# Etude mathématique et numérique de quelques modèles multi-échelles issus de la mécanique des matériaux

Marc Josien

UNIVERSITÉ PARIS-EST

**Ecole Doctorale Mathématiques et Sciences et Technologies de l'Information et de la Communication (MSTIC)**

# THÈSE DE DOCTORAT

Discipline : Mathématiques Appliquées

présentée par

**Marc JOSIEN**

---

## Etude mathématique et numérique de quelques modèles multi-échelles issus de la mécanique des matériaux

---

*Thèse dirigée par Claude Le Bris au CERMICS, École des Ponts Paristech*

*Soutenue le 20 Novembre 2018 devant le jury composé de :*

| | | |
|---|---|---|
| Felix Otto | Institut Max Planck MIS | Président de Jury |
| Anne-Laure Dalibard | Université Pierre et Marie Curie | Rapporteur |
| Gilles Francfort | Université Paris-Nord | Rapporteur |
| Sonia Fliss | ENSTA | Examinatrice |
| Yves-Patrick Pellegrini | CEA/DAM DIF | Examinateur |
| Claude Le Bris | École des Ponts ParisTech | Directeur de thèse |

*Horloge ! dieu sinistre, effrayant, impassible,*
*Dont le doigt nous menace, et nous dis : «Souviens-toi !*
*De vibrantes Douleurs dans ton cœur plein d'effroi*
*Se planteront bientôt comme dans une cible»*

*Charles Baudelaire*

## Remerciements

Je tiens tout d'abord à remercier mes encadrants Xavier Blanc, Claude Le Bris, Frédéric Legoll et Yves-Patrick Pellegrini. Merci de m'avoir soutenu et prodigué votre amitié, votre savoir, et surtout votre temps. Car s'il est long d'écrire un texte scientifique, il est encore plus long de le relire et de l'amender.

Je remercie Anne-Laure Dalibard et Gilles Francfort d'avoir bien voulu être rapporteurs de ce travail, ainsi que Sonia Fliss, Felix Otto et Yves-Patrick Pellegrini d'avoir accepté de faire partie du jury.

Monter un dossier de candidature pour faire une thèse au sein du Corps des Ingénieurs des Ponts, des Eaux et des Forêts est une entreprise délicate. Je suis reconnaissant envers Claude Le Bris et à Françoise Prêteux pour leur aide, et envers Charles Lion pour son soutien. Je remercie le Corps de m'avoir donné la chance de faire une thèse de mathématiques appliquées.

Le CERMICS est un environnement très stimulant. Je remercie l'ensemble de l'équipe et particulièrement Gabriel Stoltz, pour m'avoir fait confiance au point de me confier une classe d'étudiants, et Antoine Levitt, pour des discussions très intéressantes sur le numérique "efficace". Le petit groupe des doctorants forme une communauté sympathique, et je garde le meilleur souvenir de Julien, Grégoire, Atman, Boris, François, Pierre-Loïk, Laura, Marion, Frédéric, Alexandre, Henri, Adrien, Ling-Ling, Rafaël...

J'ai passé une part importante de mon temps au sein du Département de Physique Théorique et Appliquée, à la Direction des Applications Militaires du CEA. Cette rencontre avec des chercheurs en physique a été très enrichissante et je remercie l'équipe de Christophe Denoual pour son accueil, et particulièrement Ronan Madec pour de nombreuses discussions sur la dynamique des dislocations.

Parfois, les choses ne se passent pas si bien. Merci à Isabelle Simunic, Xavier Blanc, Yves-Patrick Pellegrini, Gabriel Stoltz et Eric Cancès pour m'avoir aidé dans des moments plus durs. Merci aussi à mes fidèles amis Martin et Alban.

Enfin, je souhaite remercier ceux qui ont su allumer et entretenir en moi la flamme des mathématiques : mon père et mes professeurs Didier Plichard, Anne-Laure Biolley, Alain Pommellet, Grégoire Allaire et Felix Otto. Chacun, à sa manière, m'a beaucoup apporté.

Merci enfin à Apolline et à ma famille, à qui je dédie cette thèse.

**Sujet** Etude mathématique et numérique de quelques modèles multi-échelles issus de la mécanique des matériaux

**Résumé** Le travail de cette thèse a porté sur l'étude mathématique et numérique de quelques modèles multi-échelles issus de la physique des matériaux.

La première partie de ce travail est consacrée à l'homogénéisation mathématique d'un problème elliptique avec une petite échelle. Nous étudions le cas particulier d'un matériau présentant une structure périodique avec un défaut. En adaptant la théorie classique d'Avellaneda et Lin pour les milieux périodiques, on démontre qu'on peut approximer finement la solution d'un tel problème, notamment à l'échelle microscopique. Nous obtenons des taux de convergence dépendant de l'étalement du défaut. On démontre aussi quelques propriétés des fonctions de Green d'un problème elliptique périodique avec conditions de bord périodiques.

Les dislocations sont des lignes de défaut de la matière responsables du phénomène de plasticité. Les deuxième et troisième parties de ce mémoire portent sur la simulation de dislocations, d'abord en régime stationnaire puis en régime dynamique. Nous utilisons le modèle de Peierls, qui couple échelle atomique et échelle mésoscopique. Dans le cadre stationnaire, on obtient une équation intégrodifférentielle non-linéaire avec un laplacien fractionnaire : l'équation de Weertman. Nous en étudions les propriétés mathématiques et proposons un schéma numérique pour en approximer la solution. Dans le cadre dynamique, on obtient une équation intégrodifférentielle à la fois en temps et en espace. Nous en faisons une brève étude mathématique, et comparons différents algorithmes pour la simuler.

Enfin, dans la quatrième partie, nous étudions la limite macroscopique d'une chaîne d'atomes soumis à la loi de Newton. Des arguments formels suggèrent que celle-ci devrait être décrite par une équation des ondes non-linéaires. Or, nous démontrons –sous certaines hypothèses– qu'il n'en est rien lorsque des chocs apparaissent.

**Mots-clefs** multi-échelle, homogénéisation, équation elliptique, fonction de Green, dislocations, équation de Peierls-Nabarro, équation de réaction-diffusion, équation intégrodifférentielle, laplacien fractionnaire

**Title** Mathematical and numerical study of some multi-scale models from materials science

**Abstract** In this thesis we study mathematically and numerically some multi-scale models from materials science.

First, we investigate an homogenization problem for an oscillating elliptic equation. The material under consideration is described by a periodic structure with a defect at the microscopic scale. By adapting Avellaneda and Lin's theory for periodic structures, we prove that the solution of the oscillating equation can be approximated at a fine scale. The rates of convergence depend upon the integrability of the defect. We also study some properties of the Green function of periodic materials with periodic boundary conditions.

Dislocations are lines of defects inside materials, which induce plasticity. The second part and the third part of this manuscript are concerned with simulation of dislocations, first in the stationnary regime then in the dynamical regime. We use the Peierls model,

which couples atomistic and mesoscopic scales and involves integrodifferential equations. In the stationary regime, dislocations are described by the so-called Weertman equation, which is nonlinear and involves a fractional Laplacian. We study some mathematical properties of this equation and propose a numerical scheme for approximating its solution. In the dynamical regime, dislocations are described by an equation which is integrodifferential in time and space. We compare some numerical methods for recovering its solution.

In the last chapter, we investigate the macroscopic limit of a simple chain of atoms governed by the Newton equation. Surprisingly enough, under technical assumptions, we show that it is not described by a nonlinear wave equation when shocks occur.

**Keywords** multi-scale, homogenization, elliptic equation, Green function, dislocations, Peierls-Nabarro equation, reaction-diffusion equation, integrodifferential equation, fractional laplacian

**Mathematical subject classification (2010)** 26A33, 35B27, 35J15, 35K57, 35R11, 45E05, 65R20, 65T50

# Table des matières

# Introduction générale

L'approche multi-échelle consiste à modéliser, étudier théoriquement et simuler numériquement des systèmes *macroscopiques* en gardant à l'esprit que ceux-ci sont issus de modèles d'abord *microscopiques*. Un enjeu majeur (en particulier pour l'homogénéisation) est d'extraire les ingrédients pertinents des mécanismes microscopiques afin de décrire le comportement macroscopique du système étudié.

Cette thèse porte sur quelques problèmes, issus de la physique des matériaux, faisant intervenir différentes échelles spatiales, et parfois temporelles :

— un problème théorique d'homogénéisation déterministe non-périodique,
— la simulation de dislocations en régime stationnaire et en régime dynamique,
— l'étude théorique d'un système linéique d'atomes soumis à la loi de Newton.

L'homogénéisation est un champ des mathématiques qui vise à extraire la limite d'une équation aux dérivées partielles lorsqu'une échelle du problème est très petite. En pratique, elle sert à approximer le comportement macroscopique (par exemple thermique ou mécanique) d'un matériau complexe présentant une microstructure. Pour ce faire, on se ramène artificiellement à un matériau simple équivalent. Pour réaliser cette approximation et en contrôler la pertinence, il est nécessaire d'avoir des informations sur la *structure* microscopique du matériau étudié. Dans le cadre déterministe, on suppose habituellement que cette structure est périodique ; on peut alors contrôler finement l'approximation, notamment grâce aux travaux d'Avellaneda et Lin. L'objet de la première partie de cette thèse est d'étudier dans quelle mesure on peut se passer de l'hypothèse de périodicité tout en conservant des résultats d'approximation. Le but est d'accéder à des structures plus riches, comme les *défauts* microscopiques. Cette recherche s'inscrit dans la continuité des travaux de X. Blanc, C. Le Bris et P.-L. Lions.

Un large pan de la science des matériaux consiste à caractériser et à simuler numériquement les phénomènes de plasticité. Ces derniers peuvent s'expliquer par la présence de dislocations dans les cristaux. Les dislocations sont des lignes de défauts qui se meuvent, se multiplient et s'annihilent dans le matériau au gré des contraintes mécaniques. Dans une deuxième partie, nous étudions l'équation de Weertman, une équation intégro-différentielle non-linéaire modélisant une dislocation isolée en régime stationnaire. Cette équation est issue du couplage entre un modèle mésoscopique continu et un modèle microscopique discret. On peut l'interpréter comme décrivant les fronts de propagation d'une équation de réaction-diffusion non-locale, ce qui se révèle utile pour la simuler numériquement.

Récemment, Y.-P. Pellegrini a généralisé l'équation de Weertman grâce à une nouvelle

équation –dite de Peierls-Nabarro Dynamique– décrivant l'évolution d'une dislocation en régime dynamique. Cette équation semi-linéaire est intégrodifférentielle à la fois en temps et en espace. La troisième partie de cette thèse étudie la simulation numérique d'une telle équation, en mettant l'accent sur une difficulté particulière : gérer efficacement la *mémoire* de la dislocation.

Le sujet concernant les dislocations a été fait en étroite collaboration avec des membres du Département de Physique Théorique et Appliquée de la Direction des Applications Militaires (DAM) du CEA où je me suis rendu régulièrement. Leurs objectifs propres en physique des matériaux ont motivé la construction de schémas, et l'implémentation de deux codes en MATLAB : pour étudier les dislocations en régime stationnaire, puis en régime dynamique.

Il est des cas où le passage d'une échelle à une autre peut être particulièrement complexe. La quatrième partie de cette thèse est consacrée à l'étude d'un système linéique d'atomes soumis à la loi de Newton et interagissant non-linéairement avec leur plus proche voisin. Grâce aux travaux de X. Blanc, C. Le Bris et P.-L. Lions, il est possible décrire le comportement macroscopique du système dans certains régimes grâce à une équation des ondes non-linéaires. Mais l'irruption de chocs change radicalement ce comportement macroscopique, qui n'est plus décrit par cette équation.

Les contributions originales de cette thèse sont :

— la généralisation de preuves d'estimations fines en homogénéisation dans le cas de certaines structures déterministes non périodiques (voir le Chapitre 2),
— l'étude de propriétés la fonction de Green avec conditions de bord périodiques dans des problèmes d'homogénéisation périodique (voir le Chapitre 3),
— une étude mathématique de l'équation de Weertman (voir le Chapitre 4),
— la construction d'un schéma pour approximer numériquement les solutions de l'équation de Weertman (voir le Chapitre 5),
— l'implémentation et la comparaison de divers schémas numériques pour la simulation de l'équation de Peierls-Nabarro Dynamique (voir les Chapitres 6, 7 et 8),
— la démonstration de la non-convergence en cas de chocs d'un système linéique d'atomes vers l'équation des ondes non-linéaire associée (voir le Chapitre 9).

# Liste des publications


[1] X. Blanc and M. Josien. From the Newton equation to the wave equation : the case of shock waves. *Applied Mathematics Research eXpress*, 2017 :338–385, 2017.

[2] X. Blanc, M. Josien, and C. Le Bris. Approximation locale précisée dans des problèmes multi-échelles avec défauts localisés. in preparation.

[3] X. Blanc, M. Josien, and C. Le Bris. Local approximation of the gradient for multiscale problems with defects. in preparation.

[4] M. Josien. Mathematical properties of the Weertman equation. Preprint arXiv :1709.0678, accepted by Comm. Math. Sci.

[5] M. Josien. Decomposition and pointwise estimates of periodic Green functions of some elliptic equations with periodic oscillatory coefficients. Preprint arXiv :1807.09062, accepted by Asymptotic Analysis.

[6] M. Josien, Y.-P. Pellegrini, F. Legoll, and C. Le Bris. Fourier-based numerical approximation of the Weertman equation for moving dislocations. *International Journal for Numerical Methods in Engineeging*, 113 :1827–1850, 2018.


Les références [5], [4], [6], [1], et [2] sont reproduites (avec des adaptations typographiques mineures) dans les Chapitres 3, 4, 5, et 9, et l'Annexe B, respectivement. On retrouve le matériau de [3] dans le Chapitre 2.

# Chapitre 1

# Introduction

Dans ce premier chapitre, nous introduisons et remettons en perspective le travail effectué durant cette thèse. Nous avons travaillé sur trois grandes thématiques :

— l'homogénéisation mathématique de problèmes elliptiques multi-échelles déterministes, en collaboration avec Xavier Blanc, Claude Le Bris et Frédéric Legoll ;

— la simulation de dislocations dans le cadre du modèle de Peierls, dans un régime stationnaire et dynamique, en collaboration avec Claude Le Bris, Frédéric Legoll et Yves-Patrick Pellegrini ;

— la limite macroscopique d'une chaîne d'atomes soumis à la loi de Newton, en collaboration avec Xavier Blanc [1].

Nous présentons successivement ces trois thèmes dans les Sections 1.1, 1.2 et 1.3. En Section 1.4, nous indiquons quelques questions ouvertes intéressantes.

Le lecteur trouvera en Annexe A.2 quelques précisions sur le sens physique de l'équation de Weertman. Par ailleurs, l'Annexe B propose un point de vue sur l'homogénéisation de problèmes elliptique sur des matériaux périodiques avec défauts très semblable à celui qui est développé dans ce chapitre.

1. ce sujet a été exploré pendant le stage de Master 2 de l'année 2015, mais nous avons rédigé l'article correspondant au Chapitre 9 pendant les premiers mois de la thèse

## 1.1 Etude de deux problèmes d'homogénéisation

Les équations aux dérivées partielles qui régissent la physique et la mécanique, bien que souvent étudiées dans des milieux homogènes, sont en réalité généralement issues de modèles faisant intervenir plusieurs échelles spatiales. Imaginons par exemple un fluide s'écoulant dans un matériau poreux : le milieu traversé est constitué d'une multitude de surfaces infranchissables et d'interstices dans lequel le fluide peut pénétrer. D'un point de vue numérique, il est très coûteux de mailler chaque pore du matériau, car ils sont de très petite taille. Par ailleurs, quand les quantités qui intéressent l'ingénieur sont macroscopiques, il est inutile de vouloir simuler précisément ces quantités à l'échelle microscopique. Il faut donc changer de méthode et proposer une formulation macroscopique du problème. Cette approche multi-échelles des problèmes d'équations aux dérivées partielles s'est développée à partir des années 1970, avec les contributions de J.-L. Lions [18], Murat et Tartar [145], Babuska, Papanicolaou et Varadhan [128], Jikov et Koslov [85]. L'*homogénéisation* d'une équation différentielle est l'opération qui consiste à transformer une équation posée sur un domaine présentant des hétérogénéités microscopiques en une équation macroscopique sur un milieu *homogène* équivalent.

Ce processus a été en particulier théorisé pour l'équation elliptique :

$$-\mathrm{div}(A_\varepsilon \cdot \nabla u^\varepsilon) = f, \tag{1.1}$$

où $A_\varepsilon(x) = A(x/\varepsilon)$ est une matrice elliptique, $\varepsilon$ est une petite échelle et $f$ est un forçage donné. Pour pouvoir expliciter le calcul des coefficients modélisant le milieu homogène, il faut que le milieu hétérogène possède des propriétés particulières comme par exemple la périodicité. Or, d'un point de vue mécanique, les matériaux possèdent certes plusieurs échelles, mais leur structure microscopique est rarement parfaitement périodique (excepté pour les mono-cristaux, qui sont en général de petite taille).

Pour se libérer de l'hypothèse de périodicité, un premier axe d'exploration est de se placer dans un cadre aléatoire, où la loi du matériau vérifie des hypothèses de stationnarité (voir par exemple [128]). L'analyse théorique de l'homogénéisation stochastique a connu une recrudescence d'activité ces dernières années, avec notamment les travaux de Gloria, Neukamm et Otto [68], Armstrong et Smart [9], Mourrat... La simulation numérique de tels problèmes demeure très coûteuse et constitue aussi un champ de recherche actif (voir la revue [4]). L'hypothèse de stationnarité, bien que séduisante, est cependant l'analogue probabiliste de l'hypothèse de périodicité : aussi est-elle relativement rigide.

Aussi, certains auteurs [27, 28, 138] ont étudié d'autres cadres. En particulier, Blanc, Le Bris et Lions (voir [27, 28]) se sont posé la question fondamentale suivante : quelles sont les hypothèses de structure nécessaires pour pouvoir faire de l'homogénéisation *utilisable* ? La notion d'*utilisabilité* est ici d'une grande importance. En effet, on sait depuis les travaux de Tartar (voir [145, Th. 6.5 p. 82]) qu'il est possible d'homogénéiser l'équation (1.1) sous des hypothèses très faibles sur $A$. Malheureusement, ce résultat abstrait n'indique pas quel est le problème limite et ne quantifie par la qualité de l'approximation du problème originel (1.1) par la solution du problème limite. En particulier, on ne sait pas construire de stratégie

numérique à partir d'un tel résultat. Au contraire, dans le cas où $A$ est périodique, cas que nous détaillons brièvement dans la Section 1.1.1 ci-dessous, on peut *construire* le problème limite et exprimer de manière quantifiée dans quelle mesure il approxime le problème originel.

A notre connaissance, il existait deux grands cas déterministes où une telle construction était possible en pratique : le cas où $A$ est périodique, et le cas où $A$ est quasi-périodique. Dans le cadre stochastique, on fait une hypothèse d'invariance par translation de la loi des coefficients : la matrice homogénéisée et les correcteurs sont alors construits de manière approchée. Pour ce faire, on résout des problèmes elliptiques sur un grand domaine, pour un grand nombre d'échantillons de la loi du coefficient originel (cette construction est donc très coûteuse). Les auteurs de [27, 28] ont entrepris l'étude de deux nouveaux cas déterministes : le cas d'un coefficient périodique, mais perturbé à l'échelle locale par un défaut, et le cas d'une interface entre deux milieux périodiques.

L'objet des articles [27, 28] était de construire des fonctions appelées *correcteurs*. Dans le cas d'un coefficient périodique, ces correcteurs servent à approximer finement le gradient $\nabla u^\varepsilon$ de la solution du problème oscillant (1.1). Dans le cas d'un coefficient périodique avec défaut, un exemple numérique de [27] suggère que ces correcteurs permettent effectivement d'obtenir une bonne approximation du $\nabla u^\varepsilon$. Notre objectif est la démonstration mathématique de cette observation empirique. Nous formalisons un cadre abstrait qui permet de généraliser les résultats d'Avellaneda et Lin [11] et de Kenig, Lin et Shen [94], qui ont été démontrés sous des hypothèses de périodicité. L'*étalement* du défaut, encodé dans son intégrabilité $\mathrm{L}^r(\mathbb{R}^d)$, apparaît alors comme déterminant.

Ce cadre théorique répond à notre but initial puisqu'il englobe le cas des coefficients périodiques perturbés par un défaut tel qu'il est formalisé dans [28]. Mais il permet aussi de retrouver des résultats pour les coefficients quasi-périodiques. Nous avons constaté a posteriori que ce cadre est proche de celui proposé récemment par Gloria, Neukamm et Otto dans [67] et raffiné par Bella, Giunti et Otto dans [17]. Nos travaux diffèrent essentiellement de ceux de Otto et ses coauteurs par l'approche employée, qui reprend la démonstration historique de [11] puis l'article [94], et nos objectifs, qui visent à établir de la régularité lipschitzienne afin de l'employer dans le cadre des défauts –tandis que Otto et ses coauteurs ont à l'esprit un cadre stochastique et démontrent des estimations à plus grande échelle. Par ailleurs, cette étude fait écho à des travaux récents d'Armstrong, Smart, Kuusi et Mourrat (voir [6]).

Nous avons aussi considéré les fonctions de Green associées à l'opérateur $-\mathrm{div}\left(A\left(\frac{\cdot}{\varepsilon}\right)\cdot\nabla\right)$ avec conditions de bord périodiques dans le cas où la matrice $A$ est elle-même périodique. Nous exhibons une décomposition de ces fonctions de Green analogue à celle de la fonction de Green du laplacien sur un domaine périodique (voir [42]), et nous démontrons que les estimations classiques point par point d'Avellaneda et Lin demeurent satisfaites pour ces fonctions de Green.

Voici notre plan pour les sections suivantes : D'abord, nous rappelons brièvement des résultats classiques d'homogénéisation périodique. Puis nous énonçons les résultats obtenus pour le cas d'un coefficient périodique perturbé par un défaut (voir aussi l'Annexe B). Ensuite, nous motivons un cadre plus abstrait dans lequel on peut obtenir des résultats

similaires. Enfin, nous introduisons notre étude sur les fonctions de Green associées à des coefficients et des conditions de bord périodiques. Cette dernière étude est indépendante de celle qui précède mais repose sur les mêmes techniques mathématiques.

### 1.1.1 Homogénéisation périodique

On se place désormais sur $\mathbb{R}^d$, pour $d \geq 3$ (le cas $d = 2$ peut aussi être étudié, mais présente des difficultés techniques liées au fait que la solution fondamentale du Laplacien en dimension 2 ne tend pas vers 0 à l'infini). Soit $\Omega \subset \mathbb{R}^d$ un ouvert borné régulier et $A$ un champ de matrices elliptiques donné. On s'intéresse au problème

$$\begin{cases} -\operatorname{div}\left(A\left(\dfrac{x}{\varepsilon}\right) \cdot \nabla u^\varepsilon(x)\right) = f(x) & \text{dans} \quad \Omega, \\ u^\varepsilon(x) = 0 & \text{sur} \quad \partial\Omega, \end{cases} \tag{1.2}$$

où $f \in \mathrm{L}^2(\Omega)$ est un second membre donné, éventuellement régulier. Ici, $\varepsilon$ est un petit paramètre. La problématique est la suivante : on souhaite approximer précisément $u^\varepsilon$ et $\nabla u^\varepsilon$.

Dans le cas où la matrice $A$ est périodique, il existe un cadre théorique bien établi, dont nous décrivons ici les traits principaux.

Si $A$ est périodique (par abus de langage, on dit "périodique" pour $\mathbb{Z}^d$-périodique), elliptique et bornée, il est bien connu (voir, *e.g.*, [2, Chap. 1, p. 1-15]) que, dans la limite où $\varepsilon \to 0$, le problème (1.2) s'homogénéise en le problème suivant :

$$\begin{cases} -\operatorname{div}\left(A^\star \cdot \nabla u^\star(x)\right) = f(x) & \text{dans} \quad \Omega, \\ u^\star = 0 & \text{sur} \quad \partial\Omega, \end{cases} \tag{1.3}$$

où $A^\star$ est une matrice constante. La matrice $A^\star$ est appelée matrice *homogénéisée* associée à $A$. L'homogénéisation se traduit par le fait que $u^\varepsilon$ converge faiblement vers $u^\star$ dans $\mathrm{H}^1_0(\Omega)$. Mais cette convergence n'est pas forte, sauf dans des cas triviaux. En effet, le gradient $\nabla u^\varepsilon$ de la solution oscille fortement avec une longueur d'onde de l'ordre de $\varepsilon$. Pour approximer le gradient $\nabla u^\varepsilon$, on doit introduire les *correcteurs* $w_j \in \mathrm{H}^1_{\mathrm{loc}}(\mathbb{R}^d)$, pour $j \in [\![1, d]\!]$, relatifs à la matrice $A$. Ils sont définis comme étant les solutions, uniques à l'ajout d'une constante près, de l'équation

$$\begin{cases} -\operatorname{div}\left(A(x) \cdot (e_j + \nabla w_j(x))\right) = 0 & \text{dans} \quad \mathbb{R}^d, \\ \dfrac{|w_j(x)|}{1 + |x|} \underset{|x| \to +\infty}{\to} 0, \end{cases} \tag{1.4}$$

où les vecteurs $e_j$ sont les vecteurs de la base canonique de $\mathbb{R}^d$. Comme $A$ est périodique, les correcteurs sont eux-mêmes périodiques. Cela facilite la résolution numérique de (1.4), qui peut être reformulé comme un problème avec des conditions au bord périodiques. Grâce à ces correcteurs, on construit une approximation $u^{\varepsilon,1}$ de $u^\varepsilon$ dans $\mathrm{H}^1(\Omega)$ définie par

$$u^{\varepsilon,1}(x) := u^\star(x) + \varepsilon \sum_{j=1}^{d} w_j\left(\frac{x}{\varepsilon}\right) \partial_j u^\star(x), \tag{1.5}$$

où $u^\star$ est la solution de (1.3). On peut alors démontrer par des arguments classiques que le reste

$$R^\varepsilon(x) := u^\varepsilon(x) - u^{\varepsilon,1}(x) = u^\varepsilon(x) - u^\star(x) - \varepsilon \sum_{j=1}^{d} w_j\left(\frac{x}{\varepsilon}\right) \partial_j u^\star(x), \tag{1.6}$$

satisfait $R^\varepsilon \to 0$ dans $\mathrm{H}^1(\Omega)$. On peut même majorer la vitesse de convergence (voir [85, (1.51) p. 28]) :

$$\|\nabla R^\varepsilon\|_{\mathrm{L}^2(\Omega)} \leq C\varepsilon^{1/2}. \tag{1.7}$$

Le contrôle sur l'approximation de $\nabla u^\varepsilon$ ci-dessus peut être encore raffiné, et les articles [11, 94] permettent d'approximer $u^\varepsilon$ dans $\mathrm{W}^{1,\infty}(\Omega)$, avec une estimation quantitative en $\varepsilon$ de l'erreur. Pour arriver à un tel résultat, deux grandes étapes théoriques sont nécessaires : une étape d'*estimation* (voir [11]) et une étape d'*approximation* (voir [94]).

En effet, dès lors que $A$ est elliptique, périodique et hölderienne, Avellaneda et Lin ont démontré dans [11] que l'on peut obtenir des estimations fines sur $u^\varepsilon$ solution de

$$-\mathrm{div}\left(A(x/\varepsilon)\cdot\nabla u^\varepsilon(x)\right) = \mathrm{div}(H(x)), \tag{1.8}$$

où $H$ est une fonction vectorielle quelconque, d'une certaine régularité ou d'une certaine intégrabilité. Ces estimations *uniformes en* $\varepsilon$ sont d'abord hölderiennes, puis lipschitziennes (pour peu que $H$ soit suffisamment régulier). Pour ce faire, ils développent une méthode originale de compacité faisant usage de la propriété d'homogénéisation de $A(\cdot/\varepsilon)$ (ou H-convergence, voir [2, Def. 1.2.15 p. 25]). Grâce à ces estimations lipschitziennes, ils parviennent à borner les gradients $\nabla_x G^\varepsilon$, $\nabla_y G^\varepsilon$ et le gradient croisé $\nabla_x\nabla_y G^\varepsilon$ de la fonction de Green $G^\varepsilon$ de (1.2) sur $\mathbb{R}^d$, comme si $G^\varepsilon$ était la solution fondamentale de l'équation de Laplace sur $\mathbb{R}^d$ (voir [64, Chap. II, p. 13-30]). C'est à dire qu'il existe une constante $C$ telle que, pour tous $x \neq y$,

$$|\nabla_x G^\varepsilon(x,y)| \leq C|x-y|^{-d+1}, \qquad |\nabla_y G^\varepsilon(x,y)| \leq C|x-y|^{-d+1}, \tag{1.9}$$

$$|\nabla_x\nabla_y G^\varepsilon(x,y)| \leq C|x-y|^{-d}. \tag{1.10}$$

Dans un second temps, les auteurs de [94] ont utilisé les résultats d'estimation précédents pour contrôler $R^\varepsilon$ dans des normes fines, avec un taux optimal en $\varepsilon$ (à des facteurs en $\log(\varepsilon)$ près). Pour ce faire, ils doivent introduire des correcteurs adaptés au domaine $\Omega$ étudié –ces nouveaux correcteurs satisfont des conditions de bord de Dirichlet au bord du domaine $\Omega$ (voir la Section 2.2.5). En effet, la fonction $u^{\varepsilon,1}$ définie par (1.5) n'est pas nulle sur le bord du domaine $\Omega$, alors que la fonction $u^\varepsilon$ est nulle sur le bord du domaine $\Omega$ : pour cette raison, le gradient $\nabla u^{\varepsilon,1}$ n'approxime par correctement le gradient $\nabla u^\varepsilon$ jusqu'au bord. Mais, en remplaçant les correcteurs dans (1.5) par ces nouveaux correcteurs, la fonction $u^{\varepsilon,1}$ ainsi obtenue satisfait des conditions de Dirichlet au bord de $\Omega$. Ainsi, ils démontrent que, modulo le fait de prendre les correcteurs adaptés décrits ci-dessus, si $f$ est suffisamment régulière, on obtient l'estimation suivante : $\|\nabla R^\varepsilon\|_{\mathrm{L}^\infty(\Omega)} \leq C\varepsilon\ln\varepsilon$.

Ils utilisent ensuite cette estimation pour approximer les gradients et le gradient croisé de la fonction de Green $G^\varepsilon$ associée au problème (1.2) grâce à la fonction de Green $G^\star$ associée au problème (1.3), convenablement combinée avec les correcteurs $w_j$ (voir [94, Th. 3.6 et Th. 3.11]). Outre les résultats de [11], l'ingrédient essentiel des démonstrations de [94] est le fait que les correcteurs $w_j$ et un potentiel $B$ sont bornés. Un potentiel $B_k^{ij}$ est une solution antisymétrique en $i$ et $j$ de l'équation suivante :

$$\sum_{i=1}^{d} \partial_i B_k^{ij} = M_k^j(x) := A_{jk}^\star - \sum_{i=1}^{d} A_{ji}(x)\left(\delta_{ik} + \partial_i w_k(x)\right).$$

En réalité, le caractère périodique de la matrice $A$ n'est pas nécessaire pour montrer des estimations lipschitziennes uniformes en $\varepsilon$ sur la solution $u^\varepsilon$ de (1.8). Par exemple, il est souligné dans [11] que de telles estimations sont aussi vraies si $A$ est quasi-périodique. Notre objectif est de formaliser un cadre dans lequel de telles estimations sont encore vraies.

### 1.1.2 Homogénéisation d'un matériau périodique avec défaut

A l'instar de [27, 28], nous considérons tout d'abord un matériau qui présente à l'échelle microscopique une structure périodique perturbée localement par un "défaut" (voir par exemple la Figure 1.1). On se convainc aisément que le comportement macroscopique d'un tel matériau est dicté par sa structure périodique sous-jacente. Toutefois, si on cherche à recouvrer des informations *uniformes*, notamment au voisinage du défaut, ce dernier ne peut plus être négligé.

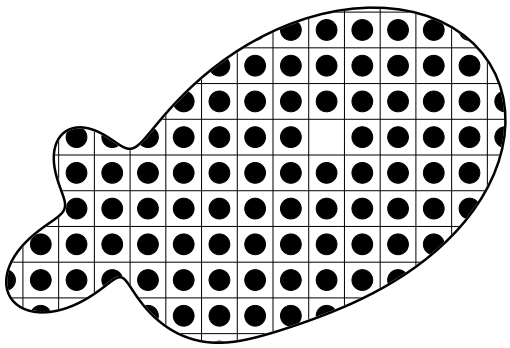


FIGURE 1.1 – Exemple stylisé de matériau présentant une structure microscopique périodique perturbée par un défaut

Les auteurs de [27] ont démontré dans un cadre hilbertien (c'est à dire pour un défaut d'intégrabilité $\mathrm{L}^2\left(\mathbb{R}^d\right)$) qu'il est en effet nécessaire de construire un *correcteur* prenant en

compte le défaut pour approximer correctement le comportement des solutions uniformément sur le domaine. Dans [28], ces mêmes auteurs ont généralisé cette construction au cas d'un défaut d'intégrabilité $\mathrm{L}^r\left(\mathbb{R}^d\right)$, pour $r \in [1, +\infty[$. Toutefois, il demeurait la question de l'utilité dudit correcteur pour approximer la solution de l'équation. C'est l'objet de cette étude : démontrer que "le correcteur corrige", en exhibant les taux de convergence.

Les auteurs de [25, 27, 28] ont étudié le cas d'un coefficient

$$A = A_{\mathrm{per}} + \widetilde{A}, \tag{1.11}$$

somme d'un coefficient périodique $A_{\mathrm{per}} \in \mathrm{L}^\infty\left(\mathbb{R}^d, \mathbb{R}^{d\times d}\right)$ et d'un défaut $\widetilde{A}$ borné et d'intégrabilité donnée

$$\widetilde{A} \in \mathrm{L}^r\left(\mathbb{R}^d, \mathbb{R}^{d\times d}\right), \tag{1.12}$$

pour $r \in [1, +\infty[$. Les coefficients sont supposés bornés et elliptiques : il existe une constante $\mu > 0$ telle que

$$\mu^{-1}|\xi|^2 \leq \left(\widetilde{A} + A_{\mathrm{per}}\right)(x)\cdot\xi\cdot\xi \leq \mu|\xi|^2 \quad \text{et} \quad \mu^{-1}|\xi|^2 \leq A_{\mathrm{per}}(x)\cdot\xi\cdot\xi \leq \mu|\xi|^2, \tag{1.13}$$

pour tous $x, \xi \in \mathbb{R}^d$. On fait ensuite l'hypothèse de régularité suivante : il existe $\alpha > 0$ tel que

$$A_{\mathrm{per}}, \widetilde{A} \in \mathrm{C}^{0,\alpha}_{\mathrm{unif}}\left(\mathbb{R}^d, \mathbb{R}^{d\times d}\right). \tag{1.14}$$

L'espace $\mathrm{C}^{0,\alpha}_{\mathrm{unif}}\left(\mathbb{R}^d, \mathbb{R}^{d\times d}\right)$ englobe l'ensemble des champs de matrices $A$ continus sur $\mathbb{R}^d$ à valeur dans $\mathbb{R}^{d\times d}$ tels que

$$\sup_{x\neq y\in\mathbb{R}^d} |x-y|^{-\alpha}\,|A(x)-A(y)| + \sup_{x\in\mathbb{R}^d}|A(x)| < +\infty.$$

Dans ce cadre, Blanc, Le Bris et Lions ont démontré l'existence de correcteurs $w_j$ associés à $A$ (voir [25, 28]).

Comme le défaut $\widetilde{A}$ est petit à l'échelle macroscopique, il n'affecte pas l'homogénéisation qui a lieu pour les matrices périodiques $A_{\mathrm{per}}(\cdot/\varepsilon)$ quand $\varepsilon \to 0$. Par conséquent, si $u^\varepsilon$ est la solution de (1.2), alors $u^\varepsilon \rightharpoonup u^\star$ dans $\mathrm{H}^1(\Omega)$, pour $u^\star$ satisfaisant (1.3) et $A^\star$ étant la matrice homogénéisée relative à $A_{\mathrm{per}}$. Cependant, comme le défaut est grand à l'échelle microscopique, il joue un rôle non négligeable lorsqu'on s'intéresse à la quantité $\nabla u^\varepsilon$ à proximité du défaut. Cet argument, développé dans [27], démontre que l'on ne peut se contenter de correcteurs $w_i^{\mathrm{per}}$ correspondant seulement au champ périodique $A_{\mathrm{per}}$, mais qu'il faut utiliser les correcteurs $w_j$ associés à $A$.

En adaptant la démarche de [11, 94], on peut quantifier dans quelle mesure le gradient $\nabla u^{\varepsilon,1}$, pour $u^{\varepsilon,1}$ défini par (1.5), approxime $\nabla u^\varepsilon$ :

**Théorème 1.1.1.** *Soient $r \in [1, +\infty[$, avec $r \neq d$, et $\nu_r$ défini par*

$$\nu_r := \min\left(1, \frac{d}{r}\right) \in ]0, 1]. \tag{1.15}$$

*Supposons que $A$, $A_{\rm per}$ et $\widetilde{A}$ satisfont* (1.11), (1.12), (1.13) *et* (1.14), *et que $A_{\rm per}$ est périodique. Soient $\Omega$ un domaine borné régulier de classe $\mathrm{C}^{1,1}$ et $\Omega_1 \subset\subset \Omega$. Soient $f \in \mathrm{L}^2(\Omega)$ et $u^\varepsilon, u^\star, R^\varepsilon$ respectivement définies par* (1.2), (1.3), *et* (1.6).

*(i) Alors $R^\varepsilon \in \mathrm{H}^1(\Omega)$ et*

$$\|R^\varepsilon\|_{\mathrm{L}^2(\Omega)} \leq C\varepsilon^{\nu_r} \|f\|_{\mathrm{L}^2(\Omega)},$$

*où $C$ ne dépend que de $d$, $A$, $r$, et $\Omega$. De plus*

$$\|\nabla R^\varepsilon\|_{\mathrm{L}^2(\Omega_1)} \leq C\varepsilon^{\nu_r} \|f\|_{\mathrm{L}^2(\Omega)},$$

*où $C$ ne dépend que de $d$, $A$, $r$, $\Omega$, et $\Omega_1$.*

*(ii) Si $f \in \mathrm{L}^p(\Omega)$, pour $p \geq 2$, alors $R^\varepsilon \in \mathrm{W}^{1,p}(\Omega)$ et*

$$\|\nabla R^\varepsilon\|_{\mathrm{L}^p(\Omega_1)} \leq C\varepsilon^{\nu_r} \|f\|_{\mathrm{L}^p(\Omega)},$$

*où $C$ ne dépend que de $d$, $A$, $r$, $p$, $\Omega$, et $\Omega_1$.*

*(iii) Enfin, pour tout $\beta \in ]0,1[$, si $f \in \mathrm{C}^{0,\beta}\left(\bar{\Omega}\right)$, on a $R^\varepsilon \in \mathrm{W}^{1,\infty}_{\rm loc}(\Omega)$ et*

$$\|\nabla R^\varepsilon\|_{\mathrm{L}^\infty(\Omega_1)} \leq C\varepsilon^{\nu_r} \ln\left(2+\varepsilon^{-1}\right) \|f\|_{\mathrm{C}^{0,\beta}(\Omega)},$$

*où $C$ ne dépend que de $d$, $A$, $r$, $\beta$, $\Omega$, et $\Omega_1$.*

Un point mis en lumière par le Théorème 1.1.1 est l'existence de deux régimes suivant l'étalement du défaut, c'est à dire l'exposant $r$ dans (1.12). Dans le premier cas, si le défaut est "peu étalé", c'est à dire que $r < d$, alors tout se passe comme dans le cas périodique, en ce sens que les taux en $\varepsilon$ des estimations du Théorème 1.1.1 demeurent les mêmes. Ceci est dû au fait que les correcteurs sont bornés (comme dans le cas périodique). En revanche, si le défaut est "très étalé", c'est à dire que $r > d$, alors les approximations obtenues grâce aux correcteurs sont moins précises que dans le cas périodique, et ce, d'autant moins que $r$ est grand. C'est dû au fait que les correcteurs ne sont plus bornés, mais seulement fortement sous-linéaire à l'infini, c'est à dire qu'ils vérifient

$$\max_{j\in[\![1,d]\!]} \sup_{x\neq y} \frac{|w_j(x)-w_j(y)|}{|x-y|^{\nu_r}} < +\infty.$$

Dans le cas $r = d$, on ne sait pas si les correcteurs $w_j$ sont bornés (voir [28]), d'où l'impossibilité de faire la démonstration comme dans les autres cas. Mais on peut se ramener (de façon sous-optimale) aux résultats prouvés pour $r > d$.

Le théorème ci-dessus permet de quantifier l'approximation $\nabla u^\varepsilon$ à proximité du défaut grâce à une application immédiate de l'inégalité de Hölder (la notation $\mathrm{B}(x,R)$ désigne la boule de centre $x$ et de rayon $R$) :

**Corollaire 1.1.2.** *Soient $2 \leq q \leq p < +\infty$. Sous les hypothèses du Théorème 1.1.1, si $f \in \mathrm{L}^p(\Omega)$, alors, pour tout $\varepsilon$ tel que $\mathrm{B}(0,2\varepsilon) \subset \Omega_1 \subset\subset \Omega$, on a*

$$\left(\frac{1}{|\mathrm{B}(0,\varepsilon)|}\int_{\mathrm{B}(0,\varepsilon)} |\nabla R^\varepsilon|^q\right)^{1/q} \leq C\varepsilon^{\nu_r-\frac{d}{p}} \|f\|_{\mathrm{L}^p(\Omega)}, \tag{1.16}$$

*où $C$ est une constante ne dépendant que de $d$, $A$, $r$, $p$, $\Omega$, et $\Omega_1$.*

Cette approximation se dégrade quand l'intégrabilité du second membre $f$ diminue, ce qui précise le résultat [27, Lem. 2].

La preuve du Théorème 1.1.1 fait naturellement intervenir la fonction de Green de Dirichlet $G^\varepsilon(x,y)$ associée à l'opérateur $-\mathrm{div}\left(A\left(\frac{\cdot}{\varepsilon}\right)\cdot\nabla\right)$ sur un domaine borné régulier. Cette fonction est nulle au bord de ce domaine et satisfait au sens des distributions l'équation aux dérivées partielles suivante :

$$-\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla_x G^\varepsilon(x,y)\right)=\delta_y(x).$$

On peut démontrer des résultats analogues aux inégalités (1.20) de [94] :

**Théorème 1.1.3.** *Soient $r\in[1,+\infty[$, avec $r\neq d$, et $\nu_r$ défini par (1.15). Supposons que $A$, $A_{\mathrm{per}}$ et $\widetilde{A}$ satisfont (1.11), (1.12), (1.13) et (1.14), et que $A_{\mathrm{per}}$ est périodique. Soit $\Omega$ un domaine borné régulier de classe $\mathrm{C}^{1,\beta}$ pour $\beta>0$. Soient $G^\varepsilon$ la fonction de Green de Dirichlet sur $\Omega$ de $-\mathrm{div}\left(A\left(\frac{\cdot}{\varepsilon}\right)\cdot\nabla\right)$. Alors il existe une constante $C$ ne dépendant que de $d$, $A$, $r$, et $\Omega$, telle que, pour tout $\varepsilon\in]0,1[$, on a les estimations suivantes :*

$$\begin{aligned}
&|\nabla_x G^\varepsilon(x,y)|\leq C|x-y|^{1-d} && \forall x\neq y\in\Omega,\\
&|\nabla_y G^\varepsilon(x,y)|\leq C|x-y|^{1-d} && \forall x\neq y\in\Omega,\\
&|\nabla_x\nabla_y G^\varepsilon(x,y)|\leq C|x-y|^{-d} && \forall x\neq y\in\Omega.
\end{aligned}$$

En outre, par une adaptation directe de [94, Th. 3.3, Th. 3.6, et Th. 3.11], il est possible d'approximer $G^\varepsilon$ ainsi que ses gradients et gradients croisés :

**Théorème 1.1.4.** *Sous les hypothèses du Théorème 1.1.3, si $\Omega$ est de classe $\mathrm{C}^{1,1}$, il existe une constante $C$ ne dépendant que de $d$, $A$, $r$, et $\Omega$, telle que*

$$|G^\varepsilon(x,y)-G^\star(x,y)|\leq C\frac{\varepsilon^{\nu_r}}{|x-y|^{d-2+\nu_r}}\qquad\forall x\neq y\in\Omega,$$

*et il existe une constante $C$ ne dépendant que de $d$, $A$, $r$, $\Omega$ et $\Omega_1\subset\subset\Omega$, telle que les estimations suivantes sont satisfaites : pour tout $i\in[\![1,d]\!]$,*

$$\begin{aligned}
&\left|\partial_{x_i}G^\varepsilon(x,y)-\sum_{j=1}^d\left(\delta_{ij}+\partial_i w_j\left(\frac{x}{\varepsilon}\right)\right)\partial_{x_j}G^\star(x,y)\right|\\
&\leq C\varepsilon^{\nu_r}\frac{\ln\left(2+\varepsilon^{-1}\right)}{|x-y|^{d-1+\nu_r}}\qquad\forall x\in\Omega_1,\forall y\in\Omega,x\neq y,
\end{aligned}$$

*et, pour tout $i\in[\![1,d]\!]$,*

$$\begin{aligned}
&\left|\partial_{y_i}G^\varepsilon(x,y)-\sum_{j=1}^d\left\{\delta_{ij}+\partial_i w_j^T\left(\frac{y}{\varepsilon}\right)\right\}\partial_{y_j}G^\star(x,y)\right|\\
&\leq C\varepsilon^{\nu_r}\frac{\ln\left(2+\varepsilon^{-1}\right)}{|x-y|^{d-1+\nu_r}}\qquad\forall x\in\Omega,\forall y\in\Omega_1,x\neq y,
\end{aligned}$$

*et enfin, pour tous* $i, j \in [\![1, d]\!]$,

$$\left| \partial_{x_i} \partial_{y_j} G^\varepsilon(x,y) - \sum_{k,l=1}^{d} \left( \delta_{ik} + \partial_i w_k \left( \frac{x}{\varepsilon} \right) \right) \partial_{x_k} \partial_{y_l} G^\star(x,y) \left( \delta_{lj} + \partial_j w_l^T \left( \frac{y}{\varepsilon} \right) \right) \right|$$
$$\leq C \varepsilon^{\nu_r} \frac{\ln\left(2 + \varepsilon^{-1}\right)}{|x-y|^{d+\nu_r}} \qquad \forall x, y \in \Omega_1, x \neq y.$$

Dans les estimations ci-dessus, la notation $w_i^T$ désigne les correcteurs associés à la matrice transposée $A^T$.

### 1.1.3 Un cadre abstrait pour l'estimation et l'approximation des solutions d'un problème oscillant

Les démonstrations des articles [11,94] peuvent en fait être adaptées dans un cadre théorique dépassant largement le cas de coefficients périodiques, ou de coefficients périodiques perturbés par un défaut. Ainsi, nous avons construit un ensemble d'hypothèses abstraites permettant la démonstration d'estimations lipschitziennes sur un problème elliptique à coefficients oscillants (à la manière de [11]). Ces hypothèses portent sur le champ de matrices à homogénéiser ainsi que sur les correcteurs associés et sur le potentiel $B$ (défini plus bas dans cette Section). Elles englobent naturellement les cas des coefficients périodiques, des coefficients périodiques avec un défaut et des coefficients quasi-périodiques (sous des hypothèses adéquates). Le cadre que nous proposons est proche de celui d'Otto et coauteurs (voir [17,67]). Nous commenterons plus précisément les similitudes et les différences dans le Chapitre 2.

Rappelons la raison pour laquelle $u^{\varepsilon,1}$ défini par (1.5) approxime efficacement $u^\varepsilon$ défini par (1.2) : cela repose sur un argument algébrique. En effet, supposons que $A(\cdot/\varepsilon)$ admet une matrice $A^\star$ *constante* pour matrice homogénéisée (si $A^\star$ n'est pas constante, il faut modifier la définition des correcteurs pour que la conclusion du calcul ci-dessous demeure valide, voir [28]). Posons

$$M_k^i(x) := A_{ik}^\star - \sum_{j=1}^{d} A_{ij}(x) \left( \delta_{jk} + \partial_j w_k(x) \right). \tag{1.17}$$

Par définition des correcteurs $w_j$, $M_k^i$ défini par (1.17) est un champ de vecteurs à divergence nulle, c'est à dire que

$$\operatorname{div}(M_k) = 0, \qquad \forall k \in [\![1, d]\!].$$

A partir de $M_k^i$, on peut construire un potentiel $B_k^{ij}$ antisymétrique par rapport à ses indices $i$ et $j$ satisfaisant

$$\Delta B_k^{ij} = -\partial_j M_k^i + \partial_i M_k^j \qquad \forall i, j, k \in [\![1, d]\!], \tag{1.18}$$

$$M_k^j = \sum_{i=1}^{d} \partial_i B_k^{ij} \qquad \forall j, k \in [\![1, d]\!]. \tag{1.19}$$

En dimension $d = 3$, $M_k^j$ prend ainsi la forme d'un rotationnel.

En suivant [85, p. 26], on observe que si $u^\varepsilon$, $u^\star$, $R^\varepsilon$ sont respectivement définies par (1.2), (1.3) et (1.6), alors,

$$-\operatorname{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla R^\varepsilon(x)\right)=\operatorname{div}\left(H^\varepsilon(x)\right), \tag{1.20}$$

où

$$H_i^\varepsilon(x):=\sum_{j,k=1}^{d}\left(\varepsilon A_{ij}\left(\frac{x}{\varepsilon}\right)w_k\left(\frac{x}{\varepsilon}\right)-\varepsilon B_k^{ij}\left(\frac{x}{\varepsilon}\right)\right)\partial_{jk}u^\star(x). \tag{1.21}$$

A l'instar de [94], l'identité (1.20) joue un rôle central dans notre approche. En effet, une fois que l'on a démontré des estimations sur le problème oscillant (1.8), on déduit de (1.20) que $\|\nabla R^\varepsilon\|\leq\|H^\varepsilon\|$ dans des normes appropriées.

La forme particulière de $H^\varepsilon$ défini par (1.21) suggère qu'il faut exercer un contrôle sur les quantités $\varepsilon w_j(\cdot/\varepsilon)$ et $\varepsilon B(\cdot/\varepsilon)$ pour obtenir un taux de convergence sur $\|\nabla R^\varepsilon\|$ quantifié en $\varepsilon$. On introduit donc naturellement les hypothèses suivantes :

$$|w_j(x)-w_j(y)|\leq C\,|x-y|^{1-\nu}\qquad\text{pour tous}\quad j\in[\![1,d]\!]\quad\text{et}\quad x,y\in\mathbb{R}^d, \tag{1.22}$$

$$|B(x)-B(y)|\leq C|x-y|^{1-\nu}\qquad\text{pour tous}\quad x,y\in\mathbb{R}^d. \tag{1.23}$$

où $\nu\in]0,1]$ et $C>0$ sont fixés. Dans ce cas, on a par exemple l'estimation suivante :

$$\|H^\varepsilon\|_{\mathrm{L}^p(\Omega)}\leq C\varepsilon^\nu\left\|\nabla^2u^\star\right\|_{\mathrm{L}^p(\Omega)}.$$

Les estimations (1.22) et (1.23) sont contraignantes pour $|x-y|$ grand ; ainsi, elles sont des versions "renforcées" de la notion de sous-linéarité stricte à l'infini. Si la matrice $A$ est périodique, on a $\nu=1$ dans (1.22) et (1.23) : les correcteurs et le potentiel sont bornés. Dans le cas d'une matrice de la forme $A=A_{\mathrm{per}}+\widetilde{A}$ satisfaisant les hypothèses du Théorème 1.1.1, on a $\nu=\nu_r$ dans (1.22) et (1.23) (pour $\nu_r$ défini par (1.15)).

Les estimations sur le problème oscillant (1.8) reposent elles-mêmes sur des estimations lipschitziennes à la manière de [11]. Celles-ci concernent les fonctions $u^\varepsilon$ "harmoniques" pour $A(\cdot/\varepsilon)$, c'est à dire qui satisfont

$$-\operatorname{div}\left(A(x/\varepsilon)\cdot\nabla u^\varepsilon(x)\right)=0\quad\text{dans}\quad\mathrm{B}(0,1). \tag{1.24}$$

Elles affirment alors que

$$\|\nabla u^\varepsilon\|_{\mathrm{L}^\infty(\mathrm{B}(0,1/2))}\leq C\,\|u^\varepsilon\|_{\mathrm{L}^2(\mathrm{B}(0,1))}, \tag{1.25}$$

où la constante $C>0$ est indépendante de $\varepsilon$ (c'est un point crucial). Nous généralisons la preuve des estimations lipschitziennes de [11] en utilisant deux ingrédients fondamentaux :

(i) le fait que les correcteurs $w_j$ soient uniformément strictement sous-linéaires ;

(ii) le fait que, lorsque $\varepsilon$ tend vers 0, le champ de matrices $A(\cdot/\varepsilon)$ s'homogénéise uniformément sur tout l'espace en une matrice $A^\star$ constante.

(Nous donnerons un sens précis à ces notions dans le Chapitre 2.) La démonstration repose sur un argument de compacité. Grâce au processus d'homogénéisation, la solution $u^\varepsilon$ du problème multi-échelle (1.24) hérite des estimations à l'échelle mésoscopique que l'on obtient sur la solution du problème homogénéisé (à coefficient constant).

Il se trouve que les deux points (i) et (ii) ci-dessus sont satisfaits si les correcteurs et le potentiel satisfont les estimations (1.22) et (1.23). Ainsi, ces deux estimations fournissent un cadre théorique dans lequel on peut démontrer des résultats d'approximation fins.

Muni de (1.22) et (1.23), on démontre par exemple le Théorème suivant, dont découle une partie du Théorème 1.1.1 :

**Théorème 1.1.5** (Analogue du Théorème 3.7 de [94]). *Soit $A \in \mathrm{C}^{0,\alpha}_{\mathrm{unif}}\left(\mathbb{R}^d, \mathbb{R}^{d\times d}\right)$ un champ de matrices uniformément elliptiques sur $\mathbb{R}^d$ tel que*

*(i) il existe des correcteurs $w_j$ associés à $A$ définis par (1.4) et satisfaisant (1.22) ;*

*(ii) il existe une matrice constante $A^\star \in \mathbb{R}^{d\times d}$ et un potentiel $B$ (c'est à dire que $B_k^{ij}$ est antisymétrique par rapport à ses indices $i$ et $j$, et c'est une solution de (1.18) et (1.19), pour $M$ défini par (1.17)) qui satisfait (1.23).*

*Soient $\Omega$ un ouvert borné régulier de classe $\mathrm{C}^{1,1}$, de $\mathbb{R}^d$, $\Omega_1 \subset\subset \Omega$, $f \in \mathrm{L}^2(\Omega)$, et $\varepsilon \in ]0,1[$. Soient $u^\varepsilon$, $u^\star$, et $R^\varepsilon$ respectivement définies par (1.2), (1.3), et (1.6).*

*(i) Si $f \in \mathrm{L}^p(\Omega)$, pour $p > d$, alors $R^\varepsilon \in \mathrm{W}^{1,p}(\Omega)$ et*

$$\|\nabla R^\varepsilon\|_{\mathrm{L}^p(\Omega_1)} \leq C\varepsilon^\nu \|f\|_{\mathrm{L}^p(\Omega)}, \tag{1.26}$$

*où $C$ ne dépend que de $d$, $A$, $\nu$, $p$, $\Omega_1$ et $\Omega$.*

*(ii) Si $f \in \mathrm{C}^{0,\beta}\left(\bar{\Omega}\right)$, pour $\beta \in ]0,1[$ alors on a $R^\varepsilon \in \mathrm{W}^{1,\infty}_{\mathrm{loc}}(\Omega)$ et*

$$\|\nabla R^\varepsilon\|_{\mathrm{L}^\infty(\Omega_1)} \leq C\varepsilon^\nu \ln\left(2+\varepsilon^{-1}\right) \|f\|_{\mathrm{C}^{0,\beta}(\Omega)}, \tag{1.27}$$

*où $C$ ne dépend que de $d$, $A$, $\nu$, $\beta$, $\Omega_1$ et $\Omega$.*

De même, on généralise le Corollaire 1.1.2 et les Théorèmes 1.1.3 et 1.1.4 au cadre d'un coefficient elliptique, uniformément hölderien, admettant une matrice homogénéisée constante, dont les correcteurs et le potentiel satisfont (1.22) et (1.23).

### 1.1.4 Estimations sur des fonctions de Green en homogénéisation périodique

Depuis les travaux d'Avellaneda et Lin [11,12], il est connu que si $A$ est une matrice elliptique, périodique et hölderienne alors la fonction de Green de l'opérateur $-\mathrm{div}\left(A(\cdot/\varepsilon)\cdot\nabla\right)$ satisfait les estimations (1.9), (1.10), lorsque le problème est posé sur tout $\mathbb{R}^d$, ou sur un ouvert borné suffisamment régulier, avec conditions de Dirichlet. Plus récemment dans [93], des estimations similaires ont été démontrées pour la fonction de Green des mêmes opérateurs, avec conditions de Neumann. Toutefois, de telles estimations, bien que très vraisemblables, n'ont jamais été démontrées pour le cas de conditions de bord périodiques (à notre connaissance). C'est l'objet de cette étude, qui a aussi été motivée par des considérations numériques issues des travaux de F. Legoll et P.-L. Rothé [100].

Plus précisément, on considère la fonction de Green $G_n(x,y)$ associée au problème oscillant suivant :

$$\begin{cases} -\operatorname{div}\left(A(nx)\cdot\nabla u_n(x)\right) = f(x) - \displaystyle\int_{\mathrm{Q}} f \qquad \text{pour} \quad x\in\mathbb{R}^d, \\ \displaystyle\int_{\mathrm{Q}} u_n = 0 \quad \text{et} \quad u_n \text{ est Q-périodique,} \end{cases} \tag{1.28}$$

où $n\in\mathbb{N}$ est un nombre arbitrairement grand (la petite échelle est $\varepsilon = n^{-1}$), $\mathrm{Q} = [-1/2,1/2]^d$ est le cube unité, et $f$ est une fonction Q-périodique.. La matrice $A$ satisfait les hypothèses habituelles d'ellipticité et de périodicité (voir [11])

$$\mu|\xi|^2 \le A(x)\cdot\xi\cdot\xi \le \mu^{-1}|\xi|^2 \qquad \forall x,\xi\in\mathbb{R}^d, \tag{1.29}$$

$$A(x+z) = A(x) \qquad \forall x\in\mathbb{R}^d, z\in\mathbb{Z}^d. \tag{1.30}$$

En utilisant la méthode de [94, Th. 3.3], on peut établir le théorème suivant, qui est une extension du résultat classique [72, Th. 1.1] :

**Proposition 1.1.6.** *Soit $d\ge 2$. Supposons que $A\in \mathrm{L}^\infty_{\mathrm{per}}\left(\mathrm{Q},\mathbb{R}^{d\times d}\right)$ satisfait* (1.29)*. Soit $G_n$ la fonction de Green associée à l'équation* (1.28)*. Alors, il existe une constante $C>0$ dépendant seulement de $d$ et $\mu$ telle que, pour tous $x\in\mathbb{R}^d$ et $y\in x+\mathrm{Q}$, avec $x\neq y$ :*

$$\text{si } d\ge 3, \qquad |G_n(x,y)| \le C|x-y|^{-d+2}, \tag{1.31}$$

$$\text{si } d=2, \qquad |G_n(x,y)| \le C\log(2+|x-y|). \tag{1.32}$$

Grâce à ce résultat, en utilisant les estimations lipschitziennes de [11], on retrouve les estimations (1.9) et (1.10) :

**Proposition 1.1.7.** *Soit $d\ge 2$. Supposons que $A\in \mathrm{L}^\infty_{\mathrm{per}}\left(\mathrm{Q},\mathbb{R}^{d\times d}\right)$ est un champ de matrices périodique et hölderien qui satisfait* (1.29)*. Soit $G_n$ la fonction de Green associée à l'équation* (1.28)*. Alors, il existe une constante $C>0$ indépendante de $n$ telle que, pour tous $x\in\mathbb{R}^d$ et $y\in x+\mathrm{Q}$, avec $x\neq y$ :*

$$|\nabla_x G_n(x,y)| \le C|x-y|^{-d+1}, \tag{1.33}$$

$$|\nabla_y G_n(x,y)| \le C|x-y|^{-d+1}, \tag{1.34}$$

$$|\nabla_x\nabla_y G_n(x,y)| \le C|x-y|^{-d}. \tag{1.35}$$

En utilisant une autre approche, inspirée de [42, p. 130-131], nous démontrons que l'on peut décomposer la fonction de Green $G_n$ grâce à la fonction de Green $\mathcal{G}_n$ de l'opérateur $-\operatorname{div}\left(A(n\cdot)\cdot\nabla\right)$ sur $\mathbb{R}^d$ :

**Proposition 1.1.8.** *Soit $d\ge 3$. Supposons que $A\in \mathrm{L}^\infty_{\mathrm{per}}\left(\mathrm{Q},\mathbb{R}^{d\times d}\right)$ est un champ de matrices périodique et hölderien qui satisfait* (1.29)*. Soit $G_n$ la fonction de Green associée à*

*l'équation* (1.28)*, et* $\mathcal{G}_n$ *la fonction de Green de l'opérateur* $-\mathrm{div}\,(A(n\cdot)\cdot\nabla)$ *sur* $\mathbb{R}^d$*. Alors, on peut décomposer* $G_n$ *comme*

$$G_n(x,y) = \sum_{m=0}^{+\infty} \left( \sum_{k\in\Gamma_m} H_n^k(x,y) \right), \tag{1.36}$$

*où les fonctions* $H_n^k$ *sont définies par*

$$\begin{aligned} H_n^k(x,y) :=& \mathcal{G}_n(x,y-k) - \int_{\mathrm{Q}} \mathcal{G}_n(x,y+y'-k)\mathrm{d}y' - \int_{\mathrm{Q}} \mathcal{G}_n(x+x',y-k)\mathrm{d}x' \\ &+ \int_{\mathrm{Q}}\int_{\mathrm{Q}} \mathcal{G}_n(x+x',y+y'-k)\mathrm{d}y'\mathrm{d}x'. \end{aligned} \tag{1.37}$$

*et les ensembles* $\Gamma_m$ *par*

$$\Gamma_m = \left\{ k\in\mathbb{Z}^d, 2^m - 1 \le k\cdot(A_{\mathrm{s}}^\star)^{-1}\cdot k < 2^{m+1}-1 \right\},$$

*où* $A_{\mathrm{s}}^\star$ *est la partie symétrique de la matrice homogénéisée* $A^\star$ *associée à la matrice* $A$*.*

Les Propositions 1.1.7 et 1.1.8 se généralisent à des systèmes.

## 1.2 Dislocations

Dans cette section, nous introduisons notre étude théorique et numérique sur certaines équations intégrodifférentielles décrivant des dislocations.

Nous décrivons succinctement les dislocations, qui sous-tendent le comportement plastique des matériaux cristallins. Une façon de modéliser celles-ci a été proposée par Peierls dans [129], dans un modèle hybride qui couple échelle atomique et échelle mésoscopique de la matière. Ce modèle induit plusieurs équations, selon les régimes considérés :

- l'équation classique de Peierls-Nabarro, qui décrit des dislocations statiques,
- l'équation de Weertman (voir [135, 154]), qui décrit des dislocations en mouvement en régime stationnaire,
- l'équation de Peierls-Nabarro Dynamique proposée dans [130], qui décrit des dislocations en régime dynamique.

Nous étudions séparément les deux dernières équations, en nous attachant tout particulièrement à une de leur spécificités mathématiques : leur caractère non-local ou intégrodifférentiel. Le but est de proposer un algorithme permettant d'approximer numériquement leur solution.

### 1.2.1 Les dislocations en physique des matériaux

Cette étude est motivée par des applications en physique des matériaux, dont nous introduisons ici quelques concepts fondamentaux. Les explications qui suivent sont inspirées de [148]. Le lecteur pourra aussi consulter [81].

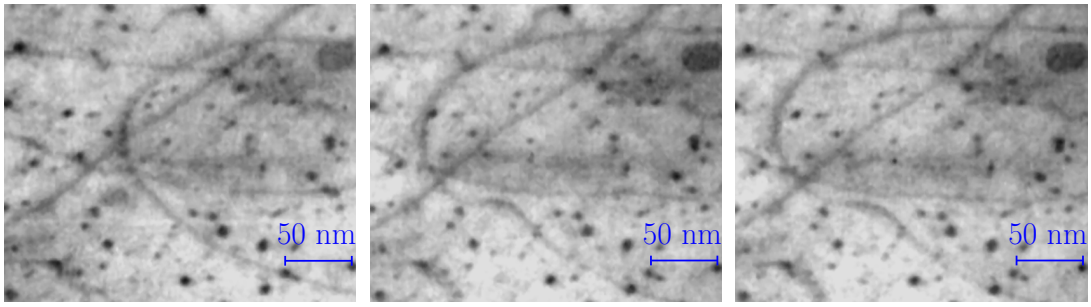



FIGURE 1.2 – Lignes de dislocation en mouvement dans un échantillon de Zircaloy-4 aux temps $t = -11.5$s, $t = 0$s et $t = 8.1$s. Nous remercions L. Dupuy pour avoir fourni les photographies, publiées originellement dans [56].

### Brève description des dislocations

Une éprouvette métallique soumise à un essai de traction passe successivement par deux régimes de déformation. Tout d'abord, elle se déforme de manière *élastique*. Lors de cette première phase, l'allongement de la barre évolue linéairement avec la contrainte mécanique imposée. Cette déformation est réversible : une fois la contrainte relâchée, l'éprouvette revient à son état initial. Puis, si la contrainte mécanique dépasse le seuil de dureté, l'éprouvette se déforme alors de manière *plastique* jusqu'à la rupture. Dans cette seconde phase, la déformation n'évolue plus linéairement avec la contrainte. Si cette dernière est relâchée, l'éprouvette demeure déformée : c'est une transformation irréversible.

Pour comprendre le phénomène de plasticité, il faut se placer à une échelle intermédiaire entre celle de l'éprouvette, et celle de la liaison inter-atomique : l'échelle du réseau cristallin. En effet, les cristaux qui constituent l'éprouvette ne sont pas parfaits ; au contraire, ils sont traversés par des lignes de défauts : les dislocations (voir Figure 1.2). Lors de la phase de déformation plastique, ces dislocations se meuvent, se tordent, nucléent et s'annihilent ; ainsi, le matériau est durablement déformé.

Considérons de plus près l'exemple idéalisé de ligne de dislocation représenté sur la Figure 1.3(a). C'est une dislocation coin. Elle correspond à ajouter virtuellement un demi-plan atomique supplémentaire (les atomes en bleu) au réseau cristallin initial, supposé parfait (représenté en pointillé sur la Figure 1.3(b)). La zone représentée en rouge est appelée *cœur* de la dislocation : c'est l'endroit où le cristal est le plus fortement déformé. Cette zone matérialise la ligne de dislocation. La ligne de dislocation peut se déplacer dans le plan $(Oxz)$, qui est appelé *plan de glissement*.

La présence de cette zone de défaut a des répercussions à longue portée dans le réseau cristallin. En effet, chaque atome du cristal tend à minimiser son énergie. Ainsi, les atomes du cœur de la dislocation sont décalés par rapport au cristal parfait. Mais leur décalage induit encore un décalage des atomes voisins, et ainsi de suite, le décalage d'un atome étant d'autant plus petit que celui-ci est loin du cœur de la dislocation (en l'absence de contrainte extérieure). Par conséquent, les dislocations sont des objets non-locaux. Leur forme est représentée grâce à la fonction de glissement $\boldsymbol{\eta}(x)$ (appelée *slip function* en anglais), qui

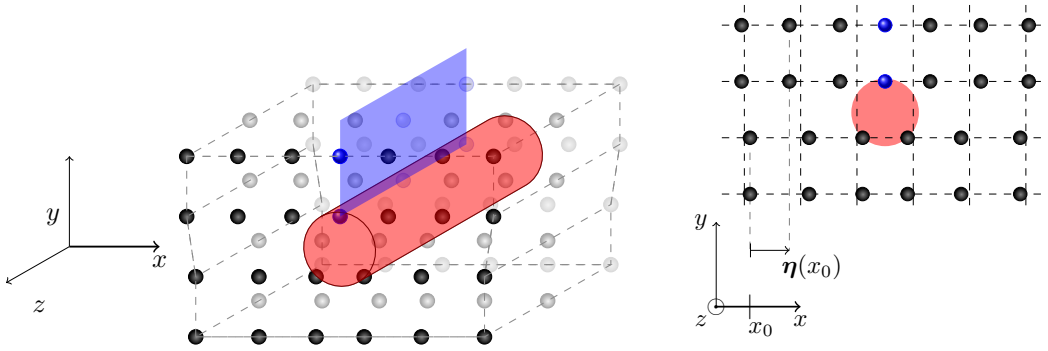



FIGURE 1.3 – (a) Représentation cristallographique d'un dislocation coin (en foncé, les atomes du plan $(x, y, z = 0)$, en bleu le demi-plan atomique surnuméraire), (b) illustration de la définition de la fonction de glissement $\boldsymbol{\eta}$ sur le même cristal, dans un plan orthogonal à la ligne de dislocation.

mesure le déplacement de chaque atome juste au-dessus du plan de glissement par rapport à l'atome qui lui est associé, juste en-dessous du plan de glissement [2]. Le *vecteur de Burgers* $\boldsymbol{b}$ mesure le glissement total

$$\boldsymbol{b} = \boldsymbol{\eta}(-\infty) - \boldsymbol{\eta}(+\infty),$$

et constitue une caractéristique importante de la dislocation. Si le vecteur de Burgers est un vecteur du réseau cristallin, les dislocations sont dites parfaites (comme sur la Figure 1.3) ; dans le cas contraire, on parle de dislocations partielles.

Nous représentons sur la Figure 1.4 un point de vue issu de la mécanique des milieux continus. Une dislocation coin y est induite par un cisaillement qui contraint la partie supérieure du matériau à se déplacer par rapport à la partie inférieure (le déplacement est noté $\boldsymbol{u}(x, y)$). Le profil de la dislocation est alors représenté par la discontinuité de déplacement $\boldsymbol{\eta}(x) = \boldsymbol{u}(x, 0^+) - \boldsymbol{u}(x, 0^-)$ le long de l'interface $(x, y = 0, z)$. Soulignons que les Figures 1.3 et 1.4 représentent le même objet.

Les lignes de dislocations étudiées ici sont rectilignes et de longueur infinie, mais les dislocations réelles se présentent souvent sous la forme de boucles (comme sur la Figure 1.2), le long desquelles la direction du vecteur de Burgers varie. Ainsi, le glissement $\boldsymbol{\eta}$ peut avoir lieu selon plusieurs directions :

— selon l'axe des $x$, on parle alors de dislocation *coin* ;
— selon l'axe des $y$, on parle alors de dislocation *coin de montée* ;
— selon l'axe des $z$, on parle alors de dislocation *vis*.

Par analogie avec la théorie de la rupture, on parlera aussi de mode II, I et III respectivement (voir par exemple [89, Chap. 3 p. 138]). Les dislocations réelles sont généralement

2. stricto sensu, $\boldsymbol{\eta}$ n'est pas une fonction, mais est seulement définie en les positions $x_i$ des atomes. Par la suite, on la considérera cependant comme un fonction dont l'argument est $x \in \mathbb{R}$.

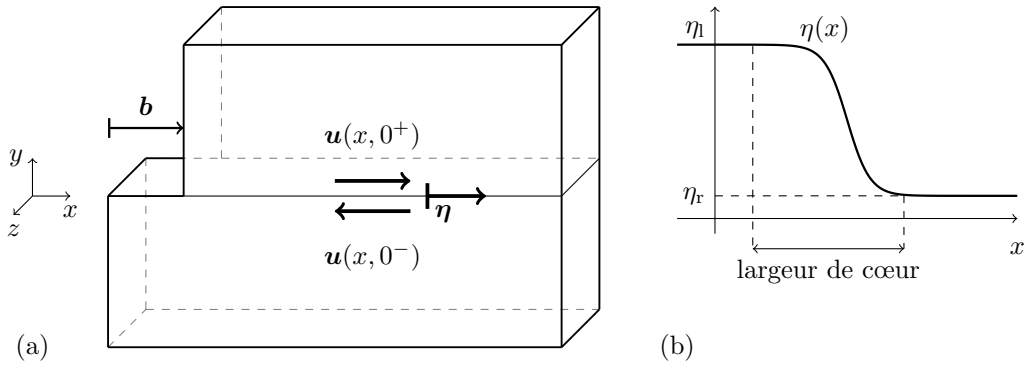



FIGURE 1.4 – (a) Représentation d'une dislocation du point de vue des milieux continus, (b) exemple de fonction $\eta$.

des combinaisons des trois types de dislocations, c'est à dire que la fonction $\boldsymbol{\eta}(t,x)$ est à valeur dans $\mathbb{R}^3$. Toutefois, sauf mention contraire, on se ramène toujours par la suite à un cas scalaire où $\eta$ est à valeur dans $\mathbb{R}$. Et ce, pour deux raisons essentielles :

— une part importante de l'analyse mathématique que nous sommes en mesure de faire requiert l'utilisation d'outils spécifiques aux équations scalaires (notamment le principe du maximum) ;
— d'un point de vue purement numérique, les équations scalaires ainsi construites concentrent une grande partie des difficultés des équations vectorielles associées.

**Méthodes actuelles de simulation**

Par ses nombreuses applications en science de matériaux (notamment en métallurgie) et à cause de la complexité des mécanismes impliqués, la compréhension des phénomènes de plasticité est un enjeu scientifique important. C'est pourquoi de nombreux laboratoires étudient actuellement des modèles multi-échelles de plasticité [38, 159], afin de calibrer des lois de comportement élasto-plastique sur les petites échelles de la matière. La finalité de ces modèles est de fournir à l'ingénieur des codes de calcul par éléments finis pour étudier des structures macroscopiques (dont l'échelle est supérieure au $\mathrm{cm}^3$).

A chaque échelle de la matière correspond un certain nombre de méthodes de simulation.

**Simulation à l'échelle microscopique** A l'échelle atomique, les dislocations se manifestent comme des objets non-locaux. On distingue deux classes principales de méthodes

— les méthodes *ab initio*, qui tendent à résoudre les équations de la physique quantique,
— les méthodes de dynamique moléculaire, où les atomes sont soumis aux lois de la mécanique classique, mais avec des potentiels d'interaction calibrés sur des modèles *ab initio*.

Grâce à ces simulations, on construit le potentiel de $\gamma$-surface (voir [47]), qui traduit les interactions inter-atomiques au niveau du plan de glissement de la dislocation. En outre,

on peut en tirer de la dynamique moléculaire des données essentielles sur la mobilité d'une dislocation individuelle comme la viscosité effective, ou les relations contrainte-vitesse (ou lois de mobilité). Les lois de mobilité sont cruciales, car elles viennent nourrir les codes de simulation à l'échelle mésoscopique décrits ci-dessous. Nos travaux sur les dislocations en mouvement à vitesse constante ont pour objectif principal la prédiction des lois de mobilité à partir du modèle de Peierls.

**Simulation à l'échelle mésoscopique** A l'échelle du $\mu$m$^3$, les dislocations sont représentées comme un réseau de lignes enchevêtrées. Les simulations de Dynamique des Dislocations Discrètes (DDD) font le lien entre l'échelle microscopique et l'échelle macroscopiques (voir [38, 53]). Elles requièrent d'une part des règles spécifiques d'interaction entre dislocations et des relations contrainte-vitesse. Celles-ci sont généralement issues de modèles simples, dont les paramètres sont calibrés à l'aide de simulations et d'expériences physiques.

**Simulation à une échelle mixte** On peut décrire les dislocations comme des objets nonlocaux dans un milieu continu élastique (voir par exemple [51]). L'ingrédient atomistique est fourni par la $\gamma$-surface, et la dislocation se manifeste par des champs qu'elle engendre dans le matériau. Le modèle de Peierls et ses multiples dérivés s'inscrivent dans ce cadre. En général, on simule quelques dislocations dans un petit volume grâce à des éléments finis. Mais il est aussi possible se ramener à la résolution d'équations intégrodifférentielles sur le plan de glissement.

Afin de pouvoir étudier des processus rapides, avec des chocs (c'est à dire des ondes de contrainte se déplaçant rapidement dans le matériau), il est nécessaire de caractériser des aspects élastodynamiques des dislocations. Ceci motive notre travail sur les dislocations en régime dynamique. L'objectif à long terme est d'intégrer dans les simulations DDD des lois de vitesse prenant en compte l'inertie des dislocations (ceci dépasse le cadre de cette thèse).

### Le modèle de Peierls

Les équations étudiées dans cette thèse sont issues d'une catégorie de modèles introduite dans l'article historique de Peierls [129]. Il s'agit de modèles hybrides, qui couplent une description mésoscopique de mécanique des milieux continus (en l'occurrence, l'équation d'élasticité linéaire) et une description microscopique de la matière (où les interactions ont lieu au niveau atomique).

Pour fixer les idées, considérons un matériau homogène isotrope scindé en deux parties (en fait des demi-espaces), qui sont au contact le long d'une interface $y = 0$ (voir Figure 1.5). Cette interface constitue le plan de glissement de la dislocation. Le matériau présente un déplacement $\boldsymbol{u}(t, x, y) \in \mathbb{R}^3$ indépendant de $z$, avec une discontinuité correspondant au glissement $\boldsymbol{\eta}(t, x) := \boldsymbol{u}(x, 0^+) - \boldsymbol{u}(x, 0^-)$ au niveau de l'interface. Dans chacun des demi-espaces environnants, le matériau est soumis à l'équation d'élasticité linéaire homogène $\rho \partial_{tt} \boldsymbol{u} = \operatorname{div}(\boldsymbol{\sigma})$, où $\rho$ est la masse volumique (uniforme) du matériau considéré, et $\boldsymbol{\sigma}$ est le tenseur des contraintes. Ce dernier satisfait la loi de Hooke dans les demi-espaces environnants et dépend donc linéairement de $\nabla \boldsymbol{u}$. Par la loi de Cauchy, on a la

relation $\sigma_{2j}(t,x,0^{\pm}) = f_j(t,x)$ sur l'interface $y = 0$, où $f$ est le chargement du matériau, issu à la fois des forces d'interactions atomiques (via la $\gamma$-surface) et du chargement imposé.

Intuitivement, les seuls degrés de liberté du modèle se réduisent à la discontinuité de déplacement $\boldsymbol{\eta}(t,x)$, qui est suffisante pour donner à $\boldsymbol{u}$ des conditions aux limites. De façon imagée, $\boldsymbol{\eta}(t,x)$ est sous influence de $\boldsymbol{\eta}(t',x')$ pour $t' < t$ via les ondes qui se propagent dans les demi-plans inférieurs et supérieurs (dans un cadre statique, ce sont simplement des interactions à longue portée). Ainsi, l'équation construite sur $\boldsymbol{\eta}$ est naturellement intégro-différentielle (nous renvoyons au Chapitre 6 pour plus de détails sur la modélisation).

Cette approche, originellement conçue pour modéliser une dislocation dans un matériau au repos (en découle alors l'équation classique de Peierls-Nabarro, voir [129]), a ensuite été généralisée par Weertman dans [154] au cas d'une dislocation se propageant à vitesse constante dans un matériau soumis à un chargement uniforme (voir aussi [135]). Récemment, Pellegrini a construit sur ce modèle une équation décrivant le comportement d'une dislocation en régime dynamique : l'équation de Peierls-Nabarro Dynamique (voir [130]).

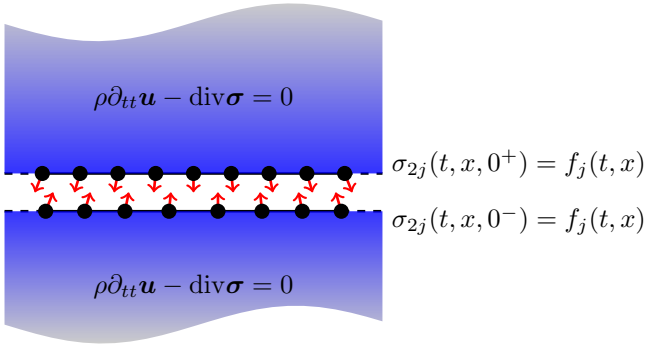



FIGURE 1.5 – Représentation du modèle de Peierls.

### 1.2.2 Dislocations en régime stationnaire

#### L'équation de Weertman

Supposons que la dislocation considérée ne se déforme pas, mais se meut à vitesse constante sous l'effet d'un chargement uniforme $\sigma$. C'est à dire que la fonction $\eta$, supposée à valeur scalaire, est un front progressif

$$\eta(t,x) = \phi(x - vt), \qquad \text{pour} \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R},$$

où $v \in \mathbb{R}$ est une certaine vitesse. Dans ce cas, le modèle de Peierls amène à écrire l'équation de Weertman (voir [135]) sur $\phi$. Convenablement adimensionnée (voir Annexe A.2), cette équation s'écrit sous la forme

$$-|\partial_x|\,\phi(x) + c\phi'(x) = F'(\phi(x)) \qquad \text{pour} \quad x \in \mathbb{R}, \tag{1.38}$$

où $F$ est un potentiel non-linéaire, et où intervient un opérateur intégrodifférentiel $|\partial_x|$. Cet opérateur, dont le symbole en Fourier est $|k|$, est aussi appelé *laplacien fractionnaire* et noté $(-\Delta)^{1/2}$. Il peut s'exprimer comme

$$|\partial_x|\,\phi(x) = -\frac{1}{\pi}\int_0^{+\infty} \frac{\phi'(x+y)-\phi'(x-y)}{y}\mathrm{d}y. \tag{1.39}$$

La formule ci-dessus illustre le caractère *non-local* de $|\partial_x|$, qui constitue une difficulté importante de (1.38).

A l'infini à droite et à gauche, les atomes sont décalés de manière parfaite et sont dans une position stable. Cela se traduit par le fait que la solution $\phi$ de (1.38) satisfait les conditions à l'infini

$$\lim_{x\to-\infty}\phi(x) = \phi_\mathrm{l} \quad \text{et} \quad \lim_{x\to+\infty}\phi(x) = \phi_\mathrm{r}, \tag{1.40}$$

où les positions $\phi_\mathrm{r}$ et $\phi_\mathrm{l}$ sont des minimiseurs locaux de $F$

$$F'(\phi_\mathrm{r}) = F'(\phi_\mathrm{l}) = 0, \tag{1.41}$$
$$F''(\phi_\mathrm{l}) > 0 \quad \text{et} \quad F''(\phi_\mathrm{r}) > 0. \tag{1.42}$$

On parle alors de potentiel *bistable*. Dans le cas particulier où $\sigma = 0$, on montre que $v = 0$ et on retrouve alors la célèbre équation de Peierls-Nabarro [129].

Soulignons que, dans l'équation (1.38), à la fois la fonction $\phi$ *et* le scalaire $c$ sont des inconnues. Ce fait important a des conséquences à la fois théoriques et numériques. L'application qui à $v$ associe $c$ (voir (A.2) dans l'Annexe A.2) est surjective sur $\mathbb{R}$, mais pas injective en général. Par conséquent, à une unique solution $(\phi, c)$ de (1.38) correspondent plusieurs vitesses $v$ de dislocation possibles.

### Quelques propriétés mathématiques de l'équation de Weertman

Des questions naturelles se posent vis-à-vis de l'équation de Weertman : existe-t-il des solutions $(\phi, c)$ à (1.38) et (1.40) ? Sont-elles uniques ? Comment les caractériser ? Comment les approximer numériquement ?

L'opérateur $|\partial_x|$, tout comme l'opérateur $-\Delta$, est un opérateur symétrique positif, diagonalisé par la transformation de Fourier. Par conséquent, l'équation (1.38) est conceptuellement proche du problème classique

$$\Delta\phi(x) + c\partial_x\phi(x) = F'(\phi(x)) \qquad \text{pour} \quad x \in \mathbb{R}, \tag{1.43}$$

couplé avec (1.40), dans le cas où $F$ est un potentiel bistable. En se ramenant à l'étude du portrait de phase d'un système autonome en dimension $d = 2$, il est facile d'établir l'existence d'une solution $(\phi, c)$ à l'équation (1.43) (voir par exemple [151, Th. 1.5 p. 208]). Une telle manipulation est impossible dans le cas de l'équation (1.38) : il est nécessaire de recourir à des propriétés plus subtiles de l'opérateur $|\partial_x|$.

Au cours des cinq dernières années, la compréhension des équations à laplacien fractionnaire du type de (1.38) a beaucoup progressé (voir notamment [40,41]). Dans [73] (voir aussi l'article de Chmaj [46]), Gui et Zhao ont étudié une large classe d'équations de réaction-diffusion du type de (1.38), où l'opérateur différentiel n'est plus $|\partial_x|$ mais un laplacien fractionnaire général $|\partial_x|^\alpha$ (de symbole $|k|^\alpha$), avec $\alpha \in ]0,2[$. Leurs résultats impliquent le :

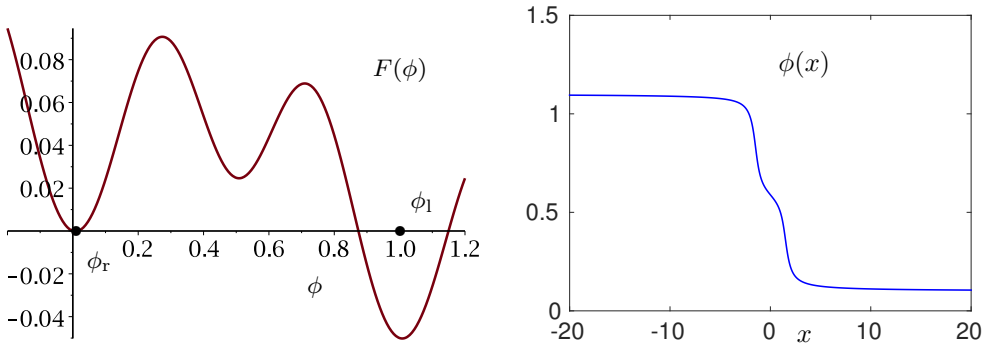
**Théorème 1.2.1** (Cas particulier du Théorème 1.1 de [73]). *Soit $\phi_l < \phi_r$. Soit un potentiel $F \in C^3(\mathbb{R})$ satisfaisant (1.41) et (1.42). Supposons que $F$ est tel que :*

$$F(\phi_l) < F(u), \qquad \text{pour tout } u \in ]\phi_r, \phi_l[, \tag{1.44}$$

$$F'(u) < 0, \qquad \text{pour tout } u \text{ tel que } F(u) < F(\phi_r). \tag{1.45}$$

*Alors, il existe une fonction $\phi$ monotone et unique à translation près, et un unique scalaire $c \in \mathbb{R}$ satisfaisant (1.38) et (1.40).*

Les conditions (1.41), (1.42), (1.44) et (1.45) induisent que le potentiel $F$ est bistable et admet en $\phi_l$ et $\phi_r$ des minima locaux. En outre, tous les points critiques de $F$ entre $\phi_r$ et $\phi_l$ sont situés au-dessus de ces derniers. Ces hypothèses sont physiquement réalistes. Les Figures 1.6 (a) et (b) illustrent un exemple de potentiel bistable $F$ et une approximation numérique de la solution $\phi$ de (1.38) associée.




(a) (b)

FIGURE 1.6 – (a) Un potentiel bistable $F$ "en bosse de chameau" (voir (5.54) plus bas), (b) reconstruction numérique de la solution $\phi(x)$ de (1.38) associée.

Dans le cas classique du laplacien, il est bien connu que l'équation (1.43) décrit les fronts progressifs $u(t,x) = \phi(x - ct)$ de l'équation de réaction-diffusion suivante :

$$\partial_t u(t,x) = \Delta u(t,x) - F'(u(t,x)) \qquad \text{pour} \quad x \in \mathbb{R}. \tag{1.46}$$

L'article historique [60] démontre la stabilité globale de ces fronts progressifs pour (1.46) si $F$ est bistable et ne présente pas de minimum intermédiaire entre $\phi_r$ et $\phi_l$. C'est à dire que

toute solution de (1.46) avec une donnée initiale $u(0,\cdot)$ "raisonnable" et satisfaisant (1.40) converge en temps long vers un front progressif. En réutilisant des ingrédients de [60], Chen a démontré dans [45] qu'une telle convergence avait lieu pour des opérateurs diffusifs très généraux, à la condition qu'ils satisfassent un principe de comparaison (c'est le cas de l'opérateur $\partial_t - \Delta$) ainsi que des hypothèses techniques.

Comme l'opérateur $\partial_t + |\partial_x|$ satisfait aussi un principe de comparaison, il est donc légitime de se demander dans quelle mesure l'équation (1.38) décrit les fronts progressifs de l'équation suivante :

$$\begin{cases} \partial_t u(t,x) + |\partial_x|\, u(t,x) = -F'(u(t,x)) & \text{pour} \quad x \in \mathbb{R}, \\ u(0,x) = u_0(x) & \text{pour} \quad x \in \mathbb{R}. \end{cases} \tag{1.47}$$

Nous avons établi le théorème suivant :

**Théorème 1.2.2.** *Soit $\phi_{\rm l} > \phi_{\rm r}$, et $F \in \mathrm{C}^3(\mathbb{R})$ satisfaisant* (1.41) *et* (1.42). *Supposons que*

$$u \in \mathrm{C}\left((0,+\infty), \mathrm{C}^2(\mathbb{R}) \cap \mathrm{W}^{2,\infty}(\mathbb{R})\right) \cap \mathrm{C}^1\left((0,+\infty), \mathrm{C}(\mathbb{R})\right) \tag{1.48}$$

*est une solution de* (1.47) *dont la donnée initiale $u_0$ est à valeur dans $[\phi_{\rm l}, \phi_{\rm r}]$ et satisfait* (1.40).

*Si $(\phi, c)$ est une solution de* (1.38) *et* (1.40), *où $\phi \in \mathrm{C}^2(\mathbb{R})$ est une fonction décroissante satisfaisant $\phi' < 0$ et*

$$\lim_{|x|\to+\infty} \phi'(x) = 0, \tag{1.49}$$

*alors, il existe des constantes $\kappa > 0$, $K > 0$ et $\xi \in \mathbb{R}$ telles que*

$$\|u(t,\cdot) - \phi(\cdot - ct + \xi)\|_{\mathrm{L}^\infty(\mathbb{R})} \leq K \mathrm{e}^{-\kappa t}, \tag{1.50}$$

*pour tout $t \in \mathbb{R}_+$.*

Le Théorème 1.2.2 est une conséquence des résultats de [45]. La preuve repose sur la méthode de *squeezing* de [45], qui consiste à construire des sur-solutions et des sous-solutions de (1.47) qui convergent vers le front progressif solution de (1.38) et qui encadrent la solution $u$ de (1.47). En utilisant le principe de comparaison, on coince cette dernière contre le front progressif. Un argument itératif induit la convergence exponentielle (1.50).

Des travaux récents [1] ont établi une convergence similaire à celle du Théorème 1.2.2, dans le cas où l'opérateur diffusif n'est pas $|\partial_x|$, mais $|\partial_x|^\alpha$ avec $\alpha \in ]1,2[$. Leurs auteurs soulignaient que le cas où $\alpha \in ]0,1]$ était alors un problème encore ouvert.

Grâce au Théorème 1.2.2, la simulation en temps long de (1.47) permet de retrouver le couple $(\phi, c)$ solution de (1.38) pour une très large gamme de données initiales. Nous insistons sur le fait que l'équation (1.47) est *artificielle* dans le sens où elle ne peut pas s'interpréter en termes de dynamique des dislocations. Dans la section suivante, nous utilisons cet outil pour approximer numériquement les solutions de (1.38).

**Résolution numérique de l'équation de Weertman**

Sauf pour certains potentiels $F$ particuliers, l'équation (1.38) ne possède pas de solution analytique connue. Nous proposons au Chapitre 5 une méthode numérique possible pour approximer numériquement sa solution. Nous en détaillons ici les aspects les plus importants.

Le problème non-linéaire (1.38) présente plusieurs spécificités :

— à la fois $\phi$ et $c$ sont des inconnues,

— l'opérateur $|\partial_x|$ est un opérateur intégrodifférentiel de convolution qui est raide (ses valeur propres sont grandes, voir [79] pour la notion de raideur),

— le domaine sur lequel est posé l'équation est la droite réelle $\mathbb{R}$ (et n'est donc pas borné),

— la solution $\phi$ n'est unique qu'à translation près.

Notre approche a été la suivante : tout d'abord poser un problème discrétisé

$$-|D_x|\,\boldsymbol{\phi} + cD_x\boldsymbol{\phi} = F'(\boldsymbol{\phi}), \tag{1.51}$$

où $\boldsymbol{\phi}$ et $F'(\boldsymbol{\phi})$ sont des vecteurs constitués des valeurs $\phi_i \simeq \phi(x_i)$, respectivement $F'(\phi(x_i))$ pour des points $x_i$ équirépartis dans un segment $[-L, L]$, et où $|D_x|$, respectivement $D_x$, sont des discrétisations des opérateurs $|\partial_x|$ et $\partial_x$. Puis, nous bâtissons un système dynamique sur le modèle de (1.47), de telle sorte qu'il converge vers un point fixe $(\boldsymbol{\phi}, c)$ qui satisfasse (1.51).

**Construction d'un problème discrétisé** Il existe plusieurs manières de discrétiser un opérateur tel que $|\partial_x|$. Certaines approches reposent sur la formulation très simple de l'opérateur $|\partial_x|$ en variables de Fourier : par exemple, la technique «*approximate approximation*» de Maz'ya *et al.* (voir [91]), ou l'utilisation de polynômes de Hermite [113]. D'autres auteurs au contraire utilisent des méthodes de quadratures de la formulation intégrodifférentielle (1.39) (par exemple [96]).

Nous avons opté pour la première manière : on scinde $\phi = \phi_{\text{ref}} + \phi_{\text{dyn}}$, où $\phi_{\text{ref}}$ est une fonction analytique de référence compatible avec les conditions à l'infini (1.40), et où $\phi_{\text{dyn}}$ est donc une fonction tendant vers 0 à l'infini. On évalue analytiquement $|\partial_x|\,\phi_{\text{ref}}$, et $|\partial_x|\,\phi_{\text{dyn}}$ est calculée en périodisant $\phi_{\text{dyn}}$, puis en appliquant une discrétisation spectrale de $|\partial_x|$. Ainsi, si $\widehat{\phi}_{\text{dyn}}(k_p)$ sont les composantes de la transformée de Fourier $\widehat{\phi}_{\text{dyn}}$ en les variables de Fourier $k_p$ (issues de la grille duale à celle des $x_j$), on définit :

$$\mathcal{F}\left\{|\partial_x|\,\phi_{\text{dyn}}\right\}(k_p) \simeq |k_p|\widehat{\phi}_{\text{dyn}}(k_p),$$

où $\mathcal{F}$ désigne la transformation de Fourier. L'approximation de $\widehat{\phi}_{\text{dyn}}(k_p)$ repose sur la transformation de Fourier discrète. Outre sa simplicité, une telle discrétisation présente l'avantage de s'effectuer rapidement, car elle s'effectue à l'aide de la Fast Fourier Transform (FFT).

**Méthode de résolution** Une fois l'équation discrète (1.51) fixée, il faut encore choisir une méthode de résolution. Nous proposons une méthode basée sur la convergence du système dynamique (1.47) vers (1.38) et inspirée de [91]. L'enjeu est de simuler (1.47) en atteignant rapidement les temps longs, de telle sorte que le point fixe de l'algorithme de simulation satisfasse (1.51).

L'opérateur $-|\partial_x|$ est raide. Aussi est-il nécessaire d'utiliser un schéma A-stable (voir [79, Chap. IV.3 p. 40]), c'est à dire qui converge vers 0 lorsqu'il approxime l'équation différentielle ordinaire $y' = \lambda y$, pour $\mathrm{Re}\,(\lambda) < 0$. Dans le cas contraire (comme pour Euler explicite), il faut prendre un pas de temps $\Delta t$ petit devant le pas spatial $h$ : atteindre les temps longs devient alors coûteux.

Nous proposons le schéma suivant :

$$\begin{cases} \boldsymbol{\phi}^{n+1} = \boldsymbol{\phi}^n - \Delta t M^n \left( |D_x|\, \boldsymbol{\phi} - c^n D_x \boldsymbol{\phi} + F'(\boldsymbol{\phi}) \right) \\ c^{n+1} = c\left[\boldsymbol{\phi}^{n+1}\right], \end{cases} \tag{1.52}$$

où $M^n$ est un opérateur de préconditionnement. Un point fixe $(\boldsymbol{\phi}, c)$ de (1.52) satisfait naturellement (1.51). Pour certains préconditionneurs $M^n$ bien choisis, le schéma (1.52) est stable inconditionnellement en le pas de discrétisation spatiale. La relation liant $c$ à $\boldsymbol{\phi}$ est construite de telle sorte que le front $\boldsymbol{\phi}$ soit correctement centré dans la boîte de simulation, en imposant que

$$\frac{1}{2L}\int_{-L}^{L} \phi(x)\mathrm{d}x \simeq \frac{1}{2}\left(\phi_\mathrm{l} + \phi_\mathrm{r}\right).$$

Cette relation supplémentaire supprime l'invariance par translation et permet de rétablir l'unicité de la solution (monotone) de (1.38).

Nous avons constaté le bon fonctionnement de l'algorithme en le testant sur des cas de potentiels $F$ où la solution analytique est connue. En comparant solutions analytiques et solutions numériques, on a inféré des taux de convergences empiriques en fonction des paramètres de discrétisation.

Nous avons aussi testé un schéma de splitting de Strang entre les opérateurs linéaires $|\partial_x|$ et $\partial_x$ et l'opérateur non-linéaire $F'(\cdot)$. Lorsque l'on utilise une méthode adéquate pour évaluer $\mathrm{e}^{\Delta t|\partial_x|}$, une tel schéma est stable. Il est même d'ordre supérieur en temps par rapport au schéma (1.52), et d'une rapidité d'exécution comparable. Or, à cause de la non-commutation des opérateurs considérés, le point fixe d'un schéma de splitting dépend du pas de temps $\Delta t$ choisi. Il ne satisfait donc pas (1.51) : il y a une erreur résiduelle qui tend vers 0 avec $\Delta t$. Par conséquent, on aura tendance à prendre $\Delta t$ petit, non pas pour des raisons de stabilité, mais pour des raisons de précision. Cela rend les simulations d'autant plus coûteuses ! Ainsi, dans la reconstruction de l'état asymptotique (1.51), le schéma (1.52) est bien plus efficace qu'un splitting de Strang, toutes propriétés de stabilité et de convergence étant similaires par ailleurs. Dans cette méthode numérique, la dynamique artificielle (1.47) est seulement un moyen, et non une fin.

**Application** Grâce à notre méthode, on peut obtenir des lois de vitesse pour des potentiels physiquement réalistes quelconques. A notre connaissance, cela n'avait jamais été fait pour des dislocations en mouvement satisfaisant l'équation de Weertman.

A titre d'illustration, nous montrons un exemple de résultat d'intérêt physique qu'il est possible de produire avec la méthode numérique décrite ci-dessus, à savoir des lois de vitesse des dislocations en fonction de la contrainte appliquée. On regarde le cas d'une

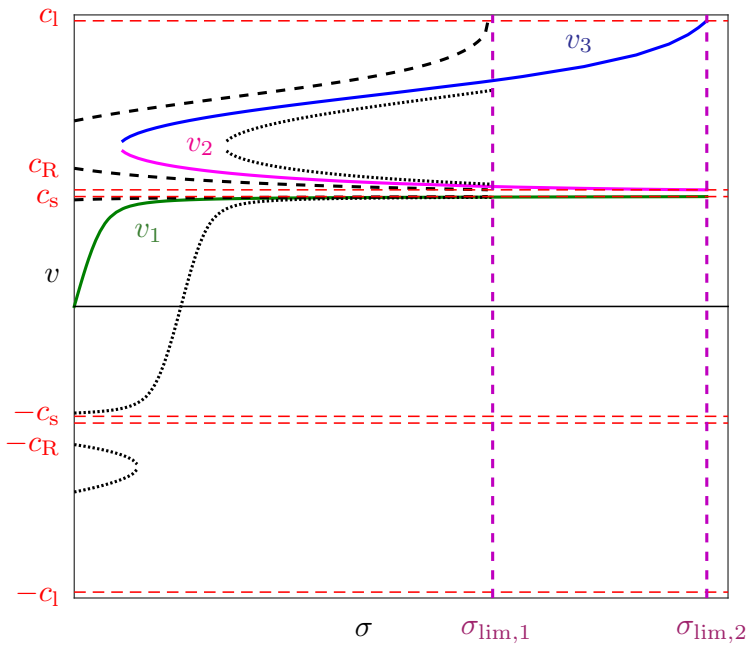



FIGURE 1.7 – Loi de vitesse $v$ en fonction du chargement $\sigma$ pour une dislocation coin soumise à un potentiel "en bosse de chameau" ($\alpha = 0.1$, $\gamma = c_\mathrm{l}/c_\mathrm{s} = \sqrt{6}$). En noir, les lois de vitesse des dislocations partielles (en tirets, la partielle gauche, et en points, la partielle droite). En rouge, les différentes vitesses limites $c_\mathrm{s}$, $c_\mathrm{R}$ et $c_\mathrm{l}$ (voir l'Annexe A.2). Les chargements $\sigma_{\mathrm{lim},1}$ et $\sigma_{\mathrm{lim},2}$ délimitent les régimes où il existe des dislocations partielles, respectivement totales, en mouvement stationnaire. Par convention, les conditions aux limites de $\phi$ satisfont $\phi_\mathrm{l} > \phi_\mathrm{r}$.

dislocation coin soumise à potentiel en "bosse de chameau" (voir Figure 1.6). On obtient alors les courbes de la Figure 1.7, qui décrit les vitesses possibles pour des dislocations parfaites et partielles. Les branches de vitesse $v_1$ et $v_3$ correspondent aux vitesses stables (subsoniques, respectivement transsoniques) et la branche de vitesse $v_2$ est la branche des vitesses transsoniques instables.

La résolution numérique de l'équation de Weertman est un prérequis pour la simulation de l'équation de Peierls-Nabarro Dynamique. En effet, cette dernière requiert une donnée initiale (dans un sens qui sera précisé dans le Chapitre 6). Une dislocation en mouvement à vitesse constante (éventuellement nulle), c'est à dire une solution de l'équation de Weertman, constitue une donnée initiale raisonnable.

### 1.2.3 Dislocations en régime dynamique

#### L'équation de Peierls-Nabarro Dynamique

Si on ne fait aucune hypothèse sur le comportement de la dislocation, on déduit du modèle de Peierls une équation intégrodifférentielle en temps et en espace, appelée équation de Peierls-Nabarro Dynamique (voir [130]). C'est une équation récente dans le domaine des dislocations, dont la phénoménologie et les implications physiques ont été peu explorées. Dans [131] avait été entreprise une résolution approchée de cette équation, en utilisant des Ansatz particuliers. Ce travail avait des limitations sévères : par exemple, on ne pouvait étudier qu'une seule dislocation à la fois. Or, l'équation de Peierls-Nabarro Dynamique est bien plus riche. Le but de cette étude est de construire un outil de simulation numérique permettant d'étudier cette équation dans les régimes d'intérêt que sont la nucléation (apparition de nouvelles dislocations), l'annihilation, la mise en mouvement d'une dislocation, l'étude des chocs...

La construction de l'équation de Peierls-Nabarro Dynamique repose sur un pilier fondamental : l'évaluation d'un opérateur au bord Neumann vers Dirichlet pour une équation d'élasticité (voir Section 1.2.1). Cela se manifeste naturellement par l'apparition d'un opérateur intégrodifférentiel en temps et en espace, que l'on peut exprimer à l'aide d'une fonction de Green (voir [130]). Ce terme constitue une difficulté majeure de l'équation (difficulté théorique *et* numérique). Des termes du même type se retrouvent dans de nombreux champs scientifiques.

En particulier, la communauté géophysique s'est confrontée à un problème similaire depuis les années 1990, avec notamment les travaux initiés par Geubelle, Rice et coauteurs [48, 49, 62, 97, 125, 126]. Dans leur cas, un tel opérateur intervient dans la propagation de fissures dans les roches. Plus généralement, de telles équations apparaissent dans les modèles de zone cohésive (voir par exemple [117]) qui s'intéressent à des interfaces (notamment les dislocations et les fissures). Ce type de modèle permet de coupler la mécanique des milieux continus loin de l'interface, avec des phénomènes de plus petite échelle (ici atomistiques), souvent non-linéaires au niveau de l'interface.

Le même problème apparaît lorsqu'il s'agit de construire des *conditions de bord transparentes* (voir [66] et la revue [75]). On considère un problème posé dans un certain domaine d'intérêt (ici, réduit à une interface) inclus dans un milieu infini. Ces conditions de bord permettent de ramener la résolution d'une équation aux dérivées partielles linéaire dans un milieu environnant à la résolution d'une équation intégrodifférentielle posée sur le bord du domaine considéré. Les conditions de bord transparentes ont beaucoup d'applications en physique des ondes : accoustique, électromagnéstisme (voir [75], où la dérivation de telles conditions est faite dans de nombreuses situations)... Un des enjeux majeur est la simulation numérique de telles équations, qui est rendue ardue précisément à cause du terme intégro-différentiel.

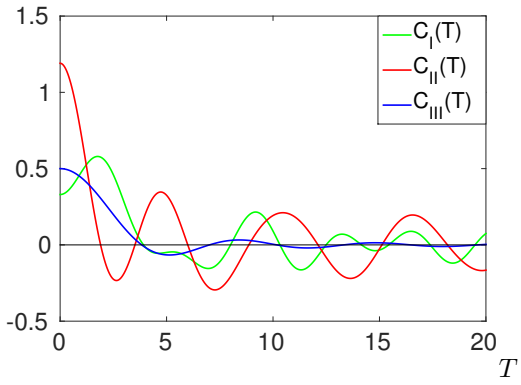



FIGURE 1.8 – Tracé de $C_{\mathrm{I}}(T)$, $C_{\mathrm{II}}(T)$ et $C_{\mathrm{III}}(T)$, pour $\gamma = c_{\mathrm{l}}/c_{\mathrm{s}} = \sqrt{3} \simeq 1.7$ (voir le Chapitre 6).

### Quelques propriétés élémentaires de l'équation de Peierls-Nabarro Dynamique

Avant de discuter plus en détail des enjeux et des difficultés de la simulation de l'équation de Peierls-Nabarro Dynamique, nous formalisons cette dernière, et décrivons brièvement quelques unes de ses propriétés mathématiques.

L'équation de Peierls-Nabarro Dynamique peut s'écrire sous la forme adimensionnée suivante :

$$\begin{cases} \kappa \partial_t \widehat{u}(t,k) = -k^2 \displaystyle\int_0^t C(|k|(t-t'))\widehat{u}(t',k)\mathrm{d}t' + \widehat{G}(t,k), \\ u(t,\cdot) = 0, \end{cases} \tag{1.53a}$$

où $k \in \mathbb{R}$ est la variable de Fourier spatiale, $t > 0$ la variable temporelle et $\widehat{u}$ désigne la transformée de Fourier de $u$ par rapport à la variable $x$. L'inconnue de l'équation (1.53a) est la fonction $u(t,x)$, pour $t > 0$ et $x \in \mathbb{R}$. Dans (1.53a), $\kappa > 0$ est une constante, $C$ est un noyau donné (égal à une des fonctions $C_i$ tracées sur la Figure 1.8), et $G(t,x)$ est une fonction dépendant non-linéairement de $u(t,x)$ et d'un chargement $\sigma^{\mathrm{a}}(t,x)$ imposé de manière dynamique.

L'équation (1.53a) peut se lire comme une famille d'équations intégrodifférentielles linéaires de Volterra du second type à noyau convolutif (voir [105]) indicées par le mode de Fourier $k$. Autrement dit, la transformation de Fourier a diagonalisé la partie linéaire de (1.53a). Un simple changement de variable $\tau = |k|t$ ramène (1.53a) à une seule et même équation, avec un second membre dépendant du mode de Fourier $|k|$.

La structure de convolution en temps et la linéarité de (1.53a) traduisent une invariance en temps de l'équation. Cette propriété apparaît sur la formule de Duhamel que satisfait $\widehat{u}$

$$\widehat{u}(t,k) = \int_0^t \mathfrak{R}(|k|(t-t'))\widehat{G}(t',k)\mathrm{d}t', \tag{1.53b}$$

où la fonction $\mathfrak{R}$, appelée *résolvante*, est solution de l'équation homogène

$$\begin{cases} \kappa \dfrac{\mathrm{d}}{\mathrm{d}\tau}\mathfrak{R}(\tau) = -\displaystyle\int_0^\tau C(\tau-\tau')\mathfrak{R}(\tau')\mathrm{d}\tau', \\ \mathfrak{R}(0) = 1. \end{cases} \tag{1.54}$$

La résolvante $\mathfrak{R}$ est essentielle pour la compréhension de l'équation de Peierls-Nabarro Dynamique. Elle joue un rôle analogue à celui du noyau gaussien vis-à-vis de l'équation de la chaleur : elle traduit un effet régularisant et dissipatif de l'équation (1.53a). En pratique, l'ampleur de cette dissipation est pilotée par un paramètre visco-plastique $\alpha$ qui est petit (l'équation est faiblement dissipative). Sous des hypothèses adéquates, on démontre un résultat d'existence et d'unicité de la solution de l'équation (1.53a) en appliquant un théorème de point-fixe de Banach à la formulation (1.53b).

### Enjeux de la simulation numérique

La simulation numérique efficace et précise de l'équation de Peierls-Nabarro Dynamique est complexe, principalement à cause du terme intégrodifférentiel présent dans l'équation.

La première difficulté provient du fait que l'équation (1.53a) est posée sur un domaine *infini* et fait intervenir *une convolution spatiale*. Cette dernière opération est potentiellement coûteuse et délicate. Forts de notre expérience sur l'équation de Weertman, nous avons naturellement opté pour l'utilisation de la transformation de Fourier discrète (voir la Section 1.2.2). A l'instar de la transformation de Fourier continue, cette approche permet de traiter les différents modes de Fourier de manière quasiment indépendante : la seule communication entre les différents modes de Fourier discrétisés $k_p$ a lieu via le terme non-linéaire $G$. Ainsi, on se ramène à simuler à une famille finie d'équations du type de (1.53), indicées par un nombre fini de modes de Fourier $k_p$.

D'où la seconde difficulté, qui est un réel obstacle numérique : le caractère intégrodifférentiel *en temps* de l'équation de Peierls-Nabarro Dynamique. Cela pose deux problèmes : le stockage et l'utilisation optimaux de la *mémoire* de l'équation, mémoire qui est nécessaire pour faire avancer la dynamique. Nous avons envisagé ces problèmes d'un point de vue purement *algorithmique*, mais ils sont aussi des problèmes informatiques : il est notable que la saturation de la mémoire machine apparaît rapidement en pratique.

Comme $\mathfrak{R}$ et $C$ sont des fonctions régulières (analytiques), les équations (1.53a) et (1.53b) sont bien posées, à second membre $G(t,x)$ régulier donné. La difficulté n'est pas tant leur résolution numérique que leur résolution numérique *efficace*, c'est à dire précise, stable et rapide. Pour cela, il faut surmonter trois obstacles :

1. la valeur de $G(t,x)$ dépend de $u(t,x)$ de façon non-linéaire ;
2. les noyaux de convolution $C(T)$ et $\mathfrak{R}(T)$ tendent lentement vers 0 à l'infini (en loi de puissance en $T$) et oscillent (voir Figure 1.8) ;
3. chaque mode de Fourier de $u(\cdot,k)$ évolue avec un temps dilaté par $|k|$.

Le premier point est davantage une difficulté pratique qu'un verrou conceptuel. Il induit qu'il est nécessaire de calculer conjointement et successivement les valeurs de $u$ et de $G$ jusqu'au temps $t$, afin d'obtenir une approximation de $u(t,\cdot)$. Ainsi, calculer $u(t,\cdot)$ via (1.53b) n'est pas plus rapide qu'utiliser (1.53a) et ne peut se faire en une seule étape –ce qui serait le cas si $G(t'<t,\cdot)$ était connue *a priori*.

Le second point est crucial. Il implique qu'il faut conserver une mémoire précise du passé pour avancer d'un pas de temps. Or, la dépendance en le passé est non-triviale car les intégrales de (1.53a) et (1.53b) encodent des annulations (dues aux oscillations des noyaux) qu'il est délicat de reproduire avec une discrétisation grossière.

Le troisième point traduit une propriété de *raideur* de (1.53a) (voir [79]), qui se manifeste par le préfacteur $|k|^2$ devant l'intégrale englobant la mémoire. Cette raideur induit la nécessité d'utiliser des schémas qui soient *stables*, afin d'éviter que le pas de temps $\Delta t$ ne soit contraint par des conditions de stabilité de type Courant–Friedrichs–Lewy (CFL). En outre, dans les régimes physiquement intéressants, l'équation de Peierls-Nabarro Dynamique est relativement peu dissipative.

### Schémas et méthodes de calcul

Nous insistons sur une trait singulier des équations intégrodifférentielles : le *schéma numérique* de l'équation n'est pas le seul élément dimensionnant de l'*algorithme* utilisé ; la *méthode de calcul* est aussi déterminante. Nous désignons par le mot de *schéma* les équations algébriques satisfaites par la discrétisation $u_n$ d'une fonction $u(t_n)$ à approximer, et par l'expression *méthode de calcul* l'ensemble des opérations algorithmiques grâce auxquelles on résout un schéma numérique donné. Suivant la méthode de calcul, le temps d'exécution est plus ou moins long.

En ce qui concerne les équation différentielles ordinaires, ces deux notions sont généralement indépendantes. Au contraire, pour les équations intégrales de Volterra, la *méthode de calcul* influe de manière non-linéaire sur la complexité temporelle de l'algorithme et sur la quantité de mémoire nécessaire. En tirant parti de certaines structures algébriques, il est possible d'accélérer le calcul d'un schéma fixé ou d'utiliser moins de mémoire. Nous avons étudié les deux structures suivantes :

1. la structure de convolution, qui permet de réorganiser plus efficacement les calculs de quadrature ;
2. la structure de noyau *dégénéré*, grâce à laquelle on peut transformer une équation intégrodifférentielle en équation différentielle ordinaire.

**Des schémas** L'équation (1.53a) appartient à la grande classe des équations intégrodifférentielles

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t) = -\int_0^t K(t,t')u(t')\mathrm{d}t' + f(t), \tag{1.55}$$

où $K$ est un noyau prenant deux arguments et n'ayant pas nécessairement la structure convolutive $K(t,t') = C(t-t')$. Par souci de simplicité, on suppose ici que $u$ est à valeur

scalaire (et ne dépend pas d'une seconde variable $x$). Il existe de nombreux schémas d'intégration de (1.55), parmi lesquel on peut citer les méthodes de Galerkine ou de collocation (voir [10, Chap. 3 p. 49]). Nous nous concentrons sur des schémas itératifs, lesquels reposent en général sur une méthode de quadrature pour calculer l'une ou l'autre intégrale présente dans (1.55). Notre point de départ bibliographique se situe dans la littérature géophysique ; ainsi, nous avons considéré des schémas issus de [62]. Puis, nous avons choisi d'étudier des schémas "bloc-par-bloc" d'ordre 4 (voir [105]). A titre de comparaison, nous proposons aussi une méthode de splitting de Strang.

**Des méthodes de calcul** On peut tout d'abord tirer parti de la structure de convolution grâce à la méthode d'*accélération* de [76]. Cette méthode de calcul permet de calculer rapidement un schéma donné, à condition que celui-ci présente une structure de convolution discrète. On passe ainsi d'une complexité $O(N^2)$ pour une implémentation naïve à une complexité de $O\left(N(\log N)^2\right)$ ($N$ est le nombre de pas de temps). Cette méthode de calcul est remarquable car elle est *exacte*.

Toute équation différentielle peut se mettre sous la forme d'une équation intégrale. L'inverse n'est pas vraie en général. En ce qui concerne (1.55), il faut que le noyau $K$ présente une structure de noyau *dégénéré* pour pouvoir la transformer en une famille d'équations différentielles indicée par $k$, c'est à dire que $K$ se décompose sous la forme suivante :

$$K(t,t') = \sum_{j=0}^{d} a_j(t) b_j(t').$$

(Nous renvoyons à [10, Chap. 2 p. 23] pour cette notion ; le terme de noyau *séparable* est aussi utilisé dans la littérature [74, p. 56]). Or, ce n'est pas le cas pour les noyaux $K(t,t') = C(t-t')$ étudiés. Comme on souhaite préserver la structure convolutive (1.53), on doit donc approximer le noyau originel par un noyau à la fois dégénéré et convolutif. Nous étudions les décompositions suivantes :

1. en somme de polynômes de Laguerre pondérés (voir [44, 104]) ;
2. en somme d'exponentielles. Celles-ci ont été notamment proposées dans les articles [3, 75] de Hagstrom, mais nous utilisons cependant les articles plus récents de la mouvance de Lubich *et al.*, *e.g.*, [15, 16, 111, 112, 137], qui reposent sur l'utilisation d'une transformation de Laplace inverse efficace datant des travaux de Talbot [143].

Ces dernières méthodes, au prix d'une erreur d'approximation sur le noyau, permettent de réduire la mémoire nécessaire, tout en demeurant relativement rapides.

Nous avons implémenté dans un code MATLAB différents algorithmes mêlant des schémas et des méthodes de calcul évoqués ci-dessus. Nous les avons comparés sur plusieurs exemples, et avons étudié la dépendance de l'erreur en les différents paramètres de discrétisation que sont la taille de la boîte de simulation, le pas de discrétisation spatiale, le pas de discrétisation temporelle. Ces tests permettent d'assurer la validité d'expériences numériques en cours, en particulier sur la nucléation de dislocations, sur le croisement de dislocations, ou l'imposition de chocs localisés.

## 1.3 Limite macroscopique d'un système de particules

Dans cette section, nous étudions la limite macroscopique d'un système de particules soumises à la deuxième loi de Newton.

### 1.3.1 Modèle microscopique

Le système discret que nous étudions est constitué d'une chaîne de particules interagissant chacune avec ses plus proches voisins. Chaque particule est soumise à la deuxième loi de Newton. Il n'y a pas de dissipation. Ce modèle très simple peut se voir comme une chaîne de masselottes reliées entre elles par des ressorts (voir Figure 1.9).

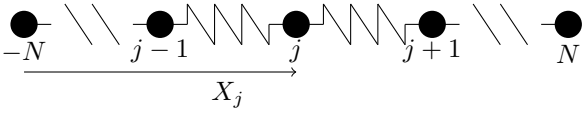



FIGURE 1.9 – Une chaîne de particules

Plus précisément, on se donne $2N+1$ particules indexées par $j \in [\![-N, N]\!]$ dont le mouvement est régi par l'équation suivante :

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} X_j(t) = W'\left(X_{j+1}(t) - X_j(t)\right) - W'\left(X_j(t) - X_{j-1}(t)\right), \tag{1.56}$$

où $X_j(t)$ est la position de la $j^{\text{ème}}$ particule au temps $t$, et $W$ est un potentiel d'interaction (pair, mais pas nécessairement quadratique). Les conditions initiales et les conditions au bord sont imposées de manière macroscopique par

$$X_j(0) = N\Phi_1\left(j/N\right) \quad \text{et} \quad \frac{d}{dt} X_j(0) = \Phi_0\left(j/N\right), \tag{1.57}$$

$$X_{-N}(t) = N\Phi_1(-1) \quad \text{et} \quad X_N(t) = N\Phi_1(1), \tag{1.58}$$

où $\Phi_1$ et $\Phi_0$ sont des fonctions données satisfaisant $\Phi_0(-1) = \Phi_0(1) = 0$.

La question que l'on se pose est la suivante : peut-on construire une limite macroscopique à ce système lorsque le nombre de particules devient infiniment grand ? L'article [26] avait répondu à cette question dans certains cas. Notre étude vise à compléter quelques-un de leurs résultats. Ce type de question a par ailleurs été abordé dans différents cadres par [20,34,57].

### 1.3.2 De l'équation de Newton à l'équation des ondes

L'équation des ondes est une limite macroscopique vraisemblable pour l'équation de Newton (1.56). En effet, fixons les scalings suivants :

$$x_j = \frac{j}{N}, \qquad \tau = \frac{t}{N}, \qquad X_j(t) = N\phi(x_j, \tau).$$

Si on suppose que $\phi$ est régulière, alors, *formellement*, un développement limité en $1/N$ sur (1.56) donne l'équation suivante :

$$\partial_\tau^2 \phi(\tau, x) = \partial_x \left[ W'(\partial_x \phi(\tau, x)) \right], \tag{1.59}$$

c'est à dire l'équation des ondes, avec des conditions initiales et des conditions au bord correspondant à (1.57) et (1.58).

Les auteurs de [26] justifient ces manipulations formelles : sous des hypothèses adéquates, ils démontrent que l'équation (1.59) est bien la limite macroscopique en temps long de (1.56), tant que la fonction $\phi$ demeure suffisamment régulière (voir [26, Prop. 2]).

### 1.3.3 Chocs dans l'équation des ondes

L'équation des ondes (1.59) est une équation hyperbolique qui est potentiellement sujette à des ondes de choc dès lors que le potentiel $W$ n'est pas quadratique. Dans ce cas, la solution $\phi$ de (1.59) n'est plus régulière, mais seulement continue par morceaux (voir [139]). Or, le résultat [26, Prop. 2] se limite aux cas où la fonction $\phi$ est suffisamment régulière (de classe $C^4$ en espace). Par conséquent, il ne recouvre pas toute la phénoménologie du modèle considéré.

Dans le cas où $W$ est un potentiel quadratique, nous démontrons que l'hypothèse de régularité sur $\phi$ n'est pas nécessaire ; pour que les conclusions de [26] demeurent valides, il suffit que les données initiales $\Phi_0$ et $\Phi_1$ soient continues par morceaux, respectivement continues et continûment dérivables par morceaux .

En revanche, si le potentiel $W$ est non-quadratique, on constate qu'en cas de choc, le comportement macroscopique du système de particules est *différent* de celui de l'équation des ondes non-linéaires. En réalité, sous certaines hypothèses techniques, et sous réserve que l'écart inter-particulaire est uniformément borné, nous démontrons qu'en cas d'ondes de choc, la solution entropique de l'équation des ondes (1.59) *n'est plus* la limite macroscopique de (1.56). De manière surprenante, pour une donnée initiale régulière, on observe donc la chose suivante :

- pendant un premier laps de temps, le système discret (1.56) est très semblable au système continu (1.59) (grâce aux résultats de [26]),
- au bout d'un certain temps, lorsque qu'un choc survient dans le système continu (sauf cas particuliers, un choc apparaît au bout d'un laps de temps fini), le comportement macroscopique du système discret s'éloigne peu à peu de la solution entropique de l'équation des ondes.

Notre démonstration repose sur deux ingrédients fondamentaux :

(i) en cas de choc, le système continu est dissipatif irréversible, tandis que le système discret demeure convervatif et réversible ;

(ii) pour que le système discret converge vers le système continu, il est *formellement* nécessaire que l'opération de moyennisation $\langle \cdot \rangle$ spatiale sur des intervalles mésoscopiques et l'opérateur non-linéaire $W'[\cdot]$ commutent comme suit :

$$\lim_{N \to +\infty} \left\langle W'(\partial_x \phi^N(\tau, x)) \right\rangle = \lim_{N \to +\infty} W' \left( \left\langle \partial_x \phi^N(\tau, x) \right\rangle \right),$$

(ici, $\phi^N$ représente une fonction correspondant à $\phi$ reconstruite à partir du système discret de $2N+1$ particules).

Du premier point, on déduit qu'il est impossible que le système discret converge *fortement* vers le système continu : l'éventuelle convergence de $\partial_x \phi^N(\tau, x)$ vers $\partial_x \phi(\tau, x)$ ne peut être qu'une convergence *faible* (ou convergence en moyenne sur des intervalles mésoscopiques) mais pas une convergence forte. Or, comme la fonction $W'$ est non-linéaire (on la suppose même strictement convexe), on déduit du second point que, si le système continu est limite du système discret, alors $\partial_x \phi^N(\tau, x)$ converge fortement vers sa moyenne $\partial_x \phi(\tau, x)$. Ainsi, il est impossible que le système discret converge *–même faiblement–* vers le système continu, lorsque le nombre de particules $2N+1$ devient infini.

Il est notable que, dans le cas où le potentiel $W$ est quadratique, alors le système continu est toujours conservatif. En outre, la fonction $W'$ est linéaire : par conséquent, l'opérateur de moyennisation et l'opérateur $W'[\cdot]$ commutent.

La démonstration utilise une propriété importante du système discret : à savoir une version affaiblie du cône de causalité. En effet, modulo une perturbation exponentiellement petite en le nombre de particules $N$, la position $X_i(t)$ et la vitesse $\frac{\mathrm{d}}{\mathrm{d}t} X_i(t)$ d'une particule $i$ à un temps $t$ donné ne dépendent que des positions et des vitesses des particules $j$ au temps $t'$ présentes dans le cône spatio-temporel $|j - i| \leq c|t - t'|$. Le scalaire $c > 0$ est une vitesse qui dépend du potentiel $W$, et de l'écart maximal entre deux particules voisines. Paradoxalement, la propriété de cône de causalité est caractéristique des systèmes d'équations aux dérivées partielles hyperboliques.

Le résultat ci-dessus présente cependant une limitation importante : il nous est nécessaire de supposer *a priori* que l'écart entre deux particules voisines demeure uniformément borné, indépendamment du nombre de particules. Nous avons observé sur des simulations numériques que c'est effectivement le cas, dès lors que la donnée initiale est suffisamment régulière. Mais nous n'avons pas réussi à le démontrer.

## 1.4 Perspectives

Cette thèse explore plusieurs axes de recherche ; toutefois, de nombreuses questions intéressantes demeurent ouvertes. Nous en citons quelques unes dans cette section.

En ce qui concerne l'homogénéisation, nous avons construit un cadre abstrait pour obtenir des résultats d'approximation fine et appliqué ces résultats aux cas de coefficients périodiques avec défaut.

Cela soulève deux types de questions :

1. peut-on construire explicitement une plus grande classe d'application ?
2. ce cadre d'hypothèses est-il le plus général possible ?

Il semble que le cadre d'hypothèses que nous posons dépasse les cas d'application proposés. Vis-à-vis du programme de recherche de Blanc, Le Bris et Lions, nous sommes donc face à un paradoxe : on sait maintenant que l'hypothèse de périodicité n'est pas nécessaire pour obtenir des résultats d'approximation fine, mais on ne sait appliquer ces résultats que dans

des cas très liés au cadre périodique. Il serait très intéressant du point de vue des applications numériques de pouvoir construire des classes *explicites* plus générales de problèmes. Le terme *explicite* signifie ici qu'il serait possible de construire numériquement des représentants de telles classes, ainsi que les correcteurs associés.

Cela rejoint la seconde question. En effet, Blanc, Le Bris et Lions avaient proposé dans [28] l'étude d'une interface entre deux milieux périodiques. Ils avaient notamment construit des correcteurs bornés dans certains cas. Le cas d'une interface sort du cadre de la théorie classique et n'entre pas non plus dans le cadre théorique proposé dans ce manuscrit (ni dans le cadre de [67]). En effet, le coefficient considéré est généralement discontinu au travers de l'interface et sa matrice homogénéisée est seulement constante par morceaux (mais pas constante globalement).

Le fait que la matrice homogénéisée ne soit pas constante soulève des difficultés techniques que nous décrivons formellement. Le principal obstacle est que la solution $u^\star$ du problème homogénéisé n'est plus infiniment régulière. En effet, son gradient $\nabla u^\star$ est généralement discontinu au travers de l'interface de discontinuité du coefficient $A^\star$, à cause de la condition de transmission de flux. Ainsi, l'identité (1.20) a comme membre de droite un terme qui se ramène (formellement, car le sens mathématique d'une telle expression n'est pas clair) à la divergence d'une mesure portée par l'interface. Or, cette identité joue un rôle central dans le raisonnement ci-dessus lorsqu'il s'agit de contrôler l'erreur de l'homogénéisation $R^\varepsilon$ et son gradient $\nabla R^\varepsilon$. Aussi semble-t-il difficile de contrôler cette dernière à proximité de l'interface. En réalité, il ne semble même pas certain que le gradient $\nabla u^\varepsilon$ soit approximé localement au voisinage de l'interface par la quantité $\nabla u^\star - \sum_{j=1}^d w_j\left(\frac{\cdot}{\varepsilon}\right)\partial_j u^\star$.

Cependant, même si $u^\star$ n'est pas aussi régulier que dans le cas où le coefficient $A^\star$ est constant, il demeure "relativement" régulier, comme le montrent notamment les travaux de Li et Vogelius [103]. Nous pensons que cette régularité moindre est cependant suffisante pour adapter la preuve des estimations lipschitziennes à la Avellaneda et Lin dans certains cas d'interfaces entre des milieux périodiques. L'étude mathématique de ces problèmes est actuellement en cours.

Nous avons proposé des algorithmes pour simuler des dislocations en régime stationnaire et en régime dynamique. L'aboutissement de cette recherche est l'utilisation de tels algorithmes pour faire des expériences numériques afin d'explorer des phénomènes physiques de manière quantitative. En ce qui concerne l'équation de Weertman, on peut désormais étudier la propagation de dislocations en régime stationnaire avec des potentiels issus de $\gamma$-surfaces obtenues expérimentalement (au moins dans le cas scalaire). En ce qui concerne l'équation de Peierls-Nabarro Dynamique, nous avons construit un outil de simulation numérique. Celui-ci permettra de valider ou d'infirmer les résultats du modèle réduit construit dans [131], mais aussi d'explorer de nouveaux phénomènes tels que la nucléation, l'annihilation ou le croisement de dislocations en régime dynamique. De telles expériences sont en cours au CEA-DAM en collaboration avec Yves-Patrick Pellegrini.

Nous n'avons traité dans ce document que des équations scalaires. Or, il est plus réaliste d'étudier des équations vectorielles. A l'heure actuelle, nous n'avons fait ce passage numérique du scalaire vers le vectoriel que dans certains cas de l'équation de Weertman en milieu anisotrope. Des résultats encourageants ont été obtenus en vue d'applications physiques.

Le fait d'étudier des fonctions à valeur vectorielle impose une plus grande technicité numérique, mais ouvre aussi des questions théoriques intéressantes :

1. sous quelles hypothèses existe-t-il une solution à l'équation de Weertman ?
2. cette solution éventuelle est-elle unique ?
3. peut-elle encore s'interpréter comme la limite en temps long d'une équation de réaction-diffusion avec laplacien fractionnaire ?

Certains indices numériques semblent suggérer que l'on peut répondre positivement aux deux dernières questions dans des cas simples. Mais il n'existe pas de principe de comparaison pour les systèmes d'équations (sauf cas très particuliers). Aussi ne semble-t-il pas possible d'aborder les questions 2 et 3 ci-dessus par les outils du Chapitre 4.

Une approche raisonnable serait déjà de comprendre dans quelle mesure de tels résultats ont été obtenus pour les équations classiques de réaction-diffusion (c'est à dire, avec un laplacien au lieu d'un laplacien fractionnaire). Je compte m'investir dans cette recherche dans les années qui viennent.

Le résultat que nous avons démontré dans le Chapitre 9 est négatif : la limite macroscopique du système atomique étudié n'est pas l'équation des ondes non-linéaires, dans le cas d'un choc. Or, des arguments théoriques et numériques suggèrent que cette limite macroscopique existe et est unique (au moins dans un sens faible). Il serait très intéressant d'identifier cette limite.

# Chapitre 2

# Homogénéisation d'un problème périodique avec défaut

Ce chapitre expose les résultats de l'étude d'un problème multi-échelle elliptique de la forme

$$\begin{cases} -\operatorname{div}\left(A\left(\dfrac{x}{\varepsilon}\right)\cdot\nabla u^{\varepsilon}(x)\right) = f(x) & \text{dans} \quad \Omega, \\ u^{\varepsilon}(x) = 0 & \text{sur} \quad \partial\Omega, \end{cases}$$

où $A$ est une matrice elliptique non-périodique. On construit un cadre général dans lequel on peut estimer et approximer précisément les quantités $u^{\varepsilon}$ et $\nabla u^{\varepsilon}$. Ce cadre s'applique au cas d'un matériau présentant une structure périodique perturbée par un défaut.

Cette étude a été réalisée en collaboration avec Xavier Blanc et Claude Le Bris. Elle fera l'objet de publications [23, 24] dont la rédaction est en cours.

Le lecteur trouvera en Annexe A.3 des matériaux théoriques additionnels, et en Annexe B une reproduction du document de travail [23].

## 2.1 Introduction

Cette étude a pour objet l'homogénéisation de certains opérateurs elliptiques multi-échelles $-\mathrm{div}\left(A\left(\frac{\cdot}{\varepsilon}\right)\cdot\nabla\right)$, où la matrice $A$ est déterministe mais pas nécessairement périodique. Nous nous basons pour cela sur les travaux d'Avellaneda et Lin [11] et leurs développements, notamment l'article [94]. Notre but est de construire un cadre théorique dans lequel on puisse démontrer des estimations sur la solution du problème oscillant, et approximer celle-ci dans des espaces de Sobolev $\mathrm{W}^{1,p}$, pour $p \in [2, +\infty]$. Nous proposons ainsi un ensemble d'hypothèses pour des opérateurs elliptiques rescalés afin de remplacer l'hypothèse classique de périodicité. Dans ce cadre, nous démontrons des estimations point par point sur la fonction de Green de Dirichlet de tels opérateurs, ainsi que sur ses dérivées premières et croisées. Aussi, nous quantifions la qualité de l'approximation sur le gradient des solutions de $-\mathrm{div}\left(A\left(\frac{\cdot}{\varepsilon}\right)\nabla u^\varepsilon\right) = f$ par le gradient corrigé de la solution du problème homogénéisé.

Un des cas possibles d'application est le cas d'un matériau périodique multi-échelle perturbé par un défaut (voir [28]). Le coefficient $A$ prend alors la forme $A = A_{\mathrm{per}} + \widetilde{A}$, où $A_{\mathrm{per}}$ est périodique et $\widetilde{A} \in \mathrm{L}^r\left(\mathbb{R}^d\right)$. L'intégrabilité du défaut $\widetilde{A}$ apparaît alors comme déterminante. On peut aussi appliquer nos résultats une certaine classe de coefficients quasi-périodiques.

L'approche est la suivante : nous formalisons d'abord un cadre général abstrait dans lequel sont démontrés les résultats souhaités, puis nous appliquons ces résultats au cas particulier de coefficients périodiques perturbés par des défauts. Nous avons constaté a posteriori que ce cadre est proche de celui proposé récemment par Gloria, Neukamm et Otto dans [67] et raffiné par Bella, Giunti et Otto dans [17]. Par ailleurs, la démarche employée fait aussi écho à d'autre travaux d'Armstrong, Smart, Kuusi et Mourrat (voir [5, 9]). Nous donnerons quelques éléments de comparaison dans la Section 2.2.4.

Nous avons voulu faire une rédaction aussi auto-consistante que possible. Pour cette raison, une grande partie des démonstrations et des résultats de ce chapitre sont des adaptations plus ou moins techniques des preuves de [11, 94].

### 2.1.1 Théorie de l'homogénéisation périodique

Nous rappelons tout d'abord des résultats classiques déjà évoqués en introduction de la thèse.

On se place sur $\mathbb{R}^d$, pour $d \geq 3$, jusqu'à la fin du chapitre. Soit $\Omega \subset \mathbb{R}^d$ un ouvert borné régulier et $A$ un champ de matrices elliptiques donné. On s'intéresse au problème

$$\begin{cases} -\mathrm{div}\left(A\left(\dfrac{x}{\varepsilon}\right)\cdot\nabla u^\varepsilon(x)\right) = f(x) & \text{dans} \quad \Omega, \\ u^\varepsilon(x) = 0 & \text{sur} \quad \partial\Omega, \end{cases} \tag{2.1}$$

où $f \in \mathrm{L}^2(\Omega)$ est un second membre donné, éventuellement régulier. Ici, $\varepsilon$ est un petit paramètre. La problématique est la suivante : on souhaite approximer précisément $u^\varepsilon$ et $\nabla u^\varepsilon$.

Dans le cas où la matrice $A$ est périodique, il existe un cadre théorique bien établi, dont nous décrivons ici les traits principaux. En effet, si $A$ est périodique, elliptique et bornée, il

est bien connu (voir, *e.g.*, [2, Chap. 1, p. 1-15]) que, dans la limite où $\varepsilon \to 0$, le problème (2.1) s'homogénéise en le problème suivant :

$$\begin{cases} -\operatorname{div}\left(A^\star \cdot \nabla u^\star(x)\right) = f(x) & \text{dans} \quad \Omega, \\ u^\star = 0 & \text{sur} \quad \partial\Omega, \end{cases} \tag{2.2}$$

où $A^\star$ est une matrice constante. C'est à dire que $u^\varepsilon \rightharpoonup u^\star$ dans $\mathrm{H}^1_0(\Omega)$. Mais cette convergence n'est pas forte, sauf dans des cas triviaux. Pour approximer le gradient $\nabla u^\varepsilon$, on doit introduire les *correcteurs* $w_j \in \mathrm{H}^1_{\mathrm{loc}}(\mathbb{R}^d)$, pour $j \in [\![1, d]\!]$, relatifs à la matrice $A$. Ils sont définis comme étant les solutions, uniques à l'ajout d'une constante près, de

$$\begin{cases} -\operatorname{div}\left(A(x) \cdot (e_j + \nabla w_j(x))\right) = 0 & \text{dans} \quad \mathbb{R}^d, \\ \dfrac{|w_j(x)|}{1+|x|} \underset{|x|\to+\infty}{\to} 0, \end{cases} \tag{2.3}$$

où les vecteurs $e_j$ sont les vecteurs de la base canonique de $\mathbb{R}^d$. Grâce à ces correcteurs, on construit une approximation de $u^\varepsilon$ dans $\mathrm{H}^1(\Omega)$. En effet, en posant

$$u^{\varepsilon,1}(x) := u^\star(x) + \varepsilon \sum_{j=1}^{d} w_j\left(\frac{x}{\varepsilon}\right) \partial_j u^\star(x), \tag{2.4}$$

et le reste

$$R^\varepsilon(x) := u^\varepsilon(x) - u^{\varepsilon,1}(x) = u^\varepsilon(x) - u^\star(x) - \varepsilon \sum_{j=1}^{d} w_j\left(\frac{x}{\varepsilon}\right) \partial_j u^\star(x), \tag{2.5}$$

on peut alors démontrer par des arguments classiques que $R^\varepsilon \to 0$ dans $\mathrm{H}^1(\Omega)$. On peut même majorer la vitesse de convergence (voir [85, (1.51) p. 28]) :

$$\|\nabla R^\varepsilon\|_{\mathrm{L}^2(\Omega)} \leq C\varepsilon^{1/2}. \tag{2.6}$$

Quand $A$ est périodique, les correcteurs sont eux-mêmes périodiques. Cela facilite la résolution numérique de (2.3), qui peut être reformulé comme un problème avec des conditions au bord périodiques.

Le contrôle sur l'approximation de $\nabla u^\varepsilon$ ci-dessus peut être encore raffiné, et les articles [11, 94] permettent d'approximer $u^\varepsilon$ dans $\mathrm{W}^{1,\infty}(\Omega)$, avec une estimation quantitative en $\varepsilon$ de l'erreur. Pour arriver à un tel résultat, deux grandes étapes théoriques sont nécessaires : une étape d'*estimation* (voir [11]) et une étape d'*approximation* (voir [94]).

En effet, dès lors que $A$ est elliptique, périodique et hölderienne, Avellaneda et Lin ont démontré dans [11] que l'on peut obtenir des estimations fines sur $u^\varepsilon$ solution de

$$-\operatorname{div}\left(A(x/\varepsilon) \cdot \nabla u^\varepsilon(x)\right) = \operatorname{div}(H(x)). \tag{2.7}$$

Ces estimations *uniformes en* $\varepsilon$ sont d'abord hölderiennes, puis lipschitziennes (pour peu que $H$ soit suffisamment régulier). Pour ce faire, ils développent une méthode originale de

compacité faisant usage de la propriété de H-convergence de $A(\cdot/\varepsilon)$. Grâce à ces estimations lipschitziennes, ils démontrent des estimations point par point sur la fonction de Green $G^\varepsilon$ de (2.1), ses gradients $\nabla_x G^\varepsilon$, $\nabla_y G^\varepsilon$ et son gradient croisé $\nabla_x \nabla_y G^\varepsilon$.

Dans un second temps, les auteurs de [94] ont utilisé les résultats d'estimation précédents pour prouver que l'on pouvait contrôler $R^\varepsilon$ dans des normes fines, avec un taux optimal en $\varepsilon$ (à des facteurs en $\log(\varepsilon)$ près). Plus précisément, ils démontrent que, modulo le fait de prendre des correcteurs adaptés au domaine $\Omega$ étudié (c'est à dire avec une définition légèrement différente de (2.3), voir la Section 2.2.5) et si $f$ est suffisamment régulière, on obtient l'estimation suivante : $\|\nabla R^\varepsilon\|_{\mathrm{L}^\infty(\Omega)} \leq C\varepsilon \ln \varepsilon$. Ils utilisent ensuite cette estimation pour approximer les gradients et le gradient croisé de la fonction de Green $G^\varepsilon$ associée au problème (2.1) grâce à la fonction de Green $G^\star$ associée au problème (2.2), convenablement modifiée par les correcteurs $w_j$ (voir [94, Th. 3.6 et Th. 3.11]). Outre les résultats de [11], les ingrédients essentiels des démonstrations de [94] sont le fait que les correcteurs $w_j$ sont bornés, de même que le potentiel (à savoir $B$ défini plus bas).

### 2.1.2 Un cadre abstrait pour l'estimation et l'approximation des solutions d'un problème oscillant

En réalité, le caractère périodique de la matrice $A$ n'est pas nécessaire pour montrer des estimations lipschitziennes uniformes en $\varepsilon$ sur $u^\varepsilon$ solution de (2.7). Par exemple, il est souligné dans [11] que de telles estimations sont aussi vraies si $A$ est quasi-périodique. De même, les résultats d'approximation de [94] peuvent encore s'appliquer dans un cadre plus général que celui des coefficients périodiques. Notre objectif est de formaliser un cadre qui permettent ces généralisations. Nous avons scindé la construction en deux parties : d'une part, les ingrédients nécessaires à la démonstration d'*estimations*, puis les ingrédients pour quantifier la qualité de l'*approximation* de $u^\varepsilon$ par $u^{\varepsilon,1}$. Dans les sections ci-dessous, nous introduisons peu à peu les différentes hypothèses et expliquons formellement les grandes étapes menant à la démonstration d'un théorème d'approximation quantifiée en $\varepsilon$ de $\nabla u^\varepsilon$ dans $\mathrm{L}^p$, pour $p \in [2, \infty]$.

#### Hypothèses abstraites pour l'estimation

Nous proposons un cadre théorique qui permet d'obtenir des estimations régularisantes dans les espaces de Sobolev $\mathrm{W}^{1,p}$, pour $p \in [2, +\infty]$ sur des solutions de (2.1). Ce cadre doit être suffisamment souple pour pouvoir traiter une grande variété de problèmes faisant intervenir un coefficient rescalé, tout en permettant d'adapter les preuves de [11] et [94]. Pour ce faire, on troque l'hypothèse de périodicité contre des hypothèses traduisant une H-convergence uniforme vers une matrice constante (voir [2, Def. 1.2.15 p. 25]). Les hypothèses de H-convergence uniforme permettent de traduire le fait que les solutions de (2.1) tendent vers une solution de (2.2) lorsque $\varepsilon$ tend vers 0 ; cette convergence doit être uniforme en le domaine spatial $\Omega$ considéré. Ces hypothèses peuvent être facilement vérifiées en pratique pour une matrice $A$ donnée dès lors qu'il existe des correcteurs $w_j$ associés.

Nous nous plaçons tout d'abord dans le cadre classique des équations elliptiques :

*Hypothèse* 1 (Ellipticité). Il existe $\mu > 0$ tel que, pour tous $x, \xi \in \mathbb{R}^d$, la matrice $A(x) \in \mathbb{R}^{d\times d}$ est inversible et

$$\xi \cdot A(x) \cdot \xi \geq \mu \left|\xi\right|^2 \text{ et } \xi \cdot A^{-1}(x) \cdot \xi \geq \mu \left|\xi\right|^2.$$

Puis, nous faisons une hypothèse technique de régularité afin de pouvoir utiliser la théorie classique de Schauder :

*Hypothèse* 2 (Régularité). Il existe $\alpha \in ]0,1[$ tel que $A \in \mathrm{C}^{0,\alpha}_{\mathrm{unif}}\left(\mathbb{R}^d, \mathbb{R}^{d\times d}\right)$.

Ensuite, nous développons une série de quatre hypothèses permettant d'établir l'uniforme H-convergence de $A(\cdot/\varepsilon)$ vers une matrice constante $A^\star$.

*Hypothèse* 3 (Existence d'un correcteur). Pour tout $j \in [\![1,d]\!]$, il existe $w_j \in \mathrm{H}^1_{\mathrm{loc}}\left(\mathbb{R}^d\right)$ un correcteur associé à $A$, c'est à dire satisfaisant (2.3).

Si la matrice $A$ vérifie l'Hypothèse 3, on lui associe ses correcteurs $w_j$ (de même, on note $w_j^T$, les correcteurs associés à $A^T$). Elle peut alors satisfaire les Hypothèses 4, 5 et 6 suivantes :

*Hypothèse* 4 (Caractère $\mathrm{L}^2_{\mathrm{unif}}$ du gradient du correcteur). Pour tout $j \in [\![1,d]\!]$,

$$\nabla w_j \in \mathrm{L}^2_{\mathrm{unif}}\left(\mathbb{R}^d, \mathbb{R}^d\right). \tag{2.8}$$

L'espace $\mathrm{L}^2_{\mathrm{unif}}(\mathbb{R}^d, \mathbb{R}^d)$ ci-dessus est l'espace des fonctions $v$ de $\mathbb{R}^d$ à valeur dans $\mathbb{R}^d$ satisfaisant

$$\sup_{y\in\mathbb{R}^d} \int_{B(y,1)} |v(x)|^2 \mathrm{d}x < +\infty.$$

*Hypothèse* 5 (Moyenne macroscopique uniforme). Pour toutes suites $y_n \in \mathbb{R}^d$, $\varepsilon_n \to 0$, et pour tout $j \in [\![1,d]\!]$,

$$\int_Q \nabla w_j\left(\frac{x}{\varepsilon_n} + y_n\right) \mathrm{d}x \underset{n\to+\infty}{\to} 0. \tag{2.9}$$

*Hypothèse* 6 (Convergence macroscopique uniforme vers la matrice homogénéisée). Il existe une matrice constante $A^\star$ telle que, pour toutes suites $\varepsilon_n \to 0$ et $y_n \in \mathbb{R}^d$, pour tous $i, j \in [\![1,d]\!]$,

$$\int_Q e_i \cdot A\left(\frac{x}{\varepsilon_n} + y_n\right) \cdot \left(e_j + \nabla w_j\left(\frac{x}{\varepsilon_n} + y_n\right)\right) \mathrm{d}x \underset{n\to+\infty}{\to} A^\star_{ij}. \tag{2.10}$$

On appelle $A^\star$ la matrice homogénéisée de $A$.

Dans les Hypothèses 4, 5 et 6, l'ingrédient qui remplace la périodicité de la théorie classique est le caractère uniforme sur $\mathbb{R}^d$ de l'estimation (2.8) et des convergences macroscopiques (2.9) et (2.10).

Il est notable que les Hypothèses 4 et 5 sont équivalentes à la sous-linéarité stricte uniforme des correcteurs (voir Section 2.2.1) :

$$\sup_{y\in\mathbb{R}^d} \frac{|w_j(x+y)-w_j(y)|}{1+|x|} \underset{|x|\to+\infty}{\to} 0. \tag{2.11}$$

En outre, si les correcteurs $w_j$ sont uniformément strictement sous-linéaires, l'Hypothèse 6 est impliquée par la sous-linéarité stricte uniforme du potentiel $B$ introduit plus bas (voir Section 2.2.1) :

$$\sup_{y\in\mathbb{R}^d} \frac{|B(x+y)-B(y)|}{1+|x|} \underset{|x|\to+\infty}{\to} 0. \tag{2.12}$$

*Remarque* 1. Les Hypothèses 4, 5 et 6 sont apparentées aux hypothèses de sous-linéarité de [67], qui reviennent à supposer une forme non-local de sous-linéarité correspondant à (2.11) et (2.34). Nous discuterons cela plus en détail dans la Section 2.2.4.

**Estimations à la Avellaneda-Lin sur la solution du problème oscillant**

Les Hypothèses 1, 2, 3, 4, 5 et 6 fournissent un cadre dans lequel on peut démontrer des estimations lipschitziennes. Pour ce faire, on adapte les preuves de [11].

Nous établissons tout d'abord l'*uniforme* H-convergence (voir [2, Def. 1.2.15 p. 25]) de $A\left(\frac{\cdot}{\varepsilon}\right)$ vers $A^\star$ :

**Proposition 2.1.1.** *Soit $A$ satisfaisant les Hypothèses 1, 3, 4, 5 et 6. Soient ensuite $y_n \in \mathbb{R}^d$ et $\varepsilon_n \to 0$. Alors, la suite $A_n(x) := A\left(\frac{x}{\varepsilon_n}+y_n\right)$ H-converge vers $A^\star$ sur tout ouvert borné régulier de $\mathbb{R}^d$.*

Rappelons que la fonction $u^\star$ est très régulière, car elle solution d'un problème elliptique à coefficients constants. Or, l'uniforme H-convergence implique que $u^\varepsilon$ est proche de $u^\star$ lorsque $\varepsilon$ est petit. D'où l'idée fondamentale des résultats d'Avellaneda et Lin [11] : *la solution du problème multi-échelle hérite de la régularité de la solution du problème homogénéisé.*

Ainsi, on démontre à la manière de [11] un résultat de régularité hölderienne, le Théorème 2.1.2 ci-dessous, où on note

$$\Omega(x,R) = \Omega \cap \mathrm{B}(x,R), \qquad \Gamma_\Omega(x,R) = \partial\Omega \cap \overline{\mathrm{B}(x,R)}. \tag{2.13}$$

**Théorème 2.1.2** (Analogue du Théorème 1 de [11])**.** *Soit $y \in \mathbb{R}^d$ fixé. Supposons que $A$ satisfait les Hypothèses 1, 2, 3, 4, 5 et 6. Soit $\Omega$ un ouvert borné régulier de classe $\mathrm{C}^{1,\gamma}$, pour $\gamma > 0$. Soient $\beta \in ]0,1[$ et $g \in \mathrm{C}^{0,\beta}\left(\overline{\mathrm{B}(0,1)}\right)$, pour $\beta > 0$. Supposons que $u^\varepsilon$ est solution de*

$$\begin{cases} -\operatorname{div}\left(A\left(\frac{x}{\varepsilon}+y\right)\cdot\nabla u^\varepsilon(x)\right) = 0 & \text{dans} \quad \Omega(0,1), \\ u^\varepsilon = g & \text{sur} \quad \Gamma_\Omega(0,1). \end{cases} \tag{2.14}$$

*Alors, il existe une constante $C$ ne dépendant que de $d$, $A$, $\beta$ et $\Omega$ telle que*

$$\|u^\varepsilon\|_{\mathrm{C}^{0,\beta}(\Omega(0,1/2))} \leq C\,\|g\|_{\mathrm{C}^{0,\beta}(\Gamma_\Omega(0,1))} + C\,\|u^\varepsilon\|_{\mathrm{L}^2(\Omega(0,1))}. \tag{2.15}$$

On a vu que le couple d'Hypothèses 4 et 5 est équivalent au fait que les correcteurs sont uniformément strictement sous-linéaires (voir (2.11) et le Lemme 2.2.1 plus bas). Ce fait, avec la Proposition 2.1.1, constitue un ingrédient fondamental pour la preuve des estimations lipschitziennes, dont la démonstration est l'adaptation directe de la méthode de compacité de [11]. Le théorème suivant constitue la pierre angulaire dans la démonstration d'estimations lipschitziennes :

**Théorème 2.1.3** (Analogue du Lemme 16 de [11])**.** *Supposons que $A$ satisfait les Hypothèses 1, 2, 3, 4, 5 et 6. Soit $y \in \mathbb{R}^d$, $R > 0$ et $\mathrm{B} := \mathrm{B}(0,R)$. Supposons que $u^\varepsilon \in \mathrm{H}^1(2\mathrm{B})$ est une solution de*

$$-\mathrm{div}\left(A\left(\frac{x}{\varepsilon}+y\right)\cdot\nabla u^\varepsilon(x)\right)=0 \quad \textit{dans} \quad 2\mathrm{B}.$$

*Alors, il existe une constante $C$ ne dépendant que de $d$ et $A$ telle que*

$$\sup_{x\in\mathrm{B}}|\nabla u^\varepsilon(x)| \leq \frac{C}{R}\left(\fint_{2\mathrm{B}}|u^\varepsilon|^2\right)^{1/2}. \tag{2.16}$$

Ces estimations lipschitziennes peuvent s'étendre jusqu'au bord ; toutefois, pour des raisons techniques, cela requiert une hypothèse de sous-linéarité renforcée sur les correcteurs $w_j$. On se donne donc $\nu \in ]0,1]$, et on suppose que :

*Hypothèse* 7 (Sous-linéarité renforcée des correcteurs). Il existe $C > 0$ telle que pour tous $x, y \in \mathbb{R}^d$, $|x-y| > 1$, et pour tout $j \in [\![1,d]\!]$, on a

$$|w_j(x) - w_j(y)| \leq C\,|x-y|^{1-\nu}. \tag{2.17}$$

Nous insistons sur le fait que dans cette section, l'Hypothèse 7 est une hypothèse technique : dans cette section, seule l'existence d'un exposant $\nu$, *quel qu'il soit*, est nécessaire. En anticipant sur la suite, soulignons que, au contraire, la valeur du paramètre $\nu$ sera d'une importance capitale pour quantifier la qualité de l'approximation sur $u^\varepsilon$.

**Théorème 2.1.4** (Analogue du Théorème 2 de [11])**.** *Soit $y \in \mathbb{R}^d$ fixé. Supposons que $A$ satisfait les Hypothèses 1, 2, 3 4, 5, 6 et 7, pour un certain $\nu \in ]0,1]$. Soit $\Omega$ un ouvert borné de $\mathbb{R}^d$ dont le bord est de régularité $\mathrm{C}^{1,\beta}$, et $g \in \mathrm{C}^{1,\beta}\left(\overline{\mathrm{B}(0,2)}\right)$, pour $\beta > 0$. Supposons que $u^\varepsilon \in \mathrm{H}^1(\Omega(0,2))$ est une solution de*

$$\begin{cases} -\mathrm{div}\left(A\left(\dfrac{x}{\varepsilon}+y\right)\cdot\nabla u^\varepsilon(x)\right)=0 & \textit{dans} \quad \Omega(0,2), \\ u^\varepsilon = g & \textit{sur} \quad \Gamma_\Omega(0,2). \end{cases} \tag{2.18}$$

*Alors, il existe une constante $C$ ne dépendant que de $d$, $A$, $\beta$, $\nu$, et du degré de régularité $\mathrm{C}^{1,\beta}$ du bord $\Gamma_\Omega(0,2)$ telle que*

$$\sup_{x\in\Omega(0,1)}|\nabla u^\varepsilon(x)| \leq C\left(\int_{\Omega(0,2)}|u^\varepsilon|^2\right)^{1/2} + C\,\|g\|_{\mathrm{C}^{1,\beta}(\Gamma_\Omega(0,2))}. \tag{2.19}$$

Il est bien connu qu'une matrice $A(x)$ satisfaisant l'Hypothèse 1 possède une fonction de Green $G^\varepsilon$ de Dirichlet associée à l'opérateur $-\mathrm{div}\left(A\left(\frac{\cdot}{\varepsilon}\right)\cdot\nabla\right)$ sur un domaine $\Omega$ régulier borné (voir [72, Th. 1.1]). Cette fonction de Green satisfait l'estimation suivante :

$$|G^\varepsilon(x,y)| \leq C|x-y|^{2-d} \qquad \forall x \neq y \in \Omega. \tag{2.20}$$

En revanche, si on ne dispose pas d'hypothèses supplémentaires sur $A$, la fonction $G^\varepsilon$ est potentiellement très oscillante, et ce, d'autant plus que $\varepsilon$ est petit. Toutefois, grâce au Théorème 2.1.3, on démontre à partir de (2.20) des estimations point par point sur les gradients et gradients croisés de $G^\varepsilon$ (voir [11, 94]) :

**Théorème 2.1.5.** *Supposons que $A$ satisfait les Hypothèses 1, 2, 3, 4, 5 et 6. Soit $\Omega$ un ouvert borné régulier de classe $\mathrm{C}^{1,\beta}$, pour $\beta > 0$, et $G^\varepsilon$ la fonction de Green de Dirichlet associée à l'opérateur $-\mathrm{div}\left(A\left(\frac{\cdot}{\varepsilon}\right)\cdot\nabla\right)$ sur $\Omega$.*

*(i) Alors, pour tout $\Omega_1 \subset\subset \Omega$, $G^\varepsilon$ satisfait l'estimation suivante :*

$$|\nabla_x G^\varepsilon(x,y)| \leq C|x-y|^{1-d} \qquad \forall x \in \Omega_1, y \in \Omega, x \neq y, \tag{2.21}$$

*où $C$ est une constante ne dépendant que de $d$, $A$, $\Omega$ et $\Omega_1$.*

*(i') S'il existe $\nu > 0$ tel que $A$ satisfait aussi l'Hypothèse 7, alors l'estimation (2.21) est vraie pour tous $x, y \in \Omega$.*

*(ii) Si $A^T$ satisfait les Hypothèses 3, 4, 5 et 6, alors*

$$|\nabla_y G^\varepsilon(x,y)| \leq C|x-y|^{1-d} \qquad \forall x \in \Omega, y \in \Omega_1, x \neq y, \tag{2.22}$$

$$|\nabla_x \nabla_y G^\varepsilon(x,y)| \leq C|x-y|^{-d} \qquad \forall x \neq y \in \Omega_1, \tag{2.23}$$

*où $C$ est une constante ne dépendant que de $d$, $A$, $\Omega$ et $\Omega_1$.*

*(ii') Si en outre $A$ et $A^T$ satisfont l'Hypothèse 7 pour un certain $\nu \in ]0,1]$, alors les estimations (2.22) et (2.23) sont vraies pour tous $x, y \in \Omega$.*

Grâce au Théorème 2.1.3 et à un lemme de mesure à la Calderon-Zygmund [140, Th. 2.4], on peut alors démontrer la Proposition suivante, qui permet d'estimer $\nabla u^\varepsilon$ dans le cas d'un problème non homogène :

**Proposition 2.1.6.** *Supposons que $A$ satisfait les Hypothèses 1, 2, 3, 4, 5 et 6. Soient $p \in ]2, +\infty[$, $y \in \mathbb{R}^d$, $R > 0$, $\mathrm{B} := \mathrm{B}(0,R)$ et $H \in \mathrm{L}^p\left(2\mathrm{B}, \mathbb{R}^d\right)$. Supposons que $u^\varepsilon \in \mathrm{H}^1(2\mathrm{B})$ est une solution de*

$$-\mathrm{div}\left(A\left(\frac{x}{\varepsilon}+y\right)\cdot\nabla u^\varepsilon(x)\right) = \mathrm{div}(H(x)) \quad \textit{dans} \quad 2\mathrm{B}. \tag{2.24}$$

*Alors, il existe une constante $C > 0$ ne dépendant que de $d$, $A$ et $p$ telle que*

$$\left(\fint_{\mathrm{B}} |\nabla u^\varepsilon|^p\right)^{1/p} \leq C\left\{\left(\fint_{2\mathrm{B}} |H|^p\right)^{1/p} + \left(\fint_{2\mathrm{B}} |\nabla u^\varepsilon|^2\right)^{1/2}\right\}. \tag{2.25}$$

Notons que la Proposition 2.1.6 ne couvre pas le cas $p = +\infty$. Mais, muni de la fonction de Green $G^\varepsilon$, et plus précisément grâce à l'estimation (2.23), on peut aussi démontrer des estimations lipschitziennes sur $u^\varepsilon$, dans le cas d'une équation inhomogène :

**Proposition 2.1.7.** *Supposons que $A$ satisfait les Hypothèses 1, 2, 3, 4, 5 et 6. Soient $\beta > 0$, $y \in \mathbb{R}^d$, $R > \varepsilon > 0$, et $\mathrm{B} := \mathrm{B}(0, R)$. Soit $H \in \mathrm{C}^{0,\beta}\left(2\mathrm{B}, \mathbb{R}^d\right)$. Alors, il existe une constante $C$ ne dépendant que de $d$, $A$ et $\beta$ telle que, si $u^\varepsilon$ satisfait (2.24), alors :*

$$\begin{aligned}\left\|\nabla u^\varepsilon\right\|_{\mathrm{L}^\infty(\mathrm{B})} \leq & C\left(\fint_{2\mathrm{B}} \left|\nabla u^\varepsilon\right|^2\right)^{1/2} + C\varepsilon^\beta \left\|H\right\|_{\dot{\mathrm{C}}^{0,\beta}(2\mathrm{B})} \\ & + C\ln\left(1 + R\varepsilon^{-1}\right)\left\|H\right\|_{\mathrm{L}^\infty(2\mathrm{B})}. \end{aligned} \tag{2.26}$$

Dans (2.26), on désigne par $\|\cdot\|_{\dot{\mathrm{C}}^{0,\beta}(\Omega)}$ la norme de Hölder homogène

$$\|f\|_{\dot{\mathrm{C}}^{0,\beta}(\Omega)} := \sup_{x \neq y \in \Omega} \frac{|f(x) - f(y)|}{|x - y|^\beta}.$$

La preuve de la Proposition 2.1.7 est basée sur la preuve de [94, Lem. 3.5].

### Hypothèses abstraites pour l'approximation

La raison pour laquelle $u^{\varepsilon,1}$ défini par (2.4) approxime efficacement $u^\varepsilon$ défini par (2.1) repose sur un argument algébrique. En effet, supposons que $A(\cdot/\varepsilon)$ H-converge vers une matrice $A^\star$ *constante* (si $A^\star$ n'est pas constante, il faut modifier la définition des correcteurs pour que la conclusion du calcul ci-dessous demeure valide, voir [28]) et définissons

$$M_k^i(x) := A_{ik}^\star - \sum_{j=1}^d A_{ij}(x)\left(\delta_{jk} + \partial_j w_k(x)\right). \tag{2.27}$$

Par définition des correcteurs $w_j$ et par l'Hypothèse 4, $M_k^i$ défini par (2.27) est un champ de vecteurs de classe $\mathrm{L}^2_{\mathrm{unif}}\left(\mathbb{R}^d\right)$ à divergence nulle, indicé par $k \in [\![1, d]\!]$, c'est à dire satisfaisant

$$\mathrm{div}\left(M_k\right) = 0, \qquad \forall k \in [\![1, d]\!].$$

En dimension $d = 3$, $M$ prend la forme d'un rotationnel (voir (2.29) ci-dessous). Plus généralement, à partir de $M_k^i$, on construit un potentiel $B$ dont la définition est la suivante :

**Définition 2.1.1** (Potentiel)**.** *Un potentiel $B$ associé à $A$ satisfait les identités suivantes :*

$$B_k^{ij} = -B_k^{ji} \qquad \forall i, j, k \in [\![1, d]\!], \tag{2.28}$$

$$M_k^j = \sum_{i=1}^d \partial_i B_k^{ij} \qquad \forall j, k \in [\![1, d]\!], \tag{2.29}$$

$$\Delta B_k^{ij} = \partial_i M_k^j - \partial_j M_k^i \qquad \forall i, j, k \in [\![1, d]\!], \tag{2.30}$$

*où $M$ est défini par* (2.27).

Le potentiel $B$ joue un rôle important dans l'identité suivante :

**Proposition 2.1.8** (p. 26 de [85]). *Soit $A$ satisfaisant les Hypothèses 1, 3 et 4. Soit $\Omega$ un domaine borné régulier. Soient $u^\varepsilon \in \mathrm{H}^1(\Omega)$, $u^\star \in \mathrm{H}^2(\Omega)$ satisfaisant*

$$-\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla u^\varepsilon(x)\right) = -\mathrm{div}\left(A^\star\cdot\nabla u^\star(x)\right) \qquad \text{dans} \quad \Omega. \tag{2.31}$$

*Alors, pour $R^\varepsilon$ défini par* (2.5)*, on a*

$$-\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla R^\varepsilon(x)\right) = \mathrm{div}\left(H^\varepsilon(x)\right), \tag{2.32}$$

*où*

$$H_i^\varepsilon(x) := \sum_{j,k=1}^{d}\left(\varepsilon A_{ij}\left(\frac{x}{\varepsilon}\right)w_k\left(\frac{x}{\varepsilon}\right) - \varepsilon B_k^{ij}\left(\frac{x}{\varepsilon}\right)\right)\partial_{jk}u^\star(x). \tag{2.33}$$

A l'instar de [94], l'identité (2.32) joue un rôle central dans notre approche. En effet, une fois que l'on a démontré des estimations sur le problème oscillant, on déduit de (2.32) que $\|\nabla R^\varepsilon\| \leq \|H^\varepsilon\|$ dans des normes appropriées. La définition (2.33) de $H^\varepsilon$ suggère qu'il faut exercer a priori un contrôle sur les quantités $\varepsilon w_j(\cdot/\varepsilon)$ et $\varepsilon B(\cdot/\varepsilon)$ pour obtenir un taux de convergence sur $\|\nabla R^\varepsilon\|$ quantifié en $\varepsilon$.

En conséquence de quoi, nous requerrons des estimations supplémentaires sur les correcteurs $w_j$ et sur le potentiel $B$ associés à $A$. On fait alors des hypothèses de « sous-linéarité renforcée » par rapport à (2.11) et (2.12). On se donne ainsi $\nu \in ]0,1]$ et on suppose que l'Hypothèse 7 et l'hypothèse suivante suivante sont satisfaites :

*Hypothèse* 8 (Sous-linéarité renforcée du potentiel). Il existe $B$ un potentiel satisfaisant (2.28), (2.29), (2.30) et il existe $C$ ne dépendant que de $d$ et $A$ tel que, pour tous $x, y \in \mathbb{R}^d$, $|x-y|>1$, on a

$$|B(x)-B(y)| \leq C|x-y|^{1-\nu}. \tag{2.34}$$

Les Hypothèses 7 et 8 sont couplées : c'est à dire qu'elle partagent le même exposant $\nu$. Elles contraignent le comportement à longue portée des correcteurs et du potentiel. Cette contrainte est d'autant plus grande que $\nu$ est proche de 1 : lorsque $\nu = 1$, elles induisent alors que les correcteurs et le potentiel sont bornés. On verra par la suite que l'exposant $\nu$ mesure la qualité de l'approximation sur la solution du problème oscillant $u^\varepsilon$.

L'Hypothèse 7 est équivalente à l'inégalité suivante :

$$\sup_{y\in\mathbb{R}^d}\sup_{|x|\in[\varepsilon,1]}\varepsilon\left|w_j\left(\frac{x}{\varepsilon}+y\right)-w_j(y)\right| \leq C\varepsilon^\nu, \tag{2.35}$$

vraie pour tout $\varepsilon \in ]0,1[$. On peut obtenir une estimation similaire sur le potentiel. D'où le contrôle souhaité sur $H^\varepsilon$, à savoir : $\|H^\varepsilon\|_{\mathrm{L}^p(\Omega)} \leq C\varepsilon^\nu \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^p(\Omega)}$, où $\Omega$ est un ouvert borné.

*Remarque* 2. Les Hypothèses 7 et 8 impliquent les Hypothèses 4, 5 et 6 (voir Section 2.2.1).

*Remarque* 3. Les Hypothèses 7 et 8 sont apparentées aux hypothèses de [17]. Nous discuterons cela plus en détail dans la Section 2.2.4.

**Approximation de la solution du problème oscillant**

Nous expliquons maintenant comment quantifier la qualité de l'approximation $u^{\varepsilon,1}$ de la solution du problème oscillant $u^\varepsilon$. Cela revient à estimer $R^\varepsilon$ et $\nabla R^\varepsilon$, pour $R^\varepsilon$, $u^\varepsilon$ et $u^\star$ respectivement définies par (2.5), (2.1), et (2.2). Les arguments développés ci-dessous sont essentiellement adaptés de [94]. Ils reposent fondamentalement sur les Hypothèses 7 et 8 énoncées ci-dessus, qui sont plus fortes que les Hypothèses 4, 5 et 6 (voir Section 2.2.1).

On déduit de (2.33) et des Hypothèses 7 et 8 que, pour tout $p \in [1, +\infty[$,

$$\|H^\varepsilon\|_{\mathrm{L}^p} \leq C\varepsilon^\nu \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^p} \leq C\varepsilon^\nu \|f\|_{\mathrm{L}^p}.$$

Par ricochet, via (2.32) et la Proposition 2.1.6, cela permet d'établir $\|\nabla R^\varepsilon\|_{\mathrm{L}^p} \leq C\|H^\varepsilon\|_{\mathrm{L}^p} \leq C\varepsilon^\nu \|f\|_{\mathrm{L}^p}$. Par un argument de dualité, on peut alors estimer l'écart entre les fonctions de Green $G^\varepsilon$ et $G^\star$ ($G^\star$ étant la fonction de Green de (2.2)) :

**Théorème 2.1.9** (Analogue du Théorème 3.3 de [94])**.** *Soit $A$ satisfaisant les Hypothèses 1, 2, 3, 7 et 8. Soit $\Omega$ un domaine borné régulier de classe $\mathrm{C}^{1,1}$. Soient $G^\varepsilon$ et $G^\star$ les fonctions de Green de Dirichlet sur $\Omega$ de $-\mathrm{div}\left(A\left(\frac{\cdot}{\varepsilon}\right)\cdot\nabla\right)$ et $-\mathrm{div}\left(A^\star\cdot\nabla\right)$, respectivement.*

*(i) Soit $p < d/(d-1)$. Alors il existe une constante $C$ ne dépendant que de $d$, $\Omega$, $A$, $\nu$ et $p$ telle que, pour tous $\varepsilon > 0$, $x \in \Omega$,*

$$\left(\int_\Omega |G^\varepsilon(x,y) - G^\star(x,y)|^p\right)^{1/p} \mathrm{d}y \leq C\varepsilon^\nu. \tag{2.36}$$

*(ii) Si on suppose en outre que $A^T$ satisfait les Hypothèses 3, 7 et 8, alors il existe une constante $C$ ne dépendant que de $d$, $\Omega$, $A$ et $\nu$ telle que, pour tous $\varepsilon > 0$, $x$, $y \in \Omega$,*

$$|G^\varepsilon(x,y) - G^\star(x,y)| \leq C\varepsilon^\nu |x-y|^{2-d-\nu}. \tag{2.37}$$

Le Théorème 2.1.9 a un corollaire immédiat (le point étant qu'il n'y a pas de restriction sur l'intégrabilité $p \in [1,\infty]$ de $f \in \mathrm{L}^p\left(\mathbb{R}^d\right)$ pour (ii)) :

**Corollaire 2.1.10.** *Soit $A$ satisfaisant les Hypothèses 1, 2, 3, 7 et 8. Soit $\Omega$ un domaine borné régulier de classe $\mathrm{C}^{1,1}$.*

*(i) Si $p > d$, il existe une constante $C$ ne dépendant que de $d$, $A$, $\nu$, $\Omega$, et $p$, telle que pour tout $f \in \mathrm{L}^p(\Omega)$, si $u^\varepsilon$ et $u^\star$ sont respectivement solutions de (2.1) et (2.2)*

$$\|u^\varepsilon - u^\star\|_{\mathrm{L}^\infty(\Omega)} \leq C\varepsilon^\nu \|f\|_{\mathrm{L}^p(\Omega)}. \tag{2.38}$$

*(ii) Si $A^T$ satisfait aussi les Hypothèses 3, 7 et 8, alors il existe une constante $C$ ne dépendant que de $d$, $A$, $\nu$, $\Omega$, $p$ et $q \in [1,+\infty]$, telle que pour tout $f \in \mathrm{L}^q(\Omega)$, si $u^\varepsilon$ et $u^\star$ sont respectivement solutions de (2.1) et (2.2)*

$$\|u^\varepsilon - u^\star\|_{\mathrm{L}^p(\Omega)} \leq C\varepsilon^\nu \|f\|_{\mathrm{L}^q(\Omega)}, \tag{2.39}$$

*dès lors que*

$$p < +\infty \quad et \quad \frac{1}{q} \leq \frac{2-\nu}{d} + \frac{1}{p}, \tag{2.40}$$

*ou que*

$$p = +\infty \quad et \quad \frac{1}{q} < \frac{2-\nu}{d}. \tag{2.41}$$

Le Corollaire 2.1.10 permet de généraliser le résultat [94, Th. 3.4], et permet notamment de majorer $\|R^\varepsilon\|_{\mathrm{L}^2(\Omega)}$. Puis, en utilisant successivement l'inégalité de Cacciopoli et la théorie hilbertienne, la Proposition 2.1.6 et la Proposition 2.1.7, on estime $\nabla R^\varepsilon$ dans des normes de plus en plus fines à l'intérieur de $\Omega$. Ainsi, on démontre le Théorème 2.1.11, dont le Théorème 1.1.1 est un cas particulier :

**Théorème 2.1.11** (Analogue du Théorème 3.7 de [94])**.** *Supposons que $A$ satisfait les Hypothèses 1, 2, 3, 7 et 8. Soient $\Omega$ un ouvert borné régulier de classe $\mathrm{C}^{1,1}$, de $\mathbb{R}^d$, $\Omega_1 \subset\subset \Omega$, $f \in \mathrm{L}^2(\Omega)$, et $\varepsilon \in ]0,1[$. Soient $u^\varepsilon$, $u^\star$, et $R^\varepsilon$ respectivement définies par (2.1), (2.2), et (2.5).*

*(i) Si $f \in \mathrm{L}^p(\Omega)$, pour $p > d$, alors $R^\varepsilon \in \mathrm{W}^{1,p}(\Omega)$ et*

$$\|\nabla R^\varepsilon\|_{\mathrm{L}^p(\Omega_1)} \leq C\varepsilon^\nu \|f\|_{\mathrm{L}^p(\Omega)}, \tag{2.42}$$

*où $C$ ne dépend que de $d$, $A$, $\nu$, $p$, $\Omega_1$ et $\Omega$.*

*(i') Dans le cas où $A^T$ satisfait aussi les Hypothèses 3, 7 et 8, si $f \in \mathrm{L}^p(\Omega)$, pour $p \geq 2$, alors $R^\varepsilon \in \mathrm{W}^{1,p}(\Omega)$ et (2.42) est satisfaite, pour une constante $C$ ne dépendant que de $d$, $A$, $\nu$, $p$, $\Omega_1$ et $\Omega$.*

*(ii) Si $f \in \mathrm{C}^{0,\beta}\left(\bar{\Omega}\right)$, pour $\beta \in ]0,1[$ alors on a $R^\varepsilon \in \mathrm{W}^{1,\infty}_{\mathrm{loc}}(\Omega)$ et*

$$\|\nabla R^\varepsilon\|_{\mathrm{L}^\infty(\Omega_1)} \leq C\varepsilon^\nu \ln\left(2+\varepsilon^{-1}\right) \|f\|_{\mathrm{C}^{0,\beta}(\Omega)}, \tag{2.43}$$

*où $C$ ne dépend que de $d$, $A$, $\nu$, $\beta$, $\Omega_1$ et $\Omega$.*

Comme conséquence du Théorème 2.1.11, on obtient une approximation des gradients et du gradient croisé de $G^\varepsilon$. En appliquant la démonstration de [94], on démontre à partir du Théorème 2.1.9 et de la Proposition 2.1.7 le :

**Théorème 2.1.12** (Analogue des Théorèmes 3.6 et 3.11 de [94])**.** *Soit $\Omega$ un domaine borné régulier de classe $\mathrm{C}^{1,1}$ et $\Omega_1 \subset\subset \Omega$. Supposons que $A$ et $A^T$ satisfont les Hypothèses 1, 2, 3, 7 et 8. Soient $G^\varepsilon$ et $G^\star$ les fonctions de Green de Dirichlet sur $\Omega$ de $-\mathrm{div}\left(A\left(\frac{\cdot}{\varepsilon}\right)\cdot\nabla\right)$, respectivement $-\mathrm{div}\left(A^\star\cdot\nabla\right)$.*

*Alors il existe une constante $C > 0$ ne dépendant que de $d$, $A$, $\nu$, $\Omega$ et $\Omega_1$ et telle que,*

*pour tout* $\varepsilon \in ]0,1[$*, et tout* $i \in [\![1,d]\!]$*, on a*

$$\left| \partial_{x_i} G^\varepsilon(x,y) - \sum_{j=1}^{d} \left( \delta_{ij} + \partial_i w_j \left( \frac{x}{\varepsilon} \right) \right) \partial_{x_j} G^\star(x,y) \right|$$
$$\leq C\varepsilon^\nu \frac{\ln\left(2+\varepsilon^{-1}\right)}{|x-y|^{d-1+\nu}} \qquad \forall x \in \Omega_1, \forall y \in \Omega, x \neq y, \tag{2.44}$$
$$\left| \partial_{y_i} G^\varepsilon(x,y) - \sum_{j=1}^{d} \left\{ \delta_{ij} + \partial_i w_j^T \left( \frac{y}{\varepsilon} \right) \right\} \partial_{y_j} G^\star(x,y) \right|$$
$$\leq C\varepsilon^\nu \frac{\ln\left(2+\varepsilon^{-1}\right)}{|x-y|^{d-1+\nu}} \qquad \forall x \in \Omega, \forall y \in \Omega_1, x \neq y, \tag{2.45}$$

*et pour tous* $i,j \in [\![1,d]\!]$*,*

$$\left| \partial_{x_i}\partial_{y_j} G^\varepsilon(x,y) - \sum_{k,l=1}^{d} \left( \delta_{ik} + \partial_i w_k \left( \frac{x}{\varepsilon} \right) \right) \partial_{x_k}\partial_{y_l} G^\star(x,y) \left( \delta_{lj} + \partial_j w_l^T \left( \frac{y}{\varepsilon} \right) \right) \right|$$
$$\leq C\varepsilon^\nu \frac{\ln\left(2+\varepsilon^{-1}\right)}{|x-y|^{d+\nu}} \qquad \forall x,y \in \Omega_1, x \neq y. \tag{2.46}$$

### 2.1.3 Applications

Les hypothèses proposées sont satisfaites pour des coefficients périodiques, des coefficients périodiques avec un défaut, et des coefficients quasi-périodiques (sous des hypothèses adéquates détaillées plus bas).

#### Cas d'un coefficient périodique avec défaut

On se place dans le cadre établi par [28]. La Proposition suivante permet d'insérer le cas $A = A_{\rm per} + \widetilde{A}$ dans le cadre théorique développé plus haut :

**Proposition 2.1.13.** *Soient* $\alpha \in ]0,1[$*,* $\mu > 0$ *et* $r \geq 1$*, avec* $r \neq d$*. Supposons que* $A_{\rm per}$ *et* $\widetilde{A}$ *satisfont*

$$\widetilde{A} \in \mathrm{L}^r\left(\mathbb{R}^d, \mathbb{R}^{d\times d}\right), \tag{2.47}$$
$$\mu^{-1} \leq \widetilde{A} + A_{\rm per} \leq \mu \qquad et \qquad \mu^{-1} \leq A_{\rm per} \leq \mu, \tag{2.48}$$
$$A_{\rm per}, \widetilde{A} \in \mathrm{C}^{0,\alpha}_{\rm unif}\left(\mathbb{R}^d, \mathbb{R}^{d\times d}\right). \tag{2.49}$$

*Supposons que* $A_{\rm per}$ *est périodique, et posons* $A = A_{\rm per} + \widetilde{A}$*. Alors* $A$ *satisfait les Hypothèses 1, 2, 3, 4, 5, 6, 7, et 8, pour* $\nu = \nu_r$ *défini par*

$$\nu_r := \min\left(1, \frac{d}{r}\right) \in ]0,1]. \tag{2.50}$$

*Sa matrice homogénéisée* $A^\star$ *est égale à celle de* $A_{\rm per}$*.*

La Proposition 2.1.13 est la conséquence directe de deux faits : $A_{\rm per}$ vérifie elle-même les Hypothèses 3, 4, 5, 6, 7, et 8, et $\widetilde{A}$ ne fait que perturber légèrement $A_{\rm per}$. La Proposition 2.1.13 permet d'appliquer au cas d'un coefficient $A = A_{\rm per} + \widetilde{A}$ tous les résultats énoncés plus haut. Les Théorèmes 1.1.1, 1.1.3 et 1.1.4 de l'introduction de la thèse sont ainsi les conséquences des Théorèmes 2.1.11, 2.1.5 et 2.1.12.

**Cas d'un coefficient quasi-périodique**

On se donne une matrice $A_{\rm per}$ satisfaisant

$$A_{\rm per} \in \mathrm{C}^\infty\left(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^{d\times d}\right) \text{ est } \mathrm{Q}^2 - \text{périodique} \tag{2.51}$$

où $\mathrm{Q} := [-1/2, 1/2]^d$ est le cube unité. On s'intéresse au coefficient suivant :

$$A(x) = A_{\rm per}(R \odot x, S \odot x),$$

où le symbole $\odot$ désigne la multiplication composante par composante (ou produit de Hadamard).

Nous avons besoin de la notion suivante d'incommensurabilité :

**Définition 2.1.2** (Nombre de Louville-Roth, voir Définition 5.4 de [28])**.** *Un nombre $x \in \mathbb{R}$ est un nombre de Liouville-Roth si, pour tout $n \in \mathbb{N}$, il existe $k_n \in \mathbb{Z}$ et $j_n \in \mathbb{Z}^*$ tels que*

$$0 < \left| x - \frac{k_n}{j_n} \right| \leq \frac{1}{|j_n|^n}.$$

Un nombre de Liouville-Roth est un irrationnel très proche des nombres rationnels. On s'intéresse à des quasi-périodes fortement irrationnelles, c'est à dire satisfaisant l'hypothèse suivante :

*Hypothèse* 9 (Quasi-périodes "fortement" irrationnelles). Les vecteurs $R \in \left(\mathbb{R}_+^*\right)^d$ et $S \in \left(\mathbb{R}_+^*\right)^d$ sont tels que leurs rapports $R_i/S_i$ ne sont ni rationnels ni des nombres de Liouville-Roth, pour tout $i \in [\![1, d]\!]$.

Cette hypothèse est motivée par la nécessité technique d'avoir une estimation régularisante, en l'occurrence, l'inégalité de Gårding (voir Section 2.7.2), qui remplace l'inégalité de Poincaré. Elle permet d'établir la

**Proposition 2.1.14.** *Soient $A_{\rm per}$ satisfaisant* (2.51) *et l'Hypothèse 1, et $R$ et $S$ satisfaisant l'Hypothèse 9. Alors $A(x) = A_{\rm per}(R \odot x, S \odot x)$ satisfait les Hypothèses 1, 2, 3, 4, 5, 6, 7, et 8, pour $\nu = 1$. Sa matrice homogénéisée $A^\star$ est la matrice homogénéisée relative à $A_{\rm per}$.*

Ainsi, tous les résultats de la Section 2.1.2 sont valides pour une telle matrice $A$. Le fait que les estimations d'Avellaneda et Lin s'appliquent au cas de coefficients quasi-périodiques était annoncé dans [11]. Mais la Proposition 2.1.14 montre cependant que le cadre théorique proposé est souple.

### 2.1.4 Plan

Notre plan s'articule en 6 sections. Dans la Section 2.2, nous mettons en lumière certains aspects des résultats proposés. Les autres Sections sont dévolues aux démonstrations des résultats annoncés dans l'introduction ; parfois, elles contiendront des remarques sur la portée de ceux-ci ou la technique de preuve. Dans la Section 2.3, nous rappelons et démontrons quelques faits élémentaires ; ensuite, dans les Sections 2.4 et 2.5, nous démontrons des estimations sur la solution du problème oscillant dans le cas homogène, puis dans le cas inhomogène. Dans la Section 2.6, nous quantifions la qualité de l'approximation de la solution du problème oscillant ; dans la Section 2.7, ces résultats sont appliqués au cas périodique avec défaut, et au cas quasi-périodique. Nous avons relégué en Annexe A.3 un certain nombre de résultats de la littérature et des lemmes techniques. On y trouvera aussi un graphique illustrant les liens logiques entre les différents résultats d'estimation.

**Dans la Section 2.2,** nous énonçons quelques considérations générales sur le cadre théorique proposé dans ce chapitre. En particulier, nous relions les Hypothèses 4, 5 et 6 à des notions de convergence faible, et nous démontrons dans la Section 2.2.1 que les Hypothèses 7 et 8 sont plus fortes que les Hypothèses 4, 5, et 6. Dans la Section 2.2.2, nous discutons du rôle joué par le rescaling dans le cadre proposé. Puis, dans la Section 2.2.3, nous insistons sur la notion d'*uniformité* de la H-convergence, qui est cruciale. Dans la Section 2.2.4, nous comparons le cadre ainsi formulé avec la littérature, et notamment avec le cadre proposé par Otto et ses collaborateurs dans [17,67]. Nous discutons dans la Section 2.2.5 de l'importance du bord dans les différents résultats énoncés, et introduisons les correcteurs adaptés au bord. Nous justifions dans la Section 2.2.6 en quoi le paramètre $\nu$ des Hypothèses 7 et 8 est représentatif de la qualité de l'homogénéisation. Dans la Section 2.2.7, nous commentons un aspect technique du cadre proposé, à savoir le fait qu'il est parfois nécessaire de faire des hypothèses non seulement sur $A$, mais aussi sur $A^T$ pour pouvoir conclure à des propriétés d'estimation ou d'approximation. Enfin, dans la Section 2.2.8, nous proposons des extensions possibles à cette étude.

**Dans la Section 2.3,** nous énonçons un certain nombre de résultats élémentaires issus de la théorie classique des équations elliptiques. Nous indiquons notamment comment construire le potentiel $B$ et démontrons des résultats de régularité sur les correcteurs et le potentiel. Nous démontrons ensuite la Proposition 2.1.8, et expliquons comment tirer parti des Hypothèses 7 et 8 pour majorer $H^\varepsilon$ définie par (2.33).

**Dans la Section 2.4,** nous adaptons la théorie de [11]. Dans le cadre des hypothèses abstraites énoncées plus haut, nous démontrons des estimations hölderiennes et lipschitziennes, à la fois internes et au bord, sur $u^\varepsilon$ solution de

$$-\mathrm{div}\left(A(x/\varepsilon)\cdot\nabla u^\varepsilon(x)\right)=0.$$

La méthode de compacité, originellement développée dans [11], est d'abord expliquée dans la Section 2.4.1. Dans la Section 2.4.3, nous en démontrons le prérequis fondamental, la

Proposition 2.1.1 d'uniforme H-convergence de $A(\cdot/\varepsilon)$. Nous appliquons ensuite la méthode de compacité pour obtenir des estimations hölderiennes dans la Section 2.4.4, à la fois internes et au bord (Sections 2.4.4 et 2.4.4).

Puis, nous démontrons des estimations lipschitziennes internes sur $u^\varepsilon$ dans la Section 2.4.5 (c'est à dire le Théorème 2.1.3). Prouver des estimations lipschitziennes au bord se révèle plus délicat. Grâce aux résultats de la Section 2.4.4, nous montrons tout d'abord des propriétés de sous-linéarité stricte uniforme des correcteurs adaptés dans la Section 2.4.6. Munis de ces correcteurs, nous pouvons finalement démontrer les estimations lipschitziennes voulues (à savoir le Théorème 2.1.4) dans la Section 2.4.7.

**Dans la Section 2.5,** nous nous intéressons au cas où

$$-\mathrm{div}\left(A(x/\varepsilon)\cdot\nabla u^\varepsilon(x)\right) = f(x),$$

où $f$ est un second membre non nul ; en particulier $f = \delta_y$ et $f = \mathrm{div}(H)$.

Pour $f = \delta_y$, on peut étudier le comportement de la fonction de Green $G^\varepsilon$ et démontrer des estimations point par point sur celle-ci, ses gradients et son gradient croisé (c'est à dire le Théorème 2.1.5, prouvé dans la Section 2.5.1). Ces estimations reposent sur l'observation que, hormis sur la singularité $x = y$ on a $f(x) = 0$. Ainsi, on peut utiliser les résultats établis dans le cas homogène dans des boules évitant la singularité (c'est à dire les Théorèmes 2.1.3 et 2.1.4).

Grâce aux résultats de la Section 2.4, si $f = \mathrm{div}(H)$ avec $H \in \mathrm{L}^p$, on peut aussi montrer des estimations intérieures dans $\mathrm{L}^p$ sur $\nabla u^\varepsilon$ (c'est à dire la Proposition 2.1.6, démontrée dans la Section 2.5.2), pour $p \in [2, +\infty[$. Enfin, on se sert des estimations sur $\nabla_x G^\varepsilon$, $\nabla_y G^\varepsilon$, et $\nabla_x\nabla_y G^\varepsilon$, pour prouver dans la Section 2.5.3 des estimations lipschitziennes intérieures sur $\nabla u^\varepsilon$ (c'est à dire la Proposition 2.1.7), dès lors que $H$ est de classe $\mathrm{C}^{0,\beta}$.

**Dans la Section 2.6,** nous utilisons les estimations précédemment démontrées pour approximer $\nabla u^\varepsilon$ et la fonction de Green $G^\varepsilon$, ainsi que ses gradients et gradients croisés. Le principe de l'approximation repose sur un calcul qui permet d'estimer la différence $R^\varepsilon$ définie par (2.5) à partir des correcteurs $w_i$ et du potentiel $B$ (voir la Proposition 2.1.8). Les résultats et démonstrations de cette section sont des adaptations de [94].

Dans la Section 2.6.1, nous démontrons le Théorème 2.1.9 qui permet d'approximer la fonction de Green $G^\varepsilon$. La preuve repose sur le Théorème de De Giorgi-Nash-Moser et un argument de dualité. Muni de cet outil, on en déduit une estimation sur $\|R^\varepsilon\|_{\mathrm{L}^p}$ dans la Section 2.6.2 (c'est à dire le Corollaire 2.1.10). Puis on estime $\|\nabla R^\varepsilon\|_{\mathrm{L}^p}$, pour $p \in [2, \infty]$, dans la Section 2.6.3 ; ceci démontre le Théorème 2.1.11. Enfin, on approxime $\nabla_x G^\varepsilon$, $\nabla_y G^\varepsilon$ et $\nabla_x\nabla_y G^\varepsilon$ dans la Section 2.6.4.

**Dans la Section 2.7,** nous appliquons les résultats précédents au cas d'un champ de matrices périodique perturbé par un défaut $A(x) = A_{\mathrm{per}}(x) + \widetilde{A}(x)$, puis au cas d'un champ de matrices quasi-périodique. Pour ce faire, il suffit de démontrer que les correcteurs $w_j$ et le potentiel $B$ relatifs à de tels champs satisfont les Hypothèses 7 et 8.

## 2.2 Discussion sur le cadre théorique proposé

### 2.2.1 Formulation alternative des Hypothèses

Si $A$ satisfait les Hypothèses 1 et 3, alors les Hypothèses 4 et 5, respectivement 4 et 6, sont équivalentes aux convergences faibles suivantes (2.52), respectivement (2.53) :

$$\nabla w_i\left(\frac{x}{\varepsilon_n}+y_n\right) \underset{n\to+\infty}{\rightharpoonup} 0 \qquad \text{dans} \quad \mathrm{L}^2(\Omega), \tag{2.52}$$

$$e_i\cdot A\left(\frac{x}{\varepsilon_n}+y_n\right)\cdot\left(e_j+\nabla w_j\left(\frac{x}{\varepsilon_n}+y_n\right)\right) \underset{n\to+\infty}{\rightharpoonup} A^\star_{ij} \qquad \text{dans} \quad \mathrm{L}^2(\Omega), \tag{2.53}$$

vraies pour tout ouvert borné $\Omega$ et toutes suites $\varepsilon_n\to 0$ et $y_n\in\mathbb{R}^d$ (dans (2.53), $A^\star\in\mathbb{R}^{d\times d}$ est constant).

En effet, il est immédiat que (2.52) implique l'Hypothèse 5. Si l'Hypothèse 4 n'est pas satisfaite, on peut construire une suite $\varepsilon_n\to 0$, et $y_n\in\mathbb{R}^d$ telle que $\nabla w_i\left(\cdot/\varepsilon_n\right)$ n'est pas bornée dans $\mathrm{L}^2\left(\mathrm{B}(y,1)\right)$, ce qui est en contradiction avec la convergence faible (2.52) par [36, Prop. 3.5(iii) p. 58]. Ainsi, par contraposée, (2.52) implique l'Hypothèse 4. Le fait que les Hypothèses 4 et 5 impliquent (2.52) est une application directe du Lemme A.3.10. De même, on démontre que les Hypothèses 4 et 6, sont équivalentes à (2.53).

Les Hypothèses 4 et 5 sont liées à une propriété de stricte sous-linéarité uniforme du correcteur, grâce au Théorème de De Giorgi-Nash-Moser.

**Lemme 2.2.1.** *Soit $A$ satisfaisant les Hypothèses 1 et 3. Alors $A$ satisfait les Hypothèses 4 et 5 si et seulement si, pour tout $j\in[\![1,d]\!]$,*

$$\sup_{y\in\mathbb{R}^d}\frac{\left|w_j\left(x+y\right)-w_j(y)\right|}{1+|x|} \underset{|x|\to+\infty}{\to} 0. \tag{2.54}$$

*Remarque* 4 (Architecture logique). Cette propriété de sous-linéarité des correcteurs est fondamentale pour la démonstration d'estimations sur le problème oscillant et sera utilisée ultérieurement pour démontrer les Théorèmes 2.1.3 et 2.1.4.

De même, on peut démontrer que si $A$ satisfait les Hypothèses 1 et 3, et si le potentiel $B$ est strictement sous-linéaire, c'est à dire si

$$\sup_{y\in\mathbb{R}^d}\frac{|B(x+y)-B(y)|}{1+|x|} \underset{|x|\to+\infty}{\to} 0, \tag{2.55}$$

alors $A$ satisfait l'Hypothèse 6. En revanche, nous ne savons pas si l'implication réciproque est vraie.

Ainsi, il suffit d'établir l'uniforme stricte sous-linéarité des correcteurs $w_j$ et du potentiel $B$ pour déduire que $A$ satisfait les Hypothèses 4, 5 et 6, d'où le résultat suivant :

**Corollaire 2.2.2.** *Supposons que $A$ satisfait les Hypothèses 1 et 3. Alors :*

*(i) Si $A$ satisfait aussi l'Hypothèse 7, alors $A$ satisfait les Hypothèses 4 et 5.*

*(ii) Si $A$ satisfait l'Hypothèse 8 pour $A^\star$ constant, alors $A$ satisfait l'Hypothèse 6.*

Nous démontrons maintenant le résultat annoncé ci-dessus :

*Démonstration du Lemme 2.2.1.* Supposons tout d'abord que $A$ satisfait les Hypothèses 4 et 5. On se donne $y \in \mathbb{R}^d$, $i \in [\![1, d]\!]$. Posons

$$u_y^\varepsilon(z) = z_i + \varepsilon w_i\left(\frac{z}{\varepsilon} + y\right) - \varepsilon \fint_{\mathrm{B}(0,2)} w_i\left(\frac{z'}{\varepsilon} + y\right) \mathrm{d}z'.$$

Alors

$$\operatorname{div}\left(A\left(\frac{z}{\varepsilon} + y\right) \cdot \nabla u_y^\varepsilon(z)\right) = 0.$$

Par définition, $u_y^\varepsilon$ est à moyenne nulle sur $\mathrm{B}(0,2)$. Ainsi, en utilisant l'inégalité de Poincaré-Wirtinger, puis l'Hypothèse 4, il existe une constante $C$ indépendante de $\varepsilon$ et de $y$ telle que

$$\fint_{\mathrm{B}(0,2)} \left|u_y^\varepsilon(z)\right|^2 \mathrm{d}z \leq C \fint_{\mathrm{B}(0,2)} \left|\nabla u_y^\varepsilon(z)\right|^2 \mathrm{d}z \leq C. \tag{2.56}$$

Grâce au Théorème de De Giorgi-Nash-Moser [64, Th. 8.2 p. 202], on déduit de (2.56) qu'il existe $\beta > 0$ ne dépendant que de $\mu$ telle que

$$\left\|u_y^\varepsilon\right\|_{\mathrm{C}^{0,\beta}(\mathrm{B}(0,1))} \leq C \fint_{\mathrm{B}(0,2)} \left|u_y^\varepsilon(z)\right|^2 \mathrm{d}z \leq C. \tag{2.57}$$

On se fixe maintenant $y_n \in \mathbb{R}^d$ , et $\varepsilon_n \to 0$. Comme les Hypothèses 4 et 5 impliquent (2.52), on en déduit que

$$\nabla u_{y_n}^{\varepsilon_n} = e_i + \nabla w_i\left(\frac{\cdot}{\varepsilon} + y_n\right) \rightharpoonup e_i \quad \text{dans} \quad \mathrm{L}^2(\mathrm{B}(0,1)).$$

Donc, par injection compacte de $\mathrm{H}^1(\mathrm{B}(0,1))$ dans $\mathrm{L}^2(\mathrm{B}(0,1))$,

$$u_{y_n}^{\varepsilon_n}(z) \to z_i \quad \text{dans} \quad \mathrm{L}^2(\mathrm{B}(0,1)).$$

Or, par (2.57), cette convergence a lieu dans $\mathrm{C}^0\left(\mathrm{B}(0,1)\right)$. Par définition de $u_y^\varepsilon$, on en déduit que

$$\sup_{y\in\mathbb{R}^d} \sup_{z\in\mathrm{B}(0,1)} \left|u_y^\varepsilon(z) - z_i\right| = \sup_{y\in\mathbb{R}^d} \sup_{z\in\mathrm{B}(0,1)} \left|\varepsilon w_i\left(\frac{z}{\varepsilon} + y\right) - \varepsilon \fint_{\mathrm{B}(0,2)} w_i\left(\frac{z'}{\varepsilon} + y\right) \mathrm{d}z'\right| \underset{\varepsilon\to 0}{\to} 0.$$

En posant $\varepsilon = (1 + |x|)^{-1}$, on obtient alors (2.54).

Réciproquement, supposons que (2.54) est satisfaite. Par l'inégalité de Cacciopoli,

$$\int_{\mathrm{B}(x,1)} |\nabla w_i(y) + e_i|^2 \,\mathrm{d}y \leq C \int_{\mathrm{B}(x,2)} |w_i(y) + y_i - (w_i(x) + x_i)|^2 \,\mathrm{d}y \leq C.$$

Donc, par inégalité triangulaire, l'Hypothèse 4 est satisfaite. Soient à présent $\varepsilon_n \to 0$ et $y_n \in \mathbb{R}^d$. Par le théorème de la divergence

$$\left|\int_Q \nabla w_i\left(\frac{x}{\varepsilon_n}+y_n\right)\mathrm{d}x\right| = \left|\int_{\partial Q}\varepsilon_n\left(w_i\left(\frac{x}{\varepsilon_n}+y_n\right)-w_i(y_n)\right)\mathrm{d}\vec{S}(x)\right| \\ \leq \varepsilon_n \sup_{x\in Q}\left|w_i\left(\frac{x}{\varepsilon_n}+y_n\right)-w_i(y_n)\right|\int_{\partial Q}\mathrm{d}S(x).$$

D'où, grâce à (2.54), on en déduit (2.9). Ainsi, l'Hypothèse 5 est satisfaite. □

*Démonstration du Corollaire 2.2.2.* La démonstration du Corollaire 2.2.2(i) est une conséquence directe du Lemme 2.2.1.

On démontre ensuite le point (ii) du Corollaire 2.2.2. Par définition de $B$ puis par le théorème de la divergence

$$\left|\int_{\mathrm{Q}}\left\{\sum_{j=1}^d A_{ij}\left(\frac{x}{\varepsilon_n}+y_n\right)\left(\delta_{jk}+\partial_j w_k\left(\frac{x}{\varepsilon_n}+y_n\right)\right)-A^\star_{ik}\right\}\mathrm{d}x\right| \\ = \left|\int_{\mathrm{Q}}\sum_{j=1}^d \partial_j B_k^{ji}\left(\frac{x}{\varepsilon_n}+y_n\right)\mathrm{d}x\right| \\ = \left|\sum_{j=1}^d\int_{\partial\mathrm{Q}}\varepsilon_n\left(B_k^{ji}\left(\frac{x}{\varepsilon_n}+y_n\right)-B_k^{ji}\left(y_n\right)\right)e_j\cdot\mathrm{d}\vec{S}\right|.$$

Alors, en invoquant l'Hypothèse 8, il existe une constante $C>0$ indépendante de $n$ telle que

$$\left|\sum_{j=1}^d\int_{\partial\mathrm{Q}}\varepsilon_n\left(B_k^{ji}\left(\frac{x}{\varepsilon_n}+y_n\right)-B_k^{ji}\left(y_n\right)\right)e_j\cdot\mathrm{d}\vec{S}\right| \leq C\varepsilon_n^\nu \underset{n\to 0}{\to} 0.$$

Ainsi, $A$ satisfait l'Hypothèse 6. □

### 2.2.2 Quelques remarques sur le rescaling

Le fait d'imposer que les matrices $A_\varepsilon$ considérées aient la structure de rescaling suivante $A_\varepsilon(x) = A(x/\varepsilon)$ induit que $A^\star(x)$, définie comme la limite faible de $A_{ij}\left(\frac{x}{\varepsilon}\right)\left(\delta_{jk}+\partial_j w_k\left(\frac{x}{\varepsilon}\right)\right)$, ne dépend que de $x/|x|$ (si $x\neq 0$). Si on suppose en outre que $A^\star(x)$ est continue en 0, alors elle est constante.

Malheureusement, notre compréhension actuelle ne nous permet pas d'englober le cas intéressant d'une interface entre deux milieux périodiques (voir [28]), sauf dans le cas extrêmement particulier où la matrice homogénéisée $A^\star$ est constante (voir Section 2.7). Mais, en général, cette dernière est discontinue.

Notons cependant que l'utilisation d'un rescaling n'est nécessaire ni à la preuve d'estimations, ni à la construction d'une approximation fine pour la solution du problème oscillant.

On pourrait imaginer un problème faisant intervenir des matrices $A_\varepsilon$, avec des correcteurs associés $w_j^\varepsilon$ qui ne seraient pas nécessairement des rescalings les uns des autres. Si les correcteurs $w_j^\varepsilon$ et les potentiels $B^\varepsilon$ satisfont une propriété du type suivant :

$$\sup_{y\in\mathbb{R}^d} \varepsilon \left|w_j^\varepsilon(x+y) - w_j^\varepsilon(y)\right| \leq C\varepsilon^{1-\nu}|x|^\nu,$$
$$\sup_{y\in\mathbb{R}^d} \varepsilon \left|B^\varepsilon(x+y) - B^\varepsilon(y)\right| \leq C\varepsilon^{1-\nu}|x|^\nu,$$

c'est à dire des estimations analogues aux Hypothèses 7 et 8, alors il semble envisageable de démontrer des résultats analogues aux Théorèmes 2.1.2, 2.1.3, 2.1.4, 2.1.5, 2.1.9, 2.1.11 et 2.1.12. Des cas concrets qui pourraient rentrer dans ce cadre seraient par exemple une microstructure périodique sur laquelle viennent se rajouter deux (ou plusieurs) défauts microscopiques séparés d'une distance macroscopique, c'est à dire

$$A_\varepsilon(x) = A_{\mathrm{per}}\left(\frac{x}{\varepsilon}\right) + \widetilde{A}_1\left(\frac{x-x_1}{\varepsilon}\right) + \widetilde{A}_2\left(\frac{x-x_0}{\varepsilon}\right),$$

où $\widetilde{A}_1$ et $\widetilde{A}_2$ satisfont chacune (2.47).

### 2.2.3 Uniformité sur l'espace

Dans les Hypothèses 4, 5 et 6, l'ingrédient qui remplace la périodicité de la théorie classique est le caractère uniforme dans $\mathbb{R}^d$ de l'Estimation (2.8) et des convergences macroscopiques (2.9) et (2.10). On retrouve cette uniformité sur $\mathbb{R}^d$ dans les Hypothèses 7 et 8.

Cette uniformité des estimations et convergences est un élément fondamental. Elle permet de traduire mathématiquement le fait que la H-convergence a lieu en tout point de l'espace, et à toute échelle de façon *uniformément contrôlée*. Elle empêche par exemple le cas où $A$ aurait un comportement à l'infini qui la conduirait à H-converger vers plusieurs matrices différentes suivant la suite de points et d'échelles considérés (voir le contre-exemple ci-dessous). Ce genre de comportement est pathologique en ce qu'il ne permet pas d'appliquer la preuve de [11], car on ne dispose plus alors d'un unique problème homogénéisé de référence. En quelque sorte, cette uniformité évite que 0 ne joue le rôle particulier qui lui échoit naturellement à cause du rescaling.

*Remarque* 5. Les dimensions $d = 1$ et $d = 2$ constituent des cas très particuliers de la théorie elliptique (notamment à cause de la forme de la fonction de Green du laplacien sur $\mathbb{R}$ et sur $\mathbb{R}^2$). Pour cette raison, nous n'avons pas traité ces cas plus techniques. Toutefois, par abus, nous utilisons à l'occasion le cas monodimensionnel pour éclairer certains aspects des résultats présentés dans ce chapitre.

*Contre-exemple.* On se place dans le cadre monodimensionnel et on pose

$$a(x) = \begin{cases} 2 & \text{si} \quad x \in [2^n, 2^n + 2^n/\log(n)] \quad \text{pour} \quad n \in \mathbb{Z}, \\ 1 & \text{sinon.} \end{cases} \tag{2.58}$$

Clairement, pour tous $x_0 < x_1$, on a

$$⨍_{x_0}^{x_1} \frac{1}{a(x/\varepsilon)} \mathrm{d}x \underset{\varepsilon \to 0}{\to} 1.$$

Donc le coefficient homogénéisé relatif à $a$ est $a^\star = 1$. Ainsi, le correcteur associé à $a$ satisfait

$$a(x)(1 + w'(x)) = 1.$$

Donc, si on se donne la suite $y_n = 2^n$, et $\varepsilon_n = \log(n)2^{-n}$, alors, par définition de $a$,

$$\begin{aligned} ⨍_0^1 w'\left(y_n + \frac{x}{\varepsilon_n}\right) \mathrm{d}x =& ⨍_{2^n}^{2^n + 2^n/\log(n)} w'(z) \mathrm{d}z \\ =& ⨍_{2^n}^{2^n + 2^n/\log(n)} \left(\frac{1}{a(z)} - 1\right) \mathrm{d}z \\ =& -\frac{1}{2}. \end{aligned}$$

Donc, $a$ ne vérifie pas l'Hypothèse 5.

Toutefois, la solution de

$$(a(x/\varepsilon)(u^\varepsilon)'(x))' = f(x) \quad \text{et} \quad u^\varepsilon(0) = u^\varepsilon(1) = 0$$

est lipschitzienne si $f$ est bornée. Donc, ce contre-exemple n'éclaire pas sur la possible irrégularité que pourrait avoir $u^\varepsilon$ si $A$ ne satisfaisait par les Hypothèses 4, 5 et 6, à cause du caractère très spécifique de la dimension 1. □

### 2.2.4 Elements de comparaison avec la littérature

Le cadre théorique développé dans ce chapitre est proche de celui proposé récemment par Gloria, Neukamm et Otto dans [67] et raffiné par Bella, Giunti et Otto dans [17]. Par ailleurs, la démarche adoptée fait écho aux travaux d'Armstrong, Smart, Kuusi et Mourrat (voir [5, 9]).

On observe une similitude avec les travaux d'Otto et de ses coauteurs du point de vue des hypothèses formulées et des résultats obtenus. La démarche globale est proche ; en effet, l'article [67] effectue une analogue de l'étape d'estimation et l'article [17] effectue une analogue de l'étape d'approximation. Les résultats et les hypothèses sont apparentés (nous effectuons ci-dessous une comparaison plus précise), mais les démonstrations sont différentes. Nos travaux diffèrent de ceux de Otto et ses coauteurs par :

— la technique de preuve employée, qui reprend la démonstration historique de [11] puis l'article [94],

— nos objectifs, qui visent à établir des résultats d'approximation point par point sur des problèmes posés en domaine borné, afin de les employer dans le cadre des défauts.

Notons que les travaux de Otto et coauteurs visent à démontrer des estimations *en moyenne* à l'échelle locale, car ils souhaitent appliquer leur résultats dans le cadre de l'homogénéisation stochastique. C'est pourquoi ils ne font pas d'hypothèse de régularité sur la matrice $A$ considérée dans [17, 67]. Au contraire, nous souhaitons obtenir des estimations point par point (et non en moyenne locale), ce qui requiert une certaine régularité de la matrice $A$ (ce qui motive notre Hypothèse 2).

Une autre différence importante entre nos travaux et les articles [17,67] est le traitement des bords. En effet, nous avons choisi de nous placer dans le cas de problèmes avec des conditions au bord de Dirichlet, car c'est un problème qui se pose souvent en pratique (mais qui est un peu plus difficile techniquement que le cas sans bord). Au contraire, les articles [17, 67] s'intéressent à des problèmes posés sur tout l'espace $\mathbb{R}^d$.

Pour plus de précision, nous reproduisons ici un résultat central de [67] :

**Lemme 2.2.3** (Conséquence du Lemme 2 de [67])**.** *Soit $\varepsilon_0 > 0$ une longueur caractéristique. Il existe une constante $C(d,\mu)$ avec les propriétés suivantes : Supposons que la matrice $A \in \mathrm{L}^\infty\left(\mathbb{R}^d, \mathbb{R}^{d\times d}\right)$ satisfait l'Hypothèse 1, s'homogénéise en une matrice $A^\star$ constante, et possède des correcteurs $w_j$ et un potentiel $B$ satisfaisant*

$$\sup_{R>\varepsilon_0} \fint_{\mathrm{B}(0,R)} \left\{ \left| w(x) - \fint_{\mathrm{B}(0,R)} w(y)\mathrm{d}y \right|^2 + \left| B(x) - \fint_{\mathrm{B}(0,R)} B(y)\mathrm{d}y \right|^2 \right\} \mathrm{d}x \leq \frac{1}{C(d,\mu)}.$$

*Alors, pour tous rayons $R_1 > R > \varepsilon_0$, pour toute solution $u$ de*

$$-\mathrm{div}\left(A\cdot\nabla u\right) = 0 \quad \textit{dans} \quad \mathrm{B}(0,R_1),$$

*on a l'estimation suivante :*

$$\fint_{\mathrm{B}(0,R)} |\nabla u|^2 \leq C(d,\mu) \fint_{\mathrm{B}(0,R_1)} |\nabla u|^2 . \tag{2.59}$$

Ce Lemme est semblable au Théorème 2.1.3. On remarque que, si une matrice $A$ satisfaisant l'Hypothèse 1 admet des correcteurs $w_j$ et un potentiel $B$ strictement sous-linéaires au sens de (2.54) et (2.55), alors, pour un paramètre $\lambda > 0$ adéquat, la matrice $A(\lambda\cdot)$ satisfait les Hypothèses du Lemme 2.2.3. Par ailleurs, si cette même matrice $A$ est uniformément hölderienne (*i.e.*, elle satisfait l'Hypothèse 2), alors on peut remplacer l'Estimation (2.59) dans le Lemme 2.2.3 par l'estimation suivante :

$$|\nabla u(0)| \leq C \fint_{\mathrm{B}(0,R_1)} |\nabla u|^2 ,$$

pour une constante $C$ dépendant de $d, \mu, \alpha$, et $\|A\|_{\mathrm{C}^{0,\alpha}\left(\mathbb{R}^d\right)}$. Ainsi, on arrive à une conclusion similaire au Théorème 2.1.3 avec des hypothèses apparentées à celles que l'on a posées dans ce document (les Hypothèses 3, 4, 5 et 6, dont on indique dans la Section 2.2.1 le lien avec la sous-linéarité forte des correcteurs et du potentiel).

Les auteurs de [67] démontrent le Lemme 2.2.3 par une preuve directe reposant sur l'équation (2.32), et non la preuve par compacité de [11].

Le lien entre une sous-linéarité "renforcée" et une bonne approximation de la solution du problème multi-échelle (2.1) par le développement à deux échelles (2.4) avait été remarqué dans l'article [17], dont voici un résultat important [1] :

**Théorème 2.2.4** (Corollaire 3 de [17]). *Soit $A \in \mathrm{L}^\infty\left(\mathbb{R}^d, \mathbb{R}^{d\times d}\right)$ un champ de matrices symétriques satisfaisant l'Hypothèse 1, s'homogénéisant vers une matrice $A^\star$ constante, et admettant des correcteurs $w_j$ et un potentiel $B$. Supposons qu'il existe un exposant $\nu \in ]0,1[$ et une longueur caractéristique $\varepsilon_0 > 0$ tels que pour tout $x \in \{x_0, y_0\}$ et tout $R > \varepsilon_0$, on a*

$$\left( \fint_{\mathrm{B}(x,R)} \left\{ \left| w(y) - \fint_{\mathrm{B}(x,R)} w \right|^2 + \left| B(y) - \fint_{\mathrm{B}(x,R)} B \right|^2 \right\} \mathrm{d}y \right)^{1/2} \leq \left( \frac{R}{\varepsilon_0} \right)^{1-\nu}. \tag{2.60}$$

*Alors, on peut approximer la fonction de Green $\mathcal{G}$ sur $\mathbb{R}^d$ de l'opérateur $-\mathrm{div}\,(A\cdot\nabla)$ grâce à la fonction de Green $\mathcal{G}^\star$ sur $\mathbb{R}^d$ de l'opérateur $-\mathrm{div}\,(A^\star\cdot\nabla)$ de la manière suivante :*

$$\begin{aligned} &\fint_{\mathrm{B}\left(x_0,\frac{\varepsilon_0}{2}\right)} \fint_{\mathrm{B}\left(y_0,\frac{\varepsilon_0}{2}\right)} \left| \partial_{x_i}\partial_{y_j}\mathcal{G}(x,y) - \sum_{k,l=1}^{d} \left(\delta_{ik} + \partial_i w_k\left(x\right)\right) \partial_{x_k}\partial_{y_l}\mathcal{G}^\star(x,y) \left(\delta_{lj} + \partial_j w_l\left(y\right)\right) \right|^2 \mathrm{d}y\mathrm{d}x \\ &\leq C(d,\mu,\alpha)\varepsilon_0^\nu \frac{\ln\left(\left|x_0-y_0\right|\varepsilon_0^{-1}\right)}{\left|x_0-y_0\right|^{d+\nu}}. \end{aligned} \tag{2.61}$$

Ce résultat est proche du Théorème 2.1.12. En effet, l'hypothèse (2.60) est très proche des Hypothèses 7 et 8 ; elles sont même équivalentes dans le sens suivant : Si on suppose que (2.60) est satisfaite en tout point (avec une longueur $\varepsilon_0$ indépendante du point considéré), et si on suppose en outre que le champ $A$ satisfait l'Hypothèse 2, alors, par la caractérisation de Campanato, on en déduit que $A$ satisfait les Hypothèses 7 et 8. Réciproquement, les Hypothèses 7 et 8 impliquent que (2.60) est satisfaite. Par ailleurs, la formule (2.61) est une analogue de (2.46) (l'homogénéité de l'estimation est la même), à la nuance près qu'elle porte sur des moyennes locales, et non sur des majorations point par point.

La technique de preuve de [17] fait appel à des notions de moments, tandis que nous utilisons les démonstrations de [94].

D'autre part, les estimations lipschitziennes à la Avellaneda et Lin (ici, le Théorème 2.1.3) ont fait l'objet d'adaptations récentes au cadre stochastique (stationnaire ergodique) par Armstrong et ses coauteurs [5,7,9] (voir les notes de cours [6]). Si la philosophie générale (qui est d'utiliser la régularité du problème homogénéisé pour obtenir des estimations sur le problème multi-échelle) ainsi que les objectifs globaux demeurent les mêmes, ces travaux sont techniquement plus éloignés de notre cadre. Ces auteurs soulignent cependant que des hypothèses similaires aux Hypothèses 4, 5 et 6 sont essentielles pour obtenir des résultats de convergence en homogénéisation (voir [6, (1.37) p. 16]).

1. il y a une coquille dans l'article [17], corrigée dans la formule (2.61)

Leur technique de preuve est d'une nature variationnelle. Elle repose sur l'introduction d'une énergie qui sert à quantifier l'écart entre le problème homogénéisé et le problème multi-échelle (voir [6, (2.9) p. 38]) :

$$J(\Omega,p,q) := \sup_{\substack{-\operatorname{div}(A(\cdot/\varepsilon)\cdot\nabla u)=0 \\ \text{dans} \quad \Omega}} \int_\Omega \left( -\frac{1}{2}\nabla u \cdot A(\cdot/\varepsilon)\cdot\nabla u - p\cdot A(\cdot/\varepsilon)\cdot\nabla u + q\cdot\nabla u \right),$$

pour $(p,q) \in \mathbb{R}^d \times \mathbb{R}^d$. Grâce à un processus de bootstrap (qui est le pendant de l'étape d'itération pour la preuve d'Avellaneda et Lin), ils contrôlent cette énergie, qui est sous-additive (voir [6, Lem. 2.2 p. 39]) et proche d'une énergie homogénéisée [6, Th. 2.4]. Ainsi, ils obtiennent des taux de convergence sur la différence $u^\varepsilon - u^\star$ [6, Th. 2.16 p. 63].

### 2.2.5 Correction au bord : correcteurs adaptés et régularité

Les estimations du Théorème 2.1.3 ou du Théorème 2.1.5 sont valides non seulement à l'intérieur du domaine $\Omega$ considéré, mais aussi jusqu'au bord. Les preuves des estimations au bord sont plus délicates, et nécessitent l'introduction de correcteurs adaptés $w_j^{\varepsilon,\Omega}$, définis comme solutions de

$$\begin{cases} -\operatorname{div}\left(A\left(\dfrac{x}{\varepsilon}\right).\left(\varepsilon\nabla w_j^{\varepsilon,\Omega}(x)+e_j\right)\right)=0 & \text{dans} \quad \Omega, \\ w_j^{\varepsilon,\Omega}=0 & \text{sur} \quad \partial\Omega. \end{cases} \tag{2.62}$$

En revanche, les Estimations (2.42) et (2.43) ne sont pas vraies en général jusqu'au bord, car les correcteurs $\varepsilon w_j(\cdot/\varepsilon)$ sont a priori non nuls sur le bord $\partial\Omega$. Pour obtenir une approximation de $\nabla u^\varepsilon$ sur tout le domaine $\Omega$, il faut donc remplacer les correcteurs $w_j$ dans (2.4) par les correcteurs adaptés définis ci-dessus. Dans le cas périodique, les auteurs de [94] obtiennent alors des estimations du type de (2.42) et (2.43), qui sont vraies sur tout le domaine (c'est à dire pour $\Omega_1 = \Omega$). Bien que nous ayons rassemblé un certain nombre d'éléments, nous n'avons pas adapté leur démonstration.

Soulignons que la construction de correcteurs $w_j$ est numériquement peu chère dans le cadre d'un champ périodique présentant un défaut $a = A_{\text{per}} + \widetilde{A}$ (voir [27]). En revanche la construction de correcteurs adaptés est a priori beaucoup plus chère, car faisant intervenir un problème de couche limite.

En première lecture, on pourra omettre les détails concernant la régularité du bord (sauf pour le Théorème 2.1.4, où elle joue un rôle de premier plan). Nos hypothèses sur la régularité du bord se scindent en deux groupes, selon que l'on énonce un résultat d'estimation ou un résultat d'approximation.

Les résultats d'estimations, à savoir les Théorèmes 2.1.2, 2.1.4 et 2.1.5, nécessitent que le domaine $\Omega$ considéré soit de classe $\mathrm{C}^{1,\beta}$, pour $\beta > 0$ (comme c'est le cas pour les résultats de [11]). C'est dû à l'utilisation d'estimations de Schauder (voir [64, Cor. 8.36 p. 212]).

Les résultats d'approximation, à savoir le Théorème 2.1.9, le Corollaire 2.1.10, et les Théorèmes 2.1.11 et 2.1.12, nécessitent que le domaine $\Omega$ considéré soit de classe $\mathrm{C}^{1,1}$. Cette

régularité est requise pour obtenir des estimations $\mathrm{W}^{2,p}$ proches du bord (voir [64, Th. 9.13 p. 239]). Ces estimations sont notamment utilisées pour estimer $\nabla_x^2 G^\star$ (voir Théorème A.3.7), ce dont nous avons besoin pour démontrer le Lemme 2.6.2. L'article [94] fait les mêmes hypothèses de régularité du domaine (voir par exemple [94, Th. 1.1 & Th. 1.2]).

### 2.2.6 Optimalité de $\nu_r$

Les estimations du Théorème 1.1.1 semblent optimales en $\varepsilon$, à des facteurs en $\log(\varepsilon)$ près. Nous justifions cette assertion d'une part en remarquant que ces résultats coïncident avec ceux de [94] dans le cas périodique (alors $\nu = 1$), et d'autre part en étudiant un exemple monodimensionnel.

*Exemple en dimension 1.* Soit $r \in ]1, +\infty[$. On pose $\widetilde{A} \in \mathrm{L}^r(\mathbb{R}^d)$ positive, paire et bornée par 1. On résout

$$\left(\{1 + \widetilde{A}(\cdot/\varepsilon)\}(u^\varepsilon)'\right)' = 1 \qquad \text{dans} \quad [-1,1], \quad \text{et} \quad u^\varepsilon(-1) = u^\varepsilon(1) = 0,$$

dont le problème homogénéisé correspondant est

$$(u^\star)''(x) = 1 \qquad \text{dans} \quad [-1,1], \quad \text{et} \quad u^\star(-1) = u^\star(1) = 0.$$

Clairement, on a

$$u^\star(x) = \frac{x^2}{2} - 1/2 \quad \text{et} \quad (u^\varepsilon)'(x) = \frac{x + C_\varepsilon}{1 + \widetilde{A}(x/\varepsilon)}.$$

Comme $\widetilde{A}$ est paire et que $u^\varepsilon(1) - u^\varepsilon(-1) = 0$, alors $C_\varepsilon = 0$. De plus, le correcteur $w$ associé à $1 + \widetilde{A}$ est défini par

$$w(x) = -\int_0^x \frac{\widetilde{A}(z)}{1 + \widetilde{A}(z)} \mathrm{d}z.$$

Donc, le reste $R^\varepsilon$ du problème étudié a pour gradient

$$\begin{aligned}
(R^\varepsilon)'(x) =& (u^\varepsilon)'(x) - (u^\star)'(x) - \varepsilon w(x/\varepsilon)(u^\star)''(x) - w'(x/\varepsilon)(u^\star)'(x) \\
=& \frac{x}{1 + \widetilde{A}(x/\varepsilon)} - x + \varepsilon \int_0^{x/\varepsilon} \frac{\widetilde{A}(z)}{1 + \widetilde{A}(z)} \mathrm{d}z + x \frac{\widetilde{A}(x/\varepsilon)}{1 + \widetilde{A}(x/\varepsilon)} \\
=& \varepsilon \int_0^{x/\varepsilon} \frac{\widetilde{A}(z)}{1 + \widetilde{A}(z)} \mathrm{d}z.
\end{aligned}$$

D'où, comme $\widetilde{A} \leq 1$,

$$\left|(R^\varepsilon)'(x)\right| \geq \frac{\varepsilon}{2} \int_0^{x/\varepsilon} \widetilde{A}(z) \mathrm{d}z.$$

Prenons alors par exemple

$$\widetilde{A}(x) = \left[(1+|x|)\,(1+\log(1+|x|))^{1+\delta}\right]^{-1/r}.$$

avec $\delta > 0$. Alors, $\widetilde{A} \in \mathrm{L}^r(\mathbb{R})$. En outre,

$$\begin{aligned}\int_0^{x/\varepsilon} \widetilde{A}(z)\mathrm{d}z \geq& \left(\log(1+\varepsilon^{-1})^{1+\delta}\right)^{-1/r} \int_0^{x/\varepsilon} \frac{1}{(1+|z|)^{1/r}}\mathrm{d}z \\ \geq& C\varepsilon^{1/r-1}\log(\varepsilon^{-1})^{-(1+\delta)/r}\end{aligned}$$

Ainsi, on obtient dans ce cas

$$\left|(R^\varepsilon)'(x)\right| \geq C\varepsilon^{1/r}\log(\varepsilon^{-1})^{-(1+\delta)/r},$$

alors que, si les conclusions du Théorème 1.1.1 sont vraies en dimension $d = 1$, on a l'estimation suivante :

$$\left|(R^\varepsilon)'(x)\right| \leq C\varepsilon^{1/r}, \tag{2.63}$$

qui est donc quasi-optimale. □

### 2.2.7 $A$ et $A^T$

Le cadre théorique développé dans ce chapitre s'applique à des matrices qui ne sont pas nécessairement symétriques. Mais il sera parfois nécessaire de supposer que non seulement $A$, mais aussi $A^T$ satisfait des hypothèses sur ses correcteurs associés . En pratique, comme on considère souvent des classes de matrices invariantes par transposition (par exemple les matrices périodiques), la nuance entre les hypothèses sur $A$ et $A^T$ ne joue aucun rôle.

Insistons cependant sur le fait suivant : si $A$ satisfait les Hypothèses 1, 3, 4, 5 et 6, on en déduit par la Proposition 2.1.1 que $A$ s'homogénéise en $A^\star$. Alors il est classique $A^T$ s'homogénéise en $\left(A^T\right)^\star = (A^\star)^T$ (voir [145, Lem. 10.2 p. 118]). Toutefois, rien n'indique que l'existence de correcteurs $w^i$ strictement sous-linéaires associés à la matrice $A$ implique que l'on puisse construire des correcteurs $w_i^T$ strictement sous-linéaires associés à la matrice $A^T$, ce que nous illustrons dans le contre-exemple ci-dessous.

*Contre-exemple.* Considérons la matrice

$$A(x_1,x_2) := \begin{pmatrix} 1 & f(x_2) \\ 0 & 1 \end{pmatrix}.$$

Calculons

$$\mathrm{div}\,(A(x_1,x_2)\cdot\nabla v(x_1,x_2)) = \partial_{11}v(x_1,x_2) + \partial_{22}v(x_1,x_2) + f(x_2)\partial_{12}v(x_1,x_2),$$

et

$$\begin{aligned}&\mathrm{div}\left(A^T(x_1,x_2)\cdot\nabla v(x_1,x_2)\right)\\ &= \partial_{11}v(x_1,x_2) + \partial_{22}v(x_1,x_2) + f(x_2)\partial_{12}v(x_1,x_2) + f'(x_2)\partial_1 v(x_1,x_2).\end{aligned}$$

Ainsi, pour tout $i \in \{1,2\}$,

$$\operatorname{div}\left(A(x_1,x_2)\cdot e_i\right)=0,$$

d'où $w_i = 0$. En revanche,

$$\operatorname{div}\left(A^T(x_1,x_2)\cdot e_1\right)=f'(x_2),$$

donc $w_1^T$ n'est pas nul. Comme $A$ ne dépend que de $x_2$, on peut démontrer que tout correcteur strictement sous-linéaire $w_1^T$ de $A^T$ ne dépend que de $x_2$ (pour cela, on dérive l'équation par rapport à $x_1$ et on utilise le théorème de Liouville). Donc, par définition de $w_1^T$,

$$\left(w_1^T\right)''(x_2)+f'(x_2)=0.$$

D'où

$$w_1^T(x_2)=C_2-\int_0^{x_2}(f(z)+C_1)\mathrm{d}z.$$

Il est alors possible de construire une fonction $f$ telle que $w_1^T$ ne soit pas strictement sous-linéaire, typiquement

$$f(x_2)=\begin{cases}0 & \text{si il existe } n\in\mathbb{N} \text{ tel que } |x_2|\in\left[2^{2n},2^{2n+1}\right],\\ 1/2 & \text{si il existe } n\in\mathbb{N} \text{ tel que } |x_2|\in\left[2^{2n+1},2^{2n+2}\right].\end{cases}$$

□

### 2.2.8 Extensions possibles

Les auteurs de [11] et [94] étudient des systèmes, et non des équations, comme nous le faisons ici. Cette restriction a permis de simplifier un certain nombre de preuves (en utilisant notamment le principe du maximum). Nous ne pouvons pas affirmer avec certitude que l'ensemble des théorèmes que nous énonçons tient encore dans le cas de systèmes, quoique cela semble vraisemblable. En particulier, des résultats récents [25] démontrent l'existence de correcteurs dans le cas de coefficients périodiques perturbés par un défaut.

Les interfaces (voir [28]) constituent un autre problème qui n'est pas inclus dans le cadre théorique présenté dans ce chapitre. La principale difficulté de ce cas est le fait que le coefficient homogénéisé $A^\star$ n'est pas constant a priori (voir Section 2.2.2). Par conséquent, on ne peut pas effectuer telle quelle la preuve du Théorème 2.1.3, qui repose intimement sur le fait que les solutions de $-\operatorname{div}\left(A^\star\cdot\nabla u^\star\right)=0$ sont de classe $\mathrm{C}^2$ (en effet, si $A^\star$ est constant par morceaux, alors $u^\star$ est seulement de classe $\mathrm{W}^{1,\infty}$). Des recherches ultérieures seront entreprises afin d'établir quels résultats tiennent encore dans ce cas.

## 2.3 Résultats élémentaires

Dans cette section, nous construisons un potentiel $B$. Puis nous démontrons des résultats de régularité $\mathrm{C}^{1,\alpha}$ sur les correcteurs $w_j$ et le potentiel $B$ grâce à la théorie elliptique classique. Ensuite, nous prouvons la Proposition 2.1.8. Enfin, nous démontrons des estimations sur la quantité $H^\varepsilon$ définie par (2.33).

### 2.3.1 Construction d'un potentiel

Nous revenons maintenant sur la construction d'un potentiel $B$, que nous formalisons ainsi :

**Proposition 2.3.1** (Potentiel associé à un champ à divergence nulle)**.** *Soit $M_k^i$ un champ de vecteurs de classe* $\mathrm{BMO}(\mathbb{R}^d)$ *à divergence nulle, indicé par $k \in [\![1,d]\!]$, c'est à dire satisfaisant*

$$\mathrm{div}\,(M_k) = 0, \qquad \forall k \in [\![1,d]\!].$$

*A $M_k^i$, on peut associer un potentiel $B_k^{ij}$ satisfaisant* (2.28)*,* (2.29) *et* (2.30) *et tel que* $\nabla B \in \mathrm{BMO}\left(\mathbb{R}^d, \mathbb{R}^{d^4}\right)$.

L'espace BMO est le sous-espace de $\mathrm{L}^1_{\mathrm{loc}}\left(\mathbb{R}^d\right)$ induit par la norme

$$\|u\|_{\mathrm{BMO}(\mathbb{R}^d)} = \sup_{\mathrm{Q}} \fint_{\mathrm{Q}} \left| u(x) - \fint_{\mathrm{Q}} u \right| \mathrm{d}x,$$

où le supremum ci-dessus est pris sur tous les cubes Q de $\mathbb{R}^d$. La preuve ci-dessous suit la référence [114].

*Démonstration.* Tout d'abord, on réécrit (2.30) (de manière formelle) comme

$$\partial_l B_k^{ij}(x) = \lim_{R\to 0^+} \int_{\mathbb{R}^d\setminus \mathrm{B}(x,R)} \Big( \partial_{lj}\mathcal{G}_\Delta(x-y) M_k^i(y) - \partial_{li}\mathcal{G}_\Delta(x-y) M_k^j(y) \Big)\,\mathrm{d}y, \tag{2.64}$$

où $\mathcal{G}_\Delta$ est la fonction de Green sur $\mathbb{R}^d$ du Laplacien. Rappelons (voir [64, (2.12) p. 17]) que cette fonction s'exprime comme

$$\mathcal{G}_\Delta(x) = \frac{\Gamma(d/2)}{2(d-2)\pi^{d/2}} |x|^{2-d}.$$

Si $M$ est régulier à support compact, l'expression (2.64) ci-dessus a un sens. Nous justifions que cette expression a un sens plus général.

Pour ce faire, on introduit l'opérateur $T_j^i$ défini par

$$T_j^i : \begin{cases} \mathrm{C}_{\mathrm{c}}^\infty(\mathbb{R}^d) & \to \quad \mathrm{L}^2(\mathbb{R}^d), \\ f & \mapsto \quad \lim\limits_{R\to 0^+} \int_{\mathbb{R}^d\setminus \mathrm{B}(x,R)} \partial_{ij}\mathcal{G}_\Delta(x-y) f(y) \mathrm{d}y. \end{cases}$$

L'opérateur $T_j^i$ est un opérateur de Calderon-Zygmund (voir [114, Déf. 1 p. 224]). Par définition, on a

$$-\mathrm{div}\left(T^i(f)\right) = -\mathrm{div}\,(f e_i) \quad \text{et} \quad \partial_k T_j^i - \partial_j T_k^i = 0,$$

pour tout $k \in [\![1, d]\!]$. La dernière équation signifie simplement que $T^i$ s'écrit sous la forme d'un gradient. L'opérateur $T^i_j$ se prolonge continûment sur $\mathrm{L}^2(\mathbb{R}^d)$ grâce au théorème de Lax-Milgram.

Notons ensuite $F_R := T^i_j \left(x \mapsto \exp\left(-R^2x^2\right)\right)$. Alors, par un simple changement d'échelle, on a $F_R(x) = F_1(Rx)$. Ainsi :

$$T^i_j \left(x \mapsto \exp\left(-R^2x^2\right)\right) - F_1(0) \underset{R\to 0^+}{\longrightarrow} 0 \qquad \text{au sens des distributions.}$$

Autrement dit, l'opérateur $T^i_j$ satisfait la condition $T^i_j(1) = 0$ (voir [114, p. 222]). Par conséquent, grâce à [114, Cor. p. 239], on peut prolonger $T^i_j$ de $\mathrm{BMO}\left(\mathbb{R}^d\right)$ dans lui-même (aux constantes près).

Ainsi, l'expression à droite (2.64) est définie à une constante près pour un champ $M^i_k$ de vecteurs de classe $\mathrm{BMO}(\mathbb{R}^d)$. En outre, elle est à rotationnel nul : c'est donc un gradient. Ainsi, on donne le sens suivant à (2.64) pour un second membre $M^i_k$ de classe $\mathrm{BMO}(\mathbb{R}^d)$ :

$$\partial_l B^{ij}_k(x) = T^i_l M^j_k - T^j_l M^i_k + b^{ij}_{lk}$$

où $b^{ij}_{lk}$ sont des constantes que l'on va fixer. Afin que $B$ satisfasse (2.28), on impose que $b^{ij}_{lk}$ soit antisymétrique en $i$ et $j$. Comme $M_k$ est à divergence nulle, en dérivant (2.29), on obtient

$$\Delta\left(\sum_{i=1}^d \partial_i B^{ij}_k\right) = \Delta M^j_k.$$

Par le théorème de Liouville, cela signifie qu'il existe une constante $C^j_k$ telle que

$$\sum_{i=1}^d \partial_i B^{ij}_k = M^j_k + C^j_k.$$

On fixe alors $b^{ij}_{lk}$ de telle sorte que $C^j_k = 0$ pour tous $j, k \in [\![1, d]\!]$. Finalement $B$ ainsi construit satisfait (2.29). Notons cependant que $B^{ij}_k$ n'est pas unique. On peut lui ajouter les fonctions affines suivantes $x \mapsto c^{ij}_k + \sum_{l=1}^d b^{ij}_{lk} x_l$, pour toutes constantes $c^{ij}_k$ antisymétriques en $i$ et $j$, et $b^{ij}_{kl}$ antisymétriques en $i$ et $j$ satisfaisant $\sum_{i=1}^d b^{ij}_{ik} = 0$. □

Dans le cas où le champ de vecteur $M$ possède des propriétés d'intégrabilité, on peut démontrer des estimations plus fines sur le potentiel $B$ associé :

**Proposition 2.3.2.** *Supposons que $M \in \mathrm{L}^p\left(\mathbb{R}^d, \mathbb{R}^{d^2}\right)$ pour $p \in ]1, +\infty[$ est un champ de vecteur à divergence nulle. On peut alors construire un potentiel $B$ (satisfaisant* (2.28), (2.29) *et* (2.30)*) ayant les propriétés suivantes : Il existe une constante $C > 0$ tel que*

$$\|\nabla B\|_{\mathrm{L}^p\left(\mathbb{R}^d\right)} \leq C \|M\|_{\mathrm{L}^p\left(\mathbb{R}^d\right)}. \tag{2.65}$$

*Si* $M \in \mathrm{L}^{p_1}\left(\mathbb{R}^d, \mathbb{R}^{d^2}\right) \cap \mathrm{L}^{p_2}\left(\mathbb{R}^d, \mathbb{R}^{d^2}\right)$ *pour* $p_1 < d < p_2$*, alors, pour tout* $\rho > 0$ *:*

$$\|B\|_{\mathrm{L}^\infty\left(\mathbb{R}^d\right)} \leq C\rho^{\beta_{p_1}} \|M\|_{\mathrm{L}^{p_1}\left(\mathbb{R}^d\right)} + C\rho^{\beta_{p_2}} \|M\|_{\mathrm{L}^{p_2}\left(\mathbb{R}^d\right)}, \tag{2.66}$$

*où* $\beta_p$ *est définie par*

$$\beta_p := 1 - \frac{d}{p}. \tag{2.67}$$

*Démonstration de la Proposition 2.3.2.* On reprend la construction de la Proposition 2.3.1. L'estimation (2.65) est une conséquence de [114, p. 233].

Démontrons l'estimation (2.66). L'intégrabilité de $M$ permet la représentation suivante (issue de l'intégration de (2.64)) :

$$B_k^{ij}(x) := \int_{\mathbb{R}^d} \left(\partial_j \mathcal{G}_\Delta(x-y) M_k^i(y) - \partial_i \mathcal{G}_\Delta(x-y) M_k^j(y)\right) \mathrm{d}y, \tag{2.68}$$

dont nous justifions la convergence. La fonction $\nabla\mathcal{G}_\Delta(x)$ se décompose comme une somme de fonctions de classe $\mathrm{L}^{p_1'}\left(\mathbb{R}^d\right)$ et $\mathrm{L}^{p_2'}\left(\mathbb{R}^d\right)$

$$\nabla\mathcal{G}_\Delta(x) = \mathbb{1}_{[\rho,+\infty[}(|x|)\nabla\mathcal{G}_\Delta(x) + \mathbb{1}_{[0,\rho]}(|x|)\nabla\mathcal{G}_\Delta(x).$$

En effet, comme $|\nabla\mathcal{G}_\Delta(x)| \leq C|x|^{-d+1}$, on obtient

$$\left(\int_{\mathbb{R}^d} \left|\nabla\mathcal{G}_\Delta(x)\mathbb{1}_{[\rho,+\infty[}(|x|)\right|^{p_1'} \mathrm{d}x\right)^{1/p_1'} \leq C\rho^{\frac{p_1-d}{p_1}} \tag{2.69}$$

et

$$\left(\int_{\mathbb{R}^d} \left|\nabla\mathcal{G}_\Delta(x)\mathbb{1}_{[0,\rho]}(|x|)\right|^{p_2'} \mathrm{d}x\right)^{1/p_2'} \leq C\rho^{\frac{p_2-d}{p_2}}. \tag{2.70}$$

L'estimation (2.66) découle alors de l'inégalité de Hölder appliquée à (2.68), grâce à (2.69) et (2.70). □

### 2.3.2 Régularité des correcteurs et du potentiel

Les Hypothèses 7 et 8 n'informent en rien, a priori, sur la régularité des correcteurs $w_j$ et du potentiel $B$. En réalité, ceux-ci sont réguliers, ce qui permettra ultérieurement de démontrer des estimations fines sur $\nabla R^\varepsilon$.

**Proposition 2.3.3.** *Soit* $A$ *satisfaisant les Hypothèses 1, 2, 3 et 4. Alors les correcteurs* $w_j$ *satisfont, pour tout* $j \in [\![1, d]\!]$*,*

$$\nabla w_j \in \mathrm{C}^{0,\alpha}_{\mathrm{unif}}\left(\mathbb{R}^d, \mathbb{R}^d\right). \tag{2.71}$$

*Supposons que* $A$ *satisfait en outre l'Hypothèse 8. Alors*

$$\nabla B \in \mathrm{C}^{0,\alpha}_{\mathrm{unif}}\left(\mathbb{R}^d, \mathbb{R}^{d^4}\right). \tag{2.72}$$

*Remarque* 6 (Architecture logique). La Proposition 2.3.3 est utilisée pour démontrer l'Estimation (2.79), le Théorème 2.1.11 et la Proposition 2.7.4.

*Remarque* 7. Les Estimations (2.71) et (2.72) impliquent que les inégalités (2.17) et (2.34) demeurent valables pour $|x-y|<1$.

*Démonstration de la Proposition 2.3.3.* Soit $x\in\mathbb{R}^d$. Posons, pour $j\in[\![1,d]\!]$,

$$\chi_j(y):=y_j-x_j+(w_j(y)-w_j(x))\,.$$

La fonction $\chi_j$ vérifie :

$$\operatorname{div}\left(A\left(y\right)\cdot\nabla\chi_j(y)\right)=0 \qquad\qquad \text{dans} \quad \mathrm{B}(x,2).$$

Grâce à l'Hypothèse 1, on peut appliquer [64, Th. 8.2 p. 202], donc il existe $\beta>0$ tel que $\chi_j\in \mathrm{C}^{0,\beta}(\mathrm{B}(x,3/2))$, avec :

$$\|\chi_j\|_{\mathrm{C}^{0,\beta}(\mathrm{B}(x,3/2))}\leq C\,\|\chi_j\|_{\mathrm{L}^2(\mathrm{B}(x,2))}\,. \tag{2.73}$$

Puis, grâce à [64, Cor. 8.36 p. 212] ($A$ satisfait les Hypothèses 1 et 2), on obtient

$$\|\nabla\chi_j\|_{\dot{\mathrm{C}}^{0,\alpha}(\mathrm{B}(x,1))}\leq C\,\|\chi_j\|_{\mathrm{L}^\infty(\mathrm{B}(x,3/2))}\leq C\,\|\chi_j\|_{\mathrm{L}^2(\mathrm{B}(x,2))}\,. \tag{2.74}$$

En utilisant (2.73), une inégalité triangulaire, puis l'inégalité de Poincaré, il découle de l'inégalité précédente que

$$\|\nabla w_j\|_{\dot{\mathrm{C}}^{0,\alpha}(\mathrm{B}(x,1))}\leq C\,\|\nabla w_j\|_{\mathrm{L}^2(\mathrm{B}(x,2))}+C. \tag{2.75}$$

L'Hypothèse 4 permet alors d'établir (2.71).

Démontrons maintenant (2.72). Considérons le potentiel $B$ sur $\mathrm{B}(x,2)$. Il satisfait

$$-\Delta B_k^{ij}=\partial_j M_k^i-\partial_i M_k^j,$$

où $M$ est défini par (2.27). Donc, grâce à [64, Cor. 8.32 p. 210], $B\in \mathrm{C}^{1,\alpha}\left(\mathrm{B}(x,1),\mathbb{R}^{d^3}\right)$ et

$$\|\nabla B\|_{\dot{\mathrm{C}}^{0,\alpha}(\mathrm{B}(x,1))}\leq C\sup_{z\in\mathrm{B}(x,2)}|B(z)-B(x)|+C\,\|M\|_{\dot{\mathrm{C}}^{0,\alpha}(\mathrm{B}(x,2))}\,,$$

Grâce (2.71) et à l'Hypothèse 2, il existe une constante $C$ indépendante de $x$ telle que

$$\|M\|_{\dot{\mathrm{C}}^{0,\alpha}(\mathrm{B}(x,2))}\leq C. \tag{2.76}$$

Par conséquent, en utilisant l'Hypothèse 8, on obtient que

$$\|\nabla B\|_{\dot{\mathrm{C}}^{0,\alpha}(\mathrm{B}(x,1))}\leq C,$$

où $C$ est indépendant de $x$. D'où le résultat. □

### 2.3.3 Calcul algébrique justifiant la forme de $R^\varepsilon$

Dans cette section, nous démontrons la Proposition 2.1.8, grâce à des manipulations algébriques élémentaires (voir [85, p. 26] ou [27]).

*Remarque* 8 (Architecture logique). La Proposition 2.1.8 est utilisée pour borner $R^\varepsilon$ et son gradient $\nabla R^\varepsilon$. En particulier, elle sert à démontrer les Lemmes 2.6.1 et 2.6.2, le Théorème 2.1.11 et le Lemme 2.6.3.

*Démonstration de la Proposition 2.1.8.* Comme la matrice $A$ satisfait l'Hypothèse 4, alors on a $w_j\left(\frac{x}{\varepsilon}\right) \in \mathrm{H}^1(\Omega)$, et comme le potentiel $B$ est dans $\mathrm{L}^2_{\mathrm{loc}}\left(\mathbb{R}^d, \mathbb{R}^{d^3}\right)$, les manipulations qui suivent sont donc bien justifiées.

Par définition, et en utilisant (2.31),

$$\begin{aligned}
-\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla R^\varepsilon(x)\right) &= -\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla u^\varepsilon(x)\right)\\
&\quad + \mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla\left(u^\star(x)+\varepsilon\sum_{k=1}^d w_k\left(\frac{x}{\varepsilon}\right)\partial_k u^\star(x)\right)\right)\\
&= -\mathrm{div}\left(A^\star\cdot\nabla u^\star(x)\right)\\
&\quad + \mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla\left(u^\star(x)+\varepsilon\sum_{k=1}^d w_k\left(\frac{x}{\varepsilon}\right)\partial_k u^\star(x)\right)\right)\\
&= \sum_{i=1}^d \partial_i\left(\sum_{k=1}^d\left(\sum_{j=1}^d A_{ij}\left(\frac{x}{\varepsilon}\right)\left(\delta_{jk}+\partial_j w_k\left(\frac{x}{\varepsilon}\right)\right)-A^\star_{ik}\right)\partial_k u^\star(x)\right)\\
&\quad + \varepsilon\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right).\sum_{k=1}^d w_k\left(\frac{x}{\varepsilon}\right)\nabla\partial_k u^\star(x)\right). \qquad (2.77)
\end{aligned}$$

Constatons que

$$\sum_{i=1}^d \partial_i\left(\sum_{j=1}^d A_{ij}\left(\frac{x}{\varepsilon}\right)\left(\delta_{jk}+\partial_j w_k\left(\frac{x}{\varepsilon}\right)\right)-A^\star_{ik}\right) = \mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right).\left(\nabla w_k\left(\frac{x}{\varepsilon}\right)+e_k\right)\right)=0.$$

On injecte alors le potentiel $B$ associé à (2.27) dans l'équation précédente (2.77). Ainsi, en utilisant (2.29) et (2.28), on obtient

$$\begin{aligned}
&\sum_{i=1}^d \partial_i\left(\sum_{k=1}^d\left(\sum_{j=1}^d A_{ij}\left(\frac{x}{\varepsilon}\right)\left(\delta_{jk}+\partial_j w_k\left(\frac{x}{\varepsilon}\right)\right)-A^\star_{ik}\right)\partial_k u^\star(x)\right)\\
&= -\sum_{i=1}^d \partial_i\left(\sum_{k=1}^d\sum_{j=1}^d \partial_j B^{ji}_k\left(\frac{x}{\varepsilon}\right)\partial_k u^\star(x)\right)\\
&= -\sum_{i=1}^d \partial_i\left(\sum_{j,k=1}^d \varepsilon B^{ij}_k\left(\frac{x}{\varepsilon}\right)\partial_{jk} u^\star(x)\right).
\end{aligned}$$

D'où finalement (2.32). □

### 2.3.4 Bornes sur $H^\varepsilon$

Les Hypothèses 7 et 8 permettent de contrôler la quantité $H^\varepsilon$ définie par (2.33) dans différentes normes. En effet, supposons que $A$ satisfait les Hypothèses 1, 2, 3, 7 et 8. Quitte à enlever une constante, on suppose aussi que $w(0) = 0$ et que $B(0) = 0$. Alors $H^\varepsilon$ satisfait les estimations suivantes pour tous $R > \varepsilon > 0$, et $p \in [1, +\infty]$ :

$$\|H^\varepsilon\|_{\mathrm{L}^p(\Omega(0,R))} \leq C\varepsilon^\nu R^{1-\nu} \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^p(\Omega(0,R))}. \tag{2.78}$$

En outre, si $0 < \beta \leq \alpha$,

$$\|H^\varepsilon\|_{\dot{\mathrm{C}}^{0,\beta}(\Omega(0,R))} \leq C\varepsilon^\nu R^{1-\nu} \left\|\nabla^2 u^\star\right\|_{\dot{\mathrm{C}}^{0,\beta}(\Omega(0,R))} + C\varepsilon^{\nu-\beta} R^{1-\nu} \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^\infty(\Omega(0,R))}. \tag{2.79}$$

Dans (2.78) et (2.79), les constantes $C$ ne dépendent pas de $\varepsilon$.

*Remarque* 9 (Architecture logique). Les estimations (2.78) et (2.79), combinées avec la Proposition 2.1.8, seront utilisées pour démontrer des résultats d'approximation sur la solution du problème oscillant. Elle servent à démontrer les Lemmes 2.6.1 et 2.6.2, le Théorème 2.1.11, et le Lemme 2.6.3.

*Démonstration des estimations* (2.78) *et* (2.79). Par définition (2.33),

$$\|H^\varepsilon\|_{\mathrm{L}^p(\Omega(0,R))} \leq C \left\| \left| \varepsilon w\left(\frac{\cdot}{\varepsilon}\right) \right| + \left| \varepsilon B\left(\frac{\cdot}{\varepsilon}\right) \right| \right\|_{\mathrm{L}^\infty(\Omega(0,R))} \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^p(\Omega(0,R))}.$$

Les Hypothèses 7 et 8 impliquent alors (2.78).

On démontre (2.79) par un calcul direct (pour ne pas surcharger les notations, on omet l'espace $\Omega(0,R)$ sur lequel sont pris toutes les normes ci-dessous) :

$$\begin{aligned}
\|H^\varepsilon\|_{\dot{\mathrm{C}}^{0,\beta}} \leq & C \left\| \left| \varepsilon w\left(\frac{\cdot}{\varepsilon}\right) \right| + \left| \varepsilon B\left(\frac{\cdot}{\varepsilon}\right) \right| \right\|_{\mathrm{L}^\infty} \left\|\nabla^2 u^\star\right\|_{\dot{\mathrm{C}}^{0,\beta}} \\
& + C \left\| A\left(\frac{\cdot}{\varepsilon}\right) \right\|_{\dot{\mathrm{C}}^{0,\beta}} \left\| \varepsilon w\left(\frac{\cdot}{\varepsilon}\right) \right\|_{\mathrm{L}^\infty} \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^\infty} \\
& + C \left\| \varepsilon w\left(\frac{\cdot}{\varepsilon}\right) \right\|_{\dot{\mathrm{C}}^{0,\beta}} \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^\infty} + \left\| \varepsilon B\left(\frac{\cdot}{\varepsilon}\right) \right\|_{\dot{\mathrm{C}}^{0,\beta}} \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^\infty}.
\end{aligned} \tag{2.80}$$

Or, par les Hypothèses 7 et 8,

$$\left\| \left| \varepsilon w\left(\frac{\cdot}{\varepsilon}\right) \right| + \left| \varepsilon B\left(\frac{\cdot}{\varepsilon}\right) \right| \right\|_{\mathrm{L}^\infty} \leq C\varepsilon^\nu R^{1-\nu}, \tag{2.81}$$

puis, par l'Hypothèse 2,

$$\left\| A\left(\frac{\cdot}{\varepsilon}\right) \right\|_{\dot{\mathrm{C}}^{0,\beta}} \leq C\varepsilon^{-\beta}. \tag{2.82}$$

En invoquant la Proposition 2.3.3, on a, pour tout $x \neq y \in \Omega(0,R)$,

$$\frac{|\varepsilon w(x/\varepsilon) - \varepsilon w(y/\varepsilon)|}{|x-y|^\beta} \leq C \|\nabla w\|_{\mathrm{L}^\infty(\mathbb{R}^d)} |x-y|^{1-\beta} \leq C|x-y|^{1-\beta}, \tag{2.83}$$

puis, grâce à l'Hypothèse 7,

$$\frac{|\varepsilon w(x/\varepsilon) - \varepsilon w(y/\varepsilon)|}{|x-y|^\beta} \leq C\varepsilon^\nu |x-y|^{1-\nu-\beta}. \tag{2.84}$$

On utilise l'Inégalité (2.83) pour $|x-y| \in [0,\varepsilon]$, et l'Inégalité (2.84) pour $|x-y| \in [\varepsilon, R]$, d'où

$$\left\| \varepsilon w\left(\frac{\cdot}{\varepsilon}\right) \right\|_{\dot{\mathrm{C}}^{0,\beta}} \leq \max\left(\varepsilon^{1-\beta}, \varepsilon^\nu R^{1-\nu-\beta}\right).$$

Comme $\varepsilon < R$, on a donc

$$\left\| \varepsilon w\left(\frac{\cdot}{\varepsilon}\right) \right\|_{\dot{\mathrm{C}}^{0,\beta}} \leq C\varepsilon^{\nu-\beta} R^{1-\nu}. \tag{2.85}$$

On montre de la même manière que

$$\left\| \varepsilon B\left(\frac{\cdot}{\varepsilon}\right) \right\|_{\dot{\mathrm{C}}^{0,\beta}} \leq C\varepsilon^{\nu-\beta} R^{1-\nu}. \tag{2.86}$$

Alors, (2.79) découle de (2.80), (2.81), (2.82), (2.85), et (2.86). □

## 2.4 Estimations dans le cas homogène

### 2.4.1 Présentation de la méthode compacité à la Avellaneda et Lin

Dans toute la Section 2.4, on fera usage de la méthode de compacité de [11]. Le but de cette méthode est de démontrer que si une fonction $u^\varepsilon$ satisfait

$$-\mathrm{div}\left(A\left(\frac{x}{\varepsilon} + y\right) \cdot \nabla u^\varepsilon(x)\right) = 0, \tag{2.87}$$

alors $u^\varepsilon$ est régulière (hölderienne ou lipschitzienne), et ce, *indépendamment de* $\varepsilon$. Cette régularité est a priori surprenante, car on s'attend formellement à ce que $u^\varepsilon(x) = u(x/\varepsilon)$, ce qui induirait alors que son gradient $\nabla u^\varepsilon(x) = \varepsilon^{-1}\nabla u(x/\varepsilon)$ exploserait quand $\varepsilon \to 0$. Or, dans les cas considérés (par exemple, le cas périodique pour [11]), il n'en est rien.

Cette propriété est due au fait que, quand $\varepsilon \to 0$, $u^\varepsilon$ est de plus en plus proche de la solution d'un problème homogénéisé $u^\star$ qui satisfait

$$-\mathrm{div}\left(A^\star \cdot \nabla u^\star\right) = 0. \tag{2.88}$$

Or, comme $A^\star$ est constante, une solution $u^\star$ d'un tel problème est naturellement très régulière, et notamment lipschitzienne. L'essence de l'approche de [11] est de faire en sorte que $u^\varepsilon$ hérite de la régularité du problème homogénéisé (2.88). Nous démontrons donc tout d'abord un résultat d'uniforme H-convergence (la Proposition 2.1.1), afin de pouvoir adapter la preuve de [11].

Puis, trois étapes sont nécessaires pour démontrer un résultat de régularité sur $u^\varepsilon$ : une initialisation, une itération et un blow-up (voir Figure 2.1). Nous illustrons ce déroulement sur le schéma de preuve du Théorème 2.1.3 ci-dessous. Par souci de simplicité, nous prenons $y = 0$ dans le schéma de preuve. La généralisation au cas où $y \neq 0$ ne présente pas de difficulté, et utilise simplement l'uniformité en $y$ des Hypothèses 4, 5 et 6 (voir Remarque 10 ci-dessous).

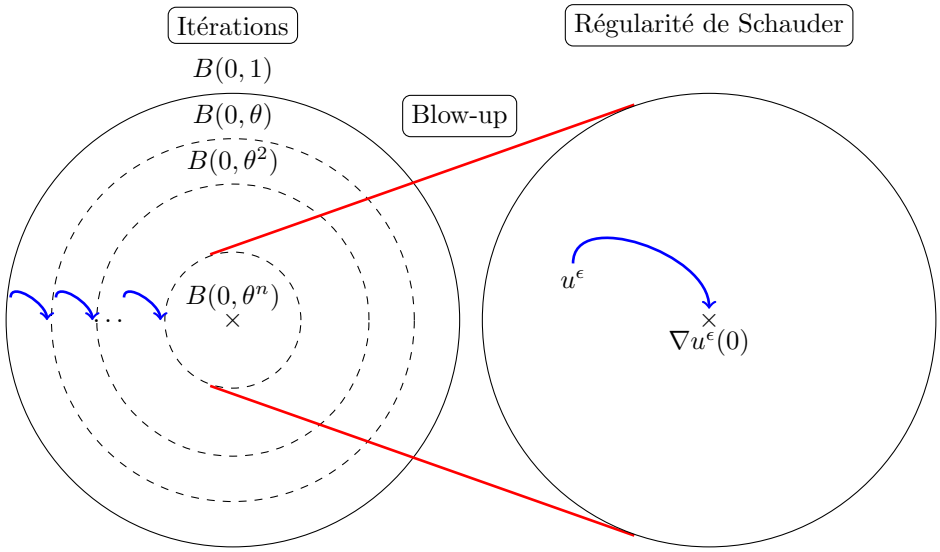



FIGURE 2.1 – Principe de la méthode de compacité de [11]

**Initialisation :** (voir [11, Lem. 14], avec un second membre nul) on démontre tout d'abord l'estimation

$$\sup_{x\in \mathrm{B}(0,\theta)} \left| u^\varepsilon(x) - u^\varepsilon(0) - \sum_{i=1}^{d} \left\{ x_i + \varepsilon w_i\left(\frac{x}{\varepsilon}\right)\right\} \fint_{\mathrm{B}(0,\theta)} \partial_i u^\varepsilon \right| \leq \theta^{1+\gamma} \left\| u^\varepsilon \right\|_{\mathrm{L}^2(\mathrm{B}(0,1))}, \qquad (2.89)$$

uniforme en $\varepsilon$ suffisamment petit, à une échelle fixée $\theta \in ]0,1[$ suffisamment petite, et pour $\gamma > 0$ fixé.

Pour ce faire, on suppose par l'absurde que (2.89) n'est pas satisfaite pour une suite de fonctions $u^\varepsilon$ avec $\varepsilon \to 0$. L'uniforme H-convergence et le Lemme A.3.5 impliquent alors que la solution $u^\star$ du problème limite (2.88) ne satisfait pas

$$\sup_{x\in \mathrm{B}(0,\theta)} \left| u^\star(x) - u^\star(0) - \sum_{i=1}^{d} x_i \fint_{\mathrm{B}(0,\theta)} \partial_i u^\star \right| \leq \theta^{1+\gamma} \left\| u^\star \right\|_{\mathrm{L}^2(\mathrm{B}(0,1))}, \qquad (2.90)$$

ce qui est contradictoire avec le fait que $u^\star$ est régulier (notons que (2.89) et (2.90) peuvent être vus comme des développements de Taylor). Une propriété essentielle à cette étape est que les correcteurs $w_i$ sont strictement sous-linéaires (voir Lemme 2.2.1). Ainsi, leur contribution à (2.89) disparaît dans la limite $\varepsilon \to 0$.

**Itération :** (voir [11, Lem. 15]) on répète l'étape précédente sur les boules $\mathrm{B}(0,\theta^2)$, $\mathrm{B}(0,\theta^3)$, etc., jusqu'à l'échelle $\theta^n$ d'ordre $\varepsilon$. Le point clef de cette étape est que la fonc-

tion

$$v(x) := u^\varepsilon(x) - u^\varepsilon(0) - \sum_{i=1}^{d} \left\{ x_i + \varepsilon w_i \left(\frac{x}{\varepsilon}\right) \right\} \fint_{\mathrm{B}(0,\theta)} \partial_i u^\varepsilon$$

est elle-même une solution de l'équation (2.87), par définition des correcteurs.

**Blow-up :** (voir [11, Lem. 16]) on déduit de l'étape d'itération précédente et de la sous-linéarité des correcteurs que « $u^\varepsilon$ est sous-linéaire jusqu'à l'échelle $\theta^n \propto \varepsilon$ ». c'est à dire que

$$\sup_{|x|>\theta^n} \frac{|u^\varepsilon(x) - u^\varepsilon(0)|}{|x|} \leq C. \tag{2.91}$$

Pour conclure, il faut pouvoir majorer le taux d'accroissement de $u^\varepsilon$, c'est à dire s'autoriser $|x| \to 0$ dans la formule précédente. Pour ce faire, on utilise la théorie classique de Schauder à l'échelle $\theta^n$ pour estimer

$$|\nabla u^\varepsilon(0)| \leq C\theta^{-n} \left\| u^\varepsilon(x) - u^\varepsilon(0) \right\|_{\mathrm{L}^\infty(\mathrm{B}(0,\theta^n))}.$$

On obtient alors la majoration souhaitée grâce à (2.91). Toutefois, lors du *blow-up*, il faut assurer un contrôle en norme $\mathrm{L}^\infty$ sur les correcteurs rescalés $\varepsilon w_j(\cdot/\varepsilon)$. Ce contrôle se fait via la propriété de stricte sous-linéarité sur les correcteurs $w_j$ indiquée par le Lemme 2.2.1. Rappelons que, dans le cas périodique, les correcteurs $w_j$ sont bornés dans $\mathrm{L}^\infty\left(\mathbb{R}^d\right)$, ce qui n'est pas satisfait en général dans les cas que l'on étudie (voir Section 2.7.1).

*Remarque* 10. Il est bon d'insister sur un point technique de la preuve des Lemmes d'initialisations 2.4.1, 2.4.3, 2.4.5 et 2.4.8 : les estimations énoncées sont indépendantes de $y$, lorsque l'on considère l'opérateur $-\mathrm{div}\left(A(\cdot/\varepsilon + y) \cdot \nabla\right)$. Ce point est fondamental pour conclure. Il motive la nécessité d'avoir une H-convergence des matrices $A\left(\frac{\cdot}{\varepsilon} + y\right)$ *uniforme* sur tout l'espace. Il induit ainsi l'uniformité en $y$ dans la formulation des Hypothèses 4, 5 et 6.

Les démonstrations de chacun des trois Théorèmes 2.1.2, 2.1.3, et 2.1.4, reprennent ces trois étapes avec des degrés de technicité divers, et des estimations adaptées à la régularité souhaitée. A une exception près, ces adaptations sont directes, dans le sens où on remplace le caractère périodique des correcteurs des preuves originelles par leur caractère strictement sous-linéaire ici (cela implique généralement de se donner un point de repère $x_0$ pour fixer la constante des correcteurs ; en pratique, on considère la fonction $w_j(\cdot) - w_j(x_0)$ plutôt que la fonction $w_j$ elle-même). L'exception notable concerne la démonstration du Théorème 2.1.4, qui nécessite de faire appel à des correcteurs $w_j^\Omega$ adaptés au domaine (voir la Section 2.4.6). Nous devons alors quantifier la stricte sous-linéarité des correcteurs afin d'obtenir des propriétés satisfaisantes sur les correcteurs adaptés $w_j^\Omega$.

### 2.4.2 Notations

Nous présentons ici quelques notations concernant certains ouverts particuliers au bord d'un domaine. Soit $\Omega$ un ouvert régulier borné de $\mathbb{R}^d$. Nous rappelons les notations (2.13) :

$$\Omega(x,R) = \Omega \cap \mathrm{B}(x,R), \quad \text{et} \quad \Gamma_\Omega(x,R) = \partial\Omega \cap \overline{\mathrm{B}(x,R)}.$$

Puis nous définissons la notion de bord d'ouvert délimité par un graphe. Soit $\phi \in \mathrm{C}^{1,\beta}\left(\mathbb{R}^{d-1}\right)$ satisfaisant

$$\phi(0) = 0, \quad \nabla\phi(0) = 0 \quad \text{et} \quad \|\phi\|_{\mathrm{C}^{1,\beta}\left(\mathbb{R}^{d-1}\right)} \leq K_0, \tag{2.92}$$

pour $\beta \in ]0,1[$, $K_0 > 0$. On pose

$$D_\phi(R) := \left\{x \in \mathbb{R}^d, x_d < \phi\left((x_1, \cdots, x_{d-1})\right)\right\} \cap \mathrm{B}(0,R), \tag{2.93}$$

$$\Delta_\phi(R) := \Gamma_{D_\phi(R)}(0,R) = \left\{x \in \mathbb{R}^d, x_d = \phi\left((x_1, \cdots, x_{d-1})\right)\right\} \cap \overline{\mathrm{B}(0,R)}. \tag{2.94}$$

Enfin, nous introduisons le cône tronqué

$$C_{K_0}(R) := \left\{x \in \mathrm{B}(0,R), x_d < 0, |x - x_d e_d| < K_0 |x_d|\right\}.$$

Notons que, grâce à (2.92), $C_{K_0}(R) \subset D_\phi(R)$.

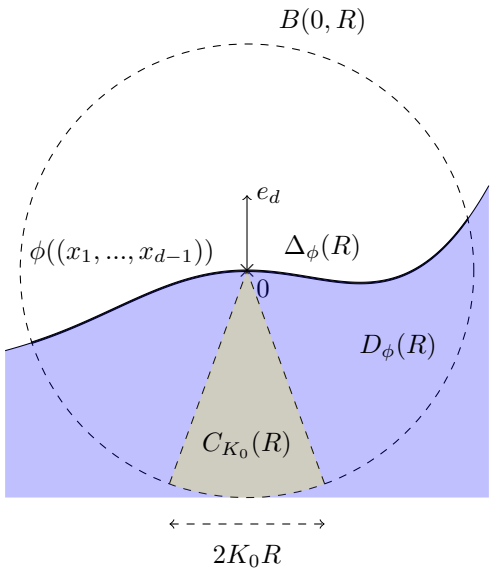



FIGURE 2.2 – Ouverts utiles au bord d'un domain $\Omega$

### 2.4.3 Uniforme H-convergence

Dans cette section, nous discutons et démontrons la Proposition 2.1.1.

*Remarque* 11. Grâce aux propriétés de la H-convergence (voir [145, Th. 6.5 p. 82]), $A^\star$ satisfait l'Hypothèse 1.

*Remarque* 12. Comme on peut le remarquer dans la démonstration, il suffit que *l'un des deux* champs de matrices $A$ ou $A^T$ satisfasse les Hypothèses 1, 3, 4, 5 et 6 pour que les conclusions de la Proposition 2.1.1 soient valides.

*Remarque* 13 (Architecture logique). La Proposition 2.1.1 est la pierre angulaire de la méthode de compacité de Avellaneda et Lin. C'est lui qui a motivé la construction des Hypothèses 4, 5 et 6, qui le démontrent, via le Lemme div-rot (Lemme A.3.1 en Annexe). Grâce à lui, on démontre les étapes d'initialisation des Théorèmes 2.1.2, 2.1.3 et 2.1.4.

*Démonstration de la Proposition 2.1.1.* On se donne $y_n \in \mathbb{R}^d$, $\varepsilon_n \to 0$, et on note $A_n(x) := A(y_n + x/\varepsilon_n)$. Par [145, Lem. 10.2 p. 118], $A_n$ H-converge vers $A^\star$ si et seulement si $A_n^T$ H-converge vers $(A^\star)^T$. Nous allons donc montrer que $A_n^T$ H-converge vers $(A^\star)^T$.

Soit $\Omega \subset \mathbb{R}^d$ un ouvert borné régulier et $f \in \mathrm{H}^{-1}(\Omega)$. Supposons que $u^n \in \mathrm{H}^1(\Omega)$ satisfait

$$-\mathrm{div}\left(A_n^T(x) \cdot \nabla u^n(x)\right) = f(x) \quad \text{dans} \quad \Omega,$$

et que la convergence faible suivante a lieu :

$$u^n \rightharpoonup u^\star \quad \text{dans} \quad \mathrm{H}^1(\Omega).$$

Montrons qu'alors on a

$$A_n^T \cdot \nabla u^n \rightharpoonup (A^\star)^T \cdot \nabla u^\star \qquad \text{dans} \quad \mathrm{L}^2\left(\Omega, \mathbb{R}^d\right). \tag{2.95}$$

Quitte à extraire, il existe $g \in \mathrm{L}^2\left(\Omega, \mathbb{R}^d\right)$ tel que

$$A_n^T \cdot \nabla u^n \rightharpoonup g \quad \text{dans} \quad \mathrm{L}^2(\Omega, \mathbb{R}^d).$$

Soit $\phi \in \mathrm{C}_0^1(\Omega)$, c'est à dire que $\phi \in \mathrm{C}^1(\Omega)$ est nulle sur $\partial\Omega$. Considérons

$$Q_i(n) := \int_\Omega \phi(x) A_n^T(x) \cdot \nabla u^n(x) \cdot \nabla \left(x_i + \varepsilon_n w_i\left(\frac{x}{\varepsilon_n} + y_n\right)\right) \mathrm{d}x.$$

Cette dernière quantité est bien définie car la fonction

$$\psi^n(x) := \nabla \left(x_i + \varepsilon_n w_i\left(\frac{x}{\varepsilon_n} + y_n\right)\right) \tag{2.96}$$

est dans $\mathrm{L}^2(\Omega)$ grâce à l'Hypothèse 4. Par ailleurs, on a

$$-\mathrm{div}\left(A_n^T(x) \cdot \nabla u^n(x)\right) = f(x) \qquad \text{et} \qquad -\mathrm{div}\left(A_n(x) \cdot \psi^n(x)\right) = 0,$$

et, grâce à (2.52) et (2.53) (corollaire des Hypothèses 5 et 6 respectivement),

$$A_n \cdot \psi_i^n \rightharpoonup A^\star \cdot e_i \quad \text{dans} \quad \mathrm{L}^2\left(\Omega, \mathbb{R}^d\right) \qquad \text{et} \qquad \psi^n \rightharpoonup e_i \quad \text{dans} \quad \mathrm{L}^2\left(\Omega, \mathbb{R}^d\right).$$

En utilisant deux fois le Lemme A.3.1 -avec $A_n \cdot \psi^n$ et $\nabla u^n$ puis avec $A_n^T \cdot \nabla u^n$ et $\psi^n$- on obtient que :

$$Q_i(n) \underset{n\to+\infty}{\to} \int_\Omega \phi(x) A^\star \cdot e_i \cdot \nabla u^\star(x) \mathrm{d}x,$$
$$Q_i(n) \underset{n\to+\infty}{\to} \int_\Omega \phi(x) g(x) \cdot e_i \mathrm{d}x.$$

Ceci étant vrai pour tout $\phi \in \mathrm{C}_0^1(\Omega)$, alors

$$g(x) = (A^\star)^T \cdot \nabla u^\star(x).$$

D'où (2.95). Par conséquent, $A_n^T$ H-converge vers $(A^\star)^T$. □

### 2.4.4 Estimations hölderiennes

L'objet de cette section est de discuter et démontrer le Théorème 2.1.2, par la méthode de compacité de [11].

*Remarque* 14. Notons que dans le cas d'une équation, le Théorème 2.1.2 est une version plus puissante du Théorème de De Giorgi-Nash-Moser en ce sens qu'il affirme que $u^\varepsilon$ est hölderienne pour un exposant *égal* à celui des éventuelles conditions de bord. Au contraire, le Théorème de De Giorgi-Nash-Moser affirme seulement l'existence d'un exposant $\beta \in ]0,1[$ tel que $u^\varepsilon$ est $\beta$-hölderienne.

*Remarque* 15 (Architecture logique). Le Théorème 2.1.2 repose essentiellement sur le Théorème de De Giorgi-Nash-Moser (voir Lemme A.3.5) et sur la Proposition 2.1.1. Il est ensuite utilisé pour démontrer des estimations lipschitziennes jusqu'au bord : la Proposition 2.4.7 et le Théorème 2.1.4.

La preuve du Théorème 2.1.2, correspond à [11, Th. 1] et suit exactement la preuve de cette référence. Vis-à-vis du schéma de démonstration exposé dans la Section 2.4.1, elle est plus simple, et ne fait pas intervenir de correcteurs (mais requiert cependant la Proposition 2.1.1).

*Remarque* 16. Dans le Théorème 2.1.2, on peut avoir éventuellement $\Gamma_\Omega(0,1) = \emptyset$.

La démonstration se fait en deux temps : tout d'abord, on démontre que le théorème est valide si $\Gamma_\Omega(0,1) = \emptyset$ (c'est à dire un cas où il n'y a pas de frontière). Puis, on démontre qu'il est aussi vrai si $\Gamma_\Omega(0,1) \neq \emptyset$.

#### Estimations hölderiennes intérieures

Nous commençons par l'étape d'initialisation :

**Lemme 2.4.1** (Analogue du Lem. 7 de [11])**.** *Supposons que $A$ satisfait les Hypothèses 1, 3, 4, 5, et 6. Soit $\beta \in ]0,1[$. Il existe une constante $\theta \in ]0,1/4[$ ne dépendant que de $\mu$ et $\beta$, et une constante $\varepsilon_0$ dépendant de $A$, $\beta$ et $\theta$ telle que, si $u^\varepsilon$ est solution de*

$$-\operatorname{div}\left(A\left(\frac{x}{\varepsilon}+y\right)\cdot\nabla u^\varepsilon(x)\right)=0 \qquad \text{dans} \quad \mathrm{B}(0,1), \tag{2.97}$$

*alors*

$$\fint_{\mathrm{B}(0,\theta)}\left|u^\varepsilon(x)-\fint_{\mathrm{B}(0,\theta)}u^\varepsilon\right|^2\mathrm{d}x\leq\theta^{2\beta}\fint_{\mathrm{B}(0,1)}|u^\varepsilon(x)|^2\,\mathrm{d}x. \tag{2.98}$$

*Démonstration.* Si $u^\star$ est solution de

$$-\operatorname{div}\left(A^\star\cdot\nabla u^\star\right)=0 \qquad \text{dans} \quad \mathrm{B}(0,1/2), \tag{2.99}$$

alors, comme $u^\star$ est $A^\star$-harmonique sur $\mathrm{B}(0,1)$ (avec $A^\star$ une matrice constante),

$$\fint_{\mathrm{B}(0,\theta)}\left|u^\star(x)-\fint_{\mathrm{B}(0,\theta)}u^\star\right|^2\mathrm{d}x\leq C\theta^2\left\|\nabla u^\star\right\|_{\mathrm{L}^\infty(\mathrm{B}(0,2/3))}\leq C\theta^2\fint_{\mathrm{B}(0,1/2)}|u^\star(x)|^2\,\mathrm{d}x.$$

D'où l'existence d'une constante $\theta > 0$ suffisamment petite telle que

$$\fint_{\mathrm{B}(0,\theta)} \left| u^\star(x) - \fint_{\mathrm{B}(0,\theta)} u^\star \right|^2 \mathrm{d}x \leq \frac{\theta^{2\beta}}{2^{d+1}} \fint_{\mathrm{B}(0,1/2)} |u^\star(x)|^2 \,\mathrm{d}x. \tag{2.100}$$

Par ailleurs, l'ensemble de fonctions $u^\varepsilon$ satisfaisant (2.97), où $y \in \mathbb{R}^d$, et

$$\fint_{\mathrm{B}(0,1)} |u^\varepsilon(x)|^2 \,\mathrm{d}x \leq 1, \tag{2.101}$$

est borné dans $\mathrm{H}^1\left(\mathrm{B}(0,1/2)\right)$ grâce à l'inégalité de Cacciopoli (voir Lemme A.3.2). Grâce à la Proposition 2.1.1, on sait en outre qu'il admet dans son adhérence, sur $\mathrm{B}(0,1/2)$ pour la topologie $H^1$-faible, quand $\varepsilon \to 0$, des solutions $u^\star$ de (2.99). Par conséquent, il découle de (2.100) qu'il existe $\varepsilon_0$ tel que, pour tout $\varepsilon < \varepsilon_0$, si $u^\varepsilon$ est solution de (2.97), alors $u^\varepsilon$ satisfait

$$\fint_{\mathrm{B}(0,\theta)} \left| u^\varepsilon(x) - \fint_{\mathrm{B}(0,\theta)} u^\varepsilon \right|^2 \mathrm{d}x \leq 2^{-d}\theta^{2\beta} \fint_{\mathrm{B}(0,1/2)} |u^\varepsilon(x)|^2 \,\mathrm{d}x \leq \theta^{2\beta} \fint_{\mathrm{B}(0,1)} |u^\varepsilon(x)|^2 \,\mathrm{d}x.$$

□

On peut alors itérer le Lemme précédent :

**Lemme 2.4.2** (Analogue du Lem. 8 de [11])**.** *Sous les hypothèses du Lemme 2.4.1, en se donnant $\theta$ et $\varepsilon_0$ issus du Lemme 2.4.1, si $\varepsilon \leq \theta^k \varepsilon_0$, toute solution $u^\varepsilon$ de* (2.97) *satisfait alors*

$$\fint_{\mathrm{B}(0,\theta^k)} \left| u^\varepsilon(x) - \fint_{\mathrm{B}(0,\theta^k)} u^\varepsilon \right|^2 \mathrm{d}x \leq \theta^{2k\beta} \fint_{\mathrm{B}(0,1)} |u^\varepsilon(x)|^2 \,\mathrm{d}x. \tag{2.102}$$

*Démonstration.* En appliquant le Lemme 2.4.1, on obtient (2.102) pour $k = 1$. On procède ensuite par récurrence et on suppose le Lemme 2.4.2 vrai jusqu'au rang $k$. On considère ensuite

$$v(z) = u^\varepsilon(\theta^k z) - \fint_{\mathrm{B}(0,\theta^k)} u^\varepsilon,$$

qui est solution de

$$-\operatorname{div}\left(A\left(\frac{z}{\theta^{-k}\varepsilon}\right) \cdot \nabla v(z)\right) = 0 \qquad \text{dans} \quad \mathrm{B}\left(0,1\right).$$

En appliquant alors le Lemme 2.4.1 à $v$, on obtient

$$\fint_{\mathrm{B}(0,\theta)} \left| v(z) - \fint_{\mathrm{B}(0,\theta)} v \right|^2 \mathrm{d}z \leq \theta^{2\beta} \fint_{\mathrm{B}(0,1)} |v(z)|^2 \,\mathrm{d}z,$$

c'est à dire

$$\fint_{\mathrm{B}(0,\theta^{k+1})} \left| u^\varepsilon(x) - \fint_{\mathrm{B}(0,\theta^{k+1})} u^\varepsilon(x) \right|^2 \mathrm{d}x \leq \theta^{2\beta} \fint_{\mathrm{B}(0,\theta^k)} \left| u^\varepsilon(x) - \fint_{\mathrm{B}(0,\theta^k)} u^\varepsilon \right|^2 \mathrm{d}x.$$

En utilisant l'hypothèse de récurrence, on en déduit alors que le Lemme 2.4.2 est vrai jusqu'au rang $k+1$. D'où le résultat par récurrence. □

*Démonstration du Théorème 2.1.2 dans le cas où* $\Gamma_\Omega(0,1) = \emptyset$. La démonstration ci-dessous suit la preuve de [11, Lem. 9].

On utilise pour ce faire la caractérisation de Campanato (voir Lemme A.3.6). Ainsi, on souhaite borner la quantité

$$\sup_{x\in\mathrm{B}(0,1/4), r>0} \left( r^{-d-2\beta} \int_{\widetilde{\mathrm{B}}(x,r)} \left| u^\varepsilon - \fint_{\widetilde{\mathrm{B}}(x,r)} u^\varepsilon \right|^2 \right)^{1/2} \quad x \geq C^{-1} \left\| u^\varepsilon \right\|_{\mathrm{C}^{0,\beta}(\mathrm{B}(0,1/4))}, \qquad (2.103)$$

où $\widetilde{\mathrm{B}}(x,r) := \mathrm{B}(x,r) \cap \mathrm{B}(0,1/4)$ et $C > 0$.

On se donne $n$ tel que $\theta^{n+1}\varepsilon_0 \leq \varepsilon \leq \theta^n \varepsilon_0$. Si tout d'abord $\theta^{k+1} \leq 2r \leq \theta^k$, pour $k \leq n$, alors, par le Lemme 2.4.2, pour tout $x \in \mathrm{B}(0,1/4)$,

$$\theta^{-2k\beta} \fint_{\mathrm{B}(x,\theta^k)} \left| u^\varepsilon(y) - \fint_{\mathrm{B}(x,\theta^k)} u^\varepsilon \right|^2 \mathrm{d}y \leq C \fint_{\mathrm{B}(x,1/2)} |u^\varepsilon(y)|^2 \,\mathrm{d}y \leq C \fint_{\mathrm{B}(0,1)} |u^\varepsilon(y)|^2 \,\mathrm{d}y.$$

D'où

$$r^{-2\beta} \fint_{\mathrm{B}(x,r)} \left| u^\varepsilon(y) - \fint_{\mathrm{B}(x,r)} u^\varepsilon \right|^2 \mathrm{d}y \leq C \fint_{\mathrm{B}(0,1)} |u^\varepsilon(y)|^2 \,\mathrm{d}y.$$

Si au contraire $2r < \theta^{n+1}$, alors il découle de la régularité elliptique classique (en appliquant successivement [64, Cor. 8.36 p. 212] puis le Lemme A.3.4) que, comme $v : z \mapsto u^\varepsilon(\theta^n z)$ satisfait

$$-\mathrm{div}\left( A\left(\frac{z}{\theta^{-n}\varepsilon}\right) \cdot \nabla v(z) \right) = 0,$$

alors

$$\fint_{\mathrm{B}(x,r)} \left| u^\varepsilon(y) - \fint_{\mathrm{B}(x,r)} u^\varepsilon \right|^2 \mathrm{d}y \leq C r^{2\beta} \theta^{-2n\beta} \fint_{\mathrm{B}(x,\theta^n)} \left| u^\varepsilon(y) - \fint_{\mathrm{B}(x,\theta^n)} u^\varepsilon \right|^2 \mathrm{d}y.$$

Or, vu ce qui précède,

$$\fint_{\mathrm{B}(x,\theta^n)} \left| u^\varepsilon(y) - \fint_{\mathrm{B}(x,\theta^n)} u^\varepsilon \right|^2 \mathrm{d}y \leq C \theta^{2n\beta} \fint_{\mathrm{B}(0,1)} |u^\varepsilon(y)|^2 \,\mathrm{d}y.$$

Ainsi, dans tous les cas

$$\sup_{x\in \mathrm{B}(0,1/4), r>0} \left( r^{-d-2\beta} \int_{\widetilde{\mathrm{B}}(x,r)} \left| u^\varepsilon - \fint_{\widetilde{\mathrm{B}}(x,r)} u^\varepsilon \right|^2 \right)^{1/2} \leq C \left\| u^\varepsilon \right\|_{\mathrm{L}^2(\mathrm{B}(0,1))},$$

et enfin, grâce à (2.103),

$$\left\| u^\varepsilon \right\|_{\mathrm{C}^{0,\beta}(\mathrm{B}(0,1/4))} \leq C \left\| u^\varepsilon \right\|_{\mathrm{L}^2(\mathrm{B}(0,1))},$$

dont découle le résultat. □

**Estimations hölderiennes au bord**

Dans cette section, nous discutons et démontrons le Théorème 2.1.2 dans le cas général, c'est à dire où $\Omega(0,1)$ n'est pas nécessairement une boule. Pour cela, nous suivons la démonstration de [11, Lem. 24]. Les Lemmes 2.4.3 et 2.4.4, d'initialisation et d'itération, sont très semblables aux Lemmes 2.4.1 et 2.4.2 ; le rôle joué par la moyenne de $u^\varepsilon$ est cependant remplacé par la condition de bord. Dans le déroulement de la preuve, nous supposons tout d'abord que $u^\varepsilon$ satisfait des conditions de bord nulles.

*Remarque* 17. En toute rigueur, il faudrait prendre un domaine $\Omega$ variable dans la preuve par l'absurde du Lemme 2.4.3. Nous ne le faisons pas ici, mais renvoyons le lecteur à la preuve du Lemme 2.4.8, où cette subtilité est traitée dans la preuve d'une estimation lipschitzienne.

**Lemme 2.4.3** (Analogue du Lem. 22 de [11])**.** *Supposons que $A$ satisfait les Hypothèses 1, 3, 4, 5, et 6. Soit $\beta > 0$ et $\Omega$ un ouvert borné régulier, avec $0 \in \partial\Omega$. Il existe une constante $\theta \in ]0, 1/4[$ ne dépendant que de $\mu$, $\Omega$ et $\beta$, et une constante $\varepsilon_0$ dépendant de $A$, $\Omega$, $\beta$ et $\theta$ telle que, si $u^\varepsilon$ est solution de*

$$\begin{cases} -\operatorname{div}\left( A\left(\dfrac{x}{\varepsilon} + y\right) \cdot \nabla u^\varepsilon(x) \right) = 0 & \text{dans} \quad \Omega(0,1), \\ u^\varepsilon = 0 & \text{sur} \quad \Gamma_\Omega(0,1), \end{cases} \tag{2.104}$$

*alors*

$$\fint_{\Omega(0,\theta)} |u^\varepsilon(x)|^2 \,\mathrm{d}x \leq \theta^{2\beta} \fint_{\Omega(0,1)} |u^\varepsilon(x)|^2 \,\mathrm{d}x. \tag{2.105}$$

*Démonstration.* Si $u^\star$ est solution de

$$\begin{cases} -\operatorname{div}\left( A^\star \cdot \nabla u^\star \right) = 0 & \text{dans} \quad \Omega(1/2), \\ u^\star = 0 & \text{sur} \quad \Gamma_\Omega(1/2), \end{cases} \tag{2.106}$$

alors, grâce à [64, Cor. 8.36 p. 212], on en déduit que, si $\theta < 1/4$,

$$\fint_{\Omega(0,\theta)} |u^\star(x)|^2 \,\mathrm{d}x \leq C\theta^2 \left\| \nabla u^\star \right\|_{\mathrm{L}^\infty(\Omega(0,1/4))} \leq C\theta^2 \left\| u^\star \right\|_{\mathrm{L}^\infty(\Omega(0,1/3))}.$$

Donc, grâce au Lemme A.3.4,

$$⨍_{\Omega(0,\theta)} |u^\star(x)|^2 \,\mathrm{d}x \leq C\theta^2 ⨍_{\Omega(0,1/2)} |u^\star(x)|^2 \,\mathrm{d}x.$$

D'où l'existence d'une constante $\theta > 0$ suffisamment petite telle que

$$⨍_{\Omega(0,\theta)} |u^\star(x)|^2 \,\mathrm{d}x \leq C_0 \frac{\theta^{2\beta}}{2} ⨍_{\Omega(0,1/2)} |u^\star(x)|^2 \,\mathrm{d}x, \tag{2.107}$$

où, pour la régularité du bord considérée, on a

$$C_0 < \frac{|\Omega(0,1/2)|}{|\Omega(0,1)|}.$$

Par ailleurs, l'ensemble de fonctions $u^\varepsilon$ satisfaisant (2.104), où $y \in \mathbb{R}^d$, et

$$⨍_{\Omega(0,1)} |u^\varepsilon(x)|^2 \,\mathrm{d}x \leq 1, \tag{2.108}$$

est borné dans $\mathrm{H}^1(\Omega(0,1/2))$ grâce à l'inégalité de Cacciopoli (voir Lemme A.3.2). Grâce à la Proposition 2.1.1, on sait en outre qu'il admet dans son adhérence, sur $\Omega(0,1/2)$ pour la topologie $H^1$-faible, quand $\varepsilon \to 0$, des solutions $u^\star$ de (2.106). Par conséquent, il découle de (2.107) qu'il existe $\varepsilon_0$ tel que, pour tout $\varepsilon < \varepsilon_0$, si $u^\varepsilon$ est solution de (2.104), alors $u^\varepsilon$ satisfait

$$⨍_{\Omega(0,\theta)} |u^\varepsilon(x)|^2 \,\mathrm{d}x \leq C_0\theta^{2\beta} ⨍_{\Omega(0,1/2)} |u^\varepsilon(x)|^2 \,\mathrm{d}x \leq \theta^{2\beta} ⨍_{\Omega(0,1)} |u^\varepsilon(x)|^2 \,\mathrm{d}x.$$

□

On peut alors itérer le Lemme précédent :

**Lemme 2.4.4** (Analogue du Lemme 23 de [11])**.** *Supposons que $A$ satisfait les Hypothèses 1, 3, 4, 5, et 6. Soit $\Omega$ un ouvert borné régulier, avec $0 \in \partial\Omega$. Si on se donne $\theta$ et $\varepsilon_0$ issus du Lemme 2.4.3, si $\varepsilon \leq \theta^k \varepsilon_0$, pour toute solution $u^\varepsilon$ de* (2.104)*, on a*

$$⨍_{\Omega(0,\theta^k)} |u^\varepsilon(x)|^2 \,\mathrm{d}x \leq \theta^{2k\beta} ⨍_{\Omega(0,1)} |u^\varepsilon(x)|^2 \,\mathrm{d}x. \tag{2.109}$$

*Démonstration.* En appliquant le Lemme 2.4.3, on obtient (2.109) pour $k = 1$. On procède ensuite par récurrence et on suppose le Lemme 2.4.4 vrai jusqu'au rang $k$. On considère ensuite $v(z) = u^\varepsilon(\theta^k z)$, qui est solution de

$$\begin{cases} -\operatorname{div}\left(A\left(\dfrac{z}{\theta^{-k}\varepsilon} + y\right) \cdot \nabla v(z)\right) = 0 & \text{dans} \quad \theta^{-k}\Omega\left(0,\theta^k\right), \\ v(z) = 0 & \text{sur} \quad \theta^{-k}\Gamma_\Omega(0,\theta^k). \end{cases} \tag{2.110}$$

En appliquant alors le Lemme 2.4.3 à $v$, on obtient

$$⨍_{\Omega(0,\theta)} |v(z)|^2 \,\mathrm{d}z \leq \theta^{2\beta} ⨍_{\Omega(0,1)} |v(z)|^2 \,\mathrm{d}z.$$

Puis, en remettant à l'échelle et en utilisant l'hypothèse de récurrence, on en déduit alors que le Lemme 2.4.4 est vrai jusqu'au rang $k+1$. D'où le résultat par récurrence. □

Nous pouvons maintenant effectuer l'étape de blow-up, dans le cas où $g = 0$. Le cas où $g \neq 0$ sera traité juste après.

*Démonstration du Théorème 2.1.2 dans le cas où $g = 0$.* On utilise pour ce faire la caractérisation de Campanato (voir Lemme A.3.6). Ainsi, on souhaite borner la quantité

$$\sup_{x\in\Omega(0,1/32),r>0} \left( r^{-d-2\beta} \int_{\widetilde{\Omega}(x,r)} \left| u^\varepsilon - ⨍_{\widetilde{\Omega}(x,r)} u^\varepsilon \right|^2 \right)^{1/2} \geq C^{-1} \left\| u^\varepsilon \right\|_{\mathrm{C}^{0,\beta}(\Omega(0,1/32))}, \tag{2.111}$$

où $\widetilde{\Omega}(x,r) := \Omega(x,r) \cap \Omega(0,1/32)$ et $C > 0$.

On se donne $n$ tel que $\theta^{n+1}\varepsilon_0 \leq \varepsilon \leq \theta^n \varepsilon_0$. Supposons tout d'abord que $x \in \Gamma_\Omega(0,1/4)$, et $r \in ]0,1/4[$. Si tout d'abord $\theta^{k+1} \leq 2r \leq \theta^k$, pour $k \leq n$, alors, par le Lemme 2.4.4, pour tout $x \in \Omega(0,1/4)$,

$$\theta^{-2k\beta} ⨍_{\Omega(x,\theta^k)} |u^\varepsilon(y)|^2 \,\mathrm{d}y \leq ⨍_{\Omega(x,1/2)} |u^\varepsilon(y)|^2 \,\mathrm{d}y \leq C ⨍_{\Omega(0,1)} |u^\varepsilon(y)|^2 \,\mathrm{d}y. \tag{2.112}$$

D'où

$$r^{-2\beta} ⨍_{\Omega(x,r)} \left| u^\varepsilon(y) - ⨍_{\Omega(x,r)} u^\varepsilon \right|^2 \mathrm{d}y \leq r^{-2\beta} ⨍_{\Omega(x,r)} |u^\varepsilon(y)|^2 \,\mathrm{d}y \leq C ⨍_{\Omega(0,1)} |u^\varepsilon(y)|^2 \,\mathrm{d}y. \tag{2.113}$$

Si au contraire $2r < \theta^{n+1}$, alors il découle de la régularité elliptique classique (en appliquant successivement [64, Cor. 8.36 p. 212] puis le Lemme A.3.4) et du fait que $u^\varepsilon(x) = 0$ que

$$⨍_{\Omega(x,r)} |u^\varepsilon(y)|^2 \,\mathrm{d}y \leq C r^{2\beta} \theta^{-2n\beta} ⨍_{\Omega(x,\theta^n)} |u^\varepsilon(y)|^2 \,\mathrm{d}y.$$

Par conséquent, vu (2.112),

$$r^{-2\beta} ⨍_{\Omega(x,r)} \left| u^\varepsilon(y) - ⨍_{\Omega(x,r)} u^\varepsilon \right|^2 \mathrm{d}y \leq C ⨍_{\Omega(0,1)} |u^\varepsilon(y)|^2 \,\mathrm{d}y.$$

Ainsi, dans tous les cas, si $x \in \Gamma_\Omega(0,1/4)$,

$$r^{-2\beta} ⨍_{\Omega(x,r)} \left| u^\varepsilon(y) - ⨍_{\Omega(x,r)} u^\varepsilon \right|^2 \mathrm{d}y \leq C ⨍_{\Omega(0,1)} |u^\varepsilon(y)|^2 \,\mathrm{d}y. \tag{2.114}$$

On prend ensuite $x \in \Omega(0,1/32)\backslash\Gamma_\Omega(0,1)$. Clairement, si $\Gamma_\Omega(0,1/16) = \emptyset$, on peut se ramener au cas sans frontière traité ci-dessus. Ainsi, on suppose que $\Gamma_\Omega(0,1/16) \neq \emptyset$. On peut alors choisir $x' \in \Gamma_\Omega(0,1/16)$ tel que $r_0 := \mathrm{d}(x, \Gamma_\Omega(0,1/16)) = |x-x'| \leq 1/8$, où $\mathrm{d}(y,F)$ désigne la distance du point $y$ à l'ensemble $F$. Ainsi, si $r \geq r_0$, on a $\Omega(x,r) \subset \Omega(x',2r)$, d'où

$$\fint_{\Omega(x,r)} \left| u^\varepsilon(y) - \fint_{\Omega(x,r)} u^\varepsilon \right|^2 \mathrm{d}y \leq C \fint_{\Omega(x',2r)} \left| u^\varepsilon(y) - \fint_{\Omega(x',2r)} u^\varepsilon \right|^2 \mathrm{d}y.$$

Vu (2.114), on obtient alors

$$\fint_{\Omega(x,r)} \left| u^\varepsilon(y) - \fint_{\Omega(x,r)} u^\varepsilon \right|^2 \mathrm{d}y \leq C r^{2\beta} \fint_{\Omega(0,1)} |u^\varepsilon(y)|^2 \,\mathrm{d}y. \tag{2.115}$$

Traitons maintenant le cas $r < r_0$. Comme $\mathrm{B}(x,r_0) \subset \Omega(0,1)$, en utilisant le Théorème 2.1.2 dans le cas sans frontière, on obtient

$$\fint_{\mathrm{B}(x,r)} \left| u^\varepsilon(y) - \fint_{\mathrm{B}(x,r)} u^\varepsilon \right|^2 \mathrm{d}y \leq C (r/r_0)^{2\beta} \fint_{\mathrm{B}(x,r_0)} \left| u^\varepsilon(y) - \fint_{\mathrm{B}(x,r_0)} u^\varepsilon \right|^2 \mathrm{d}y.$$

D'où, grâce à (2.115),

$$\fint_{\mathrm{B}(x,r)} \left| u^\varepsilon(y) - \fint_{\mathrm{B}(x,r)} u^\varepsilon \right|^2 \mathrm{d}y \leq C r^{2\beta} \fint_{\Omega(0,1)} |u^\varepsilon(y)|^2 \,\mathrm{d}y.$$

Par conséquent, on a démontré que, dans tous les cas,

$$\sup_{\substack{x \in \Omega(0,1/32) \\ 0<r<1/32}} r^{-2\beta-d} \int_{\Omega(x,r)} \left| u^\varepsilon(y) - \fint_{\Omega(x,r)} u^\varepsilon \right|^2 \mathrm{d}y \leq C \fint_{\Omega(0,1)} |u^\varepsilon(y)|^2 \,\mathrm{d}y.$$

D'où, grâce à (2.111),

$$\|u^\varepsilon\|_{\mathrm{C}^{0,\beta}(\Omega(0,1/32))} \leq C \left( \fint_{\Omega(0,1)} |u^\varepsilon(y)|^2 \,\mathrm{d}y \right)^{1/2}.$$

Quitte à faire un autre recouvrement par des boules plus petites, on a établi le Théorème 2.1.2 dans le cas où $g = 0$. □

La démonstration dans le cas où $g$ est non nul est faite ci-dessous :

*Démonstration du Théorème 2.1.2 dans le cas où $g \neq 0$.* Quitte à prolonger $g$, on peut faire en sorte que $g$ soit défini sur $\mathrm{B}(0,1)$, avec

$$\|g\|_{\mathrm{C}^{0,\beta}(\mathrm{B}(0,1))} \leq C \|g\|_{\mathrm{C}^{0,\beta}(\Gamma_\Omega(0,1))}.$$

On utilise encore la caractérisation de Campanato (voir Lemme A.3.6), et on va montrer récursivement sur $r$ que

$$r^{-2\beta-d}\int_{\widetilde{\Omega}(x,r)}\left|u^\varepsilon-\fint_{\widetilde{\Omega}(x,r)}u^\varepsilon\right|^2\leq C\fint_{\Omega(0,1)}\left|u^\varepsilon-\fint_{\Omega(0,1)}u^\varepsilon\right|^2+C\left\|g\right\|^2_{\mathrm{C}^{0,\beta}(\mathrm{B}(0,1))}, \quad (2.116)$$

où $\widetilde{\Omega}(x,r):=\Omega(x,r)\cap\Omega(0,1/4)$. Si tel est le cas, alors le Lemme A.3.6 permet de conclure la démonstration. La démonstration de (2.116) se fait par récurrence et réutilise le Théorème 2.1.2 dans le cas où $g=0$.

**Etape 1 : Initialisation** Soit $\delta=(1-\beta)/2$. Nous prétendons qu'il existe des constantes $C$ et $\theta\in]0,1[$ indépendantes de $\varepsilon$ telles que, pour tout $x\in\Omega(0,1/4)$ et $r\leq 1/4$,

$$\theta^{-2\beta}\fint_{\Omega(x,\theta r)}\left|u^\varepsilon-\fint_{\Omega(x,\theta r)}u^\varepsilon\right|^2\leq\theta^\delta\fint_{\Omega(x,r)}\left|u^\varepsilon-\fint_{\Omega(x,r)}u^\varepsilon\right|^2+Cr^{2\beta}\left\|g\right\|^2_{\mathrm{C}^{0,\beta}(\mathrm{B}(0,1))}. \quad (2.117)$$

Pour démontrer (2.117), on décompose $u^\varepsilon=u_1^\varepsilon+u_2^\varepsilon$ sur $\Omega(x,2r)$, où

$$\begin{cases}-\operatorname{div}\left(A\left(\dfrac{x}{\varepsilon}+y\right)\cdot\nabla u_1^\varepsilon(x)\right)=0 & \text{dans}\quad \Omega(x,r),\\ u_1^\varepsilon=0 & \text{sur}\quad \Gamma_\Omega(x,r),\end{cases}$$

et

$$\begin{cases}-\operatorname{div}\left(A\left(\dfrac{x}{\varepsilon}+y\right)\cdot\nabla u_2^\varepsilon(x)\right)=0 & \text{dans}\quad \Omega(x,r),\\ u_2^\varepsilon=g & \text{sur}\quad \partial\left(\Omega(x,r)\right).\end{cases}$$

Grâce au cas où $g=0$ du Théorème 2.1.2 (démontré ci-dessus) remis à l'échelle $r$, on sait que

$$\begin{aligned}\fint_{\Omega(x,\theta r)}\left|u_1^\varepsilon-\fint_{\Omega(x,\theta r)}u_1^\varepsilon\right|^2&\leq(\theta r)^{2(\beta+\delta)}\left\|u_1^\varepsilon\right\|^2_{\dot{\mathrm{C}}^{0,\beta+2\delta}(\Omega(x,\theta r))}\\&\leq C\theta^{2\beta+2\delta}\fint_{\Omega(x,r)}\left|u_1^\varepsilon-\fint_{\Omega(x,r)}u_1^\varepsilon\right|^2.\end{aligned}$$

Ainsi, quitte à prendre $\theta$ suffisamment petit, on obtient

$$\theta^{-2\beta}\fint_{\Omega(x,\theta r)}\left|u_1^\varepsilon-\fint_{\Omega(x,\theta r)}u_1^\varepsilon\right|^2\leq\theta^\delta\fint_{\Omega(x,r)}\left|u_1^\varepsilon-\fint_{\Omega(x,r)}u_1^\varepsilon\right|^2.$$

Alors, par inégalité triangulaire,

$$\begin{aligned}\theta^{-2\beta}\fint_{\Omega(x,\theta r)}\left|u_1^\varepsilon-\fint_{\Omega(x,\theta r)}u_1^\varepsilon\right|^2\leq&\,\theta^\delta\fint_{\Omega(x,r)}\left|u^\varepsilon-\fint_{\Omega(x,r)}u^\varepsilon\right|^2\\&+\theta^\delta\fint_{\Omega(x,r)}\left|u_2^\varepsilon-\fint_{\Omega(x,r)}u_2^\varepsilon\right|^2. \quad (2.118)\end{aligned}$$

Par ailleurs, grâce au principe du maximum appliqué à $u_2^\varepsilon - g(x_0)$, pour $x_0 \in \partial(\Omega(x,r))$,

$$\fint_{\Omega(x,r)} \left| u_2^\varepsilon - \fint_{\Omega(x,r)} u_2^\varepsilon \right|^2 \leq C r^{2\beta} \|g\|^2_{\mathrm{C}^{0,\beta}(\mathrm{B}(0,1))} . \tag{2.119}$$

L'Estimation (2.117) découle de (2.118) et (2.119).

**Etape 2 : Itération** En itérant l'estimation (2.117), on obtient alors, pour tout $k \in \mathbb{N}^\star$ que

$$\begin{aligned} \theta^{-2k\beta} \fint_{\Omega(x,\theta^k/4)} \left| u^\varepsilon - \fint_{\Omega(x,\theta^k/4)} u^\varepsilon \right|^2 \leq & C \sum_{j=0}^{k-1} \theta^{j\delta} \|g\|^2_{\mathrm{C}^{0,\beta}(\mathrm{B}(0,1))} \\ & + C \fint_{\Omega(x,1/4)} \left| u^\varepsilon - \fint_{\Omega(x,1/4)} u^\varepsilon \right|^2 . \end{aligned} \tag{2.120}$$

D'où

$$\begin{aligned} \theta^{-2k\beta-d} \int_{\widetilde{\Omega}(x,\theta^{k+1}/4)} \left| u^\varepsilon - \fint_{\widetilde{\Omega}(x,\theta^{k+1}/4)} u^\varepsilon \right|^2 \leq & C \fint_{\Omega(0,1)} \left| u^\varepsilon - \fint_{\Omega(0,1)} u^\varepsilon \right|^2 \\ & + C \|g\|^2_{\mathrm{C}^{0,\beta}(\mathrm{B}(0,1))} . \end{aligned} \tag{2.121}$$

**Etape 3 : Conclusion** De (2.121), on déduit que, pour tout $x \in \Omega(0,1/4)$, et tout $r < 1/4$, on a (2.116). D'où le résultat, grâce à la caractérisation de Campanato. □

### 2.4.5 Estimations lipschitziennes intérieures

Dans cette section, nous discutons et démontrons le Théorème 2.1.3, en suivant la preuve du Lemme 16 de [11].

*Remarque* 18. Dans (2.16), la constante $C$ dépend de manière subtile des Hypothèses 5 et 6, qui ne sont pas quantitatives, mais qualitatives. Ainsi, on ne peut pas expliciter la dépendance de $C$ en les paramètres, avec une preuve par l'absurde. Toutefois, il est possible de le faire si on ne considère plus une seule matrice, mais une classe de matrices, satisfaisant uniformément, dans un certain sens, les Hypothèses 1, 2, 3, 4, 5 et 6 (voir par exemple [94], où on considère des matrices périodiques uniformément régulières, elliptiques et bornées). Par souci de simplicité, nous n'utilisons pas cette approche.

*Remarque* 19 (Architecture logique). Le Théorème 2.1.3 est central ; en découlent toutes les estimations sur le gradient de la solution du problème oscillant (2.1). Il repose sur la méthode de compacité d'Avellanedda et Lin, et emploie donc la Proposition 2.1.1, ainsi que le Théorème de De Giorgi-Nash-Moser (remplaçable par le Théorème 2.1.2). Il sert à démontrer :

1. sa propre extension jusqu'au bord (le Théorème 2.1.4),

2. des estimations sur les gradients et gradient croisé de la fonction de Green $G^\varepsilon$ (le Théorème 2.1.5),
3. des estimations $\mathrm{L}^p$ (la Proposition 2.1.6).

Pour démontrer le Théorème 2.1.3, nous rappelons tout d'abord des bornes uniformes et hölderiennes en faisant appel au Théorème de De Giorgi-Nash-Moser (on peut aussi invoquer le Théorème 2.1.2, mais qui est superflu dans le cas des équations). En effet, grâce au Lemme A.3.5 et à une remise à l'échelle, il existe $\beta \in ]0,1[$, $C > 0$ ne dépendant que de $\mu$ telles que

$$\sup_{x\in \mathrm{B}(0,R)} |u^\varepsilon(x)| \leq C \left( \fint_{\mathrm{B}(0,2R)} |u^\varepsilon|^2 \right)^{1/2}, \tag{2.122}$$

et

$$\sup_{x,x'\in \mathrm{B}(0,R)} \frac{|u^\varepsilon(x) - u^\varepsilon(x')|}{|x-x'|^\beta} \leq CR^{-\beta} \left( \fint_{\mathrm{B}(0,2R)} |u^\varepsilon|^2 \right)^{1/2}. \tag{2.123}$$

Puis, nous utilisons le schéma de preuve explicité dans la Section 2.4.1.

L'initialisation consiste en le lemme suivant :

**Lemme 2.4.5** (Analogue du Lemme 14 de [11])**.** *Soit $A$ satisfaisant les Hypothèses 1, 3, 4, 5 et 6. Soit $\gamma \in ]0,1[$. Alors il existe $\theta \in ]0,1/4[$ ne dépendant que de $\mu$ et $\gamma$, et $\varepsilon_0$ ne dépendant que de $A$, $\gamma$ et $\theta$ tels que, si $u^\varepsilon \in \mathrm{H}^1(\mathrm{B}(0,1))$ satisfait*

$$-\mathrm{div}\left( A\left(\frac{x}{\varepsilon} + y\right) \cdot \nabla u^\varepsilon(x) \right) = 0 \quad \textit{dans} \quad \mathrm{B}(0,1), \tag{2.124}$$

*pour $\varepsilon \leq \varepsilon_0$ et $y \in \mathbb{R}^d$, alors*

$$\begin{aligned} &\sup_{x\in \mathrm{B}(0,\theta)} \left| u^\varepsilon(x) - u^\varepsilon(0) - \sum_{i=1}^d \left\{ x_i + \varepsilon \left( w_i\left(\frac{x}{\varepsilon} + y\right) - w_i(y) \right) \right\} \fint_{\mathrm{B}(0,\theta)} \partial_i u^\varepsilon(z) \mathrm{d}z \right| \\ &\leq \theta^{1+\gamma} \left( \fint_{\mathrm{B}(0,1)} |u^\varepsilon(z)|^2 \, \mathrm{d}z \right)^{1/2}. \end{aligned} \tag{2.125}$$

*Remarque* 20. Comme on le voit dans la preuve ci-dessous, on peut supprimer le correcteur $w$ de l'Estimation (2.125). Toutefois, lorsqu'on voudra itérer le Lemme 2.4.5, comme fait plus bas dans le Lemme 2.4.6, sa présence va se révéler cruciale. En effet, le terme à l'intérieur de la valeur absolue à gauche de (2.125) est dans le noyau de $-\mathrm{div}\left(A\left(\frac{\cdot}{\varepsilon} + y\right) \cdot \nabla\right)$.

La preuve se fait par l'absurde. Une inégalité plus forte que (2.125) est en effet satisfaite par les solutions d'une équation elliptique à coefficient constant. On se sert alors de la H-convergence de $A\left(\frac{\cdot}{\varepsilon} + y\right)$ vers $A^\star$ et d'une compacité dans les fonctions hölderiennes due à l'estimation (2.123) pour démontrer qu'il est impossible que $u^\varepsilon$ ne vérifie pas (2.126) pour $\varepsilon$ suffisamment petit.

*Démonstration du Lemme 2.4.5.* On commence par établir qu'il existe $\theta \in ]0,1/4[$ ne dépendant que de $\mu$ et de $\gamma$, et $\varepsilon_1$ ne dépendant que de $A$ et $\theta$ tel que pour tout $\varepsilon < \varepsilon_1$

$$\sup_{x\in \mathrm{B}(0,\theta)} \left| u^\varepsilon(x) - u^\varepsilon(0) - \sum_{i=1}^d x_i \fint_{\mathrm{B}(0,\theta)} \partial_i u^\varepsilon(z) \mathrm{d}z \right| \leq \frac{\theta^{1+\gamma}}{2} \left( \fint_{\mathrm{B}(0,1)} |u^\varepsilon(x)|^2 \, \mathrm{d}x \right)^{1/2}. \quad (2.126)$$

Puis, on montre qu'il existe un $\varepsilon_0 < \varepsilon_1$ ne dépendant que de $A$ tel que pour tout $\varepsilon < \varepsilon_0$

$$\begin{aligned} &\sup_{x\in \mathrm{B}(0,\theta)} \left| \sum_{i=1}^d \varepsilon \left( w_i \left( \frac{x}{\varepsilon} + y \right) - w_i(y) \right) \fint_{\mathrm{B}(0,\theta)} \partial_i u^\varepsilon(z) \mathrm{d}z \right| \\ &\leq \frac{\theta^{1+\gamma}}{2} \left( \fint_{\mathrm{B}(0,1)} |u^\varepsilon(x)|^2 \, \mathrm{d}x \right)^{1/2}. \end{aligned} \quad (2.127)$$

Ensemble, (2.126) et (2.127) concluent la preuve.

Pour établir (2.126), on utilise le problème homogénéisé et la compacité induite par la Proposition 2.1.1. Parallèlement, on établit (2.127) en utilisant la propriété de sous-linéarité stricte des correcteurs.

**Etude du problème homogénéisé** Soit $u^\star \in \mathrm{H}^1(\mathrm{B}(0,1/2))$ satisfaisant

$$-\mathrm{div}\left(A^\star \cdot \nabla u^\star(x)\right) = 0 \quad \text{dans} \quad \mathrm{B}(0,1/2). \quad (2.128)$$

Par développement de Taylor, on a

$$\sup_{x\in \mathrm{B}(0,r)} \left| u^\star(x) - u^\star(0) - x \cdot \fint_{\mathrm{B}(0,r)} \nabla u^\star(z) \mathrm{d}z \right| \leq Cr^2 \sup_{x\in \mathrm{B}(0,r)} \left| \nabla^2 u^\star(x) \right|. \quad (2.129)$$

Or, comme $A^\star$ est constante et satisfait l'Hypothèse 1 (voir [64, Th. 2.10 p. 23]), il existe une constante $C$ ne dépendant que de $\mu$ telle que

$$\sup_{x\in \mathrm{B}(0,1/4)} \left| \nabla^2 u^\star(x) \right| \leq C \int_{\mathrm{B}(0,1/2)} |u^\star(x)|^2 \, \mathrm{d}x. \quad (2.130)$$

Donc il existe une constante $C_0$ ne dépendant que de $\mu$ telle que, pour tout $\theta \in ]0,1/4[$,

$$\sup_{x\in \mathrm{B}(0,\theta)} \left| u^\star(x) - u^\star(0) - x \cdot \fint_{\mathrm{B}(0,\theta)} \nabla u^\star(z) \mathrm{d}z \right| \leq C_0 \theta^2 \left( \fint_{\mathrm{B}(0,1/2)} |u^\star(x)|^2 \, \mathrm{d}x \right)^{1/2}. \quad (2.131)$$

On se fixe alors $\theta$ tel que

$$2^{-d/2-1}\theta^{1+\gamma} = 2C_0\theta^2 \Longleftrightarrow \theta = \left( \frac{1}{C_0 2^{d/2+2}} \right)^{\frac{1}{1-\gamma}}, \quad (2.132)$$

et qui ne dépend donc que de $\gamma$ et de $\mu$.

**Compacité** Par l'absurde, supposons qu'il existe une suite $y_n \in \mathbb{R}^d$, $\varepsilon_n \to 0$ et $u^n \in \mathrm{H}^1(\mathrm{B}(0,1))$ satisfaisant

$$-\mathrm{div}\left(A\left(\frac{x}{\varepsilon_n}+y_n\right)\cdot\nabla u^n(x)\right)=0 \quad \text{dans} \quad \mathrm{B}(0,1), \tag{2.133}$$

avec

$$\sup_{x\in\mathrm{B}(0,\theta)}\left|u^n(x)-u^n(0)-x\cdot\fint_{\mathrm{B}(0,\theta)}\nabla u^n(z)\mathrm{d}z\right|>\frac{\theta^{1+\gamma}}{2}\left(\fint_{\mathrm{B}(0,1)}|u^n(x)|^2\,\mathrm{d}x\right)^{1/2}. \tag{2.134}$$

Quitte à renormaliser $u^n$, on suppose que

$$\fint_{\mathrm{B}(0,1)}|u^n(x)|^2\,\mathrm{d}x=1.$$

Par l'inégalité de Cacciopoli (Lemme A.3.2), les fonctions $u^n$ sont uniformément bornées dans $\mathrm{H}^1(\mathrm{B}(0,1/2))$. Donc, quitte à extraire,

$$u^n \rightharpoonup u^\star \quad \text{dans} \quad \mathrm{H}^1(\mathrm{B}(0,1/2)).$$

Grâce à (2.133), la Proposition 2.1.1 implique que $u^\star$ est solution de (2.128). Ainsi, $u^\star$ satisfait (2.131).

Par ailleurs, grâce à l'Estimation (2.123), il existe $\beta>0$ tel que $u^n$ est uniformément bornée dans $\mathrm{C}^{0,\beta}(\mathrm{B}(0,1/2))$. Ainsi $u^n$ converge uniformément vers $u^\star$ au sens des fonctions continues de $\mathrm{B}(0,1/2)$. Or, par le théorème de la divergence, on a

$$\fint_{\mathrm{B}(0,\theta)}\nabla u^n(z)\mathrm{d}z=\frac{1}{|\mathrm{B}(0,\theta)|}\int_{S(0,\theta)}u^n d\vec{S},$$

où $S(x,\theta)$ désigne la sphère de centre $x$ et de rayon $\theta$. Ainsi, on a la convergence suivante :

$$\fint_{\mathrm{B}(0,\theta)}\nabla u^n(z)\mathrm{d}z\to\fint_{\mathrm{B}(0,\theta)}\nabla u^\star(z)\mathrm{d}z. \tag{2.135}$$

Donc, par passage à la limite dans (2.134) et en utilisant l'inégalité de Cauchy-Schwarz, on déduit que $u^\star$ satisfait l'inégalité suivante :

$$\begin{aligned}&\sup_{x\in\mathrm{B}(0,\theta)}\left|u^\star(x)-u^\star(0)-x\cdot\fint_{\mathrm{B}(0,\theta)}\nabla u^\star(z)\mathrm{d}z\right|\\&\geq 2^{-1}\theta^{1+\gamma}\left(\fint_{\mathrm{B}(0,1)}|u^\star(x)|^2\,\mathrm{d}x\right)^{1/2}\\&\geq 2^{-d/2-1}\theta^{1+\gamma}\left(\fint_{\mathrm{B}(0,1/2)}|u^\star(x)|^2\,\mathrm{d}x\right)^{1/2}.\end{aligned} \tag{2.136}$$

Les Inégalités (2.131) et (2.136) sont incompatibles avec (2.132). D'où l'existence de $\varepsilon_1$ indépendant de $y$ tel que (2.126) soit satisfaite pour tout $\varepsilon<\varepsilon_1$.

**Ajout du correcteur** Montrons maintenant (2.127). Soit $y > 0$, et $u^\varepsilon$ satisfaisant (2.124). Grâce à l'estimation (2.123), il existe $C, \beta > 0$ ne dépendant que de $\mu$ tels que

$$\|u^\varepsilon\|_{\mathrm{C}^{0,\beta}(\mathrm{B}(0,1/2))} \leq C \|u^\varepsilon\|_{\mathrm{L}^2(\mathrm{B}(0,1))}.$$

Ainsi :

$$\left| \fint_{\mathrm{B}(0,\theta)} \nabla u^\varepsilon(z) \mathrm{d}z \right| \leq \left| \frac{1}{\mathrm{B}(0,\theta)} \int_{S(0,\theta)} u^\varepsilon d\vec{S} \right| \leq C\theta^{-d} \left( \fint_{\mathrm{B}(0,1)} |u^\varepsilon(x)|^2 \, \mathrm{d}x \right)^{1/2}. \tag{2.137}$$

En invoquant le Lemme 2.2.1,

$$\sup_{y\in\mathbb{R}^d} \sup_{x\in\mathrm{B}(0,1)} \left| \varepsilon w_i \left( \frac{x}{\varepsilon} + y \right) - \varepsilon w_i(y) \right| \underset{\varepsilon\to 0}{\to} 0.$$

Donc il existe $\varepsilon_0 < \varepsilon_1$ indépendant de $y$ tel que (2.127) est vérifiée. □

On peut alors itérer le Lemme 2.4.5 précédent pour obtenir le Lemme suivant :

**Lemme 2.4.6** (Analogue du Lemme 15 de [11])**.** *Soit $A$ satisfaisant les Hypothèses 1, 3, 4, 5, et 6, et $\gamma \in ]0,1[$. Soient $\theta$ et $\varepsilon_0$ donnés par le Lemme 2.4.5. Il existe alors $C > 0$ ne dépendant que de $\theta$ telle que pour tout $y \in \mathbb{R}^d$, si $0 < \varepsilon \leq \varepsilon_0 \theta^n$, $n \in \mathbb{N}^\star$, et si $u^\varepsilon \in \mathrm{H}^1(\mathrm{B}(0,1))$ est solution de*

$$-\mathrm{div}\left( A\left( \frac{x}{\varepsilon} + y \right) \cdot \nabla u^\varepsilon(x) \right) = 0 \qquad \text{dans} \quad \mathrm{B}(0,1), \tag{2.138}$$

*alors il existe $\kappa_j(n)$, $j \in [\![1,d]\!]$ telle que*

$$\sup_{x\in\mathrm{B}(0,\theta^{n+1})} \left| u^\varepsilon(x) - u^\varepsilon(0) - \sum_{j=1}^d \left\{ x_j + \varepsilon \left( w_j \left( \frac{x}{\varepsilon} + y \right) - w_j(y) \right) \right\} \kappa_j(n) \right|$$
$$\leq \theta^{(1+n)(1+\gamma)} \|u^\varepsilon\|_{\mathrm{L}^\infty(\mathrm{B}(0,1))}, \tag{2.139}$$

$$\sup_{j\in[\![1,d]\!]} |\kappa_j(n)| \leq C \left( \sum_{j=0}^n \theta^{j\gamma} \right) \|u^\varepsilon\|_{\mathrm{L}^\infty(\mathrm{B}(0,1))}. \tag{2.140}$$

La preuve se fait par récurrence. A chaque itération, on utilise le Lemme 2.4.5 sur une échelle qui décroît géométriquement.

*Démonstration.* Si $n = 0$, on pose :

$$\kappa_j(0) = \fint_{\mathrm{B}(0,\theta)} \partial_j u^\varepsilon(z) \mathrm{d}z.$$

Le Lemme 2.4.5 donne alors (2.139). Par le théorème de la divergence (de manière analogue à (2.137)), on en déduit que $\kappa_j(0)$ satisfait (2.140).

On suppose ensuite que le Lemme 2.4.6 est vrai pour $n \geq 0$. Soit $0 < \varepsilon \leq \theta^{n+1}\varepsilon_0$ et $u^\varepsilon$ satisfaisant (2.138). On se donne $\kappa_j(n)$ associé à $u^\varepsilon$. On pose alors $\widetilde{\varepsilon} := \varepsilon\theta^{-n-1}$ et

$$v(z) = u^\varepsilon\left(\theta^{n+1}z\right) - u^\varepsilon(0) - \sum_{j=1}^{d}\left\{\theta^{n+1}z_j + \theta^{n+1}\widetilde{\varepsilon}\left(w_j\left(\frac{z}{\widetilde{\varepsilon}} + y\right) - w_j(y)\right)\right\}\kappa_j(n).$$

Par (2.3) et (2.138), alors $v$ est une solution de

$$-\mathrm{div}\left(A\left(\frac{z}{\widetilde{\varepsilon}} + y\right)\cdot\nabla v(z)\right) = 0 \qquad \text{dans} \quad \mathrm{B}(0,1).$$

De plus, par hypothèse, $\widetilde{\varepsilon} \leq \varepsilon_0$. Donc, grâce au Lemme 2.4.5,

$$\begin{aligned}&\sup_{z\in\mathrm{B}(0,\theta)}\left|v(z) - v(0) - \sum_{i=1}^{d}\left\{z_i + \widetilde{\varepsilon}\left(w_i\left(\frac{z}{\widetilde{\varepsilon}} + y\right) - w_i(y)\right)\right\}\fint_{\mathrm{B}(0,\theta)}\partial_i v\right| \\ &\leq \theta^{1+\gamma}\left\|v\right\|_{\mathrm{L}^\infty(\mathrm{B}(0,1))}. \end{aligned} \tag{2.141}$$

Or, par hypothèse de récurrence, on a

$$\begin{aligned}\left\|v\right\|_{\mathrm{L}^\infty(\mathrm{B}(0,1))} &= \sup_{x\in\mathrm{B}(0,\theta^{n+1})}\left|u^\varepsilon(x) - u^\varepsilon(0) - \sum_{j=1}^{d}\left\{x_j + \varepsilon\left(w_j\left(\frac{x}{\varepsilon} + y\right) - w_j(y)\right)\right\}\kappa_j(n)\right| \\ &\leq \theta^{(1+n)(1+\gamma)}\left\|u^\varepsilon\right\|_{\mathrm{L}^\infty(\mathrm{B}(0,1))}. \end{aligned} \tag{2.142}$$

Posons alors

$$\kappa_j(n+1) := \kappa_j(n) + \theta^{-n-1}\fint_{\mathrm{B}(0,\theta)}\partial_j v,$$

d'où

$$\begin{aligned}&v(z) - v(0) - \sum_{i=1}^{d}\left\{z_i + \widetilde{\varepsilon}\left(w_i\left(\frac{z}{\widetilde{\varepsilon}} + y\right) - w_i(y)\right)\right\}\fint_{\mathrm{B}(0,\theta)}\partial_i v \\ &= u^\varepsilon(\theta^{n+1}z) - u^\varepsilon(0) - \sum_{j=1}^{d}\theta^{n+1}\left\{z_j + \widetilde{\varepsilon}\left(w_j\left(\frac{z}{\widetilde{\varepsilon}} + y\right) - w_j(y)\right)\right\}\kappa_j(n+1). \end{aligned} \tag{2.143}$$

Remettons cela à l'échelle de $x$. Ensemble, (2.141), (2.142), (2.143) démontrent alors (2.139) pour l'étape $n+1$. Par ailleurs, grâce au théorème de la divergence et à (2.142),

$$\begin{aligned}|\kappa_j(n+1)| &\leq |\kappa_j(n)| + \theta^{-n-1}\left|\frac{1}{|\mathrm{B}(0,\theta)|}\int_{S(0,\theta)} v d\vec{S}\right| \\ &\leq |\kappa_j(n)| + C\theta^{-n-2}\left\|v\right\|_{\mathrm{L}^\infty(\mathrm{B}(0,1))} \\ &\leq |\kappa_j(n)| + C(\theta)\theta^{(1+n)\gamma}\left\|u^\varepsilon\right\|_{\mathrm{L}^\infty(\mathrm{B}(0,1))}, \end{aligned}$$

où $C(\theta)$ est une constante dépendant seulement de $\theta$, mais pas de $n$. Cela démontre donc (2.140) et conclut la preuve. □

Nous sommes maintenant en mesure de terminer la preuve du Théorème 2.1.3, en faisant l'étape de blow-up.

*Démonstration du Théorème 2.1.3.* On ne montre l'estimation (2.16) que pour $R = 2$. A partir du moment où elle est vraie pour $R = 2$, on l'obtient alors pour $R$ quelconque par simple remise à l'échelle. On pose $\gamma := 1/2$ on se donne les $\varepsilon_0$ et $\theta$ associés par le Lemme 2.4.5.

Supposons que $\varepsilon > \varepsilon_0$. Comme $A$ satisfait l'Hypothèse 2, grâce au Corollaire A.3.3, l'Estimation (2.16) est satisfaite avec une constante ne dépendant que de $A$ et $\varepsilon_0$.

Supposons donc maintenant qu'il existe $n \geq 0$ tel que :

$$\theta^{n+1}\varepsilon_0 \leq \varepsilon \leq \theta^n \varepsilon_0. \tag{2.144}$$

Grâce au Lemme 2.4.6 appliqué pour $\gamma = 1/2$, on déduit de (2.139) et (2.140) que

$$\begin{aligned}\sup_{|x|\leq\theta^{n+1}} |u^\varepsilon(x) - u^\varepsilon(0)| &\leq \theta^{\frac{3(n+1)}{2}} \|u^\varepsilon\|_{\mathrm{L}^\infty(\mathrm{B}(0,1))} \\ &+ C \sup_{|x|\leq\theta^{n+1}} \left\{ \theta^{n+1} + \varepsilon \max_{j\in[\![1,d]\!]} \left| w_j\left(\frac{x}{\varepsilon} + y\right) - w_j(y) \right| \right\} \|u^\varepsilon\|_{\mathrm{L}^\infty(\mathrm{B}(0,1))}. \end{aligned} \tag{2.145}$$

Grâce au Lemme 2.2.1, il existe une constante $C$ ne dépendant que de $A$ telle que, pour tout $j \in [\![1, d]\!]$, $y \in \mathbb{R}^d$,

$$\sup_{|x|\leq\theta^{n+1}} \varepsilon \left| w_j\left(\frac{x}{\varepsilon} + y\right) - w_j(y) \right| \leq C\theta^{n+1}. \tag{2.146}$$

De (2.144), (2.145) et (2.146), on obtient l'existence de $C$ dépendant de $A$, $\theta$ et $\varepsilon_0$ telle que

$$\sup_{|x|\leq\theta^{n+1}} |u^\varepsilon(x) - u^\varepsilon(0)| \leq C\theta^{n+1} \|u^\varepsilon\|_{\mathrm{L}^\infty(\mathrm{B}(0,1))}. \tag{2.147}$$

On pose ensuite

$$v(z) := u^\varepsilon\left(\theta^{n+1} z\right) - u^\varepsilon(0). \tag{2.148}$$

Alors, on obtient

$$-\mathrm{div}\left( A\left(\frac{\theta^{n+1}}{\varepsilon} z + y\right) \cdot \nabla v(z) \right) = 0 \quad \text{dans} \quad \mathrm{B}\left(0, \theta^{-n-1}\right).$$

Par (2.144), $\theta^{n+1}/\varepsilon \leq 1/\varepsilon_0$. Par conséquent, comme $A$ satisfait l'Hypothèse 2, on peut appliquer le Corollaire A.3.3 à $v$. Ainsi, il existe une constante $C$ dépendant seulement de $\varepsilon_0$ et $A$ telle que

$$|\nabla v(0)| \leq C \left( \fint_{\mathrm{B}(0,1)} |v|^2 \right)^{1/2}. \tag{2.149}$$

Or, par définition (2.148) de $v$, on déduit de (2.147) et de (2.149) que

$$|\nabla u^\varepsilon(0)| \leq C \|u^\varepsilon\|_{\mathrm{L}^\infty(\mathrm{B}(0,1))}, \tag{2.150}$$

où $C$ ne dépend que de $A$, $\theta$ et $\varepsilon_0$. Puis, en appliquant (2.122) à $u^\varepsilon$, on déduit

$$|\nabla u^\varepsilon(0)| \leq C \left\| u^\varepsilon \right\|_{\mathrm{L}^2(\mathrm{B}(0,2))}. \tag{2.151}$$

Clairement le point 0 ne joue pas de rôle particulier dans (2.151) et peut être remplacé par un point quelconque $x \in \mathrm{B}(0,1)$. Ainsi, on obtient le résultat voulu (2.16). □

### 2.4.6 Sous-linéarité des correcteurs adaptés

La démonstration d'estimations lipschitziennes jusqu'au bord requiert l'introduction de correcteurs adaptés (voir Section 2.2.5), dont nous établissons des propriétés de sous-linéarité renforcée.

En remarquant que la fonction $u^\varepsilon(x) := w_j^{\varepsilon,\Omega}(x) - w_j\left(\frac{x}{\varepsilon}\right)$ satisfait

$$-\operatorname{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla u^\varepsilon(x)\right) = 0 \qquad \text{dans} \quad \Omega, \quad \text{et} \quad u^\varepsilon(x) = -w_j(x/\varepsilon) \qquad \text{sur} \quad \partial\Omega,$$

on déduit du principe du maximum que

$$\left\| w_j^{\varepsilon,\Omega} \right\|_{\mathrm{L}^\infty(\Omega)} \leq \left\| w_j\left(\cdot/\varepsilon\right) \right\|_{\mathrm{L}^\infty(\partial\Omega)}.$$

Ainsi, grâce au Lemme 2.2.1 (quitte à retirer une constante, on suppose que $w_j(\cdot/\varepsilon)$ s'annule sur $\Omega$), cela implique l'existence d'une fonction $\Xi$ ne dépendant que de $A$, qui tend vers 0 en 0, et telle que

$$\left\| w_j^{\varepsilon,\Omega} \right\|_{\mathrm{L}^\infty(\Omega)} \leq \frac{\mathrm{Diam}(\Omega)}{\varepsilon}\Xi\left(\frac{\varepsilon}{\mathrm{Diam}(\Omega)}\right). \tag{2.152}$$

Puis, on démontre que si les correcteurs $w_j$ eux-même sont fortement sous-linéaires alors les correcteurs adaptés le sont aussi :

**Proposition 2.4.7.** *Supposons que $A$ satisfait les Hypothèses 1, 2, 3, 4, 5, 6 et 7, pour un certain $\nu$ donné. Soit $\Omega$ un ouvert régulier borné. Alors, les correcteurs adaptés $w_j^{\varepsilon,\Omega(0,2)}$ satisfont l'inégalité suivante :*

$$\varepsilon\left|w_j^{\varepsilon,\Omega(0,2)}(x) - w_j^{\varepsilon,\Omega(0,2)}(y)\right| \leq C\varepsilon^\nu\left|x-y\right|^{1-\nu}, \qquad \forall x,y \in \Omega(0,1), \tag{2.153}$$

*où $C$ ne dépend pas de $\varepsilon$.*

*Remarque* 21 (Architecture logique). Cette proposition repose sur le Théorème 2.1.2. Elle est ensuite employée pour démontrer le Théorème 2.1.4. Celui-ci nécessite de disposer de correcteurs qui soient sous-linéaires au bord, c'est à dire satisfaisant

$$\varepsilon\left|w_j^{\varepsilon,\Omega(0,2)}(x)\right| \leq C\mathrm{d}(x,\Gamma_\Omega(0,2)), \tag{2.154}$$

si $\mathrm{d}(x,\Gamma_\Omega(0,2)) > \varepsilon$. L'inégalité ci-dessus est effectivement impliquée par (2.153). Toutefois, nous ne savons pas si l'Hypothèse 7 est nécessaire pour obtenir (2.154).

*Démonstration de la Proposition 2.4.7.* Soit $j \in [\![1,d]\!]$. On pose $u^\varepsilon = w_j^{\varepsilon,\Omega(0,2)}(x) - w_j\left(x/\varepsilon\right)$. Par définition (voir (2.3) et (2.62)),

$$\begin{cases} -\operatorname{div}\left(A\left(\dfrac{x}{\varepsilon}\right)\cdot\nabla u^\varepsilon(x)\right) = 0 & \text{dans} \quad \Omega(0,2), \\ u^\varepsilon(x) = -w_j\left(x/\varepsilon\right) & \text{sur} \quad \partial\left(\Omega(0,2)\right). \end{cases}$$

On distingue deux cas : le cas $\nu = 1$ et le cas $\nu < 1$.

Dans tous les cas, le principe du maximum implique que

$$\|u^\varepsilon\|_{\mathrm{L}^\infty(\Omega(0,2))} \leq \|w_j\left(\cdot/\varepsilon\right)\|_{\mathrm{L}^\infty(\Omega(0,2))} \leq C\varepsilon^{\nu-1}.$$

Si $\nu = 1$, on en déduit (2.153) par inégalité triangulaire.

Si $\nu < 1$, grâce au Théorème 2.1.2, on a

$$\|u^\varepsilon\|_{\mathrm{C}^{0,1-\nu}(\Omega(0,1))} \leq C\|w_j\left(\cdot/\varepsilon\right)\|_{\mathrm{C}^{0,1-\nu}(\Omega(0,2))} + C\|u^\varepsilon\|_{\mathrm{L}^2(\Omega(0,2))} \leq C\varepsilon^{\nu-1}.$$

De même, on en déduit (2.153) par inégalité triangulaire. □

### 2.4.7 Estimations lipschitziennes jusqu'au bord

Nous démontrons maintenant le Théorème 2.1.4 qui généralise le Théorème 2.1.3.

*Remarque* 22. Comme elle est requise pour démontrer la Proposition 2.4.7, l'Hypothèse 7 est nécessaire à la démonstration du Théorème 2.1.4. Toutefois, nous ne savons pas si elle est nécessaire pour que la conclusion du Théorème soit valide (voir Remarque 21).

*Remarque* 23 (Architecture logique). La démonstration du Théorème 2.1.4 emploie notamment les Théorèmes 2.1.2 et 2.1.3, et les Propositions 2.1.1 et 2.4.7. Ce Théorème sert ensuite à établir les points (i') et (ii') du Théorème 2.1.5.

La démonstration du Théorème 2.1.4 suit la preuve classique de [11]. Elle est donc très semblable à la preuve du Théorème 2.1.3 ci-dessus ; toutefois la présence du bord requiert un certain nombre de subtilités techniques. Dans tout ce qui suit, quitte à prolonger $g$, on va supposer que $g \in \mathrm{C}^{1,\beta}(\mathrm{B}(0,1))$ satisfait

$$\|g\|_{\mathrm{C}^{1,\beta}(\mathrm{B}(0,1))} \leq C\|g\|_{\mathrm{C}^{1,\beta}(\Gamma_\Omega(0,1))}.$$

**Lemme 2.4.8** (Analogue du Lemme 18 de [11])**.** *Supposons que $A$ satisfait les Hypothèses 1, 3, 4, 5 et 6. Soit $K_0 > 0$ et $\beta \in ]0,1[$. Pour toute constante $\gamma \in ]0,\beta[$, il existe une constante $\theta \in ]0,1/4[$ ne dépendant que de $\mu$, $\beta$, $\gamma$ et $K_0$, et une constante $\varepsilon_0 > 0$ ne dépendant que de $A$, $\beta$, $\gamma$, $K_0$ et $\theta$ telles que, si on a les conditions suivantes : soit $\phi \in \mathrm{C}^{1,\beta}\left(\mathbb{R}^{d-1}\right)$ satisfaisant (2.92), $g \in \mathrm{C}^{1,\beta}(\mathrm{B}(0,1))$, $\varepsilon < \varepsilon_0$ et $u^\varepsilon \in \mathrm{H}^1\left(D_\phi(1)\right)$ satisfaisant*

$$\begin{cases} -\operatorname{div}\left(A\left(\dfrac{x}{\varepsilon}\right)\cdot\nabla u^\varepsilon(x)\right) = 0 & \textit{dans} \quad D_\phi(1), \\ u^\varepsilon = g & \textit{sur} \quad \Delta_\phi(1), \end{cases} \tag{2.155}$$

*pour $D_\phi$ et $\Delta_\phi$ définis par* (2.93)*, avec*

$$g(0) = 0, \quad \nabla g(0) \in \mathbb{R}e_d \tag{2.156}$$

$$\|g\|_{\mathrm{C}^{1,\beta}(\mathrm{B}(0,1))} \leq 1, \tag{2.157}$$

$$\|u^\varepsilon\|_{\mathrm{L}^2\left(D_\phi(1)\right)} \leq 1, \tag{2.158}$$

*alors, on a*

$$\sup_{x\in D_\phi(\theta)} \left| u^\varepsilon(x) - \left\{ x_d + \varepsilon w_d^{\varepsilon, D_\phi(1)}(x) \right\} \fint_{C_{K_0}(\theta)} \partial_d u^\varepsilon(z) \mathrm{d}z \right| \leq \theta^{1+\gamma}. \tag{2.159}$$

*Remarque* 24 (Orientation). Il est notable que, dans l'estimation (2.159), la seule orientation dont on a besoin pour approximer le gradient est celle correspondant à la normale de la surface du domaine considéré. Sur cet aspect, l'estimation (2.159) peut être comparée avec l'estimation (2.125), où aucune orientation préférentielle n'est déterminée. Ceci est dû à l'utilisation de conditions de Dirichlet au bord dans le premier cas, tandis qu'il n'y a pas de bord dans le second. Ce fait est très important d'un point de vue technique, car, lors de l'itération qui va suivre (voir Lemme 2.4.9), on va faire face à des conditions de Dirichlet non homogènes dues au terme en $x_d \fint \partial_d u^\varepsilon$. Mais on constate que la contribution de la fonction $x \mapsto x_d$ est petite au bord car celui-ci est orthogonal à $e_d$ en 0 (ce n'est pas le cas des fonctions $x \mapsto x_i$, pour $i \neq d$) ; ainsi, le reste induit par la correction au bord demeure contrôlé.

*Remarque* 25. En toute rigueur, il faudrait considérer $A(y + x/\varepsilon)$ en lieu et place de $A(x/\varepsilon)$ dans (2.155) et démontrer que les constantes du Lemme 2.4.8 sont bien indépendantes du $y$ ainsi introduit (voir Remarque 10), mais aussi de la normale au bord du domaine $\partial\Omega$ en 0 (qui se trouve être ici fixée et égale à $e_d$). Ces ajouts sont aisés mais alourdiraient les notations.

*Démonstration du Lemme 2.4.8.* La démonstration se fait par l'absurde, et s'articule en quatre étapes. Dans l'Etape 1, on démontre que la solution du problème homogénéisé correspondant à (2.155) satisfait une estimations plus forte que (2.159). Puis, dans l'Etape 2, on suppose par l'absurde que $u^\varepsilon$ ne satisfait pas (2.159). En utilisant la compacité induite par la Proposition 2.1.1, on en déduit dans l'Etape 3 une contradiction sur la solution du problème homogénéisé. Enfin, dans l'Etape 4, on démontre que l'on peut rajouter la contribution du terme avec le correcteur dans (2.159), celle-ci étant négligeable.

**Etape 1** Supposons que $u^\star$ satisfait

$$\begin{cases} -\operatorname{div}\left(A^\star \cdot \nabla u^\star(x)\right) = 0 \quad \text{dans} \quad D_{\phi^\star}(1/2), \\ u^\star = g^\star \quad \text{sur} \quad \Delta_{\phi^\star}(1/2), \end{cases} \tag{2.160}$$

où $\phi^\star$, $g^\star$, respectivement $u^\star$, satisfont (2.92), (2.156) et (2.157), et

$$\|u^\star\|_{\mathrm{L}^2\left(D_{\phi^\star}(1/2)\right)} \leq 1. \tag{2.161}$$

Par conséquent, on peut appliquer à $u^\star - g^\star$ le résultat [64, Cor. 8.36 p. 212], d'où

$$\|u^\star - g^\star\|_{\mathrm{C}^{1,\beta}\left(D_{\phi^\star}(1/4)\right)} \leq C \|g^\star\|_{\mathrm{C}^{1,\beta}\left(D_{\phi^\star}(1/3)\right)} + C \|u^\star - g^\star\|_{\mathrm{L}^\infty\left(D_{\phi^\star}(1/3)\right)},$$

puis le Lemme A.3.5 à $u^\star$. Ainsi, il existe une constante $C$ ne dépendant que de $K_0$, $\beta$ et $\mu$ telle que

$$\|u^\star - g^\star\|_{\mathrm{C}^{1,\beta}\left(D_{\phi^\star}(1/4)\right)} \leq C \|g^\star\|_{\mathrm{C}^{1,\beta}\left(D_{\phi^\star}(1/2)\right)} + C \|u^\star\|_{\mathrm{L}^2\left(D_{\phi^\star}(1/2)\right)}.$$

Par inégalité triangulaire, et en prenant en compte (2.157) et(2.161), on obtient

$$\|u^\star\|_{\mathrm{C}^{1,\beta}\left(D_{\phi^\star}(1/4)\right)} \leq C. \tag{2.162}$$

Par développement limité, on en déduit donc l'existence d'une constante $C$ ne dépendant que de $\mu$ et $K_0$ telle que

$$\sup_{x \in D_{\phi^\star}(\theta)} \left| u^\star(x) - u^\star(0) - \sum_{i=1}^{d} x_i \partial_i u^\star(0) \right| \leq C\theta^{1+\beta}.$$

En outre, par (2.92), (2.160) et (2.156) on sait que $u^\star(0) = 0$ et que $\nabla u^\star(0)$ est colinéaire à $e_d$. En approximant alors $\nabla u^\star(0)$ par une intégrale moyennée, on obtient

$$\sup_{x \in D_{\phi^\star}(\theta)} \left| u^\star(x) - x_d \fint_{C_{K_0}(\theta)} \partial_d u^\star(z) \mathrm{d}z \right| \leq C\theta^{1+\beta}.$$

Comme $\gamma < \beta$, on peut fixer $\theta \in ]0, 1/4[$ tel que

$$\sup_{x \in D_{\phi^\star}(\theta)} \left| u^\star(x) - x_d \fint_{C_{K_0}(\theta)} \partial_d u^\star(z) \mathrm{d}z \right| \leq \frac{1}{3}\theta^{1+\gamma}. \tag{2.163}$$

**Etape 2** Supposons par l'absurde qu'il existe $\varepsilon_n \to 0$, et $\phi^n \in \mathrm{C}^{1,\beta}\left(\mathbb{R}^{d-1}\right)$, $g^n \in \mathrm{C}^{1,\beta}(\mathrm{B}(0,1))$, et $u^n$ satisfaisant respectivement (2.92), (2.156) et (2.157), (2.155) et (2.158), telles que pour tout $n$, on a

$$\sup_{x \in D_{\phi^n}(\theta)} \left| u^n(x) - x_d \fint_{C_{K_0}(\theta)} \partial_d u^n(z) \mathrm{d}z \right| \geq \frac{\theta^{1+\gamma}}{2}. \tag{2.164}$$

On souhaite borner uniformément $u^n$ dans $\mathrm{H}^1(\mathrm{B}(0,1/2))$. On ne peut appliquer directement le Lemme A.3.2 à $u^n$, car il satisfait une condition de bord inhomogène. On scinde donc $u^n = u_1^n + u_2^n$, où

$$\begin{cases} -\operatorname{div}\left(A_n(x/\varepsilon_n) \cdot \nabla u_1^n(x)\right) = -\operatorname{div}\left(A_n(x/\varepsilon_n) \cdot \nabla u_2^n(x)\right) = 0 \quad \text{dans} \quad D_{\phi^n}(1), \\ u_1^n = 0 \quad \text{sur} \quad \Delta_{\phi^n}(1) \quad \text{et} \quad u_2^n = g^n \quad \text{sur} \quad \partial\left(D_{\phi^n}(1)\right). \end{cases}$$

Par [64, Cor. 8.7 p. 183], puis grâce à (2.157), on a

$$\|u_2^n\|_{\mathrm{H}^1\left(D_{\phi^n}(1)\right)} \leq C\,\|g\|_{\mathrm{H}^1(\mathrm{B}(0,1))} \leq C\,\|g\|_{\mathrm{C}^{1,\beta}(\mathrm{B}(0,1))} \leq C.$$

Par ailleurs, le Lemme A.3.2 implique, grâce à (2.158), qu'il existe une constante $C$ ne dépendant que de $\mu$ telle que

$$\|\nabla u_1^n\|_{\mathrm{L}^2\left(D_{\phi^n}(1/2)\right)} \leq C\,\|u_1^n\|_{\mathrm{L}^2\left(D_{\phi^n}(1)\right)} \leq C\,\|u_2^n\|_{\mathrm{L}^2\left(D_{\phi^n}(1)\right)} + C\,\|u^n\|_{\mathrm{L}^2\left(D_{\phi^n}(1)\right)}.$$

Ainsi, on en déduit par inégalité triangulaire que

$$\|\nabla u^n\|_{\mathrm{L}^2\left(D_{\phi^n}(1/2)\right)} \leq C, \tag{2.165}$$

où $C$ est indépendante de $n$. Vu les conditions au bord de (2.155), il est possible de prolonger $u^n$ par $g^n$ sur $\mathrm{B}(0,1)$ (par abus de notation, ce prolongement sera aussi noté $u^n$), ce prolongement étant dans $\mathrm{H}^1(\mathrm{B}(0,1/2))$. On déduit donc naturellement de (2.165) et de (2.157) que

$$\|\nabla u^n\|_{\mathrm{L}^2(\mathrm{B}(0,1/2))} \leq C.$$

On obtient alors par compacité l'existence de $u^\star$, $g^\star$ et $\phi^\star$ telles que les convergences suivantes soient satisfaites à extraction près :

$$g^n \to g^\star \quad \text{dans} \quad \mathrm{C}^0(\mathrm{B}(0,1)), \quad \text{par équicontinuité de } (g^n), \tag{2.166}$$

$$\phi^n \to \phi^\star \quad \text{dans} \quad \mathrm{C}^0\left(\mathbb{R}^{d-1}\right), \quad \text{par équicontinuité de } (\phi^n), \tag{2.167}$$

$$u^n \rightharpoonup u^\star \quad \text{dans} \quad \mathrm{L}^2(\Omega(0,1)), \tag{2.168}$$

$$\nabla u^n \rightharpoonup \nabla u^\star \quad \text{dans} \quad \mathrm{L}^2(\Omega(0,1/2)). \tag{2.169}$$

Par le Théorème 2.1.2, on obtient de (2.92), (2.155), (2.157), et (2.158) que $(u^n)$ est équicontinue sur $\mathrm{B}(0,1/2)$ (elle est même équi-hölderienne pour tout exposant). Ainsi, quitte à extraire, on a la convergence suivante :

$$u^n \to u^\star \quad \text{dans} \quad \mathrm{C}^0(\mathrm{B}(0,1/2)). \tag{2.170}$$

Or, par le théorème de la divergence, on a l'identité suivante :

$$⨍_{C_{K_0}(\theta)} \partial_d u^n(z)\mathrm{d}z = \frac{1}{|C_{K_0}(\theta)|}\int_{\partial C_{K_0}(\theta)} u^n(z) e_d \cdot \mathrm{d}\vec{S} \tag{2.171}$$

Ainsi, on déduit de (2.164) et (2.170) que $u^\star$ satisfait

$$\sup_{x\in D_{\phi^\star}(\theta)} \left| u^\star(x) - \sum_{i=1}^{d} x_i ⨍_{C_{K_0}(\theta)} \partial_i u^\star(z)\mathrm{d}z \right| \geq \frac{\theta^{1+\gamma}}{2}. \tag{2.172}$$

**Etape 3** Grâce à la Proposition 2.1.1, $u^\star$ satisfait l'équation (2.160). En outre, par les propriétés de la convergence faible, $u^\star$ satisfait aussi (2.161). Naturellement, $\phi^\star$ et $g^\star$ satisfont (2.92), respectivement (2.156) et (2.157). Par conséquent, l'inégalité (2.172) est en contradiction avec l'estimation (2.163). D'où notre hypothèse de l'Etape 2 était absurde. Ainsi, sous les hypothèses du Lemme 2.4.8, il existe $\varepsilon_1 > 0$ tel que, pour tout $\varepsilon < \varepsilon_1$, on a

$$\sup_{x\in D_\phi(\theta)} \left| u^\varepsilon(x) - x_d \fint_{C_{K_0}(\theta)} \partial_d u^\varepsilon(z)\mathrm{d}z \right| \leq \frac{1}{2}\theta^{1+\gamma}. \tag{2.173}$$

**Etape 4** En appliquant le théorème de la divergence, on obtient

$$\left| \fint_{C_{K_0}(\theta)} \partial_d u^\varepsilon(z)\mathrm{d}z \right| = C \left| \int_{\partial C_{K_0}(\theta)} u^\varepsilon(z)\mathrm{d}S \right|.$$

Alors, grâce au Lemme A.3.4, à (2.157) et à (2.158),

$$\left| \fint_{C_{K_0}(\theta)} \partial_d u^\varepsilon(z)\mathrm{d}z \right| \leq C \left( \int_{D_\phi(1)} |u^\varepsilon(z)|^2 \,\mathrm{d}z \right)^{1/2} + C \left\| g \right\|_{\mathrm{L}^\infty(\mathrm{B}(0,1))} \leq C.$$

L'inégalité (2.152) implique alors l'existence d'un $\varepsilon_0 \leq \varepsilon_1$ tel que

$$\sup_{x\in D_\phi(\theta)} \left| \varepsilon w_d^{\varepsilon, D_\phi(1)}(x) \fint_{C_{K_0}(\theta)} \partial_d u^\varepsilon(z)\mathrm{d}z \right| \leq \frac{1}{2}\theta^{1+\gamma}. \tag{2.174}$$

**Conclusion** L'estimation (2.159) découle de (2.173) et (2.174). □

Nous pouvons maintenant énoncer le lemme d'itération :

**Lemme 2.4.9** (Analogue du Lemme 19 de [11])**.** *Soit $A$ satisfaisant les Hypothèses 1, 3, 4, 5 et 6. Soient $\beta \in ]0,1[$, $\gamma \in ]0, \beta/2[$ et $K_0 > 0$. Soient $\theta$ et $\varepsilon_0$ donnés par le Lemme 2.4.8. Il existe alors une constante $C > 0$ ne dépendant que de $A$, $\theta$ et $K_0$ (quitte à prendre $K_0$ plus petit) telle que : si $0 < \varepsilon \leq \varepsilon_0 \theta^n$, si $\phi$, $g$, et $u^\varepsilon$ satisfont respectivement* (2.92)*,* (2.156)*, et* (2.155)*, il existe une suite $\xi_k$ telle que*

$$\begin{aligned} \sup_{x\in D_\phi(\theta^{n+1})} & \left| u^\varepsilon(x) - \sum_{k=0}^{n} \theta^{\gamma k} \left\{ x_d + \varepsilon w_d^{\theta^{-k}\varepsilon, D_\phi(\theta^k)} \left( \varepsilon y + \theta^{-k} x \right) \right\} \xi_k \right| \\ & \leq \theta^{(1+n)(1+\gamma)} \left( \left\| u^\varepsilon \right\|_{\mathrm{L}^\infty(D_\phi(1))} + \left\| g \right\|_{\mathrm{C}^{1,\beta}(\mathrm{B}(0,1))} \right), \end{aligned} \tag{2.175}$$

$$|\xi_k| \leq C \left\| u^\varepsilon \right\|_{\mathrm{L}^\infty(D_\phi(1))}. \tag{2.176}$$

Comme pour le Lemme 2.4.6, la preuve se fait par récurrence, de la même façon que la preuve originale de [11]. Notons qu'un point important de la démonstration, par rapport à celle du Lemme 2.4.6, est la présence d'une donnée au bord $\widetilde{g}$ issue des problèmes rencontrés en itérant les remises à l'échelle (car $x \mapsto x_d$ n'est pas nul au bord, en général). Il est

crucial que $\widetilde{g}$ soit aussi petite que possible pour la mise à l'échelle se fasse correctement. Ceci motive d'imposer (2.156), de prendre des correcteurs adaptés, et de n'utiliser que la direction orthogonale au bord $e_d$ dans le calcul de l'approximation du gradient, dans l'estimation (2.159) du Lemme 2.4.8 (voir la Remarque 24).

*Démonstration.* Quitte à renormaliser, on suppose que

$$\|g\|_{\mathrm{C}^{0,\beta}(\mathrm{B}(0,1))} \leq \varepsilon_0/2 \quad \text{et} \quad \|u^\varepsilon\|_{\mathrm{L}^\infty\left(D_\phi(1)\right)} \leq 1. \tag{2.177}$$

Grâce au Lemme 2.4.8, (2.175) est satisfaite pour $n = 0$ avec

$$\xi_0 := \fint_{C_{K_0}(\theta)} \partial_d u^\varepsilon(z)\mathrm{d}z,$$

qui, par le théorème de la divergence, satisfait (2.176).

Supposons à présent que le Lemme 2.4.9 est vrai pour $n \geq 0$, et soit $0 < \varepsilon \leq \varepsilon_0\theta^{n+1}$. On pose alors $\widetilde{\varepsilon} := \varepsilon\theta^{-n-1}$, $z = \theta^{-n-1}x$ et

$$v(z) = \theta^{(-n-1)(1+\gamma)}\left(u^\varepsilon(\theta^{n+1}z) - \sum_{k=0}^{n}\theta^{\gamma k}\left\{\theta^{n+1}z_d + \varepsilon w_d^{\theta^{-k}\varepsilon, D_\phi\left(\theta^k\right)}\left(\theta^{n+1-k}z\right)\right\}\xi_k\right).$$

Par hypothèse de récurrence

$$\|v\|_{\mathrm{L}^\infty\left(\theta^{-n-1}D_\phi(\theta^{n+1})\right)} \leq 1. \tag{2.178}$$

En outre,

$$\begin{cases} -\operatorname{div}\left(A\left(\dfrac{x}{\widetilde{\varepsilon}}\right)\cdot\nabla v(z)\right) = 0 & \text{dans} \quad \theta^{-n-1}D_\phi\left(\theta^{n+1}\right), \\ v(z) = \widetilde{g}(z) & \text{sur} \quad \theta^{-n-1}\Delta_\phi\left(\theta^{n+1}\right), \end{cases}$$

avec

$$\widetilde{g}(z) := \theta^{(-n-1)(1+\gamma)}g(\theta^{n+1}z) - \theta^{(-n-1)(1+\gamma)}\sum_{k=0}^{n}\theta^{\gamma k+n+1}z_d\xi_k =: \widetilde{g}_1 + \widetilde{g}_2.$$

Grâce à (2.156) et (2.177), on déduit par remise à l'échelle et en utilisant le fait que $\gamma < \beta/2$ que

$$\|\widetilde{g}_1\|_{\mathrm{C}^{1,\beta}\left(\theta^{-n-1}D_\phi(\theta^{n+1})\right)} \leq \theta^{(\beta-\gamma)(n+1)}\varepsilon_0/2 \leq \varepsilon_0/2, \tag{2.179}$$

Par définition,

$$\|\widetilde{g}_2\|_{\mathrm{C}^{1,\beta}\left(\theta^{-n-1}D_\phi(\theta^{n+1})\right)} \leq \theta^{-(n+1)\gamma}\frac{1}{1-\theta^\gamma}\sup_{x,y\in\theta^{-n-1}\Delta_\phi(\theta^{n+1})}\frac{|x_d - y_d|}{|x-y|^{1+\beta}}.$$

De plus, $e_d$ est le vecteur normal à la surface $\Delta_\phi\left(\theta^{n+1}\right)$ en 0. Donc, si on note $\bar{x} := (x_1, ..., x_{d-1})$, on obtient

$$\begin{aligned}\sup_{x,y\in\theta^{-n-1}\Delta_\phi(\theta^{n+1})}\frac{|x_d-y_d|}{|x-y|^{1+\beta}} &\leq \sup_{x,y\in \mathrm{B}(0,\theta)}\frac{\theta^{-n-1}\left|\phi\left(\theta^{n+1}\bar{x}\right)-\phi\left(\theta^{n+1}\bar{y}\right)\right|}{|x-y|^{1+\beta}}\\ &\leq \theta^{-n-1}\left\|\phi\right\|_{\mathrm{C}^{1,\beta}}\sup_{x,y\in\mathrm{B}(0,\theta)}\frac{\left|\theta^{n+1}\left(\bar{x}-\bar{y}\right)\right|^{1+\beta}}{|x-y|^{1+\beta}}\\ &\leq K_0\theta^{\beta(n+1)}.\end{aligned}$$

Ainsi, comme $\beta > 2\gamma$,

$$\left\|\widetilde{g}_2\right\|_{\mathrm{C}^{1,\beta}\left(\theta^{-n-1}D_\phi(\theta^{n+1})\right)} \leq \frac{K_0}{1-\theta^\gamma}\theta^{\gamma(n+1)}.$$

Quitte à faire une dilatation des variables initiales, on peut supposer que $K_0$ est suffisamment petit pour que $K_0\theta^\gamma/(1-\theta^\gamma) \leq \varepsilon_0/2$. Par conséquent

$$\left\|\widetilde{g}\right\|_{\mathrm{C}^{1,\beta}\left(\theta^{-n-1}D_\phi(\theta^{n+1})\right)} \leq \varepsilon_0,$$

ce qui permet d'appliquer le Lemme 2.4.8, d'où

$$\sup_{z\in\theta^{-n-1}D_\phi(\theta^{n+2})}\left|v(z)-\left\{z_d+\theta^{-n-1}\varepsilon w_d^{\theta^{-n-1}\varepsilon,\theta^{-n-1}D_\phi\left(\theta^{n+1}\right)}(z)\right\}\right.$$
$$\left.\fint_{\theta^{-n-1}C_{K_0}(\theta^{n+2})}\partial_d v\right| \leq \theta^{1+\gamma}.$$

La formule précédente, réécrite avec $u^\varepsilon$ et $x$, donne

$$\begin{aligned}\sup_{x\in D_\phi(\theta^{n+2})}\Bigg|u^\varepsilon(x)-\sum_{k=0}^{n}\theta^{\gamma k}\left\{x_d+\varepsilon w_d^{\theta^{-k}\varepsilon,D_\phi(\theta^k)}\left(\theta^{-k}x\right)\right\}\xi_k\\ -\theta^{\gamma(n+1)}\left\{x_d+\varepsilon w_d^{\theta^{-n-1}\varepsilon,D_\phi(\theta^{n+1})}\left(\theta^{-n-1}x\right)\right\}\\ \left.\fint_{\theta^{-n-1}C_{K_0}(\theta^{n+2})}\partial_d v\right| \leq \theta^{(1+\gamma)(n+2)}.\end{aligned}$$

On pose alors

$$\xi_{n+1} := \fint_{\theta^{-n-1}C_{K_0}(\theta^{n+2})}\partial_d v,$$

qui, par le théorème de la divergence, puis grâce à (2.178), satisfait

$$|\xi_{n+1}| \leq C\left\|v\right\|_{\mathrm{L}^\infty\left(\theta^{-n-1}D_\phi(\theta^{n+1})\right)} \leq C.$$

Ceci conclut la démonstration de (2.175) et (2.176) pour $n+1$, d'où la preuve par récurrence du Lemme 2.4.9. □

Nous pouvons à présent effectuer la :

*Démonstration du Théorème 2.1.4.* La démonstration du Théorème 2.1.4 se fait en deux étapes. Tout d'abord, dans l'Etape 1, on démontre une propriété de sous-linéarité au bord. C'est à dire que, si $l \leq n$, on a

$$\sup_{x \in \Omega\left(0,\theta^l\right)} |u^\varepsilon(x) - u^\varepsilon(0)| \leq C\theta^l. \tag{2.180}$$

Puis, dans l'Etape 2, on cherche à majorer le gradient $\nabla u^\varepsilon(x)$ en distinguant deux cas :

1. soit $x$ est loin du bord (situation $x_1$ sur la Figure 2.3), à l'échelle $\varepsilon/\varepsilon_0$. A ce moment là, on utilise le Lemme 2.4.9, qui implique que $|u^\varepsilon(x)| \lesssim \mathrm{d}(x, \partial\Omega)$. En utilisant le Théorème 2.1.3, on obtient alors la majoration souhaitée sur $\nabla u^\varepsilon$ (car on peut dessiner autour du point $x$ un boule suffisamment grande).
2. soit $x$ est proche du bord (situation $x_2$ sur la Figure 2.3), à l'échelle $\varepsilon/\varepsilon_0$. Alors, on fait un blow-up au niveau de $x$, et on utilise la régularité classique, qui permet de majorer $\nabla u^\varepsilon$.

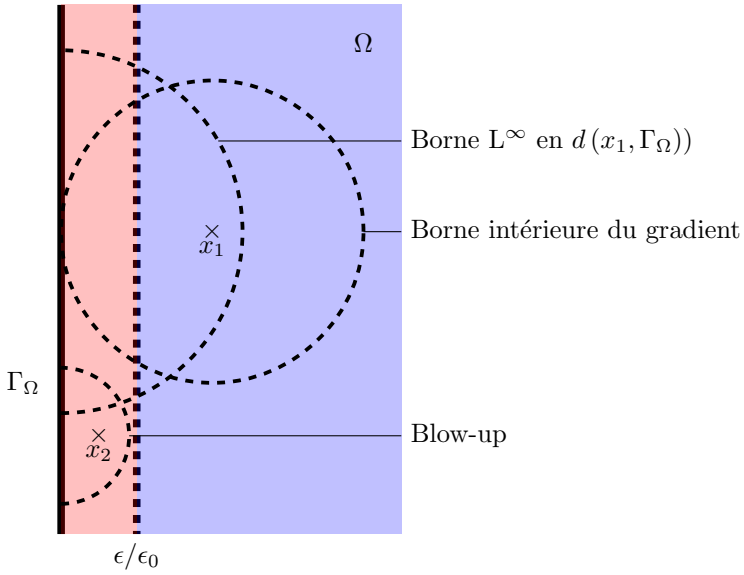



FIGURE 2.3 – La stratégie de démonstration du Théorème 2.1.4

Pour simplifier la démonstration, on suppose que $0 \in \Gamma_\Omega(0,1)$ et $y = 0$. On suppose en outre que

$$\|u^\varepsilon\|_{\mathrm{L}^2(\Omega(0,2))} \leq 1 \quad \text{et} \quad \|g\|_{\mathrm{C}^{1,\beta}(\Omega(0,2))} \leq 1. \tag{2.181}$$

Tout d'abord, grâce au Lemme A.3.4, on peut estimer $u^\varepsilon$ sur $\Omega(0,1)$

$$\|u^\varepsilon\|_{\mathrm{L}^\infty(\Omega(0,1))} \leq C\,\|u^\varepsilon\|_{\mathrm{L}^2(\Omega(0,2))} + C\,\|g\|_{\mathrm{L}^\infty(\mathrm{B}(0,2))} \leq C. \tag{2.182}$$

Puis, on se donne $\theta$, $\gamma$, $\varepsilon_0$ tels que dans le Lemme 2.4.9. On peut donc encadrer

$$\theta^{n+1}\varepsilon_0 < \varepsilon \leq \theta^n \varepsilon_0.$$

**Etape 1** Démontrons (2.180). Pour cela, on définit

$$v(x) := u^\varepsilon(x) - g(0) - \sum_{j=1}^{d} \partial_j g(0) \left( x_j + \varepsilon w_j^{\varepsilon,\Omega(0,2)}(x) \right). \tag{2.183}$$

Ainsi,

$$\begin{cases} -\operatorname{div}\left( A\left(\frac{x}{\varepsilon}\right) \cdot \nabla v(x) \right) = 0 & \text{dans} \quad \Omega(0,2), \\ v(x) = \widetilde{g}(x) & \text{sur} \quad \Gamma_\Omega(0,2), \end{cases} \tag{2.184}$$

où

$$\widetilde{g}(x) := g(x) - g(0) - \nabla g(0) \cdot x.$$

On déduit de (2.152), (2.181), et (2.182), que

$$\|v\|_{\mathrm{L}^\infty(\Omega(0,1))} \leq C. \tag{2.185}$$

Par ailleurs,

$$\|\widetilde{g}\|_{\mathrm{C}^{1,\beta}(\mathrm{B}(0,1))} \leq C \|g\|_{\mathrm{C}^{1,\beta}(\mathrm{B}(0,1))} \leq C. \tag{2.186}$$

Remarquons aussi que $\widetilde{g}(0) = 0$ et $\nabla\widetilde{g}(0) = 0$. Si $l \leq n$, on applique le Lemme 2.4.9 à $v$, d'où

$$\begin{aligned} &\sup_{x\in\Omega(0,\theta^l)} \left| v(x) - \sum_{k=0}^{l-1} \theta^{\gamma k} \left\{ x_d + \varepsilon w_d^{\theta^{-k}\varepsilon,\Omega(0,\theta^k)} \left(\theta^{-k}x\right) \right\} \xi_k \right| \\ &\leq \theta^{l(1+\gamma)} \left( \|v\|_{\mathrm{L}^\infty(\Omega(0,1))} + \|\widetilde{g}\|_{\mathrm{C}^{1,\beta}(\mathrm{B}(0,1))} \right), \end{aligned} \tag{2.187}$$

où $|\xi_k| \leq C \|v\|_{\mathrm{L}^\infty(\Omega(0,1))}$. De plus, grâce à la Proposition 2.4.7 et à (2.185),

$$\begin{aligned} &\sup_{x\in\Omega(0,\theta^l)} \left| \sum_{k=0}^{l-1} \theta^{\gamma k} \left\{ x_d + \varepsilon w_d^{\theta^{-k}\varepsilon,\Omega(0,\theta^k)} \left(\theta^{-k}x\right) \right\} \xi_k \right| \\ &\leq C \sum_{k=0}^{l-1} \theta^{\gamma k} \left( \theta^l + \varepsilon + \varepsilon^{1-\nu}\theta^{\nu l} \right) \|v\|_{\mathrm{L}^\infty(\Omega(0,1))} \\ &\leq C\theta^l. \end{aligned} \tag{2.188}$$

Ainsi, il découle de (2.185), (2.186), (2.187) et (2.188), que

$$\sup_{x\in\Omega\left(0,\theta^l\right)} |v(x)| \leq C\theta^l. \tag{2.189}$$

Puis, comme $u(0) = g(0)$ et que $g \in \mathrm{C}^{1,\beta}(\Omega(0,2))$, ceci établit donc (2.180), grâce à la Proposition 2.4.7.

**Etape 2** Soit maintenant $x \in \Omega(0,1/2)$ ; quitte à re-paramétrer, on peut prendre $x = x_d e_d$, avec $x_d > 0$ et $0 \in \partial\Omega$. Alors, on a l'alternative suivante : soit $x$ est loin du bord à l'échelle $\varepsilon/\varepsilon_0$, c'est à dire qu'il existe $l \in [\![0, n-1]\!]$ tel que

$$\theta^{l+1} \leq x_d \leq \theta^l, \tag{2.190}$$

soit $x$ est proche du bord à l'échelle $\varepsilon/\varepsilon_0$, c'est à dire que

$$0 \leq x_d \leq \theta^n. \tag{2.191}$$

On se place tout d'abord dans le cas (2.190). On applique alors le Théorème 2.1.3 à $u^\varepsilon$ sur $\mathrm{B} := \mathrm{B}\left(x, \mathrm{d}\left(x, \Gamma_\Omega(0,1/2)\right)\right)$ et on déduit de (2.180) que

$$|\nabla u^\varepsilon(x)| \leq C\theta^{-l} \sup_{x\in\mathrm{B}} |u^\varepsilon(y) - u^\varepsilon(0)| \leq C, \tag{2.192}$$

Au contraire, dans le cas (2.191), on considère

$$v(z) := u^\varepsilon(\theta^n z) - u^\varepsilon(0),$$

qui satisfait $-\mathrm{div}\left(A\left(z/(\varepsilon\theta^{-n})\right)\nabla v(z)\right) = 0$ dans $\theta^{-n}\Omega(0,\theta^n)$. Par un résultat classique [64, Cor. 8.36 p. 212], on obtient que

$$\|\nabla v\|_{\mathrm{L}^\infty(\theta^{-n}\Omega(0,\theta^{n+1}))} \leq C\,\|v\|_{\mathrm{L}^\infty(\theta^{-n}\Omega(0,\theta^n))} + C\,\|g\left(\theta^n\cdot\right)\|_{\mathrm{C}^{1,\beta}(\mathrm{B}(0,1))}.$$

En remettant alors à l'échelle 1, et grâce à (2.189),

$$\begin{aligned}\|\nabla u^\varepsilon\|_{\mathrm{L}^\infty(\Omega(0,\theta^{n+1}))} \leq & C\theta^{-n-1}\,\|u^\varepsilon - u^\varepsilon(0)\|_{\mathrm{L}^\infty(\Omega(0,\theta^n))} + C\,\|g\|_{\mathrm{C}^{1,\beta}(\mathrm{B}(0,1))} \\ \leq & C.\end{aligned} \tag{2.193}$$

Les estimations (2.192) et (2.193) permettent alors de conclure que, dans tous les cas,

$$\|\nabla u^\varepsilon\|_{\mathrm{L}^\infty(\Omega(0,1/2))} \leq C. \tag{2.194}$$

En recouvrant ensuite $\Omega(0,1)$ de boules plus petites, on arrive au résultat désiré. □

## 2.5 Estimations dans le cas inhomogène

Les preuves de cette Section sont des adaptations des arguments de [94]. La différence majeure vis-à-vis de cette référence concerne le cas où on ne fait des hypothèses que relativement à la matrice $A$ et non à la matrice $A^T$. En effet, les correcteurs de la matrice $A$ et $A^T$ ont des propriétés potentiellement différentes (voir le contre-exemple de la Section 2.2.7). C'est pourquoi nous avons tenu à utiliser le moins possible les correcteurs de la matrice $A^T$.

### 2.5.1 Estimations sur la fonction de Green $G^\varepsilon$

Dans cette section, nous discutons et démontrons le Théorème 2.1.5.

Soit un domaine régulier borné $\Omega \subset \mathbb{R}^d$. Rappelons tout d'abord (voir [72, Th. 1.1 et Th. 1.3]) que, si $A$ satisfait l'Hypothèse 1, alors l'équation suivante :

$$-\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla_x G^\varepsilon(x,y)\right) = \delta_y(x) \tag{2.195}$$

possède une unique solution $G^\varepsilon$ telle que, pour tout $R>0$,

$$G^\varepsilon(\cdot,y) \in \mathrm{H}^1\left(\Omega\backslash \mathrm{B}(y,R)\right) \cap \mathrm{W}_0^{1,1}(\Omega). \tag{2.196}$$

La fonction $G^\varepsilon$ est appelée fonction de Green de Dirichlet sur $\Omega$ associée à l'opérateur $-\mathrm{div}\left(A\left(\frac{\cdot}{\varepsilon}\right)\cdot\nabla\right)$. Elle satisfait (2.20) et

$$\left\|\nabla_x G^\varepsilon(\cdot,y)\right\|_{L^{\frac{d}{d-1},\infty}(\Omega)} \leq C \qquad \forall y\in\Omega, \tag{2.197}$$

$$\left\|\nabla_y G^\varepsilon(x,\cdot)\right\|_{L^{\frac{d}{d-1},\infty}(\Omega)} \leq C \qquad \forall x\in\Omega, \tag{2.198}$$

où $C$ dépend seulement de $\mu$. Ici $\mathrm{L}^{p,\infty}(\Omega)$ désigne l'espace de Marcinkiewicz équipé de la semi-norme (voir [21])

$$\|f\|_{\mathrm{L}^{p,\infty}} := \sup_{t>0} t\left|\{x\in\Omega, |f(x)|>t\}\right|^{1/p},$$

pour $p>1$.

*Remarque* 26. Dans le cadre de notre démonstration, il devient nécessaire de supposer que $A^T$ vérifie les Hypothèses 3, 4, 5 et 6 pour démontrer les inégalités (2.22) et (2.23). En effet, le gradient $\nabla_y G$ est la solution d'une équation elliptique qui fait intervenir $A^T$ et non $A$.

*Remarque* 27 (Architecture logique). La démonstration du Théorème 2.1.5 repose sur les Théorèmes 2.1.3 et 2.1.4. Les estimations sur les fonctions de Green et leurs dérivées du Théorème 2.1.5 seront utiles par la suite, et serviront à démontrer la Proposition 2.1.7, les Lemmes 2.6.1, et les Théorèmes 2.1.9, 2.1.12, 1.1.3, et 1.1.4.

La démonstration du Théorème 2.1.5 repose sur le constat suivant : les fonctions $x\mapsto G^\varepsilon(x,y)$, et $x\mapsto \nabla_y G^\varepsilon(x,y)$ sont toutes deux solutions d'une équation elliptique du type

$$-\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla u^\varepsilon(x)\right) = 0$$

sur tout ouvert ne contenant pas $y$. En appliquant le Théorème 2.1.3 sur des boules bien choisies, on peut alors majorer $\nabla_x G^\varepsilon(x,y)$ grâce à l'estimation établie sur $G^\varepsilon(x,y)$, et on estime de même $\nabla_x\nabla_y G^\varepsilon(x,y)$.

*Démonstration du Théorème 2.1.5.* Nous supposons tout d'abord que $A$ ne satisfait que les Hypothèses 1, 2, 3, 4, 5, et 6.

**Preuve de (i)** Nous démontrons tout d'abord que $G^\varepsilon$ vérifie bien (2.21). Soit $x_0 \in \Omega_1$, $y_0 \in \Omega$, $x_0 \neq y_0$. On pose

$$R := \frac{\min\left(\mathrm{d}(x_0, \partial\Omega), |x_0 - y_0|\right)}{2}. \tag{2.199}$$

Comme $\mathrm{d}(x_0, \partial\Omega) \geq \mathrm{d}(\Omega_1, \partial\Omega)$ et que $|x_0 - y_0| \leq \mathrm{Diam}(\Omega)$, alors il existe $C$ ne dépendant que de $\Omega$ et $\Omega_1$ telle que

$$|x_0 - y_0| \leq CR. \tag{2.200}$$

Par définition

$$-\mathrm{div}_x\left(A\left(\frac{x}{\varepsilon}\right) \cdot \nabla_x G^\varepsilon\left(x, y_0\right)\right) = 0 \quad \text{dans} \quad \mathrm{B}\left(x_0, R\right). \tag{2.201}$$

On peut alors appliquer le Théorème 2.1.3 à $x \mapsto G^\varepsilon(x, y_0)$, d'où

$$\left|\nabla_x G^\varepsilon\left(x_0, y_0\right)\right| \leq \frac{C}{R}\left(\fint_{\mathrm{B}(x_0,R)} \left|G^\varepsilon(x, y_0)\right|^2 \mathrm{d}x\right)^{1/2}.$$

Par (2.20) puis (2.200), on obtient alors

$$\left(\fint_{\mathrm{B}(x_0,R)} \left|G^\varepsilon(x, y_0)\right|^2 \mathrm{d}x\right)^{1/2} \leq C\left(\fint_{\mathrm{B}(x_0,R)} \left(\frac{1}{R}\right)^{2(d-2)} \mathrm{d}x\right)^{1/2} \leq \frac{C}{|x_0 - y_0|^{d-2}},$$

ce qui donne

$$\left|\nabla_x G^\varepsilon\left(x_0, y_0\right)\right| \leq C|x_0 - y_0|^{1-d},$$

c'est à dire (2.21).

**Preuve de (i')** Supposons que $A$ satisfait aussi l'Hypothèse 7 pour un certain $\nu \in ]0, 1]$. La fonction $G^\varepsilon(\cdot, y_0)$ satisfait (2.201) et la condition au bord $G^\varepsilon(x, y_0) = 0$ si $x \in \partial\Omega$. Par conséquent, grâce au Théorème 2.1.4, et plus précisément, en remettant (2.19) à l'échelle sur des domaines $\Omega(x_0, R)$, avec $R = |x_0 - y_0|/2$, on démontre que l'Estimation (2.21) est valide pour tous $x, y \in \Omega$.

**Preuve de (ii)** Supposons maintenant que $A^T$ satisfait aussi les Hypothèses 3, 4, 5 et 6. Par [72, Th. 1.3], $G^\varepsilon(x, y) = G_T^\varepsilon(y, x)$, où $G_T^\varepsilon$ est la fonction de Green de Dirichlet de l'opérateur $-\mathrm{div}\left(A^T\left(\frac{\cdot}{\varepsilon}\right) \cdot \nabla\right)$ sur $\Omega$. Donc en utilisant (2.21) sur $G_T^\varepsilon(y, x)$, on obtient (2.22).

Soient $x_0 \neq y_0 \in \Omega_1$, et $R$ défini par (2.199) (on a encore (2.200)). En dérivant (2.201) par rapport à $y_0$, on obtient

$$-\mathrm{div}_x\left(A\left(\frac{x}{\varepsilon}\right) \cdot \nabla_x \nabla_y G^\varepsilon\left(x, y_0\right)\right) = 0 \quad \text{dans} \quad \mathrm{B}\left(x_0, R\right).$$

Ainsi, par le Théorème 2.1.3,

$$\left|\nabla_x \nabla_y G^\varepsilon\left(x_0, y_0\right)\right| \leq CR^{-1}\left(\fint_{\mathrm{B}(x_0,R)} \left|\nabla_y G^\varepsilon(x, y_0)\right|^2 \mathrm{d}x\right)^{1/2}.$$

En utilisant (2.22), on en déduit alors (2.23).

**Preuve de (ii')** Supposons enfin qu'il existe $\nu > 0$ tel que $A$ et $A^T$ satisfont aussi l'Hypothèse 7 pour un certain $\nu \in ]0,1]$. Alors on peut utiliser le Théorème 2.1.4 sur $G^\varepsilon$ et $\nabla_y G^\varepsilon$, car $G^\varepsilon(x,y) = 0$ si $y \in \partial\Omega$, et $\nabla_y G^\varepsilon(x,y) = 0$ si $x \in \partial\Omega$. Par conséquent, on peut remplacer dans la preuve ci-dessus les boules par des ensembles $\Omega(x_0, R)$ (touchant le bord $\partial\Omega$ mais évitant cependant les singularités $x = y_0$). On établit ainsi les estimations voulues (2.22) et (2.23) pour tous $x, y \in \Omega$. □

### 2.5.2 Estimation sur $u^\varepsilon$ dans $\mathrm{W}^{1,p}$

Dans cette section, nous discutons et démontrons la Proposition 2.1.6.

*Remarque* 28 (Architecture logique). La Proposition 2.1.6 repose sur le Théorème 2.1.3 et un lemme de mesure (le Lemme A.3.9 en Annexe). Soulignons que la démonstration de la Proposition 2.1.6 ne fait pas usage des estimations obtenues sur la fonction de Green $G^\varepsilon$ et ses dérivées (c'est à dire le Théorème 2.1.5). Elle est ensuite utilisée pour démontrer le Théorème 2.1.11.

*Démonstration de la Proposition 2.1.6.* On se donne $\mathrm{B}_0$ une boule incluse dans 2B telle que $8\mathrm{B}_0 \subset 2\mathrm{B}$. On veut appliquer le Lemme A.3.9 à $\nabla u^\varepsilon$ dans $\mathrm{B}_0$. Pour simplifier les notations, on suppose que $y = 0$.

Soit $\widetilde{\mathrm{B}} := \mathrm{B}(y_0, \widetilde{R}) \subset 2\mathrm{B}_0$. Posons

$$u^\varepsilon = u_1^\varepsilon + u_2^\varepsilon,$$

où $u_1^\varepsilon$ satisfait

$$\begin{cases} -\operatorname{div}\left(A\left(\dfrac{x}{\varepsilon}\right)\cdot\nabla u_1^\varepsilon(x)\right) = \operatorname{div}(H(x)) & \text{dans} \quad \mathrm{B}\left(y_0, 4\widetilde{R}\right), \\ u_1^\varepsilon(x) = 0 & \text{sur} \quad S\left(y_0, 4\widetilde{R}\right). \end{cases} \tag{2.202}$$

En testant simplement (2.202) contre $u_1^\varepsilon$, on obtient

$$\left(\fint_{4\widetilde{\mathrm{B}}} |\nabla u_1^\varepsilon|^2\right)^{1/2} \leq C\left(\fint_{4\widetilde{\mathrm{B}}} |H|^2\right)^{1/2}. \tag{2.203}$$

Par ailleurs, on a

$$-\operatorname{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla u_2^\varepsilon(x)\right) = 0 \quad \text{dans} \quad \mathrm{B}\left(y_0, 4\widetilde{R}\right).$$

Donc, en appliquant le Théorème 2.1.3 à $u_2^\varepsilon$ puis l'inégalité de Poincaré-Wirtinger, on déduit

$$\|\nabla u_2^\varepsilon\|_{\mathrm{L}^\infty(2\widetilde{\mathrm{B}})} \leq \frac{C}{\widetilde{R}}\left(\fint_{4\widetilde{\mathrm{B}}} |u_2^\varepsilon - \fint_{4\widetilde{\mathrm{B}}} u_2^\varepsilon|^2\right)^{1/2} \leq C\left(\fint_{4\widetilde{\mathrm{B}}} |\nabla u_2^\varepsilon|^2\right)^{1/2}.$$

Or, par inégalité triangulaire et en utilisant (2.203),

$$\begin{aligned}\left(\fint_{4\widetilde{\mathrm{B}}}|\nabla u_2^\varepsilon|^2\right)^{1/2} &\leq \left(\fint_{4\widetilde{\mathrm{B}}}|\nabla u_1^\varepsilon|^2\right)^{1/2}+\left(\fint_{4\widetilde{\mathrm{B}}}|\nabla u^\varepsilon|^2\right)^{1/2}\\ &\leq \left(\fint_{4\widetilde{\mathrm{B}}}|\nabla u^\varepsilon|^2\right)^{1/2}+\left(\fint_{4\widetilde{\mathrm{B}}}|H|^2\right)^{1/2}.\end{aligned}$$

D'où

$$\|\nabla u_2^\varepsilon\|_{\mathrm{L}^\infty(2\widetilde{\mathrm{B}})} \leq C\left(\fint_{4\widetilde{\mathrm{B}}}|\nabla u^\varepsilon|^2\right)^{1/2}+C\left(\fint_{4\widetilde{\mathrm{B}}}|H|^2\right)^{1/2}. \tag{2.204}$$

Grâce à (2.203) et (2.204), on peut appliquer le Lemme A.3.9, d'où l'existence d'une constante $C$ ne dépendant pas de $\mathrm{B}_0$ telle que

$$\left(\fint_{\mathrm{B}_0}|\nabla u^\varepsilon|^p\right)^{1/p} \leq C\left\{\left(\fint_{4\mathrm{B}_0}|H|^p\right)^{1/p}+\left(\fint_{4\mathrm{B}_0}|\nabla u^\varepsilon|^2\right)^{1/2}\right\}. \tag{2.205}$$

On se ramène à B grâce à un recouvrement fini par des boules $\mathrm{B}_0$, ce qui conclut la démonstration. □

### 2.5.3 Estimations lipschitziennes intérieures

Dans cette section, nous discutons et démontrons la Proposition 2.1.7, qui permet d'obtenir des estimations lipschitziennes.

*Remarque* 29 (Architecture logique). La Proposition 2.1.7 repose sur le Théorème 2.1.5. Elle est ensuite utilisée pour démontrer les Théorèmes 2.1.12 et 2.1.11.

*Remarque* 30. Un point technique de la preuve ci-dessous est qu'on ne fait pas d'hypothèse sur $A^T$ dans l'énoncé de la Proposition 2.1.7 ; par conséquent, on s'interdit d'utiliser les estimations (2.22) et (2.23).

La démonstration de la Proposition 2.1.7 reprend des ingrédients de la démonstration du Lemme 3.5 de [94].

*Démonstration de la Proposition 2.1.7.* La démonstration se fait en deux étapes : dans l'Etape 1, on cherche à exprimer $u^\varepsilon$ grâce à la fonction de Green $G^\varepsilon$. On règle la question des données au bord grâce à un cut-off. Puis, dans l'Etape 2, on tire parti des estimations sur la fonction de Green démontrées dans le Théorème 2.1.5 pour borner uniformément $\nabla u^\varepsilon$.

Pour simplifier la preuve, on suppose que $y = 0$ et que $u^\varepsilon$ est à moyenne nulle (quitte à rajouter une constante). Ainsi, par l'inégalité de Poincaré-Wirtinger, il existe une constante $C$ ne dépendant que de la dimension telle que

$$\fint_{2\mathrm{B}}|u^\varepsilon|^2 \leq CR^2\fint_{2\mathrm{B}}|\nabla u^\varepsilon|^2. \tag{2.206}$$

**Etape 1** On note $A_\varepsilon(x) := A\left(\frac{x}{\varepsilon}\right)$. On choisit une fonction de cut-off $\phi \in \mathrm{C}_\mathrm{c}^\infty(3/2\mathrm{B})$ à valeur dans $[0,1]$ satisfaisant

$$\phi = 1 \quad \text{dans} \quad \mathrm{B}, \qquad \|\nabla\phi\|_{\mathrm{L}^\infty(2\mathrm{B})} \leq CR^{-1} \quad \text{et} \quad \left\|\nabla^2\phi\right\|_{\mathrm{L}^\infty(2\mathrm{B})} \leq CR^{-2}. \tag{2.207}$$

Remarquons que $\|\nabla(\phi u^\varepsilon)\|_{\mathrm{L}^\infty(\mathrm{B})} = \|\nabla u^\varepsilon\|_{\mathrm{L}^\infty(\mathrm{B})}$. Par ailleurs, grâce à (2.24),

$$-\mathrm{div}\left(A_\varepsilon \cdot \nabla(\phi u^\varepsilon)\right) = -\mathrm{div}\left(u^\varepsilon A_\varepsilon \cdot \nabla\phi\right) - A_\varepsilon \cdot \nabla u^\varepsilon \cdot \nabla\phi - \phi\mathrm{div}(H).$$

On utilise la représentation par fonction de Green de Dirichlet de $-\mathrm{div}\left(A_\varepsilon \cdot \nabla\right)$ sur 2B, que l'on note $G^\varepsilon$, et une intégration par parties pour obtenir

$$\begin{aligned}
\phi(x)u^\varepsilon(x) =& -\int_{2\mathrm{B}} G^\varepsilon(x,y)A_\varepsilon(y)\cdot\nabla u^\varepsilon(y)\cdot\nabla\phi(y)\mathrm{d}y \\
&+\int_{2\mathrm{B}} \nabla_y G^\varepsilon(x,y).\left(u^\varepsilon(y)A_\varepsilon(y)\cdot\nabla\phi(y)\right)\mathrm{d}y \\
&+\int_{2\mathrm{B}} \nabla_y\left(G^\varepsilon(x,y)\phi(y)\right)\cdot H(y)\mathrm{d}y \\
=:& u_1^\varepsilon(x) + u_2^\varepsilon(x) + u_3^\varepsilon(x).
\end{aligned} \tag{2.208}$$

Nous allons montrer séparément que $u_1^\varepsilon$, $u_2^\varepsilon$ et $u_3^\varepsilon$ sont lipschitziennes sur B.

**Etape 2** Soit $x \in \mathrm{B}$. Comme $A_\varepsilon \cdot \nabla u^\varepsilon \cdot \nabla\phi$ est nulle sur B, on peut dériver $u_1^\varepsilon$. Grâce à (2.21) et (2.207), on obtient alors

$$\begin{aligned}
|\nabla u_1^\varepsilon(x)| &\leq \left|\int_{2\mathrm{B}\setminus\mathrm{B}} \nabla_x G^\varepsilon(x,y)\left(A_\varepsilon(y)\cdot\nabla u^\varepsilon(y)\cdot\nabla\phi(y)\right)\mathrm{d}y\right| \\
&\leq \frac{C}{R^d}\left|\int_{2\mathrm{B}\setminus\mathrm{B}} |\nabla u^\varepsilon|\right| \\
&\leq C\left(\fint_{2\mathrm{B}} |\nabla u^\varepsilon|^2\right)^{1/2}.
\end{aligned} \tag{2.209}$$

De même comme $A_\varepsilon \cdot \nabla u^\varepsilon \cdot \nabla\phi$ est nulle sur B et hors de $3/2\mathrm{B}$, on peut dériver $u_2^\varepsilon$ et

$$|\nabla u_2^\varepsilon(x)| \leq \left|\int_{3/2\mathrm{B}\setminus\mathrm{B}} \nabla_x\nabla_y G^\varepsilon(x,y)\left(u^\varepsilon(y)A_\varepsilon(y)\cdot\nabla\phi(y)\right)\mathrm{d}y\right|.$$

En utilisant l'inégalité de Cauchy-Schwarz, on obtient

$$|\nabla u_2^\varepsilon(x)| \leq CR^{-1}\left(\int_{3/2\mathrm{B}\setminus\mathrm{B}} |\nabla_x\nabla_y G^\varepsilon(x,y)|^2\,\mathrm{d}y\right)^{1/2}\left(\int_{3/2\mathrm{B}} |u^\varepsilon(y)|^2\,\mathrm{d}y\right)^{1/2}. \tag{2.210}$$

Alors, l'inégalité (2.23) permettrait de conclure immédiatement. Mais elle n'est pas nécessaire. En effet, grâce à l'inégalité de Cacciopoli, puis grâce à (2.21),

$$\left(\int_{3/2\mathrm{B}\setminus\mathrm{B}} |\nabla_x\nabla_y G^\varepsilon(x,y)|^2\,\mathrm{d}y\right)^{1/2} \leq CR^{-1}\left(\int_{7/4\mathrm{B}\setminus 1/4\mathrm{B}} |\nabla_x G^\varepsilon(x,y)|^2\,\mathrm{d}y\right)^{1/2} \leq CR^{-d}.$$

Ainsi, on déduit de (2.210) et (2.206) que

$$|\nabla u_2^\varepsilon(x)| \leq C\left(\fint_{2\mathrm{B}} |\nabla u^\varepsilon|^2\right)^{1/2}. \tag{2.211}$$

Montrer que $u_3^\varepsilon$ est lipschitzienne se révèle plus délicat. Mais $u_3^\varepsilon$ est aussi le terme le plus significatif de (2.208) (on peut s'en convaincre en prenant $\phi = 1$, ce qui supprime les termes $u_1^\varepsilon$ et $u_2^\varepsilon$ dus au bord). Remarquons tout d'abord que, par le théorème de la divergence, comme $\phi = 0$ sur $S(0,2R)$,

$$\int_{2\mathrm{B}} \nabla_y\left(G^\varepsilon(x,y)\phi(y)\right)\mathrm{d}y = 0, \qquad \forall x \in \mathrm{B}. \tag{2.212}$$

Ainsi, on a

$$u_3^\varepsilon(x) = \int_{2\mathrm{B}} \nabla_y\left(G^\varepsilon(x,y)\phi(y)\right)\cdot(H(y)-H(x))\mathrm{d}y. \tag{2.213}$$

En dérivant (2.213), et grâce à l'égalité (2.212) précédente, on obtient alors

$$\begin{aligned}\nabla u_3^\varepsilon(x) &= \int_{2\mathrm{B}} \nabla_y\left(\nabla_x G^\varepsilon(x,y)\phi(y)\right)\cdot(H(y)-H(x))\mathrm{d}y - \int_{2\mathrm{B}} \nabla_y\left(G^\varepsilon(x,y)\phi(y)\right)\cdot\nabla H(x)\mathrm{d}y\\ &= \int_{2\mathrm{B}} \nabla_y\left(\nabla_x G^\varepsilon(x,y)\phi(y)\right)\cdot(H(y)-H(x))\mathrm{d}y.\end{aligned}$$

D'où

$$|\nabla u_3^\varepsilon(x)| \leq \int_{2\mathrm{B}} |\phi(y)\nabla_y\nabla_x G^\varepsilon(x,y) + \nabla\phi(y)\nabla_x G^\varepsilon(x,y)|\,|H(y)-H(x)|\,\mathrm{d}y.$$

Grâce à (2.21),

$$\begin{aligned}\int_{2\mathrm{B}} |\nabla\phi(y)\nabla_x G^\varepsilon(x,y)|\,|H(y)-H(x)|\,\mathrm{d}y &\leq CR^{-1}\,\|H\|_{\mathrm{L}^\infty(2\mathrm{B})}\int_{3/2\mathrm{B}\setminus\mathrm{B}} \frac{1}{|x-y|^{d-1}}\mathrm{d}y\\ &\leq C\,\|H\|_{\mathrm{L}^\infty(2\mathrm{B})}.\end{aligned}$$

En outre, comme $H$ est bornée et $\beta$-hölderienne, alors

$$\begin{aligned}&\int_{2\mathrm{B}} |\phi(y)\nabla_y\nabla_x G^\varepsilon(x,y)|\,|H(y)-H(x)|\,\mathrm{d}y\\ &\leq \|H\|_{\dot{C}^{0,\beta}(2\mathrm{B})}\int_{\mathrm{B}(x,\varepsilon)} |\nabla_y\nabla_x G^\varepsilon(x,y)|\,|x-y|^\beta\mathrm{d}y\\ &\quad + 2\,\|H\|_{\mathrm{L}^\infty(2\mathrm{B})}\int_{3/2\mathrm{B}\setminus\mathrm{B}(x,\varepsilon)} |\nabla_y\nabla_x G^\varepsilon(x,y)|\,\mathrm{d}y.\end{aligned} \tag{2.214}$$

Si on supposait que $A^T$ satisfaisait aussi les Hypothèses 3, 4, 5 et 6, alors il suffirait d'utiliser (2.23) pour obtenir l'estimation voulue sur $\nabla u_3^\varepsilon(x)$, à savoir :

$$\|\nabla u_3^\varepsilon\|_{\mathrm{L}^\infty(\mathrm{B})} \leq C \ln\left(1 + R\varepsilon^{-1}\right) \|H\|_{\mathrm{L}^\infty(2\mathrm{B})} + C\varepsilon^\beta \|H\|_{\dot{\mathrm{C}}^{0,\beta}(2\mathrm{B})}. \tag{2.215}$$

De (2.209), (2.211) et (2.215), on obtient le résultat désiré (2.26). □

Toutefois, une estimation telle que (2.23) n'est pas nécessaire pour conclure. Comme expliqué ci-dessous, on peut malgré tout démontrer (2.215) sans rien supposer sur $A^T$. On emploie pour cela l'inégalité de Cacciopoli sur une décomposition annulaire (cette preuve est largement inspirée de [28]).

*Démonstration alternative de* (2.215). On note la couronne

$$\mathcal{C}(R_1, R_2) := \mathrm{B}(x, R_2) \backslash \mathrm{B}(x, R_1). \tag{2.216}$$

On a alors, par les inégalités de Cauchy-Schwarz puis de Cacciopoli, et enfin grâce à (2.21),

$$\begin{aligned}
\int_{\mathcal{C}(2^{-j}\delta, 2^{-j+1}\delta)} |\nabla_y \nabla_x G^\varepsilon(x,y)| \,\mathrm{d}y &\leq C\left(2^{-j}\delta\right)^{d/2} \left(\int_{\mathcal{C}(2^{-j}\delta, 2^{-j+1}\delta)} |\nabla_y \nabla_x G^\varepsilon(x,y)|^2 \,\mathrm{d}y\right)^{1/2} \\
&\leq C\left(2^{-j}\delta\right)^{d/2-1} \left(\int_{\mathcal{C}(2^{-j-1}\delta, 2^{-j+2}\delta)} |\nabla_x G^\varepsilon(x,y)|^2 \,\mathrm{d}y\right)^{1/2} \\
&\leq C\left(2^{-j}\delta\right)^{d/2-1} \left(\int_{\mathcal{C}(2^{-j-1}\delta, 2^{-j+2}\delta)} |x-y|^{-2(d-1)} \mathrm{d}y\right)^{1/2} \\
&\leq C.
\end{aligned}$$

Ainsi

$$\begin{aligned}
\int_{\mathrm{B}(x,\delta)} |\nabla_y \nabla_x G^\varepsilon(x,y)| \,|x-y|^\beta \mathrm{d}y &= \sum_{j=1}^{+\infty} \int_{\mathcal{C}(2^{-j}\delta, 2^{-j+1}\delta)} |\nabla_y \nabla_x G^\varepsilon(x,y)| \,|x-y|^\beta \mathrm{d}y \\
&\leq C\delta^\beta \sum_{j=1}^{+\infty} 2^{-j\beta} \leq C(\beta)\delta^\beta
\end{aligned}$$

et

$$\int_{3/2\mathrm{B}\backslash\mathrm{B}(x,\delta)} |\nabla_y \nabla_x G^\varepsilon(x,y)| \,\mathrm{d}y \leq \sum_{j=0}^{\frac{\log(2R/\delta)}{\log(2)}} \int_{\mathcal{C}(2^j\delta, 2^{j+1}\delta)} |\nabla_y \nabla_x G^\varepsilon(x,y)| \,\mathrm{d}y \leq C\frac{1+\log(2R/\delta)}{\log(2)}.$$

Par conséquent, on déduit (2.215) de (2.214). □

## 2.6 Approximations

Le but principal de la section est démontrer que $u^\varepsilon$ et son gradient $\nabla u^\varepsilon$ peuvent être approximés finement dans différentes normes $\mathrm{L}^p$, ou, de manière équivalente, que $R^\varepsilon$ et $\nabla R^\varepsilon$ peuvent être bornés (c'est à dire le Théorème 2.1.11). Au passage, nous approximons la fonction de Green $G^\varepsilon$, ses gradients $\nabla_x G^\varepsilon$ et $\nabla_y G^\varepsilon$, et son gradient croisé $\nabla_x \nabla_y G^\varepsilon$ (c'est à dire les Théorèmes 2.1.9 et 2.1.12).

*Remarque* 31 (Architecture logique). Les arguments essentiels sur lesquels reposent ces résultats sont les suivants :

1. $R^\varepsilon$ satisfait (2.32), avec $H^\varepsilon$ définie par (2.33) (c'est à dire la Proposition 2.1.8),
2. de ce fait, les différentes estimations démontrées dans les Sections 2.4 et 2.5 permettent d'estimer $R^\varepsilon$ à partir de $H^\varepsilon$, soit directement via les Propositions 2.1.6 et 2.1.7, soit grâce aux estimations sur la fonction de Green et ses dérivées données par le Théorème 2.1.5,
3. grâce aux Hypothèses 7 et 8, on peut contrôler $H^\varepsilon$ (voir (2.78) et (2.79)).

### 2.6.1 Approximation de $G^\varepsilon$

En utilisant la stratégie de la Section 3.1 de [94], nous démontrons maintenant une estimation point par point sur $G^\varepsilon - G^\star$ sur tout le domaine (y compris près du bord), à savoir le Théorème 2.1.9. Ici, $G^\star$ est la fonction de Green de l'opérateur $-\mathrm{div}(A^\star \cdot \nabla)$ avec conditions de Dirichlet sur un ouvert borné régulier $\Omega$.

*Remarque* 32. L'estimation (2.37) est homogène : le gain de précision en $\varepsilon$ est compensé par une perte d'intégrabilité en $x = y$. Elle correspond à l'estimation du Théorème 3.3 de [94] dans le cas particulier où $A$ est périodique (alors $\nu = 1$).

*Remarque* 33 (Architecture logique). La démonstration du Théorème 2.1.9 repose sur les éléments donnés dans la Remarque 31. Le Théorème 2.1.9 sert à la démonstration du Corollaire 2.1.10, du Théorème 2.1.12 et du Théorème 1.1.4.

Pour démontrer le Théorème 2.1.9, on utilise un Lemme de De Giorgi-Nash-Moser (Lemme A.3.4) jusqu'au bord du domaine. Puis, on majore $\|u^\varepsilon - u^\star\|_{\mathrm{L}^\infty}$ en fonction de $u^\star$ (voir le Lemme 2.6.1 ci-dessous). Par un argument de dualité (voir le Lemme 2.6.2 ci-dessous, qui estime $\|u^\varepsilon - u^\star\|_{\mathrm{L}^\infty}$ en fonction de $f \in \mathrm{L}^p$), on majore $\|G^\varepsilon(x,\cdot) - G^\star(x,\cdot)\|_{\mathrm{L}^{p'}}$. Puis, en remarquant que $G^\star$ joue le même rôle vis à vis de $G^\varepsilon$ que $u^\star$ vis-à-vis de $u^\varepsilon$, on en déduit une majoration point par point de la différence $G^\varepsilon - G^\star$ par la même technique que celle qui aura conduit à estimer $\|u^\varepsilon - u^\star\|_{\mathrm{L}^\infty}$.

**Lemme 2.6.1** (Extension du Lemme 3.2 de [94]). *Soit $A$ satisfaisant les Hypothèses 1, 2, 3, 7 et 8. Soit $\Omega$ un domaine borné régulier de classe $\mathrm{C}^{1,\beta}$, pour $\beta > 0$. Soit $x_0 \in \bar{\Omega}$, $R > 0$, $p > d$ et $q \in ]1, \infty[$. Supposons que $u^\varepsilon \in \mathrm{H}^1(\Omega(x_0, 4R))$ et $u^\star \in \mathrm{W}^{2,p}(\Omega(x_0, 4R))$ satisfont*

$$\begin{cases} -\mathrm{div}\left(A\left(\dfrac{x}{\varepsilon}\right) \cdot \nabla u^\varepsilon(x)\right) = -\mathrm{div}\left(A^\star \cdot \nabla u^\star(x)\right) & \text{dans} \quad \Omega(x_0, 4R), \\ u^\varepsilon = u^\star & \text{sur} \quad \Gamma_\Omega(x_0, 4R). \end{cases}$$

*Alors il existe $C$ ne dépendant que de $A$, $\Omega$, $q$ et $p$ telle que*

$$\begin{aligned}\|u^\varepsilon - u^\star\|_{\mathrm{L}^\infty(\Omega(x_0,R))} \leq & CR^{-d/q}\|u^\varepsilon - u^\star\|_{\mathrm{L}^q(\Omega(x_0,4R))} + C\varepsilon^\nu R^{1-\nu}\|\nabla u^\star\|_{\mathrm{L}^\infty(\Omega(x_0,4R))} \\ & + C\varepsilon^\nu R^{2-\frac{d}{p}-\nu}\left\|\nabla^2 u^\star\right\|_{\mathrm{L}^p(\Omega(x_0,4R))}. \end{aligned} \tag{2.217}$$

La preuve du Lemme 2.6.1 repose sur le Théorème de De Giorgi-Nash-Moser, sur le Théorème 2.1.5, et sur la Proposition 2.1.8 couplée à l'estimation (2.78).

*Démonstration.* Pour simplifier la démonstration, on suppose que $x_0 = 0$, que les correcteurs $w_j$ satisfont tous $w_j(0) = 0$ et que le potentiel $B$ donné par l'Hypothèse 8 satisfait $B(0) = 0$. Soit tout d'abord $\widetilde{\Omega}$ un domaine régulier de classe $\mathrm{C}^{1,\beta}$ tel que $\Omega(0,2R) \subset \widetilde{\Omega} \subset \Omega(0,4R)$.

On se donne $R^\varepsilon$ défini par (2.5). Alors, par la Proposition 2.1.8,

$$-\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla R^\varepsilon(x)\right) = \mathrm{div}\left(H^\varepsilon(x)\right) \quad \text{dans} \quad \Omega(0,4R), \tag{2.218}$$

où $H^\varepsilon$ est défini par (2.33). On scinde alors $R^\varepsilon = R_1^\varepsilon + R_2^\varepsilon$, où $R_1^\varepsilon$ satisfait

$$\begin{cases} -\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla R_1^\varepsilon(x)\right) = \mathrm{div}\left(H^\varepsilon(x)\right) & \text{dans} \quad \widetilde{\Omega}, \\ R_1^\varepsilon = 0 & \text{sur} \quad \partial\widetilde{\Omega}, \end{cases}$$

ce qui implique que $R_2^\varepsilon$ est solution de

$$\begin{cases} -\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla R_2^\varepsilon(x)\right) = 0 & \text{dans} \quad \widetilde{\Omega}, \\ R_2^\varepsilon(x) = -\varepsilon\displaystyle\sum_{j=1}^d w_j\left(\frac{x}{\varepsilon}\right)\partial_j u^\star(x) & \text{sur} \quad \partial\Omega\cap\partial\widetilde{\Omega}. \end{cases}$$

Par le Lemme A.3.4,

$$\|R_2^\varepsilon\|_{\mathrm{L}^\infty(\Omega(0,R))} \leq C\left\|\varepsilon\sum_{j=1}^d w_j\left(\frac{\cdot}{\varepsilon}\right)\partial_j u^\star\right\|_{\mathrm{L}^\infty(\widetilde{\Omega})} + \frac{C}{R^{d/q}}\|R_2^\varepsilon\|_{\mathrm{L}^q(\widetilde{\Omega})},$$

puis grâce à l'Hypothèse 7 et à une inégalité triangulaire,

$$\|R_2^\varepsilon\|_{\mathrm{L}^\infty(\Omega(0,R))} \leq C\varepsilon^\nu R^{1-\nu}\|\nabla u^\star\|_{\mathrm{L}^\infty(\widetilde{\Omega})} + \frac{C}{R^{d/q}}\|R^\varepsilon\|_{\mathrm{L}^q(\widetilde{\Omega})} + C\|R_1^\varepsilon\|_{\mathrm{L}^\infty(\widetilde{\Omega})}. \tag{2.219}$$

Or, par définition de $R^\varepsilon$ (2.5) et grâce à l'Hypothèse 7,

$$\begin{aligned}\|R^\varepsilon\|_{\mathrm{L}^q(\widetilde{\Omega})} &\leq \|u^\varepsilon - u^\star\|_{\mathrm{L}^q(\widetilde{\Omega})} + CR^{d/q}\left\|\sum_{j=1}^d \varepsilon w_j\left(\frac{\cdot}{\varepsilon}\right)\partial_j u^\star(\cdot)\right\|_{\mathrm{L}^\infty(\widetilde{\Omega})} \\ &\leq \|u^\varepsilon - u^\star\|_{\mathrm{L}^q(\widetilde{\Omega})} + C\varepsilon^\nu R^{d/q+1-\nu}\|\nabla u^\star\|_{\mathrm{L}^\infty(\widetilde{\Omega})}. \end{aligned} \tag{2.220}$$

On déduit de (2.219) et de (2.220) que

$$\|R_2^\varepsilon\|_{\mathrm{L}^\infty(\Omega(0,R))} \leq C\varepsilon^\nu R^{1-\nu} \|\nabla u^\star\|_{\mathrm{L}^\infty(\widetilde{\Omega})} + \frac{C}{R^{d/q}} \|u^\varepsilon - u^\star\|_{\mathrm{L}^q(\widetilde{\Omega})} + C \|R_1^\varepsilon\|_{\mathrm{L}^\infty(\widetilde{\Omega})}. \quad (2.221)$$

En posant $G^\varepsilon$ la fonction de Green de Dirichlet de $-\mathrm{div}\left(A\left(\frac{\cdot}{\varepsilon}\right)\cdot\nabla\right)$ sur $\widetilde{\Omega}$, on exprime explicitement $R_1^\varepsilon$

$$R_1^\varepsilon(x) = \int_{\widetilde{\Omega}} G^\varepsilon(x,y)\mathrm{div}\left(H^\varepsilon(y)\right)\mathrm{d}y = -\int_{\widetilde{\Omega}} \nabla_y G^\varepsilon(x,y)\cdot H^\varepsilon(y)\mathrm{d}y.$$

Ainsi, par l'inégalité de Hölder,

$$|R_1^\varepsilon(x)| \leq \|\nabla_y G^\varepsilon(x,\cdot)\|_{\mathrm{L}^{p'}(\widetilde{\Omega})} \|H^\varepsilon\|_{\mathrm{L}^p(\widetilde{\Omega})},$$

puis, grâce à l'Estimation (2.78), on obtient

$$|R_1^\varepsilon(x)| \leq C\varepsilon^\nu R^{1-\nu} \|\nabla_y G^\varepsilon(x,\cdot)\|_{\mathrm{L}^{p'}(\widetilde{\Omega})} \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^p(\widetilde{\Omega})}.$$

Comme $p > d$, alors $p' < \frac{d}{d-1}$. Donc, grâce au Lemme A.3.8 et au Théorème 2.1.5,

$$\|\nabla_y G^\varepsilon(x,\cdot)\|_{\mathrm{L}^{p'}(\widetilde{\Omega})} \leq C\left|\widetilde{\Omega}\right|^{\frac{1}{p'}-\frac{d-1}{d}} \|\nabla_y G^\varepsilon(x,\cdot)\|_{\mathrm{L}^{\frac{d}{d-1},\infty}(\widetilde{\Omega})} \leq CR^{1-\frac{d}{p}}.$$

Finalement on obtient

$$\|R_1^\varepsilon\|_{\mathrm{L}^\infty(\widetilde{\Omega})} \leq C\varepsilon^\nu R^{2-\frac{d}{p}-\nu} \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^p(\widetilde{\Omega})}. \quad (2.222)$$

De (2.221) et (2.222), on déduit que

$$\begin{aligned}\|R^\varepsilon\|_{\mathrm{L}^\infty(\Omega(0,R))} \leq & \frac{C}{R^{d/q}} \|u^\varepsilon - u^\star\|_{\mathrm{L}^q(\Omega(0,4R))} + CR^{1-\nu}\varepsilon^\nu \|\nabla u^\star\|_{\mathrm{L}^\infty(\Omega(0,4R))} \\ & + C\varepsilon^\nu R^{2-\frac{d}{p}-\nu} \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^p(\Omega(0,4R))}. \end{aligned} \quad (2.223)$$

En utilisant l'Hypothèse 7 pour contrôler le terme $\varepsilon\sum_{j=1}^d w_j\left(\frac{x}{\varepsilon}\right)\partial_j u^\star(x)$, l'Estimation (2.223) implique à son tour (2.217). □

Muni du Lemme 2.6.1, on peut alors majorer en fonction du second membre la différence entre $u^\varepsilon$ et $u^\star$ solutions de

$$\begin{cases} -\mathrm{div}\left(A\left(\dfrac{x}{\varepsilon}\right)\cdot\nabla u^\varepsilon(x)\right) = -\mathrm{div}\left(A^\star\cdot\nabla u^\star(x)\right) = f(x) & \text{dans} \quad \Omega, \\ u^\varepsilon = u^\star = 0 & \text{sur} \quad \partial\Omega. \end{cases} \quad (2.224)$$

**Lemme 2.6.2.** *Soit $A$ satisfaisant les Hypothèses 1, 2, 3, 7 et 8, et $\Omega$ un domaine borné régulier de classe $\mathrm{C}^{1,1}$. Soient $p > d$, $x_0 \neq y_0 \in \Omega$, $R := |x_0 - y_0|/16$. Fixons $f \in \mathrm{C}_{\mathrm{c}}^{\infty}(\Omega(y_0, 4R))$, et $u^\varepsilon$ et $u^\star$ solutions de* (2.224)*. Alors*

$$\|u^\varepsilon - u^\star\|_{\mathrm{L}^\infty(\Omega(x_0,R))} \leq C R^{2-\frac{d}{p}-\nu} \varepsilon^\nu \|f\|_{\mathrm{L}^p(\Omega)}. \tag{2.225}$$

*En outre, on a l'estimation globale suivante :*

$$\|u^\varepsilon - u^\star\|_{\mathrm{L}^\infty(\Omega)} \leq C \varepsilon^\nu \|f\|_{\mathrm{L}^p(\Omega)}. \tag{2.226}$$

*Remarque* 34. Le Lemme 2.6.2 fait écho à [94, Th. 3.4].

La démonstration du Lemme 2.6.2 repose d'une part sur le Lemme 2.6.1 et d'autre part sur l'utilisation de la théorie hilbertienne pour majorer $\|u^\varepsilon - u^\star\|_{\mathrm{L}^q}$. La subtilité technique de la preuve ci-dessous est qu'il faut s'assurer de l'homogénéité en $R$ de chacune des estimations effectuées. Cela implique notamment de faire des injections de Sobolev homogènes et de démontrer des estimations fines sur $\nabla u^\star$ et $\nabla^2 u^\star$.

*Démonstration de l'estimation* (2.225) *du Lemme 2.6.2.* On invoque le Lemme 2.6.1, lequel implique que $u^\varepsilon - u^\star$ satisfait (2.217) pour $q < 2$, dont nous allons contrôler précisément chaque terme à droite afin d'établir (2.225).

Le raisonnement se fait en 3 étapes. Dans l'Etape 1, nous commençons par démontrer des estimations fines sur $\nabla u^\star(x)$ et $\nabla^2 u^\star(x)$, selon que $x$ est proche ou loin de $y_0$. Puis, dans l'Etape 2, nous employons la Proposition 2.1.8 pour estimer $\|u^\varepsilon - u^\star\|_{\mathrm{L}^q}$. Pour ce faire, nous allons invoquer la Proposition 2.1.8 pour estimer $R^\varepsilon$ définie par (2.5). Malheureusement, comme $R^\varepsilon$ ne satisfait pas des conditions de Dirichlet homogènes, il faut scinder $R^\varepsilon$ en deux parties, chacune résolvant un problème elliptique ayant respectivement une condition au bord de Dirichlet homogène, et un second membre nul. Enfin, par la sous-linéarité renforcée des correcteurs, on parvient à transformer le contrôle sur $R^\varepsilon$ en un contrôle sur $u^\varepsilon - u^\star$. Enfin, dans l'Etape 3, on utilise les estimations des Etapes 1 et 2 et le Lemme 2.6.1 pour démontrer l'estimation voulue.

**Etape 1 :** Majorons tout d'abord $\left|\nabla^2 u^\star\right|$ et $|\nabla u^\star|$.

On majore différemment $\nabla u^\star(x)$ selon que $x$ est proche ou loin de $y_0$. Tout d'abord,

$$\begin{aligned} |\nabla u^\star(x)| \leq & \left| \int_{\Omega(y_0,4R)} \nabla_x G^\star(x,y) f(y) \mathrm{d}y \right| \\ \leq & C \|\nabla_x G^\star(x,\cdot)\|_{\mathrm{L}^{p'}(\Omega(y_0,4R))} \|f\|_{\mathrm{L}^p(\Omega)}. \end{aligned}$$

Donc, par le Théorème A.3.7, grâce à (A.9),

$$|\nabla u^\star(x)| \leq C \left( \int_{\mathrm{B}(y_0,4R)} \frac{1}{|x-y|^{p'(d-1)}} \mathrm{d}y \right)^{1/p'} \|f\|_{\mathrm{L}^p(\Omega)}.$$

Comme $p > d$, alors $p' < \frac{d}{d-1}$. Ainsi,

$$\left(\int_{\mathrm{B}(y_0,4R)} \frac{1}{|x-y|^{p'(d-1)}}\mathrm{d}y\right)^{1/p'} \leq \begin{cases} CR^{\frac{d-p'(d-1)}{p'}} & \text{si } |x-y_0| < 8R, \\ CR^{\frac{d}{p'}}|x-y_0|^{-d+1} & \text{si } |x-y_0| \geq 8R. \end{cases}$$

D'où

$$|\nabla u^\star(x)| \leq C \|f\|_{\mathrm{L}^p(\Omega)} \frac{R^{d-\frac{d}{p}}}{\max\left(R^{d-1}, |x-y_0|^{d-1}\right)}. \tag{2.227}$$

Notamment, on a l'estimation suivante :

$$\|\nabla u^\star\|_{\mathrm{L}^\infty(\Omega)} \leq CR^{1-\frac{d}{p}} \|f\|_{\mathrm{L}^p(\Omega)}. \tag{2.228}$$

De même, on majore différemment $\nabla^2 u^\star(x)$ selon que $x$ est proche ou loin de $y_0$. Grâce à [64, Th. 9.15 p. 241] et [64, Lem. 9.17 p. 232],

$$\left\|\nabla^2 u^\star\right\|_{\mathrm{L}^p(\Omega)} \leq C \|f\|_{\mathrm{L}^p(\Omega)}. \tag{2.229}$$

Pour $x \in \Omega \backslash \mathrm{B}(y_0, 8R)$, en utilisant l'Estimation (A.10) du Théorème A.3.7,

$$\begin{aligned} \left|\nabla^2 u^\star(x)\right| &= \left|\int_{\Omega(y_0,4R)} \nabla_x^2 G^\star(x,y) f(y) \mathrm{d}y\right| \\ &\leq \left\|\nabla_x^2 G^\star\right\|_{\mathrm{L}^{p'}(\Omega(y_0,4R))} \|f\|_{\mathrm{L}^p(\Omega)} \\ &\leq C \frac{R^{d-d/p}}{|x-y_0|^d} \|f\|_{\mathrm{L}^p(\Omega)} \end{aligned} \tag{2.230}$$

**Etape 2 :** Bornons maintenant $\|u^\varepsilon - u^\star\|_{\mathrm{L}^q(\Omega(x_0,4R))}$. Quitte à enlever une constante, on suppose que

$$w(y_0) = 0, \quad \text{et} \quad B(y_0) = 0. \tag{2.231}$$

On pose alors $R^\varepsilon$ définie par (2.5) et $H^\varepsilon$ est défini par (2.33). Pour majorer $R^\varepsilon$, on scinde $R^\varepsilon = R_1^\varepsilon + R_2^\varepsilon$, où

$$\begin{cases} -\operatorname{div}\left(A\left(\frac{x}{\varepsilon}\right) \cdot \nabla R_1^\varepsilon(x)\right) = \operatorname{div}\left(H^\varepsilon(x)\right) & \text{dans} \quad \Omega, \\ R_1^\varepsilon = 0 & \text{sur} \quad \partial\Omega, \end{cases} \tag{2.232}$$

et

$$\begin{cases} -\operatorname{div}\left(A\left(\frac{x}{\varepsilon}\right) \cdot \nabla R_2^\varepsilon(x)\right) = 0 & \text{dans} \quad \Omega, \\ R_2^\varepsilon(x) = -\varepsilon \sum_{j=1}^d w_j\left(\frac{x}{\varepsilon}\right) \partial_j u^\star(x) & \text{sur} \quad \partial\Omega. \end{cases} \tag{2.233}$$

**Estimation sur $R_1^\varepsilon$ :** En testant (2.232) contre $R_1^\varepsilon$, on obtient

$$\|\nabla R_1^\varepsilon\|_{\mathrm{L}^2(\Omega)} \leq C \|H^\varepsilon\|_{\mathrm{L}^2(\Omega)} . \tag{2.234}$$

Nous prétendons que

$$\|H^\varepsilon\|_{\mathrm{L}^2(\Omega)} \leq C\varepsilon^\nu R^{\frac{d}{2}-\frac{d}{p}+1-\nu} \|f\|_{\mathrm{L}^p(\Omega)} . \tag{2.235}$$

En effet, si $\nu = 1$, (2.235) est une conséquence de l'Estimation (2.78), de [64, Th. 9.15 p. 241] et [64, Lem. 9.17 p. 232] et de l'inégalité de Hölder. Si maintenant $\nu \neq 1$, il va falloir utiliser les majorations plus fines précédemment montrées sur $\nabla^2 u^\star$. Grâce à (2.230) et aux Hypothèses 7 et 8 (et grâce à (2.231)), on a

$$\begin{aligned}
\|H^\varepsilon\|_{\mathrm{L}^2(\Omega\setminus \mathrm{B}(y_0,8R))} \leq & C\varepsilon^\nu R^{d-d/p} \|f\|_{\mathrm{L}^p(\Omega)} \left( \int_{\Omega\setminus \mathrm{B}(y_0,8R)} \left( \frac{|x-y_0|^{1-\nu}}{|x-y_0|^d} \right)^2 \mathrm{d}x \right)^{1/2} \\
\leq & C\varepsilon^\nu R^{d-d/p} \|f\|_{\mathrm{L}^p(\Omega)} \left( \int_R^{\mathrm{Diam}(\Omega)} z^{1-2\nu-d} \mathrm{d}z \right)^{1/2} \\
\leq & C\varepsilon^\nu R^{d-d/p} \|f\|_{\mathrm{L}^p(\Omega))} R^{1-\nu-d/2} \\
\leq & C\varepsilon^\nu R^{d/2-d/p+1-\nu} \|f\|_{\mathrm{L}^p(\Omega)} .
\end{aligned} \tag{2.236}$$

De plus, grâce à (2.78), à l'inégalité de Hölder et à (2.229),

$$\begin{aligned}
\|H^\varepsilon\|_{\mathrm{L}^2(\Omega(y_0,8R))} \leq & C\varepsilon^\nu R^{1-\nu} \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^2(\Omega(y_0,8R))} \\
\leq & C\varepsilon^\nu R^{d/2-d/p+1-\nu} \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^p(\Omega(y_0,8R))} \\
\leq & C\varepsilon^\nu R^{d/2-d/p+1-\nu} \|f\|_{\mathrm{L}^p(\Omega)} .
\end{aligned} \tag{2.237}$$

De (2.236) et (2.237), on déduit (2.235).

De (2.234) et (2.235) découle l'estimation suivante :

$$\|\nabla R_1^\varepsilon\|_{\mathrm{L}^2(\Omega)} \leq C\varepsilon^\nu R^{\frac{d}{2}-\frac{d}{p}+1-\nu} \|f\|_{\mathrm{L}^p(\Omega)} . \tag{2.238}$$

Par l'inégalité de Hölder, puis par injection de Sobolev de $\mathrm{H}_0^1(\Omega)$ dans $\mathrm{L}^{\frac{2d}{d-2}}(\Omega)$,

$$\begin{aligned}
\|R_1^\varepsilon\|_{\mathrm{L}^q(\Omega(x_0,4R))} \leq & CR^{\frac{d}{q}-(\frac{d}{2}-1)} \|R_1^\varepsilon\|_{\mathrm{L}^{\frac{2d}{d-2}}(\Omega(x_0,4R))} \\
\leq & CR^{\frac{d}{q}+1-\frac{d}{2}} \|\nabla R_1^\varepsilon\|_{\mathrm{L}^2(\Omega)} ,
\end{aligned}$$

d'où, grâce à (2.238),

$$\|R_1^\varepsilon\|_{\mathrm{L}^q(\Omega(x_0,4R))} \leq C\varepsilon^\nu R^{2+\frac{d}{q}-\frac{d}{p}-\nu} \|f\|_{\mathrm{L}^p(\Omega)} . \tag{2.239}$$

**Estimation sur $R_2^\varepsilon$ :** Appliquons le principe du maximum sur $R_2^\varepsilon$ (voir [64, Th. 8.1 p. 179])

$$\|R_2^\varepsilon\|_{\mathrm{L}^\infty(\Omega)} \leq \left\| \varepsilon \sum_{j=1}^{d} w_j \left(\frac{x}{\varepsilon}\right) \partial_j u^\star(x) \right\|_{\mathrm{L}^\infty(\partial\Omega)} . \tag{2.240}$$

Ainsi, par l'Hypothèse 7, on déduit de (2.240) et (2.227)

$$\|R_2^\varepsilon\|_{\mathrm{L}^\infty(\Omega)} \leq C\varepsilon^\nu R^{d-\frac{d}{p}} \|f\|_{\mathrm{L}^p(\Omega)} \sup_{x\in\partial\Omega} \left\{ \frac{|x-y_0|^{1-\nu}}{\max\left(R^{d-1}, |x-y_0|^{d-1}\right)} \right\}.$$

Or

$$\frac{|x-y_0|^{1-\nu}}{\max\left(R^{d-1}, |x-y_0|^{d-1}\right)} = \begin{cases} \dfrac{|x-y_0|^{1-\nu}}{R^{d-1}} \leq R^{2-d-\nu} & \text{si } |x-y_0| < R, \\ \dfrac{|x-y_0|^{1-\nu}}{|x-y_0|^{d-1}} \leq R^{2-d-\nu} & \text{si } |x-y_0| > R. \end{cases}$$

Par conséquent,

$$\|R_2^\varepsilon\|_{\mathrm{L}^\infty(\Omega)} \leq C\varepsilon^\nu R^{2-\nu-\frac{d}{p}} \|f\|_{\mathrm{L}^p(\Omega)} . \tag{2.241}$$

**Bornons $u^\varepsilon - u^\star$** Enfin, par inégalité triangulaire,

$$\begin{aligned} \|u^\varepsilon - u^\star\|_{\mathrm{L}^q(\Omega(x_0,4R))} \leq & \left\| \varepsilon \sum_{j=1}^{d} w_j \left(\frac{\cdot}{\varepsilon}\right) \partial_j u^\star \right\|_{\mathrm{L}^q(\Omega(x_0,4R))} \\ & + \|R_1^\varepsilon\|_{\mathrm{L}^q(\Omega(x_0,4R))} + \|R_2^\varepsilon\|_{\mathrm{L}^q(\Omega(x_0,4R))} . \end{aligned} \tag{2.242}$$

Or, par l'Hypothèse 7 et par (2.228) (rappelons que $|x_0 - y_0| = 16R$) :

$$\begin{aligned} & \left\| \varepsilon \sum_{j=1}^{d} w_j \left(\frac{\cdot}{\varepsilon}\right) \partial_j u^\star \right\|_{\mathrm{L}^q(\Omega(x_0,4R))} \\ & \leq CR^{d/q}\varepsilon \left\| w_j \left(\frac{\cdot}{\varepsilon}\right) \right\|_{\mathrm{L}^\infty(\Omega(x_0,4R))} \|\nabla u^\star\|_{\mathrm{L}^\infty(\Omega(x_0,4R))} \\ & \leq CR^{2+\frac{d}{q}-\frac{d}{p}-\nu}\varepsilon^\nu \|f\|_{\mathrm{L}^p(\Omega)} . \end{aligned} \tag{2.243}$$

Donc, on déduit de (2.242), (2.241), (2.239), et (2.243) que

$$\|u^\varepsilon - u^\star\|_{\mathrm{L}^q(\Omega(x_0,4R))} \leq C\varepsilon^\nu R^{2+\frac{d}{q}-\frac{d}{p}-\nu} \|f\|_{\mathrm{L}^p(\Omega)} . \tag{2.244}$$

**Etape 3 :** Nous rassemblons maintenant les estimations démontrées dans les Etapes 1 et 2. De (2.217), (2.228), (2.229) et (2.244) découle (2.225), ce qui conclut la preuve. □

La démonstration de (2.226) est beaucoup plus simple :

*Démonstration de l'estimation* (2.226) *du Lemme 2.6.2.* Soit $p > d$. On reprend la démonstration ci-dessus, qui nous amène donc aux estimations (2.238) et (2.241). Cependant, on utilise ensuite l'injection de Sobolev sur tout $\Omega$, ce qui induit l'inégalité suivante à la place de (2.239) :

$$\|R_1^\varepsilon\|_{\mathrm{L}^q(\Omega)} \leq C\varepsilon^\nu \|f\|_{\mathrm{L}^p(\Omega)}. \tag{2.245}$$

On en déduit donc

$$\|u^\varepsilon - u^\star\|_{\mathrm{L}^q(\Omega)} \leq C\varepsilon^\nu \|f\|_{\mathrm{L}^p(\Omega)}. \tag{2.246}$$

à la place de (2.244). Alors, l'Estimation (2.226) découle de (2.217), (2.228), (2.229) et (2.246). □

Nous sommes à présent en mesure de démontrer le Théorème 2.1.9, qui repose sur un argument de dualité.

*Démonstration du Théorème 2.1.9.* Comme la démonstration du point (i) est une version simplifiée de la preuve du point (ii), on commence par démontrer le point (ii).

**Démonstration du point (ii)** Soient $p > d$, $x_0, y_0 \in \Omega$ et $R := |x_0 - y_0|/16$. Soit $f \in \mathrm{C}_\mathrm{c}^\infty(\Omega(y_0, 4R))$ et $u^\varepsilon$ et $u^\star$ satisfaisant (2.224). Par définition :

$$u^\varepsilon(x) = \int_\Omega G^\varepsilon(x,y) f(y) \mathrm{d}y \qquad \text{et} \qquad u^\star(x) = \int_\Omega G^\star(x,y) f(y) \mathrm{d}y.$$

Ainsi le Lemme 2.6.2 implique que

$$\left| \int_\Omega \left( G^\varepsilon(x_0,y) - G^\star(x_0,y) \right) f(y) \mathrm{d}y \right| \leq C\varepsilon^\nu R^{2-\frac{d}{p}-\nu} \|f\|_{\mathrm{L}^p(\Omega(y_0,R))}.$$

D'où, par dualité,

$$\|G^\varepsilon(x_0,\cdot) - G^\star(x_0,\cdot)\|_{\mathrm{L}^{p'}(\Omega(y_0,4R))} \leq C\varepsilon^\nu R^{2-\frac{d}{p}-\nu}. \tag{2.247}$$

Supposons maintenant que $A^T$ satisfait les Hypothèses 4, 5 et 6. Comme

$$\begin{cases} -\operatorname{div}_y \left( A^T\left(\frac{y}{\varepsilon}\right) \cdot \nabla_y G^\varepsilon(x_0,y) \right) = 0 & \text{dans} \quad \Omega(y_0,4R), \\ -\operatorname{div}_y \left( (A^\star)^T \cdot \nabla_y G^\star(x_0,y) \right) = 0 & \text{dans} \quad \Omega(y_0,4R), \\ G^\varepsilon(x_0,\cdot) = G^\star(x_0,\cdot) = 0 & \text{sur} \quad \Gamma_\Omega(y_0,4R), \end{cases}$$

on peut appliquer le Lemme 2.6.1 à $G^\varepsilon(x_0,\cdot) - G^\star(x_0,\cdot)$ avec $q = p'$. D'où

$$\begin{aligned}|G^\varepsilon(x_0,y_0) - G^\star(x_0,y_0)| \leq & CR^{-d/p'} \left\|G^\varepsilon(x_0,\cdot) - G^\star(x_0,\cdot)\right\|_{\mathrm{L}^{p'}(\Omega(y_0,4R))} \\ & + C\varepsilon^\nu R^{1-\nu} \left\|\nabla_y G^\star(x_0,\cdot)\right\|_{\mathrm{L}^\infty(\Omega(y_0,4R))} \\ & + C\varepsilon^\nu R^{2-\frac{d}{p}-\nu} \left\|\left(\nabla_y\right)^2 G^\star(x_0,\cdot)\right\|_{\mathrm{L}^p(\Omega(y_0,4R))}\end{aligned}$$

Or, grâce aux Estimations (A.9) et (A.10) (on peut intervertir les rôles de $x$ et $y$ car $A^T$ satisfait les Hypothèses 3, 7 et 8), et en invoquant (2.247), on obtient

$$|G^\varepsilon(x_0,y_0) - G^\star(x_0,y_0)| \leq C\varepsilon^\nu R^{2-d-\nu}.$$

Comme $16R = |x_0 - y_0|$, cela conclut la démonstration du point (ii) du Théorème 2.1.9.

**Démonstration du point (i)** Le point (i) se démontre par dualité à la manière de (2.247), en utilisant l'estimation (2.226) à la place de l'estimation (2.225). □

### 2.6.2 Approximation de $u^\varepsilon$ dans $\mathrm{L}^p$

Nous justifions dans cette section le Corollaire 2.1.10.

*Remarque* 35 (Architecture logique). Le Corollaire 2.1.10 repose entièrement sur le Théorème 2.1.9 et sur un théorème de convolution dans des espaces de Marcinkiewicz. Il permet ensuite d'établir le Théorème 2.1.11.

L'Estimation (2.226) du Lemme 2.6.2 établit le point (i) du Corollaire 2.1.10. Le second point découle du Théorème 2.1.9.

*Démonstration du Corollaire 2.1.10(ii).* L'Estimation (2.39) une conséquence de l'inégalité de convolution [157, Th. 3.4] (voir aussi [127]). En effet, grâce au Théorème 2.1.9,

$$|G^\varepsilon(x,y) - G^\star(x,y)| \leq \varepsilon^\nu |x-y|^{2-d-\nu},$$

et $g : x \mapsto |x|^{2-d-\nu}$ satisfait

$$g \in \mathrm{L}^{\frac{d}{d-2+\nu},\infty}(\Omega).$$

Supposons que $p < +\infty$ et que (2.40) est satisfaite. Ainsi, par l'inégalité de convolution [157, Th. 3.4], si $f \in \mathrm{L}^q(\Omega)$, la fonction suivante

$$u := x \mapsto \int_\Omega |G^\varepsilon(x,y) - G^\star(x,y)|\, f(y)\mathrm{d}y$$

est dans $\mathrm{L}^{p,q}(\Omega) \subset \mathrm{L}^{p,p}(\Omega) = \mathrm{L}^p(\Omega)$ car $q \leq p$ (voir [21, p. 8]), d'où (2.39).

Le cas (2.41) se traite en utilisant le théorème de Young classique. □

### 2.6.3 Approximation $u^\varepsilon$ dans $\mathrm{W}^{1,p}$

Dans cette Section, on majore $\nabla R^\varepsilon$, pour $R^\varepsilon$, $u^\varepsilon$ et $u^\star$ respectivement définis par (2.5), (2.1) et (2.2). Selon l'intégrabilité et la régularité de $f$ présente dans (2.1) et (2.2), on estime $\nabla R^\varepsilon$ dans une topologie plus ou moins fine.

*Remarque* 36. Dans le cas (iii) du Théorème 2.1.11, si $\Omega$ est de classe $\mathrm{C}^{2,\gamma}$ pour $\gamma > 0$, on a en outre que $R^\varepsilon \in \mathrm{W}^{1,\infty}(\Omega)$. C'est une conséquence du Théorème [64, Th. 6.14 p. 107] appliqué à $u^\star$.

*Remarque* 37 (Architecture logique). Le Théorème 2.1.11 permet de démontrer le Théorème 1.1.1 et le Corollaire 1.1.2. Il repose sur le Corollaire 2.1.10 et sur les Propositions 2.1.6 et 2.1.7.

Procédons maintenant à la preuve du Théorème 2.1.11. Le Corollaire 2.1.10 fournit un point de départ, en ce qu'il permet d'estimer $\|u^\varepsilon - u^\star\|_{\mathrm{L}^2(\Omega)}$. Puis, on utilise la Proposition 2.1.8 qui fournit une équation elliptique sur $R^\varepsilon$. Alors, les différentes Propositions d'estimations 2.1.6, et 2.1.7, permettent d'estimer $\nabla R^\varepsilon$ dans les normes $\mathrm{L}^p$ pour $p \in [2, +\infty[$, respectivement dans la norme $\mathrm{L}^\infty$. La preuve est compliquée par le fait que $R^\varepsilon$ n'est pas nul sur $\partial\Omega$. Pour gérer cette difficulté, le principe est de scinder $R^\varepsilon$ en deux fonctions $R_1^\varepsilon$ et $R_2^\varepsilon$, où chacune satisfait un problème elliptique, avec respectivement un second membre nul ou des conditions au bord de Dirichlet homogènes.

*Démonstration du Théorème 2.1.11.* On pose $\Omega_1 \subset\subset \Omega_2 \subset\subset \Omega$. Montrons tout d'abord les régularités annoncées sur $R^\varepsilon$.

Si $f \in \mathrm{L}^p(\Omega)$, pour $p \in [2, +\infty[$, par [64, Th. 9.15 p. 241] et [64, Lem. 9.17 p. 232], alors $u^\star \in \mathrm{W}^{2,p}(\Omega)$, avec

$$\left\|\nabla^2 u^\star\right\|_{\mathrm{L}^p(\Omega)} \leq C \left\|f\right\|_{\mathrm{L}^p(\Omega)}, \tag{2.248}$$

En outre, une application directe de (2.197) et des estimations de Marcinkiewicz implique que $u^\varepsilon \in \mathrm{W}^{1,p}(\Omega)$ Par ailleurs, par la Proposition 2.3.3, $w_j(\cdot/\varepsilon) \in \mathrm{C}^{1,\alpha}\left(\bar{\Omega}\right)$, pour tout $j \in [\![1, d]\!]$. Donc $R^\varepsilon \in \mathrm{W}^{1,p}(\Omega)$.

Si maintenant $f \in \mathrm{C}^{0,\beta}\left(\bar{\Omega}\right)$, quitte à prendre $\beta < \alpha$, par [64, Cor. 8.36 p. 212], on a $u^\star \in \mathrm{C}^{1,\beta}\left(\bar{\Omega}\right)$. En outre, grâce à [64, Cor. 6.3 p. 93],

$$\|u^\star\|_{\mathrm{C}^{2,\beta}(\Omega_1)} \leq C \left\|f\right\|_{\mathrm{C}^{0,\beta}(\Omega)}. \tag{2.249}$$

Donc $R^\varepsilon \in \mathrm{W}^{1,\infty}(\Omega)$.

**Preuve de (i) :** Soit $f \in \mathrm{L}^p(\Omega)$, pour $p > d$. Par la Proposition 2.1.8, $R^\varepsilon$ satisfait (2.32), pour $H^\varepsilon$ définie par (2.33). Pour pouvoir estimer $\nabla R^\varepsilon$, il faut traiter à part la donnée au bord. On scinde donc $R^\varepsilon = R_1^\varepsilon + R_2^\varepsilon$ où

$$\begin{cases} -\operatorname{div}\left(A\left(\dfrac{x}{\varepsilon}\right) \cdot \nabla R_1^\varepsilon(x)\right) = 0 & \text{dans} \quad \Omega, \\ R_1^\varepsilon(x) = -\varepsilon \displaystyle\sum_{k=1}^{d} w_k\left(\frac{x}{\varepsilon}\right) \partial_k u^\star(x) & \text{sur} \quad \partial\Omega. \end{cases}$$

et

$$\begin{cases} -\operatorname{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla R_2^\varepsilon(x)\right)=\operatorname{div}\left(H^\varepsilon(x)\right) & \text{dans} \quad \Omega,\\ R_2^\varepsilon=0 & \text{sur} \quad \partial\Omega. \end{cases}$$

En testant l'équation précédente contre $R_2^\varepsilon$, on obtient

$$\|\nabla R_2^\varepsilon\|_{\mathrm{L}^2(\Omega)} \leq C\,\|H^\varepsilon\|_{\mathrm{L}^2(\Omega)}\,.$$

Alors, grâce à (2.78) et à (2.248),

$$\|\nabla R_2^\varepsilon\|_{\mathrm{L}^2(\Omega)} \leq C\varepsilon^\nu\left\|\nabla^2 u^\star\right\|_{\mathrm{L}^2(\Omega)} \leq C\varepsilon^\nu\,\|f\|_{\mathrm{L}^2(\Omega)}\,. \tag{2.250}$$

L'estimation ci-dessus n'est pas totalement suffisante ; il faut encore raboter les termes de bord se manifestant par $R_1^\varepsilon$. En invoquant l'inégalité de Cacciopoli (Lemme A.3.2), on a

$$\|\nabla R_1^\varepsilon\|_{\mathrm{L}^2(\Omega_1)} \leq C\,\|R_1^\varepsilon\|_{\mathrm{L}^2(\Omega)} \leq C\,\|R_2^\varepsilon\|_{\mathrm{L}^2(\Omega)} + C\,\|R^\varepsilon\|_{\mathrm{L}^2(\Omega)}\,.$$

Grâce au Corollaire 2.1.10(i), on a

$$\|R^\varepsilon\|_{\mathrm{L}^2(\Omega)} \leq C\varepsilon^\nu\,\|f\|_{\mathrm{L}^p(\Omega)}\,. \tag{2.251}$$

Par conséquent, grâce à l'inégalité de Poincaré appliquée à $R_2^\varepsilon$, à (2.250) et à (2.251), on démontre que

$$\|\nabla R_1^\varepsilon\|_{\mathrm{L}^2(\Omega_1)} \leq C\,\|\nabla R_2^\varepsilon\|_{\mathrm{L}^2(\Omega)} + C\varepsilon^\nu\,\|f\|_{\mathrm{L}^p(\Omega)} \leq C\varepsilon^\nu\,\|f\|_{\mathrm{L}^p(\Omega)}\,. \tag{2.252}$$

Alors, (2.250) et (2.252) impliquent que

$$\|\nabla R^\varepsilon\|_{\mathrm{L}^2(\Omega_1)} \leq C\varepsilon^\nu\,\|f\|_{\mathrm{L}^2(\Omega)}\,, \tag{2.253}$$

.

Soit $f \in \mathrm{L}^p(\Omega)$. On recouvre $\Omega_1$ avec un nombre fini de boules $\mathrm{B}_j$ telles que $2\mathrm{B}_j \subset \Omega_2$ et on applique la Proposition 2.1.6 à $R^\varepsilon$. Ainsi,

$$\|\nabla R^\varepsilon\|_{\mathrm{L}^p(\Omega_1)} \leq C\,\|\nabla R^\varepsilon\|_{\mathrm{L}^2(\Omega_2)} + C\,\|H^\varepsilon\|_{\mathrm{L}^p(\Omega)}\,.$$

De (2.253) et (2.78), puis (2.248), découle alors l'inégalité suivante :

$$\begin{aligned}\|\nabla R^\varepsilon\|_{\mathrm{L}^p(\Omega_1)} &\leq C\varepsilon^\nu\,\|f\|_{\mathrm{L}^2(\Omega)} + C\varepsilon^\nu\left\|\nabla^2 u^\star\right\|_{\mathrm{L}^p(\Omega)}\\ &\leq C\varepsilon^\nu\,\|f\|_{\mathrm{L}^p(\Omega)}\,.\end{aligned}$$

d'où (2.42).

**Preuve de (i') :** C'est une adaptation immédiate de la démonstration du Théorème 2.1.11(i) ci-dessus. Il suffit de remplacer l'estimation (2.251) par l'estimation suivante :

$$\|R^\varepsilon\|_{\mathrm{L}^2(\Omega)} \leq C\varepsilon^\nu\,\|f\|_{\mathrm{L}^2(\Omega)}\,,$$

qui est une conséquence du point (ii) du Corollaire 2.1.10.

**Preuve de (ii) :** Soit enfin $f \in \mathrm{C}^{0,\beta}(\Omega)$. Quitte à changer $\beta$, on suppose que $0 < \beta \leq \alpha$. En recouvrant $\Omega_1$ d'un nombre fini de boules $\mathrm{B}_j$ telles que $2\mathrm{B}_j \subset \Omega_2$ (pour $\Omega_2$ tel que $\Omega_1 \subset\subset \Omega_2 \subset\subset \Omega$) et en appliquant la Proposition 2.1.7 à $R^\varepsilon$ on obtient

$$\begin{aligned}\|\nabla R^\varepsilon\|_{\mathrm{L}^\infty(\Omega_1)} \leq & C \|\nabla R^\varepsilon\|_{\mathrm{L}^2(\Omega_2)} + C\varepsilon^\beta \|H^\varepsilon\|_{\dot{\mathrm{C}}^{0,\beta}(\Omega)} \\ & + C \ln\left(1+\varepsilon^{-1}\right) \|H^\varepsilon\|_{\mathrm{L}^\infty(\Omega)}.\end{aligned}$$

En invoquant alors (2.253), (2.78) et (2.79), on déduit que

$$\begin{aligned}\|\nabla R^\varepsilon\|_{\mathrm{L}^\infty(\Omega_1)} \leq & C\varepsilon^\nu \|f\|_{\mathrm{L}^2(\Omega)} + C\varepsilon^{\nu+\beta} \left\|\nabla^2 u^\star\right\|_{\dot{\mathrm{C}}^{0,\beta}(\Omega)} \\ & + C\varepsilon^\nu \ln\left(2+\varepsilon^{-1}\right) \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^\infty(\Omega)}. \end{aligned} \tag{2.254}$$

De (2.249) et (2.254) découle (2.43). □

### 2.6.4 Approximation de $\nabla_x G^\varepsilon$, $\nabla_y G^\varepsilon$ et $\nabla_x \nabla_y G^\varepsilon$.

Dans cette Section, nous discutons et démontrons le Théorème 2.1.12, qui permet d'approximer $\nabla_x G^\varepsilon$, $\nabla_y G^\varepsilon$ et $\nabla_x \nabla_y G^\varepsilon$ à l'aide des correcteurs et de la fonction de Green homogénéisée $G^\star$. Le Théorème 2.1.12 est en fait une version du Théorème 2.1.11 qui serait fait pour un membre singulier (en l'occurrence un Dirac).

*Remarque* 38 (Architecture logique). Le Théorème 2.1.12 est démontré grâce aux éléments évoqués dans la Remarque 31, et sur la Proposition 2.3.3. Il est ensuite utilisé pour établir le Théorème 1.1.4.

La démonstration du Théorème 2.1.12 repose sur la Proposition 2.1.7, qui permet de construire l'approximation sur $\nabla_x G^\varepsilon$ à partir de celle sur $G^\varepsilon$, puis pour $\nabla_x \nabla_y G^\varepsilon$. Le Théorème 2.1.9 permet d'initialiser ce processus, en approximant $G^\varepsilon$. La preuve ci-dessous reprend celle de [94].

La Proposition 2.1.7 et les Estimations (2.78) et (2.79) permettent de démontrer le :

**Lemme 2.6.3** ( Analogue du Lemme 3.5 de [94])**.** *Soit $A$ satisfaisant les Hypothèses 1, 2, 3, 7 et 8. Soit $x_0 \in \mathbb{R}^d$ et $R \in ]0,1[$. Soit $\varepsilon \in ]0,1[$. Supposons que, sur $\mathrm{B} := \mathrm{B}(x_0, R)$,*

$$-\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right) \cdot \nabla u^\varepsilon(x)\right) = -\mathrm{div}\left(A^\star \cdot \nabla u^\star(x)\right),$$

*où $u^\star \in \mathrm{C}^{2,\alpha}(\mathrm{B})$. Alors, il existe une constante $C$ ne dépendant que de $A$ et de $\alpha$ telle que*

$$\begin{aligned}&\left|\nabla u^\varepsilon(x_0) - \nabla u^\star(x_0) - \sum_{k=1}^d \nabla w_k\left(\frac{x_0}{\varepsilon}\right) \partial_k u^\star(x_0)\right| \\ &\leq CR^{-1} \|u^\varepsilon - u^\star\|_{\mathrm{L}^\infty(\mathrm{B})} + CR^{-\nu}\varepsilon^\nu \|\nabla u^\star\|_{\mathrm{L}^\infty(\mathrm{B})} \\ &\quad + C\varepsilon^\nu R^{1-\nu} \ln\left(2+\varepsilon^{-1}\right) \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^\infty(\mathrm{B})} + CR^{1-\nu}\varepsilon^{\nu+\alpha} \left\|\nabla^2 u^\star\right\|_{\dot{\mathrm{C}}^{0,\alpha}(\mathrm{B})}. \end{aligned} \tag{2.255}$$

*Démonstration.* Sans perdre en généralité, on suppose que $x_0 = 0$. Soit $B$ le potentiel associé à (2.27) (quitte à retirer une constante, on suppose que $B(0) = 0$), et on pose $H^\varepsilon$ défini par (2.33). Soit $R^\varepsilon$ définie par (2.5). Alors, par la Proposition 2.1.8, $R^\varepsilon$ vérifie :

$$-\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla R^\varepsilon(x)\right) = \mathrm{div}\left(H^\varepsilon(x)\right).$$

Donc, par la Proposition 2.1.7, avec $\beta = \alpha$ :

$$\begin{aligned}\left\|\nabla R^\varepsilon\right\|_{\mathrm{L}^\infty\left(\frac{1}{4}\mathrm{B}\right)} \leq & CR^{-\frac{d}{2}}\left\|\nabla R^\varepsilon\right\|_{\mathrm{L}^2\left(\frac{1}{2}\mathrm{B}\right)} + C\varepsilon^\alpha\left\|H^\varepsilon\right\|_{\dot{\mathrm{C}}^{0,\alpha}\left(\frac{1}{2}\mathrm{B}\right)} \\ & + C\ln\left(1+R\varepsilon^{-1}\right)\left\|H^\varepsilon\right\|_{\mathrm{L}^\infty\left(\frac{1}{2}\mathrm{B}\right)} \end{aligned} \tag{2.256}$$

Grâce aux Estimations (2.78) et (2.79), on déduit que

$$\left\|H^\varepsilon\right\|_{\mathrm{L}^\infty(\mathrm{B})} \leq C\varepsilon^\nu R^{1-\nu}\left\|\nabla^2 u^\star\right\|_{\mathrm{L}^\infty(\mathrm{B})}, \tag{2.257}$$

$$\left\|H^\varepsilon\right\|_{\dot{\mathrm{C}}^{0,\alpha}(\mathrm{B})} \leq C\varepsilon^\nu R^{1-\nu}\left\|\nabla^2 u^\star\right\|_{\dot{\mathrm{C}}^{0,\alpha}(\mathrm{B})} + C\varepsilon^{\nu-\alpha}R^{1-\nu}\left\|\nabla^2 u^\star\right\|_{\mathrm{L}^\infty(\mathrm{B})}. \tag{2.258}$$

Ainsi, on déduit de (2.256) que

$$\begin{aligned}\left\|\nabla R^\varepsilon\right\|_{\mathrm{L}^\infty\left(\frac{1}{4}\mathrm{B}\right)} \leq & CR^{-\frac{d}{2}}\left\|\nabla R^\varepsilon\right\|_{\mathrm{L}^2\left(\frac{1}{2}\mathrm{B}\right)} + C\varepsilon^{\nu+\alpha}R^{1-\nu}\left\|\nabla^2 u^\star\right\|_{\dot{\mathrm{C}}^{0,\alpha}(\mathrm{B})} \\ & + C\varepsilon^\nu R^{1-\nu}\left(1+\ln\left(1+\varepsilon^{-1}\right)\right)\left\|\nabla^2 u^\star\right\|_{\mathrm{L}^\infty(\mathrm{B})}. \end{aligned} \tag{2.259}$$

Majorons $\left\|\nabla R^\varepsilon\right\|_{\mathrm{L}^2\left(\frac{1}{2}\mathrm{B}\right)}$. De même que précédemment, on scinde $R^\varepsilon$ en deux parties, à savoir $R^\varepsilon = R_1^\varepsilon + R_2^\varepsilon$, où

$$\begin{cases} -\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla R_1^\varepsilon(x)\right) = \mathrm{div}\left(H^\varepsilon(x)\right) & \text{dans} \quad \mathrm{B}, \\ R_1^\varepsilon(x) = 0 & \text{sur} \quad \partial\mathrm{B}, \end{cases} \tag{2.260}$$

et :

$$\begin{cases} -\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla R_2^\varepsilon(x)\right) = 0 & \text{dans} \quad \mathrm{B}, \\ R_2^\varepsilon(x) = R^\varepsilon(x) & \text{sur} \quad \partial\mathrm{B}. \end{cases} \tag{2.261}$$

En testant (2.260) contre $R_1^\varepsilon$ et grâce à (2.257),

$$\left\|\nabla R_1^\varepsilon\right\|_{\mathrm{L}^2(\mathrm{B})} \leq C\left\|H^\varepsilon\right\|_{\mathrm{L}^2(\mathrm{B})} \leq C\varepsilon^\nu R^{\frac{d}{2}+1-\nu}\left\|\nabla^2 u^\star\right\|_{\mathrm{L}^\infty(\mathrm{B})}. \tag{2.262}$$

En utilisant l'inégalité de Cacciopoli (voir le Lemme A.3.2), on déduit de (2.261) que

$$\left\|\nabla R_2^\varepsilon\right\|_{\mathrm{L}^2\left(\frac{1}{2}\mathrm{B}\right)} \leq CR^{-1}\left\|R_2^\varepsilon\right\|_{\mathrm{L}^2(\mathrm{B})} \leq CR^{-1}\left\|R_2^\varepsilon\right\|_{\mathrm{L}^\infty(\mathrm{B})}.$$

Puis, grâce au principe du maximum [64, Th. 8.1 p. 179], on déduit

$$\left\|\nabla R_2^\varepsilon\right\|_{\mathrm{L}^2\left(\frac{1}{2}\mathrm{B}\right)} \leq CR^{\frac{d}{2}-1}\left\|R^\varepsilon\right\|_{\mathrm{L}^\infty(\mathrm{B})}. \tag{2.263}$$

Les estimations (2.262) et (2.263) impliquent à leur tour

$$\|\nabla R^\varepsilon\|_{\mathrm{L}^2\left(\frac{1}{2}\mathrm{B}\right)} \leq R^{\frac{d}{2}-1} \|R^\varepsilon\|_{\mathrm{L}^\infty(\mathrm{B})} + CR^{\frac{d}{2}+1-\nu}\varepsilon^\nu \left\|\nabla^2 u^\star\right\|_{\mathrm{L}^\infty(\mathrm{B})}. \tag{2.264}$$

Grâce à l'Hypothèse 7,

$$\left\|\varepsilon \sum_{k=1}^d w_k\left(\frac{x}{\varepsilon}\right)\partial_k u^\star(x)\right\|_{\mathrm{L}^\infty(\mathrm{B})} \leq CR^{1-\nu}\varepsilon^\nu \|\nabla u^\star\|_{\mathrm{L}^\infty(\mathrm{B})}.$$

D'où

$$\|R^\varepsilon\|_{\mathrm{L}^\infty(\mathrm{B})} \leq CR^{1-\nu}\varepsilon^\nu \|\nabla u^\star\|_{\mathrm{L}^\infty(\mathrm{B})} + C\|u^\varepsilon - u^\star\|_{\mathrm{L}^\infty(\mathrm{B})}. \tag{2.265}$$

L'estimation (2.255) découle alors de (2.259), (2.264) et (2.265). □

La démonstration de (2.44) et (2.45) découle du Lemme 2.6.3 du Théorème 2.1.9.

*Démonstration des estimations* (2.44) *et* (2.45). On se contente de ne montrer que (2.44), comme (2.45) se montre exactement de la même manière. Soit $x_0 \neq y_0$, avec $x_0 \in \Omega_1$, $y_0 \in \Omega$. Si $|x_0 - y_0| \leq \varepsilon$, alors par le Théorème 2.1.5, (2.44) est vérifiée. On se concentre donc sur le cas où $|x_0 - y_0| > \varepsilon$.

On pose :

$$R := \frac{\min\left(\mathrm{d}(x_0, \partial\Omega), |x_0 - y_0|\right)}{8}. \tag{2.266}$$

Remarquons que, comme $\mathrm{d}(x_0, \partial\Omega) \geq \mathrm{d}(\Omega_1, \partial\Omega) > 0$ et que $|x_0 - y_0| < \mathrm{Diam}(\Omega)$, alors il existe une constante $C > 0$ ne dépendant que de $\Omega$ et $\Omega_1$ telle que

$$R \geq C|x_0 - y_0|. \tag{2.267}$$

Par définition, sur $\mathrm{B}(x_0, R)$ :

$$-\mathrm{div}_x\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla_x G^\varepsilon(x, y_0)\right) = -\mathrm{div}\left(A^\star\cdot\nabla_x G^\star(x, y_0)\right) = 0.$$

Grâce au Lemme 2.6.3,

$$\begin{aligned}
&\left|\partial_{x_i} G^\varepsilon(x_0, y_0) - \sum_{j=1}^d \left\{\delta_{ij} + \partial_i w_j\left(\frac{x_0}{\varepsilon}\right)\right\}\partial_{x_j} G^\star(x_0, y_0)\right| \\
&\leq CR^{-1}\|G^\varepsilon(\cdot, y_0) - G^\star(\cdot, y_0)\|_{\mathrm{L}^\infty(\mathrm{B})} + CR^{-\nu}\varepsilon^\nu \|\nabla_x G^\star(\cdot, y_0)\|_{\mathrm{L}^\infty(\mathrm{B})} \\
&\quad + CR^{1-\nu}\varepsilon^\nu \ln\left(2 + \varepsilon^{-1}\right)\left\|(\nabla_x)^2 G^\star(\cdot, y_0)\right\|_{\mathrm{L}^\infty(\mathrm{B})} \\
&\quad + CR^{1-\nu}\varepsilon^{\nu+\alpha}\left\|(\nabla_x)^2 G^\star(\cdot, y_0)\right\|_{\dot{\mathrm{C}}^{0,\alpha}(\mathrm{B})}.
\end{aligned}$$

Grâce au Théorème A.3.7,

$$\begin{aligned}
&\|\nabla_x G^\star(x_0,\cdot)\|_{\mathrm{L}^\infty(\mathrm{B})} \leq CR^{-d+1},\\
&\left\|(\nabla_x)^2 G^\star(x_0,\cdot)\right\|_{\mathrm{L}^\infty(\mathrm{B})} \leq CR^{-d},\\
&\left\|(\nabla_x)^2 G^\star(x_0,\cdot)\right\|_{\dot{\mathrm{C}}^{0,\alpha}(\mathrm{B})} \leq CR^{-d-\alpha}.
\end{aligned}$$

En outre, grâce au Théorème 2.1.9, l'estimation suivante est satisfaite :

$$\|G^\varepsilon(\cdot,y_0) - G^\star(\cdot,y_0)\|_{\mathrm{L}^\infty(\mathrm{B})} \leq C\varepsilon^\nu R^{2-d-\nu}.$$

Par conséquent,

$$\begin{aligned}
&\left|\partial_{x_i} G^\varepsilon(x_0,y_0) - \sum_{j=1}^d \left\{\delta_{ij} + \partial_i w_j\left(\frac{x_0}{\varepsilon}\right)\right\} \partial_{x_j} G^\star(x_0,y_0)\right|\\
&\leq C\varepsilon^\nu R^{1-d-\nu} \ln\left(2+\varepsilon^{-1}\right),
\end{aligned}$$

d'où (2.44). □

Nous sommes donc en mesure de montrer la dernière inégalité (2.46), qui repose sur le Lemme 2.6.3 et sur (2.45).

*Démonstration de l'estimation* (2.46). Soit $x_0 \neq y_0 \in \Omega_1$. Si $|x_0 - y_0| \leq \varepsilon$, le Théorème 2.1.5 donne le résultat souhaité. On se restreint donc à $|x_0 - y_0| > \varepsilon$. Nous avons précédemment démontré que (2.45) est vérifiée. De même que précédemment, on pose

$$R := \frac{\min\left(\mathrm{d}(x_0,\partial\Omega), |x_0-y_0|\right)}{8}. \tag{2.268}$$

Remarquons qu'alors (2.267) est satisfaite.

Définissons

$$\begin{aligned}
u^\varepsilon(x) &:= \partial_{y_j} G^\varepsilon(x,y_0),\\
u^\star(x) &:= \sum_{l=1}^d \partial_{y_l} G^\star(x,y_0)\left(\delta_{lj} + \partial_j w_l^T\left(\frac{y_0}{\varepsilon}\right)\right).
\end{aligned}$$

Alors,sur $\mathrm{B} := \mathrm{B}(x_0,R)$,

$$-\mathrm{div}\left(A\left(\frac{x}{\varepsilon}\right)\cdot\nabla u^\varepsilon(x)\right) = -\mathrm{div}\left(A^\star\cdot\nabla u^\star(x)\right) = 0.$$

On peut donc appliquer le Lemme 2.6.3, grâce auquel

$$\begin{aligned}
&\left|\nabla u^\varepsilon(x_0) - \nabla u^\star(x_0) - \sum_{k=1}^d \nabla w_k\left(\frac{x_0}{\varepsilon}\right)\partial_k u^\star(x_0)\right|\\
&\leq CR^{-1}\|u^\varepsilon - u^\star\|_{\mathrm{L}^\infty(\mathrm{B})} + C\varepsilon^\nu R^{1-\nu}\ln\left(2+\varepsilon^{-1}\right)\left\|\nabla^2 u^\star\right\|_{\mathrm{L}^\infty(\mathrm{B})}\\
&\quad + CR^{-\nu}\varepsilon^\nu\|\nabla u^\star\|_{\mathrm{L}^\infty(\mathrm{B})} + CR^{1-\nu}\varepsilon^{\nu+\alpha}\left\|\nabla^2 u^\star\right\|_{\dot{\mathrm{C}}^{0,\alpha}(\mathrm{B})}.
\end{aligned}$$

Or, grâce à (2.45), on a,

$$\|u^\varepsilon - u^\star\|_{\mathrm{L}^\infty(\mathrm{B})} \leq C\varepsilon^\nu R^{1-\nu-d} \ln\left(2+\varepsilon^{-1}\right).$$

En outre, grâce au Théorème A.3.7 et comme $\delta_{lj} + \partial_j w_l\left(\frac{y_0}{\varepsilon}\right)$ est majorée indépendamment de $\varepsilon$ et $y_0$ (voir la Proposition 2.3.3), $u^\star$ satisfait

$$\|\nabla u^\star\|_{\mathrm{L}^\infty(\mathrm{B})} \leq CR^{-d}, \tag{2.269}$$

$$\left\|\nabla^2 u^\star\right\|_{\mathrm{L}^\infty(\mathrm{B})} \leq CR^{-d-1}, \tag{2.270}$$

$$\left\|\nabla^2 u^\star\right\|_{\dot{\mathrm{C}}^{0,\alpha}(\mathrm{B})} \leq CR^{-d-1-\alpha}. \tag{2.271}$$

Ainsi

$$\left|\nabla u^\varepsilon(x_0) - \nabla u^\star(x_0) - \sum_{k=1}^{d} \nabla w_k\left(\frac{x_0}{\varepsilon}\right)\partial_k u^\star(x_0)\right|$$
$$\leq C\varepsilon^\nu \ln\left(2+\varepsilon^{-1}\right) R^{-d-\nu},$$

c'est à dire (2.46). □

## 2.7 Deux cas d'application

Dans cette section, nous appliquons les résultats précédents à deux cas particuliers : le cas où le champ de matrices $A$ est périodique perturbé par un défaut, et le cas où il est quasi-périodique (avec une certaine restriction diophantienne).

*Remarque* 39 (Cas très restreint d'interfaces). On peut aussi appliquer nos résultats au cas d'une interface $A$ entre deux milieux périodiques $A(x\cdot e_1 > 0) = A_+(x)$ et $A(x\cdot e_1 < 0) = A_-$ le long de l'interface $\{0\}\times\mathbb{R}^{d-1}$ si :

(i) $A$ est uniformément elliptique et bornée,
(ii) $A_+$ et $A_-$ sont uniformément $\alpha$-hölderiennes,
(iii) $A_+$ et $A_-$ sont périodiques de périodes commensurables dans $\mathbb{Q}$ (voir [28, Th. 5.1]),
(iv) les deux matrices homogénéisées relatives à $A_+$ et $A_-$ sont identiques,
(v) la matrice $A$ est uniformément hölderienne au travers de l'interface $\{0\}\times\mathbb{R}^{d-1}$.

En effet, dans ce cas, grâce à [28, Th. 5.1], il existe des correcteurs $w_i$ bornés (on montre de même qu'il existe un potentiel $B$ borné) : on obtient donc les conclusions des Théorèmes 1.1.1, 1.1.3 et 1.1.4 (où en remplace $\nu_r$ par 1). Autant les hypothèses (i), (ii) et (iii) (voire éventuellement l'hypothèse (v)) sont raisonnables, autant l'hypothèse (iv) est extrêmement restrictive : elle n'est pas satisfaite pour une interface générique (hors cas très particuliers). C'est pourquoi nous ne nous détaillerons pas ce cas.

### 2.7.1 Cas d'un coefficient périodique avec défaut

Le but de cette section est de démontrer que le cas périodique avec défaut rentre dans le cadre théorique établi plus haut.

La question de la construction de correcteurs associés à $A$ est traitée dans [28, Th. 3.1 & Th. 4.1] et [25, Th. A.2] :

**Théorème 2.7.1** (Voir [28] et [25]). *Soit $r \in [1, +\infty[$. Supposons que $A_{\mathrm{per}}$ et $\widetilde{A}$ satisfont* (2.47), (2.48) *et* (2.49). *Supposons que $A_{\mathrm{per}}$ est périodique, et posons $A = A_{\mathrm{per}} + \widetilde{A}$. Alors, il existe une solution $w_j$ au problème*

$$\begin{cases} -\operatorname{div}\left(A \cdot (e_j + \nabla w_j)\right) = 0 \qquad \text{dans} \quad \mathbb{R}^d, \\ \dfrac{|w_j(x)|}{1+|x|} \underset{|x|\to+\infty}{\to} 0, \end{cases} \tag{2.272}$$

*qui s'écrit $w_j = w_j^{\mathrm{per}} + \widetilde{w}_j$, où $w_j^{\mathrm{per}}$ est l'unique -à l'ajout d'une constante près- solution périodique de*

$$-\operatorname{div}\left(A_{\mathrm{per}}\left(e_j + \nabla w_j^{\mathrm{per}}\right)\right) = 0. \tag{2.273}$$

*Si $r > 1$, il existe $C > 0$ tel que*

$$\|\nabla \widetilde{w}_j\|_{\mathrm{L}^r(\mathbb{R}^d)} \leq C, \qquad \forall j \in [\![1, d]\!]. \tag{2.274}$$

*En outre, si $1 \leq r < d$, alors il existe $C > 0$ tel que*

$$\|\widetilde{w}_j\|_{\mathrm{L}^\infty(\mathbb{R}^d)} \leq C, \qquad \forall j \in [\![1, d]\!]. \tag{2.275}$$

*Remarque* 40. Les résultats de l'article [28] étaient initialement énoncés dans le cas de champs scalaires $A_{\mathrm{per}}$ et $\widetilde{A}$. Toutefois, ils ont été généralisés à des champs de matrices dans l'article [25].

**Définition 2.7.1** (Correcteurs). *A toute matrice $A = A_{\mathrm{per}} + \widetilde{A}$ satisfaisant les hypothèses du Théorème 2.7.1, on associe des correcteurs $w_j = w_j^{\mathrm{per}} + \widetilde{w}_j$ définis grâce au Théorème 2.7.1. Quitte à rajouter une constante à $w^{\mathrm{per}}$, on impose que $w_j(0) = 0$.*

Ce théorème a un corollaire évident en dimension $d \geq 2$ :

**Corollaire 2.7.2.** *Soit $d \geq 2$. Sous les hypothèses du Théorème 2.7.1, il existe une constante $C > 0$ telle que pour tout $\widetilde{r} \in [r, +\infty]$ :*

$$\|\nabla \widetilde{w}\|_{\mathrm{L}^{\widetilde{r}}(\mathbb{R}^d)} \leq C. \tag{2.276}$$

*Remarque* 41. Ce corollaire est issu du fait que $\nabla \widetilde{w} \in \mathrm{L}^\infty(\mathbb{R})$ (voir [25, Th. A.2]).

Les correcteurs vérifient la propriété de sous-linéarité forte suivante :

**Proposition 2.7.3.** *Soit $A = A_{\mathrm{per}} + \widetilde{A}$ satisfaisant les hypothèses du Théorème 2.7.1. Si $r \neq d$, alors il existe $C > 0$ telle que pour tout $j \in [\![1, d]\!]$ tel que pour tout $x$, $y \in \mathbb{R}^d$, on a*

$$|w(x) - w(y)| \leq C|x - y|^{1-\nu_r}, \tag{2.277}$$

*pour $\nu_r$ est défini par* (2.50).

*Démonstration.* Notons tout d'abord que $w^{\mathrm{per}}$ satisfait naturellement ces estimations. En effet, par [64, Cor. 8.36 p. 212], pour tout $j \in [\![1,d]\!]$, $w_j^{\mathrm{per}} \in \mathrm{C}^{1,\alpha}(Q)$. Il suffit donc de montrer (2.277) pour $\widetilde{w}$.

Si $r < d$, $\nu_r = 1$, l'estimation (2.277) provient directement du Théorème 2.7.1. Supposons donc que $r > d$. Par le Théorème de Morrey [59, Th. 4.10 p. 167],

$$\frac{|\widetilde{w}_i(x) - \widetilde{w}_i(y)|}{|x-y|^{1-\nu_r}} \leq \|\nabla \widetilde{w}_i\|_{\mathrm{L}^r(\mathbb{R}^d)}.$$

D'où (2.277) dans le cas où $r > d$. □

On peut à présent construire un potentiel pour $A$ (voir la Définition 2.1.1).

**Proposition 2.7.4.** *Soient $A = A_{\mathrm{per}} + \widetilde{A}$ satisfaisant les hypothèses du Théorème 2.7.1, $A^\star$ la matrice homogénéisée et $w_i$ les correcteurs associés à $A$. Il existe un potentiel $B_k^{ij}$ associé à (2.27). En outre, si $r \neq d$, il satisfait, pour tout $x, y \in \mathbb{R}^d$,*

$$|B(x) - B(y)| \leq C|x-y|^{1-\nu_r}, \tag{2.278}$$

*où $C$ ne dépend que de $A_{\mathrm{per}}$, $\widetilde{A}$ et $r$, et où $\nu_r$ est défini par (2.50).*

*Remarque* 42. Grâce au principe du maximum et à une estimation lipschitzienne, le potentiel exhibé est l'unique potentiel sous-linéaire à l'infini associé à (2.27) (à l'ajout d'une constante près).

*Remarque* 43. Les estimations (2.277) et (2.278) sont optimales pour $r < d$ (car le potentiel du cas périodique sature ces estimations). On observe que dès que le défaut est étalé (c'est à dire que $r$ est grand), le contrôle sur $w$ et sur $B$ se dégrade. Le cas $r = d$ n'est pas traité de façon optimale, puisqu'on est forcé de se ramener au cas $r > d$.

Pour démontrer la Proposition 2.7.4, on va séparer d'abord la partie périodique de (2.27), puis traiter le reste $\widetilde{w}$ grâce aux estimations (2.65) et (2.66).

*Démonstration.* D'après le Lemme A.3.11, il existe un potentiel périodique $B_{\mathrm{per}}$, associé à

$$M_k^i(x) := A_{ik}^\star - \sum_{j=1}^d \left(A_{\mathrm{per}}\right)_{ij}(x)\left(\delta_{jk} + \partial_j w_k(x)\right). \tag{2.279}$$

Ce potentiel est de classe $\mathrm{C}^{1,\alpha}$ (voir la Proposition 2.3.3). Donc, il est immédiat que, pour tous $x, y \in \mathbb{R}^d$,

$$|B_{\mathrm{per}}(x) - B_{\mathrm{per}}(y)| \leq C|x-y|^{1-\nu_r}. \tag{2.280}$$

Il reste donc à trouver $\widetilde{B}$ un potentiel associé à

$$\widetilde{M}_k^i(x) = \widetilde{A}(x)\left(\delta_{ik} + \partial_i w_k(x)\right) + A_{\mathrm{per}}(x)\,\partial_i \widetilde{w}_k(x). \tag{2.281}$$

Quand on l'aura construit, $B := B_{\mathrm{per}} + \widetilde{B}$ sera le potentiel recherché.

Quitte à le prendre plus grand, supposons que $r > 1$. Soit $p \in ]1, +\infty]$. Alors

$$\left\|\widetilde{M}\right\|_{\mathrm{L}^p\left(\mathbb{R}^d\right)} \leq C \left\|\nabla \widetilde{w}\right\|_{\mathrm{L}^p\left(\mathbb{R}^d\right)} + \left\|\widetilde{A}\right\|_{\mathrm{L}^p\left(\mathbb{R}^d\right)} \left\|\nabla w\right\|_{\mathrm{L}^\infty\left(\mathbb{R}^d\right)}. \tag{2.282}$$

Or, si $p \in [r, +\infty]$, alors, grâce au Théorème 2.7.1, on a, pour tout $j \in [\![1, d]\!]$, $\nabla \widetilde{w}_j \in \mathrm{L}^p\left(\mathbb{R}^d, \mathbb{R}^d\right)$ et $\widetilde{A} \in \mathrm{L}^p\left(\mathbb{R}^d\right)$. Donc, grâce à l'inégalité (2.65), il existe un potentiel $\widetilde{B} \in \mathrm{W}^{1,p}_{\mathrm{loc}}\left(\mathbb{R}^d, \mathbb{R}^{d^3}\right)$ associé à (2.281), et il existe une constante $C > 0$ telle que

$$\left\|\nabla \widetilde{B}\right\|_{\mathrm{L}^p\left(\mathbb{R}^d\right)} \leq C. \tag{2.283}$$

Si $r > d$, on choisit $p = r$ dans (2.283). Alors, par le Théorème de Morrey, on a, pour tous $x, y \in \mathbb{R}^d$

$$\left|\widetilde{B}(x) - \widetilde{B}(y)\right| \leq C|x - y|^{\nu_r}, \tag{2.284}$$

ce qui, joint à (2.280), implique (2.278) pour $r > d$.

Dans le cas où $r < d$, on prend $p_1 = r$ et $p_2 = 2d$ dans (2.282). Alors, en prenant $\rho = 1$ dans (2.66), on obtient

$$\left\|\widetilde{B}\right\|_{\mathrm{L}^\infty\left(\mathbb{R}^d\right)} \leq C.$$

D'où (2.278) pour $\nu_r = 1$, pour $r < d$. □

On peut maintenant démontrer la Proposition 2.1.13.

*Démonstration de la Proposition 2.1.13.* Par hypothèse, $A$ et $A_{\mathrm{per}}$ vérifient les Hypothèses 1 et 2. Il existe des correcteurs $w_i^{\mathrm{per}}$ relatifs à $A_{\mathrm{per}}$ satisfaisant l'Hypothèse 3. De même, par le Théorème 2.7.1, il existe des correcteurs $w_i$ relatifs à $A$ satisfaisant l'Hypothèse 3.

Par les Propositions 2.7.3 et 2.7.4, $A$ satisfait les Hypothèses 7 et 8 pour $\nu = \nu_r$. Alors, grâce au Corollaire 2.2.2, on en déduit que $A$ satisfait aussi les Hypothèses 4, 5, et 6. □

### 2.7.2 Cas d'un coefficient quasi-périodique

Dans cette section, nous démontrons que le cas d'un coefficient quasi-périodique elliptique et régulier rentre dans le cadre théorique bâti plus haut, sous des conditions d'incommensurabilité entre les quasi-périodes. Notons que les résultats ne sont pas nouveaux, et étaient déjà annoncés dans [11] (nous renvoyons aussi à [8] pour le cas de coefficients presque-périodiques).

On peut énoncer l'inégalité de de Gårding :

**Lemme 2.7.5** (Lemme 5.5 de [28])**.** *Soient $R$ et $S$ satisfaisant l'Hypothèse 9. Alors il existe $s \in [1, +\infty[$ et $C > 0$ tels que, pour toute fonction $\mathrm{Q}$-périodique $f^{\mathrm{per}}$,*

$$\left(\int_{\mathrm{Q}^2} \left|f^{\mathrm{per}}(x, y) - \fint_{\mathrm{Q}^2} f^{\mathrm{per}}\right|^2 \mathrm{d}x\mathrm{d}y\right)^{1/2} \leq C \left\|(R \odot \nabla_x + S \odot \nabla_y) f^{\mathrm{per}}\right\|_{\mathrm{H}^s(\mathrm{Q}^2)}. \tag{2.285}$$

**Proposition 2.7.6.** *Soient $A_{\rm per}$ satisfaisant (2.51) et l'Hypothèse 1, et $R$ et $S$ satisfaisant l'Hypothèse 9. On pose $A(x) = A_{\rm per}(R\odot x, S\odot x)$. Soit $f^{\rm per}\in \mathrm{C}^\infty_{\rm per}\left(\mathrm{Q}^2\right)$ de moyenne nulle. Alors l'équation suivante :*

$$-\mathrm{div}\left(A(x)\cdot\nabla u(x)\right) = f^{\rm per}(R\odot x, S\odot x) \quad \textit{dans} \quad \mathbb{R}^d, \tag{2.286}$$

*possède une solution $u$ quasi-périodique. Celle-ci s'écrit*

$$u(x) = u^{\rm per}(R\odot x, S\odot x),$$

*où $u^{\rm per}\in \mathrm{C}^\infty(\mathrm{Q}^2)$ est $\mathrm{Q}^2$-périodique de moyenne nulle. Cette solution est, à l'ajout d'une constante près, l'unique solution de (2.286) parmi les fonctions quasi-périodiques bornées.*

La preuve de cette Proposition est une généralisation immédiate de la preuve de [28, Th. 5.8], dont l'énoncé est ci-dessous :

**Théorème 2.7.7** (Théorème 5.8 de [28])**.** *Soient $A_{\rm per}$ satisfaisant (2.51) et l'Hypothèse 1, et $R$ et $S$ satisfaisant l'Hypothèse 9. On pose $A(x) = A_{\rm per}(R\odot x, S\odot x)$. Alors, pour tout $j\in[\![1,d]\!]$, il existe une solution $w_j\in \mathrm{C}^\infty\left(\mathbb{R}^d\right)$ à l'équation suivante :*

$$\begin{cases} -\mathrm{div}\left(A(x)\cdot(\nabla w_j(x)+e_j)\right) = 0 \qquad \textit{dans} \quad \mathbb{R}^d, \\ \dfrac{|w(x)|}{1+|x|} \underset{|x|\to+\infty}{\longrightarrow} 0, \end{cases} \tag{2.287}$$

*qui s'exprime comme*

$$w_j(x) = w_j^{\rm per}(R\odot x, S\odot x), \tag{2.288}$$

*où $w_j^{\rm per}\in \mathrm{C}^\infty_{\rm per}(\mathrm{Q}^2)$.*

Désormais, on appelle de tels $w_j$ des correcteurs quasi-périodiques associés à $A$. Remarquons que ces correcteurs permettent de donner une formule explicite à la matrice homogénéisée $A^\star$ relative à $A$ :

$$A^\star_{ik} = \int_{\mathrm{Q}^2}\sum_{j=1}^d \left(A_{\rm per}\right)_{ij}(x,y)\left(\delta_{jk}+\partial_j w_k^{\rm per}(x,y)\right)\mathrm{d}x\mathrm{d}y. \tag{2.289}$$

*Remarque* 44. Si on ne fait pas d'autre hypothèse sur $R$ et $S$ que l'irrationalité des tous les rapports $R_i/S_i$ (ils peuvent éventuellement être des nombres de Liouville-Roth), le résultat [28, Th. 5.8] permet de construire des correcteurs $w_j$ dont les *gradients* $\nabla w_j$ sont quasi-périodiques et à moyenne nulle -en revanche, on ne sait pas si les correcteurs eux-mêmes sont quasi-périodiques bornés. Dans ce cas, on peut malgré tout définir $A^\star$, et appliquer les résultats des Sections 2.4 et 2.5.

On peut ensuite construire un potentiel associé à $A$ :

**Proposition 2.7.8.** *Soient $A_{\mathrm{per}}$ satisfaisant* (2.51) *et l'Hypothèse 1, et $R$ et $S$ satisfaisant l'Hypothèse 9. On pose $A(x) = A_{\mathrm{per}}(R\odot x, S\odot x)$, et $w_j$ des correcteurs quasi-périodiques bornés associés. Soient ensuite $A^\star$ la matrice homogénéisée relative à $A$ et $M_k^i$ défini par* (2.27). *Alors le potentiel $B$ défini par*

$$B_k^{ij}(x) = \int_{\mathbb{R}^d} \Big(\partial_j \mathcal{G}_\Delta(x-y) M_k^i(y) - \partial_i \mathcal{G}_\Delta(x-y) M_k^j(y)\Big)\,\mathrm{d}y$$

*est quasi-périodique et s'exprime comme*

$$B(x) = b_{\mathrm{per}}(R\odot x, S\odot x),$$

*où* $b_{\mathrm{per}} \in \mathrm{C}^\infty_{\mathrm{per}}\left(\mathrm{Q}^2, \mathbb{R}^{d^3}\right)$.

*Remarque* 45. La fonction $b_{\mathrm{per}}$ dans le Théorème 2.7.8 ci-dessus n'est pas le potentiel relatif à $A_{\mathrm{per}}$ (d'où notre choix typographique).

Nous procédons à la démonstration du Théorème 2.7.8.

*Démonstration du Théorème 2.7.8.* Cherchons à résoudre

$$-\Delta N_k^i = M_k^i. \tag{2.290}$$

Par (2.289), et comme $A_{\mathrm{per}}$ et $w^{\mathrm{per}}$ sont de classe $\mathrm{C}^\infty_{\mathrm{per}}(\mathrm{Q}^2)$, on peut appliquer la Proposition 2.7.6, d'où l'existence de $N_{\mathrm{per}} \in \mathrm{C}^\infty_{\mathrm{per}}(\mathrm{Q}^2)$ tel que

$$N(x) = N_{\mathrm{per}}(R\odot x, S\odot x)$$

est solution de (2.290). On pose alors

$$(b_{\mathrm{per}})_k^{ij}(x,y) = (R_i\partial_{x_i} + S_i\partial_{y_i})\,(N_{\mathrm{per}})_k^j(x,y) - \left(R_j\partial_{x_j} + S_j\partial_{y_j}\right)(N_{\mathrm{per}})_k^i(x,y).$$

Ainsi, il apparaît que $B(x) := b_{\mathrm{per}}(R\odot x, S\odot x)$ est le potentiel recherché, car régulier, borné, de moyenne nulle et satisfaisant

$$-\Delta B_k^{ij} = \partial_j M_k^i(x) - \partial_i M_k^j(x).$$

□

La Proposition 2.1.14 découle du Théorème 2.7.7 et de la Proposition 2.7.8.

# Chapitre 3

# Fonctions de Green d'un problème d'homogénéisation périodique

Ce chapitre reproduit une prépublication en anglais [86].

Nous y étudions les fonctions de Green périodiques $G_n$ d'opérateurs elliptiques multi-échelles $Lu = -\mathrm{div}\,(A(n\cdot)\cdot\nabla u)$, où $A$ est un champ $\mathbb{Z}^d$-périodique de matrices coercives et hölderiennes, et $n$ est un entier arbitrairement grand. Nous y démontrons des estimations locales sur ces fonctions $G_n(x,y)$, ainsi que leur gradients $\nabla_x G_n(x,y)$ et $\nabla_y G_n(x,y)$ et leur gradient croisé $\nabla_x\nabla_y G_n(x,y)$. Ces résultats sont valides en dimension $d \geq 2$, et s'appliquent aussi à des systèmes. En outre, nous construisons une décomposition de telles fonctions Green à partir de la solution fondamentale, dans le cas où la matrice homogénéisée est symétrique.

Cette étude a été réalisée à partir de suggestions de Frédéric Legoll et Claude Le Bris.

# Decomposition and pointwise estimates of periodic Green functions of some elliptic equations with periodic oscillatory coefficients

Marc Josien

**Abstract** This article is about the $\mathbb{Z}^d$-periodic Green function $G_n(x,y)$ of the multiscale elliptic operator $Lu = -\mathrm{div}\,(A(n\cdot)\cdot\nabla u)$, where $A(x)$ is a $\mathbb{Z}^d$-periodic, coercive, and Hölder continuous matrix, and $n$ is a large integer. We prove here pointwise estimates on $G_n(x,y)$, $\nabla_x G_n(x,y)$, $\nabla_y G_n(x,y)$ and $\nabla_x\nabla_y G_n(x,y)$ in dimensions $d \geq 2$. Moreover, we derive an explicit decomposition of this Green function, which is of independent interest. These results also apply for systems.

**Keywords :** Green function, periodic homogenization, multiscale problems.

## 3.1 Introduction

In this article, we consider the periodic Green function $G_n(x,y)$ associated with the multiscale problem

$$\begin{cases} -\mathrm{div}\,(A(nx)\cdot\nabla u_n(x)) = f(x) - \displaystyle\int_{\mathrm{Q}} f \qquad \text{for} \quad x \in \mathbb{R}^d, \\ \displaystyle\int_{\mathrm{Q}} u_n = 0 \quad \text{and} \quad u_n \text{ is Q-periodic,} \end{cases} \tag{3.1}$$

where $n \in \mathbb{N}$ is expected to be very large and $\mathrm{Q} = [-1/2, 1/2]^d$ is the unit cube in dimensions $d \geq 2$. Hereafter, we write "periodic" for "Q-periodic". Here $A$ satisfies the classical assumptions of ellipticity, periodicity and Hölder continuity (see [11]). Our purpose is to derive pointwise estimates for $G_n$ and its derivatives $\nabla_x G_n$, $\nabla_y G_n$ and $\nabla_x\nabla_y G_n$. In specific cases, we also express the periodic Green function $G_n$ in terms of the Green function of the operator $-\mathrm{div}\,(A(n\cdot)\cdot\nabla)$ in $\mathbb{R}^d$. We motivate this study by the fact that, although some results that we show here are believed to be true and the theoretical material we use is now classical, these results have not been, to the best of our knowledge, clearly established. In particular, we refer the reader to [29], which collects several similar results, but concerning the Green function of elliptic problems with periodic coefficients in $\mathbb{R}^d$ (and not the periodic Green function).

Estimating the behavior of the Green function of elliptic problems has attracted much attention, as the Green functions are a useful tool for getting estimates. Indeed, the solution $u_n$ to (3.1) can be written as an integral of the forcing term (*e.g.*, $f$ in (3.1), which can be in the form $f = \mathrm{div}(H)$), multiplied by the Green function. Thus, if the Green function (or its derivatives) is controlled, then one can estimate the solution $u_n$ (or its derivatives) directly from the forcing term, using the Young inequality (see, *e.g.*, [21, Chap. I p. 7-12], and see [94] for such manipulations). However, let us remark that, by duality, such estimates

can also be used to get back to the properties of the Green function. We refer the reader to [94], which goes back and forth from properties of the Green function to estimates on the solution to the oscillating problem.

The behavior of the Green function $G$ of the following Dirichlet problem :

$$\begin{cases} -\operatorname{div}(A(x)\cdot\nabla u(x)) = f(x) & \text{in} \quad \Omega, \\ u = 0 & \text{on} \quad \partial\Omega, \end{cases}$$

with elliptic and bounded matrix $A$ has been explored in the seminal article [72] (here $\Omega$ is a sufficiently regular bounded domain). Without any regularity assumption on $A$ (and without any hypothesis about the structure of $A$), the authors derive optimal pointwise estimates on $G$. But they need to assume that the matrix $A$ is continuous and sufficiently regular in order to obtain pointwise estimates on the gradients $\nabla_x G$, $\nabla_y G$, and on the second derivatives $\nabla_x \nabla_y G$. Loosely speaking, they show under suitable assumptions that these quantities behave as if $G$ was the fundamental solution to the Laplace equation (see [64, Chap. II, p. 13-30]), namely (in dimension $d \geq 3$) :

$$|G(x,y)| \leq C|x-y|^{-d+2}, \tag{3.2}$$

$$|\nabla_x G(x,y)| \leq C|x-y|^{-d+1}, \qquad |\nabla_y G(x,y)| \leq C|x-y|^{-d+1}, \tag{3.3}$$

$$|\nabla_x \nabla_y G(x,y)| \leq C|x-y|^{-d}. \tag{3.4}$$

Since the domain of interest $\Omega$ is bounded, the above quantity $|x-y|$ is bounded ; hence, the difficulty in the above estimates is obviously when $x$ is close to $y$. Their results have been generalized to systems of elliptic equations. In particular, the same type of results is proved in [55], provided that the matrix $A$ is sufficiently regular.

On the opposite side, problems like (3.1) have the specificity that the coefficients $A(n\cdot)$ are more and more oscillating when $n$ increases, since the caracteristic scale $1/n$ of the microstructure goes smaller and smaller. Therefore, the results that rely on the regularity of the coefficient do not apply *uniformly* : the constants $C$ of the estimates (3.3) and (3.4) blows up when $n$ goes to infinity. With a totally different approach than above, Avellaneda and Lin have proved that the solutions to oscillatory elliptic problems enjoy Hölder and Lipschitz regularity properties, if the matrix $A$ is elliptic, periodic, and Hölder continuous (see [11]). For that purpose, they introduced a so-called compactness method, showing that the oscillatory problems inherit regularity from the homogenized problem. Applying their results to the Green function in $\mathbb{R}^d$ itself, they derived the same type of estimates as (3.2), (3.3), and (3.4). We refer the reader to [29] for a review on the pointwise estimates on multiscale Green functions in $\mathbb{R}^d$, for matrices $A$ that are elliptic, bounded, periodic and sufficiently regular.

In [12], Avellaneda and Lin described the asymptotic behavior, in the limit where the small scale vanishes, of the Green function in $\mathbb{R}^d$ of periodic elliptic equations by using the Green function of the homogenized problem. Using the same techniques, the authors of [93, 94] established the same kind of asymptotics for the Green function of the multiscale problem set in a bounded domain, for Dirichlet and Neumann boundary conditions.

Periodic Green functions can sometimes be expressed thanks to the associated Green functions in the whole space $\mathbb{R}^d$. For example, such a decomposition can be found in [42] for the case of the Laplacian. This consists in a series involving the Green function in $\mathbb{R}^d$, translated on the grid $\mathbb{Z}^d$. The main difficulty of this decomposition is to ensure that the series actually converges ; in the case of the Laplacian, this is shown by resorting to the local symmetries of the Green function of the Laplacian. We address the question of building a similar decomposition for the case (3.1).

Most of the theoretical material and ideas used in the present article are borrowed from [11, 12, 29, 94] for the homogenization aspect, and from [42] for the decomposition of the periodic Green function.

### 3.1.1 Main results

Before getting to the oscillatory problem, we first establish the existence and the uniqueness of the periodic Green function for general periodic, elliptic and bounded coefficients. Henceforth, we denote by a subscript "per" the functional spaces of periodic functions : for example, $\mathrm{L}^2_{\mathrm{per}}(\mathrm{Q})$ is the set of functions defined on $\mathbb{R}^d$ that are periodic and square integrable on the cube Q. We consider the operator

$$T : f \mapsto u,$$

where $f \in \mathrm{L}^2_{\mathrm{per}}(\mathrm{Q})$ and $u$ is the unique periodic solution with zero mean to

$$-\mathrm{div}\left(A(x)\cdot\nabla u(x)\right) = f(x) - \int_{\mathrm{Q}} f \quad \text{for} \quad x \in \mathbb{R}^d, \tag{3.5}$$

where $A$ is periodic, elliptic and bounded. Namely, there exists $\mu > 0$ such that $A$ satisfies

$$\mu|\xi|^2 \leq A(x)\cdot\xi\cdot\xi \leq \mu^{-1}|\xi|^2 \qquad \forall x, \xi \in \mathbb{R}^d, \tag{3.6}$$

$$A(x+z) = A(x) \qquad \forall x \in \mathbb{R}^d, z \in \mathbb{Z}^d. \tag{3.7}$$

The operator $T$ admits the following integral formulation, involving the so-called periodic Green function $G$ associated with the operator $-\mathrm{div}\left(A\cdot\nabla\right)$ :

$$Tf(x) = \int_{\mathrm{Q}} G(x,y)f(y)\mathrm{d}y. \tag{3.8}$$

By classical arguments (see, *e.g.*, [72]), we first justify the existence and the uniqueness of the Green function $G$. It satisfies the following equation :

$$-\mathrm{div}_x\left(A(x)\cdot\nabla_x G(x,y)\right) = \delta_y(x) - 1 \quad \text{in} \quad \mathrm{Q}, \tag{3.9}$$

with periodic boundary conditions and lies in the functional space $E$ containing all the functions $G(x,y)$ satisfying, for all $p \in \left[1, \frac{d}{d-2}\right)$ (by convention, if $d = 2$, then $d/(d-2) = +\infty$) and $q \in \left[1, \frac{d}{d-1}\right)$,

$$\sup_{y\in\mathrm{Q}} \|G(\cdot,y)\|_{\mathrm{L}^p(\mathrm{Q})} < +\infty \qquad \sup_{y\in\mathrm{Q}} \|\nabla_x G(\cdot,y)\|_{\mathrm{L}^q(\mathrm{Q})} < +\infty, \tag{3.10}$$

$$\sup_{x\in\mathrm{Q}} \|G(x,\cdot)\|_{\mathrm{L}^p(\mathrm{Q})} < +\infty, \qquad \sup_{x\in\mathrm{Q}} \|\nabla_y G(x,\cdot)\|_{\mathrm{L}^q(\mathrm{Q})} < +\infty. \tag{3.11}$$

**Proposition 3.1.1.** *Let the dimension be $d \geq 2$. Assume that $A \in \mathrm{L}^\infty_{\mathrm{per}}\left(\mathrm{Q}, \mathbb{R}^{d^2}\right)$ satisfies* (3.6) *and* (3.7)*. Then there exists a unique periodic Green function $G(x,y)$ associated with the operator $-\mathrm{div}\,(A\cdot\nabla)$ -namely, $G$ satisfies* (3.8)*- that is in the space $E$, defined by* (3.10) *and* (3.11)*. Moreover, the function $G^\dagger(x,y) := G(y,x)$ is the periodic Green function associated with the operator $-\mathrm{div}\left(A^T\cdot\nabla\right)$. Last, $G$ is the unique periodic solution in $E$ to* (3.9) *with periodic boundary conditions satisfying*

$$\int_{\mathrm{Q}} G(x,y)\mathrm{d}y = 0 \qquad \forall x \in \mathrm{Q}, \tag{3.12}$$

$$\textit{and} \quad \int_{\mathrm{Q}} G(x,y)\mathrm{d}x = 0 \qquad \forall y \in \mathrm{Q}. \tag{3.13}$$

The proof of Proposition 3.1.1 is postponed until Section 3.2.

Actually, if $d \geq 3$, the Green function is expected to satisfy :

$$\sup_{y\in\mathrm{Q}} \|G(\cdot,y)\|_{\mathrm{L}^{\frac{d}{d-2},\infty}(\mathrm{Q})} < +\infty, \qquad \sup_{y\in\mathrm{Q}} \|\nabla_x G(\cdot,y)\|_{\mathrm{L}^{\frac{d}{d-1},\infty}(\mathrm{Q})} < +\infty, \tag{3.14}$$

$$\sup_{x\in\mathrm{Q}} \|G(x,\cdot)\|_{\mathrm{L}^{\frac{d}{d-2},\infty}(\mathrm{Q})} < +\infty, \qquad \sup_{x\in\mathrm{Q}} \|\nabla_y G(x,\cdot)\|_{\mathrm{L}^{\frac{d}{d-1},\infty}(\mathrm{Q})} < +\infty, \tag{3.15}$$

when homogeneous Dirichlet boundary conditions are considered (see [72, Th. 1.1]). Here $\mathrm{L}^{p,\infty}$ denote the Marcinkiewicz spaces (see [21, Chap. I p. 7-11] for a reference on such functional spaces). Thus, the proposed space $E$ is not optimal. However, for the purpose of the present article, it not useful to find the optimal function space, since (3.14) and (3.15) are a straightforward corollary of Propositions 3.1.2 and 3.1.3 below.

Using a method that can be found in [11, Th. 13] (see also the proof of [94, Th. 3.3]), we show a pointwise estimate on the periodic Green function $G$ associated with the operator $-\mathrm{div}\,(A\cdot\nabla)$. In dimension $d = 2$, this estimate on $G(x,y)$ is logarithmic, which introduces some technicalities.

**Proposition 3.1.2.** *Let the dimension be $d \geq 2$. Assume that $A \in \mathrm{L}^\infty_{\mathrm{per}}\left(\mathrm{Q}, \mathbb{R}^{d^2}\right)$ satisfies* (3.6) *and* (3.7)*. Let $G$ be the periodic Green function associated with the operator $-\mathrm{div}\,(A\cdot\nabla)$. Then there exists a constant $C > 0$ that only depends on $d$ and $\mu$ such that the following estimates are satisfied, for all $x \in \mathbb{R}^d$ and $y \in x + \mathrm{Q}$, with $x \neq y$ :*

$$\textit{if } d \geq 3, \qquad |G(x,y)| \leq C|x-y|^{-d+2}, \tag{3.16}$$

$$\textit{if } d = 2, \qquad |G(x,y)| \leq C\log(2+|x-y|). \tag{3.17}$$

The proof of Proposition 3.1.2 is postponed until Section 3.3.

One can apply the above result to the multiscale problem (3.1). Thus, if $A \in \mathrm{L}^\infty_{\mathrm{per}}\left(\mathrm{Q}, \mathbb{R}^{d^2}\right)$ satisfies (3.6) and (3.7), then the periodic Green function $G_n$ associated with the operator $-\mathrm{div}\,(A(n\cdot)\cdot\nabla)$ satisfies

$$\text{if } d \geq 3, \qquad |G_n(x,y)| \leq C|x-y|^{-d+2}, \tag{3.18}$$

$$\text{if } d = 2, \qquad |G_n(x,y)| \leq C\log(2+|x-y|), \tag{3.19}$$

In (3.18) and (3.19), the constant $C$ only depends on $d$ and $\mu$.

Now, we consider a matrix $A$ that is elliptic, periodic, and also Hölder continuous :

$$A \in \mathrm{C}^{0,\alpha}\left(\mathrm{Q}, \mathbb{R}^{d^2}\right), \tag{3.20}$$

for $\alpha \in (0,1)$. Using the results of [11], we derive pointwise estimates on the gradients $\nabla_x G_n$ and $\nabla_y G_n$ and on the second derivatives $\nabla_x \nabla_y G_n$ of the periodic Green function $G_n$ associated with the operator $-\mathrm{div}\,(A(n\cdot)\cdot\nabla)$.

**Proposition 3.1.3.** *Let the dimension be $d \geq 2$. Assume that $A \in \mathrm{L}^\infty_{\mathrm{per}}\left(\mathrm{Q}, \mathbb{R}^{d^2}\right)$ satisfies* (3.6)*,* (3.7) *and* (3.20)*. Let $G_n$ be the periodic Green function associated with the operator $-\mathrm{div}\,(A(n\cdot)\cdot\nabla)$. Then, there exists a constant $C > 0$ such that, for all $n \in \mathbb{N}\backslash\{0\}$, $x \in \mathbb{R}^d$ and $y \in x + \mathrm{Q}$, with $x \neq y$,*

$$|\nabla_x G_n(x,y)| \leq C|x-y|^{-d+1}, \tag{3.21}$$

$$|\nabla_y G_n(x,y)| \leq C|x-y|^{-d+1}, \tag{3.22}$$

$$|\nabla_x \nabla_y G_n(x,y)| \leq C|x-y|^{-d}. \tag{3.23}$$

The proof of Proposition 3.1.3 is postponed until Section 3.4.

Let us underline that the salient point of Proposition 3.1.3 is that the constant $C$ does *not* depend on the characteristic scale $1/n$ of the microstructure. The latter estimates are not unexpected ; see, *e.g.*, [29, Prop. 8] for similar estimates on the Green function in the whole space $\mathbb{R}^d$.

On the first hand, as is shown in [72, Th. 1.1], in the case of Dirichlet boundary conditions, Estimate (3.18) does not require any regularity assumption on $A$. As expected, it is also the case for periodic boundary conditions. On the other hand, Estimates (3.21), (3.22) and (3.23) critically rely on the fact that $A$ is periodic and sufficiently regular.

Using another approach, reminiscent of [42, p. 130-131], we show a decomposition for the periodic Green function $G$. This formula extensively uses the corresponding Green function $\mathcal{G}$ in $\mathbb{R}^d$ of the operator $-\mathrm{div}\,(A\cdot\nabla)$, which satisfies

$$-\mathrm{div}\,(A(x)\cdot\nabla_x \mathcal{G}(x,y)) = \delta_y(x) \qquad \text{in} \quad \mathbb{R}^d. \tag{3.24}$$

**Proposition 3.1.4.** *Let the dimension be $d \geq 3$. Assume that $A \in \mathrm{L}^\infty_{\mathrm{per}}\left(\mathrm{Q}, \mathbb{R}^{d^2}\right)$ satisfies* (3.6)*,* (3.7) *and* (3.20)*. Let $G$ be the periodic Green function associated with the operator $-\mathrm{div}\,(A(\cdot)\cdot\nabla)$. Then, the function $G$ can be decomposed as*

$$G(x,y) = \sum_{m=0}^{+\infty}\left(\sum_{k\in\Gamma_m} H^k(x,y)\right), \tag{3.25}$$

*where the functions $H^k$ are defined by*

$$\begin{aligned} H^k(x,y) := \mathcal{G}(x,y-k) - \int_{\mathrm{Q}} \mathcal{G}(x,y+y'-k)\mathrm{d}y' - \int_{\mathrm{Q}} \mathcal{G}(x+x',y-k)\mathrm{d}x' \\ + \int_{\mathrm{Q}}\int_{\mathrm{Q}} \mathcal{G}(x+x',y+y'-k)\mathrm{d}y'\mathrm{d}x', \end{aligned} \tag{3.26}$$

*and the function* $\mathcal{G}$ *by* (3.24)*, and the sets* $\Gamma_m$ *by*

$$\Gamma_m = \left\{ k \in \mathbb{Z}^d, 2^m - 1 \leq k \cdot (A_{\rm s}^\star)^{-1} \cdot k < 2^{m+1} - 1 \right\}, , \tag{3.27}$$

*where* $A_{\rm s}^\star$ *is the symmetric part of the homogenized matrix* $A^\star$ *associated with the matrix* $A$*.*

The proof of Proposition 3.1.4 is postponed until Section 3.5.

The above decomposition (3.25) naturally appears as a reasonable candidate, being close (but not equivalent) to the decomposition [42, p. 130-131]. But the difficulty is to ensure that the series actually converges, in the sense that

$$\sum_{m=0}^{+\infty} \left| \sum_{k \in \Gamma_m} H^k(x,y) \right| < +\infty \quad \text{for} \quad x \neq y. \tag{3.28}$$

In [42, p. 130-131], where the Laplacian with periodic boundary conditions is studied, the convergence is obtained by appealing to the *local* symmetries of the Green function of the Laplacian. This cannot be applied to our case. Here, the convergence is a consequence of the *long-range* behavior of the Green function $\mathcal{G}$. Thanks to the periodicity of $A$, the function $\mathcal{G}$ can be efficiently approximated at large scale by the Green function of the homogenized problem (see [12, 94]). Hence, taking advantage of the long-range symmetries of the Green function of the homogenized problem, one can as well prove the convergence of the series in (3.25). In this regard, we underline that, in general, the series (3.25) does not converge absolutely with respect to $k$ :

$$\sum_{k \in \mathbb{Z}^d} \left| H^k(x,y) \right| = +\infty \quad \text{for} \quad x \neq y.$$

This fact appears as a byproduct of the proof.

Last but not least, it should be underlined that (3.18) is an obvious corollary of the proof of (3.25), when considering the Green function $G_n$ of the operator $-\mathrm{div}\,(A(n\cdot)\cdot\nabla)$. Also, it can be seen in the proof (in Section 3.5) that the constant $C$ in (3.25) does not depend on $n$. Hence, the above decomposition provides an alternative way for showing pointwise estimates on the periodic Green function $G_n$.

### 3.1.2 Extension to systems

Our proof of Proposition 3.1.1 concerning existence and uniqueness of the Green function uses the De Giorgi-Nash-Moser theorem. In dimension $d \geq 3$, this ingredient can be replaced by the $\mathrm{W}^{1,p}$ and $\mathrm{L}^\infty$ estimates in [55, Lem. 2 & Lem. 3]. Therefore, the conclusions of Proposition 3.1.1 also hold for the periodic Green function of the operator $Lu := (L^\alpha u)_{\alpha \in [\![1,m]\!]}$ defined by

$$L^\alpha u := -\mathrm{div} \left( \sum_{\beta=1}^m A^{\alpha\beta} \cdot \nabla u^\beta \right) = - \sum_{i,j=1}^d \partial_i \left( \sum_{\beta=1}^m A_{ij}^{\alpha\beta} \partial_j u^\beta \right),$$

where $A = \left(A_{ij}^{\alpha\beta}\right)$, for $i, j \in [\![1, d]\!]$ and $\alpha, \beta \in [\![1, m]\!]$, $m \in \mathbb{N}$, is *continuous*, periodic, and elliptic in the following sense :

$$\mu|\xi|^2 \leq \sum_{i,j=1}^{d} \sum_{\alpha,\beta=1}^{m} A_{ij}^{\alpha\beta}(x)\xi_i^\alpha \xi_j^\beta \leq \mu^{-1}|\xi|^2 \qquad \forall x \in \mathbb{R}^d, \xi = (\xi_i^\alpha) \in \mathbb{R}^{dm}, \tag{3.29}$$

In this case, the periodic Green function $G$ (which is a matrix) satisfies, for all $\alpha, \gamma \in [\![1, m]\!]$,

$$-\mathrm{div}_x \left( \sum_{\beta=1}^{m} A^{\alpha\beta}(x) \cdot \nabla_x G^{\beta\gamma}(x, y) \right) = \delta^{\alpha\gamma} \left(\delta_y(x) - 1\right) \quad \text{in} \quad \mathrm{Q}, \tag{3.30}$$

with periodic boundary conditions, where $\delta^{\alpha\beta}$ is the Kronecker symbol.

As can be seen in Sections 3.3 and 3.4, the proofs of Propositions 3.1.2 and 3.1.3 involve arguments that are also valid if we study periodic oscillatory systems instead of equations (note that the seminal article [11] dealt with systems). More precisely, if $d \geq 3$, the periodic Green function $G_n$ associated with the operator $-\mathrm{div}\left(A(n\cdot) \cdot \nabla\right)$ satisfies (3.18), (3.19), (3.21), (3.22), and (3.23), provided that $A = \left(A_{ij}^{\alpha\beta}\right)$ is periodic, satisfies (3.29), and is Hölder continuous. Notably, the Hölder estimate [11, Lem. 9] can be used instead of the De Giorgi-Nash-Moser theorem in the proof of Proposition 3.1.2. Besides, the Lipschitz estimate borrowed from [11, Lem. 16] that we use in the proof of Proposition 3.1.3 also applies.

Finally, the decomposition (3.25) can also be generalized to the case of systems, using appropriate sets $\Gamma_m^{\alpha,\beta}$ while decomposing $G^{\alpha\beta}$ (see Section 3.5.3).

### 3.1.3 Outline

Our article is articulated as follows. In Section 3.2, we prove by classical arguments that there exists a unique periodic Green function, and that it is the solution to (3.9). We thus establish Proposition 3.1.1. Next, in Section 3.3, we proceed with the proof of Proposition 3.1.2. In dimension $d \geq 3$, the proof is based on a duality argument involving the De Giorgi-Nash-Moser theorem. In dimension $d = 2$, using a trick from [11], it reduces to expressing the 2-dimensional periodic Green function as the integral of a 3-dimensional periodic Green function. In Section 3.4, combining Estimate (3.18) of Proposition 3.1.2 and the Lipschitz estimates of [11], we show (3.21) (and similarly (3.22)), from which we deduce (3.23). Finally, in Section 3.5, we prove Proposition 3.1.4, which, under suitable hypotheses, yields a decomposition for the periodic Green function. For the sake of simplicity, we first study the case where the homogenized matrix is the identity in Section 3.5.1, and then the general case in Section 3.5.2. Additional materials about such a decomposition in the case of systems can be found in the Section 3.5.3.

## 3.2 Existence, uniqueness and basic properties of the Green function

We prove Proposition 3.1.1 by standard arguments (see [72]). First, we establish existence and uniqueness by using the regularizing properties of (3.5). Then, we show that $(x, y) \mapsto G(y, x)$ is the Green function of the transposed problem by considering the adjoint operator of $T$. In a third step, we use the variational formulation of (3.5) and establish that the Green function $G$ satisfies (3.9), (3.12) and (3.13). Finally, we show the uniqueness of the solution to (3.9), (3.12) and (3.13), using a variational argument.

*Proof.* The proof falls in four steps.

**Step 1 : Existence and uniqueness** Since $A$ is elliptic, the Lax-Milgram theorem yields that $T$ is continuous from $\mathrm{L}^2_{\mathrm{per}}(\mathrm{Q})$ to $\mathrm{L}^2_{\mathrm{per}}(\mathrm{Q})$. Moreover, if $f \in \mathrm{L}^p_{\mathrm{per}}(\mathrm{Q})$ for $p > d/2$, then, by the De Giorgi-Nash-Moser theorem (see [64, Th. 8.24 p. 202]) there holds

$$\|u\|_{\mathrm{L}^\infty(\mathrm{Q})} \leq C \|f\|_{\mathrm{L}^p(\mathrm{Q})}.$$

Therefore, by duality, there exists a unique $G(x, \cdot)$ such that, for all $x \in \mathrm{Q}$, $u(x) = Tf(x)$ can be represented as

$$u(x) = \int_{\mathrm{Q}} G(x, y) f(y) \mathrm{d}y, \tag{3.31}$$

and, for any $p' \in [1, d/(d-2))$, there holds

$$\sup_{x \in \mathrm{Q}} \|G(x, \cdot)\|_{\mathrm{L}^{p'}(\mathrm{Q})} \leq C(p').$$

Clearly, $G$ can be defined as a periodic function of $x$ and $y$. Hence, (3.31) can be generalized as

$$Tf(x) = \int_{y_0 + Q} G(x, y) f(y) \mathrm{d}y, \tag{3.32}$$

the latter formula being true for all $x, y_0 \in \mathbb{R}^d$.

Let us now consider a regular function $\phi \in \mathrm{C}^\infty_{\mathrm{per}}\left(\mathrm{Q}, \mathbb{R}^d\right)$, and $u = T\mathrm{div}(\phi)$. Once more, by [64, Th. 8.24 p. 202], for all $q > d$, there holds

$$\|u\|_{\mathrm{L}^\infty(\mathrm{Q})} \leq C \|u\|_{\mathrm{L}^2(\mathrm{Q})} + C \|\phi\|_{\mathrm{L}^q(\mathrm{Q})} \leq C \|\phi\|_{\mathrm{L}^q(\mathrm{Q})}. \tag{3.33}$$

Yet, $u$ can also be represented as

$$u(x) = \int_{\mathrm{Q}} G(x, y) \mathrm{div}(\phi(y)) \mathrm{d}y,$$

and, by the theory of distributions, as

$$u(x) = -\left\langle \nabla_y G(x,\cdot), \phi \right\rangle_{\left[\mathrm{C}^\infty_{\mathrm{per}}(Q)\right]',\mathrm{C}^\infty_{\mathrm{per}}(Q)}.$$

By (3.33), we deduce that $\nabla_y G(x,\cdot) \in \mathrm{L}^{q'}\left(\mathrm{Q},\mathbb{R}^d\right)$, and finally that

$$\sup_{x\in\mathrm{Q}} \|\nabla_y G(x,\cdot)\|_{\mathrm{L}^{q'}(\mathrm{Q})} \leq C(q'),$$

for any $q' \in [1, d/(d-1))$.

So far, we have established that there exists a unique Green function $G$, and that this function satisfies (3.11).

**Step 2 : Transposition** The proof given here relies on an argument involving the kernels of adjoint operators. We define the operator

$$T^\dagger : f \mapsto u, \tag{3.34}$$

where $f \in \mathrm{L}^2_{\mathrm{per}}(\mathrm{Q})$ and $u$ is the periodic solution to

$$-\mathrm{div}\left(A^T \cdot \nabla u\right) = f - \int_{\mathrm{Q}} f,$$

with zero mean. We denote by $G^\dagger$ the Green function of $T^\dagger$.

We claim that $T : \mathrm{L}^2_{\mathrm{per}}(\mathrm{Q}) \to \mathrm{L}^2_{\mathrm{per}}(\mathrm{Q})$ and $T^\dagger : \mathrm{L}^2_{\mathrm{per}}(\mathrm{Q}) \to \mathrm{L}^2_{\mathrm{per}}(\mathrm{Q})$ are adjoint of each other. Indeed, for all $f, g \in \mathrm{L}^2_{\mathrm{per}}(\mathrm{Q})$, if $u = Tf$ and $v = T^\dagger g$, we have

$$\begin{aligned}
\int_{\mathrm{Q}} g(x)u(x)\mathrm{d}x &= \int_{\mathrm{Q}} A^T(x)\cdot\nabla v(x)\cdot\nabla u(x)\mathrm{d}x \\
&= \int_{\mathrm{Q}} A(x)\cdot\nabla u(x)\cdot\nabla v(x)\mathrm{d}x \\
&= \int_{\mathrm{Q}} f(x)v(x)\mathrm{d}x.
\end{aligned}$$

In other words, for any $f, g \in \mathrm{L}^2_{\mathrm{per}}(\mathrm{Q})$,

$$\int_{\mathrm{Q}} Tf(x)g(x)\mathrm{d}x = \int_{\mathrm{Q}} f(x)T^\dagger g(x)\mathrm{d}x$$

which is equivalent to

$$\int_{\mathrm{Q}^2} f(y)G(x,y)g(x)\mathrm{d}x\mathrm{d}y = \int_{\mathrm{Q}^2} f(x)G^\dagger(x,y)g(y)\mathrm{d}x\mathrm{d}y.$$

Therefore, $G^\dagger(x,y) = G(y,x)$. As a consequence, $G$ also satisfies (3.10). Hence, $G \in E$.

**Step 3 : Partial differential equation satisfied by $G$** Let $\phi$ and $\psi$ be two periodic smooth functions. By definition of $G$,

$$\begin{aligned}&\int_{\mathrm{Q}} A(x)\cdot\nabla_x\left(\int_{\mathrm{Q}} G(x,y)\psi(y)\mathrm{d}y\right)\cdot\nabla\phi(x)\mathrm{d}x\\&=\int_{\mathrm{Q}}\psi(y)\phi(y)\mathrm{d}y-\int_{\mathrm{Q}}\psi(y)\mathrm{d}y\int_{\mathrm{Q}}\phi(x)\mathrm{d}x.\end{aligned}$$

The above identity can also be written in the sense of distributions as (the inversion of integrals is justified since $G$ satisfies (3.10))

$$\begin{aligned}&\left\langle\left\{\int_{\mathrm{Q}} A(x)\cdot\nabla_x\left(G(x,\cdot)\right)\cdot\nabla\phi(x)\mathrm{d}x\right\},\psi\right\rangle_{\left[\mathrm{C}^\infty_{\mathrm{per}}(\mathrm{Q})\right]',\mathrm{C}^\infty_{\mathrm{per}}(\mathrm{Q})}\\&=\left\langle\phi-\int_{\mathrm{Q}}\phi,\psi\right\rangle_{\left[\mathrm{C}^\infty_{\mathrm{per}}(\mathrm{Q})\right]',\mathrm{C}^\infty_{\mathrm{per}}(\mathrm{Q})}.\end{aligned}$$

Therefore, there holds

$$\int_{\mathrm{Q}} A(x)\cdot\nabla_x\left(G(x,y)\right)\cdot\nabla\phi(x)\mathrm{d}x=\phi(y)-\int_{\mathrm{Q}}\phi(z)\mathrm{d}z,$$

which implies that, in the sense of distributions,

$$-\mathrm{div}\left(A(x)\cdot\nabla G(x,y)\right)=\delta_y(x)-1.$$

Thus, the function $G$ satisfies (3.9). Since $T1=0$, where 1 is the function which is identically equal to 1, the function $G$ satisfies (3.12). Moreover, this latter property, written for the Green function associated with $-\mathrm{div}\left(A^T\cdot\nabla\right)$, implies that $G$ also satisfies (3.13).

**Step 4 : Equivalence of formulations** Assume that $\widetilde{G}\in E$ satisfies (3.9), (3.12) and (3.13). Let $\phi,\psi\in\mathrm{C}^\infty_{\mathrm{per}}(\mathrm{Q})$. Then, by definition of $\widetilde{G}$, there holds

$$\int_{\mathrm{Q}} A(x)\cdot\nabla_x\widetilde{G}(x,y)\cdot\nabla\phi(x)\mathrm{d}x=\phi(y)-\int_{\mathrm{Q}}\phi.$$

Testing the above equation against $\psi$, we obtain

$$\int_{\mathrm{Q}}\int_{\mathrm{Q}} A(x)\cdot\nabla_x\widetilde{G}(x,y)\cdot\nabla\phi(x)\psi(y)\mathrm{d}x\mathrm{d}y=\int_{\mathrm{Q}}\psi(y)\left(\phi(y)-\int_{\mathrm{Q}}\phi\right)\mathrm{d}y,$$

or, equivalently,

$$\int_{\mathrm{Q}} A(x)\cdot\nabla_x\left(\int_{\mathrm{Q}}\widetilde{G}(x,y)\psi(y)\mathrm{d}y\right)\cdot\nabla\phi(x)\mathrm{d}x=\int_{\mathrm{Q}}\psi(y)\left(\phi(y)-\int_{\mathrm{Q}}\phi\right)\mathrm{d}y.$$

Now, since $\widetilde{G}$ satisfies (3.12), one can replace $\psi$ by $\psi - \int_{\mathrm{Q}} \psi$ in the above equation and obtain

$$\begin{aligned}&\int_{\mathrm{Q}} A(x) \cdot \nabla_x \left( \int_{\mathrm{Q}} \widetilde{G}(x,y)\psi(y)\mathrm{d}y \right) \cdot \nabla\phi(x)\mathrm{d}x \\ &= \int_{\mathrm{Q}} \left( \psi(y) - \int_{\mathrm{Q}} \psi \right) \left( \phi(y) - \int_{\mathrm{Q}} \phi \right) \mathrm{d}y,\end{aligned}$$

which yields, by expanding the right-hand side term of the above equation,

$$\int_{\mathrm{Q}} A(x) \cdot \nabla_x \left( \int_{\mathrm{Q}} \widetilde{G}(x,y)\psi(y)\mathrm{d}y \right) \cdot \nabla\phi(x)\mathrm{d}x = \int_{\mathrm{Q}} \left( \psi(x) - \int_{\mathrm{Q}} \psi \right) \phi(x)\mathrm{d}x.$$

Therefore, the function

$$u : x \mapsto \int_{\mathrm{Q}} \widetilde{G}(x,y)\psi(y)\mathrm{d}y$$

is periodic, and obviously satisfies the variational formulation associated with the equation

$$-\mathrm{div}\left(A(x) \cdot \nabla u(x)\right) = \psi(x) - \int_{\mathrm{Q}} \psi.$$

Finally, as $\widetilde{G}$ satisfies (3.13), then $u$ has zero mean. Hence, $u = T\psi$. As a consequence, $\widetilde{G}$ is the Green function of $T$. This concludes the proof. □

## 3.3 Pointwise estimates on the periodic Green function

This section is devoted to the proof of Proposition 3.1.2. For technical reasons, we proceed first with the case of dimension $d \geq 3$, and then with the case of dimension $d = 2$.

### 3.3.1 The case of $d \geq 3$

Pointwise estimates on the Green functions of elliptic problems with Dirichlet boundary conditions have been established in the seminal article [72] of Grüter and Widman. Their proof makes use of the comparison principle. The latter is an appropriate tool for an elliptic *equation* with homogeneous Dirichlet boundary conditions : in this case, the Green function is positive. But, such an argument fails when considering the periodic Green functions, the sign of which varies (they have zero mean). As a consequence, here, we resort to a duality argument and to the De Giorgi-Nash-Moser theorem (see [11, Th. 13]). Also, when considering multiscale periodic elliptic *systems*, the latter theorem does not hold and the Hölder estimates of [11] are necessary for concluding the proof (see Section 3.1.2).

The proof below is a straightforward adaptation of the proof of [11, Th. 13]. The fact that we study periodic boundary conditions do not raise substantial difficulties, since the strategy involves local estimates.

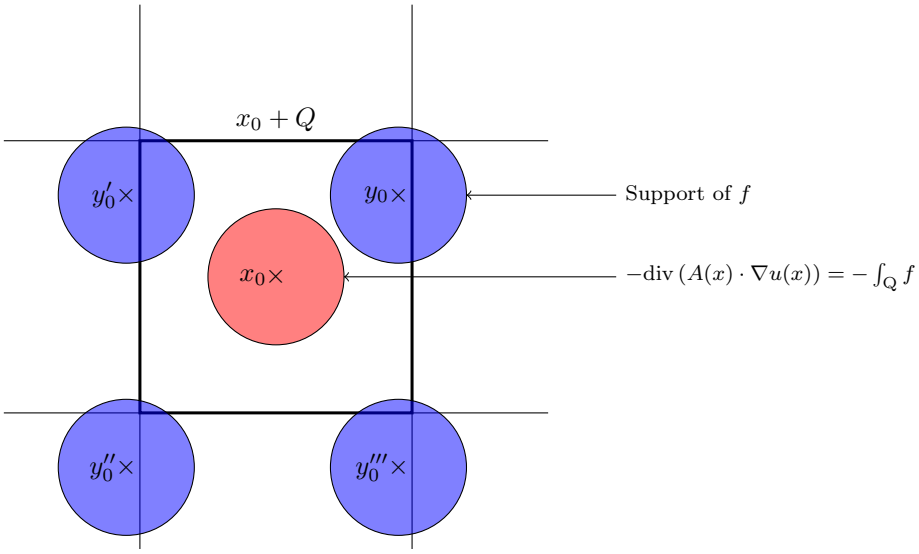



FIGURE 3.1 – Illustration of Step 1 of proof of Proposition 3.1.2.

Let us first explain in a few words the ingredients of the proof of Proposition 3.1.2 in the case $d \geq 3$. The first step of the proof consists in combining the De Giorgi-Nash-Moser theorem, the classical Hilbert theory and the Sobolev injections in order to obtain an optimal $\mathrm{L}^\infty$ estimate on the periodic solution $u$ to (3.5) for localized right-hand terms $f$. By a duality argument used in [11, Th. 13], this provides a local $\mathrm{L}^2$ bound on the Green function $G(x,y)$. Using once more the De Giorgi-Nash-Moser theorem, this proves Estimate (3.16).

*Proof of Proposition 3.1.2 in dimension $d \geq 3$.* The proof falls in two steps.

Let $x_0 \in \mathbb{R}^d, y_0 \in x_0 + Q$, $x_0 \neq y_0$, and $2r := |x_0 - y_0|$ (by periodicity, it is the only relevant case).

**Step 1 :** We recall first that the space $\mathrm{H}^1_{\mathrm{per}}(\mathrm{Q})$ is continuously embedded in $\mathrm{L}^{\frac{2d}{d-2}}_{\mathrm{per}}(\mathrm{Q})$ by Sobolev embedding (see [36, Th. 9.9 p. 278]). By a duality argument, $\mathrm{L}^{\frac{2d}{d+2}}_{\mathrm{per}}(\mathrm{Q})$ is continuously embedded in the dual of $\mathrm{H}^1_{\mathrm{per}}(\Omega)$.

Let us now consider $f \in \mathrm{L}^{\frac{2d}{d+2}}_{\mathrm{per}}(\mathrm{Q})$ that has a support contained in $\mathrm{B}(y_0, r/4) + \mathbb{Z}^d$ (see Figure 3.1). Define $u$ as the periodic solution with zero mean to (3.5). Since $\mathrm{L}^{\frac{2d}{d+2}}_{\mathrm{per}}(\mathrm{Q})$ is in the dual of $\mathrm{H}^1_{\mathrm{per}}(\mathrm{Q})$, then, by the Lax-Milgram theorem, $u \in \mathrm{H}^1_{\mathrm{per}}(\mathrm{Q})$ and there obviously holds

$$\|\nabla u\|_{\mathrm{L}^2(\mathrm{Q})} \leq C \, \|f\|_{\mathrm{L}^{\frac{2d}{d+2}}(\mathrm{Q})} . \tag{3.35}$$

As the support of $f$ is contained in $\mathrm{B}(y_0, r/4) + \mathbb{Z}^d$, then $u$ satisfies in $\mathrm{B}(x_0, r)$ the following

equation :

$$-\mathrm{div}\left(A(x)\cdot\nabla u(x)\right) = -\int_{\mathrm{Q}} f = -\left\{\int_{\mathrm{B}(y_0,r/4)} f\right\}\mathrm{div}\left(\frac{x-x_0}{d}\right).$$

Therefore, as a consequence of the De Giorgi-Nash-Moser theorem (see [64, Th. 8.24 p. 202]), there exists $\beta\in(0,1)$ and $C>0$ depending only on $d$ and $\mu$ such that

$$\begin{aligned}\sup_{x,y\in\mathrm{B}(x_0,r/2)}\frac{|u(x)-u(y)|}{r^{-\beta}|x-y|^\beta} \leq & C\left(\fint_{\mathrm{B}(x_0,r)}|u(x)|^2\,\mathrm{d}x\right)^{1/2}\\ &+C\left|\int_{\mathrm{B}(y_0,r/4)}f\right|\left(\fint_{\mathrm{B}(x_0,r)}(r|x-x_0|)^{2d}\mathrm{d}x\right)^{1/(2d)}\\ \leq & C\left(\fint_{\mathrm{B}(x_0,r)}|u(x)|^2\,\mathrm{d}x\right)^{1/2}+Cr^2\left|\int_{\mathrm{B}(y_0,r/4)}f\right|.\end{aligned}$$

Therefore,

$$\begin{aligned}|u(x_0)| \leq & \left|\fint_{\mathrm{B}(x_0,r/2)}u(x)\mathrm{d}x\right|+Cr^\beta\sup_{x,y\in\mathrm{B}(x_0,r/2)}\frac{|u(x)-u(y)|}{|x-y|^\beta}\\ \leq & Cr^{-d/2}\left(\int_{\mathrm{B}(x_0,r)}|u(x)|^2\,\mathrm{d}x\right)^{1/2}+Cr^2\left|\int_{\mathrm{B}(y_0,r/4)}f\right|.\end{aligned}\tag{3.36}$$

We now bound the two terms on the right-hand side. For the second term, by the Hölder inequality, we obtain

$$\left|\int_{\mathrm{B}(y_0,r/4)}f\right|\leq Cr^{\frac{d-2}{2}}\left(\int_{\mathrm{B}(y_0,r/4)}|f(x)|^{\frac{2d}{d+2}}\,\mathrm{d}x\right)^{\frac{d+2}{2d}}.\tag{3.37}$$

For the first term, again using the Hölder inequality, we have

$$\left(\int_{\mathrm{B}(x_0,r)}|u(x)|^2\,\mathrm{d}x\right)^{1/2}\leq Cr\left(\int_{\mathrm{Q}}|u(x)|^{\frac{2d}{d-2}}\,\mathrm{d}x\right)^{\frac{d-2}{2d}}.\tag{3.38}$$

By Sobolev injection of $\mathrm{H}^1(\mathrm{Q})$ in $\mathrm{L}^{\frac{2d}{d-2}}(\mathrm{Q})$ (and since $u$ has zero mean),

$$\left(\int_{\mathrm{Q}}|u(x)|^{\frac{2d}{d-2}}\,\mathrm{d}x\right)^{\frac{d-2}{2d}}\leq C\left(\int_{\mathrm{Q}}|\nabla u(x)|^2\,\mathrm{d}x\right)^{1/2},\tag{3.39}$$

whence, we deduce from (3.38), (3.39) and (3.35) that

$$\left(\int_{\mathrm{B}(x_0,r)}|u(x)|^2\,\mathrm{d}x\right)^{1/2}\leq Cr\left(\int_{\mathrm{B}(y_0,r/4)}|f(x)|^{\frac{2d}{d+2}}\,\mathrm{d}x\right)^{\frac{d+2}{2d}}.\tag{3.40}$$

Finally, since $2r \leq \sqrt{d}$, (3.36), (3.37), and (3.40) yield

$$|u(x_0)| \leq Cr^{-(d-2)/2} \left( \int_{\mathrm{B}(y_0,r/4)} |f(x)|^{\frac{2d}{d+2}} \, \mathrm{d}x \right)^{\frac{d+2}{2d}}. \tag{3.41}$$

**Step 2 :** The function $u$ can be expressed thanks to the Green function as

$$u(x) = \int_{\mathrm{Q}} G(x,y) f(y) \mathrm{d}y.$$

As a consequence, by duality, (3.41) implies that there exists a constant $C > 0$ such that

$$\left( \int_{\mathrm{B}(y_0,r/4)} |G(x_0,y)|^{\frac{2d}{d-2}} \, \mathrm{d}y \right)^{\frac{d-2}{2d}} \leq Cr^{-(d-2)/2}. \tag{3.42}$$

As $G(x_0,\cdot)$ satisfies

$$-\mathrm{div}_y \left( A^T(y) \cdot \nabla_y G(x_0,y) \right) = -1, \tag{3.43}$$

in $\mathrm{B}(y_0, r/4)$, then, using once more [64, Th. 8.24 p. 202], we obtain (in the same manner as (3.36))

$$\|G(x_0,\cdot)\|_{\mathrm{L}^\infty(\mathrm{B}(y_0,r/8))} \leq r^{-d/2} \left( \int_{\mathrm{B}(y_0,r/4)} |G(x_0,y)|^2 \, \mathrm{d}y \right)^{1/2} + Cr^2.$$

By the Hölder inequality, we deduce from the above estimate that

$$\|G(x_0,\cdot)\|_{\mathrm{L}^\infty(\mathrm{B}(y_0,r/8))} \leq r^{\frac{2-d}{2}} \left( \int_{\mathrm{B}(y_0,r/4)} |G(x_0,y)|^{\frac{2d}{d-2}} \, \mathrm{d}y \right)^{\frac{d-2}{2d}} + Cr^2.$$

Finally, since $2r = |x_0 - y_0| \leq \sqrt{d}$ and thanks to (3.42), we deduce (3.18). This concludes the proof of Proposition 3.1.2 in dimension $d \geq 3$. □

### 3.3.2 The case $d = 2$

We now turn to the dimension $d = 2$. Thanks to a trick that can be found in [11, Th. 13] (see also [29, Prop. 4]), we show that the 2-dimensional periodic Green function $G$ can be expressed as

$$G(x,y) = \int_0^1 \widetilde{G}((x,t),(y,0)) \mathrm{d}t \tag{3.44}$$

where $x, y \in \mathbb{R}^2$ and $t \in \mathbb{R}$ in the equation above, and $\widetilde{G}$ is the 3-dimensional periodic Green function of the following operator :

$$\widetilde{L}u := -\mathrm{div}_x \left( A \cdot \nabla_x u \right) - \partial_{tt} u. \tag{3.45}$$

Next, applying Proposition 3.1.2 –which we have already proved in dimension $d \geq 3$– to the Green function $\widetilde{G}$, we deduce Estimate (3.19).

*Proof of Proposition 3.1.2 in dimension $d = 2$.* The Green function $\widetilde{G}$ of (3.45) satisfies

$$\begin{aligned} &- \operatorname{div}_x \left( A(x) \cdot \nabla_x \widetilde{G}((x,t),(y,s)) \right) - \partial_{tt} \widetilde{G}((x,t),(y,s)) \\ &= \delta_y(x)\delta_s(t) - 1 \quad \text{in} \quad [0,1]^2 \times [0,1], \end{aligned} \tag{3.46}$$

and, for all $x \in \mathbb{R}^2$, $t \in \mathbb{R}$,

$$\int_{[0,1]^2} \int_0^1 \widetilde{G}((x,t),(y,s)) \mathrm{d}s \mathrm{d}y = 0. \tag{3.47}$$

The coefficients of the operator $\widetilde{L}$ defined by (3.45) do not depend on $(t,s)$. Therefore, by uniqueness of the Green function, the following identity holds :

$$\widetilde{G}((x,t),(y,s)) = \widetilde{G}((x,t-s),(y,0)). \tag{3.48}$$

By periodicity of $\widetilde{G}$, integrating (3.46) along the $t$ variable gives

$$- \operatorname{div}_x \left( A(x) \cdot \nabla_x \left\{ \int_0^1 \widetilde{G}((x,t),(y,0)) \mathrm{d}t \right\} \right) = \delta_y(x) - 1 \quad \text{in} \quad [0,1]^2.$$

Moreover, (3.47) and (3.48) imply that, for all $x \in \mathbb{R}^2$,

$$\int_{[0,1]^2} \left( \int_0^1 \widetilde{G}((x,t),(y,0)) \mathrm{d}t \right) \mathrm{d}y = 0,$$

and, for all $y \in \mathbb{R}^2$,

$$\int_{[0,1]^2} \left( \int_0^1 \widetilde{G}((x,t),(y,0)) \mathrm{d}t \right) \mathrm{d}x = 0,$$

Thus, by uniqueness of the Green function (see Proposition 3.1.1), we have established that $G$ satisfies (3.44).

Let $x \in \mathbb{R}^2$ and $y \in x + Q$, $x \neq y$. Invoking (3.18), $\widetilde{G}$ satisfies

$$\left| \widetilde{G}((x,t),(y,0)) \right| \leq C \frac{1}{|t| + |x-y|},$$

for all $t \in \mathbb{R} \backslash \{0\}$. Thus,

$$|G(x,y)| \leq C \int_0^1 \frac{1}{|t| + |x-y|} \mathrm{d}t \leq C \log\left(2 + |x-y|\right),$$

which is (3.19). This concludes the proof of Proposition 3.1.2 in the case $d = 2$. □

## 3.4 Pointwise estimates on the derivatives $\nabla_x G_n$, $\nabla_y G_n$ and $\nabla_x \nabla_y G_n$ of the multiscale periodic Green function

Proposition 3.1.3 relies on the Lipschitz theory of [11]. Indeed, if $u_n$ satisfies

$$-\mathrm{div}\left(A(nx)\cdot\nabla u_n(x)\right) = a \quad \text{in} \quad \mathrm{B}(x_0,r), \tag{3.49}$$

for some $a \in \mathbb{R}$, then [11, Lem. 16] shows that

$$|\nabla u_n(x_0)| \leq C r^{-1-d/2}\left(\int_{\mathrm{B}(x_0,r)} |u_n|^2\right)^{1/2} + C|a|r, \tag{3.50}$$

where $C > 0$ is a constant independent of $n$. As is detailed below, we treat separately the 2-dimensional case, in the same manner as in the proof of Proposition 3.1.2.

*Proof of Proposition 3.1.3.* Assume first that the dimension satisfies $d \geq 3$. Let $x_0 \in \mathbb{R}^d$, $y_0 \in x_0 + Q$, $x_0 \neq y_0$, and $2r := |x_0 - y_0|$ (once more, by periodicity, this is the only relevant case).

Since $G_n(x, y_0)$ satisfies

$$-\mathrm{div}_x\left(A(nx)\cdot\nabla_x G_n(x,y_0)\right) = -1 \quad \text{in} \quad \mathrm{B}(x_0,r),$$

and thanks to [11, Lem. 16], *id est* (3.50), there holds

$$|\nabla_x G_n(x_0,y_0)| \leq C r^{-1}\left\|G_n(\cdot,y_0)\right\|_{\mathrm{L}^\infty(\mathrm{B}(x_0,r))} + Cr. \tag{3.51}$$

As a consequence, Estimate (3.18) and (3.51) yield (3.21) (for $x = x_0$ and $y = y_0$). By transposition, (3.22) is also established.

By differentiating (3.9) with respect to $y$, we obtain

$$-\mathrm{div}_x\left(A(nx)\cdot\nabla_x\nabla_y G_n(x,y_0)\right) = 0 \quad \text{in} \quad \mathrm{B}(x_0,r/2).$$

Therefore, thanks to (3.50),

$$|\nabla_x\nabla_y G_n(x_0,y_0)| \leq C r^{-1}\left\|\nabla_y G(\cdot,y_0)\right\|_{\mathrm{L}^\infty(\mathrm{B}(x_0,r/2))}.$$

By using (3.22) and since $2r = |x_0 - y_0|$, we conclude that

$$|\nabla_x\nabla_y G_n(x_0,y_0)| \leq C|x_0 - y_0|^{-d},$$

and establish (3.23).

Now, let the dimension $d = 2$. As shown in the proof of Proposition 3.1.2, the 2-dimensional periodic Green function $G_n$ can be expressed as

$$G_n(x,y) = \int_0^1 \widetilde{G}_n((x,t),(y,0))\mathrm{d}t \tag{3.52}$$

where $x, y \in \mathbb{R}^2$ and $t \in \mathbb{R}$ in the equation above, and $\widetilde{G}_n$ is the 3-dimensional periodic Green function of the following operator :

$$\widetilde{L}u := -\mathrm{div}_x\left(A(n\cdot)\cdot\nabla_x u\right) - \partial_{tt}u. \tag{3.53}$$

Hence, we deduce the 2-dimensional versions of (3.21), (3.22) and (3.23) by integrating their 3-dimensional versions applied to $\widetilde{G}_n$. □

## 3.5 A decomposition of the periodic Green function

In this section, we prove Proposition 3.1.4. For the sake of simplicity, we first assume that the homogenized matrix $A^\star$ is the identity. We postpone the proof in the general case until Section 3.5.2.

### 3.5.1 Case where the homogenized matrix is the identity

For convenience, we have split the proof of Proposition 3.1.4 in two parts : first, we show that the series in (3.25) actually converges ; second, we check that its limit is the periodic Green function $G$.

The two main steps of the proof of convergence are the following : first a Taylor expansion allows for expressing the terms $H^k$ in (3.25) as functions of $\nabla_x \nabla_y \mathcal{G}$. Then, we approximate the Green function of the multiscale problem with the Green function of the homogenized problem (see [12]). Second, we take advantage of the long-range symmetries of the Green function of the homogenized problem and establish the convergence of the series in (3.25). There, the sets $\Gamma_m$ are crucial, since the convergence in (3.25) is not uniform in $k$.

*Proof of convergence of the series in* (3.25). We denote by $\mathcal{G}_\star(x,y)$ the fundamental solution in $\mathbb{R}^d$ to the homogenized problem. The homogenized matrix is the identity ; therefore $\mathcal{G}_\star$ is explicitly expressed as

$$\mathcal{G}_\star(x,y) = C_d |x-y|^{2-d}, \tag{3.54}$$

where $C_d$ is a constant (see [64, (2.12) p. 17]). Since $\mathcal{G}_\star(x,y)$ only depends on $|x-y|$, we henceforth redefine

$$\mathcal{G}_\star(x) := \mathcal{G}_\star(x,0).$$

**Step 1 :** Assume that $x \in \mathrm{Q}, y-x \in \mathrm{Q}$ and $k \notin 4\mathrm{Q}$. We reformulate $H^k$ using the Taylor formula :

$$\begin{aligned} H^k(x,y) = &- \int_{\mathrm{Q}} x' \cdot \int_0^1 \Bigg( \nabla_x \mathcal{G}(x+tx', y-k) \\ &\qquad - \int_{\mathrm{Q}} \nabla_x \mathcal{G}(x+tx', y+y'-k) \mathrm{d}y' \Bigg) \mathrm{d}t \mathrm{d}x' \\ = &\int_{\mathrm{Q}} x' \cdot \Bigg( \int_0^1 \int_{\mathrm{Q}} y' \\ &\qquad \cdot \int_0^1 \nabla_y \nabla_x \mathcal{G}(x+tx', y+\tau y'-k) \mathrm{d}\tau \mathrm{d}y' \mathrm{d}t \Bigg) \mathrm{d}x'. \end{aligned} \tag{3.55}$$

We approximate $\nabla_x\nabla_y\mathcal{G}$ by $\nabla_x\nabla_y\mathcal{G}_\star$. More precisely, thanks to [12, Corollary p. 905], there exists constants $C>0$ and $\beta\in(0,1)$ such that, for all $x'\neq y'\in\mathbb{R}^d$,

$$\left|\nabla_x\nabla_y\mathcal{G}(x',y')-\sum_{i,j=1}^{d}\partial_{x_i}\partial_{y_j}\mathcal{G}_\star(x'-y')\left(e_i+\nabla w_i\left(x'\right)\right)\otimes\left(e_j+\nabla w_j^\dagger\left(y'\right)\right)\right| \leq C|x'-y'|^{-d-\beta}. \tag{3.56}$$

In (3.56), $w_i$ and $w_j^\dagger$ denote the correctors associated to the matrix $A$, respectively $A^T$. That is, $w_i$ is the periodic function of zero mean satisfying

$$-\mathrm{div}\left(A(x)\cdot(\nabla w_i(x)+e_i)\right)=0,\quad \text{for}\quad x\in\mathrm{Q}.$$

Identity (3.55) and Estimate (3.56) imply

$$H^k(x,y)=H^{1,k}(x,y)+H^{2,k}(x,y), \tag{3.57}$$

where

$$\left|H^{1,k}(x,y)\right|\leq C|k|^{-d-\beta}, \tag{3.58}$$

and

$$\begin{aligned}H^{2,k}(x,y):=\sum_{i,j=1}^{d}\int_{\mathrm{Q}^2}\int_{[0,1]^2}&\left(x'\cdot\left(e_i+\nabla w_i(x+tx')\right)\right)\\&\left(y'\cdot\left(e_j+\nabla w_j^\dagger(y+\tau y'-k)\right)\right)\\&\partial_{x_i}\partial_{y_j}\mathcal{G}_\star\left(x+tx'-(y+\tau y'-k)\right)\mathrm{d}\tau\mathrm{d}t\mathrm{d}y'\mathrm{d}x'.\end{aligned}$$

All the correctors $w_i$ and $w_j^\dagger$ are bounded; furthermore, Formula (3.54) implies that the third-order derivatives of $\mathcal{G}_\star$ evaluated at $x-y$ are bounded by $C|x-y|^{-d-1}$, where $C$ is a constant independent of $x$ and $y$. Therefore, a Taylor expansion yields a constant $C$ such that

$$\left|H^{2,k}(x,y)-\sum_{i,j=1}^{d}\partial_{x_i}\partial_{y_j}\mathcal{G}_\star(k)Q_{ij}(x,y)\right|\leq C|k|^{-d-1}, \tag{3.59}$$

where

$$\begin{aligned}Q_{ij}(x,y):=\int_{\mathrm{Q}^2}\int_{[0,1]^2}&\left(x'\cdot\left(e_i+\nabla w_i(x+tx')\right)\right)\\&\left(y'\cdot\left(e_j+\nabla w_j^\dagger(y+\tau y')\right)\right)\mathrm{d}\tau\mathrm{d}t\mathrm{d}y'\mathrm{d}x'.\end{aligned}$$

Then, a straightforward integration yields

$$Q_{ij}(x,y)=\int_{\mathrm{Q}}\left(x_i'+w_i(x+x')-w_i(x)\right)\mathrm{d}x'\int_{\mathrm{Q}}\left(y_j'+w_j^\dagger(y+y')-w_j^\dagger(y)\right)\mathrm{d}y',$$

which can be simplified as

$$Q_{ij}(x,y) = w_i(x) w_j^\dagger(y),$$

since the correctors $w_i$ and $w_j^\dagger$ are of zero mean. Note that $Q_{ij}$ defined above does not depend on $k \in \mathbb{Z}^d$ because the correctors $w_i$ and $w_j^\dagger$ are periodic. As a consequence, collecting (3.57), (3.58), and (3.59) yields

$$\left| \sum_{k\in\Gamma_m} H^k(x,y) \right| \leq C 2^{-m\beta} + |Q_{ij}(x,y)| \sum_{i,j=1}^{d} \left| \sum_{k\in\Gamma_m} \partial_{x_i}\partial_{y_j} \mathcal{G}_\star(k) \right|. \tag{3.60}$$

**Step 2 :** Remark that $Q_{ij}(x,y) \neq 0$ in general and that $\left|\partial_{x_i}\partial_{y_j}\mathcal{G}_\star(k)\right|$ scales like $|k|^{-d}$. Therefore, by (3.59), in general, the series in (3.25) does not converge absolutely with respect to $k$.

Invoking once more (3.54), we obtain

$$\partial_{x_i}\partial_{y_j}\mathcal{G}_\star(k) = \begin{cases} C_d d(d-2) \dfrac{k_i k_j}{|k|^{d+2}} & \text{if} \quad i \neq j, \\ C_d (d-2) \dfrac{d k_i^2 - |k|^2}{|k|^{d+2}} & \text{if} \quad i = j. \end{cases}$$

Thanks to the symmetry of $\Gamma_m$ with respect to the hyperplane $x_i = 0$, in the case $i \neq j$, and thanks to the invariance of $\Gamma_m$ under the relabeling of the components of the vector $k$, in the case $i = j$, we deduce that

$$\sum_{k\in\Gamma_m} \partial_{x_i}\partial_{y_j}\mathcal{G}_\star(k) = 0, \qquad \forall i,j \in [\![1,d]\!]. \tag{3.61}$$

As a consequence, recalling (3.60),

$$\left| \sum_{k\in\Gamma_m} H^k(x,y) \right| \leq C 2^{-m\beta}. \tag{3.62}$$

Moreover, by [29, Prop. 4], there exists a constant $C$ such that for any $x' \neq y'$, there holds

$$\left|\mathcal{G}(x',y')\right| \leq C |x'-y'|^{-d+2}. \tag{3.63}$$

Hence, for any $k \in \mathbb{Z}^d$,

$$\left|H^k(x,y)\right| \leq C |x-y|^{-d+2}. \tag{3.64}$$

As a consequence of (3.62) and (3.64), the series (3.25) converges absolutely in $m$ as follows :

$$\sum_{m=0}^{+\infty} \left| \sum_{k\in\Gamma_m} H^k(x,y) \right| \leq C|x-y|^{-d+2},$$

for all $x \neq y$, $y - x \in \mathrm{Q}$. Thus, we have recovered (3.16) by an approach different from Proposition 3.1.2. □

Now that we have justified that the series in (3.25) converges, we prove that its limit, that we denote by $\overline{G}$ for the moment, *id est*,

$$\overline{G}(x,y) := \sum_{m=0}^{+\infty} \left( \sum_{k\in\Gamma_m} H^k(x,y) \right), \tag{3.65}$$

is actually equal to $G$. It can easily be checked that $\overline{G}$ is periodic in $x$ and $y$ and satisfies (3.9). The technical point is (3.12), the proof of which relies on the former Taylor expansion and on the classical result [29, Prop. 8]. According to the latter, there exists a constant $C > 0$ such that, for all $x' \neq y'$,

$$\left|\nabla_x \nabla_y \mathcal{G}(x',y')\right| \leq C|x'-y'|^{-d}. \tag{3.66}$$

*Proof of Identity* (3.25). Obviously, $\overline{G}$ is periodic in $y$. Since $A$ is periodic, there even holds

$$\mathcal{G}(x, y-k) = \mathcal{G}(x+k, y),$$

for any $k \in \mathbb{Z}^d$. Therefore $\overline{G}$ is also periodic in $x$. Moreover, we check that

$$\begin{aligned} -\mathrm{div}\left(A(x)\cdot\nabla H^k(x,y)\right) =& \delta_0(x-y+k) \\ & - 2\chi_{\mathrm{Q}}(x-y+k) + \int_{\mathrm{Q}} \chi_{\mathrm{Q}}(x-y-y'+k)\mathrm{d}y', \end{aligned}$$

where $\chi_{\mathrm{Q}}$ is the characteristic function of the set Q. Whence

$$-\mathrm{div}\left(A(x)\cdot\left(\sum_{m=0}^{+\infty}\left(\sum_{k\in\Gamma_m}\nabla_x H^k(x,y)\right)\right)\right) = \sum_{k\in\mathbb{Z}^d}\delta_0(x-y+k) - 1 \quad \text{in} \quad \mathbb{R}^d.$$

To summarize, $\overline{G}(x,y)$ defined by (3.65) is $x$-periodic and $y$-periodic, and satisfies (3.9).

Next, we justify that $\overline{G}$ satisfies (3.12). By integrating (3.55) along the $y$ variable, there holds

$$\begin{aligned} \int_Q H^k(x,y)\mathrm{d}y = & \int_0^1\int_0^1\int_{\mathrm{Q}}\int_{y\in\partial Q} \\ & x'\cdot\left(\int_{\mathrm{Q}}\nabla_x\mathcal{G}(x+tx', y+\tau y'-k)\otimes y'\mathrm{d}y'\right)\cdot\mathrm{d}\vec{S}(y)\mathrm{d}x'\mathrm{d}\tau\mathrm{d}t. \end{aligned}$$

Hence, due to cancellations on the boundaries of the translated cubes $k+Q$,

$$\int_{\mathrm{Q}} \sum_{|k|<2^m} H^k(x,y)\mathrm{d}y = \int_0^1 \int_0^1 \int_{\mathrm{Q}} \int_{y\in\Xi_m} x' \cdot \left( \int_{\mathrm{Q}} \nabla_x \mathcal{G}(x+tx', y+\tau y'-k) \otimes y' \mathrm{d}y' \right) \cdot \mathrm{d}\vec{S}(y) \mathrm{d}x' \mathrm{d}\tau \mathrm{d}t,$$

where $\Xi_m$ is the boundary of the following set :

$$\bigcup_{|k|<2^{m+1}} (k+Q)\,.$$

Now, by Taylor expansion, and thanks to (3.66), for all $\tau \in [0,1]$, $x, x', y' \in \mathrm{Q}$, and $y \in \Xi_m$,

$$\left|\nabla_x \mathcal{G}(x+tx', y+\tau y') - \nabla_x \mathcal{G}(x+tx', y)\right| \leq C 2^{-md}.$$

Therefore

$$\left| \int_{\mathrm{Q}} \nabla_x \mathcal{G}(x+tx', y+\tau y') \otimes y' \mathrm{d}y' \right| \leq \left| \int_{\mathrm{Q}} \nabla_x \mathcal{G}(x+tx', y) \otimes y' \mathrm{d}y' \right| + C 2^{-md}.$$

The integral in the right-hand term of the above estimate vanishes since, by symmetry, $\int_{\mathrm{Q}} y' \mathrm{d}y' = 0$. As a consequence,

$$\left| \int_{\mathrm{Q}} \nabla_x \mathcal{G}(x+tx', y+\tau y') \otimes y' \mathrm{d}y' \right| \leq C 2^{-md}.$$

Whence, since the surface area of $\Xi_m$ is bounded by $C 2^{m(d-1)}$, we have

$$\left| \int_{\mathrm{Q}} \sum_{|k|<2^{m+1}} H_k(x,y) \mathrm{d}y \right| \leq C 2^{-m}.$$

As a consequence, letting $m \to +\infty$, we deduce that $\overline{G}$ satisfies (3.12). By the same arguments transposed from $\overline{G}(x,y)$ to $\overline{G}(y,x)$, it can be shown that $\overline{G}$ also satisfies (3.13). Therefore, by Proposition 3.1.1, we have

$$\overline{G}(x,y) = G_n(x,y),$$

which concludes the proof. □

### 3.5.2 General case

As is easily seen, in Proposition 3.1.4, the fact that the homogenized matrix $A^\star$ is the identity is only used for establishing (3.61). One also realizes that, would (3.61) be replaced by the following estimates :

$$\left| \sum_{k\in\Gamma_m} \partial_{x_i} \partial_{y_j} \mathcal{G}_\star(k) \right| \leq C_m \text{ for all } m \in [\![1, +\infty), \quad \text{and} \quad \sum_{m=1}^{+\infty} C_m < +\infty, \tag{3.67}$$

for well-chosen sets $\Gamma_m$, then the conclusions of Proposition 3.1.4 would also apply.

We show that the sets $\Gamma_m$ defined by (3.27) are such that Estimates (3.67) are satisfied. Hence, the conclusions of Proposition 3.1.4 are true without any assumption on the homogenized matrix $A^\star$ of $A$, when using the sets $\Gamma_m$ defined above.

The homogenized matrix $A^\star$ is (constant) coercive. Then the Green function in $\mathbb{R}^d$ associated with the operator $-\mathrm{div}\,(A^\star \cdot \nabla)$ is

$$\mathcal{G}_\star(x) = \frac{C\left(A^\star_{\mathrm{s}}\right)}{\left(x \cdot \left(A^\star_{\mathrm{s}}\right)^{-1} \cdot x\right)^{(d-2)/2}},$$

where $C\left(A^\star_{\mathrm{s}}\right)$ is a constant and $A^\star_{\mathrm{s}}$ is the symmetric part of the matrix $A^\star$. Whence

$$\nabla^2 \mathcal{G}_\star(x) = C(d-2)\frac{d\left(\left(A^\star_{\mathrm{s}}\right)^{-1} \cdot x\right) \otimes \left(\left(A^\star_{\mathrm{s}}\right)^{-1} \cdot x\right) - \left(x \cdot \left(A^\star_{\mathrm{s}}\right)^{-1} \cdot x\right)\left(A^\star_{\mathrm{s}}\right)^{-1}}{\left(x \cdot \left(A^\star_{\mathrm{s}}\right)^{-1} \cdot x\right)^{(d+2)/2}}.$$

Besides, there exists an orthogonal matrix $O$ and positive scalars $\lambda_i$, $i \in [\![1, d]\!]$ such that

$$A^\star_{\mathrm{s}} = O^{-1} \cdot \mathrm{diag}\left(\lambda_1^{-2}, \cdots, \lambda_d^{-2}\right) \cdot O.$$

Therefore, denoting by $f_j$ the orthonormal base related to $O$, and decomposing

$$x = \sum_{j=1}^{d} \lambda_j^{-1} \widetilde{x}_j f_j,$$

one obtains

$$\nabla^2 \mathcal{G}_\star(x) = C(d-2)\frac{d\sum_{i,j=1}^d \lambda_i\lambda_j \widetilde{x}_i\widetilde{x}_j f_i \otimes f_j - \left(\sum_{i=1}^d \widetilde{x}_i^2\right)\left(\sum_{i=1}^d \lambda_i^2 f_i \otimes f_i\right)}{\left(\sum_{i=1}^d \widetilde{x}_i^2\right)^{(d+2)/2}}.$$

For $r \in \mathbb{R}_+$, we define the following set :

$$\Omega_r := \left\{ x \in \mathbb{R}^d, \sum_{j=1}^{d} |\widetilde{x}_j|^2 = r^2 \right\}. \tag{3.68}$$

Remark that there obviously holds

$$\Omega_r = \left\{ x \in \mathbb{R}^d, \left(x \cdot \left(A^\star_{\mathrm{s}}\right)^{-1} \cdot x\right) = r^2 \right\}.$$

On the one hand, if $i \neq j$, $f_i \otimes f_j : \nabla^2 \mathcal{G}_\star(x)$ changes sign with respect to the transformation $\widetilde{x}_j \mapsto -\widetilde{x}_j$. Therefore, if $i \neq j$, there holds

$$\int_{\Omega_r} f_i \otimes f_j : \nabla^2 \mathcal{G}_\star(x) \mathrm{d}S(x) = 0. \tag{3.69}$$

On the other hand,

$$f_i \otimes f_i : \nabla^2 \mathcal{G}_\star(x) = C(d-2) \frac{\lambda_i^2 \left( d\widetilde{x}_i^2 - \left( \sum_{j=1}^d \widetilde{x}_j^2 \right) \right)}{\left( \sum_{k=1}^d \widetilde{x}_k^2 \right)^{(d+2)/2}}.$$

By invariance of $\Omega_r$ under the relabeling of the coordinates $\widetilde{x}_j$,

$$\int_{\Omega_r} \frac{\sum_{k=1}^d \widetilde{x}_k^2}{\left( \sum_{k=1}^d \widetilde{x}_k^2 \right)^{(d+2)/2}} \mathrm{d}S(x) = d \int_{\Omega_r} \frac{\widetilde{x}_i^2}{\left( \sum_{k=1}^d \widetilde{x}_k^2 \right)^{(d+2)/2}} \mathrm{d}S(x).$$

As a consequence,

$$\int_{\Omega_r} f_i \otimes f_i : \nabla^2 \mathcal{G}_\star(x) \mathrm{d}S(x) = 0. \tag{3.70}$$

Hence, we define

$$\Lambda_m := \bigcup_{r \in [2^m - 1, 2^{m+1} - 1)} \Omega_r,$$

and $\Gamma_m := \Lambda_m \cap \mathbb{Z}^d$. By convergence of Riemann integrals, we deduce that

$$\begin{aligned} \left| \sum_{k \in \Gamma_m} \partial_i \partial_j \mathcal{G}_\star(k) \right| \leq & C \left| \partial \Lambda_m \right| \sup_{x \in \Lambda_m} \left| \partial_i \partial_j \mathcal{G}_\star(x) \right| \\ & + C \left| \Lambda_m \right| \sup_{x \in \Lambda_m} \left| \partial_i \partial_j \nabla \mathcal{G}_\star(x) \right| \\ & + C \left| \int_{\Lambda_m} \partial_i \partial_j \mathcal{G}_\star(x) \mathrm{d}x \right|. \end{aligned}$$

By (3.69) and (3.70), we deduce that the last integral in the above estimate vanishes. Furthermore, straightforward estimates on the derivatives of $\mathcal{G}_\star$ yield

$$\left| \sum_{k \in \Gamma_m} \partial_i \partial_j \mathcal{G}_\star(k) \right| \leq C 2^{-m}.$$

This implies the convergence of the series in (3.25) in the general case.

### 3.5.3 Case of systems

In the case of systems, the Green function $\mathcal{G}_\star$ of $-\mathrm{div}\,(A^\star \cdot \nabla)$ in $\mathbb{R}^d$ reads :

$$\mathcal{G}_\star^{\alpha\beta}(x) = C \left( (A_{\mathrm{s}}^\star)^{\alpha\beta} \right) \left( \sum_{i,j=1}^d x_i \left( (A_{\mathrm{s}}^\star)^{\alpha\beta} \right)_{ij}^{-1} x_j \right)^{-\frac{d-2}{2}}, \tag{3.71}$$

where

$$(A_{\mathrm{s}}^{\star})_{ij}^{\alpha\beta} = \frac{1}{2}\left((A^{\star})_{ij}^{\alpha\beta} + (A^{\star})_{ji}^{\alpha\beta}\right).$$

Whence, by the above arguments of Sections 3.5.1 and 3.5.2, we have the following decomposition :

$$G^{\alpha\beta}(x,y) = \sum_{m=0}^{+\infty}\left(\sum_{k\in\Gamma_m^{\alpha\beta}}\left(H^k\right)^{\alpha\beta}(x,y)\right), \tag{3.72}$$

where the meaning of each term will be made precise below.

The functions $H^k$ are defined by

$$\begin{aligned}\left(H^k\right)^{\alpha\beta}(x,y) := \mathcal{G}^{\alpha\beta}(x,y-k) &- \int_{\mathrm{Q}}\mathcal{G}^{\alpha\beta}(x,y+y'-k)\mathrm{d}y' \\ &- \int_{\mathrm{Q}}\mathcal{G}^{\alpha\beta}(x+x',y-k)\mathrm{d}x' \\ &+ \int_{\mathrm{Q}}\int_{\mathrm{Q}}\mathcal{G}^{\alpha\beta}(x+x',y+y'-k)\mathrm{d}y'\mathrm{d}x',\end{aligned} \tag{3.73}$$

where the function $\mathcal{G}$ is the Green function in $\mathbb{R}^d$ of the operator $-\mathrm{div}\,(A\cdot\nabla)$. Last, we define the sets $\Gamma_m^{\alpha\beta}$ by :

$$\Gamma_m^{\alpha\beta} = \left\{k\in\mathbb{Z}^d, 2^m - 1 \leq k\cdot\left((A_{\mathrm{s}}^{\star})^{\alpha\beta}\right)^{-1}\cdot k < 2^{m+1}-1\right\},$$

where $(A_{\mathrm{s}}^{\star})^{\alpha\beta}$ is considered as a matrix in $\mathbb{R}^{d\times d}$.

## Acknowledgement

The author would like to thank both Frédéric Legoll and Pierre-Loïc Rothé, who have brought his attention to this question. Frédéric Legoll has also helped for the proof in the 2-dimensional case. Moreover, the author acknowledges Claude Le Bris for his suggestions concerning the decomposition of the Green functions. Finally, the author gratefully thanks Xavier Blanc for reading the first version of this article, and the anonymous reviewer for his or her accurate suggestions.

# Chapitre 4

# Quelques propriétés mathématiques de l'équation de Weertman

Ce chapitre reprend la prépublication [87] en anglais.

Nous y étudions quelques propriétés mathématiques de l'équation de Weertman. Nous démontrons en particulier qu'elle est la limite en temps long d'une équation de réaction-diffusion avec laplacien fractionnaire. Outre l'intérêt théorique de ces résultats, ils fournissent une assise sur laquelle nous construirons un algorithme de résolution numérique dans le Chapitre 5.

Le lecteur trouvera en Annexe A.4 des matériaux théoriques additionnels.

# Some mathematical properties of the Weertman equation

Marc Josien

**Abstract** We derive here some mathematical properties of the Weertman equation and show that it is the limit of an evolution equation. The Weertman equation is a semilinear integrodifferential equation involving a fractional Laplacian. In addition to this purely theoretical interest, the results proven here give a solid ground to a numerical approach that we have implemented in [88].

**Keywords** Reaction-advection-diffusion equation, traveling waves, integrodifferential equation, the Weertman equation, fractional Laplacian

## 4.1 Introduction

**Motivation** We derive here some mathematical properties of the Weertman equation and show that it is the limit of an evolution equation. Our motivation comes from our interest in materials science. The problem we consider, however classical, enjoys the following specificity that it involves the dissipative integrodifferential operator $-|\partial_x|$ (also denoted as $-(-\Delta)^{1/2}$), which has $-|k|$ as Fourier symbol. In addition to this purely theoretical interest, the results proven here give a solid ground to a numerical approach that we have implemented in [88].

Our starting point is the so-called Weertman equation (see [135]) :

$$-|\partial_x|\,\eta(x) + c\eta'(x) = F'(\eta(x)) \qquad \text{for} \quad x \in \mathbb{R}, \tag{4.1}$$

with boundary conditions

$$\lim_{x\to-\infty} \eta(x) = \eta_{\mathrm{l}} \quad \text{and} \quad \lim_{x\to+\infty} \eta(x) = \eta_{\mathrm{r}}, \tag{4.2}$$

where *both* the scalar $c \in \mathbb{R}$ (called velocity) and the function $\eta \in \mathrm{C}^2(\mathbb{R})$ are the unknowns, and where $\eta_{\mathrm{l}} < \eta_{\mathrm{r}}$. The function $F \in \mathrm{C}^3(\mathbb{R})$ is a bistable potential ; namely, it satisfies

$$F'(\eta_{\mathrm{l}}) = F'(\eta_{\mathrm{r}}) = 0, \quad F''(\eta_{\mathrm{l}}) > 0, \quad \text{and} \quad F''(\eta_{\mathrm{r}}) > 0. \tag{4.3}$$

From a physical point of view, Equation (4.1) can be seen as a nondimensionalized form of the Weertman equation for steadily-moving dislocations in materials science (see [135]). The latter equation is a generalization of the classical Peierls-Nabarro equation [129]. Dislocations are linear defects in crystals, the motion of which is responsible for the plasticity of metals. From a physical standpoint, the function $\eta$ represents a discontinuity between the local relative material displacement $u(x, y)$ in the upper and in the lower half-spaces surrounding the glide plane on which moves the dislocation line (see Figure 4.1) ; see, *e.g.*, [81] for a classical reference on dislocations. In (4.1), the term $|\partial_x|\,\eta$ accounts for the long-range

elastic self-interactions that tend to spread the core. This repulsive interaction is counterbalanced by the nonlinear pull-back force $F'(\eta)$, which binds together the upper and lower half-spaces. Moreover, the moving dislocation is subjected to various drag mechanisms encoded into the term $c\eta'$.

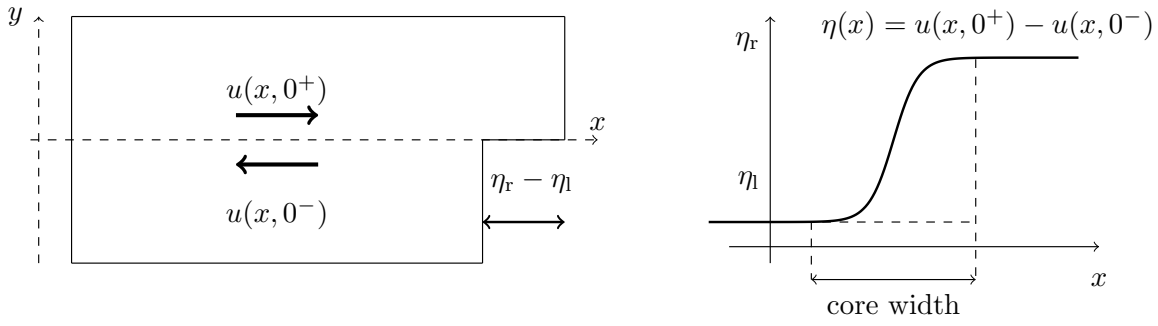



FIGURE 4.1 – Typical shape of $\eta(x)$, solution to (4.1) ; here, $u(x,y)$ is the material displacement.

From a broader perspective, the function $\eta$ can be understood as a moving phase-transformation front between the states $\eta_l$ and $\eta_r$ (see Figure 4.1), which are local minimizers of the potential $F$. In this regard, Equation (4.1) can also be found in other fields of physics. For example, it can be used to model complex media such as living cells, in which the operator $|\partial_x|$ accounts for an anomalous diffusion. See, *e.g.*, [121], where the more general operator $|\partial_x|^\alpha$ is considered (the latter operator has $|k|^\alpha$ as Fourier symbol).

**Traveling wave of reaction-diffusion equation** Equation (4.1) is a special case of the general equation

$$\begin{cases} \mathcal{A}[\eta](x) + c\eta'(x) = 0 \quad \text{for} \quad x \in \mathbb{R}, \\ \eta(-\infty) = \eta_l \quad \text{and} \quad \eta(+\infty) = \eta_r, \end{cases} \tag{4.4}$$

where $\mathcal{A}[\eta] = L\eta - F'(\eta)$ is a nonlinear operator, in which $L$ is a diffusive operator and $F$ is a bistable potential. As is easily seen, Equation (4.4) describes the traveling waves of the following reaction-diffusion equation :

$$\begin{cases} \partial_t u(t,x) = \mathcal{A}[u(t,\cdot)](x) \quad \text{for} \quad x \in \mathbb{R}, \\ u(0,x) = u_0(x) \qquad\qquad\quad \text{for} \quad x \in \mathbb{R}, \end{cases} \tag{4.5}$$

in the sense that, if $u(t,x) = \eta(x - ct)$ is a traveling wave satisfying (4.5), then $(\eta, c)$ solves (4.4). *Ipso facto*, finding a solution to (4.4) amounts to finding traveling wave solutions to (4.5). Natural questions thus arise :

(i) Does Equation (4.4) have one and only one solution $(\eta, c)$ ?

(ii) Which properties does the solution to Equation (4.4) enjoy ?

(iii) Is the traveling wave $\eta(x-ct)$, for $(\eta, c)$ solution to (4.4), an attractor of the dynamical system (4.5) ?

These questions have been addressed by many authors for various operators $L$ and for bistable potential $F$ satisfying -most of the times- the extra condition that $F$ does not admit any local minimum between $\eta_{\rm l}$ and $\eta_{\rm r}$. Other types of nonlinearities, not considered here, have attracted much attention. See [150] for the classification of traveling waves and an overview of reaction-diffusion equations.

In the seminal article [136], Sattinger remarked that if $(\eta, c)$ is a solution to (4.4), then $(\eta(\cdot+\xi), c)$ is also a solution to (4.4), for arbitrary $\xi \in \mathbb{R}$. Therefore, solutions to (4.4) can at most be unique *up to a translation*. In this regard, he introduced the notion of asymptotic stability of traveling waves and proved that the solution $(\eta, c)$ to (4.4), if it exists, is asymptotically stable under general assumptions about the spectrum of $\mathcal{A}$.

In the celebrated article [60], Fife and McLeod answered the Questions (i) and (iii) for the case where $L$ is the Laplacian. They proved indeed that if $F$ satisfies (4.3) and has no local minimum between $\eta_{\rm l}$ and $\eta_{\rm r}$, then there exists a solution $(\eta, c)$ to (4.4), which is unique up to a translation, and that this solution is globally asymptotically stable. Namely, for all initial conditions $u_0$ taking values in $[\eta_{\rm l}, \eta_{\rm r}]$ such that $u_0(-\infty) = \eta_{\rm l}$ and $u_0(+\infty) = \eta_{\rm r}$, there exist $\xi \in \mathbb{R}$, $K > 0$ and $\kappa > 0$ such that the solution $u$ of (4.5) satisfies

$$\|u(t,\cdot) - \eta(\cdot - \xi - ct)\|_{\mathrm{L}^\infty(\mathbb{R})} \leq K \mathrm{e}^{-\kappa t}, \tag{4.6}$$

for all $t \in \mathbb{R}_+$. Among other important concepts, all amenable to a wide class of dissipative operators, it is observed in [60] that $\mathcal{A}$ satisfies a comparison principle. Thus, any solution $u(t,x)$ of (4.5) can be squeezed between a super-solution $w_{+1}(t,x)$ and a sub-solution $w_{-1}(t,x)$, both at a controlled distance from $\eta(x-\xi-ct)$.

In a more recent article [45], Chen combined this squeezing approach with an iterative technique. Under technical assumptions about the operator $\mathcal{A}$, he proved the global asymptotic stability of the traveling waves of (4.5), provided that there exists a monotonic solution $\eta$ to (4.4). In this context, a positive answer to Question (i) and technical assumptions imply, using Chen's *squeezing* technique, a positive answer to Question (iii). We use Chen's approach in the present article.

The article [45] also provides tools for establishing the existence and the uniqueness of a solution to (4.4). They have been used in [46] to positively answer to Question (i) in the case where $L$ is the fractional Laplacian $-|\partial_x|^\alpha$, for $\alpha \in (0,2)$. Also, in [1], the authors have adapted Chen's squeezing technique to prove that the solution $(\eta, c)$ to (4.4) is globally asymptotically stable in the sense of (4.6), in a general framework including the case $L = -|\partial_x|^\alpha$, for $\alpha \in (1,2)$. However, they underlined the fact that the case $\alpha \leq 1$ (and in particular $\alpha = 1$), is still an open question. This motivates our study.

With an approach different from [45], the existence and the uniqueness of a solution to (4.4), for $L = -|\partial_x|^\alpha$ and $\alpha \in (0,2)$, has been proved in [40, 41] in the special case where $c = 0$ (the so-called balanced case). These results have been generalized by [73].

Assuming that $F \in \mathrm{C}^3(\mathbb{R})$ satisfies (4.3) and the following extra conditions :

$$\begin{cases} F(u) > F(\eta_\mathrm{l}), & \forall u \in (\eta_\mathrm{l}, \eta_\mathrm{r}), \\ F'(u) > 0 \quad \text{or} \quad F(u) > F(\eta_\mathrm{r}), & \forall u \in (\eta_\mathrm{l}, \eta_\mathrm{r}), \end{cases} \tag{4.7}$$

it is shown in [73, Th. 1.1] that there exist a unique $c \in \mathbb{R}$ and an increasing function $\eta \in \mathrm{C}^2(\mathbb{R})$, which is unique up to a translation, that solve (4.1). Conditions (4.3) and (4.7) mean that the potential $F(u)$ has two major wells in $u = \eta_\mathrm{l}$ and $u = \eta_\mathrm{r}$ (the states $\eta_\mathrm{l}$ and $\eta_\mathrm{r}$ are therefore stable), and that its behavior is controlled between these wells ; for example, the potential can have minor wells (see Figure 4.2). The proof of [73] relies on special solutions to (4.4) built in [40,41], which, by homotopy techniques, are used to find the solutions to the general case. The result of [73] will be our starting point for proving the global asymptotic stability of the traveling waves of (4.5) for $L = -|\partial_x|$.

Additionally, the authors of [40,41,73] have also studied some properties of the solutions to (4.4). In particular, when $L = -|\partial_x|$, they have shown that $\eta' > 0$ and that there exist constants $B > A > 0$ such that, for all $|x| \geq 1$,

$$A|x|^{-2} \leq \eta'(x) \leq B|x|^{-2}. \tag{4.8}$$

See [41, Th. 2.7] for the special case where $c = 0$ and [73, Prop. 3.2] for the general case. Moreover, the following identity is proved (see [73, Prop. 4.1]) :

$$c = [F(\eta_\mathrm{r}) - F(\eta_\mathrm{l})] \left( \int_{\mathbb{R}} |\eta'|^2 \right)^{-1}. \tag{4.9}$$

Formula (4.9) is useful because it provides the sign of $c$ just by considering the values $F(\eta_\mathrm{l})$ and $F(\eta_\mathrm{r})$.

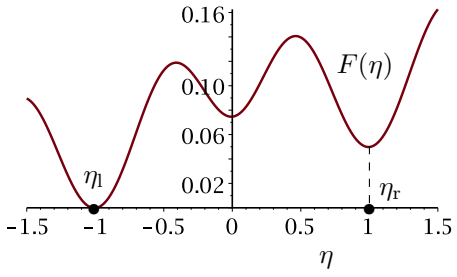



FIGURE 4.2 – A double-well "camel-hump" potential $F$, $\eta_\mathrm{l} = -1$, $\eta_\mathrm{r} = 1$.

As remarked in [40], if $c = 0$, then (4.1) can be interpreted as the restriction to the boundary of an elliptic problem with Neumann boundary condition. If indeed $u$ solves the following problem :

$$\begin{cases} \Delta u(x,y) + c\partial_y u(x,y) = 0 & \text{in} \quad \mathbb{R} \times \mathbb{R}_+, \\ \partial_y u(x,0) = F'(u(x,0)) & \text{on} \quad \mathbb{R} \times \{0\}, \end{cases} \tag{4.10}$$

for $c = 0$, then $\eta(x) := u(x, 0)$ is a solution to (4.1). However, we stress that, when $c \neq 0$, Equation (4.10) describes a diffusive traveling wave in the half-space. In this case, (4.1) is *not* the restriction to the boundary of the problem (4.10), which instead reads

$$(-\Delta - c\partial_x)^{1/2}\,\eta(x) = -F'(\eta(x)). \tag{4.11}$$

We refer to [39] for a mathematical study of (4.11).

But, we mention for completeness that (4.1) is in fact the restriction to the boundary of the following elliptic equation

$$\begin{cases} \Delta u(x,y) = 0 & \text{in} \quad \mathbb{R}\times\mathbb{R}_+, \\ \partial_y u(x,0) + c\partial_x u(x,0) = F'(u(x,0)) & \text{on} \quad \mathbb{R}\times\{0\}. \end{cases} \tag{4.12}$$

In a physical context, the latter is envisioned as an elastic equation in the half-plane with a nonlinear boundary condition. We briefly justify it. If indeed we take the Fourier transform with respect to $x$, denoted as $\mathcal{F}_x$, of the first equation of (4.12), and if we restrict on bounded solutions, we obtain

$$\mathcal{F}_x\left\{u(\cdot,y)\right\}(k) = \mathrm{e}^{-|k|y}\mathcal{F}_x\left\{u(\cdot,0)\right\}(k).$$

Injecting the above information in the second equation of (4.12) then yields (4.1) if we denote $\eta(x) = u(x, 0)$ (recall that $|\partial_x|$ is an operator which has $|k|$ as Fourier symbol).

**Main results** Our first result concerns the asymptotic expansion of the solution $\eta$ to (4.1). The following proposition is a refinement of results of [40, 41, 73] :

**Proposition 4.1.1.** *Let $\eta_{\mathrm{l}} < \eta_{\mathrm{r}}$ and $F \in \mathrm{C}^3(\mathbb{R})$ satisfy (4.3). Assume that $(\eta, c)$ is a solution to (4.1) and (4.2) such that $\eta \in \mathrm{C}^2(\mathbb{R})$ is an increasing function satisfying $\eta' > 0$ and (4.8). Then $\eta$ has the following asymptotes :*

$$\eta(x) - \eta_{\mathrm{r}} \underset{x\to+\infty}{\sim} \frac{\eta_{\mathrm{l}} - \eta_{\mathrm{r}}}{\pi F''(\eta_{\mathrm{r}})} x^{-1}, \qquad \text{and} \qquad \eta(x) - \eta_{\mathrm{l}} \underset{x\to-\infty}{\sim} \frac{\eta_{\mathrm{l}} - \eta_{\mathrm{r}}}{\pi F''(\eta_{\mathrm{l}})} x^{-1}. \tag{4.13}$$

In addition to their theoretical interest, these asymptotes also allow for getting more accurate numerical approximations of $\eta$, as shown in [88].

Our second result is :

**Proposition 4.1.2.** *Under the hypotheses of Proposition 4.1.1, $(c, \eta)$ satisfies the following identity :*

$$c = \frac{1}{\eta_{\mathrm{r}} - \eta_{\mathrm{l}}} \lim_{R\to+\infty} \int_{-R}^{R} F'(\eta). \tag{4.14}$$

The above identity is formally obtained by integrating Equation (4.1) over $\mathbb{R}$ ; we rigorously prove it. Notice that, by Proposition 4.1.1 and using a Taylor expansion, $F'(\eta) \notin \mathrm{L}^1(\mathbb{R})$.

As mentioned above in the concise form (4.5), we consider the following dynamical system :

$$\begin{cases} \partial_t u(t,x) + |\partial_x|\, u(t,x) = -F'(u(t,x)) & \text{for} \quad x \in \mathbb{R}, \\ u(0,x) = u_0(x) & \text{for} \quad x \in \mathbb{R}, \end{cases} \tag{4.15}$$

for an initial condition $u_0 \in \mathrm{L}^\infty\left(\mathbb{R}^d\right)$. We say that $u \in \mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+, \mathrm{L}^\infty\left(\mathbb{R}^d\right)\right)$ is a weak solution to (4.15) if, for all $T > 0$, for all $\phi \in \mathrm{C}^1_{\mathrm{c}}\left([0,T), \mathrm{C}^\infty_{\mathrm{c}}(\mathbb{R})\right)$, there holds

$$\begin{aligned} &\int_0^T \int_{\mathbb{R}} u(t,x)\left(-\partial_t + |\partial_x|\right)\phi(t,x)\mathrm{d}x\,\mathrm{d}t \\ &= \int_{\mathbb{R}} u_0(x)\phi(0,x)\mathrm{d}x - \int_0^T \int_{\mathbb{R}} F'(u(t,x))\phi(t,x)\mathrm{d}x\,\mathrm{d}t. \end{aligned} \tag{4.16}$$

Our third and final result is that (4.1) is the long-time limit of (4.15), for general initial conditions $u_0$ with suitable behavior at infinity (see Figure 4.3 for an example). We prove the following :

**Theorem 4.1.1.** *Let $\eta_{\mathrm{l}} < \eta_{\mathrm{r}}$, $F \in \mathrm{C}^3(\mathbb{R})$ satisfy (4.3) and $\Delta_0 > 0$ be such that*

$$F'' > 0 \quad \text{on} \quad [\eta_{\mathrm{l}} - \Delta_0, \eta_{\mathrm{l}} + \Delta_0] \cup [\eta_{\mathrm{r}} - \Delta_0, \eta_{\mathrm{r}} + \Delta_0]. \tag{4.17}$$

*Assume that $u_0 \in \mathrm{L}^\infty(\mathbb{R})$ takes values in $[\eta_{\mathrm{l}} - \Delta_0, \eta_{\mathrm{r}} + \Delta_0]$ and satisfies*

$$\limsup_{x \to -\infty} u_0(x) < \eta_{\mathrm{l}} + \Delta_0 \quad \text{and} \quad \liminf_{x \to +\infty} u_0(x) > \eta_{\mathrm{r}} - \Delta_0. \tag{4.18}$$

*Then :*

*(i) Equation (4.15) has a unique weak solution $u \in \mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+, \mathrm{L}^\infty(\mathbb{R})\right)$. Moreover, for all $T_0 > 0$,*

$$u \in \mathrm{C}\left((T_0, +\infty), \mathrm{C}^2(\mathbb{R})\right) \cap \mathrm{C}^1\left((T_0, +\infty), \mathrm{C}(\mathbb{R})\right). \tag{4.19}$$

*(ii) In addition, for all $t > 0$ and $x \in \mathbb{R}$, $u(t,x) \in [\eta_{\mathrm{l}} - \Delta_0, \eta_{\mathrm{r}} + \Delta_0]$.*

*(iii) Assume that $(\eta, c)$ is a solution to (4.1) and (4.2) such that $\eta \in \mathrm{C}^2(\mathbb{R})$ is an increasing function satisfying $\eta' > 0$ and*

$$\lim_{|x| \to +\infty} \eta'(x) = 0. \tag{4.20}$$

*Then, there exist constants $\kappa > 0$, $K > 0$ and $\xi$ such that*

$$\|u(t,\cdot) - \eta(\cdot - ct + \xi)\|_{\mathrm{L}^\infty(\mathbb{R})} \leq K\mathrm{e}^{-\kappa t}, \tag{4.21}$$

*for all $t \in \mathbb{R}_+$. In the above estimate, $\kappa$ only depends on $F$, $\eta$ and $\Delta_0$, whereas $K$ and $\xi$ depend also on $u_0$.*

Theorem 4.1.1 suggests that simulating (4.15) is sufficient to obtain in the long time a numerical approximation of the solution $(\eta, c)$ to (4.1). In this regard, it is significant that $c$, which is an unknown of (4.1), does *not* appear in (4.15). This in particular allows for constructing an approximation of the traveling wave velocity, which is unknown before the end of the simulation. We refer the reader to our study [88], where we explain the details of the numerical strategy, and to a forthcoming article [132] for the multi-dimensional case. In this regard, we shall stress that Theorem 4.1.1 stated above unfortunately only holds in the case where $\eta$ is scalar-valued, because its proof relies on a comparison principle. However, it empirically appears that such a convergence is also achieved in many cases where $\eta$ is vector-valued.

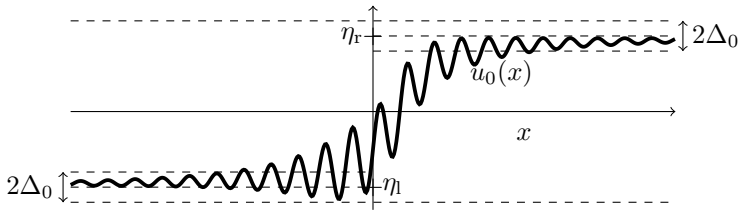



FIGURE 4.3 – A possible initial condition $u_0$ in Theorem 4.1.1.

**Outline** Our contribution is organized as follows. In Section 4.2, we introduce notations and give essential properties of the operator $|\partial_x|$. In Section 4.3, we prove Propositions 4.1.1 and 4.1.2. In Section 4.4, we justify the existence and the uniqueness of a weak solution to Equation (4.15), which satisfies (4.19), thus establishing (i) of Theorem 4.1.1. In Section 4.5, we use Chen's approach for proving (ii) and (iii) of Theorem 4.1.1. The key ingredients are a comparison principle and specific sub-solutions and super-solutions. Although we could check the technical assumptions and apply Chen's theorem, we prefer to restrict Chen's proof to our special case for self-consistency and simplicity.

## 4.2 Notations and definitions

**Notations** We denote by $\mathrm{C}_{\mathrm{c}}^{\infty}(\mathbb{R})$ the space of smooth functions with compact supports in $\mathbb{R}$ and by $\mathcal{D}'$ the space of distributions over $\mathbb{R}$. For $u \in \mathrm{C}_{\mathrm{c}}^{\infty}(\mathbb{R})$, we denote the Fourier transform by $\mathcal{F}\{u\}(k) := \int_{\mathbb{R}} \mathrm{e}^{-ikx} u(x) \mathrm{d}x$. For two functions $u$ and $v$, we denote by $*$ the convolution. Henceforth, the Fourier transform and the convolution are only taken with respect to the *space* variable $x$ (and never with respect to the *time* variable $t$). We make use of the principal value of $\frac{1}{x}$, denoted by $\mathrm{p.v.}\left(\frac{1}{x}\right)$, which is the distribution defined by

$$\left\langle \mathrm{p.v.}\left(\frac{1}{x}\right), u \right\rangle = \lim_{\varepsilon \to 0^+} \left\{ \int_{\varepsilon}^{+\infty} \frac{u(y)}{y} \mathrm{d}y + \int_{-\infty}^{-\varepsilon} \frac{u(y)}{y} \mathrm{d}y \right\},$$

for $u \in \mathrm{C}_{\mathrm{c}}^{\infty}(\mathbb{R})$.

**Definition and properties of the operator $|\partial_x|$** For convenience, we recall some elementary properties of the operator $|\partial_x|$. The Hilbert transform $\mathcal{H}$ of $u \in \mathrm{L}^2(\mathbb{R})$ is defined by

$$\mathcal{H}\{u\} := \mathcal{F}^{-1}\{-\mathrm{i}\,\mathrm{sign}(k)\mathcal{F}\{u\}(k)\}. \tag{4.22}$$

It is immediate that, if $u \in \mathrm{L}^2(\mathbb{R})$, then $\mathcal{H}^2\{u\} = -u$. Next, the operator $|\partial_x|$ is defined as

$$|\partial_x|\, u(x) := \mathcal{H}\left\{u'\right\}(x) = \mathcal{F}^{-1}\{|k|\mathcal{F}\{u\}(k)\}(x), \tag{4.23}$$

for $u \in \mathrm{C}_{\mathrm{c}}^{\infty}(\mathbb{R})$. As $\mathcal{F}\{\mathrm{p.v.}(1/x)\}(k) = -\mathrm{i}\pi\mathrm{sign}(k)$, the operator $|\partial_x|$ can be rewritten as

$$|\partial_x|\, u(x) = -\frac{1}{\pi}\int_0^{+\infty} \frac{u'(x+y) - u'(x-y)}{y}\mathrm{d}y \tag{4.24}$$

$$= -\frac{1}{\pi}\int_0^{+\infty} \frac{u(x+y) - 2u(x) + u(x-y)}{y^2}\mathrm{d}y, \tag{4.25}$$

the last expression being obtained from the previous one by integrating by parts. We see from (4.23) that the operator $|\partial_x|$ is symmetric and positive, like the Laplacian. But, unlike the Laplacian, it is clear from (4.25) that $|\partial_x|\, u(x)$ does not only depend on $u$ in the neighborhood of $x$ but also on each value $u(y)$, for $y \in \mathbb{R}$; put differently, $|\partial_x|$ is *non-local*.

A straightforward computation yields that $|\partial_x|\,\phi \in \mathrm{L}^1(\mathbb{R})$ whenever $\phi \in \mathrm{C}_{\mathrm{c}}^{\infty}(\mathbb{R})$. Hence, one can extend $|\partial_x|$ over $\mathrm{L}^{\infty}(\mathbb{R})$ by duality, defining $|\partial_x|\, u$ as the following distribution :

$$|\partial_x|\, u : \phi \in \mathrm{C}_{\mathrm{c}}^{\infty}(\mathbb{R}) \mapsto \int_{\mathbb{R}} u(y)\,|\partial_x|\,\phi(y)\mathrm{d}y. \tag{4.26}$$

When $u$ is sufficiently regular, explicit expressions for $|\partial_x|\, u$ are available. Namely, if $u \in \mathrm{L}^{\infty}(\mathbb{R}) \cap \mathrm{C}^2(\mathbb{R})$, then Expression (4.25) is valid. In particular, $|\partial_x|\, u \in \mathrm{C}(\mathbb{R}) \cap \mathrm{L}^{\infty}(\mathbb{R})$. The proof can be done by density of $\mathrm{C}_{\mathrm{c}}^{\infty}(\mathbb{R})$ in $\mathrm{C}^2_{\mathrm{loc}}(\mathbb{R})$, using the fact that (4.25) is true for smooth functions. If we assume furthermore that $u' \in \mathrm{L}^1(\mathbb{R})$, then Expression (4.24) is also valid; this is deduced from (4.25) by integration by parts.

## 4.3 Asymptotes and an identity about velocity

The proof of Proposition 4.1.1 relies on the asymptotic behavior of Cauchy integrals (see [119, p. 267]) and involves the following technical lemma :

**Lemma 4.3.1.** *Under the hypotheses of Proposition 4.1.1, there holds*

$$\eta'' \in \mathrm{L}^{\infty}(\mathbb{R}). \tag{4.27}$$

*Remark* 1. Note that it is also possible to establish by technical arguments that there exists a constant $C > 0$ such that, for all $|x| > 1$,

$$\left|\eta''(x)\right| \leq C|x|^{-2}\left(1 + \ln(|x|)\right). \tag{4.28}$$

(We refer the reader to Annex A.4 for the proof of (4.28)). However, (4.27) is sufficient to prove Proposition 4.1.1. Yet, if $F$ is sinusoidal, one can derive analytical solutions $\eta$ to (4.1), as is shown in [135], which are of the form

$$\eta(x) = \eta_{\rm l} + \frac{\eta_{\rm r} - \eta_{\rm l}}{\pi}\left(\frac{\pi}{2} + \arctan\left(ax\right)\right),$$

for $a > 0$. Whence

$$\eta''(x) = -\frac{\eta_{\rm r} - \eta_{\rm l}}{\pi}\frac{2a^3x}{\left(a^2x^2+1\right)^2} \underset{x\to+\infty}{\sim} -\frac{2(\eta_{\rm r} - \eta_{\rm l})}{\pi a x^3}.$$

Thus (4.28) is probably not optimal.

We postpone the proof of Lemma 4.3.1 until the end of the proof of Proposition 4.1.1 and temporarily admit Lemma 4.3.1.

*Proof of Proposition 4.1.1.* We focus on the case $x \to +\infty$. Provided that

$$x\left|\partial_x\right|\eta(x) \underset{x\to+\infty}{\longrightarrow} \frac{1}{\pi}\int_{-\infty}^{+\infty}\eta'(y)\mathrm{d}y = \frac{1}{\pi}\left(\eta_{\rm r} - \eta_{\rm l}\right), \tag{4.29}$$

then, using (4.1), (4.2), (4.3) and (4.8), we get by Taylor expansion

$$F''(\eta_{\rm r})\left(\eta(x) - \eta_{\rm r}\right) \underset{x\to+\infty}{\sim} -\frac{1}{\pi}x^{-1}\left(\eta_{\rm r} - \eta_{\rm l}\right),$$

which is (4.13).

Let us now prove (4.29). By assumption, $\eta' \in \mathrm{C}^1(\mathbb{R})$, and by (4.8), we have $\eta' \in \mathrm{L}^1(\mathbb{R})$. As a consequence, there holds

$$x\left|\partial_x\right|\eta(x) = \frac{x}{\pi}\lim_{\varepsilon\to0^+}\left(\int_{-\infty}^{x-\varepsilon} + \int_{x+\varepsilon}^{+\infty}\right)\frac{\eta'(y)}{x-y}\mathrm{d}y.$$

Let $R > 0$ and $x > 2R$. We split the integral into three parts

$$\begin{aligned}\pi x\left|\partial_x\right|\eta(x) = &\int_{-\infty}^{R}\frac{\eta'(y)}{1-y/x}\mathrm{d}y + x\left(\int_{R}^{x-R} + \int_{x+R}^{+\infty}\right)\frac{\eta'(y)}{x-y}\mathrm{d}y\\ &+ x\lim_{\varepsilon\to0^+}\left(\int_{x-R}^{x-\varepsilon} + \int_{x+\varepsilon}^{x+R}\right)\frac{\eta'(y)}{x-y}\mathrm{d}y.\end{aligned} \tag{4.30}$$

The first right-hand term in (4.30) is dealt with by using the dominated convergence theorem, the second one avoids the singularity of $|x-y|^{-1}$ and is bounded thanks to (4.8), and the third one is on the singularity of $|x-y|^{-1}$ and is controlled thanks to (4.27) and (4.8).

As $\eta' \in \mathrm{L}^1(\mathbb{R})$ and since (recall that $x > 2R$)

$$\left|\frac{\eta'(y)}{1-y/x}\right| \le 2\left|\eta'(y)\right| \quad \text{if} \quad y < R, \quad \text{and} \quad \frac{\eta'(y)}{1-y/x} \underset{x\to+\infty}{\longrightarrow} \eta'(y),$$

then, by the dominated convergence theorem

$$\int_{-\infty}^{R} \frac{\eta'(y)}{1-y/x}\mathrm{d}y \underset{x\to+\infty}{\longrightarrow} \int_{-\infty}^{R} \eta'(y)\mathrm{d}y. \tag{4.31}$$

Next, we split the second integral of (4.30) into three parts. Invoking (4.8), we deduce that, as $|x| > 2R$,

$$\begin{aligned}
&\left|\left(\int_{R}^{x-R} + \int_{x+R}^{+\infty}\right) \frac{\eta'(y)}{x-y}\mathrm{d}y\right| \\
&\leq Cx^{-1}\int_{R}^{x/2} \left|\eta'(y)\right| \mathrm{d}y + CR^{-1}\left(\int_{x/2}^{x-R} + \int_{x+R}^{+\infty}\right) \left|\eta'(y)\right| \mathrm{d}y \\
&\leq Cx^{-1}\int_{R}^{x/2} \frac{\mathrm{d}y}{y^2} + CR^{-1}\left(\int_{x/2}^{x-R} + \int_{x+R}^{+\infty}\right) \frac{\mathrm{d}y}{y^2} \\
&\leq CR^{-1}x^{-1}.
\end{aligned} \tag{4.32}$$

Last, we split the last integral of (4.30) into two parts, namely :

$$\left|\left(\int_{x-R}^{x-\varepsilon} + \int_{x+\varepsilon}^{x+R}\right) \frac{\eta'(y)}{x-y}\mathrm{d}y\right| = \left|\left(\int_{\varepsilon}^{x^{-2}} + \int_{x^{-2}}^{R}\right) \frac{\eta'(x-z)-\eta'(x+z)}{z}\mathrm{d}z\right|.$$

The first part of the right-hand side of the above equation is dealt with by using (4.27), and the second one by using (4.8). Whence, as $x > 2R$ and for $x^{-2} \geq \varepsilon$,

$$\begin{aligned}
\left|\left(\int_{x-R}^{x-\varepsilon} + \int_{x+\varepsilon}^{x+R}\right) \frac{\eta'(y)}{x-y}\mathrm{d}y\right| &\leq Cx^{-2}\int_{x^{-2}}^{R} z^{-1}\mathrm{d}z + C\int_{\varepsilon}^{x^{-2}} \mathrm{d}z \\
&\leq Cx^{-2}\left(\ln(Rx^2)+1\right).
\end{aligned}$$

Therefore

$$\left|x \lim_{\varepsilon\to 0^+}\left(\int_{x-R}^{x-\varepsilon} + \int_{x+\varepsilon}^{x+R}\right) \frac{\eta'(y)}{x-y}\mathrm{d}y\right| \leq Cx^{-1}\left(\ln(Rx^2)+1\right). \tag{4.33}$$

Convergence (4.31) and Estimates (4.32) and (4.33) finally yield

$$\limsup_{x\to+\infty} \left|\pi x\left|\partial_x\right|\eta(x) - \int_{-\infty}^{+\infty}\eta'(y)\mathrm{d}y\right| \leq \left|\int_{R}^{+\infty}\eta'(y)\mathrm{d}y\right| + CR^{-1},$$

which, thanks to (4.8), implies (4.29) upon letting $R \to +\infty$. □

We then proceed with the :

*Proof of Lemma 4.3.1.* We first remark that if $g \in \mathrm{L}^2(\mathbb{R})$ and if

$$-\left|\partial_x\right| h(x) + ch'(x) = g(x), \tag{4.34}$$

then $h' \in \mathrm{L}^2(\mathbb{R})$. Indeed, the Fourier transform turns (4.34) into

$$(-|k| + cik)\mathcal{F}\{h\}(k) = \mathcal{F}\{g\}(k).$$

Therefore, $k\mathcal{F}\{h\}(k) \in \mathrm{L}^2(\mathbb{R})$, whence $h' \in \mathrm{L}^2(\mathbb{R})$. We use this result to prove that $\eta'' \in \mathrm{L}^\infty(\mathbb{R})$.

Upon differentiating (4.1), we obtain

$$-|\partial_x|\,\eta'(x) + c\eta''(x) = F''(\eta(x))\eta'(x). \tag{4.35}$$

As $\eta \in \mathrm{L}^\infty(\mathbb{R})$, $F \in \mathrm{C}^3(\mathbb{R})$ and $\eta' \in \mathrm{L}^2(\mathbb{R})$ (thanks to (4.8)), then the right-hand side of (4.35) is in $\mathrm{L}^2(\mathbb{R})$. Therefore $\eta'' \in \mathrm{L}^2(\mathbb{R})$. Differentiating (4.35) yields

$$-|\partial_x|\,\eta''(x) + c\eta'''(x) = F''(\eta(x))\eta''(x) + F'''(\eta(x))\left(\eta'(x)\right)^2. \tag{4.36}$$

As $\eta \in \mathrm{L}^\infty(\mathbb{R})$, $F \in \mathrm{C}^3(\mathbb{R})$, $\eta'' \in \mathrm{L}^2(\mathbb{R})$, and, thanks to (4.8), $(\eta')^2 \in \mathrm{L}^2(\mathbb{R})$, then the right-hand side of (4.36) is in $\mathrm{L}^2(\mathbb{R})$. Therefore $\eta''' \in \mathrm{L}^2(\mathbb{R})$. As a consequence, since $\eta'' \in \mathrm{L}^2(\mathbb{R})$, we deduce by Sobolev injection that $\eta'' \in \mathrm{L}^\infty(\mathbb{R})$, whence (4.27).

□

We now focus on Proposition 4.1.2. Both the Identities (4.9) and (4.14) are formally obtained by testing (4.1) against a certain function $g$, namely $g = \eta'$ for (4.9), and $g = 1$ for (4.14). We justify below this formal integration.

*Proof of Proposition 4.1.2.* Let $R > 2$. We integrate (4.1) over $[-R, R]$ :

$$-\int_{-R}^{R} |\partial_x|\,\eta(x)\mathrm{d}x + c\left(\eta(R) - \eta(-R)\right) = \int_{-R}^{R} F'(\eta(x))\mathrm{d}x. \tag{4.37}$$

Thus, Identity (4.14) stems from (4.37), provided that

$$\lim_{R\to+\infty} \int_{-R}^{R} |\partial_x|\,\eta(x)\mathrm{d}x = 0. \tag{4.38}$$

We prove (4.38) using (4.8) and (4.27). As $\eta' \in \mathrm{L}^1(\mathbb{R}) \cap \mathrm{C}^1(\mathbb{R})$, there holds

$$\begin{aligned} |\partial_x|\,\eta(x) =& \frac{1}{\pi}\int_0^{+\infty} \frac{\eta'(x-y) - \eta'(x+y)}{y}\mathrm{d}y \\ =& \frac{1}{\pi}\int_{|y|<R} \frac{(\eta'(x-y) - \eta'(x))}{y}\mathrm{d}y + \frac{1}{\pi}\int_{|y|>R} \frac{\eta'(x-y)}{y}\mathrm{d}y. \end{aligned} \tag{4.39}$$

Remark that

$$\int_{|y|>R} \left|\frac{\eta'(x-y)}{y}\right| \mathrm{d}y \le 2R^{-1}\left\|\eta'\right\|_{\mathrm{L}^1(\mathbb{R})},$$

and that, using (4.27),

$$\int_{|y|<R}\left|\frac{(\eta'(x-y)-\eta'(x))}{y}\right|\mathrm{d}y \leq 2R\left\|\eta''\right\|_{\mathrm{L}^\infty(\mathbb{R})} \leq CR.$$

Therefore, integrating (4.39) thanks to Fubini's theorem yields

$$\begin{aligned}\int_{-R}^{R}|\partial_x|\,\eta(x)\mathrm{d}x =& \frac{1}{\pi}\int_{|y|<R}\frac{(\eta(R-y)-\eta(R))-(\eta(-R-y)-\eta(-R))}{y}\mathrm{d}y\\ &+\frac{1}{\pi}\int_{|y|>R}\frac{\eta(R-y)-\eta(-R-y)}{y}\mathrm{d}y\\ =&: T_1+T_2.\end{aligned} \tag{4.40}$$

First, we bound $T_1$. If $|y|<R/2$, thanks to (4.8), we obtain

$$|\eta(R-y)-\eta(R)|\leq C|y|R^{-2} \quad \text{and} \quad |\eta(-R-y)-\eta(-R)|\leq C|y|R^{-2}.$$

As a consequence,

$$\int_{|y|<R/2}\left|\frac{(\eta(R-y)-\eta(R))-(\eta(-R-y)-\eta(-R))}{y}\right|\mathrm{d}y \leq CR^{-1}. \tag{4.41}$$

Note that, as underlined in [73], a consequence of (4.8) is that

$$\begin{cases} A|x|^{-1}\leq \eta_{\mathrm{r}}-\eta(x)\leq B|x|^{-1}, & \text{if} \quad x>1,\\ A|x|^{-1}\leq \eta(x)-\eta_{\mathrm{l}}\leq B|x|^{-1}, & \text{if} \quad x<-1. \end{cases} \tag{4.42}$$

Therefore, if $|y|<R$

$$|\eta(R-y)-\eta(R)|\leq \frac{C}{R-|y|+1} \quad \text{and} \quad |\eta(-R-y)-\eta(-R)|\leq \frac{C}{R-|y|+1}.$$

Whence

$$\begin{aligned}&\int_{R/2<|y|<R}\left|\frac{(\eta(R-y)-\eta(R))-(\eta(-R-y)-\eta(-R))}{y}\right|\mathrm{d}y\\ &\leq \frac{C}{R}\int_{R/2}^{R}\frac{\mathrm{d}y}{R+1-y}\leq CR^{-1}\ln(R).\end{aligned} \tag{4.43}$$

We deduce from (4.41) and (4.43) that

$$|T_1(R)|\leq CR^{-1}(1+\ln(R)). \tag{4.44}$$

Thanks to (4.8) and since $\eta\in\mathrm{L}^\infty(\mathbb{R})$, if $|y|>R$, we have

$$|\eta(R-y)-\eta(-R-y)|\leq C\min\left\{R\left(|y|-R\right)^{-2},1\right\}.$$

Whence, splitting $T_2$ into two parts,

$$\begin{aligned} |T_2(R)| \leq & C\left(\int_R^{R+\sqrt{R}} \frac{\mathrm{d}y}{y} + \int_{R+\sqrt{R}}^{+\infty} \frac{R}{y(y-R)^2}\mathrm{d}y\right) \\ \leq & C\frac{\sqrt{R}}{R} + C\int_{\sqrt{R}}^{+\infty} \frac{\mathrm{d}z}{z^2} \leq CR^{-1/2}. \end{aligned} \tag{4.45}$$

Bearing (4.40) in mind, we observe that (4.44) and (4.45) imply (4.38). □

## 4.4 Existence, uniqueness and regularity of the solution to the evolution equation (4.15)

We now justify the existence, the uniqueness and the regularity of a weak solution $u$ to (4.15). We proceed in the classical way ; as the methods as well as the type of results are well-known, we only give a few hints of proofs. We refer the interested reader to Annex A.4 for some technical details and extra materials about the proofs, and to [43] for a reference on evolution equations involving m-dissipative operators.

Using the Fourier transform, the solution to the homogeneous linear equation

$$\partial_t u(t,x) + |\partial_x|\, u(t,x) = 0 \quad \text{for} \quad x \in \mathbb{R}, \qquad \text{and} \quad u(0,x) = u_0(x) \tag{4.46}$$

is given by $u(t,x) = \{K_t * u_0\}(x)$, where the kernel $K_t$ is defined by

$$K_t(x) = \frac{t}{\pi(t^2+x^2)} \qquad \text{if} \quad t > 0 \quad \text{and} \quad K_0 = \delta_0, \tag{4.47}$$

the Fourier transform of which is $\mathrm{e}^{-|k|t}$. Before getting to the inhomogeneous linear equation, we underline some interesting properties of the kernel $K_t$. First, for all $t \geq 0$, $K_t$ is a probability measure. Moreover, for all $t > 0$, $K_t$ is a smooth function. In particular, the space derivative of $K_t$ satisfies

$$\left\| \frac{\mathrm{d}}{\mathrm{d}x} K_t \right\|_{\mathrm{L}^1(\mathbb{R})} \leq Ct^{-1}, \tag{4.48}$$

where $C$ is a universal constant. In all these aspects, $K_t$ is similar to the Gaussian kernel $\mathcal{K}_t(x) = \mathrm{e}^{-\frac{x^2}{2t^2}}/(t\sqrt{2\pi})$.

The semi-group generated by $K_t$ allows for solving the inhomogeneous equation

$$\begin{cases} \partial_t u(t,x) + |\partial_x|\, u(t,x) = g(t,x) & \text{for} \quad x \in \mathbb{R}, \\ u(0,x) = u_0(x) & \text{for} \quad x \in \mathbb{R}. \end{cases} \tag{4.49}$$

Indeed, let $T > 0$, $u_0 \in \mathrm{L}^\infty(\mathbb{R})$ and $g \in \mathrm{L}^\infty([0,T]\times\mathbb{R})$. Then there exists a unique weak solution $u \in \mathrm{L}^\infty([0,T]\times\mathbb{R})$ to (4.49) in the sense that, for all $\phi \in \mathrm{C}^1_{\mathrm{c}}([0,T), \mathrm{C}^\infty_{\mathrm{c}}(\mathbb{R}))$, the

following identity holds :

$$\int_0^T \int_{\mathbb{R}} u(t,x) \left(-\partial_t + |\partial_x|\right) \phi(t,x) \mathrm{d}x\mathrm{d}t - \int_{\mathbb{R}} u_0(x)\phi(0,x)\mathrm{d}x = \int_0^T \int_{\mathbb{R}} g(t,x)\phi(t,x)\mathrm{d}x\mathrm{d}t. \tag{4.50}$$

This solution can be written thanks to the Duhamel formula as

$$u(t,x) = K_t * u_0(x) + \left\{ \int_0^t K_{t-s} * g(s,\cdot)\mathrm{d}s \right\}(x), \tag{4.51}$$

with the convention that $K_0 * u_0 = u_0$, even if $u_0$ is not regular. The existence of a solution $u$ to (4.49) is a consequence of the fact that (4.51) is well-defined ; its uniqueness is shown by using the adjoint problem of (4.49).

We now turn to the semi-linear equation (4.15). Let $F \in \mathrm{W}^{2,\infty}(\mathbb{R})$ and $u_0 \in \mathrm{L}^\infty(\mathbb{R})$. Then there exists a unique weak solution $u \in \mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+, \mathrm{L}^\infty(\mathbb{R})\right)$ to (4.15). Moreover, $u$ can be expressed as

$$u(t,x) = \left\{K_t * u_0\right\}(x) - \left\{ \int_0^t K_{t-s} * F'(u(s,\cdot))\mathrm{d}s \right\}(x). \tag{4.52}$$

The proof is done by a classical fixed-point argument on (4.52) (see for example [43, Sec. 4.3 p. 56]).

Finally, we justify that the evolution equation (4.15) has a regularizing effect ; in other words, the weak solution to (4.15) becomes instantly a classical solution. Assume indeed that $F \in \mathrm{C}^3(\mathbb{R}) \cap \mathrm{W}^{3,\infty}(\mathbb{R})$ and $u_0 \in \mathrm{L}^\infty(\mathbb{R})$, and let $u \in \mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+, \mathrm{L}^\infty\left(\mathbb{R}^d\right)\right)$ be the weak solution to (4.15). Then, for all $T_0 > 0$, we have (4.19). Therefore, for all $t > 0$, there holds

$$\partial_t u(t,x) + |\partial_x|\, u(t,x) = -F'(u(t,x)), \tag{4.53}$$

in the strong sense. Finally

$$u \in \mathrm{C}\left([0,+\infty[\,, \text{weak-}*\text{-}\mathrm{L}^\infty(\mathbb{R})\right). \tag{4.54}$$

The proof of (4.19) relies on an iterative argument based on (4.52), using the fact that, for all $t > 0$, $K_t$ is a smooth probability measure that satisfies (4.48). Last, a straightforward adaptation of the proof of [36, Ex. 4.24 p. 126] yields (4.54).

Note that, similarly to [1, Th. 2.9.], $t \mapsto u(t,\cdot)$ may not be continuous at 0 in $\mathrm{L}^\infty(\mathbb{R})$. If indeed $u_0$ is discontinuous, then $u(t,\cdot)$ cannot tend to $u_0$ in $\mathrm{L}^\infty(\mathbb{R})$ when $t \to 0$ because, invoking (4.19), $u(t,\cdot)$ is a continuous function for all $t > 0$.

## 4.5 Convergence of the evolution equation (4.15) to the Weertman equation (4.1)

In this section, we prove (ii) and (iii) of Theorem 4.1.1. The proof can be summarized in two steps : first, we show that (4.15) satisfies a comparison principle, then we use Chen's

method of squeezing, establishing respectively (ii) and (iii) of Theorem 4.1.1. For the sake of self-consistency, simplicity and conciseness (Chen's theory being quite general), we prefer to restrict the whole proof of [45, Th. 3.1] to our specific case rather than to check that the hypotheses of Chen's theory are satisfied (precisely Hypotheses (A1), (A2), (A3), (B1), (B2) and (B3) of [45], which are indeed satisfied in our case).

We henceforth assume that $F \in \mathrm{C}^3(\mathbb{R}) \cap \mathrm{W}^{3,\infty}(\mathbb{R})$ satisfies (4.3) and (4.17). We introduce the non-linear operator $\mathcal{A}[u] := -|\partial_x| u - F'(u)$, of which we now discuss some immediate properties. By the results of Section 4.4, $\mathcal{A}$ generates a semi-group on the Banach space $\mathrm{L}^\infty\left(\mathbb{R}^d\right)$. $\mathcal{A}$ is translation invariant ; namely, for all $h \in \mathbb{R}$, and for any function $u(x)$, there holds

$$\mathcal{A}[u(h+\cdot)](x) = \mathcal{A}[u](x+h), \qquad \forall x \in \mathbb{R}. \tag{4.55}$$

Moreover, $\mathcal{A}$ maps constant functions to constant functions ; namely

$$\mathcal{A}[\alpha \cdot 1] = -F'(\alpha) \cdot 1,$$

for all $\alpha \in \mathbb{R}$, where 1 above denotes the function identically equal to 1.

The operator $\mathcal{A}$ satisfies the following comparison principle :

**Proposition 4.5.1.** *Let $F \in \mathrm{W}^{2,\infty}(\mathbb{R})$. Let $\overline{u}$ and $\underline{u} \in \mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+, \mathrm{L}^\infty\left(\mathbb{R}^d\right)\right)$ be such that*

$$\begin{cases} \partial_t \underline{u}(t,x) - \mathcal{A}[\underline{u}(t,\cdot)](x) = \underline{g}(t,x) \leq 0, \\ \partial_t \overline{u}(t,x) - \mathcal{A}[\overline{u}(t,\cdot)](x) = \overline{g}(t,x) \geq 0, \end{cases} \tag{4.56}$$

*where $\underline{g}$ and $\overline{g} \in \mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+, \mathrm{L}^\infty\left(\mathbb{R}^d\right)\right)$, and $\underline{u}(0,\cdot) \leq \overline{u}(0,\cdot)$ with $\underline{u}(0,\cdot) \neq \overline{u}(0,\cdot)$ on a non-negligible set. Then, for almost every $t > 0$, $x \in \mathbb{R}$, there holds*

$$\underline{u}(t,x) < \overline{u}(t,x). \tag{4.57}$$

*Remark* 2. Proposition 4.5.1 has an immediate corollary : Assume that $u_0 \in \mathrm{L}^\infty\left(\mathbb{R}^d\right)$ takes values in $[\eta_{\mathrm{l}} - \Delta_0, \eta_{\mathrm{r}} + \Delta_0]$ and let $u \in \mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+, \mathrm{L}^\infty\left(\mathbb{R}^d\right)\right)$ be the unique solution to (4.15) with the initial condition $u_0$. By (4.3), $\overline{u} := x \mapsto \eta_{\mathrm{r}} + \Delta_0$ and $\underline{u} : x \mapsto \eta_{\mathrm{l}} - \Delta_0$ are respectively supersolutions and subsolutions to (4.15). Therefore, Proposition 4.5.1 implies that $u(t,x) \in [\eta_{\mathrm{l}} - \Delta_0, \eta_{\mathrm{r}} + \Delta_0]$, for almost every $t \in \mathbb{R}_+$ and $x \in \mathbb{R}$, thus establishing (ii) of Theorem 4.1.1.

Proposition 4.5.1 is shown thanks to the Duhamel Formula (4.51) and Grönwall's Lemma.

*Proof.* Let $M > \|F''\|_{\mathrm{L}^\infty(\mathbb{R})}$. We set

$$v(t,x) := \mathrm{e}^{Mt}\left(\overline{u}(t,x) - \underline{u}(t,x)\right), \tag{4.58}$$

and prove that $v \geq 0$. In view of (4.56), we have

$$\partial_t v(t,x) + |\partial_x|\, v(t,x) = M v(t,x) - \mathrm{e}^{Mt}\left(F'(\overline{u}(t,x)) - F'(\underline{u}(t,x)) + g(t,x\right),$$

where $g(t,x) := e^{Mt}\left(\overline{g}(t,x) - \underline{g}(t,x)\right) \geq 0$. Since the right-hand side of the latter equation is in $\mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+, \mathrm{L}^\infty\left(\mathbb{R}^d\right)\right)$, then, using (4.51), one can express $v(t,x)$ as

$$v(t,x) = \left\{K_t * v(0,\cdot)\right\}(x) + \left\{ \int_0^t K_{t-s} * \Big[ Mv(s,\cdot) - \mathrm{e}^{Ms}\left(F'(\overline{u}(s,\cdot)) - F'(\underline{u}(s,\cdot)) + g(s,\cdot)\right)\Big]\mathrm{d}s \right\}(x). \tag{4.59}$$

We introduce $v_-(t) := -\text{ess inf}\left\{\min(0, v(t,x)), x \in \mathbb{R}\right\}$. By Taylor expansion, for almost every $(s,y) \in [0,t]\times\mathbb{R}$, there holds

$$\begin{aligned} Mv(s,y) - \mathrm{e}^{Ms}\left(F'(\overline{u}) - F'(\underline{u})\right)(s,y) \geq & Mv(s,y) - \left\|F''\right\|_{\mathrm{L}^\infty\left(\mathbb{R}^d\right)} \mathrm{e}^{Ms}\left|\overline{u}(s,y) - \underline{u}(s,y)\right| \\ \geq & Mv(s,y) - M|v(s,y)| \\ \geq & -2Mv_-(s). \end{aligned} \tag{4.60}$$

Therefore, using (4.59), since $K_t$, $g$ and $v(0,\cdot)$ are nonnegative, and since $K_t$ is a probability measure for all $t \geq 0$, we obtain, for almost every $t \in \mathbb{R}_+$ and $x \in \mathbb{R}$,

$$-v(t,x)) \leq 2M \int_0^t v_-(s)\mathrm{d}s,$$

whence

$$v_-(t) \leq 2M \int_0^t v_-(s)\mathrm{d}s. \tag{4.61}$$

Since $\underline{u}$, $\overline{u} \in \mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+, \mathrm{L}^\infty\left(\mathbb{R}^d\right)\right)$, then, $v_- \in \mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+\right)$. Hence, by Grönwall's Lemma, we deduce from (4.61) that $v_-(t) = 0$, for almost every $t > 0$. Injecting this information in (4.60), and next in (4.59), yields $v(t,x) \geq \left\{K_t * v(0,\cdot)\right\}(x)$. As a consequence, as $K_t$ is positive if $t > 0$ and as $v(0,\cdot) = \overline{u}(0,\cdot) - \underline{u}(0,\cdot)$ is nonnegative and positive on a non-negligible set, we deduce that $v(t,x) > 0$, for almost every $t > 0$ and $x \in \mathbb{R}$. This implies (4.57). □

Then, we establish a stronger version of the comparison principle, the proof of which mimicks that of Proposition 4.5.1 :

**Corollary 4.5.2.** *Under the assumptions of Proposition 4.5.1, there exists a positive decreasing function $\rho$ such that, for all $R > 1$,*

$$\text{ess inf}_{x\in[-R,R]}\left(\overline{u}(1,x) - \underline{u}(1,x)\right) \geq \rho(R) \int_0^1 \left(\overline{u}(0,y) - \underline{u}(0,y)\right)\mathrm{d}y. \tag{4.62}$$

*Proof.* Introducing $v$ defined by (4.58), and using (4.59) and (4.60), we obtain

$$v(t,x) \geq \left\{K_t * v(0,\cdot)\right\}(x),$$

since $v$ is nonnegative. By definition of $K_t$ and of $v$, it implies (4.62). □

The proof of (iii) of Theorem 4.1.1 follows [45, Th. 3.1], the proof of which we restrict here to our particular case. By the previous steps, we already know that there exists a unique weak solution $u(t,x)$ to (4.15) (see Remark 2), and we aim at establishing (4.21).

Namely, we build special sub-solutions and super-solutions to (4.15) that are based on the existing solution to (4.1) (see Lemma 4.5.3 below). Then, we prove that both the vertical and the horizontal distances between these solutions are controlled (respectively $2\delta_j$ and $2l_j$ on Figure 4.5, see also Lemma 4.5.4 below). Using the fact that (4.15) is an autonomous system, we use the established control to iteratively build successive sub-solutions $w^j_{-1}$ and super-solutions $w^j_{+1}$ surrounding the actual solution $u(t_j,x)$ to (4.15). At each step $j$, the distance between these sub-solutions and super-solutions is lowered. Thus, the solution $u$ is *squeezed* between these sub-solutions and super-solutions. As a consequence, when $t$ goes to infinity, the solution $u(t,\cdot)$ tends, up to a translation, to the solution to (4.1). Because of the iterative nature of the squeezing, this convergence is achieved with exponential speed.

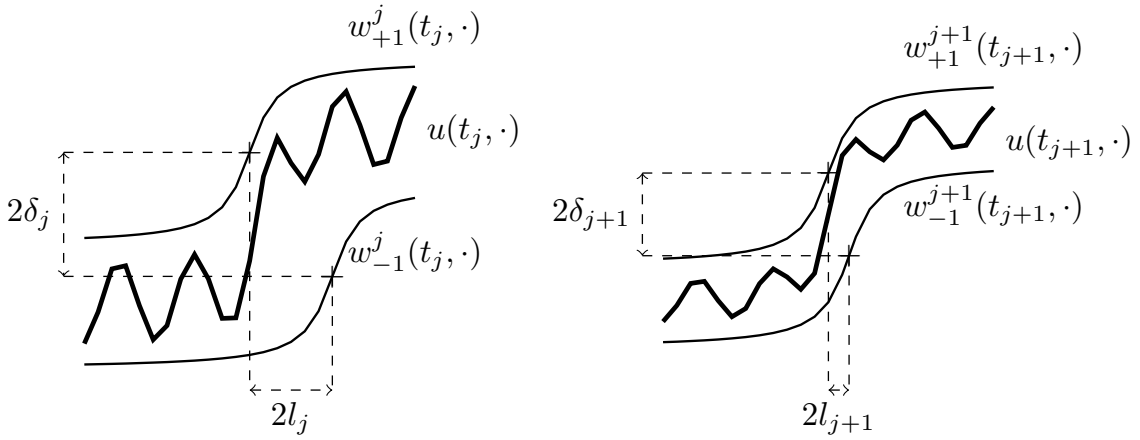



FIGURE 4.4 – Squeezing of $u(t,x)$ solution to (4.5).

**Lemma 4.5.3** (Lemma 2.2 of [45]). *Under the hypotheses of Theorem 4.1.1, let $\Delta_1 < \Delta_0$. Then, there exist positive constants $\sigma$ and $\beta$ such that, for all $\delta \in (0,\Delta_1)$ and $l \in \mathbb{R}$, the functions $w_{-1}(t,x)$ and $w_{+1}(t,x)$ defined by*

$$w_i(t,x) = \eta\left(\zeta_i(t,x)\right) + i\delta \mathrm{e}^{-\beta t} \qquad \text{for} \quad i \in \{-1,+1\}, \tag{4.63}$$

*where*

$$\zeta_i(t,x) = x - ct + il + i\sigma\delta\left[1 - \mathrm{e}^{-\beta t}\right] \qquad \text{for} \quad i \in \{-1,+1\}, \tag{4.64}$$

*are respectively a sub-solution and a super-solution to* (4.15).

*Proof of Lemma 4.5.3.* As $\mathcal{A}$ is invariant by translation, the variable $l$ in the definition of $\zeta$ plays no role. Hence, we take $l = 0$ in the proof below. We also impose for the moment $\sigma \leq 1$.

A straightforward computation yields

$$(\partial_t + |\partial_x|)\, w_i(t,x) = \left(i\sigma\beta\delta \mathrm{e}^{-\beta t/2} - c\right) \eta'\left(\zeta_i(t,x)\right) - i\delta\beta \mathrm{e}^{-\beta t} + |\partial_x|\, \eta\left(\zeta_i(t,x)\right),$$

and, as $\eta$ satisfies (4.1),

$$(\partial_t + |\partial_x|)\, w_i(t,x) = i\sigma\beta\delta \mathrm{e}^{-\beta t}\eta'\left(\zeta_i(t,x)\right) - i\beta\delta \mathrm{e}^{-\beta t} - F'\left(\eta\left(\zeta_i(t,x)\right)\right).$$

Thus

$$(\partial_t - \mathcal{A})\, w_i(t,x) = i\beta\delta \mathrm{e}^{-\beta t}\left[\sigma\eta'\left(\zeta_i(t,x)\right) - 1\right] + F'\left(w_i(t,x)\right) - F'\left(\eta\left(\zeta_i\left(t,x\right)\right)\right).$$

By Taylor-Lagrange expansion, there exists a convex combination $\theta^i(t,x)$ of $\eta\left(\zeta_i(t,x)\right)$ and $w_i(t,x)$ such that

$$(\partial_t - \mathcal{A})\, w_i(t,x) = i\delta \mathrm{e}^{-\beta t}\left[\beta\sigma\eta'\left(\zeta_i(t,x)\right) - \beta + F''\left(\theta^i(t,x)\right)\right]. \tag{4.65}$$

Recall that $\Delta_0 > \Delta_1$. Then, by (4.2), there exists $R_0 > 0$ such that

$$\begin{cases} |\eta(y) - \eta_{\mathrm{l}}| < (\Delta_0 - \Delta_1)/2 & \text{if} \quad y < -R_0, \\ |\eta(y) - \eta_{\mathrm{r}}| < (\Delta_0 - \Delta_1)/2 & \text{if} \quad y > R_0. \end{cases}$$

Therefore, as $\sigma\delta < \Delta_1$, we also have

$$\begin{cases} |w_i(t,x) - \eta_{\mathrm{l}}| \leq (\Delta_0 + \Delta_1)/2 < \Delta_0 & \text{if} \quad \zeta_i(t,x) < -R_0, \\ |w_i(t,x) - \eta_{\mathrm{r}}| \leq (\Delta_0 + \Delta_1)/2 < \Delta_0 & \text{if} \quad \zeta_i(t,x) > R_0. \end{cases}$$

Let

$$\beta := \inf\left\{F''(v), |v - \eta_{\mathrm{r}}| \leq \Delta_0 \quad \text{or} \quad |v - \eta_{\mathrm{l}}| \leq \Delta_0\right\} \tag{4.66}$$

Thus, if $|\zeta_i(t,x)| > R_0$, by definition of $\beta$ and of $\theta^i(t,x)$, $F''\left(\theta^i(t,x)\right) - \beta \geq 0$. Moreover, $\eta' > 0$. Therefore

$$\beta\sigma\eta'\left(\zeta_i(t,x)\right) + F''\left(\theta^i(t,x)\right) - \beta \geq 0. \tag{4.67}$$

Now, we set

$$\sigma := \min\left\{\beta^{-1}\left(\inf_{|y|<R_0} \eta'(y)\right)^{-1}\left(\left\|F''\right\|_{\mathrm{L}^\infty\left(\mathbb{R}^d\right)} + \beta\right), 1\right\}. \tag{4.68}$$

Therefore, if $|\zeta_i(t,x)| \leq R_0$, we also have (4.67). As a conclusion, in any case, (4.65) and (4.67) yield

$$i\,(\partial_t - \mathcal{A})\, w_i(t,x) \geq 0,$$

which implies that $w_{+1}$ and $w_{-1}$ are respectively a super-solution and a sub-solution to (4.15). □

Using Lemma 4.5.3, it is eventually possible to squeeze a solution $u(t,x)$ of (4.15) between a sub-solution and a super-solution. The following Lemma explains how this squeezing is tightened :

**Lemma 4.5.4** (Lemma 3.3 of [45]). *Under the hypotheses of Theorem 4.1.1, let $\Delta_1 < \Delta_0$. Assume that there exist $\xi \in \mathbb{R}$, $\delta \in (0, \Delta_1)$ and $l \in [0, L]$ for fixed $L$ such that, for all $x \in \mathbb{R}$,*

$$\eta(x-l) - \delta \leq u(0,x) \leq \eta(x+l) + \delta. \tag{4.69}$$

*Then, taking $\beta$ and $\sigma$ as in Lemma 4.5.3, there exist a positive constant $\varepsilon_*$ depending only on $\eta$, $F'$, and $L$, and parameters $\widetilde{\xi}$, $\widetilde{\delta}$, $\widetilde{l}$ satisfying*

$$\left|\widetilde{\xi}\right| \leq l, \quad \widetilde{\delta} = \delta + \varepsilon_* \min(1,l) \mathrm{e}^{\beta}, \quad \text{and} \quad 0 \leq \widetilde{l} \leq l + \sigma\delta - \frac{\sigma\varepsilon_* \min(1,l)}{2},$$

*such that, for all $t \geq 1$ and $x \in \mathbb{R}$,*

$$\eta(x - \widetilde{\xi} - \widetilde{l} - ct) - \widetilde{\delta}\mathrm{e}^{-\beta t} \leq u(t,x) \leq \eta(x - \widetilde{\xi} + \widetilde{l} - ct) + \widetilde{\delta}\mathrm{e}^{-\beta t}. \tag{4.70}$$

*Proof.* Thanks to Lemma 4.5.3, the functions $w_{+1}$ and $w_{-1}$ defined by (4.63), for $\zeta_i$ defined by (4.64), are respectively a super-solution and a sub-solution to (4.15). Using (4.69), it follows from Proposition 4.5.1 that, for all $t \geq 0$ and $x \in \mathbb{R}$,

$$w_{-1}(t,x) \leq u(t,x) \leq w_{+1}(t,x). \tag{4.71}$$

Let $\hat{l} := \min(1,l)$ and $\varepsilon_1 := \inf_{x\in[-1,2]} \eta'(x)$. Since $\eta$ is increasing, a Taylor expansion yields

$$\int_0^1 (\eta(x+l) - \eta(x-l))\,\mathrm{d}x \geq \int_0^1 \left(\eta\left(x+\hat{l}\right) - \eta\left(x-\hat{l}\right)\right)\mathrm{d}x \geq 2\varepsilon_1\hat{l}.$$

Therefore, at least one of the following estimates is true

$$\int_0^1 (u(0,x) - \eta(x-l))\,\mathrm{d}x \geq \varepsilon_1\hat{l} \quad \text{or} \quad \int_0^1 (\eta(x+l) - u(0,x))\,\mathrm{d}x \geq \varepsilon_1\hat{l}.$$

Hereafter, we only consider the first case, as the second one is similar. First, using (4.20), there exists $R_1$ such that

$$2\sigma\eta'(x) \leq 1 \quad \text{if} \quad |x| > R_1 \tag{4.72}$$

Let $R_2 := R_1 + L + |c| + 1 + \sigma\Delta_0$. On the one hand, invoking Proposition 4.5.1, we compare $u$ and $w_{-1}$ on $[-R_2, R_2]$

$$\begin{aligned}\inf_{x\in[-R_2,R_2]} \left\{u(1,x) - \eta\left(\zeta_{-1}(1,x)\right) + \delta\mathrm{e}^{-\beta}\right\} &\geq \rho(R_2)\int_0^1 [u(0,y) - \eta(y-l) + \delta]\,\mathrm{d}y \\ &\geq \rho(R_2)\varepsilon_1\hat{l}. \end{aligned} \tag{4.73}$$

We define

$$\varepsilon_* := \min\left(\Delta_1\left(1-\mathrm{e}^{-\beta}\right), \frac{1}{2\sigma}, \frac{\rho(R_2)\varepsilon_1}{2\sigma}\left\|\eta'\right\|^{-1}_{\mathrm{L}^\infty\left(\mathbb{R}^d\right)}\right). \tag{4.74}$$

As a consequence, if $|x| < R_2$, (4.73) yields

$$u(1,x) - \eta\left(\zeta_{-1}(1,x) + 2\varepsilon_*\sigma\hat{l}\right) + \delta\mathrm{e}^{-\beta} \geq \rho(R_2)\varepsilon_1\hat{l} - 2\varepsilon_*\sigma\hat{l}\left\|\eta'\right\|_{\mathrm{L}^\infty\left(\mathbb{R}^d\right)} \geq 0. \tag{4.75}$$

On the other hand, if $|x| > R_2$, then $|\zeta_{-1}(1,x)| \geq R_1 + 1$ whence, by definition of $\varepsilon^*$, $\left|\zeta_{-1}(1,x) + 2\varepsilon_*\sigma\hat{l}\right| \geq R_1$. Inequality (4.71) and Definition (4.72) then imply that

$$u(1,x) - \eta\left(\zeta_{-1}(1,x) + 2\varepsilon_*\sigma\hat{l}\right) + \delta\mathrm{e}^{-\beta} \geq u(1,x) - w_{-1}(1,x) - \varepsilon_*\hat{l} \quad \geq -\varepsilon_*\hat{l}. \tag{4.76}$$

Therefore, from (4.75) and from (4.76), it appears that, for all $x \in \mathbb{R}$, there holds

$$u(1,x) \geq \eta\left(\zeta_{-1}(1,x) + 2\varepsilon_*\sigma\hat{l}\right) - \left(\varepsilon_*\hat{l} + \delta\mathrm{e}^{-\beta}\right).$$

We set

$$\widetilde{\delta} := \left(\delta\mathrm{e}^{-\beta} + \varepsilon_*\hat{l}\right)\mathrm{e}^{\beta},$$

which, thanks to (4.74), satisfies $\widetilde{\delta}\mathrm{e}^{-\beta} \leq \Delta_1$. Applying once more Lemma 4.5.3 yields, for all $t \geq 1$,

$$u(t,x) \geq \eta\left(\zeta_{-1}(1,x) - c(t-1) + 2\varepsilon_*\sigma\hat{l} - \sigma\widetilde{\delta}\mathrm{e}^{-\beta}\left[1-\mathrm{e}^{-\beta(t-1)}\right]\right) - \widetilde{\delta}\mathrm{e}^{-\beta t}. \tag{4.77}$$

By definition of $\widetilde{\delta}$ and $\zeta_{-1}$, the argument of $\eta$ in the above estimate is

$$\begin{aligned} &x - ct - l - \sigma\delta\left[1-\mathrm{e}^{-\beta}\right] + 2\varepsilon_*\sigma\hat{l} - \sigma\widetilde{\delta}\mathrm{e}^{-\beta}\left[1-\mathrm{e}^{-\beta(t-1)}\right] \\ &\geq x - ct - \left[l + \sigma\delta - \sigma\varepsilon_*\hat{l}\right]. \end{aligned} \tag{4.78}$$

Defining now

$$\widetilde{\xi} := -\frac{\sigma\varepsilon_*\hat{l}}{2}, \quad \text{and} \quad \widetilde{l} := l + \sigma\delta - \frac{\sigma\varepsilon_*\hat{l}}{2},$$

and bearing in mind that $\eta$ is increasing, we deduce from (4.77) and (4.78) that

$$u(t,x) \geq \eta\left(x - ct - \widetilde{\xi} - \widetilde{l}\right) - \widetilde{\delta}\mathrm{e}^{-\beta t}. \tag{4.79}$$

Moreover, recalling (4.71), we have

$$u(t,x) \leq \eta\left(x - ct + l + \sigma\delta\right) + \delta\mathrm{e}^{-\beta t} \leq \eta\left(x - ct + \widetilde{l} - \widetilde{\xi}\right) + \widetilde{\delta}\mathrm{e}^{-\beta t}. \tag{4.80}$$

As a consequence, we obtain the desired result (4.70) from (4.79) and (4.80). □

We are now in a position to finish the proof of Theorem 4.1.1. The proof is done while iterating Lemma 4.5.4, which gradually tightens the squeezing around $u(t,x)$.

*Proof of (iii) of Theorem 4.1.1 (restriction of the proof of Theorem 3.1 of [45]).* We proceed in four steps, lowering iteratively in time the values $\delta$ and $l$ such that, for all $x \in \mathbb{R}$, there holds

$$\eta\left(x - ct - \xi - l\right) - \delta \leq u(t,x) \leq \eta\left(x - ct - \xi + l\right) + \delta. \tag{4.81}$$

**Step 1** By assumption (4.18) and since $\eta$ is increasing from $\eta_{\rm l}$ to $\eta_{\rm r}$, there exist $\Delta_1 < \Delta_0$, and $L > 1$ sufficiently large such that (4.81) holds with

$$t = t_1 := 0, \quad \delta = \delta_1 := \Delta_1, \quad \xi = \xi_1 := 0, \quad \text{and} \quad l = l_1 := L - \sigma\Delta_0.$$

**Step 2** We define

$$\delta_* := \min\left(\Delta_1, \varepsilon_*/4\right) \quad \text{and} \quad \kappa_* := \sigma\varepsilon_*/2 - \sigma\delta_* \geq \sigma\varepsilon_*/4 > 0. \tag{4.82}$$

Also, we set $t_* \geq 2$ such that

$$\mathrm{e}^{-\beta t_*}\left(1 + \varepsilon_*/\delta_*\right)\mathrm{e}^{\beta} \leq 1 - \kappa_*. \tag{4.83}$$

Using Lemma 4.5.3, we deduce from the previous step that there exists $t_2$ sufficiently high such that (4.81) holds with $t = t_2$, $\delta = \delta_2 = \delta_*$ and for a certain $\xi \in \mathbb{R}$ and $l = l_2 \leq L$ (as $\varepsilon_*$ implicitly depends on $L$, we further ensure that $l_j \leq L$, for all $j$).

If $l_2 \leq 1$, one directly goes to Step 3. Otherwise, as long as $l_j > 1$, one applies Lemma 4.5.4 at time $t_j = t_2 + (j-2)t_*$ (recall that (4.15) is an autonomous evolution equation), $\delta_j = \delta_*$ and get, by (4.82) and (4.83), that (4.81) holds for $t = t_{j+1}$ with $l \leq l_j - \kappa_*$ and $\delta \leq (1-\kappa_*)\delta_*$. Therefore, one can take $\delta_{j+1} := \delta_*$, $l_{j+1} := l_j - \kappa_*$ and iterate until $l_j < 1$.

**Step 3** By Step 2, we have an index $j_0$ such that (4.81) holds for $t = t_{j_0}$, $\delta = \delta_{j_0} = \delta_*$, $\xi_{j_0} \in \mathbb{R}$ and $l = l_{j_0} = 1$. Using Lemma 4.5.4 and Definitions (4.82) and (4.83), a straightforward computation inductively shows that, for all $j \geq 0$, Inequalities (4.81) hold for $t = t_{j+j_0}$, $\delta = \delta_{j+j_0}$, and $l = l_{j+j_0}$ being defined by

$$t_{j+j_0} := t_{j_0} + jt_*, \quad \delta_{j_0+j} := (1-\kappa_*)^j\delta_* \quad \text{and} \quad l_{j+j_0} := (1-\kappa_*)^j, \tag{4.84}$$

and for $\xi = \xi_j$ such that $|\xi_{j_0+j+1} - \xi_{j_0+j}| \leq (1-\kappa_*)^j$.

**Step 4** We have shown that (4.81) holds for $(t,\delta,l) = (t_j, \delta_j, l_j)$, for all $j \geq 0$. For $t > 0$, we associate $j$ implicitly defined by $t \in [t_j, t_{j+1})$. Thus, we deduce from Lemma 4.5.3 that (4.81) also holds for $t$, $\delta = \delta_j$, $\xi = \xi_j$, and $l = l_j + \sigma\delta_j$. Taking Step 3 into account yields, for all $j > j_0$,

$$\delta \leq \delta_*(1-\kappa_*)^{j-j_0} \quad \text{and} \quad l \leq \left(1 + \sigma\delta_*\right)(1-\kappa_*)^{j-j_0}.$$

Moreover, $\xi_j$ converges to $\xi_\infty$ and, for all $j > j_0$,

$$|\xi_j - \xi_\infty| \leq \kappa_*^{-1}(1-\kappa_*)^{j-j_0}.$$

Yet, a simple calculation shows

$$(1-\kappa_*)^{j-j_0} = (1-\kappa_*)^{-j_0-t_{j_0}/t_*} \exp\left(t \ln(1-\kappa_*)/t_*\right).$$

Setting $\kappa := -\ln(1-\kappa_*)/t_* > 0$ and

$$K := \left[\delta_* + \left(1 + \sigma\delta_* + \kappa_*^{-1}\right) \left\|\eta'\right\|_{\mathrm{L}^\infty\left(\mathbb{R}^d\right)}\right] (1-\kappa_*)^{-j_0-t_{j_0}/t_*},$$

we obtain, for all $t \geq t_{j_0}$, $t \in [t_j, t_{j+1})$,

$$\sup_{x\in\mathbb{R}} |\eta(x-ct-\xi_\infty) - u(t,x)| \leq \delta_j + \left(|l_j| + |\xi_j - \xi_\infty|\right) \left\|\eta'\right\|_{\mathrm{L}^\infty\left(\mathbb{R}^d\right)} \leq K\mathrm{e}^{-\kappa t}.$$

This concludes the proof of Theorem 4.1.1. □

**Acknowledgements** We would like to thank Claude Le Bris for his advice, and Yves-Patrick Pellegrini for his kindness and for providing a physical insight into the Weertman equation.

# Chapitre 5

# Résolution numérique de l'équation de Weertman

Ce chapitre reprend la publication en anglais [88]. Il propose un algorithme pour approximer numériquement les solutions de l'équation de Weertman.

Ce travail a été fait en collaboration avec Claude Le Bris, Frédéric Legoll et Yves-Patrick Pellegrini.

# Fourier-based numerical approximation of the Weertman equation for moving dislocations

Marc Josien [1], Yves-Patrick Pellegrini [2], Frédéric Legoll [1], Claude Le Bris [1]

**Abstract** This work addresses the numerical approximation of solutions to a dimensionless form of the Weertman equation, which models a steadily-moving dislocation and is an important extension (with advection term) of the celebrated Peierls-Nabarro equation for a static dislocation. It belongs to the class of nonlinear reaction-advection-diffusion integro-differential equations with Cauchy-type kernel, thus involving an integration over an unbounded domain. In the Weertman problem, the unknowns are the shape of the core of the dislocation *and* the dislocation velocity. The proposed numerical method rests on a time-dependent formulation that admits the Weertman equation as its long-time limit. Key features are : (i) time iterations are carried out by means of a new, robust, and inexpensive *Preconditioned Collocation Scheme* in the Fourier domain, which allows for *explicit* time evolution but amounts to implicit time integration, thus allowing for large time steps ; (ii) as the integration over the unbounded domain induces a solution with slowly-decaying tails of important influence on the overall dislocation shape, the action of the operators at play is evaluated with *exact* asymptotic estimates of the tails, combined with Discrete Fourier-Transform operations on a finite computational box of size $L$ ; (iii) a specific device is developed to compute the moving solution in a co-moving frame, to minimize the effects of the finite-box approximation. Applications illustrate the efficiency of the approach for different types of nonlinearities, with systematic assessment of numerical errors. Converged numerical results are found insensitive to the time step, and scaling laws for the combined dependence of the numerical error with respect to $L$ and to the spatial step size are obtained. The method proves fast and accurate, and could be applied to a wide variety of equations with moving fronts as solutions ; notably, Weertman-type equations with the Cauchy-type kernel replaced by a fractional Laplacian.

**Keywords** Weertman equation, Peierls-Nabarro equation, dislocations, Cauchy-type nonlinear integrodifferential equation, reaction-advection-diffusion equation, fractional Laplacian, preconditioned scheme, discrete Fourier transform

## 5.1 Introduction

This article addresses the numerical approximation of the following nonlinear integro-differential equation, with Cauchy-type singular kernel :

$$\begin{cases} -\left|\partial_x\right| \eta(x) + c_\eta \, \partial_x \eta(x) = F'_\sigma(\eta(x)) \quad \text{for} \quad x \in \mathbb{R}, \\ \eta(-\infty) = \eta_{\mathrm{l}} \quad \text{and} \quad \eta(+\infty) = \eta_{\mathrm{r}}, \end{cases} \tag{5.1}$$

1. École des Ponts and INRIA, 6 et 8 avenue Blaise Pascal, 77455 Marne-La-Vallée Cedex 2, France.
2. CEA, DAM, DIF, F-91297 Arpajon, France.

where both the real-valued function $\eta$ *and* the scalar constant $c_\eta$ are the unknowns. The potential $F_\sigma$ is a nonlinear bistable function of $\eta$ with (at least) two local minima at values $\eta = \eta_\mathrm{l}$ and $\eta = \eta_\mathrm{r}$. The meaning of the subscript $\sigma$ is explained below. The operator $|\partial_x|$ is linear, and defined in terms of the Hilbert transform $\mathcal{H}$ [95] as

$$|\partial_x|\eta(x) = \mathcal{H}(\partial_x\eta)(x) := \frac{1}{\pi}\mathrm{p.v.}\int_{-\infty}^{+\infty}\frac{\partial_x\eta(x')}{x-x'}\mathrm{d}x' = \lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{|x-x'|>\varepsilon}\frac{\partial_x\eta(x')}{x-x'}\mathrm{d}x', \tag{5.2}$$

where p.v. denotes the principal value [90] at $x$. In the context of singular integral equations, the above kernel of the Hilbert transform is known as the Cauchy kernel. The operator $-|\partial_x|$, also denoted $-(-\Delta)^{1/2}$ by some authors, is diffusive [36, p. 181]. Another useful representation of (5.2) is (by integration by parts)

$$|\partial_x|\eta(x) = -\frac{1}{\pi}\int_0^{+\infty}\frac{\eta(x+y)+\eta(x-y)-2\eta(x)}{y^2}\mathrm{d}y. \tag{5.3}$$

Let the Fourier transform in the continuum (FT) of a function $f$ be defined at wavemode $k$ as

$$\mathcal{F}[f](k) = \widehat{f}(k) := \int_{-\infty}^{+\infty}\mathrm{e}^{-\mathrm{i}kx}f(x)\,\mathrm{d}x. \tag{5.4}$$

One has $\mathcal{F}[\mathrm{p.v.}\,x^{-1}](k) = -\mathrm{i}\,\pi\,\mathrm{sgn}(k)$ [70, p. 1118], whence $\mathcal{F}[|\partial_x|\,\eta](k) = |k|\widehat{\eta}(k)$. Thus, the non-local operator $|\partial_x|$ is symmetric and positive.

Equation (5.1) is a dimensionless form of the Weertman equation [135, 154, 155] (simply referred to as 'the Weertman equation' in the following), which models straight dislocations traveling with steady velocity, thus generalizing the Peierls-Nabarro (PN) equation for static dislocations [120] :

$$\begin{cases} -|\partial_x|\,\eta(x) = F'_\sigma(\eta(x)) \quad \text{for} \quad x\in\mathbb{R}, \\ \eta(-\infty) = \eta_\mathrm{l} \quad \text{and} \quad \eta(+\infty) = \eta_\mathrm{r}. \end{cases} \tag{5.5}$$

The obtention of Equation (5.1) from the original Weertman equation is discussed in Appendix 5.7. Dislocations are linear defects in crystals, the motion of which is responsible for the plasticity of metals [81]. Dislocation lines have a non-vanishing sectional area, i.e., they possess a 'core' of finite width. The derivative $\partial_x\eta(x)$ (the so-called dislocation density) of the unknown function $\eta$ in (5.1) describes the shape function of a *flat* finite-width dislocation on its glide plane, along the $x$-direction. The core is the region of space where $\partial_x\eta(x)$ develops peaks. From a physical standpoint, the function $\eta$ represents a local relative material displacement discontinuity between the upper and lower half-spaces surrounding the glide plane on which moves the dislocation line ; see, e.g., [81] for details. From a broader perspective, the function $\eta$ can be understood as a moving phase-transformation front between the states $\eta_\mathrm{r}$ and $\eta_\mathrm{l}$ (Fig. 5.1).

In (5.1) the term $|\partial_x|\eta$ accounts for the long-range elastic self-interactions that tend to spread the core. This repulsive interaction is counterbalanced by the nonlinear pull-back force $F'_\sigma(\eta)$, which binds together the upper and lower half-spaces, thus giving the dislocation

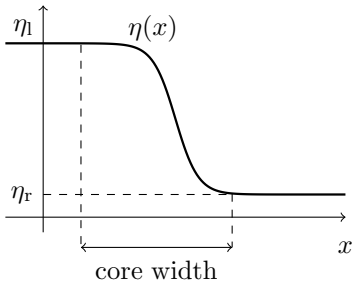



FIGURE 5.1 – Typical shape of $\eta(x)$ in Equation (5.1) when $F$ is a sinusoidal function.

core its finite width. Throughout this article, we consider that $F'_\sigma(\eta)$ includes a constant externally applied loading $\sigma$ (that is, $F_\sigma(\eta) = F(\eta) - \sigma\eta$ where $F(\eta)$ is an energy potential intrinsic to the material, tilted by the adjunction of a linear term $-\sigma\eta$). Moreover, the moving dislocation is subjected to various drag mechanisms encoded into the term $c_\eta\,\partial_x\eta$. As recalled below, the Weertman equation admits an analytical solution [135] if $F(\eta)$ is assumed sinusoidal. For more realistic potentials, a numerical approach is required.

Since its inception, the original PN model has been enriched in various directions. For instance, it most often requires being generalized to vector-valued $\eta$ to be quantitatively predictive [51, 52, 108, 109, 152]. Also, the model has been extended to two dimensions of space, to study planar dislocation loops [51, 52, 156, 161]. Quite generally, methods to compute the shape of static or moving cores encompass variational approaches and involve the finite-element and/or phase-field-type implementations [51, 52, 69, 84, 115, 160]. Yet, in spite of such a wealth of enrichments of the PN model and its associated numerical methods of solution, the one-dimensional Weertman equation —a comparatively simpler extension— has not been investigated as thoroughly, while the specific problem of determining the allowed velocities of steadily-moving dislocations for general force laws $F'_\sigma$ remains an open question of major practical interest [135]. For this reason, the present work focuses on solving the simplest, scalar, and one-dimensional case. A generalization to the vector case of the Weertman equation is the subject of ongoing work, and will be presented elsewhere [132].

It must be emphasized that, in the dimensionless form (5.1) of the equation, $c_\eta$ is *not* the physical velocity of the dislocation. The latter is deduced from $c_\eta$ in a post-processing step, which in the present scalar case is independent from the numerical task of solving the equation. Therefore the physical velocity is not further considered hereafter. *Moreover, we stress that* (5.1) *applies only to* non-supersonic *motion, for which the coefficient of* $|\partial_x|$ *does not vanish in the original Weertman equation* (Appendix 5.7).

Numerical methods for solving integrodifferential equations such as (5.1) have been proposed by many authors. The method employed in [101] uses properties of the Hilbert transform to recast Equation (5.5) into a form amenable to fixed-point approaches. In [96] the authors consider a simpler version of (5.1) on a bounded domain, in which the nonlinear term $F'_\sigma(\eta(x))$ is replaced by some given $\eta$-independent function $g(x)$, and where $c_\eta$ is also given. The solution is then obtained by means of a collocation method with quadratic

interpolation. Those works make use of the expression of the operator $|\partial_x|$ in the direct space. More recently, Karlin *et al.* presented [91] a general iterative method for solving (5.5), based on the expression of $|\partial_x|\,\eta$ in the Fourier space. Our work borrows from the latter reference. The interested reader can also refer to [30, 37] and references therein for Fourier-based numerical schemes applied to the fractional Laplacian operator $|\partial_x|^\alpha$ with $\alpha > 0$ (see also [113] for Hermite spectral methods, and [83] for finite-difference methods, applied to this operator). Fourier-based methods become increasingly popular nowadays to address dislocation problems in engineering [54] but most often address static problems.

The present article proposes a numerical method to approximate solutions to (5.1) in the case where $F_\sigma$ is bistable. As in [91], we build a dynamical system that admits (5.1) as its long-time limit, namely,

$$\begin{cases} \partial_t u(t,x) - c(t)\,\partial_x u(t,x) + |\partial_x| u(t,x) = -F'_\sigma(u(t,x)), \\ u(t=0,x) = u_0(x), \end{cases} \tag{5.6}$$

where $x \in \mathbb{R}$, and $u_0$ is a regular initial data taking values in the interval $[\eta_{\rm r}, \eta_{\rm l}]$ such that

$$u_0(-\infty) = \eta_{\rm l} \quad \text{and} \quad u_0(+\infty) = \eta_{\rm r}, \tag{5.7}$$

where $\eta_{\rm l,r}$ are zeros of $F'_\sigma(\eta)$.

The iterative numerical approach introduced below uses (5.6) to approximate the solution of (5.1). It is immediate that if $(\eta, c_\eta)$ solves (5.1) and if we impose $c(t) := c_\eta$, then $u(t,x) := \eta(x)$ satisfies (5.6) for the initial data $u_0 = \eta$. It is proved in [87] that if $F_\sigma$ is bistable, then *for any* initial data $u_0$ with values in $[\eta_{\rm r}, \eta_{\rm l}]$ that satisfies (5.7) and for *any* continuous function $c(t)$, the solution $u$ of (5.6) converges to the solution of (5.1) in a sense explained in Sec. 5.8.1 below. This leads to the following procedure :

1. consider an initial condition $u_0(x)$ such that (5.7) holds ;
2. approximate the solution $u(t,x)$ to (5.6) ;
3. while evaluating $u(t,x)$, choose $c(t)$ such that the core of $u(t,x)$ remains within the computational box and that $c(t)$ converges to $c_\eta$ (see Equation (5.17) below) ;
4. for $t = t_{\rm f}$ sufficiently large so that $u\,(t,x)$ has numerically converged, return $u\,(t_{\rm f}, x)$ as a numerical approximation to $\eta$, and deduce the velocity $c_\eta$ again using (5.17).

As will be shown, this strategy proves efficient in cases of physical interest. However there might exist alternative strategies for solving (5.1). In particular, recent attempts aim at solving nonlinear partial-differential equations with fractional Laplacian on an infinite domain, by projecting the equation onto a basis of Hermite polynomials and using a Galerkin scheme [113]. Instead, the present work uses asymptotic information on the boundary conditions to handle the issue of the infinite domain of integration.

Numerical approximation of (5.1) paves the way to investigating numerically the Dynamic PN equation [130], which generalizes the Weertman equation to transient regimes. Indeed, the initial conditions and long-time steady-state regimes of the Dynamic PN equation are solutions to the latter equation [131]. However, we emphasize that (5.6) is only an

algorithmic tool that has no relationship whatsoever with the actual dynamics of dislocations.

The article is organized as follows. In Sec. 5.2, we briefly study the dimensionless Weertman equation, and discuss both the uniqueness of its solution and its interpretation as the long-time limit of the dynamical system (5.6). We formally derive asymptotes of solutions to (5.1) and state identities about the velocity $c_\eta$ in general cases. Also, we explain how to choose $c(t)$ in (5.6) to solve this equation in a *comoving frame*, namely, one which follows the translational motion of the core. Technical elements of mathematical proofs are gathered in Appendix 5.8. An analytical solution to (5.1) that exists in a particular case is recalled. In Sec. 5.3, we introduce our numerical representation for $\eta$ and discuss corresponding implementations of the diffusive operator $-|\partial_x|$ and the advective operator $\partial_x$, as well as methods to evaluate $c(t)$. Also, we make use of the asymptotic behavior when $|x| \to +\infty$ of the solution to (5.1) to circumvent the issue of the unbounded domain of integration in (5.2). Once these fundamental elements have been introduced, we build in Sec. 5.4 a Preconditioned Collocation Scheme (PCS) that applies to our problem, and justify this denomination. In Sec. 5.5, we use this numerical approach on two test cases : one with a simple potential $F_\sigma$, for which the exact solution is known, and one with a more physically relevant potential $F_\sigma$. We also illustrate the robustness of our approach, concluding that the algorithm presented is unconditionally stable with respect to the time step $\Delta t$. We empirically derive error scalings with respect to the parameters involved in the discretization. A concluding discussion closes the article, underlining some limitations of our approach, and proposing a few possible extensions. Appendix 5.9 is devoted to examining further one such extension.

## 5.2 Some properties of the Weertman equation

This section is devoted to an overview of some important properties of the Weertman Equation (5.1) and of the companion dynamical Equation (5.6).

### 5.2.1 Invariances

As the PN equation, Equation (5.1) is obviously invariant by translation. When invoking uniqueness of the solutions, we shall henceforth implicitly refer to 'uniqueness up to arbitrary translations'. This invariance has consequences on the numerical solution, which usually undergoes an undesirable drift during calculations if no corrective action is undertaken. A special procedure is developed in Section 5.2.4 below to eliminate this difficulty.

Moreover, Equation (5.1) is invariant by reflection in the sense that if $(\eta(x), c_\eta)$ is a solution for boundary conditions $\eta(\pm\infty) = \eta_{\mathrm{l,r}}$, then $(\eta(-x), -c_\eta)$ is another solution for boundary conditions $\eta(\pm\infty) = \eta_{\mathrm{r,l}}$. Therefore, without loss of generality, we always assume throughout this article that $\eta_{\mathrm{l}} > \eta_{\mathrm{r}}$.

### 5.2.2 Existence and uniqueness of solutions to the Weertman equation.

There exists a unique solution to (5.1) when $F_\sigma$ has a bistable nonlinearity. More precisely it can be shown rigorously (the proof relies on a recent result [73]) that for $F_\sigma$ sufficiently

regular, if $\eta_\mathrm{l} > \eta_\mathrm{r}$ are such that : (i) $F'_\sigma(\eta_\mathrm{l,r}) = 0$ and $F''_\sigma(\eta_\mathrm{l,r}) > 0$ ; (ii) any local minimizer $u$ of $F_\sigma$ between $\eta_\mathrm{r}$ and $\eta_\mathrm{l}$ satisfies $F_\sigma(u) > F_\sigma(\eta_\mathrm{r})$ and $F_\sigma(u) > F_\sigma(\eta_\mathrm{l})$, then there exists a unique velocity $c_\eta$ and a decreasing function $\eta$ satisfying (5.1), which is unique up to translation. Condition (i) means that $F_\sigma$ is a 'bistable potential'. A typical example of such $F_\sigma$, to be used in Sec. 5.5.5, is drawn in Fig 5.2. Possible non-decreasing solutions to (5.1)

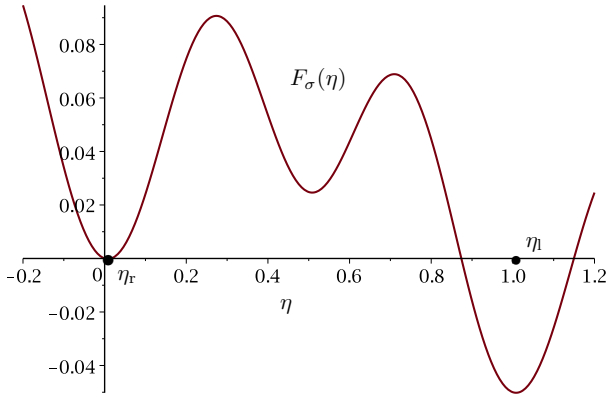



FIGURE 5.2 – Camel-hump potential $F_\sigma$ defined by Equation (5.54), with parameters $\sigma = 0.05$ and $r = 5$.

that might exist as in the PN equation [120, Equation (16)], e.g., for $\eta_\mathrm{l} = \eta_\mathrm{r}$, are disregarded in the present work. Hereafter, we always assume that $\eta_\mathrm{l,r}$ obey the above conditions, and we term such values of $\eta(x)$ at infinity *consistent boundary conditions* (CBCs).

### 5.2.3 Asymptotic behavior and characteristic lengths

By letting $|x| \to \infty$, we formally deduce the leading-order asymptotic expansions

$$\eta(x) \underset{x\to\pm\infty}{\sim} \eta_\mathrm{r,l} + \frac{\eta_\mathrm{l} - \eta_\mathrm{r}}{\pi F''_\sigma(\eta_\mathrm{r,l})} x^{-1}, \tag{5.8}$$

where subscript 'r' (right) goes along with the limit $x \to +\infty$, and 'l' (left) goes along with $-\infty$. They are proved rigorously in [87]. The key ingredient of the proof is the following asymptotic behavior of integrals with Cauchy kernel (see [119, p. 267]) :

$$|\partial_x|\eta(x) \underset{x\to+\infty}{\sim} \frac{1}{\pi x} \int_{-\infty}^{+\infty} \partial_x \eta(x')\mathrm{d}x' = \frac{1}{\pi x}(\eta_\mathrm{r} - \eta_\mathrm{l}). \tag{5.9}$$

It can be formally retrieved from the leading term of a 'series expansion' of (5.2) at $x$ large (note however that the integral involved in the next-to-leading term of such a formal expansion may not exist). Substituting expression (5.9) into (5.1), using the Taylor expansions at boundary values

$$F'_\sigma(\eta) = F'_\sigma(\eta_\mathrm{l,r}) + F''_\sigma(\eta_\mathrm{l,r})(\eta - \eta_\mathrm{l,r}) + \ldots \simeq F''_\sigma(\eta_\mathrm{l,r})(\eta - \eta_\mathrm{l,r}), \tag{5.10}$$

and noting that $\partial_x \eta(x)$ vanishes like $O\left(|x|^{-2}\right)$ [73], we obtain (5.8).

Introducing the characteristic lengths

$$a_{\mathrm{l,r}} = \frac{2\pi}{F''_\sigma(\eta_{\mathrm{l,r}})}, \tag{5.11}$$

Equation (5.8) can be rewritten as

$$\eta(x) \underset{x\to\pm\infty}{\sim} \eta_{\mathrm{r,l}} + (\eta_{\mathrm{l}} - \eta_{\mathrm{r}}) \frac{a_{\mathrm{r,l}}}{2\pi^2} x^{-1}. \tag{5.12}$$

It will be argued in Sec. 5.2.5 that $a_{\mathrm{l,r}}$ represent typical scales of variation of the dislocation density on both sides of the solution. For further purposes, it is useful to introduce a mean characteristic length as

$$\overline{a} := \frac{1}{2}(a_{\mathrm{l}} + a_{\mathrm{r}}) = \pi \left[ \frac{1}{F''_\sigma(\eta_l)} + \frac{1}{F''_\sigma(\eta_r)} \right]. \tag{5.13}$$

Finally, depending on the potential $F_\sigma(\eta)$ at hand, the solution can be either *symmetric*, in the sense that $\partial_x \eta(x)$ is an *even* function (an example will be given in Sec. 5.2.5), or *non-symmetric* if the latter property does not hold.

### 5.2.4 Convergence, velocity determination, centering and choice of $c(t)$

Under our working hypotheses, all the solutions to (5.6) converge towards the unique solution $\eta$ of (5.1) *at exponential rate*. Elements of the proof are briefly sketched in Appendix 5.8.1. A simple way to determine the velocity $c_\eta$ associated with this solution is to multiply (5.1) by $\partial_x \eta(x)$ and to integrate over $x$ [73]. The contribution of the diffusion operator vanishes by symmetry, and one obtains

$$c_\eta = \frac{F_\sigma(\eta_{\mathrm{r}}) - F_\sigma(\eta_{\mathrm{l}})}{\int_{-\infty}^{+\infty} [\partial_x \eta(x)]^2 \mathrm{d}x}. \tag{5.14}$$

The rest of the section essentially relies on formal arguments. To prescribe $c(t)$ in Equation (5.6), we need first to center at $x = 0$ the core of the solution $\eta(x)$ of (5.1) by imposing the supplementary condition

$$\frac{1}{2L} \int_{-L}^{L} \eta(x) \mathrm{d}x = \overline{\eta}, \tag{5.15}$$

where we have introduced the quantity

$$\overline{\eta} := (\eta_{\mathrm{l}} + \eta_{\mathrm{r}})/2, \tag{5.16}$$

and where the constant $L > 0$ represents the half-size of the computational box in the numerical calculations. Imposing condition (5.15) makes the solution of (5.1) unique (and not only unique up to translations), which is crucial for numerical purposes.

Next, on the basis of expression (5.14) for $c_\eta$ , we prescribe the function $c(t)$ as

$$c(t) := \frac{F_\sigma(\eta_\mathrm{r}) - F_\sigma(\eta_\mathrm{l})}{\int_{-\infty}^{+\infty}[\partial_x u(t,x)]^2\,\mathrm{d}x} + \frac{\kappa}{(\eta_\mathrm{l}-\eta_\mathrm{r})} I(t), \quad \text{where} \quad I(t) := \int_{-L}^{L} [u(t,x) - \overline{\eta}]\,\mathrm{d}x, \quad (5.17)$$

and $\kappa > 0$ is a fixed parameter (the reciprocal of some characteristic time), that will ultimately be taken inversely proportional to the algorithmic time step (see Appendix 5.8.2 and Section 5.3.6). Since by (5.63) $u(t,x)$ in (5.6) converges to $\eta(x)$ up to a translation, we formally have by comparing the following expression to (5.14) :

$$\frac{F_\sigma(\eta_\mathrm{r}) - F_\sigma(\eta_\mathrm{l})}{\int_{-\infty}^{+\infty}[\partial_x u(t,x)]^2\,\mathrm{d}x} \underset{t\to+\infty}{\longrightarrow} c_\eta. \quad (5.18)$$

Substituting this limit into definition (5.17) one deduces that at large times

$$c(t) \simeq c_\eta + \frac{\kappa}{(\eta_\mathrm{l}-\eta_\mathrm{r})} I(t). \quad (5.19)$$

We justify in Appendix 5.8.2 that $I(t) \to 0$ when $t \to \infty$, which is equivalent to

$$\frac{1}{2L}\int_{-L}^{L} u(t,x)\,\mathrm{d}x \to \overline{\eta}. \quad (5.20)$$

Thus, in the same limit, $c(t) \to c_\eta$ by (5.19). The limit (5.20) indicates that the choice (5.17) forces the dynamical solution $u(t,x)$ to obey asymptotically the same centering as $\eta(x)$. Put differently, this amounts to computing $u(t,x)$ in a comoving frame. The rightmost term in $(5.17)_1$ can thus be called a *centering correction* to the velocity.

### 5.2.5 An analytical solution

For $|\sigma| < 1$ and

$$F'_\sigma(\eta) = \sin(2\pi\eta) - \sigma, \quad (5.21)$$

with CBCs $\eta_\mathrm{r} = \arcsin(\sigma)/(2\pi)$, and $\eta_\mathrm{l} = 1 + \eta_\mathrm{r}$, the dimensionless Equation (5.1) admits the following analytical solution, easily deduced from the well-known solution [135] to the original Weertman equation :

$$\eta(x) = \eta_\mathrm{r} + \frac{\eta_\mathrm{l} - \eta_\mathrm{r}}{\pi}\left[\frac{\pi}{2} - \arctan\left(\frac{2\pi x}{a}\right)\right], \quad (5.22a)$$

$$\text{with} \quad c_\eta(\sigma) = \tan(2\pi\eta_\mathrm{r}) = \sigma/\sqrt{1-\sigma^2} \quad \text{and} \quad a = 1/\cos(2\pi\eta_\mathrm{r}) = 1/\sqrt{1-\sigma^2}. \quad (5.22b)$$

This solution is symmetric in the sense of Section 5.2.3. We will use this test case in Sec. 5.5 as a benchmark. So, when $|\sigma| \to 1$ the velocity $c_\eta$ and the core width $a$ blow up as $(1-|\sigma|)^{-1/2}$. Since $c_\eta$ and $a$ are not the physical velocity and core width, this behavior is not the hallmark of a physical pathology of the model. It however implies that the computational box should be taken wider and wider to contain the core of $\eta$, and that the latter moves with nearly infinite velocity, which has numerical consequences to be examined in Sec. 5.5.1.

In the form (5.12), the asymptotic behaviors deduced from a direct expansion of $(5.22a)_1$ read

$$\eta(x) \underset{x\to\pm\infty}{\sim} \eta_{\mathrm{r,l}} + (\eta_{\mathrm{l}} - \eta_{\mathrm{r}})\frac{a}{2\pi^2 x}, \tag{5.23}$$

where the length scales (5.11) are $a_{\mathrm{l}} = a_{\mathrm{r}} = a$. Thus, in this particular example where the solution is symmetric the asymptotic behaviors provide a connection between the core width and the next-to-leading terms in the expansion. This supports the interpretation put forward in Sec. 5.2.3 that in generic asymmetric situations $a_{\mathrm{l,r}}$ represent characteristic scales of variation of the dislocation density.

## 5.3 Building blocks

Our numerical scheme to solve the dynamical system (5.6) crucially rests on evaluating the action of the operator $|\partial_x|$ in the Fourier domain, by means of the Discrete Fourier Transform (DFT). This section explains the underlying spatial discretization procedure, and outlines the key features of the implementation.

### 5.3.1 Temporal and spatial discretization

At discrete times $t_n = n\Delta t$ with step $\Delta t > 0$, we need a suitable representation $U^n(x)$ of $u(t_n, x)$ over the whole $x$-axis. To this aim, we define a computational box $[-L, L]$, discretized into $2m$ elementary intervals of width $h = L/m$. Starting from a decomposition similar to equation (18) in Karlin et al. [91], we write the function $U^n(x)$ as

$$U^n(x) = \eta^{\mathrm{ref}}(x) + \delta U^n(x), \tag{5.24}$$

where $\eta^{\mathrm{ref}}(x)$ is a fixed reference function that complies with the asymptotic behaviors (5.8), and $\delta U^n(x)$ is a time-evolving correction, the support of which is contained *within* the box $[-L, L]$. The function $\eta^{\mathrm{ref}}(x)$ is prescribed once for all, and plays the role of boundary conditions for the operators $|\partial_x|$ and $\partial_x$. Decomposition (5.24) is motivated by the fact that the operator $|\partial_x|$ is non-local, and that the solution $\eta$ of (5.1) does not vanish at infinity. Thus, the tail contributions of $u(t, x)$ at infinity should be taken into account when computing $|\partial_x|\, u$. They are represented in (5.24) by those of $\eta^{\mathrm{ref}}(x)$. The need to properly account for tail contributions will be illustrated by means of numerical examples in Section 5.5.4.

In this article, we take $\eta^{\mathrm{ref}}$ as the linear combination

$$\eta^{\mathrm{ref}}(x) = \sum_{\beta=1}^{4} A_\beta f_\beta(x/a^{\mathrm{ref}}), \tag{5.25a}$$

where the basis functions $f_\beta(x)$ are chosen empirically such that $|\partial_x|\, f_\beta(x)$ can be computed

analytically, as

$$f_1(x) := 1, \qquad \text{with} \quad |\partial_x|\, f_1(x) = 0, \tag{5.25b}$$

$$f_2(x) := -\frac{1}{\pi}\arctan(2\pi x), \qquad \text{with} \quad |\partial_x|\, f_2(x) = -\frac{4\pi x}{4\pi^2 x^2+1}, \tag{5.25c}$$

$$f_3(x) := \frac{x}{x^2+1}, \qquad \text{with} \quad |\partial_x|\, f_3(x) = \frac{2x}{(x^2+1)^2}, \tag{5.25d}$$

$$f_4(x) := \frac{1}{\sqrt{x^2+1}}, \qquad \text{with} \quad |\partial_x|\, f_4(x) = \frac{2}{\pi}\frac{x\ln\left[(x^2+1)^{1/2}-x\right]+\sqrt{x^2+1}}{(x^2+1)^{3/2}}, \tag{5.25e}$$

and where $a^{\rm ref}$ is an extra arbitrary scaling parameter. The four coefficients $A_\beta$ are determined so as to satisfy the four constraints expressed by (5.8). We have designed the $f_\beta$ to make this possible : $f_1$ accounts for the constant offset (see Fig. 5.1), while $f_2$ allows for a transition between left and right states, and $f_{3,4}$ implement corrections to $f_2$ to match arbitrary coefficients of the $1/x$ terms in the asymptotic expansions (5.8) ($f_2$ already has $1/x$ next-to-leading asymptotic behavior, but does not give enough freedom). Comparing the asymptotic expansions of $\eta^{\rm ref}(x)$ for $x \to \pm\infty$ directly deduced from (5.25) to the generic expressions (5.8) leads to

$$A_1 = \overline{\eta}, \quad A_2 = \eta_{\rm l} - \eta_{\rm r}, \quad A_3 = \frac{A_2}{2\pi^2}\left(\frac{\overline{a}}{a^{\rm ref}} - 1\right), \quad A_4 = \frac{A_2}{2\pi^2}\frac{\overline{a}}{a^{\rm ref}}\frac{F''_\sigma(\eta_{\rm l}) - F''_\sigma(\eta_{\rm r})}{F''_\sigma(\eta_{\rm l}) + F''_\sigma(\eta_{\rm r})}, \tag{5.26}$$

where $\overline{a}$ and $\overline{\eta}$ have been defined, respectively, in Equations (5.13) and (5.16). Thus, choosing $a^{\rm ref} = \overline{a}$ makes $A_3$ vanish. In the exact case of Sec. 5.2.5, where furthermore $F''_\sigma(\eta_{\rm l}) = F''_\sigma(\eta_{\rm l})$, only $A_1$ and $A_2$ are nonzero and the solution is already completely retrieved at the level of $\eta^{\rm ref}(x)$. Therefore we shall need to take $a^{\rm ref}$ different from $\overline{a}$ to be able to use the exact solution of Sec. 5.2.5 as a non-trivial benchmark of the algorithm in Sec. 5.5. We observe that the asymptotic expansion of $\eta^{\rm ref}(x)$ in (5.25) involves only *odd* inverse powers of $|x|$. It must be emphasized that, however convenient, representation (5.25) is largely arbitrary. Indeed, any other smooth function $\eta^{\rm ref}(x)$ obeying the asymptotic conditions (5.8), mostly varying inside the computational box, and such that $|\partial_x|\,\eta^{\rm ref}(x)$ can be accurately computed once for all (either analytically, or even numerically), would equally well fit our purpose.

Finally, our discretization involves the following restricted sets of $2m$ integers ($m \geq 2$) :

$$\mathcal{K}_{(2m)} = \{-m, \ldots, m-1\}. \tag{5.27}$$

Introducing discrete positions $x_j = j\,h$, the function $U^n(x)$ is represented *inside* the box by the $2m$-vector $\mathbf{u}^n$ of components $u^n_j = U^n(x_j)$. It is decomposed according to (5.24) as

$$u^n_j = \eta^{\rm ref}(x_j) + \delta u^n_j, \quad j \in \mathcal{K}_{(2m)}, \tag{5.28}$$

where the $2m$-vector $\delta\mathbf{u}^n$ of components $\delta u^n_j$ is now the unknown of the problem. We shall also need to consider Equation (5.28) on an extended grid of coordinates $x_l$ with indices

$l \in \mathcal{K}_{(4m)}$. Accordingly, we introduce the injection $\mathcal{I}$ that maps $\mathbb{R}^{2m}$ into $\mathbb{R}^{4m}$, and the projector $\mathcal{P}$ that maps $\mathbb{R}^{4m}$ onto $\mathbb{R}^{2m}$. These operators are defined by

$$\mathcal{I}\mathbf{v} = (0, \ldots, 0, v_{-m}, \ldots, v_{m-1}, 0, \ldots, 0)\,, \tag{5.29a}$$

$$\mathcal{P}\mathbf{w} = (w_{-m}, \ldots, w_{m-1})\,, \tag{5.29b}$$

where $m$ zeros have been added on each side of the $2m$-vector $\mathbf{v} = (v_{-m}, \ldots, v_{m-1})$ in (5.29a), and $\mathbf{w} = (w_{-2m}, \ldots, w_{2m-1})$ in (5.29b) is a $4m$-vector. The operator $\mathcal{P}\mathcal{I}$ leaves $\mathbb{R}^{2m}$ invariant. On the extended grid where $U^n(x_l)$ is represented by the constant values $\eta^{\mathrm{ref}}(x_l)$ outside the computational box, equation (5.28) takes the form

$$U^n(x_l) = \eta^{\mathrm{ref}}_l + (\mathcal{I}\delta\mathbf{u}^n)_l, \quad l \in \mathcal{K}_{(4m)}, \tag{5.30}$$

where we have introduced $\boldsymbol{\eta}^{\mathrm{ref}}$, the $4m$-vector of components $\eta^{\mathrm{ref}}_l := \eta^{\mathrm{ref}}(x_l)$ with $l \in \mathcal{K}_{(4m)}$.

### 5.3.2 Discrete Fourier transforms and zero-padding

Our algorithm makes systematic use of Fourier transforms in DFT form, which allows one to benefit from Fast-Fourier-Transform (FFT) numerical packages. We denote by $\widehat{\mathbf{v}} = \mathcal{F}_{(2m)}[\mathbf{v}]$ (or $\mathcal{F}_{(2m)}[v_j]$ to emphasize the vector components) the $2m$-point DFT that operates on a $2m$-vector $\mathbf{v}$. We define it componentwise as

$$\widehat{v}_p := \left(\mathcal{F}_{(2m)}[\mathbf{v}]\right)_p := \sum_{j=-m}^{m-1} v_j \mathrm{e}^{-\mathrm{i}x_j k_p}, \qquad k_p = \frac{2\pi}{2mh}p, \qquad p \in \mathcal{K}_{(2m)}. \tag{5.31}$$

We shall also use (5.31) with $m$ replaced by $2m$ (keeping fixed the step size $h$ that defines the spatial resolution). As is well-known, carrying out by means of DFTs the convolution of a translation-invariant discretized kernel $O$ with a vector $\mathbf{v}$ requires using the injection and projection operators (5.29). This approach, called *zero-padding*, prevents DFT-induced periodization artifacts (sometimes called *aliasing* artifacts [153]) from showing up near both extremities of the interval of interest in the direct space [33]. For details on zero-padding, the reader is referred to the classical reference [134, p. 643]. Thus, in terms of DFTs,

$$(O\mathbf{v})_i = \sum_{j=-m}^{m-1} O_{i-j} v_j \quad = \left(\mathcal{P}\mathcal{F}^{-1}_{(4m)}\left[\widehat{O}_p\,\left(\mathcal{F}_{(4m)}[\mathcal{I}\mathbf{v}]\right)_p\right]\right)_i, \qquad p \in \mathcal{K}_{(4m)},\ i \in \mathcal{K}_{(2m)}, \tag{5.32}$$

where the second equality follows from definitions (5.29) and (5.31). In the direct space the kernel $O$ is naturally discretized into $4m-1$ components $O_k$ with $-(2m-1) \le k \le (2m-1)$, and must be turned into a $4m$-vector $\mathbf{O}$ by introducing an extra component $O_{-2m} := 0$ before computing in (5.32) the DFT $\widehat{\mathbf{O}} = \mathcal{F}_{(4m)}[\mathbf{O}]$ according to definition (5.31). It should be noted that, for ease of exposition, definition (5.31) differs from the one employed by DFT routines in FFT packages (however, the conversion is easy). For this reason, the zero-padding procedure described in textbooks such as [134] uses component indexings different from ours.

### 5.3.3 Discretization of the advection operator

Using Equation (5.30), we discretize the advection operator on the extended grid by means of an upwind scheme [61] suitably adapted to the extended grid. Introduce first the operators $D^{\pm}$ defined for $\mathbf{w} \in \mathbb{R}^{4m}$ by the following expressions with $\mathrm{O}(h^3)$ error (see [61, p. 297] with $q = 1/2$) :

$$\left(D^{+}[\mathbf{w}]\right)_l := \left(-w_{l+2} + 6w_{l+1} - 3w_l - 2w_{l-1}\right)/6h, \qquad l \in \mathcal{K}_{(4m)}, \tag{5.33a}$$

$$\left(D^{-}[\mathbf{w}]\right)_l := \left(2w_{l+1} + 3w_l - 6w_{l-1} + w_{l-2}\right)/6h, \qquad l \in \mathcal{K}_{(4m)}. \tag{5.33b}$$

Then, $c\,\partial_x U^n(x_j)$ is discretized spatially as $c\,\left(D(c)\mathbf{u}^n\right)_j$, for $j \in \mathcal{K}_{(2m)}$, with

$$D(c)\mathbf{u}^n := \mathcal{P}D^c[\boldsymbol{\eta}^{\mathrm{ref}} + \mathcal{I}\delta\mathbf{u}^n], \tag{5.34}$$

where $D^c := D^+$ if $c \geq 0$ and $D^c := D^-$ if $c < 0$. Note that $D(c)$ only depends on $c$ via its sign. The choice of third-order finite differences in (5.33) is motivated in Sec. 5.5.4 by numerical considerations (see Ref. [61] for schemes of orders 1 and 2 and Ref. [80, p. 111] for order 4). In (5.34) operator $D(c)$ operates in $\mathbb{R}^{4m}$ so that $\eta^{\mathrm{ref}}$ plays the role of boundary conditions when approximating $\partial_x U$ at the extremities of the computational box of size $2L$. Complementing Equations (5.33) with the periodic conventions $w_{2m+k} := w_{-2m+k}$ for $k = 0, 1$ and $w_{-2m-k} := w_{2m-k}$ for $k = 1, 2$ at the boundaries of the extended domain of size $4L$ turns $D^{\pm}$ into diagonal operators in the Fourier space, which allows for straightforward DFT inversions, such as in Equation (5.47b) below (note, however, that $D(c)$ is a non-diagonal operator due to the occurrence of $\mathcal{I}$ and $\mathcal{P}$ in its definition (5.34)). Since $D^c$ is local, and the result in (5.34) has been projected back onto $\mathbb{R}^{2m}$, this convenient periodization has no impact on the final result. To carry out DFTs we envisage $D^{\pm}$ as convolution kernels, which we represent as the $4m$-vectors :

$$\mathbf{D}^{+} = (0, \ldots, 0, -1, 6, -3, -2, 0, \ldots, 0)/6h, \tag{5.35a}$$

$$\mathbf{D}^{-} = (0, \ldots, 0, 2, 3, -6, 1, 0, \ldots, 0)/6h, \tag{5.35b}$$

with $2m-2$ zeros on the left and $2m-2$ zeros on the right in (5.35a), and $2m-1$ zeros on the left and $2m-3$ zeros on the right in (5.35b). The vector representation $\mathbf{D}^c$ of $D^c$ follows, and we recast definition (5.34) of operator $D(c)$ in terms of the DFT $\widehat{\mathbf{D}}^c = \mathcal{F}_{(4m)}[\mathbf{D}^c]$ of components $\widehat{D}^c_p$, as

$$D(c)\mathbf{u}^n = \mathcal{P}\mathcal{F}^{-1}_{(4m)}\left[\widehat{D}^c_p\left(\mathcal{F}_{(4m)}[\boldsymbol{\eta}^{\mathrm{ref}} + \mathcal{I}\delta\mathbf{u}^n]\right)_p\right], \quad p \in \mathcal{K}_{(4m)}. \tag{5.36}$$

### 5.3.4 Discretization of the diffusion operator

We turn next to the discretization of the linear integro-differential operator $|\partial_x|$. In view of (5.2), the latter involves a convolution by a pseudofunction [90], which can be done in the Fourier representation. Crucially, the present approach uses the *continuous* Fourier form of the operator because this representation is straightforwardly diagonal, and versatile

in the sense that it could as well be employed to address the fractional Laplacian $|\partial_x|^\alpha$. Its discretization, hereafter denoted by $|D|^\alpha$, is implemented as follows. We define for any $2m$-vector $\mathbf{v}$ the operator $|D|_0^\alpha$ such that

$$(|D|_0^\alpha \mathbf{v})_j := \left(\mathcal{F}_{(2m)}^{-1}\left[|k_p|^\alpha \left(\mathcal{F}_{(2m)}[\mathbf{v}]\right)_p\right]\right)_j, \qquad p, j \in \mathcal{K}_{(2m)}. \tag{5.37}$$

Then, from (5.24) and (5.37) with $\alpha = 1$, we compute the discretized form of $|\partial_x|\, U^n(x_j) = |\partial_x|\, \eta^{\mathrm{ref}}(x_j) + |\partial_x|\, \delta U^n(x_j)$ as

$$(|D|\mathbf{u}^n)_j := |\partial_x|\, \eta^{\mathrm{ref}}(x_j) + (|D|_0 \delta \mathbf{u}^n)_j \qquad j \in \mathcal{K}_{(2m)}, \tag{5.38}$$

in which $|\partial_x|\, \eta^{\mathrm{ref}}$ is evaluated analytically at $x_j$ by means of Equations (5.25).

### 5.3.5 Alternative formulations

Equations (5.36) and (5.37) define our reference formulations, in which the advection and diffusion operators are treated differently. On the one hand, (5.36) can be considered as a zero-padded (ZP) implementation of the advection. Alternatively, one might consider removing $m$ zeros on each side of $\mathbf{D}^+$ and $\mathbf{D}^-$ in (5.35), defining thus new $2m$-vectors $\underline{\mathbf{D}}^\pm$, and associated $2m$-vectors $\underline{\mathbf{D}}^c$ and $\widehat{\underline{\mathbf{D}}}^c = \mathcal{F}_{(2m)}[\underline{\mathbf{D}}^c]$. Equation (5.36) would then transform into the following non zero-padded (NZP) implementation, which is a diagonal operator in the Fourier representation :

$$\underline{D}(c)\mathbf{u}^n = \mathcal{F}_{(2m)}^{-1}\left[\widehat{\underline{D}}_p^c \left(\mathcal{F}_{(2m)}[\mathcal{P}\boldsymbol{\eta}^{\mathrm{ref}} + \delta \mathbf{u}^n]\right)_p\right], \quad p \in \mathcal{K}_{(2m)}. \tag{5.39}$$

On the other hand, the diffusion operator has been implemented in (5.37) and (5.38) in NZP form, using $2m$-point DFTs. The ZP counterpart to (5.37) would read instead, with now $k_p = 2\pi p/(4mh)$,

$$(|\underline{D}|_0^\alpha \mathbf{v})_j := \left(\mathcal{P}\mathcal{F}_{(4m)}^{-1}\left[|k_p|^\alpha \left(\mathcal{F}_{(4m)}[\mathcal{I}\mathbf{v}]\right)_p\right]\right)_j, \qquad p \in \mathcal{K}_{(4m)},\, j \in \mathcal{K}_{(2m)}, \tag{5.40}$$

which is non-diagonal in the Fourier representation, while (5.38) would become

$$(|\underline{D}|\mathbf{u}^n)_j := |\partial_x|\, \eta^{\mathrm{ref}}(x_j) + (|\underline{D}|_0 \delta \mathbf{u}^n)_j, \qquad j \in \mathcal{K}_{(2m)}. \tag{5.41}$$

As recalled in Section 5.3.2, the application of the (long-range) diffusion operator by means of the NZP expression (5.37) could involve undesired periodization effects on side regions of the computational box. However, such artifacts turn out insignificant because by construction, as a direct consequence of decomposition (5.24), $\delta \mathbf{u}^n$ is small in those regions. Use of the above alternative formulations is further investigated in section 5.5.1 below.

### 5.3.6 Velocity computation

As seen above, Equation (5.6) can be solved in the comoving frame by using the velocity $c(t)$ given by (5.17). In this way, the dislocation core lies as remote as possible from the box boundaries to minimize the influence of the approximations made in handling the tails. To proceed, we substitute in (5.36), at each time $t_n$, the quantity $c$ by $c_n = c(t_n)$ (with the convention that $c_{-1} = 0$) computed from a discretized version over $2m-1$ points of expression (5.17), in which $L$ has been replaced by $L-h$ (since $x_{-(m-1)} = -L+h$ and $x_{m-1} = L-h$); namely,

$$c_n := [F_\sigma(\eta_\mathrm{r}) - F_\sigma(\eta_\mathrm{l})] \left[ h \sum_{j=-(m-1)}^{m-1} \omega_j \left| \left(D(c_{n-1})\boldsymbol{u}^n\right)_j \right|^2 + \frac{\eta_{\mathrm{l},1}^2 + \eta_{\mathrm{r},1}^2}{3(L-h)^3} \right]^{-1} + \frac{\kappa}{\eta_\mathrm{l} - \eta_\mathrm{r}} I_n, \tag{5.42a}$$

$$I_n := h \sum_{j=-(m-1)}^{m-1} w_j \left(u_j^n - \overline{\eta}\right). \tag{5.42b}$$

In these expressions $\eta_{\rho,1} = (\eta_\mathrm{l} - \eta_\mathrm{r}) a_\rho/(2\pi^2)$ for $\rho = \mathrm{l}, \mathrm{r}$, and the $\omega_j$ are the weights 1/3, 4/3, 2/3, …, 2/3, 4/3, 1/3 of the Simpson integration rule. Remark that $c_n$ depends only on $c_{n-1}$ via its sign, which is of little consequence except in calculations at $\sigma$ small where $c_n$ is close to 0 and may oscillate during iterations. The term within brackets results from a straightforward discretization of the integral in $(5.17)_1$, in which the tail contributions have been evaluated analytically from the asymptotic expansions of $\eta^{\mathrm{ref}}$ (for simplicity, we have not evaluated the full tail contributions of $\eta^{\mathrm{ref}}$).

Moreover, a suitable value of $\kappa$ stems from the empirical consideration that if we discretize in explicit Euler form Equation (5.65) of Appendix 5.8.2 as $I_{n+1} = (1 - \kappa\Delta t) I_n$, monotone convergence towards 0 of $I_n$ is ensured by taking $\kappa < 1/\Delta t$. Correspondingly, the value of $\kappa$ used henceforth in (5.42a) is $\kappa = 1/(2\Delta t)$.

## 5.4 Algorithm

This section describes the iterative numerical scheme used to compute $\eta$ and $c_\eta$. This algorithm is explicit in time, is consistent with the dynamical system (5.6), combines the above-discretized operators, and remains stable even with a large time step $\Delta t$.

### 5.4.1 Procedure

Computations go as follows. First, for given local minimizers $\eta_\mathrm{l}$ and $\eta_\mathrm{r}$ of $F_\sigma$, the algorithm is typically initialized by choosing arbitrarily the values $u_j^0$ inside the box, preferentially not too far from the expected solution. This can be done in a number of ways; notably, by using as an initial condition the function $\eta^{\mathrm{ref}}$ with $a^{\mathrm{ref}}$ chosen large enough to encompass the expected overall width of the dislocation density (typically, a few times the characteristic scale $\overline{a}$), whence $u_j^0 \equiv 0$. A few low-resolution runs may help adjusting

$a^{\rm ref}$. Obviously, to get a reasonable representation of the solution, the discretization step $h$ must necessarily be less than any characteristic size of the core (i.e., $\ll \min(a_{\rm l}, a_{\rm r})$). Also, when carrying out incremental parametric studies, the solution computed from the previous value of the parameter under consideration can be used as an initial condition, to save CPU time. However, for studying stability issues, we shall purposely take initial data far from the expected shape of a dislocation.

Denoting by $\Phi$ the scheme presented below, we iterate

$$\mathbf{u}^{n+1} = \Phi(\mathbf{u}^n) \tag{5.43}$$

until the difference between the results of two successive iterations is small, in the sense that

$$\|\Delta\mathbf{u}^n\| := \max_{j\in\{-m,\cdots,m-1\}} \left|u_j^n - u_j^{n+1}\right| \le \Delta_0\Delta t, \tag{5.44}$$

where $\Delta_0$ is a user-defined stopping criterion. Upon completion at some $n$, the algorithm returns $\boldsymbol{\eta} := \mathbf{u}_n$ and the associated $c_{\boldsymbol{\eta}} = c_n$, evaluated thanks to (5.42a). Unless otherwise stated, $\Delta_0 = 10^{-11}$ in the numerical calculations of Sec. 5.5 below.

### 5.4.2 The Preconditioned Collocation Scheme

The scheme $\Phi$ described hereafter, which we call the *Preconditioned Collocation Scheme* (PCS), is based on the requirement that the long-time limit $\boldsymbol{\eta}$ of $\mathbf{u}^n$ should solve the following static equation :

$$-\left(|D|\boldsymbol{\eta}\right)_j + c\left(D(c)\boldsymbol{\eta}\right)_j = F'_\sigma(\eta_j), \qquad j\in\mathcal{K}_{(2m)}, \tag{5.45}$$

which is the discretized form of (5.1). This is a collocation method. A first naive way to proceed would be to attempt solving (5.6) by writing

$$\mathbf{u}^{n+1} = \mathbf{u}^n + \Delta t\big[-|D|\mathbf{u}^n + c_n D(c_n)\mathbf{u}^n - F'_\sigma(\mathbf{u}^n)\big], \tag{5.46}$$

where $\Delta t$ should be adjusted in order to achieve convergence. At convergence, the solution $\boldsymbol{\eta}$ obeys (5.45) and depends on $\Delta t$ only through the evaluation of the numerical velocity $c_n$ defined by (5.42a). As will be justified in Section 5.5.3 this dependence is of little relevance, which is an advantage.

However, system (5.46) is ill-conditioned, because it involves the stiff operators $|D|$ and $D(c_n)$. As a consequence, (5.46) does not converge if $\Delta t$ is not small enough. This issue is dealt with by preconditioning (5.46) in the following way :

$$\mathbf{u}^{n+1} = \Phi(\mathbf{u}^n) := \mathbf{u}^n + \Delta t\, M^n(\Delta t)\left[-|D|\mathbf{u}^n + c_n D(c_n)\mathbf{u}^n - F'_\sigma(\mathbf{u}^n)\right], \tag{5.47a}$$

where $M^n(\Delta t) := M_1^n(\Delta t)M_2(\Delta t)$ is defined in terms of the operator $M_1^n(\Delta t)$, which depends on $\mathbf{u}^n$ via $c_n$, and acts on a $2m$-vector $\mathbf{v}$ as

$$M_1^n(\Delta t)\mathbf{v} := \mathcal{P}\mathcal{F}_{(4m)}^{-1}\left[\frac{\left(\mathcal{F}_{(4m)}[\mathcal{I}\mathbf{v}]\right)_p}{1-\Delta t\, c_n\widehat{D}_p^{c_n}}\right], \qquad p\in\mathcal{K}_{(4m)}, \tag{5.47b}$$

and of the $n$-independent operator $M_2(\Delta t)$ defined on a $2m$-vector $\mathbf{v}$ as

$$M_2(\Delta t)\mathbf{v} := \mathcal{F}_{(2m)}^{-1}\left[\frac{\left(\mathcal{F}_{(2m)}[\mathbf{v}]\right)_p}{1+\Delta t|k_p|}\right], \qquad p \in \mathcal{K}_{(2m)}. \tag{5.47c}$$

By von Neumann analysis, one checks that the operator $1-\Delta t\, cD(c)$ is invertible, which ensures that $M_1^n$ (and thereby $M$) is well-defined. A preconditionner similar to (5.47c) is used in [37]. With the PCS (5.47) the long-time limit of $\mathbf{u}^n$ also satisfies (5.45). Therefore, the preconditioning achieved by introducing $M^n(\Delta t)$ is just a procedure to enable and speed up convergence.

Following Section 5.3.5, we see that the $M_1^n$ operator that corresponds to NZP advection simply derives from removing $\mathcal{I}$ and $\mathcal{P}$, replacing $\widehat{D}_p^c$ by $\underline{\widehat{D}}_p^c$, and using $\mathcal{F}_{(2m)}$ instead of $\mathcal{F}_{(4m)}$, with $p \in \mathcal{K}_{(2m)}$ in definition (5.47b). Likewise, the $M_2$ operator that corresponds to ZP diffusion stems from replacing $\mathcal{F}_{(2m)}$ by $\mathcal{F}_{(4m)}$, replacing $\mathbf{v}$ by $\mathcal{I}\mathbf{v}$, and applying $\mathcal{P}$ to the result in (5.47c), with $p \in \mathcal{K}_{(4m)}$. One advantage of the multiplicative splitting $M^n = M_1^n M_2$ is that it allows such variants to be examined separately to assess their individual influence.

The working of the preconditioning is understood as follows. If we ignore the differences between ZP and NZP versions of the operators and assume $\Delta t$ small enough, the operator $M^n(\Delta t)$ is close to the effective inverse of $1-\Delta t[-|D|+c_n D(c_n)]$. Hence, (5.47a) preconditioned by $M^n(\Delta t)$ is tantamount to the following semi-implicit scheme (see [79, p. 102] for a reference on semi-implicit schemes) :

$$\mathbf{u}^{n+1} \simeq \mathbf{u}^n + \Delta t\left[-|D|\mathbf{u}^{n+1} + c_n\, D(c_n)\mathbf{u}^{n+1} - F'_\sigma\left(\mathbf{u}^n\right)\right]. \tag{5.48}$$

As is well-known, treating stiff operators in an implicit way yields a stable scheme. Hence, this preconditioning naturally leads to stability (this assertion will be exemplified in Section 5.5). We note that (5.48) is a consistent discretization of (5.6) —just as (5.46).

We now justify the preconditioned character of $\Phi$ by analogy with the task of solving iteratively a linear system. Indeed, ignoring nonlinearities by replacing $F'_\sigma(\mathbf{u}^n)$ by a constant $\mathbf{b}$ and by setting $c_n = 0$, (5.45) can be put in the form $\mathsf{A}\mathbf{u} = \mathbf{b}$, so that (5.46) reduces to a scheme of the type

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \Delta t\left(\mathsf{A}\mathbf{u}^n - \mathbf{b}\right), \tag{5.49}$$

where $\mathsf{A}$ is a *positive* symmetric matrix. The latter scheme converges for general $\mathbf{b}$ if and only if all eigenvalues of $(1-\Delta t\,\mathsf{A})$ belong to the interval $(-1,1)$ ; this can require $\Delta t$ to be very small. In a similar way, the scheme (5.47a) can be abstracted as

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \Delta t\mathsf{M}\left(\mathsf{A}\mathbf{u}^n - \mathbf{b}\right), \tag{5.50}$$

with $\mathsf{M}$ close to $(1+\Delta t\mathsf{A})^{-1}$. Then, the eigenvalues of the latter scheme are close to that of $(1+\Delta t\,\mathsf{A})^{-1}$, which unconditionally belong to $(0,1)$ if $\Delta t > 0$. Equation (5.50) amounts to solving $\mathsf{MA}\mathbf{u} = \mathsf{M}\mathbf{b}$. This is the classical preconditioning method (in the spirit of a modified Richardson iteration, see [92, p. 6]).

## 5.5 Numerical results

The above algorithm has been implemented as a MATLAB® code, and our results have been obtained on a 2.3GHz standard laptop computer, with typical computation times of order one second to a few minutes per run, depending on the case at hand. Except in Sec. 5.5.5, the calculations concern benchmark comparisons with the exact solution of Sec. 5.2.5, and have been carried out with $F'_\sigma$ as defined by (5.21), for $\sigma \in [0, 1)$. Moreover, in Sections 5.5.1 to 5.5.4, a parameter value $a^{\rm ref} = \overline{a}/2$ (see Section 5.3.1) is employed in $\eta^{\rm ref}(x)$.

### 5.5.1 Convergence

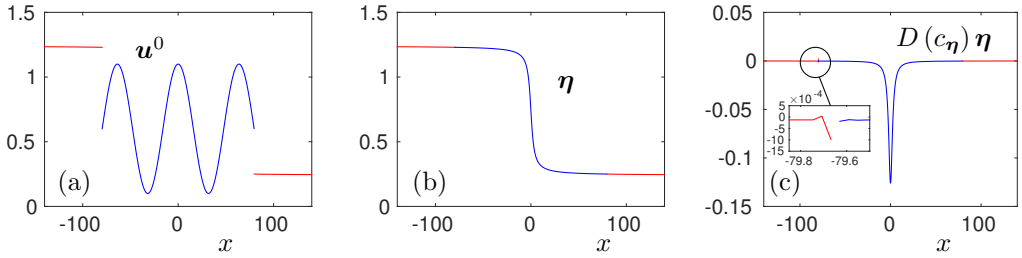



FIGURE 5.3 – (a) Initial data $\mathbf{u}^0$ ; (b) output $\boldsymbol{\eta}$ ; (c) discrete derivative of $\boldsymbol{\eta}$. Blue : solution $\boldsymbol{\eta}$ ; red : parts of $\eta^{\rm ref}$ outside the box. Discretization parameters $2m = 4096$, step $h = L/m \simeq 0.039$, and $\Delta t = 0.1$.

As stated in Sec. 5.2.5, the core width $a(\sigma)$ and the velocity $c_\eta(\sigma)$ blow up when $\sigma \to 1$. The first problem is easily solved by running one preliminary low-resolution run to provide a rough numerical estimate for the core width $\widetilde{a}(\sigma)$. The half computational box size $L$ is then adjusted to a value $L \gg \widetilde{a}(\sigma)$. Figure 5.3 displays the initial data $\mathbf{u}^0$, the converged result $\boldsymbol{\eta}$ and its discrete derivative. The figure illustrates the robustness of the PCS with respect to initial conditions in the sense that the initial data $\mathbf{u}^0$ can be non-monotone, irregular, and far from the solution $\eta$ to (5.1). In this calculation, the applied loading is $\sigma = 1 - 1.973 \times 10^{-3}$, which induces large values close to one another for the converged velocity and core width, $c_\eta(\sigma) \simeq a(\sigma) \simeq 15.9$ ; see (5.22b). The half box size is $L = 5\,a(\sigma) \simeq 80$. As seen in Fig. 5.3(c) the discretized derivative of $\boldsymbol{\eta}$ is regular inside the box $[-L, L]$, but the inset shows that some artifacts take place near the matching points between the solution inside the box and the tails of $\eta^{\rm ref}(x)$ outside it. Quite generally, it is observed that these artifacts diminish when either $L$ or $\sigma$ are increased (results not shown). As they are related to the artificial discontinuity between the numerical solution (constantly moving and brought back to the center of the box by the velocity correction term) and the reference function (fixed), their presence could not be avoided.

Under the same initial conditions, Figure 5.4 illustrates the convergence properties with time, via $\|\Delta\mathbf{u}^n\|$ defined in (5.44). The semi-logarithmic plot of Fig. 5.4(a) shows that, up to high-frequency oscillations, convergence is exponential with time, in agreement with the arguments of Section 5.8.1. Moreover, the convergence rate (the slopes in Fig. 5.4(a))

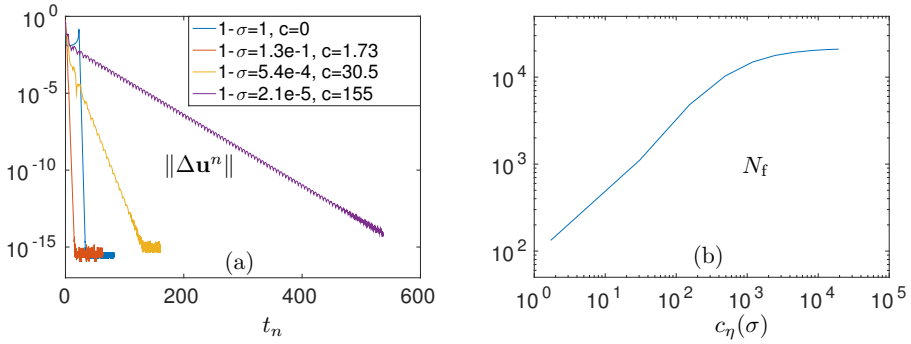



FIGURE 5.4 – (a) Convergence indicator $\|\Delta\mathbf{u}^n\|$ defined by (5.44) vs. time $t_n$, and (b) number of iterations $N_{\rm f}$ until $\|\Delta\mathbf{u}^n\|/\Delta t \leq \Delta_0$ vs. velocity $c_\eta(\sigma)$. Parameters $L = 10\,a(\sigma)$, $2m = 1024$, and $\Delta t = 0.1$.

depends on the applied loading $\sigma$, convergence being impeded when $\sigma$ approaches 1. Not unexpectedly, this loss of performance coincides with $F_\sigma$ being 'less and less bistable' in the sense that the minima of $F_\sigma$ at $\eta = \eta_{\rm r}$ and $\eta = \eta_{\rm l}$ become less and less deep. Bistability is a crucial requirement for the existence of a solution to (5.1), and for proving convergence as expressed by (5.62). In this connection, it should be remarked that the tail expressions (5.8) involve the second derivatives $F''_\sigma(\eta)$, which presumably leads to pathologies when the latter are small. However, Fig. 5.4(b), which displays the data in parametric form of parameter $\sigma$, shows that the PCS copes well with high velocities, which is important from a physical standpoint. Thus, our preconditioning does a nice job of avoiding stability issues when $c_\eta$ is large. In this respect, recourse to ZP advection proves crucial. Indeed, Fig. 5.5 exemplifies what happens upon using the NZP advection (5.39) and associated $M_1^n$ preconditionner built as described in section 5.4.2. For $\sigma$ close to its theoretical limit 1, waves coming from the right boundary of the box cause $u(t,x)$ to oscillate badly. Although convergence might eventually take place at larger times, no sign of entering a regime of convergence have shown up in this particular calculation up to $10^6$ iterations, at which point execution was stopped. Strikingly, the same calculation with ZP advection converged in 1420 iterations. In the rest of the paper, all calculations are carried out with our reference formulation (namely, ZP advection and NZP diffusion). However, extra tests were performed, leading us to conclude that using ZP diffusion with the present algorithm does not improve the accuracy of the results in any way worth reporting in detail, nor provides any advantage over using NZP diffusion, which we attribute to the circumstance already mentioned in section 5.3.5.

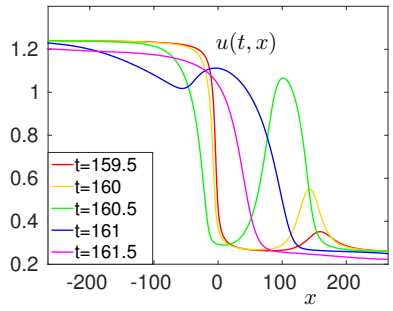



FIGURE 5.5 – Oscillations of $u(t,x)$ when using both NZP diffusion and advection operators for $\sigma = 0.9998$, which induces $c_\eta \simeq 53.0$. Discretization parameters : $2m = 1024$, $2L = 531$, $\Delta t = 0.1$.

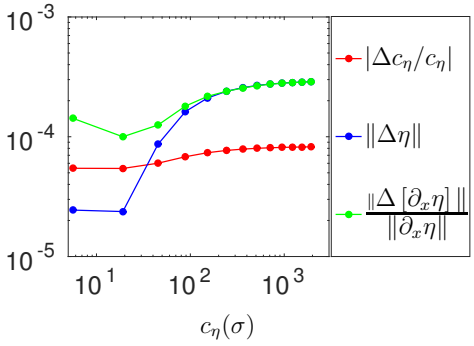



FIGURE 5.6 – Parametric plot with $\sigma$ as the variable parameter, of the error indicators (5.51) vs. $c_\eta(\sigma)$. Parameters $L = 20a(\sigma)$, $2m = 2048$ and $\Delta t = 0.1$.

### 5.5.2 Error indicators and overall accuracy

Figure 5.6 compares the output of the PCS with the exact solution of Sec. 5.2.5 in terms of the following indicators :

$$\|\Delta\eta\| := \inf_{\xi\in\mathbb{R}} \max_{j\in\{-m,\cdots,m-1\}} |\boldsymbol{\eta}_j - \eta(x_j+\xi)|\,, \tag{5.51a}$$

$$\frac{\|\Delta[\partial_x\eta]\|}{\|\partial_x\eta\|} := \inf_{\xi\in\mathbb{R}} \max_{j\in\{-m,\cdots,m-1\}} \frac{|D(c_{\boldsymbol{\eta}})\boldsymbol{\eta}_j - \partial_x\eta(x_j+\xi)|}{\|\partial_x\eta\|_{\mathrm{L}^\infty(\mathbb{R})}}, \tag{5.51b}$$

$$|\Delta c_\eta/c_\eta| := |c_{\boldsymbol{\eta}}/c_\eta - 1|. \tag{5.51c}$$

The absolute error (5.51a) can as well be understood as a relative error since in the cases considered $\|\eta\|_{\mathrm{L}^\infty(\mathbb{R})}$ is of order 1. In (5.51a) and (5.51b) the $\inf_\xi$ operation, motivated by the translation invariance of (5.1), takes care of the approximate character of the centering of the numerical solution, due to discretization errors. Formulated in this way, the error indicators are insensitive to small shifts in the position of the computed solution. The figure indicates that the outputs of the PCS accurately approximate the exact results, with errors

of a similar order of magnitude for the three quantities represented. In addition, the small errors observed on Figure 5.6 depend only weakly on $c_\eta$.

### 5.5.3 Discretization parameters and error scaling

This section closely investigates the scalings of the error at convergence with respect to the time step $\Delta t$, the half box size $L$ and the space discretization step $h$. The same case as above is considered, with a variety of applied loadings. Addressing first the influence of

TABLE 5.1 – Errors as a function of $\Delta t$ ($2L = 638$, $2m = 4096$, $h = 0.156$, $\sigma_1 = 0.9921$).

| $\Delta t$ | 0.25 | 0.1 | 0.01 | 0.001 |
|---|---|---|---|---|
| $\|\Delta\eta\|$ | $5.400822 \times 10^{-5}$ | $5.400780 \times 10^{-5}$ | $5.400754 \times 10^{-5}$ | $5.400752 \times 10^{-5}$ |
| $\frac{\|\Delta[\partial_x\eta]\|}{\|\partial_x\eta\|}$ | $1.223510 \times 10^{-4}$ | $1.223571 \times 10^{-4}$ | $1.223607 \times 10^{-4}$ | $1.223611 \times 10^{-4}$ |
| $\|\Delta c/c\|$ | $2.022173 \times 10^{-5}$ | $2.021742 \times 10^{-5}$ | $2.021482 \times 10^{-5}$ | $2.021456 \times 10^{-5}$ |

$\Delta t$, we observe that the PCS solution depends on $\Delta t$ only through the centering correction term in expression (5.42a) of $c_n$. However, as explained in Section 5.3.6, this term vanishes in the limit of infinite times, and is therefore expected to be small at convergence. This is confirmed by the errors reported in Table 5.1 for applied loading $\sigma = \sigma_1$ (see caption) so that $c_\eta(\sigma_1) \simeq 7.91$, and decreasing values of $\Delta t$. Errors are quasi-constant, which shows that the dependence on $\Delta t$ of the converged result is negligible. For definiteness, the rest of the calculations in the present paragraph is made with $\Delta t = 0.1$.

Figure 5.7 illustrates how $\|\Delta\eta\|$ depends on $L$ and $h = L/m$, for the two very different velocities obtained with loadings $\sigma_2 = 0.951$ so that $c_\eta(\sigma_2) \simeq 3.08$, and $\sigma_3 = 1 - 1.97 \times 10^{-5}$ so that $c_\eta(\sigma_3) \simeq 159$. Raw results are displayed in Figs. 5.7(a) and (b). Computations for $\sigma = \sigma_3$ did not converge with $h/a = 0.31$, which is why this value is not considered in plots (b) and (d). It turns out that for $L$ large, $\|\Delta\eta\|$ scales as $h^3$ for both loadings. Besides, it is approximately proportional to $L^{-3}$ for $L$ small. This is demonstrated in the data-collapse plots (c) and (d) where the individual datasets of Figs. 5.7(a) and (b), respectively, are merged into one single master curve by means of appropriate rescalings of abscissas and ordinates. The data collapse of Fig. 5.7(b) for $\sigma = \sigma_3$ is only partially successful, as noticeable corrections to scaling arise for $Lh \ll a(\sigma_3)^2$.

The scalings can be understood as follows. On the one hand, the scaling in $h$ is consistent with our choice of a third-order advection scheme in (5.33). On the other hand, since the error at any point is spread over the whole domain by the integro-differential operator $|\partial_x|$, the error $\eta(L) - \eta^{\rm ref}(L)$ at the boundary point $x = L$ (for instance) can be used to estimate the overall error. It behaves as $L^{-3}$ in the present case where $\eta(x)$ and $\eta^{\rm ref}(x)$ are symmetric, and where the next nonzero term in expansion (5.12) is $\propto x^{-3}$. This is consistent with the error scaling observed in the plots. Still, these elementary arguments do not explain the downwards bending of the high-velocity plots in Fig. 5.7(b), which indicates either that the

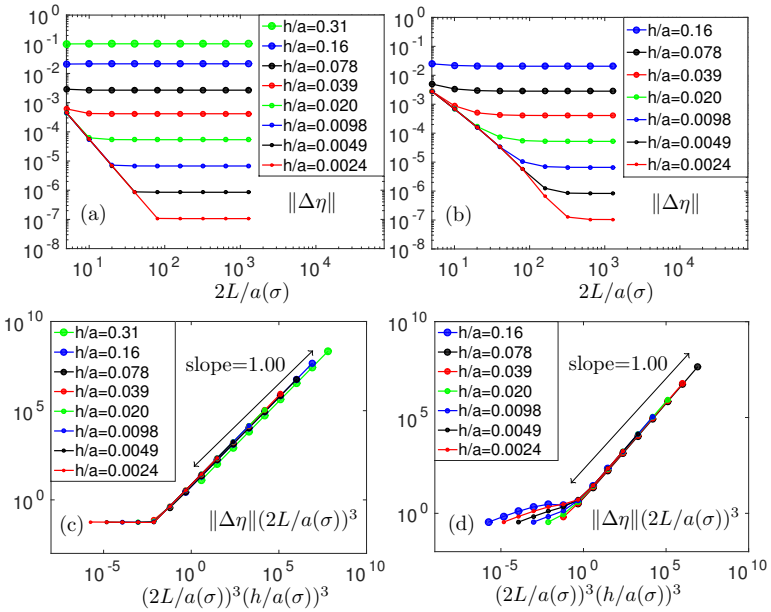



FIGURE 5.7 – Raw data and corresponding data-collapse plots for error $\|\Delta\eta\|$. Loading $\sigma = \sigma_2$ in (a) and (c) (moderate velocity), and $\sigma = \sigma_3$ in (b) and (d) (high velocity) ; see text. In the legends, $a$ stands for $a(\sigma)$.

$L^{-3}$ scaling regime has not been reached, or that it may not hold exactly. This bending causes deviations from ideal scaling, made conspicuous in Fig. 5.7(d).

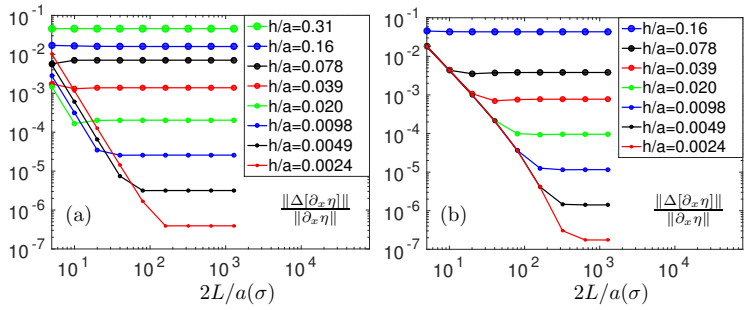



FIGURE 5.8 – Raw data for the error (5.51b) on $\partial_x\eta(x)$. (a) $\sigma = \sigma_2$, (b) $\sigma = \sigma_3$ (see text).

Likewise, Figures 5.8 and 5.9 display the errors (5.51b) and (5.51c), respectively, for the loadings $\sigma_2$ and $\sigma_3$. Whereas Fig. 5.8(a) resembles Fig. 5.7(a), the dependence of the error with respect to $h$ is more involved. In particular for $\sigma = \sigma_2$ (Fig. 5.8(a)), there exists at fixed $L$ an optimal value of $h = L/m$ that minimizes the error. The reason for this behavior is unclear. However it should be realized that approximating $\partial_x\eta$ in the sense of the $L^\infty$ norm is quite demanding. Figs. 5.8(b) and 5.9(b) display bendings similar as in Fig. 5.7(b). No data collapse is presented, as notable deviations from scaling take place in all figures.

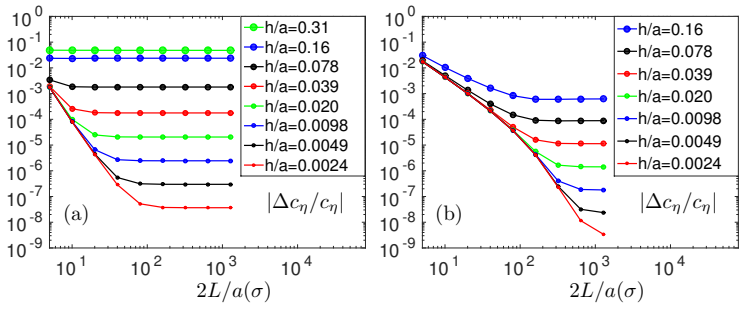



FIGURE 5.9 – Raw data for the error (5.51c) on $c_\eta$. (a) $\sigma = \sigma_2$, (b) $\sigma = \sigma_3$ (see text).

TABLE 5.2 – Error on $\eta$ minimized over $L$ for $\sigma = \sigma_2$, as a function of $m$.

| $2m$ | 32 | 128 | 512 | 2048 | 8192 |
|---|---|---|---|---|---|
| min. value of $\|\Delta\eta\|$ | $2.1 \times 10^{-2}$ | $6.1 \times 10^{-4}$ | $6.2 \times 10^{-5}$ | $7.2 \times 10^{-6}$ | $8.8 \times 10^{-7}$ |
| optimum $hL/a(\sigma_2)^2$ | 0.781 | 0.195 | 0.195 | 0.195 | 0.195 |

For practical matters one needs, e.g., to determine the optimal value of $L$ at fixed number $2m$ of discretization points, which somehow corresponds to a fixed cost to go from $\boldsymbol{u}^n$ to $\boldsymbol{u}^{n+1}$ in (5.47a). Table 5.2 shows the minimal error $\|\Delta\eta\|$ deduced from the datasets of Fig. 5.7, together with the corresponding optimal value of $L$ via the ratio $(hL)/a^2 = (L/a)^2/m$. One observes that the optimal value of the latter quantity does not depend on $m$ for $m$ sufficiently large. This can be understood from the above scaling arguments. Indeed, the data collapse in Fig. 5.7(c) indicates that for $Lh \ll C$, where $C$ is some constant, $\|\Delta\eta\| \propto L^{-3}$, and that for $Lh \gg C$, $\|\Delta\eta\| \propto h^3 = L^3/m^3$. Thus the optimum takes place at the crossover between these regimes in which the error first decreases, then increases with $L$. Balancing the two regimes, it follows that $Lh = L^2/m \simeq C$ at the optimum. The analysis still qualitatively holds for higher $\sigma$ values. The optimum just discussed is not the one evoked above in connection with Fig. 5.8. However, in both cases, one sees that increasing the number of discretization points at fixed $L$ does not necessarily improve the accuracy : one also needs to increase $L$.

The same arguments as above suggest that, in the general case of a nonsymmetric solution where the next nonzero term in expansion (5.12) can be an $\mathrm{O}(x^{-2})$ (see Sec. 5.2.3), the error on $\eta$ would behave as $L^{-2}$ instead of $L^{-3}$, and if the error scales as $h^3$ for $h$ small the optimum would take place for $Lh^{3/2} \simeq C$.

### 5.5.4 Influence of the tails and of the order of the advection scheme

At given $L$ the quality of the numerical solution also depends on how tail contributions are accounted for while handling the operator $|\partial_x|$. As the quantity $|\partial_x|\,\eta(x)$ enters the equation for $\eta$, errors on the former affect the latter via the nonlinear term. To illustrate this point, we focus on the point $x = 0$ where $\eta(0) = \eta^{\mathrm{ref}}(0) = \overline{\eta}$. Since both the function

$\eta^{\mathrm{ref}}$ and the solution $\eta$ to (5.1) satisfy the same asymptotic behavior (5.8), their difference behaves as $\eta(x) - \eta^{\mathrm{ref}}(x) = O\left(|x|^{-\tau}\right)$, where $\tau = 2$ in the general case, and $\tau = 3$ in the present, symmetric, benchmark case. Hence by (5.3), *the long-range contribution to the error* in $|\partial_x|\,\eta(0)$ due to the replacement of $\eta$ by $\eta^{\mathrm{ref}}$ outside the computational box is bounded by

$$\left|\int_L^{+\infty} \frac{\eta(y) + \eta(-y) - [\eta^{\mathrm{ref}}(y) + \eta^{\mathrm{ref}}(-y)]}{y^2}\mathrm{d}y\right| \lesssim \int_L^{+\infty} \frac{\mathrm{d}y}{y^{\tau+2}} \propto L^{-(\tau+1)}. \tag{5.52}$$

Thus, in the results of the previous Section 5.5.3 for which $\tau = 3$ the observed scaling stems from local errors at boundary points, which scale as $L^{-3}$, and not from the present long-range contributions of errors in tails, which scale at most as $L^{-4}$. In the general non-symmetric case where $\tau = 2$, the scalings of these errors are presumably changed into $L^{-2}$ (see end remark in the previous section) and $L^{-3}$, respectively. Therefore, using a function $\eta^{\mathrm{ref}}(x)$ with faithful tails (in the sense of Section 5.3.1) should make local errors dominant over tail errors in all cases. In contrast, upon not using such a reference function, the error on $|\partial_x|\,\eta(0)$ would instead scale as

$$\left|\int_L^{+\infty} \frac{\eta(y) + \eta(-y)}{y^2}\mathrm{d}y\right| \lesssim \int_L^{+\infty} \frac{\mathrm{d}y}{y^2} = L^{-1}. \tag{5.53}$$

We indeed obtained such a scaling while carrying out some preliminary studies (results not shown). To summarize, disregarding tails deteriorates the accuracy of the calculations. On the contrary, using a reference function $\eta^{\mathrm{ref}}$ with faithful tails to handle properly the operator $|\partial_x|$ strongly improves it.

The overall error also depends on the order in $h$ of the discrete advection operator $D_\pm$ of Sec. 5.3.3. Along with the calculations of the previous Section with an advection scheme of order 3, we also carried out similar calculations with schemes of order 2 and 4. In both latter cases, the power of the scaling in $h$ was found identical to the order of the scheme for $h$ small enough (results not shown). Thus, at least for orders 2 to 4, the scaling in $h$ is closely related to the order or the discretization scheme of the advection operator, the discretization errors involved by the DFT in the computation of the operator $|\partial_x|$ presumably being subdominant.

Figures 5.10 further illustrate these points by means of the indicators (5.51a) and (5.51c). The plots have been made with advection upwind schemes of order 1, 2, 3 and 4, and two different types of asymptotes ; namely, 'refined tails' (r.t.) (5.8), and 'constant tails' (c.t.) in which the inverse power-law correction in (5.8) is dropped. Both figures show that using refined tails provides in most cases orders-of-magnitude gains of accuracy over using constant tails. The dependence of the error on the order of the upwind scheme is more difficult to interpret, although orders 3 and 4 lead to better accuracy. As the second-order scheme behaves irregularly in Fig. 5.10 (b), the third-order scheme is used in the rest of the article.

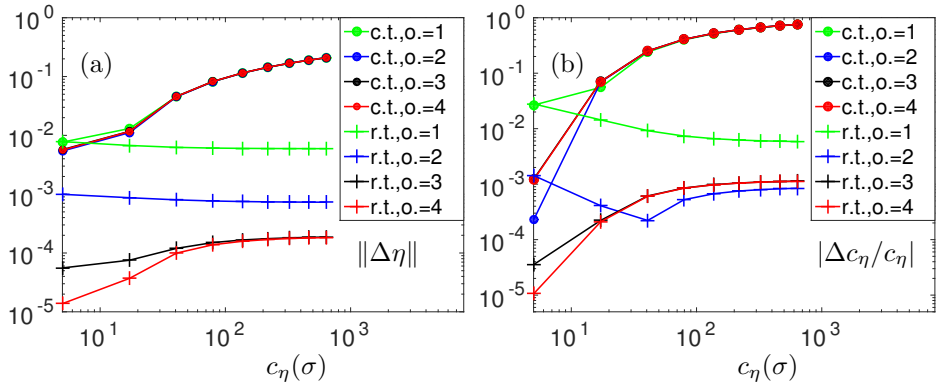



FIGURE 5.10 – Parametric plots with $\sigma$ as the variable parameter. (a) $\|\Delta\eta\|$, (b) $|\Delta c_\eta/c_\eta|$ vs. $c_\eta(\sigma)$. Legends : o. : order of the advection scheme ; r.t. 'refined tails' ; c.t. 'constant tails' (see text). Discretization parameters $L = 10a(\sigma)$, $2m = 1024$. The plots for o. = 2, 3, and 4 cannot be distinguished for constant tails.

### 5.5.5 A generalized example : the camel-hump potential

We now evaluate our method with the 'camel-hump' type potential $F_\sigma$ of Fig. 5.2, defined by

$$F_\sigma(\eta) := \frac{1}{4\pi}\left[1-\theta^2-\left(\theta\sqrt{1-\theta^2}+\arcsin(\theta)-\phi\right)\cot\phi\right]-\sigma\,\eta, \tag{5.54}$$
$$\theta = \sin(\phi)\cos(2\pi\eta), \qquad \phi = \arctan r.$$

We have derived it by means of Lejček's method [102] so as to provide the following *exact* dissociated dislocation solution to the PN equation when $\sigma = 0$ :

$$\eta(x) = \frac{1}{2\pi}\left[\pi - \arctan\left(2\pi x - r\right) - \arctan\left(2\pi x + r\right)\right], \tag{5.55}$$

where the parameter $r \geq 0$ sets the dissociation width between the partial dislocations. Depending on $\sigma$ and $r$, $F_\sigma$ can feature between its two main minima at $\eta_{\rm l}$ and $\eta_{\rm r}$, and its humps, an intermediate local minimum of depth controlled by $r$, leading thus to a more or less dissociated solution $\eta$ to (5.1). For $r = 0$, the derivative of (5.54) reduces to (5.21) and the dislocation is non-dissociated.

In this section, calculations have been carried out with $a^{\rm ref} = 5\,\overline{a}$, because the overall width of the dislocation is much larger than that of the individual peaks in the density. Figure 5.11 shows that the algorithm correctly recovers solution (5.55). Due to the presence of a secondary hollow in the potential, the method takes longer to converge.

We finally present an application to a more physically relevant case, for which the exact solution is unknown. Indeed, when $\sigma > 0$ and $r > 0$ no analytical solution to (5.1) is available. However, as discussed in Sec. 5.2.2, there exists a solution to (5.1) that can be computed with our algorithm for any $\sigma \in (0, \sigma_{\rm lim})$, where $\sigma_{\rm lim} = \max_\eta F'_\sigma(\eta)$. Figure 5.12 displays in (a) the solution $\boldsymbol{\eta}$, and in (b) its derivative (the dislocation density). In this

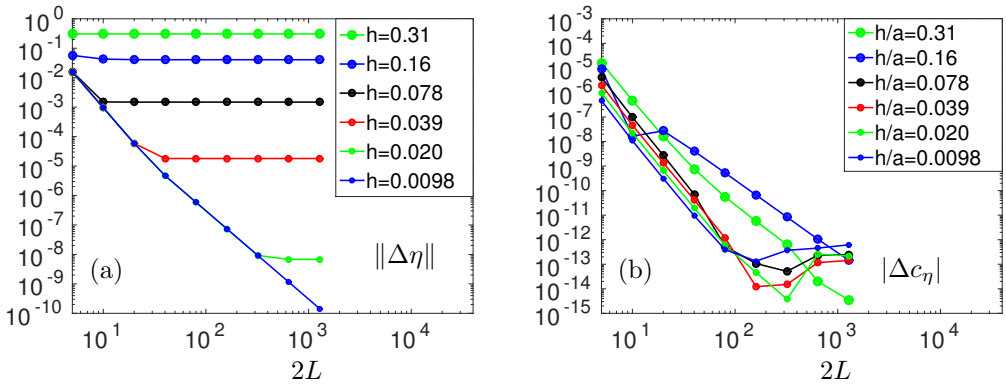



FIGURE 5.11 – (a) $\|\Delta\eta\|$ vs. $2L$ ; (b) $|\Delta c_\eta|$ vs. $2L$, for $F_\sigma$ as in (5.54). Parameters $r = 5$, and $\sigma = 0$.

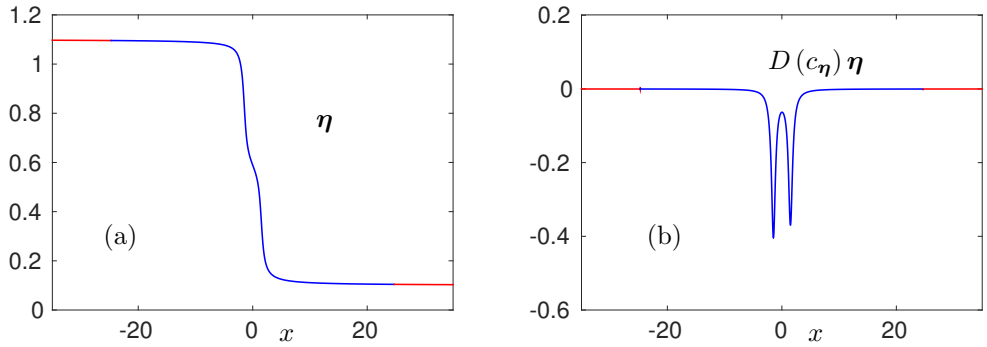



FIGURE 5.12 – (a) Numerical solution $\boldsymbol{\eta}$ and (b) discrete derivative of $\boldsymbol{\eta}$ for $F_\sigma$ defined by (5.54) with parameters $r = 5$ and $\sigma = 0.5274$. Blue : solution $\boldsymbol{\eta}$ ; red : parts of $\eta^{\mathrm{ref}}$ outside the box. Discretization parameters : $2L = 49$ and $2m = 4096$.

example, the latter features two peaks that represent partial dislocations. The midpoint between the two partial dislocations corresponds to the local minimum of $F_\sigma$ in $\eta_{\mathrm{m}} \in (\eta_{\mathrm{r}}, \eta_{\mathrm{l}})$. The asymmetry of the dislocation density can be interpreted as a consequence of the nonzero driving force $\sigma$ coupled to the dissociation process induced by the camel-hump character of the potential. Figure 5.13 displays two quantities of physical interest, namely, the velocity $c_\eta(\sigma)$ and the effective core width $a(\sigma)$, for $\sigma$ varying between 0 and $\sigma_{\mathrm{lim}} = \max(F_0') = 0.5902$. Here, the quantity $a(\sigma)$ has been computed by minimizing the $\mathrm{L}^2$ norm of the difference between the numerical solution $\boldsymbol{\eta}$ and the function, parametrized by $a$ and $x_0$ :

$$f_{a,x_0}(x) := \eta_{\mathrm{r}} + \frac{\eta_{\mathrm{l}} - \eta_{\mathrm{r}}}{\pi}\left[\frac{\pi}{2} - \arctan\left(\frac{2\pi(x - x_0)}{a}\right)\right]. \tag{5.56}$$

As expected from an analogy with the simpler case of Section 5.2.5, both $a(\sigma)$ and $c_\eta(\sigma)$ increase with $\sigma$ and blow up when $\sigma \to \sigma_{\mathrm{lim}}$.

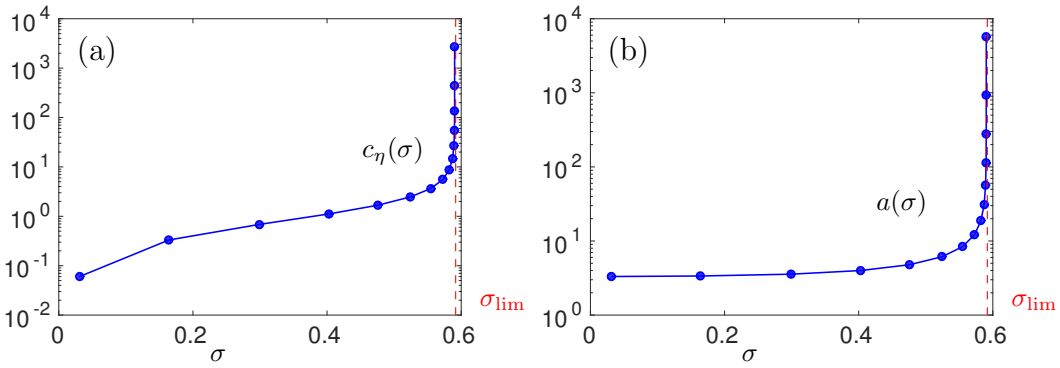



FIGURE 5.13 – (a) Velocity $c_\eta(\sigma)$, (b) core width $a(\sigma)$ for $F_\sigma$ defined by (5.54) with $r = 5$. Discretization parameters : $2m = 2048$ and $L \simeq 20a(\sigma)$.

## 5.6 Concluding discussion

To summarize, we have proposed the Preconditioned Collocation Scheme (PCS), which is a numerical procedure to approximate solutions to (5.1), based on the dynamical system (5.6). The PCS uses the continuous FT of the operator $|\partial_x|$ and takes advantage of the FFT in its implementation, *in spite of the strong constraint that the desired solution is not periodic, but has boundary conditions at infinity.* We have taken advantage from the exact asymptotic expansion of the solution to improve the accuracy of the numerical approximation. Also, we have shown that an overall $\mathrm{O}(h^3)$ error in the space discretization could be achieved by means of a third-order advection scheme. We should add that no appreciable Gibbs oscillations were observed in our numerical solutions, probably owing to the intrinsically continuous character of the expected solution, and in spite of the small artifacts at box boundaries.

The method employed remains stable when the (a priori unknown) advection part scaled by the velocity $c_\eta$ dominates over the diffusion part in (5.1), which allows to investigate the asymptotic behavior of (5.1) when $c_\eta$ is large. The PCS solves a discretized version of (5.1). Being preconditioned, it can be used with a large time step $\Delta t$, nonetheless delivering outputs that depend very weakly on $\Delta t$. Although this was not illustrated, we add that if $\eta_{\rm l}$ and $\eta_{\rm r}$ are not exactly computed as exact local minimizers of $F_\sigma$ (e.g., if they suffer from slight numerical errors), the algorithm converges as well. As Fig. 5.4(b) indicates, our method has some limitations, however : the more advection dominates diffusion, the more iterations the method requires to converge. Zero-padding proves crucial in handling advection at high velocities.

Still, the time and space complexities of the algorithm give satisfactory accuracy at reasonable computational cost. Indeed, the PCS requires $\mathrm{O}(m)$ memory space. Moreover DFTs have been speeded up by means of Fast Fourier Transform routines. Therefore, each iteration step takes $\mathrm{O}(m \log m)$ CPU time. Given that the number of iterations to convergence obviously scales as $1/\Delta t$, the PCS therefore has an overall time complexity of order $\mathrm{O}(m \log m/\Delta t)$. Then, achieving $10^{-4}$ accuracy on $\eta$ and $c_\eta$ with 2048 discretization points

takes about one second on a standard laptop with CPU running at 2.3GHz, except in demanding cases when $c_\eta \to +\infty$, where it is slower. This study involved runs with up to $10^5$ discretization points.

We point out that alternative schemes could be used to simulate the dynamical system (5.6). First, other time integrators can be appealed to. For instance, classical splitting schemes such as Strang or Lie splittings [77] aim at simulating a dynamical system that involves a sum of operators, and turn the latter into a composition of evolution operators. Although these investigations were not reported for conciseness, we have checked that such methods do indeed apply to our problem, and prove stable and robust. However, the solution they produce is inherently $\Delta t$-dependent. Hence, achieving high accuracy requires small $\Delta t$ values, which makes them expensive. Also, as mentioned in Section 5.4.2, an explicit Euler scheme could as well be employed. However, the latter approach proves unstable if $\Delta t/h$ is not small. In contrast, the alternative semi-implicit scheme sketched in (5.48) is stable. However, it is costly as it requires the inversion of the (non-diagonal) operator $1 + \Delta t\,(|D| - c_n D(c_n))$. To summarize, among all the schemes we have investigated, we deem the PCS the most robust and least expensive one. Second, we have deliberately chosen to implement the operator $|\partial_x|$ in continuous Fourier form. It would be equally possible to discretize the integral representation of the operator, e.g., in the convenient form (5.3). However, this may require adapted integration rules [91]. In this respect, a method taking advantage of FTs proves more straightforward. It also proves more versatile because the analytical form of the kernel under consideration might be unavailable in cases involving a different integrodifferential operator. The main constraint for diagonalization by the FT is that the operator be translation-invariant.

As a perspective, although definitive conclusions about the validity of the approach for more general equations are yet to be obtained, we have all reasons to believe that this method applies as well to equations of the type [73]

$$\begin{cases} -|\partial_x|^\alpha \eta(x) + c_\eta \partial_x \eta(x) = F'(\eta(x)) \quad \text{for } x \in \mathbb{R}, \\ \eta(-\infty) = \eta_{\rm l} \quad \text{and} \quad \eta(+\infty) = \eta_{\rm r}, \end{cases} \tag{5.57}$$

where $F$ is a bistable nonlinearity, $\alpha > 0$, and $|\partial_x|^\alpha$ is the operator of Fourier symbol $|k|^\alpha$. Adapting the method to values $\alpha \neq 1$ would require examining the asymptotic behavior of the solution, finding new suitable basis functions $f_\alpha$ to reproduce it, and computing the associated function $|\partial_x|^\alpha \eta^{\rm ref}(x)$. This, we did not do, except for the case $\alpha = 2$ below. However, if one does not insist on accuracy, merely checking algorithmic convergence does not require endowing $\eta^{\rm ref}(x)$ with its correct asymptotic behavior. Indeed, if one reformulates (5.47a) as an evolution equation for $\delta \mathbf{u}^n$, the contributions of $\eta^{\rm ref}$ only redefine the force. As the velocity equation (5.14) holds as well for (5.57), the only modifications therefore consist in replacing $|D|$ by $|D|^\alpha$ in (5.38) and (5.47a), and $|k_p|$ by $|k_p|^\alpha$ in (5.47c). In a series of tests carried out for 40 equispaced values of $\alpha \in (0, 4)$, with sinusoidal force and applied loading $\sigma = 0.99999$, we actually found that the PCS converged well (not shown). We explored in more detail (see Appendix 5.9) the classical advection-reaction-diffusion case $\alpha = 2$, where $|\partial_x|^2 = -\Delta$, for a force function of parameter $r$ (no applied stress)

$$F'(\eta) = -\,(r + 2\eta)\left(1 - \eta^2\right), \qquad (|r| < 2). \tag{5.58}$$

In this case, (5.57) admits the analytical solution $\eta(x) = -\tanh(x)$ and $c_\eta = r$ [71, p. 291]. Employing a variant of the above-described method with an adapted $\eta^{\text{ref}}$ function, we have recovered a numerical approximation of this analytical solution. Our method presumably applies as well to the modified Weertman equation with gradient term [135], in which the operator $|\partial_x|$ in (5.1) is replaced by $|\partial_x| - \lambda\Delta$, where $\lambda > 0$.

Yet, as already mentioned in Section 5.1, other approaches exist for solving partial differential equations with fractional Laplacian of type related to (5.1) that could as well apply to our problem. Most notably, Mao and Shen [113] recently used a decomposition of the unknown solution on a basis of Hermite functions to alleviate the problem of the infinite domain. They applied it to various examples including nonlinear equations of the Schrödinger type. However, they do not specifically address the 1-dimensional Weertman equation. An interesting perspective would thus consist in conducting in-depth numerical comparisons between Mao and Shen's method and ours for 1-dimensional fractional equations of the type (5.57).

**Acknowledgements** M. Josien thanks for its hospitality the CEA-DAM Île-de-France where part of this work was carried out. Thanks are also due to anonymous reviewers for stimulating questions and useful suggestions.

## 5.7 Appendix : The Weertman equation and its dimensionless form

With definition (5.2), the Weertman equation reads in dimensional form [135]

$$-\mu A(v)|\partial_x|\eta(x) + \mu B(v)\partial_x\eta(x) = f'\big(\eta(x)\big) - \sigma, \tag{5.59}$$

where $f$ is bounded, and the physical parameters are as follows : $\mu$ (the shear modulus), the applied stress $\sigma$, and $f'$ have the same dimension (force by unit surface) ; and $A(v)$ and $B(v)$ are dimensionless functions of the physical velocity $v$ originally deduced from elasticity theory. Equation (5.59) requires that $|\sigma| < \max_\eta |f'(\eta)|$ [135]. Well-behaved models need the adjunction of an additional drag mechanism that makes $B(v)$ non-zero (of same sign as $v$) except at $v = 0$ where it vanishes [135]. Introducing the characteristic stress $\sigma_{\text{th}} = \mu b/(2\pi d)$, where $d$ (the interplane distance) and $b$ (the Burgers vector modulus) are both lengths, the dimensionless Equation (5.1) immediately follows from the substitutions $x \to 2\pi d A(v)x$, $\eta \to b\eta$, $\sigma \to \sigma_{\text{th}}\sigma$, and $f' \to \sigma_{\text{th}}F'$ in (5.59), and from letting $c_\eta = B(v)/A(v)$ thereafter. This requires that $v$ lie in the range where $A(v) > 0$ (note that $A(v) = 0$ for supersonic velocities, i.e., velocities larger than the upper wave speed of the medium). Given $\sigma$, once $c_\eta$ has been determined by the method described in the main text, the physical velocity follows from solving for $v$ the equation $c_\eta = B(v)/A(v)$. Several branches may exist, as the model allows one to consider intersonic regimes [131, 135]. Physical results on $v$ lie outside the scope of the article, and will be reported elsewhere [132].

## 5.8 Appendix : Mathematical details

### 5.8.1 Convergence towards solutions to the Weertman equation

The first author (M.J.) proves in [87] that, under our working hypotheses, all the solutions to (5.6) converge towards the unique solution $\eta$ of (5.1) *at exponential rate*. The proof is based on the following arguments. Consider the equation :

$$\partial_t \varphi(t,x) + |\partial_x| \varphi(t,x) = -F'_\sigma(\varphi(t,x)) \quad \text{with} \quad \varphi(t=0,x) = u_0(x), \tag{5.60}$$

where $u_0(x)$ is the initial condition of (5.6). The connection between (5.60) and (5.6) resides in that $\varphi(t,x)$ solves (5.60) if and only if

$$u(t,x) = \varphi\left(t, x + \int_0^t c(s)\mathrm{d}s\right) \tag{5.61}$$

solves (5.6). Now, under mild requirements similar to those of Sec. 5.2.2, Equation (5.60) can be shown [87] to have a unique solution with the following property : if $(\eta, c_\eta)$ is the solution to (5.1) with same boundary conditions at infinity as $u_0$, then there exist constants $\kappa > 0$, $K > 0$ and $\xi \in \mathbb{R}$ such that

$$\sup_{x\in\mathbb{R}} |\varphi(t,x) - \eta(x - c_\eta t + \xi)| \leq K \mathrm{e}^{-\kappa t}. \tag{5.62}$$

The proof relies on a comparison principle, which is a generic property of operators $\partial_t - D + F'_\sigma[\cdot]$ where $D$ is a dissipative operator (see the classical references [45, 60]).

Combining Equations (5.61) and (5.62), one deduces that at large times and uniformly in $x$,

$$u(t,x) \simeq \eta(x + \zeta(t)), \quad \text{where} \quad \zeta(t) = \int_0^t c(s)\mathrm{d}s - c_\eta t + \xi. \tag{5.63}$$

Thus, given CBCs at infinity, the long-time limit of the solution to (5.6) is the solution to (5.1) with the same CBCs, up to a time-dependent drift. In the next section, we show how a suitable choice of $c(t)$ eliminates this undesirable effect.

### 5.8.2 Limit $I(t) \to 0$

To show that $I(t) \to 0$, we differentiate the large-time expression $(5.63)_1$ of $u(t,x)$ with respect to time. Further invoking (5.19), we obtain at large times (the dot denotes a total time derivative)

$$\partial_t u(t,x) \simeq \partial_x \eta(x+\zeta(t))\dot{\zeta}(t) = \partial_x \eta(x+\zeta(t))[c(t) - c_\eta] \simeq \partial_x \eta(x+\zeta(t)) \frac{\kappa}{(\eta_\mathrm{l} - \eta_\mathrm{r})} I(t). \tag{5.64}$$

Integrating (5.64) over $x \in [-L, L]$ then yields the approximate first-order differential equation

$$\dot{I}(t) \simeq -\kappa \frac{\eta(-L+\zeta(t)) - \eta(L+\zeta(t))}{\eta_\mathrm{l} - \eta_\mathrm{r}} I(t) \simeq -\kappa\, I(t), \tag{5.65}$$

where the rightmost expression follows from having neglected the $1/x$ term in the asymptotic expansions (5.12) of $\eta(x)$. From (5.12) this is legitimate if $L \gg \max(a_\mathrm{l}, a_\mathrm{r})/(2\pi^2)+|\zeta(t)|$. The latter condition is compatible with $L$ being fixed in numerical computations if $\zeta(t)$ tends to a finite value at large times. The latter property follows from a simple self-consistent argument. Indeed, Equation (5.65) implies that $I(t) \simeq I(T)\mathrm{e}^{-\kappa(t-T)}$ for $t > T$ where $T$ is some time above which (5.63) holds, so that effectively $I(t) \to 0$. Substituting this expression of $I(t)$ into (5.19), one deduces an approximate analytical expression for $c(t)$ that tends to $c_\eta$. Writing $(5.63)_2$ in the form $\zeta(t) = \zeta(T) + c_\eta(T-t) + \int_T^t c(s)\,\mathrm{d}s$, and substituting the obtained expression of $c(t)$ finally entails the desired saturation property in the form $\zeta(t \to \infty) \simeq \zeta(T) + I(T)/(\eta_\mathrm{l} - \eta_\mathrm{r})$. No quantitative estimate of the latter quantity is available.

As already mentioned, $\kappa$ is taken inversely proportional to the algorithmic time step. Thus, in practice $\kappa \to \infty$ in the limit of continuous times. Because of the $\exp(-\kappa t)$ dependence of $I(t)$, we observe that the centering correction in $(5.17)_1$ remains well-behaved in this limit.

## 5.9 Appendix : Laplacian case

In this section we compare the exact solution $\eta(x) = -\tanh(x)$, $c_\eta = r$ of the Laplacian case (Equation (5.57) with $\alpha = 2$, and potential (5.58)), with its numerical approximation. The function $\eta^{\mathrm{ref}}$ can be easily dealt with : since the next-to-leading terms in the asymptotic behaviors of the exact solution are now exponentially small, we impose the coefficients in the definition of $\eta^{\mathrm{ref}}(x)$ to cancel out the $1/x$ term in the asymptotic expansion of $\eta^{\mathrm{ref}}(x)$, namely, $A_1 = \overline{\eta}$, $A_2 = \eta_\mathrm{l} - \eta_\mathrm{r}$, $A_3 = -A_2/(2\pi^2)$, and $A_4 = 0$. Moreover, $|\partial_x|\,\eta^{\mathrm{ref}}(x)$ is replaced by $-\partial^2_{xx}\eta^{\mathrm{ref}}(x)$ in (5.38). Fig. 5.14 illustrates the good accuracy of our numerical solution.

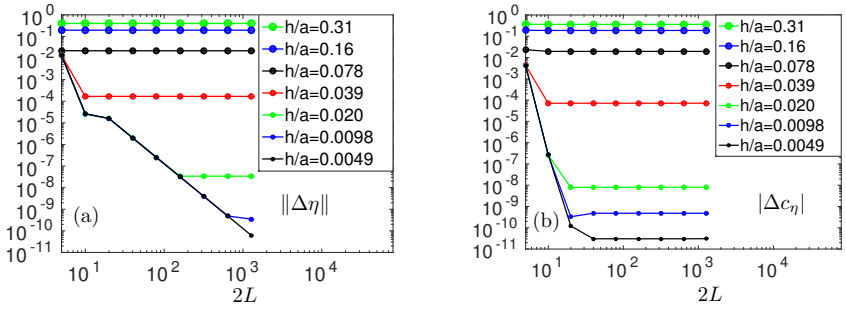



FIGURE 5.14 – (a) Raw data for the error $\|\Delta\eta\|$ and (b) for error (5.51c) on $c_\eta$. Force parameter in (5.58) : $r = 0.6$.

# Chapitre 6

# Construction de l'équation de Peierls-Nabarro Dynamique

Dans ce court chapitre, nous expliquons comment l'équation de Peierls-Nabarro Dynamique a été construite, et nous en définissons précisément chacun des termes.

Ce travail a été fait en collaboration avec Yves-Patrick Pellegrini.

Le lecteur trouvera un récapitulatif des formules utiles en Annexe A.5.

## 6.1 Introduction

Récemment, l'équation de Peierls-Nabarro Dynamique a été proposée dans [130] pour généraliser l'équation de Peierls-Nabarro à des dislocations en régime dynamique. Cette nouvelle équation s'obtient à partir du modèle de Peierls [129] décrit en introduction. C'est un modèle hybride, qui couple mécanique des milieux continus –ici, l'équation d'élasticité linéaire– et description au niveau microscopique de la matière –où les interactions ont lieu au niveau atomique. En tirant parti de la géométrie particulière du modèle, on déduit une équation sur la forme de la dislocation sur le plan de glissement : l'équation de Peierls-Nabarro Dynamique. Cette équation non-linéaire présente la particularité d'être une équation intégrodifférentielle, à la fois en temps et en espace. Le but de ce chapitre est d'en expliciter la construction et d'en définir précisément chacun des termes, afin de pouvoir formaliser un problème mathématique.

La construction de l'équation de Peierls-Nabarro Dynamique est faite dans [130] (hors le terme visco-plastique, tiré de [131]). Nous y renvoyons le lecteur pour plus de précisions. Pour notre part, nous en décrivons les ingrédients mathématiques essentiels en nous ramenant à une équation scalaire. Pour plus de clarté, nous avons séparé la construction du terme intégrodifférentiel de celle des termes non-linéaires de l'équation de Peierls-Nabarro Dynamique. En ce qui concerne la partie intégrodifférentielle de l'équation, les explications qui suivent sont largement inspirées d'une approche détaillée dans [62] (voir aussi [3, 75]). Le contexte physique de [62] est cependant différent du nôtre, puisqu'il s'agit de décrire des phénomènes de rupture.

Il serait plus réaliste d'un point de vue physique d'envisager un modèle couplé tridimensionnel (au sens de la Section 6.6 plus bas). Toutefois, l'équation *scalaire* de Peierls-Nabarro Dynamique permet déjà d'étudier des phénomènes physiques intéressants au moins qualitativement, sinon quantitativement (voir [131]). Par ailleurs, nous ne voyons pas d'obstacle à ce que les considérations –tant théoriques que numériques– des Chapitres 6, 7 et 8 s'adaptent, mutatis mutandis, à l'analogue tridimensionnel de l'équation de Peierls-Nabarro Dynamique [1].

L'objectif sous-jacent à la construction de l'équation de Peierls-Nabarro Dynamique dans [130] est de permettre la simulation de phénomènes dynamiques complexes à l'échelle de la dislocation. Par exemple : la mise en mouvement d'une dislocation en prenant en compte les effets d'inertie, la nucléation (la création de nouvelles dislocations), le comportement de deux dislocations qui se croisent, les effets d'un choc (c'est à dire ici une contrainte forte, localisée en espace, et se déplaçant rapidement) sur une dislocation. Dans le Chapitre 8, nous proposerons des stratégies numériques permettant de résoudre l'équation de Peierls-Nabarro Dynamique et de simuler de tels phénomènes.

Le plan de ce chapitre est le suivant : tout d'abord nous détaillons le modèle de Peierls dans la Section 6.2, puis nous construisons l'équation de Peierls-Nabarro Dynamique dans la Section 6.3. Cette construction se fait en trois temps : d'abord, nous déduisons du modèle le terme intégrodifférentiel (pour le mode III et ensuite pour les modes I et II) ; puis, nous

1. Ce qui n'est pas le cas de l'équation de Weertman, où le cas pluridimensionnel pose de sérieux problèmes théoriques, à cause de la perte du principe de comparaison.

définissons précisément la réponse non-linéaire au niveau du plan de glissement ; enfin, nous rajoutons un terme phénoménologique de visco-plasticité. La donnée initiale est décrite dans la Section 6.4. On aboutit ainsi dans la Section 6.5 à un problème mathématique dont chaque terme est bien défini. Dans la Section 6.6, on décrit brièvement l'équation de Peierls-Nabarro Dynamique dans un cadre vectoriel.

## 6.2 Modèle

### 6.2.1 Le modèle de Peierls

Nous détaillons le modèle de Peierls [129] présenté brièvement dans l'introduction de la thèse.

Considérons un matériau homogène élastiquement isotrope scindé en deux parties (en fait des demi-espaces), qui sont au contact le long d'une interface $y = 0$ (voir Figure 6.1). Dans chacun des demi-espaces environnants, le matériau est soumis à l'équation d'élasticité linéaire homogène. On suppose que le modèle est planaire, c'est à dire que le déplacement est indépendant de la troisième composante $z$ (désormais, on omettra d'écrire cette composante), néanmoins, le déplacement dans la *direction* $(Oz)$ peut être non nul. L'interface constitue le plan de glissement de la dislocation. Le matériau présente un déplacement $\boldsymbol{u}(t, x, y) \in \mathbb{R}^3$, avec une discontinuité

$$\boldsymbol{\eta}(t, x) := \lim_{y \to 0^+} \boldsymbol{u}(t, x, y) - \boldsymbol{u}(t, x, -y)$$

au niveau de l'interface.

La liaison entre les demi-plans est assurée au niveau de l'interface plan $y = 0$ par une force $f$. Cette force est issue de deux contributions : d'une part un potentiel inter-atomique, qui fait intervenir la discontinuité $\boldsymbol{\eta}$, et d'autre part un chargement imposé, que l'on note $\boldsymbol{\sigma}^{\mathbf{a}}$. Nous précisons dans la Section 6.3.3 son expression.

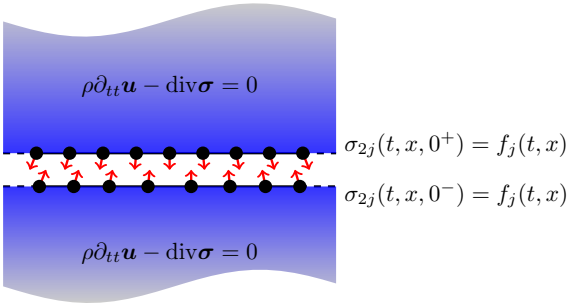



FIGURE 6.1 – Modèle de Peierls

*Remarque* 46. De manière plus réaliste, on modéliserait le matériau tout entier comme un ensemble d'atomes reliés par des potentiels (c'est à dire par un modèle de dynamique moléculaire). Aussi, le fait de supposer que le matériau est homogène élastique hors de l'interface

constitue une sorte de limite mésoscopique. Néanmoins, cette hypothèse simplificatrice sera d'une grande utilité pour se ramener à un problème posé seulement sur le plan de glissement.

### 6.2.2 Equations fondamentales

Mathématisons maintenant ces considérations physiques. Dans cette section seulement, on note $\partial_1$, $\partial_2$ et $\partial_3$ les dérivées partielles par rapport à $x$, $y$ et $z$, respectivement [2].

On note $\boldsymbol{\sigma}(t,x,y) \in \mathbb{R}^{3\times 3}$ le tenseur des contraintes. Par la loi de Cauchy, on a, sur l'interface $y=0$,

$$\sigma_{2j}(t,x,0^{\pm}) = f_j(t,x), \tag{6.1}$$

où $f$ est un chargement que l'on précisera plus tard. Puis, par la loi fondamentale de la dynamique,

$$\rho\partial_{tt}\boldsymbol{u} = \operatorname{div}(\boldsymbol{\sigma}), \tag{6.2}$$

où $\rho$ est la masse volumique (uniforme) du matériau considéré.

Le matériau est isotrope et présente un comportement élastique à l'intérieur des deux demi-plans inférieur et supérieur. Par la loi de Hooke, on peut donc fermer le système d'équations (6.1) et (6.2) par :

$$\sigma_{ij} = \mu(\partial_i u_j + \partial_j u_i) + \lambda \sum_{k=1}^{3} \partial_k u_k \delta_{ij}, \tag{6.3}$$

où $\lambda$ et $\mu$ sont les coefficients de Lamé, et $\delta$ est le symbole de Kronecker. Introduisons aussi, pour la suite de l'exposé, la vitesse des ondes de cisaillement $c_{\mathrm{s}} = \sqrt{\mu/\rho}$ et la vitesse des ondes longitudinales $c_{\mathrm{l}} = \sqrt{(2\mu+\lambda)/\rho}$, et leur rapport adimensionnel [3] :

$$\gamma = \frac{c_{\mathrm{l}}}{c_{\mathrm{s}}}. \tag{6.4}$$

## 6.3 L'équation intégrodifférentielle

Intuitivement, les seuls degrés de liberté du modèle se réduisent à la discontinuité de déplacement $\boldsymbol{\eta}(t,x)$ (pour $x \in \mathbb{R}$), qui est suffisante pour donner à $\boldsymbol{u}$ des conditions aux limites : l'information sur $\boldsymbol{u}$ loin du plan de glissement est redondante. Grâce à la géométrie très simple du modèle de Peierls, il est possible de construire explicitement une équation intégrodifférentielle sur $\boldsymbol{\eta}$.

2. Comme les fonctions considérées sont indépendantes de $z$, l'opérateur $\partial_3$ est trivial. Cependant, on a voulu utiliser cette notation pour conserver une certaine simplicité des formules de mécanique.

3. Ce choix est différent de celui de [48, p. 1416], où est défini $\gamma := c_{\mathrm{s}}/c_{\mathrm{l}}$. Comme $c_{\mathrm{s}} < c_{\mathrm{l}}$, on aura ici $\gamma > 1$.

### 6.3.1 Partie linéaire dans le cas du mode III

On suppose ici que seule la composante $u_3$ est non nulle, ce qui correspond à une dislocation vis, aussi appelée mode III. Par abus de notation, on identifie $\eta$ et $f$ à $\eta_3$, respectivement $f_3$ (sachant que les autres composantes sont nulles).

D'une part, d'après (6.2), dans chacun des demi-espaces, on a l'équation des ondes induite par l'élasticité linéaire

$$\partial_{tt} u_3(t,x,y) - c_{\rm s}^2 \Delta u_3(t,x,y) = 0, \tag{6.5}$$

pour $t, x \in \mathbb{R}$, $y \in ]-\infty, 0[\cup]0, +\infty[$. D'autre part, grâce à (6.1), on a

$$\mu \partial_y u_3\left(t,x,0^{\pm}\right) = f(t,x), \tag{6.6}$$

pour $t \in \mathbb{R}, x \in \mathbb{R}$, c'est à dire une condition de Neumann. On fait l'hypothèse que l'on part d'un matériau au repos, avec $u_3(t \le 0, \cdot, \cdot) = 0$ et $\partial_t u_3(t \le 0, \cdot, \cdot) = 0$.

Pour résoudre (6.5), on emploie naturellement la transformation de Laplace $\mathcal{L}$ par rapport au temps $t$, et la transformation de Fourier $\mathcal{F}$ par rapport à la première variable spatiale $x$. Ainsi, on réécrit (6.5) comme

$$\partial_{yy} \mathcal{L}\mathcal{F} u_3(p,k,y) = \left(c_{\rm s}^{-2} p^2 + k^2\right) \mathcal{L}\mathcal{F} u_3(p,k,y).$$

En ne considérant que les solutions qui tendent vers 0 en lorsque $|y| \to +\infty$, on obtient

$$\mathcal{L}\mathcal{F} u_3(p,k,y) = \begin{cases} \mathcal{L}\mathcal{F} u_3(p,k,0^+) \mathrm{e}^{-|y|\sqrt{c_{\rm s}^{-2}p^2+k^2}} & \text{si} \quad y > 0, \\ \mathcal{L}\mathcal{F} u_3(p,k,0^-) \mathrm{e}^{-|y|\sqrt{c_{\rm s}^{-2}p^2+k^2}} & \text{si} \quad y < 0. \end{cases} \tag{6.7}$$

Par définition de $\eta$, on déduit de (6.6) et (6.7) que

$$-\frac{\mu}{2}\sqrt{c_{\rm s}^{-2}p^2+k^2}\, \mathcal{L}\widehat{\eta}(p,k) = \mathcal{L}\widehat{f}(p,k), \qquad \text{pour} \quad (p,k) \in \mathbb{R}_+ \times \mathbb{R},$$

où on emploie la notation $\mathcal{F} f = \widehat{f}$.

En adimensionnant ce qui précède via $t \mapsto c_{\rm s} t$ et $f \mapsto -2\mu^{-1} f$, et en utilisant la transformation de Laplace inverse, on vérifie que $\widehat{\eta}(t,k)$ satisfait alors l'équation intégrodifférentielle suivante :

$$\kappa_i \partial_t \widehat{\eta}(t,k) + k^2 \int_{-\infty}^{t} C_i(|k|(t-t'))\widehat{\eta}(t',k)\mathrm{d}t' = \widehat{f}(t,k), \tag{6.8}$$

pour $(t,k) \in \mathbb{R} \times \mathbb{R}$. Dans (6.8), on a $i = \mathrm{III}$ ; la constante $\kappa_{\rm III}$ et la fonction $C_{\rm III}$ sont définies par :

$$\kappa_{\rm III} := 1, \tag{6.9}$$

$$C_{\rm III}(T) := \mathcal{L}^{-1}\left\{\sqrt{p^2+1} - p\right\}(T) = \frac{J_1(T)}{T}. \tag{6.10}$$

Dans l'expression ci-dessus, $J_1$ désigne une fonction de Bessel du premier type. La fonction $C_{\rm III}$ est tracée sur la Figure 6.2. D'autres expressions de $C_{\rm III}$, et des formulations équivalentes à (6.8) peuvent être employées (voir l'Annexe A.5.1).

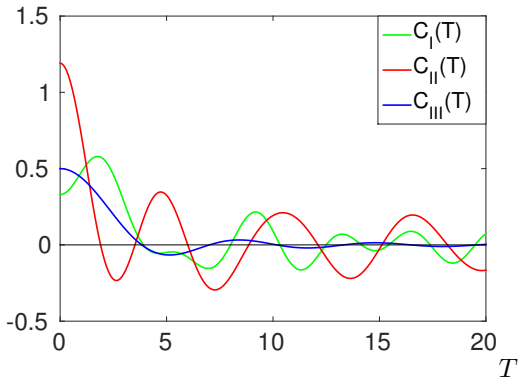



FIGURE 6.2 – Tracé de $C_{\mathrm{I}}(T)$, $C_{\mathrm{II}}(T)$ et $C_{\mathrm{III}}(T)$, pour $\gamma = \sqrt{3} \simeq 1.7$.

### 6.3.2 Partie linéaire pour les modes I et II

Dans la Section 6.3.1, nous avons construit une équation (6.8) dans le cas des dislocations-vis, en supposant que seule la composante $u_3$ du déplacement $\boldsymbol{u}$ du matériau était non-nulle. Pareille construction est possible en supposant successivement que seule la composante $u_1$, puis $u_2$, est non-nulle. On retrouve alors l'équation (6.8), respectivement avec

1. $i = \mathrm{I}$, pour le mode I ou coin de montée,
2. $i = \mathrm{II}$, pour le mode II ou coin de glissement.

Nous donnons maintenant l'expression des coefficients $\kappa_{\mathrm{I}}$ et $\kappa_{\mathrm{II}}$ et des fonctions $C_{\mathrm{I}}$ et $C_{\mathrm{II}}$ obtenus (voir [48, 62, 130]). Comme on peut le constater, ces dernières ont des expressions plus complexes que $C_{\mathrm{III}}$, et font intervenir le rapport adimensionnel $\gamma$ défini par (6.4).

On a les formules suivantes en ce qui concerne le mode I :

$$\begin{aligned}
&\kappa_{\mathrm{I}} = \gamma,\\
&C_{\mathrm{I}}(T) = \gamma^3 \frac{J_1(\gamma T)}{\gamma T} + 4T\left(W(\gamma T) - W(T)\right) + (4\gamma - \gamma^3)J_0(\gamma T) - 4J_0(T),\\
&\mathcal{L}C_{\mathrm{I}}(p) = -4p^{-2}\sqrt{1+p^2} - \gamma p + p^{-2}\frac{\left(2+p^2\right)^2}{\sqrt{1+\left(\frac{p}{\gamma}\right)^2}}.
\end{aligned}$$

où $W(T)$ est la fonction

$$W(T) := \int_0^T \frac{J_1(T')}{T'} dT'.$$

Pour le mode II, on a les formules suivantes :

$$\begin{aligned}
&\kappa_{\mathrm{II}} = 1,\\
&C_{\mathrm{II}}(T) = \frac{J_1(T)}{T} + 3J_0(T) - \frac{4}{\gamma}J_0(\gamma T) - 4T\left(W(\gamma T) - W(T)\right),\\
&\mathcal{L}C_{\mathrm{II}}(p) = \frac{p^2}{\sqrt{1+p^2}} + 4p^{-2}\left(\sqrt{1+p^2} - \sqrt{1+\left(\frac{p}{\gamma}\right)^2}\right) - p.
\end{aligned}$$

Les fonctions $C_{\mathrm{I}}$ et $C_{\mathrm{II}}$ sont représentées sur la Figure 6.2.

En Annexe A.5, le lecteur trouvera d'autres formules utiles pour les modes I et II.

*Remarque* 47. Du point de vue des applications, les modes II et III sont très intéressants. Notons que le mode II a une phénoménologie plus riche que le mode III. Le mode I sera très peu abordé dans cette étude.

*Remarque* 48 (Comparaison avec la littérature). Les fonctions $C_{\mathrm{I}}$, $C_{\mathrm{III}}$ sont en accord avec [62, (26)]. En revanche, la formule [62, (26)] donnant $C_{\mathrm{II}}$ présentait un faute de frappe, corrigée par [48, (A.1)].

### 6.3.3 Partie non-linéaire

Une spécificité de l'équation de Peierls-Nabarro Dynamique par rapport aux équations de la rupture (comme dans [62]) réside dans le terme de forçage $f$ de (6.6). Celui-ci synthétise deux contributions : d'une part, la force de rappel inter-atomique, et d'autre part un chargement $\sigma^{\mathrm{a}}$, imposé de manière exogène (en pratique, ce chargement $\sigma^{\mathrm{a}}$ est fixé par l'expérimentateur).

La force de rappel inter-atomique est issue d'un potentiel $F$. En général, on reconstruit $F$, aussi appelé $\gamma$-surface, de façon empirique ou par calcul *ab initio* (voir par exemple [108]). Le potentiel $F$ est périodique : cette périodicité est héritée de la structure cristalline sous-jacente du matériau étudié.

Il faut prendre garde à une subtilité lorsque l'on cherche à évaluer la force inter-atomique. En effet, $\eta$ mesure un défaut d'élasticité ; par conséquent, il ne représente pas exactement le glissement inter-atomique mais la partie *plastique* de ce dernier (voir [131, (5)], où $\widetilde{\eta}$ correspond à ce que l'on note ici $\eta$). Ainsi, selon [131, (8)] le terme de forçage $f$ (adimensionné) s'exprime comme

$$f(t,x) := -2\mu^{-1}\left(F'(\eta(t,x) + \eta_{\mathrm{e}}(t,x)) - \sigma^{\mathrm{a}}(t,x)\right), \tag{6.11}$$

où $\eta_{\mathrm{e}}(t,x)$ est la réponse élastique *non-linéaire* à $\sigma^{\mathrm{a}}$. Si $|\sigma^{\mathrm{a}}(t,x)|$ n'est pas trop grand, cette réponse élastique est définie implicitement par $F'(\eta_{\mathrm{e}}(t,x)) = \sigma^{\mathrm{a}}(t,x)$ (voir [131, (4)]). En pratique, $F'$ est une fonction bornée Par conséquent l'équation $F'(u) = \sigma^{\mathrm{a}}(t,x)$ n'a pas de solution si $|\sigma^{\mathrm{a}}(t,x)|$ est trop grand : à ce moment-là, $\eta_{\mathrm{e}}(t,x)$ sature à une valeur en laquelle $F'$ admet un extremum. On formalise cela par la définition suivante :

$$\eta_{\mathrm{e}}(t,x) = \arg\min\left\{|\eta|, \left|F'(\eta) - \sigma^{\mathrm{a}}(t,x)\right| = \inf\left\{\left|F'(u) - \sigma^{\mathrm{a}}(t,x)\right|, u \in \mathbb{R}\right\}\right\}, \tag{6.12}$$

qui a une unique solution si le potentiel $F$ est pair (ce qui est physiquement réaliste).

Dans le cas de l'équation de Weertman, une telle distinction entre $\eta_{\mathrm{e}}$ et $\eta$ n'a pas d'intérêt. En effet, on a la partie élastique $\eta_{\mathrm{e}}$ constante et égale à $\eta(+\infty)$, dans le formalisme de l'équation de Weertman adopté dans l'introduction.

*Remarque* 49. Dans ce modèle, on considère une relation d'élasticité non-linéaire au niveau du plan de glissement. Toutefois, celle-ci se ramène à de l'élasticité linéaire pour des petites déformations.

### 6.3.4 Visco-plasticité

De la même manière que [135, Model 1], la référence [131, (9)] ajoute à l'équation un terme phénoménologique de viscosité plastique, lequel traduit une dissipation due aux phonons (voir par exemple [65]). Ainsi, nous remplaçons $f(t,x)$ dans (6.6) par

$$\widetilde{f}(t,x) := F'(\eta(t,x)+\eta_{\mathrm{e}}(t,x)) - \sigma^{\mathrm{a}}(t,x) + \frac{\alpha\mu c_s}{2}\partial_t\eta(t,x),$$

où la constante adimensionnelle $\alpha \in ]0,1]$ est en général petite, de l'ordre de $10^{-3}$ à $10^{-1}$ (voir [131,135]). Comme le terme visco-plastique est linéaire en $\eta$, on l'englobe dans la partie linéaire, et on transforme ainsi (6.8) en

$$\kappa_i^{\alpha}\partial_\tau\widehat{\eta}(\tau,k) + k^2\int_{-\infty}^{\tau} C_i(|k|(\tau-\tau'))\eta(\tau',k)\mathrm{d}\tau' = \widehat{f}(\tau,k), \tag{6.13}$$

où

$$\kappa_i^{\alpha} := \kappa_i + \alpha, \tag{6.14}$$

et $f$ est définie par (6.11). Désormais, nous désignons l'équation (6.13), couplée avec (6.11) et (6.12), par le nom d'équation de Peierls-Nabarro Dynamique (voir [130]).

L'ajout du terme visco-plastique se traduit par un amortissement (voir Section 7.3.2). Le phénomène de visco-plasticité est à distinguer de la visco-élasticité, qui se manifeste par l'ajout à (6.5) d'un nouveau terme :

$$\partial_{tt}u_3(t,x,y) + \alpha_{\mathrm{e}}\partial_t u_3(t,x,y) - c_{\mathrm{s}}^2\Delta u_3(t,x,y) = 0, \tag{6.15}$$

Par un calcul similaire à celui entrepris dans la Section 6.3.1, il est possible (voir [110, p. 372]) de calculer la noyau associé la partie linéaire, et de l'insérer à la place de $C_{\mathrm{III}}$ dans (6.8). Ce phénomène, qui possède aussi des propriétés stabilisatrices, ne sera pas étudié dans ce document.

## 6.4 Donnée initiale

L'équation (6.13) est insuffisante à elle seule pour déterminer $\eta(t,x)$. Il faut lui adjoindre une "donnée initiale", c'est à dire prescrire la fonction $\eta(t,x)$ pour tous les temps passés. En un certain sens, il s'agit d'avoir *déjà* à disposition une solution de référence de l'équation de Peierls-Nabarro Dynamique. Les fronts progressifs solutions de l'équation de Weertman sont pour cela de bons candidats (et ils présentent en outre un intérêt physique [131]).

*Remarque* 50. Il peut sembler paradoxal de vouloir imposer une donnée initiale à (6.13), dans la mesure où la construction de cette équation dans la Section 6.3.1 présuppose que l'on part d'un matériau au repos (ce qui implique en particulier que $\eta(t \leq 0, \cdot) = 0$). Dans ce premier cas, l'intégrale de (6.13) porte sur $\tau' \in [0, \tau[$, et non $\tau' \in [-\infty, \tau[$, ce qui suffit pour construire une unique solution $\eta$. Cependant, dans le cas où la donnée initiale est un front progressif $\eta_0(t, x)$, le raisonnement de la Section 6.3.1 s'applique en fait de manière perturbative par rapport à cette "donnée initiale" (la perturbation étant nulle pour les temps négatifs).

### 6.4.1 Dislocations à vitesse constante et équation de Weertman

Il est montré dans [130] que les fronts progressifs pour l'équation de Peierls-Nabarro Dynamique satisfont l'équation de Weertman. Rappelons que cette dernière décrit une dislocation [4] en mouvement à vitesse constante $\eta_{\mathrm{e}} + \eta(t, x) = \phi(x - vt)$, et s'écrit

$$-\mu A(v) \left|\partial_x\right| \phi(x) + \mu \left( B(v) + \frac{\alpha v}{2c_{\mathrm{s}}} \right) \phi'(x) = F'(\phi(x)) - \sigma \quad \text{pour} \quad x \in \mathbb{R}, \tag{6.16}$$

où $|\partial_x|$ est l'opérateur $(-\Delta)^{1/2}$ (dont le symbole en Fourier est $|k|$), et $A(v)$ et $B(v)$ sont des scalaires donnés par les formules [130, (46), (47) et (48)]. La fonction $\phi$ satisfait les conditions à l'infini

$$\lim_{x \to -\infty} \phi(x) = \phi_{\mathrm{l}} \quad \text{et} \quad \lim_{x \to +\infty} \phi(x) = \phi_{\mathrm{r}} \tag{6.17}$$

qui sont telles que

$$F'(\phi_{\mathrm{r}}) - \sigma = F'(\phi_{\mathrm{l}}) - \sigma = 0. \tag{6.18}$$

Pour plus de détails, nous renvoyons à l'introduction, aux Chapitres 4 et 5, et à l'Annexe A.2.

L'application $v \mapsto c(v)$ n'est pas injective. Par conséquent, pour un chargement donné $\sigma$ constant et sous-critique (c'est à dire que $F'(u) = \sigma$ possède au moins une solution), il existe en général plusieurs vitesses possibles $v$. En particulier, même si on suppose que $\sigma^{\mathrm{a}}(t, x)$ est constant et uniforme pour $t < 0$, il n'existe pas de donnée initiale "naturelle" associée pour l'équation de Peierls-Nabarro Dynamique.

Une étude numérique de [131] établit que les solutions de l'équation de Peierls-Nabarro Dynamique tendent vers des fronts progressifs $\phi(x - tv)$, pour $\phi$ et $v$ satisfaisant (6.16), pour une classe particulière d'Ansatz, dans certains cas où le chargement $\sigma^{\mathrm{a}}$ est stationnaire égal à $\sigma$ à partir d'un certain temps. Cette même étude montre que la vitesse $v$ du front progressif limite ne dépend pas seulement de la valeur de $\sigma$, mais bien de l'historique de la dislocation. Elle suggère au passage qu'il existe des branches de vitesses stables, et d'autres branches qui sont instables, au sens de l'équation de Peierls-Nabarro Dynamique.

L'équation (6.16) ne possède en général pas de solution analytique connue, sauf dans certains cas particuliers de potentiel $F$. Aussi, la résolution numérique de l'équation (6.16)

4. On note que l'on somme ici la partie élastique $\eta_{\mathrm{e}}$ satisfaisant (6.12) et la partie plastique $\eta$.

est un prérequis pour la simulation de l'équation de Peierls-Nabarro Dynamique dans un cadre général. Le rôle de l'équation de Weertman dans la résolution et l'étude de l'équation de Peiers-Nabarro Dynamique a constitué une motivation importante pour le développement d'une méthode numérique capable d'approximer ses solutions, pour des potentiels $F$ bistables généraux (voir Chapitre 5).

### 6.4.2 Définition de la donnée initiale

Une manière simple de construire une donnée initiale à partir d'une solution de l'équation de Weertman est de se donner trois constantes $\sigma_0^{\mathrm{a}}$ et $\eta_{0,\mathrm{e}} = \phi_{\mathrm{r}}$ et $\phi_{\mathrm{r}}$ telles que

$$F'(\eta_{0,\mathrm{e}}) = F'(\phi_1) = \sigma_0^{\mathrm{a}}. \tag{6.19}$$

On en déduit une fonction $\phi_0(x-vt)$ satisfaisant (6.16) et (6.17). Ainsi, la fonction $\eta_0(t,x) := \phi_0(x-vt) - \eta_{0,\mathrm{e}}$ satisfait

$$\begin{aligned} &\kappa_i^\alpha \partial_t \widehat{\eta_0}(t,k) + k^2 \int_{-\infty}^{t} C_i(|k|(t-t'))\widehat{\eta_0}(t',k)\mathrm{d}t' \\ &= -\frac{2}{\mu}\mathcal{F}\left\{F'(\eta_0(t,\cdot)+\eta_{0,\mathrm{e}}) - \sigma_0^{\mathrm{a}}\right\}(k), \end{aligned} \tag{6.20}$$

pour tout temps $t \in \mathbb{R}$, et mode de Fourier $k \in \mathbb{R}$.

On utilise $\eta_0$ comme donnée initiale en prescrivant la solution pour tous les temps passés, c'est à dire en imposant la condition suivante :

$$\eta(t,\cdot) = \eta_0(t,\cdot), \qquad \forall t < 0. \tag{6.21}$$

On pose alors [5]

$$u(t,x) := \eta(t,x) - \eta_0(t,x),$$

et on déduit de (6.13) et de (6.20) l'équation suivante sur $u$ :

$$\begin{cases} \kappa_i^\alpha \partial_t \widehat{u}(t,k) + k^2 \displaystyle\int_0^t C_i(|k|(t-t'))\widehat{u}(t',k)\mathrm{d}t' = \mathcal{F}\left\{f_{\sigma^a}[u](t,\cdot)\right\}(k), \\ u(0,\cdot) = 0. \end{cases} \tag{6.22}$$

pour tout $t \geq 0$, $k \in \mathbb{R}$, où

$$\begin{aligned} f_{\sigma^a}[u](t,x) = -\frac{2}{\mu}\Bigg( &F'(u(t,x)+\eta_0(t,x)+\eta_{\mathrm{e}}(t,x)) - F'(\eta_0(t,x)+\eta_{0,\mathrm{e}}) \\ &- \sigma^{\mathrm{a}}(t,x) + \sigma_0^{\mathrm{a}}\Bigg). \end{aligned} \tag{6.23}$$

5. $u$ défini ci-dessus n'a pas de rapport avec le déplacement $\boldsymbol{u}$ de la Section 6.2

*Remarque* 51 (Donnée initiale générale). Il n'est pas nécessaire que la donnée initiale soit une dislocation en régime stationnaire. En fait, on peut prendre n'importe quelle solution de (6.13) (avec le chargement $\sigma^{\mathrm{a}}(t,x)$) comme donnée initiale du problème. En Annexe A.5.2, on construit par exemple une donnée initiale approchée correspondant au croisement de deux dislocations.

## 6.5 Formalisation du problème

Nous avons à présent tous les ingrédients pour énoncer le problème mathématique que nous nous proposons de résoudre. Rappelons que nous avons besoin de trois types d'informations :

1. des données sur la physique du problème étudié ;
2. une donnée initiale ;
3. un chargement.

La physique du problème est liée à des caractéristiques mécaniques du matériau et à la géométrie de la dislocation. L'équation (6.22) est une forme adimensionnelle de l'équation de Peierls-Nabarro Dynamique, et fait donc intervenir un petit nombre de constantes et de fonctions sans dimension. Cela se résume ici à un type de dislocation $i \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$, une constante de phonons $\alpha > 0$, un rapport de vitesses $\gamma > 0$, et un potentiel $\mu^{-1}F \in \mathrm{C}^1_{\mathrm{per}}(\mathbb{R})$. Alors, on peut construire le coefficient $\kappa_i^\alpha > 0$ et la fonction $C_i$ grâce aux formules de la Section 6.3.

L'état initial correspond au passé de la dislocation. Il est encapsulé dans les fonctions $\eta_0(t,x)$ et $\mu^{-1}\sigma_0^{\mathrm{a}}(t,x)$, et la partie élastique associée $\eta_{0,\mathrm{e}}(t,x)$.

Enfin, un chargement $\mu^{-1}\sigma^{\mathrm{a}}(t,x)$, est imposé pour $t \in \mathbb{R}_+$ et $x \in \mathbb{R}$, d'où on déduit $\eta_{\mathrm{e}}(t,x)$ satisfaisant (6.12).

Grâce à (6.23), on construit alors une fonction $f_{\sigma^a}[u](t,x)$. Finalement, résoudre l'équation de Peierls-Nabarro Dynamique, c'est trouver $u$ satisfaisant (6.22) pour tous $t \geq 0$ et $k \in \mathbb{R}$.

L'équation (6.22) n'est qu'une *reformulation* compacte du problème originel (6.5) et (6.6). Son intérêt réside en la réduction du problème initial, posé sur $(t,(x,y)) \in \mathbb{R}_+ \times \mathbb{R}^2$, à un problème de dimension plus petite, posé sur $(t,x) \in \mathbb{R}_+ \times \mathbb{R}$. Le gain dimensionnel est contrebalancé par l'apparition du terme de mémoire, qui devient une nouvelle dimension du problème. C'est un choix délibéré de notre part que de chercher à résoudre le problème sous sa forme (6.22), et non sous la forme du système (6.5) avec les conditions de bord (6.6). D'autres angles d'approche (numériques, car les problèmes sont équivalents) sont possibles, par exemple en utilisant des conditions au bord artificielles pour résoudre (6.5) et (6.6) (voir par exemple [66]). Nous ne les avons toutefois pas étudiés.

## 6.6 Equation de Peierls-Nabarro Dynamique vectorielle

Dans la Section 6.3, on suppose qu'une seule des trois composantes du déplacement $u$ est non nulle. A l'instar ce que se fait en statique (voir par exemple [57]), un modèle plus

complet couplerait les trois modes d'une dislocation. Une telle équation prendrait la forme suivante :

$$\boldsymbol{\kappa}^{\alpha} \odot \partial_t \widehat{\boldsymbol{u}}(t,k) + k^2 \int_0^t \boldsymbol{C}(|k|(t-t')) \odot \widehat{\boldsymbol{u}}(t',k)\mathrm{d}t' = \widehat{\boldsymbol{f}}(t,k), \tag{6.24}$$

où le symbole $\odot$ est utilisé pour le produit de Hadamard

$$\begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} \odot \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} = \begin{pmatrix} f_1 g_1 \\ f_2 g_2 \\ f_3 g_3 \end{pmatrix}.$$

Les fonctions à valeur dans $\mathbb{R}^3$ $\boldsymbol{\eta}_{\mathrm{e}}$, $\boldsymbol{\eta}_{0,\mathrm{e}}$, $\boldsymbol{\sigma}^{\mathrm{a}}$, $\boldsymbol{\sigma}_0^{\mathrm{a}}$, satisfont des analogues tridimensionnels de (6.12), (6.20) et (6.19), où $\boldsymbol{C} = (C_{\mathrm{I}}, C_{\mathrm{II}}, C_{\mathrm{III}})$, $\boldsymbol{\kappa}^{\alpha} = (\kappa_{\mathrm{I}}^{\alpha}, \kappa_{\mathrm{II}}^{\alpha}, \kappa_{\mathrm{III}}^{\alpha})$ et

$$\begin{aligned} \boldsymbol{f}(t,x) = -\frac{2}{\mu}\Bigg( & \nabla F(\boldsymbol{u}(t,x) + \boldsymbol{\eta}_0(t,x) + \boldsymbol{\eta}_{\mathrm{e}}(t,x)) - \nabla F\left(\boldsymbol{\eta}_0(t,x) + \boldsymbol{\eta}_{0,\mathrm{e}}\right) \\ & - \boldsymbol{\sigma}^{\mathrm{a}}(t,x) + \boldsymbol{\sigma}_0^{\mathrm{a}} \Bigg), \end{aligned}$$

Le modèle pourrait en outre être enrichi en prenant en compte le caractère anisotrope d'un matériau.

Nous ne mentionnons cependant (6.24) que comme un objectif à long terme ; nous ne l'étudierons pas dans ce document. Soulignons au passage que la construction de l'équation (6.24) requiert a priori d'utiliser de l'élasticité anisotrope pour être complète, et que la construction d'une telle équation n'a jamais été entreprise dans la littérature physique, à notre connaissance. Néanmoins, nous pensons que, même si cette hypothétique équation (6.24) est indéniablement plus complexe que (6.22), elle n'introduit pas de difficulté conceptuelle supplémentaire du point de vue numérique.

# Chapitre 7

# Quelques propriétés mathématiques de l'équation de Peierls-Nabarro Dynamique

Dans ce court chapitre, nous indiquons quelques propriétés mathématiques utiles de l'équation de Peierls-Nabarro Dynamique.

Ce travail a été effectué en collaboration avec Claude Le Bris, Frédéric Legoll et Yves-Patrick Pellegrini.

Le lecteur trouvera en Annexe 7 des matériaux théoriques additionnels.

## 7.1 Introduction

Dans ce chapitre, on s'intéresse aux équations du type

$$\begin{cases} \partial_t \widehat{u}(t,k) = -k^2 \displaystyle\int_0^t C(|k|(t-t'))\widehat{u}(t',k)\mathrm{d}t' + \widehat{f}(t,k), \\ \widehat{u}(0,k) = \widehat{u}_0(k), \end{cases} \tag{7.1}$$

écrites en la variable de Fourier $k \in \mathbb{R}$ et le temps $t \geq 0$, où la fonction $u(t,x)$ est l'inconnue, et où la fonction régulière $C$ est un noyau, et la fonction $f(t,x)$ un forcage régulier. On cherche à en caractériser quelques propriétés mathématiques utiles. On garde évidemment en tête les cas $C = C_i/\kappa_i^\alpha$ (voir (6.13)), mais les considérations développées ci-dessous sont plus générales. En outre, dans le cas de l'équation de Peierls-Nabarro Dynamique (6.22), la fonction $f(t,x)$ dépend non-linéairement de $u(t,x)$ (voir (6.23)) et on a $\widehat{u}_0 = 0$. *In fine*, l'objectif est de préparer la résolution numérique de (7.1) qui sera faite dans le Chapitre 8 suivant.

On peut reformuler en espace-temps l'équation (7.1), écrite en variable de Fourier. Nous commentons quelques aspects de cette reformulation dans la Section 7.2, qui semble cependant moins facile à manier que (7.1).

En effet, l'équation (7.1) peut se lire comme une famille d'équations intégrodifférentielles linéaires de Volterra du second type à noyau convolutif (voir [10, 105]) indicées par le mode de Fourier $k$. Autrement dit, la transformation de Fourier a diagonalisé la partie linéaire de (7.1). Un simple changement de variable

$$\tau = |k|t \tag{7.2}$$

ramène (7.1) à une seule et même équation, avec un second membre dépendant de $|k|$. En posant $\widetilde{u}_k(\tau) := \widehat{u}(t,k)$ et $\widetilde{f}_k(\tau) := \widehat{f}(t,k)$, on obtient en effet

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\widetilde{u}_k(\tau) = -\int_0^\tau C(\tau-\tau')\widetilde{u}_k(\tau')\mathrm{d}\tau' + |k|^{-1}\widetilde{f}_k(\tau). \tag{7.3}$$

Ainsi, chaque mode de Fourier $\widehat{u}(\cdot,k)$ évolue de la même manière, mais avec un temps dilaté par un facteur $|k|$. Le problème à résoudre (7.3) demeure non-trivial si $C = C_i$, car les noyaux $C_i(T)$ sont oscillants et tendent lentement vers 0 à l'infini (voir Figure 6.2), c'est à dire comme $T^{-3/2}$ pour le mode III, et comme $T^{-1/2}$ pour le mode II.

La linéarité et la structure de convolution de (7.1) permettent de représenter sa solution grâce à une formule de Duhamel

$$\widehat{u}(t,k) = \mathfrak{R}(|k|t)\widehat{u}_0(k) + \int_0^t \mathfrak{R}(|k|(t-t'))\widehat{f}(t',k)\mathrm{d}t'. \tag{7.4}$$

La formule (7.4) fait intervenir une fonction particulière : la résolvante $\mathfrak{R}$ (définie plus bas par (7.7)), qui traduit des propriétés d'amortissement de (7.1). Dans la Section 7.3, nous étudions en détail cette fonction pour le mode III de dislocation, et plus brièvement pour les modes I et II.

Comme les équations (7.1) et (7.4) partagent la même solution, elles ne sont que deux expressions d'un même problème. On parle de *forme directe* pour désigner (7.1) et de *forme résolue* pour désigner (7.4). Néanmoins, l'expression (7.4) est très utile, et permet notamment de démontrer (sous des hypothèses simples) l'existence et l'unicité d'une solution de (7.1), dans le cas où le second membre $f(t,x)$ dépend nonlinéairement de $u(t,x)$. C'est l'objet de la Section 7.4.

Nous concluons ce chapitre par une remarque sur les dérivées fractionnaires temporelles (voir Section 7.5).

## 7.2 Reformulation espace-temps

Pour fixer les idées, nous rappelons ici la forme originelle (voir [130]) de l'expression du noyau intégrodifférentiel dans l'équation de Peierls-Nabarro Dynamique (en mode III). Cette formulation spatio-temporelle est malcommode numériquement, et ne sera pas utilisée par la suite.

La transformation de Fourier diagonalise l'opérateur linéaire

$$\mathcal{O}_i : u \mapsto v : (t,x) \mapsto \mathcal{F}^{-1}\left\{-k^2 \int_0^t C_i(|k|(t-t'))\widehat{u}(t',k)\mathrm{d}t'\right\}(x).$$

C'est donc un opérateur de convolution en *espace*-temps. C'est ainsi qu'il a été initialement défini dans [130].

Si $i = \mathrm{III}$, cet opérateur s'écrit (voir [130, (30) et (35)])

$$\mathcal{O}_{\mathrm{III}}\{u\}(t,x) = -\frac{2}{\pi}\int_0^t \int_{\mathbb{R}} K(t-t', x-x')\partial_x \eta(t',x')\mathrm{d}x'\mathrm{d}t' \tag{7.5}$$

où

$$K(t,x) = \frac{x}{2t^2}\frac{\mathbb{1}_{\mathbb{R}_+}(t-|x|)}{\sqrt{t^2-x^2}}. \tag{7.6}$$

Des formules similaires existent pour les modes I et II dans [130, (38), (39) et (43)]. Toutes ces formules ont en commun une propriété de localité en espace : la valeur $\mathcal{O}_i\{u\}(t,x)$ ne dépend que de $u(t',x')$ à l'intérieur du cône $t' \leq t$ et $|x-x'| \leq c|t-t'|$, où $c = 1$ pour le mode III et $c = \gamma$ pour les modes I et II. Cette propriété, peu visible sur la transformée de Fourier de l'équation de Peierls-Nabarro Dynamique, se manifeste sur (7.6) par le fait que $K(t,\cdot)$ est à support dans $[-t,t]$.

## 7.3 Résolvantes de l'équation de Peiers-Nabarro Dynamique

La structure de convolution en temps et la linéarité de (7.1) traduisent une invariance en temps de l'équation, propriété qui apparaît sur la formule de Duhamel (7.4) satisfaite par $\widehat{u}$. Nous en explicitons la construction dans la Section 7.3.1. Cette formule est très utile car elle permet de lire des propriétés de stabilité de (7.1) sur le comportement asymptotique de la résolvante $\mathfrak{R}$ associée (définie par l'équation (7.7) ci-dessous).

### 7.3.1 Résolvante et formule de Duhamel

Nous suivons à présent la démarche de [105, Sec. 3.2 pp. 35-37] (la différence ici est que nous considérons une équation intégrodifférentielle et non intégrale), afin de construire une formule de Duhamel.

On définit la *résolvante* $\mathfrak{R}$ associée à (7.3), comme étant la solution du problème homogène suivant :

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}\tau}\mathfrak{R}(\tau) = -\displaystyle\int_0^\tau C(\tau-\tau')\mathfrak{R}(\tau')\mathrm{d}\tau', \\ \mathfrak{R}(0) = 1. \end{cases} \tag{7.7}$$

Grâce à la transformation de Laplace (qui transforme une convolution en une multiplication), on obtient l'identité suivante :

$$\mathcal{L}\mathfrak{R}(p) = \frac{1}{p + \mathcal{L}C(p)}. \tag{7.8}$$

Dans les cas particuliers où $C = C_i/\kappa_i^\alpha$ (voir (6.13)), on note la résolvante $\mathfrak{R}_i^\alpha$. Cette fonction est très utile pour caractériser les propriétés de stabilité de l'équation (7.3).

Munis de la résolvante, nous pouvons à présent exprimer de façon explicite $u$ à partir de $f$ et $u(0,\cdot)$ :

**Proposition 7.3.1.** *Supposons que le noyau $C$ est régulier, et que $\widehat{f}(t,k)$ est donnée et régulière en temps. Alors, la solution de l'équation* (7.1) *satisfait* (7.4).

*Démonstration de la Proposition 7.3.1.* Comme le noyau $C$ est régulier, grâce au [105, Th. 3.1 p. 30], il existe une unique solution $\widehat{u}$ à (7.1), laquelle est continue en $t$. Par conséquent, il suffit de vérifier que $\widehat{u}$ définie par (7.4) satisfait bien (7.1). De plus, il suffit de considérer le mode de Fourier $k = 1$, les autres cas s'y ramenant par simple dilatation.

On suppose donc que $u$ est définie par (7.4). On déduit de (7.7) que

$$\begin{aligned} \partial_t\widehat{u}(t,1) = &-\left(\int_0^t C(t-t')\mathfrak{R}(t')\mathrm{d}t'\right)\widehat{u}_0(1) \\ &-\int_0^t\left(\int_0^{(t-t')} C(t-t'-t'')\mathfrak{R}(t'')\mathrm{d}t''\right)\widehat{f}(t',1)\mathrm{d}t' + \mathfrak{R}(0)\widehat{f}(t,1). \end{aligned}$$

Puis, comme $\mathfrak{R}(0) = 1$ et par le changement de variables $s' = t' + t''$ et $s'' = t'$ dans la seconde intégrale ci-dessus, on obtient

$$\begin{aligned} \partial_t\widehat{u}(t,1) &= -\int_0^t C(t-t')\mathfrak{R}(t')\widehat{u}_0(1)\mathrm{d}t' \\ &\quad -\int_0^t C(t-s')\left(\int_0^{s'}\mathfrak{R}(s'-s'')\widehat{f}(s'',1)\mathrm{d}s''\right)\mathrm{d}s' + \widehat{f}(t,1) \\ &= -\int_0^t C(t-t')\widehat{u}(t',1)\mathrm{d}t' + \widehat{f}(t,1). \end{aligned}$$

Ainsi, $\widehat{u}(t,k)$ définie par (7.4) est solution de l'équation (7.1), ce qui conclut la démonstration, par unicité de la solution de (7.1). □

### 7.3.2 Résolvante et amortissement

La reformulation (7.4) de l'équation (7.1) traduit des propriétés de stabilité éventuelles de l'opérateur intégrodifférentiel considéré.

Considérons en effet la solution $u$ de (7.1) où $C$ est une noyau régulier quelconque. Grâce à (7.4), on déduit que la fonction $u$ satisfait l'estimation suivante :

$$|\widehat{u}(t,k)| \leq |\mathfrak{R}(|k|t)|\,|\widehat{u}(0,k)| + |k|^{-1}\left(\int_0^{|k|t} \left|\mathfrak{R}(t')\right| \mathrm{d}t'\right) \left\|\widehat{f}(\cdot,k)\right\|_{\mathrm{L}^\infty(\mathbb{R}_+)}. \tag{7.9}$$

Ainsi, si

$$\mathfrak{R}(\tau) \underset{\tau\to+\infty}{\longrightarrow} 0, \tag{7.10}$$

et si $\widehat{f}(t,k)$ et $\widehat{u}(0,k)$ sont bornées uniformément, alors, pour tout $t \in \mathbb{R}_+$ fixé,

$$|\widehat{u}(t,k)| \underset{|k|\to+\infty}{\longrightarrow} 0.$$

C'est à dire que l'équation (7.1) est régularisante (elle amortit les hauts modes de Fourier).

La condition (7.10) a une interprétation naturelle en accord avec l'intuition physique : la contribution d'un temps passé au temps présent est d'autant moindre que ce temps passé est lointain. Cela retranscrit une propriété d'*oubli*. Dans le cas de l'équation de Peierls-Nabarro Dynamique, la lenteur de la convergence vers 0 des résolvantes (voir Proposition 7.3.2 ci-dessous) fait que cette propriété d'oubli est très atténuée par rapport au cas d'une résolvante à convergence exponentielle.

Cette propriété d'amortissement est fondamentale d'un point de vue théorique, mais aussi très importante du point de vue de la simulation numérique. Ainsi, si la résolvante tend vers 0 en $+\infty$, on pourra éventuellement utiliser des méthodes de quadrature en temps sur (7.4) et conserver certaines propriétés de stabilité. A contrario, si l'on opte pour une discrétisation explicite en temps sur (7.1), cette stabilité est vraisemblablement conditionnelle en le pas de temps $\Delta t$ (dépendant du mode de Fourier $k$ maximal considéré). Des tests numériques ultérieurs vont étayer cette affirmation.

Pour comprendre le phénomène, faisons une petite digression sur l'exemple simple d'une équation de diffusion où intervient le laplacien fractionnaire $(-\Delta)^{1/2}$. Pour cela, on fixe $C = \delta_0$ dans (7.1), qui s'écrit alors

$$\begin{cases} \partial_t \widehat{u}(t,k) = -|k|\widehat{u}(t,k) + \widehat{f}(t,k), \\ \widehat{u}(0,\cdot) = \widehat{u}_0, \end{cases} \tag{7.11}$$

et dont la résolvante est

$$\mathfrak{R}(\tau) := \mathrm{e}^{-\tau}.$$

Dans ce cas, la formule de Duhamel (7.4) s'écrit

$$\widehat{u}(t,k) = \mathrm{e}^{-|k|t}\widehat{u}(0,k) + \int_0^t \mathrm{e}^{-|k|(t-t')}\,\widehat{f}(t',k)\mathrm{d}t'. \tag{7.12}$$

La formule (7.12) encode de façon visible l'effet d'amortissement de (7.11), puisque la résolvante $\mathfrak{R}(T)$ tend vers 0 exponentiellement vite lorsque $T \to +\infty$. On se convainc aisément que la plupart des méthodes numériques appliquées à (7.12), même si elles sont explicites, respecteront les propriétés de stabilité du noyau diffusif. Au contraire, il est bien connu qu'une discrétisation explicite en temps de (7.11) nécessite des conditions de type CFL sur le pas de temps $\Delta t$ pour être stable.

### 7.3.3 Etude de $\mathfrak{R}^\alpha_{\mathrm{III}}$

Dans cette section technique, on étudie quelques propriétés qualitatives de la résolvante $\mathfrak{R}^\alpha_{\mathrm{III}}$, en utilisant notamment la transformation de Laplace. Cette fonction $\mathfrak{R}^\alpha_{\mathrm{III}}$ (et ses analogues $\mathfrak{R}^\alpha_{\mathrm{I}}$ et $\mathfrak{R}^\alpha_{\mathrm{II}}$) sera utilisée dans certaines méthodes numériques du Chapitre 8. C'est pourquoi nous insistons aussi sur les formules et les représentations que l'on peut en avoir et notamment sa transformée de Laplace, dont les méthodes numériques des Sections 8.5.6 et 8.5.7 font usage.

Grâce aux formules (6.9), (6.10), (6.14) et (7.8), on déduit l'expression de la transformée de Laplace de la résolvante relative au mode III que

$$\mathcal{L}\mathfrak{R}^\alpha_{\mathrm{III}}(p) = \frac{1+\alpha}{\alpha p + \sqrt{1+p^2}}. \tag{7.13}$$

Nous ne savons pas exprimer $\mathfrak{R}^\alpha_{\mathrm{III}}(T)$ à partir de fonctions analytiques, sauf dans deux cas particuliers : $\alpha = 0$ et $\alpha = 1$. En effet, grâce à des tables de transformées de Laplace usuelles [70, Id. 103 et Id. 105 p. 1116], on obtient

$$\mathfrak{R}^0_{\mathrm{III}}(T) = J_0(T), \tag{7.14}$$

$$\mathfrak{R}^1_{\mathrm{III}}(T) = 2\frac{J_1(T)}{T}. \tag{7.15}$$

En général, il est possible d'approximer $\mathfrak{R}^\alpha_{\mathrm{III}}$ pour des valeurs de $\alpha \geq 0$ quelconques via la simulation numérique de (7.7) (ou par inversion numérique de la transformation de Laplace). Comme on peut l'observer sur la Figure 7.1, la fonction $\mathfrak{R}^\alpha_{\mathrm{III}}(T)$ tend vers 0 en oscillant, la période caractéristique de ces oscillations étant d'ordre 1 (voir Proposition 7.3.2 ci-dessous). On constate aussi que plus $\alpha$ est grand, plus $\mathfrak{R}^\alpha_{\mathrm{III}}$ tend vite vers 0.

L'expression (7.13) permet de recouvrer la fonction $\mathfrak{R}^\alpha_{\mathrm{III}}$ grâce à la transformation de Laplace inverse

$$\mathfrak{R}^\alpha_{\mathrm{III}}(T) = \frac{1}{2\mathrm{i}\pi}\int_\Gamma \mathcal{L}\mathfrak{R}^\alpha_{\mathrm{III}}(\lambda)\mathrm{e}^{\lambda T}\mathrm{d}\lambda, \tag{7.16}$$

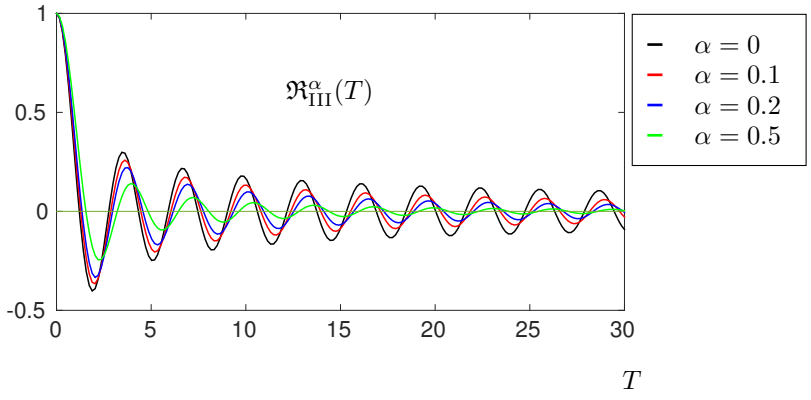



FIGURE 7.1 – Tracé de $\mathfrak{R}^{\alpha}_{\mathrm{III}}(T)$ pour $\alpha \in \{0, 0.1, 0.2, 0.5\}$.

où $\Gamma$ est un contour du plan complexe qui passe à droite des coupures et des singularités de $\mathcal{L}\mathfrak{R}^{\alpha}_{\mathrm{III}}$. Encore faut-il définir un prolongement analytique de $\mathcal{L}\mathfrak{R}^{\alpha}_{\mathrm{III}}(p)$, initialement définie pour $p \in \mathbb{R}_+$, à tout le plan complexe $\mathbb{C}$ hormis certaines lignes de coupure.

Pour ce faire, nous définissons un relèvement de la fonction $z \to \sqrt{1+z^2}$ en prolongeant analytiquement cette fonction définie sur $\mathbb{R}$ sur le plan complexe $\mathbb{C}$ privé des demi-droites $\mathrm{i} + \mathbb{R}^*_-$ et $-\mathrm{i} + \mathbb{R}^*_-$, ce qui donne un sens à l'expression (7.13) sur ce domaine. Les demi-droites $\mathrm{i} + \mathbb{R}^*_-$ et $-\mathrm{i} + \mathbb{R}^*_-$ sont appelées *lignes de coupure*. En utilisant ce relèvement, si $\alpha > 0$, la fonction $\mathfrak{R}^{\alpha}_{\mathrm{III}}(z)$ ne possède pas de singularité, mais est seulement discontinue le long des deux lignes de coupure (voir Figure 7.2). Dans le cas particulier où $\alpha = 0$, les extrémités des lignes de coupure, à savoir $\pm\mathrm{i}$, coïncident avec des singularités d'ordre $1/2$ de la fonction $\mathfrak{R}^{0}_{\mathrm{III}}$.

Les lignes de coupure peuvent être déformées continûment, induisant ainsi d'autres relèvements possibles pour $\sqrt{1+z^2}$ ; cependant les points de branchement $\pm\mathrm{i}$ sont nécessairement à l'extrémité d'une ligne de coupure. Par exemple, on peut proposer un autre relèvement de $\mathfrak{R}^{\alpha}_{\mathrm{III}}$ qui n'a qu'une seule ligne de coupure, à savoir $[-\mathrm{i}, \mathrm{i}]$ (voir la Figure 7.3). En pratique, le choix des lignes de coupure dépend souvent des contours $\Gamma$ que l'on peut alors tracer lorsqu'on utilise la transformation de Laplace inverse (7.16). D'un point de vue numérique, ce choix est lié à la vitesse de convergence d'une discrétisation de l'intégrale (7.16).

Nous justifions rigoureusement que $\mathfrak{R}^{\alpha}_{\mathrm{III}}$ tend bien vers 0 à l'infini par la :

**Proposition 7.3.2.** *Dans le cas où $\alpha = 0$, la fonction $\mathfrak{R}^{0}_{\mathrm{III}}$ est dominée de la façon suivante :*

$$\mathfrak{R}^{0}_{\mathrm{III}}(T) = \underset{T\to+\infty}{O}\left(T^{-1/2}\right), \tag{7.17}$$

*et, si $\alpha \in ]0, 1]$,*

$$\mathfrak{R}^{\alpha}_{\mathrm{III}}(T) = \underset{T\to+\infty}{O}\left(T^{-3/2}\right). \tag{7.18}$$

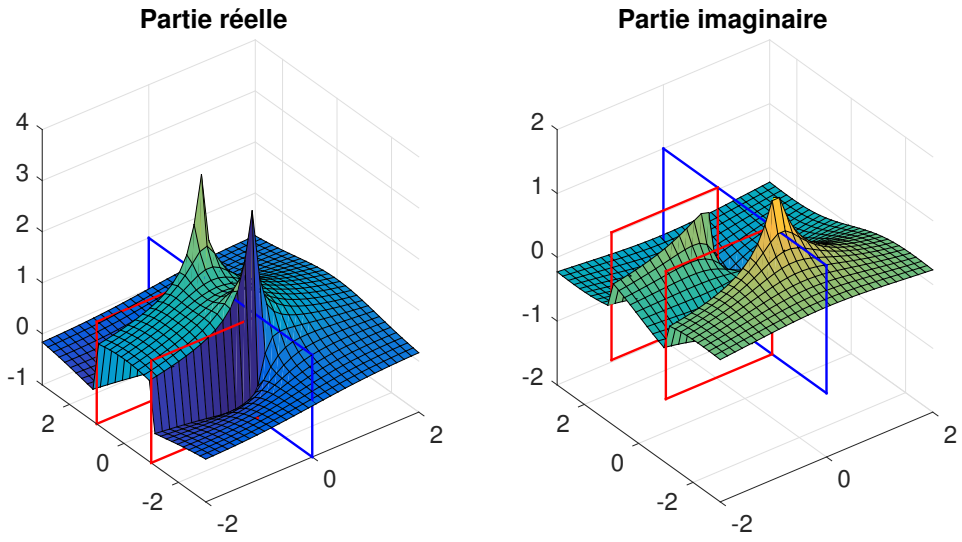



FIGURE 7.2 – Partie réelle et partie imaginaire de $\mathcal{L}\mathfrak{R}^{\alpha}_{\mathrm{III}}(z)$, où $z$ est la variable du plan complexe, et $\alpha = 0.2$. En bleu, le plan $\mathrm{Re}\,(z) = 0$, et en rouge, les plans correspondant aux lignes de coupure $\pm\mathrm{i} + \mathbb{R}^*_-$.

*Remarque* 52. La Proposition 7.3.2 semble suggérer que le paramètre visco-plastique $\alpha$ joue effectivement un rôle d'amortissement. En effet, plus $\alpha$ est proche de 0, plus $\mathfrak{R}^{\alpha}_{\mathrm{III}}$ décroît lentement en $+\infty$ (car, à $t$ fixé, $\mathfrak{R}^{\alpha}_{\mathrm{III}}(t)$ est continu en $\alpha \geq 0$).

Dans les deux cas $\alpha = 1$ et $\alpha = 0$, par les formules (7.14) et (7.15), les ordres de domination dans la Proposition 7.3.2 sont optimaux. En ce qui concerne le cas $\alpha > 0$, la preuve repose sur l'étude de la transformée de Laplace de $\mathfrak{R}^{\alpha}_{\mathrm{III}}$ (immédiatement déduite de celle de $C_{\mathrm{III}}$ via (7.8)). C'est un argument de géométrie harmonique tiré de [107, 137].

*Démonstration de la Proposition 7.3.2 dans le cas $\alpha > 0$.* Le relèvement $\mathfrak{R}^{\alpha}_{\mathrm{III}}$ défini plus haut ne présente pas de singularités et a pour coupures les demi-droites $\pm\mathrm{i} + \mathbb{R}^*_-$.

Grâce à [137, (1.4)–(1.6)] (voir aussi [107]), on sait que si une fonction $h$ est à transformée de Laplace sectorielle, c'est à dire qu'elle est telle que $\mathcal{L}h$ est analytique dans un domaine $z \in \{c + r\mathrm{e}^{\mathrm{i}\theta}, r > 0, \theta \in ]-\theta_0, \theta_0[\}$, pour $c \in \mathbb{R}$ et $\theta_0 > \pi/2$ fixés, et satisfait sur ce domaine

$$|\mathcal{L}h(z)| = O\left(|z|^{-\mu}\right), \tag{7.19}$$

alors $h$ vérifie

$$|h(T)| = \underset{T\to+\infty}{O}\left(e^{cT}T^{\mu-1}\right).$$

La fonction $\mathfrak{R}^{\alpha}_{\mathrm{III}}$ est à transformée de Laplace sectorielle, mais il faut pour cela prendre $c > 0$ dans la définition ci-dessus, ce qui ne permet même pas de démontrer qu'elle tend vers 0

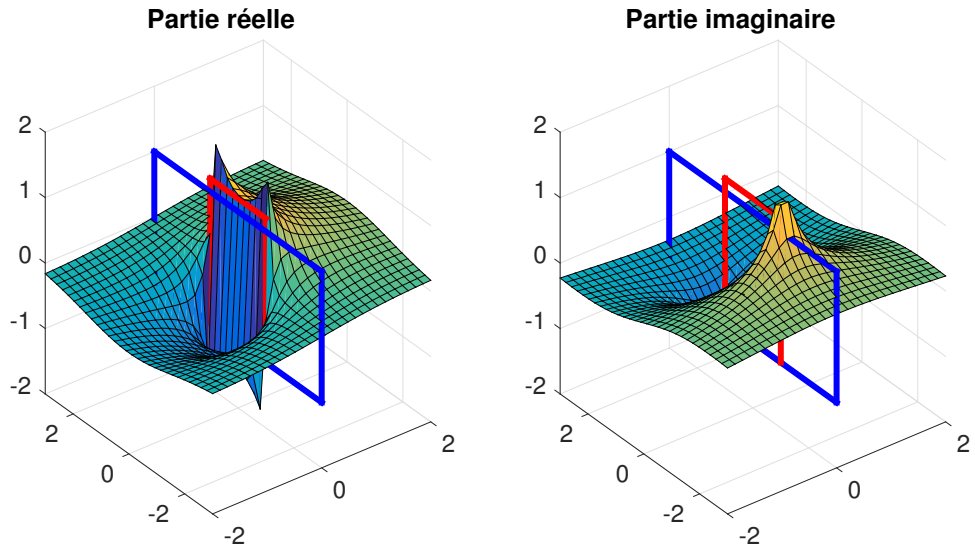



FIGURE 7.3 – Partie réelle et partie imaginaire de $\mathcal{L}\mathfrak{R}^{\alpha}_{\mathrm{III}}(z)$, où $z$ est la variable du plan complexe, et $\alpha = 0.2$. Le plan délimité par la ligne bleue est déterminé par $\mathrm{Re}\,(z) = 0$, et la ligne rouge correspond à la ligne de coupure $[-\mathrm{i}, \mathrm{i}]$.

en $+\infty$. Toutefois, à l'extrémité droite des lignes de coupure qui frappent l'axe des imaginaires purs en $z = \pm\mathrm{i}$, on observe que, par développement limité,

$$\mathfrak{R}^{\alpha}_{\mathrm{III}}(\pm\mathrm{i} + z) = \frac{1+\alpha}{\pm\alpha\mathrm{i}} + O\left(|z|^{1/2}\right). \tag{7.20}$$

En intégrant la transformation inverse de Laplace sur les contours $\Gamma_{\pm}$ définis sur la Figure 7.4, on obtient

$$\mathfrak{R}^{\alpha}_{\mathrm{III}}(T) = \frac{1}{2\mathrm{i}\pi}\left(\int_{\Gamma_+} \mathcal{L}\mathfrak{R}^{\alpha}_{\mathrm{III}}(z)\mathrm{e}^{zT}\mathrm{d}z + \int_{\Gamma_-} \mathcal{L}\mathfrak{R}^{\alpha}_{\mathrm{III}}(z)\mathrm{e}^{zT}\mathrm{d}z\right).$$

On sépare les intégrales ci-dessus en une zone proche des extrémités $z = \pm\mathrm{i}$ et une zone plus éloignée. Ainsi

$$\begin{aligned}\mathfrak{R}^{\alpha}_{\mathrm{III}}(T) =& \frac{1}{2\mathrm{i}\pi}\left(\int_{\Gamma_+\cap\{z,\mathrm{Re}\,(z)>-\varepsilon\}} + \int_{\Gamma_-\cap\{z,\mathrm{Re}\,(z)>-\varepsilon\}}\right)\mathcal{L}\mathfrak{R}^{\alpha}_{\mathrm{III}}(z)\mathrm{e}^{zT}\mathrm{d}z \\ &+ \frac{1}{2\mathrm{i}\pi}\left(\int_{\Gamma_+\cap\{z,\mathrm{Re}\,(z)<-\varepsilon\}} + \int_{\Gamma_-\cap\{z,\mathrm{Re}\,(z)<-\varepsilon\}}\right)\mathcal{L}\mathfrak{R}^{\alpha}_{\mathrm{III}}(z)\mathrm{e}^{zT}\mathrm{d}z.\end{aligned}$$

Sur la partie proche des extrémités, on utilise le développement limité (7.20), tandis qu'on

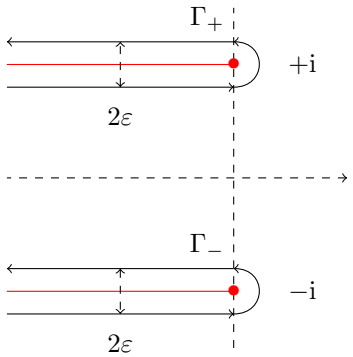



FIGURE 7.4 – En noir, le contour d'intégration pour calculer $\mathfrak{R}^\alpha_{\mathrm{III}}$ par transformation de Laplace inverse. En rouge, les lignes de coupure $\pm\mathrm{i}+\mathbb{R}_-$.

utilise le fait que $\mathcal{L}\mathfrak{R}^\alpha_{\mathrm{III}}$ est bornée ailleurs. Ainsi,

$$\begin{aligned}|\mathfrak{R}^\alpha_{\mathrm{III}}(T)| \leq & C\int_\varepsilon^{+\infty} \min\left(|z|^{1/2},1\right)\mathrm{e}^{-zT}\mathrm{d}z + C\varepsilon\varepsilon^{1/2}\mathrm{e}^{\varepsilon T}\\ \leq & C\left(T^{-3/2}\int_0^{+\infty} y^{1/2}\mathrm{e}^{-y}\mathrm{d}y + \varepsilon^{3/2}\mathrm{e}^{\varepsilon T}\right).\end{aligned}$$

En imposant $\varepsilon T = 1$, on déduit alors (7.18) de l'estimation précédente. □

### 7.3.4 Description de $\mathfrak{R}^\alpha_{\mathrm{I}}$ et $\mathfrak{R}^\alpha_{\mathrm{II}}$

Dans cette section technique, on décrit brièvement les résolvantes $\mathfrak{R}^\alpha_{\mathrm{I}}$ et $\mathfrak{R}^\alpha_{\mathrm{II}}$ (nous n'avons pas fait d'étude aussi systématique que pour $\mathfrak{R}^\alpha_{\mathrm{III}}$).

Grâce à la formule (7.8), on peut exprimer leur transformée de Laplace

$$\mathcal{L}\mathfrak{R}^\alpha_{\mathrm{I}}(p) = \frac{(\gamma+\alpha)p^2\sqrt{1+\gamma^{-2}p^2}}{\alpha p^3\sqrt{1+\gamma^{-2}p^2}+(p^2+2)^2-4\sqrt{1+p^2}\sqrt{1+\gamma^{-2}p^2}},$$

et

$$\mathcal{L}\mathfrak{R}^\alpha_{\mathrm{II}}(p) = \frac{(1+\alpha)p^2\sqrt{1+p^2}}{\alpha p^3\sqrt{1+p^2}+(p^2+2)^2-4\sqrt{1+p^2}\sqrt{1+\gamma^{-2}p^2}}.$$

Sur le même principe que dans la section précédente, on peut définir des coupures sur la partie du plan complexe (voir Figure 7.5) :

$$\mathbb{C}\backslash\left((\mathrm{i}+\mathbb{R}_-)\cup(\mathrm{i}+\mathbb{R}_-)\cup(-\gamma\mathrm{i}+\mathbb{R}_-)\cup(\gamma\mathrm{i}+\mathbb{R}_-)\right).$$

On constate numériquement que, si $\alpha > 0$, les pôles de $\mathcal{L}\mathfrak{R}^\alpha_{\mathrm{II}}$ sont à partie réelle négative.

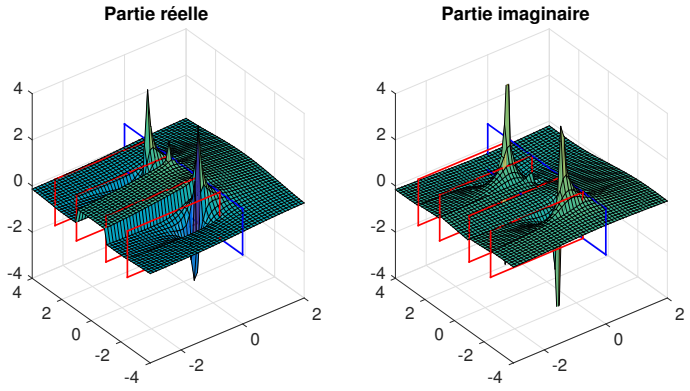



FIGURE 7.5 – Partie réelle et partie imaginaire de $\mathcal{L}\mathfrak{R}^{\alpha}_{\mathrm{II}}(z)$, où $z$ est la variable du plan complexe, pour $\alpha = 0.2$ et $\gamma = \sqrt{6}$. Le plan délimité par la ligne bleue est déterminé par $\mathrm{Re}\,(z) = 0$, et les plans rouges correspondent aux lignes de coupure $\pm\mathrm{i}+\mathbb{R}_-$ et $\pm\mathrm{i}\gamma+\mathbb{R}_-$.

Par exemple, pour $\alpha = 0.2$ et $\gamma = \sqrt{6} \simeq 2.4495$, on trouve les quatre pôles suivants

$$z_1^{\pm} \simeq -0.4193 \pm 1.8495\mathrm{i} \quad \text{et} \quad z_2^{\pm} \simeq -0.0043 \pm 0.9425\mathrm{i}.$$

Nous ne connaissons pas de valeur $\alpha$ telle qu'il soit possible d'exprimer $\mathfrak{R}^{\alpha}_{\mathrm{I}}$ et $\mathfrak{R}^{\alpha}_{\mathrm{II}}$ à partir de fonctions usuelles. Toutefois, il est possible de recourir à une approximation numérique (voir Figures 7.6) pour les évaluer. De même que pour le mode III, on constate que les résolvantes $\mathfrak{R}^{\alpha}_{\mathrm{I}}$ et $\mathfrak{R}^{\alpha}_{\mathrm{II}}$ oscillent sur une période caractéristique d'ordre 1, tout en tendant vers 0 (sauf, semble-t-il, si $\alpha = 0$). En outre, plus le paramètre $\alpha$ est grand, plus elles tendent vite vers 0 à l'infini (voir Figure 7.6).

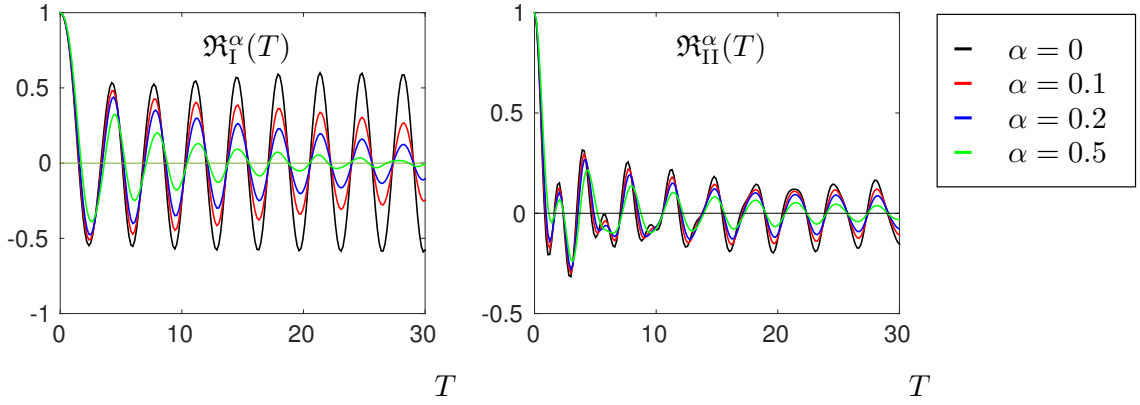



FIGURE 7.6 – Tracés de $\mathfrak{R}^{\alpha}_{\mathrm{I}}(T)$ et de $\mathfrak{R}^{\alpha}_{\mathrm{II}}(T)$, pour $\gamma = \sqrt{3} \simeq 1.73$ et $\alpha \in \{0, 0.1, 0.2, 0.5\}$.

## 7.4 Un théorème d'existence et d'unicité dans un cas non-linéaire

Afin de justifier le fait que l'équation *non-linéaire* (6.22) que nous nous proposons de résoudre est bien posée, nous démontrons le résultat très simple suivant :

**Théorème 7.4.1.** *Soient $\mathfrak{R}(t)$ une fonction mesurable bornée et $f(t,x,u)$ une fonction satisfaisant les estimations suivantes :*

$$\sup_{t\in\mathbb{R}_+} \|f(t,\cdot,0)\|_{\mathrm{L}^2(\mathbb{R})} = M_1 < +\infty, \tag{7.21}$$

$$\sup_{t\in\mathbb{R}_+ x,u\in\mathbb{R}} |\partial_u f(t,x,u)| = M_2 < +\infty. \tag{7.22}$$

*Soit $u_0 \in \mathrm{L}^2(\mathbb{R})$. Alors, il existe une unique fonction $u \in \mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+, \mathrm{L}^2(\mathbb{R})\right)$ solution du problème suivant :*

$$\widehat{u}(t,k) = \mathfrak{R}(|k|t)\widehat{u}_0(k) + \int_0^t \mathfrak{R}(|k|(t-t'))\mathcal{F}\left\{f(t',\cdot,u(t',\cdot))\right\}(k)\mathrm{d}t'. \tag{7.23}$$

*Si la fonction $\mathfrak{R}$ est continue, alors $u \in \mathrm{C}\left(\mathbb{R}_+, \mathrm{L}^2(\mathbb{R})\right)$.*

La preuve de ce théorème repose sur le principe de la preuve du Théorème de Cauchy-Lipschitz [36, Th. 7.3 p. 184], c'est à dire sur le classique Théorème du point-fixe de Banach [36, Th. 5.7 p. 138].

*Démonstration du Théorème 7.4.1.* La preuve se déroule en quatre parties. La première partie consiste à démontrer une estimation de contraction, dont découle les deux parties suivantes, qui concernent l'existence, puis l'unicité de $u$. La preuve d'existence se scinde elle-même en deux étapes. On démontre tout d'abord un résultat d'existence locale, puis on en déduit un résultat d'existence globale. Enfin, on démontre le résultat de continuité.

**Propriétés de contraction de l'intégrale** Soit $T > 0$. Pour $u \in \mathrm{L}^\infty\left([0,T], \mathrm{L}^2(\mathbb{R})\right)$ et $t \leq T$, on définit l'application $\Phi[u](t,x)$ de transformée de Fourier :

$$\widehat{\Phi}[u](t,k) := \int_0^t \mathfrak{R}(|k|(t-t'))\mathcal{F}\left\{f(t',\cdot,u(t',\cdot))\right\}(k)\mathrm{d}t'.$$

Alors, par le théorème de Plancherel, puis l'inégalité de Cauchy-Schwarz,

$$\begin{aligned}\|\Phi[u]\|^2_{\mathrm{L}^2(\mathbb{R})} =& 2\pi \int_{\mathbb{R}} \left| \int_0^t \mathfrak{R}(|k|(t-t'))\mathcal{F}\left\{f(t',\cdot,u(t',\cdot))\right\}(k)\mathrm{d}t' \right|^2 \mathrm{d}k \\ \leq& 2\pi \int_{\mathbb{R}} \left( \int_0^t \left|\mathfrak{R}(|k|(t-t'))\right|^2 \mathrm{d}t' \right) \\ & \left( \int_0^t \left|\mathcal{F}\left\{f(t',\cdot,u(t',\cdot))\right\}(k)\right|^2 \mathrm{d}t' \right) \mathrm{d}k \\ \leq& 2\pi t \left\|\mathfrak{R}\right\|^2_{\mathrm{L}^\infty(\mathbb{R}_+)} \int_0^t \int_{\mathbb{R}} \left|\mathcal{F}\left\{f(t',\cdot,u(t',\cdot))\right\}(k)\right|^2 \mathrm{d}k\mathrm{d}t' \\ \leq& Ct \left\|f(t,x,u(t,x))\right\|_{\mathrm{L}^1([0,T],\mathrm{L}^2(\mathbb{R}))}.\end{aligned}$$

On remarque que, par inégalité triangulaire, grâce à (7.21) et (7.22), pour tout $t > 0$

$$\left\|f(t,\cdot,u(t,\cdot)\right\|_{\mathrm{L}^2(\mathbb{R})} \leq M_1 + M_2 \left\|u(t,\cdot)\right\|_{\mathrm{L}^2(\mathbb{R})}. \tag{7.24}$$

Par conséquent, il existe une constante $C > 0$ ne dépendant que de $M_1$, $M_2$ et $\|\mathfrak{R}\|_{\mathrm{L}^\infty(\mathbb{R}_+)}$ telle que, pour tout $t \leq T$,

$$\left\|\Phi[u](t,\cdot)\right\|_{\mathrm{L}^2(\mathbb{R})} \leq C\sqrt{T}\left(1 + \|u\|_{\mathrm{L}^1([0,T],\mathrm{L}^2(\mathbb{R}))}\right). \tag{7.25}$$

En outre, si $u_1, u_2 \in \mathrm{L}^\infty\left([0,T],\mathrm{L}^2(\mathbb{R})\right)$, il existe une constante $C > 0$ ne dépendant que de $M_1$, $M_2$ et $\|\mathfrak{R}\|_{\mathrm{L}^\infty(\mathbb{R}_+)}$ telle que, pour tout $t \leq T$,

$$\left\|\Phi[u_1](t,\cdot) - \Phi[u_2](t,\cdot)\right\|_{\mathrm{L}^2(\mathbb{R})} \leq C\sqrt{T}\left\|u_1 - u_2\right\|_{\mathrm{L}^1([0,T],\mathrm{L}^2(\mathbb{R}))} \tag{7.26}$$

**Existence** Pour $u \in \mathrm{L}^\infty\left([0,T],\mathrm{L}^2(\mathbb{R})\right)$, on pose alors $\Psi[u](t,x)$ définie par

$$\widehat{\Psi}[u](t,k) = \mathfrak{R}(|k|t)\widehat{u}_0(k) + \widehat{\Phi}[u](t,k).$$

Résoudre (7.23), c'est trouver un point fixe à $\Psi$ dans $\mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+,\mathrm{L}^2(\mathbb{R})\right)$.

Par (7.26), l'application $\Psi$ est contractante sur l'ensemble $\mathrm{L}^\infty\left([0,T],\mathrm{L}^2(\mathbb{R})\right)$ –qui est un espace de Banach– à condition que $T$ soit suffisamment petit, et admet donc sur cet ensemble un unique point fixe par [36, Th. 5.7 p. 138] : on a ainsi établi un résultat d'existence local.

On prolonge ensuite la solution locale $u$ construite sur l'espace $\mathrm{L}^\infty\left([0,T],\mathrm{L}^2(\mathbb{R})\right)$. Pour ce faire, on définit pour $v \in \mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+,\mathrm{L}^2(\mathbb{R})\right)$ l'application $\widetilde{\Psi}$ par

$$\begin{aligned}\widetilde{\Psi}[v](t,x) =& \mathfrak{R}(|k|(T+t))\widehat{u}_0(k) + \int_0^T \mathfrak{R}(|k|(T+t-t'))\mathcal{F}\left\{f(t',\cdot,u(t',\cdot))\right\}(k)\mathrm{d}t' \\ & + \int_0^t \mathfrak{R}(|k|(t-t'))\mathcal{F}\left\{f(t',\cdot,v(t',\cdot))\right\}(k)\mathrm{d}t'\end{aligned}$$

De même que précédemment, on démontre que l'application $\widetilde{\Psi}$ est une contraction sur l'espace de Banach $\mathrm{L}^\infty\left([0,T],\mathrm{L}^2(\mathbb{R})\right)$, et y admet donc un unique point fixe. En prolongeant alors $u$ sur $[0,2T] \times \mathbb{R}$ par $u(t+T,\cdot) := v(t,\cdot)$, on a construit une solution de (7.23) sur $[0,2T] \times \mathbb{R}$. En procédant itérativement, on construit ainsi une solution globale $u \in \mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+,\mathrm{L}^2(\mathbb{R})\right)$.

**Unicité** Supposons que $u_1$ et $u_2 \in \mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+, \mathrm{L}^2(\mathbb{R})\right)$ soient deux solutions de (7.23). Alors, par (7.26), on a, pour tout $t \in \mathbb{R}$,

$$\|u_1 - u_2\|_{\mathrm{L}^\infty([0,t],\mathrm{L}^2(\mathbb{R}))} \leq C\sqrt{t}\,\|u_1 - u_2\|_{\mathrm{L}^1([0,t],\mathrm{L}^2(\mathbb{R}))}\,.$$

Introduisons $T$ le supremum des points $t$ tels que, en presque en tout $t' \in [0,t]$, $u_1(t',\cdot) = u_2(t',\cdot)$. Si $T < +\infty$, alors,

$$\begin{aligned}\|u_1 - u_2\|_{\mathrm{L}^\infty([0,T+\varepsilon],\mathrm{L}^2(\mathbb{R}))} \leq & C\sqrt{T+\varepsilon}\,\|u_1 - u_2\|_{\mathrm{L}^1([T,T+\varepsilon],\mathrm{L}^2(\mathbb{R}))} \\ \leq & C\varepsilon\sqrt{T+\varepsilon}\,\|u_1 - u_2\|_{\mathrm{L}^\infty([0,T+\varepsilon],\mathrm{L}^2(\mathbb{R}))}\,.\end{aligned}$$

Pour $\varepsilon > 0$ suffisamment petit, on obtient une contradiction, donc $T = +\infty$. Ainsi, $u_1$ et $u_2$ coïncident dans $\mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+, \mathrm{L}^2(\mathbb{R})\right)$.

**Continuité** La preuve de continuité se fait grâce au théorème de convergence dominée. □

Par la Proposition 7.3.2, si $\alpha \geq 0$, la fonction $\mathfrak{R}^\alpha_{\mathrm{III}}$ est bornée ; par ailleurs elle est continue. En outre, si $F \in \mathrm{C}^2_{\mathrm{per}}(\mathbb{R})$, et si la différence de chargements $\sigma^{\mathrm{a}}(t,x) - \sigma^{\mathrm{a}}_0$ et la différence des parties élastiques associées $\eta_{\mathrm{e}}(t,x) - \eta_{0,\mathrm{e}}$ sont dans $\mathrm{C}\left(\mathbb{R}_+, \mathrm{C_c}(\mathbb{R})\right)$, alors $f(t,x,u) = f_{\sigma^a}[u](t,x)$ (définie par (6.23)) satisfait (7.21) et (7.22). Dans ce cas on peut appliquer le Théorème 7.4.1 à la forme résolue associée à (6.22).

Le théorème que nous énonçons est minimaliste. Toutefois, avec des hypothèses adéquates sur les dérivées partielles de $f$, et sur la régularité de $u_0$, on peut vraisemblablement remplacer l'espace $\mathrm{L}^2(\mathbb{R})$ par un espace $\mathrm{H}^s(\mathbb{R})$ en ce qui concerne la variable spatiale $x$. Cependant, la démonstration ci-dessus utilise le fait que la transformée de Fourier est une isométrie à un facteur multiplicatif près de $\mathrm{L}^2(\mathbb{R})$ sur $\mathrm{L}^2(\mathbb{R})$. Obtenir des résultats dans d'autres espaces fonctionnels que les espaces $\mathrm{H}^s(\mathbb{R})$ (comme par exemple $\mathrm{L}^\infty(\mathbb{R})$ ou $\mathrm{C}(\mathbb{R})$) semble donc complexe. Par ailleurs, on peut légitimement se poser la question de la régularité en temps des solutions. A cet égard, observons que l'on n'a utilisé que le fait que la résolvante est bornée. Si on suppose en outre qu'elle tend vers 0 à une certaine vitesse, on peut en outre s'attendre à ce que la solution $u$ deviennent de plus en plus régulière.

Les hypothèses du Théorème 7.4.1 sont parfois trop restrictives. Nous réalisons notamment des expériences numériques qui sortent du cadre de l'Hypothèse (7.21). Ainsi, on peut imposer un chargement brusque sur une dislocation statique, c'est à dire $\sigma^{\mathrm{a}}(t > 0, x) = \sigma$, pour $f(t,x,u) = f_{\sigma^a}[u](t,x)$ définie par (6.23) : cela implique que $f(t,\cdot,0) \notin \mathrm{L}^2(\mathbb{R})$. Dans ce cas, on ne peut pas appliquer le Théorème 7.4.1 à (6.22).

Etant davantage motivés par la simulation numérique et les résultats que l'on peut en tirer, nous laissons de côté ces questions intéressantes.

## 7.5 Une remarque sur les dérivées fractionnaires

Certains auteurs [44, 124] se sont intéressés à des équations à dérivées fractionnaires à la fois temporelles et spatiales, c'est à dire

$$\partial_t^\nu u(t,x) = Lu(t,x), \tag{7.27}$$

où $L$ est un opérateur de dérivée fractionnaire spatiale, par exemple $L = (-\Delta)^{\beta/2}$, pour $\beta \in ]0, 2]$ qui est un opérateur de symbole $|k|^\beta$ en Fourier. La dérivée fractionnaire temporelle $\partial_t^\nu$ est la "dérivée de Caputo à gauche" (voir [124]), définie par

$$\partial_t^\nu u(t,x) := \frac{1}{\Gamma(1-\nu)} \int_0^t \frac{1}{(t-t')^\nu} \partial_t u(t',x) \mathrm{d}t', \tag{7.28}$$

pour $\nu \in ]0, 1[$. Grâce aux transformation de Laplace et de Fourier, comme les opérateurs $\partial_t^\nu$ et $L$ possèdent des structures de convolution respectivement temporelle et spatiale, on diagonalise l'équation (7.27), qui s'écrit alors

$$p^{\nu-1} \left( p\mathcal{L}\widehat{u}(p,k) - \widehat{u}(0,k) \right) = |k|^\beta \mathcal{L}\widehat{u}(p,k),$$

car

$$\mathcal{L}\left\{ t^{-\nu} \right\}(p) = \Gamma(1-\nu) p^{\nu-1}.$$

En prenant d'autre définitions de $L$ et $\partial_t^\nu$, de telles manipulations amènent à réécrire (7.27) comme

$$a(p)\mathcal{L}\widehat{u}(p,k) + \sum_{j=1}^{d} a_j(p) \left(\partial_t\right)^j \widehat{u}(0,k) = b(k)\mathcal{L}\widehat{u}(p,k), \tag{7.29}$$

où $a$, $b$ et $c$ sont des fonctions correspondant aux opérateurs.

Or, la solution homogène de (7.1) ne peut satisfaire (7.29). Par conséquent, l'équation de Peierls-Nabarro Dynamique ne peut s'écrire simplement à l'aide d'un nombre fini d'opérateurs fractionnaires. Dans le cas contraire, cela aurait permis d'utiliser des techniques de simulation s'appliquant à de telles équations. Par ailleurs, cela souligne le fait que l'équation de Peierls-Nabarro Dynamique est d'une nature très différente de l'équation dynamique construite au Chapitre 4. Ces deux équations n'ont en commun que leurs états stationnaires.

# Chapitre 8

# Résolution numérique de l'équation de Peierls-Nabarro Dynamique

Dans ce chapitre, nous comparons différentes approches numériques pour simuler l'équation de Peierls-Nabarro Dynamique. Ces comparaisons sont à la fois théoriques et fondées sur des tests numériques.

Ce travail a été effectué en collaboration avec Claude Le Bris, Frédéric Legoll et Yves-Patrick Pellegrini.

Le lecteur trouvera en Annexe 8 des précisions techniques sur les algorithmes implémentés.

## 8.1 Introduction

Le but de ce Chapitre est de proposer des algorithmes pour simuler l'équation de Peierls-Nabarro Dynamique (6.22) dont nous rappelons ici l'expression :

$$\begin{cases} \kappa_i^\alpha \partial_t \widehat{u}(t,k) + k^2 \displaystyle\int_0^t C_i(|k|(t-t'))\widehat{u}(t',k)\mathrm{d}t' = \mathcal{F}\left\{f_{\sigma^a}[u](t,\cdot)\right\}(k), \\ u(0,\cdot) = 0. \end{cases} \tag{8.1}$$

Pour ce faire, nous étudions divers schémas et diverses méthodes numériques, issues de la littérature, que nous testons ensuite sur plusieurs cas particuliers. Nous rappelons dans cette section plusieurs points déjà abordés dans l'introduction de la thèse.

### 8.1.1 Cadre du problème

Notre parti pris a été d'interpréter la transformation de Fourier continue comme une transformation de Fourier discrète, en réutilisant l'approche développée dans le Chapitre 5. Ainsi, la discrétisation spatiale de (8.1) fait intervenir des équations intégrodifférentielles du type (8.2) ci-dessous

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}u(t) = -\displaystyle\int_0^t C(t-t')u(t')\mathrm{d}t' + f(t), \\ u(0) = 0, \end{cases} \tag{8.2a}$$

dont la solution est donnée par la formule de Duhamel

$$u(t) = \int_0^t \mathfrak{R}(t-t')f(t')\mathrm{d}t', \tag{8.2b}$$

où $\mathfrak{R}$ est la résolvante associée à (8.2a).

Dans notre cadre d'application, la fonction $C$ de (8.2a) est la fonctions $(\kappa_i^\alpha)^{-1}C_i$ et la fonction $\mathfrak{R}$ de (8.2b) est la fonctions $\mathfrak{R}_i^\alpha$ (on se ramène au cas où $k = 1$), pour $i \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$ et $\alpha > 0$ (voir l'Annexe A.5.1 rassemblant des formules sur $\kappa_i^\alpha$, $C_i$ et $\mathfrak{R}_i^\alpha$). Comme les fonctions $\mathfrak{R}_i^\alpha$ et $C_i^\alpha$ sont régulières, les équations (8.2a) et (8.2b) sont bien posées. La difficulté n'est pas tant la résolution numérique de ces équations que leur résolution numérique *efficace*, c'est à dire précise, stable et rapide.

### 8.1.2 Enjeux et difficultés

Nous choisissons de présenter les différents algorithmes de ce chapitre sur le modèle (8.2). En réalité, il faut résoudre un *système* d'équations du type de (8.2), couplées non-linéairement. Gardons donc à l'esprit que l'équation originale (8.1) présente les traits particuliers suivants :

1. la valeur de $f(t,x)$ dépend non-linéairement de $u(t,x)$ (voir Section 8.2),
2. les noyaux de convolution $C_i(T)$ et $\mathfrak{R}_i(T)$ tendent lentement vers 0 (en loi de puissance de $T$), et oscillent,

3. chaque mode de Fourier de $u(\cdot, k)$ évolue avec une échelle de temps dilatée par un facteur $|k|$.

Le premier point induit qu'il est nécessaire de calculer conjointement et successivement les valeurs de $u$ et de $f$ jusqu'au temps $t$, afin d'obtenir une approximation de $u(t)$ dans (8.2a). Ainsi, calculer $u(t)$ via (8.2b) n'est pas plus rapide qu'utiliser (8.2a) et ne peut se faire en une seule étape –ce qui serait le cas si $f(t' < t)$ était connue *a priori*.

Le second point est crucial. Il signifie qu'il faut conserver une mémoire précise du passé pour avancer d'un pas de temps. En outre, la dépendance en le passé est non-triviale ; en particulier, les intégrales de (8.2a) et (8.2b) encodent des annulations (dues aux oscillations du noyau) qu'il est délicat de reproduire avec une discrétisation grossière.

Le troisième point traduit une propriété de raideur de (8.1) (voir [79]), qui se manifeste par le préfacteur $|k|^2$ devant l'intégrale englobant la mémoire. Cette raideur induit la nécessité d'utiliser des schémas qui soient inconditionnellement *stables*, c'est à dire que le pas de temps $\Delta t$ n'est pas contraint par des conditions de stabilité de type CFL.

### 8.1.3 Schémas et méthodes de calcul

Nous insistons sur une trait singulier des équations intégrodifférentielles : le *schéma numérique* de l'équation n'est pas le seul élément dimensionnant d'une méthode d'intégration ; la *méthode de calcul* est aussi déterminante. Nous précisons ces deux notions : Un *schéma numérique* est une série d'équations algébriques satisfaites par la discrétisation $u_n$ d'une fonction $u(t_n)$ à approximer. Une *méthode de calcul* est l'ensemble des opérations algorithmiques grâce auxquelles on résout un schéma numérique donné. Cette résolution peut être éventuellement approchée. On désigne par le nom d'*algorithme* l'ensemble constitué d'un schéma numérique implémenté par une méthode de calcul fixée.

En ce qui concerne les équation différentielles ordinaires, ces deux notions sont généralement indépendantes. Par exemple, les méthodes d'intégration de Runge-Kutta reposent sur la simulation de systèmes dynamiques en temps discret qui s'écrivent formellement comme $\boldsymbol{u}^{n+1} = f(\Delta t, t_n, \boldsymbol{u}^n)$. Alors la *méthode de calcul* se résume à résoudre à chaque itération une équation éventuellement non-linéaire, laquelle se déduit du *schéma*. Par conséquent, la complexité de l'algorithme est de la forme $O(JN)$, où $N$ est le nombre total d'itérations, $J$ est le coût d'une itération élémentaire (où intervient donc la méthode de calcul, et la taille $m$ du vecteur $\boldsymbol{u}^n$ considéré).

Au contraire, pour les équations intégrales de Volterra, la *méthode de calcul* influe de façon non-linéaire sur la complexité temporelle de la méthode de l'algorithme, et aussi sur la quantité de mémoire nécessaire. En effet, il est possible d'accélérer le calcul d'un schéma fixé, ou d'utiliser moins de mémoire, en tirant parti de certaines structures algébriques.

Durant les 40 dernières années, la simulation numérique des équations du type de (8.2) a suscité un grand intérêt à cause de la diversité des phénomènes qu'elles peuvent modéliser, des conditions de bord transparentes aux fissures. Un effort particulier a été porté sur la rapidité des méthodes de calcul (voir [76, 112]). Par conséquent, au vu de la richesse de la littérature, nous n'avons pas pour objectif de proposer une nouvelle méthode de calcul plus efficace que celles déjà existantes. En revanche, nous avons comparé quelques algorithmes de

la littérature, et d'autres algorithmes, issus du couplage de différents schémas à différentes méthodes de calcul.

Nous présentons d'abord des schémas s'appliquant à des équations intégrodifférentielles sous des formes générales, et ensuite des méthodes qui permettent de les évaluer plus ou moins rapidement, de manière exacte ou approchée.

### 8.1.4 Schémas étudiés

Les équations (8.2a) et (8.2b), appartiennent à la grande classe des équations intégrodifférentielles de Volterra

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t) = -\int_0^t K(t,t')u(t')\mathrm{d}t' + f(t), \tag{8.3a}$$

et respectivement

$$u(t) = \int_0^t K(t,t')f(t')\mathrm{d}t', \tag{8.3b}$$

où $K$ est un noyau prenant deux arguments, et n'ayant pas nécessairement la structure convolutive $K(t,t') = C(t-t')$. Il existe de nombreux schémas d'intégration de (8.3a), parmi lesquels on peut citer les méthodes de Galerkine ou de collocation (voir [10, Chap. 3 p. 49]). Nous nous concentrons sur des schémas itératifs qui reposent en général sur une méthode quadrature pour calculer l'intégrale présente dans (8.3a) ou (8.3b).

Notre point de départ bibliographique se situe dans la littérature géophysique ; ainsi, nous avons considéré des schémas issus de [62]. Puis, nous avons choisi d'étudier en particulier des schémas "bloc-par-bloc" (recommandés par [134, p. 994]). Nous en détaillons deux versions : la première discrétise directement la forme (8.3a), tandis que la seconde discrétise (8.3b). Enfin, l'étude de méthodes dites *oublieuses*, introduites dans la section suivante, nous a amené à proposer un schéma de splitting entre la partie linéaire, et la partie non-linéaire de l'équation (8.1).

Un des critères déterminants dans le choix d'un schéma, outre sa précision, est sa stabibilité en temps long (cela correspond, dans notre cas, à sa capacité à gérer la raideur de l'équation (8.1)). Dans le cas d'équations différentielles ordinaires raides, le caractère implicite ou explicite d'un schéma donné est déterminant pour établir sa stabilité (voir [79]). En ce qui concerne les équations intégrodifférentielles, ce caractère implicite ou explicite du schéma est moins important (car le caractère implicite ou explicite du schéma n'a un impact que sur une petite portion de l'intégrale à approximer). En revanche, de manière heuristique, le choix de la forme à intégrer, (8.2a) ou (8.2b), semble crucial lorsqu'il s'agit de stabilité. En effet, la première forme est mal conditionnée à cause du facteur en $|k|^2$, alors que la seconde encode naturellement la stabilité de l'équation via la résolvante. Ainsi, un schéma reposant sur une méthode de quadrature de l'intégrale de (8.2b) est naturellement plus stable qu'un schéma discrétisant directement (8.2a). Cela semble lié au fait que la résolvante $\mathfrak{R}$ encode directement les propriétés de stabilité de (8.2a), lesquelles semblent difficiles à caractériser seulement à partir d'évaluations du noyau $C$. On étaye cet argument sur un exemple simple dans la Section 7.3.2 et par des tests numériques dans la Section 8.6.

### 8.1.5 Méthodes de calcul étudiées

En tirant parti de certaines structures des schémas, on peut accélérer l'exécution d'un schéma donné. Les structures que nous avons étudiées sont les suivantes :

1. la structure de convolution, qui permet de réorganiser plus efficacement les calculs de quadrature ;
2. la structure de noyau *dégénéré* (voir la Section 8.5.1), grâce à laquelle on peut transformer une équation intégrodifférentielle en équation différentielle ordinaire.

Nous montrons d'abord comment tirer parti de la structure de convolution grâce à la méthode d'*accélération* de [76]. Dès lors que le schéma numérique voulu présente lui-même une structure de convolution discrète, cette méthode de calcul permet de l'implémenter de manière plus rapide, en passant d'une complexité $O(N^2)$ pour une implémentation naïve à une complexité de $O\left(N(\log N)^2\right)$. Cette méthode de calcul est remarquable car elle est *exacte*.

Nous présentons ensuite la structure de noyau dégénéré, qui permet de transformer en équation différentielle ordinaire l'équation intégrale étudiée. Cette structure n'est pas naturellement présente dans (8.2a) et (8.2b), aussi faut-il se restreindre à la classe des noyaux qui sont à la fois convolutifs et dégénérés. Ensuite, il faut approximer le noyau convolutif originel par de tels noyaux. Nous étudions les décompositions suivantes :

1. en somme de polynômes de Laguerre pondérés (voir [44, 104]) ;
2. en somme d'exponentielles, qui ont été étudiées dans les articles [3, 75] de Hagstrom et dans les articles de la mouvance de Lubich *et al.*, *e.g.*, [15, 16, 111, 112, 137].

Ces méthodes, au prix d'une erreur d'approximation sur le noyau, permettent de réduire la mémoire nécessaire, tout en demeurant relativement rapides.

### 8.1.6 Tests numériques

Nous avons implémenté 6 algorithmes différents, en utilisant les schémas et les méthodes numériques citées ci-dessus. Nous les avons alors soumis à différents tests, en évaluant :

— tout d'abord, leur capacité à résoudre une seule équation (8.2) scalaire, pour les noyaux intervenant dans l'équation de Peierls-Nabarro Dynamique ;
— puis, leur capacité à résoudre l'équation complète (8.1) dans certains cas.

Nous avons été particulièrement sensibles aux critères suivants : la stabilité, la précision, la rapidité d'exécution. En revanche, nous n'avons pas menés de tests systématiques sur la quantité de mémoire machine requise par chacun des algorithme. En testant les algorithmes sur l'équation complète, nous avons aussi évalué la dépendance en les paramètres de discrétisation spatiale du schéma proposé.

### 8.1.7 Plan

Ce chapitre s'articule en 6 Sections. La Section 8.2 détaille la discrétisation spatiale employée. La Section 8.3 introduit les trois schémas issus de la littérature que sont le schéma de [97], et deux schémas bloc-par-bloc. La Section 8.4 expose la méthode d'accélération de

[76]. La Section 8.5 étudie la structure de noyau dégénéré. Elle décrit ensuite deux méthodes oublieuses basées sur des noyaux dégénérés. La première, est issue de [44,104], et utilise une décomposition en polynômes de Laguerre pondérés. La seconde est tirée de [112]. Dans la Section 8.6, on propose des combinaisons d'algorithmes et de méthodes, et on effectue des tests numériques afin d'évaluer les différents algorithmes. Enfin, dans la Section 8.7, on conclut cette étude en proposant deux meilleures méthodes parmi les méthodes proposées. La première, nommée ici BBD-A, combine un schéma bloc-par-bloc appliqué à l'équation (8.2a) (voir Section 8.3.2) avec une méthode de calcul accélérée (voir Section 8.4). La seconde, nommée ici BBR-O, combine un schéma bloc-par-bloc appliqué à l'équation (8.2b) (voir Section 8.3.3) avec une méthode de calcul oublieuse (voir Section 8.5.7).

## 8.2 Détails de la discrétisation spatiale

Nous avons calqué la discrétisation spatiale de (8.1) sur celle que nous avons proposée pour l'équation de Weertman, au Chapitre 5.

On se fixe un nombre $2m = 2^P$ de points de discrétisation dont les indices sont dans l'ensemble $\mathcal{K}_{2m} = \{-m, \cdots, m-1\}$. On introduit alors les abscisses des points de discrétisation, qui sont équiréparties dans $[-L, L]$ :

$$x_j = jh, \quad \text{pour} \quad j \in \mathcal{K}_{2m},$$

où $h = L/m$. A cette grille spatiale correspond la grille duale de Fourier

$$k_p = \frac{2\pi}{2mh} p, \quad \text{pour} \quad p \in \mathcal{K}_{2m}.$$

On associe la discrétisation spatiale suivante à une fonction $u$ définie sur $\mathbb{R}$ :

$$u_j = u(x_j) \quad \text{pour} \quad j \in \mathcal{K}_{2m}.$$

A sa transformée de Fourier $\widehat{u}$, on associe la transformée de Fourier discrète $\mathcal{F}_d$ du vecteur $\boldsymbol{u} = (u_j)_{j \in \mathcal{K}_{2m}}$, c'est à dire

$$\widehat{u}(k_p) \simeq \left(\mathcal{F}_d\{\boldsymbol{u}\}\right)_p := \sum_{j=-m}^{m-1} u_j \mathrm{e}^{-\mathrm{i}x_j k_p}, \quad \text{pour} \quad p \in \mathcal{K}_{2m}.$$

Ainsi, on transforme (8.1) en l'équation semi-discrétisée suivante :

$$\kappa_i^\alpha \frac{\mathrm{d}}{\mathrm{d}t} \left(\mathcal{F}_d\{\boldsymbol{u}(t)\}\right)_p + k_p^2 \int_0^t C_i(|k_p|(t-t')) \left(\mathcal{F}_d\{\boldsymbol{u}(t')\}\right)_p \mathrm{d}t' = \left(\mathcal{F}_d\left\{\boldsymbol{f}(t)\right\}\right)_p, \tag{8.4}$$

pour tout $p \in \mathcal{K}_{2m}$, où

$$f_j(t) = f_{\sigma^a}[u](t, x_j).$$

On emploie les différentes méthodes d'intégration temporelle développées dans les Sections ultérieures sur l'équation (8.4).

Cette discrétisation préserve la structure de l'équation, car elle transforme une opération de convolution (continue) en espace, par exemple

$$\mathcal{O}_{\mathrm{III}}\{u\}(t,x) = -\int_{\mathbb{R}} \mathcal{F}^{-1}\left\{k^2 C_i(|k|(t-t'))\right\}(x_l - x')u(t',x')\mathrm{d}x'$$

en une opération de convolution discrète, qui s'écrit ici comme

$$h\sum_{j=-m}^{m-1}\left(\mathcal{F}_d^{-1}\left\{k_p^2 C_i(|k_p|(t-t'))\right\}\right)_j u_{l-j}(t').$$

Au noyau de convolution spatio-temporel $\mathcal{F}^{-1}\left\{k^2 C_i(|k|t)\right\}(x)$ correspond naturellement un noyau de convolution discrétisé en espace $\mathcal{F}_d^{-1}\left\{k_p^2 C_i(|k_p|(t-t'))\right\}_j$. Par ailleurs, la simulation de cette équation semi-discrétisée est facilitée par le fait que le terme de mémoire est découplé selon de modes de Fourier discret.

Mais une telle discrétisation périodise artificiellement l'équation (8.1), laquelle est originellement posée sur toute la droite réelle $\mathbb{R}$. On observe un phénomène de "réplications" qui se manifeste par des artefacts numériques au bord du domaine de simulation –lequel n'est pas de taille infinie (voir Chapitre 5). Il semble possible de recourir à des techniques de zero-padding pour atténuer cet effet (voir [134, Chap. 12 p. 624]). On pourrait aussi utiliser la technique de [48]. Nous n'explorons pas ces possibilités dans ce document.

*Remarque* 53 (Parallélisation). Une telle discrétisation spatiale est naturellement parallélisable, en distribuant la mémoire relative à chaque mode de Fourier $k$ à des processeurs différents.

*Remarque* 54 (Symétrie de la transformation de Fourier). La transformée de Fourier discrète d'un vecteur à coordonnées réelles est symétrique. En pratique, on tire parti de cette symétrie en n'effectuant les calculs numériques que sur les modes de Fourier $k_p$ pour $p \in \{-m, \cdots, 0\}$. Cela permet de diviser par 2 la mémoire utilisée et le temps de calcul.

## 8.3 Trois schémas

Dans cette section, nous décrivons trois schémas possibles pour simuler (8.2) :
— un schéma issu de [97] ;
— un schéma bloc-par-bloc issu de [105] appliqué à (8.2a) ;
— un schéma bloc-par-bloc issu de [105] appliqué à (8.2b).
Une méthode de calcul amènera par ailleurs un schéma de splitting, décrit dans la Section 8.5.5.

**Discrétisation temporelle** Nous fixons maintenant les notations utilisées dans toute la section. On souhaite résoudre (8.2a), ou son équivalent (8.2b), sur l'intervalle $[0, T]$. Par simplicité, on travaille avec une discrétisation temporelle à pas constant $\Delta t$ (même si chacun

des schémas présentés ci-dessous a été conçu pour un pas de temps variable). On définit ainsi

$$T = N\Delta t \quad \text{et} \quad t_n := n\Delta t.$$

Pour une fonction $g$, on désignera alors par $g_n$ une approximation de $g(t_n)$.

### 8.3.1 Un schéma d'ordre 2 de Lapusta et coauteurs

La première méthode que nous avons utilisée est l'algorithme de [62]. Nous décrivons ici les ingrédients d'une de ses variantes (c'est à dire le cœur de schéma de [97]), dont les versions les plus récentes sont, semble-t-il, toujours utilisées (voir la publication [126]). Le but originel de ce schéma était résoudre des problèmes proches du mode III de l'équation de Peierls-Nabarro Dynamique, en ce sens que seul le terme de forçage $f$ est différent.

Le schéma proposé dans [97, Sec. 5] utilise la forme directe (8.2a). Pour ce faire, les auteurs recourent à une formulation en position $u(t)$ et vitesse $v(t) = u'(t)$. On introduit une autre variable

$$z(t) := -\int_0^t C(t-t')u(t')\mathrm{d}t'.$$

Ainsi, (8.2a) se réécrit

$$v(t) = z(t) + f(t),$$

avec, par intégration par parties,

$$z(t) = -W(t)u(0) + W(0)u(t) + \int_0^t W(t-t')u(t')\mathrm{d}t',$$

où

$$W(t) = \int_t^{+\infty} C(t')\mathrm{d}t'.$$

On calcule alors itérativement $u_n$, $v_n$ et $z_n$ comme suit.

On fait tout d'abord une prédiction sur $u_{n+1}$ :

$$v_n := z_n + f_n, \tag{8.5a}$$

$$u_{n+1}^{[1]} := u_n + \Delta t v_n. \tag{8.5b}$$

Puis on évalue une première prédiction de $z_{n+1}$ grâce à une formule de points-milieux

$$z_{n+1}^{[1]} := -W(t_{n+1})u(0) + W(0)u_{n+1}^{[1]} + \Delta t W_{1/2}v_n + \Delta t \sum_{j=0}^{n-1} W_{n+1/2-j}v_{j+1/2}, \tag{8.5c}$$

avec

$$W_{j+1/2} := W\left(t_j + \frac{\Delta t}{2}\right) \tag{8.5d}$$

$$v_{j+1/2} := \frac{1}{2}\left(v_j + v_{j+1}\right), \tag{8.5e}$$

On en déduit une première prédiction de $v_{n+1}$ :

$$v_{n+1}^{[1]} := z_{n+1}^{[1]} + f_{n+1}. \tag{8.5f}$$

Puis, on calcule toutes les autres quantités à partir de ces prédictions, par une sorte de point-milieu :

$$v_{n+1/2} := \frac{1}{2}\left(v_n + v_{n+1}^{[1]}\right), \tag{8.5g}$$

$$u_{n+1} := u_n + \Delta t v_{n+1/2}, \tag{8.5h}$$

$$z_{n+1} := -W(t_{n+1})u(0) + W(0)u_{n+1} + \Delta t \sum_{j=0}^{n} W_{n+1/2-j} v_{j+1/2}, \tag{8.5i}$$

Le schéma ci-dessus est d'ordre 2 en $\Delta t$. Il semble qu'on puisse l'interpréter comme une première itération pour évaluer la solution du schéma implicite suivant :

$$v_{n+1/2} := \frac{1}{2}\left(v_n + v_{n+1}\right), \tag{8.6a}$$

$$u_{n+1} := u_n + \Delta t v_{n+1/2}, \tag{8.6b}$$

$$v_{n+1} := W(0)u_{n+1} - W(t_{n+1})u(0) + \Delta t \sum_{j=0}^{n} W_{n+1/2-j} v_{j+1/2} + f_{n+1}. \tag{8.6c}$$

*Remarque* 55 (Troncature des intégrales). Dans le cadre très général de (8.3a), l'absence d'information sur $K$ nécessite a priori de garder en mémoire tout le passé pour calculer les variations de $u$. Or, si l'on observe que

$$K(t,t') \underset{t-t'\to+\infty}{\longrightarrow} 0, \tag{8.7}$$

on peut éventuellement tronquer l'intégrale de (8.3a) par

$$\int_0^t K(t,t')u(t')\mathrm{d}t' \simeq \int_{\max(t-t_{\max},0)}^t K(t,t')u(t')\mathrm{d}t'$$

où $t_{\max}$ est un temps de troncature. L'erreur de troncature est alors d'autant plus grande que la convergence (8.7) est lente. Cette manière de faire est proposée dans [97] pour le noyau $C(t) = C_{\mathrm{III}}(t)$ qui décroît comme $t^{-3/2}$. Elle est avantageuse lors de la simulation de (8.1), car, si l'on emploie le même temps de référence $t_{\max}$ pour tous les modes $k$ de Fourier, on a besoin d'autant moins de mémoire que la longueur d'onde $1/|k|$ est petite.

*Remarque* 56. A la lecture de [97,125,126], il semble que les méthodes de calcul décrites dans les Sections 8.4 et 8.5.7 ne soient pas encore utilisées dans la communauté des chercheurs en géophysique.

### 8.3.2 Le schéma bloc-par-bloc appliquée à la forme directe

Les schémas bloc-par-bloc sont des schémas multi-pas s'appliquant à des équations intégrales ; elles peuvent être vues comme une généralisation des schémas de Runge-Kutta implicites. Nous suivons ici la présentation [105, p. 186], pour une méthode bloc-par-bloc utilisant la règle de quadrature de Simpson qui est précise à l'ordre 4 en temps.

On réécrit (8.3a) sous une forme intégrale, comme

$$\begin{cases} u(t_{n+p}) = u(t_n) + \displaystyle\int_{t_n}^{t_{n+p}} (z(t) + f(t))\,\mathrm{d}t, \\ z(t) = -\displaystyle\int_0^t C(t-t')u(t')\mathrm{d}t'. \end{cases} \tag{8.8}$$

Alors, en employant la règle de quadrature de Simpson sur (8.8) pour $p \in \{1, 2\}$, on obtient

$$u_{2n+1} = u_{2n} + \frac{\Delta t}{6}\left(z_{2n} + 4z_{2n+1/2} + z_{2n+1}\right) + \frac{\Delta t}{6}\left(f_{2n} + 4f_{2n+1/2} + f_{2n+1}\right), \tag{8.9a}$$

$$u_{2n+2} = u_{2n} + \frac{\Delta t}{3}\left(z_{2n} + 4z_{2n+1} + z_{2n+2}\right) + \frac{\Delta t}{3}\left(f_{2n} + 4f_{2n+1} + f_{2n+2}\right), \tag{8.9b}$$

et :

$$\begin{aligned} z_{2n+1} = &- \frac{\Delta t}{3}\sum_{j=0}^{2n} w_j^{2n} C\left((2n+1-j)\Delta t\right) u_j \\ &- \frac{\Delta t}{6}\left(C\left(\Delta t\right) u_{2n} + 4C\left(\Delta t/2\right) u_{2n+1/2} + C\left(0\right) u_{2n+1}\right), \end{aligned} \tag{8.9c}$$

$$z_{2n+2} = - \frac{\Delta t}{3}\sum_{j=0}^{2n+2} w_j^{2n+2} C\left((2n+2-j)\Delta t\right) u_j. \tag{8.9d}$$

Ici, les $w_j^{2n}$ sont les $2n+1$ poids donnés par $(1, 4, 2, 4, 2, ..., 2, 4, 1)$. Cependant, on se rend compte qu'il est nécessaire de définir des valeurs intermédiaires $u_{2n+1/2}$ et $z_{2n+1/2}$ (notons qu'on a $f_{2n+1/2} = f(t_{2n+1/2})$). Pour ce faire, on recourt à l'interpolation quadratique suivante :

$$u_{2n+1/2} = \frac{3}{8}u_{2n} + \frac{3}{4}u_{2n+1} - \frac{1}{8}u_{2n+2}, \tag{8.9e}$$

$$z_{2n+1/2} = \frac{3}{8}z_{2n} + \frac{3}{4}z_{2n+1} - \frac{1}{8}z_{2n+2}. \tag{8.9f}$$

Les valeurs

$$U_n := \left(u_{2n+1/2}; u_{2n+1}; u_{2n+2}; z_{2n+1/2}; z_{2n+1}; z_{2n+2}\right)$$

constituent le "bloc". Le système (8.9) se réécrit comme

$$(1 - \Delta t M_{\Delta t}) \cdot U_n = V_n + \Delta t W_n, \tag{8.10}$$

(voir l'Annexe A.7.1). Dans le schéma ci-dessus, $M_{\Delta t}$ est une matrice carrée $6 \times 6$ construite à partir de l'évaluation de $C$ pour les premiers temps 0, $\Delta t/2$, et $\Delta t$, le vecteur $V_n$ dépend linéairement de $u_{m \leq 2n}$ et de $z_{2n}$, et le vecteur $W_n$ est un forçage construit à partir de $f$. En pratique, $f$ dépend de $u$ ; par conséquent $W_n$ dépend de $U_n$ et on fait des itérations de point fixe pour résoudre l'équation (non-linéaire) implicite (8.10) (voir (A.49) en Annexe A.7.1). Comme on peut le constater, le schéma (8.9) joue sur les pas de temps pairs et impairs qui ont des rôles différents. Mais il ne nécessite pas de calculer des valeurs initiales, comme cela peut être le cas pour d'autres méthodes multi-pas.

### 8.3.3 La méthode bloc-par-bloc appliquée à la forme résolue

Nous avons présenté dans la Section 8.3.2 la méthode bloc-par-bloc pour résoudre l'équation intégrodifférentielle (8.2a). Il existe aussi une méthode bloc-par-bloc pour résoudre l'équation intégrale (8.2b) (voir [105, p. 114]). Celle-ci repose sur une quadrature de l'intégrale dans (8.2b) par la méthode de Simpson.

Plus précisément, on résout :

$$\begin{aligned} u_{2n+1} =& \frac{\Delta t}{3} \sum_{j=0}^{2n} w_j^{2n} \mathfrak{R}((2n+1-j)\Delta t) f_j \\ &+ \frac{\Delta t}{6} \left( \mathfrak{R}(\Delta t) f_{2n} + 4\mathfrak{R}\left(\frac{\Delta t}{2}\right) f_{2n+1/2} + \mathfrak{R}(0) f_{2n+1} \right), \end{aligned} \tag{8.11a}$$

$$u_{2n+2} = \frac{\Delta t}{3} \sum_{j=0}^{2n+2} w_j^{2n+2} \mathfrak{R}((2n+2-j)\Delta t) f_j, \tag{8.11b}$$

où interviennent les poids de Simpson $w_j^{2n}$, c'est à dire $(1, 4, 2, 4, 2, \cdots, 2, 4, 1)$. Comme $f(t)$ dépend en réalité de $u(t)$, il est nécessaire de définir $u_{2n+1/2}$, ce qui est fait grâce à une interpolation quadratique :

$$u_{2n+1/2} = \frac{3}{8} u_{2n} + \frac{3}{4} u_{2n+1} - \frac{1}{8} u_{2n+2}. \tag{8.11c}$$

Comme le schéma (8.9), le schéma bloc-par-bloc défini par (8.11) est d'ordre 4 en $\Delta t$ (voir [105, p. 116]).

Le Schéma (8.11) présuppose que l'on peut représenter la résolvante $\mathfrak{R}(t)$. Mais, dans les cas qui nous intéressent, on ne sait pas en général exprimer cette fonction à l'aide de fonctions usuelles : il n'est pas simple d'évaluer la fonction $\mathfrak{R}(t)$. A priori, il faut donc l'approximer numériquement. Néanmoins, en pratique, on va utiliser le schéma (8.11) dans le cadre des méthodes reposant sur la transformation de l'équation intégrodifférentielle (8.2b) en une équation différentielle (voir les Sections 8.5.6 et 8.5.7). Dans ce cas, il n'est pas nécessaire d'évaluer la résolvante $\mathfrak{R}$ (mais on utilise l'expression analytique de sa tranformée de Laplace).

### 8.3.4 Complexité

Les schémas présentés ci-dessus sont de complexité comparable. En effet, lorsque le nombre $N$ de pas de temps est grand, la majeure partie du coût de calcul réside dans l'évaluation des sommes (à savoir (8.5c) et (8.5i), (8.9c) et (8.9d), (8.11a) et (8.11b)), et non dans la résolution d'éventuelles équations implicites non-linéaires. Selon [105, p. 124], cette considération est générique pour les schémas d'intégration des équations intégrodifférentielles reposant sur une méthode de quadrature. Par conséquent, il d'autant plus recommandé d'utiliser des schémas implicites –dès que ceux-ci sont plus stables.

Une implémentation naïve du calcul de des sommes ci-dessus consisterait à les évaluer indépendamment à chaque pas de temps (d'où une complexité $O(N^2)$). Une telle implémentation stockerait aussi toutes les valeurs $u_n$ (c'est à dire $O(N)$ valeurs). L'utilisation d'une *méthode de calcul* efficace permet de réduire ces coûts en temps et/ou en mémoire. Toutes les méthodes de calcul que nous présentons ci-après sont applicables aux schémas de cette Section (si ceux-ci discrétisent (8.2)).

## 8.4 Une méthode de calcul accélérée

Nous présentons ici le principe de la méthode de calcul proposée dans [76]. Celle-ci s'applique à des équations intégrodifférentielles avec une structure de convolution, c'est à dire (8.2). Nous désignons cette méthode, originellement qualifiée de *fast*, par le nom de méthode *accélérée*. En effet, elle est *purement algorithmique*, en ce sens qu'elle permet de calculer *plus rapidement* un schéma *donné*, dès lors que celui-ci nécessite d'évaluer une convolution discrète –typiquement issue d'une méthode de quadrature et dont le pas de temps est constant (cette restriction sur le pas de temps peut être limitante). Contrairement à d'autres méthodes de calcul "rapides" (telles que celles présentées dans [31,74], par exemple les méthodes multipolaires), aucune approximation n'est faite vis-à-vis du schéma initial : les résultats donnés par un schéma accéléré et par un schéma non accéléré sont rigoureusement les mêmes (aux erreurs machine près).

L'accélération de [76] est obtenue en exploitant la structure de convolution de (8.2a) grâce à des techniques de sommation rapide reposant sur l'utilisation de la Transformation de Fourier Rapide (ou Fast Fourier Transform, FFT). Elle permet de réduire le nombre d'opérations d'un schéma choisi a priori, de $O(N^2)$ pour une implémentation naïve à $O\left(N \log N\right)$ ou à $O\left(N(\log N)^2\right)$ suivant les cas. En revanche, ne changeant pas le schéma auquel elle s'applique, elle nécessite de conserver la même quantité de mémoire : c'est sa principale faiblesse. Sa force est d'utiliser à bon escient cette mémoire.

### 8.4.1 Principe de la méthode

Plus précisément, supposons que l'on cherche à évaluer les sommes suivantes :

$$z_n = \sum_{j=0}^{n-1} c_{n-j} u_j, \tag{8.12}$$

pour $n \in [\![0, N]\!]$, et $(c_j)$ et $(u_j)$ des suites *données*. Une telle somme apparaît par exemple dans (8.9c) (où, par souci de simplicité, on a remplacé $2n+1$ par $n$, $w_j^{2n} C(2n+1-j)$ par $c_{n-j}$, et où on ne prend pas en compte le second terme à droite de (8.9c)) ; rappelons que c'est précisément son évaluation qui concentre la majeure partie du coût numérique d'une méthode de résolution par quadrature.

Une procédure naïve consisterait à évaluer indépendamment, pour tout $n \in [\![0, N]\!]$, la somme $z_n$ via (8.12). Elle aurait une complexité de $O(N^2)$. Or, (8.12) n'est rien d'autre que le produit matrice-vecteur suivant :

$$\begin{pmatrix} z_0 \\ z_1 \\ \cdots \\ \cdots \\ z_N \end{pmatrix} = \begin{pmatrix} 0 & \cdot & \cdots & \cdots & 0 \\ c_1 & 0 & \cdots & \cdots & \cdots \\ c_2 & c_1 & 0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ c_N & c_{N-1} & \cdots & c_1 & 0 \end{pmatrix} \cdot \begin{pmatrix} u_0 \\ u_1 \\ \cdots \\ \cdots \\ u_N \end{pmatrix}. \tag{8.13}$$

La matrice ci-dessus est de Toeplitz, c'est à dire qu'elle satisfait la propriété

$$M_{i+k,j+k} = M_{i,j}$$

pour tous entiers les $i, j, k$ tels que la formule ci-dessus ait un sens. Donc, le calcul (8.13) être effectué en $O(N \log N)$ opérations, grâce à la FFT (voir [134, Sec. 13]). On calcule ainsi *simultanément* tous les $z_0, ..., z_N$ en une étape peu coûteuse.

### 8.4.2 Description de la méthode pour un second membre variable

Il n'est pas possible d'utiliser directement la technique ci-dessus pour un schéma discrétisant (8.1). En effet, dans ce cas, on ne connaît pas $u_{n+k}$, pour $k > 0$, *avant* d'avoir calculé $z_n$, mais *après*. En revanche, remarquons que le calcul de $z_{n+j}$, pour $j \in [\![1, n]\!]$ nécessite d'évaluer la somme partielle

$$z_j^{[n]} = \sum_{k=0}^{n} c_{n+j-k} u_k, \tag{8.14}$$

dans le sens où :

$$z_{n+j} = z_j^{[n]} + \sum_{k=n+1}^{n+j-1} c_{n+j-k} u_k.$$

La somme (8.14) est elle-même une convolution discrète ne faisant appel qu'aux termes $u_k$ qui appartiennent déjà au passé au moment où on souhaite calculer $z_{n+j}$. Ainsi, on peut calculer les $z_j^{[n]}$, pour $j \in [\![1, n]\!]$, à partir de la $n$-ème itération. Ce dernier calcul peut être effectué en $O(n \log n)$ opérations élémentaires grâce à la FFT. En construisant un pavage dyadique des temps comme illustré sur la Figure 8.1, on peut ainsi accélérer le calcul des $z_n$ de façon systématique, et ainsi réduire le temps de calcul de $O(N^2)$ à $O\left(N(\log N)^2\right)$.

De façon plus imagée, on utilise le passé sur un laps de temps de longueur variable pour préparer l'avenir sur un laps de temps de même longueur.

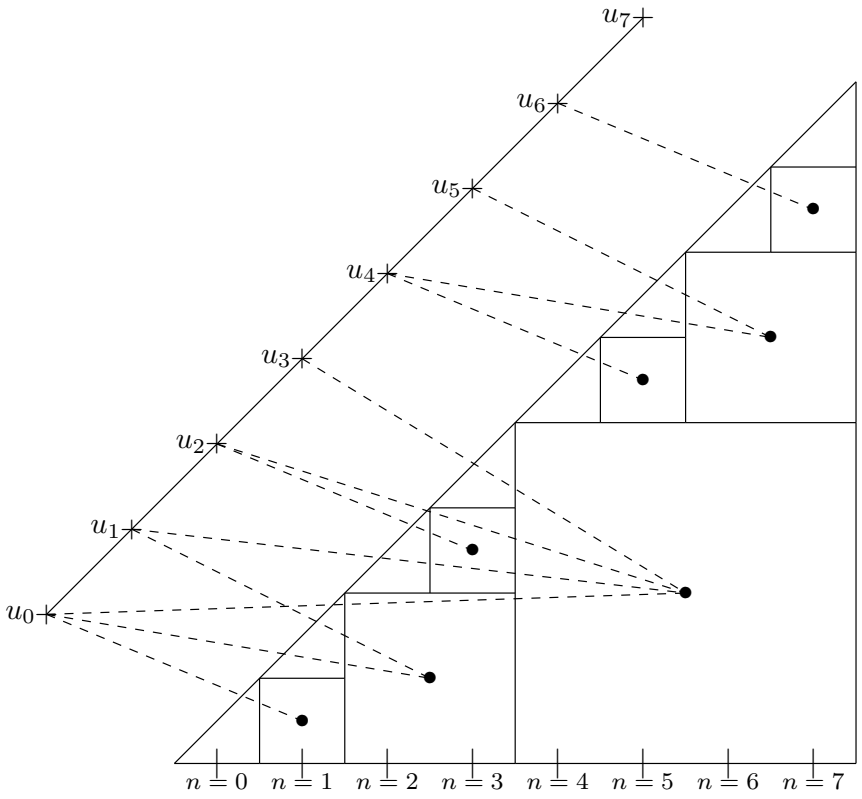



FIGURE 8.1 – Utilisation des termes $u_i$ pour calculer les valeurs $z_i$ via un pavage dyadique. Les pois noirs désignent les calculs de convolution par FFT.

*Remarque* 57 (Explication graphique). Sur la Figure 8.1 (à mettre en regard avec le descriptif des étapes ci-dessous), on remarque qu'on a fait appel 3 fois au terme $u_0$ pour calculer les valeurs $z_1$ jusqu'à $z_7$. Si l'on avait utilisé une méthode naïve, il aurait été utilisé 7 fois. On se figure qu'il sera utilisé $\lceil \log_2(N+1) \rceil$ fois pour calculer toutes les valeurs $z_1, ..., z_N$.

Pour ce faire, à chaque pas de temps $n \in [\![2^{M-1}, 2^M - 1]\!]$, on suppose que l'on dispose des valeurs de stockage $z_{j,n}$, qui doivent satisfaire :

$$z_{j,n} = \begin{cases} z_j & \text{si} \quad j \leq n, \\ 0 & \text{si} \quad j > 2^M. \end{cases} \tag{8.15}$$

Les termes $z_{j,n}$, pour $j \in [\![n+1, 2^M]\!]$ sont des sommes partielles de (8.12) de la forme

$$z_{j,n} = \sum_{k=0}^{i[j,n]} c_{j-k} f_k,$$

où la suite $i[j,n]$ est issue du fonctionnement de l'algorithme.

On raisonne itérativement en $n$. Pour $n=0$, on fixe $z_{j,0}=0$, pour tout $j\in\mathbb{N}$, et on vérifie que (8.15) est satisfaite. Puis, à chaque temps $n\geq 1$, on considère la décomposition dyadique

$$n=\sum_{l=0}^{+\infty} b_l 2^l.$$

On choisit le plus petit indice $l_n$ tel que $b_l\neq 0$ ($b_l$ ne peut prendre que les valeurs 0 ou 1). Puis, pour tout $j\in[\![0,2^{l_n}-1]\!]$, on alloue

$$z_{n+j,n}:=z_{n+j,n-1}+\sum_{k=n-2^l}^{n-1} c_{n+j-k}u_k,$$

calcul qui, comme on l'a précisé plus haut, s'effectue avec $O(l_n 2^{l_n})$ opérations grâce à l'utilisation de la FFT. Pour tous les autres $z_{n,j}$, on pose $z_{n,j}=z_{n-1,j}$. On observe au passage qu'il est inutile de stocker les valeurs $z_{j,n}$ pour $j>2^M$.

Montrons ce qui se passe sur les quatre premières itérations. Au départ, on a $z_{j,0}=0$. Puis :

— A l'étape $n=1$, on a $l_1=0$. On remplit donc la case 1 par

$$z_{1,1}=0+c_1f_0=c_1f_0,$$

ce qui donne donc $z_{1,1}=z_1$. Aucune autre valeur $z_{j>1,1}$ n'est affectée.

— A l'étape $n=2$, on a $l_2=1$. On remplit donc les cases 2 et 3

$$z_{2,2}=0+\sum_{k=0}^{1} c_{2-k}u_k \quad\text{et}\quad z_{3,2}=0+\sum_{k=1}^{2} c_{3-k}u_k.$$

Aucune autre valeur $z_{j>2,2}$ n'est affectée. On a donc $z_{2,2}=z_2$, et on a stocké une somme partielle sur la case 3.

— A l'étape $n=3$, on a $l_2=0$. On remplit donc la case 3

$$z_{3,3}=z_{3,2}+\sum_{k=0}^{0} c_{3-k}u_k=\sum_{k=0}^{2} c_{3-k}u_k.$$

Aucune autre valeur $z_{j>3,3}$ n'est affectée. On a donc $z_{3,3}=z_3$.

— A l'étape $n=4$, on a $l_2=2$. On remplit donc les cases $4,5,6,7$

$$z_{4,4}=0+\sum_{k=0}^{3} c_{4-k}u_k,\quad z_{5,4}=0+\sum_{k=0}^{3} c_{5-k}u_k,$$
$$z_{6,4}=0+\sum_{k=0}^{3} c_{6-k}u_k \quad\text{et}\quad z_{7,4}=0+\sum_{k=0}^{3} c_{7-k}u_k.$$

Aucune autre valeur $z_{j>4,4}$ n'est affectée. On a donc $z_{4,4}=z_4$, et des sommes partielles ont été stockées dans les cases de 5 à 7.

## 8.5 Dégénérescence et méthodes oublieuses

Nous étudions maintenant des méthodes de calcul qui, au contraire de la section précédente, *approximent* un schéma donné. Elles font usage de structures particulières afin de transformer les équations intégrodifférentielles (8.2) en équations différentielles ordinaires. Lubich et coauteurs ont qualifié de telles méthodes d'*oublieuses* (voir [137]) car elles ne font appel qu'à un nombre restreint de valeurs au temps présent $t$ pour calculer l'évolution de $u(t)$ ; elles sont donc moins gourmandes en mémoire. Nous présentons dans cette section les principaux ingrédients de cette approche.

### 8.5.1 Les noyaux dégénérés

Toute équation différentielle peut s'écrire sous une forme intégrale. L'inverse n'est pas vraie en général, sauf si le noyau $K(t,t')$ de (8.3a) se développe de la façon suivante :

$$K(t,t') = \sum_{j=0}^{d} a_j(t) b_j(t'), \qquad \text{pour} \quad 0 \le t' \le t \in \mathbb{R}_+. \tag{8.16}$$

On parle alors de noyau "dégénéré" de rang $d+1$ (voir [10, Chap. 2 p. 23], ou [105, p. 9] pour des équations intégrales de Volterra du 2$^{\text{ème}}$ type) ; le terme "séparable" est aussi utilisé dans la littérature (voir [74, p. 56]).

On suppose dans tout ce qui suit que les fonctions $a_j$ et $b_j$ sont régulières. Dans ce cas, on obtient un système de $d+1$ équations différentielles couplées. En effet, en injectant (8.16) dans (8.3a), et en posant

$$I_j(t) = \int_0^t b_j(t') u(t') \mathrm{d}t', \tag{8.17}$$

on trouve le système suivant :

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t} u(t) = -\displaystyle\sum_{j=0}^{d} a_j(t) I_j(t) + f(t), \\ \dfrac{\mathrm{d}}{\mathrm{d}t} I_j(t) = b_j(t) u(t), \qquad\qquad \forall j \in [\![1, d]\!], \end{cases} \tag{8.18}$$

avec les données initiales adéquates

$$I_j(0) = 0, \qquad\qquad \forall j \in [\![1, d]\!].$$

Le gain numérique lorsque l'on traite des équations intégrodifférentielles à noyau dégénéré de rang faible est évident : on troque en effet une équation avec mémoire contre un système d'équations différentielles ordinaires de faible dimension.

Cette décomposition peut sembler *a priori* inadaptée en ce qui concerne (8.1) pour deux raisons :

1. la structure de dégénérescence est très différente de la structure de convolution, qui semble être une propriété à préserver ;
2. les noyaux $K(t,t') := C_i(t-t')$ ne sont pas dégénérés.

Dans la section suivante, nous étudions donc les noyaux qui sont simultanément dégénérés et convolutifs. Ensuite, nous décrivons deux approches possibles pour construire une approximation dégénérée d'un noyau qui ne l'est pas au départ.

### 8.5.2 Dégénerescence et structure convolutive

Les noyaux convolutifs dégénérés présentent la double propriété d'engendrer des équations intégrodifférentielles réductibles à des équations différentielles ordinaires et de respecter la propriété d'invariance en temps des noyaux de convolution. Ils ont une forme simple, décrite par la :

**Proposition 8.5.1.** *Soit un noyau $C$ dégénéré s'écrivant*

$$C(t,s) = \sum_{j=0}^{d} a_j(t) b_j(s)$$

*où les fonctions $a_j$ et $b_j$ sont régulières. Le noyau $C$ est convolutif si et seulement s'il existe un polynôme $Q \in \mathbb{R}[X]$ tel que*

$$Q\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) C(t) = 0, \tag{8.19}$$

*c'est à dire que $C$ s'écrit sous la forme*

$$C(t) = \sum_{j=0}^{d} P_j(t) \mathrm{e}^{\lambda_j t}, \tag{8.20}$$

*faisant intervenir les polynômes $P_j \in \mathbb{R}[X]$ et les scalaires $\lambda_j \in \mathbb{C}$.*

Nous démontrons la Proposition 8.5.1 en Annexe A.6.1.

Si $C$ est convolutif et dégénéré, le système d'équations différentielles (8.18) n'est pas à coefficients constants (sauf cas triviaux). Toutefois, grâce à la Proposition 8.5.1, on peut proposer un autre système d'équations différentielles, qui est à coefficients constants, lui. Dans la section suivante, on explicite la construction de ce système d'équations différentielles.

### 8.5.3 D'une équation intégrodifférentielle à une équation différentielle ordinaire à coefficients constants

Par souci de simplicité dans les calculs, nous introduisons les polynômes de Laguerre $L_i$, qui forment une base échelonnée de $\mathbb{R}[X]$ (c'est à dire, de degré croissant). Ils ont les

propriétés utiles suivantes :

$$L_i(0) = 1, \tag{8.21}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} L_i(t) = -\sum_{k=0}^{i-1} L_k(t). \tag{8.22}$$

On introduit alors

$$y_i(t) := \int_0^t L_i(t-t')\mathrm{e}^{\lambda(t-t')} g(t')\mathrm{d}t', \tag{8.23}$$

où $g$ est une fonction continue. Grâce à (8.21) et (8.22), on montre aisément que, pour tout entier $d$ fixé,

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} y_0(t) \\ y_1(t) \\ \cdots \\ y_d(t) \end{pmatrix} = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ -1 & \lambda & 0 & \cdots \\ \cdots & \cdots & \cdots & 0 \\ -1 & -1 & -1 & \lambda \end{pmatrix} \begin{pmatrix} y_0(t) \\ y_1(t) \\ \cdots \\ y_d(t) \end{pmatrix} + \begin{pmatrix} g(t) \\ g(t) \\ \cdots \\ g(t) \end{pmatrix}.$$

**Application à la forme directe** (8.2a) Supposons qu'on a la décomposition suivante :

$$C(t) = \sum_{j=0}^{d} \sum_{i=0}^{\deg P_j} a_{ji} L_i(t) \mathrm{e}^{\lambda_j t}. \tag{8.24}$$

On pose les fonctions $y_{ji}$ construites sur le modèle de (8.23), c'est à dire

$$y_{ji}(t) = \int_0^t L_i(t-t')\mathrm{e}^{\lambda_j(t-t')} u(t')\mathrm{d}t' \qquad \text{pour} \quad j \in [\![0, d]\!] \quad \text{et} \quad j \in [\![0, \deg P_j]\!].$$

Alors la forme directe (8.2a) peut s'écrire comme :

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} u \\ Y \end{pmatrix} = M_{\mathrm{D}} \cdot \begin{pmatrix} u \\ Y \end{pmatrix} + \begin{pmatrix} f \\ \vec{0} \end{pmatrix}, \tag{8.25}$$

où $Y$ est le vecteur de dimension

$$D = \sum_{j=0}^{d} (1 + \deg P_j) \tag{8.26}$$

qui est constitué par $(y_{00}, \cdots, y_{0\deg P_0}, y_{20}, \cdots, y_{d\deg P_d})$, et où $M_{\mathrm{D}}$ est une matrice constante s'écrivant

$$M_{\mathrm{D}} = \left(\begin{array}{c|ccc|c|ccc} 0 & -a_{0,0} & \cdots & -a_{0,\deg P_0} & \cdots & -a_{d,0} & \cdots & -a_{d,\deg P_d} \\ \hline 1 & & & & & & & \\ \cdots & & B_0 & & 0 & & 0 & \\ 1 & & & & & & & \\ \hline 1 & & & & & & & \\ \cdots & & 0 & & \cdots & & 0 & \\ 1 & & & & & & & \\ \hline 1 & & & & & & & \\ \cdots & & 0 & & 0 & & B_d & \\ 1 & & & & & & & \end{array}\right)$$

où chacun des blocs $B_i$ est une matrice carrée de taille $(\deg P_i + 1) \times (\deg P_i + 1)$ de la forme

$$B_i = \begin{pmatrix} \lambda_i & 0 & \cdots & 0 \\ -1 & \lambda_i & 0 & \cdots \\ \cdots & \cdots & \cdots & 0 \\ -1 & -1 & -1 & \lambda_i \end{pmatrix}.$$

**Application à la forme résolue** (8.2b) Si l'on a la décomposition suivante :

$$\mathfrak{R}(t) = \sum_{j=0}^{d} \sum_{i=0}^{\deg P_j} a_{ji} L_i(t) \mathrm{e}^{\lambda_j t}, \tag{8.27}$$

on définit des fonctions

$$y_{ji}(t) = \int_0^t L_i(t - t') \mathrm{e}^{\lambda_j (t-t')} f(t') \mathrm{d}t',$$

où $j$ parcourt $[\![0, d]\!]$, et où, à $j$ donné $i$ parcourt $[\![0, \deg P_j]\!]$. Alors la forme directe (8.2b) peut s'écrire comme :

$$\frac{\mathrm{d}}{\mathrm{d}t} Y = M_{\mathrm{R}} \cdot Y + \begin{pmatrix} f \\ \cdots \\ f \end{pmatrix}, \tag{8.28}$$

où $Y$ est le vecteur $(y_{01}, \cdots, y_{0\deg P_0}, y_{20}, \cdots, y_{d\deg P_d})$, où $M_{\mathrm{R}}$ est une matrice constante, diagonale par blocs, et dont chacun des blocs est de taille $(\deg P_j + 1) \times (\deg P_j + 1)$ et de la forme

$$\begin{pmatrix} \lambda_j & 0 & \cdots & 0 \\ -1 & \lambda_j & 0 & \cdots \\ \cdots & \cdots & \cdots & 0 \\ -1 & -1 & -1 & \lambda_j \end{pmatrix}.$$

On retrouve alors $u$ grâce à la formule

$$u = \sum_{j=0}^{d} \sum_{i=0}^{\deg P_j} a_{ji} y_{ji}(t).$$

Nous étudions par la suite deux décompositions possibles. Dans la Section 8.5.6, nous imposons que la somme (8.24) (ou (8.27)) n'ait qu'un seul terme, avec $\lambda_0 = -1/2$ et nous construisons des polynômes $P_0$ de degré croissant. Au contraire, dans la Section 8.5.7, chacun des polynômes $P_j$ est constant, et on cherche des valeurs $\lambda_j$ efficaces. L'enjeu est le suivant : proposer une méthode rapide requerrant peu de mémoire tout en étant précise.

### 8.5.4 Stabilité des méthodes avec noyaux convolutifs dégénérés

Il est très simple d'étudier la stabilité en temps long des équations différentielles ordinaires (8.25) et (8.28). En effet, elles sont stables si et seulement si la matrice $\exp(tM)$ est bornée indépendamment de $t$, pour $M = M_{\mathrm{D}}$, respectivement $M = M_{\mathrm{R}}$. Une condition suffisante est que la matrice $M$ ait des valeurs propres qui soient toutes à partie réelle strictement négative. Les valeurs propres de la matrice $M_{\mathrm{R}}$ sont très simples : ce sont en effet les nombres $\lambda_j$ qui interviennent dans l'Expression (8.27). Ainsi, (8.28) est stable si tous les nombres $\lambda_j$ sont à partie réelle strictement négative.

Il est avantageux d'approximer directement la résolvante $\mathfrak{R}$ par une résolvante dégénérée $\mathfrak{R}_d$ satisfaisant (8.27), où les $\lambda_j$ sont tous à partie réelle strictement négative –cette dernière étant nécessairement stable. En revanche, l'approximation d'un noyau $C$ par des noyaux dégénérés (respectivement $C_d$ et $\mathfrak{R}_d$) peut engendrer une équation différentielle ordinaire instable (8.25), ou (8.28), alors même que l'équation originelle (8.2) était stable. En particulier, imposer que les $\lambda_j$ soient tous à partie réelle strictement négative dans (8.24) n'est pas suffisant pour assurer la stabilité de (8.25).

### 8.5.5 Schémas de splitting

Le forme des noyaux $C$ et $\mathfrak{R}$ que nous avons choisie suggère un *schéma* a priori inattendu pour une équation intégrodifférentielle. En effet, les équations différentielles ordinaires (8.25) et (8.28) peuvent être intégrées numériquement par un schéma de splitting (voir [77, p. 49]) entre l'opérateur linéaire d'une part, et le second membre induit par $f$ d'autre part. On peut par exemple proposer un splitting de Strang, d'ordre 2 en $\Delta t$. Nous en explicitons l'expression sur (8.25). Notons $\Psi_1(t,s)$ le propagateur du système

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} u(t') \\ Y(t') \end{pmatrix} = \begin{pmatrix} f(t') \\ \vec{0} \end{pmatrix} \quad \text{pour} \quad t' \in [s,t],$$

et $\Psi_2(t-s)$ le propagateur du système

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} u(t') \\ Y(t') \end{pmatrix} = M_{\mathrm{D}} \cdot \begin{pmatrix} u(t') \\ Y(t') \end{pmatrix} \quad \text{pour} \quad t' \in [s,t].$$

Il est immédiat que

$$\Psi_1(t,s)\cdot\begin{pmatrix} u \\ Y \end{pmatrix} = \begin{pmatrix} u \\ Y \end{pmatrix} + \begin{pmatrix} \int_s^t f(t')\mathrm{d}t' \\ \vec{0} \end{pmatrix},$$
$$\Psi_2(t-s)\cdot\begin{pmatrix} u \\ Y \end{pmatrix} = \exp((t-s)M)\cdot\begin{pmatrix} u \\ Y \end{pmatrix}.$$

Alors, le schéma de Strang revient à approximer

$$\begin{pmatrix} u(t+\Delta t) \\ Y(t+\Delta t) \end{pmatrix} \simeq \Psi_1\left(t+\Delta t, t+\Delta\frac{t}{2}\right)\cdot\Psi_2\left(\Delta t\right)\cdot\Psi_1\left(t+\frac{\Delta t}{2}, t\right)\cdot\begin{pmatrix} u(t) \\ Y(t) \end{pmatrix}.$$

Si l'on revient à l'équation (8.1), une telle dissociation des opérateurs linéaire et non-linéaire est appréciable, car on peut ainsi calculer le premier à l'aide des variables $k$ de Fourier, et le second à l'aide des variables $x$ physiques. Seul le propagateur $\Psi_2$ peut être calculé de manière exacte (par exponentiation). En ce qui concerne le propagateur $\Psi_1$, nous nous contentons en pratique de l'approximer via un schéma explicite de Runge-Kutta d'ordre 4 (voir [78, Tab. 1.2 p. 138]). Si l'on voulait préserver la symétrie du schéma, il faudrait utiliser alternativement des opérateurs discrétisés adjoints (voir [77, p. 49]), ce qui nécessite des méthodes non-linéaires implicites.

### 8.5.6 Une première méthode oublieuse : approximation par des polynômes de Laguerre pondérés

Dans cette section, nous indiquons une manière d'approximer la résolvante $\mathfrak{R}$ de (8.2a) par une suite de résolvantes dégénérées $\mathfrak{R}_d$ de la forme (8.27). Cette manière est tirée de [104] et repose sur l'utilisation des polynômes de Laguerre. Notons que [44] propose une méthode de résolution d'équations intégrodifférentielles avec dérivées fractionnaires temporelles (voir Section 7.5) à l'aide de polynômes de Laguerre généralisés.

Soit

$$\mathfrak{R}(t) \simeq \mathfrak{R}_d(t) := \sum_{j=0}^{d} a_j L_j(t)\mathrm{e}^{-t/2}, \tag{8.29}$$

où les $L_j$ sont les polynômes de Laguerre. On pose alors, pour tout $j \in \mathbb{N}$,

$$y_j(t) := \int_0^t L_j(t-t')\mathrm{e}^{-(t-t')/2} f(t')\mathrm{d}t'. \tag{8.30}$$

Par les calculs de la Section 8.5.3

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} y_0(t) \\ y_1(t) \\ \cdots \\ y_d(t) \end{pmatrix} = \begin{pmatrix} -1/2 & 0 & \cdots & 0 \\ -1 & -1/2 & 0 & \cdots \\ \cdots & \cdots & \cdots & 0 \\ -1 & -1 & -1 & -1/2 \end{pmatrix}\begin{pmatrix} y_0(t) \\ y_1(t) \\ \cdots \\ y_d(t) \end{pmatrix} + \begin{pmatrix} f(t) \\ f(t) \\ \cdots \\ f(t) \end{pmatrix}, \tag{8.31}$$

et on approxime la solution de (8.2b) par $u_d$ qui se décompose de la manière suivante :

$$u_d(t) = \mathfrak{R}_d(t)u_d(0) + \sum_{j=0}^{d} a_j y_j(t). \tag{8.32}$$

Une manière de construire une approximation (8.29) est proposée dans [104]. Nous l'utilisons ci-après. Comme les fonctions $L_j(t)\mathrm{e}^{-t/2}$ forment une base hilbertienne de $\mathrm{L}^2\left(\mathbb{R}, \mathrm{e}^{-t}\mathrm{d}t\right)$, alors si $\mathfrak{R} \in \mathrm{L}^2(\mathbb{R})$, on peut la décomposer comme

$$\mathfrak{R}(t) = \sum_{j=0}^{+\infty} a_j L_j(t)\mathrm{e}^{-t/2}, \tag{8.33}$$

où

$$a_j = \int_0^{+\infty} \mathfrak{R}(t)L_j(t)\mathrm{e}^{-t/2}\mathrm{d}t.$$

Les approximations dégénérées $\mathfrak{R}_d$ sont naturellement les sommes partielles de la série (8.33). Or, les polynômes de Laguerre constituent la série génératrice de

$$\exp\left(\frac{tz}{z-1}\right)(1-z)^{-1} = \sum_{j=0}^{+\infty} L_j(t)z^j.$$

D'où

$$\begin{aligned}\sum_{j=0}^{+\infty} a_j z^j &= (1-z)^{-1}\int_0^{+\infty} \mathfrak{R}(t)\exp\left(-\frac{t(1+z)}{2(1-z)}\right)\mathrm{d}t \\ &= (1-z)^{-1}\,\mathcal{L}\mathfrak{R}\left(\frac{1+z}{2(1-z)}\right).\end{aligned}$$

Il en découle

$$a_j = \frac{1}{j!}\left(\frac{\mathrm{d}^j}{(\mathrm{d}z)^j}\left[(1-z)^{-1}\,\mathcal{L}\mathfrak{R}\left(\frac{1+z}{2(1-z)}\right)\right]\right)(z=0),$$

ce qui fournit une manière de calculer les coefficients $a_j$ (on peut par exemple s'aider de Maple pour le faire).

*Remarque* 58. La matrice figurant dans (8.31) est une matrice de Toeplitz, dont l'exponentielle est aussi une matrice de Toeplitz. Par conséquent, la simulation de l'équation différentielle ordinaire (8.31) a une complexité qui croît comme $O(d\log d)$, et non comme $O(d^2)$.

### 8.5.7 Une seconde méthode oublieuse : utilisation d'une transformation de Laplace inverse numérique

Nous approximons maintenant la résolvante $\mathfrak{R}$ par une somme finie d'exponentielles. On pose

$$\mathfrak{R}_d(t) := \sum_{j=0}^{d} a_j \mathrm{e}^{\lambda_j t}, \tag{8.34}$$

où les $\lambda_j$ sont tous distincts (nous renvoyons le lecteur à [75, Sec. 3.1] pour de plus amples explications, que nous reproduisons partiellement). L'Identité (8.34) se traduit sur la transformée de Laplace de $\mathfrak{R}_d$ :

$$\mathcal{L}\mathfrak{R}_d(p) = \sum_{j=0}^{d} \frac{a_j}{p-\lambda_j}.$$

Plusieurs manières de construire cette approximation du noyau $\mathfrak{R}$ sont proposées dans [75], mais nous avons choisi d'utiliser les récentes méthodes de calcul oublieuses développées par Lubich et Schädle (voir [112]). Nous en expliquons les ressorts principaux dans cette section, en faisant notamment références aux articles [106, 112, 137], qui déploient cette méthode sur différents schémas ; nous nous concentrons ici sur le schéma simple issu de [112]. Notre angle d'attaque est purement numérique, mais il existe aussi un corpus théorique (voir, *e.g.*, [111]).

Les méthodes de calcul oublieuses de Lubich et Schädle interprètent la formule (8.34) comme une méthode de quadrature s'appliquant à l'intégrale de contour correspondant à l'inversion de la transformation le Laplace. En effet, on peut écrire

$$\mathfrak{R}(t) = \frac{1}{2\mathrm{i}\pi} \int_{\Gamma} \mathcal{L}\mathfrak{R}(\lambda) \mathrm{e}^{\lambda t} \mathrm{d}\lambda,$$

où $\Gamma$ est un contour qui englobe par la droite toutes les singularités et les coupures de $\mathfrak{R}$. Une fois discrétisée, l'intégrale ci-dessus devient alors

$$\mathfrak{R}(t) \simeq \frac{1}{2\mathrm{i}\pi} \sum_{j=0}^{d} \mathcal{L}\mathfrak{R}(\lambda_j) w_j \mathrm{e}^{\lambda_j t} =: \mathfrak{R}_d(t), \tag{8.35}$$

où les $w_j$ sont des poids d'intégration relatifs aux points $\lambda_j$ du contour $\Gamma$. Ainsi, on recouvre une décomposition du type de (8.34). Sur un autre plan, comme souligné dans [112], la transformée de Laplace $\mathcal{L}\mathfrak{R}$ est un forme qui apparaît naturellement dans de nombreuses équations intégrodifférentielles, qui s'écrivent plus facilement en variables de Laplace-Fourier (voir Section 6.3.1). Au contraire, l'expression analytique de la fonction $\mathfrak{R}$ est le fruit de calculs complexes, voire ne peut s'écrire avec des fonctions usuelles.

L'approximation (8.35) n'est précise que pour des temps $t$ situés sur un intervalle isolé de 0 et de $+\infty$. C'est pourquoi une deuxième technique est utilisée : un recouvrement de la

ligne temporelle par des intervalles de taille croissante, proche d'une décomposition dyadique (on parle d'une *mosaïque* ou en anglais *tessellation*). Sur chaque intervalle, on utilise un contour et une discrétisation adaptés afin de recouvrer une approximation précise de la transformation de Laplace.

Nous expliquons dans un premier temps comment des portions de l'intégrale de convolution (8.2b) sont approximées, puis comment ces approximations sont concaténées par un processus de mosaïque.

### La méthode de base

**Discrétisation d'une portion d'intégrale** Nous suivons maintenant [112, Sec. 2.3]. On souhaite approximer

$$\int_{t_1}^{t_0} \mathfrak{R}(t-t')f(t')\mathrm{d}t'.$$

Pour ce faire, on peut utiliser la formule d'inversion de la transformation de Laplace sur un contour $\Gamma$ bien choisi. Dans ce cas,

$$\int_{t_1}^{t_0} \mathfrak{R}(t-t')f(t')\mathrm{d}t' = \frac{1}{2\mathrm{i}\pi}\int_{\Gamma} \mathcal{L}\mathfrak{R}(\lambda)\mathrm{e}^{(t-t_0)\lambda}\left(\int_{t_1}^{t_0}\mathrm{e}^{(t_0-t')\lambda}f(t')\mathrm{d}t'\right)\mathrm{d}\lambda. \tag{8.36}$$

On remarque alors que la seconde intégrale ci-dessus, à savoir

$$y(t_0,t_1,\lambda) := \int_{t_1}^{t_0}\mathrm{e}^{(t_0-t')\lambda}f(t')\mathrm{d}t', \tag{8.37}$$

peut se réinterpréter comme étant la solution de l'équation suivante, évaluée au temps $t_0$,

$$\frac{\mathrm{d}}{\mathrm{d}t}y = \lambda y + f \qquad\qquad y(t_1) = 0. \tag{8.38}$$

Donc, après discrétisation de l'intégrale sur $\Gamma$, on réécrit (8.36) comme

$$\int_{t_1}^{t_0} \mathfrak{R}(t-t')f(t')\mathrm{d}t' \simeq \sum_{j=0}^{d} w_j \mathcal{L}\mathfrak{R}(\lambda_j)\mathrm{e}^{\lambda_j(t-t_0)}y(t_0,t_1,\lambda_j), \tag{8.39}$$

La méthode des trapèzes proposée dans [112] est une méthode de quadrature de l'intégrale sur $\Gamma$ particulièrement efficace. Elle converge de manière spectrale (voir [107]) ; c'est à dire que l'erreur de quadrature est dominée par $\mathrm{e}^{-\nu d}$, où $\nu$ est une constante strictement positive.

Une fois rendu à (8.39), il suffit de choisir une méthode d'intégration pour calculer $y(t_0,t_1,\lambda_j)$ –qui est solution de (8.38). A titre d'exemple, nous reproduisons le schéma proposé dans [112, Sec. 2.3] qui est d'ordre 2. Il correspond à approximer $f$ par une fonction affine entre deux temps d'intégration $t_n$ et $t_{n+1}$ :

$$y_{n+1} = y_n + \frac{\mathrm{e}^{\Delta t\lambda}-1}{\Delta t\lambda}\left(\Delta t\lambda y_n + \Delta t f_n + \Delta t\frac{f_{n+1}-f_n}{\Delta t\lambda}\right) - \Delta t\frac{f_{n+1}-f_n}{\Delta t\lambda}. \tag{8.40}$$

On peut aussi employer les schémas de Runge-Kutta de [137]. Pour notre part, nous avons implémenté cette *méthode de calcul* sur le *schéma* (8.40) et sur le *schéma* bloc-par-bloc (8.11).

**Contours et stabilité** Dans les cas que nous étudions, nous pouvons choisir un prolongement analytique de la fonction $\mathfrak{R}$ tel que ses lignes de coupures et ses singularités soient à partie réelle négative ou nulle sur le plan complexe (voir Figure 7.2). Comme le contour $\Gamma$ contourne par la droite ces singularités et lignes de coupure, les valeurs de $\lambda$ intervenant dans (8.39) sont donc à partie réelle soit négative (éventuellement de valeur absolue très grande), soit positive mais aussi petite que l'on veut.

Tout d'abord, dans le cas où $\mathrm{Re}\,(\lambda) < 0$, il est important de préserver numériquement la stabilité du système continu (8.38). Pour ce faire, on emploie des schémas numériques A-stables, c'est à dire qui convergent vers 0 lorsqu'ils approximent l'équation différentielle ordinaire $y' = \lambda y$ dès que $\mathrm{Re}\,(\lambda) < 0$ (voir [79, Chap. IV.3 p. 40]). On pourra consulter l'étude récente [15] pour l'emploi des méthodes de Runge-Kutta A-stables sur une équation intégrale à mémoire.

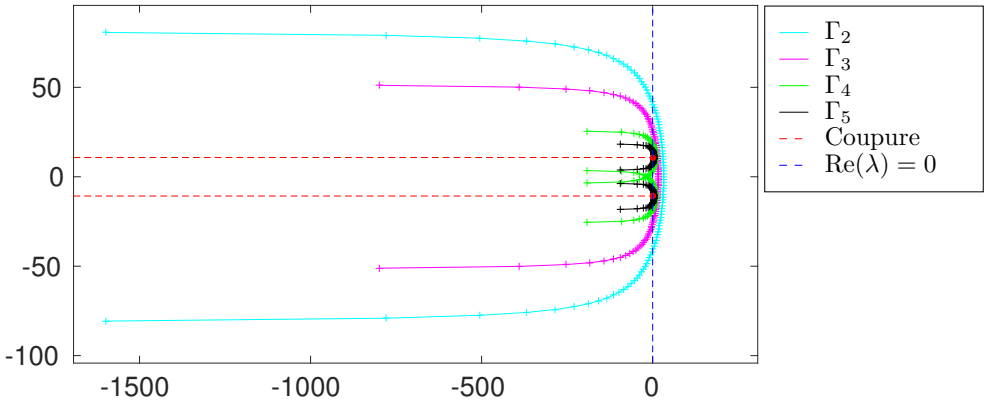



FIGURE 8.2 – Différents contours discrétisés $\Gamma_j$ relatifs à $\mathfrak{R}^{\alpha}_{\mathrm{III}}(|k|t)$, avec $|k| = 11.0$

Dès que $\lambda$ est à partie réelle strictement positive, la solution $y$ de (8.38) est instable. Par conséquent, deux contraintes antagonistes pèsent sur le contour $\Gamma$ :

— le contour $\Gamma$ doit "coller" autant que possible les lignes de coupure de $\mathfrak{R}$, afin que l'instabilité qui se développe comme $\exp\left(t \sup \mathrm{Re}\,(\Gamma)\right)$ soit contrôlée ;

— le contour $\Gamma$ et la fonction $\mathfrak{R}$ restreinte à $\Gamma$ doivent être aussi réguliers que possible pour que l'approximation soit précise.

Afin d'éviter le développement d'instabilités dues à $\mathrm{Re}\,(\lambda > 0)$, on contrôle la partie réelle positive du produit $\lambda_j t$. Pour ce faire, on n'emploie un contour $\Gamma_l$ que sur une plage de temps donnée

$$I_l := \left[B^{l-1}\Delta t, \left(2B^l - 1\right)\Delta t\right], \tag{8.41}$$

pour une constante $B \geq 2$ et une discrétisation temporelle $\Delta t > 0$ fixées. Lorsque $l$ croît, le contour $\Gamma_l$ est "contracté" autour des lignes de coupure de $\mathfrak{R}$, de telle sorte que la partie réelle

de chacun de ses points soit majorée par une quantité proportionnelle à $(2B^l\Delta t)^{-1}$. Cette "contraction" est faite de telle sorte que le contour $\Gamma_l$ reste régulier, et se ramène en général à une homothétie sur chacune des composantes de $\Gamma_l$. Pour maintenir un nombre constant de points de discrétisation sur le contours $\Gamma_l$, on tire parti du fait que la contribution des points $\lambda_j$ est insignifiante si $t\mathrm{Re}\,(\lambda_j) < 0$ et $|t\mathrm{Re}\,(\lambda_j)|$ est grande.

Dans [137, Sec. 4], le lecteur trouvera une discussion sur des formes de contours que l'on peut employer, et comment les paramétrer en fonction de $l$ de telle sorte que la méthode conserve une certaine stabilité (voir Figure 8.2 pour un exemple de contours, appelés "contours de Talbot", et l'Annexe A.7.2).

**Discrétisation de l'intégrale complète** Pour des $\lambda_j$ fixés, la discrétisation (8.39) n'est efficace, c'est à dire précise et stable, que sur les intervalles de temps $t-t' \in I_l$ isolés à la fois de 0 et de $+\infty$. Sur chacun de ces intervalles $I_l$ définis par (8.41), on se fixe un contour $\Gamma_l$. Ainsi, pour $0 = \tau_{l_{\max}+1} < ... < \tau_0 = t$, on décompose de la même manière que dans (8.36) :

$$\begin{aligned}\int_0^t \mathfrak{R}(t-t')f(t')\mathrm{d}t' = &\sum_{l=1}^{l_{\max}} \frac{1}{2\mathrm{i}\pi}\int_{\Gamma_l} \mathcal{L}\mathfrak{R}(\lambda)\mathrm{e}^{(t-\tau_l)\lambda}y(\tau_l,\tau_{l+1},\lambda)\mathrm{d}\lambda \\ &+ \int_{\tau_1}^{\tau_0} \mathfrak{R}(t-t')f(t')\mathrm{d}t', \end{aligned} \tag{8.42}$$

où $y$ est définie par (8.37).

Dans (8.42), l'intégrale entre $\tau_1$ et $\tau_0 = t$ a été séparée du reste car elle ne vérifie par la propriété que $t-t'$ est isolé de 0. Il faut en fait utiliser une autre méthode d'intégration que ce qui suit pour la traiter (voir par exemple [112, Sec. 2.4]).

Ainsi, après discrétisation de l'intégrale sur les contours $\Gamma_l$, on obtient :

$$\int_0^{\tau_1} \mathfrak{R}(t-t')f(t')\mathrm{d}t' \simeq \sum_{l=1}^{l_{\max}}\sum_{j=0}^{d} W_{lj}\mathrm{e}^{\lambda_{lj}(t-\tau_l)}y(\tau_l,\tau_{l+1},\lambda_{lj}), \tag{8.43}$$

où $\lambda_{lj} \in \Gamma_l$ sont les points de discrétisation du contour $\Gamma_l$, et où

$$W_{lj} := \mathcal{L}\mathfrak{R}(\lambda_{lj})w_{lj},$$

avec $w_{lj}$ les poids relatifs à la méthode des trapèzes.

L'article [112] explique comment fixer astucieusement les temps intermédiaires $\tau_l$, afin de tirer parti du fait que $y(\cdot,\cdot,\lambda_{lj})$ est la solution d'une équation différentielle ordinaire et comment faire en sorte qu'il ne soit pas nécessaire de garder en mémoire le passé $f(t' < t)$. Nous le détaillons dans la section suivante.

**La mosaïque**

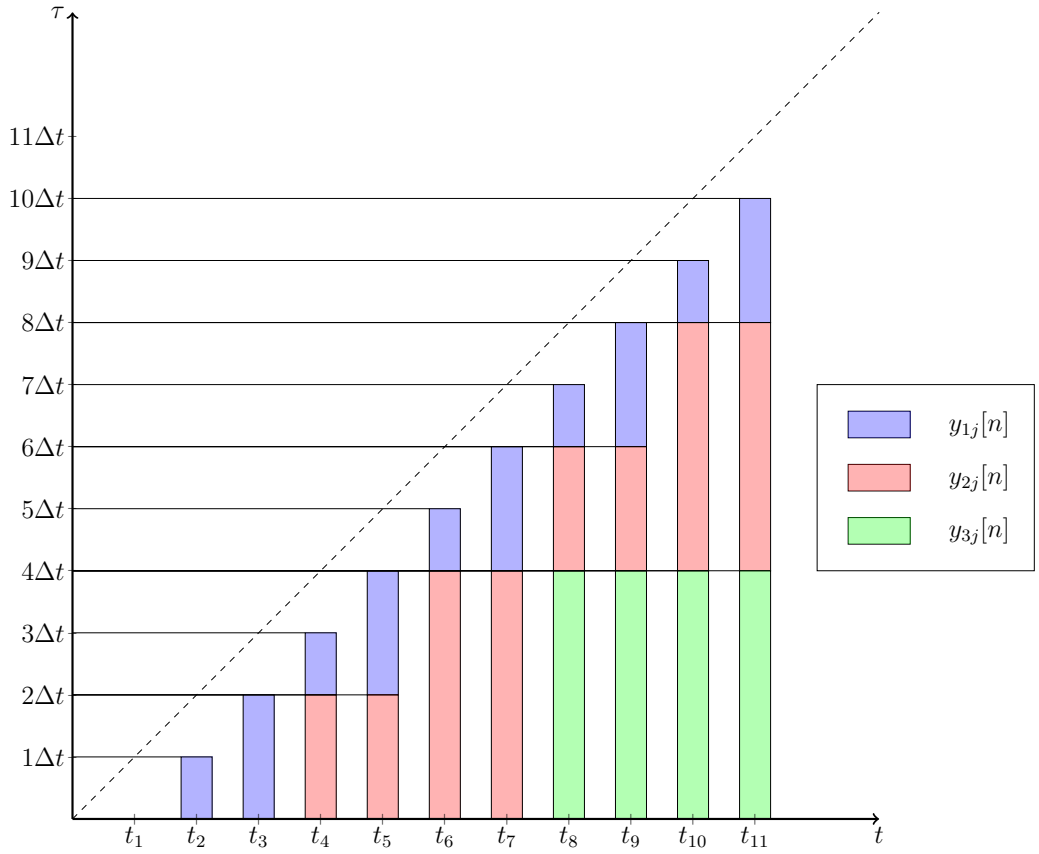



FIGURE 8.3 – Utilisation des différentes quantités $y_{lj}[n]$ dans (8.44), pour le cas $B = 2$.

Nous décrivons maintenant la technique des "mosaïques" de [112]. Cette technique algorithmique est apparentée à la méthode d'accélération présentée dans la Section 8.4.

La décomposition temporelle $(\tau_l)$ dans (8.42) dépend en fait implicitement de $t$. Afin de le mettre en évidence, précisons nos notations. On définit $\tau_l^n$ la décomposition relative à $t_n$ et

$$y_{lj}[n] := y\left(\tau_l^n, \tau_{l+1}^n, \lambda_{lj}\right).$$

Ainsi, (8.43) se réécrit :

$$\int_0^{\tau_1^n} \mathfrak{R}(t_n - t')f(t')\mathrm{d}t' \simeq \sum_{l=1}^{l_{\max}} \sum_{j=0}^{d} W_{lj} r_{lj}[n] y_{lj}[n], \tag{8.44}$$

où

$$r_{lj}[n] := \exp\left(\left(t_n - \tau_l^n\right)\lambda_{lj}\right).$$

Les quantités $W_{lj}$ et $r_{lj}[n]$ présentes dans (8.44) sont aisément calculables dès lors que la décomposition $(\tau_l^n)$ est fixée. La détermination de $\tau_l^n$ et de $y_{lj}[n]$ est plus délicate. De la même manière que dans la Section 8.4, la manière la plus transparente de procéder est récursive.

**Détermination de $\tau_l^n$** Au début, $\tau_l^0 = 0$, pour tout $l \in \mathbb{N}$. Puis, pour déterminer $\tau_l^{n+1}$, on effectue l'opération suivante (voir le pseudo-code de [137, Sec. 4]) :

1. pour $l \geq 1$ croissant, tant que $B^{l-1}$ divise $n$ et que $n > B^{l-1}$, on fixe $\tau_l^{n+1} := \tau_l^n + B^{l-1}\Delta t$ ;
2. puis, pour les $l$ plus grands on laisse $\tau_l^{n+1} := \tau_l^n$.

Par conséquent, pour $n$ fixé, la suite $(\tau_l^n)$ est décroissante en $l$ et stationnaire à 0. Par ailleurs, on a

1. si $n \leq 2B^l - 1$, $\tau_{l+1}^n = 0$ ;
2. sinon, $\left[t_n - \tau_l^n, t_n - \tau_{l+1}^n\right] \subset \left[B^{l-1}\Delta t, \left(2B^l - 1\right)\Delta t\right] = I_l$.

**Détermination de $y_{lj}[n]$** Par définition, pour $l$ fixé et $n$ variable, la suite des intervalles $\left[\tau_l^n, \tau_{l+1}^n\right]$ est constante sur des paliers de longueur $B^{l-1}$. Ainsi, la valeur de $y_{lj}[n]$ n'est mise à jour que tous les $B^{l-1}$ pas de temps. Parallèlement, la différence $\tau_{l+1}^n - \tau_l^n$ vaut périodiquement, lorsque $n$ varie, $B^{l-1}$, $2B^{l-1}$, ..., $B^l$, $B^{l-1}$, $2B^{l-1}$, ..., *etc.* Pour mener à bien ce processus, trois versions de $y_{lj}[n]$ sont nécessaires :

1. $y_{lj}[n]$ lui-même, employé pour évaluer (8.44) ;
2. $y_{lj}^{\mathrm{p}}[n]$, qui constitue la prochaine mise à jour de $y_{lj}[n]$ ;
3. $y_{lj}^{\mathrm{c}}[n]$, qui est calculée au fur et à mesure à partir des valeurs $f(t')$, pour $t' \in [t_{n-1}, t_n]$, grâce à (8.38).

La quantité la plus variable est $y_{lj}^{\mathrm{c}}[n]$, qui est mise à jour sur chaque intervalle de temps $[t_n, t_{n+1}]$ comme étant la solution (approchée) de

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_{lj}\right) y = f.$$

C'est elle qui agrège la mémoire. Lorsque l'on atteint un temps $t_n = \tau_l^{n+m}$ pour un certain $m$, alors on stocke $y_{lj}^{\mathrm{p}}[n] := y_{lj}^{\mathrm{c}}[n]$, qui sera la prochaine valeur employée dans $y_{lj}[n+m]$. Si jamais $\tau_{l+1}^{n+m} - \tau_l^{n+m} = B^l$, alors on remet à zéro $y_{lj}^{\mathrm{c}}[n] := 0$.

Les Figures 8.3 et 8.4 illustrent sur un exemple comment est utilisée $y_{lj}[n]$ au fil du temps, et comment est géré le calcul et la mise à jour des trois avatars de $y_{lj}[n]$.

*Remarque* 59. Pour que la méthode soit effectivement efficace, il faut se donner a priori une borne supérieure $T > t$ sur les temps $t$ considérés, et on fixe ensuite

$$l_{\max} := \left\lfloor \frac{\log(T)}{\log(B)} \right\rfloor + 1.$$

Seuls les $l \in [\![0, l_{\max} + 1]\!]$ jouent alors un rôle dans (8.44). En pratique, cela signifie qu'il faut avoir une estimation sur le moment de fin de l'expérience numérique avant de commencer celle-ci.

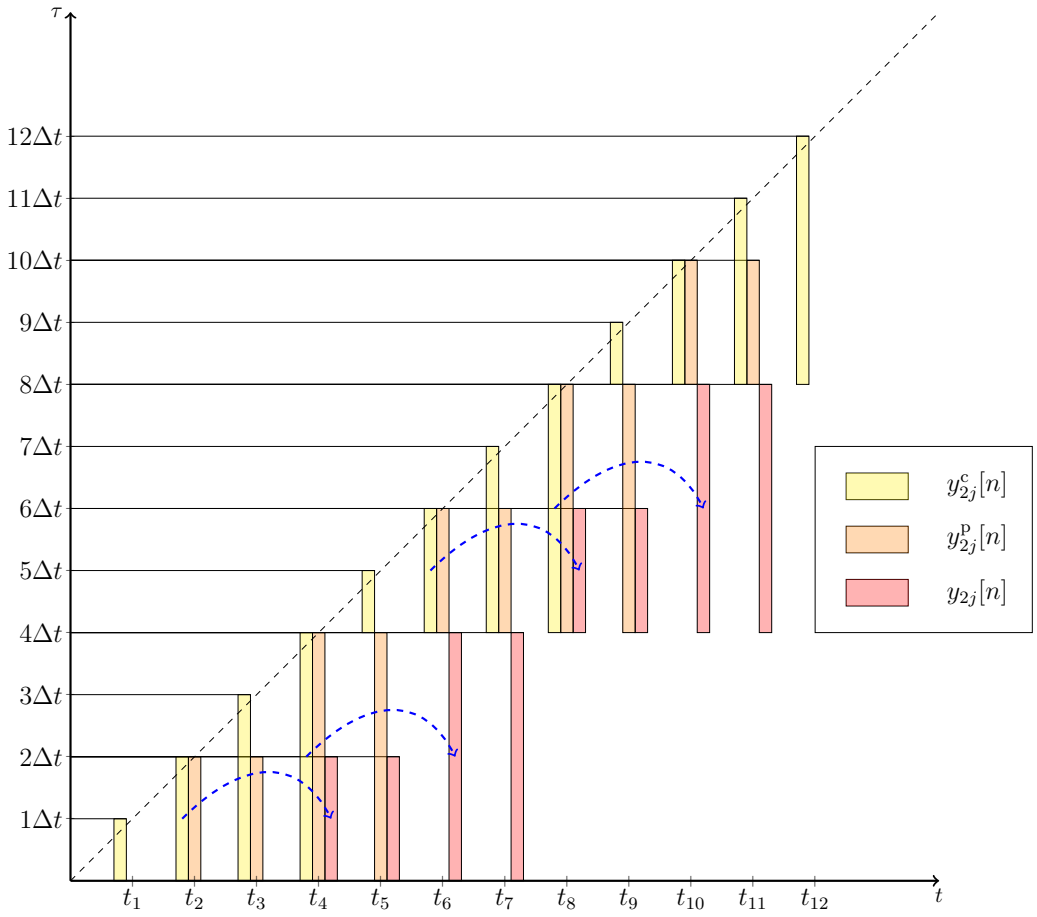



FIGURE 8.4 – Mise à jour de $y_{lj}[n]$, $y^{p}_{lj}[n]$, et $y^{c}_{lj}[n]$ dans le cas où $B = 2$ et $l = 2$. Les flèches en tiret figurent comment la mémoire est transférée de $y^{c}_{lj}[n]$ jusqu'à $y_{lj}[n]$, via $y^{p}_{lj}[n]$

### Considérations sur la méthode

Une rapide analyse démontre que la méthode décrite ci-dessus a une complexité temporelle en $O(dN \log N)$ où $\log N$ correspond à $l_{\max}$ (le nombre d'intervalles temporels $I_l$) et $d+1$ est le nombre de points de discrétisation de l'intégrale sur le contour $\Gamma_l$. En outre, elle utilise une quantité de mémoire en $O(d \log N)$. Par conséquent, elle est asymptotiquement *plus économe* en mémoire qu'une méthode non dégénérée, qui requiert une mémoire en $O(N)$.

*Remarque* 60. Il n'est pas nécessaire que le pas de temps soit fixe : une méthode à pas adaptatif est proposée dans [106].

## 8.6 Tests numériques

Dans cette Section, nous présentons plusieurs tests numériques sur la base desquels nous comparons les différents algorithmes implémentés.

Tout d'abord, nous évaluons leur capacité à résoudre efficacement l'équation sur *un seul mode* de Fourier, que nous rappelons ci-dessous :

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}u(t) = -\displaystyle\int_0^t \frac{C_i(t-t')}{\kappa_i^\alpha}u(t')\mathrm{d}t' + \frac{f(t)}{\kappa_i^\alpha}, \\ u(0) = u_0, \end{cases} \tag{8.45}$$

ou sa forme avec résolvante :

$$u(t) = \mathfrak{R}_i^\alpha(t)u_0 + \frac{1}{\kappa_i^\alpha}\int_0^t \mathfrak{R}_i^\alpha(t-t')f(t')\mathrm{d}t'. \tag{8.46}$$

On observe dans un premier temps dans quelle mesure les méthodes considérées peuvent approximer la résolvante $\mathfrak{R}$ elle-même (c'est à dire que l'on prend $f = 0$). Dans le cas des méthodes reposant sur la forme résolue (8.46), la précision sur $\mathfrak{R}_i^\alpha$ ainsi calculée est évidemment indépendante du pas temporel ; toutefois cette précision est limitée par d'autres paramètres qui entrent en jeu lorsqu'on utilise des méthodes oublieuses. Dans un second temps, nous ajoutons un forçage régulier $f$ à (8.45). Enfin, nous considérons l'équation (8.1) dans toute sa complexité. Ces dernières expériences permettent de valider les différentes approches.

*Remarque* 61. Tous les tests effectués ci-dessous ont été effectués sur un code Matlab tournant sur un ordinateur personnel. Cela soulève deux points. D'une part, les spécificités du langage Matlab font que les temps de calcul ne sont pas forcément très représentatifs. D'autre part, les tests ont été effectués sur des systèmes de taille relativement petite, et sur des durées relativement courtes.

### 8.6.1 Nomenclature

Afin de parler précisément des algorithmes que nous allons employer, nous proposons la nomenclature suivante. Celle-ci basée sur deux blocs de lettres. Le premier bloc désigne le *schéma* employé. Il est constitué de trois lettres :

- les deux premières lettres désignent le type de schéma :
  - LR, pour schéma de Lapusta, Rice et coauteurs, décrit dans la Section 8.3.1,
  - BB, pour les schémas Blocs-par-Blocs, décrits dans les Section 8.3.2 et 8.3.3,
  - SS, pour les schémas de Splitting de Strang décrits dans la Section 8.5.5,
  - LS, pour le schéma de Lubich et Schädle, décrit dans la Section 8.5.7 (voir (8.40)).
- la troisième lettre désigne la forme de l'équation utilisée, c'est à dire :
  - D pour la forme *directe* de type (8.2a),
  - R forme *résolue* de type (8.2b),

Le second bloc est constitué d'une lettre et désigne la *méthode de calcul* de l'algorithme, à savoir :

— aucune lettre si l'algorithme est implémenté de manière naïve,
— A pour l'*Accélération* décrite dans la Section 8.4,
— L pour la méthode utilisant la décomposition en polynômes de *Laguerre* de la Section 8.5.6,
— O pour la méthode *Oublieuse* reposant sur l'inversion numérique des transformées de Laplace décrite dans la Section 8.5.7.

La philosophie étant que l'on peut se fixer indépendamment un schéma (lié en général à la forme de l'équation) et une méthode de calcul.

Nous avons implémenté 7 algorithmes principaux (tous ne sont pas implémentés pour tous les types de dislocations). Le tableau suivant résume l'implémentation effective des algorithmes (dans un code MATLAB) :

TABLE 8.1 – Implémentation effective des méthodes (le "X" signifie que la méthode a été implémentée)

| Algorithme | Schéma | Méthode de calcul | Mode I | Mode II | Mode III |
|---|---|---|---|---|---|
| LRD- | LRD- | Naïve | - | - | X |
| BBD- | BBD- | Naïve | X | X | X |
| BBD-A | BBD- | Accélérée | X | X | X |
| SSD-L | SSD- | Laguerre | X | X | X |
| SSR-L | SSR- | Laguerre | X | X | X |
| LSR-O | LSR- | Oublieuse | - | X | X |
| BBR-O | BBR- | Oublieuse | - | X | X |

### 8.6.2 Méthodologie

Nous procédons à plusieurs tests numériques de la manière suivante :

1. nous fixons une équation intégrodifférentielle (E) à résoudre sur un intervalle de temps $[0, T]$ ;
2. nous calculons une solution numérique de référence $u_{\text{ref}}$ de cette équation (avec la méthode bloc-par-bloc directe BBD-A, et un pas de temps très petit) ;
3. nous choisissons un algorithme, et l'employons pour approximer la solution de (E), en faisant varier le pas de temps $\Delta t$, et éventuellement d'autres paramètres de l'algorithme (notamment lorsqu'elle nécessite une approximation du noyau). On obtient ainsi des fonctions $u_{\text{num}}^{\Delta t}(t)$ ;

4. nous comparons les sorties de l'algorithme considéré avec la solution de référence. En particulier, nous étudions l'erreur

$$e(t) = \left| u_{\text{ref}}(t) - u_{\text{num}}^{\Delta t} \right|, \tag{8.47}$$

et l'indicateur aggrégé :

$$E = \max_{n \in [\![1, T/\Delta t]\!]} \left| u_{\text{ref}}(n\Delta t) - u_{\text{num}}^{\Delta t}(n\Delta t) \right|, \tag{8.48}$$

en faisant varier divers paramètres.

Ainsi, nous pouvons non seulement déterminer numériquement les propriétés des algorithmes proposés, mais aussi comparer ces derniers entre eux. Nous sommes particulièrement sensibles aux critères suivants :

— l'ordre,
— la stabilité,
— le coût (en termes de mémoire et de calcul).

*Remarque* 62 (Saturation de l'erreur). Comme notre solution de référence est numérique, celle-ci comporte une erreur intrinsèque. Par conséquent, on peut avoir un phénomène apparent de saturation de l'erreur : celle-ci semble rester stationnaire alors que l'on raffine le calcul. Cela est dû au fait que l'on obtient une précision du calcul inférieure à l'erreur de la solution numérique de référence.

### 8.6.3 Simulation de l'équation réduite

#### Equation homogène : reconstruction de la résolvante

Nous comparons tout d'abord les différents algorithmes lorsqu'il s'agit de reconstruire les résolvantes $\mathfrak{R}_{\text{III}}^{\alpha}$ et $\mathfrak{R}_{\text{II}}^{\alpha}$. Pour ce faire, nous proposons les deux tests suivants :

**Test numérique 1.** *Approximer numériquement la solution de* (8.45) *sur l'intervalle temporel* $[0, 200]$*, avec*

— $i = \text{III}$ *et* $\alpha = 0.1$*,*
— $f = 0$ *et* $u_0 = 1$*.*

**Test numérique 2.** *Approximer numériquement la solution de* (8.45) *sur l'intervalle temporel* $[0, 200]$*, avec*

— $i = \text{II}$*,* $\alpha = 0.1$ *et* $\gamma = \sqrt{3} \simeq 1.73$*,*
— $f = 0$ *et* $u_0 = 1$*.*

Pour ces deux premiers tests, les algorithmes utilisant la forme directe (8.2a), et ceux utilisant la forme résolue (8.2b) se comportent très différemment. En effet, les premiers, sauf SSD-, n'intègrent pas exactement la partie linéaire de l'équation (8.2a) et produisent un résultat dont la précision augmente lorsque le pas de temps $\Delta t$ diminue. Au contraire, les seconds possèdent *a priori* la résolvante $\mathfrak{R}_i^{\alpha}$ de l'équation. Aussi, leur précision ne dépend pas du pas de temps $\Delta t$. Toutefois, en ce qui concerne les méthodes oublieuses, la précision avec laquelle (8.2b) est résolue dépend directement de la qualité de l'approximation faite sur la résolvante.

**Les schémas d'intégration de la forme directe : BBD-A, LRD-, et SSD-L** On effectue les Tests 1 et 2. Alors, on constate sur la Figure 8.5(a) que le schéma bloc-par-bloc BBD-, et le schéma LRD- sont effectivement d'ordre 4 et 2, respectivement. Mais il apparaît que le schéma LRD- est instable, tandis que BBD- demeure stable, lorsque $\Delta t$ est grand. Ceci disqualifie clairement le schéma LRD- : il est de coût similaire à BBD-, mais d'ordre inférieur et moins stable. Nous n'utiliserons désormais LRD- que dans certains cas du mode III, à titre de comparaison.

En ce qui concerne le mode II, on constate sur la Figure 8.5(a) que le schéma BBD- est à nouveau d'ordre 4, mais est instable pour certains pas de temps $\Delta t > 1$. Cette différence de comportement entre les modes II et III s'explique vraisemblablement par le fait que le noyau $C_{\mathrm{III}}(T)$ converge plus rapidement vers 0 à l'infini que le noyau $C_{\mathrm{II}}(T)$ (en $O(T^{-3/2})$ et en $O(T^{-1/2})$ respectivement).

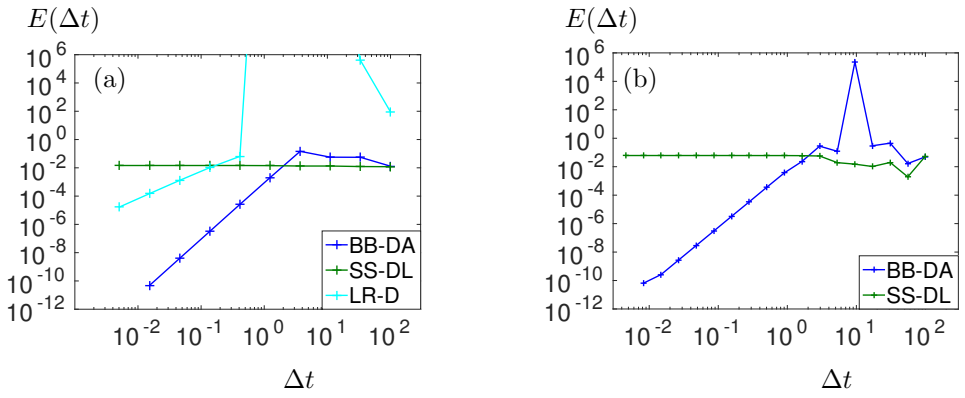



FIGURE 8.5 – Erreur $E(\Delta t)$ définie par (8.48) pour les algorithmes BBD-A, LRD-, SSD-L, dans le cas des (a) Test 1 et (b) Test 2. Le nombre de modes pour SSD-L est $d = 198$.

Au contraire des deux algorithmes précédents, l'algorithme SSD-L produit un résultat qui est indépendant de $\Delta t$ car le schéma sous-jacent est exact en l'absence de second membre. Cependant, comme on peut le constater, la précision reste relativement faible vis-à-vis des deux autres méthodes, à cause de la qualité médiocre de l'approximation du noyau $C_i$ (l'erreur est donc entièrement due à la méthode de calcul). En outre, pour les paramètres du Test 2 (et $d = 198$), la matrice d'évolution possède une valeur propre de partie réelle strictement positive $0.0068 > 0$. Donc, l'algorithme est instable. Cette instabilité apparaît de manière sensible pour les temps grands (pour $T > 500$ dans le cas du Test 2).

**Les schémas d'intégration de la forme résolue : SSR-L, BBR-O, LSR-O** Par définition, les schémas d'intégration de la forme résolue (8.2b) sont exacts en l'absence de forçage. Mais les méthodes indirectes que nous avons présentées ont toutes un caractère oublieux : ainsi, elles reposent fondamentalement sur une approximation de la résolvante. Nous discutons ici de l'efficacité des mécanismes d'approximation de la résolvante.

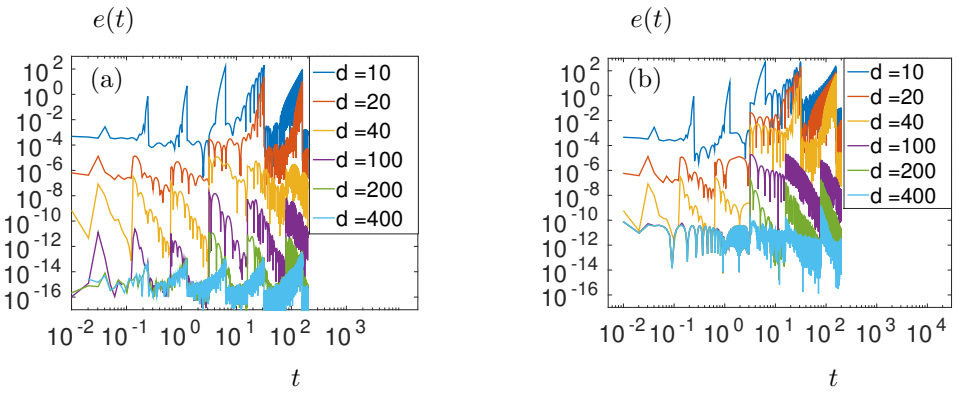



FIGURE 8.6 – Erreur $e(t)$ définie par (8.47) pour BBR-O en fonction de $d$, pour $\Delta t = 0.0051$, dans le cas (a) du Test 1 et (b) du Test 2.

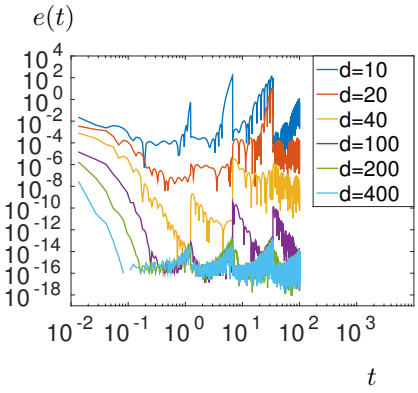



FIGURE 8.7 – Précision sur la reconstruction de $\mathfrak{R}^0_{\mathrm{III}}$ grâce à la méthode d'inversion de Laplace de la Section 8.5.7, pour différents nombres de modes. Ici $\Delta t = 0.13$.

**Inversion numérique de la transformation de Laplace** Les méthodes BBR-O et LSR-O partagent un même mécanisme d'approximation de la résolvante basé sur la transformation de Laplace numérique inverse de [107]. On constate empiriquement que sur l'intervalle $[\Delta t, +\infty[$, on obtient une méthode précise et d'erreur faible comme en témoignent les Figures 8.6(a) et 8.6(b), où est tracée $e(t)$ pour $d$ variable. En outre, on observe que cette erreur décroît très rapidement lorsque l'on augmente le nombre de modes, comme l'illustre la Table 8.2. Avec un nombre relativement restreint de modes ($d = 100$), on obtient des précisions de l'ordre de $10^{-8}$ pour le mode III, et de $10^{-5}$ pour le mode II (voir Table 8.2). On remarque aussi la forme caractéristique en dent de scie de l'erreur : la pointe des dents correspond au moment où on change de contour d'intégration (voir Section 8.5.7). Si on ne changeait pas de contour, l'erreur exploserait exponentiellement avec le temps. Inversement,

si, à contour donné, on cherchait à employer la méthode pour un temps $t$ trop petit (ce qu'on ne fera pas en pratique), l'erreur augmenterait aussi, comme on peut l'observer pour les temps $t < \Delta t$ sur les les Figures 8.6(a) et 8.6(b). Nous renvoyons à l'Annexe A.7.2 pour une description des paramètres que nous avons choisis.

A nombre de modes $d$ donné, la méthode d'inversion de la transformée de Laplace est plus précise pour recouvrer $\mathfrak{R}^{\alpha}_{\mathrm{III}}$ que $\mathfrak{R}^{\alpha}_{\mathrm{II}}$. Cela s'explique par un argument géométrique simple. Le prolongement analytique de $\mathcal{L}\mathfrak{R}^{\alpha}_{\mathrm{II}}$ possède 4 pôles et 4 lignes de coupures, tandis que celui de $\mathcal{L}\mathfrak{R}^{\alpha}_{\mathrm{III}}$ ne possède que 2 lignes de coupures. D'où un contour $\Gamma$ plus complexe pour calculer l'inversion de la transformation de Laplace de $\mathcal{L}\mathfrak{R}^{\alpha}_{\mathrm{II}}$ que celle de $\mathcal{L}\mathfrak{R}^{\alpha}_{\mathrm{III}}$. Ainsi, à précision égale, on a besoin de plus de points sur le contour dans le premier cas que dans le second.

**Approximation par des polynômes de Laguerre** Nous effectuons maintenant les Tests 1 et 2 sur l'algorithme SSR-L. On constate que les résolvantes $\mathfrak{R}_{\mathrm{II}}$ et $\mathfrak{R}_{\mathrm{III}}$ sont d'autant mieux restituées que le nombre de modes $d+1$ est grand. Toutefois, cette approximation converge relativement plus lentement en $d$ que celle de la transformation de Laplace inverse (voir Table 8.2). En outre, à cause de la forte décroissance imposée par le terme exponentiel de $L_j(t)\mathrm{e}^{-t/2}$, elle impose un cut-off exponentiel sur les queues (voir Figure 8.8(a) et (b) et 8.9).

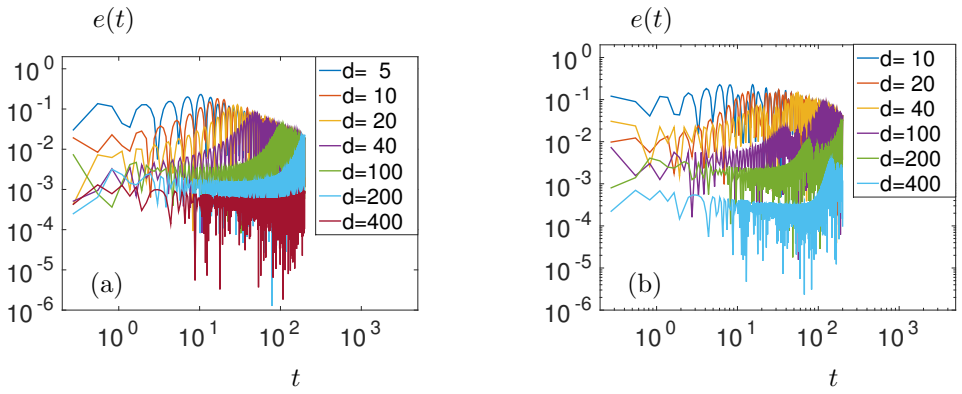



FIGURE 8.8 – Erreur $e(t)$ définie par (8.47) pour SSR-L en fonction de $d$, dans le cas (a) du Test 1 et (b) du Test 2.

### Equation inhomogène

Nous effectuons maintenant un test où, au contraire, la donnée initiale est égale à 0, et où on impose un chargement régulier $f(t)$.

**Test numérique 3.** *Approximer numériquement la solution de* (8.45) *sur l'intervalle temporel* $[0, 200]$, *avec*

— $i = \mathrm{III}$ *et* $\alpha = 0.1$,

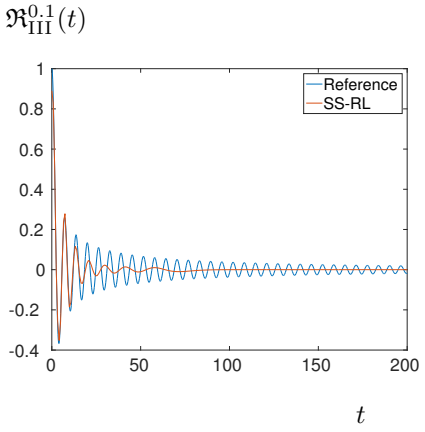



FIGURE 8.9 – Approximation numérique de $\mathfrak{R}^{0.1}_{\mathrm{III}}$ par SSR-L, $d = 20$.

— $f(t) = \sin(t/\pi)$ *et* $u_0 = 0$.

**Test numérique 4.** *Approximer numériquement la solution de* (8.45) *sur l'intervalle temporel* $[0, 200]$, *avec*
— $i = \mathrm{II}$, *et* $\alpha = 0.1$,
— $f(t) = \sin(t/\pi)$ *et* $u_0 = 0$.

On observe à nouveau des comportements différents entre les algorithmes reposant sur une formule exacte du noyau (LRD- et BBD-A) et ceux reposant sur une approximation du noyau (SSD-L, SSR-L, BBR-O et LSR-O), illustrés par les Figures 8.10(a) et 8.10(b). Les algorithmes LRD- et BBD-A ont deux régimes :

1. un premier régime pour $\Delta t$ grand, où la méthode est au mieux stable mais donne une réponse peu précise ;
2. un second régime pour $\Delta t$ plus petit, où la méthode donne une réponse de plus en plus précise lorsque $\Delta t$ diminue. On obtient alors un ordre correspondant à l'ordre du schéma sous-jacent.

Pour les algorithmes SSD-L, SSR-L, BBR-O et LSR-O, un troisième régime apparaît : pour $\Delta t$ est très petit, l'erreur sature à une valeur $E_d > 0$ (c'est particulièrement visible pour SSD-L, SSR-L et BBR-O sur la Figure 8.10(b)). Cette saturation de l'erreur est due à l'erreur résiduelle sur les noyaux $\mathfrak{R}^{\alpha}_{i}$ considérés.

On observe par ailleurs que LSR-O est plus précis que BBR-O lorsque $\Delta t$ est grand. Nous interprétons cela comme la conséquence de deux faits :

1. le schéma BBR- a tendance à sur-amortir le forçage lorsque $\Delta t$ est grand (non montré) ;
2. le schéma LSR- est exact pour des fonctions affines par morceaux sur le maillage $\Delta t\mathbb{Z}$. Or, dans les Tests 3 et 4, on a pris une fonction de forçage lisse.

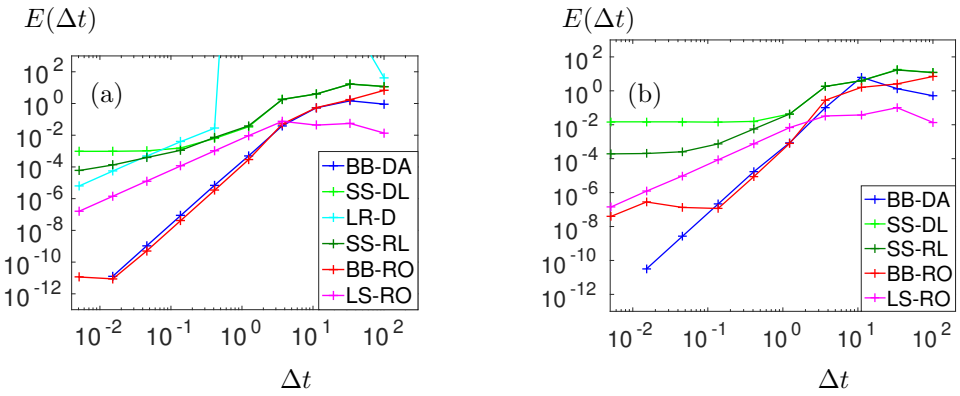



FIGURE 8.10 – Erreur sur le calcul de $u$ par les algorithmes BBD-A, SSD-L, LRD-, SSR-L, BBR-O, LSR-O pour le (a) Test 3 et (b) Test 4. Le nombre de modes pour pour SSD-L est $d = 100$, et SSR-L, $d = 498$, et pour BBR-O et LSR-O, $d = 200$.

**Méthodes $d$-dépendantes** Nous faisons à nouveau varier le nombre de modes $d+1$ pour les méthodes SSR-L et BBR-O afin d'observer son influence sur le résultat dans les Tests 3 et 4. On constate alors (voir Table 8.2) des erreurs qui sont cohérentes avec les Tests 1 et 2. Une fois de plus, la méthode BBR-O permet d'avoir un résultat très précis à condition d'utiliser un nombre de modes $d + 1$ relativement grand, mais est très peu précise si le nombre de modes est trop faible. Au contraire, la méthode SSR-L est plus précise pour un nombre faible de modes, mais cette précision augmente relativement moins avec le nombre de modes que pour BBR-O.

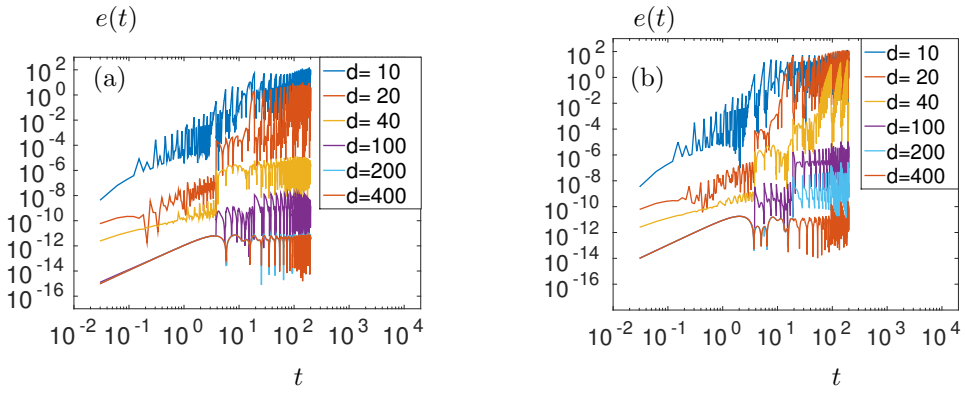



FIGURE 8.11 – Erreur résiduelle pour (a) le Test 3 par BBR-O, et (b) le Test 4, pour $d$ variable et $\Delta t = 0.0051$.

TABLE 8.2 – Erreur résiduelle $E_d$ définie par (8.48) en fonction de $d$ pour les Tests 1, 2, 3 et 4

| | $d$ | 10 | 20 | 40 | 100 | 200 | 400 |
|---|---|---|---|---|---|---|---|
| Test 1 | BBR-O $\Delta t = 0.0051$ | 210 | 24 | $1.2 \cdot 10^{-5}$ | $1.8 \cdot 10^{-8}$ | $1.4 \cdot 10^{-11}$ | $3.6 \cdot 10^{-13}$ |
| | SSR-L $\Delta t = 0.0051$ | $1.8 \cdot 10^{-1}$ | $1.2 \cdot 10^{-1}$ | $8.5 \cdot 10^{-2}$ | $4.3 \cdot 10^{-2}$ | $2.3 \cdot 10^{-2}$ | $1.1 \cdot 10^{-3}$ |
| Test 2 | BBR-O $\Delta t = 0.0051$ | 620 | 230 | 43 | $1.1 \cdot 10^{-5}$ | $3.8 \cdot 10^{-7}$ | $1.9 \cdot 10^{-9}$ |
| | SSR-L $\Delta t = 0.0051$ | $2.3 \cdot 10^{-1}$ | $1.7 \cdot 10^{-1}$ | $1.5 \cdot 10^{-1}$ | $9.2 \cdot 10^{-2}$ | $3.4 \cdot 10^{-2}$ | $4.1 \cdot 10^{-3}$ |
| Test 3 | BBR-O $\Delta t = 0.0051$ | 150 | 9.5 | $1.2 \cdot 10^{-5}$ | $1.6 \cdot 10^{-8}$ | $9.9 \cdot 10^{-12}$ | $7.5 \cdot 10^{-12}$ |
| | SSR-L $\Delta t = 0.0051$ | 0.16 | 0.11 | $7.5 \cdot 10^{-2}$ | $9.6 \cdot 10^{-3}$ | $5 \cdot 10^{-3}$ | $1.1 \cdot 10^{-4}$ |
| Test 4 | BBR-O $\Delta t = 0.0051$ | 130 | 120 | 22 | $1.1 \cdot 10^{-5}$ | $2.8 \cdot 10^{-7}$ | $2.6 \cdot 10^{-10}$ |
| | SSR-L $\Delta t = 0.0051$ | 0.13 | $5.3 \cdot 10^{-2}$ | $1.2 \cdot 10^{-1}$ | $2.4 \cdot 10^{-2}$ | $8.4 \cdot 10^{-3}$ | $2.8 \cdot 10^{-4}$ |

### 8.6.4 Simulation de l'équation de Peierls-Nabarro Dynamique

Nous comparons maintenant les différentes stratégies dans des cas particuliers de l'équation non-linéaire complète (8.1).

**Test numérique 5** (Poinçonnement sinusoïdal en temps en mode II)**.** *On fixe* $i = \mathrm{II}$ *et les constantes*

$$\alpha = 0.1, \qquad \mu = 1, \qquad \gamma = \sqrt{3} \simeq 1.731. \tag{8.49}$$

*On se donne un potentiel satisfaisant*

$$F'(u) = \sin(2\pi u). \tag{8.50}$$

*Soit une donnée initiale*

$$\sigma_0^{\mathrm{a}} = 0, \quad \textit{et} \quad \eta_0(t,x) = \eta_0(x), \tag{8.51}$$

*où* $\eta_0(x)$ *est la solution de* (6.16) *associée à* $\sigma_0^{\mathrm{a}} = 0$*, avec*

$$\eta_0(+\infty) = 0 \quad \textit{et} \quad \eta_0(-\infty) = 0, \tag{8.52}$$

*qui satisfait* $\eta_0(-4) = 1/2$*. On impose un chargement*

$$\sigma^{\mathrm{a}}(t > 0, x) = 0.9 \sin(t) \exp(-x^2/8), \tag{8.53}$$

*et on cherche la solution* $\eta$ *de* (8.1) *sur l'intervalle temporel* $t \in [0, 40]$*.*

**Test numérique 6** (Poinçonnement sinusoïdal en mode III)**.** *On fixe* $i = \mathrm{III}$*. On se donne des constantes* (8.49) *un potentiel satisfaisant* (8.50)*. Soit une donnée initiale* (8.51)*, où* $\eta_0(x)$ *est la solution de* (6.16) *associée à* $\sigma_0^{\mathrm{a}} = 0$*, avec* (8.52)*, qui satisfait* $\eta_0(-4) = 1/2$*. On impose un chargement* (8.53) *et on cherche la solution* $\eta$ *de* (6.13) *sur l'intervalle temporel* $t \in [0, 40]$*.*

Les paramètres de discrétisation spatiale sont :
— la taille de boîte $2L = 20$,
— le nombre de modes de Fourier $2m = 2^{10} = 1024$, d'où $h \simeq 0.195$.

A présent, la solution de référence est donc une solution de référence "à discrétisation spatiale fixée", et souffre donc du fait que l'on est sur une boîte de taille finie, avec un pas de discrétisation spatiale qui n'est pas infiniment petit.

Les Tests 5 et 6 sont dans un cas "favorable", dans le sens où :
— le forçage $\sigma^{\mathrm{a}}$ est régulier en temps et en espace, et quasiment local en espace (à cause de la décroissance très fort de la gaussienne),
— le forçage $\sigma^{\mathrm{a}}$ induit une fonction $\eta_{\mathrm{e}}$ elle aussi régulière en temps et en espace (comme $|\sigma^{\mathrm{a}}| < 1$, il n'y a pas saturation de (6.12)), et quasiment locale en espace.

*Remarque* 63 (Erreur due à la discrétisation spatiale). Les tests que nous effectuons sont d'abord à discrétisation spatiale fixée, avec les effets de périodisation décrits dans la Section 8.2.

*Remarque* 64 (Paramètre visco-plastique $\alpha$). Par souci de simplicité, on a toujours pris $\alpha = 0.1$ dans les expériences ci-dessus. Changer la valeur de ce paramètre ne semble pas changer nos conclusions qualitatives.

### Accord des algorithmes

Nous dessinons sur la Figure 8.12 la variation de $\eta(t,x)+\eta_{\mathrm{e}}(t,x)$ au fil du temps. On constate tout d'abord qu'on n'observe pas d'oscillations spatiales, qui pourraient être dues à un mauvais traitement des hauts modes de Fourier (voir Figure 8.12 et la Figure 8.13).

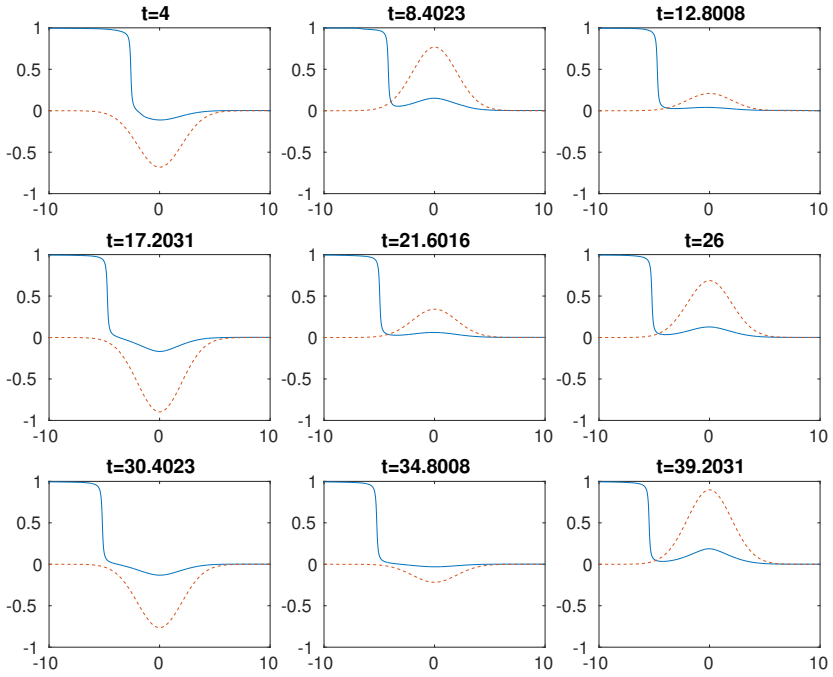



FIGURE 8.12 – Evolution de $\eta(t,x)+\eta_{\mathrm{e}}(t,x)$ en bleu, et de $\sigma^{\mathrm{a}}(t,x)$ en tirets rouges, pour le Test 6.

En testant les différentes stratégies, on constate qu'elles fonctionnent à divers degrés d'efficacité et de précision, mais convergent vers une même solution quand le pas de temps tend vers 0 –aux erreurs dues au noyau près (en ce qui concerne les algorithmes BBR-O et LSR-O, ces erreurs sont très petites pour le paramètre $d$ employé). On peut ainsi étudier l'erreur des différentes méthodes par rapport à une solution de référence numérique (ici, elle est fournie par BBD-A, pour $\Delta t/h = 10^{-1}$ voir Figure 8.14). Dans le cas du Test 6, il est notable que toutes les statégies sont stables pour tous les pas de temps, sauf LRD-, qui nécesssite un petit pas de temps pour être stable. Au contraire, pour le Test 5, seules les stratégies basées sur la forme résolue sont stables pour tous les pas de temps. C'est cohérent avec les Tests 1 et 2 effectués précédemment.

Sans surprise, on constate sur la Figure 8.14 que les schémas sont d'autant plus efficaces lorsque le pas de temps tend vers 0 qu'elles sont d'ordre élevé. En outre, pour différentes stratégies d'ordre 2 en temps, l'algorithme LSR-O est plus efficace que LRD-, puis que les algorithmes SSD-L, SSR-L. Cette remarque est cohérente avec la Figure 8.10(a).

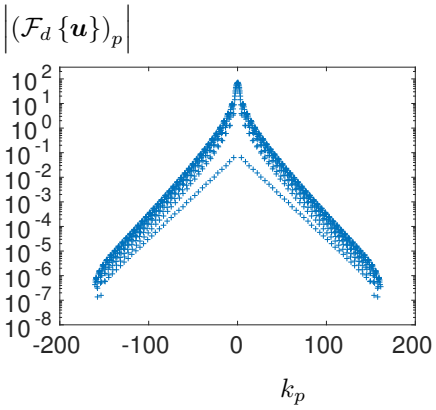



FIGURE 8.13 – Importance des différents modes de Fourier discrets $|\mathcal{F}_d\{\boldsymbol{u}(T=40)\}|$ dans le cadre du Test 6, où on utilise BBD-A avec $\Delta t = 0.002$.

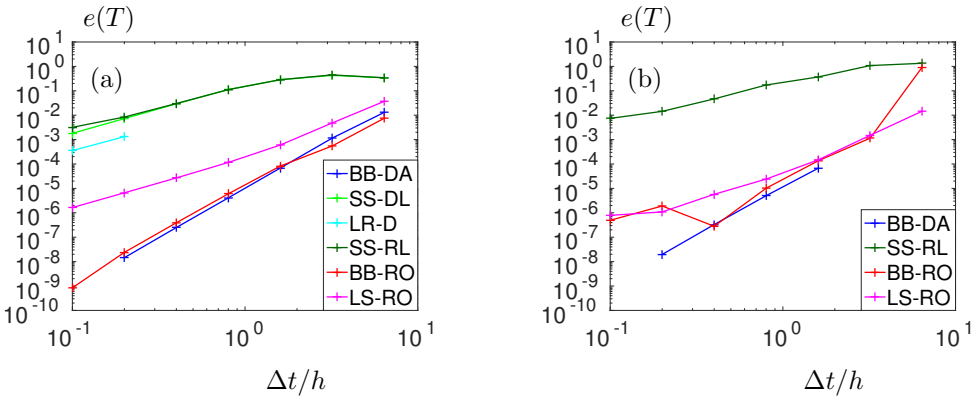



FIGURE 8.14 – Erreurs $e(T)$ des différentes méthodes pour (a) le Test 6 et (b) le Test 5, en fonction de $\Delta t/h$. Ici, pour SSD-L, on a $d+1 = 100$, pour SSR-L, on a $d+1 = 250$, pour BBR-O et LSR-O, on a $d = 150$ (les points absents à *droite* des lignes sont dus à la non-convergence des algorithmes ; LRD- et SSD-L n'ont pas été testés dans le cas (b)).

En ce qui concerne les temps d'exécution, on observe empiriquement que les algorithmes LRD- et SSR-L et BBD-A sont les plus rapides, les algorithmes BBR-O et LSR-O venant ensuite (voir Figure 8.15). Les algorithmes BBD- et LRD- avec une méthode de calcul naïve voient leurs temps d'exécution augmenter plus rapidement que les algorithmes accélérés ou oublieux.

Une régression linéaire indique que, pour l'ensemble des stratégies hors LRD-, le temps d'exécution est environ proportionnel au nombre de pas d'itération $N$, ce qui est cohérent avec la théorie (voir Table 8.3 plus bas). Au contraire, les algorithmes LRD- (que nous n'avons implémenté qu'avec une méthode de calcul naïve) et BBD- voient leurs temps d'exé-

cution augmenter plus rapidement que les autres algorithmes lorsque $N \to +\infty$. En théorie, cette dernière devrait augmenter en $N^2$, on observe sur cet exemple une augmentation empirique en $N^{1.6}$, respectivement $N^{1.8}$.

Cela démontre qu'il est très intéressant de recourir à la méthode accélérée décrite dans la Section 8.4 pour effectuer les calculs. Sur les exemples des Tests 5 et 6, on obtient un facteur d'accélération de 15 à 18 lorsqu'on effectue 10000 itérations. En revanche, l'intérêt des méthodes oublieuses est moins évident, pour le code et l'ordinateur utilisés. En effet, on n'atteint pas un régime dans lequel elles sont vraiment plus économes en mémoire que les méthodes naïves ou accélérées.

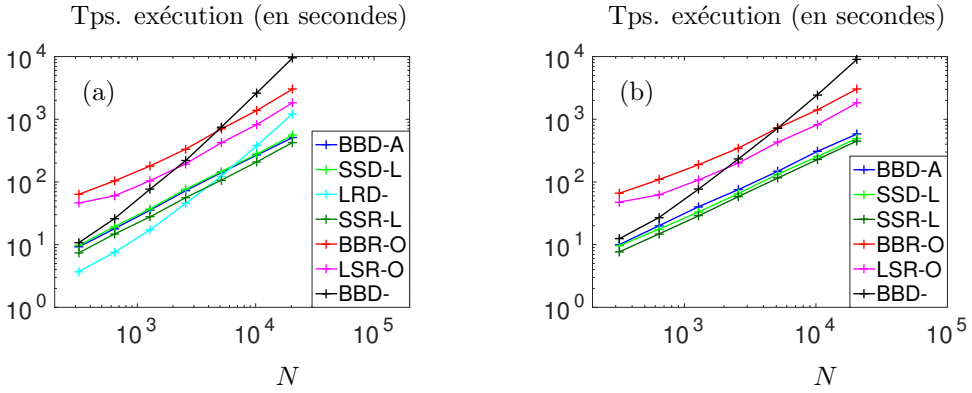



FIGURE 8.15 – Temps d'exécution (en secondes) des différentes méthodes pour (a) le Test 6 et (b) le Test 5, en fonction du nombre d'itérations $N$. Ici, pour SSD-L, on a $d+1 = 100$, pour SSR-L, on a $d+1 = 250$, pour BBR-O et LSR-O, on a $d = 150$. Les expériences ont été faites avec un code MATLAB, sur un PC portable avec 4 processeurs de 2.30GHz, et une mémoire de 7.7GiO.

Commentons deux points particuliers :

— l'algorithme LRD- est initialement plus rapide que tous les autres algorithmes car il n'utilise pas de pas de temps intermédiaire. Aussi doit-il stocker 2 fois moins de mémoire. Au contraire, l'algorithme BBD- utilise un pas de temps intermédiaire. Par ailleurs, il nécessite d'évaluer 2 grandes sommes au lieu d'une seule. Ainsi est-il (approximativement) 8 fois plus lent que l'algorithme LRD-, à nombre d'itérations fixés, ce qu'on observe asymptotiquement sur la Figure 8.15(a) (pour 10000 itérations, ce rapport est de 7.8),

— les méthodes oublieuses sont sous-optimales pour les petits pas de temps. Par exemple, ici, pour les 5 premiers pas de temps, la méthode oublieuse génère un contour de 150 points, et fait donc évoluer 150 variables à chaque pas de temps, au lieu de n'en conserver que 5 (on peut faire le même commentaire pour les 25 et 125 variables suivantes). Par simplicité, on n'a pas optimisé cela. C'est une des raisons pour lesquelles elles semblent ici si peu avantageuses.

*Remarque* 65. Le langage MATLAB est beaucoup plus rapide dans l'exécution de calculs

vectoriels que dans l'exécution de boucles, et notre code n'a été optimisé que dans une limite raisonnable. Aussi, dans notre code, des méthodes de calcul algorithmiquement plus subtiles, telles que la méthode d'accélération, et celle d'inversion de la transformée de Laplace, sont artificiellement ralenties par rapport à des stratégies plus simples telles que la méthode de calcul naïve, ou celle reposant sur l'utilisation des polynômes de Laguerre. Cependant, nous pensons que, à stratégie donnée, l'augmentation relative du temps d'exécution est bien représentative.

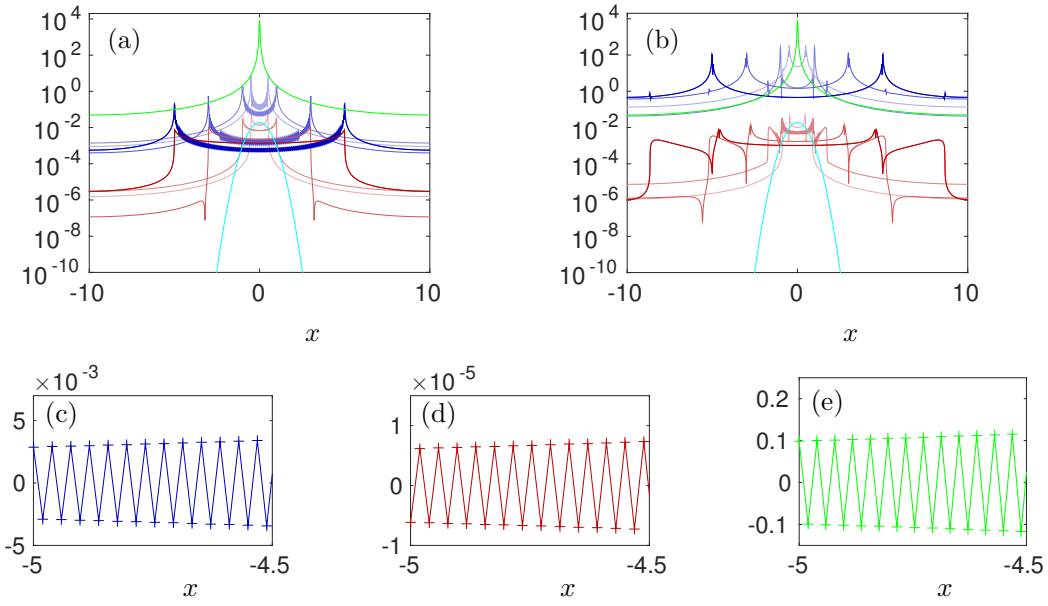



FIGURE 8.16 – Valeur absolue des noyaux de convolution discrétisés dans l'espace physique : (a) $|\mathcal{F}^{-1}\left\{(\kappa_{\mathrm{III}}^{\alpha})^{-1}\,|k|^2 C_{\mathrm{III}}\right\}(|k|t)|(x)$ (en bleu) et $|\mathcal{F}^{-1}\left\{\mathfrak{R}_{\mathrm{III}}^{\alpha}\right\}(t,x)|$ (en rouge) et (b) $|\mathcal{F}^{-1}\left\{(\kappa_{\mathrm{II}}^{\alpha})^{-1}\,k^2 C_{\mathrm{II}}(|k|t)\right\}(x)|$ (en bleu) et $|\mathcal{F}^{-1}\left\{\mathfrak{R}_{\mathrm{II}}^{\alpha}\right\}(t,x)|$ (en rouge). A titre de référence, on fait figurer en vert, et en cyan, sur (a) et (b) le noyau de convolution discrétisé correspondant au laplacien, c'est à dire $|\mathcal{F}_d^{-1}\left\{k_p^2\right\}|$ et $|\mathcal{F}_d^{-1}\left\{\exp(-1/12 * k_p^2)\right\}|$. Zooms (c) sur $\mathcal{F}^{-1}\left\{(\kappa_{\mathrm{III}}^{\alpha})^{-1}\,k^2 C_{\mathrm{III}}(k)\right\}(x)$, (d) sur $\mathcal{F}^{-1}\left\{\mathfrak{R}_{\mathrm{III}}^{\alpha}\right\}(t=1,x)$ et (e) sur $\mathcal{F}_d^{-1}\left\{k_p^2\right\}$. Les différentes courbes figurent les noyaux aux temps $t \in \{0.5, 1, 3, 5\}$, dans un dégradé de clair vers foncé pendant que $t$ croît. Ici, $\gamma = \sqrt{3}$, $\alpha = 0.1$, et $2L = 20$, $h = 0.0195$ ; on utilise $d = 150$ modes pour l'inversion de la transformation de Laplace.

### Discrétisation spatiale des noyaux

La discrétisation spatiale choisie revient à effectuer des convolutions discrètes en utilisant des noyaux discrétisés (voir Section 8.2). Nous traçons ces derniers sur la Figure 8.16.

Les noyaux spatio-temporels considérés $\mathcal{F}^{-1}\{k^2 C_i(|k|t)\}(x)$ sont à support compact en

espace à un temps $t$ fixé (voir Section 7.2). Nous pensons que les noyaux $\mathcal{F}^{-1}\left\{\mathfrak{R}_i^\alpha(|k|t)\right\}(x)$ le sont aussi (à cause de la nature hyperbolique du système d'équations inital, voir Chapitre 6) ; a minima, ils doivent être d'autant plus piqués lorsque $t$ est petit. Or, on remarque sur les Figures 8.16(a) et (b) que ce n'est pas exactement le cas pour les discrétisations spatiales des noyaux de convolution. Celles-ci sont certes de faible valeur, mais cependant éloignées de 0 en valeur absolue dans les zones où les noyaux originaux sont nuls (on a vérifié que ces effets ne sont ni dus à l'erreur machine, ni à un trop faible nombre de modes pour l'inversion de la transformation de Laplace). C'est un trait caractéristique des méthodes spectrales basées sur la transformation de Fourier. En effet, la discrétisation classique du laplacien via sa transformée de Fourier $\mathcal{F}\{\Delta\}(k) = k^2$ donne des résultats similaires (mais pas pour la gaussienne, résolvante associée au laplacien). Le caractère local des opérateurs se traduit néanmoins par des oscillations violentes autour de 0 à l'échelle de la discrétisation spatiale $h$ (voir Figures 8.16(a), (b) et (c)).

**Influence des paramètres de discrétisation spatiale sur la solution globale**

Effectuons maintenant une dernière expérience sur la base du Test 5, afin d'observer l'influence des deux paramètres de discrétisation spatiale (le pas $h$ ; et la largeur de boîte $L$). Ceux-ci jouent des rôles très différents : le premier pilote la précision *locale* de la méthode, tandis que le second pilote la précision *globale* de la méthode.

La méthode que nous employons est fondée sur l'utilisation de la transformation de Fourier : elle converge donc très rapidement (les fonctions étudiées étant régulières) lorsque le pas de discrétisation tend vers 0. On le constate sur la Figure 8.17(a), où on atteint une erreur de l'ordre de $10^{-7}$ avec un pas de discrétisation spatial de $1.95 \cdot 10^{-2}$.

Rappelons que la discrétisation que nous avons choisie périodise naturellement le problème, qui est initialement posé sur la droite réelle. Naturellement, cela induit des effets de bord. Mais, en cas de chargement $\sigma^{\mathrm{a}}(t,x)$ à support compact en $x$ (comme c'est quasiment le cas pour le Test 5), la fonction $\eta(t,x)$ n'est sensiblement affectée en un point éloigné du support du chargement qu'au bout d'un certain temps. On constate que la discrétisation que nous employons préserve cette propriété dans une large mesure (voir Figure 8.18). Néanmoins, à largeur $2L$ de boîte fixé, il arrive nécessairement un instant à partir duquel les effets de bord se font ressentir. Toutefois, tant que la boîte est suffisamment large, ceux-ci sont négligeables (voir Figure 8.17(b)).

Cela peut se révéler dommageable dans le cas où le chargement imposé à la dislocation n'est *pas* local, par exemple dans le cas d'un chargement sous la forme d'un front progressif, ou d'un chargement violent. Nous proposons un test où le chargement est régulier en temps est constant en espace (donc non-local) :

**Test numérique 7** (Mise en mouvement de dislocations de type $i = \mathrm{II}$)**.** *On fixe* $i = \mathrm{II}$ *et les constantes* (8.49)*. On se donne un potentiel satisfaisant* (8.50)*. Soit une donnée initiale*

$$\sigma_0^{\mathrm{a}} = 0, \quad \textit{et} \quad \eta_0(t,x) = \eta_0(x),$$

*où* $\eta_0(x)$ *est la solution de* (6.16) *associée à* $\sigma_0^{\mathrm{a}} = 0$*, avec*

$$\eta_0(+\infty) = 0 \quad \textit{et} \quad \eta_0(-\infty) = 0,$$

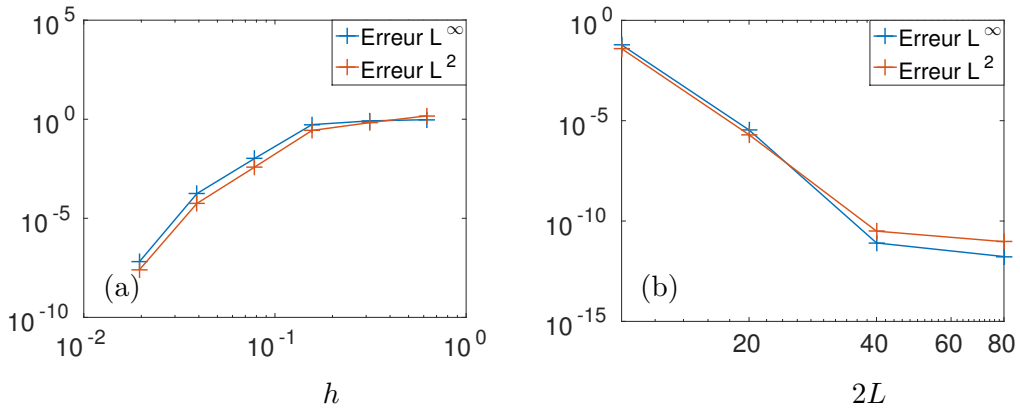



FIGURE 8.17 – Erreur sur $\eta(5,\cdot)$ pour le Test 5. Toutes les expériences sont réalisées avec BBD-A, avec $\Delta t = 1 \cdot 10^{-3}$. Dans le cas (a), on a pris $2L = 40$ et la solution de référence est calculée avec $h = 0.0098$, et dans le cas (b), on a pris $h = 0.0195$, et la solution de référence est calculée avec $2L = 160$.

*qui satisfait* $\eta_0(0) = 1/2$. *On impose un chargement*

$$\sigma^{\mathrm{a}}(t > 0, x) = 0.5 \sin(t),$$

*et on cherche la solution* $\eta$ *de* (6.13) *sur l'intervalle temporel* $t \in [0, 5]$.

On compare alors les erreurs commises sur la solution $\eta(t, x)$, pour BBD-A, à pas spatial et temporel fixés, et en faisant varier les largeurs de boîtes $2L$. Contrairement aux cas précédents, où l'effet de périodisation n'avait pas d'impact sensible, cet effet apparaît maintenant. En effet, comme le montre la Figure 8.19(a), la solution converge maintenant très lentement en la taille de la boîte (empiriquement, l'erreur est proportionnelle à $L^{-1}$, dès lors que $L$ est suffisamment grand pour que les ondes émises par le coeur de la dislocation n'atteignent pas le bord). En fait, l'erreur se propage à partir du bord du domaine, à cause de la périodisation induite par la discrétisation des opérateurs.

*Remarque* 66. Le zero-padding (voir [134, Chap. 12 p. 624]), qui revient à imposer artificiellement $\eta(t, x)$ sur $[-L, -L/2] \cup [L/2, L]$ peut être une première réponse face à ces effets indésirables. Néanmoins, cela revient à gommer l'effet des contributions à longue portée sur le coeur de la boîte de simulation, et induit par conséquent un autre type d'erreur. Nous n'avons pas étudié cette stratégie numérique.

## 8.7 Comparaison des algorithmes

Après avoir effectué les tests de la Section 8.6, et munis des considérations théoriques des Sections 8.1, 8.2, 8.3, 8.4 et 8.5, nous pouvons maintenant évaluer dans une certaine mesure l'intérêt de chaque algorithme. Les critères sont les suivants :

1. la stabilité ;

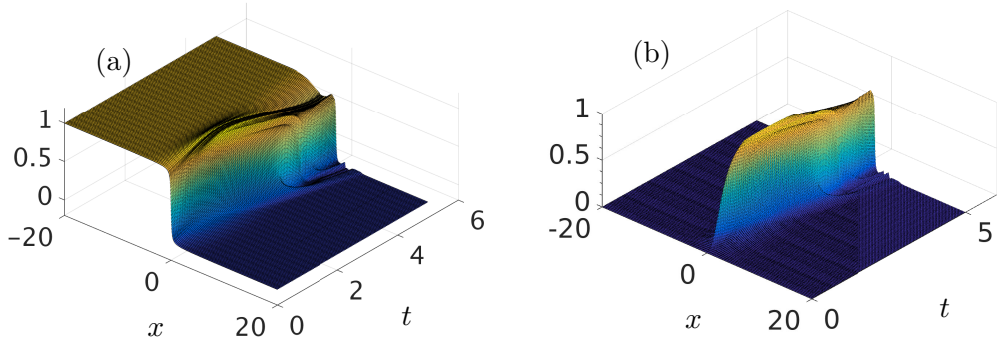



FIGURE 8.18 – Evolution de (a) $\eta_e(t,x) + \eta(t,x)$ et (b) $|\eta(t,x) - \eta_{\text{ref}}(t,x)|$ de en fonction de $t \in [0,5]$ et de $x \in [-20,20]$ pour le Test 5. La fonction $\eta$ est calculée grâce à BBD-A, pour $2L = 40$, $h = 0.0195$, $\Delta t = 1 \cdot 10^{-3}$ ,et $\eta_{\text{ref}}$ est une solution numérique de référence calculée avec BBD-A, pour $2L = 40$, $h = 0.0098$, $\Delta t = 1 \cdot 10^{-3}$.

2. la précision ;
3. la rapidité d'exécution et la mémoire utilisée.

Toutefois, comme pointé dans la Section 8.6.4, les tests de rapidité d'exécution des différents algorithmes sont dépendants des spécificités du langage informatique et de l'ordinateur utilisés. En outre, on n'a pas mesuré la mémoire concrètement utilisée par la machine.

D'après la Section 8.1, on peut caractériser a priori l'ordre, la complexité et le coût en mémoire des algorithmes considérés. Nous résumons ces considérations dans le Tableau 8.3, où on note les différents paramètres :

— $N$ : nombre de pas de temps,
— $d$ : nombre de modes pour approximer le noyau $C_i$ (ou la résolvante $\mathfrak{R}_i^\alpha$),
— $E_d$ : erreur due au fait que le noyau est approximé (cette erreur dépend éventuellement de $\Delta t$, mais ne tend pas vers 0 lorsque $\Delta t \to 0$),
— $P$ : nombre d'itérations de point fixe pour résoudre une éventuelle partie implicite non-linéaire de l'algorithme.
— $2m$ : nombre de points de discrétisation spatiale.

On y rajoute la stabilité observée empiriquement ("Cond." signifie "conditionnellement stable" et "Incond." signifie "inconditionnellement instable").

### 8.7.1 Forme directe et forme résolue

Comme les arguments théoriques le suggéraient (voir Section 7.3.2), un schéma simulant la forme résolue (8.2b) est en général stable, tandis qu'un schéma simulant la forme directe (8.2a) est potentiellement instable. Plus précisément, tous les algorithmes que nous avons testés et qui reposent sur (8.2b) sont stables, tandis que tous les algorithmes reposant sur la forme directe (8.2a) souffrent d'instabilité. Cette instabilité est :

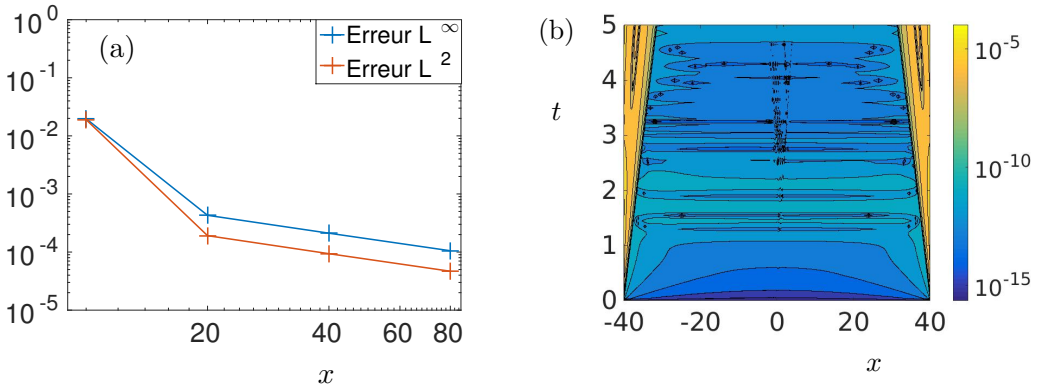



FIGURE 8.19 – (a) Erreur sur $\eta(5,x)$ en fonction de $L$ (b) Erreur en fonction de $(t,x) \in [0,5] \times [-40,40]$ sur $\eta(t,x)$ pour BBD-A, dans le cas où $L = 80$, $h = 0.0195$.

1. soit une instabilité inconditionnelle en temps long en ce qui concerne SSD-L (empiriquement, on observe une instabilité inconditionnelle pour les mode II et III) ;
2. soit une instabilité conditionnelle en $\Delta t/h$ en ce qui concerne LRD- et BBD-A. Il semble cependant que le schéma BBD-A soit inconditionnellement stable pour le mode III.

A cause du rescaling en $1/|k|$ de l'équation (6.13), le premier type d'instabilité n'est pas tolérable. Le second type d'instabilité contraint le rapport $\Delta t/h$ à être petit. Dans notre cas, il n'y a en général pas beaucoup d'intérêt à prendre $h$ très petit, à cause du fait que la méthode de discrétisation spatiale converge très rapidement lors $h$ tend vers 0 (voir Section 8.6.4). Par ailleurs, en pratique, il peut être nécessaire de prendre $\Delta t \ll 1$ à cause du forçage exogène (et de la non-linéarité en général $f_{\sigma^a}$). Ainsi, une stabilité conditionnelle en $\Delta t/h$ n'est pas nécessairement handicapante.

En conclusion, il est préférable d'utiliser un schéma reposant sur la forme résolue (8.2b), indépendamment de la méthode de calcul utilisée, si l'on souhaite obtenir de la stabilité inconditionnelle. Mais les schémas reposant sur (8.2a) peuvent aussi être compétitifs dans certains cas.

### 8.7.2 Précision et ordre du schéma

Ayant tranché la question de la stabilité, tournons-nous maintenant vers la question de l'ordre (en temps) du schéma utilisé. Il semble préférable d'utiliser un schéma d'ordre élevé (ici 4). Toutefois, ce point est à tempérer par les remarques suivantes :

1. le schéma BBR- (d'ordre 4) est moins précis que le schéma LSR- (d'ordre 2) lorsque le pas de temps est grand. Cela pourrait être dommageable ;
2. le forçage induit par $\sigma^{\mathrm{a}}$ sur (6.13) est potentiellement peu régulier en temps, ce qui rend l'utilisation d'un schéma d'ordre élevé moins utile.

TABLE 8.3 – Caractéristiques des algorithmes

| Algorithme | Coût de calcul | Mémoire | Précision | Stabilité empirique |
|---|---|---|---|---|
| LRD- | $O\left(m \log mN^2\right)$ | $O(mN)$ | $O(\Delta t^2)$ | Cond. |
| BBD-A | $O\left(m \log mN (\log N)^2\right)$ $+O(mNP)$ | $O(mN)$ | $O(\Delta t^4)$ | Cond. |
| SSD-L | $O(m \log md \log dN)$ | $O(md)$ | $O(\Delta t^2)+ E_d$ | Instable |
| SSR-L | $O(m \log md \log dN)$ | $O(md)$ | $O(\Delta t^2)+ E_d$ | Incond. |
| LSR-O | $O(m \log mdN \log N)$ $+O(mNP)$ | $O(md \log N)$ | $O(\Delta t^2)+ E_d$ | Incond. |
| BBR-O | $O(m \log mdN \log N)$ $+O(mNP)$ | $O(md \log N)$ | $O(\Delta t^4)+ E_d$ | Incond. |

Ainsi, de tous les schémas que nous avons testés, le schéma BBR- semble être le plus efficace, à la fois stable et précis, sachant que, dans le cas du mode II, le schéma LSR permet d'avoir des meilleurs résultats lorsque le pas de temps est grand.

D'autres schémas –que nous n'avons pas testés– peuvent être employés. Néanmoins, si l'on souhaite discrétiser la forme résolue (8.2b), il faut nécessairement utiliser un schéma A-stable (voir [14]). On peut par exemple utiliser des méthodes multi-pas comme les *backward differential formulas* (ou BDF, voir [78, Chap. III.1 p. 356]). Mais, dès lors que l'on souhaite avoir une méthode d'ordre strictement supérieur à 2, il n'est plus possible de recourir à une méthode multi-pas (à cause de la barrière de Dahlquist, voir [79, Th. 1.4 p. 247]) ; aussi, des méthodes de Runge-Kutta sont étudiées dans [13,14]. En particulier, nous avons implémenté la méthode RadauIIA d'ordre 5 recommandée par [13], qui donne des résultats meilleurs que BBR-A, étant d'ordre 5, toutes choses étant égales du point de vue qualitatif (non montré sur les figures).

Enfin, n'occultons pas un aspect qui peut être coûteux dans certains cas : à savoir la résolution d'une équation *implicite* non-linéaire induite par les schémas choisis. Si jamais ce coût devenait prohibitif, on pourrait alors re-considérer la question sous un autre angle. Par ailleurs, on pourrait sans doute accélerer cette étape en utilisant une méthode plus efficace que le point-fixe utilisé ici. Notons que, dans les Tests 5 et 6, l'étape de résolution du système non-linéaire prend un temps non-négligeable.

### 8.7.3 Méthodes oublieuses et méthodes accélérées : précision et efficacité

Discutons enfin le choix de la méthode de calcul. Nous avons considéré la méthode d'accélération (Section 8.4), et la méthode oublieuse basée sur des polynômes de Laguerre de la Section 8.5.6, et la méthode oublieuse de Lubich et Schädle basée sur l'inversion de la transformée de Laplace (voir Section 8.5.7). Deux critères doivent être considérés : la rapidité d'exécution et la mémoire nécessaire.

Les méthodes reposant sur l'utilisation de polynômes de Laguerre sont rapides, et économes en mémoire, à nombre de mode $d$ fixé (voir Table 8.3). Mais leur précision est mauvaise. Aussi ne doivent-elles être utilisées, éventuellement, que pour effectuer des précalculs et des étalonnages. Nous les écartons donc de la discussion ci-dessous.

Sur les expériences effectuées, la méthode accélérée est la plus avantageuse en termes de temps d'exécution : elle diminue d'un ordre de grandeur le temps de simulation pour 10000 itérations par rapport à une méthode naïve. Si l'on se place dans un cadre où le pas de temps et constant et où chaque mode de Fourier est mis à jour à chaque pas de temps, quel que soit l'algorithme utilisé, on aura un coût en temps minimal en $O(mN)$. Donc, l'algorithme reposant sur la méthode accélérée est quasi-optimal en termes de temps de calcul (à des termes logarithmiques près), car il a un coût en $O(m \log mN (\log N)^2)$ (le terme $m \log m$ est dû à l'utilisation de la FFT pour la transformation de Fourier spatiale).

La méthode accélérée est empiriquement plus rapide que la méthode oublieuse de Lubich et Schädle de la Section 8.5.7. Cela se justifie formellement en considérant que la méthode accélérée a une complexité en $O(m \log mN (\log N)^2)$ tandis que la méthode oublieuse de Lubich et Schädle a une complexité en $O(m \log mdN \log N / \log B)$. On a donc deux régimes différents grossièrement délimités par $N_0 = \exp(d/\log B)$. Ainsi, la méthode accélérée est plus rapide que la méthode oublieuse de Lubich et Schädle si $N \ll N_0$, et elle est plus lente si $N \gg N_0$. Or, si $B = 5$, on constate que pour avoir une erreur petite (voir Table 8.2), il faut environ $d = 150$ modes pour avoir une bonne précision lorsque l'on emploie la méthode oublieuse. Or, cela donne $N_0 \simeq 10^{13}$. L'énormité du chiffre justifie donc que la méthode accélérée peut être considérée comme la méthode la plus rapide (notons que ce ne sont que des arguments formels, car on ne compare que des ordres de grandeur, où les constantes jouent un rôle crucial). Empiriquement, les méthodes oublieuses se révèlent plus lentes que la méthode accélérée sur les tests effectués, et ne sont compétitives par rapport à un méthode naïve qu'au-delà de 1000 itérations.

La méthode d'accélération est infiniment précise, tandis que les deux autres méthodes de calcul souffrent d'un terme d'erreur dépendant du nombre de modes $d$ constituant l'approximation du noyau. Mais l'erreur $E_d$ inhérente à la méthode de calcul décroît très rapidement pour la méthode oublieuse de Lubich et Schädle (voir Table 8.2) et devient infime lorsque $d$ est de l'ordre de plusieurs centaines.

En se référant à la Table 8.3, on constate par ailleurs que la méthode d'accélération est aussi gourmande en mémoire qu'une méthode naïve. Au contraire, la méthode Lubich et Schädle est plus économe. Par un argument formel, on constate qu'elle devient plus intéressante (dans les cas étudiés) lorsque $N >> N_0$, pour $d \log N_0 = N_0$, c'est à dire $N_0$ de l'ordre de 1000 (pour $d = 100$).

*Remarque* 67 (Souplesse). Il existe une différence d'un autre type entre la méthode accélérée et la méthode oublieuse de Lubich et Schädle. La première semble liée à l'utilisation d'un pas de discrétisation constant, ce qui peut être prohibitif en pratique. Au contraire, la méthode oublieuse de Lubich et Schädle a été généralisée à un pas adaptatif (voir [106]).

Bien que cela ne soit pas un critère purement scientifique, soulignons que la méthode de Lubich et Schädle est plus délicate à implémenter que la méthode accélérée.

### 8.7.4 Brève discussion sur la méthode LRD-

Nous n'avons trouvé aucune raison théorique ni pratique de préférer le schéma LRD- aux schémas BBD- et BBR-. En effet, ces derniers sont d'ordre supérieur (à savoir 4, versus 2) pour un coût légèrement supérieur en mémoire et en temps d'exécution (à un facteur multiplicatif constant près). Empiriquement, sur les cas testés, la zone de stabilité en $\Delta t$ de BBD- est en outre plus large que celle de LRD-.

Pour tempérer ce constat, il faut cependant souligner que l'article [97] propose des améliorations techniques :

1. l'algorithme LRD- est fait originellement pour gérer des pas de temps très variables. En effet, les problématiques rencontrées par les géophysiciens font qu'il est parfois nécessaire de faire varier le pas de temps de plusieurs ordres de grandeur ;
2. dans [97], il est suggéré un traitement différencié de la mémoire selon le mode de Fourier $|k|$ considéré, lorsque l'on cherche à résoudre l'équation complète (8.1). Notamment, tirant parti de la décroissance en $t^{-3/2}$ du noyau $C(t)$, les auteurs tronquent l'évaluation de l'intégrale de convolution à un temps proportionnel à $(T_{\mathrm{f}}|k|)^{-3/2}$ suivant les modes de Fourier $|k|$. Cela permet de réduire significativement la mémoire utilisée.

Autant la première amélioration nous semble superflue dans notre cas, autant il est possible que la seconde permette de rendre l'algorithme plus rapide sans trop perdre en précision. Toutefois, la difficulté informatique engendrée par le fait de gérer des tableaux dont la taille des colonnes est variable et l'absence d'analyse numérique sur l'erreur commise nous ont conduit à reporter *sine die* l'implémentation d'une telle amélioration.

### 8.7.5 Conclusion

Nous avons construit plusieurs algorithmes simulant l'équation de Peierls-Nabarro Dynamique, et démontré par l'exemple qu'il était possible de la simuler sur un ordinateur personnel.

Concernant la discrétisation temporelle, on peut tout d'abord utiliser les algorithmes SSR-L pour obtenir rapidement des résultats grossiers. Toutefois, la méthode oublieuse reposant sur les polynômes de Laguerre est relativement peu efficace pour ce problème. Quand on souhaite obtenir des résultats précis, on peut utiliser l'algorithme BBR-O avec des coûts en temps et en mémoire accrus. Si l'on n'est pas limité par la mémoire de la machine, et si $\Delta t < h$, l'algorithme BBD-A est une excellente alternative.

Une autre possibilité [1] consisterait à hybrider les méthodes oublieuses et accélérées. On utiliserait la méthode de transformation inverse de Laplace afin de reconstruire la résolvante. Pour les premiers pas de temps (environ les 1000 premiers), on utiliserait une méthode accélérée. Puis, on entrerait dans le régime où la méthode oublieuse est réellement efficace, et on gérerait les pas de temps suivants grâce à cette dernière. Un schéma intéressant serait naturellement le schéma BBR-, ou un schéma de Runge-Kutta A-stable d'ordre élevé (comme RadauIIA, par exemple d'ordre 5). Ainsi, on aurait un algorithme stable, rapide, et relativement moins gourmand en mémoire qu'une méthode naïve. [2]

Le fait de pouvoir reconstruire la résolvante grâce à la transformation numérique de Laplace de [107] est un avantage appréciable lorsque le modèle sera enrichi. En particulier, cela pourra être exploité dans les cas suivants :

1. l'ajout d'un "terme de gradient" (c'est à dire d'un Laplacien) à l'équation de Peierls-Nabarro Dynamique, à la manière de [135],
2. la résolution de l'équation de Peierls-Nabarro dynamique vectorielle en milieu anisotrope (où les noyaux analytiques généralisant $C_i$ n'ont pas encore été obtenus).

La discrétisation spatiale que nous employons utilise le fait que la transformation de Fourier diagonalise spatialement l'équation (6.13), et converge rapidement lorsque $h$ tend vers 0. Mais elle périodise artificiellement la solution numérique, et nécessite donc de prendre de grandes tailles de boîte pour faire des simulations fiables. La taille de boîte doit être d'autant plus grande que le forçage $\sigma^{\mathrm{a}}$ est non-local, et que l'expérience est longue. De surcroît, cette discrétisation nécessite l'utilisation d'un pas spatial constant, ce qui rend les simulations sur des larges boîtes coûteuses.

L'équation de Peierls-Nabarro est une équation récente dans le domaine des dislocations dont la phénoménologie et les implications physiques ont été peu explorées. L'étape suivante de notre recherche consiste à employer les algorithmes ainsi construits pour explorer des situations physique d'intérêt, et en particulier :

1. les chocs, c'est à dire imposer un chargement $\sigma(t, x)$ sous la forme d'une onde progressive,
2. les nucléations, c'est à dire la création de paires de dislocations de signes opposés,
3. les régimes transitoires lors de la mise en mouvement d'une dislocation (à mettre en perspective avec les résultats de [131]).

Cela fait l'objet d'un travail en cours, en collaboration avec Yves-Patrick Pellegrini.

A titre d'illustration, on montre sur la Figure 8.20 une image de nucléation, réalisée grâce à l'algorithme BBD-A.

1. je remercie Lehel Banjai pour une discussion à ce sujet
2. une telle méthode a été implémentée après la soumission du manuscrit. Les performances sont effectivement légèrement meilleures que celle du schéma BBR-O, mais il semble que les gains de performances soient minimes sur un ordinateur portable, pour l'équation traitée.

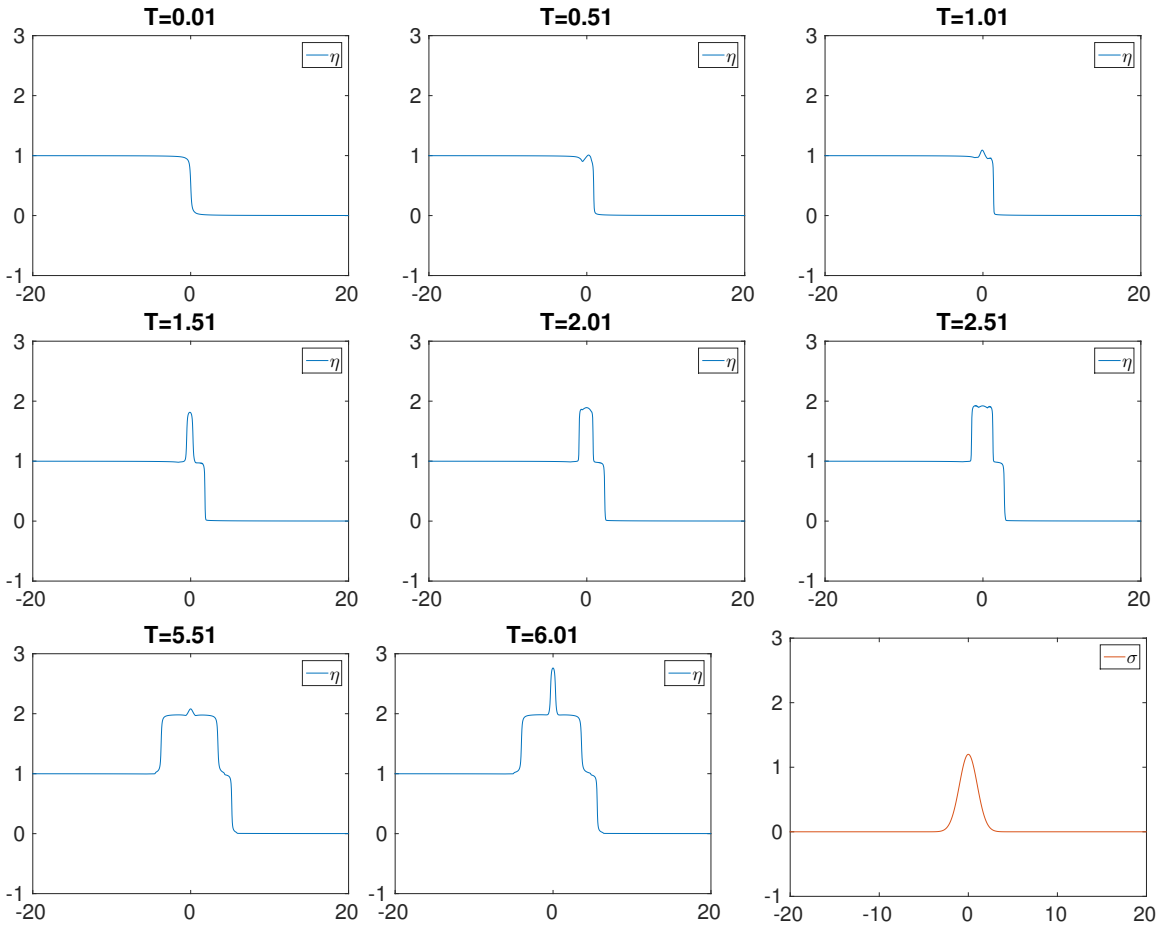



FIGURE 8.20 – Nucléation de dislocations en mode II, $\alpha = 0.01$, $\gamma = 2$, potentiel sinusoïdal. On soumet une dislocation au repose à un chargement $\sigma(x)$ brusque et plus fort que la contrainte limite en $x = 0$ (dessiné en rouge). Rapidement, de nouvelles dislocations se créent (c'est la nucléation) pour compenser la contrainte. Ces nouvelles dislocations s'éloignent ensuite dans des directions opposées.

# Chapitre 9

# Limite macroscopique d'un système de particules

Ce chapitre reprend l'article en anglais [22] coécrit avec Xavier Blanc.

Nous y étudions la limite macroscopique d'une chaîne de particules soumises à l'équation de Newton, dans le cas particulier où des chocs se propagent. Sous certaines hypothèses, nous démontrons que cette limite macroscopique n'est pas décrite par l'équation des ondes non-linéaire.

# From the Newton equation to the wave equation : the case of shock waves

Xavier Blanc [1], Marc Josien [2 3]

**Abstract** We study the macroscopic limit of a chain of atoms governed by the Newton equation. It is known from the work of Blanc, Le Bris, Lions, that this limit is the solution of a nonlinear wave equation, as long as this solution remains smooth. We show numerically and mathematically that, if the distances between particles remain bounded, it is not the case any more when there are shocks -at least for a convex nearest-neighbour interaction potential with convex derivative.

**Keywords** Newton equation, nonlinear wave equation, discrete-to-continuum limit

## 9.1 Introduction

**Motivation** We investigate here the macroscopic limit of the time-dependent Newton equation ruling the evolution of a set of particles at the microscopic scale. We perform our study in a simplified context : the particles form a one-dimension chain and we suppose that the interactions between the particles are nearest-neighbour interactions. It has been proven in [26] that, when the potential is convex, this system tends to a wave equation, provided that the solution of this wave equation is regular. However, non-linear wave equations are known to develop shocks in finite time. Our aim is to examine how this phenomenon impacts the convergence of Newton equations to wave equation.
Consider $2N+1$ particles, indexed by $j \in [\![-N, N]\!]$ and with positions $X_j$ which interact through the Newton equation, for $j \in [\![-N+1, N-1]\!]$ :

$$\frac{d^2}{dt^2} X_j(t) = W'\left(X_{j+1}(t) - X_j(t)\right) - W'\left(X_j(t) - X_{j-1}(t)\right), \tag{9.1}$$

where $W$ is the interaction potential. Throughout the article, we assume that $W$ is even. The initial and boundary conditions are :

$$X_{j+1}(0) - X_j(0) = N \int_{j/N}^{(j+1)/N} \phi_0^x(y)\, dy \quad \text{and} \quad \frac{d}{dt} X_j(0) = \phi_0^\tau\left(\frac{j}{N}\right), \tag{9.2}$$

$$X_{-N}(t) = N\phi_l \text{ and } X_N(t) = N\phi_r, \tag{9.3}$$

with functions $\phi_0^x, \phi_0^\tau$ and $\phi_l, \phi_r \in \mathbb{R}$ being compatible in the following sense :

$$\int_{-1}^{1} \phi_0^x(x)dx = \phi_r - \phi_l, \qquad \phi_0^\tau(-1) = 0, \qquad \text{and} \quad \phi_0^\tau(1) = 0. \tag{9.4}$$

1. Univ. Paris Diderot, Sorbonne Paris Cité, Laboratoire Jacques-Louis Lions, UMR 7598, UPMC, CNRS, F-75205 Paris, France
2. Université Paris-Est, Cermics (ENPC), F-77455 Marne-la-Vallée
3. INRIA Paris, 2 Rue Simone Iff, F-75012 Paris

We introduce the following rescaling :

$$t = N\tau, \qquad\qquad j = Nx.$$

The time $t$ is the microscopic time while $\tau$ is the macroscopic time. Then, the semi-discrete equation (9.1) is consistent with the wave equation :

$$\partial_\tau^2 \phi(\tau, x) = \partial_x \left[ W'(\partial_x \phi(\tau, x)) \right], \tag{9.5}$$

with initial and boundary conditions :

$$\partial_x \phi(\tau = 0, x) = \phi_0^x(x) \text{ and } \partial_\tau \phi(\tau = 0, x) = \phi_0^\tau(x), \tag{9.6}$$

$$\phi(\tau, -1) = \phi_l \text{ and } \phi(\tau, 1) = \phi_r. \tag{9.7}$$

*Remark* 3. It is worth pointing out that the natural variables in the hyperbolic system (9.5) are $\partial_\tau \phi$ and $\partial_x \phi$, in the sense that :

$$\partial_\tau \begin{pmatrix} \partial_\tau \phi \\ \partial_x \phi \end{pmatrix} = \partial_x \begin{pmatrix} W'(\partial_x \phi) \\ \partial_\tau \phi \end{pmatrix},$$

which is a $p$-system (see [139], p 127-131). We therefore introduce their discrete analogues :

$$U_j = X_{j+1} - X_j \qquad \text{and} \qquad V_j = \frac{dX_j}{dt}. \tag{9.8}$$

*Remark* 4 (About inversion). One could a priori think that (9.1) may lead to some inversions of atom positions, especially when shocks occur (see [35]). Put differently, one could have $X_{j+1}(t) < X_j(t)$ for certain $t$ and $j$, even if $X_j(t = 0)$ was increasing. This would question the physical relevance of (9.1), for the $j$-th particle is supposed to interact with its nearest neighbours (which are the $j-1$-th and the $j+1$-th particles if and only if $X_j$ is monotone). However, numerical simulations show that, for many interesting initial conditions (including many of those that lead to shocks), such inversions never occur. We therefore assume throughout the article that condition $X_j(t) < X_{j+1}(t)$ holds for all $t, j$.

In the regular case and if $W$ is convex, it has been proven in [26] that (9.1) converges to (9.5) in the following sense :

**Theorem 9.1.1** (Proposition 2. in [26]). *Assume that $W \in C^4(\mathbb{R})$, and that $W'' \geq \alpha > 0$. Suppose $\phi_l = -1$ and $\phi_r = 1$. Let $\phi_0^x \in C^3(\mathbb{R})$ and $\phi_0^\tau \in C^4(\mathbb{R})$ such that (9.4) holds. Assume that $\phi \in C\left([0, T[, C^4([-1, 1])\right)$ is a solution to (9.5) for the initial and boundary conditions (9.6) and (9.7). Let $X_j(t)$ be the unique solution to (9.1) for the initial and boundary conditions (9.2) and (9.3). Then we have the following convergences :*

$$\forall \tau \in [0, T[, \qquad \sup_{-N \leq j \leq N} \left| \frac{1}{N} X_j(N\tau) - \phi\left(\tau, \frac{j}{N}\right) \right| \underset{N \to \infty}{\to} 0, \tag{9.9}$$

$$\forall \tau \in [0, T[, \qquad \sup_{-N \leq j \leq N} \left| \frac{dX_j}{dt}(N\tau) - \partial_\tau \phi\left(\tau, \frac{j}{N}\right) \right| \underset{N \to \infty}{\to} 0. \tag{9.10}$$

When $W$ is convex but not quadratic, even if $\phi_0^x$ and $\phi_0^\tau$ are smooth, shocks generally occur in finite time for solutions of (9.5). By shock, we mean that the solution $\phi$ of (9.5) becomes irregular (see [139] for examples). An interesting question is what happens after such shocks for the discrete system (9.1), and in particular if there is still a link between (9.1) and (9.5). To answer this question, we will consider Riemann-like initial conditions, as is customary in the study of hyperbolic systems.
Let us underline that (9.1), which can be seen as a semi-discrete numerical scheme, is taken for granted, as it comes from a physical model. Some authors take the opposite way, and modify given schemes (adding viscosity for example) in order to go from the discrete system to the continuous one (see [116]), or to help the numerical computation of hyperbolic systems ( [142]).
Let us also mention that their exists a quite detailed study on discrete systems ruled by (9.1) in the particular case where :

$$W(u) = \exp(-u). \tag{9.11}$$

In that case, called the Toda lattice (see [50], [82], [147], [149]), the discrete Hamiltonian system is completely integrable : this allows for a detailed description of the solutions. It is well-known that (9.5) does not describe well the limiting system and that the solutions are dispersive waves. This is linked with Lax pairs, and helps to make the connection with the Korteweg-de Vries equation (see [98]). We will not investigate in this article this particular case, which is, in our understanding, closely linked with the special structure induced by the potential (9.11). We shall however demonstrate that the solutions associated with more general potentials globally display the same features as the dispersive waves of the Toda lattice (see Section 9.5).

**Numerics** In order to have a better understanding of (9.1), we perform some numerical experiments. To do so, we use a Verlet scheme (see [99], p 111) on the variables $U_j$ and $Z_j := \frac{dU_j}{dt}$. More explicitly, we simulate :

$$\begin{cases} U_j^{n+1/2} = U_j^n + \frac{\delta t}{2} Z_j^n, \\ Z_j^{n+1/2} = Z_j^n + \frac{\delta t}{2}\left(W'\left(U_{j+1}^{n+1/2}\right) - 2W'\left(U_j^{n+1/2}\right) + W'\left(U_{j-1}^{n+1/2}\right)\right), \\ U_j^{n+1} = U_j^n + \delta t Z_j^{n+1/2}, \\ Z_j^{n+1} = Z_j^n + \delta t\left(W'\left(U_{j+1}^{n+1/2}\right) - 2W'\left(U_j^{n+1/2}\right) + W'\left(U_{j-1}^{n+1/2}\right)\right), \end{cases} \tag{9.12}$$

where $X_j^n$ is an approximation for $X_j(n\delta t)$. We take an initial condition corresponding to a Riemann problem or a smooth initial condition that develops shocks in finite time (for the sake of simplicity, we only use Riemann problems for illustrations in this article). The crucial feature of (9.12) is that it preserves the Hamiltonian properties of (9.1) (for (9.12) is symplectic). The error we make on $U_j$ in $L^2$ norm is of order $O(NT\delta t)$ (see [77] p13), where $T$ is the final macroscopic time of simulation, which allows to simulate (9.1) for a reasonably large number $2N$ of particles ($N \simeq 10^4$), and thus to have a fair experimental knowledge of the system (9.1).

**Outline of the article** In Section 9.2, we introduce the notations and collect some classical facts about (9.1) and (9.5). In particular, we focus on the initial and boundary conditions, that are supposed to mimic the Riemann problem. We also focus on the natural energy of these systems.
In Section 9.3, we state and next illustrate our main results. We focus first on the simple quadratic potential $W(u) = u^2/2$ and claim that the convergence of (9.1) to (9.5) is true for a large class of initial conditions. This is proved in Section 9.4. Then we examine the case where both $W$ and $W'$ are strongly convex. We show that, if the distances between neighbouring particles remain bounded and if the energy of the continuous system (9.5) is not preserved, solutions of (9.1) do not converge to solutions of (9.5). It is based on the fact that the system (9.1) displays the property of light cone : the perturbations propagate with a finite speed at macroscopic level. This is proved in Section 9.5. We state next a conjecture about a uniform bound on the distances between particles of the system (9.1), that we justify with numerics and that we question through a study of the linear case. This conjecture is motivated by the fact that the assumption of boundedness of the distances between particle is a major assumption in every result of Section 9.5. We discuss it in Section 9.6. Finally, we state that discrete shock waves do not exist, either when $W(u) = \frac{u^2}{2}$ or when $W'$ and $W''$ are strictly convex. It is proved in Section 9.7.

## 9.2 Preliminaries

### 9.2.1 General notations

Let $q \in [1, \infty]$. For $Y_j$, with $j \in [\![-N, N-1]\!]$, we denote by :

$$\|(Y_j)\|_{l_j^q} = \begin{cases} \left( \sum_{j=-N}^{N-1} Y_j^q \right)^{1/q} & \text{if } q < \infty, \\ \max_{j \in [\![-N,N-1]\!]} |Y_j| & \text{if } q = \infty. \end{cases}$$

We denote by $C_p$ the set of piecewise continuous functions on $[-1, 1]$, and $C_p^1$ the set of piecewise continuous functions on $[-1, 1]$ that have piecewise continuous derivatives. We use the subscript *per* for functional spaces to indicate that we intersect these spaces with the space of 2-periodic functions. We use the subscripts $x$, $\tau$, $t$ for functional spaces to indicate that these spaces have their variables $x$ in $[-1, 1]$, $\tau$ in $[0, T]$, respectively $t$ in $[0, NT]$. For example :

$$H_\tau^1(C_x) := H^1([0, T], C([-1, 1])).$$

### 9.2.2 Initial data and boundary conditions

In the present article, we mainly use Dirichlet boundary conditions. They have the advantage of being consistent with Riemann problems. In Section 9.6, we will also use periodic boundary conditions for technical reasons ; more specifically, when the potential $W$

is quadratic, it allows for an explicit resolution of (9.1).
We say that (9.1) (respectively (9.5)) is set with Dirichlet boundary conditions if (9.2) and (9.3) (respectively (9.6) and (9.7)) are satisfied and if compatibility condition (9.4) holds. We say that the system (9.1) is set with periodic boundary conditions if (9.1) is satisfied for all $j$ with the convention that $X_{N+j} = X_{-N+j}$. The associated initial conditions are (9.3) with $\phi_0^\tau, \phi_0^x \in C_p$ such that the following compatibility condition :

$$\int_{-1}^{1} \phi_1^x(x)dt = 0$$

is satisfied.

### 9.2.3 Hypotheses on $W$

We suppose that $W$ is $C^2$ and strongly convex :

$$W'' \geq \alpha > 0. \tag{9.13}$$

Indeed, this assumption implies that (9.5) is a strictly hyperbolic system (if not, the theory for (9.5) is far more complex). A very particular case is when $W$ is quadratic :

$$W(u) = \frac{u^2}{2}. \tag{9.14}$$

When we consider a non-quadratic potential, we also assume that $W$ is $C^3$ and that $W'$ is strictly convex :

$$W''' > 0. \tag{9.15}$$

Let us emphasize that (9.5) is genuinely non-linear when $W''' > 0$ (see [139] p 113 and p 127). We speak about the *linear case* (respectively the *nonlinear case*) when (9.14) is satisfied (respectively when (9.13) and (9.15) are satisfied). The terminology may seem ambiguous, but it is justified by (9.5), which involves $W'$ and not $W$.
The convexity (9.13), and *a fortiori* (9.15), is obviously a strong and non-physical simplification, as a physical potential should be even (and non-constant even potentials with other minima than 0 cannot satisfy (9.13) on $\mathbb{R}$). For example, our results do not directly cover this "quadratic" potential :

$$W(u) = (|u| - 1)^2. \tag{9.16}$$

Our numerical experiments suggest that for given initial conditions, the distances between particles is bounded from below and from above (see Remark 4 and Section 9.6). Hence one can require (9.13) or (9.15) to be true only on the corresponding intervals. For example, if we know a priori that the order of the particles is preserved, one can apply our results with the potential (9.16).

### 9.2.4 The discrete system

**Notations** For the discrete system (9.1), we denote :

$$V_j(t) = \frac{dX_j}{dt}(t), \qquad U_j(t) = X_{j+1}(t) - X_j(t), \qquad Z_j(t) = \frac{dU_j}{dt}(t).$$

*Remark* 5 (Dependence on $N$). $X_j$ and the other discrete quantities implicitly depend on $N$. When necessary, we write $X_j^N$, $U_j^N$, *et cetera*.

The correspondence between the discrete system and the continuous system is encoded in the following notations :

$$\begin{aligned}
&k^N(x) := \lfloor Nx \rfloor,\\
&\theta^N(x) := Nx - k^N(x),\\
&\phi^N(\tau,x) := \frac{1-\theta^N(x)}{N} X_{k^N(x)}(t) + \frac{\theta^N(x)}{N} X_{k^N(x)+1}(t),\\
&\zeta^N(\tau,x) := \partial_\tau \phi^N(\tau,x) = \left( \left(1-\theta^N(x)\right) \frac{d}{dt} X_{k^N(x)} + \theta^N(x) \frac{d}{dt} X_{k^N(x)+1} \right)(t),\\
&\xi^N(\tau,x) := \frac{d}{dt} X_{k^N(x)}(t).
\end{aligned}$$

In any case, we extend the functions $\phi^N, \xi^N$ by continuity with constant branches on $]-\infty,-1] \cup [1,+\infty[$. For example, we have $\phi(x) = \phi_l$ if $x < -1$.

Remark that $\partial_\tau \phi^N$ is the linear interpolation of $V_j$, and is therefore not equal to $\xi^N(\tau,x)$, which corresponds to $V_j(t)$. However, they are very close to each other, as is stated in the following Lemma (the proof is given in the Appendix below) :

**Lemma 9.2.1.** *Let $T > 0$. We have the following estimates :*

$$\left\| \zeta^N(\tau,.) \right\|_{L^2_x} \leq \left\| \xi^N(\tau,.) \right\|_{L^2_x} \leq 6 \left\| \zeta^N(\tau,.) \right\|_{L^2_x}, \tag{9.17}$$

*and the following equivalences :*

$$\zeta^N \to \xi^\infty \text{ in } L^2_{\tau,x} \qquad \Longleftrightarrow \qquad \xi^N \to \xi^\infty \text{ in } L^2_{\tau,x}, \tag{9.18}$$

$$\zeta^N \rightharpoonup \xi^\infty \text{ in } L^2_{\tau,x} \qquad \Longleftrightarrow \qquad \xi^N \rightharpoonup \xi^\infty \text{ in } L^2_{\tau,x}. \tag{9.19}$$

**Properties of the discrete system** The discrete system (9.1) is an Hamiltonian system, with the energy :

$$\mathcal{E}_D\left( \{U_j\}_{-N\leq j\leq N}, \{V_j\}_{-N\leq j\leq N} \right) := \frac{1}{2N} \sum_{j=-N}^{N-1} V_j^2 + \frac{1}{N} \sum_{j=-N}^{N-1} W(U_j). \tag{9.20}$$

The energy (9.20) is the total mechanical energy of the system. The kinetic energy is the first term and the potential energy is the second term. Either in Dirichlet or in periodic setting,

if $X_j(t)$ satisfies (9.1) and $\{U_j\}, \{V_j\}$ are defined by (9.8), an elementary calculation shows that the discrete energy $\mathcal{E}_D$ is preserved :

$$\frac{d}{dt}\left[\mathcal{E}_D\left(\{U_j(t)\}_{-N\leq j\leq N}, \{V_j(t)\}_{-N\leq j\leq N}\right)\right] = 0. \tag{9.21}$$

A direct application of the Cauchy-Lipschitz theorem implies that (9.1), with the initial and boundary conditions (9.2) and (9.3), has a unique solution, locally in time. Moreover the energy (9.20) being a coercive function of $U_i$ and $V_i$ (as $W$ satisfies (9.13)), this solution is uniformly bounded and is a global one. For the sake of simplicity, we will hereafter use the following abuse of notation :

$$\mathcal{E}_D(t) := \mathcal{E}_D\left(\{U_j(t)\}_{-N\leq j\leq N}, \{V_j(t)\}_{-N\leq j\leq N}\right). \tag{9.22}$$

For later purpose, we define the notion of discrete compatibility.

**Definition 9.2.1** ($D$-compatibility)**.** *We say that $T > 0$ is $D$-compatible with $\phi_0^x$ and $\phi_0^\tau$ if there exist $\delta > 0$ and $C > 0$ such that :*

$$\begin{aligned} &\left|\partial_x\phi^N(\tau,x) - \phi_0^x(-1)\right| \leq CN^{-1} && \forall(\tau,x)\in[0,T]\times[-1,-1+\delta],\\ &\left|\partial_x\phi^N(\tau,x) - \phi_0^x(1)\right| \leq CN^{-1} && \forall(\tau,x)\in[0,T]\times[1-\delta,1], \end{aligned}$$

*and $\partial_\tau\phi^N$ converges uniformly to 0 on $[0,T]\times\{[-1,-1+\delta]\cup[1-\delta,1]\}$, as $N$ goes to infinity.*

$D$-compatibility means that the solution of (9.1) is almost not perturbed near the boundary $x = -1$ and $x = 1$, until time $T$.

### 9.2.5 The continuous system

Let $T > 0$. Following [139], p 28, we say that $\phi \in W^{1,\infty}_{\tau,x}$ is a weak solution of (9.5) with initial conditions (9.6) if, for all $g \in C_c^\infty([0,T[\times]-1,1[)$ :

$$\int_0^T\int_{-1}^1 \left\{\partial_x g W'(\partial_x\phi) - \partial_\tau g\partial_\tau\phi\right\}(\tau,x)dxd\tau = \int_{-1}^1 g(0,x)\phi_0^\tau(x)dx, \tag{9.23}$$

$$\int_0^T\int_{-1}^1 \left\{\partial_x g\partial_\tau\phi - \partial_\tau g\partial_x\phi\right\}(\tau,x)dxd\tau = \int_{-1}^1 g(0,x)\phi_0^x(x)dx. \tag{9.24}$$

We say that a weak solution $\phi$ of (9.5) is an entropy solution if it also satisfies in the weak sense (see [139], p 82) :

$$\frac{d}{d\tau}\mathcal{E}_C(\tau) \leq 0, \tag{9.25}$$

where $\mathcal{E}_C$ is the continuous energy associated with $\phi$ :

$$\mathcal{E}_C(\tau) = \int_{-1}^1 \left\{\frac{1}{2}(\partial_\tau\phi)^2 + W(\partial_x\phi)\right\}(\tau,x)dx. \tag{9.26}$$

We are interested in weak entropy solutions $\phi$ of (9.5) satisfying

$$\mathcal{E}_C(T) < \mathcal{E}_C(0). \tag{9.27}$$

Shocks satisfy (9.27). We recall now the definition of the Riemann problems :

**Definition 9.2.2** (Riemann problem)**.** *Let $u_l, u_r, v_l, v_r \in \mathbb{R}$, and :*

$$\phi_0^x(x) := \begin{cases} u_l & \text{if } x < 0, \\ u_r & \text{if } x \geq 0, \end{cases} \qquad \phi_0^\tau(x) := \begin{cases} v_l & \text{if } x < 0, \\ v_r & \text{if } x \geq 0. \end{cases} \tag{9.28}$$

*Solving the Riemann problem associated with $(u_l, u_r, v_l, v_r)$ consists in finding $\phi$ an entropy solution of* (9.5) *on $[0,T] \times \mathbb{R}$ with initial conditions* (9.6)*.*

This is the classical Riemann problem. However, it is possible to use weaker assumptions on $\phi_0^x$ and $\phi_0^\tau$, that simulate what we call a *boundary* Riemann problem. This second definition is more flexible and allows to work with a very large class of initial conditions (for example, smooth initial data that develop discontinuities in finite times, in system (9.5)). Namely :

**Definition 9.2.3** (Boundary Riemann problem)**.** *Let $u_l, u_r \in \mathbb{R}^2$, and :*

$$\phi_0^x(x) := \begin{cases} u_l & \text{if } x < -1/2, \\ u_r & \text{if } x > 1/2, \end{cases} \qquad \phi_0^\tau(x) := \begin{cases} 0 & \text{if } x < -1/2, \\ 0 & \text{if } x > 1/2, \end{cases} \tag{9.29}$$

*without further requirement on $\phi_0^x$ and $\phi_0^\tau$ between $-1/2$ and $1/2$. Solving this boundary Riemann problem consists in finding $\phi(\tau, x)$, entropy solution of* (9.5) *with initial and boundary conditions* (9.6) *and* (9.7)*, and $X_j(t)$, solution of* (9.1) *with initial and boundary conditions* (9.2) *and* (9.3)*, with $\phi_l$ and $\phi_r$ being constant so that* (9.4) *is satisfied.*

We impose $\phi_0^\tau$ to vanish near the boundary in the boundary Riemann problem (9.29) so that $\phi_l$ and $\phi_r$ are constant ; this is useful to avoid some technicalities about boundary conditions.
The solutions of the Riemann problem (9.28) are combinations of rarefaction waves and shock waves. One does not change the solution of (9.5) (for $T$ sufficiently small) if one restricts $\phi$ to $x \in [-1, 1]$ and solves (9.5) with Dirichlet boundary conditions (9.7).
For example (see [139], p 127-131), if $u_r > u_l$ and if the following Rankine-Hugoniot condition is satisfied :

$$v_r - v_l = \sigma\,(u_r - u_l) \qquad \text{and} \qquad W'(u_r) - W'(u_l) = \sigma(v_r - v_l), \tag{9.30}$$

then the entropy solution of the Riemann problem reads as :

$$\partial_x \phi(\tau, x) = \begin{cases} u_l & \text{if } x < -\sigma\tau \\ u_r & \text{if } -\sigma\tau \leq x \end{cases}, \qquad \partial_\tau \phi(\tau, x) = \begin{cases} v_l & \text{if } x < -\sigma\tau \\ v_r & \text{if } -\sigma\tau \leq x \end{cases},$$

and satisfies (9.27). We are interested in boundary Riemann problems. As a consequence, we focus on weak solutions $\phi \in W^{1,\infty}_{\tau,x}$ of (9.5) in the Dirichlet setting that can be continued by a constant on the right and on the left :

**Definition 9.2.4** ($C$-compatibility)**.** *Let $\phi_0^x$ and $\phi_0^\tau \in C_p$, $\phi_l$ and $\phi_r \in \mathbb{R}$ satisfying* (9.4)*. Assume that $\phi$ is an entropy solution of* (9.5) *on $[0,T]\times[-1,1]$ with initial and boundary conditions* (9.6) *and* (9.7)*. Let $\phi_0^x$ and $\phi_0^\tau$ satisfy* (9.29)*. We say that $T$ is $C$-compatible with $\phi_0^x$ and $\phi_0^\tau$ if there exists $\delta>0$ such that :*

$$\begin{cases} \phi(\tau,x) = \phi_l + u_l(x+1) & \text{if } (\tau,x)\in[0,T]\times[-1,-1+\delta], \\ \phi(\tau,x) = \phi_r + u_r(x-1) & \text{if } (\tau,x)\in[0,T]\times[1-\delta,1]. \end{cases} \tag{9.31}$$

If $T$ is $D$-compatible and $C$-compatible with $\phi_0^x$ and $\phi_0^\tau$, we say that it is $DC$-compatible. Basically, $DC$-compatibility provides a strong control on the solutions of (9.1) and (9.5) near the boundary $x=-1$ and $x=1$, until time $T$.

For the linear system, we have the following theorem of existence and uniqueness (see Theorem 3 p 384 and Theorem 4 p 385 of [58]) :

**Theorem 9.2.1** (Existence and uniqueness in the linear case)**.** *Let $\phi_0^x, \phi_x^\tau \in L^2_x$. Suppose that $A, B \in C^1_x$, and $A \geq \alpha > 0$. Then, there exists one and only one solution $\phi \in H^1_{\tau,x}$ to :*

$$\partial_\tau^2\phi(\tau,x) = \partial_x\left(A(x)\partial_x\phi(\tau,x) + B(x)\right),$$

*with initial and boundary conditions* (9.6) *and* (9.7)*. In addition, the energy $\widetilde{\mathcal{E}}_C$ of the system is preserved :*

$$\widetilde{\mathcal{E}}_C = \int_{-1}^{1}\left\{\frac{1}{2}\left(\partial_\tau\phi(\tau,x)\right)^2 + \frac{A(x)}{2}\left(\partial_x\phi(\tau,x)\right)^2 + B(x)\partial_x\phi(\tau,x)\right\}dx.$$

It is clear that this energy extends the above definition (9.26).

### 9.2.6 Discrete shock waves

**Definition 9.2.5.** *We say that $X_j(t)$, $j\in\mathbb{Z}, t\in\mathbb{R}_+$, is a discrete shock wave of* (9.1) *associated with $(u_l,u_r)\in\mathbb{R}^2$, $u_l\neq u_r$, if $X_j(t)$ satisfies* (9.1) *and if there exist $\phi\in C^2(\mathbb{R})$ and $c\in\mathbb{R}$ such that :*

$$\begin{aligned} &X_j(t) = \phi(j-ct) && \forall t\in\mathbb{R}_+, \forall j\in\mathbb{Z}, \\ &\lim_{x\to-\infty}\phi'(x) = u_l, \\ &\lim_{x\to+\infty}\phi'(x) = u_r. \end{aligned}$$

The definition implies that :

$$c^2\phi''(x) = W'(\phi(x+1)-\phi(x)) - W'(\phi(x)-\phi(x-1)). \tag{9.32}$$

## 9.3 Results

We state here our main results and illustrate them with some numerical results.

### 9.3.1 The linear case

When $W(u) = u^2/2$, one observes that $\phi^N$ converges in $H^1_{\tau,x}$ to $\phi$. One can even see that for regular initial conditions, this convergence seems to hold in every $W^{1,p}_{\tau,x}$. This is illustrated by Figure 9.1.

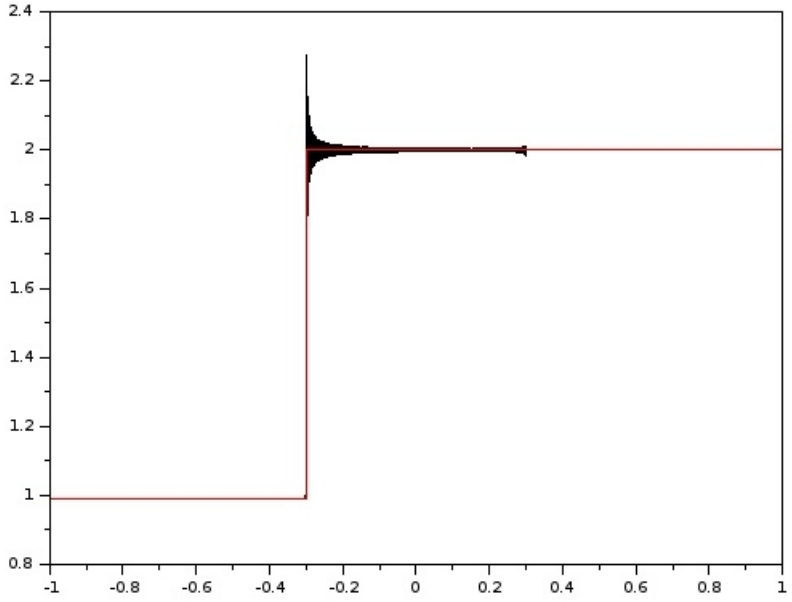



FIGURE 9.1 – Comparison between $\partial_x \phi^N$ (black curve) and $\partial_x \phi$ (red curve) for Riemann shock initial conditions. $W(u) = \frac{u^2}{2}$, $N = 10000$, $\tau = 0.3$.

We prove this convergence in a generalized framework, where the quadratic potential $W$ depends not only on $u$ but also on $x$ :

**Theorem 9.3.1.** *Let $T > 0$, $\phi_l, \phi_r \in \mathbb{R}$, $\phi_0^x$ and $\phi_0^\tau \in C_p$ satisfy (9.4). Assume that :*

$$W(x,u) = \frac{1}{2}A(x)u^2 + B(x)u.$$

*with $A, B \in C^1_x$ and $A$ satisfying :*

$$A \geq \alpha > 0, \tag{9.33}$$

*for $\alpha$ a given positive constant. Consider the solution $X_j^N(t)$ to :*

$$\frac{d^2}{dt^2}X_j^N(t) = \partial_u W\left(\frac{j}{N}, U_j^N(t)\right) - \partial_u W\left(\frac{j-1}{N}, U_{j-1}^N(t)\right) \tag{9.34}$$

*for all $j \in [\![-N+1, N-1]\!]$ for the initial and boundary conditions (9.2) and (9.3). Then the associated $\phi^N$ converges :*

$$\phi^N \underset{N\to+\infty}{\to} \phi \text{ in } H^1_{\tau,x},$$

*where $\phi$ is the unique solution of :*

$$\partial_\tau^2 \phi(\tau, x) = \partial_x \left( \partial_u W(x, \partial_x \phi(\tau, x)) \right), \tag{9.35}$$

*for the initial and boundary conditions* (9.6) *and* (9.7).

It has a direct corollary :

**Corollary 9.3.1.** *Let $T > 0$, $\phi_l, \phi_r \in \mathbb{R}$, $\phi_0^x$ and $\phi_0^\tau \in C_p$ satisfying* (9.4). *Assume that $W$ satisfies* (9.14). *Let $\phi$ be the solution of* (9.5) *for the initial and boundary conditions* (9.6) *and* (9.7), *and $X_j^N$ be the solution of* (9.1) *for the initial and boundary conditions* (9.2) *and* (9.3). *We have the following convergence :*

$$\phi^N \underset{N \to +\infty}{\to} \phi \text{ in } H^1_{\tau,x}.$$

*Remark* 6 (Less restrictive assumptions). For both Theorem 9.3.1 and Corollary 9.3.1, it is sufficient to assume that $\phi_0^x$ and $\phi_0^\tau \in L^2_x$ as long as $X_j^N$ satisfies the initial condition :

$$X_{j+1}(0) - X_j(0) = N \left( \phi^N \left( \tau = 0, \frac{j+1}{N} \right) - \phi^N \left( \tau = 0, \frac{j+1}{N} \right) \right)$$
$$\text{and} \quad \frac{d}{dt} X_j(0) = \partial_\tau \phi^N \left( \tau = 0, \frac{j}{N} \right),$$

such that :

$$\partial_x \phi^N(\tau = 0, .) \to \phi_0^x \text{ in } L^2_x, \qquad \xi^N(\tau = 0, .) \to \phi_0^\tau \text{ in } L^2_x.$$

### 9.3.2 The non-linear case

If the potential $W$ is convex but not quadratic, when there is a shock, we observe on numerical simulations that $\partial_x \phi^N$ does not converge strongly to the associated $\partial_x \phi$. It does not even converge weakly. Actually $\partial_x \phi^N$ oscillates with a high frequency and an amplitude that does not decrease when $N$ grows. We believe that this situation is generic for basically any potential such that $W'$ is not affine on the zone where $U_j$ evolves. We can to prove this non-convergence, under the extra-hypothesis that $W'$ is strictly convex, and under the assumption that the distance between particles $U_j^N$ is bounded uniformly in $N$, $j$ and $t \in [0, NT]$ (the latter assumption is discussed in Section 9.6).

We illustrate this non-convergence with Figure 9.2 comparing $\partial_x \phi^N$ and $\partial_x \phi$. Remarkably enough, even if there are large oscillations, let us remark that distances between particles remain bounded. We check numerically that $\partial_x \phi^N$ coincides with $\partial_x \phi$ outside a region of space away from the shock that grows linearly in macroscopic time ; this property is known for Toda lattice [82]. It is also known that the Korteweg-de-Vries equation [98] has a similar behaviour.

**Theorem 9.3.2.** *Let $W \in C^3([a, b])$ satisfy* (9.13) *and* (9.15). *Assume that $\phi_0^x, \phi_0^\tau \in C^1_{p,x}$ and $\phi_l, \phi_r \in \mathbb{R}$ satisfy* (9.4) *and* (9.29), *for $u_l, u_r \in \mathbb{R}$. Let $T_0 > 0$.*

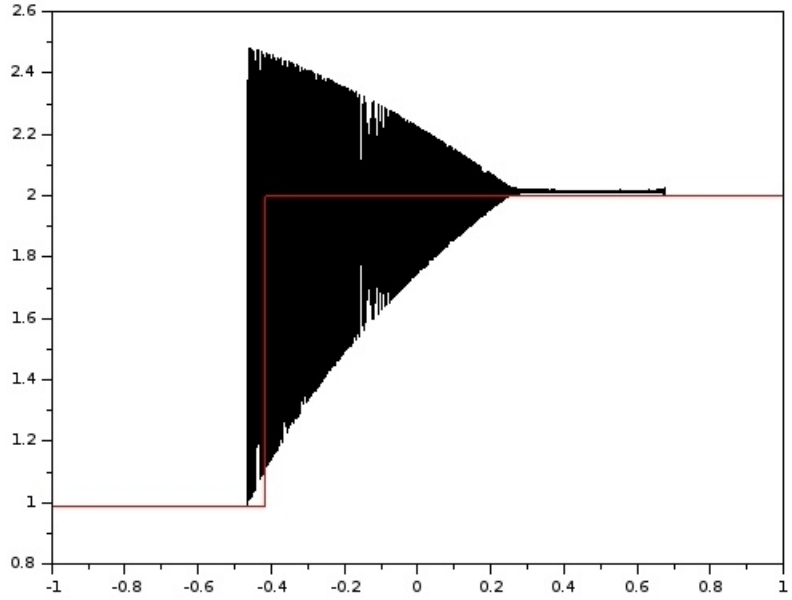



FIGURE 9.2 – Comparison between $\partial_x \phi^N$ (black curve) and $\partial_x \phi$ (red curve) for Riemann shock initial conditions, and a non-linear potential. $N = 5000$, $\tau = 0.075$, $W(u) = \frac{u^6}{6}$.

*Let $X_j^N(t)$ be the solution of* (9.1) *for the initial and boundary conditions* (9.2) *and* (9.3). *Suppose that :*

$$U_j^N(t) \in [a, b] \qquad \forall j \in [\![-N, N-1]\!], \forall N > 0, \forall t \in [0, NT_0]. \tag{9.36}$$

*Then there exists a D-compatible $T \leq T_0$.*
*Assume that $\phi \in W^{1,\infty}_{\tau,x}$, is an entropy solution of* (9.5) *for the initial and boundary conditions* (9.6) *and* (9.7), *that $T$ is $C$-compatible and that there exists $T_1 < T$ satisfying :*

$$\mathcal{E}_C(T_1) < \mathcal{E}_C(0). \tag{9.37}$$

*Then $\phi^N$ does* ***not*** *converge to $\phi$ in the sense of distribution in space and time $D'_{\tau,x}$.*

*Remark* 7 (Applicability to shocks). Let us remark that the only restriction we have on $\phi_0^x$ and $\phi_0^\tau$ is to be piecewise continuous, with piecewise continuous derivatives (and some technical assumptions around the boundary). Thus, one can design $\phi_0^x$ and $\phi_0^\tau$ so that the initial configuration leads instantly to (local) shock waves. In other words, for such initial configuration, we have that (9.37) is satisfied, for all $T_1 > 0$. Therefore, if one can apply Theorem 9.3.2, that is, if we *a priori* know that (9.36) holds (see Conjecture 9.3.1 below), then we have that (9.1) do not approximate (9.5) for such initial data.

*Remark* 8 (Entropy solution). In Theorem 9.3.2, we only compare $\phi^N$ to the *entropy* solution $\phi$ of (9.5). We think that $\phi^N$ cannot converge to *any* weak solution of (9.5). Indeed, if $\phi^N$

converges to $\phi$, which is a solution of (9.5), Lemma 9.5.2 below implies that $\partial_x \phi^N$ converges strongly to $\partial_x \phi$. But numerical experiments show that $\partial_x \phi^N$ oscillates too much so that it cannot converge strongly to anything : this justifies our conclusion.

*Remark* 9 (Reversibility). (9.1) is a reversible system, but (9.5) is not when shocks occur, whereas both systems are reversible as long as the solution $\phi$ of (9.5) remains smooth enough. In the first case, the discrete system does not converge to the continuous one (Theorem 9.3.2), but it does in the second case (Theorem 9.1.1).

*Remark* 10 (Convergence breakdown). Suppose that $W$ satisfies (9.13) and (9.15). Let $\phi_0^x$ and $\phi_0^\tau$ be smooth functions. Define $X_j^N$ and $\phi$ as in Theorem 9.3.2. Suppose that there exists a $DC$-compatible $T_f > 0$ and $T_0 < T_f$ such that $\mathcal{E}_C(T_0) < \mathcal{E}_C(0)$ -in other words, a shock occurs. If Conjecture 9.3.1 below holds for initial data $\phi_0^x$, $\phi_0^\tau$ and for given time $T = T_1 \in ]T_0, T_f]$, it leads to the following paradoxical situation :

1. until time $T_0$, $\phi(\tau, .)$ is sufficiently smooth, so that Theorem 9.1.1 applies. Thus :
$$\phi^N \to \phi \text{ in } C\left([0, T_0] \times [-1, 1]\right),$$
2. as $\mathcal{E}_C(T_1) < \mathcal{E}_C(0)$, applying Theorem 9.3.2, we get that $\phi^N$ does not converge to $\phi$ in $D'([0, T_1] \times [-1, 1])$ .

Therefore, shocks break the discrete-to-continuum convergence of (9.1) to (9.5).

In the discrete system (9.1), perturbations are propagating instantly. It can be proven by linearizing (9.1) and assuming a small perturbation $\varepsilon$ on a fixed $j_0$-th particle :

$$\widetilde{X}_j(0) = X_j(0) + \varepsilon \delta^j_{j_0}, \qquad \frac{dX_j}{dt}(0) = \frac{d\widetilde{X}_j}{dt}(0).$$

We assume that both $X_j$ and $\widetilde{X}_j$ satisfy (9.1). Integrating iteratively (9.1) for small time $\Delta t$, we get (for $j > 0$) at leading order in $\varepsilon$ (the proof of this formal expansion is in the spirit of the proof of Proposition 9.5.1 below) :

$$\widetilde{X}_{j_0+j}(\Delta t) - X_{j_0+j}(\Delta t) \simeq \varepsilon \frac{(\Delta t)^{2j}}{(2j)!} \prod_{k=0}^{j-1} W''\left(U_{j_0+k}(0)\right).$$

However, on the macroscopic level, this propagation has a finite speed. This paradox is due to the fact that the influence of perturbation on $x_0$ at $t_0$ decays exponentially outside a cone $|x - x_0| \leq c\,|t - t_0|$. It is noticeable that this light cone property is an important feature of hyperbolic systems. It is however a key ingredient to prove that (9.1) does *not* converge to (9.5).

We formalize the fact that perturbations of the discrete system propagate with a finite speed on the macroscopic level by the following theorem :

**Theorem 9.3.3.** *Let $W \in C^2(\mathbb{R})$ satisfy (9.13). Let $T > 0$. Assume that $X_j^N(t)$ and $\widetilde{X}_j^N(t)$ satisfy (9.1), for $j \in [\![0, N-1]\!]$, $t \in [0, NT]$, with right boundary condition $\widetilde{X}_N^N = X_N^N = N\phi_r$. We denote by :*

$$\widetilde{V}_j^N(t) = \frac{d}{dt}\widetilde{X}_j^N(t), \qquad \widetilde{U}_j^N(t) := \widetilde{X}_{j+1}^N(t) - \widetilde{X}_j^N(t).$$

*Suppose that*

$$X_j^N(t=0) = \widetilde{X}_j^N(t=0), \forall j \in [\![1, N-1]\!],$$
$$V_j^N(t=0) = \widetilde{V}_j^N(t=0), \forall j \in [\![1, N-1]\!].$$

*Assume that there exists* $C \in \mathbb{R}_+$ *such that,* $\forall N > 0$ *:*

$$\sup_{t \in \mathbb{R}_+} \left( \sum_{j=0}^{N-1} W\left(U_j^N(t)\right) + \frac{1}{2} \sum_{j=0}^{N-1} \left(V_j^N(t)\right)^2 \right) \leq CN, \tag{9.38}$$

*and :*

$$K = \sup_{u \in ]u_1, u_2[} \left| W''(u) \right| < +\infty, \tag{9.39}$$

*where :*

$$u_1 := \inf_{\substack{N>0,\\ j \in [\![0, N-1]\!],\\ t \in [0, NT]}} \left\{ \min\left( U_j^N(t), \widetilde{U}_j^N(t) \right) \right\},$$
$$u_2 := \sup_{\substack{N>0,\\ j \in [\![0, N-1]\!],\\ t \in [0, NT]}} \left\{ \max\left( U_j^N(t), \widetilde{U}_j^N(t) \right) \right\}.$$

*Let* $c = \exp(2)\sqrt{K}$*. Let* $x \in ]0,1[$*. Then, for all* $\tau < \frac{x}{c}$*, we have :*

$$\lim_{N \to +\infty} \left\{ N \sup_{\substack{0 \leq t < N\tau,\\ j > Nx}} \left| U_j^N(t) - \widetilde{U}_j^N(t) \right| \right\} = 0, \tag{9.40}$$

$$\lim_{N \to +\infty} \sup_{\substack{0 \leq t < N\tau,\\ j > Nx}} \left| V_j^N(t) - \widetilde{V}_j^N(t) \right| = 0. \tag{9.41}$$

A few remarks are in order :

*Remark* 11. The speed $c$ in Theorem 9.3.3 is not optimal -we see it from numerical experiments- but it has the same order as the natural speed of (9.5), given by (9.30). Formally :

$$\exp(2)\sqrt{K} \propto \sup_{u \in [u_l, u_r]} \sqrt{|W''(u)|} \propto \sqrt{\left| \frac{W'(u_r) - W'(u_l)}{u_r - u_l} \right|}.$$

The above formal calculation is exact when $W$ is quadratic.

*Remark* 12. Assumption (9.39) of Theorem 9.3.3 is automatically fulfilled if there exists $\alpha, \beta \in \mathbb{R}$ such that $\alpha \leq W''(u) \leq \beta, \forall u \in \mathbb{R}$. This is the case when the potential $W$ is quadratic. However, such a bound cannot hold if $W'$ is strongly convex : one needs to know a priori that the distances $U_j^N$ between particles is bounded uniformly in $N$.

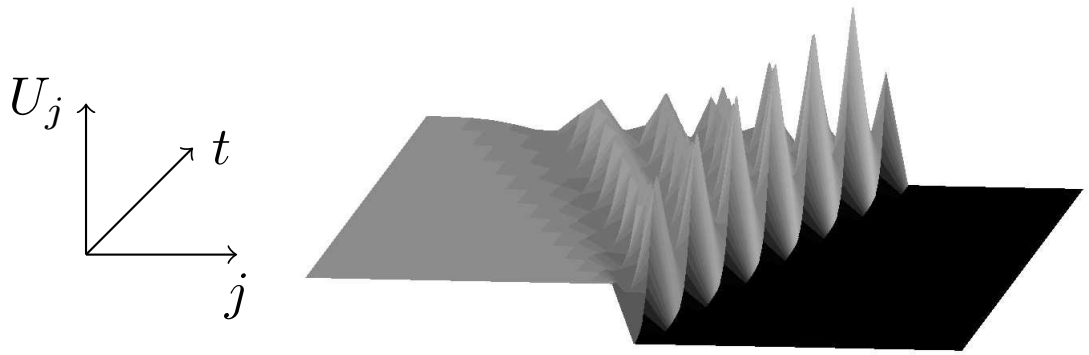



FIGURE 9.3 – Light cone on the surface $\left(t, x, U^N_{k^N(x)}(t)\right)$, for Riemann initial conditions corresponding to a shock wave, for a non-linear potential $W(u) = \frac{u^6}{6}$

### 9.3.3 Uniform $L^\infty_t\left(l^\infty_j\right)$ bound

Most of the results we are able to prove in the non-linear case require the assumption that, for given initial data, the distances between particles remain bounded uniformly in $N$ (that is, (9.36) is satisfied). We have not been able to prove that this assumption is fulfilled. We formulate the following conjecture :

*Conjecture* 9.3.1. Suppose that $W \in C^2(\mathbb{R})$ satisfies (9.13). Assume that $\phi^x_0$ and $\phi^\tau_0 \in C^1_{p,x}$. Then, there exist $T > 0$ and $a < b \in \mathbb{R}$ depending only on $W$, $\phi^x_0$ and $\phi^\tau_0$ such that, for $X_j(t)$ satisfying (9.1) for the initial and boundary conditions (9.2) and (9.3) :

$$a \leq U^N_j(N\tau) \leq b, \qquad \forall j \in [\![-N, N-1]\!], \forall N \in \mathbb{N}, \forall \tau \in [0, T]. \tag{9.42}$$

Note that in the case of Riemann problem (9.28), the initial conditions satisfy the hypotheses of Conjecture 9.3.1. We checked Conjecture 9.3.1 numerically for a large set of piecewise smooth initial data, with potentials of the form $W(u) = Au^\gamma + Bu^2$, $\gamma > 2$, $A, B \in \mathbb{R}_+$. When $\phi$ is sufficiently smooth, Conjecture 9.3.1 can be proven (by Theorem 9.1.1).

Let us point out the fact that it seems necessary to require some smoothness on the initial conditions in Conjecture 9.3.1. In other words, one cannot hope that, for $X^N_j$ solution of (9.1), for given $T > 0$, $\left\|U^N_j\right\|_{l^\infty_{j,t}}$ is controlled by $\left\|U^N_j(t=0)\right\|_{l^\infty_j}$ and $\left\|V^N_j(t=0)\right\|_{l^\infty_j}$ uniformly in $N$.

Indeed, we can prove the following proposition, illustrated by Figure 9.4 above :

**Proposition 9.3.2.** *Let $W(u) = \frac{u^2}{2}$, and $\tau_0 > 0$. There exists a sequence of initial conditions $U^N_j(t=0)$, $V^N_j(t=0)$ such that, for $X^N_j(t)$ the corresponding solutions of* (9.1) *with periodic boundary conditions, we have :*

$$\left\|U^N_j(t=0)\right\|_{l^\infty_j} \leq 1, \qquad \left\|V^N_j(t=0)\right\|_{l^\infty_j} \leq 1, \qquad \left\|U^N_j(N\tau_0)\right\|_{l^\infty_j} \underset{N\to\infty}{\to} +\infty.$$

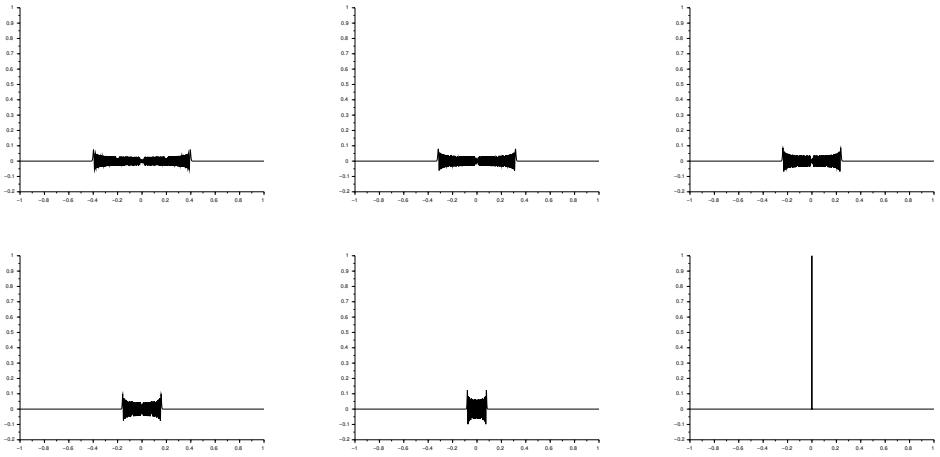


FIGURE 9.4 – Successive pictures of $U_j(N\tau)$. $N = 1000$, $\tau = 0, 0.08, ..., 0.4$

### 9.3.4 Non-existence of discrete shock waves

A natural question is whether or not there exist non-trivial discrete shock waves for the Newton equation (9.1). Should such discrete progressive waves exist, one could expect that they would describe an important feature of the limit of (9.1) system when $N \to +\infty$. Unfortunately, we prove that discrete shock waves do not exist, even when the potential is quadratic. More specifically, we prove the following propositions :

**Proposition 9.3.3.** *Suppose $W$ satisfies* (9.14)*. Then there exists no discrete shock wave for* (9.1)*, in the sense of Definition 9.2.5.*

**Proposition 9.3.4.** *Suppose $W$ satisfies* (9.13) *and* (9.15)*. Then there exists no discrete shock wave for* (9.1)*, in the sense of Definition 9.2.5.*

*Remark* 13. It is straightforward from the proof that there does not exist any other discrete wave than the constant ones in the linear case. In the non-linear case, we do not know if there exists solitons, that is $X_j(t)$ satisfying Definition 9.2.5, with the slight modification that $u_l = u_r$.

## 9.4 The linear case

When the potential is quadratic, the corresponding wave equation (9.5) is linear. Its characteristic lines do not cross, therefore, when the initial conditions are regular, shocks never occur. Furthermore, the energy is preserved : the continuous system (9.5) is thus conservative, as the discrete one (9.1). This is the reason why the discrete system naturally tends to the continuous one, and we show it with simple arguments, essentially using weak compactness of $H^1_{\tau,x}$. This extends the results of [26].

Let us prove Theorem 9.3.1. We first prove that the discrete energy is preserved :

**Lemma 9.4.1.** *Under the hypotheses of Theorem 9.3.1, the following generalized discrete energy is preserved :*

$$\widetilde{\mathcal{E}}_D^N(t) := \sum_{k=-N}^{N-1} W\left(\frac{k}{N}, U_k(t)\right) + \frac{1}{2}\sum_{k=-N}^{N-1} \left(V_k(t)\right)^2. \tag{9.43}$$

*Proof of Lemma 9.4.1.* Using (9.34), we get that :

$$\begin{aligned}
\frac{d}{dt}\widetilde{\mathcal{E}}_D^N(t) &= \sum_{k=-N}^{N-1} \left\{\partial_u W\left(\frac{k}{N}, U_k\right)\left(V_{k+1} - V_k\right)\right\}(t) + \sum_{k=-N+1}^{N-1} \left\{V_k \frac{d^2}{dt^2} X_k\right\}(t), \\
&= \sum_{k=-N}^{N-1} \left\{\partial_u W\left(\frac{k}{N}, U_k\right)\left(V_{k+1} - V_k\right)\right\}(t) \\
&\quad + \sum_{k=-N+1}^{N-1} \left\{V_k\left(\partial_u W\left(\frac{k}{N}, U_k\right) - \partial_u W\left(\frac{k-1}{N}, U_{k-1}\right)\right)\right\}(t).
\end{aligned}$$

Reorganizing the sum, we obtain :

$$\frac{d}{dt}\widetilde{\mathcal{E}}_D^N(t) = \left\{\partial_u W\left(\frac{N-1}{N}, U_{N-1}\right) V_N - \partial_u W\left(-1, U_{-N}\right) V_{-N}\right\}(t).$$

Yet, as (9.3) is satisfied, we have $V_N = V_{-N} = 0$. Therefore :

$$\frac{d}{dt}\widetilde{\mathcal{E}}_D^N(t) = 0,$$

which implies the desired result. □

Next we prove that (9.34) is consistent with (9.35) :

**Lemma 9.4.2.** *Under the hypotheses of Theorem 9.3.1, we have the following convergences for all $g \in C_c^\infty(]-\infty, T[\times]-1,1[)$ :*

$$\begin{aligned}
&\int_0^T \int_{-1}^1 \left\{\partial_x g \partial_u W\left(x, \partial_x \phi^N\right) - \partial_\tau g \partial_\tau \phi^N\right\}(\tau, x) dx d\tau \\
&- \int_{-1}^1 g(0,x)\phi_0^\tau(x) dx \underset{N\to+\infty}{\to} 0, \qquad (9.44) \\
&\int_0^T \int_{-1}^1 \left\{\partial_x g \partial_\tau \phi^N - \partial_\tau g \partial_x \phi^N\right\}(\tau, x) dx d\tau - \int_{-1}^1 g(0,x)\phi_0^x(x) dx \underset{N\to+\infty}{\to} 0. \qquad (9.45)
\end{aligned}$$

*Démonstration.* It is easy to prove (9.45) by an integration by parts :

$$\begin{aligned}
&\int_0^T \int_{-1}^1 \left\{\partial_x g \partial_\tau \phi^N - \partial_\tau g \partial_x \phi^N\right\}(\tau,x) dx d\tau \\
&= -\int_0^T \int_{-1}^1 \left\{g \partial_x \partial_\tau \phi^N - g \partial_x \partial_\tau \phi^N\right\}(\tau,x) dx d\tau + \int_{-1}^1 g(0,x) \partial_x \phi^N(0,x) dx \\
&\underset{N\to+\infty}{\to} \int_{-1}^1 g(0,x) \phi_0^x(x) dx.
\end{aligned}$$

Before proving (9.44), let us introduce the operators :

$$D_N^- : f(x) \mapsto N\left(f(x) - f\left(x - \frac{1}{N}\right)\right) \qquad D_N^+ : f(x) \mapsto N\left(f(x) - f\left(x + \frac{1}{N}\right)\right),$$

which are adjoint of each other.
We integrate by parts :

$$\begin{aligned}
&\int_0^T \int_{-1}^1 \left\{\partial_x g \partial_u W\left(x, \partial_x \phi^N\right) - \partial_\tau g \partial_\tau \phi^N\right\}(\tau,x) dx d\tau \\
&= \int_0^T \int_{-1}^1 \left\{\partial_x g \partial_u W\left(x, \partial_x \phi^N\right) + g \partial_\tau^2 \phi^N\right\}(\tau,x) dx d\tau \\
&\quad + \int_{-1}^1 g(0,x) \partial_\tau \phi^N(0,x) dx. \qquad (9.46)
\end{aligned}$$

It is clear that :

$$\int_{-1}^1 g(0,x) \partial_\tau \phi^N(0,x) dx \underset{N\to+\infty}{\to} \int_{-1}^1 g(0,x) \phi_0^\tau(x) dx. \qquad (9.47)$$

We focus on the other integrals. By definition, if $x \in [-1+1/N, 1-1/N]$ :

$$\begin{aligned}
\partial_\tau^2 \phi^N(\tau,x) &= N \frac{d^2}{dt^2}\left(\left(1-\theta^N(x)\right) X_{k^N(x)} + \theta^N(x) X_{k^N(x)+1}\right)(t) \\
&= \left(1-\theta^N(x)\right) D_N^- \left\{A\left(\frac{k^N(x)}{N}\right) U_{k^N(x)}(t) + B\left(\frac{k^N(x)}{N}\right)\right\} \\
&+ \theta^N(x) D_N^- \left\{A\left(\frac{k^N(x)+1}{N}\right) U_{k^N(x)+1}(t) + B\left(\frac{k^N(x)+1}{N}\right)\right\} \\
&= \left(1-\theta^N(x)\right) D_N^- \left\{A\left(\frac{k^N(x)}{N}\right) \partial_x \phi^N(\tau,x) + B\left(\frac{k^N(x)}{N}\right)\right\} \\
&+ \theta^N(x) D_N^- \left\{A\left(\frac{k^N(x)+1}{N}\right) \partial_x \phi^N(\tau,x+1/N) + B\left(\frac{k^N(x)+1}{N}\right)\right\}.
\end{aligned}$$

Remark that $D_N^\pm$ and $\theta^N$ commute, in the sense that :

$$D_N^\pm \left\{\theta^N(x) f(x)\right\} = \theta^N(x) D_N^\pm \left\{f(x)\right\}.$$

Hence :

$$\partial_\tau^2\phi^N(\tau,x) = D_N^-\Bigg\{ \left(1-\theta^N(x)\right)\left(A\left(\frac{k^N(x)}{N}\right)\partial_x\phi^N(\tau,x) + B\left(\frac{k^N(x)}{N}\right)\right) + \theta^N(x)\left(A\left(\frac{k^N(x)+1}{N}\right)\partial_x\phi^N(\tau,x+1/N) + B\left(\frac{k^N(x)+1}{N}\right)\right)\Bigg\},$$

if $|x| < 1-1/N$. We assume that $N$ is sufficiently large, so that Supp(g) $\subset [-1+2/\mathrm{N}, 1-2/\mathrm{N}]$. Then :

$$\begin{aligned}
&\int_0^T\int_{-1}^1 \partial_\tau^2\phi^N(\tau,x)g(\tau,x)dxd\tau \\
&= \int_0^T\int_{-1}^1 g(\tau,x)D_N^-\Bigg\{ \left(1-\theta^N(x)\right)\left(A\left(\frac{k^N(x)}{N}\right)\partial_x\phi^N(\tau,x) + B\left(\frac{k^N(x)}{N}\right)\right) \\
&\quad + \theta^N(x)\left(A\left(\frac{k^N(x)+1}{N}\right)\partial_x\phi^N(\tau,x+1/N) + B\left(\frac{k^N(x)+1}{N}\right)\right)\Bigg\}dxd\tau.
\end{aligned}$$

As $D_N^-$ and $D_N^+$ are adjoint of each other :

$$\begin{aligned}
&\int_0^T\int_{-1}^1 \partial_\tau^2\phi^N(\tau,x)g(\tau,x)dxd\tau \\
&= \int_0^T\int_{-1}^1 D_N^+\left\{g(\tau,x)\right\}\Bigg\{ \left(1-\theta^N(x)\right)\left(A\left(\frac{k^N(x)}{N}\right)\partial_x\phi^N(\tau,x) + B\left(\frac{k^N(x)}{N}\right)\right) \\
&\quad + \theta^N(x)\left(A\left(\frac{k^N(x+1/N)}{N}\right)\partial_x\phi^N(\tau,x+1/N) + B\left(\frac{k^N(x+1/N)}{N}\right)\right)\Bigg\}dxd\tau \\
&= \int_0^T\int_{-1}^1 \left[\left(1-\theta^N(x)\right)D_N^+\left\{g(\tau,x)\right\} + \theta^N(x)D_N^+\left\{g(\tau,x-1/N)\right\}\right] \\
&\quad \left(A\left(\frac{k^N(x)}{N}\right)\partial_x\phi^N(\tau,x) + B\left(\frac{k^N(x)}{N}\right)\right)dxd\tau.
\end{aligned}$$

Now, since $A$, $B$ and $g$ are regular :

$$\begin{aligned}
D_N^+\left\{g(\tau,x)\right\} &= -\partial_x g(\tau,x) + \frac{h_g^N(\tau,x)}{N}, \\
A\left(\frac{k^N(x)}{N}\right) &= A(x) + \frac{h_A^N(x)}{N}, \\
B\left(\frac{k^N(x)}{N}\right) &= B(x) + \frac{h_B^N(x)}{N},
\end{aligned}$$

where :

$$\sup_N \sup_{\tau,x}\left\{\left|h_g^N(\tau,x)\right| + \left|h_A^N(\tau,x)\right| + \left|h_B^N(\tau,x)\right|\right\} < +\infty.$$

Therefore :

$$\begin{aligned}&\int_0^T \int_{-1}^1 \partial_\tau^2 \phi^N(\tau,x) g(\tau,x) dx d\tau \\ &= -\int_0^T \int_{-1}^1 \partial_x g(\tau,x) \left( A(x) \partial_x \phi^N(\tau,x) + B(x) \right) dx d\tau + C^N, \end{aligned} \tag{9.48}$$

where, by the Cauchy-Schwarz inequality :

$$C^N \leq \frac{C}{\sqrt{N}} \left( 1 + \sqrt{\int_0^T \int_{-1}^1 \left( \partial_x \phi^N(\tau,x) \right)^2 dx d\tau} \right). \tag{9.49}$$

Using Lemma 9.4.1 and (9.33), we get that :

$$\int_{-1}^1 \left( \partial_x \phi^N(\tau,x) \right)^2 dx \leq C. \tag{9.50}$$

Whence, from (9.48), (9.49) and (9.50) we deduce :

$$\left| \int_0^T \int_{-1}^1 \left\{ \partial_\tau^2 \phi^N g + \partial_x g \partial_u \left( W(x, \partial_x \phi^N) \right) \right\} (\tau,x) dx d\tau \right| \underset{N \to +\infty}{\to} 0. \tag{9.51}$$

From (9.46), (9.47) and (9.51), we obtain (9.44). □

We are now able to prove Theorem 9.3.1.

*Proof of Theorem 9.3.1.* Using Lemma 9.4.1 and (9.33), we estimate :

$$\begin{aligned}&\sum_{k=-N}^{N-1} \left\{ A\left(\frac{k}{N}\right) U_k^2 + V_k^2 \right\} (t) \\ &\overset{(9.33)}{\leq} C + C \sum_{k=-N}^{N-1} \left\{ A\left(\frac{k}{N}\right) (U_k)^2 + B\left(\frac{k}{N}\right) U_k + (V_k)^2 \right\} (t) \\ &\leq C \left( 1 + \widetilde{\mathcal{E}}_D^N(t) \right) \\ &\leq C \left( 1 + \widetilde{\mathcal{E}}_D^N(0) \right). \end{aligned}$$

Since $\phi_0^x$ and $\phi_0^\tau$ are sufficiently regular, we have :

$$\widetilde{\mathcal{E}}_D^N(0) \underset{N \to \infty}{\to} \int_{-1}^1 W\left(x, \phi_0^x(x)\right) dx + \frac{1}{2} \int_{-1}^1 \left( \phi_0^\tau(x) \right)^2 dx.$$

We therefore obtain that (using (9.17)), for all $\tau > 0$ :

$$\int_{-1}^1 \left\{ A\left( \frac{k^N(x)}{N} \right) \left( \partial_x \phi^N(\tau,x) \right)^2 + \left( \partial_\tau \phi^N(\tau,x) \right)^2 \right\} dx \leq C.$$

And by smoothness of $A$ and thanks to the fact that $A \geq \alpha > 0$, we obtain :

$$\int_{-1}^{1} \left\{ A(x) \left(\partial_x \phi^N(\tau,x)\right)^2 + \left(\partial_\tau \phi^N(\tau,x)\right)^2 \right\} dx \leq C. \tag{9.52}$$

We take the following scalar product on $\dot{H}^1_{\tau,x}$ :

$$\langle g,h \rangle_{\widetilde{\dot{H}}^1_{\tau,x}} := \int_0^T \int_{-1}^1 \left\{ A(x)\partial_x g \partial_x h + \partial_\tau g \partial_\tau h \right\}(\tau,x) dx d\tau.$$

This scalar product induces a norm which is equivalent to the classical one on $\dot{H}^1_{\tau,x}$, as $A \geq \alpha > 0$ and $A$ is bounded. We denote $\widetilde{\dot{H}}^1_{\tau,x}$ for $\dot{H}^1_{\tau,x}$ endowed with this scalar product, respectively $\widetilde{H}^1_{\tau,x}$ for $H^1_{\tau,x}$ endowed with the scalar product :

$$\langle g,h \rangle_{\widetilde{H}^1_{\tau,x}} = \langle g,h \rangle_{\widetilde{\dot{H}}^1_{\tau,x}} + \int_0^T \int_{-1}^1 g(\tau,x)h(\tau,x) dx d\tau.$$

From (9.52) and the fact that $\phi^N(-1) = \phi_l$, we get by the Poincaré inequality that $\phi^N$ is bounded in $\widetilde{H}^1_{\tau,x}$. By weak compactness of this space, we extract :

$$\phi^N \rightharpoonup \phi_\infty \text{ in } \widetilde{H}^1_{\tau,x}. \tag{9.53}$$

Lemma 9.4.2 implies that, for all $g \in C_c^\infty([0,T[,]-1,1[)$ :

$$\int_0^T \int_{-1}^1 \left\{ \partial_x g \partial_u W\left(x, \partial_x \phi^\infty\right) - \partial_\tau g \partial_\tau \phi_\infty \right\}(\tau,x) dx d\tau = \int_{-1}^1 g(0,x)\phi_0^\tau(x) dx,$$
$$\int_0^T \int_{-1}^1 \left\{ \partial_x g \partial_\tau \phi_\infty - \partial_\tau g \partial_x \phi_\infty \right\}(\tau,x) dx d\tau = \int_{-1}^1 g(0,x)\phi_0^x(x) dx.$$

Therefore, $\phi_\infty$ is $\phi$, the unique solution of (9.5) for the initial and boundary conditions (9.6) and (9.7) given by Theorem 9.2.1.
We now prove that this convergence is strong. As $\widetilde{\mathcal{E}}_D^N$ is preserved (and thanks to (9.17)), we have :

$$\begin{aligned} \left\|\phi^N\right\|^2_{\widetilde{\dot{H}}^1_{\tau,x}} \leq & 2T\left(\widetilde{\mathcal{E}}_D^N(0)\right)^2 + 2\int_0^T \int_{-1}^1 \left\{ \left( A\left(\frac{k^N(x)}{N}\right) - A(x)\right)\left(\partial_x \phi^N\right)^2 \right. \\ & \left. - B\left(\frac{k^N(x)}{N}\right)\partial_x \phi^N \right\}(\tau,x) dx \end{aligned} \tag{9.54}$$

Yet, as $A$ and $B$ are $C^1$, we have :

$$\lim_{N\to+\infty} \sup_{x\in[-1,1]} \left| A\left(\frac{k^N(x)}{N}\right) - A(x) \right| = 0, \qquad \text{and} \qquad B\left(\frac{k^N(x)}{N}\right) \to B(x) \text{ in } L^2_x.$$

Moreover :

$$\widetilde{\mathcal{E}}_D^N(0) \underset{N\to\infty}{\to} \int_{-1}^{1} \left\{ \frac{A(x)}{2} \left(\phi_0^x(x)\right)^2 + B(x)\phi_0^x(x) + \frac{1}{2}\left(\phi_0^\tau(x)\right)^2 \right\} dx.$$

Therefore, thanks to (9.54) :

$$\begin{aligned} \limsup_{N\to+\infty} \left\|\phi^N\right\|^2_{\widetilde{H}^1_{\tau,x}} \leq & 2T \int_{-1}^{1} \left\{ \frac{A(x)}{2} \left(\phi_0^x(x)\right)^2 + B(x)\phi_0^x(x) + \frac{1}{2}\left(\phi_0^\tau(x)\right)^2 \right\} dx \\ & - 2\int_0^T \int_{-1}^{1} \left\{ B\left(x\right)\partial_x\phi \right\}(\tau,x)dx. \end{aligned} \tag{9.55}$$

But, from Theorem 9.2.1, the continuous energy $\mathcal{E}_C$ is also preserved. This implies :

$$\begin{aligned} & 2T \int_{-1}^{1} \left\{ \frac{A(x)}{2} \left(\phi_0^x(x)\right)^2 + B(x)\phi_0^x(x) + \frac{1}{2}\left(\phi_0^\tau(x)\right)^2 \right\} dx \\ & - 2\int_0^T \int_{-1}^{1} \left\{ B\left(x\right)\partial_x\phi \right\}(\tau,x)dx \\ & = \int_0^T \int_{-1}^{1} \left\{ \frac{A(x)}{2} \left(\partial_x\phi\right)^2 + \frac{1}{2}\left(\partial_\tau\phi\right)^2 \right\}(\tau,x)dx. \end{aligned} \tag{9.56}$$

From (9.55) and (9.56), we obtain :

$$\limsup_{N\to+\infty} \left\|\phi^N\right\|_{\widetilde{H}^1_{\tau,x}} \leq \|\phi\|_{\widetilde{H}^1_{\tau,x}}.$$

Whence $\phi^N$ strongly converges to $\phi$ in $\widetilde{H}^1_{\tau,x}$ (and as a consequence, in $H^1_{\tau,x}$). □

## 9.5 The non-linear case

This section is devoted to the proof of Theorem 9.3.2.

*Remark* 14 (Boundedness of $\partial_x\phi$). Under the hypotheses of Theorem 9.3.2, if $\partial_x\phi$ does not belong to $[a,b]$ on a non-zero measure set, then $\partial_x\phi^N$ cannot converge weakly to $\partial_x\phi$. Therefore, we henceforth assume that $\partial_x\phi(\tau,x) \in [a,b]$, $\forall x \in [-1,1], \forall \tau < T_0$.

### 9.5.1 Light cone

The system (9.1) has the property that perturbations propagate at a finite speed on the macroscopic level. This is stated in Theorem 9.3.3, but before proving it, we have to derive a Grönwall-type estimate :

**Proposition 9.5.1.** *Under the hypotheses of Theorem 9.3.3, if, for fixed $N$ :*

$$M = \left\| U_0(.) - \widetilde{U}_0(.) \right\|_{L^\infty_t} < +\infty, \tag{9.57}$$

*then we have, for all* $j \in [\![1, N-1]\!]$, $t \in \mathbb{R}_+$ *:*

$$\left|U_j(t) - \widetilde{U}_j(t)\right| \leq M \frac{\left(2t\sqrt{K}\right)^{2j}}{(2j)!} \exp\left(2t\sqrt{K}\right), \tag{9.58}$$

$$\left|V_j(t) - \widetilde{V}_j(t)\right| \leq M\sqrt{K} \exp\left(2t\sqrt{K}\right) \left[\frac{\left(2t\sqrt{K}\right)^{2j+1}}{(2j+1)!} + \frac{\left(2t\sqrt{K}\right)^{2j-1}}{(2j-1)!}\right]. \tag{9.59}$$

*Proof.* Remark first that it is straightforward to get (9.59) from (9.58) ((9.58) also holds for $j = 0$) by integrating (9.1). Indeed, using (9.1), we have, for $j \in [\![1, N-2]\!]$ :

$$\begin{aligned}
\left|V_j(t) - \widetilde{V}_j(t)\right| &\leq \int_0^t \left|W'\left(U_j(s)\right) - W'\left(\widetilde{U}_j(s)\right) + W'\left(\widetilde{U}_{j-1}(s)\right) - W'\left(U_{j-1}(s)\right)\right| ds \\
&\overset{(9.39)}{\leq} K \int_0^t \left\{\left|\widetilde{U}_j(s) - U_j(s)\right| + \left|\widetilde{U}_{j-1}(s) - U_{j-1}(s)\right|\right\} ds \\
&\overset{(9.58)}{\leq} MK \exp\left(2t\sqrt{K}\right) \left[\frac{t\left(2t\sqrt{K}\right)^{2j}}{(2j+1)!} + \frac{t\left(2t\sqrt{K}\right)^{2(j-1)}}{(2j-1)!}\right].
\end{aligned}$$

We now prove the estimate (9.58). We do it by induction on $j$ in the expression :

$$S_j(t) := \max_{k \in [\![j, N-1]\!]} \left|U_k(t) - \widetilde{U}_k(t)\right|.$$

Using (9.1) and (9.39), we get, for $j \geq 1$ :

$$\begin{aligned}
\left|U_j(t) - \widetilde{U}_j(t)\right| \leq& \int_0^t \left|V_{j+1}(s) - \widetilde{V}_{j+1}(s) + \widetilde{V}_j(s) - V_j(s)\right| ds \\
\leq& K \int_0^t \int_0^s \left\{\left|U_{j+1} - \widetilde{U}_{j+1}\right| + 2\left|U_j - \widetilde{U}_j\right| + \left|U_{j-1} - \widetilde{U}_{j-1}\right|\right\}(r) dr ds, \\
\leq& 4K \int_0^t \int_0^s S_{j-1}(r) dr ds.
\end{aligned}$$

Thus :

$$S_j(t) \leq 4K \int_0^t \int_0^s S_{j-1}(r) dr ds. \tag{9.60}$$

From hypothesis (9.57), we have that $S_0(0) \leq M$. Using the same argument as above, we get for $j = 0$ :

$$S_0(t) \leq M + 4K \int_0^t \int_0^s S_0(r) dr ds.$$

Using the Grönwall Lemma, we obtain :

$$S_0(t) \leq M \exp\left(2t\sqrt{K}\right). \tag{9.61}$$

Therefore, we obtain from (9.60) that :

$$\begin{aligned} S_j(t) \leq & (4K)^j \int_0^t \int_0^{s_j} \int_0^{t_{j-1}} \int_0^{s_{j-1}} ... \int_0^{t_1} \int_0^{s_1} S_0(t_0) dt_0 ds_1 dt_1 ... ds_{j-1} dt_{j-1} ds_j \\ \leq & (4K)^j \int_0^t \frac{(t-t_0)^{2j-1}}{(2j-1)!} S_0(t_0) dt_0 \\ \leq & M (4K)^j \frac{t^{2j}}{(2j)!} \exp\left(2t\sqrt{K}\right), \end{aligned}$$

which concludes the proof. □

We are now able to prove Theorem 9.3.3 :

*Proof of Theorem 9.3.3.* Let $j > Nx$, $t \in [0, N\tau[$. By strong convexity of $W$ and by the Cauchy-Schwarz inequality, (9.38) implies :

$$M_N := \sup_{t \in \mathbb{R}_+} \left| U_0^N(t) \right| \leq C\sqrt{N}.$$

By Proposition 9.5.1, we have :

$$\left| U_j(t) - \widetilde{U}_j(t) \right| \leq M_N \frac{\left(2t\sqrt{K}\right)^{2j}}{(2j)!} \exp\left(2t\sqrt{K}\right).$$

Using logarithm and Stirling formula, we obtain :

$$\begin{aligned} \ln\left(N \left| U_j(t) - \widetilde{U}_j(t) \right|\right) = & \ln(M_N) + 2j \ln\left(2t\sqrt{K}\right) - \ln((2j)!) + 2t\sqrt{K} + \ln(N) \\ \leq & C(1 + \ln(N)) + 2j \ln\left(\frac{2et\sqrt{K}}{2j}\right) + 2t\sqrt{K}, \\ \leq & C(1 + \ln(N)) + 2j \ln\left(\frac{\exp(2)t\sqrt{K}}{j}\right) + (2t\sqrt{K} - 2j) \\ \leq & C(1 + \ln(N)) + 2j \ln\left(\frac{cN\tau}{Nx}\right) + 2\left(\frac{c}{\exp(2)} N\tau - Nx\right) \\ \leq & C(1 + \ln(N)) + 2N \left(x \ln\left(\frac{c\tau}{x}\right) + \frac{c}{\exp(2)}\tau - x\right). \end{aligned}$$

As $c\tau < x$, we obtain that :

$$\lim_{N \to +\infty} \sup_{0 \leq t < N\tau} \max_{j > Nx} \ln\left(N \left| U_j^N(t) - \widetilde{U}_j^N(t) \right|\right) = -\infty,$$

which gives (9.40). The same method applies for (9.41). □

### 9.5.2 Strengthened convergence

In this section, we prove that if $\phi^N$ converges weakly to $\phi$ in $H^1_{\tau,x}$, then convergence holds in a strong sense for $\partial_x\phi^N$ to $\partial_x\phi$ in $L^2_{\tau,x}$.

**Lemma 9.5.2.** *Under the hypotheses of Theorem 9.3.2, if $T$ is DC-compatible, then the following implication is true :*

$$\phi^N \rightharpoonup \phi \text{ in } H^1_{\tau,x} \qquad \Longrightarrow \qquad \partial_x\phi^N \to \partial_x\phi \text{ in } L^2_{\tau,x}.$$

To get it, we first prove two parallel integral identities, for the continuous system (9.5) and the discrete one (9.1).

**Lemma 9.5.3.** *Assume that $W \in C^1(\mathbb{R})$. Assume that $\phi \in W^{1,\infty}_{\tau,x}$ is a weak solution of (9.5) with initial and boundary conditions (9.6) and (9.7), such that $\phi_0^x$ and $\phi_0^\tau$ satisfy (9.29). Then, if $T$ is C-compatible, we have :*

$$\begin{aligned}&\int_0^T \int_{-1}^1 (1-x)\left(\partial_\tau\phi(\tau,x) - \phi_0^\tau(x)\right) dxd\tau \\ &= \int_0^T \int_{-1}^1 (T-\tau)\left(W'\left(\partial_x\phi(\tau,x)\right) - W'\left(\phi_0^x(-1)\right)\right) dxd\tau. \end{aligned} \tag{9.62}$$

A similar identity can be derived for the discrete system :

**Lemma 9.5.4.** *Let $W \in C^1(\mathbb{R})$. Let $X_j(t)$ be the solution of (9.1) for the initial and boundary conditions (9.2) and (9.3). Suppose that $T$ is D-compatible. Assume that :*

$$\partial_\tau\phi^N \rightharpoonup \xi_\infty \text{ in } L^2_{\tau,x}, \tag{9.63}$$

$$\partial_x\phi^N \rightharpoonup \psi_\infty \text{ in } L^2_{\tau,x}. \tag{9.64}$$

*Assume that $\partial_x\phi^N$ is uniformly bounded in $\tau, x, N$. Let $\nu_{\tau,x}$ be the Young measure associated with the convergence of $\partial_x\phi^N$ to $\psi_\infty$. Then, we have :*

$$\begin{aligned}&\int_0^T \int_{-1}^1 (1-x)\left(\xi_\infty(\tau,x) - \phi_0^\tau(x)\right) dxd\tau \\ &= \int_0^T \int_{-1}^1 (T-\tau)\left\{\int_{\mathbb{R}} W'(\lambda) d\nu_{\tau,x}(\lambda) - W'(\phi_0^x(-1))\right\} d\tau dx. \end{aligned} \tag{9.65}$$

See Theorem 3.1 p 31 of [118] for the definition of Young measures.

We temporarily admit Lemma 9.5.3 and 9.5.4 and proceed with the proof of Lemma 9.5.2.

*Proof of Lemma 9.5.2.* Using Lemmata 9.5.3 and 9.5.4, we get that, if $\nu_{\tau,x}$ characterizes the weak convergence of $\partial_x\phi^N$ to $\partial_x\phi$ :

$$\int_0^T \int_{-1}^1 (T-\tau)\left\{W'\left(\partial_x\phi(\tau,x)\right) - \int_{\mathbb{R}} W'(\lambda) d\nu_{\tau,x}(\lambda)\right\} dxd\tau = 0.$$

But, as $W'$ is strongly convex, using Theorem 1.1.8 p 47 of [123], we obtain :

$$\nu_{\tau,x} = \delta_{\partial_x \phi(\tau,x)}.$$

Then, Corollary 3.2 p 34 of [118] implies :

$$\partial_x \phi^N \to \partial_x \phi \text{ in } L^2_{\tau,x}.$$

□

Next, we prove Lemma 9.5.3 :

*Proof of Lemma 9.5.3.* We claim that, for all $g \in C^\infty([0,T] \times [-1,1])$ such that $g(T,\cdot) = 0$, we have :

$$\begin{aligned} &\int_0^T \int_{-1}^1 \left\{ \partial_x g W'(\partial_x \phi) - \partial_\tau g \partial_\tau \phi \right\}(\tau,x) d\tau dx \\ &= \int_{-1}^1 g(0,x) \phi_0^\tau(x) dx + \int_0^T \left( g(1,\tau) W'(\phi_0^x(1)) - g(-1,\tau) W'(\phi_0^x(-1)) \right) d\tau. \end{aligned} \tag{9.66}$$

Inserting $g(\tau,x) := (T-\tau)(1-x)$ in (9.66) implies (9.62).

We prove now (9.66) by a density argument. For that purpose, we introduce a small parameter $\delta > 0$, $v_\delta \in C_c^\infty(]-1,1[)$ that is equal to 1 in $[-1+\delta, 1-\delta]$, such that $0 \leq v_\delta \leq 1$ and $\|v_\delta'\|_{L^\infty} \leq C\delta^{-1}$, and $w_\delta \in C_c^\infty([0,T[)$ that is equal to 1 in $[0, T-\delta]$ such that $0 \leq w_\delta \leq 1$ and $\|w_\delta'\|_{L^\infty} \leq C\delta^{-1}$. We set $\widetilde{g} : (\tau,x) \mapsto g(\tau,x) v_\delta(x) w_\delta(\tau)$. Hence, $\widetilde{g} \in C_c^\infty([0,T[\times]-1,1[)$. Using it as a test function in (9.23), we obtain :

$$\begin{aligned} &\int_0^T \int_{-1}^1 v_\delta(x) w_\delta(\tau) \left\{ \partial_x g W'(\partial_x \phi) - \partial_\tau g \partial_\tau \phi \right\}(\tau,x) d\tau dx \\ &+ \int_0^T \int_{-1}^1 \left( v_\delta'(x) w_\delta(\tau) \left\{ g W'(\partial_x \phi) \right\}(\tau,x) - v_\delta(x) w_\delta'(\tau) \left\{ g \partial_\tau \phi \right\}(\tau,x) \right) d\tau dx \\ &= \int_{-1}^1 g(0,x) v_\delta(x) \phi_0^\tau(x) dx. \end{aligned} \tag{9.67}$$

By the dominated convergence theorem, we have :

$$\int_{-1}^1 g(0,x) v_\delta(x) \phi_0^\tau(x) dx \underset{\delta \to 0}{\to} \int_{-1}^1 g(0,x) \phi_0^\tau(x) dx, \tag{9.68}$$

and :

$$\begin{aligned} &\int_0^T \int_{-1}^1 v_\delta(x) w_\delta(\tau) \left\{ \partial_x g W'(\partial_x \phi) - \partial_\tau g \partial_\tau \phi \right\}(\tau,x) d\tau dx \\ &\underset{\delta \to 0}{\to} \int_0^T \int_{-1}^1 \left\{ \partial_x g W'(\partial_x \phi) - \partial_\tau g \partial_\tau \phi \right\}(\tau,x) d\tau dx. \end{aligned} \tag{9.69}$$

By $C$-compatibility of $T$, one can take $\delta$ such that $\partial_x\phi(\tau,x)$ is constant on $[1-\delta,1]$ and on $[-1,-1+\delta]$. Moreover, $v'_\delta(x)=0$ if $x\in[-1+\delta,1-\delta]$. For such $\delta$, one has, by integration by parts :

$$\begin{aligned}\int_{-1}^{1} v'_\delta(x)\left\{gW'(\partial_x\phi)\right\}(\tau,x)dx =& g(\tau,-1+\delta)W'\left(\phi_0^x(-1)\right)-g(\tau,1-\delta)W'\left(\phi_0^x(1)\right)\\ &-\int_{-1}^{-1+\delta}\partial_x g(\tau,x)v_\delta(x)W'\left(\phi_0^x(-1)\right)dx\\ &-\int_{1-\delta}^{1}\partial_x g(\tau,x)v_\delta(x)W'\left(\phi_0^x(1)\right)dx.\end{aligned}$$

Applying the dominated convergence theorem again, one obtains from the above expression that :

$$\int_{-1}^{1} v'_\delta(x)\left\{gW'(\partial_x\phi)\right\}(\tau,x)dxd\tau \underset{\delta\to 0}{\to} g(\tau,-1)W'\left(\phi_0^x(-1)\right)-g(\tau,1)W'\left(\phi_0^x(1)\right). \quad (9.70)$$

Newt, we deal with the last integral of (9.67) :

$$\int_0^T\int_{-1}^{1}\left|v_\delta(x)w'_\delta(\tau)g(\tau,x)\partial_\tau\phi(\tau,x)\right|dxd\tau \le \int_{T-\delta}^{T}\int_{-1}^{1} C\delta^{-1}\left|g(\tau,x)\right|\left\|\partial_\tau\phi\right\|_{L^\infty_{\tau,x}}dxd\tau.$$

Using the fact that $g(T,\cdot)=0$ and that $g$ is continuous, we infer :

$$\int_0^T\int_{-1}^{1}\left|v_\delta(x)w'_\delta(\tau)g(\tau,x)\partial_\tau\phi(\tau,x)\right|dxd\tau\to 0. \quad (9.71)$$

Therefore, we deduce (9.66) from (9.67), (9.68), (9.69), (9.70) and (9.71), and conclude the proof.

□

We now prove Lemma 9.5.4 :

*Proof of Lemma 9.5.4.* We sum up (9.1) and get :

$$\sum_{j=-N}^{k}\frac{d^2}{dt^2}X_j(t)=\sum_{j=-N+1}^{k}W'(U_j)(t)-W'(U_{j-1})(t)=W'(U_k)(t)-W'(U_{-N})(t).$$

Then we multiply the above expression by $(NT-t)$ and integrate with respect to $t$ :

$$\begin{aligned}\int_0^{NT}(NT-t)\left(W'(U_k)-W'(U_{-N})\right)(t)dt &= \int_0^{NT}\int_0^t\sum_{i=-N}^{k}\frac{d^2}{dt^2}X_j(s)dsdt\\ &=\sum_{j=-N}^{k}\int_0^{NT}\left\{\frac{dX_j}{dt}(t)-\frac{d}{dt}X_j(0)\right\}dt.\end{aligned}$$

If we rescale it and sum over $k \in [\![-N, N-1]\!]$, we obtain :

$$\begin{aligned}
&\frac{1}{N}\sum_{k=-N}^{N-1}\int_0^T (T-\tau)\left(W'\left(\partial_x\phi^N\left(\tau,\frac{k}{N}\right)\right) - W'\left(U_{-N}(N\tau)\right)\right)d\tau \\
&= \frac{1}{N^2}\sum_{k=-N}^{N-1}\sum_{j=-N}^{k}\int_0^T \left\{\xi^N\left(\tau,\frac{j}{N}\right) - \xi^N\left(0,\frac{j}{N}\right)\right\}d\tau \\
&= \frac{1}{N}\sum_{k=-N}^{N-1}\frac{(N-k)}{N}\int_0^T \left\{\xi^N\left(\tau,\frac{k}{N}\right) - \xi^N\left(0,\frac{k}{N}\right)\right\}d\tau \\
&= \int_0^T\int_{-1}^1 \left(1-\frac{\lfloor Nx\rfloor}{N}\right)\left(\xi^N(\tau,x) - \xi^N(0,x)\right)dxd\tau. \qquad (9.72)
\end{aligned}$$

Remark first that :

$$1-\frac{\lfloor Nx\rfloor}{N} \to 1-x \text{ in } L^\infty_x.$$

From initial conditions (9.2), from (9.19) and from (9.63), we get that :

$$\xi^N(\tau,.) - \xi^N(0,.) \rightharpoonup \xi^\infty(\tau,.) - \phi_0^\tau(.) \text{ in } L^2_x.$$

Therefore :

$$\begin{aligned}
&\int_0^T\int_{-1}^1 \left(1-\frac{\lfloor Nx\rfloor}{N}\right)\left(\xi^N(\tau,x) - \xi^N(0,x)\right)dxd\tau \\
&\to \int_0^T\int_{-1}^1 (1-x)\left(\xi_\infty(\tau,x) - \phi_0^\tau(x)\right)dxd\tau. \qquad (9.73)
\end{aligned}$$

As $T$ is $D$-compatible, we have the following convergence, uniform for $\tau \in [0,T]$ :

$$W'\left(U_{-N}(N\tau)\right) \to W'\left(\phi_0^x(-1)\right). \qquad (9.74)$$

From Hypothesis (9.64), we get that, by definition of $\nu$ :

$$\begin{aligned}
&\frac{1}{N}\sum_{k=-N}^{N-1}\int_0^T (T-\tau)W'\left(\partial_x\phi^N\left(\tau,\frac{k}{N}\right)\right)d\tau \\
&\to \int_0^T\int_{-1}^1 (T-\tau)\left(\int_{\mathbb{R}} W'(\lambda)d\nu_{\tau,x}(\lambda)\right)d\tau dx. \qquad (9.75)
\end{aligned}$$

Finally, (9.72), (9.73), (9.74) and (9.75) imply (9.65). □

### 9.5.3 From strong convergence of $\partial_x\phi^N$ to strong convergence of $\partial_\tau\phi^N$

We now prove that the strong convergence of $\partial_x\phi^N$ in $L^2_{\tau,x}$ implies the strong convergence of $\partial_\tau\phi^N$ in $L^2_{\tau,x}$.

**Lemma 9.5.5.** *Under the hypotheses of Theorem 9.3.2, if $T$ is DC-compatible, then the following implication is true :*

$$\left\{\begin{array}{ll}\phi^N \to \phi & \text{in } L^2_{\tau,x}\\ \partial_x\phi^N \to \partial_x\phi & \text{in } L^2_{\tau,x}\\ \partial_\tau\phi^N \rightharpoonup \partial_\tau\phi & \text{in } L^2_{\tau,x}\end{array}\right. \quad\Longrightarrow\quad \left\{\begin{array}{ll}\partial_\tau\phi^N \to \partial_\tau\phi & \text{in } L^2_{\tau,x}\\ \xi^N \to \partial_\tau\phi & \text{in } L^2_{\tau,x}\end{array}\right. ,$$

*for $\xi^N(\tau,x) = V^N_{k^N(x)}(N\tau)$.*

We can now proceed with the proof of Lemma 9.5.5.

*Proof of Lemma 9.5.5.* By definition :

$$\zeta^N(\tau,x) = \left(\left(1-\theta^N(x)\right)\frac{d}{dt}X_{k^N(x)} + \theta^N(x)\frac{d}{dt}X_{k^N(x)+1}\right)(t). \tag{9.76}$$

Using (9.1), we obtain :

$$\begin{aligned}\partial_\tau\zeta^N(\tau,x) =& N\left(1-\theta^N(x)\right)\left(W'\left(U_{k(x)}\right) - W'\left(U_{k(x)-1}\right)\right)(t)\\ &+ N\theta^N(x)\left(W'\left(U_{k(x)+1}\right) - W'\left(U_{k(x)}\right)\right)(t).\end{aligned}$$

We define :

$$\Xi(x) := \left\{\begin{array}{lll}0 & \text{if} & x<0,\\ \frac{x^2}{2} & \text{if} & x\in[0,1],\\ \frac{1}{2} - (x-1)^2 + (x-1) & \text{if} & x\in[1,2],\\ \frac{1}{2} + \frac{(x-2)^2}{2} - (x-2) & \text{if} & x\in[2,3],\\ 0 & \text{if} & x>3.\end{array}\right.$$

Let $\Xi^N_j(x) := \Xi(Nx-j)$. We have :

$$\frac{d}{dx}\Xi^N_j(x) := \left\{\begin{array}{lll}0 & \text{if} & x<j/N,\\ N\theta^N(x) & \text{if} & x\in[j/N,(j+1)/N],\\ N\left(1-2\theta^N(x)\right) & \text{if} & x\in[(j+1)/N,(j+2)/N],\\ N\left(\theta^N(x)-1\right) & \text{if} & x\in[(j+2)/N,(j+3)/N],\\ 0 & \text{if} & x>(j+3)/N.\end{array}\right.$$

Next, we define :

$$\Psi^N(\tau,x) := \sum_{j=-N}^{N-1}\Xi^N_{j-1}(x)W'\left(U_j(N\tau)\right).$$

We have :

$$\partial_x\Psi^N(\tau,x) = \partial_\tau\zeta^N(\tau,x), \qquad \forall\tau\in[0,T], \forall x\in[-1,1].$$

Since $\partial_x\phi^N \to \partial_x\phi$ in $L^2_{\tau,x}$, and since $\partial_x\phi^N$ is bounded in $L^\infty$, we have :

$$W'\left(\partial_x\phi^N\right) \to W'\left(\partial_x\phi\right) \text{ in } L^2_{\tau,x}.$$

We claim that it implies :

$$\left\|\Psi^N - W'\left(\partial_x\phi\right)\right\|_{L^2_{\tau,x}} \underset{N\to+\infty}{\to} 0. \tag{9.77}$$

Indeed, as $\partial_x\phi^N$ is bounded in $L^\infty_{\tau,x}$ uniformly in $N$, it suffices to bound as follows :

$$\begin{aligned}
&\int_{-1}^{1}\left|\Psi^N(\tau,x) - W'\left(\partial_x\phi(\tau,x)\right)\right|dx\\
&\leq \int_{-1}^{1}\left|\sum_{j=-N}^{N-1}\Xi^N_{j-1}(x)\left(W'\left(U_j(N\tau)\right) - W'\left(\partial_x\phi(\tau,x)\right)\right)\right|dx\\
&\leq \int_{-1}^{1}\sum_{j=-N}^{N-1}\Xi^N_{j-1}(x)\left|W'\left(U_j(N\tau)\right) - W'\left(\partial_x\phi(\tau,x)\right)\right|dx\\
&\leq \int_{-1}^{1}\sum_{j=-N}^{N-1}\mathbb{1}_{\left[\frac{j-1}{N},\frac{j+2}{N}\right]}(x)\left|W'\left(U_j(N\tau)\right) - W'\left(\partial_x\phi(\tau,x)\right)\right|dx\\
&\leq \sum_{j=-1}^{1}\int_{-1}^{1}\left|W'\left(U_{k(x)+j}(N\tau)\right) - W'\left(\partial_x\phi(\tau,x)\right)\right|dx\\
&\leq \sum_{j=-1}^{1}\int_{-1}^{1}\left|W'\left(\partial_x\phi^N\left(\tau,x+\frac{j}{N}\right)\right) - W'\left(\partial_x\phi\left(\tau,x\right)\right)\right|dx\\
&\to 0.
\end{aligned}$$

Interpolating with $L^\infty_{\tau,x}$, this gives (9.77). We define :

$$\alpha^N := \zeta^N - \zeta, \qquad \beta^N := \Psi - W'\left(\partial_x\phi\right), \qquad \gamma^N := \partial_x\phi^N - \partial_x\phi.$$

Remark that, by definition, we have :

$$\partial_\tau\alpha^N = \partial_x\beta^N, \tag{9.78}$$

$$\partial_x\alpha^N = \partial_\tau\gamma^N. \tag{9.79}$$

$D$-compatibility of $T$ implies the following convergences :

$$\int_0^T\left(\alpha^N(\tau,-1)\right)^2 \to 0, \quad \int_0^T\left(\alpha^N(\tau,1)\right)^2 d\tau \to 0, \quad \text{and} \quad \int_{-1}^1\left(\alpha^N(0,x)\right)^2 dx \to 0. \tag{9.80}$$

From (9.77), and by $D$-compatibility, we deduce that :

$$\lim_{N\to+\infty}\left\|\beta^N\right\|_{L^2_{\tau,x}} = 0, \quad \text{and} \quad \lim_{N\to+\infty}\left\{\left\|\beta^N(.,1)\right\|_{L^2_\tau} + \left\|\beta^N(.,-1)\right\|_{L^2_\tau}\right\} = 0. \tag{9.81}$$

The energy estimates for (9.1) and (9.5) give :

$$\sup_N\left\|\gamma^N\right\|_{L^2_{\tau,x}} \leq C, \quad \text{and} \quad \sup_N\left\{\left\|\gamma^N(0,.)\right\|_{L^2_x} + \left\|\gamma^N(T,.)\right\|_{L^2_x}\right\} \leq C. \tag{9.82}$$

We now claim that (9.78), (9.79), (9.81), (9.82), (9.80) imply :

$$\limsup_{N\to+\infty} \left\|\alpha^N\right\|_{L^2_{\tau,x}} = 0, \tag{9.83}$$

which gives the desired result, thanks to (9.18). To prove (9.83), we write :

$$\alpha^N(\tau,x) \overset{(9.78)}{=} \alpha^N(\tau,-1) + \int_{-1}^{x} \partial_\tau \gamma^N(\tau,y)dy$$
$$\text{and} \qquad \alpha^N(\tau,x) \overset{(9.79)}{=} \alpha^N(0,x) + \int_0^{\tau} \partial_x \beta^N(\nu,x)d\nu.$$

Therefore :

$$\begin{aligned}
&\int_0^T \int_{-1}^1 \left(\alpha^N(\tau,x)\right)^2 dxd\tau \\
&= \int_0^T \int_{-1}^1 \left(\alpha^N(\tau,-1) + \int_{-1}^x \partial_\tau\gamma^N(\tau,y)dy\right)\left(\alpha^N(0,x) + \int_0^\tau \partial_x\beta^N(\nu,x)d\nu\right)dxd\tau \\
&= \int_0^T \int_{-1}^1 \alpha^N(\tau,-1)\alpha^N(0,x)dxd\tau + \int_0^T\int_{-1}^1 \alpha^N(\tau,-1)\int_0^\tau \partial_x\beta^N(\nu,x)d\nu dxd\tau \\
&\quad + \int_0^T\int_{-1}^1 \alpha^N(0,x)\int_{-1}^x \partial_\tau\gamma^N(\tau,y)dydxd\tau \\
&\quad + \int_0^T\int_{-1}^1\int_{-1}^x \partial_\tau\gamma^N(\tau,y)dy\int_0^\tau \partial_x\beta^N(\nu,x)d\nu dxd\tau \\
&=: T_1^N + T_2^N + T_3^N + T_4^N.
\end{aligned}$$

We deal separately with $T_1^N, T_2^N, T_3^N, T_4^N$. By the Cauchy-Schwarz inequality :

$$T_1^N \le \left(\int_0^T \left|\alpha^N(\tau,-1)\right|^2 d\tau\right)^{1/2}\left(\int_{-1}^1 \left|\alpha^N(0,x)\right|^2 dx\right)^{1/2} \overset{(9.80)}{\to} 0. \tag{9.84}$$

Integrating over $x$ in $T_2^N$, we obtain :

$$T_2^N = \int_0^T \alpha^N(\tau,-1)\int_0^\tau \left(\beta^N(\nu,1) - \beta^N(\nu,-1)\right)d\nu d\tau \overset{(9.81),(9.80)}{\to} 0. \tag{9.85}$$

Next, integrating over $\tau$ in $T_3^N$, we get :

$$T_3^N = \int_{-1}^1 \alpha^N(0,x)\int_{-1}^x \left(\gamma^N(T,y) - \gamma^N(0,y)\right)dydx \overset{(9.82),(9.80)}{\to} 0. \tag{9.86}$$

We deal with $T_4^N$ by a double integration by parts :

$$\begin{aligned} T_4^N &= \int_0^T \int_{-1}^1 \partial_\tau \gamma^N(\tau, y) dy \int_0^\tau \beta^N(\nu, 1) d\nu d\tau - \int_0^T \int_{-1}^1 \partial_\tau \gamma^N(\tau, x) \int_0^\tau \beta^N(\nu, x) d\nu dx d\tau \\ &= \int_{-1}^1 \gamma^N(T, y) dy \int_0^T \beta^N(\tau, 1) d\tau - \int_0^T \int_{-1}^1 \gamma^N(\tau, x) \beta^N(\tau, 1) dx d\tau \\ &\quad - \int_0^T \int_{-1}^1 \beta^N(\tau, x) \gamma^N(T, x) d\tau dx + \int_0^T \int_{-1}^1 \beta^N(\tau, x) \gamma^N(\tau, x) d\tau dx. \\ &\overset{(9.81),(9.82)}{\to} 0. \end{aligned} \tag{9.87}$$

From (9.84), (9.85), (9.86), (9.87), we obtain (9.83), which concludes the proof. □

### 9.5.4 Proof of Theorem 9.3.2

We are now in position to prove Theorem 9.3.2.

*Proof of Theorem 9.3.2.* We first prove the existence of a $D$-compatible $T > 0$. $\phi_0^x$ and $\phi_0^\tau$ satisfies (9.29), and :

$$\left\| W''(U_j(t)) \right\|_{L_t^\infty(l_j^\infty)} \leq \left\| W'' \right\|_{L^\infty([a,b])}.$$

Moreover, the discrete energy is preserved, which implies that :

$$\begin{aligned} \mathcal{E}_D(t) \leq & \sum_{j=-N+1}^{N} \left\{ \frac{1}{2} \left( \phi_0^\tau \left( \frac{j}{N} \right) \right)^2 + W \left( N \int_{j/N}^{(j+1)/N} \phi_0^x(y)\, dy \right) \right\} \\ \leq & N \left\| \phi_\tau \right\|_{L_x^\infty}^2 + 2N \left\| W \right\|_{L^\infty([a,b])}. \end{aligned}$$

Therefore, we can apply Theorem 9.3.3. Let $\widetilde{X}_j(t)$ satisfy (9.1) with initial conditions $\widetilde{U}_j(t = 0) = u_r$, $\widetilde{V}_j(t = 0) = 0$ and Dirichlet boundary conditions (remark that this means that $U_j(t)$ and $V_j(t)$ do not depend on time). We compare $\widetilde{U}_j(t) = u_r$ and $\widetilde{V}_j(t) = 0$ with $U_j(t)$ and $V_j(t)$, respectively. Using Theorem 9.3.3, there exists $c > 0$ such that, if $T < 1/(4c)$ and $T < T_0$ :

$$\sup_{t\in[0,NT]} \max_{j\in[3N/4,N]} N \left| U_j(t) - u_r \right| \underset{N\to+\infty}{\to} 0, \quad \text{and} \quad \sup_{t\in[0,NT]} \max_{j\in[3N/4,N]} N \left| V_j(t) - 0 \right| \underset{N\to+\infty}{\to} 0.$$

The same argument applies for $j < -3N/4$. Therefore, there exists a $D$-compatible $T > 0$.

We now prove that $\phi^N$ does not converge to $\phi$. We argue by contradiction and assume that :

$$\phi^N \to \phi \text{ in } D'_{\tau,x}. \tag{9.88}$$

Since the discrete energy $\mathcal{E}_D$ is preserved and as $W$ is strongly convex, we get an $H^1$ estimate over $\phi^N$ :

$$\int_{-1}^1 \left\{ \left( \partial_x \phi^N(\tau, x) \right)^2 + \left( \partial_\tau \phi^N(\tau, x) \right)^2 \right\} \leq C.$$

which directly implies :

$$\phi^N \rightharpoonup \phi \text{ in } H^1_{\tau,x}. \tag{9.89}$$

By Lemma 9.5.2, we get that :

$$\partial_x \phi^N \to \partial_x \phi \text{ in } L^2_{\tau,x}. \tag{9.90}$$

Whence, by Lemme 9.5.5, we have :

$$\xi^N \to \partial_\tau \phi \text{ in } L^2_{\tau,x}. \tag{9.91}$$

$W$ is continuous. Therefore (9.90), (9.36) and (9.91) imply :

$$\begin{aligned} &\int_0^T \int_{-1}^1 \left\{ \frac{1}{2} \left( \partial_\tau \phi^N \right)^2 + W \left( \partial_x \phi^N \right) \right\} (\tau, x) dx d\tau \\ &\to \int_0^T \int_{-1}^1 \left\{ \frac{1}{2} \left( \partial_\tau \phi \right)^2 + W \left( \partial_x \phi \right) \right\} (\tau, x) dx d\tau. \end{aligned} \tag{9.92}$$

But the left-hand term of (9.92) also converges, by discrete energy conservation :

$$\begin{aligned} &\int_0^T \int_{-1}^1 \left\{ \frac{1}{2} \left( \partial_\tau \phi^N \right)^2 + W \left( \partial_x \phi^N \right) \right\} (\tau, x) dx d\tau \\ &= T \int_{-1}^1 \left\{ \frac{1}{2} \left( \partial_\tau \phi^N \right)^2 + W \left( \partial_x \phi^N \right) \right\} (0, x) dx \\ &\underset{N \to \infty}{\longrightarrow} T \int_{-1}^1 \left\{ \frac{1}{2} \left( \phi_0^\tau(x) \right)^2 + W \left( \phi_0^x(x) \right) \right\} d\tau = T \mathcal{E}_C(0), \end{aligned}$$

and the right-hand term of (9.92) is nothing but the energy $\mathcal{E}_C(\tau)$. As $\phi$ is an entropy solution, we have :

$$\int_0^T \mathcal{E}_C(\tau) d\tau = \int_0^{T_1} \mathcal{E}_C(\tau) d\tau + \int_{T_1}^T \mathcal{E}_C(\tau) d\tau \overset{(9.37)}{<} T \mathcal{E}_C(0).$$

Therefore, we reach a contradiction, and $\phi^N$ cannot converge to $\phi$. □

## 9.6 A uniform bound on the distance between particles

Notice that it is important to assume some regularity on the initial conditions in Conjecture 9.3.1. It is indeed possible to build some initial conditions that are small in $l^\infty_j$ such that the associated solutions of (9.1) are not bounded uniformly in $N$ at a fixed macroscopic time $\tau > 0$. The following proof of Proposition 9.3.2 uses the reversibility of equation (9.1) and an explicit spectral decomposition.

We first derive explicit formulae for solution of linear periodic system (9.1).
Let $I \in \mathcal{M}_{2N}(\mathbb{R})$ be the identity matrix, and $J \in \mathcal{M}_{2N}(\mathbb{R})$ the circular permutation :

$$J_{jk} := \delta^j_{k+1}.$$

When the potential is quadratic and satisfies (9.14), system (9.1) with periodic boundary conditions is equivalent to :

$$\frac{d^2X}{dt^2} + \left(2I - J - J^{-1}\right) X = 0.$$

We diagonalize this system using its eigenvectors :

$$\Omega_j = \frac{1}{\sqrt{2N}} \begin{pmatrix} 1 \\ \omega_j \\ .. \\ \omega_j^{2N-1} \end{pmatrix},$$

where $\omega_j = \exp\left(\frac{ij\pi}{N}\right)$. The associated eigenvalues are :

$$\lambda_j = 2\left(1 - \cos\left(\frac{j\pi}{N}\right)\right) = 4\sin^2\left(\frac{j\pi}{2N}\right). \tag{9.93}$$

Thus, a solution of (9.1) with periodic boundary conditions satisfies :

$$\begin{aligned} X_j(t) = &\sum_{k=1}^{2N-1} \cos\left(t\sqrt{\lambda_k}\right) (\Omega_k|X(0))\,(\Omega_k)_j + \sum_{k=1}^{2N-1} \frac{1}{\sqrt{\lambda_k}} \sin\left(t\sqrt{\lambda_k}\right) (\Omega_k|V(0))\,(\Omega_k)_j \\ &+ \frac{1}{\sqrt{2N}} \left[(\Omega_0|X(0)) + (\Omega_0|V(0))\,t\right], \end{aligned} \tag{9.94}$$

where, for two vectors $Y$, $Z \in \mathbb{C}^{2N}$, $(|)$ denotes the hermitian product :

$$(Y|Z) := \sum_{k=-N}^{N-1} Y_k Z_k^*.$$

One easily derives such formulae for $V_j$, $U_j$ and $Z_j$ by linearity.

We can now prove Proposition 9.3.2.

*Proof of Proposition 9.3.2.* Using the reversibility of (9.1), it is enough to prove that, if $X_j^N$ is a solution of (9.1) with periodic boundary condition such that $U_j^N(t=0) = \delta_j^0$ and $V_j^N(t=0) = 0$, then :

$$\left\|U_j^N(N\tau)\right\|_{l_j^\infty} \to 0, \tag{9.95}$$

$$\left\|V_j^N(N\tau)\right\|_{l_j^\infty} \to 0. \tag{9.96}$$

Indeed, let $\widetilde{X}_i$ be the solution of (9.1) with periodic boundary conditions and the following initial conditions :

$$\begin{aligned} \widetilde{U}_j^N(t=0) &:= K_N U_j^N(N\tau), \\ \widetilde{V}_j^N(t=0) &:= -K_N V_j^N(N\tau). \end{aligned}$$

By linearity and reversibility of (9.1), we get :

$$\widetilde{U}_j^N(N\tau) = K_N U_j^N(0),$$
$$\widetilde{V}_j^N(N\tau) = -K_N V_j^N(0).$$

Setting $K_N = \left\{ \max\left( \left\| U_j^N(N\tau) \right\|_{l_j^\infty}, \left\| V_j^N(N\tau) \right\|_{l_j^\infty} \right) \right\}^{-1}$ gives the desired result.

We only show (9.95), as the proof of (9.96) is similar. Thanks to (9.94) :

$$U_k^N(N\tau) = \frac{1}{2N} \sum_{j=-N}^{N-1} \cos\left( 2N\tau \sin\left( \frac{j\pi}{2N} \right) \right) e^{ijk\pi/N}.$$

To simplify the proof, we suppose $N = nm$ (it can be generalized with a few technicalities). Thus :

$$U_k^N(N\tau) = \frac{1}{2N} \sum_{l=-n}^{n-1} \sum_{j=0}^{m-1} \cos\left( 2N\tau \sin\left( \frac{(lm+j)\pi}{2N} \right) \right) e^{i(lm+j)k\pi/N}.$$

Let us bound terms of the type :

$$Q_\pm^{l,k} := \left| \frac{1}{N} \sum_{j=0}^{m-1} \exp\left( \pm i 2N\tau \sin\left( \frac{(lm+j)\pi}{2N} \right) \right) e^{i(lm+j)k\pi/N} \right|.$$

We expand :

$$\sin\left( \frac{(lm+j)\pi}{2N} \right) = \sin\left( \frac{lm\pi}{2N} \right) + \frac{j\pi}{2N} \cos\left( \frac{lm\pi}{2N} \right) + \frac{m^2}{N^2} r(N,m,l,j),$$

where $r(N,m,l,j) < C$, independently of $N$, $m$, $l$, $j$. As a consequence :

$$\begin{aligned} Q_\pm^{l,k} \leq & \left| \frac{1}{N} \sum_{j=0}^{m-1} \exp\left( \pm 2i \left( \frac{\tau j\pi}{2} \cos\left( \frac{lm\pi}{2N} \right) + \frac{m^2\tau}{N} r(N,m,l,j) \right) \right) e^{ijk\pi/N} \right| \\ \leq & \frac{1}{N} \sum_{j=0}^{m-1} \frac{Cm^2}{N} + \left| \frac{1}{N} \sum_{j=0}^{m-1} \exp\left( \pm i\tau j\pi \cos\left( \frac{lm\pi}{2N} \right) + \frac{ijk\pi}{N} \right) \right| \\ \leq & C\frac{m^3}{N^2} + \frac{1}{N} \frac{2}{|1 - \exp(i\gamma_\pm(k,l,m,N,\tau))|}, \end{aligned}$$

where :

$$\gamma_\pm(k,l,m,N,\tau) = \frac{k\pi}{N} \pm \tau\pi \cos\left( \frac{lm\pi}{2N} \right).$$

Without loss of generality, we focus only on $Q_+^{l,k}$. There exist at most two solutions $s_1$ and $s_2 \in [-\pi/2, \pi/2]$ to the equation :

$$\tau\pi \cos(s) + \frac{k\pi}{N} = 0.$$

Let $1 > \delta > 0$. If $\frac{lm\pi}{2N} \notin ]s_1 - \delta, s_1 + \delta[ \cup ]s_2 - \delta, s_2 + \delta[$, then :

$$|1 - \exp(i\gamma_+(k, l, m, N, \tau))| > C\delta^2. \tag{9.97}$$

for some universal constant $C$. Whence, if $\frac{lm\pi}{2N} \notin ]s_1 - \delta, s_1 + \delta[ \cup ]s_2 - \delta, s_2 + \delta[$, we have :

$$Q_+^{l,k} \leq C\frac{m^3}{N^2} + \frac{1}{N}\frac{C}{\delta^2}.$$

Moreover, it is immediate from the definition of $Q_\pm^{l,k}$ that for all $l, k$ :

$$Q_\pm^{l,k} \leq \frac{m}{N}.$$

We denote :

$$E := [\![-n, n-1]\!] \cap \left( \left] \frac{2N}{m\pi}(s_1 - \delta), \frac{2N}{m\pi}(s_1 + \delta) \right[ \cup \left] \frac{2N}{m\pi}(s_2 - \delta), \frac{2N}{m\pi}(s_2 + \delta) \right[ \right),$$

and bound the sum :

$$\begin{aligned}
\sum_{l=-n}^{n-1} Q_+^{l,k} =& \sum_{l \in E} Q_+^{l,k} + \sum_{l \notin E} Q_+^{l,k} \\
\leq& \sum_{l \in E} \frac{m}{N} + C \sum_{l \notin E} \frac{m^3}{N^2} + \frac{1}{N\delta^2} \\
\leq& \frac{8N}{m\pi}\delta\frac{m}{N} + 2Cn\left(\frac{m^3}{N^2} + \frac{1}{N\delta^2}\right) \\
\leq& C\left(\delta + \frac{m^2}{N} + \frac{1}{m\delta^2}\right).
\end{aligned}$$

Let $\delta = (m)^{-1/3}$, $m = N^{3/7}$. We get :

$$\sum_{l=-n}^{n-1} Q_+^{l,k} \leq \frac{C}{N^{1/7}}.$$

Doing the same manipulations on $Q_-^{l,k}$, we get that, for all $k \in [\![-N, N-1]\!]$ :

$$\left|U_k^N(N\tau)\right| \leq \sum_{l=-n}^{n-1} Q_+^{l,k} + Q_-^{l,k} \leq \frac{C}{N^{1/7}},$$

whence (9.95). □

*Remark* 15. One can remove the technical assumption $N = nm$ with $m, n \in \mathbb{N}$, by fixing $\mu := \lfloor N^{1/7} \rfloor$, and then $m := \mu^3$, $n := \mu^4$. Remarking that $N - \mu^7 < CN^{6/7}$, we can apply the same proof as above and derive the same estimates.

## 9.7 Non-existence of discrete shock waves

We prove in this section that there do not exist discrete shock waves. We use some ideas from [19], where an existence result is proven for upwind schemes.

### 9.7.1 Quadratic potential

In this section, we prove Proposition 9.3.3. We first show a lemma which is valid for a wide class of potentials $W$ :

**Lemma 9.7.1.** *Suppose $W \in C^1(\mathbb{R})$, such that $W'(u_l) \neq W'(u_r)$. Then there exists no discrete shock wave to the equation (9.1) with zero speed. That is, there does not exist $X_j(t)$ satisfying Definition 9.2.5, with associated $c = 0$.*

*Proof.* Integrating (9.32), we get :

$$c^2 \int_x^y \phi''(s)ds = \int_{y-1}^{y} W'(\phi(s+1) - \phi(s))ds - \int_{x-1}^{x} W'(\phi(s+1) - \phi(s))ds.$$

If $x \to -\infty$ and $y \to +\infty$, we obtain :

$$c^2(u_r - u_l) = W'(u_r) - W'(u_l), \tag{9.98}$$

which is the Rankine-Hugoniot equation (9.30). It cannot hold if $c = 0$. □

We can now prove Proposition 9.3.3 :

*Proof of Proposition 9.3.3.* If $c = 0$, Lemma 9.7.1 gives the desired result. Suppose $c \neq 0$. Using Fourier transform on (9.32) implies :

$$c^2\xi^2 \mathcal{F}(\phi)(\xi) = (\exp(i\xi) - 2 + \exp(-i\xi))\, \mathcal{F}(\phi)(\xi).$$

The equation :

$$c^2\xi^2 = 2\,(1 - \cos(\xi))\,, \tag{9.99}$$

has a finite number of solutions $\xi_j$, $j \in J$. Therefore, there exist $K_j \in \mathbb{N}$, $a_{jk} \in \mathbb{R}$ such that :

$$\mathcal{F}(\phi) = \sum_{j=1}^{J} \sum_{k=0}^{K_j} a_{jk} \delta_{\xi_j}^{(k)}.$$

Thus :

$$\phi(x) = \sum_{j=1}^{J} \sum_{k=0}^{K_j} a_{jk} (ix)^k \exp(ix\xi_j).$$

Since $\phi'$ has a limit for $x \to +\infty$ then $a_{jk} = 0$ if $j \neq 0$ or $k > 1$. Thus, there exists no discrete shock wave with $c \neq 0$. □

### 9.7.2 Convex non-linear potential

We now prove Proposition 9.3.4.

*Proof of Proposition 9.3.4.* Suppose $u_l \neq u_r$. Test now (9.32) with $\phi'$ :

$$\begin{aligned}\frac{c^2}{2}\left(u_r^2-u_l^2\right) =& \lim_{R\to+\infty}\int_{-R}^{R}\left\{W'(\phi(x+1)-\phi(x))-W'(\phi(x)-\phi(x-1))\right\}\phi'(x)dx\\ =& \lim_{R\to+\infty}\left\{\int_{-R}^{R}W'(\phi(x+1)-\phi(x))\left(\phi'(x+1)-\phi'(x)\right)dx\right.\\ &+\int_{R-1}^{R}\phi'(x+1)W'(\phi(x+1)-\phi(x))dx\\ &\left.-\int_{-R-1}^{-R}\phi'(x+1)W'\left(\phi(x+1)-\phi(x)\right)dx\right\}\\ =& W(u_r)-W(u_l)+W'(u_r)u_r-W'(u_l)u_l.\end{aligned}$$

Using (9.98), we get :

$$\frac{1}{2}\left(W'(u_r)-W'(u_l)\right)=\frac{W(u_r)-W(u_l)}{u_r-u_l}. \tag{9.100}$$

Yet :

$$\begin{aligned}&\frac{1}{2}\left(W'(u_r)-W'(u_l)\right)-\frac{W(u_r)-W(u_l)}{u_r-u_l}\\ &=\frac{1}{u_r-u_l}\int_{u_l}^{u_r}\left\{\frac{u_r-s}{u_r-u_l}W'(u_l)+\frac{s-u_l}{u_r-u_l}-W'(s)\right\}ds,\end{aligned}$$

and as $W'$ is strictly convex :

$$\frac{u_r-s}{u_r-u_l}W'(u_l)+\frac{s-u_l}{u_r-u_l}-W'(s)>0,\forall s\in]u_l,u_r[.$$

This is contradictory. Therefore, as $u_l \neq u_r$, there does not exist any discrete shock wave of (9.1). □

## 9.8 Appendix

*Proof of Lemma 9.2.1.* Recall that $V_{-N}(t)=0$ and $V_N(t)=0$ by (9.3), and that $\tau = t/N$. Thus :

$$\begin{aligned}\left\|\zeta^N(\tau,.)\right\|_{L^2_x}^2 =& \sum_{j=-N}^{N-1}\int_0^{1/N}\left(\left(1-\theta^N(x)\right)V_j+\theta^N(x)V_{j+1}\right)^2(t)dx\\ =&\frac{1}{N}\sum_{j=-N}^{N-1}\left\{\frac{1}{3}\left|V_j\right|^2+\frac{1}{3}\left|V_{j+1}\right|^2+\frac{1}{3}\left|V_j\right|\left|V_{j+1}\right|\right\}(t),\end{aligned}$$

and :

$$\left\|\xi^N(\tau,.)\right\|^2_{L^2_x} = \frac{1}{N}\sum_{j=-N}^{N-1} |V_j(t)|^2 .$$

Therefore, Cauchy-Schwarz inequality implies (9.17).

Assume now that $\zeta^N \to \zeta^\infty$ in $L^2_{\tau,x}$. Let $N > 0$. We can approximate first $\zeta^\infty$ by $\zeta_1^\infty$, a function which is piecewise constant with respect to $t$, and piecewise constant with respect to $x$ on intervals $[j/N, (j+1)/N]$ (we impose that $\zeta_1^\infty(-1) = \zeta_1^\infty(-1) = 0$). We denote by $\zeta_2^\infty$ the linear interpolate of $\zeta_2^\infty$ with respect to $x$ on intervals $[j/N, (j+1)/N]$. By the same proof as above, it is immediate that :

$$\left\|\xi^N(\tau,.) - \zeta_1^\infty(\tau,.)\right\|_{L^2_x} \leq 6\left\|\zeta^N(\tau,.) - \zeta_2^\infty(\tau,.)\right\|_{L^2_x} . \tag{9.101}$$

Furthermore, we can approximate $\zeta^\infty$ in such a way that, when $N \to +\infty$, we have :

$$\|\zeta^\infty - \zeta_1^\infty\|_{L^2_{\tau,x}} = o(1) \qquad \text{and} \qquad \|\zeta^\infty - \zeta_2^\infty\|_{L^2_{\tau,x}} = o(1). \tag{9.102}$$

Thus, by triangle inequality :

$$\begin{aligned}\left\|\xi^N - \zeta^\infty\right\|_{L^2_{\tau,x}} \leq& \left\|\xi^N - \zeta_1^\infty\right\|_{L^2_{\tau,x}} + \|\zeta^\infty - \zeta_1^\infty\|_{L^2_{\tau,x}} \\ \overset{(9.101)}{\leq}& 6\left\|\zeta^N - \zeta_2^\infty\right\|_{L^2_{\tau,x}} + \|\zeta^\infty - \zeta_1^\infty\|_{L^2_{\tau,x}} \\ \leq& 6\left\|\zeta^N - \zeta^\infty\right\|_{L^2_{\tau,x}} + 6\|\zeta^\infty - \zeta_2^\infty\|_{L^2_{\tau,x}} + \|\zeta^\infty - \zeta_1^\infty\|_{L^2_{\tau,x}} .\end{aligned}$$

Therefore, thanks to (9.102), we get the first implication of (9.18). The reverse implication follows by similar arguments.

We now claim that :

$$\zeta^N - \xi^N \to 0 \text{ in } H^{-1}_{\tau,x}, \tag{9.103}$$

which, with (9.17), implies (9.19). Let $v \in C^1_x$. Then, by integration by parts :

$$\int_{-1}^{1} \left(\zeta^N(\tau,x) - \xi^N(\tau,x)\right) v(x)dx = v(1)F^N(\tau,1) - \int_{-1}^{1} F^N(\tau,x)v'(x)dx. \tag{9.104}$$

where :

$$F^N(\tau,x) = \int_{-1}^{x} \left(\zeta^N(\tau,y) - \xi^N(\tau,y)\right) dy.$$

By definition :

$$\begin{aligned}\left|F^N(\tau,x)\right| =& \left|\int_{\frac{\lfloor Nx\rfloor}{N}}^{x} \left((1-\theta^N(x))V_{k^N(x)} + \theta^N(x)V_{k^N(x)+1}\right)(t)dx - \frac{1}{2N}V_{k^N(x)}(t)\right| \\ \leq& \frac{1}{N}\left(\left|V_{k^N(x)}\right| + \left|V_{k^N(x)+1}\right|\right)(t),\end{aligned}$$

and by (9.3), $F^N(\tau,1)=0$. Therefore (9.104) implies that :

$$\begin{aligned}\left|\int_{-1}^{1}\left(\zeta^N(\tau,x)-\xi^N(\tau,x)\right)v(x)dx\right| \leq & \left\|F^N\right\|_{L^2_x}\left\|v'\right\|_{L^2_x}\\ \leq & \frac{1}{\sqrt{N}}\left|\frac{1}{N}\sum_{j=-N}^{N-1}|V_j(t)|^2\right|^{1/2}\left\|v'\right\|_{L^2_x}\\ \leq & \frac{1}{\sqrt{N}}\mathcal{E}_D(t)\left\|v'\right\|_{L^2_x}.\end{aligned}$$

Therefore, as $\mathcal{E}_D(t)=\mathcal{E}_D(0)$ (see Lemma 9.4.1), we have that :

$$\left|\int_{-1}^{1}\left(\zeta^N(\tau,x)-\xi^N(\tau,x)\right)v(x)dx\right| \leq \frac{C}{\sqrt{N}}\left\|v'\right\|_{L^2_x},$$

which, thanks to (9.17), implies (9.103) and concludes the proof of (9.19). □

## Acknowledgement

We wish to thank Claude Le Bris and Frédéric Legoll for their help and Gabriel Stoltz and Frédéric Lagoutière, for fruitful discussions.

# Bibliographie


[1] F. Achleitner and C. Kuehn. Traveling waves for a bistable equation with nonlocal diffusion. *Adv. Differential Equations*, 20(9-10) :887–936, 2015.

[2] G. Allaire. *Shape optimization by the homogenization method*, volume 146 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2002.

[3] B. Alpert, L. Greengard, and T. Hagstrom. Rapid evaluation of nonreflecting boundary kernels for time-domain wave propagation. *SIAM J. Numer. Anal.*, 37(4) :1138–1164, 2000.

[4] A. Anantharaman, R. Costaouec, C. Le Bris, F. Legoll, and F. Thomines. Introduction to numerical stochastic homogenization and the related computational challenges : some recent developments. In *Multiscale modeling and analysis for materials simulation*, volume 22 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 197–272. World Sci. Publ., Hackensack, NJ, 2012.

[5] S. Armstrong, T. Kuusi, and J.-C. Mourrat. The additive structure of elliptic homogenization. *Invent. Math.*, 208(3) :999–1154, 2017.

[6] S. Armstrong, T. Kuusi, and J.-C. Mourrat. Quantitative stochastic homogenization and large-scale regularity, 2018. in preparation.

[7] S. Armstrong and J.-C. Mourrat. Lipschitz regularity for elliptic equations with random coefficients. *Arch. Ration. Mech. Anal.*, 219(1) :255–348, 2016.

[8] S. Armstrong and Z. Shen. Lipschitz estimates in almost-periodic homogenization. *Comm. Pure Appl. Math.*, 69(10) :1882–1923, 2016.

[9] S. Armstrong and C. Smart. Quantitative stochastic homogenization of convex integral functionals. *Ann. Sci. Éc. Norm. Supér. (4)*, 49(2) :423–481, 2016.

[10] K. E. Atkinson. *The numerical solution of integral equations of the second kind*, volume 4 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 1997.

[11] M. Avellaneda and F.-H. Lin. Compactness methods in the theory of homogenization. *Comm. Pure Appl. Math.*, 40(6) :803–847, 1987.

[12] M. Avellaneda and F.-H. Lin. $L^p$ bounds on singular integrals in homogenization. *Comm. Pure Appl. Math.*, 44(8-9) :897–910, 1991.

[13] L. Banjai, M. López-Fernández, and A. Schädle. Fast and oblivious algorithms for dissipative and two-dimensional wave equations. *SIAM J. Numer. Anal.*, 55(2) :621–639, 2017.

[14] L. Banjai and C. Lubich. An error analysis of Runge-Kutta convolution quadrature. *BIT*, 51(3) :483–496, 2011.

[15] L. Banjai and C. Lubich. Runge–Kutta convolution coercivity and its use for time-dependent boundary integral equations. *arXiv preprint arXiv :1702.08385*, 2017. preprint.

[16] L. Banjai and A. Rieder. Convolution quadrature for the wave equation with a nonlinear impedance boundary condition. *Math. Comp.*, 87(312) :1783–1819, 2018.

[17] P. Bella, A. Giunti, and F. Otto. Quantitative stochastic homogenization : local control of homogenization error through corrector. In *Mathematics and materials*, volume 23 of *IAS/Park City Math. Ser.*, pages 301–327. Amer. Math. Soc., Providence, RI, 2017.

[18] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic analysis for periodic structures*. AMS Chelsea Publishing, Providence, RI, 2011.

[19] S. Benzoni-Gavage. Semi-discrete shock profiles for hyperbolic systems of conservation laws. *Phys. D*, 115(1-2) :109–123, 1998.

[20] M. Berezhnyy and L. Berlyand. Continuum limit for three-dimensional mass-spring networks and discrete Korn's inequality. *J. Mech. Phys. Solids*, 54(3) :635–669, 2006.

[21] J. Bergh and J. Löfström. *Interpolation spaces. An introduction.* Springer-Verlag, Berlin-New York, 1976.

[22] X. Blanc and M. Josien. From the Newton equation to the wave equation : the case of shock waves. *Applied Mathematics Research eXpress*, 2017 :338–385, 2017.

[23] X. Blanc, M. Josien, and C. Le Bris. Approximation locale précisée dans des problèmes multi-échelles avec défauts localisés. Preprint hal-01893991.

[24] X. Blanc, M. Josien, and C. Le Bris. Precised approximations in elliptic homogenization beyond the periodic setting. Preprint hal-01958207.

[25] X. Blanc, C. Le Bris, and P.-L. Lions. On correctors for linear elliptic homogenization in the presence of local defects. In preparation.

[26] X. Blanc, C. Le Bris, and P.-L. Lions. From the Newton equation to the wave equation in some simple cases. *Netw. Heterog. Media*, 7(1) :1–41, 2012.

[27] X. Blanc, C. Le Bris, and P.-L. Lions. A possible homogenization approach for the numerical simulation of periodic microstructures with defects. *Milan J. Math.*, 80(2) :351–367, 2012.

[28] X. Blanc, C. Le Bris, and P.-L. Lions. Local profiles for elliptic problems at different scales : defects in, and interfaces between periodic structures. *Comm. Partial Differential Equations*, 40(12) :2173–2236, 2015.

[29] X. Blanc, F. Legoll, and A. Anantharaman. Asymptotic behavior of Green functions of divergence form operators with periodic coefficients. *Appl. Math. Res. Express. AMRX*, (1) :79–101, 2013.

[30] A. Bonito, J. P. Borthagaray, R. Nochetto, E. Otarola, and A. Salgado. Numerical methods for fractional diffusion. 2017. arXiv preprint arXiv :1707.01566.

[31] S. Börm. *Efficient numerical methods for non-local operators*, volume 14 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2010. $\mathcal{H}^2$-matrix compression, algorithms and analysis.

[32] B. Böttcher, R. Schilling, and J. Wang. *Lévy matters. III*, volume 2099 of *Lecture Notes in Mathematics*. Springer, Cham, 2013.

[33] R. N. Bracewell. *The Fourier transform and its Applications*. Mc Graw-Hill, Boston, 2000.

[34] J. Braun. Connecting atomistic and continuous models of elastodynamics. *Arch. Ration. Mech. Anal.*, 224(3) :907–953, 2017.

[35] Y. Brenier. Approximation of a simple Navier-Stokes model by monotonic rearrangement. *Discrete Contin. Dyn. Syst.*, 34(4) :1285–1300, 2014.

[36] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.

[37] A. Bueno-Orovio, D. Kay, and K. Burrage. Fourier spectral methods for fractional-in-space reaction-diffusion equations. *BIT*, 54(4) :937–954, 2014.

[38] V. Bulatov, F. Abraham, L. Kubin, B. Devincre, and S. Yip. Connecting atomistic and mesoscale simulations of crystal plasticity. *Nature*, 391(6668) :669, 1998.

[39] X. Cabré, N. Cónsul, and J. V. Mandé. Traveling wave solutions in a half-space for boundary reactions. *Anal. PDE*, 8(2) :333–364, 2015.

[40] X. Cabré and Y. Sire. Nonlinear equations for fractional Laplacians I. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 31(1) :23–53, 2014.

[41] X. Cabré and Y. Sire. Nonlinear equations for fractional Laplacians II. *Trans. Amer. Math. Soc.*, 367(2) :911–941, 2015.

[42] I. Catto, C. Le Bris, and P.-L. Lions. *The mathematical theory of thermodynamic limits : Thomas-Fermi type models*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 1998.

[43] T. Cazenave and A. Haraux. *An introduction to semilinear evolution equations*, volume 13 of *Oxford Lecture Series in Mathematics and its Applications*. The Clarendon Press, Oxford University Press, New York, 1998.

[44] S. Chen, J. Shen, and L.-L. Wang. Laguerre functions and their applications to tempered fractional differential equations on infinite intervals. *J. Sci. Comput.*, 74(3) :1286–1313, 2018.

[45] X. Chen. Existence, uniqueness, and asymptotic stability of traveling waves in nonlocal evolution equations. *Adv. Differential Equations*, 2(1) :125–160, 1997.

[46] A. Chmaj. Existence of traveling waves in the fractional bistable equation. *Arch. Math. (Basel)*, 100(5) :473–480, 2013.

[47] J. Christian. The relation between dislocation velocity and stress. *Scripta metallurgica*, 4(10) :811–814, 1970.

[48] A. Cochard and J. Rice. A spectral method for numerical elastodynamic fracture analysis without spatial replication of the rupture event. *Journal of the Mechanics and Physics of Solids*, 45(8) :1393–1418, 1997.

[49] S. Day, L. Dalguer, N. Lapusta, and Y. Liu. Comparison of finite difference and boundary integral solutions to three-dimensional spontaneous rupture. *Journal of Geophysical Research : Solid Earth*, 110(B12), 2005.

[50] P. Deift and K. T.-R. McLaughlin. *A continuum limit of the Toda lattice*, volume 131. 1998.

[51] C. Denoual. Dynamic dislocation modeling by combining Peierls–Nabarro and Galerkin methods. *Physical Review B*, 70(2) :024106, 2004.

[52] C. Denoual. Modeling dislocation by coupling Peierls–Nabarro and element-free Galerkin methods. *Computer methods in applied mechanics and engineering*, 196(13-16) :1915–1923, 2007.

[53] B. Devincre and L. Kubin. Mesoscopic simulations of dislocations and plasticity. *Materials Science and Engineering : A*, 234 :8–14, 1997.

[54] K. S. Djaka, A. Villani, V. Taupin, L. Capolungo, and S. Berbenni. Field dislocation mechanics for heterogeneous elastic materials : a numerical spectral approach. *Comput. Methods Appl. Mech. Engrg.*, 315 :921–942, 2017.

[55] G. Dolzmann and S. Müller. Estimates for Green's matrices of elliptic systems by $L^p$ theory. *Manuscripta Math.*, 88(2) :261–273, 1995.

[56] J. Drouet, L. Dupuy, F. Onimus, and F. Mompiou. A direct comparison between in-situ transmission electron microscopy observations and dislocation dynamics simulations of interaction between dislocation and irradiation induced loop in a zirconium alloy. *Scripta Materialia*, 119 :71–75, 2016.

[57] Wei-nan E and Ping-bing Ming. Cauchy-Born rule and the stability of crystalline solids : dynamic problems. *Acta Math. Appl. Sin. Engl. Ser.*, 23(4) :529–550, 2007.

[58] L. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[59] L. Evans and R. Gariepy. *Measure theory and fine properties of functions*. Textbooks in Mathematics. CRC Press, Boca Raton, FL, revised edition, 2015.

[60] P. C. Fife and J. B. McLeod. The approach of solutions of nonlinear diffusion equations to travelling front solutions. *Arch. Rat. Mech. Anal.*, 65(4) :335–361, 1977.

[61] C. A. J. Fletcher. *Computational techniques for fluid dynamics. 1.* Springer Series in Computational Physics. Springer-Verlag, Berlin, 1991.

[62] P. Geubelle and J. Rice. A spectral method for three-dimensional elastodynamic fracture problems. *Journal of the Mechanics and Physics of Solids*, 43(11) :1791–1824, 1995.

[63] Mariano Giaquinta. *Multiple integrals in the calculus of variations and nonlinear elliptic systems*, volume 105 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1983.

[64] D. Gilbarg and N. Trudinger. *Elliptic partial differential equations of second order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001.

[65] J. J. Gilman. Dislocation motion in a viscous medium. *Physical Review Letters*, 20(4) :157, 1968.

[66] D. Givoli. Non-reflecting boundary conditions. *Journal of computational physics*, 94(1) :1–29, 1991.

[67] A. Gloria, S. Neukamm, and F. Otto. A regularity theory for random elliptic operators. *ArXiv e-print arXiv :1409.2678*, 2014.

[68] A. Gloria, S. Neukamm, and F. Otto. Quantification of ergodicity in stochastic homogenization : optimal bounds via spectral gap on Glauber dynamics. *Invent. Math.*, 199(2) :455–515, 2015.

[69] R. Gracie and T. Belytschko. An adaptive concurrent multiscale method for the dynamic simulation of dislocations. *International Journal for Numerical Methods in Engineering*, 86(4-5) :575–597, 2011.

[70] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products*. Elsevier, Academic Press, Amsterdam, 7th. edition, 2007.

[71] G. W. Griffiths and W. E. Schiesser. *Traveling wave analysis of partial differential equations*. Elsevier/Academic Press, Amsterdam, 2012.

[72] M. Grüter and K.-O. Widman. The Green function for uniformly elliptic equations. *Manuscripta Math.*, 37(3) :303–342, 1982.

[73] C. Gui and M. Zhao. Traveling wave solutions of Allen-Cahn equation with a fractional Laplacian. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 32(4) :785–812, 2015.

[74] W. Hackbusch. *Hierarchical matrices : algorithms and analysis*, volume 49 of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2015.

[75] T. Hagstrom. Radiation boundary conditions for the numerical simulation of waves. In *Acta numerica, 1999*, volume 8 of *Acta Numer.*, pages 47–106. Cambridge Univ. Press, Cambridge, 1999.

[76] E. Hairer, C. Lubich, and M. Schlichte. Fast numerical solution of nonlinear Volterra convolution equations. *SIAM J. Sci. Statist. Comput.*, 6(3) :532–541, 1985.

[77] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006.

[78] E. Hairer, S. Nørsett, and G. Wanner. *Solving ordinary differential equations. I Nonstiff problems*, volume 8 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 1993.

[79] E. Hairer and G. Wanner. *Solving ordinary differential equations. II Stiff and differential-algebraic problems*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 1996.

[80] F. B. Hildebrand. *Introduction to numerical analysis*. Dover Publications, Inc., New York, second edition, 1987.

[81] J. P. Hirth and J. Lothe. *Theory of dislocations*. John Wiley & Sons, 1982.

[82] B. L. Holian, H. Flaschka, and D. W. McLaughlin. Shock waves in the Toda lattice : analysis. *Phys. Rev. A (3)*, 24(5) :2595–2623, 1981.

[83] Y. Huang and A. Oberman. Numerical methods for the fractional Laplacian : a finite difference–quadrature approach. *SIAM J. Numer. Anal.*, 52(6) :3056–3084, 2014.

[84] D. E. Hurtado and M. Ortiz. Finite element analysis of geometrically necessary dislocations in crystal plasticity. *International Journal for Numerical Methods in Engineering*, 93(1) :66–79, 2013.

[85] V. Jikov, S. Kozlov, and O. Oleĭnik. *Homogenization of differential operators and integral functionals*. Springer-Verlag, Berlin, 1994.

[86] M. Josien. Decomposition and pointwise estimates of periodic Green functions of some elliptic equations with periodic oscillatory coefficients. Preprint arXiv :1807.09062, accepted by *Asymptotic Analysis*.

[87] M. Josien. Mathematical properties of the Weertman equation. Preprint arXiv :1709.0678, accepted by *Communications in Mathematical Sciences*.

[88] M. Josien, Y.-P. Pellegrini, F. Legoll, and C. Le Bris. Fourier-based numerical approximation of the Weertman equation for moving dislocations. *International Journal for Numerical Methods in Engineering*, 113 :1827–1850, 2018.

[89] M. Kanninen and C. Popelar. *Advanced fracture mechanics*, volume 15 of *Oxford Engineering Science Series*. Oxford Univeristy Press, 1985.

[90] R. P. Kanwal. *Generalized functions*. Birkhäuser Boston, Inc., Boston, MA, 2004.

[91] V. Karlin, V. G. Maz′ya, A. B. Movchan, J. R. Willis, and R. Bullough. Numerical solution of nonlinear hypersingular integral equations of the Peierls type in dislocation theory. *SIAM J. Appl. Math.*, 60(2) :664–678, 2000.

[92] C. T. Kelley. *Iterative methods for linear and nonlinear equations*, volume 16 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1995.

[93] C. Kenig, F.-H. Lin, and Z. Shen. Homogenization of elliptic systems with Neumann boundary conditions. *J. Amer. Math. Soc.*, 26(4) :901–937, 2013.

[94] C. Kenig, F.-H. Lin, and Z. Shen. Periodic homogenization of Green and Neumann functions. *Comm. Pure Appl. Math.*, 67(8) :1219–1262, 2014.

[95] F. King. *Hilbert transforms. Vol. 1*, volume 124 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2009.

[96] R. Kurtz, T. Farris, and Sun C. The numerical solution of Cauchy singular integral equations with application to fracture. *Int. J. Fract.*, 66(2) :139–154, 1994.

[97] N. Lapusta, J. Rice, Y. Ben-Zion, and G. Zheng. Elastodynamic analysis for slow tectonic loading with spontaneous rupture episodes on faults with rate-and state-dependent friction. *Journal of Geophysical Research : Solid Earth*, 105(B10) :23765–23789, 2000.

[98] P. D. Lax and C. D. Levermore. The small dispersion limit of the Korteweg-de Vries equation. I. *Comm. Pure Appl. Math.*, 36(3) :253–290, 1983.

[99] C. Le Bris. *Systèmes multi-échelles [in French]*, volume 47 of *Mathématiques & Applications (Berlin)*. Springer-Verlag, Berlin, 2005.

[100] F. Legoll and P.-L. Rothé. On the numerical approximation of fluctuations in stochastic homogenization. in preparation.

[101] L. Lejček. Peierls-Nabarro model of planar dislocation cores in bcc crystals. *Czech. J. Phys.*, 22(9) :802–812, 1972.

[102] L. Lejček. Dissociated dislocations in the Peierls-Nabarro model. *Czech. J. Phys. B*, 26(3) :294–299, 1976.

[103] Y. Y. Li and M. Vogelius. Gradient estimates for solutions to divergence form elliptic equations with discontinuous coefficients. *Arch. Ration. Mech. Anal.*, 153(2) :91–151, 2000.

[104] T. N. Lien, D. D. Trong, and A. Pham Ngoc Dinh. Laguerre polynomials and the inverse Laplace transform using discrete data. *J. Math. Anal. Appl.*, 337(2) :1302–1314, 2008.

[105] P. Linz. *Analytical and numerical methods for Volterra equations*, volume 7 of *SIAM Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1985.

[106] M. López-Fernández, C. Lubich, and A. Schädle. Adaptive, fast, and oblivious convolution in evolution equations with memory. *SIAM J. Sci. Comput.*, 30(2) :1015–1037, 2008.

[107] M. López-Fernández, C. Palencia, and A. Schädle. A spectral order method for inverting sectorial Laplace transforms. *SIAM J. Numer. Anal.*, 44(3) :1332–1350, 2006.

[108] G. Lu, V. Bulatov, and N. Kioussis. A nonplanar Peierls-Nabarro model and its application to dislocation cross-slip. *Philos. Mag.*, 83(31-34) :3539–3548, 2003.

[109] G. Lu, N. Kioussis, V. Bulatov, and E. Kaxiras. Generalized-stacking-fault energy surface and dislocation properties of aluminum. *Physical Review B*, 62(5) :3099, 2000.

[110] C. Lubich. On the multistep time discretization of linear initial-boundary value problems and their boundary integral equations. *Numer. Math.*, 67(3) :365–389, 1994.

[111] C. Lubich. Convolution quadrature revisited. *BIT*, 44(3) :503–514, 2004.

[112] C. Lubich and A. Schädle. Fast convolution for nonreflecting boundary conditions. *SIAM J. Sci. Comput.*, 24(1) :161–182, 2002.

[113] Z. Mao and J. Shen. Hermite spectral methods for fractional PDEs in unbounded domains. *SIAM J. Sci. Comput.*, 39(5) :A1928–A1950, 2017.

[114] Y. Meyer. *Ondelettes et opérateurs. II [in French]*. Actualités Mathématiques. Hermann, Paris, 1990. Opérateurs de Calderón-Zygmund.

[115] J. Mianroodi, A. Hunter, I. Beyerlein, and B. Svendsen. Theoretical and computational comparison of models for dislocation dissociation and stacking fault/core formation in fcc crystals. *J. Mech. Phys. Solids*, 2016.

[116] A. Mielke and L. Truskinovsky. From discrete visco-elasticity to continuum rate-independent plasticity : rigorous results. *Arch. Ration. Mech. Anal.*, 203(2) :577–619, 2012.

[117] R. Miller, R. Phillips, G. Beltz, and M. Ortiz. A non-local formulation of the Peierls dislocation model. *Journal of the Mechanics and Physics of Solids*, 46(10) :1845–1867, 1998.

[118] S. Müller. *Variational models for microstructure and phase transitions*, volume 1713 of *Lecture Notes in Math.* Springer, Berlin, 1999.

[119] N. I. Muskhelishvili. *Some basic problems of the mathematical theory of elasticity.* Noordhoff International Publishing, Leiden, 1977.

[120] F. R. N. Nabarro. Dislocations in a simple cubic lattice. *Proc. Phys. Soc.*, 59(2) :256, 1947.

[121] Y. Nec, A. A. Nepomnyashchy, and A. A. Golovin. Front-type solutions of fractional Allen-Cahn equation. *Phys. D*, 237(24) :3237–3251, 2008.

[122] J. N. Newman. Approximations for the Bessel and Struve functions. *Math. Comp.*, 43(168) :551–556, 1984.

[123] C. P. Niculescu and L.-E. Persson. *Convex functions and their applications.* Springer, New York, 2006.

[124] R. Nochetto, E. Otárola, and A. Salgado. A PDE approach to space-time fractional parabolic problems. *SIAM J. Numer. Anal.*, 54(2) :848–873, 2016.

[125] H. Noda and N. Lapusta. Three-dimensional earthquake sequence simulations with evolving temperature and pore pressure due to shear heating : Effect of heterogeneous hydraulic diffusivity. *Journal of Geophysical Research : Solid Earth*, 115(B12), 2010.

[126] H. Noda and N. Lapusta. Stable creeping fault segments can become destructive as a result of dynamic weakening. *Nature*, 493(7433) :518, 2013.

[127] R. O'Neil. Convolution operators and $L(p, q)$ spaces. *Duke Math. J.*, 30 :129–142, 1963.

[128] G. C. Papanicolaou and S. R. S. Varadhan. Boundary value problems with rapidly oscillating random coefficients. In *Random fields, Vol. I, II (Esztergom, 1979)*, volume 27 of *Colloq. Math. Soc. János Bolyai*, pages 835–873. North-Holland, Amsterdam-New York, 1981.

[129] R. Peierls. The size of a dislocation. *Proceedings of the Physical Society*, 52(1) :34, 1940.

[130] Y.-P. Pellegrini. Dynamic Peierls-Nabarro equations for elastically isotropic crystals. *Phys. Rev. B*, 81 :024101, 2010.

[131] Y.-P. Pellegrini. Equation of motion and subsonic-transonic transitions of rectilinear edge dislocations : A collective-variable approach. *Phys. Rev. B*, 90 :054120, 2014.

[132] Y.-P. Pellegrini and M. Josien. Numerical solutions of the multidimensional Weertman equation. in preparation.

[133] L. Pillon. *Modélisations du mouvement instationnaire et des interactions de dislocations*. PhD thesis, Paris 6, 2008.

[134] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical recipes*. Cambridge University Press, Cambridge, third edition, 2007.

[135] P. Rosakis. Supersonic dislocation kinetics from an augmented Peierls model. *Phys. Rev. Lett.*, 86(1) :95, 2001.

[136] D. H. Sattinger. On the stability of waves of nonlinear parabolic systems. *Advances in Math.*, 22(3) :312–355, 1976.

[137] A. Schädle, M. López-Fernández, and C. Lubich. Fast and oblivious convolution quadrature. *SIAM J. Sci. Comput.*, 28(2) :421–438, 2006.

[138] B. Schweizer and M. Veneroni. The needle problem approach to non-periodic homogenization. *Netw. Heterog. Media*, 6(4) :755–781, 2011.

[139] D. Serre. *Systems of conservation laws. 1.* Cambridge University Press, Cambridge, 1999.

[140] Z. Shen. Bounds of Riesz transforms on $L^p$ spaces for second order elliptic operators. *Ann. Inst. Fourier (Grenoble)*, 55(1) :173–197, 2005.

[141] Z. Shen. The Calderón-Zygmund lemma revisited. In *Lectures on the analysis of nonlinear partial differential equations. Part 2*, volume 2 of *Morningside Lect. Math.*, pages 203–224. Int. Press, Somerville, MA, 2012.

[142] E. Tadmor. The numerical viscosity of entropy stable schemes for systems of conservation laws. I. *Math. Comp.*, 49(179) :91–103, 1987.

[143] A. Talbot. The accurate numerical inversion of laplace transforms. *IMA Journal of Applied Mathematics*, 23(1) :97–120, 1979.

[144] L. Tartar. *An introduction to Sobolev spaces and interpolation spaces*, volume 3 of *Lecture Notes of the Unione Matematica Italiana*. Springer, Berlin ; UMI, Bologna, 2007.

[145] L. Tartar. *The general theory of homogenization*, volume 7 of *Lecture Notes of the Unione Matematica Italiana*. Springer-Verlag, Berlin ; UMI, Bologna, 2009.

[146] T.P. Theodoulidis. Struve functions. https ://fr.mathworks.com/matlabcentral/fileexchange/37302-struve-functions, 2012. Consulté le 10/01/2017.

[147] M. Toda. *Theory of nonlinear lattices*, volume 20 of *Springer Series in Solid-State Sciences*. Springer-Verlag, Berlin, second edition, 1989.

[148] J.-C. Tolédano. *Bases physiques de la plasticité des solides*. Editions Ecole Polytechnique, 2007.

[149] S. Venakides, P. Deift, and R. Oba. The Toda shock problem. *Comm. Pure Appl. Math.*, 44(8-9) :1171–1242, 1991.

[150] A. I. Volpert, V. A. Volpert, and V. A. Volpert. *Traveling wave solutions of parabolic systems*, volume 140 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1994.

[151] V. Volpert. *Elliptic partial differential equations. Vol. 2*, volume 104 of *Monographs in Mathematics*. Birkhäuser/Springer Basel AG, Basel, 2014.

[152] B. von Sydow, J. Hartford, and G. Wahnström. Atomistic simulations and Peierls-Nabarro analysis of the Shockley partial dislocations in palladium. *Comput. Mater. Sci.*, 15(3) :367–379, 1999.

[153] J. Wang and Q. Ma. Numerical techniques on improving computational efficiency of spectral boundary integral method. *International Journal for Numerical Methods in Engineering*, 102(10) :1638–1669, 2015.

[154] J. Weertman. Dislocations in uniform motion on slip or climb planes having periodic force laws. In *Mathematical Theory of Dislocations*, pages 178–202, New York, 1969. American Society of Mechanical Engineers.

[155] J. Weertman. *Stress dependence on the velocity of a dislocation moving on a viscously damped slip plane*, chapter 7, pages 75–83. The M.I.T. Press, 1969.

[156] Y. Xiang, H. Wei, P. Ming, and Weinan E. A generalized Peierls-Nabarro model for curved dislocations and core structures of dislocation loops in Al and Cu. *Acta Materialia*, 56(7) :1447–1460, 2008.

[157] L. Y. H. Yap. Some remarks on convolution operators and $L(p, q)$ spaces. *Duke Math. J.*, 36 :647–658, 1969.

[158] K. Yosida. *Functional analysis*. Classics in Mathematics. Springer-Verlag, Berlin, 1995.

[159] V. Zhakhovsky, M. M Budzevich, N. A Inogamov, I. Oleynik, and C. White. Two-zone elastic-plastic single shock waves in solids. *Physical review letters*, 107(13) :135502, 2011.

[160] X. Zhang, A. Acharya, N. J. Walkington, and J. Bielak. A single theory for some quasi-static, supersonic, atomic, and tectonic scale applications of dislocations. *Journal of the Mechanics and Physics of Solids*, 84 :145–195, 2015.

[161] A. Zhu, C. Jin, D. Zhao, Y. Xiang, and J. Huang. A numerical scheme for generalized Peierls-Nabarro model of dislocations based on the fast multipole method and iterative grid redistribution. *Commun. Comput. Phys.*, 18(05) :1282–1312, 2015.

# Annexes A

## A.1 Notations et conventions

Dans cette section, nous rassemblons quelques notations utilisées tout au long de ce document.

**Transformation de Fourier et de Laplace** La transformée de Fourier d'une fonction $u : \mathbb{R}^d \to \mathbb{R}$ suffisamment régulière et intégrable est définie comme suit :

$$\mathcal{F}\{u\}(k) = \widehat{u}(k) := \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}k\cdot x} u(x) \mathrm{d}x.$$

La transformation de Fourier inverse est notée :

$$\mathcal{F}^{-1}\{u\}(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}k\cdot x} u(k) \mathrm{d}k.$$

Par abus de notations, on notera parfois

$$\mathcal{F}\{u(x)\} := \mathcal{F}\{x \mapsto u(x)\},$$

et de même pour $\mathcal{F}^{-1}$.

Si $u : \mathbb{R}_+ \to \mathbb{R}$, on définit sa transformée de Laplace :

$$\mathcal{L}u(p) = \int_0^{+\infty} u(t)\mathrm{e}^{-pt}\mathrm{d}t.$$

La transformation de Laplace inverse est naturellement notée $\mathcal{L}^{-1}$.

**Opérateurs particuliers** On note $|\partial_x|$ l'opérateur défini par

$$|\partial_x|\, u(x) = \mathcal{F}^{-1}\left\{|k|\hat{u}(k)\right\}(x).$$

**Espaces fonctionnels** Nous utilisons en général les notations classiques de [36]. En outre, on utilise les notations suivantes :

$$\mathrm{C}_0^1(\Omega) := \left\{\phi \in \mathrm{C}^1(\Omega), \phi = 0 \quad \text{sur} \quad \partial\Omega\right\},$$
$$\|u\|_{\dot{\mathrm{C}}^{0,\beta}(\Omega)} := \sup_{x\neq y\in\Omega} \frac{|u(x)-u(y)|}{|x-y|^\beta}.$$

Si $p > 1$, on pose $\mathrm{L}^{p,\infty}(\Omega)$ l'espace de Marcinkiewicz équipé de la semi-norme (voir [21])

$$\|f\|_{\mathrm{L}^{p,\infty}} := \sup_{t>0} t\,|\{x \in \Omega, |f(x)| > t\}|^{1/p}.$$

Par abus de notation, on omet parfois de préciser la dimension de l'espace d'arrivée des fonctions. Ainsi, il arrive que l'on note $\nabla f \in \mathrm{L}^p(\Omega)$ à la place de $\nabla f \in \mathrm{L}^p\left(\Omega, \mathbb{R}^d\right)$.

Soit $E$ un espace fonctionnel (par exemple $\mathrm{L}^2(\mathbb{R}^d)$, $\mathrm{C}^{0,\alpha}(\mathbb{R}^d)$. On utilise la notation $E_{\mathrm{per}}$ pour noter les espaces fonctionnels dont les éléments sont Q-périodiques ; par exemple :

$$\mathrm{C}^\infty_{\mathrm{per}}(\mathbb{R}^d) := \left\{u \in \mathrm{C}^\infty(\mathbb{R}^d), u \text{ est } \mathrm{Q}-\text{périodique}\right\}.$$

On utilise la notation $E_{\mathrm{unif}}$ comme suit :

$$\mathrm{L}^2_{\mathrm{unif}}(\mathbb{R}^d) := \left\{ f \in \mathrm{L}^2_{\mathrm{loc}}(\mathbb{R}^d), \sup_{x\in\mathbb{R}^d} \|f\|_{\mathrm{L}^2(\mathrm{B}(x,1))} < +\infty \right\}.$$

**Vecteurs et fonctions de plusieurs variables** Lorsqu'on considère $F \in \mathbb{R}^d$ multi-indicée, on note $F$ pour $(F_{j_1,\ldots,j_n})_{j_1,\ldots,j_n\in[\![1,d]\!]}$, et parfois $\boldsymbol{F}$ si on veut insister sur son caractère vectoriel. Lorsqu'on considère une fonction $F(x,y)$ dépendant de deux variables $x$ et $y$, on note $\nabla_x F(a,b)$, $\mathrm{div}_x(F(a,b))$, $\Delta_x F(a,b)$, respectivement $\nabla_y F(a,b)$, $\mathrm{div}_y(F(a,b))$, $\Delta_y F(a,b)$ le gradient, la divergence, le laplacien de $F$ en $(x,y) = (a,b)$ selon sa première variable $x$, respectivement sa seconde variable $y$.

**Fonctions particulières** On pose $\mathcal{G}_\Delta$ la fonction de Green sur $\mathbb{R}^d$ du Laplacien, qui vérifie :

$$\mathcal{G}_\Delta(x) = \frac{\Gamma(d/2)}{2(d-2)\pi^{d/2}}|x|^{2-d} =: C_d|x|^{2-d}.$$

On note $J_k$ la $k^{\text{ème}}$ fonction fonction de Bessel de la première espèce.

Les fonctions spéciales $C_i$, $\mathfrak{R}_i^\alpha$, $W$ and $\mathcal{W}$, et les coefficients $\kappa_i^\alpha$ sont définis dans la Section A.5.1.

**Notations sur des figures géométriques** Le cube Q, la sphère $S(x,R)$ pour $x \in \mathbb{R}^d$, $R > 0$ sont définis par

$$\mathrm{Q} := [-1/2, 1/2]^d \qquad \text{et} \qquad S(x,R) := \left\{y \in \mathbb{R}^d, |x-y| = R\right\}.$$

Soit $\Omega \subset \mathbb{R}^d$ un ouvert régulier borné. Définissons le diamètre $\mathrm{Diam}(\Omega)$ et la distance au bord $\mathrm{d}(x, \partial\Omega)$ par :

$$\mathrm{Diam}(\Omega) := \sup_{x,y\in\Omega} |x-y| \qquad\qquad \mathrm{d}(x,\partial\Omega) := \sup_{y\in\partial\Omega} |x-y|.$$

On pose

$$\Omega(x,R) = \Omega \cap \mathrm{B}(x,R), \qquad \Gamma_\Omega(x,R) = \partial\Omega \cap \overline{\mathrm{B}(x,R)}.$$

On pose aussi, pour $\phi : \mathbb{R}^{d-1} \to \mathbb{R}$,

$$D_\phi(R) := \left\{ x \in \mathbb{R}^d, x_d < \phi\left((x_1, \cdots, x_{d-1})\right) \right\} \cap \mathrm{B}(0,R),$$
$$\Delta_\phi(R) := \Gamma_{D_\phi(R)}(0,R) = \left\{ x \in \mathbb{R}^d, x_d = \phi\left((x_1, \cdots, x_{d-1})\right) \right\} \cap \overline{\mathrm{B}(0,R)}.$$

Et on introduit le cône tronqué

$$C_{K_0}(R) := \left\{ x \in \mathrm{B}(0,R), x_d < 0, |x - x_d e_d| < K_0 |x_d| \right\}.$$

**Exposants particuliers** Si $p \in [1, +\infty]$, on lui associe $p' \in [1, +\infty]$ l'exposant conjugué, défini par :

$$\frac{1}{p} + \frac{1}{p'} = 1.$$

On utilise aussi l'exposant $\nu_r$ défini par

$$\nu_r := \min\left(1, \frac{d}{r}\right) \in ]0,1].$$

## A.2 Annexes du Chapitre 1

### A.2.1 Equation de Weertman complète

Nous explicitons ici les grandeurs physiques intervenant dans l'équation de Weertman.

Supposons que la dislocation considérée ne se déforme pas, mais se meut à vitesse constante sous l'effet d'un chargement uniforme $\sigma$. C'est à dire que $\eta$, supposée à valeur scalaire, est un front progressif

$$\eta(t,x) = \phi(x - vt), \qquad\qquad \text{pour} \quad (t,x) \in \mathbb{R} \times \mathbb{R},$$

où $v \in \mathbb{R}$ est une certaine vitesse. Dans ce cas, le modèle de Peierls amène à écrire l'équation suivante sur $\phi$, dite «équation de Weertman» (voir [135]) :

$$-\mu A(v) \left|\partial_x\right| \phi(x) + \mu \left( B(v) + \frac{\alpha v}{2c_\mathrm{s}} \right) \phi'(x) = f'(\eta) - \sigma \quad \text{pour} \quad x \in \mathbb{R}. \tag{A.1}$$

Précisons les termes intervenant dans l'équation (A.1). La constante $\mu$ est le module de cisaillement ; le potentiel $f$ (aussi appelé $\gamma$-surface [108]) est borné et périodique, et

induit une force (par unité de surface) $f'$ ; $\sigma$ est une contrainte appliquée sur la dislocation de même dimension que $f'$ ; les scalaires sans dimension $A(v)$ et $B(v)$ sont donnés par les formules [130, (46), (47) et (48)] ; le symbole $c_{\mathrm{s}}$ désigne la vitesse des ondes de cisaillement. Le scalaire sans dimension $\alpha > 0$ traduit un phénomène de frottement non-radiatif de nature visco-plastique (voir [65, 135]).

Il existe trois régimes de vitesse :
- le régime subsonique pour $|v| < c_{\mathrm{s}}$ ;
- le régime transonique pour $c_{\mathrm{s}} < |v| < c_{\mathrm{l}}$, où $c_{\mathrm{l}}$ est la vitesse des ondes longitudinales (ce régime n'existe pas pour le mode III) ;
- le régime supersonique, pour $|v| > c_{\mathrm{l}}$ dans le cas des modes I et II, et pour $|v| > c_{\mathrm{s}}$ dans le cas du mode III.

Dans le cas particulier où $\sigma = 0$, on déduit que $v = 0$. On retrouve alors la célèbre équation de Peierls-Nabarro [129].

Adimensionnons l'équation (A.1). On suppose à partir de maintenant que $A(v) > 0$ (ce qui implique que les dislocations considérées sont en régime non-supersonique). On effectue les substitutions suivantes $x \mapsto bA(v)x$, $\phi \mapsto b\phi$, $f' - \sigma \mapsto \mu F'$, où $b$ est le module du vecteur de Burgers, et on pose le rapport

$$c := \frac{B(v) + \frac{\alpha v}{2c_{\mathrm{s}}}}{A(v)}. \tag{A.2}$$

Alors, l'équation (A.1) peut s'écrire de manière plus compacte sous la forme suivante :

$$-\left|\partial_x\right| \phi(x) + c\phi'(x) = F'(\phi(x)) \qquad \text{pour} \quad x \in \mathbb{R}, \tag{A.3}$$

c'est à dire l'équation (1.38) de l'Introduction.

## A.3 Annexes du Chapitre 2

### A.3.1 Résultats de la littérature

Nous rassemblons dans cette section quelques résultats classiques de la littérature, afin que le lecteur puisse les avoir sous les yeux.

**Lemme A.3.1** (Lemme 1.1 p. 4 de [85]). *Soient $u$ et $v \in \mathrm{L}^2\left(\Omega, \mathbb{R}^d\right)$. Supposons que des suites $(u_n)$ et $(v_n)$ satisfont*

$$\begin{cases} u_n \underset{n\to+\infty}{\rightharpoonup} u & \text{dans } \mathrm{L}^2(\Omega), \\ v_n \underset{n\to+\infty}{\rightharpoonup} v & \text{dans } \mathrm{L}^2(\Omega), \end{cases}$$

*et que $u_n$ et $v_n$ satisfont par ailleurs, pour tout $n \in \mathbb{N}$,*

$$\begin{cases} \mathrm{rot}(u_n) = 0, \\ \mathrm{div}(v_n) \to f & \text{dans } \mathrm{H}^{-1}(\Omega). \end{cases}$$

*Alors, pour tout $\phi \in \mathrm{C}_0^1(\Omega)$, on a la convergence suivante :*

$$\int_\Omega (u_n \cdot v_n)\, \phi \to \int_\Omega (u \cdot v)\phi.$$

**Lemme A.3.2** (Lemme de Cacciopoli (Proposition 2.1 p. 76 de [63]))**.** *Soit $M$ un champ de matrices satisfaisant l'Hypothèse 1 et $\Omega$ un ouvert lipschitzien. Si $u$ satisfait*

$$\begin{cases} -\operatorname{div}(M(x)\cdot\nabla u(x)) = 0 & \text{dans} \quad \Omega(0,R), \\ u = 0 & \text{sur} \quad \Gamma_\Omega(0,R), \end{cases} \tag{A.4}$$

*alors il existe une constante $C(\mu)$ telle que*

$$\int_{\Omega(0,R/2)} |\nabla u|^2 \leq \frac{C(\mu)}{R^2} \int_{\Omega(0,R)} |u|^2. \tag{A.5}$$

*Remarque* 68. On peut avoir éventuellement $\Gamma_\Omega(0,R) = \emptyset$ dans (A.4) ci-dessus. En fait, la Proposition 2.1 p. 76 de [63] n'inclut pas le cas où $\Gamma_\Omega(0,R) \neq \emptyset$. Toutefois, on se convainc que la preuve fonctionne exactement pareil dans le cas ci-dessus.

Les Théorèmes [64, Th. 8.2 p. 202] et [64, Cor. 8.36 p. 212] ont un corollaire utile :

**Corollaire A.3.3.** *Soit $M$ satisfaisant les Hypothèses 1 et 2. Supposons que $u \in \mathrm{H}^1(\mathrm{B}(0,1))$ satisfait*

$$-\operatorname{div}(M(x)\cdot\nabla u(x)) = 0 \quad \text{dans} \quad \mathrm{B}(0,1),$$

*dans $\mathrm{B}(0,1)$. Alors il existe une constante $C$ ne dépendant que de $M$ telle que*

$$\|\nabla u\|_{\mathrm{L}^\infty(\mathrm{B}(0,1/2))} \leq C \left( \fint_{\mathrm{B}(0,1)} |u|^2 \right)^{1/2}.$$

**Lemme A.3.4** (Conséquence du Théorème 8.25 p. 202 de [64].)**.** *Soit $\Omega$ un ouvert borné régulier de classe $\mathrm{C}^{1,\alpha}$ et $R > 0$. Soit $A \in \mathrm{L}^\infty\left(\mathrm{B}(0,2R), \mathbb{R}^{d^2}\right)$ satisfaisant l'Hypothèse 1. Supposons que $u \in \mathrm{H}^1(\Omega(0,2R))$ satisfait*

$$\begin{cases} -\operatorname{div}\left(A\left(x\right)\cdot\nabla u(x)\right) = 0 & \text{dans} \quad \Omega(0,2R), \\ u = g & \text{sur} \quad \Gamma_\Omega(0,2R). \end{cases}$$

*Alors, pour tout $p > 1$, on a*

$$\|u\|_{\mathrm{L}^\infty(\Omega(0,R))} \leq C\, \|g\|_{\mathrm{L}^\infty(\Gamma_\Omega(0,2R))} + C \left( R^{-d} \int_{\Omega(0,2R)} |u|^p \right)^{1/p}, \tag{A.6}$$

*où $C$ ne dépend que de $\mu$ et de $p$.*

Le Lemme A.3.4 joue un rôle analogue au Lemme 3.1 de [94], et a un corollaire utile :

**Lemme A.3.5** (Conséquence des Théorèmes 8.25 p. 202 et 8.29 p. 205 de [64]). *Soit $A$ satisfaisant l'Hypothèse 1, $\Omega$ un ouvert borné régulier de classe $\mathrm{C}^{1,\alpha}$ et $g \in \mathrm{C}^{0,\beta}(\mathrm{B}(0,1))$, pour $\beta > 0$. Supposons que $u \in \mathrm{H}^1(\Omega(0,1))$ satisfait*

$$\begin{cases} -\operatorname{div}(A(x)\cdot\nabla u(x)) = 0 & \text{dans} \quad \Omega(0,1), \\ u = g & \text{sur} \quad \Gamma_\Omega(0,1), \end{cases} \tag{A.7}$$

*Alors il existe $\gamma > 0$ et une constante $C$ ne dépendant que de $\mu$, de $\beta$, et de $\Omega$, tels que*

$$\|u\|_{\mathrm{C}^{0,\gamma}(\Omega(0,1/2))} \leq C\,\|u\|_{\mathrm{L}^2(\Omega(0,1))} + C\,\|g\|_{\mathrm{C}^{0,\beta}(\mathrm{B}(0,1))}\,. \tag{A.8}$$

**Lemme A.3.6** (Caractérisation de Campanato, Théorème 1.2 p. 70 de [63]). *Soit $\Omega$ un domaine borné à bord lipschitzien, alors, pour tout $\rho \in ]0,1[$, il existe $C_\rho > 0$ tel que, pour toute fonction $u : \Omega \to \mathbb{R}$, on a l'équivalence des semi-normes suivantes :*

$$C_\rho \sup_{x,y\in\Omega} \frac{|u(x)-u(y)|}{|x-y|^\rho} \geq \sup_{x\in\Omega, r>0} \left( r^{-d-2\rho} \int_{\mathrm{B}(x,r)\cap\Omega} \left| u - \fint_{\mathrm{B}(x,r)\cap\Omega} u \right|^2 \right)^{1/2},$$

$$C_\rho^{-1} \sup_{x,y\in\Omega} \frac{|u(x)-u(y)|}{|x-y|^\rho} \leq \sup_{x\in\Omega, r>0} \left( r^{-d-2\rho} \int_{\mathrm{B}(x,r)\cap\Omega} \left| u - \fint_{\mathrm{B}(x,r)\cap\Omega} u \right|^2 \right)^{1/2}.$$

**Théorème A.3.7** (Fonction de Green de Dirichlet d'une équation à coefficients constants). *Soit $A^\star$ une matrice constante satisfaisant l'Hypothèse 1, et $\Omega$ un domaine borné régulier de classe $\mathrm{C}^{1,1}$. Soit $G^\star$ la fonction de Green de Dirichlet de $-\operatorname{div}(A^\star\cdot\nabla)$ dans $\Omega$. Alors, il existe une constante $C > 0$ telle que*

$$|\nabla_x G^\star(x,y)| \leq C|x-y|^{-d+1} \qquad \text{pour tous} \quad x \neq y \in \Omega, \tag{A.9}$$

$$\left\|\nabla_x^2 G^\star\right\|_{\mathrm{L}^q(\Omega(y_0,4R))} \leq CR^{d/q-d} \qquad \text{pour tous} \quad R > 0, y_0 \in \Omega, q \in ]1,\infty[. \tag{A.10}$$

*En outre, pour tout domaine ouvert $\Omega_1 \subset\subset \Omega$, pour tout $m,n \in \mathbb{N}$, il existe une constante $C_{mn}$ telle que*

$$\left|\nabla_x^m \nabla_y^n G(x,y)\right| \leq C|x-y|^{-d+2-n-m} \qquad \text{pour tous} \quad x \neq y \in \Omega_1. \tag{A.11}$$

*Démonstration.* L'Estimation (A.9) est une conséquence de [72, Th. 3.3].

L'Estimation (A.10) est une conséquence de [64, Th. 9.13 p. 239] et de [72, Th. 1.1].

L'Estimation (A.11) est une conséquence des théorèmes de régularité classique, appliqués itérativements aux dérivées de $G^\star$. □

**Lemme A.3.8** (Formule (1.12) de [72]). *Soient $\Omega$ un domaine ouvert borné de $\mathbb{R}^d$, $p \in ]1,+\infty[$, et $1 \leq q < p$. Supposons que $f \in \mathrm{L}^{p,\infty}(\Omega)$. Alors, il existe une constante $C$ ne dépendant que de $d$, $p$, $q$ telle que,*

$$\|f\|_{\mathrm{L}^q(\Omega)} \leq C\,|\Omega|^{1/q-1/p}\,\|f\|_{\mathrm{L}^{p,\infty}(\Omega)}\,. \tag{A.12}$$

*Démonstration.* Par le principe de Cavalieri, pour tout $T \in \mathbb{R}_+^\star$,

$$\begin{aligned}\int_\Omega |f|^q =& q\int_0^{+\infty} t^{q-1}\,|\{f>t\}|\,\mathrm{d}t\\ \leq& q\int_T^{+\infty} t^{q-1-p}\,\|f\|_{\mathrm{L}^{p,\infty}}^p\,\mathrm{d}t + q\,|\Omega|\int_0^T t^{q-1}\mathrm{d}t\\ \leq& C\left(\|f\|_{\mathrm{L}^{p,\infty}}^p\,T^{q-p} + |\Omega|\,T^q\right). \end{aligned} \tag{A.13}$$

On minimise le membre de droite de (A.13) en posant :

$$T = |\Omega|^{-\frac{1}{p}}\,\|f\|_{\mathrm{L}^{p,\infty}}.$$

D'où (A.12). □

**Lemme A.3.9** (Corollaire du Théorème 2.4 de [141])**.** *Soit* $\mathrm{B}_0$ *une boule de* $\mathbb{R}^d$*,* $F \in \mathrm{L}^2(4\mathrm{B}_0)$*. Soient* $q > p > 2$*,* $f \in \mathrm{L}^p(4\mathrm{B}_0)$*. Supposons qu'il existe* $K > 0$*, telles que pour toute boule* $\mathrm{B} \subset 2\mathrm{B}_0$*, avec* $|\mathrm{B}| \leq 1/2\,|\mathrm{B}_0|$*, il existe* $F_1$*,* $F_2$ *définies sur* $2\mathrm{B}$ *telles que*

$$|F| \leq |F_1| + |F_2| \quad \textit{sur} \quad 2\mathrm{B}, \tag{A.14}$$

$$\left(⨍_{2\mathrm{B}} |F_1|^2\right)^{1/2} \leq K \sup_{\mathrm{B}\subset\mathrm{B}'\subset 4\mathrm{B}_0} \left(⨍_{\mathrm{B}'} |f|^2\right)^{1/2}, \tag{A.15}$$

$$\left(⨍_{2\mathrm{B}} |F_2|^q\right)^{1/q} \leq K\left\{\left(⨍_{3\mathrm{B}} F^2\right)^{1/2} + \sup_{\mathrm{B}\subset\mathrm{B}'\subset 4\mathrm{B}_0} \left(⨍_{\mathrm{B}'} |f|^2\right)^{1/2}\right\}, \tag{A.16}$$

*où* $\mathrm{B}'$ *désigne toujours une boule. Alors,* $F \in \mathrm{L}^p(\mathrm{B}_0)$ *et*

$$\left(⨍_{\mathrm{B}_0} |F|^p\right)^{1/p} \leq C\left\{\left(⨍_{4\mathrm{B}_0} |F|^2\right)^{1/2} + \left(⨍_{4\mathrm{B}_0} |f|^p\right)^{1/p}\right\},$$

*où* $C$ *dépend seulement de* $K$*,* $p$ *et* $q$*.*

### A.3.2 Autres résultats techniques

**Lemme A.3.10.** *Soit* $u \in \mathrm{L}^2_{\mathrm{unif}}\left(\mathbb{R}^d\right)$ *tel que, pour toutes suites* $y_n \in \mathbb{R}^d$*,* $\varepsilon_n \to 0$*,*

$$⨍_Q u\left(y_n + \frac{x}{\varepsilon_n}\right)\mathrm{d}x \to 0. \tag{A.17}$$

*Soit* $\Omega$ *un ouvert borné. Alors, pour toutes suites* $y_n \in \mathbb{R}^d$*,* $\varepsilon_n \to 0$*,*

$$u\left(y_n + \frac{x}{\varepsilon_n}\right) \rightharpoonup 0 \quad \textit{dans} \quad \mathrm{L}^2(\Omega).$$

*Démonstration.* On pose $u^n(x) = u\left(y_n + \frac{x}{\varepsilon_n}\right)$. Comme $u \in \mathrm{L}^2_{\mathrm{unif}}\left(\mathbb{R}^d\right)$, $u^n$ est bornée dans $\mathrm{L}^2(\Omega)$. Donc, quitte à extraire,

$$u^n \rightharpoonup u \quad \text{dans} \quad \mathrm{L}^2(\Omega).$$

Soit $h > 0$. On appelle $C_i, i \in [\![1, N_h]\!]$ les cubes délimités par le réseau $h\mathbb{Z}^d$ qui sont à l'intérieur de $\Omega$, et on appelle $c_i$ leur centre. On se donne une fonction $\phi \in \mathrm{L}^2(\Omega)$ qui est nulle hors de $\cup_i C_i$, et constante sur chaque $C_i$. Alors

$$\int_\Omega \phi(x) u^n(x) \mathrm{d}x = \sum_{i=1}^{N_h} \phi(c_i) \int_{C_i} u^n(x) \mathrm{d}x.$$

La convergence (A.17) implique

$$\int_{C_i} u^n(x) \mathrm{d}x \to 0, \qquad \forall i \in [\![1, N_h]\!].$$

Ainsi,

$$\int_\Omega \phi(x) u^n(x) \mathrm{d}x \to 0.$$

Comme de telles fonctions $\phi$, pour $h > 0$, forment un sous-ensemble dense de $\mathrm{L}^2(\Omega)$, on en déduit que

$$u = 0.$$

D'où le résultat souhaité. □

**Lemme A.3.11.** *Soit $A$ un champ de matrices périodique satisfaisant l'Hypothèse 1. Alors pour tout $j \in [\![1, d]\!]$, il existe $w_j$ des correcteurs périodiques associés à $A$, c'est à dire satisfaisant*

$$-\mathrm{div}\left(A(x)\left(e_j + \nabla w_j(x)\right)\right) = 0,$$

*et $w_j \in \mathrm{H}^1(Q)$ est périodique, unique si on fixe $\int_Q w_j = 0$. De plus, si $A^\star$ est la matrice homogénéisée relative à $A$, il existe un potentiel périodique $B_k^{ij}$ associé à*

$$M_k^i(x) = A^\star_{ik} - A(x)\left(\delta_{ik} + \partial_i w_k(x)\right). \tag{A.18}$$

*Démonstration.* L'existence et l'unicité à l'ajout d'une constante près des correcteurs est une conséquence directe du théorème de Lax-Milgram.

Construisons maintenant le potentiel $B_k^{ij}$. Résolvons le problème suivant

$$\Delta N_k^i = M_k^i \qquad \text{dans} \quad Q, \tag{A.19}$$

avec $N_k^i \in \mathrm{H}^1_{per}(Q)$. Comme

$$\int_Q M_i^k(x) \mathrm{d}x = 0,$$

par l'alternative de Fredholm, le Problème (A.19) a une unique solution, à l'ajout d'une constante près. On pose ensuite

$$B_k^{ij} = \partial_i N_k^j - \partial_j N_k^i, \tag{A.20}$$

qui est le potentiel recherché. □

### A.3.3 Arbres des preuves

Nous illustrons ici les liens logiques entre les résultats d'estimations sur la solution du problème oscillant. On suppose que $A$ satisfait les Hypothèses 1 et 3. Les flèches bleues et rouges indiquent qu'un résultat ou une hypothèse à leur base est utilisé pour démontrer un résultat à leur pointe.

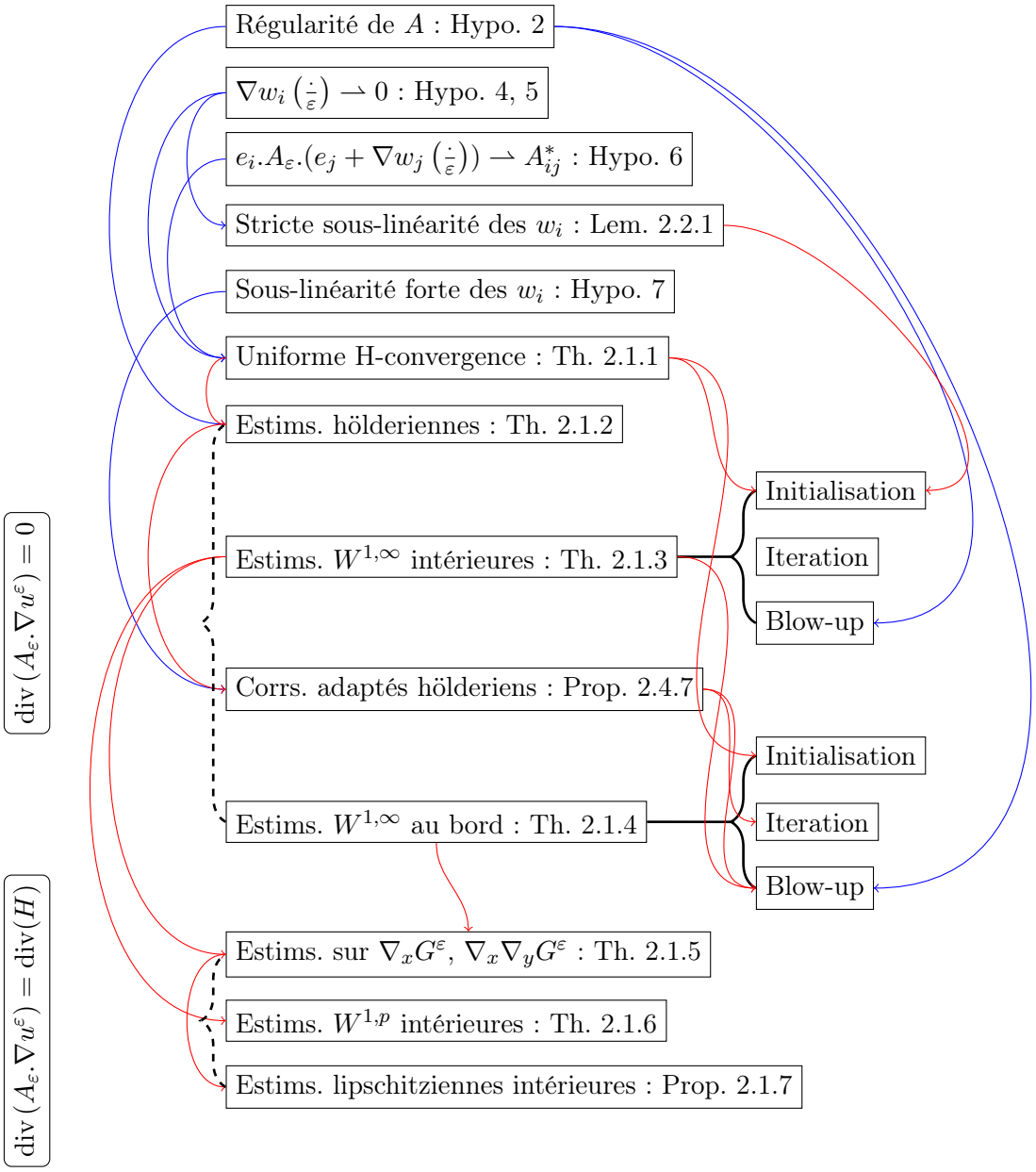



FIGURE A.1 – Structure logique entre les résultats d'estimation

## A.4 Annexes du Chapitre 4

Les annexes de ce chapitre sont en anglais. Nous remercions Gilles Francfort pour ses suggestions en vue de simplifier les preuves de cette section. La principale difficulté des preuves ci-dessous tient au fait que l'on effectue les preuves dans $\mathrm{L}^\infty(\mathbb{R})$, qui est un espace naturel où étudier les solutions de (4.15) (en effet, on s'intéresse ici à des solutions qui ne tendent pas vers 0 en $+\infty$). Dans cet espace, le semi-groupe engendré par $-|\partial_x|$ n'est pas un semi-groupe de contractions (voir [43, Def. 3.4.1]), car il n'est pas continu vis-à-vis de la norme de $\mathrm{L}^\infty$.

### A.4.1 Existence and uniqueness of the solution to the evolution equation

We prove that (4.15) has a unique weak solution.

We first show that the operator $|\partial_x|$ generates a contraction semi-group on the Banach space $\mathrm{L}^1(\mathbb{R})$, the dual of which is the space $\mathrm{L}^\infty(\mathbb{R})$ :

**Lemma A.4.1.** *The semi-group induced by $u \mapsto K_t * u$ (defined by (4.47)) on $\mathrm{L}^1(\mathbb{R})$ is a contraction semi-group, the generator of which is $-|\partial_x|$ on the domain*

$$\mathcal{D}(|\partial_x|) := \left\{ u \in \mathrm{L}^1(\mathbb{R}), |\partial_x| \, u \in \mathrm{L}^1(\mathbb{R}) \right\}. \tag{A.21}$$

*Proof.* We first show that $K_t$ induces a contraction semi-group (see [43, Def. 3.4.1]). As is well-known (see [70, Eq. (9) p. 1119]), (4.47) yields in Fourier variables :

$$\mathcal{F}\{K_t\}(t,k) = \mathrm{e}^{-|k|t}. \tag{A.22}$$

Hence, $K_t$ is indeed a semi-group (since convolution is turned into multiplication by Fourier transform). Moreover, $K_t$ is a probability measure ; therefore, for any $u \in \mathrm{L}^1(\mathbb{R})$, the Young inequality yields

$$\|K_t * u\|_{\mathrm{L}^1(\mathbb{R})} \leq \|u\|_{\mathrm{L}^1(\mathbb{R})}.$$

Next, we show that, for any $u \in \mathrm{L}^1(\mathbb{R})$, then

$$\lim_{t \to 0} K_t * u \to u \quad \text{in} \quad \mathrm{L}^1(\mathbb{R}). \tag{A.23}$$

Since $\mathrm{C}_c^\infty(\mathbb{R})$ is dense in $\mathrm{L}^1(\mathbb{R})$, it is sufficient to show the latter property only for $u \in \mathrm{C}_c^\infty(\mathbb{R})$. Then, as a consequence of the dominated convergence theorem, (A.23) is satisfied. As a conclusion, $K_t$ is a contraction semi-group on $\mathrm{L}^1(\mathbb{R})$.

By (A.22), this semi-group is generated by $|\partial_x|$. By [43, Th. 3.4.3], the operator $|\partial_x|$ is m-dissipative and with a dense domain in $\mathrm{L}^1(\mathbb{R})$ : the domain $\mathcal{D}(|\partial_x|)$ defined by (A.21). Remark that $\mathrm{C}_c^\infty(\mathbb{R}) \subset \mathcal{D}(|\partial_x|)$. □

Actually, the semi-group defined in Lemma A.4.1 can be extended on $\mathrm{L}^\infty(\mathbb{R})$ (as a strong Feller semi-group [32, Def. 11]).

**Lemma A.4.2.** *Let $T > 0$, $u_0 \in \mathrm{L}^\infty(\mathbb{R})$ and $g \in \mathrm{L}^\infty([0,T]\times\mathbb{R})$. Then there exists a unique weak solution $u \in \mathrm{L}^\infty([0,T]\times\mathbb{R})$ to (4.49) in the sense that, for all $\phi \in \mathrm{C}^1_{\mathrm{c}}([0,T), \mathcal{D}(|\partial_x|))$ (for $\mathcal{D}(|\partial_x|)$ defined by (A.21)), (4.50) holds. This solution can be written thanks to the Duhamel formula (4.51), where $K_t$ is defined by (4.47).*

Existence is due to the fact that (4.51) is well-defined ; uniqueness is showed using the adjoint problem of (4.49).

*Proof.* By Lemma A.4.1, a duality argument, *i.e.* [158, Th. p. 273], implies that $K_t$ defined above induces a semi-group on $\mathrm{L}^\infty(\mathbb{R})$. We show that this semi-group generates the unique solutions of the weak formulation (4.50).

We first show that $u$ defined by (4.51) is a solution to (4.50). As a consequence of the following inequality :

$$\left\{\int_0^t |K_{t-s} * g(s,\cdot)|\,\mathrm{d}s\right\}(x) \leq t\,\|g\|_{\mathrm{L}^\infty((0,T)\times\mathbb{R})}\,, \tag{A.24}$$

the function $u$ is well-defined. Let $\phi \in \mathrm{C}^1_{\mathrm{c}}([0,T), \mathcal{D}(|\partial_x|))$. Since $K_t$ is a probability measure, by the Young inequality, there holds

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}}\int_{\mathbb{R}} \left(|K_t(x-y)u_0(y)| + \int_0^t |K_{t-s}(x-y)g(s,y)|\,\mathrm{d}s\right) \\
&\quad (|-\partial_t\phi(t,x)| + ||\partial_x|\,\phi(t,x)|)\,\mathrm{d}y\mathrm{d}x\mathrm{d}t \\
&\leq T\left(\|u_0\|_{\mathrm{L}^\infty(\mathbb{R})} + \|g\|_{\mathrm{L}^1([0,T],\mathrm{L}^\infty(\mathbb{R}))}\right)\left(\|-\partial_t\phi\|_{\mathrm{L}^\infty(\mathrm{L}^1(\mathbb{R}))} + \||\partial_x|\,\phi\|_{\mathrm{L}^\infty(\mathrm{L}^1(\mathbb{R}))}\right) \\
&< +\infty.
\end{aligned}$$

Therefore, using Fubini's theorem, the first integral of (4.50) can be rewritten as :

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}} u(t,x)\,(-\partial_t + |\partial_x|)\,\phi(t,x)\mathrm{d}x\mathrm{d}t \\
&= \int_{\mathbb{R}} u_0(x) \int_0^T \{K_t * (-\partial_t + |\partial_x|)\,\phi(t,\cdot)\}\,(x)\mathrm{d}t\mathrm{d}x \\
&\quad + \int_0^T \int_{\mathbb{R}} g(s,y)\left(\int_s^T \{K_{t-s} * (-\partial_t + |\partial_x|)\,\phi(t,\cdot)\}\,(y)\mathrm{d}t\right)\mathrm{d}y\mathrm{d}s
\end{aligned}$$

Yet, by definition, for $\phi \in \mathcal{D}(|\partial_x|)$,

$$K_t * |\partial_x|\,\phi = -\frac{\mathrm{d}}{\mathrm{d}t}K_t * \phi.$$

Therefore,

$$\begin{aligned}
\int_0^T \{K_t * (-\partial_t + |\partial_x|)\,\phi(t,\cdot)\}\,(x)\mathrm{d}t &= -\int_0^T \frac{\mathrm{d}}{\mathrm{d}t}\{K_t * \phi(t,\cdot)\}\,(x)\mathrm{d}t \\
&= \{K_0 * \phi(0,\cdot)\}\,(x) - \{K_T * \phi(T,\cdot)\}\,(x) \\
&= \phi(0,x),
\end{aligned}$$

since $\phi(T,\cdot)=0$. Similarly, there holds

$$\int_s^T \left\{K_{t-s} * \left(-\partial_t + |\partial_x|\right)\phi(t,\cdot)\right\}(y)\mathrm{d}t = \phi(s,y).$$

As a conclusion, the function $u$ defined by (4.51) satisfies (4.50).

Proving uniqueness reduces to showing that the only weak solution to the homogeneous problem (4.46) with initial condition $u_0 = 0$ is zero. For that purpose, we introduce the adjoint problem

$$\begin{cases} -\partial_t\phi(t,x) + |\partial_x|\,\phi(t,x) = h(t,x) & \text{for} \quad x\in\mathbb{R}, \\ \phi(T,x) = 0 & \text{for} \quad x\in\mathbb{R}, \end{cases} \tag{A.25}$$

for $h \in \mathrm{C}_{\mathrm{c}}^{\infty}\left((0,T)\times\mathbb{R}\right)$. Going backward in time, thanks to [43, Prop. 4.1.6], this problem has a strong solution $\phi \in \mathrm{C}\left([0,T],\mathcal{D}(|\partial_x|)\right)\cap \mathrm{C}^1\left([0,T],\mathrm{L}^1(\mathbb{R})\right)$ :

$$\phi(t,x) = \left\{\int_t^T K_{s-t} * h(s,.)\mathrm{d}s\right\}(x). \tag{A.26}$$

While testing (4.50) with $\phi$, we obtain

$$\int_0^T\int_{\mathbb{R}} u(t,x)h(t,x)\mathrm{d}x\mathrm{d}t = 0. \tag{A.27}$$

Equation (A.27) being true for all $h \in \mathrm{C}_{\mathrm{c}}^{\infty}\left((0,T)\times\mathbb{R}\right)$, we deduce that $u=0$. □

Now, we establish the first part of Theorem 4.1.1 ; namely, (4.15) has a unique weak solution.

**Theorem A.4.1.** *Let $F \in \mathrm{C}^2(\mathbb{R})\cap \mathrm{W}^{2,\infty}(\mathbb{R})$ and $u_0 \in \mathrm{L}^{\infty}(\mathbb{R})$. Then there exists a unique weak solution $u \in \mathrm{L}^{\infty}_{\mathrm{loc}}\left(\mathbb{R}_+,\mathrm{L}^{\infty}(\mathbb{R})\right)$ to (4.15). Moreover, $u$ can be expressed by (4.52).*

*Proof.* The proof is done by a classical fixed-point argument (see for example [43, Sec. 4.3 p. 56]). We proceed by analysis and synthesis.

**Analysis** Assume that $u \in \mathrm{L}^{\infty}_{\mathrm{loc}}\left(\mathbb{R}_+,\mathrm{L}^{\infty}(\mathbb{R})\right)$ is a weak solution to (4.15). As $F'$ is bounded, then $F'(u)$ is also bounded ; thus Lemma A.4.2 yields (4.52). Next, assume that $u_1$ and $u_2$ are two weak solutions to (4.15). We prove that $w := u_1 - u_2$ is zero. Thanks to (4.52), we have

$$w(t,x) = \left\{\int_0^t K_{t-s} * \left[F'\left(u_2(s,.)\right) - F'\left(u_1(s,.)\right)\right]\mathrm{d}s\right\}(x).$$

Since $F'$ is bounded and since $K_t$ is a probability measure, then, for all $t > s \geq 0$ and $x\in\mathbb{R}$,

$$\left|K_{t-s} * \left[F'\left(u_2(s,.)\right) - F'\left(u_1(s,.)\right)\right](x)\right| \leq 2\left\|F'\right\|_{\mathrm{L}^{\infty}(\mathbb{R})}.$$

Therefore, $w$ is bounded. Furthermore, by Taylor expansion, there holds

$$\begin{aligned}\|w(t,\cdot)\|_{\mathrm{L}^\infty(\mathbb{R})} &\leq \int_0^t \|K_{t-s}\|_{\mathrm{L}^1(\mathbb{R})} \left\|F''\right\|_{\mathrm{L}^\infty(\mathbb{R})} \|w(s,\cdot)\|_{\mathrm{L}^\infty(\mathbb{R})}\,\mathrm{d}s \\ &\leq \left\|F''\right\|_{\mathrm{L}^\infty(\mathbb{R})} \int_0^t \|w(s,\cdot)\|_{\mathrm{L}^\infty(\mathbb{R})}\,\mathrm{d}s.\end{aligned}$$

Thus, by Grönwall's Lemma, $w$ is zero. As a conclusion, there exists at most only one solution to (4.15), which satisfies (4.52).

**Synthesis** We define

$$\Psi : u \mapsto \Psi[u](t,x) := K_t * u_0(x) - \left\{\int_0^t K_{t-s} * F'(u(s,\cdot))\mathrm{d}s\right\}(x).$$

We show that, if $T$ is small, then $\Psi$ is a contraction on the Banach space

$$E(T) := \mathrm{L}^\infty([0,T]\times\mathbb{R}).$$

Let $u \in E(T)$. We have $F'(u) \in \mathrm{L}^\infty([0,T]\times\mathbb{R})$, $K \in \mathrm{L}^\infty\left([0,T],\mathrm{L}^1(\mathbb{R})\right)$, and $u_0 \in \mathrm{L}^\infty(\mathbb{R})$. Therefore $\Psi[u] \in E(T)$. Moreover, if $u_1$ and $u_2 \in E(T)$, then, by Taylor expansion

$$\begin{aligned}|\Psi[u_1](t,x) - \Psi[u_2](t,x)| &\leq \left\{\int_0^t K_{t-s} * \left|F'(u_1(s,\cdot)) - F'(u_2(s,\cdot))\right|\mathrm{d}s\right\}(x) \\ &\leq \left\|F''\right\|_{\mathrm{L}^\infty(\mathbb{R})} \left\{\int_0^t K_{t-s} * |u_1(s,\cdot) - u_2(s,\cdot)|\,\mathrm{d}s\right\}(x).\end{aligned}$$

Therefore, by (A.24),

$$\|\Psi[u_1] - \Psi[u_2]\|_{E(T)} \leq T\left\|F''\right\|_{\mathrm{L}^\infty(\mathbb{R})} \|u_1 - u_2\|_{E(T)}.$$

Whence, if we set

$$T := \frac{1}{1 + \|F''\|_{\mathrm{L}^\infty(\mathbb{R})}},$$

then $\Psi$ is a contraction on the Banach space $E(T)$. As a consequence, $\Psi$ has a unique fixed point ; namely, there exists a unique solution $u$ to (4.52), for $t \in [0,T]$. As $T$ only depends on $\|F''\|_{\mathrm{L}^\infty(\mathbb{R})}$, then, one can iterate this argument and create a global solution $u$ to (4.52) for all $t \in \mathbb{R}$. Using Lemma (A.4.2), $u$ thus defined is a weak solution to (4.15). □

### A.4.2 Regularizing effect

Evolution equation (4.15) has a regularizing effect ; indeed, the weak solution to (4.15) becomes instantly a classical solutions :

**Proposition A.4.3.** *Let $F \in \mathrm{C}^3(\mathbb{R}) \cap \mathrm{W}^{3,\infty}(\mathbb{R})$ and $u_0 \in \mathrm{L}^\infty(\mathbb{R})$. Let $u \in \mathrm{L}^\infty_{\mathrm{loc}}\left(\mathbb{R}_+, \mathrm{L}^\infty\left(\mathbb{R}^d\right)\right)$ be the weak solution to* (4.15)*. Then, for all $T_0 > 0$*

$$u \in \mathrm{C}\left((T_0, +\infty), \mathrm{C}^2(\mathbb{R})\right) \cap \mathrm{C}^1\left((T_0, +\infty), \mathrm{C}(\mathbb{R})\right). \tag{A.28}$$

*Moreover, for all $t > 0$, there holds* (4.53) *in the strong sense. Finally $u$ satisfies* (4.54)*.*

The proof of Proposition A.4.3 relies on the following :

**Lemma A.4.4.** *Let $G$ be a function from $\mathbb{R}^3$ to $\mathbb{R}$. Assume that there exists a constant $C_G > 0$ such that for all $u \in \mathbb{R}$ and almost all $t \in \mathbb{R}_+$, $x \in \mathbb{R}$, we have the following estimates :*

$$|G(t,x,0)| + |\partial_3 G(t,x,u)| + \frac{|\partial_2 G(t,x,u)|}{1+|u|} \leq C_G. \tag{A.29}$$

*Next, suppose that $u \in \mathrm{L}^\infty\left([0,T] \times \mathbb{R}\right)$ satisfies*

$$u(t,x) = K_t * u_0(x) + \left\{ \int_0^t K_{t-s} * G(s,\cdot,u(s,\cdot))\mathrm{d}s \right\}(x), \tag{A.30}$$

*where $u_0 \in \mathrm{L}^\infty(\mathbb{R})$. Then we have $u \in \mathrm{C}^0_{\mathrm{loc}}\left((0,T], \mathrm{W}^{1,\infty}(\mathbb{R})\right)$.*

We prove Lemma A.4.4 by Grönwall's Lemma, taking advantage of the fact that $K_t$ is a smooth probability measure for $t > 0$, the first derivative of which scales like $t^{-1}$. More precisely, there exists a constant $C > 0$ such that, for any $t' > 0$,

$$\|K_{t'}\|_{\mathrm{L}^1(\mathbb{R})} + t' \left\| \frac{\mathrm{d}}{\mathrm{d}x} K_{t'} \right\|_{\mathrm{L}^1(\mathbb{R})} + (t')^2 \left\| \frac{\mathrm{d}^2}{(\mathrm{d}x)^2} K_{t'} \right\|_{\mathrm{L}^1(\mathbb{R})} \leq C. \tag{A.31}$$

*Proof.* For $\delta > 0$, we set

$$a_\delta(t) := t \sup_{x\in\mathbb{R}} \sup_{|y|<\delta} \frac{|u(t,x) - u(t,y)|}{\delta},$$

and we show that $a_\delta(t)$ is bounded uniformly in $t \in [0,T]$ and $\delta > 0$.

Applying the Young inequality on (A.30) yields

$$\begin{aligned} a_\delta(t) \leq & t \left\| \frac{\mathrm{d}}{\mathrm{d}x} K_t \right\|_{\mathrm{L}^1(\mathbb{R})} \|u_0\|_{\mathrm{L}^\infty(\mathbb{R})} + t \int_{t/2}^t \left\| \frac{\mathrm{d}}{\mathrm{d}x} K_{t-s} \right\|_{\mathrm{L}^1(\mathbb{R})} \|G(s,\cdot,u(s,\cdot))\|_{\mathrm{L}^\infty(\mathbb{R})} \,\mathrm{d}s \\ & + t \int_0^{t/2} \|K_{t-s}\|_{\mathrm{L}^1(\mathbb{R})} \sup_{x\in\mathbb{R}} \sup_{|y|\leq\delta} \frac{|G(s,x,u(s,x)) - G(s,x+y,u(s,x+y)|}{\delta} \mathrm{d}s. \end{aligned} \tag{A.32}$$

Invoking (A.29), we obtain for almost every $s \in \mathbb{R}$,

$$\|G(s,\cdot,u(s,\cdot)\|_{\mathrm{L}^\infty(\mathbb{R})} \leq 2C_G + C_G \|u(s,\cdot)|_{\mathrm{L}^\infty(\mathbb{R})} \leq C,$$

and

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\sup_{|y|\leq\delta}\frac{|G(s,x,u(s,x))-G(s,x+y,u(s,x+y)|}{\delta}\\
&\leq \|\partial_2 G(s,\cdot,u(s,\cdot))\|_{\mathrm{L}^\infty(\mathbb{R})}+\|\partial_3 G(s,\cdot,u(s,\cdot))\|_{\mathrm{L}^\infty(\mathbb{R})}\sup_{x\in\mathbb{R}}\sup_{|y|\leq\delta}\frac{|u(s,x)-u(s,x+y)|}{\delta}\\
&\leq C+C\frac{a_\delta(s)}{s}.
\end{aligned}$$

Therefore, using (A.31), (A.32) implies

$$a_\delta(t)\leq C+Ct\int_{t/2}^t(t-s)^{-1}\mathrm{d}s+Ct\int_0^{t/2}\left(1+\frac{a_\delta(s)}{s}\right)\mathrm{d}s\leq\quad C(1+t)+C\int_0^t a_\delta(s)\mathrm{d}s.$$

Hence, by Grönwall's Lemma and since $a_\delta(0)=0$ (as $u\in\mathrm{L}^\infty([0,T]\times\mathbb{R})$), $a_\delta$ is bounded on $[0,T]$, uniformly in $\delta$. As a conclusion, by definition of $a_\delta$, for all $T_0\in(0,T)$, we have $u\in\mathrm{L}^\infty_{\mathrm{loc}}\left((0,T],\mathrm{W}^{1,\infty}(\mathbb{R})\right)$.

Now, we prove that $t\mapsto u(t,\cdot)$ is continuous in $\mathrm{W}^{1,\infty}(\mathbb{R})$. We set $t>\Delta t>0$. By differentiating (A.30) and using the previous step, we obtain that

$$\begin{aligned}
\partial_x u(t,x)=&\frac{\mathrm{d}}{\mathrm{d}x}K_t*u_0(x)+\int_0^{t-\Delta t}\left\{\frac{\mathrm{d}}{\mathrm{d}x}K_{t-s}*g(s,\cdot)\right\}(x)\mathrm{d}s\\
&+\int_{t-\Delta t}^t\left\{K_{t-s}\partial_x g(s,\cdot)\right\}(x)\mathrm{d}s
\end{aligned}\tag{A.33}$$

where the function $g:(t,x)\mapsto G(t,x,u(t,x))$ is in $\mathrm{L}^\infty\left([0,T],\mathrm{L}^\infty(\mathbb{R})\right)\cap\mathrm{L}^\infty_{\mathrm{loc}}\left((0,T],\mathrm{W}^{1,\infty}(\mathbb{R})\right)$. Therefore, for any $t'>\Delta t$, the Hölder inequality yields

$$\begin{aligned}
&\left\|\partial_x u(t,\cdot)-\partial_x u(t',\cdot)\right\|_{\mathrm{L}^\infty(\mathbb{R})}\\
&\leq\left\|\frac{\mathrm{d}}{\mathrm{d}x}K_t-\frac{\mathrm{d}}{\mathrm{d}x}K_{t'}\right\|_{\mathrm{L}^1}\|u_0\|_{\mathrm{L}^\infty}+\int_0^{t-\Delta t}\left\|\frac{\mathrm{d}}{\mathrm{d}x}K_{t-s}-\frac{\mathrm{d}}{\mathrm{d}x}K_{t'-s}\right\|_{\mathrm{L}^1(\mathbb{R})}\|g\|_{\mathrm{L}^\infty(\mathbb{R})}\,\mathrm{d}s\\
&\quad+\int_{t-\Delta t}^t\|K_{t-s}\|_{\mathrm{L}^1(\mathbb{R}^d)}\|\partial_x g(s,\cdot)\|_{\mathrm{L}^\infty(\mathbb{R})}\,\mathrm{d}s+\int_{t-\Delta t}^{t'}\|K_{t'-s}\|_{\mathrm{L}^1(\mathbb{R}^d)}\|\partial_x g(s,\cdot)\|_{\mathrm{L}^\infty(\mathbb{R})}\,\mathrm{d}s.
\end{aligned}$$

But since the following convergence uniformly holds for $s\in[0,t-\Delta t]$ :

$$\left\|\frac{\mathrm{d}}{\mathrm{d}x}K_{t-s}-\frac{\mathrm{d}}{\mathrm{d}x}K_{t'-s}\right\|_{\mathrm{L}^1(\mathbb{R})}\underset{t'\to t}{\longrightarrow}0,$$

we obtain from the previous inequality that

$$\limsup_{t'\to t}\left\|\partial_x u(t,\cdot)-\partial_x u(t',\cdot)\right\|_{\mathrm{L}^\infty(\mathbb{R})}\leq 2\Delta t\,\|\partial_x g(s,\cdot)\|_{\mathrm{L}^\infty([t-\Delta t,t+\Delta t],\mathrm{L}^\infty(\mathbb{R}))}.$$

By imposing $\Delta t\to 0$ in the above estimate, we deduce that $u\in\mathrm{C}^0_{\mathrm{loc}}\left((0,T],\mathrm{W}^{1,\infty}(\mathbb{R})\right)$.

It remains to show that $u(t,\cdot)$ is actually in $\mathrm{C}^1(\mathbb{R})$ for any $t>0$. Let $0<\Delta t<t$. The Hölder inequality applied on (A.33) implies

$$\begin{aligned}|\partial_x u(t,x)-\partial_x u(t,y)| \leq & |x-y|\left\|\frac{\mathrm{d}^2}{(\mathrm{d}x)^2}K_t\right\|_{\mathrm{L}^1(\mathbb{R})}\|u_0\|_{\mathrm{L}^\infty(\mathbb{R})}\\ &+|x-y|\int_0^{t/2}\left\|\frac{\mathrm{d}^2}{(\mathrm{d}x)^2}K_{t-s}\right\|_{\mathrm{L}^1(\mathbb{R})}\|g(s,\cdot)\|_{\mathrm{L}^\infty(\mathbb{R})}\,\mathrm{d}s\\ &+|x-y|\int_{t/2}^{t-\Delta t}\left\|\frac{\mathrm{d}}{\mathrm{d}x}K_{t-s}\right\|_{\mathrm{L}^1(\mathbb{R})}\|\partial_x g(s,\cdot)\|_{\mathrm{L}^\infty(\mathbb{R})}\,\mathrm{d}s\\ &+2\int_{t-\Delta t}^{t}\|K_{t-1}\|_{\mathrm{L}^1(\mathbb{R})}\|\partial_x g(s,\cdot)\|_{\mathrm{L}^\infty(\mathbb{R})}\,\mathrm{d}s.\end{aligned}$$

Now, since $u_0\in\mathrm{L}^\infty(\mathbb{R})$, $g\in\mathrm{L}^\infty([0,T]\times\mathbb{R})\cap\mathrm{L}^\infty\left([T_0,T],\mathrm{W}^{1,\infty}(\mathbb{R})\right)$, using (A.31) yields

$$|\partial_x u(t,x)-\partial_x u(t,y)|\leq C|x-y|\left(t^{-2}+t^{-1}+\ln(t/\Delta t)\right)+2C\Delta t.$$

Finally, setting $\Delta t:=|x-y|$, we obtain that there exists a constant independent of $x$ and $y$ such that

$$|\partial_x u(t,x)-\partial_x u(t,y)|\leq C|x-y|\left(1+\ln\left(1+|x-y|^{-1}\right)\right).$$

As a conclusion, $x\mapsto\partial_x u(t,x)$ is a continuous function. This concludes the proof of Lemma A.4.4. □

We prove Proposition A.4.3 by applying Lemma A.4.4 on $u$ and on $\partial_x u$.

*Proof of Proposition A.4.3.* By Lemma A.4.1, $u$ satisfies (4.52). Since $F'\in\mathrm{C}^1(\mathbb{R})$, then, for $G(t,x,v):=-F'(v)$, one applies Lemma A.4.4 and gets that $u\in\mathrm{C}^0_{\mathrm{loc}}\left((0,T],\mathrm{W}^{1,\infty}(\mathbb{R})\right)$. Setting now an arbitrary $T_0\in(0,T)$ and differentiating (4.52) (replacing 0 by $T_0$) yields

$$\partial_x u(t+T_0,x)=\left(K_t*\partial_x u(T_0,\cdot)\right\}(x)+\left\{\int_0^t K_{t-s}*G\left(s,x,\partial_x u(T_0+s,\cdot)\right)\mathrm{d}s\right\}(x),$$

for

$$G(t,x,v):=-F''(u(T_0+t,x))v.$$

Since $F''\in\mathrm{C}^1(\mathbb{R})$ and since $u(T_0+t,\cdot)\in\mathrm{W}^{1,\infty}(\mathbb{R})$ for all $t\in[0,T-T_0]$, then $G$ obviously satisfies (A.29). Therefore, applying ones more Lemma A.4.4 implies that

$$u\in\mathrm{C}\left((T_0,T],\mathrm{W}^{2,\infty}(\mathbb{R})\right)\cap\mathrm{C}\left((T_0,T],\mathrm{C}^2(\mathbb{R})\right). \tag{A.34}$$

Therefore $|\partial_x|\,u\in\mathrm{C}\left((T_0,T),\mathrm{C}(\mathbb{R})\right)$.

If we choose $\phi\in\mathrm{C}^\infty_{\mathrm{c}}\left((T_0,T)\times\mathbb{R}\right)$ and insert it in (4.50), we can exchange $|\partial_x|\,\phi$ and $u$ and get

$$\int_0^T\int_{\mathbb{R}}u(t,x)\partial_t\phi(t,x)\mathrm{d}x\mathrm{d}t=\int_0^T\int_{\mathbb{R}}\left(|\partial_x|\,u(t,x)+F'(u(t,x))\right)\phi(t,x)\mathrm{d}x\mathrm{d}t.$$

Yet, as $|\partial_x|\, u + F'(u) \in \mathrm{C}((T_0,T),\mathrm{C}(\mathbb{R}))$, then the distribution $\partial_t u \in \mathrm{C}((T_0,T),\mathrm{C}(\mathbb{R}))$. Recalling that $u \in \mathrm{C}\left((T_0,T),\mathrm{C}^2(\mathbb{R})\right)$, this implies (A.28). As a consequence, (4.53) holds in the strong sense for $t > T_0$.

Last but not least, we show (4.54) : the difficult point is $t = 0$. We paraphrase [144, Lem. 3.2 p. 15], that does not apply to this particular case ($K_t$ is *not* a mollifier, for it does not have compact support) even though the result still holds. Our aim is to show that, for all $\phi \in \mathrm{C}_\mathrm{c}^\infty(\mathbb{R})$, there holds

$$\int_{\mathbb{R}} K_t * u_0(x)\phi(x)\mathrm{d}x \underset{t\to 0^+}{\longrightarrow} \int_{\mathbb{R}} u_0(x)\phi(x)\mathrm{d}x. \tag{A.35}$$

Since $u_0 \in \mathrm{L}^\infty\left(\mathbb{R}^d\right)$ and $K_t \in \mathrm{L}^1(\mathbb{R})$, Fubini's theorem yields

$$\int_{\mathbb{R}} K_t * u_0(x)\phi(x)\mathrm{d}x = \int_{\mathbb{R}} u_0(x) K_t * \phi(x)\mathrm{d}x.$$

Moreover, as $K_t$ is regular and satisfies $|K_t(x)| \leq t x^{-2}$ then, by dominated convergence theorem,

$$K_t * \phi \underset{t\to 0^+}{\longrightarrow} \phi \quad \text{in} \quad \mathrm{L}^1(\mathbb{R}).$$

This implies (A.35) and concludes the proof of Proposition A.4.3. □

### A.4.3 An asymptotic estimate

**Lemma A.4.5.** *Under the hypotheses of Proposition 4.1.1, there exists $C > 0$ such that, for all $|x| > 1$,* (4.28) *is satisfied.*

*Proof of Lemma A.4.5.* We first derive a useful expression for $\eta''$. Rewriting (4.1) with the Hilbert transform (see (4.23)) implies

$$-\mathcal{H}\{\eta'\}(x) + c\eta'(x) = F'\left(\eta(x)\right). \tag{A.36}$$

We denote

$$G(x) := F'(\eta(x)) \quad \text{and} \quad g(x) := G'(x) = F''(\eta(x))\eta'(x).$$

Recall that $\mathcal{H}^2\{u\} = -u$ if $u \in \mathrm{L}^2(\mathbb{R})$. As a consequence, applying $\mathcal{H}$ on (A.36) and using (A.36) once more yields

$$\begin{aligned}\eta'(x) &= -c\mathcal{H}\{\eta'\}(x) + \mathcal{H}\left\{G\right\}(x)\\ &= -c^2\eta'(x) + cG(x) + \mathcal{H}\left\{G\right\}(x),\end{aligned}$$

whence

$$\eta'(x) = \frac{1}{1+c^2}\left[cG(x) + \mathcal{H}G(x)\right],$$

and, differentiating the above expression,

$$\eta''(x) = \frac{1}{1+c^2}\left[cg(x) + |\partial_x|\, G(x)\right]. \tag{A.37}$$

We now use (A.37) to obtain (4.28). We have the following inequalities

$$|G(x)| \leq C(1+|x|)^{-1}, \quad |g(x)| \leq Cx^{-2}, \quad \text{and} \quad \left|g'(x)\right| \leq C, \tag{A.38}$$

where $C$ does not depend on $x \in \mathbb{R}$. Estimates (A.38) are consequences of (4.42) (using a Taylor expansion and (4.3)), of (4.8), and respectively of (4.27). As $G \in \mathrm{C}^2(\mathbb{R}) \cap \mathrm{L}^\infty\left(\mathbb{R}^d\right)$ (by assumption, $\eta \in \mathrm{C}^2(\mathbb{R})$) and as $G' = g \in \mathrm{L}^1(\mathbb{R})$, the second right-hand term of (A.37) can be rewritten as

$$|\partial_x|\, G(x) = \frac{1}{\pi}\int_0^{+\infty} \frac{g(x-y) - g(x+y)}{y}\mathrm{d}y.$$

We split this integral into three parts

$$\begin{aligned}
|\partial_x|\, G(x) &= \frac{1}{\pi}\int_0^{x^{-2}} \frac{g(x-y) - g(x+y)}{y}\mathrm{d}y + \frac{1}{\pi}\int_{x^{-2}}^{x/2} \frac{g(x-y) - g(x+y)}{y}\mathrm{d}y \\
&\quad + \frac{1}{\pi}\int_{x/2}^{+\infty} \frac{g(x-y) - g(x+y)}{y}\mathrm{d}y \\
&=: \frac{1}{\pi}\left(I_1 + I_2 + I_3\right).
\end{aligned} \tag{A.39}$$

We deal with $I_1$ while using the estimate on $g'$ of (A.38)

$$|I_1| \leq 2x^{-2} \sup_{|z-x|<x^{-2}} |g'(z)| \leq Cx^{-2}, \tag{A.40}$$

and with $I_2$ while using the estimate on $g$ of (A.38)

$$|I_2| \leq \sup_{|z-x|<x/2} |g(z)| \int_{x^{-2}}^{x/2} \frac{1}{y}\mathrm{d}y \leq Cx^{-2}\ln x. \tag{A.41}$$

To bound $I_3$, we first integrate by parts

$$\begin{aligned}
|I_3| \leq &\left|\left[-\frac{G(x-y)}{y}\right]_{x/2}^{+\infty} - \left[\frac{G(x+y)}{y}\right]_{x/2}^{+\infty}\right| \\
&+ \int_{x/2}^{+\infty}\left\{\left|\frac{G(x-y)}{y^2}\right| + \left|\frac{G(x+y)}{y^2}\right|\right\}\mathrm{d}y.
\end{aligned}$$

Thanks to the estimate on $G$ of (A.38), we obtain

$$\begin{aligned}
|I_3| \leq &Cx^{-2} + C\int_{x/2}^{+\infty}\left\{\frac{(1+|x-y|)^{-1}}{y^2} + \frac{(1+|x+y|)^{-1}}{y^2}\right\}\mathrm{d}y \\
\leq &Cx^{-2} + C\int_{x/2}^{+\infty}\frac{(1+|x-y|)^{-1}}{y^2}\mathrm{d}y.
\end{aligned}$$

We split the rightmost integral into two parts

$$\int_{x/2}^{+\infty} \frac{(1+|x-y|)^{-1}}{y^2}\mathrm{d}y \leq \int_{x/2}^{2x} \frac{(1+|x-y|)^{-1}}{x^2}\mathrm{d}y + \int_{2x}^{+\infty} \frac{x^{-1}}{y^2}\mathrm{d}y \leq Cx^{-2}(1+\ln x).$$

Whence

$$|I_3| \leq Cx^{-2}(1+\ln x). \tag{A.42}$$

Recalling (A.39), estimates (A.40), (A.41) and (A.42) yield

$$||\partial_x|\, G(x)| \leq Cx^{-2}(1+\ln x),$$

which, together with (A.38), implies in turn (4.28). □

## A.5 Annexes du Chapitre 6

### A.5.1 Formules

Le but de cette annexe est de constituer un formulaire pour les noyaux $C_i$ et les résolvantes $\mathfrak{R}_i^\alpha$ de l'équation de Peierls-Nabarro Dynamique.

**Fonctions auxiliaires** On définit les fonctions suivantes :

$$W(T) = \int_0^T \frac{J_1(T')}{T'}dT'. \tag{A.43}$$

et

$$\begin{aligned}\mathcal{W}(T) &= TW(T) + J_0(T)\\ &= T\int_0^T \frac{J_1(T')}{T'}dT' + J_0(T)\\ &= T\int_0^T J_0(T')\mathrm{d}T' - TJ_1(T) + J_0(T).\end{aligned}$$

La fonction $\mathcal{W}$ (ou encore $W$ définie par (A.43)) peut s'exprimer à l'aide des fonctions de Struve $H_0$ et $H_1$ (voir notamment [70, Sec. 8.55 p. 942] pour une référence sur les fonctions de Struve). En effet, on observe que

$$\mathcal{W}(T) = \frac{\pi}{2}T^2\left[J_1(T)H_0(T) - J_0(T)H_1(T)\right] + (1+T^2)J_0(T) - TJ_1(T).$$

Cette expression permet d'évaluer en pratique la fonction $\mathcal{W}$ (ou $W$).

Le logiciel MATLAB ne propose pas de package par défaut pour évaluer les fonctions de Struve. Nous avons tout d'abord utilisé l'approximation proposée dans [122], basée sur des fractions rationnelles. Toutefois, cette approximation présente le défaut de n'être précise

qu'à l'ordre $10^{-8}$. Cette erreur, qui semble a priori faible, s'accumule cependant via les intégrales de (8.45). Finalement, cela engendre des erreurs appréciables dans la simulation de (8.2a) sur les temps longs (de l'ordre de $10^{-4}$ pour des temps $T$ de l'ordre de 100). C'est pourquoi nous avons préféré utiliser les fonctions que l'on trouve dans [146], malgré la documentation quasiment inexistante sur lesdites fonctions. Ce choix est fondé sur des tests numériques que nous avons effectués. Par ailleurs, précisons que les fonctions issues de [146] sont les seuls fichiers du code que nous n'avons pas écrits nous-mêmes (ou qui ne soient pas implémentés dans les boîtes à outils de MATLAB).

**Résolvante mode par mode** En sus des résolvantes $\mathfrak{R}_i^\alpha$, il est intéressant de construire, mode de Fourier $|k|$ par mode de Fourier $|k|$, une résolvante

$$R_i^\alpha(t,k) := \mathfrak{R}_i^\alpha(|k|\tau). \tag{A.44}$$

En pratique, c'est la résolvante $R_i^\alpha$ qui est utilisée dans le code. De (A.44) découle l'identité suivante :

$$\mathcal{L}R_i^\alpha(p,k) = \frac{1}{|k|}\mathcal{L}\mathfrak{R}_i^\alpha\left(\frac{p}{|k|}\right) = \frac{\kappa_i^\alpha}{\kappa_i^\alpha p + |k|\mathcal{L}C_i\left(\frac{p}{|k|}\right)}. \tag{A.45}$$

**Le mode I : coin de montée**

On a les formules suivantes :

$$\begin{aligned}
&\kappa_{\mathrm{I}}^\alpha = \gamma + \alpha,\\
&C_{\mathrm{I}}(T) = \gamma^3\frac{J_1(\gamma T)}{\gamma T} + 4T\left(W(\gamma T) - W(T)\right) + (4\gamma - \gamma^3)J_0(\gamma T) - 4J_0(T),\\
&C_{\mathrm{I}}(T) = \gamma^3\frac{J_1(\gamma T)}{\gamma T} + 4\left(\gamma^{-1}\mathcal{W}(\gamma T) - \mathcal{W}(T)\right) - \gamma^{-1}(\gamma^2-2)^2J_0(\gamma T),\\
&\mathcal{L}C_{\mathrm{I}}(p) = -4p^{-2}\sqrt{1+p^2} - \gamma p + p^{-2}\frac{\left(2+p^2\right)^2}{\sqrt{1+\left(\frac{p}{\gamma}\right)^2}}.
\end{aligned}$$

**Le mode II : coin**

On a les formules suivantes :

$$\begin{aligned}
&\kappa_{\mathrm{II}}^\alpha = 1 + \alpha,\\
&C_{\mathrm{II}}(T) = \frac{J_1(T)}{T} + 3J_0(T) - \frac{4}{\gamma}J_0(\gamma T) - 4T\left(W(\gamma T) - W(T)\right),\\
&C_{\mathrm{II}}(T) = \frac{J_1(T)}{T} - 4\left(\gamma^{-1}\mathcal{W}(\gamma T) - \mathcal{W}(T)\right) - J_0(T)\\
&\mathcal{L}C_{\mathrm{II}}(p) = \frac{p^2}{\sqrt{1+p^2}} + 4p^{-2}\left(\sqrt{1+p^2} - \sqrt{1+\left(\frac{p}{\gamma}\right)^2}\right) - p.
\end{aligned}$$

La résolvante est la suivante :

$$\mathcal{L}\mathfrak{R}^{\alpha}_{\mathrm{II}}(p) = \frac{(1+\alpha)p^2\sqrt{1+p^2}}{\alpha p^3\sqrt{1+p^2} + \left(p^2+2\right)^2 - 4\sqrt{1+p^2}\sqrt{1+\gamma^{-2}p^2}}.$$

D'où

$$\mathcal{L}R^{\alpha}_{\mathrm{II}}(p,k) = \frac{(1+\alpha)p^2\sqrt{k^2+p^2}}{\alpha p^3\sqrt{k^2+p^2} + \left(p^2+2k^2\right)^2 - 4k^2\sqrt{k^2+p^2}\sqrt{k^2+\gamma^{-2}p^2}}.$$

**Le mode III : vis**

On a les formules suivantes :

$$\begin{aligned}
&\kappa^{\alpha}_{\mathrm{III}} = 1+\alpha,\\
&C_{\mathrm{III}}(T) = \frac{J_1(T)}{T},\\
&\mathcal{L}C_{\mathrm{III}}(p) = \sqrt{1+p^2} - p.
\end{aligned}$$

La résolvante est la suivante :

$$\mathcal{L}\mathfrak{R}^{\alpha}_{\mathrm{III}}(p) = \frac{1+\alpha}{\alpha p + \sqrt{1+p^2}}.$$

D'où

$$\mathcal{L}R^{\alpha}_{\mathrm{III}}(p,k) = \frac{1+\alpha}{\alpha p + \sqrt{k^2+p^2}}.$$

Il existe des expressions explicites de la résolvante. Pour $i = \mathrm{III}$ et $\alpha = 0$,

$$\mathfrak{R}^{0}_{\mathrm{III}}(T) = J_0(T),$$

laquelle est dominée par $T^{-1/2}$ en $+\infty$.

Pour $i = \mathrm{III}$ et $\alpha = 1$,

$$\mathfrak{R}^{1}_{\mathrm{III}}(T) = 2\frac{J_1(T)}{T},$$

qui est dominée par $T^{-3/2}$ en $+\infty$.

### A.5.2 Croisement de deux dislocations

Une expérience intéressante consiste à lancer deux dislocations en régime stationnaire, à même contrainte imposée, l'une contre l'autre. Selon leur vitesse initiale, il existe deux possibilités : soit elles se traversent, soit elle s'annihilent (voir [133, Chap. 5]). Nous proposons ici une manière d'imposer la donnée initiale correspondant à une telle situation physique.

On se donne la situation physique suivante : un matériau est soumis à un cisaillement $\sigma$ constant, et deux dislocations sont éloignées l'une de l'autre pour les temps passés. Elles sont en mouvement sur le même plan de glissement et se dirigent l'une vers l'autre. Cela revient à se donner

$$\eta_{0,\mathrm{e}} + \eta_{0,1}(t,x) = \phi_1(x - v_1 t) \qquad \eta_{0,\mathrm{e}} + \eta_{0,2}(t,x) = \phi_2(x - v_2 t)$$

deux dislocations progressives pour un même chargement $\sigma$, où $\eta_{0,\mathrm{e}}$ satisfait (6.19). C'est à dire que $(\phi_1, v_1)$ et $(\phi_2, v_2)$ sont solutions de l'équation de Weertman (6.16). Ces deux dislocations sont supposées éloignées l'une de l'autre ; en première approximation, elles n'interagissent donc pas. Il est alors crédible de supposer que, pour un matériau soumis à un chargement constant $\sigma$ pour tous les temps passés, une solution approximative de l'équation de Peierls-Nabarro Dynamique (6.13) est donnée par

$$\eta_0(t,x) := \eta_{0,1}(t,x) + \eta_{0,2}(t,x).$$

Par un calcul similaire à celui qui a mené à (6.22), on déduit que la fonction

$$u = \eta - \eta_{0,1} - \eta_{0,2}$$

satisfait l'équation intégrodifférentielle suivante :

$$\kappa_i^\alpha \partial_t \widehat{u}(t,k) = -k^2 \int_0^t C_i(|k|(t-t'))\widehat{u}(t',k)\mathrm{d}t' + \widehat{f}(t,k), \tag{A.46}$$

où

$$f(t,x) = -\frac{2}{\mu}\Bigg( F'(u(t,x) + \eta_{0,1}(t,x) + \eta_{0,2}(t,x) + \eta_{\mathrm{e}}(t,x)) \\ - F'(\eta_{0,1}(t,x) + \eta_{0,\mathrm{e}}) - F'(\eta_{0,2}(t,x) + \eta_{0,\mathrm{e}}) - \sigma^{\mathrm{a}}(t,x) + 2\sigma \Bigg).$$

## A.6 Annexes du Chapitre 7

### A.6.1 Démonstration de la Proposition 8.5.1

*Démonstration de la Proposition 8.5.1.* Il est immédiat que si le noyau $C$ satisfait (8.19), alors il s'écrit sous la forme (8.20) (voir [78, Chap. I.4 p. 16]) –donc il est bien dégénéré.

La réciproque est une petite extension du résultat classique affirmant que les seules fonctions $f$ satisfaisant

$$f(t+t') = f(t)f(t') \qquad \forall t, t' > 0$$

sont des exponentielles, *id est*, $f(t) = \mathrm{e}^{\lambda t}$.

Comme $C$ est un noyau à la fois convolutif et dégénéré, pour tous $t > t' > 0$ et $\Delta t > 0$, on a

$$C(t-t') = \sum_{j=0}^{d} a_j(t+\Delta t) b_j(t'+\Delta t), \tag{A.47}$$

Quitte à regrouper des termes, on suppose que les fonctions $a_j$ constituent une famille libre de fonctions régulières. Alors, il est possible de prendre des valeurs $t_0, t_1, \cdots, t_d$ telles que la matrice $(A_{jk}) = (a_k(t_j))$ soit inversible. Ainsi, il découle de (A.47) que, pour tous $t, \Delta t > 0$,

$$\begin{aligned}
&\begin{pmatrix} a_0(t_0+\Delta t) & a_1(t_0+\Delta t) & \cdots & a_d(t_0+\Delta t) \\ a_0(t_1+\Delta t) & a_1(t_1+\Delta t) & \cdots & a_d(t_1+\Delta t) \\ \cdots & \cdots & \cdots & \cdots \\ a_0(t_d+\Delta t) & a_1(t_2+\Delta t) & \cdots & a_d(t_d+\Delta t) \end{pmatrix} \cdot \begin{pmatrix} b_0(t+\Delta t) \\ b_1(t+\Delta t) \\ \cdots \\ b_d(t+\Delta t) \end{pmatrix} \\
&= \begin{pmatrix} a_0(t_0) & a_1(t_0) & \cdots & a_d(t_0) \\ a_0(t_1) & a_1(t_1) & \cdots & a_d(t_1) \\ \cdots & \cdots & \cdots & \cdots \\ a_0(t_d) & a_1(t_2) & \cdots & a_d(t_d) \end{pmatrix} \cdot \begin{pmatrix} b_0(t) \\ b_1(t) \\ \cdots \\ b_d(t) \end{pmatrix}
\end{aligned} \tag{A.48}$$

Il existe une matrice $A'$ telle que le développement limité suivant soit justifié, pour $\Delta t$ petit,

$$\begin{pmatrix} a_0(t_0+\Delta t) & a_1(t_0+\Delta t) & \cdots & a_d(t_0+\Delta t) \\ a_0(t_1+\Delta t) & a_1(t_1+\Delta t) & \cdots & a_d(t_1+\Delta t) \\ \cdots & \cdots & \cdots & \cdots \\ a_0(t_d+\Delta t) & a_1(t_2+\Delta t) & \cdots & a_d(t_d+\Delta t) \end{pmatrix} = A + \Delta t A' + O(\Delta t^2).$$

Par conséquent, on déduit de (A.48) que, pour $\Delta t$ petit,

$$\begin{pmatrix} b_0(t+\Delta t) \\ b_1(t+\Delta t) \\ \cdots \\ b_d(t+\Delta t) \end{pmatrix} = \left(I - \Delta t A^{-1} \cdot A'\right) \cdot \begin{pmatrix} b_0(t) \\ b_1(t) \\ \cdots \\ b_d(t) \end{pmatrix} + O(\Delta t^2).$$

D'où, pour tout $t > 0$, en posant $\Delta t := t/n$,

$$\begin{pmatrix} b_0(t) \\ b_1(t) \\ \cdots \\ b_d(t) \end{pmatrix} = \lim_{n \to +\infty} \left(I - \frac{t}{n} A^{-1} \cdot A'\right)^n \cdot \begin{pmatrix} b_0(0) \\ b_1(0) \\ \cdots \\ b_d(0) \end{pmatrix}.$$

Or

$$\lim_{n \to +\infty} \left(I - \frac{t}{n} A^{-1} \cdot A'\right)^n = \exp\left(-t A^{-1} \cdot A'\right).$$

Ainsi les fonctions $b_j$ s'écrivent sous la forme

$$b_j(t) = \sum_{k=1}^{d} P_{jk}(t) \mathrm{e}^{\lambda_{jk} t},$$

où les fonctions $P_{jk}$ sont des polynômes et les $\lambda_{jk}$ des nombres complexes (issus de l'exponentiation de la matrice $-tA^{-1} \cdot A'$). On démontre de façon similaire que les fonctions $a_j$ s'expriment de la même manière. Donc $C$ satisfait bien (8.20). Par voie de conséquence, il satisfait aussi (8.19). □

## A.7 Annexes du Chapitre 8

### A.7.1 Expression explicite de la méthode bloc-par-bloc

$$\Delta t M_{\Delta} t = \begin{pmatrix} 0 & \frac{3}{4} & -\frac{1}{8} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2\Delta t}{3} & \frac{\Delta t}{6} & 0 \\ 0 & 0 & 0 & 0 & \frac{4\Delta t}{3} & \frac{\Delta t}{3} \\ 0 & 0 & 0 & 0 & \frac{3}{4} & -\frac{1}{8} \\ \frac{2\Delta t}{3} C\left(\frac{\Delta t}{2}\right) & \frac{\Delta t}{6} C(0) & 0 & 0 & 0 & 0 \\ 0 & \frac{4\Delta t}{3} C\left(\Delta t\right) & \frac{\Delta t}{3} C(0) & 0 & 0 & 0 \end{pmatrix},$$

$$V_n = \begin{pmatrix} \frac{3}{8} f_{2n} \\ f_{2n} + \frac{\Delta t}{6} z_{2n} \\ f_{2n} + \frac{\Delta t}{3} z_{2n} \\ \frac{3}{8} z_{2n} \\ \frac{\Delta t}{3} \sum_{j=0}^{2n} w_j^{2n} C((2n+1-j)\Delta t) f_j + \frac{\Delta t}{6} C(\Delta t) f_{2n} \\ \frac{\Delta t}{3} \sum_{j=0}^{2n} w_j^{2n+2} C((2n+2-j)) f_j \end{pmatrix}$$

$$\Delta t W[U_n^j; V_n] = \begin{pmatrix} 0 \\ \frac{\Delta t}{6} f_{2n} + \frac{2\Delta t}{3} f_{2n+1/2} + \frac{\Delta t}{6} f_{2n+1} \\ \frac{\Delta t}{3} f_{2n} + \frac{4\Delta t}{3} f_{2n+1} + \frac{\Delta t}{3} f_{2n+2} \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

où $w_i^{2n}$ sont les poids de Simpson, c'est à dire $(1, 4, 2, 4, 2, ..., 2, 4, 1)$.

En pratique, $W_n$ dépend de $U_n$. On utilise un schéma supplémentaire pour l'estimer, en général un schéma de point fixe tel que :

$$U_n^{j+1} = (1 - \Delta t M_{\Delta t})^{-1} \left(V_n + \Delta t W[U_n^j; V_n]\right). \tag{A.49}$$

### A.7.2 Contours d'intégration pour la méthode d'inversion de Laplace

Dans cette annexe technique, nous rappelons les formules, données par [112] décrivant un contour de Talbot (voir [143]), et expliquons comment on les emploie en pratique, et avec quels paramètres. Nous employons les notations de [112] (les symboles $\mu$, $\nu$, $\sigma$ ne sont donc pas des constantes physiques dans cette section). L'article [137] offre une discussion sur des jeux de paramètres possibles.

Les contours de Talbot sont paramétrés de la manière suivante :

$$\gamma : \begin{cases} ]-\pi, \pi[ & \to \Gamma_{\text{Talbot}}, \\ \theta & \mapsto \sigma + \mu\left(\theta \cot(\theta) + \mathrm{i}\nu\theta\right), \end{cases}$$

où $\sigma, \mu$ et $\nu$ sont des paramètres choisis de telle sorte que les singularités et les lignes de coupure soient englobés à gauche du contour de Talbot. En pratique, un choix raisonnable de paramètres pour un contour de Talbot $\Gamma_{\text{Talbot}}$ relatif à un intervalle $I_l$ défini par (8.41) est le choix suivant :

$$\nu = 0.6, \qquad \mu = \mu_0 \left(2B^l \Delta t\right)^{-1}, \qquad \mu_0 = 8,$$

et $\sigma$ est un point de branchement de $\mathfrak{R}$. Nous avons pris en outre

$$B = 5.$$

Enfin, les valeurs de $\sigma$ sont déterminées :

- soit de manière analytique, en ce qui concerne les extrémités des lignes de coupure de $\mathcal{L}\mathfrak{R}$,
- soit de manière numérique par une procédure d'optimisation, en ce qui concerne les pôles de $\mathcal{L}\mathfrak{R}_{\text{II}}$.

Notons que le contour global $\Gamma_l$ est une union disjointe de plusieurs contours de Talbot. En particulier, le nombres de mode $d+1$ est la somme du nombre de *tous* les points du contour global (et non le nombre de point d'*un seul* contour de Talbot ; cette convention diffère de [112]). L'approche de base est que chacun des contours de Talbot qui constituent $\Gamma_l$ entoure un et un seul pôle ou ligne de coupure de la fonction $\mathfrak{R}$ et soit suffisamment loin des autres pôles ou lignes de coupure. Si jamais c'est impossible avec les paramètres ci-dessus, on augmente le paramètre $\nu$ afin de construire un contour plus grand qui contient alors deux et exactement deux pôles ou lignes de coupures, et ainsi de suite jusqu'à ce que le contour $\Gamma_l$ soit correctement disposé (voir notamment [112, Fig. 4]).

# Annexes B

# Approximation locale précisée dans des problèmes multi-échelles avec défauts localisés

Ce chapitre reprend la prépublication [23]. On y expose brièvement les résultats de l'étude d'un problème multi-échelle elliptique de la forme

$$\begin{cases} -\operatorname{div}\left((A_{\rm per}+\widetilde{A})\left(\dfrac{x}{\varepsilon}\right)\cdot\nabla u^\varepsilon(x)\right)=f(x) & \text{dans} \quad \Omega, \\ u^\varepsilon(x)=0 & \text{sur} \quad \partial\Omega, \end{cases}$$

où $A_{\rm per}$ est une matrice elliptique périodique et $\widetilde{A}$ représente un défaut localisé. Ce court document résume les résultats les plus importants du Chapitre 2 dans ce cadre.

Cette étude a été réalisée en collaboration avec Xavier Blanc et Claude Le Bris.

# Approximation locale précisée dans des problèmes multi-échelles avec défauts localisés

Xavier Blanc [1], Marc Josien [2], Claude Le Bris [2]

**Résumé** Nous poursuivons l'étude initiée dans [27] de problèmes multi-échelles avec défauts, dans le cadre de la théorie de l'homogénéisation, spécifiquement ici pour une équation de diffusion avec un coefficient de la forme fonction périodique perturbée par une fonction $L^r(\mathbb{R}^d)$, $1 < r < +\infty$, modélisant un défaut local. Nous esquissons la démonstration du fait que le correcteur, dont l'existence a été prouvée dans [27, 28], permet d'approcher la fonction solution de l'équation originale avec la même précision, essentiellement, que dans le cas purement périodique. Les taux de convergence varient, et sont précisés, en fonction de l'intégrabilité $L^r$ du défaut. Une extension à un cas abstrait "général" est mentionnée. Les résultats annoncés dans cette Note seront précisés dans le document [24] en préparation (voir aussi le Chapitre 2).

**Abstract**

**Local precised approximation in multiscale problems with local defects.**

We proceed here with our systematic study, initiated in [27], of multiscale problems with defects, within the context of homogenization theory. The case under consideration here is that of a diffusion equation with a diffusion coefficient of the form of a periodic function perturbed by an $L^r(\mathbb{R}^d)$ , $1 < r < +\infty$, function modeling a localized defect. We outline the proof of the following approximation result : the corrector function, the existence of which has been established in [27,28], allows to approximate the solution of the original multiscale equation with essentially the same accuracy as in the purely periodic case. The rates of convergence may however vary, and are made precise, depending upon the $L^r$ integrability of the defect. The generalization to an abstract setting is mentioned. Our proof exactly follows the pattern of the original proof of Avellaneda and Lin in [11] in the periodic case, extended in the works of Kenig and collaborators [94], and borrows a lot from it. The details of the results announced in this Note are given in our forthcoming publication [24] (see also Chapter 2).

## Abridged English version

We continue our study [27, 28] of homogenization problems in nonperiodic media. The particular setting considered here is that of a diffusion equation (B.1) with a diffusion coefficient $a$ of the form (B.5), where $\widetilde{a}$ is an $L^r(\mathbb{R}^d)$ , $1 < r < +\infty$, function modeling a localized defect, decaying at infinity in a loose sense, and perturbing the background periodic medium $a_{\rm per}$. We aim at quantitatively estimating at which rate the two-scale expansion (truncated at the first order) provided by homogenization theory approaches the

1. Univ. Paris Diderot, Sorbonne Paris Cité, Laboratoire Jacques-Louis Lions, UMR 7598, UPMC, CNRS, F-75205 Paris, FRANCE

2. Ecole des Ponts and INRIA, 6 & 8, avenue Blaise Pascal, 77455 Marne-La-Vallée Cedex 2, FRANCE

exact solution $u^\varepsilon$ as $\varepsilon$ vanishes. Put differently, we seek the rate at which the remainder term (B.4) goes to zero. In (B.4), the corrector employed is the function $w_p$ solution to the corrector equation (B.7), the existence and uniqueness (up to additive constants) of which has been established in our previous works [27, 28]. Such a corrector is different from the periodic corrector, and although intuitively one could have thought, based on the observation that the presence of $\widetilde{a}$ does not modify the homogenized equation (B.2), that the periodic corrector $w_p^{per}$ would give an equally accurate approximation, it does not. The rates of convergence obtained are made precise in our main result, namely Théorème 2.1 of the French version, and estimates (B.10) through (B.12) in various norms. Interestingly, the rates of convergence may however vary from one case to another, and also in comparison with the periodic case. They depend upon the $L^r$ integrability of the defect. Our proof exactly follows the pattern of the original proof of Avellaneda and Lin in [11] in the periodic case, extended in the works of Kenig and collaborators [94], and borrows a lot from it. Instead of having a bounded (periodic) corrector, as in those proofs, we have here a corrector function that is not necessarily bounded. The crucial ingredient of the proof is then, in fact, the strict sublinearity of the corrector function, which, given our assumptions, can be made precise, see (B.8). This suggests a generalization of our setting, beyond the "periodic +local perturbation" case, which we make precise in Section 3.1 of the French version : essentially, besides usual assumptions, the key point is that the corrector is stricly sublinear at infinity with a prescribed rate of sublinearity and (a property that is actually very much linked to the former) that the potential function associated to this corrector (defined in (B.15)-(B.16)) is also strictly sublinear at infinity in a similar quantitative manner (see (B.19)). In passing, and as is also the case in the proof of the periodic case, we establish estimates for the Green function $G^\varepsilon$ of the original problem (see (B.32), (B.33), (B.34) of the French version) and its convergence to the (possibily corrected) Green function of the homogenized problem (B.2) (see (B.28)). The details of the results announced in this Note are given in our forthcoming publication [24] (see also Chapter 2).

## B.1 Introduction

### B.1.1 Motivation

Dans cette Note, nous poursuivons l'étude initiée dans [27] de problèmes elliptiques multi-échelles, dans le cadre de la théorie de l'homogénéisation. L'équation que nous considérons possède un coefficient qui présente, à l'échelle microscopique, une structure périodique perturbée localement par un « défaut ». On se convainc aisément que le comportement macroscopique d'un tel matériau est dicté par sa seule structure périodique sous-jacente. Si, en revanche, on cherche à obtenir une information plus fine, en terme de taux de convergence, à l'échelle microscopique, pour une norme plus forte ou encore au voisinage du défaut, alors ce défaut ne peut plus être négligé.

Dans [27], il a été montré (en 1D au moins, et formellement en dimension supérieure) dans un cadre hilbertien, i.e pour un défaut dans $L^2(\mathbb{R}^d)$, qu'il est en effet nécessaire de construire un correcteur prenant en compte le défaut pour obtenir une approximation préci-

sée de la solution. L'existence d'un tel correcteur est démontrée, et ce résultat est généralisé dans [28] au cas d'un défaut d'intégrabilité $L^p(\mathbb{R}^d)$, $1 < p < +\infty$, ainsi qu'à d'autres situations "perturbatives". Il est affirmé dans [27, 28] que, formellement, un tel correcteur permet d'assurer l'approximation voulue. L'objet de cette Note est d'énoncer précisément ce résultat, et de donner les grandes lignes de sa preuve : « le correcteur corrige » dans la norme considérée, à un ordre précisé. Les résultats annoncés dans cette Note, et leurs preuves, seront détaillés dans la publication [24] en préparation (voir aussi le Chapitre 2).

### B.1.2 Le cas périodique

Nous nous donnons un champ $a \in L^\infty\left(\mathbb{R}^d\right)$, pris à valeurs scalaires pour simplifier l'exposé, et nous considérons le problème suivant :

$$-\operatorname{div}(a(x/\varepsilon)\nabla u^\varepsilon(x)) = f(x) \quad \text{dans} \quad \Omega, \quad \text{et} \quad u^\varepsilon = 0 \quad \text{sur} \quad \partial\Omega, \tag{B.1}$$

posé sur un domaine borné régulier $\Omega \subset \mathbb{R}^d$. Le champ $a$ est elliptique. Il est bien connu (voir par exemple [85]) que, dans le cas où $a$ est périodique, alors le problème (B.1) s'homogénéise en le problème suivant :

$$-\operatorname{div}\left(A^*.\nabla u^*(x)\right) = f(x) \quad \text{dans} \quad \Omega, \quad \text{et} \quad u^* = 0 \quad \text{sur} \quad \partial\Omega, \tag{B.2}$$

où $A^*$ est une matrice constante. En particulier, on observe la convergence faible $\nabla u^\varepsilon \rightharpoonup \nabla u^*$ dans $L^2(\Omega)$. Pour obtenir de la convergence forte, il faut corriger $u^*$. Pour ce faire, on définit les *correcteurs* $w_j$, $j \in [\![1, d]\!]$, associés à $a$ comme étant les solutions de l'équation suivante :

$$-\operatorname{div}\left(a\left(e_j + \nabla w_j\right)\right) = 0 \quad \text{dans} \quad \mathbb{R}^d, \quad \text{et} \quad |w_j(x)|/(1+|x|) \underset{|x|\to+\infty}{\to} 0, \tag{B.3}$$

où les $e_j$ sont les vecteurs de la base canonique de $\mathbb{R}^d$, et on introduit le reste défini par

$$R^\varepsilon(x) := u^\varepsilon(x) - u^*(x) - \varepsilon \sum_{j=1}^{d} w_j\left(\frac{x}{\varepsilon}\right) \partial_j u^*(x). \tag{B.4}$$

Les correcteurs $w_j$ sont périodiques, et on démontre classiquement que $\|\nabla R^\varepsilon\|_{L^2(\Omega)} \to 0$. Avellaneda et Lin ont aussi démontré dans [11] que, si de plus $a$ est Hölderienne, on peut obtenir des estimations lipschitziennes sur $u^\varepsilon$, et des estimations sur $R^\varepsilon$ quantifiées en $\varepsilon$. Ces travaux ont notamment été approfondis dans [94], où est démontré que, modulo le fait de prendre des correcteurs $w_j$ adaptés au domaine (c'est à dire avec une définition légèrement différente de (B.3)), on obtient l'estimation suivante : $\|\nabla R^\varepsilon\|_{L^\infty(\Omega)} \leq C\varepsilon \ln \varepsilon$ (voir [94, Lem. 3.5]). Au cours de la preuve, les auteurs approximent la fonction de Green $G^\varepsilon$ associée au problème (B.1), ses gradients $\nabla_x G^\varepsilon$ et $\nabla_y G^\varepsilon$, et son gradient croisé $\nabla_x \nabla_y G^\varepsilon$. Pour ce faire, ils emploient la fonction de Green $G^*$ du problème (B.2), convenablement modifiée par les correcteurs $w_j$ (voir [94, Th. 3.6 et Th. 3.11]).

### B.1.3 Le cas périodique perturbé par un défaut local

Dans [27, 28] est considéré le cas d'un champ scalaire

$$a = a_{\rm per} + \widetilde{a}, \tag{B.5}$$

où $a_{\rm per}$ est périodique et $\widetilde{a}$ est une perturbation locale. Plus précisément, supposons que $d \geq 3$ (en dimension 2, des détails techniques supplémentaires sont nécessaires du fait que les fonctions de Green d'opérateurs elliptiques ne sont pas bornées à l'infini) et qu'il existe $\alpha > 0$, $\mu > 0$ et $r \in ]1, +\infty[$ tels que

$$a_{\rm per}\ ,\ \widetilde{a} \in C^{0,\alpha}_{\rm unif}\left(\mathbb{R}^d\right), \quad \mu^{-1} \leq a_{\rm per} \leq \mu, \quad \mu^{-1} \leq \widetilde{a} + a_{\rm per} \leq \mu, \quad \text{et} \quad \widetilde{a} \in L^r\left(\mathbb{R}^d\right). \tag{B.6}$$

Ici, $C^{0,\alpha}_{\rm unif}\left(\mathbb{R}^d\right)$ désigne l'espace des fonctions uniformément Hölderiennes sur $\mathbb{R}^d$, de coefficient $\alpha \in ]0, 1[$. Alors par [28, Th. 4.1], il existe une solution $w_j$ au problème (B.3) qui s'écrit $w_j = w_j^{\rm per} + \widetilde{w}_j$, où $\nabla \widetilde{w}_j \in L^r\left(\mathbb{R}^d\right) \cap L^\infty\left(\mathbb{R}^d\right)$ et où $w_j^{\rm per}$ est une solution périodique de

$$-\operatorname{div}\left(a_{\rm per}\left(e_j + \nabla w_j^{\rm per}\right)\right) = 0 \quad \text{dans} \quad \mathbb{R}^d. \tag{B.7}$$

Par le Théorème de Morrey [59, Th. 4.10 p. 167], cela implique en particulier que, si $r \neq d$, les correcteurs $w_j$ satisfont

$$|w_j(x) - w_j(y)| \leq C\,|x - y|^{1-\nu}, \tag{B.8}$$

pour tous $x, y \in \mathbb{R}^d$, où

$$\nu = \nu_r := \min\left(1, d/r\right) \in ]0, 1]. \tag{B.9}$$

Dans le cas $r = d$, qui est critique, on obtient (B.8) seulement pour tout $\nu < 1$. Dans tous les cas, cette estimation est « seulement contraignante » pour $|x-y|$ grand. En effet, pour $|x-y|$ petit, $w_j \in C^{1,\alpha}_{\rm unif}$ par régularité elliptique (pour la valeur $\alpha$ de (B.6)), et (B.8) est triviale (avec $\nu = 0$). L'inégalité (B.8) traduit une « sous-linéarité renforcée » des correcteurs (qui devient, pour $r < d$, une borne $L^\infty$ sur lesdits correcteurs). Comme nous allons le voir, l'existence d'un correcteur ainsi que l'estimation (B.8) sont des ingrédients essentiels pour contrôler le gradient du reste $\nabla R^\varepsilon$.

## B.2 Résultats

Nous démontrons un résultat qui étend au cas ci-dessus une partie de l'analyse faite dans [11] et [94] (en particulier le Théorème 5 de [11] et les Théorèmes 3.4 et 3.7 de [94]) :

**Théorème B.2.1.** *Soit $d \geq 3$. Supposons qu'il existe $\alpha > 0$, $\mu > 0$ et $r \in ]1, +\infty[$, $r \neq d$ tels que $a_{\rm per}$ et $\widetilde{a}$ satisfont* (B.6)*, et soit $\nu_r$ défini par* (B.9)*. Soient $\Omega$ un domaine régulier*

*et* $\Omega_1 \subset\subset \Omega$. *Considérons* $f \in L^2(\Omega)$ *et* $u^\varepsilon, u^*, R^\varepsilon$ *respectivement définies par* (B.1), (B.2), *et* (B.4). *Alors* $R^\varepsilon \in H^1(\Omega)$ *et*

$$\|R^\varepsilon\|_{L^2(\Omega)} \leq C_1 \varepsilon^{\nu_r} \|f\|_{L^2(\Omega)}\,, \tag{B.10}$$

$$\|\nabla R^\varepsilon\|_{L^2(\Omega_1)^d} \leq C_2 \varepsilon^{\nu_r} \|f\|_{L^2(\Omega)^d}\,. \tag{B.11}$$

*En outre, pour tout* $\beta \in ]0, \alpha]$, *si* $f \in C^{0,\beta}\left(\bar{\Omega}\right)$, *on a* $R^\varepsilon \in W^{1,\infty}(\Omega)$ *et*

$$\|\nabla R^\varepsilon\|_{L^\infty(\Omega_1)^d} \leq C_3 \varepsilon^{\nu_r} \ln\left(2 + \varepsilon^{-1}\right) \|f\|_{C^{0,\beta}(\Omega)^d}\,, \tag{B.12}$$

*où les constantes* $C_1, C_2, C_3$ *sont indépendantes de* $\varepsilon$ *et* $f$.

L'intégrabilité $L^r$ du défaut détermine la qualité de l'approximation. En particulier, l'exposant $r = d$ est critique, ce qui est naturel (voir [28, Section 3]). Il constitue la charnière entre deux régimes. En effet, si $r < d$, les résultats sont les mêmes que dans le cadre de l'homogénéisation périodique, et ce parce que $w_j \in L^\infty\left(\mathbb{R}^d\right)$. En revanche, si $r > d$, alors le correcteur n'est pas borné a priori, d'où un moindre contrôle sur la quantité $R^\varepsilon$ : le défaut devient suffisamment "gros" pour avoir un impact macroscopique (pour l'approximation de $u^\varepsilon$ dans des normes assez fines et/ou à l'ordre $\varepsilon$). Dans le cas $r = d$, il n'est pas prouvé, ni même certain, que les correcteurs $w_j$ sont bornés (voir [28]). Mais on peut se ramener (de façon sous-optimale) aux résultats prouvés pour $r > d$, puisque $L^d \cap L^\infty \subset L^r \cap L^\infty$ pour $d < r < +\infty$. Notons enfin que la présence du correcteur dans (B.10) est superflue, et qu'on peut avoir la même estimation sur $u^\varepsilon - u^*$ (au lieu de $R^\varepsilon$).

## B.3 Remarques et extensions possibles

Nous faisons ici quelques remarques et renvoyons à la publication en préparation [24] et au Chapitre 2 pour plus de précisions.

### B.3.1 Cadre abstrait général

La démonstration du Théorème B.2.1 ne fait usage de l'hypothèse de la structure "périodique + défaut" que pour démontrer l'existence de correcteurs et d'un potentiel (à savoir la fonction $B$ définie en (B.15)-(B.16) ci-dessous) fortement sous-linéaires. Ainsi, les conclusions du Théorème B.2.1 sont en fait valides sous les hypothèses suivantes, plus générales que celles utilisées ici :

1. la matrice $a$ est elliptique, bornée, uniformément Hölderienne ;
2. elle admet un correcteur $w_j$, c'est-à-dire une solution de (B.3) ;
3. ce correcteur est fortement sous-linéaire à l'infini, c'est-à-dire qu'il vérifie (B.8), pour un certain $\nu \in ]0, 1]$ ;
4. il existe un potentiel $B$ associé (i.e une solution antisymétrique de (B.15)-(B.16) ci-dessous), qui est lui aussi fortement sous-linéaire, c'est-à-dire qu'il vérifie (B.19), pour $\nu \in ]0, 1]$.

Ces hypothèses impliquent en particulier que $a(x/\varepsilon)$ H-converge uniformément vers $A^*$, propriété que l'on définit comme suit : pour tout suite $\varepsilon_n \to 0$ et toute suite $(y_n)_{n\in\mathbb{N}}$ de $\mathbb{R}^d$,

$$A\left(\frac{x}{\varepsilon_n} + y_n\right) \quad \text{H-converge vers} \quad A^*. \tag{B.13}$$

Pour la définition de la H-convergence, nous renvoyons à [145, Definition 6.4]. Cette propriété est fondamentale dans la preuve esquissée ci-dessous. Pour les détails de cette généralisation, nous renvoyons encore à [24] et au Chapitre 2.

### B.3.2 Autres remarques

1. La preuve esquissée ici est faite dans le cas où $a$ est scalaire. Toutefois, il est possible de travailler avec un coefficient matriciel. On obtient alors des résultats analogues.
2. La preuve originale de [11] fonctionne pour des systèmes. Par conséquent, dans la mesure où l'existence des correcteurs $w_j$ est aussi prouvée dans [25] pour le cas des systèmes, il semble a priori possible de démontrer un résultat analogue au Théorème B.2.1 dans le cadre d'un système d'équations. Une telle adaptation n'a cependant pas été entreprise. Voir à ce sujet la Remarque 69 ci-dessous. Notons que, dans le cas d'une équation, le principe du maximum et le Théorème de De Giorgi-Nash-Moser [64, Th. 8.24 p. 202] permettent de simplifier certains aspects de la démonstration.
3. De la même manière que dans [94], il est possible d'approximer la fonction de Green $G^\varepsilon$ relative à l'Equation (B.1), ainsi que ses gradients $\nabla_x G^\varepsilon$ et $\nabla_y G^\varepsilon$, et son gradient croisé $\nabla_x \nabla_y G^\varepsilon$.
4. Les estimations (B.11) et (B.12) sont des estimations à l'intérieur du domaine. Toutefois, dans le cas d'une matrice périodique, on obtient des estimations jusqu'au bord (voir [94]). Cela requiert d'introduire des correcteurs adaptés au domaine, lesquels peuvent être construits à partir des correcteurs définis sur tout $\mathbb{R}^d$ (voir [94, Prop. 2.4]). Ces correcteurs sont également bien définis dans le cas présent, ce qui fournit un point de départ pour adapter la preuve de [94].
5. On peut aussi montrer dans le cadre du Théorème B.2.1 que, si $f \in L^p(\Omega)$, pour tout $p \in [2, +\infty[$, alors

$$\|R^\varepsilon\|_{L^p(\Omega)} \leq C\varepsilon^{\nu_r} \|f\|_{L^p(\Omega)} \quad \text{et} \quad \|\nabla R^\varepsilon\|_{L^p(\Omega_1)} \leq C\varepsilon^{\nu_r} \|f\|_{L^p(\Omega)}.$$

   L'estimation sur $R^\varepsilon$ est immédiate vu le schéma de preuve ci-dessous. L'estimation sur $\nabla R^\varepsilon$ découle d'un Lemme de mesure à la Calderón-Zygmund (voir [141, Th. 2.4]).

## B.4 Schéma de preuve

Notre schéma de preuve suit celui des articles [11,94]. L'idée repose sur le fait que, pour $\varepsilon = 0$, l'équation est à coefficients constants, donc vérifie des estimations de régularité elliptique, à la fois de type Schauder (en normes $C^{k,\alpha}$), et de type Sobolev (en normes $W^{k,p}$).

Pour $\varepsilon$ petit, par compacité, on arrive à obtenir des propriétés similaires. Ceci est l'idée fondamentale des preuves de [11], laquelle repose sur le fait que le correcteur $w_j$ est borné, et donne, via l'expression (B.4), une bonne approximation de $u_\varepsilon$. Ici, le correcteur n'est plus borné a priori, mais le fait que $\nabla \widetilde{w}_j \in L^r(\mathbb{R}^d)$ nous permet d'adapter les preuves de [11, 94]. Notons que le cas où le défaut est d'intégrabilité $r < d$ ne nécessite que des modifications très minimes, car dans ce cas on sait que les $w_j$ sont bornés, ce qui est un ingrédient fondamental des démonstrations de [11]. En revanche, si $r > d$, il faut faire quelques modifications ponctuelles et techniques, qui changent notamment le taux d'approximation (d'où la présence de l'exposant $\nu_r$ dans le Théorème B.2.1). Nous donnons ci-dessous les grandes lignes de la démonstration, en indiquant en particulier les points où le fait que $w_j$ n'est pas borné nécessite des adaptations.

### B.4.1 Justification de l'introduction de la quantité $R^\varepsilon$

Le point de départ de cette étude est un calcul de [85, p. 26-27] (voir aussi [27]), qui indique que pour $u^\varepsilon$, $u^*$, $R^\varepsilon$ respectivement définis par (B.1), (B.2), et (B.4), on a

$$-\operatorname{div}\left(a\left(x/\varepsilon\right)\nabla R^\varepsilon(x)\right) = \varepsilon \operatorname{div}\left(a\left(\frac{x}{\varepsilon}\right)\sum_{k=1}^{d} w_k\left(\frac{x}{\varepsilon}\right)\nabla\partial_k u^*(x)\right) - \sum_{i=1}^{d}\sum_{k=1}^{d} M_k^i\left(\frac{x}{\varepsilon}\right)\partial_{ik}u^*(x), \tag{B.14}$$

où

$$M_k^i(x) := A_{ik}^* - a\left(x\right)\left(\delta_{ik} + \partial_i w_k\left(x\right)\right). \tag{B.15}$$

Pour tout $k \in [\![1,d]\!]$, comme $\operatorname{div}(M_k) = 0$, il existe un potentiel $B_k = \left[B_k^{ij}\right]_{1\le i,j\le d}$ (voir [85, p. 26-27]) associé à $M_k$, c'est-à-dire une fonction antisymétrique par rapport aux indices $i$ et $j$ qui satisfait

$$\operatorname{div}\left(B_k\right) = M_k. \tag{B.16}$$

Grâce à (B.14), on déduit que

$$-\operatorname{div}\left(a\left(x/\varepsilon\right)\nabla R^\varepsilon(x)\right) = \operatorname{div}\left(H^\varepsilon(x)\right) \quad \text{dans} \quad \Omega, \tag{B.17}$$

où

$$H_i^\varepsilon(x) = \varepsilon\sum_{k=1}^{d} a\left(\frac{x}{\varepsilon}\right) w_k\left(\frac{x}{\varepsilon}\right)\partial_{ik}u^*(x) - \varepsilon\sum_{j,k=1}^{d} B_k^{ij}\left(\frac{x}{\varepsilon}\right)\partial_{jk}u^*(x). \tag{B.18}$$

Comme $a_{\text{per}}$ est périodique et $\widetilde{a} \in L^r\left(\mathbb{R}^d\right)$, le potentiel $B_k$ se construit, pour chaque $k \in [\![1,d]\!]$, en séparant sa partie périodique $B_{k,\text{per}}$ et sa partie due au défaut $\widetilde{B}_k$, pour laquelle on démontre que $\nabla\widetilde{B}_k \in L^r\left(\mathbb{R}^d\right)^{d\times d}$ (par la théorie de Calderón-Zygmund, voir [114, p. 233]). Par le Théorème de Morrey [59, Th. 4.10 p. 167], cela implique alors que, pour tout $i,j,k \in [\![1,d]\!]$,

$$\left|B_k^{ij}(x) - B_k^{ij}(y)\right| \le C|x-y|^{1-\nu_r}, \qquad \forall x,y \in \mathbb{R}^d, \tag{B.19}$$

pour $\nu_r$ défini par (B.9). Grâce à (B.6), (B.8) et à (B.19), on déduit de (B.18) que, pour tout $p \in [1, +\infty]$, si $f \in L^p(\Omega)$,

$$\|H^\varepsilon\|_{L^p(\Omega)^d} \leq C\varepsilon^{\nu_r} \left\|\nabla^2 u^*\right\|_{L^p(\Omega)^{d\times d}} \leq C\varepsilon^{\nu_r} \|f\|_{L^p(\Omega)} \,. \tag{B.20}$$

Comme on sait par ailleurs par régularité elliptique que $\nabla w_j \in L^\infty\left(\mathbb{R}^d\right)$, si $f \in C^{0,\beta}$ avec $\beta \leq \alpha$, on peut démontrer, toujours à partir de (B.18), (B.19), (B.8), que

$$\|H^\varepsilon\|_{C^{0,\beta}(\Omega)^d} \leq C\varepsilon^{\nu_r-\beta} \|f\|_{C^{0,\beta}(\Omega)} \quad \text{et} \quad \|H^\varepsilon\|_{L^\infty(\Omega)^d} \leq C\varepsilon^{\nu_r} \|f\|_{C^{0,\beta}(\Omega)} \,. \tag{B.21}$$

L'enjeu de la suite de la démonstration consiste à tirer parti de (B.17) et du contrôle sur $H^\varepsilon$, à savoir (B.20) et (B.21), pour borner $R^\varepsilon$.

### B.4.2 Convergence dans $H^1(\Omega_1)$

Nous esquissons dans cette section la preuve de (B.10) et (B.11). Supposons momentanément que $f \in L^p\left(\mathbb{R}^d\right)$ avec $p > d$. La fonction $R^\varepsilon$ satisfait (B.17) dans $\Omega$. Néanmoins, elle n'est pas nulle au bord, ce qui empêche de faire un simple raisonnement variationnel. On la scinde donc en deux parties $R^\varepsilon = R_1^\varepsilon + R_2^\varepsilon$ telles que

$$-\operatorname{div}\left(a\left(x/\varepsilon\right)\nabla R_1^\varepsilon(x)\right) = 0 \quad \text{dans} \quad \Omega \quad \text{et} \quad R_1^\varepsilon = -\varepsilon \sum_{j=1}^d w_j\left(\cdot/\varepsilon\right)\partial_j u^* \quad \text{sur} \quad \partial\Omega, \tag{B.22}$$

$$-\operatorname{div}\left(a\left(x/\varepsilon\right)\nabla R_2^\varepsilon(x)\right) = \operatorname{div}\left(H^\varepsilon(x)\right) \quad \text{dans} \quad \Omega \quad \text{et} \quad R_2^\varepsilon = 0 \quad \text{sur} \quad \partial\Omega. \tag{B.23}$$

Grâce au principe du maximum et à (B.8), on peut estimer $R_1^\varepsilon$

$$\|R_1^\varepsilon\|_{L^\infty(\Omega)} \leq \left\|\varepsilon \sum_{j=1}^d w_j\left(\frac{\cdot}{\varepsilon}\right)\partial_j u^*\right\|_{C^0(\overline{\Omega})} \leq C\varepsilon^{\nu_r} \|u^*\|_{C^1(\overline{\Omega})} \,, \tag{B.24}$$

puis, grâce à une injection de Sobolev de $W^{2,p}(\Omega)$ dans $C^{1,\gamma}(\Omega)$ (pour un certain $\gamma > 0$) et à l'estimation de régularité elliptique classique [64, Lem. 9.17 p. 242], on obtient

$$\|R_1^\varepsilon\|_{L^\infty(\Omega)} \leq C\varepsilon^{\nu_r} \|u^*\|_{W^{2,p}(\Omega)} \leq C\varepsilon^{\nu_r} \|f\|_{L^p(\Omega)} \,. \tag{B.25}$$

On étudie maintenant $R_2^\varepsilon$. Rappelons que, grâce à [72, Th. 1.1], si $G^\varepsilon$ est la fonction de Green relative à l'Equation (B.1) (donc avec conditions de Dirichlet homogènes au bord), alors la fonction $\nabla_y G^\varepsilon(x,\cdot)$ est bornée dans l'espace de Marcinkiewicz $L^{\frac{d}{d-1},\infty}(\Omega)$, uniformément en $x \in \Omega$ et en $\varepsilon > 0$. Ainsi, en réécrivant

$$R_2^\varepsilon(x) = -\int_\Omega \nabla_y G^\varepsilon(x,y) H^\varepsilon(y) \mathrm{d}y,$$

et en utilisant (B.20) (rappelons que $p > d$), on obtient

$$\|R_2^\varepsilon\|_{L^\infty(\Omega)} \leq C \|H^\varepsilon\|_{L^p(\Omega)} \leq C\varepsilon^{\nu_r} \|f\|_{L^p(\Omega)} \,, \tag{B.26}$$

Par conséquent, (B.25), (B.26) et la deuxième inégalité de (B.24) impliquent

$$\|u^\varepsilon - u^*\|_{L^\infty(\Omega)} \leq C\varepsilon^{\nu_r} \|f\|_{L^p(\Omega)}\,. \tag{B.27}$$

Ceci étant vrai pour tout $f \in L^p(\Omega)$, un argument de dualité (voir [94, Th. 3.3] dans le cas $\nu_r = 1$) permet d'estimer $G^\varepsilon - G^*$, où $G^*$ est la fonction de Green de l'Equation (B.2), sur le domaine $\widetilde{\Omega}(x,y) := \Omega \cap B(y, |x-y|/16)$. On obtient ainsi

$$\|G^\varepsilon(x,\cdot) - G^*(x,\cdot)\|_{L^{p'}\left(\widetilde{\Omega}(x,y)\right)} \leq C\varepsilon^{\nu_r}|x-y|^{2-\frac{d}{p}-\nu_r} \qquad \forall x, y \in \Omega,$$

On applique alors la preuve de [94, Lem. 3.2], qui démontre une version au bord du résultat (B.27), que l'on applique à la fonction $G^\varepsilon$

$$\begin{aligned}|G^\varepsilon(x,y) - G^*(x,y)| \leq & C|x-y|^{-d/p'} \|G^\varepsilon(x,.) - G^*(x,.)\|_{L^{p'}\left(\widetilde{\Omega}(x,y)\right)} \\ & + C\varepsilon^{\nu_r}|x-y|^{1-\nu_r} \|\nabla_y G^*(x,.)\|_{L^\infty\left(\widetilde{\Omega}(x,y)\right)} \\ & + C\varepsilon^{\nu_r}|x-y|^{2-\frac{d}{p}-\nu_r} \left\|(\nabla_y)^2 G^*(x,.)\right\|_{L^p\left(\widetilde{\Omega}(x,y)\right)}.\end{aligned}$$

Le résultat de [55, Th. 1] permet de borner $\nabla_y G^*$ et $(\nabla_y)^2 G_*$, point par point. Ainsi, grâce à l'inégalité de Hölder, on obtient

$$|G^\varepsilon(x,y) - G^*(x,y)| \leq C\varepsilon^{\nu_r}|x-y|^{2-d-\nu_r} \qquad \forall x, y \in \Omega. \tag{B.28}$$

Si on suppose maintenant, conformément à l'hypothèse du Théorème B.2.1, que $f \in L^2(\Omega)$, l'inégalité (B.28) implique, via l'inégalité de Young et le fait que $|x|^{2-d-\nu_r} \in L^1_{\text{loc}}(\mathbb{R}^d)$,

$$\|u^\varepsilon - u^*\|_{L^2(\Omega)} \leq C\varepsilon^{\nu_r} \|f\|_{L^2(\Omega)}\,, \tag{B.29}$$

d'où (B.10), grâce à (B.8) et au fait que $u^* \in W^{1,\infty}(\Omega)$. Cette estimation (B.10) de la norme $L^2$ de $R_\varepsilon$ se transmet en l'estimée (B.11) de son gradient en utilisant l'inégalité de Caccioppoli et des arguments similaires à ceux ci-dessus.

### B.4.3 Estimation $L^\infty$ sur le gradient : le cas homogène

Nous adaptons la preuve du résultat [11, Lem. 16] qui implique que, si $-\operatorname{div}(a(\cdot/\varepsilon)\nabla u^\varepsilon) = 0$ dans $B(0,2)$, alors

$$\|\nabla u^\varepsilon\|_{L^\infty(B(0,1))} \leq C\,\|u^\varepsilon\|_{L^\infty(B(0,2))}\,. \tag{B.30}$$

Dans [11], l'ingrédient essentiel de la preuve est le caractère borné du correcteur, propriété impliquée par la périodicité. Cependant, l'uniforme H-convergence et la sous-linéarité du correcteur sont en fait suffisants pour appliquer leur preuve. La démonstration de (B.30) se fait en trois étapes :

1. Initialisation (voir [11, Lem. 14], avec un second membre nul) : on obtient l'estimation

$$\sup_{|x|\leq\theta}\left|u^\varepsilon(x)-u^\varepsilon(0)-\sum_{i=1}^d\left\{x_i+\varepsilon w_i\left(\frac{x}{\varepsilon}\right)\right\}\fint_{B(0,\theta)}\partial_i u^\varepsilon\right|$$
$$\leq\theta^{1+\gamma}\left(\fint_{B(0,2)}|u^\varepsilon|^2\right)^{1/2}, \tag{B.31}$$

   uniforme en $\varepsilon$ suffisamment petit, à une échelle $\theta\in]0,1[$ fixée. Cette étape repose sur l'existence des correcteurs et sur une propriété d'uniforme H-convergence de $a(\cdot/\varepsilon)$ vers $A^*$, c'est-à-dire (B.13).
2. Itération (voir [11, Lem. 15]) : on répète l'étape précédente sur les boules $B(0,\theta^2)$, $B(0,\theta^3)$, etc., jusqu'à l'échelle $\theta^k$ d'ordre $\varepsilon$. Cette étape se déroule dans notre cas comme dans le cas périodique.
3. *Blow-up* (voir [11, Lem. 16]) : au cours de cette étape, on utilise la théorie classique de Schauder pour estimer $\nabla u^\varepsilon(0)$. (Ici, le point 0 ne joue pas de rôle particulier.) Il faut pour cela assurer un contrôle en norme $L^\infty$ sur les correcteurs rescalés $\varepsilon w_j(\cdot/\varepsilon)$. De nouveau, ce contrôle est immédiat dans le cas périodique ou si $r<d$, parce que les correcteurs sont alors bornés ; dans le cas où $r>d$, il se fait via la propriété de sous-linéarité sur les correcteurs $w_i$ présente dans (B.8).

*Remarque* 69. La preuve ci-dessus ne dépend pas du fait qu'on traite un système ou une équation. Ce n'est pas le cas de la section B.4.2, où nous avons utilisé le principe du maximum et des estimations de de Giorgi-Nash. En conséquence, pour la preuve dans le cas d'un système, il serait nécessaire d'inverser l'ordre des arguments : d'abord démontrer (B.30), puis démontrer les résultats de la section B.4.2, en utilisant, en lieu et place des estimations de de Giorgi-Nash et du principe du maximum, l'estimation (B.30).

### B.4.4 Estimation sur les fonctions de Green

L'estimation suivante est démontrée dans [72, Th. 1.1], sous des hypothèses beaucoup plus générales que celles supposées ici :

$$|G^\varepsilon(x,y)|\leq C|x-y|^{2-d}\qquad\forall x\neq y\in\Omega. \tag{B.32}$$

D'autre part, (B.30), après changement d'échelle, implique que

$$\|\nabla G^\varepsilon(\cdot,y)\|_{L^\infty(B_{2R})}\leq CR^{-1}\|G^\varepsilon(\cdot,y)\|_{L^\infty(B_R)},$$

pour toutes boules concentriques $B_R$ et $B_{2R}$ telles que $y\notin B_{2R}$. Ceci permet de démontrer, en utilisant $R=|x-y|/2$,

$$|\nabla_x G^\varepsilon(x,y)|\leq C|x-y|^{1-d}\quad\text{et}\quad|\nabla_y G^\varepsilon(x,y)|\leq C|x-y|^{1-d}\qquad\forall x\neq y\in\Omega_1, \tag{B.33}$$

et, en remarquant que, en dérivant par rapport à $y$ l'équation $-\operatorname{div}_x(a(x)\nabla_x G(x,y)=\delta(x-y)$, on a $-\operatorname{div}_x\left(a(x/\varepsilon)\nabla_y\nabla_x G^\varepsilon(x,y_0)\right)=0$ dans $\Omega\backslash B(y_0,R)$, pour tous $y_0\in\Omega$ et $R>0$. On obtient donc, en utilisant (B.30) et (B.33)

$$|\nabla_x\nabla_y G^\varepsilon(x,y)|\leq C|x-y|^{-d}\qquad\forall x\neq y\in\Omega_1. \tag{B.34}$$

### B.4.5 Estimation $L^\infty$ sur le gradient : le cas non-homogène

On suppose désormais que $f \in C^{0,\beta}(\overline{\Omega})$, avec $\beta \leq \alpha$. Nous appliquons alors la preuve du résultat [94, Lem. 3.5], dont nous esquissons la démonstration. Si $G^\varepsilon$ satisfaisait (B.34) sur tout le domaine, et si $R^\varepsilon$ était nul sur $\partial\Omega$ (ce qui n'est pas le cas en général), alors on déduirait de (B.17) que

$$\nabla R^\varepsilon(x) = -\int_\Omega \nabla_x \nabla_y G^\varepsilon(x,y) H^\varepsilon(y) \mathrm{d}y, \tag{B.35}$$

d'où, en isolant la singularité de $\nabla_x \nabla_y G^\varepsilon$ et en utilisant (B.20) (pour $p = +\infty$) et (B.21),

$$\begin{aligned} \|\nabla R^\varepsilon\|_{L^\infty(\Omega)} \leq & C\left[\varepsilon^2 \|H^\varepsilon\|_{C^{0,\beta}(\Omega)} + \ln\left(2+\varepsilon^{-1}\right) \|H^\varepsilon\|_{L^\infty(\Omega)}\right] \\ \leq & C\varepsilon^{\nu_r} \ln\left(2+\varepsilon^{-1}\right) \|f\|_{C^{0,\beta}(\Omega)}. \end{aligned} \tag{B.36}$$

Pour prendre en compte que (B.34) est une estimée intérieure et que $R^\varepsilon$ est non nul au bord de $\partial\Omega$, on procède à une localisation en multipliant $R^\varepsilon$ par une fonction de cut-off, et on démontre ainsi (B.12).

**Résumé** Le travail de cette thèse a porté sur l'étude mathématique et numérique de quelques modèles multi-échelles issus de la physique des matériaux.

La première partie de ce travail est consacrée à l'homogénéisation mathématique d'un problème elliptique avec une petite échelle. Nous étudions le cas particulier d'un matériau présentant une structure périodique avec un défaut. En adaptant la théorie classique d'Avellaneda et Lin pour les milieux périodiques, on démontre qu'on peut approximer finement la solution d'un tel problème, notamment à l'échelle microscopique. Nous obtenons des taux de convergence dépendant de l'étalement du défaut. On démontre aussi quelques propriétés des fonctions de Green d'un problème elliptique périodique avec conditions de bord périodiques.

Les dislocations sont des lignes de défaut de la matière responsables du phénomène de plasticité. Les deuxième et troisième parties de ce mémoire portent sur la simulation de dislocations, d'abord en régime stationnaire puis en régime dynamique. Nous utilisons le modèle de Peierls, qui couple échelle atomique et échelle mésoscopique. Dans le cadre stationnaire, on obtient une équation intégrodifférentielle non-linéaire avec un laplacien fractionnaire : l'équation de Weertman. Nous en étudions les propriétés mathématiques et proposons un schéma numérique pour en approximer la solution. Dans le cadre dynamique, on obtient une équation intégrodifférentielle à la fois en temps et en espace. Nous en faisons une brève étude mathématique, et comparons différents algorithmes pour la simuler.

Enfin, dans la quatrième partie, nous étudions la limite macroscopique d'une chaîne d'atomes soumis à la loi de Newton. Des arguments formels suggèrent que celle-ci devrait être décrite par une équation des ondes non-linéaires. Or, nous démontrons –sous certaines hypothèses– qu'il n'en est rien lorsque des chocs apparaissent.

**Abstract** In this thesis we study mathematically and numerically some multi-scale models from materials science.

First, we investigate an homogenization problem for an oscillating elliptic equation. The material under consideration is described by a periodic structure with a defect at the microscopic scale. By adapting Avellaneda and Lin's theory for periodic structures, we prove that the solution of the oscillating equation can be approximated at a fine scale. The rates of convergence depend upon the integrability of the defect. We also study some properties of the Green function of periodic materials with periodic boundary conditions.

Dislocations are lines of defects inside materials, which induce plasticity. The second part and the third part of this manuscript are concerned with simulation of dislocations, first in the stationnary regime then in the dynamical regime. We use the Peierls model, which couples atomistic and mesoscopic scales and involves integrodifferential equations. In the stationary regime, dislocations are described by the so-called Weertman equation, which is nonlinear and involves a fractional Laplacian. We study some mathematical properties of this equation and propose a numerical scheme for approximating its solution. In the dynamical regime, dislocations are described by an equation which is integrodifferential in time and space. We compare some numerical methods for recovering its solution.

In the last chapter, we investigate the macroscopic limit of a simple chain of atoms governed by the Newton equation. Surprisingly enough, under technical assumptions, we show that it is not described by a nonlinear wave equation when shocks occur.