BE SXS-Galerkin Discontinu

Pierre RAYNAL, Paulin HUGUET, Jean-Pierre MANSOUR

Novembre 2023

Introduction

L'objectif de ce BE est de résoudre le système des équations de Maxwell dans un domaine bidimensionnel $\Omega \subset \mathbb{R}^2$.

Plus précisément, nous cherchons à déterminer $\mathbf{E}:\Omega\times[0,T]\to\mathbb{R}^2$ et $H:\Omega\times[0,T]\to\mathbb{R}$ solution de

$$\varepsilon(x)\frac{\partial \mathbf{E}}{\partial t} - \mathbf{curl}H(x,t) = \mathbf{J}(x,t),$$

$$\mu(x)\frac{\partial H}{\partial t}(x,t) + \mathbf{curl}\mathbf{E}(x,t) = 0,$$

$$\mathbf{E}(x,0) = \mathbf{E_0}(x),$$

$$H(x,0) = H_0(x),$$

avec:

•
$$\operatorname{curl} H = \begin{pmatrix} \frac{\partial H}{\partial y} \\ -\frac{\partial H}{\partial x} \end{pmatrix}$$

• curl**E**=
$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}$$

1 Analyse Numérique de la méthode de Galerkin Discontinu pour un problème de cavité

Dans un premier temps, nous fermons le problème en imposant sur la frontière Γ une condition de type métal parfait avec $\mathbf{n} \times \mathbf{E} = 0$ avec \mathbf{n} la normale unitaire sortante du domaine Ω .

Q1 : On construit un maillage fini du domaine Ω que l'on notera τ_h .

On travaillera sur les espaces $\mathbf{V}_{h,k}^d = \{\mathbf{v}_h \in [\mathbf{L}^2(\Omega)]^d \text{ telles que } \forall T \in \tau_h, \mathbf{v}_{h|} \in [P^k(T)]^d\}$, avec $P^k(T)$ l'espace des polynômes de degré au plus égal à $k \in \mathbb{N}$.

On travaille alors sur le système équivalent en formulation faible avec $\forall (\mathbf{v}_h, w_h) \in (\mathbf{V}_{h,k}^2, \mathbf{V}_{h,k}^1)$ et $\forall T \in \tau_h$:

$$\frac{d}{dt} \int_{T} \varepsilon \mathbf{E}_{h} \cdot \mathbf{v}_{h} dx - \int_{T} \operatorname{curl} H_{h} \cdot \mathbf{v}_{h} dx = \int_{T} \mathbf{J} \cdot \mathbf{v}_{h} dx ,$$

$$\frac{d}{dt} \int_{T} \mu H_{h} w_{h} dx + \int_{T} \operatorname{curl} \mathbf{E}_{h} w_{h} dx = 0 .$$

En utilisant les techniques d'intégration par parties sur les équations de notre système, on obtient alors :

$$\frac{d}{dt} \int_{T} \varepsilon \mathbf{E}_{h} \cdot \mathbf{v}_{h} dx - \int_{T} \operatorname{curl} H_{h} \cdot \mathbf{v}_{h} dx + \frac{1}{2} \int_{\partial T} |[\mathbf{n} \times \mathbf{H}_{h}]| \cdot \mathbf{v}_{h} d\gamma = \int_{T} \mathbf{J} \cdot \mathbf{v}_{h} dx}{\left[\frac{d}{dt} \int_{T} \mu H_{h} w_{h} dx + \int_{T} \mathbf{E}_{h} \cdot \operatorname{curl} w_{h} dx - \int_{\partial T} \{\{\mathbf{E}_{h}\}\} \cdot (\mathbf{n}_{T} \cdot w_{h}) d\gamma = 0\right]}.$$

avec, sur F la frontière entre T et T':

- $|[\mathbf{n} \times z]|_F = \mathbf{n}_T \times z_T + \mathbf{n}_{T'} \times z_{T'}$,
- $\{\{z\}\}_F = \frac{1}{2}(z_T + z_{T'})$

On a ainsi ici un schéma centré.

On étudiera aussi le **Schéma décentré** obtenu en ajoutant les termes de pénalisation pondérées par α_E , $\alpha_H \ge 0$:

$$\frac{d}{dt} \int_{T} \varepsilon \mathbf{E}_{h} \cdot \mathbf{v}_{h} dx - \int_{T} \operatorname{curl} H_{h} \cdot \mathbf{v}_{h} dx + \frac{1}{2} \int_{\partial T} |[\mathbf{n} \times \mathbf{H}_{h}]| \cdot \mathbf{v}_{h} d\gamma + \alpha_{E} \int_{\partial T} |[\mathbf{n} \times \mathbf{E}_{h}]| (\mathbf{n} \times \mathbf{v}_{h}) d\gamma = \int_{T} \mathbf{J} \cdot \mathbf{v}_{h} dx}{\int_{T} \mu H_{h} w_{h} dx + \int_{T} \mathbf{E}_{h} \cdot \operatorname{curl} w_{h} dx - \int_{\partial T} \{\{\mathbf{E}_{h}\}\} \cdot (\mathbf{n}_{T} \cdot w_{h}) d\gamma + \alpha_{H} \int_{\partial T} |[\mathbf{n} \times \mathbf{H}_{h}]| (\mathbf{n} \times \mathbf{w}_{h}) d\gamma = 0}.$$

Ces deux schémas construits correspondent bien à la formulation de Galerkin discontinue dans le cas $\mathbf{n} \times \mathbf{E} = 0$.

En notant $E \in \mathbb{R}^{\dim(\mathbf{V}_{h,k}^2)}$ et $H \in \mathbb{R}^{\dim(\mathbf{V}_{h,k}^1)}$, les vecteurs contenant respectivement les degrés de liberté des champs \mathbf{E}_h et H_h , on peut alors réécrire la formulation générale de Galerkin discontinu sous la forme matricielle suivant :

$$M^{E} \frac{dE}{dt} - R^{E}H + P^{E}E = F$$
,
$$M^{H} \frac{dH}{dt} + R^{H}E + P^{H}E = 0$$
,
$$E(0) = E_{0} \text{ et } H(0) = H_{0},$$

avec:

• $M^E \in \mathbb{R}^{\dim(\mathbf{V}_{h,k}^2) \times (\mathbf{V}_{h,k}^2)}$ et $M^H \in \mathbb{R}^{\dim(\mathbf{V}_{h,k}^1) \times (\mathbf{V}_{h,k}^1)}$, les matrices de masses "diagonales" respectives pour les deux équations.

$$M^E = \begin{pmatrix} M^{E,1} & & & & \\ & M^{E,2} & & & \\ & & \ddots & & \\ & & & M^{E,N} \end{pmatrix} \text{ et } M^H = \begin{pmatrix} M^{H,1} & & & & \\ & M^{H,2} & & & \\ & & & \ddots & & \\ & & & & M^{H,N} \end{pmatrix}, \text{ avec N le nom-}$$

bre d'éléments du maillage, et $\forall i \in |[1,N]|, M_{k,l}^{E,i} = \int_{T_i} \varepsilon(x) v_l^i(x) v_k^i(x) dx$ produit scalaire des composantes d'une base de $\mathbf{V}_{h,k}^2$ adapté à l'élement du maillage T_i . Et de même $M_{k,l}^{H,i} = \int_{T_i} \mu(x) w_l^i(x) w_k^i(x) dx$ sur $\mathbf{V}_{h,k}^1$.

- $P^E \in \mathbb{R}^{\dim(\mathbf{V}_{h,k}^2) \times (\mathbf{V}_{h,k}^2)}$ et $P^H \in \mathbb{R}^{\dim(\mathbf{V}_{h,k}^1) \times (\mathbf{V}_{h,k}^1)}$ sont des matrices de pénalisation, construites sur un modèle permettant d'ajouter des effets dissipants sur les frontières des éléments.
- $R^E \in \mathbb{R}^{\dim(\mathbf{V}_{h,k}^2) \times (\mathbf{V}_{h,k}^1)}$ et $R^H = (R^E)^T \in \mathbb{R}^{\dim(\mathbf{V}_{h,k}^1) \times (\mathbf{V}_{h,k}^2)}$ sont les matrices de rigidité, composées de termes diagonaux relatifs aux effets sur le domaine ouvert de l'élément, et d'autres éléments "quasi-diagonaux" incorporant les effets sur les frontières des éléments.

Q2 : Dans le cas de la discrétisation temporelle du système, on va utiliser un schéma de type saute-mouton ou leap-frog qui peut se résumer ainsi :

$$\boxed{ M^E \frac{E^{n+1} - E^n}{\Delta t} - R^E H^{n+1/2} + P^E E^n = F^{n+1/2} }, \\ \boxed{ M^H \frac{H^{n+1/2} - H^{n-1/2}}{\Delta t} - R^H E^n + P^H H^n = 0 },$$

Avec $E^n = E(t_n), H^{n+1/2} = H(t_{n+1/2})$ et $t_i = i\Delta t$.

Dans le cas de ce système, E et H ne sont pas calculés au même moment. L'intialisation de E^0 reste la même que précédemment, mais pour $H^{1/2}$, on est obligé de passer par un développement de Taylor en t=0 à l'ordre 1 :

$$H^{1/2} = H^0 + \frac{\Delta t}{2} \frac{dH^0}{dt} + O(\Delta t^2)$$
$$\approx H^0 - \frac{\Delta t}{2} (M^H)^{-1} R^H E^0$$

Définission maintenant l'énergie du système discrétisé au temps n par :

$$\varepsilon_h^n = (M^E E^n, E^n) + (M^H H^{n+1/2}, H^{n-1/2}).$$
(1)

Si on se place dans le cas où F=0 et sans pénalité, on a alors le système qui devient :

$$M^{E}(E^{n+1} - E^{n}) = \Delta t R^{E} H^{n+1/2},$$

$$M^{H}(H^{n+1/2} - H^{n-1/2}) = -\Delta t (R^{E})^{T} E^{n}$$

Ce qui donne alors :

$$\begin{split} (E^{n+1})^T M^E (E^{n+1} - E^n) &= (E^{n+1})^T \Delta t R^E H^{n+1/2} \,, \\ (H^{n+1/2})^T M^H (H^{n+3/2} - H^{n+1/2}) &= -(H^{n+1/2})^T \Delta t (R^E)^T E^{n+1} = -(E^{n+1})^T \Delta t R^E H^{n+1/2} \,, \end{split}$$

et:

$$\begin{split} (E^n)^T M^E (E^{n+1} - E^n) &= (E^n)^T \Delta t R^E H^{n+1/2} \,, \\ (H^{n+1/2})^T M^H (H^{n+1/2} - H^{n-1/2}) &= -(H^{n+1/2})^T \Delta t (R^E)^T E^n = -(E^n)^T \Delta t R^E H^{n+1/2} \,, \end{split}$$

En sommant alors les deux sets d'équations, on obtient alors :

$$(M^{E}E^{n+1}, E^{n+1}) - (M^{E}E^{n}, E^{n+1}) + (M^{H}H^{n+3/2}, H^{n+1/2}) - (M^{H}H^{n+1/2}, H^{n+1/2}) = 0,$$

$$(M^{E}E^{n+1}, E^{n}) - (M^{E}E^{n}, E^{n+1}) + (M^{H}H^{n+1/2}, H^{n+1/2}) - (M^{H}H^{n-1/2}, H^{n+1/2}) = 0,$$

$$(M^{E}E^{n}, E^{n+1}) - (M^{E}E^{n}, E^{n+1}) + (M^{H}H^{n+1/2}, H^{n+1/2}) - (M^{H}H^{n-1/2}, H^{n+1/2}) = 0,$$

Ce qui donne ultimement en sommant les deux dernières équations :

$$\varepsilon_h^{n+1} - \varepsilon_h^n = 0 \,,$$

Et ce $\forall n \in \mathbb{N}$.

L'énergie discrétisée est donc bien conservée dans le cas sans source et sans pénalité (centré).

Néanmoins, dans ce cas, l'énergie discrétisée n'est pas forcément positive, ce qui vient imposer une contrainte supplémentaire sur le temps :

$$\begin{split} \varepsilon_h^n &= (M^E E^n, E^n) + (M^H H^{n+1/2}, H^{n-1/2}) \\ &= \underbrace{(M^E E^n, E^n)}_{||E^n||^2 \geq 0} + \underbrace{(M^H H^{n-1/2}, H^{n-1/2}) \geq 0}_{||H^{n-1/2}||^2} - \Delta t (R^H E^n, V^{n-1/2}) \\ &= ||E^n||^2 + ||H^{n-1/2}||^2 - \Delta t (R^T (M^E)^{-1/2} (M^E)^{1/2} E^n, (M^H)^{-1/2} (M^H)^{1/2} H^{n-1/2}) \\ &\geq ||E^n||^2 + ||H^{n-1/2}||^2 - \Delta t ||(M^E)^{-1/2} R^H (M^H)^{-1/2} || ||E^n|| ||H^{n-1/2}|| \\ &\geq (1 - \frac{\Delta t}{2} ||(M^E)^{-1/2} R^H (M^H)^{-1/2} ||) (||E^n||^2 + ||H^{n-1/2}||^2) \,, \end{split}$$

Et donc si $\Delta t < \frac{2}{||(M^E)^{-1/2}R^H(M^H)^{-1/2}||}$, l'énergie est donc bien positive et le système est bien stable ici.

Q3.1: Nous allons tout d'abord étudier l'algorithme de résolution sur un exemple simple correspondant à $\Omega =]0,1[\times]0,1[$, avec $\varepsilon = \varepsilon_0$ permittié diélectrique du vide, et $\mu = \mu_0$, sans aucun mouvement de charge J = 0, et avec les conditions initiales suivantes $(m \in \mathbb{N}*)$:

$$\mathbf{E_0}(x,y) = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$H_0(x,y) = -\cos(m\pi x)\cos(m\pi y).$$

Ré-injecter les champs proposés dans l'énoncé, dans les différentes équations du système, permet de voir que la solution suivante convient bien au système de propagation d'une onde électromagnétique dans le vide :

$$\mathbf{E}(x, y, t) = \frac{m\pi}{\omega\varepsilon_0} sin(\omega t) \begin{pmatrix} cos(m\pi x) sin(m\pi y) \\ -sin(m\pi x) cos(m\pi y) \end{pmatrix}$$
$$H(x, y, t) = -cos(\omega t) cos(m\pi x) cos(my) ,$$

avec $\omega = c_0 \pi m \sqrt{2}$ et $c_0 = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}$.

Q3.3 : (a) Numériquement, on va d'abord travailler sur un schéma centré sans pénalisation avec $\alpha_E = \alpha_H = 0$ avec le maillage cavity pas01.mat.

On impose une CFL issue de l'analyse précédente de la forme suivante :

$$CFL = \frac{c_0}{h_{min}} \frac{2}{||(M^H)^{-1/2}R^H(M^E)^{-1/2}||},$$

avec h_{min} de l'ordre du plus petit diamètre du cercle circonscrit aux triangles du maillage.

Pour les ordres d'approximation entre k=1 et k=3, on obtient :

• Pour k=1, on a une CFL=0.4367 très grande, donnant un schéma très rapide à calculer mais un résultat très discontinu et peu satisfaisant pour une simulation d'une fonction continue.

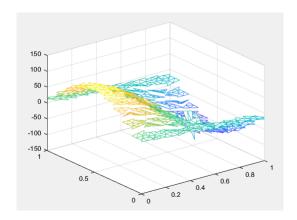


Figure 1: Résultat E_x final pour k=1 ($T_f=10^{-9}$) (schéma centré)

• Pour k=2, on a une CFL=0.2421, bien plus petite mais donnant un résultat "lissé" bien plus satisfaisant pour une approximation d'un champ continu

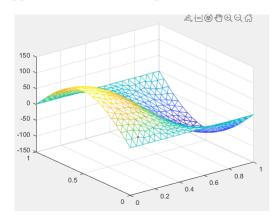


Figure 2: Résultat E_x final pour k=2 ($T_f=10^{-9}$) (schéma centré)

• Pour k=3, on a une CFL=0.1547, toujours plus petite et donannt des résultats s'approchant toujours plus du champ continu, mais au prix d'un plus grand nombre d'opérations (le pas temporel étant bien plus petit)

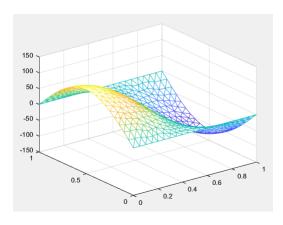


Figure 3: Résultat E_x final pour k=3 ($T_f=10^{-9}$) (schéma centré)

De tout cela, on en déduit que pour un même maillage, la CFL évolue approximativement de manière proportionnelle à $\frac{1}{k}$.

(b) Maintenant, on vient rajouter des valeurs de pénalité correspondant aux flux upwind avec $\alpha_E = \sqrt{\frac{\mu_0}{\varepsilon_0}}$ et $\alpha_H = \alpha_E^{-1}$.

En utilisant le même maillage que précédemment, et la même CFL, même pour un temps de simulation très petit $(T_f = 10^{-9})$, on observe très rapidement une divergence violente des valeurs de notre schéma $(10^9 \text{ contre } 10^2 \text{ obtenu précédemment})$.

En effet, avec ces termes de pénalité, on n'est plus dans le cas du schéma centré, la CFL utilisée calculée sur cette base n'est plus valide.

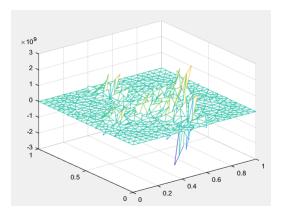


Figure 4: Divergence de la simulation $(T_f = 10^{-9})$ pour E_x avec CFL du schéma centré mais matrices de pénalisation

Ce résultat était donc assez prévisible.

(c) En divisant la CFL précédante par 4, on prend en compte le passage à un ordre différent d'approximation causé par les matrices de pénalisation. On retrouve ainsi une simulation qui converge bien vers le résultat souhaité sans divergence.

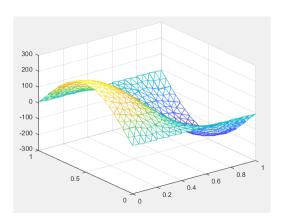


Figure 5: Résultat final E_x ($T_f = 10^{-9}$) avec CFL mise à jour et matrices de pénalisation non nulles

(d) On compare maintenant les ordres de convergence en espace de notre schéma simulé dans les cas avec et sans matrices de pénalisation.

Afin d'avoir une erreur de discrétisation en temps qui permet de rendre l'ordre étudié. On impose ainsi une condition sur les pas de temps :

$$\Delta t = \frac{1}{4} \text{ CFL } \frac{h_{min}^{\beta}}{c_0}.$$

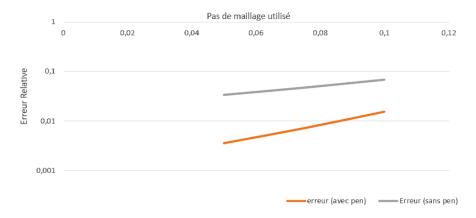


Figure 6: Erreurs relatives dans le cas k=1, m=1, β =2, pour un temps final $T_f = 10^{-9}$

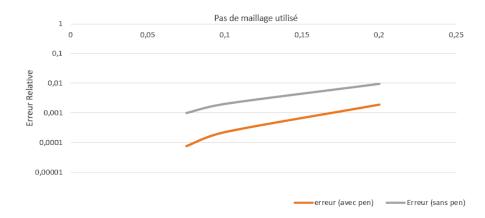


Figure 7: Erreurs relatives dans le cas k=2, m=1, β =3, pour un temps final $T_f = 10^{-9}$

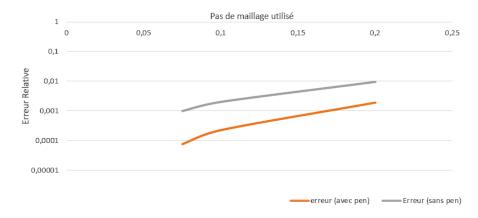


Figure 8: Erreurs relatives dans le cas k=3, m=2, β =4, pour un temps final $T_f = 10^{-10}$

On observe en effet que les ordres de convergence évoluent de manière très similaire par rapport au pas de temps pour les cas avec et sans pénalité.

On remarque aussi, qu'augmenter l'ordre d'approximation de k=2 à k=3 n'a que très peu d'effet sur l'erreur relative contrairement au passage de k=1 à k=2.

(e) On trace ensuite l'évolution de cette erreur relative (par rapport à la solution réelle de notre problème) en fonction de $\sqrt{N_{tri} N}$ avec N_{tri} le nombre de triangle du maillage utilisé, et donc $N_{tri} * N$

représentant le nombre de degrés de liberté de notre schéma, et permettant de comparer les différentes approximations k entre elles ici.

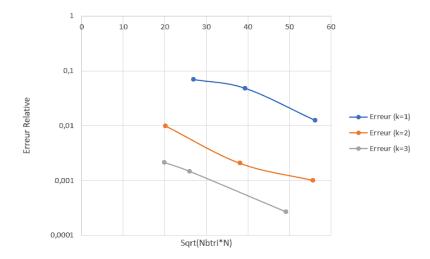


Figure 9: Erreurs relatives en fonction de la racine du nombre de degrés de libertés de notre schéma (pour m=1)

2 Simulation en domaine non-borné

Q4: On peut reprendre exactement la même approche faite pour la question 1. La seule véritable différence vient du fait que lorsque l'on intègre par partie le terme $\int_T curl E_h w_h$, l'on obtient :

$$\int_{T} curl E_h \cdot w_h = \int_{T} E_h curl w_h + \int_{\partial T} (n \times E_h) \cdot w_h$$

Puis en remplaçant $n \times E$ par $-\sqrt{\mu/\epsilon}H$ sur Γ_{SM} , on obtient bien la formulation variationnelle attendue. On peut ensuite lui ajouter des termes de décentrement pour obtenir les formulations (13a) et (13b).

 ${f Q5}$: Voici après simulation une visualisation de l'état de la composante E_y en temps $T_f=2E-9$

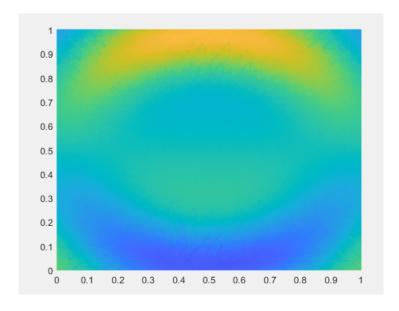


Figure 10: Etat de E_y en temps $T_f=2E-9$

On constate que les conditions de Sliver-Müller sont une bonnes approximations d'une transparence à l'infini. C'est à dire qu'avec ces conditions, le bord du domaine Ω est en quelque sorte "effacé" et l'onde continue à se propager comme si elle vivait en domaine non borné. C'est évidemment très pratique car les conditions de Sliver-Müller permettent de simuler avec une bonne précision un comportement non borné, quelque chose qui n'est en pratique pas réalisable.

 ${f Q6}$: Le but de cette question est de représenté un signal électromagnétique qui subirait un changement discontinu de milieu. On lance donc le programme comme demandé et voici ce que l'on obtient pour les composantes du champ electrique E:

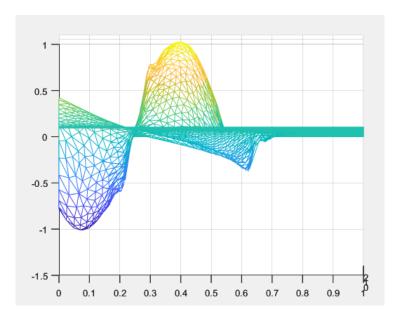


Figure 11: E_x en temps $t_f = 1E - 9$

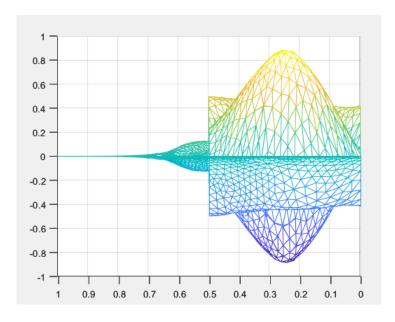


Figure 12: E_y en temps $t_f = 1E - 9$

Un résultat frappant est que la composante E_x reste continue contrairement à E_y . Ceci peut paraître étonnant mais pas tant que ça en réalité. En effet, la physique nous enseigne que au passage d'une discontinuité de milieu, le champ éléctrique ne retient sa continuité uniquement dans la direction

tangentielle au changement de milieu. On peut également l'expliquer mathématiquement :

Appelons Ω_1 (resp. Ω_2) le sous-domaine de Ω dans lequel $\epsilon = \epsilon_0$ (resp. $\epsilon = 4\epsilon_0$), ainsi que Γ la frontière les séparants. Les équations nous disent que $E \in H(curl)$, et ceci est équivalent (d'après le cours) à ce que :

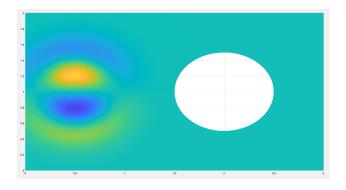
$$n \times E_1 = n \times E_2 \operatorname{sur} \Gamma$$

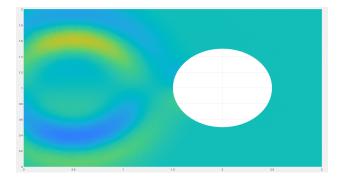
(où E_1 et E_2 sont le champ E restreint aux deux sous domaines) ainsi que $E_1 \in H(\Omega_1, curl)$ et $E_2 \in H(\Omega_2, curl)$. Dans ce cas précis, ceci revient à dire que $(E_x)_1 = (E_x)_2$ d'où la continuité de E_x mais sans plus de contrainte sur E_y .

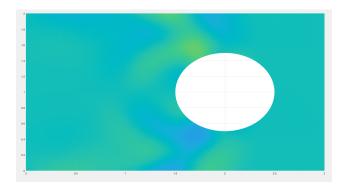
On peut d'ores et déjà remarquer un avantage de la méthode GD : en relaxant les continuités, on autorise notre solution à atteindre des espaces plus généraux qui dans ce cas-ci sont nécessaires pour atteindre la solution physique. On imagine qu'une méthode plus restrictive sur la continuité de ses fonctions donnerait des résultats aberrants dans cette situation.

 $\mathbf{Q7}$: On ne souhaite aucune pénétration du champ éléctrique à l'intérieur du disque. Notons Γ le bord extérieur du domaine (à l'infini) et C le cercle délimitant le disque unité. On a donc $\partial\Omega = \Gamma \cup C$. Pour éviter les perturbations dûes au bord du domaine et d'éventuelles réflexions non désirables on choisit des conditions SM sur Γ , pour prendre en compte la conductivité parfaite du disque, on impose au champ E de vérifier $n \times E = 0$ sur C (condition de type métal parfait).

On peut donc reprendre le système (11) en remplaçant Γ_{SM} par Γ ainsi que Γ_D par C pour obtenir un modèle physique probant. L'approximation GD qui en découle est donc simplement celle qui nous vient de (13a) et (13b). Voici les simulations que l'on obtient :







On voit que l'onde traverse le bord Γ sans perturbation et vient se réfléchir sur C (c'est particulièrement visible sur la troisième image). Il s'agit exactement de ce que l'on pourrait attendre d'un conducteur parfait soumis à un champ électrique.

Q8 : Le but de ce BE a été d'étudier divers aspects (numériques et qualitatifs) du schéma de Galerkin Discontinu. On propose de lister ici quelques avantages notables de l'utilisation de cette méthode.

Tout d'abord, il faut remarquer que cette méthode confère l'avantage d'avoir des matrices de masses relativement simples à inverser car ces dernières sont diagonalisables par blocs (ce qui est agréable, lorsque l'on compare par exemple à une méthode d'éléments finis). Autre chose dans le même registre : les maillages de la méthode GD ne requiert pas de conformité, ce qui peut simplifier grandement les affinages de maillages.

Il est également quantitativement fonctionnel car comme vu en question 2, ce schéma fait décroitre l'énergie L^2 ce qui nous assure une bonne stabilité.

On peut à présent mentionner l'essentiel : la possibilité de considérer des fonctions discontinues dans nos espaces d'approximations. A première vue, ceci pourrait donner l'impression que l'on relaxe artificiellement les continuités simplement pour un confort mathématique. Comme vu en question 6, il n'en ai en réalité rien car certaines situations physiques régits par les équations de Maxwell (qui forment l'objet de ce BE) requiert cette relaxation des régularités.