 S_4 . \left(\frac{S}{S_4} . (\vec{\nu}_1, \vec{\nu}_2, \vec{\nu}_3) . \frac{16\sqrt{3}}{9} \right)^2 \\ &= \frac{\omega^8}{16} . \frac{1}{S_1 S_2 S_3 S_4} \left(S . (S_1 \vec{\nu}_1, S_2 \vec{\nu}_2, S_3 \vec{\nu}_3) . \frac{32\sqrt{3}}{18} \right)^2 \end{aligned}$$

De plus, d'après la relation (III.E.67) du lemme 42

$$S_1 S_2 S_3 (\vec{\nu}_1, \vec{\nu}_2, \vec{\nu}_3)_{mixte} = 36/8 V^2 \ ,$$

on a :

(III.E.74)
$$\det D = \frac{\omega^8}{S_1 S_2 S_3 S_4} \left(\frac{S}{4} \frac{36}{8} V^2 \frac{32\sqrt{3}}{18} \right)^2$$

d'où l'on tire la relation III.E.72. □

e) Majoration du conditionnement.

Théorème 22 *Des ondes planes équiréparties assurent la majoration suivante du conditionnement des matrices D et D^r :*

(III.E.75)
$$\boxed{\frac{\lambda_{max}}{\lambda_{min}} \leq \frac{4}{3} \left(\frac{48}{\pi} \right)^4 \sigma^{12}}$$

où S et V sont respectivement la surface et le volume du tétraèdre, (S_1, S_2, S_3, S_4) les quatre surfaces des faces. Cette majoration est obtenue pour un élément régulier, c'est-à-dire vérifiant les hypothèses de régularité H1, H2 et H3 de la section I.2.1.2.

Preuve. On note λ_{min} et λ_{max} les valeurs propres minimale et maximale de la matrice D (toutes deux positives car D_k est définie positive). On définit le conditionnement par :

(III.E.76)
$$K(D) = \frac{\lambda_{max}}{\lambda_{min}}$$

1. Majoration des caractéristiques géométriques de Ω_k par h. On suppose que le maillage vérifie les hypothèses d'uniforme régularité. On définit h le diamètre de l'élément Ω_k considéré et ρ le diamètre de la plus grande sphère de Ω_k. On suppose l'hypothèse 2 section I.2.1.2 :

   $$\exists \sigma \text{ tel que } h \leq \sigma \rho \ .$$

   Alors, il est clair que :
   – Les faces du tétraèdre ont des arêtes de longueur inférieure au diamètre du tétraèdre :

   (III.E.77)
   $$S_n \leq \frac{1}{2} h^2$$

– Le volume du tétraèdre est supérieur au volume d'une sphère contenue dans ce tétraèdre :

$$V \geq \frac{\pi}{6}\rho^3 \geq \frac{\pi h^3}{6\sigma^3} \qquad \text{(III.E.78)}$$

– La surface totale du tétraèdre est supérieure à la surface d'une sphère contenue dans ce tétraèdre (inégalité de droite dans (III.E.79)). D'après (III.E.77), on a trivialement l'inégalité de gauche.

$$2h^2 \geq S \geq \pi\rho^2 \geq \frac{\pi h^2}{\sigma^2} \qquad \text{(III.E.79)}$$

2. Minoration du déterminant de D :

$$\mathrm{det}D = \frac{12\omega^8 S^2 V^4}{S_1 S_2 S_3 S_4} \qquad \text{(III.E.80)}$$

On utilise les inégalités (III.E.79) de droite pour S, (III.E.78) pour V, (III.E.77) pour S_n. On a alors :

$$\mathrm{det}D \geq 8\frac{12\omega^8 \frac{\pi^2 h^4}{\sigma^4}\frac{\pi^4 h^{12}}{6^4\sigma^{12}}}{h^8} \qquad \text{(III.E.81)}$$

qui se simplifie en :

$$\mathrm{det}D \geq \frac{2\pi^6}{3^3}\frac{\omega^8 h^8}{\sigma^{16}} \ . \qquad \text{(III.E.82)}$$

3. Majoration de λ_{max}. On remarque que D a tous ses coefficients bornés par $2\omega^2(S_1+S_2+S_3+S_4)$, ce qui implique évidemment que :

$$\lambda_{max} \leq 8\omega^2 S \leq 16\omega^2 h^2 \qquad \text{(III.E.83)}$$

4. Minoration de λ_{min}. On remarque que :

$$\lambda_{min}\lambda_{max}^3 \geq |\mathrm{det}D| \qquad \text{(III.E.84)}$$

donc :

$$\lambda_{min} \geq \frac{|\mathrm{det}D|}{\lambda_{max}^3} \ . \qquad \text{(III.E.85)}$$

On remplace dans (III.E.85) par (III.E.83) et (III.E.82). Cela donne :

$$\lambda_{min} \geq \frac{2\pi^6}{3^3}\frac{\omega^8 h^8}{\sigma^{16} 2^{12}\omega^6 h^6} \qquad \text{(III.E.86)}$$

ou :

$$\lambda_{min} \geq \frac{2}{27}\left(\frac{\pi}{4}\right)^6 \frac{\omega^2 h^2}{\sigma^{16}} \qquad \text{(III.E.87)}$$

5. Minoration du conditionnement. De la définition du conditionnement (III.E.76) et de (III.E.85), puis en utilisant la valeur de det$D$, on a :

$$K(D) = \frac{\lambda_{max}^4}{\mathrm{det}D} \leq \frac{S_1 S_2 S_3 S_4 (8\omega^2 S)^4}{12\omega^8 S^2 V^4} \qquad \text{(III.E.88)}$$

qui se simplifie d'abord en :

$$K(D) \leq \frac{S_1 S_2 S_3 S_4 8^4 S^2}{12 V^4} \ . \qquad \text{(III.E.89)}$$

On utilise les inégalités (III.E.79) de gauche pour S, (III.E.78) pour V, (III.E.77) pour S_n. On a alors :

$$K(D) \leq \frac{8^2}{3}\frac{h^8 S^2}{V^4} \qquad \text{(III.E.90)}$$

et :

$$K(D) \leq \frac{4.6^4 8^2}{3\pi^4} \frac{h^8 h^4}{h^{12}} \sigma^{12} \ . \tag{III.E.91}$$

On simplifie cette expression en :

$$K(D) \leq \frac{4}{3} \left(\frac{48}{\pi} \right)^4 \sigma^{12} \ . \tag{III.E.92}$$

□

Remarque 59 Il est trivial de remarquer que la matrice limite ($h \to 0$) avec cinq fonctions de base ou plus (p) n'est pas inversible. En effet la nouvelle matrice D obtenue s'écrira toujours sous la forme :

$$\begin{aligned} D &= \omega^2 (S_1 D_1 + S_2 D_2 + S_3 D_3 + S_4 D_4) \\ D_n &= [\vec{w}_n \otimes \vec{w}_n] \end{aligned}$$

mais :

$$\vec{w}_n = \begin{pmatrix} 1 - \vec{\nu}_n.\vec{v}_1 \\ 1 - \vec{\nu}_n.\vec{v}_2 \\ 1 - \vec{\nu}_n.\vec{v}_3 \\ 1 - \vec{\nu}_n.\vec{v}_4 \\ \vdots \\ 1 - \vec{\nu}_n.\vec{v}_p \end{pmatrix}$$

Le rang de la matrice D est au plus de quatre, pour une matrice de taille $p \times p$.

III.E.3 Problème de Maxwell tridimensionnel.

Rappelons que h désigne le paramètre de taille du maillage. Dans toute cette partie, dès que cela est possible, nous omettrons de noter l'indice k de la maille de travail Ω_k. Nous utiliserons les notations de la section III.E.2 à la différence que les notations réduites ne sont plus issues des termes géométriques divisés par $(h\omega^2)^2$ mais par h^2.

Théorème 23 *Soit la matrice limite $\mathcal{D}$ construite à l'aide des sous-blocs D_k de D par*

$$\mathcal{D} = \lim_{h \to 0} \frac{D_k}{h} \tag{III.E.93}$$

1. *Le terme $\mathcal{D}^{l,m}$ de la matrice limite $\mathcal{D}$, pour l et m variant de 1 à p fonctions de base, est la matrice de produit scalaire*

   $$\mathcal{D}^{l,m} = \sum_{n=1}^{4} \frac{S^r_{k,j}}{\sqrt{\varepsilon_{kj}\mu_{kj}}} \mathcal{Z}^0_{k,m} \overline{\mathcal{Z}}^0_{k,l} \text{ où } j = j(k,n) \ , \tag{III.E.94}$$

   où $\mathcal{Z}^0_{k,m}$ est donné par (III.B.19 p. 179).
2. *Le déterminant de la matrice limite pour strictement plus de six fonctions de base est nul.*
3. *Pour six fonctions $[\mathbf{E}^{'}_m, \mathbf{H}^{'}_m]$ ondes planes définissant les fonctions de base du problème variationnel, le déterminant de $\mathcal{D}$ est proportionnel au carré du déterminant de la matrice dont la colonne d'indice m est le vecteur $[\mathbf{E}^{'}_m, \mathbf{H}^{'}_m]$. La constante ne dépend que des caractéristiques géométriques de l'élément Ω_k (non nécessairement polygonal) et des coefficients de face ε et μ. On note $|\mathcal{D}|$ le déterminant de $\mathcal{D}$. Il existe C un facteur ne dépendant que de la géométrie du problème, indépendant du choix des fonctions $[\mathbf{E}^{'}_m, \mathbf{H}^{'}_m]$, tel que*

   $$|\mathcal{D}| = C |(\mathbf{E}^{'}_m, \mathbf{H}^{'}_m)_{m=1\ldots6}|^2 \ .$$

4. *Dans le cas de coefficients ε et μ constants réels sur Ω_k et ses quatre voisins, et pour trois fonctions de base de type $\mathbf{F}$ (données par les ondes planes de la forme $\mathbf{E}^{\mathbf{F}}_{k,l}$ (II.8.22)) numérotées de 1 à 3 et trois fonctions de base de type $\mathbf{G}$ (données par les ondes planes de la forme $\mathbf{E}^{\mathbf{G}}_{k,l}$ (II.8.23)) numérotées de 4 à 6 (se reporter à la définition des fonctions de base $\mathcal{Z}_{k,l}$ II.8.29 p. 106), il existe C un facteur ne dépendant que de la géométrie du problème, indépendant du choix des fonctions $\mathbf{F}_{k,l}$ (II.8.21 p. 104), tel que le déterminant de $\mathcal{D}$ soit égal à*

$$|\mathcal{D}| = C|(\mathbf{F}_{k,l})_{l=1\ldots3}|^4 \ .$$

Le déterminant de $\mathcal{D}$ est maximal pour des fonctions de base aux directions de propagation $V_{k,l}$ (II.8.20) équiréparties (dans un plan puisqu'il n'y en a que trois).

Nous avons essayé de suivre la même logique de démonstration que pour le problème de Helmholtz tridimensionnel. Nous n'avons pas pu cependant calculer le déterminant, nous avons réussi à montrer sa proportionnalité à un facteur que l'on sait estimer. Les preuves sont moins techniques que dans l'étude du problème de Helmholtz.

Preuve. **a) Découplage du déterminant.**

i) Forme de la matrice limite. Rappelons que l'on a, d'après (III.B.19 p. 179),

$$D_k^{l,m} = \sum_{n=1}^{4} S_n^k Z_n^k D^0_{l,m} \frac{e^{ih_{n+1}}\dfrac{\sin h_{n+1}}{h_{n+1}} - e^{-ih_n}\dfrac{\sin h_n}{h_n}}{2i(h_n+h_{n+1})} \tag{III.E.95}$$

avec

$$\begin{cases} h_n = \omega(V_{k,m}-V_{k,l})\,\dfrac{(\vec{x}_{n+1}-\vec{x}_n)}{2} \\ S_n^k = 2|Face_n| \\ Z_n^k = e^{i\omega((V_{k,m}-V_{k,l})\,.\,\mathbf{X}_n)} \end{cases} \tag{III.E.96}$$

Lorsque h tend vers zéro, on a

$$\begin{cases} h_n \to 0 \\ Z_n \to 1 \\ \dfrac{e^{ih_2}\dfrac{\sin h_2}{h_2} - e^{-ih_1}\dfrac{\sin h_1}{h_1}}{2i(h_1+h_2)} \to \dfrac{1}{2} \end{cases} \tag{III.E.97}$$

donc,

$$D_k^{l,m} \to \frac{1}{2}\sum_{n=1}^{4} S_n^k D^0_{l,m} \tag{III.E.98}$$

avec

$$D^0_{l,m} = \mathcal{Z}^0_{k,m}\overline{\mathcal{Z}}^0_{k,l} \ .$$

ii) La matrice $\mathcal{D}$ est une matrice de Gram. La matrice $\mathcal{D}$ est la matrice dont le terme $\mathcal{D}_{l,m}$ est donné par

$$< \mathbf{E}^{'}_m, \mathbf{H}^{'}_m], [\mathbf{E}^{'}_l, \mathbf{H}^{'}_l] >$$

où l'application

$$< \ldots, \ldots > \begin{cases} ([\mathbf{E}^{'}_m, \mathbf{H}^{'}_m], [\mathbf{E}^{'}_l, \mathbf{H}^{'}_l]) \mapsto \displaystyle\sum_{n=1}^{4} \frac{S_n}{2\sqrt{\varepsilon_n\mu_n}} Q^n_m \overline{Q^n_l} \\ \qquad Q^n_m = \left(\sqrt{\varepsilon_n}\mathbf{E}^{'}_m \wedge \nu_n + \sqrt{\mu_n}\left(\mathbf{H}^{'}_m \wedge \nu_n\right) \wedge \nu_n\right) \\ \qquad \varepsilon_n = \sqrt{|\varepsilon_k \varepsilon_{j=j(k,n)}|} \\ ((\mathbb{C}^3)\times(\mathbb{C}^3))^2 \to \mathbb{C} \end{cases}$$

est sesquilinéaire hermitienne.

Proposition 18 *La forme sesquilinéaire $< \ldots, \ldots >$ est définie (on sait déjà qu'elles est positive) si et seulement si le polyèdre définissant les normales ν_n n'est pas dégénéré.*

Montrons que si $< [\mathbf{E}_m^{'}, \mathbf{H}_m^{'}], [\mathbf{E}_m^{'}, \mathbf{H}_m^{'}] >= 0$, alors $[\mathbf{E}_m^{'}, \mathbf{H}_m^{'}] = (0,0)$.
L'assertion $< [\mathbf{E}_m^{'}, \mathbf{H}_m^{'}], [\mathbf{E}_m^{'}, \mathbf{H}_m^{'}] >= 0$ est équivalente à

$$\begin{aligned}
&\forall n = 1 \ldots 4,\ Q_m^n = 0 \\
&\Leftrightarrow \forall n = 1 \ldots 4,\ \left(\sqrt{\varepsilon_n}\mathbf{E}_m^{'} \wedge \nu_n + \sqrt{\mu_n}\left(\mathbf{H}_m^{'} \wedge \nu_n\right) \wedge \nu_n\right) = 0 \\
&\Leftrightarrow \forall n = 1 \ldots 4,\ \left(\sqrt{\varepsilon_n}\mathbf{E}_m^{'} \wedge \nu_n + \sqrt{\mu_n}\left(-\mathbf{H}_m^{'} + (\mathbf{H}_m^{'}\nu_n)\nu_n\right)\right) = 0
\end{aligned}$$

on multiplie par $\dfrac{S_n}{\sqrt{\varepsilon_n}}\overline{\mathbf{H}_m^{'}}$ chacun des quatre termes puis l'on somme. On remarque de plus que

$$\sum_{n=1}^{4} S_n(\mathbf{E}_m^{'} \wedge \nu_n)\overline{\mathbf{H}_m^{'}} = \sum_{n=1}^{4}(\overline{\mathbf{H}_m^{'}} \wedge \mathbf{E}_m^{'})S_n\nu_n = (\overline{\mathbf{H}_m^{'}} \wedge \mathbf{E}_m^{'})\sum_{n=1}^{4} S_n\nu_n = 0$$

puisque $\sum_{n=1}^{4} S_n\nu_n = 0$. Donc :

$$\sum_{n=1}^{4} \frac{\sqrt{\mu_n}}{\sqrt{\varepsilon_n}} S_n(|(\mathbf{H}_m^{'}\nu_n)|^2 - |\mathbf{H}_m^{'}|^2) = 0$$

Ceci n'est possible, puisque $\forall n = 1 \ldots 4,\ \dfrac{\sqrt{\mu_n}}{\sqrt{\varepsilon_n}} S_n > 0$ que si $\forall n = 1 \ldots 4,\ \mathbf{H}_m^{'} = (\mathbf{H}_m^{'}\nu_n)\nu_n$. Si $\mathbf{H}_m^{'} \neq 0$ cette condition ne peut être vérifiée que si les vecteurs ν_n sont parallèles, donc que si le tétraèdre est dégénéré. On a donc bien $\mathbf{H}_m^{'} = 0$ qui implique

$$\forall n = 1 \ldots 4,\ \mathbf{E}_m^{'} \wedge \nu_n = 0 \ .$$

En supposant que le tétraèdre n'est pas dégénéré, on peut écrire $\mathbf{E}_m^{'}$ dans le repère (ν_1, ν_2, ν_3) sous la forme $\mathbf{E}_m^{'} = \alpha\nu_1 + \beta\nu_2 + \gamma\nu_3$. Alors la relation $\mathbf{E}_m^{'} \wedge \nu_1 = 0$ projetée selon ν_3 donne $\beta = 0$ et projetée selon ν_2 donne $\gamma = 0$. La relation $\mathbf{E}_m^{'} \wedge \nu_2 = 0$ projetée selon ν_3 donne $\alpha = 0$.
L'application $< \ldots, \ldots >$ est donc bien un produit scalaire de $\mathbb{C}^6$, ce qui assure que le déterminant $|\mathcal{D}|$ est nul si et seulement si la famille $[\mathbf{E}_m^{'}, \mathbf{H}_m^{'}]_{m=1\ldots6}$ est liée (résultat élémentaire sur les matrices de Gram, cf M. Monasse, cours de Mathématiques Spéciales).

iii) Découplage du déterminant de $\mathcal{D}$. Prenons $[e] = (e_1, e_2, e_3, e_4, e_5, e_6)$ une base orthonormée de $\mathbb{C}^6$. Définissons M la matrice dans la base $[e]$ de l'endomorphisme de $\mathbb{C}^6$ qui à tout e_m associe $[\mathbf{E}_m^{'}, \mathbf{H}_m^{'}]$. La matrice M est donc la matrice dont les colonnes sont constituées des coordonnées de chaque $[\mathbf{E}_m^{'}, \mathbf{H}_m^{'}]$ dans la base $[e]$. Puisque la base $[e]$ est orthonormée, on a

$$< [\mathbf{E}_m^{'}, \mathbf{H}_m^{'}], [\mathbf{E}_l^{'}, \mathbf{H}_l^{'}] >= \sum_{n=1}^{6} M_{n,m}\overline{M_{n,l}}$$

pour tout l et m donc

$$\mathcal{D} = M^{\top}\overline{M} \ .$$

Exprimé dans une autre base $[h]$, on a

$$|\mathcal{D}| = |P_{[h]-[e]}|^2|([\mathbf{E}_m^{'}, \mathbf{H}_m^{'}])_{m=1\ldots6}|^2$$

où $P_{[h]-[e]}$ dénote la matrice de passage entre les base $[e]$ et $[h]$.

iv) Conclusion. On a

$$\boxed{|\mathcal{D}| = C|([\mathbf{E}_m^{'}, \mathbf{H}_m^{'}])_{m=1\ldots6}|^2}$$

où C ne dépend que de la géométrie de Ω_k et des coefficients ε et μ.

b) Cas à coefficients réels constants.

Montrons le point 4 du théorème 23 dans le cas où les fonctions de base sont des ondes planes aux polarisations complexes conjuguées deux à deux. Ces fonctions sont soit de type **F** (II.8.22), soit e type **G** (II.8.23). On sait que les termes de la matrice de produit scalaire correspondant à leur couplage sont nuls, et par passage à la limite, il en est de même pour la matrice D_r. Remarquons de plus que toujours dans le cas $\varepsilon = \mu = 1$ du vide et pour $h \to 0$, l'on passe d'un type à l'autre par simple conjugaison (le terme intégrale tend vers une limite réelle), donc tous les résultats obtenus sur les fonctions de type (II.8.22) se généralisent aux fonctions de type (II.8.23) par conjugaison.

La matrice $\mathcal{D}$ est donc de la forme :

$$\mathcal{D} = \begin{bmatrix} D & 0 \\ 0 & \overline{D} \end{bmatrix} ,$$

en notant encore par D, par abus de langage, la matrice limite issue des quatre fonctions de base de type **F**.

La matrice D est une matrice de produit scalaire sur l'espace des fonctions de type **F** donc c'est une matrice de Gram définie positive pour trois fonctions de base linéairement indépendantes. On sait qu'alors le déterminant de D est proportionnel au carré du déterminant dans la base canonique de la matrice $[F]$ dont les colonnes sont les fonctions de base :

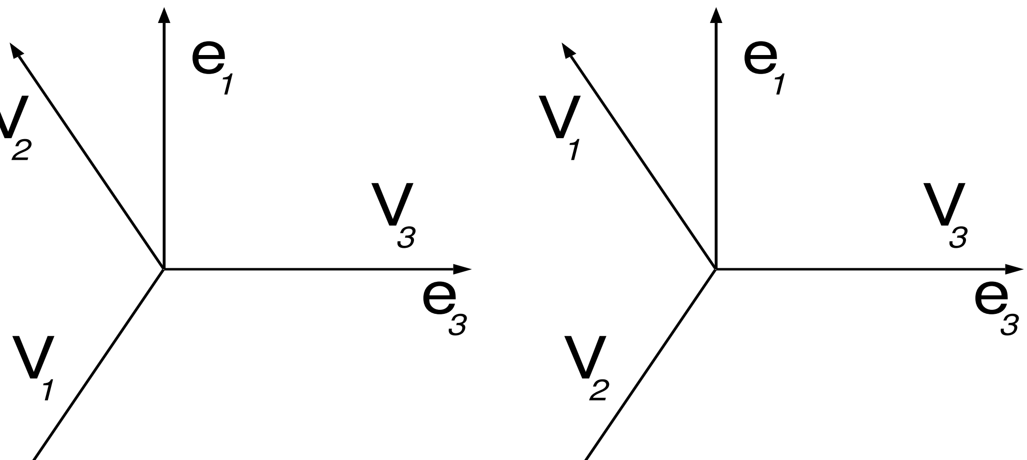
$$det[D] = \lambda |[F]|^4 \qquad \text{(III.E.99)}$$

où $[F]$ est la matrice dont les colonnes F_m sont des vecteurs de la forme (II.8.21) : $\mathbf{F}_{k,m} = \left(\mathbf{E}^0_{k,m} + i\mathbf{E}^0_{k,m} \wedge V_{k,m}\right)$.

c) Maximisation du déterminant de la matrice $[F]$.

Nous construisons la matrice carrée dont les colonnes sont les vecteurs de base. Nous considérons donc une matrice de taille 3 à coefficients complexes. Nous allons chercher à maximiser le module du déterminant de cette matrice.

Lemme 43 *On note $[F] = (\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3)$ la matrice dont les vecteurs colonnes sont les vecteurs complexes $\mathbf{F}_n$. On suppose les modules des vecteurs $\mathbf{F}_n$ égaux à $\sqrt{2}$. Les parties réelles et imaginaires des fonctions $\mathbf{F}_n$ sont égales en module et orthogonales. Nous montrons que le déterminant de la matrice $[F]$ est maximal pour des fonctions $\mathbf{F}_n$ orthogonales à des vecteurs directeurs V_n équirépartis dans un plan de l'espace $\mathbb{R}^3$ (cf figure III.E.1).*



Sens négatif, sens positif.

FIG. III.E.1 – Forme des directions.

Preuve. On effectue le calcul dans le repère orthonormé direct canonique (e_1, e_2, e_3). Nous avons toute liberté pour le choix de F_3. Nous décidons de prendre F_3 tel que $E_3 = e_2$ et $H_3 = e_1$: sa partie réelle et sa partie imaginaires sont orthogonales à e_3. Une rotation d'angle α dans le plan orthogonal à e_3 pour le choix de E_3 donne un nouveau vecteur $F_3^{'} = E_3^{'} + iH_3^{'}$ tel que

$$\begin{cases} E_3^{'} = \cos(\alpha)e_2 - \sin(\alpha)e_1 \\ H_3^{'} = \sin(\alpha)e_2 + \cos(\alpha)e_1 \ . \end{cases}$$

On a donc

$$\begin{aligned} F_3^{'} &= (\cos(\alpha) + i\sin(\alpha))e_2 - (\sin(\alpha) - i\cos(\alpha))e_1 \\ &= e^{i\alpha}(e_2 + ie_1) \\ &= e^{i\alpha}F_3 \end{aligned}$$

ce qui garde inchangé le module du déterminant : cette propriété, valable pour la matrice D de produit scalaire (proposition 14), est valable pour la matrice limite, et aussi pour la matrice $[F]$ des polarisations. Nous l'avons montré ici sur le premier vecteur F_1, dans la suite, nous le montrons de la même façon sur les deux autres vecteurs F_2 et F_3. On continue la construction précédente en choisissant de prendre F_2 dans le plan orthogonal à $\cos(\zeta)e_3 + \sin(\zeta)e_1$. On a donc

$$\begin{cases} E_2 = \cos(\eta)e_2 + \sin(\eta)\left(-\sin(\zeta)e_3 + \cos(\zeta)e_1\right) \\ H_2 = -\sin(\eta)e_2 + \cos(\eta)\left(-\sin(\zeta)e_3 + \cos(\zeta)e_1\right) \end{cases}$$

et on vérifie que le déterminant ne dépend pas de ζ :

$$\begin{aligned} F_2 &= (\cos(\eta) - i\sin(\eta))e_2 + \left(-\sin(\zeta)e_3 + \cos(\zeta)e_1\right)\left(\sin(\eta) + i\cos(\eta)\right) \\ &= e^{-i\eta}(e_2 + i\left(-\sin(\zeta)e_3 + \cos(\zeta)e_1\right)) \\ &= e^{-i\eta}\begin{bmatrix} i\cos(\zeta) \\ 1 \\ -i\sin(\zeta) \end{bmatrix} \ . \end{aligned}$$

Enfin, nous choisissons F_1 dans le plan orthogonal à

$$\begin{bmatrix} \cos(\theta)\cos(\phi) \\ \cos(\theta)\sin(\phi) \\ \sin(\theta) \end{bmatrix}$$

donc par exemple

$$E_1 = \begin{bmatrix} \cos(\lambda)\sin(\theta)\cos(\phi) + \sin(\lambda)\sin(\phi) \\ \cos(\lambda)\sin(\theta)\sin(\phi) - \sin(\lambda)\cos(\phi) \\ -\cos(\lambda)\cos(\theta) \end{bmatrix} \text{ et } H_1 = \begin{bmatrix} -\sin(\lambda)\sin(\theta)\cos(\phi) + \cos(\lambda)\sin(\phi) \\ -\sin(\lambda)\sin(\theta)\sin(\phi) - \cos(\lambda)\cos(\phi) \\ +\sin(\lambda)\cos(\theta) \end{bmatrix}$$

et on vérifie que le déterminant ne dépend pas de λ :

$$\begin{aligned} F_1 &= \begin{bmatrix} (\cos(\lambda) - i\sin(\lambda))\sin(\theta)\cos(\phi) + (\sin(\lambda) + i\cos(\lambda))\sin(\phi) \\ (\cos(\lambda) - i\sin(\lambda))\sin(\theta)\sin(\phi) - (\sin(\lambda) + i\cos(\lambda))\cos(\phi) \\ -(\cos(\lambda) - i\sin(\lambda))\cos(\theta) \end{bmatrix} \\ &= \begin{bmatrix} e^{-i\lambda}\sin(\theta)\cos(\phi) + ie^{-i\lambda}\sin(\phi) \\ e^{-i\lambda}\sin(\theta)\sin(\phi) - ie^{-i\lambda}\cos(\phi) \\ -e^{-i\lambda}\cos(\theta) \end{bmatrix} \\ &= e^{-i\lambda}\begin{bmatrix} \sin(\theta)\cos(\phi) + i\sin(\phi) \\ \sin(\theta)\sin(\phi) - i\cos(\phi) \\ -\cos(\theta) \end{bmatrix} \ . \end{aligned}$$

On doit donc maximiser le module du déterminant de la matrice

$$\begin{bmatrix} \sin(\theta)\cos(\phi) + i\sin(\phi) & i\cos(\zeta) & i \\ \sin(\theta)\sin(\phi) - i\cos(\phi) & 1 & 1 \\ -\cos(\theta) & -i\sin(\zeta) & 0 \end{bmatrix}$$

que l'on peut comparer à la matrice des vecteurs par rapport auxquels on a choisi les fonctions de base :

$$\begin{bmatrix} \cos(\theta)\cos(\phi) & \sin(\zeta) & 0 \\ \cos(\theta)\sin(\phi) & 0 & 0 \\ \sin(\theta) & \cos(\zeta) & 1 \end{bmatrix} .$$

La fonction à optimiser atteint son maximum pour ses dérivées par rapport aux trois variables θ, ϕ et ζ toutes trois nulles (le domaine d'étude est périodique). Cela se traduit, en effectuant le changement de variable qui consiste à tout exprimer en fonction de l'arc moitié (t_1 est la tangente de $\theta/2$, t_2 est la tangente de $\phi/2$ et t_3 est la tangente de $\zeta/2$), en réduisant au même dénominateur, en utilisant le fait que t_1, t_2 et t_3 sont réels, par :

$$\begin{cases} 0 = -4t_3^2(t_1-1)(t_1^3t_2^2t_3 - 2t_1^2t_2^2t_3^2 - t_1^3t_2^2 + 3t_1^2t_2^2t_3 - 2t_1t_2^2t_3^2 - t_1^3t_3 \\ \quad + t_1^2t_2^2 - 2t_1^2t_3^2 - 3t_1t_2^2t_3 - t_1^3 - 3t_1^2t_3 + t_1t_2^2 - 2t_1t_3^2 \\ \quad - t_2^2t_3 + t_1^2 + 3t_1t_3 - t_2^2 + t_1 + t_3 - 1) \\ 0 = -8(t_1-1)^3(t_1+1)t_2t_3^3 \\ 0 = 2t_3(t_1-1)^2(t_1^2t_2^2t_3^3 + t_1^2t_2^2t_3^2 - 3t_1^2t_2^2t_3 - t_1^2t_3^3 + 6t_1t_2^2t_3^2 - t_2^2t_3^3 \\ \quad + t_1^2t_2^2 + t_1^2t_3^2 + t_2^2t_3^2 + 3t_1^2t_3 - 2t_1t_2^2 + 6t_1t_3^2 \\ \quad + 3t_2^2t_3 + t_3^3 + t_1^2 + t_2^2 + t_3^2 - 2t_1 - 3t_3 + 1) . \end{cases}$$

La deuxième équation ci-dessus implique que l'on a au choix (non exclusif) :

$$\left| \begin{array}{l} t_1 = 1 \\ t_1 = -1 \end{array} \right. \quad \left| \begin{array}{l} t_2 = 0 \\ t_3 = 0 \end{array} \right.$$

Nous éliminons les choix $t_1 = 1$ (respectivement $t_3 = 0$) car ils donnent un déterminant nul (c'est-à-dire un minimum). Cela correspond évidemment à F_1 (respectivement F_2) colinéaire à F_3.

Pour $t_1 = -1$ on a :

$$\begin{cases} 0 = -32({t_2}^2+1){t_3}^3({t_2}^2-1) \\ 0 = -32t_3({t_2}^2+1)^2({t_3}^2-1) \end{cases}$$

et le déterminant est de module $2|\sin(\zeta)| \leq 2$ maximal pour $\zeta = +\pi/2$ ou $\zeta = -\pi/2$ qui donne respectivement $t_3 = 1$ ou $t_3 = -1$. Cela implique, dans les deux cas, $t_2 = 1$ ou $t_2 = -1$, c'est-à-dire $\phi = -\pi/2$ ou $\phi = +\pi/2$. Remarquons que changer ϕ en $-\phi$ dans le cas $\theta = -\pi/2$ correspond à une rotation du vecteur F_1 dans le plan orthogonal à F_3. Il s'agit donc de la même configuration, puisqu'on s'intéresse aux vecteurs de base modulo les rotations par rapport à leur direction normale. La conclusion de cette étude est que l'on obtient deux maxima locaux qui correspondent à deux positions optimales du vecteur F_2, lorsque l'on choisit F_1 et F_3 orthogonaux. Remarquons que ces deux positions sont symétriques par rapport au plan (e_1, e_2) qui définit F_3 (comme toujours modulo une rotation).

Pour $t_2 = 0$, on a :

$$\begin{cases} 0 = 4(t_1-1)(t_1^2 + 2t_3t_1 - 1)(t_3t_1 + t_1 + t_3 - 1)t_3^2 \\ 0 = -2(t_1-1)^2(t_3t_1 + t_1 + t_3 - 1)(t_3^2t_1 - t_3^2 - 2t_3 - 2t_3t_1 + 1 - t_1)t_3 \end{cases}$$

ou, sachant par hypothèse que $t_3 \neq 0$ et $t_1 \neq 1$,

$$\begin{cases} 0 = (t_3t_1 + t_1 + t_3 - 1)(t_1^2 + 2t_3t_1 - 1) \\ 0 = (t_3t_1 + t_1 + t_3 - 1)(t_3^2t_1 - t_3^2 - 2t_3 - 2t_3t_1 + 1 - t_1) \end{cases}$$

et le déterminant vaut

$$2|\frac{t_3(-2t_1+1+t_1^2+t_3t_1^2-t_3)}{(1+t_3^2)(1+t_1^2)}|\ .$$

Nous éliminons le cas $(t_3t_1+t_1+t_3-1)=0$. En effet, puisqu'on suppose $t_1 \neq -1$ (qui nous reporterait au cas précédent), on aurait $t_3 = \frac{1-t_1}{1+t_1}$. On vérifie qu'alors le déterminant est nul : cela correspond au cas F_1 colinéaire à F_2.

On a donc les deux équations

$$\begin{cases} 0 = t_1^2 + 2t_3t_1 - 1 \\ 0 = t_3^2t_1 - t_3^2 - 2t_3 - 2t_3t_1 + 1 - t_1 \end{cases}$$

et le déterminant est donné par

$$det = \cos(\theta+\zeta) + \sin(\zeta) - \cos(\theta)$$

qui est maximal en la valeur $\frac{3\sqrt{3}}{2} \approx 2,598076$. En effet, le maximum est obtenu pour

$$\begin{cases} 0 = \sin(\theta+\zeta) - \sin(\theta) \\ 0 = \sin(\theta+\zeta) - \cos(\zeta)\ , \end{cases}$$

c'est-à-dire

$$\begin{cases} \sin(\theta) = \sin(\theta+\zeta) \\ \sin(\theta) = \cos(\zeta)\ . \end{cases}$$

1. Si $\zeta = \pi/2 - \theta$, alors $\sin(\theta) = \sin(\pi/2) = 1$ donc $\theta = \pi/2$ qui est une situation déjà rencontrée : le déterminant est nul.
2. Si $\zeta = \theta - \pi/2$, alors $\sin(\theta) = \sin(2\theta - \pi/2) = -\cos(2\theta)$ soit $(-\sin(\theta)+1)(2\sin(\theta)+1)$ donc :
   (a) si $\theta = +\pi/2$, on est ramené au cas précédent où le déterminant est nul,
   (b) si $\theta = -5\pi/6$, on a $\zeta = 2\pi/3$ et le déterminant est $\frac{3\sqrt{3}}{2}$,
   (c) si $\theta = -\pi/6$, on a $\zeta = -2\pi/3$ et le déterminant est $\frac{3\sqrt{3}}{2}$.

Dans les deux cas ci-dessus, on a trouvé deux extrema locaux qui sont des maxima globaux équivalents puisque supérieurs à tous les maxima trouvés précédemment. On vérifie que la matrice des dérivées secondes en θ, ϕ et ζ,

$$-\sqrt{3}\begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 3/8 & 0 \\ 1/2 & 0 & 1 \end{bmatrix}$$

est bien définie négative. On a alors les configurations

$$\begin{bmatrix} -1/2 & -i/2 & i \\ -i & 1 & 1 \\ +\sqrt{3}/2 & -i\sqrt{3}/2 & 0 \end{bmatrix} \text{ et } \begin{bmatrix} -1/2 & -i/2 & i \\ -i & 1 & 1 \\ -\sqrt{3}/2 & i\sqrt{3}/2 & 0 \end{bmatrix}$$

qui correspondent aux directions :

$$\begin{bmatrix} -\sqrt{3}/2 & \sqrt{3}/2 & 0 \\ 0 & 0 & 0 \\ -1/2 & -1/2 & 1 \end{bmatrix} \text{ et } \begin{bmatrix} \sqrt{3}/2 & -\sqrt{3}/2 & 0 \\ 0 & 0 & 0 \\ -1/2 & -1/2 & 1 \end{bmatrix}$$

Dans le plan (e_3, e_1), cela donne la configuration de la figure III.E.1.

□

□

Bibliographie

[1] Abboud (Toufic). – *Etude mathématique et numérique de quelques problèmes de diffraction d'ondes électromagnétiques.* – Thèse de Doctorat, FRANCE/Ecole Polytechnique, 1991.

[2] Angélini (J. J.), Soize (Ch.) et Soudais (P.). – Méthode numérique mixte pour la résolution des équations de Maxwell harmoniques. *La recherche aérospatiale*, vol. 4, 1992, pp. 27–72 (en trois parties).

[3] Angélini (J. J.), Soize (Ch.) et Soudais (P.). – Hybrid numerical method for harmonic 3D Maxwell equations : scattering by a mixed conducting and inhomogeneous anisotropic dielectric medium. *IEEE Trans. Ant. Prop.*, vol. 41, n° 1, 1993, pp. 66–76.

[4] Benamou (J.D.). – A domain decomposition method for the optimal control of system governed by the Helmholtz equation. *Third international conference on mathematical and numerical wave propagation phenomena.*, vol. SIAM, 1995. – Cannes-Mandelieu.

[5] Bérenger (J.P.). – A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, vol. 114, 1994, pp. 185–200.

[6] Bernardi (C.) et Maday (Y.). – *Approximations spectrales de problèmes aux limites elliptiques.* – SMAI. Springer Verlag, 1992.

[7] Bonnefoy (Jean-Louis). – Communication privée, CEA/Cesta, 1991.

[8] Bonnemason (P.) et Stupfel (B.). – Modeling high frequency scattering by axisymmetric perfectly or imperfectly conducting scatterers. *Electromagnetics*, vol. 13, 1993, pp. 111–129.

[9] Bouche (D.) et Molinet (M.). – *Méthodes asymptotiques en électromagnétisme.* – SMAI. Springer Verlag, 1994.

[10] Bouche (Daniel). – *La méthode des courants asymptotiques.* – Thèse de Doctorat, FRANCE/Université Bordeaux I, 1992.

[11] Brézis (Haïm). – *Analyse fonctionnelle - Théorie et applications.* – Paris, Masson, 1987.

[12] Brezzi (F.) et Fortin (M.). – *Mixed and Hybrid Finite Element Methods.* – New York, Springer Verlag, 1991.

[13] Cai (X.C.) et Widlund (O.B.). – Domain decomposition algorithm for indefinite elliptic problems. *SIAM, J. Sci. Stat. Comput*, vol. 13, 1992, pp. 243–258.

[14] Cessenat (Michel). – *Mathematical Methods in Electromagnetism, Linear Theory and Applications*, vol. 41 of *Series on Advances in Mathematics for Applied Sciences.* – World Scientific, 1996.

[15] Cessenat (O.) et Després (B.). – *Une nouvelle formulation variationnelle des équations d'onde en fréquence. Application au problème de Helmholtz* 2*D.* – Note n° 2779, CEA, décembre 1994.

[16] Cessenat (O.) et Després (B.). – Application of an Ultra Weak Variational Formulation of Elliptic PDEs to the 2D Helmholtz Problem. *SIAM Journal of Numerical Analysis*, vol. Accepted for publication, 1995.

[17] Ciarlet (P.G.). – *The Finite Element Method for Elliptic Problems.* – Paris, North Holland, 1979.

[18] Clement (F.), Kern (M.) et Rubin (C.). – Solution of the 3D Helmholtz equation by conjugate gradients. *Copper mountain conference on iterative methods*, 1990.

[19] Collino (F.). – *Boundary Conditions and Layer Technique for the Simulation of Electromagnetic Waves above a Lossy Medium.* – Rapport technique n° 2698, INRIA, Domaine de Voluceau Rocquencourt, BP 105, 78 153 Le Chesnay Cedex France, novembre 1995.

[20] Colton (D.) et Kreiss (R.). – *Integral Equation Methods in Scattering Theory.* – Wiley-Interscience, 1983.

[21] Costabel (M.). – A remark on the regularity of solutions of Maxwell's equations on Lipschitz domains. *Math. meth. in the appl. sci.*, vol. 12, 1990.

[22] Crouzet (Laurent). – *Résolution des équations de Maxwell tridimensionnelles en régime fréquentiel par éléments finis conformes, multiplicateurs de Lagrange et méthodes itératives.* – Thèse de Doctorat, FRANCE/Paris VI, 1994.

[23] Dautray (R.) et Lions (J.L.). – *Analyse mathématique et calcul numérique*, vol. 3 : Transformations, Sobolev, Opérateurs. – Paris, Masson, 1987.

[24] Després (B.). – *Méthode de décomposition de domaine pour les problèmes de propagation d'ondes en régime harmonique. Le théorème de Borg pour l'équation de Hill vectorielle.* – Thèse de Doctorat, FRANCE/Paris IX Dauphine, 1991.

[25] Després (B.). – *Un procédé de discrétisation des équations d'ondes en fréquence.* – Note n° 2726, CEA, mai 1993.

[26] Després (B.). – *Implementation of a non overlapping domain decomposition method on a Cray T3D for solving the 3D harmonic Maxwell's equations.* – Communication, IMA international symposium on Maxwell's equations, 1994.

[27] Després (B.). – Sur une formulation variationnelle de type ultra-faible. *C.R. Acad. Sci. Paris*, vol. 318, n° I, 1994, pp. 939–944.

[28] Després (B.) et Cessenat (O.). – A new variational formulation of elliptic linear PDEs. Application to the 2D Helmholtz problem. *In : Workshop on Approximations and Numerical Methods for the Solution of the Maxwell Equations*, éd. par Laboratory (Oxford University Computing). – Oxford, mars 1995.

[29] Després (B.), Joly (P.) et Roberts (J.E.). – A domain decomposition method for the harmonic Maxwell equations. *Iterative methods in linear algebra.*, vol. IMACS, 1992, pp. 475–484.

[30] Engquist (B.) et Majda (A.). – Absorbing boundary conditions for the numerical simulation of waves. *Math. of Comp.*, vol. 31, 1977, pp. 629–651.

[31] Freund (R. W.) et Nachtigal (N. M.). – QMR : A quasi-minimal residual method for non hermitian linear systems. *Numerische Mathematik.*, vol. 60, 1991, pp. 315–339.

[32] Golub (G.) et al. – International symposium on domain decomposition methods for partial differential equations. *SIAM.*, 1990.

[33] Grégoire (J.), Nédelec (J.C.) et Planchard (J.). – *Problèmes relatifs à l'équation de Helmholtz.* – Rapport technique n° Bulletin de la DER, EDF, 1974.

[34] Hackbusch (Wolfgang). – *Multi-Grid Methods and Applications*, vol. 4 of *Springer Series in Computational Mathematics.* – Heidelberg, Springer Verlag, 1985.

[35] Johnson (G.) et Nédelec (J. C.). – On the coupling of boundary integral and finite element methods. *Math. Comput.*, vol. 35, 1980, pp. 1063–1079.

[36] Joly (P.) et Trouvé (P.). – *Méthodes Mathématiques et Numériques de Propagation d'Ondes en Régime Harmonique.* – Rapport technique n° Cours de DEA, Université Paris Dauphine, Paris IX, 1994.

[37] La Bourdonnaye (A. De). – High frequency approximation of integral equations modeling scattering phenomena. *Mod. Math. et Anal. Num.*, vol. 28, 1994, pp. 223–241.

[38] La Bourdonnaye (A. De). – Some formulations coupling volumic and integral equation methods for Helmholtz equation and electromagnetism. *Num. Math.*, vol. 69, 1995, pp. 257–268.

[39] La Bourdonnaye (A. De). – *A substructuring method for a harmonic wave propagation problem : Analysis of the conditioning number of the problem on the interface.* – Rapport technique, INRIA BP 93, 06 902 Sophia Antipolis Cx, CERMICS, 1995.

[40] Lafitte (Olivier). – *Quelques résultats asymptotiques pour des problèmes hyperboliques.* – Cours de dea, Université de Villetaneuse, 1996.

[41] Le Martret (R.) et Cessenat (O.). – An integral equation formulation. *In : Workshop on Approximations and Numerical Methods for the Solution of the Maxwell Equations*, éd. par Laboratory (Oxford University Computing). – Oxford, mars 1995.

[42] Le Potier (C.). – *Evolutions du code "SUMER-T" : Matériaux à caractéristiques réelles. Condition aux limites absorbantes d'ordre* 1. – Note n° 2582, CEA, 1994.

[43] Le Potier (C.) et Le Martret (R.). – Finite volume solution of Maxwell's equations in nonsteady mode. *La Recherche Aérospatiale*, vol. 5, 1994, pp. 329–342.

[44] Lions (J.L.) et Magenes (E.). – *Problèmes aux limites non homogènes et applications*, vol. 1 et 2. – Paris, Dunod, 1968.

[45] Lions (P.L.). – On the Schwarz alternating method. *SIAM*, vol. 3, 1990, p. Third international symposium on domain decomposition methods for partial differential equations.

[46] Mitrea (M.). – The method of layer potentials in electromagnetic scattering theory on nonsmooth domains. *Duke Math. J.*, vol. 77, n° 1, janvier1995, pp. 111–133.

[47] Monk (P.). – A finite element method for approximating the time-harmonic Maxwell equations. *Numer. Math.*, vol. 63, 1992, pp. 243–261.

[48] Nédelec (J. C.). – Mixed finite element in $\mathbb{R}^3$. *Numer. Math.*, vol. 35, 1980, pp. 315–341.

[49] Nédelec (J. C.). – A new family of mixed finite element in $\mathbb{R}^3$. *Numer. Math.*, vol. 50, 1986, pp. 57–81.

[50] Nédelec (J.C.). – *Approximation des équations intégrales en Mécanique et en Physique.* – Cours de l'école d'été d'analyse numérique CEA-EDF-INRIA, Ecole Polytechnique, 1977.

[51] Opfer (G.) et Schober (G.). – Richardson's iteration for nonsymmetric matrices. *Linear Algebra and its Applications*, vol. 58, 1984, pp. 343–361.

[52] Pascual (J.). – *Applications des éléments finis discontinus aux équations de la neutronique et de la photonique.* – Note n° 2798, CEA, 1995.

[53] Press (W.H), Vettering (W.K), Teukolsky (S.) et Flannery (B.). – *Numerical Recipes in Fortran, the Art of Scientific Computing.* – Cambridge University Press, 1992, 194–198p.

[54] Raviart (P.A) et Thomas (J.M). – *Introduction à l'analyse numérique des équations aux dérivées partielles.* – Collection Mathématiques Appliquées pour la maîtrise. Paris, Masson, 1992, 3 édition.

[55] Ruck (George T.). – *Radar Cross Section Handbook.* – Plenum Press, 1970.

[56] Saad (Y.) et Schultz (M.). – GMRes : A generalized minimal residual algorithm for solving nonsymmetric linear systems. *Siam J. Sci. Stat. Comput.*, vol. 7, n° 3, juillet 1986, pp. 856–869.

[57] Theodor (R.) et Lascaux (P.). – *Analyse numérique matricielle appliquée à l'art de l'ingénieur*, vol. 2. – Paris, Masson, 1986.

[58] Van Der Vorst (H. A.). – Bi-CGStab : A fast and smoothly converging variant of BI-CG for the solution of nonsymmetric linear systems. *Siam J. Sci. Stat. Comput.*, vol. 13, n° 2, mars 1992, pp. 631–644.

[59] Wang (Dau Sing). – Limits and validity of the impedance boundary condition on penetrable surfaces. *IEEE Transactions on Antennas and Propagation*, vol. AP-35, n° 4, avril 1987, pp. 453–457.

[60] Yee (K.S.). – Numerical solution of initial boundary value problem in isotropic media. *IEEE T.A.P.*, vol. AP-14, n° 3, 1966, pp. 302–307.

Index