Quelques méthodes performantes pour la simulation des phénomènes de propagation et de diffraction d'ondes

Sebastien Pernet

Habilitation à Diriger des Recherches

Université Toulouse III - Paul Sabatier

Spécialité : Mathématiques appliquées

Quelques méthodes performantes pour la simulation des phénomènes de propagation et de diffraction d'ondes

par

Sébastien Pernet

soutenue le 11/12/2017 devant le jury composé de

François Alouges	Professeur Ecole Polytechnique	examinateur
Xavier Antoine	Professeur IECL, Nancy	rapporteur
Hélène Barucq	DR INRIA MAGIQUE3D, Pau	rapporteur
Abderrahmane Bendali	Professeur Emerite, INSA Toulouse	examinateur
Christophe Besse	Professeur université Paul Sabatier, Toulouse	examinateur
Anne-Sophie Bonnet-BenDhia	DR CNRS, POEMS, ENSTA Paristech	examinateur
Christophe Geuzaine	Professeur université de Liège	rapporteur
Jean-Pierre Raymond	Professeur université Paul Sabatier, Toulouse	examinateur

TABLE DES MATIÈRES

Introduction générale

Ce document contient trois parties : une partie scientifique présentant une synthèse de mes travaux de recherche les plus significatifs ainsi que des perspectives, une partie administrative donnant des éléments sur ma carrière, et enfin, en annexe, il y a cinq de mes publications. Ces travaux sont bien évidemment le résultat de collaborations ainsi que d'encadrements de thèse et de stages. Ces contributeurs seront nommés dans la partie bilan de chaque chapitre de la partie scientifique. Le thème général de mes travaux est la recherche, l'analyse et la mise en oeuvre de méthodes numériques pour la simulation des phénomènes de propagation et diffraction d'ondes acoustiques, électromagnétiques et élastiques. Je tiens à préciser que le contexte professionnel dans lequel j'évolue a naturellement orienté de façon significative mes recherches. En effet, que ce soit au CERFACS ou à l'ONERA, la recherche est en partie guidée par des problématiques remontées par des industriels (au CERFACS par exemple Dassault, MBDA, AIRBUS, TOTAL, CNES) ou des départements métiers à l'ONERA (électromagnétisme et radar par exemple). Cet environnement est clairement un atout pour un mathématicien appliqué.

La partie scientifique est découpée en trois chapitres correspondant chacun à des problématiques physiques différentes et donc à des difficultés numériques spécifiques.

La première partie s'intéresse à la recherche de méthodes performantes pour la simulation des phénomènes de propagation d'ondes électromagnétiques transitoires. Par performante, on entend capable de modéliser la complexité des scènes (haute fréquence, géométrie, physique complexe $\cdots$ etc $\cdots$), de garantir la précision de la solution et d'avoir une complexité de calcul leur permettant d'être compétitives dans le futur. J'ai débuté ces travaux durant ma thèse au sein du Département Electromagnétisme et Radar de l'ONERA Toulouse sous la direction de Gary Cohen et Xavier Ferrières. Nous nous intéressions à des problèmes de compatibilité électromagnétique. Il s'agissait de déterminer les champs électromagnétiques induits par une agression large bande proche des structures. Ceci se modélise en utilisant les équations de Maxwell dans le domaine temporel. On était au début des années 2000 et la méthode numérique la plus répandue dans le monde industriel était basée sur le schéma de Yee qui correspond à une discrétisation de type différences finies d'ordre deux. Pour pallier aux défauts de cette méthode (difficultés de simuler des phénomènes en temps long et à tenir compte de la géométrie des objets), nous avons proposé une méthode de Galerkin discontinue en utilisant des éléments spectraux et une formulation mixte. Le schéma qui en découle est clairement non standard et produit des solutions numériques de haute précision à un coût réduit par rapport à une approche classique. J'ai poursuivi ces travaux au CERFACS après ma thèse en collaboration étroite avec Xavier Ferrières avec pour objectif d'améliorer notre méthode en terme de performance afin de décrire fidèlement des scènes complexes. En effet, les maillages non-structurés utilisés dans ces situations possèdent inévitablement des zones sur-raffinées qui peuvent induire une grande disparité de taille entre les cellules. La condition de stabilité du schéma temporel utilisé implique donc un pas de temps très petit et un calcul prohibitif. Ceci est d'autant plus pénalisant que nous utilisons des ordres élevés et que cette condition se dégrade avec l'ordre d'approximation. Afin d'améliorer les performances de la méthode, nous avons proposé une stratégie basée sur une méthode de pas de temps local combinée avec l'utilisation d'une approximation non-conforme en espace et en degré polynomial. Ces travaux ont été réalisés en partie dans le projet de recherche fédérateur MAHPSO (dénommé Modèles d'Approximation de Haute Précision pour les Systèmes de propagation d'Ondes) piloté par l'équipe Modélisation Mathématique et Simulation Numérique de l'ONERA dirigé par Francois Rogier. Parallèlement, j'ai réalisé une analyse d'erreur *a posteriori* de notre schéma lors du projet REI DIGATOP. Notre objectif a été d'apporter un outil quantitatif permettant de s'assurer de la qualité de la solution numérique et le cas échéant de savoir où il faut

agir (raffinement et déraffinement de maillage, augmentation de l'ordre de l'approximation, raffinement en temps) pour améliorer la précision globale de la solution. Enfin, j'évoquerai à la fin de cette partie des résultats que nous avons obtenus dans la thèse de Yohann Dudouit que j'ai co-dirigé avec Luc Giraud (INRIA) et dans laquelle nous avons apporté une solution à une problématique de TOTAL qui est la détermination de l'effet d'un réseau d'hydrofractures sur la propagation élastique. Bien que visant un domaine d'application différent, ces résultats me semblent clairement connexes aux précédents et rentrent dans un thème plus global de recherche de méthodes performantes pour simuler la propagation des ondes transitoires.

La deuxième partie concerne la recherche de méthodes intégrales performantes en terme de coût de calcul et de précision pour simuler la diffraction d'ondes électromagnétiques en régime harmonique. Il s'agit du domaine dans lequel je suis le plus actif actuellement. J'ai débuté ces travaux lors de mon arrivée dans l'équipe électromagnétisme et acoustique du CERFACS en m'intéressant à des problèmes de furtivité dans le cadre d'un contrat avec Dassault. Il s'agissait de proposer des méthodes de calcul rapide de la signature radar d'un objet. Le but étant de prévoir avant la phase de conception si oui ou non la forme et les matériaux utilisés pour construire cet objet assurent une bonne discrétion de ce dernier en milieu hostile. Ce type d'étude se fait dans le domaine fréquentiel et la plupart du temps en utilisant des équations intégrales. Il m'a alors été demandé de proposer de nouvelles formulations intégrales pour le calcul de la diffraction d'ondes électromagnétiques par un obstacle avec condition d'impédance variable. Cette condition ayant pour but de modéliser l'effet de peintures appliquées à la surface des avions de combat afin d'absorber les ondes radar et ainsi d'éviter leur détection. Nous avons tout d'abord proposé et analysé une formulation intégrale de type CFIE efficace pour la prise en compte de la condition de Léontovich qui a ensuite été étendue à des conditions aux limites impédantes d'ordre élevé. Ces travaux m'ont amené à m'intéresser à une nouvelle approche pour le préconditionnement des équations intégrales. En effet, de nombreux travaux (par exemple les travaux de J.C. Nédélec, S. Christiansen, D. Levadoux, X. Antoine, M. Darbas pour les objets parfaitement conducteurs) ont montré que cette problématique peut être traitée en amont de la discrétisation. J'ai réalisé alors une extension de ce type de méthodologie pour les objets imparfaitement conducteurs en proposant une nouvelle formulation intégrale intrinsèquement bien conditionnée. Ces travaux m'ont permis d'initier un nouveau thème de recherche dans l'équipe, ainsi qu'une collaboration étroite avec David Levadoux de l'ONERA spécialiste dans le domaine. Cette collaboration a en particulier abouti au projet ANR ARTHEMIS (Méthode de décomposition de domaine et solveurs bien conditionnés en électromagnétisme) (CERFACS-ONERA-Ecole Polytechnique) dans lequel nous avons étendu en particulier nos résultats aux problèmes de transmission. Les méthodes développées ont aussi été partie prenante dans la réussite de contrats industriels pour le CNES et MBDA. Tous ces travaux ont concerné la recherche de nouvelles formulations intégrales en vue d'une résolution plus efficace de problèmes réalistes. Bien que l'utilisation de ces méthodes soit incontournable pour de nombreuses applications, comparativement aux méthodes de type éléments finis, elles demeurent pourtant insuffisamment popularisées et leur utilisation reste généralement l'affaire de spécialistes expérimentés. Nous pensons que l'un des obstacles à une plus grande exploitation de ces méthodes est l'absence d'outils automatiques, fournissant entre autre une aide à l'utilisateur pour réaliser ces simulations en automatisant la création d'un espace d'approximation adapté à son problème en terme de précision tout en optimisant la taille du système à résoudre. Nous avons donc développé un nouveau thème de recherche sur l'analyse d'erreur *a posteriori* et le raffinement auto-adaptatif avec nos partenaires industriels via le projet ANR RAFFINE et la thèse de Marc Bakry. Ces techniques sont quasiment inexistantes dans le domaine des équations intégrales, alors que leur importance n'a fait que croître dans le domaine des méthodes de type éléments finis. Je tiens à préciser que par "quasiment inexistantes", nous voulons dire qu'il y a très peu de travaux pour les méthodes basées sur des noyaux oscillants, néanmoins nous savons que l'équipe de Dirk Preatorius en Autriche travaille activement sur ce sujet pour des problèmes statiques et

a produit de nombreux résultats très intéressants. Notre contribution la plus importante a été de proposer une technique originale de localisation permettant de dériver des indicateurs d'erreur *a posteriori* fiables et efficaces pour une grande classe de problèmes.

La troisième partie s'intéresse au calcul de la propagation du son dans un écoulement complexe en régime harmonique. En 2006, Anne-Sophie Bonnet-BenDhia a pris la direction de l'équipe électromagnétisme et acoustique du CERFACS et par la même occasion a souhaité développer un thème de recherche sur l'aéroacoustique. J'ai donc diversifié mes activités en m'intéressant plus particulièrement aux nuisances sonores produites par les turboréacteurs d'un avion. Les mesures montrent que ces dernières sont causées par un bruit de raies (ou bruit tonal) qui se caractérise par des spectres ayant des pics d'amplitude à des fréquences isolées correspondant aux harmoniques de la fréquence de passage des aubes. La détermination du rayonnement acoustique pour ces fréquences spécifiques représente de toute évidence un enjeu pour la réduction des nuisances sonores. Notre objectif principal a été de répondre aux deux questions suivantes :

1. Faut-il tenir compte du couplage entre la propagation d'onde et le transport hydrodynamique lorsque l'on calcule le rayonnement acoustique en écoulement?
2. Si oui, comment réaliser cela?

En effet, ce couplage est souvent négligé dans les applications industrielles bien évidemment pour limiter les coûts de calcul mais aussi, à cause du traitement de domaines infinis qui est toujours sujet à des difficultés car la présence d'un écoulement non uniforme (et notamment l'existence de modes de vorticité et d'entropie) dissymétrise le problème : les conditions doivent savoir gérer différemment les bords amont et aval, ce qui rend l'écriture de condition de rayonnement complexe. C'est pour que cela la plupart des outils numériques commerciaux traitant le régime harmonique, se placent dans des cas d'écoulements simples. Ils supposent que la vitesse de l'écoulement porteur est à rotationnel nul (on parle d'écoulement irrotationnel ou potentiel). Dans le projet ANR AEROSON et la thèse d'Emilie Peynaud, nous avons proposé une méthode basée sur le modèle de Galbrun qui quant à lui est une alternative intéressante au système d'Euler linéarisé. Ce modèle a été initialement proposé par le projet POEMS (INRIA/CNRS) pour des écoulements simplifiés et nous avons réalisé une extension à des écoulements complexes. Dans notre cas, l'utilisation d'une formulation augmentée par l'introduction d'une inconnue hydrodynamique pour stabiliser la discrétisation du modèle permet de réintroduire explicitement la distinction entre l'amont et l'aval de l'écoulement, ce qui est seulement implicite dans une formulation variationnelle classique et donc de faire fonctionner des couches PML. La discrétisation du problème augmenté est basée sur une approximation par des éléments finis classiques pour le déplacement et sur une approche de Galerkin discontinue pour l'inconnue hydrodynamique.

PARTIE I

PARTIE SCIENTIFIQUE

CHAPITRE 1

RECHERCHE DE MÉTHODES PERFORMANTES POUR LA SIMULATION DES PHÉNOMÈNES DE PROPAGATION D'ONDES ÉLECTROMAGNÉTIQUES TRANSITOIRES

Ce premier chapitre concerne la recherche de méthodes performantes pour la simulation des phénomènes de propagation d'ondes électromagnétiques transitoires. Je présente tout d'abord une méthode de type Galerkin discontinue basée sur des éléments spectraux et un choix non standard pour l'espace d'approximation. Ces choix permettent de calculer des approximations numériques de haute précision des champs électromagnétiques à un coût réduit par rapport à une approche classique. J'évoque ensuite une stratégie de pas de temps local combinée avec des approximations non-conformes en espace et en ordre d'approximation qui permet d'améliorer notre méthode en terme de performances afin de décrire fidèlement des scènes complexes. Dans la partie suivante, je donne des résultats d'une analyse d'erreur *a posteriori* de notre schéma. Cette dernière a conduit à un indicateur d'erreur *a posteriori* de type résidu qui permet de s'assurer de la qualité de la solution numérique et le cas échéant de savoir où il faut agir (raffinement et déraffinement de maillage, augmentation de l'ordre de l'approximation, raffinement en temps) pour améliorer la précision globale de la solution. Je termine ce chapitre en présentant des résultats que nous avons obtenus en élastodynamique dans la thèse de Yohann Dudouit. Bien qu'en dehors du thème électromagnétisme, les méthodologies peuvent être appliquées aux deux physiques et de plus, ces travaux rentrent dans un thème plus global de recherche de méthodes performantes pour simuler la propagation des ondes transitoires. Enfin, la dernière partie de ce chapitre présente un bilan de ces travaux et propose quelques perspectives de travail.

1.1. Introduction

Depuis longtemps, les équations des ondes en général et les équations de Maxwell en particulier, sont résolues dans le domaine fréquentiel par des méthodes d'éléments finis (FEM). Le constat est différent dans le domaine temporel où elles ont été principalement résolues par des méthodes de différences finies (FDM) d'ordre deux dont une des plus connues est le schéma de Yee [**Yee66**] (appelé aussi FDTD pour Finite Difference Time Domain method) utilisé en électromagnétisme depuis 1966.

Malheureusement, à cause de l'erreur de dispersion, les FDM d'ordre deux trouvent leur limite lors de simulations en temps long que l'on rencontre souvent dans les problèmes physiques soit parce que la longueur d'onde est petite par rapport à la taille du domaine de calcul, soit parce que les ondes sont piégées dans des cavités. En effet, le nombre de points nécessaires pour obtenir une solution précise augmente avec l'intervalle de résolution en temps (en terme de longueur d'onde). Une solution à ce problème semble être l'utilisation d'une FDM d'ordre élevé qui permet d'augmenter le pas de discrétisation en espace tout en gardant une bonne précision. Cependant avec ce type de méthode, il est très difficile de modéliser les géométries complexes à cause de la taille importante des cellules constituant la grille de discrétisation du domaine.

Les gens sont restés néanmoins longtemps réticents à l'utilisation des FEM dans le domaine temporel (appelés FETD pour Finite Element Time Domaine method) qui permet pourtant d'assurer une bonne approximation de ces géométries. La raison venait principalement de la présence d'une matrice de masse qui est naturellement diagonale pour les FDM, mais n-diagonale pour les FEM, avec n qui augmente avec la dimension d'espace et l'ordre de la méthode. Cette matrice doit être inversée à chaque pas de temps et ralentit donc les performances de la méthode et cela même si on utilise un algorithme itératif pour l'inversion.

Une réponse à cette difficulté a été proposée en adoptant deux approches différentes. Une première introduite par G. Cohen *et al.* [**CJT94**] pour l'équation des ondes, est basée sur une technique de condensation de masse pour les FEM sur des maillages en quadrilatères et en hexaèdres en utilisant des points de Gauss-Lobatto. En fait, cette idée a été tout d'abord utilisée pour la simulation de réservoir [**You78**] et la neutronique [**HSV79**] et a été ensuite étendue à d'autres problèmes sous le nom "d'éléments spectraux" [**MP89**]. Cette méthode a alors été combinée aux formulations mixtes, tout d'abord en électromagnétisme (utilisant une formulation $H(curl)$-L^2) [**CM99**] puis à l'acoustique [**CF00, CF05**] et conduit à des approximations efficaces et peu coûteuses en mémoire. Ces approximations sont appelées "méthodes d'éléments spectraux mixtes".

Ma thèse a d'ailleurs débuté par une étude détaillée de la méthode proposée par G. Cohen et P. Monk [**CJT94**] dans le but de l'appliquer aux problèmes de compatibilité électromagnétique. Nous avons choisi de travailler avec des éléments finis hexaédriques plutôt que tétraédriques car ils permettent une condensation de masse naturelle et une montée en ordre simple et efficace [**PFC05**]. En effet, la condensation de masse peut néanmoins être envisagée sur des tétraèdres (pour la discrétisation des équations de Maxwell) mais les éléments d'ordre 3 conduisent par exemple, à précision égale, à plus de degrés de liberté que les éléments d'ordre 2 [**EJ97**]. Les espaces d'éléments finis utilisés sont construits à partir de la seconde famille de Nédélec [**Né86**]. En gardant la conformité H-rot **localement** sur chaque cellule pour l'approximation L^2 (ce qui n'est pas classique) et en construisant les fonctions de base à partir des points de quadrature de Gauss-Lobatto, on aboutit au système:

$$\begin{cases} M_\varepsilon \dfrac{d\mathbf{E}}{dt} - R_h\mathbf{H} + M_\sigma\mathbf{E} + \mathbf{J} = 0 \\ M_\mu \dfrac{d\mathbf{H}}{dt} + R_h^*\mathbf{E} = 0 \end{cases}$$

où

- M_ε, M_σ : matrices diagonales par bloc $N \times N$ où N est le nombre de degrés de liberté autour d'un point d'interpolation,
- M_μ : matrice diagonale par bloc 3×3,
- R_h : matrice très creuse qui nécessite uniquement un stockage sur l'élément de référence.

Nous obtenons donc un important gain de stockage et substantiel gain en temps de calcul.
Cette méthode a été évaluée par comparaison systématique avec des méthodes éprouvées dans le domaine temporel (FDTD et volumes finis) et pour différents types de problèmes (cavités, géométries courbes, problèmes de grandes tailles). Les premiers résultats en 2D ont montré l'intérêt d'utiliser cette méthode du point de vue de la précision, du stockage et du temps de calcul [**PFC05**]. Nous avons donc décidé de passer au cas 3D. Nous avons alors vite été confrontés à la problématique du maillage. Les mailleurs de type hexaédrique sont en effet peu courant. C'est pourquoi nous avons décidé de construire nos maillages à partir de maillages tétraédriques dont on découpe chaque tétraèdre en quatre hexaèdres. Malheureusement, pour ce type de maillage, des comparaisons numériques ont montré que la solution obtenue par la méthode d'éléments finis est entachée d'ondes "parasites". Il n'existe pas à notre connaissance d'explication théorique de l'apparition de ces ondes pour cette formulation mixte. En effet les nombreux travaux portant sur le sujet (voir par exemple [**DNR78a, DNR78b, CFR00**] et les travaux de Boffi *et al.* [**Bof00, Bof01, BCD$^+$09, BFP99, BG02**]) s'intéressent principalement aux formulations du second ordre. Ils montrent néanmoins

que la seconde famille de Nédélec peut être sujette à des modes parasites lorsqu'elle est utilisée sur des quadrilatères ou des hexaèdres même réguliers. Une réponse partielle et numérique a été fournie par M. Duruflé [**Dur06**]. Il a en effet montré en résolvant sur un cube unité le problème aux valeurs propres associé à la formulation mixte considérée et en comparant les résultats avec la solution analytique que pour un maillage régulier les valeurs propres numériques correspondent bien à celles physiques mais avec un défaut de multiplicité (qui se traduit par une erreur d'amplitude) et que pour un maillage hexaédrique non-structuré, il y a bien apparition de modes non-physiques qui peuvent polluer sérieusement la solution temporelle. La méthode proposée reste néanmoins consistante sur ce type de maillage. Malheureusement, il faut augmenter de façon importante le nombre de point de discrétisation par longueur d'onde nécessaire à l'obtention d'une solution acceptable. On perd donc tout l'intérêt de la méthode.

La seconde approche pour éviter l'inversion d'une matrice de masse n-diagonale a été proposée par Hesthaven *et al.* [**HT00, HW08**]. Elle est basée sur l'utilisation d'une méthode de Galerkin Discontinue (DGM pour Discontinuous Galerkin Method) sur maillages triangulaires ou tétraédriques et conduit naturellement à des matrices de masse diagonales par bloc. L'utilisation de l'ordre élevé est essentielle pour les équations de Maxwell car elles produisent des ondes parasites qui doivent être supprimées ou contrôlées en ajoutant un terme de pénalisation à la formulation. Ce terme est dissipatif et l'approche d'ordre élevé réduit de façon substantielle cet effet. De plus, Hesthaven a réussi à réduire de façon importante la mémoire nécessaire à ce type d'approche induisant une augmentation raisonnable du temps calcul en utilisant une reconstruction à chaque pas de temps de la matrice de rigidité. Enfin, il a calculé des points d'interpolation quasi-optimaux afin d'éviter les phénomènes de Runge pouvant apparaitre lors de l'utilisation de polynômes de degré élevé pour construire l'approximation [**Hes98**]. Il faut tout de même noter que ces méthodes induisent plus de degrés de liberté qu'une approche de type éléments finis.

Ces deux approches nous ont suggéré de construire une DGM en utilisant des éléments spectraux et une formulation mixte. Dans ce chapitre, nous décrirons premièrement brièvement cette méthode en précisant ses propriétés. Nous présenterons ensuite les améliorations que nous avons proposées afin d'assurer de bonnes performances dans de nombreuses situations ainsi que des résultats connexes obtenus en élastodynamique. Nous terminerons par un bilan sur ces travaux de recherche et nous donnerons quelques perspectives.

1.2. Méthode de Galerkin Discontinue pour l'électromagnétisme transistoire

Dans cette partie, nous allons décrire la DGM que nous avons proposée. Cette approche pouvant s'appliquer dans un cadre plus large (acoustique, élasticité), nous allons donc la présenter brièvement à partir d'un problème hyperbolique linéaire abstrait.

1.2.1. Problème et notations. — Soit A un opérateur différentiel linéaire du premier ordre tel que $D(A) \subset \left[L^2(\Omega)\right]^q$ et $Im(A) \subset \left[L^2(\Omega)\right]^p$ où $D(A)$ et $Im(A)$ sont respectivement le domaine et l'image de A, Ω un domaine ouvert de $\mathbb{R}^d$ et A^* l'opérateur adjoint de A. On peut décomposer A de la façon suivante:

$$A = \sum_{i=1}^{d} A_i \frac{\partial}{\partial x_i}, \tag{1}$$

où A_i est une matrice réelle de taille $p \times q$.

Prenons $\underline{\psi} \in \mathcal{D}(\Omega)^q$ et $\underline{\psi'} \in \mathcal{D}(\Omega)^p$ (où $\mathcal{D}(\Omega)$ est l'espace des fonctions qui sont C^∞ et à support compact dans Ω). En utilisant une intégration par parties dans chaque direction, nous obtenons :

$$\begin{aligned} \int_\Omega A\underline{\psi} \cdot \underline{\psi'} \, \mathrm{d}\underline{x} &= \sum_{i=1}^{d} \int_\Omega A_i \frac{\partial \underline{\psi}}{\partial x_i} \cdot \underline{\psi'} \, \mathrm{d}\underline{x} = -\sum_{i=1}^{d} \int_\Omega \underline{\psi} \cdot \frac{\partial (A_i^T \underline{\psi'})}{\partial x_i} \, \mathrm{d}\underline{x} \\ &= -\int_\Omega \underline{\psi} \cdot \sum_{i=1}^{d} \left(\frac{\partial A_i^T}{\partial x_i} \underline{\psi'} + A_i^T \frac{\partial \underline{\psi'}}{\partial x_i} \right) \mathrm{d}\underline{x} = \int_\Omega \underline{\psi} \cdot A^* \underline{\psi'} \, \mathrm{d}\underline{x}, \end{aligned} \tag{2}$$

où A_i^T est la matrice transposée de A_i.

En particulier, lorsque A_i ne dépend pas de la variable d'espace, nous avons :

$$A^* = -\sum_{i=1}^{d} A_i^T \frac{\partial}{\partial x_i}. \tag{3}$$

Nous considérons à présent le problème hyperbolique suivant :

$$\underline{\underline{\lambda}}(\underline{x}) \frac{\partial \underline{u}}{\partial t}(\underline{x}, t) + A\underline{v}(\underline{x}, t) = \underline{f}(\underline{x}, t) \text{ dans } \Omega, \tag{4a}$$

$$\underline{\underline{\mu}}(\underline{x}) \frac{\partial \underline{v}}{\partial t}(\underline{x}, t) - A^* \underline{u}(\underline{x}, t) = 0 \text{ dans } \Omega, \tag{4b}$$

$$A^*(n)\, \underline{u}(\underline{x}, t) = 0 \text{ sur } \partial\Omega, \tag{4c}$$

$$\underline{u}(\underline{x}, 0) = \underline{u}_0(\underline{x}), \ \underline{v}(\underline{x}, 0) = \underline{v}_0(\underline{x}) \text{ dans } \Omega, \tag{4d}$$

où $\underline{u}(\underline{x}, t) \in I\!R^p$, $\underline{v}(\underline{x}, t) \in I\!R^q$, $\underline{f}(\underline{x}, t) \in I\!R^p$, $\underline{\underline{\lambda}}(\underline{x}) \in I\!R^p \times I\!R^p$ et $\underline{\underline{\mu}}(\underline{x}) \in I\!R^q \times I\!R^q$ sont des matrices symétriques definies positives, $\underline{n} = (n_i)_{i=1}^d$ est la normale unitaire extérieure à Ω, $A(\underline{n}) = \sum_{i=1}^{d} A_i \, n_i$, $A^*(\underline{n}) = -\sum_{i=1}^{d} A_i^T n_i$ et A ne dépend pas de la variable d'espace.

Nous pouvons réecrire (4) dans le cadre de l'application du théorème de Hille-Yosida [**Bre10**]

$$\frac{\partial \underline{w}}{\partial t} + B\, \underline{w} = F \tag{5}$$

où B est un opérateur maximal monotone défini par

$$B = \begin{bmatrix} 0 & \underline{\underline{\lambda}}^{-1} A \\ -\underline{\underline{\mu}}^{-1} A^* & 0 \end{bmatrix}$$

et le domaine de l'opérateur B est $D(B) = H_0(A^*\,, \Omega) \times H(A\,, \Omega) \subset H = L^2(\Omega)^{p+q}$ avec

$$H(A\,, \Omega) = \{\underline{w} \in L^2(\Omega)^q \,:\, A\,\underline{w} \in L^2(\Omega)^p\},$$

$$H_0(A^*\,, \Omega) = \{\underline{w} \in L^2(\Omega)^p \,:\, A^*\,\underline{w} \in L^2(\Omega)^q \text{ et } A^*(\underline{n})\underline{w} = 0 \text{ on } \Gamma\}.$$

De plus, nous supposons que l'opérateur $I + B$ est surjectif sur H et si $(\underline{u}_0, \underline{v}_0) \in D(B)$, nous obtenons que l'unique solution de (4) a les propriétés de régularité suivantes :

$$\underline{u} \in C^1([0, T], L^2(\Omega)^p) \cap C^0([0, T], H_0(A^*\,, \Omega)), \tag{7a}$$

$$\underline{v} \in C^1([0, T], L^2(\Omega)^q) \cap C^0([0, T], H(A\,, \Omega)). \tag{7b}$$

Remarques :

1. Les espaces $H(A\,,\,\Omega)$ et $H(A^*\,,\,\Omega)$ sont des espaces de Hilbert pour les produits scalaires:

$$(\underline{v}\,,\,\underline{w})_A = (\underline{v}\,,\,\underline{w})_0 + (A\underline{v}\,,\,A\underline{w})_0,$$

$$(\underline{v}\,,\,\underline{w})_{A^*} = (\underline{v}\,,\,\underline{w})_0 + (A^*\underline{v}\,,\,A^*\underline{w})_0.$$

2. Les opérateurs de trace

$$\begin{aligned} A(\underline{n}) \,:\, H(A\,,\,\Omega) &\rightarrow [H^{-1/2}(\partial\Omega)]^p \\ \underline{w} &\mapsto A(\underline{n})\underline{w}, \end{aligned}$$

$$\begin{aligned} A^*(\underline{n}) \,:\, H(A^*\,,\,\Omega) &\rightarrow [H^{-1/2}(\partial\Omega)]^q \\ \underline{w} &\mapsto A^*(\underline{n})\underline{w}, \end{aligned}$$

   sont linéaires, continus mais non nécessairement surjectifs[**Jen06**].

3. Nous avons la formule d'intégration par parties : pour tout $\underline{v} \in H(A^*\,,\,\Omega)$ et $\underline{w} \in H^1(\Omega)^q$,

$$< A^*(\underline{n})\underline{v}\,,\,\underline{w} >= \int_\Omega A^*\underline{v}\cdot\underline{w}\,\mathrm{d}\underline{x} - \int_\Omega \underline{v}\cdot A\underline{w}\,\mathrm{d}\underline{x},$$

   où $<,>$ est le crochet de dualité entre $H(A^*\,,\,\Omega)$ et $[H^{-1/2}(\partial\Omega)]^q$. Cette dernière égalité peut être étendue à $H(A^*\,,\,\Omega)\times H(A\,,\,\Omega)$.

Considérons à présent un maillage d'un domaine ouvert (polyhédrique) Ω

$$\mathcal{T}_h = \bigcup_{\ell=1}^{N_e} K_\ell, \tag{10}$$

où K_ℓ est un élément de n'importe quelle forme.

Le principe d'une méthode de Galerkin discontinue est de chercher une solution définie sur ce maillage en ne lui imposant aucune continuité à travers les faces de $\mathcal{T}_h$. La solution est donc cherchée dans un sous-espace de L^2. Cependant, ce défaut de continuité est compensé par l'introduction de sauts des traces des inconnues à travers les faces du maillage. Par souci de simplicité de l'exposé, nous supposons une régularité supplémentaire (voir la première remarque ci-dessous) sur la solution exacte (4) afin d'assurer que les traces $A^*(\underline{n}_\Gamma)\underline{u}$ et $A(\underline{n}_\Gamma)\underline{v}$ appartiennent à L^2 pour toutes les faces Γ du maillage $\mathcal{T}_h$, $\underline{n}_\Gamma$ étant une normale unitaire à Γ. Nous définissons alors l'espace

$$H^s(\mathcal{T}_h) = \left\{ u \in L^2(\Omega) \text{ tel que, } u_{|_{K_\ell}} \in H^s(K_\ell) \right\}, \tag{11}$$

où $H^s(K_\ell)$ est l'espace de Sobolev d'ordre $s > \frac{1}{2}$. L'espace $H^s(\mathcal{T}_h)$ est muni de la norme

$$||u||_{s,h} = \left(\sum_{K_\ell\in\mathcal{T}_h} ||u||^2_{s,\ell} \right)^{\frac{1}{2}}, \tag{12}$$

où $||\cdot||_{s,\ell}$ est la norme usuelle de $H^s(K_\ell)$.

Pour une fonction vectorielle $\underline{v} \in \underline{H}^s_n(\mathcal{T}_h) = [H^s(\mathcal{T}_h)]^n$, nous definissons la norme :

$$||\underline{v}||_{s,h} = \left(\sum_{i=1}^{n} ||v_i||^2_{s,h} \right)^{\frac{1}{2}}. \tag{13}$$

Par conséquent, nous supposons qu'il existe $s > \dfrac{1}{2}$,

$$\underline{u}(\cdot,t) \in H_0(A^*\,,\,\Omega) \cap \underline{H}^s_p(\mathcal{T}_h),\ \forall t, \tag{14a}$$

$$\underline{v}(\cdot,t) \in H(A\,,\,\Omega) \cap \underline{H}^s_q(\mathcal{T}_h),\ \forall t. \tag{14b}$$

Nous définissons maintenant respectivement par $\mathcal{F}_I$ et $\mathcal{F}_B$ les ensembles des faces (arêtes en 2D) intérieures et extérieures du maillage $\mathcal{T}_h$. Sur chaque face $\Gamma_{\ell,m} = K_\ell \cap K_m \in \mathcal{F}_I$, nous définissons par :

$$[\underline{v}]^{K_\ell}_{\Gamma_{\ell,m}} = (\underline{v}_{|K_m})_{|\Gamma_{\ell,m}} - (\underline{v}_{|K_\ell})_{|\Gamma_{\ell,m}} \tag{15}$$

le saut de $\underline{v}$ à travers $\Gamma_{\ell,m}$, $(\underline{v}_{|K_\ell})_{|\Gamma_{\ell,m}}$ étant la trace de $\underline{v}_{|K_\ell}$ sur $\Gamma_{\ell,m}$. De plus, sur une face $\Gamma = \partial K \cap \partial\Omega \in \mathcal{F}_B$, nous définissons:

$$[\underline{v}]^{K}_{\Gamma} = -(\underline{v}_{|K})_{|\Gamma}. \tag{16}$$

Remarque : En pratique, $\underline{v}_{|K_\ell}$ appartient (ou dérive) généralement à (d') un espace polynomial et est donc localement très régulière.

1.2.2. Formulation de Galerkin discontinue. — En utilisant les notations définies précédemment, nous introduisons le problème suivant :

Trouver $\underline{u}(.,t) \in U_p$, $\underline{v}(.,t) \in U_q$ tel que $\forall K_\ell \in \mathcal{T}_h$ et $\forall \underline{\varphi} \in U_p$, $\forall \underline{\psi} \in U_q$,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{K_\ell} \underline{\underline{\lambda}}\,\underline{u}_\ell \cdot \underline{\varphi}_\ell \,\mathrm{d}\underline{x} = &- \int_{K_\ell} A\underline{v}_\ell \cdot \underline{\varphi}_\ell \,\mathrm{d}\underline{x} + \int_{\partial K_\ell} \alpha\, A(\underline{n})[\underline{v}]^{K_\ell}_{\partial K_\ell} \cdot \underline{\varphi}_\ell \,\mathrm{d}\sigma \\ &+ \int_{\partial K_\ell} \gamma\, C[\underline{u}]^{K_\ell}_{\partial K_\ell} \cdot \underline{\varphi}_\ell \,\mathrm{d}\sigma + \int_{K_\ell} \underline{f} \cdot \underline{\varphi}_\ell \,\mathrm{d}\underline{x},\end{aligned} \tag{17a}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{K_\ell} \underline{\underline{\mu}}\,\underline{v}_\ell \cdot \underline{\psi}_\ell \,\mathrm{d}\underline{x} = &\int_{K_\ell} A^*\underline{u}_\ell \cdot \underline{\psi}_\ell \,\mathrm{d}\underline{x} + \int_{\partial K_\ell} \beta\, A^*(\underline{n})[\underline{u}]^{K_\ell}_{\partial K_\ell} \cdot \underline{\psi}_\ell \,\mathrm{d}\sigma \\ &+ \int_{\partial K_\ell} \delta\, C'[\underline{v}]^{K_\ell}_{\partial K_\ell} \cdot \underline{\psi}_\ell \,\mathrm{d}\sigma,\end{aligned} \tag{17b}$$

$$\underline{u}(\underline{x},0) = \underline{u}_0(\underline{x}),\ \underline{v}(\underline{x},0) = \underline{v}_0(\underline{x}) \text{ in } \Omega, \tag{17c}$$

où $\underline{u}_\ell = \underline{u}_{|K_\ell}$, $\underline{v}_\ell = \underline{v}_{|K_\ell}$, $\underline{\varphi}_\ell = \underline{\varphi}_{|K_\ell}$, $\underline{\psi}_\ell = \underline{\psi}_{|K_\ell}$. De plus, $C = A(\underline{n})\,A^*(\underline{n})$ et $C' = A^*(\underline{n})\,A(\underline{n})$ sont des matrices symétriques positives, α, β, γ, δ sont des fonctions constantes par morceaux sur chaque face de $\mathcal{T}_h$ et U_p et U_q sont des espaces fonctionnels à définir.

Remarque : Dans (17a) et (17b), les intégrales contenant C et C' peuvent être vues comme des termes de pénalisation qui imposent les conditions de transmission $A(\underline{n})[\underline{v}]^{K_\ell}_{\partial K_\ell} = 0$ et $A^*(\underline{n})[\underline{u}]^{K_\ell}_{\partial K_\ell} = 0$ en un sens de type moindres carrés. En particulier, ces termes sont consistants avec la solution exacte puisqu'ils sont nuls pour cette dernière.

Avant d'expliciter le problème approché, nous présentons une proposition qui donne des conditions pour lesquelles la formulation (17a)-(17c) est consistante avec le problème fort. En effet, nous avons

Proposition 1. — *Pour $U_p = \underline{H}^s_p(\mathcal{T}_h)$, $U_q = \underline{H}^s_q(\mathcal{T}_h)$ et si α, β, γ, δ vérifient*

1. *γ, $\beta \neq 0$ sur $\mathcal{F}_I$,*

2. *$\alpha = \delta = 0$ et γ ou $\beta \neq 0$ sur $\mathcal{F}_B$,*

alors les problèmes (4a)-(4d) et (17a)-(17c) sont équivalents.

"L'approximation" GD (17a)-(17c) est définie sous une forme continue dans la proposition ci-dessus, en prenant $U_p = \underline{H}^s_p(\mathcal{T}_h)$ et $U_q = \underline{H}^s_q(\mathcal{T}_h)$. Cependant, en pratique la discrétisation est réalisée en utilisant des espaces d'approximation de dimension finie $\underline{V}_{h,n} \subset \underline{H}^s_n(\mathcal{T}_h)$ avec $n = p, q$ qui seront précisés ci-dessous pour les problèmes en électromagnétisme. Nous renvoyons à **[CP16]** pour d'autres exemples.

Enfin, une condition nécessaire pour avoir un problème (17a)-(17c) bien-posé est l'existence d'une énergie discrète $\mathcal{E}_h$ du système qui ne soit pas croissante. Dans notre contexte, cette énergie est définie de la façon suivante. Puisque $\underline{\underline{\lambda}}$ et $\underline{\underline{\mu}}$ sont des matrices symétriques, définies positives, on peut donc écrire $\underline{\underline{\lambda}} = \tilde{\underline{\underline{\lambda}}}^T \tilde{\underline{\underline{\lambda}}}$ et $\underline{\underline{\mu}} = \tilde{\underline{\underline{\mu}}}^T \tilde{\underline{\underline{\mu}}}$. Alors

$$\underline{\underline{\lambda}} \frac{\partial \underline{u}_\ell}{\partial t} \cdot \underline{u}_\ell = \frac{\partial (\tilde{\underline{\underline{\lambda}}} \underline{u}_\ell)}{\partial t} \cdot \tilde{\underline{\underline{\lambda}}} \underline{u}_\ell = \frac{1}{2} \frac{\partial ||\tilde{\underline{\underline{\lambda}}} \underline{u}_\ell||^2}{\partial t} \tag{18}$$

et, de la même manière

$$\underline{\underline{\mu}} \frac{\partial \underline{v}_\ell}{\partial t} \cdot \underline{v}_\ell = \frac{1}{2} \frac{\partial ||\tilde{\underline{\underline{\mu}}} \underline{v}_\ell||^2}{\partial t} \tag{19}$$

Finalement,

$$\mathcal{E}_h = \sum_{K_\ell \in \mathcal{T}_h} \int_{K_\ell} \left(||\tilde{\underline{\underline{\lambda}}} \underline{u}_\ell||^2 + ||\tilde{\underline{\underline{\mu}}} \underline{v}_\ell||^2 \right) \mathrm{d}\underline{x} \tag{20}$$

Des calculs élémentaires nous permettent d'écrire :

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{E}_h = \frac{\mathrm{d}}{\mathrm{d}t} \sum_{\Gamma_{\ell,m} \in \mathcal{F}_I} R_E^{\ell,m} + \frac{\mathrm{d}}{\mathrm{d}t} \sum_{\Gamma_\ell \in \mathcal{F}_B} R_E^\ell. \tag{21}$$

où

$$\begin{aligned} R_E^{\ell,m} &= (1 + \alpha - \beta) \int_{\Gamma_{\ell,m}} A(\underline{n}) (\underline{v}_m \cdot \underline{u}_m - \underline{v}_\ell \cdot \underline{u}_\ell) \, \mathrm{d}\sigma \\ &+ (\alpha + \beta) \int_{\Gamma_{\ell,m}} A(\underline{n}) (\underline{v}_m \cdot \underline{u}_l - \underline{v}_\ell \cdot \underline{u}_m) \, \mathrm{d}\sigma \\ &- \gamma \int_{\Gamma_{\ell,m}} ||A^*(\underline{n}) (\underline{u}_m - \underline{u}_\ell)||^2 \, \mathrm{d}\sigma - \delta \int_{\Gamma_{\ell,m}} ||A(\underline{n}) (\underline{v}_m - \underline{v}_\ell)||^2 \, \mathrm{d}\sigma. \end{aligned} \tag{22}$$

et

$$R_E^\ell = (1 - \beta) \int_{\Gamma_\ell} A(\underline{n}) \underline{v}_m \cdot \underline{u}_m \, \mathrm{d}\sigma - \gamma \int_{\Gamma_\ell} ||A^*(\underline{n}) \underline{u}_m||^2 \, \mathrm{d}\sigma - \delta \int_{\Gamma_\ell} ||A(\underline{n}) \underline{v}_m||^2 \, \mathrm{d}\sigma. \tag{23}$$

Les relations (22) et (23) impliquent que, pour avoir $\mathrm{d}\mathcal{E}_h/\mathrm{d}t \leq 0$, on doit prendre :

1. $\alpha = -\beta = -1/2$ pour $\Gamma_\ell \in \mathcal{F}_I$,
2. $\beta = 1$ pour $\Gamma_\ell \in \mathcal{F}_B$,
3. Si $\gamma > 0$ et $\delta > 0$, on a $\mathrm{d}\mathcal{E}_h/\mathrm{d}t < 0$ et l'énergie décroit. Le schéma est alors *dissipatif*.
4. Si $\gamma = 0$ et $\delta = 0$, on a $\mathrm{d}\mathcal{E}_h/\mathrm{d}t = 0$ et l'énergie est constante. Le schéma est alors *conservatif* : nous avons la conservation de l'énergie au cours du temps.

Nous allons à présent nous focaliser sur les problèmes d'électromagnétisme transitoire qui correspondent au domaine dans lequel nous avons proposé un schéma original. Néanmoins, nous renvoyons à **[CP16]** pour des applications en acoustique et en élastodynamique. Nous présentons aussi dans ce livre plusieurs approximations pour ces problèmes.

1.2.3. Application aux équations de Maxwell transitoires. — Le système des équations de Maxwell dans le domaine temporel est de : trouver $\underline{E}$, $\underline{H} : \Omega \to I\!R^3$ solution de

$$\underline{\underline{\varepsilon}} \frac{\partial \underline{E}}{\partial t}(\underline{x},t) - \nabla \times \underline{H}(\underline{x},t) = -\underline{J}(\underline{x},t), \tag{24a}$$

$$\underline{\underline{\mu}} \frac{\partial \underline{H}}{\partial t}(\underline{x},t) + \nabla \times \underline{E}(\underline{x},t) = 0. \tag{24b}$$

Les conditions initiales sont alors

$$\underline{E}(\underline{x},0) = \underline{E_0}(\underline{x}), \quad \underline{H}(\underline{x},0) = \underline{H_0}(\underline{x}). \tag{25}$$

Le système (24a)-(24b) s'écrit dans le cadre abstrait précédent en prenant : $p = 3$, $q = 3$, $\underline{u} = \underline{E}$, $\underline{v} = \underline{H}$, $A = A^* = \underline{curl}$,

$$A(\underline{n}) = A^*(\underline{n}) = \begin{pmatrix} 0 & n_3 & -n_2 \\ -n_3 & 0 & n_1 \\ n_2 & -n_1 & 0 \end{pmatrix}.$$

De plus, il vient

$$C = C' = \begin{pmatrix} n_2^2 + n_3^2 & -n_1 n_2 & -n_1 n_3 \\ -n_1 n_2 & n_1^2 + n_3^2 & -n_2 n_3 \\ -n_1 n_3 & -n_2 n_3 & n_1^2 + n_2^2 \end{pmatrix}.$$

En fait si $\underline{V} \in I\!R^3$, nous avons $A(\underline{n})\underline{V} = \underline{V} \times \underline{n}$ et $C\,\underline{V} = \underline{n} \times \underline{V} \times \underline{n}$. Ces définitions nous donnent la formulation GD suivante :

Trouver $\underline{E}_h(.,t) \in \underline{V}_{h,3} \subset \underline{H}^s_3(\mathcal{T}_h)$, $\underline{H}_h(.,t) \in \underline{V}_{h,3}$ tel que $\forall K_\ell \in \mathcal{T}_h$ et $\forall \underline{\varphi}_h \in \underline{V}_{h,3}$, $\forall \underline{\psi}_h \in \underline{V}_{h,3}$,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \int_{K_\ell} \underline{\underline{\varepsilon}}\,\underline{E}_\ell \cdot \underline{\varphi}_\ell \,\mathrm{d}\underline{x} = & \int_{K_\ell} \nabla \times \underline{H}_\ell \cdot \underline{\varphi}_\ell \,\mathrm{d}\underline{x} + \int_{\partial K_\ell} \alpha \,[\underline{H} \times \underline{n}]^{K_\ell}_{\partial K_\ell} \cdot \underline{\varphi}_\ell \,\mathrm{d}\sigma \\ & + \int_{\partial K_\ell} \gamma \,[\underline{n} \times \underline{E} \times \underline{n}]^{K_\ell}_{\partial K_\ell} \cdot \underline{\varphi}_\ell \,\mathrm{d}\sigma - \int_{K_\ell} \underline{J} \cdot \underline{\varphi}_\ell \,\mathrm{d}\underline{x}, \end{aligned} \tag{26a}$$

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \int_{K_\ell} \underline{\underline{\mu}}\,\underline{H}_\ell \cdot \underline{\psi}_\ell \,\mathrm{d}\underline{x} = & - \int_{K_\ell} \nabla \times \underline{E}_\ell \cdot \underline{\psi}_\ell \,\mathrm{d}\underline{x} + \int_{\partial K_\ell} \beta \,[\underline{E} \times \underline{n}]^{K_\ell}_{\partial K_\ell} \cdot \underline{\psi}_\ell \,\mathrm{d}\sigma \\ & + \int_{\partial K_\ell} \delta \,[\underline{n} \times \underline{H} \times \underline{n}]^{K_\ell}_{\partial K_\ell} \cdot \underline{\psi}_\ell \,\mathrm{d}\sigma, \end{aligned} \tag{26b}$$

$$\underline{E}_h(\underline{x},0) = \underline{E}_{0h}(\underline{x}),\ \underline{H}_h(\underline{x},0) = \underline{H}_{0h}(\underline{x}) \text{ in } \Omega \tag{26c}$$

avec

1. $\alpha = -\beta = -1/2$ et γ, $\delta \geq 0$ pour $\Gamma_\ell \in \mathcal{F}_I$,
2. $\beta = 1$, $\alpha = \delta = 0$ et $\gamma \geq 0$ pour $\Gamma_\ell \in \mathcal{F}_B$.

Remarques :

1. La formulation GD ci-dessus traite la condition aux limites de type conducteur parfait *i.e.* $\underline{E} \times \underline{n} = 0$ sur $\partial\Omega$.
2. Ces configurations englobent les formulations classiques utilisant les flux centrés et les flux décentrés. Une des différences entre ces formulations est la présence ou non de dissipation.

Fort de notre expérience précédente sur les éléments finis [**PFC05**], nous avons choisi de construire l'espace d'approximation $\underline{V}_{h,3}$ en utilisant des quadrilatères ou des hexaèdres et une approximation locale conforme H-rot. Ce choix n'est pas standard mais il permet de supprimer le stockage des matrices de rigidité et de sauts (en partie pour cette dernière).

Plus précisément, nous définissons tout d'abord un maillage $\mathcal{T}_h$ composé de quadrilatères ou d'hexaèdres qui seront notés K_ℓ et $\widehat{K} = [0,1]^3$ est l'élément de référence . $\underline{F}_\ell$ est la transformation telle que $K_\ell = \underline{F}_\ell(\widehat{K})$. L'espace $\underline{V}_{h,3}$ est alors défini par

$$\underline{V}_{h,3} = \left\{ \underline{w}_h \in [L^2(\Omega)]^3 \text{ such that } \forall K_\ell \in \mathcal{T}_h,\ DF_\ell^{-T} \underline{w}_{h|_{K_\ell}} \circ \underline{F}_\ell \in [Q_r]^3 \right\} \tag{27}$$

où DF_ℓ est la matrice jacobienne de F_ℓ.

Ce choix implique les deux propriétés importantes : pour tout $\underline{u} \in \underline{V}_{h,3}$ et $\forall K_\ell \in \mathcal{T}_h$

$$\begin{aligned} (\nabla \times \underline{u}) \circ \underline{F}_\ell &= (DF_\ell^{-T} \hat{\nabla}) \times (DF_\ell^{-T} \hat{\underline{u}}) \\ &= J_\ell^{-1} DF_\ell \, (\hat{\nabla} \times \hat{\underline{u}}), \end{aligned} \tag{28}$$

et

$$\begin{aligned} (\underline{n} \times \underline{u}) \circ \underline{F}_\ell &= \frac{DF_\ell^{-T} \hat{\underline{n}}}{||DF_\ell^{-T} \hat{\underline{n}}||} \times (DF_\ell^{-T} \hat{\underline{u}}) \\ &= \frac{J_\ell^{-1} DF_\ell}{||DF_\ell^{-T} \hat{\underline{n}}||} (\hat{\underline{n}} \times \hat{\underline{u}}). \end{aligned} \tag{29}$$

où $DF_\ell^{-T} \underline{u}_{|K_\ell} \circ \underline{F}_\ell = \hat{\underline{u}}$ et $J_\ell := det(DF_\ell)$.

En effet, nous avons alors que $\forall \underline{u},\ \underline{v} \in \underline{V}_{h,3}$,

$$\begin{aligned} \int_{K_\ell} \nabla \times \underline{u} \cdot \underline{v} \, \mathrm{d}\underline{x} &= \int_{\widehat{K}} |J_\ell| \, J_\ell^{-1} DF_\ell \, (\hat{\nabla} \times \hat{\underline{u}}) \cdot \, DF_\ell^{-T} \hat{\underline{v}} \, \mathrm{d}\hat{\underline{x}} \\ &= s_\ell \int_{\widehat{K}} (\hat{\nabla} \times \hat{\underline{u}}) \cdot \hat{\underline{v}} \, \mathrm{d}\hat{\underline{x}}, \end{aligned} \tag{30}$$

et

$$\int_{\partial K_\ell} \underline{n} \times \underline{u}_l \cdot \underline{v}_l \, \mathrm{d}\sigma = s_l \int_{\partial \widehat{K}} (\hat{\underline{n}} \times \hat{\underline{u}}) \cdot \hat{\underline{v}} \, \mathrm{d}\hat{\sigma} \tag{31}$$

où s_ℓ est le signe de J_ℓ qui est constant par cellule.

Les fonctions de base de $[Q_r]^3$ sur l'élément de référence sont $\hat{\varphi}_i \, \underline{e}_n$, $i = 1..(r+1)^3$, $n = 1..3$ où

$$\hat{\varphi}_i(\hat{x}_1, \hat{x}_2, \hat{x}_3) = \hat{\varphi}_{i_1}(\hat{x}_1) \, \hat{\varphi}_{i_2}(\hat{x}_2) \, \hat{\varphi}_{i_3}(\hat{x}_3) \tag{32}$$

avec $i = (r+1)[(r+1)(i_3 - 1) + i_2 - 1] + i_1$ et $\hat{\Xi} = (\hat{\xi}_j)_{j=1}^{r+1}$ est un ensemble de points de quadrature sur $[0,1]$ dont les poids sont $(\hat{\omega}_k)_{k=1}^{r+1}$ tel que

$$\forall i = 1..(r+1)^3,\ \forall j = 1..(r+1)^3,\ \hat{\varphi}_i(\hat{\underline{\xi}}_j) = \delta_{ij}, \tag{33}$$

où $\hat{\underline{\xi}}_j \in \hat{\Xi}^3$.

Nous avons deux choix naturels possibles pour les points d'intégration, ceux de Gauss et de Gauss-Lobatto. Les points de Gauss correspondent aux racines du polynôme de Legendre et sont donc tous situés à l'intérieur du segment $[0,1]$ tandis que ceux de Gauss-Lobatto contiennent toujours les deux extrémités 0 et 1 et les autres correspondent aux racines de la dérivée du polynôme de Legendre (notre choix est précisé à la fin de cette partie).

Nous obtenons alors la formulation semi-discrète

$$B_\varepsilon \frac{d\,\mathbf{E}}{dt} - R_h\,\mathbf{H} + D_h^\gamma\,\mathbf{E} + S_h^\alpha\,\mathbf{H} + \mathbf{J} = 0, \tag{34a}$$

$$B_\mu \frac{d\,\mathbf{H}}{dt} + R_h\,\mathbf{E} + S_h^\beta\,\mathbf{E} + D_h^\delta\,\mathbf{H} = 0. \tag{34b}$$

où

- B_ε, B_μ sont des matrices de masse diagonales par blocs dont la taille des blocs est 3×3,
- R_h est une matrice très creuse nécessitant un faible stockage,
- S_h^α, S_h^β sont des matrices de saut diagonales par blocs (avec des blocs plus larges si on choisit des degrés de liberté aux points de Gauss que s'ils sont localisés aux points de Gauss-Lobatto) nécessitant un faible stockage,
- D_h^γ, D_h^δ sont des matrices de saut symétriques diagonales par blocs qui nécessitent le stockage d'une matrice symétrique de taille 2×2 pour tous les éléments du maillage, ce qui correspond à un stockage additionnel raisonnable.

Les avantages principaux des points de Gauss sont :

1. Ils sont exacts pour les polynômes Q_{2r+1}, ce qui implique en particulier que les intégrales de masse peuvent être calculées exactement (lorsqu'il n'y a pas de fonction poids dépendant de la variable d'espace),
2. ils induisent de plus de très bonnes propriétés de dispersion numérique au schéma **[CFP06]**,

tandis que leurs principaux défauts sont :

1. les sauts sont calculés par une interpolation 1D,
2. tous les degrés de liberté produisent des sauts,
3. l'approximation en temps du terme dissipatif nécessite un décentrement en temps qui induites une réduction (raisonnable) de la condition de stabilité CFL.

D'un autre côté, les principaux avantages des points de Gauss-Lobatto sont:

1. Les sauts sont calculés sans interpolation puisque des degrés de liberté sont situés sur les faces,
2. seul les degrés de liberté sur les faces produisent des sauts,
3. on peut traiter les termes dissipatifs de manière centrée,

tandis que les principaux défauts sont :

1. ils sont exacts pour Q_{2r-1}, ce qui implique un calcul approché des termes de masse,
2. ils produisent un schéma plus dispersif.

Pour conclure, au vue des très bonnes propriétés de dispersion et de dissipation induit par l'utilisation de points de quadrature de Gauss, nous avons choisi de construire notre schéma à partir de ceux-ci. Néanmoins, on peut se poser la question suivante : est-ce que le coût calcul plus faible des points de Gauss-Lobatto ne permet-il pas de contrecarrer le défaut de précision de ce schéma?

1.3. Estimation d'erreur a priori

Dans cette partie, nous présentons les résultats de convergence *hp* obtenus pour notre schéma. Nous renvoyons à **[PF07]** pour les démonstrations. Les principales difficultés de cette étude sont que contrairement aux schémas basés sur des maillages tétraédriques, la transformation F_K n'est pas affine et les fonctions définies par la transformation présentée dans (27) ne sont pas nécessairement

polynomiales. Pour cette raison, des outils spécifiques ont dû être utilisés pour le cas de cellules hexaédriques.

Soient $(\mathbf{E}, \mathbf{H})$ et $(\mathbf{E}_h, \mathbf{H}_h)$ les solutions exacte et DG obtenues pour notre schéma. Notre but est d'estimer $\|\mathbf{E} - \mathbf{E}_h\|_{0,\Omega}$ et $\|\mathbf{H} - \mathbf{H}_h\|_{0,\Omega}$.
Pour cela, nous introduisons la norme énergie :

$$\|(\mathbf{E}, \mathbf{H})\|_*^2 = \|\mathbf{E}\|_{0,\Omega,\underline{\underline{\varepsilon}}}^2 + \|\mathbf{H}\|_{0,\Omega,\underline{\underline{\mu}}}^2. \tag{35}$$

avec $\|\mathbf{u}\|_{0,\Omega,\underline{\underline{\theta}}}^2 := \displaystyle\int_\Omega \underline{\underline{\theta}}\,\mathbf{u}\cdot\,\mathbf{u}\,\mathrm{d}x$ avec $\underline{\underline{\theta}} = \underline{\underline{\varepsilon}}$ ou $\underline{\underline{\mu}}$.

Cette norme (35) est bien adpatée pour nos estimations car elle apparaît naturellement lorsqu'on étudie l'énergie du système de Maxwell. Nous préférons donc estimer :

$$\|(\mathbf{E} - \mathbf{E}_h, \mathbf{H} - \mathbf{H}_h)\|_* = \sqrt{\|\mathbf{E} - \mathbf{E}_h\|_{0,\Omega,\underline{\underline{\varepsilon}}}^2 + \|\mathbf{H} - \mathbf{H}_h\|_{0,\Omega,\underline{\underline{\mu}}}^2}. \tag{36}$$

Nous introduisons une projection de la solution exacte $(\mathbf{E}, \mathbf{H})$ sur $\underline{V}_{h,3}$ *i.e.* $(\pi_h^1\mathbf{E}, \pi_h^1\mathbf{H})$ et il vient

$$\begin{aligned}
&\|(\mathbf{E} - \mathbf{E}_h, \mathbf{H} - \mathbf{H}_h)\|_*^2 = \\
&\|\mathbf{E} - \pi_h^1\mathbf{E} + \pi_h^1\mathbf{E} - \mathbf{E}_h\|_{0,\Omega,\underline{\underline{\varepsilon}}}^2 + \|\mathbf{H} - \pi_h^1\mathbf{H} + \pi_h^1\mathbf{H} - \mathbf{H}_h\|_{0,\Omega,\underline{\underline{\mu}}}^2 \\
&\leq \|\Delta_{\mathbf{E}}^P\|_{0,\Omega,\underline{\underline{\varepsilon}}}^2 + \|\Delta_{\mathbf{E}}^I\|_{0,\Omega,\underline{\underline{\varepsilon}}}^2 + 2\|\Delta_{\mathbf{E}}^P\|_{0,\Omega,\underline{\underline{\varepsilon}}}\|\Delta_{\mathbf{E}}^I\|_{0,\Omega,\underline{\underline{\varepsilon}}} \\
&+\|\Delta_{\mathbf{H}}^P\|_{0,\Omega,\underline{\underline{\mu}}}^2 + \|\Delta_{\mathbf{H}}^I\|_{0,\Omega,\underline{\underline{\mu}}}^2 + 2\|\Delta_{\mathbf{H}}^P\|_{0,\Omega,\underline{\underline{\mu}}}\|\Delta_{\mathbf{H}}^I\|_{0,\Omega,\underline{\underline{\mu}}}
\end{aligned} \tag{37}$$

où $\Delta_{\mathbf{E}}^P = \mathbf{E} - \pi_h^1\mathbf{E}$ (appelée erreur de projection) et $\Delta_{\mathbf{E}}^I = \mathbf{E}_h - \pi_h^1\mathbf{E}$ (appelée erreur d'interpolation).
Nous avons la même chose pour $\mathbf{H}$. L'opérateur π_h^1 que nous avons utilisé, est basé sur un projecteur local H^1 **[PF07]**.

En utilisant l'identité : $2ab \leq a^2 + b^2$, (37) devient :

$$\|(\mathbf{E} - \mathbf{E}_h, \mathbf{H} - \mathbf{H}_h)\|_*^2 \leq 2\left(\|(\Delta_{\mathbf{E}}^P, \Delta_{\mathbf{H}}^P)\|_*^2 + \|\left(\Delta_{\mathbf{E}}^I, \Delta_{\mathbf{H}}^I\right)\|_*^2\right). \tag{38}$$

Afin d'estimer l'erreur introduite par l'approximation spatiale, nous devons évaluer $\|(\Delta_{\mathbf{E}}^P, \Delta_{\mathbf{H}}^P)\|_*$ et $\|\left(\Delta_{\mathbf{E}}^I, \Delta_{\mathbf{H}}^I\right)\|_*$. Nous avons alors établi les résultats suivants :

Premièrement dans le cas d'un schéma sans dissipation *i.e* $\gamma = \delta = 0$, on a

Théorème 1. — *Soit r un entier positif. Supposons que la solution exacte vérifie :*

- $\mathbf{E}$, $\mathbf{H} \in \mathbf{L}^\infty\left([0,T], \mathbf{H}^{s+1}(\mathcal{T}_h)\right)$,
- $\mathbf{E}_t$, $\mathbf{H}_t \in \mathbf{L}^\infty\left([0,T], \mathbf{H}^{s'+1}(\mathcal{T}_h)\right)$

pour des réels $s, s' \geq 0$ avec $\mathbf{v}_t := \partial\mathbf{v}/\partial t$ et $0 < h_K \leq 1\ \forall K \in \mathcal{T}_h$.

Alors, nous avons l'estimation d'erreur d'interpolation suivante :

$$\|(\Delta_E^I, \Delta_H^I)\|_*(T) \leq \|(\Delta_E^I, \Delta_H^I)\|_*(0) + CT\frac{h^{\min(s-1,s',r-1)}}{r^{\min(s-\frac{1}{2},s'+1)}}\beta(\mathbf{E}, \mathbf{H}), \tag{39}$$

où $C > 0$ est une constante indépendante de r et $h = \max\limits_{K\in\mathcal{T}_h} h_K$ et

$$\beta(\mathbf{E}, \mathbf{H}) = \sup_{t\in(0,T)}\left(\|\mathbf{E}\|_{s+1,h}(t), \|\mathbf{H}\|_{s+1,h}(t), \|\mathbf{E}_t\|_{s'+1,h}(t), \|\mathbf{H}_t\|_{s'+1,h}(t)\right). \tag{40}$$

Si on revient à l'erreur du schéma, en utilisant (38), nous avons :

$$
\begin{aligned}
&\|(\mathbf{E}-\mathbf{E}_h,\mathbf{H}-\mathbf{H}_h)\|_*(T) \leq \sqrt{2}(\|(\Delta^P_{\mathbf{E}},\Delta^P_{\mathbf{H}})\|_* + \|\big(\Delta^I_{\mathbf{E}},\Delta^I_{\mathbf{H}}\big)\|_*)(T)\\
&\leq \sqrt{2}\|(\Delta^I_E,\Delta^I_H)\|_*(0) + C\sqrt{2}\big[h^s \max\big(\|\mathbf{E}\|_{s+1,h}(T),\|\mathbf{H}\|_{s+1,h}(T)\big)\\
&+T\frac{h^{\min(s-1,s',r-1)}}{r^{\min(s-\frac{1}{2},s'+1)}}\beta(\mathbf{E},\mathbf{H})\big].
\end{aligned}
\tag{41}
$$

Remarques :

1. Nous voyons que l'erreur semble être sous-optimale et croit linéairement en temps. De plus, pour $r=1$, l'estimation précédente ne prouve pas la consistance du schéma. Dans **[PF07]**, nous montrons avec un simple exemple numérique qu'il n'est pas clair que ce schéma soit consistant pour certains maillages non-structurés.

2. Si le maillage utilisé est orthogonal ou presque parallélépipédique, nous trouvons un exposant h^s dans l'estimation. Dans ce cas, nous sommes soit dans un cas affine soit en présence de dérivées secondes de F_K bornées par Ch_K^2.

3. Dans **[PF07]**, nous étudions aussi l'effet sur l'estimation d'erreur de l'utilisation de la formule de quadrature de Gauss pour calculer les intégrales. Nous avons conclu qu'elle pouvait induire une détérioration de l'ordre de convergence en espace lorsque la solution exacte n'est pas assez régulière à l'intérieur d'au moins une cellule. Néanmoins, si les données du problème sont régulières, la condensation de masse n'induit pas de perte d'ordre de convergence en h (*i.e.* h_K^{r-1}).

Deuxièmement dans **[MPFC08]**, nous avons étudié la convergence en espace de notre schéma DG avec dissipation *i.e.* γ, $\delta > 0$. Nous avons établi

Théorème 2. — *Soit r un entier positif. Supposons que la solution exacte vérifie :*

- $\mathbf{E}$, $\mathbf{H} \in \mathbf{L}^\infty\left([0,T],[H^{s+1}(\mathcal{T}_h)]^3\right)$,
- $\mathbf{E}_t$, $\mathbf{H}_t \in \mathbf{L}^\infty\left([0,T],[H^{s'+1}(\mathcal{T}_h)]^3\right)$

pour des réels $s, s' \geq 0$ et $0 < h_K \leq 1$.

Alors, nous avons l'estimation de l'erreur d'interpolation suivante

$$
\left\|(\Delta^I_{\mathbf{E}},\Delta^I_{\mathbf{H}})\right\|_*(T) \leq \left\|(\Delta^I_{\mathbf{E}},\Delta^I_{\mathbf{H}})\right\|_*(0) + CTh^{\min(s-\frac{1}{2},s'-\frac{1}{2},r-\frac{1}{2})}\beta(\mathbf{E},\mathbf{H})
\tag{42}
$$

où $C > 0$ est une constante indépendante de r et $h = \max\limits_{K\in\mathcal{T}_h} h_K$ et $\beta(\mathbf{E},\mathbf{H})$ est défini comme précédemment.

Finalement, on obtient l'estimateion d'erreur pour le schéma GD :

$$
\|(\mathbf{E}-\mathbf{E}_h,\mathbf{H}-\mathbf{H}_h)\|_*^2 \leq 2(\left\|(\Delta^P_{\mathbf{E}},\Delta^P_{\mathbf{H}})\right\|_*^2 + \left\|(\Delta^I_{\mathbf{E}},\Delta^I_{\mathbf{H}})\right\|_*^2).
\tag{43}
$$

En conclusion, si la solution exacte est suffisament régulière, l'ordre de convergence pour le schéma pénalisé est $r-1/2$ alors que l'on a $r-1$ pour celui non-dissipatif. Les termes de dissipation assurent donc la convergence L^2 du schéma même pour une approximation Q_1. De plus, l'ajout d'un peu de dissipation peut être bénéfique pour contrôler les modes parasites inhérents aux méthodes d'ordre élevé du fait du nombre important de degrés de liberté **[MPFC08]**.

1.4. Approximation en temps

L'approximation en temps est basée sur le schéma saute-mouton d'ordre deux *i.e* $\mathbf{E}$ et $\mathbf{H}$ sont calculés respectivement aux temps $t_n = n\Delta t$ et $t_{n+1/2} = (n+1/2)\Delta t$ où Δt est le pas de discrétisation en temps. De plus, si on considère les termes dissipatifs dans la formulation, alors on a décidé de décentrer leur approximation en temps afin de garder un schéma explicite. Il vient alors:

$$B_\varepsilon \frac{\mathbf{E^{n+1}} - \mathbf{E^n}}{\Delta t} + R_h \mathbf{H^{n+1/2}} + D_h^\alpha \mathbf{E^n} + S_h^\beta \mathbf{H^{n+1/2}} + \mathbf{J^n} = 0, \tag{44a}$$

$$B_\mu \frac{\mathbf{H^{n+1/2}} - \mathbf{H^{n-1/2}}}{\Delta t} + R_h \mathbf{E^n} + \left(S_h^\gamma\right)^* \mathbf{E^n} + \left(D_h^\delta\right)^* \mathbf{H^{n-1/2}} = 0. \tag{44b}$$

La stabilité du schéma a été obtenue à la fois par une analyse de type onde plane sur maillage cartésien (qui permet d'avoir des conditions de stabilité nécessaires et suffisantes) et par une technique énergétique sur maillage non structuré (qui permet d'avoir des conditions suffisantes pouvant être parfois restrictives). Cette dernière technique met en évidence le caractère non dissipatif du schéma sur maillages non-structurés lorsque $\alpha = \delta = 0$. Nous renvoyons à [**CFP06, MPFC08**] pour l'ensemble de ces résultats.

1.5. Quelques optimisations/améliorations

Bien que ce schéma présente des performances intéressantes, son utilisation pour décrire fidèlement une scène complexe nécessite des améliorations. En effet, les maillages non-structurés utilisés possèdent inévitablement des zones sur-raffinées et il peut y avoir une grande disparité de taille entre les cellules. La condition de stabilité du schéma temporel utilisé implique donc un pas de temps très petit et un calcul prohibitif. Ceci est d'autant plus pénalisant que l'on utilise des ordres élevés. En effet, cette condition décroît avec l'ordre d'approximation. Afin d'améliorer les performances de la méthode, nous avons proposé les optimisations/améliorations suivantes :

1.5.1. Pas de temps local. — Nous avons vu que le schéma (44a-44b) est stable sous une condition CFL (Courant-Friedrichs-Lewy) de la forme $\Delta t/h_{\min} \leq C(r)$ où Δt est un pas de temps global, $h_{\min}$ est le plus petit diamètre des cellules du maillage et $C(r)$ est une constante positive qui dépend de l'ordre d'approximation polynomiale en espace r. Si l'on veut tenir compte de détails géométriques ou capturer les singularités dans la solution, il est alors naturel d'utiliser localement une technique de raffinement en espace. L'utilisation de ce type de maillages hétérogènes aura pour conséquence immédiate de diminuer significativement le pas de temps Δt afin de garantir la stabilité du schéma. De plus, si ces zones correspondent à une petite fraction du domaine de calcul alors elles induisent dans la partie complémentaire un effort de calcul inutile et très pénalisant du point de vue du coût de la méthode. Pour pallier à ce problème, l'idée est de considérer une méthode de pas de temps local qui considère des pas de temps adaptés aux différentes contraintes spatiales pour réaliser l'avancée en temps. Cet objectif est difficile à réaliser dans le contexte de la propagation d'onde. En particulier, trois points importants doivent être considérés lors du développement d'une stratégie de pas de temps local :

- assurer la stabilité avec le meilleur pas de temps dans chaque sous-domaine,
- garantir la précision de la discrétisation en temps,
- obtenir un schéma explicite ou "quasi" explicite peu coûteux.

Il existe de nombreux travaux qui s'intéressent à cette problématique. Les méthodes proposées sont souvent basées sur la conservation d'une énergie discrète qui est une propriété pertinente pour assurer la stabilité d'un schéma. On peut citer par exemple les schémas basés sur une approche symplectique proposée par S. Piperno [**Pip06**] qui sont faciles à implémenter et généralement efficaces en pratique mais dont la question de la stabilité n'est pas totalement tranchée. Il y a aussi l'approche proposée

par P. Joly *et al* [**CFJ06, BJR05, Rod08, EJ09**] qui proposent un schéma qui est optimal en terme de stabilité mais dont l'implémentation est plus complexe et exige la résolution d'un petit système linéaire. M. Grote *et al* proposent des schémas (voir par exemple [**DG09**]) de pas de temps local explicites, stables, précis, efficaces et relativement faciles à implémenter adaptés aux équations des ondes (ondes, électromagnétisme, élastodynamique) d'ordre deux en temps. Nous avons d'ailleurs éprouvé ce type de schéma dans la thèse de Y. Dudouit [**Dud14**] pour des problèmes de couplage élasto-acoustique. Enfin, nous pouvons aussi conseiller de regarder les méthodes de pas de temps local basées sur l'approche ADER [**DKT07**].

Dans [**MPFC08**], nous avons proposé une technique de pas de temps local qui est construite à partir de l'approche symplectique de S. Piperno. Contrairement à son approche qui est basée sur le schéma de Verlet (qui est une réorganisation du schéma de saute-mouton en 3 étapes), nous avons décidé de construire notre schéma en restant sur une approche purement saute-mouton qui sera plus performante en terme de temps calcul.

Ce schéma utilise une approche multi-classes dans laquelle les cellules (ou les degrés de liberté dans un contexte éléments finis) sont distribuées dans N ensembles ou classes $1, 2, ..., N-1, N$ qui sont respectivement associés aux pas de temps $\Delta t/2^{N-1}$, $\Delta t/2^{N-2}$, $\cdots$, $\Delta t/2$, Δt. Les cellules de petite taille sont donc dans la classe 1 et celles de plus grande taille dans la classe N.

Nous allons présenter succinctement l'approche pour un schéma DG non-dissipatif *i.e* $\alpha = \delta = 0$. Dans ce cas, (34a-34b) peut s'écrire sous la forme générique :

$$B_h^1 \frac{d\,\mathbf{E}}{dt} - R_h\,\mathbf{H} = 0, \tag{45a}$$

$$B_h^2 \frac{d\,\mathbf{H}}{dt} + R_h^*\,\mathbf{E} = 0. \tag{45b}$$

Par exemple, pour $N = 2$, nous proposons le schéma multi-classes suivant pour traiter les cellules situées à l'interface entre les classes 1 et 2 :

$$B_{h2}^2 \frac{V_2^{n+\frac{1}{2}} - V_2^{n-\frac{1}{2}}}{\Delta t} = -R_{h2}^*\,U_2^n + A_{12}^*\,U_1^n \tag{46a}$$

$$B_{h1}^2 \frac{V_1^{n+\frac{1}{6}} - V_1^{n-\frac{1}{6}}}{\Delta t/3} = -R_{h1}^*\,U_1^n + A_{21}^*\,U_2^n \tag{46b}$$

$$B_{h1}^1 \frac{U_1^{n+\frac{2}{6}} - U_1^n}{\Delta t/3} = R_{h1}\,V_1^{n+\frac{1}{6}} - A_{12}\left(V_2^{n+\frac{1}{2}}\right)^\star \tag{46c}$$

$$B_{h1}^2 \frac{V_1^{n+\frac{1}{2}} - V_1^{n+\frac{1}{6}}}{\Delta t/3} = -R_{h1}^*\,U_1^{n+\frac{2}{6}} + A_{21}^*\,(U_2^n)^\star \tag{46d}$$

$$B_{h2}^1 \frac{U_2^{n+1} - U_2^n}{\Delta t} = R_{h2}\,V_2^{n+\frac{1}{2}} - A_{21}\,V_1^{n+\frac{1}{2}} \tag{46e}$$

$$B_{h1}^1 \frac{U_1^{n+\frac{4}{6}} - U_1^{n+\frac{2}{6}}}{\Delta t/3} = R_{h1}\,V_1^{n+\frac{1}{2}} - A_{12}\,V_2^{n+\frac{1}{2}} \tag{46f}$$

$$B_{h1}^2 \frac{V_1^{n+\frac{5}{6}} - V_1^{n+\frac{1}{2}}}{\Delta t/3} = -R_{h1}^*\,U_1^{n+\frac{4}{6}} + A_{21}^*\,\left(U_2^{n+1}\right)^\star \tag{46g}$$

$$B_{h1}^1 \frac{U_1^{n+1} - U_1^{n+\frac{4}{6}}}{\Delta t/3} = R_{h1} V_1^{n+\frac{5}{6}} - A_{12} \left(V_2^{n+\frac{1}{2}} \right)^\star \tag{46h}$$

où U_i et V_i représentent les degrés de liberté de U et V dans la classe i et nous avons décomposé les matrices de la façon suivante :

$$B_h^k W = \begin{bmatrix} B_{h1}^k & 0 \\ 0 & B_{h1}^k \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \end{bmatrix}, \tag{47a}$$

$$R_h W = \begin{bmatrix} R_{h1} & -A_{12} \\ -A_{21} & R_{h2} \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \end{bmatrix}. \tag{47b}$$

Comme S. Piperno, nous remplaçons les champs inconnus (notés *) par leur dernière valeur connue.

Remarque : Pour les cellules qui ne sont pas situées à l'interface entre les cellules de type 1 et 2, c'est le schéma de saute-mouton classique qui s'applique (en utilisant le pas de temps associé à la classe).

Plus généralement, si nous dénotons $LeapFrogV(N, \Delta t)$ et $LeapFrogU(N, \Delta t)$ les deux étapes classiques d'intégration d'un schéma saute-mouton appliquées aux cellules de la classe N avec un pas Δt, nous pouvons alors définir une méthode saute-mouton multi-classes par un processus récursif. Une étape de l'intégration en temps est alors définie :

$$\begin{cases} 1. \ ComputeV(N, \Delta t) \\ 2. \ ComputeU(N, \Delta t) \end{cases} \tag{48}$$

où les fonctions récursives $ComputeV(N, \Delta t)$ et $ComputeU(N, \Delta t)$ sont respectivement définies par

$$\underline{ComputeV(N, \Delta t)} : \quad \begin{cases} - \ LeapFrogV\,(N, \Delta t) \\ - \ ComputeV\left(N-1, \frac{\Delta t}{3}\right) \\ - \ ComputeU\left(N-1, \frac{\Delta t}{3}\right) \\ - \ ComputeV\left(N-1, \frac{\Delta t}{3}\right) \end{cases} \qquad \underline{ComputeU(N, \Delta t)} : \quad \begin{cases} - \ LeapFrogU\,(N, \Delta t) \\ - \ ComputeU\left(N-1, \frac{\Delta t}{3}\right) \\ - \ ComputeV\left(N-1, \frac{\Delta t}{3}\right) \\ - \ ComputeU\left(N-1, \frac{\Delta t}{3}\right) \end{cases}$$

avec $ComputeV\,(1, \Delta t)$ est défini par $LeapFrogV\,(1, \Delta t)$ et $ComputeU\,(1, \Delta t)$ est défini par $LeapFrogU\,(1, \Delta t)$, où Δt est le pas de temps.

Remarque : Puisque le schéma saute-mouton est composé de seulement deux étapes (contre trois pour le schéma de Verlet), cette méthode implique un gain de 33% par rapport au schéma de S. Piperno tout en gardant les mêmes avantages/défauts c'est-à-dire le schéma est explicite et donne de bons résultats avec un temps CPU significativement réduit par rapport à la méthode sans pas de temps local mais comme le schéma basé sur l'approche Verlet, il nécessite de réduire la constante CFL afin d'assurer la stabilité lors de simulations en temps long.

Pour finir, nous donnons un résultat de comparaison (voir **[MPFC08]** pour plus de détails) entre les schémas récursifs de Verlet (R-V) et saute-mouton (R-LF). Nous avons considéré un missile générique (voir fig. 1) agressé par une onde plane (voir **[MPFC08]**).
Nous donnons dans la table 1 la répartition des cellules du maillage par classe pour les deux approches. On peut noter d'une part un faible pourcentage de "petites" cellules et d'autre part une disparité importante de la taille des cellules du maillage. Par exemple, il y a au moins un facteur 2^{10} entre la plus petite et la plus grosse cellule (car le schéma utilise 10 classes). Ceci explique la grande efficacité de ces approches sur ce type de maillage. Dans **[MPFC08]**, nous montrons le gain en temps CPU obtenu par notre schéma en comparaison avec les approches saute-mouton classiques *i.e.* sans pas de temps local et Verlet. Pour cet exemple, l'amélioration obtenue par les techniques

Scheme/Class	1	2	3	4	5	6	7	8	9	10
R-LF	10	200	1400	14300	71600	3500	×	×	×	×
R-V	8	16	160	550	1500	5800	46000	33500	3300	200

TABLE 1. Répartition des cellules par classe pour un maillage d'un missile (∼91000 cellules)

récursives est très importante : R-V et R-LF sont respectivement 11 et 15 fois plus rapides que le schéma standard.

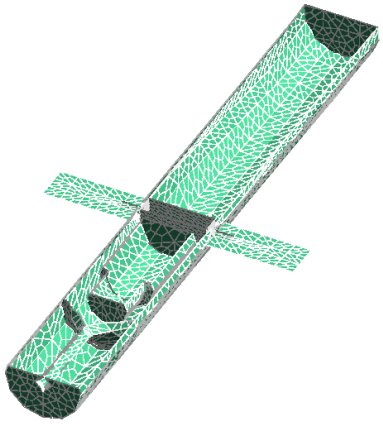


FIGURE 1. Géométrie du missile

1.5.2. Approximation non-conforme. — Nous avons étudié deux voies complémentaires qui permettent d'améliorer la méthode. Premièrement, nous avons exploité une des souplesses du GD qui est la discontinuité de la solution entre les cellules du maillage. On peut en effet choisir de fixer un ordre d'approximation différent pour chaque cellule et puisqu'on travaille avec des cellules hexaédriques, on peut même considérer un ordre différent dans chaque direction. Nous parlons respectivement d'approximation polynomiale isotrope et anisotrope. Ces ordres peuvent être déterminés en fonction de la taille de la cellule (par rapport à la longueur d'onde). Dans les très petites cellules, il n'est en effet pas nécessaire d'utiliser des ordres élevés pour obtenir une solution précise. Par ce biais, nous avons obtenu une diminution du nombre de degrés de liberté (pour une précision comparable au schéma avec ordre fixe) et une augmentation de la condition CFL, donc du pas de temps (contraint par les cellules de petites dimensions) [**PMFP09, MFP$^+$13**].

Deuxièmement, nous avons envisagé une autre voie complémentaire à l'approche précédente qui est l'utilisation de maillage non-conforme. L'utilisation de cellules non-structurées permet de relâcher rapidement le maillage (passer de petites cellules à de grosses cellules), mais la conformité de celui-ci implique forcément la présence de "grosses" cellules possédant au moins une face de petite taille. Le diamètre de ces cellules nécessite l'utilisation d'une approximation d'ordre élevé afin d'avoir une bonne précision tout en imposant un petit pas de temps (conditionné par les petites faces et l'ordre d'approximation) pour obtenir la stabilité du schéma. Une solution pour limiter l'impact de ce problème serait de réaliser une montée en ordre progressif entre la zone contenant les petites cellules (bas ordre) et celle contenant les grosses (ordre élevé). Néanmoins, ceci semble difficile à gérer dans le cas d'une scène complexe. La solution la plus adaptée est donc d'utiliser des maillages possédant des zones de non-conformité (les intersections entre hexaèdres voisins peuvent être des portions de faces de ces hexaèdres). On peut ainsi relâcher le maillage en utilisant des cellules dont les faces d'intersection non nulles avec la zone raffinée ne sont pas forcément contraintes par les éléments de cette zone. Conceptuellement, les méthodes GD permettent de réaliser cela naturellement.

Néanmoins, une mise en oeuvre efficace tant du point de vue de la stabilité du schéma, de la qualité de la solution que du coût de la méthode nécessite au préalable une analyse fine du schéma. Nous avons étudié la stabilité et la convergence du schéma GD appliqué à des maillages non-conformes pour une approximation polynomiale anisotrope [**Per08**]. Cette étude nous a permis d'exhiber des formules d'intégrations numériques qui permettent d'assurer la stabilité du schéma. De plus, ce choix permet de conserver une énergie discrète dans le cas de l'utilisation de "flux centrés" bien que les intégrales de surfaces ne soient pas calculées exactement dans les zones de non-conformité. Pour finir, sur la fig. 2, nous présentons un exemple numérique qui compare une approche mixte $Q_2 - Q_1$ couplée avec la stratégie de pas de temps local décrite précédemment avec une approche classique Q_2. Sur cet exemple, nous avons une division par 7 du temps calcul tout en garantissant une précision équivalente à la solution numérique.

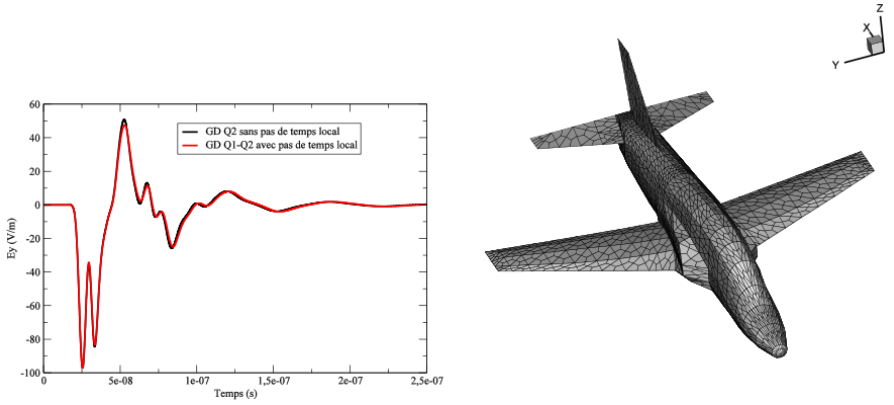


FIGURE 2. Comparaison d'une approche mixte $Q_2 - Q_1$ couplée avec un pas de temps local avec une approche classique Q_2.

1.6. Développements connexes

1.6.1. Analyse d'erreur *a posteriori.* — Aujourd'hui, l'analyse d'erreur a posteriori et l'adaptation automatique de maillage sont devenues des outils importants dans l'analyse numérique des équations aux dérivées partielles (EDP). En effet, la performance d'une méthode de résolution numérique d'une EDP est étroitement liée à la qualité du maillage utilisé. L'adaptation de maillage automatique a largement prouvé son efficacité. Cette technique permet, d'une part de réduire notablement le coût de calcul (en réduisant le nombre de degrés de liberté) et d'autre part, d'atteindre une solution numérique à la précision désirée. La qualité de la solution obtenue est évaluée à travers des techniques d'estimateurs d'erreur a posteriori. Ce type d'estimateur représente une quantité calculable (ne dépendant que de la solution discrète et des données du problème) qui est équivalente à l'erreur entre la solution exacte (qui est inconnue) et la solution approchée. Toute la difficulté est d'exhiber à partir d'un schéma numérique un estimateur fiable et efficace. Le développement de tels outils reste un challenge dans de nombreuses situations. En particulier, peu de résultats existent pour l'adaptativité espace-temps. Ceci est d'autant plus vrai pour les équations de Maxwell pour lesquelles, il n'existe que des résultats dans le domaine fréquentiel ou pour des problèmes de Maxwell coercifs [**HPS05, CDN07, Sch08, IHvdV08, Mon98, Rep07, BHHW00, HPS07, CWZ07, CD07, BS08, SDCD08, RD00, RD02, CDR03**]. Pour le domaine temporel, il y très peu de résultats et uniquement pour des modèles particuliers [**Hof00, ZCL06, Li09**].

Dans le projet REI DIGATOP, nous avions pour but d'initier une action sur la recherche d'estimateurs a posteriori dans le cadre des équations de Maxwell dans le domaine temporel. L'objectif suprême étant d'apporter un outil quantitatif permettant de s'assurer de la qualité de la solution

numérique et le cas échéant de savoir où il faut agir (raffinement et déraffinement de maillage, augmentation de l'ordre de l'approximation, raffinement en temps) pour améliorer la précision globale de la solution. Pour dériver une estimation d'erreur a posteriori pour notre schéma, nous avons considéré la même technique énergétique que celle utilisée pour l'analyse d'erreur a priori du schéma et pour (espérer) obtenir une estimation optimale, nous nous sommes inspirés de **[MN03]** en combinant cette technique avec une représentation de l'erreur basée sur un opérateur de reconstruction. Cette approche est résumée sur la figure 3. De part ces choix, notre analyse d'erreur a

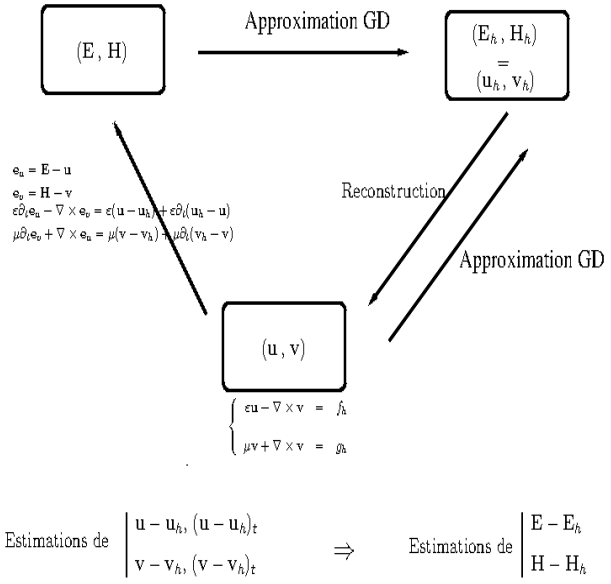


FIGURE 3. Démarche de construction de l'estimateur d'erreur *a posteriori*

posteriori est ramenée à la détermination d'un estimateur d'un problème "plus simple": Trouver $(\mathbf{u}, \mathbf{v}) = \mathrm{H}_{\mathbf{0}}(\mathrm{curl}, \mathbf{\Omega}) \times \mathrm{H}(\mathrm{curl}, \mathbf{\Omega})$ solution du problème

$$\left\{ \begin{array}{rcl} \varepsilon \mathbf{u} - \nabla \times \mathbf{v} & = & f \text{ dans } \Omega \\ \mu \mathbf{v} + \nabla \times \mathbf{u} & = & g \text{ dans } \Omega \\ \mathbf{n} \times \mathbf{u} & = & 0 \text{ sur } \Gamma \end{array} \right. \tag{49}$$

On s'est donc focalisé sur l'analyse d'erreur a posteriori de (49) discrétisé par notre approximation GD. Nous avons alors dérivé un estimateur d'erreur **[Per09]** qui est décrit à l'aide des quantités locales : $\forall K \in \mathcal{T}_h$,

$$(\eta_K^s)^2 := h_K^{2s} \|(f - \varepsilon \mathbf{u_h} + \nabla \times \mathbf{v_h})\|_{\mathbf{0,K}}^{\mathbf{2}} + \mathbf{h_K^{2s}} \|(\mathbf{g} - \mu \mathbf{v_h} + \nabla \times \mathbf{u_h})\|_{\mathbf{0,K}}^{\mathbf{2}}$$

$$+ h_K^{2s-1} \|(\beta [\![\mathbf{v_h} \times \mathbf{n}]\!]_{\partial K}^K + \alpha [\![\mathbf{n} \times \left(\mathbf{u_h} \times \mathbf{n} \right)]\!]_{\partial \mathbf{K}}^{\mathbf{K}}) \|_{\mathbf{0},\partial \mathbf{K}}^{\mathbf{2}}$$

$$+ h_K^{2s-1} \|((\gamma [\![\mathbf{u_h} \times \mathbf{n}]\!]_{\partial K}^K + \delta [\![\mathbf{n} \times \left(\mathbf{v_h} \times \mathbf{n} \right)]\!]_{\partial \mathbf{K}}^{\mathbf{K}}) \|_{\mathbf{0},\partial \mathbf{K}}^{\mathbf{2}}$$

$$+ h_K^2 \|\varepsilon^{-1/2} \nabla \cdot (f_h - \varepsilon \mathbf{u_h})\|_{\mathbf{0,K}}^{\mathbf{2}} + \mathbf{h_K^2} \|\mu^{\mathbf{-1/2}} \nabla \cdot (\mathbf{g_h} - \varepsilon \mathbf{v_h})\|_{\mathbf{0,K}}^{\mathbf{2}}$$

$$+ \frac{h_K}{2} \|\varepsilon^{-1/2} [\![(f_h - \varepsilon \mathbf{u_h}) \cdot \mathbf{n}]\!] \|_{\mathbf{0},\partial \mathbf{K} \cap \mathcal{E}_\mathbf{i}}^{\mathbf{2}} + \frac{\mathbf{h_K}}{\mathbf{2}} \|\mu^{\mathbf{-1/2}} [\![(\mathbf{g_h} - \mu \mathbf{v_h}) \cdot \mathbf{n}]\!] \|_{\mathbf{0},\partial \mathbf{K} \cap \mathcal{E}_\mathbf{i}}^{\mathbf{2}}$$

Sous des hypothèses convenables, on a montré qu'il existe $s \in]1/2, 1]$ et des constantes C, $C_1 > 0$ indépendantes de h telles que

$$\|(\mathbf{u}, \mathbf{v}) - (\mathbf{u_h}, \mathbf{v_h})\|^{\mathbf{2}} \leq \mathbf{C} \sum_{\mathbf{K} \in \mathcal{T}_\mathbf{h}} (\eta_{\mathbf{K}}^{\mathbf{s}})^{\mathbf{2}} + \underbrace{C_1(\|(f - f_h)\|_0^2 + \|(g - g_h)\|_0^2)}_{=\nu}$$

Cette estimation implique ce qu'on appelle la fiabilité de l'estimateur c'est-à-dire qu'une valeur petite de l'estimateur implique une erreur petite.

Dans le rapport **[Per10a]**, on a démontré l'éfficacité de cet estimateur pour $s = 1$ c'est-à-dire

$$\eta := \left(\sum_{K \in \mathcal{T}_h} (\eta_K^s)^2 \right)^{1/2} \leq C\|(\mathbf{u}, \mathbf{v}) - (\mathbf{u_h}, \mathbf{v_h})\| + \nu \tag{50}$$

ou encore autrement dit, si l'indicateur est grand alors l'erreur est grande.

Cet estimateur a été testé numériquement dans la thèse de JB Laurent **[Lau13]**. Il a montré que l'estimateur proposé convenait bien pour la localisation des zones de raffinement/déraffinement pour la méthode DG proposée ci-dessus.

1.6.2. Couplage elasto-acoustique. — Dans le cadre de la thèse de Yohann Dudouit **[Dud14]** financée par Total, nous nous sommes entre autres intéressés à la propagation d'ondes dans des milieux élastiques hétérogènes dans lesquels on trouve des inclusions/fractures remplies d'eau. L'objectif était de mettre en oeuvre une méthode numérique capable de simuler rigoureusement l'impact sur la propagation des ondes d'une multitude d'hétérogénéités de petite taille. On peut trouver des milieux équivalents pour réaliser cela et ainsi éviter de tenir compte explicitement de ces hétérogénéités, malheureusement, ces derniers ont souvent du mal à restituer la totalité des phénomènes. En particulier, une partie des ondes diffractées ne sont pas prise en compte alors qu'elles peuvent être cruciales pour la caractérisation du milieu. Cette caractérisation est essentielle pour la prospection pétrolière.

L'approche que l'on a choisie est basée sur un modèle purement en déplacement correspondant au couplage de l'équation de l'élastodynamique du second ordre en temps et celle de l'acoustique. L'approximation de ce modèle est basée sur une formulation de Galerkin discontinue de type IPDG (Interior Penalty Discontinuous Galerkin) dans laquelle nous traitons de manière unifiée les domaines acoustique et élastique. L'unification est réalisée par le choix de termes de flux adéquats aux deux physiques et à leurs raccords (voir **[DGMP16]** pour les détails). Par ailleurs, nous avons proposé un terme de pénalisation optimisé dans la méthode IPDG qui est mieux adapté à l'équation de l'élastodynamique, conduisant à moins de dispersion numérique et à une meilleure condition CFL (elle est améliorée de 30% par rapport à la méthode classique). Nous avons aussi amélioré une formulation PML du second ordre pour laquelle nous avons proposé une nouvelle discrétisation temporelle qui rend la formulation plus stable. En tirant parti de la p-adaptativité et des maillages non-conformes des méthodes de Galerkin discontinues combinés à une méthode de pas de temps local (basée sur l'approche de Marcus Grote et Julien Diaz), nous avons réduit significativement le coût du raffinement local. Sur les figures 4 et 5 nous présentons un cas illustratif.

1.7. Bilan et perspectives

Nous avons présenté une méthode numérique basée sur une approche GD qui possède de très bonnes caractéristiques pour résoudre des scènes complexes en électromagnétisme. De nombreuses simulations numériques ont confirmé ces bonnes propriétés. De façon générale (qui dépasse le cadre de ce document), les connaissances sur les méthodes d'ordre élevé ont atteint une maturité importante. Néanmoins, il est souvent difficile d'exploiter ces méthodes en situation réelle. En

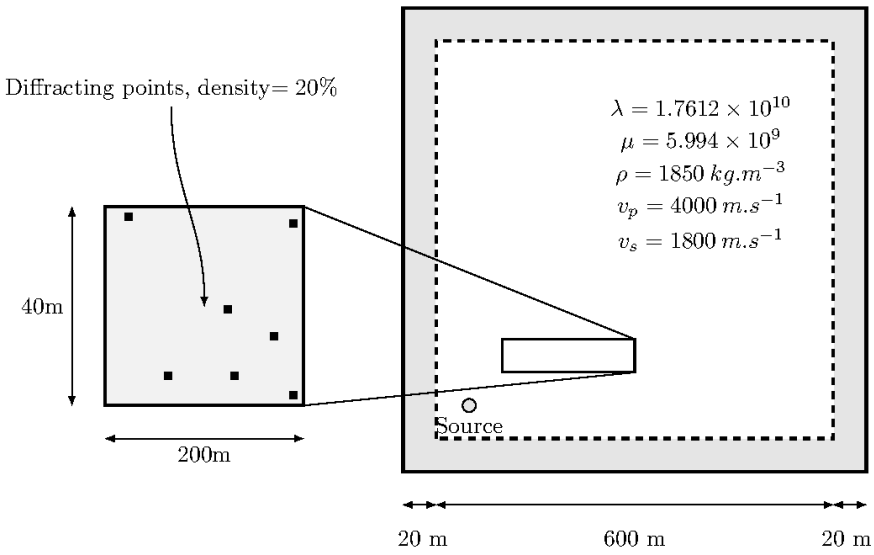


FIGURE 4. Caractéristiques du milieu diffractant

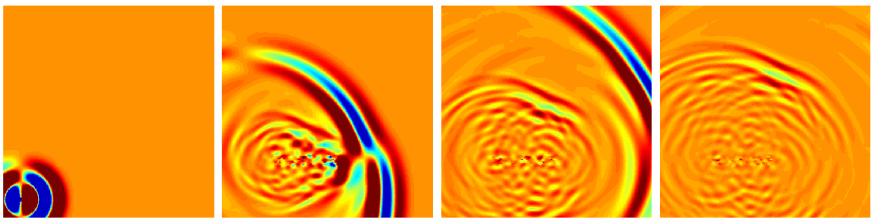


FIGURE 5. Snapshots de la composante $\underline{u}_x$ du déplacement à différents temps pour un milieu avec des points diffractants

particulier, la gestion de la géométrie des objets reste toujours problématique malgré les différentes stratégies (p-adaptation, pas de temps local) mises en place pour limiter leur impact. Il est apparu ces dernières années de nombreux travaux sur la construction de méthodes basées sur des maillages courbes ou même directement construits à partir de la CAO des objets. Nous pensons que notre approche GD serait bien adaptée à ce type de stratégie. En effet, notre espace d'approximation permet d'éliminer naturellement les informations géométriques au niveau des termes de rigidité et de la partie sauts centrés, ce qui devrait aider grandement à l'étude et à la mise en place d'une telle méthode.

Un autre aspect complémentaire au développement de méthodes d'ordre élevé est leur exploitation. En effet, une fois que nous avons à notre disposition ce type de solutions numériques, la question de leur utilisation se pose. Deux types de traitement sont généralement effectués : on effectue une représentation graphique de la solution et/ou on en extrait des informations par post-traitement (vorticité, streamline, coupe ...etc...). Le premier semble *a priori* anodin, il pose néanmoins de nombreux problèmes. En effet, il faut avoir à l'esprit que les logiciels de visualisation ne savent représenter sans perte d'information que des fonctions affines sur simplexes et donc la représentation exacte de notre solution numérique est un voeu pieux (à part si elle est affine sur simplexe). On a alors deux options. La première est de réaliser soi-même la mise au format en réalisant une interpolation affine sur simplexes de la solution. C'est la plus couramment utilisée. Elle pose néanmoins de nombreuses questions: Comment mesurer l'erreur introduite par le procédé? Qu'en est-il du contrôle sur les extractions d'information faites à partir des logiciels de visualisation? Comment éviter une trop grande augmentation des données à manipuler? La seconde option est

d'utiliser la fonctionnalité "ordre élevé" que possèdent certains logiciels de visualisation et qui permet de fournir à ce dernier une description exacte de certaines solutions d'ordre élevé. Néanmoins, le logiciel va créer à partir d'un algorithme "invisible" une représentation approchée que l'utilisateur pourra uniquement contrôler en fournissant une précision cible. Ceci pose évidemment la question du sens du contrôle de l'erreur : quelle précision cible dois-je choisir pour observer les phénomènes importants? Nous avons récemment lancé un axe de recherche sur ce sujet. Ce travail se situe entre le code de calcul et le logiciel de visualisation et a pour objectif de proposer des méthodes de post-traitement des résultats bruts issus du code de calcul afin de pouvoir réaliser ensuite à partir du logiciel de visualisation une représentation à la précision désirée à la fois de la solution et des extractions que l'utilisateur décidera alors de faire. Par exemple, une première question intéressante à se poser est : les schémas numériques nous donnent des solutions approchées dont le contrôle de l'erreur est dans des normes dérivant de la norme L^2, que signifie donc une représentation ponctuelle de ces dernières par rapport aux solutions exactes? Un exemple concret de cette problématique sont les sauts numériques créés par les méthodes GD. Hormis pour mesurer la qualité du schéma GD, ils n'ont aucun intérêt pour la visualisation et de plus, ils sont néfastes pour certaines extractions (par exemple problème de convergence lors de l'extraction de streamline par une méthode Runge-Kutta). Donc la recherche d'une solution plus pertinente pour la représentation (solution et extractions) dans une classe de solutions admissibles est primordiale. Il est connu par exemple que des méthodes de projection dans un espace d'approximation conforme à la régularité de la solution exacte ou encore un "lissage" par une méthode de type SIAC [**SCKR08**] sont très bénéfiques. Nous avons proposé une première méthodologie pour réaliser une représentation linéaire optimisée de fonctions d'ordre élevé [**HMP16**] et nous poursuivons actuellement sur les différents aspects présentés ci-dessus dans la thèse de Matthieu Mounaury et le projet de recherche ONERA PREVISIO.

Ces travaux ont donné lieu à l'écriture d'un livre chez Springer avec G. Cohen et à sept publications dans des journaux : 2 Journal of Computational Physics, 1 Mathematics of COMPutation, 1 IEEE Trans. on Antennas and Propagation, 1 IET Science, Measurement & Technology, 1 CRAS Physique et Journal of Computational Methods in Physics. Ils ont été réalisés en collaborant avec X. Ferrieres, G. Cohen, Y. Dudouit (doctorant), F. Millot, L. Giraud ainsi que des étudiants en stage niveau master E. Montseny et B. Mallet. Enfin, ils ont été en partie réalisés dans le cadre des projets suivants : Projet de Recherche Fédérateur ONERA MAHPSO et le projet REI DIGATOP.

3

CHAPITRE 2

FORMULATIONS INTEGRALES POUR LE CALCUL RAPIDE DE LA SURFACE EQUIVALENTE RADAR

Dans ce deuxième chapitre, je présente une synthèse de mes travaux sur le thème de la recherche de méthodes intégrales performantes en terme de coût calcul et de précision pour simuler la diffraction d'ondes électromagnétiques en régime harmonique. Plus précisément, ce chapitre contient deux aspects principaux. Premièrement, les constructions de formulations de type GCSIE (Generalized Combined Source Integral Equation) pour des problèmes avec conditions d'impédance et de transmission sont présentées. Ces formulations induisent des systèmes linéaires bien adaptés à une résolution itérative. Deuxièmement, une nouvelle technique de localisation permettant de dériver des indicateurs d'erreur *a posteriori* fiables, efficaces et permettant un contrôle quantitatif de l'erreur induise par un schéma construit à partir d'une formulation intégrale est décrite. Enfin, la dernière partie de ce chapitre présente un bilan de ces travaux et propose quelques perspectives de travail.

2.1. Problématique et difficultés

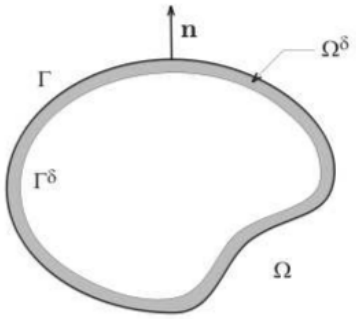
Une partie de mes activités de recherche porte sur la modélisation des phénomènes de diffraction d'ondes électromagnétiques en régime harmonique et plus particulièrement sur la recherche de nouvelles méthodes ou formulations pour le calcul rapide de la Surface Equivalente Radar (SER). Les méthodes intégrales sont très populaires et ont remporté beaucoup de succès durant ces trente dernières années. Une des raisons de leur popularité est qu'elles posent le problème sur les surfaces des objets, diminuant ainsi la dimension spatiale des problèmes à résoudre. Toutefois, les matrices résultant de leurs approximations sont denses et de taille typiquement proportionnelle à la puissance quatrième de la fréquence. Ceci a limité l'application de ce type de méthodologie jusqu'aux années 1990. L'apparition des techniques multipôles (FMM pour "Fast Multipole Methods") a permis la compression de ces matrices ainsi que l'accélération des produits matrice-vecteur. Cette avancée révolutionnaire a permis de traiter des cas auparavant inenvisageables (on est maintenant capable de résoudre des problèmes impliquant plusieurs millions d'inconnues sur la surface de l'objet).

Lors de mon arrivée dans l'équipe Electromagnétisme et Acoustique du CERFACS en 2005, je me suis tout d'abord intéressé à la modélisation par des équations intégrales du phénomène de diffraction d'onde électromagnétique par des objets recouverts d'une fine couche de matériaux diélectriques. C'est par exemple une problématique très importante dans le domaine de la furtivité où pour éviter leur détection par un radar, les objets (avions de combat par exemple) sont souvent partiellement recouverts par une fine couche de matériaux (peinture) afin de réduire la SER de l'onde qu'ils diffractent. Pour évaluer la SER d'objet, les équations intégrales en régime harmonique sont l'outil privilégié car les radars émettent à des fréquences fixées. Ces phénomènes sont décrits par le champ électromagnétique $(\mathbf{E}, \mathbf{H}) : \Omega \cup \Omega^\delta \to \mathbb{C}^3$ vérifiant les équations de Maxwell, les conditions

de transmission et les conditions aux limites suivantes :

$$
(51)\quad \begin{cases}
\nabla \times \mathbf{E} - ik\mathbf{H} = 0 & \text{dans } \Omega \\
\nabla \times \mathbf{H} + ik\mathbf{E} = 0 & \text{dans } \Omega \\
\nabla \times \mathbf{E} - ik\mu\mathbf{H} = 0 & \text{dans } \Omega^\delta \\
\nabla \times \mathbf{H} + ik\varepsilon\mathbf{E} = 0 & \text{dans } \Omega^\delta \\
\mathbf{n} \times \mathbf{E} = 0 & \text{sur } \Gamma^\delta \\
\mathbf{n} \times \mathbf{E}_{|\Omega} = \mathbf{n} \times \mathbf{E}_{|\Omega^\delta} & \text{sur } \Gamma \\
\mathbf{n} \times \mathbf{H}_{|\Omega} = \mathbf{n} \times \mathbf{H}_{|\Omega^\delta} & \text{sur } \Gamma \\
+ \text{ Condition de radiation à l'infini}
\end{cases}
$$

où les notations sont décrites sur la figure 2.1, k est le nombre d'onde et ε, μ sont les caractéristiques diélectriques de la couches Ω^δ.



Les couches utilisées sont généralement petites par rapport à la longueur d'onde du problème et des modèles approchés sont souvent préférés pour éviter des problèmes numériques ainsi qu'un coût de calcul trop important. On trouve alors dans la littérature diverses conditions aux limites effectives permettant de restituer le comportement de la couche. Ces conditions peuvent être obtenues en utilisant des développements asymptotiques dont le petit paramètre est l'épaisseur de la couche δ. Ceci nous permet de considérer le modèle approché suivant : Trouver $(\mathbf{E}, \mathbf{H}) : \Omega \to \mathbb{C}^3$ tel que

$$
(52)\quad \begin{cases}
\nabla \times \mathbf{E} - ik\mathbf{H} = 0 & \text{dans } \Omega \\
\nabla \times \mathbf{H} + ik\mathbf{E} = 0 & \text{dans } \Omega \\
\mathbf{n} \times (\mathbf{E} \times \mathbf{n}) + Z^\delta_{eff}\mathbf{n} \times \mathbf{H} = 0 & \text{sur } \Gamma \\
+ \text{ Condition de radiation à l'infini}
\end{cases}
$$

où Z^δ_{eff} est un opérateur surfacique (composé d'opérateurs différentiels la plupart du temps) modélisant l'effet de la couche mince.

Dans ce manuscrit, nous ne parlerons pas de la dérivation de ces conditions, nous donnons simplement quelques exemples:

- $Z^\delta_{eff} = ik\mu\delta$ correspondant à la condition de Léontovitch [**Leo78**]. Cette condition est exacte pour les ondes arrivant avec une incidence normale. C'est une des conditions les plus utilisées.

- $Z^{\delta}_{eff} = ik\mu\delta\left(1 + \frac{1}{k_d^2}\nabla_\Gamma \operatorname{div}_\Gamma\right)$ a été introduite par Engquist et Nédélec [**EN93**] et correspond à un développement asymptotique d'ordre 1. Ici, k_d est le nombre d'onde de la fine couche de diélectrique.
- $Z^{\delta}_{eff} = ik\mu\delta\left(1 - \delta(\mathcal{C}-\mathcal{H}) + \frac{1}{k_d^2}\nabla_\Gamma\left(1+\delta\mathcal{H}\right)\operatorname{div}_\Gamma\right)$ avec $\mathcal{C}$ l'opérateur de courbure relativement à $\mathbf{n}$ et $\mathcal{H}$ la courbure moyenne. Cette condition d'ordre 2 a été introduite par Haddar et Joly [**HJ02**].
- Terminons par une condition d'ordre 3 proposée par A. Bendali et K. Lemrabet [**BL08**] :

$$\begin{aligned} Z^{\delta}_{eff} &= ik\mu\delta\Big(\big(1+\delta(\mathcal{C}-\mathcal{H}) + \frac{2}{3}\delta^2(\mathcal{C}-\mathcal{H})(\mathcal{C}-2\mathcal{H}) \\ &+ \frac{1}{k_d^2}\nabla_\Gamma(1+\delta\mathcal{H}+\frac{\delta^2}{3}(4\mathcal{H}^2-G))\operatorname{div}_\Gamma \\ &+ \frac{\delta^2}{3}(k_d^2(1+\frac{1}{k_d^2}\nabla_\Gamma\operatorname{div}_\Gamma)^2 - \nabla_\Gamma\times\operatorname{curl}_\Gamma)\Big) \end{aligned}$$

Nous nous sommes principalement intéressés à la résolution numérique du problème (52) en utilisant des équations intégrales pour un opérateur d'impédance $Z^{\delta}_{eff} = \eta(x)$ de type Léontovitch où η permet de tenir compte de la présence de plusieurs type de matériaux, en particulier les parties parfaitement conductrices correspondent à $\eta = 0$. Cette condition englobe donc le cas mixte conducteur-impédant (voir figure (1)). Le problème que l'on désire traiter est bien posé si $\eta \in \mathrm{L}^\infty(\Gamma)$ est borné inférieurement par une constante strictement positive et pour une régularité lipschitzienne de la frontière Γ. Je tiens à préciser dès à présent (puisque que je n'en parlerai pas dans la suite) que dans [**CMP08**], nous avons proposé une méthode CFIE efficace pour traiter le problème (52) pour la condition de Léontovitch qui a été étendue aux conditions d'impédance d'ordre élevé présentées ci-dessus [**BFLP09**].

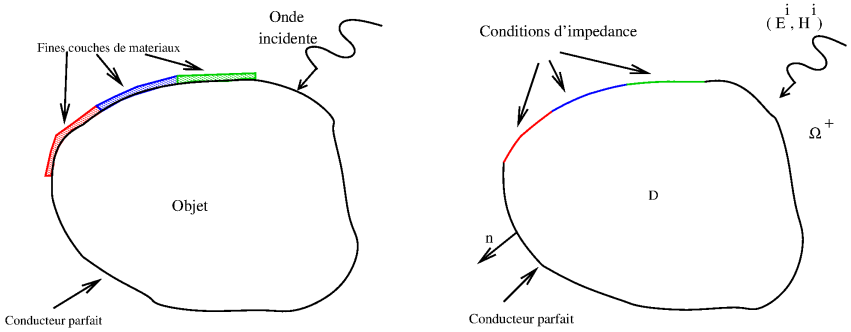


FIGURE 1. Description du problème

Bien qu'en apparence simple, la condition d'impédance de Léontovitch induit un certain nombre de problèmes lorsque l'on veut utiliser un formalisme basé sur des équations intégrales. Les principaux sont :

- Cette condition fait apparaître une relation entre deux quantités de nature fonctionnelle différente. En effet, les quantités $\mathbf{n}\times(\mathbf{E}\times\mathbf{n})$ et $\mathbf{n}\times\mathbf{H}$ appartiennent, respectivement, aux espaces $\mathrm{H}^{-1/2}(\operatorname{div}_\Gamma, \Gamma)$ et $\mathrm{H}^{-1/2}(\operatorname{rot}_\Gamma, \Gamma)$ qui sont duaux. Ceci rend difficile l'élimination d'un des courants en utilisant la condition d'impédance. En effet, l'idéal serait d'utiliser une approximation conforme dans l'espace $\mathrm{H}^{-1/2}(\operatorname{div}_\Gamma, \Gamma)\cap\mathrm{H}^{-1/2}(\operatorname{rot}_\Gamma, \Gamma)$ qui n'existe pas actuellement lorsque la surface n'est pas régulière.

- L'approximation des formulations intégrales en électromagnétisme est généralement sur les éléments finis de Raviart-Thomas d'ordre le plus bas (RT0). Or, l'évaluation de l'opérateur de rotation $\mathbf{n}\times$ n'est pas consistante du point de vue mathématique pour cet espace. En effet, la dualité L^2 entre les espaces RT0 et $\mathbf{n}\times$RT0 ne vérifie pas une condition inf-sup uniforme **[CN02b]**. Néanmoins, Buffa-Christiansen ont proposé récemment un espace d'élément finis basé sur un raffinement barycentrique **[BC07]** qui permet de réaliser une discrétisation stable de cette dualité. Cet espace alourdit considérablement l'assemblage des problèmes.
- Après discrétisation, les formulations de type équations intégrales de frontière conduisent à la résolution d'un système linéaire dense qui nécessite l'utilisation d'un solveur itératif lorsque la taille du problème devient trop importante. On sait que le taux de convergence de la plupart des méthodes itératives dépend, entre autre, du conditionnement et de la répartition dans le plan complexe, des valeurs propres de la matrice représentant le système. Nous avons constaté que la présence de l'opérateur d'impédance surtout s'il est variable, provoque une dispersion du spectre du système.
- Les formulations classiques **[BFG99, Lan95]** pour résoudre ce type de problèmes dégénèrent en une équation en champs électrique (EFIE) sur les parties métalliques de la surface de l'objet. Il est connu que la convergence des solveurs itératifs est lente lorsque l'on résout le système linéaire provenant de la discrétisation de l'EFIE.

Nous avons donc travaillé à l'amélioration des techniques intégrales pour résoudre ce type de problème. Plus précisément, notre objectif est de proposer des formulations adaptées à une résolution itérative. Dans ce rapport, je présente uniquement les résultats les plus significatifs.

2.2. Synthèse sur la méthodologie

On s'intéresse à l'aide d'une équation intégrale à la résolution d'un problème de diffraction par un objet imparfaitement conducteur (52). Les formulations intégrales utilisées pour résoudre ce problème sont construites à partir des formules de représentation de Stratton-Chu :

$$\left\{ \begin{array}{rcl} \mathbf{E}(x) & = & \mathcal{T}\mathbf{J}(x) + \mathcal{K}\mathbf{M}(x) \\ \mathbf{H}(x) & = & -\mathcal{K}\mathbf{J}(x) + \mathcal{T}\mathbf{M}(x) \end{array} \right. \tag{53}$$

où $\mathbf{J}$ et $\mathbf{M}$ sont respectivement les courants électrique et magnétique et

$$\mathcal{T} = \frac{1}{ik}\nabla\times\nabla\times G, \quad \mathcal{K} = \nabla\times G \quad \text{avec} \quad G\mathbf{u}(\mathbf{x}) = -\frac{\mathbf{1}}{\mathbf{4}\pi}\int_{\mathbf{\Gamma}} \frac{\mathbf{e}^{\mathbf{ik}\|\mathbf{x}-\mathbf{y}\|}}{\|\mathbf{x}-\mathbf{y}\|}\mathbf{u}(\mathbf{y})\mathbf{dy}$$

ainsi qu'en utilisant les formules de trace :

$$\left\{ \begin{array}{rcl} \mathbf{E}^{\pm}_{tan} & = & T\,\mathbf{J} \pm \frac{1}{2}\mathbf{n}\times\mathbf{M} + K\mathbf{M} \\ \mathbf{H}^{\pm}_{tan} & = & T\,\mathbf{M} - \big(\pm\frac{1}{2}\mathbf{n}\times\mathbf{J} + K\mathbf{J}\big) \end{array} \right. \tag{54}$$

où T et K sont les opérateurs de surface classiques obtenus en déterminant les composantes tangentielles intérieure et extérieure des formules de Stratton-Chu et $\mathbf{v}_{tan}$ signifie $\mathbf{n}\times(\mathbf{v}\times\mathbf{n})$ *i.e* la composante tangentielle de $\mathbf{v}$.

En utilisant les formules (54) et la condition aux limites, on peut construire de nombreuses formulations intégrales conduisant à des propriétés numériques différentes. Le choix de la "bonne" formulation n'est pas une tâche évidente mais il existe néanmoins des critères pour guider la construction :

- La formulation est-elle valable pour les surfaces ouvertes, fermées ou les deux? *i.e* quelle est son spectre d'application?
- l'équation est-elle bien posée pour toutes les fréquences c'est-à-dire y a t-il existence et unicité de la solution quelle que soit la fréquence? Cette propriété est important pour éviter la présence de courants parasites polluant la solution mais aussi pour éviter la détérioriation du conditionnement du système lorsque l'on travaille près d'une fréquence de résonance qui elle n'est pas connue.
- l'équation est-elle une perturbation compacte de l'opérateur identité c'est-à-dire de la forme I+C avec un opérateur C dont le spectre est fini ou formé d'une suite tendant vers 0? Le rayon spectral de C est-il suffisament petit pour avoir une bonne répartition des valeurs propres dans le plan complexe et conduire à une résolution rapide du système linéaire par une méthode de Krylov?
- La solution numérique est-elle précise?

Pour ce qui est de la première question, la réponse est simple puisque, à part l'équation en champ électrique (EFIE) qui est valable à la fois pour les surfaces ouvertes et fermées, toutes les autres sont par construction restreintes aux surfaces fermées. Le dernier point est quant à lui souvent le plus problématique et n'est généralement vérifié qu'*a posteriori* lors de simulations numériques sur des cas-tests de référence. Ce sont donc les points 2 et 3 qui servent généralement de critère de choix.

Le point 2 a, par exemple, permis de construire la fameuse équation en champs combinés CFIE (cf. Tab. 1) qui est actuellement la méthode la plus utilisée dans l'industrie. Cette méthode a néanmoins montré des limitations qui deviennent de plus en plus problématiques à cause de la complexité croissante des problèmes à résoudre. Un premier défaut de cette formulation est la précision insuffisante de la solution obtenue dans certaines situations. Ceci se manifeste généralement lorsque la géométrie présente des singularités (par exemple une pointe) et lorsque l'on veut detecter des faibles niveaux d'énergie diffactractée. Le second défaut de cette formulation CFIE est la détérioration du taux de convergence des solveurs itératifs lorsque la taille du problème augmente par rapport à la longueur d'onde considérée. La solution la plus immédiate pour pallier à ce problème est l'utilisation d'un préconditionneur algébrique de type SPAI (SParse Approximate Inverse). Néanmoins, l'efficacité de celui-ci demande de répondre à une question non triviale : quel taux de remplissage dois-je choisir pour un problème donné?

La table 1 résume les propriétés des formulations intégrales classiques. On voit qu'aucune ne satisfait les points 2 et 3 simultanément. Pour arriver à satisfaire ces deux critères, il est apparu ces dernières années que la problématique du conditionnement pouvait être prise en compte dès la conception des équations, c'est-à-dire au niveau continu de la formulation du problème plutôt qu'après sa discrétisation **[CN02a, ABL07, Dar06, AD07, SW98]**. Des opérateurs régularisant jouant le rôle de préconditionneurs sont ainsi construits à partir de l'examen mathématique des opérateurs fondamentaux de l'électromagnétisme. Les succès remportés récemment par ces nouvelles techniques de préconditionnement nous ont naturellement conduits à les étudier dans le cadre de la diffraction d'une onde électromagnétique par des objets imparfaitement conducteurs (modélisés par une condition d'impédance). Nous avons ainsi proposé une nouvelle famille d'équations intégrales baptisée GCSIE appartenant aux méthodes indirectes (on parle aussi d'équations en sources). Le formalisme GCSIE est dépendant du choix d'un opérateur R qui tient lieu de préconditionneur continu. La particularité de cette approche est que l'opérateur optimal à adopter est clairement connu. Par exemple, dans le cas d'un problème métallique, il s'agit de l'opérateur Dirichlet-to-Neuman (DtN) (ou admittance pour les physiciens) couplant sur la surface d'un objet la trace tangentielle du champ électrique à celle du champ magnétique. En effet, lorsque R est égal à DtN le formalisme GCSIE conduit à une équation dont l'opérateur est l'identité. S'il n'est pas possible en dehors de quelques géométries canoniques de connaître DN, on peut approcher "assez précisément" cet opérateur, particulièrement dans le régime des hautes fréquences.

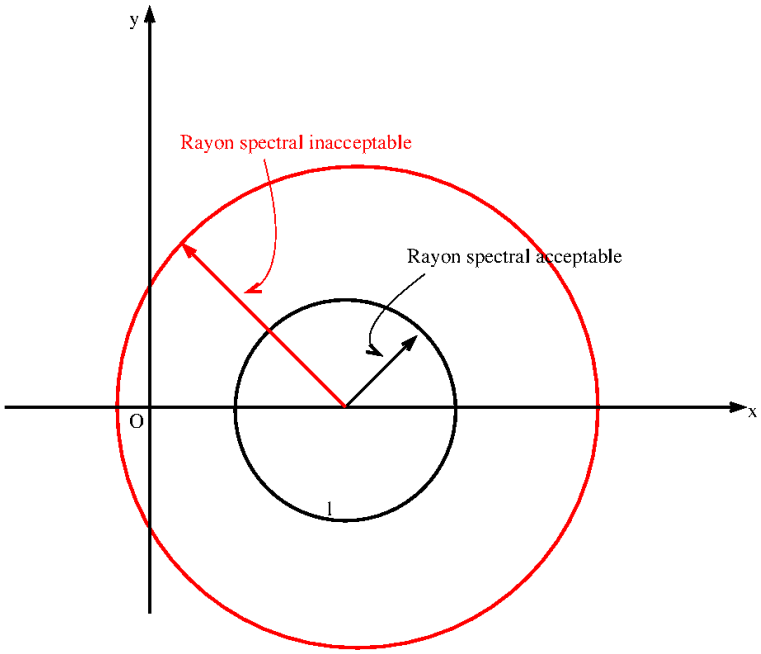


FIGURE 2. Comportement spectral désiré

Acronyme	Equation	Bien posée pour toutes les fréquences	Perturbation compacte de l'identité
EFIE	$\mathrm{T}\,\mathbf{J} = -\mathbf{E}^i_{tan}$	Non	Non
MFIE	$(\frac{1}{2}Id + \mathbf{n} \times \mathrm{K})\mathbf{J} = \mathbf{n} \times \mathbf{H}^i$	Non	Oui
CFIE	$[(1-\alpha)\mathrm{T} + \alpha(\frac{1}{2}Id + \mathbf{n} \times \mathrm{K})]\mathbf{J}$ $= -(1-\alpha)\mathbf{E}^i_{tan} + \alpha\mathbf{n} \times \mathbf{H}^i$ avec $\alpha \in [0,1]$ ou encore $[\beta\mathrm{T} + (\frac{1}{2}Id + \mathbf{n} \times \mathrm{K})]\mathbf{J}$ $= \beta\mathbf{E}^i_{tan} + \mathbf{n} \times \mathbf{H}^i$	Oui	Non

TABLE 1. Les formulations classiques

2.3. Dérivation formelle de formulations GCSIE

Dans cette partie, nous présentons une ébauche dans la dérivation d'une formulation GCSIE. Les problèmes qui nous intéressent peuvent s'écrire formellement :

$$\text{Trouver } u \in W \text{ tel que } \gamma\, u = u_0$$

où

- $u_0 \in \mathcal{D}'(\Gamma)$ est une distribution de Γ la frontière d'un ouvert borné D,
- W est l'espace des solutions admissibles,
- γ représente la condition aux limites.

Nous supposons à présent que l'on a à notre disposition une formule de reconstruction qui nous permet de reconstruire chaque champ $w \in W$ à partir de ses données de Cauchy $\gamma_c w$ via un potentiel de Caldéron $\mathcal{C} : \mathcal{D}'(\Gamma) \to W$:

$$w = \mathcal{C}(\gamma_c w)$$

C'est exactement ce que réalise les formules de Stratton-Chu avec $\gamma_c : (\mathbf{E}, \mathbf{H}) \to (\mathbf{n}\times\mathbf{E}, \mathbf{n}\times\mathbf{H})$.

Maintenant, si on suppose que notre problème est bien posé alors il existe un opérateur R défini par

$$R : \gamma w \mapsto \gamma_c w$$

Tout ceci nous permet d'écrire la relation fondamentale :

$$\gamma\, \mathcal{C} R = Id$$

C'est pour ces raisons que l'on dit que R est l'opérateur régularisant optimal associé à notre problème. Il est évident que l'opérateur R est généralement inconnu. Néanmoins, si on est capable de calculer une approximation $\widetilde{R}$ de celui-ci, on peut considérer une nouvelle paramétrisation de la solution de notre problème :

$$w = \mathcal{C}\widetilde{R}u$$

où u est solution de l'équation intégrale:

$$\text{Trouver } u \text{ tel que } \gamma\, \mathcal{C}\widetilde{R}u = u_0 \tag{55}$$

Il est à noter que bien que $\widetilde{R}$ soit une approximation de R, w est toujours la solution exacte de notre problème. L'opérateur $\widetilde{R}$ agit sur les propriétés de l'équation intégrale. En effet, rappelons que si $\widetilde{R} = R$ alors la nouvelle équation est triviale et si on est capable d'expliciter une "bonne" approximation de R (qui permet de réaliser la décomposition "identité + compact" de l'opérateur intégral avec un bon regroupement du spectre) alors on s'attend à ce que le système linéaire obtenu après discrétisation se prête bien à une résolution itérative. Par exemple, Fig. 2 illustre le type de regroupement du spectre auquel on veut aboutir.

2.4. GCSIE pour la condition de Léontovitch

Nous considérons à présent le problème (52) pour une condition aux limites de Léontovitch. Dans ce cas, on utilise la méthodologie précédente avec :

1. $w = (\mathbf{E}, \mathbf{H})$,
2. $\gamma\, w = \mathbf{n} \times \mathbf{E}_{|\Gamma} + \eta \mathbf{n} \times (\mathbf{H}_{|\Gamma} \times \mathbf{n})$,
3. $\gamma_c : (\mathbf{E}, \mathbf{H}) \to (\mathbf{n} \times \mathbf{E}, \mathbf{n} \times \mathbf{H})$,
4. $W = \{(\mathbf{E}, \mathbf{H}) : \Omega \to \mathbb{C}^3 : \nabla \times \mathbf{E} - ik\mathbf{H} = 0, \nabla \times \mathbf{H} + ik\mathbf{E} = 0$ et condition de radiation$\}$
5. $\mathcal{C}(\mathbf{u}, \mathbf{v}) = \mathcal{L}\mathbf{v} - \mathcal{K}\mathbf{u}$ est donné par le formule de Stratton-Chu (53).

Dans **[Per10b]**, on a montré que l'opérateur régularisant optimal R associé au problème impédant peut être défini à l'aide de l'opérateur d'admittance extérieur de Γ c'est-à-dire de l'opérateur Y_+ qui à $\mathbf{n} \times \mathbf{H}$ associe $\mathbf{n} \times \mathbf{E}$ par la relation $\mathbf{n} \times \mathbf{E} = -Y_+(\mathbf{n} \times \mathbf{H})$:

$$R = (Id + \eta \mathbf{n} \times R_H\, ,\, R_H) \tag{56}$$

où

$$R_H = \begin{cases} (\eta \mathbf{n} \times Id - Y_+)^{-1} \text{ si } \eta \neq 0 \\ Y_+ \text{ si } \eta = 0 \end{cases} \tag{57}$$

L'approximation $\widetilde{R}$ de R est alors construite à partir d'une approximation de l'opérateur d'admittance $\widetilde{Y}_+$ en remplaçant R_H par

$$\widetilde{R}_H = \begin{cases} \left(\eta \mathbf{n} \times Id - \widetilde{Y}_+\right)^{-1} \text{ si } \eta \neq 0 \\ \widetilde{Y}_+ \text{ si } \eta = 0 \end{cases} \tag{58}$$

L'approximation $\widetilde{Y}_+$ est quant-à-elle construite à partir de l'admittance exacte sur un plan P qui est donnée simplement par $-2\mathbf{n} \times \mathcal{T}_{tan}$ et qui peut encore s'exprimer dans les potentiels de Helmholtz (*i.e* les potentiels scalaires donnés par la décomposition de Helmholtz d'un champ de vecteur tangent à P) par :

$$Y_+ = \frac{1}{k}(-\mathbf{n} \times \nabla_P \;\; \nabla_P) \begin{pmatrix} 0 & -(\Delta_P + k^2\, Id)^{1/2} \\ k^2(\Delta_P + k^2\, Id)^{-1/2} & 0 \end{pmatrix} \begin{pmatrix} P_{loop} \\ P_{star} \end{pmatrix} \tag{59}$$

où les opérateurs P_{loop} et P_{star} sont définis naturellement par la décomposition de Helmholtz

$$\mathbf{u} = -\mathbf{n} \times \nabla_P\, P_{loop}\mathbf{u} + \nabla_P P_{star}\mathbf{u} \tag{60}$$

avec $P_{loop}\mathbf{u} = -\Delta_P^{-1}\, curl_{\P}\, \mathbf{u}$ et $P_{star}\mathbf{u} = \Delta_P^{-1}\, div_{\P}\, \mathbf{u}$, Δ_P est l'opérateur de Laplace-Beltrami associé à P et Δ_P^{-1} est l'inverse Δ_P au sens de Moore-Penrose. Plus précisément pour une surface quelconque Γ, nous avons choisi :

$$\widetilde{Y}_+ = \frac{1}{k}(-\mathbf{n} \times \nabla_\Gamma \;\; \nabla_\Gamma) \begin{pmatrix} 0 & -(\Delta_\Gamma + k^2\, Id)^{1/2} \\ k^2(\Delta_\Gamma + k^2\, Id)^{-1/2} & 0 \end{pmatrix} \begin{pmatrix} P_{loop} \\ P_{star} \end{pmatrix} \tag{61}$$

On obtient ainsi l'équation GCSIE suivante :

$$\left(\frac{1}{2} + \mathbf{n} \times T\, \widetilde{R}_H - \mathbf{n} \times K\, \widetilde{R}_E - \eta T\, \widetilde{R}_E - \eta K\, \widetilde{R}_H\right) u = -\mathbf{n} \times \mathbf{E}^{inc} + \eta \mathbf{H}^{inc}_{tan} \tag{62a}$$

$$\widetilde{R}_H = \left(\eta \mathbf{n} \times Id - \widetilde{Y}_+\right)^{-1} \text{ et } \widetilde{R}_E = Id + \eta \mathbf{n} \times \widetilde{R}_H \tag{62b}$$

L'analyse de l'équation (62) est basée sur le calcul symbolique. Nous avons donc supposé pour cela que Γ est $\mathcal{C}^\infty$ et nous avons montré que cette équation est bien posée pour toutes les fréquences et qu'elle se décompose en une perturbation compacte de l'identité (plus précisément de la forme αId) [**Per10b**]. Pour assurer l'injectivité de cet opérateur, il faut néanmoins remplacer dans $\widetilde{Y}_+$ le nombre d'onde k par $k + i\varepsilon$ pour un $\varepsilon > 0$. Numériquement, le système linéaire obtenu après discrétisation est résolu par un solveur itératif de type GMRES sans utiliser de préconditionneur. En particulier, les opérateurs de type racine carrée sont localisés en utilisant une approximation de Padé avec rotation de branche [**MZB97**]. Cette rotation permet d'améliorer la convergence de cette localisation du fait de la spécificité du Laplacien Beltrami qui induit des valeurs propres proches de la branche de coupure classique. Nous avons aussi constaté numériquement qu'il n'était pas nécessaire de prendre un ordre élevé pour l'approximation de Padé *i.e* $p = 2$ est souvent suffisant. De plus, le taux de convergence du solveur itératif est quasi indépendant de la fréquence et de la finesse de la discrétisation utilisée. La grande souplesse d'utilisation (aucun paramètre n'est à régler contrairement à un préconditionneur algébrique) et la robustesse de ce type de formulation en font un candidat sérieux pour la résolution des applications futures en électromagnétisme. Ce résultat est mis en lumière sur la figure 3 où notre équation est comparée avec une approche standard

dénommée BGL [**Lan95**]. Nous renvoyons à [**LMP14**] pour la discrétisation de cette équation et de nombreux autres résultats numériques.

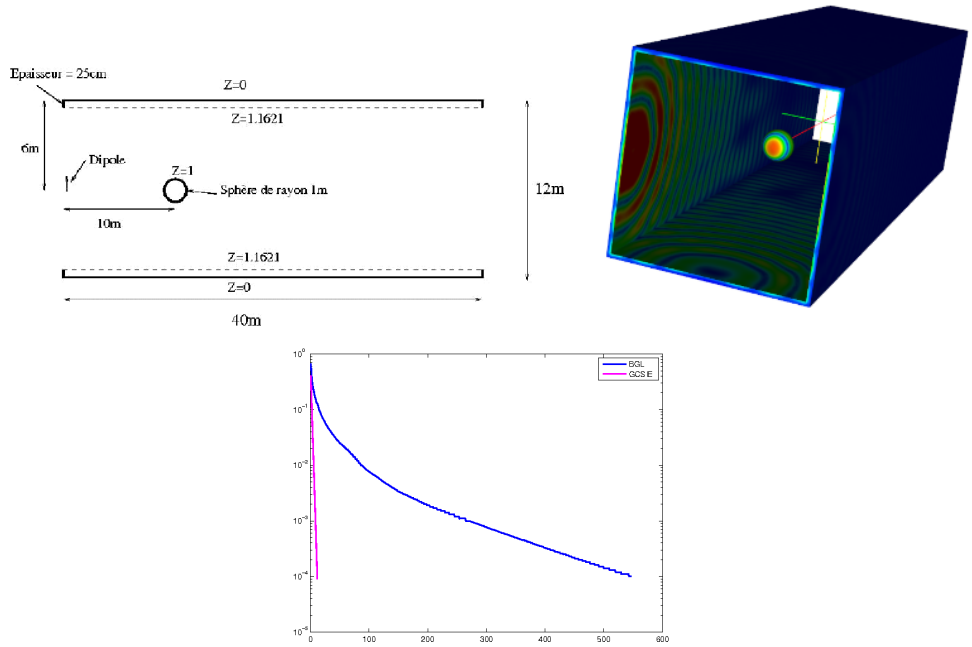


FIGURE 3. Calcul du rayonnement électromagnétique dans une chambre anéchoïque modélisée par une condition de Léontovitch variable et comparaison des historiques GMRES avec une approche standard (BGL)

2.5. Extension aux problèmes de transmission

Fort du succès pour les problèmes de diffraction avec conditions impédantes ou d'objets parfaitement conducteurs, nous avons proposé dans [**LMP15**] une GCSIE pour un problème de transmission. Dans ce cas, on considère un domaine borné Ω_2 composé d'un diélectrique homogène $(\varepsilon_2\,,\,\mu_2)$ et son complémentaire Ω_1 de caractéristique $(\varepsilon_1\,,\,\mu_1)$. La méthodologie précédente s'applique alors avec :

1. $w = (\mathbf{E}\,,\,\mathbf{H})$,
2. $W = W^+ \oplus W^-$ avec
   - $(\mathbf{E}\,,\,\mathbf{H}) \in W^+$ si et seulement si $(\mathbf{E}\,,\,\mathbf{H})$ vérifie:

$$\begin{cases} \nabla \times \mathbf{E} - i\omega\mu_1\mathbf{H} = 0 \text{ dans } \Omega_1, \\ \nabla \times \mathbf{H} + i\omega\varepsilon_1\mathbf{E} = 0 \text{ dans } \Omega_1, \\ \gamma_1\mathbf{E} := \mathbf{n}_1 \times \mathbf{E} \text{ existe}, \\ (\mathbf{E}\,,\,\mathbf{H}) \text{ vérifie la condition de radiation à l'infini.} \end{cases} \tag{63}$$

où $\mathbf{n}_1$ est la normale unitaire extérieure à Ω_1.

- $(\mathbf{E}, \mathbf{H}) \in W^-$ si et seulement si $(\mathbf{E}, \mathbf{H})$ vérifie:

$$\begin{cases} \nabla \times \mathbf{E} - i\omega\mu_2\mathbf{H} = 0 \text{ dans } \Omega_2, \\ \nabla \times \mathbf{H} + i\omega\varepsilon_2\mathbf{E} = 0 \text{ dans } \Omega_2, \\ \gamma_2\mathbf{E} := \mathbf{n}_2 \times \mathbf{E} \text{ existe.} \end{cases} \tag{64}$$

  où $\mathbf{n}_2$ est la normale unitaire extérieure à Ω_2.

3. $\gamma(\mathbf{E}, \mathbf{H}) = (\gamma_1\mathbf{E} - \gamma_2\mathbf{E}, \gamma_1\mathbf{H} - \gamma_2\mathbf{H})$,
4. $u_0 = -(\mathbf{n} \times \mathbf{E}^{inc}, \mathbf{n} \times \mathbf{H}^{inc})$ avec $\mathbf{n} = \mathbf{n}_2$ et $(\mathbf{E}^{inc}, \mathbf{H}^{inc})$ une onde plane incidente,
5. $\gamma_c : (\mathbf{E}, \mathbf{H}) \to (\mathbf{n} \times \mathbf{E}, \mathbf{n} \times \mathbf{H})$ et le potentiel de reconstruction est toujours défini à partir de la formule de stratton-Chu.

Dans [**LMP15**], nous avons proposé une approximation de l'opérateur R sous-jacent au problème de transmission considéré qui conduit à la formulation intégrale bien posée pour toutes les fréquences suivante :

$$\left(\begin{bmatrix} \frac{1}{2}Id - \mathbf{n}\times K_1 & iZ_1\mathbf{n}\times T_1 \\ -\frac{i}{Z_1}\mathbf{n}\times T_1 & \frac{1}{2}Id - \mathbf{n}\times K_1 \end{bmatrix} \widetilde{R}_1 + \begin{bmatrix} \frac{1}{2}Id + \mathbf{n}\times K_2 & -iZ_2\mathbf{n}\times T_2 \\ \frac{i}{Z_2}\mathbf{n}\times T_2 & \frac{1}{2}Id + \mathbf{n}\times K_2 \end{bmatrix} \left(Id - \widetilde{R}_1\right) \right) \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} \tag{65}$$

où

$$\widetilde{R}_1 = \begin{pmatrix} \widetilde{A} & -Z_1^2\widetilde{Y}_1\widetilde{B} \\ \widetilde{Y}_1\widetilde{A} & \widetilde{B} \end{pmatrix} \tag{66}$$

avec $\widetilde{Y}_1$ est $\widetilde{Y}_+$ avec $k = k_1$ (voir équation (61))

$$\widetilde{A} = \frac{Z_1k_2}{Z_1k_2 + Z_2k_1}\Pi_{loop} + \frac{Z_1k_1}{Z_1k_1 + Z_1k_1}\Pi_{star} \tag{67a}$$

$$\widetilde{B} = \frac{Z_1k_1}{Z_1k_1 + Z_2k_2}\Pi_{loop} + \frac{Z_2k_1}{Z_1k_2 + Z_2k_1}\Pi_{star} \tag{67b}$$

où $\Pi_{loop} := -\mathbf{n} \times \nabla_\Gamma P_{loop}$ et $\Pi_{star} = \nabla_\Gamma P_{star}$. De plus, on montre que l'équation (65) peut être vue comme une perturbation compacte de l'opérateur identité. On voit par exemple sur la figure 4 le regroupement du spectre de la GCSIE autour du point $(1., 0)$, qui montre que l'opérateur $\widetilde{R}$ est bien une bonne approximation de l'opérateur optimal R. De plus, sur cette même figure, on voit que la formulation classique PMCHWT [**JSC02**] a un spectre qui tourne autour de l'origine et qui est plus dispersé dans le plan complexe. Ce type de distribution est moins adapté à une résolution itérative. Enfin la figure 5 confirme bien la supériorité de la GCSIE lors d'une résolution itérative pour une géométrie singulière. Nous renvoyons à [**LMP15**] pour les détails de ces comparaisons.

2.6. Analyse d'erreur *a posteriori* et algorithmes auto-adaptatifs

Jusqu'à présent, la majorité des travaux que j'ai présentés ont concerné l'accélération de la résolution des équations intégrales de frontière pour traiter des problèmes réalistes. Elles sont aujourd'hui utilisées aussi bien pour calculer le rayonnement d'une antenne spatiale ou la Signature Equivalente Radar d'un avion, que pour déterminer la fréquence de résonance d'un filtre à résonateurs couplés ou pour simuler la propagation des ondes sismiques dans un bassin sédimentaire. Comparativement

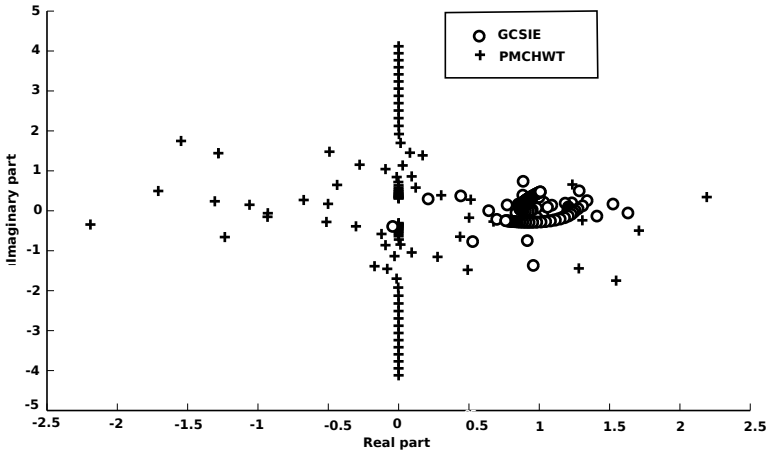


FIGURE 4. Comparaison des spectres des opérateurs GCSIE et PMCHWT pour une sphère diélectrique de caractéristiques $\varepsilon_2 = 4$ et $\mu_2 = 1$ et pour une fréquence de $225Mhz$.

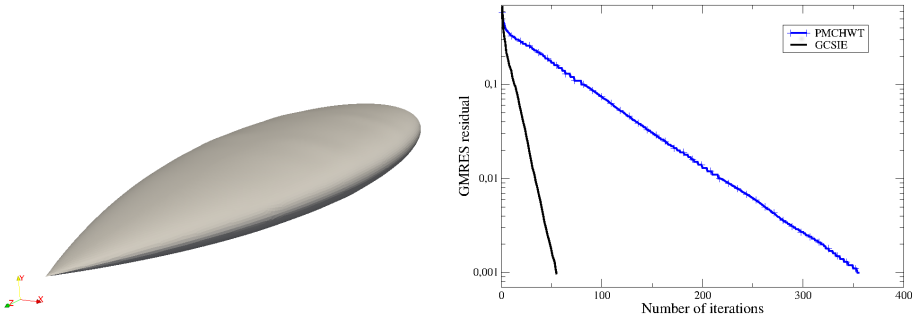


FIGURE 5. Comparaison de l'historique de convergence GMRES entre la GCSIE et la PMCHWT (approche classique préconditionnée) pour une amande diélectrique de caractéristiques $\varepsilon_2 = 4$ et $\mu_2 = 1$ pour un maillage composé de 134648 triangles soit 403944 degrés de liberté.

aux méthodes de type éléments finis, les méthodes intégrales demeurent pourtant insuffisamment popularisées, leur utilisation restant généralement l'affaire de spécialistes expérimentés.

Nous pensons que l'un des obstacles à une plus grande exploitation de ces méthodes est l'absence d'outils automatiques, permettant de garantir la précision de la solution calculée. L'outil le plus adéquat pour réaliser cette tâche est ce qu'on appelle un estimateur d'erreur a posteriori. Ce type d'estimateur représente une quantité calculable (ne dépendant que de la solution discrète et des données du problème) qui est équivalente à l'erreur entre la solution exacte (qui est inconnue) et la solution approchée. L'adaptation automatique de maillage via l'analyse d'erreur a posteriori a prouvé son efficacité pour de nombreux problèmes en permettant une réduction notable du coût de calcul (en réduisant le nombre de degrés de liberté) tout en atteignant une solution approchée à la précision désirée. La principale difficulté est que ce type d'approche dépend fortement du problème traité et du schéma numérique utilisé. Le développement de tels outils reste donc un challenge dans de nombreuses situations. En particulier, ces techniques sont quasiment inexistantes dans le domaine des équations intégrales, alors que leur importance n'a fait que croître dans le domaine des méthodes de type éléments finis par exemple. Il existe pourtant un certain nombre de résultats théoriques montrant la possibilité d'établir des estimations d'erreur *a posteriori* pour

les formulations intégrales, mais ces estimateurs ne semblent pas avoir été expérimentés, et encore moins exploités, de façon pratique et systématique dans les grands codes de calcul. De plus, peu de ces résultats traitent le cas des noyaux oscillants qui sont caractéristiques des phénomènes de propagation. Le développement d'estimateurs d'erreur est donc un point essentiel pour ensuite proposer des procédures d'adaptation de maillage, qui constituent une méthode complémentaire à la FMM pour réduire le coût de calcul des BEMs. En choisissant le maillage optimal pour obtenir un niveau voulu de précision, on peut réduire le nombre de degrés de liberté.

Dans la thèse de Marc Bakry **[Bak16]** et dans le projet ANR RAFFINE, nous nous sommes intéressés à cette problématique pour des équations intégrales en acoustique et en électromagnétisme. La principale difficulté dans l'analyse d'erreur *a posteriori* pour les équations intégrales est la non-localité des opérateurs et des normes. Or, pour être utilisé dans un algorithme de raffinement auto-adaptatif, un indicateur d'erreur doit se décomposer en contributions locales aux cellules du maillage. Il existe néanmoins des techniques de localisation standard permettant de contourner le problème. Nous avons dans un premier temps généralisé ces indicateurs classiques en acoustique et en particulier, nous avons démontré un résultat de convergence quasi-optimale d'un algorithme autoadaptatif guidé pour un estimateur de type résidu **[Bak16]**. Ce résultat est important car il confirme bien que l'on peut raffiner localement dans le cadre d'opérateurs non-locaux tout en assurant la convergence de la méthode. C'est un résultat que l'on constate en pratique mais qui n'est pas intuitif car l'erreur associée à un triangle du maillage pourrait provenir de zones éloignées de celui-ci.

Un des défauts des indicateurs d'erreur *a posteriori* standards est qu'ils ne permettent pas de donner une valeur suffisamment précise de l'erreur induite par un schéma numérique. Ils sont de bons indicateurs pour déterminer les endroits où l'erreur est la plus grande et donc où il faut raffiner un maillage, par contre ils ne donnent généralement pas une information suffisamment précise pour être utilisés comme critère d'arrêt d'une méthode de raffinement auto-adaptative. Il subsiste, en effet, une constante de proportionnalité entre l'indicateur et l'erreur exacte dont on perd généralement le contrôle lors de l'utilisation des techniques de localisation standards. Notre contribution la plus importante a été de proposer une technique originale de localisation **[BPC78, BPC17]** permettant de dériver des indicateurs d'erreur a posteriori fiables et efficaces pour une grande classe de problèmes. Ces indicateurs sont de plus asymptotiquement exacts sous une condition de régularité.

Décrivons succinctement cette méthodologie sur un problème abstrait avec lequel on peut exprimer la plupart des équations intégrales en acoustique et électromagnétisme. Considérons $\mathcal{A} : H \to H^*$ un opérateur linéaire agissant sur un espace de Hilbert H et à valeur dans le dual topologique H^* de ce dernier vérifiant $\mathcal{A} = \mathcal{A}_0 + \mathcal{K}$ où $\mathcal{A}_0$ est un opérateur continu et T-coercif *ie* il existe $T \in \mathcal{L}(H, H)$ un opérateur bijectif tel qu'il existe $\alpha > 0$, $\langle \mathcal{A}_0 \mathbf{v}\,,\, T\mathbf{v}\rangle \geq \alpha \|\mathbf{v}\|_H^2$ et $\langle \mathcal{K}\cdot\,,\, T\cdot\rangle$ est une forme bilinéaire compacte où $\langle\cdot\,,\,\cdot\rangle$ représente les crochets de dualité. Nous nous intéressons à des approximations de type Galerkin du problème $\mathcal{A}\,u = b$ dans une suite d'espaces emboités $(V_l)_{l\in I\!N}$ *i.e* $V_l \subset V_{l+1} \subset H$: trouver $u_l \in V_l$ tel que pour tout $v_l \in V_l$, $< \mathcal{A}\,u_l\,,\, v_l >=< b\,,\, v_l >$. Nous supposons que ces problèmes discrets sont bien posés, que $\lim\limits_{l\to+\infty} u_l = u$ dans H et $V_\infty := \overline{\cup_{l\in\mathbf{N}} V_l} = H$.

Nous avons alors le résultat fondamental suivant qui va guider la construction de nos estimateurs *a posteriori*:

Théorème 3. — *Soit $\Lambda : H^* \to V \subset [L^2(\Gamma)]^d$ un isomorphisme sur un sous-espace fermé V de $[L^2(\Gamma)]^d$. Alors l'estimateur d'erreur a posteriori défini par $\eta_\Lambda := \|\Lambda\, r_l\|_{0,\Gamma}$ où $r_l := b - \mathcal{A}u_l$ est efficace, fiable et local.*

De plus, s'il existe Λ tel que l'identité $\Lambda^*\Lambda\mathcal{A}_0 = T + \mathcal{K}_1$ soit vérifiée où Λ^* est l'opérateur adjoint de Λ et $\mathcal{K}_1$ est une perturbation compacte, alors η_Λ est asymptotiquement exact pour la norme de l'erreur $|||u - u_l|||^2 = \langle \mathcal{A}_0(u - u_l)\,,\, T(u - u_l)\rangle$.

Remarques :

1. La première partie de ce théorème est triviale. En effet, puisque Λ est un isomorphisme, on a $\|\Lambda^{-1}\|_{op}\|r_l\|_{H^*} \leq \eta_\Gamma \leq \|\Lambda\|_{op}\|r_l\|_{H^*}$. Nous concluons en utilisant le fait que $\|r_l\|_{H^*}$ est toujours équivalent à $\|u-u_l\|_H$ lorsque le problème continu considéré est bien posé.
2. La seconde partie découle des propriétés suivantes : $\lim_{l\to+\infty} u_l = u$ dans H et $V_\infty := \overline{\cup_{l\in\mathbf{N}} V_l} = H$ impliquent $\bar{e}_l := (u-u_l)/\|u-u_l\|_H \rightharpoonup 0$ faiblement dans H lorsque $l \to +\infty$ et par conséquence, pour tout opérateur compact $\widetilde{\mathcal{K}} : H \to W$, $\widetilde{\mathcal{K}}\,\bar{e}_l \to 0$ fortement dans W, en d'autres termes, $\widetilde{\mathcal{K}}\,(u-u_l)$ tend vers zéro plus vite que $\|u-u_l\|_H$.

Expliquons à présent comment est construit un opérateur Λ en pratique. Pour cela, on considère la diffraction d'une onde électromagnétique incidente par un objet parfaitement conducteur modélisée par l'équation intégrale EFIE. Cette équation peut être écrite dans le contexte abstrait précédent de la façon suivante : $H = H^{-1/2}(div_\Gamma, \Gamma)$,

$$\mathcal{A}\mathbf{u} = \mathbf{S}_k\mathbf{u} + \frac{1}{k^2}\nabla_\Gamma S_k\, div_\Gamma \mathbf{u} \tag{68}$$

où G_k et $\mathbf{G}_k$ sont respectivement les potentiels simple couche scalaire et vectoriel, k est le nombre d'onde et ∇_Γ est l'opérateur gradient de surface.

L'opérateur T est défini à partir de la décomposition de Helmholtz des champs de vecteurs de H: $\forall \mathbf{v} = \Pi_{loop}\mathbf{v} + \Pi_{star}\mathbf{v} = \mathbf{n}\times\nabla_\Gamma\psi + \nabla_\Gamma\varphi \in H^{-1/2}(div_\Gamma,\Gamma)$, $T\mathbf{v} = \Pi_{loop}\mathbf{v} - \Pi_{star}\mathbf{v}$. Finalement, l'opérateur T-coercif $\mathcal{A}_0$ est défini par (voir **[BH03]**)

$$\mathcal{A}_0\mathbf{v} = \mathbf{S}_0\mathbf{v} + \frac{1}{k^2}\nabla_\Gamma S_0\, div_\Gamma\mathbf{v} - 2\Pi^*_{star}\mathbf{S}_0\mathbf{v}. \tag{69}$$

La construction d'un candidat pour Λ est basée sur des calculs symboliques. Afin d'utiliser cet outil, la surface Γ est supposée $\mathcal{C}^\infty$ pour la phase de construction. De plus, la décomposition de Helmholtz nous permet d'écrire un opérateur M agissant sur les champs de vecteurs tangents à Γ comme une matrice 2×2 d'opérateurs scalaires M_{ij}:

$$M = [\mathbf{n}\times\nabla_\Gamma\ \nabla_\Gamma]\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22}\end{bmatrix}\begin{bmatrix}\Pi_{loop}\\ \Pi_{star}\end{bmatrix}$$

Nous proposons de chercher Λ comme un opérateur pseudodifférentiel tel que son symbole principal vérifie $\sigma_p(\Lambda^*\Lambda\mathcal{A}_0) = \sigma_p(\Lambda)^2\sigma_p(\mathcal{A}_0) = \sigma_p(T)$. Dans ce contexte, nous avons

$$\sigma_p(\mathcal{A}_0) = \begin{bmatrix}\frac{1}{2\|\xi\|} & 0\\ 0 & -\frac{\|\xi\|}{2k^2}\end{bmatrix},\ \sigma_p(T) = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$$

où ξ représente un vecteur cotangent à Γ.

La structure de ces symboles principaux suggère de chercher Λ sous la forme

$$\Lambda = [\mathbf{n}\times\nabla_\Gamma\ \nabla_\Gamma]\begin{bmatrix}\Lambda^+ & 0\\ 0 & \Lambda^-\end{bmatrix} \tag{70}$$

avec $\sigma_p(\Lambda^+) = 1/\sqrt{2\|\xi\|}$ et $\sigma_p(\Lambda^-) = k\sqrt{2/\|\xi\|}$

Finalement, nous construisons les candidats pour $\Lambda^\pm$ en utilisant l'opérateur de Laplace-Beltrami :

$$\Lambda^\pm = \sqrt{2k}\left(Id - \frac{\Delta_\Gamma}{k^2}\right)^{\mp 1/4} \tag{71}$$

En effet, $\sigma_p(\Delta_\Gamma) = -\|\xi\|^2$ et nous avons ajouté l'opérateur identité Id afin d'obtenir un isomorphisme.

Remarques :

1. L'opérateur Λ défini par (70) et (71) est un isomorphisme **[GMMM10]** sur $L^2_t(\Gamma) = \{\mathbf{v}\in \left[L^2(\Gamma)\right]^3 : \mathbf{v}\cdot\mathbf{n} = 0\}$ pour des surfaces lipschitziennes.

2. Par construction, η_Λ est asymptotiquement exacte lorsque la surface est lisse néanmoins nous nous attendons à ce que la constante d'efficacité soit proche de 1 lorsque la surface possédera quelques singularités géométriques comme nous avons pu le constater en acoustique.
3. L'implémentation de η_Λ est basée sur l'utilisation de l'algorithme proposé dans [**HHT08**] pour calculer efficacement les opérateurs (71).

Sur la figure 6, nous observons, d'une part, l'effet bénéfique sur la convergence en présence de singularités du raffinement auto-adaptatif. En particulier, ce dernier restitue le taux de convergence optimal. De plus, on constate que l'estimateur η_Λ est quasiment confondu avec la courbe de référence qui correspond à une erreur calculée à partir d'une solution numérique très précise. Ceci confirme la capacité de η_Λ de restituer à une information quantitative fiable sur l'erreur d'approximation. La figure 7 met en évidence le fait que l'estimateur η_Λ est asymptotiquement exact.

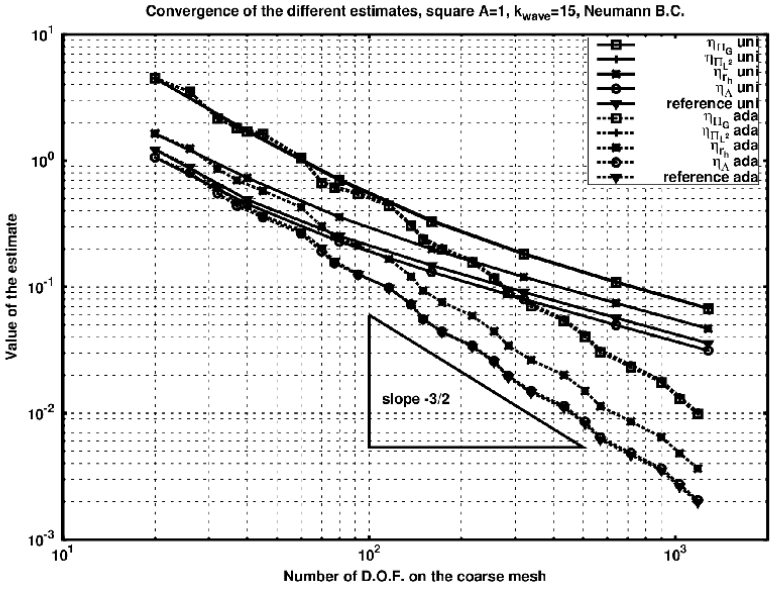


FIGURE 6. Diffraction acoustique par un carré pour un problème de Neumann : convergence de différents estimateurs pour un raffinement uniforme (uni) ou adaptatif (ada).

2.7. Bilan et perspectives

Dans une première partie, nous venons de décrire brièvement une méthode pour construire des formulations intégrales intrinsèquement bien conditionnées. Nous avons étendu avec succès aux problèmes impédants et de transmission les résultats obtenus pour la diffraction par des obstacles parfaitement conducteurs. Rappelons, que grâce à cette technique de régularisation, nous savons qu'il est possible de construire de nouvelles formulations intégrales capables de résoudre des problèmes de plusieurs millions d'inconnues en des temps records, et cela sans préconditionneur.

Fort de ce succès, nous entrevoyons deux perspectives à ce travail. Premièrement, nous voulons appliquer cette approche à des problèmes plus hétérogènes dans le sens mixant diffraction/transmission, obstacles ouverts et fermés et possédant des parties filaires. Deuxièmement, nous pensons que l'approximation de l'opérateur régularisant proposée pour les problèmes de transmission devrait nous fournir un candidat naturel pour construire des conditions de transmission dans une méthode de décomposition de domaine pour des matériaux hétérogènes. Nous allons donc investiguer cette voie.

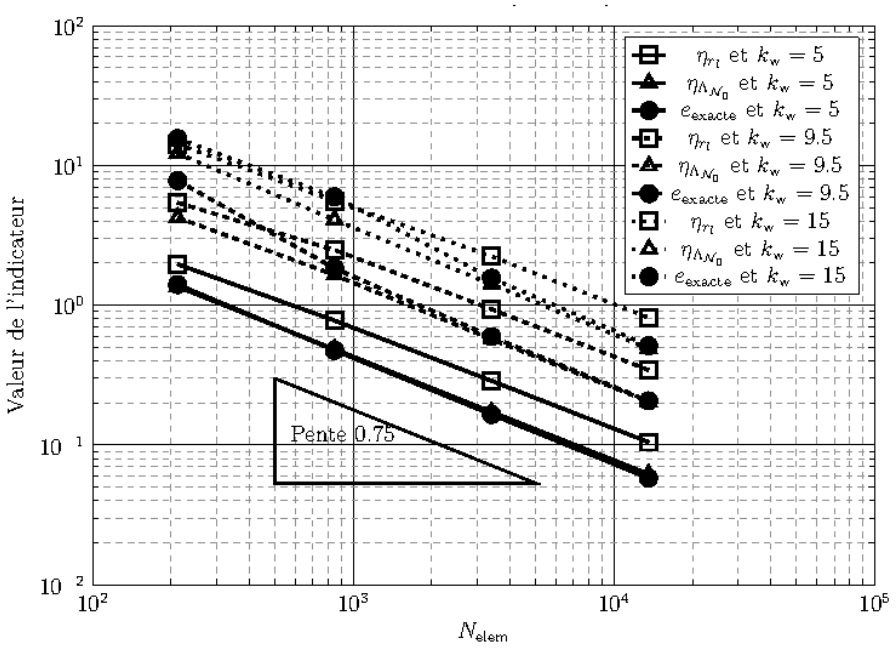


FIGURE 7. Diffraction acoustique par une sphère pour un problème de Neumann et pour différentes fréquences : évolution de l'estimateur η_Λ par rapport à l'erreur exacte pour un raffinement uniforme.

Dans la seconde partie, nous avons évoqué les résultats que nous avons obtenu dans le domaine de l'analyse d'erreur *a posteriori* et le calcul autoadaptatif pour les équations intégrales en acoustique et électromagnétisme. Nous allons poursuivre sur la voie de ces premiers résultats très prometteurs. Rappelons qu'un de nos objectifs est de fournir une aide à l'utilisateur pour réaliser ces simulations en automatisant la création d'un espace d'approximation adapté à son problème en terme de précision tout en optimisant la taille du système à résoudre. En résumé, nous avons pour ambition de concevoir un algorithme auto-adaptatif permettant :

- de garantir la précision d'un calcul par rapport à une prescription de l'utilisateur,
- de construire un maillage "idéal" du point de vue de la simulation,
- de réduire le coût global du calcul en économisant du temps ingénieur lors de la génération du maillage.

Tout d'abord, l'estimateur η_Λ devra être discrétisé, implémenté et éprouvé numériquement dans le cadre de l'électromagnétisme 3D. Néanmoins, on peut facilement constater que sa structure implique un coût non négligeable. Pour contourner ce problème, nous envisageons deux voies. Premièrement, nous avons testé en acoustique une version localisée de l'opérateur Λ qui numériquement semble garder une très bonne constante d'efficacité, il faudrait donc investiguer cette approche. Deuxièmement, nous proposons d'utiliser η_Λ uniquement comme critère d'arrêt de l'algorithme de raffinement autoadaptatif qui lui pourrait être guidé par un estimateur classique moins coûteux. On pourrait penser par exemple à l'estimateur basé sur la localisation standard du résidu pour lequel on a de plus prouvé le guidage quasi-optimale du raffinement.

En parallèle, nous nous intéresserons à la mise en place de structures de données et d'un produit matrice-vecteur rapide adaptés aux raffinements auto-adpatatifs de maillages. Le schéma classique de résolution par équations intégrales est l'utilisation d'un solveur itératif dont le produit matrice-vecteur est accéléré par une méthode FMM (Fast Multipole Method). La mise en place d'un algorithme de raffinement auto-adaptatif nécessite naturellement un certain nombre de modifications dans ce schéma afin d'assurer la performance du code de calcul. En effet, cet algorithme va

produire localement des zones raffinées, voire très raffinées (i.e. un maillage multi-échelle) qui sont généralement néfastes pour les performances des solveurs basés sur une méthode FMM standard. Il est donc impératif de posséder une méthode d'accélération efficace pour ce type de configurations. Des solutions dans la littérature (FMM bases fréquences **[CF15]**, H-matrix **[Hac15]**) seront le point de départ pour concevoir un produit matrice-vecteur efficace pour ce type de maillage. D'autre part, à chaque itération, l'algorithme de raffinement produira uniquement des modifications locales du maillage donc de la matrice à inverser. Il sera donc primordial de mettre en place des structures évitant la reconstruction globale à chaque raffinement de toutes les données nécessaires pour effectuer les produits matrice-vecteur (voir **[Djo06]**).

De plus la technique basée sur l'estimateur η_Λ nous offre aussi une approche naturelle pour déterminer une estimation de l'erreur algébrique induite par l'utilisation d'un solveur itératif. Cette estimation sera de même nature que l'erreur de discrétisation (norme sous-jacente à la méthode d'approximation) et devrait ainsi permettre de proposer des critères d'arrêt efficaces. Plus précisément, les estimateurs d'erreur a posteriori que l'on a évoqué précédemment sont utilisés pour répondre à la question : quelle est la distance entre la solution approchée u_h et la solution exacte u de notre problème ? Leur construction est basée sur l'hypothèse que le problème discret a été résolu de façon exacte. Or en pratique, on utilise souvent un solveur itératif pour inverser le système linéaire, et donc, pour pouvoir utiliser les indicateurs d'erreur dérivés pour u_h , il faut utiliser un critère d'arrêt très petit afin de s'assurer que la solution calculée $\tilde{u}_h$ soit proche de u_h . Ceci peut induire une augmentation substantielle du coût de calcul sans aucune amélioration supplémentaire de l'erreur globale. Notre objectif est de donner une estimation précise de l'erreur $u - \tilde{u}_h$ par le biais d'un estimateur d'erreur a posteriori de la forme $\mu_d + \mu_a$ où μ_d et μ_a estiment respectivement l'erreur de discrétisation $u - u_h$ et l'erreur algébrique $u_h - \tilde{u}_h$. On peut alors choisir comme critère d'arrêt pour le solveur itératif une condition du type $\mu_a \leq \beta\mu_d$ avec $0 < \beta < 1$ (voir **[JSV10]**). Autrement dit, on utilise un équilibrage entre l'erreur de discrétisation et l'erreur algébrique pour déterminer un critère d'arrêt. Le but est d'éviter de faire un grand nombre d'itérations à partir du moment où l'erreur algébrique passe en dessous de l'erreur de discrétisation et ainsi de faire une économie importante en terme de coût calcul.

Enfin, un dernier axe s'intéressera à la preuve théorique de la convergence quasi-optimale de l'algorithme de raffinement auto-adaptatif c'est à dire que nous tenterons de démontrer que ce dernier produit la meilleure suite de maillages possibles pour un problème donné. Nous avons déjà prouvé ce type de résultat dans **[Bak16]** pour un algorithme guidé par une localisation standard du résidu et pour une résolution directe du système linéaire. Ici, on s'attachera à étendre le résultat aux algorithmes guidés par les indicateurs a posteriori construits à partir de la nouvelle technique de localisation proposée dans **[BPC78]**[**Bak16**] et en considérant une résolution itérative du système linéaire contrôlée comme prévu dans le point précédent. Ce résultat non trivial assurera la robustesse du nouvel algorithme.

Ces travaux ont donné lieu à six publications dans des journaux : Computers and Mathematics with Applications, J. Integral Equations Applications, Communication In Computational Physics, Mathematical Modelling and Numerical Analysis, In Mathematics in industry-scientific computing in electical engineering, Progress In ElectRomagnetism. Ils ont été réalisés en collaborant avec M. Bakry (doctorant), A. Bendali, F. Collino, M.B Fares, D. Levadoux, F. Millot ainsi que des étudiants en stage niveau master T. Vanolmen, G. Allai, J. Lee, A. Dececco et E. Queirolo. Enfin, ils ont été en grande partie réalisés dans le cadre des projets et contrats suivants : ANR ARTHEMIS, ANR RAFFINE, contrat CERFACS-Dassault-Aviation, contrat CERFACS-EADS/MBDA, contrat CERFACS-CNES et enfin contrat CERFACS-CEA.

CHAPITRE 3

PROPAGATION DU SON DANS UN ÉCOULEMENT COMPLEXE EN RÉGIME HARMONIQUE

Dans ce dernier chapitre, je présente un schéma numérique basé sur l'équation de Galbrun qui permet de tenir compte du couplage entre les phénomènes acoustiques et hydrodynamiques lors du calcul de la propagation du son dans un écoulement complexe en régime harmonique. Il correspond à la discrétisation de deux équations couplées de nature différente : d'un côté, l'équation de Galbrun et de l'autre une équation de transport harmonique d'ordre deux. La première équation étant traitée par une approche éléments finis et la seconde par une approximation de type Galerkin discontinu. Des éléments d'analyse du modèle et du schéma sont aussi donnés ainsi qu'un certain nombre de résultats numériques. Enfin, la dernière partie de ce chapitre présente un bilan de ces travaux et propose quelques perspectives de travail.

3.1. Introduction et problématique

La réduction des nuisances sonores (trafic aérien ou routier) est un enjeu sociétal crucial qui impose aux gouvernements de renforcer les normes de certification acoustique. Les industriels sont donc obligés d'investir des moyens importants pour réduire les nuisances sonores.

L'acoustique s'intéresse à ces phénomènes physiques et a pour objectifs principaux l'identification des sources de bruit, l'étude des mécanismes de la propagation du son ainsi que l'évaluation de l'impact sur le milieu environnant. Les modèles de propagation acoustique usuels sont obtenus par linéarisation des équations de la mécanique des fluides et considèrent l'écoulement comme une donnée du problème. La complexité des systèmes d'équations obtenus dépend des hypothèses faites sur l'écoulement. En particulier, la prise en compte d'écoulements généraux complexifie les modèles car elle s'accompagne de couplages qui ont lieu entre l'acoustique et les phénomènes hydrodynamiques [**Ast09**] et qui se traduisent par des systèmes d'équations vectorielles. Les équations d'Euler linéarisées forment, dans ce cadre, le système de référence. Cependant il existe d'autres approches basées par exemple sur les équations de Navier-Stokes linéarisées, ou encore sur l'équation de Galbrun [**Poi85**] qui permet aussi de prendre en compte les effets de l'écoulement moyen sur la propagation acoustique.

Nous nous sommes intéressés plus particulièrement aux nuisances sonores produites par les turboréacteurs d'un avion. Les mesures montrent que ces dernières sont causées par un bruit de raies (ou bruit tonal) qui se caractérise par des spectres ayant des pics d'amplitude à des fréquences isolées correspondant aux harmoniques de la fréquence de passage des aubes. La détermination du rayonnement acoustique pour ces fréquences spécifiques représente de toute évidence un enjeu pour la réduction des nuisances sonores.

Pour la propagation en régime fréquentiel, des approches basées sur la résolution du système d'Euler linéarisé ont été envisagées mais les références disponibles sont beaucoup plus rares qu'en régime temporel. Par exemple [**OL08**] adapte au cas fréquentiel des schémas aux différences finies de type DRP (Dispersion Relation Preserving). Sur maillage non structuré, l'approximation de

Galerkin classique n'étant pas stable [**ANE81**], il convient d'utiliser des méthodes stabilisées de type SUPG (Streamline Upwind Petrov Galerkin) ou Galerkin discontinu [**IAS10, RM06, Gab07**]. Notons cependant que les équations résolues dans ces travaux ont une forme simplifiée : des termes de gradient de l'écoulement ont été négligés afin de limiter les coûts de calcul. Mais surtout, le traitement de domaines infinis est toujours sujet à des difficultés car la présence d'un écoulement non uniforme (et notamment l'existence de modes de vorticité et d'entropie) dissymétrise le problème : les conditions doivent savoir gérer différemment les bords amont et aval, ce qui rend l'écriture de conditions de rayonnement complexe.

Notons aussi que la plupart des outils numériques commerciaux traitant le régime harmonique, se placent dans des cas d'écoulements simples. Ils supposent que la vitesse de l'écoulement porteur est à rotationnel nul (on parle d'écoulement irrotationnel ou potentiel). On peut alors montrer que les équations d'Euler linéarisées se simplifient en une unique équation scalaire portant sur le potentiel de vitesse. Il s'agit de l'équation de Helmholtz convectée. Un avantage important concerne les conditions de rayonnement qui peuvent être déterminées de façon explicite si l'écoulement est uniforme loin de la source. Dans ce cas, l'équation se ramène par transformation de Lorentz à l'équation de Helmholtz standard et un couplage avec une méthode intégrale permet alors de faire rayonner la solution à l'infini [**Dup05**]. L'analyse de cette approche est bien maîtrisée [**CK98**] mais son domaine de validité est très restreint. En particulier, elle ne s'applique qu'aux situations où les phénomènes hydrodynamiques sont absents.

Le modèle de Galbrun [**Gal31**] apparaît comme une alternative intéressante au système d'Euler linéarisé et comme une approche améliorée par rapport à celle de Helmholtz convectée.

Soit un fluide en écoulement stationnaire subsonique arbitraire dans un domaine infini qui est supposé parfait et en évolution adiabatique. Ce fluide est caractérisé par ses champs non uniformes de pression p_0, vitesse $\underline{v}_0$, densité ρ_0 et vitesse du son c_0 qui vérifient les équations d'Euler stationnaires et l'équation d'état suivantes :

$$div(\rho_0\,\underline{v}_0) = 0, \tag{72a}$$

$$\rho_0\,(\underline{v}_0\cdot\nabla)\;\underline{v}_0 + \nabla\,p_0 = 0, \tag{72b}$$

$$p_0 = K\,\rho_0^{\gamma} \text{ et } c_0^2 = \gamma\,\frac{p_0}{\rho_0}, \tag{72c}$$

où K et γ sont des constantes.

L'équation de Galbrun est alors obtenue par linéarisation des équations de conservation de la mécanique des fluides [**Poi85**] et porte sur la perturbation du déplacement lagrangien $\underline{\xi}(x,t)$. En régime harmonique, nous définissons $\underline{u}$ tel que $\underline{\xi}(x,t) := \Re\left(\underline{u}(x)\,e^{-i\omega\,t}\right)$ et l'équation de Galbrun s'écrit alors

$$\rho_0\frac{D^2\,\underline{u}}{D\,t^2} - \nabla\left(\rho_0 c_0^2\,div(\underline{u})\right) + div(\underline{u})\,\nabla\,p_0 - \nabla\underline{u}^T\cdot\nabla\,p_0 = 0 \tag{73}$$

où $D/D\,t$ désigne l'opérateur de transport qui en régime harmonique s'écrit :

$$\frac{D}{D\,t} = -i\omega + \underline{v}_0\cdot\nabla(\cdot) \tag{74}$$

Il est possible de vérifier [**Leg03**] qu'à partir de la solution de l'équation de Galbrun nous retrouvons la solution des équations d'Euler linéarisées en utilisant le lien

$$p = -\rho_0\,c_0^2\,div(\underline{u}) - \underline{u}\cdot\nabla\,p_0 \tag{75a}$$

$$\underline{v} = \frac{D\,\underline{u}}{D\,t} - (\underline{u}\cdot\nabla)\,\underline{v}_0 \tag{75b}$$

L'équation de Galbrun est donc une équation vectorielle qui porte uniquement sur la perturbation du déplacement qui est une inconnue lagrangienne que l'on observe selon un point de vue eulérien (c'est-à-dire à un instant donné et en un point de l'écoulement porteur non perturbé), on parle de représentation mixte Euler-Lagrange. Cette équation aux dérivées partielles du second ordre présente des similarités avec les équations des ondes traitées en électromagnétisme et se prête bien à l'écriture d'une formulation variationnelle. Cependant, il est maintenant connu qu'en régime fréquentiel, une discrétisation par éléments finis de Lagrange est instable [**PE01**]. Deux stratégies de résolution alternatives ont été proposées.

Une première alternative exposée dans [**TGT03**] consiste en une formulation mixte en pression et déplacement approchée par une méthode d'éléments finis. En absence d'écoulement, une condition inf-sup est démontrée et assure ainsi la convergence de la méthode mais ce résultat ne peut pas être étendu dans le cas d'un écoulement non nul. Une étude de stabilité et de dispersion du schéma a cependant été faite dans [**GAT05**] et ce modèle a par exemple été exploité pour des applications dans des guides à parois traitées avec écoulement cisaillé [**NTPD11**]. En domaine infini par contre, cette formulation rencontre des difficultés pour la mise en place de condition de radiation pour borner le domaine de calcul. En particulier, l'usage de couches PML (perfectly matched layer) conduit à des solutions erronées même pour un écoulement uniforme [**BBL06**].

Une autre manière de résoudre l'équation de Galbrun proposée dans [**BLL01**] se base sur une méthode de régularisation inspirée de l'électromagnétisme. Elle consiste à ajouter à l'équation un terme qui ne change pas la valeur de la solution mais qui rend possible l'utilisation d'éléments finis nodaux pour la détermination du déplacement. Cette nouvelle écriture dite formulation augmentée fait intervenir une inconnue supplémentaire : le rotationnel du déplacement lagrangien appelé, par abus de langage, inconnue hydrodynamique ou vorticité. La formulation de Galbrun augmentée a fait l'objet de deux thèses successives dans des cas d'écoulement simplifiés pour lesquels l'inconnue vorticité peut être déterminée soit de façon exacte, soit par un calcul préalable. La première thèse [**Leg03**] pose le cadre théorique permettant de prouver la stabilité de l'approximation par éléments finis nodaux et montre un résultat d'existence et d'unicité dans le cas d'un écoulement uniforme dans un guide. Le domaine de calcul est borné par des couches PML et la convergence exponentielle de la méthode, en fonction de l'épaisseur des couches, est démontrée. La deuxième thèse [**Duc07**] étend cette étude au cas d'écoulement parallèle cisaillé dans un guide d'onde. Une formule de convolution le long des lignes de courant relie alors l'inconnue vorticité au déplacement [**BDLM07, BDM07**]. Cette formule a l'inconvénient de devenir fortement oscillante lorsque l'écoulement est lent. Dans ce cas, un modèle approché a aussi été proposé, qui consiste à remplacer l'expression exacte de la vorticité par une formule approchée locale : il s'agit de l'approximation dite faible Mach [**BMMP10**].

Notre objectif a été d'étendre ces travaux aux cas d'écoulements plus généraux, pour lesquels il n'est plus possible d'obtenir l'inconnue hydrodynamique par une formule simple. Ce travail a été réalisé essentiellement en 2D néanmoins la formulation en 3D ne pose pas de difficulté. Nous discuterons par contre de la mise place du schéma numérique dans les perspectives.

3.2. Formulation augmentée

Commençons par le cas simple $\underline{v}_0 = \underline{0}$. Dans ce cas l'équation de Galbrun s'écrit :

$$-\omega^2 \underline{u} - c_0^2 \nabla \left(div(\underline{u}) \right) = 0. \tag{76}$$

En prenant le rotationnel de cette équation, il vient que les solutions $\underline{u}$ vérifient

$$rot(\underline{u}) = 0 \tag{77}$$

Le procédé d'augmentation consiste à remplacer le terme en $\nabla\, div$ par un laplacien en ajoutant formellement à l'équation (76) le terme

$$c_0^2 rot\, rot\, (\underline{u}) = 0 \tag{78}$$

qui s'identifie à zéro.

Remarque : Si on est en 2D, alors dans (78), le premier rotationnel est vectoriel et le second scalaire.

Ainsi, par l'identité

$$rot(rot) - \nabla\, div = -\Delta \tag{79}$$

le champ $\underline{u}$ solution de (76) vérifie une équation elliptique (l'équation de Helmholtz vectorielle) de la forme

$$-\omega^2 \underline{u} - c_0^2 \Delta\, \underline{u} = 0. \tag{80}$$

On peut donc envisager une résolution stable de l'equation (80) par des éléments finis nodaux (*i.e* conforme H^1).

Remarque : L'équation (76) peut évidemment être discrétisée directement par des éléments finis conformes $H - div$ (par ceux de Raviart-Thomas par exemple). L'exemple choisi est seulement utilisé pour expliquer simplement la technique d'augmentation. De plus, dès que le fluide porteur n'est plus au repos, l'opérateur n'est plus stabe dans $H(div)$ et les élements de Raviart-Thomas ne sont donc plus une bonne solution. En effet, si on considère par exemple le cas d'un écoulement uniforme $\underline{v}_0 = v_0 \underline{e}_1$ où $\underline{e}_1$ est le premier vecteur de la base canonique de $\mathbb{R}^d$, on peut montrer que l'on a toujours $rot(\underline{u}) = 0$ et la formulation variationnelle associée à l'équation de Galbrun augmentée est de la forme :

$$\int \left(-\omega^2 \underline{u} \cdot \overline{\underline{v}} - 2i\omega v_0 \frac{\partial \underline{u}}{\partial x_1} \cdot \overline{\underline{v}} - v_0^2 \frac{\partial \underline{u}}{\partial x_1} \cdot \frac{\partial \overline{\underline{v}}}{\partial x_1} + c_0^2 \left(div\underline{u}\, div\overline{\underline{v}} + rot\underline{u}\, rot\overline{\underline{v}} \right) \right) \mathrm{d}x = 0 \tag{81}$$

où $\underline{v}$ est une fonction-test à support compact. Cette nouvelle formulation du problème initial nous permet d'écrire : si Ω est un ouvert borné et si on suppose que l'écoulement porteur est subsonique *i.e* $|v_0|/c_0 < 1$, alors la forme bilinéaire

$$b(\underline{u}\,,\, \underline{v}) = \int_\Omega \left(\underline{u} \cdot \overline{\underline{v}} - v_0^2 \frac{\partial \underline{u}}{\partial x_1} \cdot \frac{\partial \overline{\underline{v}}}{\partial x_1} + c_0^2 \left(div\underline{u}\, div\overline{\underline{v}} + rot\underline{u}\, rot\overline{\underline{v}} \right) \right) \mathrm{d}x \tag{82}$$

est coercive sur l'espace $V := \{\underline{v} \in H^1(\Omega)^d \,:\, \underline{v} \cdot \underline{n} = 0 \text{ sur } \Gamma\}$ où $\underline{n}$ est la normale unitaire extérieure à Ω. Ceci permet de montrer que le problème augmenté avec la condition aux limites de bords rigides $\underline{u} \cdot \underline{n} = 0$ relève de l'alternative de Fredholm car les termes de (81) autres que la forme bilinéaire $b(\underline{u}, \underline{v})$ constituent une perturbation compacte de cette dernière. De plus, on peut constater que si on enlève le terme $rot\underline{u} \cdot rot\overline{\underline{v}}$ qui provient de la technique d'augmentation, alors le terme positif restant $|div\underline{u}|^2$ est incapable de contrôler $-\left|\dfrac{\partial \underline{u}}{\partial x_1}\right|^2$ qui provient de la dérivée convective d'ordre deux.

Dans le cas d'un fluide quelconque, on n'obtient pas une relation aussi simple que (77). En effet, si on introduit la variable $\underline{\psi} = rot(\underline{u})$, on a montré dans **[BMMP10]** que la formulation augmentée prend la forme suivante :

$$\rho_0 \frac{D^2\, \underline{u}}{D\, t^2} - \nabla \left(\rho_0 c_0^2\, div(\underline{u}) \right) + div(\underline{u})\, \nabla\, p_0 - \nabla \underline{u}^T \cdot \nabla\, p_0 + rot(\rho_0 c_0^2 (rot(\underline{u}) - \underline{\psi}) = \underline{f} \tag{83}$$

où dans le cas 2D, on a

$$\frac{D^2\, \underline{\psi}}{D\, t^2} = -2 \frac{D}{D\, t} \left(\mathcal{B} \underline{u} \right) - \mathcal{C} \underline{u} + \frac{1}{\rho_0} rot(\underline{f}) \tag{84}$$

avec

$$\mathcal{B}\underline{u} = \sum_{j=1}^{2} \nabla\, v_{0,j} \wedge \frac{\partial \underline{u}}{\partial x_j} \tag{85}$$

et

$$\begin{aligned} \mathcal{C}\underline{u} = & \sum_{j,k=1}^{2} \left(\frac{\partial\, v_{0,k}}{\partial x_j} \nabla\, v_{0,j} \wedge \frac{\partial \underline{u}}{\partial x_k} - v_{0,j} \nabla \left(\frac{\partial\, v_{0,k}}{\partial x_j} \right) \wedge \frac{\partial \underline{u}}{\partial x_k} \right) \\ & + \frac{1}{\rho_0} \sum_{j=1}^{2} \left(\frac{1}{\rho_0 c_0^2} \frac{\partial p_0}{\partial x_j} \nabla p_0 - \nabla \left(\frac{\partial p_0}{\partial x_j} \right) \right) \wedge \nabla u_j \end{aligned} \tag{86}$$

Remarque : Les opérateurs $\mathcal{B}$ et $\mathcal{C}$ font intervenir des gradients de l'écoulement et prennent des formes très simplifiées en écoulement parallèle cisaillé de la forme $\underline{v}_0 = v_0(x_2)\underline{e}_1$. Nous avons

$$\mathcal{B}\underline{u} = -\frac{\partial v_0}{\partial x_2} \frac{\partial u_1}{\partial x_1}, \tag{87a}$$

$$\mathcal{C}\underline{u} = 0. \tag{87b}$$

La technique d'augmentation conduit alors à un système couplé de deux équations (83-84) à deux inconnues $(\underline{u}\,,\, \underline{\psi})$. La difficulté pour résoudre ce système vient du fait que les deux équations sont de natures très différentes : une équation des ondes d'une part et une équation de transport d'autre part.

3.3. L'équation de transport en régime harmonique

Cette première section est dédiée à l'étude du probléme de transport en régime harmonique (qui est la brique de base pour la résolution de l'équation hydrodynamique). Notre objectif a été de trouver une méthode numérique stable pour résoudre ce problème sur maillage non structuré. Il s'agit d'une équation aux dérivées partielles d'ordre un à laquelle une condition de Dirichlet est classiquement adjointe et qui décrit le transport d'une quantité physique par un écoulement porteur donné. L'originalité de notre travail est que nous considérons cette équation en régime fréquentiel et non en temps. Il est alors naturel de s'inspirer des nombreuses références bibliographiques qui traitent de l'équation très similaire d'advection-réaction. Mais le cas que nous avons traité est plus délicat. En effet, d'une part, les difficultés connues pour l'équation d'advection-réaction sont amplifiées dans le cas de l'équation de transport en régime harmonique, car cette dernière n'est pas dissipative. D'autre part, du point de vue théorique, l'estimation fondamentale de stabilité est plus difficile à établir et requiert des conditions particulières sur l'écoulement (comme l'absence de zones de recirculation par exemple). Enfin, son étude est un préalable à l'étude de tous les modèles d'aéroacoustique prenant en compte les phénomènes hydrodynamiques. Nous décrivons à présent les résulats obtenus pour cette équation. Nous renvoyons à la thèse d'Emilie Peynaud **[Pey13]**[**BMPP11**] pour plus de détails et pour les preuves.

3.3.1. Position du problème. — Soit Ω un sous-domaine borné de $\mathbb{R}^d$ $(d \geq 2)$ de frontière $\partial\Omega$ Lipschitzienne et de normale sortante $\underline{n}$. Soit un champ de vecteurs $\underline{v}_0 : \Omega \to \mathbb{R}^d$ satisfaisant les hypothèses suivantes :

1. $\underline{v}_0 \in C^1(\Omega)$,
2. $div\underline{v}_0 = 0$ dans Ω.

Remarques : Cette dernière hypothèse nous permet de simplifier les estimations dans les démonstrations en évitant des termes en $div\underline{v}_0$. Néanmoins, un simple changement d'inconnue permet de se ramener à l'hypothèse physique d'incompressibilité du fluide traduite par l'équation $div(\rho_0\underline{v}_0) = 0$.

Nous définissons de plus les frontières à flux entrant et à flux sortant respectivement par :

$$\Gamma^- := \{x \in \partial\Omega \,:\, \underline{v}_0 \cdot \underline{n} < 0\}, \tag{88a}$$

$$\Gamma^+ := \{x \in \partial\Omega \,:\, \underline{v}_0 \cdot \underline{n} > 0\}, \tag{88b}$$

et les bords glissants par

$$\Gamma_0 := \{x \in \partial\Omega \,:\, \underline{v}_0 \cdot \underline{n} = 0\}. \tag{89}$$

Soient $f \in L^2(\Omega)$ et $g \in L^2(\Gamma^-)$. Enfin, notons α un paramètre complexe non nul tel que $\alpha = a + i\omega$ avec a et ω réels. Nous considérons le problème suivant : trouver $\psi : \Omega \to \mathbb{C}$ solution de

$$\mathcal{P}_\alpha \,:\, \left\{ \begin{array}{rcl} \alpha\psi + \underline{v}_0 \cdot \nabla\psi & = & f \text{ dans } \Omega \\ \psi & = & g \text{ sur } \Gamma^- \end{array} \right. \tag{90}$$

Considérons à présent un exemple pour illustrer la spécificité du transport harmonique. Pour cela, on considère un écoulement bi-dimensionnel dirigé selon le premier vecteur $\underline{e}_1$ de la base canonique de $\mathbb{R}^2$ défini par $\underline{v}_0(x_1, x_2) = v_0(x_2)\underline{e}_1$ (voir fig. 1). L'advection-réaction correspond à la dissipation d'une quantité selon l'écoulement porteur pour $a > 0$. La figure 1 montre clairement ce comportement. Au contraire, pour le transport harmonique (voir fig. 2), il n'y a pas de phénomène dissipatif. En fait, l'inconnue ψ est transportée sans perte dans le sens de l'écoulement. De plus, la solution a un comportement oscillatoire induit par le régime harmonique dont la longueur d'onde est $2\pi v_0(x_2)/\omega$ (soit la longueur parcourue par une particule convectée par l'écoulement pendant une période). Comme le montre la figure 2, plus la vitesse du fluide v_0 est faible, plus les oscillations sont rapides.

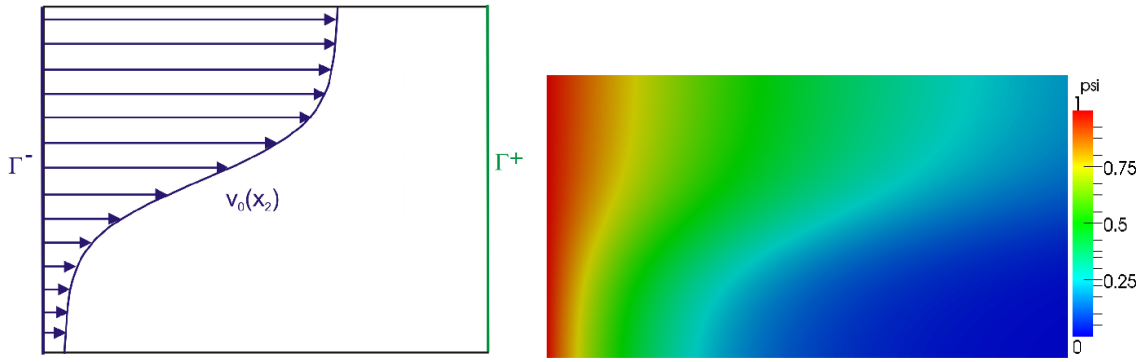


FIGURE 1. Schéma du domaine avec écoulement porteur cisaillé (gauche) et solution pour $f = 0$, $g = 1$ et $\alpha = 0.5$ (droite).

3.3.2. Analyse du probème continu. — Nous supposons à présent que $f \in L^2(\Omega)$. L'approche naturelle est de chercher la solution de $\mathcal{P}_\alpha$ dans l'espace du graphe de l'opérateur c'est-à-dire

$$H(\Omega, \underline{v}_0) := \left\{\varphi \in L^2(\Omega) \,:\, \underline{v}_0 \cdot \nabla\varphi \in L^2(\Omega)\right\} \tag{91}$$

qui est un espace de Hilbert pour le produit scalaire sous-jacent à la norme du graphe :

$$\|\varphi\|_H^2 := |\alpha|^2 \, \|\varphi\|_{0,\Omega}^2 + \|\underline{v}_0 \cdot \nabla\varphi\|_{0,\Omega}^2 \tag{92}$$

Pour définir la trace des fonctions de $H(\Omega, \underline{v}_0)$ et ainsi pouvoir considérer des conditions aux limites de type Dirichlet, il nous faut supposer que :

1. L'espace $\mathcal{C}_0^1(\overline{\Omega})$ (*i.e* fonctions de classe $\mathcal{C}^1$ à support compact dans $\overline{\Omega}$) est dense dans $H(\Omega, \underline{v}_0)$,
2. Les frontières de Ω à flux entrant et sortant sont bien séparées *i.e* $dist(\Gamma^-, \Gamma^+) > 0$.

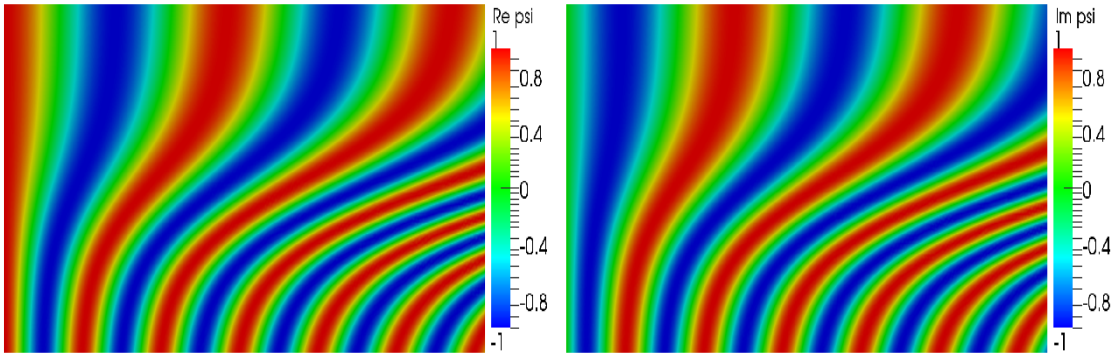


FIGURE 2. Solution pour $f = 0$, $g = 1$ et $\alpha = 4 * i$: partie réelle (gauche) et imaginaire (droite).

Remarque : La dernière hypothèse est nécessaire pour assurer l'existence de la trace (voir **[Pey13]** pour des contre-exemples).

Nous pouvons énoncer le résulat suivant (voir **[EG04]**) :

Lemme 1. — *L'opérateur trace* $\gamma^{\pm} : \varphi \mapsto \varphi_{|\Gamma^{\pm}}$ *est continu et surjectif de* $H(\Omega, \underline{v}_0)$ *dans*

$$L^2(\Gamma^{\pm}, |\underline{v}_0 \cdot \underline{n}|) := \{\varphi \in L^2(\Gamma^{\pm}) \ : \ \int_{\Gamma^{\pm}} |\varphi|^2 \, ||\underline{v}_0 \cdot \underline{n}| \, \mathrm{d}\sigma < +\infty\}$$

et nous avons la formule d'intégration par parties : $\forall \varphi, \, \psi \in H(\Omega, \underline{v}_0,$

$$\int_{\omega} (\underline{v}_0 \cdot \nabla \varphi) \, \overline{\psi} \, \mathrm{d}x = - \int_{\Omega} \varphi \, (\underline{v}_0 \cdot \nabla \overline{\psi}) \, \mathrm{d}x + \int_{\Gamma^{+}} (\underline{v}_0 \cdot \underline{n}) \varphi \, \overline{\psi} \tag{93}$$

D'après ce lemme, l'espace $H(\Omega, \underline{v}_0, \Gamma^{\pm}) := \{\varphi \in H(\Omega, \underline{v}_0) \ : \ \varphi = 0 \text{ sur } \Gamma^{\pm}\}$ est un espace de Hilbert. De plus, la surjectivité de l'application trace nous permet de restreindre l'analyse au problème $\mathcal{P}_{\alpha}$ avec une condition aux limites homogène c'est-à-dire trouver $\psi \in H(\Omega, \underline{v}_0, \Gamma^{-})$ telle que

$$\alpha \psi + \underline{v}_0 \cdot \nabla \psi = f \text{ dans } \Omega, \tag{94}$$

que l'on peut encore écrire sous la forme faible

$$p(\alpha; \psi, \varphi) := \int_{\Omega} (\alpha \psi + \underline{v}_0 \cdot \nabla \psi) \ \varphi \, \mathrm{d}x = \int_{\Omega} f \, \varphi \, \mathrm{d}x, \ \forall \varphi \in L^2(\Omega). \tag{95}$$

En utilisant une technique standard pour étudier l'équation d'advection-réaction **[EG04]**, on peut "facilement" montrer que le problème (95) est bien-posé dès que $\Re \alpha > 0$ et en particulier, on a la condition de inf-sup suivante :

$$\inf_{\psi \in H(\Omega, \underline{v}_0, \Gamma^{-})} \sup_{\varphi \in L^2(\Omega)} \frac{\Re p(\alpha; \psi, \varphi)}{\|\psi\|_H \, \|\varphi\|_{0,\Omega}} \geq \frac{\Re \alpha}{|\alpha|} > 0 \tag{96}$$

Les résultats que l'on peut donc obtenir en utilisant une approche classique de la littérature (voir **[EG04]** par exemple) ne nous permettent pas de traiter le cas où la partie réelle du coefficient de réaction est nulle. A ce stade, nous pourrions penser que la difficulté est simplement technique et que nous ne savons pas choisir correctement la fonction test permettant de dériver une condition inf-sup. Cependant, dans le thèse de E. Peynaud **[Pey13]** nous avons exhibé des configurations pour lesquelles le problème est mal posé. Il semble donc bien que les difficultés liées à l'obtention d'un résultat de stabilité ne soient pas seulement d'ordre technique. Pour assurer le caractère bien-posé du problème quel que soit α dans $\mathbf{C}$, nous introduisons une hypothèse supplémentaire sur

l'écoulement porteur : la notion d'écoulement $\Omega-$remplissant. Pour cela, il nous faut introduire le problème aux caractéristiques suivant : Trouver Φ solution de l'équation différentielle ordinaire,

$$\left\{ \begin{array}{rcl} \dfrac{\partial \Phi}{\partial s}(s,b) & = & \underline{v}_0(\Phi(s,b)) \\ \Phi(0,b) & = & b \end{array} \right. \tag{97}$$

où $(s,b) \in \widetilde{\Omega} := \{(s,b) \,:\, b \in \Gamma^-,\, s \in I\!R^+\}$.

Les hypothèses sur l'écoulement porteur nous permettent de conclure à l'existence d'une solution maximale unique pour le problème (97) en utilisant le théorème de Cauchy-Lipschitz. De plus, Φ définit une $\mathcal{C}^1$-difféomorphisme de $\widetilde{\Omega}$ dans Ω.

Nous pouvons maintenant énoncer la définition d'écoulement $\Omega-$remplissant proposée dans [**Aze96**].

Définition 1. — *Un écoulement est dit $\Omega-$remplissant s'il existe $t^* > 0$ tel que pour tout $x \in \Omega$, il existe $s < t^*$ et $b \in \Gamma^-$ tels que $x = \Phi(s,b)$ où $\Phi(\cdot, b)$ est l'unique solution de (97).*

Remarque : La définition précédente veut dire que les caractéristiques du champ de vitesse issues du bord entrant atteignent le bord à flux sortant en un temps fini borné, sauf éventuellement en un ensemble de mesure nulle.

On peut définir aussi

Définition 2. — *(temps de Ω-parcours) Soit $\underline{v}_0$ est un écoulement Ω-remplissant et $b \in \Gamma^-$. Le temps de Ω-parcours τ_b est le temps que met la ligne de courant issue de b pour sortir du domaine Ω (ou atteindre Γ^+).*

Si $\underline{v}_0$ est un écoulement Ω-remplissant alors il existe $\tau < +\infty$ tel que $\sup\limits_{b \in \Gamma^-} \tau_b \leq \tau$.

Nous énonçons à présent le résultat principal de cette section :

Proposition 2. — *Si $\underline{v}_0 \in C^1(\Omega)$ est un écoulement Ω-remplissant vérifiant $div \underline{v}_0 = 0$ dans Ω, alors le problème $\mathcal{P}_\alpha$ est bien posé dans $H(\Omega, \underline{v}_0, \Gamma^-)$ pour tout $\alpha \in \mathbf{C}$. En particulier,*

$$\|\psi\|^2_{0,\Omega} \leq C \|f\|^2_{0,\Omega} \tag{98}$$

où

$$C = \left\{ \begin{array}{l} \dfrac{\tau^2}{2} \ si\ a = 0 \\ \dfrac{1}{2a}\left(\tau + \dfrac{e^{-2a\tau} - 1}{2a}\right) \ si\ a \neq 0 \end{array} \right. \tag{99}$$

avec $\tau = \sup\limits_{b \in \Gamma^-} \tau_b$ et $a = \Re\alpha$.

3.3.3. Approximation. — Nous avons envisagé plusieurs approches pour discrétiser le problème (94). Premièrement, la méthode de Galerkin continue qui consiste à discrétiser directement (95) en utilisant des éléments finis de classe $\mathcal{C}_0$ de type Lagrange. Il est connu que de ce type d'approche n'est pas appropriée pour les équations aux dérivées partielles d'ordre un. En effet, la condition inf-sup discrète associée n'est pas uniforme en h et conduit à un schéma sous-optimal. Ceci se traduit numériquement par l'apparition de signaux parasites dans la solution approchée (voir fig. 3).

Une parade couramment employée consiste à stabiliser la formulation de Galerkin en introduisant de la dissipation dans le schéma. Ces techniques de stabilisation correspondent à des méthodes de Galerkin-Petrov et nous avons en particulier testé des formulations SUPG (Streamline Upwind Petrov Galerkin) et moindres carrés (LS pour least-square) pour lesquelles le caractère elliptique de la formulation est restauré en ajoutant à la fonction test un terme de dérivée convective. Les résultats d'analyse montrent en effet des améliorations au niveau de la stabilité des schémas numériques.

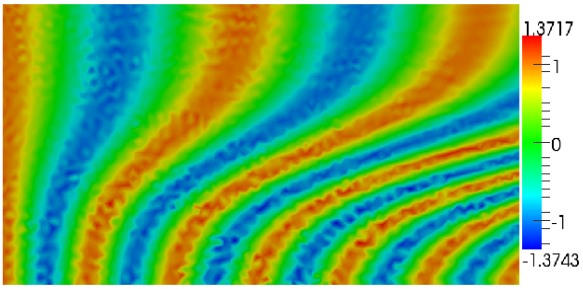


FIGURE 3. Résolution de l'équation de transport harmonique par une approche éléments finis : solution numérique parasitée

Néanmoins, l'utilisation de ce type d'approche est rédhibitoire pour le transport harmonique (voir fig. 4). En effet, la dissipation induite par ces schémas nécessite généralement l'utilisation de maillages fins générant un coût de calcul prohibitif.

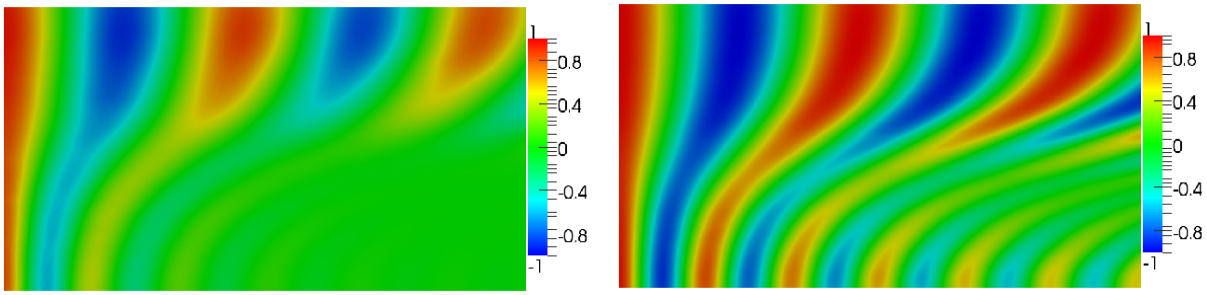


FIGURE 4. Dissipation numérique d'un schéma LS lors de la résolution de l'équation de transport harmononique : pas de maillage h (à gauche) et $h/4$ (à droite)

La meilleure solution a été l'utilisation d'une méthode de Galerkin discontinue décentrée qui est très adaptée aux problèmes de transport. Plus précisément, nous avons choisi le schéma proposé par W. Reed and T. Hill [**RH73**]. C'est le seul des schémas que l'on a testés avec l'approche LS qui n'induit pas d'ondes parasites. Il a de plus de très bonnes propriétés de dispersion (en $O(h^4)$) et de dissipation (en $O(h^3)$). Nous revoyons à la thèse d'Emilie Peynaud [**Pey13**] pour le détails de cette analyse.

Hormis la sélection de la méthode numérique, notre principal apport a été l'analyse du schéma GD décentré dans le cadre de l'équation de transport harmonique. En effet, lors de l'analyse du problème continu, le contrôle en norme $\|\cdot\|_0$ n'est plus immédiat lorsque $\Re\alpha = 0$.

L'approximation GD décentrée du problème $\mathcal{P}_\alpha$ s'écrit : trouver $\psi_h \in W_h^k := \{v_h \in L^2(\Omega) : \forall K \in \mathcal{T}_h,\, v_{h|K} \in \mathcal{P}^k(K)\}$ solution de

$$p_h(\alpha;\psi_h,\varphi_h) = l_h(\varphi_h),\ \forall \varphi_h \in W_h^k \tag{100}$$

où

$$p_h(\alpha;\psi_h,\varphi_h) := (\alpha\psi_h + \underline{v}_0\cdot\nabla\psi_h\,,\,\varphi_h)_{0,\mathcal{T}_h} + (|\underline{v}_0\cdot\underline{n}|[\![\psi_h]\!]\,,\,\varphi_h^+)_{0\,,\mathcal{E}_h^{i,\pm}\cup\mathcal{E}_h^{b,-}} \tag{101}$$

et

$$l_h(\varphi_h) := (f\,,\,\varphi_h)_{0,\Omega} + (|\underline{v}_0\cdot\underline{n}|g\,,\,\varphi_h)_{0\,,\mathcal{E}_h^{b,-}} \tag{102}$$

avec

- $(u\,,\,v)_{0,X_h} = \sum_{K\in X_h} (u\,,\,v)_{0,K}$,

- $\mathcal{T}_h$ un maillage en triangles ou en tétraèdres de Ω,
- $\mathcal{P}^k(K)$ l'espace des polynômes de degré total inférieur ou égal à $k \in I\!N$ sur K,
- l'ensemble des faces de $\mathcal{T}_h$ appartenant à $\partial\Omega$ est noté $\mathcal{E}_h^b$,
- l'ensemble des faces intérieures à Ω est noté $\mathcal{E}_h^i$,
- les ensembles des faces appartenant au bord à flux entrant et sortant $\Gamma^\pm$ sont définis par

$$\mathcal{E}_h^{b,\pm} := \{e \in \mathcal{E}_h^b \,:\, e \subset \Gamma^\pm\}, \tag{103}$$

- l'ensemble des faces appartenant au bord glissant Γ^0 est défini par

$$\mathcal{E}_h^{b,0} := \{e \in \mathcal{E}_h^b \,:\, \underline{v}_0 \cdot \underline{n} = 0 \text{ sur } e \text{ avec } \underline{n} \text{ une normale de } e\}, \tag{104}$$

- les ensembles des faces intérieures sur lesquelles le flux de l'écoulement $\underline{v}_0$ est respectivement non nul et nul sont définis par (voir remarque ci-dessous)

$$\mathcal{E}_h^{i,\pm} := \{e \in \mathcal{E}_h^i \,:\, \exists k > 0 \text{ tel que } |\underline{v}_0 \cdot \underline{n}| \geq k \text{ sur } e \text{ avec } \underline{n} \text{ une normale de } e\}, \tag{105}$$

et

$$\mathcal{E}_h^{i,0} := \{e \in \mathcal{E}_h^i \,:\, \underline{v}_0 \cdot \underline{n} = 0 \text{ sur } e \text{ avec } \underline{n} \text{ une normale de } e\}. \tag{106}$$

- Si $e \in \mathcal{E}_h^{i,\pm}$, alors nous notons $K_e^\pm$ les deux éléments de $\mathcal{T}_h$ tels que $e = K_e^+ \cap K_e^-$ et $\underline{v}_{0|e}$ est dirigé de K_e^- vers K_e^+. De plus, on note $\varphi_h^\pm = (\varphi_{h|K_e^\pm})_{|e}$.

Remarques : Pour simplifier les écritures, nous avons supposé qu'il existe $k > 0$ tel que $\underline{v}_0 \cdot \underline{n} \geq k$ ou $\underline{v}_0 \cdot \underline{n} = 0$ sur chaque face du maillage. Autrement dit, les bords des éléments sont soit à flux d'écoulement entrant, soit sortant, soit glissant. Il ne peut pas y avoir sur une même face une zone à flux entrant et une autre à flux sortant.

Nous avons montré que p_h vérifie, dans le cas du transport harmonique, une condition inf-sup uniforme en h :

Théorème 4. — *Soit $\alpha \in \mathbb{C}^*$. Il existe $h_0 > 0$ et une constante $\beta > 0$ indépendante de h tels que pour tout $h \leq h_0$,*

$$\inf_{\psi_h \in W_h^h} \sup_{\varphi_h \in W_h^k} \frac{\Re\, p_h(\alpha; \psi_h, \varphi_h)}{\|\psi_h\|_{h,\underline{v}_0,\nabla} \|\varphi_h\|_{h,\underline{v}_0,\nabla}} \geq \beta > 0 \tag{107}$$

où

$$\|\varphi_h\|^2_{h,\underline{v}_0,\nabla} = \|\varphi_h\|^2_{0,\Omega} + \sum_{K \in \mathcal{T}_h} h_K \|\underline{v}_0 \cdot \nabla\, \varphi_h\|^2_{0,\Omega} + \sum_{e \in \mathcal{E}_h^{i,\pm} \cup \mathcal{E}_h^{b,\pm}} \frac{1}{2} \left\| |\underline{v}_0 \cdot \underline{n}|^{1/2} [\![\varphi_h]\!] \right\|^2_{0,e} \tag{108}$$

avec h_K le diamètre de l'élément $K \in \mathcal{T}_h$ et $h = \max_{K \in \mathcal{T}_h} h_K$.

Remarque : Ce résultat constitue une extension du résultat classique *i.e* $\alpha \in I\!R^{+,*}$.

Nous pouvons maintenant donner une estimation d'erreur associée au schéma GD :

Théorème 5. — *Soit $\alpha \in \mathbb{C}^*$. Pour h suffisament petit, l'estimation d'erreur entre la solution exacte, $\psi \in H_{conv}(\mathcal{T}_h)$, et la solution GD, ψ_h, est donnée par*

$$\|\psi - \psi_h\|_{h,\underline{v}_0,\nabla} \leq C \inf_{\varphi_h \in W_h^k} \|\psi - \varphi_h\|_{h,\underline{v}_0,\nabla,1/2} \tag{109}$$

où

$$\|u\|^2_{h,\underline{v}_0,\nabla,1/2} = \|u\|^2_{h,\underline{v}_0,\nabla} + \sum_{K \in \mathcal{T}_h} h_K^{-1} \|u\|^2_{0,K} + \sum_{e \in \mathcal{E}_h^{i,\pm}} \frac{1}{2} \||\underline{v}_0 \cdot \underline{n}|^{1/2} u^+\|^2_{0,e}, \tag{110}$$

$$H_{conv}(\mathcal{T}_h) := \{\varphi : \Omega \to \mathbb{C} : \forall K \in \mathcal{T}_h,\ \underline{v}_0 \cdot \nabla\varphi \in L^2(K) \text{ et } \exists \varepsilon > 0,\ \varphi \in H^{1/2+\varepsilon}(K)\} \tag{111}$$

et $C > 0$ est une constante indépendante de h.

De plus, si $\psi \in H^{k+1}(\mathcal{T}_h)$ alors il vient

$$\|\psi - \psi_h\|_{h,\underline{v}_0,\nabla} \leq C\, h^{k+1/2} \|\psi\|_{k+1,\mathcal{T}_h} \tag{112}$$

où $\|\psi\|^2_{k+1,\mathcal{T}_h} = \sum_{K\in\mathcal{T}_h} \|\psi\|^2_{k+1,K}$.

Remarque : Pour démontrer les résultats précédents, nous avons supposé que l'écoulement porteur est Ω-remplissant. Techniquement, cette hypothèse sert uniquement à garantir le caractère bien posé du problème de transport adjoint dont nous nous servons pour obtenir le résultat de stabilité. En particulier, les résultats d'analyse de l'approximation qu'on vient de présenter ne concernent pas les écoulements ayant des zones de recirculation (voir fig. 5) ou des points d'arrêt (voir Fig. 8). Néanmoins, nous montrons dans **[Pey13]** qu'une extension aux écoulements recirculants est possible (voir Fig. 6 et 7) et que les résultats numériques montrent que la méthode reste stable en présence de points d'arrêt (voir Fig. 9).

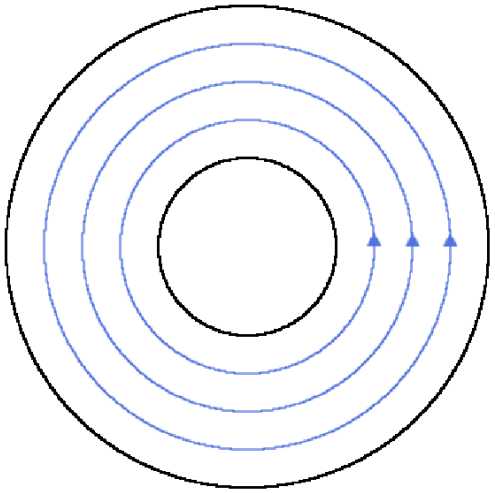


FIGURE 5. Domaine de calcul Ω et écoulement recirculant de la forme $\underline{v}_0(r\cos(\theta), r\sin(\theta)) = v(r)\underline{e}_\theta$.

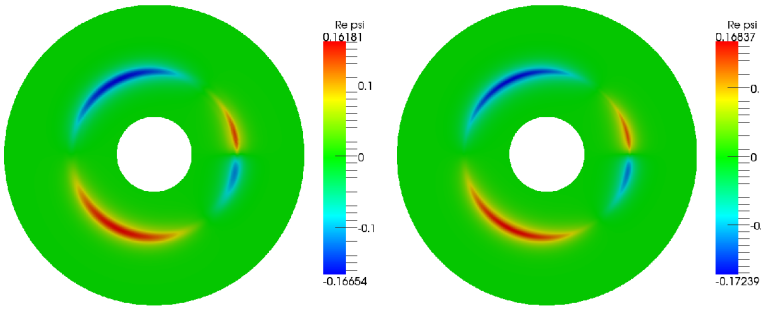


FIGURE 6. Stabilité de la solution GD pour un maillage de pas h (gauche) et un maillage de pas $h/2$ lorsque le problème est bien posé pour l'écoulement recirculant et $\alpha = -0.7i$.

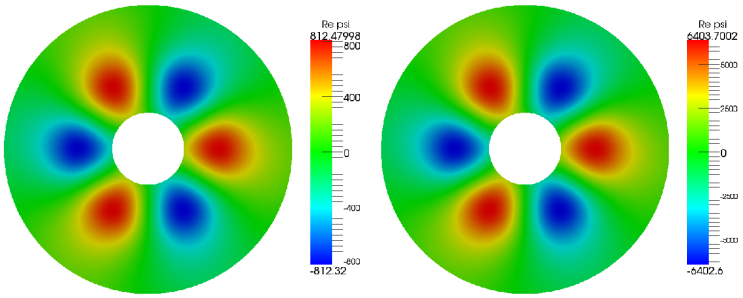


FIGURE 7. Non stabilité de la solution GD (il faut regarder la colorbar!) pour un maillage de pas h (gauche) et un maillage de pas $h/2$ lorsqu'il n'y a pas unicité de la solution pour l'écoulement recirculant et $\alpha = -1.5i$.

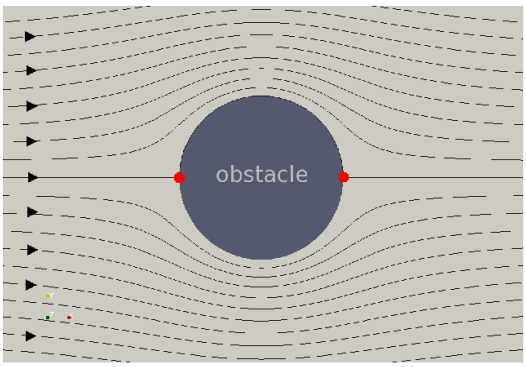


FIGURE 8. Domaine de calcul Ω et écoulement avec points d'arret *i.e* la vitesse est nulle aux points rouges.

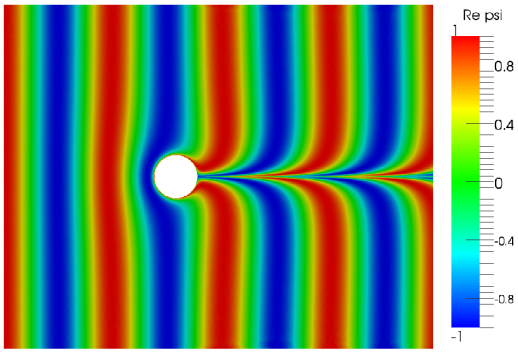


FIGURE 9. Ecoulement avec points d'arret: partie réelle de la solution pour $\alpha = -2i$, $f = 0$ et $g = 1$

3.4. Prise en compte des phénomèmes hydrodynamiques en utilisant l'équation de Galbrun

Maintenant que nous appréhendons correctement le problème de transport harmonique ainsi que son approximation, nous nous intéressons à la résolution de l'acoustique en écoulement en utilisant le système augmenté (83) et (84).

De plus nous nous intéressons à des problèmes dans des domaines non bornés, ou plus exactement dans des domaines très grands devant la zone d'intérêt (le voisinage du réacteur de l'avion par exemple) de sorte qu'il est jusifié de les considérer infinis. Nous appellerons Ω_∞ ce domaine dans la suite. Nous avons supposé que l'écoulement devient "simple" (uniforme ou parallèle cisaillé) à une distance finie de la zone d'intérêt qui contient quant à elle toutes les perturbations de géométrie et d'écoulement. Il est nécessaire de réduire le calcul à la zone d'intérêt et donc d'imposer au bord du domaine des conditions qui rendent compte le mieux possible du caractère non-borné du problème. La méthode que nous avons retenue consiste à utiliser des couches PML. Il est connu que ce type d'approche peut poser de nombreuses difficultés en aéroacoustique, néanmoins dans notre cas, elle fonctionne grâce à l'utilisation de la formulation augmentée **[BBL06, Duc07]**. C'est en effet l'introduction de l'inconnue hydrodynamique ψ qui résout les problèmes en réintroduisant explicitement la distinction entre l'amont et l'aval de l'écoulement, ce qui est seulement implicite dans une formulation variationnelle classique.

Nous supposons pour la présentation que les perturbations d'écoulement et de géométrie sont contenues dans le domaine $\{(x_1,x_2)\in I\!R^2 \,:\, |x_1|<R \text{ et } |\mathbf{x_2}|<\mathbf{R}\}$ *i.e* en dehors de ce domaine l'écoulement est uniforme ou cisaillé. Le domaine de calcul est alors défini par $\Omega_L=\Omega_\infty\cap B_L$ où B_L est le domaine rectangulaire $B_L:=\{(x_1,x_2)\in I\!R^2 \,:\, |x_1|<R+L \text{ et } |x_2|<R+L\}$. Ici, L correspond à la largeur des couches PML et le domaine d'intérêt physique est $\Omega_R=\Omega_\infty\cap B_R$ (voir fig. 10). La méthode des PML est basée sur l'introduction d'un paramètre complexe α vérifiant $\Re\alpha>0$ et $\Im\alpha<0$ et consiste à modifier les opérateurs de dérivation selon la substitution

$$\frac{\partial}{\partial x_i}\rightarrow \alpha_i(x)\frac{\partial}{\partial x_i} \tag{113}$$

avec

$$\alpha_i(x)=\begin{cases} 1 \text{ si } |x_i|<R \\ \alpha \text{ si } |x_i|>R \end{cases} \tag{114}$$

Sur les bords externes des PML, nous imposons à $\underline{u}$ une condition de bord rigide. L'inconnue ψ, est de nature différente (ce n'est pas une onde). D'après l'étude de l'équation de transport, il est naturel d'imposer la valeur de ψ sur le bord à flux d'écoulement entrant $\Gamma_L^-=\{(x_1,x_2)\,:\,x_1=-(R+L)\}$. D'autre part, il n'y a pas de raison pour que la vorticité ψ associée à l'onde sortante soit non nulle à l'amont de la zone de perturbation **[BBL06]** puisque nous avons supposé lorsque nous avons défini l'onde sortante qu'il n'y a pas de source à l'infini. En fait, les tourbillons hydrodynamiques sont produits par la source et par le couplage entre l'acoustique et l'hydrodynamique due à un écoulement non uniforme, et sont ensuite transportés à l'aval par le fluide. Il apparaît alors naturel d'imposer $\psi=D\psi/Dt=0$ sur Γ_L^-, assurant ainsi la causalité de ψ. Finalement le problème à résoudre s'écrit :

$$\rho_0\frac{D_\alpha^2\underline{u}}{Dt^2}-\nabla_\alpha(\rho_0c_0^2 div_\alpha\underline{u})+rot_\alpha(\rho_0c_0^2(rot_\alpha\underline{u}-\psi))+div_\alpha\underline{u}\nabla p_0-{}^t\nabla_\alpha\underline{u}\cdot\nabla p_0=\underline{f} \text{ dans } \Omega_L, \tag{115a}$$

$$\frac{D_\alpha^2\psi}{Dt^2}=-2\frac{D_\alpha}{Dt}\left(\mathcal{B}_\alpha\underline{u}\right)-\mathcal{C}\underline{u}+\frac{1}{\rho_0}rot\underline{f} \text{ dans } \Omega_L, \tag{115b}$$

$$\underline{u}\cdot\underline{n}=rot_\alpha(\underline{u})-\psi=0 \text{ sur } \partial\Omega_L, \tag{115c}$$

$$\psi=\frac{D_\alpha\psi}{Dt}=0 \text{ sur } \Gamma_L^-. \tag{115d}$$

Remarques :

1. La condition $rot_\alpha\underline{u}-\psi=0$ sur $\partial\Omega_L$ est une traduction de la condition $rot\underline{u}-\psi=0$ sur $\partial\Omega$ qui est nécessire pour assurer l'équivalence entre la formulation de Galbrun et sa version augmentée.

2. Nous n'avons pas mis d'indice α pour les $\mathcal{C}\underline{u}$, $rot\underline{f}$ et ∇p_0 car ils sont nuls dans les PML à cause des hypothèses sur l'écoulement porteur.

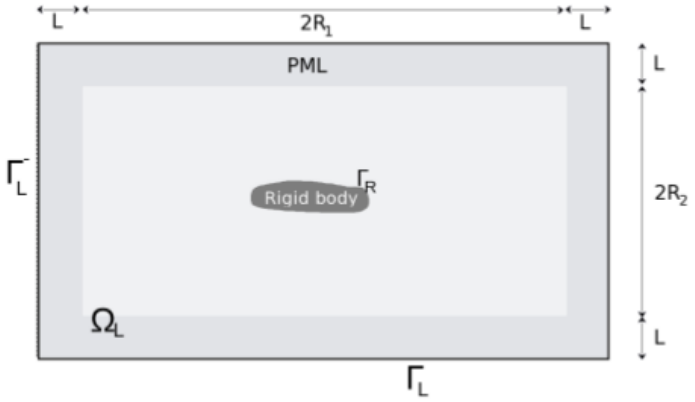


FIGURE 10. Domaine de calcul avec PML

L'équation (115b) et les conditions aux limites (115d) peuvent facilement être récrites à l'aide de simples équations de transport harmoniques. En effet, on peut décomposer ψ sous la forme $\psi = \psi_{\mathcal{B}} + \psi_{\mathcal{C}} + \psi_{\underline{f}}$ avec

$$\left\{ \begin{array}{rcl} \dfrac{D_\alpha \psi_{\mathcal{B}}}{Dt} & = & -2\,\mathcal{B}_\alpha \underline{u} \text{ dans } \Omega_L \\ \psi_{\mathcal{B}} & = & 0 \text{ sur } \Gamma_L^- \end{array} \right. , \tag{116}$$

$$\left\{ \begin{array}{rcl} \dfrac{D_\alpha \psi_{\mathcal{C}}}{Dt} & = & \tilde{\psi}_{\mathcal{C}} \text{ dans } \Omega_L \\ \dfrac{D_\alpha \tilde{\psi}_{\mathcal{C}}}{Dt} & = & -\mathcal{C}\underline{u} \text{ dans } \Omega_L \\ \psi_{\mathcal{C}} & = & 0 \text{ sur } \Gamma_L^- \\ \tilde{\psi}_{\mathcal{C}} & = & 0 \text{ sur } \Gamma_L^- \end{array} \right. \tag{117}$$

et

$$\left\{ \begin{array}{rcl} \dfrac{D_\alpha \psi_{\underline{f}}}{Dt} & = & \tilde{\psi}_{\underline{f}} \text{ dans } \Omega_L \\ \dfrac{D_\alpha \tilde{\psi}_{\underline{f}}}{Dt} & = & \dfrac{1}{\rho_0}\, rot\, \underline{f} \text{ dans } \Omega_L \\ \psi_{\underline{f}} & = & 0 \text{ sur } \Gamma_L^- \\ \tilde{\psi}_{\underline{f}} & = & 0 \text{ sur } \Gamma_L^-. \end{array} \right. \tag{118}$$

En particulier, sous les hypothèses de la proposition 2, on sait qu'il existe un opérateur $\mathcal{T}$ qui à $\underline{u} \in V = \{\underline{u} \in H^1(\Omega_L)^2 \,:\, \underline{u}\cdot\underline{n} = 0 \text{ sur } \partial\Omega_L\}$ va associer $\mathcal{T}\underline{u} = \psi_{\mathcal{B}} + \psi_{\mathcal{C}}$ l'unique solution de (116 et 117).

Nous pouvons alors écrire notre problème couplé sous la forme variationnelle compacte suivante : trouver $\underline{u} \in V$ tel que pour tout $\underline{v} \in V$

$$a(\underline{u},\underline{v}) + b(\underline{u},\underline{v}) = \ell(\underline{v}) \tag{119}$$

où

$$
\begin{aligned}
a(\underline{u},\underline{v}) &= \int_{\Omega_L} \frac{\rho_0}{\alpha_1\alpha_2}\left(c_0^2 div_\alpha \underline{u}\, div_\alpha \bar{\underline{v}} + c_0^2 rot_\alpha \underline{u}\, rot_\alpha \bar{\underline{v}} - (\underline{v}_0\cdot\nabla_\alpha)\underline{u}\,(\underline{v}_0\cdot\nabla_\alpha)\bar{\underline{v}}\right)\,\mathrm{d}x \\
&\quad - \int_{\Omega_L} \frac{\rho_0\, c_0^2}{\alpha_1}\mathcal{T}\,\underline{u}\, rot_\alpha \bar{\underline{v}}\,\mathrm{d}x
\end{aligned} \tag{120}
$$

$$
b(\underline{u},\underline{v}) = \int_{\Omega_L} \frac{-\rho_0\omega}{\alpha_1\alpha_2}(2i\underline{v}_0\cdot\nabla_\alpha)\underline{u} + \omega\underline{u})\cdot\bar{\underline{v}}\,\mathrm{d}x + \int_{\Omega_R}(div\underline{u})\nabla p_0 - {}^t\nabla\underline{u}\cdot\nabla p_0)\,\bar{\underline{v}}\,\mathrm{d}x \tag{121}
$$

et

$$
\ell(\underline{v}) = \int_{\Omega_R} \underline{f}\cdot\ \underline{v}\,\mathrm{d}x + \int_{\Omega_L}\frac{\rho_0 c_0^2}{\alpha_1}\psi_{\underline{f}}\, rot_\alpha\bar{\underline{v}}\,\mathrm{d}x \tag{122}
$$

Nous avons effectué l'analyse de (115) en introduisant un cadre un peu restrictif mais qui permet d'aller jusqu'au bout d'une démonstration complète sur le caractère bien posé du problème couplé et de son approximation. Nous présenterons néanmoins des cas tests qui sortent de ce cadre théorique et qui montrent que la méthode d'approximation mixte proposée fournit toujours des résultats satisfaisants.

Plus précisément, nous avons considéré deux situations :

- Le domaine de propagation est un sous-espace de $\mathbb{R}^2$ défini par

$$
\Omega_\infty = \{(x_1,x_2)\in\mathbb{R}^2\ :\ x_2 > h(x_1)\} \tag{123}
$$

  où h est une fonction continue positive telle que $h(x_1)=0$ pour tout $|x_1|>r$ (voir fig. 11).
- Le domaine de propagation $\Omega_\infty = \mathbb{R}^2$ est occupé par un fluide dont le champ de vitesse est subsonique et dirigé selon la direction $\underline{e}_1$. La composante horizontale de $\underline{v}_0$ est connue analytiquement et dépend uniquement de la variable verticale x_2 *i.e* $\underline{v}_0 = v_1(x_2)\underline{e}_1$ où $v_1(x_2) > 0, \forall x_2$. De plus, nous supposons qu'en dehors du domaine perturbé $\{(x_1,x_2)\in\mathbb{R}^2\ :\ |x_2|<R\}$ l'écoulement est uniforme (voir fig. 12).

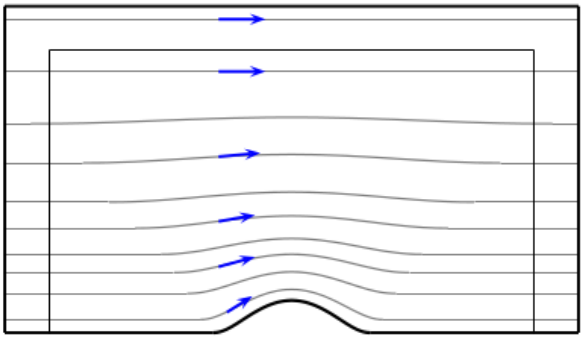


FIGURE 11. Cas d'une perturbation géométrique.

Finalement, nous avons le résultat suivant ([**BMMP10, Pey13**]) :

Théorème 6. — *Supposons que $\underline{v}_0$ est $\Omega-$remplissant. Le problème (119) est bien posé si*

$$
\inf_{x\in\Omega_L}\left(1-\frac{|\underline{v}_0|}{c_0^2}\right) > \frac{\tilde{c}(\tau,\underline{v}_0,\alpha_1)}{K_\alpha}(\tau^{-1}\|\mathcal{B}_\alpha\| + \|\mathcal{C}\|) \tag{124}
$$

où $\tilde{c}(\tau,\underline{v}_0,\alpha_1)$ est une constante définie par :

$$
\tilde{c}(\tau,\underline{v}_0,\alpha_1) = \frac{1}{|\alpha_1|}\max\left(\tilde{C}(\tau,\underline{v}_0), \tau\,\tilde{C}(\tau,\underline{v}_0)^{1/2}\right) \tag{125}
$$

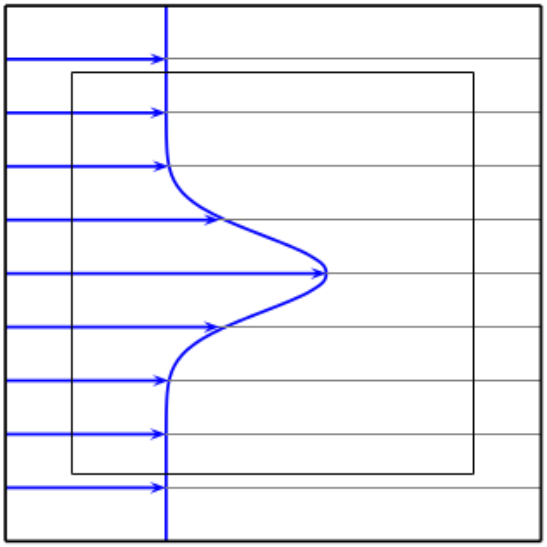


FIGURE 12. Cas d'un jet analytique.

$$\text{avec } \tilde{C}(\tau, \underline{v}_0) = \frac{\tau^2 e^{2\tau \|div\, \underline{v}_0\|_\infty}}{2},$$

$$\|\mathcal{B}_\alpha\| = \sup_{\underline{v}\in V,\, \underline{v}\neq 0} \frac{\|\rho_0^{1/2} c_0 \mathcal{B}_\alpha \underline{v}\|_{0,\Omega_L}}{\|\rho_0^{1/2} c_0 \nabla \underline{v}\|_{0,\Omega_L}} \text{ et } \|\mathcal{C}\| = \sup_{\underline{v}\in V,\, \underline{v}\neq 0} \frac{\|\rho_0^{1/2} c_0 \mathcal{C} \underline{v}\|_{0,\Omega_L}}{\|\rho_0^{1/2} c_0 \nabla \underline{v}\|_{0,\Omega_L}}. \tag{126}$$

Remarque : Ce théorème constitue un premier résultat sur le caractère bien-posé du système de Gabrun augmenté en présence de PML. On peut néanmoins facilement être critique vis-à-vis de celui-ci. En effet, les hypothèses nécessaires sont restrictives : l'écoulement doit être $\Omega-$remplissant et vérifier la condition (124) ce qui exclut *de facto* de nombreux écoulements subsoniques réalistes. Pour voir cela, il suffit de constater que la condition (124) devient par exemple plus restrictive lorsque le plus grand temps de parcours τ (voir partie transport pour sa définition) augmente ou encore par l'augmentation éventuelle des normes $\|\mathcal{B}_\alpha\|$ et $\|\mathcal{C}\|$ avec la taille du domaine. Des estimations plus fines des différents termes lors de la démonstration permettraient peut-être d'étendre ce résultat.

La discrétisation du problème augmenté est basée sur une approximation par des éléments finis classiques pour le déplacement et sur une approche de Galerkin discontinue pour l'inconnue hydrodynamique ψ. Plus précisément, si on considère $\mathcal{T}_h$ une triangulation de Ω_L, les espaces d'approximation pour $\underline{u}$ et ψ sont respectivement :

$$V_h^k := \{\underline{v}_h \in V \,:\, \forall K \in \mathcal{T}_h,\, \underline{v}_{h|K} \in (\mathcal{P}^k(K))^2\} \tag{127}$$

et

$$W_h^k := \{\varphi_h \in L^2(\Omega_L) \,:\, \forall K \in \mathcal{T}_h,\, \varphi_{h|K} \in \mathcal{P}^k(K))\} \tag{128}$$

La formulation variationnelle discrète s'écrit alors : trouver $\underline{u}_h \in V_h^k$ tel que pour tout $\underline{v}_h \in V_h^k$,

$$a_h(\underline{u}_h, \underline{v}_h) + b(\underline{u}_h, \underline{v}_h) = \ell_h(\underline{v}_h) \tag{129}$$

où les formes a_h et ℓ_h sont définies en remplaçant respectivement $\mathcal{T}$ et $\psi_{\underline{f}}$ par l'opérateur approché $\mathcal{T}_h$ et $\psi_{\underline{f},h}$ qui sont obtenus par des résolutions successives d'équations de transport harmoniques (voir la décomposition (116, 117 et 118) en utilisant la méthode de Galerkin discontinue proposée dans la première partie.

Sous les conditions du théorème 6, nous avons montré **[BMM+12, Pey13]** que le schéma proposé vérifie une condition inf-sup uniforme : il existe une constante $\delta > 0$ indépendante de h et $h_0 > 0$ telle que pour tout $h \in]0, h_0]$,

$$\inf_{\underline{v}_h \in V_h^k} \inf_{\underline{w}_h \in V_h^k} \frac{|a_h(\underline{v}_h, \underline{w}_h) + b(\underline{v}_h, \underline{w}_h)|}{\|\underline{v}_h\|_{1,\Omega_L} \, \|\underline{w}_h\|_{1,\Omega_L}} \geq \delta > 0 \tag{130}$$

De plus, si $\underline{u} \in [H^{k+1}(\Omega_L)]^2$, on a l'estimation d'erreur *a priori*

$$\|\underline{u}_h - \underline{u}\|_{1,\Omega_L} \leq C \, h^k |\underline{u}|_{k+1,\Omega_L} \tag{131}$$

Regardons à présent quelques résultats numériques. Le premier qui correspond à la figure 13 nous montre que lorsque l'on est dans une situation purement potentielle (*i.e.* écoulement potentiel et source acoustique ($rot\underline{f} = 0$)) alors la méthode basée sur le système augmenté de Galbrun redonne bien une solution similaire au modèle potentiel. En effet, aucun couplage entre l'acoustique et l'hydrodynamique n'existe dans ce cas. Le second (fig. 14) montre que lorsque l'écoulement n'est plus potentiel, les phénomènes hydrodynamiques induits par le jet (on ne les voit pas sur la figure car ils sont d'une échelle plus basse **[Pey13]**) provoquent une modification de la directivité de l'onde (il suffit de comparer les traits noirs sur la figure). Le modèle potentiel ne peut naturellement pas restituer ce phénomène. Nous renvoyons à la thèse d'Emilie Peynaud pour de nombreux resultats numériques. En particulier, on y trouve un estimateur *a posteriori* défini par l'évaluation de $\psi_h - rot\underline{u}_h$ pour vérifier que la solution obtenue par le modèle augmenté est cohérente avec le modèle de Galbrun initial. C'est un outil précieux pour valider nos résultats numériques.

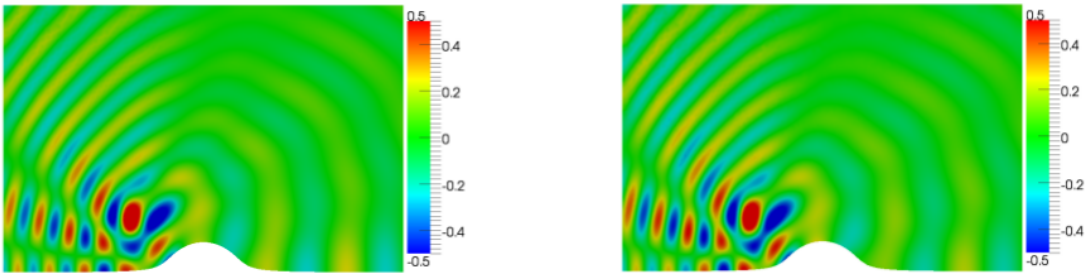


FIGURE 13. Rayonnement acoustique dans un écoulement potentiel (voir fig. 11) et pour une source acoustique : comparaison de la perturbation de vitesses euleriennes obtenues par Galbrun (gauche) et le modèle potentiel (droite)

3.5. Bilan et perspectives

L'objectif principal de ces travaux était de répondre aux deux questions suivantes :

1. Faut-il tenir compte du couplage entre la propagation d'onde et le transport hydrodynamique lorsque l'on calcule le rayonnement acoustique en écoulement?
2. Si oui, comment réaliser cela?

Les travaux que nous avons réalisés dans l'ANR AEROSON et la thèse d'Emilie Peynaud nous ont permis de répondre positivement à la première question et de proposer une méthode numérique basée sur l'équation de Galbrun pour réaliser cet objectif. Cette méthode a été dérivée et analysée en 2D mais il n'y a *a priori* pas de difficulté pour l'étendre en 3D. Nous nous attendons à ce que les effets hydrodynamiques soient plus marqués qu'en 2D car en 3D l'inconnue hydrodynamique ψ devient vectorielle. Sa mise en place en 3D va poser néanmoins de nombreux problèmes. Premièrement, la taille du système linéaire à résoudre va considérablement augmenter, non seulement à

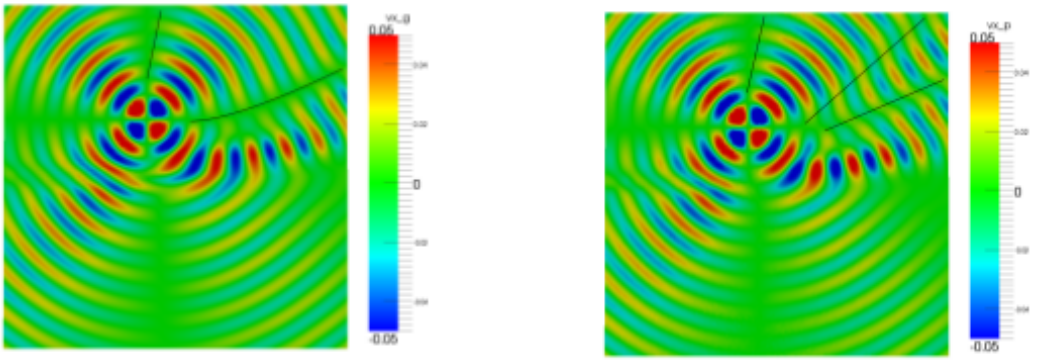


FIGURE 14. Rayonnement d'une source acoustique dans un jet (voir fig. 12) : comparaison de la perturbation des vitesses euleriennes obtenues par Galbrun (gauche) et le modèle potentiel (droite)

cause du passage à une dimension supérieure, mais aussi à cause du fait que ψ est désormais une inconnue vectorielle qui est approchée par des éléments finis discontinus. Une résolution par une inversion directe sera donc inenvisageable et la mise en place de stratégies de résolution efficace du système linéaire s'avère alors inévitable. L'établissement de telles méthodes est d'autant plus difficile que notre modèle couple deux types de phénomènes physiques : la propagation d'onde d'une part et les convections hydrodynamiques d'autre part. On pourrait évidemment penser à utiliser simplement une méthode itérative mais malheureusement l'équation de Galbrun est proche d'une équation de Helmholtz convectée qui conduit donc à des systèmes linéaires indéfinis qui sont difficiles à préconditionner efficacement. En particulier, à notre connaissance, la construction d'un préconditionneur efficace pour ce type d'équation est encore une question ouverte. C'est pourquoi, nous envisagerions plutôt de nous tourner vers des méthodes de type décomposition de domaine. Nous avons d'ailleurs déjà obtenu des résultats prometteurs en 2D lors d'un stage au cours duquel des conditions de transmission adaptées à notre modèle ont été proposées.

Un autre aspect qui pourrait être utilisé pour diminuer la complexité du modèle est d'utiliser dans les endroits où cela est possible le modèle faible Mach proposé dans **[BMMP10]**. Ce dernier donne en effet un lien explicite entre le déplacement $\underline{u}$ et ψ. Il s'agirait dans ce cas de faire une analyse d'erreur *a posteriori* sur le modèle faible Mach par rapport au modèle augmenté complet afin de dériver un indicateur local de l'erreur du modèle. Ce dernier pourrait être utilisé pour choisir localement dans le domaine de calcul quelle précision dans le modèle est nécessaire pour obtenir un résultat correct.

De plus, nous jugeons que l'analyse de la méthode proposée est assez restrictive par rapport aux écoulements porteurs. Néanmoins, les résultats numériques montrent que l'extension à des écoulements plus complexes (possédant des zones de recirculations, des points d'arrêt, voire instables) semble possible. Il est donc complémentaire de poursuivre l'effort de compréhension théorique de l'approche.

Enfin, notre travail montre la faisabilité du couplage entre l'équation de transport harmonique et le modèle de Galbrun augmenté et propose une méthode de résolution par éléments finis mixtes continus/discontinus avec l'usage de couches PML. Ce couplage avec le transport hydrodynamique pourrait désormais être envisagé pour des modèles de propagation acoustique autres que le modèle de Galbrun augmenté comme c'est actuellement le cas dans la thèse de Antoine Bensalah encadrée

par P. Joly et J.F Mercier dans laquelle l'équation de Golstein est utilisée.

Ces travaux ont donné lieu à deux publications dans des journaux : 1 Journal of Computational and Applied Mathematics et 1 Communication In Computational Physics. Ils ont été réalisés en collaborant avec E. Peynaud (doctorante), A.S Bonnet, F. Millot, J.F Mercier ainsi que des étudiants en stage niveau master V. Lebris, G. Pigeron, M. A. Bouhlel et J.M Bart. Enfin, ces travaux font aussi partie du projet ANR AEROSON.

PARTIE II

PARTIE ADMINISTRATIVE

Curriculum Vitae

Sébastien PERNET
Né le 19/10/1977
ONERA, 2 avenue Edouard Belin 31055 Toulouse
Tél. : 0562252634
E-mail : Sebastien.Pernet@onera.fr

FORMATION :

Octobre 2001-septembre 2004 : Thèse en Mathématiques Appliquées (Office National d'Etudes et de Recherches Aérospatiales (ONERA) de Toulouse)

Titre : Etude de méthodes d'ordre élevé pour la résolution des équations de Maxwell dans le domaine temporel.

Date de soutenance : 22 novembre 2004

Jury :

M. Gary Cohen (directeur de thèse: Chercheur HDR, INRIA)
M. Abderrahmane Bendali (rapporteur: PR INSA Toulouse)
M. Bruno Després (rapporteur: PR Paris VI)
Mme Christine Bernardi (présidente: DR CNRS)
M. Jean Dolbeault (examinateur: DR CNRS, Paris IX)
M. Xavier Ferrieres (examinateur: chercheur ONERA, Toulouse)
M. Patrick Joly (examinateur: DR INRIA)

Mention : Très honorable avec les félicitations du jury

Thèmes : Equations de Maxwell, méthodes d'ordre élevé, éléments finis, Galerkin discontinu, méthodes spectrales, condensation de masse, dispersion, dissipation, estimation d'erreur, stabilité.

DEA de Mathématiques appliquées, Maîtrise et licence de Mathématiques Pures, université de Metz.

EXPERIENCES PROFESSIONNELLES :

Juin 2012 - x : Ingénieur de recherche sénior ONERA à Toulouse

Mars 2009 - mai 2012 : Chercheur sénior au Centre Européen de Recherche de Formation Avancée en Calcul Scientifique (CERFACS) à Toulouse

Mars 2007- février 2009 : Chercheur junior au CERFACS

Mars 2005- février 2007 : Post-doctorant au CERFACS

Octobre 2001- Septembre 2004 : Doctorant à l'ONERA

Encadrement d'étudiants et jury

- **Thèses :**
  1. Mathieu Mounaury (2015- x) : *Méthodes de visualisation adaptées au calcul de haute précision : analyse et exploitation rigoureuses des résultats* (Directeur de thèse : Christophe Besse), financement ONERA . (encadrement 25%)
  2. Marc Bakry (2013-2016) : *Fiabilité et optimisation des calculs obtenus par des formulations intégrales en propagation d'ondes* (Directeur de thèse : Patrick Ciarlet), Financement ANR et DGA. (encadrement 80%)
  3. Yohann Dudouit (2010-2014) : *Spatio-temporal refinement using a discontinuous Galerkin approach for elastodynamic in a high performance computing framework* (Directeur de thèse Luc Giraud). Financement Total. (encadrement 50%)
  4. Emilie Peynaud (2009-2013) : *modélisation de la propagation du son en écoulements via l'équation de Galbrun* (Directrice de thèse : Anne-Sophie Bonnet BenDhia). Financement ANR. (encadrement 50%)
- **Stages niveau master 1 et 2 :**
  1. Vincent Lebris, 4ème année INSA Toulouse, *Calcul du rayonnement acoustique et champ lointain en présence d'écoulement*, 2007 (avec F. Millot).
  2. Benoit Mallet, stage de fin d'étude Math-Meca, *Introduction d'approximations spatiales non identiques par cellules dans un schéma Galerkin Discontinu*, 2008 (avec X. Ferrieres).
  3. Emilie Peynaud, 4ème année INSA Toulouse, *Résolution d'une équation de transport pour des applications en aéroacoustique*, 2008 (avec F. Millot).
  4. Tyra Vanolmen, 4ème année INSA Toulouse, *Etude numérique d'un espace d'éléments finis de surface basé sur un raffinement barycentrique*, 2008.
  5. Gregory Pigeron, stage de fin d'étude Math-Meca, *Meilleur prise en compte des phénomènes hydrodynamiques: couplage de l'équation de Galbrun avec une équation de convection*, 2009 (avec F. Millot).
  6. Mohamed Amine Bouhlel, 4ème année INSA, *Résolution itérative et préconditionnement d'une méthode d'éléments finis en aéroacoustique*, 2010 (avec F. Millot et E. Peynaud).
  7. Ghalem Allai et Jaipyo Lee, stage d'initiation à la recherche de 2ème année SUPAERO, *Estimation d'erreur a posteriori pour les équations intégrales: les techniques de reconstruction*, 2011.
  8. Alexandra Dececco, stage Master 1, *indicateur d'erreur a posteriori pour les équations intégrales*, 2011.
  9. Bart Jean-Michel, 4ème année INSA, *Méthode de décompostion de domaine pour les problèmes d'acoustique en écoulement*, 2011.
  10. Elena Queirolo, EPFL niveau master 1, *Etude d'un indicateur d'erreur a posteriori pour l'EFIE*, 2013.
  11. Pol André Hass, Master pro ingénierie mathématique Toulouse, *Méthode pour une visualisation adaptée au calcul haute précision en acoustique*, 2014 (avec V. Mouysset).
  12. Maxime Alexandre, Master 2 université de Lorraine, *Etude de la tomographie par impédance électrique pour une application en génie civil*, 2015 (avec V. Mouysset).
  13. Nadir-Alexandre Messai, PFE ISAE, *Mise en place d'algorithmes pour la résolution autoadaptative de problèmes de diffraction d'ondes*, 2017 (avec D. Levadoux).

Participation à des jurys de thèse :
• Invité au jury de thèse de Laura Pebernet soutenue le 30/11/2010. Etude d'un modèle PIC dans une approximation GD pour les équations de Maxwell/Vlasov-Recherche d'une solution Hybride non conforme efficace.

• Examinateur au jury de thèse de Jennifer Bourguignon-Mirebeau soutenue le 12 décembre 2011, Préconditionnement de méthodes de décomposition de domaine pour les problèmes de diffraction d'ondes électromagnétiques impliquant une cavité profonde.

• Examinateur au jury de thèse de Cyril Agut soutenue le 13 décembre 2011, Schéma numériques d'ordre élevé en temps et en espace pour l'équation des ondes.

• Examinateur au jury de thèse de Emilie Peynaud soutenue le 21/06/2013. Modélisation de la propagation du son en écoulements via l'équation de Galbrun.

• Examinateur au jury de thèse de Yohann Dudouit soutenue le 08/11/2014. Spatio-temporal refinement using a discontinuous Galerkin approach for elastodynamic in a high performance computing framework.

• Examinateur au jury de thèse de Marc Bakry soutenue le 03/10/2016. Fiabilité et optimisation des calculs obtenus par des formulations intégrales en propagation d'ondes.

Publications, séminaires, conférences et rapports

Livre :

Gary Cohen and Sébastien Pernet, *Finite Element and Discontinuous Galerkin Methods for Transient Wave Equations*, Scientific Computation, Springer-verlag, 381 pages, 2016.

Revues avec comités de lecture :

[A1] M. Bakry, S. Pernet, F. Collino, *A new accurate residual-based a posteriori error indicator for BEM in 2D-acoustics*, acceptée pour publication dans Computers and Mathematics with Applications.

[A2] D. Levadoux, F. Millot, S. Pernet, *A well-conditioned boundary integral equation for transmission problems of electromagnetism*, J. Integral Equations Applications, vol. 27, no 3, pp. 431-454, 2015.

[A3] D. Levadoux, F. Millot, S. Pernet, *An Unpreconditioned Boundary-Integral for Iterative Solution of Scattering Problems with Non-Constant Leontovitch Impedance Boundary Conditions*, Communication In Computational Physics, vol. 15, issue 5, pp. 1431-1460, 2014.

[A4] B.Mallet, X.Ferrieres, S.pernet, J.B.Laurent, B.Pecqueux and P.Seimandi, *A p-strategy with a time-stepping method in a Discontinuous Galerkin Approach to solve Electromagnetic Problems*, Journal of Computational Methods in Physics, Vol. 2013, Article ID 563480, p. 1-13, 2013

[A5] A.S. Bonnet-Ben Dhia, J.F. Mercier, F. Millot, S. Pernet and E. Peynaud, *Time-Harmonic Acoustic Scattering in a Complex Flow: a Full Coupling Between Acoustics and Hydrodynamics*, CICP, vol 11, no 2, pp 555-572, 2012.

[A6] S. Pernet, *A well-conditioned integral equation for iterative solution of scattering problems with a variable leontovitch boundary condition*, Mathematical Modelling and Numerical Analysis vol.44,

no 4 p781-801, 2010.

[A7] A. S. Bonnet, J. Mercier, F. Millot and S. Pernet, *A low Mach model for time harmonic acoustics in arbitrary flows*, Journal of Computational and Applied Mathematics vol.234 no 6, p.1868-1875, 2010.

[A8] D. Levadoux, F. Millot, and S. Pernet, *New trends in the preconditioning of integral equations of electromagnetism*, Springer-Verlag Berlin Heifelberg, Scientific Computing in Electrical Engineering SCEE 2008 by Janne Roos,Luis R. J. Costa(Mathematics in industry 14):383-394, 2010.

[A9] F. Collino, F. Millot, and S. Pernet, *Boundary-Integral Methods for iterative solution of scattering problems with variable impedance surface condition*, PIER, vol. 80, p. 1-28 , 2008.

[A10] E. Montseny, S. Pernet, X. Ferrieres, and G. Cohen, *Dissipative terms and local time-stepping improvements in a spatial high order Discontinuous Galerkin scheme for the time domain Maxwell's equations*, Journal of Computational Physics, 227, p. 6795-6820, 2008.

[A11] L. Pebernet, X. Ferrieres, S. Pernet, B. Michielsen, F. Rogier, and P. Degond, *Discontinuous Galerkin method applied to electromagnetic compatibility problems: introduction of thin wire and thin resistive material models*, IET Science, Measurement & Technology, 2, 395-401, 2008.

[A12] S. Pernet and X. Ferrieres, *hp a-priori error estimates for a non-dissipative spectral discontinuous Galerkin method to solve the Maxwell equations in the time domain*, Mathematics of Computation, 76, p. 1801-1832, 2007.

[A13] G. Cohen, X. Ferrieres, and S. Pernet, *A spatial high-order hexahedral discontinuous Galerkin method to solve Maxwell's equations in time domain*, J. Comput. Phys., 217, p. 340-363, 2006.

[A14] G. Cohen, X. Ferrieres, and S. Pernet, *Une méthode de Galerkin Discontinue d'ordre élevé pour résoudre les équations de Maxwell dans le domaine temporel*, CRAS Physique, 7, p. 494-500, 2006.

[A15] S. Pernet, X. Ferrieres, and G. Cohen, *High Order Finite Element Method to solve Maxwell's equations in time domain*, IEEE Trans. on Antennas and Propagation, 53, p. 2889-2899, 2005.

Papiers Soumis :

[S1] Y. Dudouit, L. Giraud, F. Millot, S; Pernet, *Interior penalty discontinuous Galerkin method for coupled elasto-acoustic media*, https://dumas.ccsd.cnrs.fr/UNIV-BORDEAUX/hal-01406158v1, under review.

Séminaires et invitations :

[B1] S. Pernet , *Adaptive Post-Processing Method to Represent High-order Numerical Solutions*, Journées Ondes du Sud Ouest à Pau, 10 mars 2016.

[B2] S. Pernet, *Quelques résultats sur une méthode de Galerkin discontinu en électromagnétisme instationnaire*, Journées Ondes du Sud Ouest à Toulouse, 6 février 2014.

[B3] S. Pernet, *Nouvelle tendance pour le préconditionnement des équations intégrales en électromagnétisme*, Séminaire au groupe de travail "applications des mathématiques" de ENS Cachan, antenne Bretagne, 25 février 2009.

[B4] S. Pernet, *Une famille d'équations intégrales en source pour un calcul efficace de la diffraction d'une onde électromagnétique*, Invitation à l'atelier Melina, du mardi 12 au vendredi 15 Mai 2009 à Dinard.

[B5] S. Pernet, *Une équation intégrale bien conditionnée pour résoudre des problèmes de diffractions d'ondes électromagnétiques avec condition de Léontovitch non constante*, Séminaire mensuel de l'UMR POEMS, 25 Octobre 2007.

[B6] S. Pernet, *Méthodes volumiques d'ordre élevées pour résoudre les équations de Maxwell dans le domaine temporel*, Juin 2004, séminaire INRIA-ONERA organisé pour la venue de JS Hesthaven (Brown university).

[B7] S. Pernet, *Etude d'une méthode de Galerkin discontinue pour résoudre les équations de Maxwell dans le domaine temporel*, Juin 2004, séminaire projet CAIMAN (INRIA Sophia Antipolis).

[B8] S. Pernet, *Application de méthodes de résolution des équations de Maxwell dans le domaine temporel à des problèmes d'électromagnétisme*, invitation GDR onde (groupe thématique 1), 27 novembre 2002, Ecole Polytechnique, Palaiseau.

Conférences avec actes :

[CA1] M. Bakry, S. Pernet, F. Collino, *Reliable and efficient a posteriori error estimate for EFIE in electromagnetism*, waves 2017, Minneapolis, 2017.

[CA2] M. Bakry, S. Pernet, F. Collino, *A new a posteriori error estimate for BEM in 2D-acoustics*, waves 2015, Karlsruhe Germany, 2015.

[CA3] M. Bakry, S. Pernet, *A posteriori error control for BEM in 2D-acoustics*, ECCOMAS, Barcelona Spain, 2014, 11 pages (www.wccm-eccm-ecfd2014.org/admin/files/filePaper/p3680.pdf).

[CA4] A.S. Bonnet-BenDhia, J.F. Mercier, F. Millot, S. Pernet, E. Peynaud, *Time-harmonic acoustic scattering in a complex flow*, proceeding of the Acoustics 2012 Nantes conference, p.1312-1316, 23-27 april 2012.

[CA5] D. Levadoux, F. Millot, and S. Pernet, *Intrinsically well-conditioned integral equations for scattering by homogeneous bodies*, Numelec 2012, Marseille, 2012.

[CA6] A.S. Bonnet-BenDhia, J.F. Mercier, F. Millot, S. Pernet et E. Peynaud, *Galbrun based numerical scheme to compute time-harmonic scattering in an arbitrary mean flow*, proceeding of 17th AIAA/CEAS Aeroacoustics Conference (32nd AIAA Aeroacoustics Conference), American Institute of Aeronautics and Astronautics, 6 pages, 2011.

[CA7] E. Peynaud, F. Millot, S. Pernet, A.-S. Bonnet-Ben Dhia and J.-F. Mercier, *A mixed continuous/discontinuous finite element method for acoustics in an arbitrary flow*, waves 2011, Vancouver, July 25-29, 2011.

[CA8] J.B Laurent, V. Mouysset, S. Pernet, and X. Ferrieres, *Development of an adaptive mesh in a discontinuous Galerkin method for time Maxwell's equations*, waves 2011, Vancouver, July 25-29, 2011.

[CA9] S. Pernet, X. Ferrieres, L. Pebernet, B. Pecqueux, *Some techniques for an efficient use of a discontinuous Galerkin approximation of time domain Maxwell's equations on hexahedral meshes*, Mathematical and Numerical Aspects of Wave Propagation, Pau, june 2009.

[CA10] A.S. Bonnet-Ben Dhia, J.F. Mercier, F. Millot and S. Pernet, *Time-harmonic acoustic scattering in a complex flow : a full coupling between acoustics and hydrodynamics*, Waves2009, Mathematical and Numerical Aspects of Wave Propagation, Pau, june 2009.

[CA11] Bendali, M. Fares, K. Lemrabet and S. Pernet, *Recent Developments in the Scattering of an Electromagnetic Wave by a Coated Perfectly Conducting Obstacle*, Waves2009, Mathematical and Numerical Aspects of Wave Propagation, Pau, june 2009.

[CA12] S. Pernet, B. Mallet, and X. Ferrieres and B. Pecqueux, *A local order spatial approximation combined with a local time stepping method in a Discontinuous Galerkin approach to solve time domain Maxwell's equations*, ACES, Monterey, USA, mars 2009.

[CA13] S. Pernet, D. Levadoux and F. Millot, *Formulations intégrales bien conditionnées en électromagnétisme*, NUMELEC, Liège, décembre 2008.

[CA14] S. Pernet and F. Millot, *An Inherently Well-Conditioned Integral Equation to Solve the Scattering Problems by Partially Coated Objects*, ACES, Niagara Falls, Canada, april 2008.

[CA15] E. Montseny, S. Pernet, X. Ferrieres, M. Zweers and G. Cohen, *A Discontinuous Galerkin Method to Solve Maxwell Equation in Time Domain*, The 23rd Annual Review of Progress in Applied Computational Electromagnetics, Verona, Italy, mars, 2007.

[CA16] G. Cohen, X. Ferrieres and S. Pernet, *Hexahedral Discontinuous Galerkin MethodsFor Maxwell's Equations in Time Domain*, WAVES2005, Brown university, 20-24 juin 2005.

[CA17] S. Pernet, X. Ferrieres, G. Cohen, *A hp-discontinuous Galerkin method to solve Maxwell's equations in time domain*, Proceedings of ECCOMAS 2004 (http://www.mit.jyu.fi/eccomas2004/-proceedings/proceed.html), 10 pages, 24-28 juillet 2004, Jyväskylä, Finlande.

[CA18] S. Pernet, X. Ferrieres, G. Cohen, *Méthode éléments finis d'ordre élevé pour résoudre les équations de Maxwell dans le domaine temporel-Comparaison avec des schémas de Yee et volumes finis*, Proceedings de Numélec'03, 4 ème Conférence Européenne sur les Méthodes Numériques en Electromagnétisme, 28-30 octobre 2003, Toulouse.

[CA19] G. Cohen, X. Ferrieres, P. Monk et S. Pernet, *Mass-Lumped Edge Elements for the Lossy Maxwell's Equations*, Proceedings of WAVES 2003, The sixth international conference on mathematical and numerical aspects of wave propagation, 30 juin-4 juillet 2003, Jyväskylä, Finlande, p. 383-388.

[CA20] G. Cohen, X. Ferrieres, P. Monk et S. Pernet, *Efficient mixed finite elements for the lossy Maxwell's equations in time domain*, Proceedings of IEEE2003, International Symposium on Electromagnetic Compatibility, 11-16 mai 2003, Istanbul, Turquie.

[CA21] S. Pernet, X. Ferrieres, G. Cohen, *An original finite element method to solve Maxwell's equations in time domain*, Proceedings of EMC Zurich 2003, 18-20 février 2003, Zurich, Suisse.

[CA22] S. Pernet, X. Ferrieres, G. Cohen, *Une méthode d'éléments finis pour la résolution des équations de Maxwell dans le domaine temporel. Comparaison en terme de temps CPU et de précision avec le schéma de Yee*, Proceedings du JINA 2002, 12-14 novembre 2002, Nice.

[CA23] S. Pernet, X. Ferrieres, G. Cohen, *Efficient mixed finite elements for Maxwell's equation in time domain*, in Proceedings of the AMEREM Conference 2002, 3-7 juin 2002, Annapolis, Etats-Unis.

[CA24] C. Bauer, G. Cohen, X. Ferrieres, P. Monk, P. Borderies et S. Pernet, *Comparison between a Finite Element Method and the Yee's Scheme to Solve Maxwell's Equations*, in Proceedings of JEE02, European Symposium on Numerical Methods in Electromagnetics, 6-8 mars 2002, Toulouse.

Conférences :

[C1] M. Bakry, S. Pernet, *Un nouvel indicateur d'erreur a posteriori pour la BEM en acoustique 2D et 3D*, CANUM, Obernai, 2016.

[C2] M. Bakry, S. Pernet, F. Collino, *A new a posteriori error estimate for the BEM in 3D-acoustics*, ECCOMAS, Hersonissos Grèce, 2016.

[C3] P.A. Hass, V. Mouysset, S. Pernet, *Adaptive Post-Processing Method to Represent High-Order Numerical Solutions*, ECCOMAS, Hersonissos Grèce, 2016.

[C4] Y. Dudouit, J-L Boelle, L. Giraud, F. Millot, S. Pernet, *High Performance Computing Using Local Time-Stepping Methods for Elastodynamics*, SIAM conference on Mathematics and Computational Issues in the Geosciences, Padua Italy, 2013.

[C5] S. Pernet, *Application du formalisme GCSIE à des problèmes hétérogènes en électromagnétisme*, Canum 2013, Seignosse le Penon, 2013.

[C6] E. Peynaud, A.-S. Bonnet-Ben Dhia, J.-F. Mercier, F. Millot, S. Pernet, *Propagation acoustique dans un fluide en écoulement. Couplage avec le transport hydrodynamique*, Canum 2012, Super-Besse Puy-de-Dôme, 2012.

[C7] A. Bendali, M'B. Fares, K. Lemrabet, Florence Millot, Sebastien Pernet, *A Combined Field Integral Equation for Higher-order Generalized Impedance Conditions*, PIERS 2011, 20-23 March, Marrakech, Morocco.

[C8] D. Levadoux, Florence Millot, S. Pernet, *Intrinsically well-conditioned integral equations for scattering by inhomogeneous bodies*, SCEE 2010, September 19-24, 2010.

[C9] D. Levadoux, Florence Millot, S. Pernet, *Comparison between the Classical Integral Equations and a Well Conditioned Integral Equation*, PIERS 2010 in Cambridge, USA, 5-8 July, 2010.

[C10] A.S. Bonnet-Ben Dhia, J.F. Mercier, F. Millot and S. Pernet, *A finite element method for time harmonic acoustics in arbitrary flows*, Acoustics'08, juillet 2008.

[C11] D. Levadoux, F. Millot and S. Pernet, *New trends in the preconditioning of integral equations of electromagnetism*, SCEE 2008, Helsinki, October 2008.

[C12] A.S. Bonnet-Ben Dhia, J.F. Mercier, F. Millot and S. Pernet, *A low mach model for time hramonic acoustics in arbitrary flows*, Waves2007, Mathematical and Numerical Aspects of Wave Propagation, Reading, 2007.

[C13] F. Collino, F. Millot and S. Pernet, *Boundary-Integral Methods for iterative solution of scattering problems with variable impedance surface condition*, congrés BETEQ Paris, du 4 au 6 septembre 2006.

[C14] Ferrières X., Mouysset V., Pernet S. et Alliot J.C., *Méthodes temporelles pour la résolution de Maxwell : stratégie multi-domaines/multi- méthodes*, CEM 2006, St Malo, du 4 au 6 avril 2006.

[C15] E. Bachelier, G. Cohen, X. Ferrieres, B. Pecqueux et S. Pernet, *Méthode de Galerkin Discontinue d'ordre élevé pour résoudre les équations de Maxwell en instationnaire*, CEM 2006, St Malo, du 4 au 6 avril 2006.

Rapports techniques et contractuels :

[R1] F. Collino, F. Millot, S. Pernet, *Boundary-Integral Methods for iterative solution of scattering problems with variable impedance surface condition*", Technical Report CERFACS, TR/EMC/07/18, 2007.

[R2] F. Millot, S. Pernet, *Rapport de synthèse sur la résolution des problèmes de diffraction d'ondes électromagnétiques avec condition d'impédance*, Technical Report CERFACS, TR/EMC/07/77, 2007.

[R3] F. Collino, M'B. Fares, F. Millot, S. Pernet, *Application de la GEFIE au calcul de diagramme de rayonnement d'un microsatellite* , Technical Report CERFACS, 2008.

[R4] F. Millot, S. Pernet, *Formulations en source et en champ généralisées pour le calcul de la diffraction d'une onde électromagnétique par un obstacle parfaitement conducteur* , Contract Report CERFACS, CR/EMC/08/118, 2008.

[R5] S. Pernet, *Résultats de Convergence and de stabilité d'une méthode de Galerkin discontinue pour des maillages non-conformes et une approximation polynomiale anisotropique*, Technical Report CERFACS, TR/EMC/08/107, 2008.

[R6] S. Pernet, *Rapport d'avancement T0 +6 pour la tâche 2 du projet DIGATOP: Mise au point d'un estimateur a posteriori pour FEMGD*, Contract Report CERFACS, CR/EMA/09/134, 2009.

[R7] S. Pernet, *Rapport d'avancement T0 +12 pour la tâche 2 du projet DIGATOP: Mise au point d'un estimateur a posteriori pour FEMGD*, Contract Report CERFACS, 2010.

[R8] A. Bendali, F. Collino, S. Pernet, *Méthodes numériques pour le calcul de la diffraction par une cavité revêtue*, Contract Report CERFACS, CR-EMA-10-135, 2010.

[R9] A.S Bonnet, F. Millot, S. Pernet et E. Peynaud, *Analysis of a discontinuous galerkin method for a time harmonic convection equation*, Techinical Report, TR/EMC/11/132, 2011.

[R10] F. Millot et S. Pernet, *Analyse des performances d'une formulation intégrale de type GSIE pour des problèmes de diffractions avec condititions d'impédance*, Contract Report, CR/EMC/11/131, 2011.

[R11] V. Mouysset, S. Pernet, *Rapport d'avancement 2015 - PR PREVISIO*, Technical Report, RA 1/23878 DTIM, 2016.

[R12] M. Alexandre, V. Mouysset, S. Pernet, S. Mefire, *ANR Continus Tâche 2.2 : Modèle direct de mesure de résistivité. Rapport d'avancement*, Technical Report ONERA, RT 2/22623 DTIM, 2016.

Développement de codes de calcul

Durant ma thèse, j'ai écrit entièrement deux solveurs permettant de résoudre les équations de Maxwell dans un milieu 3D hétérogène. Le premier est basé sur une méthode d'éléments finis d'ordre élevé construits sur des maillages hexaédriques non structurés et utilisant un formalisme PML pour borner le domaine de calcul. Le second est quant à lui basé sur une approche de type Galerkin discontinue d'ordre élevé construite sur des maillages hexaédriques non structuré, incluant une stratégie de pas de temps local et un formalisme PML. Ce solveur est le noyau qui a permis la construction du code FEMGD utilisé et développé depuis par l'ONERA et qui a été livré à la DGA.

Au CERFACS, j'ai participé au développement du code CESC (Cerfacs Electromagnetic Solver Code) en l'enrichissant par exemple des méthodes qui sont présentées dans la partie scientifique de ce document. CESC est un code généraliste massivement parallèle permettant de modéliser de nombreuses situations en électromagnétisme via des formalismes intégraux en régime harmonique. En particulier, les résolutions itératives sont basées sur le couplage de solveurs GMRES (GMRES, GMRES-DR et flexible GMRES) et de méthodes de calcul multipôle rapide multi-niveaux permettant ainsi une résolution très rapide de grands systèmes linéaires. J'ai participé aussi activement au développement d'un code pour le calcul de l'acoustique en écoulement basé sur l'approximation de l'équation de Galbrun.

Enfin, je participe actuellement au développement d'outil C++ permettant une exploitation rigoureuse des résulats numériques issus des codes de haute précision (méthodes d'ordre élevé par exemple) par visualisation en utilisation des logiciels standards (gmsh, tecplot, paraview ... etc ...).

Activités d'enseigement

- ISAE-SUPAERO :
    - Cours d'EDP (Théorie des systèmes hyperboliques (linéaires et non-linéaires), méthodes de volumes finis, théorie des équations elliptiques, méthodes des éléments finis) en 2ème anné : 2015/2016, 2016/2017 avec un volume horaire de 12h de cours/an et 8h de TD/an.
    - Analyse fonctionnelle et harmonique en première année: 2015/2016, 2016/2017 avec un volume horaire de 20h de TD/an.

- ISAE-ENSICA :
  - Analyse fonctionnelle (Théorie de la mesure et de l'intégration, théorie des distributions, analyse complexe, espaces de Hilbert) en 1ere année Ensica : 2008/2009, 2009/2010, 2010/2011, 2011/2012, 2012/2013, 2013/2014 et 2014/2015 avec un volume horaire de 30h de cours/an et 10h de TD/an.
  - Résolution numérique des EDP (Formulations variationnelles, éléments finis, différences finies, synthèse modale, stabilité, convergence) en 1ere année : 2006/2007, 2007/2008, 2008/2009, 2009/2010, 2010/2011 avec un volume horaire de 18.75h de cours/an et 8.75h de TD/an.
  - Analyse Matricielle et Optimisation en 1ere année : 2006/2007 et 2007/2008 avec un volume horaire de 17.5h de cours/an et 7.5h de TD/an.
- ENAC 1ère année, Analyse (mesure et intégration, analyse complexe) : 2011/2012 avec un volume horaire de 30h de cours et 10h de TD.
- INSA 5ème année option GMM, Equations intégrales en électromagnétisme : 2009/2010 avec un volume horaire de 12.5h de cours.

Collaborateurs

A. Bendali (INSA), A.S. Bonnet (CNRS), G. Cohen (INRIA) , F. Collino, X. Ferrieres (ONERA), L. Giraud (INRIA) , D. Levadoux (ONERA), J.F. Mercier (CNRS), F. Millot (CERFACS) et V. Mouysset (ONERA)

Participation à projets

- PRF COFEBELF (Couplage FEm-BEm pour l'ELectromagnétisme dans le domaine Fréquentiel): Projet de Recherche Fédérateur ONERA, 01/2017-12/2020, coordinateur : F.X. Roux.
- ANR CONTINUS : Contrôle non destructif et Inversion Numérique pour la Surveillance des structures de grandes dimensions (ARKOGEOS, ONERA, EDF, IECL, LMDC) octobre 2014 (54 mois), coordinateur : Gilles Klysz (INSA).
- PR PREVISIO (P1 Representation for Enhanced Visualization by Interpolating Solution of Increased Order) : Projet de Recherche ONERA, 01/2015-12/2017, coordinateur : S. Pernet.
- ANR AEROSON projet blanc : Simulation numérique du rayonnement sonore dans des géométries complexes en présence d'écoulements réalistes (EADS-IW, CERFACS, Laboratoire d'Acoustique de l'Université du Maine), 01/01/2009 - 02/28/2013, coordinateur : Jean-François Mercier (CNRS).
- ANR ARTHEMIS projet MN : Méthode de décomposition de domaine et solveurs bien conditionnés en électromagnétisme : une réponse aux enjeux industriels actuels (ONERA, CERFACS, Ecole Polytechnique), novembre 2011 - octobre 2015, coordinateur : David Levaoux (ONERA).

- REI DIGATOP : DIscontinuous GAlerkin Time OPtimization (NUCLETUDES, ONERA, CERFACS, XLIM, INRIA), 2008-2011.

- ANR RAFFINE projet MN : Robustesse, Automatisation et Fiabilité des Formulations INtégrales en propagation d'ondes : Estimateurs a posteriori et adaptivité (CERFACS, EADS, IMACS, ONERA, Thales), janvier 2013 - septembre 2017, coordinateur : Marc Bonnet (CNRS).

- PRF MAHPSO : Projet de Recherche Fédérateur ONERA sur les modèle d'approximation de haute précision pour les systèmes d'équations de propagation d'onde, 2006 - 2010, coordinateur : Francois Rogier (ONERA).

De plus, l'environnement du CERFACS m'a permis d'avoir des relations étroites avec de grands groupes industriels sous forme de contrat. En particulier, j'ai réalisé des nombreuses études amont pour le compte de Dassault, EADS-MBDA, CNES, ONERA et CEA-DAM.

BIBLIOGRAPHIE

[ABL07] F. Alouges, S. Borel & D. Levadoux – "A stable well conditioned integral equation for electromagnetism scattering", *J. Comput. Appl. Math.* **204** (2007), no. 2, p. 440–451.

[AD07] X. Antoine & M. Darbas – "Generalized combined field integral equations for the iterative solution of the three-dimensional helmholtz equation", *Mathematical Modelling and Numerical Analysis* **41** (2007), p. 147–167.

[ANE81] R. Astley, N.Walkington & W. Eversman – "Accuracy and stability of finite element schemes for the duct transmission problem", *In AIAA, Astrodynamics Specialist Conference* **1** (1981).

[Ast09] R.-J. Astley – "Numerical methods for noise propagation in moving flows, with application to turbofan engines", *Acoustical science and technology* **30** (2009), no. 4, p. 227–239.

[Aze96] P. Azerad – "Analyse des équations de navier-stokes en bassin peu profond et de l'équation de transort", Thèse, Université de Neuchâtel, 1996.

[Bak16] M. Bakry – "Fiabilité et optimisation des calculs obtenus par des formulations intégrales en propagation d'ondes", Thèse, université Paris-Saclay, 2016.

[BBL06] E. Bécache, A.-S. Bonnet & G. Legendre – "Perfectly matched layers for time harmonic acoustics in the presence of a uniform flow", *SIAM J. Numer. Anal.* **44** (2006), p. 1191–1217.

[BC07] A. Buffa & S.-H. Christiansen – "A dual finite element complex on the barycentric refinement", *Math. Comp.* **76** (2007), p. 1743–176.

[BCD+09] D. Boffi, M. Costabel, M. Dauge, L. Demkowicz & R. Hiptmair – "Discrete compactness for the p-version of discrete differential forms", *SIAM J. Numer. Anal.* **49** (2009), no. 1, p. 135–158.

[BDLM07] A.-S. Bonnet, E.-M. Duclairoir, G. Legendre & J.-F. Mercier – "Acoustic propagation in a flow: numerical simulation of the time-harmonic regime", *J. of Comp. and Appl. Math.* **204** (2007), no. 2, p. 428–439.

[BDM07] A.-S. Bonnet, E.-M. Duclairoir & J.-F. Mercier – "Acoustic propagation in a flow: numerical simulation of the time-harmonic regime", *ESAIM Proceedings* **22** (2007).

[BFG99] A. Bendali, M.-B. Fares & J. Gay – "A boundary-element solution of the leontovitch problem", *IEEE Transaction on Antennas and Propagation* **47** (1999), no. 10, p. 1597–1605.

[BFLP09] A. Bendali, M. Fares, K. Lemrabet & S. Pernet – "Recent developments in the scattering of an electromagnetic wave by a coated perfectly conducting obstacle", *Waves2009, Mathematical and Numerical Aspects of Wave Propagation, Pau* (2009).

[BFP99] D. Boffi, P. Fernandez & I. Perugia – "Computational models of electromagnetic resonators: Analysis of edge element approximation", *SIAM J. Numer. Anal.* **36** (1999), no. 4, p. 1264–1290.

[BG02] D. Boffi & L. Gastaldi – "Edge finite elements for the approximation of maxwell resolvent operator", *RAIRO - Math. Model. Numer. Anal.* **36** (2002), no. 2, p. 293–305.

[BH03] A. Buffa & R. Hiptmair – "Galerkin boundary element methods for electromagnetic scattering", In Topics in computational wave propagation (S. B. Heidelberg, éd.), Springer Berlin Heidelberg, 2003, p. 85–126.

[BHHW00] R. Beck, R. Hiptmair, R.-H.-W. Hoppe & B. Wohlmuth – "Residual based a posteriori error estimators for eddy current computation", *RAIRO* **34** (2000), no. 1, p. 159–182.

[BJR05] E. Bécache, P. Joly & J. Rodriguez – "Space-time mesh refinement for elastodynamics. numerical results", *Comput. Methods Appl. Mech. Engrg.* **194** (2005), p. 355–366.

[BL08] A. Bendali & K. Lemrabet – "Asymptotic analysis of the scattering of a time-harmonic wave by a perfectly conducting metal coated with a thin dielectric shell", *Asymptotic Analysis* **57** (2008), p. 199–227.

[BLL01] A.-S. Bonnet, G. Legendre & E. Lunéville – "Analyse mathématique de l'équation de galbrun en écoulement uniforme", *Comptes Rendus de l'Académie des Sciences, Series IIB-Mechanics* **329** (2001), no. 8, p. 601–606.

[BMM+12] A.-S. Bonnet, J.-F. Mercier, F. Millot, S. Pernet & E. Peynaud – "Time-harmonic acoustic scattering in a complex flow: a full coupling between acoustics and hydrodynamics", *CICP* **11** (2012), no. 2, p. 555–572.

[BMMP10] A.-S. Bonnet, J.-F. Mercier, F. Millot & S. Pernet – "A low mach model for time harmonic acoustics in arbitrary flows", *J. of Comp. and Appl. Math.* **234** (2010), no. 6, p. 1868–1875.

[BMPP11] A.-S. Bonnet, F. Millot, S. Pernet & E. Peynaud – "Analysis of a discontinuous galerkin method for a time harmonic convection equation", Tech. Report TR/EMC/11/132, CERFACS, 2011.

[Bof00] D. Boffi – "Fortin operator and discrete compactness for edge elements", *Numer. Math.* **87** (2000), p. 229–246.

[Bof01] _____, "A note on the de rham complex and a discrete compactness property", *Appl. Math. Lett.* **14** (2001), p. 33–38.

[BPC78] M. Bakry, S. Pernet & F. Collino – "A new accurate residual-based a posteriori error indicator for bem in 2d-acoustics", *acceptée pour publication dans Computers and Mathematics with Applications* (1978).

[BPC17] ———, "Reliable and efficient a posteriori error estimate for efie in electromagnetism", *Proc. of WAVES2017, Minneapolis, May 15-19* (2017).

[Bre10] H. Brezis – *Functional analysis, sobolev spaces and pdes*, Springer, 2010.

[BS08] D. Braess & J. Schöberl – "Equilibrated residual error estimator for edge elements", *Math. Comp.* **77** (2008), p. 651–672.

[CD07] S. Cochez-Dhondt – "Méthodes d'éléments finis et estimations d'erreur a posteriori", Thèse, université de Valenciennes et Hainaut-Cambrésis, 2007.

[CDN07] S. Cochez-Dhondt & S. Nicaise – "Robust a posteriori error estimation for the maxwell equations", *Methods Appl. Mech. Engrg.* **196** (2007), p. 2583–2595.

[CDR03] W. Cecot, L. Demkowicz & W. Rachowicz – "Three dimensional infinite element for maxwell's equations", *International Journal for Numerical Methods in Engineering* **57** (2003), p. 899–921.

[CF00] G. Cohen & S. Fauqueux – "Mixed finite elements with mass-lumping for the transient wave equation", *J. Comput. Acoust.* **8** (2000), no. 1, p. 171–188.

[CF05] ———, "Mixed spectral finite elements for the linear elasticity system in unbounded domains", *SIAM J. Sci. Comput.* **26** (2005), no. 3, p. 864–884.

[CF15] S. Chaillat & F.Collino – "A wideband fast multipole method for the helmholtz kernel : Theoretical developments", *Computers and Mathematics with Applications* **70** (2015), p. 660–678.

[CFJ06] F. Collino, T. Fouquet & P. Joly – "Conservative space-time mesh refinement methods for the fdtd solution of maxwell's equations", *J. Comput. Phys.* **211** (2006), no. 1, p. 9–35.

[CFP06] G. Cohen, X. Ferrieres & S. Pernet – "A spatial high-order hexahedral discontinuous galerkin method to solve maxwell's equations in time domain", *J. Comput. Phys.* **217** (2006), no. 2, p. 340–363.

[CFR00] S. Caorsi, P. Fernandes & M. Raffetto – "On the convergence of galerkin finite element approximations of electromagnetic eigenproblems", *SIAM J. Numer. Anal.* **38** (2000), no. 2, p. 580–607.

[CJT94] G. Cohen, P. Joly & N. Tordjman – "Higher-order finite elements with mass lumping for the 1-D wave equation", *Finite Elem. Anal. Des.* **17** (1994), no. 3-4, p. 329–336.

[CK98] D. Colton & R. Kress – *Inverse acoustic and electromagnetic scattering theory*, Springer, 1998.

[CM99] G. Cohen & P. Monk – "Mur-nédélec finite element schemes for maxwell's equations", *Comput. Methods Appl. Mech. Engrg.* **169** (1999), no. 3-4, p. 197–217.

[CMP08] F. Collino, F. Millot & S. Pernet – "Boundary-integral methods for iterative solution of scattering problems with variable impedance surface condition", *PIER* **80** (2008), p. 1–28.

[CN02a] S. Christiansen & J.-C. Nédélec – "A preconditioner for the electric field integral equation based on calderon formulas", *SIAM J. Numer. Anal.* **40** (2002), no. 3, p. 1100–1135.

[CN02b] S.-H. Christiansen & J.-C. Nédélec – "A preconditioner for the electric field integral equation based on calderon formulas", *SIAM J. Numer. Anal.* **40** (2002), no. 3, p. 1100–1135.

[CP16] G. Cohen & S. Pernet – *Finite element and discontinuous galerkin methods for transient wave equations*, Scientific Computation, Springer-verlag, 2016.

[CWZ07] Z. Chen, L. Wang & W. Zheng – "An adaptive multilevel method for time-harmonic maxwell equations with singularities", *SIAM J. SCI. COMPUT.* **29** (2007), no. 1, p. 118–138.

[Dar06] M. Darbas – "Generalized cfie for the iterative solution of 3-d maxwell equations", *Applied Mathematics Letters* **19** (2006), no. 8, p. 834–839.

[DG09] J. Diaz & M.-J. Grote – "Energy conserving explicit local time stepping for second-order wave equations", *SIAM J. Sci. Comput.* **31** (2009), no. 3, p. 1985–2014.

[DGMP16] Y. Dudouit, L. Giraud, F. Millot & S. Pernet – "Interior penalty discontinuous galerkin method for coupled elasto-acoustic media", *https://dumas.ccsd.cnrs.fr/UNIV-BORDEAUX/hal-01406158v1* (2016).

[Djo06] J. Djokic – "Efficient update of hierarchical matrices in the case of adaptive discretisation schemes", Thèse, University of Leipzig, 2006.

[DKT07] M. Dumbser, M. Käser & E.-F. Toro – "An arbitrary high-order discontinuous galerkin method for elastic waves on unstructured meshes - v. local time stepping and p-adaptivity", *Geophys. J. Int.* **171** (2007), no. 2, p. 695–717.

[DNR78a] J. Descloux, N. Nassif & J. Rappaz – "On spectral approximation part 1. the problem of convergence", *RAIRO Numer. Anal.* **12** (1978), p. 97–112.

[DNR78b] ———, "On spectral approximation part 2. error estimates for the galerkin method convergence", *RAIRO Numer. Anal.* **12** (1978), p. 113–119.

[Duc07] E.-M. Duclairoir – "Rayonnement acoustique dans un écoulement cisaillé : Une méthode d'éléments finis pour la simulation du régime harmonique", Thèse, Ecole Doctorale de l'Ecole Polytechnique, 2007.

[Dud14] Y. Dudouit – "Spatio-temporal refinement using a discontinuous galerkin approach for elastodynamic in a high performance computing framework", Thèse, Université de Bordeaux, 2014.

[Dup05] S. Duprey – "Analyse mathématique et numérique du rayonnement acoustique des turboréacteurs", Thèse, Institut Elie Cartan-Université Poincaré Nancy, 2005.

[Dur06] M. Duruflé – "Integration numérique et éléments finis d'ordre élevé appliqués aux équations de maxwell en régime harmonique", Thèse, Université de Paris-Dauphine, 2006.

[EG04] A. Ern & J.-L. Guermond – *Theory and practice of finite elements*, Applied Mathematical Sciences, Springer, New-York, 2004.

[EJ97] A. Elmkies & P. Joly – "éléments finis d'arête et condensation de masse pour les équations de maxwell: le cas de dimension 3.", *C. R. Acad. Sci. Paris SÃ©r. I Math.* **325** (1997), no. 11, p. 1217–1222.

[EJ09] A. Ezziani & P. Joly – "Space-time mesh refinement for discontunuous galerkin methods for symmetric hyperbolic systems", *J. Comput. Appl. Math.* **234** (2009), no. 6, p. 1886–1895.

[EN93] B. Engquist & J.-C. Nédélec – "Effective boundary conditions for acoustic and electro-magnetics scattering in thin layers", Tech. Report research report CMAP 278, Ecole Polytechnique, 1993.

[Gab07] G. Gabard – "Discontinuous galerkin methods with plane waves for time-harmonic problems", *J. Comp. Phys.* **225** (2007), no. 2, p. 1961–1984.

[Gal31] H. Galbrun – "Propagation d'une onde sonore dans l'atmosphÃ¨re et théorie des zones de silence", *Gauthier-Villars, Paris* (1931).

[GAT05] G. Gabard, R. Astley & M. B. Tahar – "Stability and accuracy of finite element methods for flow acoustics. ii : Two-dimensional effects", *International journal for numerical methods in engineering* **63** (2005), no. 7, p. 974–987.

[GMMM10] F. Gesztesy, I. Mitrea, D. Mitrea & M. Mitrea – "On the nature of the laplace-beltrami operator on lipschitz manifolds", *Journal of Mathematical Sciences* **172** (2010), no. 3, p. 279–346.

[Hac15] W. Hackbusch – *Hierarchical matrices : algorithms and analysis*, computational mathematics, Springer, 2015.

[Hes98] J.-S. Hesthaven – "From electrostatics to almost optimal nodal sets for polynomial interpolation in a simplex", *SIAM J. Numer. Anal.* **35** (1998), no. 2, p. 655–676.

[HHT08] N. Hale, N.-J. Higham & L.-N. Trefethen – "Computing a^{α} , $log(a)$, and related matrix functions by contour integrals", *SIAM Journal on Numerical Analysis* **46** (2008), no. 5, p. 2505–2523.

[HJ02] H. Haddar & P. Joly – "Stability of thin layer approximation of electromagnetic waves scattering by linear and nonlinear coatings", *Journal of Computational and Applied Mathematics* **143** (2002), no. 2, p. 201–236.

[HMP16] P.-A. Hass, V. Mouysset & S. Pernet – "Adaptive post-processing method to represent high-order numerical solutions", *ECCOMAS, Hersonissos Grèce* (2016).

[Hof00] J. Hoffman – "Adaptive finite element methods for the unsteady maxwell's equations", *Chalmers Finite Element Center, PREPRINT 2000-001l* (2000).

[HPS05] P. Houston, I. Perugia & D. Schötzau – "Energy norm a posteriori error estimation for mixed discontinuous galerkin approximations of the maxwell operator", *Comput. Meth. Appl. Mech. Engrg.* **194** (2005), p. 499–510.

[HPS07] ———, "An a posteriori error indicator for discontinuous galerkin discretizations of h(curl)-elliptic partial differential equations", *IMA J. Numer. Anal.* **27** (2007), p. 122–150.

[HSV79] J.-P. Hennart, E. Sainz & M. Villegas – "On the efficient use of the finite element method in static neutron diffusion calculations", *Computational Methods in Nuclear Engineering* **1** (1979), p. 3–87.

[HT00] J.-S. Hesthaven & C.-H. Teng – "Stable spectral methods on tetrahedral elements", *SIAM J. Sci. Comp.* **21** (2000), no. 6, p. 2352–2380.

[HW08] J.-S. Hesthaven & T. Warburton – *Nodal discontinuous galerkin methods: Algorithms, analysis, and applications*, 1st éd., Springer Publishing Company, Incorporated, 2008.

[IAS10] A. Iob, R. Arina, & C. Schipani – "Frequency-domain linearized euler model for turbomachinery noise radiation through engine exhaust", *AIAA journal* **48** (2010), no. 4, p. 848–858.

[IHvdV08] F. Izsák, D. Harutyunyan & J. van der Vegt – "Implicit a posteriori error estimates for the maxwell equations", *Math. Comp.* **77** (2008), p. 1355–1386.

[Jen06] M. Jensen – "Remarks on duality in graph spaces of first-order linear operators", *PAMM-Proceedings in Applied Mathematics and Mechanics* **6** (2006), no. 1, p. 31–34.

[JSC02] B. Jung, T. Sarkar & Y. Chung – "A survey of various frequency domain integral equations for the analysis of scattering from three-dimensional dielectric objects", *PIER* **36** (2002), p. 193–246.

[JSV10] P. Jiránez, Z. Strakos & M. Vohralík – "A posteriori error estimates including algebraic error and stopping criteria for iterative solvers", *SIAM J. Sci. Comput.* **32** (2010), p. 1567–1590.

[Lan95] V. Lange – "Equations intéegrales espace-temps pour les équations de maxwell. calcul du champ diffracté par un obstacle dissipatif", Thèse, Universitée de Bordeaux, 1995.

[Lau13] J.-B. Laurent – "Raffinements locaux auto-adaptatifs dans une méthode galerkin discontinu pour la résolution des équations de maxwell", Thèse, Université de Toulouse, 2013.

[Leg03] G. Legendre – "Rayonnement acoustique dans un fluide en écoulement : analyse mathématique et numérique de l'équation de galbrun", Thèse, Université Paris VI, 2003.

[Leo78] M.-A. Leontovitch – "Approximate boundary conditions for the electromagnetic field on the surface of a good conductor", *Investigations Radiowave Propagation Part II, Academy of Sciences, Moscow* (1978).

[Li09] J. Li – "A posteriori error estimation for an interior penalty discontinuous galerkin method for maxwell's equations in cold plasma", *Adv. Appl. Math. Mech.* **1** (2009), no. 1, p. 107–124.

[LMP14] D. Levadoux, F. Millot & S. Pernet – "An unpreconditioned boundary-integral for iterative solution of scattering problems with non-constant leontovitch impedance boundary conditions", *CICP* **15** (2014), no. 5, p. 1431–1460.

[LMP15] ———, "A well-conditioned boundary integral equation for transmission problems of electromagnetism", *J. Integral Equations Applications* **27** (2015), no. 3, p. 431–454.

[MFP+13] B. Mallet, X. Ferrieres, S. Pernet, J.-B. Laurent, B. Pecqueux & P. Seimandi – "A p-strategy with a time-stepping method in a discontinuous galerkin approach to solve electromagnetic problems", *Journal of Computational Methods in Physics* **2013** (2013), p. 1–13.

[MN03] C. Makridakis & R. Nochetto – "Elliptic reconstruction and a posteriori error estimates for parabolic problems", *SIAM J. Numer. Anal.* **41** (2003), no. 4, p. 1585–1594.

[Mon98] P. Monk – "A posteriori error indicators for maxwell's equations", *J. Comp. Appl. Math.* **100** (1998), p. 173–190.

[MP89] Y. Maday & A.-T. Patera – "Spectral element methods for the imcompressible navier-stokes equations", *State of the Art Survey in Computational Mechanics, ed. A. K. Noor* (1989), p. 71–143.

[MPFC08] E. Montseny, S. Pernet, X. Ferrières & G. Cohen – "Dissipative terms and local time-stepping improvements in a spatial high order discontinuous galerkin scheme for the time-domain maxwell's equations", *J. Comput. Phys.* **227** (2008), no. 14, p. 6795–6820.

[MZB97] F.-A. Milinazzo, C.-A. Zala & G.-H. Brooke – "Rational square-root approximations for parabolic equation algorithms", *J. Acoust. Soc. Am.* **101** (1997), no. 2, p. 760–766.

[Néd86] J.-C. Nédélec – "A new family of mixed finite elements in $\mathbb{R}^3$", *Numer.Math.* **50** (1986), no. 1, p. 57–81.

[NTPD11] B. Nennig, M. B. Tahar & E. Perrey-Debain – "On the acoustic boundary condition in the presence of flow", *J. Acoust. Soc. Am.* **130** (2011), no. 1, p. 42–51.

[OL08] Y. Özyörük & S. Lidoine – "Numerical analysis of noise radiation from a turbofan exhaust cowl with an extended liner in flight", *14th AIAA/CEAS Aeroacoustics Conference (29th AIAA Aeroacoustics Conference)* (2008).

[PE01] C. Peyret & G. Élias – "Finite-element method to study harmonic aeroacoustics problems", *J. Acoust. Soc. Amer.* **110** (2001), no. 2, p. 661–668.

[Per08] S. Pernet – "Résultats de convergence et de stabilité d'une méthode de galerkin discontinue pour des maillages non-conformes et une approximation polynomiale anisotropique", Tech. Report TR/EMC/08/107, CERFACS, 2008.

[Per09] ———, "Rapport d'avancement t0+6 pour la tâche 2 du projet digatop: Mise au point d'un estimateur a posteriori pour femgd", Tech. Report CR/EMA/09/134, CERFACS, 2009.

[Per10a] ———, "Rapport d'avancement t0+12 pour la tâche 2 du projet digatop: Mise au point d'un estimateur a posteriori pour femgd", Tech. report, CERFACS, 2010.

[Per10b] ———, "A well-conditioned integral equation for iterative solution of scattering problems with a variable leontovitch boundary condition", *Mathematical Modelling and Numerical Analysis* **44** (2010), no. 4, p. 781–801.

[Pey13] E. Peynaud – "modélisation de la propagation du son en écoulements via l'équation de galbrun", Thèse, Université de Toulouse, 2013.

[PF07] S. Pernet & X. Ferrieres – "hp a-priori error estimates for a non-dissipative spectral discontinuous galerkin method to solve the maxwell equations in the time domain", *Mathematics of Computation* **76** (2007), no. 260, p. 1801–1832.

[PFC05] S. Pernet, X. Ferrieres & G. Cohen – "High order finite element method to solve maxwell's equations in time domain", *IEEE Trans. on Antennas and Propagation* **53** (2005), no. 9, p. 2889–2899.

[Pip06] S. Piperno – "Symplectic local time-stepping in non-dissipative dgtd methods applied to wave propagation problems", *ESAIM Math. Model. Numer. Anal.* **40** (2006), no. 5, p. 815–841.

[PMFP09] S. Pernet, B. Mallet, X. Ferrieres & B. Pecqueux – "A local order spatial approximation combined with a local time stepping method in a discontinuous galerkin approach to solve time domain maxwell's equations", *ACES, Monterey, USA* (2009).

[Poi85] B. Poirée – "Les équations de l'acoustique linéaire et non-linéaire dans les fluides en mouvement", *Acoustica* **57** (1985), p. 5–25.

[RD00] W. Rachowicz & L. Demkowicz – "An hp-adaptive finite element method for electromagnetics. part 1: Data structure and constrained approximation", *Computer Methods in Applied Mechanics and Engineering* **187** (2000), no. 1-2, p. 307–337.

[RD02] ———, "An hp-adaptive finite element method for electromagnetics. part 2: A 3d implementation", *International Journal for Numerical Methods in Engineering* **53** (2002), p. 147–180.

[Rep07] S. Repin – "Functional a posteriori estimates for maxwell's equation", *Journal of Mathematical Sciences* **142** (2007), no. 1, p. 1821–1827.

[RH73] W. Reed & T. Hill – "Triangular mesh methods for the neutron transport equation", Tech. Report LA-UR-73-479, Los Alamos Scientific Laboratory, 1973.

[RM06] P.-P. Rao & P. J. Morris – "Use of finite element methods in frequency domain aeroacoustics", *AIAA journal* **44** (2006), no. 7, p. 1643–1652.

[Rod08] J. Rodriguez – "A spurious-free space-time mesh refinement for elastodynamics", *Int. J Multiscale Comput. Eng.* **6** (2008), no. 3, p. 263–279.

[Sch08] J. Schöberl – "A posteriori error estimates for maxwell equations", *Math. Comp.* **77** (2008), p. 633–649.

[SCKR08] M. Steffen, S. Curtis, R.-M. Kirby & J.-K. Ryan – "Investigation of smoothness-increasing accuracy-conserving filters for improving streamline integration through discontinuous fields", *IEEE Transactions on Visualization and Computer Graphics* **14** (2008), no. 3, p. 680–692.

[SDCD08] P. Solin, L. Dubcova, J. Cerveny & I. Dolezel – "Adaptive hp-fem with arbitrary-level hanging nodes for maxwell's equations", *Research Report No. 2008-01, The University of Texas at El Paso* (2008).

[SW98] O. Steinbach & W.-L. Wendland – "The construction of some efficient preconditioners in the boundary element method", *Adv. Comput. Math.* **9** (1998), no. 1-2, p. 191–216.

[TGT03] F. Treyssède, G. Gabard & M. B. Tahar – "A mixed finite element method for acoustic wave propagation in moving fluids based on an eulerian-lagrangian description", *J. Acoust. Soc. Am.* **113** (2003), no. 2, p. 705–716.

[Yee66] K. Yee – "Numerical solutions of initial boundary value problems involving maxwell's equations in isotropic media", *IEEE Trans. Antennas Propag.* **14** (1966), no. 3, p. 302–307.

[You78] L.-C. Young – "An efficient finite element method for reservoir simulation", *Proc. of the 53rd Annual Fall Technical Conference and Exhibition of the Society of Petroleum Engineers of AIME, Houston, Texas Oct. 1-3* (1978).

[ZCL06] W. Zheng, Z. Chen & W. L – "An adaptive finite element method for the $h - \psi$ formulation of time-dependent eddy current problems", *Numerische Mathematik* **103** (2006), no. 4, p. 667–689.

PARTIE III

ANNEXES : 5 PUBLICATIONS

A NEW ACCURATE RESIDUAL–BASED A POSTERIORI ERROR INDICATOR FOR THE BEM IN 2D-ACOUSTICS

Marc Bakry[a,*], Sébastien Pernet[b], Francis Collino[c]

[a] *INRIA Saclay – Ile de France, bâtiment Alan Turing, 1 rue Honoré d'Estienne d'Orves, Campus de l'Ecole Polytechnique, FR-91120 Palaiseau*
[b] *Office National d'Etudes et de Recherches Aérospatiales (ONERA), Département Traitement de l'Information et Modélisation (DTIM), 2 avenue Edouard Belin, FR-31055 Toulouse Cedex 4*
[c] *CERFACS, 42 avenue Gaspard Coriolis, FR-31057 Toulouse Cedex 1*

Abstract

In this work we construct a new reliable, efficient and local *a posteriori* error estimate for the single layer and hyper-singular boundary integral equations associated to the Helmholtz equation in two dimensions. It uses a localization technique based on a generic operator Λ which is used to transport the residual into L^2. Under appropriate conditions on the construction of Λ, we show that it is asymptotically exact with respect to the energy norm of the error. The single layer equation and the hyper-singular equation are treated separately. While the current analysis requires the boundary to be smooth, numerical experiments show that the new error estimators also perform well for non-smooth boundaries.

1. Introduction

The Boundary Element Method (BEM) is a method, based on boundary integral formulations, that can be used for the resolution of wave propagation problems. It features strong advantages since only the boundary Γ of the domain is meshed, the radiation condition at infinity is intrinsically taken into account and it is more accurate than other common methods like the Finite Element Method (FEM). The main disadvantages of the BEM are the difficult implementation (singular integrals) and the manipulation of fully populated matrices. This last problem has been partially bypassed thanks to recent improvements on the acceleration of the BEM like the Fast Multipole Method [18] or $\mathcal{H}$-matrices [17].

In this paper we study the propagation of an acoustic wave with wave number k in an infinite medium. This wave is diffracted by a scatterer represented by a simply-connected bounded Lipschitz domain with boundary Γ. In the following,

*Corresponding author

a discretization of this boundary will be named $\mathcal{T}_h$. We apply either a Dirichlet boundary condition for the global field u such that $u|_\Gamma = 0$ or a Neumann boundary condition $\left.\dfrac{\partial u}{\partial n}\right|_\Gamma = 0$ where n is the outward pointing normal of the scatterer. The two integral equations we are going to study are

$$(\mathcal{S}_k\varphi)(x) := \int_\Gamma G_k(x,y)\,\varphi(y)\,d\gamma_y = -u_i(x), \tag{1}$$

which solves the scattering problem with Dirichlet boundary condition and

$$(\mathcal{N}_k\psi)(x) := \text{f.p.}\int_\Gamma \frac{\partial^2 G_k}{\partial n_x \partial n_y}(x,y)\,\psi(y)\,d\gamma_y = -\partial_n u_i(x), \tag{2}$$

which solves the scattering problem with a Neumann boundary condition for an incoming wave u_i and is given as a finite parts integral. The kernel G_k is the Green function associated to the Helmholtz equation. For 2D wave propagation problems, it reads

$$G_k(x,y) = \frac{i}{4}H_0^{(1)}(k|x-y|),$$

with $H_0^{(1)}$ the Hankel function of the first kind and of order 0.

The operator $\mathcal{S}_k$ is named *single layer integral operator* and the operator $\mathcal{N}_k$ is commonly named *hyper–singular operator*. These equations are solved using a Galerkin method. The variational formulations are recalled (see [22]) below:

- equation (1), find $\varphi \in H^{-1/2}(\Gamma)$ such that for all $\varrho \in H^{-1/2}(\Gamma)$,

$$\int_\Gamma\int_\Gamma \varrho(x)\,G_k(x,y)\,\varphi(y)\,d\gamma_x\,d\gamma_y = -\int_\Gamma \varrho(x)\,u_i(x)\,d\gamma_x, \tag{3}$$

- equation (2), find $\psi \in H^{1/2}(\Gamma)$ such that for all $\varsigma \in H^{1/2}(\Gamma)$,

$$\begin{aligned}&\int_\Gamma\int_\Gamma \overrightarrow{\text{rot}}_\Gamma\varsigma(x)\,G_k(x,y)\,\overrightarrow{\text{rot}}_\Gamma\psi(y)\,d\gamma_x\,d\gamma_y - \\ &k^2\int_\Gamma\int_\Gamma \varsigma(x)\,G_k(x,y)\,\psi(y)\,n_x\cdot n_y\,d\gamma_x\,d\gamma_y = -\int_\Gamma \varsigma(x)\,\partial_n u_i(x)\,d\gamma_x.\end{aligned} \tag{4}$$

Despite its strong advantages, the BEM for wave propagation problems still lacks reliable, efficient and automatic tools for the control of the error. Such tools are called *a posteriori error estimates*. They are used in the context of *auto-adaptive refinement* in order to ensure the accuracy of a computation. An

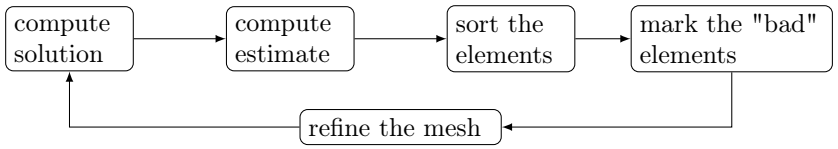


Figure 1: Example of an algorithm for autoadaptive mesh-refinement

auto-adaptive loop for mesh refinement can be described as pictured in Figure 1.

Three properties are required. Let η be such an estimate and e_h the numerical error, it must be *reliable* and *efficient*, *i.e.*, there exist two constants $C_{\text{rel}}, C_{\text{eff}} > 0$ such that

$$C_{\text{eff}}\,\eta \leq \|e_h\| \leq C_{\text{rel}}\,\eta.$$

The third property is the *locality* of η. It means that the total value of the estimate can be decomposed as the sum of local contributions of the estimate. This can be expressed by

$$\eta^2 = \sum_{\tau \in \mathcal{T}_h} \eta_\tau^2$$

where η_τ represents the value of η restricted to τ. This definition *does not imply* any local efficiency with respect to the error! Here we only mean that we need some value which can be defined locally in order to guide an autoadaptive refinement loop.

A posteriori estimates for the BEM have already been widely investigated, although not as much as for the FEM. Moreover, the numerical analysis remains to our knowledge rather limited. We can cite the pioneering work of B. Faermannn [7, 8, 9] which is based on the localization of the norms associated to Sobolev spaces of non-integer order on patches of elements of the mesh. In the field of acoustic scattering, we can have a look at the multilevel error indicator in [19] for general Fredholm operators. There is also the work of Chen et al. [20] in acoustics where a simple residual–based error estimate is used. Some investigations in electromagnetism have been made. For example Nochetto and Stamm [21] developed a weighted residual–based error estimate for the Electric Field Integral Equation. Finally, the most successful work (to our knowledge) is the one of the team of D. Praetorius at TU Wien and its collaborators who have deeply studied the *a posteriori* error estimates and the adaptive methods in the context of BEM for the Laplace problems. The first examples [4, 1] use the fact that the norm of the residual is an equivalent norm of the error. They are based on a localization of this norm. Other examples (the list being clearly non-exhaustive) are the averaging error estimates [2, 3] featuring hierarchical meshes and localizations of the Sobolev norms. A nice review of most of the

existing estimates can be found in [13]. For some of the estimates (for example [4]), there exists a proof for quasi-optimal convergence rates [12]. The topic of optimal convergence of autoadaptive refinement algorithms is tackled using an abstract setting in [11] and will not be further recalled in this paper.

The process of localization induces generally a loss of control between the "value of the error" and the value of the estimate. The corresponding multiplicative constant (equivalently named *efficiency constant*) will essentially depend on Γ. We seek a way to set the value of this constant independently on Γ while keeping the three first properties. The result is a new error estimate which features a localization technique based on the use of a localization operator Λ. Under appropriate conditions (in particular if Γ is $\mathcal{C}^\infty$), it is asymptotically exact with respect to the Galerkin norm of the error in the sense that when the mesh is refined it reflects the value of the error up to a higher order term. In the case of acoustic scattering and more particularly equations (1) and (2), the estimates read as follows: For r_h the residual of the equation which is solved,

$$\eta_{\mathcal{S}_0} = \|\Lambda_{\mathcal{S}_0} r_h\|_0 \quad \text{and} \quad \eta_{\mathcal{N}_0} = \|\Lambda_{\mathcal{N}_0} r_h\|_0,$$

with

$$\Lambda_{\mathcal{S}_0} r_h(x) := \frac{2}{\sqrt{\pi}} \nabla_\Gamma \cdot \int_\Gamma \sqrt{|x-y|}\, \nabla_\Gamma r_h(y)\, d\gamma_y,$$

and

$$\Lambda_{\mathcal{N}_0} r_h(x) := \frac{1}{\sqrt{\pi}} \int_\Gamma \frac{r_h(y)}{\sqrt{|x-y|}}\, d\gamma_y.$$

The paper is organized as follows. We first give an abstract setting. We then make a simple observation which is the first step to the construction of the new *a posteriori* error estimate. We state the theorem defining the estimate and we suggest a method for its construction. We build the estimates corresponding to the equations (1) and (2). We conclude with two numerical examples in order to confirm the expected properties and we give an opening on another way to build the operators Λ.

2. Abstract setting

In this section we establish the abstract setting which we are going to use. We also **slightly change the notations for the error and the residual**.

The equations resulting from the BEM are of the form

$$\mathcal{A}u = b, \tag{5}$$

where $\mathcal{A} : H \to H^\star$ is a linear bounded operator, H is some separable Hilbert space and $H^\star$ its topological dual space. We denote by $\|\cdot\|_H$ the norm on H which is based upon the inner product $(\cdot\,,\cdot)_H$.

We write $(\,.\,,\,.\,)$ for the L^2 inner product and $\langle\,.\,,\,.\,\rangle$ for the classical duality brackets.

In the following, for any operator $\mathcal{L}$ the adjoint operator is written $\mathcal{L}^\star$.

In this paper we are interested in the class of linear *Fredholm* operators which can be decomposed in the form

$$\mathcal{A} = \mathcal{A}_0 + \mathcal{K}, \tag{6}$$

where $\mathcal{A}_0$ is the continuous, symmetric, and coercive part and $\mathcal{K}$ is a compact perturbation. The Fredholm alternative [23] yields a solvability criterion for (5) and we suppose that this problem admits a unique solution. The coercive part $\mathcal{A}_0$ allows us to define the *Galerkin norm*.

Proposition 2.1 (Galerkin norm)**.** *Using the notation of (6), the quantity* $|||u|||^2 = \langle \mathcal{A}_0 u, u\rangle$ *defines an equivalent norm to* $\|\cdot\|_H$ *on the Hilbert space* H*. This norm is called the Galerkin norm.*

Proof. The proof is trivial since the operator $\mathcal{A}_0$ is bounded and coercive on H. □

In this paper we do not concern ourselves with the convergence of the autoadaptive algorithm but only on the *a posteriori* error estimation. Consequently, we suppose that the autoadaptive algorithm produces a sequence $(\mathcal{T}_l)_{l\in\mathbb{N}}$ of successively refined meshes and corresponding nested approximation spaces $V_l \subset V_{l+1} \subset H$ such that

$$V_\infty := \overline{\bigcup_{l=0\ldots\infty} V_l} = H. \tag{7}$$

Remark 2.1. *We do* ***not necessarily*** *have $V_\infty = H$ when we use a standard autoadaptive algorithm like the so-called Dörfler marking strategy [11]. Nevertheless, the paper [28] proposes a simple modification of the latter which does not impact the optimal convergence of the algorithm and ensures the property (7).*

We suppose furthermore that for each $l \in \mathbb{N}$ the associated discrete problem is well-posed. The corresponding discrete solution is written u_l and satisfies $\langle \mathcal{A}u_l, v_l\rangle = \langle b, v_l\rangle$ for all $v_l \in V_l$. The *approximation error* is then

$$e_l = u - u_l.$$

We also suppose that $\|\!|u-u_l|\!\| \xrightarrow[l\to\infty]{} 0$, *i.e.*, the discrete solution u_l converges to the *exact* solution u of (5). In practice, all these assumptions are satisfied if the mesh $\mathcal{T}_0$ corresponding to the initial space V_0 is sufficiently fine. However, [28] shows that some assumptions still hold if the initial mesh is coarse.

We define the *residual*

$$r_l = b - \mathcal{A}u_l, \qquad r_l \in H^\star. \tag{8}$$

Finally, for sake of simplicity in the presentation, we assume that for all $l \in \mathbb{N}$, $u_l \neq u$.

We introduce finally the notion of *higher order term* which we will extensively use in the following.

Definition 2.1 (Higher order term)**.** *We say that a function $a(\cdot)$ depending on b is a* higher order term *compared to b in 0 if*

$$\lim_{b\to 0} \frac{a(b)}{b} = 0.$$

In particular, *higher order terms* are often, in the following, a consequence of some compact operator applied to "something" converging to zero, typically the error e_l. In fact, we have the two following lemmas.

Lemma 2.1. *Let $(w_l)_{l\in\mathbb{N}}$ be the sequence defined as*

$$w_l = \frac{e_l}{\|\!|e_l|\!\|}, \qquad l \in \mathbb{N},$$

then $(w_l)_{l\in\mathbb{N}}$ is bounded and converges weakly to zero in V_∞.

Proof. We follow the same pattern as the proof of **Lemma 6** in [10].

We prove the result by using the following argument: if there exists $w \in H$ such that for every subsequence $(z_{k_l})_{l\in\mathbb{N}}$ of a sequence $(z_l)_{l\in\mathbb{N}}$, there exists a subsequence of these subsequences which converges weakly to w, then $(z_l)_{l\in\mathbb{N}}$ converges weakly to w.

Any subsequence of the sequence $(w_l)_{l\in\mathbb{N}}$ is obviously bounded since for all $l \in \mathbb{N}$, $\|\!|w_l|\!\| = 1$. Consequently, there exists for every of these subsequences, a weakly convergent subsequence $(w_{l_j})_{j\in\mathbb{N}}$ converging to some w. We want to prove that $w = 0$.

On the other side, the Galerkin orthogonality of the error $\langle \mathcal{A}e_l, v_l\rangle = 0$ for all $v_l \in V_l$ and the nestedness of the subspaces V_l implies that for all $j \in \mathbb{N}$ such that $l_j \geq l$, we have for all $v_l \in V_l$

$$\begin{aligned}\langle \mathcal{A}w_{l_j}, v_l\rangle &= \|\!|u-u_{l_j}|\!\|^{-1}\langle \mathcal{A}e_{l_j}, v_l\rangle \\ &= 0\end{aligned}$$

Moreover, for all $\varepsilon > 0$, there exists $j_0 \in \mathbb{N}$ such that for all $j \geq j_0$, we have $l_j \geq l$ and $|\langle \mathcal{A}(w - w_{l_j}), v_l \rangle| \leq \varepsilon$ by using the weak convergence of $(w_{l_j})_{j \in \mathbb{N}}$ to w and the continuity of the linear operator $\mathcal{A}$. Therefore, we have for all $\varepsilon > 0$, for all $l \in \mathbb{N}$, and sufficiently large l_j that

$$|\langle \mathcal{A}w, v_l \rangle| = \langle \mathcal{A}(w - w_{l_j}), v_l \rangle + \langle \mathcal{A}w_{l_j}, v_l \rangle \leq 0 + \varepsilon = \varepsilon,$$

which implies that for all $l \in \mathbb{N}$, we have

$$\langle \mathcal{A}w, v_l \rangle = 0, \qquad \forall v_l \in V_l.$$

Now, by density and continuity, the hypothesis (7) leads to

$$\langle \mathcal{A}w, v \rangle = 0, \qquad \forall v \in H.$$

Finally, the invertibility of the operator $\mathcal{A}$ gives $w = 0$ and the sequence $(w_l)_l$ converges weakly to $w = 0$. □

We then have the following lemma which enlightens the reason why compact operators can yield higher order terms.

Lemma 2.2. *If $\widetilde{\mathcal{K}} : H \to H^\star$ is a compact operator, then $\langle \widetilde{\mathcal{K}} e_l, e_l \rangle$ gives rise to a higher order term compared to $\|\!|e_l|\!\|^2$.*

Proof. Let $(w_l)_{l \in \mathbb{N}}$ be the sequence defined in **Lemma** 2.1. We know that this sequence converges weakly to 0. As a consequence, using the fact that $\widetilde{\mathcal{K}}$ is compact, we have the following strong convergence in $H^\star$

$$\|\widetilde{\mathcal{K}} w_l\|_{H^\star} \xrightarrow[l \to \infty]{} 0.$$

Finally, using Proposition 2.1, there exists a constant $C > 0$ such that for all $v \in H$, $\|v\|_H \leq C \|\!|v|\!\|$ and we can write:

$$\begin{aligned} |\langle \widetilde{\mathcal{K}} e_l, e_l \rangle| &= \|\!|e_l|\!\| \, |\langle \mathcal{K} w_l, e_l \rangle| \\ &\leq C \|\widetilde{\mathcal{K}} w_l\|_{H^\star} \, \|\!|e_l|\!\|^2 \end{aligned}$$

which means in particular that $\dfrac{|\langle \widetilde{\mathcal{K}} e_l, e_l \rangle|}{\|\!|e_l|\!\|^2} \xrightarrow[l \to \infty]{} 0.$ □

3. Construction of a new *a posteriori* error estimate

The aim of this section is to build a new efficient, reliable, local and possibly asymptotically exact (with respect to the Galerkin norm of the error) *a posteriori* error estimate in the sense of exactness up to some higher order terms.

As for many estimates of the literature (for example [4]), we use the fact that under the assumption that equation (5) is well–posed, the norm of the residual $\|r_l\|_{H^\star}$ is a reliable and efficient *a posteriori* error estimate of the norm of the error $|\!|\!|e_l|\!|\!|$. Actually, the well–posedness character of the continuous problem implies the existence of an inf-sup condition, *i.e.*, there exists $\alpha > 0$ such that

$$\inf_{v\in H}\sup_{w\in H}\frac{|\langle \mathcal{A}v, w\rangle|}{\|v\|_H\|w\|_H} \geq \alpha.$$

In particular, for all $u_l \in V_l$ (not necessarily solution of the Galerkin problem), we have:

$$\|r_l\|_{H^\star} = \sup_{w\in H}\frac{|\langle \mathcal{A}(u-u_l), w\rangle|}{\|w\|_H} \geq \alpha\|u-u_l\|_H.$$

Finally, the continuity of $\mathcal{A}$ leads to $\|r_l\|_{H^\star} \leq \|\mathcal{A}\|\|u-u_l\|_H$ and we have

$$\|A\|^{-1}\|r_l\|_{H^\star} \leq \|u-u_l\|_H \leq \alpha^{-1}\|r_l\|_{H^\star}. \tag{9}$$

We have the following theorem.

Theorem 3.1 (New estimate – Strong form)**.** *Let $\Lambda : H^\star \to L^2(\Gamma)$ be an isomorphism. Then, the* a posteriori *error estimate defined by*

$$\eta_\Lambda = \|\Lambda r_l\|_0,$$

is reliable, efficient and local (following our definition given in the introduction).

Proof. Since Λ is an isomorphism, then Λ and Λ^{-1} are bounded and we have $\|\Lambda^{-1}\|\|r_l\|_{H^\star} \leq \eta_\Lambda \leq \|\Lambda\|\|r_l\|_{H^\star}$. Consequently, $\|\Lambda r_l\|_0$ is a reliable and efficient *a posteriori* error estimate of $\|r_l\|_{H^\star}$. We conclude by using (9). □

In the following we discuss the possible asymptotic exactness of the estimate introduced in **Theorem 3.1** with respect to the Galerkin norm of the error.

We have the following proposition.

Proposition 3.1. *Let η_Λ be the estimate introduced in **Theorem 3.1**. If there exists an operator Λ such that $\Lambda^\star\Lambda\mathcal{A}_0 = \mathcal{I} + \mathcal{K}_2$ where $\mathcal{I}$ is the identity operator and $\mathcal{K}_2 : H \to H$ some compact perturbation, then η_Λ is asymptotically exact with respect to the Galerkin norm of the error.*

Proof. Let us start from the definition of $\|\Lambda r_l\|_0^2$: using the decomposition of $\mathcal{A}$ as the sum of a coercive part $\mathcal{A}_0$ and a compact part $\mathcal{K}$, we have

$$\begin{aligned}\|\Lambda r_l\|_0^2 &= (\Lambda\mathcal{A}e_l, \Lambda\mathcal{A}e_l)\\ &= (\Lambda(\mathcal{A}_0+\mathcal{K})e_l, \Lambda(\mathcal{A}_0+\mathcal{K})e_l)\\ &= (\Lambda\mathcal{A}_0e_l, \Lambda\mathcal{A}_0e_l) + (\Lambda\mathcal{K}e_l, \Lambda\mathcal{A}_0e_l) + (\Lambda\mathcal{A}_0e_l, \Lambda\mathcal{K}e_l) + (\Lambda\mathcal{K}e_l, \Lambda\mathcal{K}e_l)\\ &= \langle\mathcal{A}_0e_l, \Lambda^\star\Lambda\mathcal{A}_0e_l\rangle + \langle\mathcal{K}_1e_l, e_l\rangle, \end{aligned} \tag{10}$$

where $\mathcal{K}_1 = \mathcal{A}_0^\star\Lambda^\star\Lambda\mathcal{K} + \mathcal{K}^\star\Lambda^\star\Lambda\mathcal{A}_0 + \mathcal{K}^\star\Lambda^\star\Lambda\mathcal{K}$ is a compact operator from H to $H^\star$ as the composition of bounded and compact operators.

If $\Lambda^\star\Lambda\mathcal{A}_0 = \mathcal{I} + \mathcal{K}_2$, the identity (10) becomes

$$\begin{aligned}\|\Lambda r_l\|_0^2 &= \langle\mathcal{A}_0e_l, e_l\rangle + \langle(\mathcal{K}_1 + \mathcal{K}_2^\star\mathcal{A}_0)e_l, e_l\rangle\\ &= |\!|\!| e_l |\!|\!|^2 + \langle\mathcal{K}_3e_l, e_l\rangle, \end{aligned} \tag{11}$$

where the operator $\mathcal{K}_3 = \mathcal{K}_1 + \mathcal{K}_2^\star\mathcal{A}_0 : H \to H^\star$ is compact. Using **Lemma 2.2** with $\mathcal{K}_3$, we have that the quantity $\delta_l := \langle\mathcal{K}_3e_l, e_l\rangle / |\!|\!| e_l |\!|\!|^2$ converges to zero when $l \to +\infty$ and the equality (11) leads to

$$\frac{\|\Lambda r_l\|_0^2}{|\!|\!| e_l |\!|\!|^2} = 1 + \delta_l \xrightarrow[l\to+\infty]{} 1. \tag{12}$$

In other words, by carefully choosing Λ, the corresponding estimate η_Λ is asymptotically exact with respect to the Galerkin norm of the error.

□

The last remaining question is: how do we choose Λ? In many applications of the BEM, H and $H^\star$ are Sobolev spaces such that

$$H = H^s(\Gamma) \qquad \text{and} \qquad H^\star = H^{-s}(\Gamma), \qquad s \in \mathbb{R}$$

which means that $\mathcal{A}_0$ is an operator of order $2s$. The operator Λ is consequently of order $-s$: it is an "approximation of the inverse of the square root of $\mathcal{A}_0$". In a way, this problem is really close to analytical preconditioning techniques [24, 25] where one is looking for an approximate inverse of $\mathcal{A}$.

Such an operator Λ is not easily build and we suggest here an approach based on *microlocal analysis*. We perfectly acknowledge the fact that microlocal analysis techniques require Γ to be $\mathcal{C}^\infty$. That leads us to the following remark.

Remark 3.1 (Important!)**.** *The analysis we will be conducting in the following* **requires** Γ *to be* $\mathcal{C}^\infty$*. However, this is almost never the case in practice. In fact, the analysis will help us to find* ***candidates*** *for* Λ *which will have the required properties on such boundaries. Once we selected a potential candidate, there are no objections to trying it for domains with singularities or general Lipschitz boundaries. We are just not sure of the expected properties. This is what will be done in the numerical applications.*

On $\mathcal{C}^\infty$ surfaces/contours, $\mathcal{A}$ is a *pseudo-differential operator*. The *class of symbols of* $\mathcal{A}$ may be seen as the set of all its Fourier representations (see [15], chapters 6–8, for all the details). There is at least one representative of this class whose expansion is a sum of pseudo-homogeneous terms. The leading term is called *principal homogeneous symbol* $\sigma^0_{\mathcal{A},q}$ and its degree of homogeneity q yields the differentiation order of $\mathcal{A}$. The principal homogeneous symbol of $\mathcal{A}_0$, homogeneous of order $2s$, is then written $\sigma^0_{\mathcal{A}_0,2s}(x,\xi)$ where ξ is a variable in the Fourier domain.

An interesting property is that we can manipulate the symbols the same way we would manipulate the operators. Consequently, the principal homogeneous symbol for the operator Λ is

$$\boxed{\sigma^0_{\Lambda,-s}(x,\xi) = \frac{1}{\sqrt{\sigma^0_{\mathcal{A}_0,2s}(x,\xi)}}}. \tag{13}$$

The knowledge of $\sigma^0_{\Lambda,-s}$ gives us an abstract candidate. We can express a representative of all the candidates as an integral operator whose kernel can be compute from $\sigma^0_{\Lambda,-s}$.

In the following section we particularize the operator $\mathcal{A}$ for equations (1) and (2) and consequently the functional spaces H and $H^\star$ and the Galerkin norm $|\!|\!| \,.\, |\!|\!|$. We then build the corresponding operators Λ.

4. Construction of the operator Λ for two acoustic integral operators in 2D

In this section we give a method for the construction of the Λ–operator for each equation (1) and (2) and compute the related explicit operators $\Lambda_{\mathcal{S}_0}$ and $\Lambda_{\mathcal{N}_0}$.

In the case of equation (1), we have $\mathcal{A} \equiv \mathcal{S}_k$ which means that $H = H^{-1/2}(\Gamma)$ and $H^\star = H^{1/2}(\Gamma)$. Assuming that the logarithmic capacity of Γ is lower than one, the operator $\mathcal{S}_0$ is elliptic (see [29], part **5** and **6**). The Galerkin norm may be defined as

$$|\!|\!| u |\!|\!|^2 = \langle \mathcal{S}_0 u, u \rangle, \qquad \forall u \in H^{-1/2}(\Gamma).$$

It is known from the literature [15], p. 514, that the principal homogeneous symbol of $\mathcal{S}_0$ is

$$\sigma^0_{\mathcal{S}_0,-1}(x,\xi) = \frac{1}{2|\xi|}.$$

We deduce the principal homogeneous symbol of $\Lambda_{\mathcal{S}_0}$

$$\boxed{\sigma^0_{\Lambda_{\mathcal{S}_0},1/2}(x,\xi) = \sqrt{2|\xi|}}. \tag{14}$$

In the case of equation (2), we have $\mathcal{A} \equiv \mathcal{N}_k$ with $H = H^{1/2}(\Gamma)$ and $H^\star = H^{-1/2}(\Gamma)$. The Galerkin norm in this case is defined as

$$|\!|\!| u |\!|\!|^2 = \langle (\mathcal{N}_0 + \mathcal{S}_0)u, u \rangle, \qquad \forall u \in H^{1/2}(\Gamma).$$

Once again, we find in the literature [15], p. 526, that the principal homogeneous symbol of $\mathcal{N}_0$ is

$$\sigma^0_{\mathcal{N}_0,1}(x,\xi) = \frac{|\xi|}{2}.$$

We then deduce the principal homogeneous symbol of $\Lambda_{\mathcal{N}_0}$

$$\boxed{\sigma^0_{\Lambda_{\mathcal{N}_0},-1/2}(x,\xi) = \sqrt{\frac{2}{|\xi|}}}. \tag{15}$$

The effect of the operator $\Lambda_{\mathcal{S}_0}$ can then be seen as a "half-differentiation" while the effect of the operator $\Lambda_{\mathcal{N}_0}$ is a "half-integration".

We must now re-construct a practical form of $\Lambda_{\mathcal{S}_0}$ and $\Lambda_{\mathcal{N}_0}$. We will seek these operators as integral operators as we have explicit formulas for the conversion symbol $\rightarrow$ kernel and kernel $\rightarrow$ symbol. These formulas are summarized on p. 392–394 in [15].

4.1. Construction of $\Lambda_{\mathcal{S}_0}$

The formulas for the conversion symbol $\rightarrow$ kernel are unfortunately impractical and it is easier to first guess the final form of the operator and show that it has the right symbol. Let

$$\Lambda_{\mathcal{S}_0} u(x) = -\sqrt{\pi}\,\text{f.p.} \int_\Gamma \frac{u(y)}{|x-y|^{3/2}}\, d\gamma_y, \tag{16}$$

where "f.p." means "Hadamard finite part integral". The kernel $k(x, x-y) = |x-y|^{-3/2}$ is a pseudo-homogeneous function of order $-3/2$. Using formula (7.1.82) in [15], we compute the principal homogeneous symbol of order $1/2$.

$$\sigma^0_{\Lambda_{\mathcal{S}_0},1/2}(x,\xi) = -\frac{1}{2\sqrt{\pi}}\,\text{f.p.}\int_{\mathbb{R}} k(x,z)\,e^{-i\xi z}\,d\gamma_y$$
$$= -\frac{1}{2\sqrt{\pi}}\lim_{\epsilon\to 0}\left(\int_{|z|>0}\frac{e^{-i\xi z}}{|z|^{3/2}}\,dz - 2\frac{e^{-i\xi\epsilon}+e^{i\xi\epsilon}}{\sqrt{\epsilon}}\right).$$

Using the fact that $\Re(e^{-i\xi z})$ is even and $\Im(e^{-i\xi z})$ is odd, we have

$$\tilde{\sigma}^0_{\Lambda_{\mathcal{S}_0},1/2}(x,\xi) = 2\int_\epsilon^\infty \frac{\cos(\xi z)}{z^{3/2}}\,dz$$
$$= 2\int_\epsilon^\infty \frac{\cos(\xi z)-1}{z^{3/2}}\,dz + \frac{4}{\sqrt{\epsilon}}.$$

As $\epsilon \to 0$,

$$\sigma^0_{\Lambda_{\mathcal{S}_0},1/2}(x,\xi) = -\frac{1}{\sqrt{\pi}}\int_0^\infty \frac{\cos(\xi z)-1}{z^{3/2}}\,dz$$
$$= -\frac{1}{\sqrt{\pi}}\int_0^\infty \sqrt{|\xi|}\frac{\cos(\zeta)-1}{\zeta^{3/2}}\,d\zeta, \qquad \zeta = |\xi| z$$
$$\boxed{\sigma^0_{\Lambda_{\mathcal{S}_0},1/2}(x,\xi) = \sqrt{2|\xi|}}.$$

Consequently, the operator defined in (16) is a good candidate. However, we will use a slightly modified version obtained by integration by parts [6]. We set

$$\boxed{\Lambda_{\mathcal{S}_0}u(x) = \frac{2}{\sqrt{\pi}}\nabla_\Gamma \cdot \int_\Gamma \sqrt{|x-y|}\,\nabla_\Gamma u(y)\,d\gamma_y}. \tag{17}$$

The kernel is modified with respect to the previous version since it contains all constant functions.

Proposition 4.1. *Let $\Gamma = \mathcal{C}_R(O)$ be the circle centered in the origin O with radius R and*

$$V = \left\{\varphi \in H^{1/2}(\Gamma),\ \langle\varphi, 1_\Gamma\rangle = 0\right\}$$

the subspace of all functions of $H^{1/2}(\Gamma)$ with zero mean, then the operator

$$\Lambda_{\mathcal{S}_0} : V \to L^2(\Gamma)$$

defined in (17) is an isomorphism.

Proof. The operator $\Lambda_{\mathcal{S}_0}$ is Fredholm since it corresponds, by construction, to $\mathcal{S}_0^{-1/2}$ up to more regular operators which are obviously compact as a consequence of the Sobolev embeddings. Consequently, we only need to prove that $\Lambda_{\mathcal{S}_0}$ is injective. We will show that the only eigenfunction associated to the zero eigenvalue is the constant function on Γ.
The eigenfunctions are $\varphi_n(\theta) = e^{in\theta}$ for $n \in \mathbb{Z}$ and we have

$$\begin{aligned}&\Lambda_{\mathcal{S}_0}\varphi_n\\ &= \frac{2}{\sqrt{\pi}}\frac{\partial}{\partial\theta_x}\int_\Gamma \sqrt{R}\left((\cos(\theta)-\cos(\theta_x))^2+(\sin(\theta)-\sin(\theta_x))^2\right)^{1/4}\frac{\partial e^{in\theta}}{\partial\theta}\,R\,d\theta\end{aligned}$$

$$\Lambda_{\mathcal{S}_0}\varphi_n = -\frac{2R^{3/2}n^2}{\sqrt{\pi}}e^{in\theta_x}I_n,$$

with $I_n = \displaystyle\int_0^{2\pi}\sqrt{2\left|\sin(\frac{\tilde\theta}{2})\right|}\,e^{in\tilde\theta}\,d\tilde\theta$ where we made the change of variable $\tilde\theta = \theta - \theta_x$. Therefore, we have obviously

$$\lambda_n = -\frac{2R^{3/2}n^2}{\sqrt{\pi}}I_n.$$

We want to prove that $I_{n\geq 1} < 0$. The parity of the integrand in I_n yields

$$I_n = 2\int_0^{\pi}\sqrt{2\left|\sin(\frac{\tilde\theta}{2})\right|}\cos(n\tilde\theta)\,d\tilde\theta.$$

Let $f(t), t \in [0,\pi]$ be a positive concave increasing function, and we set $I_n(f) = \displaystyle\int_0^{\pi} f(t)\,\cos(nt)\,dt$. Integration by parts yields

$$\begin{aligned}I_n(f) &= -\frac{1}{n}\int_0^{\pi} f'(t)\sin(nt)\,dt\\ &= -\frac{1}{n^2}\int_0^{n\pi} f'(\frac{t}{n})\sin(t)\,dt\\ &= -\frac{1}{n^2}\sum_{k=0}^{n-1}\int_{k\pi}^{(k+1)\pi} f'(\frac{t}{n})\sin(t)\,dt\\ &= -\frac{1}{n^2}\sum_{k=0}^{n-1} J_k.\end{aligned}$$

If $n = 1$, obviously $I_n(f) \leq 0$. In fact, we only need to prove that for all even numbers k in $[0, n-2]$, the quantity $J_k + J_{k+1}$ is a positive real number. Let $f'_{k+j} = f'(\frac{k+j}{n})$ and set k even, then $\sin(t) \geq 0$ and

$$\begin{cases} J_k & \geq \displaystyle\int_{k\pi}^{(k+\frac{1}{2})\pi} f'_{k+\frac{1}{2}} \sin(t)\, dt + \int_{(k+\frac{1}{2})\pi}^{(k+1)\pi} f'_{k+1} \sin(t)\, dt \\ J_{k+1} & \geq \displaystyle\int_{(k+1)\pi}^{(k+\frac{3}{2})\pi} f'_{k+1} \sin(t)\, dt + \int_{(k+\frac{3}{2})\pi}^{(k+2)\pi} f'_{k+\frac{3}{2}} \sin(t)\, dt. \end{cases}$$

Finally,

$$J_k + J_{k+1} \geq f'_{\frac{1}{2}} - f'_{\frac{3}{2}} \geq 0$$

with strict inequality if f' is decreasing. We can easily check that $f(t) = 2\sqrt{2\sin(\frac{t}{2})}$ complies with all the hypotheses and for all $n \geq 1$, $I_n < 0$. The eigenspace associated to $\lambda_0 = 0$ is spanned by $\varphi_0(\theta) = 1$ and $\lambda_{n\geq 1} > 0$ which concludes the proof. □

It is not yet sufficient to prove that it is an isomorphism on any contour. In fact, we will use a weakened hypothesis on $\Lambda_{\mathcal{S}_0}$ but we still suppose that Γ is a $\mathcal{C}^\infty$ curve.

Theorem 4.1. *Let Γ be a $\mathcal{C}^\infty$ curve, the functional space V and $\Lambda_{\mathcal{S}_0}$ being defined as in* ***Proposition*** *4.1. Let also $(\mathcal{T}_l)_{l\in\mathbb{N}}$ and $(V_l)_{l\in\mathbb{N}}$ be the sequences of meshes and nested approximation spaces introduced in* ***section 2*** *such that the function $v = 1$ belongs to V_l, then there exists a rank l_0 such that for all $l \geq l_0$, the* a posteriori *error estimate for $|||u - u_l|||$ defined by*

$$\eta_{\Lambda_{\mathcal{S}_0}} = \|\Lambda_{\mathcal{S}_0} r_l\|_0, \tag{18}$$

is ***reliable***, ***efficient***, ***local*** *and* ***asymptotically exact****.*

Proof. We can decompose $\Lambda_{\mathcal{S}_0}$ as the sum of an isomorphism $\Lambda_0 : V \to L^2(\Gamma)$ and a compact perturbation $K : V \to L^2(\Gamma)$. For example, by construction, $\Lambda_{\mathcal{S}_0}$ has the same principal symbol as $\mathcal{S}_0^{-1/2}$ which is an isomorphism on V and consequently, we can take $\Lambda_0 = \mathcal{S}_0^{-1/2}$ and $K = \Lambda_{\mathcal{S}_0} - \mathcal{S}_0^{-1/2}$. More generally, the structure of the principal symbol of $\Lambda_{\mathcal{S}_0}$ implies that the latter is a strongly elliptic operator and consequently, it verifies a Garding inequality (see [15]). The hypothesis $v = 1 \in V_l$ implies that $r_l \in V$ and we have

$$\begin{aligned} \|\Lambda_{\mathcal{S}_0} r_l\|_0^2 &= ((\Lambda_0 + K) r_l, (\Lambda_0 + K) r_l) \\ &= \langle \Lambda_0^\star \Lambda_0 r_l, r_l \rangle_{-1/2,1/2} + \langle \tilde{K} r_l, r_l \rangle_{-1/2,1/2}, \end{aligned}$$

where $\tilde{K} = \Lambda_0^\star K + K^\star \Lambda_0 + K^\star K$ is compact.
Since Λ_0 is an isomorphim, there exist two constants C_1, $C_2 > 0$ such that

$$\begin{aligned}\left| C_1 \|r_l\|_{1/2}^2 - \|\tilde{K} r_l\|_{-1/2} \|r_l\|_{1/2} \right| &\leq \|\Lambda_{\mathcal{S}_0} r_l\|_0^2 \\ &\leq C_2 \|r_l\|_{1/2}^2 + \|\tilde{K} r_l\|_{-1/2} \|r_l\|_{1/2},\end{aligned}$$

and we would like $\|\tilde{K} r_l\|$ to converge faster than $\|r_l\|_{1/2}$ as $l \to \infty$. Using **Lemma 2.2** and the fact that $\|r_l\|_{1/2}$ is a reliable and efficient *a posteriori* error estimate of the norm of the error $|||e_l|||$, *i.e.*, there exist β_1, $\beta_2 > 0$ such that $\beta_1 |||e_l||| \leq \|r_l\|_{1/2} \leq \beta_2 |||e_l|||$, we have

$$\delta_l := \frac{\|\tilde{K} r_l\|_{-1/2}}{\|r_l\|_{1/2}} = \frac{\|\tilde{K} \mathcal{A} e_l\|_{-1/2}}{\|r_l\|_{1/2}} \leq \frac{\|\tilde{K} \mathcal{A} w_l\|_{-1/2}}{\beta_1} \xrightarrow{l \to \infty} 0.$$

Finally, there exist two constants $C_3 > 0$ and $C_4 > 0$ and a rank $l_0 \in \mathbb{N}$ such that for all $l \geq l_0$,

$$C_3 |||e_l||| \leq \|\Lambda_{\mathcal{S}_0} r_l\|_0 \leq C_4 |||e_l|||,$$

i.e., the reliability and the efficiency where $C_3 = \beta_1 C_1 - \delta_l$ and $C_4 = \beta_2 C_2 + \delta_l$. The asymptotic exactness is a result from **Proposition 3.1**. □

Remark 4.1. *The rank l_0 introduced in **Theorem 4.1** corresponds to the moment when the compact parts resulting from the decomposition $\Lambda_{\mathcal{S}_0} = \Lambda_0 + K$ are under control.*

4.2. Construction of $\Lambda_{\mathcal{N}_0}$

We construct the operator Λ associated to $\mathcal{N}_0$. We will follow the same approach as for $\Lambda_{\mathcal{S}_0}$. We already know from (15) the principal symbol of $\Lambda_{\mathcal{N}_0}$. We seek the associated Schwartz-kernel. From formula (7.1.39) in [15],

$$\begin{aligned}k(x,z) &= \frac{\sqrt{2}}{2\pi} \int_{\mathbb{R}} \frac{e^{i\xi z}}{\sqrt{|\xi|}} \, d\xi \\ &= \frac{1}{\sqrt{2}\pi} \left(\lim_{\xi \to \infty} \frac{\sqrt{\pi} \operatorname{erf}(\sqrt{-iz}\sqrt{\xi})}{\sqrt{-iz}} + \lim_{\xi \to -\infty} \frac{\sqrt{\pi} \operatorname{erf}(\sqrt{iz}\sqrt{-\xi}}{\sqrt{iz}} \right) \\ &= \sqrt{\frac{2}{\pi}} \Re\left(\frac{1}{\sqrt{iz}}\right) \\ &= \sqrt{\frac{2}{\pi}} \sqrt{\frac{|z|}{2}} \frac{1}{|z|} \\ &= \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{|z|}}.\end{aligned}$$

We define

$$\boxed{\Lambda_{\mathcal{N}_0} u(x) = \frac{1}{\sqrt{\pi}} \int_\Gamma \frac{u(y)}{\sqrt{|x-y|}}\, d\gamma_y}. \tag{19}$$

We have a theorem similar to **Theorem 4.1**.

Theorem 4.2. *Let Γ be a $\mathcal{C}^\infty$ curve, and $\Lambda_{\mathcal{N}_0}$ as defined in (19). Let also $(\mathcal{T}_l)_{l\in\mathbb{N}}$ and $(V_l)_{l\in\mathbb{N}}$ be the sequences of meshes and nested approximation spaces introduced in* ***section 2****, then there exists a rank l_0 such that for all $l \geq l_0$, the* a posteriori *error estimate for $|||u - u_l|||$ defined by*

$$\eta_{\Lambda_{\mathcal{N}_0}} = \|\Lambda_{\mathcal{N}_0} r_l\|_0,$$

is ***reliable****,* ***efficient****,* ***local*** *and* ***asymptotically exact****.*

Proof. The proof is carried in the same fashion as for Theorem 4.1. □

5. Numerical examples

In this section we aim at validating the properties of $\Lambda_{\mathcal{S}_0}$ and $\Lambda_{\mathcal{N}_0}$. We only consider the exterior Dirichlet and Neumann problems. The curve Γ is closed. In this paper we suppose that the incident wave u_i is a plane wave propagating along the $-\mathbf{e}_x$ axis. For all the simulations, the wave number is set to

$$k = 15.$$

We focus on two geometries: a circle with radius $R = 0.9$ and a square with side length $a = 1$. The justifications for the choice of these geometries are given in the corresponding subsections.

The initial meshes contain 20 elements which correspond to approximately 2 elements per wavelength. This is really low as the rule of thumb recommends both at least 6 elements per wavelength and additional refinement at the edges/corners [27]. We choose such a low initial discretization in order to "get the estimate's back against the wall"[1].

For each case, the autoadaptive refinement algorithm is guided using η_Λ. The value of η_Λ is compared to the value of some estimates of the literature:

- The residual–based estimate by C. Carstensen et al [4, 1], written η_{r_h}.

[1] If the initial number of elements is too low, the estimate "may not detect" the oscillations and yields a very low error

- An averaging–based estimate by D. Praetorius et al [2, 3]. When the projector is the classical L^2 projector, we write the corresponding estimate $\eta_{\Pi_{L^2}}$. In the case of equation (1), we also tried the Galerkin projection and the corresponding estimate is written $\eta_{\Pi_{\mathrm{G}}}$.
- The reference solution which can either be the exact solution when available (on the circle), or an "exact estimate". The latter consists in comparing the solution u_l with the solution $\widehat{u}_l$ for some refinement $\widehat{\mathcal{T}}_l$ of $\mathcal{T}_l$.

The rest of this section is organized as follows: we first give informations on the implementation of the BEM and the different estimates. We then explain the refinement algorithm which is used. Finally, we introduce the test cases and produce the result for the estimate $\eta_{\Lambda_{\mathcal{S}_0}}$ and for $\eta_{\Lambda_{\mathcal{N}_0}}$.

5.1. Implementation of the BEM and the estimates

The approximation/test space for equation (3) is chosen to be $V_l = \mathcal{P}^0(\mathcal{T}_l)$, the space of constant functions on $\mathcal{T}_l$. Assuming that the mesh is quasi-uniform and that the exact solution is smooth enough, then the best possible convergence rate with respect to the size of the elements can be estimated using the Bramble–Hilbert lemma (see [5], Theorem 4.1.3) from the polynomial degree of the approximation spaces. We expect the *best possible* convergence rate

$$\|e_l\|_{-1/2} = \mathcal{O}(h_l^{3/2}) = \mathcal{O}(N_{\mathrm{elem}}^{-3/2})$$

where N_{elem} is the number of elements in $\mathcal{T}_l$. The reason why the convergence rate is expressed with respect to N_{elem} is that the mesh size is not relevant anymore. On the contrary, the computation cost is directly linked to N_{elem}. We only *hope* that by using autoadaptive refinement we are able to obtain the convergence rate for a smooth solution, even if it is not.

For equation (4), the approximation/test space is $V_l = \mathcal{P}^1(\mathcal{T}_l)$ the continuous space of linear functions on $\mathcal{T}_l$. As for equation (3), we expect $\|e_l\|_{1/2} = \mathcal{O}(h_l^{3/2}) = \mathcal{O}(N_{\mathrm{elem}}^{-3/2})$.

The BEM matrix elements are integrated using semi-analytical integration. For couples of elements for which the integral is singular, the kernel is decomposed in a singular part and a regular part. The regular part is integrated using classical Gauss-Legendre quadrature rules.

The singular part consists in integrating two times a logarithmic function times some polynomial. The inner integral is computed analytically while the outer integral is computed using once again a Gauss-Legendre quadrature rule.

However, it has been noticed that when the degree of refinement is high, the numerical solution has instabilities. In that case, one must increase the number of quadrature points.

The implementation of the estimates depends on the ability to compute the residual. The basis remains the same: we compute a polynomial approximation $\mathbf{R}_l$ of the residual, we assemble the matrix $\mathbf{L}$ corresponding to the variational formulation associated to the operator Λ, we assemble the classical L^2–mass matrix, and we compute a polynomial approximation of Λr_l under the form

$$\mathbf{E} = \mathbf{M}^{-1}\mathbf{L}\mathbf{R}_l.$$

The estimate is then computed as $\eta_\Lambda^2 = \mathbf{E}^{\mathrm{T}}\mathbf{M}\mathbf{E}$.

For both equations (1) and (2), the residual is projected using a classical L^2 projection on a polynomial space of higher order. The main difference lies in the mesh on which the residual is projected.

In the case of equation (1), the projection on the same mesh $\mathcal{T}_l$ fails to be accurate and leads to a poor behavior for the convergence with a lot of oscillations. To remedy this problem, we must project the residual on a uniform refinement $\widehat{\mathcal{T}}_l$ of $\mathcal{T}_l$. To achieve good accuracy, we had to refine $\mathcal{T}_l$ at least three times (each "old" segment contains 2^3 new segments)! The projection space is then $\mathcal{P}^1(\widehat{\mathcal{T}}_l)$.

In the case of equation (2), the projection on the mesh $\mathcal{T}_l$ proves sufficient to achieve a good accuracy. The projection space is $\mathcal{P}^2(\mathcal{T}_l)$ which is the space of piecewise quadratic, globally continuous, functions.

The entries of the matrix $\mathbf{L}$ are computed analytically when we consider the integration of an element over itself. In the other cases, classical Gauss-Legendre quadrature are sufficient.

5.2. The autoadaptive refinement algorithm

The sequence of meshes $(\mathcal{T}_l)_{l\in\mathbb{N}}$ is generated using the autoadaptive loop described in Figure 1. The criterion used for the marking is the Dörfler marking (see e.g. [11]) with parameter θ_d. We mark the elements using the squared indicator, *i.e.*, we refine the minimum set $\mathcal{M}(\mathcal{T}_l)$ of elements $\tau \in \mathcal{T}_l$ such that

$$\theta_\mathrm{d} \sum_{\tau\in\mathcal{T}_l} \eta_{\Lambda,\tau}^2 \leq \sum_{\tau\in\mathcal{M}(\mathcal{T}_l)} \eta_{\Lambda,\tau}^2. \tag{20}$$

In other words, we mark the elements of $\mathcal{T}_l$ contributing the most to $100\cdot\theta_\mathrm{d}\%$ of the total squared error. If θ_d is close to 0, only a few elements are refined. On the contrary, θ_d close to 1 induces "nearly uniform" refinement. In the following we choose $\boxed{\theta_\mathrm{d} = 0.5}$ as it seems to be a good compromise.

The elements being linear, the refinement is carried out by bisecting each element.

The algorithm stops when a prescribed number of elements in the mesh is reached.

5.3. Numerical application – Circle with radius $R = 0.9$

The circle is probably the most simple geometry. The exact solution for both equations (1) and (2) is perfectly smooth and the convergence rate for uniform refinement is already the best possible convergence rate. This test case fulfills all the hypotheses of the development made in the **sections 2**–**4**.

The convergence curves for equation (1) are presented in Figure 2. The uniform convergence rate is already the best possible for uniform refinement and autoadaptive refinement only slightly improves the value of the error. We observe that the curves for $\eta_{\Lambda_{\mathcal{S}_0}}$ and the reference overlap perfectly.

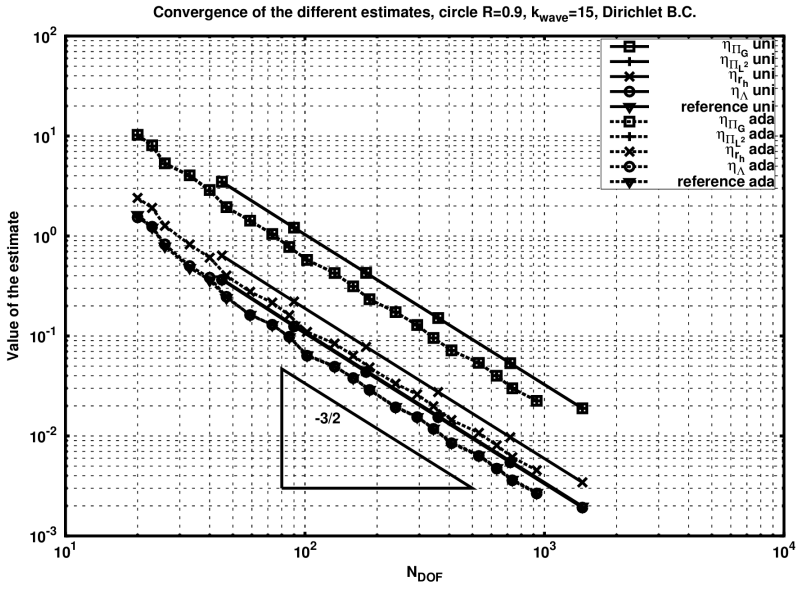


Figure 2: Convergence of different estimates, circle $R = 0.9$, $k = 15$, Dirichlet b.c.

We observe the same behavior for equation (2) in Figure 3. In that case, the improvement coming from autoadaptive refinement is negligible.

As expected, the estimate η_Λ behaves perfectly well on a smooth geometry. Unfortunately, smooth geometries are barely useful, nor met, in practical applications. We wish to know how the Λ–based estimate behaves when it is **pulled out of its theoretical environment**.

5.4. Numerical application – Unit square

We choose now the geometry to be a square with side length $a = 1$. It is now a Lipschitz curve and we fail to meet the hypothesis for the development of **section 4**. It is also known from the literature [26] that the solution in the case of the Dirichlet problem is singular at the corners while it is weakly continuous

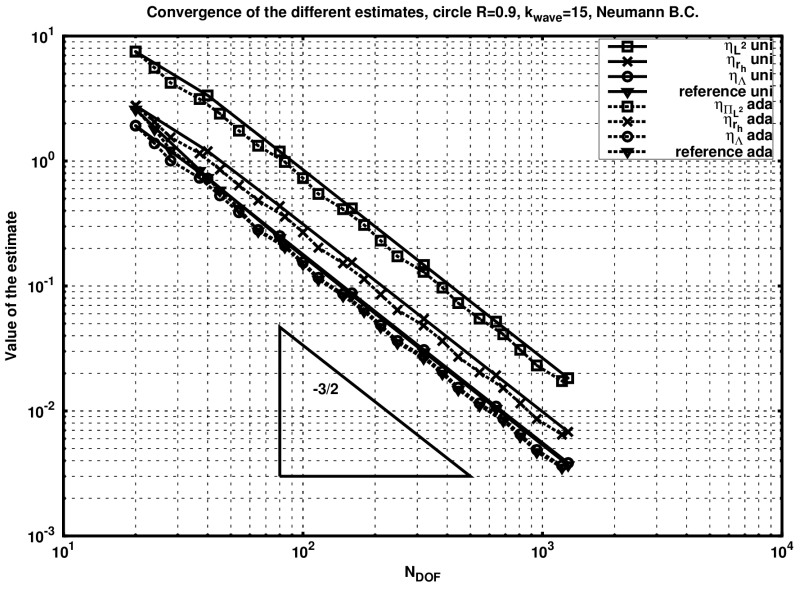


Figure 3: Convergence of different estimates, circle $R = 0.9$, $k = 15$, Neumann b.c.

in the case of the Neumann problem. Consequently, the convergence rate is not the best possible when using uniform mesh refinement.

However, we expect that the estimates keep their properties in this more general case.

The convergence curves for the Dirichlet problem are represented in Figure 4. Uniform refinement is clearly suboptimal as the convergence rate is approximately[2] $\mathcal{O}(N_{\text{elem}}^{-0.66})$. The curves for $\eta_{\Lambda_{S_0}}$ and the reference overlap. It means that the "exact" error in the Galerkin norm is also accurately estimated by $\eta_{\Lambda_{S_0}}$. The autoadaptive algorithm is also efficiently guided by $\eta_{\Lambda_{S_0}}$ since the convergence rate for autoadaptive refinement is the best possible.

We make the same observation for the Neumann problem, in Figure 5, for which the error is accurately estimated and the convergence rate is the best possible.

6. Conclusion

We introduced a new *a posteriori* error estimate for which we were able to prove the reliability and efficiency when Γ is a $\mathcal{C}^\infty$ curve for both equations (1) and (2). In this case, the Λ–estimate is asymptotically exact for any Γ with respect to the Galerkin norm of the error in the sense that it is exact up to

[2]Using [26] and the Bramble–Hilbert lemma, we can prove this value.

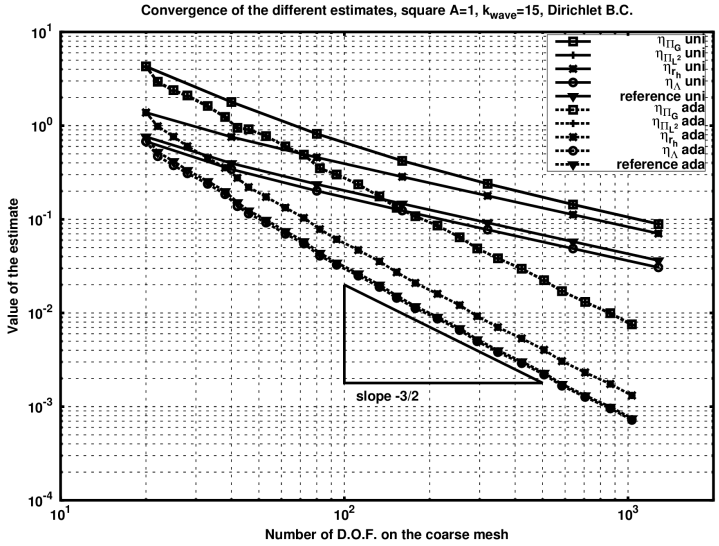


Figure 4: Convergence of different estimates, square $a = 1$, $k = 15$, Dirichlet b.c.

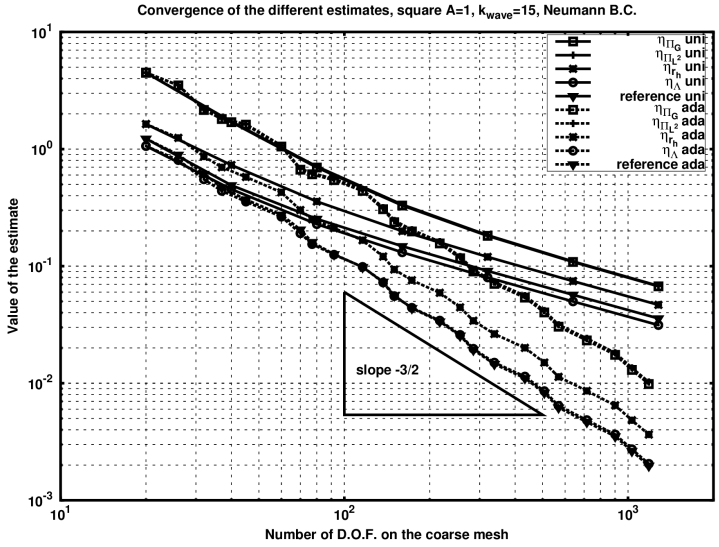


Figure 5: Convergence of different estimates, square $a = 1$, $k = 15$, Neumann b.c.

higher order terms. This is confirmed by numerical simulations on the circle. We also verify numerically that it is reliable, efficient and asymptotically exact on Γ with corners, but we are not yet able to prove it.

Its main advantages are that it does not require the computation of the

solution on a finer mesh (as required by the averaging techniques) or the use of complementary functional subspaces (see [16]). The first tricky point will be the accurate computation of the residual which can be costly. However this can be accelerated with a fast-multipole algorithm. The second difficulty will be the computation of the Λ–operator as it may require the computation of singular integrals.

In this paper, the Λ–estimate has been introduced for oscillatory problems but it remains of course valid for the classical problem of the Laplace equation. Much better, we provide a generic pattern for the construction of an *a posteriori* error estimate for the BEM in general. The "only" difficulty will be to find an appropriate Λ–operator.

We understand that the use of an integral operator for *a posteriori* estimation is costly. In fact, it is totally impractical as soon as we are are dealing with 3D–acoustics since it requires specific integration technique on triangles for the singular kernels. Consequently, we are currently investigating the use of Λ under the form $\Lambda \equiv \sqrt{2}\,(\mathcal{I} - \Delta)^{\pm 1/4}$ for which it has already been proven that it is an isomorphism on generic Lipschitz boundaries and for which it is easier to define a behavior in the case of screen problems.

7. Acknowledgements

This work is funded by the *Ministère français de la Défense* and the *Agence Nationale de la Recherche* (ANR-12-MONU-0021).

8. Bibliography

[1] C. Carstensen, M. Maischak, D. Praetorius, and E. P. Stephan, *Residual-based a posteriori error estimate for hypersingular equation on surfaces*, Numer. Math. **97** (2004), no. 3, 397–425.

[2] C. Carstensen and D. Praetorius, *Averaging Techniques for the Effective Numerical Solution of Symm's Integral Equation of the First Kind*, SIAM Journal on Scientifi computing **27** (2006), no. 4, 1226–1260

[3] C. Carstensen and D. Praetorius, *Averaging techniques for a posteriori error control in finite element and boundary element analysis*, Boundary Element Analysis (O. Steinbach M. Schanz, ed.), vol. 29, Springer, 2007, pp. 29–59.

[4] C. Carstensen, *An a posteriori error estimate for a first-kind integral equation*, Math. Comput. **66** (1997), no. 217, 139–155.

[5] P. G. Ciarlet, *Finite element method for elliptic problems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2002.

[6] M. Lecouvez F. Collino, P. Joly and B. Stupfel, *Quasi-local transmission conditions for non-overlapping domain decomposition methods for the Helmholtz equation*, Comptes Rendus Physique **15** (2014), no. 5, 403–414.

[7] B. Faermann, *Local a-posteriori error indicators for the Galerkin discretization of boundary integral equations*, Numer. Math. **79** (1998), 43–76.

[8] B. Faermann, *Localization of the Aronszajn-Slobodeckij norm and application to adaptive boundary elements methods. Part I. The two-dimensional case*, IMA J. Numer. Anal. **20** (2) (2000), 203–234.

[9] B. Faermann, *Localization of the Aronszajn-Slobodeckij norm and application to adaptive boundary elements methods. Part II. The three-dimensional case*, Numer. Math. **92** (2002), 467–499.

[10] M. Feischl, T. Führer, and D. Praetorius, *Adaptive FEM with optimal convergence rates for a certain class of nonsymmetric and possibly nonlinear problems*, SIAM Journal on Numerical Analysis **52** (2014), no. 2, 601–625.

[11] C. Carstensen and M. Feischl, M. Page and D. Praetorius, *Axioms of adaptivity*, Comp. Methods Appl. Math. **67** (2014), no. 6, 1195–1253.

[12] M. Feischl, M. Karkulik, J.M. Melenk and D. Praetorius, *Quasi-optimal Convergence Rate for an Adaptive Boundary Element Method*, SIAM Journal on Numerical Analysis **51** (2013), no. 2, 1327–1348.

[13] M. Feischl, T. Führer, N. Heuer, M. Karkulik and D. Praetorius, *Adaptive Boundary Element Methods : A posteriori error estimators, adaptivity, convergence, and implementation*, Achives of Computational Methods in Engineering **22** (2014), no. 3, 309–389.

[14] N. Hale, N. J. Higham, and L. N. Trefethen, *Computing* $A^{\alpha}, \log(A)$*, and related matrix functions by contour integrals*, SIAM Journal on Numerical Analysis **46** (2008), no. 5, 2505–2523.

[15] G. C. Hsiao and W. L. Wendland, *Boundary integral equations*, Applied Mathematical Sciences, Springer, Berlin, Heidelberg, 2008.

[16] J. Jou and J.-L. Liu, *A posteriori boundary element error estimation*, Journal of Computational and Applied Mathematics **106** (1999), no. 1, 1–19.

[17] W. Hackbusch, *Hierarchical Matrices : Algorithms and Analysis*, Springer–Verlag Berlin Heidelberg, 49, 2015.

[18] J. Carrier, L. Greengard and V. Rokhlin, *A Fast Adaptive Multipole Algorithm for Particle Simulations*, SIAM J. Sci. Stat. Comput. 9, 669 (1988).

[19] M. Maischak, P. Mund and E.P. Stephan, *Adaptive multilevel BEM for acoustic scattering*, Comput. Methods Appl. Mech. Engrg. 150 (1997) 351–367.

[20] J.T. Chen, K.H. Chen and C.T. Chen, *Adaptive boundary element method for time-harmonic exterior acoustics in two dimensions*, Comput. Methods Appl. Mech. Engrg. 191 (2002) 3331–3345.

[21] R. H. Nochetto and B. Stamm, *A posteriori error estimates for the Electric Field Integral Equation on Polyhedra*, ArXiv e-prints arXiv:1204.3930, 2012.

[22] J.-C. Nédélec, *Acoustic and electromagnetic equations : integral representations for harmonic problems*, Applied Mathematical Science, Springer New-York, 2001.

[23] W. McLean, *Strongly elliptic systems and boundary integral equations*, Cambridge University Press, 2000.

[24] X. Antoine and M. Darbas, *Generalized Combined Field Integral Equations for the Iterative Solution of the Three-Dimensional Helmholtz Equation*, Mathematical Modelling and Numerical Analysis **41** (2007), no. 1, 147–167.

[25] D. Levadoux, F. Millot and S. Pernet, *A well-conditioned boundary integral equation for transmission problems of electromagnetism*, J. Integral Equations Applications **27** (2015), no. 3, 431–454.

[26] P. Grisvard, *Singularities in boundary value problems*, Recherches en mathématiques appliquées (1992), Masson.

[27] S. Marburg and B. Nolte, *Discretization Requirements : How many Elements per Wavelength are Necessary ?*, chpt. 12 in *Computational Acoustics of Noise Propagation in Fluids – Finite and Boundary Element Methods*, 2008, Springer Berlin Heidelberg

[28] A. Bespalov, A. Haberl and D. Praetorius, *Adaptive FEM with Coarse Initial Mesh Guarantees Optimal Convergence Rates for Compactly Perturbed Elliptic Problems*, Comput. Methods Appl. Mech. Engrg. **317** (2017), 318–340.

[29] I. H. Sloan and A. Spence, *The Galerkin Method for Integral Equations of the First Kind with Logarithmic Kernel: Theory*, IMA J. Numer. Anal. **8** (1988), no. 1, 105–122.

Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 217 (2006) 340–363
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

A spatial high-order hexahedral discontinuous Galerkin method to solve Maxwell's equations in time domain

G. Cohen [a], X. Ferrieres [b,*], S. Pernet [b]

[a] INRIA, Domaine ds Voluceau, Rocquencourt – BP 105, 78153 Le Chesnay Cedex, France
[b] ONERA DEMR, unité CDE, 2 Avenue Edouard Belin, 31055 Toulouse, France

Received 19 May 2005; received in revised form 10 November 2005; accepted 4 January 2006
Available online 24 February 2006

Abstract

In this paper, we present a non-dissipative spatial high-order discontinuous Galerkin method to solve the Maxwell equations in the time domain. The non-intuitive choice of the space of approximation and the basis functions induce an important gain for mass, stiffness and jump matrices in terms of memory. This spatial approximation, combined with a leapfrog scheme in time, leads also to a fast explicit and accurate method. A study of the dispersive error is carried out and a stability condition for the proposed scheme is established. Some comparisons with other schemes are presented to validate the new scheme and to point out its advantages. Finally, in order to improve the efficiency of the method in terms of CPU time on general unstructured meshes, a strategy of local time-stepping is proposed.

Keywords: Numerical methods; Discontinuous Galerkin methods; Maxwell's equations in time domain; Conservative spatial centered scheme; Dispersive error; Stability analysis; Local time step

1. Introduction

The most widely used method for solving Maxwell's equations in the time domain is the finite difference time domain method (FDTD) based on the well-known scheme of Yee [1] and Taflove and Hagness [2]. This method involves an orthogonal Cartesian grid and is based on a second order leapfrog approximation in space and time. However, the FDTD method suffers a certain number of limitations such as, for example, difficulty in the treatment of curved objects. In such a case, the staircase approximation can generate spurious diffraction phenomena, which strongly damage the accuracy of the solution [3].

Numerous researchers and engineers have tried to develop efficient methods, which make it possible to take into account the complex shape of objects [4–9]. Moreover, the growing need to accurately model the propagation of electromagnetic waves over a large number of wavelengths (more than 100) has forced

* Corresponding author. Tel.: +33 05 62 25 28 02; fax: +33 05 62 25 25 77.
E-mail address: ferriere@onecert.fr (X. Ferrieres).

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2006.01.004

them to develop high-order or spectral methods [16,10,12]. Their first choice naturally turned towards finite element methods (FEM), which are a powerful tool for developing new numerical techniques [13]. However, one of the difficulties encountered in using finite element for Maxwell's equations is that of constructing a finite dimensional subspace of the continuous space $H(curl,\Omega)$. This functional space is natural for the solution of this problem, because the tangential components of a function belonging to $H(curl,\Omega)$ are continuous across any surface and the normal components of the same function may be discontinuous. It is well known that the use of classical continuous Langrage finite elements, which provide a suitable approximation of the space $[H^1(\Omega)]^3$, leads to spurious solutions. The appropriate finite element space – of so called edge element – was introduced by Nédélec in the 1980s [14,15]. Unfortunately, the classical version of these elements leads to a high computational cost since a matrix inversion is required at each time-step. This drawback increases with the order of approximation. Mass-lumping techniques seemed to be the right approach to avoid this inversion. An attractive method based on Nédélec's second family of edge elements was introduced by Cohen and Monk [16]. In this method, the use of the Gauss–Lobatto quadrature formulas yields a block-diagonal mass matrix, which enables us to obtain an explicit scheme for polynomial approximation at all orders. Unfortunately, this method produces important parasitic waves for large distortions of the cells. In the same idea, first and second order tetrahedral mass-lumped edge elements, which have no parasitic wave problems, were constructed in [17]. However, this approach seems to be efficient only for second-order elements.

The second choice is the use of discontinuous Galerkin methods (DGM). These methods were introduced in the first half of the 1970s by Reed and Hill [18] for the scalar neutron transport equation. Following this first study, many DGM were developed and analyzed by a large number of researchers in order to solve a large range of problems. One can find in [19] an exhaustive review of these methods from their beginning. In this survey, one can notice that few papers are devoted to the resolution of Maxwell's equations. In fact, the use of this kind of method to solve this problem is relatively recent. For the frequency domain, one can quote [21,20] and, for the time domain, one can quote [22,23,12] among numerous other papers. The main drawback of these methods is the large number of unknowns in each cell for high-order approximation schemes, which implies that much memory is needed for the local matrices (the mass matrix for example). Thus, in order to be efficient, the order of approximation in these methods must be limited. Hesthaven and Warburton [12] recently developed a low-storage, high-order, discontinuous Galerkin method for tetrahedral meshes with a judicious choice of the location of the degrees of freedom. However, his approach provides an algorithm that is $O(r^6)$ instead of $O(r^4)$ for hexahedra, r being the order of the method. One can notice that, before the use of these high-order methods, finite volume methods (that can be viewed as low order DGM schemes) were used to solve the Maxwell equations but these methods suffer from the presence of dissipation [24] or dispersion [26], which makes their use inaccurate for large-sized problems.

There are two approaches for implementing the discontinuous Galerkin methods: the h-version and the p-version. The h-version uses mesh refinement to achieve convergence to a fixed order, which keep the polynomial degree of the approximation fixed. The alternative p-version allows the order of polynomials to increase on a fixed mesh. A hybrid h–p version can also be considered. This paper concerns a method which takes to the third point of view and which enables us to reduce the storage requirement for local matrices, even for high-order approximation.

The outline of the paper is as follows. In Section 2, we describe a discontinuous Galerkin formulation for solving Maxwell's equations. In Section 3, we present some comparisons with other methods to validate this method and to show its advantages over other methods. Finally, in Section 4, a local time-stepping technique is proposed to enhance the performance of the method in terms of computational time.

2. The continuous formulation

2.1. The continuous problem

Let Ω be a domain on which the electric and magnetic fields $(\mathbf{E},\mathbf{H})$ satisfy:

$$\begin{cases} \varepsilon \frac{\partial \mathbf{E}}{\partial t} + \sigma \mathbf{E} = \nabla \times \mathbf{H}, \\ \mu \frac{\partial \mathbf{H}}{\partial t} = -\nabla \times \mathbf{E}, \\ \mathbf{E}(t=0) = 0, \quad \mathbf{H}(t=0) = 0. \end{cases} \tag{1}$$

On the boundary $\partial\Omega$ of the domain, we impose $\mathbf{n} \times \mathbf{E} = 0$ for cavity problems. This condition is also applied after PML [29] to simulate unbounded domains.

2.2. Definition of the discontinuous Galerkin framework

Let a set $\mathscr{T}_h$ of hexahedral elements $(K_i)_{i=1,N}$ be a partition of Ω. In our approach, Maxwell's equations are rewritten by adding two terms – which are equal to zero for the continuous problem – to each equation of (1). These terms define jumps of the electric and magnetic tangential components fields across the hexahedron K_i.

On each $K \in \mathscr{T}_h$, we get

$$\begin{cases} \varepsilon \frac{\partial \mathbf{E}}{\partial t} + \sigma \mathbf{E} = \nabla \times \mathbf{H} + \alpha [\![\mathbf{H} \times \mathbf{n}]\!]_{\partial K}^{K} \delta_{\partial K} + \beta [\![\mathbf{n} \times (\mathbf{E} \times \mathbf{n})]\!]_{\partial K}^{K} \delta_{\partial K}, \\ \mu \frac{\partial \mathbf{H}}{\partial t} = -\nabla \times \mathbf{E} + \gamma [\![\mathbf{E} \times \mathbf{n}]\!]_{\partial K}^{K} \delta_{\partial K} + \delta [\![\mathbf{n} \times (\mathbf{H} \times \mathbf{n})]\!]_{\partial K}^{K} \delta_{\partial K}, \end{cases} \tag{2}$$

where $[\![\mathbf{u}]\!]_{\partial K}^{K}$ defines the jump across the boundary ∂K of the volume K. More precisely, the jump is given by $[\![\mathbf{u}]\!]_{\partial K}^{K} = \mathbf{u}_K^+ - \mathbf{u}_K$ where $\mathbf{u}_K$ is the boundary value taken inside the volume K and $\mathbf{u}_K^+$ the same boundary value taken inside the other volume adjacent to ∂K. When $\Gamma = \partial K \cap \partial\Omega \neq \emptyset$, then $(\mathbf{u}_K^+)_{|\Gamma} = 0$. The term $\delta_{\partial K}$ is the Kronecker symbol for ∂K which is equal to 1 on ∂K and 0 elsewhere. It denotes the fact that these jump terms are added exclusively on the boundary of the elements. Now, we have to choose α, β, γ and δ so that (1) and (2) are equivalent problems.

Let the energy be defined by $\int_\Omega \varepsilon \mathbf{E} \cdot \mathbf{E}\,dx + \int_\Omega \mu \mathbf{H} \cdot \mathbf{H}\,dx$. For (1), this energy is constant when $\sigma = 0$. Now, we would like (2) also to satisfy an energy conservation principle. By using a weak formulation of Eqs. (2) when $\sigma = 0$, we can write for each element $K \in \mathscr{T}_h$:

$$\begin{cases} \int_K \mu \frac{\partial \mathbf{H}}{\partial t} \cdot \mathbf{H}\,dx = -\int_K \nabla \times \mathbf{E} \cdot \mathbf{H}\,dx + \gamma \int_{\partial K} [\![\mathbf{E} \times \mathbf{n}]\!]_{\partial K}^{K} \cdot \mathbf{H}\,ds + \delta \int_{\partial K} [\![\mathbf{n} \times (\mathbf{H} \times \mathbf{n})]\!]_{\partial K}^{K} \cdot \mathbf{H}\,ds, \\ \int_K \varepsilon \frac{\partial \mathbf{E}}{\partial t} \cdot \mathbf{E}\,dx = \int_K \nabla \times \mathbf{H} \cdot \mathbf{E}\,dx + \alpha \int_{\partial K} [\![\mathbf{H} \times \mathbf{n}]\!]_{\partial K}^{K} \cdot \mathbf{E}\,ds + \beta \int_{\partial K} [\![\mathbf{n} \times (\mathbf{E} \times \mathbf{n})]\!]_{\partial K}^{K} \cdot \mathbf{E}\,ds. \end{cases} \tag{3}$$

By adding the two equations over all the elements K and by integrating by part, we obtain:

$$\begin{aligned} &\sum_{K \in \mathscr{T}_h} \left(\int_K \varepsilon \frac{\partial \mathbf{E}}{\partial t} \cdot \mathbf{E}\,dx + \int_K \mu \frac{\partial \mathbf{H}}{\partial t} \cdot \mathbf{H}\,dx \right) \\ &\quad = \sum_{K \in \mathscr{T}_h} \int_{\partial K} (1 + \alpha - \gamma)(\mathbf{E}_K \times \mathbf{n}) \cdot \mathbf{H}_K\,ds - \sum_{\partial K \in \mathscr{F}_h^i} \int_{\partial K} \delta((\mathbf{E}_K + \mathbf{E}_K^+) \times \mathbf{n} \cdot (\mathbf{E}_K + \mathbf{E}_K^+) \times \mathbf{n})\,ds \\ &\qquad + \sum_{\partial K \in \mathscr{F}_h^i} \int_{\partial K} \beta((\mathbf{H}_K + \mathbf{H}_K^+) \times \mathbf{n} \cdot (\mathbf{H}_K + \mathbf{H}_K^+) \times \mathbf{n})\,ds + \sum_{\partial K \in \mathscr{F}_h^i} \int_{\partial K} (\alpha + \gamma)(\mathbf{E}_K^+ \times n \cdot \mathbf{H}_K + \mathbf{H}_K^+ \times \mathbf{n} \cdot \mathbf{E}_K)\,ds \\ &\qquad - \sum_{\partial K \in \mathscr{F}_h^b} \int_{\partial \hat{K}} (\delta(\mathbf{E}_K \times \mathbf{n} \cdot \mathbf{E}_K \times \mathbf{n}) + \beta(\mathbf{H}_K \times \mathbf{n} \cdot \mathbf{H}_K \times \mathbf{n}))\,ds, \end{aligned} \tag{4}$$

where $\mathscr{F}_h^i$ and $\mathscr{F}_h^b$ define, respectively, the faces inside the computational domain Ω and the faces of the boundary $\partial\Omega$.

To derive

$$\int_\Omega \varepsilon \frac{\partial \mathbf{E}}{\partial t} \mathbf{E}\,dx + \int_\Omega \mu \frac{\partial \mathbf{H}}{\partial t} \mathbf{H}\,dx = 0,$$

from Eq. (4), the values α, β, δ and γ must be such that $\beta = \delta = 0$, $1 + \alpha - \gamma = 0$ and $\alpha + \gamma = 0$ on faces belong to $\mathscr{F}_h^i$. On the other faces, energy conservation is ensured only if $\beta = \delta = 0$ and $1 + \alpha - \gamma = 0$. Hence, to guarantee energy conservation, we obtain $-\alpha = \gamma = \frac{1}{2}$ for faces belonging to $\mathscr{F}_h^i$ and we have different possibilities for α and γ for faces belonging to $\mathscr{F}_h^b$. For theses faces, our choice is guided by the equivalence between

problems (1) and (2). In particular, we set these coefficients in problem (2) so that boundary conditions of problem (1) are correctly taken into account. For example, in the case of a metallic boundary condition $\mathbf{E}_{|\partial\Omega}\times\mathbf{n}=0$ on a face in (1), we take $\gamma=1$ and then $\alpha=0$ on this face in (2) to have equivalence between the two problems.

So, an equivalent conservative formulation of (1) in each volume $K\in\mathscr{T}_h$ is given by

$$\begin{cases}\varepsilon\frac{\partial\mathbf{E}}{\partial t}+\sigma\mathbf{E}=\nabla\times\mathbf{H}+\alpha[\![\mathbf{H}\times\mathbf{n}]\!]_{\partial K}^{K}\delta_{\partial K},\\ \mu\frac{\partial\mathbf{H}}{\partial t}=-\nabla\times\mathbf{E}+\gamma[\![\mathbf{E}\times\mathbf{n}]\!]_{\partial K}^{K}\delta_{\partial K},\end{cases}\tag{5}$$

with the values of α and γ as previously defined.

3. Construction of the approximation

3.1. Approximate formulation

For the following of this paper, we assume each cell K in the $\mathscr{T}_h$ partition to be constituted by a homogeneous material, where the electric and magnetic fields are sufficiently regular to be considered in $(H^1(K))^3$. In order to define an approximation (5), we must first introduce the space

$$\boldsymbol{H}^1(\mathscr{T}_h)=\left\{\boldsymbol{v}\in(L^2(\Omega))^3;\forall K\in\mathscr{T}_h,\boldsymbol{v}_{|K}\in(H^1(K))^3\right\}.$$

Then, we can define the following variational formulation of (5).

Find $(\mathbf{E},\mathbf{H})\in(\boldsymbol{H}^1(\mathscr{T}_h))^2$ such that

$$\begin{cases}\varepsilon\frac{\partial}{\partial t}\int_\Omega\mathbf{E}\cdot\boldsymbol{\varphi}\,dx=-\int_\Omega\sigma\mathbf{E}\cdot\boldsymbol{\varphi}\,dx+\sum\limits_{K\in\mathscr{T}_h}\int_K\nabla\times\mathbf{H}\cdot\boldsymbol{\varphi}\,dx+\sum\limits_{K\in\mathscr{T}_h}\alpha\int_{\partial K}[\![\mathbf{H}\times\mathbf{n}]\!]_{\partial K}^{K}\cdot\boldsymbol{\varphi}\,ds,\\ \mu\frac{\partial}{\partial t}\int_\Omega\mathbf{H}\cdot\boldsymbol{\psi}\,dx=\sum\limits_{K\in\mathscr{T}_h}\left(-\int_K\nabla\times\mathbf{E}\cdot\boldsymbol{\psi}\,dx+\gamma\int_{\partial K}[\![\mathbf{E}\times\mathbf{n}]\!]\cdot\boldsymbol{\psi}\,ds\right),\end{cases}\tag{6}$$

where $\boldsymbol{\varphi}\in\boldsymbol{H}^1(\mathscr{T}_h)$ and $\boldsymbol{\psi}\in\boldsymbol{H}^1(\mathscr{T}_h)$.

In a second step, we define the approximation space of $\boldsymbol{H}^1(\mathscr{T}_h)$:

$$U_h=\left\{\boldsymbol{v}\in(L^2(\Omega))^3;\forall K\in\mathscr{T}_h,DF_K^*\boldsymbol{v}_{|K}o\boldsymbol{F}_K\in[Q_r(\hat{K})]^3\right\},$$

where $Q_r(\hat{K})$ is the set of polynomials of $\hat{K}=[0,1]^3$ whose order is less or equal to r in each variable. For any $K\in\mathscr{T}_h$, $\boldsymbol{F}_K$ is the conform mapping such that $\boldsymbol{F}_K(\hat{K})=K$ and DF_K the Jacobian matrix of $\boldsymbol{F}_K$. In the following, we shall denote $J_K=\det(DF_K)$ the Jacobian of $\boldsymbol{F}_K$. The definition of the approximate space U_h is not classical for discontinuous Galerkin methods (generally the solution is approximate by a polynom on each cell), but, as we shall see later, the use of the curl-conforming mapping in our approximation will be important to imply a low storage for the stiffness and jump matrices and a substantial gain on CPU time.

In this space, the following approximate formulation holds:

Find $(\mathbf{E}_h,\mathbf{H}_h)\in(U_h)^2$ such that

$$\begin{cases}\varepsilon\frac{\partial}{\partial t}\int_\Omega\mathbf{E}_h\cdot\boldsymbol{\varphi}_h\,dx=-\int_\Omega\sigma\mathbf{E}_h\cdot\boldsymbol{\varphi}_h\,dx+\sum\limits_{K\in\mathscr{T}_h}\int_K\nabla\times\mathbf{H}_h\cdot\boldsymbol{\varphi}_h\,dx+\sum\limits_{K\in\mathscr{T}_h}\alpha\int_{\partial K}[\![\mathbf{H}_h\times\mathbf{n}]\!]_{\partial K}^{K}\cdot\boldsymbol{\varphi}_h\,ds,\\ \mu\frac{\partial}{\partial t}\int_\Omega\mathbf{H}_h\cdot\boldsymbol{\psi}_h\,dx=\sum\limits_{K\in\mathscr{T}_h}\left(-\int_K\nabla\times\mathbf{E}_h\cdot\boldsymbol{\psi}_h\,dx+\gamma\int_{\partial K}[\![\mathbf{E}_h\times\mathbf{n}]\!]\cdot\boldsymbol{\psi}_h\,ds\right).\end{cases}\tag{7}$$

3.2. Basis functions and degrees of freedom

3.2.1. Basis functions on the unit cube

In order to define the basis functions of U_h, we first define the basis functions on the unit cube $\hat{K}$ (Fig. 1). Let $\vec{\hat{\xi}}_{ijk}=(\hat{\xi}_i,\hat{\xi}_j,\hat{\xi}_k)$, $1\leqslant i\leqslant r+1$, $1\leqslant j\leqslant r+1$, $1\leqslant k\leqslant r+1$, be a set of points of $\hat{K}$, where $\hat{\xi}_\ell$ represents the abscissa of a Gauss quadrature point on the interval $[0,1]$. On the other hand, we define the set of the $(r+1)^3$

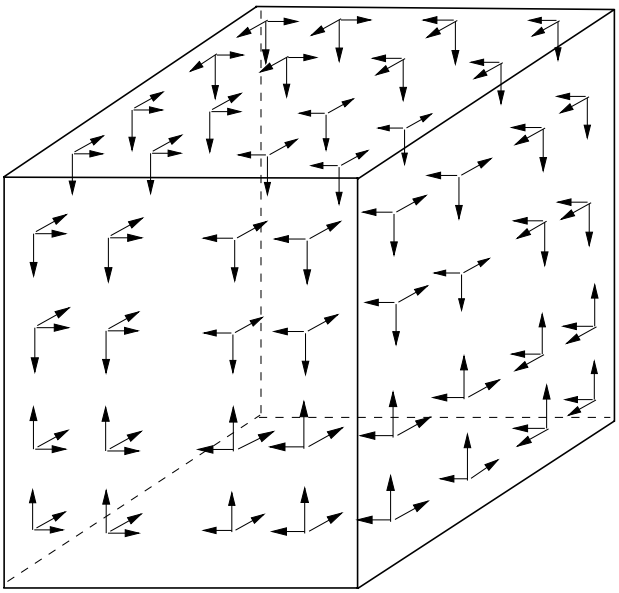


Fig. 1. Basis functions located on the unit cube for a Q_3 approximation.

Lagrange interpolation polynomials $\hat{\varphi}_{ijk} \in Q_r$ such that $\hat{\varphi}_{ijk}(\vec{\hat{\xi}}_{\ell,m,n}) = \delta_{i\ell}\delta_{jm}\delta_{kn}$, where δ_{ij} is the Kronecker symbol. We finally define the following set $\widehat{\mathscr{B}}$ of $3(r+1)^3$ vector-valued functions basis functions on $\hat{K}$: $\hat{\boldsymbol{\varphi}}^{(1)}_{ijk} = (\hat{\varphi}_{ijk}, 0, 0)^{\mathrm{T}}$, $\hat{\boldsymbol{\varphi}}^{(2)}_{ijk} = (0, \hat{\varphi}_{ijk}, 0)^{\mathrm{T}}$, $\hat{\boldsymbol{\varphi}}^{(3)}_{ijk} = (0, 0, \hat{\varphi}_{ijk})^{\mathrm{T}}$.

3.2.2. Basis functions on any hexahedron K

Following the definition of U_h, we can now deduce from $\widehat{\mathscr{B}}$ a basis $\mathscr{B}$ of this space. On each element K, we define a set of $3(r+1)^3$ basis functions $\boldsymbol{\varphi}^{(\ell)}_{ijk,K}$ such that $\boldsymbol{\varphi}^{(\ell)}_{ijk,K} = DF_K^{*-1}\hat{\boldsymbol{\varphi}}^{(\ell)}_{ijk}$, for all $\ell = 1, \ldots, 3$. So,

$$\mathscr{B} = \left\{ \boldsymbol{\varphi}^{(\ell)}_{ijk,K}, \forall K \in \mathscr{T}_h, \forall (ijk) \in \{1, \ldots, r+1\}^3, \forall \ell \in \{1,2,3\} \right\}.$$

In their definition, the L^2-character of the functions implies that the support of each basis function is reduced to one element. Thus, it is obvious that $\dim U_h = 3(r+1)^3 N_e$ for any mesh whose number of elements is N_e.

3.2.3. Mass matrices

In the weak formulation, we need to evaluate for each basis function $\boldsymbol{\varphi}^l_{ijk,K}(x)$ the discrete form of the terms $\int_\Omega \mathbf{E}(t,x) \cdot \boldsymbol{\varphi}^l_{ijk,K}(x)\,\mathrm{d}x$, $\int_\Omega \mathbf{H}(t,x) \cdot \boldsymbol{\varphi}^l_{ijk,K}(x)\,\mathrm{d}x$ and $\int_\Omega \sigma(x)\mathbf{E}(t,x) \cdot \boldsymbol{\varphi}^l_{ijk,K}(x)\,\mathrm{d}x$. These three integrals have a similar relation between basis functions and we apply the same developments to evaluate them.

So, we only explain how to obtain the discrete form of $\int_\Omega \mathbf{E}(t,x) \cdot \boldsymbol{\varphi}^l_{ijk,K}(x)\,\mathrm{d}x$.

By using $\mathbf{E}(x,t) o F_K = \sum_{K \in \mathscr{T}_h} \sum_{l=1}^{3} \sum_{i,j,k=1,r+1} E^l_{ijk,K}(t) (DF^*_K(\hat{x}))^{-1} \hat{\boldsymbol{\varphi}}^l_{ijk}(\hat{x})$ we can write

$$\int_\Omega \mathbf{E}(t,x) \cdot \boldsymbol{\varphi}^l_{ijk,K}(x)\,\mathrm{d}x = \int_K \mathbf{E}(t,x) \cdot \boldsymbol{\varphi}^l_{ijk,K}(x)\,\mathrm{d}x,$$

since $K = \mathrm{Supp}(\boldsymbol{\varphi}^l_{ijk,K})$.

So, we obtain

$$\begin{aligned}\int_K \mathbf{E}(t,x) \cdot \boldsymbol{\varphi}^l_{ijk,K}\,\mathrm{d}x &= \sum_{p=1}^{3} \sum_{m,n,q=1,N} E^p_{mnq,K}(t) \int_{\hat{K}} (DF^*_K)^{-1}(\hat{x}) \hat{\boldsymbol{\varphi}}^p_{mnq}(\hat{x}) \cdot (DF^*_K)^{-1}(\hat{x}) \hat{\boldsymbol{\varphi}}^l_{ijk}(\hat{x}) |J_K(\hat{x})|\,\mathrm{d}\hat{x} \\ &= \sum_{p=1}^{3} \sum_{m,n,q=1,N} E^p_{mnq,K}(t) \int_{\hat{K}} DF_K^{-1}(\hat{x}) (DF^*_K)^{-1}(\hat{x}) \hat{\boldsymbol{\varphi}}^p_{mnq}(\hat{x}) \cdot \hat{\boldsymbol{\varphi}}^l_{ijk}(\hat{x}) |J_K(\hat{x})|\,\mathrm{d}\hat{x}.\end{aligned}$$

By using the Gauss quadrature rule for the integral and the fact that $\hat{\boldsymbol{\varphi}}^l_{ijk}(\hat{x}_{mnq}) = \delta_i(\hat{x}_m)\delta_j(\hat{y}_n)\delta_k(\hat{z}_q)$, we finally get

$$\int_K \mathbf{E}(t,x)\cdot\boldsymbol{\varphi}^l_{ijk,K}\mathrm{d}x = \omega_{ijk}\sum_{p=1}^{3} E^p_{ijk,K}(DF_K^{-1}(\hat{x}_{ijk})(DF_K^*)^{-1}(\hat{x}_{ijk}))_{pl}|J_K(\hat{x}_{ijk})|$$

where ω_{ijk} is the quadrature weight at the point $\hat{x}_{ijk}$ and $(M)_{pl}$ defines the (p,l) term in the matrix M.

So, in the matrix of the discrete problem for a degree of freedom, we have three non-zero terms for the basis function which are defined at the same given quadrature point. By choosing a numbering of the unknowns around the points, we obtain a 3×3 block-diagonal mass matrix which can be diagonal for regular elements.

3.2.4. Stiffness matrix

The terms of the weak formulation considered here are $\int_\Omega \nabla\times E(t,x)\cdot\boldsymbol{\varphi}^l_{ijk,K}(x)\,\mathrm{d}x$ for the electric equation and $\int_\Omega \nabla\times H(t,x)\cdot\boldsymbol{\varphi}^l_{ijk,K}(x)\,\mathrm{d}x$ for the magnetic equation. As in the previous section, we explain only the construction of the discrete form of the term related to the electric equation. A similar demonstration holds for the magnetic equation related terms. As for the evaluation of the mass matrix, we have

$$\int_\Omega \nabla\times E(t,x)\cdot\boldsymbol{\varphi}^l_{ijk,K}(x)\,\mathrm{d}x = \int_K \nabla\times E(t,x)\cdot\boldsymbol{\varphi}^l_{ijk,K}(x)\,\mathrm{d}x,$$

where K is the unique element on which $\boldsymbol{\varphi}^l_{ijk,K}(x)$ is not identically equal to zero.

$$\begin{aligned}
\int_K \nabla\times E(t,x)\cdot\boldsymbol{\varphi}^l_{ijk,K}(x)\,\mathrm{d}x &= \int_{\hat{K}} ((DF^*)^{-1}(\hat{x})\hat{\nabla})\times((DF^*)^{-1}(\hat{x})\hat{E}(t,\hat{x}))\cdot(DF^*)^{-1}(\hat{x})\hat{\boldsymbol{\varphi}}^l_{ijk}(\hat{x})|J(\hat{x})|\,\mathrm{d}\hat{x}\\
&= \int_{\hat{K}} \frac{DF(\hat{x})}{J(\hat{x})}(\hat{\nabla}\times\hat{E}(t,\hat{x}))\cdot(DF^*)^{-1}(\hat{x})\hat{\boldsymbol{\varphi}}^l_{ijk}|J(\hat{x})|\,\mathrm{d}\hat{x}\\
&= \int_{\hat{K}} \mathrm{sign}(J(\hat{x}))(\hat{\nabla}\times\hat{E}(t,\hat{x}))\cdot\hat{\boldsymbol{\varphi}}^l_{ijk}(\hat{x})\,\mathrm{d}\hat{x}\\
&= \sum_{m,n,q=1,N}\sum_{p=1}^{3}\mathrm{sign}(J(\hat{x}))\int_{\hat{K}} E^p_{mnq,K}(t)(\hat{\nabla}\times\hat{\boldsymbol{\varphi}}^p_{mnq}(\hat{x}))\cdot\hat{\boldsymbol{\varphi}}^l_{ijk}(\hat{x})\,\mathrm{d}\hat{x}\\
&= \sum_{m,n,q=1,N}\sum_{p=1}^{3}\mathrm{sign}(J(\hat{x}_{mnq}))\omega_{mnq}E^l_{mnq,K}(t)(\hat{\nabla}\times\hat{\boldsymbol{\varphi}}^p_{mnq})^l(\hat{x}_{ijk}),
\end{aligned}$$

where $(\nabla\times u)^l$ is the l component of $\nabla\times u$.

We can see on this formula that, for all given elements K, the stiffness matrix is obtained only by the knowledge of the derivative term $\hat{\nabla}\times\hat{\boldsymbol{\varphi}}^l_{ijk}$ for all components and points on the reference element and by the sign of the Jacobian at each point on the element K. Because we assume that the inverse of the Jacobian always exists, the sign of the Jacobian must be the same on the element. So we only need to know the sign of the Jacobian at a given point on each element to have it for all points on the element. Then, for the stiffness matrix, we only need to store the derivative terms on the reference element and a sign for each element K. This implies a very small storage and a fast process to obtain the full stiffness matrix of the scheme.

3.2.5. Jump matrix

We denote the set of faces of $\mathscr{T}_h$ by $\mathscr{F}_h = \mathscr{F}^i_h\cup\mathscr{F}^b_h$, where $\mathscr{F}^i_h = \{\Gamma\in\mathscr{F}^i_h, \Gamma = K'\cap K\}$ and $\mathscr{F}^b_h = \{\Gamma\in\mathscr{F}^b_h, \Gamma = K\cap\partial\Omega\}$ are the sets of the interior and boundary faces.

The computation of the jump or flux terms is one of the most expensive parts in the time domain algorithms using finite volume or discontinuous Galerkin spatial approximation. An inappropriate computational approach or a bad formulation can dramatically penalize the computational code. We will see that the approximate space as well as the DGM formulation used leads to an efficient computation of the jump terms. Actually, a detailed computation of these terms will show that they need a negligible storage. Moreover, we will see that the number of operations to determine these terms is dramatically reduced, thanks to the correspondence of the basis functions with the selected quadrature points.

Let $\boldsymbol{\varphi}^l_{ijk,K}$ be a basis function defined by $\boldsymbol{\varphi}^l_{ijk,K}oF_K = DF_K^{*-1}\hat{\boldsymbol{\varphi}}^l_{ijk}$. The terms considered here are $\int_{\partial K}[\![\mathbf{H}\times\mathbf{n}]\!]\cdot\boldsymbol{\varphi}^l_{ijk,K}\,\mathrm{d}s$ and $\int_{\partial K}[\![\mathbf{E}\times\mathbf{n}]\!]\cdot\boldsymbol{\varphi}^l_{ijk,K}\,\mathrm{d}s$.

One can decompose the boundary, ∂K, of K in $\partial K = \bigcup_{i=1}^{6} \Gamma_i$, with $\Gamma_i \in \mathscr{F}$ for which $F_{K|\Gamma_i}(\hat{\Gamma}_i) = \Gamma_i$ with $\hat{\Gamma}_i \subset \partial\hat{K}$ for $i = 1, \ldots, 6$.

We can write

$$\int_{\partial K} [\![H_h \times \mathbf{n}_K]\!]^K_{\partial K} \cdot \boldsymbol{\varphi}^l_{ijk,K} \, \mathrm{d}\sigma = \sum_{i=1}^{6} \int_{\Gamma_i} [\![H_h \times \mathbf{n}^K_i]\!]^K_{\Gamma_i} \cdot \boldsymbol{\varphi}^l_{ijk,K} \, \mathrm{d}\sigma_i, \tag{8}$$

where $\mathrm{d}\sigma_i$ is the surface element of the face Γ_i and n^K_i the unit outward normal to K associated with Γ_i.

Let Γ_i be a face of K so that $\mathbf{F}_{K|\Gamma_i}(\hat{\Gamma}_i) = \Gamma_i$, we have $\mathrm{d}\sigma_i = |J_{K|\hat{\Gamma}_i}| \| DF^{*-1}_{K|\hat{\Gamma}_i} \hat{n}_i \| \, \mathrm{d}\hat{\eta}_i \, \mathrm{d}\hat{\chi}_i$ where $\hat{n}_i$ is the unit outward normal to $\hat{K}$ associated with the reference face $\hat{\Gamma}_i$ and $\hat{\eta}$, $\hat{\chi}$ are the tangential components of this face (i.e., $\hat{x}$, $\hat{y}$ or $\hat{z}$).

The definition of the basis functions also provides the following property (see [11]):

Let $\mathbf{u}_h \in U_h$, we have

$$(\mathbf{u}_{h|K} \times \mathbf{n}^K_i) o F_{K|\hat{\Gamma}_i} = \frac{1}{J_{K|\hat{\Gamma}_i} \| DF^{*-1}_{K|\hat{\Gamma}_i} \hat{\mathbf{n}}_i \|} DF_{K|\hat{\Gamma}_i} (\hat{\mathbf{u}}_{K|\hat{\Gamma}_i} \times \hat{\mathbf{n}}_i), \tag{9}$$

where $\hat{\mathbf{u}}_K = \sum_{l=1}^{3} \sum_{i,j,k}^{r+1} u^l_{ijk} \hat{\boldsymbol{\varphi}}^l_{ijk}$ and $\mathbf{n}^K_i$ is the unit outward normal to K associated with the face Γ_i.

By using these two previous properties, we can prove the following proposition:

Proposition 1. *$\forall \Gamma_i \in \mathscr{F}^b_h$, we have*

$$\int_{\Gamma_i} [\![\mathbf{H}_h \times \mathbf{n}_i]\!]^K_{\Gamma_i} \cdot \boldsymbol{\varphi}^l_{ijk} \, \mathrm{d}\sigma_i = -\mathrm{sign}(J_K) \int_{\hat{\Gamma}_i} (\hat{\mathbf{H}}_{K|\hat{\Gamma}_i} \times \hat{\mathbf{n}}_i) \cdot \hat{\boldsymbol{\varphi}}^l_{ijk} \, \mathrm{d}\hat{\eta}_i \, \mathrm{d}\hat{\chi}_i. \tag{10}$$

In this proposition, evaluating the term $\int_{\hat{\Gamma}_i} (\hat{\mathbf{H}}_{K|\hat{\Gamma}_i} \times \hat{n}_i) \cdot \hat{\boldsymbol{\varphi}}^l_{ijk} \, \mathrm{d}\hat{\eta}_i \, \mathrm{d}\hat{\chi}_i$ needs a very small number of operations. Take for example a face $\hat{\Gamma}_i = \{\hat{z} = 0\}$, we obtain by using a Gauss quadrature rule:

- if $l = 3$, then $\int_{\hat{\Gamma}_i} (\hat{\mathbf{H}}_{K|\hat{\Gamma}_i} \times \hat{\mathbf{n}}_i) \cdot \hat{\boldsymbol{\varphi}}^l_{ijk} \, \mathrm{d}\hat{x}_1 \, \mathrm{d}\hat{x}_2 = 0$,
- if $l = 1$, then $\int_{\hat{\Gamma}_i} (\hat{\mathbf{H}}_{K|\hat{\Gamma}_i} \times \hat{\mathbf{n}}_i) \cdot \hat{\boldsymbol{\varphi}}^l_{ijk} \, \mathrm{d}\hat{x}_1 \mathrm{d}\hat{x}_2 = -\sum_{l_3=1}^{r+1} H^2_{ijl_3,K} \hat{\omega}_i \hat{\omega}_j \hat{\varphi}_{l_3}(0) \hat{\varphi}_k(0)$,
- if $l = 2$, then $\int_{\hat{\Gamma}_i} (\hat{\mathbf{H}}_{K|\hat{\Gamma}_i} \times \hat{\mathbf{n}}_i) \cdot \hat{\boldsymbol{\varphi}}^l_{ijk} \, \mathrm{d}\hat{x}_1 \, \mathrm{d}\hat{x}_2 = \sum_{l_3=1}^{r+1} H^1_{ijl_3,K} \hat{\omega}_i \hat{\omega}_j \hat{\varphi}_{l_3}(0) \hat{\varphi}_k(0)$,

where $\{\hat{\omega}_i, \hat{\omega}_j\}_{l,m=1,\ldots,r+1}$ are the quadrature weights.

One has the same type of results for all the other faces of reference. One can see that the computation of the surface term coming from a boundary face requires little storage. Indeed, only the sign of Jacobian and negligible basis functions interactions on the reference faces must be known. Finally, the previous expressions show that the number of interactions between the basis functions is limited (only $r_k + 1$ interactions). This induces a substantial gain of CPU time for the computation of these quantities.

Now, we are going to see that the same conclusions are reached when computing the jump terms for the interior faces.

Let $\Gamma_i \in \mathscr{F}^i_h$ where $\Gamma_i = K \cap K'$, then we have

$$\int_{\Gamma_i} [\![H_h \times \mathbf{n}_i]\!]^K_{\Gamma_i} \cdot \boldsymbol{\varphi}^l_{ijk,K} \, \mathrm{d}\sigma_i = \int_{\hat{\Gamma}_i} |J_{K|\hat{\Gamma}_i}| \| DF^{*-1}_{K|\hat{\Gamma}_i} \hat{\mathbf{n}}_i \| (\mathbf{H}_{hK'} \times \mathbf{n}_i - \mathbf{H}_{hK} \times \mathbf{n}_i) o F_{K|\hat{\Gamma}_i} \cdot DF^{*-1}_K \hat{\boldsymbol{\varphi}}^l_{ijk} \, \mathrm{d}\hat{\eta}_i \, \mathrm{d}\hat{\chi}_i. \tag{11}$$

In (11), the computation of the part of the integral containing $\mathbf{H}_{hK}$ is done in the same way as for a boundary face. So, the previous proposition can be applied to this case and we obtain the same conclusions. Thus, we are only interested in the $\mathbf{H}_{hK'}$ term, i.e.,

$$I = \int_{\hat{\Gamma}_i} |J_{K|\hat{\Gamma}_i}| \| DF^{*-1}_{K|\hat{\Gamma}_i} \hat{\mathbf{n}}_i \| (\mathbf{H}_{hK'} \times \mathbf{n}_i) \cdot DF^{*-1}_K \hat{\boldsymbol{\varphi}}^l_{ijk} \, \mathrm{d}\hat{\eta}_i \, \mathrm{d}\hat{\chi}_i. \tag{12}$$

Let $i' \in \{1, \ldots, 6\}$ such that: $F_{K'}(\hat{\Gamma}_{i'}) = F_K(\hat{\Gamma}_i) = \Gamma_{i'} = \Gamma_i$. Let us consider the following change of variables:

$$\mathscr{G}_{K' \to K} = F^{-1}_{K|\Gamma_i} o F_{K'|\hat{\Gamma}_{i'}} : \hat{\Gamma}_{i'} \to \hat{\Gamma}_i. \tag{13}$$

Then $\forall \hat{x}' \in \hat{\Gamma}_{i'}$, we have

$$\left(|J_{K'|\hat{\Gamma}_{i'}}|\|DF^{*-1}_{K'|\hat{\Gamma}_{i'}}\hat{\mathbf{n}}_{i'}\|\right)(\hat{x}') = \left(|J_{K|\hat{\Gamma}_i}|\|DF^{*-1}_{K|\hat{\Gamma}_i}\hat{\mathbf{n}}_i\|\right) o\mathscr{G}_{K'\to K}(\hat{x}'). \tag{14}$$

The change of variables (13) and (9) gives

$$I = -\int_{\hat{\Gamma}_{i'}} \frac{\left(|J_{K|\hat{\Gamma}_i}|\|DF^{*-1}_{K|\hat{\Gamma}_i}\hat{\mathbf{n}}_i\|\right) o\mathscr{G}_{K'\to K}}{J_{K'|\hat{\Gamma}_{i'}}\|DF^{*-1}_{K'|\hat{\Gamma}_{i'}}\hat{\mathbf{n}}_{i'}\|} DF_{K'|\hat{\Gamma}_{i'}}(\hat{\mathbf{H}}_{K'|\hat{\Gamma}_{i'}} \times \hat{\mathbf{n}}_{i'}) \cdot (DF^{*-1}_K \hat{\boldsymbol{\varphi}}^l_{ijk}) o\mathscr{G}_{K'\to K}\, \mathrm{d}\hat{\eta}_{i'}\, \mathrm{d}\hat{\chi}_{i'}. \tag{15}$$

Property (14) leads to the simplification:

$$I = -\mathrm{sign}(J_{K'}) \int_{\hat{\Gamma}_{i'}} (DF^{-1}_K o\mathscr{G}_{K'\to K} DF_{K'})(\hat{\mathbf{H}}_{K'} \times \hat{\mathbf{n}}_{i'}) \cdot \hat{\boldsymbol{\varphi}}^l_{ijk} o\mathscr{G}_{K'\to K}\, \mathrm{d}\hat{\eta}_{i'}\, \mathrm{d}\hat{\chi}_{i'}. \tag{16}$$

Remark 1. $(DF^{-1}_K o\mathscr{G}_{K'\to K} DF_{K'})$ and $\mathscr{G}_{K'\to K}$ provide the connection with the local numbering of the degrees of freedom of K and K'. Both are permutation matrices constant per face.

So, in order to compute (16), it is sufficient to know the matrix $DF^{-1}_K o\mathscr{G}_{K'\to K} DF_{K'}$ on each internal face of $\mathscr{T}_h$. Its constant character implies that, to evaluate it, we only have to compute it at one point of this face.

3.2.6. Semi-discrete numerical scheme and 3D spatial dispersion analysis

In the previous sections, we described the different integral terms of the weak formulation and we showed that our choice of approximation spaces and basis functions led to a low storage algorithm, whenever using high order spatial approximations. In this section, we first give a full matrix representation of the semi-discrete numerical scheme obtained, then we provide an analysis of the spatial dispersive error of the scheme.

The semi-discrete numerical scheme proposed in this paper can be written as

$$\begin{aligned} M_\varepsilon \frac{\partial E}{\partial t} + M_\sigma E &= RH + S_E E, \\ M_\mu \frac{\partial H}{\partial t} &= -RE + S_H H \end{aligned} \tag{17}$$

where M_ε, M_μ, M_σ are 3×3 block-diagonal matrices, R is the stiffness matrix and S_E, S_H are the jump matrices whose terms are given in the previous sections. In these matrices, only the mass matrices must be stored because they depend on the element K. For the stiffness and jump matrices, we just have to store the sign of the Jacobian and a permutation matrix for each element K. Finally, in terms of storage required by the method we roughly have

- $6\times(r+1)^3$ real values per cell for the unknowns,
- $3\times 6\times(r+1)^3$ real values per cell for the mass matrices ($6\times(r+1)^3$ values for an homogeneous non-lossy experiment),
- 1 value per cell for the sign of the Jacobian,
- $4\times(r+1)^3$ values per face for the fluxes.

This implies a total storage between $24\times(r+1)^3+1$ and $36\times(r+1)^3+1$ values per cell, where r is the spatial order of the scheme.

We also choose to consider a spatial conservative numerical scheme in order to avoid errors due to numerical dissipation. To complete the analysis of our spatial scheme, let us now study the spatial dispersive error of the scheme. We prove that the 3D dispersion is deduced from the 1D one. In particular, this result allows us to easily determine the rate of convergence in space on Cartesian grid of the GD scheme studied here.

For this purpose, we consider an infinite regular mesh of $\mathbb{R}^3$ with a space-step equal to h. For $\mathbf{p} = (p_1, p_2, p_3) \in \mathbb{Z}^3$, we denote the cell $[p_1 h, (p_1+1)h]\times[p_2 h, (p_2+1)h]\times[p_3 h, (p_3+1)h]$ by $I_{\mathbf{p}}$. On this mesh, we assume that the discrete solution can be written in the form of the following numerical plane wave (see for example [11]):

$$E^l_{ijk,\mathbf{p}} = E^l_{ijk} e^{i(\omega_h t - k_1 h p_1 - k_2 h p_2 - k_3 h p_3)} e^{-i(\hat{x}_i k_1 h + \hat{y}_j k_2 h + \hat{z}_k k_3 h)},$$
$$H^l_{ijk,\mathbf{p}} = H^l_{ijk} e^{i(\omega_h t - k_1 h p_1 - k_2 h p_2 - k_3 h p_3)} e^{-i(\hat{x}_i k_1 h + \hat{y}_j k_2 h + \hat{z}_k k_3 h)}, \tag{18}$$

where $\hat{x}_i$, $\hat{y}_j$ and $\hat{z}_k$ are Gauss–Lobatto quadrature points and $l = 1,2,3$ are the three vector components.

Now, by reporting (18) into the discrete system, we get for the first electric component an equation of the form:

$$w_h h \varepsilon E^1_{ijk} = \sum_{l_3=1}^{r+1} H^2_{i,j,l_3} B_{h,r}[k_3](k, l_3) - \sum_{l_2=1}^{r+1} H^3_{i,l_2,k} B_{h,r}[k_2](j, l_2), \tag{19}$$

where $B_{h,r}[k_3](k,l_3)$ and $B_{h,r}[k_2](j,l_2)$ are geometric terms given by the discrete scheme.

Let A and B be $m \times n$ and $p \times q$ matrices, respectively. The Kronecker product is the $mp \times nq$ matrix $C = A \otimes B$ given by: $C_{(i-1)p+l,(j-1)q+r} = A_{i,j} B_{l,r}$ By using the Kronecker product, we can still write Eq. (19):

$$w_h h \varepsilon E^1 = (I_{r+1} \otimes I_{r+1} \otimes B_{h,r}[k_3]) H^2 - (I_{r+1} \otimes B_{h,r}[k_2] \otimes I_{r+1}) H^3, \tag{20}$$

where $E^l = E^l_{ijk,p}$ and $H^l = H^l_{ijk,p}$ for $l = 1,2,3$, and $i,j,k = 1,\ldots,r+1$.

By working in the same way for the other components, one obtains the matrix system:

$$\omega_h h \begin{pmatrix} \varepsilon E \\ \mu H \end{pmatrix} = \begin{pmatrix} 0 & A \\ -A & 0 \end{pmatrix} \begin{pmatrix} E \\ H \end{pmatrix}, \tag{21}$$

with $E = (E^l)_{l=1,3}$, $H = (H^l)_{l=1,3}$ and

$$A = \begin{pmatrix} 0 & a_3 & -a_2 \\ -a_3 & 0 & a_1 \\ a_2 & -a_1 & 0 \end{pmatrix}.$$

The terms a_1, a_2 and a_3 are, respectively, given by $B_{h,r}[k_1] \otimes I_{r+1} \otimes I_{r+1}$, $I_{r+1} \otimes B_{h,r}[k_2] \otimes I_{r+1}$ and $I_{r+1} \otimes I_{r+1} \otimes B_{h,r}[k_3]$ Finally, by using (21), we obtain the fundamental relations $\forall l = 1,\ldots,3$

$$\varepsilon\mu h^2 \omega_h^2 E^l = \alpha E^l,$$
$$\varepsilon\mu h^2 \omega_h^2 H^l = \alpha H^l, \tag{22}$$

where $\alpha = ((B_{h,r}[k_1])^2 \otimes I_{r+1} \otimes I_{r+1}) + (I_{r+1} \otimes (B_{h,r}[k_2])^2 \otimes I_{r+1}) + (I_{r+1} \otimes I_{r+1} \otimes (B_{h,r}[k_3])^2)$. Then, we obtain the following theorem:

Theorem 1. *Let ω_h be an eigenvalue of the spectral problem (22), then we have*

$$\omega_h^2 = \frac{1}{c^2 h^2} (\omega_h[k_1]^2 + \omega_h[k_2]^2 + \omega_h[k_3]^2), \tag{23}$$

where $\omega_h[k_i]^2$ are not only the eigenvalues of the matrix $(B_{h,r}[k_i])^2$, but also the eigenvalues of the 1D dispersion problem associated with the wavevectors k_i and $c = \frac{1}{\sqrt{\varepsilon\mu}}$.

In conclusion, the 3D dispersive properties are similar to those established in 1D. In particular, we obtain the same orders of approximation. Table 1 lists the rate of convergence ($\mathcal{O}(h^{p_r})$) obtained by using a 1D dispersion analysis of our DG approximation.

In this table, we can notice that dispersive error of the scheme decreases as the order of the scheme increases. However, these values seem to indicate that the accuracy of the scheme is the same for orders 2 and 3 and for orders 4 and 5. This is not exactly true. Indeed, we can write the dispersion error as kh^n where k, h and n are, respectively, a constant, the spatial step and the order of the scheme. For the different orders

Table 1
Numerical dispersion orders for different orders of approximation

Order of approximation	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$
Order of dispersive error	2	6	6	10	10

$n = 1,2,3,4,5$, the evaluated constant k is, respectively, equal to 3.2898, −2670.1817, 312.48, −931215.53 and 47988.2293. Taking into account these constant values, we can see that order 3 (respectively, 5) is better than order 2 (respectively, 4).

The pattern of the dispersive orders obtained in Table 1, is related to the choice of the centered fluxes. A study of the dispersive errors was also made by using non-centered fluxes ($\beta = \delta = \frac{1}{2}$) with our approximation. The values obtained are shown in Table 2. These results show that the approach of the non-centered fluxes is more attractive in terms of dispersive errors, but we also have an important additional cost storage for jump matrices due to the terms $[\![n \times (E \times n)]\!]$ and $[\![n \times (H \times n)]\!]$ and the scheme becomes dissipative. In these conditions, the centered fluxes scheme of the stays to be more interesting.

3.2.7. Numerical approximation in time and stability analysis

The most natural way for time discretization would be to use higher-order finite difference schemes. Generally, such explicit schemes are unstable [27]. In the domain of centered schemes (which are not dissipative and coherent with our choice), very few alternatives remain. The stable modified equation approach or symmetric schemes are complicated and are very difficult to adapt to unbounded domain conditions such as ABC or PML [11]. For these reasons, we decided to use a simple leapfrog scheme in time. The problem is to know if this low approximation in time compared to the one in space can introduce errors, which annihilate the advantage to use spatial schemes of order higher than 2. Numerical experiments have shown that it is not the case. Although of second-order, the time scheme appeared to be accurate for relatively long time-requiring experiments ($\simeq$50 wavelengths). This is mainly due to the fact that we do not use it with its maximal CFL on non-regular meshes, in which the size of the elements can vary with a ratio of 10 or even more and the CFL is adapted to the smallest cell. Of course, for longer experiments, a phase-shift, which increases with time, can be observed.

To illustrate this purpose, Fig. 2 shows the evolution of the L^2 error for different spatial orders. We can see that increasing the order of the spatial approximation allows us to have a more accurate solution. The L^2 error defines here the L^2 norm of the difference between the exact solution and the solution computed for each degree of freedom in the problem. The configuration and the analytic solution of the example used in this figure are the same as the example proposed for the evaluation of modes inside cavity in numerical results section. Another possibility to quantitatively evaluate the accuracy of our approximation consists in seeing the ability of the scheme to restitute the free-divergence condition for Maxwell equations ($\nabla \cdot E = 0$). To consider this point, we computed, for different meshes of the same example of cavity and different orders of approximation, an equivalent H^{-1} norm of the divergence of the electric discrete fields E_h, proposed by Cockburn et al. [25]:

$$\|\mathbf{E}_h\|_{*,h} = \sum_{\Gamma \in \mathbf{F}} \int_{\Gamma}^{G} |[\![\mathbf{E}_h \cdot \mathbf{n}]\!]| + \sum_{\Gamma \in \mathscr{T}_h} \int_{K}^{G} |\nabla \cdot \mathbf{E}_h|.$$

Figs. 3 and 4 show the results obtained by using different sizes of cells in the mesh and different orders of approximation, respectively. These results confirm the interest in using high spatial approximation order despite order 2 in time.

In the previous results, the error is principally due to the dispersive phenomena, since it can be easily be proven that our numerical scheme, with a leap-frog scheme in time, remains conservative. The energetic quantity taken into account for this proof is given by

$$\sum_{K \in \mathscr{T}_h} \int_K \varepsilon \mathbf{E}_K^n \cdot \mathbf{E}_K^n \, \mathrm{d}K + \int_K \mu_0 \mathbf{H}_K^{n+\frac{1}{2}} \cdot \mathbf{H}_K^{n-\frac{1}{2}} \mathrm{d}K,$$

where $(\mathbf{E}_K^n, \mathbf{H}_K^n)$ are electric and magnetic fields values taken on K at time t_n (see [26]).

Table 2
Numerical dispersion and dissipation orders by using non-centered fluxes

Order of approximation	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$
Order of dispersive error	4	6	8	10	12
Order of dissipative error	3	5	7	9	11

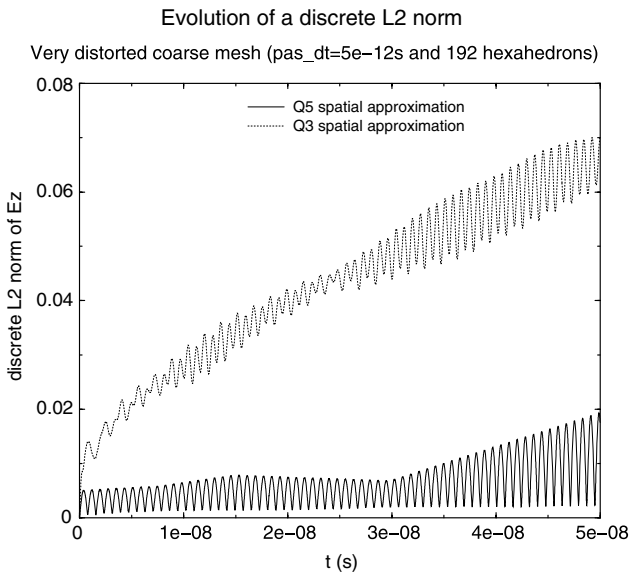


Fig. 2. Evolution of the L^2 error in time for spatial order equal to 3 and 5.

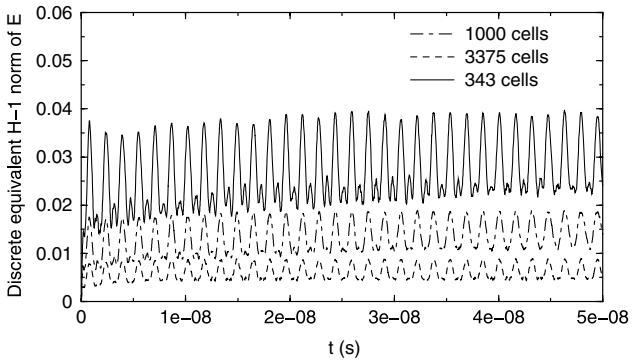


Fig. 3. h-Convergence by using a Q_3 order.

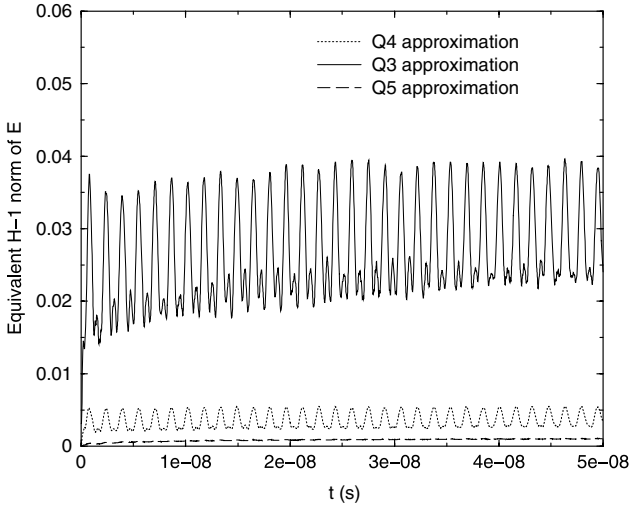


Fig. 4. p-Convergence by using mesh constituted of 343 cells.

Now, we are interested in the stability of the fully discrete scheme. In a first step, we study the stability by a plane wave technique on a homogeneous infinite grid. This technique provides necessary and sufficient stability conditions of the scheme on a regular mesh. In a second step, we use an energy technique to prove the stability on non-structured hexahedral meshes. We get a sufficient stability condition. Moreover, we show that our scheme conserves a discrete energy. This property confirms the non-dissipative nature of the scheme.

3.2.7.1. CFL conditions obtained by a plane waves technique. For a Cartesian homogeneous grid, we proceed as for the spatial dispersion analysis except that we add the time discretization. Then, the numerical plane wave becomes:

$$\begin{aligned}(E^n_{i,j,k,\mathbf{p}})^l &= E^l_{ijk} e^{\mathrm{i}(\omega_h n\Delta t - k_1 h p_1 - k_2 h p_2 - k_3 h p_3)} e^{-\mathrm{i}(\hat{x}_i k_1 h + \hat{y}_j k_2 h + \hat{z}_k k_3 h)},\\ (H^{n-\frac{1}{2}}_{ijk,\mathbf{p}})^l &= H^l_{ijk} e^{\mathrm{i}(\omega_h \left(n-\frac{1}{2}\right)\Delta t - k_1 h p_1 - k_2 h p_2 - k_3 h p_3)} e^{-\mathrm{i}(\hat{x}_i k_1 h + \hat{y}_j k_2 h + \hat{z}_k k_3 h)}.\end{aligned} \tag{24}$$

So, spectral problem (22) is rewritten $\forall l = 1,\ldots,3$:

$$\begin{aligned}\frac{4h^2}{c^2\Delta t^2}\sin\left(\frac{\omega_h \Delta t}{2}\right) E^l &= \alpha E^l,\\ \frac{4h^2}{c^2\Delta t^2}\sin\left(\frac{\omega_h \Delta t}{2}\right) H^l &= \alpha H^l.\end{aligned} \tag{25}$$

So, we deduce the space-time dispersion relations:

$$\sin^2\left(\frac{\omega_h \Delta t}{2}\right) = \frac{\Delta t^2 c^2}{4h^2}\left(\omega_h[k_1]^2 + \omega_h[k_2]^2 + \omega_h[k_3]^2\right), \tag{26}$$

where $\omega_h[k_i]^2$ is defined as in Theorem 1.

Finally, the numerical scheme is stable if and only if

$$\frac{\Delta t}{h} \leqslant \frac{2}{c}\frac{1}{\max\left(\sqrt{\omega_h[k_1]^2 + \omega_h[k_2]^2 + \omega_h[k_3]^2}\right)} = \frac{c_r}{\sqrt{3}}, \tag{27}$$

where the max is taken on all the eigenvalues of the 1D problem for $i = 1,2,3$ and c_r corresponds to CFL number for the same 1D scheme using an order r of polynomial approximation.

The values of c_r can be numerically determined and for $r = 1$, 2, 3, 4 and 5 we obtain 0.5, 0.247, 0.15, 0.101 and 0.0732, respectively.

It is thus sufficient to divide by $\sqrt{3}$ to obtain 3D CFL conditions.

3.2.7.2. Stability by energy technique. To prove the stability of the DG scheme by an energy technique, we use the same approach as the one used by Piperno et al. in [26] on a non-dissipative finite volume scheme.

The discrete energy, which naturally appears when one uses a leapfrog scheme for the time domain approximation, is defined by

$$\mathscr{E}^n_h = \sum_{K\in\mathscr{T}_h} \mathscr{E}^n_K, \tag{28}$$

where $\forall K \in \mathscr{T}_h$,

$$\mathscr{E}^n_K = \int_K^G \varepsilon_K \mathbf{E}^n_{hK}\cdot\mathbf{E}^n_{hK}\,\mathrm{d}x + \int_K^G \mu_K \mathbf{H}^{n+\frac{1}{2}}_{hK}\cdot\mathbf{H}^{n-\frac{1}{2}}_{hK}\,\mathrm{d}x. \tag{29}$$

We easily show that this energy is conserved during the discrete time, i.e., $\forall n \geqslant 0$, we have

$$\mathscr{E}^{n+1}_h - \mathscr{E}^n_h = 0, \tag{30}$$

and that it can only be expressed as a function of the variables $\tilde{\mathbf{E}}^n_h$ and $\tilde{\mathbf{H}}^{n+\frac{1}{2}}_h$:

$$\mathscr{E}_h^n = \sum_{K\in\mathscr{T}_h}\left[\int_K^G \tilde{\mathbf{E}}_{hK}^n\cdot\tilde{\mathbf{E}}_{hK}^n\,\mathrm{d}x + \int_K^G \tilde{\mathbf{H}}_{hK}^{n+\frac{1}{2}}\cdot\tilde{\mathbf{H}}_{hK}^{n+\frac{1}{2}}\,\mathrm{d}x + \frac{\Delta t}{2\sqrt{\varepsilon_K\mu_K}}\int_K^G \nabla\times\tilde{\mathbf{E}}_{hK}^n\cdot\tilde{\mathbf{H}}_{hK}^{n+\frac{1}{2}}\,\mathrm{d}x\right.$$
$$+\frac{\Delta t}{2\sqrt{\varepsilon_K\mu_K}}\int_K^G \tilde{\mathbf{E}}_{hK}^n\cdot\nabla\times\tilde{\mathbf{H}}_{hK}^{n+\frac{1}{2}}\,\mathrm{d}x - \frac{\Delta t}{2\sqrt{\varepsilon_K\mu_K}}\sum_{i=1}^{\mathrm{nbfi}_K}\int_{\Gamma_{\rho(i,K)}}^G(\tilde{\mathbf{E}}_{hV(i,K)}^n\times\mathbf{n}_K)\cdot\tilde{\mathbf{H}}_{hK}^{n+\frac{1}{2}}\,\mathrm{d}\sigma_{\rho(i,K)}$$
$$\left.+\frac{\Delta t}{2\sqrt{\varepsilon_K\mu_K}}\sum_{i=1}^{6}\int_{\Gamma_{\rho(i,K)}}^G(\tilde{\mathbf{E}}_{hK}^n\times\mathbf{n}_K)\cdot\tilde{\mathbf{H}}_{hK}^{n+\frac{1}{2}}\,\mathrm{d}\sigma_{\rho(i,K)}\right], \tag{31}$$

where $\tilde{\mathbf{E}}_{hK}^n = \mathbf{E}_{hK}^n\sqrt{\varepsilon_K}$, $\tilde{\mathbf{H}}_{hK}^{n+\frac{1}{2}} = \mathbf{H}_{hK}^{n+\frac{1}{2}}\sqrt{\mu_K}$, $V(i,K)$ is the neighbor of K containing the face $\Gamma_{\rho(i,K)}$ and nbfi_K is the number of faces of K belonging to $\mathscr{F}_h^i$.

From now, we can eliminate the exponent in n and the tildes in the notation, since we study the quadratic form (31) for all the variables $\mathbf{E}_h$, $\mathbf{H}_h$. To prove the L^2-stability, it would suffices if we determine a condition for which this quadratic form is definite positive. In this case, $\mathscr{E}_h^n$ will define a norm.

Proposition 2. *We have the estimate*

$$\mathscr{E}_h \geqslant \hat{\mathscr{E}}_h, \tag{32}$$

where

$$\hat{\mathscr{E}}_h = \sum_{K\in\mathscr{T}_h}\left[\Lambda_K\int_{\hat{K}}^G\hat{\mathbf{E}}_K\cdot\hat{\mathbf{E}}_K\,\mathrm{d}\hat{x} + \Lambda_K\int_{\hat{K}}^G\hat{\mathbf{H}}_K\cdot\hat{\mathbf{H}}_K\,\mathrm{d}\hat{x} + \frac{\Delta t}{2\sqrt{\varepsilon_K\mu_K}}\mathrm{sign}(J_K)\int_{\hat{K}}^G\hat{\nabla}\times\hat{\mathbf{E}}_K\cdot\hat{\mathbf{H}}_K\,\mathrm{d}\hat{x}\right.$$
$$+\frac{\Delta t}{2\sqrt{\varepsilon_K\mu_K}}\mathrm{sign}(J_K)\int_{\hat{K}}^G\hat{\mathbf{E}}_K\cdot\hat{\nabla}\times\hat{\mathbf{H}}_K\,\mathrm{d}\hat{x} + \sum_{i=1}^{\mathrm{nbfi}_K}\frac{\Delta t}{2\sqrt{\varepsilon_{V(i,K)}\mu_K}}\mathrm{sign}(J_{V(i,K)})$$
$$\left.\times\int_{\hat{\Gamma}_{i'}}^G(\hat{\mathbf{E}}_{V(i,K)}\times\hat{\mathbf{n}}_{i'})\cdot N_{V(i,K)\to K}^*\hat{\mathbf{H}}_K o\mathscr{G}_{V(i,K)\to K}\,\mathrm{d}\hat{\sigma}_{i'} + \frac{\Delta t}{2\sqrt{\varepsilon_K\mu_K}}\sum_{i=\mathrm{nbfi}_K+1}^{6}\mathrm{sign}(J_K)\int_{\hat{\Gamma}_i}^G(\hat{\mathbf{E}}_K\times\hat{\mathbf{n}}_i)\cdot\hat{\mathbf{H}}_K\,\mathrm{d}\hat{\sigma}_i\right] \tag{33}$$

and

$$\Lambda_K = \min_{1\leqslant i,j,k\leqslant r+1}\left(\frac{|J_K(\hat{x}_i,\hat{y}_j,\hat{z}_k)|}{\lambda_{\max}((DF_K^*DF_K)(\hat{x}_i,\hat{y}_j,\hat{z}_k))}\right) \tag{34}$$

with $\lambda_{\max}(DF_K^*DF_K)$ the greatest eigenvalue of $DF_K^*DF_K$.

Proof. To show Λ_k, it would suffices if we to use the expression of the mass terms and take the minimum of these terms on the Gauss points. Then, we use the properties of stiffness and jump matrices previously, which were presented to prove the result. □

Let $\hat{\mathscr{R}}$, $\hat{\mathscr{D}}$ and $\hat{\mathscr{B}}$ be the $3(r+1)^3\times 3(r+1)^3$ matrices defined by: $\forall i,i'\in\{1,2,3\}$ and $\forall\mathbf{l},\mathbf{l}'\in\{1,\ldots,r+1\}^3$

$$\hat{\mathscr{R}}((i,\mathbf{l}),(i',\mathbf{l}')) = \int_{\hat{K}}\hat{\nabla}\times\hat{\boldsymbol{\varphi}}_{\mathbf{l}}^i\cdot\hat{\nabla}\times\hat{\boldsymbol{\varphi}}_{\mathbf{l}'}^{i'}\,\mathrm{d}\hat{\mathbf{x}},$$
$$\hat{\mathscr{D}}((i,\mathbf{l}),(i',\mathbf{l}')) = \delta_{ii'}\delta_{\mathbf{l}\mathbf{l}'}\hat{\omega}_{\mathbf{l}}, \tag{35}$$
$$\hat{\mathscr{B}}((i,\mathbf{l}),(i',\mathbf{l}')) = \int_{\partial\hat{K}}(\hat{\boldsymbol{\varphi}}_{\mathbf{l}}^i\times\hat{\mathbf{n}})\cdot(\hat{\boldsymbol{\varphi}}_{\mathbf{l}'}^{i'}\times\hat{\mathbf{n}})\,\mathrm{d}\hat{\sigma}.$$

We are now going to estimate the terms of (33) in the function of $\|\hat{\mathbf{E}}_K\|_{0,\hat{K}}$ and $\|\hat{\mathbf{H}}_K\|_{0,\hat{K}}$ and to prove the L^2-stability:

Theorem 2. *The condition given by*

$$\frac{\Delta t}{\Lambda_K} < \frac{2}{c_K}\frac{1}{\sqrt{\lambda_{\max}\left(\hat{\mathscr{D}}^{-\frac{1}{2}}\hat{\mathscr{R}}\hat{\mathscr{D}}^{-\frac{1}{2}}\right)} + \frac{1}{2}\max\limits_{1\leqslant i\leqslant\mathrm{nbfi}_K}\left(\sqrt{\frac{\mu_K}{\mu_{V(i,K)}}},\sqrt{\frac{\varepsilon_K}{\varepsilon_{V(i,K)}}}\right)\lambda_{\max}\left(\hat{\mathscr{D}}^{-\frac{1}{2}}\hat{\mathscr{B}}\hat{\mathscr{D}}^{-\frac{1}{2}}\right)}, \tag{36}$$

$\forall K\in\mathscr{T}_h$, is sufficient to ensure the stability of the DG scheme, where $c_K = \frac{1}{\sqrt{\varepsilon_K\mu_K}}$

Proof. We have

$$\int_{\hat{K}}^{G} \hat{\nabla} \times \hat{\mathbf{E}}_K \cdot \hat{\mathbf{H}}_K \, \mathrm{d}\hat{x} \leqslant \|\hat{\nabla} \times \hat{\mathbf{E}}_K\|_{0,\hat{K}} \|\hat{\mathbf{H}}_K\|_{0,\hat{K}}. \tag{37}$$

Using the matrices defined in (35), we easily see that:

$$\int_{\hat{K}}^{G} \hat{\nabla} \times \hat{\mathbf{E}}_K \cdot \hat{\mathbf{H}}_K \, \mathrm{d}\hat{x} \leqslant \sqrt{\lambda_{\max}\left(\hat{\mathscr{D}}^{-\frac{1}{2}} \hat{\mathscr{R}} \hat{\mathscr{D}}^{-\frac{1}{2}}\right)} \|\hat{\mathbf{E}}_K\|_{0,\hat{K}} \|\hat{\mathbf{H}}_K\|_{0,\hat{K}}. \tag{38}$$

Now we estimate the surface terms:

$$\begin{aligned}
&\frac{1}{\sqrt{\varepsilon_{V(i,K)}\mu_K}} \int_{\hat{\Gamma}_{i'}}^{G} (\hat{\mathbf{E}}_{V(i,K)} \times \hat{\mathbf{n}}_{i'}) \cdot N^*_{V(i,K)\to K} \hat{\mathbf{H}}_K o \mathscr{G}_{V(i,K)\to K} \, \mathrm{d}\hat{\sigma}_{i'} \\
&\quad \leqslant \frac{1}{\sqrt{\varepsilon_{V(i,K)}\mu_K}} \|(\hat{\mathbf{E}}_{V(i,K)} \times \hat{\mathbf{n}}_{i'})\|_{0,\hat{\Gamma}_{i'}} \|(\hat{\mathbf{n}}_{i'} \times (N^*_{V(i,K)\to K} \hat{\mathbf{H}}_K o \mathscr{G}_{V(i,K)\to K} \times \hat{\mathbf{n}}_{i'}))\|_{0,\hat{\Gamma}_{i'}} \\
&\quad \leqslant \frac{1}{\sqrt{\varepsilon_{V(i,K)}\mu_K}} \|(\hat{\mathbf{E}}_{V(i,K)} \times \hat{\mathbf{n}}_{i'})\|_{0,\hat{\Gamma}_{i'}} \|(\hat{\mathbf{H}}_K \times \hat{\mathbf{n}}_i)\|_{0,\hat{\Gamma}_i} \\
&\quad \leqslant \frac{1}{2}\left[c_{V(i,K)} \sqrt{\frac{\mu_{V(i,K)}}{\mu_K}} \|(\hat{\mathbf{E}}_{V(i,K)} \times \hat{\mathbf{n}}_{i'})\|^2_{0,\hat{\Gamma}_{i'}} + c_K \sqrt{\frac{\varepsilon_K}{\varepsilon_{V(i,K)}}} \|(\hat{\mathbf{H}}_K \times \hat{\mathbf{n}}_i)\|^2_{0,\hat{\Gamma}_i} \right],
\end{aligned} \tag{39}$$

where $c_K = \frac{1}{\sqrt{\varepsilon_K \mu_K}}$.

In the same way, we obtain:

$$\int_{\hat{\Gamma}_i}^{G} (\hat{\mathbf{E}}_K \times \hat{\mathbf{n}}_i) \cdot \hat{\mathbf{H}}_K \, \mathrm{d}\hat{\sigma}_i \leqslant \frac{1}{2}\left[\|(\hat{\mathbf{E}}_K \times \hat{\mathbf{n}}_i)\|^2_{0,\hat{\Gamma}_i} + \|(\hat{\mathbf{H}}_K \times \hat{\mathbf{n}}_i)\|^2_{0,\hat{\Gamma}_i} \right]. \tag{40}$$

Finally, we have

$$\begin{aligned}
&\sum_{K\in\mathscr{T}_h} \left[-\sum_{i=1}^{\mathrm{nbfi}_K} \frac{1}{\sqrt{\varepsilon_{V(i,K)}\mu_K}} \mathrm{sign}(J_{V(i,K)}) \int_{\hat{\Gamma}_{i'}}^{G} (\hat{\mathbf{E}}_{V(i,K)} \times \hat{\mathbf{n}}_{i'}) \cdot N^*_{V(i,K)\to K} \hat{\mathbf{H}}_K o \mathscr{G}_{V(i,K)\to K} \, \mathrm{d}\hat{\sigma}_{i'} \right. \\
&\qquad \left. + \frac{1}{\sqrt{\varepsilon_K \mu_K}} \sum_{i=\mathrm{nbfi}_K+1}^{6} \mathrm{sign}(J_K) \int_{\hat{\Gamma}_i}^{G} (\hat{\mathbf{E}}_K \times \hat{\mathbf{n}}_i) \cdot \hat{\mathbf{H}}_K \, \mathrm{d}\hat{\sigma}_i \right] \\
&\qquad \leqslant \frac{1}{2} \sum_{K\in\mathscr{T}_h} c_K \max_{1\leqslant i\leqslant \mathrm{nbfi}_K} \left(\sqrt{\frac{\mu_K}{\mu_{V(i,K)}}}, \sqrt{\frac{\varepsilon_K}{\varepsilon_{V(i,K)}}} \right) \left[\|(\hat{\mathbf{E}}_K \times \hat{\mathbf{n}})\|^2_{0,\partial\hat{K}} + \|(\hat{\mathbf{H}}_K \times \hat{\mathbf{n}})\|^2_{0,\partial\hat{K}} \right] \\
&\qquad \leqslant \frac{\lambda_{\max}\left(\hat{\mathscr{D}}^{-\frac{1}{2}} \hat{\mathscr{B}} \hat{\mathscr{D}}^{-\frac{1}{2}}\right)}{2} \sum_{K\in\mathscr{T}_h} c_K \max_{1\leqslant i\leqslant \mathrm{nbfi}_K} \left(\sqrt{\frac{\mu_K}{\mu_{V(i,K)}}}, \sqrt{\frac{\varepsilon_K}{\varepsilon_{V(i,K)}}} \right) \left[\|\hat{\mathbf{E}}_K\|^2_{0,\hat{K}} + \|\hat{\mathbf{H}}_K\|^2_{0,\hat{K}} \right].
\end{aligned} \tag{41}$$

Eqs. (38) and (41) then lead to

$$\begin{aligned}
\hat{\mathscr{E}}_h \geqslant \sum_{K\in\mathscr{T}_h} \Lambda_K &\left[\left(1 - \frac{\Delta t c_K}{4\Lambda_K} \max_{1\leqslant i\leqslant \mathrm{nbfi}_K} \left(\sqrt{\frac{\mu_K}{\mu_{V(i,K)}}}, \sqrt{\frac{\varepsilon_K}{\varepsilon_{V(i,K)}}} \right) \lambda_{\max}\left(\hat{\mathscr{D}}^{-\frac{1}{2}} \hat{\mathscr{B}} \hat{\mathscr{D}}^{-\frac{1}{2}}\right) \right) \|\hat{\mathbf{E}}_K\|^2_{0,\hat{K}} \right. \\
&+ \left(1 - \frac{\Delta t c_K}{4\Lambda_K} \max_{1\leqslant i\leqslant \mathrm{nbfi}_K} \left(\sqrt{\frac{\mu_K}{\mu_{V(i,K)}}}, \sqrt{\frac{\varepsilon_K}{\varepsilon_{V(i,K)}}} \right) \lambda_{\max}\left(\hat{\mathscr{D}}^{-\frac{1}{2}} \hat{\mathscr{B}} \hat{\mathscr{D}}^{-\frac{1}{2}}\right) \right) \|\hat{\mathbf{H}}_K\|^2_{0,\hat{K}} \\
&\left. - \frac{\Delta t c_K}{\Lambda_K} \sqrt{\lambda_{\max}\left(\hat{\mathscr{D}}^{-\frac{1}{2}} \hat{\mathscr{R}} \hat{\mathscr{D}}^{-\frac{1}{2}}\right)} \|\hat{\mathbf{E}}_K\|_{0,\hat{K}} \|\hat{\mathbf{H}}_K\|_{0,\hat{K}} \right],
\end{aligned} \tag{42}$$

which ends the proof of the theorem. □

4. Numerical results

To validate and to illustrate the advantages of this method, several comparisons on simple scattering problems with different other methods are made. In these comparisons, we are interested in the fields near the structure (EMC problems) or by far fields (RCS) for scattering objects and by the fields inside cavities. The methods used to compare the results with our method are finite difference time domain (FDTD) [1], finite volume time domain (FVTD) [24] and marching on time (MOT) [4] methods.

4.1. Near fields and far fields for curved objects

The first example consists in evaluating the field scattered by a perfectly plate cone, at two test-points A and B located near the structure. Fig. 5 shows the location of the test-points around the object. The cone is illuminated with a plane wave given by $f(t) = E_y e^{-\gamma^2}$, with $\gamma = 3e8t - z + 6$. The incidence is along the axis z and the minimum wavelength is approximately 300 MHz.

Fig. 6 shows a comparison between the solutions obtained with the different methods.

In this comparison, to obtain an accurate result for the FDTD method, it is necessary to use a mesh with a spatial step size smaller than $0.025m = \lambda/40$. The classical $\lambda/10$ criterion is not sufficient to obtain a correct solution in this case.

The need to use a very small spatial step size implies a small time step (4.e−11 s) and a more important number of cells in the computational domain (596,232), which dramatically increases the memory storage (27 Mo) and the CPU time consumed in the FDTD method (99 min on a Pentium 4 at 3 GHz). In these conditions, the discontinuous Galerkin approach is more interesting (25 min with a Q_3-approximation for 15 Mo of memory storage).

For the FVTD method, the results obtained with a spatial averaged step size equal to $\lambda/10$ are correct due to the fact of taking accurately into account the shape of the object with a reasonable number of cells in the domain (46,292), which implies a low memory storage (6 Mo). But the apparition of little cells in the unstructured mesh induces a little time step (2.e−7 s) to guarantee the stability and a supplementary cost in the CPU time for the FVTD solution (35 min). In this condition, the discontinuous Galerkin method again remains more interesting in terms of CPU time.

The second example consists in evaluating the backscattered far field of a perfectly metallic sphere of radius 1 m. Fig. 7 shows a comparison of the solutions obtained by the discontinuous Galerkin approach, the FDTD method and a time domain EFIE method (MOT), which is considered to be the reference solution. We can see

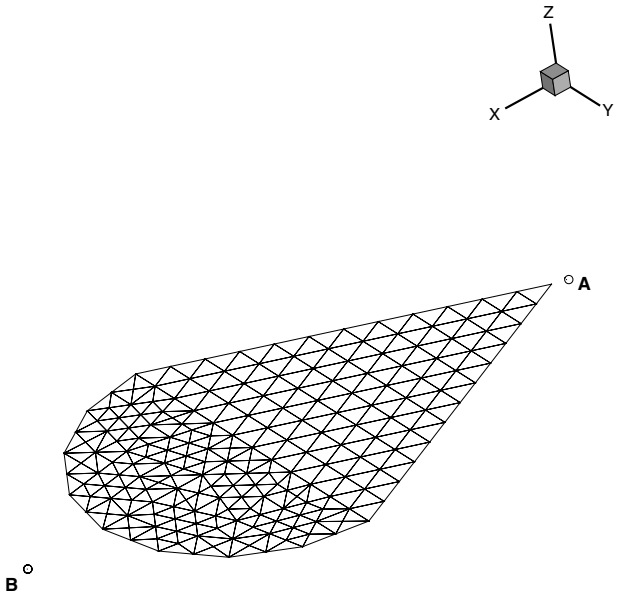


Fig. 5. Location of the test-points on the cone.

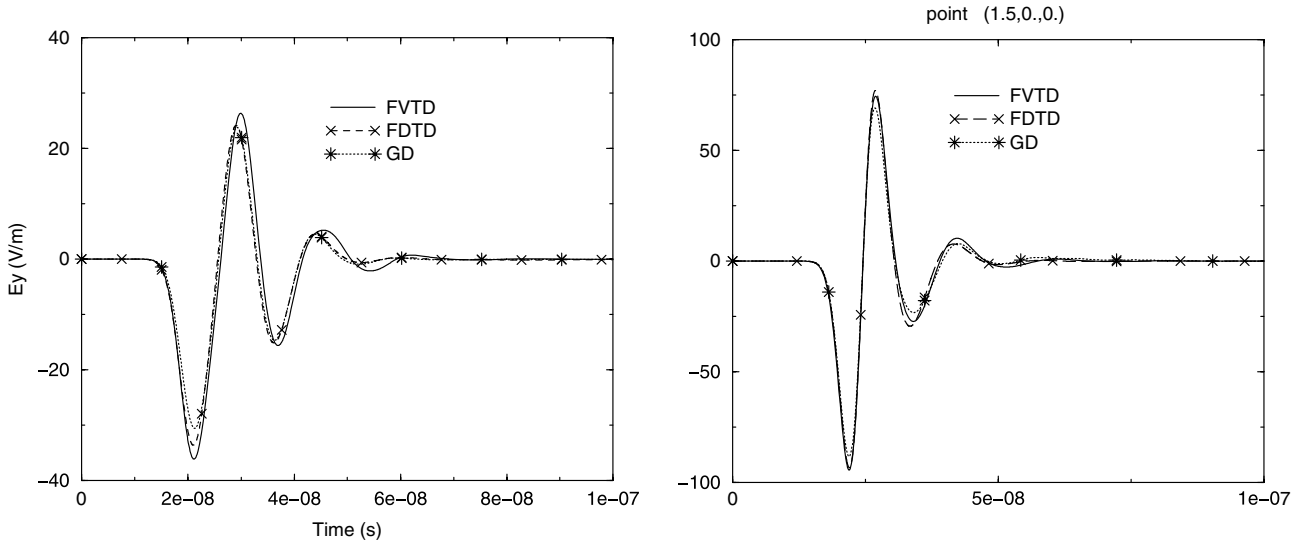


Fig. 6. Comparisons of the solutions obtained with different methods at the test-points A and B.

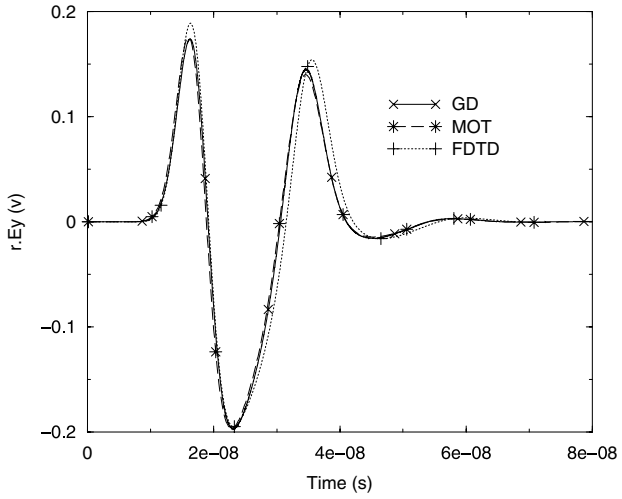


Fig. 7. Backscattered far fields obtained with different methods. The point is located at $r = 100$ m of the center of the sphere.

the perfect agreement between the DG and MOT solutions and the difficulty to coincide with the FDTD method. This is probably due to the staircase approach of the semi-disk. Again, in this example, the GDM allows us to obtain an accurate solution with the same advantage of memory and CPU time as for the previous example.

In these methods, the far fields are computed by an integral formula using electric and magnetic currents taken on a fictitious surface which encloses the scattered object [30].

4.2. Cavity problems and behavior of the solution at long time

For cavities or for long time experiments, the dispersion and the dissipation errors of the numerical scheme play a very important role. In these kinds of problems, the high order character of the method is crucial to obtain an accurate solution. To illustrate this purpose, we first study the propagation of a mode inside a perfectly metallic cubic cavity ($\mathbf{E} \times \mathbf{n} = 0$ on the wall of the cavity) with an edge of $1m$. The propagative mode studied is a mode $(m,n,0)$ given by

$$\begin{cases} E_x = 0; \quad E_y = 0; \quad H_z = 0, \\ E_z = \sin(m\pi x)\sin(n\pi y)\cos(\omega t), \\ H_x = \frac{1}{\omega\mu_0}\pi n \sin(m\pi x)\cos(n\pi y)\sin(\omega t), \\ H_y = \frac{1}{\omega\mu_0}\pi m \cos(m\pi x)\sin(n\pi y)\sin(\omega t). \end{cases} \tag{43}$$

Fig. 8 shows the comparison between the exact and the computed solutions for a given mode ($m = n = 3$) inside the cavity by using different orders of approximation and different mesh sizes. The advantage of using high order schemes in this kind of problem appears clearly. We also notice the reduction of the number of cells needed to get an accurate solution when the order increases. This also induces a gain of CPU time. Fig. 9 shows a comparison between computed solutions obtained on different meshes with the FDTD and the discontinuous Galerkin (Q_5-approximation) methods. We notice on this example the good behavior of the discontinuous Galerkin method after 180 wavelengths.

The next example is the evaluation of scattered fields for a long time of observation. Generally, for perfectly metallic objects with dimension of a few wavelengths (<10), the time of observation of the scattered signals is short. This is not the case for dielectric objects or for large objects (>100 wavelengths). So it is worth perform-

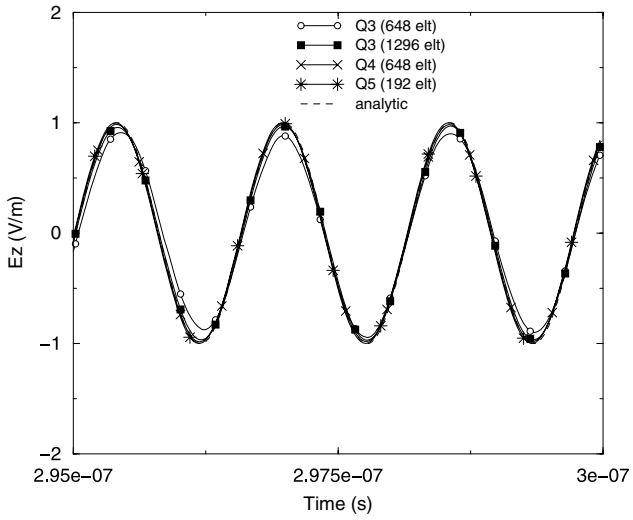


Fig. 8. Electric field taken inside a cavity after 180 wavelengths.

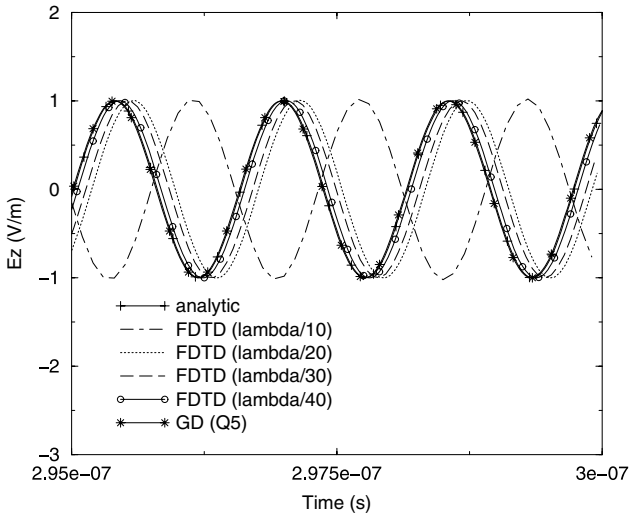


Fig. 9. Comparison FDTD and discontinuous Galerkin methods inside a cavity.

ing this kind of experiment with our method. The configuration proposed is defined by a metallic sphere of internal radius equal to 50 cm coated with a layer (thickness: 25 cm) of a dielectric material ($\varepsilon_r = 10$). This object is illuminated by a plane wave defined by $(k_x = 1,\ E_y = 377f(t,x),\ H_z = f(t,x))$ with $f(t,x) = \exp(-(5e8(t-(x+20/3e8))^2))$ and we evaluate the field at a test-point A located at 1 cm of the object (see Fig. 10).

The presence of the dielectric material makes the solution be unsteady for long observation time. In Fig. 11, one can see the behavior of the solutions of the problems obtained by FDTD and ours for different sizes of cells and different orders of approximation.

We can see in these figures the advantage of having a high order method to obtain a solution avoiding dispersion error. In the results shown in the previous figures, we use for FDTD ($\lambda/20$) a mesh of 6,751,269 cells (309 Mo) compared to 16,984 cells (42 Mo) for a Galerkin discontinuous Q_3-approximation (we use the same number of cells for the Q_5-approximation (192 Mo) with a quite similar solution as for Q_3-approximation). The CPU times are similar between FDTD ($\lambda/40$) and the Q_3-approximation, but at long time of observation, the solutions are very different. To improve the FDTD solution, we need to use a smaller spatial step and in this case, the values of CPU time and memory storage obtained clearly show the efficiency of our method in such experiments.

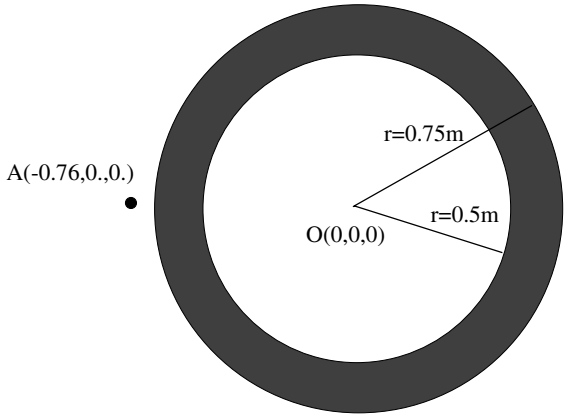


Fig. 10. Cut of the object at the plane $z = 0$.

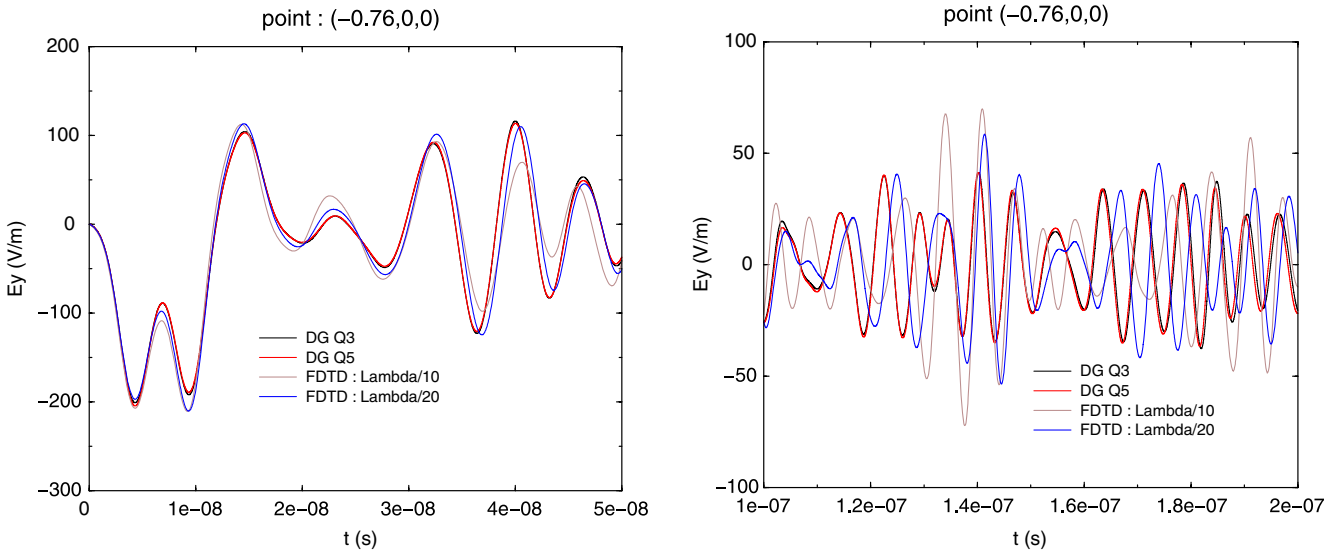


Fig. 11. Evolution of the fields at observation times corresponding to 20 wavelengths and to 80 wavelengths.

5. Local time-stepping strategy

All these examples clearly prove the capacity of our method to take into account curved geometries and to decrease the dispersion error in cavity problems or at long times. Of course, the efficiency of the method lies in the possibility to get high order spatial approximation with reasonable memory storage and CPU time. However, the difficulty of this approach consists in obtaining a mesh only based on hexahedral cells. An efficient method consists in constructing an initial mesh composed of tetrahedral cells and then, in splitting each cell into four hexahedral cells. The drawback of this method is the very distorted character of the final mesh and the fact that the size of some cells can be very little (as in tetrahedral meshes). Moreover, we have noticed some problems of stability at high order in the classical PML [29], on some examples by using very distorted meshes. To avoid this problem, another approach consists in using, for instance, an absorbing boundary condition strategy [31]. However, generally, this approach needs to have the boundary of the computational domain far from the object in order to get reasonable reflections. This condition can increase the computational domain and the CPU time.

For all these reasons, the use of a local time-stepping strategy seems to be interesting.

The local time-stepping strategy that we propose has been applied to some scattering problems and gives very good results for our actual experiments where the largest time of observation is on 100 wavelengths.

5.1. Proposed strategy

We can write our discontinuous Galerkin method as

$$\begin{cases} H^{n+\frac{1}{2}} - H^{n-\frac{1}{2}} = \mathrm{d}t\, f_E(E^n), \\ E^{n+1} - E^n = \mathrm{d}t\, f_H(H^{n+\frac{1}{2}}), \end{cases} \tag{44}$$

where E, H are the electric and magnetic fields, respectively, and f_E, f_H are linear functions. $\mathrm{d}t$ defines the time step verifying the stability condition:

$$\mathrm{d}t < cfl \frac{\mathrm{d}l}{v} \tag{45}$$

where v is the velocity of light in the medium, $\mathrm{d}l$ the size of the cell and cfl a strict positive number.

When the condition (45) is applied to each cell of the mesh, the value of $\mathrm{d}t$ varies from cell to cell. To ensure the stability, the minimum value $\mathrm{d}t_{\min} = \min \mathrm{d}t$ is taken as the time-step of the method. In our strategy, we propose to define a value $\mathrm{d}t_c = 2(n+1)\mathrm{d}t_{\min}$, where n is a given integer and take this value as the time step of the method. Then, to ensure the stability condition, we obtain two sets of cells where the fields in the first one labelled 1 are computed by using a time step equal to $\mathrm{d}t_{\min}$ (the values of $\mathrm{d}t$ of each cell in this set are smaller than $\mathrm{d}t_c$) and a second, labelled 2, where the fields are computed by using a time step equal to $\mathrm{d}t_c$.

In the time process, we obtain for one step in the set of cells labelled 2, $2(n+1)$ local steps in the set of cells labelled 1. To compute the fields located in the two sets, we need some values of fields located in the other set. The difficulty of the local time-stepping process is to have the fields at the appropriate time in each step. This is generally obtained by doing interpolations, but this is more difficult in the leap-frog scheme presented above.

The strategy proposed to obtained this coincidence of time is the following:

- mark with number 3 the cells labelled 2 which are neighbors of the cells labelled 1;
- mark with number 4 the cells labelled 2 which are neighbors of the cells labelled 3. At this step, we obtain a configuration of labelled cells as represented in Fig. 12.
- assume we know the magnetic and electric fields in the computational domain, respectively, for the time $(m-\frac{1}{2})\mathrm{d}t_c$ and $t_m = m\,\mathrm{d}t_c$. At each step $\mathrm{d}t_c$, we have the following sequence:
  - save at the time $t_m - \frac{\mathrm{d}t_c}{2}$ in H_s the H fields and at the time t_m in E_s the electric fields for the cells marked 3 and 4;
  - compute H for all the cells marked 2,3 and 4 at the time $t_m + \frac{\mathrm{d}t_c}{2}$;
  - put $t_0 = t_m$;

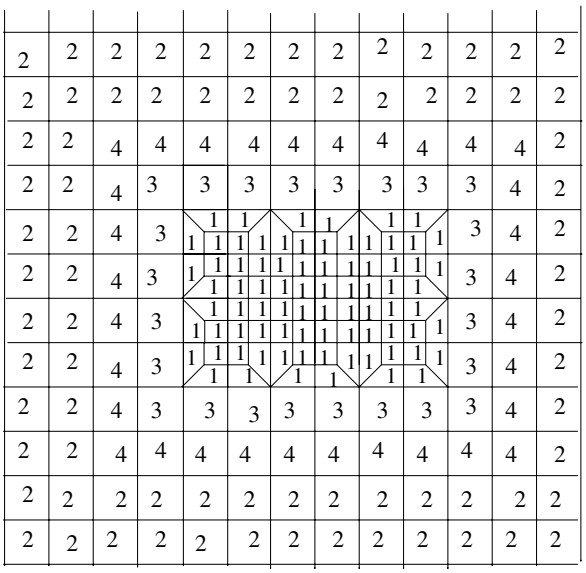


Fig. 12. Example of local time step cell marking.

- for each local step 1 to n
  - compute H for all cells marked 1 at the time $t_0 + \frac{\mathrm{d}t_{\min}}{2}$;
  - interpolate H for the cells marked 3 and 4 at the time $t_0 + \frac{\mathrm{d}t_{\min}}{2}$ by using H_s fields and the H fields previously evaluated at the time $t_m + \frac{\mathrm{d}t_c}{2}$;
  - compute E for the cells marked 1 and 3 at the time $t_0 + \mathrm{d}t_{\min}$;
  - increase $t_0 = t_0 + \mathrm{d}t_{\min}$;
- evaluate the H fields in cells marked 1 at the time $t_0 + \frac{\mathrm{d}t_{\min}}{2}$, which is equal to $t_m + \frac{\mathrm{d}t_c}{2}$ by the choice made on $\mathrm{d}t_c$;
- restore E_s values in E for the cells marked 3 and 4;
- save at the time t_m, the electric fields of the cells marked 3 and 4 in E_s and the magnetic fields H at the time of the cells marked 3 and 4 in H_s;
- evaluate the electric fields E for the cells marked 2,3 and 4 at the time $t_m + \mathrm{d}t_c$;
- put $t_0 = t_m + \frac{\mathrm{d}t_c}{2}$;
- for each local step 1 to n
  - evaluate the electric fields E for the cells marked 1 at the time $t_0 + \frac{\mathrm{d}t_{\min}}{2}$;
  - interpolate the electric fields E for the cells marked 3 and 4 at the time $t_0 + \frac{\mathrm{d}t_{\min}}{2}$ with the values E_s and E previously computed at the time $t_m + \mathrm{d}t_c$;
  - evaluate the magnetic fields H for the cells marked 1 and 3 at the time $t_0 + \mathrm{d}t_{\min}$;
  - increase $t_0 = t_0 + \mathrm{d}t_{\min}$;
- evaluate the electric fields E for the cells marked 1 at the time $t_0 + \frac{\mathrm{d}t_{\min}}{2}$;
- restore the H_s values in H for the cells marked 3 and 4;
- increase $t_m = t_m + \mathrm{d}t_c$;

The strategy presented here has two different time steps, but it is also possible to have more of them. In this case, we only need to mark more cells with different numbers. For our applications, we do not use more local time steps because the most important point is to eliminate the very small cells in order to have a time step for the process large enough.

5.2. Numerical examples

To show the interest of this strategy, we consider a plane wave illuminating a cone, as previously studied in the numerical results section. The scattered fields are observed at the test-point A (see Fig. 5).

To obtain the unstructured mesh of the computational domain, we enclose the cone inside a sphere where an absorbing condition of order 1 is applied (condition of Silver-Muller). The domain between the cone and the sphere is given by a set of tetrahedric cells, which are split into four hexaehdric cells. This kind of mesh implies that the size of a lot of cells cannot be easily controlled. Consequently, the obtained mesh contains cells with large variation of size. Therefore, it is interesting to apply in this case a local time-stepping strategy in our method.

Fig. 13 shows a part of the unstructured mesh where we can see the significant difference between the size of the cells. In particular, in this mesh, the smallest size of the cell is equal to 0.02 m, the largest size to 0.45 m and the averaged size to 0.2 m . In our Galerkin method, these values imply, for a Q_3-approximation, a time step dt equal to 5e−12 s to ensure the stability. If we consider a time step equal to 3 times this minimal value, the number of cells where the condition of stability is locally verified is equal to 535 compared to 13,160 for all the computational domains. Then, only 4% of the cells in the mesh require the application of a local time step equal to dt. The others are evaluated by using a time step equal to 3 dt. If we consider time steps equal to 5 and 7 times dt, the number of cells where dt must be applied increases, respectively, to 25% and 43%.

With a simulation time equal to 3.5e−7 s, we compare the method by using dt as the time step on the whole domain and a strategy of local time steps by using time step given by 3, 5 and 7 times the value of dt. Fig. 13 presents the results obtained in each simulation. We can see the perfect coincidence of the different curves with, for the simulation without local time steps, a CPU time of 337 min 24 s (on a Pentium 4 at 3 GHz) and for the others, a CPU time of 159 min 59 s, 180 min and 194 min 14 s, respectively. In these results, we notice the important gain due to the use of a strategy of local time-stepping. However, this gain is limited by the fact that additional interpolations on cells located around the little cells are introduced in the calculation. This additional cost implies a limit of the time step under which we waste CPU time because the number of small cells considered becomes too large in the mesh. In this example, we can see this effect by using a time step larger than $3 \times 5\text{e}{-12}$ s. Nevertheless, in this example, we notice that the CPU time is still smaller than the time required when we do not use a strategy of local time-stepping.

This strategy has also been applied in the case of a larger size example (see Fig. 14) for which a comparison with the classical FDTD method has been done. In this example, an aircraft is illuminated by a plane wave given by ($k_x = 1$, $E_y = 377 * f(t)$, $H_z = f(t)$) with $f(t) = 3.\text{e}8 * t - x$. We evaluate the fields at the test-point $A = (6,0,1.5)$ located as shown in Fig. 14. To obtain the same solution, we need to have 6,301,008 cells in the FDTD method and 180,076 in the Galerkin discontinuous method (Q_2-approximation). The amount of memory storage is more interesting in the Galerkin discontinuous method, however, because of the very small size of several cells in our unstructured mesh, we need a very small time step (2.e−12 s) in the Galerkin dis-

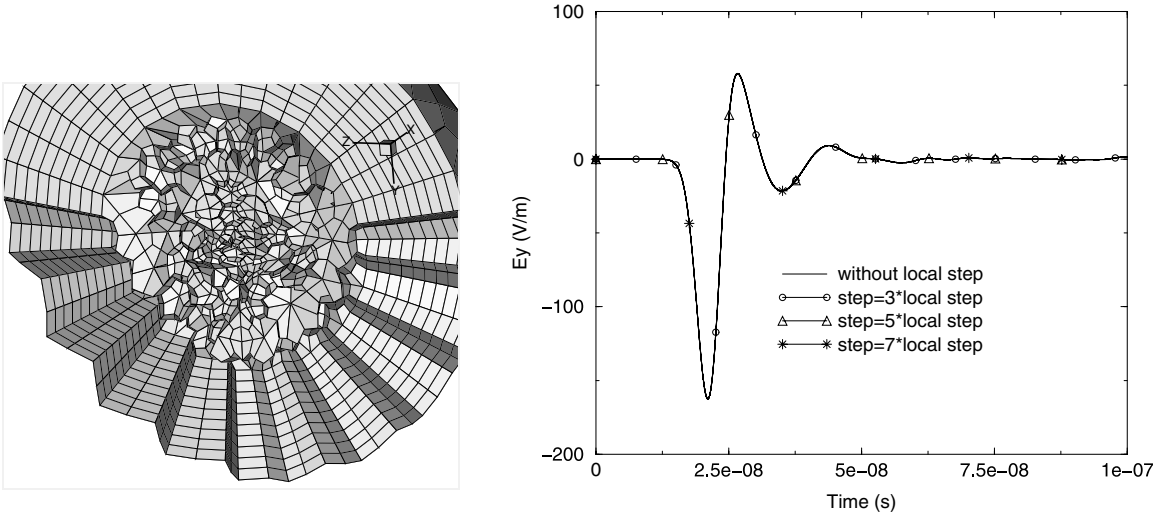


Fig. 13. Part of the mesh of the computational domain and solutions.

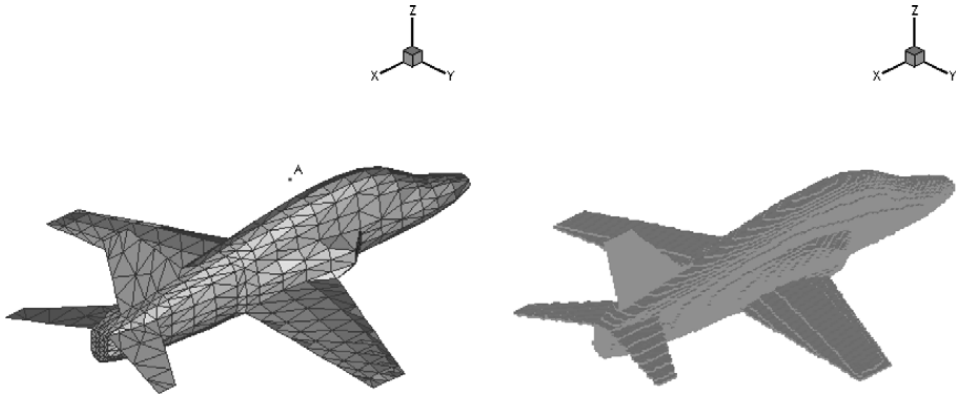


Fig. 14. Discontinuous Galerkin and FDTD surfaces meshing object.

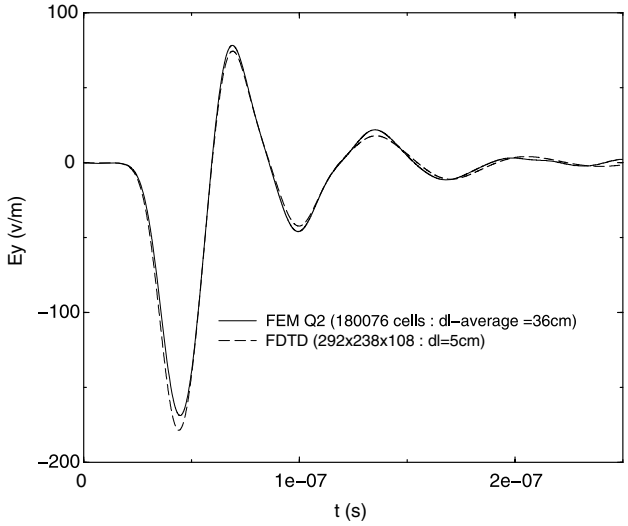


Fig. 15. Comparison between the solutions obtained.

continuous method compared to 8.e−11 s in the FDTD method. In this case, the FDTD is more attractive than the Galerkin discontinuous method in terms of CPU time. Nevertheless, by using the previous local time-stepping strategy, we use only $dt = 2.e-12$ s on 18,211 cells (10%) of the mesh and a time step equal to $11dt$ on all the others. In these conditions, the Galerkin discontinuous method is equivalent in terms of CPU time to the FDTD method with a smaller memory storage. In this case, globally, the Galerkin discontinuous method becomes once again more interesting than FDTD. Fig. 15 shows the comparison between the solutions obtained by the two methods.

6. Conclusion

In this paper, we have presented a non-dissipative spatial high order Galerkin discontinuous method to solve the Maxwell equations in the time domain. This method has the advantage of requiring small memory storage with a high order spatial scheme. In particular, the use of a centered formulation allows us to obtain an important gain in storage for jump matrices. A study of the dispersion and the stability of the method has been done and some numerical results have shown the advantage of this method compared to other classical methods such as FDTD, for the same level of accuracy of the solutions. In particular, in our experiments we do not see an error impact due to the spurious modes, on the solution.

However, for complex geometry problems, the drawback to this method is that it requires the use of a mesh composed of hexahedric cells. A solution to obtain this kind of mesh from tetrahedric cells can be easily obtained by splitting each cell into four hexahedric cells. But this implies a significant difference in terms of the size of the cells and the size of the time step which must be very small. A strategy for local time-stepping has been proposed in order to reduce considerably the cost related to the little cells in the mesh. Some examples of scattering for time of observation on at most 100 wavelengths justify the interest of this approach. However, for this method, conservative energy quantity is not ensured and for long time-period the method could be unstable. Then, for some specific examples such as cavities, this strategy could not be efficient and will need to be improved in the future.

References

[1] K.S. Yee, Numerical solution of initial boundary value problems involving Maxwell's equation in isotropic media, IEEE Trans. Antennas Prop. 14 (3) (1966) 302–307.
[2] A. Taflove, S.C. Hagness, Computational Electrodynamics: The Finite-Difference Time-Domain Method, second ed., Artech, Norwood, MA, 2000.
[3] A.C. Cangellaris, D.B. Wright, Analysis of the numerical error caused by the stair-stepped approximation of a conducting boundary in FDTD simulations of electromagnetic phenomena, IEEE Trans. Antennas Prop. AP-39 (10) (1991) 1518–1525.
[4] S.M. Rao, in: D. Irwin (Ed.), Time Domain Electromagnetics, Auburn University Academic Press, 1999.
[5] G. Rodrigue, D. White, A vector finite element time-domain method for solving Maxwell's equations on unstructured hexahedral grids, SIAM J. Sci. Comput. 23 (2001) 683–706.
[6] F. Edelvik, G. Ledfelt, Explicit hybrid time domain solver for the maxwell equations in 3D, J. Sci. Comput. 15 (1) (2000).
[7] K.S. Yee, J.S. Chen, The finite-difference time-domain (FDTD) and the finite-volume time-domain (FVTD) methods in solving Maxwell equation, IEEE Trans. Antennas Prop. 45 (3) (1997) 354–363.
[8] T. Rylander, Stable FEM-FDTD hybrid method for Maxwell's equations, Ph.D. dissertation, Chalmers University of Technology, Goteborg, Sweden, 2002.
[9] X. Ferrieres, J.P. Parmantier, S. Bertuol, A. Ruddle, Application of an hybrid finite difference/finite volume method to solve an automotive EMC problem, IEEE Trans. EMC 46 (4) (2005) 624–634.
[10] S. Pernet, X. Ferrieres, G. Cohen, An original finite element method to solve Maxwell's equations in time domain, in: Proceedings of EMC Zurich' 2003, 18–20 February 2003, Zurich, Switzerland.
[11] G. Cohen, Higher-Order Numerical Methods for Transient Wave Equations, Springer Verlag, Berlin, 2001.
[12] J.S. Hesthavens, T. Warburton, High-order nodal methods on unstructured grids. I. Time-domain solution of Maxwell's equations, J. Comput. Phys. 181 (2002) 1–34.
[13] J. Jin, The Finite Element Method in Electromagnetics, John Wiley & Sons, New York, 1993.
[14] J.-C. Nédélec, Mixed finite elements in $\mathbb{R}^3$, Numer. Math. 35 (3) (1980) 315–341.
[15] J.-C. Nédélec, A new familly of mixed finite elements in $\mathbb{R}^3$, Numer. Math. 50 (1) (1986) 57–81.
[16] G. Cohen, P. Monk, Mur-Nedelec finite element schemes for Maxwell's equations, Comp. Meth. Appl. Mech. Eng. 169 (3–4) (1999) 197–217.
[17] A. Elmkies, P. Joly, Éléments finis d'arête et condensation de masse pour les équations de Maxwell: le cas de la dimension 3, C.R.A.S., Math. 325 (série I) (1997) 1217–1222.
[18] W. Reed, T. Hill, Triangular mesh methods for the neutron transport equation, Tech. Report LA-UR-73-479, Los Alamos National Laboratory, Los Alamos, NM, USA, 1973.
[19] B. Cockburn, G.E. Karniadakis, C.-W. Shu, The Development of Discontiunuous Galerkin MethodsLecture Notes in Computational Science and Engineering, vol. 11, Springer, Berlin, 2000.
[20] P. Houston, I. Perugia, D. Shötzau, Mixed discontinuous Galerkin approximation of the Maxwell operator, SIAM J. Numer. Anal. 42 (1) (2004) 434–459.
[21] I. Perugia, D. Schötzau, P. Monk, Stabilized interior penalty methods for time-harmonic Maxwell equations, Comput. Meth. Appl. Mech. Eng. 191 (2002) 4675–4697.
[22] P. Monk, G.R. Richter, A discontinuous Galerkin method for linear symmetric hyperbolic systems in inhomogeneous media, J. Sci. Comput. 22 (2005) 443–477.
[23] B. Cockburn, F. Li, C.-W. Shu, Locally divergence-free discontinuous Galerkin methods for the Maxwell equations, J. Comput. Phys. 194 (2004) 588–610.
[24] P. Bonnet, X. Ferrieres, Numerical modeling of scattering problems using a time domain finite volume method, JEWA 11 (1997) 1165–1189.
[25] B. Cockburn, F. Li, C.W. Shu, Locally divergence-free discontinuous Galerkin methods for the Maxwell equations, J. Comput. Phys. 194 (2004) 588–610.
[26] S. Piperno, M. Remaki, L. Fezoui, A non-diffusive finite volume scheme for the 3D Maxwell equations on unstructured meshes, SIAM J. Numer. Anal. 39 (6) (2002) 2089–2108.
[27] E. Hairer, S.P. Nørsett, G. Wanner, Solving Ordinary Differential Equations, Springer Series in Computational Mathematics, 1991.

[29] J.P. Berenger, A perfectly matched layer for the absorption of electromagnetic waves, J. Comput. Phys. 114 (1994) 185–200.
[30] K.S. Yee, Time domain extrapolation to the far field based on FDTD calculations, IEEE Trans. Antennas Prop. 39 (3) (1991) 410–413.
[31] W.F. Hall, A.V. Kabakian, A sequence of absorbing boundary conditions for Maxwell's equations, J. Comput. Phys. 194 (2004) 140–155.

MATHEMATICS OF COMPUTATION
Volume 76, Number 260, October 2007, Pages 1801–1832
S 0025-5718(07)01991-6
Article electronically published on April 20, 2007

HP A-PRIORI ERROR ESTIMATES FOR A NON-DISSIPATIVE SPECTRAL DISCONTINUOUS GALERKIN METHOD TO SOLVE THE MAXWELL EQUATIONS IN THE TIME DOMAIN

S. PERNET AND X. FERRIERES

ABSTRACT. In this paper, we present the *hp*-convergence analysis of a non-dissipative high-order discontinuous Galerkin method on unstructured hexahedral meshes using a mass-lumping technique to solve the time-dependent Maxwell equations. In particular, we underline the spectral convergence of the method (in the sense that when the solutions and the data are very smooth, the discretization is of unlimited order). Moreover, we see that the choice of a non-standard approximate space (for a discontinuous formulation) with the absence of dissipation can imply a loss of spatial convergence. Finally we present a numerical result which seems to confirm this property.

1. INTRODUCTION

The most widely used time domain method for solving Maxwell equations is the Finite Difference Time Domain method (FD-TD) based on the well known Yee scheme [5], [6]. This method uses an orthogonal Cartesian grid and is based on a centered difference approximation in space and a leap-frog approximation in time. That provides a second order accurate scheme. However the FD-TD method suffers from a certain number of drawbacks. For example, to treat curved objects, the staircase approximation of the boundary generates parasitic diffraction phenomena which can seriously damage the accuracy of the solution [7].

Scientists and engineers have tried to develop several efficient methods which make it possible to take into account the complex shapes of the objects [25], [9]. Moreover, the growing need to solve accurately propagating electromagnetic waves over many wavelengths has forced them to develop high-order or spectral methods [27], [8].

Their first choice has naturally turned to the Finite Element Method (FEM) which is a powerful tool to develop new numerical techniques [26]. One of the difficulties in using an FEM in the Maxwell types of problems is the construction of a finite dimension subspace of the continuous space $H(\mathrm{curl}, \Omega)$. Indeed, the tangential components of a function belonging to $H(\mathrm{curl}, \Omega)$ are continuous across any surface, but the normal components of the same function may be discontinuous. It is well know that the use of classical Lagrange finite elements of the space $[H^1(\Omega)]^3$ leads to spurious solutions. The appropriate finite element space was introduced by Nedelec in the 1980s [21], [22]. Unfortunately, the classical version of the edge finite elements leads to a high computational cost since a matrix inversion is needed at each time

Received by the editor June 20, 2005 and, in revised form, June 4, 2006.
2000 *Mathematics Subject Classification.* Primary 35B45; Secondary 65M12.

step. This drawback becomes more and more important when the order of the approximation increases. The mass-lumping technique is used in order to use this type of method in transient problems. One of the most efficient methods for solving the Maxwell equations was developed by Cohen and Monk in [23]. In this method, the use of the Gauss-Lobatto quadrature formulae yields a block diagonal mass matrix which allows one to obtain an explicit scheme for all polynomial orders of approximation.

The second choice is the use of Discontinuous Galerkin Methods (DGM). These methods were born in the first half of the Seventies throughout the work of Reed and Hill [18] on the scalar neutron transport equation. The first mathematical analysis was carried out by Lesaint and Raviart in 1974 [19]. One of the basic ideas came from certain authors who weakly imposed the Dirichlet boundary condition in the FEMs instead of taking it into account directly in functional spaces. Then they decided to use this technique not only on the boundary of the computational domain but directly on the boundary of each element of the mesh in order to restore certain continuities of the solution of the studied problem (for example tangential, normal continuity, etc.). Following these first studies, many DGMs were developed and analyzed by many scientists in order to solve a large variety of problems (hyperbolic, parabolic, elliptic, etc.). An exhaustive review of these methods since their beginning is presented in [14]. However, one will note that few papers deal with the resolution of the Maxwell equations. In fact the use of this type of method to solve electromagnetism problems is relatively recent. For the frequency domain, one can quote the works of [17], [16] and for the time domain, one can quote the works of [24] (space-time discontinuous approximation), [12] (efficient local divergence-free basis functions), [15] (refinements on cartesian grid), [8] (very efficient spectral discontinuous spatial approximation with low storage Runge-Kutta scheme for time approximation : high order RKDG scheme). One can notice that before the use of these high-order methods, Finite Volume methods (that can be viewed as low order DG schemes) were used to solve the Maxwell equations. These methods suffer from the too important presence of dissipation [10] or dispersion [11] which makes their use inaccurate in problems of big size in terms of wavelength.

The DGM have the following advantages:

- arbitrary order which is chosen according to the precision on the desired exact solution.
- methods easily parallelisable: discontinuous elements, mass matrices which are diagonal per blocks (= number of degrees of freedom in the cell).
- to treat complicated geometries and simple ways to treat the boundary conditions.
- adaptive strategies: space refinements natural (without taking account of the continuities as in finite elements), order of approximation different from one cell to the others.

Moreover, there are two approaches in implementing the DGMs, namely, the h-version and the p-version. The h-version allows the mesh size to be decreased to achieve convergence at a rate of the employed polynomial basis. The alternative p-version allows the order of polynomials to be increased with the sizes of the elements kept at an initial triangulation. A hybrid hp-version can also be considered. This paper is devoted to the study of the convergence study of this type of method.

The outline of the paper is as follows. In section 2, we describe the discontinuous Galerkin formulation that we have chosen to solve the Maxwell equations. In section 3, we justify the choice of an H^1-type projector to carry out our analysis and we derive some *hp*-projection errors for this one. In section 4, first we determine the a-priori error estimates of the DGM for the spatial semi-discrete approximation without numerical integration; second, we study the effect of the use of the Gauss quadrature rule to compute the integrals on the previous error estimates. Finally, in section 5, a numerical example which confirms the theoretical analysis is given.

2. Presentation of the discontinuous Galerkin method

2.1. **Time-dependent Maxwell's equations.** Let Ω be a bounded open subset of $\mathbb{R}^3$ whose boundary is $\partial\Omega$ and $\mathbf{n}$ denotes the unit outward normal to Ω. Let $\underline{\underline{\varepsilon}}(x)$, $\underline{\underline{\mu}}(x)$ and $\underline{\underline{\sigma}}(x)$ denote, respectively, the permittivity, the permeability and the conductivity tensors of the medium.

We consider the problem described by the Maxwell equations: Find $(\mathbf{E}, \mathbf{H}) : \Omega \times]0, T[\to \mathbb{R}^3 \times \mathbb{R}^3$ such that:

$$
\left\{
\begin{array}{c}
\underline{\underline{\varepsilon}} \dfrac{\partial \mathbf{E}}{\partial t} - \nabla \times \mathbf{H} + \underline{\underline{\sigma}} \mathbf{E} + \mathbf{J}_s = 0 \ \text{ in } \Omega, \\
\underline{\underline{\mu}} \dfrac{\partial \mathbf{H}}{\partial t} + \nabla \times \mathbf{E} = 0 \ \text{ in } \Omega, \\
\mathbf{E} \times \mathbf{n}(x) = 0 \ \text{ on } \partial\Omega, \\
\mathbf{E}(x,0) = \mathbf{E}_0(x) \text{ and } \mathbf{H}(x,0) = \mathbf{H}_0(x) \ \text{ in } \Omega,
\end{array}
\right. \tag{2.1}
$$

where $\mathbf{E}$, $\mathbf{H}$ denote the electric and magnetic field intensities, $\mathbf{J}_s$ specifies the applied current and $\mathbf{E}_0$, $\mathbf{H}_0$ are the initial conditions.

We assume that $\underline{\underline{\varepsilon}}$, $\underline{\underline{\mu}}$, $\underline{\underline{\sigma}} \in [L^\infty(\Omega)]^{3\times 3}$ are symmetric definite positive matrices and $\exists C_1, C_2 > 0$ such that:

$$
\forall \xi \in \mathbb{R}^3 \, : \, C_1|\xi|^2 \le \underline{\underline{\varepsilon}}\xi.\xi \le C_2|\xi|^2, \, C_1|\xi|^2 \le \underline{\underline{\mu}}\xi.\xi \le C_2|\xi|^2, \, C_1|\xi|^2 \le \underline{\underline{\sigma}}\xi.\xi \le C_2|\xi|^2.
$$

Moreover if we assume $\mathbf{J}_s \in C^0(0, T; [L^2(\Omega)]^3)$, we have the existence and the uniqueness of the solution $(\mathbf{E}, \mathbf{H}) \in [C^1(0, T; [L^2(\Omega)]^3) \cap C^0(0, T; H_0(\mathrm{curl}, \Omega))]^2$ [3].

2.2. **Discontinuous formulation.** We assume that the computational domain, Ω, is split into a set of cells, $\mathscr{T}_h$ such that $\Omega = \bigcup_{i=1}^{N_e} K_i$, where $K_i \in \mathscr{T}_h$, $\dot{K}_i \cap \dot{K}_j = \emptyset$, $\forall i \neq j$ and K_i is a hexahedron. We denote the set of faces of $\mathscr{T}_h$ by $\mathscr{F}_h = \mathscr{F}_h^i \cup \mathscr{F}_h^b$ where $\mathscr{F}_h^i$ ($\Gamma \in \mathscr{F}_h^i$, $\Gamma = K' \cap K$) and $\mathscr{F}_h^b$ ($\Gamma \in \mathscr{F}_h^b$, $\Gamma = K \cap \partial\Omega$) are the sets of the interior and boundary faces. To each element $K \in \mathscr{T}_h$, we associate the outward unit normal $\mathbf{n}_K$.

For a real $s \ge 0$, we define the classical broken space:

$$
H^s(\mathscr{T}_h) = \{v \in L^2(\Omega) \, : \, \forall K \in \mathscr{T}_h, \, v_{|K} \in H^s(K)\}. \tag{2.2}
$$

$H^s(\mathscr{T}_h)$ is equipped with the natural norm: Let $v \in H^s(\mathscr{T}_h)$,

$$
\|v\|_{s,h} = \Big(\sum_{K \in \mathscr{T}_h} \|v\|_{s,K}^2 \Big)^{\frac{1}{2}}, \tag{2.3}
$$

where $\|.\|_{s,K}$ is the usual Sobolev norm of H^s on K.

For $s > \frac{1}{2}$, we define the jump of a function $v \in H^s(\mathscr{T}_h)$:

$$\begin{aligned} \forall \Gamma \in \mathscr{F}_h^i \text{ such that } \Gamma = K' \cap K,\ [\![v]\!]_\Gamma^K &= (v_{|K'})_{|\Gamma} - (v_{|K})_{|\Gamma} \\ \forall \Gamma \in \mathscr{F}_h^b \text{ such that } \Gamma \subset \partial K,\ [\![v]\!]_\Gamma^K &= -(v_{|K})_{|\Gamma}. \end{aligned} \tag{2.4}$$

We denote $\mathbf{H}^s(\mathscr{T}_h)$ as the vectorial broken space $[H^s(\mathscr{T}_h)]^3$ and its norm is defined by

$$\|v\|_{s,h} = \Big(\sum_{i=1}^{3} \|v_i\|_{s,h}^2\Big)^{\frac{1}{2}} \tag{2.5}$$

where $v = (v_1, v_2, v_3) \in \mathbf{H}^s(\mathscr{T}_h)$.

We rewrite the problem (2.1) under the following discontinuous form:

Find $(\mathbf{E}(\cdot,t), \mathbf{H}(\cdot,t)) \in \mathbf{H}^1(\mathscr{T}_h) \times \mathbf{H}^1(\mathscr{T}_h)$ such that, $\forall K \in \mathscr{T}_h$ and $\forall \boldsymbol{\phi}_1,\ \boldsymbol{\phi}_2 \in \mathbf{H}^1(\mathscr{T}_h)$,

$$\left\{ \begin{aligned} &\frac{d}{dt}\int_K \underline{\underline{\varepsilon}}\mathbf{E}_K \cdot \boldsymbol{\phi}_{1K} dx - \int_K \nabla \times \mathbf{H}_K \cdot \boldsymbol{\phi}_{1K} dx \\ &\qquad + \int_K \underline{\underline{\sigma}}\mathbf{E}_K \cdot \boldsymbol{\phi}_{1K} dx + \int_K \mathbf{J}_s \cdot \boldsymbol{\phi}_{1K} dx \\ &= \int_{\partial K} \alpha [\![\mathbf{n}_K \times (\mathbf{E} \times \mathbf{n}_K)]\!]_{\partial K}^K \cdot \boldsymbol{\phi}_{1K} d\sigma + \int_{\partial K} \beta [\![\mathbf{H} \times \mathbf{n}_K]\!]_{\partial K}^K \cdot \boldsymbol{\phi}_{1K} d\sigma \\ &\frac{d}{dt}\int_K \underline{\underline{\mu}}\mathbf{H}_K \cdot \boldsymbol{\phi}_{2K} dx + \int_K \nabla \times \mathbf{E}_K \cdot \boldsymbol{\phi}_{2K} dx \\ &= \int_{\partial K} \gamma [\![\mathbf{E} \times \mathbf{n}_K]\!]_{\partial K}^K \cdot \boldsymbol{\phi}_{2K} d\sigma + \int_{\partial K} \delta [\![\mathbf{n}_K \times (\mathbf{H} \times \mathbf{n}_K)]\!]_{\partial K}^K \cdot \boldsymbol{\phi}_{2K} d\sigma \end{aligned} \right. \tag{2.6}$$

where $\mathbf{E}_K = \mathbf{E}_{|K}$, $\mathbf{H}_K = \mathbf{H}_{|K}$, $\boldsymbol{\phi}_{jK} = \boldsymbol{\phi}_{j|K}$, $d\sigma$ is the surface measurement associated with ∂K and α, β, γ, δ are four reals that could be different from one face to another.

We get a non-dissipative formulation. For that we choose the parameters :

- $\forall \Gamma \in \mathscr{F}_h^i$, α, $\delta = 0$, $\beta = -\frac{1}{2}$ and $\gamma = \frac{1}{2}$,
- $\forall \Gamma \in \mathscr{F}_h^b$, α, $\delta = 0$, $\beta = 0$ and $\gamma = 1$.

Indeed, by using this choice, the classical electromagnetic energy $\mathscr{E}(t) = \int_\Omega \underline{\underline{\epsilon}}\mathbf{E}(t) \cdot \mathbf{E}(t) d\mathbf{x} + \int_\Omega \underline{\underline{\mu}}\mathbf{H}(t) \cdot \mathbf{H}(t) d\mathbf{x}$ is time-conserved, i.e. $\mathscr{E}(t) = \mathscr{E}(0)$, $\forall t$.

2.3. **Spatial approximation.** Given a non-negative integer r and $E \subset \mathbb{R}^d$, $Q_r(E)$ is the space of polynomials of degree at most equal to r in each variable on E. Let us introduce the standard unit cube $\hat{K} = [0,1]^3$. $\forall K \in \mathscr{T}_h$, $F_K : \hat{K} \to K$ denotes the trilinear mapping which associates the vertices of each element. $(\hat{x}_1, \hat{x}_2, \hat{x}_3)$ are the coordinates on the reference element and (x_1, x_2, x_3) the coordinates on the elements of the mesh. DF_K and J_K are the Jacobian matrix and its determinant associated with the map F_K.

We use the discontinuous finite element space:

$$U_h = \{\mathbf{v}_h \in \mathbf{L}^2(\Omega) : \forall K \in \mathscr{T}_h,\ DF_K^* \mathbf{v}_{h|K} o F_K \in [Q_r(\hat{K})]^3\} \tag{2.7}$$

where $r \in \mathbb{N}$.

In (2.7), the Jacobian matrix is the essential ingredient to build a conform Hing-curl approximation [21]. In our case, it allows us to reduce the storage of the stiffness and the jump matrices [34]. We do not detail this point here because the aim of this paper is only the study of the convergence of this approximation. For more details on this point, we can see [34] or [4].

The first step to define the basis functions of U_h is to construct a vector valued polynomial basis of $[Q_r]^3$, $\forall K \in \mathscr{T}_h$. We denote by $(\hat{\xi}_l, \hat{\omega}_l)_{l=1}^{r+1}$ the Gauss quadrature rule on $[0,1]$ where $(\hat{\xi}_l)_{l=1}^{r+1}$ are the quadrature points and $(\hat{\omega}_l)_{l=1}^{r+1}$ are the associated quadrature weights. The quadrature points and weights of the corresponding rules on $\hat{K}$ are the cartesian product of 1D points $\{\hat{\xi}_{l,m,n} = (\hat{\xi}_l, \hat{\xi}_m, \hat{\xi}_n) : \forall 1 \leq l, m, n \leq r+1\}$ and the set $\{\hat{\omega}_{l,m,n} = \hat{\omega}_l\hat{\omega}_m\hat{\omega}_n : \forall 1 \leq l, m, n \leq r+1\}$ respectively. Let $(\hat{\varphi}_l)_{l=1}^{r+1}$ be the set of Lagrange polynomials associated with the set of points $(\hat{\xi}_l)_{l=1}^{r+1}$.

We have $\hat{\varphi}_l(\hat{\xi}_j) = \delta_{l,j}$ and $(\hat{\varphi}_l)_{l=1}^{r+1}$ is a set of basis functions of $P_r([0,1]) = Q_r([0,1])$.

Now, we define the basis functions of $[Q_r(\hat{K})]^3$ in the following way:

$$\boldsymbol{\varphi}^i_{l,m,n}(\hat{x}_1, \hat{x}_2, \hat{x}_3) = \hat{\varphi}_l(\hat{x}_1)\hat{\varphi}_m(\hat{x}_2)\hat{\varphi}_n(\hat{x}_3)\mathbf{e}_i \tag{2.8}$$

where $i = 1, 2$ or 3 and $(\mathbf{e}_i)_{i=1,2,3}$ is the canonical basis of $\mathbb{R}^3$.

We have $\boldsymbol{\varphi}^i_{l,m,n}(\hat{\xi}_{l',m',n'}) = \delta_{l,l'}\delta_{m,m'}\delta_{l,l'}\vec{e}_s$. The choice of the basis functions at the quadrature points allows us to mass-lump the mass matrix [4].

Let $\mathscr{B}_h$ be a set of basis functions of U_h. We define an element $\boldsymbol{\psi}_h \in \mathscr{B}_h$ in the following way: $\boldsymbol{\psi}_h \in \mathscr{B}_h \Leftrightarrow \text{supp}(\boldsymbol{\psi}_h) = K \in \mathscr{T}_h$, $\exists \boldsymbol{\varphi}^i_{l,m,n}$ such that

$$\boldsymbol{\psi}_h o F_K = DF_K^{*-1}\boldsymbol{\varphi}^i_{l,m,n}.$$

Let $\mathbf{v}_h \in U_h$. So, we have the decomposition:

$$\forall K \in \mathscr{T}_h,\ \mathbf{v}_{h|K} o F_K = \sum_{i=1}^{3}\sum_{l,m,n=1}^{r+1} v^i_{K,l,m,n} DF_K^{*-1}\boldsymbol{\varphi}^i_{l,m,n} \tag{2.9}$$

where $v^i_{K,l,m,n}$ are the degrees of freedom of $\mathbf{v}_h$.

Finally, we obtain the following semi-discrete discontinuous Galerkin problem: find $(\mathbf{E}_h(\cdot,t), \mathbf{H}_h(\cdot,t)) \in U_h \times U_h$ such that, $\forall K \in \mathscr{T}_h$ and $\forall \boldsymbol{\phi}_{h1}, \boldsymbol{\phi}_{h2} \in \mathscr{B}_h$,

$$\begin{cases} \dfrac{d}{dt}\displaystyle\int_K^G \underline{\underline{\varepsilon}}\mathbf{E}_{hK}\cdot\boldsymbol{\phi}_{h1K}dx - \int_K^G \nabla\times\mathbf{H}_{hK}\cdot\boldsymbol{\phi}_{h1K}dx \\ \qquad + \displaystyle\int_K^G \underline{\underline{\sigma}}\mathbf{E}_{hK}\cdot\boldsymbol{\phi}_{h1K}dx + \int_K^G \mathbf{J}_s\cdot\boldsymbol{\phi}_{h1K}dx \\ \qquad\qquad = \displaystyle\int_{\partial K}^G \beta[\![\mathbf{H}_h\times\mathbf{n}_K]\!]^K_{\partial K}\cdot\boldsymbol{\phi}_{h1K}d\sigma, \\ \\ \dfrac{d}{dt}\displaystyle\int_K^G \underline{\underline{\mu}}\mathbf{H}_{hK}\cdot\boldsymbol{\phi}_{h2K}dx + \int_K^G \nabla\times\mathbf{E}_{hK}\cdot\boldsymbol{\phi}_{h2K}dx \\ \qquad\qquad = \displaystyle\int_{\partial K}^G \gamma[\![\mathbf{E}_h\times\mathbf{n}_K]\!]^K_{\partial K}\cdot\boldsymbol{\phi}_{h2K}d\sigma \end{cases} \tag{2.10}$$

where $\int_K^G$ and $\int_{\partial K}^G$ denote the integrals computed with the quadrature rule G after a change of variables on the unit cube $\hat{K}$.

Remark 2.1. Recall that the orders of the Gauss quadrature rule is $2r+1$, i.e. exact for $[Q_r(\hat{K})]^3$.

3. Study of a projector on the approximate space

In this part, we choose a projector on U_h and we carry out its *hp*-convergence analysis. In particular, we prove some error estimates for this projector on a hexahedral mesh.

3.1. **Definitions and properties of meshes.** We assume that all hexahedrons K are convex in order to ensure the existence of the diffeomorphism $F_K \in [Q_1(\hat{K})]^3$. Now let us give some definitions and properties on the quadrilateral finite elements (for more details see [1], [2]) and on the transformation F_K: To characterize an element $K \in \mathscr{T}_h$, we define:

$$\begin{aligned} h_K &= \text{diameter of K}, \\ \sigma_K &= \frac{h_K}{\rho_K} = \text{regularity parameter} \end{aligned} \tag{3.1}$$

where $\rho_K = \|J_{F_K^{-1}}\|_{\infty,K}^{\frac{1}{3}}$ with $J_{F_K^{-1}}$ as the determinant of the Jacobian matrix of F_K^{-1}.

Remark 3.1. In two dimensions, we can give a geometric characterization of ρ_K (see [33]). Indeed, in this case, ρ_K is the minimum of the diameters of the inscribed circles in the four triangles being able to be built with the nodes of the quadrangle K.

We note that

$$\begin{aligned} |F_K|_{m,\infty,\hat{K}} &= \sup_{\hat{\mathbf{x}}\in\hat{K}} \|D^m F_K(\hat{\mathbf{x}})\|_{\mathscr{L}_m(\mathbb{R}^3,\mathbb{R}^3)}, \\ |F_K^{-1}|_{m,\infty,K} &= \sup_{\mathbf{x}\in K} \|D^m F_K^{-1}(\mathbf{x})\|_{\mathscr{L}_m(\mathbb{R}^3,\mathbb{R}^3)} \end{aligned} \tag{3.2}$$

where $\mathscr{L}_m(\mathbb{R}^3,\mathbb{R}^3)$ is the set of the m-linear applications of $\mathbb{R}^3$ in $\mathbb{R}^3$, $D^m F_K(\hat{\mathbf{x}})$ and $D^m F_K^{-1}(\mathbf{x})$ are respectively the mth derivatives of F_K and F_K^{-1} at the points $\hat{\mathbf{x}}$ and $\mathbf{x}$. We will use the following estimates given in [2]:

$$\begin{aligned} &|F_K|_{1,\infty,\hat{K}} \le Ch_K,\ \|J_K\|_{\infty,\hat{K}} \le Ch_K^3, \\ &|F_K^{-1}|_{1,\infty,K} \le C\frac{h_K^2}{\rho_K^3},\ \|J_{F_K^{-1}}\|_{\infty,K} = \rho_K^{-3}, \\ &|F_K|_{2,\infty,\hat{K}} \le Ch_K,\ |F_K|_{2,\infty,\hat{K}} \le Ch_K^2 \text{ if } K \text{ is almost a parallelepiped} \end{aligned} \tag{3.3}$$

where $C > 0$ is independent of K and r.

Remark 3.2. By the expression "almost a parallelepiped", one wants to say a small deformation of a parallelepipedic cell. In this case, the second derivatives of F_K are zero.

Remark 3.3. We have by definition

$$\begin{aligned} D(F_K^{-1})(\mathbf{x}) &= \left(DF_K(F^{-1}(\mathbf{x}))\right)^{-1}, \\ J_{F_K^{-1}} o F_K &= \frac{1}{J_K} \end{aligned} \tag{3.4}$$

where $D(F_K^{-1})$ is the Jacobian matrix of F_K^{-1}.

Using the properties (3.3), it is easy to deduce the following proposition:

Proposition 1. *We have the following estimates:* $\forall \hat{x} \in \hat{K}$,

$$\begin{aligned} \lambda((DF_K DF_K^*)(\hat{x})) &\leq C h_K^2, \\ \lambda((DF_K^{-1} DF_K^{*-1})(\hat{x})) &\leq C \frac{h_K^4}{\rho_K^6} \end{aligned} \tag{3.5}$$

where $\lambda(A)$ *belongs to the spectrum of* A *and* $C > 0$ *is independent of* K *and* r.

Proof. Let $\hat{x} \in \hat{K}$. As $(DF_K DF_K^*)(\hat{x})$ and $(DF_K^{-1} DF_K^{*-1})(\hat{x})$ are symmetrical matrices, we can write:

$$\begin{aligned} \rho((DF_K DF_K^*)(\hat{x})) &= \sup_{\mathbf{v} \in \mathbb{R}^{*3}} \frac{\|(DF_K DF_K^*)(\hat{x})\mathbf{v}\|}{\|\mathbf{v}\|} = \|(DF_K DF_K^*)(\hat{x})\| \\ &\leq \|(DF_K)(\hat{x})\| \|(DF_K^*)(\hat{x})\| \leq |F_K|^2_{1,\infty,\hat{K}}; \end{aligned} \tag{3.6}$$

$\rho(A)$ is the spectral radius of A. Using (3.3), we immediately obtain the first inequality of (3.5). A similar reasoning allows us to prove the second estimate of (3.5). □

Finally, we define the regularity of a mesh:

Definition 3.4. A family $\mathscr{T}_h$ of triangulation of Ω is known as regular when h tends toward 0, if there exists a number $\sigma > 0$, independent of h, such that:

$$\sigma_K \leq \sigma, \ \forall K \in \mathscr{T}_h. \tag{3.7}$$

3.2. **Choice of a projector.** When deriving error estimates, an important point is the choice of a "good" projector on the approximate space used for discretization. Indeed, the use of an inappropriate projector can lead to sub-optimal estimates which give any interesting information about the numerical scheme. This part aims at justifying our choice.

For our DG scheme, the first idea is to use an L^2 projector. In particular, one can use the projector defined in the following way:

First, we can split the approximate space U_h in the following way:

$$U_h = \bigoplus_{K \in \mathscr{T}_h} U_K \tag{3.8}$$

where $U_K = \{\mathbf{v} \in \mathbf{L}^2(K) \, : \, DF_K^* v o F_K \in [Q_r(\hat{K})]^3\}$.

Then, in the first step, we define the L^2 projector $\hat{\pi}_r^0$ on $[Q_r(\hat{K})]^3$:

Definition 3.5 (Projector L^2)**.** Let $\hat{\mathbf{v}} \in \mathbf{L}^2(\hat{K})$ and $r \geq 0$. We define the projector L^2, $\hat{\pi}_r^0 \hat{\mathbf{v}}$, of $\hat{\mathbf{v}}$ on $[Q_r(\hat{K})]^3$ by : $\forall \hat{\boldsymbol{\varphi}} \in [Q_r(\hat{K})]^3$, we have

$$\int_{\hat{K}} \hat{\pi}_r^0 \hat{\mathbf{v}} \cdot \hat{\boldsymbol{\varphi}} d\hat{\mathbf{x}} = \int_{\hat{K}} \hat{\mathbf{v}} \cdot \hat{\boldsymbol{\varphi}} d\hat{\mathbf{x}}. \tag{3.9}$$

In the second step, we come back to U_K by defining the projector π_K^0.

Definition 3.6 (Projector on U_K)**.** Let $\mathbf{v} \in \mathbf{L}^2(K)$. We define the projection $\pi_K^0\mathbf{v}$ of $\mathbf{v}$ on U_K by

$$\big(\pi_K^0\mathbf{v}\big)oF_K = DF_K^{*-1}\hat{\pi}_r^0\hat{\mathbf{v}} \tag{3.10}$$

where $\hat{\mathbf{v}} = DF_K^*\mathbf{v} \circ F_K$.

Finally we define the projection operator on U_h.

Definition 3.7 (Projector on U_h)**.** Let $\mathbf{v} \in \mathbf{L}^2(\Omega)$. We define the projection $\pi_h^0\mathbf{v}$ of $\mathbf{v}$ on U_h by: For $K \in \mathscr{T}_h$,

$$\big(\pi_h^0\mathbf{v}\big)_{|K} = \pi_K^0\mathbf{v}_{|K}. \tag{3.11}$$

When examining the DG scheme in more detail, one sees that it is necessary to know error estimates of the first order derivatives of the projector used (because of the presence of the rational terms). So, an H^1 type projector on U_h can be a possibility for this study. In particular, we have considered the projector defined as:

First, we define the H^1 projector $\hat{\pi}_r^1$ on $[Q_r(\hat{K})]^3$.

Definition 3.8 (Projector H^1)**.** Let $\hat{\mathbf{v}} \in \mathbf{H}^1(\hat{K})$ and $r \geq 0$. We define the H^1 projection, $\hat{\pi}_r^1\hat{\mathbf{v}}$, of $\hat{\mathbf{v}}$ on $[Q_r(\hat{K})]^3$ by $\forall\hat{\boldsymbol{\varphi}} \in [Q_r(\hat{K})]^3$, we have

$$\int_{\hat{K}} \big(\hat{\pi}_r^1\hat{\mathbf{v}} - \hat{\mathbf{v}}\big) \cdot \hat{\boldsymbol{\varphi}} d\hat{\mathbf{x}} + \sum_{k=1}^{3} \int_{\hat{K}} \frac{\partial}{\partial \hat{x}_k}\big(\hat{\pi}_r^1\hat{\mathbf{v}} - \hat{\mathbf{v}}\big) \cdot \frac{\partial}{\partial \hat{x}_k}\hat{\boldsymbol{\varphi}} d\hat{\mathbf{x}} = 0. \tag{3.12}$$

Remark 3.9. In (3.12), $\dfrac{\partial \mathbf{w}}{\partial \hat{x}_k}$ means $\Big(\dfrac{\partial w_1}{\partial \hat{x}_k}, \dfrac{\partial w_2}{\partial \hat{x}_k}, \dfrac{\partial w_3}{\partial \hat{x}_k}\Big)^*$.

Then, we come back to U_K. Let $K \in \mathscr{T}_h$ and $\mathbf{v} \in \mathbf{H}^s(K)$ with $s \geq 1$. We define the projector π_K^1 on U_K by

$$\big(\pi_K^1\mathbf{v}\big)oF_K = DF_K^{*-1}\big(\hat{\pi}_r^1\hat{\mathbf{v}}\big) \tag{3.13}$$

where $\hat{\mathbf{v}} = DF_K^*(\mathbf{v}oF_K)$.

Finally we define the projection operator on U_h.

Definition 3.10 (Projector on U_h)**.** Let $\mathbf{v} \in \mathbf{L}^2(\Omega)$. We define the projection $\pi_h^1\mathbf{v}$ of $\mathbf{v}$ on U_h by: For $K \in \mathscr{T}_h$,

$$\big(\pi_h^1\mathbf{v}\big)_{|K} = \pi_K^1\mathbf{v}_{|K}. \tag{3.14}$$

We must be able to discriminate against these two projectors. The following subsection ("*hp*-projection errors") shows that the study identically applies to the two projectors and consequently gives the same interpolation error estimates. Moreover, section 4 shows that the two projectors lead to the same h convergence rate. However, the study of the spectral or of the *hp* convergence shows that these projectors do not give the same result:

Using theorem 57 of [31] as well as a tensorisation argument (i.e. $\hat{\pi}_r^0 = \hat{\pi}_{r,\hat{x}_3}^0 o\hat{\pi}_{r,\hat{x}_2}^0 o\hat{\pi}_{r,\hat{x}_1}^0$), we obtain the projection errors for $\hat{\pi}_r^0$:

Theorem 3.11. *$\forall\hat{\mathbf{u}} \in \mathbf{H}^p(\hat{K})$, it exists a constant C such that*

$$\|\hat{\mathbf{u}} - \hat{\pi}_r^0\hat{\mathbf{u}}\|_{q,\hat{K}} \leq C r^{\sigma(p,q)} \|\hat{\mathbf{u}}\|_{p,\hat{K}} \tag{3.15}$$

where

$$\sigma(p,q) = \begin{cases} \dfrac{3}{2}q - p,\ 0 \le q \le 1, \\ 2q - p - \dfrac{1}{2}q,\ q \ge 1, \end{cases} \tag{3.16}$$

and $0 \le q \le p$.

As already mentioned, we need the H^1 projection error to estimate the error of the GD scheme. The previous theorem gives us:

$$\|\hat{\mathbf{u}} - \hat{\pi}_r^0 \hat{\mathbf{u}}\|_{1,\hat{K}} \le C r^{\frac{3}{2}-p} \|\hat{\mathbf{u}}\|_{p,\hat{K}}. \tag{3.17}$$

(3.17) shows that we do not have the optimality for the H^1 norm.

However, for $\hat{\pi}_r^1$, we can find in [30] the following estimate: $\forall t, s \in \mathbb{R}$ verifying $0 \le t \le 1 \le s$, then for $\hat{\mathbf{v}} \in \mathbf{H}^s(\hat{K})$, there exists a constant $C > 0$ independent of r such that:

$$\|\hat{\mathbf{v}} - \hat{\pi}_r^1 \hat{\mathbf{v}}\|_{t,\hat{K}} \le C r^{t-s} \|\hat{\mathbf{v}}\|_{s,\hat{K}}. \tag{3.18}$$

In particular, we will use the two estimates: ($t = 0, 1$ in (3.18)).

Proposition 2. *For* $\hat{\mathbf{v}} \in \mathbf{H}^s(\hat{K})$, $s \ge 1$,

$$\begin{aligned} \|\hat{\mathbf{v}} - \hat{\pi}_r^1 \hat{\mathbf{v}}\|_{0,\hat{K}} &\le \frac{C}{r^s} \|\hat{\mathbf{v}}\|_{s,\hat{K}}, \\ \|\hat{\mathbf{v}} - \hat{\pi}_r^1 \hat{\mathbf{v}}\|_{1,\hat{K}} &\le \frac{C}{r^{s-1}} \|\hat{\mathbf{v}}\|_{s,\hat{K}} \end{aligned} \tag{3.19}$$

where $C > 0$ *is a constant independent of* r.

In this case, we obtain the optimal projection errors ($1/r^s$ and $1/r^{s-1}$ for the L^2 and the H^1 norms respectively). In conclusion, we have decided to use the projector π_h^1 to analyze the convergence properties of the DG scheme in the *hp*-version.

3.3. ***hp*-projection errors.** To study the projection error introduced by $\hat{\pi}_r^1$, we use the bracket semi-norm: Let $u \in W^{m,p}(\hat{K})$,

$$[u]_{m,p,\hat{K}}^2 = \sum_{i=1}^{3} \left\| \frac{\partial^m u}{\partial \hat{x}_i^m} \right\|_{p,\hat{K}}^2. \tag{3.20}$$

and the Bramble-Hilbert lemma adapted to Q_r (see [33], [1], [2]):

Lemma 3.12 (Bramble-Hilbert)**.** *Let* p, q *be two numbers such that* $1 \le p,\ q \le \infty$ *and let* r, m *be two integers such that* $r \ge 0$ *and* $m \le r + 1$,

$$W^{r+1,p}(\hat{K}) \hookrightarrow W^{m,q}(\hat{K}). \tag{3.21}$$

Let $\Pi \in \mathscr{L}(W^{r+1,p}(\hat{K}); W^{m,q}(\hat{K}))$ *be an operator which verifies*

$$\forall p \in Q_r,\ \Pi p = p. \tag{3.22}$$

Then there exists C *dependent on* $\hat{K}$ *and* r *such that*

$$\forall v \in W^{r+1,p}(\hat{K}),\ |v - \Pi v|_{m,q,\hat{K}} \le C [v]_{r+1,p,\hat{K}}. \tag{3.23}$$

In (3.23), $|\cdot|_{m,q,\hat{K}}$ is the semi-norm defined by: Let $v \in W^{m,q}(\hat{K})$,

$$|v|_{m,q,\hat{K}} = \left(\sum_{|\alpha|=m} \int_{\hat{K}} |\frac{\partial^{|\alpha|}}{\partial \hat{\mathbf{x}}^{\alpha}} v|^q d\hat{\mathbf{x}} \right)^{\frac{1}{q}}.$$

The Bramble-Hilbert lemma applied to the operator $\hat{\pi}_r^1$, immediately leads to:

Proposition 3. *For $r \geq 0$ and $m \leq r+1$, there exists C dependent on $\hat{K}$ and r such that:*

$$\forall \hat{\mathbf{v}} \in \mathbf{H}^{r+1}(\hat{K}),\ |\hat{\mathbf{v}} - \hat{\pi}_r^1 \hat{\mathbf{v}}|_{m,\hat{K}} \leq C[\mathbf{v}]_{r+1,\hat{K}}. \tag{3.24}$$

In order to derive the *hp*-projection error estimates for π_h^1, we must specify the exact r-dependence of the constant C of (3.24). To do so, we come back to the proof of the Bramble-Hilbert lemma but directly considering π_h^1. The first step, to prove this type of result, is to write [1]: $\forall \hat{\mathbf{v}} \in \mathbf{H}^{r+1}(\hat{K})$,

$$\begin{aligned} |\hat{\mathbf{v}} - \hat{\pi}_r^1 \hat{\mathbf{v}}|_{m,\hat{K}} &\leq \|I - \hat{\pi}_r^1\|_{\mathscr{L}(\mathbf{H}^{r+1}(\hat{K}),H^m(\hat{K}))} \inf_{\hat{p}\in[Q_r(\hat{K})]^3} \|\hat{\mathbf{v}} + p\|_{r+1,\hat{K}} \\ &\leq C_1 \|I - \hat{\pi}_r^1\|_{\mathscr{L}(\mathbf{H}^{r+1}(\hat{K}),H^m(\hat{K}))} [\mathbf{v}]_{r+1,\hat{K}} \end{aligned} \tag{3.25}$$

where C_1 is independent of r.

By using (3.18), (3.25) we immediately get:

$$|\hat{\mathbf{v}} - \hat{\pi}_r^1 \hat{\mathbf{v}}|_{m,\hat{K}} \leq \frac{C_2(\hat{K})}{r^{r+1-m}} [\mathbf{v}]_{r+1,\hat{K}},\ 0 \leq m \leq r+1. \tag{3.26}$$

In order to determine the projector errors, we will need the following estimate :

Lemma 3.13. *Let $K \in \mathscr{T}_h$ and $v \in W^{m,p}(K)$. We have the estimate:*

$$[v \circ F_K]_{m,p,\hat{K}} \leq C \frac{h_K^m}{\rho_K^{\frac{3}{p}}} |v|_{m,p,K}. \tag{3.27}$$

If $\mathscr{T}_h$ belongs to a regular family of triangulation, we give:

$$[v \circ F_K]_{m,p,\hat{K}} \leq C \sigma^{\frac{3}{p}} h_K^{m-\frac{3}{p}} |v|_{m,p,K} \tag{3.28}$$

where $C > 0$ independent of K and r.

Proof. Note $F_K = (F_K^1, F_K^2, F_K^3)$. To prove this lemma, we use the property:

$$\partial^2_{\hat{x}_k^2} F_K^i = 0 \text{ for } i = 1,2,3, \tag{3.29}$$

because $F_K^i \in Q_1(\hat{K})$. □

Let $\mathbf{v} \in \mathbf{H}^{r+1}(K)$, $r \geq 0$.

Lemma 3.14. *There exists C independent of K and r such that:*

$$\begin{aligned} \|\mathbf{v} - \pi_K^1 \mathbf{v}\|_{0,K} &\leq C \frac{h_K^{\frac{1}{2}}}{r^{r+1}} [\hat{\mathbf{v}}]_{r+1,\hat{K}}, \\ |\mathbf{v} - \pi_K^1 \mathbf{v}|_{1,K} &\leq \frac{C}{h_K^{\frac{1}{2}} r^r} [\hat{\mathbf{v}}]_{r+1,\hat{K}}. \end{aligned} \tag{3.30}$$

Proof. We prove only the second inequality. It suffices to use the same process to obtain the first. Write $\mathbf{w} = \mathbf{v} - \pi_K^1 \mathbf{v} = (w_1, w_2, w_3)^*$ ($*$ reads for the transposition operator). We have

$$|\mathbf{w}|_{1,K}^2 = \sum_{i=1}^{3}\sum_{l=1}^{3} \int_{\hat{K}} |J_K||(\partial_{x_l} w_i) o F_K|^2 d\hat{\mathbf{x}} \tag{3.31}$$

where the notation ∂_{x_l} means $\dfrac{\partial}{\partial x_l}$. By definition, we have $\mathbf{w} = DF_K^{*-1} o F_K^{-1} \hat{\mathbf{w}} o F_K^{-1}$ where $\hat{\mathbf{w}} = \hat{\mathbf{v}} - \hat{\pi}_r^1 \hat{\mathbf{v}}$ and DF_K^* writes:

$$DF_K^* = \begin{pmatrix} \partial_{\hat{x}_1} x_1 & \partial_{\hat{x}_1} x_2 & \partial_{\hat{x}_1} x_3 \\ \partial_{\hat{x}_2} x_1 & \partial_{\hat{x}_2} x_2 & \partial_{\hat{x}_2} x_3 \\ \partial_{\hat{x}_3} x_1 & \partial_{\hat{x}_3} x_2 & \partial_{\hat{x}_3} x_3 \end{pmatrix} \tag{3.32}$$

where $x_i = F_K^i(\hat{\mathbf{x}})$ for $i = 1, 2, 3$.

Inverting this matrix with the help of the co-factors formula, we obtain:

$$DF_K^{*-1} = \frac{1}{J_K}$$

$$\begin{pmatrix} \partial_{\hat{x}_2} x_2 \partial_{\hat{x}_3} x_3 - \partial_{\hat{x}_2} x_3 \partial_{\hat{x}_3} x_2 & -\partial_{\hat{x}_2} x_1 \partial_{\hat{x}_3} x_3 + \partial_{\hat{x}_2} x_3 \partial_{\hat{x}_3} x_1 & \partial_{\hat{x}_2} x_1 \partial_{\hat{x}_3} x_2 - \partial_{\hat{x}_2} x_2 \partial_{\hat{x}_3} x_1 \\ -\partial_{\hat{x}_1} x_2 \partial_{\hat{x}_3} x_3 + \partial_{\hat{x}_1} x_3 \partial_{\hat{x}_3} x_2 & \partial_{\hat{x}_1} x_1 \partial_{\hat{x}_3} x_3 - \partial_{\hat{x}_1} x_3 \partial_{\hat{x}_3} x_1 & -\partial_{\hat{x}_1} x_1 \partial_{\hat{x}_3} x_2 + \partial_{\hat{x}_1} x_2 \partial_{\hat{x}_3} x_1 \\ \partial_{\hat{x}_1} x_2 \partial_{\hat{x}_2} x_3 - \partial_{\hat{x}_1} x_3 \partial_{\hat{x}_2} x_2 & -\partial_{\hat{x}_1} x_1 \partial_{\hat{x}_2} x_3 + \partial_{\hat{x}_1} x_3 \partial_{\hat{x}_2} x_1 & \partial_{\hat{x}_1} x_1 \partial_{\hat{x}_2} x_2 - \partial_{\hat{x}_1} x_2 \partial_{\hat{x}_2} x_1 \end{pmatrix}.$$

Note that $DF_K^{*-1} = \dfrac{1}{J_K}(m_{i,j})_{i,j=1,\cdots,3}$, so we have $w_i = \displaystyle\sum_{j=1}^{3} \frac{m_{i,j} o F_K^{-1}}{J_K o F_K^{-1}} \hat{w}_j o F_K^{-1}$.

Now, we derive the last expression with respect to x_l:

$$\begin{aligned} \partial_{x_l} w_i = \sum_{j=1}^{3} \Big[& \frac{\partial_{x_l}(m_{i,j} o F_K^{-1}) J_K o F_K^{-1} - m_{i,j} o F_K^{-1} \partial_{x_l}(J_K o F_K^{-1})}{(J_K o F_K^{-1})^2} \hat{w}_j o F_K^{-1} \\ & + \frac{m_{i,j} o F_K^{-1}}{J_K o F_K^{-1}} \partial_{x_l}(\hat{w}_j o F_K^{-1}) \Big] \end{aligned} \tag{3.33}$$

$$\begin{aligned} = \sum_{j=1}^{3} \Big[\sum_{k=1}^{3} & \frac{(\partial_{\hat{x}_k} m_{i,j}) o F_K^{-1} \partial_{x_l} \hat{x}_k J_K o F_K^{-1} - m_{i,j} o F_K^{-1} (\partial_{\hat{x}_k} J_K) o F_K^{-1} \partial_{x_l} \hat{x}_k}{(J_K o F_K^{-1})^2} \hat{w}_j o F_K^{-1} \\ & + \frac{m_{i,j} o F_K^{-1}}{J_K o F_K^{-1}} (\partial_{\hat{x}_k} \hat{w}_j) o F_K^{-1} \partial_{x_l} \hat{x}_k \Big] \end{aligned}$$

Note that

$$T_{i,j}^{k,l} = \frac{(\partial_{\hat{x}_k} m_{i,j}) \partial_{x_l} \hat{x}_k o F_K J_K - m_{i,j} (\partial_{\hat{x}_k} J_K) \partial_{x_l} \hat{x}_k o F_K}{(J_K)^2}$$

$$\tilde{T}_{i,j}^{k,l} = \frac{m_{i,j}}{J_K} \partial_{x_l} \hat{x}_k o F_K \tag{3.34}$$

so we can write:

$$(\partial_{x_l} w_i) o F_K = \sum_{j,k=1}^{3} \Big[T_{i,j}^{k,l} \hat{w}^j + \tilde{T}_{i,j}^{k,l} \partial_{\hat{x}_k} \hat{w}^j \Big]. \tag{3.35}$$

The mesh regularity leads to:

$$\begin{aligned} |T_{i,j}^{k,l}| &\leq \frac{C}{h_K^2}, \\ |\tilde{T}_{i,j}^{k,l}| &\leq \frac{C}{h_K^2} \end{aligned} \tag{3.36}$$

where $C > 0$ independent of K and r. Indeed, the definition of $m_{i,j}$ gives us $|m_{i,j}| \leq C h_K^2$ and $|\partial_{\hat{x}_k} m_{i,j}| \leq C h_K^2$ (keep in mind that $x_i = F_K^i(\mathbf{x})$ for $i \in [\![1,3]\!]$). Moreover, the estimates (3.3) imply $|\partial_{x_l} \hat{x}_k o F_K| \leq C/h_K$, $|J_K| \leq C h_K^3$, $|\partial_{\hat{x}_k} J_K| \leq C h_K^3$ and $|J_K| \geq C' h_K^3$. That allows us to obtain:

$$|(\partial_{x_l} w_i) o F_K|^2 \leq \frac{C}{h_K^4} \sum_{j,k=1}^{3} \Big[|\hat{w}_j|^2 + |\partial_{\hat{x}_k} \hat{w}_j|^2 \Big]. \tag{3.37}$$

Return to our semi-norm: Using (3.37), (3.31) leads to

$$\begin{aligned} |\mathbf{w}|_{1,K}^2 &\leq C \frac{\|J_K\|_{\infty,\hat{K}}}{h_K^4} \sum_{i=1}^{3} \sum_{l=1}^{3} \sum_{j,k=1}^{3} \int_{\hat{K}} \Big[|\hat{w}_j|^2 + |\partial_{\hat{x}_k} \hat{w}_j|^2 \Big] d\hat{x} \\ &\leq \frac{C}{h_K} \|\hat{\mathbf{w}}\|_{1,\hat{K}}^2. \end{aligned} \tag{3.38}$$

Finally (3.26) gives the lemma. □

The following step is to increase $[\hat{\mathbf{v}}]_{m,\hat{K}}$ by a power of h_K and $\|\mathbf{v}\|_{m,K}$.

Lemma 3.15. *Let $\mathbf{v} \in \mathbf{H}^m(K)$. We have the following estimate:*

$$[\hat{\mathbf{v}}]_{m,\hat{K}} \leq C \sum_{l=0}^{1} |F_K|_{l+1,\infty,\hat{K}} [\mathbf{v} o F_K]_{m-l,\hat{K}} \tag{3.39}$$

where $C > 0$ independent of K and r.

Proof. We have $\hat{\mathbf{v}} = DF_K^* \mathbf{v} o F_K$ and $[\hat{\mathbf{v}}]_{m,\hat{K}}^2 = \sum_{i=1}^{3} \sum_{j=1}^{3} \int_{\hat{K}} |\frac{\partial^m \hat{v}_j}{\partial \hat{x}_i^m}|^2 d\hat{x}$. We can write $\hat{v}_j = \sum_{k=1}^{3} J_{j,k} v_k o F$ where $DF_K^* = (J_{j,k})_{j,k=1,\cdots,3}$. The Leibniz formula leads to:

$$\frac{\partial^m \hat{v}_j}{\partial \hat{x}_i^m} = \sum_{k=1}^{3} \sum_{l=0}^{m} \begin{pmatrix} l \\ m \end{pmatrix} \frac{\partial^l (J_{j,k})}{\partial \hat{x}_i^l} \frac{\partial^{m-l} (v_k o F)}{\partial \hat{x}_i^{m-l}}. \tag{3.40}$$

For $l \geq 2$, we have $\dfrac{\partial^l (J_{j,k})}{\partial \hat{x}_i^l} = 0$ (indeed $F_K \in [Q_1(\hat{K})]^3$). That implies:

$$\begin{aligned} \int_{\hat{K}} |\frac{\partial^m \hat{v}_j}{\partial \hat{x}_i^m}|^2 d\hat{x} &\leq C \sum_{k=1}^{3} \sum_{l=0}^{1} |F_K|^2_{l+1,\infty,\hat{K}} \int_{\hat{K}} |\frac{\partial^{m-l}(v_k o F)}{\partial \hat{x}_i^{m-l}}|^2 d\hat{x} \\ &\leq C \sum_{k=1}^{3} \sum_{l=0}^{1} |F_K|^2_{l+1,\infty,\hat{K}} [v_k o F_K]^2_{m-l,\hat{K}}. \end{aligned} \tag{3.41}$$

So, we obtain the following result:

$$\begin{aligned} [\hat{\mathbf{v}}]^2_{m,\hat{K}} &\leq C \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{k=1}^{3} \sum_{l=0}^{1} |F_K|^2_{l+1,\infty,\hat{K}} [v_k o F_K]^2_{m-l,\hat{K}} \\ &\leq C \sum_{l=0}^{1} |F_K|^2_{l+1,\infty,\hat{K}} [\mathbf{v} o F_K]^2_{m-l,\hat{K}}. \quad \square \end{aligned} \tag{3.42}$$

Finally, by grouping (3.30), (3.28) and (3.39) together, we obtain the following error estimates:

Proposition 4. *Let $\mathbf{v} \in \mathbf{H}^{r+1}(K)$. Then there exists C independent of the cell K and r such that:*

$$\begin{aligned} \|\mathbf{v} - \pi_K^1 \mathbf{v}\|_{0,K} &\leq C \frac{h_K^r}{r^{r+1}} \big(|\mathbf{v}|_{r,K} + h_K |\mathbf{v}|_{r+1,K} \big), \\ |\mathbf{v} - \pi_K^1 \mathbf{v}|_{1,K} &\leq C \frac{h_K^{r-1}}{r^r} \big(|\mathbf{v}|_{r,K} + h_K |\mathbf{v}|_{r+1,K} \big). \end{aligned} \tag{3.43}$$

Now, by using the interpolation Theorem 1.4 of [33], we extend the result to the real exponents.

Proposition 5. *Let $\mathbf{v} \in \mathbf{H}^{s+1}(K)$, for $0 \leq s \leq r$ real and assume that $0 < h_K \leq 1$. Then there exists C independent of the cell K and r and such that:*

$$\begin{aligned} \|\mathbf{v} - \pi_K^1 \mathbf{v}\|_{0,K} &\leq C \frac{h_K^s}{r^{s+1}} \|\mathbf{v}\|_{s+1,K}, \\ |\mathbf{v} - \pi_K^1 \mathbf{v}|_{1,K} &\leq C \frac{h_K^{s-1}}{r^s} \|\mathbf{v}\|_{s+1,K}. \end{aligned} \tag{3.44}$$

Proof. Let $r1 < r2$ be two positive integers and $\theta \in [0,1]$. Assume that $\pi_K^0 \in \mathscr{L}(\mathbf{H}^{r1+1}(K), H^m(K)) \cap \mathscr{L}(\mathbf{H}^{r2+1}(K), H^m(K))$ for $m = 0, 1$. Then we have:

$$\begin{aligned} &\|I - \pi_K^1\|_{\mathscr{L}(\mathbf{H}^{\theta r1 + (1-\theta) r2 + 1}(K), H^m(K))} \\ &\leq C \|I - \pi_K^1\|^{\theta}_{\mathscr{L}(\mathbf{H}^{r1+1}(K), H^m(K))} \|I - \pi_K^1\|^{1-\theta}_{\mathscr{L}(\mathbf{H}^{r2+1}(K), H^m(K))}. \end{aligned}$$

The inequalities (3.43) lead to:

$$\|I - \pi_K^1\|_{\mathscr{L}(\mathbf{H}^{r1+1}(K), H^m(K))} \leq C \frac{h_K^{r1-m}}{r^{r1+1-m}},$$

$$\|I - \pi_K^1\|_{\mathscr{L}(\mathbf{H}^{r2+1}(K), H^m(K))} \leq C \frac{h_K^{r2-m}}{r^{r2+1-m}}.$$

So we obtain:

$$\|I-\pi_K^1\|_{\mathscr{L}(\mathbf{H}^{\theta r1+(1-\theta)r2+1}(K),H^m(K))}\leq C\frac{h_K^{\theta r1+(1-\theta)r2-m}}{r^{\theta r1+(1-\theta)r2+1-m}}.$$

Finally, take $r1=0$, $r2=r$ and $s=(1-\theta)$. We can write the inequality:

$$\begin{aligned}\|\mathbf{v}-\pi_K^1\mathbf{v}\|_{m,K}&\leq\|I-\pi_K^1\|_{\mathscr{L}(\mathbf{H}^{s+1}(K),H^m(K))}\|\mathbf{v}\|_{s+1,K}\\&\leq C\frac{h_K^{s-m}}{r^{s+1-m}}\|\mathbf{v}\|_{s+1,K}.\quad\square\end{aligned}$$

Now, if we take $\mathbf{v}\in\mathbf{H}^s(K)$ with $s\geq r+1$, we prove easily the error estimates:

$$\begin{aligned}\|\mathbf{v}-\pi_K^1\mathbf{v}\|_{0,K}&\leq C\frac{h_K^r}{r^s}\|\mathbf{v}\|_{s,K},\\|\mathbf{v}-\pi_K^1\mathbf{v}|_{1,K}&\leq C\frac{h_K^{r-1}}{r^{s-1}}\|\mathbf{v}\|_{s,K}.\end{aligned}\tag{3.45}$$

Finally, (3.43) and (3.45) lead to the global result: Let $\mathbf{v}\in\mathbf{H}^{s+1}(K)$ with $s\geq 0$:

$$\begin{aligned}\|\mathbf{v}-\pi_K^1\mathbf{v}\|_{0,K}&\leq C\frac{h_K^{\min(s,r)}}{r^{s+1}}\|\mathbf{v}\|_{s+1,K},\\|\mathbf{v}-\pi_K^1\mathbf{v}|_{1,K}&\leq C\frac{h_K^{\min(s-1,r-1)}}{r^s}\|\mathbf{v}\|_{s+1,K}\end{aligned}\tag{3.46}$$

where C is independent of the cell K and r.

4. A-priori error estimates for the spatial semi-discrete approximation

In this part, we consider that all the integrals are computed in an exact way. Let $(\mathbf{E},\mathbf{H})$ and $(\mathbf{E}_h,\mathbf{H}_h)$ be respectively the solutions of (2.1) and (2.10). Our goal is to estimate $\|\mathbf{E}-\mathbf{E}_h\|_{0,\Omega}$ and $\|\mathbf{H}-\mathbf{H}_h\|_{0,\Omega}$. For that, we introduce the energy norm:

$$\|(\mathbf{E},\mathbf{H})\|_*^2=\|\mathbf{E}\|_{0,\Omega,\underline{\underline{\varepsilon}}}^2+\|\mathbf{H}\|_{0,\Omega,\underline{\underline{\mu}}}^2.\tag{4.1}$$

The norm (4.1) is more adapted to our estimations because it appears naturally in the Maxwell equations. So, we prefer to estimate:

$$\|(\mathbf{E}-\mathbf{E}_h,\mathbf{H}-\mathbf{H}_h)\|_*=\sqrt{\|\mathbf{E}-\mathbf{E}_h\|_{0,\Omega,\underline{\underline{\varepsilon}}}^2+\|\mathbf{H}-\mathbf{H}_h\|_{0,\Omega,\underline{\underline{\mu}}}^2}.\tag{4.2}$$

Introduce the projection of the exact solution $(\mathbf{E},\mathbf{H})$ i.e. $(\pi_h^1\mathbf{E},\pi_h^1\mathbf{H})$ (we assume that $\mathbf{E}$ and $\mathbf{H}$ have the regularity necessary for the definition of projections in (4.2)):

$$\begin{aligned}&\|(\mathbf{E}-\mathbf{E}_h,\mathbf{H}-\mathbf{H}_h)\|_*^2\\&=\|\mathbf{E}-\pi_h^1\mathbf{E}+\pi_h^1\mathbf{E}-\mathbf{E}_h\|_{0,\Omega,\underline{\underline{\varepsilon}}}^2+\|\mathbf{H}-\pi_h^1\mathbf{H}+\pi_h^1\mathbf{H}-\mathbf{H}_h\|_{0,\Omega,\underline{\underline{\mu}}}^2\\&\leq\|\Delta_{\mathbf{E}}^P\|_{0,\Omega,\underline{\underline{\varepsilon}}}^2+\|\Delta_{\mathbf{E}}^I\|_{0,\Omega,\underline{\underline{\varepsilon}}}^2+2\|\Delta_{\mathbf{E}}^P\|_{0,\Omega,\underline{\underline{\varepsilon}}}\|\Delta_{\mathbf{E}}^I\|_{0,\Omega,\underline{\underline{\varepsilon}}}\\&+\|\Delta_{\mathbf{H}}^P\|_{0,\Omega,\underline{\underline{\mu}}}^2+\|\Delta_{\mathbf{H}}^I\|_{0,\Omega,\underline{\underline{\mu}}}^2+2\|\Delta_{\mathbf{H}}^P\|_{0,\Omega,\underline{\underline{\mu}}}\|\Delta_{\mathbf{H}}^I\|_{0,\Omega,\underline{\underline{\mu}}}\end{aligned}\tag{4.3}$$

where $\Delta^P_{\mathbf{E}} = \mathbf{E} - \pi^1_h\mathbf{E}$ (projection error) and $\Delta^I_{\mathbf{E}} = \mathbf{E}_h - \pi^1_h\mathbf{E}$ (interpolation error). We have the same thing for $\mathbf{H}$. Using the inequality $2ab \leq a^2 + b^2$, (4.3) becomes:

$$\|(\mathbf{E}-\mathbf{E}_h, \mathbf{H}-\mathbf{H}_h)\|^2_* \leq 2(\|(\Delta^P_{\mathbf{E}}, \Delta^P_{\mathbf{H}})\|^2_* + \|\big(\Delta^I_{\mathbf{E}}, \Delta^I_{\mathbf{H}}\big)\|^2_*). \tag{4.4}$$

To estimate the error introduced by the spatial approximation, we have to evaluate $\|(\Delta^P_{\mathbf{E}}, \Delta^P_{\mathbf{H}})\|_*$ and $\|\big(\Delta^I_{\mathbf{E}}, \Delta^I_{\mathbf{H}}\big)\|_*$. The estimation of the first term does not pose any problem, it is sufficient to use the projection errors of the previous section; on the other hand, the second term requires more work. This will be done in three steps: first we will set up the equations which will make it possible to evaluate this error, then we will present two trace lemmas which will be used to estimate the surface integrals and finally we will consecutively evaluate the interpolation error for the study in h and r.

4.1. **Orthogonal property.** Introducing $\big(\pi^1_h\mathbf{E}, \pi^1_h\mathbf{H}\big)$ in the semi-discrete DG system (without numerical integration) and taking $\boldsymbol{\phi}_{1h} = \Delta^I_{\mathbf{E}}$, we obtain:

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\big(\Delta^I_{\mathbf{E}}, \Delta^I_{\mathbf{E}}\big)_{0,K,\underline{\underline{\varepsilon}}} &= -\big(\frac{\partial}{\partial t}\pi^1_h\mathbf{E}, \Delta^I_{\mathbf{E}}\big)_{0,K,\underline{\underline{\varepsilon}}} + \big(\nabla\times\Delta^I_{\mathbf{H}}, \Delta^I_{\mathbf{E}}\big)_{0,K} \\
&+\big(\nabla\times\pi^1_h\mathbf{H}, \Delta^I_{\mathbf{E}}\big)_{0,K} - \big(\Delta^I_{\mathbf{E}}, \Delta^I_{\mathbf{E}}\big)_{0,K,\underline{\underline{\sigma}}} - \big(\pi^1_h\mathbf{E}, \Delta^I_{\mathbf{E}}\big)_{0,K,\underline{\underline{\sigma}}} \\
-\big(\mathbf{J}_s, \Delta^I_{\mathbf{E}}\big)_{0,K} &+ \big(\beta[\![\Delta^I_{\mathbf{H}}\times\mathbf{n}_K]\!]^K_{\partial K}, \Delta^I_{\mathbf{E}|K}\big)_{0,\partial K} + \big(\beta[\![\pi^1_h\mathbf{H}\times\mathbf{n}_K]\!]^K_{\partial K}, \Delta^I_{\mathbf{E}|K}\big)_{0,\partial K}.
\end{aligned} \tag{4.5}$$

It is easy to see that the exact solution verifies:

$$\begin{aligned}
&(\frac{\partial}{\partial t}\Delta^P_{\mathbf{E}}, \Delta^I_{\mathbf{E}})_{0,K,\underline{\underline{\varepsilon}}} + (\frac{\partial}{\partial t}\pi^1_h\mathbf{E}, \Delta^I_{\mathbf{E}})_{0,K,\underline{\underline{\varepsilon}}} - (\nabla\times\Delta^P_{\mathbf{H}}, \Delta^I_{\mathbf{E}})_{0,K} \\
&-(\nabla\times\pi^1_h\mathbf{H}, \Delta^I_{\mathbf{E}})_{0,K} + (\Delta^P_{\mathbf{E}}, \Delta^I_{\mathbf{E}})_{0,K,\underline{\underline{\sigma}}} \\
&+(\pi^1_h\mathbf{E}, \Delta^I_{\mathbf{E}})_{0,K,\underline{\underline{\sigma}}} + \big(\mathbf{J}_s, \Delta^I_{\mathbf{E}}\big)_{0,K} = 0.
\end{aligned} \tag{4.6}$$

Combine (4.5) and (4.6):

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\big(\Delta^I_{\mathbf{E}}, \Delta^I_{\mathbf{E}}\big)_{0,K,\underline{\underline{\varepsilon}}} &= -\big(\Delta^I_{\mathbf{E}}, \Delta^I_{\mathbf{E}}\big)_{0,K,\underline{\underline{\sigma}}} + (\frac{\partial}{\partial t}\Delta^P_{\mathbf{E}}, \Delta^I_{\mathbf{E}})_{0,K,\underline{\underline{\varepsilon}}} \\
&+\big(\nabla\times\Delta^I_{\mathbf{H}}, \Delta^I_{\mathbf{E}}\big)_{0,K} - (\nabla\times\Delta^P_{\mathbf{H}}, \Delta^I_{\mathbf{E}})_{0,K} + (\Delta^P_{\mathbf{E}}, \Delta^I_{\mathbf{E}})_{0,K,\underline{\underline{\sigma}}} \\
&+\big(\beta[\![\Delta^I_{\mathbf{H}}\times\mathbf{n}_K]\!]^K_{\partial K}, \Delta^I_{\mathbf{E}|K}\big)_{0,\partial K} + \big(\beta[\![\pi^1_h\mathbf{H}\times\mathbf{n}_K]\!]^K_{\partial K}, \Delta^I_{\mathbf{E}|K}\big)_{0,\partial K}.
\end{aligned} \tag{4.7}$$

Applying the same reasoning for the $\mathbf{H}$ equation, we have:

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\big(\Delta^I_{\mathbf{H}}, \Delta^I_{\mathbf{H}}\big)_{0,K,\underline{\underline{\mu}}} &= (\frac{\partial}{\partial t}\Delta^P_{\mathbf{H}}, \Delta^I_{\mathbf{H}})_{0,K,\underline{\underline{\mu}}} - \big(\nabla\times\Delta^I_{\mathbf{E}}, \Delta^I_{\mathbf{H}}\big)_{0,K} \\
&+(\nabla\times\Delta^P_{\mathbf{E}}, \Delta^I_{\mathbf{H}})_{0,K} + \big(\gamma[\![\Delta^I_{\mathbf{E}}\times\mathbf{n}_K]\!]^K_{\partial K}, \Delta^I_{\mathbf{H}|K}\big)_{0,\partial K} \\
&+\big(\gamma[\![\pi^1_h\mathbf{E}\times\mathbf{n}_K]\!]^K_{\partial K}, \Delta^I_{\mathbf{H}|K}\big)_{0,\partial K}.
\end{aligned} \tag{4.8}$$

The Green formula gives:

$$(\nabla\times\Delta^I_{\mathbf{H}}, \Delta^I_{\mathbf{E}})_{0,K} = (\Delta^I_{\mathbf{H}}, \nabla\times\Delta^I_{\mathbf{E}})_{0,K} + (\Delta^I_{\mathbf{H}|K}, \Delta^I_{\mathbf{E}|K}\times\mathbf{n}_K)_{0,\partial K}. \tag{4.9}$$

Adding (4.7) and (4.8), we obtain:

$$
\begin{aligned}
&\frac{1}{2}\frac{d}{dt}\big[(\Delta^I_{\mathbf{E}},\Delta^I_{\mathbf{E}})_{0,K,\underline{\underline{\varepsilon}}}+(\Delta^I_{\mathbf{H}},\Delta^I_{\mathbf{H}})_{0,K,\underline{\underline{\mu}}}\big]\\
&=\big[(\frac{\partial}{\partial t}\Delta^P_{\mathbf{E}},\Delta^I_{\mathbf{E}})_{0,K,\underline{\underline{\varepsilon}}}+(\frac{\partial}{\partial t}\Delta^P_{\mathbf{H}},\Delta^I_{\mathbf{H}})_{0,K,\underline{\underline{\mu}}}\big]+(\nabla\times\Delta^P_{\mathbf{E}},\Delta^I_{\mathbf{H}})_{0,K}\\
&-(\nabla\times\Delta^P_{\mathbf{H}},\Delta^I_{\mathbf{E}})_{0,K}-\big(\Delta^I_{\mathbf{E}},\Delta^I_{\mathbf{E}}\big)_{0,K,\underline{\underline{\sigma}}}+(\Delta^P_{\mathbf{E}},\Delta^I_{\mathbf{E}})_{0,K,\underline{\underline{\sigma}}}\\
&+\big(\beta[\![\Delta^I_{\mathbf{H}}\times\mathbf{n}_K]\!]^K_{\partial K},\Delta^I_{\mathbf{E}|K}\big)_{0,\partial K}+\big(\beta[\![\pi^1_h\mathbf{H}\times\mathbf{n}_K]\!]^K_{\partial K},\Delta^I_{\mathbf{E}|K}\big)_{0,\partial K}\\
&+(\Delta^I_{\mathbf{H}|K},\Delta^I_{\mathbf{E}|K}\times\mathbf{n}_K)_{0,\partial K}-\big(\gamma[\![\Delta^I_{\mathbf{E}}\times\mathbf{n}_K]\!]^K_{\partial K},\Delta^I_{\mathbf{H}|K}\big)_{0,\partial K}\\
&+\big(\gamma[\![\pi^1_h\mathbf{E}\times\mathbf{n}_K]\!]^K_{\partial K},\Delta^I_{\mathbf{H}|K}\big)_{0,\partial K}.
\end{aligned}
\tag{4.10}
$$

We know that $\forall t\in(0,T)$, $(\mathbf{E},\mathbf{H})(t)\in H_0(rot,\Omega)\times H(rot,\Omega)$, so we have $\forall\Gamma=(K\cap K')\in\mathscr{F}^i_h$, $[\![\mathbf{E}\times\mathbf{n}_K]\!]^{K\text{ ou }K'}_\Gamma=0$ and $[\![\mathbf{H}\times\mathbf{n}_K]\!]^{K\text{ ou}K'}_\Gamma=0$. Moreover, keep in mind that $\forall\Gamma\in\mathscr{F}^b_h$, $\beta=0$.

By summing (4.10) over all the cells of the mesh and using the previous properties, we can write:

$$
\begin{aligned}
&\frac{1}{2}\frac{d}{dt}\|(\Delta^I_E,\Delta^I_H)\|^2_*=\frac{1}{2}\frac{d}{dt}\sum_{K\in\mathscr{T}_h}\big[(\Delta^I_{\mathbf{E}},\Delta^I_{\mathbf{E}})_{0,K,\underline{\underline{\varepsilon}}}+(\Delta^I_{\mathbf{H}},\Delta^I_{\mathbf{H}})_{0,K,\underline{\underline{\mu}}}\big]\\
&\le\sum_{K\in\mathscr{T}_h}\Big[\big|(\frac{\partial}{\partial t}\Delta^P_{\mathbf{E}},\Delta^I_{\mathbf{E}})_{0,K,\underline{\underline{\varepsilon}}}\big|+\big|(\frac{\partial}{\partial t}\Delta^P_{\mathbf{H}},\Delta^I_{\mathbf{H}})_{0,K,\underline{\underline{\mu}}}\big|+|(\nabla\times\Delta^P_{\mathbf{E}},\Delta^I_{\mathbf{H}})_{0,K}|\\
&+|\big(\nabla\times\Delta^P_{\mathbf{H}},\Delta^I_{\mathbf{E}}\big)_{0,K}|+|(\Delta^P_{\mathbf{E}},\Delta^I_{\mathbf{E}})_{0,K,\underline{\underline{\sigma}}}|+|\big(\beta[\![\Delta^P_{\mathbf{H}}\times\mathbf{n}_K]\!],\Delta^I_{\mathbf{E},K}\big)_{0,\partial K}|\\
&+|\big(\gamma[\![\Delta^P_{\mathbf{E}}\times\mathbf{n}_K]\!],\Delta^I_{\mathbf{H},K}\big)_{0,\partial K}|\Big]
\end{aligned}
\tag{4.11}
$$

To obtain (4.11), we have used the fact that $\big(\Delta^I_{\mathbf{E}},\Delta^I_{\mathbf{E}}\big)_{0,K,\underline{\underline{\sigma}}}\ge 0$ and to eliminate surface terms in $\Delta^I_{\mathbf{E}}$ and $\Delta^I_{\mathbf{H}}$, we have used the identity:

$$
\begin{aligned}
\sum_{K\in\mathscr{T}_h}\big(&\big(\beta[\![\Delta^I_{\mathbf{H}}\times\mathbf{n}_K]\!]^K_{\partial K},\Delta^I_{\mathbf{E}|K}\big)_{0,\partial K}-\big(\gamma[\![\Delta^I_{\mathbf{E}}\times\mathbf{n}_K]\!]^K_{\partial K},\Delta^I_{\mathbf{H}|K}\big)_{0,\partial K}\\
&+(\Delta^I_{\mathbf{H}|K},\Delta^I_{\mathbf{E}|K}\times\mathbf{n}_K)_{0,\partial K}\big)=0.
\end{aligned}
$$

4.2. **Trace lemmas.** To estimate the surface integrals, we will need several intermediate results:

Lemma 4.1. *Let $\mathbf{u}_h\in U_h$ and $K\in\mathscr{T}_h$; then there exists a constant $C>0$ independent of K and r such that:*

$$
(\mathbf{u}_{h|K},\mathbf{u}_{h|K})_{0,\partial K}\le C\sigma^{11}_K\frac{r^2}{\rho_K}(\mathbf{u}_h,\mathbf{u}_h)_{0,K}.
\tag{4.12}
$$

Moreover, if $\mathscr{T}_h$ belongs to a regular family of meshes, we have:

$$(\mathbf{u}_{h|K}, \mathbf{u}_{h|K})_{0,\partial K} \leq C\frac{r^2}{h_K}(\mathbf{u}_h, \mathbf{u}_h)_{0,K}. \tag{4.13}$$

Proof. We have:

$$\frac{\displaystyle\int_{\partial K} |\mathbf{u}_{hK}|^2 d\sigma}{\displaystyle\int_K |\mathbf{u}_{hK}|^2 dx} = \frac{\displaystyle\int_{\partial \hat{K}} |J_K| \|DF_K^{*-1}\hat{n}\| \big(DF_K^{-1}DF_K^{*-1}\hat{\mathbf{u}}_K\big)\cdot\hat{\mathbf{u}}_K d\hat{\sigma}}{\displaystyle\int_{\hat{K}} |J_K| \big(DF_K^{-1}DF_K^{*-1}\hat{\mathbf{u}}_K\big)\cdot\hat{\mathbf{u}}_K d\hat{x}}. \tag{4.14}$$

The estimations (3.3) lead to:

$$\frac{\displaystyle\int_{\partial K} |\mathbf{u}_{hK}|^2 d\sigma}{\displaystyle\int_K |\mathbf{u}_{hK}|^2 dx} \leq C\frac{\sigma_K^{11}}{\rho_K}\frac{\displaystyle\int_{\partial \hat{K}} \hat{\mathbf{u}}_K\cdot\hat{\mathbf{u}}_K d\hat{\sigma}}{\displaystyle\int_{\hat{K}} \hat{\mathbf{u}}_K\cdot\hat{\mathbf{u}}_K d\hat{x}}. \tag{4.15}$$

In [29], we can find the estimation:

$$\frac{\displaystyle\int_{\partial \hat{K}} \hat{\mathbf{u}}_K\cdot\hat{\mathbf{u}}_K d\hat{\sigma}}{\displaystyle\int_{\hat{K}} \hat{\mathbf{u}}_K\cdot\hat{\mathbf{u}}_K d\hat{x}} \leq Cr^2. \tag{4.16}$$

So, we obtain the wanted result. □

We will need the trace inequality also:

Lemma 4.2. *Let $K \in \mathscr{T}_h$. There exists $C > 0$ independent of K and r such that $\forall v \in H^1(K)$,*

$$\|v\|^2_{0,\partial K} \leq C(\|v\|_{0,K}\|\nabla v\|_{0,K} + \rho_K^{-1}\sigma_K^{-1}\|v\|^2_{0,K}). \tag{4.17}$$

Moreover, if $\mathscr{T}_h$ belongs to a regular family of meshes, we have:

$$\|v\|^2_{0,\partial K} \leq C(\|v\|_{0,K}\|\nabla v\|_{0,K} + h_K^{-1}\|v\|^2_{0,K}). \tag{4.18}$$

Proof. Let $K \in \mathscr{T}_h$ and $\mathbf{v} \in H^1(K)$. Pose $\hat{v} = voF_K$. So, we have the trace inequality:

$$\|\hat{v}\|^2_{0,\partial \hat{K}} \leq C(\|\hat{v}\|_{0,\hat{K}}\|\hat{\nabla}\hat{v}\|_{0,\hat{K}} + \|\hat{v}\|^2_{0,\hat{K}}). \tag{4.19}$$

See for example the annexes of [28] to obtain a proof of this result.

Now, we are going to return to the cell K. We have the estimations:

$$\bullet\|\hat{v}\|^2_{0,\partial\hat{K}} = \int_{\partial\hat{K}} \hat{v}^2 d\hat{\sigma} = \int_{\partial K} \frac{1}{|J_K|\|DF_K^{*-1}\hat{\mathbf{n}}\|} v^2 d\sigma$$

$$\geq \frac{1}{\|J_K\|_{\infty,\hat{K}}|F_K^{-1}|_{1,\infty,K}}\|v\|^2_{0,\partial K} \tag{4.20}$$

$$\geq C\frac{\sigma_K^3}{h_K^2}\|v\|^2_{0,\partial K} \text{ by using (3.3)},$$

$$\bullet\|\hat{v}\|^2_{0,\hat{K}} = \int_{\hat{K}} \hat{v}^2 d\hat{\mathbf{x}} = \int_K \frac{1}{|J_K|} v^2 d\mathbf{x} \leq \|J_K^{-1}\|_{\infty,K}\|v\|^2_{0,K} = \rho_K^{-3}\|v\|^2_{0,K}, \tag{4.21}$$

$$\bullet\|\hat{\nabla}\hat{v}\|^2_{0,\hat{K}} = \int_{\hat{K}} \hat{\nabla}\hat{v}\cdot\hat{\nabla}\hat{v} d\hat{\mathbf{x}} = \int_K \frac{1}{|J_K|} DF_K^*\nabla v \cdot DF_K^*\nabla v d\mathbf{x}$$

$$\leq C\frac{\sigma_K^2}{\rho_K}\|\nabla v\|^2_{0,K} \text{ by using (3.3)}. \tag{4.22}$$

(4.20) becomes:

$$C_1\frac{\sigma_K^3}{h_K^2}\|v\|^2_{0,\partial K} \leq C(C_2\frac{\sigma_K}{\rho_K^2}\|v\|_{0,K}\|\nabla v\|_{0,K} + C_3\frac{1}{\rho_K^3}\|v\|^2_{0,K}). \tag{4.23}$$

We obtain the wanted result. □

4.3. **Error estimates.** The use of (4.12) allows us to establish the following estimations of the surface terms:

$$\sum_{K\in\mathscr{T}_h} [|\left(\beta[\![\Delta^P_{\mathbf{H}}\times\mathbf{n}_K]\!],\Delta^I_{\mathbf{E}|K}\right)_{0,\partial K}| + |\left(\gamma[\![\Delta^P_{\mathbf{E}}\times\mathbf{n}_K]\!],\Delta^I_{\mathbf{H}|K}\right)_{0,\partial K}|]$$

$$\leq \sum_{K\in\mathscr{T}_h} [|\beta|\|[\![\Delta^P_{\mathbf{H}}\times\mathbf{n}_K]\!]\|_{0,\partial K}\|\Delta^I_{\mathbf{E}|K}\|_{0,\partial K}$$

$$+|\gamma|\|[\![\Delta^P_{\mathbf{E}}\times\mathbf{n}_K]\!]\|_{0,\partial K}\|\Delta^I_{\mathbf{H}|K}\|_{0,\partial K}] \tag{4.24}$$

$$\leq C\sum_{K\in\mathscr{T}_h} \sigma_K^{\frac{11}{2}}\frac{r}{\rho_K^{\frac{1}{2}}}[|\beta|\|[\![\Delta^P_{\mathbf{H}}\times\mathbf{n}_K]\!]\|_{0,\partial K}\|\Delta^I_{\mathbf{E}}\|_{0,K,\underline{\underline{\varepsilon}}}$$

$$+|\gamma|\|[\![\Delta^P_{\mathbf{E}}\times\mathbf{n}_K]\!]\|_{0,\partial K}\|\Delta^I_{\mathbf{H}}\|_{0,K,\underline{\underline{\mu}}}]$$

where C is a constant independent of K and r, but dependent on materials in the event here of B_1 defined in the first section.

(4.17) gives:

$$\begin{aligned}(4.25)\qquad &\|\Delta^P_{\mathbf{E}|K}\times\mathbf{n}_K\|^2_{0,\partial K}\le C(\|\Delta^P_{\mathbf{E}}\|_{0,K}\|\Delta^P_{\mathbf{E}}\|_{1,K}+\sigma_K^{-1}\rho_K^{-1}\|\Delta^P_{\mathbf{E}}\|^2_{0,K}),\\ &\|\Delta^P_{\mathbf{H}|K}\times\mathbf{n}_K\|^2_{0,\partial K}\le C(\|\Delta^P_{\mathbf{H}}\|_{0,K}\|\Delta^P_{\mathbf{H}}\|_{1,K}+\sigma_K^{-1}\rho_K^{-1}\|\Delta^P_{\mathbf{H}}\|^2_{0,K})\end{aligned}$$

where C is a constant independent of K and r.

Indeed, note that $\mathbf{v}=\Delta^P_{\mathbf{E}|K}=(v_1,v_2,v_3)^*$ and $\mathbf{n}_K=(n_1,n_2,n_3)^*$. We can then write:

$$(4.26)\qquad \Delta^P_{\mathbf{E}|K}\times\mathbf{n}_K=\mathbf{v}\times\mathbf{n}_K=(v_2n_3-v_3n_2,v_3n_1-v_1n_3,v_1n_2-v_2n_1)^*.$$

Now, developing $\|\Delta^P_{\mathbf{E}|K}\times\mathbf{n}_K\|^2_{0,\partial K}$, we get:

$$\begin{aligned}(4.27)\qquad &\|\Delta^P_{\mathbf{E}|K}\times\mathbf{n}_K\|^2_{0,\partial K}\\ &=\int_{\partial K}\big((v_2n_3-v_3n_2)^2+(v_3n_1-v_1n_3)^2+(v_1n_2-v_2n_1)^2\big)d\sigma\\ &\le 2\int_{\partial K}\big(v_2^2n_3^2+v_3^2n_2^2+v_3^2n_1^2+v_1^2n_3^2+v_1^2n_2^2+v_2^2n_1^2\big)d\sigma\\ &\le 4\int_{\partial K}\big(v_1^2+v_2^2+v_3^2\big)d\sigma=4\big(\|v_1\|^2_{0,\partial K}+\|v_2\|^2_{0,\partial K}+\|v_3\|^2_{0,\partial K}\big).\end{aligned}$$

To obtain the last inequality, we have used the fact that $n_1^2+n_2^2+n_3^2=1$. Applying (4.18) to $v_i\in H^1(K)$ (for $1\le i\le 3$), one deduces the inequalities:

$$(4.28)\qquad \|v_i\|^2_{0,\partial K}\le C(\|v_i\|_{0,K}\|\nabla v_i\|_{0,K}+\sigma_K^{-1}\rho_K^{-1}\|v_i\|^2_{0,K}).$$

Finally, introducing (4.28) into (4.27), we have:

$$\begin{aligned}(4.29)\qquad \|\Delta^P_{\mathbf{E}|K}\times\mathbf{n}_K\|^2_{0,\partial K}&\le 4C\sum_{i=1}^{3}(\|v_i\|_{0,K}\|\nabla v_i\|_{0,K}+\sigma_K^{-1}\rho_K^{-1}\|v_i\|^2_{0,K})\\ &\le 4C\sum_{i=1}^{3}(\|\mathbf{v}\|_{0,K}\|\mathbf{v}\|_{1,K}+\sigma_K^{-1}\rho_K^{-1}\|\mathbf{v}\|^2_{0,K})\\ &\le 12C(\|\mathbf{v}\|_{0,K}\|\mathbf{v}\|_{1,K}+\sigma_K^{-1}\rho_K^{-1}\|\mathbf{v}\|^2_{0,K}).\end{aligned}$$

We obtain the wanted result. For $\|\Delta^P_{\mathbf{H}|K}\times\mathbf{n}_K\|^2_{0,\partial K}$ it is obviously the same thing.

For the other terms of (4.11), we have the following estimates:

$$
\begin{aligned}
&\bullet \sum_{K\in\mathscr{T}_h} \left[|(\nabla\times\Delta^P_{\mathbf{E}},\Delta^I_{\mathbf{H}})_{0,K}| + |\left(\nabla\times\Delta^P_{\mathbf{H}},\Delta^I_{\mathbf{E}}\right)_{0,K}|\right]\\
&\leq C\sum_{K\in\mathscr{T}_h}\left[\|\Delta^P_{\mathbf{E}}\|_{1,K}\|\Delta^I_{\mathbf{H}}\|_{0,K,\underline{\underline{\mu}}} + \|\Delta^P_{\mathbf{H}}\|_{1,K}\|\Delta^I_{\mathbf{E}}\|_{0,K,\underline{\underline{\varepsilon}}}\right],\\
&\bullet \sum_{K\in\mathscr{T}_h} |\left(\frac{\partial}{\partial t}\Delta^P_{\mathbf{E}},\Delta^I_{\mathbf{E}}\right)_{0,K,\underline{\underline{\varepsilon}}}| + |\left(\frac{\partial}{\partial t}\Delta^P_{\mathbf{H}},\Delta^I_{\mathbf{H}}\right)_{0,K,\underline{\underline{\mu}}}|\\
&\leq C\sum_{K\in\mathscr{T}_h}\left(\|\Delta^P_{\mathbf{E}_t}\|_{0,K}\|\Delta^I_{\mathbf{E}}\|_{0,K,\underline{\underline{\varepsilon}}} + \|\Delta^P_{\mathbf{H}_t}\|_{0,K}\|\Delta^I_{\mathbf{H}}\|_{0,K,\underline{\underline{\varepsilon}}}\right),\\
&\bullet \sum_{K\in\mathscr{T}_h} |(\Delta^P_{\mathbf{E}},\Delta^I_{\mathbf{E}})_{0,K,\underline{\underline{\sigma}}}| \leq C\sum_{K\in\mathscr{T}_h}\|\Delta^P_{\mathbf{E}}\|_{0,K}\|\Delta^I_{\mathbf{E}}\|_{0,K,\underline{\underline{\varepsilon}}}
\end{aligned}
\tag{4.30}
$$

where C is a constant independent of K and r but dependent on the dielectric values of the medium and $\mathbf{u}_t = \frac{\partial}{\partial t}\mu$.

Remark 4.3. To obtain the second inequality of (4.30), we have used the property:

$$
\frac{\partial}{\partial t}\Delta^P_{\mathbf{E}} = \Delta^P_{\frac{\partial}{\partial t}\mathbf{E}} = \Delta^P_{\mathbf{E}_t}. \tag{4.31}
$$

We have the same thing for $\Delta^P_{\mathbf{H}}$.

Now, we are going to recombine the established estimates and use the projection errors of the previous section. Keep in mind that we use a regular family, $(\mathscr{T}_h)_{h>0}$, of meshes. We assume that $\mathbf{E}$, $\mathbf{H}\in\mathbf{H}^{s+1}(\mathscr{T}_h)\cap H(\mathrm{rot},\Omega)$, $\mathbf{E}_t$, $\mathbf{H}_t\in\mathbf{H}^{s'+1}(\mathscr{T}_h)$ and $\mathbf{J}_s\in\mathbf{H}^{s''+1}(\mathscr{T}_h)$ with $0\leq s,s',s''\leq r$ and $0<h_K\leq 1$, $\forall K\in\mathscr{T}_h$.

Using (3.46), (4.25) becomes:

$$
\begin{aligned}
\|\Delta^P_{\mathbf{E}|K}\times\mathbf{n}_K\|^2_{0,\partial K} &\leq C\frac{h_K^{\min(2s-1,2r-1)}}{r^{2s+1}}\|\mathbf{E}\|_{s+1,K},\\
\|\Delta^P_{\mathbf{H}|K}\times\mathbf{n}_K\|^2_{0,\partial K} &\leq C\frac{h_K^{\min(2s-1,2r-1)}}{r^{2s+1}}\|\mathbf{H}\|_{s+1,K}.
\end{aligned}
\tag{4.32}
$$

Thus, the boundary terms are bounded by:

$$
\begin{aligned}
&\sum_{K\in\mathscr{T}_h}[|\left(\beta[\![\Delta^P_{\mathbf{H}}\times\mathbf{n}_K]\!],\Delta^I_{\mathbf{E},K}\right)_{0,\partial K}| + |\left(\gamma[\![\Delta^P_{\mathbf{E}}\times\mathbf{n}_K]\!],\Delta^I_{\mathbf{H},K}\right)_{0,\partial K}|]\\
&\leq C\sum_{K\in\mathscr{T}_h}\frac{h_K^{\min(s-1,r-1)}}{r^{s-\frac{1}{2}}}[\|\mathbf{E}\|_{s+1,K}\|\Delta^I_{\mathbf{E},K}\|_{0,K,\underline{\underline{\varepsilon}}} + \|\mathbf{H}\|_{s+1,K}\|\Delta^I_{\mathbf{H},K}\|_{0,K,\underline{\underline{\mu}}}].
\end{aligned}
\tag{4.33}
$$

Here C depends on r.

We also have the estimates of (4.30):

$$
\begin{aligned}
&\bullet \sum_{K\in\mathscr{T}_h} \big[|(\nabla\times\Delta^P_{\mathbf{E}},\Delta^I_{\mathbf{H}})_{0,K}|+|\big(\nabla\times\Delta^P_{\mathbf{H}},\Delta^I_{\mathbf{E}}\big)_{0,K}|\big]\\
&\le C\sum_{K\in\mathscr{T}_h}\frac{h_K^{\min(s-1,r-1)}}{r^s}\big[\|\mathbf{E}\|_{s+1,K}\|\Delta^I_{\mathbf{H}}\|_{0,K,\underline{\underline{\mu}}}+\|\mathbf{H}\|_{s+1,K}\|\Delta^I_{\mathbf{E}}\|_{0,K,\underline{\underline{\varepsilon}}}\big],\\
&\bullet \sum_{K\in\mathscr{T}_h}\big|\big(\frac{\partial}{\partial t}\Delta^P_{\mathbf{E}},\Delta^I_{\mathbf{E}}\big)_{0,K,\underline{\underline{\varepsilon}}}\big|+\big|\big(\frac{\partial}{\partial t}\Delta^P_{\mathbf{H}},\Delta^I_{\mathbf{H}}\big)_{0,K,\underline{\underline{\mu}}}\big| \qquad (4.34)\\
&\le C\sum_{K\in\mathscr{T}_h}\frac{h_K^{\min(s',r)}}{r^{s'+1}}\big[\|\mathbf{E}_t\|_{s'+1,K}\|\Delta^I_{\mathbf{E}}\|_{0,K,\underline{\underline{\varepsilon}}}+\|\mathbf{H}_t\|_{s'+1,K}\|\Delta^I_{\mathbf{H}}\|_{0,K,\underline{\underline{\mu}}}\big],\\
&\bullet \sum_{K\in\mathscr{T}_h}|(\Delta^P_{\mathbf{E}},\Delta^I_{\mathbf{E}})_{0,K,\underline{\underline{\sigma}}}|\le C\sum_{K\in\mathscr{T}_h}\frac{h_K^{\min(s,r)}}{r^{s+1}}\|\mathbf{E}\|_{s+1,K}\|\Delta^I_{\mathbf{E}}\|_{0,K,\underline{\underline{\varepsilon}}}.
\end{aligned}
$$

Now, by using the fact that $\frac{\|\Delta^I_{\mathbf{E}}\|_{0,K,\underline{\underline{\varepsilon}}}}{\|(\Delta^I_E,\Delta^I_H)\|_*}\le 1$ and $\frac{\|\Delta^I_{\mathbf{H}}\|_{0,K,\underline{\underline{\mu}}}}{\|(\Delta^I_E,\Delta^I_H)\|_*}\le 1$, (4.11) leads to:

$$
\begin{aligned}
\frac{d}{dt}\|(\Delta^I_E,\Delta^I_H)\|_* \le C\sum_{K\in\mathscr{T}_h}\Big[&\frac{h_K^{\min(s-1,r-1)}}{r^{s-\frac12}}\big(\|\mathbf{E}\|_{s+1,K}+\|\mathbf{H}\|_{s+1,K}\big)\\
&+\frac{h_K^{\min(s',r)}}{r^{s'+1}}\big(\|\mathbf{E}_t\|_{s'+1,K}+\|\mathbf{H}_t\|_{s'+1,K}\big)\Big].
\end{aligned}
\qquad (4.35)
$$

Finally, the Gronwall lemma on the interval $(0,T)$ gives the following theorem:

Theorem 4.4. *Let r be a positive integer. Assume that the exact solution verifies $(\mathbf{E},\mathbf{H})\in\mathbf{H}^{s+1}(\mathscr{T}_h)$ and $(\mathbf{E}_t,\mathbf{H}_t)\in\mathbf{H}^{s'+1}(\mathscr{T}_h)$ for $s,s'\ge 0$ real and $0<h_K\le 1$, $\forall K\in\mathscr{T}_h$. Then,we have the global estimate of the interpolation error:*

$$
\begin{aligned}
&\|(\Delta^I_E,\Delta^I_H)\|_*(T)\le\|(\Delta^I_E,\Delta^I_H)\|_*(0)\\
&+CT\frac{h^{\min(s-1,s',r-1)}}{r^{\min(s-\frac12,s'+1)}}\max_{t\in(0,T)}\big(\|\mathbf{E}\|_{s+1,h}(t),\|\mathbf{H}\|_{s+1,h}(t),\\
&\cdot\|\mathbf{E}_t\|_{s'+1,h}(t),\|\mathbf{H}_t\|_{s'+1,h}(t)\big)
\end{aligned}
\qquad (4.36)
$$

where $C>0$ is a constant independent of K and r and $h=\max_{K\in\mathscr{T}_h}h_K$.

Return to the error of the scheme: According to (4.4), we have

$$\|(\mathbf{E}-\mathbf{E}_h,\mathbf{H}-\mathbf{H}_h)\|_*(T) \leq \sqrt{2}(\|(\Delta^P_{\mathbf{E}},\Delta^P_{\mathbf{H}})\|_* + \|\big(\Delta^I_{\mathbf{E}},\Delta^I_{\mathbf{H}}\big)\|_*)(T)$$
$$\leq \sqrt{2}\|(\Delta^I_E,\Delta^I_H)\|_*(0) + C\sqrt{2}\big[h^s \max\big(\|\mathbf{E}\|_{s+1,h}(T),\|\mathbf{H}\|_{s+1,h}(T)\big)$$
$$+T\frac{h^{\min(s-1,s',r-1)}}{r^{\min(s-\frac{1}{2},s'+1)}}\max_{t\in(0,T)}\big(\|\mathbf{E}\|_{s+1,h}(t),\|\mathbf{H}\|_{s+1,h}(t),\|\mathbf{E}_t\|_{s'+1,h}(t),\|\mathbf{H}_t\|_{s'+1,h}(t)\big)\big].$$

We see that the error seems to be sub-optimal and it increases at most linearly in time. Moreover, for $r=1$, the previous estimate does not prove the consistence of the scheme. In the last section of this paper, we will see with a simple numerical example that it is not clear that this scheme is consistent for a certain type of mesh.

Remark 4.5. If the mesh used is orthogonal or almost parallelepipedic, we find an exponent h^s. Indeed, we are respectively in an affine case and with second derivatives of F_K bounded by Ch_K^2.

4.4. **Error due to the numerical integration.** In this sub-section, we assume that the dielectric tensors are constant by cells and that we have conformal meshes. This last assumption allows us to have all discrete jump integrals (i.e. computed by using the Gauss rule) which are exact [34] i.e. $\forall \mathbf{u}_h$, $\mathbf{v}_h \in U_h$ we have:

$$\int_{\partial K}^{G}[\![\mathbf{u}_h\times\mathbf{n}_K]\!]\cdot\mathbf{v}_h d\sigma = \int_{\partial K}[\![\mathbf{u}_h\times\mathbf{n}_K]\!]\cdot\mathbf{v}_h d\sigma. \tag{4.37}$$

For technical reasons due to the use of a quadrature formula, we will need the interpolation operator I_h on U_h defined by: Let $\mathbf{v}\in[C^0(\mathscr{T}_h)]^3$ (i.e. $\mathbf{v}\in\mathbf{L}^2(\Omega)$ be such that $\forall K\in\mathscr{T}_h$ we have $\mathbf{v}_{|K}\in[C^0(K)]^3$); then $\forall K\in\mathscr{T}_h$,

$$I_{h|K}\mathbf{v}oF_K(\boldsymbol{\xi}_\mathbf{l}) = \mathbf{v}oF_K(\boldsymbol{\xi}_\mathbf{l}) \tag{4.38}$$

$\forall \mathbf{l}\in\{1,\cdots,r+1\}^3$. We can easily transpose the error estimates of the operator π_h^1 to I_h and we obtain, in particular: Let $\mathbf{v}\in\mathbf{H}^{s+1}(K)$, $s>\dfrac{1}{2}$ ($s>\dfrac{3}{2}$ to ensure the inclusion of $\mathbf{H}^s(K)$ in $[C^0(K)]^3$); then there exists C independent of the element K and r such that:

$$\|\mathbf{v}-I_{h|K}\mathbf{v}\|_{0,K} \leq C\frac{h_K^{\min(s,r)}}{r^{s+1}}\|\mathbf{v}\|_{s+1,K}. \tag{4.39}$$

To prove the r-dependence of (4.39), we have used the result in [31], $\mathbf{v}\in\mathbf{H}^{s+1}(K)$, $s>\dfrac{1}{2}$, then there exists a constant C independent of r such that:

$$\|\mathbf{v}-I_{h|K}\mathbf{v}\|_{0,K} \leq \frac{C}{r^{s+1}}\|\mathbf{v}\|_{s+1,K}. \tag{4.40}$$

Now, let us rewrite equation (4.10) by taking into account the Gauss quadrature rule:

$$
\begin{aligned}
\frac{1}{2}\frac{d}{dt}\int_K^G \big[\underline{\underline{\varepsilon}}\Delta_I^{\mathbf{E}}\cdot\Delta_I^{\mathbf{E}}+\underline{\underline{\mu}}\Delta_I^{\mathbf{H}}\cdot\Delta_I^{\mathbf{H}}\big]d\mathbf{x} &= \int_K \big[\underline{\underline{\varepsilon}}\frac{\partial}{\partial t}\Delta_P^{\mathbf{E}}\cdot\Delta_I^{\mathbf{E}}+\underline{\underline{\mu}}\frac{\partial}{\partial t}\Delta_P^{\mathbf{H}}\cdot\Delta_I^{\mathbf{H}}\big]d\mathbf{x}\\
&+\Big[\int_K \underline{\underline{\varepsilon}}\frac{\partial}{\partial t}\pi_h\mathbf{E}\cdot\Delta_I^{\mathbf{E}}d\mathbf{x}-\int_K^G \underline{\underline{\varepsilon}}\frac{\partial}{\partial t}\pi_h\mathbf{E}\cdot\Delta_I^{\mathbf{E}}\Big]+\Big[\int_K \underline{\underline{\mu}}\frac{\partial}{\partial t}\pi_h\mathbf{H}\cdot\Delta_I^{\mathbf{H}}d\mathbf{x}\\
&-\int_K^G \underline{\underline{\mu}}\frac{\partial}{\partial t}\pi_h\mathbf{H}\cdot\Delta_I^{\mathbf{H}}\Big]+\Big[\int_K \underline{\underline{\sigma}}\pi_h\mathbf{E}\cdot\Delta_I^{\mathbf{E}}d\mathbf{x}-\int_K^G \underline{\underline{\sigma}}\pi_h\mathbf{E}\cdot\Delta_I^{\mathbf{E}}\Big]\\
&+\Big[\int_K \mathbf{J}_s\cdot\Delta_I^{\mathbf{E}}d\mathbf{x}-\int_K^G \mathbf{J}_s\cdot\Delta_I^{\mathbf{E}}\Big]+\int_{\partial K}\beta[\![\Delta_I^{\mathbf{H}}\times\mathbf{n}_K]\!]\cdot\Delta_I^{\mathbf{E}}\\
&+\int_{\partial K}\beta[\![\pi_h\mathbf{H}\times\mathbf{n}_K]\!]\cdot\Delta_I^{\mathbf{E}}+\int_{\partial K}\gamma[\![\pi_h\mathbf{E}\times\mathbf{n}_K]\!]\cdot\Delta_I^{\mathbf{H}}\\
&+\int_{\partial K}\gamma[\![\Delta_I^{\mathbf{E}}\times\mathbf{n}_K]\!]\cdot\Delta_I^{\mathbf{H}}+\int_K \nabla\times\Delta_I^{\mathbf{H}}\cdot\Delta_I^{\mathbf{E}}d\mathbf{x}-\int_K \nabla\times\Delta_I^{\mathbf{E}}\cdot\Delta_I^{\mathbf{H}}d\mathbf{x}\\
&-\int_K \nabla\times\Delta_P^{\mathbf{H}}\cdot\Delta_I^{\mathbf{E}}d\mathbf{x}+\int_K \nabla\times\Delta_P^{\mathbf{E}}\cdot\Delta_I^{\mathbf{H}}d\mathbf{x}\\
&-\int_K^G \underline{\underline{\sigma}}\Delta_I^{\mathbf{E}}\cdot\Delta_I^{\mathbf{E}}+\int_K \underline{\underline{\sigma}}\Delta_P^{\mathbf{E}}\cdot\Delta_I^{\mathbf{E}}d\mathbf{x}.
\end{aligned}
\tag{4.41}
$$

To obtain (4.41), we have used the fact that the stiffness and the jump integrals are exact for all the elements belonging to the approximate space U_h. Indeed, for the stiffness terms we have the classical result: Let $\mathbf{u}_h$, $\mathbf{v}_h \in U_h$; we have $\forall K \in \mathscr{T}_h$,

$$
\begin{aligned}
\int_K^G \nabla\times\mathbf{u}_h\cdot\mathbf{v}_h &= \int_{\hat{K}}^G |J_K|\frac{DF_K}{J_K}\hat{\nabla}\times\hat{\mathbf{u}}_h\cdot DF_K^{*-1}\hat{\mathbf{v}}_h\\
&= \operatorname{sign}(J_K)\int_{\hat{K}}^G \hat{\nabla}\times\hat{\mathbf{u}}_h\cdot\hat{\mathbf{v}}_h\\
&= \operatorname{sign}(J_K)\int_{\hat{K}} \hat{\nabla}\times\hat{\mathbf{u}}_h\cdot\hat{\mathbf{v}}_h d\hat{\mathbf{x}} = \int_K \nabla\times\mathbf{u}_h\cdot\mathbf{v}_h d\mathbf{x}.
\end{aligned}
\tag{4.42}
$$

To write the last line of (4.42) we use the fact that the Gauss formula used is exact for all the polynomials in $Q_{2r+1}(\hat{K})$. This is why we have omitted the symbol G (for "Gauss") in these integrals.

As regards the discrete energy norm of the first line, one shows easily that it is equivalent to the energy norm without numerical integration. Indeed, let $\mathbf{u}_h \in U_h$; then we have:

$$
\int_K^G \mathbf{u}_h\cdot\mathbf{u}_h d\mathbf{x} = \int_{\hat{K}}^{\mathrm{G}} |J_K| DF^{-1}DF_K^{*-1}\hat{\mathbf{u}}_h\cdot\hat{\mathbf{u}}_h d\hat{x}.
\tag{4.43}
$$

By using the estimates (3.3) and (3.5), we can write immediately these two inequalities:

$$
C_1\frac{\rho_K^3}{h_K^2}\int_{\hat{K}}^{\mathrm{G}} \hat{\mathbf{u}}_h\cdot\hat{\mathbf{u}}_h d\hat{x} \le \int_K^G \mathbf{u}_h\cdot\mathbf{u}_h d\mathbf{x} \le C_2\frac{h_K^7}{\rho_K^6}\int_{\hat{K}}^{\mathrm{G}} \hat{\mathbf{u}}_h\cdot\hat{\mathbf{u}}_h d\hat{x}
\tag{4.44}
$$

where $C_1, C_2 > 0$ are independent of K and r.

Now, as the Gauss formula is exact to the order $2r+1$ when we use $(r+1)^3$ quadrature points, we have:

$$\int_{\hat{K}}^{\mathrm{G}} \hat{\mathbf{u}}_h \cdot \hat{\mathbf{u}}_h d\hat{x} = \int_{\hat{K}} \hat{\mathbf{u}}_h \cdot \hat{\mathbf{u}}_h d\hat{x}. \tag{4.45}$$

Returning to the cell K of the mesh:

$$\begin{aligned}\int_{\hat{K}} \hat{\mathbf{u}}_h \cdot \hat{\mathbf{u}}_h d\hat{x} &= \int_{\hat{K}} DF_K^* \mathbf{u}_h oF_K \cdot DF_K^* \mathbf{u}_h oF_K d\hat{x} \\ &= \int_{\hat{K}} |J_K| \frac{DF_K DF_K^*}{|J_K|} \mathbf{u}_h oF_K \cdot \mathbf{u}_h oF_K d\hat{x}.\end{aligned} \tag{4.46}$$

Again using (3.3) and (3.5), one obtains the following inequalities:

$$\frac{\rho_K^6}{C_2 h_K^7} \int_K \mathbf{u}_h \cdot \mathbf{u}_h d\mathbf{x} \le \int_{\hat{K}} \hat{\mathbf{u}}_h \cdot \hat{\mathbf{u}}_h d\hat{x} \le \frac{h_K^2}{C_1 \rho_K^3} \int_K \mathbf{u}_h \cdot \mathbf{u}_h d\mathbf{x}. \tag{4.47}$$

Combining (4.44) and (4.47), we get the wanted result i.e.,

$$\frac{C_1}{C_2} \sigma_K^9 \int_K \mathbf{u}_h \cdot \mathbf{u}_h d\mathbf{x} \le \int_K^G \mathbf{u}_h \cdot \mathbf{u}_h d\mathbf{x} \le \frac{C_2}{C_1} \sigma_K^9 \int_K \mathbf{u}_h \cdot \mathbf{u}_h d\mathbf{x}. \tag{4.48}$$

From this equivalence, one deduces the following result: The assumption of regularity of the mesh gives:

$$C\sigma^9 \int_K \left[\underline{\underline{\varepsilon}} \Delta_I^{\mathbf{E}} \cdot \Delta_I^{\mathbf{E}} + \underline{\underline{\mu}} \Delta_I^{\mathbf{H}} \cdot \Delta_I^{\mathbf{H}}\right] d\mathbf{x} \le \int_K^G \left[\underline{\underline{\varepsilon}} \Delta_I^{\mathbf{E}} \cdot \Delta_I^{\mathbf{E}} + \underline{\underline{\mu}} \Delta_I^{\mathbf{H}} \cdot \Delta_I^{\mathbf{H}}\right] d\mathbf{x} \tag{4.49}$$

where $C > 0$ is independent of K and r.

Now, we are going to estimate the first integration error of the second line of (4.41):

$$\begin{aligned}&\int_K \underline{\underline{\varepsilon}} \frac{\partial}{\partial t} \pi_h \mathbf{E} \cdot \Delta_I^{\mathbf{E}} d\mathbf{x} - \int_K^G \underline{\underline{\varepsilon}} \frac{\partial}{\partial t} \pi_h \mathbf{E} \cdot \Delta_I^{\mathbf{E}} = \int_K \underline{\underline{\varepsilon}} \pi_h \mathbf{E}_t \cdot \Delta_I^{\mathbf{E}} d\mathbf{x} \\ &- \int_K^G \underline{\underline{\varepsilon}} \pi_h \mathbf{E}_t \cdot \Delta_I^{\mathbf{E}} = \int_K \underline{\underline{\varepsilon}} \big(\pi_h \mathbf{E}_t - \mathbf{E}_t\big) \cdot \Delta_I^{\mathbf{E}} d\mathbf{x} + \int_K \underline{\underline{\varepsilon}} \mathbf{E}_t \cdot \Delta_I^{\mathbf{E}} d\mathbf{x} \\ &- \int_K^G \underline{\underline{\varepsilon}} \mathbf{E}_t \cdot \Delta_I^{\mathbf{E}} + \int_K^G \underline{\underline{\varepsilon}} \big(I_h \mathbf{E}_t - \pi_h \mathbf{E}_t\big) \cdot \Delta_I^{\mathbf{E}} \\ &\le \|\pi_h \mathbf{E}_t - \mathbf{E}_t\|_{0,K,\underline{\underline{\varepsilon}}} \|\Delta_I^{\mathbf{E}}\|_{0,K,\underline{\underline{\varepsilon}}} + \Big| \int_K \underline{\underline{\varepsilon}} \mathbf{E}_t \cdot \Delta_I^{\mathbf{E}} d\mathbf{x} - \int_K^G \underline{\underline{\varepsilon}} \mathbf{E}_t \cdot \Delta_I^{\mathbf{E}} \Big| \\ &+ C \|\pi_h \mathbf{E}_t - I_h \mathbf{E}_t\|_{0,K,\underline{\underline{\varepsilon}}} \|\Delta_I^{\mathbf{E}}\|_{0,K,\underline{\underline{\varepsilon}}} \\ &\le (1+C) \|\pi_h \mathbf{E}_t - \mathbf{E}_t\|_{0,K,\underline{\underline{\varepsilon}}} \|\Delta_I^{\mathbf{E}}\|_{0,K,\underline{\underline{\varepsilon}}} + C \|I_h \mathbf{E}_t - \mathbf{E}_t\|_{0,K,\underline{\underline{\varepsilon}}} \|\Delta_I^{\mathbf{E}}\|_{0,K,\underline{\underline{\varepsilon}}} \\ &+ \Big| \int_K \underline{\underline{\varepsilon}} \mathbf{E}_t \cdot \Delta_I^{\mathbf{E}} d\mathbf{x} - \int_K^G \underline{\underline{\varepsilon}} \mathbf{E}_t \cdot \Delta_I^{\mathbf{E}} \Big|\end{aligned} \tag{4.50}$$

where C is independent of K and r.

We obtain the previous inequalities by combining the Schwarz discrete inequality and the previous equivalence property. To estimate the last line of (4.50), only for the last term, additional work is necessary. First, we develop this term:

(4.51)
$$\Big|\int_K \underline{\underline{\varepsilon}}\mathbf{E}_t\cdot\Delta_I^{\mathbf{E}}d\mathbf{x}-\int_K^G \underline{\underline{\varepsilon}}\mathbf{E}_t\cdot\Delta_I^{\mathbf{E}}\Big|$$
$$=\Big|\int_{\hat K}|J_K|DF_K^{-1}\underline{\underline{\varepsilon}}oF_K DF_K^{*-1}\hat{\mathbf{E}}_t\cdot\hat\Delta_I^{\mathbf{E}}d\mathbf{x}-\int_{\hat K}^G|J_K|DF_K^{-1}\underline{\underline{\varepsilon}}oF_K DF_K^{*-1}\hat{\mathbf{E}}_t\cdot\hat\Delta_I^{\mathbf{E}}\Big|$$

where $\hat{\mathbf{E}}_t=DF_K^*\mathbf{E}_toF_K$ and $\hat\Delta_I^{\mathbf{E}}=\hat{\mathbf{E}}_h-\hat\pi_r^1\hat{\mathbf{E}}\in[Q_r(\hat K)]^3$.

Let $\hat w=|J_K|DF_K^{-1}\underline{\underline{\varepsilon}}oF_K DF_K^{*-1}\hat{\mathbf{E}}_t$. Introduce the interpolation polynomial $\hat I^r\hat{\mathbf{w}}$ in (4.51) and using the fact that the Gauss quadrature rule is exact for the polynomial space Q_{2r+1} when we take $(r+1)^3$ quadrature points, we get:

(4.52)
$$\Big|\int_K \underline{\underline{\varepsilon}}\mathbf{E}_t\cdot\Delta_I^{\mathbf{E}}d\mathbf{x}-\int_K^G \underline{\underline{\varepsilon}}\mathbf{E}_t\cdot\Delta_I^{\mathbf{E}}\Big|=\Big|\int_{\hat K}\hat{\mathbf{w}}\cdot\hat\Delta_I^{\mathbf{E}}d\mathbf{x}-\int_K\hat I^r\hat{\mathbf{w}}\cdot\hat\Delta_I^{\mathbf{E}}d\mathbf{x}\Big|$$
$$=\Big|\int_{\hat K}\big(\hat{\mathbf{w}}-\hat I^r\hat{\mathbf{w}}\big)\cdot\hat\Delta_I^{\mathbf{E}}d\mathbf{x}\Big|\le C\|\hat{\mathbf{w}}-\hat I^r\hat{\mathbf{w}}\|_{0,\hat K}\|\hat\Delta_I^{\mathbf{E}}\|_{0,\hat K,\underline{\underline{\varepsilon}}}$$

where C depends on C_1.

Using the Bramble-Hilbert lemma and the theory of the spectral methods [31], we can write the two following estimates for the interpolation operator $\hat I^r$:

(4.53)
$$\|\hat I^r(\hat{\mathbf{w}})-\hat{\mathbf{w}}\|_{0,\hat K}\le\frac{C(\hat K)}{r^{r+1}}[\hat{\mathbf{w}}]_{r+1,\hat K},$$
$$\|\hat I^r(\hat{\mathbf{w}})-\hat{\mathbf{w}}\|_{0,\hat K}\le\frac{C}{r^s}\|\hat{\mathbf{w}}\|_{s,\hat K}.$$

First, we are going to estimate the term $[\hat{\mathbf{w}}]_{r+1,\hat K}$. The definition of $\hat{\mathbf{w}}$ leads to:

(4.54)
$$[\hat{\mathbf{w}}]_{r+1,\hat K}=[|J_K|DF_K^{-1}\underline{\underline{\varepsilon}}oF_K\mathbf{E}_toF_K]_{r+1,\hat K}=[M_K\underline{\underline{\varepsilon}}oF_K\mathbf{E}_toF_K]_{r+1,\hat K}$$

where $M_K=(m_{i,j}^K)\in\mathscr{M}(3,3)$, the cofactor matrix of DF_K. Developing (4.54), we get:

(4.55)
$$[M_K\underline{\underline{\varepsilon}}oF_K\mathbf{E}_toF_K]_{r+1,\hat K}^2=\sum_{i=1}^3[\sum_{j=1}^3 m_{i,j}^K\sum_{k=1}^3\varepsilon_{j,k}\mathbf{E}_t^koF_K]_{r+1,\hat K}^2$$
$$\le 6\sum_{i=1}^3\sum_{j=1}^3\sum_{k=1}^3[m_{i,j}^K\varepsilon_{j,k}\mathbf{E}_t^koF_K]_{r+1,\hat K}^2$$
$$\le 6\sum_{i=1}^3\sum_{j=1}^3\sum_{k=1}^3\sum_{l=1}^3\int_{\hat K}\Big|\frac{\partial^{r+1}}{\partial\hat x_l^{r+1}}(m_{i,j}^K\varepsilon_{j,k}\mathbf{E}_t^koF_K)\Big|^2d\hat{\mathbf{x}}$$
$$\le 6\sum_{i=1}^3\sum_{j=1}^3\sum_{k=1}^3\sum_{l=1}^3\varepsilon_{j,k}\int_{\hat K}\Big|\sum_{m=0}^{r+1}\begin{pmatrix}m\\r+1\end{pmatrix}\frac{\partial^m}{\partial\hat x_l^m}(m_{i,j}^K)\frac{\partial^{r+1-m}}{\partial\hat x_l^{r+1-m}}(\mathbf{E}_t^koF_K)\Big|^2d\hat{\mathbf{x}}$$
(because $\underline{\underline{\varepsilon}}$ is constant within a cell).

It is easy to see that $\forall m \geq 3$, $\frac{\partial^m}{\partial \hat{x}_k^m}(m_{i,j}^K) = 0$ and that $0 \leq m \leq 2$, $\left|\frac{\partial^m}{\partial \hat{x}_k^m}(m_{i,j}^K)\right| \leq Ch_K^2$. That implies the following estimate:

$$
\begin{aligned}
&[M_K \underline{\underline{\varepsilon}} oF_K \mathbf{E}_t oF_K]^2_{r+1,\hat{K}} \\
&\leq C \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\sum_{l=1}^{3}\sum_{m=0}^{2}\int_{\hat{K}} \left|\frac{\partial^m}{\partial \hat{x}_l^m}(m_{i,j}^K)\frac{\partial^{r+1-m}}{\partial \hat{x}_l^{r+1-m}}(\mathbf{E}_t^k oF_K)\right|^2 d\hat{\mathbf{x}} \qquad (4.56)\\
&\leq Ch_K^4 \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\sum_{l=0}^{3}\sum_{m=0}^{2}\int_{\hat{K}} \left|\frac{\partial^{r+1-m}}{\partial \hat{x}_l^{r+1-m}}(\mathbf{E}_k^j oF_K)\right|^2 d\hat{\mathbf{x}}.
\end{aligned}
$$

Finally, we have:

$$
[\hat{\mathbf{w}}]^2_{r+1,\hat{K}} \leq Ch_K^4([\mathbf{E}_t oF_K]^2_{r+1,\hat{K}} + [\mathbf{E}_t oF_K]^2_{r,\hat{K}} + [\mathbf{E}_t oF_K]^2_{r-1,\hat{K}}). \qquad (4.57)
$$

Now, using (3.28), we deduce the following three estimates:

$$
\begin{aligned}
[\mathbf{E}_t oF_K]_{r+1,\hat{K}} &\leq Ch_K^{r-\frac{1}{2}}|\mathbf{E}_t|_{r+1,K},\\
[\mathbf{E}_t oF_K]_{r,\hat{K}} &\leq Ch_K^{r-\frac{3}{2}}|\mathbf{E}_t|_{r,K}, \qquad (4.58)\\
[\mathbf{E}_t oF_K]_{r-1,\hat{K}} &\leq Ch_K^{r-\frac{5}{2}}|\mathbf{E}_t|_{r-1,K}.
\end{aligned}
$$

Injecting (4.58) in (4.57), we obtain:

$$
[\hat{\mathbf{w}}]_{r+1,\hat{K}} \leq C\big(h_K^{r+\frac{3}{2}}|\mathbf{E}_t|_{r+1,K} + h_K^{r+\frac{1}{2}}|\mathbf{E}_t|_{r,K} + h_K^{r-\frac{1}{2}}|\mathbf{E}_t|_{r-1,K}\big). \qquad (4.59)
$$

Moreover, we have:

$$
\|\hat{\Delta}_I^E\|_{0,\hat{K}} \leq \frac{C}{h_K^{\frac{1}{2}}}\|\Delta_I^{\mathbf{E}}\|_{0,K,\underline{\underline{\varepsilon}}}. \qquad (4.60)
$$

Combining (4.59) and (4.60), we have the following error estimate for the interpolation operator:

$$
\begin{aligned}
\Big|\int_K^G \mathbf{E}_t\cdot\Delta_I^{\mathbf{E}}d\mathbf{x} - \int_K \mathbf{E}_t\cdot\Delta_I^{\mathbf{E}}d\mathbf{x}\Big| &\leq \frac{C}{r^{r+1}}\big(h_K^{r+1}|\mathbf{E}_t|_{r+1,K} + h_K^r|\mathbf{E}_t|_{r,K} \qquad (4.61)\\
&\quad + h_K^{r-1}|\mathbf{E}_t|_{r-1,K}\big)\|\Delta_I^{\mathbf{E}}\|_{0,K,\underline{\underline{\varepsilon}}}.
\end{aligned}
$$

Finally, using (4.61) and (4.39), (4.50) gives: For $0 < h_K \leq 1$,

$$
\Big|\int_K^G \underline{\underline{\varepsilon}}\pi_h\mathbf{E}_t\cdot\Delta_E^I - \int_K \underline{\underline{\varepsilon}}\pi_h\mathbf{E}_t\cdot\Delta_E^I d\mathbf{x}\Big| \leq C\frac{h_K^{r-1}}{r^{r+1}}\|\mathbf{E}_t\|_{r+1,K}\|\Delta_E^I\|_{0,K,\ tens\varepsilon}. \qquad (4.62)
$$

Proceeding in the same way, we prove:

$$
\begin{aligned}
&\bullet\Big|\int_K^G \underline{\underline{\mu}}\pi_h\mathbf{H}_t\cdot\Delta_H^I - \int_K \underline{\underline{\mu}}\pi_h\mathbf{H}_t\cdot\Delta_H^I d\mathbf{x}\Big| \le C\frac{h_K^{r-1}}{r^{r+1}}\|\mathbf{H}_t\|_{r+1,K}\|\Delta_H^I\|_{0,K,\underline{\underline{\varepsilon}}},\\
&\bullet\Big|\int_K^G \underline{\underline{\sigma}}\pi_h\mathbf{E}\cdot\Delta_E^I - \int_K \underline{\underline{\sigma}}\pi_h\mathbf{E}\cdot\Delta_E^I d\mathbf{x}\Big| \le C\frac{h_K^{r-1}}{r^{r+1}}\|\mathbf{E}\|_{r+1,K}\|\Delta_E^I\|_{0,K,\underline{\underline{\varepsilon}}},\\
&\bullet\Big|\int_K^G \mathbf{J}_s\cdot\Delta_E^I - \int_K \mathbf{J}_s\cdot\Delta_E^I d\mathbf{x}\Big| \le C\frac{h_K^{r-1}}{r^{r+1}}\|\mathbf{J}_s\|_{r+1,K}\|\Delta_E^I\|_{0,K,\underline{\underline{\varepsilon}}}.
\end{aligned}
\tag{4.63}
$$

From (4.62) and (4.63), we deduce that it suffices to add the error (after the temporal integration from 0 to T):

$$
C\frac{h^{\min(s-1,s'-1,s"-1,r-1)}}{r^{\min(s+1,s'+1,s''+1)}}\Big(\max_{t\in[0,T]}\big(\|\mathbf{E}_t\|_{s'+1,h},\|\mathbf{H}_t\|_{s'+1,h},\|\mathbf{E}\|_{s+1,h},\|\mathbf{J}_s\|_{s"+1,h}\big)\Big)T \tag{4.64}
$$

from $\dfrac{1}{2} < s, s', s" \le r$ (one had $s, s', s" \ge 0$ when the numerical integration was not used) for the space error estimate using the mass-lumping technique. Here C is a positive constant independent of K and r.

We conclude that the use of the Gauss quadrature formula can generate a deterioration of the spatial convergence ($s'-1$, $s''-1$) when the exact solution of the problem is not very regular inside at least a cell. Nevertheless, if the data of the treated problem are regular, then the mass-lumping does not generate a deterioration of the h convergence (i.e. h_K^{r-1}). Moreover, this reinforces the risk of inconsistency of the scheme using $r = 1$. Lastly, the behavior remains linear in time.

5. Numerical results

The aim of this part is to numerically verify whether the h-convergence rates obtained in the previous sections are optimal or not. To carry out this purpose, we study the propagation of a mode inside a perfectly metallic cubic cavity ($\mathbf{E}\times\mathbf{n} = 0$ on the wall of the cavity) with an edge of $a = 0.25$m. The propagative mode that we study is a mode $(m, n, 0)$ given by:

$$
\begin{cases}
E_x = 0;\ E_y = 0;\ H_z = 0,\\
E_z = \sin(m\pi\dfrac{x}{a})\sin(n\pi\dfrac{y}{a})\cos(\omega t),\\
H_x = \dfrac{\pi n}{a\omega\mu_0}\sin(m\pi\dfrac{x}{a})\cos(n\pi\dfrac{y}{a})\sin(\omega t),\\
H_y = \dfrac{\pi m}{a\omega\mu_0}\cos(m\pi\dfrac{x}{a})\sin(n\pi\dfrac{y}{a})\sin(\omega t),
\end{cases}
\tag{5.1}
$$

where $\omega = 3\cdot 10^8\sqrt{\big(\frac{m\pi}{a}\big)^2+\big(\frac{n\pi}{a}\big)^2}$.

By imposing this mode as an initial condition (i.e. for $t = 0$), the DG scheme gives an approached solution of (5.1). Hence, one knows the exact solution of our

problem. We can then compute the errors due to the DG scheme for some appropriated norms. More precisely, we have used two norms. The first is the classical L^2 norm ($\|\cdot\|_{0,\Omega}$) and the second is the norm

$$\|\mathbf{u}\|_h^2 = \|\mathbf{u}\|_{0,\Omega}^2 + \sum_{K\in\mathscr{T}_h} \|\nabla\times\mathbf{u}\|_{0,K}^2 + \sum_{\Gamma\in\mathscr{F}_h} \|[\![\mathbf{u}\times\mathbf{n}]\!]\|_{0,\Gamma}^2$$

which gives the classical $H(\mathrm{curl},\Omega)$ norm when μ is curl-conforming.

Moreover, we have chosen to take a sufficiently small time step in order to have a negligible time error.

Finally, we have used two types of mesh. The first is a "slightly deformed" cartesian type and the second is obtained by cutting each tetrahedron of a tetrahedrique mesh in four hexahedrons.

Remark 5.1. The simulations are carried out for the approximation Q_1 and Q_2. For highers orders, the computational cost rapidly becomes too important when one wants to obtain the asymptotic behaviour.

a- "slightly deformed" Cartesian grids:

The meshes are composed of $N\times N\times N$ cells, N being the number of subdivisions in each direction. For the Q_1 and Q_2 approximations, we have respectively taken $m=n=1$ and $m=n=3$. We have determined both the projection errors (i.e. for $t=0$) and the DG errors obtained after having covered one period (i.e. $t=\dfrac{\omega}{2\pi}$).

Tables 1 and 3 contain the results obtained for the L^2 projection of the initial condition. We find the theoretical rates, i.e., h^r for the L^2 norm and h^{r-1} for the norm $\|\cdot\|_h$. Indeed, the slight deformation has been made in order to obtain the estimation $|F_K|_{2,\infty,\hat{K}} \le Ch_K$ (and not h_K^2). Moreover, under this hypothesis, the theoretical results predicate that the L^2 error of the DG scheme is bounded by $O(h^{r-1})$. However, the results contained in Tables 2 and 3 show that the h-convergence rates are h^r for the norm $\|\cdot\|_{0,\Omega}$ and h^{r-1} for the norm $\|\cdot\|_h$. To conclude, for this type of mesh, the theoretical convergence rates obtained seem to be sub-optimals.

TABLE 1. Projection errors on "slightly deformed" Cartesian grids for Q_1.

Q_1	$8\times8\times8$	$16\times16\times16$	$32\times32\times32$	$64\times64\times64$
L^2 error	$1.2812\cdot10^{-2}$	$5.775\cdot10^{-3}$	$2.835\cdot10^{-3}$	$1.439\cdot10^{-3}$
L^2 order	X	1.14	1.026	0.98
$\|\cdot\|_h$ error	0.1938	0.1754	0.172	0.1736
$\|\cdot\|_h$ order	X	≈ 0	≈ 0	≈ 0

TABLE 2. Error after one period on "slightly deformed" Cartesian grids for Q_1.

Q_1	$8\times8\times8$	$16\times16\times16$	$32\times32\times32$	$64\times64\times64$
L^2 error	$2.794\cdot10^{-2}$	$1.611\cdot10^{-2}$	$8.342\cdot10^{-3}$	$4.189\cdot10^{-3}$
L^2 order	X	0.7	0.95	0.993
$\|\cdot\|_h$ error	0.25	0.263	0.267	0.268
$\|\cdot\|_h$ order	X	≈ 0	≈ 0	≈ 0

TABLE 3. Projection errors on "slightly deformed" Cartesian grids for Q_2.

Q_2	$16 \times 16 \times 16$	$32 \times 32 \times 32$	$64 \times 64 \times 64$
L^2 error	$1.708 \cdot 10^{-3}$	$3.658 \cdot 10^{-4}$	$8.772 \cdot 10^{-5}$
L^2 order	X	2.22	2.06
$\|\cdot\|_h$ error	$8.2951 \cdot 10^{-2}$	$3.513 \cdot 10^{-2}$	$1.675 \cdot 10^{-2}$
$\|\cdot\|_h$ order	X	1.24	1.06

TABLE 4. Error after one period on "slightly deformed" Cartesian grids for Q_2.

Q_2	$16 \times 16 \times 16$	$32 \times 32 \times 32$	$64 \times 64 \times 64$
L^2 error	$5.721 \cdot 10^{-3}$	$1.206 \cdot 10^{-3}$	$3.12 \cdot 10^{-4}$
L^2 order	X	2.24	1.95
$\|\cdot\|_h$ error	0.16	$5.947 \cdot 10^{-2}$	$2.607 \cdot 10^{-2}$
$\|\cdot\|_h$ order	X	1.42	1.18

b- General unstructured hexahedral meshes:
For this numerical experiment, we have used meshes obtained by cutting each tetrahedron of a tetrahedrique mesh in four hexahedrons. We have taken an initial mesh that we have successively refined. As for the previous example, we have the estimation $|F_K|_{2,\infty,\hat{K}} \leq Ch_K$. For the Q_1 and Q_2 approximations, we have taken $m = n = 1$.

Tables 5-6 and 7-8 contain the results obtained for the projection and the DG errors respectively. The first line of each table corresponds to the maximal spatial step (h). For the projection, the convergence rates conform to the theoretical results i.e. h^r for $\|\cdot\|_{0,\Omega}$ and h^{r-1} for $\|\cdot\|_h$. With regard to the DG errors, the Q_1 and Q_2 approximations seem to have convergence rates for the L^2- norm equal to h^0 and h^1 respectively i.e. h^{r-1}. This result seems to be confirmed by the errors obtained for $\|\cdot\|_h$. Indeed for this norm, the convergence rates are $h^{-\alpha}$ for Q_1 and h^{β} for Q_2, α and β seemingly tending respectively towards 1 and 0. For this type of mesh, the theoretical convergence rates seem to be optimal.

TABLE 5. Projection errors on general unstructured hexahedral meshes for Q_1.

Q_1	0.039	0.021	0.011
L^2 error	$8.4622 \cdot 10^{-3}$	$4.3179 \cdot 10^{-3}$	$2.1568 \cdot 10^{-3}$
L^2 order	X	1.08	1.07
$\|\cdot\|_h$ error	0.3575	0.3567	0.3544
$\|\cdot\|_h$ order	X	≈ 0	≈ 0

TABLE 6. Projection errors on general unstructured hexahedral meshes for Q_2.

Q_2	0.078	0.039	0.021
L^2 error	$8.0761 \cdot 10^{-4}$	$1.8373 \cdot 10^{-4}$	$4.9105 \cdot 10^{-5}$
L^2 order	X	2.13	2.13
$\|\cdot\|_h$ error	$2.4309 \cdot 10^{-2}$	$1.136 \cdot 10^{-2}$	$5.8978 \cdot 10^{-3}$
$\|\cdot\|_h$ order	X	1.09	1.05

TABLE 7. Error after one period on general unstructured hexahedral meshes for Q_1.

Q_1	0.039	0.021	0.011
L^2 error	$6.2637 \cdot 10^{-2}$	$3.6486 \cdot 10^{-2}$	$2.8541 \cdot 10^{-2}$
L^2 order	X	0.87	0.37
$\|\cdot\|_h$ error	1.168	1.326	2.04
$\|\cdot\|_h$ order	X	-0.208	-0.67

TABLE 8. Error after one period on general unstructured hexahedral meshes for Q_2.

Q_2	0.078	0.039	0.021
L^2 error	$6.6374 \cdot 10^{-3}$	$2.6737 \cdot 10^{-3}$	$1.1924 \cdot 10^{-3}$
L^2 order	X	1.31	1.3
$\|\cdot\|_h$ error	0.1568	0.1152	$9.8245 \cdot 10^{-2}$
$\|\cdot\|_h$ order	X	0.44	0.25

REFERENCES

1. Philippe G. Ciarlet, *The finite element method for elliptic problems*, North-Holland, 1978. MR0520174 (58:25001)
2. Achdou Yves, *The Finite Element Methods*, www.ann.jussieu.fr/achdou/enseignement.
3. Malika Remaki, *Méthodes numériques pour les équations de Maxwell instationnnaires en Milieu hétérogène*, Doctorat de Mathématiques Appliquées de l'Ecole Nationale des Ponts et Chaussées, 1999.
4. G. Cohen, *Higher-order numerical methods for transient wave equations*. Springer-Verlag, 2002. MR1870851 (2002m:65069)
5. K.S. Yee, Numerical solution of initial boundary value problems involving Maxwell's equation in isotropic media. *IEEE Trans. Antennas Prop.*, **14**, 302–307, 1966.
6. A. Taflove (ed.), *Advances in computational electrodynamics: The Finite-Difference Time-Domain*, Artech House, Boston, 1998. MR1639352 (99c:78001)
7. Andreas C. Cangellaris and Diana B. Wright, *Analysis of the Numerical Error Caused by the Stair-Stepped Approximation of a Conducting Boundary in FDTD Simulations of Electromagnetic Phenomena*, IEEE Trans. Antennas Prop., vol. AP-39, No. 10, pp. 1518-1525, October 1991.
8. J.S. Hesthavens and T. Warburton, *High-Order Nodal Methods on Unstructured Grids. I. Time-domain Solution of Maxwell's Equations*, J. Comput. Phys., vol. 181, pp. 1-34, 2002. MR1925981 (2003f:78034)

9. Garry Rodrigue and Daniel White, *A vector Finite Element Time-Domain Method for solving Maxwell's equations on unstructured hexahedral grids*, SIAM J. Sci. Comput., vol. 23, No. 3, pp. 683-706, 2001. MR1860960 (2002h:78036)
10. P. Bonnet and X. Ferrieres, *Numerical modeling of scattering problems using a time domain finite volume method*, JEWA, vol. 11, pp. 1165-1189, 1997.
11. S. Piperno and M. Remaki, and L. Fezoui, *A non-diffusive finite volume scheme for the 3D Maxwell equations on unstructured meshes*, SIAM J. Numer. Anal., vol. 39, No. 6, pp. 2089-2108, 2002. MR1897951 (2003e:65153)
12. Bernardo Cockburn, Fengyan Li, and Chi-Wang Shu, *Locally divergence-free discontinuous Galerkin methods for the Maxwell equations*, J. of Comput. Phys., vol. 194, pp. 588-610, 2004. MR2034859 (2004j:78024)
13. Bernardo Cockburn and Chi-Wang Shu, *The Runge-Kutta Discontinuous Galerkin Method for conservation law V*, J. Comput. Phys., vol. 141, pp. 199-224, 1998. MR1619652 (99c:65181)
14. B. Cockburn, G.E. Karniadakis, and C-W. Shu, *The Development of Discontinuous Galerkin Methods*, Lecture Notes in Computational Science and Engineering, vol. 11, Springer, 2000. MR1842161 (2002e:65002)
15. Nicolas Canouet, *Méthodes de Galerkin Discontinu pour la résolution du système de Maxwell sur des maillages localement raffinés non-conforme*,Doctorat de Mathématiques Appliquées de l'Ecole Nationale des Ponts et Chaussées, December 2003.
16. Paul Houston, Ilaria Perugia, and Dominik Shötzau, *Mixed Discontinuous Galerkin approximation of the Maxwell operator*, SIAM J. Numer. Anal., vol. 42, No. 1, pp. 434-459, 2004. MR2051073 (2005b:65128)
17. I. Perugia, D. Schötzau, and P. Monk, *Stabilized interior penalty methods for time-harmonic Maxwell equations*, Comput. Methods Appl. Mech. Eng., 191 (2002), pp.4675-4697. MR1929626 (2003j:78058)
18. W. Reed and T. Hill, *Triangular mesh methods for the neutron transport equation*, Tech. Report LA-UR-73-479, Los Alamos National Laboratory, Los Alamos, New Mexico, USA, 1973.
19. P. Lesaint and P. Raviart, *On a finite element method for solving the neutron transport equation*, in Mathematical Aspects of Finite Element Methods in Partial Differential Equations, C. deBoor, ed., Academic Press, New York, 1974, pp. 89-123. MR0658142 (58:31918)
20. C. Johnson and J. Pitkäranta, *An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation*, Math. Comp. 46 (1986), pp. 1–26. MR815828 (88b:65109)
21. J.-C. Nédélec, *Mixed finite elements in* $\mathbb{R}^3$, Numer. Math., 35(3), pp. 315-341, 1980. MR592160 (81k:65125)
22. J.-C. Nédélec, *A new family of mixed finite elements in* $\mathbb{R}^3$, Numer. Math, 50(1), pp. 57-81, 1986. MR864305 (88e:65145)
23. G. Cohen and P. Monk, *Mur-Nedelec finite element schemes for Maxwell's equations*, Comp. Meth. in Appl. Mech. Eng., 169(3-4), pp. 197-217, 1999. MR1675684 (99k:78002)
24. Peter Monk and Gerald R. Richter, *A discontinuous Galerkin method for linear symmetric hyperbolic systems in inhomogeneous media*. J. Sci. Comp. 22/23 (2005), 443–477. MR2142205 (2006b:65144)
25. S. M. Rao, *Time domain electromagnetics*, Series Editor, David Irwin, Auburn University, Academic Press, 1999.
26. J. Jin, *The finite Element Method in Electromagnetics*, John Wiley & Sons, New York, 1993. MR1903357 (2004b:78019)
27. S. Pernet, X. Ferrieres, and G. Cohen, *An original finite element method to solve Maxwell's equations in time domain*, Proceedings of EMC Zurich'2003, 18-20 February 2003, Zurich, Switzerland.
28. S. Prudhomme, F. Pascal, T. Oden, and A. Romkes, *Review of a priori error estimation for Discontinuous Galerkin*, Orsay, 2000, 2000-02.
29. B. Rivière, M.F. Wheeler, and V. Girault, *A priori error estimates for finite element methods based on discontinuous approximation spaces for elliptic problems*, SIAM J. Numer. Anal., 2001, 39, 3, pp. 902-931. MR1860450 (2002g:65149)
30. C. Bernardi and Y. Maday, *Spectral Methods*, in Handbook of Numerical Analysis, vol. V by P.G. Ciarlet and J.L. Lions (eds.), Elsevier Sciences, North-Holland, Amsterdam, 1997. MR1470226

31. D. Gottlieb and J.S. Hesthaven, *Spectral methods for time-dependent problems*, Cambridge Press.
32. A. Elmkies, *Sur les éléments finis d'arête pour la résolution des équations de Maxwell en milieu anisotrope et pour des maillages quelconques*, Université Paris IX-Dauphine, 1998, Thèse de mathématiques appliquées à l'ingéniérie.
33. V. Girault and P-A. Raviart, *Finite element methods for Navier-Stoke equations*, 1986, Springer-Verlag, New York. MR851383 (88b:65129)
34. S. Pernet, *Etude de méthodes d'ordre élevé pour résoudre les équations de Maxwell dans le domaine temporel. Application à la détection et à la compatibilité électromagnétique*, Thesis, University of Paris, IX, November 2004.

CERFACS (European Centre for Research and Advanced Training in Scientific Computation) 42, Avenue Gaspard Coriolis, 31057 Toulouse Cedex 01, France
E-mail address: pernet@cerfacs.fr

ONERA, 2 avenue Edouard Belin, 31055 Toulouse, France
E-mail address: ferrieres@onecert.fr

ESAIM: M2AN 44 (2010) 781–801
DOI: 10.1051/m2an/2010023

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

A WELL-CONDITIONED INTEGRAL EQUATION FOR ITERATIVE SOLUTION OF SCATTERING PROBLEMS WITH A VARIABLE LEONTOVITCH BOUNDARY CONDITION

Sébastien Pernet[1]

Abstract. The construction of a well-conditioned integral equation for iterative solution of scattering problems with a variable Leontovitch boundary condition is proposed. A suitable parametrix is obtained by using a new unknown and an approximation of the transparency condition. We prove the well-posedness of the equation for any wavenumber. Finally, some numerical comparisons with well-tried method prove the efficiency of the new formulation.

Mathematics Subject Classification. 65R20, 15A12, 65N38, 65F10, 65Z05.

Received September 15, 2007. Revised June 22, 2009.
Published online March 17, 2010.

1. Introduction

This paper deals with the solution of electromagnetic scattering problems by an obstacle whose surface is covered by thin layers of imperfectly conductor materials. This type of materials is generally taken into account by imposing an impedance boundary condition like the Leontovitch condition [21] on the surface of the object. It was recognized that this type of boundary condition can be extensively used to get a tractable problem in numerous complex situations. A first example can be found in radar applications: objects are often partially coated by a thin dielectric layer to reduce the radar cross section of scattering waves; in this case, the direct scattering problem amounts to a mixed boundary value problem with Maxwell's equations posed on an unbounded domain and where on the coated part of the boundary the electromagnetic field satisfies an impedance boundary condition while on the remaining part of the boundary the tangential component of the total electric field vanishes. Another domain of application is the use of this condition as an absorbing boundary condition to limit the computational domain of a finite elements method [25]. This condition plays also a major role in the domains decomposition method for Maxwell's equations [5,13,31]. Thus, it appears crucial to have efficient numerical methods well suited for such boundary conditions.

In the frequency domain, the Boundary Integral Methods (BIM) are a very attractive tool to solve electromagnetic problems. Numerous studies have been conducted about the choice and resolution of an integral equation for many years. Among significant accomplishments (this list is not exhaustive): the development of equations which are well-posed for any frequency [22,24], the development of the Fast Multipole Method (FMM) which allows to reduce the matrix-vector product to $O(N\log(N))$ complexity and consequently gives the possibility to treat big problems (several millions degrees of freedom) [29] and more recently the emergence of "naturally"

Keywords and phrases. Electromagnetic scattering, boundary integral equations, impedance boundary condition, preconditioner.

[1] CERFACS, 42 avenue Gaspard Coriolis, 31057 Toulouse Cedex 01, France. sebastien.pernet@cerfacs.fr

well-conditioned formulations which are derived by finding a analytical preconditioner or parametrix (in sense of the pseudo-differential calculus) of the underlying integral operator [1,12,17,30]. The latter technique allows to obtain Fredholm integral equations of second kind which decompose under the form "Identity + C" where C is a compact operator and which lead to good rates of convergence when an iterative solver is used to solved the linear system. In particular, [1,17] have shown that the convergence rate is wavenumber (k) and mesh-size (h) independent (for the smooth surfaces). That is why this technique is attractive, compared to the classical algebraic preconditioners like SPAI (SParse Approximate Inverse) where the k and h-dependences are not obvious to take into account. The aim of this paper is to propose a well-conditioned integral equation to solve for the scattering problems with a variable Leontovitch boundary condition. This construction is based on an adequate parametrix which is obtained by using an approximation of the transparent condition [2,18]. The formulation is an extension to the impedance problems of this one proposed in [18] with furthermore first numerical results.

The outline of the paper is as follows. In Section 2, we present the mathematical model and the basic tools to derive the integral equations in electromagnetism. Section 3 is devoted to the construction of the well-conditioned integral equation. In Section 4, we prove that this equation corresponds to a compact perturbation of the identity operator and is one-to-one when the impedance operator and the surface are smooth. These last assumptions are necessary to use the pseudo-differential calculus. In Section 5, we present the discretization and the resolution of the integral formulation. Finally, in Section 6, some numerical test-cases allow us to show that we obtain good convergences rates for constant and piecewise constant impedance operators. Moreover, some comparisons with other integral method allow us to see that the proposed formulation is also competitive in terms of accuracy.

2. The scattering problem and integral representation

Let Ω^- be a Lipschitz polyhedron with a boundary Γ which is assumed to be simply connected and connected. The open complement of Ω^- in $\mathbb{R}^3$ is Ω^+. Vector $\mathbf{n}$ denotes the unit normal to Γ pointing into the exterior domain Ω^+ of Ω^-. The problem is to find the scattering electromagnetic fields $\mathbf{E}$ and $\mathbf{H}$ solution to the Maxwell system

$$\begin{cases} \mathrm{curl}\mathbf{E} - ikZ_0\mathbf{H} = 0 \text{ in } \Omega^+, \\ \mathrm{curl}\mathbf{H} + ikZ_0^{-1}\mathbf{E} = 0 \text{ in } \Omega^+, \end{cases} \tag{2.1}$$

completed with both the Silver-Müller radiation condition at infinity

$$\lim_{|x|\to\infty} |x| \left(\mathbf{E}(x) + Z_0 \frac{x}{|x|} \times \mathbf{H}(x) \right) = 0 \tag{2.2}$$

and the boundary conditions on the surface Γ

$$\mathbf{n} \times (\mathbf{E}_{|\Gamma} \times \mathbf{n}) - Z_0\eta\,(\mathbf{n} \times \mathbf{H}_{|\Gamma}) = \mathbf{g} \tag{2.3}$$

where $k > 0$ is the wavenumber, Z_0 is the intrinsic impedance of the vacuum, $\mathbf{g}$ is a data and $\eta(x)$ is an impedance function. The variations of $\eta(x)$ allows us to take into account the presence of different materials on the surface Γ of the obstacle.

We have the following existence and the uniqueness results (see [8] for a proof and [25] for details on this kind of techniques):

Theorem 2.1. *We assume that:*

- Ω^- *is a connected Lipschitz polyhedral domain.*
- $\mathbf{g} \in \mathbf{H}_T^0(\Gamma)$.
- $\eta \in L^\infty(\Gamma)$ *and is assumed to be a strictly-positive real-valued function.*

Then the exterior mixed boundary value problem (2.1)-(2.2)-(2.3) *has a unique solution which belongs to the space* $X_{\mathrm{loc}}(\Omega^+,\Gamma) := \{\mathbf{u} \in \mathbf{H}_{\mathrm{loc}}(\mathrm{curl},\Omega^+) : \mathbf{n} \times \mathbf{u}_{|\Gamma} \in \mathbf{L}_T^2(\Gamma)\}$, *where* $\mathbf{L}_T^2(\Gamma) = \mathbf{H}_T^0(\Gamma) = \{\mathbf{u} \in [L^2(\Gamma)]^3 : \mathbf{u}\cdot\mathbf{n} = 0\}$.

Remark 2.2. In what follows, we assume that the assumptions of Theorem 2.1 are verified. Moreover, the impedance function $\eta(x)$ is assumed to be piecewise constant.

Now, we are presenting some material from the classical scattering theory. First, we recall some functional spaces which allow to correctly define the trace and integral operators in the context of Lipschitz polyhedral domains [7]: the tangential trace operator γ_t defined by:

$$\begin{array}{rccc} \gamma_t \,:\, \mathbf{H}(\mathrm{curl},\Omega) & \to & \mathbf{H}_{\times}^{-1/2}(\mathrm{div}_\Gamma,\Gamma) \\ \mathbf{u} & \mapsto & \mathbf{n}\times\mathbf{u} \end{array} \tag{2.4}$$

is continuous, surjective and possesses a right inverse where $\mathbf{H}_{\times}^{-1/2}(\mathrm{div}_\Gamma,\Gamma) = \{\mathbf{v} \in \mathbf{H}_{\times}^{-1/2}(\Gamma) \,:\, \mathrm{div}_\Gamma \mathbf{v} \in \mathrm{H}^{-1/2}(\Gamma)\}$ with $\mathbf{H}_{\times}^{-1/2}(\Gamma)$ is the dual space of the Hilbert space $\mathbf{H}_{\times}^{1/2}(\Gamma) = \gamma_t(\mathrm{H}^1(\Omega))$ with respect to the pairing $\langle \gamma_t\mathbf{v}\,,\,\gamma_t\mathbf{u}\rangle_{\tau,\Gamma} = \displaystyle\int_\Gamma (\mathrm{curl}\mathbf{v}\cdot\mathbf{u} - \mathbf{v}\cdot\mathrm{curl}\mathbf{u})\mathrm{d}x$.

Any electromagnetic field in Ω^+ which is a sum of a plane wave $(\mathbf{E}^{\mathrm{inc}},\mathbf{H}^{\mathrm{inc}})$ and of a radiating field $(\mathbf{E},\mathbf{H})$ is uniquely determined by the knowledge of the two equivalent currents,

$$\mathbf{J}(x) = \mathbf{n}\times\mathbf{H}_{|\Gamma}(x) \text{ and } \mathbf{M}(x) = -\mathbf{n}\times\mathbf{E}_{|\Gamma}(x), \tag{2.5}$$

through the well known Stratton-Chu formulae [7,15,25]

$$\left\{\begin{array}{lll} \mathbf{E}(x) = & iZ_0\tilde{\mathrm{T}}\mathbf{J}(x) + \tilde{\mathrm{K}}\mathbf{M}(x) & x\in\Omega^+ \\ \mathbf{H}(x) = & -\tilde{\mathrm{K}}\mathbf{J}(x) + iZ_0^{-1}\tilde{\mathrm{T}}\mathbf{M}(x) & x\in\Omega^+, \end{array}\right. \tag{2.6}$$

where the respective potentials $\tilde{T}$ and $\tilde{K}$ are defined by

$$\begin{array}{rcl} \tilde{\mathrm{T}} \,:\, \mathbf{H}_{\times}^{-1/2}(\mathrm{div}_\Gamma,\Gamma) & \to & \mathbf{H}_{\mathrm{loc}}(\mathrm{curl}^2,\Omega^+\cup\Omega^-)\cap\mathbf{H}_{\mathrm{loc}}(\mathrm{div}0,\Omega^+\cup\Omega^-) \\ \mathbf{J} & \mapsto & \tilde{\mathrm{T}}\mathbf{J}(x) = k\displaystyle\int_\Gamma G(x,y)\mathbf{J}(y)\mathrm{d}\Gamma(y) + \frac{1}{k}\int_\Gamma \vec{\nabla}_x G(x,y)\mathrm{div}_\Gamma\mathbf{J}(y)\mathrm{d}\Gamma(y) \\ \tilde{\mathrm{K}} \,:\, \mathbf{H}_{\times}^{-1/2}(\mathrm{div}_\Gamma,\Gamma) & \to & \mathbf{H}_{\mathrm{loc}}(\mathrm{curl}^2,\Omega^+\cup\Omega^-)\cap\mathbf{H}_{\mathrm{loc}}(\mathrm{div}0,\Omega^+\cup\Omega^-) \\ \mathbf{J}(x) & \mapsto & \tilde{\mathrm{K}}\mathbf{J}(x) = \displaystyle\int_\Gamma \vec{\nabla}_y G(x,y)\times\mathbf{J}(y)\mathrm{d}\Gamma(y) \end{array} \tag{2.7}$$

and $G(x,y)$ is the fundamental solution for the radiating solution of the 3-D Helmholtz equation

$$G(x,y) = \frac{\exp(ik|x-y|)}{4\pi|x-y|}. \tag{2.8}$$

The tangential traces on Γ of the potentials $\tilde{\mathrm{T}}$ and $\tilde{\mathrm{K}}$ are known and one has [7,25]

$$\begin{aligned} \mathbf{n}\times(\mathbf{E}_{|\Gamma}\times\mathbf{n})(x) &= iZ_0\mathrm{T}\mathbf{J}(x) + \mathrm{K}\mathbf{M}(x) + \frac{1}{2}\mathbf{n}\times\mathbf{M}(x), \\ \mathbf{n}\times(\mathbf{H}_{|\Gamma}\times\mathbf{n})(x) &= -\mathrm{K}\mathbf{J}(x) - \frac{1}{2}\mathbf{n}\times\mathbf{J}(x) + iZ_0^{-1}\mathrm{T}\mathbf{M}(x), \end{aligned} \tag{2.9}$$

where T and K are defined by

$$\begin{array}{rccc} \mathrm{T}: & \mathbf{H}_\times^{-1/2}(\mathrm{div}_\Gamma,\Gamma) & \to & \mathbf{H}_\times^{-1/2}(\mathrm{curl}_\Gamma,\Gamma) \\ & \mathbf{J} & \mapsto & \{\gamma_t(\tilde{\mathrm{T}}\mathbf{J})\times\mathbf{n}\}_\Gamma \\ \mathrm{K}: & \mathbf{H}_\times^{-1/2}(\mathrm{div}_\Gamma,\Gamma) & \to & \mathbf{H}_\times^{-1/2}(\mathrm{curl}_\Gamma,\Gamma) \\ & \mathbf{J} & \mapsto & \{\gamma_t(\tilde{\mathrm{K}}\mathbf{J})\times\mathbf{n}\}_\Gamma \end{array} \tag{2.10}$$

where $\{\gamma_t A\times\mathbf{n}\} = \frac{1}{2}(\mathbf{n}^+\times(A\times\mathbf{n}^+)+\mathbf{n}^-\times(A\times\mathbf{n}^-))$ and $\mathbf{H}_\times^{-1/2}(\mathrm{curl}_\Gamma,\Gamma) = \{\mathbf{v}\in\mathbf{H}_\perp^{-1/2}(\Gamma) : \mathrm{curl}_\Gamma\mathbf{v}\in\mathrm{H}^{-1/2}(\Gamma)\}$ is the dual space $\mathbf{H}_\times^{-1/2}(\mathrm{div}_\Gamma,\Gamma)$ with $\mathbf{H}_\perp^{-1/2}(\Gamma)$ is the dual space of the Hilbert space $\mathbf{H}_\perp^{1/2}(\Gamma) = \mathbf{n}\times\gamma_t(\mathrm{H}^1(\Omega))$.

Using (2.5), we obtain

$$\begin{array}{rcll} \mathbf{M} & = & \frac{1}{2}\mathbf{M} - \mathbf{n}\times\mathrm{K}\mathbf{M} - iZ_0\mathbf{n}\times\mathrm{T}\mathbf{J} & \text{(11-a)} \\ \mathbf{n}\times\mathbf{J} & = & \frac{1}{2}\mathbf{n}\times\mathbf{J} + \mathrm{K}\mathbf{J} - iZ_0^{-1}\mathrm{T}\mathbf{M} & \text{(11-b).} \end{array} \tag{2.11}$$

These two relations hold whatever the boundary condition on Γ is. There are not independent: except for some exceptional values of k (interior resonance), they are indeed equivalent. When impedance boundary condition is considered, we have to add the boundary condition (2.3) or equivalently

$$\mathbf{n}\times\mathbf{M}(x) = \eta Z_0\mathbf{J}(x) + \mathbf{g}(x) \quad \text{for } x\in\Gamma. \tag{2.12}$$

The two unknowns $\mathbf{J},\mathbf{M}$ have to be determined using the four previous equations. Several boundary integral equations can be constructed to determine the currents, all of them amounts to combine (2.11) and (2.12) to get an equation with a unique solution. The derivations of some of these equations can be found for example in [4].

3. Well-conditioned integral equation

3.1. A general approach

In this part, we present a general approach to construct an inherently well-conditioned integral equation. For that, we follow the reasoning which is used in [1,17].

We consider the generic scattering problem: find the radiating electromagnetic field $(\mathbf{E},\mathbf{H})$ in Ω^+

- solution of the Maxwell equations (2.1) with radiation condition (2.2);
- and subjected to a boundary condition $\mathbf{B}(\mathbf{J},\mathbf{M}) = \mathbf{g}$ on the boundary Γ.

For example, the perfect conducting object corresponds to $\mathbf{B}(\mathbf{J},\mathbf{M}) = \mathbf{M} = -\mathbf{E}^{\mathrm{inc}}\times\mathbf{n}$ and the purely impedance case corresponds to $\mathbf{B}(\mathbf{J},\mathbf{M}) = \mathbf{M}+\eta Z_0\mathbf{n}\times\mathbf{J} = -\mathbf{n}\times\mathbf{g} = \tilde{\mathbf{g}}$.

Now, if we assume that our generic problem is well-posed, then one can formally define an operator $\mathbf{Y}^{ex}$ which links the data $\mathbf{g}$ to the electromagnetic currents $(\mathbf{J},\mathbf{M})$ solution of the problem in this way:

$$(\mathbf{J},\mathbf{M}) = \mathbf{Y}^{ex}\mathbf{g} = (\mathbf{Y}_J^{ex},\mathbf{Y}_M^{ex})\mathbf{g}. \tag{3.1}$$

Recall that the electromagnetic fields solution of the Maxwell equations (2.1) with radiation condition (2.2) can be parameterized by the currents $(\mathbf{J}, \mathbf{M})$ *via* (2.6) and so if we introduce (3.1) in these equations, we obtain:

$$\left\{ \begin{array}{rcll} \mathbf{E}(x) & = & \big(iZ_0\tilde{\mathrm{T}} \circ \mathbf{Y}_J^{ex} + \tilde{\mathrm{K}} \circ \mathbf{Y}_M^{ex}\big)\mathbf{g}(x), & x \in \Omega^+ \\ \mathbf{H}(x) & = & \big(-\tilde{\mathrm{K}} \circ \mathbf{Y}_J^{ex} + iZ_0^{-1}\tilde{\mathrm{T}} \circ \mathbf{Y}_M^{ex}\big)\mathbf{g}(x), & x \in \Omega^+. \end{array} \right. \tag{3.2}$$

Finally, if we form the boundary condition by using (3.2), we obviously obtain:

$$\mathbf{B} \circ \mathbf{P} \circ \mathbf{Y}^{ex}\,\mathbf{g} \;=\; \mathbf{g} \tag{3.3}$$

where $\mathbf{P}$ is the Calderón projector defined by (2.11):

$$\mathbf{P} = \left[\begin{array}{cc} \frac{1}{2}\mathrm{Id} - \mathbf{n} \times \mathrm{K} & iZ_0^{-1}\mathbf{n} \times \mathrm{T} \\ -iZ_0\mathbf{n} \times \mathrm{T} & \frac{1}{2}\mathrm{Id} - \mathbf{n} \times \mathrm{K} \end{array} \right]$$

where Id is the identity operator.

So, we obtain the crucial relation:

$$\mathbf{B} \circ \mathbf{P} \circ \mathbf{Y}^{ex} \;=\; \mathrm{Id}. \tag{3.4}$$

In conclusion, one can parameterized the electromagnetic field by a density $\mathbf{u}$,

$$\left\{ \begin{array}{rcl} \mathbf{E} & = & \big(iZ_0\tilde{\mathrm{T}} \circ \mathbf{Y}_J^{ex} + \tilde{\mathrm{K}} \circ \mathbf{Y}_M^{ex}\big)\mathbf{u} \\ \mathbf{H} & = & \big(-\tilde{\mathrm{K}} \circ \mathbf{Y}_J^{ex} + iZ_0^{-1}\tilde{\mathrm{T}} \circ \mathbf{Y}_M^{ex}\big)\mathbf{u} \end{array} \right. \tag{3.5}$$

which is the solution of the integral equation:

$$\mathbf{B} \circ \mathbf{P} \circ \mathbf{Y}^{ex}\mathbf{u} \;=\; \mathbf{g}. \tag{3.6}$$

In this particular case, $\mathbf{u} = \mathbf{g}$ because $\mathbf{B} \circ \mathbf{P} \circ \mathbf{Y}^{ex} = \mathrm{Id}$.

Now, we are going to exploit the properties (3.4), (3.5) and (3.6) to formally construct well-conditioned integral equations. Since the operator $\mathbf{Y}^{ex}$ is generally unknown in practice, we assume that we possess an approximation $\mathbf{Y}$ of this operator. The idea is to substitute $\mathbf{Y} = (\mathbf{Y}_J, \mathbf{Y}_M)$ to $\mathbf{Y}^{ex}$ in (3.5) and (3.6). For that, we choose to parameterize the radiating electromagnetic field by a density $\mathbf{u}$,

$$\left\{ \begin{array}{rcl} \mathbf{E} & = & \big(iZ_0\tilde{\mathrm{T}} \circ \mathbf{Y}_J + \tilde{\mathrm{K}} \circ \mathbf{Y}_M\big)\mathbf{u} \\ \mathbf{H} & = & \big(-\tilde{\mathrm{K}} \circ \mathbf{Y}_J + iZ_0^{-1}\tilde{\mathrm{T}} \circ \mathbf{Y}_M\big)\mathbf{u} \end{array} \right. \tag{3.7}$$

which is the solution of the integral equation:

$$\mathbf{B} \circ \mathbf{P} \circ \mathbf{Y}\mathbf{u} \;=\; \mathbf{g}. \tag{3.8}$$

Now, let us give some important remarks:

- The electromagnetic field defined by (3.7) is always a radiating field *i.e.* verifying the Maxwell equations and the radiation condition at the infinity.

- For all operator $\mathbf{Y}$ such that the integral equation (3.8) is well-posed, the electromagnetic fields ($\mathbf{E}$, $\mathbf{H}$) defined by (3.7) and (3.8) is the exact field of our initial problem. It is true even if $\mathbf{Y}$ is a "bad" approximation of the optimal operator $\mathbf{Y}^{ex}$.
- If $\mathbf{Y}$ is "sufficiently close" on $\mathbf{Y}^{ex}$, one expects that the integral operator $\mathbf{B} \circ \mathbf{P} \circ \mathbf{Y}$ leads to a well-conditioned linear system after discretization.
- The currents defined by $\mathbf{J}_u = \mathbf{Y}_J \mathbf{u}$ and $\mathbf{M}_u = \mathbf{Y}_M \mathbf{u}$ are fictitious.

3.2. Application to the impedance problem

3.2.1. *Formal derivation*

In our problem, the boundary condition is:

$$\mathbf{B}(\mathbf{J},\mathbf{M}) \quad = \quad \mathbf{M} + \eta Z_0 \mathbf{n} \times \mathbf{J} \quad = \quad \mathbf{g} \quad \text{on } \Gamma. \tag{3.9}$$

The existence and the uniqueness (Thm. 2.1) of the solution of problem (2.1)-(2.2)-(2.3) induce the existence of the operator $\mathbf{Y}^{ex} = (\mathbf{Y}_M^{ex}, \mathbf{Y}_J^{ex})$ and in particular, we have

$$\mathbf{M} + \eta Z_0 \mathbf{n} \times \mathbf{J} \quad = \quad (\mathbf{Y}_M^{ex} + \eta Z_0 \mathbf{n} \times \mathbf{Y}_J^{ex})\mathbf{g} \quad = \quad \mathbf{g} \quad \text{on } \Gamma. \tag{3.10}$$

(3.10) leads to a new expression of the operator $\mathbf{Y}^{ex}$

$$\mathbf{Y}^{ex} \quad = \quad (\mathrm{Id} - \eta Z_0 \mathbf{n} \times \mathbf{Y}_J^{ex}\,,\, \mathbf{Y}_J^{ex}). \tag{3.11}$$

Consequently, a candidate for the approximation operator $\mathbf{Y}$ can be chosen as:

$$\mathbf{Y} \quad = \quad (\mathrm{Id} - \eta Z_0 \mathbf{n} \times \mathbf{Y}_J\,,\, \mathbf{Y}_J) \tag{3.12}$$

where $\mathbf{Y}_J$ is an approximation of $\mathbf{Y}_J^{ex}$.

Now, we want to define an approximation of $\mathbf{Y}_J^{ex}$. For that, we introduce the so-called exact exterior admittance or Stecklov-Poincaré operator $\mathscr{Y}^{ex}$ of the surface Γ. Recall that the operator is defined in this way: let $(\mathbf{u},\mathbf{v})$ be the solution of the well-posed problem:

$$\begin{cases} \mathrm{curl}\mathbf{u} - ikZ_0\mathbf{v} & = & 0 & \text{in } \Omega^+ \\ \mathrm{curl}\mathbf{v} + ikZ_0^{-1}\mathbf{u} & = & 0 & \text{in } \Omega^+ \\ \mathbf{n} \times \mathbf{u} \in \mathbf{H}_\times^{-1/2}(\mathrm{div}_\Gamma, \Gamma) \text{ fixed on } \Gamma & & & \\ + \text{ radiation condition at infinity.} & & & \end{cases} \tag{3.13}$$

The operator $\mathscr{Y}^{ex}$ corresponds to the operator which from $\mathbf{n} \times \mathbf{v}$ gives $\mathbf{n} \times \mathbf{u} = -\mathscr{Y}^{ex}(\mathbf{n} \times \mathbf{v})$. In particular, the current ($\mathbf{M}$, $\mathbf{J}$) on Γ induce by the field solution of the problem (2.1)-(2.3)-(2.2) are linked by the relation:

$$\mathbf{M} \quad = \quad \mathscr{Y}^{ex}(\mathbf{J}). \tag{3.14}$$

By using (3.9) and (3.14), one obtains a new expression of $\mathbf{Y}_J^{ex}$:

$$\mathbf{Y}_J^{ex} \quad = \quad \left(\mathscr{Y}^{ex} + \eta Z_0 \mathbf{n} \times \mathrm{Id}\right)^{-1}. \tag{3.15}$$

In conclusion, if one possesses an approximation $\mathscr{Y}$ of the admittance, we choose as candidate for the approximation of $\mathbf{Y}^{ex}$:

$$\begin{cases} \mathbf{Y} & = & (\mathrm{Id} - \eta Z_0 \mathbf{n} \times \mathbf{Y}_J\,,\, \mathbf{Y}_J) \\ \mathbf{Y}_J & = & \left(\mathscr{Y} + \eta Z_0 \mathbf{n} \times \mathrm{Id}\right)^{-1} \end{cases} \tag{3.16}$$

and the integral equation (3.8) becomes: find $\mathbf{u} \in \mathbf{H}_{\times}^{-1/2}(\mathrm{div}_\Gamma, \Gamma)$ such that

$$\mathscr{Z}\mathbf{u} \equiv \mathbf{B} \circ \mathbf{P} \circ \mathbf{Y}\mathbf{u} = \left(\frac{1}{2}\mathrm{Id} - \mathbf{n} \times \mathrm{K} \circ \mathbf{Y}_M - iZ_0\mathbf{n} \times \mathrm{T} \circ \mathbf{Y}_J + \eta Z_0 \mathrm{K} \circ \mathbf{Y}_J - i\eta \mathrm{T} \circ \mathbf{Y}_M\right)\mathbf{u} = \mathbf{g}. \tag{3.17}$$

3.2.2. *An approximation of the admittance for a smooth surface*

We choose an approximation based on by a microlocal analysis of the problem (3.13). This kind of approach has been already used by Antoine and Darbas for the construction of analytical preconditioners in electromagnetism for perfectly conductor objects. It is based on the knowledge of principal symbol of the admittance of a smooth surface and on an efficient approximation of the square root operator found in this symbol. They obtain an approximate operator which act on all zones: hyperbolic, elliptic and creeping zones. It is a regularization of the square root by an adequate small damping parameter which allows to correctly take into account the creeping modes. We will give the results which we briefly need and we refer to [16] for more details.

The principal symbol $\sigma(\mathscr{Y}^{ex})$ of $\mathscr{Y}^{ex}$ is [2]

$$\sigma(\mathscr{Y}^{ex})(\theta_2, \theta_3, k) = -Z_0 \left(1 - \frac{|\theta|^2}{k^2}\right)^{-1/2} \begin{pmatrix} -\dfrac{\theta_2\theta_3}{k^2} & 1 - \dfrac{\theta_3^2}{k^2} \\ -1 + \dfrac{\theta_2^2}{k^2} & \dfrac{\theta_2\theta_3}{k^2} \end{pmatrix} \tag{3.18}$$

where (θ_2, θ_3) is the dual variable of a particular parameterization of Γ, $|\theta|^2 = \theta_2^2 + \theta_3^2$ and k is the dual variable of the time t.

Now, if one returns in the primal variables by a inverse Fourier transform in space, one obtains the following approximation of $\mathscr{Y}^{ex}$:

$$\begin{array}{rcl} \mathscr{Y} \ : \ \mathbf{H}^{-1/2}(\mathrm{div}_\Gamma, \Gamma) & \to & \mathbf{H}^{-1/2}(\mathrm{div}_\Gamma, \Gamma) \\ \mathbf{v} & \mapsto & \mathscr{Y}\mathbf{u} = -Z_0 \left(\mathrm{Id} + \dfrac{\vec{\Delta}_\Gamma}{k^2}\right)^{-\frac{1}{2}} \left(\mathrm{Id} - \dfrac{1}{k^2}\vec{\mathrm{curl}}_\Gamma \mathrm{curl}_\Gamma\right) \circ \mathbf{n} \times \mathbf{v} \end{array} \tag{3.19}$$

where $\mathbf{H}^{-1/2}(\mathrm{div}_\Gamma, \Gamma) = \{\mathbf{u} \in \mathbf{H}_t^{-1/2} \ : \ \mathrm{div}_\Gamma \mathbf{u} \in H^{-1/2}\}$ is the classical space for a smooth surface.

For our convenience, we define another expression of $\mathscr{Y}$. For that, we consider the Helmholtz decomposition of tangential fields belonging to $\mathbf{H}^{-1/2}(\mathrm{div}_\Gamma, \Gamma)$: For each $\mathbf{v} \in \mathbf{H}^{-1/2}(\mathrm{div}_\Gamma, \Gamma)$, there exists two scalar potential $\psi_v \in H^{1/2}(\Gamma)$ and $\phi_v \in H^{3/2}(\Gamma)$ such that:

$$\mathbf{v} \quad = \quad \mathbf{n} \times \nabla_\Gamma \psi_v + \nabla_\Gamma \phi_v \tag{3.20}$$

and by reading $\mathscr{Y}$ as 2×2 matrix of operator and by posing $\mathbf{w} = \mathbf{n} \times \nabla_\Gamma \psi_w + \nabla_\Gamma \phi_w = \mathscr{Y}\mathbf{v}$, one can write:

$$\begin{pmatrix} \psi_w \\ \phi_w \end{pmatrix} \quad = \quad Z_0 \begin{pmatrix} 0 & \left(1 + \dfrac{\Delta_\Gamma}{k^2}\right)^{\frac{1}{2}} \\ -\left(1 + \dfrac{\Delta_\Gamma}{k^2}\right)^{-\frac{1}{2}} & 0 \end{pmatrix} \begin{pmatrix} \psi_v \\ \phi_v \end{pmatrix}. \tag{3.21}$$

The operator $ik\left(1 + \dfrac{\Delta_\Gamma}{k^2}\right)^{\frac{1}{2}}$ corresponds to the Dirichlet-To-Neumann operator for the acoustic scattering problem [3]. For an efficient approximation of this one, we introduce a damping parameter ε in order to perturb the wavenumber k by $k_\varepsilon = k + \mathrm{i}\varepsilon$ and so we regularize the singularity of the square root at the level

of the transition region of glancing rays. A suitable value of ε has been determined in [16]: $\varepsilon = 0.4k^{-1/3}\mathscr{H}^{-2/3}$ where $\mathscr{H}$ is the mean curvature of Γ. This value is optimum for the spherical geometries.

Finally, we explain how to compute accurately the square root. Recall that the square-root operator $SQR = \left(1 + \frac{\Delta_\Gamma}{k_\varepsilon^2}\right)^{\frac{1}{2}}$ is a non-local pseudo-differential operator. To realize an efficient estimation of SQR, we use the technique based on a Padé expansion of the square root and a rotating branch-cut technique [23] (θ corresponds to the angle of the rotation):

$$\left(1 + \frac{\Delta_\Gamma}{k_\varepsilon^2}\right)^{\frac{1}{2}} \approx A_0 + \sum_{j=1}^{p} A_j \frac{\Delta_\Gamma}{k_\varepsilon^2}\left(I + B_j \frac{\Delta_\Gamma}{k_\varepsilon^2}\right)^{-1} \tag{3.22}$$

where $A_0 = \mathrm{e}^{\mathrm{i}\theta/2} R_p(\mathrm{e}^{-\mathrm{i}\theta} - 1)$, $A_j = \mathrm{e}^{-\mathrm{i}\theta/2} a_j/(1 + b_j(\mathrm{e}^{-\mathrm{i}\theta} - 1))^2$, $B_j = \mathrm{e}^{-\mathrm{i}\theta/2} b_j/(1 + b_j(\mathrm{e}^{-\mathrm{i}\theta} - 1))$ with $R_p(z) = 1 + \sum_{j=1}^{p} a_j z/(1 + b_j z)$, $a_j = 2/(2p+1)\sin^2(j\pi/(2p+1))$ and $b_j = \cos^2(j\pi/(2p+1))$.

3.2.3. *More tractable approximations of $\mathscr{Y}_{\mathbf{J}}$ and $\mathscr{Y}_{\mathbf{M}}$ on a smooth surface*

By pursuing our work with the Helmholtz potentials, we are going to give a new approximation $\mathbf{Y}$. For that, we look at the action of $\mathbf{Y}$ (3.16) on tangential vector field $\mathbf{u} \in \mathbf{H}_T^0(\Gamma)$ by using the Helmholtz decomposition.

First, if one poses $\mathbf{J}_u = \mathbf{Y}_J \mathbf{u} \in \mathbf{H}^{-1/2}(\mathrm{div}_\Gamma, \Gamma) \cap \mathbf{H}_T^0(\Gamma)$ and $\mathbf{M}_u = \mathbf{Y}_M \mathbf{u} \in \mathbf{H}^{-1/2}(\mathrm{div}_\Gamma, \Gamma) \cap \mathbf{H}_T^0(\Gamma)$ (these regularities are suggested by Thm. 2.1), it is easy to see that (3.16) can be write in this way:

$$\begin{cases} \mathbf{u} &= \mathbf{M}_u + Z_0 \eta \mathbf{n} \times \mathbf{J}_u \\ \mathbf{M}_u &= \mathscr{Y}\, \mathbf{J}_u. \end{cases} \tag{3.23}$$

Secondly, by using the Helmholtz decompositions:

$$\begin{cases} \mathbf{u} \in \mathbf{H}_T^0(\Gamma) & \mapsto & \mathbf{u} = \mathbf{n} \times \nabla_\Gamma \psi_u + \nabla_\Gamma \phi_u & \text{with } \psi_u,\, \phi_u \in H^1(\Gamma) \\ \mathbf{M}_u \in \mathbf{H}^{-1/2}(\mathrm{div}_\Gamma, \Gamma) \cap \mathbf{H}_T^0(\Gamma) & \mapsto & \mathbf{M}_u = \mathbf{n} \times \nabla_\Gamma \psi_{\mathbf{M}_u} + \nabla_\Gamma \phi_{\mathbf{M}_u} & \text{with } (\psi_{\mathbf{M}_u},\, \phi_{\mathbf{M}_u}) \in H^1(\Gamma) \times H^{3/2}(\Gamma) \\ \mathbf{J}_u \in \mathbf{H}^{-1/2}(\mathrm{div}_\Gamma, \Gamma) \cap \mathbf{H}_T^0(\Gamma) & \mapsto & \mathbf{J}_u = \mathbf{n} \times \nabla_\Gamma \psi_{\mathbf{J}_u} + \nabla_\Gamma \phi_{\mathbf{J}_u} & \text{with } (\psi_{\mathbf{J}_u},\, \phi_{\mathbf{J}_u}) \in H^1(\Gamma) \times H^{3/2}(\Gamma) \end{cases} \tag{3.24}$$

the first equation of (3.23) leads to

$$\mathbf{n} \times \nabla_\Gamma \psi_{\mathbf{u}} + \nabla_\Gamma \phi_{\mathbf{u}} = \mathbf{n} \times \nabla_\Gamma \psi_{\mathbf{M}_u} + \nabla_\Gamma \phi_{\mathbf{M}_u} + Z_0 \eta\big(- \nabla_\Gamma \psi_{\mathbf{J}_u} + \mathbf{n} \times \nabla_\Gamma \phi_{\mathbf{J}_u}\big) \tag{3.25}$$

and by formally applying curl_Γ and div_Γ to (3.25), we obtain:

$$\begin{aligned} \psi_{\mathbf{u}} &= \psi_{\mathbf{M}_u} + Z_0 \Delta_\Gamma^{-1}(\mathrm{curl}_\Gamma(\eta \nabla_\Gamma \psi_{\mathbf{J}_u})) - Z_0 \Delta_\Gamma^{-1}(\mathrm{curl}_\Gamma(\eta \mathbf{n} \times \nabla_\Gamma \phi_{\mathbf{J}_u})) \\ \phi_{\mathbf{u}} &= \phi_{\mathbf{M}_u} - Z_0 \Delta_\Gamma^{-1}(\mathrm{div}_\Gamma(\eta \nabla_\Gamma \psi_{\mathbf{J}_u})) + Z_0 \Delta_\Gamma^{-1}(\mathrm{div}_\Gamma(\eta \mathbf{n} \times \nabla_\Gamma \phi_{\mathbf{J}_u})) \end{aligned} \tag{3.26}$$

where Δ_Γ^{-1} is the Moore-Penrose pseudo-inverse of the Laplace-Beltrami operator.

Now by using the second equation of (3.23) as well as (3.21), we get the system of two equations:

$$\begin{cases} \dfrac{\psi_{\mathbf{u}}}{Z_0} &= \left(1 + \dfrac{\Delta_\Gamma}{k_\varepsilon^2}\right)^{\frac{1}{2}} \phi_{\mathbf{J}_u} + \Delta_\Gamma^{-1}(\mathrm{rot}_\Gamma(\eta \nabla_\Gamma \psi_{\mathbf{J}_u})) - \Delta_\Gamma^{-1}(\mathrm{rot}_\Gamma(\eta \mathbf{n} \times \nabla_\Gamma \phi_{\mathbf{J}_u})) \\ -\dfrac{\phi_{\mathbf{u}}}{Z_0} &= \left(1 + \dfrac{\Delta_\Gamma}{k_\varepsilon^2}\right)^{-\frac{1}{2}} \psi_{\mathbf{J}_u} + \Delta_\Gamma^{-1}(\mathrm{div}_\Gamma(\eta \nabla_\Gamma \psi_{\mathbf{J}_u})) - \Delta_\Gamma^{-1}(\mathrm{div}_\Gamma(\eta \mathbf{n} \times \nabla_\Gamma \phi_{\mathbf{J}_u})). \end{cases} \tag{3.27}$$

We remark that if the impedance function η is constant, then these two equations are decoupled in the sense that the first equation only depends on $\phi_{\mathbf{J}_u}$ and the second one on $\psi_{\mathbf{J}_u}$ since $\mathrm{rot}_\Gamma(\eta\nabla_\Gamma\psi_{\mathbf{J}_u}) = \mathrm{div}_\Gamma(\eta\mathbf{n}\times\nabla_\Gamma\phi_{\mathbf{J}_u}) = 0$. So, by remembering of the second important remark made at the end of the section entitled "A general approach" and since η is generally chosen to be piecewise constant, we consider a new "less complicated" approximation $\mathbf{Y}_J$ which is defined by:

$$\left\{\begin{array}{rcl} \dfrac{\psi_{\mathbf{u}}}{Z_0} & = & \left(1+\dfrac{\Delta_\Gamma}{k_\varepsilon^2}\right)^{\frac{1}{2}}\phi_{\mathbf{J}_u} - \Delta_\Gamma^{-1}(\mathrm{rot}_\Gamma(\eta\mathbf{n}\times\nabla_\Gamma\phi_{\mathbf{J}_u})) \\ -\dfrac{\phi_{\mathbf{u}}}{Z_0} & = & \left(1+\dfrac{\Delta_\Gamma}{k_\varepsilon^2}\right)^{-\frac{1}{2}}\psi_{\mathbf{J}_u} + \Delta_\Gamma^{-1}(\mathrm{div}_\Gamma(\eta\nabla_\Gamma\psi_{\mathbf{J}_u})). \end{array}\right. \tag{3.28}$$

Finally, we obtain $\mathbf{M}_u$ (or $\mathbf{Y}_M$) by taking $\mathbf{M}_u = \mathbf{u} - Z_0\eta\mathbf{n}\times\mathbf{J}_u = \mathbf{u} - Z_0\eta\mathbf{n}\times\mathbf{Y}_{\mathbf{J}}\mathbf{u}$.

Remark 3.1. The approximation (3.28) does not change the initial problem. Indeed, the representation formulae (2.6) always imply that $(\mathbf{E},\mathbf{H})$ is solution of the problem (2.1)-(2.2) and the trace formulas derived of (2.11), the integral equation (3.17) and the operator $\mathbf{Y}_M$ give the boundary condition (2.3). The operator $\mathbf{Y}_J$ and consequently $\mathbf{Y}_M$ must be sufficiently close on the exact one in order to obtain an integral operator $\mathscr{Z}$ (defined in (3.17)) close to the identity operator. Numerically, these operators will particularly have an influence on the rate of convergence of the iterative solver.

3.2.4. *Case of a non-smooth surface*

At the begin of this paper, we have assumed that the domain Ω is a Lipschitz polyhedron which is simply connected and connected. Moreover, the integral framework and Theorem 2.1 has been given for this regularity assumption. Unfortunately, we are unable to currently determine an approximation $\mathbf{Y}$ of the optimal operator $\mathbf{Y}^{ex}$ in this context and that is why we have derived it for the smooth surface. Nevertheless, as we have already it said in many place, for all operator $\mathbf{Y}$, (3.7) and (3.8) define always the solution of the initial problem (2.1)-(2.2)-(2.3). Consequently, it is not absurd to consider the approximation defined in the smooth case for the non-smooth situation. It should all the same be made sure that (3.28) can be defined for the Lipschitz polyhedron. This does not pose any problem since Buffa *et al.* have proved that the space $\mathbf{H}_\times^{-1/2}(\mathrm{div}_\Gamma,\Gamma)$ admits the following Hodge decomposition [7]:

$$\mathbf{H}_\times^{-1/2}(\mathrm{div}_\Gamma,\Gamma) \quad = \quad \mathbf{n}\times\nabla_\Gamma\left(H^{1/2}(\Gamma)/\mathbb{R}\right) \oplus \nabla_\Gamma\mathscr{H}(\Gamma) \tag{3.29}$$

where $\mathscr{H}(\Gamma) = \{v\in H^1(\Gamma)/\mathbb{R} \,:\, \Delta_\Gamma\, v \in H^{-1/2}(\Gamma)\}$.

Remark 3.2. An Hodge decomposition of the space $\mathbf{H}_\times^{-1/2}(\mathrm{div}_\Gamma,\Gamma)$ can be also obtained in the multi-connected case [6].

Finally, an open question is the well-posedness of the integral equation (3.8) for the non-smooth surface. It is the same problem for the very popular Combined Field Integral Equation (CFIE) but this does not prevent its successful use in industrial contexts.

4. Well-posedness for smooth surface and impedance operator

We assume that the surface Γ is smooth and connected and that η is constant in order to use the pseudo-differential calculus. The results can be extended to the case of a non-constant smooth impedance in a straightforward way.

In this section, we first prove that our choice leads to an equation which can be written under the form: a*identity + one compact operator where a is closed on 1 *i.e.* a*identity $\approx$ identity. It is well known that this type of equation is well-adapted to an iterative resolution and that if the spectral behavior of the equation is well restored to the discrete level, then the convergence rate is independent to space and frequency refinement [9,10].

This result has already been proved in [18] for the metallic case *i.e.* $\eta = 0$ and by using the same kind of techniques, we easily prove this property for $\eta \neq 0$. Then, we will prove the integral operator (3.17) is one-to-one and consequently, a Fredholm alternative will allow us to assert the well-posed nature of the formulation.

As η is assumed to be constant, (3.28) gives the simplified system:

$$\begin{cases} \psi_{\mathbf{J}_u} = -\dfrac{1}{Z_0}\left(\eta + \left(1 + \dfrac{\Delta_\Gamma}{k_\epsilon^2}\right)^{-\frac{1}{2}}\right)^{-1} \phi_{\mathbf{u}} = Y_{\mathbf{J}}^1 \psi_{\mathbf{u}} \\ \phi_{\mathbf{J}_u} = \dfrac{1}{Z_0}\left(\eta + \left(1 + \dfrac{\Delta_\Gamma}{k_\epsilon^2}\right)^{\frac{1}{2}}\right)^{-1} \psi_{\mathbf{u}} = Y_{\mathbf{J}}^2 \psi_{\mathbf{u}} \end{cases} \tag{4.1}$$

and

$$\begin{cases} \psi_{\mathbf{M}_u} = \left(1 + \eta\left(1 + \dfrac{\Delta_\Gamma}{k_\varepsilon^2}\right)^{-\frac{1}{2}}\right)^{-1} \psi_{\mathbf{u}} = Y_{\mathbf{M}}^1 \psi_{\mathbf{u}} \\ \phi_{\mathbf{M}_u} = \left(1 + \eta\left(1 + \dfrac{\Delta_\Gamma}{k_\varepsilon^2}\right)^{\frac{1}{2}}\right)^{-1} \phi_{\mathbf{u}} = Y_{\mathbf{M}}^2 \phi_{\mathbf{u}}. \end{cases} \tag{4.2}$$

Proposition 4.1. *We have*

$$\begin{aligned} Y_{\mathbf{M}}^1 &\sim \mathrm{Id} \bmod \Psi^{-1}(\Gamma) & Y_{\mathbf{J}}^1 &\sim -\frac{1}{Z_0\eta}\mathrm{Id} \bmod \Psi^{-1}(\Gamma) \\ Y_{\mathbf{M}}^2 &\sim \ \mathrm{op}\left(\frac{k_\varepsilon}{i\eta|\theta|}\right) \bmod \Psi^{-2}(\Gamma) & Y_{\mathbf{J}}^2 &\sim \ \mathrm{op}\left(\frac{k_\varepsilon}{Z_0 i|\theta|}\right) \bmod \Psi^{-2}(\Gamma) \end{aligned} \tag{4.3}$$

where $\Psi^s(\Gamma)$ *denotes the class of pseudo-differential operators of order* s *defined on the closed surface* Γ, $\mathrm{op}(\sigma)$ *defines the pseudo-differential operator whose the symbol is* σ *and* $A \sim B \bmod \Psi^{-m}(\Gamma)$ *with* $m \in \mathbb{Z}$ *means* $\exists C \in \Psi^{-m}(\Gamma)$ *such that* $A - (B + C) \in \Psi^{-\infty}(\Gamma)$.

Proof. All these results are obtained in the same way. We only detail the calculus for $Y_{\mathbf{M}}^2$:

$$\begin{aligned} \sigma(Y_{\mathbf{M}}^2) &= \frac{1}{1 + \eta\sqrt{1 - \frac{|\theta|^2}{k_\varepsilon^2}}} = \frac{1}{1 + i\eta\frac{|\theta|}{k_\varepsilon}\sqrt{1 - \frac{k_\varepsilon^2}{|\theta|^2}}} = \frac{1}{1 + i\eta\frac{|\theta|}{k_\varepsilon}\left(1 + \sum_{j=1}^{+\infty}\frac{\lambda_j}{|\theta|^{2j}}\right)} \\ &= \frac{1}{i\eta\frac{|\theta|}{k_\varepsilon}\left(\frac{k_\varepsilon}{i\eta|\theta|} + 1 + \sum_{j=1}^{+\infty}\frac{\lambda_j}{|\theta|^{2j}}\right)} = \frac{k_\varepsilon}{i\eta|\theta|}\left(1 + \sum_{j=1}^{+\infty}(-1)^j X^j\right) \end{aligned}$$

where $X = \dfrac{k_\varepsilon}{i\eta|\theta|} + \displaystyle\sum_{j=1}^{+\infty}\frac{\lambda_j}{|\theta|^{2j}}$.

This expansion is valid in the elliptic zone and for $|\theta| \to +\infty$. □

Proposition 4.2. *We have*

$$\begin{aligned} \mathbf{n} \times T\mathscr{Y}_{\mathbf{J}} &\sim \frac{k_\varepsilon}{2iZ_0 k}\mathbf{n} \times \nabla_\Gamma \Delta_\Gamma^{-1} \ \mathrm{curl}_\Gamma \bmod \Psi^{-1}(\Gamma) \\ T\mathscr{Y}_{\mathbf{M}} &\sim -\frac{k_\varepsilon}{2i\eta k}\nabla_\Gamma \Delta_\Gamma^{-1} \ \mathrm{div}_\Gamma \bmod \Psi^{-1}(\Gamma). \end{aligned} \tag{4.4}$$

Proof. We only prove the first result. The second is obtained in the same way. First, we write T under this straightforward form $T = kG + \frac{1}{k}\nabla_\Gamma G_s \mathrm{div}_\Gamma$ and $\mathbf{J}_u, \mathbf{M}_u$ in this way

$$\begin{aligned}\mathbf{J}_u &= \big(\mathbf{n}\times\nabla_\Gamma Y^1_{\mathbf{J}}\Delta_\Gamma^{-1}\mathrm{div}_\Gamma - \nabla_\Gamma Y^2_{\mathbf{J}}\Delta_\Gamma^{-1}\mathrm{curl}_\Gamma\big)\mathbf{u}\\ \mathbf{M}_u &= \big(-\mathbf{n}\times\nabla_\Gamma Y^1_{\mathbf{M}}\Delta_\Gamma^{-1}\mathrm{curl}_\Gamma + \nabla_\Gamma Y^2_{\mathbf{M}}\Delta_\Gamma^{-1}\mathrm{div}_\Gamma\big)\mathbf{u}.\end{aligned} \tag{4.5}$$

We then get:

$$\begin{aligned}\mathbf{n}\times T\mathbf{J}_u &= \big(k\mathbf{n}\times G + \frac{1}{k}\mathbf{n}\times\nabla_\Gamma G_s\mathrm{div}_\Gamma\big)\big(\mathbf{n}\times\nabla_\Gamma Y^1_{\mathbf{J}}\Delta_\Gamma^{-1}\mathrm{div}_\Gamma - \nabla_\Gamma Y^2_{\mathbf{J}}\Delta_\Gamma^{-1}\ \mathrm{curl}_\Gamma\big)\mathbf{u}\\ &= k\mathbf{n}\times G\mathbf{n}\times\nabla_\Gamma Y^1_{\mathbf{J}}\Delta_\Gamma^{-1}\mathrm{div}_\Gamma\mathbf{u}\\ &\quad - k\mathbf{n}\times G\nabla_\Gamma Y^2_{\mathbf{J}}\Delta_\Gamma^{-1}\ \mathrm{curl}_\Gamma\mathbf{u} + \frac{1}{k}\mathbf{n}\times\nabla_\Gamma G_s\Delta_\Gamma Y^2_{\mathbf{J}}\Delta_\Gamma^{-1}\ \mathrm{curl}_\Gamma\mathbf{u}.\end{aligned}$$

Finally, by using these classical properties: $G_s \sim -\frac{1}{2|\theta|} \bmod \Psi^{-2}(\Gamma)$, $G \sim -\frac{1}{2|\theta|}\mathrm{Id} \bmod \Psi^{-2}(\Gamma)$ and $\Delta_\Gamma^{-1} \sim -4G_s^2 \bmod \Psi^{-3}(\Gamma)$, we immediately obtain:

$$\begin{aligned}\mathbf{n}\times T\mathscr{Y}_{\mathbf{J}} &\sim \frac{1}{k}\mathbf{n}\times\nabla_\Gamma G_s\Delta_\Gamma Y^2_{\mathbf{J}}\Delta_\Gamma^{-1}\ \mathrm{curl}_\Gamma \bmod \Psi^{-1}(\Gamma)\\ &\sim -\frac{2k_\varepsilon}{iZ_0k}\mathbf{n}\times\nabla_\Gamma G_s\Delta_\Gamma G_s\Delta_\Gamma^{-1}\ \mathrm{curl}_\Gamma \bmod \Psi^{-1}(\Gamma)\\ &\sim -\frac{2k_\varepsilon}{iZ_0k}\mathbf{n}\times\nabla_\Gamma G_s^2\ \mathrm{curl}_\Gamma \bmod \Psi^{-1}(\Gamma)\\ &\sim \frac{k_\varepsilon}{2iZ_0k}\mathbf{n}\times\nabla_\Gamma\Delta_\Gamma^{-1}\ \mathrm{curl}_\Gamma \bmod \Psi^{-1}(\Gamma).\end{aligned}$$

□

Introducing (4.4) in (3.17) and as K is pseudo-differential operator of order -3 [27], we obtain:

$$\begin{aligned}\mathscr{Z} &\sim \tfrac{1}{2}\mathrm{Id} - \tfrac{k_\varepsilon}{2k}\mathbf{n}\times\nabla_\Gamma\Delta_\Gamma^{-1}\ \mathrm{curl}_\Gamma + \tfrac{k_\varepsilon}{2k}\nabla_\Gamma\Delta_\Gamma^{-1}\mathrm{div}_\Gamma \bmod \Psi^{-1}(\Gamma)\\ &\sim \left(\tfrac{1}{2} + \tfrac{k_\varepsilon}{2k}\right)\mathrm{Id} \bmod \Psi^{-1}(\Gamma)\end{aligned}$$

where $a = \frac{1}{2} + \frac{k_\varepsilon}{2k} = 1 + \frac{\varepsilon}{2k}$.

We still have to prove the injectivity of the operator in order to use a Fredholm alternative. For that, we again follow [18] by using the spectral decomposition of Laplace-Beltrami operator Δ_Γ on Γ: let $(Y_i)_i$ be a basis of eigenvectors of Δ_Γ, then we have $-\Delta_\Gamma Y_i = \lambda_i Y_i$ with $\lambda_i \geq 0$. Moreover $(Y_i)_i$ and $\left(\frac{\nabla_\Gamma Y_i}{\sqrt{\lambda_i}}, \frac{\mathbf{n}\times\nabla_\Gamma Y_i}{\sqrt{\lambda_i}}\right)_{i\geq 1}$ are orthogonal Hilbertian basis of $L^2(\Gamma)$ and $L^2_T(\Gamma)$ respectively.

Proposition 4.3. *If the operator* $\mathbf{Y} = (\mathbf{Y}_M, \mathbf{Y}_J)$ *verifies the condition*

$$\Re\left(\int_\Gamma \mathbf{n}\times\mathbf{Y}_J\mathbf{u}\cdot\overline{\mathbf{Y}_M\mathbf{u}}\,\mathrm{d}\Gamma\right) > 0 \text{ for all } \mathbf{u}\in\mathbf{H}^0_T(\Gamma) \text{ and } \mathbf{u}\neq 0 \tag{4.6}$$

then the equation (3.17) *is one-to-one.*

Proof. First, if one takes a zero incident field, then one obviously has $\mathbf{E}_{|\Omega^+} = \mathbf{H}_{|\Omega^+} = 0$. Since $\mathbf{n}\times(\mathbf{E}_\Gamma^-\times\mathbf{n}) = -1/2\mathbf{n}\times\mathbf{Y}_M\mathbf{u} + iZ_0T\mathbf{Y}_J\mathbf{u} + K\mathbf{Y}_M\mathbf{u}$ ($-$ denotes the interior trace), we immediately get $(\mathbf{E}_\Gamma^-\times\mathbf{n}) = -\mathbf{Y}_M\mathbf{u}$. In the same way, we have $\mathbf{n}\times(\mathbf{H}_\Gamma^-\times\mathbf{n}) = 1/(ikZ_0)(\mathbf{n}\times(\mathrm{rot}(\mathbf{E}^-)_\Gamma\times\mathbf{n})) = \mathbf{n}\times\mathbf{Y}_J\mathbf{u}$.

By using the Green formula on Ω^-, we can derive $\|\mathrm{rot}\mathbf{E}^-\|^2_{0,\Omega^-} - k^2\|\mathbf{E}^-\|^2_{0,\Omega^-} = \displaystyle\int_\Gamma(\mathbf{n}\times(\mathrm{rot}(\mathbf{E}^-)_\Gamma\times\mathbf{n}))\cdot\overline{(\mathbf{E}^-\times\mathbf{n})}\mathrm{d}\Gamma = -(ikZ_0)\int_\Gamma(\mathbf{n}\times\mathbf{Y}_J\mathbf{u})\cdot\overline{\mathbf{Y}}_M\mathbf{u}\mathrm{d}\Gamma$. So, since k is a real number, we obtain $\Re\left(\displaystyle\int_\Gamma(\mathbf{n}\times\mathscr{Y}_\mathbf{J}\mathbf{u})\cdot\overline{\mathscr{Y}_\mathbf{M}\mathbf{u}}\mathrm{d}\Gamma\right) = 0$ and (4.6) leads to $\mathbf{u} = 0$. □

Now, we prove (4.6) for our choice of operator $\mathbf{Y}$.

Proposition 4.4. *If one chooses the operator* $\mathbf{Y}$ *defined in Section 3.2.3, then the condition*

$$\Re\left(\int_\Gamma \mathbf{n}\times\mathbf{Y}_J\mathbf{u}\cdot\overline{\mathbf{Y}_M\mathbf{u}}\,\mathrm{d}\Gamma\right) > 0 \text{ for all } \mathbf{u}\in\mathbf{H}^0_T(\Gamma) \text{ and } \mathbf{u}\neq 0 \tag{4.7}$$

holds.

Proof. Let $\mathbf{u}\in\mathbf{H}^0_T(\Gamma)$ such that $\mathbf{u}\neq 0$. One can decompose $\mathbf{u}$ in this way: $\mathbf{u} = \displaystyle\sum_{i=1}^\infty\left(\alpha_i\frac{\mathbf{n}\times\nabla_\Gamma Y_i}{\sqrt{\lambda_i}} + \beta_i\frac{\nabla_\Gamma Y_i}{\sqrt{\lambda_i}}\right)$ where α_i,β_i are complex numbers. In particular, one can take $\psi_\mathbf{u} = \displaystyle\sum_{i=1}^\infty\alpha_i\frac{Y_i}{\sqrt{\lambda_i}}$ and $\phi_\mathbf{u} = \displaystyle\sum_{i=1}^\infty\beta_i\frac{Y_i}{\sqrt{\lambda_i}}$. (4.1) and (4.2) give us:

$$\begin{aligned}
\psi_{\mathbf{J}_u} &= -\frac{1}{Z_0}\sum_{i=1}^\infty\beta_i\left(\eta+\left(1-\frac{\lambda_i}{k_\varepsilon^2}\right)^{-\frac12}\right)^{-1}\frac{Y_i}{\sqrt{\lambda_i}} = -\frac{1}{Z_0}\sum_{i=1}^\infty\beta_i(\psi_{\mathbf{J}_u})_i\frac{Y_i}{\sqrt{\lambda_i}}\\
\phi_{\mathbf{J}_u} &= \frac{1}{Z_0}\sum_{i=1}^\infty\alpha_i\left(\eta+\left(1-\frac{\lambda_i}{k_\varepsilon^2}\right)^{\frac12}\right)^{-1}\frac{Y_i}{\sqrt{\lambda_i}} = \frac{1}{Z_0}\sum_{i=1}^\infty\alpha_i(\phi_{\mathbf{J}_u})_i\frac{Y_i}{\sqrt{\lambda_i}}\\
\psi_{\mathbf{M}_u} &= \sum_{i=1}^\infty\alpha_i\left(1+\eta\left(1-\frac{\lambda_i}{k_\varepsilon^2}\right)^{-\frac12}\right)^{-1}\frac{Y_i}{\sqrt{\lambda_i}} = \sum_{i=1}^\infty\alpha_i(\psi_{\mathbf{M}_u})_i\frac{Y_i}{\sqrt{\lambda_i}}\\
\phi_{\mathbf{M}_u} &= \sum_{i=1}^\infty\beta_i\left(1+\eta\left(1-\frac{\lambda_i}{k_\varepsilon^2}\right)^{\frac12}\right)^{-1}\frac{Y_i}{\sqrt{\lambda_i}} = \sum_{i=1}^\infty\beta_i(\phi_{\mathbf{M}_u})_i\frac{Y_i}{\sqrt{\lambda_i}}
\end{aligned}\tag{4.8}$$

(4.8) leads to $\displaystyle\int_\Gamma(\mathbf{n}\times\mathbf{Y}_J\mathbf{u})\cdot\mathbf{Y}_M\mathbf{u}\mathrm{d}\Gamma = 1/Z_0\sum_{i=1}^\infty\left(|\beta_i|^2(\psi_{\mathbf{J}_u})_i\overline{(\phi_{\mathbf{M}_u})_i} + |\alpha_i|^2(\phi_{\mathbf{J}_u})_i\overline{(\psi_{\mathbf{M}_u})_i}\right)$. Since $(\psi_{\mathbf{J}_u})_i\overline{(\phi_{\mathbf{M}_u})_i} = \left(1-\frac{\lambda_i}{k_\varepsilon^2}\right)^{\frac12}\left|\left(1+\eta\left(1-\frac{\lambda_i}{k_\varepsilon^2}\right)^{\frac12}\right)\right|^{-2}$, $(\phi_{\mathbf{J}_u})_i\overline{(\psi_{\mathbf{M}_u})_i} = \left(1-\frac{\lambda_i}{k_\varepsilon^2}\right)^{-\frac12}\left|\left(1+\eta\left(1-\frac{\lambda_i}{k_\varepsilon^2}\right)^{-\frac12}\right)\right|^{-2}$, $\eta>0$ (see assumption Thm. 2.1) and $\Re\left(1-\frac{\lambda_i}{k_\varepsilon^2}\right)^{\pm\frac12} > 0$ (for $k>0$ and $\varepsilon>0$), we finally get the result. □

Finally, we have:

Theorem 4.5. *If one chooses the operator* $\mathbf{Y}$ *defined in Section 3.2.3, then the integral equation (3.17) is well-posed for any frequency.*

Proof. We use a basic Fredholm argument. □

5. Discretization and resolution

After discretization, we want to iteratively solve (3.17). For that, we have to be able to form the matrix-vector products defined by the operator $\mathscr{Z}$ and required by the iterative solver. But, the choice made for $\mathbf{Y}_J$ and $\mathbf{Y}_M$ does not allow us to directly discretize the operators $\mathrm{T}\circ\mathbf{Y}_{OP}$ and $\mathrm{K}\circ\mathbf{Y}_{OP}$ where $OP=\mathbf{Y}_J$ or $\mathbf{Y}_M$ since one do not know the Schwarz kernel of these compositions. That is why we proceed in several steps: one knows $\mathbf{u}\in\mathrm{RT}$ (RT is well-known lower-order Raviart-Thomas space [28]) at the iteration n and one wants to compute $\mathscr{Z}\mathbf{u}$.

- First, we determine two potentials $(\phi_{\mathbf{u}},\psi_{\mathbf{u}})$ of $\mathbf{u}$ such that $\mathbf{u}\approx\mathbf{n}\times\nabla_\Gamma\psi_{\mathbf{u}}+\nabla_\Gamma\phi_{\mathbf{u}}$.
- We compute $(\phi_{\mathbf{J}_u},\psi_{\mathbf{J}_u})$ by solving (3.28) and we derive $\mathbf{J}_u\in\mathrm{RT}\approx\mathbf{n}\times\nabla_\Gamma\psi_{\mathbf{J}_u}+\nabla_\Gamma\phi_{\mathbf{J}_u}$.
- We compute $\mathbf{M}_u=\mathbf{u}-Z_0\eta\mathbf{n}\times\mathbf{J}_u\in\mathrm{RT}$.
- Finally, we compute the "classical" matrix-vector products coming from integral operators T and K.

In this section, we explain how we treat these four steps.

5.1. First step

First, recall some properties related to the Raviart-Thomas space [11]. Let W_h be the subspace of RT defined by $W_h:=\{\mathbf{v}\in\mathrm{RT}\ :\ \nabla_\Gamma\cdot\mathbf{v}=0\}$. There is $C>0$ such that $\forall\mathbf{v}\in\mathrm{RT}$ verifying $\forall\mathbf{w}\in W_h\ (\mathbf{v},\mathbf{w})=0$, the solution p of $p\in H^1(\Gamma)^\bullet$ and $\Delta_\Gamma p=\nabla_\Gamma\cdot\mathbf{v}$ satisfies $|\mathbf{v}-\nabla_\Gamma p|_{0,\Gamma}\le Ch|\nabla_\Gamma\cdot\mathbf{v}|_{0,\Gamma}$. Moreover, each $\mathbf{v}\in\mathrm{RT}$ can be written in a unique way $\mathbf{v}=\mathbf{n}\times\nabla_\Gamma\psi+\mathbf{v}_\perp$, with $\psi\in\overset{o}{P}_1$ ($\overset{o}{P}_1$ denotes the space of scalar functions, piecewise linear on each triangle of the mesh and whose mean value is null on Γ) and $\mathbf{v}_\perp$ in the L^2-orthogonal of the kernel of the divergence operator in RT.

We begin by the computation of ψ: for that it is sufficient to project $\mathbf{v}$ onto the space defined by:

$$V_{\mathrm{loop}} \quad=\quad \mathrm{span}\{\mathbf{n}\times\nabla_\Gamma\phi_i\ :\ i=1,\dots,nbnoe-1\} \tag{5.1}$$

where ϕ_i is the Lagrange basis function associated to the node i of the mesh and $nbnoe$ denotes the number of vertices. More precisely, we find $\mathbf{v}_{\mathrm{loop}}=\sum_{i=1}^{nbnoe-1}\mathbf{v}^i_{\mathrm{loop}}\mathbf{n}\times\nabla_\Gamma\phi_i$ such that $\forall\mathbf{w}\in V_{\mathrm{loop}}$,

$$\int_\Gamma\mathbf{v}_{\mathrm{loop}}\cdot\mathbf{w}\mathrm{d}\Gamma \quad=\quad \int_\Gamma\mathbf{v}\cdot\mathbf{w}\mathrm{d}\Gamma. \tag{5.2}$$

Finally, we simply have $\psi=\sum_{i=1}^{nbnoe-1}\mathbf{v}^i_{\mathrm{loop}}\phi_i$ where ψ is obviously determined modulo one constant. We can remark that this computation only implies the inversion of a small sparse linear system whose $((nbnoe-1)\times(nbnoe-1))$.

Now we have to determine a potential $\phi\in\overset{o}{P}_1$ such that $\mathbf{v}_\perp\approx\nabla_\Gamma\phi$. For that, we suggest to solve: find $\phi\in\overset{o}{P}_1$ such that $\forall\phi'\in\overset{o}{P}_1$

$$-\int_\Gamma\nabla_\Gamma\phi\cdot\nabla_\Gamma\phi'\mathrm{d}\Gamma \quad=\quad \int_\Gamma\nabla_\Gamma\cdot\mathbf{v}\phi'\mathrm{d}\Gamma \quad=\quad \int_\Gamma\nabla_\Gamma\cdot\mathbf{v}_\perp\phi'\mathrm{d}\Gamma. \tag{5.3}$$

As $\forall\mathbf{w}\in W_h\ (\mathbf{v}_\perp,\mathbf{w})_{0,\Gamma}=0$, ϕ solution of (5.3) is a finite element approximation of the problem: find $p\in H^1(\Gamma)^\bullet$ such that $\Delta_\Gamma p=\nabla_\Gamma\cdot\mathbf{v}=\nabla_\Gamma\cdot\mathbf{v}_\perp$. One easily deduces from it that the *a priori* error estimate $|\mathbf{v}_\perp-\nabla_\Gamma\phi|_{0,\Gamma}\le Ch|\nabla_\Gamma\cdot\mathbf{v}|_{0,\Gamma}$ holds.

In conclusion, the determination of (ψ,ϕ) requires the resolution of two sparse systems. We also use the MUltifrontal Massively Parallel sparse direct Solver (MUMPS) [26] to solve them. MUMPS provides a quick and low cost inversion.

5.2. **Second step**

The system (3.28) can be written as two discrete weak formulations:
find $\left(\phi_{\mathbf{J}_u}, V_j, U\right) \in [P_1(\mathscr{T}_h)]^2 \times \mathring{P}_1(\mathscr{T}_h)$ such that for all $\left(\phi', V_j', U'\right) \in [P_1(\mathscr{T}_h)]^2 \times \mathring{P}_1(\mathscr{T}_h)$

$$\left\{\begin{array}{rcl} A_0(\phi_{\mathbf{J}_u}, \phi') - \displaystyle\sum_{j=1}^{p} \frac{A_j}{k_\varepsilon^2}(\nabla_\Gamma V_j, \nabla_\Gamma \phi') + (U, \phi') & = & \dfrac{1}{Z_0}(\psi_{\mathbf{u}}, \phi') \\ (V_j, V_j') - \dfrac{B_j}{k_\varepsilon^2}(\nabla_\Gamma V_j, \nabla_\Gamma V_j') - (\phi_{\mathbf{J}_u}, V_j') & = & 0 \text{ for all } j = 1, \ldots, p \\ (\nabla_\Gamma U, \nabla_\Gamma U') + (\eta \mathbf{n} \times \nabla_\Gamma \phi_{\mathbf{J}_u}, \mathbf{n} \times \nabla_\Gamma U') & = & 0 \end{array}\right. \tag{5.4}$$

and find $\left(\psi_{\mathbf{J}_u}, V_j, U\right) \in [P_1(\mathscr{T}_h)]^2 \times \mathring{P}_1(\mathscr{T}_h)$ such that for all $\left(\psi', V_j', U'\right) \in [P_1(\mathscr{T}_h)]^2 \times \mathring{P}_1(\mathscr{T}_h)$

$$\left\{\begin{array}{rcl} (\psi_{\mathbf{J}_u}, \psi') + A_0(U, \psi') - \displaystyle\sum_{j=1}^{p} \frac{A_j}{k_\varepsilon^2}(\nabla_\Gamma V_j, \nabla_\Gamma \psi') & = & -\dfrac{1}{Z_0}\left(\left(1 + \dfrac{\Delta_\Gamma}{k_\varepsilon^2}\right)^{\frac{1}{2}} \phi_{\mathbf{u}}, \psi'\right) \\ (V_j, V_j') - \dfrac{B_j}{k_\varepsilon^2}(\nabla_\Gamma V_j, \nabla_\Gamma V_j') - (U, V_j') & = & 0 \text{ for all } j = 1, \ldots, p \\ -(\nabla_\Gamma U, \nabla_\Gamma U') + (\eta \nabla_\Gamma \psi_{\mathbf{J}_u}, \nabla_\Gamma U') & = & 0 \end{array}\right. \tag{5.5}$$

where $(\cdot, \cdot)$ denotes the classical L^2-scalar product.

Remark 5.1. In (5.5), we also must determine the square root term $\left(\left(1 + \dfrac{\Delta_\Gamma}{k^2}\right)^{\frac{1}{2}} \phi_{\mathbf{u}}, \psi'\right)$. For that, we proceed in this way:

$$\left(\left(1 + \frac{\Delta_\Gamma}{k_\varepsilon^2}\right)^{\frac{1}{2}} \phi_{\mathbf{u}}, \psi'\right) \approx A_0(\phi_{\mathbf{u}}, \psi') - \sum_{j=1}^{p} \frac{A_j}{k_\varepsilon^2}(\nabla_\Gamma V_j, \nabla_\Gamma \psi') \tag{5.6}$$

where $V_j \in P_1(\mathscr{T}_h)$ $(j = 1, \ldots, p)$ is solution of the discrete weak formulation $(V_j, V_j') - \dfrac{B_j}{k_\varepsilon^2}(\nabla_\Gamma V_j, \nabla_\Gamma V_j') = (\phi_{\mathbf{u}}, V_j')$ for all $V_j' \in P_1(\mathscr{T}_h)$.

All these sparse linear systems are solved for by using MUMPS. The size of these systems is about $(2+p)ntri/2$ (with $ntri$ is the number of triangle of the mesh and we have used the approximation $nbnoe \approx ntri/2$). In the next section, we show that $p = 2$ is a relevant choice. In this case, the size becomes $2ntri$. It is not a big system in the case of surface meshes and we have remarked [14] that the cost of the resolution of these kind of system by MUMPS is always negligible compared to the one of the resolution of the integral equation (dense system), even when we use a Fast Multipole Method.

Finally, $\mathbf{J}_u \in \mathrm{RT}$ is obtained by a simple projection of $\mathbf{n} \times \nabla_\Gamma \psi_{\mathbf{J}_u} + \nabla_\Gamma \phi_{\mathbf{J}_u}$ on RT.

5.3. **Third step**

The determination of $\mathbf{M}_u$ is less obvious than one could believe. Indeed, the straightforward way to derive $\mathbf{M}_u \in \mathrm{RT}$ is to do a simple projection of $\mathbf{M}_u = \mathbf{u} - Z_0 \eta \mathbf{n} \times \mathbf{J}_u$ on RT. But, it seems that the property proved in the previous section (*i.e.* $\mathscr{Z} \sim a\mathrm{Id} + \mathrm{Compact}$) is not "sufficiently" preserved after the discretization. To avoid this problem, it is more judicious to compute $\mathbf{M}_u$ by using $\mathbf{n} \times \nabla_\Gamma \psi_{\mathbf{u}} + \nabla_\Gamma \phi_{\mathbf{u}}$. Indeed, if η is constant, one can see that $\mathbf{M}_u \approx \mathbf{n} \times \nabla_\Gamma(\psi_{\mathbf{u}} - Z_0 \eta \phi_{\mathbf{J}_u}) + \nabla_\Gamma(\phi_{\mathbf{u}} + Z_0 \eta \psi_{\mathbf{J}_u})$ and $\mathbf{Y}_M$ is better closed on the continuous operator $\mathbf{Y}_M^{ex}$ than when we directly use $\mathbf{u}$.

This choice has a direct impact on the solutions. Indeed, let us look how the impedance condition is verified to the discrete level: $\forall \mathbf{w}_h \in \mathrm{RT}$, we have

$$\begin{aligned}(\mathbf{E}^h \times \mathbf{n} + Z_0\eta\mathbf{n} \times (\mathbf{H}^h \times \mathbf{n}),\, \mathbf{w}_h) &= \left(\mathbf{g} + \frac{1}{2}(-\mathbf{u} + \mathbf{M}_u + \eta Z_0 \mathbf{n} \times \mathbf{J}_u),\, \mathbf{w}_h\right) \\ &= \left(\mathbf{g} + \frac{1}{2}(-\mathbf{u} + \mathbf{n} \times \nabla_\Gamma \psi_\mathbf{u} + \nabla_\Gamma \phi_\mathbf{u} - Z_0\eta\mathbf{n} \times \mathbf{J}_u + Z_0\eta\mathbf{n} \times \mathbf{J}_u),\, \mathbf{w}_h\right) \\ &= \left(\mathbf{g} + \frac{1}{2}(-\mathbf{n} \times \nabla_\Gamma \psi_\mathbf{u} - \mathbf{u}_\perp + \mathbf{n} \times \nabla_\Gamma \psi_\mathbf{u} + \nabla_\Gamma \phi_\mathbf{u}),\, \mathbf{w}_h\right) \\ &= \left(\mathbf{g} + \frac{1}{2}(\nabla_\Gamma \phi_\mathbf{u} - \mathbf{u}_\perp),\, \mathbf{w}_h\right).\end{aligned} \tag{5.7}$$

We can make some remarks:

- For the continuous case, the space $\mathbf{H}_\times^{-1/2}(\mathrm{div}_\Gamma, \Gamma)$ admits a hodge decomposition and we have $\mathbf{u}_\perp = \nabla_\Gamma \phi_\mathbf{u}$.
- For the discrete case, if one choose $\mathbf{M}_u = \mathbf{u} - Z_0\eta\mathbf{n} \times \mathbf{J}_u$, then the boundary condition is verified in this classical sense:

$$(\mathbf{E}^h \times \mathbf{n} + Z_0\eta\mathbf{n} \times (\mathbf{H}^h \times \mathbf{n})\,,\, \mathbf{w}_h) = (\mathbf{g}\,,\, \mathbf{w}_h), \quad \forall \mathbf{w}_h \in \mathrm{RT} \tag{5.8}$$

whereas by using our choice $\mathbf{M}_u = \mathbf{n} \times \nabla_\Gamma \psi_\mathbf{u} + \nabla_\Gamma \phi_\mathbf{u} - Z_0\eta\mathbf{n} \times \mathbf{J}_u$, (5.7) leads to

$$\sup_{\mathbf{w}_h \in \mathrm{RT}} \frac{(\mathbf{E}_h \times \mathbf{n} + Z_0\eta\mathbf{n} \times (\mathbf{H}_h \times \mathbf{n}) - \mathbf{g}, \mathbf{w}_h)_{0,\Gamma}}{\|\mathbf{w}_h\|_{0,\Gamma}} \leq \frac{1}{2}\|\mathbf{u}_\perp - \nabla_\Gamma \phi_\mathbf{u}\|_{0,\Gamma} \leq Ch|\nabla_\Gamma \cdot \mathbf{u}|_{0,\Gamma} \tag{5.9}$$

since the Raviart-Thomas space does not subject to an exact Helmholtz decomposition. Nevertheless, the numerical results will show that this perturbation is under control.

5.4. **Fourth step**

The discretization of the operators T and K by the Raviart-Thomas finite elements is well-known and that is why we do not detail this here. What is less obvious is the computation of $\mathbf{n} \times \mathrm{T}$. We have already treated this difficulty in [14] when we wanted to construct a Combined Field Integral Equation for the impedance problems. The technique is based on a discrete analogue of the Helmholtz decomposition and has been proposed by Christiansen and Nédélec in [12]. For the completeness of the paper, we briefly describe it in this paper:

Let l_h be a linear form defined on RT; the problem is to construct some $\Theta^h l_h$ that mimics $\mathbf{n} \times l_h$. In what follows, P_0^o denotes the space of scalar functions whose restriction to any triangle is constant and whose mean value is null on Γ and P_1^o denotes the space of scalar functions, piecewise linear on each triangle of the mesh and whose mean value is null on Γ. We consider the following saddle-point problem:

$$\begin{gathered}\textit{Find } (u^h, q^h) \in \mathrm{RT} \times P_0^o \textit{ such that for all } (u', q') \in \mathrm{RT} \times P_0^o, \\ \begin{cases} \displaystyle\int_\Gamma u^h \cdot u' \mathrm{d}\Gamma + \int_\Gamma q^h \mathrm{div} u' \mathrm{d}\Gamma &= l^h(u') \\ \displaystyle\int_\Gamma q' \mathrm{div} u^h \mathrm{d}\Gamma &= 0. \end{cases}\end{gathered} \tag{5.10}$$

Once this system is solved, we associate to (u^h, q^h) the following element of RT:

$$v^h = \Theta^h l^h = \mathscr{P}_{\mathrm{RT}}(u^h \times \mathbf{n}) - \mathrm{curl}\mathscr{P}_{P_1^o}(q^h), \tag{5.11}$$

where $\mathscr{P}_X$ denotes the L^2-projection on the finite element space X.

The map Θ^h is analyzed in [12] and we have improved its determination by using the Loop-Star basis functions [20].

Now, we are going to explain how to reduce the cost of the (dense) matrix-vector products. There are normally four vector products to form the two terms: $T1 = K\mathbf{M}_u + iZ_0T\mathbf{J}_u$ and $T2 = iZ_0K\mathbf{J}_u + T\mathbf{M}_u$. We remark that if one computes the two quantities $Q1 = (K+T)(\mathbf{M}_u + iZ_0\mathbf{J}_u)$ and $Q2 = (K-T)(\mathbf{M}_u - iZ_0\mathbf{J}_u)$, we simply have $T1 = (Q1+Q2)/2$ and $T2 = (Q1-Q2)/2$ *i.e.* only two matrix-vector products. Moreover [14] proposed an algorithm which only uses two fast multipole calculations to determine $K\mathbf{u}+T\mathbf{v}$ and $K\mathbf{v}+T\mathbf{u}$.

The last steps are:

- Calculation of $\Theta^h(K\mathbf{M}_u + iZ_0T\mathbf{J}_u) \approx \mathbf{n}\times(K\mathbf{M}_u + iZ_0T\mathbf{J}_u)$: [14] proved that this step is negligible compared to the (dense) matrix-vector products even when FMM is used.
- Calculation of $i\eta[iZ_0K\mathbf{J}_u + T\mathbf{M}_u]$: first we compute $\mathbf{v} \in \mathrm{RT} \approx [iZ_0K\mathbf{J}_u + T\mathbf{M}_u]$ by using a sparse inversion of the RT mass matrix (*i.e.* we make a simple projection of this term on the RT finite elements space) and then we form the sparse matrix-vector product $(i\eta\mathbf{v}, \mathbf{v}')$ with $\mathbf{v}' \in \mathrm{RT}$.

6. Numerical results

The aim of this part is to numerically prove the good iterative behavior of the method. We will also show that the solutions are accurate. Each result will be compared to the one obtained by using a Impedance Combined Field Equation (ICFIE) that we have proposed in [14]. We have proved that ICFIE is a relevant method to treat constant and variable impedance problems. The goal of this first study is to show that it is possible to construct a cheap naturally well-conditioned formulation for the partially coated object and to give a first numerical issue. The GMRES solver [19] is used to solve for the linear system. In the following, Γ_m and Γ_i correspond respectively to parts of Γ where the impedance operator is equal or not to zero.

We consider four scattering objects illuminated by a plane wave with a wavenumber k:

- A sphere (smooth surface) of diameter 1 m. The meshes are composed of 1500, 6000 and 13 500 edges which respectively correspond to wavenumbers k equal to 4.83, 11 and 16.4. The discretization complies with the criterion of around 10 points per wavelength. We have simulated two situations: the first is a constant impedance $\eta = 0.34$ *i.e.* $\Gamma_m = \emptyset$ and the second is defined by $\eta = 0.34$ on $\Gamma_i = \{(x,y,z) : x^2+y^2+z^2 = 1 \text{ and } z > 0\}$.
- A cube (non-smooth surface). The meshes are composed of 2178, 4860, 8460 and 13 374 edges which respectively correspond to wavenumbers k equal to 10, 15, 20 and 25. We have also considered two situations: the first is a constant impedance $\eta = 0.34$ and for the second $\eta = 1$ on Γ_i which corresponds to a face of the cube. The incident plane wave goes on a corner of the cube where the material is discontinuous.
- A gap-toothed cube (non-convex and non-smooth surface). The mesh is composed of 20 628 edges and the wavenumber k is equal to 14 m^{-1}. Γ_i is composed of three faces as depicted in Figure 1. On each face, we impose a different impedance value (piecewise constant impedance). In particular, we have taken $\eta = 0.3, 0.5, 0.8$.
- A Channel cavity (non-convex, non-smooth surface and cavity effect) which is depicted in Figure 2. The mesh is composed of 153 033 edges and the wavenumber k is equal to 105 m^{-1}. We take $\eta = 0.035 + 0.035\,i$ on Γ_i which corresponds to the circular part to the bottom of the cavity. For this example, we have coupled the ICFIE with a SParse Approximate Inverse (SPAI) preconditioner.

First, we focus on the influence of p (*i.e.* the truncation of the Padé expansion) on the convergence of iterative solver and the accuracy of the solution. In [16], it is proved that $p = 8$ and $\theta = \pi/3$ yield to a satisfactory approximation of the square-root operator. These values are derived in order to construct an efficient On-surface Radiation Condition (OSRC) in the high-frequency domain. So, the degree of the approximation has a direct impact on the accuracy of the solution. In our case, this impact can be qualified of "indirect" (in comparison of the previous) in the sense that it is firstly the integral formulation (3.17) which is modified and then it is its discretization which gives the quality of the solution. So, one can have an accurate solution for $p = 0$.

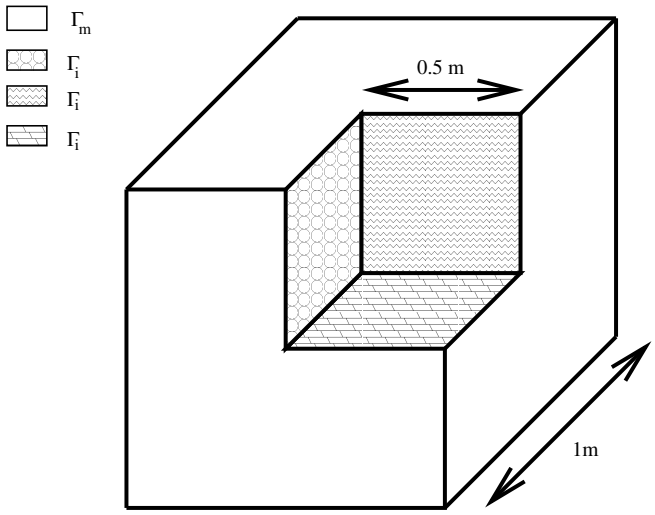


FIGURE 1. Gap-toothed cube.

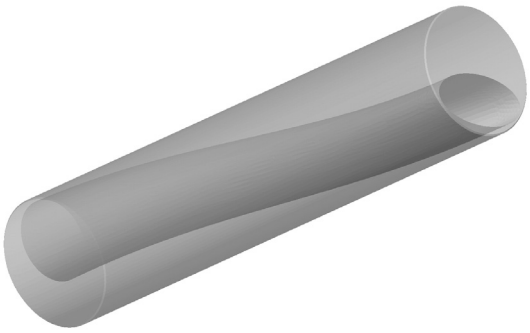


FIGURE 2. Channel cavity.

We clarify this point in order to avoid all confusion. One thus can hope to find a compromise between the robustness (number of iterations) and the accuracy. To illustrate that, we consider the case of the partially coated cube with 8460 edges and $k = 20$. Figure 4 represents the Radar Cross Section (RCS) in function of p. We can see that the solutions are similar. With regards to the robustness, Figure 3 shows that for this configuration the number of iterations reaches its minimum (25 for a residual equals to 10^{-8}) as soon as $p = 2$. For the rest of the numerical experiments, we have taken $p = 2$.

Figures 7 give the RCS diagrams calculated with the new formulation and ICFIE for the box and the cavity. It shows that the accuracy of the proposed formulation is comparable to the one of ICFIE for RCS computation.

Figures 5 respectively show the evolution of the number of iterations according to the discretization (number of points per wavelength) and the wavenumber k for the totally and partially coated sphere. For the new formulation, the number of matrix-vector product is independent of the discretization step and the frequency. Only 9 and 10 iterations are necessary to obtain the solution. For the ICFIE, one can observe a linear dependence which is most important for the space refinement. One also sees that the discontinuity of the material has a direct impact on the number of iterations for the two formulations.

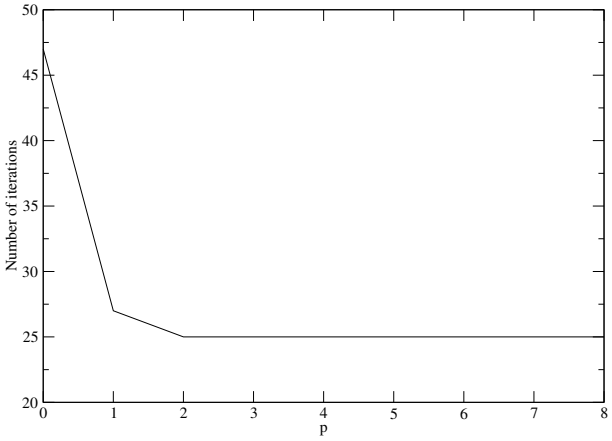


FIGURE 3. Number of iterations in function of p for the partially coated cube, $k = 20$ and the residue equals to 10^{-8}.

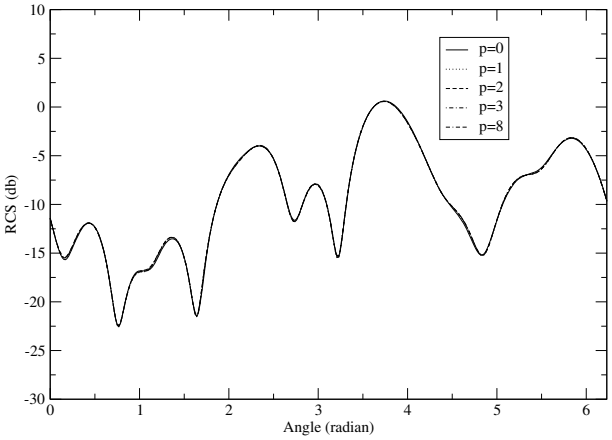


FIGURE 4. RCS in function of p for the partially coated cube, $k = 20$ and the residue equals to 10^{-8}.

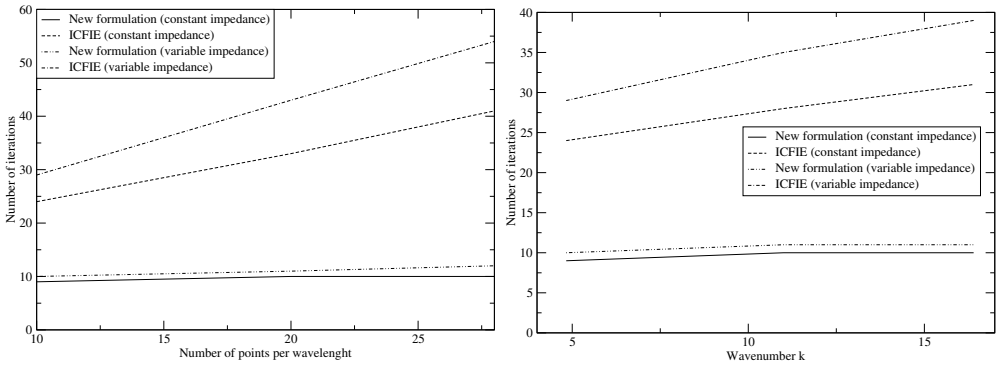


FIGURE 5. Behavior for the space refinement (left-hand side) and the frequency increase (right-hand side) for the sphere.

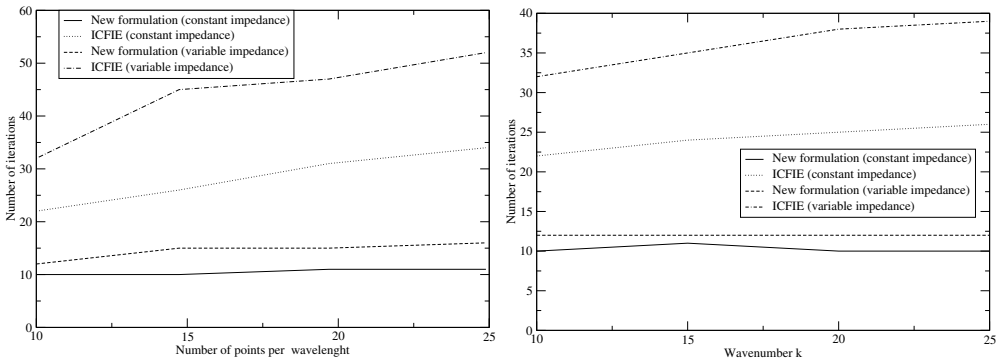


FIGURE 6. Behavior for the space refinement (left-hand side) and the frequency increase (right-hand side) for the cube.

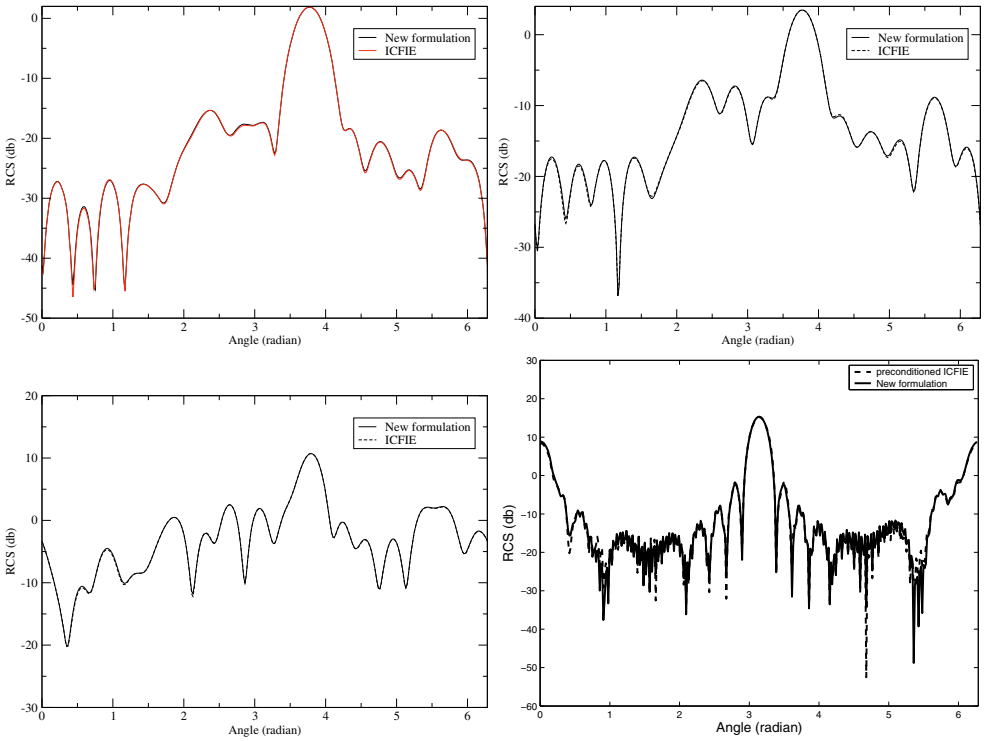


FIGURE 7. Comparison of the RCS between the new formulation and the ICFIE for the cube (up) (13 374 edges and $k = 25$) with constant (left-hand side) and variable (right-hand side) impedance, the gap-toothed cube (down-left-hand side) and the channel cavity (down-right-hand side).

We have carried out the same analysis for the cube. The results are given in Figures 6. For ICFIE, the conclusion is the same one as for the sphere. For the new formulation, we observe a slight discretization dependence. Nevertheless, the stability of the method stays good and the gain in terms of iteration is still important.

Finally, we have tested the method for two partially coated non-convex geometries: the gap-toothed cube and the channel cavity. Figure 8 shows the evolution of residues according to the number of matrix-vector products

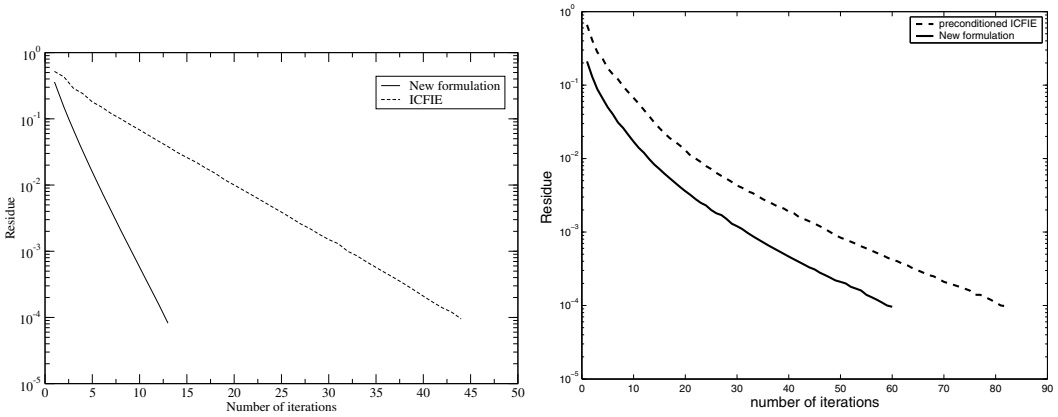


FIGURE 8. Evolution of the residue according to the number of matrix-vector products for the gap-toothed cube (left-hand side) and the channel cavity (right-hand side).

for a GMRES tolerance equals to 10^{-4}. We observe an good convergence rate for these types of geometry. With concern the channel cavity, we have used a SPAI preconditioner for the ICFIE and we have speed up the CPU time of a factor 3 by using the new formulation.

7. Conclusion

We have proposed a well-conditioned integral formulation for the electromagnetic scattering with the variable Leontovitch condition. Several numerical test-cases have proved that this formulation leads to convergence rates (for a GMRES solver) significantly better than the one of ICFIE. Moreover, these rates are almost frequency and discretization-step independent. The accuracy of solutions (for the RCS) are of the same order than ICFIE. Finally, this method seems to be a good candidate to efficiently treat more complicated problems (in particular a greater number of degrees of freedom). Indeed, as we have already said, it only requires two fast multipole calculations whereas we use four of this one for ICFIE. Moreover, the choice of $p = 2$ for the truncation of the Padé expansion implies small sparse matrices that one can treat at lower cost by MUMPS. This study is actually in progress. A comparison with ICFIE preconditioned by a SPAI will be interesting to really evaluate the cost of the method.

References

[1] F. Alouges, S. Borel and D. Levadoux, A stable well conditioned integral equation for electromagnetism scattering. *J. Comput. Appl. Math.* **204** (2007) 440–451.

[2] X. Antoine and H. Barucq, Microlocal diagonalization of strictly hyperbolic pseudodifferential systems and application to the design of radiation conditions in electromagnetism. *SIAM J. Appl. Math.* **61** (2001) 1877–1905.

[3] X. Antoine and M. Darbas, Generalized combined field integral equations for the iterative solution of the three-dimensional Helmholtz equation. *ESAIM: M2AN* **41** (2007) 147–167.

[4] A. Bendali, M'B Fares and J. Gay, A boundary-element solution of the Leontovitch problem. *IEEE Trans. Antennas Propagat.* **47** (1999) 1597–1605.

[5] Y. Boubendir, *Techniques de décomposition de domaine et méthode d'équations intégrales*. Ph.D. Thesis, INSA, France (2002).

[6] A. Buffa, Hodge decomposition on the boundary of a polyhedron: the multi-connected case. *Math. Mod. Meth. Appl. Sci.* **11** (2001) 1491–1504.

[7] A. Buffa and R. Hiptmair, Galerkin Boundary Element Methods for Electromagnetic Scattering, in *Computational Methods in Wave Propagation*, M. Ainsworth, P. Davies, D.B. Duncan, P.A. Martin and B. Rynne Eds., *Lecture Notes in Computational Science and Engineering* **31**, Springer-Verlag (2003) 83–124.

[8] F. Cakoni, D. Colton and P. Monk, The electromagnetic inverse scattering problem for partially coated Lipschitz domains. *Proc. Royal. Soc. Edinburgh* **134A** (2004) 661–682.

[9] S.L. Campbell, I.C.F. Ipsen, C.T. Kelley, C.D. Meyer and Z.Q. Xue, *Convergence estimates for solution of integral equations with GMRES*. Tech. Report CRSC-TR95-13, North Carolina State University, Center for Research in Scientific Computation, USA (1995).
[10] S.L. Campbell, I.C.F. Ipsen, C.T. Kelley and C.D. Meyer, GMRES and the Minimal Polynomial. *BIT Numerical Mathematics* **36** (1996) 664–675.
[11] H.S. Christiansen, *Résolution des équations intégrales pour la diffraction d'ondes acoustiques et électromagnétiques – Stabilisation d'algorithmes itératifs et aspects de l'analyse numérique*. Ph.D. Thesis, Centre de Mathématiques Appliquées, UMR 7641, CNRS/École polytechnique, France (2002).
[12] S. Christiansen and J.C. Nédélec, A preconditioner for the electric field integral equation based on Calderon formulas. *SIAM J. Numer. Anal.* **40** (2002) 1100–1135.
[13] F. Collino, S. Ghanemi and P. Joly, Domain decomposition method for the Helmholtz equation: a general presentation. *Comput. Methods Appl. Mech. Eng.* **184** (2000) 171–211.
[14] F. Collino, F. Millot and S. Pernet, Boundary-integral methods for iterative solution of scattering problems with variable impedance surface condition. *PIER* **80** (2008) 1–28.
[15] D. Colton and R. Kress, *Inverse acoustic and electromagnetic scattering theory*, *Applied Mathematical Sciences* **93**. Springer, Berlin, Germany (1992).
[16] M. Darbas, *Préconditionneurs analytiques de type Calderon pour les formulations intégrales des problèmes de direction d'ondes*. Ph.D. Thesis, INSA Toulouse, France (2004).
[17] M. Darbas, Generalized CFIE for the Iterative Solution of 3-D Maxwell Equations. *Appl. Math. Lett.* **19** (2006) 834–839.
[18] M. Darbas, *Some second-kind integral equations in electromagnetism*. Preprint, Cahiers du Ceremade 2006-15 (2006) http://www.ceremade.dauphine.fr/preprints/CMD/2006-15.pdf.
[19] V. Frayssé, L. Giraud, S. Gratton and J. Langou, *A Set of GMRES Routines for Real and Complex Arithmetics on High Performance Computers*. CERFACS Technical Report, TR/PA/03/3 (2003) http://www.cerfacs.fr/algor/Softs/GMRES/index.html.
[20] J.-F. Lee, R. Lee and R.J. Burkholder, Loop star basis functions and a robust preconditioner for EFIE scattering problems. *IEEE Trans. Antennas Propagat.* **51** (2003) 1855–1863.
[21] M.A. Leontovitch, *Approximate boundary conditions for the electromagnetic field on the surface of a good conductor, Investigations Radiowave Propagation part II*. Academy of Sciences, Moscow, Russia (1978).
[22] J.R. Mautz and R.F. Harrington, A combined-source solution for radiation and scattering from a perfectly conducting body. *IEEE Trans. Antennas Propag.* **AP-27** (1979) 445–454.
[23] F.A. Milinazzo, C.A. Zala, G.H. Brooke, Rational square-root approximations for parabolic equation algorithms. *J. Acoust. Soc. Am.* **101** (1997) 760–766.
[24] K.M. Mitzner, *Numerical solution of the exterior scattering problem at eigenfrequencies of the interior problem*. Int. Scientific Radio Union Meeting, Boston, USA (1968).
[25] P. Monk, *Finite Element Methods for Maxwell's Equations*, *Numerical Mathematics and Scientific Computation*. Oxford Science Publication, UK (2003).
[26] Multifrontal Massively Parallel Solver, www.enseeiht.fr/lima/apo/MUMPS.
[27] J.C. Nédélec, *Acoustic and Electromagnic Equations Integral Representation for Harmonic Problems*. Springer, New York, USA (2001).
[28] S.M. Rao, D.R. Wilton and A.W. Glisson, Electromagnetic scattering by surfaces of arbitrary shape. *IEEE Trans. Antennas Propagat.* **AP-30** (1982) 409–418.
[29] V. Rokhlin, Diagonal form of translation operators for the Helmholtz equation in three dimensions. *Appl. Comput. Harmon. Anal.* **1** (1993) 82–93.
[30] O. Steinbach and W.L. Wendland, The construction of some efficient preconditioners in the boundary element method. *Adv. Comput. Math.* **9** (1998) 191–216.
[31] B. Stupfel, A hybrid finite element and integral equation domain decomposition method for the solution of the 3-D scattering problem. *J. Comput. Phys.* **172** (2001) 451–471.

Time-Harmonic Acoustic Scattering in a Complex Flow: a Full Coupling Between Acoustics and Hydrodynamics.

A.-S. Bonnet-Ben Dhia[1,2,*], J.-F. Mercier[1], F. Millot[2], S. Pernet[2] and E. Peynaud[2]

[1] *POEMS, CNRS-INRIA-ENSTA UMR 7231, 32 Boulevard Victor, 75015 Paris, France.*
[2] *CERFACS, 42, Avenue Gaspard Coriolis, 31057 Toulouse Cedex 01, France.*

Abstract. For the numerical simulation of time harmonic acoustic scattering in a complex geometry, in presence of an arbitrary mean flow, the main difficulty is the coexistence and the coupling of two very different phenomena: acoustic propagation and convection of vortices. We consider a linearized formulation coupling an augmented Galbrun equation (for the perturbation of displacement) with a time harmonic convection equation (for the vortices). We first establish the well-posedness of this time harmonic convection equation in the appropriate mathematical framework. Then the complete problem, with Perfectly Matched Layers at the artificial boundaries, is proved to be coercive + compact, and a hybrid numerical method for the solution is proposed, coupling finite elements for the Galbrun equation and a Discontinuous Galerkin scheme for the convection equation. Finally a 2D numerical result shows the efficiency of the method.

AMS subject classifications: 35Q35, 65N30

Key words: aeroacoustics, scattering of sound in flows, Galbrun equation, advection equation, finite elements, Discontinuous Galerkin method

1 Introduction

The reduction of noise is becoming today a main objective whose progress is, in particular, related to a better understanding of the complex phenomena occurring when acoustic waves propagate in presence of a mean flow. For instance, the radiation of the sound produced by aircraft engines is strongly influenced by the presence of the flow around the airplane. Several methods have been developed to solve the time-domain Linearized Euler Equations, but the treatment of the artificial boundaries still raises open questions.

*Corresponding author. *Email address:* Anne-Sophie.Bonnet-Bendhia@ensta.fr (A.-S. Bonnet-Ben Dhia)

On the other hand, the time-harmonic problem has been considered only in the simplest case of a potential mean flow, apart for some attempts to solve the model of Galbrun in a general flow [9]. Galbrun's system corresponds to a linearized model whose unknown $\mathbf{u}$ is the perturbation of the Lagrangian displacement. It results in second order equations in time and in space, at first sight similar to more classical wave models. Contrary to the Linearized Euler Equations, Galbrun's system does not involve any derivatives of the mean flow quantities.

Our objective is to develop a numerical method to solve time-harmonic Galbrun's system, in a quite general case in the sense that the geometry, and therefore the mean flow, can be complex. As a consequence, discretization methods written on an unstructured mesh will be privileged. It is now well-known that a direct resolution using finite elements combined with Perfectly Matched Layers does not work. Extending an approach originally applied to time-harmonic Maxwell equations, we have shown that the difficulties can be overcome by writing a so-called augmented equation. This augmented equation requires the evaluation of $\psi = \mathrm{curl}\,\mathbf{u}$, which becomes the main difficulty.

This approach has been developed in 2D and applied successively to the case of a uniform flow and to the case of a non-vanishing parallel shear flow. In the first case, ψ can be computed a priori [8] and in the second case, it is explicitly related to $\mathbf{u}$ by a non-local convolution formula [4]. A simplified approach has been proposed in the case of a low Mach flow [3]: we can then replace the exact non-local expression of ψ by a simple local formula. This low Mach approach has been validated in the case of both a potential and a parallel flow, for which reference solutions are available.

The objective of the present paper is to get rid of the low Mach hypothesis. The main part of the paper is devoted to the theoretical study of the time-harmonic advection equation satisfied by ψ. The well-posedness results that we establish cannot be directly deduced from known results on the classical advection equation [7], but the techniques we use are inspired from [1].

The outline of the paper is the following. The model is briefly described in section 2, including the augmented equation for $\mathbf{u}$, the hydrodynamic equation for ψ and the Perfectly Matched Layers. Details can be found in [3]. Section 3 is devoted to the theoretical study of the time-harmonic advection equation. Well-posedness is deduced from an inf-sup condition, which is proved for a flow which "fills" the domain, in the sense of [1]. These results are used in section 4 to prove that the complete problem in $(\mathbf{u},\psi)$ with Perfectly Matched Layers is of Fredholm type if the flow varies slowly. A numerical method, coupling classical finite elements for $\mathbf{u}$ with a Discontinuous Galerkin scheme for ψ is finally described in section 5 and some numerical results are presented.

2 A model for acoustic scattering in a complex flow

2.1 Geometry and flow

Let $\Omega_\infty = \{(x_1,x_2) \in \mathbb{R}^2; x_2 > h(x_1)\}$ where h is a continuous function such that, for some positive r, $h(x_1) = 0$ for $|x_1| > r$. We suppose that Ω_∞ is filled with a compressible inviscid fluid and that the boundary $\Gamma_\infty = \{(x_1,x_2) \in \mathbb{R}^2; x_2 = h(x_1)\}$ is rigid. The fluid is moving and the flow, which is stationary, is characterized by its non uniform fields of velocity $\mathbf{v}_0$, density ρ_0, pressure p_0 and sound velocity c_0, which solve in Ω_∞ the stationary Euler equations:

$$\begin{cases} \operatorname{div}(\rho_0 \mathbf{v}_0) = 0 \\ \rho_0 \mathbf{v}_0 \cdot \nabla \mathbf{v}_0 + \nabla p_0 = 0 \end{cases} \tag{2.1}$$

On the rigid boundary:

$$\mathbf{v}_0 \cdot \mathbf{n} = 0 \quad (\Gamma_\infty) \tag{2.2}$$

where $\mathbf{n}$ denotes the normal vector to Γ_∞ pointing to the exterior of Ω_∞. Finally, for a barotropic fluid, the state law reads:

$$\nabla p_0 = c_0^2 \nabla \rho_0 \tag{2.3}$$

We suppose that the flow is subsonic and uniform far from the perturbation:

$$\exists R > 0 / \text{ for } |x| > R, \ \mathbf{v}_0(x) = v_\infty \mathbf{e}_1 \text{ and } \quad (\rho_0(x), p_0(x), c_0(x)) = (\rho_\infty, p_\infty, c_\infty)$$

This means that the half disk $D_R = \{(x_1,x_2); x_2 > 0 \text{ and } x_1^2 + x_2^2 < R^2\}$ of radius R contains the perturbed area of the propagation domain. We suppose for instance that $v_\infty > 0$. Finally, we assume for simplicity that all quantities related to the mean flow are regular

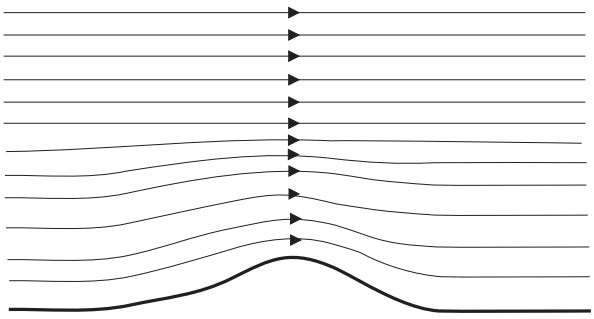


Figure 1: Mean flow

enough in the sense that ρ_0, p_0, c_0 and $\mathbf{v}_0$ are in $\mathcal{C}^2(\overline{\Omega}_\infty)$. This regularity will be used to define the operators $\mathcal{B}$ and $\mathcal{C}$ below.

2.2 The augmented Galbrun equation

The Galbrun equation is a linear equation which models the propagation of small perturbations of the previous mean flow, which can be produced by an acoustic source. In time harmonic regime (with a $e^{-i\omega t}$ time dependence, $\omega > 0$), this equation takes the following form:

$$\rho_0 \frac{D^2\mathbf{u}}{Dt^2} - \nabla(\rho_0 c_0^2 \mathrm{div}\,\mathbf{u}) + \mathrm{div}\,\mathbf{u}\nabla p_0 - {}^t\nabla\mathbf{u}\cdot\nabla p_0 = \mathbf{f} \quad (\Omega_\infty) \tag{2.4}$$

where the convective derivative $\dfrac{D\mathbf{u}}{Dt}$ is defined by:

$$\frac{D\mathbf{u}}{Dt} = -i\omega\mathbf{u} + \nabla\mathbf{u}\cdot\mathbf{v}_0$$

and f is a source term, compactly supported in $D_R \cap \Omega_\infty$, such that $\mathrm{curl}\,\mathbf{f} \in L^2(\Omega_\infty)$. The unknown $\mathbf{u}$ is the perturbation of displacement. Its normal component vanishes on the rigid boundary:

$$\mathbf{u}\cdot\mathbf{n} = 0 \quad (\Gamma_\infty) \tag{2.5}$$

Let us notice that outside the perturbed area D_R, equation (2.4) takes the following simplified form:

$$\frac{D^2\mathbf{u}}{Dt^2} - c_\infty^2 \nabla(\mathrm{div}\,\mathbf{u}) = \frac{1}{\rho_\infty}\mathbf{f} \tag{2.6}$$

It is well-known that a direct finite element discretization of (2.4) (using Lagrange elements) leads to a polluted result, due to a lack of H^1 coerciveness. A way to restore coerciveness is to consider the following "augmented" formulation:

$$\begin{aligned} &\rho_0 \frac{D^2\mathbf{u}}{Dt^2} - \nabla(\rho_0 c_0^2 \mathrm{div}\,\mathbf{u}) + \mathbf{curl}(\rho_0 c_0^2(\mathrm{curl}\,\mathbf{u} - \psi)) \\ &+ \mathrm{div}\,\mathbf{u}\nabla p_0 - {}^t\nabla\mathbf{u}\cdot\nabla p_0 = \mathbf{f} \end{aligned} \tag{2.7}$$

where we have introduced a new unknown

$$\psi = \mathrm{curl}\,\mathbf{u}$$

called here the "vorticity" (in the literature, the vorticity is usually defined as the curl of the Eulerian velocity). It has been proved in [3] that ψ satisfies the following equation

$$\frac{D^2\psi}{Dt^2} = -2\frac{D}{Dt}(\mathcal{B}\mathbf{u}) - \mathcal{C}\mathbf{u} + \frac{1}{\rho_0}\mathrm{curl}\,\mathbf{f} \tag{2.8}$$

with

$$\mathcal{B}\mathbf{u} = \sum_{j=1}^{2} \nabla v_{0,j} \wedge \frac{\partial\mathbf{u}}{\partial x_j} \tag{2.9}$$

and

$$\begin{aligned}\mathcal{C}\mathbf{u} = &\sum_{j,k=1}^{2}\left(\frac{\partial v_{0,k}}{\partial x_j}\nabla v_{0,j}\wedge\frac{\partial \mathbf{u}}{\partial x_k} - v_{0,j}\nabla\frac{\partial v_{0,k}}{\partial x_j}\wedge\frac{\partial \mathbf{u}}{\partial x_k}\right)\\ &+\frac{1}{\rho_0}\sum_{j=1}^{2}\left(\frac{1}{\rho_0 c_0^2}\frac{\partial p_0}{\partial x_j}\nabla p_0 - \nabla\left(\frac{\partial p_0}{\partial x_j}\right)\right)\wedge\nabla u_j\end{aligned} \tag{2.10}$$

Notice that $\mathcal{C}\mathbf{u} = \mathcal{B}\mathbf{u} = 0$ outside the perturbed area D_R and that $\mathcal{C}\mathbf{u}$ vanishes everywhere for a parallel shear flow.

For the coupled problem (2.7, 2.5, 2.8) to be equivalent with the initial problem (2.4, 2.5), the following additional boundary condition must be imposed:

$$\mathrm{curl}\,\mathbf{u} - \psi = 0 \quad (\Gamma_\infty) \tag{2.11}$$

Moreover let us point out that the equivalence (with $\mathbf{u}\in H^1(\Omega_\infty)^2$) requires the regularity of $\partial\Omega_\infty$ (see for instance the remark 3.5 in [5]), and the treatment of reentrant corners still raises open questions of modelling.

2.3 The Perfectly Matched Layers

Our objective is the computation of the "outgoing" solution $(\mathbf{u},\psi)$ of the coupled problem (2.7 ,2.5, 2.8, 2.11). Following [2], for this outgoing solution, ψ must vanish upstream of the perturbation area: indeed, the vortices are produced by the source and by the coupling between acoustics and hydrodynamics in the case of a non uniform flow, and are then convected downstream by the flow. In practice, we use PMLs to select this

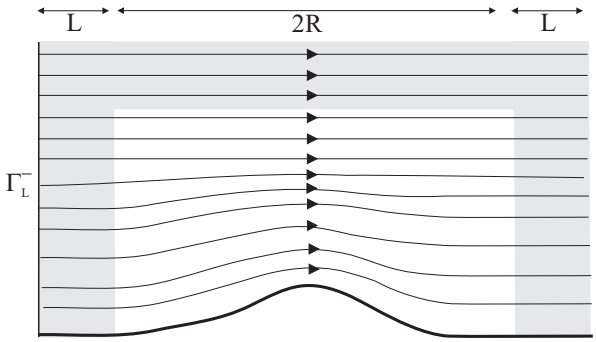


Figure 2: The computational domain with the Perfectly Matched layers

outgoing solution. The computational domain is defined by $\Omega_L = B_L \cap \Omega_\infty$ where B_L is the following square

$$B_L = \{(x_1,x_2)\,/\,|x_1| < R+L \text{ and } 0 < x_2 < R+L\}$$

and L denotes the width of the absorbing layers. The model in the PMLs involves a complex parameter α such that $\Re e(\alpha) > 0$ and $\Im m(\alpha) < 0$. In the following the index α means that the corresponding operator has been modified according to the substitution:

$$\frac{\partial}{\partial x_i} \to \alpha_i(x) \frac{\partial}{\partial x_i}$$

with α_i defined by $\alpha_i(x) = 1$ if $|x_i| < R$ and $\alpha_i(x) = \alpha$ if $|x_i| > R$. For example,

$$\mathrm{div}_\alpha \mathbf{u} = \alpha_1(x) \frac{\partial u_1}{\partial x_1} + \alpha_2(x) \frac{\partial u_2}{\partial x_2}$$

Finally, the problem that we solve is the following:

$$\begin{array}{lll}
\rho_0 \dfrac{D_\alpha^2 \mathbf{u}}{Dt^2} - \nabla_\alpha(\rho_0 c_0^2 \mathrm{div}_\alpha \mathbf{u}) + \mathbf{curl}_\alpha(\rho_0 c_0^2 (\mathrm{curl}_\alpha \mathbf{u} - \psi)) & & \\
+ \mathrm{div}_\alpha \mathbf{u} \nabla p_0 - {}^t\nabla_\alpha \mathbf{u} \cdot \nabla p_0 = \mathbf{f} & \text{in} & \Omega_L \\
\dfrac{D_\alpha^2 \psi}{Dt^2} = -2 \dfrac{D_\alpha}{Dt}(\mathcal{B}\mathbf{u}) - \mathcal{C}\mathbf{u} + \dfrac{1}{\rho_0} \mathrm{curl} \mathbf{f} & \text{in} & \Omega_L \\
\mathbf{u} \cdot \mathbf{n} = \mathrm{curl}_\alpha \mathbf{u} - \psi = 0 & \text{on} & \Gamma_\infty^L \\
\mathbf{u} = 0 & \text{on} & \Gamma_0^L \\
\psi = \dfrac{D_\alpha \psi}{Dt} = 0 & \text{on} & \Gamma_L^-
\end{array} \tag{2.12}$$

where $\Gamma_\infty^L = \Gamma_\infty \cap \partial\Omega_L$, $\Gamma_0^L = \partial\Omega_L \backslash \Gamma_\infty^L$ and Γ_L^- is the inflow boundary of the computational domain:

$$\Gamma_L^- = \{(x_1, x_2) / x_1 = -R - L \text{ and } 0 < x_2 < R + L\}$$

The boundary condition on Γ_L^- will allow to ensure the causality of ψ, as described below.

Notice that it is useless to introduce the notations $\mathcal{B}_\alpha$ or $\mathcal{C}_\alpha$ since $\mathcal{B}$ and $\mathcal{C}$ vanish in the absorbing layers. For the same reason, we wrote ∇p_0 instead of $\nabla_\alpha p_0$ in the first equation of (2.12).

It has been already proved (see for instance [3]) that for a given ψ, the problem in $\mathbf{u}$ is of Fredholm type in $H^1(\Omega_L)^2$. In the next section, we consider the problem in ψ for a given $\mathbf{u}$. Results for the coupled problem (2.12) are finally discussed in section 4.

3 The time-harmonic convective equation

3.1 A model problem

Let us first consider the following model problem :

$$\begin{array}{lll}
-i\omega\psi + \mathbf{v} \cdot \nabla\psi = g & \text{in} & \Omega \\
\psi = 0 & \text{on} & \Gamma^-
\end{array} \tag{3.1}$$

where Ω is a bounded domain of $\mathbb{R}^2$, $\mathbf{v}$ is a vector field defined on Ω such that

$$\mathbf{v} \in C^1(\overline{\Omega})^2 \quad \text{and} \quad \operatorname{div}\mathbf{v} = 0$$

and Γ^- (resp. Γ^+) is the inflow (resp. outflow) boundary ($\mathbf{n}$ is the exterior normal vector to Ω):

$$\Gamma^{\pm} = \{x \in \partial\Omega; \pm\mathbf{v}\cdot\mathbf{n} > 0\}$$

Contrary to the standard advection equation with $\omega = ia$ with $a \geq 0$ (see [7]), this time-harmonic equation seems not having been studied before. Following [7], we notice that $\operatorname{div}(\psi\mathbf{v}) = \mathbf{v}\cdot\nabla\psi$ and we consider the Hilbert space

$$H(\Omega,\mathbf{v}) = \{\psi \in L^2(\Omega); \operatorname{div}(\psi\mathbf{v}) \in L^2(\Omega)\} \tag{3.2}$$

equipped with the following norm:

$$\|\psi\|_{H(\Omega,\mathbf{v})} = \sqrt{\int_\Omega \omega^2|\psi|^2 + |\mathbf{v}\cdot\nabla\psi|^2} \tag{3.3}$$

where ω is introduced for homogeneity reasons. Then, we deduce from classical properties of the space $H(\Omega,\operatorname{div})$ the existence of a continuous trace application:

$$\psi \in H(\Omega,\mathbf{v}) \longmapsto \psi\,\mathbf{v}\cdot\mathbf{n} \in H^{-1/2}(\Gamma^-)$$

As a consequence, the space

$$H(\Omega,\mathbf{v},\Gamma^-) = \{\psi \in L^2(\Omega); \mathbf{v}\cdot\nabla\psi \in L^2(\Omega) \text{ and } \psi = 0 \text{ on } \Gamma^-\} \tag{3.4}$$

is a closed subspace of $H(\Omega,\mathbf{v})$. Naturally, the space $H(\Omega,\mathbf{v},\Gamma^+)$ can be defined in the same manner. Finally, we have the:

Lemma 3.1. *For all $\psi \in H(\Omega,\mathbf{v},\Gamma^-)$, $|\psi|^2\mathbf{v}\cdot\mathbf{n} \in L^1(\Gamma^+)$ and the following identities hold:*

$$\forall\psi \in H(\Omega,\mathbf{v},\Gamma^-)\,\forall\varphi \in H(\Omega,\mathbf{v}) \quad \int_\Omega (\mathbf{v}\cdot\nabla\psi)\,\varphi = -\int_\Omega \psi\,(\mathbf{v}\cdot\nabla\varphi) + \int_{\Gamma^+} \mathbf{v}\cdot\mathbf{n}\,\psi\varphi\,d\gamma \tag{3.5}$$

$$\forall\psi \in H(\Omega,\mathbf{v},\Gamma^-) \quad \int_{\Gamma^+} \mathbf{v}\cdot\mathbf{n}\,|\psi|^2 d\gamma \leq \frac{1}{\omega}\|\psi\|^2_{H(\Omega,\mathbf{v})} \tag{3.6}$$

We can now specify the functional framework well-suited for problem (3.1): for a data g in $L^2(\Omega)$, the solution ψ is sought in $H(\Omega,\mathbf{v},\Gamma^-)$.

3.2 Explicit solution in the case of a uniform flow

Let us first consider the case of a uniform vector field $\mathbf{v} = v\mathbf{e}_1$ ($v > 0$) in a rectangular domain $\Omega =]0,d[\times]0,\ell[$, so that $\Gamma^- = \{(0,x_2); 0 < x_2 < \ell\}$. Then problem (3.1) consists in finding $\psi \in H(\Omega,\mathbf{v})$ such that

$$\begin{array}{lll} v\dfrac{\partial \psi}{\partial x_1}(x_1,x_2) - i\omega\psi(x_1,x_2) = g(x_1,x_2) & \text{in} & \Omega \\ \psi(0,x_2) = 0 & \text{for} & 0 < x_2 < \ell \end{array} \tag{3.7}$$

where $g \in L^2(\Omega)$. This is a family of first order differential equations in x_1 (with constant coefficients) parametrized by x_2, whose solution $\psi = \psi_g$ is given by the following convolution formula

$$\psi_g(x_1,x_2) = \frac{1}{v}\int_0^{x_1} g(s,x_2) e^{i\frac{\omega}{v}(x_1-s)} ds \tag{3.8}$$

Notice the oscillating behavior (the smaller v, the smaller the wavelength). From formula (3.8) results the following stability estimate:

$$\|\psi_g\|_{H(\Omega,\mathbf{v})} \leq \sqrt{1+\sqrt{2}\frac{\omega d}{v}+\frac{\omega^2 d^2}{v^2}}\,\|g\|_{L^2(\Omega)} \tag{3.9}$$

Indeed, we have

$$\begin{array}{rcl} \|\psi_g\|^2_{H(\Omega,\mathbf{v})} & = & \|\omega\psi_g\|^2_{L^2(\Omega)} + \|v\dfrac{\partial \psi_g}{\partial x_1}\|^2_{L^2(\Omega)} \\ & = & \|\omega\psi_g\|^2_{L^2(\Omega)} + \|g + i\omega\psi_g\|^2_{L^2(\Omega)} \\ & \leq & 2\|\omega\psi_g\|^2_{L^2(\Omega)} + \|g\|^2_{L^2(\Omega)} + 2\|g\|_{L^2(\Omega)}\|\omega\psi_g\|_{L^2(\Omega)} \end{array} \tag{3.10}$$

and we directly obtain (3.9) by using the estimate

$$\begin{array}{rcl} \|\omega\psi_g\|^2_{L^2(\Omega)} & \leq & \displaystyle\int_0^d \int_0^\ell \frac{\omega^2}{v^2}\Big(\int_0^{x_1} |g(s,x_2)|ds\Big)^2 dx_2 dx_1 \\ & \leq & \displaystyle\frac{\omega^2}{v^2}\int_0^d \int_0^\ell \Big(\int_0^{x_1} |g(s,x_2)|^2 ds\Big)\Big(\int_0^{x_1} 1^2 ds\Big) dx_2 dx_1 \\ & \leq & \displaystyle\frac{\omega^2 d^2}{2v^2}\|g\|^2_{L^2(\Omega)} \end{array} \tag{3.11}$$

Let us point out that the estimate (3.9) deteriorates when the length d of the domain increases (with a linear dependence in d) or when the velocity v decreases.

3.3 Well-posedness in the general case

In the case of an arbitrary vector field $\mathbf{v}$, the previous approach is not generalizable and an explicit solution of problem (3.1) is not available. We will use instead a variational approach for problem (3.1) written in the following weak form:

$$\psi \in H(\Omega,\mathbf{v},\Gamma^-) \quad / \quad a(\psi,\phi) = \int_\Omega g\bar{\phi} \quad \forall \phi \in L^2(\Omega) \tag{3.12}$$

where $a(\psi,\phi) = \int_\Omega (-i\omega\psi + \mathbf{v}\cdot\nabla\psi)\bar{\phi}$. Following [1], we suppose that the vector field $\mathbf{v}$ satisfies the following additional hypothesis:

$$v^- = \inf_{x\in\Omega} \mathbf{v}(x)\cdot\mathbf{e}_1 > 0 \tag{3.13}$$

This condition implies that $\mathbf{v}$ is Ω-filling (see [1]) in the sense that every point in Ω can be reached by the flow associated to $\mathbf{v}$ in a finite time. In particular, recirculation zones are forbidden.

It is proved in [1] that, under condition (3.13), problem (3.1) with $\omega = 0$ is well-posed. The same result holds if Γ^- is replaced by Γ^+. In particular, there exists a unique real-valued function τ such that

$$\tau \in H(\Omega,\mathbf{v},\Gamma^+) \quad \text{and} \quad \mathbf{v}\cdot\nabla\tau = -2 \quad \text{in } \Omega \tag{3.14}$$

Moreover $\tau \in L^\infty(\Omega)$ and $\|\tau\|_{L^\infty(\Omega)}$ is twice the maximum time necessary for a particule convected by the flow $\mathbf{v}$ to go across the domain Ω, so that:

$$\|\tau\|_{L^\infty(\Omega)} \leq 2\frac{d(\Omega)}{v^-} \tag{3.15}$$

where $d(\Omega) = \max_{x\in\Omega} x_1 - \min_{x\in\Omega} x_1$. The function τ will be used to establish the following result:

Proposition 3.1. Under condition (3.13), the following inf-sup condition holds:

$$\inf_{\psi\in H(\Omega,\mathbf{v},\Gamma^-)} \sup_{\phi\in L^2(\Omega)} \Re e(a(\psi,\phi)) \geq \frac{1}{\beta}\|\psi\|_{H(\Omega,\mathbf{v})}\|\phi\|_{L^2(\Omega)}$$

where

$$\beta = 2\sqrt{2+4\left(\frac{\omega d(\Omega)}{v^-}\right)^2}$$

Proof. Let $\psi \in H(\Omega,\mathbf{v},\Gamma^-)$. Taking $\phi = \omega^2\tau\psi + \mathbf{v}\cdot\nabla\psi$, we get

$$\begin{aligned} \Re e(a(\psi,\phi)) &= \Re e\left(\int_\Omega -i\omega^3\tau|\psi|^2 + |\mathbf{v}\cdot\nabla\psi|^2 + \omega^2(\mathbf{v}\cdot\nabla\psi)\tau\bar{\psi} - i\omega\psi(\mathbf{v}\cdot\nabla\bar{\psi})\right) \\ &\geq \int_\Omega |\mathbf{v}\cdot\nabla\psi|^2 + \omega^2\Re e\left(\int_\Omega (\mathbf{v}\cdot\nabla\psi)\tau\bar{\psi}\right) - \omega\|\psi\|_{L^2(\Omega)}\|\mathbf{v}\cdot\nabla\psi\|_{L^2(\Omega)} \end{aligned} \tag{3.16}$$

Then applying (3.5), we get:

$$\int_\Omega (\mathbf{v}\cdot\nabla\psi)\tau\bar{\psi} = -\int_\Omega \tau\psi(\mathbf{v}\cdot\nabla\bar{\psi}) - \int_\Omega |\psi|^2\mathbf{v}\cdot\nabla\tau$$

which gives, using the properties of τ:

$$\Re e\left(\int_\Omega (\mathbf{v}\cdot\nabla\psi)\tau\bar{\psi}\right) = \|\psi\|^2_{L^2(\Omega)} \tag{3.17}$$

Combining (3.16) and (3.17), we obtain finally the following inequality:

$$\Re e(a(\psi,\phi)) \geq \frac{1}{2}\|\psi\|^2_{H(\Omega,\mathbf{v})} \tag{3.18}$$

On the other hand:

$$\|\phi\|^2_{L^2(\Omega)} = \|\mathbf{v}\cdot\nabla\psi\|^2_{L^2(\Omega)} + \omega^4\|\tau\psi\|^2_{L^2(\Omega)} + 2\omega^2\Re e\left(\int_\Omega (\mathbf{v}\cdot\nabla\psi)\tau\bar{\psi}\right)$$

which leads, using (3.17) to the following estimate:

$$\|\phi\|^2_{L^2(\Omega)} \leq (2+\omega^2\|\tau\|^2_{L^\infty(\Omega)})\|\psi\|^2_{H(\Omega,\mathbf{v})} \tag{3.19}$$

The theorem results from (3.18) and (3.19). □

Well-posedness of problem (3.1) is then a simple consequence of the previous proposition:

Theorem 3.1. *Under condition (3.13), (3.1) is well-posed and its solution ψ satisfies the following estimate:*

$$\|\psi\|_{H(\Omega,\mathbf{v})} \leq \beta\|g\|_{L^2(\Omega)} \tag{3.20}$$

where β has been defined in proposition 3.1.

Proof. Let us consider the operator A from $H(\Omega,\mathbf{v},\Gamma^-)$ to $L^2(\Omega)$ defined by $A\psi = -i\omega\psi + \mathbf{v}\cdot\nabla\psi$. From proposition 3.1, it results that A is injective and has a closed range. To prove the surjectivity, notice that the adjoint A^* of A which is defined from $H(\Omega,\mathbf{v},\Gamma^+)$ into $L^2(\Omega)$ by $A^*\psi = i\omega\psi + \mathbf{v}\cdot\nabla\psi$ is also injective (by a similar argument), so that the range of A is dense. □

Remark 3.1. In the particular case studied in subsection 3.2, estimate (3.20) becomes:

$$\|\psi\|_{H(\Omega,\mathbf{v})} \leq 2\sqrt{2+4\frac{\omega^2 d^2}{v^2}}\|g\|_{L^2(\Omega)}$$

which is in accordance with (3.9).

3.4 Some straightforward generalizations

Some simple extensions are required in order to apply the previous results to the acoustic problem (2.12).

3.4.1 The case of a compressible flow

The flow which is considered in the acoustic problem is a solution of Euler's equations (2.1). In particular, the velocity field $\mathbf{v}_0$ satisfies $\mathrm{div}(\rho_0\mathbf{v}_0)=0$ but not $\mathrm{div}\,\mathbf{v}_0=0$. As a consequence, the above results cannot be directly applied to the following problem

$$\begin{aligned} -i\omega\psi+\mathbf{v}_0\cdot\nabla\psi&=g \quad \text{in} \quad \Omega \\ \psi&=0 \qquad\quad\;\; \text{on} \quad \Gamma^- \end{aligned} \tag{3.21}$$

The idea is to write the first equation of (3.21) in the following equivalent form:

$$-i\omega\rho_0\psi+\rho_0\mathbf{v}_0\cdot\nabla\psi=\rho_0 g \quad \text{in} \quad \Omega \tag{3.22}$$

in order to use the equation $\mathrm{div}(\rho_0\mathbf{v}_0)=0$, and then to introduce a modified definition of function τ:

$$\tau\in H(\Omega,\rho_0\mathbf{v}_0,\Gamma^+) \quad \text{and} \quad \rho_0\mathbf{v}_0\cdot\nabla(\rho_0\tau)=-2\rho_0^2 \quad \text{in } \Omega \tag{3.23}$$

and a modified definition of the norm in the space $H(\Omega,\rho_0\mathbf{v}_0)$:

$$\|\psi\|_{H(\Omega,\rho_0\mathbf{v}_0)}=\sqrt{\int_\Omega \omega^2\rho_0^2|\psi|^2+|\rho_0\mathbf{v}_0.\nabla\psi|^2}$$

If we assume that the density ρ_0 belongs to $\mathrm{L}^\infty(\Omega)$ and is bounded from below by a strictly positive constant $\rho_0^{\inf}>0$, then we obtain with a very similar approach as above the following result:

Theorem 3.2. *If $\mathbf{v}_0$ satisfies condition (3.13), (3.21) is well-posed and its solution ψ satisfies the following estimate:*

$$\|\psi\|_{H(\Omega,\rho_0\mathbf{v}_0)}\leq\beta_0\|\rho_0 g\|_{L^2(\Omega)} \tag{3.24}$$

where

$$\beta_0=2\sqrt{2+4\left(\frac{\omega d(\Omega)}{v_0^-}\frac{\rho_0^{\sup}}{\rho_0^{\inf}}\right)^2}$$

and $\rho_0^{\sup}=\|\rho_0\|_\infty$.

3.4.2 The case of a second order time-harmonic convective equation

As the hydrodynamic equation (2.8) is a second order one, it will be useful to notice that the following problem

$$\begin{aligned} &\frac{D^2\psi}{Dt^2} = (-i\omega + \mathbf{v}_0\cdot\nabla)^2\psi = g \quad \text{in} \quad \Omega \\ &\psi = \frac{D\psi}{Dt} = 0 \qquad\qquad\qquad\qquad \text{on} \quad \Gamma^- \end{aligned} \tag{3.25}$$

can be very simply solved by introducing the intermediary unknown $\tilde{\psi} = \frac{D\psi}{Dt}$. By theorem 3.2, there exists a unique $\tilde{\psi} \in H(\Omega,\rho_0\mathbf{v}_0,\Gamma^-)$ solution of $\dfrac{D\tilde{\psi}}{Dt} = g$ and a unique $\psi \in H(\Omega,\rho_0\mathbf{v}_0,\Gamma^-)$ solution of $\dfrac{D\psi}{Dt} = \tilde{\psi}$, and therefore solution of (3.25), which satisfy the following estimates:

$$\|\tilde{\psi}\|_{H(\Omega,\rho_0\mathbf{v}_0)} \leq \beta_0 \|\rho_0 g\|_{L^2(\Omega)} \text{ and } \|\psi\|_{H(\Omega,\rho_0\mathbf{v}_0)} \leq \beta_0 \|\rho_0\tilde{\psi}\|_{L^2(\Omega)}$$

Summing up, we obtain the following estimate:

$$\|\psi\|_{H(\Omega,\rho_0\mathbf{v}_0)} \leq \frac{{\beta_0}^2}{\omega}\|\rho_0 g\|_{L^2(\Omega)} \tag{3.26}$$

Indeed, by using the definition of the norm $\|\cdot\|_{H(\Omega,\rho_0\mathbf{v}_0)}$, we immediately obtain:

$$\|\rho_0\tilde{\psi}\|_{L^2(\Omega)} \quad \leq \quad \frac{1}{\omega}\|\tilde{\psi}\|_{H(\Omega,\rho_0\mathbf{v}_0)} \tag{3.27}$$

4 Well-posedness of the coupled problem

We will now use the previous results on the time harmonic convective equation to prove, under some conditions on the mean flow, the well-posedness of the coupled problem (2.12). We suppose that $\mathbf{v}_0$ satisfies condition (3.13).

4.1 Estimates on ψ

Suppose first that $\mathbf{u} \in H_0^1(\Omega_L)^2$ and let us consider the following problem for ψ, which is part of problem (2.12):

$$\begin{aligned} &\frac{D_\alpha^2\psi}{Dt^2} = -2\frac{D_\alpha}{Dt}(\mathcal{B}\mathbf{u}) - \mathcal{C}\mathbf{u} + \frac{1}{\rho_0}\mathrm{curl}\mathbf{f} \quad \text{in} \quad \Omega_L \\ &\psi = \frac{D_\alpha\psi}{Dt} = 0 \qquad\qquad\qquad\qquad\qquad\quad \text{on} \quad \Gamma_L^- \end{aligned} \tag{4.1}$$

Simple considerations show that ψ vanishes in $\Omega_L \backslash (\Omega_R \cup Q_L^+)$ where we have set

$\Omega_R = \{(x_1,x_2)/|x_1|<R \text{ and } 0<x_2<R\} \cap \Omega_\infty$ and $Q_L^+ = \{(x_1,x_2)/R<x_1<R+L \text{ and } 0<x_2<R\}$

Problem (4.1) can be solved by solving first the problem in Ω_R:

$$\begin{aligned} &\frac{D^2\psi}{Dt^2} = -2\frac{D}{Dt}(\mathcal{B}\mathbf{u}) - \mathcal{C}\mathbf{u} + \frac{1}{\rho_0}\mathrm{curl}\mathbf{f} \quad \text{in} \quad \Omega_R \\ &\psi = \frac{D\psi}{Dt} = 0 \qquad\qquad\qquad\qquad\quad \text{on} \quad \Gamma_R^- \end{aligned} \tag{4.2}$$

where $\Gamma_R^\pm = \{(x_1,x_2)/\pm x_1 = R \text{ and } 0<x_2<R\}$. The solution of (4.2) then provides initial conditions on Γ_R^+ (which is the inflow boundary of Q_L^+) to compute ψ in Q_L^+, which is a solution of the following homogeneous equation (since $\mathcal{B}$, $\mathcal{C}$, and $\mathbf{f}$ are supported in Ω_R) with constant coefficients:

$$\frac{D_\alpha^2\psi}{Dt^2} = \left(-i\omega + \alpha v_\infty \frac{\partial}{\partial x_1}\right)^2 \psi = 0 \quad \text{in} \quad Q_L^+ \tag{4.3}$$

By linearity, the solution ψ of (4.2) is given by $\psi = \psi_\mathcal{B} + \psi_\mathcal{C} + \psi_\mathbf{f}$ where $\psi_\mathcal{B}$ is a solution of

$$\begin{aligned} &\frac{D\psi_\mathcal{B}}{Dt} = -2\mathcal{B}\mathbf{u} \quad \text{in} \quad \Omega_R \\ &\psi_\mathcal{B} = 0 \qquad\qquad\quad \text{on} \quad \Gamma_R^- \end{aligned} \tag{4.4}$$

and $\psi_\mathcal{C}$ and $\psi_\mathbf{f}$ satisfy the same homogeneous initial conditions on Γ_R^- as ψ and the following equations in Ω_R:

$$\frac{D^2\psi_\mathcal{C}}{Dt^2} = -\mathcal{C}\mathbf{u} \quad \text{and} \quad \frac{D^2\psi_\mathbf{f}}{Dt^2} = \frac{1}{\rho_0}\mathrm{curl}\mathbf{f}$$

Results of subsection 3.4 prove the existence of $\psi_\mathcal{B}$ and $\psi_\mathcal{C}$ ($\psi_\mathbf{f}$ can be treated like $\psi_\mathcal{C}$) and the following estimates:

$$\begin{aligned} &\|\psi_\mathcal{B}\|_{H(\Omega_R,\rho_0\mathbf{v}_0)} \le 2\beta_0 \|\rho_0\mathcal{B}\mathbf{u}\|_{L^2(\Omega_R)} \\ &\|\psi_\mathcal{C}\|_{H(\Omega_R,\rho_0\mathbf{v}_0)} \le \frac{{\beta_0}^2}{\omega}\|\rho_0\mathcal{C}\mathbf{u}\|_{L^2(\Omega_R)} \text{ and } \|\frac{D\psi_\mathcal{C}}{Dt}\|_{H(\Omega_R,\rho_0\mathbf{v}_0)} \le \beta_0 \|\rho_0\mathcal{C}\mathbf{u}\|_{L^2(\Omega_R)} \end{aligned} \tag{4.5}$$

Then $\psi_\mathcal{B}$ and $\psi_\mathcal{C}$ can be extended in Q_L^+ by solving (4.3):

$$\begin{aligned} &\psi_\mathcal{B}(x_1,x_2) = \psi_\mathcal{B}(R,x_2) e^{i\frac{\omega}{\alpha v_\infty}(x_1-R)} \\ &\psi_\mathcal{C}(x_1,x_2) = \left(\psi_\mathcal{C}(R,x_2) + \frac{x_1-R}{\alpha v_\infty}\frac{D\psi_\mathcal{C}}{Dt}(R,x_2)\right) e^{i\frac{\omega}{\alpha v_\infty}(x_1-R)} \qquad R<x_1<R+L \end{aligned} \tag{4.6}$$

Combining (3.6), (4.5) and (4.6) and setting $\gamma_L = \omega L/v_\infty$, we finally get (after some calculations using the rough estimate $|e^{i\frac{\omega}{\alpha v_\infty}(x_1-R)}| \leq 1$):

$$\begin{aligned}
&\|\rho_0\psi_{\mathcal{B}}\|_{L^2(\Omega_L)} \leq \frac{2\beta_0}{\omega}\sqrt{1+\frac{\rho_\infty}{\rho_0^{\inf}}\gamma_L}\,\|\rho_0\mathcal{B}\mathbf{u}\|_{L^2(\Omega_R)} \\
&\|\rho_0\psi_{\mathcal{C}}\|_{L^2(\Omega_L)} \leq \frac{\beta_0}{\omega^2}\sqrt{\beta_0^2+2\frac{\rho_\infty}{\rho_0^{\inf}}\gamma_L\left(\beta_0^2+\frac{\gamma_L^2}{3|\alpha|^2}\right)}\,\|\rho_0\mathcal{C}\mathbf{u}\|_{L^2(\Omega_R)}
\end{aligned} \tag{4.7}$$

4.2 Coercivity condition

The main result that will be proved now is the well-posedness of problem (2.12). We denote by V the functional space for the fields $\mathbf{u}$:

$$V = \{\mathbf{u} \in H^1(\Omega_L)^2;\, \mathbf{u} = 0 \text{ on } \Gamma_0^L \text{ and } \mathbf{u}\cdot\mathbf{n} = 0 \text{ on } \Gamma_\infty^L\}$$

Let us introduce now the operator $\mathcal{T}$, defined from V into $L^2(\Omega_L)$, such that $\mathcal{T}\mathbf{u} = \psi_{\mathcal{B}} + \psi_{\mathcal{C}}$. Then the solution ψ of (4.2) is given by

$$\psi = \mathcal{T}\mathbf{u} + \psi_{\mathbf{f}} \tag{4.8}$$

By (4.7), $\mathcal{T}$ is a bounded operator satisfying:

$$\|\rho_0^{1/2}c_0\mathcal{T}\mathbf{u}\|_{L^2(\Omega_L)} \leq \frac{\beta_0}{\omega}\left(K_1\|\mathcal{B}\| + \frac{K_2\beta_0 + K_3}{\omega}\|\mathcal{C}\|\right)\|\rho_0^{1/2}c_0\nabla\mathbf{u}\|_{L^2(\Omega_R)} \tag{4.9}$$

where K_i are dimensionless constants depending only on γ_L and ρ_0, and where the norms of operators $\mathcal{B}$ and $\mathcal{C}$ are defined by (the weight $\rho_0^{1/2}c_0$ will be well-suited in what follows):

$$\|\mathcal{B}\| = \sup_{\mathbf{u}\in V,\,\mathbf{u}\neq 0}\frac{\|\rho_0^{1/2}c_0\mathcal{B}\mathbf{u}\|_{L^2(\Omega_R)}}{\|\rho_0^{1/2}c_0\nabla\mathbf{u}\|_{L^2(\Omega_R)}} \quad \text{and} \quad \|\mathcal{C}\| = \sup_{\mathbf{u}\in V,\,\mathbf{u}\neq 0}\frac{\|\rho_0^{1/2}c_0\mathcal{C}\mathbf{u}\|_{L^2(\Omega_R)}}{\|\rho_0^{1/2}c_0\nabla\mathbf{u}\|_{L^2(\Omega_R)}}$$

Now using (4.8), we can eliminate ψ in problem (2.12) which is then rewritten as follows:

Find $\mathbf{u} \in V$ such that:

$$\begin{aligned}
&\rho_0\frac{D_\alpha^2\mathbf{u}}{Dt^2} - \nabla_\alpha(\rho_0c_0^2\mathrm{div}_\alpha\mathbf{u}) + \mathbf{curl}_\alpha(\rho_0c_0^2(\mathrm{curl}_\alpha\mathbf{u} - \mathcal{T}\mathbf{u})) \\
&+\mathrm{div}_\alpha\mathbf{u}\nabla p_0 - {}^t\nabla_\alpha\mathbf{u}\cdot\nabla p_0 = \mathbf{f} + \mathbf{curl}_\alpha(\rho_0c_0^2\psi_{\mathbf{f}}) \qquad \text{in} \quad \Omega_L \\
&\mathrm{curl}_\alpha\mathbf{u} - \mathcal{T}\mathbf{u} = \psi_{\mathbf{f}} \qquad \text{on} \quad \Gamma_\infty^L
\end{aligned} \tag{4.10}$$

Problem (4.10) has the following variational form:

$$\text{Find } \mathbf{u} \in V \text{ such that } \forall \mathbf{v} \in V \quad a(\mathbf{u},\mathbf{v}) + b(\mathbf{u},\mathbf{v}) = \ell(v)$$

where

$$
\begin{aligned}
a(\mathbf{u},\mathbf{v}) = &\int_{\Omega_L} \frac{\rho_0}{\alpha_1\alpha_2}\left(c_0^2 \mathrm{div}_\alpha \mathbf{u}\, \mathrm{div}_\alpha \bar{\mathbf{v}} + c_0^2 \mathrm{curl}_\alpha \mathbf{u}\, \mathrm{curl}_\alpha \bar{\mathbf{v}} - (\mathbf{v}_0\cdot\nabla_\alpha)\mathbf{u}\,(\mathbf{v}_0\cdot\nabla_\alpha)\bar{\mathbf{v}}\right)\\
&-\int_{\Omega_L} \frac{\rho_0 c_0^2}{\alpha_1}\mathcal{T}\mathbf{u}\,\mathrm{curl}_\alpha \bar{\mathbf{v}}\\
b(\mathbf{u},\mathbf{v}) = &\int_{\Omega_L} \frac{-\rho_0\omega}{\alpha_1\alpha_2}(2i(\mathbf{v}_0\cdot\nabla_\alpha)\mathbf{u}+\omega\mathbf{u})\cdot\bar{\mathbf{v}} + \int_{\Omega_R}\left(\mathrm{div}\,\mathbf{u}\nabla p_0 - {}^t\nabla\mathbf{u}\cdot\nabla p_0\right)\bar{\mathbf{v}},\\
\ell(v) = &\int_{\Omega_R}\mathbf{f}\cdot\bar{\mathbf{v}} + \int_{\Omega_L}\frac{\rho_0 c_0^2}{\alpha_1}\psi_{\mathbf{f}}\,\mathrm{curl}_\alpha\bar{\mathbf{v}}
\end{aligned}
$$

Theorem 4.1. *Suppose $\mathbf{v}_0$ satisfies condition (3.13). Problem (4.10) is of Fredholm type if*

$$
\inf_{x\in\Omega_R}\left(1-\frac{|\mathbf{v}_0|^2}{c_0^2}\right) > \frac{\sqrt{2}}{K_\alpha}\frac{\beta_0}{\omega}\left(K_1\|\mathcal{B}\| + \frac{K_2\beta_0+K_3}{\omega}\|\mathcal{C}\|\right) \tag{4.11}
$$

where the constant K_i are given in (4.9) and $K_\alpha = \min(1,|\alpha|)\min\left(\Re e\,\alpha, \Re e\frac{1}{\alpha}\right)$.

Proof. Following the proof of theorem 1 of [3], we will prove that, under hypothesis (4.11), the bilinear form $a(\mathbf{u},\mathbf{v})$ has a coercive+compact decomposition on V. This proves the theorem since $b(\mathbf{u},\mathbf{v})$ is clearly compact (i.e. associated to a compact operator on V).

It is established in [3] that $\forall \mathbf{u},\mathbf{v}\in V$:

$$
\int_{\Omega_L}\frac{\rho_0 c_0^2}{\alpha_1\alpha_2}(\mathrm{div}_\alpha\mathbf{u}\,\mathrm{div}_\alpha\bar{\mathbf{v}} + \mathrm{curl}_\alpha\mathbf{u}\,\mathrm{curl}_\alpha\bar{\mathbf{v}}) = \int_{\Omega_L}\frac{\rho_0 c_0^2}{\alpha_1\alpha_2}\nabla_\alpha\mathbf{u}\cdot\nabla_\alpha\bar{\mathbf{v}} + d(\mathbf{u},\mathbf{v})
$$

with $d(\mathbf{u},\mathbf{v}) = \int_{\Omega_L}\left(\frac{\partial(\rho_0 c_0^2)}{\partial x_1}\frac{\partial\mathbf{u}}{\partial x_2} - \frac{\partial(\rho_0 c_0^2)}{\partial x_2}\frac{\partial\mathbf{u}}{\partial x_1}\right)\times\bar{\mathbf{v}} - \int_{\partial\mathcal{O}}\rho_0 c_0^2[(\mathbf{u}\cdot\nabla)\mathbf{n}\times\mathbf{n}](\mathbf{n}\times\bar{\mathbf{v}})$, so that $d(\mathbf{u},\mathbf{v})$ is compact. The theorem then requires the existence of a constant $\delta>0$ such that $\forall\mathbf{u}\in V$:

$$
\left|\int_{\Omega_L}\frac{\rho_0 c_0^2}{\alpha_1\alpha_2}\nabla_\alpha\mathbf{u}\cdot\nabla_\alpha\bar{\mathbf{u}} - \frac{\rho_0}{\alpha_1\alpha_2}(\mathbf{v}_0\cdot\nabla_\alpha)\mathbf{u}\cdot(\mathbf{v}_0\cdot\nabla_\alpha)\bar{\mathbf{u}} - \frac{\rho_0 c_0^2}{\alpha_1}\mathcal{T}\mathbf{u}\,\mathrm{curl}_\alpha\bar{\mathbf{u}}\right| \geq \delta\int_{\Omega_L}\rho_0 c_0^2|\nabla\mathbf{u}|^2
$$

The existence of $\delta>0$ is obtained under hypothesis (4.11) by using (4.9) and the following inequality:

$$
\left|\int_{\Omega_L}\frac{\rho_0}{\alpha_1\alpha_2}\left(c_0^2\nabla_\alpha\mathbf{u}\cdot\nabla_\alpha\bar{\mathbf{u}} - (\mathbf{v}_0\cdot\nabla_\alpha)\mathbf{u}\cdot(\mathbf{v}_0\cdot\nabla_\alpha)\bar{\mathbf{u}}\right)\right| \geq \min\left(\Re e\,\alpha, \Re e\frac{1}{\alpha}\right)\int_{\Omega_L}\rho_0(c_0^2-|\mathbf{v}_0|^2)|\nabla\mathbf{u}|^2
$$

□

Remark 4.1. 1. Coerciveness is obtained for small values of $\|\mathcal{B}\|$ and $\|\mathcal{C}\|$, that is for a slowly varying flow.

2. The right member of (4.11) diverges as $\omega \to 0$. This is due to the dependence versus ω of the norm (3.3) which is not appropriate at low frequency. The divergence can be easily removed by replacing ω in the definition (3.3) by some arbitrary value $\omega_0 > \omega$.

3. Estimates (4.7) can be improved by taking into account the decreasing behavior in the PMLs, leading to constants K_i, $i = 1,2,3$, depending only on ρ_0, α and v_∞, and therefore independent of ω. As β_0/ω is a decreasing function of ω, we see then that condition (4.11) is easier to satisfy when ω increases.

5 Numerical solution of the coupled problem

5.1 The numerical scheme

Numerical results for the coupled problem can be obtained by combining the Finite Element solution of (2.7) and the Discontinuous Galerkin solution of (2.8).
Let $\mathcal{M}_h$ be a triangulation [6] of the computational domain Ω_L . The construction of the approximation is based on the following approximate spaces:

$$\begin{aligned} V_h^k &:= \{\mathbf{v}_h \in V; \forall T \in \mathcal{M}_h, \mathbf{v}_{h|T} \in (\mathcal{P}^k(T))^2\} \\ W_h^k &:= \{\varphi_h \in L^2(\Omega_L); \forall T \in \mathcal{M}_h, \varphi_{h|T} \in \mathcal{P}^k(T)\} \end{aligned} \tag{5.1}$$

where $k \in \mathbb{N}^*$ and $\mathcal{P}^k(T)$ is the space of polynomial functions of total degree at most k.
Moreover, for each $T \in \mathcal{M}_h$, we denote by $\mathbf{n}_T$ the outward unit normal to ∂T and $\mathcal{F}_T^-$ is the subset of ∂T where $\mathbf{v}_0 \cdot \mathbf{n}_T < 0$. Finally, $\mathcal{F}_T^{-,i}$ is the subset of $\mathcal{F}_T^-$ corresponding to the interior faces i.e. $\forall F \in \mathcal{F}_T^{-,i}$, $F \cap \Gamma_L^- = \emptyset$.

The approximate formulation is then written as follows: find $\mathbf{u}_h \in V_h^k$ such that $\forall \mathbf{v}_h \in V_h^k$,

$$a_h(\mathbf{u}_h, \mathbf{v}_h) + b(\mathbf{u}_h, \mathbf{v}_h) = \ell_h(\mathbf{v}_h) \tag{5.2}$$

where a_h and ℓ_h are defined as the bilinear and linear forms a and ℓ (see subsection 4.2) by replacing $\mathcal{T}$ by an approximate operator $\mathcal{T}_h$ and $\psi_{\mathbf{f}}$ by an approximation $\psi_{\mathbf{f},h}$.

The discrete operator $\mathcal{T}_h$ and function $\psi_{\mathbf{f},h}$ are then defined by the successive solution of the following first order problem:

$$\left\{ \begin{array}{rcl} \dfrac{D_\alpha \psi}{Dt} & = & g \text{ in } \Omega_L \\ \psi & = & 0 \text{ on } \Gamma_L^- \end{array} \right. \tag{5.3}$$

by using a classical discontinuous Galerkin method [7] well-adapted to take into account transport phenomena: find $\psi_h \in W_h^k$ such that $\forall \phi_h \in W_h^k$ and $\forall T \in \mathcal{M}_h$,

$$a_{DG,T}(\psi_h, \phi_h) \quad = \quad \ell_{T,g}(\phi_h) \tag{5.4}$$

where

$$a_{DG,T}(\psi_h,\phi_h) = -i\omega\int_T \frac{1}{\alpha_1\alpha_2}\psi_h\bar{\phi}_h dx + \int_T \frac{1}{\alpha_1\alpha_2}\mathbf{v}_0\cdot\nabla_\alpha\psi_h\bar{\phi}_h dx + \sum_{E=T\cap\Gamma_L^-}\int_E |\mathbf{v}_0\cdot\mathbf{n}_{T,\alpha}|\psi_{h|T}\bar{\phi}_h d\sigma$$
$$- \sum_{E=T\cap T'}\frac{1}{2}\int_{E\cap\mathcal{F}_T^{-,i}}((\mathbf{v}_{0|T'}\cdot\mathbf{n}_{T',\alpha})\psi_{h|T'} + (\mathbf{v}_{0|T}\cdot\mathbf{n}_{T,\alpha})\psi_{h|T})\bar{\phi}_h d\sigma$$

with $\mathbf{n}_{T,\alpha} = \dfrac{1}{\alpha_{1|T}\alpha_{2|T}}(\alpha_{1|T}n_T^1, \alpha_{2|T}n_T^2)^T$ and $\ell_{T,g}(\phi_h) = \displaystyle\int_T \frac{1}{\alpha_1\alpha_2} g\bar{\phi}_h dx$.

Now, we define the operator $\mathcal{D}_h$ from $L^2(\Omega_L)$ into W_h^k by $\mathcal{D}_h g = \psi_h$ where ψ_h is the solution of (5.4). The operator $\mathcal{T}_h$ is then constructed in the following way:

$$\mathcal{T}_h \quad = \quad \mathcal{D}_h(-2\mathcal{B}\mathbf{u}_h) + \mathcal{D}_h\circ\mathcal{D}_h(-\mathcal{C}\mathbf{u}_h) \tag{5.5}$$

Finally, $\psi_{\mathbf{f},h}$ is defined by $\mathcal{D}_h\circ\mathcal{D}_h\left(\dfrac{\mathrm{curl}\mathbf{f}}{\rho_0}\right)$.

5.2 Numerical results

We present here a numerical result where the depth function h introduced in subsection 2.1 is defined by:

$$h(x_1)\left(x_1^2 + (h(x_1)+b)^2 - a^2\right) = a^2 b$$

and the mean flow is a potential incompressible flow (with a constant density $\rho_0 = \rho_\infty$) given by:

$$\mathbf{v}_0(x_1,x_2) = v_\infty \begin{pmatrix} 1 + \dfrac{a^2}{x_1^2+(x_2+b)^2} - \dfrac{2a^2x_1^2}{(x_1^2+(x_2+b)^2)^2} \\ \dfrac{-2a^2x_1(x_2+b)}{(x_1^2+(x_2+b)^2)^2} \end{pmatrix} \quad \text{and} \quad \nabla p_0 = -\frac{1}{2}\rho_\infty\nabla\left(|\mathbf{v}_0|^2\right)$$

where a and b are strictly positive constants such that $a < R$. This is the potential flow around a cylinder of center $(0,-b)$ and radius a, and the definition of h is such that Γ_∞ is a stream line of the flow located strictly above the cylinder. As a consequence, $\mathbf{v}\cdot\mathbf{e}_1$ does not vanish and condition (3.13) is fulfilled.

Notice that the hypotheses of uniformity of the geometry ($h(x_1) = 0$ for $|x_1| > r$) and of the flow ($\mathbf{v}_0(x) = v_\infty$ for $x_1^2 + x_2^2 > R^2$) are not satisfied for finite values of r and R, but the variations of h and $\mathbf{v}_0$ will be supposed negligible far enough. Also this incompressible flow does not satisfy the state law (2.3), except asymptotically for $c_0 \to +\infty$, but we take advantage of its analytical expression (similar expressions for compressible potential flows do not exist). Here we take:

$$a = 0.5, \quad b = 0.1, \quad v_\infty = 0.4c_0, \quad R = 3, \quad L = 1, \quad \alpha = 0.65(1-i)$$

The frequency is such that $\frac{\omega}{c_0} = \frac{4\pi}{3}$ and we consider a source term f of the following form:

$$f = \mu \nabla \varphi + \beta \mathbf{curl}\, \varphi \quad \text{where} \quad \varphi(x_1, x_2) = e^{-\left((x_1 - x_1^s)^2 + (x_2 - x_2^s)^2\right)/r_s^2}$$

Here we take:

$$\mu = 100, \quad \beta = 3, \quad x_1^s = -1.5, \quad x_2^s = 1.05, \quad r^s = 0.3$$

Moreover, we have neglected the term $\psi_{\mathcal{C}}$ in first approximation to evaluate the proposed method.

Below are the isovalues of the real part of both components of $\mathbf{u}$ in the domain Ω_R. One can observe two kinds of structures, corresponding to acoustic and hydrodynamic

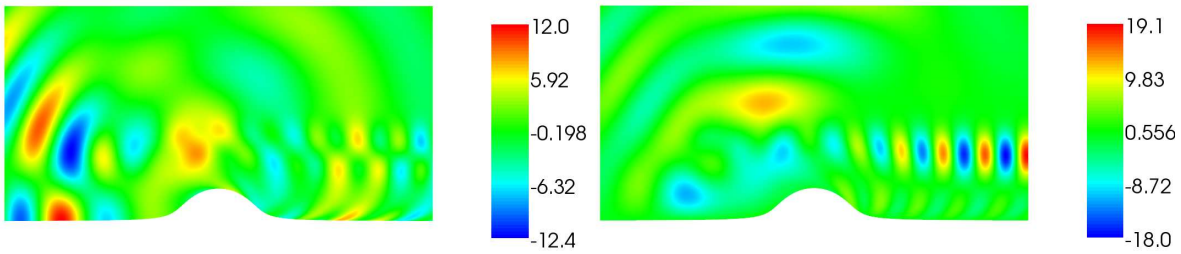


Figure 3: Real part of u_1 and u_2

phenomena. The acoustic wave is particularly visible upstream of the source and partially hidden downstream by the vortices. These ones are mainly produced by the source and convected along the stream lines of the flow, but we can also notice some vortices generated by the perturbed part of the mean flow (where $\mathcal{B}\mathbf{u}$ and $\mathcal{C}\mathbf{u}$ take significant values) and convected along the rigid boundary. This interpretation is confirmed by the representation of ψ_f and $\mathcal{T}u$ below:

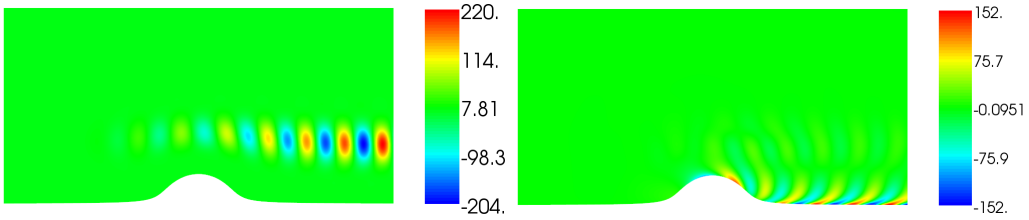


Figure 4: Real part of ψ_f and $\mathcal{T}u$

Let us recall that we have solved the augmented equation (2.7) instead of (2.4). Equivalence is achieved if $\mathrm{curl}\,\mathbf{u} = \psi$, which can be checked a posteriori as illustrated by the following results (note the different scales):

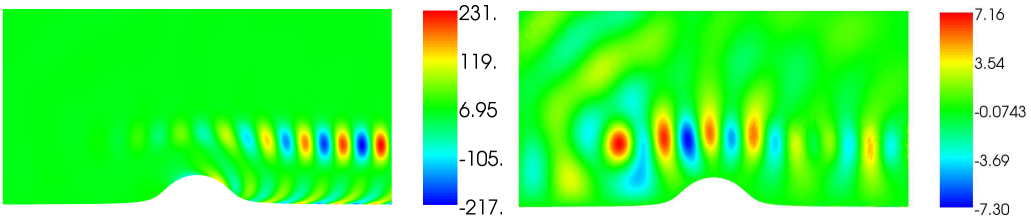


Figure 5: Real part of curl$\mathbf{u}$ and of curl$\mathbf{u} - \psi$

References

[1] P. Azerad, *Analyse des équations de Navier-Stokes en bassin peu profond et de l'équation de transport*, PhD thesis, Neuchâtel, 1996.

[2] E. Bécache, A.-S. Bonnet-Ben Dhia, and G. Legendre, *Perfectly matched layers for time-harmonic acoustics in the presence of a uniform flow*, SIAM J. Numer. Anal., 44, pp. 1191-1217, 2006.

[3] A. S. Bonnet-Ben Dhia, J. F. Mercier, F. Millot and S. Pernet, *A low Mach model for time harmonic acoustics in arbitrary flows*, J. of Comp. and Appl. Math., vol. 234(6), pp. 1868-1875, 2010.

[4] A. S. Bonnet-Ben Dhia, E. M. Duclairoir, G. Legendre and J. F. Mercier, *Time-harmonic acoustic propagation in the presence of a shear flow*, J. of Comp. and App. Math., vol. 204(2), pp. 428-439, 2007.

[5] A. S. Bonnet-Ben Dhia, E. M. Duclairoir and J. F. Mercier, *Acoustic propagation in a flow: numerical simulation of the time-harmonic regime*, ESAIM Procs., vol. 22, pp. 1-14, 2007.

[6] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, Series *Studies in Mathematics and its Applications*, North-Holland, Amsterdam, 1978.

[7] A. Ern and J.-L. Guermond, *Theory and practice of finite elements*, Applied Mathematical Sciences, Springer-Verlag, New York, 2004.

[8] G. Legendre, *Rayonnement acoustique dans un fluide en écoulement : analyse mathématique et numérique de l'équation de Galbrun*, PhD thesis, Paris VI University, 2003.

[9] F. Treyssede, G. Gabard, and M. B. Tahar, *A mixed finite element method for acoustic wave propagation in moving fluids based on an Eulerian-Lagrangian description*, JASA, 113, pp. 705-716, 2003.