



**HAL**  
open science

# Estimateurs d'erreur a posteriori et critères de convergence en homogénéisation numérique par FFT

Renaud Ferrier, C Bellis

► **To cite this version:**

Renaud Ferrier, C Bellis. Estimateurs d'erreur a posteriori et critères de convergence en homogénéisation numérique par FFT. 15ème Colloque National en Calcul des Structures, May 2022, Hyères, France. hal-03681031

**HAL Id: hal-03681031**

**<https://hal.science/hal-03681031>**

Submitted on 30 May 2022

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

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

# Estimateurs d'erreur *a posteriori* et critères de convergence en homogénéisation numérique par FFT

R. Ferrier<sup>1</sup>, C. Bellis<sup>2</sup>

<sup>1</sup> Mines Saint-Etienne, Univ Lyon, CNRS, UMR 5307 LGF, Saint-Etienne, renaud.ferrier@emse.fr

<sup>2</sup> Aix-Marseille Univ, CNRS, Centrale Marseille, LMA UMR 7031, Marseille, bellis@lma.cnrs-mrs.fr

**Résumé** — Un critère d'arrêt pour les schémas itératifs d'homogénéisation périodique par FFT est proposé et évalué numériquement. Ce critère est basé sur la distinction et la comparaison entre les erreurs de discrétisation et de convergence. Des estimateurs de ces erreurs sont proposés et leurs performances sont comparées sur des cas tests 2D. On évalue ensuite un critère de convergence basé sur l'un de ces estimateurs d'erreur de discrétisation et on montre que cela permet d'aboutir à une erreur totale mieux contrôlée qu'un critère basé uniquement sur le résidu du schéma itératif considéré.

**Mots clés** — homogénéisation numérique, estimateurs d'erreur, Transformée de Fourier Rapide, filtre médian.

## 1 Introduction

### 1.1 Contexte et motivations

L'homogénéisation est une approche classique permettant une modélisation macroscopique et des calculs numériques efficaces pour des domaines de grandes tailles présentant des hétérogénéités microscopiques. Il s'agit de pré-calculer un modèle matériau homogène équivalent macroscopiquement à partir de la description de la microstructure et des propriétés physiques des différentes phases le composant aux petites échelles. Dans le cadre de l'homogénéisation périodique, le calcul de ce modèle effectif peut se faire via la résolution d'une équation intégrale de convolution en volume appelée équation de Lippmann-Schwinger périodique.

Dans [7], les auteurs proposent une discrétisation de l'équation de Lippmann-Schwinger s'appuyant sur la Transformée de Fourier Rapide (FFT), et utilisent un algorithme du point fixe pour résoudre le système (linéaire ou non) résultant. Depuis, de nombreuses variantes de cette méthode ont été proposées, soit en modifiant la méthode de discrétisation [11, 2, 10], soit en utilisant d'autres algorithmes itératifs [3, 12, 4], tels que des méthodes de descente de gradient.

Toutes ces variantes de la méthode d'homogénéisation par FFT reposent sur une approche itérative associée à une discrétisation de l'équation de Lippmann-Schwinger. Dans ce contexte, l'indicateur utilisé le plus couramment pour quantifier la convergence de la solution ne s'appuie que sur le résidu associé au schéma, c'est à dire uniquement sur un estimateur de l'erreur d'itération. Ainsi, l'objectif de cette étude est de proposer un cadre dans lequel utiliser conjointement des estimateurs de l'erreur d'itération et de l'erreur de discrétisation, pour aboutir à un contrôle de l'erreur totale sur le paramètre matériau effectif que l'on cherche à calculer. Cette démarche s'inspire notamment de ce qui a été proposé dans [9] dans le contexte de la décomposition de domaine.

Pour l'estimation de l'erreur de discrétisation, on s'est inspiré de deux familles de méthodes utilisées en éléments finis. Il s'agit d'une part de la méthode ZZ1 [13], que l'on peut rapprocher, dans le contexte des approches FFT, de méthodes de filtrage utilisées en traitement d'image, et d'autre part de la méthode de l'Erreur en Relation de Comportement [5], pour laquelle une première tentative d'adaptation à l'homogénéisation par FFT a été proposée dans [1].

### 1.2 Le problème et sa discrétisation

Soit un milieu périodique caractérisé par le volume élémentaire représentatif  $\Omega \subset \mathbb{R}^d$  et un champ inhomogène de conductivité isotrope  $\gamma \in L_{\text{per}}^{\infty}(\Omega)$  avec  $\gamma > 0$ . L'objectif est de quantifier le comportement

macroscopique du milieu considéré en calculant le tenseur effectif  $\gamma_{\text{eff}}$ , définissant une énergie  $W_{\text{eff}}(\bar{e}) = \frac{1}{2} \bar{e} \cdot \gamma_{\text{eff}} \cdot \bar{e}$  pour tout  $\bar{e} \in \mathbb{R}^d$  satisfaisant :

$$W_{\text{eff}}(\bar{e}) = \min_{e^* \in \mathcal{E}_0} W(\bar{e} + e^*) \quad \text{avec} \quad W(e) = \frac{1}{2} \langle e(\mathbf{x}) \gamma(\mathbf{x}) e(\mathbf{x}) \rangle, \quad (1)$$

où  $e(\mathbf{x}) = \bar{e} + \tilde{e}(\mathbf{x})$ , et  $\langle \cdot \rangle$  est l'opérateur de moyenne défini comme :

$$\langle f \rangle = \frac{1}{|\Omega|} \int_{\Omega} f(\mathbf{x}) d\mathbf{x}. \quad (2)$$

Par ailleurs, on considère l'espace fonctionnel de champs « cinématiquement » admissibles suivant :

$$\mathcal{E}_0 = \{ e^* \in \mathbf{L}_{\text{per}}^2(\Omega) \mid \exists w^* \in H_{\text{per}}^1(\Omega), e^* = \nabla w^* \},$$

où  $\mathbf{L}_{\text{per}}^2(\Omega)$  est le sous-espace des champs de tenseurs dont les composantes sont dans  $L_{\text{per}}^2(\Omega)$ .

Pour  $\bar{e} \in \mathbb{R}^d$  et avec  $\tilde{e}$  minimisant (1), le problème de minimisation (1) est équivalent au système d'équations locales :

$$\begin{cases} e(\mathbf{x}) = \bar{e} + \tilde{e}(\mathbf{x}), & \tilde{e} = \nabla u, & u \text{ périodique sur } \partial\Omega, \\ j(\mathbf{x}) = \gamma(\mathbf{x}) e(\mathbf{x}), \\ \nabla \cdot j(\mathbf{x}) = 0, & j \cdot n \text{ anti-périodique sur } \partial\Omega, \end{cases} \quad (3)$$

où  $n$  est la normale sortante unitaire à  $\partial\Omega$ .  $e$  peut être physiquement interprété comme un champ électrostatique dont la moyenne  $\bar{e}$  est imposée, et  $j$  est le flux électrique. Par ailleurs, la fluctuation périodique  $\tilde{e}$  est exprimée comme gradient d'un potentiel scalaire  $u \in H_{\text{per}}^1(\Omega)$  est à moyenne nulle  $\langle \tilde{e} \rangle = \mathbf{0}$ .

Le tenseur effectif  $\gamma_{\text{eff}}$  peut être calculé composante par composante en résolvant le problème (3)  $d$  fois en utilisant des chargements  $\bar{e}$  linéairement indépendants et en calculant les produits scalaires des solutions correspondantes. Dans ce contexte, pour tout  $\alpha \in L_{\text{per}}^{\infty}(\Omega)$  avec  $\alpha > 0$ , on introduit le produit scalaire et la norme associée comme suit :

$$(\mathbf{f}_1, \mathbf{f}_2)_{\alpha} = \langle \mathbf{f}_1(\mathbf{x}) \cdot \alpha(\mathbf{x}) \mathbf{f}_2(\mathbf{x}) \rangle \quad \text{and} \quad \|\mathbf{f}\|_{\alpha} = (\mathbf{f}, \mathbf{f})_{\alpha}^{1/2}, \quad (4)$$

pour tous vecteurs  $\mathbf{f}_1, \mathbf{f}_2 \in \mathbf{L}_{\text{per}}^2(\Omega)$ . La fonctionnelle quadratique minimisée dans (1) s'exprime alors comme  $W(\mathbf{f}) = \frac{1}{2} \|\mathbf{f}\|_{\gamma}^2$ . Enfin, le produit scalaire classique sur  $L_{\text{per}}^2(\Omega)$  sera noté  $(\cdot, \cdot)$  et sa norme associée  $\|\cdot\|$ .

Soit un milieu de référence homogène de conductivité  $\gamma_0 > 0$ , on introduit maintenant l'opérateur de Green associé  $\Gamma_0$  sur  $\mathbf{L}_{\text{per}}^2(\Omega)$ , défini comme suit :

$$\Gamma_0 : \tau \mapsto e^* = \Gamma_0 \tau \quad \text{avec} \quad e^* \in \mathcal{E}_0 \text{ and } s = (\gamma_0 e^* - \tau) \in \mathcal{S}, \quad (5)$$

où l'espace fonctionnel  $\mathcal{S}$  est défini par :

$$\mathcal{S} = \{ s \in L_{\text{per}}^2(\Omega) \mid \nabla \cdot s(\mathbf{x}) = 0, s \cdot n \text{ anti-périodique sur } \partial\Omega \}.$$

Le problème local (3) est équivalent à la formulation faible suivante : trouver  $u \in H_{\text{per}}^1(\Omega)$  tel que :

$$a(u, v) = \ell(v) \quad \forall v \in H_{\text{per}}^1(\Omega), \quad (6)$$

où  $a$  et  $\ell$  sont respectivement les formes bilinéaire et linéaire de  $H_{\text{per}}^1(\Omega)$  suivantes :

$$a(u, v) = \int_{\Omega} \gamma \nabla u \cdot \nabla v d\mathbf{x} \quad \text{et} \quad \ell(v) = - \int_{\Omega} \gamma \bar{e} \cdot \nabla v d\mathbf{x}.$$

Notons que la formulation variationnelle (6) est équivalente à

$$(\gamma(\bar{e} + \tilde{e}), e^*) = 0 \quad \forall e^* = \nabla v \in \mathcal{E}_0, \quad (7)$$

avec  $\tilde{e} = \nabla u \in \mathcal{E}_0$  et l'on peut alors montrer que l'identité (7) peut alors être réécrite via l'opérateur de Green (5) sous la forme suivante :

$$\mathbf{A}\tilde{e} = \mathbf{b}, \quad \text{avec} \quad \mathbf{A} = \Gamma_0\gamma: \mathcal{E}_0 \rightarrow \mathcal{E}_0 \quad \text{et} \quad \mathbf{b} = -\Gamma_0\gamma\bar{e} \in \mathcal{E}_0. \quad (8)$$

Enfin, avec les conditions limites périodiques considérées, l'opérateur de Green (5) peut être exprimé à l'aide de la transformée de Fourier  $\mathcal{F}$  :

$$\Gamma_0\tau(\mathbf{x}) = \mathcal{F}^{-1} [\hat{\Gamma}_0(\boldsymbol{\xi}) \cdot \mathcal{F}[\tau](\boldsymbol{\xi})](\mathbf{x}) \quad \forall \mathbf{x} \in \Omega, \quad (9)$$

où le tenseur d'ordre 2 symétrique  $\hat{\Gamma}_0(\boldsymbol{\xi})$  est défini dans l'espace de Fourier par :

$$\hat{\Gamma}_0(\mathbf{0}) = \mathbf{0} \quad \text{et} \quad \hat{\Gamma}_0(\boldsymbol{\xi}) = \frac{\boldsymbol{\xi} \otimes \boldsymbol{\xi}}{\gamma_0|\boldsymbol{\xi}|^2} \quad \forall \boldsymbol{\xi} \in \mathcal{R}^* \setminus \{\mathbf{0}\}. \quad (10)$$

où  $\mathcal{R}^*$  est le réseau réciproque associé aux vecteurs de périodicité décrivant le milieu considéré. Dans le cas (standard) où ces vecteurs sont orthogonaux,  $\mathcal{R}^*$  est un réseau orthogonal.

Pour discrétiser le problème (8), on introduit une grille régulière de taille  $h$  et l'ensemble de points d'interpolation correspondant  $\{\mathbf{x}_i\}$ , où  $i$  est un multi-indice de dimension  $d$ , et le sous-espace d'approximation  $\mathcal{T}^h \subset L^2_{\text{per}}(\Omega)$  généré par les polynômes trigonométriques associés à la Transformée de Fourier Discrète (DFT) sur la grille considérée. Cette grille est assimilable à une image constituée de pixels centrés sur les points  $\{\mathbf{x}_i\}$ . On considère également le sous-espace discret  $\mathcal{E}_0^h = \mathcal{E}_0 \cap \mathcal{T}^h$  des champs de gradient à moyenne nulle, au sens de l'opérateur de moyenne  $\langle \cdot \rangle_h$ , discrétisation de (2) :

$$\langle \mathbf{f} \rangle_h = \sum_i h \mathbf{f}(\mathbf{x}_i), \quad (11)$$

De la même façon, on introduit une version discrète  $(\cdot, \cdot)_{h,\alpha}$  du produit scalaire (4) :

$$(\mathbf{f}_1, \mathbf{f}_2)_{h,\alpha} = \sum_i h \mathbf{f}_1(\mathbf{x}_i) \cdot \alpha(\mathbf{x}_i) \mathbf{f}_2(\mathbf{x}_i)$$

Par ailleurs, une version DFT  $\Gamma_0^h$  de l'opérateur de Green de (8) peut être définie formellement en transposant (9–10) sous forme DFT. On introduit alors la forme discrète de la formulation faible (7) avec le schéma d'intégration inexact (11), ce qui mène à l'équation suivante pour la solution discrète  $\tilde{e}^h \in \mathcal{E}_0^h$  :

$$(\gamma(\bar{e} + \tilde{e}^h), e^{*h})_h = 0 \quad \forall e^{*h} \in \mathcal{E}_0^h, \quad (12)$$

Enfin, on introduit aussi le flux discret  $\mathbf{j}^h = \gamma(\bar{e} + \tilde{e}^h)$ , et on montre que (12) mène au système linéaire suivant :

$$\mathbf{A}^h \tilde{e}^h = \mathbf{b}^h, \quad \text{avec} \quad \mathbf{A}^h = \Gamma_0^h \gamma: \mathcal{E}_0^h \rightarrow \mathcal{E}_0^h \quad \text{et} \quad \mathbf{b}^h = -\Gamma_0^h \gamma \bar{e} \in \mathcal{E}_0^h. \quad (13)$$

qui constitue le système étudié.

### 1.3 Objectifs de l'étude

L'approximation de  $\tilde{e}$ , la solution de (1), est basée sur (i) la discrétisation du domaine à l'aide d'une taille de grille de taille  $h$ , des fonctions de base et d'un schéma d'intégration, et (ii) un schéma itératif donné dont le nombre d'itérations est noté  $k$ . On aboutit alors à une solution approchée  $\tilde{e}_k^h$ . Dans ce contexte, notre objectif est de proposer un critère de convergence pour la propriété effective  $\gamma_{\text{eff}}$ , ce qui revient, d'après (1), à quantifier l'erreur  $\mathcal{E}_{k,h}^{\text{eff}}$  définie comme suit :

$$\mathcal{E}_{k,h}^{\text{eff}} \stackrel{\text{def}}{=} |W(e_k^h) - W(e)| \quad \text{avec} \quad e_k^h = \bar{e} + \tilde{e}_k^h,$$

pour  $\bar{e} \in \mathbb{R}^d$  donné. En supposant que le schéma itératif converge, on introduit la limite  $e_\infty^h = \bar{e} + \tilde{e}_\infty^h$ . Dans la suite, nous nous intéresserons à la relation entre l'erreur ci-dessus et l'erreur totale  $\mathcal{E}_{k,h}^{\text{tot}}$  sur le champ, qui est telle que :

$$\mathcal{E}_{k,h}^{\text{tot}} \stackrel{\text{def}}{=} \|e_k^h - e\|_\gamma \leq \underbrace{\|e_k^h - e_\infty^h\|_\gamma}_{\mathcal{E}_k^h} + \underbrace{\|e_\infty^h - e\|_\gamma}_{\mathcal{E}^h},$$

où  $\mathcal{E}_k^h$  est l'erreur d'*itération* par rapport à la solution discrète convergée  $e_\infty^h$ , et  $\mathcal{E}^h$  est l'erreur de *discrétisation* par rapport à la solution exacte de (1). Remarquons que les définitions des erreurs ci-dessus utilisent formellement la norme *continue* définie à l'équation (4).

Comme ni la propriété effective  $\gamma_{\text{eff}}$  ni la solution continue ne sont connues *a-priori*, ces erreurs ne sont pas accessibles directement. C'est pourquoi nous allons proposer des estimateurs *a-posteriori*, notés  $\mathcal{E}_k^h$  et  $\mathcal{E}^h$  pour les erreurs d'itération et de discrétisation respectivement. Pour le premier, on utilisera simplement le résidu du système linéaire (13) (ce qui sera justifié dans la présentation), tandis que la proposition de méthodes pour déterminer le second est l'objet de la section 2.

**Remarque 1** Dans cette étude, on ne prend pas explicitement en compte l'erreur de géométrie, mais elle peut être considérée comme faisant partie de l'erreur d'intégration puisqu'elle participe à l'écart le calcul des produits scalaires  $(\bullet, \bullet)_\gamma$  et  $(\bullet, \bullet)_{h,\gamma}$ .

## 1.4 Cas-tests numériques

Le premier cas-test numérique étudié est celui d'une unique inclusion carrée de fraction surfacique 1/4. Cette géométrie a été choisie car la propriété effective correspondante a été déterminée analytiquement dans [8]. Il est à noter que la solution du problème (3) sur cette géométrie présente des singularités (intégrables) de coins. Ainsi, une deuxième géométrie plus régulière est également considérée. Il s'agit d'une distribution périodique de disques, obtenus par génération aléatoire, telle que la fraction surfacique des inclusions soit elle aussi 1/4. Ces géométries sont présentées sur la Figure 1. Le rapport, ou contraste, de conductivité entre les inclusions (en rouge) et la matrice (en bleu) est noté  $c$ .

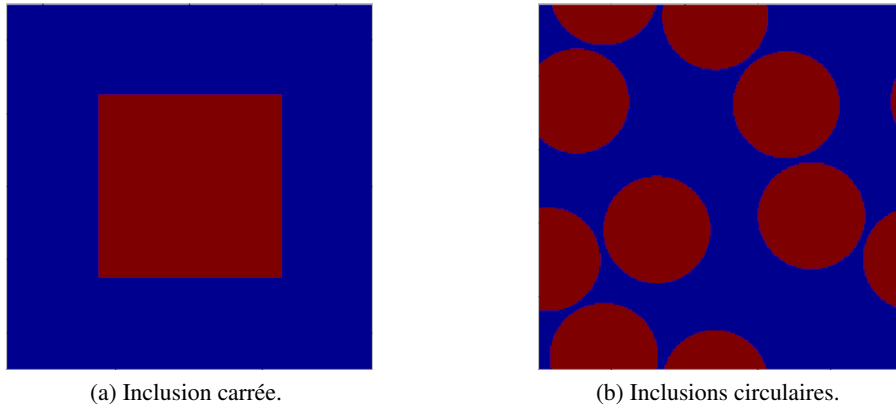


FIGURE 1 – Géométries utilisées.

## 2 Estimateurs d'erreur de discrétisation

Pour estimer l'erreur  $\mathcal{E}^h = \|\tilde{e} - \tilde{e}_\infty^h\|_\gamma = \|\gamma(e - e_\infty^h)\|_{\gamma^{-1}}$ , nous proposons trois méthodes s'inspirant de celles qui ont été développées dans le cadre de la méthode des éléments finis. Ces trois méthodes reposent sur la construction d'un flux modifié noté  $j_*^h$ , et sur l'idée que  $\mathcal{E}^h = \|j_*^h - \gamma e^h\|_{\gamma^{-1}}$  puisse fournir une approximation satisfaisante de  $\mathcal{E}^h$ . Les trois méthodes sont présentées ci-après mais leurs justifications théoriques seront discutées lors de la présentation.

### 2.1 Filtre moyen

Cette méthode, inspirée de la méthode ZZ1, consiste à déterminer  $j_*^h$  indépendamment en chaque point d'échantillonnage (pixel) comme la moyenne pondérée des valeurs du flux discret  $j^h$  aux points d'échantillonnage adjacents :

$$j_*^h(\mathbf{g}) = \sum_{\mathbf{g}' \in \mathcal{G}(A_g)} \alpha_{g,g'} j^h(\mathbf{g}')$$

où  $\mathcal{G}(A_g)$  est l'ensemble des points d'échantillonnage adjacents au point  $\mathbf{g}$ , et  $\alpha_{g,g'}$  un poids défini selon une règle simple.

## 2.2 Filtre médian

Ici, chaque composante de  $j_*^h, j_*^h{}_\zeta$  est estimée indépendamment en chaque point comme la médiane pondérée des valeurs de  $j^h$  aux points d'échantillonnage adjacents :

$$j_*^h(\mathbf{g})_\zeta = \arg \min_{j_* \in \mathbb{R}^2} \sum_{\mathbf{g}' \in \mathcal{G}(A_g)} \alpha_{g,g'} |j^h(\mathbf{g}')_\zeta - j_*| \quad \text{avec} \quad \zeta = x \text{ or } y.$$

## 2.3 Erreur en Relation de Comportement

Enfin, on propose d'utiliser la méthode d'équilibrage de [6], qui consiste en deux étapes :

- Déterminer un flux normal équilibré  $F_*^h$  sur les bords de chaque pixel  $E_g$  à partir du flux  $j^h$ .
- Calculer  $j_*^h$  par le calcul éléments finis sur un maillage raffiné d'un sous-problème, indépendamment sur chaque pixel, à savoir :

$$\begin{cases} j_*(\mathbf{x}) = \gamma^g \nabla u_*(\mathbf{x}) & \mathbf{x} \in E_g \\ \nabla \cdot j_*(\mathbf{x}) = 0 & \mathbf{x} \in E_g \\ j_*(\mathbf{x}) \cdot \nu_g(\mathbf{x}) = F_*^h(\mathbf{x}) & \mathbf{x} \in \partial E_g \end{cases}$$

où  $\gamma^g$  est la conductivité *homogène* dans le pixel, et  $\nu_g$  est la normale sortante unitaire à  $\partial E_g$ .

**Remarque 2** Dans le cas usuel où la grille est régulière, le coût algorithmique de la construction de  $j_*^h$  dans le cadre de la méthode de l'Erreur en Relation de Comportement est comparable à celui de la méthode de filtrage moyen, alors que cette première méthode est connue pour être bien plus coûteuse en temps de calcul dans le cas d'un maillage quelconque.

## 2.4 Évaluation des estimateurs

On peut évaluer les trois estimateurs proposés plus haut sur des cas à faible contraste  $c = 10$ , et avec une discrétisation grossière de  $2^6$  pixels par côté. L'erreur « de référence » est obtenue en comparant la solution obtenue à celle issue d'un calcul sur une grille de  $2^{10} \times 2^{10}$  pixels. On trace les cartes logarithmiques de l'erreur locale pour le cas de l'inclusion carrée sur la Figure 2a, et pour les inclusions circulaires sur la Figure 2b. Les erreurs globales correspondantes (norme  $L^2$  sur l'ensemble du domaine du champ d'erreur locale) sont données sur le Tableau 1. On observe notamment que la borne usuelle de l'erreur en relation de comportement (l'erreur estimée devrait être supérieure à l'erreur vraie) n'est pas respectée ici. Ceci est dû au fait que le problème discret (12) est intégré de façon inexacte, avec le produit scalaire discret  $(\bullet, \bullet)_{h,\gamma}$ .

TABLE 1 – Erreurs globales estimées et de référence

	Référence	Filtre moyen	Filtre médian	Err. en Relation de Comp.
Inclusion carrée	0.0196	0.0948	0.0325	0.0227
Inclusions circulaires	0.131	0.327	0.246	0.161

Il apparaît que la méthode d'estimation d'erreur par filtre moyen a tendance à sur-estimer grandement l'erreur. Ceci est dû au fait que ce filtrage a tendance à lisser les discontinuités de flux, même quand celles-ci sont présentes dans la solution exacte (voir les côtés supérieur et inférieur de l'inclusion carrée). L'estimateur par filtre médian conduit à des résultats de meilleure qualité, et enfin l'erreur en relation de comportement semble dans les deux cas présents être la méthode la plus précise.

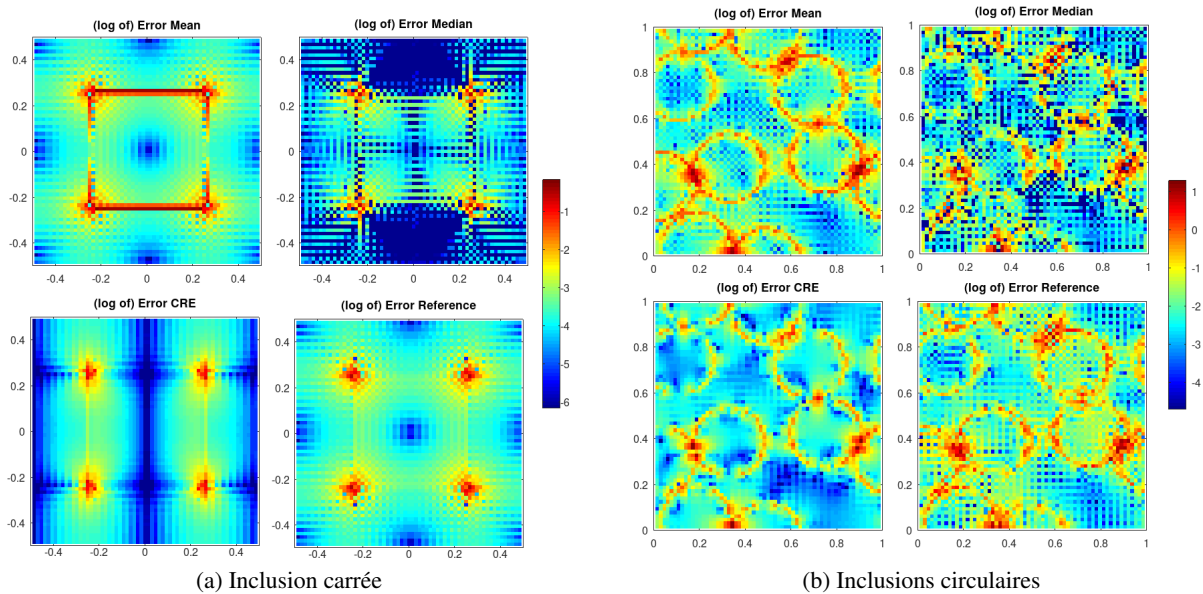


FIGURE 2 – Cartes d’erreurs estimées et de référence (échelle logarithmique).

Il sera néanmoins expliqué dans la présentation que l’estimateur d’erreur basé sur l’application d’un filtre médian est apparu comme le plus robuste les cas de fort contraste. C’est pourquoi le principe du critère d’arrêt de la section suivante est illustré sur cette méthode. Remarquons toutefois que d’autres méthodes de reconstruction d’un flux équilibré plus adaptées aux forts contrastes pourraient être développées pour profiter du fait que, pour des grilles régulières, la méthode de l’erreur en relation de comportement est beaucoup moins coûteuse en temps de calcul.

### 3 Critère d’arrêt basé sur l’estimateur d’erreur par filtre médian

On construit deux critères, notés  $R_5$  et  $\{E_r\}_{r=1,10,100}$ , qui seront évalués sur les cas-tests considérés :

- $R_5$  consiste à stopper les itérations lorsque  $\varepsilon_k^h \leq 10^{-5} \varepsilon_0^h$ , où  $\varepsilon_0^h$  est le résidu initial. Il s’agit du critère d’arrêt classiquement utilisé.
- $E_r$  consiste à stopper les itérations lorsque  $\varepsilon_k^h \leq r^{-1} \varepsilon^h$ . Il s’agit d’attendre que l’estimateur d’erreur d’itération atteigne une fraction de l’estimateur de l’erreur de discrétisation.

On introduit alors un indicateur de l’amélioration possible, notée  $\delta_k^h$ , qui permettra d’évaluer la performance des critères d’arrêt considérés. Cette quantité (qui n’est pas disponible si la solution exacte n’est pas connue) consiste en un rapport entre erreur d’itération et erreur totale sur la propriété effective. Elle permet de quantifier à quel point l’erreur sur la propriété effective aurait pu être réduite si on avait laissé le schéma itérer indéfiniment.

$$\delta_k^h = \sqrt{\frac{|W_h(e_k^h) - W_h(e_\infty^h)|}{|W(e) - W_h(e_\infty^h)|}}$$

En pratique,  $e_\infty^h$  est estimé en utilisant un nombre d’itérations égal au quadruple du nombre maximal d’itérations donné par les quatre critères évalués. Lorsque  $\delta_k^h$  est très petit devant 1, cela signifie qu’il est devenu inutile d’itérer car l’erreur sur la propriété effective, qu’elle soit grande ou petite, ne pourra plus évoluer. Un critère d’arrêt satisfaisant devrait donc idéalement permettre de contrôler  $\delta_k^h$ .

On considère un fort contraste de conductivités  $c = 10^3$ , et une discrétisation allant de  $2^5$  à  $2^{12}$  pixels par côté. Le problème linéaire (13), qui est symétrique, est résolu par un algorithme du Gradient Conjugué. Pour chaque problème et chaque critère d’arrêt, si  $k$  est l’itération pour laquelle le critère d’arrêt est atteint, on donne sur la figure 3 la valeur correspondante de  $\delta_k^h$ .

On observe alors que, pour les deux géométries étudiées, quand  $k$  est déterminé par le critère d’arrêt usuel  $R_5$  basé uniquement sur le résidu,  $\delta_k^h$  augmente lorsque la discrétisation devient plus fine. Ceci signifie que l’on n’a aucun contrôle effectif sur le niveau d’erreur total. À l’inverse, pour les critères d’arrêt de type  $E_r$ , utilisant l’estimateur d’erreur de discrétisation,  $\delta_k^h$  dépend peu de la discrétisation, et a

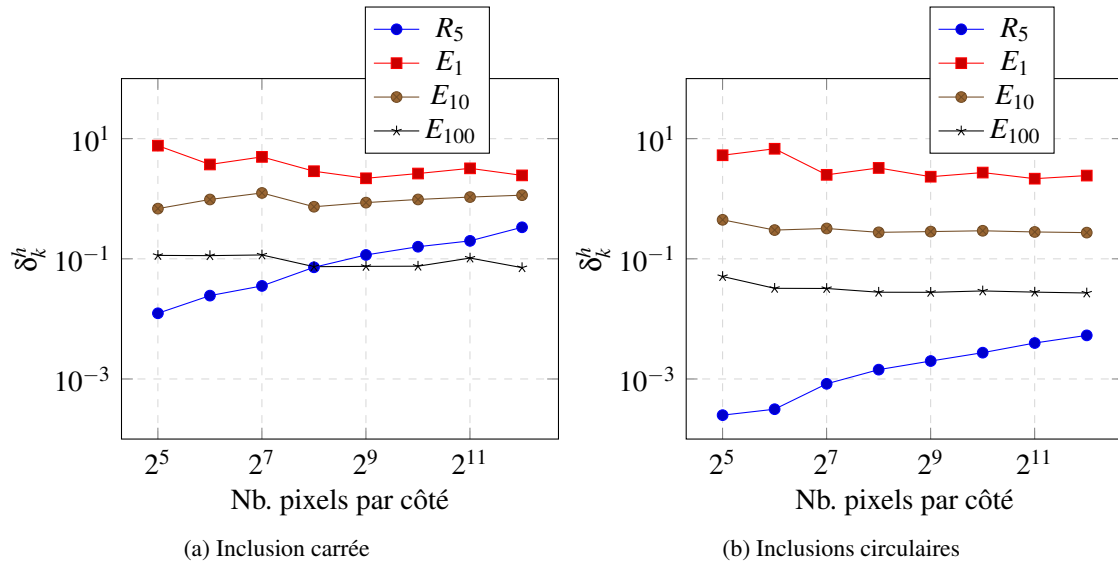


FIGURE 3 – Amélioration possible  $\delta_k^h$  à l'issue du procédé itératif pour différentes discrétisations et différents critères d'arrêt

des niveaux proches entre les deux cas-tests. Par conséquent, si  $r$  est suffisamment grand, l'erreur totale n'est pilotée que par l'erreur de discrétisation. Tout repose alors sur le choix de la bonne valeur de  $r$  assurant  $\delta_k^h \ll 1$ . On montrera au cours de la présentation des résultats permettant d'aboutir à des règles heuristiques pour choisir  $r$  en fonction notamment du contraste de conductivité.

## 4 Conclusion

On propose dans cette contribution un critère d'arrêt pour les calculs d'homogénéisation numérique par FFT et schémas itératifs. Il s'agit d'estimer l'erreur de discrétisation sur la solution courante, et d'arrêter le processus itératif lorsque cette erreur est suffisamment petite devant le résidu, qui quantifie l'erreur d'itération. On montre, au travers d'expériences numériques, que cette procédure permet de contrôler l'erreur totale sur la propriété effective. Par ailleurs, une fois que le calcul terminé, on dispose alors d'une estimation de l'erreur associée pouvant être utilisée pour évaluer sa qualité.

Ce critère d'arrêt doit s'appuyer sur un estimateur d'erreur de discrétisation aussi fiable et robuste que possible. Trois ont été développés en s'inspirant des estimateurs ZZ1 et de l'erreur en relation de comportement, utilisés classiquement dans le cadre des éléments finis. Dans les cas considérés, le filtre médian est apparu comme le plus performant, tandis que l'estimateur basé sur l'erreur en relation de comportement paraît prometteur.

Enfin, plusieurs perspectives à ce travail émergent naturellement. D'abord, lorsque le produit scalaire énergétique n'est pas calculé de façon exacte il n'est pas possible d'encadrer strictement la propriété effective calculée numériquement. C'est pourquoi il pourrait être intéressant de chercher à appliquer notre procédure à une variante de la méthode d'homogénéisation numérique par FFT pour laquelle les produits scalaires sont calculés de façon exacte. Une deuxième perspective est liée au coût CPU de l'estimation d'erreur. Pour éviter que celui-ci n'alourdisse inutilement les itérations, il peut être intéressant de ne calculer l'estimation d'erreur que toutes les  $n$  itérations, ou bien à chaque fois que le résidu passe en dessous d'une valeur donnée.

## Références

- [1] S. Brisard and L. Chamoin. Constitutive relation error for fft-based methods. In *VII European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS Congress2016)*. Crete Island, Greece, 2016.



- [2] S. Brisard and L. Dormieux. Combining Galerkin approximation techniques with the principle of Hashin and Shtrikman to derive a new FFT-based numerical method for the homogenization of composites. *Comput Methods Appl Mech Eng.*, 217–220 :197–212, 2012.
- [3] D. J. Eyre and G. W. Milton. A fast numerical scheme for computing the response of composites using grid refinement. *The European Physical Journal Applied Physics*, 6(1) :41–47, 1999.
- [4] M. Kabel, T. Böhlke, and M. Schneider. Efficient fixed point and Newton–Krylov solvers for FFT-based homogenization of elasticity at large deformations. *Computational Mechanics*, 54(6) :1497–1514, 2014.
- [5] P. Ladeveze and D. Leguillon. Error estimate procedure in the finite element method and applications. *SIAM Journal on Numerical Analysis*, 20(3) :485–509, 1983.
- [6] P. Ladevèze and P. Rougeot. New advances on a posteriori error on constitutive relation in f.e. analysis. *Computer Methods in Applied Mechanics and Engineering*, 150(1-4) :239–249, 1997.
- [7] H. Moulinec and P. Suquet. A numerical method for computing the overall response of nonlinear composites with complex microstructure. *Computer methods in applied mechanics and engineering*, 157(1-2) :69–94, 1998.
- [8] Y. V. Obnosov. Periodic heterogeneous structures : new explicit solutions and effective characteristics of refraction of an imposed field. *SIAM Journal on Applied Mathematics*, 59(4) :1267–1287, 1999.
- [9] V. Rey, C. Rey, and P. Gosselet. A strict error bound with separated contributions of the discretization and of the iterative solver in non-overlapping domain decomposition methods. *Computer Methods in Applied Mechanics and Engineering*, 270 :293–303, 2014.
- [10] J. Vondřejc, J. Zeman, and I. Marek. An FFT-based Galerkin method for homogenization of periodic media. *Computers & Mathematics with Applications*, 68(3) :156–173, 2014.
- [11] F. Willot, B. Abdallah, and Y.-P. Pellegrini. Fourier-based schemes with modified Green operator for computing the electrical response of heterogeneous media with accurate local fields. *International Journal for Numerical Methods in Engineering*, 98(7) :518–533, 2014.
- [12] J. Zeman, J. Vondřejc, J. Novák, and I. Marek. Accelerating a FFT-based solver for numerical homogenization of periodic media by conjugate gradients. *Journal of Computational Physics*, 229(21) :8065–8071, 2010.
- [13] O. C. Zienkiewicz and J. Z. Zhu. A simple error estimator and adaptive procedure for practical engineering analysis. *International journal for numerical methods in engineering*, 24(2) :337–357, 1987.