Effective models and numerical homogenization methods for long time wave propagation in heterogeneous media

THÈSE N^{O} 7881 (2017)

PRÉSENTÉE LE 27 OCTOBRE 2017
À LA FACULTÉ DES SCIENCES DE BASE
CHAIRE D'ANALYSE NUMÉRIQUE ET MATHÉMATIQUES COMPUTATIONNELLES
PROGRAMME DOCTORAL EN MATHÉMATIQUES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Timothée Noé POUCHON

acceptée sur proposition du jury:

Prof. F. Nobile, président du jury
Prof. A. Abdulle, directeur de thèse
Prof. P. Joly, rapporteur
Prof. B. Schweizer, rapporteur
Prof. A. Buffa, rapporteuse

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Suisse
2017

À Nany et Papé.

Acknowledgements

Me voici arrivé à la fin de mon doctorat. Quatre années de recherche en mathématiques. Quatre années de recherche et, finalement, ce document. Ce document, mais forcément bien plus que ça. Au-delà des résultats mathématiques, qui représentent l'aboutissement du travail de recherche, il y a le travail de recherche en lui-même. Ce travail ne s'écrit pas, ou, en tout cas, il ne se lit pas. Ce travail est une image mentale, qui prend forme lentement, pièce par pièce, jusqu'à former un ensemble cohérent. Bien sûr, ce processus est influencé par une multitude de facteurs. Parmi les principaux figurent l'état d'esprit et les discussions avec autrui. Si ces deux points ont été si positifs pour moi, c'est grâce aux personnes qui m'entourent. Pour cette raison, je souhaite remercier ces personnes au travers de ces quelques lignes.

Premièrement, je souhaite remercier mon directeur de thèse, Prof. Assyr Abdulle. Je vous remercie d'abord de m'avoir donné l'opportunité d'effectuer ce doctorat. Ensuite et surtout, je vous remercie d'avoir partagé avec moi votre immense expérience et votre passion pour les maths et la recherche. Finalement, merci de m'avoir poussé à comprendre les phénomènes en profondeur, à persévérer, à être exigeant et critique, et de m'avoir enseigné quantité d'autres compétences primordiales en mathématiques.

Je remercie également les autres membres de mon jury de thèse. Pour avoir pris le temps de lire et comprendre mon travail, je remercie Prof. Annalisa Buffa, Prof. Patrick Joly et Prof. Ben Schweizer. Merci également au Prof. Fabio Nobile d'avoir accepté de présider le jury.

J'aimerais ensuite remercier ma famille et particulièrement mes parents. Mes parents, sans qui, il est certain, je n'aurais pas accompli ce travail. Je ne saurais exprimer toutes les raisons pour lesquelles je vous suis reconnaissant. Vous avez toujours été présents et m'avez guidé sur le chemin de la vie. Et puis vous m'avez poussé à trouver quelque chose qui me plaise. Si je termine ma thèse aujourd'hui, c'est en premier lieu grâce à vous. C'est aussi grâce à vous, Nany et Papé, qui avez mis votre grain de sel dans mon éducation. Toujours avec bienveillance. Merci aussi à Julie, Fanny et Manon, pour la solide fratrie que nous avons construite.

Pour ma motivation au travail et mon envie de me lever le matin, mon entourage à l'EPFL a été d'une importance capitale. J'aimerais donc remercier les membres de la chaire ANMC. En premier, mes collègues de bureau. Merci Ondrej, pour nos conversations, pour tes conseils et pour ton expertise en programmation. Merci Giacomo, pour ta bonne humeur, ta musique et ton intérêt de tout. Ensuite mon collègue de repas. Merci Simon, pour ton humour, ta régularité sur 10 km et les panpizzas. Finalement, je remercie (par ordre d'apparition) : Yun, Adrian, Martin, Antti, Patrick, Andrea, Giacomino, Edoardo. Sans oublier Virginie, bien sûr, toujours là pour les paniques administratives.

Pour des raisons diverses et variées, j'aimerais aussi remercier Robin, Jean-Luc, Malvin, Mikaël, Mathieu et Alex. De près ou de loin, souvent ou pas, votre amitié m'a beaucoup donné.

Pour terminer, je remercie Orane, qui m'a accompagné durant 94.7% de mon doctorat. Autant dire que tu as été là pour son intégralité. Merci pour ce que tu m'as apporté. Merci pour ce que tu m'apportes. Et merci pour ce que tu m'apporteras encore.

Lausanne, 6 juillet 2017
Timothée Pouchon

Abstract

Modeling wave propagation in highly heterogeneous media is of prime importance in engineering applications of diverse nature such as seismic inversion, medical imaging or the design of composite materials. The numerical approximation of such multiscale physical models is a mathematical challenge. Indeed, to reach an acceptable accuracy, standard numerical methods require the discretization of the whole medium at the microscopic scale, which leads to a prohibitive computational cost. Homogenization theory ensures the existence of a homogenized wave equation, obtained from the original problem by a limiting process. As this equation does not depend on the microscopic scale, it is a good target for numerical methods. Unfortunately, for general media, the homogenized equation may not be unique and no formulas are available for its effective data. Nevertheless, such formulas are known for media described by a locally periodic tensor. In that case, or more generally for problems with scale separation, methods such as the finite element heterogeneous multiscale method (FE-HMM) are proved to efficiently approximate the homogenized solution. For wave propagation in heterogeneous media, however, it is known that at large timescales the homogenized solution fails to describe the dispersive behavior of the original wave. Hence, a new equation that captures this dispersion is needed. In this thesis, we study such effective equations for long time wave propagation in heterogeneous media.
The first result that we present holds in periodic media. Using the technique of asymptotic expansion, we obtain the characterization of a whole family of equations that describes the long time dispersive effects of the oscillating wave. The validity of our derivation is ensured by rigorous a priori error estimates. We also derive a numerical procedure for the computation of the tensors involved in the first order effective equations. This leads to a numerical homogenization method for long time wave propagation in periodic media. The second result that we present generalizes the procedure for deriving effective equations to arbitrary timescales. This generalization is also useful, for example, for the homogenization of the wave equation with high frequency initial data. We also provide a numerical procedure allowing to compute effective tensors of arbitrary order. The third result is the generalization of the family of first order effective equations from periodic to locally periodic media. A rigorous a priori error analysis is also derived in this situation. This constitutes the first analysis of effective models for the long time approximation of the wave equation in locally periodic media. In a second part of the thesis, we derive numerical homogenization methods for the long time approximation of the wave equation in locally periodic media. In one dimension, we analyze a modification of the FE-HMM called the FE-HMM-L. In higher dimensions, we design a spectral homogenization method. For both methods, we prove error estimates valid for large timescales and in arbitrarily large spatial domains. In particular, we show that these numerical homogenization methods converge to effective solutions that approximate the highly oscillatory wave equation over long time.

Key words: homogenization, wave equation, multiscale, long time behavior, dispersive waves, numerical homogenization, finite element, spectral method, heterogeneous multiscale method, a priori error analysis.

Résumé

Modéliser la propagation d'ondes dans des milieux hétérogènes est indispensable dans diverses applications en ingénierie telles que l'inversion sismique, l'imagerie médicale ou encore la manufacture de matériaux composites. L'approximation numérique de tels modèles physiques multi-échelles est un défi mathématique. En effet, pour atteindre une précision satisfaisante, les méthodes numériques standards nécessitent une discrétisation globale à l'échelle microscopique, ce qui entraîne un coût computationnel prohibitif. La théorie de l'homogénéisation garantit l'existence d'une équation homogène, obtenue du problème initial par un procédé de passage à la limite. Comme cette équation ne dépend pas de l'échelle microscopique, elle est une bonne cible pour les méthodes numériques. Cependant, dans le cas général, l'équation homogène peut ne pas être unique et aucune formule n'est disponible pour calculer ses données effectives. Néanmoins, une telle formule est connue lorsque le milieu peut être décrit par un tenseur localement périodique. Dans ce cas, et plus généralement lorsque les échelles sont séparées, il est démontré que des méthodes telles que la méthode d'éléments finis hétérogène multi-échelles (FE-HMM) approximent efficacement la solution homogène. Toutefois, pour la propagation d'ondes dans des milieux hétérogènes, sur des temps longs, la solution homogène échoue à décrire le comportement dispersif de l'onde originale. De ce fait, une nouvelle équation capable de capturer cette dispersion est nécessaire. Dans cette thèse, nous étudions de telles équations effectives pour la propagation d'ondes en milieux hétérogènes sur des temps longs.
Le premier résultat que nous présentons est valable dans des milieux périodiques. En utilisant un développement asymptotique, nous obtenons la caractérisation de toute une famille d'équations effectives qui décrivent les effets dispersifs de l'onde originale sur des temps longs. La validité de notre dérivation est attestée par des estimations rigoureuses de l'erreur. Nous élaborons également une procédure numérique pour calculer les tenseurs impliqués dans les équations effectives de premier ordre. Il en résulte une méthode efficace d'homogénéisation numérique pour la propagation d'ondes à temps longs dans des milieux périodiques. Le deuxième résultat que nous présentons généralise la procédure de dérivation d'équations effectives à des temps arbitrairement longs. Cette généralisation est aussi utile, par exemple, pour l'homogénéisation de l'équation des ondes avec des données initiales à hautes fréquences. Nous fournissons également une procédure numérique pour calculer des tenseurs effectifs d'ordre arbitraire. Le troisième résultat est une généralisation de la famille d'équations effectives de premier ordre, de milieux périodiques à localement périodiques. Une analyse rigoureuse de l'erreur à priori est également prouvée dans ce cas. Ce résultat constitue la première analyse de modèles effectifs pour l'approximation de l'équation des ondes dans des milieux localement périodiques sur des temps longs. Dans une seconde partie de la thèse, nous dérivons des méthodes d'homogénéisation numérique pour l'approximation à temps longs de l'équation des ondes dans des milieux localement périodiques. En une dimension, nous analysons une modification de la méthode FE-HMM appelée FE-HMM-L. En dimension plus élevée, nous élaborons une méthode spectrale d'homogénéisation. Pour les deux méthodes, nous démontrons des estimations d'erreur à priori, valables pour des temps longs et des domaines arbitrairement grands. En particulier, nous prouvons que ces méthodes d'homogénéisation numérique convergent vers des solutions effectives qui approximent l'équation des ondes en milieux localement périodiques sur des temps longs.

Mots clefs : homogénéisation, équation des ondes, multi-échelles, comportement sur des temps

longs, ondes dispersives, homogénéisation numérique, éléments finis, méthode spectrale, méthode hétérogène multi-échelles, analyse d'erreur a priori.

Contents

Notation

Abbreviations

ODE	ordinary differential equation
PDE	partial differential equation
FE / FEM	finite element / finite element method
HMM	heterogeneous multiscale method
DOF	degrees of freedom

Standard sets of numbers

$\mathbb{N}$	set of positive integers $\{0, 1, 2, \ldots\}$
$\mathbb{N}_{>0}$	set of strictly positive integers $\{1, 2, \ldots\}$
$\mathbb{Z}$	set of integers
$\mathbb{R}$	set of real numbers

Differentials

∂_t	partial differential with respect to the time t
∂_{x_i} / ∂_i	partial differential with respect to the i-th space variable x_i
∂^α	for $\alpha \in \mathbb{N}^d$, conventional multiindex notation : $\partial^\alpha = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d}$

Functional spaces

Let $\mathcal{O}$ be an open domain of $\mathbb{R}^d$ and consider functions $\mathcal{O} \to \mathbb{R}$.

$\mathcal{C}^k(\mathcal{O})$	k-times continuously differentiable functions
$\mathcal{D}(\mathcal{O})$	functions of class $\mathcal{C}^\infty$ with a compact support in $\mathcal{O}$
$\mathcal{D}'(\mathcal{O})$	space of distributions, dual space of $\mathcal{D}(\mathcal{O})$
$\mathrm{L}^p(\mathcal{O})$	usual Lebesgue space with $p \in [1, \infty]$
$\mathrm{W}^{k,p}(\mathcal{O})$	usual Sobolev space with $k \in \mathbb{N}$ and $p \in [1, \infty]$
$\mathrm{H}^k(\mathcal{O})$	Sobolev space $W^{k,2}(\mathcal{O})$
$\mathrm{L}^2_{\mathrm{per}}(\mathcal{O})$	$\mathcal{O}$-periodic functions in $\mathrm{L}^2(\mathcal{O})$
$\mathrm{H}^1_{\mathrm{per}}(\mathcal{O})$	closure of $C^\infty_{\mathrm{per}}(\mathcal{O})$ for the H^1 norm
$\mathcal{L}^2(\mathcal{O})$	quotient space $\mathrm{L}^2(\mathcal{O})/\mathbb{R}$
$\mathcal{W}_{\mathrm{per}}(\mathcal{O})$	quotient space $\mathrm{H}^1_{\mathrm{per}}(\mathcal{O})/\mathbb{R}$
$\mathrm{L}^2_0(\mathcal{O})$	functions of $\mathrm{L}^2(\mathcal{O})$ with zero mean
$\mathrm{W}_{\mathrm{per}}(\mathcal{O})$	functions of $\mathrm{H}^1_{\mathrm{per}}(\mathcal{O})$ with zero mean
$\mathcal{P}_k(\mathcal{O})$	space of polynomials of degree $\leq k$

Other notations for functions

X^*	dual space of the vector space X
$\langle \cdot, \cdot \rangle_{X^*,X}$	dual evaluation
$\|\cdot\|_X$	standard norm in a normed space X
$(\cdot, \cdot)_X$	standard inner product in a pre-Hilbert space X
$\langle \cdot \rangle_{\mathcal{O}}$	integral mean on $\mathcal{O}$: for $v \in \mathrm{L}^1(\mathcal{O})$, $\langle v \rangle_{\mathcal{O}} = \lvert\mathcal{O}\rvert^{-1} \int_{\mathcal{O}} v(x)\,\mathrm{d}x$
$(\cdot, \cdot)_{\mathcal{O}}$	shorthand for $(\cdot, \cdot)_{\mathrm{L}^2(\mathcal{O})}$
$\|\cdot\|_{\mathrm{L}^p}, \|\cdot\|_{\mathrm{L}^p(X)}$	shorthand for $\|\cdot\|_{\mathrm{L}^p(\mathcal{O})}, \|\cdot\|_{\mathrm{L}^p(0,T;X)}$
$\lvert\cdot\rvert_{\mathrm{H}^k(\mathcal{O})}$	seminorm in $\mathrm{H}^k(\mathcal{O})$
$\lvert\cdot\rvert_{\mathrm{L}^\infty(0,T;\mathrm{H}^k(\mathcal{O}))}$	$\lvert v\rvert_{\mathrm{L}^\infty(0,T;\mathrm{H}^k(\mathcal{O}))} = \operatorname{ess\,sup}_{t\in(0,T)} \lvert v(t)\rvert_{\mathrm{H}^k(\mathcal{O})}$

Tensors

The notation of the sum symbol for the dot product between two tensors is dropped. By convention, repeated indices are summed.

e_i	i-th canonical basis vector in $\mathbb{R}^d$
$\mathrm{Ten}^n(\mathbb{R}^d)$	vector space of tensors of order n
$\mathrm{Sym}^n(\mathbb{R}^d)$	vector subspace of symmetric tensors of order n
$b \otimes c$	standard tensor product
$b\partial^n$	shorthand for $b_{i_1\cdots i_n}\partial^n_{i_1\cdots i_n}$ (for $b \in \mathrm{Ten}^n(\mathbb{R}^d)$)
$b\eta : \xi$	for $b \in \mathrm{Ten}^3(\mathbb{R}^d)$, $\eta, \xi \in \mathrm{Ten}^2(\mathbb{R}^d)$, $b\eta : \xi = b_{ijkl}\eta_{ij}\xi_{kl}$
$S^n(b)$	symmetrization of the tensor $b \in \mathrm{Ten}^n(\mathbb{R}^d)$: $\big(S^n(b)\big)_{i_1\cdots i_n} = \frac{1}{n!}\sum_{\sigma\in S_n} b_{i_{\sigma(1)}\cdots i_{\sigma(n)}}$
$S^n_{i_1\cdots i_n}\{b_{i_1\cdots i_n}\}$	$(i_1, \cdots, i_d)$-th component of $S^n(b)$: $S^n_{i_1\cdots i_n}\{b_{i_1\cdots i_n}\} = \big(S^n(b)\big)_{i_1\cdots i_n}$
$S^{2,2}(b)$	for $b \in \mathrm{Ten}^4(\mathbb{R}^d)$: $S^{2,2}_{ij,kl}\{b_{ijkl}\} = S^2_{ij}\{S^2_{kl}b_{ijkl}\}$
$\lvert b\rvert_\infty$	infinity norm of $b \in \mathrm{Ten}^n(\mathbb{R}^d)$: $\lvert b\rvert_\infty = \max_{1\le i_1,\ldots,i_n\le d} \lvert b_{i_1\cdots i_n}\rvert$
$I(d,n)$	set of multiindices of the distinct entries of a tensor in $\mathrm{Sym}^n(\mathbb{R}^d)$: $I(d,n) = \big\{i = (i_1,\ldots,i_n) : 1 \le i_1 \le \ldots \le i_n \le d\big\}$

Miscellaneous

d	dimension of the problem: $d \in \{1,2,3\}$
ε	characteristic length of the heterogeneous tensor a^ε
T^ε	long timescale: $T^\varepsilon = \varepsilon^{-2}T$, $T = \mathcal{O}(1)$
$[w]$	equivalence class of $w \in \mathrm{L}^2(\mathcal{O})$ in $\mathcal{L}^2(\mathcal{O})$
$\boldsymbol{w}$	equivalence class in $\mathcal{W}_{\mathrm{per}}(\mathcal{O})$
$\lambda_{\min}(\cdot)$	minimal eigenvalue of a symmetric matrix
$\{\cdot\}_+$	positive part of a real number: $\{\cdot\}_+ = \max\{0, \cdot\}$
$\|\cdot\|_W$	norm on $\mathrm{W}_{\mathrm{per}}(\mathcal{O})$ defined as $\|w\|_W = \inf_{\substack{w=w_1+w_2 \\ w_1,w_2\in \mathrm{W}_{\mathrm{per}}(\mathcal{O})}} \Big\{\|w_1\|_{\mathrm{L}^2(\mathcal{O})} + \|\nabla w_2\|_{\mathrm{L}^2(\mathcal{O})}\Big\}$

1 Introduction

Multiscale models are ubiquitous when considering a physical phenomenon that takes place in a heterogeneous medium. Indeed, while in applications we are often interested in a macroscopic quantity, the microscopic structure of the medium influences the physical process and must be incorporated in the model. Many engineering applications involve wave phenomena in a multiscale framework. Assume, for example, that we want to design a wall to filter the noise generated by the vehicles on a highway. For this task, the aspects that must be studied include the shape and the size of the wall as well as the constituent material, as its microscopic composition affects the macroscopic propagation of the acoustic wave (see e.g., [77]). Another situation where modeling multiscale wave propagation is needed is the simulation of an earthquake (see e.g. [73, 86]). For example, in the planning of a construction, it is crucial to predict the consequences of seismic tremors on the future building. The composition of the ground consists of large and small rocks as well as microscopic fissures. Since we want to predict the displacement of the macroscopic seismic wave, we are again in a multiscale regime. Inverse problems involve multiscale wave propagation as well. For instance, we may be interested in sending a wave in an object whose composition is unknown and reconstructing it from output measurements. In seismic inversion, for example, seismic waves are emitted and an approximate description of the geology of the underground is obtained through measurements of the response (see e.g., [93]). Likewise, in medical imaging, ultrasounds are used to reveal the internal body structure (see e.g., [28, 67]).

As mentioned, a large variety of wave phenomena such as highway noise, earthquake, medical ultrasounds can be modeled by the wave equation. Mathematically, waves are described by the function $u^\varepsilon : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ such that

$$\partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \big(a^\varepsilon(x) \nabla_x u^\varepsilon(t,x)\big) = f(t,x) \quad \text{in } (0,T] \times \mathbb{R}^d, \tag{1.1}$$

with given initial conditions $u^\varepsilon(0,x), \partial_t u^\varepsilon(0,x)$, where $T > 0$ is the final time, f is a source term, and a^ε is the tensor describing the medium. The scalar $\varepsilon > 0$ is the characteristic length of the spatial variation of the tensor a^ε. As we consider (1.1) in a multiscale regime, we assume that the highest wavelength of the initial conditions and f are of order $\mathcal{O}(1)$ and that $\varepsilon \ll 1$. We refer to these two scales as the microscopic (micro) scale $\mathcal{O}(\varepsilon)$ and the macroscopic (macro) scale $\mathcal{O}(1)$. When it comes to approximating numerically u^ε, we first need to truncate the infinite domain $\mathbb{R}^d$. We thus consider a hypercube $\Omega \subset \mathbb{R}^d$ and impose Ω-periodic boundary conditions on $x \mapsto u^\varepsilon(t,x)$. Note that Ω must be sufficiently large for the waves not to reach the boundaries in the time interval $[0,T]$ (we call Ω a pseudoinfinite domain). To be accurate, standard numerical methods such as finite elements (FE) or finite differences (FD) require a grid that resolve the whole domain Ω at the scale $\mathcal{O}(\varepsilon)$. These methods thus require to solve linear systems of size $\mathcal{O}(\varepsilon^{-d})$ at every time iteration. More precisely, if for example 10 points per wavelength are used, the computational cost of each iteration is $\mathcal{O}(10^d \varepsilon^{-d})$. Hence, as $\varepsilon \to 0$, or as the time increases (i.e. Ω increases), the computational cost becomes prohibitive. Therefore, more sophisticated

numerical methods that do not require scale resolution are needed.

The study of multiscale problems such as (1.1) is tied to homogenization theory (see [24, 66, 37, 76] for general theory and [27] for the wave equation). The general homogenization result for (1.1) with $T = \mathcal{O}(1)$ ensures the convergence of u^ε to a function u^0 as $\varepsilon \to 0$ (in the appropriate functional space, see Section 3.2), where $u^0 : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ solves the so-called homogenized equation:

$$\partial_t^2 u^0(t,x) - \nabla_x \cdot \big(a^0(x)\nabla_x u^0(t,x)\big) = f(t,x) \quad \text{in } (0,T] \times \mathbb{R}^d, \tag{1.2}$$

with the same initial conditions as u^ε. The homogenized tensor a^0 in (1.2) is obtained as the so-called G-limit of the sequence $\{a^\varepsilon\}_{\varepsilon>0}$ as $\varepsilon \to 0$. Hence, a^0 does not depend on the microscopic scale $\mathcal{O}(\varepsilon)$ and thus the homogenized solution u^0 is a good target for numerical methods. However, for a general tensor a^ε, no explicit formula is available for a^0, which may not even be unique. Nevertheless, if a^ε is endowed with some specific structure, a^0 is unique and can be computed. Namely, if a^ε is uniformly periodic, i.e.,

$$a^\varepsilon(x) = a\big(\tfrac{x}{\varepsilon}\big) \quad \text{where} \quad y \mapsto a(y) \text{ is } Y\text{-periodic}, \tag{1.3}$$

where Y is a reference cell (typically $Y = (0,1)^d$), then the homogenized tensor is computed as $a^0_{ij} = \big\langle e_i^T a(\nabla_y \chi_j + e_j)\big\rangle_Y$, where χ_j are the Y-periodic functions solving the cell problem

$$-\nabla_y \cdot \big(a(y)\chi_j(y)\big) = \nabla_y \cdot \big(a(y)e_j\big) \quad \text{in } Y. \tag{1.4}$$

An example of a two-dimensional uniformly periodic medium is displayed in Figure 1.1. If we assume that a^ε is locally periodic, i.e.,

$$a^\varepsilon(x) = a\big(x, \tfrac{x}{\varepsilon}\big) \quad \text{where} \quad y \mapsto a(x,y) \text{ is } Y\text{-periodic}, \tag{1.5}$$

then $a^0(x)$ can still be computed for any $x \in \Omega$ with the solutions of (1.4) where $a(y)$ is replaced by $a(x,y)$. In this case the function $\chi_j(x,y)$ thus depends on the slow variable x. An example of a two-dimensional locally periodic medium is displayed in Figure 1.2.

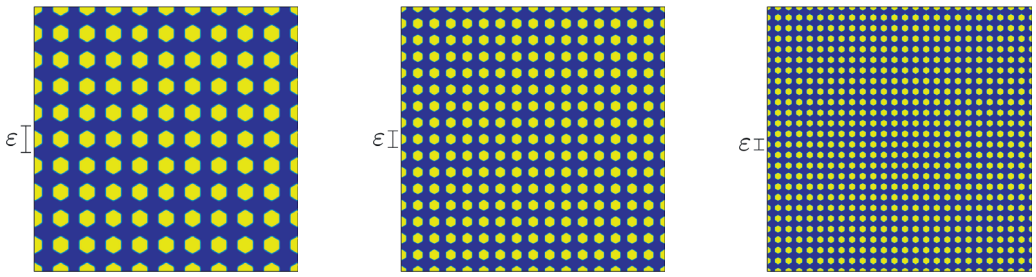

Figure 1.1: Example of a two-dimensional uniformly periodic medium (1.3) displayed in $(0,1)^2$. From left to right: $\varepsilon = 1/10$, $1/16$, and $1/25$.

In the last few years, several multiscale methods have been developed for the approximation of (1.1). All the methods rely on an upscaling procedure that somehow extracts the micro information of the medium and use it at the macro scale. The physical origin of (1.1) motivates the choice of an appropriate method. We can divide the problems in two classes, depending whether a^ε has, or not, scale separation. We refer to scale separation if there are two clearly separated scales. Conversely, we say that the problem does not have scale separation if the medium depends on a continuum of scales. On the one hand, problems with scale separation derive mainly from cases where the medium is artificially designed, which is often the case in material science and can concern geoscience in certain applications. On the other hand, problems

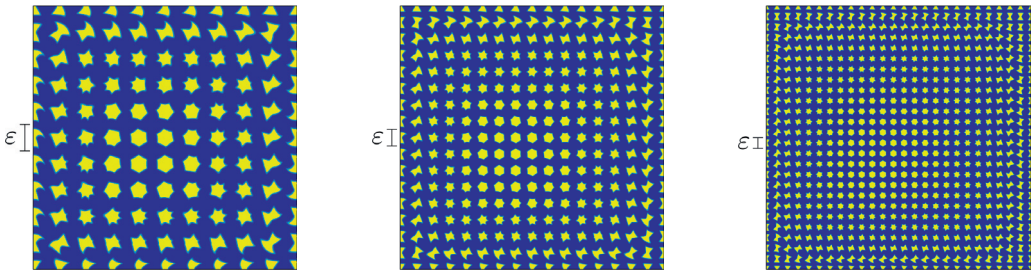

Figure 1.2: Example of a two-dimensional locally periodic medium (1.5) displayed in $(0,1)^2$. From left to right: $\varepsilon = 1/10$, $1/16$, and $1/25$.

without scale separation arise when the medium is natural, as for example the ground or the human body. Logically, the methods dealing with problems without scale separation are more general, and thus more costly. Indeed, a method that takes advantage of the specific structure of the medium is expected to have a smaller cost. Let us shortly review the multiscale methods available for the approximation of (1.1) (it is done in more details in Section 3.1).

Let us begin with the methods that are suited for problems with scale separation. In this case, heterogeneous multiscale methods (HMM) use a sampling strategy to approximate an effective medium at the macro scale (see [7]). As the fine scale has to be resolved only locally in small sampling domains, the cost of a HMM is proportional to the number of degrees of freedom at the macro scale. Furthermore, as it involves small independent problems, the sampling procedure can be efficiently parallelized. The finite difference HMM (FD-HMM), defined in [45] and analyzed in [20], relies on a FD method at the macro scale. The effective flux is approximated by solving micro problems in space-time sampling domains of size $\tau \times \eta^d$, where $\tau, \eta \geq \varepsilon$. The finite element HMM (FE-HMM), defined and analyzed in [8], relies on the FEM on a macro mesh of Ω to approximate the homogenized solution. The homogenized tensor is approximated by solving micro problems in spatial sampling domains of size δ^d ($\delta \geq \varepsilon$) that are localized at the quadrature points of the macro mesh. In the case of a locally periodic tensor (assumption (1.5), Figure 1.2), the FD-HMM and the FE-HMM are proved to converge to the homogenized solution u^0. The sampling strategy of the HMM is in general not conclusive in applications without scale separations. Indeed, some important features of the micro structure can be missed.

Let us then present the methods that are suited for problems without scale separation (see [11] for a detailed review). Owhadi and Zhang presented in [79] a multiscale method based on a harmonic change of coordinates G^ε. Once G^ε is available, the cost of the method is independent of ε. Furthermore, under a so-called Cordes type condition, the approximation is proved to converge to u^ε. The major drawback of this method is the one time overhead involved by the computation of the fine scale fields G^ε. Indeed, this step requires to solve d elliptic PDEs at the fine scale, globally in Ω. In the multiscale finite element method using limited global information presented by Jiang, Efendiev, and Ginting [65, 64], a multiscale method is defined under the assumption that there exists n global fields $\{G_k^\varepsilon\}_{k=1}^n$ such that u^ε can be approximated as $u^\varepsilon(t,x) \approx v(t, G_1^\varepsilon(x), \ldots, G_n^\varepsilon(x))$. Verifying the validity of this assumption is problematic. Furthermore, in the best case, the method still endures the one time overhead involved by the global computations of fine scale fields. In [80], Owhadi and Zhang used a flux-transfer property to localize the computations at the fine scale to portions of the macro mesh. In particular, the space of approximation involves the solutions of elliptic problems in small domains of size $\mathcal{O}(H^{1/2}|\log(H)|)$, where H is the diameter of a macro element. As these problems are independent, they can be solved in parallel, which represents a valuable gain of the execution time. Note that the coarse basis functions used in the definition of the multiscale space

are required to have sufficient smoothness (e.g., B-splines). Finally, Abdulle and Henning defined in [12] a multiscale method in the framework of the localized orthogonal decomposition (LOD) method. In this method, the construction of the space of approximation requires to solve fine scale problems in patches of size $\mathcal{O}(H|\log(H)|)$, where H is the diameter of a macro element. As these problems are independent, this construction can be efficiently parallelized.

In this thesis, we concentrate on problems with scale separation (see (1.3) and (1.5), Figures 1.1 and 1.2) and design numerical methods for the approximation of (1.1) at timescales $T = \mathcal{O}(\varepsilon^{-2})$ and beyond. As mentioned, in the case of scale separation, the FE-HMM and the FD-HMM provide an accurate approximation of the homogenized solution for a cheap cost. However, these methods must only be used for short time approximation. Indeed, at large timescales $T = \mathcal{O}(\varepsilon^{-2})$, u^ε is described at the macro scale by a superposition of several waves moving with different speeds (left plot in Figure 1.3). This phenomenon is known as dispersion. As this dispersion is not captured by the homogenized solution u^0 (right plot in Figure 1.3), the approximation provided by either method is inaccurate. Hence, a new effective solution that describes the macro behavior of u^ε at timescales of order $\mathcal{O}(\varepsilon^{-2})$ is needed for the development of long time homogenization methods. Calling the homogenized solution u^0 a zero-th order effective equation (valid for timescales $\mathcal{O}(\varepsilon^0)$), we look for a higher order effective equation (valid for timescales $\mathcal{O}(\varepsilon^{-2})$). This equation must agree with (1.2) at order $\mathcal{O}(\varepsilon^0)$ and have additional higher order differential operators that describe the dispersion.

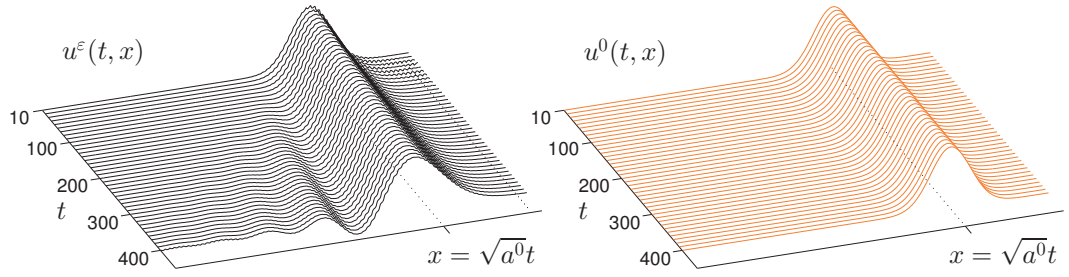

Figure 1.3: Comparison between u^ε and u^0 in a moving frame for a one-dimensional example (replica of Figure 4.1, see Section 4.1 for the data of the problem).

Finding a higher order effective equation for (1.1) in the case of a uniformly periodic tensor (1.3) is an active field of research and the topic of numerous papers (see [85, 52, 51, 72, 42, 43, 13, 18, 14]). Let us mention the main recent results (discussed in more details in Chapter 4). Santosa and Symes formally derived in [85] an equation of the form (for $f = 0$)

$$\partial_t^2 u(t,x) - a^0_{ij}\partial^2_{ij} u(t,x) + \varepsilon^2 c_{ijkl}\partial^4_{ijkl} u(t,x) = 0 \quad \text{in } (0,\varepsilon^{-2}T] \times \mathbb{R}^d. \tag{1.6}$$

However, due to the negative sign of the tensor c, (1.6) is ill-posed. Recently, Lamacz proposed in [72], a well-posed Boussinesq type equation in the one-dimensional case given by (for $f = 0$)

$$\partial_t^2 u(t,x) - a^0\partial^2_x u(t,x) - \varepsilon^2 b\partial^2_x\partial^2_t u(t,x) = 0 \quad \text{in } (0,\varepsilon^{-2}T] \times \mathbb{R}. \tag{1.7}$$

This derivation is supported by an error estimate for $u^\varepsilon - u$. Then, Dohnal, Lamacz and Schweizer defined in [42, 43] a well-posed equation (for $f = 0$)

$$\partial_t^2 u(t,x) - a^0_{ij}\partial^2_{ij} u(t,x) + \varepsilon^2 d_{ijkl}\partial^4_{ijkl} u(t,x) - \varepsilon^2 e_{ij}\partial^2_{ij}\partial^2_t u(t,x) = 0 \quad \text{in } (0,\varepsilon^{-2}T] \times \mathbb{R}^d. \tag{1.8}$$

This derivation is also validated by an error estimate for $u^\varepsilon - u$. Finally, Allaire, Briane, and Vanninathan [18] formally derived an equation of the form (1.8). These different derivations for

the effective equations are done in two different frameworks: [85] and [42, 43] use the expansion of u^ε in Bloch waves, while [72] uses asymptotic expansions (and [18] compares this two approaches, focusing on the elliptic case). The well-posed effective equations in [72] and [42, 43] are obtained after transforming the ill-posed equation (1.6) from [85]. To transform the ill-posed equation into a well-posed one, a Boussinesq trick is performed (as designated in [18]). Formally, this trick consists in using the effective equation at order $\mathcal{O}(1)$ (i.e., the homogenized equation) to replace space derivatives into time derivatives. While it is an easy task in one dimension, the general case in [42, 43] requires a complicated algebraic procedure to decompose the operator $c_{ijkl}\partial^4_{ijkl}$ as

$$c_{ijkl}\partial^4_{ijkl} = d_{ijkl}\partial^4_{ijkl} - (a^0_{ij}\partial^2_{ij})(e_{kl}\partial^2_{kl}), \tag{1.9}$$

which leads to one possible pair of effective tensors d_{ijkl}, e_{ij}.

In a recent paper, Benoit and Gloria [23] proposed an effective equation of arbitrary order for the wave equation. Their derivation is based on the so-called Bloch–Taylor expansion of u^ε, which is a generalization of the Bloch expansion used in [85] and [42, 43]. Their effective equation has the form (for $f = 0$)

$$\partial_t^2 u - a^0\partial^2 u - \sum_{r=1}^{\lfloor \alpha/2 \rfloor} \varepsilon^{2r}\bar{a}^{2r}\partial^{2r+2}u - (\mathrm{i}\varepsilon)^{2(\lfloor \frac{\alpha}{2} \rfloor+1)}\gamma \mathrm{Id}\partial^{2(\lfloor \frac{\alpha}{2} \rfloor+1)+2}u = 0 \quad \text{in } (0,\varepsilon^{-\alpha}T]\times\mathbb{R}^d, \tag{1.10}$$

where $\bar{a}^{2r}$ are effective tensors defined via so-call extended correctors and γ is a regularization parameter. While their analysis holds for more general tensors (almost periodic, quasiperiodic and random), they avoid the crucial question of well-posedness of the equations by introducing a regularization term. As no practical procedure for computing γ is available, this result does not yes translate into a numerical method.

In the literature, apart from the methods studied in this thesis, one numerical methods has been introduced in [46] for the long time approximation of the wave equation. The method is a modification of the FD-HMM from [45]. It is built to capture the effective flux of the ill-posed equation (1.6) from [85]. However, to capture the long time effects, the space-time sampling strategy requires larger sampling domains as $\varepsilon \to 0$. Furthermore, as the target equation (1.6) is ill-posed, a regularization process is needed. Nevertheless, in one dimension and for uniformly periodic tensors, the method is shown in [19] to capture the effective flux of (1.6).

Before presenting the details of the contributions of this thesis, let us summarize our main results and present the organization of the thesis. Chapter 2 is an introduction on the analysis of numerical methods for hyperbolic equations. We prove the well-posedness of the partial differential equations used in the thesis and discuss their numerical approximations. In Chapter 3, we discuss homogenization results for the wave equations (1.1) and introduce the tool of asymptotic expansion, which is central for most results of this thesis. Chapter 4 contains the first main result. In particular, we derive a new family of effective equations for timescales $\mathcal{O}(\varepsilon^{-2})$ and uniformly periodic tensors. Furthermore, we provide an algorithm for the computation of the first order effective tensors. In Chapter 5, we present the second main result generalizing the family of effective equations for timescales $\mathcal{O}(\varepsilon^{-2})$ to arbitrarily large timescales $\mathcal{O}(\varepsilon^{-\alpha})$ (in the case of a uniformly periodic tensor (1.3)). In this case as well, we present an algorithm for the computation of the arbitrary order effective tensors. In Chapter 6, we provide the third main result. In particular, we derive a family of effective equations for timescales $\mathcal{O}(\varepsilon^{-2})$, in the case of a locally periodic tensor (1.5). Based on this result, in Chapter 7, we construct and analyze numerical homogenization methods for the long time approximation of (1.1), in the case of locally periodic tensors (1.5). In particular, two methods are analyzed. The first one, based on the FE-HMM, is designed specifically for the one-dimensional case, and the second one, based on a spectral approximation, is valid in arbitrary dimensions.

We now describe in more details the main contributions of the thesis.

Effective equations for timescales $\mathcal{O}(\varepsilon^{-2})$ in periodic media

In Chapter 4, we present the first main result of this thesis. We derive a new family of effective equations in the case of a periodic tensor (1.3) (Figure 1.1). The derivation is validated by a rigorous a priori error estimate. Furthermore, we provide a numerical procedure for the computation of the effective tensors. This result has been published in [14] ([13] for the one-dimensional case). Note that various additional results are presented in this work.

The family is composed of equations of the form

$$\partial_t^2 \tilde{u}(t,x) - a^0_{ij}\partial^2_{ij}\tilde{u}(t,x) + \varepsilon^2 a^2_{ijkl}\partial^4_{ijkl}\tilde{u}(t,x) - \varepsilon^2 b^2_{ij}\partial^2_{ij}\partial_t^2\tilde{u}(t,x) = f(t,x) \quad \text{in } (0,\varepsilon^{-2}T]\times\Omega, \tag{1.11}$$

where Ω is an arbitrarily large hypercube in $\mathbb{R}^d$. Let us emphasize that whereas in [42, 43], a single effective equation (1.8) is obtained, we define a whole family that is characterized by a constraint on the tensors a^2, b^2. Furthermore, we prove an error estimate that guarantees the error $u^\varepsilon - \tilde{u}$ to be of order $\mathcal{O}(\varepsilon)$ in the norm $\mathrm{L}^\infty(0,\varepsilon^{-2}T;W)$, where the norm

$$\|w\|_W = \inf_{\substack{w=w_1+w_2 \\ w_1,w_2\in \mathrm{W}_{\mathrm{per}}(\Omega)}} \Big\{ \|w_1\|_{\mathrm{L}^2(\Omega)} + \|\nabla w_2\|_{\mathrm{L}^2(\Omega)} \Big\}, \tag{1.12}$$

is equivalent to the $\mathrm{L}^2(\Omega)$ norm through the Poincaré constant. For the analysis to hold, the hypercube Ω and the reference cell Y have to satisfy the assumption

$$\Omega \text{ is the union of cells of volume } \varepsilon|Y|. \tag{1.13}$$

Essentially, this assumption ensures the Ω-periodicity of $v\big(\frac{\cdot}{\varepsilon}\big)$ for any Y-periodic function v. As we track the dependence of the error estimate on Ω, our result is comparable to the ones obtained in [72] and [42, 43], which hold in the whole space $\mathbb{R}^d$.

In our derivation, we use asymptotic expansions, as it was done in one dimension in [72]. We construct an adaptation $\mathcal{B}^\varepsilon\tilde{u}$ of the candidate effective solution $\tilde{u}$ (the form of (1.11) is an ansatz). The adaptation takes the general form

$$\mathcal{B}^\varepsilon\tilde{u}(t,x) = \tilde{u}(t,x) + \sum_{k=1}^{K} \varepsilon^k \chi^k_{i_1\cdots i_k}\big(\tfrac{x}{\varepsilon}\big)\partial^k_{i_1\cdots i_k}\tilde{u}(t,x), \tag{1.14}$$

and is built to solve the same equation as u^ε with an additional remainder. The timescale dictates the accuracy of $\mathcal{B}^\varepsilon\tilde{u}$, i.e., the order of the remainder. In the case $T = \mathcal{O}(\varepsilon^{-2})$, four correctors are sufficient, i.e. $K = 4$. This process leads to the definition of the correctors $\chi^k_{i_1\cdots i_k}$ as the solutions of elliptic PDEs in Y with periodic boundary conditions: the so-called cell problems (χ^1_i solves (1.4)). The well-posedness of the cell problems for $\chi^4_{i_1\cdots i_4}$ characterizes the family of effective equations by providing a constraint on the tensors a^2, b^2. From this constraint, we elaborate a constructive process to obtain pairs of tensors a^2, b^2 of effective equations in the family. Our algorithm involves the solutions of $d + \binom{d+1}{2}$ cell problems, while in [42, 43] $d + \binom{d+1}{2} + \binom{d+2}{3}$ cell problems need to be solved.

Let us mention an important difference in our approach. In [72] and [42, 43], the well-posed effective equations are obtained after transforming the ill-posed equation (1.6). In our derivation, we directly start with an ansatz that enables the effective equation to be well-posed. We are thus naturally led to the understanding that the effective equations are characterized by the constraint obtained on a^2, b^2.

While the effective equations (1.6), (1.7), (1.8), and (1.11) are derived in different frameworks, we verify that the tensors involved in these equations are in fact the same. As a second result, we prove that the well-posed equations (1.7) and (1.8) belongs to our family of effective equations. To see it, note that the constraint characterizing the family reads

$$a^2 - a^0 \otimes b^2 =_S c, \tag{1.15}$$

where c is the tensor in (1.6) (the notation $=_S$ means that the equality must hold up to symmetries). First, in one dimension, we verify that the pair $(a^2, b^2) = (0, b)$, where b is the coefficient in (1.7), satisfies (1.15). Hence, (1.7) belongs to the family. Furthermore, recall that the tensors d, e in (1.8) are built such that (1.9) holds. Hence, the pair $(a^2, b^2) = (d, e)$ satisfies (1.15) and (1.8) thus belongs to the family.

Effective equations for arbitrary large timescales in periodic media

In Chapter 5, we present the second main result of this thesis. The family of effective equations for timescales $\mathcal{O}(\varepsilon^{-2})$, derived in Chapter 4, is generalized to arbitrary timescales $\mathcal{O}(\varepsilon^{-\alpha})$. We thus obtain effective equations of arbitrary order for the wave equation in periodic media.

Effective equations of arbitrary order are not only necessary for large timescales $\mathcal{O}(\varepsilon^{-\alpha})$, but also, for example, when dealing with high-frequency initial data. Indeed, in this situation the dispersive effects appear at shorter time. Furthermore, in some cases, additional effects that are visible are not described by the family of first order effective equations derived in Chapter 4. Hence, higher order effective equations are needed.

The family of effective equations for arbitrary timescales is composed of equations of the form: $\tilde{u} : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ such that

$$\partial_t^2 \tilde{u} - a^0 \partial^2 \tilde{u} - \sum_{r=1}^{\lfloor \alpha/2 \rfloor} (-1)^r \varepsilon^{2r} \left(a^{2r} \partial^{2r+2} \tilde{u} - b^{2r} \partial^{2r} \partial_t^2 \tilde{u} \right) = f \quad \text{in } (0, \varepsilon^{-\alpha} T] \times \Omega, \tag{1.16}$$

where a^0 is the homogenized tensor and $a^{2r} \in \mathrm{Ten}^{2r+2}(\mathbb{R}^d)$, $b^{2r} \in \mathrm{Ten}^{2r}(\mathbb{R}^d)$ are pairs of non-negative tensors that satisfy some symmetry. The domain Ω is an arbitrarily large hypercube satisfying assumption (1.13).

The derivation of the family follows the same process as in Chapter 4. We formulate the ansatz that the effective equations have the form (1.16) and construct an adaptation of $\tilde{u}$ that takes the form (1.14). As the timescale is now of order $\mathcal{O}(\varepsilon^{-\alpha})$, $K = \alpha + 2$ correction terms are needed. After technical developments, we obtain the definition of the cell problems for χ^1 to $\chi^{\alpha+2}$. The well-posedness of these cell problems provides constraints on the pairs $\{a^{2r}, b^{2r}\}$ that characterize the family of effective equations.

While the family is defined implicitly by the constraint, we provide a numerical procedure for the computation of tensors a^{2r}, b^{2r} defining effective equations in the family. As a second result of the chapter, we prove a relation between the correctors that allows to substantially reduce the cost of this computation.

The regularized effective equation (1.10), derived in [23], is connected to (1.16). Indeed, we prove a relation between the tensors $\bar{a}^{2r}$ in (1.10) and the constraints on a^{2r}, b^{2r} characterizing our family. In particular, the extended correctors defined in [23] are the same functions as the solutions of our cell problems $\{\chi^k\}$. Nevertheless, (1.16) has the fundamental advantage of being well-posed without any regularization process and can be used in practice

Effective equations for timescales $\mathcal{O}(\varepsilon^{-2})$ in locally periodic media

In Chapter 6, we present the third main result of this thesis. We generalize the family of effective equations for uniformly periodic tensors, derived in Chapter 4, to locally periodic tensors (see (1.5) and Figure 1.2). This is the first result regarding the long time homogenization of the wave equation in locally periodic media.

The family is constituted of equations of the form

$$\partial_t^2 \tilde{u}(t,x) - \partial_i\big(a_{ij}^0(x)\partial_j \tilde{u}(t,x)\big) + \varepsilon L^1 \tilde{u}(t,x) + \varepsilon^2 L^2 \tilde{u}(t,x) = f(t,x) \quad \text{in } (0,\varepsilon^{-2}T]\times\Omega, \quad (1.17)$$

where $a^0(x)$ is the homogenized tensor and Ω is an arbitrarily large hypercube of $\mathbb{R}^d$ satisfying the assumption (1.13). The correction operators L^1 and L^2 are given by

$$L^1 = -\partial_i\big(a_{ij}^{12}(x)\partial_j\,\cdot\big) + b^{10}\partial_t^2, \quad L^2 = \partial_{ij}^2\big(a_{ijkl}^{24}(x)\partial_{kl}^2\,\cdot\big) - \partial_i\big(b_{ij}^{22}(x)\partial_j\partial_t^2\,\cdot\big) - \partial_i\big(a_{ij}^{22}(x)\partial_j\,\cdot\big) + b^{20}\partial_t^2,$$

where the involved tensors are defined via the solutions of local cell problems (for all $x\in\Omega$). We prove an error estimate ensuring that $u^\varepsilon - \tilde{u}$ is of order $\mathcal{O}(\varepsilon)$ in the norm $\mathrm{L}^\infty(0,\varepsilon^{-2}T;W)$ (see (1.12)). This result is a direct generalization of the estimate obtained in Chapter 4. Indeed, we verify that if the tensor is constant in the slow variable, i.e., $a(x,y)=a(y)$, we recover the family of effective equations (1.11) defined for uniformly periodic tensors (with $a^{24}=a^2$ and $b^{22}=b^2$).

The derivation of the family follows a similar process as in the uniformly periodic case. We start with the ansatz that the effective equation has the form (1.17), where L^1, L^2 are unknown. Indeed, in this case the exact form of the correction operators is not known a priori. We thus build them as we cancel each term in the expansion. Then, the dependence of $a(x,y)$ on the slow variable leads to an adaptation of the form (for a timescale $\mathcal{O}(\varepsilon^{-2})$)

$$\mathcal{B}^\varepsilon \tilde{u}(t,x) = \tilde{u}(t,x) + \sum_{k=1}^{4}\varepsilon^k \sum_{\ell=1}^{k} \chi^{k,\ell}_{i_1\cdots i_{k-\ell+1}}\big(x,\tfrac{x}{\varepsilon}\big)\partial^{k-\ell+1}_{i_1\cdots i_{k-\ell+1}}\tilde{u}(t,x).$$

Imposing that $\mathcal{B}^\varepsilon\tilde{u}$ must solve the same equation as u^ε up to a remainder, we obtain the definitions of the correctors as the solutions of cell problems. The cell problems are defined locally, i.e., for all $x\in\Omega$, $\chi^{k,\ell}_{i_1\cdots i_{k-\ell+1}}(x,\cdot)$ solves an elliptic PDE in Y, with periodic boundary conditions. The well-posedness of these cell problems imposes quantitative constraints on L^1 and L^2. We thus design L^1 and L^2 so that these constraints are satisfied whilst (1.17) is well-posed.

Compared to the effective equations (1.11) in the uniformly periodic case, (1.17) contains the additional operators εL^1 and $\varepsilon^2 L^{2,1} = -\partial_i\big(a_{ij}^{22}(x)\partial_j\,\cdot\big) + b^{20}\partial_t^2$. The presence of εL^1 suggests that the homogenized equation already needs to be corrected for timescales $\mathcal{O}(\varepsilon^{-1})$. However, in all the numerical examples that we considered, the effect of εL^1 is not significant. Furthermore, the only examples where $\varepsilon^2 L^{2,1}$ is important is when the variation of $x\mapsto a(x,y)$ is sharp. These considerations indicate that εL^1 and $\varepsilon^2 L^{2,1}$ could, in certain cases, be removed from the effective equation. This possibility is especially attractive as the corresponding cost of approximation would be significantly lighter. Nevertheless, no practical criterion could be found to support such simplification.

Numerical homogenization methods for long time wave propagation in locally periodic media

Chapter 7 contains the main results for the numerical homogenization of the wave equation in locally periodic media at long times $\mathcal{O}(\varepsilon^{-2})$ (see (1.5) and Figure 1.2). Based on the effective equations derived in Chapter 6 for locally periodic tensors, we define numerical homogenization methods.

In the first part of the chapter, we consider the FE-HMM-L: a modification of the FE-HMM for long time applications that was introduced in [10, 9]. This method is well-suited for the one-dimensional case. Indeed, this case is particular as one equation in the family does not contain a fourth order differential operator. Furthermore, one single corrector is sufficient to compute the effective coefficients. The dispersion effects in the numerical model is built in by an appropriate modification of the mass matrix of an effective FEM solution. The effective solution is obtained by approximating the effective coefficients at each quadrature point of a macroscopic mesh of Ω by solving a micro problem. Two different a priori analyses for the error between u^ε and the approximation of the FE-HMM-L are presented. First, we state an error estimate valid for timescales $\mathcal{O}(\varepsilon^{-2})$ and small domains Ω such that $\mathrm{diam}(\Omega) = \mathcal{O}(1)$, which was published in [13]. Second, we prove an error estimate valid for timescales $\mathcal{O}(\varepsilon^{-2})$ and arbitrarily large domains.

In the second part of the chapter, we define a numerical homogenization method for the multidimensional wave equation in locally periodic media at timescales $\mathcal{O}(\varepsilon^{-2})$. In the multidimensional case, the effective equations are of the form (1.17). In order to handle the fourth order differential operator, we define a spectral homogenization method. The effective tensors are computed locally at the nodes of a macro grid of Ω by approximating the cell problems with the FEM. They are then used in a spectral method defined on the macro grid (or on a subgrid). For this method, we prove an error estimate between u^ε and the approximation, which is valid for timescales $\mathcal{O}(\varepsilon^{-2})$ and arbitrarily large hypercubes Ω. In particular, the method converges to an effective equation of the family derived in Chapter 6.

2 Numerical methods for linear hyperbolic equations

This chapter is a survey of some classical numerical methods used in the approximation of hyperbolic equations. We concentrate on the two partial differential equations that are studied in this thesis: the wave equation and the linear Boussinesq equation. For the first equation, we consider a heterogeneous medium described by a tensor a^ε in an open hypercube $\Omega \subset \mathbb{R}^d$. The number $\varepsilon > 0$ is the characteristic length of variation of the tensor. We consider the equation on a long time interval $[0, T^\varepsilon]$, where $T^\varepsilon = \varepsilon^{-2} T$ and $T = \mathcal{O}(1)$. The wave equation is: $u^\varepsilon : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ such that

$$\partial_t^2 u^\varepsilon(t,x) + \nabla_x \cdot \big(a^\varepsilon(x) \nabla_x u^\varepsilon(t,x)\big) = f(t,x) \quad \text{in } (0, T^\varepsilon] \times \Omega, \tag{2.1}$$

given with periodic boundary conditions and initial conditions for $u^\varepsilon(0), \partial_t u^\varepsilon(0)$. As we will see in Chapters 4 and 6, if the tensor has a periodic or locally periodic structure, the macroscopic long time behavior of u^ε can be described by the solution of a linear Boussinesq equation. Namely, for tensors a^0, a^2, b^2, that will be specified, we consider $\tilde{u} : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ such that

$$\partial_t^2 \tilde{u} + \partial_i \big(a^0_{ij}(x) \partial_j \tilde{u}\big) + \varepsilon^2 \partial^2_{ij} \big(a^2_{ijkl}(x) \partial^2_{kl} \tilde{u}\big) - \varepsilon^2 \partial_i \big(b^2_{ij}(x) \partial_j \partial_t^2 \tilde{u}\big) = f \quad \text{in } (0, T^\varepsilon] \times \Omega, \tag{2.2}$$

with periodic boundary conditions and initial conditions for $\tilde{u}(0), \partial_t \tilde{u}(0)$. In this chapter, we investigate some numerical methods that can be used to approximate the solutions of (2.1) and (2.2). In particular, we introduce three classical numerical methods: the finite element method (FEM), the spectral method and the Fourier method. First, we prove a priori error estimates for the FEM with numerical integration for the approximation of (2.2), in the case $a^2 = 0$. This result uses classical techniques but is not found in the literature. Furthermore, the estimates and the technique will be useful for the analysis of a numerical homogenization method in Chapter 7. The FE approximation of (2.1) is postponed to Chapter 3, where we discuss the multiscale character of the equation. Second, we analyze the spectral method for the approximation of (2.1) and provide some indications for its application. Indeed, this method will be used to approximate (2.1) in a few possible cases (in one dimension and in small two-dimensional domains). Note that in Chapter 7, we will design a numerical homogenization method for the long time approximation of (2.1), that relies on a spectral approximation of an effective model of the form (2.1). Finally, we define and analyze the Fourier method for the approximation of (2.2) in the case of constant tensors a^0, a^2, b^2. This method will be used extensively in Chapters 4 and 5, where the obtained effective models are hyperbolic equations with constant coefficients. Note that in all the error estimates that are derived, the dependence on Ω and the final time T^ε is tracked. Indeed, as these quantities are assumed to be large in the main results of this thesis, their influence needs to be clarified.

The Chapter is organized as follows. In Section 2.1, we prove the well-posedness of (2.1) and (2.2). In Section 2.2, we discuss the FEM for the approximation of (2.2) (with $a^2 = 0$). In particular,

we provide a priori error estimates in the $\mathrm{L}^\infty(\mathrm{H}^1)$ and $\mathrm{L}^\infty(\mathrm{L}^2)$ norms. Next, in Section 2.3, we introduce the spectral method for the approximation of (2.1) and proceed to the corresponding analysis. Finally, in Section 2.4, we present the Fourier method for the approximation of (2.2) in the case of constant coefficients.

2.1 Well-posedness and energy estimates for linear hyperbolic equations

In this section, we prove the well-posedness of the wave equation (2.1) and of the linear Boussinesq equation (2.2). In particular, we define the appropriate functional spaces and prove the existence and uniqueness of a weak solution using the Faedo–Galerkin method.

Let us briefly summarize the Faedo–Galerkin method. Assume that V is an appropriate functional space for the equation. We want to prove the existence and uniqueness of a weak solution $u : [0, T^\varepsilon] \to V$. The first step is the construction of a sequence of approximate solutions $\{u_m(t)\} \subset V^m$, where $V^m \subset V$ is finite-dimensional. As the coefficients of u_m in the basis of V^m solve an ODE, standard theory provides the existence and uniqueness of u_m, for all $m \in \mathbb{N}$. The second step is to prove a bound for the sequence $\{u_m\}$ in V norm (independently of m). Then, functional analysis results provide a weak limit u, which is proved to be the unique weak solution. Note that the techniques used to derive the uniform bound for the sequence—the energy estimates—will be employed repeatedly in this thesis.

To define the finite dimensional subspace V^m, we use an orthonormal basis of $\mathrm{L}^2_0(\Omega)$. As we are in a periodic setting, we select the Fourier basis, introduced in Appendix A.4.1. Indeed, this basis satisfies useful properties. Let $\{w_\ell\}_{\ell\in\mathbb{N}}$ be the Fourier basis of $\mathrm{L}^2(\Omega)$ and define the finite dimensional space $\tilde{V}^m = \operatorname{span}\{w_\ell : 0 \le \ell \le m\}$. An important property of V^m is that the differentiation with respect to x_ν is a linear map $\tilde{V}^m \to \tilde{V}^m$, i.e., $\tilde{V}^m$ is closed under differentiation: $\partial_\nu w_\ell = D_\nu^{\Omega,\ell} w_\ell$, where $D_\nu^{\Omega,\ell} \in \mathrm{i}\mathbb{R}$. We define the orthogonal projection onto V^m as

$$P^m : \mathrm{L}^2(\Omega) \to \tilde{V}^m, \quad v \mapsto P^m v = \sum_{\ell=0}^{m} (v, w_\ell)_{\mathrm{L}^2(\Omega)} w_\ell. \tag{2.3}$$

We verify that the differentiation map permutes with P^m: for $v \in \mathrm{W}_{\mathrm{per}}(\Omega)$,

$$\begin{aligned} \partial_\nu\big(P^m v\big) &= \sum_{\ell=0}^{m} (v, w_\ell)_{\mathrm{L}^2} D_\nu^{\Omega,\ell} w_\ell = \sum_{\ell=0}^{m} (v, \overline{D_\nu^{\Omega,\ell}} w_\ell)_{\mathrm{L}^2} w_\ell = -\sum_{\ell=0}^{m} (v, D_\nu^{\Omega,\ell} w_\ell)_{\mathrm{L}^2} w_\ell \\ &= -\sum_{\ell=0}^{m} (v, \partial_\nu w_\ell)_{\mathrm{L}^2} w_\ell = \sum_{\ell=0}^{m} (\partial_\nu v, w_\ell)_{\mathrm{L}^2} w_\ell = P^m\big(\partial_\nu v\big), \end{aligned} \tag{2.4}$$

where we used integration by parts and the fact that $\overline{D_\nu^{\Omega,\ell}} = -D_\nu^{\Omega,\ell}$. Thanks to Plancherel formula (A.52), we verify that the projection P^m is stable in $\mathrm{L}^2(\Omega)$

$$\|P^m v\|_{\mathrm{L}^2}^2 = \sum_{\ell=0}^{m} |(v, w_\ell)_{\mathrm{L}^2}|^2 \le \sum_{\ell=0}^{\infty} |(v, w_\ell)_{\mathrm{L}^2}|^2 = \|v\|_{\mathrm{L}^2}^2.$$

Furthermore, thanks to (2.4), P^m is stable in $\mathrm{H}^1(\Omega)$

$$|P^m v|_{\mathrm{H}^1}^2 = \sum_{\nu=1}^{d} \|\partial_\nu P^m v\|_{\mathrm{L}^2}^2 = \sum_{\nu=1}^{d} \|P^m(\partial_\nu v)\|_{\mathrm{L}^2}^2 \le \sum_{\nu=1}^{d} \|\partial_\nu v\|_{\mathrm{L}^2}^2 = |v|_{\mathrm{H}^1}^2,$$

and similarly in any $\mathrm{H}^n(\Omega)$:

$$|P^m v|_{\mathrm{H}^n} \le |v|_{\mathrm{H}^n}. \tag{2.5}$$

Note that the mean of a function $v \in \mathrm{L}^2(\Omega)$ is given by $\langle v \rangle_\Omega = (v, w_0)_{\mathrm{L}^2(\Omega)}$ (see Appendix A.4.1). Hence, $\{w_\ell\}_{\ell \in \mathbb{N}_{>0}}$ is an orthonormal basis of $\mathrm{L}^2_0(\Omega)$ and we set $V^m = \mathrm{span}\{w_\ell : 1 \le \ell \le m\}$. We verify that $V^m \subset \mathrm{W}_{\mathrm{per}}(\Omega)$ and the previous properties still holds in this space.

2.1.1 The wave equation in heterogeneous media

In this section, we prove the well-posedness of the wave equation (2.1). The proof follows the Faedo–Galerkin method and can be found in [74, 48].

Let $\Omega \subset \mathbb{R}^d$ be a hypercube and let $a^\varepsilon \in [\mathrm{L}^\infty_{\mathrm{per}}(\Omega)]^{d\times d}$ be a symmetric, uniformly elliptic and bounded tensor, i.e.,

$$\lambda|\xi|^2 \le a^\varepsilon(x)\xi \cdot \xi \le \Lambda|\xi|^2 \quad \forall \xi \in \mathbb{R}^d \ \text{ for a.e. } x \in \Omega. \tag{2.6}$$

We consider the wave equation: find $u^\varepsilon : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ such that

$$\begin{aligned} &\partial_t^2 u^\varepsilon(t,x) + \nabla_x \cdot \big(a^\varepsilon(x)\nabla_x u^\varepsilon(t,x)\big) = f(t,x) && \text{in } (0,T^\varepsilon]\times\Omega,\\ &x \mapsto u^\varepsilon(t,x) \ \ \Omega\text{-periodic} && \text{in } [0,T^\varepsilon],\\ &u^\varepsilon(0,x) = g^0(x), \quad \partial_t u^\varepsilon(0,x) = g^1(x) && \text{in } \Omega, \end{aligned} \tag{2.7}$$

where g^0, g^1 are given initial conditions and f is a source term. In order to prove the well-posedness of (2.7), let us introduce its variational formulation. We define the following bilinear form

$$A^\varepsilon : \mathrm{W}_{\mathrm{per}}(\Omega) \times \mathrm{W}_{\mathrm{per}}(\Omega) \to \mathbb{R} : \quad (v,w) \mapsto A^\varepsilon(v,w) = (a^\varepsilon \nabla v, \nabla w)_{\mathrm{L}^2(\Omega)}. \tag{2.8}$$

Note that the assumptions on a^ε ensures that A^ε is coercive and bounded:

$$A^\varepsilon(v,v) \ge \lambda \|\nabla v\|^2_{\mathrm{L}^2(\Omega)}, \qquad A^\varepsilon(v,w) \ge \Lambda \|\nabla v\|_{\mathrm{L}^2(\Omega)} \|\nabla w\|_{\mathrm{L}^2(\Omega)} \qquad \forall v, w \in \mathrm{W}_{\mathrm{per}}(\Omega).$$

We call a function

$$u^\varepsilon \in \mathrm{L}^\infty(0,T^\varepsilon; \mathrm{W}_{\mathrm{per}}(\Omega)), \quad \partial_t u^\varepsilon \in \mathrm{L}^\infty(0,T^\varepsilon; \mathrm{L}^2_0(\Omega)), \quad \partial_t^2 u^\varepsilon \in \mathrm{L}^2(0,T^\varepsilon; \mathrm{W}^*_{\mathrm{per}}(\Omega)),$$

a weak solution of (2.7) if

$$\begin{aligned} &\big\langle \partial_t^2 u^\varepsilon(t), v\big\rangle + A^\varepsilon\big(u^\varepsilon, v\big) = \big(f(t), w\big)_{\mathrm{L}^2(\Omega)} \quad \forall v \in \mathrm{W}_{\mathrm{per}}(\Omega) \quad \text{for a.e. } t \in [0,T^\varepsilon],\\ &u^\varepsilon(0) = g^0, \ \ \partial_t u^\varepsilon(0) = g^1, \end{aligned} \tag{2.9}$$

where we denoted the dual evaluation $\langle\cdot,\cdot\rangle_{\mathrm{W}^*_{\mathrm{per}}(\Omega),\mathrm{W}_{\mathrm{per}}(\Omega)}$ as $\langle\cdot,\cdot\rangle$. Note that if we assume the regularity (2.10) to hold, (2.9) makes sense thanks to the embeddings

$$\begin{aligned} \big\{v \in \mathrm{L}^2(0,T^\varepsilon; \mathrm{W}_{\mathrm{per}}(\Omega)),\ \partial_t v \in \mathrm{L}^2(0,T^\varepsilon; \mathrm{W}^*_{\mathrm{per}}(\Omega))\big\} &\hookrightarrow \mathcal{C}([0,T]; \mathrm{L}^2_0(\Omega)),\\ \big\{v \in \mathrm{L}^2(0,T^\varepsilon; \mathrm{L}^2_0(\Omega)),\ \partial_t v \in \mathrm{L}^2(0,T^\varepsilon; \mathrm{W}^*_{\mathrm{per}}(\Omega))\big\} &\hookrightarrow \mathcal{C}([0,T]; \mathrm{W}^*_{\mathrm{per}}(\Omega)). \end{aligned}$$

Finally, note that the choice of normalization, $\langle u^\varepsilon(t)\rangle_\Omega = 0$, is arbitrary. In fact, the well-posedness can be proved for any normalization $\langle u^\varepsilon(t)\rangle_\Omega = \langle g^0\rangle_\Omega$. The following theorem ensures the existence and uniqueness of a weak solution of (2.7).

Theorem 2.1.1. *Assume that the data satisfy*

$$g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega), \quad g^1 \in \mathrm{L}^2_0(\Omega), \quad f \in \mathrm{L}^2(0,T^\varepsilon; \mathrm{L}^2_0(\Omega)). \tag{2.10}$$

Then, there exists a unique weak solution u^ε of (2.7). Furthermore, the following estimate holds

$$\|\partial_t u^\varepsilon\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} + \|u^\varepsilon\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^1(\Omega))} \le C\big(\|g^1\|_{\mathrm{L}^2(\Omega)} + |g^0|_{\mathrm{H}^1(\Omega)} + \|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{L}^2(\Omega))}\big), \tag{2.11}$$

where C depends only on λ, Λ and the Poincaré constant C_Ω.

Remark 2.1.2. Note that the energy bound (2.11) depends on T^ε through the quantity

$$\|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{L}^2(\Omega))} = \int_0^{T^\varepsilon} \|f(t)\|_{\mathrm{L}^2(\Omega)} \mathrm{d}t.$$

Furthermore, C depends on the domain Ω through the Poincaré constant C_Ω. However, this dependence can be avoided if we keep the estimate in the H^1 seminorm (which is a norm on $\mathrm{W}_{\mathrm{per}}(\Omega)$), i.e.,

$$\|\partial_t u^\varepsilon\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} + |u^\varepsilon|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^1(\Omega))} \leq C\big(\|g^1\|_{\mathrm{L}^2(\Omega)} + |g^0|_{\mathrm{H}^1(\Omega)} + \|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{L}^2(\Omega))}\big), \quad (2.12)$$

where C depends only on λ and Λ. The bound (2.12) thus depends on T^ε and Ω only through the norms of the data.

Remark 2.1.3. Note that u^ε can be proved to satisfy the stronger regularity $u^\varepsilon \in \mathcal{C}^0([0,T^\varepsilon];\mathrm{W}_{\mathrm{per}}(\Omega))$ and $\partial_t u^\varepsilon \in \mathcal{C}^0([0,T^\varepsilon];\mathrm{L}^2_0(\Omega))$ (see e.g. [74]).

Proof. Let $\{w_\ell\}_{\ell=1}^\infty \subset \mathrm{W}_{\mathrm{per}}(\Omega)$ be the Fourier basis of $\mathrm{L}^2_0(\Omega)$. We define the finite dimensional space $V^m = \mathrm{span}\{w_\ell : 1 \leq \ell \leq m\}$. Let P^m be the projection onto V^m defined by the restriction of (2.3) to $\mathrm{L}^2_0(\Omega)$. Define $u^m(t) = \sum_{\ell=1}^m \alpha_\ell^m(t) w_\ell$, as the solution of the problem

$$\begin{array}{l} \big(\partial_t^2 u^m(t), w_k\big)_{\mathrm{L}^2} + A^\varepsilon\big(u^m(t), w_k\big) = \big(f(t), w_k\big)_{\mathrm{L}^2} \quad 1 \leq k \leq m \quad \text{for a.e. } t \in [0,T^\varepsilon], \\ u^m(0) = P^m g^0, \quad \partial_t u^m(0) = P^m g^1. \end{array} \quad (2.13)$$

Problem (2.13) can be rewritten as a second order ordinary differential equation on $[0,T^\varepsilon]$ for $\alpha^m(t) = \big(\alpha_1^m(t), \ldots, \alpha_m^m(t)\big)^T$:

$$\begin{array}{l} \bar{M}(\alpha^m)''(t) + \bar{A}\alpha^m(t) = F(t), \\ \alpha^m(0) = G^0, \quad (\alpha^m)'(0) = G^1, \end{array}$$

where $G_k^i = (g^i, w_k)_{\mathrm{L}^2}$, $(F(t))_k = \big(f(t), w_k\big)_{\mathrm{L}^2}$, and the $m \times m$ matrices $\bar{M}$ and $\bar{A}$ are defined as $\bar{M}_{k\ell} = (w_\ell, w_k)_{\mathrm{L}^2}$, $\bar{A}_{k\ell} = A^\varepsilon(w_\ell, w_k)_{\mathrm{L}^2}$. As M is positive definite, classical theory on ordinary differential equations ensures the existence and unicity of a solution $\alpha^m \in \mathcal{C}^1([0,T^\varepsilon];\mathbb{R}^m)$ with $(\alpha^m)'' \in \mathrm{L}^2([0,T^\varepsilon];\mathbb{R}^m)$ (see e.g. [38]). We thus have $u^m \in \mathcal{C}^1([0,T^\varepsilon];V^m)$ and $\partial_t^2 u^m \in \mathrm{L}^2(0,T^\varepsilon;V^m)$. Let us now derive an energy estimate for u^m that does not depend on m. For $t \in [0,T^\varepsilon]$, we multiply (2.13) by $(\alpha_k^m)'(t)$ and sum up over $1 \leq k \leq m$, to obtain for a.e. $t \in [0,T^\varepsilon]$

$$\big(\partial_t^2 u^m(t), \partial_t u^m(t)\big)_{\mathrm{L}^2} + A^\varepsilon\big(u^m(t), \partial_t u^m(t)\big) = \big(f(t), \partial_t u^m(t)\big)_{\mathrm{L}^2}.$$

Using the symmetry of $(\cdot,\cdot)_{\mathrm{L}^2}$ and A^ε, we rewrite this equality as

$$\tfrac{1}{2}\tfrac{\mathrm{d}}{\mathrm{d}t}\Big(\|\partial_t u^m(t)\|_{\mathrm{L}^2}^2 + A^\varepsilon\big(u^m(t), u^m(t)\big)\Big) = \big(f(t), \partial_t u^m(t)\big)_{\mathrm{L}^2}.$$

Defining $Eu^m(t) = \|\partial_t u^m(t)\|_{\mathrm{L}^2}^2 + A^\varepsilon(u^m(t), u^m(t))$, we integrate over $[0,\xi]$ and get for any $\xi \in [0,T^\varepsilon]$

$$Eu^m(\xi) = Eu^m(0) + 2\int_0^\xi \big(f(t), \partial_t u^m(t)\big)_{\mathrm{L}^2} \mathrm{d}t. \quad (2.14)$$

We bound the second term of the right hand side using the Cauchy–Schwartz, Hölder, and Young inequalities:

$$2\int_0^\xi \big(f(t), \partial_t u^m(t)\big)_{\mathrm{L}^2} \mathrm{d}t \leq 2\|f\|_{\mathrm{L}^1(\mathrm{L}^2)}\|\partial_t u^m\|_{\mathrm{L}^\infty(\mathrm{L}^2)} \leq 2\|f\|_{\mathrm{L}^1(\mathrm{L}^2)}^2 + \tfrac{1}{2}\|\partial_t u^m\|_{\mathrm{L}^\infty(\mathrm{L}^2)}^2.$$

As $A^\varepsilon\big(u^m(\xi), u^m(\xi)\big) \geq 0$, we have $\|\partial_t u^m(\xi)\|_{\mathrm{L}^2}^2 \leq Eu^m(\xi)$ and we obtain from (2.14)

$$\tfrac{1}{2}\|\partial_t u^m\|_{\mathrm{L}^\infty(\mathrm{L}^2)}^2 \leq Eu^m(0) + 2\|f\|_{\mathrm{L}^1(\mathrm{L}^2)}^2. \quad (2.15)$$

Then, again using (2.14), (2.8), and the Poincaré inequality, we deduce that for any ξ,

$$\|u^m(\xi)\|^2_{\mathrm{H}^1} \leq CA^\varepsilon\big(u^m(\xi),u^m(\xi)\big) \leq CEu^m(\xi) \leq C\big(Eu^m(0) + \|f\|^2_{\mathrm{L}^1(\mathrm{L}^2)}\big). \tag{2.16}$$

Note that $Eu^m(0) \leq \|g^1\|^2_{\mathrm{L}^2} + \Lambda|g^0|_{\mathrm{H}^1}$. Hence, combining (2.15) and (2.16), we obtain the energy estimate

$$\|\partial_t u^m\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \|u^m\|_{\mathrm{L}^\infty(\mathrm{H}^1)} \leq C\big(\|g^1\|_{\mathrm{L}^2} + |g^0|_{\mathrm{H}^1} + \|f\|_{\mathrm{L}^1(\mathrm{L}^2)}\big). \tag{2.17}$$

We next derive an estimate for $\|\partial_t^2 u^m\|_{\mathrm{L}^2(0,T^\varepsilon;\mathrm{W}^*_{\mathrm{per}}(\Omega))}$. For $v \in \mathrm{W}_{\mathrm{per}}(\Omega)$ such that $\|v\|_{\mathrm{H}^1} \leq 1$, as the basis $\{w_\ell\}_{\ell=1}^\infty$ is orthogonal and using (2.13), we find that

$$\big\langle \partial_t^2 u^m(t), v\big\rangle = \big(\partial_t^2 u^m(t), v\big)_{\mathrm{L}^2} = \big(\partial_t^2 u^m(t), P^m v\big)_{\mathrm{L}^2} = \big(f(t), P^m v\big)_{\mathrm{L}^2} - A^\varepsilon\big(u^m(t), P^m v\big).$$

As $|P^m v|_{\mathrm{H}^1} \leq |v|_{\mathrm{H}^1}$, we obtain

$$\|\partial_t^2 u^m(t)\|_{\mathrm{W}^*_{\mathrm{per}}} \leq \|f(t)\|_{\mathrm{L}^2} + \Lambda|u^m(t)|_{\mathrm{H}^1},$$

and thus, as $|u^m|_{\mathrm{L}^2(\mathrm{H}^1)} \leq \sqrt{T^\varepsilon}|u^m|_{\mathrm{L}^\infty(\mathrm{H}^1)}$, using (2.17) leads to

$$\|\partial_t^2 u^m\|_{\mathrm{L}^2(\mathrm{W}^*_{\mathrm{per}})} \leq C\sqrt{T^\varepsilon}\big(\|g^1\|_{\mathrm{L}^2} + |g^0|_{\mathrm{H}^1} + \|f\|_{\mathrm{L}^1(\mathrm{L}^2)}\big) + \|f\|_{\mathrm{L}^2(\mathrm{L}^2)}. \tag{2.18}$$

Estimates (2.17) and (2.18) imply that $\{u^m\}$, $\{\partial_t u^m\}$, and $\{\partial_t^2 u^m\}$ are bounded sequences in the spaces $\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega)) = [\mathrm{L}^1(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega)^*)]^*$, $\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2_0(\Omega)) = [\mathrm{L}^1(0,T^\varepsilon;\mathrm{L}^2_0(\Omega)^*)]^*$, and $\mathrm{L}^2(0,T^\varepsilon;\mathrm{W}^*_{\mathrm{per}}(\Omega))$, respectively. As $\mathrm{L}^2(0,T^\varepsilon;\mathrm{W}^*_{\mathrm{per}}(\Omega))$ is reflexive, and as the spaces $\mathrm{L}^1(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega)^*)$ and $\mathrm{L}^1(0,T^\varepsilon;\mathrm{L}^2_0(\Omega)^*)$ are separable, standard functional analysis results (see e.g. [94]) ensure the existence of subsequences of $\{u^m\}$, $\{\partial_t u^m\}$, $\{\partial_t^2 u^m\}$, still indexed by m, such that

$$\begin{aligned}
&u^m \rightharpoonup u^\varepsilon && \text{weakly}^* \text{ in } \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega)),\\
&\partial_t u^m \rightharpoonup \partial_t u^\varepsilon && \text{weakly}^* \text{ in } \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2_0(\Omega)),\\
&\partial_t^2 u^m \rightharpoonup \partial_t^2 u^\varepsilon && \text{weakly } \text{ in } \mathrm{L}^2(0,T^\varepsilon;\mathrm{W}^*_{\mathrm{per}}(\Omega)),
\end{aligned} \tag{2.19}$$

as $m \to \infty$. Furthermore, the limits u^ε, $\partial_t u^\varepsilon$, and $\partial_t^2 u^\varepsilon$ satisfy the same estimates as the sequences (2.17). Using the weak convergences (2.19), we can verify that u^ε satisfies (2.9) and is a weak solution of (2.7). To prove the uniqueness, we use the estimate (2.17). □

2.1.2 The Boussinesq equation

In this section, we prove the well-posedness of the Boussinesq equation (2.2). In a first part, we provide three results of well-posedness of the equation. The first one is more general as it requires less regularity of the data. The second and third ones ensures more regularity of the weak solution. In a second part, we state the corresponding results for the equation without fourth order operator. Indeed, this case is important for the one-dimensional study of long time effective models in locally periodic medium in Chapters 6 and 7. Finally, in the last part, we provide error estimates ensuring a higher regularity of the solution, in the case of constant tensors.

Let $\Omega \subset \mathbb{R}^d$ be a hypercube. Let $a^0 \in [\mathrm{L}^\infty_{\mathrm{per}}(\Omega)]^{d\times d}$ be a symmetric tensor, uniformly elliptic, and bounded tensor, i.e.,

$$\lambda|\xi|^2 \leq a^0(x)\xi\cdot\xi \leq \Lambda|\xi|^2 \quad \forall \xi \in \mathbb{R}^d \ \text{ for a.e. } x \in \Omega. \tag{2.20}$$

Let also $b^2 \in [\mathrm{L}^\infty_{\mathrm{per}}(\Omega)]^{d\times d}$ and let a^2 be a fourth order tensor function such that $a^2_{ijkl} \in \mathrm{L}^\infty_{\mathrm{per}}(\Omega)$. We assume that b^2 and a^2 satisfy

$$b^2_{ij} = b^2_{ji}, \qquad b^2(x)\xi\cdot\xi \geq 0 \quad \forall \xi \in \mathbb{R}^d \ \text{ for a.e. } x \in \Omega, \tag{2.21a}$$

$$a^2_{ijkl} = a^2_{klij}, \qquad a^2(x)\eta : \eta \geq 0 \quad \forall \eta \in \mathrm{Sym}^2(\mathbb{R}^d) \ \text{ for a.e. } x \in \Omega, \tag{2.21b}$$

where we recall the notation

$$a^2\eta : \xi = a^2_{ijkl}\eta_{kl}\xi_{ij} \quad \forall \eta, \xi \in \mathrm{Sym}^2(\mathbb{R}^d).$$

We consider the Boussinesq equation: $\tilde{u} : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ such that

$$\begin{array}{ll} \partial_t^2 \tilde{u} + \partial_i\big(a^0_{ij}\partial_j \tilde{u}\big) + \varepsilon^2 \partial^2_{ij}\big(a^2_{ijkl}\partial^2_{kl}\tilde{u}\big) - \varepsilon^2\partial_i\big(b^2_{ij}\partial_j\partial_t^2\tilde{u}\big) = f & \text{in } (0,T^\varepsilon]\times\Omega, \\ x \mapsto \tilde{u}(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,T^\varepsilon], \\ \tilde{u}(0,x) = g^0(x), \quad \partial_t\tilde{u}(0,x) = g^1(x) & \text{in } \Omega. \end{array} \tag{2.22}$$

We now present three different well-posedness results for (2.22). The first one is the more general. It ensures the existence and uniqueness of a weak solution under minimum regularity requirements. In the second result, under some stronger requirement on the tensors a^2, b^2, the weak solution is proved more regular. Finally, the third result provides conditions on the data for the weak solution to be even more regular.

We begin with the general well-posedness result. Let us introduce a first definition of a weak solution of (2.22). We define the forms

$$\begin{aligned} &\big(v,w\big)_{\mathcal{S}} = \big(v,w\big)_{\mathrm{L}^2} + \varepsilon^2\big(b^2\nabla v, \nabla w\big)_{\mathrm{L}^2}, \quad v,w \in \mathrm{L}^2_0(\Omega)\cap\mathrm{H}^1(\Omega), \\ &\tilde{A}\big(v,w\big) = \big(a^0\nabla v, \nabla w\big)_{\mathrm{L}^2} - \varepsilon^2\big(\partial_i v, \partial_j(a^2_{ijkl}\partial^2_{kl}w)\big)_{\mathrm{L}^2} \quad v \in \mathrm{W}_{\mathrm{per}}(\Omega), w \in \mathrm{W}_{\mathrm{per}}(\Omega)\cap\mathrm{H}^3(\Omega), \\ &A\big(v,w\big) = \big(a^0\nabla v, \nabla w\big)_{\mathrm{L}^2} + \varepsilon^2\big(a^2\nabla^2 v, \nabla^2 w\big)_{\mathrm{L}^2} \quad v,w \in \mathrm{W}_{\mathrm{per}}(\Omega)\cap\mathrm{H}^2(\Omega). \end{aligned}$$

Note that in the definition of $\tilde{A}$, we assume that $a^2 \in \mathrm{W}^{1,\infty}(\Omega)$. Furthermore, if v and w are sufficiently regular, an integration by parts ensures that $\tilde{A}\big(v,w\big) = A\big(v,w\big)$ We call a function $\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega))$, with $\partial_t\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2_0(\Omega))$, a weak solution of (2.22) if, for all test functions $v \in \mathcal{C}^2([0,T^\varepsilon];\mathrm{W}_{\mathrm{per}}(\Omega)(\Omega)\cap\mathrm{H}^3(\Omega))$ with $v(T^\varepsilon) = \partial_t v(T^\varepsilon) = 0$,

$$\int_0^{T^\varepsilon}\big(\tilde{u}(t),\partial_t^2 v(t)\big)_{\mathcal{S}} + \tilde{A}\big(\tilde{u}(t),v(t)\big)\,\mathrm{d}t = \int_0^{T^\varepsilon}\big(f(t),v(t)\big)_{\mathrm{L}^2(\Omega)}\,\mathrm{d}t + \big(g^1,v(0)\big)_{\mathcal{S}} - \big(g^0,\partial_t v(0)\big)_{\mathcal{S}}. \tag{2.23}$$

Using integration by parts (in space and time), we verify that a sufficiently regular weak solution $\tilde{u}$ satisfies (2.22) in a L^2 sense.

The following theorem is the more general well-posedness result for the Boussinesq equation (2.22) (the proof is provided below).

Theorem 2.1.4. *Assume that $a^2 \in \mathrm{W}^{1,\infty}(\Omega)$ and that the data satisfy the regularity*

$$g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega)\cap\mathrm{H}^2(\Omega), \quad g^1 \in \mathrm{L}^2_0(\Omega)\cap\mathrm{H}^1(\Omega), \quad f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{L}^2_0(\Omega)). \tag{2.24}$$

Then there exists a unique weak solution $\tilde{u}$ (in the sense of (2.23)), and the following estimate holds

$$\|\partial_t\tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} + \|\tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega))} \leq C\big(\|g^1\|_{\mathrm{H}^1(\Omega)} + \|g^0\|_{\mathrm{H}^2(\Omega)} + \|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{L}^2(\Omega))}\big), \tag{2.25}$$

where C depends only on $\lambda, \Lambda, \|b^2\|_{\mathrm{L}^\infty(\Omega)}, \varepsilon^2\|a^2\|_{\mathrm{L}^\infty(\Omega)}$ and the Poincaré constant C_Ω.

Let us now show that, under additional requirements on the tensors, we obtain a weak solution which is more regular. In this direction, let us define the natural functional spaces associated to (2.22). Note that as $b^2(x)$ is a symmetric positive semidefinite matrix for a.e. x, the square root $\sqrt{b^2(x)}$ is well defined (using the diagonalization of $b^2(x)$). We define in a similar manner the square root of a^2 as follows. Referring to Section 4.3.3, there exist a bijective map

$\nu : \mathrm{Sym}^2(\mathbb{R}^d) \to \mathbb{R}^{N(d)}$, where $N(d) = \binom{d+1}{2}$, and a symmetric $N(d) \times N(d)$ matrix $M(a^2)$ such that

$$a^2\eta : \xi = M(a^2)\nu(\eta) \cdot \nu(\xi) \quad \forall \eta, \xi \in \mathrm{Sym}^2(\mathbb{R}^d).$$

In particular, thanks to (2.21b), $M(a^2)$ is positive semidefinite. The square root of a^2 is then defined as $\sqrt{a^2} = \sqrt{M(a^2)}$. We define the functional spaces

$$\begin{aligned} \mathcal{S}(\Omega) &= \big\{ v \in \mathrm{L}^2_0(\Omega) : \sqrt{b^2}\nabla v \in [\mathrm{L}^2(\Omega)]^d \big\}, \\ \mathcal{V}(\Omega) &= \big\{ v \in \mathrm{W}_{\mathrm{per}}(\Omega) : \sqrt{a^2}\nabla^2 v \in [\mathrm{L}^2(\Omega)]^{d\times d} \big\}, \end{aligned} \tag{2.26}$$

where $\big(\sqrt{a^2}\nabla^2 v\big)_{ij} = \big(\sqrt{a^2}\big)_{ijkl}\partial^2_{kl} v$. Note that the spaces of definition of the forms $(\cdot,\cdot)_{\mathcal{S}}$ and $A(\cdot,\cdot)$ can be extended to $\mathcal{S}(\Omega)$ and $\mathcal{V}(\Omega)$, respectively. In order to ensure the spaces $\mathcal{S}(\Omega)$ and $\mathcal{V}(\Omega)$ to be complete, we require either of the following assumptions to hold:

$$\text{the tensors } a^2, b^2 \text{ are constant} \tag{H1}$$

$$\text{the tensors } a^2, b^2 \text{ are strictly positive definite} \tag{H2}$$

Then, if (H1) or (H2) holds, we verify that, equipped with the inner products $(\cdot,\cdot)_{\mathcal{S}}$ and $A(\cdot,\cdot)$, respectively, $\mathcal{S}(\Omega)$ and $\mathcal{V}(\Omega)$ are Hilbert spaces. On $\mathcal{V}(\Omega)$, we define the inner product $(v,w)_{\mathcal{V}} = (v,w)_{\mathcal{S}} + A(v,w)$. Thanks to the Poincaré–Wirtinger inequality, the ellipticity of a^0 and (2.21), the norms $\|v\|_{\mathcal{V}} = \sqrt{(v,v)_{\mathcal{V}}}$ and $\|v\| = \sqrt{A(v,v)}$ are equivalent. Using Riesz representation theorem, we obtain the following characterization for the dual $\mathcal{V}^*(\Omega)$. For $F \in \mathcal{V}^*(\Omega)$, there exist $f^0 \in \mathrm{L}^2_0(\Omega)$, $f^1_l, f^2_{kl} \in \mathrm{L}^2(\Omega)$, $1 \le k,l \le d$ such that

$$\big\langle F, v\big\rangle_{\mathcal{V}^*,\mathcal{V}} = \big(f^0, v\big)_{\mathrm{L}^2} + \big((a^0_{kl} + \varepsilon^2 b^2_{kl}) f^1_l, \partial_k v\big)_{\mathrm{L}^2} + \big(\varepsilon^2 a^2_{ijkl} f^2_{kl}, \partial^2_{ij} v\big)_{\mathrm{L}^2}. \tag{2.27}$$

We thus obtain the following embeddings

$$\mathcal{V}(\Omega) \hookrightarrow \mathcal{S}(\Omega) \hookrightarrow \mathrm{L}^2_0(\Omega) \hookrightarrow \mathcal{V}^*(\Omega).$$

Furthermore, $\mathcal{V}(\Omega)$ is dense in $\mathrm{L}^2_0(\Omega)$.

We now define a more regular weak solution of (2.22). If (H1) or (H2) holds, a function $\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{V}(\Omega))$ with $\partial_t \tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{S}(\Omega))$ is a weak solution of (2.22) if, for all test functions $v \in \mathcal{C}^2([0,T^\varepsilon];\mathcal{V}(\Omega))$ with $v(T^\varepsilon) = \partial_t v(T^\varepsilon) = 0$,

$$\int_0^{T^\varepsilon} \big(\tilde{u}(t), \partial^2_t v(t)\big)_{\mathcal{S}} + A\big(\tilde{u}(t), v(t)\big)\,\mathrm{d}t = \int_0^{T^\varepsilon} \big(f(t), v(t)\big)_{\mathrm{L}^2(\Omega)}\,\mathrm{d}t + \big(g^1, v(0)\big)_{\mathcal{S}} - \big(g^0, \partial_t v(0)\big)_{\mathcal{S}}. \tag{2.28}$$

Again, we verify that a sufficiently regular weak solution satisfies (2.22) in a L^2 sense.

The following theorem is the second well-posedness result for the Boussinesq equation. It ensures the existence and uniqueness of a weak solution in the sense of (2.28) (the proof is provided below).

Theorem 2.1.5. *Assume that either* (H1) *or* (H2) *holds and that the data satisfy the regularity*

$$g^0 \in \mathcal{V}(\Omega) \cap \mathrm{H}^2(\Omega), \quad g^1 \in \mathcal{S}(\Omega) \cap \mathrm{H}^1(\Omega), \quad f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{L}^2_0(\Omega)). \tag{2.29}$$

Then there exists a unique weak solution $\tilde{u}$ (in the sense of (2.28)*), and the following estimate holds*

$$\|\partial_t \tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{S})} + \|\tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{V})} \le C\big(\|g^1\|_{\mathrm{H}^1(\Omega)} + \|g^0\|_{\mathrm{H}^2(\Omega)} + \|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{L}^2(\Omega))}\big), \tag{2.30}$$

where C depends only on $\lambda, \Lambda, \|b^2\|_{\mathrm{L}^\infty(\Omega)}, \varepsilon^2\|a^2\|_{\mathrm{L}^\infty(\Omega)}$ and the Poincaré constant C_Ω.

Finally, the following result provides requirements on the data to obtain more regularity of the weak solution.

Theorem 2.1.6. *Assume that either* (H1) *or* (H2) *holds and that the data satisfy the regularity*

$$\begin{gathered} a^0 \in \mathrm{W}^{1,\infty}(\Omega), \quad a^2 \in \mathrm{W}^{2,\infty}(\Omega), \\ g^0 \in \mathcal{V}(\Omega) \cap \mathrm{H}^4(\Omega), \quad g^1 \in \mathcal{S}(\Omega) \cap \mathrm{H}^2(\Omega), \quad f \in \mathrm{H}^1(0,T^\varepsilon;\mathrm{L}^2_0(\Omega)). \end{gathered} \tag{2.31}$$

Then the unique weak solution $\tilde{u}$ (in the sense of (2.28)*) satisfies $\partial_t \tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{V}(\Omega))$, $\partial_t^2 \tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{S}(\Omega))$ and the following estimate holds:*

$$\|\partial_t^2 \tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{S})} + \|\partial_t \tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{V})} \leq C\big(\|g^1\|_{\mathrm{H}^2(\Omega)} + \|g^0\|_{\mathrm{H}^4(\Omega)} + \|f\|_{\mathrm{W}^{1,1}(0,T^\varepsilon;\mathrm{L}^2(\Omega))}\big), \tag{2.32}$$

where C depends only on $\lambda, \|b^2\|_{\mathrm{L}^\infty(\Omega)}, \|a^0\|_{\mathrm{W}^{1,\infty}(\Omega)}, \varepsilon^2\|a^2\|_{\mathrm{W}^{2,\infty}(\Omega)}$ and the Poincaré constant C_Ω.

Thanks to Theorem 2.1.6, under the regularity (2.31), $\tilde{u}$ satisfies the regularity

$$\tilde{u} \in \mathrm{W}^{1,\infty}(0,T^\varepsilon;\mathcal{V}(\Omega)) \hookrightarrow \mathcal{C}^0([0,T^\varepsilon];\mathcal{V}(\Omega)), \quad \partial_t \tilde{u} \in \mathrm{W}^{1,\infty}(0,T^\varepsilon;\mathcal{S}(\Omega)) \hookrightarrow \mathcal{C}^0([0,T^\varepsilon];\mathcal{S}(\Omega)). \tag{2.33}$$

Then, integrating by parts twice with respect to t in (2.28), and using the test function $w\varphi(t)$, where $w \in \mathcal{V}(\Omega)$ and $\varphi \in \mathcal{C}^2_c([0,T^\varepsilon])$, we obtain, by density of $\mathcal{C}^2_c([0,T^\varepsilon])$ in $\mathrm{L}^2(0,T^\varepsilon)$,

$$\int_0^{T^\varepsilon} \Big(\big(\partial_t^2 \tilde{u}(t), w\big)_{\mathcal{S}} + A\big(\tilde{u}(t), w\big) - \big(f(t), w\big)_{\mathrm{L}^2} \Big) \psi(t)\,\mathrm{d}t = 0 \quad \forall \psi \in \mathrm{L}^2(0,T^\varepsilon).$$

Hence, we verify that $\tilde{u}$ is the unique solution of the following variational formulation of (2.22),

$$\begin{aligned} &\big(\partial_t^2 \tilde{u}(t), w\big)_{\mathcal{S}} + A\big(\tilde{u}(t), w\big) = \big(f(t), w\big)_{\mathrm{L}^2} \quad \forall w \in \mathcal{V}(\Omega) \quad \text{ for a.e. } t \in [0,T^\varepsilon], \\ &\tilde{u}(0) = g^0, \quad \partial_t \tilde{u}(0) = g^1, \end{aligned} \tag{2.34}$$

where the initial conditions make sense in $\mathcal{V}(\Omega)$ and $\mathcal{S}(\Omega)$, respectively (thanks to (2.33)). Note that the variational formulation (2.34) is suited for the development and analysis of numerical methods such as the finite element method or the spectral method.

Proof of Theorem 2.1.5 (and Theorem 2.1.4). We prove here Theorem 2.1.5. The proof of Theorem 2.1.4 is very similar and the few necessary changes are specified.

Let $\{w_\ell\}_{\ell=1}^\infty$ be the Fourier basis of $\mathrm{L}^2_0(\Omega)$. We define the finite dimensional space $V^m = \mathrm{span}\{w_\ell : 1 \leq \ell \leq m\}$. Let P^m be the projection onto V^m defined by the restriction of (2.3) to $\mathrm{L}^2_0(\Omega)$. We define $u^m(t) = \sum_{\ell=1}^m \alpha_\ell^m(t) w_\ell$ as the solution of the problem

$$\begin{aligned} &\big(\partial_t^2 u^m(t), w_k\big)_{\mathcal{S}} + A\big(u^m(t), w_k\big) = \big(f(t), w_k\big)_{\mathrm{L}^2} \quad 1 \leq k \leq m \quad \text{ for a.e. } t \in [0,T^\varepsilon], \\ &u^m(0) = P^m g^0, \quad \partial_t u^m(0) = P^m g^1. \end{aligned} \tag{2.35}$$

Problem (2.35) can be rewritten as a second order ordinary differential equation on $[0,T^\varepsilon]$ for $\alpha^m(t) = \big(\alpha_1^m(t), \dots, \alpha_m^m(t)\big)^T$:

$$\begin{aligned} &\bar{M}(\alpha^m)''(t) + \bar{A}\alpha^m(t) = F(t), \\ &\alpha^m(0) = G^0, \quad (\alpha^m)'(0) = G^1, \end{aligned}$$

where $G^i_k = (g^i, w_k)_{\mathrm{L}^2}$, $(F(t))_k = \big(f(t), w_k\big)_{\mathrm{L}^2}$, and the $m \times m$ matrices $\bar{M}$ and $\bar{A}$ are defined as $\bar{M}_{k\ell} = (w_\ell, w_k)_{\mathcal{S}}$, $\bar{A}_{k\ell} = A(w_\ell, w_k)_{\mathrm{L}^2}$. As b^2 is positive semidefinite, $\bar{M}$ is positive definite. Consequently, classical theory on ordinary differential equations ensures the existence and unicity

of a solution $\alpha^m \in \mathcal{C}^1([0,T^\varepsilon];\mathbb{R}^m)$ with $(\alpha^m)'' \in \mathrm{L}^2(0,T^\varepsilon;\mathbb{R}^m)$ (see e.g. [38]). Hence, we have $u^m \in \mathcal{C}^1([0,T^\varepsilon];V^m)$ and $\partial_t^2 u^m \in \mathrm{L}^2(0,T^\varepsilon;V^m)$. Let us now derive an energy estimate for u^m, independently of m. For a $t \in [0,T^\varepsilon]$, we multiply (2.35) by $(\alpha_k^m)'(t)$ and sum over $1 \le k \le m$, to obtain for a.e. $t \in [0,T^\varepsilon]$

$$\big(\partial_t^2 u^m(t), \partial_t u^m(t)\big)_{\mathcal{S}} + A(u^m(t), \partial_t u^m(t)) = \big(f(t), \partial_t u^m(t)\big)_{\mathrm{L}^2}.$$

Using the symmetry of the forms, this equality can be rewritten as

$$\tfrac{1}{2}\tfrac{\mathrm{d}}{\mathrm{d}t}\Big(\|\partial_t u^m(t)\|_{\mathcal{S}}^2 + A(u^m(t),u^m(t))\Big) = \big(f(t),\partial_t u^m(t)\big)_{\mathrm{L}^2}.$$

Defining $Eu^m(t) = \|\partial_t u^m(t)\|_{\mathcal{S}}^2 + A(u^m(t),u^m(t))$, we integrate over $[0,\xi]$ for $\xi \in [0,T^\varepsilon]$ and get

$$Eu^m(\xi) = Eu^m(0) + 2\int_0^\xi \big(f(t),\partial_t u^m(t)\big)_{\mathrm{L}^2}\,\mathrm{d}t. \tag{2.36}$$

We bound the second term of the right hand side using Cauchy–Schwartz, Hölder, and Young inequalities:

$$2\int_0^\xi \big(f(t),\partial_t u^m(t)\big)_{\mathrm{L}^2}\,\mathrm{d}t \le 2\|f\|_{\mathrm{L}^1(\mathrm{L}^2)}\|\partial_t u^m\|_{\mathrm{L}^\infty(\mathrm{L}^2)} \le 2\|f\|_{\mathrm{L}^1(\mathrm{L}^2)}^2 + \tfrac{1}{2}\|\partial_t u^m\|_{\mathrm{L}^\infty(\mathcal{S})}^2.$$

As $A(u^m(\xi),u^m(\xi)) \ge 0$, we have $\|\partial_t u^m(\xi)\|_{\mathcal{S}}^2 \le Eu^m(\xi)$ and we obtain from (2.36)

$$\tfrac{1}{2}\|\partial_t u^m\|_{\mathrm{L}^\infty(\mathcal{S})}^2 \le Eu^m(0) + 2\|f\|_{\mathrm{L}^1(\mathrm{L}^2)}^2. \tag{2.37}$$

Then, again using (2.36), we deduce that for any $\xi \in [0,T^\varepsilon$,

$$\|u^m(\xi)\|_{\mathcal{V}}^2 \le CA\big(u^m(\xi),u^m(\xi)\big) \le CEu^m(\xi) \le C\big(Eu^m(0) + \|f\|_{\mathrm{L}^1(\mathrm{L}^2)}^2\big). \tag{2.38}$$

Let us now bound $Eu^m(0) = \|P^m g^1\|_{\mathcal{S}}^2 + A\big(P^m g^0, P^m g^0\big)$. For the first term, recalling the stability of P^m (2.5), we have

$$\|P^m g^1\|_{\mathcal{S}}^2 \le \|P^m g^1\|_{\mathrm{L}^2}^2 + C\varepsilon^2\|\nabla(P^m g^1)\|_{\mathrm{L}^2}^2 \le C\big(\|g^1\|_{\mathrm{L}^2}^2 + \varepsilon^2\|\nabla g^1\|_{\mathrm{L}^2}^2\big).$$

For the second term, we have

$$A\big(P^m g^0, P^m g^0\big) \le C\big(\|\nabla(P^m g^0)\|_{\mathrm{L}^2}^2 + \varepsilon^2\|\nabla^2(P^m g^0)\|_{\mathrm{L}^2}^2 \le C\big(\|\nabla g^0\|_{\mathrm{L}^2}^2 + \varepsilon^2\|\nabla^2 g^0\|_{\mathrm{L}^2}^2\big).$$

Combining the two last estimates with (2.37) and (2.38), we obtain the energy estimate

$$\|\partial_t u^m\|_{\mathrm{L}^\infty(\mathcal{S})} + \|u^m\|_{\mathrm{L}^\infty(\mathcal{V})} \le C\big(\|g^1\|_{\mathrm{H}^1} + \|g^0\|_{\mathrm{H}^2} + \|f\|_{\mathrm{L}^1(\mathrm{L}^2)}\big). \tag{2.39}$$

Estimate (2.39) implies that $\{u^m\}$ and $\{\partial_t u^m\}$ are bounded sequences in the spaces $\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{V}(\Omega)) = [\mathrm{L}^1(0,T^\varepsilon;\mathcal{V}^*(\Omega))]^*$ and $\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{S}(\Omega)) = [\mathrm{L}^1(0,T^\varepsilon;\mathcal{S}^*(\Omega))]^*$, respectively. As the spaces $\mathrm{L}^1(0,T^\varepsilon;\mathcal{V}^*(\Omega))$ and $\mathrm{L}^1(0,T^\varepsilon;\mathcal{S}^*(\Omega))$ are separable, standard functional analysis results (see e.g. [94]) ensure the existence of subsequences of $\{u^m\}$, $\{\partial_t u^m\}$, still indexed by m, such that

$$\begin{aligned} &u^m \rightharpoonup \tilde{u} && \text{weakly}^* \text{ in } \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{V}(\Omega)),\\ &\partial_t u^m \rightharpoonup \partial_t \tilde{u} && \text{weakly}^* \text{ in } \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{S}(\Omega)), \end{aligned} \tag{2.40}$$

as $m \to \infty$. Furthermore, the limits $\tilde{u}$ and $\partial_t \tilde{u}$ satisfy the same estimate as u^m (2.39).

In the context of Theorem 2.1.4, we note that (2.39) implies the estimate

$$\|\partial_t u^m\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \|u^m\|_{\mathrm{L}^\infty(\mathrm{H}^1)} \le C\big(\|g^1\|_{\mathrm{H}^1} + \|g^0\|_{\mathrm{H}^2} + \|f\|_{\mathrm{L}^1(\mathrm{L}^2)}\big).$$

The sequences $\{u^m\}$ and $\{\partial_t u^m\}$ are thus bounded in the spaces $\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega)) = [\mathrm{L}^1(0,T^\varepsilon;\mathrm{W}^*_{\mathrm{per}}(\Omega))]^*$ and $\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2_0(\Omega)) = [\mathrm{L}^1(0,T^\varepsilon;\mathrm{L}^2_0(\Omega))]^*$, and we obtain a weak limit of subsequence in these spaces.

Finally, we prove that the weak limit $\tilde{u}$ is a weak solution, in the sense of (2.28) (for Theorem 2.1.4, we prove similarly that the weak limit is a weak solution in the sense of (2.23)). Note that the space of functions of the form φw, with $\varphi \in \mathcal{C}^2(0,T^\varepsilon)$, $\varphi(T^\varepsilon) = \varphi'(T^\varepsilon) = 0$, and $w \in \mathcal{V}(\Omega)$ is dense in the space of test functions. It is thus sufficient to verify that (2.28) holds for $v = \varphi w_k$. Multiplying (2.35) with $\varphi(t)$, integrating over $[0,T^\varepsilon]$ and integrating by parts, we obtain

$$\int_0^{T^\varepsilon} \big(u^m(t), \varphi''(t)w_k\big)_{\mathcal{S}}\,\mathrm{d}t + \int_0^{T^\varepsilon} A\big(u^m(t),\varphi(t)w_k\big)\,\mathrm{d}t = \int_0^{T^\varepsilon} \big(f(t),\varphi(t)w_k\big)_{\mathrm{L}^2}\,\mathrm{d}t + \big(P^m g^1, v(0)\big)_{\mathcal{S}} - \big(P^m g^0, \partial_t v(0)\big)_{\mathcal{S}}.$$

Thanks to the weak* convergences (2.40) and as $\lim_{m\to\infty} P^m g^i = g^i$, we verify that $\tilde{u}$ satisfies (2.28) for $v = \varphi w_k$ ($\partial_t v = \varphi' w_k$) and that completes the proof. □

Proof of Theorem 2.1.6. Thanks to the regularity and symmetry of the tensors a^0 and a^2, the proof of Theorem 2.1.5 can be performed with the orthonormal basis of $\mathrm{L}^2_0(\Omega)$ formed by the eigenfunctions of the elliptic operator $\mathcal{A}v = -\partial_i(a^0_{ij}\partial_j v) + \varepsilon^2\partial_{ij}(a^2_{ijkl}\partial_{kl}v)$. We still denote the basis $\{w_k\}_{k\in\mathbb{N}} \subset \mathcal{V}(\Omega)$. From the time differentiation of (2.35), similarly as (2.39) yields

$$\|\partial_t^2 u^m\|_{\mathrm{L}^\infty(\mathcal{S})} + \|\partial_t u^m\|_{\mathrm{L}^\infty(\mathcal{V})} \le C\big(\|\partial_t^2 u^m(0)\|_{\mathcal{S}} + \|g^1\|_{\mathrm{H}^2} + \|\partial_t f\|_{\mathrm{L}^1(\mathrm{L}^2(\Omega))}\big). \tag{2.41}$$

Let us estimate the term $\|\partial_t u^m(0)\|_{\mathcal{S}}$. Using (2.35), we get

$$\|\partial_t^2 u^m(0)\|^2_{\mathcal{S}} = \big(f(0) - \mathcal{A}u^m(0), \partial_t^2 u^m(0)\big)_{\mathrm{L}^2}.$$

Hence

$$\|\partial_t^2 u^m(0)\|_{\mathcal{S}} \le \|f(0)\|_{\mathrm{L}^2} + \|\mathcal{A}u^m(0)\|_{\mathrm{L}^2}. \tag{2.42}$$

The embedding $\mathrm{W}^{1,1}(0,T^\varepsilon;\mathrm{L}^2(\Omega)) \hookrightarrow \mathcal{C}^0([0,T^\varepsilon];\mathrm{L}^2(\Omega))$ implies that

$$\|f(0)\|_{\mathrm{L}^2} \le \max\{1, 1/T^\varepsilon\}\|f\|_{\mathrm{W}^{1,1}(\mathrm{L}^2)}.$$

We still have to estimate $\|\mathcal{A}u^m(0)\|_{\mathrm{L}^2}$. Integrating by parts we have

$$\|\mathcal{A}u^m(0)\|^2_{\mathrm{L}^2} = \big(\mathcal{A}u^m(0), \mathcal{A}u^m(0)\big)_{\mathrm{L}^2} = \big(P^m g^0, \mathcal{A}^2 u^m(0)\big)_{\mathrm{L}^2}.$$

As $\mathcal{A}w_k = \lambda_k w_k$ for any $k \in \mathbb{N}$, we verify that $\mathcal{A}^2 u^m(0) \in V^m$ and thus

$$\|\mathcal{A}u^m(0)\|^2_{\mathrm{L}^2} = \big(g^0, \mathcal{A}^2 u^m(0)\big)_{\mathrm{L}^2} = \big(\mathcal{A}g^0, \mathcal{A}u^m(0)\big)_{\mathrm{L}^2} \le C\|g^0\|_{\mathrm{H}^4}\|\mathcal{A}u^m(0)\|_{\mathrm{L}^2}, \tag{2.43}$$

where C depends on $\|a^0\|_{\mathrm{W}^{1,\infty}(\Omega)}$ and $\varepsilon^2\|a^2\|_{\mathrm{W}^{2,\infty}(\Omega)}$. Combining estimates (2.41), (2.42) and (2.43) and passing to the limit $m \to \infty$ proves estimate (2.32). □

Special case: no fourth order operator

In the case without the fourth order tensor a^2 in the Boussinesq equation 2.22, the well-posedness can be proved under weaker regularity of the data. We state here the results, the proofs follow the same lines as for Theorems 2.1.5 and 2.1.6.

We consider the following equation: find $\bar{u} : [0,T^\varepsilon] \times \Omega \to \mathbb{R}$ such that

$$\begin{array}{ll} \partial_t^2 \bar{u} - \partial_i\big(a^0_{ij}\partial_j \bar{u}\big) - \varepsilon^2\partial_i\big(b^2_{ij}\partial_j\partial_t^2\bar{u}\big) = f & \text{in } (0,T^\varepsilon]\times\Omega, \\ x \mapsto \bar{u}(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,T^\varepsilon], \\ \bar{u}(0) = g^0, \quad \partial_t\bar{u}(0) = g^1 & \text{in } \Omega. \end{array} \tag{2.44}$$

We define a weak solution of (2.44) similarly as in (2.28) (with $\mathcal{V}(\Omega) = \mathrm{W}_{\mathrm{per}}(\Omega)$ and $a^2 = 0$). Define the bilinear forms

$$\big(v, w\big)_{\mathcal{S}} = \big(v, w\big)_{\mathrm{L}^2} + \varepsilon^2 \big(b^2 \nabla v, \nabla w\big)_{\mathrm{L}^2}, \quad v, w \in \mathcal{S}(\Omega), \tag{2.45}$$

$$A(v, w) = \big(a^0 \nabla v, \nabla w\big)_{\mathrm{L}^2} \quad v, w \in \mathrm{W}_{\mathrm{per}}(\Omega). \tag{2.46}$$

We call a function $\bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{W}_{\mathrm{per}}(\Omega))$, with $\partial_t \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathcal{S}(\Omega))$, a weak solution of (2.44) if, for all test functions $v \in \mathcal{C}^2([0, T^\varepsilon]; \mathrm{W}_{\mathrm{per}}(\Omega))$ with $v(T^\varepsilon) = \partial_t v(T^\varepsilon) = 0$,

$$\int_0^{T^\varepsilon} \big(\bar{u}(t), \partial_t^2 v(t)\big)_{\mathcal{S}} + A\big(\tilde{u}(t), v(t)\big)\, \mathrm{d}t = \int_0^{T^\varepsilon} \big(f(t), v(t)\big)_{\mathrm{L}^2(\Omega)}\, \mathrm{d}t + \big(g^1, v(0)\big)_{\mathcal{S}} - \big(g^0, \partial_t v(0)\big)_{\mathcal{S}}. \tag{2.47}$$

The following result provides the existence and uniqueness of a weak solution $\bar{u}$ to (2.44).

Theorem 2.1.7. *Assume that the data satisfy the following regularity*

$$g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega), \quad g^1 \in \mathcal{S}(\Omega) \cap \mathrm{H}^1(\Omega), \quad f \in \mathrm{L}^2(0, T^\varepsilon; \mathrm{L}^2_0(\Omega)). \tag{2.48}$$

Then there exists a unique weak solution $\bar{u}$ of (2.44) and the following estimate holds

$$\|\partial_t \bar{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{S}(\Omega))} + \|\bar{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^1(\Omega))} \leq C\big(\|g^1\|_{\mathrm{H}^1(\Omega)} + \|g^0\|_{\mathrm{H}^1(\Omega)} + \|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{L}^2(\Omega))}\big), \tag{2.49}$$

where C depends only on $\lambda, \Lambda, \|b^2\|_{\mathrm{L}^\infty(\Omega)}$ and the Poincaré constant C_Ω.

Under stronger regularity assumptions on the data, we prove a higher regularity of the weak solution.

Theorem 2.1.8. *Assume that*

$$a^0 \in \mathrm{W}^{1,\infty}(\Omega), \quad g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^2(\Omega), \quad g^1 \in \mathcal{S}(\Omega) \cap \mathrm{H}^1(\Omega), \quad f \in \mathrm{H}^1(0, T^\varepsilon; \mathrm{L}^2_0(\Omega)). \tag{2.50}$$

Then $\partial_t \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^1(\Omega))$, $\partial_t^2 \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathcal{S}(\Omega))$ and the following estimate holds:

$$\|\partial_t^2 \bar{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{S}(\Omega))} + \|\partial_t \bar{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^1(\Omega))} \leq C\big(\|g^1\|_{\mathrm{H}^1(\Omega)} + \|g^0\|_{\mathrm{H}^2(\Omega)} + \|f\|_{\mathrm{W}^{1,1}(0,T^\varepsilon;\mathrm{L}^2(\Omega))}\big), \tag{2.51}$$

where C depends only on $\lambda, \|b^2\|_{\mathrm{L}^\infty(\Omega)}, \|a^0\|_{\mathrm{W}^{1,\infty}(\Omega)}$ and the Poincaré constant C_Ω.

Theorem 2.1.8 ensures that if (2.50) holds, $\bar{u}$ is the unique solution of the following variational formulation of (2.44): $\bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{W}_{\mathrm{per}}(\Omega))$, with $\partial_t \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathcal{S}(\Omega))$ and $\partial_t^2 \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathcal{S}(\Omega))$, such that

$$\begin{aligned} &\big(\partial_t^2 \tilde{u}(t), w\big)_{\mathcal{S}} + A\big(\tilde{u}(t), w\big) = \big(f(t), w\big)_{\mathrm{L}^2} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(\Omega) \quad \text{ for a.e. } t \in [0, T^\varepsilon], \\ &\tilde{u}(0) = g^0, \quad \partial_t \tilde{u}(0) = g^1, \end{aligned} \tag{2.52}$$

where the initial conditions makes sense in $\mathrm{W}_{\mathrm{per}}(\Omega)$ and $\mathcal{S}(\Omega)$, respectively (thanks to (2.33) with $\mathcal{V}(\Omega) = \mathrm{W}_{\mathrm{per}}(\Omega)$).

Energy estimate for higher regularity of the solution (constant tensors)

In the last part of this section, we prove energy estimates that ensure a higher regularity of the weak solution of the Boussinesq equation 2.22, in the case of constant tensors a^0, b^2, a^2.

Theorem 2.1.9. *Assume that the tensors a^0, b^2 and a^2 are constant, that f is Ω-periodic and assume that the assumptions of Theorem 2.1.5 holds.*

i) If we assume in addition that for some $k \geq 0$ the data satisfy the regularity

$$g^0 \in \mathrm{H}^{k+2}(\Omega), \quad g^1 \in \mathrm{H}^{k+1}(\Omega), \quad f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{H}^k(\Omega)),$$

then the weak solution $\tilde{u}$ of (2.28) satisfies the estimate

$$|\partial_t \tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^k(\Omega))} + |\tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^{k+1}(\Omega))} \leq C\big(\|g^1\|_{\mathrm{H}^{k+1}(\Omega)} + \|g^0\|_{\mathrm{H}^{k+2}(\Omega)} + \|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^k(\Omega))}\big), \tag{2.53}$$

where the constant depends only on λ, a^0, b^2, a^2.

ii) If we assume in addition that for some $k \geq 0$ the data satisfy the regularity

$$g^0 \in \mathrm{H}^{k+4}(\Omega), \quad g^1 \in \mathrm{H}^{k+2}(\Omega), \quad f \in \mathrm{H}^1(0,T^\varepsilon;\mathrm{H}^k(\Omega)),$$

then the weak solution $\tilde{u}$ of (2.28) satisfies the estimate

$$|\partial_t^2 \tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^k(\Omega))} + |\partial_t \tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^{k+1}(\Omega))} \leq C\big(\|g^1\|_{\mathrm{H}^{k+2}(\Omega)} + \|g^0\|_{\mathrm{H}^{k+4}(\Omega)} + \|f\|_{\mathrm{W}^{1,1}(0,T^\varepsilon;\mathrm{H}^k(\Omega))}\big), \tag{2.54}$$

where the constant depends only on λ, a^0, b^2, a^2 and $\max\{1, 1/T^\varepsilon\}$.

Proof. We prove estimate (2.53) and (2.54) for $k = 1$. The proof for a general $k \geq 0$ follows the same line with obvious modifications. For the sake of simplicity, we assume that $\Omega = (0, L_1) \times \cdots \times (0, L_d)$. Let $\{w_\ell\}_{\ell \in \mathbb{N}}$ be the Fourier basis and consider the approximated problem in V^m given in (2.35).

i) Recall that the Fourier basis functions satisfy $\partial_{x_\nu} w_\ell = D_\nu^{\Omega,\ell} w_\ell$, where $D_\nu^{\Omega,\ell} \in \mathrm{i}\mathbb{R}$. Hence, multiplying (2.35) by $-D_\nu^{\Omega,\ell}$, we obtain

$$-\big(\partial_t^2 u^m(t), \partial_{x_\nu} w_\ell\big)_{\mathcal{S}} - A\big(u^m(t), \partial_{x_\nu} w_\ell\big) = -\big(f(t), \partial_{x_\nu} w_\ell\big)_{\mathrm{L}^2} \quad 1 \leq \ell \leq m \quad \text{for a.e. } t \in [0, T^\varepsilon].$$

As the tensors are assumed to be constant, we integrate by parts and get

$$\big(\partial_t^2 u_\nu^m(t), w_\ell\big)_{\mathcal{S}} + A\big(u_\nu^m(t), w_\ell\big) = \big(f_\nu(t), w_\ell\big)_{\mathrm{L}^2} \quad 1 \leq \ell \leq m \quad \text{for a.e. } t \in [0, T^\varepsilon]. \tag{2.55}$$

where we used the short hand notation $u_\nu^m(t) = \partial_{x_\nu} u^m(t) = \sum_{\ell=1}^m \alpha_\ell(t) \partial_{x_\nu} w_\ell$ and $f_\nu = \partial_{x_\nu} f$. Multiplying this equality by $\dot{\alpha}_\ell(t) D_\nu^{\Omega,\ell}$ and summing over $1 \leq \ell \leq m$, we obtain

$$\big(\partial_t^2 u_\nu^m(t), \partial_t u_\nu^m(t)\big)_{\mathcal{S}} + A\big(u_\nu^m(t), \partial_t u_\nu^m(t)\big) = \big(f_\nu(t), \partial_t u_\nu^m(t)\big)_{\mathrm{L}^2} \quad \text{for a.e. } t \in [0, T^\varepsilon],$$

which can be rewritten as

$$\tfrac{1}{2} \tfrac{\mathrm{d}}{\mathrm{d}t} \Big(\|\partial_t u_\nu^m(t)\|_{\mathcal{S}}^2 + A\big(u_\nu^m(t), u_\nu^m(t)\big) \Big) = \big(f_\nu(t), \partial_t u_\nu^m(t)\big)_{\mathrm{L}^2} \quad \text{for a.e. } t \in [0, T^\varepsilon].$$

Denoting $E u_\nu^m(t) = \|\partial_t u_\nu^m(t)\|_{\mathcal{S}}^2 + A\big(u_\nu^m(t), u_\nu^m(t)\big)$, we integrate the equality over $[0, \xi]$ and get

$$E u_\nu^m(\xi) = E u_\nu^m(0) + 2\big(f_\nu(t), \partial_t u_\nu^m(t)\big)_{\mathrm{L}^2}. \tag{2.56}$$

The second term of the right hand side is bounded using Cauchy–Schwartz, Hölder and Young inequalities:

$$2\int_0^\xi \big(f_\nu(t), \partial_t u_\nu^m(t)\big)_{\mathrm{L}^2} \,\mathrm{d}t \leq 2\|f_\nu\|_{\mathrm{L}^1(\mathrm{L}^2)} \|\partial_t u_\nu^m\|_{\mathrm{L}^\infty(\mathrm{L}^2)} \leq 2\|f_\nu\|_{\mathrm{L}^1(\mathrm{L}^2)}^2 + \tfrac{1}{2}\|\partial_t u_\nu^m\|_{\mathrm{L}^\infty(\mathcal{S})}^2. \tag{2.57}$$

Taking the L^∞ norm with respect to ξ in (2.56), we thus obtain

$$\tfrac{1}{2}\|\partial_t u_\nu^m\|_{\mathrm{L}^\infty(\mathcal{S})}^2 \leq E u_\nu^m(0) + 2\|f_\nu\|_{\mathrm{L}^1(\mathrm{L}^2)}^2,$$

which combined with (2.56) and (2.57) gives

$$\operatorname*{ess\,sup}_{\xi\in[0,T^\varepsilon]} A\big(u_\nu^m(\xi),u_\nu^m(\xi)\big) \le Eu_\nu^m(0) + 2\|f_\nu\|^2_{\mathrm{L}^1(\mathrm{L}^2)}.$$

Thanks to the properties of $A(\cdot,\cdot)$ and the definition of $\|\cdot\|_{\mathcal{S}}$, we thus have

$$\operatorname*{ess\,sup}_{\xi\in[0,T^\varepsilon]} \|\partial_t u_\nu^m(\xi)\|_{\mathrm{L}^2} + \operatorname*{ess\,sup}_{\xi\in[0,T^\varepsilon]} |u_\nu^m(\xi)|_{\mathrm{H}^1} \le C\big(\|\partial_t u_\nu^m(0)\|_{\mathrm{H}^1} + \|u_\nu^m(0)\|_{\mathrm{H}^2} + \|f_\nu\|_{\mathrm{L}^1(\mathrm{L}^2)}\big), \quad (2.58)$$

where C depends only on a^0, a^2, b^2. As $u_\nu^m(0) = \partial_{x_\nu}(P^m g^0) = P^m(\partial_{x_\nu} g^0)$ and $\partial_t u_\nu^m(0) = P^m(\partial_{x_\nu} g^1)$, using the stability of P^m and applying (2.58) for $\nu = 1,\dots,d$, we obtain the estimate

$$\operatorname*{ess\,sup}_{\xi\in[0,T^\varepsilon]} |\partial_t u^m(\xi)|_{\mathrm{H}^1} + \operatorname*{ess\,sup}_{\xi\in[0,T^\varepsilon]} |u^m(\xi)|_{\mathrm{H}^2} \le C\big(\|g^1\|_{\mathrm{H}^2} + \|g^0\|_{\mathrm{H}^3} + \|f\|_{\mathrm{L}^1(\mathrm{H}^1)}\big). \quad (2.59)$$

Taking the limit $m\to\infty$, we obtain (2.53) for $k=1$.

ii) Let us now prove (2.54) for $k=1$. From the time differentiation of (2.35), we obtain in a similar manner as (2.58) (we keep the $\mathcal{S}$ norm for the first term of $E\partial_t^2 u_\nu^m(0)$)

$$\operatorname*{ess\,sup}_{\xi\in[0,T^\varepsilon]} |\partial_t^2 u^m(\xi)|_{\mathrm{H}^1} + \operatorname*{ess\,sup}_{\xi\in[0,T^\varepsilon]} |\partial_t u^m(\xi)|_{\mathrm{H}^2} \le C\big(\|\partial_t^2 u_\nu^m(0)\|_{\mathcal{S}} + \|g^1\|_{\mathrm{H}^3} + \|\partial_t f\|_{\mathrm{L}^1(\mathrm{H}^1)}\big). \quad (2.60)$$

Let us bound $\|\partial_t^2 u_\nu^m(0)\|_{\mathcal{S}}$. Using (2.55) , we have

$$\|\partial_t^2 u_\nu^m(0)\|_{\mathcal{S}}^2 = \big(\partial_t^2 u_\nu^m(0), \partial_t^2 u_\nu^m(0)\big)_{\mathcal{S}} = \big(f_\nu(0) - \mathcal{A}u_\nu^m(0), \partial_t^2 u_\nu^m(0)\big)_{\mathrm{L}^2},$$

which, using Cauchy–Schwartz, implies

$$\|\partial_t^2 u_\nu^m(0)\|_{\mathcal{S}} \le \|f_\nu(0)\|_{\mathrm{L}^2} + \|\mathcal{A}u_\nu^m(0)\|_{\mathrm{L}^2}. \quad (2.61)$$

To bound the first term of the right hand side, we use the continuous embedding $\mathrm{W}^{1,1}(0,T^\varepsilon;\mathrm{L}^2(\Omega)) \hookrightarrow \mathcal{C}([0,T^\varepsilon];\mathrm{L}^2(\Omega))$ which implies that

$$\|f_\nu(0)\|_{\mathrm{L}^2} = \|\partial_{x_\nu} f(0)\|_{\mathrm{L}^2} \le \max\{1, 1/T^\varepsilon\}\|f\|_{\mathrm{W}^{1,1}(\mathrm{H}^1)}.$$

For the second term, note that as the tensors are constant we have $\mathcal{A}^2 u_\nu^m(0) \in V^m$. Hence, integrating by parts and using that $\partial_{x_\nu} P^m g^0 = P^m(\partial_{x_\nu} g^0)$ and the definition of P^m, we have

$$\|\mathcal{A}u_\nu^m(0)\|^2_{\mathrm{L}^2} = \big(P^m\partial_{x_\nu} g^0, \mathcal{A}^2 u_\nu^m(0)\big)_{\mathrm{L}^2} = \big(\partial_{x_\nu} g^0, \mathcal{A}^2 u_\nu^m(0)\big)_{\mathrm{L}^2} = \big(\mathcal{A}(\partial_{x_\nu} g^0), \mathcal{A}u_\nu^m(0)\big)_{\mathrm{L}^2},$$

which implies via Cauchy–Schwartz inequality that $\|\mathcal{A}u_\nu^m(0)\|_{\mathrm{L}^2} \le C\|g^0\|_{\mathrm{H}^5}$, where C depends only on a^0 and a^2. Finally, combining (2.60), (2.61) with the two last bounds, we obtain

$$\operatorname*{ess\,sup}_{\xi\in[0,T^\varepsilon]} |\partial_t^2 u^m(\xi)|_{\mathrm{H}^1} + \operatorname*{ess\,sup}_{\xi\in[0,T^\varepsilon]} |\partial_t u^m(\xi)|_{\mathrm{H}^2} \le C\big(\|g^1\|_{\mathrm{H}^3} + \|g^0\|_{\mathrm{H}^5} + \|f\|_{\mathrm{W}^{1,1}(\mathrm{H}^1)}\big).$$

Taking the limit $m\to\infty$, we obtain (2.54) for $k=1$. □

2.2 The finite element method for hyperbolic equations

In this section, we present the finite element method with numerical integration for the approximation of the Boussinesq equation (without fourth order differential operator). The results and techniques presented in this section will be used in the analysis of a numerical homogenization method in Chapter 7. Note that the study of the effects of numerical integration in the finite element method is essential in numerical homogenization methods (as discussed in [4]). We

refer to Appendix A.3 for an introduction on this topic. Note that the analysis of the finite element method for the multiscale wave equation is postponed to Section 3.1, where we discuss its multiscale character.

The a priori error analysis of the finite element method with numerical integration for the Boussinesq equation follows the technique of elliptic projection (see [44, 21, 22]). Even though the proof is classical, it is not found in the literature. The standard error estimates involve a constant that depends on the domain Ω. Indeed, we verify that it depends on the Poincaré constant for the $\mathrm{L}^\infty(\mathrm{H}^1)$ estimate and on the H^2 regularity constant for the $\mathrm{L}^\infty(\mathrm{L}^2)$ estimate (coming from the Aubin–Nitsche argument). As we want to avoid this dependence to consider pseudoinfinite domains, this issue is settled in Chapter 7, Section 7.1.4. In particular, we modify the elliptic projection and obtain an error estimate in the norm $\|\nabla\cdot\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))}$ (it is a norm on $\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega))$) with a constant independent of Ω.

Let Ω be a periodic hypercube in $\mathbb{R}^d$, let $a^0 \in \mathrm{W}^{1,\infty}(\Omega)$ be a symmetric tensor, elliptic and bounded and let b^2 be a symmetric, positive semidefinite tensor (see the assumptions (2.20) and (2.21)). We consider the Boussinesq equation: $\bar{u} : [0,T^\varepsilon]\times\Omega$ such that

$$\begin{array}{ll} \partial_t^2\bar{u} - \partial_i\big(a^0_{ij}\partial_j\bar{u}\big) - \varepsilon^2\partial_i\big(b^2_{ij}\partial_j\partial_t^2\bar{u}\big) = f & \text{in } (0,T^\varepsilon]\times\Omega, \\ x\mapsto\bar{u}(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,T^\varepsilon], \\ \bar{u}(0,x) = g^0(x), \quad \partial_t\bar{u}(0,x) = g^1(x) & \text{in } \Omega. \end{array} \tag{2.62}$$

Recall the definition of the functional space

$$\mathcal{S}(\Omega) = \big\{v\in\mathrm{L}^2_0(\Omega) : \sqrt{b^2}\nabla v\in[\mathrm{L}^2(\Omega)]^d\big\},$$

and define the bilinear forms

$$\begin{aligned} A^0\big(v,w\big) &= \big(a^0\nabla v,\nabla w\big)_{\mathrm{L}^2}, \\ \big(v,w\big)_{\mathcal{S}} &= \big(v,w\big)_{\mathrm{L}^2(\Omega)} + \varepsilon^2 B^2\big(v,w\big), \\ B^2\big(v,w\big) &= \big(b^2\nabla v,\nabla w\big)_{\mathrm{L}^2}. \end{aligned}$$

If the initial conditions and the right hand side satisfy the regularity (2.50), Theorem 2.1.8 ensures that there exists a unique weak solution $\bar{u}\in\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega))$, with $\partial_t\bar{u}\in\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{S}(\Omega))$ and $\partial_t^2\bar{u}\in\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{S}(\Omega))$, such that

$$\begin{array}{l} \big(\partial_t^2\bar{u}(t),w\big)_{\mathcal{S}} + A^0\big(\bar{u}(t),w\big) = \big(f(t),w\big)_{\mathrm{L}^2} \quad \forall w\in\mathrm{W}_{\mathrm{per}}(\Omega) \quad \text{for a.e. } t\in[0,T^\varepsilon], \\ \bar{u}(0) = g^0, \quad \partial_t\bar{u}(0) = g^1. \end{array} \tag{2.63}$$

Let us define the finite element method with numerical integration for the approximation of $\bar{u}$. Let $\mathcal{T}_H$ be a regular shape regular mesh of Ω with simplicial elements. For an integer $\ell\geq 1$, we define the finite element space

$$V_H(\Omega) = \{v_H\in\mathrm{W}_{\mathrm{per}}(\Omega) : v_H|_K\in\mathcal{P}_\ell(K)\ \forall K\in\mathcal{T}_H\}. \tag{2.64}$$

Let $\{\hat{\omega}_j,\hat{x}_j\}_{j=1}^J$ be the quadrature formula used in the computation of the stiffness matrix. We assume that it satisfies the following hypotheses

$$\begin{array}{ll} \textit{(i)} & \hat{\omega}_j>0, \qquad j=1,\dots,J, \\ \textit{(ii)} & \int_{\hat{K}}\hat{p}(\hat{x})\,\mathrm{d}\hat{x} = \sum_{j=1}^J\hat{\omega}_j\hat{p}(\hat{x}_j) \qquad \forall\hat{p}\in\mathcal{P}^\sigma(\hat{K}),\ \sigma=\max\{2\ell-2,1\}. \end{array} \tag{2.65}$$

We emphasize that (2.65) are the requirements for simplicial elements and that for quadrilaterals they are different (see e.g. [4]). Furthermore, we assume that the quadrature formula $\{\hat{\omega}'_j,\hat{x}'_j\}_{j=1}^{J'}$, required for the computation of the mass matrix, fulfills the following hypothesis

$$\textit{(iii)} \quad \sum_{j=1}^{J'}\hat{\omega}'_j|\hat{p}(\hat{x}'_j)|^2 \geq \hat{\lambda}'\|\hat{p}\|_{\mathrm{L}^2(\hat{K})} \quad \forall\hat{p}\in\mathcal{P}^\ell(\hat{K}),\ \text{for a } \hat{\lambda}'>0. \tag{2.66}$$

We define the following bilinear forms, for $v_H, w_H \in V_H(\Omega)$,

$$\begin{aligned}
A_H^0(v_H, w_H) &= \sum_{K\in\mathcal{T}_H}\sum_{j=1}^{J} \omega_{K_j} a^0(x_{K_j}) v_H(x_{K_j}) w_H(x_{K_j}),\\
\big(v_H, w_H\big)_Q &= \big(v_H, w_H\big)_H + \varepsilon^2 B_H^2(v_H, w_H),\\
\big(v_H, w_H\big)_H &= \sum_{K\in\mathcal{T}_H}\sum_{j=1}^{J'} \omega'_{K_j} v_H(x'_{K_j}) w_H(x'_{K_j}),\\
B_H^2(v_H, w_H) &= \sum_{K\in\mathcal{T}_H}\sum_{j=1}^{J} \omega_{K_j} b^2(x_{K_j}) v_H(x_{K_j}) w_H(x_{K_j}).
\end{aligned}$$

The finite elements approximation of $\bar{u}$ is defined as follows: find $u_H : [0, T^\varepsilon] \to V_H(\Omega)$ such that

$$\begin{aligned}
&\big(\partial_t^2 u_H(t), v_H\big)_Q + A_H^0\big(u_H(t), v_H\big) = \big(f(t), v_H\big)_{\mathrm{L}^2} \quad \forall v_H \in V_H(\Omega) \quad \text{for a.e. } t \in [0,T],\\
&u_H(0) = g_H^0, \quad \partial_t u_H(0) = g_H^1,
\end{aligned} \tag{2.67}$$

where g_H^0, g_H^1 are approximations of the initial conditions in $V_H(\Omega)$.

Let us show the well-posedness of (2.67). Let $\{\varphi_i(x)\}_{i=1}^N$ be a basis of $V_H(\Omega)$ (e.g., the Lagrangian basis) and write the initial conditions and the solution as

$$g_H^k = \sum_{j=1}^N G_j^k(t)\varphi_j(x), \qquad u_H(t,x) = \sum_{j=1}^N U_j(t)\varphi_j(x).$$

We verify that (2.67) is equivalent to the following well-posed second order ODE in $\mathbb{R}^N$:

$$\begin{aligned}
&M\ddot{U}(t) + AU(t) = F(t) \qquad \text{for a.e. } t \in [0, T^\varepsilon],\\
&U(0) = G^0, \quad \dot{U}(0) = G^1,
\end{aligned} \tag{2.68}$$

where

$$A_{ij} = A_H^0(\varphi_j, \varphi_i), \quad M_{ij} = \big(\varphi_j, \varphi_i\big)_Q, \quad \big(F(t)\big)_i = \big(f(t), \varphi_i\big)_{\mathrm{L}^2}.$$

Hence, standard theory ensures the existence and uniqueness of $u_H \in \mathcal{C}^1([0, T^\varepsilon]; V_H(\Omega))$ (see e.g. [38]).

In practice, we need a fully discretized scheme to implement the method. Let us apply the leap frog method for the time discretization of (2.68) (see in Appendix A.5). Consider a uniform discretization of the time interval $[0, T^\varepsilon]$: $t^n = n\Delta t$, $n = 1, \dots, N$, where $\Delta t = T^\varepsilon/N$. The fully discretized method is defined as

$$\begin{aligned}
&V^{n+1/2} = V^{n-1/2} + \Delta t M^{-1}\big(F(t^n) - AU^n\big) && n = 1, \dots, N-1,\\
&U^{n+1} = U^n + \Delta t V^{n+1/2} && n = 0, \dots, N-1,\\
&U^0 = G^0, \qquad V^{1/2} = G^1 + \tfrac{\Delta t}{2}\big(F(0) - AU^0\big).
\end{aligned} \tag{2.69}$$

Observe that at each time iteration, we have to solve a linear system involving the matrix M. As M is sparse, symmetric, positive definite, this can be done with an iterative solver such as the conjugate gradient method. The performance can be improved by computing a Cholesky decomposition of M in a preprocessing step.

We prove the following error estimates for $\bar{u} - u_H$.

Theorem 2.2.1. *Assume that the quadrature formulas satisfy the assumptions* (2.65) *and* (2.66). *Let* $\bar{u}$ *and* u_H *be the solution of* (2.63) *and* (2.67), *respectively.*

i) Assume that $a^0, b^2 \in \mathrm{W}^{\ell,\infty}(\Omega)$ *and* $\partial_t^k \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^{\ell+1}(\Omega))$ *for* $0 \le k \le 4$. *Then the error satisfies* $\|\bar{u} - u_H\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^1(\Omega))} \le \mathrm{e}^{\mathrm{FE}}_{\mathrm{H}^1}$, *where*

$$\begin{aligned}\mathrm{e}^{\mathrm{FE}}_{\mathrm{H}^1} = {} & C_1\big(\|g^1 - g^1_H\|_{\mathrm{H}^1(\Omega)} + \|g^0 - g^0_H\|_{\mathrm{H}^1(\Omega)}\big) \\ & + C_2\big(H^\ell + T^\varepsilon H^{\ell+1} + T^\varepsilon(1+\varepsilon)\varepsilon H^\ell\big)\textstyle\sum_{k=0}^4 \|\partial_t^k \bar{u}\|_{\mathrm{L}^\infty(\mathrm{H}^{\ell+1})},\end{aligned}$$

where C_1, C_2 *are independent of* H *and* ε *but depend on* Ω.

ii) Assume that $a^0 \in \mathrm{W}^{\ell+1,\infty}(\Omega)$, $b^2 \in \mathrm{W}^{\ell,\infty}(\Omega)$ *and* $\partial_t^k \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^{\ell+1}(\Omega))$ *for* $0 \le k \le 3$. *Then the error satisfies* $\|\bar{u} - u_H\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} \le \mathrm{e}^{\mathrm{FE}}_{\mathrm{L}^2}$, *where*

$$\begin{aligned}\mathrm{e}^{\mathrm{FE}}_{\mathrm{L}^2} = {} & C_1\big(\|g^0 - g^0_H\|_{\mathrm{L}^2(\Omega)} + \varepsilon\|g^0 - g^0_H\|_{\mathrm{H}^1(\Omega)} + \|g^1 - g^1_H\|_{\mathrm{L}^2(\Omega)} + \varepsilon\|g^1 - g^1_H\|_{\mathrm{H}^1(\Omega)}\big) \\ & + C_2(1+T^\varepsilon)\big(H^{\ell+1} + \varepsilon H^\ell\big)\textstyle\sum_{k=0}^3 \|\partial_t^k \bar{u}\|_{\mathrm{L}^\infty(\mathrm{H}^{\ell+1})},\end{aligned}$$

where C_1, C_2 *are independent of* H *and* ε *but depend on* Ω.

Note that both estimates depends linearly on the final time T^ε. This dependence indicates that the error accumulates as the time increases. Assuming $\varepsilon, H \le 1$, $T^\varepsilon = \mathcal{O}(\varepsilon^{-2})$, and if the initial conditions are chosen as $g^i_H = I_H g^i$ with I_H as in (2.73), we have (under sufficient regularity of the data)

$$\mathrm{e}^{\mathrm{FE}}_{\mathrm{H}^1} = \mathcal{O}\big(\varepsilon^{-2}H^{\ell+1} + \varepsilon^{-1}H^\ell\big), \quad \mathrm{e}^{\mathrm{FE}}_{\mathrm{L}^2} = \mathcal{O}\big(\varepsilon^{-2}H^{\ell+1} + \varepsilon^{-1}H^\ell\big).$$

Hence, for large timescales and in small domains, the errors in the H^1 and in the L^2 norms have the same asymptotic behavior.

Proof of the a priori error estimates

The proof of Theorem 2.2.1 is divided into three Lemmas. We split the error $\bar{u} - u_H$ as

$$\bar{u} - u_H = (\bar{u} - \pi_H \bar{u}) - (u_H - \pi_H \bar{u}) = \eta - \zeta_H, \tag{2.70}$$

where $\pi_H \bar{u}$ is the elliptic projection defined below. First, we derive estimates for η in the $\mathrm{L}^\infty(\mathrm{L}^2)$ and the $\mathrm{L}^\infty(\mathrm{H}^1)$ norms in Lemma 2.2.2. Second, we estimate ζ_H in the $\mathrm{L}^\infty(\mathrm{H}^1)$ norm in Lemma 2.2.3 and in the $\mathrm{L}^\infty(\mathrm{L}^2)$ norm in Lemma 2.2.4.

Let us first give some preliminary results. In all the proof, C denotes a generic constant that is independent of H, ε, and T^ε. First, the assumption (2.66) on the quadrature formula $\{\hat{\omega}'_j, \hat{x}'_j\}_{j=1}^{J'}$ ensures that $\|v_H\|_H = (v_H, v_H)_H^{1/2}$ is a norm on V_H, equivalent to the L^2 norm independently of H. Hence, as b^2 is positive semidefinite, the norm $\|v_H\|_Q = (v_H, v_H)_Q^{1/2}$ satisfies

$$c_Q\|v_H\|_{\mathrm{L}^2} \le \|v_H\|_Q \le C_Q\big(\|v_H\|_{\mathrm{L}^2} + \varepsilon\|v_H\|_{\mathrm{H}^1}\big), \tag{2.71}$$

for some constants c_Q, C_Q independent of H and ε. Thanks to assumptions (2.65) on $\{\hat{\omega}_j, \hat{x}_j\}_{j=1}^J$, provided sufficient regularity of a^0 and b^2, we the following estimates hold (see Theorems A.3.6 and A.3.9):

$$\begin{aligned}
&|A^0(v_H, w_H) - A^0_H(v_H, w_H)| \le CH^{\ell+\mu} \max_{ij} \|a^0_{ij}\|_{\mathrm{W}^{\ell+\mu,\infty}} \|v_H\|_{\bar{\mathrm{H}}^\ell} \|w_H\|_{\bar{\mathrm{H}}^{1+\mu}}, \\
&|A^0(v_H, w_H) - A^0_H(v_H, w_H)| \le CH \max_{ij} \|a^0_{ij}\|_{\mathrm{W}^{1,\infty}} \|v_H\|_{\mathrm{H}^1} \|w_H\|_{\mathrm{H}^1}, \\
&|B^2(v_H, w_H) - B^2_H(v_H, w_H)| \le CH^\ell \max_{ij} \|b^2_{ij}\|_{\mathrm{W}^{\ell,\infty}} \|v_H\|_{\bar{\mathrm{H}}^\ell} \|w_H\|_{\bar{\mathrm{H}}^1}, \\
&|(v_H, w_H)_{\mathrm{L}^2} - (v_H, w_H)_H| \le CH^{\ell+\mu} \|v_H\|_{\bar{\mathrm{H}}^\ell} \|w_H\|_{\bar{\mathrm{H}}^{1+\mu}},
\end{aligned} \tag{2.72}$$

for any $v_H, w_H \in V_H$ and $\mu = 0, 1$. Furthermore, recall that the projection operator I_H satisfies, for any for $v \in \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^{s+1}(\Omega)$ with $1 \leq s \leq \ell$ (see (A.19)):

$$\Big(\sum_{K \in \mathcal{T}_H} \|v - I_H v\|^2_{\mathrm{H}^m(K)} \Big)^{1/2} \leq C H^{s+1-m} \|v\|_{\mathrm{H}^{s+1}(\Omega)} \quad 0 \leq m \leq s+1. \tag{2.73}$$

Combining the two last estimates in (2.72) with (2.73), we verify that for $v \in \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^{\ell+1}$, $w_H \in V_H$

$$|(v, w_H)_{\mathcal{S}} - (I_H v, w_H)_Q| \leq C\big(H^{\ell+1} + \varepsilon^2 H^{\ell}\big) \|v\|_{\bar{\mathrm{H}}^{\ell+1}} \|w_H\|_{\bar{\mathrm{H}}^2}. \tag{2.74}$$

Let us now define the elliptic projection $\pi_H \bar{u}(t) \in V_H(\Omega)$ as the solution of

$$A^0_H\big(\pi_H \bar{u}(t), v_H\big) = \big(f(t), v_H\big)_{\mathrm{L}^2} - \big(I_H \partial_t^2 \bar{u}(t), v_H\big)_Q \quad \forall v_H \in V_H(\Omega) \text{ for a.e. } t \in [0, T^\varepsilon]. \tag{2.75}$$

As A^0_H is elliptic and bounded, Lax–Milgram theorem ensures that $\pi_H \bar{u}(t)$ exists and is unique for a.e. $t \in [0, T^\varepsilon]$. Moreover, using equation (2.63), we have

$$A^0_H\big(\pi_H \bar{u}(t), v_H\big) = A^0\big(\bar{u}(t), v_H\big) + \big(I_H \partial_t^2 \bar{u}(t), v_H\big)_{\mathcal{S}} - \big(I_H \partial_t^2 \bar{u}(t), v_H\big)_Q,$$

and thus the following estimate is obtained

$$\|\pi_H \bar{u}(t)\|_{\mathrm{H}^1} \leq C\big(\|\bar{u}(t)\|_{\mathrm{H}^1} + \|\partial_t^2 \bar{u}(t)\|_{\mathrm{H}^1}\big) \quad \text{for a.e. } t \in [0, T^\varepsilon]. \tag{2.76}$$

Hence, provided $\partial_t^2 u$ belongs to $\mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^1(\Omega))$, we get $\pi_H \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^1(\Omega))$.

In the three following lemmas, we provide error estimates for $\eta = \bar{u} - u_H$ and $\zeta_H = u_H - \pi_H \bar{u}$ in the $\mathrm{L}^\infty(\mathrm{L}^2)$ and the $\mathrm{L}^\infty(\mathrm{H}^1)$ norms.

Lemma 2.2.2. *Assume that for $1 \leq p \leq \infty$, $\partial_t^k \bar{u}, \partial_t^{k+2} \bar{u} \in \mathrm{L}^p(0, T^\varepsilon; \mathrm{H}^{\ell+1}(\Omega))$ for $k \geq 0$ and that $a^0, b^2 \in \mathrm{W}^{\ell,\infty}(\Omega)$. Then $\partial_t^k \pi_H \bar{u} \in \mathrm{L}^p(0, T^\varepsilon; \mathrm{H}^1(\Omega))$ and the following estimate holds for $\eta = \bar{u} - \pi_H \bar{u}$,*

$$\|I_H \partial_t^k \eta\|_{\mathrm{L}^p(\mathrm{H}^1)} + \|\partial_t^k \eta\|_{\mathrm{L}^p(\mathrm{H}^1)} \leq C H^{\ell} \big(\|\partial_t^k \bar{u}\|_{\mathrm{L}^p(\mathrm{H}^{\ell+1})} + \|\partial_t^{k+2} \bar{u}\|_{\mathrm{L}^p(\mathrm{H}^{\ell+1})}\big), \tag{2.77}$$

where C is independent of $H, \varepsilon, T^\varepsilon$ but depends on the Poincaré constant. If in addition we assume $a^0 \in \mathrm{W}^{\ell+1,\infty}(\Omega)$, then

$$\|I_H \partial_t^k \eta\|_{\mathrm{L}^p(\mathrm{L}^2)} + \|\partial_t^k \eta\|_{\mathrm{L}^p(\mathrm{L}^2)} \leq C\big(H^{\ell+1} + \varepsilon^2 H^{\ell}\big) \big(\|\partial_t^k \bar{u}\|_{\mathrm{L}^p(\mathrm{H}^{\ell+1})} + \|\partial_t^{k+2} \bar{u}\|_{\mathrm{L}^p(\mathrm{H}^{\ell+1})}\big), \tag{2.78}$$

where C is independent of $H, \varepsilon, T^\varepsilon$ but depends on Ω.

Proof. First, note that the forms A^0, $(.,.)_{\mathcal{S}}$, A^0_H and $(.,.)_Q$ are time independent, and hence the time differentiation of equations (2.75) and (2.63) yields, similarly to (2.76), the estimate

$$\|\partial_t^k \pi_H \bar{u}(t)\|_{\mathrm{H}^1} \leq C\big(\|\partial_t^k \bar{u}(t)\|_{\mathrm{H}^1} + \|\partial_t^{k+2} \bar{u}(t)\|_{\mathrm{H}^1}\big) \quad \text{for a.e. } t \in [0, T^\varepsilon].$$

We thus verify that $\partial_t^k \pi_H \bar{u} \in \mathrm{L}^p(0, T^\varepsilon; \mathrm{H}^1(\Omega))$. Second, we prove estimates (2.77) and (2.78) for $k = 0$. The proof for $k > 0$ is obtained in the same way by differentiating (2.75) and (2.63) with respect to t. Using (2.75) and (2.63) we have almost everywhere in $[0, T^\varepsilon]$ and for any $v_H \in V_H(\Omega)$

$$\begin{aligned} A^0_H\big(I_H \eta, v_H\big) = {} & A^0_H\big(I_H \bar{u}, v_H\big) - A^0\big(I_H \bar{u}, v_H\big) + A^0\big(I_H \bar{u} - \bar{u}, v_H\big) \\ & + \big(\partial_t^2 \bar{u} - I_H \partial_t^2 \bar{u}, v_H\big)_{\mathcal{S}} + \big(I_H \partial_t^2 \bar{u}, v_H\big)_{\mathcal{S}} - \big(I_H \partial_t^2 \bar{u}, v_H\big)_Q. \end{aligned}$$

Using (2.72) and (2.73), we obtain for a.e. $t \in [0, T^\varepsilon]$

$$A_H^0\big(I_H\eta(t), v_H\big) \le CH^\ell\big(\|\bar u(t)\|_{\mathrm{H}^{\ell+1}} + \|\partial_t^2\bar u(t)\|_{\mathrm{H}^{\ell+1}}\big)\|v_H\|_{\mathrm{H}^1}.$$

We let $v_H = I_H\eta(t)$ in this inequality and, using the ellipticity of A_H^0 and taking the L^p norm with respect to t, we obtain the estimate for $\|I_H\eta\|_{\mathrm{L}^p(\mathrm{H}^1)}$. As $\eta = \bar u - I_H\bar u + I_H\eta$, estimate (2.77) for $k = 0$ follows, thanks to (2.73). Next, we prove (2.78) using a standard Aubin–Nitsche argument. For a.e. $t \in [0, T^\varepsilon]$, we write

$$\|\eta(t)\|_{\mathrm{L}^2} = \sup_{g\in\mathrm{L}^2(\Omega)} \frac{1}{\|g\|_{\mathrm{L}^2}}\big|\big(\eta(t), g\big)_{\mathrm{L}^2}\big|. \tag{2.79}$$

Let $g \in \mathrm{L}^2(\Omega)$ be fixed and define φ_g as the solution of the elliptic problem $A^0(v, \varphi_g) = (g, v)_{\mathrm{L}^2}\ \forall v \in \mathrm{W}_{\mathrm{per}}(\Omega)$. An elliptic regularity result ensures that $\|\varphi_g\|_{\mathrm{H}^2} \le C\|g\|_{\mathrm{L}^2}$ (thanks to the regularity of a^0 and as the domain Ω is polygonal, see [71]). Using (2.75) and (2.63), we verify that for any $v_H \in V_H$ and a.e. $t \in [0, T^\varepsilon]$

$$\begin{aligned}|A^0\big(\eta(t), \varphi_g\big)| =& |A^0\big(\eta(t), \varphi_g - v_H\big)| + |\big(I_H\partial_t^2\bar u(t), v_H\big)_Q - \big(\partial_t^2\bar u(t), v_H\big)_{\mathcal S}| \\ &+ |A_H^0\big(\pi_H\bar u(t), v_H\big) - A^0\big(\pi_H\bar u(t), v_H\big)|.\end{aligned} \tag{2.80}$$

We bound the last term of the right hand side, using (2.72) and (2.73), as

$$\begin{aligned}|A_H^0\big(\pi_H\bar u, v_H\big) - A^0(\pi_H\bar u, v_H)| \le& |A^0\big(I_H\eta, v_H\big) - A_H^0\big(I_H\eta, v_H\big)| + |A_H^0\big(I_H\bar u, v_H\big) - A^0\big(I_H\bar u, v_H\big)| \\ \le& C\big(H\|I_H\eta\|_{\mathrm{H}^1} + H^{\ell+1}\|\bar u\|_{\mathrm{H}^\ell}\big)\|v_H\|_{\mathrm{H}^2}.\end{aligned} \tag{2.81}$$

In (2.80), we let $v_H = I_H\varphi_g$, so that using (2.73), we have

$$|A^0\big(\eta(t), \varphi_g - I_H\varphi_g\big)| \le \Lambda\|\eta(t)\|_{\mathrm{H}^1}\|\varphi_g - I_H\varphi_g\|_{\mathrm{H}^1} \le CH\|\eta(t)\|_{\mathrm{H}^1}\|\varphi_g\|_{\mathrm{H}^2}. \tag{2.82}$$

We combine then (2.80) with (2.82), (2.74) and (2.81), and we obtain (also using again (2.73))

$$|A^0\big(\eta(t), \varphi_g\big)| \le C\Big(H\|I_H\eta\|_{\mathrm{H}^1} + H\|\eta\|_{\mathrm{H}^1} + H^{\ell+1}\|\bar u\|_{\mathrm{H}^\ell} + \big(H^{\ell+1} + \varepsilon^2H^\ell\big)\|\partial_t^2\bar u\|_{\mathrm{H}^{\ell+1}}\Big)\|\varphi_g\|_{\mathrm{H}^2}.$$

Finally, we use this estimate in (2.79) together with (2.77), the definition of φ_g and the bound $\|\varphi_g\|_{\mathrm{H}^2} \le C\|g\|_{\mathrm{L}^2}$ to prove (2.78) for $k = 0$. The proof of Lemma 2.2.2 is complete. □

Lemma 2.2.3. *The following estimate holds for $\zeta_H = u_H - \pi_H\bar u$,*

$$\begin{aligned}\|\partial_t\zeta_H\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \|\zeta_H\|_{\mathrm{L}^\infty(\mathrm{H}^1)} \le C\big(&\mathrm{e}_{\mathrm{H}^1}^{\mathrm{data}} + \|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|\partial_t\eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \varepsilon\|\partial_t\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} \\ &+ \|I_H\partial_t^2\eta\|_{\mathrm{L}^1(\mathrm{L}^2)} + \varepsilon\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(\mathrm{H}^1)}\big),\end{aligned} \tag{2.83}$$

where $\mathrm{e}_{\mathrm{H}^1}^{\mathrm{data}} = \|g^0 - g_H^0\|_{\mathrm{H}^1} + \|g^1 - g_H^1\|_{\mathrm{L}^2} + \varepsilon\|g^1 - g_H^1\|_{\mathrm{H}^1}$ and C is independent of H, ε and T^ε.

Proof. Using equations (2.67), (2.63) and (2.75), we verify that for any $v_H \in V_H(\Omega)$ and a.e. $t \in [0, T^\varepsilon]$,

$$\begin{aligned}\big(\partial_t^2\zeta_H(t), v_H\big)_Q + A_H^0\big(\zeta_H(t), v_H\big) &= \big(f(t), v_H\big)_{\mathrm{L}^2} - \big(\partial_t^2\pi_H\bar u(t), v_H\big)_Q - A_H^0\big(\pi_H\bar u(t), v_H\big) \\ &= \big(I_H\partial_t^2\eta(t), v_H\big)_Q.\end{aligned} \tag{2.84}$$

We let $v_H = \partial_t\zeta_H(t)$ and use the symmetry of the forms $(\cdot,\cdot)_Q$ and A_H^0 to get for a.e. $t \in [0, T^\varepsilon]$

$$\tfrac12\tfrac{\mathrm{d}}{\mathrm{d}t}\Big(\|\partial_t\zeta_H(t)\|_Q^2 + A_H^0\big(\zeta_H(t), \zeta_H(t)\big)\Big) = \big(I_H\partial_t^2\eta(t), \partial_t\zeta_H(t)\big)_Q.$$

We denote the discrete energy as $E_H\zeta_H(t) = \|\partial_t\zeta_H(t)\|_Q^2 + A_H^0\big(\zeta_H(t),\zeta_H(t)\big)$ and integrate the last equality to get $\forall\xi\in[0,T]$

$$E_H\zeta_H(\xi) = E_H\zeta_H(0) + 2\int_0^\xi \big(I_H\partial_t^2\eta(t),\partial_t\zeta_H(t)\big)_Q\,\mathrm{d}t. \tag{2.85}$$

Applying Cauchy–Schwartz, Hölder, and Young inequalities, we obtain the following bound on the second term of the right hand side

$$2\int_0^\xi \big(I_H\partial_t^2\eta(t),\partial_t\zeta_H(t)\big)_Q\,\mathrm{d}t \le 2\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(Q)}\|\partial_t\zeta_H\|_{\mathrm{L}^\infty(Q)} \le 2\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(Q)}^2 + \tfrac12\|\partial_t\zeta_H\|_{\mathrm{L}^\infty(Q)}^2. \tag{2.86}$$

As $A_H^0\big(\zeta_H(\xi),\zeta_H(\xi)\big)\ge 0$, we obtain successively from (2.85) and (2.86)

$$\begin{aligned} &\tfrac12\|\partial_t\zeta_H\|_{\mathrm{L}^\infty(Q)}^2 \le E_H\zeta_H(0) + 2\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(Q)}^2,\\ &\|\nabla\zeta_H\|_{\mathrm{L}^\infty(\mathrm{L}^2)}^2 \le 2/\lambda\big(E_H\zeta_H(0) + 2\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(Q)}^2\big). \end{aligned} \tag{2.87}$$

where λ is the ellipticity constant of $A_H^0(\cdot,\cdot)$. Note that $\zeta_H(0) = (\bar u - u_H)(0) + \eta(0)$, so that

$$E_H\zeta_H(0) \le \|g_H^1 - g^1\|_Q + \Lambda\|g_H^0 - g^0\|_{\mathrm{H}^1} + \Lambda\|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|\partial_t\eta\|_{\mathrm{L}^\infty(Q)}. \tag{2.88}$$

Combining (2.87), (2.88) and (2.71), we obtain estimate (2.83) and the proof of the lemma is complete. □

Lemma 2.2.4. *The following estimate holds for $\zeta_H = u_H - \pi_H\bar u$,*

$$\|\zeta_H\|_{\mathrm{L}^\infty(\mathrm{L}^2)} \le C\big(\mathrm{e}_{\mathrm{L}^2}^{\mathrm{data}} + \|\eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \varepsilon\|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|I_H\partial_t\eta\|_{\mathrm{L}^1(\mathrm{L}^2)} + \varepsilon\|I_H\partial_t\eta\|_{\mathrm{L}^1(\mathrm{H}^1)}\big), \tag{2.89}$$

where $\mathrm{e}_{\mathrm{L}^2}^{\mathrm{data}} = \|g^0 - g_H^0\|_{\mathrm{L}^2} + \varepsilon\|g^0 - g_H^0\|_{\mathrm{H}^1} + \|I_Hg^1 - g_H^1\|_{\mathrm{L}^2} + \varepsilon\|I_Hg^1 - g_H^1\|_{\mathrm{H}^1}$ and C is independent of H,ε and T^ε.

Proof. We use (2.84) with $v_H = w_H(t)$, where $w_H\in\mathrm{H}^1(0,T^\varepsilon;V_H(\Omega))$, and have almost everywhere in $[0,T^\varepsilon]$

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\big(\partial_t\zeta_H,w_H\big)_Q - \big(\partial_t\zeta_H,\partial_tw_H\big)_Q + A_H^0\big(\zeta_H,w_H\big) = \tfrac{\mathrm{d}}{\mathrm{d}t}\big(\partial_tI_H\eta,w_H\big)_Q - \big(\partial_tI_H\eta,\partial_tw_H\big)_Q.$$

Denoting $e = u - u_H = \eta - \zeta_H$, we rewrite this equality as

$$-\big(\partial_t\zeta_H,\partial_tw_H\big)_Q + A_H^0\big(\zeta_H,w_H\big) = \tfrac{\mathrm{d}}{\mathrm{d}t}\big(\partial_tI_He,w_H\big)_Q - \big(\partial_tI_H\eta,\partial_tw_H\big)_Q. \tag{2.90}$$

For $\xi\in[0,T^\varepsilon]$, we define $\hat w_H(t) = \int_t^\xi \zeta_H(\tau)\mathrm{d}\tau$, which satisfies $\hat w_H\in\mathrm{H}^1(0,T^\varepsilon;V_H(\Omega))$, $\hat w_H(\xi) = 0$ and $\partial_t\hat w_H = -\zeta_H$. We set $w_H = \hat w_H$ in (2.90) and thanks to the symmetry of the forms A_H^0 and $(\cdot,\cdot)_Q$, we get almost everywhere in $[0,T^\varepsilon]$

$$\tfrac12\tfrac{\mathrm{d}}{\mathrm{d}t}\Big(\|\zeta_H\|_Q^2 + A_H^0\big(\hat w_H,\hat w_H\big)\Big) = \tfrac{\mathrm{d}}{\mathrm{d}t}\big(\partial_tI_He,\hat w_H\big)_Q + \big(I_H\partial_t\eta,\zeta_H\big)_Q.$$

We integrate over $[0,\xi]$ and obtain for all $\xi\in[0,T^\varepsilon]$,

$$\|\zeta_H(\xi)\|_Q^2 + A_H^0\big(\hat w_H(0),\hat w_H(0)\big) = \|\zeta_H(0)\|_Q^2 - 2\big(I_H\partial_te(0),\hat w_H(0)\big)_Q + 2\int_0^\xi \big(I_H\partial_t\eta(t),\zeta_H(t)\big)_Q\,\mathrm{d}t. \tag{2.91}$$

The first term of the right hand side is bounded using the triangle inequality as

$$\|\zeta_H(0)\|_Q \le \|\bar u(0) - u_H(0)\|_Q + \|\eta(0)\|_Q \le \|g^0 - g_H^0\|_Q + \|\eta\|_{\mathrm{L}^\infty(Q)}.$$

The second term is bounded using Cauchy-Schwartz and Young inequalities as

$$2\big(I_H\partial_t e(0),\hat{w}_H(0)\big)_Q \leq \tfrac{2C_Q^2}{\lambda_\Omega}\|I_H\partial_t e(0)\|_Q^2 + \tfrac{\lambda_\Omega}{2C_Q^2}\|\hat{w}_H(0)\|_Q^2 \leq \tfrac{2C_Q^2}{\lambda_\Omega}\|I_H\partial_t e(0)\|_Q^2 + \tfrac{\lambda_\Omega}{2}\|\hat{w}_H(0)\|_{\mathrm{H}^1}^2,$$

where C_Q is the constant in (2.71) and $\lambda_\Omega = \lambda/(1+C_\Omega^2)$, where λ is the ellipticity constant of a^0 and C_Ω is the Poincaré constant. The third term is bounded using Cauchy–Schwartz, Hölder, and Young inequality as

$$2\int_0^\xi \big(I_H\partial_t\eta(t),\zeta_H(t)\big)_Q\,\mathrm{d}t \leq 2\|I_H\partial_t\eta\|_{\mathrm{L}^1(Q)}\|\zeta_H\|_{\mathrm{L}^\infty(Q)} \leq 2\|I_H\partial_t\eta\|_{\mathrm{L}^1(Q)} + \tfrac{1}{2}\|\zeta_H\|_{\mathrm{L}^\infty(Q)}^2.$$

Thus, we obtain from the combination of (2.91) with the last three bounds and the ellipticity of $A_H^0(\cdot,\cdot)$

$$\tfrac{1}{2}\|\zeta_H\|_{\mathrm{L}^\infty(Q)}^2 + \tfrac{\lambda_\Omega}{2}\|\hat{w}_H(0)\|_{\mathrm{H}^1}^2 \leq C\big(\|g^0-g_H^0\|_Q^2 + \|I_Hg^1-g_H^1\|_Q^2 + \|\eta\|_{\mathrm{L}^\infty(Q)}^2 + \|I_H\partial_t\eta\|_{\mathrm{L}^1(Q)}^2\big).$$

This estimate and (2.71) implies (2.89) and the proof of the lemma is complete. □

Proof of Theorem 2.2.1. Let $e=\bar{u}-u_H$ and denote the norm $\|v\| = \|\partial_t v\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \|v\|_{\mathrm{L}^\infty(\mathrm{H}^1)}$. Recalling the splitting (2.70), we use the triangle inequality and Lemma 2.2.3 and obtain

$$\|e\| \leq \|\eta\| + \|\zeta_H\| \leq C\big(\mathrm{e}_{\mathrm{H}^1}^{\mathrm{data}} + \|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|\partial_t\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|I_H\partial_t^2\eta\|_{\mathrm{L}^1(\mathrm{L}^2)} + \varepsilon\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(\mathrm{H}^1)}\big).$$

Using Hölder inequality gives

$$\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(\mathrm{L}^2)} + \varepsilon\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(\mathrm{H}^1)} \leq T^\varepsilon\|I_H\partial_t^2\eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + T^\varepsilon\varepsilon\|I_H\partial_t^2\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)},$$

hence, applying Lemma 2.2.2, we obtain the estimate of Theorem 2.2.1 i). Let us prove the second estimate. Using the splitting (2.70), the triangle inequality and (2.71) we have

$$\begin{aligned}\|e\|_{\mathrm{L}^\infty(\mathrm{L}^2)} &\leq \|\eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \|\zeta_H\|_{\mathrm{L}^\infty(\mathrm{L}^2)}\\ &\leq C\big(\mathrm{e}_{\mathrm{L}^2}^{\mathrm{data}} + \|\eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \varepsilon\|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|I_H\partial_t\eta\|_{\mathrm{L}^1(\mathrm{L}^2)} + \varepsilon\|I_H\partial_t\eta\|_{\mathrm{L}^1(\mathrm{H}^1)}\big).\end{aligned}$$

Using Hölder inequality and the L^2 estimate of Lemma 2.2.2, we obtain the estimate of Theorem 2.2.1 ii). The proof of the theorem is complete.

2.3 The spectral method for hyperbolic equations

In this section, we analyze the spectral method for the approximation of the multiscale wave equation. Spectral methods are appropriate numerical methods for the approximation of linear, time dependent PDEs with smooth solutions. Indeed, if the solution is smooth, the method reaches so-called spectral accuracy. However, in the case of the multiscale wave equation, the grid must resolve globally the fine scale to capture the features of the tensor. Hence, the method is extremely costly and can be used only if the tensor is smooth. In this thesis, the only applications where we approximate the multiscale wave equation are either in one dimension or in small two-dimensional domains.

The analysis of the method relies on the interpolation of smooth periodic functions by trigonometric polynomials. For further details on this topic, we refer to Appendix A.4. For the complete theory on the spectral method, we refer to [59, 58, 68, 69, 89, 29, 25, 63] ([91] for the implementation).

The spectral method is judicious for the approximation of the Boussinesq equation, introduced in Section 2.1.2, when the tensors have a spatial variation. In particular, as long as the solution is smooth, the method is capable of handling the fourth order differential operator in space. The analysis of the spectral method for the Boussinesq equation follows the same techniques as for the wave equation. Note that a spectral homogenization method for the long time approximation of wave propagation in locally periodic media is analyzed in Chapter 7, Section 7.2. In particular, the effective model on which the method relies is a Boussinesq equation.

Analysis of the spectral method for the wave equation

Let $\Omega \subset \mathbb{R}^d$ be a periodic hypercube, $\Omega = (a_1, b_1) \times \cdots \times (a_d, b_d)$ and denote F_Ω the bijective affine mapping

$$F_\Omega : (0, 2\pi)^d \to \Omega, \quad \bar{x} \mapsto F_\Omega(\bar{x}) = B_\Omega \bar{x} + a, \tag{2.92}$$

where B_Ω is the diagonal matrix defined as $(B_\Omega)_{jj} = (b_j - a_j)/(2\pi)$. Let $a^\varepsilon \in [\mathrm{L}^\infty_{\mathrm{per}}(\Omega)]^{d\times d}$ be a symmetric, uniformly elliptic and bounded tensor, i.e.,

$$\lambda|\xi|^2 \le a^\varepsilon(x)\xi\cdot\xi \le \Lambda|\xi|^2 \quad \forall \xi \in \mathbb{R}^d \ \text{ for a.e. } x \in \Omega. \tag{2.93}$$

We consider the wave equation: find $u^\varepsilon : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ such that

$$\begin{aligned} &\partial_t^2 u^\varepsilon(t,x) + \nabla_x \cdot \big(a^\varepsilon(x)\nabla_x u^\varepsilon(t,x)\big) = f(t,x) && \text{in } (0,T^\varepsilon]\times\Omega,\\ &x \mapsto u^\varepsilon(t,x) \ \ \Omega\text{-periodic} && \text{in } [0,T^\varepsilon],\\ &u^\varepsilon(0,x) = g^0(x), \quad \partial_t u^\varepsilon(0,x) = g^1(x) && \text{in } \Omega, \end{aligned} \tag{2.94}$$

where g^0, g^1 are given initial conditions and f is a source term. Theorem 2.1.1 ensures that if the data satisfy $g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega)$, $g^1 \in \mathrm{L}^2_0(\Omega)$, and $f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{L}^2_0(\Omega))$, then there exists a unique weak solution $u^\varepsilon \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega))$, with $\partial_t u^\varepsilon \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2_0(\Omega))$, $\partial_t^2 u^\varepsilon \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{W}^*_{\mathrm{per}}(\Omega))$ such that

$$\begin{aligned} &\big\langle \partial_t^2 u^\varepsilon(t), v\big\rangle + A^\varepsilon\big(u^\varepsilon, v\big) = \big(f(t), w\big)_{\mathrm{L}^2(\Omega)} \quad \forall v \in \mathrm{W}_{\mathrm{per}}(\Omega) \quad \text{ for a.e. } t \in [0,T^\varepsilon],\\ &u^\varepsilon(0) = g^0, \ \ \partial_t u^\varepsilon(0) = g^1, \end{aligned} \tag{2.95}$$

where we denoted the dual evaluation $\langle\cdot,\cdot\rangle_{\mathrm{W}^*_{\mathrm{per}}(\Omega),\mathrm{W}_{\mathrm{per}}(\Omega)}$ as $\langle\cdot,\cdot\rangle$ and the bilinear form A^ε is defined as $A^\varepsilon(v,w) = (a^\varepsilon\nabla v, \nabla w)_{\mathrm{L}^2(\Omega)}$.

Let us introduce the spectral method for the approximation of u^ε. For $N \in \mathbb{N}^d_{>0}$, let $h_\nu = (b_\nu - a_\nu)/N_\nu$ and let G_N be the uniform grid of Ω given by

$$G_N = \{x_n = (n_1 h_1, \ldots, n_d h_d) : 0 \le n_\nu \le 2N_\nu - 1\}.$$

We define the space of trigonometric polynomials of order N as (see Appendix A.4.4)

$$\begin{aligned} &V_N(\Omega) = \mathrm{span}(B_N),\\ &B_N = \big\{w_{k_1\cdots k_d}(x) = \textstyle\prod_{\nu=1}^d \bar{w}^\nu_{k_\nu} \circ F_\Omega^{-1}(x) : \bar{w}^\nu_{k_\nu} \in B^1_{N_\nu}\big\},\\ &\text{where} \quad B^1_{N_\nu} = \big\{\bar{w}^\nu_{k_\nu}(\bar{x}) = e^{\mathrm{i}k_\nu\bar{x}} : |k_\nu| \le N_\nu - 1\big\} \cup \big\{\bar{w}^\nu_{N_\nu}(\bar{x}) = \tfrac{1}{2}\big(e^{\mathrm{i}N_\nu\bar{x}} + e^{\mathrm{i}N_\nu\bar{x}}\big)\big\}. \end{aligned}$$

We define the following inner product on $V_N(\Omega)$:

$$(p,q)_N = h^{\mathbb{1}} \sum_{x_n \in G_N} p(x_n)\overline{q(x_n)} = h_1 \sum_{n_1=0}^{2N_1-1} \cdots h_d \sum_{n_d=0}^{2N_d-1} p(x_{n_1\cdots n_d})\overline{q(x_{n_1\cdots n_d})} \quad \forall p,q \in V_N(\Omega),$$

where $h^{\mathbb{1}} = h_1 \cdots h_d$ and $\overline{z}$ denote the complex conjugate of $z \in \mathbb{C}$. The corresponding norm is denoted $\|\cdot\|_N = \sqrt{(\cdot,\cdot)_N}$. We verify that $p \in V_N(\Omega)$ is uniquely determined by its values on the grid G_N and

$$\big(p,q\big)_N = \big(p,q\big)_{\mathrm{L}^2(\Omega)} \qquad \forall p,q \in V_N(\Omega). \tag{2.96}$$

Let $I_N : \mathrm{L}^2_{\mathrm{per}}(\Omega) \to V_N(\Omega)$ be the interpolation operator defined in (A.74). Theorem A.4.4 states that if $v \in \mathrm{L}^2_{\mathrm{per}}(\Omega) \cap \mathrm{H}^s(\Omega)$, for some $s \ge (d+1)/2$, then, for any $\sigma \le s$,

$$\big|v - I_N v\big|_{\mathrm{H}^\sigma(\Omega)} \le C \frac{r(N)^{s-\sigma}}{\big|B_\Omega^{-1}N\big|^{s-\sigma}} |v|_{\mathrm{H}^s(\Omega)}, \tag{2.97}$$

where B_Ω is the matrix in (2.92) and C is a constant depending only on d, s, and $r(N) = N_{\max}/N_{\min}$. Let us introduce the *convolution* of two trigonometric polynomials $p, q \in V_N(\Omega)$ as the unique trigonometric polynomial $p * q \in V_N(\Omega)$ such that $p * q(x_n) = p(x_n)q(x_n)$ for all $x_n \in G_N$ (the name comes from the fact that the coefficients of $p * q$ are given as a finite convolution of the coefficients of p and q). For $b \in \mathrm{L}^\infty_{\mathrm{per}}(\Omega), v \in \mathrm{L}^2_{\mathrm{per}}(\Omega)$, we verify that for all $x_n \in G_N$,

$$I_N b * I_N v(x_n) = I_N b(x_n) I_N v(x_n) = bv(x_n) = I_N(bv)(x_n), \tag{2.98}$$

which implies the equality $I_N b * I_N v = I_N(bv)$. For the approximation of u^ε, we introduce the subspace

$$\mathring{V}_N(\Omega) = V_N(\Omega) \cap \mathrm{W}_{\mathrm{per}}(\Omega),$$

and the corresponding interpolation operator $\mathring{I}_N : \mathrm{L}^2_{\mathrm{per}}(\Omega) \to \mathring{V}_N(\Omega)$, defined in (A.82). Theorem A.4.5 ensures that if $v \in \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^s(\Omega)$, for some $s \geq (d+1)/2$, then for any $\sigma \leq s$,

$$\left|v - \mathring{I}_N v\right|_{\mathrm{H}^\sigma(\Omega)} \leq C \frac{r(N)^{s-\sigma}}{\left|B_\Omega^{-1} N\right|^{s-\sigma}} |v|_{\mathrm{H}^s(\Omega)}, \tag{2.99}$$

where C is a constant depending only on d, s, and $r(N) = N_{\max}/N_{\min}$. Finally, we define the bilinear form $A^\varepsilon_N : \mathring{V}_N(\Omega) \times \mathring{V}_N(\Omega) \to \mathbb{R}$, as

$$A^\varepsilon_N(v_N, w_N) = \left(I_N a^\varepsilon_{ij} * \partial_j v_N, \partial_i w_N\right)_N = h^{\mathbb{1}} \sum_{x_n \in G_N} a^\varepsilon_{ij}(x_n) \partial_j v(x_n) \overline{\partial_i w(x_n)}. \tag{2.100}$$

The spectral method for the approximation of (2.94) is defined as follows: find $u_N : [0, T^\varepsilon] \to \mathring{V}_N(\Omega)$ such that

$$\begin{aligned} &\left(\partial_t^2 u_N(t), v_N\right)_N + A^\varepsilon_N\left(u_N(t), v_N\right) = \left(\mathring{I}_N f(t), v_N\right)_N \quad \forall v_N \in \mathring{V}_N(\Omega) \ \text{ for a.e. } t \in [0, T^\varepsilon], \\ &u_N(0) = \mathring{I}_N g^0, \ \ \partial_t u_N(0) = \mathring{I}_N g^1. \end{aligned} \tag{2.101}$$

To prove the well-posedness of (2.101), we need the following lemma.

Lemma 2.3.1. *The bilinear form A^ε_N is symmetric, elliptic, and bounded. In particular, for all $v_N, w_N \in \mathring{V}_N(\Omega)$*

$$A^\varepsilon_N(v_N, v_N) \geq \lambda \|\nabla v_N\|^2_{\mathrm{L}^2(\Omega)}, \quad A^\varepsilon_N(v_N, w_N) \leq \Lambda \|\nabla v_N\|_{\mathrm{L}^2(\Omega)} \|\nabla w_N\|_{\mathrm{L}^2(\Omega)},$$

where λ and Λ are given in (2.93).

Proof. First, the symmetry of A^ε_N is a direct consequence of the symmetry of a^ε. Next, using the ellipticity of $a^\varepsilon(x)$ and (2.96), we have

$$A^\varepsilon_N(v_N, v_N) \geq \lambda h^{\mathbb{1}} \sum_{x_n \in G_N} |\nabla v_N(x_n)|^2 = \lambda \|\nabla v_N\|^2_{\mathrm{L}^2}.$$

Similarly, the bound on $a^\varepsilon(x)$, the Cauchy–Schwartz inequality, and (2.96) gives

$$A^\varepsilon_N(v_N, w_N) \leq \Lambda \left(h^{\mathbb{1}} \sum_{x_n \in G_N} |\nabla v_N(x_n)|^2\right)^{1/2} \left(h^{\mathbb{1}} \sum_{x_n \in G_N} |\nabla w_N(x_n)|^2\right)^{1/2} = \Lambda \|\nabla v_N\|_{\mathrm{L}^2} \|\nabla w_N\|_{\mathrm{L}^2}.$$

The proof of the lemma is complete. □

Let us show that (2.101) is equivalent to a well-posed second order ODE. Recall that for any $t \in [0, T^\varepsilon]$, the trigonometric polynomial $u_N(t) \in \mathring{V}_N(\Omega)$ is uniquely determined by its values on the grid G_N. Define the following elements of $\mathbb{C}^{2N_1 \times \cdots \times 2N_d}$:

$$(F(t))_n = f(t, x_n), \quad G^i_n = g^i(x_n). \qquad 0 \leq n_\nu \leq 2N_\nu - 1.$$

Furthermore, let D_i be the spectral differentiation map, defined in (A.81) (Appendix A.4.5). We denote the convolution product $A_{ij} * V \in \mathbb{C}^{2N_1 \times \cdots \times 2N_d}$, where $(A_{ij} * V)_n = a^\varepsilon_{ij}(x_n) V_n$. We can then rewrite (2.101) as an evolution equation: $U : [0, T^\varepsilon] \to \mathbb{C}^{2N_1 \times \cdots \times 2N_d}$, where $(U(t))_n = u(t, x_n)$, satisfies

$$\begin{array}{l} \ddot{U}(t) = F(t) + D_i\big(A_{ij} * D_j U(t)\big), \quad \text{for a.e. } t \in [0, T^\varepsilon], \\ U(0) = G^0, \quad \dot{U}(0) = G^1. \end{array} \tag{2.102}$$

Thanks to Lemma 2.3.1, a standard result on ODE ensures that (2.102) is well-posed and we obtain a unique solution of (2.101) $u_N \in \mathcal{C}^1(0, T; \mathring{V}_N(\Omega))$ (see e.g. [38]).

Let us proceed to the time discretization of (2.102) with the leap frog method (see Appendix A.5). Consider a uniform discretization of the time interval $[0, T^\varepsilon]$, $t^k = k\Delta t$, $k = 1, \ldots, K$, where $\Delta t = T^\varepsilon / K$. We obtain the fully discretized method

$$\begin{array}{ll} V^{k+1/2} = V^{k-1/2} + \Delta t\big(F(t^k) + D_i\big(A_{ij} * D_j U^k\big) & k = 1, \ldots, K-1, \\ U^{k+1} = U^k + \Delta t V^{k+1/2} & k = 0, \ldots, K-1, \\ U^0 = G^0, \qquad V^{1/2} = G^1 + \frac{\Delta t}{2}\big(F(0) + D_i\big(A_{ij} * D_j U^0\big). & \end{array} \tag{2.103}$$

An implementation of (2.103) is given in Appendix A.4.7. Observe that at each time iteration, we need to compute $V \mapsto D_i\big(A_{ij} * D_j V\big)$. As discussed in Appendix A.4.5, this can be done using the Fast Fourier Transform algorithm (see [62], [56]). Hence, the construction of the corresponding full matrices is avoided and we can apply the method with large N.

Let us prove that the spectral method (2.101) is stable. Using $\partial_t u_N(t)$ as a test function in (2.101) and thanks to the symmetry of A^ε_N, we obtain for a.e. $t \in [0, T^\varepsilon]$

$$\tfrac{1}{2}\tfrac{\mathrm{d}}{\mathrm{d}t}\Big(\|u_N(t)\|^2_{\mathrm{L}^2} + A^\varepsilon_N\big(u_N(t), u_N(t)\big)\Big) = \big(\mathring{I}_N f(t), \partial_t u_N(t)\big)_{\mathrm{L}^2}.$$

Integrating the equality over $[0, \xi]$, and using Lemma 2.3.1, we get

$$\|\partial_t u_N(\xi)\|^2_{\mathrm{L}^2} + \lambda\|\nabla u_N(\xi)\|^2_{\mathrm{L}^2} \leq \|\mathring{I}_N g^1\|^2_{\mathrm{L}^2} + \Lambda\|\mathring{I}_N g^0\|^2_{\mathrm{H}^1} + 2\int_0^\xi \big(\mathring{I}_N f(t), \partial_t u_N(t)\big)_{\mathrm{L}^2} \mathrm{d}t$$

for any $\xi \in [0, T]$. Using Cauchy–Schwartz, Hölder, and Young inequalities on the last term, we obtain the estimate

$$\tfrac{1}{2}\|\partial_t u_N\|^2_{\mathrm{L}^\infty(\mathrm{L}^2)} + \lambda\|\nabla u_N\|^2_{\mathrm{L}^\infty(\mathrm{L}^2)} \leq \|\mathring{I}_N g^1\|^2_{\mathrm{L}^2} + \Lambda\|\mathring{I}_N g^0\|^2_{\mathrm{H}^1} + 4\|\mathring{I}_N f\|^2_{\mathrm{L}^1(\mathrm{L}^2)}. \tag{2.104}$$

Using the stability of $\mathring{I}_N$ we verify that the method is stable.

We prove the following a priori error estimate.

Theorem 2.3.2. *Let u^ε be the solution of* (2.94) *and u_N its approximation defined by* (2.101). *Assume that the data and the solution satisfy for some $s \geq (d+1)/2$:*

$$a^\varepsilon_{ij} \in \mathrm{W}^{s,\infty}(\Omega),\ g^0 \in \mathrm{H}^{s+1}(\Omega),\ g^1 \in \mathrm{H}^s(\Omega),\ f \in \mathrm{L}^1(0, T^\varepsilon; \mathrm{H}^s(\Omega)),$$
$$u^\varepsilon, \partial_t u^\varepsilon, \partial_t^2 u^\varepsilon \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^{s+1}(\Omega)).$$

Then the following estimate holds for $e = u^\varepsilon - u_N$

$$\|\partial_t e\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \|e\|_{\mathrm{L}^\infty(\mathrm{H}^1)} \leq C\frac{r(N)^{s+1}}{|B_\Omega^{-1}N|^s}\Big(\|g^1\|_{\mathrm{H}^s(\Omega)} + \|g^0\|_{\mathrm{H}^{s+1}(\Omega)} + \|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^s(\Omega))} \\ (1+T^\varepsilon)\|a^\varepsilon\|_{\mathrm{W}^{s,\infty}(\Omega)}\sum_{k=0}^{2}\|\partial_t^k u^\varepsilon\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^{s+1}(\Omega))}\Big), \tag{2.105}$$

where $r(N) = \max_\nu N_\nu / \min_\nu N_\nu$*,* B_Ω *is the matrix in* (2.92)*, and the constant* C *depends only on the Poincaré constant* C_Ω*,* d*,* λ*, and* Λ*.*

Estimate (2.105) ensures the convergence of u_N to u^ε as $N \to \infty$. Observe that the limit has to be taken in all dimension simultaneously, i.e., $N_\nu \to \infty$ at the same time so that $r(N)$ stays bounded (this is rigorously expressed in Remark A.4.1). Even though the method converges, (2.105) cannot be used in applications to choose N and target an order of tolerance for the error. Indeed, as we are in a regime where $\varepsilon \ll 1$, the $\mathrm{L}^\infty(\mathrm{H}^s)$ norms of u^ε and its time derivatives cannot be computed and are probably extremely large quantities (and so is $\|a^\varepsilon\|_{\mathrm{W}^{s,\infty}(\Omega)}$). Nevertheless, we verify that the method is accurate only if $h_\nu \leq \varepsilon$. Indeed, this is the critical value for the grid G_N to capture the frequencies of a^ε. This condition is supported by the presence of B_Ω^{-1} in (2.105), implying that if the domain grows, N must be increased accordingly to keep the same accuracy. In addition, the stability of the method depends strongly on the size h_ν. Indeed, we observe that if h_ν is too large, the approximation of (2.102) is unstable. An instability that cannot be acceptably compensated by the reduction of the time step. Altogether, if the tensor a^ε is smooth, the method gives an accurate approximation for $h_\nu \sim \varepsilon/16$. In one-dimension, the corresponding cost of the method is tolerable (even for large domains). However, in two dimensions, the cost of the method is not affordable in large domains.

Proof of the a priori error estimate

The proof of Theorem 2.3.2 is structured as follows. We split the error as

$$u^\varepsilon - u_N = (u^\varepsilon - \pi_N u) - (u_N - \pi_N u^\varepsilon) = \eta - \zeta_N,$$

where $\pi_N u^\varepsilon$ is the elliptic projection defined below. First, we prove a preliminary result on the approximation of the form A^ε (Lemma 2.3.3). Then, the terms η and ζ_N are estimated separately (Lemmas 2.3.4 and 2.3.5).

Let us define the elliptic projection: for $t \in [0, T^\varepsilon]$, $\pi_N u^\varepsilon(t) \in \mathring{V}_N(\Omega)$ is the solution of the elliptic equation

$$A_N^\varepsilon\big(\pi_N u^\varepsilon(t), v_N\big) = A^\varepsilon\big(u^\varepsilon(t), v_N\big) \quad \forall v_N \in \mathring{V}_N(\Omega). \tag{2.106}$$

Thanks to the properties of the form A_N^ε, in Lemma 2.3.1, Lax–Milgram theorem ensures the existence and uniqueness of $\pi_N u^\varepsilon(t)$.

We prove the following preliminary result on the approximation of A^ε by A_N^ε.

Lemma 2.3.3. *Assume that* $a^\varepsilon \in [\mathrm{W}^{s,\infty}(\Omega)]^{d\times d}$*. Then the bilinear form* A_N^ε *satisfies for any* $v \in \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^{s+1}(\Omega)$*,* $w_N \in \mathring{V}_N(\Omega)$*,*

$$\big|A^\varepsilon\big(v, w_N\big) - A_N^\varepsilon\big(\mathring{I}_N v, w_N\big)\big| \leq C\|a^\varepsilon\|_{\mathrm{W}^{s,\infty}(\Omega)}\frac{r(N)^{s+1}}{|B_\Omega^{-1}N|^s}\|v\|_{\mathrm{H}^{s+1}(\Omega)}\|\nabla w_N\|_{\mathrm{L}^2(\Omega)},$$

where $r(N) = \max_\nu N_\nu / \min_\nu N_\nu$ and the constant C depends s and d.

Proof. As $\mathring{I}_N v = I_N v - \langle I_N v\rangle_\Omega$, it holds $\nabla(\mathring{I}_N v) = \nabla(I_N v)$ and we prove the result for $I_N v$. We thus split the error as

$$\mathrm{e}_{A^\varepsilon} = \left|\left(a_{ij}^\varepsilon \partial_j v, \partial_i w_N\right)_{\mathrm{L}^2} - \left(I_N a_{ij}^\varepsilon * \partial_j(I_N v), \partial_i w_N\right)_N\right| \leq \mathrm{e}^1_{A^\varepsilon} + \mathrm{e}^2_{A^\varepsilon},$$

where

$$\begin{aligned}
\mathrm{e}^1_{A^\varepsilon} &= \left|\left(a_{ij}^\varepsilon \partial_j v, \partial_i w_N\right)_{\mathrm{L}^2} - \left(I_N a_{ij}^\varepsilon * I_N(\partial_j v), \partial_i w_N\right)_N\right|,\\
\mathrm{e}^2_{A^\varepsilon} &= \left|\left(I_N a_{ij}^\varepsilon * I_N(\partial_j v), \partial_i w_N\right)_N - \left(I_N a_{ij}^\varepsilon * \partial_j(I_N v), \partial_i w_N\right)_N\right|.
\end{aligned}$$

Thanks to (2.98), it holds $I_N a_{ij}^\varepsilon * I_N(\partial_j v) = I_N(a_{ij}^\varepsilon \partial_j v)$. Hence, using (2.96) and (2.97), we have

$$\begin{aligned}
\mathrm{e}^1_{A^\varepsilon} &= \left|\left(a_{ij}^\varepsilon \partial_j v - I_N(a_{ij}^\varepsilon \partial_j v), \partial_i w_N\right)_{\mathrm{L}^2}\right|\\
&\leq C\frac{r(N)^{s+1}}{|B_\Omega^{-1}N|^s}\|a_{ij}^\varepsilon \partial_j v\|_{\mathrm{H}^s}\|w_N\|_{\mathrm{H}^1} \leq C\frac{r(N)^{s+1}}{|B_\Omega^{-1}N|^s}\|a^\varepsilon\|_{\mathrm{W}^{s,\infty}}\|v\|_{\mathrm{H}^{s+1}}\|w_N\|_{\mathrm{H}^1}.
\end{aligned}$$

For the second term, the triangle inequality and (2.97) give

$$\begin{aligned}
\mathrm{e}^2_{A^\varepsilon} &\leq \|a_{ij}^\varepsilon\|_{\mathrm{L}^\infty}\|I_N(\partial_j v) - \partial_j(I_N v)\|_{\mathrm{L}^2}\|w_N\|_{\mathrm{H}^1}\\
&\leq \|a_{ij}^\varepsilon\|_{\mathrm{L}^\infty}\left(\|I_N(\partial_j v) - \partial_j v\|_{\mathrm{L}^2} + \|\partial_j v - \partial_j(I_N v)\|_{\mathrm{L}^2}\right)\|w_N\|_{\mathrm{H}^1}\\
&\leq C\frac{r(N)^{s+1}}{|B_\Omega^{-1}N|^s}\|a^\varepsilon\|_{\mathrm{L}^\infty}\|v\|_{\mathrm{H}^{s+1}}\|w_N\|_{\mathrm{H}^1}.
\end{aligned}$$

Combining the estimates for $\mathrm{e}^1_{A^\varepsilon}$ and $\mathrm{e}^2_{A^\varepsilon}$, we obtain the desired bound and the proof of the lemma is complete. □

The next lemma provides an estimate of $\eta = u^\varepsilon - \pi_N u^\varepsilon$.

Lemma 2.3.4. *Assume that $a^\varepsilon \in [\mathrm{W}^{s,\infty}(\Omega)]^{d\times d}$ and that for $p \in [1,\infty]$ and some $k \geq 0$ we have $\partial_t^k u^\varepsilon \in \mathrm{L}^p(0,T^\varepsilon;\mathrm{H}^s(\Omega))$. Then $\partial_t^k \pi_N u^\varepsilon \in \mathrm{L}^p(0,T^\varepsilon;\mathrm{H}^s(\Omega))$ and the following estimate holds*

$$\|\partial_t^k \eta\|_{\mathrm{L}^p(0,T^\varepsilon;\mathrm{H}^1(\Omega))} + \|\mathring{I}_N \partial_t^k \eta\|_{\mathrm{L}^p(0,T^\varepsilon;\mathrm{H}^1(\Omega))} \leq C\frac{r(N)^{s+1}}{|B_\Omega^{-1}N|^s}\|a^\varepsilon\|_{\mathrm{W}^{s,\infty}(\Omega)}\|u^\varepsilon\|_{\mathrm{L}^p(0,T^\varepsilon;\mathrm{H}^{s+1}(\Omega))},$$

where the constant C depends on the Poincaré constant C_Ω, d, s, and λ.

Proof. We prove the result for $k = 0$. The result for $k > 0$ is obtained in the same way, starting from the time differentiations of (2.95) and (2.106). Using (2.95) and (2.106), we have for any $v_N \in \mathring{V}_N(\Omega)$ and a.e. $t \in [0,T^\varepsilon]$,

$$\left|A_N^\varepsilon\left(\mathring{I}_N\eta(t), v_N\right)\right| = \left|A_N^\varepsilon\left(\mathring{I}_N u^\varepsilon(t), v_N\right) - A^\varepsilon\left(u^\varepsilon(t), v_N\right)\right|.$$

Using Lemma 2.3.3 gives

$$\left|A_N^\varepsilon\left(\mathring{I}_N\eta(t), v_N\right)\right| \leq \frac{r(N)^{s+1}}{|B_\Omega^{-1}N|^s}\|u^\varepsilon(t)\|_{\mathrm{H}^{s+1}}\|w_N\|_{\mathrm{H}^1}.$$

The ellipticity of A_N^ε (Lemma 2.3.1) thus implies

$$\lambda/(1+C_\Omega^2)\|\mathring{I}_N\eta(t)\|_{\mathrm{H}^1}^2 \leq \left|A_N^\varepsilon\left(\mathring{I}_N\eta(t), \mathring{I}_N\eta(t)\right)\right| \leq \frac{r(N)^{s+1}}{|B_\Omega^{-1}N|^s}\|u^\varepsilon(t)\|_{\mathrm{H}^{s+1}}\|\mathring{I}_N\eta(t)\|_{\mathrm{H}^1}.$$

Dividing both side of the inequality by $\|\mathring{I}_N\eta(t)\|_{\mathrm{H}^1}$ and taking the L^p norm in time proves the estimate for $\mathring{I}_N\eta$. To obtain the estimate for η, we use the relation $\eta = u^\varepsilon - \mathring{I}_N u^\varepsilon + \mathring{I}_N\eta$ and (2.99). The proof of the lemma is complete. □

Next, the following lemma provides an estimate for $\zeta_N = u_N - \pi_N u^\varepsilon$.

Lemma 2.3.5. *Under the assumptions of Theorem 2.3.2, the following estimate holds*

$$\|\partial_t \zeta_N\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \|\zeta_N\|_{\mathrm{L}^\infty(\mathrm{H}^1)} \leq C \frac{r(N)^{s+1}}{|B_\Omega^{-1} N|^s} \Big(\|g^1\|_{\mathrm{H}^s} + \|g^0\|_{\mathrm{H}^{s+1}} + \|f\|_{\mathrm{L}^1(\mathrm{H}^s)} \Big) + C \Big(\|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|\partial_t \eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \|\partial_t^2 \eta\|_{\mathrm{L}^1(\mathrm{L}^2)} \Big), \tag{2.107}$$

where the constant C depends on d, s, Λ and λ.

Proof. Let us denote the L^2 inner product as $(\cdot,\cdot) = (\cdot,\cdot)_{\mathrm{L}^2(\Omega)}$. Using (2.96) and equations (2.101) and (2.106), we verify that for any $v_N \in \mathring{V}_N(\Omega)$ and a.e. $t \in [0, T^\varepsilon]$,

$$\big(\partial_t^2 \zeta_N(t), v_N\big) + A_N^\varepsilon\big(\zeta_N(t), v_N\big) = \big(\partial_t^2 \eta(t), v_N\big) + \big(\mathring{I}_N f(t) - f(t), v_N\big).$$

Using the test function $v_N = \partial_t \zeta_N(t)$, we obtain

$$\tfrac{1}{2} \tfrac{\mathrm{d}}{\mathrm{d}t} \Big(\|\zeta_N(t)\|_{\mathrm{L}^2}^2 + A_N^\varepsilon\big(\zeta_N(t), \zeta_N(t)\big) \Big) = \big(\partial_t^2 \eta(t) + \mathring{I}_N f(t) - f(t), \partial_t \zeta_N(t)\big).$$

Defining $E_N \zeta(\xi) = \|\partial_t \zeta_N(t)\|_{\mathrm{L}^2}^2 + A_N^\varepsilon\big(\zeta_N(t), \zeta_N(t)\big)$ and integrating over $[0, \xi]$, we find that for any $\xi \in [0, T^\varepsilon]$

$$E_N \zeta(\xi) = E_N \zeta(0) + 2 \int_o^\xi \big(\partial_t^2 \eta(t) + \mathring{I}_N f(t) - f(t), v_N\big) \,\mathrm{d}t. \tag{2.108}$$

Using Cauchy–Schwartz, Hölder and Young inequalities, we bound the second term of the right hand side as

$$2 \int_o^\xi \big(\partial_t^2 \eta(t) + \mathring{I}_N f(t) - f(t), v_N\big) \,\mathrm{d}t \leq 4 \|\partial_t^2 \eta\|_{\mathrm{L}^1(\mathrm{L}^2)}^2 + 4 \|f - \mathring{I}_N f\|_{\mathrm{L}^1(\mathrm{L}^2)}^2 + \tfrac{1}{2} \|\partial_t \zeta_N\|_{\mathrm{L}^\infty(\mathrm{L}^2)}^2. \tag{2.109}$$

Combining this estimate with (2.108), where we take the L^∞ norm with respect to ξ, and because $A_N\big(\zeta_N(\xi), \zeta_N(\xi)\big) \geq 0$, we get

$$\tfrac{1}{2} \|\partial_t \zeta_N\|_{\mathrm{L}^\infty(\mathrm{L}^2)}^2 \leq E_N \zeta_N(0) + 4 \|\partial_t^2 \eta\|_{\mathrm{L}^1(\mathrm{L}^2)}^2 + 4 \|f - \mathring{I}_N f\|_{\mathrm{L}^1(\mathrm{L}^2)}^2.$$

Using (2.108) and (2.109) with the ellipticity of A_N^ε and the last estimate, we obtain

$$\lambda \|\nabla \zeta_N\|_{\mathrm{L}^\infty(\mathrm{L}^2)}^2 \leq 2 E_N \zeta_N(0) + 8 \|\partial_t^2 \eta\|_{\mathrm{L}^1(\mathrm{L}^2)}^2 + 8 \|f - \mathring{I}_N f\|_{\mathrm{L}^1(\mathrm{L}^2)}^2.$$

We bound the term $E_N \zeta_N(0)$ using the equality $\zeta_N = \eta - e$, where $e = u^\varepsilon - u_N$, the triangle inequality, and Lemma 2.3.1:

$$\begin{aligned} E_N \zeta_N(0) &\leq \|\partial_t e(0)\|_{\mathrm{L}^2}^2 + \|\partial_t \eta(0)\|_{\mathrm{L}^2}^2 + \Lambda \|e(0)\|_{\mathrm{H}^1}^2 + \Lambda \|\eta(0)\|_{\mathrm{H}^1}^2 \\ &\leq \|g^1 - \mathring{I}_N g^1\|_{\mathrm{L}^2}^2 + \|\partial_t \eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)}^2 + \Lambda \|g^0 - \mathring{I}_N g^0\|_{\mathrm{H}^1}^2 + \Lambda \|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}^2. \end{aligned}$$

Estimate (2.107) is obtained by the combination of the three last estimates, (2.99), and the Poincaré inequality. □

Proof of Theorem 2.3.2. Recall that $e = u^\varepsilon - u_N = \eta - \zeta_N$. Applying the triangle inequality and Lemma 2.3.5, we have

$$\begin{aligned} \|\partial_t e\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \|e\|_{\mathrm{L}^\infty \mathrm{H}^1} &\leq \|\partial_t \eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \|\eta\|_{\mathrm{L}^\infty \mathrm{H}^1} + \|\partial_t \zeta_N\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \|\zeta_N\|_{\mathrm{L}^\infty \mathrm{H}^1} \\ &\leq C r(N)^{s+1} |B_\Omega^{-1} N|^{-s} \big(\|g^1\|_{\mathrm{H}^s} + \|g^0\|_{\mathrm{H}^{s+1}} + \|f\|_{\mathrm{L}^1(\mathrm{H}^s)} \big) \\ &\quad + C \big(\|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|\partial_t \eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \|\partial_t^2 \eta\|_{\mathrm{L}^1(\mathrm{L}^2)} \big). \end{aligned}$$

Hölder inequality implies $\|\partial_t^2 \eta\|_{\mathrm{L}^1(\mathrm{L}^2)} \leq T^\varepsilon \|\partial_t^2 \eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)}$, and, applying Lemma 2.3.4, we have

$$\|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|\partial_t \eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + T^\varepsilon \|\partial_t^2 \eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} \leq C \frac{r(N)^{s+1}}{|B_\Omega^{-1} N|^s} (1 + T^\varepsilon) \|a^\varepsilon\|_{\mathrm{W}^{s,\infty}} \sum_{k=0}^{2} \|\partial_t^k u^\varepsilon\|_{\mathrm{L}^\infty(\mathrm{H}^{s+1})}.$$

Combining the three estimates, we obtain (2.105) and the proof of the theorem is complete. □

2.4 Fourier method for constant coefficients hyperbolic equations

When the tensors in the Boussinesq equation (Section 2.1.2) are constant, an explicit form of the solution is available in the Fourier basis. In this section, we take advantage of this explicit form to derive a Fourier method for the approximation of this equation and perform its a priori error analysis. In particular, the method does not require a discretization in time and is extremely accurate when the data are smooth. Note that the method and its analysis rely on the interpolation of smooth periodic function by trigonometric polynomials, which is analyzed in Appendix A.4.

Let $\Omega \subset \mathbb{R}^d$ be a periodic hypercube, $\Omega = (a_1, b_1) \times \cdots \times (a_d, b_d)$ and denote F_Ω the bijective affine mapping

$$F_\Omega : (0, 2\pi)^d \to \Omega, \quad \bar{x} \mapsto F_\Omega(\bar{x}) = B_\Omega \bar{x} + a, \tag{2.110}$$

where B_Ω is the diagonal matrix defined as $(B_\Omega)_{jj} = (b_j - a_j)/(2\pi)$. We consider the Boussinesq equation from Section 2.1.2 with constant coefficients: $\tilde{u} : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ such that

$$\begin{array}{ll} \partial_t^2 \tilde{u} + a^0_{ij} \partial^2_{ij} \tilde{u} - \varepsilon^2 b^2_{ij} \partial^2_{ij} \partial^2_t u + \varepsilon^2 a^2_{ijkl} \partial^4_{ijkl} \tilde{u} = 0 & \text{in } (0, T^\varepsilon] \times \Omega, \\ x \mapsto \tilde{u}(t, x) \ \ \Omega\text{-periodic} & \text{in } [0, T^\varepsilon], \\ \tilde{u}(0, x) = g^0(x), \quad \partial_t \tilde{u}(0, x) = g^1(x) & \text{in } \Omega. \end{array} \tag{2.111}$$

The tensors a^0, b^2, a^2 are constant and satisfy for some $\lambda, \Lambda > 0$:

$$\begin{array}{ll} a^0_{ij} = a^0_{ji}, & \lambda |\xi|^2 \le a^0 \xi \cdot \xi \le \Lambda |\xi|^2 \quad \forall \xi \in \mathbb{R}^d, \\ b^2_{ij} = b^2_{ji}, & b^2 \xi \cdot \xi \ge 0 \quad \forall \xi \in \mathbb{R}^d, \\ a^2_{ijkl} = a^2_{klij}, & a^2 \eta : \eta \ge 0 \quad \forall \eta \in \mathrm{Sym}^2(\mathbb{R}^d), \end{array} \tag{2.112}$$

where we recall the notation

$$a^2 \eta : \xi = a^2_{ijkl} \eta_{kl} \xi_{ij} \quad \forall \eta, \xi \in \mathrm{Sym}^2(\mathbb{R}^d).$$

Referring to Theorem 2.1.5, if the data satisfy $g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^2(\Omega), g^1 \in \mathrm{W}_{\mathrm{per}}(\Omega), f \in \mathrm{L}^2(0, T^\varepsilon; \mathrm{L}^2_0(\Omega))$, then there exists a unique weak solution of (2.111). In particular, $\tilde{u}$ belongs to $\mathrm{L}^\infty(0, T^\varepsilon; \mathrm{W}_{\mathrm{per}}(\Omega))$.

Let us formally find an explicit form for $\tilde{u}$ in the Fourier basis (see Appendix A.4.1). Recall that $v \in \mathrm{L}^2_{\mathrm{per}}(\Omega)$ can be expanded in the Fourier basis as

$$v(x) \overset{\mathrm{L}^2}{=} \sum_{k \in \mathbb{Z}^d} \hat{v}(k) e^{\mathrm{i} k \cdot F_\Omega^{-1}(x)}, \qquad \hat{v}(k) = \frac{1}{|\Omega|} \int_\Omega v(x) e^{-\mathrm{i} k \cdot F_\Omega^{-1}(x)} \, \mathrm{d}x.$$

The initial conditions can thus be expressed as $g^i(x) \overset{\mathrm{L}^2}{=} \sum_{k \in \mathbb{Z}^d} \widehat{g^i}(k) e^{\mathrm{i} k \cdot F_\Omega^{-1}(x)}$. As the map $x \mapsto \tilde{u}(t, x)$ is Ω-periodic, we look for a solution of the form

$$\tilde{u}(t, x) \overset{\mathrm{L}^2}{=} \sum_{\mathbb{Z}^d \setminus \{0\}} \hat{u}(t, k) e^{\mathrm{i} k \cdot F_\Omega^{-1}(x)}, \tag{2.113}$$

where we used that $\tilde{u} \in \mathrm{W}_{\mathrm{per}}(\Omega)$ implies $\hat{u}(t, k = 0) = \langle \tilde{u}(t) \rangle_\Omega = 0$. Inserting this ansatz in equation (2.111), we obtain for any $k \in \mathbb{Z}^d \setminus \{0\}$, the second order ODE

$$\begin{array}{l} \frac{\mathrm{d}^2}{\mathrm{d}t^2} \hat{u}(t, k) = -p(k) \hat{u}(t, k) \qquad t \in (0, T^\varepsilon], \\ \hat{u}(0, k) = \widehat{g^0}(k), \quad \frac{\mathrm{d}}{\mathrm{d}t} \hat{u}(0, k) = \widehat{g^1}(k), \end{array} \tag{2.114}$$

where, defining $k_\Omega = B_\Omega^{-1} k$,

$$p(k) = \frac{a^0 k_\Omega \cdot k_\Omega + \varepsilon^2 a^2 k_\Omega k_\Omega^T : k_\Omega k_\Omega^T}{1 + \varepsilon^2 b^2 k_\Omega \cdot k_\Omega}.$$

Note that the assumptions on the tensors (2.112) ensures that $p(k) > 0$ for any $k \in \mathbb{Z}^d \backslash \{0\}$. Consequently, (2.114) admits the unique solution

$$\hat{u}(t,k) = \widehat{g^0}(k) \cos\Big(\sqrt{p(k)}t\Big) + \frac{\widehat{g^1}(k)}{\sqrt{p(k)}} \sin\Big(\sqrt{p(k)}t\Big). \tag{2.115}$$

We can show that $\tilde{u}$, defined in (2.113) with the coefficients in (2.115), is the unique solution of (2.111).

With the tools introduced in Appendix A.4.4, we can define an efficient numerical method for the approximation of $\tilde{u}$. For $N \in \mathbb{N}_{>0}^d$, let $h_\nu = (b_\nu - a_\nu)/N_\nu$ and let G_N be the uniform grid of Ω given by

$$G_N = \big\{x_n = (n_1 h_1, \dots, n_d h_d) : 0 \le n_\nu \le 2N_\nu - 1\big\}.$$

The approximation is defined for any $t \in [0, T^\varepsilon]$ as

$$\begin{aligned}
&u_N(t,x) = \sum_{|k_1| \le N_1}{}' \cdots \sum_{|k_d| \le N_d}{}' \hat{u}_k(t) e^{\mathrm{i} k \cdot F_\Omega^{-1}(x)}, \qquad (2.116)\\
&\hat{u}_k(t) = \widehat{I_N g^0}(k) \cos\Big(\sqrt{p(k)}t\Big) + \frac{\widehat{I_N g^1}(k)}{\sqrt{p(k)}} \sin\Big(\sqrt{p(k)}t\Big) \quad k \in \big\{|k_\nu| \le N_\nu,\ 1 \le \nu \le d\big\} \backslash \{0\},\\
&\hat{u}_0(t) = 0,
\end{aligned}$$

where I_N and its coefficients are defined for $v \in \mathrm{L}^2_{\mathrm{per}}(\Omega)$ as (see (A.74))

$$\begin{aligned}
&I_N v(x) = \sum_{|k_1| \le N_1}{}' \cdots \sum_{|k_d| \le N_d}{}' \hat{v}_{k_1 \cdots k_d} e^{\mathrm{i} k \cdot F_\Omega^{-1}(x)},\\
&\hat{v}_{k_1 \cdots k_d} = \frac{1}{2N_1} \sum_{n_1=0}^{2N_1-1} \cdots \frac{1}{2N_d} \sum_{n_d=0}^{2N_d-1} v(x_{n_1 \cdots n_d}) e^{-i k_1 n_1 \bar{h}_1} \cdots e^{-i k_d n_d \bar{h}_d},
\end{aligned}$$

where the notation $\sum'$ indicates that the terms $k_\nu \in \{-N_\nu, N_\nu\}$ are halved. We emphasize that the method (2.116) is explicit in time. In particular, no time discretization is needed, which represents a huge saving of computational time. Furthermore, the coefficients $\widehat{I_N g^i}(k)$ can be computed with a Fast Fourier Transform (FFT) algorithm and the value of $u_N(t,\cdot)$ on the grid G_N are computed with an inverse FFT algorithm. An implementation of the method (2.116) for a two-dimensional example is provided in Appendix A.4.7 (see also Appendix A.4.5). Note that the data used in the implementation are defined in Section 4.4.3, in the context of the long time homogenization of the wave equation in periodic media.

We prove the following a priori error estimate for the Fourier method (2.116).

Theorem 2.4.1. *Assume that $\tilde{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{W}_{\mathrm{per}}(\Omega)) \cap \mathrm{H}^s(\Omega))$, $g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^s(\Omega)$ and $g^1 \in \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^s(\Omega)$ for some $s \ge (d+1)/2$. Then, for any $t \in [0, T^\varepsilon]$ and $\sigma \le s$,*

$$|\tilde{u}(t) - u_N(t)|_{\mathrm{H}^\sigma(\Omega)} \le C \frac{r(N)^{s-\sigma}}{|B_\Omega^{-1} N|^{s-\sigma}} \Big(|\tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^s(\Omega))} + |g^0|_{\mathrm{H}^s(\Omega)} + |g^1|_{\mathrm{H}^s(\Omega)}\Big),$$

where $r(N) = \max_\nu N_\nu / \min_\nu N_\nu$ and C depends on $r(N)$, d, s, a^0 and b^2.

Proof. The proof is done in the context introduced in Appendix A.4.4. In particular, it is similar to the proof of Theorem A.4.4. We prove the result in the case $\Omega = (0, 2\pi)^d$. The proof in the general case follows the same line by replacing k with $k_\Omega = B_\Omega^{-1} k$. We split the $\tilde{\mathrm{H}}^\sigma$ norm (see (A.56)) of the error as in (A.66)

$$|u(t) - u_N(t)|^2_{\tilde{\mathrm{H}}^\sigma} \leq \sum_{k \in K_\geq} |k|^{2\sigma} |\hat{u}(t,k)|^2 + \sum_{k \in K_\leq} |k|^{2\sigma} |\hat{u}(t,k) - \widehat{u_N}(t,k)|^2 =: E_1 + E_2,$$

where the sets of indices are given as

$$K_\geq = \big\{k \in \mathbb{Z}^d : |k_\nu| \geq N_\nu \text{ for at least one } \nu\big\}, \quad K_\leq = \big\{k \in \mathbb{Z}^d : |k_\nu| \leq N_\nu \text{ for all } \nu\big\}.$$

For the first term, similarly as in (A.78), we obtain

$$E_1 = \sum_{k \in K_\geq} \frac{1}{|k|^{2(s-\sigma)}} |k|^{2s} |\hat{u}(t,k)|^2 \leq C \frac{r(N)^{2(s-\sigma)}}{|N|^{2(s-\sigma)}} |u(t)|^2_{\tilde{\mathrm{H}}^s(\Omega)}.$$

For the second term, note that thanks to (2.112), we have $1/p(k) \leq C$ for some constant C depending on a^0 and b^2. Hence,

$$|\hat{u}(t,k) - \widehat{u_N}(t,k)|^2 \leq |\widehat{g^0}(k) - \widehat{I_N g^0}(k)|^2 + C_r |\widehat{g^1}(k) - \widehat{I_N g^1}(k)|^2,$$

and thus using Theorem A.4.4 and (A.55), we obtain

$$E_2 \leq |g^0 - I_N g^0|^2_{\tilde{\mathrm{H}}^\sigma} + C_r |g^1 - I_N g^1|^2_{\tilde{\mathrm{H}}^\sigma} \leq C \frac{r(N)^{2(s-\sigma)}}{|N|^{2(s-\sigma)}} \big(\|g^0\|_{\mathrm{H}^s} + \|g^1\|_{\mathrm{H}^s}\big).$$

Combining the estimates for E_1 and E_2 gives the proof of the theorem. □

3 Homogenization theory and multiscale methods for the wave equation

In this chapter, we discuss the approximation of the multiscale wave equation. We consider the solution $u^\varepsilon : [0,T] \times \Omega \to \mathbb{R}$ of

$$\partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \big(a^\varepsilon(x) \nabla_x u^\varepsilon(t,x)\big) = f(t,x) \quad \text{in } (0,T] \times \Omega, \tag{3.1}$$

where we impose periodic boundary conditions and initial conditions on $u^\varepsilon(0,x), \partial_t u^\varepsilon(0,x)$. In the multiscale regime, the wavelengths of the initial data and of the source term are macroscopic $\mathcal{O}(1)$, while the tensor varies at the microscopic scale $\mathcal{O}(\varepsilon)$, $\varepsilon \ll 1$. As we will see, approximating (3.1) with standard numerical methods such as the finite element method (FEM) or the finite difference method (FDM) leads to an expensive cost that becomes prohibitive as $\varepsilon \to 0$. Indeed, for such method to attain a tolerable accuracy, the discretization must resolve the medium at the microscopic scale $\mathcal{O}(\varepsilon)$ in the whole domain Ω. Hence, more sophisticated methods that do not require scale resolution are needed. The methods available in the literature for the approximation of (3.1) are reviewed in Section 3.1.

To elaborate more sophisticated numerical methods, one possibility is to look for a function that describes well the effective behavior of u^ε, i.e., its macroscopic behavior without the variations occurring at the microscopic scale. We call such function an effective solution. The homogenization theory is the study of effective equations: it deals with the existence and uniqueness of effective solutions and, in certain cases, provides formulas for their computation. A vast literature is available on the homogenization of elliptic operators for which several different mathematical techniques are used. The basics can be found in [24, 84, 81, 78, 66, 37, 17], with a focus on periodic homogenization, essential in this thesis and introduced in Section 3.3. For the homogenization of elliptic operators driven by general symmetric tensors, the theory on G-convergence is used to prove the existence of effective equations [87, 41] (see Section 3.2). This theory was generalized to non necessarily symmetric tensors and called H-convergence in [76, 90]. Finally, let us cite the two-scale convergence method [16] and the homogenization by unfolding [35, 36], which are widely used in homogenization processes.

Among the homogenization processes, we can dissociate two approaches. The first one is to study the problem at the limit $\varepsilon \to 0$. In fact, the majority of the available homogenization results are obtained through a limiting process. The second approach, is to fix $\varepsilon > 0$ and prove error estimates between the solution u^ε and an effective solution. This approach is less general as it naturally requires stronger assumptions on the data and on the structure of the tensor. Asymptotic expansions, which are extensively used for homogenization processes in periodic media, goes in the direction of this second approach. In this chapter, both approaches are introduced in the context of the wave equation. Nevertheless, we focus on asymptotic expansions and adaptation techniques, as these tools will be imperative in the derivation of long time effective models in Chapters 4, 5, and 6.

The chapter is organized as follows. First, in Section 3.1, we discuss the numerical approximation of the multiscale wave equation (3.1). In particular, we review the multiscale methods that are available in the literature. Then, in Section 3.2, we present the general homogenization result for the wave equation, obtained via G-convergence. Furthermore, we discuss a particularity of the wave equation connected to the convergence of the associated energy. In Section 3.3, we introduce the technique of asymptotic expansions and how to use it to prove rigorous error estimates. Finally, in Section 3.4, we present the finite element heterogeneous method (FE-HMM), which is a numerical homogenization method that will be adapted to the long time approximation of the wave equation in Chapter 7.

3.1 Numerical approximation of the wave equation in heterogeneous media

In this section, we discuss the numerical approximation of the multiscale wave equation. In particular, we justify why standard numerical methods can not be used and review the multiscale methods available in the literature.

Let a^ε be a symmetric, elliptic and bounded tensor and consider $u^\varepsilon : [0,T] \times \Omega \to \mathbb{R}$ such that

$$\begin{aligned} &\partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \big(a^\varepsilon(x) \nabla_x u^\varepsilon(t,x)\big) = f(t,x) && \text{in } (0,T] \times \Omega, \\ &x \mapsto u^\varepsilon(t,x) \ \ \Omega\text{-periodic} && \text{in } [0,T], \\ &u^\varepsilon(0,x) = g^0(x), \quad \partial_t u^\varepsilon(0) = g^1(x) && \text{in } \Omega, \end{aligned} \tag{3.2}$$

where the data are such that the equation is well-posed (see Section 2.1.1). We assume that we are in a multiscale setting, i.e., the initial data and f have wavelengths of order $\mathcal{O}(1)$ and a^ε varies at the scale $\mathcal{O}(\varepsilon)$, with $\varepsilon \ll 1$. Intuitively, the multiscale character of (3.2) is a problem for its numerical approximation. Indeed, a standard numerical method, such as the finite element method or the finite difference method, accurately approximates u^ε only if the discretization of Ω is sufficiently fine to capture the microscopic features of a^ε. Hence, such methods have an expensive cost, which is prohibitive as $\varepsilon \to 0$ or if the domain grows.

Let us now follow [11] and mathematically justify why the finite element method is not suited to approximate (3.2). Referring to Section 2.1.1, let us assume that the weak solution of (3.2) $u^\varepsilon \in \mathrm{L}^\infty(0,T;\mathrm{W}_{\mathrm{per}}(\Omega))$ satisfies the variational formulation

$$\begin{aligned} &\langle \partial_t^2 u^\varepsilon(t), v\rangle + A^\varepsilon\big(u^\varepsilon(t), v\big) = \big(f(t), v\big)_{\mathrm{L}^2} \quad \forall v \in \mathrm{W}_{\mathrm{per}}(\Omega) \quad \text{ for a.e. } t \in [0,T], \\ &u(0) = g^0, \quad \partial_t u(0) = g^1, \end{aligned} \tag{3.3}$$

where A^ε is the bilinear form defined by

$$A^\varepsilon : \mathrm{W}_{\mathrm{per}}(\Omega) \times \mathrm{W}_{\mathrm{per}}(\Omega) \to \mathbb{R}, \quad (v,w) \mapsto A^\varepsilon(v,w) = (a^\varepsilon \nabla v, \nabla w)_{\mathrm{L}^2}.$$

Let $V_h \subset \mathrm{W}_{\mathrm{per}}(\Omega)$ be a finite element space (see Appendix A.3) and consider the approximation of u^ε defined as the unique solution of

$$\begin{aligned} &\big(\partial_t^2 u_h(t), v_h\big)_{\mathrm{L}^2} + A^\varepsilon\big(u_h(t), v_h\big) = \big(f(t), v_h\big)_{\mathrm{L}^2} \quad \forall v_h \in V_h \quad \text{ for a.e. } t \in [0,T], \\ &u_h(0) = g_h^0, \quad \partial_t u_h(0) = g_h^1. \end{aligned} \tag{3.4}$$

In order to derive estimates for $u^\varepsilon - u_h$, we introduce the Riesz projection onto V_h: $\pi_h u^\varepsilon : [0,T] \to V_h$, where $\pi_h u^\varepsilon(t) \in V_h$ solves

$$A^\varepsilon\big(\pi_h u^\varepsilon(t), v_h\big) = A^\varepsilon\big(u^\varepsilon(t), v_h\big) \quad \forall v_h \in V_h.$$

We verify that if $\partial_t^k u^\varepsilon \in \mathrm{L}^\infty(0,T;\mathrm{H}^1(\Omega))$, for $k \geq 0$, then the projection satisfies $\pi_h \partial_t^k u^\varepsilon \in \mathrm{L}^\infty(0,T;\mathrm{H}^1(\Omega))$ and for any $t \in [0,T]$, the following estimate holds for $\eta = u^\varepsilon - \pi_h u^\varepsilon$:

$$\|\partial_t^k \eta(t)\|_{\mathrm{H}^1} \leq C \inf_{v_h \in V_h} \|\partial_t^k u^\varepsilon(t) - v_h\|_{\mathrm{H}^1}. \tag{3.5}$$

Then, following the analysis from [44], we verify that the error satisfies

$$\|u^\varepsilon - u_h\|_{\mathrm{L}^\infty(\mathrm{H}^1)} \leq C\big(\|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|\partial_t\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + T\|\partial_t^2\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}\big), \tag{3.6}$$

where we assumed the best case scenario where the initial conditions in (3.4) are $g_h^0 = \pi_h g^0$, $g_h^1 = \pi_h g^1$. Combining (3.5) with (3.6), we conclude that for u_h to approximate accurately u^ε in $\mathrm{H}^1(\Omega)$, we need $\inf_{v_h\in V_h}\|\partial_t^k u^\varepsilon(t) - v_h\|_{\mathrm{H}^1}$ to be small for $k = 0, 1, 2$. This requirement means that V_h must be able to describe u^ε and its time derivatives in the H^1 norm. If V_h is a standard piecewise polynomial finite element space, we thus need $h \sim \varepsilon$. This conclusion agrees with the instinctive fact that to capture the gradient ∇u^ε, the finite element space must be able to describe its variations at the microscopic scale $\mathcal{O}(\varepsilon)$. Assuming that the size of the domain is of order $\mathcal{O}(1)$, the computational cost of the method is then of order $\mathcal{O}(\varepsilon^{-d})$. As ε is a small quantity, the cost of the FEM is extremely large. Let us then verify that the error estimate in the $\mathrm{L}^\infty(\mathrm{L}^2)$ norm does not change this conclusion. Assuming that $\|a^\varepsilon\|_{\mathrm{W}^{1,\infty}} \leq C\varepsilon^{-1}$, Aubin–Nitsche duality argument implies

$$\|\partial_t^k\eta(t)\|_{\mathrm{L}^2} \leq C\|a^\varepsilon\|_{\mathrm{W}^{1,\infty}}h\|\partial_t^k\eta(t)\|_{\mathrm{H}^1} \leq C\frac{h}{\varepsilon}\inf_{v_h\in V_h}\|\partial_t^k u^\varepsilon(t) - v_h\|_{\mathrm{H}^1}, \tag{3.7}$$

and, following [21], the error satisfies the following optimal error estimate

$$\|u^\varepsilon - u_h\|_{\mathrm{L}^\infty(\mathrm{L}^2)} \leq C\big(\|\eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + T\|\partial_t\eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)}\big). \tag{3.8}$$

Hence, (3.7) and (3.8) leads to the same conclusion as (3.6): to ensure a small $\mathrm{L}^\infty(\mathrm{L}^2)$ error we need $h \sim \varepsilon$ and again the corresponding cost is of order $\mathcal{O}(\varepsilon^{-d})$. We conclude that to reach an acceptable accuracy, the FEM requires extremely large computational resources. Hence, to approximate the wave equation (3.2), more sophisticated numerical methods are needed.

Literature review on multiscale methods for the wave equation

Let us review the multiscale methods available in the literature to approximate the wave equation in heterogeneous media (3.2) at short time. We refer to [11] for a more detailed survey. The methods can be divided in two categories and it is the physical nature of the problem that motivates their selection. The determining criterion is whether the medium has or not scale separation. The medium has scale separation, if the involved scales can clearly be distinguished. Such structure mainly happens in the study of composite materials, or in other material science applications, where the medium is artificially designed. It can also concern geoscience, when the medium is fictional (i.e., not from natural data). The medium does not have scale separation, if it involves a continuum of scales. This is the case when the medium is a natural datum. For example, in geoscience, when considering the ground or in medical imaging, when considering the human body. The methods suited for problems with scale separation derive from homogenization results. They are cheaper but less general than the methods suited for problems without scale separation. Indeed, on the one hand they make use of the specific structures of the medium to reduce the cost of approximation, while on the second hand if the structure is not respected they provide an approximation of poor accuracy. Let us describe in some more details the numerical methods that are available for the short time approximation of (3.2).

We begin with the methods from the first category, suited when the medium has scale separation. In this case, the numerical homogenization methods provide an affordable approximation of u^ε, accurate in the $\mathrm{L}^\infty(\mathrm{L}^2)$ norm. The two main numerical homogenization methods have been developed in the framework of the heterogeneous multiscale method (HMM). Let us concisely explain the HMM. Considering an abstract incomplete physical system, the HMM is constituted of two components: first, a model at the macro scale, with a preferred solver, and second, a

numerical procedure at the micro scale, whose purpose is to extract the missing information by a sampling strategy. As the computations at the fine scale are performed independently in small sampling domains, the procedure can be efficiently parallelized thus granting a significant gain of time. In the case of the wave equation (3.2), the HMM aims to solve the homogenized equation at the macro scale and the missing datum is the homogenized tensor, which is approximated by solving micro problems. Two HMM are available for the wave equation.

The finite difference HMM (FD-HMM), introduced in [46, 19], relies on a finite difference method at the macro scale. The missing data, the effective flux, is approximated by solving micro problems in space-time sampling domains of size $\tau \times \eta^d$, where $\tau, \eta \geq \varepsilon$. The method is formally analyzed in [20], where an a priori error estimate for the approximate effective flux is shown. In particular, if the medium is locally periodic, the method converges to the homogenized solution.

In the finite element HMM (FE-HMM) from [8], finite elements are used at the macro scale (the FE-HMM is studied in Section 3.4). The homogenized tensor is approximated at the quadrature points of the macro mesh by solving micro problems in sampling domains of volume δ^d, where $\delta \geq \varepsilon$. The a priori error analysis provides a priori error estimates in the $\mathrm{L}^\infty(\mathrm{H}^1)$ and $\mathrm{L}^\infty(\mathrm{L}^2)$ norms. In particular, when the medium has a locally periodic structure, the FE-HMM converges to the homogenized solution. Note that in [6], a reduced basis approach is described to lighten the cost of the computation of the micro problems (see Section 3.4).

In a different framework, let us also mention the method presented in [70, 30]. At the macro scale, a spectral method is applied. The approximation of the homogenized medium is based on asymptotic expansions (in the periodic case) and computed in a preprocessing step. When applied to non periodic media, this step includes a filtering process. Despite a lack of theoretical support, the method appears to give satisfying numerical results.

Let us also cite the method from [53, 54]. It is designed to deal with wave problems in infinite periodic media that are locally perturbed. The procedure allows to find artificial boundary conditions in order to limit the computational domain to a neighborhood of the perturbation of the medium. In particular, the method provides the construction of discrete Dirichlet to Neumann operators for the coupling between the small domain, where the medium is perturbed, and the exterior domain, where the medium is periodic.

We continue with the second category of methods, suited when the medium does not have scale separation. In this case, the microscopic features of the medium are extracted and used in the construction a low dimensional space. The methods differ both in the way to acquire the microscopic informations and in the way to use it in the definition of the approximation space.

The first such method is found in [79]. It relies on a harmonic change of coordinates $G^\varepsilon = (G_1^\varepsilon, \ldots, G_d^\varepsilon)$. The space of approximation is a FE space with basis functions defined as compositions of coarse FE basis functions with G^ε. Under a so-called Cordes type condition, they provide a rigorous analysis of the method. Even though this condition is difficult to verify in practice, numerical experiments suggests that it might in fact not be necessary for the method to perform well. The main drawback of the method is the computation of G^ε, which requires to solve d elliptic PDEs at the fine scale, globally in Ω. This computation is indeed extremely costly if not impossible.

In the multiscale finite element methods using limited global information, introduced in [65, 64], it is assumed that u^ε can be approximated as $u^\varepsilon(t,x) \approx v(t, G_1^\varepsilon(x), \ldots, G_m^\varepsilon(x))$, where $(G_1^\varepsilon, \ldots, G_m^\varepsilon)$ are available fine scale fields. The approximation space is constructed by products of coarse FE basis functions with the fields G_k^ε. As the support of the obtained basis functions is local, the obtained matrices are sparse and the cost of the method is low. However, apart from the harmonic

change of coordinate introduced in [79], the only available procedures for the computation of the fields $(G_1^\varepsilon, \dots, G_m^\varepsilon)$ are heuristic. Furthermore, they also involve extremely costly global computations at the fine scale.

The multiscale method defined in [80] brings a notable improvement. Thanks to a flux-transfer transformation, the computation of the fine scale information is localized to portions of the domain Ω. To ensure optimal convergence rates, the diameter of these portions must be of order $\mathcal{O}(H^{1/2}|\log(H)|)$, where H is the size of a macroscopic mesh. As these problems can be solved independently, the procedure can be efficiently parallelized, thus ensuring a considerable gain of time. Note however, that in this approach, the locally supported coarse basis functions have to be sufficiently smooth (e.g., B-splines).

Finally, the multiscale method from [12] enters the framework of the localized orthogonal decomposition (LOD, see [75]). The method is based on a decomposition of a fine scale finite element space into a coarse part and a fine part. The fine part is computed by approximating the Riesz projection with respect to $(a^\varepsilon \nabla\cdot, \nabla\cdot)$. This approximation can be done locally by solving elliptic problems on patches of size $\mathcal{O}(H|\log(H)|)$. As these problems are independent, the preprocessing step for the construction of the multiscale basis can be efficiently parallelized

3.2 Homogenization of the wave equation in general media

In this section, we discuss different results for the homogenization of the wave equation. First, we introduce the general homogenization result for the wave equation via G-convergence. Furthermore, we discuss the convergence of the energy and its connection to a corrector result. Second, we explain the process of asymptotic expansion in the cases of the elliptic and wave equations in periodic media. In particular, we rigorously prove error estimates for the homogenized solutions following adaptation techniques. This part is an essential prerequisite to Chapters 4, 5, and 6, where we derive effective models for the wave over long time.

3.2.1 General homogenization of the wave equation by G-convergence

We state here the homogenization of the wave equation by G-convergence, proved in [27]. We refer to [87, 41] for the theory on G-convergence.

We define $\mathcal{M}(\lambda, \Lambda, \Omega)$ as the set of symmetric matrix functions $a \in [\mathrm{L}^\infty(\Omega)]^{d\times d}$ that are uniformly elliptic and bounded, i.e.,

$$\lambda|\xi|^2 \le a(x)\xi\cdot\xi \le \Lambda|\xi|^2 \qquad \forall \xi \in \mathbb{R}^d \quad \text{for a.e. } x \in \Omega. \tag{3.9}$$

Note that as a is symmetric, (3.9) implies that a is bounded is the classical sense: $|a(x)\xi| \le \|a(x)\|_2|\xi| \le \Lambda|\xi|$ ($\|\cdot\|_2$ denotes the spectral norm).

For a given $f \in \mathrm{W}^*_{\mathrm{per}}(\Omega)$, we consider the elliptic equation

$$\begin{aligned} &-\nabla_x \cdot \big(a^\varepsilon(x)\nabla_x u^\varepsilon(x)\big) = f(x) \qquad \text{in } \Omega,\\ &u^\varepsilon \ \ \Omega\text{-periodic}, \end{aligned} \tag{3.10}$$

where $\{a^\varepsilon\}_{\varepsilon>0}$ is a sequence of matrices in $\mathcal{M}(\lambda, \Lambda, \Omega)$. For $\varepsilon > 0$, as $a^\varepsilon \in \mathcal{M}(\lambda, \Lambda, \Omega)$, Lax–Milgram theorem ensures the existence and uniqueness of a weak solution of (3.10) $u^\varepsilon \in \mathrm{W}_{\mathrm{per}}(\Omega)$. To study the behavior of (3.10) and of its solution in the limit $\varepsilon \to 0$, we introduce the notion of G-convergence.

Definition 3.2.1. A sequence of matrices $\{a^\varepsilon\} \subset \mathcal{M}(\lambda, \Lambda, \Omega)$ *G-converges* to the matrix $a^0 \in \mathcal{M}(\lambda, \Lambda, \Omega)$ if, for every $f \in \mathrm{W}^*_{\mathrm{per}}(\Omega)$, the solution of (3.10) u^ε weakly converges in $\mathrm{W}_{\mathrm{per}}(\Omega)$ to

the solution u^0 of

$$\begin{array}{ll} -\nabla_x \cdot \big(a^0(x)\nabla_x u^0(x)\big) = f(x) & \text{in } \Omega, \\ u^0 \ \ \Omega\text{-periodic.} & \end{array} \tag{3.11}$$

The main result on G-convergence is its compactness property: for any sequence $\{a^\varepsilon\} \subset \mathcal{M}(\lambda, \Lambda, \Omega)$, there exists a subsequence $\{a^{\varepsilon'}\}$ and a matrix $a^0 \in \mathcal{M}(\lambda, \Lambda, \Omega)$ such that $\{a^{\varepsilon'}\}$ G-converges to a^0. This property implies the following result for (3.10): there exists $a^0 \in \mathcal{M}(\lambda, \Lambda, \Omega)$ and a subsequence $\{u^{\varepsilon'}\}$ of $\{u^\varepsilon\}$ that weakly converges in $\mathrm{W}_{\mathrm{per}}(\Omega)$ to the solution u^0 of (3.11). However, without additional assumption on $\{a^\varepsilon\}$, the theory does not provide an explicit formula for a^0. Furthermore, a^0 may not be unique as nothing ensures in general that different G-converging subsequences have the same limit. In other words, we have the existence of a limit equation, but it might not be unique and we have no way of computing its solution.

We consider now the wave equation in heterogeneous media. Let $\Omega \subset \mathbb{R}^d$ be an open hypercube and let $a^\varepsilon \in \mathcal{M}(\lambda, \Lambda, \Omega)$ be a symmetric, uniformly elliptic, bounded tensor (see (3.9)) . We consider the following equation : find $u^\varepsilon : [0,T] \times \Omega \to \mathbb{R}$ such that

$$\begin{array}{ll} \partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \big(a^\varepsilon(x)\nabla_x u^\varepsilon(t,x)\big) = f(t,x) & \text{in } (0,T] \times \Omega, \\ x \mapsto u^\varepsilon(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,T], \\ u^\varepsilon(0,x) = g^0(x), \quad \partial_t u^\varepsilon(0) = g^1(x) & \text{in } \Omega, \end{array} \tag{3.12}$$

where g^0, g^1 are initial conditions and f is a source term. As proved in Section 2.1.1, if $g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega), g^1 \in \mathrm{L}^2_0(\Omega)$ and $f \in \mathrm{L}^2(0,T;\mathrm{L}^2_0(\Omega))$, then there exists a unique weak solution of (3.12) such that $u^\varepsilon \in \mathrm{L}^\infty(0,T;\mathrm{W}_{\mathrm{per}}(\Omega))$, $\partial_t u^\varepsilon \in \mathrm{L}^\infty(0,T;\mathrm{L}^2_0(\Omega))$ and $\partial_t^2 u^\varepsilon \in \mathrm{L}^\infty(0,T;\mathrm{W}^*_{\mathrm{per}}(\Omega))$.

The general result of homogenization for the wave equation (3.12) is proved in [27]. In particular, we have the following theorem.

Theorem 3.2.2. *If $\{a^\varepsilon\} \subset \mathcal{M}(\lambda, \Lambda, \Omega)$ G-converges to a^0, then the solution u^ε of* (3.12) *satisfies*

$$u^\varepsilon \rightharpoonup^* u^0 \ \ \textit{in } \mathrm{L}^\infty(0,T;\mathrm{W}_{\mathrm{per}}(\Omega)), \qquad \partial_t u^\varepsilon \rightharpoonup^* \partial_t u^0 \ \ \textit{in } \mathrm{L}^\infty(0,T;\mathrm{L}^2_0(\Omega)),$$

where $u^0 : [0,T] \times \Omega \to \mathbb{R}$ is the solution of

$$\begin{array}{ll} \partial_t^2 u^0(t,x) - \nabla_x \cdot \big(a^0(x)\nabla_x u^0(t,x)\big) = f(t,x) & \textit{in } (0,T] \times \Omega, \\ x \mapsto u^0(t,x) \ \ \Omega\textit{-periodic} & \textit{in } [0,T], \\ u^0(0,x) = g^0(x), \quad \partial_t u^0(0) = g^1(x) & \textit{in } \Omega. \end{array} \tag{3.13}$$

Theorem 3.2.2 ensures the existence of an effective equation, namely the homogenized equation (3.13). However, for a general tensor a^ε, neither do we have a formula for the computation of a^0 nor do we even have its uniqueness.

In section 3.3, we discuss the homogenization of (3.12) in the particular case of a periodic tensor $a^\varepsilon(x) = a\big(\frac{x}{\varepsilon}\big)$. In particular, in this case the homogenized solution u^0 is unique and we derive an explicit formula for the homogenized tensor a^0, which is constant. We also prove a homogenization result under the form of an error estimate for $u^\varepsilon - u^0$.

3.2.2 Convergence of the energy and well-prepared initial data

We present here a particularity of the wave equation connected to the convergence of the energy. In particular, we present a corrector result proved in [27].

The homogenization of the wave equation has a particularity. Indeed, as presented in [27], the energy associated to u^ε does not converge in general to the energy associated to u^0 as $\varepsilon \to 0$. As a consequence, it is not possible to derive a standard corrector result.

For simplicity, assume that $f = 0$ and consider the energies associated to (3.12) and (3.13):

$$E^\varepsilon u^\varepsilon(t) = \tfrac{1}{2}\|\partial_t u^\varepsilon(t)\|^2_{\mathrm{L}^2(\Omega)} + \tfrac{1}{2}A^\varepsilon\big(u^\varepsilon(t), u^\varepsilon(t)\big),$$
$$E^0 u^0(t) = \tfrac{1}{2}\|\partial_t u^0(t)\|^2_{\mathrm{L}^2(\Omega)} + \tfrac{1}{2}A^0\big(u^0(t), u^0(t)\big).$$

As $f = 0$, we verify that $E^\varepsilon u^\varepsilon(t)$ and $E^0 u^0(t)$ are conserved and thus, for any $t \in [0,T]$,

$$E^\varepsilon u^\varepsilon(t) = E^\varepsilon u^\varepsilon(0) = \tfrac{1}{2}\|g^1\|^2_{\mathrm{L}^2(\Omega)} + \tfrac{1}{2}A^\varepsilon\big(g^0, g^0\big),$$
$$E^0 u^0(t) = E^0 u^0(0) = \tfrac{1}{2}\|g^1\|^2_{\mathrm{L}^2(\Omega)} + \tfrac{1}{2}A^0\big(g^0, g^0\big).$$

Let us verify that in general, $A^\varepsilon(g^0, g^0)$ does not converge to $A^0(g^0, g^0)$ as $\varepsilon \to 0$. Let $d = 1$ and consider a smooth Y-periodic tensor $a\big(\frac{x}{\varepsilon}\big)$. It can be proved that $a\big(\frac{\cdot}{\varepsilon}\big) \rightharpoonup^* \langle a\rangle_Y$ in $\mathrm{L}^\infty(\mathbb{R}^d)$ as $\varepsilon \to 0$ (see e.g. [37]). Hence, in this case we have $A^\varepsilon(g^0, g^0) \to \big(\langle a\rangle_Y \partial_x g^0, \partial_x g^0\big)_\Omega$. If the tensor is not constant, we verify that $a^0 \neq \langle a\rangle_Y$ (see Section 3.3.2) and thus $A^\varepsilon(g^0, g^0)$ does not converge to $A^0(g^0, g^0)$. Therefore, the energy of $E^\varepsilon u^\varepsilon(t)$ does not converge to $E^0 u^0(t)$.

The heart of the problem is in fact an incompatibility between the initial condition g^0 and the tensor a^ε. Indeed, we show that if the initial condition is well prepared, we obtain the desired convergence of the energy. Let $\tilde{u}^\varepsilon : [0,T] \times \Omega \to \mathbb{R}^d$ be the solution of

$$\begin{array}{ll} \partial_t^2 \tilde{u}^\varepsilon(t,x) - \nabla_x \cdot \big(a^\varepsilon(x)\nabla_x \tilde{u}^\varepsilon(t,x)\big) = f(t,x) & \text{in } (0,T]\times\Omega, \\ x \mapsto \tilde{u}^\varepsilon(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,T], \\ \tilde{u}^\varepsilon(0,x) = \tilde{g}^0(x), \quad \partial_t \tilde{u}^\varepsilon(0) = g^1(x) & \text{in } \Omega, \end{array} \tag{3.14}$$

where the prepared initial position $\tilde{g}^0$ is the solution of the elliptic equation

$$\begin{array}{ll} -\nabla_x \cdot \big(a^\varepsilon(x)\nabla_x \tilde{g}^0(x)\big) = -\nabla_x \cdot \big(a^0(x)\nabla_x g^0(x)\big) & \text{in } \Omega, \\ \tilde{g}^0 \ \ \Omega\text{-periodic.} & \end{array} \tag{3.15}$$

Then, $\tilde{u}^\varepsilon$ still satisfies the weak convergences $\tilde{u}^\varepsilon \rightharpoonup^* u^0$ in $\mathrm{L}^\infty(0,T;\mathrm{W}_{\mathrm{per}}(\Omega))$ and $\partial_t \tilde{u}^\varepsilon \rightharpoonup^* \partial_t u^0$ in $\mathrm{L}^\infty(0,T;\mathrm{L}^2_0(\Omega))$. Furthermore, the homogenization of (3.15) ensures the convergence (see e.g. [37])

$$\lim_{\varepsilon\to 0} A^\varepsilon(\tilde{g}^0, \tilde{g}^0) \to A^0(g^0, g^0).$$

We thus obtain $E^\varepsilon \tilde{u}^\varepsilon(t) \to E^0 u^0(t)$ as $\varepsilon \to 0$, as desired. This preparation of the initial condition allows to prove a corrector result. In particular, [27] proves that

$$\nabla_x \tilde{u}^\varepsilon - C^\varepsilon \nabla_x u^0 \to 0 \ \text{ in } \mathcal{C}^0([0,T];[\mathrm{L}^1(\Omega)]^d),$$

where C^ε is a corrector matrix associated to a^ε. The incompatibility between g^0 and a^ε is then contained in the residual $u^\varepsilon - \tilde{u}^\varepsilon$. The residual satisfies $u^\varepsilon - \tilde{u}^\varepsilon \rightharpoonup^* 0$ in $\mathrm{L}^\infty(0,T;\mathrm{W}_{\mathrm{per}}(\Omega))$, but its energy does not vanish in general:

$$E^\varepsilon(u^\varepsilon - \tilde{u}^\varepsilon)(t) = \tfrac{1}{2}A^\varepsilon\big(\tilde{g}^0, \tilde{g}^0\big) - \tfrac{1}{2}A^\varepsilon\big(g^0, g^0\big) \neq 0.$$

Note that in the case of a periodic tensor $a\big(\frac{x}{\varepsilon}\big)$, the corrector matrix C^ε can be defined explicitly. This is done in Section 3.3.3, where we prove an error estimate for $\|\nabla_x \tilde{u}^\varepsilon - C^\varepsilon \nabla_x u^0\|_{\mathrm{L}^\infty(\mathrm{L}^2)}$. Furthermore, we show that under sufficient regularity, $\tilde{g}^0$ in (3.14) can be replaced by $\bar{g}^0 = g^0 + \varepsilon\chi_i\big(\frac{\cdot}{\varepsilon}\big)\partial_i g^0$, where χ_i are the standard correctors in periodic homogenization.

3.3 Homogenization in periodic media using asymptotic expansion

Asymptotic expansion is a widely used technique in homogenization processes (see e.g. [24, 66, 37]). This technique plays a fundamental role in the derivation of the main results obtained in this thesis. In this section, we introduce asymptotic expansions in two different contexts. We first proceed to the homogenization of the elliptic equation in periodic media. Second, we perform the short time homogenization of the wave equation in periodic media. In particular, in both cases, we use the process to prove a rigorous error estimate between the homogenized solution and the original solution.

3.3.1 Error estimate for the homogenization of elliptic equations

We proceed here to the homogenization of the elliptic equation in periodic media using asymptotic expansion. In particular, we derive the homogenized equation and prove an error estimate ensuring that the homogenized solution is ε-close to the oscillating solution in the $\mathrm{L}^\infty(\mathrm{L}^2)$ norm.

Let $\Omega \subset \mathbb{R}^d$ be an open hypercube. Let $a \in \mathcal{M}(\lambda, \Lambda, Y)$ be a symmetric tensor that is periodic on the reference cell Y, i.e., $y \mapsto a(y)$ is Y-periodic. We consider the elliptic equation : find $u^\varepsilon : \Omega \to \mathbb{R}$ such that

$$\begin{aligned} &-\nabla_x \cdot \big(a\big(\tfrac{x}{\varepsilon}\big)\nabla_x u^\varepsilon(x)\big) = f(x) \qquad \text{in } \Omega, \\ &u^\varepsilon \quad \Omega\text{-periodic.} \end{aligned} \tag{3.16}$$

The well-posedness of (3.16) is ensured by Lax–Milgram theorem. The compactness property of the G-convergence, introduced in Section 3.2.1, ensures that the sequence $\{a\big(\tfrac{\cdot}{\varepsilon}\big)\}_{\varepsilon>0}$ admits a subsequence that G-converges to a tensor $a^0 \in \mathcal{M}(\lambda, \Lambda, Y)$. Hence, there exists a homogenized solution u^0 that solves (3.16) such that $u^\varepsilon \rightharpoonup u^0$ in $\mathrm{W}_{\mathrm{per}}(\Omega)$. In what follows, we use asymptotic expansion to derive a formula for a^0 and thus identify the homogenized solution.

The asymptotic expansion starts with the ansatz that u^ε can be expanded under the form

$$u^\varepsilon(x) = u^0\big(x, \tfrac{x}{\varepsilon}\big) + \varepsilon u^1\big(x, \tfrac{x}{\varepsilon}\big) + \varepsilon^2 u^2\big(x, \tfrac{x}{\varepsilon}\big) + \dots, \tag{3.17}$$

where the $u^i(x, y)$ are Ω-periodic in x and Y-periodic in y. Let us denote the operator $\mathcal{A}^\varepsilon = -\nabla_x \cdot \big(a\big(\tfrac{x}{\varepsilon}\big)\nabla_x \cdot\big)$. We introduce the differential operators

$$\mathcal{A}_{yy} = -\nabla_y \cdot \big(a(y)\nabla_y \cdot\big), \quad \mathcal{A}_{xy} = -\nabla_y \cdot \big(a(y)\nabla_x \cdot\big) - \nabla_x \cdot \big(a(y)\nabla_y \cdot\big),$$
$$\mathcal{A}_{xx} = -\nabla_x \cdot \big(a(y)\nabla_x \cdot\big),$$

and the chain rule implies that $\mathcal{A}^\varepsilon \psi\big(x, \tfrac{x}{\varepsilon}\big) = \big(\varepsilon^{-2}\mathcal{A}_{yy} + \varepsilon^{-1}\mathcal{A}_{xy} + \mathcal{A}_{xx}\big)\psi\big(x, \tfrac{x}{\varepsilon}\big)$ for any sufficiently smooth function $\psi(x, y)$. Inserting the ansatz (3.17) in (3.16), we obtain for any $x \in \Omega$, with $y = \tfrac{x}{\varepsilon}$,

$$\begin{aligned} 0 = \mathcal{A}^\varepsilon u^\varepsilon(x) - f(x) = \quad & \varepsilon^{-2}\big(\mathcal{A}_{yy}u^0(x,y) \qquad\qquad\qquad\qquad\qquad\quad\ \big) \\ & + \varepsilon^{-1}\big(\mathcal{A}_{yy}u^1(x,y) + \mathcal{A}_{xy}u^0(x,y) \qquad\qquad\quad\ \big) \\ & + \varepsilon^{0} \ \ \big(\mathcal{A}_{yy}u^2(x,y) + \mathcal{A}_{xy}u^1(x,y) + \mathcal{A}_{xx}u^0(x,y) - f(x)\big) \\ & + \mathcal{O}(\varepsilon). \end{aligned} \tag{3.18}$$

The right hand side of (3.18) is a polynomial of infinite degree in the variable ε. As this equality must hold for any $\varepsilon > 0$, all the coefficients of the polynomial have to vanish. Hence, the successive cancellations of the terms of order ε^i, i increasing, provide equations for $u^i(x, y)$. At order ε^{-2}, the equation reads : for all $x \in \Omega$, find a Y-periodic function $y \mapsto u^0(x, y)$ such that $\mathcal{A}_{yy}u^0(x, y) = 0$ for all $y \in Y$. We verify that any solution to this problem is of the form

$u^0(x,y) = u^0(x)$, where the dependence in x is still to be determined. Canceling the term of order ε^{-1} in (3.18), and taking into account the definition of u^0, we obtain the following equation for u^1: for all $x \in \Omega$, find a Y-periodic function $y \mapsto u^1(x,y)$ such that

$$-\partial_{y_m}\Big(a_{mn}(y)\big(\partial_{y_n}u^1(x,y) + \partial_{x_n}u^0(x)\big)\Big) = 0 \quad \forall y \in Y. \tag{3.19}$$

To prove the well-posedness of this elliptic PDE, we apply Lax–Milgram theorem in the space $\mathrm{W}_{\mathrm{per}}(Y)$. As the bilinear form $(v,w) \mapsto (a\nabla_y v, \nabla_y w)_Y$ is elliptic and bounded, we have to verify that the right hand side belong to $\mathrm{W}^*_{\mathrm{per}}(Y)$. We refer to Appendix A.2 for a characterization of $\mathrm{W}^*_{\mathrm{per}}(Y)$. In particular, $f \in [\mathrm{H}^1_{\mathrm{per}}(Y)]^*$ given by

$$\langle f, w\rangle = \big(f^0, w\big)_{\mathrm{L}^2(Y)} + \big(f^1_k, \partial_k w\big)_{\mathrm{L}^2(Y)},$$

for some $f^0, f^1_1, \ldots, f^1_d \in \mathrm{L}^2(Y)$ belongs to $\mathrm{W}^*_{\mathrm{per}}(Y)$ if and only if

$$\big(f^0, 1\big)_{\mathrm{L}^2(Y)} = 0. \tag{3.20}$$

We thus write the weak formulation of (3.19) : for all $x \in \Omega$, we look for $u^1(x) = u^1(x,\cdot) \in \mathrm{W}_{\mathrm{per}}(Y)$ such that

$$\big(a\nabla_y u^1(x), \nabla_y w\big)_{\mathrm{L}^2(Y)} = -\big(a\nabla_x u^0(x), \nabla_y w\big)_{\mathrm{L}^2(Y)} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(Y). \tag{3.21}$$

Using the characterization of $\mathrm{W}^*_{\mathrm{per}}(Y)$, we verify that the right hand side of (3.21) belongs to $\mathrm{W}^*_{\mathrm{per}}(Y)$ and thus (3.19) is well-posed. Looking for a solution of the form $u^1(x,y) = \chi_i(y)\partial_{x_i}u^0(x) + \tilde{u}^1(x)$, (3.19) can be rewritten as

$$\Big(-\nabla_y \cdot \big(a(y)\big(\nabla_y\chi_i(y) + e_i\big)\Big)\partial_{x_i}u^0(x) = 0 \quad y \in Y.$$

We thus define $u^1(x,y) = \chi_i(y)\partial_{x_i}u^0(x)$, where $\chi_i \in \mathrm{W}_{\mathrm{per}}(Y)$ is the solution of the cell problem

$$-\nabla_y \cdot \Big(a(y)\big(\nabla_y\chi_i(y) + e_i\big)\Big) = 0 \quad \forall y \in Y. \tag{3.22}$$

Note that we chose $\tilde{u}^1(x) = 0$ for simplicity. The process can indeed be carried on with $\tilde{u}^1(x)$ unknown, which is then constrained by the cancellation of higher order terms. Finally, we cancel the term of order ε^0 in (3.18), obtaining the following equation for u^2: for all $x \in \Omega$, find a Y-periodic function $y \mapsto u^2(x,y)$ such that

$$-\partial_{y_m}\Big(a_{mn}(y)\big(\partial_{y_n}u^2(x,y) + e_i\chi_j(y)\partial^2_{x_{ij}}u^0(x)\big)\Big) - e_i^T a(y)\big(\nabla_y\chi_j(y) + e_j\big)\partial^2_{x_{ij}}u^0(x) = f(x) \quad \forall y \in Y.$$

This elliptic PDE is well-posed if the right hand side belongs to $\mathrm{W}^*_{\mathrm{per}}(Y)$. Imposing the constraint (3.20), we verify that this equation is well-posed if

$$-\big\langle e_i^T a(\nabla_y\chi_j + e_j)\big\rangle_Y \partial^2_{x_{ij}}u^0(x) = f(x) \quad \forall x \in \Omega. \tag{3.23}$$

If (3.23) holds, then the term of order ε^0 in (3.18) vanishes if $u^2(x,y) = \theta_{ij}(y)\partial^2_{x_{ij}}u^0(x)$, where $\theta_{ij} \in \mathrm{W}_{\mathrm{per}}(Y)$ solves the cell problem

$$-\nabla_y \cdot \Big(a(y)\big(\nabla_y\theta_{ij}(y) + e_i\chi_j(y)\big)\Big) - e_i^T a(y)\big(\nabla_y\chi_j(y) + e_j\big) + a^0_{ij} = 0 \quad \forall y \in Y, \tag{3.24}$$

where we denoted

$$a^0_{ij} = \big\langle e_i^T a(\nabla_y\chi_j + e_j)\big\rangle_Y.$$

Hence, the equation characterizing u^0 (3.23), is an elliptic problem. To guarantee its well-posedness, we prove that a^0 is symmetric, elliptic and bounded in the following lemma.

Lemma 3.3.1. *Let a be a Y-periodic tensor that belongs to $\mathcal{M}(\lambda,\Lambda,Y)$ and define the homogenized tensor $a^0_{ij} = \langle e_i^T a(\nabla_y\chi_j + e_j)\rangle_Y$, where $\chi_j \in \mathrm{W}_{\mathrm{per}}(Y)$ solves the cell problem (weak formulation of (3.22))*

$$\big(a(\nabla_y\chi_j + e_j), \nabla_y w\big)_{\mathrm{L}^2(Y)} = 0 \quad \forall w \in \mathrm{W}_{\mathrm{per}}(Y). \tag{3.25}$$

Then a^0 can be alternatively written as

$$a^0_{ij} = \big\langle a(\nabla_y\chi_j + e_j)\cdot(\nabla_y\chi_i + e_i)\big\rangle_Y = -\big\langle a\nabla_y\chi_j\cdot\nabla_y\chi_i\big\rangle_Y + \big\langle ae_j\cdot e_i\big\rangle_Y, \tag{3.26}$$

and satisfies $a^0 \in \mathcal{M}(\lambda,\Lambda,Y)$.

Proof. First, note that Lax–Milgram theorem ensures the existence and uniqueness of χ_j. Note that the choice of normalization $\langle\chi_j\rangle_Y = 0$ is arbitrary and has no influence on the definition of a^0. Next, using (3.25) with the test function $w = \chi_i$ gives $\big(a(\nabla_y\chi_j + e_j), \nabla_y\chi_i\big) = 0$ and we can thus rewrite

$$|Y|a^0_{ij} = \big(a(\nabla_y\chi_j + e_j), e_i)\big)_Y + \big(a(\nabla_y\chi_j + e_j), \nabla_y\chi_i\big)_Y = \big(a(\nabla_y\chi_j + e_j), (\nabla_y\chi_i + e_i)\big)_Y,$$

which proves the first equality in (3.26) and the symmetry of a^0. Note that (3.25) can also be used to rewrite a^0 as

$$|Y|a^0_{ij} = \big(a\nabla_y\chi_j, e_i\big)_Y + \big(ae_j, e_i\big)_Y = -\big(a\nabla_y\chi_j, \nabla_y\chi_i\big)_Y + \big(ae_j, e_i\big)_Y,$$

proving the second equality in (3.26). Let us now prove that a^0 is λ-elliptic and Λ-bounded. For $\xi \in \mathbb{R}^d$, we have

$$|Y|a^0\xi\cdot\xi = \sum_{ij=1}^{d} \big(a(\nabla_y\chi_j + e_j)\cdot(\nabla_y\chi_i + e_i)\big)_Y \xi_i\xi_j = \big(aF_\xi, F_\xi\big)_Y, \tag{3.27}$$

where we denoted the field $F_\xi = \sum_{i=1}^d (\nabla_y\chi_i + e_i)\xi_i$. As χ_i is Y-periodic, it satisfies $\big(\nabla_y\chi_i, e_j\big)_Y = \int_Y \partial_{y_j}\chi_i \,\mathrm{d}y = 0$, and thus

$$\begin{aligned}\|F_\xi\|^2_{\mathrm{L}^2(Y)} = \big(F_\xi, F_\xi\big)_Y &= \big(\nabla_y\chi_i, \nabla_y\chi_j\big)_Y\xi_i\xi_j + \big(\nabla_y\chi_i, e_j\big)_Y\xi_i\xi_j + \big(e_i, \nabla_y\chi_j\big)_Y\xi_i\xi_j + \big(e_i, e_j\big)_Y\xi_i\xi_j \\ &= \big\|\textstyle\sum_i \nabla_y\chi_i\xi_i\big\|^2_{\mathrm{L}^2} + |Y||\xi|^2 \geq |Y||\xi|^2.\end{aligned}$$

Using (3.27) and the ellipticity of a, this estimate implies $|Y|a^0\xi\cdot\xi \geq \lambda\|F_\xi\|^2_{\mathrm{L}^2(Y)} \geq |Y|\lambda|\xi|^2$, which proves the λ-ellipticity of a^0. Using again (3.25) with the test function $w = \chi_i$, and the ellipticity of a, we write

$$\begin{aligned}\big(aF_\xi, F_\xi\big)_Y &= \big(a\nabla_y\chi_i, \nabla_y\chi_j\big)_Y\xi_i\xi_j + \big(a\nabla_y\chi_i, e_j\big)_Y\xi_i\xi_j + \big(ae_i, \nabla_y\chi_j\big)_Y\xi_i\xi_j + \big(ae_i, e_j\big)_Y\xi_i\xi_j \\ &= -\big(a\nabla_y\chi_i, \nabla_y\chi_j\big)_Y\xi_i\xi_j + \big(ae_i, e_j\big)_Y\xi_i\xi_j \\ &= -\Big(a\big(\textstyle\sum_i \nabla_y\chi_i\xi_i\big)\cdot\big(\sum_i \nabla_y\chi_i\xi_i\big)\Big)_Y + (a\xi,\xi)_Y \leq (a\xi,\xi)_Y.\end{aligned}$$

Using (3.27) and the bound on a, we thus get $|Y|a^0\xi\cdot\xi \leq (a\xi,\xi)_Y \leq |Y|\Lambda|\xi|^2$. This estimate proves the Λ-boundedness of a^0 and ends the proof of the lemma. □

Let us synthesize the conclusion of the asymptotic expansion. We have found the equation (3.23) which is well-posed and characterizes u^0, the first term in the expansion (3.17). The function u^0 is independent of ε and is an effective solution for u^ε. To support this last point, we prove a rigorous error estimate in the following theorem.

Theorem 3.3.2. *Assume that $d \leq 3$, $Y = (-1/2, 1/2)^d$ and that $\Omega = (a_1, b_1) \times \cdots \times (a_d, b_d)$ satisfies $(b_i - a_i)/\varepsilon \in \mathbb{N}_{>0}$. Let $u^\varepsilon \in \mathrm{W}_{\mathrm{per}}(\Omega)$ be the solution of (3.16) and let $u^0 \in \mathrm{W}_{\mathrm{per}}(\Omega)$ be the unique weak solution of*

$$\begin{array}{ll} -\nabla_x \cdot \big(a^0 \nabla_x u^0(x)\big) = f(x) & x \in \Omega, \\ u^0 \quad \Omega\text{-periodic}, & \end{array} \tag{3.28}$$

where $a^0_{ij} = \big\langle e_i^T a(\nabla_y \chi_j + e_j)\big\rangle_Y$ and χ_j is the solution of (3.22). Then, if $a \in \mathrm{W}^{2,\infty}(Y)$ and $f \in \mathrm{L}^2_0(\Omega) \cap \mathrm{H}^1(\Omega)$, the following error estimate holds

$$\|u^\varepsilon - u^0\|_{\mathrm{L}^2(\Omega)} \leq C\varepsilon \|u^0\|_{\mathrm{H}^3(\Omega)}, \tag{3.29}$$

where the constant C depends on the Poincaré constant, λ and $\max_{ij} \|a_{ij}\|_{\mathrm{W}^{2,\infty}(Y)}$.

Proof. Note that thanks to the definition of a^0, (3.24) is well-posed and θ_{ij} exists and is unique. Thanks to the assumption $f \in \mathrm{L}^2_0(\Omega) \cap \mathrm{H}^1(\Omega)$, a regularity result ensures that $u^0 \in \mathrm{H}^3(\Omega)$ (see Theorem A.2.2). Similarly, $a \in \mathrm{W}^{2,\infty}(Y)$ ensures that $\chi_i, \theta_{ij} \in \mathrm{H}^3(Y)$. Furthermore, as $d \leq 3$, the Sobolev embedding $\mathrm{H}^2_{\mathrm{per}}(Y) \hookrightarrow \mathcal{C}^0_{\mathrm{per}}(\bar{Y})$ (see Appendix A.2) ensures $\chi_i, \theta_{ij} \in \mathcal{C}^1_{\mathrm{per}}(\bar{Y})$. We can now define the following adaptation of u^0,

$$\boldsymbol{\mathcal{B}}^\varepsilon u^0(x) = \big[u^0(x) + \varepsilon \chi_i\big(\tfrac{x}{\varepsilon}\big)\partial_i u^0(x) + \varepsilon^2 \theta_{ij}\big(\tfrac{x}{\varepsilon}\big)\partial^2_{ij} u^0(x)\big] \quad x \in \Omega,$$

where $[\cdot]$ denotes the equivalence class in the quotient $\mathcal{W}_{\mathrm{per}}(\Omega) = \mathrm{H}^1_{\mathrm{per}}(\Omega)/\mathbb{R}$. Thanks to the assumption on Ω, we verify that $\boldsymbol{\mathcal{B}}^\varepsilon u^0$ is Ω-periodic (χ_i and θ_{ij} are extended to Ω by periodicity). Furthermore, the regularity of u^0, χ_j, θ_{ij} ensures that $\boldsymbol{\mathcal{B}}^\varepsilon u^0$ belongs to $\mathcal{W}_{\mathrm{per}}(\Omega)$. Using (3.16), we verify that $\boldsymbol{\mathcal{B}}^\varepsilon u^0$ satisfies $\mathcal{A}^\varepsilon \boldsymbol{\mathcal{B}}^\varepsilon u^0 - \big[f\big] = \boldsymbol{\mathcal{R}}^\varepsilon u^0$, where

$$\begin{aligned} \big\langle \boldsymbol{\mathcal{R}}^\varepsilon u^0, \boldsymbol{w}\big\rangle = \Big(\big[\quad & \varepsilon^{-1}\big(- \nabla_y \cdot (a(\nabla_y \chi_i + e_i))\big)\partial_i u^0 \\ & + \quad \big(- \nabla_y \cdot (a(\nabla_y \theta_{ij} + e_i \chi_j)) - e_i^T a(\nabla_y \chi_j + e_j) + a^0_{ij}\big)\partial^2_{ij} u^0\big], \boldsymbol{w}\Big)_{\mathcal{L}^2} \\ & + \varepsilon\Big(\big[e_i^T a(\nabla_y \theta_{ij} + e_i \chi_j)\partial^3_{ijk} u^0\big], \boldsymbol{w}\Big)_{\mathcal{L}^2} + \varepsilon\Big(a_{mi}\theta_{jk}\partial^3_{ijk} u^0, \partial_m \boldsymbol{w}\Big)_{\mathrm{L}^2}. \end{aligned}$$

The cell problems for χ_i and θ_{ij} imply that the two first terms of the right hand side vanish, and we thus verify $\|\boldsymbol{\mathcal{R}}^\varepsilon u^0\|_{\mathcal{W}^*_{\mathrm{per}}} \leq C\varepsilon \|u^0\|_{\mathrm{H}^3}$. Defining now $\boldsymbol{\eta} = \big[u^\varepsilon\big] - \boldsymbol{\mathcal{B}}^\varepsilon u^0 \in \mathcal{W}_{\mathrm{per}}(\Omega)$, we verify that $\boldsymbol{\eta}$ satisfies $\mathcal{A}^\varepsilon \boldsymbol{\eta} = \boldsymbol{\mathcal{R}}^\varepsilon u^0$ in $\mathcal{W}^*_{\mathrm{per}}(\Omega)$. Using the estimate provided by Lax–Milgram theorem, we obtain

$$\|\nabla_x \boldsymbol{\eta}\|_{\mathrm{L}^2} \leq C\varepsilon \|u^0\|_{\mathrm{H}^3}. \tag{3.30}$$

Hence, as $u^\varepsilon - u^0 \in \mathrm{W}_{\mathrm{per}}(\Omega)$, using the triangle and the Poincaré–Wirtinger inequalities, we have

$$\|u^\varepsilon - u^0\|_{\mathrm{L}^2} = \|\big[u^\varepsilon - u^0\big]\|_{\mathcal{L}^2} \leq \|\boldsymbol{\eta}\|_{\mathcal{L}^2} + \|\big[u^0\big] - \boldsymbol{\mathcal{B}}^\varepsilon u^0\|_{\mathcal{L}^2} \leq C_\Omega \|\nabla_x \boldsymbol{\eta}\|_{\mathrm{L}^2} + C\varepsilon \|u^0\|_{\mathrm{H}^2} \leq C\varepsilon \|u^0\|_{\mathrm{H}^3},$$

where we used the trivial estimate $\|\big[u^0\big] - \boldsymbol{\mathcal{B}}^\varepsilon u^0\|_{\mathcal{L}^2} \leq C\varepsilon \|u^0\|_{\mathrm{H}^2}$. We have proved estimate (3.29) and the proof of the theorem is complete. □

3.3.2 Error estimate for the homogenization of the wave equation

Following the process used in the previous section for the elliptic equation, we proceed here to the short time homogenization of the wave equation in periodic media using asymptotic expansions. In particular, we derive the homogenized equation and prove an error estimate ensuring that the homogenized solution describes well the oscillating wave in the $\mathrm{L}^\infty(\mathrm{L}^2)$ norm.

Let $\Omega \subset \mathbb{R}^d$ be an open hypercube and let $a \in \mathcal{M}(\lambda, \Lambda, Y)$ (see (3.9)), where $Y \subset \mathbb{R}^d$ is the reference cell. We consider the following equation : find $u^\varepsilon : [0,T] \times \Omega \to \mathbb{R}$ such that

$$\begin{array}{ll} \partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \big(a\big(\frac{x}{\varepsilon}\big)\nabla_x u^\varepsilon(t,x)\big) = f(t,x) & \text{in } (0,T]\times\Omega, \\ x \mapsto u^\varepsilon(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,T], \\ u^\varepsilon(0,x) = g^0(x), \quad \partial_t u^\varepsilon(0) = g^1(x) & \text{in } \Omega, \end{array} \tag{3.31}$$

where g^0, g^1 are initial conditions and f is a source term. Referring to Section 2.1.1, if $g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega), g^1 \in \mathrm{L}^2_0(\Omega)$ and $f \in \mathrm{L}^2(0,T;\mathrm{L}^2_0(\Omega))$, then there exists a unique weak solution of (3.31) such that $u^\varepsilon \in \mathrm{L}^\infty(0,T;\mathrm{W}_{\mathrm{per}}(\Omega))$, $\partial_t u^\varepsilon \in \mathrm{L}^\infty(0,T;\mathrm{L}^2_0(\Omega))$ and $\partial_t^2 u^\varepsilon \in \mathrm{L}^\infty(0,T;\mathrm{W}^*_{\mathrm{per}}(\Omega))$. The compactness property of the G-convergence ensures the existence of a subsequence of $\{a\big(\frac{\cdot}{\varepsilon}\big)\}_{\varepsilon>0}$ that G-converges to a tensor $a^0 \in \mathcal{M}(\lambda,\Lambda,\Omega)$ (see Section 3.2.1). Hence, Theorem 3.2.2 ensures that there exists a homogenized solution u^0 that solves (3.11) such that $u^\varepsilon \rightharpoonup^* u^0$ in $\mathrm{L}^\infty(0,T;\mathrm{W}_{\mathrm{per}}(\Omega))$. In what follows, we use asymptotic expansion to characterize a^0 and u^0.

In order to introduce a systematic methodology for the derivation of effective equations, let us change a little the process used in the previous section in the elliptic case. We start with two ansatz. The first ansatz is the form of the effective equation. In particular, we assume that $u^0 : [0,T] \times \Omega \to \mathbb{R}$ solves the equation

$$\begin{array}{ll} \partial_t^2 u^0(t,x) - a^0_{ij}\partial^2_{ij} u^0(t,x)\big) = f(t,x) & \text{in } (0,T]\times\Omega, \\ x \mapsto u^0(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,T], \\ u^0(0,x) = g^0(x), \quad \partial_t u^0(0,x) = g^1(x) & \text{in } \Omega, \end{array} \tag{3.32}$$

where the tensor a^0_{ij} has to be defined. The second ansatz is that u^ε can be approximated by an adaptation of u^0 of the form

$$\mathcal{B}^\varepsilon u^0(t,x) = u^0(t,x) + \varepsilon u^1\big(t,x,\tfrac{x}{\varepsilon}\big) + \varepsilon^2 u^2\big(t,x,\tfrac{x}{\varepsilon}\big), \tag{3.33}$$

where u^1 and u^2 are bounded operators of u^0 to be defined. The asymptotic expansion consists now in imposing $(\partial_t^2 - \mathcal{A}^\varepsilon)(\mathcal{B}^\varepsilon u^0 - u^\varepsilon) = \mathcal{O}(\varepsilon)$, where $\mathcal{A}^\varepsilon = -\nabla_x \cdot \big(a\big(\frac{x}{\varepsilon}\big)\nabla_x \cdot\big)$, to find explicitly u^1 and u^2 and link their definitions with the definition of a^0 in (3.32). Using equations (3.31) and (3.32) and the form of the adaptation (3.33), we obtain for any t, x,

$$\begin{aligned} (\partial_t^2 - \mathcal{A}^\varepsilon)\big(\mathcal{B}^\varepsilon u^0(t,x) - u^\varepsilon(t,x)\big) &= (\partial_t^2 - \mathcal{A}^\varepsilon)\mathcal{B}^\varepsilon u^0(t,x) - f(t,x) \\ &= \ \varepsilon^{-1}\big(\mathcal{A}_{yy}u^1(t,x,y) + \mathcal{A}_{xy}u^0\big(t,x,y\big) \qquad\qquad\qquad\qquad\qquad\qquad\quad\big) \\ &\quad + \varepsilon^0 \ \ \big(\mathcal{A}_{yy}u^2(t,x,y) + \mathcal{A}_{xy}u^1(t,x,y) + \mathcal{A}_{xx}u^0(t,x) - a^0_{ij}\partial^2_{ij}u^0(t,x)\big) \\ &\quad + \mathcal{O}(\varepsilon). \end{aligned} \tag{3.34}$$

where $y = \frac{x}{\varepsilon}$. Canceling the term of order $\mathcal{O}(\varepsilon^{-1})$ leads to defining $u^1(t,x,y) = \chi_i\big(\frac{x}{\varepsilon}\big)\partial_i u^0(t,x)$, where χ_i solves (3.22). To cancel the term of order $\mathcal{O}(\varepsilon^0)$, we define $u^2(t,x,y) = \theta_{ij}\big(\frac{x}{\varepsilon}\big)\partial^2_{ij}u^0(t,x)$ where $\theta_{ij} \in \mathrm{W}_{\mathrm{per}}(Y)$ solves (compare with (3.24))

$$\big(a\nabla_y\theta_{ij}, \nabla_y w\big)_{\mathrm{L}^2(Y)} = -\big(ae_i\chi_j, \nabla_y w\big)_{\mathrm{L}^2(Y)} + \big(a(\nabla_y\chi_j + e_j) - a^0 e_j, e_i\nabla_y w\big)_{\mathrm{L}^2(Y)} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(Y). \tag{3.35}$$

Imposing the solvability condition (3.20) on (3.35), we obtain the same definition for the homogenized tensor as obtained in the elliptic case in the previous section: $a^0_{ij} = \big\langle e_i^T a(\nabla_y\chi_j + e_j)\big\rangle_Y$. Lemma 3.3.1 ensures that a^0 is elliptic and bounded. Hence, the solution u^0 of (3.31) exists and is unique.

We are now able to prove the desired error estimate. For the sake of simplicity, we require here the regularity $u^0 \in \mathrm{L}^\infty(0,T;\mathrm{H}^4(\Omega))$, $\partial_t^2 u^0 \in \mathrm{L}^\infty(0,T;\mathrm{H}^2(\Omega))$. Note, however, that the lower regularity $u^0 \in \mathrm{L}^\infty(0,T;\mathrm{H}^3(\Omega))$, $\partial_t^2 u^0 \in \mathrm{L}^\infty(0,T;\mathrm{H}^1(\Omega))$ is sufficient to prove an error estimate (using the same technique as in the proof of Theorem 4.2.4, in Section 4.2.5).

Theorem 3.3.3. *Assume that $d \leq 3$, $Y = (-1/2, 1/2)^d$ and that $\Omega = (a_1, b_1) \times \cdots \times (a_d, b_d)$ satisfies $(b_i - a_i)/\varepsilon \in \mathbb{N}_{>0}$. Let u^ε be the solution of (3.31) and let $u^0 \in \mathrm{W}_{\mathrm{per}}(\Omega)$ be the unique weak solution of (3.32), where $a^0_{ij} = \langle e_i^T a(\nabla_y \chi_j + e_j)\rangle_Y$ and χ_j is the solution of (3.22). We assume that $a \in \mathrm{W}^{2,\infty}(Y)$ and that u^0 satisfies the regularity $u^0 \in \mathrm{L}^\infty(0,T;\mathrm{H}^4(\Omega))$, $\partial_t^2 u^0 \in \mathrm{L}^\infty(0,T;\mathrm{H}^2(\Omega))$. Then the following error estimate holds*

$$\begin{aligned}\|u^\varepsilon - u^0\|_{\mathrm{L}^\infty(0,T;\mathrm{L}^2(\Omega))} \leq C\varepsilon\Big(&\|g^1\|_{\mathrm{H}^2(\Omega)} + \|g^0\|_{\mathrm{H}^2(\Omega)} \\ &+ \|u^0\|_{\mathrm{L}^\infty(0,T;\mathrm{H}^4(\Omega))} + \|\partial_t^2 u^0\|_{\mathrm{L}^\infty(0,T;\mathrm{H}^2(\Omega))}\Big),\end{aligned} \tag{3.36}$$

where the constant C depends on the Poincaré constant T, λ and $\max_{ij} \|a_{ij}\|_{\mathrm{W}^{2,\infty}(Y)}$.

Proof. First, note that the regularity of the tensor ensures that $\chi_i, \theta_{ij} \in \mathcal{C}^1_{\mathrm{per}}(\bar{Y})$. We define the adaptation

$$\boldsymbol{\mathcal{B}}^\varepsilon u^0(t,x) = \big[u^0(t,x) + \varepsilon\chi_i\big(\tfrac{x}{\varepsilon}\big)\partial_i u^0(t,x) + \varepsilon^2\theta_{ij}\big(\tfrac{x}{\varepsilon}\big)\partial^2_{ij}u^0(t,x)\big],$$

and verify that the assumptions ensure $\boldsymbol{\mathcal{B}}^\varepsilon u^0(t) \in \mathcal{W}_{\mathrm{per}}(\Omega)$. Thanks to the regularity of u^0, the following equality holds in $\mathcal{L}^2(\Omega)$:

$$\big[\partial_t^2 u^0(t)\big] = \big[f(t)\big] + \big[a^0_{ij}\partial^2_{ij}u^0(t)\big].$$

We thus compute

$$\partial_t^2 \boldsymbol{\mathcal{B}}^\varepsilon u^0(t) = \big[f(t)\big] + \big[a^0_{ij}\partial^2_{ij}u^0(t)\big] + \boldsymbol{\mathcal{R}}^\varepsilon_1 u^0(t), \tag{3.37}$$

where $\boldsymbol{\mathcal{R}}^\varepsilon_1 u^0(t) = \big[\varepsilon\chi_i\partial_i\partial_t^2 u^0(t) + \varepsilon^2\theta_{ij}\partial^2_{ij}\partial_t^2 u^0(t)\big]$. For any $t \in [0,T]$,

$$\begin{aligned}\mathcal{A}^\varepsilon \boldsymbol{\mathcal{B}}^\varepsilon u^0(t) = \big[\quad &\varepsilon^{-1}\big(-\nabla_y\cdot(a(\nabla_y\chi_k + e_k))\big)\partial_k u^0(t) \\ + \quad &\big(-\nabla_y\cdot(a(\nabla_y\theta_{ij} + e_i\chi_j)) - e_i^T a(\nabla_y\chi_j + e_j)\big)\partial^2_{ij}u^0(t)\big] + \boldsymbol{\mathcal{R}}^\varepsilon_2 u^0(t),\end{aligned} \tag{3.38}$$

where $\boldsymbol{\mathcal{R}}^\varepsilon_2 u^0(t) = \big(-\nabla_y\cdot(ae_i\theta_{jk}) - e_i^T a(\nabla_y\theta_{jk} + e_j\chi_k)\big)\partial^3_{ijk}u^0(t) - \varepsilon^2 a_{ij}\theta_{kl}\partial^4_{ijkl}u^0(t)$. Combining (3.37) and (3.38) and equation (3.31), we verify that $\boldsymbol{\eta} = \big[u^\varepsilon\big] - \boldsymbol{\mathcal{B}}^\varepsilon u^0$ satisfies

$$\begin{aligned}&(\partial_t^2 + \mathcal{A}^\varepsilon)\boldsymbol{\eta}(t) = \boldsymbol{\mathcal{R}}^\varepsilon u^0(t) \quad \text{in } \mathcal{W}_{\mathrm{per}}(\Omega) \quad \text{for a.e. } t \in [0,T],\\ &\boldsymbol{\eta}(0) = \big[g^0 - \boldsymbol{\mathcal{B}}^\varepsilon g^0\big], \quad \partial_t\boldsymbol{\eta}(0) = \big[g^1 - \boldsymbol{\mathcal{B}}^\varepsilon g^1\big].\end{aligned}$$

where $\boldsymbol{\mathcal{R}}^\varepsilon u^0 = \boldsymbol{\mathcal{R}}^\varepsilon_1 u^0 + \boldsymbol{\mathcal{R}}^\varepsilon_2 u^0$. Using the error estimate from Lemma 4.2.1 gives the bound

$$\|\boldsymbol{\eta}\|_{\mathrm{L}^\infty(\mathcal{W})} \leq C\big(\|\partial_t\boldsymbol{\eta}(0)\|_{\mathcal{L}^2} + \|\boldsymbol{\eta}(0)\|_{\mathcal{L}^2} + \|\boldsymbol{\mathcal{R}}^\varepsilon u^0\|_{\mathrm{L}^1(\mathcal{L}^2)}\big), \tag{3.39}$$

where the norm $\|\cdot\|_{\mathcal{W}}$ is defined in (A.3) and satisfies $\|\cdot\|_{\mathcal{L}^2} \leq (1 + C_\Omega)\|\cdot\|_{\mathcal{W}}$ (C_Ω is the Poincaré constant). As Hölder's inequality gives $\|\boldsymbol{\mathcal{R}}^\varepsilon u^0\|_{\mathrm{L}^1(\mathcal{L}^2)} \leq T\|\boldsymbol{\mathcal{R}}^\varepsilon u^0\|_{\mathrm{L}^\infty(\mathcal{L}^2)}$, the definition of $\boldsymbol{\mathcal{B}}^\varepsilon$ in (3.37) and the regularity of the correctors leads to

$$\|\boldsymbol{\eta}\|_{\mathrm{L}^\infty(\mathcal{L}^2)} \leq C\|\boldsymbol{\eta}\|_{\mathrm{L}^\infty(\mathcal{L}^2)} \leq C\varepsilon\big(\|g^1\|_{\mathrm{H}^2} + \|g^0\|_{\mathrm{H}^2} + \|u^0\|_{\mathrm{L}^\infty(\mathrm{H}^4)} + \|\partial_t^2 u^0\|_{\mathrm{L}^\infty(\mathrm{H}^2)}\big).$$

As $(u^\varepsilon - u^0)(t) \in \mathrm{W}_{\mathrm{per}}(\Omega)$, we have $\|u^\varepsilon - u^0\|_{\mathrm{L}^\infty(\mathrm{L}^2)} = \|\,[u^\varepsilon - u^0]\,\|_{\mathrm{L}^\infty(\mathcal{L}^2)}$ and thus the triangle inequality and the definition of $\boldsymbol{\mathcal{B}}^\varepsilon u^0$ give

$$\begin{aligned}\|u^\varepsilon - u^0\|_{\mathrm{L}^\infty(\mathrm{L}^2)} &\leq \|\boldsymbol{\eta}\|_{\mathrm{L}^\infty(\mathcal{L}^2)} + \|\,[\boldsymbol{\mathcal{B}}^\varepsilon u^0 - u^0]\,\|_{\mathrm{L}^\infty(\mathcal{L}^2)} \\ &\leq C\varepsilon\big(\|u^0\|_{\mathrm{L}^\infty(\mathrm{H}^4)} + \|\partial_t^2 u^0\|_{\mathrm{L}^\infty(\mathrm{H}^2)}\big).\end{aligned}$$

We have proved (3.36) and the proof of the theorem is complete. □

3.3.3 A corrector result for the wave equation

As discussed in Section 3.2.2, the corrector result that can be proved for the wave equation is not standard. Namely, it requires a preparation of the initial condition. This issue is connected to the convergence of the energy and to an incompatibility between the tensor and the initial wave. In this section, in the particular case of a periodic tensor, we prove a corrector result for a prepared initial condition. In particular, we show that the corrected initial condition corresponds to the first order adaptation obtained in the previous section.

Let $\bar{u}^\varepsilon : [0,T] \times \Omega \to \mathbb{R}$ be the solution of the wave equation

$$\begin{array}{ll} \partial_t^2 \bar{u}^\varepsilon(t,x) - \nabla_x \cdot \big(a\big(\frac{x}{\varepsilon}\big)\nabla_x \bar{u}^\varepsilon(t,x)\big) = f(t,x) & \text{in } (0,T]\times\Omega, \\ x \mapsto \bar{u}^\varepsilon(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,T], \\ \bar{u}^\varepsilon(0,x) = \bar{g}^0(x), \quad \partial_t \bar{u}^\varepsilon(0) = g^1(x) & \text{in } \Omega, \end{array} \tag{3.40}$$

where the prepared initial position is

$$\bar{g}^0 = g^0 + \varepsilon \chi_i\big(\tfrac{\cdot}{\varepsilon}\big)\partial_i g^0 - \big\langle \varepsilon \chi_i\big(\tfrac{\cdot}{\varepsilon}\big)\partial_i g^0 \big\rangle_\Omega,$$

and χ_i are the solution of the cell problem (3.32). Note that $\bar{u}^\varepsilon$ is not defined in the same way as $\tilde{u}^\varepsilon$ in (3.14). Let us show that they are in fact ε-close in $\mathrm{W}_{\mathrm{per}}(\Omega)$. Indeed, recall that $\tilde{u}^\varepsilon(0) = \tilde{g}^0$, where $\tilde{g}^0$ is the solution in $\mathrm{W}_{\mathrm{per}}(\Omega)$ of the elliptic equation

$$-\nabla_x \cdot \big(a^\varepsilon \nabla_x \tilde{g}^0\big) = -\nabla_x \cdot \big(a^0 \nabla_x g^0\big).$$

This PDE matches the homogenization result of Theorem 3.3.2. in particular, (3.30) reads

$$\| \big[\tilde{g}^0\big] - \boldsymbol{\mathcal{B}}^\varepsilon g^0 \|_{\mathrm{H}^1} \leq C\varepsilon \|g^0\|_{\mathrm{H}^3(\Omega)}.$$

As we also have $\| \big[\bar{g}^0\big] - \boldsymbol{\mathcal{B}}^\varepsilon g^0 \|_{\mathrm{H}^1} \leq C\varepsilon \|g^0\|_{\mathrm{H}^3(\Omega)}$, applying the standard energy estimate for the wave equation solved by $\tilde{u}^\varepsilon - \bar{u}^\varepsilon$ (see Theorem 2.1.1), we obtain

$$\|\tilde{u}^\varepsilon - \bar{u}^\varepsilon\|_{\mathrm{H}^1} \leq C \| \big[\tilde{g}^0 - \bar{g}^0\big] \|_{\mathrm{H}^1} \leq C\big(\| \big[\tilde{g}^0\big] - \boldsymbol{\mathcal{B}}^\varepsilon g^0 \|_{\mathrm{H}^1} + \| \big[\bar{g}^0\big] - \boldsymbol{\mathcal{B}}^\varepsilon g^0 \|_{\mathrm{H}^1} \big) \leq C\varepsilon \|g^0\|_{\mathrm{H}^3(\Omega)}.$$

We prove the following corrector result.

Theorem 3.3.4. *Assume that the hypotheses of Theorem 3.3.3 hold and define the corrector matrix* $C^\varepsilon_{ij} = \delta_{ij} + \partial_{y_i}\chi_j\big(\frac{\cdot}{\varepsilon}\big)$. *Then the following estimate holds*

$$\begin{aligned} \|\nabla_x \bar{u}^\varepsilon - C^\varepsilon \nabla_x u^0\|_{\mathrm{L}^\infty(0,T;\mathrm{L}^2(\Omega))} \leq C\varepsilon \Big(& \|g^1\|_{\mathrm{H}^2(\Omega)} + \|g^0\|_{\mathrm{H}^3(\Omega)} \\ & + \|u^0\|_{\mathrm{L}^\infty(0,T;\mathrm{H}^4(\Omega))} + \|\partial_t^2 u^0\|_{\mathrm{L}^\infty(0,T;\mathrm{H}^2(\Omega))} \Big), \end{aligned} \tag{3.41}$$

where the constant C *depends on the Poincaré constant,* T, λ *and* $\max_{ij} \|a_{ij}\|_{\mathcal{C}^2(Y)}$.

Proof. Define $\boldsymbol{\eta} = \big[\bar{u}^\varepsilon\big] - \boldsymbol{\mathcal{B}}^\varepsilon u^0$, and use (3.37) and (3.38) to verify that it satisfies $(\partial_t^2 + \mathcal{A}^\varepsilon)\boldsymbol{\eta}(t) = \boldsymbol{\mathcal{R}}^\varepsilon u^0(t)$ in $\mathcal{W}_{\mathrm{per}}(\Omega)$, where $\boldsymbol{\mathcal{R}}^\varepsilon u^0 = \boldsymbol{\mathcal{R}}_1^\varepsilon u^0 + \boldsymbol{\mathcal{R}}_2^\varepsilon u^0$. The standard error estimate for the wave equation (see Theorem 2.1.1) ensures that (compare with (3.39))

$$\|\partial_t \boldsymbol{\eta}\|_{\mathrm{L}^\infty(\mathcal{L}^2)} + \|\nabla_x \boldsymbol{\eta}\|_{\mathrm{L}^\infty(\mathcal{L}^2)} \leq C\big(\|\partial_t \boldsymbol{\eta}(0)\|_{\mathcal{L}^2} + \|\nabla_x \boldsymbol{\eta}(0)\|_{\mathrm{L}^2} + \|\boldsymbol{\mathcal{R}}^\varepsilon u^0\|_{\mathrm{L}^1(\mathcal{L}^2)} \big).$$

Thanks to the definition of $\bar{u}^\varepsilon(0)$, we verify that $\|\nabla_x \boldsymbol{\eta}(0)\|_{\mathrm{L}^2} \leq C\varepsilon \|g^0\|_{\mathrm{H}^3}$ and thus

$$\|\nabla_x \boldsymbol{\eta}\|_{\mathrm{L}^\infty(\mathcal{L}^2)} \leq C\varepsilon \big(\|g^1\|_{\mathrm{H}^2(\Omega)} + \|g^0\|_{\mathrm{H}^3(\Omega)} + \|u^0\|_{\mathrm{L}^\infty(0,T;\mathrm{H}^4(\Omega))} + \|\partial_t^2 u^0\|_{\mathrm{L}^\infty(0,T;\mathrm{H}^2(\Omega))} \big). \tag{3.42}$$

Thanks to the equality $C^\varepsilon \nabla_x u^0 = \nabla_x \big(\boldsymbol{\mathcal{B}}^\varepsilon u^0 - \varepsilon^2 \big[\theta_{ij} \partial_{ij}^2 u^0\big] \big) - \varepsilon e_j \chi_i \partial_{ij}^2 u^0$, we have

$$\|\nabla_x \bar{u}^\varepsilon - C^\varepsilon \nabla_x u^0\|_{\mathrm{L}^2} \leq \|\nabla_x \boldsymbol{\eta}\|_{\mathrm{L}^\infty(\mathcal{L}^2)} + C\varepsilon \|u^0\|_{\mathrm{H}^3},$$

which, combined with (3.42), gives (3.41) and the proof of the theorem is complete. □

3.3.4 Numerical experiments

In this section, we illustrate the homogenization of the wave equation in periodic media in various numerical experiments. First, we compare the oscillating wave and the homogenized solution in a simple example. Then, we verify the corrector result of the previous section. In particular, we compute the remaining wave corresponding to the incompatibility between the tensor and the initial position.

Let the reference cell be $Y = (-1/2, 1/2)$ and consider to following oscillating tensor

$$a\big(\tfrac{x}{\varepsilon}\big) = \sqrt{2} - \cos\big(2\pi\tfrac{x}{\varepsilon}\big), \tag{3.43}$$

where $\varepsilon = 1/20$. We consider the solution u^ε of (3.16), where the initial conditions are given as $g^0(x) = e^{-10x^2}$ and $g^1(x) = 0$ and the source $f = 0$. Furthermore, we assume that Ω is large enough to have no influence on u^ε on the time interval $t \in [0, 10]$ (see below). Theorem 3.3.3 predicts that u^ε is close to the homogenized solution u^0 in the $\mathrm{L}^\infty(\mathrm{L}^2)$ norm. Hence, we study the homogenized solution. We compute the zero mean corrector and the homogenized tensor corresponding to $a\big(\frac{x}{\varepsilon}\big)$ are (see Theorem (3.3.2))

$$\chi(y) = \frac{1}{\pi}\mathrm{atan}\big((1+\sqrt{2})\tan(\pi y)\big) - y, \qquad a^0 = 1.$$

Therefore, the homogenized equation (3.32) is the wave equation with constant wave speed $\sqrt{a^0} = 1$ For $x \in \mathbb{R}$, its solution is given by d'Alembert's formula:

$$u^0(t,x) = \tfrac{1}{2}\big(g^0(x-t) + g^0(x+t)\big) + \tfrac{1}{2}\int_{x-t}^{x+t} g^1(s)\,\mathrm{d}s = \tfrac{1}{2}\big(g^0(x-t) + g^0(x+t)\big).$$

Hence, u^0 is the sum of two traveling waves of speed ± 1, i.e., one moving to the right and one moving to the left. In the left plot of Figure 3.1, we display the solution u^0 for $(t,x) \in [0,1] \times [-1.5, 1.5]$. We verify that the behavior of u^0 follows d'Alembert's formula. From this

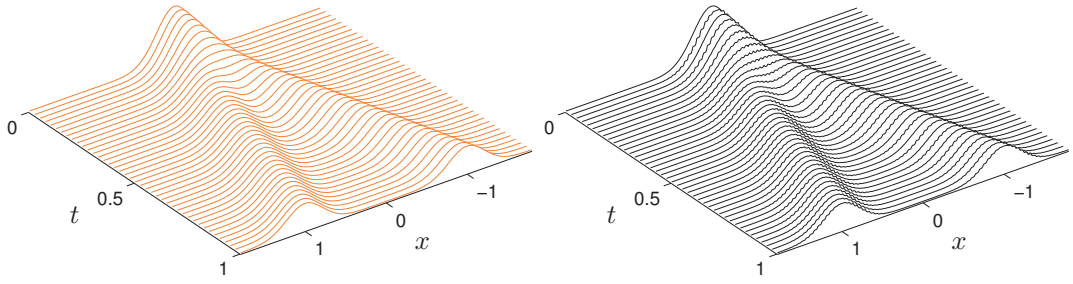

Figure 3.1: Left : the homogenized solution u^0 for $(t,x) \in [0,1] \times [-1.8, 1.8]$. Right : the fine scale solution u^ε for $(t,x) \in [0,1] \times [-1.8, 1.8]$.

information, we deduce that $\Omega = (-11, 11)$ is a sufficiently large domain for the two waves composing u^0 never to reach the boundary. We compute now an accurate approximation of u^ε using the pseudospectral method, introduced in Section 2.3, on a grid of size $\Delta x = \varepsilon/16$. The leap frog method is used for the time integration of the obtained second order ODE, with time step $\Delta t = \Delta x/50$. In the right plot of Figure 3.1, we displayed the evolution of u^ε for $(t,x) \in [0,1] \times [-1.5, 1.5]$. Furthermore, in Figure 3.2, we display the right going waves of u^ε and u^0 at $t = 10$. As predicted by Theorem 3.3.3, we observe that u^ε is close to u^0 in the $\mathrm{L}^\infty(\mathrm{L}^2)$ norm. Furthermore, we see that the oscillations at the micro scale $\mathcal{O}(\varepsilon)$ in u^ε are not described

by u^0. As discussed in Section 3.2.2, the gradient of u^ε can be captured by the correction $C^\varepsilon \nabla_x u^0$ only if the initial condition is prepared with respect to the oscillatory tensor. Denote $\bar{u}^\varepsilon$ the solution of the equation (3.3.3), where the initial position is prepared as

$$\bar{u}^\varepsilon(0) = \bar{g}^0 = g^0 + +\varepsilon\chi_i\left(\tfrac{\cdot}{\varepsilon}\right)\partial_i g^0 - \left\langle \varepsilon\chi_i\left(\tfrac{\cdot}{\varepsilon}\right)\partial_i g^0\right\rangle_\Omega.$$

Theorem 3.3.4 ensures that $\nabla_x \bar{u}^\varepsilon$ is close to $C^\varepsilon \nabla_x u^0$ in $\mathrm{L}^\infty(\mathrm{L}^2)$. However, as discussed in Section 3.2.2, the remainder $u^\varepsilon - \bar{u}^\varepsilon$ has a non vanishing energy. In Figure 3.3, we display $u^\varepsilon - \bar{u}^\varepsilon$ at $t = 0$ and its right going wave at $t = 10$. We observe that $u^\varepsilon - \bar{u}^\varepsilon$ is close to zero in $\mathrm{L}^\infty(\mathrm{L}^2)$. However, as

$$\lambda\|\nabla_x(u^\varepsilon - \bar{u}^\varepsilon)(t)\|^2_{\mathrm{L}^\infty(\mathrm{L}^2)} \leq A^\varepsilon\big((u^\varepsilon - \bar{u}^\varepsilon)(t), (u^\varepsilon - \bar{u}^\varepsilon)(t)\big) \leq E^\varepsilon(u^\varepsilon - \bar{u}^\varepsilon)(t),$$

and as $u^\varepsilon - \bar{u}^\varepsilon$ oscillates at the scale ε, we verify that the energy $E^\varepsilon(u^\varepsilon - \bar{u}^\varepsilon)(t)$ is positive. This testifies the incompatibility between the initial position g^0 and the tensor $a\left(\frac{x}{\varepsilon}\right)$. Let us now define the errors

$$\mathrm{e}(\bar{u}^\varepsilon) = \|\nabla_x u^\varepsilon - C^\varepsilon \nabla_x u^0\|_{\mathrm{L}^\infty(\mathrm{L}^2)}, \qquad \mathrm{e}(u^\varepsilon) = \|\nabla_x \bar{u}^\varepsilon - C^\varepsilon \nabla_x u^0\|_{\mathrm{L}^\infty(\mathrm{L}^2)},$$

where the correction C^ε is defined in Theorem 3.3.4. In Figure 3.4, $\mathrm{e}(u^\varepsilon)$ and $\mathrm{e}(\bar{u}^\varepsilon)$ are displayed for several values of ε (same settings as in the previous example). On the one hand, we observe that $\mathrm{e}(\bar{u}^\varepsilon)$ converges with a linear rate, as predicted by Theorem 3.3.4. On the other hand, $\mathrm{e}(u^\varepsilon)$ stagnates to an error of order $\mathcal{O}(1)$, as expected.

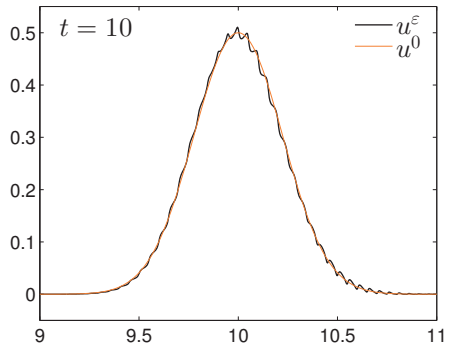

Figure 3.2: Comparison of the right-going waves of u^ε and u^0 at $t = 10$.

3.4 The finite element heterogeneous multiscale method (FE-HMM) for the wave equation

In this section, we follow [8] and define the finite element heterogeneous multiscale method (FE-HMM) for the wave equation and give its a priori error analysis. We refer to [8] for the missing proofs and further explanations. Note that the FE-HMM-L, studied in Chapter 7, is a modification of the FE-HMM designed for the long time approximation of the wave equation in one dimension.

Let $\Omega \subset \mathbb{R}^d$ be an open hypercube and let $a^\varepsilon \in \mathcal{M}(\lambda, \Lambda, \Omega)$ be a symmetric, uniformly elliptic, bounded tensor (see (3.9)). We consider the wave equation : find $u^\varepsilon : [0,T] \times \Omega \to \mathbb{R}$ such that

$$\begin{array}{ll} \partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \big(a^\varepsilon(x)\nabla_x u^\varepsilon(t,x)\big) = f(t,x) & \text{in } (0,T]\times\Omega, \\ x \mapsto u^\varepsilon(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,T], \\ u^\varepsilon(0,x) = g^0(x), \quad \partial_t u^\varepsilon(0) = g^1(x) & \text{in } \Omega, \end{array} \tag{3.44}$$

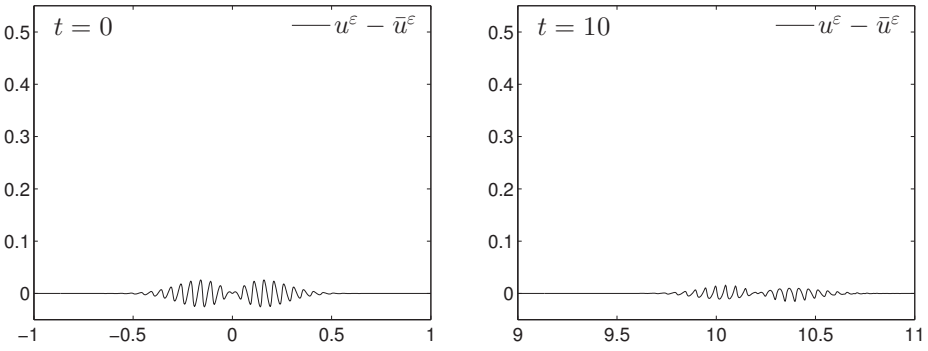

Figure 3.3: Illustration of the incompatibility between the initial position g^0 and the tensor $a\left(\frac{x}{\varepsilon}\right)$.

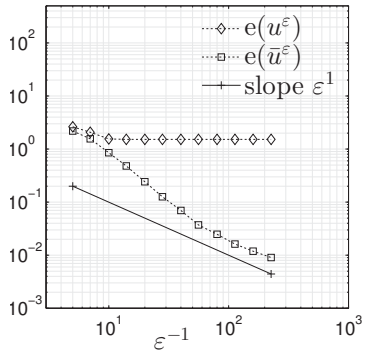

Figure 3.4: Loglog plots of the errors $\mathrm{e}(\bar{u}^\varepsilon)$ and $\mathrm{e}(\bar{u}^\varepsilon)$ for different size ε.

where g^0, g^1 are initial conditions and f is a source term. As proved in Section 2.1.1, if $g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega), g^1 \in \mathrm{L}^2_0(\Omega)$ and $f \in \mathrm{L}^2(0,T;\mathrm{L}^2_0(\Omega))$, then there exists a unique weak solution of (3.44) such that $u^\varepsilon \in \mathrm{L}^\infty(0,T;\mathrm{W}_{\mathrm{per}}(\Omega))$, $\partial_t u^\varepsilon \in \mathrm{L}^\infty(0,T;\mathrm{L}^2_0(\Omega))$ and $\partial_t^2 u^\varepsilon \in \mathrm{L}^2(0,T;\mathrm{W}^*_{\mathrm{per}}(\Omega))$.

Recall that if we assume that a^ε G-converges to a^0, Theorem 3.2.2 ensures the weak convergence $u^\varepsilon \rightharpoonup^* u^0$ in $\mathrm{L}^\infty(0,T;\mathrm{W}_{\mathrm{per}}(\Omega))$, where u^0 is the solution of the homogenized equation

$$\begin{array}{ll} \partial_t^2 u^0(t,x) - \nabla_x \cdot \big(a^0(x)\nabla_x u^0(t,x)\big) = f(t,x) & \text{in } (0,T]\times\Omega, \\ x \mapsto u^0(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,T], \\ u^0(0,x) = g^0(x), \quad \partial_t u^0(0) = g^1(x) & \text{in } \Omega. \end{array} \tag{3.45}$$

Recall that for general tensors a^ε, there is no explicit formula to compute the homogenized tensor a^0 in (3.45). The FE-HMM approximates a^0 on the fly by a sampling strategy. If the tensor a^ε has a locally periodic structure, the method is proved to converge to the homogenized solution u^0 (see Section 3.4.2).

3.4.1 The FE-HMM for the wave equation

Following [8], we define the FE-HMM for the numerical approximation of the wave equation (3.44). For simplicity, we restrict the definition of the method to simplicial meshes. Note that the analysis in [8] also holds for meshes with quadrilateral elements.

Let $\mathcal{T}_H$ be a partition of Ω into d-simplices. Denote by H_K the diameter of the element $K \in \mathcal{T}_H$ and define $H = \max_{K\in\mathcal{T}_H} H_K$. The macro finite element space is defined, for a given $\ell \in \mathbb{N}_{>0}$, as

$$V_H(\Omega) = \big\{ v_H \in \mathrm{W}_{\mathrm{per}}(\Omega) : v_H|_K \in \mathcal{P}^\ell(K) \ \forall K \in \mathcal{T}_H \big\}. \tag{3.46}$$

Let $\hat{K}$ be a reference element and, for every $K \in \mathcal{T}_H$, let F_K be the unique continuous mapping such that $F_K(\hat{K}) = K$ with $\det(\nabla F_K) > 0$, where ∇F_K denotes the Jacobian matrix of F_K. We are given a quadrature formula on $\hat{K}$ by a set of weights and quadrature points $\{\hat{\omega}_j, \hat{x}_j\}_{j=1}^J$. Note that it naturally induces a quadrature formula on K, whose weights and quadrature points are given by $\{\omega_{K_j} = \nabla F_K \hat{\omega}_j,\ x_{K_j} = F_K(\hat{x}_j)\}_{j=1}^J$. The following assumptions are required for the construction of the stiffness matrix to ensure the optimal convergence rate of FEM with numerical quadrature [34, 33] (see Theorems A.3.6 and A.3.9 in Section A.3):

$$\begin{aligned} &\textit{(i)} && \hat{\omega}_j > 0, \quad j = 1, \dots, J, \\ &\textit{(ii)} && \textstyle\int_{\hat{K}} \hat{p}(\hat{x})\,\mathrm{d}\hat{x} = \sum_{j=1}^J \hat{\omega}_j \hat{p}(\hat{x}_j) \quad \forall \hat{p} \in \mathcal{P}^\sigma(\hat{K}),\ \sigma = \max\{2\ell - 2, 1\}. \end{aligned} \tag{3.47}$$

Furthermore, we assume that the quadrature formula $\{\hat{\omega}'_j, \hat{x}'_j\}_{j=1}^{J'}$, required for the computation of the mass matrix, fulfills the following hypothesis

$$\textit{(iii)} \quad \textstyle\sum_{j=1}^{J'} \hat{\omega}'_j |\hat{p}(\hat{x}'_j)|^2 \geq \hat{\lambda}' \|\hat{p}\|_{\mathrm{L}^2(\hat{K})} \quad \forall \hat{p} \in \mathcal{P}^\ell(\hat{K}),\ \text{for a } \hat{\lambda}' > 0. \tag{3.48}$$

The quadrature formula $\{\hat{\omega}'_j, \hat{x}'_j\}_{j=1}^{J'}$ defines a scalar product (and associated norm) on $V_H(\Omega) \times V_H(\Omega)$, equivalent to the standard L^2 scalar product. For every macro element $K \in \mathcal{T}_H$ and every $j \in \{1, \dots, J\}$, we define around the quadrature point x_{K_j} a sampling domain $K_{\delta j} = x_{K_j} + \delta Y$, where δ is a positive real number such that $\delta \geq \varepsilon$. Each sampling domain $K_{\delta j}$ is discretized in a partition $\mathcal{T}_h$, where $h = \max_{Q\in\mathcal{T}_h} h_Q$ is the maximal diameter of the elements $Q \in \mathcal{T}_h$. The micro finite element space is defined, for a $q \in \mathbb{N}_{>0}$, as

$$V_h(K_{\delta j}) = \{ z_h \in W(K_{\delta j}) : z_h|_Q \in \mathcal{P}^q(Q) \ \ \forall Q \in \mathcal{T}_h \}, \tag{3.49}$$

where we let $W(K_{\delta j}) = \mathrm{W}_{\mathrm{per}}(K_{\delta j})$ for a periodic coupling and $W(K_{\delta j}) = \mathrm{H}^1_0(K_{\delta j})$ for a coupling with Dirichlet boundary conditions.

The FE-HMM is then defined as follows: find $u_H : [0, T] \to V_H(\Omega)$ such that

$$\begin{aligned} &\big(\partial_t^2 u_H(t), v_H\big)_H + A_H\big(u_H(t), v_H\big) = \big(f(t), v_H\big)_{\mathrm{L}^2} \quad \forall v_H \in V_H(\Omega) \text{ for a.e. } t \in [0, T], \\ &u_H(0) = g_H^0,\ \partial_t u_H(0) = g_H^1, \end{aligned} \tag{3.50}$$

where g_H^0, g_H^1 are appropriate approximations of the initial conditions g^0, g^1 in $V_H(\Omega)$, and the bilinear forms are defined for $v_H, w_H \in V_H(\Omega)$ as

$$\begin{aligned} A_H(v_H, w_H) &= \sum_{K\in\mathcal{T}_H} \sum_{j=1}^J \frac{\omega_{K_j}}{|K_{\delta j}|} \int_{K_{\delta j}} a^\varepsilon(x) \nabla v_{h,K_j}(x) \nabla w_{h,K_j}(x)\,\mathrm{d}x, \\ \big(v_H, w_H\big)_H &= \sum_{K\in\mathcal{T}_H} \sum_{j=1}^{J'} \omega'_{K_j} v_H(x'_{K_j}) w_H(x'_{K_j}). \end{aligned} \tag{3.51}$$

The micro functions v_{h,K_j} for v_H (resp. w_H) are the solutions of the following micro problems in $K_{\delta j}$: find v_{h,K_j} such that $(v_{h,K_j} - v^{\mathrm{lin}}_{H,K_j}) \in V_h(K_{\delta j})$ and

$$\big(a^\varepsilon \nabla v_{h,K_j}, \nabla z_h\big)_{\mathrm{L}^2(K_{\delta j})} = 0 \quad \forall z_h \in V_h(K_{\delta j}), \tag{3.52}$$

where the piecewise linear approximation of v_H (resp. w_H) around x_{K_j} is given by

$$v_{H,K_j}^{\mathrm{lin}}(x) = v_H(x_{K_j}) + (x - x_{K_j}) \cdot \nabla v_H(x_{K_j}).$$

We reformulate the FE-HMM (3.50) to understand its connection with the homogenization theory (see [8, 1, 3]). For every $(K, j) \in \mathcal{T}_H \times \{1, \dots, J\}$ and $1 \le n \le d$, define $\psi_{h,n}^{K,j} \in V_h(K_{\delta j})$ as the solution of the cell problem in the sampling domain $K_{\delta j}$:

$$\big(a^\varepsilon \nabla \psi_{h,n}^{K,j}, \nabla z_h\big)_{\mathrm{L}^2(K_{\delta j})} = -\big(a^\varepsilon e_n, \nabla z_h\big)_{\mathrm{L}^2(K_{\delta j})} \quad \forall z_h \in V_h(K_{\delta j}), \tag{3.53}$$

and define the tensor a_K^0 at the quadrature point x_{K_j} as

$$\big(a_K^0(x_{K_j})\big)_{mn} = \big\langle e_m^T a^\varepsilon \big(e_n + \nabla \psi_{h,n}^{K,j}\big)\big\rangle_{K_{\delta j}}, \quad 1 \le m, n \le d. \tag{3.54}$$

The following lemma is proved in [1, 3].

Lemma 3.4.1. *The bilinear form A_H can be rewritten for $v_H, w_H \in V_H(\Omega)$ as*

$$A_H(v_H, w_H) = \sum_{K \in \mathcal{T}_H} \sum_{j=1}^{J} \omega_{K_j} a_K^0(x_{K_j}) \nabla v_H(x_{K_j}) \nabla w_H(x_{K_j}). \tag{3.55}$$

Furthermore, A_H is elliptic and bounded, i.e., such that for any $v_H, w_H \in V_H(\Omega)$,

$$A_H(v_H, v_H) \ge \lambda \|\nabla v_H\|^2_{\mathrm{L}^2(\Omega)}, \qquad A_H(v_H, w_H) \le \Lambda^2/\lambda \|\nabla v_H\|_{\mathrm{L}^2(\Omega)} \|\nabla w_H\|_{\mathrm{L}^2(\Omega)}. \tag{3.56}$$

Proof. First, we verify that the micro function v_{h,K_j} satisfies $v_{h,K_j} = v_{H,K_j}^{\mathrm{lin}} + \psi_{h,n}^{K,j} \partial_n v_{H,K_j}^{\mathrm{lin}}$. Indeed, as the right hand side is a solution of (3.52), by unicity, it is equal to v_{h,K_j}. Using this equality in definition (3.51) gives (3.55). In order to prove the ellipticity and continuity of A_H, let us follow [2] and prove that

$$\|\nabla v_{H,K_j}^{\mathrm{lin}}\|_{\mathrm{L}^2(K_{\delta j})} \le \|\nabla v_{h,K_j}\|_{\mathrm{L}^2(K_{\delta j})} \le \sqrt{\Lambda/\lambda} \|\nabla v_{H,K_j}^{\mathrm{lin}}\|_{\mathrm{L}^2(K_{\delta j})}. \tag{3.57}$$

Let us drop the notation of K_j and denote $K_\delta = K_{\delta j}$. As $v_h - v_H^{\mathrm{lin}} \in V_h(K_\delta)$, note that for both couplings it holds $v_h - v_H^{\mathrm{lin}}|_{\partial K_\delta} = 0$. Hence, using that $\nabla v_H^{\mathrm{lin}}$ is constant on K_δ,

$$\big(\nabla v_H^{\mathrm{lin}}, \nabla v_h\big)_{K_\delta} = \nabla v_H^{\mathrm{lin}} \cdot \int_{K_\delta} \nabla v_h - \nabla v_H^{\mathrm{lin}} \,\mathrm{d}x + \big(\nabla v_H^{\mathrm{lin}}, \nabla v_H^{\mathrm{lin}}\big)_{K_\delta} = \|\nabla v_H^{\mathrm{lin}}\|^2_{K_\delta}.$$

We thus have

$$0 \le \|\nabla v_h - \nabla v_H^{\mathrm{lin}}\|^2_{K_\delta} = \|\nabla v_h\|^2_{K_\delta} + \|\nabla v_h\|^2_{K_\delta} - 2\big(\nabla v_H^{\mathrm{lin}}, \nabla v_h\big)_{K_\delta} = \|\nabla v_h\|^2_{K_\delta} - \|\nabla v_H^{\mathrm{lin}}\|^2_{K_\delta},$$

which proves the first inequality in (3.57). Next, (3.52) gives $\big(\nabla v_h, \nabla v_h - \nabla v_H^{\mathrm{lin}}\big)_{K_\delta} = 0$. We thus write

$$\begin{aligned}
\big(a^\varepsilon \nabla v_h, \nabla v_h\big)_{K_\delta} &= \big(a^\varepsilon \nabla v_h, \nabla v_h\big)_{K_\delta} - \big(a^\varepsilon \nabla v_h, \nabla v_h - \nabla v_H^{\mathrm{lin}}\big)_{K_\delta} \\
&\quad + \big(a^\varepsilon \nabla v_H^{\mathrm{lin}}, \nabla v_h - \nabla v_H^{\mathrm{lin}}\big)_{K_\delta} - \big(a^\varepsilon \nabla v_H^{\mathrm{lin}}, \nabla v_h\big)_{K_\delta} + \big(a^\varepsilon \nabla v_H^{\mathrm{lin}}, \nabla v_H^{\mathrm{lin}}\big)_{K_\delta} \\
&= \big(a^\varepsilon \nabla v_H^{\mathrm{lin}}, \nabla v_H^{\mathrm{lin}}\big)_{K_\delta} - \big(a^\varepsilon \nabla v_h - \nabla v_{H,K_j}^{\mathrm{lin}}, \nabla v_h - \nabla v_H^{\mathrm{lin}}\big)_{K_\delta}.
\end{aligned}$$

Using the ellipticity and the bound on a^ε, we obtain $\lambda \|\nabla v_h\|^2_{K_\delta} \le \Lambda \|\nabla v_{H,K_j}^{\mathrm{lin}}\|^2_{K_\delta}$, proving the second inequality in (3.57). The estimates in (3.56) follow from (3.57), the fact that $\nabla v_{H,K_j}^{\mathrm{lin}}(x_{K_j}) = \nabla v_H(x_{K_j})$, and the hypothesis on the quadrature formula (3.47). □

As a consequence of Lemma 3.4.1, (3.50) is equivalent to a regular second order ordinary differential equation. Therefore, existence and uniqueness of a solution of (3.50) is ensured by classical theory for ordinary differential equations [38] and the FE-HMM is well-posed. Furthermore, the solution u_H satisfies the regularity $u_H \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{W}_{\mathrm{per}}(\Omega)), \partial_t u_H \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{L}^2_0(\Omega))$.

3.4.2 A priori error analysis of the FE-HMM for the wave equation

In this section, we present the analysis of the FE-HMM for the wave equation provided in [8]. In particular, we discuss the different contributions of the error between the approximation and the homogenized solution.

Recall that a^0 is the homogenized tensor obtained as the G-limit of $\{a^\varepsilon\}$ as $\varepsilon \to 0$, and u^0 is the solution of the corresponding homogenized equation (3.45). The error $u^0 - u_H$ is composed of three parts. First, the error made at the macro level, coming from the FEM approximation at the macro scale. Second, the error made at the micro level, coming from the FEM approximation of the micro problems. Finally, the remaining error, called the modeling error, comes from the sampling strategy for computing a^0 and can be quantified only in the case where explicit formulas for a^0 are known. It is discussed below.

The a priori error analysis of the FE-HMM is stated in the following theorems (see [8] for the proofs).

Theorem 3.4.2. *Assume that $a^0 \in \mathrm{W}^{\ell,\infty}(\Omega)$ and $\partial_t^k u^0 \in \mathrm{L}^\infty(0,T;\mathrm{H}^{\ell+1}(\Omega))$ for $0 \le k \le 4$. Then $e = u^0 - u_H$ satisfies the estimate*

$$\|\partial_t e\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \|e\|_{\mathrm{L}^\infty(\mathrm{H}^1)} \le C\mathrm{e}^{\mathrm{data}}_{\mathrm{H}^1} + C\big(H^\ell + \mathrm{e}_{\mathrm{HMM}}\big) \sum_{k=0}^{4} \|\partial_t^k u^0\|_{\mathrm{L}^\infty(0,T;\mathrm{H}^{\ell+1}(\Omega)},$$

where $\mathrm{e}^{\mathrm{data}}_{\mathrm{H}^1} = \|g^1 - g^1_H\|_{\mathrm{L}^2} + \|g^0 - g^0_H\|_{\mathrm{H}^1}$,

$$\mathrm{e}_{\mathrm{HMM}} = \sup_{K\in\mathcal{T}_H, 1\le j\le J} \|a^0(x_{K_j}) - a^0_K(x_{K_j})\|_F,$$

and C is a constant independent of ε, δ, h, and H

Theorem 3.4.3. *Assume that $a^0 \in \mathrm{W}^{\ell+1,\infty}(\Omega)$ and $\partial_t^k u^0 \in \mathrm{L}^\infty(0,T;\mathrm{H}^{\ell+1}(\Omega))$ for $0 \le k \le 3$. Then $e = u^0 - u_H$ satisfies the estimate*

$$\|e\|_{\mathrm{L}^\infty(\mathrm{L}^2)} \le C\mathrm{e}^{\mathrm{data}}_{\mathrm{L}^2} + C\big(H^{\ell+1} + \mathrm{e}_{\mathrm{HMM}}\big) \sum_{k=0}^{3} \|\partial_t^k u^0\|_{\mathrm{L}^\infty(0,T;\mathrm{H}^{\ell+1}(\Omega)},$$

where $\mathrm{e}^{\mathrm{data}}_{\mathrm{L}^2} = \|I_H g^1 - g^1_H\|_{\mathrm{L}^2} + \|g^0 - g^0_H\|_{\mathrm{L}^2}$ *and C is a constant independent of ε, δ, h, and H (I_H denotes any projection operator onto $V_H(\Omega)$).*

In order to estimate the HMM error $\mathrm{e}_{\mathrm{HMM}}$, let us introduce the exact solution of the micro problems: for every $(K,j) \in \mathcal{T}_H \times \{1,\dots,J\}$ and $1 \le n \le d$, define $\psi_n^{K,j} \in W(K_{\delta j})$ as the solution of

$$\big(a^\varepsilon \nabla\psi_n^{K,j}, \nabla z\big)_{\mathrm{L}^2(K_{\delta j})} = -\big(a^\varepsilon e_n, \nabla z\big)_{\mathrm{L}^2(K_{\delta j})} \quad \forall z \in W(K_{\delta j}), \tag{3.58}$$

and define $\big(\bar a^0_K(x_{K_j})\big)_{mn} = \big\langle e_m^T a^\varepsilon \big(e_n + \nabla\psi_n^{K,j}\big)\big\rangle_{K_{\delta j}}$ for $1 \le m,n \le d$. We now split the HMM error into the micro error and the modeling error, $\mathrm{e}_{\mathrm{HMM}} \le \mathrm{e}_{\mathrm{mic}} + \mathrm{e}_{\mathrm{mod}}$, where

$$\mathrm{e}_{\mathrm{mic}} = \sup_{K\in\mathcal{T}_H, 1\le j\le J} \|a^0_K(x_{K_j}) - \bar a^0_K(x_{K_j})\|_F, \quad \mathrm{e}_{\mathrm{mod}} = \sup_{K\in\mathcal{T}_H, 1\le j\le J} \|a^0(x_{K_j}) - \bar a^0_K(x_{K_j})\|_F.$$

In order to estimate $\mathrm{e}_{\mathrm{mic}}$, no assumption on the structure of the tensor is required, but we assume that the solution of (3.58) satisfies

$$|\psi_n^{K,j}|_{\mathrm{H}^{q+1}(K_{\delta j})} \le C\varepsilon^{-q}\sqrt{|K_{\delta j}|}, \tag{3.59}$$

with C independent of ε, x_{K_j}, and $K_{\delta j}$. Note that for a periodic coupling, $W(K_{\delta j}) = \mathrm{W}_{\mathrm{per}}(K_{\delta j})$, (3.59) is satisfied if the tensor satisfies the regularity $a^\varepsilon \in \mathrm{W}^{q,\infty}(\Omega)$ and $|a^\varepsilon|_{\mathrm{W}^{q,\infty}} \le C\varepsilon^{-q}$. For a

coupling with homogeneous Dirichlet boundary conditions, $W(K_{\delta j}) = \mathrm{H}^1_0(K_{\delta j})$, (3.59) holds for $q = 1$ if the tensor satisfies the regularity $a^\varepsilon \in \mathrm{W}^{1,\infty}(\Omega)$ and $|a^\varepsilon|_{\mathrm{W}^{1,\infty}} \leq C\varepsilon^{-1}$. If (3.59) holds, then the micro error is proved to satisfy

$$\mathrm{e}_{\mathrm{mic}} \leq C\left(\frac{h}{\varepsilon}\right)^{2q}. \tag{3.60}$$

To analyze the modeling error $\mathrm{e}_{\mathrm{mod}}$, we assume that the tensor a^ε has a locally periodic structure, i.e., $a^\varepsilon(x) = a\big(x, \frac{x}{\varepsilon}\big)$, where $a(x, y)$ is Y-periodic in y and the reference cell $Y = (-1/2, 1/2)$. Under this assumption, the following explicit formula for a^0 can be proved for $x \in \Omega$

$$a^0_{mn}(x) = \big\langle e_m^T a(x, \cdot)(\nabla_y \chi_n(x, \cdot) + e_n)\big\rangle_Y,$$

where for all $x \in \Omega$ $\chi_n(x, \cdot) \in \mathrm{W}_{\mathrm{per}}(Y)$ is the solution of the local cell problem (compare with (3.58))

$$\big(a(x, \cdot)\nabla_y \chi_n(x, \cdot), \nabla_y w\big)_{\mathrm{L}^2(Y)} = -\big(a(x, \cdot)e_n, \nabla_y w\big)_{\mathrm{L}^2(Y)} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(Y).$$

Hence, if we assume that the tensor a^ε is collocated in the slow variable in each sampling domain, i.e., $a^\varepsilon|_{K_{\delta j}} = a\big(x_{K_j}, \frac{\cdot}{\varepsilon}\big)$, and if, in addition, we assume the regularity $a_{ij} \in \mathcal{C}^0(\bar{\Omega}; \mathrm{W}^{1,\infty}(Y))$, the following estimates can be proved

$$\mathrm{e}_{\mathrm{mod}} \leq \begin{cases} 0 & \text{if } \delta/\varepsilon \in \mathbb{N}_{>0} \text{ and } W(K_{\delta j}) = \mathrm{W}_{\mathrm{per}}(K_{\delta j}), \\ C\varepsilon/\delta & \text{if } \delta/\varepsilon \notin \mathbb{N}_{>0} \text{ and } W(K_{\delta j}) = \mathrm{H}^1_0(K_{\delta j}). \end{cases} \tag{3.61}$$

Without assuming that a^ε is collocated in the slow variable, an additional error proportional to the size of the sampling domain $C\delta$ is expected in both estimates (3.61). In the case $\delta/\varepsilon \notin \mathbb{N}_{>0}$, the error in (3.61) is called the resonance error and comes from a mismatch between the size of the sampling domains δ and the period of the tensor ε. In the case where the period is unknown, an oversampling strategy is used: use $W(K_{\delta j}) = \mathrm{H}^1_0(K_{\delta j})$ and use δ large enough to reduce the error term ε/δ. Note that this process increases notably the cost of the method. In particualr, finding an efficient method for the reduction of the resonance error is an active field of research.

Let us discuss the cost of the method. We denote $M_{\mathrm{mic}} = C(h/\varepsilon)^{-d}$ the number of degrees of freedom (DOF) in one micro problem and let N_{mic} be such that $h = \delta/N_{\mathrm{mic}}$. We first note that M_{mic} is independent of ε. Indeed, we have $\delta = C\varepsilon$, where $C = \mathcal{O}(1)$, and thus $h/\varepsilon = Ch/\delta = C/N_{\mathrm{mic}}$. As the number of DOF at the macro scale is also independent of ε, if $\mathrm{e}_{\mathrm{mod}} = 0$, the method converges independently of ε. Note that the main cost of the FE-HMM lies in solving the $\mathcal{O}(N_{\mathrm{mac}})$ micro problems (3.52), where $N_{\mathrm{mac}} = \mathcal{O}(H^{-d})$. Indeed, once the stiffness matrix is assembled, (3.50) has the standard cost of the FEM on the macro mesh $\mathcal{T}_H$. As the micro computations are independent, they can be done in parallel, which notably decreases the execution time. However, according to Theorem 3.4.3 and (3.60), to increase the accuracy both the micro and macro mesh sizes have to be decreased. Hence, increasing the accuracy substantially increases the cost to solve the micro problems: $N_{\mathrm{mac}}M_{\mathrm{mic}}$. To settle this issue, a reduced order method has been developed. Based on the reduced basis method (see [83] and the references therein), the FE-HMM was enhanced to a reduced basis FE-HMM (RB-FE-HMM) (see [6] for the elliptic case and [5] for the wave equation). Briefly, the RB-FE-HMM is divided into an offline and an online stages. The offline stage consists in the construction of a low dimensional subspace for the micro solutions. It relies on a greedy procedure based on an a posteriori error estimate for the approximation of the homogenized tensor. In the online stage, the homogenized tensor is then approximated inexpensively in the low dimensional subspace. The offline process is costly, but as it relies only on the tensor and on the domain, it can be reused for different initial data, source terms, boundary conditions, and even different physical problems.

4 Effective models for long time wave propagation in periodic media

This chapter contains the first main contributions of this thesis. In particular, we derive a new family of effective equations for the wave equation over long time. A substantial part of the content of the chapter was published in [14] (see also [13] for the one-dimensional case). Note that various additional results are presented.

We consider the wave equation in heterogeneous media over long time. Let $\Omega \subset \mathbb{R}^d$ be an arbitrarily large hypercube, let $a^\varepsilon(x) = a\left(\frac{x}{\varepsilon}\right)$ be a tensor, where $a(y)$ is periodic in a reference cell Y (e.g. $Y = (0,1)^d$), and let $T^\varepsilon = \varepsilon^{-2}T$, where $T = \mathcal{O}(1)$. We consider the solution $u^\varepsilon : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ of

$$\partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \left(a^\varepsilon(x)\nabla_x u^\varepsilon(t,x)\right) = f(t,x) \quad \text{in } (0, \varepsilon^{-2}T] \times \Omega, \tag{4.1}$$

where the initial conditions $u^\varepsilon(0,x)$ and $\partial_t u^\varepsilon(0,x)$ are given and $x \mapsto u^\varepsilon(t,x)$ is Ω-periodic. We assume that we are in a multiscale regime, i.e., the wavelengths of the initial conditions and of the source term are of order $\mathcal{O}(1)$, while the wavelength ε of the tensor is much smaller. We saw in Chapter 3 that at timescales $\mathcal{O}(1)$, the macroscopic behavior of u^ε is well described by the homogenized solution. However, at timescales $\mathcal{O}(\varepsilon^{-2})$, dispersion develops in the macroscopic behavior of u^ε. As this dispersion is not described by the homogenized solution, a new effective equation is needed. Finding such equation is an active field of research and the literature on this topic is reviewed in detail in Section 4.1. In particular, the main result available is presented in [42, 43], where one effective equation of the form (for $f = 0$)

$$\partial_t^2 u(t,x) - a^0_{ij}\partial^2_{ij} u(t,x) + \varepsilon^2 d_{ijkl}\partial^4_{ijkl} u(t,x) - \varepsilon^2 e_{ij}\partial^2_{ij}\partial_t^2 u(t,x) = f(t,x) \quad \text{in } (0, \varepsilon^{-2}T] \times \Omega, \tag{4.2}$$

is defined. To attest the validity of this equation, an error estimate is proved (the result holds in fact for the whole space $\Omega = \mathbb{R}^d$). In this chapter, we derive a new family of effective equations of the form

$$\partial_t^2 \tilde{u}(t,x) - a^0_{ij}\partial^2_{ij} \tilde{u}(t,x) + \varepsilon^2 a^2_{ijkl}\partial^4_{ijkl} \tilde{u}(t,x) - \varepsilon^2 b^2_{ij}\partial^2_{ij}\partial_t^2 \tilde{u}(t,x) = f(t,x) \quad \text{in } (0, \varepsilon^{-2}T] \times \Omega. \tag{4.3}$$

We emphasize that while [42, 43] construct one particular effective equation, we provide a characterization of an infinite set of effective equations. Under sufficient regularity of the data, we prove that any element $\tilde{u}$ of the family satisfies the estimate

$$\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} \leq C\varepsilon,$$

where the norm $\|\cdot\|_W$ is defined as (see (4.23))

$$\|w\|_W = \inf_{\substack{w = w_1 + w_2 \\ w_1, w_2 \in \mathrm{W}_{\mathrm{per}}(\Omega)}} \left\{ \|w_1\|_{\mathrm{L}^2(\Omega)} + \|\nabla w_2\|_{\mathrm{L}^2(\Omega)} \right\}, \tag{4.4}$$

and is equivalent to the $\mathrm{L}^2(\Omega)$ norm through the Poincaré constant. As the estimate holds for arbitrarily large hypercubes Ω, we can compare it to the result from [42, 43], which holds in the whole $\mathbb{R}^d$.

While the equation (4.2) is derived in [42, 43] using the Bloch wave expansion of u^ε, our derivation is done using asymptotic expansions. As a secondary result, we prove that these different approaches lead to the same effective equations. In particular, we prove that the effective equation (4.2) belongs to our family. Note that a similar comparison of these derivations was done independently in [18], with a focus on the elliptic case.

The family of effective equations is derived as follows. Assuming that an effective equation has the form (4.3), we construct an adaptation of $\tilde{u}$. The construction of the adaptation uses the asymptotic expansion technique, introduced for timescales $\mathcal{O}(1)$ in Section 3.3. In particular, the adaptation involves correctors, which are defined as the solutions of cell problems (elliptic PDEs in Y with periodic boundary conditions). However, as the timescale is now of order $\mathcal{O}(\varepsilon^{-2})$, more correctors are needed in the adaptation. Recall that in the derivation at short times, in Section 3.3, the homogenized tensor a^0 was characterized by the well-posedness of the second order cell problems. In a similar way, the well-posedness of the fourth order cell problems provides a constraint on the tensors a^2, b^2 in (4.3). Combined with the positive sign of a^2, b^2, required for the well-posedness of (4.3), this constraint characterizes the family of effective equations. Note that the ansatz on the effective equation is primordial. Indeed, we verify that if we start with the equation (4.3) without the operator $b^2_{ij}\partial^2_{ij}\partial^2_t$, we end up with an ill-posed equation (which was obtained in [85]). Hence, the starting equation needs to be general enough. In particular, this conclusion will be essential for the derivation of effective equations in locally periodic media in Chapter 6.

The implicit definition of the family through the constraint on a^2, b^2 is not usable as such in practice. We thus provide a constructive procedure to compute pairs of non-negative effective tensors a^2, b^2 defining equations of the family. Note that the algorithm requires to solve $d + \binom{d+1}{2}$ cell problems, while in [42, 43] this number is $d + \binom{d+1}{2} + \binom{d+2}{3}$.

The chapter is organized as follows. In Section 4.1, we present with an example the dispersive effects that appear at long times and review the literature available on this topic. Then, in Section 4.2, we present the main result of the chapter: we define the new family of effective equations and state the corresponding error estimate. In particular, we explain how asymptotic expansions are used to rigorously prove error estimates with adaptation techniques. Next, in Section 4.3, a procedure to construct effective equations is presented and an algorithm for the computation of the tensors is provided. Finally, we verify our theoretical findings through various numerical examples in Section 4.4.

4.1 Dispersive effects appearing at timescales $\mathcal{O}(\varepsilon^{-2})$: literature overview

In this section, we discuss the dispersive effects developed at long times by the solution of the wave equation in periodic media. In particular, we review the results available in the literature on this topic.

We consider the wave equation in an infinite periodic medium. Let $a^\varepsilon(x) = a\left(\frac{x}{\varepsilon}\right) = a(y)$ be a symmetric Y-periodic tensor in a reference cell Y (e.g., $Y = (-1/2, 1/2)^d$). We assume that a^ε is uniformly elliptic and bounded. Given initial conditions and a source f, we look for the wave displacement $u^\varepsilon : [0, T] \to \mathbb{R}$ such that

$$\partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \left(a^\varepsilon(x)\nabla_x u^\varepsilon(t,x)\right) = f(t,x) \quad \text{in } (0,T] \times \mathbb{R}^d. \tag{4.5}$$

Note that in applications (4.5) can not be approximated in the full space $\mathbb{R}^d$. One simple way to proceed is to solve the equation in a hypercube Ω that is large enough for the wave never to reach the boundary and add artificial periodic boundary conditions (we call such Ω a pseudoinfinite domain). For now, let us discuss (4.5) in the whole space $\mathbb{R}^d$. In Chapter 3 (Theorem 3.3.3), we presented the homogenization result for (4.5). In particular, we derived explicitly the homogenized equation

$$\partial_t^2 u^0(t,x) - a_{ij}^0 \partial_{ij}^2 u^0(t,x) = f(t,x) \quad \text{in } (0,T]\times\mathbb{R}^d, \tag{4.6}$$

whose solution u^0 no longer oscillates at the microscopic scale and describes the macroscopic behavior of the wave u^ε at short timescales $\mathcal{O}(1)$. Thanks to the periodic structure of a^ε, the homogenized tensor $a^0 \in \mathrm{Sym}^2(\mathbb{R}^d)$ in (4.6) can be computed explicitly via the solutions of d cell problems in Y.

However, it is known that at long timescales of order $\mathcal{O}(\varepsilon^{-2})$, dispersion effects appear in the macroscopic behavior of the wave u^ε. These effects are not described by the homogenized solution u^0. To see it, let us come back to the example from Section 3.3.4. We recall that the data are

$$a\big(\tfrac{x}{\varepsilon}\big) = \sqrt{2} - \cos\big(2\pi\tfrac{x}{\varepsilon}\big),\ \varepsilon = 1/20, \quad g^0(x) = e^{-10x^2}, \quad g^1(x) = 0, \quad f = 0.$$

As we want to approximate u^ε at the time $t = \varepsilon^{-2} = 400$, recalling that the homogenized wave speed is $\sqrt{a^0} = 1$, we let the computational domain be $\Omega = (-402, 402)$. In Figure 4.1, the time evolutions of u^ε (left) and u^0 (right) are displayed in the moving frame $x \in [\sqrt{a^0}t - 2.9, \sqrt{a^0}t + 1.1]$. We observe that, as the time increases, u^ε is macroscopically a superposition of waves moving with different speeds. This phenomenon is known as dispersion. Manifestly, the homogenized solution u^0 does not describe this dispersive behavior. Hence, a new effective solution that describes u^ε at timescales $\mathcal{O}(\varepsilon^{-2})$ is needed. Considering that u^0, valid for timescales $\mathcal{O}(\varepsilon^0)$, is a zero-th order effective equation, we are looking for a *higher order effective equation*, valid for timescales $\mathcal{O}(\varepsilon^{-2})$. Such equation must agree with (4.6) at order $\mathcal{O}(1)$ and have additional higher order constant differential operators for the description of the dispersion. The challenge lies first in exhibiting the form of these operators, then defining the coefficients driving them, and, finally, giving an efficient algorithm to compute them.

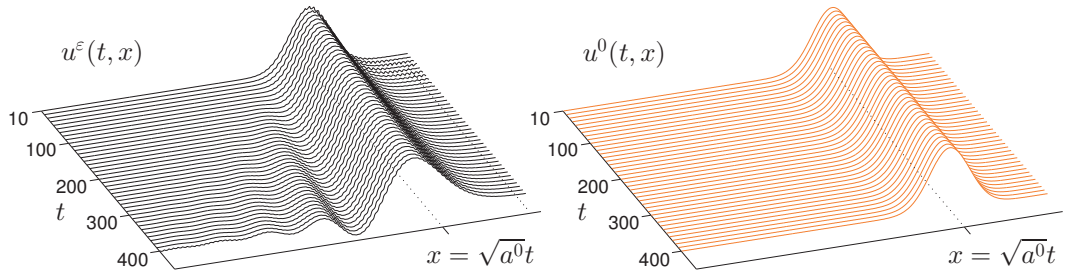

Figure 4.1: Comparison between u^ε and u^0 in a moving frame. The data of the problem are given in the text.

In the literature, several papers have addressed this problem (see [85, 52, 51, 72, 42, 43, 13, 18, 14]). Before going into details on some results, let us give a chronological review. In [85], Santosa and Symes formally built an approximation of u^ε (for $f = 0$) over times of order $\mathcal{O}(\varepsilon^{-2})$ that solves (with a higher order remainder) an equation of the form

$$\partial_t^2 u(t,x) - a_{ij}^0 \partial_{ij}^2 u(t,x) + \varepsilon^2 c_{ijkl} \partial_{ijkl}^4 u(t,x) = 0 \quad \text{in } (0,\varepsilon^{-2}T]\times\mathbb{R}^d. \tag{4.7}$$

However, due to the negative sign of the tensors c, (4.7) is ill-posed. Nevertheless, numerical experiments show that a regularized approximation of (4.7) captures the desired dispersive effects of u^ε. Recently, several authors proposed a well-posed modification of (4.7).

The first rigorous result was presented by Lamacz in [72], in the one-dimensional case. An error estimate is proved for timescales $\mathcal{O}(\varepsilon^{-2})$ between u^ε (for $f = 0$) and the solution of a Boussinesq type equation given by

$$\partial_t^2 u(t,x) - a^0 \partial_x^2 u(t,x) - \varepsilon^2 b \partial_x^2 \partial_t^2 u(t,x) = 0 \quad \text{in } (0, \varepsilon^{-2}T] \times \mathbb{R}. \tag{4.8}$$

The coefficient b in (4.8) is computed via a cascade of 3 elliptic cell problems (including the cell problem necessary for a^0). This result is discussed in detail in Section 4.1.2.

In the multidimensional case, an error estimate over long times $\mathcal{O}(\varepsilon^{-2})$ was then proved by Dohnal, Lamacz, and Schweizer in [42, 43]. The (well-posed) effective equation is of the form (for $f = 0$)

$$\partial_t^2 u(t,x) - a_{ij}^0 \partial_{ij}^2 u(t,x) + \varepsilon^2 \big(d_{ijkl} \partial_{ijkl}^4 u(t,x) - e_{ij} \partial_{ij}^2 \partial_t^2 u(t,x) \big) = 0 \quad \text{in } (0, \varepsilon^{-2}T] \times \mathbb{R}^d, \tag{4.9}$$

where the tensors d, e are computed via an algebraic decomposition of the tensor c in (4.7). Their numerical procedure involves the solution of $d + \binom{d+1}{2} + \binom{d+2}{3}$ cell problems. A summary of the derivation is given in Section 4.1.1.

In [13], we generalized the one-dimensional result from [72]. Using a similar technique, we derived a family of (well-posed) effective equations of the form (4.9) (in one dimension), where the coefficients are computed with the help of a single cell problem (the same as to compute a^0). We emphasize that while the equations (4.8) in [72] and (4.9) in [42, 43] defined single equations, we provided the characterization of infinitely many. In addition, the family is validated by the proof of an error estimate. This result contributes to this thesis and is presented in Section 4.2 (Section 4.3.1 for the particular one-dimensional case).

Next, Allaire, Briane, and Vanninathan [18] derived in a formal way an equation of the form

$$\partial_t^2 u(t,x) - a_{ij}^0 \partial_{ij}^2 u(t,x) + \varepsilon^2 \big(\tilde{d}_{ijkl} \partial_{ijkl}^4 u(t,x) - \varepsilon^2 \tilde{e}_{ij} \partial_{ij}^2 \partial_t^2 u(t,x) \big) = f(t,x) - \tilde{e}_{ij} \partial_{ij}^2 f(t,x), \tag{4.10}$$

in $(0, \varepsilon^{-2}T] \times \mathbb{R}^d$. The tensors $\tilde{d}, \tilde{e}$ are obtained by a theoretical decomposition of the tensor c in (4.7), which differs from the one in [42, 43]. Although their derivation clarifies the connection between the approaches of Bloch-wave expansion and of asymptotic expansion, no numerical procedure is provided for the computation of the tensors. Furthermore, no error estimate is provided to certify the derivation.

Then, we generalized our result from [13] to the multidimensional case in [14]. In particular, a whole family of effective equations of the form (4.9) was defined. Furthermore, we described a numerical procedure that involves the solution of only $d + \binom{d+1}{2}$ cell problems. This is the main result presented in the current chapter.

To be complete, let us finally mention a recent paper from Benoit and Gloria [23] dealing with the long time homogenization of the wave equation. However, the review of their result is postponed to the next chapter, as they provide an effective equation of arbitrary order (which is precisely the topic of Chapter 5).

4.1.1 Derivation of an effective equation via Bloch wave expansion

The first framework used to derive effective equations uses the expansion of u^ε in Bloch waves (see [92]). This approach has first been used by Santosa and Symes in [85], where an ill-posed equation

is formally derived. The approach was then carried on by Dohnal, Lamacz, and Schweizer in [42, 43], where a well-posed effective model is obtained and an error estimate is rigorously proved. In this section, we summarize the result from [42, 43] and indicate what is owed to [85]. Let us mention that Bloch waves have also been used for the homogenization of elliptic equations (see [40] and the references therein). Note that the survey given here continues in Section 4.2.7, where we compare the effective tensors obtained in this thesis via asymptotic expansion with the ones obtained in [85, 42, 43] via Bloch wave expansion. In particular, we show that the two approaches lead to the same effective tensors.

Let us present the exact settings of [85, 42, 43]. We consider the wave equation in periodic media at timescale $\mathcal{O}(\varepsilon^{-2})$ in the whole space $\mathbb{R}^d$: $u^\varepsilon : [0, \varepsilon^{-2}T] \times \mathbb{R}^d \to \mathbb{R}$ such that

$$\begin{aligned} &\partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \big(a\big(\tfrac{x}{\varepsilon}\big)\nabla_x u^\varepsilon(t,x)\big) = 0 \qquad && \text{in } (0,\varepsilon^{-2}T]\times\mathbb{R}^d, \\ &u^\varepsilon(0,x) = g(x), \quad \partial_t u^\varepsilon(0,x) = 0 && \text{in } \mathbb{R}^d, \end{aligned} \tag{4.11}$$

where we assume that $g \in \mathrm{L}^2(\mathbb{R}^d) \cap \mathrm{L}^1(\mathbb{R}^d)$ is such that its Fourier transform G has a compact support $K \subset\subset \mathbb{R}^d$. The tensor $a(y)$ is assumed to belong to $[\mathcal{C}^1_{\mathrm{per}}(\bar{Y})]^{d\times d}$, where $Y = (-\pi,\pi)^d$. The effective equation from [42, 43] is given by $w^\varepsilon : [0,\varepsilon^{-2}T]\times\mathbb{R}^d \to \mathbb{R}$

$$\begin{aligned} &\partial_t^2 w^\varepsilon(t,x) - a^0_{ij}\partial^2_{ij} w^\varepsilon - \varepsilon^2\big(E_{ij}\partial^2_{ij}\partial_t^2 w^\varepsilon - F_{ijmn}\partial^4_{ijmn} w^\varepsilon\big) = 0 \qquad && \text{in } (0,\varepsilon^{-2}T]\times\mathbb{R}^d, \\ &w^\varepsilon(0,x) = g(x), \quad \partial_t w^\varepsilon(0,x) = 0 && \text{in } \mathbb{R}^d, \end{aligned} \tag{4.12}$$

where a^0 is the homogenized tensor and E, F are defined through an algebraical procedure (see (4.18) below).

The following error estimate is then proved in [42, 43], in the norm

$$\|v\|_{\mathrm{L}^2(\mathbb{R}^d)+\mathrm{L}^\infty(\mathbb{R}^d)} = \inf_{\substack{v=v_1+v_2 \\ v_1\in\mathrm{L}^2(\mathbb{R}^d), v_2\in\mathrm{L}^\infty(\mathbb{R}^d)}} \Big\{\|v_1\|_{\mathrm{L}^2(\mathbb{R}^d)} + \|v_2\|_{\mathrm{L}^\infty(\mathbb{R}^d)}\Big\}.$$

Theorem 4.1.1 (Dohnal, Lamacz & Schweizer, [42, 43])**.** *The solutions u^ε and w^ε of respectively* (4.11) *and* (4.12) *satisfy the error estimate*

$$\|u^\varepsilon - w^\varepsilon\|_{\mathrm{L}^\infty(0,\varepsilon^{-2}T;\mathrm{L}^2(\mathbb{R}^d)+\mathrm{L}^\infty(\mathbb{R}^d))} \le C\varepsilon, \tag{4.13}$$

where C depends only on a, Y, T and g.

The starting point of [85] and [42, 43] is the expression of u^ε in Bloch waves. Let the reciprocal periodicity cell be $Z = (-1/2, 1/2)^d$. Then, for a fixed $k \in Z$, we construct $\{\mu_m(k), \psi_m(y,k)\}_{m=0}^\infty$ the eigenvalues and eigenfunctions of the problem

$$-(\nabla_y + \mathrm{i}k)\cdot\big(a(y)(\nabla_y + \mathrm{i}k)\psi_m(y,k)\big) = \mu_m(k)\psi_m(y,k),$$

where $\mu_m(k)$ are real and $\mu_{m+1}(k) \ge \mu_m(k) \ge 0$. We define then the rescaled Bloch waves $w^\varepsilon_m(x,k) = \psi_m\big(\frac{x}{\varepsilon}, \varepsilon k\big)e^{ik\cdot x}$ and the rescaled eigenvalues $\mu^\varepsilon_m(k) = \mu_m(\varepsilon k)$. In particular, $\{\mu^\varepsilon_m(k), w^\varepsilon_m(x,k)\}$ satisfy

$$-\nabla_x\cdot\big(a\big(\tfrac{x}{\varepsilon}\big)\nabla_x w^\varepsilon_m(x,k)\big) = \mu^\varepsilon_m(k) w^\varepsilon_m(x,k),$$

and the Bloch waves $\{w^\varepsilon_m(x,k)\}_{m\ge0}$ form a basis of $\mathrm{L}^2(\mathbb{R}^d)$. Then u^ε can be expressed as

$$u^\varepsilon(t,x) = \sum_{m=0}^\infty \int_{Z/\varepsilon} \hat{g}^\varepsilon_m(k) w_m(x,k) \Re\big(e^{\mathrm{i}t\sqrt{\mu^\varepsilon_m(k)}}\big)\,\mathrm{d}k, \quad \hat{g}^\varepsilon_m(k) = \int_{\mathbb{R}^d} g(x)\overline{w^\varepsilon_m(x,k)}\,\mathrm{d}x, \tag{4.14}$$

where $\overline{z}$ denotes the complex conjugate of z and $\Re(z)$ its real part. First, it is proved in [85] (and used in [42, 43]) that

$$\Big\| \sum_{m=1}^{\infty} \int_{Z/\varepsilon} \hat{g}_m^\varepsilon(k) w_m(x,k) \Re\big(e^{\mathrm{i}t\sqrt{\mu_m^\varepsilon(k)}}\big)\,\mathrm{d}k \Big\|_{\mathrm{L}^\infty(0,\varepsilon^{-2}T;\mathrm{L}^2(\mathbb{R}^d))} \leq C\varepsilon.$$

In particular, for the homogenization process the only relevant part in the expansion (4.14) is the term $m = 0$. As the rest of the derivation is done rather formally in [85] (and ends with the ill-posed equation (4.17)), we now exclusively follow the derivation in [42, 43]. The approximation

$$U^\varepsilon(t,x) = (2\pi)^{-d/2} \int_K G(k) e^{\mathrm{i}k\cdot x} \Re\big(e^{\mathrm{i}t\sqrt{\mu_0^\varepsilon(k)}}\big)\,\mathrm{d}k,$$

where G is the Fourier transform of g and K is its support, is proved to satisfy the error estimate

$$\|u^\varepsilon - U^\varepsilon\|_{\mathrm{L}^\infty(0,\infty;\mathrm{L}^2(\mathbb{R}^d)+\mathrm{L}^\infty(\mathbb{R}^d))} \leq C\varepsilon. \tag{4.15}$$

The next step is the approximation of $\Re\big(e^{\mathrm{i}t\sqrt{\mu_0^\varepsilon(k)}}\big)$ using Taylor expansion. In particular,

$$\mu_0^\varepsilon(k) = A_{ij}k_ik_j + \varepsilon^2 C_{ijmn}k_ik_jk_mk_n + \mathcal{O}(\varepsilon^4),$$

where $A_{ij} = \partial^2_{ij}\mu_0(0)$ and $C_{ijmn} = \partial^4_{ijmn}\mu_0(0)$, and we obtain the approximation

$$v^\varepsilon(t,x) = (2\pi)^{-d/2}\frac{1}{2}\sum_{\pm}\int_K G(k)e^{\mathrm{i}k\cdot x}\exp\Big(\pm \mathrm{i}t\sqrt{A_{ij}k_ik_j}\Big)\exp\Big(\pm\frac{\mathrm{i}\varepsilon^2 t}{2}\frac{C_{ijmn}k_ik_jk_mk_n}{\sqrt{A_{ij}k_ik_j}}\Big).$$

The function v^ε satisfies the error estimate

$$\|U^\varepsilon - v^\varepsilon\|_{\mathrm{L}^\infty(0,\varepsilon^{-2}T;\mathrm{L}^2(\mathbb{R}^d)+\mathrm{L}^\infty(\mathbb{R}^d))} \leq C\varepsilon. \tag{4.16}$$

As shown in [43], it holds in fact $A_{ij} = \partial^2_{ij}\mu_0(0) = a^0_{ij}$, where a^0 is the homogenized tensor defined in (4.41). Hence, v^ε satisfies

$$\partial_t^2 v^\varepsilon = a^0_{ij}\partial^2_{ij}v^\varepsilon - \varepsilon^2 C_{ijmn}\partial^4_{ijmn}v^\varepsilon - \varepsilon^4(C_{ijmn}k_ik_jk_mk_n)^2/(4a^0_{ij}k_ik_j)v^\varepsilon.$$

However, C being negative, the equation

$$\partial_t^2 v = a^0_{ij}\partial^2_{ij}v - \varepsilon^2 C_{ijmn}\partial^4_{ijmn}v \tag{4.17}$$

is ill-posed and cannot be used. Next, [42, 43] gives an algebraic procedure to build $E \in \mathrm{Ten}^2(\mathbb{R}^d), F \in \mathrm{Ten}^4(\mathbb{R}^d)$ that satisfy some symmetry and sign assumption (to ensure the well-posedness of (4.12), see (4.55)) and such that the following decomposition holds:

$$-C_{ijmn}\partial^4_{ijmn} = \big(E_{ij}\partial^2_{ij}\big)\big(a^0_{mn}\partial^2_{mn}\big) - F_{ijmn}\partial^4_{ijmn}. \tag{4.18}$$

We observe that the decomposition (4.18) is a preparation for a Boussinesq trick, i.e., to use the effective equation to replace the operator $a^0_{mn}\partial^2_{mn}$ with ∂_t^2 (plus a higher order error term). Finally, it is proved in [42, 43] that the solution w^ε of the (well-posed) equation (4.12) satisfies $\|\nabla(v^\varepsilon - w^\varepsilon)\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\mathbb{R}^d))} \leq C\varepsilon^2$, which combined with (4.15), (4.16) proves the error estimate (4.13).

4.1.2 Derivation in one dimension of an effective equation via asymptotic expansion

The second framework used for the derivation of effective equations is the technique of asymptotic expansion, that we introduced in Section 3.3. This technique was used by Lamacz in [72] to derive

an effective equation in one-dimension. Chronologically, [72] situates after the formal derivation via Bloch waves of [85] and before the rigorous one in [42, 43]. This derivation from [72] and the proof of the error estimate were the starting point of the results presented in the current chapter.

We summarize here the result from [72] and present the main ideas of the derivation of the effective equation. Given a Y-periodic, elliptic, bounded tensor $a \in \mathcal{C}^\infty_{\text{per}}(Y)$, we consider the solution $u^\varepsilon : [0, \varepsilon^{-2}T] \times \mathbb{R} \to \mathbb{R}$ of

$$\begin{aligned} &\partial_t^2 u^\varepsilon(t,x) - \partial_x\big(a\big(\tfrac{x}{\varepsilon}\big)\partial_x u^\varepsilon(t,x)\big) = 0 && \text{in } (0,\varepsilon^{-2}T]\times\mathbb{R},\\ &u^\varepsilon(0,x) = g^0(x) + \varepsilon\chi\big(\tfrac{x}{\varepsilon}\big)\partial_x g^0(x) + \varepsilon^2\theta\big(\tfrac{x}{\varepsilon}\big)\partial_x^2 g^0(x), && \text{in } \mathbb{R},\\ &\partial_t u^\varepsilon(0,x) = g^1(x) + \varepsilon\chi\big(\tfrac{x}{\varepsilon}\big)\partial_x g^1(x), && \text{in } \mathbb{R}, \end{aligned} \tag{4.19}$$

where χ, θ are Y-periodic solution of given cell problems (the same as in Section 3.3). The effective equation is given as $w^\varepsilon : [0, \varepsilon^{-2}T] \times \mathbb{R} \to \mathbb{R}$ such that

$$\begin{aligned} &\partial_t^2 w^\varepsilon(t,x) - a^0\partial_x^2 w^\varepsilon(t,x) - \frac{a^2}{a^0}\partial_x^2\partial_t^2 w^\varepsilon(t,x) = 0 && \text{in } (0,\varepsilon^{-2}T]\times\mathbb{R},\\ &w^\varepsilon(0,x) = g^0(x) && \text{in } \mathbb{R},\\ &\partial_t w^\varepsilon(0,x) = g^1(x) && \text{in } \mathbb{R}, \end{aligned} \tag{4.20}$$

where a^0 is the homogenized coefficient and a^2 is defined via a cascade of cell problems, depending on $a(y)$. The following convergence is then proved.

Theorem 4.1.2 (Lamacz, [72])**.** *Assume that $g^0, g^1 \in \mathcal{C}^\infty_c(-R,R)$ for some $R > 0$ and $\int_{-R}^R g^1(x)\,\mathrm{d}x = 0$. Then the following convergence holds*

$$\lim_{\varepsilon\to 0} \|u^\varepsilon - w^\varepsilon\|_{\mathrm{L}^\infty(0,\varepsilon^{-2}T;\mathrm{L}^\infty(\mathbb{R}))} = 0,$$

where the limit makes sense through the change of variables $\tau = \varepsilon^2 t$, i.e.

$$\lim_{\varepsilon\to 0} \|u^\varepsilon - w^\varepsilon\|_{\mathrm{L}^\infty(0,\varepsilon^{-2}T;\mathrm{L}^\infty(\mathbb{R}))} = \lim_{\varepsilon\to 0} \|u^\varepsilon(\cdot/\varepsilon^2,\cdot) - w^\varepsilon(\cdot/\varepsilon^2,\cdot)\|_{\mathrm{L}^\infty(0,T;\mathrm{L}^\infty(\mathbb{R}))} = 0.$$

Theorem 4.1.2 is the first rigorous result asserting that the long time dispersive effects could be modeled by a well-posed equation. The result, however, requires prepared initial conditions in (4.19). This issue is in fact connected to the convergence of the energy (and to the correctors problem), discussed in Sections 3.2.2 and 3.3.3. In our result, in Section 4.2, we overcome this difficulty by weakening the norm of the error estimate.

Let us summarize how (4.20) is obtained. The main step is the adaptation of a (sufficiently regular) function v to the micro structure of the medium. This is done through the definition of a linear, time independent adaptation operator $\mathcal{B}^\varepsilon$ that satisfies

$$-\partial_x\big(a\big(\tfrac{\cdot}{\varepsilon}\big)\partial_x(\mathcal{B}^\varepsilon v)\big) = -\mathcal{B}^\varepsilon\Big(\sum_{i=0}^3 \varepsilon^i a^i \partial_x^{2+i} v\Big) + R^\varepsilon v,$$

with a remainder $R^\varepsilon v$ that is sufficiently small and where $\{a^i\}_{i=0}^3$ are constant coefficients (based on the tensor a). The construction of the adaptation operator $\mathcal{B}^\varepsilon$ relies on asymptotic expansion. Its role is clarified by the following observation: if v^ε satisfies

$$\partial_t^2 v^\varepsilon(t,x) - \sum_{i=0}^3 \varepsilon^i a^i \partial_x^{2+i} v^\varepsilon(t,x) = 0, \tag{4.21}$$

then $\mathcal{B}^\varepsilon v^\varepsilon$ solves

$$\partial_t^2(\mathcal{B}^\varepsilon v^\varepsilon) - \partial_x\big(a\big(\tfrac{\cdot}{\varepsilon}\big)\partial_x(\mathcal{B}^\varepsilon v^\varepsilon)\big) = \mathcal{B}^\varepsilon\Big(\partial_t^2 v^\varepsilon - \sum_{i=0}^3 \varepsilon^i a^i \partial_x^{2+i} v^\varepsilon\Big) + R^\varepsilon v^\varepsilon = R^\varepsilon v^\varepsilon,$$

which ensures by energy techniques that $\|\partial_x(u^\varepsilon - \mathcal{B}^\varepsilon v^\varepsilon)\|_{\mathrm{L}^\infty(\mathrm{L}^2)}$ is small. However, [72] proves that the coefficients satisfy $a^0 > 0$, $a^1 = a^3 = 0$, and $a^2 \geq 0$. Hence, the same issue as in [85] is encountered: (4.21) is ill-posed due to the sign of a^2. To overcome the problem, [72] uses a Boussinesq trick to transform (4.21) into a well-posed equation. Namely, using (4.21) at order $\mathcal{O}(1)$, $\partial_x^2 v^\varepsilon$ is replaced by $\partial_t^2 v^\varepsilon / a^0$ so that v^ε satisfies

$$\varepsilon^2 a^2 \partial_x^4 v^\varepsilon(t,x) = \varepsilon^2 \frac{a^2}{a^0} \partial_x^2 \partial_t^2 v^\varepsilon(t,x) + \mathcal{O}(\varepsilon^4).$$

Combined to (4.21), this equality leads to the well-posed equation (4.20).

4.2 A new family of effective equations for long time wave propagation

In the previous section, we presented different derivations of effective equations for the long time homogenization of the wave equation. In particular, [72] and [42, 43] define effective equations and rigorously prove their validity with error estimates (in the one-dimensional case for [72] and in the general case in [42, 43]). In this section, we present the first main contribution of this thesis. We define a family of effective equations and prove an error estimate ensuring that any element of the family is ε-close to u^ε over a time interval of length $\varepsilon^{-2}T$ (see Section 4.2.4, Theorem 4.2.4).

Let us mention a fundamental difference between the results in [72] and [42, 43] and the result of this chapter. The error estimate in Theorem 4.2.4 holds in arbitrarily large periodic domains Ω, while [42, 43] deals with the whole space $\mathbb{R}^d$ (and [72] with $\mathbb{R}$). Nevertheless, as the dependence of our error estimate on the domain is explicit, our result can be compared with [42, 43]. In particular, we prove that the effective equations defined in [72] and [42, 43] belong to the family of effective equations.

The settings of our result, compared to [72] and [42, 43], are more general on the following aspects. First, we do not require prepared initial data as in [72], and, contrarily to [42, 43], we allow for a non zero initial speed. Second, we allow for a source term, which is neither the case in [72] nor in [42, 43]. Third, we obtain a result for a tensor with minimal regularity, $a \in [\mathrm{L}^\infty(Y)]^{d\times d}$ (Section 4.2.6), while it is required to be of class $\mathcal{C}^1$ in [42, 43] and $\mathcal{C}^\infty$ in [72]. Fourth, we provide a numerical procedure that is significantly cheaper than in [42, 43] and [72]. Indeed, in the one-dimensional case, we show that solving 1 single cell problem is sufficient to compute an effective equation, whereas [72] requires 3. And in the multidimensional case, our algorithm requires the solutions of $d + \binom{d+1}{2}$ cell problems, while $d + \binom{d+1}{2} + \binom{d+2}{3}$ are necessary in [42, 43].

Let us define the norms that are involved in the main result (see Appendix A.1 for the details). Recall that we denote the quotient space $\mathcal{L}^2(\Omega) = \mathrm{L}^2(\Omega)/\mathbb{R}$. A bracket $[\![v]\!]$ is used to denote the equivalence class of $v \in \mathrm{L}^2(\Omega)$ in $\mathcal{L}^2(\Omega)$. Furthermore, we denote $\mathcal{W}_{\mathrm{per}}(\Omega) = \mathrm{H}^1_{\mathrm{per}}(\Omega)/\mathbb{R}$ and a bold face letter $\boldsymbol{v}$ is used to denote the elements of $\mathcal{W}_{\mathrm{per}}(\Omega)$. The space $\mathrm{W}_{\mathrm{per}}(\Omega)$ is composed of the zero mean representatives of the equivalence classes in $\mathcal{W}_{\mathrm{per}}(\Omega)$. We define the following norm on $\mathcal{W}_{\mathrm{per}}(\Omega)$

$$\|\boldsymbol{w}\|_{\mathcal{W}} = \inf_{\substack{\boldsymbol{w}=\boldsymbol{w}_1+\boldsymbol{w}_2 \\ \boldsymbol{w}_i=[\![w_i]\!]\in\mathcal{W}_{\mathrm{per}}(\Omega)}} \left\{ \|[\![w_1]\!]\|_{\mathcal{L}^2(\Omega)} + \|\nabla w_2\|_{\mathrm{L}^2(\Omega)} \right\} \quad \forall \boldsymbol{w} \in \mathcal{W}_{\mathrm{per}}(\Omega), \tag{4.22}$$

and the corresponding norm on $\mathrm{W}_{\mathrm{per}}(\Omega)$

$$\|w\|_W = \inf_{\substack{w=w_1+w_2 \\ w_1,w_2\in\mathrm{W}_{\mathrm{per}}(\Omega)}} \left\{ \|w_1\|_{\mathrm{L}^2(\Omega)} + \|\nabla w_2\|_{\mathrm{L}^2(\Omega)} \right\} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(\Omega). \tag{4.23}$$

We verify that a function $w \in \mathrm{W}_{\mathrm{per}}(\Omega)$ satisfies $\|w\|_W = \|[\![w]\!]\|_{\mathcal{W}}$. Furthermore, using the Poincaré–Wirtinger inequality, we verify that $\|\cdot\|_W$ is equivalent to the L^2 norm:

$$\|w\|_W \leq \|w\|_{\mathrm{L}^2(\Omega)} \leq \max\{1, C_\Omega\}\|w\|_W \quad \forall w \in \mathrm{W}_{\mathrm{per}}(\Omega), \tag{4.24}$$

where C_Ω is the Poincaré constant.

4.2.1 The wave equation in an arbitrarily large periodic domain

We introduce here the precise settings of our result. Recall that in applications, we do not approximate the wave equation in the full space $\mathbb{R}^d$, but in a pseudoinfinite domain. (a sufficiently large periodic domain so that the waves never reach the boundaries). We thus assume that Ω is an arbitrarily large hypercube in $\mathbb{R}^d$. We emphasize that in the long time analysis, we track the influence of the size of Ω on the error estimates.

Let $\Omega, Y \in \mathbb{R}^d$ be open hypercubes such that Ω is a union of cells of volume $\varepsilon|Y|$, as in Figure 4.2. More precisely, letting $\ell \in \mathbb{R}^d$ be the period of the tensor a, i.e., $a(y + k \cdot \ell) = a(y)$ for all $y \in Y$ and $k \in \mathbb{Z}^d$, we assume that $\Omega = (\omega_1^l, \omega_1^r) \times \cdots \times (\omega_d^l, \omega_d^r)$ satisfies

$$\frac{\omega_i^r - \omega_i^l}{\varepsilon \ell_i} \in \mathbb{N}_{>0} \quad \forall i = 1, \dots, d. \tag{4.25}$$

Assumption (4.25) ensures that for any Y-periodic function γ, the map $x \mapsto \gamma\left(\frac{x}{\varepsilon}\right)$ is Ω-periodic (γ is extended to $\mathbb{R}^d$ by periodicity). In particular, as we assume a to be Y-periodic, $a\left(\frac{\cdot}{\varepsilon}\right)$ is Ω-periodic.

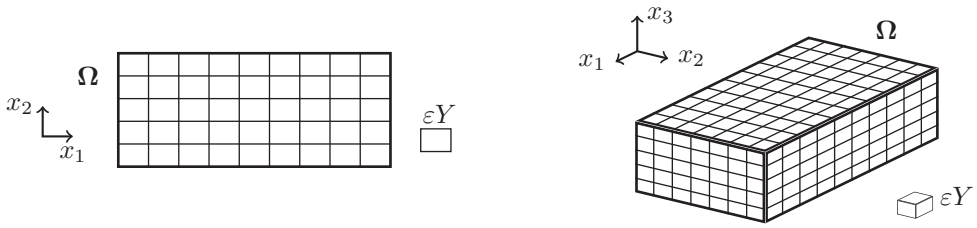

Figure 4.2: The hypercube Ω is assumed to be a union of unit cells of volume $\varepsilon|Y|$ (on the left $d = 2$, on the right $d = 3$).

For $T^\varepsilon = \varepsilon^{-2} T$, we consider the wave equation: find $u^\varepsilon : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ such that

$$\begin{array}{ll} \partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \left(a\left(\frac{x}{\varepsilon}\right) \nabla_x u^\varepsilon(t,x)\right) = f(t,x) & \text{in } (0, T^\varepsilon] \times \Omega, \\ x \mapsto u^\varepsilon(t,x) \ \ \Omega\text{-periodic} & \text{in } [0, T^\varepsilon], \\ u^\varepsilon(0,x) = g^0(x), \quad \partial_t u^\varepsilon(0,x) = g^1(x) & \text{in } \Omega, \end{array} \tag{4.26}$$

where g^0, g^1 are given initial conditions and f is a source. The following notation is used for the differential operator $\mathcal{A}^\varepsilon = -\nabla_x \cdot \left(a\left(\frac{x}{\varepsilon}\right) \nabla_x \cdot \right)$. We assume that $a \in [\mathrm{L}^\infty_{\mathrm{per}}(Y)]^{d \times d}$ is symmetric, uniformly elliptic and bounded, i.e. there exists $\lambda, \Lambda > 0$ such that

$$\lambda |\xi|^2 \le a(y)\xi \cdot \xi \le \Lambda |\xi|^2 \ \text{ for a.e. } y \in Y \ \ \forall \xi \in \mathbb{R}^d. \tag{4.27}$$

The well-posedness of (4.26) is proved in Section 2.1.1. If $g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega)$, $g^1 \in \mathrm{L}^2_0(\Omega)$, and $f \in \mathrm{L}^2(0, T^\varepsilon; \mathrm{L}^2_0(\Omega))$, then there exists a unique weak solution $u^\varepsilon \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{W}_{\mathrm{per}}(\Omega))$ with $\partial_t u^\varepsilon \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{L}^2_0(\Omega))$ and $\partial_t^2 u^\varepsilon \in \mathrm{L}^2(0, T^\varepsilon; \mathrm{W}^*_{\mathrm{per}}(\Omega))$.

4.2.2 An energy estimate to motivate asymptotic expansion

In this section, we explain how using asymptotic expansion we construct an adaptation operator that can be used to rigorously prove an error estimate for the long time homogenization of the

wave equation. Recall that this approach was used in section 3.3.2 to prove an error estimate for the homogenized equation at timescales $\mathcal{O}(1)$. We explain here the approach in the case of an arbitrary timescale $\mathcal{O}(\varepsilon^{-\alpha})$, where α is a non-negative integer. In particular, we establish the connection between the timescale α and the accuracy that has to be attained by the adaptation.

We start by proving an energy estimate for a function satisfying the wave equation (with a right hand side in $\mathcal{W}^*_{\rm per}(\Omega)$). We emphasize that the constant in the estimate neither depends on the length of the time interval nor on the domain.

Lemma 4.2.1. *Let $\tau > 0$ and assume that $\boldsymbol{\eta} \in \mathrm{L}^\infty(0,\tau;\mathcal{W}_{\rm per}(\Omega))$, with $\partial_t\boldsymbol{\eta} \in \mathrm{L}^\infty(0,\tau;\mathcal{L}^2(\Omega))$, $\partial_t^2\boldsymbol{\eta} \in \mathrm{L}^2(0,\tau;\mathcal{W}^*_{\rm per}(\Omega))$ satisfies*

$$\begin{aligned} &\partial_t^2\boldsymbol{\eta}(t) + \mathcal{A}^\varepsilon\boldsymbol{\eta}(t) = \boldsymbol{r}(t) \quad \text{in } \mathcal{W}^*_{\rm per}(\Omega) \quad \text{for a.e. } t \in [0,\tau], \\ &\boldsymbol{\eta}(0) = \boldsymbol{\eta}^0, \quad \partial_t\boldsymbol{\eta}(0) = \boldsymbol{\eta}^1, \end{aligned} \tag{4.28}$$

where $\boldsymbol{\eta}^0 \in \mathcal{W}_{\rm per}(\Omega)$, $\boldsymbol{\eta}^1 \in \mathcal{L}^2(\Omega)$ and $\boldsymbol{r} \in \mathrm{L}^2(0,\tau;\mathcal{W}^*_{\rm per}(\Omega))$ is given as

$$\big\langle \boldsymbol{r}(t), \boldsymbol{w}\big\rangle_{\mathcal{W}^*_{\rm per}(\Omega),\mathcal{W}_{\rm per}(\Omega)} = \big(\boldsymbol{r}_0(t), \boldsymbol{w}\big)_{\mathcal{L}^2(\Omega)} + \big(r_1(t), \nabla\boldsymbol{w}\big)_{\mathrm{L}^2(\Omega)},$$

with $\boldsymbol{r}_0 \in \mathrm{L}^2(0,\tau;\mathcal{L}^2(\Omega))$ and $r_1 \in [\mathrm{L}^2(0,\tau;\mathrm{L}^2(\Omega))]^d$. Then the following estimate holds

$$\|\boldsymbol{\eta}\|_{\mathrm{L}^\infty(0,\tau;\mathcal{W})} \le C(\lambda)\big(\|\boldsymbol{\eta}^1\|_{\mathcal{L}^2(\Omega)} + \|\boldsymbol{\eta}^0\|_{\mathcal{L}^2(\Omega)} + \|\boldsymbol{r}_0\|_{\mathrm{L}^1(0,\tau;\mathcal{L}^2(\Omega))} + \|r_1\|_{\mathrm{L}^1(0,\tau;\mathrm{L}^2(\Omega))}\big), \tag{4.29}$$

where $C(\lambda)$ depends only on the ellipticity constant λ and the norm $\|\cdot\|_{\mathcal{W}}$ is defined in (4.22).

Proof. We denote $\langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle_{\mathcal{W}^*_{\rm per},\mathcal{W}_{\rm per}}$ and $\mathrm{L}^p(X) = \mathrm{L}^p(0,\tau;X)$. Define $\boldsymbol{\zeta}_2 \in \mathrm{L}^\infty(\mathcal{W}_{\rm per}(\Omega))$, with $\partial_t\boldsymbol{\zeta}_2 \in \mathrm{L}^\infty(\mathcal{L}^2(\Omega))$, $\partial_t^2\boldsymbol{\zeta}_2 \in \mathrm{L}^2(\mathcal{W}^*_{\rm per}(\Omega))$ as the unique solution of the equation

$$\begin{aligned} &\big\langle \partial_t^2\boldsymbol{\zeta}_2(t) + \mathcal{A}^\varepsilon\boldsymbol{\zeta}_2(t), \boldsymbol{w}\big\rangle = \big(\boldsymbol{r}_0(t), \boldsymbol{w}\big)_{\mathcal{L}^2} \quad \forall \boldsymbol{w} \in \mathcal{W}_{\rm per}(\Omega) \quad \text{for a.e. } t \in [0,\tau], \\ &\boldsymbol{\zeta}_2(0) = \big[0\big], \quad \partial_t\boldsymbol{\zeta}_2(0) = \boldsymbol{\eta}^1. \end{aligned}$$

The standard well-posedness of the wave equation ensures the existence and uniqueness of $\boldsymbol{\zeta}_2$ (see Theorem 2.1.1). Furthermore, $\boldsymbol{\zeta}_2$ satisfies the energy estimate

$$\tfrac{1}{2}\|\partial_t\boldsymbol{\zeta}_2\|^2_{\mathrm{L}^\infty(\mathcal{L}^2)} + \lambda\|\nabla\boldsymbol{\zeta}_2\|^2_{\mathrm{L}^\infty(\mathrm{L}^2)} \le 2\|\boldsymbol{\eta}^1\|^2_{\mathcal{L}^2} + 4\|\boldsymbol{r}_0\|^2_{\mathrm{L}^1(\mathcal{L}^2)}. \tag{4.30}$$

We define then $\boldsymbol{\zeta}_1 = \boldsymbol{\eta} - \boldsymbol{\zeta}_2$ and verify that $\boldsymbol{\zeta}_1$ satisfies

$$\boldsymbol{\zeta}_1 \in \mathrm{L}^\infty(\mathcal{W}_{\rm per}(\Omega)), \quad \partial_t\boldsymbol{\zeta}_1 \in \mathrm{L}^\infty(\mathcal{L}^2(\Omega)), \quad \partial_t^2\boldsymbol{\zeta}_1 \in \mathrm{L}^2(\mathcal{W}^*_{\rm per}(\Omega)),$$

and

$$\begin{aligned} &\big\langle \partial_t^2\boldsymbol{\zeta}_1(t) + \mathcal{A}^\varepsilon\boldsymbol{\zeta}_1(t), \boldsymbol{w}\big\rangle = \big(r_1(t), \nabla\boldsymbol{w}\big)_{\mathrm{L}^2} \quad \forall \boldsymbol{w} \in \mathcal{W}_{\rm per}(\Omega) \quad \text{for a.e. } t \in [0,\tau], \\ &\boldsymbol{\zeta}_1(0) = \boldsymbol{\eta}^0, \quad \partial_t\boldsymbol{\zeta}_1(0) = \big[0\big]. \end{aligned} \tag{4.31}$$

For all $t \in [0,\tau]$, let $\hat{\boldsymbol{v}}(t) = (\mathcal{A}^\varepsilon)^{-1}\partial_t\boldsymbol{\zeta}_1(t) \in \mathcal{W}_{\rm per}(\Omega)$, i.e.,

$$\big(a^\varepsilon\nabla\hat{\boldsymbol{v}}(t), \nabla\boldsymbol{w}\big)_{\mathrm{L}^2} = \big\langle \partial_t\boldsymbol{\zeta}_1(t), \boldsymbol{w}\big\rangle \quad \forall \boldsymbol{w} \in \mathcal{W}_{\rm per}(\Omega).$$

Note that the existence and uniqueness of $\hat{\boldsymbol{v}}(t)$ are ensured by Lax–Milgram theorem. In particular, $\hat{\boldsymbol{v}}(t)$ satisfies the estimate

$$\lambda\|\nabla\hat{\boldsymbol{v}}(t)\|^2_{\mathrm{L}^2} \le \big(a^\varepsilon\nabla\hat{\boldsymbol{v}}(t), \nabla\hat{\boldsymbol{v}}(t)\big)_{\mathrm{L}^2} = \big\langle \partial_t\boldsymbol{\zeta}_1(t), \hat{\boldsymbol{v}}(t)\big\rangle. \tag{4.32}$$

Using $\hat{\boldsymbol{v}}(t)$ as a test function in (4.31), we get

$$\big\langle \partial_t^2\boldsymbol{\zeta}_1(t), (\mathcal{A}^\varepsilon)^{-1}\partial_t\boldsymbol{\zeta}_1(t)\big\rangle + \big\langle \mathcal{A}^\varepsilon\boldsymbol{\zeta}_1(t), (\mathcal{A}^\varepsilon)^{-1}\partial_t\boldsymbol{\zeta}_1(t)\big\rangle = \big(r_1(t), \nabla\hat{\boldsymbol{v}}(t)\big)_{\mathrm{L}^2}.$$

Thanks to the symmetry of $\mathcal{A}^\varepsilon$, this equality can be rewritten as

$$\tfrac{1}{2}\tfrac{\mathrm{d}}{\mathrm{d}t}\Big(\big\langle\partial_t\boldsymbol{\zeta}_1(t),\hat{\boldsymbol{v}}(t)\big\rangle+\|\boldsymbol{\zeta}_1(t)\|^2_{\mathcal{L}^2}\Big)=\big(r_1(t),\nabla\hat{\boldsymbol{v}}(t)\big)_{\mathrm{L}^2}.$$

For $\xi\in[0,\tau]$, we integrate this equality over $[0,\xi]$ and, using (4.32), we obtain

$$\lambda\|\nabla\hat{\boldsymbol{v}}(\xi)\|^2_{\mathrm{L}^2}\le\big\langle\partial_t\boldsymbol{\zeta}_1(\xi),\hat{\boldsymbol{v}}(\xi)\big\rangle+\|\boldsymbol{\zeta}_1(\xi)\|^2_{\mathcal{L}^2}=\|\boldsymbol{\eta}^0\|^2_{\mathcal{L}^2}+2\int_0^\xi\big(r_1(t),\nabla\hat{\boldsymbol{v}}(t)\big)_{\mathrm{L}^2}\,\mathrm{d}t. \tag{4.33}$$

Using Cauchy–Schwartz, Hölder and Young inequalities, we bound the second term of the right hand side as

$$2\int_0^\xi\big(r_1(t),\nabla\hat{\boldsymbol{v}}(t)\big)_{\mathrm{L}^2}\,\mathrm{d}t\le 2\|r_1\|_{\mathrm{L}^1(\mathrm{L}^2)}\|\nabla\hat{\boldsymbol{v}}\|_{\mathrm{L}^\infty(\mathrm{L}^2)}\le 2\lambda^{-1}\|r_1\|^2_{\mathrm{L}^1(\mathrm{L}^2)}+\tfrac{1}{2}\lambda\|\nabla\hat{\boldsymbol{v}}\|^2_{\mathrm{L}^\infty(\mathrm{L}^2)}. \tag{4.34}$$

Taking now the L^∞ norm with respect to ξ in (4.33) and using (4.34), we obtain the estimate

$$\tfrac{1}{2}\lambda\|\nabla\hat{\boldsymbol{v}}\|^2_{\mathrm{L}^\infty(\mathrm{L}^2)}\le\|\boldsymbol{\eta}^0\|^2_{\mathcal{L}^2}+2\lambda^{-1}\|r_1\|^2_{\mathrm{L}^1(\mathrm{L}^2)}. \tag{4.35}$$

Finally, we combine (4.33), (4.34) and (4.35) and obtain the following bound

$$\|\boldsymbol{\zeta}_1\|^2_{\mathrm{L}^\infty(\mathcal{L}^2)}\le 2\|\boldsymbol{\eta}^0\|^2_{\mathcal{L}^2}+4\lambda^{-1}\|r_1\|^2_{\mathrm{L}^1(\mathrm{L}^2)}. \tag{4.36}$$

Recalling the definition of the norm $\|\cdot\|_{\mathcal{W}}$ in (4.22) and as $\boldsymbol{\eta}=\boldsymbol{\zeta}_1+\boldsymbol{\zeta}_2$, the combination of estimates (4.30) and (4.36) gives (4.29) and the proof of the lemma is complete. □

Let us now explain what is the asymptotic expansion and how, with Lemma 4.2.1, it leads to a rigorous error estimate between u^ε and an effective solution in the $\mathrm{L}^\infty(0,\varepsilon^{-2}T;W)$ norm . In order to have a better understanding of the influence of the time interval, let us consider the equation (4.26) on the time interval $[0,\varepsilon^{-\alpha}T]$, where $\alpha>0$. We consider a candidate effective solution $\tilde{u}$. The form of the effective equation is an ansatz (discussed in the next section). Typically, $\tilde{u}$ solves an equation composed of the homogenized equation plus some higher order corrections operators, which depend on ε and whose tensors have to be defined (hence $\tilde{u}$ depends on ε). Let us then assume that $\tilde{u}$ is sufficiently smooth and satisfies $\langle\tilde{u}\rangle_\Omega=\langle u^\varepsilon\rangle_\Omega$ and the energy bounds $\sum_{k=1}^{k_n}|\partial_t^n\tilde{u}(t)|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathrm{H}^k(\Omega))}\le C$, independently of ε, for $n=0,1,2$ and for some k_n.

The asymptotic expansion is a technique to build an adaptation of $\tilde{u}$ of the form $\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}=\big[\tilde{u}\big]+\boldsymbol{\mathcal{C}}^\varepsilon\tilde{u}$. We first require the adaptation to satisfy the following conditions

$$\begin{gathered}\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}\in\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathcal{W}_{\mathrm{per}}(\Omega)),\ \partial_t\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}\in\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathcal{L}^2(\Omega)),\\ \partial_t^2\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}\in\mathrm{L}^2(0,\varepsilon^{-\alpha}T;\mathcal{W}^*_{\mathrm{per}}(\Omega)),\end{gathered} \tag{4.37a}$$

$$\|\boldsymbol{\mathcal{C}}^\varepsilon\tilde{u}\|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathcal{W})}\le C\varepsilon\sum_{k=1}^{k_0}|\tilde{u}|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathrm{H}^k(\Omega))}, \tag{4.37b}$$

$$\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}(0)=\big[u^\varepsilon(0)\big]+C\varepsilon,\quad\partial_t\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}(0)=\big[\partial_t u^\varepsilon(0)\big]+C\varepsilon. \tag{4.37c}$$

Next, we have to determine how accurately $\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}$ must approximate u^ε. The Hölder inequality leads to the following corollary of Lemma 4.2.1.

Corollary 4.2.2. *If in Lemma 4.2.1 $\tau=\varepsilon^{-\alpha}T$ and*

$$\boldsymbol{r}_0\in\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathcal{L}^2(\Omega)),\qquad r_1\in[\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathrm{L}^2(\Omega))]^d,$$

then $\boldsymbol{\eta}$ satisfies the estimate

$$\begin{aligned}\|\boldsymbol{\eta}\|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathcal{W})}\le C\big(&\|\boldsymbol{\eta}^1\|_{\mathcal{L}^2(\Omega)}+\|\boldsymbol{\eta}^0\|_{\mathcal{L}^2(\Omega)}+\varepsilon^{-\alpha}\|\boldsymbol{r}_0\|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathcal{L}^2(\Omega))}\\&+\varepsilon^{-\alpha}\|r_1\|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathrm{L}^2(\Omega))}\big),\end{aligned}$$

where C depends only on λ and T.

Hence, the last condition on the adaptation is to require that the remainder

$$\big(\boldsymbol{r}_0(t),\boldsymbol{w}\big)_{\mathcal{L}^2(\Omega)}+\big(r_1(t),\nabla\boldsymbol{w}\big)_{\mathrm{L}^2(\Omega)}=\big\langle(\partial_t^2+\mathcal{A}^\varepsilon)(\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}-\big[u^\varepsilon\big])(t),\boldsymbol{w}\big\rangle_{\mathcal{W}_{\mathrm{per}}^*,\mathcal{W}_{\mathrm{per}}}$$

satisfies

$$\|\boldsymbol{r}_0\|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathcal{L}^2(\Omega))}+\|r_1\|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathrm{L}^2(\Omega))}\leq C\varepsilon^\gamma\sum_{n=0}^{2}\sum_{k=1}^{k_n}|\partial_t^n\tilde{u}|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathrm{H}^k(\Omega))},\tag{4.38}$$

for some $\gamma>\alpha$. If conditions (4.37) and (4.38) are met, we can prove the error estimate as follows. As $(u^\varepsilon-\tilde{u})(t)\in\mathrm{W}_{\mathrm{per}}(\Omega)$, we have $\|u^\varepsilon-\tilde{u}\|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;W)}=\|\big[u^\varepsilon-\tilde{u}\big]\|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathcal{W})}$. Hence, using the triangle inequality gives

$$\|u^\varepsilon-\tilde{u}\|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;W)}\leq\|\big[u^\varepsilon\big]-\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}\|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathcal{W})}+\|\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}-\big[\tilde{u}\big]\|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathcal{W})}.$$

The second term of the right hand side is bounded using (4.37b). For the first term, we apply Corollary 4.2.2 to $\boldsymbol{\eta}=\big[u^\varepsilon\big]-\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}$ and, using the properties of $\tilde{u}$, (4.37c) and (4.38), we obtain the error estimate

$$\|u^\varepsilon-\tilde{u}\|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;W)}\leq C\varepsilon^{\min\{1,\gamma-\alpha\}}.\tag{4.39}$$

4.2.3 Asymptotic expansion and constraints on the effective tensors

Our goal is now to construct an adaptation satisfying the requirements presented in the previous section. The construction is done with asymptotic expansion. In particular, we derive cell problems for the definition of the correctors that constitute the adaptation. In addition, the well-posedness of these cell problems provides constraints on the effective tensors thus characterizing the effective equations. All the computations are done formally, i.e., we assume as much regularity as required. The rigorous result with its detailed proof is presented in the next section.

We are looking for an effective solution on a time interval $[0,T^\varepsilon]$, with $T^\varepsilon=\varepsilon^{-2}T$. As discussed in the previous section, we thus need to construct an adaptation $\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}(t)$ such that $(\partial_t^2+\mathcal{A}^\varepsilon)(\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}-\big[u^\varepsilon\big])(t)=\mathcal{O}(\varepsilon^3)$ for a.e. $t\in[0,T^\varepsilon]$. In what follows, we construct $\mathcal{B}^\varepsilon\tilde{u}(t)\in\mathrm{H}^1_{\mathrm{per}}(\Omega)$, such that $(\partial_t^2+\mathcal{A}^\varepsilon)(\mathcal{B}^\varepsilon\tilde{u}-u^\varepsilon)(t)=\mathcal{O}(\varepsilon^3)$ and we will then set $\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}=\big[\mathcal{B}^\varepsilon\tilde{u}\big]$ in $\mathcal{W}^*_{\mathrm{per}}(\Omega)$.

First, we introduce the effective solution $\tilde{u}$. Referring to [81, 43, 42, 13], we make the ansatz that the effective equation is of the form

$$\begin{array}{ll}\partial_t^2\tilde{u}-a^0_{ij}\partial^2_{ij}\tilde{u}+\varepsilon^2\big(a^2_{ijkl}\partial^4_{ijkl}\tilde{u}-b^2_{ij}\partial^2_{ij}\partial^2_t\tilde{u}\big)=f & \text{in } (0,T^\varepsilon]\times\Omega,\\ x\mapsto\tilde{u}(t,x)\ \ \Omega\text{-periodic} & \text{in } [0,T^\varepsilon],\\ \tilde{u}(0,x)=g^0(x),\quad \partial_t\tilde{u}(0,x)=g^1(x) & \text{in } \Omega,\end{array}\tag{4.40}$$

where $b^2\in\mathrm{Ten}^2(\mathbb{R}^d)$, $a^2\in\mathrm{Ten}^4(\mathbb{R}^d)$ are tensors to determine and $a^0\in\mathrm{Sym}^2(\mathbb{R}^d)$ is the homogenized tensor defined by (see Lemma 3.3.1)

$$a^0_{ij}=\big\langle e_i^Ta(\nabla\chi_j+e_j)\big\rangle_Y,\tag{4.41}$$

where χ_j belongs to the class of solutions of (4.45). Notice that in (4.40) the tensors a^0,b^2 and a^2 are constant but $\tilde{u}$ depends on ε.

The form of equation (4.40), and more particularly the form of the $\mathcal{O}(\varepsilon^2)$ order operator, is an important ansatz. In fact, performing the asymptotic expansion with an equation of the form $\partial_t^2\tilde{u}-a^0_{ij}\partial^2_{ij}\tilde{u}+\varepsilon^2c_{ijkl}\partial^4_{ijkl}\tilde{u}=f$, leads to a dead end as the tensor c that we obtain is negative. Indeed, we obtain the same ill-posed equation as the one obtained via Bloch wave techniques in

[85] (see Section 4.2.7). We will see that using equation (4.40) brings sufficient freedom in the definition of b^2, a^2 to obtain well-posed effective equations.

The second ansatz that we make is that the adaptation of $\tilde{u}$ is of the form

$$\mathcal{B}^\varepsilon \tilde{u}(t,x) = \tilde{u}\big(t,x\big) + \varepsilon u^1\big(t,x,\tfrac{x}{\varepsilon}\big) + \varepsilon^2 u^2\big(t,x,\tfrac{x}{\varepsilon}\big) + \varepsilon^3 u^3\big(t,x,\tfrac{x}{\varepsilon}\big) + \varepsilon^4 u^4\big(t,x,\tfrac{x}{\varepsilon}\big), \tag{4.42}$$

where the $u^i(t,x,y)$ are Ω-periodic in x and Y-periodic in y. We introduce the differential operators

$$\mathcal{A}_{yy} = -\nabla_y \cdot \big(a(y)\nabla_y \cdot\big), \quad \mathcal{A}_{xy} = -\nabla_y \cdot \big(a(y)\nabla_x \cdot\big) - \nabla_x \cdot \big(a(y)\nabla_y \cdot\big),$$
$$\mathcal{A}_{xx} = -\nabla_x \cdot \big(a(y)\nabla_x \cdot\big),$$

so that for $\psi(x,y)$ smooth enough, using the chain rule, we have $\mathcal{A}^\varepsilon \psi\big(x,\frac{x}{\varepsilon}\big) = \big(\varepsilon^{-2}\mathcal{A}_{yy} + \varepsilon^{-1}\mathcal{A}_{xy} + \mathcal{A}_{xx}\big)\psi\big(x,\frac{x}{\varepsilon}\big)$. We fix a $t \in [0, T^\varepsilon]$ and using equations (4.26), (4.40) and ansatz (4.42), we compute

$$\begin{aligned}
(\partial_t^2 + \mathcal{A}^\varepsilon)(\mathcal{B}^\varepsilon \tilde{u} - u^\varepsilon)(t,x) = {} & \partial_t^2 \mathcal{B}^\varepsilon \tilde{u}(t,x) + \mathcal{A}^\varepsilon \mathcal{B}^\varepsilon \tilde{u}(t,x) - f(t,x) \\
= {} & \varepsilon^{-1}\big(\quad \mathcal{A}_{yy}u^1 + \mathcal{A}_{xy}\tilde{u} \qquad\qquad\qquad\qquad\qquad\qquad \big) \\
& + \varepsilon^0 \ \big(\quad \mathcal{A}_{yy}u^2 + \mathcal{A}_{xy}u^1 + \mathcal{A}_{xx}\tilde{u} \ + a^0_{ij}\partial^2_{ij}\tilde{u} \qquad\qquad \big) \\
& + \varepsilon^1 \ \big(\partial_t^2 u^1 + \mathcal{A}_{yy}u^3 + \mathcal{A}_{xy}u^2 + \mathcal{A}_{xx}u^1 \qquad\qquad\qquad \big) \\
& + \varepsilon^2 \ \big(\partial_t^2 u^2 + \mathcal{A}_{yy}u^4 + \mathcal{A}_{xy}u^3 + \mathcal{A}_{xx}u^2 - a^2_{ijkl}\partial^4_{ijkl}\tilde{u} + b^2_{ij}\partial^2_{ij}\partial_t^2\tilde{u} \big) \\
& + \mathcal{O}(\varepsilon^3),
\end{aligned} \tag{4.43}$$

where the u^i are evaluated at $(t,x,y=\frac{x}{\varepsilon})$. We now define successively u^1 to u^4 so that the terms of order $\mathcal{O}(\varepsilon^{-1})$ to $\mathcal{O}(\varepsilon^2)$ in (4.43) cancel. At order $\mathcal{O}(\varepsilon^{-1})$, we obtain the equation $\mathcal{A}_{yy}u^1 + \mathcal{A}_{xy}\tilde{u} = 0$, which reads

$$-\nabla_y \cdot \big(a(y)(\nabla_y u^1(t,x,y) + \nabla_x \tilde{u}(t,x))\big) = 0.$$

We can show that any solution of this elliptic equation is of the form $\chi_i(y)\partial_i\tilde{u}(t,x) + \tilde{u}_1(t,x)$ (see Section 3.3.1), where $\tilde{u}_1$ is a function that is independent of y and for all $1 \le i \le d$, χ_i is Y-periodic and solves the cell problem

$$-\nabla_y \cdot \big(a(\nabla_y \chi_i + e_i)\big) = 0 \quad \text{in } Y.$$

For simplicity, we choose $\tilde{u}_1 = 0$, i.e., $u_1(t,x,y) = \chi_i(y)\partial_i\tilde{u}(t,x)$. Consider now the $\mathcal{O}(1)$ order term in (4.43), which now reads

$$-\nabla_y \cdot \big(a(y)\nabla_y u^2(t,x,y)\big) = \big(\nabla_y \cdot (a(y)e_i\chi_j(y)) + e_i^T a(y)(\nabla_y\chi_j(y) + e_j) - a^0_{ij}\big)\partial^2_{ij}\tilde{u}(t,x).$$

The solution is given by $u^2(t,x,y) = \tilde{\theta}_{ij}(y)\partial^2_{ij}\tilde{u}(t,x) + \tilde{u}_2(t,x)$, where for $1 \le i,j \le d$ $\tilde{\theta}_{ij}$ is Y-periodic and solves the cell problem

$$-\nabla_y \cdot \big(a\nabla_y\tilde{\theta}_{ij}\big) = \nabla_y \cdot \big(ae_i\chi_j\big) + e_i^T a\nabla_y\chi_j + a_{ij} - a^0_{ij} \quad \text{in } Y.$$

Once again, we let $\tilde{u}_2 = 0$ for simplicity. We note here that for sufficiently smooth $\tilde{u}$, u^2 can also be written as $\theta_{ij}(y)\partial^2_{ij}\tilde{u}(t,x)$, where $\theta_{ij} = \frac{1}{2}(\tilde{\theta}_{ij} + \tilde{\theta}_{ji}) = S^2_{ij}\{\tilde{\theta}_{ij}\}$ is the symmetrization of $\tilde{\theta}_{ij}$ and solves the cell problem

$$-\nabla_y \cdot \big(a\nabla_y\theta_{ij}\big) = S^2_{ij}\big\{\nabla_y \cdot \big(ae_i\chi_j\big) + e_i^T a\nabla_y\chi_j + a_{ij} - a^0_{ij}\big\} \quad \text{in } Y.$$

The advantage of the second form of u^2 is that there are only $\binom{d+1}{2}$ cell problems describing $\{\theta_{ij}\}$ compared to the d^2 for $\{\tilde{\theta}_{ij}\}$. Before canceling the $\mathcal{O}(\varepsilon^1)$ and $\mathcal{O}(\varepsilon^2)$ order terms, we take into

account the definition of u^1 and u^2 to rewrite (4.43). Using (4.40), we have

$$\begin{aligned}
\partial_t^2 u^1 &= \chi_i \partial_i \partial_t^2 \tilde{u} = \chi_i \partial_i f + a^0_{ij} \chi_k \partial^3_{ijk} \tilde{u} + \mathcal{O}(\varepsilon^2),\\
\partial_t^2 u^2 &= \theta_{ij} \partial^2_{ij} \partial_t^2 \tilde{u} = \theta_{ij} \partial^2_{ij} f + a^0_{ij} \theta_{kl} \partial^4_{ijkl} \tilde{u} + \mathcal{O}(\varepsilon^2),\\
b^2_{ij} \partial^2_{ij} \partial_t^2 \tilde{u} &= b^2_{ij} \partial^2_{ij} f + a^0_{ij} b^2_{kl} \partial^4_{ijkl} \tilde{u} + \mathcal{O}(\varepsilon^2),
\end{aligned}$$

hence (4.43) reads

$$\begin{aligned}
(\partial_t^2 + \mathcal{A}^\varepsilon)(\mathcal{B}^\varepsilon \tilde{u} - u^\varepsilon) = \;& \varepsilon^1 \big(\mathcal{A}_{yy} u^3 + \mathcal{A}_{xy} u^2 + \mathcal{A}_{xx} u^1 + a^0_{ij} \chi_k \partial^3_{ijk} \tilde{u} \big)\\
&+ \varepsilon^2 \big(\mathcal{A}_{yy} u^4 + \mathcal{A}_{xy} u^3 + \mathcal{A}_{xx} u^2 + \big(a^0_{ij} (b^2_{kl} + \theta_{kl}) - a^2_{ijkl} \big) \partial^4_{ijkl} \tilde{u} \big)\\
&+ \varepsilon^1 \chi_i \partial_i f + \varepsilon^2 (b^2_{ij} + \theta_{ij}) \partial^2_{ij} f + \mathcal{O}(\varepsilon^3).
\end{aligned} \tag{4.44}$$

Let us first assume that $f = 0$. To cancel the $\mathcal{O}(\varepsilon^1)$ and $\mathcal{O}(\varepsilon^2)$ order terms in (4.44), we can set $u^3(t,x,y) = \kappa_{ijk}(y) \partial^3_{ijk} \tilde{u}(t,x)$, and $u^4(t,x,y) = \rho_{ijkl}(y) \partial^4_{ijkl} \tilde{u}(t,x)$, where κ_{ijk} and ρ_{ijkl} are the solutions of cell problems obtained in a similar manner as for χ_i and θ_{ij}. As previously, in order to minimize the number of cell problems, we use the symmetrization operators S^3 and S^4. In summary, we obtain the following cell problems: for $1 \le i,j,k,l \le d$, find Y-periodic functions $\chi_i, \theta_{ij}, \kappa_{ijk}, \rho_{ijkl}$ such that

$$\varepsilon^{-1} : \big(a \nabla_y \chi_i, \nabla_y w \big)_Y = - \big(a e_i, \nabla_y w \big)_Y, \tag{4.45a}$$

$$\varepsilon^{0} : \big(a \nabla_y \theta_{ij}, \nabla_y w \big)_Y = S^2_{ij} \big\{ - \big(a e_i \chi_j, \nabla_y w \big)_Y + \big(a(\nabla_y \chi_j + e_j) - a^0 e_j, e_i w \big)_Y \big\}, \tag{4.45b}$$

$$\begin{aligned}
\varepsilon^{1} : \big(a \nabla_y \kappa_{ijk}, \nabla_y w \big)_Y = S^3_{ijk} \big\{ & - \big(a e_i \theta_{jk}, \nabla_y w \big)_Y\\
& + \big(a(\nabla_y \theta_{jk} + e_j \chi_k) - a^0 e_j \chi_k, e_i w \big)_Y \big\},
\end{aligned} \tag{4.45c}$$

$$\begin{aligned}
\varepsilon^{2} : \big(a \nabla_y \rho_{ijkl}, \nabla_y w \big)_Y = S^4_{ijkl} \big\{ & - \big(a e_i \kappa_{jkl}, \nabla_y w \big)_Y + \big(a(\nabla_y \kappa_{jkl} + e_j \theta_{kl}), e_i w \big)_Y\\
& + \big(a^2_{ijkl} - a^0_{ij} \theta_{kl} - a^0_{ij} b^2_{kl}, w \big)_Y \big\},
\end{aligned} \tag{4.45d}$$

for Y-periodic test functions $w \in \mathrm{H}^1_{\mathrm{per}}(Y)$.

Let us now explain how the well-posedness of these cell problems constrains the definition of the effective tensors a^2 and b^2. To show that (4.45a) to (4.45d) are well-posed in the quotient space $\mathcal{W}_{\mathrm{per}}(Y)$, we apply Lax–Milgram theorem (we thus obtain a solution unique up to a constant). As the bilinear form $(v,w) \mapsto (a \nabla_y v, \nabla_y w)_Y$ is elliptic and bounded, we have to verify that the right hand sides belong to $\mathcal{W}^*_{\mathrm{per}}(Y)$. We refer to Appendix A.2 for a characterization of $\mathcal{W}^*_{\mathrm{per}}(Y)$. In particular, $F \in [\mathrm{H}^1_{\mathrm{per}}(Y)]^*$ given by

$$\big\langle F, w \big\rangle = \big(f^0, w \big)_{\mathrm{L}^2(Y)} + \big(f^1_k, \partial_k w \big)_{\mathrm{L}^2(Y)},$$

for some $f^0, f^1_1, \dots, f^1_d \in \mathrm{L}^2(Y)$ belongs to $\mathcal{W}^*_{\mathrm{per}}(Y)$ if and only if

$$\big(f^0, 1 \big)_{\mathrm{L}^2(Y)} = 0. \tag{4.46}$$

Consequently, the right hand sides of the cell problems (4.45) have to satisfy the solvability condition (4.46). In particular, imposing the well-posedness of (4.45d) provides a constraint on the effective tensors a^2, b^2. Let us prove the well-posedness of (4.45a) to (4.45c) and derive explicitly the constraint on a^2, b^2 dictated by the well-posedness of (4.45d).

First, note that the right hand side of (4.45a) trivially satisfies (4.46). Next, if we let $w = 1$ in the right hand side of (4.45b), we obtain

$$S^2_{ij} \big\{ \big(a(\nabla_y \chi_j + e_j) - a^0 e_j, e_i \big)_Y \big\} = |Y| S^2_{ij} \big\{ \big\langle e_i^T a (\nabla_y \chi_j + e_j) \big\rangle_Y \big\} - |Y| S^2_{ij} \big\{ a^0_{ij} \big\} = 0, \tag{4.47}$$

where we used the definition of the homogenized tensor (4.41). Hence, the cell problem (4.45b) is well-posed. Next, letting $w = 1$ in the right hand side of (4.45c) we obtain

$$S^3_{ijk}\big\{ - \big(a\nabla_y\theta_{jk}, e_i\big)_Y - \big(e_j\chi_k, e_i\big)_Y + a^0_{ij}\big(\chi_k, 1\big)_Y\big\}, \tag{4.48}$$

and we need this quantity to vanish for any $1 \le i, j, k \le d$. Using the symmetry of a, equations (4.45a) with the test function $w = \theta_{jk}$, and (4.45b) with $w = \chi_i$, we have

$$\begin{aligned} -\big(a\nabla_y\theta_{jk}, e_i\big)_Y &= \big(a\nabla_y\theta_{jk}, \nabla_y\chi_i\big)_Y \\ &= S^2_{jk}\big\{ - \big(ae_j\chi_k, \nabla_y\chi_i\big)_Y + \big(a(\nabla_y\chi_k + e_k), e_j\chi_i\big)_Y - \big(a^0_{jk}, \chi_i\big)_Y\big\}, \end{aligned}$$

and (4.48) can thus be rewritten as

$$S^3_{ijk}\big\{ - \big(ae_j\chi_k, \nabla_y\chi_i + e_i\big)_Y + \big(a(\nabla_y\chi_k + e_k), e_j\chi_i\big)_Y - a^0_{jk}\big(1, \chi_i\big)_Y + a^0_{ij}\big(\chi_k, 1\big)_Y\big\} = 0.$$

This equality proves that the cell problem (4.45c) is well-posed. Finally, we apply the solvability condition (4.46) to the right hand side of equation (4.45d) in order to obtain a constraint on a^2 and b^2. Letting $w = 1$, we have

$$|Y|S^4_{ijkl}\big\{a^2_{ijkl} - a^0_{ij}b^2_{kl}\big\} = S^4_{ijkl}\big\{ - \big(a\nabla_y\kappa_{jkl}, e_i\big)_Y - \big(ae_j\theta_{kl}, e_i\big)_Y + \big(a^0_{ij}, \theta_{kl}\big)_Y\big\}. \tag{4.49}$$

We use the symmetry of a, equation (4.45a) with test function $w = \kappa_{jkl}$, and equation (4.45c) with $w = \chi_i$, to get

$$\begin{aligned} -\big(a\nabla_y\kappa_{jkl}, e_i\big)_Y &= \big(a\nabla_y\kappa_{jkl}, \nabla_y\chi_i\big)_Y \\ &= S^3_{jkl}\big\{ - \big(ae_j\theta_{kl}, \nabla_y\chi_i\big)_Y + \big(a(\nabla_y\theta_{kl} + e_k\chi_l), e_j\chi_i\big)_Y - a^0_{jk}\big(\chi_l, \chi_i\big)\big\}. \end{aligned}$$

Used in (4.49), this equality gives (using the symmetry of a)

$$\begin{aligned} |Y|S^4_{ijkl}\big\{a^2_{ijkl} - a^0_{ij}b^2_{kl}\big\} = S^4_{ijkl}\big\{&\big(ae_j\chi_i, \nabla_y\theta_{kl}\big)_Y - \big(a(\nabla_y\chi_i + e_i), e_j\theta_{kl}\big)_Y + \big(a^0_{ij}, \theta_{kl}\big)_Y \\ &- a^0_{jk}\big(\chi_l, \chi_i\big)_Y + \big(ae_k\chi_l, e_j\chi_i\big)_Y\big\}. \end{aligned}$$

Using equation (4.45b) with the test function $w = \theta_{kl}$, we obtain then the following constraint on a^2 and b^2

$$|Y|S^4_{ijkl}\big\{a^2_{ijkl} - a^0_{ij}b^2_{kl}\big\} = S^4_{ijkl}\big\{\big(a_{jk}\chi_l, \chi_i\big)_Y - \big(a\nabla_y\theta_{ji}, \nabla_y\theta_{kl}\big)_Y - a^0_{jk}\big(\chi_l, \chi_i\big)_Y\big\}. \tag{4.50}$$

We conclude that the cell problem (4.45d) is well-posed in $\mathcal{W}_{\rm per}(Y)$ if and only if the tensors a^2, b^2 satisfy (4.50). In particular, if this constraint is satisfied, we can define the adaptation $\mathcal{B}^\varepsilon\tilde{u}$ as in (4.42) and show that $(\partial_t^2 + \mathcal{A}^\varepsilon)(\mathcal{B}^\varepsilon\tilde{u} - u^\varepsilon) = \mathcal{O}(\varepsilon^3)$ (under sufficient regularity of $\tilde{u}$ and the correctors). Hence, following the plan described in Section 4.2.2 with $\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u} = \big[\mathcal{B}^\varepsilon\tilde{u}\big]$, we can prove the estimate $\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} \le C\varepsilon$. This result is rigorously proved in the next section.

It is interesting to note that the tensor in the right hand side of (4.50) is independent of the choice of $\chi_i \in \boldsymbol{\chi}_i$. Indeed, as this tensor must characterizes the long time effects its definition must be independent of any choice. This is proved in Section 4.3.5.

Recall that we assumed $f = 0$. It is in fact not necessary if we proceed as follows. Indeed, in order to cancel the non-vanishing terms $\varepsilon\chi_i\partial_i f + \varepsilon^2(b^2_{ij} + \theta_{ij})\partial^2_{ij}f$ in (4.44), we add a term in the adaptation (4.42). Namely, we replace (4.42) by

$$\begin{aligned} \mathcal{B}^\varepsilon\tilde{u}(t,x) = \tilde{u}(t,x) &+ \varepsilon\chi_i\big(\tfrac{x}{\varepsilon}\big)\partial_i\tilde{u}(t,x) + \varepsilon^2\theta_{ij}\big(\tfrac{x}{\varepsilon}\big)\partial^2_{ij}\tilde{u}(t,x) \\ &+ \varepsilon^3\kappa_{ijk}\big(\tfrac{x}{\varepsilon}\big)\partial^3_{ijk}\tilde{u}(t,x) + \varepsilon^4\rho_{ijkl}\big(\tfrac{x}{\varepsilon}\big)\partial^4_{ijkl}\tilde{u}(t,x) + \varphi(t,x), \end{aligned} \tag{4.51}$$

where $\varphi(t,\cdot)$ belongs to the class $\boldsymbol{\varphi}(t) \in \mathcal{W}_{\rm per}(\Omega)$ that solves

$$\begin{aligned}&(\partial_t^2 + \mathcal{A}^\varepsilon)\boldsymbol{\varphi}(t) = -\big[\varepsilon\chi_i\big(\tfrac{\cdot}{\varepsilon}\big)\partial_i f(t) + \varepsilon^2\big(b_{ij}^2 + \theta_{ij}\big(\tfrac{\cdot}{\varepsilon}\big)\big)\partial_{ij}^2 f(t)\big] \quad \text{in } \mathcal{W}_{\rm per}^*(\Omega) \ \text{ a.e. } t\in[0,T^\varepsilon],\\ &\boldsymbol{\varphi}(0) = \partial_t\boldsymbol{\varphi}(0) = \big[0\big].\end{aligned} \tag{4.52}$$

The standard well-posedness of the wave equation (Theorem 2.1.1) ensures that if $f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{H}^2(\Omega))$ and $\chi_i, \theta_{ij} \in \mathrm{L}^\infty(Y)$, there exists a unique solution $\boldsymbol{\varphi}$ of (4.52), satisfying

$$\boldsymbol{\varphi} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{W}_{\rm per}(\Omega)), \quad \partial_t\boldsymbol{\varphi} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{L}^2(\Omega)), \quad \partial_t^2\boldsymbol{\varphi} \in \mathrm{L}^2(0,T^\varepsilon;\mathcal{W}_{\rm per}^*(\Omega)). \tag{4.53}$$

Observe then that $\mathcal{B}^\varepsilon\tilde{u}$, defined in (4.51), satisfies

$$\big(\partial_t^2 + \mathcal{A}^\varepsilon\big)\big(\big[\mathcal{B}^\varepsilon\tilde{u}\big] - \big[u^\varepsilon\big]\big) = \big[r^\varepsilon(t) - \varepsilon\chi_i\big(\tfrac{x}{\varepsilon}\big)\partial_i f(t,x) - \varepsilon^2\big(b_{ij}^2 + \theta_{ij}\big(\tfrac{x}{\varepsilon}\big)\big)\partial_{ij}^2 f(t,x)\big],$$

where r^ε is the right hand side of (4.44), so that $\big(\partial_t^2 + \mathcal{A}^\varepsilon\big)\big(\big[\mathcal{B}^\varepsilon\tilde{u}\big] - \big[u^\varepsilon\big]\big) = \mathcal{O}(\varepsilon^3)$. Furthermore, we verify that under sufficient regularity of the data, $\big[\mathcal{B}^\varepsilon\tilde{u}\big](t)$ belongs to $\mathcal{W}_{\rm per}(\Omega)$ and $\big[\mathcal{B}^\varepsilon\tilde{u}\big]$ satisfies the conditions in (4.37). First, condition (4.37c) follows directly from (4.51) and the initial conditions in (4.52). Next, let us verify that (4.37b) holds. Applying the estimate from Lemma 4.2.1 to $\boldsymbol{\varphi}$, we obtain, provided $\chi_k \in \mathcal{C}^0(\bar{Y}), \theta_{ij} \in \mathcal{C}^0(\bar{Y}), f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{H}^2(\Omega))$,

$$\|\boldsymbol{\varphi}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{W})} \leq C\varepsilon\|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^2(\Omega))}, \tag{4.54}$$

where C only depends on $\lambda, \max_k \|\chi_k\|_{\mathrm{L}^\infty(Y)}$ and $\max_{ij}\|\theta_{ij}\|_{\mathrm{L}^\infty(Y)}$. Hence, provided sufficient regularity on the correctors, and if f satisfies $\|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^2(\Omega))} = \mathcal{O}(1)$, (4.54) ensures (4.37b) (more details on the requirement $\|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^2(\Omega))} = \mathcal{O}(1)$ are given in the next section).

Remark 4.2.3. The effective equation obtained in [18] is of the same form as (4.40), with the additional term $-\varepsilon^2 b_{ij}^2\partial_{ij}^2 f$ in the right hand side (see (4.10)). We can verify that this modification indeed cancels a part of the second term in (4.44) and thus leads to a slightly better equation (better in the sense that the constant in the error estimate of Theorem 4.2.4 below is smaller). Nevertheless, in the regime of our result, this correction is negligible. Indeed, denoting $\tilde{u}_1, \tilde{u}_2$ the solutions of the equations with and without the additional term, respectively , we verify that $\|\tilde{u}_1 - \tilde{u}_2\|_{\mathrm{L}^\infty(W)} \leq \varepsilon^2 2/\lambda^{1/2}|b^2|_\infty\|f\|_{\mathrm{L}^1(\mathrm{H}^2)}$. As we require f to satisfy $\|f\|_{\mathrm{L}^1(\mathrm{H}^4)} = \mathcal{O}(1)$ (see Corollary 4.2.5), the benefit of the correction of the right hand side is not significant. Note that for more general source term, this correction might be worth.

To conclude this section, let us discuss the correctors and their dependence. First, as (4.45a-4.45d) are well-posed in $\mathcal{W}_{\rm per}(Y)$, we obtain the unique (class of) solutions $\boldsymbol{\chi}_k, \boldsymbol{\theta}_{ij}, \boldsymbol{\kappa}_{ijk}, \boldsymbol{\rho}_{ijkl} \in \mathcal{W}_{\rm per}(Y)$ for $1 \leq i,j,k,l \leq d$. Note that $\boldsymbol{\theta}_{ij}$ depends on the choice $\chi_k \in \boldsymbol{\chi}_k$, κ_{ijk} depends on the choices $\chi_k \in \boldsymbol{\chi}_k$, $\theta_{ij} \in \boldsymbol{\theta}_{ij}$, etc. A natural choice for the normalization of the correctors is the zero-mean function. However, observe that the constraint (4.50) has been derived independently of the choice of normalization. Hence, any normalization can be used.

4.2.4 A priori error estimate and definition of the family of effective equations

We present here the main result of this chapter and the first contribution of the thesis. In particular, we define a family of effective equation and prove that its elements are ε-close to the oscillatory solution u^ε in the $\mathrm{L}^\infty(0,T^\varepsilon;W)$ norm.

Let $a^0 \in \mathrm{Sym}^2(\mathbb{R}^d)$ be the homogeneous tensor defined as (4.41) and let $b^2 \in \mathrm{Ten}^2(\mathbb{R}^d)$ and $a^2 \in \mathrm{Ten}^4(\mathbb{R}^d)$ be constant tensors such that

$$\begin{aligned}&i) \qquad b_{ij}^2 = b_{ji}^2, \qquad && b^2\eta\cdot\eta \geq 0 \ \ \forall\eta\in\mathbb{R}^d,\\ &ii) \qquad a_{ijkl}^2 = a_{klij}^2, \qquad && a^2\xi:\xi \geq 0 \ \ \forall\xi\in\mathrm{Sym}^2(\mathbb{R}^d).\end{aligned} \tag{4.55}$$

Consider the following linear Boussinesq equation: $\tilde{u} : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ such that

$$\begin{array}{ll} \partial_t^2 \tilde{u} - a^0_{ij}\partial^2_{ij}\tilde{u} + \varepsilon^2\big(a^2_{ijkl}\partial^4_{ijkl}\tilde{u} - b^2_{ij}\partial^2_{ij}\partial^2_t\tilde{u}\big) = f & \text{in } (0,T^\varepsilon]\times\Omega \\ x \mapsto \tilde{u}(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,T^\varepsilon], \\ \tilde{u}(0,x) = g^0(x), \quad \partial_t\tilde{u}(0,x) = g^1(x) & \text{in } \Omega, \end{array} \tag{4.56}$$

where the initial conditions g^0, g^1 and the source term f are the same as in the equation for u^ε (4.26). As proved in Section 2.1.2, if the data satisfy the regularity $g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^2(\Omega)$, $g^1 \in \mathrm{L}^2_0 \cap \mathrm{H}^1(\Omega)$, and $f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{L}^2_0(\Omega))$, (4.56) has a unique weak solution $\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega))$ with $\partial_t\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2_0(\Omega))$.

The following theorem provides a sufficient condition on the tensors a^2, b^2 such that (4.56) is an effective equation up to timescales $\mathcal{O}(\varepsilon^{-2})$.

Theorem 4.2.4. *Assume that the Y-periodic tensor satisfies $a(y) \in \mathrm{W}^{2,\infty}(Y)$. Furthermore, assume that the solution $\tilde{u}$ of (4.56), the initial conditions, and the right hand side satisfy the regularity*

$$\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^5(\Omega)), \quad \partial_t\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^4(\Omega)), \quad \partial_t^2\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^3(\Omega)),$$
$$g^0 \in \mathrm{H}^4(\Omega), \quad g^1 \in \mathrm{H}^4(\Omega), \quad f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{H}^2(\Omega)).$$

Let $\boldsymbol{\chi}_k$ be the (class of) solution of (4.45a), fix any $\chi_k \in \boldsymbol{\chi}_k$, let $\boldsymbol{\theta}_{ij}$ be the corresponding (class of) solution of (4.45b) and fix $\theta_{ij} \in \boldsymbol{\theta}_{ij}$. Assume then that b^2 and a^2 satisfy the relation

$$S^4_{ijkl}\Big\{a^2_{ijkl} - a^0_{ij}b^2_{kl}\Big\} = S^4_{ijkl}\Big\{\big\langle a_{jk}\chi_l\chi_i\big\rangle_Y - \big\langle a\nabla\theta_{ji}\cdot\nabla\theta_{kl}\big\rangle_Y - a^0_{jk}\big\langle\chi_l\chi_i\big\rangle_Y\Big\}. \tag{4.57}$$

Then the following error estimate holds

$$\begin{aligned} \|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} \le C\varepsilon\Big(&\|g^1\|_{\mathrm{H}^4(\Omega)} + \|g^0\|_{\mathrm{H}^4(\Omega)} + \|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^2(\Omega))} \\ &+ \textstyle\sum_{k=1}^5 |\tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^k(\Omega))} + |\partial_t^2\tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^3(\Omega))}\Big), \end{aligned} \tag{4.58}$$

where C depends only on $T, \lambda, \Lambda, |b^2|_\infty, |a^2|_\infty, \|a\|_{\mathrm{W}^{2,\infty}(Y)}$, and Y, and we recall the definition of the norm (see (4.23))

$$\|w\|_W = \inf_{\substack{w=w_1+w_2 \\ w_1,w_2\in\mathrm{W}_{\mathrm{per}}(\Omega)}} \Big\{\|w_1\|_{\mathrm{L}^2(\Omega)} + \|\nabla w_2\|_{\mathrm{L}^2(\Omega)}\Big\} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(\Omega).$$

Using the estimates for the higher regularity of $\tilde{u}$, we can prove an error estimate that depends only on the data of the problem.

Corollary 4.2.5. *Assume that the assumptions of Theorem 4.2.4 hold. If in addition the data satisfy the regularity*

$$g^0 \in \mathrm{H}^7(\Omega), \quad g^1 \in \mathrm{H}^5(\Omega), \quad f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega)\cap\mathrm{H}^4(\Omega)), \quad \partial_t f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{H}^3(\Omega)),$$

then the following error estimates holds

$$\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} \le C\varepsilon\Big(\|g^1\|_{\mathrm{H}^5(\Omega)} + \|g^0\|_{\mathrm{H}^7(\Omega)} + \|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^4(\Omega))} + \|\partial_t f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^3(\Omega))}\Big), \tag{4.59}$$

where C depends only on $T, \lambda, \Lambda, |b^2|_\infty, |a^2|_\infty, \|a\|_{\mathrm{W}^{2,\infty}(Y)}$ and Y.

Proof. Under the assumptions, we can show that the weak solution $\tilde{u}$ satisfies for $1 \leq k \leq 5$ the energy estimate (see Theorem 2.1.9 *i*))

$$|\tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^k(\Omega))} \leq C\big(\|g^1\|_{\mathrm{H}^k(\Omega)} + \|g^0\|_{\mathrm{H}^{k+1}(\Omega)} + \|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^{k-1}(\Omega))}\big),$$

where the constant depends only on $\lambda, \Lambda, |b^2|_\infty, |a^2|_\infty$. Similarly, we have for $0 \leq k \leq 3$, (see Theorem 2.1.9 *ii*))

$$|\partial_t^2 \tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^k(\Omega))} \leq C\big(\|g^1\|_{\mathrm{H}^{k+2}(\Omega)} + \|g^0\|_{\mathrm{H}^{k+4}(\Omega)} + \|f\|_{\mathrm{W}^{1,1}(0,T^\varepsilon;\mathrm{H}^k(\Omega))}\big).$$

Combining these energy estimates with (4.58), we obtain (4.59) and the proof is complete. □

Let us emphasize that the constant C in estimate (4.59) does not depend on Ω. Hence, for an arbitrarily large domain Ω, if the quantities

$$\|g^1\|_{\mathrm{H}^5(\Omega)}, \quad \|g^0\|_{\mathrm{H}^7(\Omega)}, \quad \|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^4(\Omega))}, \quad \|\partial_t f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^3(\Omega))}, \tag{4.60}$$

are bounded independently of ε, estimate (4.58) reads $\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} = \mathcal{O}(\varepsilon)$. In particular, a source term f with a sufficiently small support in space and decaying sufficiently fast to zero in time satisfies (4.60).

Let us discuss the case when Ω is a small periodic domain, i.e., $\mathrm{diam}(\Omega) = \mathcal{O}(1)$. In this case, $\tilde{u}$ is the superposition of reflexions of the wave and long time dispersive effects are still observed (see e.g. [13]). Combining (4.59) with (4.24) and using that the Poincaré constant C_Ω is bounded by $\mathrm{diam}(\Omega)$, we obtain an error estimate proving that $\tilde{u}$ is ε-close to u^ε in the $\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))$ norm.

Thanks to Theorem 4.2.4, we define the family of effective equations as follows.

Definition 4.2.6. The family $\mathcal{E}$ of effective equations is the set of equations (4.56) where b^2, a^2 satisfy both (4.55) and (4.57). Note that $\mathcal{E}$ is used to denote both the family of effective equations and the corresponding solutions.

4.2.5 Proof of the error estimate (Theorem 4.2.4)

We prove here Theorem 4.2.4. The proof follows two steps. First, we define the adaptation operator $\boldsymbol{\mathcal{B}}^\varepsilon$ using the correctors defined in Section 4.2.3. In particular, the existence and uniqueness of the correctors is ensured by the assumption (4.57) on the tensors a^2, b^2. We then split the error as

$$\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(W)} = \|\left[u^\varepsilon - \tilde{u}\right]\|_{\mathrm{L}^\infty(\mathcal{W})} \leq \|\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u} - \left[u^\varepsilon\right]\|_{\mathrm{L}^\infty(\mathcal{W})} + \|\left[\tilde{u}\right] - \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}\|_{\mathrm{L}^\infty(\mathcal{W})},$$

and estimate both terms separately. In particular, we prove that $\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}$ satisfies the same wave equation as u^ε up to a remainder of order $\mathcal{O}(\varepsilon^3)$ (Lemma 4.2.8).

First, note that the cell problems (4.45a), (4.45b) and (4.45c) are well-posed (a^0 is defined as (4.41)). Then, as we assume that (4.57) holds, the cell problem (4.45d) is well-posed. Let χ_i and θ_{ij} be as in Theorem 4.2.4, let $\boldsymbol{\kappa}_{ijk}$ be the corresponding solution of (4.45c), fix $\kappa_{ijk} \in \boldsymbol{\kappa}_{ijk}$, and similarly fix ρ_{ijkl} in the corresponding class $\boldsymbol{\rho}_{ijkl}$ of solution of (4.45d). As we assume $a \in \mathrm{W}^{2,\infty}(Y)$, elliptic regularity result (Theorem A.2.2) and Sobolev embeddings (see Appendix A.2) ensure that $\chi_i, \theta_{ij}, \kappa_{ijk}, \rho_{ijkl} \in \mathcal{C}^1_{\mathrm{per}}(\bar{Y})$ and for any $1 \leq i,j,k,l \leq d$ it holds

$$\|\chi_i\|_{\mathcal{C}^1(\bar{Y})}, \|\theta_{ij}\|_{\mathcal{C}^1(\bar{Y})}, \|\kappa_{ijk}\|_{\mathcal{C}^1(\bar{Y})}, \|\rho_{ijkl}\|_{\mathcal{C}^1(\bar{Y})} \leq C \max_{ij} \|a_{ij}\|_{\mathrm{W}^{2,\infty}(Y)}, \tag{4.61}$$

where C depends only on λ, Λ in (4.27) and Y. Finally, let $\boldsymbol{\varphi} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{W}_{\mathrm{per}}(\Omega))$ be the unique (class of) solution of (4.52).

We now define the adaptation operator as

$$\boldsymbol{\mathcal{B}}^\varepsilon : \mathrm{L}^2(0,T^\varepsilon;\mathrm{H}^1_{\mathrm{per}}(\Omega)\cap \mathrm{H}^3(\Omega)) \to \mathrm{L}^2(0,T^\varepsilon;\mathcal{W}^*_{\mathrm{per}}(\Omega)), \quad v \mapsto \boldsymbol{\mathcal{B}}^\varepsilon v,$$

$$\begin{aligned}\big\langle \boldsymbol{\mathcal{B}}^\varepsilon v(t), \boldsymbol{w}\big\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}} &= \Big(\big[v(t) + \varepsilon(\chi_j - \partial_{y_m}\theta_{mj})\partial_j v(t) + \varepsilon^3(\kappa_{jkl} - \partial_{y_m}\rho_{mjkl})\partial^3_{jkl} v(t)\big], \boldsymbol{w}\Big)_{\mathcal{L}^2(\Omega)} \\ &\quad - \Big(\varepsilon^2\theta_{mj}\partial_j v(t) + \varepsilon^4 \rho_{mjkl}\partial^3_{jkl} v(t), \partial_m \boldsymbol{w}\Big)_{\mathrm{L}^2(\Omega)} + \big\langle \boldsymbol{\varphi}(t), \boldsymbol{w}\big\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}},\end{aligned} \tag{4.62}$$

for a.e. $t \in [0,T^\varepsilon]$, where the correctors $\chi_i, \theta_{ij}, \kappa_{ijk}$ and ρ_{ijkl} are evaluated at $y = \frac{x}{\varepsilon}$. Using the Green formula (as in Remark 4.2.7), we verify that for $v \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{H}^1_{\mathrm{per}}(\Omega)\cap\mathrm{H}^4(\Omega))$, we have $\big\langle \boldsymbol{\mathcal{B}}^\varepsilon v(t), \boldsymbol{w}\big\rangle = \big(\big[\mathcal{B}^\varepsilon v(t)\big], \boldsymbol{w}\big)_{\mathcal{L}^2}$ where $\mathcal{B}^\varepsilon$ is defined in (4.51). Moreover, if $v \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{H}^1_{\mathrm{per}}(\Omega)\cap \mathrm{H}^5(\Omega))$, then $\boldsymbol{\mathcal{B}}^\varepsilon v(t) \in \mathcal{W}_{\mathrm{per}}(\Omega)$ and we can define

$$\big\langle \mathcal{A}^\varepsilon \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}(t), \boldsymbol{w}\big\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}} = \big\langle \mathcal{A}^\varepsilon \big[(\mathcal{B}^\varepsilon v(t)\big], \boldsymbol{w}\big\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}}.$$

Finally, note that under the assumptions of Theorem 4.2.4, $\boldsymbol{\mathcal{B}}^\varepsilon$ verifies the requirements in (4.37).

Remark 4.2.7. The following formula (applications of the Green formula) is useful: for any $c \in [\mathrm{W}^{1,\infty}_{\mathrm{per}}(\Omega)]^d$, $v \in \mathrm{H}^1_{\mathrm{per}}(\Omega)$ and $\boldsymbol{w} = [w] \in \mathcal{W}_{\mathrm{per}}(\Omega)$,

$$\begin{aligned}\big(\big[\varepsilon c_m\big(\tfrac{\cdot}{\varepsilon}\big)\partial_m v\big], \boldsymbol{w}\big)_{\mathcal{L}^2(\Omega)} &= \big(\varepsilon c_m\big(\tfrac{\cdot}{\varepsilon}\big)\partial_m v, w\big)_{\mathrm{L}^2(\Omega)} - |\Omega|\big\langle \varepsilon c_m\big(\tfrac{\cdot}{\varepsilon}\big)\partial_m v\big\rangle_\Omega \big\langle w\big\rangle_\Omega \\ &= -\big(\partial_{y_m} c_m\big(\tfrac{\cdot}{\varepsilon}\big)v, w\big)_{\mathrm{L}^2(\Omega)} - \big(\varepsilon c_m\big(\tfrac{\cdot}{\varepsilon}\big)v, \partial_m w\big)_{\mathrm{L}^2(\Omega)} + |\Omega|\big\langle \partial_{y_m} c_m\big(\tfrac{\cdot}{\varepsilon}\big)v\big\rangle_\Omega \big\langle w\big\rangle_\Omega \\ &= -\big(\big[\partial_{y_m} c_m\big(\tfrac{\cdot}{\varepsilon}\big)v\big], \boldsymbol{w}\big)_{\mathcal{L}^2(\Omega)} - \big(\varepsilon c_m\big(\tfrac{\cdot}{\varepsilon}\big)v, \partial_m \boldsymbol{w}\big)_{\mathrm{L}^2(\Omega)},\end{aligned} \tag{4.63}$$

where we recall the notation $\partial_{y_m} c_m = \sum_{m=1}^d \partial_{y_m} c_m$.

Lemma 4.2.8. *Under the assumptions of Theorem 4.2.4, $\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}$ satisfies for a.e. $t \in [0,T^\varepsilon]$*

$$(\partial_t^2 + \mathcal{A}^\varepsilon)\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}(t) = \big[f(t)\big] + \boldsymbol{\mathcal{R}}^\varepsilon \tilde{u}(t) \quad \text{in } \mathcal{W}^*_{\mathrm{per}}(\Omega) \quad \text{for a.e. } t \in [0,T^\varepsilon],$$

where the remainder $\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{W}^*_{\mathrm{per}}(\Omega))$ satisfies

$$\big\langle \boldsymbol{\mathcal{R}}^\varepsilon \tilde{u}(t), \boldsymbol{w}\big\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}} = \big((\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u})_0(t), \boldsymbol{w}\big)_{\mathcal{L}^2} + \big((\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u})_1(t), \nabla \boldsymbol{w}\big)_{\mathrm{L}^2},$$

with the bound

$$\begin{aligned}&\|(\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u})_0\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{L}^2(\Omega))} + \|(\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u})_1\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} \\ &\qquad \le C\varepsilon^3\Big(|\tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^5(\Omega))} + |\partial_t^2 \tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^3(\Omega))}\Big),\end{aligned} \tag{4.64}$$

for a constant C that depends only on $\lambda, \Lambda, |b^2|_\infty, |a^2|_\infty, \|a\|_{\mathrm{W}^{2,\infty}(Y)}$ and Y.

Proof. To simplify the notation, $\langle\cdot,\cdot\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}}$ is denoted by $\langle\cdot,\cdot\rangle$. First, using equation (4.56) and the assumptions on the regularity of $\tilde{u}$, note that the following equalities hold in $\mathrm{L}^2(\Omega)$ for a.e. $t \in [0,T^\varepsilon]$ and for $1 \le p \le d$,

$$\partial_t^2 \tilde{u} = f + a^0_{ij}\partial^2_{ij}\tilde{u} - \varepsilon^2 a^2_{ijkl}\partial^4_{ijkl}\tilde{u} + \varepsilon^2 b^2_{ij}\partial^2_{ij}\partial^2_t \tilde{u}, \tag{4.65}$$

$$\partial_p\partial_t^2 \tilde{u} = \partial_p f + a^0_{ij}\partial^3_{pij}\tilde{u} - \varepsilon^2 a^2_{ijkl}\partial^5_{pijkl}\tilde{u} + \varepsilon^2 b^2_{ij}\partial^3_{pij}\partial^2_t \tilde{u}. \tag{4.66}$$

Then, we fix $t \in [0, T^\varepsilon]$ and develop the terms $\partial_t^2 \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}(t)$ and $\mathcal{A}^\varepsilon \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}(t)$ separately. Using (4.65) and formula (4.63), we have

$$\big(\big[\partial_t^2 \tilde{u}\big], \boldsymbol{w}\big)_{\mathcal{L}^2} = \big(\big[f + a_{ij}^0 \partial_{ij}^2 \tilde{u} - \varepsilon^2 a_{ijkl}^2 \partial_{ijkl}^4 \tilde{u}\big], \boldsymbol{w}\big)_{\mathcal{L}^2} - \big(\varepsilon^2 b_{mj}^2 \partial_j \partial_t^2 \tilde{u}, \partial_m \boldsymbol{w}\big)_{\mathrm{L}^2}. \tag{4.67}$$

Using (4.62) and (4.67), we compute

$$\begin{aligned}\big\langle \partial_t^2 \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}, \boldsymbol{w}\big\rangle = \Big(\big[& f + a_{ij}^0 \partial_{ij}^2 \tilde{u} + \varepsilon(\chi_j - \partial_{y_m} \theta_{mj}) \partial_j \partial_t^2 \tilde{u} - \varepsilon^2 a_{ijkl}^2 \partial_{ijkl}^4 \tilde{u} \\ & + \varepsilon^3 (\kappa_{jkl} - \partial_{y_m} \rho_{mjkl}) \partial_{jkl}^3 \partial_t^2 \tilde{u} \big], \boldsymbol{w} \Big)_{\mathcal{L}^2} \\ & - \Big(\varepsilon^2 (\theta_{mj} + b_{mj}^2) \partial_j \partial_t^2 \tilde{u} + \varepsilon^4 \rho_{mjkl} \partial_{jkl}^3 \partial_t^2 \tilde{u}, \partial_m \boldsymbol{w} \Big)_{\mathrm{L}^2} + \big\langle \partial_t^2 \boldsymbol{\varphi}, \boldsymbol{w} \big\rangle.\end{aligned}$$

Using now (4.66) to substitute $\partial_j \partial_t^2 \tilde{u}$, we obtain

$$\begin{aligned}\big\langle \partial_t^2 \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}, \boldsymbol{w}\big\rangle = & \Big(\big[f + a_{ij}^0 \partial_{ij}^2 \tilde{u} + \varepsilon a_{ij}^0 (\chi_k - \partial_{y_m} \theta_{mk}) \partial_{ijk}^3 \partial_t^2 \tilde{u} - \varepsilon^2 a_{ijkl}^2 \partial_{ijkl}^4 \tilde{u} \big], \boldsymbol{w} \Big)_{\mathcal{L}^2} \\ & - \Big(\varepsilon^2 a_{ij}^0 (\theta_{mk} + b_{mk}^2) \partial_{ijk}^3 \tilde{u}, \partial_m \boldsymbol{w} \Big)_{\mathrm{L}^2} \\ & + \Big(\big[\varepsilon(\chi_k - \partial_{y_m} \theta_{mk}) \partial_k f \big], \boldsymbol{w} \Big)_{\mathcal{L}^2} - \Big(\varepsilon^2 (\theta_{mj} + b_{mj}^2) \partial_j f, \partial_m \boldsymbol{w} \Big)_{\mathrm{L}^2} \\ & + \big\langle \partial_t^2 \boldsymbol{\varphi}, \boldsymbol{w} \big\rangle + \big\langle \boldsymbol{\mathcal{R}}_1^\varepsilon \tilde{u}, \boldsymbol{w} \big\rangle,\end{aligned}$$

where

$$\begin{aligned}\big\langle \boldsymbol{\mathcal{R}}_1^\varepsilon \tilde{u}, \boldsymbol{w}\big\rangle = \Big(\big[& \varepsilon^3 (\kappa_{jkl} + b_{jk}^2 \chi_l - \partial_{y_m} (\rho_{mjkl} + b_{jk}^2 \theta_{ml})) \partial_{jkl}^3 \partial_t^2 \tilde{u} \\ & - \varepsilon^3 a_{ijkl}^2 (\chi_p - \partial_{y_m} \theta_{mp}) \partial_{ijklp}^5 \tilde{u} \big], \boldsymbol{w} \Big)_{\mathcal{L}^2} \\ & + \Big(\varepsilon^4 (\rho_{mijk} - b_{ij}^2 \theta_{mk} + b_{ij}^2 b_{mk}^2) \partial_{ijk}^3 \partial_t^2 \tilde{u} + \varepsilon^4 a_{ijkl}^2 (\theta_{mp} + b_{mp}^2) \partial_{ijklp}^5 \tilde{u}, \partial_m \boldsymbol{w} \Big)_{\mathcal{L}^2}.\end{aligned}$$

Finally, applying formula (4.63), we obtain

$$\begin{aligned}\big\langle \partial_t^2 \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}, \boldsymbol{w}\big\rangle = & \Big(\big[f + a_{ij}^0 \partial_{ij}^2 \tilde{u} + \varepsilon a_{ij}^0 \chi_k \partial_{ijk}^3 \tilde{u} + \varepsilon^2 \big(a_{ij}^0 \theta_{kl} + a_{ij}^0 b_{kl}^2 - a_{ijkl}^2\big) \partial_{ijkl}^4 \tilde{u} \big], \boldsymbol{w} \Big)_{\mathcal{L}^2} \\ & + \Big(\big[\varepsilon \chi_k \partial_k f + \varepsilon^2 (\theta_{ij} + b_{ij}^2) \partial_{ij}^2 f \big], \boldsymbol{w} \Big)_{\mathcal{L}^2} + \big\langle \partial_t^2 \boldsymbol{\varphi}, \boldsymbol{w} \big\rangle + \big\langle \boldsymbol{\mathcal{R}}_1^\varepsilon \tilde{u}, \boldsymbol{w} \big\rangle.\end{aligned} \tag{4.68}$$

Next, the second term is computed as

$$\begin{aligned}\big\langle \mathcal{A}^\varepsilon \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}, \boldsymbol{w}\big\rangle = \Big(\big[& \quad \varepsilon^{-1} \big(- \nabla_y \cdot (a(\nabla_y \chi_k + e_k)) \big) \partial_k \tilde{u} \\ & + \quad \big(- \nabla_y \cdot (a(\nabla_y \theta_{ij} + e_i \chi_j)) - e_i^T a(\nabla_y \chi_j + e_j) \big) \partial_{ij}^2 \tilde{u} \\ & + \varepsilon^1 \ \big(- \nabla_y \cdot (a(\nabla_y \kappa_{ijk} + e_i \theta_{jk})) - e_i^T a(\nabla_y \theta_{jk} + e_j \chi_k) \big) \partial_{ijk}^3 \tilde{u} \\ & + \varepsilon^2 \ \big(- \nabla_y \cdot (a(\nabla_y \rho_{ijkl} + e_i \kappa_{jkl})) - e_i^T a(\nabla_y \kappa_{jkl} + e_j \theta_{kl}) \big) \partial_{ijkl}^4 \tilde{u} \big], \boldsymbol{w} \Big)_{\mathcal{L}^2} \\ & + \big\langle \mathcal{A}^\varepsilon \boldsymbol{\varphi}, \boldsymbol{w} \big\rangle + \big\langle \boldsymbol{\mathcal{R}}_2^\varepsilon \tilde{u}, \boldsymbol{w} \big\rangle,\end{aligned} \tag{4.69}$$

where

$$\big\langle \boldsymbol{\mathcal{R}}_2^\varepsilon \tilde{u}, \boldsymbol{w}\big\rangle = \varepsilon^3 \big(\big[-e_i^T a(\nabla_y \rho_{jklp} + e_j \kappa_{klp}) \partial_{ijklp}^5 \tilde{u} \big], \boldsymbol{w} \big)_{\mathcal{L}^2} + \big(a_{mi} \rho_{jklp} \partial_{ijklp}^5 \tilde{u}, \partial_m \boldsymbol{w} \big)_{\mathrm{L}^2}.$$

Now, we combine (4.68) and (4.69) and use cell problems (4.45a-4.45d) and (4.52) and obtain $(\partial_t^2 + \mathcal{A}^\varepsilon) \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}(t) = \big[f(t)\big] + \boldsymbol{\mathcal{R}}^\varepsilon \tilde{u}(t)$, where $\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u} = \boldsymbol{\mathcal{R}}_1^\varepsilon \tilde{u} + \boldsymbol{\mathcal{R}}_2^\varepsilon \tilde{u}$. Thanks to the regularity of the correctors and using (4.61), we verify estimate (4.64) for the remainder $\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u}$ and the proof of the lemma is complete. □

Proof of Theorem 4.2.4. As $(u^\varepsilon - \tilde{u})(t) \in \mathrm{W}_{\mathrm{per}}(\Omega)$, we have $\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(W)} = \| \big[u^\varepsilon - \tilde{u}\big] \|_{\mathrm{L}^\infty(\mathcal{W})}$. Hence, using the triangle inequality, we split the error as

$$\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(W)} \leq \|\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u} - \big[u^\varepsilon\big] \|_{\mathrm{L}^\infty(\mathcal{W})} + \| \big[\tilde{u}\big] - \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}\|_{\mathrm{L}^\infty(\mathcal{W})}. \tag{4.70}$$

Let us bound the two terms of the right hand side. The equation for u^ε (4.26) implies that $(\partial_t^2 + \mathcal{A}^\varepsilon)\big[u^\varepsilon(t)\big] = \big[f(t)\big]$ in $\mathcal{W}^*_{\mathrm{per}}(\Omega)$ for a.e. $t \in [0, T^\varepsilon]$. Lemma 4.2.8 thus implies that

$$(\partial_t^2 + \mathcal{A}^\varepsilon)(\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u} - \big[u^\varepsilon\big])(t) = \boldsymbol{\mathcal{R}}^\varepsilon \tilde{u}(t) \text{ in } \mathcal{W}^*_{\mathrm{per}}(\Omega) \quad \text{for a.e. } t \in [0, T^\varepsilon].$$

Applying Corollary 4.2.2 to $\boldsymbol{\eta} = \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u} - \big[u^\varepsilon\big]$, using estimate (4.64) and the definition of $\boldsymbol{\mathcal{B}}^\varepsilon$ in (4.62), we obtain

$$\|\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u} - \big[u^\varepsilon\big] \|_{\mathrm{L}^\infty(\mathcal{W})} \leq C\varepsilon\Big(\|g^1\|_{\mathrm{H}^4} + \|g^0\|_{\mathrm{H}^4} + |\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^5)} + |\partial_t^2 \tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^3)}\Big), \tag{4.71}$$

where C depends only on $\lambda, \Lambda, |b^2|_\infty, |a^2|_\infty, \|a\|_{\mathrm{W}^{2,\infty}(Y)}, Y$ and T. For the second term of (4.70), we use the definition of $\boldsymbol{\mathcal{B}}^\varepsilon$ (4.62) and estimate (4.54) and obtain

$$\| \big[\tilde{u}\big] - \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}\|_{\mathrm{L}^\infty(\mathcal{W})} \leq C\varepsilon\Big(\textstyle\sum_{k=1}^4 |\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^k)} + \|f\|_{\mathrm{L}^1(\mathrm{H}^2)}\Big), \tag{4.72}$$

where C depends only on $\|a\|_{\mathrm{W}^{2,\infty}(Y)}$ and $|b^2|_\infty$. Combining (4.70), (4.71) and (4.72), we obtain estimate (4.58) and the proof of the theorem is complete. □

4.2.6 A priori error estimate for a tensor with minimal regularity

In Theorem 4.2.4, the requirement on the regularity of the tensor, $a \in \mathrm{W}^{2,\infty}(Y)$, is severe. Indeed, general homogenization results only require the tensor to be bounded, i.e., $a \in \mathrm{L}^\infty(Y)$ (see Section 3.2.1). In this section, we prove an error estimate ensuring that the family of effective equations $\mathcal{E}$, defined in Definition 4.2.6, is still valid when the tensor is only bounded. For the result to hold, the only penalty is to require more regularity on the effective solutions.

To understand the idea of the proof, let us track the need for the regularity of the tensor in the proof of Theorem 4.2.4. First, observe that $a \in \mathrm{L}^\infty(Y)$ is sufficient for the correctors to be well defined in $\mathrm{H}^1_{\mathrm{per}}(Y)$ (see Section 4.2.3). The first need for the regularity $a \in \mathrm{W}^{2,\infty}(Y)$ is found in the definition of the adaptation operator in (4.62). Indeed, under the assumption $\tilde{u}(t) \in \mathrm{W}_{\mathrm{per}}(\Omega)$, to ensure that, for example, the term $\chi_i\big(\frac{\cdot}{\varepsilon}\big)\partial_i \tilde{u}(t)$ belongs to $\mathrm{L}^2(\Omega)$, we need $\chi_i \in \mathrm{L}^\infty(Y)$ (and similarly the other terms require the corrector to belong to $\mathrm{L}^\infty(Y)$). Observe however that under the stronger regularity $\tilde{u}(t) \in \mathrm{W}^{1,\infty}(\Omega)$, $\chi_i \in \mathrm{L}^2(Y)$ is sufficient for the adaptation to make sense. The second need for $a \in \mathrm{W}^{2,\infty}(Y)$ lies in the estimation of the remainder in Lemma 4.2.8. Indeed, to bound it we need the correctors to belong to $\mathrm{W}^{1,\infty}(Y)$. We will see that the remainder can still be estimated if the correctors only belongs to $\mathrm{H}^1_{\mathrm{per}}(Y)$ (Lemma 4.2.10).

The error estimate for a bounded tensor is stated in the following theorem.

Theorem 4.2.9. *Assume that the Y-periodic tensor satisfies $a(y) \in \mathrm{L}^\infty(Y)$. Furthermore, assume that the solution $\tilde{u}$ of* (4.56)*, the initial conditions and the right hand side satisfy the regularity*

$$\tilde{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^7(\Omega)), \quad \partial_t \tilde{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^6(\Omega)), \quad \partial_t^2 \tilde{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^5(\Omega)),$$
$$g^0 \in \mathrm{H}^6(\Omega), \quad g^1 \in \mathrm{H}^6(\Omega), \quad f \in \mathrm{L}^2(0, T^\varepsilon; \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^4(\Omega)),$$

and that b^2 and a^2 satisfy the relation (4.57)*. Then the following error estimate holds*

$$\begin{aligned} \|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} \leq C\varepsilon\Big(&\|g^1\|_{\mathrm{H}^6(\Omega)} + \|g^0\|_{\mathrm{H}^6(\Omega)} + \|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^4(\Omega))} \\ &+ \textstyle\sum_{k=1}^7 |\tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^k(\Omega))} + \sum_{k=3}^5 |\partial_t^2 \tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^k(\Omega))}\Big), \end{aligned} \tag{4.73}$$

where C depends only on $T, \lambda, \Lambda, |b^2|_\infty, |a^2|_\infty$, and Y.

Proof of the error estimate (Theorem 4.2.9)

The structure of the proof of Theorem 4.2.9 is similar to that of Theorem 4.2.4. We first verify that the regularity assumptions enable to define an adaptation operator $\boldsymbol{\mathcal{B}}^\varepsilon$ using the correctors defined in Section 4.2.3. We then split the error as

$$\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(W)} = \|\left[u^\varepsilon - \tilde{u}\right]\|_{\mathrm{L}^\infty(\mathcal{W})} \leq \|\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u} - \left[u^\varepsilon\right]\|_{\mathrm{L}^\infty(\mathcal{W})} + \|\left[\tilde{u}\right] - \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}\|_{\mathrm{L}^\infty(\mathcal{W})},$$

and estimate both terms separately. In particular, we verify that Lemma 4.2.8 still holds, i.e., that $\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}$ satisfies the same wave equation as u^ε up to a remainder. Finally, we prove that the terms composing the remainder can still be estimated in a convenient way under the new regularity assumptions (Lemma 4.2.10).

Let us first define the adaptation. As the tensor $a(y)$ belongs to $\mathrm{L}^\infty(Y)$, Lax–Milgram theorem ensures that the correctors $\chi_i, \theta_{ij}, \kappa_{ijk}$, and ρ_{ijkl}, defined in Section 4.2.3, belong to $\mathrm{H}^1_{\mathrm{per}}(Y)$ and satisfy

$$\|\chi_i\|_{\mathrm{H}^1(Y)}, \|\theta_{ij}\|_{\mathrm{H}^1(Y)}, \|\kappa_{ijk}\|_{\mathrm{H}^1(Y)}, \|\rho_{ijkl}\|_{\mathrm{H}^1(Y)} \leq C, \tag{4.74}$$

for some constant C depending only on λ, Λ, $|b^2|_\infty$, $|a^2|_\infty$ and Y. Next, as we assume $d \leq 3$, the embedding $\mathrm{H}^2_{\mathrm{per}}(\Omega) \hookrightarrow \mathcal{C}^0_{\mathrm{per}}(\bar{\Omega})$ is continuous and we verify that $f \in \mathrm{L}^2(0,T^\varepsilon; \mathcal{C}^2_{\mathrm{per}}(\bar{\Omega}))$. Hence, the right hand side of (4.52) belongs to $\mathrm{L}^2(0,T^\varepsilon; \mathcal{L}^2(\Omega))$ and $\boldsymbol{\varphi}$ exists, is unique, and satisfies the regularity (4.53). Consequently, the adaptation operator defined in (4.62) defines a linear operator (still denoted $\boldsymbol{\mathcal{B}}^\varepsilon$)

$$\boldsymbol{\mathcal{B}}^\varepsilon : \mathrm{L}^2(0,T^\varepsilon; \mathcal{C}^3_{\mathrm{per}}(\bar{\Omega})) \to \mathrm{L}^2(0,T^\varepsilon; \mathcal{W}^*_{\mathrm{per}}(\Omega)).$$

Thanks to the embedding $\mathrm{H}^2_{\mathrm{per}}(\Omega) \hookrightarrow \mathcal{C}^0_{\mathrm{per}}(\bar{\Omega})$, $\tilde{u}$ satisfies the regularity

$$\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon; \mathcal{C}^5_{\mathrm{per}}(\bar{\Omega})), \quad \partial_t \tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon; \mathcal{C}^4_{\mathrm{per}}(\bar{\Omega})), \quad \partial_t^2 \tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon; \mathcal{C}^3_{\mathrm{per}}(\bar{\Omega})), \tag{4.75}$$

which ensures that

$$\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon; \mathcal{W}_{\mathrm{per}}(\Omega)), \quad \partial_t \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon; \mathcal{L}^2(\Omega)), \quad \partial_t^2 \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon; \mathcal{W}^*_{\mathrm{per}}(\Omega)).$$

Note that the result of Lemma 4.2.8 still holds: $\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}$ satisfies

$$(\partial_t^2 + \mathcal{A}^\varepsilon)\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}(t) = \left[f(t)\right] + \boldsymbol{\mathcal{R}}^\varepsilon \tilde{u}(t) \quad \text{in } \mathcal{W}^*_{\mathrm{per}}(\Omega) \quad \text{for a.e. } t \in [0,T^\varepsilon], \tag{4.76}$$

where the remainder $\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u}$ is defined in the proof of Lemma 4.2.8. Thanks to the regularity of the correctors and $\tilde{u}$, we verify that $\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon; \mathcal{W}^*_{\mathrm{per}}(\Omega))$. In order to estimate the terms in the remainder $\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u}$, we prove the following lemma.

Lemma 4.2.10. *Assume that ε is bounded independently of* $\mathrm{diam}(\Omega)$. *Then $\gamma \in \mathrm{L}^2_{\mathrm{per}}(Y)$ and $v \in \mathrm{H}^2_{\mathrm{per}}(\Omega)$ satisfy the estimate*

$$\left\|\gamma\left(\tfrac{\cdot}{\varepsilon}\right) v\right\|_{\mathrm{L}^2(\Omega)} \leq C \|\gamma\|_{\mathrm{L}^2(Y)} \|v\|_{\mathrm{H}^2(\Omega)}, \tag{4.77}$$

for some constant C that depends only on Y, d and the bound on ε.

Proof. Let us first introduce some notations. Let $\ell \in \mathbb{R}^d$ be the period of the tensor a and assume without loss of generality that $Y = (0,\ell_1) \times \cdots \times (0,\ell_d)$ and $\Omega = (0,\omega_1) \times \cdots \times (0,\omega_d)$. As Ω satisfies the assumption (4.25), the numbers $N_i = \omega_i/\ell_i\varepsilon$ are integers and the cells constituting Ω

belongs to the set $\{\varepsilon(n\cdot\ell+Y) : 0\le n_i\le N_i-1\}$. Denoting $\Xi=\{\xi=n\cdot\ell : 0\le n_i\le N_i-1\}$, the domain Ω is then given by

$$\Omega=\operatorname{int}\Big(\bigcup_{\xi\in\Xi}\varepsilon(\xi+\bar{Y})\Big). \tag{4.78}$$

Hence, almost every $x\in\Omega$ satisfies $x=\varepsilon(\xi+y)$, with $\xi\in\Xi, y\in Y$, and the Y-periodic function γ satisfies $\gamma\big(\frac{x}{\varepsilon}\big)=\gamma(\xi+y)=\gamma(y)$, for a.e. $x=\varepsilon(\xi+y)\in\Omega$. Furthermore, let $Z\subset\mathbb{R}^d$ be an open set with a $\mathcal{C}^1$ boundary, that contains Y and is contained in the neighbor cells, i.e.,

$$Y\subset Z\subset N_Y=(-\ell_1,2\ell_1)\times\cdots\times(-\ell_d,2\ell_d).$$

For example, $Z=F_Y^{-1}(S)$, where S is the open sphere of diameter $\sqrt{d}$ centered in $(1/2,\cdots,1/2)$ (recall that $d\le 3$) and $F_Y : N_Y\to(-1,2)^d$ is a smooth change of coordinates. As Z has a $\mathcal{C}^1$ boundary and $d\le 3$, Sobolev embedding theorem ensures that the embedding $\mathrm{H}^2(Z)\hookrightarrow\mathcal{C}^0(\bar{Z})$ is continuous. Hence, there exists a constant C_Y, depending only Y, such that

$$\|w\|_{\mathcal{C}^0(\bar{Y})}\le\|w\|_{\mathcal{C}^0(\bar{Z})}\le C_Y\|w\|_{\mathrm{H}^2(Z)}\le C_Y\|w\|_{\mathrm{H}^2(N_Y)}\qquad\forall w\in\mathrm{H}^2(N_Y). \tag{4.79}$$

We now prove (4.77). Using (4.78), we have

$$\big\|\gamma\big(\tfrac{\cdot}{\varepsilon}\big)v\big\|^2_{\mathrm{L}^2(\Omega)}=\sum_{\xi\in\Xi}\int_{\varepsilon(\xi+Y)}\Big|\gamma\big(\tfrac{x}{\varepsilon}\big)v(x)\Big|^2\,\mathrm{d}x=\sum_{\xi\in\Xi}\int_Y\Big|\gamma(y)v\big(\varepsilon(\xi+y)\big)\Big|^2\varepsilon^d\,\mathrm{d}y,$$

where we made the change of variables $x=\varepsilon(\xi+y)$. As $v\in\mathrm{H}^2_{\mathrm{per}}(\Omega)\hookrightarrow\mathcal{C}^0_{\mathrm{per}}(\bar{\Omega})$, we have

$$\big\|\gamma\big(\tfrac{\cdot}{\varepsilon}\big)v\big\|^2_{\mathrm{L}^2(\Omega)}\le\|\gamma\|^2_{\mathrm{L}^2(Y)}\sum_{\xi\in\Xi}\varepsilon^d\|v_{\xi,\varepsilon}\|^2_{\mathcal{C}^0(\bar{Y})}, \tag{4.80}$$

where $v_{\xi,\varepsilon}$ is the function of $\mathcal{C}^0(\bar{Y})$ defined by $v_{\xi,\varepsilon}(y)=v\big(\varepsilon(\xi+y)\big)$. Using (4.79) gives $\|v_{\xi,\varepsilon}\|_{\mathcal{C}^0(\bar{Y})}\le C_Y\|v_{\xi,\varepsilon}\|_{\mathrm{H}^2(N_Y)}$. Furthermore, we have

$$\varepsilon^d\|v_{\xi,\varepsilon}\|^2_{\mathrm{H}^2(N_Y)}=\int_{N_Y}|v_{\xi,\varepsilon}(y)|^2\varepsilon^d\mathrm{d}y+\int_{N_Y}|\nabla_y v_{\xi,\varepsilon}(y)|^2\varepsilon^d\mathrm{d}y+\int_{N_Y}|\nabla_y^2 v_{\xi,\varepsilon}(y)|^2\varepsilon^d\mathrm{d}y.$$

As $\partial_{y_i}v_{\xi,\varepsilon}=\varepsilon\partial_{x_i}v$ and $\partial^2_{y_{ij}}v_{\xi,\varepsilon}=\varepsilon^2\partial^2_{x_{ij}}v$, the change of variable $x=\varepsilon(\xi+y)$ leads to

$$\big\|\gamma\big(\tfrac{\cdot}{\varepsilon}\big)v\big\|^2_{\mathrm{L}^2(\Omega)}\le C\|\gamma\|^2_{\mathrm{L}^2(Y)}\sum_{\xi\in\Xi}\|v\|^2_{\mathrm{H}^2(\varepsilon(\xi+N_Y))}\le(2d^2+1)C\|\gamma\|^2_{\mathrm{L}^2(Y)}\sum_{\xi\in\Xi}\|v\|^2_{\mathrm{H}^2(\varepsilon(\xi+Y))},$$

where we used that every cell $\varepsilon(\xi+Y)$ belongs to the neighborhoods of $(2d^2+1)$ cells. This proves (4.77) and the proof of the lemma is complete. □

Proof of Theorem 4.2.9. Thanks to Lemma 4.2.10 and (4.74), we verify that the remainder $\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u}$ in (4.76) (defined in the proof of Lemma 4.2.8) satisfies

$$\big\langle\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u}(t),\boldsymbol{w}\big\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}}=\big((\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u})_0(t),\boldsymbol{w}\big)_{\mathcal{L}^2}+\big((\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u})_1(t),\nabla\boldsymbol{w}\big)_{\mathrm{L}^2},$$

with the bound

$$\|(\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u})_0\|_{\mathrm{L}^\infty(\mathcal{L}^2(\Omega))}+\|(\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u})_1\|_{\mathrm{L}^\infty(\mathrm{L}^2(\Omega))}\le C\varepsilon^3\Big(\textstyle\sum_{k=5}^7|\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^k)}+\sum_{k=3}^5|\partial_t^2\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^k)}\Big),$$

where C depends on λ, Λ, $|b^2|_\infty$, $|a^2|_\infty$, and Y. Defining $\boldsymbol{\eta}=\big[u^\varepsilon\big]-\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}$, we verify thanks to Lemma 4.2.10 that

$$\|\boldsymbol{\eta}(0)\|_{\mathcal{L}^2(\Omega)}\le C\varepsilon\|g^0\|_{\mathrm{H}^6(\Omega)},\qquad\|\partial_t\boldsymbol{\eta}(0)\|_{\mathcal{L}^2(\Omega)}\le C\varepsilon\|g^1\|_{\mathrm{H}^6(\Omega)}.$$

Hence, Corollary 4.2.2 ensures that

$$\|\boldsymbol{\eta}\|_{\mathrm{L}^\infty(\mathcal{W})} \leq C\varepsilon\Big(\|g^1\|_{\mathrm{H}^6(\Omega)} + \|g^0\|_{\mathrm{H}^6(\Omega)} + \textstyle\sum_{k=5}^{7} |\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^k)} + \sum_{k=3}^{5} |\partial_t^2 \tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^k)}\Big). \tag{4.81}$$

Furthermore, using once again Lemma 4.2.10, we have

$$\|\,[\tilde{u}] - \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}\|_{\mathrm{L}^\infty(\mathcal{L}^2(\Omega))} \leq C\varepsilon\Big(\textstyle\sum_{k=1}^{7} |\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^k)} + \|f\|_{\mathrm{L}^1(\mathrm{H}^4)}\Big). \tag{4.82}$$

Finally, as $(u^\varepsilon - \tilde{u})(t) \in \mathrm{W}_{\mathrm{per}}(\Omega)$, the triangle inequality gives

$$\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(W)} = \|\,[u^\varepsilon - \tilde{u}]\,\|_{\mathrm{L}^\infty(\mathcal{W})} \leq \|\boldsymbol{\eta}\|_{\mathrm{L}^\infty(\mathcal{W})} + \|\,[\tilde{u}] - \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}\|_{\mathrm{L}^\infty(\mathcal{W})},$$

which, combined with (4.81) and (4.82), gives estimate (4.73) and that completes the proof of Theorem 4.2.9. □

4.2.7 Comparison with the coefficients obtained via Bloch wave expansion

In this section, we compare the effective equation obtained in [42, 43], presented in Section 4.1.1, with the effective equations of the family $\mathcal{E}$, defined in Definition 4.2.6. In particular, we prove that the effective equation from [42, 43] belongs to the family $\mathcal{E}$. This result attests that the derivations using Bloch wave expansion and asymptotic expansion lead to the same effective equations. Note that this comparison has been done independently in [18], with a particular attention to the elliptic case.

Let us recall the result of [42, 43], presented in Section 4.1.1. The effective equation is given by

$$\begin{aligned} &\partial_t^2 w^\varepsilon = a^0_{ij}\partial^2_{ij} w^\varepsilon + \varepsilon^2\big(E_{ij}\partial^2_{ij}\partial_t^2 w^\varepsilon - F_{ijmn}\partial^4_{ijmn} w^\varepsilon\big) \qquad \text{in } (0,T^\varepsilon]\times\mathbb{R}^d,\\ &w^\varepsilon(0,x) = g(x), \quad \partial_t w^\varepsilon(0,x) = 0, \end{aligned}$$

where the tensors $E \in \mathrm{Ten}^2(\mathbb{R}^d), F \in \mathrm{Ten}^4(\mathbb{R}^d)$ are built to satisfy the symmetry and sign (4.55), and such that

$$-C_{ijmn}\partial^4_{ijmn} = E_{ij}\partial^2_{ij} a^0_{mn}\partial^2_{mn} - F_{ijmn}\partial^4_{ijmn}. \tag{4.83}$$

In this section, we prove that the decomposition (4.83) is equivalent to the constraint involved in the definition of the family $\mathcal{E}$. In particular, that proves that (4.83) belongs to $\mathcal{E}$.

Let us give explicitly the formulas from [42] to compute C_{ijkl}. We consider the following cell problems: for $1 \leq i \leq j \leq k \leq d$, let $\psi_0^{e_j}, \psi_0^{e_i+e_j}, \psi_0^{e_i+e_j+e_k}$ be the Y-periodic zero mean solutions of

$$-\nabla\cdot\big(a\nabla\psi_0^{e_j}\big) = \mathrm{i}\nabla\cdot\big(ae_j\big), \tag{4.84a}$$

$$-\nabla\cdot\big(a\nabla\psi_0^{e_i+e_j}\big) = 2S^2_{ij}\Big\{\mathrm{i}\nabla\cdot\big(ae_i\psi_0^{e_j}\big) + \mathrm{i}e_i^T a\nabla\psi_0^{e_j} - a_{ij} + a^0_{ij}\Big\}, \tag{4.84b}$$

$$\begin{aligned} -\nabla\cdot\big(a\nabla\psi_0^{e_i+e_j+e_k}\big) = 3S^3_{ijk}\Big\{&\mathrm{i}\nabla\cdot\big(ae_i\psi_0^{e_j+e_k}\big) + \mathrm{i}e_i^T a\nabla\psi_0^{e_j+e_k}\\ &- 2a_{ij}\psi_0^{e_k} + 2a^0_{ij}\psi_0^{e_k}\Big\}. \end{aligned} \tag{4.84c}$$

Then, C is defined for $1 \leq i,j,k,l \leq d$ as

$$C_{ijkl} = \tfrac{1}{2}S^4_{ijkl}\big\{\big\langle a_{ij}\psi_0^{e_k+e_l}\big\rangle_Y\big\} - \tfrac{1}{6}\mathrm{i}S^4_{ijkl}\big\{\big\langle e_i^T a\nabla\psi_0^{e_j+e_k+e_l}\big\rangle_Y\big\}. \tag{4.85}$$

The cell problems (4.84a), (4.84b) and (4.84c) are similar to the ones obtained in (4.45a), (4.45b) and (4.45c) with asymptotic expansion. Let us determine their exact relation. First, note that

$\psi_0^{e_j}$ and $\psi_0^{e_i+e_j+e_k}$ are purely complex valued and $\psi_0^{e_i+e_j}$ are real valued (this ensures that C_{ijkl} is real). Second, consider $\chi_k, \theta_{ij}, \kappa_{ijk}$ the zero mean solutions of respectively problems (4.45a), (4.45b) and (4.45c). Using the unicity of a solution of an elliptic boundary value problem, we verify that

$$\psi_0^{e_j} = \mathrm{i}\chi_j, \quad \psi_0^{e_i+e_j} = -2\theta_{ij}, \quad \psi_0^{e_i+e_j+e_k} = -6\mathrm{i}\kappa_{ijk}. \tag{4.86}$$

We now show that the computed effective quantities are in fact the same. Using (4.86), we rewrite C_{ijkl} in (4.85) as

$$C_{ijkl} = S^4_{ijkl}\big\{ -\big\langle a_{ij}\theta_{kl}\big\rangle_Y - \big\langle e_i^T a\nabla\kappa_{jkl}\big\rangle_Y\big\} = |Y|S^4_{ijkl}\big\{ -\big(a\nabla\kappa_{jkl}, e_i\big)_Y - \big(ae_j\theta_{kl}, e_i\big)_Y\big\}.$$

As $\langle\theta_{kl}\rangle_Y = 0$, this expression is equal to the right hand side of (4.49). Hence, from (4.50), we verify that

$$C_{ijkl} = S^4_{ijkl}\big\{\big\langle a_{jk}\chi_l\chi_i\big\rangle_Y - \big\langle a\nabla\theta_{ji}\cdot\nabla\theta_{kl}\big\rangle_Y - a^0_{jk}\big(\chi_l,\chi_i\big)_Y\big\}. \tag{4.87}$$

Now, as E, F defined in [43] satisfies (4.83), thanks to (4.87) we can infer that they satisfy the constraint (4.57) from Theorem 4.2.4. As E, F satisfy the symmetry and sign (4.55) by construction, the effective equation defined in [43] belongs to the family $\mathcal{E}$, defined in Definition 4.2.6.

4.3 Computing the tensors of an effective equation

The family of effective equations $\mathcal{E}$, relies on an implicit definition of the tensors a^2, b^2 through a constraint (Definition 4.2.6). However, the constraint does not provide a way to compute a^2 and b^2 nor does it even ensure the existence of an effective equation. In this section, we prove in a constructive way that there exists tensors a^2, b^2 satisfying the requirements of Definition 4.2.6, thus ensuring that the family $\mathcal{E}$ is not empty. First, in the one-dimensional case, we show that the family can be parametrized by the normalization value of the first corrector. Second, in the multidimensional case, we provide an algorithm to compute the tensors of some effective equations in $\mathcal{E}$. These results were published in [13] and [14].

4.3.1 One-dimensional case

The computation of the effective coefficients in the one-dimensional case is particular. Indeed, we proved in [13] that the coefficients a^2 and b^2 in the effective equations can be computed with the solution of one single cell problem. This property leads to an explicit parametrization of the family of effective equations $\mathcal{E}$, defined in Definition 4.2.6. We prove here how this result is obtained with Theorem 4.2.4.

We consider the result of Theorem 4.2.4 in one dimension. In particular, let us rewrite the constraint (4.57) on the coefficients a^2, b^2 as

$$|Y|(a^2 - a^0b^2) = \big(a(\partial_y\theta+\chi), \chi - \partial_y\theta\big)_Y - a^0\big(\chi,\chi\big)_Y. \tag{4.88}$$

We now derive two relations that are only valid in the one-dimensional case. Noting that $a(\partial_y\chi+1) \in \mathrm{H}(\mathrm{div}, Y)$, we use integration by parts, the periodicity of $a(\partial_y\chi+1)$ and the cell problem for χ (4.45a) to obtain for any $y_1, y_2 \in Y$

$$a(\partial_y\chi+1)\Big|_{y_1}^{y_2} = a(\partial_y\chi+1)\big(H_{y_2} - H_{y_1}\big)\Big|_{\partial Y} - \int_Y \big(H_{y_2} - H_{y_1}\big)\partial_y\big(a(\partial_y\chi+1)\big)\mathrm{d}y = 0,$$

where H_{y_i} is the Heaviside step function centered in y_i. Hence, $a(\partial_y\chi+1)$ is constant on Y. Thanks to the definition of a^0 (see (4.41)), we conclude that $a(y)\big(\partial_y\chi(y)+1\big) = a^0\ \forall y \in Y$. A similar argument proves that $a(y)\big(\partial_y\theta(y)+\chi(y)\big) = C$ is constant on Y (see the cell problem for θ

in (4.45b)). Dividing this equality by $a(y)$, taking the mean over Y and using that $\langle 1/a\rangle_Y = 1/a^0$, we verify that $C = a^0\langle\chi\rangle_Y$. This equality, used in (4.88), leads to a constraint independent of θ:

$$a^2 - a^0 b^2 = a^0\langle\chi\rangle_Y\langle\chi - \partial_y\theta\rangle_Y - a^0\langle\chi^2\rangle_Y = a^0\langle\chi\rangle_Y^2 - a^0\langle\chi^2\rangle_Y. \tag{4.89}$$

We claim that the two following sets are equal:

$$E_1 = \big\{(b^2,a^2)\in\mathbb{R}^2_{\geq 0} : a^0b^2 - a^2 = a^0\langle\chi^2\rangle_Y - a^0\langle\chi\rangle_Y^2 \ \text{ for a } \chi\in\boldsymbol{\chi}\big\},$$
$$E_2 = \big\{(b^2,a^2)\in\mathbb{R}^2 : \ b^2 = \big\langle(\chi-\langle\chi\rangle_Y)^2\big\rangle_Y + \langle\chi\rangle_Y^2, \ \ a^2 = a^0\langle\chi\rangle_Y^2 \ \text{ for a } \chi\in\boldsymbol{\chi}\big\},$$

where we observe that $b_0^2 = \big\langle(\chi-\langle\chi\rangle_Y)^2\big\rangle_Y$ is non-negative and independent of the representative $\chi\in\boldsymbol{\chi}$. To prove the equality, first verify by a direct computation that $E_2\subset E_1$. Next, we show the converse inclusion. Let $(b^2,a^2)\in E_1$ and note that we can write

$$a^0b^2 - a^2 = a^0\big\langle(\chi-\langle\chi\rangle_Y)^2\big\rangle_Y = a^0b_0^2,$$

where $b_0^2 = \big\langle(\chi-\langle\chi\rangle_Y)^2\big\rangle_Y$ depends only on the zero mean element $\chi-\langle\chi\rangle_Y$ in $\boldsymbol{\chi}$ and is thus independent of $\langle\chi\rangle_Y$. Set $\mu = \sqrt{a^2/a^0}$ (a^2 is non-negative) and fix $\chi\in\boldsymbol{\chi}$ such that $\langle\chi\rangle_Y = \mu$ so that we have $a^2 = a^0\langle\chi\rangle_Y^2$. We then obtain

$$a^0b^2 = a^0\big\langle(\chi-\langle\chi\rangle_Y)^2\big\rangle_Y + a^2 = a^0\big(\big\langle(\chi-\langle\chi\rangle_Y)^2\big\rangle_Y + \langle\chi\rangle_Y^2\big),$$

which implies $(b^2,a^2)\in E_2$. As we have proved both inclusions, we obtain the equality $E_1 = E_2$. We can then parametrize the family of effective solutions as

$$\mathcal{E} = \{\tilde{u}_{\langle\chi\rangle} \text{ solution of (4.56) where } (b^2,a^2)\in E_2\}.$$

Observe that for $\langle\chi\rangle_Y = 0$, the coefficient a^2 vanishes and hence there is no fourth order operator $a^2\partial_x^4$ in the effective equation. This particular equation is the effective equation obtained in [72]. It is also the effective model on which the FE-HMM-L from [13] relies (see Chapter 7).

It is interesting to remark that the coefficient of the ill-posed equation introduced in [85] satisfies the condition (4.89). The equation is $\partial_t^2u = a^0\partial_x^2u - \varepsilon^2c\partial_x^4u$, where c has been proved in [19] to satisfy $c = -a^0\langle\chi^2\rangle_Y$, χ being the zero mean element of $\boldsymbol{\chi}$. Hence, the pair $(b^2,a^2) = (0,c)$ satisfies (4.89). However, c being negative, $(0,c)\notin E_1$, the corresponding equation is ill-posed and hence does not belong to $\mathcal{E}$.

4.3.2 Multidimensional case

In this section, we prove in the multidimensional case that the family of effective equations $\mathcal{E}$, defined in Definition 4.2.6, is not empty. Furthermore, we design a numerical procedure to construct the tensors of effective equations in $\mathcal{E}$. In particular, we need to construct a matrix whose sign is associated to a fourth order major symmetric tensor. The details of this construction are postponed to Section 4.3.3.

Recall that the family of effective equations $\mathcal{E}$ is defined by the pairs of tensors $b^2\in\mathrm{Ten}^2(\mathbb{R}^d)$, $a^2\in\mathrm{Ten}^4(\mathbb{R}^d)$ such that (see Definition 4.2.6)

$$b^2_{ij} = b^2_{ji}, \qquad b^2\eta\cdot\eta\geq 0 \ \ \forall\eta\in\mathbb{R}^d, \tag{4.90a}$$
$$a^2_{ijkl} = a^2_{klij}, \qquad a^2\xi:\xi\geq 0 \ \ \forall\xi\in\mathrm{Sym}^2(\mathbb{R}^d), \tag{4.90b}$$
$$S^4_{ijkl}\Big\{a^2_{ijkl} - a^0_{ij}b^2_{kl}\Big\} = S^4_{ijkl}\Big\{\big\langle a_{jk}\chi_l\chi_i\big\rangle_Y - \big\langle a\nabla_y\theta_{ji}\cdot\nabla_y\theta_{kl}\big\rangle_Y - a^0_{jk}\big\langle\chi_l\chi_i\big\rangle_Y\Big\}, \tag{4.90c}$$

where $\chi_i \in \boldsymbol{\chi}_i, \theta_{ij} \in \boldsymbol{\theta}_{ij}$ and $\boldsymbol{\chi}_i, \boldsymbol{\theta}_{ij}$ are the unique (class of) solutions of the cell problems (4.45a) and (4.45b), respectively.

Let us refer to Section 4.3.3 for the following definitions. A tensor $q \in \mathrm{Ten}^4(\mathbb{R}^d)$ is called major symmetric, if it satisfies the symmetry relation in (4.90b). Furthermore, a tensor $q \in \mathrm{Ten}^4(\mathbb{R}^d)$ is positive semidefinite if

$$q\xi : \xi = q_{ijkl}\xi_{ij}\xi_{kl} \geq 0 \quad \forall \xi \in \mathrm{Sym}^2(\mathbb{R}^d). \tag{4.91}$$

Similarly, a tensor $q \in \mathrm{Ten}^4(\mathbb{R}^d)$ is positive definite if the inequality in (4.91) is strict for all $\xi \in \mathrm{Sym}^2(\mathbb{R}^d)\backslash\{0\}$.

Let us call a pair of tensors a^2, b^2 *valid* if it satisfies the requirements (4.90). In the multidimensional case, constructing a valid pair a^2, b^2 is not as straightforward as in the one-dimensional case. As mentioned several times, the issue when looking for an effective equation is to obtain a well-posed equation. In particular, the sign of the tensor a^2 is crucial. For example, considering (4.90c), it is first tempting to try the pair

$$a^2_{ijkl} = q_{ijkl}, \quad b^2_{ij} = 0, \quad \text{where } \; q_{ijkl} = S^4_{ijkl}\Big\{\big\langle a_{jk}\chi_l\chi_i\big\rangle_Y - \big\langle a\nabla_y\theta_{ji}\cdot\nabla_y\theta_{kl}\big\rangle_Y - a^0_{jk}\big\langle\chi_l\chi_i\big\rangle_Y\Big\}.$$

However, the corresponding equation is precisely the one obtained in [85], which is known to be ill-posed due to the sign of q_{ijkl} (see [39]). The second logical attempt is the choice

$$a^2_{ijkl} = q_{ijkl}, \quad b^2_{ij} = \big\langle\chi_i\chi_j\big\rangle_Y, \quad \text{where } \; q_{ijkl} = S^4_{ijkl}\Big\{\big\langle a_{jk}\chi_l\chi_i\big\rangle_Y - \big\langle a\nabla_y\theta_{ji}\cdot\nabla_y\theta_{kl}\big\rangle_Y\Big\}.$$

Indeed, this pair satisfies (4.90a) and (4.90c). However, a simple numerical example (see Section 4.4.3) shows that q does not have a non-negative sign and thus does not fulfill (4.90b) in general. In fact, to construct a valid pair of effective tensors a^2, b^2, we need to use the freedom provided by the second minus sign of the right hand side in the constraint (4.90c) and the positive definiteness of the homogenized tensor a^0.

In this direction, we have the following lemma.

Lemma 4.3.1. *Let A and R be symmetric, positive definite matrices. Then, the tensor defined by $q_{ijkl} = A_{jk}R_{il}$ is positive definite.*

Proof. As R is symmetric positive definite, the Cholesky factorization ensures the existence of an invertible matrix H such that $R = H^TH$. As A is positive definite, for $\xi \in \mathrm{Sym}^2(\mathbb{R}^d)$ we have

$$q\xi : \xi = A_{jk}R_{il}\xi_{ij}\xi_{kl} = A_{jk}\big(H_{mi}\xi_{ij}\big)\big(H_{ml}\xi_{lk}\big) = (\xi H_m)^T A \xi H_m \geq 0,$$

where we denoted $H_m = (H_{m1}, \ldots, H_{md})^T$. Now, assume that the equality holds. Then, as A is positive definite, it must hold $\xi H_m = 0$ for all $m = 1, \ldots, d$, or equivalently $\xi H^T = 0$. As H is regular so is H^T and we conclude that $\xi = 0$. This proves that the inequality is strict for $\xi \neq 0$ and the proof of the lemma is complete. □

Thanks to this result, we are able to prove the existence of effective equations in the family. Let us first define the symmetrization operator $S^{2,2}$:

$$\big(S^{2,2}(q)\big)_{ijkl} = S^{2,2}_{ij,kl}\{q_{ijkl}\} = S^2_{ij}\{S^2_{kl}\{q_{ijkl}\}\} \quad \forall q \in \mathrm{Ten}^4(\mathbb{R}^d).$$

Referring to Remark 4.3.5, if q satisfies $q_{ijkl} = q_{lkji}$ for $1 \leq i, j, k, l \leq d$, then the tensor $S^{2,2}(q)$ is major symmetric. Note that the use of $S^{2,2}$ instead of S^4 is not strictly needed and is discussed below. Then, let $\{R^\delta\}_{\delta>0} \subset \mathrm{Sym}^2(\mathbb{R}^d)$ be a sequence of parametrized symmetric, positive definite

matrices such that the smallest eigenvalue of R^δ increases as δ increases. We verify that for sufficiently large values of δ, the tensors

$$a^2_{ijkl} = S^{2,2}_{ij,kl}\Big\{\big\langle a_{jk}\chi_i\chi_l\big\rangle_Y - \big\langle a\nabla_y\theta_{ji}\cdot\nabla_y\theta_{kl}\big\rangle_Y + a^0_{jk}R^\delta_{il}\Big\}, \qquad b^2_{ij} = \langle\chi_i\chi_j\rangle_Y + R^\delta_{ij}, \tag{4.92}$$

satisfy all the requirements (4.90) (we recall that a^0 is positive definite). This construction proves that the family of effective equations $\mathcal{E}$, defined in Definition 4.2.6, is not empty (see Figure 4.6, Section 4.4.3). Note that a related construction of effective tensors has been independently described theoretically in [18].

We still need a practical way for the computation of R^δ such that the pair a^2, b^2 in (4.92) is valid. To that end, the following tool is introduced in Section 4.3.3. For a major symmetric tensor $q \in \mathrm{Ten}^4(\mathbb{R}^d)$, there exists a bijective map $\nu : \mathrm{Sym}^2(\mathbb{R}^d) \to \mathbb{R}^{N(d)}$, where $N(d) = \binom{d+1}{2}$, and a matrix $M(q) \in \mathrm{Sym}^2(\mathbb{R}^{\mathbb{N}(d)})$ such that

$$q\xi : \eta = M(q)\nu(\xi)\cdot\nu(\eta) \quad \forall \xi,\eta \in \mathrm{Sym}^2(\mathbb{R}^d). \tag{4.93}$$

In particular, q is positive (semi)definite if and only if $M(q)$ is positive (semi)definite (see Lemma 4.3.6). Thanks to this tool, we now have a constructive method to obtain effective equations. In the following lemma, we apply (4.92) with $R^\delta = \delta I$, where I is the $d \times d$ identity matrix.

Lemma 4.3.2. *Define the tensor* $\check{a}^2_{ijkl} = S^{2,2}_{ij,kl}\Big\{\big\langle a_{jk}\chi_i\chi_l\big\rangle_Y - \big\langle a\nabla_y\theta_{ji}\cdot\nabla_y\theta_{kl}\big\rangle_Y\Big\}$, *and the matrices* $A^2 = M(\check{a}^2)$, $A^0 = M\big(S^{2,2}_{ij,kl}\{a^0_{jk}I_{il}\}\big)$. *The minimal eigenvalues of* A^2 *and* A^0 *are denoted* $\lambda_{\min}(A^2)$ *and* $\lambda_{\min}(A^0)$, *respectively. Then the tensor (recall that* $\{\cdot\}_+ = \max\{0,\cdot\}$*)*

$$a^2_{ijkl} = \check{a}^2_{ijkl} + \delta S^{2,2}_{ij,kl}\big\{a^0_{jk}I_{il}\big\}, \qquad \textit{with} \;\; \delta \geq \delta^* = \left\{ -\frac{\lambda_{\min}(A^2)}{\lambda_{\min}(A^0)}\right\}_+,$$

is positive semidefinite (i.e., it satisfies $a^2\xi : \xi \geq 0 \;\; \forall \xi \in \mathrm{Sym}^2(\mathbb{R}^d)$*).*

Proof. First, as A^2 and A^0 are symmetric matrices it is clear that $\lambda_{\min}(A^2)$ and $\lambda_{\min}(A^0)$ are real and thanks to Lemma 4.3.1 and (4.93) it holds $\lambda_{\min}(A^0) > 0$. Furthermore, $\lambda_{\min}(A^2) \leq (A^2v\cdot v)/(v\cdot v)$ for any $v \in \mathbb{R}^{N(d)}$ and similarly for A^0. Now, if A^2 is positive semidefinite, then $\delta^* = 0$ and the tensor a^2 is positive semidefinite for any $\delta \geq 0$. Next, assume that $\lambda_{\min}(A^2) < 0$. We verify then that for any $v \in \mathbb{R}^{N(d)}$,

$$\delta^* = -\frac{\lambda_{\min}(A^2)}{\lambda_{\min}(A^0)} \geq -\frac{A^2v\cdot v}{A^0v\cdot v}.$$

Then, let $\xi \in \mathrm{Sym}^2(\mathbb{R}^d)$ and denote $v = \nu(\xi)$ (see (4.93)). Decomposing $\delta = \delta^* + \Delta\delta$ with $\Delta\delta \geq 0$, we obtain

$$a^2\xi : \xi = A^2v\cdot v + \delta^*A^0v\cdot v + \Delta\delta A^0v\cdot v \geq 0.$$

The proof of the lemma is complete. □

Let us discuss alternatives to the process to construct valid pairs a^2, b^2 provided by Lemma 4.3.2. Indeed, the choice of the positive definite tensor $S^{2,2}_{ij,kl}\{a^0_{jk}I_{il}\}$ is arbitrary. As an alternative, we can use, for example, $S^{2,2}_{ij,kl}\{a^0_{jk}a^0_{il}\}$, and obtain the subset of the family defined by the pairs

$$\begin{aligned} &A^2 = M(\check{a}^2), \quad A^0 = M\big(S^{2,2}_{ij,kl}\{a^0_{jk}a^0_{il}\}\big), \quad \delta \geq \delta^* = \left\{ -\frac{\lambda_{\min}(A^2)}{\lambda_{\min}(A^0)}\right\}_+, \\ &a^2_{ijkl} = \check{a}^2_{ijkl} + \delta S^{2,2}_{ij,kl}\big\{a^0_{jk}a^0_{il}\big\}, \qquad b^2_{ij} = \langle\chi_i\chi_j\rangle_Y + \delta a^0_{ij}. \end{aligned} \tag{4.94}$$

Numerical experiments indicates that this choice works equally. It is however interesting to note that in the case of a locally periodic tensor, studied in Chapter 6, this second choice is imposed naturally (see Section 6.2.2, in particular Lemma 6.2.8). Another alternative, is to use the full symmetrization S^4, instead of $S^{2,2}$. Indeed, we have the following lemma.

Lemma 4.3.3. *If A is a symmetric, positive definite matrix, then the tensor defined by $S^4_{ijkl}\{A_{ij}A_{kl}\}$ is positive definite.*

Proof. Let $\xi \in \mathrm{Sym}^2(\mathbb{R}^d)\backslash\{0\}$. As

$$A_{ij}A_{kl}\xi_{ij}\xi_{kl} = \big(A_{ij}\xi_{ij}\big)^2 \geq 0 \quad \forall \xi \in \mathrm{Sym}^2(\mathbb{R}^d),$$

the tensor $A_{ij}A_{kl}$ is positive semidefinite. Note that it is not positive definite in general because A may have zero entries. Thanks to the symmetry of A and ξ, we verify that

$$S^4_{ijkl}\{A_{ij}A_{kl}\}\xi_{ij}\xi_{kl} = \tfrac{1}{2}\big(A_{ij}A_{kl} + A_{jk}A_{il}\big)\xi_{ij}\xi_{kl}.$$

As Lemma 4.3.1 ensures the second term to be strictly positive for $\xi \neq 0$, we obtain the positive definiteness of $S^4_{ijkl}\{A_{ij}A_{kl}\}$. □

Hence, using $S^4\{a^0_{ij}a^0_{kl}\}$ in the above process also leads to effective equations. Nevertheless, the following result might be used to argue that the choices $S^{2,2}_{ij,kl}\{a^0_{jk}I_{il}\}$ and $S^{2,2}_{ij,kl}\{a^0_{jk}a^0_{il}\}$ are more natural (note that it is a complement to Lemma 4.3.4).

Lemma 4.3.4. *Let A, R be symmetric, positive semidefinite matrices with respective minimal eigenvalues $\lambda_{\min}(A), \lambda_{\min}(R)$. Then*

$$S^{2,2}_{ij,kl}\{A_{jk}R_{il}\}\xi_{ij}\xi_{kl} = A_{jk}R_{il}\xi_{ij}\xi_{kl} \geq \lambda_{\min}(A)\lambda_{\min}(R)\|\xi\|_F \quad \forall \xi \in \mathrm{Sym}^2(\mathbb{R}^d),$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

Proof. For $i = 1, \dots, d$, we denote $\lambda_i(A)$, $\lambda_i(R)$ the eigenvalues of A and R, respectively. The symmetry of A and R ensures the existence of orthogonal matrices Q, P such that $A = Q^T L(A) Q$ and $R = P^T L(R) P$, where $L(A) = \mathrm{diag}\big(\lambda_1(A), \dots, \lambda_d(A)\big)$ and $L(R)$ similarly. Using the symmetry of ξ, we thus have

$$\begin{aligned} A_{jk}R_{il}\xi_{ij}\xi_{kl} &= \lambda_n(A)\lambda_m(R)Q_{mj}Q_{mk}P_{ni}P_{nl}\xi_{ij}\xi_{kl} = \textstyle\sum_{mn} \lambda_n(A)\lambda_m(R)\big(\sum_{ij} Q_{mj}P_{ni}\xi_{ij}\big)^2 \\ &\geq \lambda_{\min}(A)\lambda_{\min}(R)\textstyle\sum_{mn}\big(\sum_{ij} Q_{mj}P_{ni}\xi_{ij}\big)^2. \end{aligned} \tag{4.95}$$

Denoting $v^m_i = \sum_j Q_{mj}\xi_{ij}$ and using the orthogonality of P, we get

$$\textstyle\sum_{mn}\big(\sum_{ij} Q_{mj}P_{ni}\xi_{ij}\big)^2 = \sum_m \sum_n \big(\sum_i P_{ni}v^m_i\big)^2 = \sum_m \big(v^i_j\big)^2 = \sum_i \sum_m \big(\sum_j Q_{mj}\xi_{ij}\big)^2.$$

Using then the orthogonality of Q, we obtain

$$\textstyle\sum_{mn}\big(\sum_{ij} Q_{mj}P_{ni}\xi_{ij}\big)^2 = \sum_{ij}\xi^2_{ij} = \|\xi\|_F.$$

Combined with (4.95), this equality concludes the proof of the lemma. □

4.3.3 Matrix associated to a major symmetric tensor of order four

In this section, from a fourth order major symmetric tensor $q \in \mathrm{Ten}^4(\mathbb{R}^d)$, we build a matrix whose eigenvalues are associated to the sign of q. This tool is used to construct pairs of tensors defining effective equations (see the previous section).

Let us first give some definitions. A tensor $q \in \mathrm{Ten}^4(\mathbb{R}^d)$ is *major symmetric* if

$$q_{ijkl} = q_{klij} \qquad 1 \le i,j,k,l \le d. \tag{4.96}$$

A tensor $q \in \mathrm{Ten}^4(\mathbb{R}^d)$ is *minor symmetric* if

$$q_{ijkl} = q_{jikl} = q_{ijlk} \qquad 1 \le i,j,k,l \le d. \tag{4.97}$$

We say that $q \in \mathrm{Ten}^4(\mathbb{R}^d)$ is *positive definite* if

$$q\xi : \xi > 0 \qquad \forall \xi \in \mathrm{Sym}^2(\mathbb{R}^d)\backslash\{0\},$$

and is *positive semidefinite* if

$$q\xi : \xi \ge 0 \qquad \forall \xi \in \mathrm{Sym}^2(\mathbb{R}^d).$$

Remark 4.3.5. Let $q \in \mathrm{Ten}^4(\mathbb{R}^d)$ be a minor symmetric tensor. Then q is major symmetric if and only if it satisfies the symmetry relation $q_{ijkl} = q_{lkji}$ $1 \le i,j,k,l \le d$. Indeed, if q is major symmetric, using the minor symmetry we have

$$q_{ijkl} = \tfrac{1}{4}\big(q_{ijkl} + q_{jikl} + q_{ijlk} + q_{jilk}\big) = \tfrac{1}{4}\big(q_{klij} + q_{klji} + q_{lkij} + q_{lkji}\big) = q_{lkji}.$$

And conversely, if $q_{ijkl} = q_{lkji}$, using the minor symmetries gives

$$q_{ijkl} = \tfrac{1}{4}\big(q_{ijkl} + q_{jikl} + q_{ijlk} + q_{jilk}\big) = \tfrac{1}{4}\big(q_{lkji} + q_{lkij} + q_{klji} + q_{klij}\big) = q_{klij}.$$

Consequently, the major symmetry relation is sometimes defined alternatively as $q_{ijkl} = q_{lkji}$.

For $q \in \mathrm{Ten}^4(\mathbb{R}^d)$, we define a matrix $M(q)$ such that q is positive (semi)definite if and only if $M(q)$ is positive (semi)definite. First, let us define the minor symmetrization of q as

$$\bar{q}_{ijkl} = S^{2,2}_{ij,kl}\big\{q_{ijkl}\big\} = \tfrac{1}{4}\big(q_{ijkl} + q_{jikl} + q_{ijlk} + q_{jilk}\big),$$

which satisfies (4.97) and $q\xi : \eta = \bar{q}\xi : \eta$ for any $\xi, \eta \in \mathrm{Sym}^2(\mathbb{R}^d)$. The tensor $\bar{q}$ defines a linear map $\mathrm{Sym}^2(\mathbb{R}^d) \to \mathrm{Sym}^2(\mathbb{R}^d)$, $\xi \mapsto \bar{q}\xi$ as

$$(\bar{q}\xi)_{ij} = \bar{q}_{ijkl}\xi_{kl} = \sum_{k=1}^{d} \bar{q}_{ijkk}\xi_{kk} + 2\sum_{k=1}^{d}\sum_{l=k+1}^{d} \bar{q}_{ijkl}\xi_{kl}. \tag{4.98}$$

In order to build a matrix associated to this linear map, we define the sets of indices

$$J = \{(i,j) : 1 \le i \le j \le d\}, \qquad I = \{1, \dots, N(d)\},$$

where $N(d) = \binom{d+1}{2}$ is the number of distinct entries of a symmetric matrix in $\mathrm{Sym}^2(\mathbb{R}^d)$. Let $\ell^{-1} : J \to I$ be the one to one map given by $\ell^{-1}(i,j) = K^d_{ij}$, where K^d is the symmetric $d \times d$ matrix given by (fill the diagonal, then successively the $d-1$ upper diagonal rows)

$$K^d = \begin{pmatrix} 1 & d+1 & \cdots & \cdots & 2d-1 \\ & 2 & 2d & \cdots & 3d-3 \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & N(d) \\ & & & & d \end{pmatrix}.$$

Define then the bijective map $\nu : \mathrm{Sym}^2(\mathbb{R}^d) \to \mathbb{R}^{N(d)}, \xi \to \nu(\xi)$, by $\big(\nu(\xi)\big)_m = \xi_{\ell(m)}$ and note that its inverse is given for $v \in \mathbb{R}^{N(d)}$ by $\big(\nu^{-1}(v)\big)_{ij} = v_{\ell^{-1}(i,j)}$. Defining the linear map $Q : \mathbb{R}^{N(d)} \to \mathbb{R}^{N(d)}$, $Q = \nu \circ \bar{q} \circ \nu^{-1}$, we verify that for $v \in \mathbb{R}^{N(d)}$

$$\big(Qv\big)_m = \sum_{k=1}^{d} \bar{q}_{\ell(m)\ell(k)} v_k + 2 \sum_{k=d+1}^{N(d)} \bar{q}_{\ell(m)\ell(k)} v_k.$$

Hence, denoting $\{e_i\}_{i=1}^{N(d)}$ the canonical basis of $\mathbb{R}^{N(d)}$, the matrix associated to the linear map (4.98) is given in the basis $\big\{e_1, \dots e_d, \frac{1}{2}e_{d+1}, \dots, \frac{1}{2}e_{N(d)}\big\}$ by $Q_{mn} = \bar{q}_{\ell(m)\ell(n)}$. We can then show that for any $\xi, \eta \in \mathrm{Sym}^2(\mathbb{R}^d)$, we have

$$\begin{aligned}\bar{q}\xi : \eta &= \sum_{ik=1}^{d} \bar{q}_{iikk}\xi_{ii}\eta_{kk} + 2\Bigg(\sum_{\substack{ik=1\\k<l}}^{d} \bar{q}_{iikl}\xi_{ii}\eta_{kl} + \sum_{\substack{jk=1\\i<j}}^{d} \bar{q}_{ijkk}\xi_{ij}\eta_{kk}\Bigg) + 4\sum_{\substack{i<j\\k<l}}^{d} \bar{q}_{ijkl}\xi_{ij}\eta_{kl} \\ &= \nu(\xi)^T P^T Q P \nu(\eta),\end{aligned}$$

where

$$P_{mn} = \delta_{mn} z_n, \qquad z_n = \begin{cases} 1 & \text{if } 1 \le n \le d, \\ 2 & \text{if } d+1 \le n \le N(d). \end{cases}$$

Hence, we define the matrix associated to $\bar{q}$ as $\tilde{M}(\bar{q}) = P^T Q P$, given by

$$\big(\tilde{M}(\bar{q})\big)_{mn} = z_m z_n \bar{q}_{\ell(m)\ell(n)}. \tag{4.99}$$

For $d = 2, 3$ $\tilde{M}(\bar{q})$ is given respectively as

$$\tilde{M}(\bar{q}) = \begin{pmatrix} \bar{q}_{1111} & \bar{q}_{1122} & 2\bar{q}_{1112} \\ & \bar{q}_{2222} & 2\bar{q}_{2212} \\ & & 4\bar{q}_{1212} \end{pmatrix}, \qquad \tilde{M}(\bar{q}) = \begin{pmatrix} \bar{q}_{1111} & \bar{q}_{1122} & \bar{q}_{1133} & 2\bar{q}_{1112} & 2\bar{q}_{1113} & 2\bar{q}_{1123} \\ & \bar{q}_{2222} & \bar{q}_{2233} & 2\bar{q}_{2212} & 2\bar{q}_{2213} & 2\bar{q}_{2223} \\ & & \bar{q}_{3333} & 2\bar{q}_{3312} & 2\bar{q}_{3313} & 2\bar{q}_{3323} \\ & & & 4\bar{q}_{1212} & 4\bar{q}_{1213} & 4\bar{q}_{1223} \\ & & & & 4\bar{q}_{1313} & 4\bar{q}_{1323} \\ & & & & & 4\bar{q}_{2323} \end{pmatrix}.$$

We summarize the results of this section in the following lemma.

Lemma 4.3.6. *Let $q \in \mathrm{Ten}^4(\mathbb{R}^d)$ be a tensor satisfying the major symmetry* (4.96) *and let $M(q) = \tilde{M}\big(S^{2,2}(q)\big)$, where $\tilde{M}$ is defined in* (4.99) *and $S^{2,2}_{ij,kl} = S^2_{ij}\{S^2_{kl}\{\cdot\}\}$. Then*

$$q\xi : \eta = M(q)\nu(\xi) \cdot \nu(\eta) \quad \forall \xi, \eta \in \mathrm{Sym}^2(\mathbb{R}^d).$$

In particular, q is positive (semi)definite if and only if $M(q)$ is positive (semi)definite.

4.3.4 Algorithm to compute the tensors of an effective equation

As discussed in Section 4.3.2, Lemma 4.3.2 provides a procedure for the construction of pairs of tensors of some effective equations in the family $\mathcal{E}$, defined in Definition 4.2.6. We present here the full algorithm to compute the effective tensors a^0, b^2 and a^2 of one of the corresponding effective equations. Note that the algorithm can easily be modified to obtain different effective equations. We emphasize that this algorithm is appropriate for dimensions $d \ge 2$ as a simpler one is given for $d = 1$ in Section 4.3.1.

Let $J(d) \subset \{1, \dots, d\}^4$ denotes the set of indices of distinct entries of a major an minor symmetric tensor of order 4. Let $M(q) = \tilde{M}\big(S^{2,2}(q)\big)$, where $\tilde{M}$ is defined in (4.99) and $S^{2,2}_{ij,kl} = S^2_{ij}\{S^2_{kl}\{\cdot\}\}$.

Note that we compute here the homogenized tensor under a naturally symmetric form (see Lemma 3.3.1).

Algorithm 4.3.7. Computation of the tensors a^0, b^2, a^2 of an effective equation in the family $\mathcal{E}$.

1. for $1 \le k \le d$ find $\chi_k \in \mathrm{W}_{\mathrm{per}}(Y)$ such that $\forall w \in \mathrm{W}_{\mathrm{per}}(Y)$
   $\big(a\nabla_y\chi_k, \nabla_y w\big)_Y = -\big(ae_k, \nabla_y w\big)_Y$,
2. for $1 \le i \le j \le d$ compute
   $a^0_{ij} = a^0_{ji} = -\big\langle a\nabla_y\chi_j \cdot \nabla_y\chi_i\big\rangle_Y + \big\langle ae_i \cdot e_j\big\rangle_Y$,
3. for $1 \le i \le j \le d$ find $\theta_{ij} = \theta_{ji} \in \mathrm{W}_{\mathrm{per}}(Y)$ such that $\forall w \in \mathrm{W}_{\mathrm{per}}(Y)$
   $\big(a\nabla_y\theta_{ij}, \nabla_y w\big)_Y = S^2_{ij}\Big\{ -\big(a\chi_j e_i, \nabla_y w\big)_Y + \big(a(\nabla_y\chi_j + e_j) - a^0 e_j, e_i w\big)_Y\Big\}$,
4. for $(i,j,k,l) \in J(d)$ compute
   $\check{a}^2_{ijkl} = S^{2,2}_{ij,kl}\Big\{\big\langle a_{jk}\chi_i\chi_l\big\rangle_Y - \big\langle a\nabla_y\theta_{ij}\cdot\nabla_y\theta_{kl}\big\rangle_Y\Big\}$,
5. build the matrices $A^2 = M(\check{a}^2)$, $A^0 = M\big(S^{2,2}_{ij,kl}\{a^0_{jk}I_{il}\}\big)$ and compute
   $\delta^* = \left\{ -\dfrac{\lambda_{\min}(A^2)}{\lambda_{\min}(A^0)}\right\}_+$,
6. for $1 \le i \le j \le d$ compute
   $b^2_{ij} = b^2_{ji} = \big\langle \chi_i\chi_j\big\rangle_Y + \delta^* I_{ij}$,
7. for $(i,j,k,l) \in J(d)$ compute
   $a^2_{ijkl} = \check{a}^2_{ijkl} + \delta^* S^{2,2}_{ij,kl}\big\{a^0_{jk}I_{il}\big\}$.

4.3.5 Subset of the family parametrizable by the mean of the first corrector

In Section 4.3.1, we have seen that in the one-dimensional case, the family $\mathcal{E}$ of effective equations, defined in Definition 4.2.6, can be parameterized by the normalization value of the first corrector. This parametrization gives a connection between the unique class of corrector $\boldsymbol{\chi} \in \mathcal{W}_{\mathrm{per}}(Y)$ and the family $\mathcal{E}$. In higher dimensions, the different subsets of the family constructed in Section 4.3.2 do not a priori satisfy a similar relation. In this section, we show that, in the general case, a subset of the family $\mathcal{E}$ can indeed be parametrized by the normalization parameter of the first correctors, but this subset might be empty.

Let us first introduce some notations. For $i = 1, \dots, d$, let us parametrize the class of the first corrector $\boldsymbol{\chi}_i$ (solution of (4.45a)) by its mean, denoted $\mu_i \in \mathbb{R}$. Explicitly, we denote χ^μ_i the element of $\boldsymbol{\chi}_i$ such that $\langle\chi^\mu_i\rangle_Y = \mu_i$. Furthermore, let θ^μ_{ij} denote the zero mean element of the corresponding class of second correctors $\boldsymbol{\theta}_{ij}$, i.e., the solution of (4.45b) that corresponds to χ^μ_i. Note that the normalization of θ^μ_{ij} has no influence and we pick the zero mean element for simplicity. In particular, $\chi^\mu_i \in \mathrm{H}^1_{\mathrm{per}}(Y)$ and $\theta^\mu_{ij} \in \mathrm{W}_{\mathrm{per}}(Y)$ satisfy for any test functions $w \in \mathrm{H}^1_{\mathrm{per}}(Y)$:

$$\big(a\nabla_y\chi^\mu_i, \nabla_y w\big)_Y = -\big(ae_i, \nabla_y w\big)_Y, \tag{4.100a}$$

$$\big(a\nabla_y\theta^\mu_{ij}, \nabla_y w\big)_Y = S^2_{ij}\big\{ -\big(ae_i\chi^\mu_j, \nabla_y w\big)_Y + \big(a(\nabla_y\chi^\mu_j + e_j) - a^0 e_j, e_i w\big)_Y\big\}. \tag{4.100b}$$

Hence, χ^0_i is the zero mean element of $\boldsymbol{\chi}_i$ and θ^0_{ij} is the zero mean element of $\boldsymbol{\theta}_{ij}$, corresponding to χ^0_i. We verify that

$$\chi^\mu_i = \chi^0_i + \mu_i, \qquad \theta^\mu_{ij} = \theta^0_{ij} + S^2_{ij}\{\mu_i\chi_j\}. \tag{4.101}$$

Indeed, plugging the right hand sides of these equalities in (4.100a) and (4.100b), respectively, and using the uniqueness of χ_i^μ and θ_{ij}^μ, we obtain (4.101). Recall now that the family of effective equations, defined in Definition 4.2.6, consists in the pairs of tensors a^2, b^2 satisfying (4.55) and (4.57). In particular, a^2, b^2 must satisfy

$$|Y|(a^2_{ijkl} - a^0_{ij}b^2_{kl}) =_S \left(a_{jk}\chi_l^\mu, \chi_i^\mu\right)_Y - \left(a\nabla_y\theta_{ji}^\mu, \nabla_y\theta_{kl}^\mu\right)_Y - a^0_{jk}\left(\chi_l^\mu, \chi_i^\mu\right)_Y, \tag{4.102}$$

where $=_S$ signifies that the equality holds up to symmetries, i.e., $b, c \in \mathrm{Ten}^n(\mathbb{R}^d)$ satisfy $b =_S c$ iff $S^n(b) = S^n(c)$. Let us prove that

$$\begin{aligned} &\left(a_{jk}\chi_l^\mu, \chi_i^\mu\right)_Y - \left(a\nabla_y\theta_{ji}^\mu, \nabla_y\theta_{kl}^\mu\right)_Y =_S \left(a_{jk}\chi_l^0, \chi_i^0\right)_Y - \left(a\nabla_y\theta_{ji}^0, \nabla_y\theta_{kl}^0\right)_Y + a^0_{jk}\mu_i\mu_l, \\ &\left(\chi_i^\mu, \chi_j^\mu\right)_Y = \left(\chi_i^0, \chi_j^0\right)_Y + \mu_i\mu_j. \end{aligned} \tag{4.103}$$

Using (4.101), the second equality is proved by a direct computation (recall that $\langle\chi_i^0\rangle_Y = 0$). Let us prove the first equality. Using (4.101), we rewrite

$$\begin{aligned} \left(a_{jk}\chi_l^\mu, \chi_i^\mu\right)_Y - \left(a\nabla_y\theta_{ji}^\mu, \nabla_y\theta_{kl}^\mu\right)_Y =_S{}& \left(a_{jk}\chi_l^0, \chi_i^0\right)_Y - \left(a\nabla_y\theta_{ji}^0, \nabla_y\theta_{kl}^0\right)_Y \\ &+ 2\mu_i\Big(\left(a\nabla_y\chi_j^0, \nabla_y\theta_{kl}^0\right)_Y - \left(ae_j, e_k\chi_l^0\right)_Y\Big) + \mu_i\mu_l\Big(\left(a\nabla_y\chi_j^0, \nabla_y\chi_k^0\right)_Y - \left(ae_j, e_k\right)_Y\Big). \end{aligned}$$

Equation (4.100b), with the test function χ_j^0, implies

$$\begin{aligned} \left(a\nabla_y\chi_j^0, \nabla_y\theta_{kl}^0\right)_Y - \left(ae_j, e_k\chi_l^0\right)_Y =_S{}& -\left(ae_k\chi_l^0, \nabla_y\chi_j^0\right)_Y + \left(a(\nabla_y\chi_l^0 + e_l) - a^0e_l, e_k\chi_j^0\right)_Y \\ &- \left(ae_j, e_k\chi_l^0\right)_Y =_S 0. \end{aligned}$$

Using (4.100a) with the test function χ_j^0 implies

$$\left(a\nabla_y\chi_j^0, \nabla_y\chi_k^0\right)_Y - \left(ae_j, e_k\right)_Y = -\left(a(\nabla_y\chi_j^0 + e_j), e_k\right)_Y = -|Y|a^0_{jk}.$$

Combining the last three equalities, we obtain the first equality in (4.103).

The first implication of (4.103) is that the fourth order tensor in the right hand side of (4.102) does not depend on the normalization of χ_i. This was expected as this tensor constrains the pair b^2, a^2 which characterizes the dispersion.

Further, thanks to (4.103), we follow the process given in (4.92), with the matrix $R^\delta_{ij} = \mu_i\mu_j$, and define the pair

$$\begin{aligned} a^2_{ijkl}(\mu) &= S^{2,2}_{ij,kl}\Big\{\left(a_{jk}\chi_l^\mu, \chi_i^\mu\right)_Y - \left(a\nabla_y\theta_{ji}^\mu, \nabla_y\theta_{kl}^\mu\right)_Y\Big\}, \\ b^2_{ij}(\mu) &= \left(\chi_i^\mu, \chi_j^\mu\right)_Y. \end{aligned} \tag{4.104}$$

Indeed, (4.103) ensure that

$$\begin{aligned} a^2_{ijkl}(\mu) &=_S \left(a_{jk}\chi_l^0, \chi_i^0\right)_Y - \left(a\nabla_y\theta_{ji}^0, \nabla_y\theta_{kl}^0\right)_Y + a^0_{jk}\mu_i\mu_l, \\ b^2_{ij}(\mu) &=_S \left(\chi_i^0, \chi_i^0\right)_Y + \mu_i\mu_l, \end{aligned}$$

and if the tensor $a^0_{jk}\mu_i\mu_l$ is "sufficiently large", the pair $a^2(\mu), b^2(\mu)$ defines an effective equation in the family. Furthermore, the pairs $a^2(\mu), b^2(\mu)$ defined by (4.104) are parametrized by the normalization parameters $(\mu_1, \ldots, \mu_d)$, $\mu_i = \langle\chi_i^\mu\rangle_Y$. Hence, the pairs $a^2(\mu), b^2(\mu)$ defines a subset of the family that is parametrized by the normalization of the correctors $\chi_1, \ldots, \chi_d$. However, we verify that this subset might be empty. Indeed, the tensor $a^0_{jk}\mu_i\mu_l$ is only positive semidefinite:

$$a^0_{jk}\mu_i\mu_l\xi_{ij}\xi_{kl} = a^0(\xi\mu)\cdot(\xi\mu) \geq 0,$$

and, for example, we have $\xi\mu = 0$ for $\xi = \mathrm{diag}(v)$ with $v \perp \mu$. Consequently, there may not exist μ such that $a^0_{jk}\mu_i\mu_l$ is sufficiently large for $a^2(\mu)$ to be positive semidefinite.

4.4 Numerical experiments

In this section, we illustrate the theoretical results obtained in this chapter through various numerical experiments. In particular, we present several examples that confirm the result of Theorem 4.2.4, which states that the family of effective equation captures the dispersion effects of u^ε (Theorem 4.2.9 if the tensor is only bounded).

First, we consider two examples in the one-dimensional case. One with a smooth tensor and one with a discontinuous tensor. Second, we consider two and three dimensional examples in layered materials. In a last experiment, we show that the dispersion effects are not due to the incompatibility of the initial condition with the tensor discussed in Section 3.2.2.

4.4.1 One-dimensional example: smooth tensor

For the first example, let us come back to the example of Section 4.1. Let the reference cell be $Y = (-1/2, 1/2)$ and consider the smooth oscillating tensor :

$$a\big(\tfrac{x}{\varepsilon}\big) = \sqrt{2} - \cos\big(2\pi\tfrac{x}{\varepsilon}\big),$$

where $\varepsilon = 1/20$. We find that the (class of) the first corrector, the homogenized tensor, and $\big\langle(\chi - \langle\chi\rangle)^2\big\rangle_Y$ are given by

$$\chi(y) = \frac{1}{\pi}\mathrm{atan}\big((1+\sqrt{2})\tan(\pi y)\big) - y + \langle\chi\rangle, \qquad a^0 = 1, \qquad \big\langle(\chi - \langle\chi\rangle)^2\big\rangle_Y \simeq 0.00909633,$$

where $\langle\chi\rangle$ is any real number. In Section 4.3.1, we proved that the family of effective equations is composed of the solutions $\tilde{u}_{\langle\chi\rangle}$ of (4.56), where b^2 and a^2 are defined as

$$b^2 = \big\langle(\chi - \langle\chi\rangle)^2\big\rangle_Y + \langle\chi\rangle^2, \qquad a^2 = a^0\langle\chi\rangle^2,$$

for any given value of the parameter $\langle\chi\rangle$. Let us illustrate this result with an example. We set $\Omega = (-402, 402)$, which is large enough for the waves never to reach the boundary. We consider the wave equation (4.26), where the initial conditions are given as $g^0(x) = e^{-10x^2}$, $g^1(x) = 0$ and the source $f = 0$. We approximate u^ε with a spectral method on a grid of size $\Delta x = \varepsilon/20$ (see Section 2.3). The leap frog method is used for the time integration of the obtained second order ODE (see Section A.5). To approximate $\tilde{u}_{\langle\chi\rangle}$ and the homogenized solution u^0, we use the Fourier method, defined in Section 2.4, on a grid of size $\Delta x = \varepsilon/8$. In Figure 4.3, we display u^ε, u^0, and $\tilde{u}_{\langle\chi\rangle}$ for 20 values of the parameter $\langle\chi\rangle \in [0, 0.38]$, at the time $t = \varepsilon^{-2} = 400$. We observe that the dispersion visible in the macroscopic behavior of u^ε is not captured by u^0. On the contrary, all the elements of the family $\mathcal{E}$ describe well this dispersive feature. This example corroborates the result of Theorem 4.2.4 and the derivation of the family in Section 4.3.1.

4.4.2 One-dimensional example: discontinuous tensor

In the previous example, the tensor was smooth. Let us now consider an example where it is bounded but not smooth. Let the reference cell be $Y = (0, 1)$ and consider the Y-periodic discontinuous tensor

$$a(y) = a_2\mathbb{1}_{[0,1/4[}\big(\{y\}_Y\big) + a_1\mathbb{1}_{[1/4,3/4[}\big(\{y\}_Y\big) + a_2\mathbb{1}_{[3/4,1[}\big(\{y\}_Y\big),$$

where $a_1, a_2 > 0$, $\mathbb{1}_X$ denotes the indicator function of the set X, and $\{y\}_Y = y - \lfloor y \rfloor$. We compute the first corrector:

$$\chi(y) = \tfrac{1}{a_1}y\mathbb{1}_{]0,1/4[}(y) + \Big(\tfrac{1}{4a_1} + \tfrac{1}{a_2}(y - 1/4)\Big)\mathbb{1}_{[1/4,3/4[}(y) + \Big(\tfrac{1}{4a_1} + \tfrac{1}{2a_2} + \tfrac{1}{a_1}(y - 3/4)\Big)\mathbb{1}_{[3/4,1[}(y) + \langle\chi\rangle,$$

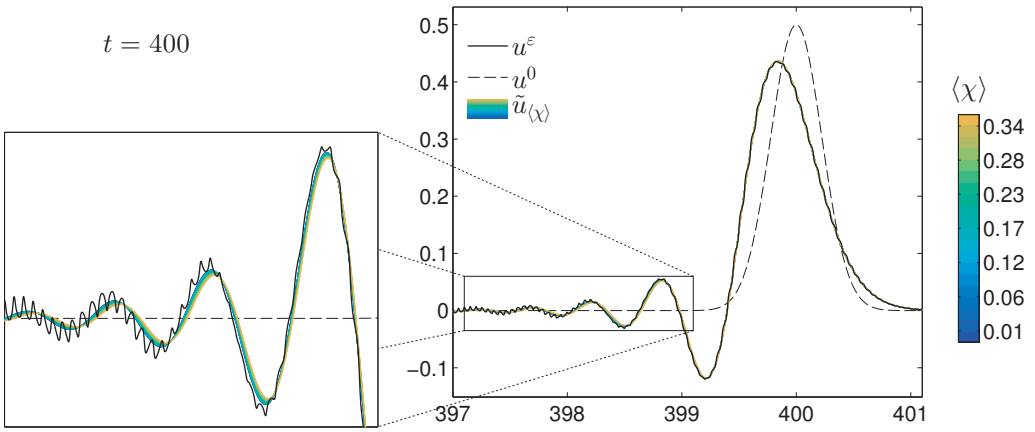

Figure 4.3: Comparison of the wave u^ε (smooth tensor) with the homogenized solution u^0 and effective solutions $\tilde{u}_{\langle\chi\rangle}$ from the family $\mathcal{E}$, for several values of the parameter $\langle\chi\rangle \in [0, 0.38]$ at $t = \varepsilon^{-2} = 400$ and zoom on $x \in [397.1, 399.1]$.

where $\langle\chi\rangle$ is any real number. We then fix $a_2 = 3.255$ and $a_1 = \left(2 - \frac{1}{a_2}\right)^{-1}$ so that $a^0 = 1$ and $\left\langle (\chi - \langle\chi\rangle)^2 \right\rangle_Y \simeq 0.00999885$. As in the example in the previous section, we consider the family of effective equations $\mathcal{E}$ defined in Section 4.3.1. We consider the same data: $g^0(x) = e^{-10x^2}$, $g^1(x) = 0$, and $f = 0$. In order to approximate u^ε, we construct the following mesh. The subintervals where a^ε is constant are discretized into Chebyshev grids of 32 nodes (i.e. the nodes are distributed on the interval as the Chebyshev nodes in $(-1, 1)$). Hence, the mesh has a node at each discontinuity of a^ε and a high concentration of nodes in its neighborhood. The wave u^ε is then approximated on this mesh with $\mathcal{P}_1$-FEM. The effective solutions $\tilde{u}_{\langle\chi\rangle}$ and the homogenized solution u^0 are approximated using the method defined in Section 2.4 on a grid of size $\Delta x = \varepsilon/8$. As the method for the approximation of u^ε is costly, we consider the small domain $\Omega = (-6, 6)$ (which verifies $|\Omega|/\varepsilon|Y| \in \mathbb{N}$). In Figure 4.4, we display u^ε, u^0, and $\tilde{u}_{\langle\chi\rangle}$ for different values of the parameter $\langle\chi\rangle \in [0, 0.38]$ at time $t = \varepsilon^{-2} = 400$. We observe that the dispersive behavior of u^ε is not captured by u^0, while it is well described by all the elements of the family $\mathcal{E}$. This example corroborates the result of Theorem 4.2.9 ensuring that even if the tensor is only bounded, the family of effective equations is still valid.

4.4.3 Two-dimensional example in small and pseudoinfinite domains

We now consider a two-dimensional example. First, we compute effective tensors using Algorithm 4.3.7. Then, we compare u^ε with the corresponding effective solution and the homogenized solution, on a large time interval $\mathcal{O}(\varepsilon^{-2})$ in a small domain. Finally, we display the effective solution and the homogenized solution in a pseudoinfinite domain. In particular, we provide visualizations of the dispersion phenomenon in two dimensions. Note that a Matlab implementation of the long time homogenization method for this example is provided in Appendix A.4.7.

Let the reference cell be $Y = (\text{-}1/2, 1/2)$ and consider the Y-periodic diagonal tensor given by

$$a(y) = \begin{pmatrix} \tilde{a}(y_2) & 0 \\ 0 & \tilde{a}(y_2) \end{pmatrix} = \begin{pmatrix} 1 - 0.5\cos(2\pi y_2) & 0 \\ 0 & 1 - 0.5\cos(2\pi y_2) \end{pmatrix}. \tag{4.105}$$

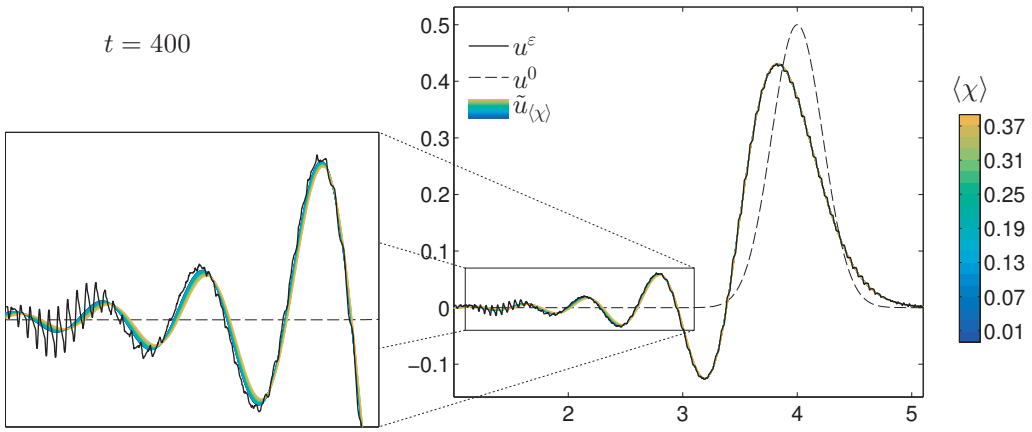

Figure 4.4: Comparison of the wave u^ε (discontinuous tensor (4.4.2)) with the homogenized solution u^0 and effective solutions $\tilde{u}_{\langle\chi\rangle}$ from the family $\mathcal{E}$ for several values of the parameter $\langle\chi\rangle \in [0, 0.38]$ at $t = \varepsilon^{-2} = 400$.

For $\varepsilon > 0$, the oscillatory tensor $a\left(\frac{x}{\varepsilon}\right)$ describes the layered material displayed in Figure 4.5. It is well known that the corresponding homogenized tensor is anisotropic and given by (see [24, 66, 37])

$$a^0 = \begin{pmatrix} \int_{-1/2}^{1/2} \tilde{a}(y_2)\,\mathrm{d}y_2 & 0 \\ 0 & \left(\int_{-1/2}^{1/2} (\tilde{a}(y_2))^{-1}\,\mathrm{d}y_2\right)^{-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{3}/2 \end{pmatrix}, \tag{4.106}$$

where $\sqrt{a^0_{ii}}$ is the homogenized wave speed in the i-th direction. Theorem 4.2.4 ensures that at timescales $\mathcal{O}(\varepsilon^{-2})$, u^ε is well described by the effective equations in the family $\mathcal{E}$ (Definition 4.2.6). To obtain an effective solution, we first compute the tensors b^2 and a^2 using Algorithm 4.3.7, and then approximate the solution of the corresponding effective equation using the Fourier method presented in Section 2.4.

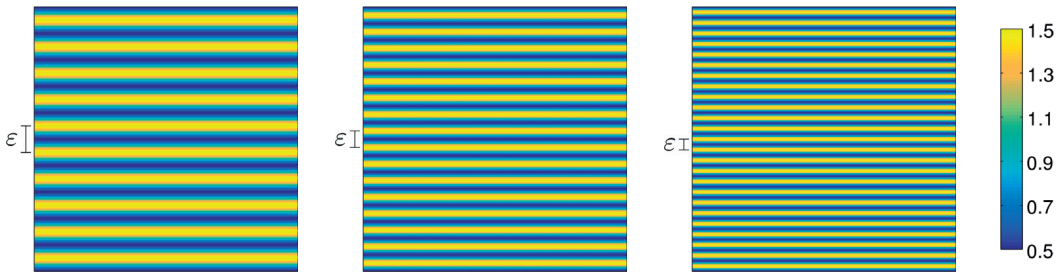

Figure 4.5: Tensor $a^\varepsilon(x) = a\left(\frac{x}{\varepsilon}\right)$ where a is defined in (4.105) displayed in $(0,1)^2$ for, respectively from left to right, $\varepsilon = 1/10$, $1/16$, and $1/25$.

Computation of the tensors of effective equations

We use Algorithm 4.3.7 to compute the tensors a^2, b^2 of an effective equation in the family $\mathcal{E}$ (Definition 4.2.6). Although an analytic expression for the first correctors χ_1, χ_2 is available for

the tensor (4.105), in order to test the numerical procedure, we approximate the cell functions $\chi_1, \chi_2, \theta_{11}, \theta_{12}, \theta_{22}$. To do so, we use a P1 finite element method on a uniform mesh of Y with 1024 points in both directions. We verify that the corresponding approximation of a^0 is accurate. Then, we compute the 6 distinct entries of the tensor $\check{a}^2$ and find

$$\begin{array}{lll} \check{a}^2_{1111} = -0.00339360, & \check{a}^2_{2222} = 0, & \check{a}^2_{1212} = 0.00086375, \\ \check{a}^2_{1122} = 0.00339360, & \check{a}^2_{1112} = 0, & \check{a}^2_{2212} = 0. \end{array}$$

From $\check{a}^2$, we construct the 3×3 symmetric matrix $A^2 = M(\check{a}^2)$ (see Section 4.3.3). The eigenvalues of A^2 are computed as

$$\operatorname{spec}(A^2) = \{-0.0054909,\ 0.0020973,\ 0.0034550\},$$

so that A^2 is not positive semidefinite. In order to compute the non-negative tensor a^2, we build the matrix $A^0 = M\big(S^{2,2}_{ij,kl}\{a^0_{jk} I_{il}\}\big)$ and obtain

$$A^0 = \begin{pmatrix} a^0_{11} & 0 & \frac{1}{2}(a^0_{12}+a^0_{21}) \\ & a^0_{22} & \frac{1}{2}(a^0_{12}+a^0_{21}) \\ & & a^0_{11}+a^0_{22} \end{pmatrix}, \qquad \operatorname{spec}(A^0) = \{a^0_{11},\ a^0_{22},\ a^0_{11}+a^0_{22}\}.$$

We then compute $\delta^* = \left\{0, -\frac{\lambda_{\min}(A^2)}{\lambda_{\min}(A^0)}\right\}_+ = 0.006340411$ and the tensors b^2, a^2 are

$$\begin{array}{ll} b^2_{11} = \langle \chi^2_1 \rangle_Y + \delta^* = 0.00634041, & a^2_{1111} = \check{a}^2_{1111} + \delta^* a^0_{11} = 0.00294681, \\ b^2_{22} = \langle \chi^2_2 \rangle_Y + \delta^* = 0.01004512, & a^2_{2222} = \check{a}^2_{2222} + \delta^* a^0_{22} = 0.00549097, \\ b^2_{12} = 0, & a^2_{1212} = \check{a}^2_{1212} + \frac{1}{4}\delta^*(a^0_{11}+a^0_{22}) = 0.0038215948, \\ & a^2_{1122} = \check{a}^2_{1122} = 0.0033935973, \\ & a^2_{1112} = \check{a}^2_{1112} + \frac{1}{4}\delta^*(a^0_{12}+a^0_{21}) = 0, \\ & a^2_{2212} = \check{a}^2_{1222} + \frac{1}{4}\delta^*(a^0_{12}+a^0_{21}) = 0. \end{array}$$

We recall that other effective equations can be obtained by defining the tensors as in (4.92), where $R^\delta \in \mathrm{Sym}^2(\mathbb{R}^d)$ is a positive definite matrix with sufficiently large eigenvalues. In order to illustrate this, we let $r = (r_1, r_2) \in \mathbb{R}^2$, $R^\delta = \operatorname{diag}(r_1, r_2)$, and denote a^2_r, b^2_r as defined in (4.92), where the subscript specifies the dependence in r. For several values of $r \in \mathbb{R}^2$, we compute the minimal eigenvalue $\lambda_{\min}(r)$ of $M(a^2_r)$. In Figure 4.6, we plot $r = (r_1, r_2)$ with a red square (■) if $\lambda_{\min}(r) < 0$ and a green square (■) if $\lambda_{\min}(r) \geq 0$. Hence, each green square corresponds to a different well-posed effective equation in the family and we call the corresponding r *valid*. We observe that there is a distinct frontier between valid and invalid values of r. The black square is (δ^*, δ^*), where δ^* is defined in Algorithm 4.3.7 (see Lemma 4.3.2). As expected, (δ^*, δ^*) lies in the domain of valid values. The subset of the diagonal in the valid values $\{(\delta, \delta) : \delta \geq \delta^*\}$ corresponds to the effective equations provided by Lemma 4.3.2. For a future experiment in this section, let us introduce the following notation:

$$\begin{array}{l} \{\tilde{u}_\delta\}_{\delta \geq \delta^*} \text{ is the solution of the effective equation (4.56) with } a^2 = a^2_\delta,\ b^2 = b^2_\delta, \\ \bar{u} = \tilde{u}_{\delta^*} \text{ is the effective solution given by Algorithm 4.3.7.} \end{array} \tag{4.107}$$

Example in a small domain

Let us fix $\varepsilon = 1/10$ and consider equation (4.26), where the initial conditions and the source term are given as

$$g^0(x) = e^{-20(x_1^2+x_2^2)}, \qquad g^1(x) = 0, \qquad f(t,x) = 0, \tag{4.108}$$

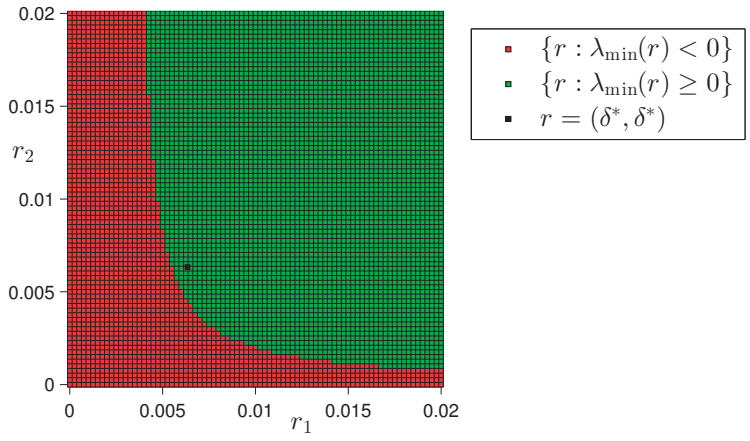

Figure 4.6: Sorting of the minimal eigenvalues of $M(a_r^2)$, where a_r^2 is defined in (4.92) with $R^\delta = \mathrm{diag}(r_1, r_2)$. Each green square corresponds to an effective equation in the family $\mathcal{E}$. The black square is (δ^*, δ^*), where δ^* is computed in Algorithm 4.3.7. The diagonal $\{(\delta, \delta), \delta \geq \delta^*\}$ corresponds to the effective solutions provided by Lemma 4.3.2 and denoted $\{\tilde{u}_\delta\}_{\delta \geq \delta^*}$ in the text.

and the periodic domain is $\Omega = (-2, 2)^2$. Even though (4.26) does not have a physical interest, on this small domain we are able to approximate u^ε, the solution of (4.26) (g^0 must be replaced with $g^0 - \langle g^0 \rangle_\Omega$ to fit the setting of (4.26)). To do so, we use a spectral method on a uniform grid of Ω of size $h = \varepsilon/10$ (see Section 2.3). The time integration of the obtained second order ordinary differential equation is done with the leap frog scheme with time step $\Delta t = h/100$ (see Section A.5). The solutions u^0 and $\bar{u}$ are approximated using the Fourier method, defined in Section 2.4, on the same grid as u^ε. On Figure 4.7, we display u^ε, u^0, and $\bar{u}$ at $t = \varepsilon^{-2} = 100$. We first observe that, as expected, the behavior of u^ε is not well described by the homogenized solution u^0. As ensured by Theorem 4.2.4, the effective equation $\bar{u}$ does describe well u^ε in the $\mathrm{L}^\infty(0, T^\varepsilon; \mathrm{L}^2(\Omega))$ norm. Next, we compute the normalized errors

$$\mathrm{err}(v)(t) = \|(u^\varepsilon - v)(t)\|_{\mathrm{L}^2(\Omega)} / \|u^\varepsilon(t)\|_{\mathrm{L}^2(\Omega)}, \quad v \in \{u^0, \bar{u}\},$$

on the time interval $[0, 100]$. The result is displayed in Figure 4.8. We observe that the homogenized solution quickly drift away from the fine scale solution u^ε. As we know, this is due to the dispersion effects developed by u^ε. On the contrary, we see that for up to $t = \varepsilon^{-2} = 100$, the error $u^\varepsilon - \bar{u}$ is small, as predicted by Theorem 4.2.4.

Example in a pseudoinfinite domain

Let us now consider the wave equation with the data (4.108) in a pseudoinfinite medium. We thus have to find a domain that is large enough for the wave not to reach the boundary. As the homogenized tensor (4.106) is diagonal, we know the form of the homogenized solution u^0: the initial pulse g^0, centered at the origin, spreads in all directions with speeds $\sqrt{a^0_{11}}$ along the x axis and $\sqrt{a^0_{22}}$ along the y axis. We thus set

$$\Omega = (-L_1, L_1) \times (-L_2, L_2), \qquad L_i = \left\lfloor \sqrt{a^0_{ii}} t \right\rfloor + 2.$$

We compute u^0 and $\bar{u}$ (see (4.107)) with the Fourier method (see Section 2.4), on a grid of size $\varepsilon/16$. In Figure 4.9, we display the global form of $\bar{u}$ at $t = 100$ and in the zooms we can observe the dispersion effects. Note that although $a\left(\frac{x}{\varepsilon}\right)$ oscillates only in the y direction, the dispersion

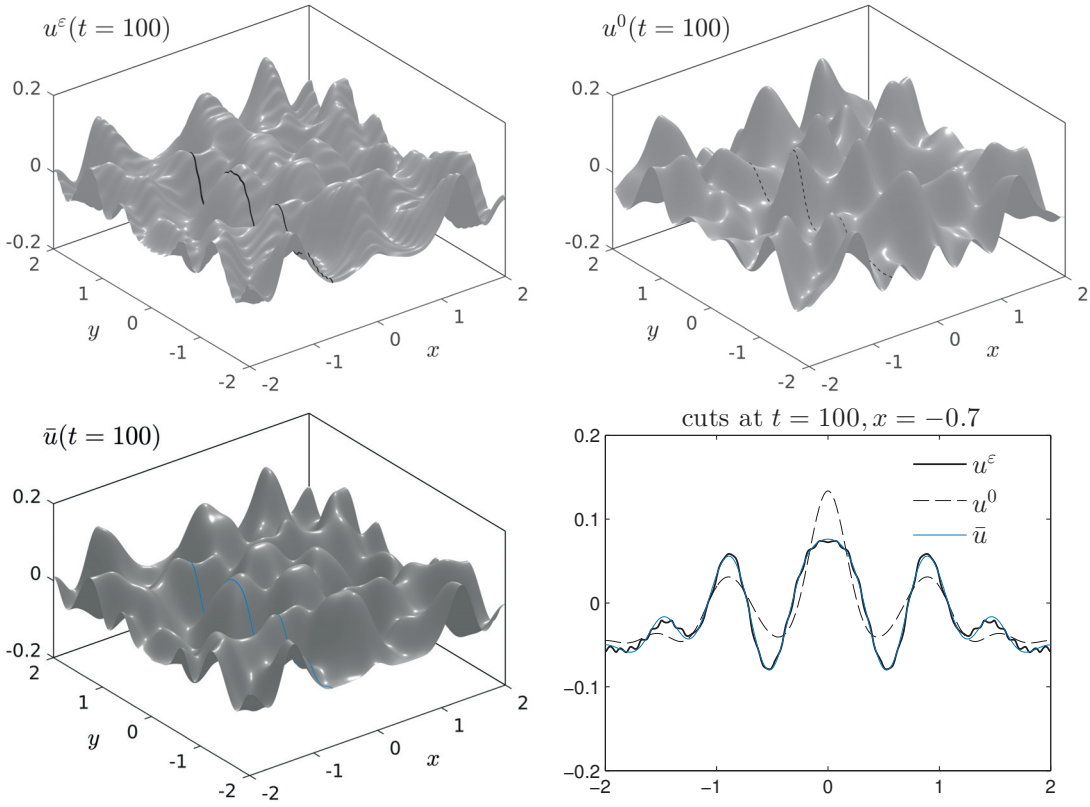

Figure 4.7: Comparison on Ω between u^ε (top-left), the homogenized solution u^0 (top-right) and the effective solution $\bar{u}$ (bottom-left) and cuts at $x = -0.7$ (bottom-right) at $t = \varepsilon^{-2} = 100$.

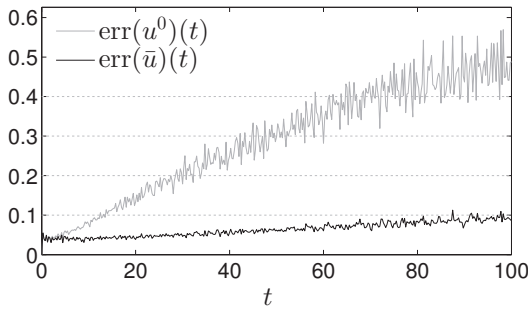

Figure 4.8: Plot of the time evolution of the normalized $\mathrm{L}^2(\Omega)$ errors $u^\varepsilon - u^0$ and $u^\varepsilon - \bar{u}$.

is as strong in the x direction as in the y direction. In the top-left plot of Figure 4.10, we can see a closer view of the dispersion developed in $\bar{u}$ at $t = 100$. Furthermore, the same view of u^0 is displayed in the top-right plot of Figure 4.10 and we see that there is no dispersion after the main pulse. In the bottom plot of Figure 4.10, we can compare cuts at $y = 0$ of $\bar{u}$, $\{\tilde{u}_\delta\}_\delta$ for several values of $\delta \in [\delta^*, 11\delta^*]$ (see (4.107)) and u^0. We see that the effective solutions $\{\tilde{u}_\delta\}_\delta$ and $\bar{u}$ have almost the same dispersive behavior. As Theorem 4.2.4 ensures that $\bar{u}$ and $\tilde{u}_\delta$ approximate well u^ε, we conclude that u^0 is a poor approximation of u^ε at $t = 100$.

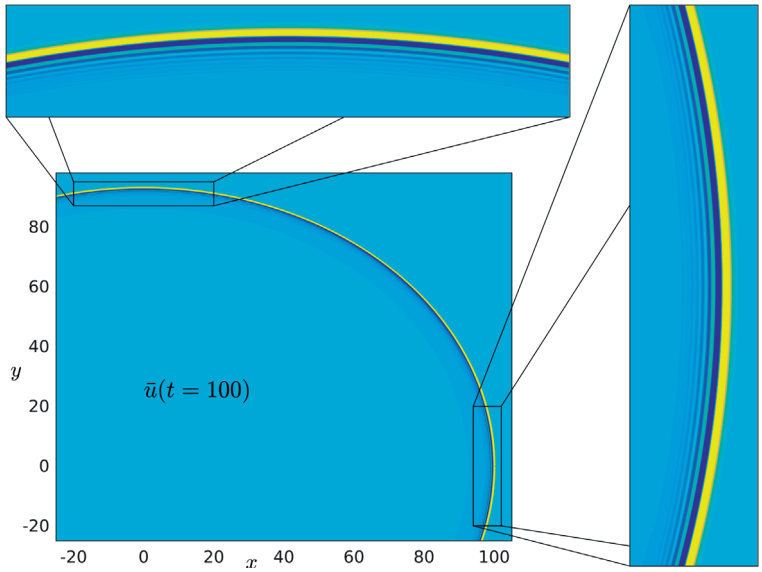

Figure 4.9: Global view of $\bar{u}$ at $t = 100$ and zooms on the subdomains $(94, 102) \times (-20, 20)$ and $(-20, 20) \times (87, 95)$.

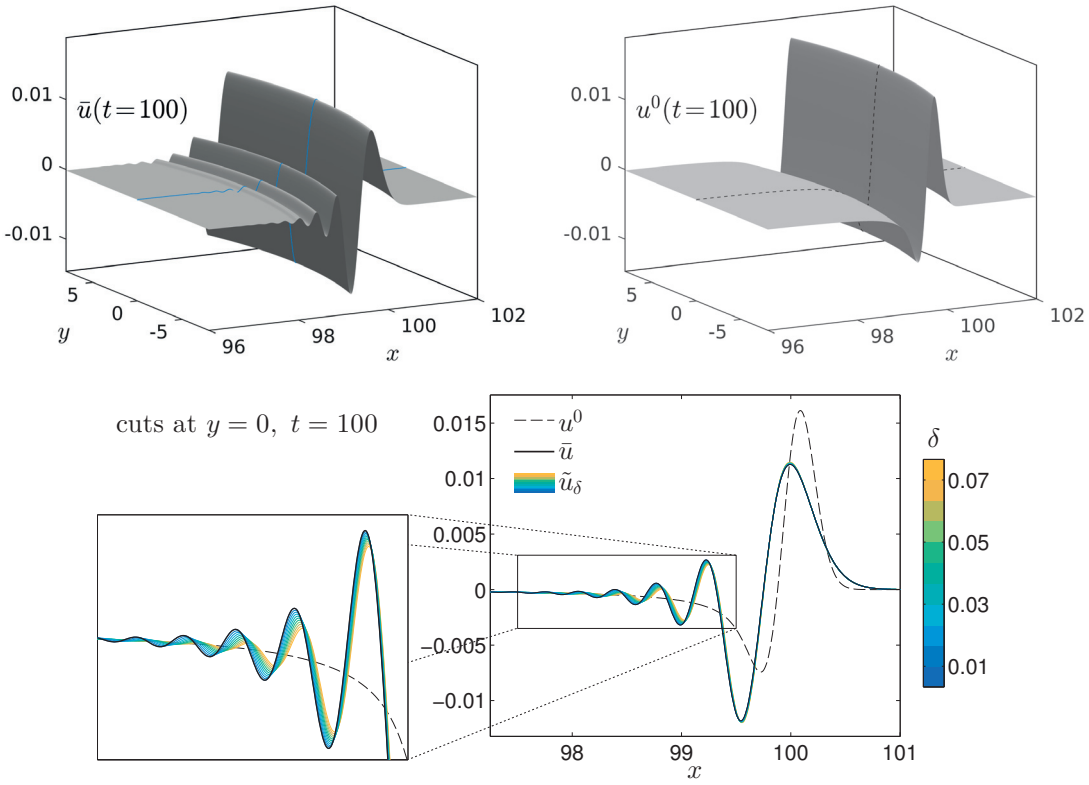

Figure 4.10: Top: 3d views of $\bar{u}$ (top-left) and u^0 (top-right) at $t = 100$ for $(x, y) \in (96, 102) \times (-8, 8)$ Bottom: 1d cuts $x \in [97.25, 101]$, $y = 0$, $t = 100$ of u^0 and the effective solutions $\bar{u}$, $\tilde{u}_\delta$ for several values of $\delta \in [\delta^*, 11\delta^*]$ (see (4.107)).

4.4.4 Three-dimensional example in a pseudoinfinite domain

Let us consider a three dimensional example. In three dimensions, approximating u^ε is not feasible. Indeed, even in small domains the computational cost of a spectral method (or FEM) is extremely large. However, using the homogenization method obtained in this chapter, we can visualize the description of the dispersion developed by u^ε at long times.

Let the reference cell be $Y = (-1/2, 1/2)^3$ and consider the Y-periodic diagonal tensor given by

$$a(y) = \tilde{a}(y_2) I_3 = \big(1 - 0.5\cos(2\pi y_3)\big) I_3, \tag{4.109}$$

where I_3 is the 3×3 identity matrix. For $\varepsilon > 0$, the oscillatory tensor $a\big(\frac{x}{\varepsilon}\big)$ describes the layered material displayed in Figure 4.11. We compute the effective tensors a^0 and a^2, b^2 corresponding to $a(y)$ using Algorithm 4.3.7. In particular, we verify that the homogenized tensor is diagonal and given by

$$a^0 = \operatorname{diag}(1, 1, \sqrt{3}/2).$$

We fix $\varepsilon = 1/5$ and consider the model problem given by the data

$$g^0(x) = e^{50|x|^2}, \quad g^1 = 0, \quad f = 0.$$

Using the Fourier method (see Section 2.4), we approximate the homogenized solution and the effective solution $\bar{u}$ at $t = \varepsilon^{-2} = 25$ in the pseudoinfinite domain defined as

$$\Omega = (-L_1, L_1) \times (-L_2, L_2) \times (-L_3, L_3), \qquad L_i = \left\lfloor \sqrt{a^0_{ii}}\, t \right\rfloor + 1.$$

Visualizations of the computed u^0 and $\bar{u}$ are displayed in Figures 4.12 and 4.13. Both solutions are waves spreading away from the origin. However, comparing the frontal waves, we see that dispersive effects are clearly visible in $\bar{u}$, while no such behavior is to be seen in u^0. Theorem 4.2.4 ensures that $\bar{u}$ describes well u^ε.

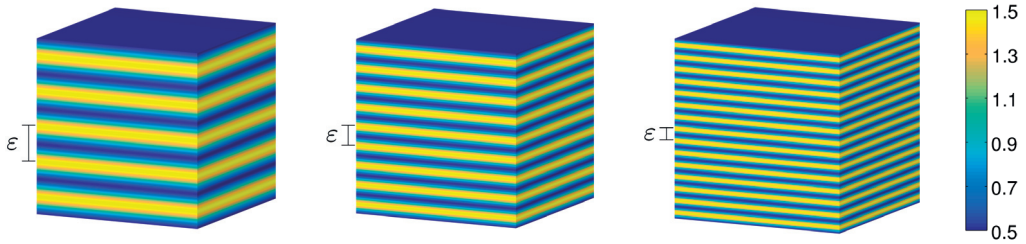

Figure 4.11: Tensor $a^\varepsilon(x) = a\big(\frac{x}{\varepsilon}\big)$ where a is defined in (4.109) displayed in $(0,1)^3$ for, respectively from left to right, $\varepsilon = 1/5$, $1/9$, and $1/15$.

4.4.5 Long time effects for a prepared initial condition

In this section, we illustrate that the long time dispersive phenomenon is not a consequence of the incompatibility between the tensor $a\big(\frac{x}{\varepsilon}\big)$ and the initial position g^0, discussed in Section 3.2.2. To do so, we show that the solution of (3.40) $\bar{u}^\varepsilon$, which has a prepared initial condition (see Section 3.3.3), still develops dispersion.

Consider the settings of Section 4.4.1:

$$Y = (-1/2, 1/2), \quad a\big(\tfrac{x}{\varepsilon}\big) = \sqrt{2} - \cos\big(2\pi \tfrac{x}{\varepsilon}\big), \quad \varepsilon = 1/20, \quad \Omega = (-402, 402),$$
$$g^0(x) = e^{-10x^2}, \quad g^1(x) = 0, \quad f = 0.$$

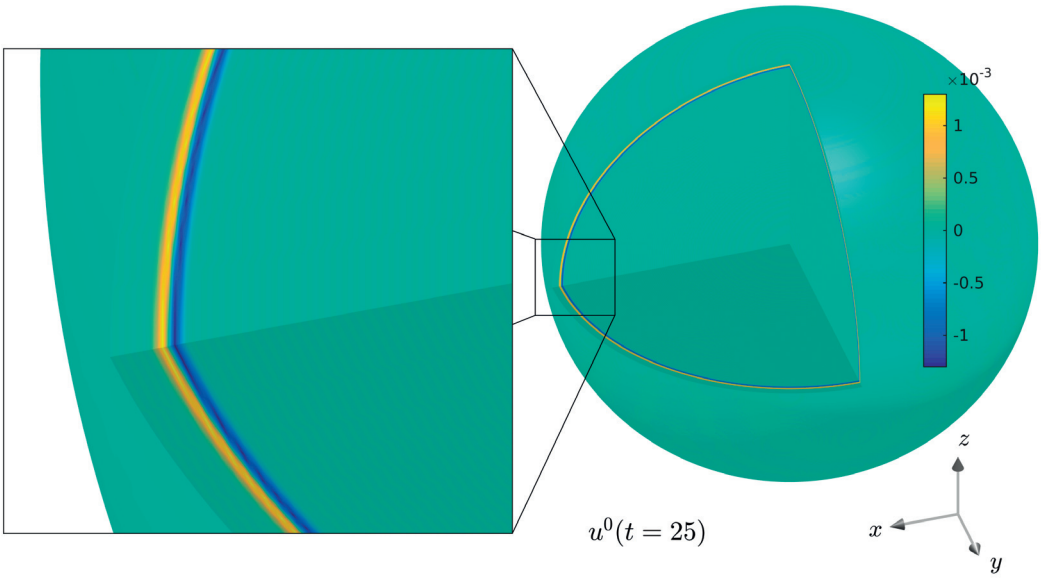

Figure 4.12: Visualization of the homogenized solution u^0 for the layered material of Figure 4.11 with $\varepsilon = 1/5$ at time $t = 25$.

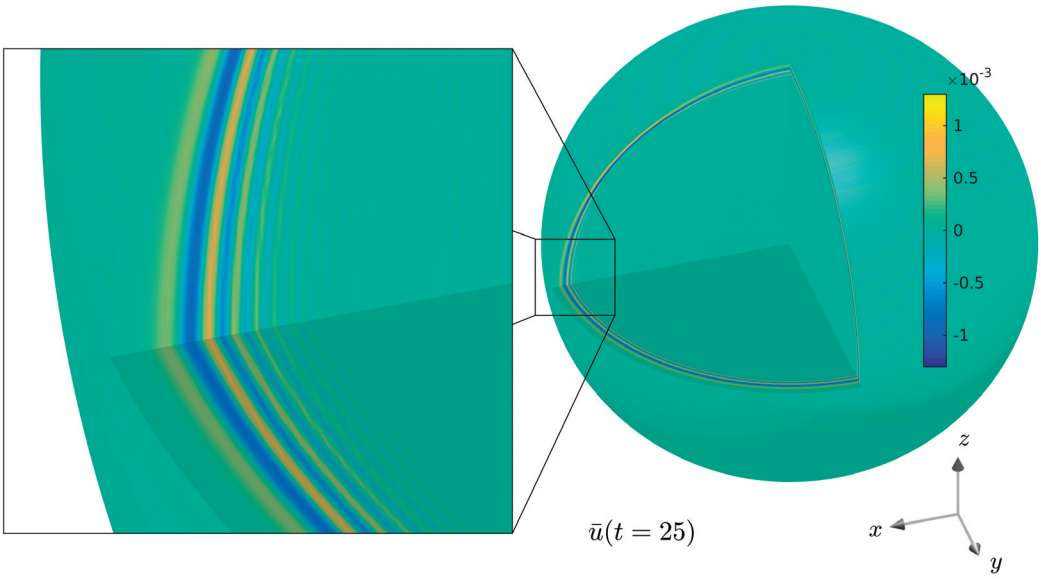

Figure 4.13: Visualization of the effective solution $\bar{u}$ for the layered material of Figure 4.11 with $\varepsilon = 1/5$ at time $t = 25$.

As we have an explicit formula for χ, we can compute the initial position of $\bar{u}^\varepsilon$ as $\bar{g}^0(x) = g^0(x) + \varepsilon\chi\left(\frac{x}{\varepsilon}\right)\partial_x g^0(x)$. The function $\bar{u}^\varepsilon$ is approximated with a spectral method on a grid of size $\Delta x = \varepsilon/20$ (see Section 2.3). The leap frog method is used for the time integration of the obtained second order ODE (see Section A.5). The functions $\tilde{u}_{\langle\chi\rangle}$ and u^0, are approximated (with

initial position g^0) using the Fourier method (Section 2.4) on a grid of size $\Delta x = \varepsilon/8$. In Figure 4.14, we display $\bar{u}^\varepsilon$, u^0, and $\tilde{u}_{\langle\chi\rangle}$ for 20 values of the parameter $\langle\chi\rangle \in [0, 0.38]$, at $t = \varepsilon^{-2} = 400$. We observe that even though its initial position is compatible with the tensor $a(\left(\frac{x}{\varepsilon}\right))$, $\bar{u}^\varepsilon$ have a similar dispersive behavior as u^ε. The dispersion is well described by the element of the family of effective equations $\tilde{u}_{\langle\chi\rangle}$ but not by u^0.

In Section 3.3.3, thanks to the preparation of the initial condition, we proved that the gradient of $\bar{u}^\varepsilon$ could be approximated by a correction of u^0 (see Theorem 3.3.4). Similarly, under sufficient regularity of the data, the adaptation $\mathcal{B}^\varepsilon\tilde{u}$, defined in (4.51), satisfies

$$\|\nabla_x \bar{u}^\varepsilon - \nabla_x \mathcal{B}^\varepsilon \tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} \leq C\varepsilon.$$

To prove it, we apply the standard energy estimate for the wave equation to the function $\boldsymbol{\eta} = \bar{u}^\varepsilon - \mathcal{B}^\varepsilon\tilde{u}$ (see lemma 4.2.8).

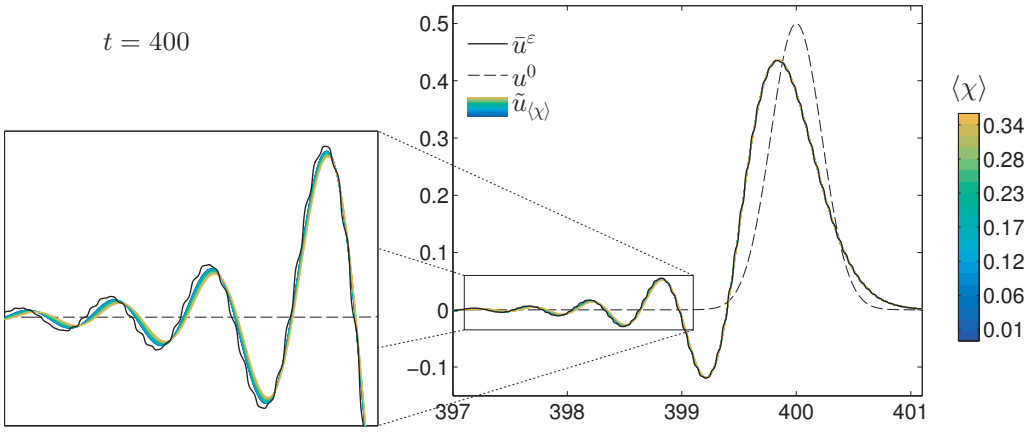

Figure 4.14: Comparison of $\bar{u}^\varepsilon$ with the homogenized solution u^0 and effective solutions $\tilde{u}_{\langle\chi\rangle}$ from the family $\mathcal{E}$ for several values of the parameter $\langle\chi\rangle \in [0, 0.38]$ at $t = \varepsilon^{-2} = 400$.

5 Effective models for wave propagation in periodic media for arbitrary timescales

In Chapter 4, we derived a family of effective equations for wave propagation in periodic media at timescales $\mathcal{O}(\varepsilon^{-2})$. In this chapter, we generalize this result to arbitrary timescales. Let $\Omega \subset \mathbb{R}^d$ be an arbitrarily large hypercube, α be a non-negative integer, and let $a^\varepsilon(x) = a\left(\frac{x}{\varepsilon}\right)$ be a symmetric, elliptic and bounded tensor, where $a(y)$ is Y-periodic (see Section 4.2.1). We consider $u^\varepsilon : [0, \varepsilon^{-\alpha}T] \times \Omega \to \mathbb{R}$ the solution of

$$\begin{array}{ll} \partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \left(a\left(\frac{x}{\varepsilon}\right)\nabla_x u^\varepsilon(t,x)\right) = f(t,x) & \text{in } (0, \varepsilon^{-\alpha}T] \times \Omega. \\ x \mapsto u^\varepsilon(t,x) \ \ \Omega\text{-periodic} & \text{in } [0, \varepsilon^{-\alpha}T], \\ u^\varepsilon(0,x) = g^0(x), \quad \partial_t u^\varepsilon(0,x) = g^1(x) & \text{in } \Omega, \end{array} \tag{5.1}$$

where g^0, g^1 are initial conditions and f is a source term (see Section 2.1.1 for the well-posedness of (5.1)).

Effective equations of arbitrary order are not only useful for long time wave propagation. Recall that the family of effective equations, defined in Chapter 4, is valid in the standard multiscale regime. In particular, we assume that the wavelengths of the initial conditions and of the source term are of order $\mathcal{O}(1)$, while the wavelength ε of the tensor is much smaller. However, in regimes where either the initial conditions or the source term have higher frequencies, the error estimate from Theorem 4.2.4 does not guarantee an acceptable bound. In particular, it is not ensured that the effective equations provide accurate approximations of u^ε. In fact, numerical experiments confirm that the provided approximations do not capture the correct macroscopic behavior of u^ε. For example, we verify that the higher the frequency of the initial wave is, the sooner the dispersive effects of u^ε appear. We also observe that for high frequency regimes, u^ε develops additional effects that are not described by the effective equations in the family from Chapter 4. Hence, in that situation, we need higher order effective equations.

The main result of this chapter is the derivation of a family of well-posed effective equations that describe u^ε for arbitrary timescales $\varepsilon^{-\alpha}T$. The family is composed of equations of the form

$$\partial_t^2 \tilde{u} - a^0 \partial^2 \tilde{u} - \sum_{r=1}^{\lfloor \alpha/2 \rfloor} (-1)^r \varepsilon^{2r} \left(a^{2r} \partial^{2r+2} \tilde{u} - b^{2r} \partial^{2r} \partial_t^2 \tilde{u}\right) = f \quad \text{in } (0, \varepsilon^{-\alpha}T] \times \Omega, \tag{5.2}$$

where a^0 is the homogenized tensor and $a^{2r} \in \mathrm{Ten}^{2r+2}(\mathbb{R}^d)$, $b^{2r} \in \mathrm{Ten}^{2r}(\mathbb{R}^d)$ are pairs of non-negative tensors satisfying some given constraints. In (5.2) and in the whole chapter, $\partial^n v$ denotes the tensor of $\mathrm{Ten}^n(\mathbb{R}^d)$ with coordinates $\partial^n_{i_1 \cdot\cdot i_n} v$. Furthermore, for $q \in \mathrm{Ten}^n(\mathbb{R}^d)$, we use the shorthand $q\partial^n v$ to denote the operator $q_{i_1 \cdot\cdot i_n} \partial^n_{i_1 \cdot\cdot i_n} v$.

The derivation of the family (5.2) follows the technique introduced in Chapter 4. First, assuming that the form of the equation is (5.2), we construct an adaptation of $\tilde{u}$. As the timescale is now

of order $\mathcal{O}(\varepsilon^{-\alpha})$, the adaptation is composed of $\alpha+2$ correction terms. After some technical developments, we obtain the cell problems of order 1 to $\alpha+2$. The well-posedness of these cell problems provides constraints on each pair of tensors a^{2r}, b^{2r}. The family is then defined implicitly by the pairs of non-negative, major symmetric tensors a^{2r}, b^{2r} satisfying these constraints.

In addition, we provide an algorithm for the computation of the tensors of effective equations in the family. In particular, we generalize the matrix construction associated to a major symmetric tensor of order four, from Chapter 4, to symmetric tensors of arbitrary even order. Following this process we obtain one possible construction of effective equations in the family.

The fact that no odd correction is needed in (5.2) is a consequence of the unconditional well-posedness of the odd order cell problems. The proof of this remarkable property relies on a technical relation that involves all the previous cell problems (i.e., to prove that the cell problem of order $2r+1$ is well-posed, we need to use the cell problems 1 to $2r$). Note that this result is already known in the context of Bloch wave (see e.g. [42], [23] and the references therein). The second result of this chapter is a new technical relation that enables to reduce significantly the cost of computation of the effective tensors. Namely, while the naive formula to compute a^{2r}, b^{2r} requires to solve the cell problems of order 1 to $2r+1$, we prove that in fact only the cell problems of order 1 to $r+1$ are necessary.

We note that recently, an effective equation of arbitrary order for wave problems was derived in [23] (the result holds in fact for more general tensors: almost periodic, quasiperiodic and random). The derivation of this effective equation significantly differs from our approach as it is based on regularization techniques. In particular, using the so-called Bloch–Taylor expansion of u^ε, an effective equation of the form (for $f=0$)

$$\partial_t^2 u - a^0\partial^2 u - \sum_{r=1}^{\lfloor \alpha/2 \rfloor} \varepsilon^{2r}\bar{a}^{2r}\partial^{2r+2}u - (\mathrm{i}\varepsilon)^{2(\lfloor \frac{\alpha}{2} \rfloor+1)}\gamma \mathrm{Id}\partial^{2(\lfloor \frac{\alpha}{2} \rfloor+1)+2}u = 0 \quad \text{in } (0,\varepsilon^{-\alpha}T]\times\mathbb{R}^d, \tag{5.3}$$

is derived, where $\bar{a}^{2r}$ are effective tensors defined via so-call extended correctors and γ is a regularization parameter (γ is large enough for (5.3) to be well-posed). Furthermore, under low regularity requirements, an error estimate for $u^\varepsilon - u$ is proved. However, no procedure for the computation of γ is available. In fact, numerical tests indicate that the range of acceptable values for γ is narrow. If γ is too small, the equation is ill-posed and if γ is too large, the solution u of (5.3) does not describe u^ε accurately. Hence, the use of (5.3) in practice is problematic. Comparatively, equation (5.2) has the evident advantage of being well-posed without regularization.

As the derivations of (5.2) and (5.3) are done in different frameworks, their comparison is not trivial. As a supplementary result, we prove that the so-called extended correctors, defined in [23], and the correctors obtained in this chapter are the same functions. Furthermore, we derive an exact relation between the tensors $\bar{a}^{2r}$ in (5.3) and the constraint imposed on the pair a^{2r}, b^{2r} in (5.2). This result attests that the two approaches of derivation, via Taylor–Bloch expansion and via asymptotic expansions, lead to the same effective quantities. However, the form of the equation is primordial to obtain well-posed equations without the need of regularization.

The chapter is organized as follows. In Section 5.1, we discuss the effective models for the specific timescales $\mathcal{O}(\varepsilon^{-1})$ and $\mathcal{O}(\varepsilon^{-3})$. In particular, we prove that the homogenized equation is still valid at timescales $\mathcal{O}(\varepsilon^{-1})$ and that the family of effective equations from Chapter 4 is still valid at timescales $\mathcal{O}(\varepsilon^{-3})$. Then, we present the main result of the chapter in Section 5.2. We define the family of effective equations for arbitrary timescales and present the complete derivation. Furthermore, we provide a numerical procedure to compute the effective tensors of arbitrary order. In Section 5.3, we illustrate numerically that in high frequency regimes, the first order effective equations from Chapter 4 do not describe all the macroscopic feature of u^ε. Finally, in Section 5.4, we test our theoretical results in diverse numerical experiments.

5.1 Effective equations for timescales $\mathcal{O}(\varepsilon^0)$ to $\mathcal{O}(\varepsilon^{-3})$

In order to have a better understanding on the derivation of effective equations for arbitrary timescales, we consider the odd timescales $\mathcal{O}(\varepsilon^{-1})$ and $\mathcal{O}(\varepsilon^{-3})$ (these results were given in one-dimension in our paper [13]). First, we prove that the effective model for timescales $\mathcal{O}(\varepsilon^0)$, the homogenized equation, is still valid at timescales $\mathcal{O}(\varepsilon^{-1})$. Second, we prove that the effective models for timescales $\mathcal{O}(\varepsilon^{-2})$, derived in Chapter 4, are still valid at timescales $\mathcal{O}(\varepsilon^{-3})$. These results are particular cases of the general rule that the effective models for even timescales $\mathcal{O}(\varepsilon^{-\alpha})$ (α even) are still valid at timescales $\mathcal{O}(\varepsilon^{-(\alpha+1)})$. This is a consequence of the general result provided in the next section.

5.1.1 The homogenized equation is still valid at timescales $\mathcal{O}(\varepsilon^{-1})$

We prove here that the classical homogenized equation, derived in Section 3.3.2, is still valid at timescales $\mathcal{O}(\varepsilon^{-1})$.

Let $u^0 : [0, \varepsilon^{-1}T] \times \Omega \to \mathbb{R}$ be the solution of the homogenized equation

$$\begin{array}{ll} \partial_t^2 u^0(t,x) - a^0_{ij}\partial^2_{ij} u^0(t,x) = f(t,x) & \text{in } (0,\varepsilon^{-1}T]\times\Omega, \\ x \mapsto u^0(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,\varepsilon^{-1}T], \\ u^0(0,x) = g^0(x), \quad \partial_t u^0(0,x) = g^1(x) & \text{in } \Omega, \end{array} \tag{5.4}$$

where a^0 is the homogenized tensor defined as $a^0_{ij} = \big\langle e_i^T a(\nabla_y \chi_j + e_j)\big\rangle_Y$, where $\chi_j \in \mathrm{W}_{\mathrm{per}}(Y)$ are the first correctors (defined in (4.45)). We prove the following result.

Theorem 5.1.1. *Assume that the Y-periodic tensor satisfies $a(y) \in \mathrm{W}^{2,\infty}(Y)$. Furthermore, assume that the solution u^0 of (5.4), the initial conditions and the right hand side satisfy the regularity*

$$u^0 \in \mathrm{L}^\infty(0,\varepsilon^{-1}T;\mathrm{H}^4(\Omega)), \quad \partial_t u^0 \in \mathrm{L}^\infty(0,\varepsilon^{-1}T;\mathrm{H}^3(\Omega)), \quad \partial_t^2 u^0 \in \mathrm{L}^\infty(0,\varepsilon^{-1}T;\mathrm{H}^2(\Omega)),$$
$$g^0 \in \mathrm{H}^3(\Omega), \quad g^1 \in \mathrm{H}^3(\Omega), \quad f \in \mathrm{L}^2(0,\varepsilon^{-1}T;\mathrm{H}^1(\Omega)).$$

Then the following estimate holds:

$$\begin{aligned} \|u^\varepsilon - u^0\|_{\mathrm{L}^\infty(0,\varepsilon^{-1}T;W)} \le C\varepsilon\Big(& \|g^1\|_{\mathrm{H}^3(\Omega)} + \|g^0\|_{\mathrm{H}^3(\Omega)} + \|f\|_{\mathrm{L}^1(0,\varepsilon^{-1}T;\mathrm{H}^1(\Omega))} \\ & + \textstyle\sum_{k=1}^4 |u^0|_{\mathrm{L}^\infty(0,\varepsilon^{-1}T;\mathrm{H}^k(\Omega))} + |\partial_t^2 u^0|_{\mathrm{L}^\infty(0,\varepsilon^{-1}T;\mathrm{H}^2(\Omega))}\Big), \end{aligned} \tag{5.5}$$

where C depends only on $T, \lambda, \|a\|_{\mathrm{W}^{2,\infty}(Y)}$, and Y, and we recall the definition of the norm (see (A.4))

$$\|w\|_W = \inf_{\substack{w=w_1+w_2 \\ w_1,w_2\in \mathrm{W}_{\mathrm{per}}(\Omega)}} \Big\{ \|w_1\|_{\mathrm{L}^2(\Omega)} + \|\nabla w_2\|_{\mathrm{L}^2(\Omega)} \Big\} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(\Omega).$$

Remark 5.1.2. Referring to Section 4.2.6, a similar result as Theorem 5.1.1 can be proved for a bounded tensor $a(y) \in \mathrm{L}^\infty(Y)$, provided

$$u^0 \in \mathrm{L}^\infty(0,\varepsilon^{-1}T;\mathrm{H}^6(\Omega)), \quad \partial_t u^0 \in \mathrm{L}^\infty(0,\varepsilon^{-1}T;\mathrm{H}^5(\Omega)), \quad \partial_t^2 u^0 \in \mathrm{L}^\infty(0,\varepsilon^{-1}T;\mathrm{H}^4(\Omega)),$$
$$g^0 \in \mathrm{H}^5(\Omega), \quad g^1 \in \mathrm{H}^5(\Omega), \quad f \in \mathrm{L}^2(0,\varepsilon^{-1}T;\mathrm{H}^3(\Omega)).$$

Proof. The proof of Theorem 5.1.1 is analogous to the proof of Theorem 4.2.4. First, we define the adaptation operator $\bar{\boldsymbol{\mathcal{B}}}^\varepsilon : \mathrm{L}^2(0,\varepsilon^{-1}T;\mathrm{H}^3(\Omega)) \to \mathrm{L}^2(0,\varepsilon^{-1}T;\mathcal{W}^*_{\mathrm{per}}(\Omega))$ as

$$\begin{aligned} \big\langle \bar{\boldsymbol{\mathcal{B}}}^\varepsilon v(t), \boldsymbol{w}\big\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}} = & \big(\big[v(t) + \varepsilon\chi_i\partial_i v(t) + \varepsilon^2(\theta_{ij} - \partial_{y_m}\kappa_{mij})\partial^2_{ij} v(t)\big], \boldsymbol{w}\big)_{\mathcal{L}^2} \\ & - \big(\varepsilon^3\kappa_{mij}\partial^2_{ij} v(t), \partial_m \boldsymbol{w}\big)_{\mathrm{L}^2}, + \big\langle \bar{\boldsymbol{\varphi}}(t), \boldsymbol{w}\big\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}}, \end{aligned}$$

where $\chi_i \in \boldsymbol{\chi}_i, \theta_{ij} \in \boldsymbol{\theta}_{ij}, \kappa_{ijk} \in \boldsymbol{\kappa}_{ijk}$ are the correctors defined in (4.45a), (4.45b), and (4.45c) and are evaluated in $y = \frac{x}{\varepsilon}$, and $\bar{\boldsymbol{\varphi}} \in \mathrm{L}^\infty(0, \varepsilon^{-1}T; \mathcal{W}_{\mathrm{per}}(\Omega))$ is the unique solution of

$$\begin{aligned} &(\partial_t^2 + \mathcal{A}^\varepsilon)\bar{\boldsymbol{\varphi}}(t) = -\big[\varepsilon\chi_i\big(\tfrac{\cdot}{\varepsilon}\big)\partial_i f(t)\big] \qquad \text{in } \mathcal{W}^*_{\mathrm{per}}(\Omega) \ \text{ for a.e. } t \in [0, \varepsilon^{-1}T], \\ &\bar{\boldsymbol{\varphi}}(0) = \partial_t\bar{\boldsymbol{\varphi}}(0) = \big[0\big]. \end{aligned} \tag{5.6}$$

Note that $\partial_t\bar{\boldsymbol{\varphi}} \in \mathrm{L}^\infty(0, \varepsilon^{-1}T; \mathcal{L}^2(\Omega))$, $\partial_t^2\bar{\boldsymbol{\varphi}} \in \mathrm{L}^\infty(0, \varepsilon^{-1}T; \mathcal{W}^*_{\mathrm{per}}(\Omega))$, and that the correctors belong to $\mathcal{C}^1_{\mathrm{per}}(\bar{\Omega})$. We thus have $\bar{\boldsymbol{\mathcal{B}}}^\varepsilon u^0(t) \in \mathcal{W}_{\mathrm{per}}(\Omega)$ and $\partial_t^2\bar{\boldsymbol{\mathcal{B}}}^\varepsilon u^0(t) \in \mathcal{W}^*_{\mathrm{per}}(\Omega)$. Let us compute explicitly the remainder $\bar{\boldsymbol{\mathcal{R}}}^\varepsilon u^0$ such that

$$(\partial_t^2 + \mathcal{A}^\varepsilon)\bar{\boldsymbol{\mathcal{B}}}^\varepsilon u^0(t) = \big[f(t)\big] + \bar{\boldsymbol{\mathcal{R}}}^\varepsilon u^0(t) \quad \text{in } \mathcal{W}^*_{\mathrm{per}}(\Omega) \quad \text{for a.e. } t \in [0, \varepsilon^{-1}T]. \tag{5.7}$$

First, thanks to the regularity of u^0, (5.4) gives the following equalities

$$\partial_t^2 u^0 = f + a^0_{ij}\partial^2_{ij}u^0 \text{ in } \mathrm{L}^2_0(\Omega), \qquad \partial_k\partial_t^2 u^0 = \partial_k f + a^0_{ij}\partial^3_{ijk}u^0 \text{ in } \mathrm{L}^2(\Omega).$$

Using these equalities, we find that

$$\begin{aligned} \big\langle \partial_t^2\bar{\boldsymbol{\mathcal{B}}}^\varepsilon u^0, \boldsymbol{w}\big\rangle = &\big(\big[f\big] + \big[a^0_{ij}\partial^2_{ij}u^0 + \varepsilon a^0_{ij}\chi_k\partial^3_{ijk}u^0\big], \boldsymbol{w}\big)_{\mathcal{L}^2} \\ &+ \big\langle \partial_t^2\bar{\boldsymbol{\varphi}}, \boldsymbol{w}\big\rangle + \big(\big[\varepsilon\chi_i\partial_i f\big], \boldsymbol{w}\big)_{\mathcal{L}^2} + \big\langle \bar{\boldsymbol{\mathcal{R}}}^\varepsilon_1 u^0, \boldsymbol{w}\big\rangle, \end{aligned}$$

where $\big\langle \bar{\boldsymbol{\mathcal{R}}}^\varepsilon_1 u^0, \boldsymbol{w}\big\rangle = \big(\big[\varepsilon^2(\theta_{ij} - \partial_{y_m}\kappa_{mij})\partial^2_{ij}\partial_t^2 u^0\big], \boldsymbol{w}\big)_{\mathcal{L}^2} - \big(\varepsilon^3\kappa_{mij}\partial^2_{ij}\partial_t^2 u^0, \partial_m\boldsymbol{w}\big)_{\mathrm{L}^2}$. Next, we compute the second term as

$$\begin{aligned} \big\langle \mathcal{A}^\varepsilon\bar{\boldsymbol{\mathcal{B}}}^\varepsilon u^0, \boldsymbol{w}\big\rangle = \Big(\big[\quad &\varepsilon^{-1}\big(-\nabla_y\cdot(a(\nabla_y\chi_k + e_k))\big)\partial_k u^0 \\ &+ \quad \big(-\nabla_y\cdot(a(\nabla_y\theta_{ij} + e_i\chi_j)) - e_i^T a(\nabla_y\chi_j + e_j)\big)\partial^2_{ij}u^0 \\ &+ \varepsilon^1 \ \big(-\nabla_y\cdot(a(\nabla_y\kappa_{ijk} + e_i\theta_{jk})) - e_i^T a(\nabla_y\theta_{jk} + e_j\chi_k)\big)\partial^3_{ijk}u^0\big], \boldsymbol{w}\Big)_{\mathcal{L}^2} \\ &+ \big\langle \mathcal{A}^\varepsilon\boldsymbol{\varphi}, \boldsymbol{w}\big\rangle + \big\langle \bar{\boldsymbol{\mathcal{R}}}^\varepsilon_2 u^0, \boldsymbol{w}\big\rangle, \end{aligned}$$

where $\big\langle \bar{\boldsymbol{\mathcal{R}}}^\varepsilon_2 u^0, \boldsymbol{w}\big\rangle = \varepsilon^2\big(\big[-e_i^T a(\nabla_y\kappa_{jkl} + e_j\theta_{kl})\partial^4_{ijkl}u^0\big], \boldsymbol{w}\big)_{\mathcal{L}^2} + \big(a_{mi}\kappa_{jkl}\partial^4_{ijkl}u^0, \partial_m\boldsymbol{w}\big)_{\mathrm{L}^2}$. Using the cell problems for χ_i, θ_{ij}, and κ_{ijk}, we find that $\bar{\boldsymbol{\mathcal{R}}}^\varepsilon u^0 = \bar{\boldsymbol{\mathcal{R}}}^\varepsilon_1 u^0 + \bar{\boldsymbol{\mathcal{R}}}^\varepsilon_2 u^0$ satisfies (5.7). Setting $\boldsymbol{\eta} = \big[u^\varepsilon\big] - \bar{\boldsymbol{\mathcal{B}}}^\varepsilon u^0$, we apply Corollary 4.2.2 and obtain

$$\|\boldsymbol{\eta}\|_{\mathrm{L}^\infty(0,\varepsilon^{-1}T;\mathcal{W})} \le C\varepsilon\Big(\|g^1\|_{\mathrm{H}^3} + \|g^0\|_{\mathrm{H}^3} + |u^0|_{\mathrm{L}^\infty(0,\varepsilon^{-1}T;\mathrm{H}^4(\Omega))} + |\partial_t^2 u^0|_{\mathrm{L}^\infty(0,\varepsilon^{-1}T;\mathrm{H}^2(\Omega))}\Big). \tag{5.8}$$

The definition of $\bar{\boldsymbol{\mathcal{B}}}^\varepsilon$ gives the estimate

$$\|\bar{\boldsymbol{\mathcal{B}}}^\varepsilon u^0 - \big[u^0\big]\|_{\mathrm{L}^\infty(0,\varepsilon^{-1}T;\mathcal{W})} \le C\varepsilon\Big(\textstyle\sum_{k=1}^4 |u^0|_{\mathrm{L}^\infty(\varepsilon^{-1}T;\mathrm{H}^k(\Omega))} + \|f\|_{\mathrm{L}^1(\varepsilon^{-1}T;\mathrm{H}^1(\Omega))}\Big), \tag{5.9}$$

where we used the standard energy estimate for the wave equation to get $\|\bar{\boldsymbol{\varphi}}\|_{\mathrm{L}^\infty(\mathcal{W})} \le \|f\|_{\mathrm{L}^1(\mathrm{H}^1)}$ from (5.6). As $(u^\varepsilon - u^0)(t) \in \mathrm{W}_{\mathrm{per}}(\Omega)$, we have $\|u^\varepsilon - u^0\|_{\mathrm{L}^\infty(W)} = \|\big[u^\varepsilon - u^0\big]\|_{\mathrm{L}^\infty(\mathcal{W})}$ and the triangle inequality gives

$$\|u^\varepsilon - u^0\|_{\mathrm{L}^\infty(0,\varepsilon^{-1}T;W)} \le \|\boldsymbol{\eta}\|_{\mathrm{L}^\infty(0,\varepsilon^{-1}T;\mathcal{W})} + \|\bar{\boldsymbol{\mathcal{B}}}^\varepsilon u^0 - \big[u^0\big]\|_{\mathrm{L}^\infty(0,\varepsilon^{-1}T;\mathcal{W})}. \tag{5.10}$$

Combining (5.10) with (5.8) and (5.9), we obtain estimate (5.5) and the proof of the theorem is complete. □

5.1.2 The family of effective equations for timescales $\mathcal{O}(\varepsilon^{-2})$ is still valid at timescales $\mathcal{O}(\varepsilon^{-3})$

We now turn to timescales $\mathcal{O}(\varepsilon^{-3})$. In particular, we prove that the family of effective equations for timescales $\mathcal{O}(\varepsilon^{-2})$, defined in Definition 4.2.6, is still valid at timescales $\mathcal{O}(\varepsilon^{-3})$.

According to the discussion from Section 4.2.2, in order to prove that a solution $\tilde{u}$ describes well u^ε on $[0, \varepsilon^{-3}T]$, we need to construct an adaptation $[\hat{\mathcal{B}}^\varepsilon \tilde{u}](t)$, which satisfies the properties (4.37) and is such that $(\partial_t^2 + \mathcal{A}^\varepsilon)([\hat{\mathcal{B}}^\varepsilon \tilde{u} - u^\varepsilon])(t) = \mathcal{O}(\varepsilon^4)$ for a.e. t (see (4.38)). To do so, let us come back to the asymptotic expansion of Section 4.2.3. The ansatz on the form of the effective equation remains (4.40) and we now assume that the adaptation has the form

$$\begin{aligned}\hat{\mathcal{B}}^\varepsilon \tilde{u}(t,x) = \tilde{u}\big(t,x\big) + \varepsilon u^1\big(t,x,\tfrac{x}{\varepsilon}\big) + \varepsilon^2 u^2\big(t,x,\tfrac{x}{\varepsilon}\big) + \varepsilon^3 u^3\big(t,x,\tfrac{x}{\varepsilon}\big) + \varepsilon^4 u^4\big(t,x,\tfrac{x}{\varepsilon}\big) \\ + \varepsilon^5 u^5\big(t,x,\tfrac{x}{\varepsilon}\big) + \hat{\varphi}(t,x). \qquad (5.11)\end{aligned}$$

Compared with (4.42), the adaptation $\hat{\mathcal{B}}^\varepsilon \tilde{u}$ contains the two additional terms $\varepsilon^5 u^5\big(t,x,\frac{x}{\varepsilon}\big)$ and $\hat{\varphi}(t,x)$. The term $\varepsilon^5 u^5\big(t,x,\frac{x}{\varepsilon}\big)$ is present to increase the accuracy of the adaptation, i.e, $(\partial_t^2 + \mathcal{A}^\varepsilon)([\hat{\mathcal{B}}^\varepsilon \tilde{u} - u^\varepsilon])(t) = \mathcal{O}(\varepsilon^4)$ (instead of $\mathcal{O}(\varepsilon^3)$), while the term $\hat{\varphi}(t,x)$ is present to cancel the terms coming from the right hand side f (as done in (4.51)). Repeating the process of Section 4.2.3, we obtain successively the definitions

$$\begin{aligned} u^1(t,x,y) &= \chi_i(y)\partial_i \tilde{u}(t,x), & u^2(t,x,y) &= \theta_{ij}(y)\partial_{ij}^2 \tilde{u}(t,x), \\ u^3(t,x,y) &= \kappa_{ijk}(y)\partial_{ijk}^3 \tilde{u}(t,x), & u^4(t,x,y) &= \rho_{ijkl}(y)\partial_{ijkl}^4 \tilde{u}(t,x), \end{aligned}$$

where the correctors solves the cell problems (4.45a), (4.45b), (4.45c), and (4.45d). Using these definitions and the effective equation leads, in place of (4.44), to

$$\begin{aligned}(\partial_t^2 + \mathcal{A}^\varepsilon)(\hat{\mathcal{B}}^\varepsilon \tilde{u} - u^\varepsilon) = \varepsilon^3 \Big(\mathcal{A}_{yy} u^5 + \mathcal{A}_{xy} u^4 + \mathcal{A}_{xx} u^3 + \big((a_{ij}^0 b_{kl}^2 - a_{ijkl}^2)\chi_m + a_{ij}^0 \kappa_{klm}\big)\partial_{ijklm}^5 \tilde{u} \Big) \\ + (\partial_t^2 + \mathcal{A}^\varepsilon)\hat{\varphi} + \varepsilon \chi_i \partial_i f + \varepsilon^2 (b_{ij}^2 + \theta_{ij})\partial_{ij}^2 f + \varepsilon^3 (b_{ij}^2 \chi_k + \kappa_{ijk})\partial_{ijk}^3 f \\ + \mathcal{O}(\varepsilon^4). \qquad (5.12)\end{aligned}$$

To cancel the term of order $\mathcal{O}(\varepsilon^3)$, we thus define $u^5(t,x,y) = \sigma_{ijklm}(y)\partial_{ijklm}^5 \tilde{u}(t,x)$, where σ_{ijklm} is Y-periodic and solves the cell problems

$$\begin{aligned}\varepsilon^3 \ : \ \big(a\nabla_y \sigma_{ijklm}, \nabla_y w\big)_Y = S_{ijklm}^5 \big\{ - \big(a e_i \rho_{jklm}, \nabla_y w\big)_Y + \big(a(\nabla_y \rho_{jklm} + e_j \kappa_{klm}), e_i w\big)_Y \\ + \big((a_{ijkl}^2 - a_{ij}^0 b_{kl}^2)\chi_m - a_{ij}^0 \kappa_{klm}, w\big)_Y \big\}, \qquad (5.13)\end{aligned}$$

for any Y-periodic test functions $w \in \mathrm{H}^1_{\mathrm{per}}(Y)$. Let us prove that the right hand side of (5.13) belongs to $\mathcal{W}^*_{\mathrm{per}}(Y)$. To do so, we prove that it satisfies the solvability condition (A.8), i.e., that the tensor

$$c_{ijklm} = S_{ijklm}^5 \big\{ \big(a(\nabla_y \rho_{jklm} + e_j \kappa_{klm}), e_i\big)_Y + \big((a_{ijkl}^2 - a_{ij}^0 b_{kl}^2)\chi_m - a_{ij}^0 \kappa_{klm}, 1\big)_Y \big\},$$

vanishes for any $1 \leq i,j,k,l,m \leq d$. Using successively the cell problem (4.45a) with the test function ρ_{jklm} and the cell problem (4.45d) with the test function χ_i, we obtain

$$\begin{aligned}\big(a\nabla_y \rho_{jklm}, e_i\big)_Y &= -\big(a\nabla_y \rho_{jklm}, \nabla_y \chi_i\big)_Y \\ &= S_{jklm}^4 \big\{ \big(a e_j \kappa_{klm}, \nabla_y \chi_i\big)_Y - \big(a(\nabla_y \kappa_{klm} + e_k \theta_{lm}), e_j \chi_i\big)_Y \\ &\qquad - \big(a_{jklm}^2 - a_{jk}^0 \theta_{lm} - a_{jk}^0 b_{lm}^2, \chi_i\big)_Y \big\},\end{aligned}$$

and we can rewrite

$$\begin{aligned}c_{ijklm} = S_{ijklm}^5 \big\{ - \big(a e_j \chi_i, \nabla_y \kappa_{klm}\big)_Y + \big(a(\nabla_y \chi_i + e_i), e_j \kappa_{klm}\big)_Y - \big(a_{ij}^0, \kappa_{klm}\big)_Y \\ - \big(a e_k \theta_{lm}, e_j \chi_i\big)_Y + \big(a_{jk}^0 \theta_{lm}, \chi_i\big)_Y \big\}.\end{aligned}$$

Using then (4.45b) with the test function κ_{klm} and (4.45c) with the test function θ_{ji}, we get

$$\begin{aligned} c_{ijklm} &= S^5_{ijklm}\big\{\big(a\nabla_y\theta_{ji},\nabla_y\kappa_{klm}\big)_Y - \big(ae_k\theta_{lm},e_j\chi_i\big)_Y + \big(a^0_{jk}\theta_{lm},\chi_i\big)_Y\big\}\\ &= S^5_{ijklm}\big\{-\big(ae_k\theta_{lm},\nabla_y\theta_{ji}\big)_Y + \big(a(\nabla_y\theta_{lm}+e_l\chi_m)-a^0e_l\chi_m,e_k\theta_{ji})\big)_Y\\ &\qquad -\big(ae_k\theta_{lm},e_j\chi_i\big)_Y + \big(a^0_{jk}\theta_{lm},\chi_i\big)_Y\big\} = 0, \end{aligned}$$

and (5.13) is well-posed in $\mathcal{W}_{\mathrm{per}}(Y)$.

To complete the asymptotic expansion, we define the function $\hat\varphi$ in (5.11) so that the terms containing the source f in (5.12) cancel (i.e. $\hat\varphi\in\hat{\boldsymbol{\varphi}}$, where $\hat{\boldsymbol{\varphi}}$ is defined in (5.14)). Assuming sufficient regularity of the data, we verify that the adaptation $\big[\hat{\mathcal{B}}^\varepsilon\tilde u\big]$ satisfies the requirements (4.37) and (4.38) on the time interval $[0,\varepsilon^{-3}T]$. We prove the following theorem.

Theorem 5.1.3. *Assume that the Y-periodic tensor satisfies $a(y)\in \mathrm{W}^{2,\infty}(Y)$ and let $\tilde u$ belongs to the family of effective equations $\mathcal{E}$ defined in Definition 4.2.6. Furthermore, assume that the data and $\tilde u$ satisfy the regularity*

$$\begin{gathered}\tilde u\in \mathrm{L}^\infty(0,\varepsilon^{-3}T;\mathrm{H}^6(\Omega)),\quad \partial_t\tilde u\in \mathrm{L}^\infty(0,\varepsilon^{-3}T;\mathrm{H}^5(\Omega)),\quad \partial_t^2\tilde u\in \mathrm{L}^\infty(0,\varepsilon^{-3}T;\mathrm{H}^4(\Omega)),\\ g^0\in\mathrm{H}^5(\Omega),\quad g^1\in\mathrm{H}^5(\Omega),\quad f\in\mathrm{L}^2(0,\varepsilon^{-3}T;\mathrm{H}^3(\Omega)).\end{gathered}$$

Then the following error estimate holds

$$\begin{aligned}\|u^\varepsilon-\tilde u\|_{\mathrm{L}^\infty(0,\varepsilon^{-3}T;W)} \le C\varepsilon\Big(&\|g^1\|_{\mathrm{H}^5(\Omega)}+\|g^0\|_{\mathrm{H}^5(\Omega)}+\|f\|_{\mathrm{L}^1(0,\varepsilon^{-3}T;\mathrm{H}^3(\Omega))}\\ &+\textstyle\sum_{k=1}^6|\tilde u|_{\mathrm{L}^\infty(0,\varepsilon^{-3}T;\mathrm{H}^k(\Omega))}+|\partial_t^2\tilde u|_{\mathrm{L}^\infty(0,\varepsilon^{-3}T;\mathrm{H}^4(\Omega))}\Big),\end{aligned}$$

where C depends only on $T,\lambda,|b^2|_\infty,|a^2|_\infty,\|a\|_{\mathrm{W}^{2,\infty}(Y)}$, and Y, and we recall the definition of the norm (see (A.4)*)*

$$\|w\|_W=\inf_{\substack{w=w_1+w_2\\ w_1,w_2\in\mathrm{W}_{\mathrm{per}}(\Omega)}}\Big\{\|w_1\|_{\mathrm{L}^2(\Omega)}+\|\nabla w_2\|_{\mathrm{L}^2(\Omega)}\Big\}\quad\forall w\in\mathrm{W}_{\mathrm{per}}(\Omega).$$

Remark 5.1.4. Referring to Section 4.2.6, a similar result as Theorem 5.1.3 can be proved for a bounded tensor $a(y)\in\mathrm{L}^\infty(Y)$, provided

$$\begin{gathered}\tilde u\in \mathrm{L}^\infty(0,\varepsilon^{-3}T;\mathrm{H}^8(\Omega)),\quad \partial_t\tilde u\in \mathrm{L}^\infty(0,\varepsilon^{-3}T;\mathrm{H}^7(\Omega)),\quad \partial_t^2\tilde u\in \mathrm{L}^\infty(0,\varepsilon^{-3}T;\mathrm{H}^5(\Omega)),\\ g^0\in\mathrm{H}^7(\Omega),\quad g^1\in\mathrm{H}^7(\Omega),\quad f\in\mathrm{L}^2(0,\varepsilon^{-3}T;\mathrm{H}^5(\Omega)).\end{gathered}$$

Proof. Define the adaptation operator $\hat{\boldsymbol{\mathcal{B}}}^\varepsilon:\mathrm{L}^2(0,\varepsilon^{-3}T;\mathrm{H}^4(\Omega))\to\mathrm{L}^2(0,\varepsilon^{-3}T;\mathcal{W}^*_{\mathrm{per}}(\Omega))$ as

$$\begin{aligned}\big\langle\hat{\boldsymbol{\mathcal{B}}}^\varepsilon v(t),\boldsymbol{w}\big\rangle =\big\langle\boldsymbol{\mathcal{B}}^\varepsilon v(t),\boldsymbol{w}\big\rangle &-\big(\varepsilon^4\big[\partial_{y_m}\sigma_{mijkl}\partial^4_{ijkl}v(t)\big],\boldsymbol{w}\big)_{\mathcal{L}^2}-\big(\varepsilon^5\sigma_{mijkl}\partial^4_{ijkl}v(t),\partial_m\boldsymbol{w}\big)_{\mathrm{L}^2}\\ &+\big\langle\hat{\boldsymbol{\varphi}}(t),\boldsymbol{w}\big\rangle,\end{aligned}$$

where $\big\langle\cdot,\cdot\big\rangle=\big\langle\cdot,\cdot\big\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}}$, $\boldsymbol{\mathcal{B}}^\varepsilon$ is the adaptation operator defined in (4.62), and $\hat{\boldsymbol{\varphi}}\in\mathrm{L}^\infty(0,\varepsilon^{-3}T;\mathcal{W}_{\mathrm{per}}(\Omega))$ is the unique solution of

$$\begin{aligned}(\partial_t^2+\mathcal{A}^\varepsilon)\hat{\boldsymbol{\varphi}} &= -\big[\varepsilon\chi_i\partial_i f+\varepsilon^2\big(b^2_{ij}+\theta_{ij}\big)\partial^2_{ij}f\\ &\qquad+\varepsilon^3\big(b^2_{ij}\chi_k+\kappa_{ijk}\big)\partial^3_{ijk}f\big]\quad\text{in }\mathcal{W}^*_{\mathrm{per}}(\Omega)\text{ a.e. }t\in[0,\varepsilon^{-3}T],\\ \hat{\boldsymbol{\varphi}}(0)=\partial_t\hat{\boldsymbol{\varphi}}(0) &= \big[0\big].\end{aligned}\tag{5.14}$$

The rest of the proof follows the same steps as the proofs of Theorems 4.2.4 and 5.1.1. □

5.2 Effective equations for arbitrary timescales

In Section 4.2.2, we described a procedure to derive effective equations for wave propagation at timescales $\mathcal{O}(\varepsilon^{-\alpha})$, where $\alpha \geq 0$ is an integer. In particular, the main task is the construction of an adaptation of an effective solution using asymptotic expansion. This procedure was used in Section 4.2.3 to derive a family of effective equations $\mathcal{E}$ valid at timescales $\mathcal{O}(\varepsilon^{-2})$. It was then used in Section 5.1, to prove that the homogenized equation is valid at timescales $\mathcal{O}(\varepsilon^{-1})$ and that the family $\mathcal{E}$ is still valid at timescales $\mathcal{O}(\varepsilon^{-3})$. In this section, the same procedure is applied at arbitrary timescales. The main result of the chapter is presented in Section 5.2.1, where we define a family of effective equations for arbitrary timescales. The technical construction of the adaptation, i.e., the derivation of the cell problems of arbitrary order, is postponed to Section 5.2.5.

As suggested by the results of the previous section, an effective equation valid at even timescales $\mathcal{O}(\varepsilon^{-\alpha})$ (α even) is still valid for timescales $\mathcal{O}(\varepsilon^{-(\alpha+1)})$. This fact is already known from the Bloch wave theory as the odd derivatives of the first Bloch eigenvalue cancel (see e.g. [42], [23]). In our derivation, it is a consequence of the unconditional well-posedness of the odd order cell problems. This important feature follows a technical relation between the solutions of the cell problems (see (5.25) and Lemma 5.2.5). The second result of this chapter is a new relation between the solutions of the cell problems that allows to reduce the cost of computation of the effective tensors. This remarkable relation is proved in Lemma 5.2.6, in Section 5.2.2, and discussed in Section 5.2.4.

In the whole section, we denote $(\cdot,\cdot) = (\cdot,\cdot)_{\mathrm{L}^2(Y)}$ and assume $|Y| = 1$ for simplicity. As we deal with tensors of arbitrary order, let us introduce some definitions and notations. A tensor $q \in \mathrm{Ten}^{2n}(\mathbb{R}^d)$ is major symmetric if

$$q_{i_1\cdots i_n i_{n+1}\cdots i_{2n}} = q_{i_{n+1}\cdots i_{2n} i_1\cdots i_n} \quad 1 \leq i_1 \cdots i_{2n} \leq d. \tag{5.15}$$

A tensor $q \in \mathrm{Ten}^{2n}(\mathbb{R}^d)$ is positive semidefinite if

$$q_{i_1\cdots i_{2n}}\xi_{i_1\cdots i_n}\xi_{i_{n+1}\cdots i_{2n}} \geq 0 \quad \forall \xi \in \mathrm{Sym}^n(\mathbb{R}^d), \tag{5.16}$$

and it is positive definite if

$$q_{i_1\cdots i_{2n}}\xi_{i_1\cdots i_n}\xi_{i_{n+1}\cdots i_{2n}} > 0 \quad \forall \xi \in \mathrm{Sym}^n(\mathbb{R}^d)\backslash\{0\}. \tag{5.17}$$

We use the standard notation for the tensor product

$$\otimes : \mathrm{Ten}^m(\mathbb{R}^d) \times \mathrm{Ten}^n(\mathbb{R}^d) \to \mathrm{Ten}^{n+m}(\mathbb{R}^d), \quad (p,q) \mapsto (p\otimes q)_{i_1\cdots i_{m+n}} = p_{i_1\cdots i_m} q_{i_{m+1}\cdots i_{m+n}}.$$

Furthermore, we use the shorthand notation

$$\otimes^s p = \underbrace{p \otimes \cdots \otimes p}_{s \text{ times}}.$$

To improve the readability, the differential operator $q_{i_1\cdots i_n}\partial^n_{i_1\cdots i_n}$ is denoted $q\partial^n$, i.e.,

$$q\partial^n = q_{i_1\cdots i_n}\partial^n_{i_1\cdots i_n}. \tag{5.18}$$

Note that for a sufficiently smooth function v, any $q \in \mathrm{Ten}^n(\mathbb{R}^d)$ satisfies $q\partial^n = S^n(q)\partial^n$, where $S^n(q) \in \mathrm{Sym}^n(\mathbb{R}^d)$ is the symmetrization of q. For this reason, in the derivation we mostly deal with symmetric tensors. We denote $=_S$ an equality that holds up to symmetries, i.e., $p, q \in \mathrm{Ten}^n(\mathbb{R}^d)$ satisfy $p =_S q$ if and only if $S^n(p) = S^n(q)$. Note that, up to symmetries, the tensor products is commutative: $p \otimes q =_S q \otimes p \ \ \forall p \in \mathrm{Ten}^n(\mathbb{R}^d), q \in \mathrm{Ten}^m(\mathbb{R}^d)$. Finally, we denote $I(d,n)$ the set of multiindices of the distinct entries of a tensor in $\mathrm{Sym}^n(\mathbb{R}^d)$, i.e.,

$$I(d,n) = \big\{i = (i_1,\ldots,i_n) : 1 \leq i_1 \leq \ldots \leq i_n \leq d\big\}.$$

We verify that the cardinality of $I(d,n)$ is given by $N(d,n) = |I(d,n)| = \binom{d+n-1}{n}$.

5.2.1 A priori error estimate and family of effective equations

We present here the main result of this chapter, contributing to the thesis. We define a family of effective equations that capture the macroscopic behavior of u^ε on arbitrarily large timescales. The family relies on constraints for the effective tensors that are imposed by the well-posedness of the cell problems of arbitrary order. The derivation of these cell problems is presented in detail in Section 5.2.5.

Let α be an integer and let $\Omega \subset \mathbb{R}^d$ be an arbitrarily large hypercube, assumed to be the union of cells of volume $\varepsilon|Y|$ (see assumption (4.25), Figure 4.2). Let a^0 be the homogenized tensor and for $r = 1, \dots, \lfloor \alpha/2 \rfloor$, let $a^{2r} \in \mathrm{Ten}^{2r+2}(\mathbb{R}^d)$, $b^{2r} \in \mathrm{Ten}^{2r}(\mathbb{R}^d)$ be positive semidefinite major symmetric tensors (see (5.15) and (5.16)). We consider the equation: $\tilde{u} : [0, \varepsilon^{-\alpha}T] \times \Omega \to \mathbb{R}$ such that

$$\begin{aligned} &\partial_t^2 \tilde{u} - a^0 \partial^2 \tilde{u} - \sum_{r=1}^{\lfloor \alpha/2 \rfloor} (-1)^r \varepsilon^{2r} \big(a^{2r} \partial^{2r+2} \tilde{u} - b^{2r} \partial^{2r} \partial_t^2 \tilde{u} \big) = f && \text{in } (0, \varepsilon^{-\alpha}T] \times \Omega, \\ &x \mapsto \tilde{u}(t,x) \quad \Omega\text{-periodic} && \text{in } [0, \varepsilon^{-\alpha}T], \\ &\tilde{u}(0,x) = g^0(x), \quad \partial_t \tilde{u}(0,x) = g^1(x) && \text{in } \Omega, \end{aligned} \tag{5.19}$$

where we recall the notation for the differential operators (5.18). The existence and uniqueness of a weak solution of (5.19) is ensured by the non-negative signs of the tensors if the data satisfy the regularity (see Section 5.2.5 for more details)

$$g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^{\lfloor \alpha/2 \rfloor + 1}(\Omega), \quad g^1 \in \mathrm{L}^2_0(\Omega) \cap \mathrm{H}^{\lfloor \alpha/2 \rfloor}(\Omega), \quad f \in \mathrm{L}^2(0, \varepsilon^{-\alpha}T; \mathrm{L}^2_0(\Omega)).$$

Remark 5.2.1. Following the argument of Remark 4.2.3, the right hand side of the effective equation (5.19) could also be corrected as $f - \mathcal{S}_1^\varepsilon f$, where $\mathcal{S}_1^\varepsilon f$ is defined in Remark 5.2.12.

Let us summarize how the cell problems of arbitrary order are obtained (see Section 5.2.5 for the full derivation). We look for an adaptation of $\tilde{u}$ of the form

$$\mathcal{B}^\varepsilon \tilde{u}(t,x) = \tilde{u}(t,x) + \sum_{k=1}^{\alpha+2} \varepsilon^k \chi^k\big(\tfrac{x}{\varepsilon}\big) \partial^k \tilde{u}(t,x), \tag{5.20}$$

where $\{\chi^k_{i_1 \cdots i_k}\}_{k=1}^{\alpha+2}$ are Y-periodic functions to be defined and we recall that $\partial^k \tilde{u}$ is the tensor of $\mathrm{Ten}^k(\mathbb{R}^d)$ with coordinates $\partial^k_{i_1 \cdots i_k} \tilde{u}$. Following the argument of Section 4.2.2, we need to build $\mathcal{B}^\varepsilon \tilde{u}$ such that $r^\varepsilon = (\partial_t^2 + \mathcal{A}^\varepsilon)(\mathcal{B}^\varepsilon \tilde{u} - u^\varepsilon)$ is of order $\mathcal{O}(\varepsilon^{\alpha+1})$. Applying inductive Boussinesq tricks, we substitute $\partial_t^2 \tilde{u}$ in the terms of order ε^{-1} to ε^α in r^ε. This technical task is postponed to Section 5.2.5. Canceling then the terms in the expansion, we obtain the cell problems of order 1 to $\alpha + 2$. They read as follows. Define the tensor $c^r \in \mathrm{Ten}^{2r}(\mathbb{R}^d)$ as

$$c^0 = a^0, \qquad c^r = a^{2r} - \sum_{\ell=0}^{r-1} b^{2(r-\ell)} \otimes c^\ell \quad r = 1, \dots, \lfloor \alpha/2 \rfloor. \tag{5.21}$$

Let then $\{\chi^k_{i_1 \cdots i_k}\}_{k=1}^{\alpha+2}$ be functions in $\mathrm{H}^1_{\mathrm{per}}(Y)$ such that for all test functions $w \in \mathrm{H}^1_{\mathrm{per}}(Y)$,

$$\big(a \nabla_y \chi^1_i, \nabla_y w\big) = -\big(a e_i, \nabla_y w\big), \tag{5.22a}$$

$$\big(a \nabla_y \chi^2_{i_1 i_2}, \nabla_y w\big) = S^2_{i_1 i_2} \big\{ -\big(a e_{i_1} \chi^1_{i_2}, \nabla_y w\big) + \big(a(\nabla_y \chi^1_{i_2} + e_{i_2}) - a^0 e_{i_2}, e_{i_1} w\big) \big\}, \tag{5.22b}$$

$$\begin{aligned} \Big(a \nabla_y \chi^{2r+1}_{i_1 \cdots i_{2r+1}}, \nabla_y w\Big) = S^{2r+1}_{i_1 \cdots i_{2r+1}} \Big\{ &-\Big(a e_{i_1} \chi^{2r}_{i_2 \cdots i_{2r+1}}, \nabla_y w\Big) \\ &+ \Big(a\big(\nabla_y \chi^{2r}_{i_2 \cdots i_{2r+1}} + e_{i_2} \chi^{2r-1}_{i_3 \cdots i_{2r+1}}\big), e_{i_1} w\Big) \\ &+ \Big(\sum_{\ell=1}^{r} (-1)^{r-\ell+1} \big(c^{r-\ell} \otimes \chi^{2\ell-1}\big)_{i_1 \cdots i_{2r+1}}, w\Big) \Big\}, \end{aligned} \tag{5.22c}$$

$$\begin{aligned}\Big(a\nabla_y\chi^{2r+2}_{i_1\cdots i_{2r+2}},\nabla_y w\Big) = S^{2r+2}_{i_1\cdots i_{2r+2}}\Big\{ &-\Big(ae_{i_1}\chi^{2r+1}_{i_2\cdots i_{2r+2}},\nabla_y w\Big)\\ &+\Big(a\big(\nabla_y\chi^{2r+1}_{i_2\cdots i_{2r+2}}+e_{i_2}\chi^{2r}_{i_3\cdots i_{2r+2}}\big),e_{i_1}w\Big)\\ &+\Big(\sum_{\ell=1}^{r}(-1)^{r-\ell+1}\big(c^{r-\ell}\otimes\chi^{2\ell}\big)_{i_1\cdots i_{2r+2}},w\Big)\\ &-\Big((-1)^r c^r_{i_1\cdots i_{2r+2}},w\Big)\Big\},\end{aligned} \tag{5.22d}$$

where $(\cdot,\cdot)=(\cdot,\cdot)_{\mathrm{L}^2(Y)}$. We recognize that χ^1 and χ^2 are the two first cell functions defined in (4.45a) and (4.45b) in Section 4.2.3. Furthermore, we can verify that χ^3,χ^4, and χ^5 are the cell functions defined in (4.45c), (4.45d), and (5.13), respectively.

We now investigate the well-posedness of the cell problems (5.22) in order to derive the constraints on $\{a^{2r},b^{2r}\}$. For simplicity, we choose the zero mean correctors $\chi^k_{i_1\cdots i_k}\in\mathrm{W}_{\mathrm{per}}(Y)$. Recall that we assume $|Y|=1$. To apply the Lax–Milgram theorem and prove that the cell problems are well-posed in $\mathrm{W}_{\mathrm{per}}(Y)$, we need the right hand sides to belong to $\mathrm{W}^*_{\mathrm{per}}(Y)$. In Appendix A.2, we provide a characterization of $\mathrm{W}^*_{\mathrm{per}}(Y)$. In particular, $F\in[\mathrm{H}^1_{\mathrm{per}}(Y)]^*$ given by

$$\big\langle F,w\big\rangle=\big(f^0,w\big)_{\mathrm{L}^2(Y)}+\big(f^1_k,\partial_k w\big)_{\mathrm{L}^2(Y)},$$

for some $f^0,f^1_1,\dots,f^1_d\in\mathrm{L}^2(Y)$ belongs to $\mathrm{W}^*_{\mathrm{per}}(Y)$ if and only if

$$\big(f^0,1\big)_{\mathrm{L}^2(Y)}=0. \tag{5.23}$$

Let us then consider the cell problems (5.22). As already seen, (5.22a) is well-posed unconditionally, while (5.22b) is well-posed provided a^0 is the homogenized tensor (see Section 4.2.3). Similarly, we verified in Section 4.2.3 that the cell problem for χ^3 is well-posed unconditionally, while the cell problem for χ^4 is well-posed provided a^2,b^2 satisfy the constraint (see (4.49))

$$c^1=a^2-b^2\otimes a^0=_S-h^1,\qquad h^1_{i_1\cdots i_4}=\big(a\big(\nabla_y\chi^3_{i_2i_3i_4}+e_{i_2}\chi^2_{i_3i_4}\big),e_{i_1}\big). \tag{5.24}$$

Let us derive the constraint imposed by the well-posedness of the higher order cell problems. Assume that the cell problems are well-posed up to order $2r$, where $r\geq 2$. The odd order cell problem (5.22c) is well-posed provided

$$\big(a\big(\nabla_y\chi^{2r}_{i_2\cdots i_{2r+1}}+e_{i_2}\chi^{2r-1}_{i_3\cdots i_{2r+1}}\big),e_{i_1}\big)=_S 0. \tag{5.25}$$

This equality is verified in Lemma 5.2.5, in Section 5.2.2. Next, the even order cell problem (5.22d) is well-posed provided the following equality holds:

$$c^r=_S(-1)^r h^r,\qquad h^r_{i_1\cdots i_{2r+2}}=\big(a\big(\nabla_y\chi^{2r+1}_{i_2\cdots i_{2r+2}}+e_{i_2}\chi^{2r}_{i_2\cdots i_{2r+2}}\big),e_{i_1}\big). \tag{5.26}$$

Using the definition of c^r in (5.21), the constraint (5.26) can be rewritten for a^{2r},b^{2r} as

$$a^{2r}-b^{2r}\otimes a^0=_S(-1)^r h^r+\sum_{\ell=1}^{r-1}b^{2(r-\ell)}\otimes c^\ell. \tag{5.27}$$

Equality (5.27) is thus the characterization of the family of effective equations. Namely, if (5.19) is well-posed and if its tensors satisfy (5.27) and (5.24), the solution describes well u^ε. Indeed, we verify the following result.

Theorem 5.2.2. *Assume that the Y-periodic tensor satisfies $a(y)\in\mathrm{W}^{2,\infty}(Y)$. Furthermore, assume that the data and the solution of* (5.19) *satisfy the regularity*

$$\begin{gathered}\tilde u\in\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathrm{H}^{\alpha+3}(\Omega)),\quad \partial_t\tilde u\in\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathrm{H}^{\alpha+2}(\Omega)),\quad \partial_t^2\tilde u\in\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathrm{H}^{\alpha+1}(\Omega)),\\ g^0\in\mathrm{H}^{\alpha+2}(\Omega),\quad g^1\in\mathrm{H}^{\alpha+2}(\Omega),\quad f\in\mathrm{L}^2(0,\varepsilon^{-\alpha}T;\mathrm{H}^{\alpha}(\Omega)).\end{gathered}$$

Let $\{\chi^k_{i_1\cdots i_k}\}_{k=1}$ *be the zero mean solutions of the cell problems* (5.22) *and assume that the tensors* $\{a^{2r}, b^{2r}\}_{r=1}^{\lfloor\alpha/2\rfloor}$ *satisfy the constraints* (5.24) *and* (5.27)*. Then the following error estimate holds*

$$\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;W)} \leq C\varepsilon\Big(\|g^1\|_{\mathrm{H}^{\alpha+2}(\Omega)} + \|g^0\|_{\mathrm{H}^{\alpha+2}(\Omega)} + \|f\|_{\mathrm{L}^1(0,\varepsilon^{-\alpha}T;\mathrm{H}^\alpha(\Omega))} \\ + \textstyle\sum_{k=1}^{\alpha+3} |\tilde{u}|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathrm{H}^k(\Omega))} + |\partial_t^2\tilde{u}|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathrm{H}^{\alpha+1}(\Omega))} \Big),$$

where C *depends only on* T, λ, $\{|b^{2r}|_\infty, |a^{2r}|_\infty\}_{r=1}^{\lfloor\alpha/2\rfloor}$, $\|a\|_{\mathrm{W}^{2,\infty}(Y)}$, *and* Y, *and we recall the definition of the norm (see* (A.4)*)*

$$\|w\|_W = \inf_{\substack{w=w_1+w_2 \\ w_1,w_2\in \mathrm{W}_{\mathrm{per}}(\Omega)}} \Big\{ \|w_1\|_{\mathrm{L}^2(\Omega)} + \|\nabla w_2\|_{\mathrm{L}^2(\Omega)} \Big\} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(\Omega).$$

Remark 5.2.3. Referring to Section 4.2.6, a similar result as Theorem 5.2.2 can be proved for a bounded tensor $a(y) \in \mathrm{L}^\infty(Y)$, provided

$$\tilde{u} \in \mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathrm{H}^{\alpha+5}(\Omega)), \quad \partial_t\tilde{u} \in \mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathrm{H}^{\alpha+4}(\Omega)), \quad \partial_t^2\tilde{u} \in \mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathrm{H}^{\alpha+3}(\Omega)),$$
$$g^0 \in \mathrm{H}^{\alpha+4}(\Omega), \quad g^1 \in \mathrm{H}^{\alpha+4}(\Omega), \quad f \in \mathrm{L}^2(0,\varepsilon^{-\alpha}T;\mathrm{H}^{\alpha+2}(\Omega)).$$

Thanks to Theorem 5.2.2, we define the family of effective equations for arbitrary timescales.

Definition 5.2.4. The family $\mathcal{E}$ of effective equations is the set of equations (5.19), where the tensors $\{b^{2r}, a^{2r}\}_{r=1}^{\lfloor\alpha/2\rfloor}$ are major symmetric, positive semidefinite (see (5.15) and (5.16)), and satisfy the constraints (5.24) and (5.27).

The proof of Theorem 5.2.2 has the same structure as for Theorems 4.2.3, 5.1.1, and 5.1.3. First, we define an adaptation operator

$$\boldsymbol{\mathcal{B}}^\varepsilon : \mathrm{L}^2(0,\varepsilon^{-\alpha}T;\mathrm{H}^{\alpha+3}(\Omega)) \to \mathrm{L}^2(0,\varepsilon^{-\alpha}T;\mathcal{W}_{\mathrm{per}}(\Omega)),$$

such that $\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u} = \big[\mathcal{B}^\varepsilon\tilde{u}\big] + \boldsymbol{\varphi}$, where $\mathcal{B}^\varepsilon\tilde{u}$ is defined in (5.20) and $\boldsymbol{\varphi}$ is the unique solution of

$$(\partial_t^2 + \mathcal{A}^\varepsilon)\boldsymbol{\varphi}(t,x) = -\big[\mathcal{S}^\varepsilon f(t,x)\big] \quad \text{in } \mathcal{W}^*_{\mathrm{per}}(\Omega) \text{ a.e. } t \in [0,\varepsilon^{-\alpha}T],$$
$$\boldsymbol{\varphi}(0) = \partial_t\boldsymbol{\varphi}(0) = \big[0\big],$$

where $\mathcal{S}^\varepsilon f$ is defined in Remark 5.2.12. Note that the assumptions on the effective tensors (5.24) and (5.27) ensure the well-posedness of the cell problems and thus $\boldsymbol{\mathcal{B}}^\varepsilon$ is well defined. Next, we define the remainder $\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u}(t) = (\partial_t^2 + \mathcal{A}^\varepsilon)(\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}(t)) - \big[f(t)\big]$ and, using the equation (5.19), we substitute $\partial_t^2\tilde{u}$ in every term of $\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u}$ up to order $\mathcal{O}(\varepsilon^\alpha)$. The cell problems ensure then that $\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u}$ can be written as

$$\big\langle \boldsymbol{\mathcal{R}}^\varepsilon\tilde{u}(t), \boldsymbol{w}\big\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}} = \big((\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u})_0(t), \boldsymbol{w}\big)_{\mathcal{L}^2} + \big((\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u})_1(t), \nabla\boldsymbol{w}\big)_{\mathrm{L}^2},$$

where $(\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u})_0(t)$ and $(\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u})_1(t)$ satisfy

$$\|(\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u})_0\|_{\mathrm{L}^\infty(\mathcal{L}^2(\Omega))} + \|(\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u})_1\|_{\mathrm{L}^\infty(\mathrm{L}^2(\Omega))} \leq C\varepsilon^{\alpha+1}\Big(|\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^{\alpha+3})} + |\partial_t^2\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^{\alpha+1})}\Big).$$

Hence, Corollary 4.2.2 ensures that $\boldsymbol{\eta} = \big[u^\varepsilon\big] - \boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}$ satisfies

$$\|\boldsymbol{\eta}\|_{\mathrm{L}^\infty(\mathcal{W})} \leq C\varepsilon\Big(|\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^{\alpha+3})} + |\partial_t^2\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^{\alpha+1})} + \|g^0\|_{\mathrm{H}^{\alpha+2}} + \|g^1\|_{\mathrm{H}^{\alpha+2}}\Big). \tag{5.28}$$

As $(u^\varepsilon - \tilde{u})(t) \in \mathrm{W}_{\mathrm{per}}(\Omega)$, the triangle inequality gives the estimate

$$\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(W)} = \|\big[u^\varepsilon - \tilde{u}\big]\|_{\mathrm{L}^\infty(\mathcal{W})} \leq \|\boldsymbol{\eta}\|_{\mathrm{L}^\infty(\mathcal{W})} + \|\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u} - \big[\tilde{u}\big]\|_{\mathrm{L}^\infty(\mathcal{W})},$$

which, combined with (5.28) and the trivial bound for $\|\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u} - \big[\tilde{u}\big]\|_{\mathrm{L}^\infty(\mathcal{W})}$, proves the theorem.

5.2.2 Two remarkable relations between the solutions of the cell problems

In this section, we present two technical relations between the solutions of the cell problems defined in (5.22). The first one, Lemma 5.2.5, guarantees that the odd order cell problems are well-posed unconditionally (see (5.25)). This fact is already known in the context of Bloch wave theory (see e.g. [42], [23] and the references therein). In particular, this feature implies that no additional correction is required in the effective equation for odd timescales. The second relation is a new result. Recall that in the case $r = 1$, in Section 4.2.3, the dependence on χ^3 of the constraint h^1, in (5.24), was removed. This remarkable fact is generalized in Lemma 5.2.6, were we show that the constraint h^r, in (5.27), can be computed with $\{\chi^k\}_{k=1}^{r+1}$ instead of $\{\chi^k\}_{k=1}^{2r+1}$. As discussed in Section 5.2.4, the consequence of this relation is an meaningful reduction of the computational cost for the computation of the effective tensors.

Lemma 5.2.5. *For any $1 \leq r \leq \alpha/2$, we have*

$$\Big(a\big(\nabla_y \chi^{2r}_{i_2\cdot\cdot i_{2r+1}} + e_{i_2}\chi^{2r-1}_{i_3\cdot\cdot i_{2r+1}}\big), e_{i_1}\Big) =_S 0.$$

Proof. First note that thanks to the symmetry of a, the terms of the following sum cancel two by two :

$$T = \sum_{k=1}^{2r} (-1)^k \big(a\nabla_y \chi^k_{i_1\cdot\cdot i_k}, \nabla_y \chi^{2r+1-k}_{i_{k+1}\cdot\cdot i_{2r+1}}\big) =_S 0. \tag{5.29}$$

We write T as

$$T = -\big(a\nabla_y \chi^1_{i_1}, \nabla_y \chi^{2r}_{i_2\cdot\cdot i_{2r+1}}\big) + \big(a\nabla_y \chi^2_{i_1 i_2}, \nabla_y \chi^{2r-1}_{i_3\cdot\cdot i_{2r+1}}\big) + T^1 + T^2,$$

where, if $r = 1$, $T^1 = T^2 = 0$ and, if $r \geq 2$, T^1 and T^2 are the sums over the odd and the even indices, respectively:

$$T^1 = -\sum_{s=1}^{r-1} \big(a\nabla_y \chi^{2s+1}_{i_1\cdot\cdot i_{2s+1}}, \nabla_y \chi^{2(r-s)}_{i_{2s+2}\cdot\cdot i_{2r+1}}\big), \qquad T^2 = \sum_{s=1}^{r-1} \big(a\nabla_y \chi^{2s+2}_{i_1\cdot\cdot i_{2s+2}}, \nabla_y \chi^{2(r-s)-1}_{i_{2s+3}\cdot\cdot i_{2r+1}}\big).$$

Using (5.22a), (5.22b), and the symmetry of a, we find that

$$\begin{aligned} T =_S & \big(a(\nabla_y \chi^{2r}_{i_2\cdot\cdot i_{2r+1}} + e_{i_2}\chi^{2r-1}_{i_3\cdot\cdot i_{2r+1}}), e_{i_1}\big) \\ & - \big(ae_{i_1}\chi^1_{i_2}, \nabla_y \chi^{2r-1}_{i_3\cdot\cdot i_{2r+1}}\big) + \big(a\nabla_y \chi^1_{i_2}, e_{i_1}\chi^{2r-1}_{i_3\cdot\cdot i_{2r+1}}\big) + T^1 + T^2. \end{aligned} \tag{5.30}$$

We claim that

$$T^1 + T^2 =_S \big(ae_{i_1}\chi^1_{i_2}, \nabla_y \chi^{2r-1}_{i_3\cdot\cdot i_{2r+1}}\big) - \big(a\nabla_y \chi^1_{i_2}, e_{i_1}\chi^{2r-1}_{i_3\cdot\cdot i_{2r+1}}\big). \tag{5.31}$$

If $r = 1$, (5.31) is trivial. Let us prove it for $r \geq 2$. We use (5.22c) and (5.22d) to write T^1 and T^2 as $T^i =_S T^i_1 + T^i_2 + T^i_3$, where

$$\begin{aligned} T^1_1 &=_S \sum_{s=1}^{r-1} \big(ae_{i_1}\chi^{2s}_{i_2\cdot\cdot i_{2s+1}}, \nabla_y \chi^{2(r-s)}_{i_{2s+2}\cdot\cdot i_{2r+1}}\big) - \sum_{s=1}^{r-1} \big(a\nabla_y \chi^{2s}_{i_2\cdot\cdot i_{2s+1}}, e_{i_1}\chi^{2(r-s)}_{i_{2s+2}\cdot\cdot i_{2r+1}}\big), \\ T^1_2 &=_S -\sum_{s=1}^{r-1} \big(ae_{i_2}\chi^{2s-1}_{i_3\cdot\cdot i_{2s+1}}, e_{i_1}\chi^{2(r-s)}_{i_{2s+2}\cdot\cdot i_{2r+1}}\big), \\ T^1_3 &=_S -\sum_{s=1}^{r-1}\sum_{\ell=1}^{s} (-1)^{s-\ell+1} \big((c^{s-\ell} \otimes \chi^{2\ell-1})_{i_1\cdot\cdot i_{2s+1}}, \chi^{2(r-s)}_{i_{2s+2}\cdot\cdot i_{2r+1}}\big), \end{aligned}$$

$$
\begin{aligned}
T_1^2 &=_S -\sum_{s=1}^{r-1} \big(a e_{i_1} \chi^{2s+1}_{i_2\cdot\cdot i_{2s+3}}, \nabla_y \chi^{2(r-s)-1}_{i_{2s+3}\cdot\cdot i_{2r+1}}\big) + \sum_{s=1}^{r-1} \big(a \nabla_y \chi^{2s+1}_{i_2\cdot\cdot i_{2s+2}}, e_{i_1} \chi^{2(r-s)-1}_{i_{2s+3}\cdot\cdot i_{2r+1}}\big),\\
T_2^2 &=_S \sum_{s=1}^{r-1} \big(a e_{i_2} \chi^{2s}_{i_3\cdot\cdot i_{2s+2}}, e_{i_1} \chi^{2(r-s)-1}_{i_{2s+3}\cdot\cdot i_{2r+1}}\big),\\
T_3^2 &=_S \sum_{s=1}^{r-1}\sum_{\ell=1}^{s} (-1)^{s-\ell+1} \big((c^{s-\ell}\otimes \chi^{2\ell})_{i_1\cdot\cdot i_{2s+3}}, \chi^{2(r-s)-1}_{i_{2s+3}\cdot\cdot i_{2r+1}}\big).
\end{aligned}
$$

Changing the index as $m = r - s$ in T_1^1 gives $T_1^1 =_S 0$. The same change of indices in T_2^1 gives $T_2^1 + T_2^2 =_S 0$. Next, in T_3^1, we make the change of indices $m = r - s$, invert the order of summation, and again change $k = r - \ell$ to get

$$
\begin{aligned}
T_3^1 &= \sum_{s=1}^{r-1}\sum_{\ell=1}^{s} (-1)^{s-\ell} c^{s-\ell}\otimes\chi^{2\ell-1}\otimes\chi^{2(r-s)} = \sum_{m=1}^{r-1}\sum_{\ell=1}^{r-m} (-1)^{r-m-\ell} c^{r-m-\ell}\otimes\chi^{2\ell-1}\otimes\chi^{2m}\\
&= \sum_{\ell=1}^{r}\sum_{m=1}^{r-\ell} (-1)^{r-m-\ell} c^{r-m-\ell}\otimes\chi^{2\ell-1}\otimes\chi^{2m} = \sum_{k=1}^{r-1}\sum_{m=1}^{k} (-1)^{k-m} c^{k-m}\otimes\chi^{2(r-k)-1}\otimes\chi^{2m} = -T_3^2,
\end{aligned}
$$

which proves that $T_3^1 + T_3^2 =_S 0$. Combining the different equalities for the T_j^i, we now have

$$
T^1 + T^2 =_S T_1^1 + (T_2^1 + T_2^2) + (T_3^1 + T_3^2) + T_1^2 =_S T_1^2.
$$

Finally, the change of indices $m = r - s - 1$ in the first term of T_1^2 leads to

$$
\begin{aligned}
T^1 + T^2 =_S T_1^2 &= -\sum_{m=0}^{r-2} \big(a e_{i_1} \chi^{2(r-m)-1}_{i_2\cdot\cdot i_{2(r-m)}}, \nabla_y \chi^{2m+1}_{i_{2(r-m)+1}\cdot\cdot i_{2r+1}}\big) + \sum_{s=1}^{r-1} \big(a \nabla_y \chi^{2s+1}_{i_2\cdot\cdot i_{2s+2}}, e_{i_1} \chi^{2(r-s)-1}_{i_{2s+3}\cdot\cdot i_{2r+1}}\big)\\
&= -\big(a e_{i_1} \chi^{2r-1}_{i_2\cdot\cdot i_{2r}}, \nabla_y \chi^{1}_{i_{2r+1}}\big) + \big(a \nabla_y \chi^{2r-1}_{i_2\cdot\cdot i_{2r}}, e_{i_1} \chi^{1}_{i_{2r+1}}\big),
\end{aligned}
$$

which proves the claim (5.31). Combining then (5.29), (5.30), and (5.31) gives the result of the lemma. □

Lemma 5.2.6. *Recall that we assume $|Y| = 1$. For $2 \le r \le \alpha/2$, the tensor*

$$
h^r_{i_1\cdot\cdot i_{2r+2}} = \big(a\big(\nabla_y \chi^{2r+1}_{i_2\cdot\cdot i_{2r+2}} + e_{i_2} \chi^{2r}_{i_3\cdot\cdot i_{2r+2}}\big), e_{i_1}\big),
$$

satisfies the decomposition

$$
h^r =_S (-1)^r k^r + \sum_{s=\lfloor\frac{r+1}{2}\rfloor}^{r-1} \sum_{\ell=1}^{\lfloor\frac{r}{2}\rfloor} p^{r,s,\ell} - \sum_{s=\lfloor\frac{r}{2}\rfloor+1}^{r} \sum_{\ell=1}^{\lfloor\frac{r+1}{2}\rfloor} q^{r,s,\ell},
$$

where $p^{r,s,\ell}$, $q^{r,s,\ell}$, and k^r are defined as

$$
\begin{aligned}
p^{r,s,\ell} &= (-1)^{s-\ell+1} \big\langle c^{s-\ell}\otimes\chi^{2\ell}\otimes\chi^{2(r-s)}\big\rangle_Y,\\
q^{r,s,\ell} &= (-1)^{s-\ell+1} \big\langle c^{s-\ell}\otimes\chi^{2\ell-1}\otimes\chi^{2(r-s)+1}\big\rangle_Y,\\
k^r_{i_1\cdot\cdot i_{2r+2}} &= -\Big(a\nabla_y \chi^{r+1}_{i_1\cdot\cdot i_{r+1}}, \nabla_y \chi^{r+1}_{i_{r+2}\cdot\cdot i_{2r+2}}\Big) + \Big(a e_{i_2} \chi^{r}_{i_3\cdot\cdot i_{r+2}}, e_{i_1} \chi^{r}_{i_{r+3}\cdot\cdot i_{2r+2}}\Big).
\end{aligned}
\tag{5.32}
$$

Proof. Defining $\sigma^k = \begin{cases} (-1)^k & \text{if } k \le r+1 \\ (-1)^{k+1} & \text{if } k > r+1 \end{cases}$, we verify thanks to the symmetry of a that

$$
T = \sum_{k=1}^{2r+1} \sigma^k \big(a\nabla_y \chi^{k}_{i_1\cdot\cdot i_k}, \nabla_y \chi^{2r+2-k}_{i_{k+1}\cdot\cdot i_{2r+2}}\big) =_S (-1)^{r+1} \big(a\nabla_y \chi^{r+1}_{i_1\cdot\cdot i_{r+1}}, \nabla_y \chi^{r+1}_{i_{r+2}\cdot\cdot i_{2r+2}}\big). \tag{5.33}
$$

We write T as

$$T = -\big(a\nabla_y\chi^1_{i_1}, \nabla_y\chi^{2r+1}_{i_2\cdot\cdot i_{2r+2}}\big) + \big(a\nabla_y\chi^2_{i_1i_2}, \nabla_y\chi^{2r}_{i_3\cdot\cdot i_{2r+2}}\big) + T^1 + T^2,$$

where, if $r = 1$, $T^1 = \sigma^{2r+1}\big(a\nabla_y\chi^{2r+1}_{i_1\cdot\cdot i_{2r+1}}, \nabla_y\chi^1_{i_{2r+2}}\big)$ and $T^2 = 0$ and, if $r \geq 2$, T^1 and T^2 are the sums over the odd and the even indices, respectively:

$$T^1 = \sum_{s=1}^{r} \sigma^{2s+1}\big(a\nabla_y\chi^{2s+1}_{i_1\cdot\cdot i_{2s+1}}, \nabla_y\chi^{2(r-s)+1}_{i_{2s+2}\cdot\cdot i_{2r+2}}\big), \qquad T^2 = \sum_{s=1}^{r-1} \sigma^{2s+2}\big(a\nabla_y\chi^{2s+2}_{i_1\cdot\cdot i_{2s+2}}, \nabla_y\chi^{2(r-s)}_{i_{2s+3}\cdot\cdot i_{2r+2}}\big).$$

Using (5.22a), (5.22b), and the symmetry of a, we find that

$$\begin{aligned} T =_S & \big(a(\nabla_y\chi^{2r+1}_{i_2\cdot\cdot i_{2r+2}} + e_{i_2}\chi^{2r}_{i_3\cdot\cdot i_{2r+2}}), e_{i_1}\big) \\ & - \big(ae_{i_1}\chi^1_{i_2}, \nabla_y\chi^{2r}_{i_3\cdot\cdot i_{2r+2}}\big) + \big(a\nabla_y\chi^1_{i_2}, e_{i_1}\chi^{2r}_{i_3\cdot\cdot i_{2r+2}}\big) + T^1 + T^2. \end{aligned} \tag{5.34}$$

Using (5.22c) and (5.22d), we write T^1 and T^2 as $T^i =_S T^i_1 + T^i_2 + T^i_3$, where

$$\begin{aligned} T^1_1 =_S & -\sum_{s=1}^{r} \sigma^{2s+1}\big(ae_{i_1}\chi^{2s}_{i_2\cdot\cdot i_{2s+1}}, \nabla_y\chi^{2(r-s)+1}_{i_{2s+2}\cdot\cdot i_{2r+2}}\big) + \sum_{s=1}^{r} \sigma^{2s+1}\big(a\nabla_y\chi^{2s}_{i_2\cdot\cdot i_{2s+1}}, e_{i_1}\chi^{2(r-s)+1}_{i_{2s+2}\cdot\cdot i_{2r+2}}\big), \\ T^1_2 =_S & \sum_{s=1}^{r} \sigma^{2s+1}\big(ae_{i_2}\chi^{2s-1}_{i_3\cdot\cdot i_{2s+1}}, e_{i_1}\chi^{2(r-s)+1}_{i_{2s+2}\cdot\cdot i_{2r+2}}\big), \qquad T^1_3 =_S \sum_{s=1}^{r} \sigma^{2s+1} \sum_{\ell=1}^{s} q^{r,s,\ell}_{i_1\cdot\cdot i_{2r+2}}, \\ T^2_1 =_S & -\sum_{s=1}^{r-1} \sigma^{2s+2}\big(ae_{i_1}\chi^{2s+1}_{i_2\cdot\cdot i_{2s+3}}, \nabla_y\chi^{2(r-s)}_{i_{2s+3}\cdot\cdot i_{2r+2}}\big) + \sum_{s=1}^{r-1} \sigma^{2s+2}\big(a\nabla_y\chi^{2s+1}_{i_2\cdot\cdot i_{2s+2}}, e_{i_1}\chi^{2(r-s)}_{i_{2s+3}\cdot\cdot i_{2r+2}}\big), \\ T^2_2 =_S & \sum_{s=1}^{r-1} \sigma^{2s+2}\big(ae_{i_2}\chi^{2s}_{i_3\cdot\cdot i_{2s+2}}, e_{i_1}\chi^{2(r-s)}_{i_{2s+3}\cdot\cdot i_{2r+2}}\big), \qquad T^2_3 =_S \sum_{s=1}^{r-1} \sigma^{2s+2} \sum_{\ell=1}^{s} p^{r,s,\ell}_{i_1\cdot\cdot i_{2r+2}}, \end{aligned}$$

where $q^{r,s,\ell}$ and $p^{r,s,\ell}$ are defined in (5.32). We claim that the following equalities hold:

$$T^1_1 + T^2_1 =_S \big(ae_{i_1}\chi^1_{i_2}, \nabla_y\chi^{2r}_{i_3\cdot\cdot i_{2r+2}}\big) - \big(a\nabla_y\chi^1_{i_2}, e_{i_1}\chi^{2r}_{i_3\cdot\cdot i_{2r+2}}\big), \tag{5.35}$$

$$T^1_2 + T^2_2 =_S (-1)^{r+1}\big(ae_{i_2}\chi^r_{i_3\cdot\cdot i_{r+2}}, e_{i_1}\chi^r_{i_{r+3}\cdot\cdot i_{2r+2}}\big), \tag{5.36}$$

$$T^1_3 =_S \sum_{s=\lfloor\frac{r}{2}\rfloor+1}^{r} \sum_{\ell=1}^{\lfloor\frac{r+1}{2}\rfloor} q^{r,s,\ell}_{i_1\cdot\cdot i_{2r+2}}, \tag{5.37}$$

$$T^2_3 =_S -\sum_{s=\lfloor\frac{r+1}{2}\rfloor}^{r-1} \sum_{\ell=1}^{\lfloor\frac{r}{2}\rfloor} p^{r,s,\ell}_{i_1\cdot\cdot i_{2r+2}}. \tag{5.38}$$

Let us first prove (5.35). In T^1_1, we separate the terms $s = r$ in both sums and then make the change of indices $m = r - s$ in the remaining sums. Summing with T^2_1, we find

$$\begin{aligned} T^1_1 + T^2_1 =_S & -\sigma^{2r+1}\big(ae_{i_1}\chi^{2r}_{i_2\cdot\cdot i_{2r+1}}, \nabla_y\chi^1_{i_{2r+2}}\big) + \sigma^{2r+1}\big(a\nabla_y\chi^{2r}_{i_2\cdot\cdot i_{2r+1}}, e_{i_1}\chi^1_{i_{2r+2}}\big) \\ & + \sum_{m=1}^{r-1} (\sigma^{2m+2} - \sigma^{2(r-m)+1})\big(ae_{i_1}\chi^{2m+1}_{i_2\cdot\cdot i_{2m+3}}, \nabla_y\chi^{2(r-m)}_{i_{2m+3}\cdot\cdot i_{2r+2}}\big) \\ & + \sum_{s=1}^{r-1} (\sigma^{2(r-m)+1} - \sigma^{2m+2})\big(a\nabla_y\chi^{2m+1}_{i_2\cdot\cdot i_{2m+2}}, e_{i_1}\chi^{2(r-m)}_{i_{2m+3}\cdot\cdot i_{2r+2}}\big). \end{aligned}$$

As $\sigma^{2r+1} = (-1)^{2(r+1)} = 1$ and

$$\sigma^{2m+2} = \sigma^{2(r-m)+1} = \begin{cases} 1 & \text{if } m \leq \lfloor\frac{r-1}{2}\rfloor, \\ -1 & \text{if } m > \lfloor\frac{r-1}{2}\rfloor, \end{cases}$$

(5.35) is proved. Let us now prove (5.36). Studying the signs of σ^{2s+1} and σ^{2s+2}, we write $T_2^i = T_{21}^i + T_{22}^i$, where

$$T_{21}^1 = -\sum_{s=1}^{\lfloor \frac{r}{2} \rfloor} \big(a e_{i_2} \chi^{2s-1}_{i_3 \cdot\cdot i_{2s+1}}, e_{i_1} \chi^{2(r-s)+1}_{i_{2s+2} \cdot\cdot i_{2r+2}}\big), \qquad T_{22}^1 = \sum_{s=\lfloor \frac{r}{2} \rfloor+1}^{r} \big(a e_{i_2} \chi^{2s-1}_{i_3 \cdot\cdot i_{2s+1}}, e_{i_1} \chi^{2(r-s)+1}_{i_{2s+2} \cdot\cdot i_{2r+2}}\big),$$

$$T_{21}^2 = \sum_{s=1}^{\lfloor \frac{r-1}{2} \rfloor} \big(a e_{i_2} \chi^{2s}_{i_3 \cdot\cdot i_{2s+2}}, e_{i_1} \chi^{2(r-s)}_{i_{2s+3} \cdot\cdot i_{2r+2}}\big), \qquad T_{22}^2 = -\sum_{s=\lfloor \frac{r-1}{2} \rfloor+1}^{r} \big(a e_{i_2} \chi^{2s}_{i_3 \cdot\cdot i_{2s+2}}, e_{i_1} \chi^{2(r-s)}_{i_{2s+3} \cdot\cdot i_{2r+2}}\big).$$

Making the change of indices $m = r - s + 1$ in T_{21}^1 and $m = r - s$ in T_{21}^2, we find

$$T_{21}^1 = -\sum_{m=\lfloor \frac{r+1}{2} \rfloor+1}^{r} \big(a e_{i_2} \chi^{2(r-m)+1}_{i_3 \cdot\cdot i_{2(r-m)+3}}, e_{i_1} \chi^{2m-1}_{i_{2(r-m)+4} \cdot\cdot i_{2r+2}}\big),$$

$$T_{21}^2 = \sum_{m=\lfloor \frac{r}{2} \rfloor+1}^{r} \big(a e_{i_2} \chi^{2(r-m)}_{i_3 \cdot\cdot i_{2(r-m)+2}}, e_{i_1} \chi^{2m}_{i_{2(r-m)+3} \cdot\cdot i_{2r+2}}\big).$$

Assume first that r is even. In this case, $\lfloor \frac{r}{2} \rfloor = \lfloor \frac{r+1}{2} \rfloor$ and that implies $T_2^1 =_S 0$. Furthermore, as $\lfloor \frac{r-1}{2} \rfloor + 1 = \lfloor \frac{r}{2} \rfloor$, we obtain $T_2^2 =_S -\big(a e_{i_2} \chi^{r}_{i_3 \cdot\cdot i_{r+2}}, e_{i_1} \chi^{r}_{i_{r+3} \cdot\cdot i_{2r+2}}\big)$, which implies (5.36) in the case where r is even. Assume then that r is odd. In this case, we verify that $T_2^2 =_S 0$ and $T_2^1 =_S \big(a e_{i_2} \chi^{r}_{i_3 \cdot\cdot i_{r+2}}, e_{i_1} \chi^{r}_{i_{r+3} \cdot\cdot i_{2r+2}}\big)$, and that concludes the proof of (5.36). Next, we prove (5.37). Let us write T_3^1 as $T_3^1 = \big\langle (\tilde{T}_{31}^1 + \tilde{T}_{32}^1)_{i_1 \cdot\cdot i_{2r+2}} \big\rangle_Y$, where

$$\tilde{T}_{31}^1 = -\sum_{s=1}^{\lfloor \frac{r}{2} \rfloor} \sum_{\ell=1}^{s} q^{r,s,\ell}, \qquad \tilde{T}_{32}^1 = \sum_{s=\lfloor \frac{r}{2} \rfloor+1}^{r} \sum_{\ell=1}^{s} q^{r,s,\ell},$$

and $q^{r,s,\ell}$ is defined in (5.32). Making the change of indices $m = r - s + 1$, exchanging the sums, changing the indices $k = r - \ell + 1$, and using the equality $q^{r-m+1,r-k+1} =_S q^{k,m}$, we rewrite $\tilde{T}_{31}^1$ as

$$\tilde{T}_{31}^1 = -\sum_{m=\lfloor \frac{r+1}{2} \rfloor+1}^{r} \sum_{\ell=1}^{r-m+1} q^{r-m+1,\ell} = -\sum_{\ell=1}^{\lfloor \frac{r}{2} \rfloor} \sum_{m=\lfloor \frac{r+1}{2} \rfloor+1}^{r-\ell+1} q^{r-m+1,\ell} =_S -\sum_{k=\lfloor \frac{r+1}{2} \rfloor+1}^{r} \sum_{m=\lfloor \frac{r+1}{2} \rfloor+1}^{k} q^{k,m}.$$

If r is even, we verify that $\lfloor \frac{r}{2} \rfloor = \lfloor \frac{r+1}{2} \rfloor$ and summing $\tilde{T}_{31}^1 + \tilde{T}_{32}^1$, we obtain (5.37). Similarly, if r is odd, we verify that $\lfloor \frac{r}{2} \rfloor + 1 = \lfloor \frac{r+1}{2} \rfloor$ and summing $\tilde{T}_{31}^1 + \tilde{T}_{32}^1$ gives (5.37). Finally, we prove (5.38). Let us write T_3^2 as $T_3^2 = \big\langle (\tilde{T}_{31}^2 + \tilde{T}_{32}^2)_{i_1 \cdot\cdot i_{2r+2}} \big\rangle_Y$, where

$$\tilde{T}_{31}^2 = \sum_{s=1}^{\lfloor \frac{r-1}{2} \rfloor} \sum_{\ell=1}^{s} p^{r,s,\ell}, \qquad \tilde{T}_{32}^2 = -\sum_{s=\lfloor \frac{r-1}{2} \rfloor+1}^{r} \sum_{\ell=1}^{s} p^{r,s,\ell},$$

and $p^{r,s,\ell}$ is defined in (5.32). Making the change of indices $m = r - s$, exchanging the sums, again changing $k = r - \ell$, and using that $p^{r-m,r-k} =_S p^{k,m}$, we rewrite $\tilde{T}_{31}^2$ as

$$\tilde{T}_{31}^2 = \sum_{m=\lfloor \frac{r}{2} \rfloor+1}^{r-1} \sum_{\ell=1}^{r-m} p^{r-m,\ell} = \sum_{\ell=1}^{\lfloor \frac{r-1}{2} \rfloor} \sum_{m=\lfloor \frac{r}{2} \rfloor+1}^{r-\ell} p^{r-m,\ell} = \sum_{k=\lfloor \frac{r}{2} \rfloor+1}^{r-1} \sum_{m=\lfloor \frac{r}{2} \rfloor+1}^{k} p^{k,m}.$$

Studying the parities of r separately, we sum $\tilde{T}_{31}^2 + \tilde{T}_{32}^2$ and obtain (5.38). Combining now (5.33), (5.34), and (5.35), we obtain

$$\big(a\big(\nabla_y \chi^{2r+1}_{i_2 \cdot\cdot i_{2r+2}} + e_{i_2} \chi^{2r}_{i_3 \cdot\cdot i_{2r+2}}\big), e_{i_1}\big) =_S (-1)^{r+1} \big(a \nabla_y \chi^{r+1}_{i_1 \cdot\cdot i_{r+1}}, \nabla_y \chi^{r+1}_{i_{r+2} \cdot\cdot i_{2r+2}}\big) - (T_2^1 + T_2^2 + T_3^1 + T_3^2).$$

This equality, combined with (5.36), (5.37), and (5.38) proves the lemma. □

5.2.3 Existence of effective equations and matrix associated to a symmetric tensor of even order

Recall that the family of effective equations for arbitrary timescales $\mathcal{E}$ is defined implicitly by constraints on the tensors $\{(a^{2r}, b^{2r})\}_{r=1}^{\lfloor \alpha/2 \rfloor}$ (Definition 5.2.4). However, these constraints do not provide a way to compute the tensors explicitly. Furthermore, we have yet no guaranty of the existence of an equation in the family $\mathcal{E}$. In this section, we prove that the family $\mathcal{E}$ is not empty. Furthermore, we describe one possible construction of effective tensors. In particular, we generalize the procedure used for fourth order tensors, presented in Section 4.3.2.

Let us recall the definitions of positive (semi)definite tensors of even order in (5.16) and (5.17). A tensor $q \in \mathrm{Ten}^{2n}(\mathbb{R}^d)$ is said positive semidefinite if

$$q_{i_1 \cdots i_{2n}} \xi_{i_1 \cdots i_n} \xi_{i_{n+1} \cdots i_{2n}} \geq 0 \quad \forall \xi \in \mathrm{Sym}^n(\mathbb{R}^d). \tag{5.39}$$

It is positive definite if the inequality in (5.39) is strict for all $\xi \in \mathrm{Sym}^n(\mathbb{R}^d) \backslash \{0\}$.

In a first time, we prove two results on the sign of even order tensors. The first is a generalization of Lemma 4.3.1, while the second generalizes Lemma 4.3.3 and ensures that the tensor $S^{2n}(\otimes^n a^0)$ is positive definite.

Lemma 5.2.7. *Let $R \in \mathrm{Ten}^{2n}(\mathbb{R}^d)$ be a positive definite tensor and let $A \in \mathrm{Sym}^2(\mathbb{R}^d)$ be a symmetric, positive definite matrix. Then the tensor of $\mathrm{Ten}^{2n+2}(\mathbb{R}^d)$ defined by $A_{i_1 i_{2n+2}} R_{i_2 \cdots i_{2n+1}}$ is positive definite.*

Proof. As A is symmetric positive definite, the Cholesky factorization gives an invertible H such that $A = H^T H$. For $\xi \in \mathrm{Sym}^{n+1}(\mathbb{R}^d)$, we thus have

$$A_{i_1 i_{2n+2}} R_{i_2 \cdots i_{2n+1}} \xi_{i_1 \cdots i_{n+1}} \xi_{i_{n+1} \cdots i_{2n+2}} = R_{i_2 \cdots i_{2n+1}} \big(H_{rj} \xi_{j i_2 \cdots i_{n+1}}\big) \big(H_{rj} \xi_{j i_{n+2} \cdots i_{2n+2}}\big) \geq 0. \tag{5.40}$$

As R is positive definite, the equality holds if and only if $H_{rj} \xi_{j i_2 \cdots i_{n+1}} = 0$ for all $r, i_2, \ldots, i_{n+1} \in \{1, \ldots, d\}$. Let $i_2, \cdots, i_{n+1}$ be arbitrarily fixed and denote $v_j = \xi_{j i_2 \cdots i_{n+1}}$. Hence, we have $H_{rj} v_j = 0$ for all r, which is equivalent to $H^T v = 0$. As H is regular, so is H^T, and thus $v = 0$. We have proved that if the equality holds in (5.40) then $\xi = 0$. Hence the tensor is positive definite and the proof of the lemma is complete. □

Lemma 5.2.8. *If $A \in \mathrm{Sym}^2(\mathbb{R}^d)$ is a symmetric, positive definite matrix, then the tensor $S^{2n}(\otimes^n A) \in \mathrm{Sym}^{2n}(\mathbb{R}^d)$ is positive definite.*

Proof. We proceed by induction. The case $n = 1$ is proved by Lemma 4.3.3. We assume that the result holds for $1, \ldots, n-1$ and prove it for n. Let $\xi \in \mathrm{Sym}^n(\mathbb{R}^d) \backslash \{0\}$. First, assume that n is odd. Then, the product $S^{2n}(\otimes^n A)\xi : \xi$ is composed of terms of the form

$$A_{jk} A_{i_1 i_2} \cdots A_{i_{2n-3} i_{2n-2}} \xi_{j i_1 \cdots i_{n-1}} \xi_{k i_n \cdots i_{2n-2}}, \tag{5.41}$$

i.e., one of the factor $A_{i_r i_s}$ share indices with both ξ. Thanks to Lemma 5.2.7, the induction assumption ensures that all these terms are strictly positive and thus $S^{2n}(\otimes^n A)$ is positive definite. Second, we assume that n is even. Then, the product $S^{2n}(\otimes^n A)\xi : \xi$ is composed of terms of two forms. First, there are terms of the form (5.41). By the same induction argument as before, they are strictly positive. Second, terms of the form

$$A_{i_1 i_2} \cdots A_{i_{n-1} i_n} A_{i_{n+1} i_{n+2}} \cdots A_{i_{2n-1} i_{2n}} \xi_{i_1 \cdots i_n} \xi_{i_{n+1} \cdots i_{2n}} = \big(A_{i_1 i_2} \cdots A_{i_{n-1} i_n} \xi_{i_1 \cdots i_n}\big)^2 \geq 0.$$

Altogether, we verify that $S^{2n}(\otimes^n A)\xi : \xi > 0$ and the proof of the lemma is complete. □

Let us now prove that the family $\mathcal{E}$, defined in Definition 5.2.4, is not empty. Recall that the existence of valid pairs a^2, b^2 was already proved in Section 4.3.2. We thus need to find pairs of positive semidefinite, major symmetric tensors $\{a^{2r}, b^{2r}\}_{r=2}^{\alpha/2}$ that satisfy the constraints (5.27), given by

$$a^{2r} - b^{2r} \otimes a^0 =_S (-1)^r h^r + \sum_{\ell=1}^{r-1} b^{2(r-\ell)} \otimes c^\ell =: \check{q}^r, \tag{5.42}$$

where h^r and c^r are the tensors defined in (5.24) and (5.21), respectively. The existence of such pairs is guaranteed by Lemma 5.2.7. Indeed, it ensures that if $R \in \mathrm{Sym}^{2r}(\mathbb{R}^d)$ is a "sufficiently large" positive definite tensor, then the pair

$$a^{2r} = S^{2r+2}\big(\check{q}^r + R \otimes a^0\big), \qquad b^{2r} = S^{2r}\big(R\big), \tag{5.43}$$

define an effective equation in the family $\mathcal{E}$. Indeed, these tensors are positive semidefinite by construction. Furthermore, they are also major symmetric and satisfy the constraint (5.42).

We now need a practical way to construct the tensor R in (5.43). To that purpose, for a given symmetric tensor of even order q, we construct a matrix $M(q)$, whose spectrum is connected to the sign of q. The construction is similar to what was done in Section 4.3.3 for fourth order major symmetric tensors. We consider the bilinear map

$$\mathrm{Sym}^n(\mathbb{R}^d) \times \mathrm{Sym}^n(\mathbb{R}^d) \to \mathbb{R}, \qquad (\xi, \eta) \mapsto q\xi : \eta = q_{i_1 \cdots i_n i_{n+1} \cdots i_{2n}} \xi_{i_1 \cdots i_n} \eta_{i_{n+1} \cdots i_{2n}}. \tag{5.44}$$

Denote $I(d,n)$ the set of multiindices of the distinct entries of a tensor in $\mathrm{Sym}^n(\mathbb{R}^d)$, i.e.,

$$I(d,n) = \big\{ i = (i_1, \ldots, i_n) : 1 \le i_1 \le \ldots \le i_n \le d \big\}.$$

We verify that the cardinality of $I(d,n)$ is given by $N(d,n) = |I(d,n)| = \binom{d+n-1}{n}$. Denote then $J(d,n) = \{1, \ldots, N(d,n)\}$ and let $\ell : J(d,n) \to I(d,n)$ be a bijection. We define then the bijective mapping

$$\nu : \mathrm{Sym}^n(\mathbb{R}^d) \to \mathbb{R}^{N(d,n)}, \qquad \xi \mapsto \nu(\xi), \quad \big(\nu(\xi)\big)_m = \xi_{\ell(m)} \quad m \in J(d,n).$$

For $i \in I(d,n)$, let $z(i)$ be the number of multiindices in $\{1, \ldots, d\}^n$ that are equivalent to i up to symmetries, i.e.,

$$z(i) = \big|\{ j \in \{1, \ldots d\}^n : \text{ there exists a permutation } \sigma \text{ s.t. } \sigma(j) = i \}\big|.$$

With these notations, we rewrite the map defined in (5.44) as

$$q\xi : \eta = \sum_{i,j \in I(d,n)} z(i) z(j) q_{ij} \xi_i \eta_j = \sum_{n,m=1}^{N(d,n)} z\big(\ell(n)\big) z\big(\ell(m)\big) q_{\ell(n)\ell(m)} \xi_{\ell(n)} \eta_{\ell(m)}.$$

Defining then the matrix associated to a tensor as

$$\begin{aligned} M : \mathrm{Sym}^{2n}(\mathbb{R}^d) &\to \mathrm{Sym}^2(\mathbb{R}^{N(d,n)}), \\ q &\mapsto M(q) \quad \big(M(q)\big)_{mn} = z\big(\ell(n)\big) z\big(\ell(m)\big) q_{\ell(n)\ell(m)} \quad m,n \in J(d,n), \end{aligned} \tag{5.45}$$

we verify that $q\xi : \eta = M(q)\nu(\xi) \cdot \nu(\eta)$. Hence, q is positive definite (resp. semidefinite) if and only if $M(q)$ is positive definite (resp. semidefinite).

We prove the following lemma.

Lemma 5.2.9. *Let $\check{q} \in \mathrm{Sym}^{2n}(\mathbb{R}^d)$ and define the matrices $Q = M(\check{q})$ and $A = M\big(S^{2n}(\otimes^n a^0)\big)$. Then the tensor*

$$q = \check{q} + \delta S^{2n}(\otimes^n a^0), \qquad \delta \geq \delta^* = \left\{ -\frac{\lambda_{\min}\big(Q\big)}{\lambda_{\min}(A)} \right\}_+,$$

is positive semidefinite, where we denoted $\lambda_{\min}(\cdot)$ the minimal eigenvalue of the matrices.

Proof. First, as Q and A are symmetric matrices by definition, $\lambda_{\min}(Q)$ and $\lambda_{\min}(A)$ are real. Next, thanks to Lemma 5.2.8 $S^{2n}(\otimes^n a^0)$ is positive definite and thus $\lambda_{\min}(A) > 0$. Note that $\lambda_{\min}(Q) \leq (Qv \cdot v)/(v \cdot v)$ for any $v \in \mathbb{R}^{N(d,n)}$ and similarly for A. If Q is positive semidefinite, then $\delta^* = 0$ and the tensor q is positive semidefinite for any $\delta \geq \delta^* = 0$. Assuming then that $\lambda_{\min}(Q) < 0$, we verify that for any $v \in \mathbb{R}^{N(d,n)}$,

$$\delta^* = -\frac{\lambda_{\min}(Q)}{\lambda_{\min}(A)} \geq -\frac{Qv \cdot v}{Av \cdot v}.$$

Hence, writing $\delta = \delta^* + \Delta\delta$ with $\Delta\delta \geq 0$ and denoting $v_\xi = \nu(\xi)$, we have

$$q\xi : \xi = Qv_\xi \cdot v_\xi + \delta^* A v_\xi \cdot v_\xi + \Delta\delta A v_\xi \cdot v_\xi \geq 0 \quad \forall \xi \in \mathrm{Sym}^n(\mathbb{R}^d),$$

and the proof of the lemma is complete. □

Thanks to Lemma 5.2.9, we can complete the construction given in (5.43) by setting $R = \delta^* S^{2r}(\otimes^r a^0)$, where

$$\delta^* = \left\{ -\frac{\lambda_{\min}\big(Q\big)}{\lambda_{\min}(A} \right\}_+, \quad Q = M(\check{q}^r), \quad A = M\big(S^{2r+2}(\otimes^{r+1} a^0)\big).$$

This process is used to compute the pairs $\{a^{2r}, b^{2r}\}$ for every $r \geq 1$ in the next section (Algorithm 5.2.10).

5.2.4 Algorithm for the computation of the tensors of an effective equation

In this section, we provide a numerical procedure for the computation of the effective tensors of an effective equation that belongs to the family $\mathcal{E}$, defined in Definition 5.2.4. Note that the procedure relies on the results of the previous section, where we presented a process to increase the sign of symmetric tensors.

The numerical procedure is given in Algorithm 5.2.10. Let us discuss it. From line 1 to line 8, we recognize Algorithm 4.3.7 for the computation of the tensors a^2, b^2 (Section 4.3.4). However, note that in Algorithm 4.3.7, (5.43) was applied with $R = \delta S^{2,2}_{ij,kl}\{a^0_{jk} I_{il}\}$, while in Algorithm 5.2.10, we use $R = \delta S^4(a^0 \otimes a^0)$. This alternative was discussed in Section 4.3.2). Next, let us verify that the tensors a^{2r}, b^{2r}, defined in lines 22 and 23, satisfy the constraint (5.26) characterizing the family $\mathcal{E}$. Indeed, we have

$$a^{2r} - b^{2r} \otimes a^0 =_S \check{a}^{2r} - \big\langle a^0 \otimes \chi^r \otimes \chi^r \big\rangle_Y =_S k^r + (-1)^r \sum_{s=\lfloor \frac{r+1}{2} \rfloor}^{r-1} \sum_{\ell=1}^{\lfloor \frac{r}{2} \rfloor} p^{r,s,\ell} + (-1)^{r+1} \sum_{s=\lfloor \frac{r}{2} \rfloor + 1}^{r} \sum_{\ell=1}^{\lfloor \frac{r+1}{2} \rfloor} q^{r,s,\ell}.$$

As Lemma 5.2.6 ensures that the right hand side equals $(-1)^r h^r$, a^{2r}, b^{2r} satisfy (5.26). Finally, observe that

$$-\big\langle a^0 \otimes \chi^r \otimes \chi^r \big\rangle_Y = \begin{cases} (-1)^r p^{r,r/2,r/2} & \text{if } r \text{ is even,} \\ (-1)^{r+1} q^{r,(r+1)/2,(r+1)/2} & \text{if } r \text{ is odd.} \end{cases} \tag{5.46}$$

Thus, the term $\langle a^0 \otimes \chi^r \otimes \chi^r \rangle_Y$ in line 18 cancels one term in one of the preceding double sums.

It is interesting to note that the construction to obtain positive semidefinite tensors is also necessary in one dimension. Indeed, recall that in one dimension, for $r = 1$ it holds $\check{a}^2 = 0$ and thus $a^2 = 0$, $b^2 = \langle (\chi^1)^2 \rangle_Y$ (see Section 4.3.1). However, for $r \geq 2$, the algorithm defines

$$\left(a^{2r}, b^{2r}\right) = \begin{cases} \left(0, \langle (\chi^r)^2 \rangle_Y + |\check{a}^{2r}|\right) & \text{if } \check{a}^{2r} \leq 0, \\ \left(\check{a}^{2r}, \langle (\chi^r)^2 \rangle_Y\right) & \text{if } \check{a}^{2r} > 0. \end{cases}$$

As the sign of $\check{a}^{2r}$ is unknown a priori, this process is needed to guarantee the well-posedness of the corresponding effective equation.

Let us discuss the complexity of Algorithm 5.2.10. Let then $\mathrm{CP}(d, k)$ be the total number of cell problems to solve to obtain χ^1 to χ^k. As χ^r has $N(d, r) = |I(d, r)| = \binom{d+r-1}{r}$ entries, we calculate

$$\mathrm{CP}(d, k) = \sum_{r=1}^{k} N(d, r) = \sum_{r=1}^{k} \binom{r + d - 1}{r} = \binom{k + d}{d} - 1.$$

The cost of Algorithm 5.2.10 is thus $\mathrm{CP}(d, \lfloor \alpha/2 \rfloor + 1)$. Note that without Lemma 5.2.6, computing the tensors requires χ^1 to $\chi^{2\lfloor \alpha/2 \rfloor + 1}$, i.e., the cost would be $\mathrm{CP}(d, 2\lfloor \alpha/2 \rfloor + 1)$. To fully appreciate the gain obtained thanks to Lemma 5.2.6, let us compare the corresponding costs for the computation of the effective tensors for a timescale $\mathcal{O}(\varepsilon^{-6})$, i.e. $\lfloor \alpha/2 \rfloor = 3$. If $d = 2$, only 14 cell problems need to be solved thanks to Lemma 5.2.6 instead of 35. If $d = 3$, 34 cell problems are sufficient thanks to Lemma 5.2.6 instead of 119.

Algorithm 5.2.10 Compute the tensors of an effective equation (5.19). Note that the matrix construction $M(\cdot)$ is defined in (5.45).

Input : tensor a, timescale α.

Output : effective tensors $a^0, \{a^{2r}, b^{2r}\}_{r=1}^{\lfloor \alpha/2 \rfloor}$.

for all $i \in I(d,1)$ $\quad \chi_i^1 \leftarrow$ solve (5.22a) with $\langle \chi_i^1 \rangle_Y = 0$

for all $i \in I(d,2)$ $\quad a^0_{i_1 i_2} = -\langle a \nabla_y \chi^1_{i_2} \cdot \nabla_y \chi^1_{i_1} \rangle_Y + \langle a e_{i_2} \cdot e_{i_1} \rangle_Y$

for all $i \in I(d,2)$ $\quad \chi^2_{i_1 i_2} \leftarrow$ solve (5.22b) with $\langle \chi^2_{i_1 i_2} \rangle_Y = 0$

for all $i \in I(d,4)$ $\quad \check{a}^2_{i_1 \cdots i_4} = S^4_{i_1 \cdots i_4} \big\{ - \langle a \nabla_y \chi^2_{i_1 i_2} \cdot \nabla_y \chi^2_{i_3 i_4} \rangle_Y + \langle a e_{i_2} \chi^1_{i_3} \cdot e_{i_1} \chi^1_{i_4} \rangle_Y \big\}$

$A^2 = M(\check{a}^2), \quad A^0 = M\big(S^4(a^0 \otimes a^0)\big)$

$\delta^* = \left\{ - \dfrac{\lambda_{\min}(A^2)}{\lambda_{\min}(A^0)} \right\}_+$

$a^2 = S^4\big(\check{a}^2 + \delta^* a^0 \otimes a^0\big)$

$b^2 = \langle \chi^1 \otimes \chi^1 \rangle_Y + \delta^* a^0$

$c^0 = a^0, \quad c^1 = a^2 - b^2 \otimes a^0$

for $r = 2, \ldots, \lfloor \alpha/2 \rfloor$ **do**

  for all $i \in I(d, r+1)$ $\quad \chi^{r+1}_{i_1 \cdots i_{r+1}} \leftarrow$ solve (5.22c) or (5.22d) with $\langle \chi^{r+1}_{i_1 \cdots i_{r+1}} \rangle_Y = 0$

  for all $i \in I(d, 2r+2)$

    $k^r_{i_1 \cdots i_{2r+2}} = -\langle a \nabla_y \chi^{r+1}_{i_1 \cdots i_{r+1}} \cdot \nabla_y \chi^{r+1}_{i_{r+2} \cdots i_{2r+2}} \rangle_Y + \langle a e_{i_2} \chi^r_{i_3 \cdots i_{r+1}} \cdot e_{i_1} \chi^r_{i_{r+2} \cdots i_{2r+2}} \rangle_Y$

  for $s = \lfloor \frac{r+1}{2} \rfloor, \ldots, r-1$ and $\ell = 1, \ldots, \lfloor \frac{r}{2} \rfloor$

    $p^{r,s,\ell} = (-1)^{s-\ell+1} \langle c^{s-\ell} \otimes \chi^{2\ell} \otimes \chi^{2(r-s)} \rangle_Y$

  for $s = \lfloor \frac{r}{2} \rfloor + 1, \ldots, r$ and $\ell = 1, \ldots, \lfloor \frac{r+1}{2} \rfloor$

    $q^{r,s,\ell} = (-1)^{s-\ell+1} \langle c^{s-\ell} \otimes \chi^{2\ell-1} \otimes \chi^{2(r-s)+1} \rangle_Y$

  $$\check{a}^{2r} = S^{2r+2}\Bigg(k^r + (-1)^r \sum_{s=\lfloor \frac{r+1}{2} \rfloor}^{r-1} \sum_{\ell=1}^{\lfloor \frac{r}{2} \rfloor} p^{r,s,\ell} + (-1)^{r+1} \sum_{s=\lfloor \frac{r}{2} \rfloor + 1}^{r} \sum_{\ell=1}^{\lfloor \frac{r+1}{2} \rfloor} q^{r,s,\ell} + \langle a^0 \otimes \chi^r \otimes \chi^r \rangle_Y + \sum_{\ell=1}^{r-1} b^{2(r-\ell)} \otimes c^\ell \Bigg)$$

  $A^{2r} = M(\check{a}^{2r}), \quad A^0 = M\big(S^{2r+2}(\otimes^{r+1} a^0)\big)$

  $\delta^* = \left\{ - \dfrac{\lambda_{\min}(A^{2r})}{\lambda_{\min}(A^0)} \right\}_+$

  $a^{2r} = \check{a}^{2r} + \delta^* S^{2r+2}(\otimes^{r+1} a^0)$

  $b^{2r} = S^{2r}\big(\langle \chi^r \otimes \chi^r \rangle + \delta^* \otimes^r a^0\big)$

  $c^r = a^{2r} - \sum_{\ell=0}^{r-1} b^{2(r-\ell)} \otimes c^\ell$

end for

5.2.5 Derivation of the cell problems of arbitrary order via asymptotic expansion

In this section, we proceed to the technical derivation of the cell problems defined in (5.22). Let us briefly recall how we proceed. First, we consider a candidate effective equation, whose solution is denoted $\tilde{u}$. As we know from Chapter 4, the form of this equation is of major importance. Indeed, if it is too restrictive, it may lead to ill-posed equations. We thus let the higher order differential operators be composed of pairs of operators: one with purely space derivatives and one with mixed space and time derivatives (see (5.47)). The main task is then the construction of an adaptation of $\tilde{u}$. In particular, the adaptation involves correctors that are solution of cell problems. To obtain the cell problems, we need to recursively apply Boussinesq tricks, i.e., use

the effective equation at first order to substitute the time derivatives in the expansion. This is a technical task whose result is given in Lemma 5.2.11.

Let us proceed to the derivation of the cell problems. For simplicity, let us consider an even timescale: $\mathcal{O}(\varepsilon^{-\alpha})$, with α even. We also let $f = 0$ (see Remark 5.2.12 for the case $f \neq 0$). Finally, recall that to simplify the notations we assume $|Y| = 1$. Let us first discuss the ansatz on the form of the effective equation. Let a^0 be the homogenized tensor and for $r = 1, \dots, \lfloor \alpha/2 \rfloor$, let $a^{2r} \in \mathrm{Ten}^{2r+2}(\mathbb{R}^d)$, $b^{2r} \in \mathrm{Ten}^{2r}(\mathbb{R}^d)$ be positive semidefinite major symmetric tensors (see (5.17) and (5.15)). Consider the following ansatz for the effective equation:

$$\begin{aligned} &\partial_t^2 \tilde{u} = a^0 \partial^2 \tilde{u} + \sum_{r=1}^{\alpha/2} (-1)^r \varepsilon^{2r} \big(a^{2r} \partial^{2r+2} \tilde{u} - b^{2r} \partial^{2r} \partial_t^2 \tilde{u} \big) && \text{in } (0, \varepsilon^{-\alpha} T] \times \Omega, \\ &x \mapsto \tilde{u}(t,x) \ \ \Omega\text{-periodic} && \text{in } [0, \varepsilon^{-\alpha} T], \\ &\tilde{u}(0,x) = g^0(x), \quad \partial_t \tilde{u}(0,x) = g^1(x) && \text{in } \Omega, \end{aligned} \tag{5.47}$$

where we recall the notation for the differential operators (5.18). Observe that thanks to the sign of the tensors, under sufficient regularity of g^0, g^1, (5.47) is well-posed. Indeed, its weak formulation is

$$\begin{aligned} \int_0^{\varepsilon^{-\alpha} T} \big(\tilde{u}(t), \partial_t^2 v(t) \big)_{\mathcal{H}} + A\big(\tilde{u}(t), v(t)\big) \, \mathrm{d}t = \int_0^{\varepsilon^{-\alpha} T} \big(f(t), v(t) \big)_{\mathrm{L}^2(\Omega)} \, \mathrm{d}t \\ + \big(g^1, v(0) \big)_{\mathcal{H}} - \big(g^0, \partial_t v(0) \big)_{\mathcal{H}}. \end{aligned}$$

for any test functions $v \in \mathcal{C}^2([0, \varepsilon^{-\alpha} T]; \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^{\alpha/2+1}(\Omega))$ with $v(\varepsilon^{-\alpha} T) = \partial_t v(\varepsilon^{-\alpha} T) = 0$, where the bilinear forms are defined as

$$\begin{aligned} \big(v, w\big)_{\mathcal{H}} &= \big(v, w\big)_{\mathrm{L}^2} + \sum_{r=1}^{\alpha/2} \varepsilon^{2r} \big(b^{2r}_{i_1 \cdots i_{2r}} \partial^r_{i_1 \cdots i_r} v, \partial^r_{i_{r+1} \cdots i_{2r}} w \big)_{\mathrm{L}^2}, \\ A\big(v, w\big) &= \big(a^0 \nabla v, \nabla w \big)_{\mathrm{L}^2} + \sum_{r=1}^{\alpha/2} \varepsilon^{2r} \big(a^{2r}_{i_1 \cdots i_{2r+2}} \partial^{r+1}_{i_1 \cdots i_{r+1}} v, \partial^{r+1}_{i_{r+2} \cdots i_{2r+2}} w \big)_{\mathrm{L}^2}. \end{aligned}$$

Thanks to the sign and major symmetry of the tensors, we can prove the existence and uniqueness of a weak solution $\tilde{u}$ of (5.48) (see also Section 2.1.2). We assume here that $\tilde{u}$ and its time derivatives are as smooth as required. Furthermore, we assume that the following quantities are bounded independently of ε

$$\textstyle\sum_{k=1}^K |\tilde{u}|_{\mathrm{L}^\infty(0, \varepsilon^{-\alpha} T; \mathrm{H}^k(\Omega))}, \sum_{k=1}^K |\partial_t^2 \tilde{u}|_{\mathrm{L}^\infty(0, \varepsilon^{-\alpha} T; \mathrm{H}^k(\Omega))} \leq C,$$

for a sufficiently large K. The ansatz on the adaptation of $\tilde{u}$ is

$$\mathcal{B}^\varepsilon \tilde{u}(t,x) = \tilde{u}(t,x) + \sum_{k=1}^{\alpha+2} \varepsilon^k u^k\big(t, x, \tfrac{x}{\varepsilon}\big) = \tilde{u}(t,x) + \sum_{k=1}^{\alpha+2} \varepsilon^k \chi^k\big(\tfrac{x}{\varepsilon}\big) \partial^k \tilde{u}(t,x). \tag{5.48}$$

We develop

$$\begin{aligned} r^\varepsilon = (\partial_t^2 + \mathcal{A}^\varepsilon)(\mathcal{B}^\varepsilon \tilde{u} - u^\varepsilon) = \ & \varepsilon^{-1} \big(\qquad \mathcal{A}_{yy} u^1 + \mathcal{A}_{xy} \tilde{u} \qquad \big) \\ & + \varepsilon^0 \ \big(\partial_t^2 \tilde{u} + \mathcal{A}_{yy} u^2 + \mathcal{A}_{xy} u^1 + \mathcal{A}_{xx} \tilde{u} \big) \\ & + \sum_{k=1}^{\alpha} \varepsilon^k \big(\partial_t^2 u^k + \mathcal{A}_{yy} u^{k+2} + \mathcal{A}_{xy} u^{k+1} + \mathcal{A}_{xy} u^k \big) \quad + \mathcal{O}(\varepsilon^{\alpha+1}), \end{aligned} \tag{5.49}$$

where $\mathcal{A}_{yy}, \mathcal{A}_{xy}, \mathcal{A}_{xx}$ are defined as

$$\mathcal{A}_{yy} = -\nabla_y \cdot \big(a(y) \nabla_y \cdot \big), \quad \mathcal{A}_{xy} = -\nabla_y \cdot \big(a(y) \nabla_x \cdot \big) - \nabla_x \cdot \big(a(y) \nabla_y \cdot \big),$$
$$\mathcal{A}_{xx} = -\nabla_x \cdot \big(a(y) \nabla_x \cdot \big).$$

We now need to substitute all the terms containing $\partial_t^2 \tilde{u}$ in this development until only space derivatives of $\tilde{u}$ are left. To do so, we apply recursive Boussinesq tricks. The result of this technical task is contained in the following lemma (the proof is postponed to the end of the section).

Lemma 5.2.11. *The solution of* (5.47) *satisfies*

$$\partial_t^2 \tilde{u} = \sum_{r=0}^{\alpha/2} \varepsilon^{2r} (-1)^r c^r \partial^{2r+2} \tilde{u} + \mathcal{O}(\varepsilon^{\alpha+2}), \tag{5.50}$$

where the tensors $c^r \in \mathrm{Ten}^{2r+2}(\mathbb{R}^d)$ *are defined by*

$$c^0 = a^0, \qquad c^r = a^{2r} - \sum_{\ell=0}^{r-1} b^{2(r-\ell)} \otimes c^\ell \quad 1 \leq r \leq \alpha/2. \tag{5.51}$$

Let us now rewrite the terms $\varepsilon^k \partial_t^2 u^k$ in the expansion (5.49). We deal with the two parities of k separately. Consider first k odd: $k = 2\ell - 1$ for some $1 \leq \ell \leq \alpha/2$. Using the definition of u^k in (5.48) and Lemma 5.2.11, we find

$$\varepsilon^{2\ell-1} \partial_t^2 u^{2\ell-1} = \sum_{r=\ell}^{\alpha/2} \varepsilon^{2r} (-1)^{r-\ell} \big(c^{r-\ell} \otimes \chi^{2\ell-1}\big) \partial^{2r+1} \tilde{u} + \mathcal{O}(\varepsilon^{\alpha+2\ell+1}).$$

Summing over the odd indices $k = 2\ell - 1 \leq \alpha$ and reordering the terms, we get

$$\begin{aligned} \sum_{\ell=1}^{\alpha/2} \varepsilon^{2\ell-1} \partial_t^2 u^{2\ell-1} &= \sum_{\ell=1}^{\alpha/2} \sum_{r=\ell}^{\alpha/2} \varepsilon^{2r-1} (-1)^{r-\ell} \big(c^{r-\ell} \otimes \chi^{2\ell-1}\big) \partial^{2r+1} \tilde{u} + \mathcal{O}(\varepsilon^{\alpha+3}) \\ &= \sum_{r=1}^{\alpha/2} \varepsilon^{2r-1} \Big(\sum_{\ell=1}^{r} (-1)^{r-\ell} c^{r-\ell} \otimes \chi^{2\ell-1} \Big) \partial^{2r+1} \tilde{u} + \mathcal{O}(\varepsilon^{\alpha+3}). \end{aligned} \tag{5.52}$$

We proceed in the same way for the even indices $k = 2\ell$, where $1 \leq \ell \leq \alpha/2$. Lemma 5.2.11 ensures that

$$\varepsilon^{2\ell} \partial_t^2 u^{2\ell} = \sum_{r=\ell}^{\alpha/2} \varepsilon^{2r} (-1)^{r-\ell} \big(c^{r-\ell} \otimes \chi^{2\ell}\big) \partial^{2r+2} \tilde{u} + \mathcal{O}(\varepsilon^{\alpha+2\ell+2}),$$

and, summing over the even indices $k = 2\ell \leq \alpha$, we obtain

$$\sum_{\ell=1}^{\alpha/2} \varepsilon^{2\ell} \partial_t^2 u^{2\ell} = \sum_{r=1}^{\alpha/2} \varepsilon^{2r} \Big(\sum_{\ell=1}^{r} (-1)^{r-\ell} c^{r-\ell} \otimes \chi^{2\ell} \Big) \partial^{2r+2} \tilde{u} + \mathcal{O}(\varepsilon^{\alpha+4}). \tag{5.53}$$

Using Lemma 5.2.11, (5.52), and (5.53) in the development (5.49) then brings

$$\begin{aligned} r^\varepsilon = \quad & \varepsilon^{-1} \big(\mathcal{A}_{yy} u^1 + \mathcal{A}_{xy} \tilde{u} \qquad\qquad\qquad\quad \big) \\ & + \varepsilon^0 \ \big(\mathcal{A}_{yy} u^2 + \mathcal{A}_{xy} u^1 + \mathcal{A}_{xx} \tilde{u} + a^0 \partial^2 \tilde{u} \big) \\ & + \sum_{r=1}^{\alpha/2} \varepsilon^{2r-1} \Big(\mathcal{A}_{yy} u^{2r+1} + \mathcal{A}_{xy} u^{2r} + \mathcal{A}_{xy} u^{2r-1} + \Big(\sum_{\ell=1}^{r} (-1)^{r-\ell} c^{r-\ell} \otimes \chi^{2\ell-1} \Big) \partial^{2r+1} \tilde{u} \Big) \\ & + \sum_{r=1}^{\alpha/2} \varepsilon^{2r} \quad \Big(\mathcal{A}_{yy} u^{2r+2} + \mathcal{A}_{xy} u^{2r+1} + \mathcal{A}_{xy} u^{2r} + \Big(\sum_{\ell=1}^{r} (-1)^{r-\ell} c^{r-\ell} \otimes \chi^{2\ell} + (-1)^r c^r \Big) \partial^{2r+2} \tilde{u} \Big) \\ & + \mathcal{O}(\varepsilon^{\alpha+1}). \end{aligned} \tag{5.54}$$

Using the definition of u^k, we compute

$$\begin{aligned}
\mathcal{A}_{yy}u^k &= -\nabla_y \cdot \big(a\nabla_y \chi^k_{i_1\cdot\cdot i_k}\big)\partial^k_{i_1\cdot\cdot i_k}\tilde{u},\\
\mathcal{A}_{xy}u^k &= -\nabla_y \cdot \big(ae_{i_1}\chi^k_{i_2\cdot\cdot i_k}\big)\partial^k_{i_1\cdot\cdot i_k}\tilde{u} - e^T_{i_1}a\nabla_y\chi^k_{i_2\cdot\cdot i_k}\partial^k_{i_1\cdot\cdot i_k}\tilde{u},\\
\mathcal{A}_{xx}u^k &= -e^T_{i_1}ae_{i_2}\chi^k_{i_3\cdot\cdot i_k}\partial^k_{i_1\cdot\cdot i_k}\tilde{u}.
\end{aligned}$$

Finally, canceling successively the term of order $\mathcal{O}(\varepsilon^k)$ for $k = -1, \dots, \alpha$, in (5.54), we obtain the cell problems given in (5.22).

Remark 5.2.12. In the case where the right hand side f is not zero, we verify, instead of (5.50), that

$$\partial_t^2\tilde{u} = \sum_{r=0}^{\alpha/2}\varepsilon^{2r}(-1)^r c^r \partial^{2r+2}\tilde{u} + \mathcal{S}_1^\varepsilon f + \mathcal{O}(\varepsilon^{\alpha+2}),$$

where $\mathcal{S}_1^\varepsilon f$ is given by

$$\mathcal{S}_1^\varepsilon f(t,x) = f(t,x) + \sum_{r=1}^{\alpha/2}(-1)^r\varepsilon^{2r}\bigg(\sum_{j=0}^{r-1}B^r(j-1)\bigg)\partial^{2r}f(t,x),$$

and $B^r(j-1)$ are the constant tensors defined in (5.56) below. Then, (5.54) contains the additional term $\mathcal{S}^\varepsilon f = \mathcal{S}_1^\varepsilon f + \mathcal{S}_2^\varepsilon f$, where

$$\mathcal{S}_2^\varepsilon f(t,x) = \sum_{k=1}^{\alpha}\chi^k\big(\tfrac{\cdot}{\varepsilon}\big)\partial^k f(t,x) + \sum_{r=1}^{\alpha/2}\sum_{k=1}^{\alpha-2r}(-1)^r\varepsilon^{2r+k}\chi^k\big(\tfrac{\cdot}{\varepsilon}\big)\bigg(\sum_{j=0}^{r-1}B^r(j-1)\bigg)\partial^{2r+k}f(t,x).$$

Proof of Lemma 5.2.11. Step 1. In a first step, let us prove that (5.50) holds with $c^r = \tilde{c}^r$ defined as

$$\tilde{c}^0 = a^0, \qquad \tilde{c}^r = a^{2r} + \sum_{j=1}^{r}\sum_{s=0}^{r-j}B^{r-s}(j-1)\otimes a^{2s} \quad 1\le r\le \alpha/2, \tag{5.55}$$

where $B^r(j)$ is defined recursively as

$$\begin{aligned}
&B^r(0) = -b^{2r} && 1\le r\le\alpha/2,\\
&B^r(j) = -\sum_{s=1}^{r-j}B^{r-s}(j-1)\otimes b^{2s} && j+1\le r\le\alpha/2, \qquad 1\le j\le\alpha/2-1.
\end{aligned} \tag{5.56}$$

We define the sequence of tensors

$$\begin{aligned}
&A^r(0) = a^{2r} && 0\le r\le\alpha/2,\\
&A^r(j) = \sum_{s=0}^{r-j}B^{r-s}(j-1)\otimes a^{2s} && j\le r\le\alpha/2, \qquad 1\le j\le\alpha/2,
\end{aligned} \tag{5.57}$$

and denote

$$\begin{aligned}
R(j) &= \sum_{r=j}^{\alpha/2}(-1)^r\varepsilon^{2r}A^r(j)\partial^{2r+2}\tilde{u} \qquad 0\le j\le\alpha/2,\\
S(j) &= \sum_{r=j+1}^{\alpha/2}(-1)^r\varepsilon^{2r}B^r(j)\partial^{2r}\partial_t^2\tilde{u} \quad 0\le j\le\alpha/2-1.
\end{aligned}$$

With this notation, the effective equation (5.47) can be written as

$$\partial_t^2\tilde{u} = \sum_{r=0}^{\alpha/2}(-1)^r\varepsilon^{2r}a^{2r}\partial^{2r+2}\tilde{u} + \sum_{r=1}^{\alpha/2}(-1)^r\varepsilon^{2r}\big(-b^{2r}\big)\partial^2\partial_t^2\tilde{u} = R(0)+S(0). \tag{5.58}$$

We claim that $R(j)$ and $S(j)$ satisfy the following inductive relation.

$$S(j) = R(j+1)+S(j+1)+\mathcal{O}(\varepsilon^{\alpha+2}) \qquad 0\le j\le \alpha/2-2, \tag{5.59a}$$

$$S(\alpha/2-1) = R(\alpha/2)+\mathcal{O}(\varepsilon^{\alpha+2}). \tag{5.59b}$$

Let us first prove (5.59a). Using (5.58) to substitute $\partial_t^2\tilde{u}$ in $S(j)$, we have

$$\begin{aligned} S(j) = &\sum_{r_1=j+1}^{\alpha/2}\sum_{r_2=0}^{\alpha/2}(-1)^{r_1+r_2}\varepsilon^{2(r_1+r_2)}\big(B^{r_1}(j)\otimes a^{2r_2}\big)\partial^{2(r_1+r_2)+2}\tilde{u}\\ &+\sum_{r_1=j+1}^{\alpha/2}\sum_{r_2=1}^{\alpha/2}(-1)^{r_1+r_2}\varepsilon^{2(r_1+r_2)}\big(-B^{r_1}(j)\otimes b^{2r_2}\big)\partial^{2(r_1+r_2)}\partial_t^2\tilde{u}. \end{aligned}$$

Changing the index $r=r_1+r_2$ and reordering the sum, we find that the two sums satisfy

$$\begin{aligned} S(i) = &\sum_{r=j+1}^{\alpha/2}(-1)^r\varepsilon^{2r}\bigg(\sum_{s=0}^{r-(j+1)}B^{r-s}(j)\otimes a^{2s}\bigg)\partial^{2r+2}\tilde{u}\\ &+\sum_{r=j+2}^{\alpha/2}(-1)^r\varepsilon^{2r}\bigg(-\sum_{s=1}^{r-(j+1)}B^{r-s}(j)\otimes b^{2s}\bigg)\partial^{2r}\partial_t^2\tilde{u}+\mathcal{O}(\varepsilon^{\alpha+2}). \end{aligned}$$

Using the definitions of $A^r(j)$ and $B^r(j)$, we verify that the right hand side equals $R(j+1)+S(j+1)+\mathcal{O}(\varepsilon^{\alpha+2})$ and (5.59a) is proved. To prove (5.59b), we use the definition of $S(j)$ and (5.58) to get

$$\begin{aligned} S(\alpha/2-1) &= (-1)^{\alpha/2}\varepsilon^\alpha B^{\alpha/2}(\alpha/2-1)\partial^\alpha\partial_t^2\tilde{u} = (-1)^{\alpha/2}\varepsilon^\alpha\big(B^{\alpha/2}(\alpha/2-1)\otimes a^0\big)\partial^{\alpha+2}\tilde{u}+\mathcal{O}(\varepsilon^{\alpha+2})\\ &= (-1)^{\alpha/2}\varepsilon^\alpha A^{\alpha/2}(\alpha/2)\partial^{\alpha+2}\tilde{u} = R(\alpha/2)+\mathcal{O}(\varepsilon^{\alpha+2}), \end{aligned}$$

and the claim (5.59) is verified. We now prove that (5.50) holds with $c^r=\tilde{c}^r$. From (5.58), applying recursively (5.59) gives

$$\partial_t^2\tilde{u} = \sum_{j=0}^{\alpha/2}R(j)+\mathcal{O}(\varepsilon^{\alpha+2}). \tag{5.60}$$

Exchanging the sums, we find that

$$\sum_{j=0}^{\alpha/2}R(j) = \sum_{j=0}^{\alpha/2}\sum_{r=j}^{\alpha/2}(-1)^r\varepsilon^{2r}A^r(j)\partial^{2r+2}\tilde{u} = \sum_{r=0}^{\alpha/2}\varepsilon^{2r}(-1)^r\bigg(\sum_{j=0}^{r}A^r(j)\bigg)\partial^{2r+2}\tilde{u}. \tag{5.61}$$

Using the definition of $A^r(i)$ in (5.57), we verify that for $r=0$, $\sum_{j=0}^r A^r(j)=a^0=\tilde{c}^0$ and for $1\le r\le\alpha/2$:

$$\sum_{j=0}^{r}A^r(j) = A^0(j)+\sum_{j=1}^{r}A^r(j) = a^{2r}+\sum_{j=1}^{r}\sum_{s=0}^{r-j}B^{r-s}(j-1)\otimes a^{2s} = \tilde{c}^r.$$

Combining this equality with (5.60) and (5.60) (5.50) holds with $c^r=\tilde{c}^r$ and Step 1 is proved.

Step 2. The second step is to prove that the sequence of tensors $\tilde{c}^r$, defined in (5.55), satisfies $\tilde{c}^r = c^r$, i.e., that $\tilde{c}^r$ satisfies the inductive relation (5.51). We prove this result by induction on r. The base case is trivially verified as (5.55) and (5.51) give $c^1 = a^2 - b^2 \otimes a^0 = \tilde{c}^1$. Let now $r \geq 2$ and assume that $\tilde{c}^s = c^s$ for $s = 1, \dots, r-1$. We have to verify that the tensor

$$c^r = a^{2r} - \sum_{\ell=0}^{r-1} b^{2(r-k)} \otimes c^\ell = a^{2r} - b^{2r} \otimes c^0 - \sum_{\ell=1}^{r-1} b^{2(r-\ell)} \otimes c^\ell,$$

equals $\tilde{c}^r$, defined in (5.55). Using the induction assumption and (5.55), we write

$$c^r = a^{2r} - \sum_{k=0}^{r-1} b^{2(r-\ell)} \otimes a^{2\ell} - \sum_{k=1}^{r-1} \sum_{j=1}^{\ell} \sum_{s=0}^{\ell-j} b^{2(r-k)} \otimes B^{\ell-s}(j-1) \otimes a^{2s}. \tag{5.62}$$

Let us denote the triple sum T and its summand $x^r_{\ell,j,s} = b^{2(r-\ell)} \otimes B^{\ell-s}(j-1) \otimes a^{2s}$. We apply the change of indices $m = r - \ell$ and exchange the sums twice to get

$$T = \sum_{m=1}^{r-1} \sum_{j=1}^{r-m} \sum_{s=0}^{r-m-j} x^r_{r-m,j,s} = \sum_{j=1}^{r-1} \sum_{m=1}^{r-j} \sum_{s=0}^{r-m-j} x^r_{r-m,j,s} = \sum_{j=1}^{r-1} \sum_{s=0}^{r-j-1} \sum_{m=1}^{r-j-s} x^r_{r-m,j,s}.$$

We claim that $B^r(j)$ satisfies (compare to (5.56))

$$B^r(j) = \sum_{m=1}^{r-j} b^{2m} \otimes B^{r-m}(j-1). \tag{5.63}$$

We proceed by induction on j. The case $j = 1$ follows the change of index $m = r - s$:

$$B^r(1) = -\sum_{s=1}^{r-1} b^{2(r-s)} \otimes b^{2s} = -\sum_{m=1}^{r-1} b^{2m} \otimes b^{2(r-m)} = -\sum_{m=1}^{r-1} b^{2m} \otimes B^{r-m}(0).$$

Assuming that (5.63) holds from 1 to $j-1$, we use (5.56) an get

$$\begin{aligned} B^r(j) &= \sum_{s=1}^{r-j} B^{r-s}(j-1) \otimes b^{2s} = \sum_{s=1}^{r-j} \sum_{m=1}^{r-s-j+1} b^{2m} \otimes B^{r-s-m}(j-2) \otimes b^{2s} \\ &= \sum_{m=1}^{r-j} b^{2m} \otimes \Big(\sum_{s=1}^{r-m-j+1} B^{r-s-m}(j-2) \otimes b^{2s} \Big) = \sum_{m=1}^{r-j} b^{2m} \otimes B^{r-m}(j-1), \end{aligned}$$

which proves (5.63). Thanks to (5.63), we then have

$$T = \sum_{j=1}^{r-1} \sum_{s=0}^{r-j-1} \Big(\sum_{m=1}^{r-s-j} b^{2m} \otimes B^{r-s-m}(j-1) \Big) \otimes a^{2s} = -\sum_{j=1}^{r-1} \sum_{s=0}^{r-j-1} B^{r-s}(j) \otimes a^{2s}.$$

We use this equality and the definition of $B^r(0)$ in (5.62), and change the index j to obtain

$$c^r = a^{2r} + \sum_{k=0}^{r-1} B^{r-k}(0) \otimes a^{2k} + \sum_{j=2}^{r} \sum_{s=0}^{r-j} B^{r-s}(j-1) \otimes a^{2s} = a^{2r} + \sum_{j=1}^{r} \sum_{s=0}^{r-j} B^{r-s}(j-1) \otimes a^{2s}.$$

This expression matches the definition of $\tilde{c}^r$ in (5.55) and Step 2 is proved. The proof of Lemma 5.2.11 is complete. □

5.2.6 Comparison with the tensors obtained via Taylor–Bloch expansion

Recently, a result for the long time homogenization of the wave equation was presented in [23]. The analysis from [23] generalizes the results for periodic tensors to quasiperiodic, almost-periodic and random tensors. In particular, an effective equation of arbitrary order is derived from the so-called Bloch–Taylor expansion of u^ε. Note that this derivation is profoundly different from our result, presented in Section 5.2.1, as the obtained equation is based on regularization techniques. In this section, we prove that, in the periodic case, the correctors defined in [23] match the correctors defined in (5.22), obtained with asymptotic expansion. This relation allows us to express the connection between the effective equation from [23] and the family of effective equations, defined in Definition 5.2.4.

Let us first summarize the derivation and the result from [23]. We consider $u^\varepsilon : [0, \varepsilon^{-\alpha}T] \times \mathbb{R}^d \to \mathbb{R}$

$$\begin{aligned} &\partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \big(a\big(\tfrac{x}{\varepsilon}\big)\nabla_x u^\varepsilon(t,x)\big) = 0 \quad \text{in } (0,\varepsilon^{-\alpha}T] \times \mathbb{R}^d, \\ &u^\varepsilon(0,x) = g(x), \quad \partial_t u^\varepsilon(0,x) = 0 \qquad\qquad \text{in } \mathbb{R}^d, \end{aligned} \tag{5.64}$$

where a is a tensor that can be of different nature: periodic, almost periodic, quasiperiodic, and random with decaying correlation at infinity (we refer to [23] for more details on these different assumptions). Recall that in [85] and [42, 43], the Bloch wave expansion of u^ε is used for the long time homogenization of (5.64) in the periodic case (see Section 4.1.1). As the Bloch theory does not apply for more general tensors, it is generalized in [23] as follows. As in that case the Bloch eigenfunctions might not exists, they are replaced by their formal Taylor expansion, that is based on the extended correctors. Then an expansion for u^ε is obtained and validated by the study of the corresponding defect. The tensors in the effective equations of arbitrary order are obtained with the definition of the extended correctors (detailed below).

Let us give some more details on their result in the periodic case. We consider the case $\alpha \geq 4$, as for $\alpha \leq 3$, [23] essentially cite the result from [18], which corresponds to what was done in Chapter 4. The effective equation is given by

$$\begin{aligned} &\partial_t^2 w^\varepsilon(t,x) - \sum_{j=0}^{\alpha} \varepsilon^j \bar{a}_j \partial^{j+1} w^\varepsilon(t,x) - \varepsilon^{2(\lfloor \frac{\alpha}{2} \rfloor + 1)} R w^\varepsilon(t,x) = 0 \qquad \text{in } (0,\varepsilon^{-\alpha}T] \times \mathbb{R}^d, \\ &w^\varepsilon(0,x) = g(x), \quad \partial_t w^\varepsilon(0,x) = 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{in } \mathbb{R}^d, \end{aligned} \tag{5.65}$$

where $\bar{a}_j$ are the effective tensors defined below (in particular $\bar{a}_j = 0$ for odd j, see (5.69)). Furthermore, Rw^ε is a regularization term given by

$$R w^\varepsilon(t,x) = \gamma(-1)^{\lfloor \frac{\alpha}{2} \rfloor + 1} \mathrm{Id} \partial^{2(\lfloor \frac{\alpha}{2} \rfloor + 1)+2} w^\varepsilon(t,x), \tag{5.66}$$

where γ is sufficiently large for (5.65) to be well-posed. They prove the following error estimate (given here in the particular periodic case).

Theorem 5.2.13 (Benoit & Gloria [23])**.** *Assume that g belongs to the Schwartz space and $a \in \mathrm{L}^\infty_{\mathrm{per}}(Y)$. Then*

$$\|u^\varepsilon - w^\varepsilon\|_{\mathrm{L}^\infty(0,\varepsilon^{-\alpha}T;\mathrm{L}^2(\mathbb{R}^d))} \leq C\varepsilon,$$

where the constant C depends on a norm of g, α, T, and γ.

The error estimate is thus obtained in a stronger norm than the preceding results ([42, 43] and Theorem 4.2.4). Nevertheless, in applications, no procedure is available to compute the regularization parameter γ in (5.66). Furthermore, numerical experiments shows that to find an acceptable value for γ is not an easy task (see the example in Section 5.4.3). This issue does not occur in the effective equation defined in Section 5.2.1. Indeed, in our effective equation as the corrections of order ε^{2r} are composed of pairs of positive operators, the well-posedness is obtained

without regularization. Nevertheless, in what follows, we verify that the tensors involved in both equations are the same. In particular, we show that it holds

$$\bar{a}_{2r} = (-1)^r c_r \qquad r = 1, \dots, \lfloor \alpha/2 \rfloor, \tag{5.67}$$

where c^r is the tensor given in (5.21) (and (5.55)) and $\bar{a}_{2r}$ are the tensors in (5.65).

Let us first define the extended correctors from [23]. For $K \geq 0$, the first K extended correctors $(\varphi_k, \sigma_k, \psi_k)_{k=0}^K$ in the direction $\eta \in \mathbb{R}^d$, $|\eta| = 1$, are defined as follows.

- $\varphi_0 = 0$ and for $k \geq 1$, $\varphi_k \in \mathrm{W}_{\mathrm{per}}(Y)$ solves

$$-\nabla_y \cdot (a \nabla_y \varphi_k) = \nabla_y \cdot (-\sigma_{k-1}\eta + a\eta\varphi_{k-1} + \nabla_y \psi_{k-1}), \tag{5.68}$$

- for all $k \geq 0$, the tensor $\bar{a}_k \in \mathrm{Sym}^{k+2}(\mathbb{R}^d)$, the symmetric matrix $\tilde{a}_k$ and the scalar λ_k are given by

$$(\bar{a}_k)_{i_1 \cdots i_{k+2}} \eta_{i_1} \cdots \eta_{i_{k+2}} = \tilde{a}_k \eta = \left\langle a(\nabla_y \varphi_{k+1} + \eta \varphi_k) \right\rangle_Y, \qquad \lambda_k = \tilde{a}_k \eta \cdot \eta,$$

- $\psi_0 = \psi_1 = 0$ and for $k \geq 2$, $\psi_1 \in \mathrm{W}_{\mathrm{per}}(Y)$ solves

$$-\Delta \psi_k = \nabla_y \psi_{k-1} \cdot \eta + \sum_{\ell=1}^{k-1} \lambda_{k-1-\ell} \varphi_\ell,$$

- for $k \geq 1$, the field q_k is given by

$$q_k = a(\nabla_y \varphi_k + \eta \varphi_{k-1}) - \tilde{a}_{k-1}\eta + \nabla_y \psi_{k-1} - \sigma_{k-1}\eta, \qquad \left\langle q_k \right\rangle_Y = 0,$$

- $\sigma_0 = 0$ and for $k \geq 1$, $\sigma_k \in \mathrm{W}_{\mathrm{per}}(Y)$ is a skew-symmetric matrix (i.e., $(\sigma_k)_{mn} = -(\sigma_k)_{nm}$) that satisfies

$$-\Delta (\sigma_k)_{mn} = \partial_m (q_k)_n - \partial_n (q_k)_m \quad \forall 1 \leq m, n \leq d, \quad \partial_n (\sigma_k)_{mn} = (q_k)_m \quad \forall 1 \leq m \leq d.$$

Let us now verify that the extended correctors are the same functions as the correctors defined in (5.22). For $k \geq 0$, let us denote the function $\chi_\eta^k = \chi^k_{i_1 \cdots i_k} \eta_{i_1} \cdots \eta_{i_k}$, where $\chi^k_{i_1 \cdots i_k}$ are the correctors defined in (5.22). Let $R^k_{i_1 \cdots i_k}$ denote the right hand side of the cell problem for $\chi^k_{i_1 \cdots i_k}$. Then, the right hand side of the cell problem for χ^k_η is given by $R^k_{i_1 \cdots i_k} \eta_{i_1} \cdots \eta_{i_k}$. Note that the following important property is proved in [23]:

$$\tilde{a}_{2j-1} = 0, \quad \lambda_{2j-1} = 0 \quad \forall j \geq 1. \tag{5.69}$$

We prove the following relation by induction:

$$\psi_k = \chi_\eta^k \quad k \geq 1. \tag{5.70}$$

Writing (5.68) for $k = 1$, we verify that $\psi_1 = \chi_\eta^1$. Then, the equality $\lambda_0 = a^0 \eta \cdot \eta$ follows and (5.70) is verified for $k = 2$. Assume now that $\psi_j = \chi_\eta^j$ for $j = 1, \dots, k-1$ for some $k \geq 3$. Comparing the definition of λ_k with the constraint that we fixed on c^j in (5.26), we verify that

$$\lambda_{2j} = (-1)^j c_\eta^j \qquad j = 0, \dots, \lfloor \tfrac{k-2}{2} \rfloor, \tag{5.71}$$

where we denoted $c_\eta^j = c^j_{i_1 \cdots i_{2j+2}} \eta_{i_1} \cdots \eta_{i_{2j+2}}$. Let us now rewrite the cell problem for φ_k. First, using the definitions of σ_k, we verify that

$$-\nabla_y \cdot (\sigma_{k-1}\eta) = -\partial_m (\sigma_{k-1})_{mn} \eta_n = \partial_m (\sigma_{k-1})_{nm} \eta_n = q_{k-1} \cdot \eta.$$

Using the definition of q_{k-1}, we thus rewrite (5.68) as

$$-\nabla_y\cdot(a\nabla_y\varphi_k) = a(\nabla_y\varphi_{k-1}+\eta\varphi_{k-2})\cdot\eta-\tilde{a}_{k-2}\eta\cdot\eta+\nabla_y\psi_{k-2}\cdot\eta-\sigma_{k-1}\eta\cdot\eta+\nabla_y\cdot(a\eta\varphi_{k-1})+\Delta\psi_{k-1}.$$

As σ_{k-1} is skew-symmetric, it satisfies $\sigma_{k-1}\eta\cdot\eta = 0$. Using then the definitions of ψ_{k-1} and λ_{k-2}, we obtain

$$-\nabla_y\cdot(a\nabla_y\varphi_k) = \nabla_y\cdot(a\eta\varphi_{k-1}) + a(\nabla_y\varphi_{k-1}+\eta\varphi_{k-2})\cdot\eta - \sum_{\ell=1}^{k-2}\lambda_{k-2-\ell}\varphi_\ell - \lambda_{k-2}.$$

Reordering the terms, separating the even and odd indices, and using (5.69) and (5.71), we have

$$\sum_{\ell=1}^{k-2}\lambda_{k-2-\ell}\varphi_\ell = \sum_{\ell=0}^{k-3}\lambda_\ell\varphi_{k-2-\ell} = \sum_{j=0}^{\lfloor\frac{k-3}{2}\rfloor}\lambda_{2j}\varphi_{k-2-2j} + \sum_{j=1}^{\lfloor\frac{k-2}{2}\rfloor}\lambda_{2j-1}\varphi_{k-1-2j} = \sum_{j=0}^{\lfloor\frac{k-3}{2}\rfloor}(-1)^j c_\eta^j\varphi_{k-2-2j}. \tag{5.72}$$

Assuming that k is odd, $k = 2r+1$, (5.72) implies

$$\sum_{\ell=1}^{k-2}\lambda_{k-2-\ell}\varphi_\ell = \sum_{\ell=1}^{r}(-1)^{r-\ell}c_\eta^{r-\ell}\varphi_{2\ell-1},$$

and, using (5.69), the cell problem for φ_{2r+1} thus reads

$$-\nabla_y\cdot(a\nabla_y\varphi_{2r+1}) = \nabla_y\cdot(a\eta\varphi_{2r}) + a(\nabla_y\varphi_{2r}+\eta\varphi_{2r-1})\cdot\eta + \sum_{\ell=1}^{r}(-1)^{r-\ell+1}c_\eta^{r-\ell}\varphi_{2\ell-1}.$$

Comparing this equation with (5.22c) proves that $\psi_k = \chi_\eta^k$. Assume then that k is even, $k = 2r+2$. Thanks to (5.72), we have

$$\sum_{\ell=1}^{k-2}\lambda_{k-2-\ell}\varphi_\ell = \sum_{\ell=1}^{r}(-1)^{r-\ell}c_\eta^{r-\ell}\varphi_{2\ell},$$

and, using (5.71), the corresponding cell problem for φ_{2r+2} thus reads

$$-\nabla_y\cdot(a\nabla_y\varphi_{2r+2}) = \nabla_y\cdot(a\eta\varphi_{2r+1}) + a(\nabla_y\varphi_{2r+1}+\eta\varphi_{2r})\cdot\eta + \sum_{\ell=1}^{r}(-1)^{r-\ell+1}c_\eta^{r-\ell}\varphi_{2\ell} - (-1)^r c_\eta^r.$$

Comparing this equation with (5.22d) proves that $\psi_k = \chi_\eta^k$. We have proved that (5.70) holds for all k. The same argument as for (5.71) thus proves that (5.67) holds.

5.3 Effective behavior of high frequency waves

In this section, we discuss the influence of high frequency waves on the dispersion phenomena occurring in long time wave propagation in periodic media. In particular, we show that the higher the frequencies of the initial position are, the sooner dispersion effects appear. Two conclusions are then drawn. First, the so-called "long time effects" are related to the wave and particularly to its high frequencies. Second, to deal with certain frequency regimes, we need higher order effective equations.

Let us consider a simple one-dimensional problem. Let $a : [0,1] \to \mathbb{R}$ be a smooth 1-periodic, positive, bounded tensor and let g^0 be a given initial wave. We consider $u_1^\varepsilon : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$, the solution of the equation

$$\begin{aligned} &\partial_t^2 u_1^\varepsilon(t,x) = \partial_x\big(a\big(\tfrac{x}{\varepsilon}\big)\partial_x u_1^\varepsilon(t,x)\big) && (t,x)\in\mathbb{R}_+\times\mathbb{R},\\ &u_1^\varepsilon(0,x) = g^0(x),\ \partial_t u_1^\varepsilon(0,x) = 0 && x\in\mathbb{R}. \end{aligned} \tag{5.73}$$

We recall the result of Theorem 4.2.4, in Section 4.3.1: on the time interval $[0, \varepsilon^{-2}T]$, u_1^ε is approximated by the solution $\tilde{u}_1 : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ of the effective equation

$$\begin{array}{ll} \partial_t^2 \tilde{u}(t,x) = a^0 \partial_x^2 \tilde{u}(t,x) + \varepsilon^2 b^2 \partial_x^2 \partial_t^2 \tilde{u}(t,x) & (t,x) \in \mathbb{R}_+ \times \mathbb{R}, \\ \tilde{u}(0,x) = g^0(x),\ \partial_t \tilde{u}(0,x) = 0 & x \in \mathbb{R}, \end{array} \tag{5.74}$$

where $b^2 = \langle \chi^2 \rangle_Y$ and $\chi \in \mathrm{W}_{\mathrm{per}}(Y)$ is the zero mean first corrector corresponding to $a(y)$. This is in fact true in any interval $\Omega \subset \mathbb{R}$ and, for example, for sufficiently regular g^0 with an $\mathcal{O}(1)$ support. This result means that, given g^0, the dispersive behavior of u_1^ε at $t = \varepsilon^{-2}$ is determined by $a(y)$ and the size ε. For that reason, let us specify the dependence of $\tilde{u}$ on ε as $\tilde{u}(\varepsilon; t, x)$. Consider now $u_\nu^\varepsilon : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$, the solution of

$$\begin{array}{ll} \partial_t^2 u_\nu^\varepsilon(t,x) = \partial_x \big(a\big(\frac{x}{\varepsilon}\big) \partial_x u_\nu^\varepsilon(t,x) \big) & (t,x) \in \mathbb{R}_+ \times \mathbb{R}, \\ u_\nu^\varepsilon(0,x) = g^0(\nu x),\ \partial_t u_\nu^\varepsilon(0,x) = 0 & x \in \mathbb{R}, \end{array} \tag{5.75}$$

where $\nu > 0$ is a given scaling parameter. Making the changes of variables $\hat{x} = \nu x$ and $\hat{t} = \nu t$ in (5.75) and introducing $\hat{u}_\nu^\varepsilon(\hat{t}, \hat{x}) = u_\nu^\varepsilon(\hat{t}/\nu, \hat{x}/\nu)$, we verify that

$$\begin{array}{ll} \partial_{\hat{t}}^2 \hat{u}_\nu^\varepsilon(\hat{t},\hat{x}) = \partial_{\hat{x}} \big(a\big(\frac{\hat{x}}{\nu\varepsilon}\big) \partial_{\hat{x}} \hat{u}_\nu^\varepsilon(\hat{t},\hat{x}) \big) & (\hat{t},\hat{x}) \in \mathbb{R}_+ \times \mathbb{R}, \\ \hat{u}_\nu^\varepsilon(0,\hat{x}) = g^0(\hat{x}),\ \partial_{\hat{t}} \hat{u}_\nu^\varepsilon(0,\hat{x}) = 0 & \hat{x} \in \mathbb{R}. \end{array} \tag{5.76}$$

Observe that (5.76) is the same equation as (5.73) up to the period of oscillation of the tensor: ε is replaced by $\nu\varepsilon$. Accordingly, the macroscopic behavior of $\hat{u}_\nu^\varepsilon$ can be described up to timescales $\mathcal{O}((\nu\varepsilon)^{-2})$ by $\tilde{u}(\nu\varepsilon; t, x)$ (the solution of (5.74) where ε is replaced by $\nu\varepsilon$). Consequently, $\hat{u}_\nu^\varepsilon(\hat{t} = (\nu\varepsilon)^{-2}, \hat{x})$, i.e., $u_\nu^\varepsilon(t = \varepsilon^{-2}/\nu^3, x/\nu)$, must have a similar dispersive behavior as $u_1^\varepsilon(t = \varepsilon^{-2}, x)$. In other words, if $\nu > 1$ (i.e., an increase of the frequencies of the initial wave), the amplitude of the dispersion developed by u_ν^ε is as important as for u_1^ε, but it occurs at a shorter time.

To illustrate this conclusion, let us consider the example introduced in Section 4.4.1. We consider the model problem given by the data

$$g^0(x) = e^{-10x^2}, \quad a(y) = \sqrt{2} - \cos(2\pi y), \quad \varepsilon = 1/20.$$

Recall that for these data, u_1^ε has a visible long time dispersive behavior at $t = \varepsilon^{-2} = 400$ (see Figure 4.3 and also Figure 5.1). We let $\nu = 2^{1/3}$ so that, based on the previous argument, a similar dispersive effect must appear in the behaviour of u_ν^ε at $t = \varepsilon^{-2}/\nu^3 = 200$. Denote $\tilde{u}_\nu$ the effective solution for u_ν^ε, i.e., $\tilde{u}_1(t,x) = \tilde{u}(\varepsilon; t, x)$ and $\tilde{u}_\nu(t,x) = \tilde{u}(\nu\varepsilon; t, x)$. To account for the scaling x/ν, we compare u_1^ε and u_ν^ε in the space intervals

$$I_1 = \sqrt{a^0}t + [-4, 1], \qquad I_\nu = \sqrt{a^0}t + [-4, 1]/\nu,$$

respectively. In Figure 5.1, $u_\nu^\varepsilon, \tilde{u}_\nu$ are displayed in their respective settings. As predicted, the graphs of $\{\tilde{u}_1 : t = \varepsilon^{-2}, x \in I_1\}$ and $\{\tilde{u}_\nu : t = \varepsilon^{-2}/\nu^3, x \in I_\nu\}$ are identical. Accordingly, the amplitudes of the dispersion in u_1^ε and u_ν^ε are the same (up to the microscopic oscillations). Let us now proceed to the same experiment for larger values of ν. We let $\nu = 40^{1/3}$ and $\nu = 80^{1/3}$ so that the dispersive effects are expected to happen at $t = 10$ and $t = 5$, respectively. In Figure 5.2, we display the solutions u_ν^ε and $\tilde{u}_\nu$ at $t = \varepsilon^{-2}/\nu^3$ and for $x \in I_\nu$. In both cases, we verify that $\{\tilde{u}_\nu : t = \varepsilon^{-2}/\nu^3, x \in I_\nu\}$ is the same as in both plots of Figure 5.1. However, this time, the tail of $\tilde{u}_\nu$ does not match the dispersion developed by u_ν^ε. We observe that the farther of the front wave we are, the worse $\tilde{u}_\nu$ is.

These experiments lead to two conclusions. First, the higher the frequencies of the initial wave is, the sooner u^ε develops dispersion. Second, for certain regimes, the effective equations obtained in Section 4.2 does not describe well the dispersion of u^ε. Note that this second issue does

not contradicts Theorem 4.2.4. Indeed, in the examples with $\nu = 40^{1/3}$ and $\nu = 80^{1/3}$, the quantities $\|g^0(\nu\cdot)\|_{\mathrm{H}^5(\Omega)}$ are considerable so that the corresponding bound on the error is large. The consequence of these conclusions is that to homogenize high frequency waves, we need higher order effective models. In Section 5.4.1, we use the higher order effective equations of the family defined in Definition 5.2.4 to capture the additional dispersion effects observed in Figure 5.2.

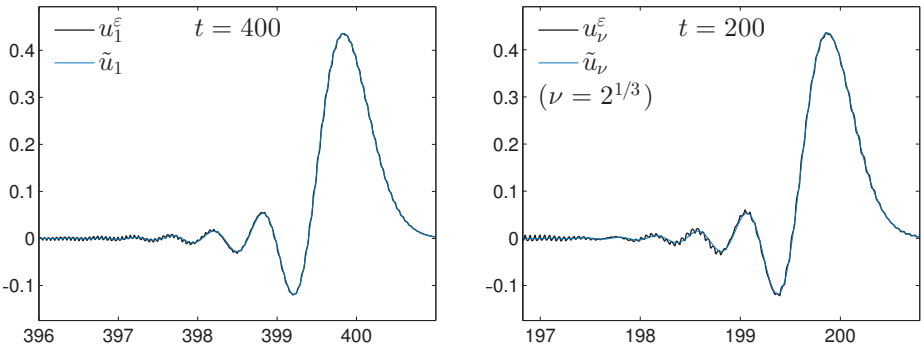

Figure 5.1: Comparison of $u_1^\varepsilon, \tilde{u}_1$ at $t = \varepsilon^{-2} = 400$ for $x \in I_1$ and $u_\nu^\varepsilon, \tilde{u}_\nu$ at $t = \varepsilon^{-2}/\nu^3 = 200$ for $x \in I_\nu$ $(\nu = 2^{1/3})$.

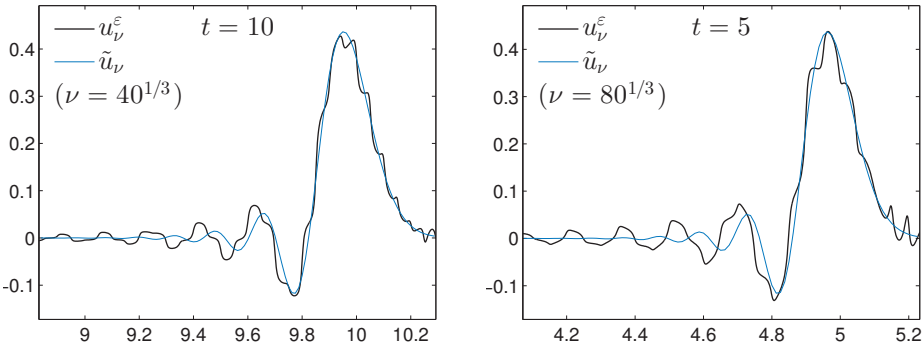

Figure 5.2: Plots of $u_\nu^\varepsilon, \tilde{u}_\nu$ at $t = \varepsilon^{-2}/\nu^3$ for $x \in I_\nu$ for $\nu = 40^{1/3}$ (left) and $\nu = 80^{1/3}$ (right).

5.4 Numerical experiments

In this section, we illustrate the use of the family of high order effective equations, defined in Section 5.2.1 (Definition 5.2.4). Instead of considering examples on large timescales, where u^ϵ is extremely costly—or impossible—to approximate, we consider examples with high frequency initial data. Indeed, we have seen in Section 5.3 that higher order effective equations are also needed for the homogenization of the wave equation in high frequency regimes. First, we consider the one-dimensional examples, considered in Section 5.3. Second, we deal with a two-dimensional example.

5.4.1 One-dimensional example

We come back to the examples presented in Section 5.3. Recall that the model problem is given by the data

$$g^0(x) = e^{-10x^2}, \quad a(y) = \sqrt{2} - \cos(2\pi y), \quad \varepsilon = 1/20,$$

and let $\nu > 0$ determine the variance of the initial pulse $g^0(\nu\cdot)$. The solution of (5.75) is denoted u_ν^ε and the solution of the effective equation of order 1 (5.76) is denoted $\tilde{u}_\nu^1$. Similarly, we let $\tilde{u}_\nu^s$ be the effective equation of order s, i.e., the solution of (5.19) with $\alpha/2 = s$, where the coefficients are computed with Algorithm 5.2.10.

Recall that for $\nu = 40^{1/3}$ and $\nu = 80^{1/3}$, $\tilde{u}_\nu^1$ does not describe all the dispersion developed by u_ν^ε at $t = \varepsilon^{-2}/\nu^3$ (see Figure 5.2). In these two cases, we compute $\tilde{u}_\nu^2, \tilde{u}_\nu^3, \tilde{u}_\nu^4$ using a Fourier method on a grid of size $h = \varepsilon/8$ (see Section 2.4). For $s = 1, \dots, 4$, we define the normalized error

$$\mathrm{err}(\tilde{u}_\nu^s)(t) = \|(u_\nu^\varepsilon - \tilde{u}_\nu^s)(t)\|_{\mathrm{L}^2(\Omega)} / \|u_\nu^\varepsilon(t)\|_{\mathrm{L}^2(\Omega)}.$$

In Figure 5.3, the computed normalized errors are displayed for $\{\tilde{u}_\nu^s\}_{s=1}^4$ on the time interval $[0, 100]$, for $\nu = 40^{1/3}$ (left) and $\nu = 80^{1/3}$ (right). In both cases, we observe that the higher the order of the effective solution is, the lower the error is. Furthermore, as already noticed, we see that for $\nu = 80^{1/3}$ the effective solutions drift away from u_ν^ε more quickly than for $\nu = 40^{1/3}$. In Figure 5.4, we compare u_ν^ε and $\{\tilde{u}_\nu^s\}_{s=1}^3$ in the interval $I_\nu = \sqrt{a^0}t + [-4, 1]/\nu$ at $t = \varepsilon^{-2}/\nu^3$. In the case $\nu = 40^{1/3}$, in the left plot, we observe that $\tilde{u}_\nu^2$ and $\tilde{u}_\nu^3$ capture well the dispersion of u_ν^ε while $\tilde{u}_\nu^1$ does not. No significant difference between $\tilde{u}_\nu^2$ and $\tilde{u}_\nu^3$ is visible. In the case $\nu = 80^{1/3}$, in the right plot, we see that $\tilde{u}_\nu^3$ does capture slightly better the tail of the dispersion than $\tilde{u}_\nu^2$. As expected, the higher the order of the effective equation is, the better the dispersion is captured. However, the improvement from $\tilde{u}_\nu^s$ to $\tilde{u}_\nu^{s+1}$ is modest (this is also visible in Figure 5.3).

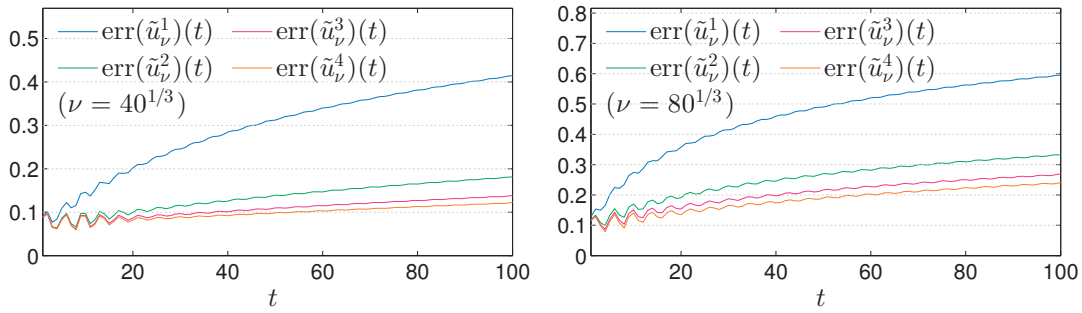

Figure 5.3: Comparison of the normalized errors of $\{\tilde{u}_\nu^s\}_{s=1}^4$ for the time interval $[0, 100]$ for $\nu = 40^{1/3}$ (left) and $\nu = 80^{1/3}$ (right).

5.4.2 Two-dimensional example

We now turn to a two dimensional example. We consider the model problem given by the data

$$g^0(x) = e^{-20|\nu x|^2}, \ \nu = 5^{1/3}, \quad a(y) = \begin{pmatrix} 1 - 0.5\cos(2\pi y_2) & 0 \\ 0 & 1 - 0.5\cos(2\pi y_2) \end{pmatrix}, \quad \varepsilon = 1/10.$$

Note that for $\nu = 1$, these are the data of the example considered in Section 4.4.3. In particular, $a^\varepsilon(x) = a\left(\frac{x}{\varepsilon}\right)$ describes a layered material (in the x_2-direction). Following the same argument as in Section 5.3, we thus expect visible dispersive effects at $T = \varepsilon^{-2}/\nu^3 = 20$. For this moderately

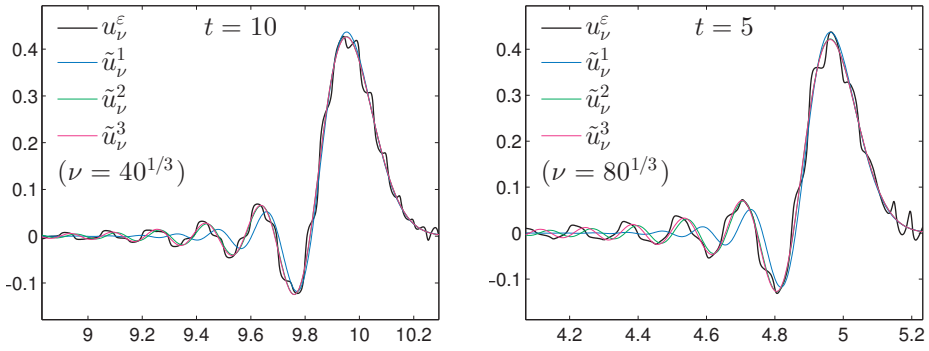

Figure 5.4: Plots of u_ν^ε and $\{\tilde{u}_\nu^s\}_{s=1}^3$ at $t = \varepsilon^{-2}/\nu^3$ for $x \in I_\nu$ for $\nu = 40^{1/3}$ (left) and $\nu = 80^{1/3}$ (right).

long time, we are able to compute u^ε (which is still extremely costly). We thus consider the pseudoinfinite domain

$$\Omega = (-L_1, L_1) \times (-L_2, L_2), \quad L_i = \left\lfloor \sqrt{a_{ii}^0} T \right\rfloor + 2.$$

For the space discretization of u^ε, we use a spectral method on a grid of size $h = \varepsilon/16$ (see Section 2.3). The time integration of the obtained second order ODE is done with the leap frog scheme with $\Delta t = h/100$. We denote $\tilde{u}^s$ the effective solution of order s (i.e., the solution of (5.19) with $\alpha/2 = s$). The higher order effective tensors are computed using Algorithm 5.2.10. To approximate $\tilde{u}^s$, we use a Fourier method on a grid of size $h = \varepsilon/8$ (see Section 2.4).

We first compare the front waves that travel in the x_2-direction, which is the oscillating direction of the medium. In Figure 5.5, we display u^ε (top-left), $\tilde{u}^1$ (top-right), and $\tilde{u}^2$ (bottom-left) on subdomains of Ω and the corresponding cuts along $x_1 = 0$ (bottom-right). We observe that u^ε oscillates at the micro scale and has a strongly dispersive behavior at the macro scale. This dispersion is not accurately described by $\tilde{u}^1$. The description of the dispersion is better for $\tilde{u}^2$. The comparison of the cuts along $x_1 = 0$ reveals that $\tilde{u}^2$ (green) indeed describes the dispersion better but further on the tail it is not accurate either. We also observe that $\tilde{u}^3$ (red) is slightly better. As in the one-dimensional case, the improvement brought by $\tilde{u}^{s+1}$ compared to $\tilde{u}^s$ is rather limited.

Second, we compare the front waves that travel in the x_1-direction. In Figure 5.6, we display u^ε (top-left), $\tilde{u}^1$ (top-right), and $\tilde{u}^2$ (bottom-left) on the subdomains of Ω and the corresponding cuts along $x_2 = 0$ (bottom-right). First, we observe that u^ε oscillates at the microscopic scale in the x_2 direction. We see that the macroscopic behavior of u^ε is well captured by both $\tilde{u}^1$ and $\tilde{u}^2$. However, a closer look reveals that the tail of the dispersion is better described by $\tilde{u}^1$ than $\tilde{u}^2$. Furthermore, in the bottom-left plot, we see that $\tilde{u}^3$ (red) has an even stronger flattening effect. Hence, while $\tilde{u}^3$ is supposed to describe more accurately the dispersion effects, it does the contrary. This negative effect must be linked to the construction of the effective tensors in Algorithm 5.2.10. It is doubtless that other effective equations in the family $\mathcal{E}$ would describe u^ε more accurately.

Let us summarize the outcome of this example. On one hand, the effects developed in the x_2 direction by u^ε are described better by the higher order effective solutions. On the other hand, in the x_1 direction u^ε is already well described by the first order effective solution and higher order effective solutions are less and less accurate. To remedy this issue, further research is needed to

find other effective equations that do not have such effect.

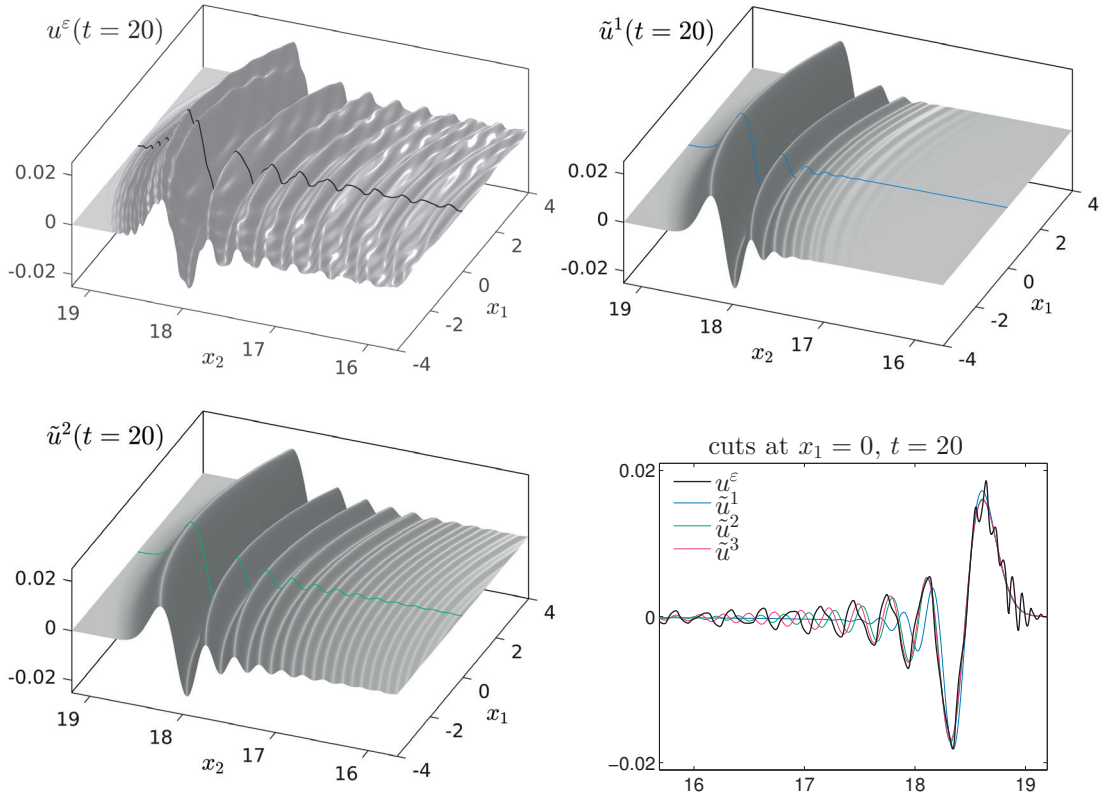

Figure 5.5: Comparison of u^ε and $\tilde{u}^s$ at $t = 20$ on the subdomains $[-4, 4] \times [\sqrt{a^0_{22}}t - 5/\nu, \sqrt{a^0_{22}}t + 1/\nu]$, and the corresponding cuts along $x_1 = 0$.

5.4.3 Attempt of regularization of the ill-posed high order effective equation

In this example, we illustrate that the regularized effective equation (5.65) from [23] is difficult to use in practice.

Consider the one-dimensional model problem of Section 5.4.1, for $\nu = 40^{1/3}$ (see the left plot of Figure 5.4). Let w^ε_γ be the solution of the regularized effective equation (5.65) of order 2, i.e., w^ε_γ solves

$$\partial_t^2 w^\varepsilon - a^0 \partial_x^2 w^\varepsilon_\gamma - \varepsilon^2 \bar{a}_2 \partial_x^2 w^\varepsilon_\gamma - \varepsilon^4 \bar{a}_4 \partial_x^4 w^\varepsilon_\gamma + \varepsilon^6 \gamma \partial_x^6 w^\varepsilon_\gamma = 0.$$

The index γ specifies the dependence of w^ε_γ on the regularization parameter γ. Recall that in Section 5.2.6, we proved that the coefficients satisfy $\bar{a}_{2r} = (-1)^r c^r$, where c^r are defined in (5.21). To approximate w^ε_γ we use the Fourier method. We notice that the grid size h has an influence on the well-posedness of the equation. Indeed, the method can lead to a stable approximation for some h and explodes for some smaller h. For our test, we fix $h = \varepsilon/4$ (the grid has to capture the frequencies of the initial position $g^0(\nu x) = e^{-10(\nu x)^2}$, $\nu = 40^{1/3}$). On this grid, we verify numerically that the equation is ill-posed for $\gamma = 10^{-5}$ and well-posed for $\gamma = \gamma^* = 2 \cdot 10^{-5}$. Then, we compute w^ε_γ for 30 values of γ in the interval $[\gamma^*, 2.5 \cdot 10^{-3}]$. The obtained solutions are displayed with u^ε in Figure 5.7. We observe that for $\gamma = \gamma_*$ and the 2 next values, w^ε_γ acceptably capture the dispersion of u^ε. However, for all higher values of γ, w^ε_γ is far from describing u^ε.

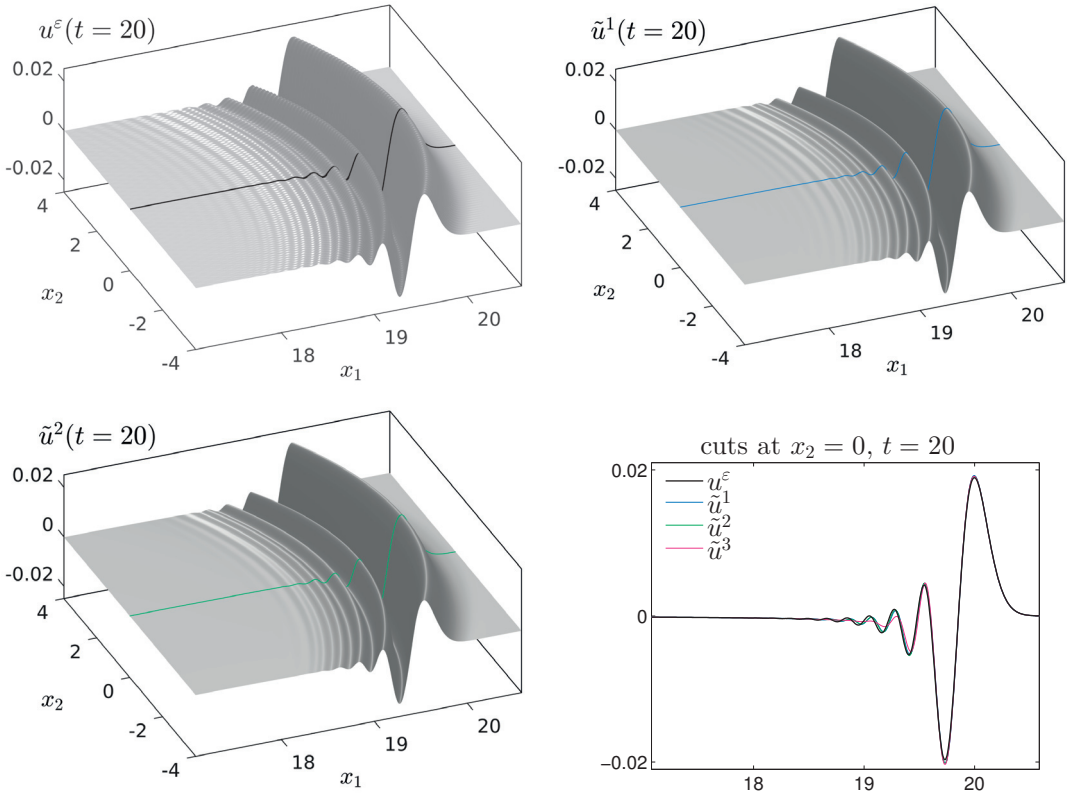

Figure 5.6: Comparison of u^ε and $\tilde{u}^s$ at $t = 20$ on the subdomains $[\sqrt{a^0_{11}}t - 5/\nu, \sqrt{a^0_{11}}t + 1/\nu] \times [-4, 4]$, and the corresponding cuts along $x_2 = 0$.

The conclusion of this experiment is that in order to use the effective equation from [23] in this application, we would need a procedure providing γ in the small window $[2 \cdot 10^{-5}, 2.7 \cdot 10^{-4}]$. In particular, γ cannot be randomly guessed to obtain a valid effective equation.

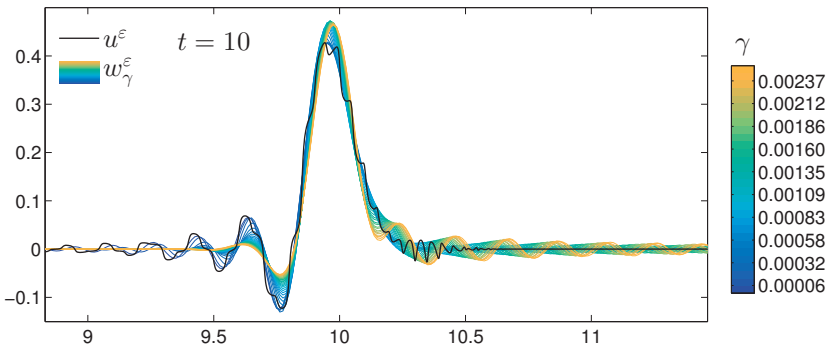

Figure 5.7: Comparison of u^ε and the regularization effective equation w^ε_γ for several value of the regularization parameter $\gamma \in [2 \cdot 10^{-5}, 2.5 \cdot 10^{-3}]$.

6 Effective models for long time wave propagation in locally periodic media

In Chapter 4, we derived a family of effective equations for wave propagation in periodic media for timescales of order $\mathcal{O}(\varepsilon^{-2})$. In practice, the periodicity assumption is often relaxed to local periodicity, i.e., a slow deformation in the tensor is allowed. Such model is useful if the features of the material are changing at the macroscopic scale. In this chapter, we generalize the technique and result from Chapter 4 and derive effective models for wave propagation in locally periodic media at timescales $\mathcal{O}(\varepsilon^{-2})$. This analysis constitutes the first result for the description of long time effects for the wave equation in locally periodic media.

Let $\Omega \subset \mathbb{R}^d$ be an arbitrarily large hypercube and let $a^\varepsilon(x)$ be a tensor with a locally periodic structure, i.e., $a^\varepsilon(x) = a\big(x, \frac{x}{\varepsilon}\big)$, where $a(x,y)$ is Ω-periodic in x and Y-periodic in y (Y is a reference cell, e.g. $Y = (0,1)^d$). For $T^\varepsilon = \varepsilon^{-2}T$, consider the wave equation: $u^\varepsilon : [0, T^\varepsilon] \times \mathbb{R}^d \to \mathbb{R}$ such that

$$\partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \big(a^\varepsilon(x)\nabla_x u^\varepsilon(t,x)\big) = f(t,x) \quad \text{in } (0,T^\varepsilon] \times \Omega, \tag{6.1}$$

with given initial position and speed $u^\varepsilon(0,x)$, $\partial_t u^\varepsilon(0,x)$ and periodic boundary conditions. For such tensor, homogenization theory still provides formula for the homogenized tensor. Namely, $a^0_{ij}(x) = \big\langle e_i^T a(x,\cdot)(\nabla_y \chi_j(x,\cdot) + e_j)\big\rangle_Y$, where $\{\chi_j(x,\cdot)\}_{j=1}^d$ are the solutions of local cell problems in Y (i.e., a different cell problems for every $x \in \Omega$). However, at timescales $\mathcal{O}(\varepsilon^{-2})$, some features of the macroscopic behavior of u^ε are not described by the homogenized solution. Hence, a new effective equation that describes these additional effects is needed.

In this chapter, we define a family of effective equations of the form

$$\partial_t^2 \tilde{u}(t,x) - \partial_i\big(a^0_{ij}(x)\partial_j \tilde{u}(t,x)\big) + \varepsilon L^1 \tilde{u}(t,x) + \varepsilon^2 L^2 \tilde{u}(t,x) = f(t,x) \quad \text{in } (0,T^\varepsilon] \times \Omega, \tag{6.2}$$

with the same initial conditions as u^ε and periodic boundary conditions. In one dimension, the operators are defined as

$$L^1 = 0, \qquad L^2 = \partial_x^2\big(a^{24}(x)\partial_x^2 \cdot\big) - \partial_x\big(b^{22}(x)\partial_x \partial_t^2 \cdot\big) - \partial_x\big(a^{22}(x)\partial_x \cdot\big) + b^{20}\partial_t^2,$$

where the formulas for $a^{24}, b^{22}, b^{20}, a^{22}$ only involve the first corrector $\chi(x,\cdot)$, the homogenized tensor $a^0(x)$, and are linked by a parameter. In the multidimensional case, we obtain the operators

$$L^1 = -\partial_i\big(a^{12}_{ij}(x)\partial_j \cdot\big) + b^{10}\partial_t^2, \quad L^2 = \partial^2_{ij}\big(a^{24}_{ijkl}(x)\partial^2_{kl}\cdot\big) - \partial_i\big(b^{22}_{ij}(x)\partial_j\partial_t^2\cdot\big) - \partial_i\big(a^{22}_{ij}(x)\partial_j\cdot\big) + b^{20}\partial_t^2,$$

where the formulas for the tensors a^{2i}, b^{2i} involve the first corrector $\{\chi_i(x,\cdot)\}_{i=1}^d$, a parameter, and two other correctors: $\{\theta^0_{ij}(x,\cdot)\}^d_{ij=1}$ and $\{\theta^1_i(x,\cdot)\}^d_{i=1}$. While $\theta^0_{ij}(x,\cdot)$ corresponds to a local version of the second order corrector obtained in the uniformly periodic case, $\theta^1_i(x,\cdot)$ is a new corrector originating from the variation $x \mapsto a(x,y)$. We verify that this family generalizes the

family obtained in the uniformly periodic case, in Chapter 4. Indeed, if the tensor has no variation in x, i.e., $a(x,y) = a(y)$, both families match.

The main result of the chapter is an error estimate validating the family of effective equations. Namely, under sufficient regularity of the data, we prove that any element $\tilde{u}$ of the family satisfies

$$\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} \leq C\varepsilon,$$

where the norm $\|\cdot\|_W$ is defined as (see (A.4))

$$\|w\|_W = \inf_{\substack{w=w_1+w_2 \\ w_1,w_2\in \mathrm{W}_{\mathrm{per}}(\Omega)}} \left\{ \|w_1\|_{\mathrm{L}^2(\Omega)} + \|\nabla w_2\|_{\mathrm{L}^2(\Omega)} \right\} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(\Omega).$$

As the dependence of the constant C on Ω is given explicitly, this result holds for arbitrarily large hypercubes Ω.

Let us explain how the operators L^1 and L^2 are derived. As in Chapter 4, we construct an adaptation of $\tilde{u}$ using asymptotic expansions. The adaptation involves correctors, which are the solutions of local (in x) cell problems in Y. While in Chapter 4 the form of the effective equation was a fixed ansatz, here we do not fix it a priori and construct L^1, L^2 as we match the different levels of the expansion. For each level, the well-posedness of the effective equation (6.2) constrains the form of the correction operators L^i, while the well-posedness of the obtained cell problems constrains them quantitatively. Compared to the uniformly periodic case, the dependency of the tensor on the slow variable $x \mapsto a(x,y)$ requires additional corrections in the adaptation. The repercussion of these new correctors is the apparition of additional operators in the effective equations.

Compared to the effective equations obtained in the uniformly periodic case, (6.2) contains the additional operators εL^1 and $\varepsilon^2 L^{2,1} = \varepsilon^2\big(b^{20}\partial_t^2 - \partial_i(a_{ij}^{22}\partial_j\cdot)\big)$ (in the general case). In particular, as $L^1 \neq 0$, a correction of the homogenized equation is already needed to obtain effective equations at timescales $\mathcal{O}(\varepsilon^{-1})$. However, in all the numerical examples that we considered, the effect of εL^1 is not significant. Furthermore, the importance of $\varepsilon^2 L^{2,1}$ is confirmed, but only in examples where the variation $x \mapsto a(x,y)$ is sharp. These facts suggest that, in certain applications, the operators εL^1 and $\varepsilon^2 L^{2,1}$ could be removed from the effective equations. This possibility is tempting as the computational cost for approximating the corresponding effective equations is significantly lower. Nevertheless, we could not derive a practical criterion to attest whether the removal of εL^1 and $\varepsilon^2 L^{2,1}$ can be done without affecting the order of accuracy.

The chapter is organized as follows. In Section 6.1, we discuss the modifications that are done in the derivation compared to the uniformly periodic case. Then, we define the family of effective equations in the one-dimensional case and present the complete derivation. Next, in Section 6.2, we state the main result of the chapter: we define the family of effective equations in the multidimensional case. In particular, we present the technical derivation of the cell problems and of the correction operators L^1 and L^2. In Section 6.3, we extend the validity of the family of effective equations to tensors with minimal regularity in the second variable. Next, in Section 6.4, the potential simplification of the effective equations is discussed. Finally, in Section 6.5, we test the different theoretical results of the chapter in various numerical examples.

6.1 Effective equations for locally periodic media in one dimension

In this section, we define a family of effective equations for locally periodic media in the one-dimensional case. The main result is presented in Section 6.1.2, where we provide an a priori error analysis ensuring that the elements of the family are ε-close to the oscillatory solution. The

derivation of the family is presented in Section 6.1.3 and the rigorous proof of the error estimate is provided in Section 6.1.4.

Let $a^\varepsilon(x) = a\big(x, \frac{x}{\varepsilon}\big)$ be a one-dimensional locally periodic tensor, where $a(x,y)$ is Y-periodic in y and Ω-periodic in x. The domain $\Omega \subset \mathbb{R}$ is arbitrarily large and assumed to be the union of cells of length $\varepsilon|Y|$ (see assumption (4.25), Figure 4.2). In particular, this assumption ensures that $a^\varepsilon(x)$ is Ω-periodic ($y \mapsto a(x,y)$ is extended by periodicity). For $T^\varepsilon = \varepsilon^{-2}T$, we consider the wave equation: $u^\varepsilon : [0,T^\varepsilon] \times \Omega \to \mathbb{R}$ such that

$$\begin{array}{ll} \partial_t^2 u^\varepsilon(t,x) - \partial_x\big(a\big(x,\frac{x}{\varepsilon}\big)\partial_x u^\varepsilon(t,x)\big) = f(t,x) & \text{in } (0,T^\varepsilon]\times\Omega, \\ x \mapsto u^\varepsilon(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,T^\varepsilon], \\ u^\varepsilon(0,x) = g^0(x), \quad \partial_t u^\varepsilon(0,x) = g^1(x) & \text{in } \Omega, \end{array} \tag{6.3}$$

where g^0, g^1 are the initial position and speed and f is a source. We denote the differential operator $\mathcal{A}^\varepsilon = -\partial_x\big(a\big(x,\frac{x}{\varepsilon}\big)\partial_x\,\cdot\,\big)$. We assume that $a(x,y)$ is uniformly elliptic and bounded, i.e. there exists $\lambda, \Lambda > 0$ such that

$$\lambda \le a(x,y) \le \Lambda \ \text{ for a.e. } (x,y) \in \Omega \times Y. \tag{6.4}$$

The well-posedness of problem (6.3) is proved in Section 2.1.1. If $g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega)$, $g^1 \in \mathrm{L}^2_0(\Omega)$, $f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{L}^2_0(\Omega))$, then there exists a unique weak solution $u^\varepsilon \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega))$ with $\partial_t u^\varepsilon \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2_0(\Omega))$ and $\partial_t^2 u^\varepsilon \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{W}^*_{\mathrm{per}}(\Omega))$.

6.1.1 Comment on the methodology for the construction of effective equation

In this section, we discuss the methodology of the derivation of the family of effective equations. In particular, we present the modifications that we operate compared to the uniformly periodic case.

Let us briefly recall how the effective coefficients are obtained for a uniformly periodic tensor in Chapter 4 (for simplicity, let us assume $f = 0$ and $d = 1$). We start with two ansatz. The first is that the effective equation has the form

$$\partial_t^2 \tilde{u} - a^0 \partial_x^2 \tilde{u} + \varepsilon^2 (a^2 \partial_x^4 \tilde{u} - b^2 \partial_x^2 \partial_t^2 \tilde{u}) = 0, \tag{6.5}$$

which is well-posed if $b^2, a^2 \ge 0$ (a^0 is the homogenized tensor). The second is that u^ε can be approximated by an adaptation of $\tilde{u}$, which takes the form

$$\mathcal{B}^\varepsilon \tilde{u}(t,x) = \tilde{u}(t,x) + \varepsilon\chi\big(\tfrac{x}{\varepsilon}\big)\partial_x\tilde{u}(t,x) + \varepsilon^2\theta\big(\tfrac{x}{\varepsilon}\big)\partial_x^2\tilde{u}(t,x) + \varepsilon^3\kappa\big(\tfrac{x}{\varepsilon}\big)\partial_x^3\tilde{u}(t,x) + \varepsilon^4\rho\big(\tfrac{x}{\varepsilon}\big)\partial_x^4\tilde{u}(t,x). \tag{6.6}$$

We impose $\mathcal{B}^\varepsilon\tilde{u}$ to solve the same equation as u^ε, up to a remainder. We thus obtain the definition of the correctors $\chi, \theta, \kappa, \rho$ as the solutions of cell problems, which are elliptic PDEs in the reference cell Y, with periodic boundary conditions. These cell problems must be well-posed in the quotient space $\mathcal{W}_{\mathrm{per}}(Y)$. We verify that the cell problems for χ and κ are well-posed unconditionally. Furthermore, the well-posedness of the cell problem for θ is guaranteed by the definition of the homogenized tensor a^0. Finally, the well-posedness of the cell problem for ρ imposes a constraint for the definition of b^2, a^2. Namely, we need the following equality to hold:

$$a^0 b^2 - a^2 = a^0\langle\chi^2\rangle_Y - a^0\langle\chi\rangle_Y^2. \tag{6.7}$$

The family of effective equations is then defined by the pairs of coefficients $b^2, a^2 \ge 0$ satisfying (6.7). To find such pairs, two different processes lead to the same family. The first, used in Section 4.3.1, is to define $b^2 = \langle\chi^2\rangle_Y$, $a^2 = a^0\langle\chi\rangle_Y^2$, and observe that each value of $\langle\chi\rangle_Y$ leads to

a different valid pair b^2, a^2. An alternative way to derive such pairs is, after rewriting (6.7) as $a^0 b^2 - a^2 = a^0 \big\langle (\chi - \langle \chi \rangle_Y)^2 \big\rangle_Y$, to define

$$b^2 = \big\langle (\chi - \langle \chi \rangle_Y)^2 \big\rangle_Y + s, \qquad a^2 = a^0 s, \tag{6.8}$$

for some parameter $s \geq 0$. In this case we can fix $\langle \chi \rangle_Y = 0$. Even though these two ways are equivalent and lead to the same parametrized family of effective equations, we can note the following differences. First, the corresponding constants C in the error estimate $\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} \leq C\varepsilon$ (Theorem 4.2.4) are different. The constant C grows in a simpler way with respect to the parameter in the case where $\langle \chi \rangle_Y = 0$ is fixed. Indeed, we verify that C depends on a^2, b^2, but also on $\|\chi\|_{\mathcal{C}^1(\bar{Y})}, \|\theta\|_{\mathcal{C}^1(\bar{Y})}, \|\kappa\|_{\mathcal{C}^1(\bar{Y})}, \|\rho\|_{\mathcal{C}^1(\bar{Y})}$ and these quantities depend on $\langle \chi \rangle_Y$. Note also that in the multidimensional case, varying the parameters $\langle \chi_i \rangle_Y$ does not necessarily lead to well-posed effective equations (see Section 4.3.5), in which case setting $\langle \chi_i \rangle_Y \neq 0$ is superfluous. Note that in the locally periodic setting, an additional question for the choice of normalization is brought by the variation in x. Indeed, as we deal with a corrector that depends on the slow variable x, $\chi(x, y)$, if the normalization is not fixed, we must make sense of $\partial_x \langle \chi(x) \rangle_Y = \langle \partial_x \chi(x) \rangle_Y$. Finally, note that setting $\langle \chi_i \rangle_Y = 0$ is more consistent in the following sense: in the special case of a constant tensor $a(y) = a$, the natural requirement $\mathcal{B}^\varepsilon \tilde{u} = u^\varepsilon = \tilde{u}$ holds if and only if $\langle \chi_i \rangle_Y = 0$. Following these considerations, in the whole section (and in the whole chapter), we make the following assumption:

$$\text{all the correctors have zero mean.} \tag{H1}$$

Let us now summarize how we construct effective equations for a locally periodic tensor in one dimension (the full derivation is presented in Section 6.1.3). We still assume for simplicity that $f = 0$. First, we make the ansatz that the effective equation has the form

$$\partial_t^2 \tilde{u} - \partial_x \big(a^0(x) \partial_x \tilde{u} \big) + \varepsilon L^1 \tilde{u} + \varepsilon^2 L^2 \tilde{u} = 0, \tag{6.9}$$

where $a^0(x)$ is the homogenized tensor and L^1, L^2 are differential operators to be defined. Then, we construct an adaptation $\mathcal{B}^\varepsilon \tilde{u}$ of $\tilde{u}$ that solves the same equation as u^ε up to a remainder of order $\mathcal{O}(\varepsilon^3)$. The adaptation takes the form

$$\begin{aligned} \mathcal{B}^\varepsilon \tilde{u}(t,x) = \tilde{u}(t,x) + \varepsilon \chi\big(x, \tfrac{x}{\varepsilon}\big) \partial_x \tilde{u}(t,x) + \varepsilon^2 \sum_{i=0}^{1} \theta_i\big(x, \tfrac{x}{\varepsilon}\big) \partial_x^{2-i} \tilde{u}(t,x) \\ + \varepsilon^3 \sum_{i=0}^{2} \kappa_i\big(x, \tfrac{x}{\varepsilon}\big) \partial_x^{3-i} \tilde{u}(t,x) + \varepsilon^4 \sum_{i=0}^{3} \rho_i\big(x, \tfrac{x}{\varepsilon}\big) \partial_x^{4-i} \tilde{u}(t,x), \end{aligned} \tag{6.10}$$

where the correctors $\chi, \theta_i, \kappa_i, \rho_i$ are solutions of cell problems. Observe that compared to (6.6), the adaptation (6.10) contains more correctors. They come from the dependence of $a(x, y)$ on the slow variable x. The differential operators L^1 and L^2 are then defined to satisfy two conditions. First, the coefficients involved in L^1, L^2 must verify the constraints given by the well-posedness of the cell problems. Second, they must ensure that (6.9) is well-posed. After some technical simplifications, we obtain $L^1 = 0$ and

$$L^2 = \partial_x^2 (a^{24}(x) \partial_x^2 \cdot) - \partial_x (b^{22}(x) \partial_x \partial_t^2 \cdot) - \partial_x (a^{22}(x) \partial_x \cdot) + b^{20} \partial_t^2,$$

where $a^{24}(x), b^{22}(x)$ must satisfy a constraint similar to (6.7) for all $x \in \Omega$ and $a^{22}(x), b^{20}$ a new constraint. Note that the cancelation of L^1 is specific to the one-dimensional case. In the multidimensional case, in Section 6.2, we verify that in general $L^1 \neq 0$.

6.1.2 Error estimate and family of effective equations in one dimension

We state here the main result of this section and define the family of effective equations in one dimension. The derivation of the cell problems and of the corresponding constraints on the

effective tensors is presented in Section 6.1.3 and the rigorous proof of the error estimate is provided in Section 6.1.4.

Let us define the effective equations. For all $x \in \Omega$, let $\chi(x) = \chi(x,\cdot) \in \mathrm{W}_{\mathrm{per}}(Y)$ be the unique solution of

$$\big(a(x)\partial_y\chi(x), \partial_y w\big)_Y = -\big(a(x), \partial_y w\big)_Y \quad \forall w \in \mathrm{W}_{\mathrm{per}}(Y), \tag{6.11}$$

and let $a^0(x)$ be the homogenized tensor defined for all $x \in \Omega$ by

$$a^0(x) = \big\langle a(x)(\partial_y\chi(x) + 1)\big\rangle_Y. \tag{6.12}$$

We emphasize that $\big\langle \chi(x)\big\rangle_Y = 0$ (assumption (H1)). Let then $b^{20}(x), a^{22}(x), a^{24}(x), b^{22}(x)$ be Ω-periodic coefficients that satisfy

$$b^{20}, a^{22}, a^{24}, b^{22} \in \mathrm{L}^\infty_{\mathrm{per}}(\Omega), \qquad b^{20}(x), a^{22}(x), a^{24}(x), b^{22}(x) \geq 0 \quad \text{for a.e. } x \in \Omega. \tag{6.13}$$

Let $\tilde{u} : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ be the solution of the equation

$$\begin{array}{ll} \partial_t^2\tilde{u} - \partial_x(a^0\partial_x\tilde{u}) + \varepsilon^2\big(\partial_x^2(a^{24}\partial_x^2\tilde{u}) - \partial_x(b^{22}\partial_x\partial_t^2\tilde{u}) - \partial_x(a^{22}\partial_x\tilde{u}) + b^{20}\partial_t^2\tilde{u}\big) = f & \text{in } (0,T^\varepsilon]\times\Omega, \\ x \mapsto \tilde{u}(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,T^\varepsilon], \\ \tilde{u}(0,x) = g^0(x), \quad \partial_t\tilde{u}(0,x) = g^1(x) & \text{in } \Omega. \end{array} \tag{6.14}$$

As the homogenized tensor is elliptic and bounded (see Lemma 3.3.1), and as (6.13) holds, if the data satisfy the regularity $a^{24} \in \mathrm{W}^{1,\infty}(\Omega)$, $g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^2(\Omega)$, $g^1 \in \mathrm{L}^2_0(\Omega) \cap \mathrm{H}^1(\Omega)$, $f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{L}^2_0(\Omega))$, then there exists a unique weak solution of (6.14) (see Section 2.1.2). The main result of this section is the error estimate provided by the following theorem.

Theorem 6.1.1. *Assume* (H1) *and that the tensor satisfies* $a \in \mathcal{C}^1(\bar{\Omega};\mathrm{W}^{1,\infty}(Y)) \cap \mathcal{C}^4(\bar{\Omega};\mathrm{L}^\infty(Y))$. *Furthermore, assume that the solution* $\tilde{u}$ *of* (6.14)*, the initial conditions and the source term satisfy the regularity*

$$\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^5(\Omega)), \quad \partial_t\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^4(\Omega)), \quad \partial_t^2\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^3(\Omega)),$$
$$g^0 \in \mathrm{H}^4(\Omega), \quad g^1 \in \mathrm{H}^4(\Omega), \quad f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{H}^2(\Omega)).$$

Let $\chi(x,\cdot) \in \mathrm{W}_{\mathrm{per}}(Y)$ *be the solution of* (6.11) *and assume that the coefficients of* (6.14) *are defined, for some* $r \geq 0$*, as*

$$\begin{array}{ll} a^{24}(x) = r a^0(x)^2, & b^{22}(x) = \big\langle \chi(x)^2\big\rangle_Y + r a^0(x), \\ b^{20} = r\max_{x\in\Omega}\{\partial_x^2 a^0(x)\}, & a^{22}(x) = -r a^0(x)\partial_x^2 a^0(x) + b^{20}a^0(x). \end{array} \tag{6.15}$$

Then the following error estimate holds

$$\begin{aligned} \|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} \leq C\varepsilon\Big(& \|g^1\|_{\mathrm{H}^4(\Omega)} + \|g^0\|_{\mathrm{H}^4(\Omega)} + \|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^2(\Omega))} \\ & + \textstyle\sum_{k=1}^5 |\tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^k(\Omega))} + \|\partial_t^2\tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^3(\Omega))}\Big), \end{aligned} \tag{6.16}$$

where C *depends only on* T, λ, Y, $\|a\|_{\mathcal{C}^1(\bar{\Omega};\mathrm{W}^{1,\infty}(Y))}$, $\|a\|_{\mathcal{C}^4(\bar{\Omega};\mathrm{L}^\infty(Y))}$, *and* r*, and we recall the definition of the norm (see* (A.4)*)*

$$\|w\|_W = \inf_{\substack{w=w_1+w_2 \\ w_1,w_2\in\mathrm{W}_{\mathrm{per}}(\Omega)}} \Big\{\|w_1\|_{\mathrm{L}^2(\Omega)} + \|\nabla w_2\|_{\mathrm{L}^2(\Omega)}\Big\} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(\Omega).$$

Thanks to Theorem 6.16, we define the family of effective equations.

Definition 6.1.2. The family of effective equations $\mathcal{E}$ is the set of equations (6.14), where a^0 is the homogenized tensor, defined in (6.12), and the coefficients $b^{20}, a^{22}, a^{24}, b^{22}$ are defined in (6.15) for some parameter $r \geq 0$.

Remark 6.1.3. As proved in the multidimensional case in Section 6.3, an error estimate still holds for a tensor $a \in \mathcal{C}^4(\bar{\Omega}; \mathrm{L}^\infty(Y))$, if we assume

$$\tilde{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^6(\Omega)), \quad \partial_t \tilde{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^5(\Omega)), \quad \partial_t^2 \tilde{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^4(\Omega)),$$
$$g^0 \in \mathrm{H}^5(\Omega), \quad g^1 \in \mathrm{H}^5(\Omega), \quad f \in \mathrm{L}^2(0, T^\varepsilon; \mathrm{H}^3(\Omega)).$$

To prove it, we use the Sobolev embedding $\mathrm{H}^1(\Omega) \hookrightarrow \mathcal{C}^0(\bar{\Omega})$ and Lemma 6.3.2.

Remark 6.1.4. The family $\mathcal{E}$, defined in Definition 6.1.2, generalizes the family obtained for a one-dimensional uniformly periodic tensor in Section 4.3.1. Indeed, if the tensor does not depend on the slow variable, i.e., $a(x, y) = a(y)$, then, we verify that $\chi(x, y)$ and $a^0(x)$ are constant in x. Hence, a^{22}, b^{20} vanish and a^{24}, b^{22} are constant. The effective equation (6.14) is thus left with the single correction $\varepsilon^2(a^{24}\partial_x^4 - b^{22}\partial_x^2\partial_t^2)$, which has the same form as in the uniformly periodic case (see (6.5)). In the uniformly periodic case, the family is defined by the pairs $b^2 = \langle \chi^2 \rangle_Y$, $a^2 = a^0 \langle \chi \rangle_Y^2$, parametrized by $\langle \chi \rangle_Y \in \mathbb{R}$ (see (6.7)). We verify that the pairs a^{24}, b^{22}, defined by (6.15), are the same ($a^{24} = a^2$, $b^{22} = b^2$) via the following relation between the parameters: $\langle \chi \rangle_Y^2 = r a^0$. Furthermore, for the alternative definition of the family given in (6.8), the pairs match via the relation $s = r a^0$.

6.1.3 Derivation of the adaptation operator and of the effective equations

In this section, we present the full derivation of the family of effective equations defined in Definition 6.1.2. The derivation follows the plan described in Section 6.1.1. In particular, we derive the cell problems for the correctors that are necessary to define the adaptation operator used in the proof of Theorem 6.1.1. Recall that the well-posedness of these cell problems provides constraints on the effective coefficients. Let us recall that we assume all the correctors to have zero mean (assumption (H1)).

The result of the section is synthesized in the following theorem.

Theorem 6.1.5. *Let $\tilde{u}$ belong to the family $\mathcal{E}$ (Definition 6.1.2). Then there exists an adaptation of the form*

$$\mathcal{B}^\varepsilon \tilde{u}(t, x) = \tilde{u}(t, x) + \varepsilon u^1\big(t, x, \tfrac{x}{\varepsilon}\big) + \varepsilon^2 u^2\big(t, x, \tfrac{x}{\varepsilon}\big) + \varepsilon^3 u^3\big(t, x, \tfrac{x}{\varepsilon}\big) + \varepsilon^4 u^4\big(t, x, \tfrac{x}{\varepsilon}\big) + \varphi(t, x), \quad (6.17)$$

such that $x \mapsto \mathcal{B}^\varepsilon \tilde{u}(t, x)$ is Ω-periodic and

$$(u^\varepsilon - \mathcal{B}^\varepsilon \tilde{u})(0) = \mathcal{O}(\varepsilon), \qquad \partial_t (u^\varepsilon - \mathcal{B}^\varepsilon \tilde{u})(0) = \mathcal{O}(\varepsilon), \quad (6.18a)$$
$$(\partial_t^2 + \mathcal{A}^\varepsilon)(u^\varepsilon - \mathcal{B}^\varepsilon \tilde{u})(t) = \mathcal{O}(\varepsilon^3) \quad \text{for a.e. } t \in [0, T^\varepsilon], \quad (6.18b)$$

where we denoted $\mathcal{A}^\varepsilon = -\partial_x\big(a\big(x, \tfrac{x}{\varepsilon}\big)\partial_x \cdot\big)$.

Thanks to Theorem 6.1.5, and in particular to (6.18), the adaptation can be used in the process described in Section 4.2.2 to prove that $\tilde{u}$ is close to u^ε in the $\mathrm{L}^\infty(0, T^\varepsilon; W)$ norm.

In the rest of the section, we proceed with the construction of the adaptation $\mathcal{B}^\varepsilon \tilde{u}$ and of the effective equations. In particular, we need to define the functions u^k and φ in (6.67) so that (6.18) holds. Note that in contrast to the uniformly periodic case in Chapter 4, we do not have an a priori knowledge on the form of the higher order operators needed in the effective equation.

Consequently, we construct the higher order operators at the same time as we cancel the levels in the asymptotic expansion.

Let us now construct explicitly the adaptation and derive the constraint on the effective operators. We make the ansatz that the effective equation has the form

$$\begin{array}{ll} \partial_t^2\tilde{u} - \partial_x\big(a^0(x)\partial_x\tilde{u}\big) + \varepsilon\tilde{L}^1\tilde{u} + \varepsilon^2\tilde{L}^2\tilde{u} = f & \text{in } (0,T^\varepsilon]\times\Omega, \\ x \mapsto \tilde{u}(t,x) \quad \Omega\text{-periodic} & \text{in } [0,T^\varepsilon], \\ \tilde{u}(0,x) = g^0(x), \quad \partial_t\tilde{u}(0,x) = g^1(x) & \text{in } \Omega, \end{array} \tag{6.19}$$

where a^0 is the homogenized tensor (defined in (6.12) and (6.26) below) and $\tilde{L}^1, \tilde{L}^2$ are linear, ε-independent differential operators to be defined. Next, we make the ansatz that the adaptation $\mathcal{B}^\varepsilon\tilde{u}$ has the form (6.17), where $u^i(t,x,y)$ are unknown operators of $\tilde{u}$, Ω-periodic in x and Y-periodic in y, and φ is an unknown operator of f. Let us introduce the differential operators

$$\mathcal{A}_{yy} = -\partial_y\big(a(x,y)\partial_y\,\cdot\,\big), \quad \mathcal{A}_{yx} = -\partial_y\big(a(x,y)\partial_x\,\cdot\,\big) - \partial_x\big(a(x,y)\partial_y\,\cdot\,\big), \quad \mathcal{A}_{xx} = -\partial_x\big(a(x,y)\partial_x\,\cdot\,\big).$$

For $\psi(x,y)$ smooth enough, we verify that $\mathcal{A}^\varepsilon\psi\big(x,\frac{x}{\varepsilon}\big) = \big(\varepsilon^{-2}\mathcal{A}_{yy} + \varepsilon^{-1}\mathcal{A}_{xy} + \mathcal{A}_{xx}\big)\psi\big(x,\frac{x}{\varepsilon}\big)$. Using (6.3), (6.19) and (6.17), we obtain the development

$$\begin{aligned} R^\varepsilon = (\partial_t^2 + \mathcal{A}^\varepsilon)(\mathcal{B}^\varepsilon\tilde{u} - u^\varepsilon)(t,x) &= \partial_t^2\mathcal{B}^\varepsilon\tilde{u}(t,x) + \mathcal{A}^\varepsilon\mathcal{B}^\varepsilon\tilde{u}(t,x) - f(t,x) \\ &= \quad \varepsilon^{-1}\big(\qquad \mathcal{A}_{yy}u^1 + \mathcal{A}_{xy}\tilde{u} \qquad\qquad\qquad\quad \big) \\ &\quad + \varepsilon^0 \ \big(\qquad \mathcal{A}_{yy}u^2 + \mathcal{A}_{xy}u^1 + \mathcal{A}_{xx}\tilde{u} + \partial_x(a^0\partial_x\tilde{u}) \big) \\ &\quad + \varepsilon^1 \ \big(\partial_t^2u^1 + \mathcal{A}_{yy}u^3 + \mathcal{A}_{xy}u^2 + \mathcal{A}_{xx}u^1 - \tilde{L}^1\tilde{u} \qquad \big) \\ &\quad + \varepsilon^2 \ \big(\partial_t^2u^2 + \mathcal{A}_{yy}u^4 + \mathcal{A}_{xy}u^3 + \mathcal{A}_{xx}u^2 - \tilde{L}^2\tilde{u} \qquad \big) \\ &\quad + (\partial_t^2 + \mathcal{A}^\varepsilon)\varphi + \mathcal{O}(\varepsilon^3), \end{aligned} \tag{6.20}$$

where the u^i are evaluated at $\big(t,x,y=\frac{x}{\varepsilon}\big)$. We now successively find $u^1,\dots,u^4$ and φ such that the terms of order $\mathcal{O}(\varepsilon^{-1})$ to $\mathcal{O}(\varepsilon^2)$ in (6.20) vanish. Note that the u^k are set to cancel the terms containing $\tilde{u}$ and φ is set to cancel the terms containing f that will appear.

Canceling the ε^{-1}, ε^0 and ε terms and derivation of the constraints defining $\tilde{L}^1$

Canceling the ε^{-1} order term in (6.20) leads to defining

$$u^1(t,x,y) = \chi(x,y)\partial_x\tilde{u}(t,x) + \tilde{u}^1(t,x), \tag{6.21}$$

where for all $x\in\Omega$, $\chi(x,y)$ is Y-periodic in y and solves the cell problem

$$-\partial_y\big(a(x,y)(\partial_y\chi(x,y)+1)\big) = 0.$$

Let us write the weak formulation of the cell problem in $\mathrm{W}_{\mathrm{per}}(Y)$: for all $x\in\Omega$, $\chi(x) = \chi(x,\cdot)\in \mathrm{W}_{\mathrm{per}}(Y)$ is the solution of

$$\varepsilon^{-1}: \quad \big(a(x)\partial_y\chi(x), \partial_y w\big)_Y = -\big(a(x), \partial_y w\big)_Y, \tag{6.22}$$

for all test functions $w\in\mathrm{W}_{\mathrm{per}}(Y)$. Observe that for a fixed $x\in\Omega$, (6.22) is the same cell problem as obtained at order ε^{-1} in the periodic case (see (4.45a)). To simplify, we let $\tilde{u}^1(t,x) = 0$ in (6.21). Using the definition of u^1, the term of order ε^0 in (6.20) reads

$$\begin{aligned} \mathcal{A}_{yy}u^2(t,x,y) &+ \Big(-\partial_y\big(a(x,y)(\chi(x,y)+1) - a(x,y)(\partial_y\chi(x,y)+1) + a^0(x)\big)\Big)\partial_x^2\tilde{u}(t,x) \\ &+ \Big(-\partial_y\big(a(x,y)(\partial_x\chi(x,y)+1) - \partial_x\big(a(x,y)(\partial_y\chi(x,y)+1)\big) + \partial_x a^0(x)\Big)\partial_x\tilde{u}(t,x). \end{aligned}$$

In order to cancel this term, it is sufficient to define

$$u^2(t,x,y) = \theta_0(x,y)\partial_x^2\tilde{u}(t,x) + \theta_1(x,y)\partial_x\tilde{u}(t,x), \tag{6.23}$$

where for all $x \in \Omega$, $\theta_0(x) = \theta_0(x,\cdot), \theta_1(x) = \theta_1(x,\cdot)$ are the solutions in $\mathrm{W}_{\mathrm{per}}(Y)$ of the cell problems

$$\varepsilon^0 :$$
$$\big(a(x)\partial_y\theta_0(x), \partial_y w\big)_Y = -\big(a(x)\chi(x), \partial_y w\big)_Y + \big(a(x)(\partial_y\chi(x)+1) - a^0(x), w\big)_Y, \tag{6.24a}$$
$$\big(a(x)\partial_y\theta_1(x), \partial_y w\big)_Y = -\big(a(x)\partial_x\chi(x), \partial_y w\big)_Y + \big(\partial_x\big(a(x)(\partial_y\chi(x)+1)\big) - \partial_x a^0(x), w\big)_Y, \tag{6.24b}$$

for all test functions $w \in \mathrm{W}_{\mathrm{per}}(Y)$. While for fixed $x \in \Omega$, (6.24a) corresponds to the cell problem obtained at order ε^0 in the periodic case (see (4.45b)), (6.24b) is a new cell problem coming from the variation of the tensor in the slow variable x. In order to verify that equations (6.24) are well-posed in $\mathrm{W}_{\mathrm{per}}(Y)$, we apply Lax–Milgram theorem. In particular, we need to show that the right hand sides belong to $\mathrm{W}^*_{\mathrm{per}}(\Omega)$. Referring to Appendix A.2, $F \in [\mathrm{H}^1_{\mathrm{per}}(Y)]^*$ given by

$$\big\langle F, w\big\rangle = \big(f^0, w\big)_{\mathrm{L}^2(Y)} + \big(f^1, \partial_x w\big)_{\mathrm{L}^2(Y)},$$

for some $f^0, f^1 \in \mathrm{L}^2(Y)$ belongs to $\mathrm{W}^*_{\mathrm{per}}(Y)$ if and only if

$$\big(f^0, 1\big)_{\mathrm{L}^2(Y)} = 0. \tag{6.25}$$

We thus have to verify that the right hand sides of the cell problems (6.24) satisfy (6.25). For (6.24a), the condition is satisfied as the definition of the homogenized tensor at $x \in \Omega$ is

$$a^0(x) = \big\langle a(x)\partial_y\chi(x) + 1\big)\big\rangle_Y. \tag{6.26}$$

To verify the solvability of (6.24b), observe that for a sufficiently regular tensor a, it holds

$$\partial_x a^0(x) = \big\langle \partial_x\big(a(x)\partial_y\chi(x)+1\big)\big)\big\rangle_Y.$$

Note that at this point we can prove the classical homogenization result at short times $T = \mathcal{O}(1)$, for a locally periodic tensor (under suitable regularity assumptions, see Section 4.2.2). Indeed, the current adaptation (6.17), with $u^3 = u^4 = 0$, solves the same equation as u^ε up to a reminder of order ε. As we look for an adaptation with a remainder of order ε^3 (see (6.18b)), we carry on with the asymptotic expansion. Taking into account the definitions of u^1, u^2, we have

$$\begin{aligned}
\partial_t^2 u^1 &= \chi\partial_x f + \chi\partial_x^2(a^0\partial_x\tilde{u}) - \varepsilon\chi\partial_x(\tilde{L}^1\tilde{u}) + \mathcal{O}(\varepsilon^2),\\
\partial_t^2 u^2 &= \theta_0\partial_x^2 f + \theta_1\partial_x f + \theta_0\partial_x^3(a^0\partial_x\tilde{u}) + \theta_1\partial_x^2(a^0\partial_x\tilde{u}) + \mathcal{O}(\varepsilon),
\end{aligned}$$

and (6.20) can be rewritten as

$$\begin{aligned}
R^\varepsilon = \quad & \varepsilon^1\big(\mathcal{A}_{yy}u^3 + \mathcal{A}_{xy}u^2 + \mathcal{A}_{xx}u^1 + \chi\partial_x^2(a^0\partial_x\tilde{u}) - \tilde{L}^1\tilde{u}\big)\\
& + \varepsilon^2\big(\mathcal{A}_{yy}u^4 + \mathcal{A}_{xy}u^3 + \mathcal{A}_{xx}u^2 + \theta_0\partial_x^3(a^0\partial_x\tilde{u}) + \theta_1\partial_x^2(a^0\partial_x\tilde{u}) - \chi\partial_x(\tilde{L}^1\tilde{u}) - \tilde{L}^2\tilde{u}\big)\\
& + (\partial_t^2 + \mathcal{A}^\varepsilon)\varphi + \varepsilon\chi\partial_x f + \varepsilon^2\big(\theta_0\partial_x^2 f + \theta_1\partial_x f\big) + \mathcal{O}(\varepsilon^3).
\end{aligned} \tag{6.27}$$

As done in Section 4.2.3, we will deal with the terms coming from the right hand side f separately. Canceling the ε^1 order term of (6.27) leads similarly as for order ε^0 to defining

$$u^3(t,x,y) = \kappa_0(x,y)\partial_x^3\tilde{u}(t,x) + \kappa_1(x,y)\partial_x^2\tilde{u}(t,x) + \kappa_2(x,y)\partial_x\tilde{u}(t,x). \tag{6.28}$$

Furthermore, in order to ensure the well-posedness of the obtained cell problems for the κ_i, we let $\tilde{L}^1$ have the form $\tilde{L}^1 = a^{13}(x)\partial_x^3 + a^{12}(x)\partial_x^2 + a^{11}(x)\partial_x$, where the coefficients $a^{1i}(x)$ are to be defined. The variational formulations of the obtained cell problems are then: for all $x \in \Omega$, $\kappa_0(x,\cdot), \kappa_1(x,\cdot), \kappa_2(x,\cdot)$ are the solutions in $\mathrm{W}_{\mathrm{per}}(Y)$ of the cell problems (we drop the notation of the evaluation at x for readability)

$$\varepsilon^1: \quad \big(a\partial_y\kappa_0, \partial_y w\big)_Y = -\big(a\theta_0, \partial_y w\big)_Y + \big(a(\partial_y\theta_0 + \chi) - a^0\chi + a^{13}, w\big)_Y, \tag{6.29a}$$

$$\begin{aligned}\big(a\partial_y\kappa_1, \partial_y w\big)_Y = &-\big(a(\partial_x\theta_0 + \theta_1), \partial_y w\big)_Y + \big(a(\partial_y\theta_1 + \partial_x\chi), w\big)_Y \\ &+ \big(\partial_x(a(\partial_y\theta_0 + \chi)), w\big)_Y + \big(-2\partial_x a^0\chi + a^{12}, w\big)_Y,\end{aligned} \tag{6.29b}$$

$$\big(a\partial_y\kappa_2, \partial_y w\big)_Y = -\big(a\partial_x\theta_1, \partial_y w\big)_Y + \big(\partial_x(a(\partial_y\theta_1 + \partial_x\chi)), w\big)_Y + \big(-\partial_x^2 a^0\chi + a^{11}, w\big)_Y, \tag{6.29c}$$

for all test functions $w \in \mathrm{W}_{\mathrm{per}}(Y)$. Again, we note that for a fixed $x \in \Omega$, (6.29a) is the same as the cell problem (4.45c) obtained in the uniformly periodic case, while (6.29b) and (6.29c) are new cell problems. The coefficients of $\tilde{L}^1$ are then defined so that the cell problems (6.29) are well-posed in $\mathrm{W}_{\mathrm{per}}(Y)$, i.e., such that the right hand sides satisfy (6.25): for all $x \in \Omega$, $a^{13}(x), a^{12}(x)$ and $a^{11}(x)$ are defined as (assumption (H1) implies $\langle\chi(x)\rangle_Y = 0$)

$$a^{13}(x) = -\big\langle a(x)(\partial_y\theta_0(x) + \chi(x))\big\rangle_Y, \tag{6.30a}$$

$$a^{12}(x) = -\big\langle a(x)(\partial_y\theta_1(x) + \partial_x\chi(x))\big\rangle_Y - \partial_x\big\langle a(x)(\partial_y\theta_0(x) + \chi(x))\big\rangle_Y, \tag{6.30b}$$

$$a^{11}(x) = -\partial_x\big\langle a(x)(\partial_y\theta_1(x) + \partial_x\chi(x))\big\rangle_Y. \tag{6.30c}$$

These constraints are simplified in the following Lemma.

Lemma 6.1.6. *Under assumption* (H1)*, the coefficients defined in* (6.30) *satisfy for all* $x \in \Omega$

$$a^{13}(x) = 0, \qquad a^{12}(x) = 0, \qquad a^{11}(x) = 0.$$

Proof. Let $x \in \Omega$ be fixed and recall that $\langle\chi(x)\rangle_Y = 0$. As $a(x,\cdot)(1 + \partial_y\chi(x,\cdot)) \in \mathrm{H}(\mathrm{div}, Y)$, using integration by parts and equation (6.22), we obtain for any $y_1, y_2 \in Y$,

$$a(x,y)(\partial_y\chi(x,y) + 1)\Big|_{y=y_1}^{y_2} = -\int_Y \big(H_{y_2} - H_{y_1}\big)\partial_y\big(a(x,y)(\partial_y\chi(x,y) + 1)\big)\mathrm{d}y = 0,$$

where H_y is the Heaviside step function centered in y. Hence, the function $y \mapsto a(x,y)(\partial_y\chi(x,y) + 1)$ is constant. The definition of a^0 in (6.26) then implies

$$a(x,y)(\partial_y\chi(x,y) + 1) = a^0(x) \quad \forall (x,y) \in \Omega \times Y. \tag{6.31}$$

Dividing this equality by $a(x,y)$ and taking the mean over Y, we obtain the expression $a^0(x) = 1/\langle 1/a(x,\cdot)\rangle_Y$. Consider now equation (6.24a). A similar argument implies that $y \mapsto a(x,y)(\partial_y\theta_0(x,y) + \chi(x,y))$ is constant, i.e.,

$$a(x,y)\big(\partial_y\theta_0(x,y) + \chi(x,y)\big) = C(x).$$

Dividing the equality by $a(x,y)$ and taking the mean in y over Y, we verify that

$$a(x,y)\big(\partial_y\theta_0(x,y) + \chi(x,y)\big) = a^0(x)\big\langle\chi(x)\big\rangle_Y = 0 \quad \forall (x,y) \in \Omega \times Y. \tag{6.32}$$

In the same way, we can prove from (6.24b) that

$$a(x,y)\big(\partial_y\theta_1(x,y) + \partial_x\chi(x,y)\big) = a^0(x)\big\langle\partial_x\chi(x)\big\rangle_Y = a^0(x)\partial_x\big\langle\chi(x)\big\rangle_Y = 0 \quad \forall (x,y) \in \Omega \times Y. \tag{6.33}$$

Using equalities (6.31), (6.32) and (6.33) in the definitions (6.30) proves the result of the lemma. □

Lemma 6.1.6 implies that $\tilde{L}^1 = 0$. Hence, there is no correction of order ε in the effective equation (6.19). This fact ensures that the standard homogenized solution u^0 approximates well u^ε for timescales $\mathcal{O}(\varepsilon^{-1})$. Note that this result is specific to the one-dimensional case. In higher dimensions, we will see in Section 6.2 that the operator $\tilde{L}^1$ obtained in the same way does not vanish in general.

Canceling the ε^2 terms and derivation of the constraints defining $\tilde{L}^2$

Let us come back to the asymptotic expansion. We need to cancel the ε^2 order term in (6.27). As for the first terms, we thus define

$$u^4(t,x,y) = \sum_{i=0}^{3} \rho_i(x,y)\partial_x^{4-i}\tilde{u}(t,x), \tag{6.34}$$

where $\rho_0, \dots, \rho_3$ are the solutions of cell problems. We now need to define the operator $\tilde{L}^2$. To design it, we focus on two points: the well-posedness of the cell problems and the well-posedness of equation (6.19). The first point is familiar by now and consists in enforcing the solvability condition (6.25) to the cell problems. The latter is connected to what was done in Chapter 4 to obtain well-posed effective equation. For the well-posedness of such hyperbolic linear equations, we refer to Section 2.1.2. We have to introduce enough terms in $\tilde{L}^2$ so that the well-posedness of the cell problems for $\{\rho_i(x,y)\}_{i=0}^3$ can be ensured. As we have 4 cell problems, the first idea is to try with $\tilde{L}^2 = \sum_{i=1}^4 a^{2i}(x)\partial_x^i$. However, doing so leads to the definition $a^{24}(x) = -a^0(x)\langle\chi(x)^2\rangle_Y$ and provokes the ill-posedness of (6.19). Hence, this definition for $\tilde{L}^2$ is not adequate. Nevertheless, a similar issue has been solved in the uniformly periodic case using a Boussinesq trick. The trick consists in using the equation (6.19) at order $\mathcal{O}(1)$, i.e., $\partial_t^2\tilde{u} = f + \partial_x(a^0\partial_x\tilde{u})$ and take advantage of the sign of a^0. Explicitly, we introduce the additional operator $-\partial_x\big(b^{22}(x)\partial_x\partial_t^2\cdot\big)$ in $\tilde{L}^2$. Then, replacing $\partial_t^2\tilde{u}$ with $f + \partial_x\big(a^0(x)\partial_x\tilde{u}\big)$, we obtain a constraint on the difference $a^{24} - a^0b^{22}$, which can be satisfied by pairs of non-negative coefficients a^{24}, b^{22}. Each pair corresponds to a well-posed effective equation. This reflexion indicates that the initial definition of $\tilde{L}^2$ must have as many liberties as possible, i.e., as many different operators as possible. Note that a similar issue appears for the term $a^{22}\partial_x^2$ of $\tilde{L}^2$. Namely, if $a^{22} > 0$ the ellipticity of the second order operator in (6.19) is weakened and could even break. Adding the term $b^{20}\partial_t^2\tilde{u}$ in $\tilde{L}^2$ and using a Boussinesq trick, we obtain a constraint on $a^{22} - a^0b^{20}$ that can be satisfied by non-negative a^{22}, b^{20}. Regarding the odd operators a^{23}, a^{21}, we will see that they can be handled conveniently later. Following the previous reasoning, we make the following ansatz:

$$\tilde{L}^2 = \partial_x^2\big(a^{24}(x)\partial_x^2\,\cdot\,\big) - \partial_x\big(b^{22}(x)\partial_x\partial_t^2\,\cdot\,\big) + a^{23}(x)\partial_x^3 - a^{22}(x)\partial_x^2 + b^{20}(x)\partial_t^2 + a^{21}(x)\partial_x. \tag{6.35}$$

Note that the sign of each term are chosen following the conventions in the theory of PDEs. Using the effective equation (6.19) to substitute the second time derivative, we obtain

$$\begin{aligned}\tilde{L}^2\tilde{u} = {}& \partial_x^2\big(a^{24}\partial_x^2\tilde{u}\big) - \partial_x\big(b^{22}\partial_x^2(a^0\partial_x\tilde{u})\big) + a^{23}\partial_x^3\tilde{u} + (a^0b^{20} - a^{22})\partial_x^2\tilde{u} + (a^{21} + \partial_x a^0b^{20})\partial_x\tilde{u} \\ & - \partial_x\big(b^{22}\partial_x f\big) + b^{20}f + \mathcal{O}(\varepsilon).\end{aligned} \tag{6.36}$$

Inserting (6.36) in (6.27), we obtain the following cell problems for the cancellation of the terms of order ε^2 in (6.27): for all $x \in \Omega$, $\rho_i(x) = \rho_i(x,\cdot)$ $0 \le i \le 3$ are the solutions in $\mathrm{W}_{\mathrm{per}}(Y)$ of (the cell functions, a, and the coefficients are evaluated in x)

$$\begin{aligned}&\varepsilon^2: \\ \big(a\partial_y\rho_0, \partial_y w\big)_Y = {}& -\big(a\kappa_0, \partial_y w\big)_Y + \big(a(\partial_y\kappa_0 + \theta_0), w\big)_Y + \big(-a^0\theta_0 + a^{13}\chi, w\big)_Y, \\ & + \big(a^{24} - a^0b^{22}, w\big)_Y,\end{aligned} \tag{6.37a}$$

$$\begin{aligned}\big(a\partial_y\rho_1,\partial_y w\big)_Y =& -\big(a(\partial_x\kappa_0+\kappa_1),\partial_y w\big)_Y+\big(a(\partial_y\kappa_1+\partial_x\theta_0+\theta_1),w\big)_Y\\ &+\big(\partial_x(a(\partial_y\kappa_0+\theta_0)),w\big)_Y+\big(-3\partial_x a^0\theta_0-a^0\theta_1+(\partial_x a^{13}+a^{12})\chi,w\big)_Y\\ &+\big(2\partial_x a^{24}-\partial_x(a^0b^{22})-2\partial_x a^0b^{22}+a^{23},w\big)_Y,\end{aligned}\tag{6.37b}$$

$$\begin{aligned}\big(a\partial_y\rho_2,\partial_y w\big)_Y =& -\big(a(\partial_x\kappa_1+\kappa_2),\partial_y w\big)_Y\\ &+\big(a(\partial_y\kappa_2+\partial_x\theta_1),w\big)_Y+\big(\partial_x(a(\partial_y\kappa_1+\partial_x\theta_0+\theta_1)),w\big)_Y\\ &+\big(-3\partial_x^2 a^0\theta_0-2\partial_x a^0\theta_1+(\partial_x a^{12}+a^{11})\chi,w\big)_Y\\ &+\big(\partial_x^2 a^{24}-2\partial_x(\partial_x a^0b^{22})-\partial_x^2 a^0b^{22}+a^0b^{20}-a^{22},w\big)_Y,\end{aligned}\tag{6.37c}$$

$$\begin{aligned}\big(a\partial_y\rho_3,\partial_y w\big)_Y =& -\big(a\partial_x\kappa_2,\partial_y w\big)_Y+\big(\partial_x(a(\partial_y\kappa_2+\partial_x\theta_1)),w\big)_Y\\ &+\big(-\partial_x^3 a^0\theta_0-\partial_x^2 a^0\theta_1+\partial_x a^{11}\chi-\partial_x(\partial_x^2 a^0b^{22})+a^{21}+\partial_x a^0b^{20},w\big)_Y,\end{aligned}\tag{6.37d}$$

for all test functions $w\in \mathrm{W}_{\mathrm{per}}(Y)$. In order for the cell problems (6.37) to be well-posed in $\mathrm{W}_{\mathrm{per}}(Y)$, their right hand sides have to satisfy (6.25). Accordingly, recalling that $\langle\chi\rangle_Y=\langle\theta_0\rangle_Y=\langle\theta_1\rangle_Y=0$ (assumption (H1)), we impose the following constraints on $a^{24},a^{23},a^{22},a^{21}$ and b^{22},b^{20}, for all $x\in\Omega$:

$$a^{24}-a^0b^{22}=-\big\langle a(\partial_y\kappa_0+\theta_0)\big\rangle_Y,\tag{6.38a}$$

$$\begin{aligned}a^{23}=&-\big\langle a(\partial_y\kappa_1+\partial_x\theta_0+\theta_1)\big\rangle_Y-\partial_x\big\langle a(\partial_y\kappa_0+\theta_0)\big\rangle_Y\\ &+\partial_x(a^0b^{22}-a^{24})-\partial_x a^{24}+2\partial_x a^0b^{22},\end{aligned}\tag{6.38b}$$

$$\begin{aligned}a^0b^{20}-a^{22}=&-\big\langle a(\partial_y\kappa_2+\partial_x\theta_1)\big\rangle_Y-\partial_x\big\langle a(\partial_y\kappa_1+\partial_x\theta_0+\theta_1)\big\rangle_Y\\ &+2\partial_x(\partial_x a^0b^{22})+\partial_x^2 a^0b^{22}-\partial_x^2 a^{24},\end{aligned}\tag{6.38c}$$

$$a^{21}=-\partial_x\big\langle a(\partial_y\kappa_2+\partial_x\theta_1)\big\rangle_Y+\partial_x(\partial_x^2 a^0b^{22})-\partial_x a^0b^{20}.\tag{6.38d}$$

These constraints are simplified in the following lemma.

Lemma 6.1.7. *Under assumption* (H1)*, if we denote* $R(x)=b^{22}(x)-\big\langle\chi(x)^2\big\rangle_Y$*, then the constraints* (6.38) *can be rewritten for all* $x\in\Omega$ *as*

$$a^{24}=a^0R,\tag{6.39a}$$

$$a^{23}=\partial_x a^0R-a^0\partial_x R,\tag{6.39b}$$

$$a^0b^{20}-a^{22}=2\partial_x^2 a^0R-a^0\partial_x^2 R,\tag{6.39c}$$

$$a^{21}=\partial_x(\partial_x^2 a^0R)-b^{20}\partial_x a^0.\tag{6.39d}$$

Proof. Let us first prove (6.39a). Using (6.22) with the test function $w=\kappa_0$ and (6.29a) with $w=\chi$, we have

$$\begin{aligned}-\big(a(\partial_y\kappa_0+\theta_0),1\big)_Y&=\big(a\partial_y\kappa_0,\partial_y\chi\big)_Y-\big(a\theta_0,1\big)_Y\\ &=-\big(a(\partial_y\chi+1),\theta_0\big)_Y+\big(a(\partial_y\theta_0+\chi)-a^0\chi+a^{13},\chi\big)_Y.\end{aligned}$$

Using (6.31) and (6.32), (6.38a) simplifies to (6.39a). Let us now prove (6.39b). Using (6.22) with the test function $w=\kappa_1$ and (6.29b) with $w=\chi$, we obtain

$$\begin{aligned}-\big(a(\partial_y\kappa_1+\partial_x\theta_0+\theta_1),1\big)_Y=&\big(a\partial_y\kappa_1,\partial_y\chi\big)_Y-\big(a(\partial_x\theta_0+\theta_1),1\big)_Y\\ =&-\big(a(\partial_y\chi+1),\partial_x\theta_0+\theta_1\big)_Y+\big(a(\partial_y\theta_1+\partial_x\chi),\chi\big)_Y\\ &+\big(\partial_x(a(\partial_y\theta_0+\chi)),\chi\big)_Y+\big(-2\partial_x a^0\chi+a^{12},\chi\big)_Y.\end{aligned}$$

Using (6.31), (6.32), and (6.33), we obtain

$$-\big(a(\partial_y\kappa_1+\partial_x\theta_0+\theta_1),1\big)_Y=-2\partial_x a^0\big(\chi,\chi\big)_Y. \tag{6.40}$$

Thanks to (6.38a), we have $-\partial_x\big\langle a(\partial_y\kappa_0+\theta_0)\big\rangle_Y=\partial_x(a^{24}-a^0b^{22})$. Using then (6.40) and (6.39a) in (6.38b), we obtain (6.39b). We now prove (6.39c). Using (6.22) with the test function $w=\kappa_1$ and (6.29c) with $w=\chi$, we get

$$\begin{aligned}-\big(a(\partial_y\kappa_2+\partial_x\theta_1),1\big)_Y&=\big(a\partial_y\kappa_2,\partial_y\chi\big)_Y-\big(a\partial_x\theta_1,1\big)_Y\\&=-\big(a(\partial_y\chi+1),\partial_x\theta_1\big)_Y+\big(\partial_x(a(\partial_y\theta_1+\partial_x\chi)),\chi\big)_Y+\big(-\partial_x^2a^0\chi+a^{11},\chi\big)_Y.\end{aligned}$$

Using (6.31), (6.32), and (6.33) brings

$$-\big(a(\partial_y\kappa_2+\partial_x\theta_1),1\big)_Y=-\partial_x^2a^0\big(\chi,\chi\big)_Y. \tag{6.41}$$

This equality combined with (6.40), (6.31), (6.32), (6.33), and (6.39a) leads after simplification to the equality (6.39c). Finally, (6.39d) is proved by combining (6.41), (6.31), (6.32), (6.33), and (6.39a). □

Including a non-zero right hand side

To complete the definition of the adaptation (6.17), we have to define the corrector φ to remove the terms coming from the right hand side f in the expansion (6.27) combined with (6.36). We thus define $\boldsymbol{\varphi}=[\varphi]\in\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{W}_{\mathrm{per}}(\Omega))$, with $\partial_t\boldsymbol{\varphi}\in\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{L}^2(\Omega))$ and $\partial_t^2\boldsymbol{\varphi}\in\mathrm{L}^2(0,T^\varepsilon;\mathcal{W}^*_{\mathrm{per}}(\Omega))$, as the unique solution of the equation

$$\begin{aligned}&(\partial_t^2+\mathcal{A}^\varepsilon)\boldsymbol{\varphi}=-\big[\varepsilon\chi(\cdot,\tfrac{\cdot}{\varepsilon})\partial_x f+\varepsilon^2\big(\partial_x\big(b^{22}\partial_x f\big)+\theta_0(\cdot,\tfrac{\cdot}{\varepsilon})\partial_x^2 f+\theta_1(\cdot,\tfrac{\cdot}{\varepsilon})\partial_x f-b^{20}f\big)\big]\\&\qquad\qquad\text{in }\mathcal{W}^*_{\mathrm{per}}(\Omega)\ \text{ a.e. } t\in[0,T^\varepsilon],\\&\boldsymbol{\varphi}(0)=\partial_t\boldsymbol{\varphi}(0)=[0].\end{aligned} \tag{6.42}$$

The standard energy estimate for the wave equation ensures the following bound:

$$\|\boldsymbol{\varphi}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{W})}\le\|\partial_x\boldsymbol{\varphi}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))}\le C\varepsilon\|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^2(\Omega))}, \tag{6.43}$$

where C depends only on λ, Λ, $\|\chi\|_{\mathcal{C}^0(\bar\Omega;\mathcal{C}^0(\bar Y))}$, $\|b^{22}\|_{\mathcal{C}^1(\bar\Omega)}$, $\|b^{20}\|_{\mathcal{C}^0(\bar\Omega)}$, $\|\theta_0\|_{\mathcal{C}^0(\bar\Omega;\mathcal{C}^0(\bar Y))}$, and $\|\theta_1\|_{\mathcal{C}^0(\bar\Omega;\mathcal{C}^0(\bar Y))}$.

Now that all the correctors have been defined, let us discuss what happens in the case of a uniformly periodic tensor, i.e., $a(x,y)=a(y)\ \forall x\in\Omega$. In Remark 6.1.4, we have shown that in this case we recover the family defined in Section 4.3.1. In addition, we verify that the adaptations are the same (if we require the correctors to have zero mean, see Section 4.2.3). Indeed, we verify that we have $\theta_1=0$, $\kappa_1=\kappa_2=0$, $\rho_1=\rho_2=\rho_3=0$ and $\chi,\theta_0,\kappa_0,\rho_0,\varphi$ are the same (zero mean) correctors as in the uniformly periodic case (defined in (4.45)).

Proof of Theorem 6.1.5

The adaptation $\mathcal{B}^\varepsilon\tilde u$ in (6.17) is defined explicitly by $u^1,\dots,u^4$ and φ (see (6.21), (6.23), (6.28), (6.34), and (6.42)). Thanks to assumption (4.25), we verify that $x\mapsto\mathcal{B}^\varepsilon\tilde u(t,x)$ is Ω-periodic. Furthermore, by construction, $\mathcal{B}^\varepsilon\tilde u$ satisfies the properties (6.18). We only need to verify that the effective coefficients in (6.15) satisfy the constraints ensuring the well-posedness of the correctors. First, we have verified that the cell problems (6.22) for $\chi(x)$ are well-posed unconditionally. Second, thanks to the definition of the homogenized tensor, we have verified that the cell problems

(6.24) for $\theta_i(x)$ are well-posed. Next, as Lemma 6.1.6 ensures that $\tilde{L}^1 = 0$, the cell problems (6.29) for $\kappa_i(x)$ are well-posed unconditionally. Finally, we verify that the coefficients in (6.15) satisfy the equalities (6.39) for $R(x) = ra^0(x)$. Hence, Lemma 6.1.7 guarantees the well-posedness of the cell problems (6.37) for $\rho_i(x)$. As all the cell problems are well-posed, $\mathcal{B}^\varepsilon \tilde{u}$ is well-defined and the proof of the theorem is complete.

6.1.4 Proof of the error estimate (Theorem 6.1.1)

In this section, we prove Theorem 6.1.1. The proof is structured as follows. First, based on the correctors derived in Section 6.1.3, we define the adaptation operator $\boldsymbol{\mathcal{B}}^\varepsilon$. In particular, we recall that the definition of the effective coefficients ensures the existence and uniqueness of the correctors. The error is then split as

$$\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(W)} = \| \left[u^\varepsilon - \tilde{u}\right] \|_{\mathrm{L}^\infty(\mathcal{W})} \leq \|\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u} - \left[u^\varepsilon\right] \|_{\mathrm{L}^\infty(\mathcal{W})} + \| \left[\tilde{u}\right] - \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}\|_{\mathrm{L}^\infty(\mathcal{W})},$$

and both terms are estimated separately. In particular, we prove that $\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}$ satisfies the same equation as u^ε up to a remainder of order $\mathcal{O}(\varepsilon^3)$ (Lemma 6.1.8).

We first introduce the correctors derived in Section 6.1.3. Let χ, $\{\theta_i\}_{i=0}^1$, $\{\kappa_i\}_{i=0}^2$ and $\{\rho_i\}_{i=0}^3$ be the correctors defined by the cell problems in (6.22), (6.24), (6.29) and (6.37), and let $\boldsymbol{\varphi}$ be the solution of (6.42). Thanks to the definition of a^0 in (6.12) and the definition of the coefficients in (6.15) ensures that all the cell problems are well-posed in $\mathrm{W}_{\mathrm{per}}(Y)$ (see Section 6.1.3 and in particular the proof of Theorem 6.1.5). Using Lemma 6.2.10, we obtain the following regularity implications, for $m, n \geq 0$:

$$\begin{aligned}
\chi, \theta_0, \kappa_0, \rho_0 \in \mathcal{C}^n(\bar{\Omega}; \mathrm{H}^{m+1}(Y)) \quad &\Leftarrow \quad a \in \mathcal{C}^n(\bar{\Omega}; \mathrm{W}^{m,\infty}(Y)), \\
\theta_1, \kappa_1, \rho_1 \in \mathcal{C}^n(\bar{\Omega}; \mathrm{H}^{m+1}(Y)) \quad &\Leftarrow \quad a \in \mathcal{C}^n(\bar{\Omega}; \mathcal{C}^m(Y)) \cap \mathcal{C}^{n+1}(\bar{\Omega}; \mathrm{W}^{\{m-1\}_+,\infty}(Y)), \\
\kappa_2, \rho_2 \in \mathcal{C}^n(\bar{\Omega}; \mathrm{H}^{m+1}(Y)) \quad &\Leftarrow \quad a \in \cap_{k=0}^2 \mathcal{C}^{n+k}(\bar{\Omega}; \mathrm{W}^{\{m-k\}_+,\infty}(Y)), \\
\rho_3 \in \mathcal{C}^n(\bar{\Omega}; \mathrm{H}^{m+1}(Y)) \quad &\Leftarrow \quad a \in \cap_{k=0}^3 \mathcal{C}^{n+k}(\bar{\Omega}; \mathrm{W}^{\{m-k\}_+,\infty}(Y)), \\
a^0 \in \mathcal{C}^n(\bar{\Omega}) \quad &\Leftarrow \quad a \in \mathcal{C}^n(\bar{\Omega}; \mathrm{L}^\infty(Y)),
\end{aligned} \tag{6.44}$$

where $\{\cdot\}_+ = \max\{0, \cdot\}$. In particular, under the assumption of Theorem 6.1.1, i.e., $a \in \mathcal{C}^1(\bar{\Omega}; \mathrm{W}^{1,\infty}(Y)) \cap \mathcal{C}^4(\bar{\Omega}; \mathrm{L}^\infty(Y))$, all the correctors belongs to $\mathcal{C}^1(\bar{\Omega}; \mathrm{H}^2(Y))$. Furthermore, $\kappa_0, \kappa_1, \kappa_2 \in \mathcal{C}^2(\bar{\Omega}; \mathrm{H}^1(Y))$. As $d = 1$, the embedding $\mathrm{H}^1(Y) \hookrightarrow \mathcal{C}^0(\bar{Y})$ holds and we have the following estimates (needed in the proof of Lemma 6.1.8 below)

$$\begin{aligned}
&\|\chi\|_{\mathcal{C}^0(\bar{\Omega}; \mathcal{C}^0(\bar{Y}))}, \|\theta_0\|_{\mathcal{C}^1(\bar{\Omega}; \mathcal{C}^1(\bar{Y}))}, \|\theta_1\|_{\mathcal{C}^0(\bar{\Omega}; \mathcal{C}^0(\bar{Y}))} \leq C(a, \lambda, Y), \\
&\|\kappa_i\|_{\mathcal{C}^2(\bar{\Omega}; \mathcal{C}^0(\bar{Y}))}, \|\rho_i\|_{\mathcal{C}^1(\bar{\Omega}; \mathcal{C}^1(\bar{Y}))}, \|a^0\|_{\mathcal{C}^4(\bar{\Omega})} \leq C(a, \lambda, Y),
\end{aligned} \tag{6.45}$$

where $C(a, \lambda, Y)$ is a constant depending only on $\lambda, Y, \|a\|_{\mathcal{C}^1(\bar{\Omega}; \mathrm{W}^{1,\infty}(Y))}$, and $\|a\|_{\mathcal{C}^4(\bar{\Omega}; \mathrm{L}^\infty(Y))}$.

Let us introduce the following useful application of the Green formula (see Remark 4.2.7): For $c \in \mathrm{W}^{1,\infty}_{\mathrm{per}}(\Omega; \mathrm{W}^{1,\infty}_{\mathrm{per}}(Y))$, $v \in \mathrm{H}^1_{\mathrm{per}}(\Omega)$ and $\boldsymbol{w} \in \mathcal{W}_{\mathrm{per}}(\Omega)$, we have

$$\left(\left[c(\cdot, \cdot/\varepsilon)\partial_x v\right], \boldsymbol{w}\right)_{\mathcal{L}^2} = -\left(\left[(\partial_x c(\cdot, \cdot/\varepsilon) + \varepsilon^{-1}\partial_y c(\cdot, \cdot/\varepsilon))v\right], \boldsymbol{w}\right)_{\mathcal{L}^2} - \left(c(\cdot, \cdot/\varepsilon)v, \partial_x \boldsymbol{w}\right)_{\mathrm{L}^2}. \tag{6.46}$$

For $i = 0, \dots, 3$, we define the operators $\boldsymbol{\mathcal{B}}_i^\varepsilon : \mathrm{H}^4_{\mathrm{per}}(\Omega) \to \mathcal{W}_{\mathrm{per}}(\Omega)$ for $v \in \mathrm{H}^4_{\mathrm{per}}(\Omega)$ as

$$
\begin{aligned}
\big\langle \boldsymbol{\mathcal{B}}_0^\varepsilon v, \boldsymbol{w} \big\rangle &= \big(\big[v \big], \boldsymbol{w} \big)_{\mathcal{L}^2}, \\
\big\langle \boldsymbol{\mathcal{B}}_1^\varepsilon v, \boldsymbol{w} \big\rangle &= \big(\varepsilon \big[\chi \partial_x v \big], \boldsymbol{w} \big)_{\mathcal{L}^2}, \\
\big\langle \boldsymbol{\mathcal{B}}_2^\varepsilon v, \boldsymbol{w} \big\rangle &= \big(\varepsilon^2 \big[(-\partial_x \theta_0 - \varepsilon^{-1} \partial_y \theta_0 + \theta_1) \partial_x v \big], \boldsymbol{w} \big)_{\mathcal{L}^2} - \big(\varepsilon^2 \theta_0 \partial_x v, \partial_x \boldsymbol{w} \big)_{\mathrm{L}^2}, \\
\big\langle \boldsymbol{\mathcal{B}}_3^\varepsilon v, \boldsymbol{w} \big\rangle &= \big(\varepsilon^3 \big[\kappa_0 \partial_x^3 v + \kappa_1 \partial_x^2 v + \kappa_2 \partial_x v \big], \boldsymbol{w} \big)_{\mathcal{L}^2}, \\
\big\langle \boldsymbol{\mathcal{B}}_4^\varepsilon v, \boldsymbol{w} \big\rangle &= \big(\varepsilon^4 \big[(-\partial_x \rho_0 - \varepsilon^{-1} \partial_y \rho_0 + \rho_1) \partial_x^3 v + \rho_2 \partial_x^2 v + \rho_3 \partial_x v \big], \boldsymbol{w} \big)_{\mathcal{L}^2} - \big(\varepsilon^4 \rho_0 \partial_x v, \partial_x \boldsymbol{w} \big)_{\mathrm{L}^2},
\end{aligned}
$$

where the correctors are evaluated at $(x, x/\varepsilon)$ and $\langle \cdot, \cdot \rangle$ denotes $\langle \cdot, \cdot \rangle_{\mathcal{W}_{\mathrm{per}}, \mathcal{W}_{\mathrm{per}}}$. The adaptation operator $\boldsymbol{\mathcal{B}}^\varepsilon : \mathrm{L}^2(0, T^\varepsilon; \mathrm{H}^4_{\mathrm{per}}(\Omega)) \to \mathrm{L}^2(0, T^\varepsilon; \mathcal{W}^*_{\mathrm{per}}(\Omega))$ is then defined for $v \in \mathrm{L}^2(0, T^\varepsilon; \mathrm{H}^4_{\mathrm{per}}(\Omega))$ as

$$
\boldsymbol{\mathcal{B}}^\varepsilon v(t) = \sum_{i=0}^{4} \boldsymbol{\mathcal{B}}_i^\varepsilon (v(t)) + \boldsymbol{\varphi}(t). \tag{6.47}
$$

Note that if $v \in \mathrm{L}^2(0, T^\varepsilon; \mathrm{H}^1_{\mathrm{per}}(\Omega) \cap \mathrm{H}^5(\Omega))$, then $\boldsymbol{\mathcal{B}}^\varepsilon v(t) \in \mathcal{W}_{\mathrm{per}}(\Omega)$ and, using (6.110), we verify that $\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}(t) = \big[\mathcal{B}^\varepsilon \tilde{u}(t) \big]$, For $\mathcal{A}^\varepsilon = -\partial_x \big(a^\varepsilon(x) \partial_x \cdot \big)$, we thus define

$$
\big\langle \mathcal{A}^\varepsilon \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}(t), \boldsymbol{w} \big\rangle_{\mathcal{W}^*_{\mathrm{per}}, \mathcal{W}_{\mathrm{per}}} = \big\langle \mathcal{A}^\varepsilon \big[\mathcal{B}^\varepsilon v(t) \big], \boldsymbol{w} \big\rangle_{\mathcal{W}^*_{\mathrm{per}}, \mathcal{W}_{\mathrm{per}}},
$$

where $\mathcal{B}^\varepsilon \tilde{u}$ is defined in (6.17). Remark that the definition of $\boldsymbol{\mathcal{B}}^\varepsilon$ in (6.47) allows to consider functions with lower regularity than $\mathcal{B}^\varepsilon$. In particular, as $\partial_t^2 \tilde{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^3_{\mathrm{per}}(\Omega))$, $\boldsymbol{\mathcal{B}}^\varepsilon(\partial_t^2 \tilde{u})$ is well-defined. This is needed to prove the following lemma, which ensures that $\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}$ solves the same equation as $\big[u^\varepsilon \big]$ with a remainder of order ε^3.

Lemma 6.1.8. *Under the assumptions of Theorem 6.1.1, $\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}$ satisfies*

$$
(\partial_t^2 + \mathcal{A}^\varepsilon) \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}(t) = \big[f(t) \big] + \boldsymbol{\mathcal{R}}^\varepsilon \tilde{u}(t) \quad \text{in } \mathcal{W}^*_{\mathrm{per}}(\Omega) \quad \text{for a.e. } t \in [0, T^\varepsilon],
$$

where the remainder $\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathcal{W}^*_{\mathrm{per}}(\Omega))$ is given as

$$
\big\langle \boldsymbol{\mathcal{R}}^\varepsilon \tilde{u}(t), \boldsymbol{w} \big\rangle_{\mathcal{W}^*_{\mathrm{per}}, \mathcal{W}_{\mathrm{per}}} = \big((\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u})_0(t), \boldsymbol{w} \big)_{\mathcal{L}^2} + \big((\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u})_1(t), \partial_x \boldsymbol{w} \big)_{\mathrm{L}^2},
$$

with the bound

$$
\begin{aligned}
&\| (\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u})_0 \|_{\mathrm{L}^\infty(0, T^\varepsilon; \mathcal{L}^2(\Omega))} + \| (\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u})_1 \|_{\mathrm{L}^\infty(0, T^\varepsilon; \mathrm{L}^2(\Omega))} \\
&\qquad \leq C \varepsilon^3 \Bigg(\sum_{k=1}^{5} |\tilde{u}|_{\mathrm{L}^\infty((0, T^\varepsilon; \mathrm{H}^k(\Omega)))} + \| \partial_t^2 \tilde{u} \|_{\mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^3(\Omega))} \Bigg),
\end{aligned} \tag{6.48}
$$

for a constant C that only depends on λ, Y, $\|a\|_{\mathcal{C}^1(\bar{\Omega}; \mathrm{W}^{1,\infty}(Y))}$, $\|a\|_{\mathcal{C}^4(\bar{\Omega}; \mathrm{L}^\infty(Y))}$, $|b^{20}|$, $\|a^{22}\|_{\mathcal{C}^2(\bar{\Omega})}$, $\|a^{24}\|_{\mathcal{C}^3(\bar{\Omega})}$ and $\|b^{22}\|_{\mathcal{C}^2(\bar{\Omega})}$.

Proof. Let us denote $\langle \cdot, \cdot \rangle_{\mathcal{W}_{\mathrm{per}}, \mathcal{W}_{\mathrm{per}}}$ as $\langle \cdot, \cdot \rangle$. For a fixed $t \in [0, T^\varepsilon]$ and $\boldsymbol{w} \in \mathcal{W}_{\mathrm{per}}(\Omega)$, we compute the remainder $\big\langle \boldsymbol{\mathcal{R}}^\varepsilon \tilde{u}(t), \boldsymbol{w} \big\rangle = \big(\big[f \big], \boldsymbol{w} \big)_{\mathcal{L}^2} - \big\langle (\partial_t^2 + \mathcal{A}^\varepsilon) \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}(t), \boldsymbol{w} \big\rangle$. Let us develop the two terms separately. For the sake of clarity, we drop the function evaluations in t and the evaluation of the correctors at $(x, x/\varepsilon)$. From the definition of the adaptation $\boldsymbol{\mathcal{B}}^\varepsilon$ in (6.47), we have $\partial_t^2 \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u} = \sum_{i=0}^{2} \boldsymbol{\mathcal{B}}_i^\varepsilon \partial_t^2 \tilde{u} + \partial_t^2 \boldsymbol{\varphi} + \boldsymbol{\mathcal{R}}_1^\varepsilon \tilde{u}$, where $\boldsymbol{\mathcal{R}}_1^\varepsilon \tilde{u} = \sum_{i=3}^{4} \boldsymbol{\mathcal{B}}_i^\varepsilon \partial_t^2 \tilde{u}$. Thanks to the regularity of $\tilde{u}$ and (6.14), we have the following equalities

$$
\begin{aligned}
\big(\big[\partial_t^2 \tilde{u} \big], \boldsymbol{w} \big)_{\mathcal{L}^2} = \big(\big[f \big] + \big[\partial_x (a^0 \partial_x \tilde{u}) - \varepsilon^2 \big(\partial_x (a^{22} \partial_x \tilde{u}) - b^{20} \partial_t^2 \tilde{u} - \partial_x^2 (a^{24} \partial_x^2 \tilde{u}) \big) \big], \boldsymbol{w} \big)_{\mathcal{L}^2} \\
- \big(\varepsilon^2 b^{22} \partial_x \partial_t^2 \tilde{u}, \partial_x \boldsymbol{w} \big)_{\mathrm{L}^2},
\end{aligned} \tag{6.49}
$$

$$
\partial_x \partial_t^2 \tilde{u} = \partial_x f + \partial_x^2 (a^0 \partial_x \tilde{u}) + \varepsilon^2 \big(\partial_x^3 (\partial_x^2 (b^{22} \partial_x \partial_t^2 \tilde{u}) - a^{24} \partial_x^2 \tilde{u}) + \partial_x^2 (a^{22} \partial_x \tilde{u}) - b^{20} \partial_x \partial_t^2 \tilde{u} \big), \tag{6.50}
$$

where (6.50) holds in $\mathrm{L}^2(\Omega)$. We use (6.49) to rewrite

$$\begin{aligned}\big\langle \partial_t^2 \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}, \boldsymbol{w}\big\rangle = &\big([\![f]\!] + \big[\partial_x(a^0\partial_x\tilde{u}) + \varepsilon\chi\partial_x\partial_t^2\tilde{u} + \varepsilon^2\big(\partial_x(a^{22}\partial_x\tilde{u}) - b^{20}\partial_t^2\tilde{u} - \partial_x^2(a^{24}\partial_x^2\tilde{u})\big)\big], \boldsymbol{w}\big)_{\mathcal{L}^2}\\ &+ \big(\varepsilon^2\big[(-\partial_x\theta_0 - \varepsilon\partial_y\theta_0 + \theta_1)\partial_x\partial_t^2\tilde{u}\big], \boldsymbol{w}\big)_{\mathcal{L}^2} - \big(\varepsilon^2(\theta_0 + b^{22})\partial_x\partial_t^2\tilde{u}, \partial_x\boldsymbol{w}\big)_{\mathrm{L}^2}\\ &+ \big\langle \partial_t^2\boldsymbol{\varphi}, \boldsymbol{w}\big\rangle + \big\langle \boldsymbol{\mathcal{R}}_1^\varepsilon\tilde{u}, \boldsymbol{w}\big\rangle,\end{aligned}$$

Then, we rewrite $\partial_x\partial_t^2\tilde{u}$ and $b^{20}\big([\![\partial_t^2\tilde{u}]\!], \boldsymbol{w}\big)_{\mathcal{L}^2}$ using (6.50) and (6.49) and obtain

$$\begin{aligned}&\big\langle \partial_t^2 \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}, \boldsymbol{w}\big\rangle\\ &\quad= \big([\![f]\!] + \big[\partial_x(a^0\partial_x\tilde{u}) + \varepsilon\chi\partial_x^2(a^0\partial_x\tilde{u}) + \varepsilon^2\big(\partial_x(a^{22}\partial_x\tilde{u}) - b^{20}\partial_x(a^0\partial_x\tilde{u}) - \partial_x^2(a^{24}\partial_x^2\tilde{u})\big)\big], \boldsymbol{w}\big)_{\mathcal{L}^2}\\ &\qquad+ \big(\varepsilon^2\big[(-\partial_x\theta_0 - \varepsilon^{-1}\partial_y\theta_0 + \theta_1)\partial_x^2(a^0\partial_x\tilde{u})\big], \boldsymbol{w}\big)_{\mathcal{L}^2} - \big(\varepsilon^2(\theta_0 + b^{22})\partial_x^2(a^0\partial_x\tilde{u}), \partial_x\boldsymbol{w}\big)_{\mathrm{L}^2}\\ &\qquad+ \big\langle \partial_t^2\boldsymbol{\varphi}, \boldsymbol{w}\big\rangle + \big(\big[(\varepsilon\chi + \varepsilon^2(-\partial_x\theta_0 - \varepsilon^{-1}\partial_y\theta_0 + \theta_1)\partial_x f - b^{20}f\big], \boldsymbol{w}\big)_{\mathcal{L}^2}\\ &\qquad\qquad- \big(\varepsilon^2(\theta_0 + b^{22})\partial_x f, \partial_x\boldsymbol{w}\big)_{\mathrm{L}^2} \quad + \big\langle \boldsymbol{\mathcal{R}}_1^\varepsilon\tilde{u} + \boldsymbol{\mathcal{R}}_2^\varepsilon\tilde{u}, \boldsymbol{w}\big\rangle,\end{aligned}$$

where $\boldsymbol{\mathcal{R}}_2^\varepsilon\tilde{u}$ is given by

$$\begin{aligned}&\big\langle \boldsymbol{\mathcal{R}}_2^\varepsilon\tilde{u}, \boldsymbol{w}\big\rangle\\ &\quad= \big(\big[(\varepsilon^3\chi - \varepsilon^4\partial_x\theta_0 - \varepsilon^3\partial_y\theta_0 + \varepsilon^4\theta_1)(-b^{20}\partial_x\partial_t^2\tilde{u} + \partial_x^2(a^{22}\partial_x\tilde{u}) - \partial_x^3(a^{24}\partial_x^2\tilde{u}) + \partial_x^2(b^{22}\partial_x\partial_t^2\tilde{u}))\\ &\qquad- \varepsilon^4 b^{20}(-b^{20}\partial_t^2\tilde{u} + \partial_x(a^{22}\partial_x\tilde{u}) - \partial_x^2(a^{24}\partial_x^2\tilde{u}) + \partial_x(b^{22}\partial_x\partial_t^2\tilde{u}))\big], \boldsymbol{w}\big)_{\mathcal{L}^2}\\ &\qquad- \big(\varepsilon^4(\theta_0 + b^{22})(-b^{20}\partial_x\partial_t^2\tilde{u} + \partial_x^2(a^{22}\partial_x\tilde{u}) - \partial_x^3(a^{24}\partial_x^2\tilde{u}) + \partial_x^2(b^{22}\partial_x\partial_t^2\tilde{u})), \partial_x\boldsymbol{w}\big)_{\mathrm{L}^2}.\end{aligned}$$

Applying formula (6.46), we obtain

$$\begin{aligned}\big\langle \partial_t^2 \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}, \boldsymbol{w}\big\rangle = &\big([\![f]\!] + \big[\partial_x(a^0\partial_x\tilde{u}) + \varepsilon\chi\partial_x^2(a^0\partial_x\tilde{u})\big], \boldsymbol{w}\big)_{\mathcal{L}^2}\\ &+ \big(\varepsilon^2\big[\theta_0\partial_x^3(a^0\partial_x\tilde{u}) + \theta_1\partial_x^2(a^0\partial_x\tilde{u}) + \partial_x(b^{22}\partial_x^2(a^0\partial_x\tilde{u}))\\ &\qquad- \partial_x^2(a^{24}\partial_x^2\tilde{u}) + \partial_x(a^{22}\partial_x\tilde{u}) - b^{20}\partial_x(a^0\partial_x\tilde{u})\big], \boldsymbol{w}\big)_{\mathcal{L}^2}\\ &+ \big(\big[\varepsilon\chi\partial_x f + \varepsilon^2\theta_0\partial_x^2 f + \varepsilon^2\theta_1\partial_x f + \partial_x(b^{22}\partial_x f) - b^{20}f\big], \boldsymbol{w}\big)_{\mathcal{L}^2} + \big\langle \partial_t^2\boldsymbol{\varphi}, \boldsymbol{w}\big\rangle\\ &+ \big\langle \boldsymbol{\mathcal{R}}_1^\varepsilon\tilde{u} + \boldsymbol{\mathcal{R}}_2^\varepsilon\tilde{u}, \boldsymbol{w}\big\rangle,\end{aligned} \tag{6.51}$$

Let us now compute the other term, $\mathcal{A}^\varepsilon\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}$. We have in $\mathcal{W}_{\mathrm{per}}^*(\Omega)$

$$\begin{aligned}\mathcal{A}^\varepsilon\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u} = \big[\quad &\varepsilon^{-1}\big(-\partial_y(a(\partial_y\chi + 1))\big)\partial_x\tilde{u}\\ &+\varepsilon^0\ \big(-\partial_y(a(\partial_y\theta_0 + \chi)) - a(\partial_y\chi + 1)\big)\partial_x^2\tilde{u}\\ &+\varepsilon^0\ \big(-\partial_y(a(\partial_y\theta_1 + \partial_x\chi)) - \partial_x(a(\partial_y\chi + 1))\big)\partial_x\tilde{u}\\ &+\varepsilon^1\ \big(-\partial_y(a(\partial_y\kappa_0 + \theta_0)) - a(\partial_y\theta_0 + \chi)\big)\partial_x^3\tilde{u}\\ &+\varepsilon^1\ \big(-\partial_y(a(\partial_y\kappa_1 + \partial_x\theta_0 + \theta_1)) - \partial_x(a(\partial_y\theta_0 + \chi)) - a(\partial_y\theta_1 + \partial_x\chi)\big)\partial_x^2\tilde{u}\\ &+\varepsilon^1\ \big(-\partial_y(a(\partial_y\kappa_2 + \partial_x\theta_1)) - \partial_x(a(\partial_y\theta_1 + \partial_x\chi))\big)\partial_x\tilde{u}\\ &+\varepsilon^2\ \big(-\partial_y(a(\partial_y\rho_0 + \kappa_0)) - a(\partial_y\kappa_0 + \theta_0)\big)\partial_x^4\tilde{u}\\ &+\varepsilon^2\ \big(-\partial_y(a(\partial_y\rho_1 + \partial_x\kappa_0 + \kappa_1)) - \partial_x(a(\partial_y\kappa_0 + \theta_0)) - a(\partial_y\kappa_1 + \partial_x\theta_0)\big)\partial_x^3\tilde{u}\\ &+\varepsilon^2\ \big(-\partial_y(a(\partial_y\rho_2 + \partial_x\kappa_1 + \kappa_2)) - \partial_x(a(\partial_y\kappa_1 + \partial_x\theta_0 + \theta_1)) - a(\partial_y\kappa_2 + \partial_x\theta_1)\big)\partial_x^2\tilde{u}\\ &+\varepsilon^2\ \big(-\partial_y(a(\partial_y\rho_3 + \partial_x\kappa_2)) - \partial_x(a(\partial_y\kappa_2 + \partial_x\theta_1))\big)\partial_x\tilde{u} \qquad \big]\\ &+\mathcal{A}^\varepsilon\boldsymbol{\varphi} + \boldsymbol{\mathcal{R}}_3^\varepsilon\tilde{u},\end{aligned} \tag{6.52}$$

where, defining the functions $R_i(x,y)$ $0 \le i \le 3$, as

$$\begin{aligned}&R_0 = a(\partial_y\rho_0 + \kappa_0), && R_1 = a(\partial_y\rho_1 + \partial_x\kappa_0 + \kappa_1),\\ &R_2 = a(\partial_y\rho_2 + \partial_x\kappa_1 + \kappa_2), && R_3 = a(\partial_y\rho_3 + \partial_x\kappa_2),\end{aligned}$$

$\boldsymbol{\mathcal{R}}_3^\varepsilon \tilde{u}$ is given by

$$\begin{aligned}\big\langle \boldsymbol{\mathcal{R}}_3^\varepsilon \tilde{u}, \boldsymbol{w}\big\rangle = &-\varepsilon^3\big(\big[\textstyle\sum_{i=0}^3 (R_i \partial_x^{5-i}\tilde{u} + \partial_x R_i \partial_x^{4-i}\tilde{u})\big], \boldsymbol{w}\big)_{\mathcal{L}^2} \\ &+ \varepsilon^4\big(a\textstyle\sum_{i=0}^3 \rho_i \partial_x^{5-i}\tilde{u} + a\sum_{i=0}^3 \partial_x \rho_i \partial_x^{4-i}\tilde{u}, \partial_x \boldsymbol{w}\big)_{\mathrm{L}^2}.\end{aligned}$$

Combining now (6.51) and (6.52) and using the definitions of the correctors (6.22), (6.24), (6.29), (6.37) and (6.42), we obtain $\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u} = \sum_{i=1}^3 \boldsymbol{\mathcal{R}}_i^\varepsilon \tilde{u}$. Thanks to (6.45), we verify that $\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u}$ satisfies estimate (6.48) and the proof of the lemma is complete. □

Proof of Theorem 6.1.1. Using that $u^\varepsilon - \tilde{u} \in \mathrm{W}_{\mathrm{per}}(\Omega)$ and the triangle inequality, we have

$$\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(W)} = \|\,[u^\varepsilon - \tilde{u}]\,\|_{\mathrm{L}^\infty(\mathcal{W})} \le \|\,[u^\varepsilon] - \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}\|_{\mathrm{L}^\infty(\mathcal{W})} + \|\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u} - [\tilde{u}]\,\|_{\mathrm{L}^\infty(\mathcal{W})}. \tag{6.53}$$

Let us estimate the two terms of the right hand side. First, note that $\boldsymbol{\eta} = [u^\varepsilon] - \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}$ satisfies $(\partial_t^2 + \mathcal{A}^\varepsilon)\boldsymbol{\eta}(t) = \boldsymbol{\mathcal{R}}^\varepsilon \tilde{u}(t)$ in $\mathcal{W}^*_{\mathrm{per}}(\Omega)$ for a.e $t \in [0, T^\varepsilon]$, where $\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u}$ is defined in Lemma 6.1.8. Hence, using Corollary 4.2.2, the first term satisfies

$$\|\,[u^\varepsilon] - \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}\|_{\mathrm{L}^\infty(\mathcal{W})} \le C\varepsilon\Big(\|g^1\|_{\mathrm{H}^4} + \|g^0\|_{\mathrm{H}^4} + \sum_{k=1}^5 |\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^k)} + \|\partial_t^2 \tilde{u}\|_{\mathrm{L}^\infty \mathrm{H}^3}\Big), \tag{6.54}$$

where C depends on T, λ, Y, $\|a\|_{\mathcal{C}^1(\bar{\Omega};\mathrm{W}^{1,\infty}(Y))}$, $\|a\|_{\mathcal{C}^4(\bar{\Omega};\mathrm{L}^\infty(Y))}$, $|b^{20}|$, $\|a^{22}\|_{\mathcal{C}^2(\bar{\Omega})}$, $\|a^{24}\|_{\mathcal{C}^3(\bar{\Omega})}$, and $\|b^{22}\|_{\mathcal{C}^2(\bar{\Omega})}$. Using the definition of the coefficients (6.15) and (6.45), we verify that

$$|b^{20}| + \|a^{22}\|_{\mathcal{C}^2(\bar{\Omega})} + \|a^{24}\|_{\mathcal{C}^3(\bar{\Omega})} + \|b^{22}\|_{\mathcal{C}^2(\bar{\Omega})} \le C_0(a, \lambda, Y) + C_1(a, \lambda, Y) r.$$

Next, using the definition of $\boldsymbol{\mathcal{B}}^\varepsilon$ (6.47) and the estimates (6.43) and (6.45), the second term of (6.53) satisfies

$$\|\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u} - [\tilde{u}]\,\|_{\mathrm{L}^\infty(\mathcal{W})} \le C\varepsilon\Big(\sum_{k=1}^5 |\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^k)} + \|f\|_{\mathrm{L}^1(\mathrm{H}^2)}\Big), \tag{6.55}$$

where C depends on $\lambda, Y, \|a\|_{\mathcal{C}^1(\bar{\Omega};\mathrm{W}^{1,\infty}(Y))}$, and $\|a\|_{\mathcal{C}^4(\bar{\Omega};\mathrm{L}^\infty(Y))}$. Combining (6.53), (6.54) and (6.55), we obtain (6.16) and the proof of the theorem is complete. □

6.2 Effective equations in several dimensions

In this section, we present the main result of the chapter. We derive effective equations for long time wave propagation in locally periodic media in the multidimensional case. In particular, in Section 6.2.1, we define a parametrized family of effective equations and present an error estimate establishing its validity. The derivation is done in a similar manner as in the one-dimensional case, in Section 6.1. We refer to Section 6.1.1 for an explanation of the process. The technical derivation of the cell problems and of the corresponding constraints for the characterization of the family is presented in Section 6.2.2 and the proof of the main result is provided in Section 6.2.4.

Let $a^\varepsilon(x) = a\big(x, \frac{x}{\varepsilon}\big)$ be a $d \times d$ locally periodic tensor, i.e., $a(x, y)$ is Y-periodic in y and Ω-periodic in x. The domain $\Omega \subset \mathbb{R}^d$ is an arbitrarily large hypercube, assumed to be the union of cells of volume $\varepsilon|Y|$ (see assumption (4.25), Figure 4.2). In particular, this assumption ensures that $a^\varepsilon(x)$ is Ω-periodic ($y \mapsto a(x, y)$ is extended by periodicity). For $T^\varepsilon = \varepsilon^{-2}T$, we consider the wave equation: find $u^\varepsilon : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ such that

$$\begin{array}{ll} \partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \big(a\big(x, \frac{x}{\varepsilon}\big)\nabla_x u^\varepsilon(t,x)\big) = f(t,x) & \text{in } (0, T^\varepsilon] \times \Omega, \\ x \mapsto u^\varepsilon(t,x) \ \ \Omega\text{-periodic} & \text{in } [0, T^\varepsilon], \\ u^\varepsilon(0,x) = g^0(x), \quad \partial_t u^\varepsilon(0,x) = g^1(x) & \text{in } \Omega, \end{array} \tag{6.56}$$

where g^0, g^1 are given initial conditions and f is a source. The tensor $a(x,y)$ is assumed to be uniformly elliptic and bounded, i.e. there exists $\lambda, \Lambda > 0$ such that

$$\lambda|\xi|^2 \leq a(x,y)\xi\cdot\xi \leq \Lambda|\xi|^2 \quad \text{for a.e. } (x,y) \in \Omega \times Y. \tag{6.57}$$

The well-posedness of (6.56) is proved in Section 2.1.1. If $g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega)$, $g^1 \in \mathrm{L}^2_0(\Omega)$, $f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{L}^2_0(\Omega))$, then there exists a unique weak solution $u^\varepsilon \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega))$ with $\partial_t u^\varepsilon \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2_0(\Omega))$ and $\partial_t^2 u^\varepsilon \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{W}^*_{\mathrm{per}}(\Omega))$.

6.2.1 Error estimate and family of effective equations

We present here the main result of this chapter, which contributes to this thesis. We define a family of effective equations for the wave equation over long time in locally periodic media. The family is validated by an error estimate ensuring that its elements are ε-close to u^ε in the $\mathrm{L}^\infty(0,T^\varepsilon;W)$ norm. The complete derivation of the family is presented in Section 6.2.2 and the proof of the main result is provided in Section 6.2.4. We refer to Section 6.1.1 for a summary on the process used for the derivation. In particular, we recall assumption (H1).

Let us first define the parametrized tensors of the family of effective equations, as obtained in Section 6.2.2. For $x \in \Omega$, $\{\chi_i(x)\}_{i=1}^d$, $\{\theta^0_{ij}(x)\}_{ij=1}^d$, $\{\theta^1_i(x)\}_{i=1}^d \subset \mathrm{W}_{\mathrm{per}}(Y)$ are the zero mean solutions of the cell problems

$$\big(a(x)\nabla_y\chi_i(x),\nabla_y w\big)_Y = -\big(a(x)e_i,\nabla_y w\big)_Y, \tag{6.58a}$$

$$\big(a(x)\nabla_y\theta^0_{ij}(x),\nabla_y w\big)_Y = -\big(a(x)e_i\chi_j(x),\nabla_y w\big)_Y + \big(a(x)(\nabla_y\chi_j(x)+e_j) - a^0(x)e_j, e_i w\big)_Y, \tag{6.58b}$$

$$\big(a(x)\nabla_y\theta^1_i(x),\nabla_y w\big)_Y = -\big(a(x)\nabla_x\chi_i(x),\nabla_y w\big)_Y + \big(\nabla_x\cdot a(x)(\nabla_y\chi_i(x)+e_i) - \nabla_x\cdot a^0(x)e_i, w\big)_Y, \tag{6.58c}$$

for all test functions $w \in \mathrm{W}_{\mathrm{per}}(Y)$, where $a^0(x)$ is the homogenized tensor defined by

$$a^0_{ij}(x) = \big\langle e_i^T a(x)(\nabla_y\chi_j(x)+e_j)\big\rangle_Y. \tag{6.59}$$

We define the differential operator

$$L^1 = -\partial_i\big(\bar a^{12}_{ij}(x)\partial_j\,\cdot\,\big) + b^{10}\partial_t^2, \tag{6.60}$$

based on the following tensors

$$\begin{aligned}
&p^{13}_{ijk}(x) = \big\langle a(x)(\nabla_y\chi_k(x)+e_k)\cdot e_j\chi_i(x)\big\rangle_Y,\\
&q^{12}_{ij}(x) = \big\langle a(x)(\nabla_y\chi_j(x)+e_j)\cdot\nabla_x\chi_i(x)\big\rangle_Y,\\
&\check a^{12}_{ij}(x) = S^2_{ij}\Big\{-\partial_r p^{13}_{rij}(x) + \partial_r p^{13}_{jir}(x) - \partial_r p^{13}_{irj}(x) + 2q^{12}_{ij}(x)\Big\},\\
&b^{10} = \max_{x\in\Omega}\left\{-\frac{\lambda_{\min}(\check a^{12}(x))}{\lambda_{\min}(a^0(x))}\right\}_+,\\
&\bar a^{12}_{ij}(x) = \check a^{12}_{ij}(x) + b^{10}a^0_{ij}(x).
\end{aligned} \tag{6.61}$$

Furthermore, we define the differential operator

$$L^2 = \partial^2_{ij}\big(\bar a^{24}_{ijkl}(x)\partial^2_{kl}\,\cdot\,\big) - \partial_i\big(b^{22}_{ij}(x)\partial_j\partial_t^2\,\cdot\,\big) - \partial_i\big(\bar a^{22}_{ij}(x)\partial_j\,\cdot\,\big) + b^{20}\partial_t^2, \tag{6.62}$$

based upon the following tensors

$$\begin{aligned}
&\check{a}^{24}_{ijkl}(x) = S^{2,2}_{ij,kl}\Big\{\big\langle a(x)\chi_i(x)e_j\cdot\chi_l(x)e_k\big\rangle_Y - \big\langle a(x)\nabla_y\theta^0_{ij}(x)\cdot\nabla_y\theta^0_{kl}(x)\big\rangle_Y\Big\},\\
&A^{24}(x) = M\Big(\check{a}^{24}(x)\Big), \quad A^0(x) = M\Big(S^{2,2}_{ij,kl}\{a^0_{jk}(x)a^0_{il}(x)\}\Big),\\
&\delta \geq \delta^* = \max_{x\in\Omega}\left\{-\frac{\lambda_{\min}(A^{24}(x))}{\lambda_{\min}(A^0(x))}\right\}_+,\\
&\bar{a}^{24}_{ijkl}(x) = \check{a}^{24}_{ijkl}(x) + \delta S^{2,2}_{ij,kl}\big\{a^0_{jk}(x)a^0_{il}(x)\big\},\\
&b^{22}_{ij}(x) = \big\langle\chi_i(x)\chi_j(x)\big\rangle_Y + \delta a^0_{ij}(x),
\end{aligned} \tag{6.63}$$

where $S^{2,2}_{ij,kl}\{\cdot\} = S^2_{ij}\{S^2_{kl}\{\cdot\}\}$ and $M(\cdot)$ is the matrix construction defined in Section 4.3.3, and

$$\begin{aligned}
&p^{23}_{ijk}(x) = \big\langle a(x)e_j\chi_i(x)\cdot\nabla_x\chi_k(x)\big\rangle_Y - \big\langle a(x)\nabla_y\theta^0_{ji}(x)\cdot\nabla_y\theta^1_k(x)\big\rangle_Y,\\
&p^{22}_{ij}(x) = \big\langle a(x)\nabla_x\chi_j(x)\cdot\nabla_x\chi_i(x)\big\rangle_Y - \big\langle a(x)\nabla_y\theta^1_i(x)\cdot\nabla_y\theta^1_j(x)\big\rangle_Y,\\
&\check{a}^{22}_{ij}(x) = S^2_{ij}\Big\{\partial_r p^{23}_{jir}(x) - \partial_r p^{23}_{rij}(x) - \partial_r p^{23}_{irj}(x) + p^{22}_{ij}(x)\Big\}\\
&\qquad\qquad + b^{10}\check{a}^{12}_{ij}(x) + \delta\partial_s a^0_{ri}(x)\partial_r a^0_{sj}(x) - \delta\partial_r\big(a^0_{rs}(x)\partial_s a^0_{ij}(x)\big),\\
&b^{20} = \max_{x\in\Omega}\left\{-\frac{\lambda_{\min}(\check{a}^{22}(x))}{\lambda_{\min}(a^0(x))}\right\}_+,\\
&\bar{a}^{22}_{ij}(x) = \check{a}^{22}_{ij}(x) + b^{20}a^0_{ij}(x).
\end{aligned} \tag{6.64}$$

Observe that the tensors of L^2 are parametrized by $\delta \geq \delta^*$. Let then $\tilde{u} : [0,T^\varepsilon]\times\Omega\to\mathbb{R}$ be the solution of

$$\begin{aligned}
&\partial_t^2\tilde{u} - \partial_i\big(a^0_{ij}(x)\partial_j\tilde{u}\big) + \varepsilon L^1\tilde{u} + \varepsilon^2 L^2\tilde{u} = f && \text{in } (0,T^\varepsilon]\times\Omega,\\
&x\mapsto\tilde{u}(t,x) \ \ \Omega\text{-periodic} && \text{in } [0,T^\varepsilon],\\
&\tilde{u}(0,x) = g^0(x), \quad \partial_t\tilde{u}(0,x) = g^1(x) && \text{in } \Omega,
\end{aligned} \tag{6.65}$$

where the initial conditions g^0, g^1 and the source f are the same as in the equation for u^ε (6.56). As the homogenized tensor is symmetric, elliptic, and bounded (see Lemma 3.3.1), if the tensors and the data satisfy the regularity

$$\begin{gathered}
a^{24}_{ijkl}\in \mathrm{W}^{1,\infty}(\Omega), \quad b^{22}_{ij}, a^{12}_{ij}, a^{22}_{ij}\in \mathrm{L}^\infty(\Omega),\\
g^0\in \mathrm{W}_{\mathrm{per}}(\Omega)\cap\mathrm{H}^2(\Omega), \quad g^1\in \mathrm{L}^2_0(\Omega)\cap\mathrm{H}^1(\Omega), \quad f\in \mathrm{L}^2(0,T^\varepsilon;\mathrm{L}^2_0(\Omega)),
\end{gathered}$$

then there exists a unique weak solution of (6.65) (see Section 2.1.2). The main result of this chapter is the following theorem.

Theorem 6.2.1. *Assume* (H1) *and that the tensor $a(x,y)$ satisfies*

$$a\in\mathcal{C}^1(\bar{\Omega};\mathrm{W}^{2,\infty}(Y))\cap\mathcal{C}^2(\bar{\Omega};\mathrm{W}^{1,\infty}(Y))\cap\mathcal{C}^4(\bar{\Omega};\mathrm{L}^\infty(Y)).$$

Furthermore, assume that the solution $\tilde{u}$ of (6.65)*, the initial conditions and the right hand side satisfy the regularity*

$$\begin{gathered}
\tilde{u}\in\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^5(\Omega)), \quad \partial_t\tilde{u}\in\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^4(\Omega)), \quad \partial_t^2\tilde{u}\in\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^3(\Omega)),\\
g^0\in\mathrm{H}^4(\Omega), \quad g^1\in\mathrm{H}^4(\Omega), \quad f\in\mathrm{L}^2(0,T^\varepsilon;\mathrm{H}^2(\Omega)).
\end{gathered}$$

Then the following estimate holds

$$\begin{aligned}
\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} \leq C\varepsilon\Big(&\|g^1\|_{\mathrm{H}^4(\Omega)} + \|g^0\|_{\mathrm{H}^4(\Omega)} + \|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^2(\Omega))}\\
&+ \textstyle\sum_{k=1}^5 |\tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^k(\Omega))} + \|\partial_t^2\tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^3(\Omega))}\Big),
\end{aligned} \tag{6.66}$$

where C depends only on T, λ, Y, $\|a\|_{\mathcal{C}^1(\bar\Omega;\mathrm{W}^{2,\infty}Y))}$, $\|a\|_{\mathcal{C}^2(\bar\Omega;\mathrm{W}^{1,\infty}(Y))}$, $\|a\|_{\mathcal{C}^4(\bar\Omega;\mathrm{L}^\infty(Y))}$, and δ, and we recall the definition of the norm (see (A.4)*)*

$$\|w\|_W = \inf_{\substack{w=w_1+w_2\\ w_1,w_2\in \mathrm{W}_{\mathrm{per}}(\Omega)}} \Big\{ \|w_1\|_{\mathrm{L}^2(\Omega)} + \|\nabla w_2\|_{\mathrm{L}^2(\Omega)} \Big\} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(\Omega).$$

Theorem 6.2.1 leads to the following definition.

Definition 6.2.2. We define the family of effective equations $\mathcal{E}$ as the set of equations (6.65), where a^0 is the homogenized tensor defined in (6.59) and L^1, L^2 are defined in (6.60) and (6.62) for some parameter $\delta \geq \delta^*$.

Remark 6.2.3. The family $\mathcal{E}$, defined in Definition 6.2.2, generalizes the family for a uniformly periodic tensor defined by (4.94), in Section 4.3.2. Indeed, if the tensor does not depend on the slow variable, i.e., $a(x,y) = a(y)$, we verify successively that $\chi_i(x,y)$, $a^0(x)$, $\theta^0_{ij}(x,y)$ are constant in x and $\theta^1_i(x,y) = 0$. Hence, we have $\bar a^{12} = \bar a^{22} = 0$, $b^{10} = b^{20} = 0$, and $\bar a^{24}, b^{22}$ are constant. Equation (6.65) is thus left with the only correction $\varepsilon^2(\bar a^{24}_{ijkl}\partial^4_{ijkl} - b^{22}_{ij}\partial^2_{ij}\partial^2_t)$, which is of the same form as in the uniformly periodic case. Furthermore, we verify that the pairs $(\bar a^{24}, b^{22})$, defined by (6.63), match the pairs defined in the uniformly periodic case in (4.94) ($\bar a^{24} = a^2$, $b^{22} = b^2$, indeed we have $S^2_{ij}\{\theta^0_{ij}(x,y)\} = \theta_{ij}(y)$).

Remark 6.2.4. Assume that ε and $a(x,y)$ are such that $a^0 + \varepsilon\check a^{12} + \varepsilon^2\check a^{22}$ is uniformly elliptic, i.e., there exists $\tilde\lambda > 0$ such that

$$\big(a^0(x) + \varepsilon\check a^{12}(x) + \varepsilon^2\check a^{22}(x)\big)\xi\cdot\xi \geq \tilde\lambda|\xi|^2 \quad \text{ for a.e. } x\in\Omega \ \ \forall \xi\in\mathbb{R}^d.$$

Then the equation (6.65) with

$$L^1 = -\partial_i\big(\check a^{12}_{ij}\partial_j\,\cdot\,\big), \qquad L^2 = \partial^2_{ij}\big(\bar a^{24}_{ijkl}\partial^2_{kl}\,\cdot\,\big) - \partial_i\big(b^{22}_{ij}\partial_j\partial^2_t\,\cdot\,\big) - \partial_i\big(\check a^{22}_{ij}\partial_j\,\cdot\,\big),$$

is well-posed and its solution also satisfies the error estimate (6.66). Indeed, the role of the operators $b^{10}\partial^2_t$ and $b^{20}\partial^2_t$ is only to ensure the ellipticity of $a^0 + \varepsilon a^{12} + \varepsilon^2 a^{22}$. In the case where $\check a^{12}$ and $\check a^{22}$ are sufficiently small, $b^{10}\partial^2_t$ and $b^{20}\partial^2_t$ are superfluous. This discussion is carried on in Section 6.4, where we discuss the necessity of the operators $-\varepsilon\partial_i\big(\check a^{12}_{ij}\partial_j\,\cdot\,\big)$ and $-\varepsilon^2\partial_i\big(\check a^{22}_{ij}\partial_j\,\cdot\,\big)$ in the effective equations.

6.2.2 Derivation of the adaptation operator and of the effective equations

In this section, we present the complete derivation of the family of effective equations (Definition 6.2.2). In particular, we build the adaptation operator, used in the proof of Theorem 6.2.1. To that end, we derive the cell problems and the corresponding constraints on the effective tensors. The technical part is the simplification of the constraints. We refer to Section 6.1.1 for the description of the procedure used for the derivation and, in particular, we recall assumption (H1).

The main result of the section is the following theorem.

Theorem 6.2.5. *Let L^1 and L^2 be defined in* (6.60) *and* (6.62)*, respectively. Then there exists an adaptation of the form*

$$\mathcal{B}^\varepsilon\tilde u(t,x) = \tilde u(t,x) + \varepsilon u^1\big(t,x,\tfrac{x}{\varepsilon}\big) + \varepsilon^2 u^2\big(t,x,\tfrac{x}{\varepsilon}\big) + \varepsilon^3 u^3\big(t,x,\tfrac{x}{\varepsilon}\big) + \varepsilon^4 u^4\big(t,x,\tfrac{x}{\varepsilon}\big) + \varphi(t,x), \quad (6.67)$$

such that $x\mapsto \mathcal{B}^\varepsilon\tilde u(t,x)$ is Ω-periodic and

$$(u^\varepsilon - \mathcal{B}^\varepsilon\tilde u)(0) = \mathcal{O}(\varepsilon), \qquad \partial_t(u^\varepsilon - \mathcal{B}^\varepsilon\tilde u)(0) = \mathcal{O}(\varepsilon), \quad (6.68a)$$

$$(\partial^2_t + \mathcal{A}^\varepsilon)(u^\varepsilon - \mathcal{B}^\varepsilon\tilde u)(t) = \mathcal{O}(\varepsilon^3) \quad \textit{for a.e. } t\in[0,T^\varepsilon], \quad (6.68b)$$

where we denoted $\mathcal{A}^\varepsilon = -\nabla_x\cdot\big(a\big(x,\tfrac{x}{\varepsilon}\big)\nabla_x\,\cdot\,\big)$.

Thanks to the properties (6.68), we can use the adaptation (6.67) in the process described in Section 4.2.2 to prove that $\tilde{u}$ is close to u^ε in the $\mathrm{L}^\infty(0,T^\varepsilon;W)$ norm. This is done in Section 6.2.4, where we prove Theorem 6.2.1.

Let us now construct explicitly the adaptation $\mathcal{B}^\varepsilon\tilde{u}$ in (6.67). To do so, we use asymptotic expansions to derive cell problems. As we know, the well-posedness of these cell problems constrains quantitatively the operators in the effective equations. Note that, as in one dimension in Section 6.1.3, we do not have an a priori knowledge on the form of the higher order operators L^1 and L^2 in the effective equation. We thus design them as we cancel the levels of the expansion. They must guarantee that the effective equations are well-posed whilst the constraints provided by the cell problems are satisfied.

We make the ansatz that the effective solution $\tilde{u} : [0,T^\varepsilon]\times\Omega \to \mathbb{R}$ solves

$$\begin{array}{ll} \partial_t^2\tilde{u} - \partial_i\big(a^0_{ij}(x)\partial_j\tilde{u}\big) + \varepsilon\tilde{L}^1\tilde{u} + \varepsilon^2\tilde{L}^2\tilde{u} = f & \text{in } (0,T^\varepsilon]\times\Omega, \\ x\mapsto\tilde{u}(t,x) \;\; \Omega\text{-periodic} & \text{in } [0,T^\varepsilon], \\ \tilde{u}(0,x) = g^0(x), \quad \partial_t\tilde{u}(0,x) = g^1(x) & \text{in } \Omega, \end{array} \tag{6.69}$$

where $a^0(x)$ is the homogeneous tensor (defined in (6.59)) and $\tilde{L}^1,\tilde{L}^2$ are linear, ε-independent differential operators to be defined. We use here the notation $\tilde{L}^1,\tilde{L}^2$ to emphasize that at this point the operators are unknown. Furthermore, we make the ansatz that $\mathcal{B}^\varepsilon\tilde{u}$ has the form (6.67), where the $u^i(t,x,y)$ are unknown operators of $\tilde{u}$, Ω-periodic in x and Y-periodic in y. We introduce the differential operators

$$\mathcal{A}_{yy} = -\nabla_y\cdot\big(a(x,y)\nabla_y\cdot\big), \quad \mathcal{A}_{xy} = -\nabla_y\cdot\big(a(x,y)\nabla_x\cdot\big) - \nabla_x\cdot\big(a(x,y)\nabla_y\cdot\big), \quad \mathcal{A}_{xx} = -\nabla_x\cdot\big(a(x,y)\nabla_x\cdot\big).$$

For a sufficiently regular function $\psi(x,y)$, we verify that $\mathcal{A}^\varepsilon\psi\big(x,\frac{x}{\varepsilon}\big) = \big(\varepsilon^{-2}\mathcal{A}_{yy} + \varepsilon^{-1}\mathcal{A}_{xy} + \mathcal{A}_{xx}\big)\psi\big(x,\frac{x}{\varepsilon}\big)$. Hence, using (6.56), (6.69) and (6.67), we obtain the development

$$\begin{aligned} R^\varepsilon = (\partial_t^2 + \mathcal{A}^\varepsilon)(\mathcal{B}^\varepsilon\tilde{u} - u^\varepsilon)(t,x) &= \partial_t^2\mathcal{B}^\varepsilon\tilde{u}(t,x) + \mathcal{A}^\varepsilon\mathcal{B}^\varepsilon\tilde{u}(t,x) - f(t,x) \\ &= \varepsilon^{-1}\big(\mathcal{A}_{yy}u^1 + \mathcal{A}_{xy}\tilde{u} \big) \\ &\quad + \varepsilon^0 \big(\mathcal{A}_{yy}u^2 + \mathcal{A}_{xy}u^1 + \mathcal{A}_{xx}\tilde{u} + \partial_i(a^0_{ij}\partial_j\tilde{u}) \big) \\ &\quad + \varepsilon^1 \big(\partial_t^2 u^1 + \mathcal{A}_{yy}u^3 + \mathcal{A}_{xy}u^2 + \mathcal{A}_{xx}u^1 - \tilde{L}^1\tilde{u} \big) \\ &\quad + \varepsilon^2 \big(\partial_t^2 u^2 + \mathcal{A}_{yy}u^4 + \mathcal{A}_{xy}u^3 + \mathcal{A}_{xx}u^2 - \tilde{L}^2\tilde{u} \big) \\ &\quad + (\partial_t^2 + \mathcal{A}^\varepsilon)\varphi + \mathcal{O}(\varepsilon^3), \end{aligned} \tag{6.70}$$

where the u^i are evaluated at $\big(t,x,y=\frac{x}{\varepsilon}\big)$. We now look for $u^1,\dots,u^4$ such that the terms of order $\mathcal{O}(\varepsilon^{-1})$ to $\mathcal{O}(\varepsilon^2)$ in (6.70) vanish. Note that the role of the u^k is to cancel the terms containing $\tilde{u}$ and the role of φ is to cancel the terms containing f that will appear.

Canceling the ε^{-1}, ε^0 and ε terms and derivation of the constraints defining $\tilde{L}^1$

The cancellation of the term of order $\mathcal{O}(\varepsilon^{-1})$ in (6.70) leads to defining

$$u^1(t,x,y) = \chi_i(x,y)\partial_i\tilde{u}(t,x), \tag{6.71}$$

where for $x\in\Omega$, $1\le i\le d$, $\chi_i(x) = \chi_i(x,\cdot)$ is Y-periodic and solves the cell problem

$$\varepsilon^{-1}: \quad \big(a(x)\nabla_y\chi_i(x),\nabla_y w\big)_Y = -\big(a(x)e_i,\nabla_y w\big)_Y, \tag{6.72}$$

for all test functions $w\in \mathrm{W_{per}}(Y)$. Referring to Appendix A.2, $F\in[\mathrm{H}^1_{\mathrm{per}}(Y)]^*$ given by

$$\big\langle F,w\big\rangle = \big(f^0,w\big)_{\mathrm{L}^2(Y)} + \big(f^1_k,\partial_k w\big)_{\mathrm{L}^2(Y)},$$

for some $f^0, f_1^1, \dots, f_d^1 \in \mathrm{L}^2(Y)$ belongs to $\mathrm{W}^*_{\mathrm{per}}(Y)$ if and only if

$$\big(f^0, 1\big)_{\mathrm{L}^2(Y)} = 0. \tag{6.73}$$

As the right hand side of (6.72) satisfies trivially the condition (6.73), it belongs to $\mathrm{W}^*_{\mathrm{per}}(Y)$ and the equation is well-posed in $\mathrm{W}_{\mathrm{per}}(Y)$. Next, the equation obtained by canceling the term of order $\mathcal{O}(1)$ reads now

$$\begin{aligned} -\nabla_y \cdot (a \nabla_y u^2) = \ & \big(\nabla_y \cdot (e_i \chi_j) + e_i^T a (\nabla_y \chi_j + e_j) - a^0_{ij}\big) \partial^2_{ij} \tilde{u} \\ & + \big(\nabla_y \cdot (\nabla_x \chi_i) + \nabla_x \cdot a (\nabla_y \chi_i + e_i) - \nabla_x \cdot (a^0 e_i)\big) \partial_i \tilde{u}. \end{aligned}$$

Compared to the uniformly periodic case, observe that a supplementary term coming from the variation in x appears in this equation. To satisfy this equality, we can define

$$u^2(t,x,y) = \theta^0_{ij}(x,y) \partial^2_{ij} \tilde{u}(t,x) + \theta^1_i(x,y) \partial_i \tilde{u}(t,x), \tag{6.74}$$

where for $x \in \Omega$, $1 \le i,j \le d$, $\theta^0_{ij}(x) = \theta^0_{ij}(x,\cdot)$ and $\theta^1_i(x) = \theta^1_i(x,\cdot)$ belong to $\mathrm{W}_{\mathrm{per}}(Y)$ and solve the cell problems

$$\varepsilon^0 :$$

$$\begin{aligned} \big(a(x) \nabla_y \theta^0_{ij}(x), \nabla_y w\big)_Y = & - \big(a(x) e_i \chi_j(x), \nabla_y w\big)_Y \\ & + \big(a(x)(\nabla_y \chi_j(x) + e_j) - a^0(x) e_j, e_i w\big)_Y, \end{aligned} \tag{6.75a}$$

$$\begin{aligned} \big(a(x) \nabla_y \theta^1_i(x), \nabla_y w\big)_Y = & - \big(a(x) \nabla_x \chi_i(x), \nabla_y w\big)_Y \\ & + \big(\nabla_x \cdot a(x)(\nabla_y \chi_i(x) + e_i) - \nabla_x \cdot (a^0(x) e_i), w\big)_Y, \end{aligned} \tag{6.75b}$$

for all test functions $w \in \mathrm{W}_{\mathrm{per}}(Y)$. In order to apply Lax–Milgram theorem and obtain the well-posedness of these equations, we need to verify that the right hand sides belong to $\mathrm{W}^*_{\mathrm{per}}(Y)$ or equivalently that they satisfy (6.73). As the homogenized tensor a^0 is defined as

$$a^0_{ij}(x) = \big\langle e_i^T a(x)(\nabla_y \chi_j(x) + e_j) \big\rangle_Y, \tag{6.76}$$

the right hand side of (6.75a) has zero mean and thus, for all $x \in \Omega$, $\theta^0_{ij}(x) \in \mathrm{W}_{\mathrm{per}}(Y)$ exists and is unique. Let us now check that (6.75b) is well-posed. Using (6.76), we have

$$\big(\nabla_x \cdot a(\nabla_y \chi_i + e_i) - \nabla_x \cdot (a^0 e_i), 1\big)_Y = |Y| \partial_m \Big(\big\langle e_m^T a (\nabla_y \chi_i + e_i) \big\rangle_Y - a^0_{mi} \Big) = 0,$$

so that the right hand side of (6.75b) satisfies (6.73) and thus belongs to $\mathrm{W}^*_{\mathrm{per}}(\Omega)$. Hence, (6.75b) is well-posed in $\mathrm{W}_{\mathrm{per}}(Y)$ and, for all $x \in \Omega$, $\theta^1_i(x) \in \mathrm{W}_{\mathrm{per}}(Y)$ exists and is unique. At this point, we have defined an adaptation such that $(\partial_t^2 + \mathcal{A}^\varepsilon)(\boldsymbol{\mathcal{B}}^\varepsilon \tilde{u} - u^\varepsilon) = \mathcal{O}(\varepsilon)$. Hence, following the process described in Section 4.2.2, we can prove the classical homogenization result at short times $T = \mathcal{O}(1)$, for a locally periodic tensor (under suitable regularity assumptions). In order to find effective equations at timescales $\mathcal{O}(\varepsilon^{-2})$, we continue and cancel the higher order terms in (6.70). Taking into account the definitions of u^1 and u^2 and the effective equation (6.69), we have

$$\begin{aligned} \partial_t^2 u^1 &= \chi_k \partial_k \partial_t^2 \tilde{u} = \chi_k \partial_k f + \chi_k \partial_{km} (a^0_{mn} \partial_n \tilde{u}) - \varepsilon \chi_i \partial_k \tilde{L}^1 \tilde{u} + \mathcal{O}(\varepsilon^2), \\ \partial_t^2 u^2 &= \theta^0_{ij} \partial^2_{ij} \partial_t^2 \tilde{u} + \theta^1_i \partial_i \partial_t^2 \tilde{u} = \theta^0_{ij} \partial^2_{ij} f + \theta^1_i \partial_i f + \theta^0_{ij} \partial^3_{ijm} (a^0_{mn} \partial_n \tilde{u}) + \theta^1_k \partial_{km} (a^0_{mn} \partial_n \tilde{u}) + \mathcal{O}(\varepsilon). \end{aligned}$$

These equalities used in (6.70) lead to

$$\begin{aligned} R^\varepsilon = \ & \varepsilon \ \big(\mathcal{A}_{yy} u^3 + \mathcal{A}_{xy} u^2 + \mathcal{A}_{xx} u^1 + \chi_k \partial^2_{km} (a^0_{mn} \partial_n \tilde{u}) - \tilde{L}^1 \tilde{u} \big) \\ & + \varepsilon^2 \big(\mathcal{A}_{yy} u^4 + \mathcal{A}_{xy} u^3 + \mathcal{A}_{xx} u^2 + \theta^0_{ij} \partial^3_{ijm} (a^0_{mn} \partial_n \tilde{u}) + \theta^1_i \partial^2_{im} (a^0_{mn} \partial_n \tilde{u}) - \chi_i \partial_i \tilde{L}^1 \tilde{u} - \tilde{L}^2 \tilde{u} \big) \\ & + (\partial_t^2 + \mathcal{A}^\varepsilon) \varphi + \varepsilon \chi_i \partial_i f + \varepsilon^2 (\theta^0_{ij} \partial^2_{ij} f + \theta^1_i \partial_i f) + \mathcal{O}(\varepsilon^3), \end{aligned} \tag{6.77}$$

We are now looking for u^3 such that the $\mathcal{O}(\varepsilon)$ order term in (6.77) cancels. We thus define

$$u^3(t,x,y) = \kappa^0_{ijk}(x,y)\partial^3_{ijk}\tilde{u}(t,x) + \kappa^1_{ij}(x,y)\partial^2_{ij}\tilde{u}(t,x) + \kappa^2_i(x,y)\partial_i\tilde{u}(t,x), \tag{6.78}$$

where $\kappa^0_{ijk}(x,\cdot), \kappa^1_{ij}(x,\cdot)$ and $\kappa^2_i(x,\cdot)$ are solutions of cell problems to define. We now need to define $\tilde{L}^1$ such that these cell problems are well-posed. The first idea is to set $\tilde{L}^1 = a^{13}_{ijk}(x)\partial^3_{ijk} - a^{12}_{ij}(x)\partial^2_{ij} + a^{11}_i(x)\partial_i$ and to define the coefficients a^{13}, a^{12}, a^{11} thanks to the constraints obtained for the solvability of the cell problems. However, we also have to ensure that $\tilde{L}^1$ allows the well-posedness of the effective equation (6.69). From the uniformly periodic case, we can anticipate that $a^{13}_{ijk}(x)\partial^3_{ijk} = 0$. Nevertheless, the operator $-\varepsilon a^{12}_{ij}(x)\partial^2_{ij}$ need not to deteriorate the ellipticity of $-\partial_i(a^0_{ij}\partial_j\cdot)$ in the effective equation and thus a^{12}_{ij} has to be positive semidefinite. This condition can not be ensured in general by the obtained tensor. As in Section 6.1.3, we thus apply a Boussinesq trick. Namely, we add the term $b^{10}\partial^2_t$ in $\tilde{L}^1$. Observe that if we formally substitute $\partial^2_t\tilde{u} = f - \partial_i(a^0_{ij}\partial_j\tilde{u})$ in $\tilde{L}^1\tilde{u}$, the constraint imposed by the well-posedness of the cell problem for κ^1_{ij} applies on $a^{12}_{ij} - b^{10}a^0_{ij}$. As a^0 is positive definite, we can then find $b^{10} \geq 0$ and a^{12} positive semidefinite satisfying it. Let then

$$\tilde{L}^1 = a^{13}_{ijk}(x)\partial^3_{ijk} - a^{12}_{ij}(x)\partial^2_{ij} + a^{11}_i(x)\partial_i + b^{10}(x)\partial^2_t. \tag{6.79}$$

Using the effective equation, we obtain

$$\tilde{L}^1\tilde{u} = \tilde{L}^{1,x}\tilde{u} + b^{10}\partial_m(a^0_{mn}\partial_n\tilde{u}) + b^{10}f + \varepsilon b^{10}\tilde{L}^1\tilde{u} + \mathcal{O}(\varepsilon^2), \tag{6.80}$$

where we denoted $\tilde{L}^{1,x} = \tilde{L}^1 - b^{10}(x)\partial^2_t$, the spatial part of $\tilde{L}^1$. Hence, we rewrite (6.77) as

$$\begin{aligned} R^\varepsilon = \;& \varepsilon\big(\mathcal{A}_{yy}u^3 + \mathcal{A}_{xy}u^2 + \mathcal{A}_{xx}u^1 + \chi_i\partial^2_{im}(a^0_{mn}\partial_n\tilde{u}) - \tilde{L}^{1,x}\tilde{u} - b^{10}\partial_m(a^0_{mn}\partial_n\tilde{u})\big) \\ & + \varepsilon^2\big(\mathcal{A}_{yy}u^4 + \mathcal{A}_{xy}u^3 + \mathcal{A}_{xx}u^2 + \theta^0_{ij}\partial^3_{ijm}(a^0_{mn}\partial_n\tilde{u}) + \theta^1_i\partial^2_{im}(a^0_{mn}\partial_n\tilde{u}) \\ & \qquad - \chi_i\partial_i(\tilde{L}^1\tilde{u}) + b^{10}\tilde{L}^1\tilde{u} - \tilde{L}^2\tilde{u}\big) \\ & + (\partial^2_t + \mathcal{A}^\varepsilon)\varphi + \varepsilon\chi_i\partial_i f + \varepsilon^2(\theta^0_{ij}\partial^2_{ij}f + \theta^1_i\partial_i f) - \varepsilon b^{10}f + \mathcal{O}(\varepsilon^3). \end{aligned} \tag{6.81}$$

Recall that u^3 is defined as (6.78), hence, rewriting explicitly the equations obtained by canceling the $\mathcal{O}(\varepsilon)$ order term in (6.81), we obtain the following cell problems: for $x \in \Omega$, $\kappa^0_{ijk}(x), \kappa^1_{ij}(x), \kappa^2_i(x) \in \mathrm{W}_{\mathrm{per}}(Y)$ satisfy (we do not specify the evaluation in x for readability)

ε^1 :

$$\begin{aligned} \big(a\nabla_y\kappa^0_{ijk}, \nabla_y w\big)_Y =& -\big(ae_i\theta^0_{jk}, \nabla_y w\big)_Y \\ & + \big(a(\nabla_y\theta^0_{jk} + e_j\chi_k), e_i w\big)_Y - \big(a^0_{ij}\chi_k, w\big)_Y + \big(a^{13}_{ijk}, w\big)_Y, \end{aligned} \tag{6.82a}$$

$$\begin{aligned} \big(a\nabla_y\kappa^1_{ij}, \nabla_y w\big)_Y =& -\big(a(\nabla_x\theta^0_{ij} + e_i\theta^1_j), \nabla_y w\big)_Y + \big(\nabla_x\cdot a(\nabla_y\theta^0_{ij} + e_i\chi_j), w\big)_Y \\ & + \big(a(\nabla_y\theta^1_j + \nabla_x\chi_j), e_i w\big)_Y - \big(\chi_i\partial_m a^0_{mj} + \chi_m\partial_m a^0_{ij}, w\big)_Y \\ & - \big(a^{12}_{ij} - b^{10}a^0_{ij}, w\big)_Y, \end{aligned} \tag{6.82b}$$

$$\begin{aligned} \big(a\nabla_y\kappa^2_i, \nabla_y w\big)_Y =& -\big(a\nabla_x\theta^1_i, \nabla_y w\big)_Y + \big(\nabla_x\cdot a(\nabla_y\theta^1_i + \nabla_x\chi_i), w\big)_Y \\ & - \big(\chi_m\partial^2_{mn}a^0_{ni}, w\big)_Y + \big(b^{10}\partial_m a^0_{mi} + a^{11}_i, w\big)_Y, \end{aligned} \tag{6.82c}$$

for all test functions $w \in \mathrm{W}_{\mathrm{per}}(Y)$. We enforce the right hand sides of these equations to satisfy the solvability condition (6.73), i.e., to belong to $\mathrm{W}^*_{\mathrm{per}}(Y)$, and that leads to the following constraints on the tensors (recall that $\langle\chi_i(x)\rangle_Y = 0$):

$$|Y|a^{13}_{ijk} = -\big(a(\nabla_y\theta^0_{jk} + e_j\chi_k), e_i\big)_Y, \tag{6.83a}$$

$$|Y|(a^{12}_{ij} - b^{10}a^0_{ij}) = \big(\nabla_x\cdot a(\nabla_y\theta^0_{ij} + e_i\chi_j), 1\big)_Y + \big(a(\nabla_y\theta^1_j + \nabla_x\chi_j), e_i\big)_Y, \tag{6.83b}$$

$$|Y|a^{11}_i = -\big(\nabla_x\cdot a(\theta^1_i + \nabla_x\chi_i), 1\big)_Y - |Y|b^{10}\partial_m a^0_{mi}. \tag{6.83c}$$

We emphasize that the constraints (6.83) must hold locally for each $x \in \Omega$. These expressions and the expression for $\tilde{L}^1$ are simplified in the two following lemmas.

Lemma 6.2.6. *The constraints on a^{13}, a^{12}, b^{10} and a^{11} in* (6.83) *can be rewritten for all $x \in \Omega$ as*

$$a^{13}_{ijk}(x) = (p^{13}_{ijk} - p^{13}_{kji})(x), \qquad p^{13}_{ijk} = \big\langle a(\nabla_y \chi_k + e_k) \cdot e_j \chi_i \big\rangle_Y, \tag{6.84a}$$

$$(a^{12}_{ij} - b^{10} a^0_{ij})(x) = -\partial_m a^{13}_{mij}(x) + p^{12}_{ij}(x), \qquad p^{12}_{ij} = \big\langle a(\nabla_y \theta^1_j + \nabla_x \chi_j) \cdot e_i \big\rangle_Y, \tag{6.84b}$$

$$a^{11}_i(x) = -\partial_m p^{12}_{mi}(x) - b^{10} \partial_m a^0_{mi}(x). \tag{6.84c}$$

Furthermore, $p^{12}(x)$ can be expressed as

$$p^{12}_{ij}(x) = -\partial_m p^{13}_{imj}(x) + q^{12}_{ij}(x) + q^{12}_{ji}(x), \qquad q^{12}_{ij} = \big\langle a(\nabla_y \chi_j + e_j) \cdot \nabla_x \chi_i \big\rangle_Y. \tag{6.84d}$$

Proof. Let us denote $(\cdot,\cdot)_Y$ as $(\cdot,\cdot)$ and $\langle\cdot\rangle_Y$ as $\langle\cdot\rangle$. We first prove (6.84a). Using (6.72) with the test function $w = \theta^0_{jk}$ and (6.75a) with $w = \chi_i$, we have

$$\begin{aligned} -\big(a(\nabla_y \theta^0_{jk} + e_j \chi_k), e_i\big) &= \big(a \nabla_y \theta^0_{jk}, \nabla_y \chi_i\big) - \big(a e_j \chi_k, e_i\big) \\ &= -\big(a e_j \chi_k, \nabla_y \chi_i + e_i\big) + \big(a(\nabla_y \chi_k + e_k), e_j \chi_i\big), \end{aligned}$$

which, thanks to the symmetry of $a(x,y)$ proves (6.84a). Let us now prove (6.84b). Thanks to (6.83a), the first term of (6.83b) is

$$\big(\nabla_x \cdot a(\nabla_y \theta^0_{ij} + e_i \chi_j), 1\big) = \partial_m \big(a(\nabla_y \theta^0_{ij} + e_i \chi_j), e_m\big) = -|Y| \partial_m a^{13}_{mij},$$

and thus (6.83b) can be rewritten as (6.84b). To rewrite a^{11}_i as in (6.84c), we simply note that $-\big(\nabla_x \cdot a(\nabla_y \theta^1_i + \nabla_x \chi_i), 1\big) = -|Y| \partial_m p^{12}_{mi}$. Finally, let us prove (6.84d). Using (6.72) with the test function $w = \theta^1_j$ and (6.75b) with $w = \chi_i$, we have

$$\big(a(\nabla_y \theta^1_j + \nabla_x \chi_j), e_i\big) = -\big(a \nabla_y \theta^1_j, \nabla_y \chi_i\big) + \big(a \nabla_x \chi_j, e_i\big) = \big(a \nabla_x \chi_j, \nabla_y \chi_i + e_i\big) - \big(\nabla_x \cdot a(\nabla_y \chi_j + e_j), \chi_i\big).$$

Furthermore,

$$-\big(\nabla_x \cdot a(\nabla_y \chi_j + e_j), \chi_i\big) = -\partial_m \big(a(\nabla_y \chi_j + e_j), e_m \chi_i\big) + \big(a(\nabla_y \chi_j + e_j), \nabla_x \chi_i\big) = |Y|(-\partial_m p^{13}_{imj} + q^{12}_{ij}),$$

and thus, combining the two last equalities gives (6.84d). The proof of the lemma is complete. □

We then verify that the two operators $\tilde{L}^1$ and L^1 coincide.

Lemma 6.2.7. *Let $\bar{a}^{12}$ and b^{10} be the tensors defined in* (6.61) *and assume that $\bar{a}^{12} \in \mathcal{C}^1(\bar{\Omega})$. Let also $\tilde{L}^1$ and L^1 be the operators defined in* (6.79) *and* (6.60)*, respectively. Then $\tilde{L}^1 v = L^1 v$ for any $v \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^3(\Omega))$ with $\partial_t^2 v \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{L}^2(\Omega))$.*

Proof. First, note that thanks to (6.84a), we have $S^3_{ijk}\{a^{13}_{ijk}\} = 0$ and thus $a^{13}_{ijk} \partial^3_{ijk} v = 0$. Furthermore, thanks to (6.84a), (6.84b), and (6.84d), we verify that $S^2_{ij}\{a^{12}_{ij}\} = \bar{a}^{12}_{ij}$. Hence, we have

$$\tilde{L}^1 v - b^{10} \partial_t^2 v = -S^2_{ij}\{a^{12}_{ij}\} \partial^2_{ij} v + a^{11}_i \partial_i v = -\partial_i \big(\bar{a}^{12}_{ij} \partial_j v\big) + \big(a^{11}_i + \partial_m (S^2_{mi}\{a^{12}_{mi}\})\big) \partial_i v. \tag{6.85}$$

We claim that $a^{11}_i + \partial_m (S^2_{mi}\{a^{12}_{mi}\}) = 0$. To prove it, note that as b^{10} is constant, using (6.84b) and (6.84c), we have

$$a^{11}_i + \partial_m (S^2_{mi}\{a^{12}_{mi}\}) = \tfrac{1}{2} \partial_m \big(p^{12}_{im} - p^{12}_{mi}\big) - \tfrac{1}{2} \partial^2_{mn} \big(a^{13}_{nmi} + a^{13}_{nim}\big).$$

Using then (6.84a) and (6.84d), we verify that

$$a^{11}_i + \partial_m (S^2_{mi}\{a^{12}_{mi}\}) = \tfrac{1}{2} \partial^2_{mn} \big(- p^{13}_{inm} + p^{13}_{mni} - p^{13}_{nmi} + p^{13}_{imn} - p^{13}_{nim} + p^{13}_{min}\big) = 0,$$

and the claim is proved. Combined with (6.85), the claim concludes the proof of the lemma. □

Canceling the ε^2 terms and derivation of the constraints defining $\tilde{L}^2$

We come back to the asymptotic expansion. The next step is to cancel the $\mathcal{O}(\varepsilon^2)$ order term containing $\tilde{u}$ in (6.81). Following the same reasoning as for u^3, we look for u^4 of the form

$$u^4(t,x,y) = \rho^0_{ijkl}(x,y)\partial^4_{ijkl}\tilde{u}(t,x) + \rho^1_{ijk}(x,y)\partial^3_{ijk}\tilde{u}(t,x) + \rho^2_{ij}(x,y)\partial^2_{ij}\tilde{u}(t,x) + \rho^3_i(x,y)\partial_i\tilde{u}(t,x), \tag{6.86}$$

for some correctors $\rho^0, \rho^1, \rho^2, \rho^3$ to be defined. The ansatz on the form of $\tilde{L}^2$ could be $\partial^2_{ij}(a^{24}_{ijkl}\partial^2_{kl}) + a^{23}_{ijk}\partial^3_{ijk} - a^{22}_{ij}\partial^2_{ij} + a^{21}_i\partial_i$. However, referring to the argument presented in Section 6.1.3 in one dimension, this choice cannot guarantee the well-posedness of the effective equation (6.69). We thus apply Boussinesq tricks. First, similarly as for $\tilde{L}^1$, we add the operator $b^{20}\partial^2_t$ in order to obtain a constraint on the difference $a^{22}_{ij} - b^{20}a^0_{ij}$. Second, we know from the uniformly periodic case that the constraint on a^{24}_{ijkl} leads to a negative tensor. Hence, we add the term $-\partial_i(b^{22}_{ij}\partial_j\partial^2_t\cdot)$ in $\tilde{L}^2$ in order to obtain a constraint on $a^{24}_{ijkl} - a^0_{jk}b^{22}_{il}$. Note that this trick is not possible for the operator of order 3. Nevertheless, we will see that we can find a tensor a^{23} such that $a^{23}_{ijk}\partial^3_{ijk} = 0$ that satisfies the corresponding constraint. We thus define

$$\tilde{L}^2 = \partial^2_{ij}(a^{24}_{ijkl}(x)\partial^2_{kl}\cdot) - \partial_i(b^{22}_{ij}(x)\partial_j\partial^2_t\cdot) + a^{23}_{ijk}(x)\partial^3_{ijk} - a^{22}_{ij}(x)\partial^2_{ij} + a^{21}_i(x)\partial_i + b^{20}(x)\partial^2_t, \tag{6.87}$$

and using (6.69), we obtain

$$\tilde{L}^2\tilde{u} = \tilde{L}^{2,x}\tilde{u} - \partial_i(b^{22}_{ij}\partial^2_{jk}(a^0_{kl}\partial_l\tilde{u})) + b^{20}\partial_m(a^0_{mn}\partial_n\tilde{u}) - \partial_i(b^{22}_{ij}\partial_j f) + b^{20}f + \mathcal{O}(\varepsilon), \tag{6.88}$$

where $\tilde{L}^{2,x} = \tilde{L}^2 + \partial_i(b^{22}_{ij}\partial_j\partial^2_t\cdot) - b^{20}\partial^2_t$ is the spatial part of $\tilde{L}^2$. Taking into account the definition of $\tilde{L}^1$ and using (6.69), we have

$$\begin{aligned}
&\chi_i\partial_i(\tilde{L}^1\tilde{u}) = \chi_i\partial_i(\tilde{L}^{1,x}\tilde{u}) + \chi_i\partial_i(b^{10}\partial_m(a^0_{mn}\partial_n\tilde{u})) + \chi_i\partial_i(b^{10}f) + \mathcal{O}(\varepsilon),\\
&b^{10}\tilde{L}^1\tilde{u} = b^{10}\tilde{L}^{1,x}\tilde{u} + (b^{10})^2\partial_m(a^0_{mn}\partial_n\tilde{u}) + (b^{10})^2 f + \mathcal{O}(\varepsilon).
\end{aligned}$$

Therefore, using (6.74), (6.78), (6.86), (6.80) and (6.88), we rewrite the $\mathcal{O}(\varepsilon^2)$ order term in (6.81) as

$$\begin{aligned}
R^\varepsilon =& \varepsilon^2\big(\mathcal{A}_{yy}u^4 + \mathcal{A}_{xy}u^3 + \mathcal{A}_{xx}u^2 + \theta^0_{ij}\partial^3_{ijm}(a^0_{mn}\partial_n\tilde{u}) + \theta^1_i\partial^2_{im}(a^0_{mn}\partial_n\tilde{u}) - \chi_i\partial_i(\tilde{L}^{1,x}\tilde{u})\\
&+ \chi_i\partial_i(b^{10}\partial_m(a^0_{mn}\partial_n\tilde{u})) - b^{10}\tilde{L}^{1,x}\tilde{u} + (b^{10})^2\partial_m(a^0_{mn}\partial_n\tilde{u})\\
&- \tilde{L}^{2,x}\tilde{u} + \partial_i(b^{22}_{ij}\partial^2_{jm}(a^0_{mn}\partial_n\tilde{u})) - b^{20}\partial_m(a^0_{mn}\partial_n\tilde{u})\big)\\
&+ \varepsilon(\chi_i\partial_i f - b^{10}f) + \varepsilon^2\big(\theta^0_{ij}\partial^2_{ij}f + \theta^1_i\partial_i f - \chi_i\partial_i(b^{10}f) + (b^{10})^2 f + \partial_i(b^{22}_{ij}\partial_j f) - b^{20}f\big)\\
&+ (\partial^2_t + \mathcal{A}^\varepsilon)\varphi + \mathcal{O}(\varepsilon^3).
\end{aligned} \tag{6.89}$$

Canceling this term leads to the following cell problems: for $x \in \Omega$, $1 \le i,j,k,l \le d$, we look for $\rho^0_{ijkl}(x), \rho^1_{ijk}(x), \rho^2_{ij}(x), \rho^3_i(x) \in \mathrm{W}_{\mathrm{per}}(Y)$ such that

ε^2 :

$$\begin{aligned}
\big(a\nabla_y\rho^0_{ijkl}, \nabla_y w\big)_Y =& -\big(ae_i\kappa^0_{jkl}, \nabla_y w\big)_Y + \big(a(\nabla_y\kappa^0_{jkl} + e_j\theta^0_{kl}), e_i w\big)_Y\\
&+ \big(a^{13}_{jkl}\chi_i - a^0_{ij}\theta^0_{kl}, w\big)_Y + \big(a^{24}_{ijkl} - a^0_{jk}b^{22}_{il}, w\big)_Y,
\end{aligned} \tag{6.90a}$$

$$\begin{aligned}
\big(a\nabla_y\rho^1_{ijk}, \nabla_y w\big)_Y =& -\big(a(e_i\kappa^1_{jk} + \nabla_x\kappa^0_{ijk}), \nabla_y w\big)_Y + \big(\nabla_x\cdot a(\nabla_y\kappa^0_{ijk} + e_i\theta^0_{jk}), w\big)_Y\\
&+ \big(a(\nabla_y\kappa^1_{jk} + \nabla_x\theta^0_{jk} + e_j\theta^1_k), e_i w\big)_Y\\
&+ \big(\chi_m\partial_m a^{13}_{ijk} + (b^{10}a^0_{ij} - a^{12}_{ij})\chi_k\\
&\qquad - \theta^0_{ij}\partial_m a^0_{mk} - \theta^0_{mi}\partial_m a^0_{jk} - \theta^0_{im}\partial_m a^0_{jk} - a^0_{ij}\theta^1_k, w\big)_Y\\
&+ \big(\partial_m(a^{24}_{imjk} + a^{24}_{mijk} - b^{22}_{mk}a^0_{ij}) - b^{22}_{im}\partial_m a^0_{jk} - b^{22}_{ij}\partial_m a^0_{mk}\\
&\qquad - b^{10}a^{13}_{ijk} + a^{23}_{ijk}, w\big)_Y,
\end{aligned} \tag{6.90b}$$

$$\begin{aligned}\big(a\nabla_y\rho^2_{ij},\nabla_y w\big)_Y =& -\big(a(e_i\kappa^2_j+\nabla_x\kappa^1_{ij}),\nabla_y w\big)+\big(\nabla_x\cdot a(\nabla_y\kappa^1_{ij}+\nabla_x\theta^0_{ij}+e_i\theta^1_j),w\big)_Y\\ &+\big(a(\nabla_y\kappa^2_j+\nabla_x\theta^1_j),e_i w\big)_Y\\ &+\big(\chi_m\partial_m(b^{10}a^0_{ij}-a^{12}_{ij})+b^{10}\chi_i\partial_m a^0_{mj}+a^{11}_i\chi_j\\ &\qquad -\theta^1_i\partial_m a^0_{mj}-\theta^1_m\partial_m a^0_{ij}-(\theta^0_{im}+\theta^0_{mi})\partial^2_{mn}a^0_{nj}-\theta^0_{mn}\partial^2_{mn}a^0_{ij},w\big)_Y\\ &+\big(\partial^2_{mn}a^{24}_{mnij}-\partial_m(b^{22}_{mi}\partial_n a^0_{nj})-\partial_m(b^{22}_{mn}\partial_n a^0_{ij})-b^{22}_{im}\partial^2_{mn}a^0_{nj}\\ &\qquad +b^{10}(a^{12}_{ij}-b^{10}a^0_{ij})-(a^{22}_{ij}-b^{20}a^0_{ij}),w\big)_Y,\end{aligned}\tag{6.90c}$$

$$\begin{aligned}\big(a\nabla_y\rho^3_i,\nabla_y w\big)_Y =& -\big(a\nabla_x\kappa^2_i,\nabla_y w\big)_Y+\big(\nabla_x\cdot a(\nabla_y\kappa^2_i+\nabla_x\theta^1_i),w\big)_Y\\ &+\big(\chi_m\partial_m a^{11}_i+\chi_m\partial_m(b^{10}\partial_n a^0_{ni})-\theta^0_{mn}\partial^3_{mnp}a^0_{pi}-\theta^1_m\partial^2_{mn}a^0_{ni},w\big)_Y\\ &+\big(b^{20}\partial_m a^0_{mi}-b^{10}(b^{10}\partial_m a^0_{mi}+a^{11}_i)-\partial_m(b^{22}_{mn}\partial^2_{np}a^0_{pi})+a^{21}_i,w\big)_Y,\end{aligned}\tag{6.90d}$$

for all test functions $w\in \mathrm{W}_{\mathrm{per}}(Y)$. We enforce the right hand sides of (6.90a-6.90d) to satisfy (6.73), which leads to the following constraints (recall that $\chi_i(x),\theta^0_{ij}(x)$ and $\theta^1_i(x)$ have zero mean):

$$|Y|(a^{24}_{ijkl}-a^0_{jk}b^{22}_{il})=-\big(a(\nabla_y\kappa^0_{jkl}+e_j\theta^0_{kl}),e_i\big)_Y,\tag{6.91a}$$

$$\begin{aligned}|Y|a^{23}_{ijk}=&-\big(\nabla_x\cdot a(\nabla_y\kappa^0_{ijk}+e_i\theta^0_{jk}),1\big)_Y-\big(a(\nabla_y\kappa^1_{jk}+\nabla_x\theta^0_{jk}+e_j\theta^1_k),e_i\big)_Y\\ &+|Y|\big(b^{22}_{im}\partial_m a^0_{jk}+b^{22}_{ij}\partial_m a^0_{mk}+b^{10}a^{13}_{ijk}+\partial_m(b^{22}_{mk}a^0_{ij}-a^{24}_{imjk}-a^{24}_{mijk})\big),\end{aligned}\tag{6.91b}$$

$$\begin{aligned}|Y|(a^{22}_{ij}-b^{20}a^0_{ij})=&\big(\nabla_x\cdot a(\nabla_y\kappa^1_{ij}+\nabla_x\theta^0_{ij}+e_i\theta^1_j),1\big)_Y+\big(a(\nabla_y\kappa^2_j+\nabla_x\theta^1_j),e_i\big)_Y\\ &+|Y|\big(\partial^2_{mn}a^{24}_{mnij}-\partial_m(b^{22}_{mi}\partial_n a^0_{nj})-\partial_m(b^{22}_{mn}\partial_n a^0_{ij})-b^{22}_{im}\partial^2_{mn}a^0_{nj}\\ &\qquad +b^{10}(a^{12}_{ij}-b^{10}a^0_{ij})\big),\end{aligned}\tag{6.91c}$$

$$\begin{aligned}|Y|a^{21}_i=&-\big(\nabla_x\cdot a(\nabla_y\kappa^2_i+\nabla_x\theta^1_i),1\big)_Y\\ &+|Y|\big(b^{10}(b^{10}\partial_m a^0_{mi}+a^{11}_i)-b^{20}\partial_m a^0_{mi}+\partial_m(b^{22}_{mn}\partial^2_{np}a^0_{pi})\big),\end{aligned}\tag{6.91d}$$

where each constraint is given locally for $x\in\Omega$. These expressions are simplified in the following Lemma.

Lemma 6.2.8. *Denote $R_{ij}(x)=b^{22}_{ij}(x)-\big\langle\chi_j(x)\chi_i(x)\big\rangle_Y$. Then the constraints on a^{24}, b^{22}, a^{23}, a^{22}, b^{20} and a^{21} given in (6.91) can be rewritten as*

$$a^{24}_{ijkl}=\big\langle a_{jk}\chi_l\chi_i\big\rangle_Y-\big\langle a\nabla_y\theta^0_{kl}\cdot\nabla_y\theta^0_{ji}\big\rangle_Y+a^0_{jk}R_{il},\tag{6.92a}$$

$$a^{23}_{ijk}=p^{23}_{ijk}-p^{23}_{kji}+b^{10}a^{13}_{ijk}-\partial_m(a^0_{mj}R_{ik})+\partial_m a^0_{mk}R_{ij}+\partial_m a^0_{jk}R_{mi},\tag{6.92b}$$

$$p^{23}_{ijk}=\big\langle ae_j\chi_i\cdot\nabla_x\chi_k\big\rangle_Y-\big\langle a\nabla_y\theta^0_{ji}\cdot\nabla_y\theta^1_k\big\rangle_Y,\tag{6.92c}$$

$$\begin{aligned}a^{22}_{ij}-b^{20}a^0_{ij}=&\partial_m(p^{23}_{jim}-p^{23}_{mij}-p^{23}_{imj})+p^{22}_{ij}+b^{10}(a^{12}_{ij}-b^{10}a^0_{ij})\\ &+\partial^2_{mn}(a^0_{ni}R_{mj})-\partial_m(\partial_n a^0_{nj}R_{mi})-\partial_m(\partial_n a^0_{ij}R_{mn})-\partial^2_{mn}a^0_{nj}R_{im},\end{aligned}\tag{6.92d}$$

$$p^{22}_{ij}=\big\langle a\nabla_x\chi_j\cdot\nabla_x\chi_i\big\rangle_Y-\big\langle a\nabla_y\theta^1_i\cdot\nabla_y\theta^1_j\big\rangle_Y,\tag{6.92e}$$

$$a^{21}_i=\partial^2_{mn}p^{23}_{mni}-\partial_m p^{22}_{mi}+b^{10}(b^{10}\partial_m a^0_{mi}+a^{11}_i)-b^{20}\partial_m a^0_{mi}+\partial_m(\partial^2_{np}a^0_{pi}R_{mn}).\tag{6.92f}$$

Proof. We simply denote $(\cdot,\cdot)_Y$ as $(\cdot,\cdot)$ and $\langle\cdot\rangle_Y$ as $\langle\cdot\rangle$. We first prove (6.92a). Using (6.72) with

the test function $w = \kappa^0_{jkl}$ and (6.82a) with $w = \chi_i$, we have

$$\begin{aligned}-\big(a(\nabla_y\kappa^0_{jkl} + e_j\theta^0_{kl}), e_i\big) &= \big(a\nabla_y\kappa^0_{jkl}, \nabla_y\chi_i\big) - \big(ae_j\theta^0_{kl}, e_i\big)\\ &= -\big(ae_j\theta^0_{kl}, \nabla_y\chi_i + e_i\big) + \big(a\nabla_y\theta^0_{kl}, e_j\chi_i\big) + \big(ae_k\chi_l, e_j\chi_i\big) - a^0_{jk}\big(\chi_l, \chi_i\big).\end{aligned}$$

Cell problem (6.75a) wiht $w = \theta^0_{kl}$ leads then to

$$-\big(a(\nabla_y\kappa^0_{jkl} + e_j\theta^0_{kl}), e_i\big) = -\big(a\nabla_y\theta^0_{kl}, \nabla_y\theta^0_{ji}\big) + \big(a_{jk}\chi_l, \chi_i\big) - a^0_{jk}\big(\chi_l, \chi_i\big),$$

which, used in (6.91a), proves (6.92a). Let us now prove (6.92b). Using (6.91a), we verify that the first term of (6.91b) is

$$-\big(\nabla_x \cdot a(\nabla_y\kappa^0_{ijk} + e_i\theta^0_{jk}), 1\big) = -\partial_m\big(a(\nabla_y\kappa^0_{ijk} + e_i\theta^0_{jk}), e_m\big) = |Y|\partial_m(a^{24}_{mijk} - a^0_{ij}b^{22}_{mk}). \quad (6.93)$$

Then, using cell problems (6.72) with $w = \kappa^1_{jk}$ and (6.82b) with $w = \chi_i$, the second term of (6.91b) satisfies

$$\begin{aligned}-\big(a(\nabla_y\kappa^1_{jk} + \nabla_x\theta^0_{jk} + e_j\theta^1_k), e_i\big) &= \big(a\nabla_y\kappa^1_{jk}, \nabla_y\chi_i\big) - \big(a(\nabla_x\theta^0_{jk} + e_j\theta^1_k), e_i\big)\\ &= -\big(a(\nabla_x\theta^0_{jk} + e_j\theta^1_k), \nabla_y\chi_i + e_i\big) + \big(\nabla_x \cdot a(\nabla_y\theta^0_{jk} + e_j\chi_k), \chi_i\big) + \big(a(\nabla_y\theta^1_k + \nabla_x\chi_k), e_j\chi_i\big)\\ &\quad - \partial_m a^0_{mk}\big(\chi_j, \chi_i\big) - \partial_m a^0_{jk}\big(\chi_m, \chi_i\big). \quad (6.94)\end{aligned}$$

From (6.92a), using cell problem (6.75a) with $w = \theta^0_{jk}$, we verify that

$$\big(a(\nabla_y\theta^0_{jk} + e_j\chi_k), e_m\chi_i\big) = \big(a(\nabla_y\chi_i + e_i), e_m\theta^0_{jk}\big) + |Y|(a^{24}_{imjk} - a^0_{mj}R_{ik}),$$

so that we can rewrite the second term of the right hand side of (6.94) as

$$\begin{aligned}&\big(\nabla_x \cdot a(\nabla_y\theta^0_{jk} + e_j\chi_k), \chi_i\big)\\ &\quad= \partial_m\big(a(\nabla_y\theta^0_{jk} + e_j\chi_k), e_m\chi_i\big) - \big(a(\nabla_y\theta^0_{jk} + e_j\chi_k), \nabla_x\chi_i\big)\\ &\quad= \partial_m\big(a(\nabla_y\chi_i + e_i), e_m\theta^0_{jk}\big) + |Y|\partial_m(a^{24}_{imjk} - a^0_{mj}R_{ik}) - \big(a(\nabla_y\theta^0_{jk} + e_j\chi_k), \nabla_x\chi_i\big). \quad (6.95)\end{aligned}$$

Note that

$$\partial_m\big(a(\nabla_y\chi_i + e_i), e_m\theta^0_{jk}\big) - \big(a(\nabla_y\chi_i + e_i), \nabla_x\theta^0_{jk}\big) = \big(\nabla_x \cdot a(\nabla_y\chi_i + e_i), \theta^0_{jk}\big), \quad (6.96)$$

hence, using (6.95) and (6.96) in (6.94), we obtain

$$\begin{aligned}&-\big(a(\nabla_y\kappa^1_{jk} + \nabla_x\theta^0_{jk} + e_j\theta^1_k), e_i\big)\\ &\quad= \big(ae_j\chi_i, \nabla_y\theta^1_k\big) - \big(a(\nabla_y\chi_i + e_i), e_j\theta^1_k\big) - \big(a\nabla_x\chi_i, \nabla_y\theta^0_{jk}\big) + \big(\nabla_x \cdot a(\nabla_y\chi_i + e_i), \theta^0_{jk}\big)\\ &\qquad+ \big(a\nabla_x\chi_k, e_j\chi_i\big) - \big(ae_j\chi_k, \nabla_x\chi_i\big)\\ &\qquad+ |Y|\big(\partial_m(a^{24}_{imjk} - a^0_{mj}R_{ik}) - \partial_m a^0_{mk}\langle\chi_j\chi_i\rangle - \partial_m a^0_{jk}\langle\chi_m\chi_i\rangle\big).\end{aligned}$$

Using cell problems (6.75a) with $w = \theta^1_k$ and (6.75b) with $w = \theta^0_{jk}$, we finally obtain the expression

$$\begin{aligned}-\big(a(\nabla_y\kappa^1_{jk} + \nabla_x\theta^0_{jk} + e_j\theta^1_k), e_i\big) = |Y|\big(&p^{23}_{ijk} - p^{23}_{kji} + \partial_m(a^{24}_{imjk} - a^0_{mj}R_{ik})\\ &- \partial_m a^0_{mk}\langle\chi_j\chi_i\rangle - \partial_m a^0_{jk}\langle\chi_m\chi_i\rangle\big), \quad (6.97)\end{aligned}$$

where p^{23}_{ijk} is defined is (6.92c). Now, using (6.93) and (6.97), the constraint (6.91b) can be rewritten as (6.92b). Let us now prove (6.92d). First, we use (6.97) to rewrite the first term of (6.91c) as

$$\begin{aligned}\big(\nabla_x \cdot a(\nabla_y\kappa^1_{ij} + \nabla_x\theta^0_{ij} + e_i\theta^1_j), 1\big) = |Y|\big(&\partial_m(p^{23}_{jim} - p^{23}_{mij}) + \partial^2_{mn}(a^0_{ni}R_{mj} - a^{24}_{mnij})\\ &+ \partial_m(\partial_n a^0_{nj}\langle\chi_i\chi_m\rangle) + \partial_m(\partial_n a^0_{ij}\langle\chi_n\chi_m\rangle)\big). \quad (6.98)\end{aligned}$$

Using cell problems (6.72) with $w = \kappa_j^2$ and (6.82c) with $w = \chi_i$, the second term of (6.91c) can be written as

$$\begin{aligned}\big(a(\nabla_y\kappa_j^2 + \nabla_x\theta_j^1), e_i\big) &= -\big(a\nabla_y\kappa_j^2, \nabla_y\chi_i\big) + \big(a\nabla_x\theta_j^1, e_i\big)\\ &= \big(a\nabla_x\theta_j^1, \nabla_y\chi_i + e_i\big) - \big(\nabla_x\cdot a(\nabla_y\theta_j^1 + \nabla_x\chi_j), \chi_i\big) + \partial_{mn}^2 a_{nj}^0\big(\chi_m, \chi_i\big).\end{aligned} \tag{6.99}$$

Now, using cell problems (6.75a) and (6.75b) with $w = \theta_j^1$, the first term in (6.99) can be written as

$$\begin{aligned}\big(a(\nabla_y\chi_i + e_i), \nabla_x\theta_j^1\big) &= \partial_m\big(a(\nabla_y\chi_i + e_i), e_m\theta_j^1\big) - \big(\nabla_x\cdot a(\nabla_y\chi_i + e_i), \theta_j^1\big)\\ &= \partial_m\big(a(\nabla_y\theta_{mi}^0 + e_m\chi_i), \nabla_y\theta_j^1\big) - \big(a(\nabla_y\theta_i^1 + \nabla_x\chi_i), \nabla_y\theta_j^1\big).\end{aligned}$$

Fruthermore, note that the second term of (6.99) can be written as

$$-\big(\nabla_x\cdot a(\nabla_y\theta_j^1 + \nabla_x\chi_j), \chi_i\big) = -\partial_m\big(a(\nabla_y\theta_j^1 + \nabla_x\chi_j), e_m\chi_i\big) + \big(a(\nabla_y\theta_j^1 + \nabla_x\chi_j), \nabla_x\chi_i\big),$$

hence, we obtain from (6.99), after simplification,

$$\begin{aligned}\big(a(\nabla_y\kappa_j^2 + \nabla_x\theta_j^1), e_i\big) &= \partial_m\big(a\nabla_y\theta_{mi}^0, \nabla_y\theta_j^1\big) - \big(a\nabla_y\theta_i^1, \nabla_y\theta_j^1\big)\\ &\quad - \partial_m\big(a\nabla_x\chi_j, e_m\chi_i\big) + \big(a\nabla_x\chi_j, \nabla_x\chi_i\big) + \partial_{mn}^2 a_{nj}^0\big(\chi_m, \chi_i\big)\\ &= |Y|\big(-\partial_m p_{imj}^{23} + p_{ij}^{22} + \partial_{mn}^2 a_{nj}^0\langle\chi_m\chi_i\rangle\big),\end{aligned} \tag{6.100}$$

where p_{ij}^{22} is defined in (6.92e). Now, starting from (6.91c) and using (6.98) and (6.100), we obtain (6.92d) (after simplification). Finally, using (6.100), we have

$$-\big(\nabla_x\cdot a(\nabla_y\kappa_i^2 + \nabla_x\theta_i^1), 1\big) = \partial_{mn}^2 p_{mni}^{23} - \partial_m p_{mi}^{22} - \partial_m(\partial_{np}^2 a_{pi}^0\langle\chi_n\chi_m\rangle),$$

and (6.92f) follows directly from (6.91d). □

We then verify that the two operators $\tilde{L}^2$ and L^2 coincide.

Lemma 6.2.9. *Let $\bar{a}^{24}$, $b^{22}, \bar{a}^{22}$ be the tensors defined in* (6.63) *and* (6.64) *and assume that $\bar{a}^{24} \in \mathcal{C}^2(\bar{\Omega})$ and $b^{22}, \bar{a}^{12} \in \mathcal{C}^1(\bar{\Omega})$. Let also L^2 be the operator defined in* (6.62) *and $\tilde{L}^2$ be the operator defined in* (6.87) *with the tensors given in* (6.92) *where $R_{ij} = \delta a_{ij}^0$ for some $\delta \in \mathbb{R}$. Then $\tilde{L}^2 v = L^2 v$ for any $v \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^4(\Omega))$ with $\partial_t^2 v \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^2(\Omega))$.*

Proof. First, inserting $R_{ij} = \delta a_{ij}^0$ in (6.92d) and using (6.84a), we verify that $S_{ijk}^3\{a_{ijk}^{23}\} = 0$ and thus $a_{ijk}^{23}\partial_{ijk}^3 v = 0$. Second, using (6.92d), (6.61), and the definition of R_{ij}, we verify that $S_{ij}^2\{a_{ij}^{22}\} = \bar{a}_{ij}^{22}$. Furthermore, it holds $S_{ij,kl}^{2,2}\{a_{ijkl}^{24}\} = \bar{a}_{ijkl}^{24}$. Hence, denoting $\tilde{L}^{2,x} = \tilde{L}^2 + \partial_i\big(b_{ij}^{22}\partial_j\partial_t^2\,\cdot\,\big) - b^{20}\partial_t^2$, we have

$$\tilde{L}^{2,x} v = \partial_{ij}^2\big(\bar{a}_{ijkl}^{24}\partial_{kl}^2 v\big) - \partial_i\big(S_{ij}^2\{a_{ij}^{22}\}\partial_j v\big) + \big(a_i^{21} + \partial_m(S_{mi}^2\{a_{mi}^{22}\})\big)\partial_i v. \tag{6.101}$$

We claim that $a_i^{21} + \partial_m(S_{mi}^2\{a_{mi}^{22}\}) = 0$. Indeed, using (6.92d), the form of R_{ij}, and the symmetry of p^{22} and a^0, we compute

$$\begin{aligned}S_{ij}^2\{a_{ij}^{22}\} =& S_{ij}^2\big\{\partial_n(p_{jin}^{23} - p_{nij}^{23} - p_{inj}^{23})\big\} + p_{ij}^{22} + b^{10}S_{ij}^2\{a_{ij}^{12}\} - (b^{10})^2 a_{ij}^0\\ &+ \delta\partial_n a_{pi}^0\partial_p a_{nj}^0 - \delta\partial_p\big(a_{pn}^0\partial_n a_{ij}^0\big) + b^{20}a_{ij}^0.\end{aligned}$$

Note that we have seen in the proof of Proposition 6.2.7 that $a_i^{11} + \partial_m(S_{mi}^2\{a_{mi}^{12}\}) = 0$. Using then (6.92f), direct computations lead to

$$a_i^{21} + \partial_m(S_{mi}^2\{a_{mi}^{22}\}) = \delta\partial_m\big(\partial_{np}^2 a_{pi}^0 a_{mn}^0\big) + \delta\partial_m\big(\partial_n a_{pm}^0\partial_p a_{ni}^0\big) - \delta\partial_{mp}^2\big(a_{pn}^0\partial_n a_{mi}^0\big) = 0,$$

which proves the claim. Combined with (6.101), the claim concludes the proof of the lemma. □

Including a non-zero right hand side

In order to ensure that $R^\varepsilon = \mathcal{O}(\varepsilon^3)$ in (6.89), we still have to define the corrector φ in the adaptation (6.67) to remove the terms coming from the right hand side f. We thus define $\boldsymbol{\varphi} = [\varphi] \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{W}_{\mathrm{per}}(\Omega))$, with $\partial_t\boldsymbol{\varphi} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{L}^2(\Omega))$ and $\partial_t^2\boldsymbol{\varphi} \in \mathrm{L}^2(0,T^\varepsilon;\mathcal{W}^*_{\mathrm{per}}(\Omega))$, as the unique solution of the equation

$$\begin{aligned} &(\partial_t^2 + \mathcal{A}^\varepsilon)\boldsymbol{\varphi}(t) = -\boldsymbol{\mathcal{S}}^\varepsilon f(t) \quad \text{in } \mathcal{W}^*_{\mathrm{per}}(\Omega) \quad \text{for a.e. } \ t \in [0,T^\varepsilon], \\ &\boldsymbol{\varphi}(0) = \partial_t\boldsymbol{\varphi}(0) = [0], \end{aligned} \tag{6.102}$$

where, denoting $\chi_i^\varepsilon = \chi_i\big(\cdot,\frac{\cdot}{\varepsilon}\big), \theta_{ij}^{0\varepsilon} = \theta_{ij}^0\big(\cdot,\frac{\cdot}{\varepsilon}\big), \theta_i^{1\varepsilon} = \theta_i^1\big(\cdot,\frac{\cdot}{\varepsilon}\big)$,

$$\boldsymbol{\mathcal{S}}^\varepsilon f = \big[\varepsilon(\chi_i^\varepsilon\partial_i f - b^{10}f) + \varepsilon^2\big(\theta_{ij}^{0\varepsilon}\partial_{ij}^2 f + \theta_i^{1\varepsilon}\partial_i f - \chi_i^\varepsilon\partial_i(b^{10}f) + (b^{10})^2 f + \partial_i(b_{ij}^{22}\partial_j f) - b^{20}f\big)\big].$$

The standard energy estimate for the wave equation ensures

$$\|\boldsymbol{\varphi}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{W})} \le \|\nabla_x\boldsymbol{\varphi}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} \le C\varepsilon\|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^2(\Omega))}, \tag{6.103}$$

where C depends only on

$$\lambda, \Lambda, \|\chi_i\|_{\mathcal{C}^0(\bar\Omega;\mathcal{C}^0(Y))}, \|b_{ij}^{22}\|_{\mathcal{C}^1(\bar\Omega)}, |b^{10}|, |b^{20}|, \|\theta_{ij}^0\|_{\mathcal{C}^0(\bar\Omega;\mathcal{C}^0(Y))}, \|\theta_i^1\|_{\mathcal{C}^0(\bar\Omega;\mathcal{C}^0(Y))}.$$

We have defined explicitly all the correctors in the adaptation (6.67). Let us show that if the tensor is uniformly periodic, i.e., $a(x,y) = a(y)$, we recover the adaptation and effective equations derived in Section 4.3.2. We already proved in Remark 6.2.3 that the family of effective equations coincides. In addition, we verify that, if we require all the correctors to have zero mean, the adaptation built in this section, is the same as in the uniformly periodic case. Indeed, we have

$$\begin{aligned} &\theta_i^1 = \kappa_{ij}^1 = \kappa_i^2 = \rho_{ijk}^1 = \rho_{ij}^2 = \rho_i^3 = 0, \\ &\chi_i = \hat\chi_i, \quad S^2\{\theta_{ij}^0\} = \hat\theta_{ij}, \quad S^3\{\kappa_{ij}^0\} = \hat\kappa_{ijk}, \quad S^4\{\theta_{ij}^0\} = \hat\rho_{ijkl}, \quad \varphi = \hat\varphi, \end{aligned}$$

where $\hat\chi_i$, $\hat\theta_{ij}$, $\hat\kappa_{ijk}$, $\hat\rho_{ijkl}$ and $\hat\varphi$ are the zero mean correctors defined in the uniformly periodic case in (4.45).

Proof of Theorem 6.2.5

To conclude this section, let us prove Theorem 6.2.5. The adaptation $\mathcal{B}^\varepsilon\tilde u$ is defined explicitly by (6.67), where $u^1,\dots,u^4$ are defined in (6.71), (6.74), (6.78), and (6.86), and $\varphi \in \boldsymbol{\varphi}$ solves (6.102). Then, combining Lemma 6.2.6 with Proposition 6.2.7, we verify that $L^1\tilde u = \tilde L^1\tilde u$, where the tensors involved in the definition of $\tilde L^1\tilde u$ satisfy the constraints (6.83). Hence, the cell problems (6.82) are well-posed and u^3 is well defined. Similarly, combining Lemma 6.2.8 with Proposition 6.2.9, we verify that $L^2\tilde u = \tilde L^2\tilde u$ and the definition of $\tilde L^2$ ensures that u^4 is well defined. Note that thanks to assumption (4.25), we verify that $x \mapsto \mathcal{B}^\varepsilon\tilde u(t,x)$ is Ω-periodic. This proves the existence of the adaptation $\mathcal{B}^\varepsilon\tilde u$. By construction (see (6.70)), $\mathcal{B}^\varepsilon\tilde u$ satisfies the properties (6.68) and the proof of the theorem is complete.

6.2.3 A regularity result for the correctors

In Section 6.2.2, we derived cell problems for the correctors involved in the adaptation As the adaptation is the main tool in the proof of Theorem 6.2.1 (Section 6.2.4), we need to establish the influence of the tensor $a(x,y)$ on the regularity of the correctors. In this section, we prove a sufficient condition for the correctors to belong to $\mathcal{C}^n(\bar\Omega;\mathrm{H}^{m+1}(Y))$.

Let $a \in [\mathcal{C}^0_{\mathrm{per}}(\bar\Omega; \mathrm{L}^\infty_{\mathrm{per}}(Y))]^{d\times d}$ be the tensor and $r \in \mathcal{C}^0_{\mathrm{per}}(\bar\Omega; \mathrm{W}^*_{\mathrm{per}}(Y))$ a right hand side. For all $x \in \Omega$, let $v(x) \in \mathrm{W}_{\mathrm{per}}(Y)$ be the solution of the cell problem

$$\big(a(x)\nabla_y v(x), \nabla_y w\big)_{\mathrm{L}^2(Y)} = \big\langle r(x), w\big\rangle_{\mathrm{W}^*_{\mathrm{per}}(Y),\mathrm{W}_{\mathrm{per}}(Y)} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(Y). \tag{6.104}$$

Thanks to the Lax–Milgram theorem, $v(x)$ exists and is unique for all $x \in \Omega$. The following result provides a sufficient conditino for v to belong to $\mathcal{C}^n(\bar\Omega; \mathrm{H}^{m+1}(Y))$.

Lemma 6.2.10. *If $a \in [\mathcal{C}^n(\bar\Omega; \mathrm{W}^{m,\infty}(Y))]^{d\times d}$ and $r \in \mathcal{C}^n(\bar\Omega; \mathrm{H}^{m-1}(Y))$ ($\mathrm{H}^0(Y) = \mathrm{L}^2_0(Y)$) for some integers $n \geq 0$ and $m \geq 1$, then v satisfies the regularity $v \in \mathcal{C}^n(\bar\Omega; \mathrm{H}^{m+1}(Y))$ and the following estimate holds*

$$\|v\|_{\mathcal{C}^n(\bar\Omega;\mathrm{H}^{m+1}(Y))} \leq C\|r\|_{\mathcal{C}^n(\bar\Omega;\mathrm{H}^{m-1}(Y))}, \tag{6.105}$$

where C depends only on Y, λ, m, n and $\max_{ij} \|a_{ij}\|_{\mathcal{C}^n(\bar\Omega;\mathrm{W}^{m,\infty}(Y))}$.

Proof. We prove the result by induction on $n \in \mathbb{N}$. Let us prove it for $n = 0$. As $a(x) \in \mathrm{W}^{m,\infty}(Y), r(x) \in \mathrm{H}^{m-1}(Y)$, the regularity result of Theorem A.2.2 ensures that $v(x) \in \mathrm{H}^{m+1}(Y)$ and

$$\|v(x)\|_{\mathrm{H}^{m+1}(Y)} \leq C\|r(x)\|_{\mathrm{H}^{m-1}(Y)}, \tag{6.106}$$

where the constant C depends on Y, λ, m and $\max_{ij} \|a_{ij}(x)\|_{\mathrm{W}^{m,\infty}(Y)}$. Further, from (6.104), $v(x+h) - v(x)$ solves the variational equation

$$\big(a(x)\nabla_y(v(x+h)-v(x)), \nabla_y w\big)_Y = \big(r(x+h)-r(x), w\big)_Y + \big(\nabla_y\cdot\big((a(x+h)-a(x))\nabla_y v(x+h)\big), w\big)_Y,$$

$\forall w \in \mathrm{W}_{\mathrm{per}}(Y)$ and thus satisfies

$$\begin{aligned}\|v(x+h) - v(x)\|_{\mathrm{H}^{m+1}(Y)} \leq C\big(&\|r(x+h) - r(x)\|_{\mathrm{H}^{m-1}(Y)} \\ &+ \|a(x+h) - a(x)\|_{\mathrm{W}^{m,\infty}(Y)}\|v(x+h)\|_{\mathrm{H}^{m+1}(Y)}\big).\end{aligned}$$

As we assume $a \in \mathcal{C}^0(\bar\Omega; \mathrm{W}^{m,\infty}(Y)), r \in \mathcal{C}^0(\bar\Omega; \mathrm{H}^{m-1}(Y))$, we conclude that $v \in \mathcal{C}^0(\bar\Omega; \mathrm{H}^{m+1}(Y))$. Estimate (6.105) for $n = 0$ follows from (6.106), and that proves the result for $n = 0$. Assume now that the result holds true for $n-1$ and let us prove that it remains true for n. Let $\alpha \in \mathbb{N}^d$ be a multi-index such that $|\alpha| = \sum_{i=1}^d \alpha_i = n$. For two functions $f, g \in \mathcal{C}^n(\bar\Omega)$, we write

$$\partial^\alpha(fg) = f\partial^\alpha g + R_\alpha(f,g), \qquad R_\alpha(f,g) = \sum_{\substack{|\gamma|+|\beta|=|\alpha|\\ |\beta|\geq 1}} b^\alpha_{\beta,\gamma}\partial^\beta f\partial^\gamma g,$$

where $b^\alpha_{\beta,\gamma}$ are coefficients in $\mathbb{R}$. Differentiating (6.104) with respect to α, we find that $\partial^\alpha v(x) \in \mathrm{W}_{\mathrm{per}}(Y)$ is the solution of the variational problem

$$\big(a(x)\nabla_y\partial^\alpha v(x), \nabla_y w\big)_Y = \big(\partial^\alpha r(x), w\big)_Y + \sum_{ij}\big(\partial_{y_i} R_\alpha\big(a_{ij}(x), \partial_{y_j} v(x)\big), w\big)_Y$$

for all $w \in \mathrm{W}_{\mathrm{per}}(Y)$. Thus it satisfies

$$\|\partial^\alpha v(x)\|_{\mathrm{H}^{m+1}(Y)} \leq C\big(\|r\|_{\mathcal{C}^n(\mathrm{H}^{m-1})} + \|a\|_{\mathcal{C}^n(\mathrm{W}^{m,\infty})}\|v\|_{\mathcal{C}^{n-1}(\mathrm{H}^{m+1})}\big). \tag{6.107}$$

Furthermore, $\partial^\alpha v(x+h) - \partial^\alpha v(x)$ solves

$$\begin{aligned}\big(a(x)\nabla_y\big(\partial^\alpha v(x+h) - \partial^\alpha v(x)\big), \nabla_y w\big)_Y = &\big(\partial^\alpha r(x+h) - \partial^\alpha r(x), w\big)_Y \\ &+ \sum_{ij}\big(\partial_{y_i} R_\alpha\big(a_{ij}(x+h) - a_{ij}(x), \partial_{y_j} v(x)\big), w\big)_Y \\ &+ \sum_{ij}\big(\partial_{y_i} R_\alpha\big(a_{ij}(x+h), \partial_{y_j} v(x+h) - \partial_{y_j} v(x)\big), w\big)_Y \\ &+ \sum_{ij}\big(\partial_{y_i}\big(a_{ij}(x+h) - a_{ij}(x), \partial_{y_j}\partial^\alpha v(x)\big), w\big)_Y\end{aligned}$$

for all $w \in \mathrm{W}_{\mathrm{per}}(Y)$, and we thus have

$$\begin{aligned}\|\partial^\alpha v(x+h)-\partial^\alpha v(x)\|_{\mathrm{H}^{m+1}(Y)} \leq C\Big(& \|\partial^\alpha r(x+h)-\partial^\alpha r(x)\|_{\mathrm{H}^{m-1}(Y)} \\ & + \Big(\sum_{|\beta|\leq n} \max_{ij} \|\partial^\beta a_{ij}(x+h)-\partial^\beta a_{ij}(x)\|_{\mathrm{W}^{m,\infty}(Y)}\Big) \sum_{|\gamma|\leq n-1} \|\partial^\gamma v(x)\|_{\mathrm{H}^{m+1}(Y)} \\ & + \Big(\sum_{|\beta|\leq n} \max_{ij} \|\partial^\beta a_{ij}(x)\|_{\mathrm{W}^{m,\infty}(Y)}\Big) \sum_{|\gamma|\leq n-1} \|\partial^\gamma v(x+h)-\partial^\gamma v(x)\|_{\mathrm{H}^{m+1}(Y)} \\ & + \max_{ij} \|a_{ij}(x+h)-a_{ij}(x)\|_{\mathrm{W}^{m,\infty}(Y)} \|v\|_{\mathcal{C}^n(\mathrm{H}^{m+1})}\Big).\end{aligned}$$

As we assume $a \in \mathcal{C}^n(\bar\Omega;\mathrm{W}^{m,\infty}(Y)), r \in \mathcal{C}^n(\bar\Omega;\mathrm{H}^{m-1}(Y))$, using (6.107) and the induction hypothesis $v \in \mathcal{C}^{n-1}(\bar\Omega;\mathrm{H}^{m+1}(Y))$, we conclude that $v \in \mathcal{C}^n(\bar\Omega;\mathrm{H}^{m+1}(Y))$. Finally, estimate (6.105) follows from (6.107) and the proof of the lemma is complete. □

6.2.4 Proof of the error estimate (Theorem 6.2.1)

In this section, we prove the main result of the chapter, Theorem 6.2.1. The proof is structured as follows. First, using the correctors derived in Section 6.2, we define the adaptation operator $\boldsymbol{\mathcal{B}}^\varepsilon$. In particular, recall that the definition of the effective tensors ensures the well-posedness of the cell problems. We then split the error as

$$\|u^\varepsilon - \tilde u\|_{\mathrm{L}^\infty(W)} = \|\,[u^\varepsilon - \tilde u]\,\|_{\mathrm{L}^\infty(\mathcal{W})} \leq \|\boldsymbol{\mathcal{B}}^\varepsilon \tilde u - \,[u^\varepsilon]\,\|_{\mathrm{L}^\infty(\mathcal{W})} + \|\,[\tilde u] - \boldsymbol{\mathcal{B}}^\varepsilon \tilde u\|_{\mathrm{L}^\infty(\mathcal{W})},$$

and both terms are estimated separately. In particular, we prove that $\boldsymbol{\mathcal{B}}^\varepsilon \tilde u$ satisfies the same equation as u^ε up to a remainder of order $\mathcal{O}(\varepsilon^3)$ (Lemma 6.2.11).

Let us first introduced the correctors involved in the definition the adaptation operator. Consider the correctors

$$\chi_i(x), \theta^0_{ij}(x), \theta^1_i(x), \kappa^0_{ijk}(x), \kappa^1_{ij}(x), \kappa^2_i(x), \rho^0_{ijkl}(x), \rho^1_{ijk}(x), \rho^2_{ij}(x), \rho^3_i(x) \in \mathrm{W}_{\mathrm{per}}(Y),$$

defined in the cell problems (6.72), (6.75), (6.82) and (6.90), and let $\boldsymbol{\varphi}$ be the solution of (6.102). Propositions 6.2.7 and 6.2.9 ensure that $L^1\tilde u = \tilde L^1 \tilde u$ and $L^2\tilde u = \tilde L^2\tilde u$, where the definitions of the tensors in $\tilde L^1$ (resp. $\tilde L^2$) guarantee the well-posedness of the cell problems (6.82) (resp. (6.90)). Let us investigate the regularity of the correctors. Using Lemma 6.2.10, we can show the following implications, for $n \geq 0$, $m \geq 0$:

$$\begin{aligned}
\chi_i, \theta^0_{ij}, \kappa^0_{ijk}, \rho^0_{ijkl} \in \mathcal{C}^n(\bar\Omega;\mathrm{H}^{m+1}(Y)) &\quad\Leftarrow\quad a \in \mathcal{C}^n(\bar\Omega;\mathrm{W}^{m,\infty}(Y)),\\
\theta^1_i, \kappa^1_{ij}, \rho^1_{ijk} \in \mathcal{C}^n(\bar\Omega;\mathrm{H}^{m+1}(Y)) &\quad\Leftarrow\quad a \in \mathcal{C}^n(\bar\Omega;\mathrm{W}^{m,\infty}(Y)) \cap \mathcal{C}^{n+1}(\bar\Omega;\mathrm{W}^{\{m-1\}_+,\infty}(Y)),\\
\kappa^2_i, \rho^2_{ij} \in \mathcal{C}^n(\bar\Omega;\mathrm{H}^{m+1}(Y)) &\quad\Leftarrow\quad a \in \cap_{k=0}^2 \mathcal{C}^{n+k}(\bar\Omega;\mathrm{W}^{\{m-k\}_+,\infty}(Y)),\\
\rho^3_i \in \mathcal{C}^n(\bar\Omega;\mathrm{H}^{m+1}(Y)) &\quad\Leftarrow\quad a \in \cap_{k=0}^3 \mathcal{C}^{n+k}(\bar\Omega;\mathrm{W}^{\{m-k\}_+,\infty}(Y)),\\
a^0_{ij}, \bar a^{24}_{ijkl}, b^{22}_{ij} \in \mathcal{C}^n(\bar\Omega) &\quad\Leftarrow\quad a \in \mathcal{C}^n(\bar\Omega;\mathrm{L}^\infty(Y)),\\
\bar a^{12}_{ij} \in \mathcal{C}^n(\bar\Omega) &\quad\Leftarrow\quad a \in \mathcal{C}^{n+1}(\bar\Omega;\mathrm{L}^\infty(Y)),\\
\bar a^{22}_{ij} \in \mathcal{C}^n(\bar\Omega) &\quad\Leftarrow\quad a \in \mathcal{C}^{n+2}(\bar\Omega;\mathrm{L}^\infty(Y)),
\end{aligned} \tag{6.108}$$

where $\{\cdot\}_+ = \max\{0,\cdot\}$. In particular, under the assumption of Theorem 6.2.1:

$$a \in \mathcal{C}^1(\bar\Omega;\mathrm{W}^{2,\infty}(Y)) \cap \mathcal{C}^2(\bar\Omega;\mathrm{W}^{1,\infty}(Y)) \cap \mathcal{C}^4(\bar\Omega;\mathrm{L}^\infty(Y)),$$

all the correctors belongs to $\mathcal{C}^1(\bar\Omega;\mathrm{H}^3(Y))\cap\mathcal{C}^2(\bar\Omega;\mathrm{H}^2(Y))$. As $d\le 3$, the Sobolev embedding $\mathrm{H}^2_{\mathrm{per}}(Y)\hookrightarrow\mathcal{C}^0_{\mathrm{per}}(\bar Y)$ holds and the correctors belongs to $\mathcal{C}^1(\bar\Omega;\mathcal{C}^1_{\mathrm{per}}(\bar Y))\cap\mathcal{C}^2(\bar\Omega;\mathcal{C}^0_{\mathrm{per}}(\bar Y))$. Hence, the following estimates (needed in the proof of Lemma 6.2.11 below) hold

$$\begin{aligned}\max_{ijkl}\Big\{&\|\chi_i\|_{\mathcal{C}^0(\mathcal{C}^0)},\|\theta^0_{ij}\|_{\mathcal{C}^1(\mathcal{C}^1)},\|\theta^1_i\|_{\mathcal{C}^0(\mathcal{C}^0)},\|\kappa^0_{ijk}\|_{\mathcal{C}^2(\mathcal{C}^0)},\|\kappa^1_{ij}\|_{\mathcal{C}^2(\mathcal{C}^0)},\\&\|\kappa^2_i\|_{\mathcal{C}^2(\mathcal{C}^0)},\|\rho^0_{ijkl}\|_{\mathcal{C}^1(\mathcal{C}^1)},\|\rho^1_{ijk}\|_{\mathcal{C}^1(\mathcal{C}^1)},\|\rho^2_{ij}\|_{\mathcal{C}^1(\mathcal{C}^1)},\|\rho^3_i\|_{\mathcal{C}^1(\mathcal{C}^1)},\\&\|\bar a^{12}_{ij}\|_{\mathcal{C}^2},|b^{10}|,\|\bar a^{24}_{ijkl}\|_{\mathcal{C}^3},\|b^{22}_{ij}\|_{\mathcal{C}^2},\|\bar a^{22}_{ij}\|_{\mathcal{C}^2},|b^{20}|\Big\}\\&\qquad\le C_1(a,\lambda,Y)+\delta C_2(a,\lambda,Y),\end{aligned}\tag{6.109}$$

where $C_i(a,\lambda,Y)$ depend only on λ, Y, $\|a\|_{\mathcal{C}^1(\mathrm{W}^{2,\infty})}$, $\|a\|_{\mathcal{C}^2(\mathrm{W}^{1,\infty})}$, and $\|a\|_{\mathcal{C}^4(\mathrm{L}^\infty)}$, and δ is the parameter.

Let us introduce the following useful application of the Green formula (see Remark 4.2.7 for a proof): for $c\in[\mathrm{W}^{1,\infty}_{\mathrm{per}}(\Omega)]^d$, $v\in\mathrm{H}^1_{\mathrm{per}}(\Omega)$, and $\boldsymbol{w}\in\mathcal{W}_{\mathrm{per}}(\Omega)$, we have

$$\big(\big[c_m\partial_m v\big],\boldsymbol{w}\big)_{\mathcal{L}^2}=-\big(\big[\partial_m c_m v\big],\boldsymbol{w}\big)_{\mathcal{L}^2}-\big(c_m,\partial_m\boldsymbol{w}\big)_{\mathrm{L}^2},\tag{6.110}$$

where we recall that $\partial_m c_m=\sum_{m=1}^d\partial_m c_m$. In order to shorten the notation, we define the following functions of $\mathcal{C}^1_{\mathrm{per}}(\bar\Omega)$: $\chi^\varepsilon_i=\chi_i\big(\cdot,\frac{\cdot}{\varepsilon}\big)$, $\theta^{0\varepsilon}_{ij}=\theta^0_{ij}\big(\cdot,\frac{\cdot}{\varepsilon}\big)$, $\theta^{1\varepsilon}_i=\theta^1_i\big(\cdot,\frac{\cdot}{\varepsilon}\big)$, and similarly $\kappa^{0\varepsilon}_{ijk},\kappa^{1\varepsilon}_{ij},\kappa^{2\varepsilon}_i,\rho^{0\varepsilon}_{ijkl},\rho^{1\varepsilon}_{ijk},\rho^{2\varepsilon}_{ij},\rho^{3\varepsilon}_i$. We define then the operators $\boldsymbol{\mathcal{B}}^\varepsilon_i:\mathrm{H}^3_{\mathrm{per}}(\Omega)\to\mathcal{W}^*_{\mathrm{per}}(\Omega)$ for $v\in\mathrm{H}^3_{\mathrm{per}}(\Omega)$ as

$$\begin{aligned}\big\langle\boldsymbol{\mathcal{B}}^\varepsilon_0 v,\boldsymbol{w}\big\rangle&=\big(\big[v\big],\boldsymbol{w}\big)_{\mathcal{L}^2},\\
\big\langle\boldsymbol{\mathcal{B}}^\varepsilon_1 v,\boldsymbol{w}\big\rangle&=\big(\varepsilon\big[\chi^\varepsilon_i\partial_i v\big],\boldsymbol{w}\big)_{\mathcal{L}^2},\\
\big\langle\boldsymbol{\mathcal{B}}^\varepsilon_2 v,\boldsymbol{w}\big\rangle&=\big(\varepsilon^2\big[(-\partial_m\theta^{0\varepsilon}_{mi}+\theta^{1\varepsilon}_i)\partial_i v\big],\boldsymbol{w}\big)_{\mathcal{L}^2}-\big(\varepsilon^2\theta^{0\varepsilon}_{mi}\partial_i v,\partial_m\boldsymbol{w}\big)_{\mathrm{L}^2},\\
\big\langle\boldsymbol{\mathcal{B}}^\varepsilon_3 v,\boldsymbol{w}\big\rangle&=\big(\varepsilon^3\big[\kappa^{0\varepsilon}_{ijk}\partial^3_{ijk}v+\kappa^{1\varepsilon}_{ij}\partial^2_{ij}v+\kappa^{2\varepsilon}_i\partial_i v\big],\boldsymbol{w}\big)_{\mathcal{L}^2},\\
\big\langle\boldsymbol{\mathcal{B}}^\varepsilon_4 v,\boldsymbol{w}\big\rangle&=\big(\varepsilon^4\big[(-\partial_m\rho^{0\varepsilon}_{mijk}+\rho^{1\varepsilon}_{ijk})\partial^3_{ijk}v+\rho^{2\varepsilon}_{ij}\partial^2_{ij}v+\rho^{3\varepsilon}_i\partial_i v\big],\boldsymbol{w}\big)_{\mathcal{L}^2}-\big(\varepsilon^4\rho^{0\varepsilon}_{mijk}\partial^3_{ijk}v,\partial_m\boldsymbol{w}\big)_{\mathrm{L}^2},\end{aligned}$$

where $\langle\cdot,\cdot\rangle$ denotes the dual evaluation $\langle\cdot,\cdot\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}}$. The adaptation operator is then defined as

$$\boldsymbol{\mathcal{B}}^\varepsilon:\mathrm{L}^2(0,T^\varepsilon;\mathrm{H}^3_{\mathrm{per}}(\Omega))\to\mathrm{L}^2(0,T^\varepsilon;\mathcal{W}^*_{\mathrm{per}}(\Omega)),\quad v\mapsto\boldsymbol{\mathcal{B}}^\varepsilon v(t)=\sum_{i=0}^4\boldsymbol{\mathcal{B}}^\varepsilon_i(v(t))+\boldsymbol{\varphi}(t).\tag{6.111}$$

Note that if $v\in\mathrm{L}^2(0,T^\varepsilon;\mathrm{H}^1_{\mathrm{per}}(\Omega)\cap\mathrm{H}^5(\Omega))$, then $\boldsymbol{\mathcal{B}}^\varepsilon v(t)\in\mathcal{W}_{\mathrm{per}}(\Omega)$ and, using (6.110), we verify that $\boldsymbol{\mathcal{B}}^\varepsilon\tilde u(t)=\big[\mathcal{B}^\varepsilon\tilde u(t)\big]$, where $\mathcal{B}^\varepsilon\tilde u$ is defined in (6.67) (with $\{u^k\}_{k=1}^4$ defined in (6.71), (6.74), (6.78), and (6.86)). For $\mathcal{A}^\varepsilon=-\nabla_x\cdot\big(a^\varepsilon(x)\nabla_x\cdot\big)$, we thus define

$$\big\langle\mathcal{A}^\varepsilon\boldsymbol{\mathcal{B}}^\varepsilon\tilde u(t),\boldsymbol{w}\big\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}}=\big\langle\mathcal{A}^\varepsilon\big[\mathcal{B}^\varepsilon v(t)\big],\boldsymbol{w}\big\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}}.$$

Remark that the definition of $\boldsymbol{\mathcal{B}}^\varepsilon$ in (6.111) allows to consider functions with lower regularity than $\mathcal{B}^\varepsilon$. In particular, as $\partial_t^2\tilde u\in\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^3_{\mathrm{per}}(\Omega))$, $\boldsymbol{\mathcal{B}}^\varepsilon(\partial_t^2\tilde u)$ is well-defined, . This is needed in the proof of the following lemma, where we prove that $\boldsymbol{\mathcal{B}}^\varepsilon\tilde u$ solves the same equation as $\big[u^\varepsilon\big]$ with a remainder of order ε^3.

Lemma 6.2.11. *Under the assumptions of Theorem 6.2.1, $\boldsymbol{\mathcal{B}}^\varepsilon\tilde u$ satisfies*

$$(\partial_t^2+\mathcal{A}^\varepsilon)\boldsymbol{\mathcal{B}}^\varepsilon\tilde u(t)=\big[f(t)\big]+\boldsymbol{\mathcal{R}}^\varepsilon\tilde u(t)\quad\text{in }\mathcal{W}^*_{\mathrm{per}}(\Omega)\quad\text{for a.e. }t\in[0,T^\varepsilon],$$

where the remainder $\boldsymbol{\mathcal{R}}^\varepsilon\tilde u\in\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{W}^*_{\mathrm{per}}(\Omega))$ is given as

$$\big\langle\boldsymbol{\mathcal{R}}^\varepsilon\tilde u(t),\boldsymbol{w}\big\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}}=\big((\boldsymbol{\mathcal{R}}^\varepsilon\tilde u)_0(t),\boldsymbol{w}\big)_{\mathcal{L}^2}+\big((\boldsymbol{\mathcal{R}}^\varepsilon\tilde u)_1(t),\nabla_x\boldsymbol{w}\big)_{\mathrm{L}^2},$$

with the bound

$$\|(\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u})_0\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{L}^2(\Omega))}+\|(\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u})_1\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} \\ \leq C\varepsilon^3\Big(\sum_{k=1}^5|\tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^k(\Omega))}+\|\partial_t^2\tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^3(\Omega))}\Big),$$

for a constant C that depends only on λ, Y, $\|a\|_{\mathcal{C}^1(\bar{\Omega};\mathrm{W}^{2,\infty}(Y))}$, $\|a\|_{\mathcal{C}^2(\bar{\Omega};\mathrm{W}^{1,\infty}(Y))}$, $\|a\|_{\mathcal{C}^4(\bar{\Omega};\mathrm{L}^\infty(Y))}$, and δ.

Proof. Let us denote $\langle\cdot,\cdot\rangle_{\mathcal{W}_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}}$ as $\langle\cdot,\cdot\rangle$. For a fixed $t\in[0,T^\varepsilon]$, we compute the remainder $\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u}(t)=(\partial_t^2+\mathcal{A}^\varepsilon)\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}(t)-\big[f(t)\big]$. Let us first compute explicitly the first term, $\partial_t^2\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}(t)$. For the sake of clarity, we drop the notation of the evaluation in t. From the definition of $\boldsymbol{\mathcal{B}}^\varepsilon$ in (6.111), it holds $\partial_t^2\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}=\sum_{i=0}^2\boldsymbol{\mathcal{B}}_i^\varepsilon\partial_t^2\tilde{u}+\partial_t^2\boldsymbol{\varphi}+\boldsymbol{\mathcal{R}}_1^\varepsilon\tilde{u}$, where $\boldsymbol{\mathcal{R}}_1^\varepsilon\tilde{u}=\sum_{i=3}^4\boldsymbol{\mathcal{B}}_i^\varepsilon\partial_t^2\tilde{u}$, i.e.,

$$\big\langle\partial_t^2\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u},\boldsymbol{w}\big\rangle=\big(\big[\partial_t^2\tilde{u}\big],\boldsymbol{w}\big)_{\mathcal{L}^2}+\big(\big[\varepsilon\chi_i^\varepsilon\partial_i\partial_t^2\tilde{u}+\varepsilon^2(-\partial_m\theta_{mi}^{0\varepsilon}+\theta_i^{1\varepsilon})\partial_i\partial_t^2\tilde{u}\big],\boldsymbol{w}\big)_{\mathcal{L}^2} \\ -\big(\varepsilon^2\theta_{mi}^{0\varepsilon}\partial_i\partial_t^2\tilde{u},\partial_m\boldsymbol{w}\big)+\big\langle\partial_t^2\boldsymbol{\varphi}+\boldsymbol{\mathcal{R}}_1^\varepsilon\tilde{u},\boldsymbol{w}\big\rangle. \tag{6.112}$$

We rewrite the three first terms of the right hand side. Note that thanks to the regularity of $\tilde{u}$ and the effective equation (6.65), we have the following equalities

$$\partial_t^2\tilde{u}=f+\partial_m(a_{mn}^0\partial_n\tilde{u})-\varepsilon L^1\tilde{u}-\varepsilon^2L^2\tilde{u}\quad\text{in }\mathrm{L}_0^2(\Omega), \tag{6.113}$$

$$\partial_i\partial_t^2\tilde{u}=\partial_i f+\partial_{im}^2(a_{mn}^0\partial_n\tilde{u})-\varepsilon\partial_i(L^1\tilde{u})-\varepsilon^2\partial_i(L^2\tilde{u})\quad\text{in }\mathrm{L}^2(\Omega). \tag{6.114}$$

Using (6.113), we rewrite the first term of (6.112) as

$$\big[\partial_t^2\tilde{u}\big]=\big[f\big]+\big[\partial_m(a_{mn}^0\partial_n\tilde{u})+\varepsilon\big(-L^{1,x}\tilde{u}-b^{10}\partial_m(a_{mn}^0\partial_n\tilde{u})\big)+\varepsilon^2\big(-L^2\tilde{u}+b^{10}L^1\tilde{u}\big)\big] \\ +\big[-\varepsilon b^{10}f+\varepsilon^3b^{10}L^2\tilde{u}\big].$$

Using the definitions of L^1 and L^2 and (6.113), we have

$$\varepsilon^2\big(\big[-L^2\tilde{u}+b^{10}L^1\tilde{u}\big],\boldsymbol{w}\big)_{\mathcal{L}^2} \\ =\varepsilon^2\big(\big[-L^{2,x}\tilde{u}-b^{10}L^{1,x}\tilde{u}+((b^{10})^2-b^{20})\partial_m(a_{mn}^0\partial_n\tilde{u})\big],\boldsymbol{w}\big)_{\mathcal{L}^2} \\ -\varepsilon^2\big(b_{mi}^{22}\partial_i\partial_t^2\tilde{u},\partial_m\boldsymbol{w}\big)_{\mathrm{L}^2}+\varepsilon^2\big(\big[((b^{10})^2-b^{20})(f+\varepsilon L^1\tilde{u}+\varepsilon^2L^2\tilde{u})\big],\boldsymbol{w}\big)_{\mathcal{L}^2},$$

where $L^{1,x}=L^1-b^{10}\partial_t^2$ and $L^{2,x}=L^2+\partial_i(b_{ij}^{22}\partial_j\partial_t^2\cdot)-b^{20}\partial_t^2$ are the spatial parts of the operators. We thus obtain

$$\big(\big[\partial_t^2\tilde{u}\big],\boldsymbol{w}\big)_{\mathcal{L}^2}=\big(\big[f\big]+\big[\partial_m(a_{mn}^0\partial_n\tilde{u})+\varepsilon\big(-L^{1,x}\tilde{u}-b^{10}\partial_m(a_{mn}^0\partial_n\tilde{u})\big) \\ +\varepsilon^2\big(-L^{2,x}\tilde{u}+b^{10}L^{1,x}\tilde{u}+((b^{10})^2-b^{20})\partial_m(a_{mn}^0\partial_n\tilde{u})\big],\boldsymbol{w}\big)_{\mathcal{L}^2} \\ -\big(\varepsilon^2b_{mi}^{22}\partial_i\partial_t^2\tilde{u},\partial_m\boldsymbol{w}\big)_{\mathrm{L}^2}+\big(\boldsymbol{\mathcal{S}}_1^\varepsilon f+\boldsymbol{\mathcal{R}}_2^\varepsilon\tilde{u},\boldsymbol{w}\big)_{\mathcal{L}^2}, \tag{6.115}$$

where

$$\boldsymbol{\mathcal{S}}_1^\varepsilon f=\big[-\varepsilon b^{10}f+\varepsilon^2((b^{10})^2-b^{20})f\big], \\ \boldsymbol{\mathcal{R}}_2^\varepsilon\tilde{u}=\big[\varepsilon^3b^{10}L^2\tilde{u}+\varepsilon^3((b^{10})^2-b^{20})(L^1\tilde{u}+\varepsilon L^2\tilde{u})\big].$$

Next, we use (6.114) and then (6.113) to write the second term of (6.112) as

$$\big[\varepsilon\chi_i^\varepsilon\partial_i\partial_t^2\tilde{u}\big]=\big[\varepsilon\chi_i^\varepsilon\partial_{im}^2(a_{mn}^0\partial_n\tilde{u})-\varepsilon^2\chi_i^\varepsilon\partial_i(L^{1,x}\tilde{u})-\varepsilon^2\chi_i^\varepsilon\partial_i(b^{10}\partial_m(a_{mn}^0\partial_n\tilde{u}))\big]+\boldsymbol{\mathcal{S}}_2^\varepsilon f+\boldsymbol{\mathcal{R}}_3^\varepsilon\tilde{u}, \tag{6.116}$$

where

$$\boldsymbol{\mathcal{S}}_2^\varepsilon f = \big[\varepsilon\chi_i^\varepsilon\partial_i f - \varepsilon^2\chi_i^\varepsilon\partial_i(b^{10}f)\big],$$
$$\boldsymbol{\mathcal{R}}_3^\varepsilon\tilde{u} = \big[-\varepsilon^3\chi_i^\varepsilon\partial_i(L^2\tilde{u}) + \varepsilon^3\chi_i^\varepsilon\partial_i(b^{10}(L^1\tilde{u}+\varepsilon L^2\tilde{u}))\big].$$

Furthermore, using (6.114) and formula (6.110), we rewrite

$$\begin{aligned}&-\big(\big[\varepsilon^2\partial_m\theta_{mi}^{0\varepsilon}\partial_i\partial_t^2\tilde{u}\big],\boldsymbol{w}\big)_{\mathcal{L}^2} - \big(\varepsilon^2(\theta_{mi}^{0\varepsilon}+b_{mi}^{22})\partial_i\partial_t^2\tilde{u},\partial_m\boldsymbol{w}\big)_{\mathrm{L}^2}\\ &\quad=\big(\big[\varepsilon^2\theta_{ij}^{0\varepsilon}\partial_{ijm}^3(a_{mn}^0\partial_n\tilde{u}) + \partial_i(b_{ij}^{22}\partial_{jm}^2(a_{mn}^0\partial_n\tilde{u}))\big],\boldsymbol{w}\big)_{\mathcal{L}^2} + \big\langle\boldsymbol{\mathcal{S}}_3^\varepsilon f + \boldsymbol{\mathcal{R}}_4^\varepsilon\tilde{u},\boldsymbol{w}\big\rangle. \qquad (6.117)\end{aligned}$$

where

$$\big\langle\boldsymbol{\mathcal{S}}_3^\varepsilon f,\boldsymbol{w}\big\rangle = \big(\varepsilon^2\big[\theta_{ij}^{0\varepsilon}\partial_{ij}^2 f + \partial_i(b_{ij}^{22}\partial_j f)\big],\boldsymbol{w}\big)_{\mathcal{L}^2},$$
$$\big\langle\boldsymbol{\mathcal{R}}_4^\varepsilon\tilde{u},\boldsymbol{w}\big\rangle = \big(\varepsilon^3\big[\partial_m\theta_{mi}^{0\varepsilon}\partial_i(L^1\tilde{u}+\varepsilon L^2\tilde{u})\big],\boldsymbol{w}\big)_{\mathcal{L}^2} + \big(\varepsilon^3(\theta_{mi}^{0\varepsilon}+b_{mi}^{22})\partial_i(L^1\tilde{u}+\varepsilon L^2\tilde{u}),\partial_m\boldsymbol{w}\big)_{\mathrm{L}^2},$$

and, using (6.114), we rewrite

$$\big[\varepsilon^2\theta_i^{1\varepsilon}\partial_i\partial_t^2\tilde{u}\big] = \big[\varepsilon^2\theta_i^{1\varepsilon}\partial_{im}^2(a_{mn}^0\partial_n\tilde{u})\big] + \boldsymbol{\mathcal{S}}_4^\varepsilon f + \boldsymbol{\mathcal{R}}_5^\varepsilon\tilde{u}, \qquad (6.118)$$

where $\boldsymbol{\mathcal{S}}_4^\varepsilon f = \big[\varepsilon^2\theta_i^{1\varepsilon}\partial_i f\big]$ and $\boldsymbol{\mathcal{R}}_5^\varepsilon\tilde{u} = \big[\varepsilon^3\theta_i^{1\varepsilon}\partial_i(L^1\tilde{u}+\varepsilon L^2\tilde{u})\big]$. Combining equalities (6.112), (6.115), (6.116), (6.117) and (6.118), we finally obtain

$$\begin{aligned}\partial_t^2\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u} = \big[f\big] &+ \big[\partial_m(a_{mn}^0\partial_n\tilde{u})\big] + \varepsilon\big[\chi_i\partial_{im}^2(a_{mn}^0\partial_n\tilde{u}) - L^{1,x}\tilde{u} - b^{10}\partial_m(a_{mn}^0\partial_n\tilde{u})\big]\\ &+\varepsilon^2\big[\theta_{ij}^{0\varepsilon}\partial_{ijm}^3(a_{mn}^0\partial_n\tilde{u}) + \theta_i^{1\varepsilon}\partial_{im}^2(a_{mn}^0\partial_n\tilde{u}) - \chi_i^\varepsilon\partial_i(L^{1,x}\tilde{u}) - \chi_i^\varepsilon\partial_i(b^{10}\partial_m(a_{mn}^0\partial_n\tilde{u}))\\ &\qquad - L^{2,x}\tilde{u} + b^{10}L^{1,x}\tilde{u} + ((b^{10})^2 - b^{20})\partial_m(a_{mn}^0\partial_n\tilde{u}) + \partial_i(b_{ij}^{22}\partial_{jm}^2(a_{mn}^0\partial_n\tilde{u}))\big]\\ &+\partial_t^2\boldsymbol{\varphi} + \textstyle\sum_{i=1}^4\boldsymbol{\mathcal{S}}_i^\varepsilon f + \sum_{i=1}^5\boldsymbol{\mathcal{R}}_i^\varepsilon\tilde{u}. \qquad (6.119)\end{aligned}$$

For the second term, $\mathcal{A}^\varepsilon\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}(t)$, we have (the correctors and a are evaluated at $\big(x, y=\frac{x}{\varepsilon}\big)$)

$$\begin{aligned}&\mathcal{A}^\varepsilon\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u} =\\ &\big[\quad\varepsilon^{-1}\big(-\nabla_y\cdot(a(\nabla_y\chi_i+e_i))\big)\partial_i\tilde{u}\\ &+\varepsilon^0\ \big(-\nabla_y\cdot(a(\nabla_y\theta_{ij}^0+e_i\chi_j)) - e_i^Ta(\nabla_y\chi_j+e_j)\big)\partial_{ij}^2\tilde{u}\\ &+\varepsilon^0\ \big(-\nabla_y\cdot(a(\nabla_y\theta_i^1+\nabla_x\chi_i)) - \nabla_x\cdot(a(\nabla_y\chi_i+e_i))\big)\partial_i\tilde{u}\\ &+\varepsilon^1\ \big(-\nabla_y\cdot(a(\nabla_y\kappa_{ijk}^0+e_i\theta_{jk}^0)) - e_i^Ta(\nabla_y\theta_{jk}^0+e_j\chi_k)\big)\partial_{ijk}^3\tilde{u}\\ &+\varepsilon^1\ \big(-\nabla_y\cdot(a(\nabla_y\kappa_{ij}^1+\nabla_x\theta_{ij}^0+e_i\theta_j^1)) - \nabla_x\cdot(a(\nabla_y\theta_{ij}^0+e_i\chi_j)) - e_i^Ta(\nabla_y\theta_j^1+\nabla_x\chi_j)\big)\partial_{ij}^2\tilde{u}\\ &+\varepsilon^1\ \big(-\nabla_y\cdot(a(\nabla_y\kappa_i^2+\nabla_x\theta_i^1)) - \nabla_x\cdot(a(\nabla_y\theta_i^1+\nabla_x\chi_i))\big)\partial_i\tilde{u}\\ &+\varepsilon^2\ \big(-\nabla_y\cdot(a(\nabla_y\rho_{ijkl}^0+e_i\kappa_{jkl}^0)) - e_i^Ta(\nabla_y\kappa_{jkl}^0+e_j\theta_{kl}^0)\big)\partial_{ijkl}^4\tilde{u}\\ &+\varepsilon^2\ \big(-\nabla_y\cdot(a(\nabla_y\rho_{ijk}^1+\nabla_x\kappa_{ijk}^0+e_i\kappa_{jk}^1)) - \nabla_x\cdot(a(\nabla_y\kappa_{ijk}^0+e_i\theta_{jk}^0))\\ &\qquad\qquad - e_i^Ta(\nabla_y\kappa_{jk}^1+\nabla_x\theta_{jk}^0)\big)\partial_{ijk}^3\tilde{u}\\ &+\varepsilon^2\ \big(-\nabla_y\cdot(a(\nabla_y\rho_{ij}^2+\nabla_x\kappa_{ij}^1+e_i\kappa_j^2)) - \nabla_x\cdot(a(\nabla_y\kappa_{ij}^1+\nabla_x\theta_{ij}^0+e_i\theta_j^1))\\ &\qquad\qquad - e_i^Ta(\nabla_y\kappa_j^2+\nabla_x\theta_j^1)\big)\partial_{ij}^2\tilde{u}\\ &+\varepsilon^2\ \big(-\nabla_y\cdot(a(\nabla_y\rho_i^3+\nabla_x\kappa_i^2)) - \nabla_x(a(\nabla_y\kappa_i^2+\nabla_x\theta_i^1))\big)\partial_i\tilde{u}\qquad\big]\\ &+\mathcal{A}^\varepsilon\boldsymbol{\varphi}+\boldsymbol{\mathcal{R}}_6^\varepsilon\tilde{u}+\boldsymbol{\mathcal{R}}_7^\varepsilon\tilde{u}, \qquad (6.120)\end{aligned}$$

where, defining the following functions of (x,y),

$$R_{ijkl}^0 = a(\nabla_y\rho_{ijkl}^0+e_i\kappa_{jkl}^0), \qquad R_{ijk}^1 = a(\nabla_y\rho_{ijk}^1+\nabla_x\kappa_{ijk}^0+e_i\kappa_{jk}^1),$$
$$R_{ij}^2 = a(\nabla_y\rho_{ij}^2+\nabla_x\kappa_{ij}^1+e_i\kappa_j^2), \qquad R_i^3 = a(\nabla_y\rho_i^3+\nabla_x\kappa_i^2),$$

the remainders $\boldsymbol{\mathcal{R}}_6^\varepsilon\tilde{u}$ and $\boldsymbol{\mathcal{R}}_7^\varepsilon\tilde{u}$ are given by

$$\begin{aligned}\boldsymbol{\mathcal{R}}_6^\varepsilon\tilde{u} &= \varepsilon^3\big[e_m^T R^0_{ijkl}\partial^5_{mijkl}\tilde{u} + \nabla_x\cdot R^0_{ijkl}\partial^4_{ijkl}\tilde{u} + e_m^T R^1_{ijk}\partial^4_{mijk}\tilde{u} + \nabla_x\cdot R^1_{ijk}\partial^3_{ijk}\tilde{u}\\ &\qquad + e_m^T R^2_{ij}\partial^3_{mij}\tilde{u} + \nabla_x\cdot R^2_{ij}\partial^2_{ij}\tilde{u} + e_m^T R^3_i\partial^2_{mi}\tilde{u} + \nabla_x\cdot R^3_i\partial_i\tilde{u}\big],\\ \big\langle\boldsymbol{\mathcal{R}}_7^\varepsilon\tilde{u},\boldsymbol{w}\big\rangle &= \varepsilon^4\big(e_m^T a(\nabla_x\rho^0_{ijkl}\partial^4_{ijkl}\tilde{u} + \nabla_x\rho^1_{ijk}\partial^3_{ijk}\tilde{u} + \nabla_x\rho^2_{ij}\partial^2_{ij}\tilde{u} + \nabla_x\rho^3_i\partial_i\tilde{u},\partial_m\boldsymbol{w}\big)_{\mathrm{L}^2}\\ &\qquad + \varepsilon^4\big(a_{mn}\rho^0_{ijkl}\partial^5_{nijkl}\tilde{u} + a_{mn}\rho^1_{ijk}\partial^4_{nijk}\tilde{u} + a_{mn}\rho^2_{ij}\partial^3_{nij}\tilde{u} + a_{mn}\rho^3_i\partial^2_{ni}\tilde{u},\partial_m\boldsymbol{w}\big)_{\mathrm{L}^2}.\end{aligned}$$

Combining now (6.119) and (6.120), and using cell problems (6.72), (6.75), (6.82), (6.90), and the definition of $\boldsymbol{\varphi}$ in (6.102) (verify that $\sum_{i=1}^4\boldsymbol{\mathcal{S}}_i^\varepsilon f = \boldsymbol{\mathcal{S}}^\varepsilon f$), the remainder is given by $\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u} = \sum_{i=1}^7\boldsymbol{\mathcal{R}}_i^\varepsilon\tilde{u}$. Using (6.109),we verify that $\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u}$ satisfies estimate (6.112) and the proof of the lemma is complete. □

Proof of Theorem 6.2.1. As $u^\varepsilon-\tilde{u}\in\mathrm{W}_{\mathrm{per}}(\Omega)$ and thanks to the triangle inequality, we have

$$\|u^\varepsilon-\tilde{u}\|_{\mathrm{L}^\infty(W)} = \|\big[u^\varepsilon-\tilde{u}\big]\|_{\mathrm{L}^\infty(\mathcal{W})} \le \|\big[u^\varepsilon\big]-\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}\|_{\mathrm{L}^\infty(\mathcal{W})} + \|\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}-\big[\tilde{u}\big]\|_{\mathrm{L}^\infty(\mathcal{W})}. \tag{6.121}$$

Let us estimate the two terms of the right hand side. First, note that $\boldsymbol{\eta} = \big[u^\varepsilon\big]-\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}$ satisfies $(\partial_t^2+\mathcal{A}^\varepsilon)\boldsymbol{\eta}(t) = \boldsymbol{\mathcal{R}}^\varepsilon\tilde{u}(t)$ in $\mathcal{W}^*_{\mathrm{per}}(\Omega)$ for a.e $t\in[0,T^\varepsilon]$, where $\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u}$ is defined in Lemma 6.2.11. Hence, using Corollary 4.2.2, the first term satisfies

$$\|\big[u^\varepsilon\big]-\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}\|_{\mathrm{L}^\infty(\mathcal{W})} \le C\varepsilon\Big(\|g^1\|_{\mathrm{H}^4}+\|g^0\|_{\mathrm{H}^4}+\textstyle\sum_{k=1}^5|\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^k)}+\|\partial_t^2\tilde{u}\|_{\mathrm{L}^\infty(\mathrm{H}^3)}\Big), \tag{6.122}$$

where C depends on T, λ, Y, $\|a\|_{\mathcal{C}^1(\bar{\Omega};\mathrm{W}^{2,\infty}(Y))}$, $\|a\|_{\mathcal{C}^2(\bar{\Omega};\mathrm{W}^{1,\infty}(Y))}$, $\|a\|_{\mathcal{C}^4(\bar{\Omega};\mathrm{L}^\infty(Y))}$, and δ. Next, using the definition of $\boldsymbol{\mathcal{B}}^\varepsilon$ (6.111) and the estimates (6.103) and (6.109), the second term of (6.121) satisfies

$$\|\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}-\big[\tilde{u}\big]\|_{\mathrm{L}^\infty(\mathcal{W})} \le C\varepsilon\Big(\textstyle\sum_{k=1}^5|\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^k)}+\|f\|_{\mathrm{L}^1(\mathrm{H}^2)}\Big), \tag{6.123}$$

where C depends on λ, Y, $\|a\|_{\mathcal{C}^1(\bar{\Omega};\mathrm{W}^{2,\infty}(Y))}$, $\|a\|_{\mathcal{C}^2(\bar{\Omega};\mathrm{W}^{1,\infty}(Y))}$, $\|a\|_{\mathcal{C}^4(\bar{\Omega};\mathrm{L}^\infty(Y))}$, and δ. Combining (6.121), (6.122), and (6.123), we obtain (6.66) and the proof of the theorem is complete. □

6.3 Effective equations for tensors with minimal regularity in the second variable

Theorem 6.2.1 provides an error estimate for a family of effective equations under the assumption that the tensor $a(x,y)$ satisfies the regularity

$$a\in\mathcal{C}^1(\bar{\Omega};\mathrm{W}^{2,\infty}(Y))\cap\mathcal{C}^2(\bar{\Omega};\mathrm{W}^{1,\infty}(Y))\cap\mathcal{C}^4(\bar{\Omega};\mathrm{L}^\infty(Y)).$$

The requirement on the regularity of $y\mapsto a(x,y)$ is severe. In this section, we adapt what was done in Section 4.2.6, for uniformly periodic tensors, and prove an error estimate for a tensor with minimal regularity in the second variable: $a\in\mathcal{C}^4(\bar{\Omega};\mathrm{L}^\infty(Y))$. In particular, this allows for discontinuities in the map $y\mapsto a(x,y)$. To enable this lower regularity of the tensor, we increase the regularity requirements on the effective solution, on the initial conditions, and on the right hand side.

Let us present the key points of the proof of the result. The lower regularity of the tensor ensures the correctors to belong to $\mathcal{C}^1(\bar{\Omega};\mathrm{H}^1(Y))$ instead of $\mathcal{C}^1(\bar{\Omega};\mathcal{C}^1(\bar{Y}))$. However, the higher regularity of the solution, combined with Sobolev embeddings, ensures $\tilde{u}\in\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{C}^5(\bar{\Omega}))$ instead of $\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^5(\Omega))$. Hence, we verify that the adaptation $\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}$ (defined in (6.111)) still belongs to $\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{W}_{\mathrm{per}}(\Omega))$. Note that in order to estimate the terms composing the remainder of Lemma 6.2.11, we generalize Lemma 4.2.10 to locally periodic functions (see Lemma 6.3.2).

Theorem 6.3.1. *Assume that the tensor $a(x,y)$ satisfies the regularity $a \in \mathcal{C}^4(\bar{\Omega};\mathrm{L}^\infty(Y))$. Furthermore, assume that the solution $\tilde{u}$ of (6.65), the initial conditions and the right hand side satisfy the regularity*

$$\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^7(\Omega)),\quad \partial_t\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^6(\Omega)),\quad \partial_t^2\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^5(\Omega)),$$
$$g^0 \in \mathrm{H}^6(\Omega),\quad g^1 \in \mathrm{H}^5(\Omega),\quad f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega)\cap\mathrm{H}^4(\Omega)).$$

Then the following estimate holds

$$\begin{aligned}\|u^\varepsilon-\tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} \le C\varepsilon\Big(&\|g^1\|_{\mathrm{H}^6(\Omega)}+\|g^0\|_{\mathrm{H}^6(\Omega)}+\|f\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^4(\Omega))}\\ &+\textstyle\sum_{k=1}^7|\tilde{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^k(\Omega))}+\|\partial_t^2\tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^5(\Omega))}\Big),\end{aligned} \tag{6.124}$$

where C depends only on T, λ, Y, $\|a\|_{\mathcal{C}^4(\bar{\Omega};\mathrm{L}^\infty(Y))}$, and δ.

The proof of Theorem 6.3.1 follows the same structure as that of Theorem 6.2.1. First, we investigate the regularity of the correctors. As $a \in \mathcal{C}^0(\bar{\Omega};\mathrm{L}^\infty(Y))$, we verify thanks to (6.108) that all the correctors belong (at least) to $\mathcal{C}^1(\bar{\Omega};\mathrm{W}_{\mathrm{per}}(\Omega))$ and that $\kappa^0,\kappa^1,\kappa^2 \in \mathcal{C}^2(\bar{\Omega};\mathrm{L}^2_0(\Omega))$. Furthermore, the following estimate (needed in the proof) hold

$$\begin{aligned}\max_{ijkl}\Big\{&\|\chi_i\|_{\mathcal{C}^0(\mathrm{H}^1)},\|\theta^0_{ij}\|_{\mathcal{C}^1(\mathrm{H}^1)},\|\theta^1_i\|_{\mathcal{C}^0(\mathrm{H}^1)},\|\kappa^0_{ijk}\|_{\mathcal{C}^2(\mathrm{H}^1)},\|\kappa^1_{ij}\|_{\mathcal{C}^2(\mathrm{H}^1)},\\ &\|\kappa^2_i\|_{\mathcal{C}^2(\mathrm{H}^1)},\|\rho^0_{ijkl}\|_{\mathcal{C}^1(\mathrm{H}^1)},\|\rho^1_{ijk}\|_{\mathcal{C}^1(\mathrm{H}^1)},\|\rho^2_{ij}\|_{\mathcal{C}^1(\mathrm{H}^1)},\|\rho^3_i\|_{\mathcal{C}^1(\mathrm{H}^1)},\\ &\|\bar{a}^{12}_{ij}\|_{\mathcal{C}^2},|b^{10}|,\|\bar{a}^{24}_{ijkl}\|_{\mathcal{C}^3},\|b^{22}_{ij}\|_{\mathcal{C}^2},\|\bar{a}^{22}_{ij}\|_{\mathcal{C}^2},|b^{20}|\Big\}\\ &\le C_1(a,\lambda,Y)+\delta C_2(a,\lambda,Y),\end{aligned} \tag{6.125}$$

where $C_i(a,\lambda,Y)$ depend only on λ, Y, $\|a\|_{\mathcal{C}^4(\mathrm{L}^\infty)}$, and δ is the parameter. Next, as $d \le 3$, the embedding $\mathrm{H}^2_{\mathrm{per}}(\Omega) \hookrightarrow \mathcal{C}^0_{\mathrm{per}}(\bar{\Omega})$ is continuous. Hence, we have $f \in \mathrm{L}^2(0,T^\varepsilon;\mathcal{C}^2_{\mathrm{per}}(\Omega))$ and the right hand side $\mathcal{S}^\varepsilon f$ of (6.102) belongs to $\mathrm{L}^2(0,T^\varepsilon;\mathcal{L}^2(\Omega))$. Consequently, $\boldsymbol{\varphi} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{W}_{\mathrm{per}}(\Omega))$ exists, is unique, and satisfies $\partial_t^2\varphi \in \mathrm{L}^2(0,T^\varepsilon;\mathcal{W}^*_{\mathrm{per}}(\Omega))$. We thus verify that (6.111) defines a linear map (still denoted $\boldsymbol{\mathcal{B}}^\varepsilon$)

$$\boldsymbol{\mathcal{B}}^\varepsilon : \mathrm{L}^2(0,T^\varepsilon;\mathrm{W}^{3,\infty}(\Omega)) \to \mathrm{L}^2(0,T^\varepsilon;\mathcal{W}^*_{\mathrm{per}}(\Omega)),\quad v \mapsto \boldsymbol{\mathcal{B}}^\varepsilon v(t)=\sum_{i=0}^4 \boldsymbol{\mathcal{B}}^\varepsilon_i(v(t))+\boldsymbol{\varphi}(t).$$

Again, the embedding $\mathrm{H}^2_{\mathrm{per}}(\Omega) \hookrightarrow \mathcal{C}^0_{\mathrm{per}}(\bar{\Omega})$ ensures that

$$\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{C}^5_{\mathrm{per}}(\Omega)),\quad \partial_t\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{C}^4_{\mathrm{per}}(\Omega)),\quad \partial_t^2\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{C}^3_{\mathrm{per}}(\Omega)),$$

and we have

$$\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{W}_{\mathrm{per}}(\Omega)),\quad \boldsymbol{\mathcal{B}}^\varepsilon\partial_t\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{L}^2(\Omega)),\quad \boldsymbol{\mathcal{B}}^\varepsilon\partial_t^2\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{W}^*_{\mathrm{per}}(\Omega)).$$

Furthermore, we verify that $\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}$ satisfies

$$(\partial_t^2+\mathcal{A}^\varepsilon)\boldsymbol{\mathcal{B}}^\varepsilon\tilde{u}(t)=\big[f(t)\big]+\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u}(t)\quad\text{in }\mathcal{W}^*_{\mathrm{per}}(\Omega)\quad\text{for a.e. } t\in[0,T^\varepsilon], \tag{6.126}$$

where the remainder $\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathcal{W}^*_{\mathrm{per}}(\Omega))$ is defined in the proof of Lemma 6.2.11. In order to estimate $\|\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{W}^*_{\mathrm{per}}(\Omega))}$, we need the following result.

Lemma 6.3.2. *Then $\gamma \in \mathcal{C}^0(\bar{\Omega};\mathrm{L}^2_{\mathrm{per}}(Y))$ and $v \in \mathrm{H}^2_{\mathrm{per}}(\Omega)$ satisfy the estimate*

$$\big\|\gamma\big(\cdot,\tfrac{\cdot}{\varepsilon}\big)v\big\|_{\mathrm{L}^2(\Omega)} \le C\|\gamma\|_{\mathcal{C}^0(\bar{\Omega};\mathrm{L}^2(Y))}\|v\|_{\mathrm{H}^2(\Omega)}, \tag{6.127}$$

for some constant C that depends only on Y, d and the bound on ε.

Proof. The proof follows the same lines as for Lemma 4.2.10. Let us first recall the notations. Let $\ell \in \mathbb{R}^d$ be the period of the tensor a and assume without loss of generality that $Y = (0,\ell_1)\times\cdots\times(0,\ell_d)$ and $\Omega = (0,\omega_1)\times\cdots\times(0,\omega_d)$. As Ω satisfies the assumption (4.25) (see Figure 4.2), the numbers $N_i = \omega_i/\ell_i\varepsilon$ are integers and the cells constituting Ω belongs to the set $\{\varepsilon(n\cdot\ell + Y) : 0 \le n_i \le N_i - 1\}$. Denoting $\Xi = \{\xi = n\cdot\ell : 0 \le n_i \le N_i - 1\}$, the domain Ω is then given by

$$\Omega = \operatorname{int}\Big(\bigcup_{\xi\in\Xi} \varepsilon(\xi + \bar{Y})\Big). \tag{6.128}$$

Furthermore, let $Z \subset \mathbb{R}^d$ be an open set with $\mathcal{C}^1$ boundary that contains Y and is contained in the neighbor cells, i.e.,

$$Y \subset Z \subset N_Y = (-\ell_1, 2\ell_1)\times\cdots\times(-\ell_d, 2\ell_d).$$

For example, $Z = F_Y^{-1}(S)$, where S is the open sphere of diameter $\sqrt{d}$ centered in $(1/2,\cdots,1/2)$ (recall that $d \le 3$) and $F_Y : N_Y \to (-1,2)^d$ is a smooth change of coordinates. As Z has a $\mathcal{C}^1$ boundary and $d \le 3$, Sobolev embedding theorem ensures that the embedding $\mathrm{H}^2(Z) \hookrightarrow \mathcal{C}^0(\bar{Y})$ is continuous. Hence, there exists a constant C_Y, depending only Y, such that

$$\|w\|_{\mathcal{C}^0(\bar{Y})} \le \|w\|_{\mathcal{C}^0(\bar{Z})} \le C_Y\|w\|_{\mathrm{H}^2(Z)} \le C_Y\|w\|_{\mathrm{H}^2(N_Y)} \qquad \forall w \in \mathrm{H}^2(N_Y). \tag{6.129}$$

We now prove the estimate. Using (6.128) and the Y-periodicity of $y \mapsto \gamma(x,y)$, we have

$$\big\|\gamma\big(\cdot,\tfrac{\cdot}{\varepsilon}\big)v\big\|^2_{\mathrm{L}^2(\Omega)} = \sum_{\xi\in\Xi}\int_{\varepsilon(\xi+Y)}\Big|\gamma\big(x,\tfrac{x}{\varepsilon}\big)v(x)\Big|^2\,\mathrm{d}x = \sum_{\xi\in\Xi}\int_Y\Big|\gamma\big(\varepsilon(\xi+y),y\big)v\big(\varepsilon(\xi+y)\big)\Big|^2\varepsilon^d\,\mathrm{d}y,$$

where we made the change of variables $x = \varepsilon(\xi+y)$. We define the function $v_{\xi,\varepsilon}(y) = v\big(\varepsilon(\xi+y)\big)$. As $v \in \mathrm{H}^2_{\mathrm{per}}(\Omega) \hookrightarrow \mathcal{C}^0_{\mathrm{per}}(\bar{\Omega})$, we have $v_{\xi,\varepsilon} \in \mathcal{C}^0(\bar{Y})$ and

$$\big\|\gamma\big(\cdot,\tfrac{\cdot}{\varepsilon}\big)v\big\|^2_{\mathrm{L}^2(\Omega)} \le \sum_{\xi\in\Xi}\varepsilon^d\|v_{\xi,\varepsilon}\|^2_{\mathcal{C}^0(\bar{Y})}\int_Y\Big|\gamma\big(\varepsilon(\xi+y),y\big)\Big|^2\,\mathrm{d}y \le \|\gamma\|^2_{\mathcal{C}^0(\bar{\Omega};\mathrm{L}^2(Y))}\sum_{\xi\in\Xi}\varepsilon^d\|v_{\xi,\varepsilon}\|^2_{\mathcal{C}^0(\bar{Y})}.$$

Using (6.129) gives $\|v_{\xi,\varepsilon}\|_{\mathcal{C}^0(\bar{Y})} \le C_Y\|v_{\xi,\varepsilon}\|_{\mathrm{H}^2(N_Y)}$. Furthermore, we have

$$\begin{aligned}\varepsilon^d\|v_{\xi,\varepsilon}\|^2_{\mathrm{H}^2(N_Y)} &= \int_{N_Y}|v_{\xi,\varepsilon}(y)|^2\varepsilon^d\mathrm{d}y + \int_{N_Y}|\nabla_y v_{\xi,\varepsilon}(y)|^2\varepsilon^d\mathrm{d}y + \int_{N_Y}|\nabla_y^2 v_{\xi,\varepsilon}(y)|^2\varepsilon^d\mathrm{d}y\\ &= \int_{N_Y}\big|v\big(\varepsilon(\xi+y)\big)\big|^2\varepsilon^d\mathrm{d}y + \varepsilon^2\int_{N_Y}\big|\nabla_x v\big(\varepsilon(\xi+y)\big)\big|^2\varepsilon^d\mathrm{d}y + \varepsilon^4\int_{N_Y}\big|\nabla_x^2 v\big(\varepsilon(\xi+y)\big)\big|^2\varepsilon^d\mathrm{d}y.\end{aligned}$$

Hence, the change of variable $x = \varepsilon(\xi+y)$ leads to

$$\begin{aligned}\big\|\gamma\big(\cdot,\tfrac{\cdot}{\varepsilon}\big)v\big\|^2_{\mathrm{L}^2(\Omega)} &\le C\|\gamma\|^2_{\mathcal{C}^0(\bar{\Omega};\mathrm{L}^2(Y))}\sum_{\xi\in\Xi}\|v\|^2_{\mathrm{H}^2(\varepsilon(\xi+N_Y))}\\ &\le (2d^2+1)C\|\gamma\|^2_{\mathcal{C}^0(\bar{\Omega};\mathrm{L}^2(Y))}\sum_{\xi\in\Xi}\|v\|^2_{\mathrm{H}^2(\varepsilon(\xi+Y))},\end{aligned}$$

where we used that every cell $\varepsilon(\xi+Y)$ belongs to the neighborhoods of $(2d^2+1)$ cells. This is (6.129) and the proof of the lemma is complete. □

Proof of Theorem 6.3.1. Applying Lemma 6.3.2 and using (6.125), we verify that the remainder $\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u}$ in (6.126) satisfies

$$\big\langle\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u}(t),\boldsymbol{w}\big\rangle_{\mathcal{W}^*_{\mathrm{per}},\mathcal{W}_{\mathrm{per}}} = \big((\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u})_0(t),\boldsymbol{w}\big)_{\mathcal{L}^2} + \big((\boldsymbol{\mathcal{R}}^\varepsilon\tilde{u})_1(t),\nabla\boldsymbol{w}\big)_{\mathrm{L}^2},$$

with the bound

$$\|(\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u})_0\|_{\mathrm{L}^\infty(\mathcal{L}^2(\Omega))} + \|(\boldsymbol{\mathcal{R}}^\varepsilon \tilde{u})_1\|_{\mathrm{L}^\infty(\mathrm{L}^2(\Omega))} \leq C\varepsilon^3\Big(\sum_{k=5}^7 |\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^k)} + \sum_{k=3}^5 |\partial_t^2 \tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^k)}\Big),$$

where C depends only on λ, Y, $\|a\|_{\mathcal{C}^4(\mathrm{L}^\infty)}$, b^{10}, b^{20}, and δ. Define now $\boldsymbol{\eta} = [u^\varepsilon] - \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}$. Using Lemma 6.3.2 and (6.125), we verify that

$$\|\boldsymbol{\eta}(0)\|_{\mathcal{L}^2(\Omega)} \leq C\varepsilon \|g^0\|_{\mathrm{H}^6(\Omega)}, \qquad \|\partial_t \boldsymbol{\eta}(0)\|_{\mathcal{L}^2(\Omega)} \leq C\varepsilon \|g^1\|_{\mathrm{H}^6(\Omega)}.$$

Hence, applying Corollary 4.2.2 gives

$$\|\boldsymbol{\eta}\|_{\mathrm{L}^\infty(\mathcal{W})} \leq C\varepsilon\Big(\|g^1\|_{\mathrm{H}^6(\Omega)} + \|g^0\|_{\mathrm{H}^6(\Omega)} + \sum_{k=1}^7 |\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^k)} + \|\partial_t^2 \tilde{u}\|_{\mathrm{L}^\infty(\mathrm{H}^5)}\Big). \tag{6.130}$$

Using once againg Lemma 6.3.2 and (6.125), we verify that

$$\|[\tilde{u}] - \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}\|_{\mathrm{L}^\infty(\mathcal{L}^2(\Omega))} \leq C\varepsilon\Big(\sum_{k=1}^7 |\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^k)} + \|f\|_{\mathrm{L}^1(\mathrm{H}^4)}\Big). \tag{6.131}$$

Finally, as $u^\varepsilon - \tilde{u} \in \mathrm{W}_{\mathrm{per}}(\Omega)$, the triangle inequality gives

$$\|u^\varepsilon - \tilde{u}\|_{\mathrm{L}^\infty(W)} = \|[u^\varepsilon - \tilde{u}]\|_{\mathrm{L}^\infty(\mathcal{W})} \leq \|\boldsymbol{\eta}\|_{\mathrm{L}^\infty(\mathcal{W})} + \|[\tilde{u}] - \boldsymbol{\mathcal{B}}^\varepsilon \tilde{u}\|_{\mathrm{L}^\infty(\mathcal{W})},$$

which, combined with (6.130) and (6.131), proves estimate (6.124). That completes the proof of Theorem 6.3.1. □

6.4 Simplified family of the effective equations

In this section, we discuss the possibility of simplifying the effective equations obtained in Section 6.2 (Definition 6.2.2). In particular, some of the effective operators seem, in practice, to be unnecessary in certain cases. As the removal of these operators leads to a significant gain of computational cost for the corresponding approximation, it has to be studied with attention.

In Section 6.2, we obtained a family of effective equations of the form

$$(1+\varepsilon b^{10}+\varepsilon^2 b^{20})\partial_t^2 \tilde{u} - \partial_i\big((a_{ij}^0 + \varepsilon a_{ij}^{12} + \varepsilon^2 a_{ij}^{22})\partial_j \tilde{u}\big) + \varepsilon^2 \partial_{ij}^2\big(\bar{a}_{ijkl}^{24}\partial_{kl}^2 \tilde{u}\big) - \varepsilon^2 \partial_i\big(b_{ij}^{22}\partial_j \tilde{u}\big) = f, \tag{6.132}$$

where the tensors depend on $x \in \Omega$. We recall that the effective equations for a uniformly periodic tensor are of the form (Chapter 4, Section 4.2).

$$\partial_t^2 \tilde{u} - a_{ij}^0 \partial_{ij}^2 \tilde{u} + \varepsilon^2 \bar{a}_{ijkl}^{24} \partial_{ijkl}^4 \tilde{u} - \varepsilon^2 b_{ij}^{22} \partial_{ij}^2 \tilde{u} = f,$$

where the tensors are constant. Comparing the two equations, we note that (6.132) contains the additional operators

$$\varepsilon L^1 = \varepsilon\big(b^{10}\partial_t^2 - \partial_i(a_{ij}^{12}\partial_j \cdot)\big), \qquad \varepsilon^2 L^{2,1} = \varepsilon^2\big(b^{20}\partial_t^2 - \partial_i(a_{ij}^{22}\partial_j \cdot)\big). \tag{6.133}$$

This difference naturally questions the role of εL^1 and $\varepsilon^2 L^{2,1}$ in (6.132). Indeed, a priori, the presence of εL^1 in (6.132) indicates that the homogenized equation must already be corrected for timescales $\mathcal{O}(\varepsilon^{-1})$. Despite several attempts, we failed to find an example of tensor for which $\tilde{u}$ exhibits a visible difference with and without εL^1 in (6.132). The use of $\varepsilon^2 L^{2,1}$ is illustrated in a numerical example in Section 6.5.1. However, for its influence to be notable, the variation of the map $x \mapsto a(x, y)$ must be sharp. These considerations interrogate the necessity of εL^1 and $\varepsilon^2 L^{2,1}$ in the effective equations. The prospect of removing these estimates from the equation is especially interesting as the corresponding cost of approximation is significantly reduced (as discussed in Chapter 7, Remark 7.2.5). For these reasons, we prove an error estimate that

quantifies the repercussion of the removal of εL^1 and $\varepsilon^2 L^{2,1}$. Nevertheless, no criterion was found to assess a priori whether these operators can be removed or not in practice.

For simplicity, we assume that ε and $a(x,y)$ are such that the matrix $a^0 + \varepsilon \check{a}^{12} + \varepsilon^2 \check{a}^{22}$ is positive definite, everywhere in Ω (see Remark 6.2.4). This assumption can be avoided, but the proof is more technical. Furthermore, recall that the role of εL^1 and $\varepsilon^2 L^{2,1}$ in the equation is precisely to replace the operator $\partial_i(\varepsilon \check{a}^{12} + \varepsilon^2 \check{a}^{22}\partial_j \cdot)$. Indeed, the decomposition into pairs of operators in (6.133) is a trick to guarantee unconditionally the well-posedness of the equation (see Section 6.2.2). Hence, to study the effects of εL^1 and $\varepsilon^2 L^{2,1}$ on the effective solution, we must study the tensors $\varepsilon \check{a}^{12}$ and $\varepsilon^2 \check{a}^{22}$. We thus consider the solution $\tilde{u} : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ the solution of

$$\begin{array}{ll} \partial_t^2 \tilde{u} - \partial_i\big((a^0_{ij} + \varepsilon \check{a}^{12}_{ij} + \varepsilon^2 \check{a}^{22}_{ij})\partial_j \tilde{u}\big) + \varepsilon^2 \partial^2_{ij}\big(\bar{a}^{24}_{ijkl}\partial^2_{kl}\tilde{u}\big) - \varepsilon^2 \partial_i\big(b^{22}_{ij}\partial_j \tilde{u}\big) = f & \text{in } (0, T^\varepsilon] \times \Omega, \\ x \mapsto \tilde{u}(t,x) \ \ \Omega\text{-periodic} & \text{in } [0, T^\varepsilon], \\ \tilde{u}(0,x) = g^0(x), \quad \partial_t \tilde{u}(0,x) = g^1(x) & \text{in } \Omega, \end{array} \tag{6.134}$$

where the tensors are defined in (6.61), (6.64), and (6.63). Referring to Remark 6.2.4, $\tilde{u}$ is an effective solution. Let then $\hat{u} : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$, be the solution of the equation

$$\begin{array}{ll} \partial_t^2 \hat{u} - \partial_i\big(a^0_{ij}\partial_j \hat{u}\big) + \varepsilon^2 \partial^2_{ij}\big(\bar{a}^{24}_{ijkl}\partial^2_{kl}\hat{u}\big) - \varepsilon^2 \partial_i\big(b^{22}_{ij}\partial_j \hat{u}\big) = f & \text{in } (0, T^\varepsilon] \times \Omega, \\ x \mapsto \hat{u}(t,x) \ \ \Omega\text{-periodic} & \text{in } [0, T^\varepsilon], \\ \hat{u}(0,x) = g^0(x), \quad \partial_t \hat{u}(0,x) = g^1(x) & \text{in } \Omega. \end{array} \tag{6.135}$$

We prove the following error estimate for $\tilde{u} - \hat{u}$.

Theorem 6.4.1. *Assume that the assumptions of Theorem 6.2.1 hold. Then the following estimate holds*

$$\|\tilde{u} - \hat{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} \leq C\big(\varepsilon^{-1} \max_{ij} \|\check{a}^{12}_{ij}\|_{\mathcal{C}^0(\bar{\Omega})} + \max_{ij} \|\check{a}^{22}_{ij}\|_{\mathcal{C}^0(\bar{\Omega})}\big)|\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^1)}, \tag{6.136}$$

where C depends only on λ, and T.

Combined with Theorem 6.2.1 and Remark 6.2.4, Theorem 6.4.1 ensures that (under the regularity assumptions of Theorem 6.2.1)

$$\|u^\varepsilon - \hat{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} \leq C\Big(\varepsilon + \varepsilon^{-1} \max_{ij} \|\check{a}^{12}_{ij}\|_{\mathcal{C}^0(\bar{\Omega})} + \max_{ij} \|\check{a}^{22}_{ij}\|_{\mathcal{C}^0(\bar{\Omega})}\Big),$$

where the constant C depends on Ω and T^ε only through the norms of the data and $\tilde{u}$. This estimate ensures that for some tensors and under some regimes of ε, the corrections εL^1 and $\varepsilon^2 L^{2,1}$ in the effective equations (6.132) are superfluous. Namely, if

$$\max_{ij} \|\check{a}^{12}_{ij}\|_{\mathcal{C}^0(\Omega)} = \mathcal{O}(\varepsilon^2), \qquad \max_{ij} \|\check{a}^{22}_{ij}\|_{\mathcal{C}^0(\Omega)} = \mathcal{O}(\varepsilon), \tag{6.137}$$

then approximating u^ε with $\hat{u}$ is accurate enough. In all the numerical examples that we considered, (6.137) was satisfied. Nevertheless, to take advantage of this fact in practice, we need a criterion to determine whether (6.137) holds without having to compute $\check{a}^{12}_{ij}$ and $\check{a}^{22}_{ij}$. Ideally, the knowledge of $a(x,y)$, and in particular of $x \mapsto a(x,y)$, should be enough to take the decision of the removal of $\varepsilon \check{a}^{12}_{ij}$ and $\varepsilon^2 \check{a}^{22}_{ij}$ in the equation. Unfortunately, we were unable to derive such criterion.

Proof of the error estimate

Let us recall the functional spaces introduced in Section 2.1.2. We define the bilinear forms

$$\begin{aligned} \big(v, w\big)_{\mathcal{S}} &= \big(v, w\big)_{\mathrm{L}^2(\Omega)} + \big(\varepsilon^2 b^{22}\nabla v, \nabla w\big)_{\mathrm{L}^2(\Omega)}, \\ A\big(v, w\big) &= \big(a^0 \nabla v, \nabla w\big)_{\mathrm{L}^2(\Omega)} + \big(\varepsilon^2 \bar{a}^{24}\nabla^2 v, \nabla^2 w\big)_{\mathrm{L}^2(\Omega)}. \end{aligned}$$

Equipped with the inner product $(\cdot,\cdot)_{\mathcal{S}}$ and $A(\cdot,\cdot)$, respectively, we verify that the spaces

$$\mathcal{S}(\Omega)=\{v\in \mathrm{L}^2_0(\Omega):\sqrt{b^{22}}\nabla v\in[\mathrm{L}^2(\Omega)]^d\},\quad \mathcal{V}(\Omega)=\{v\in \mathrm{W}_{\mathrm{per}}(\Omega):\sqrt{\bar{a}^{24}}\nabla^2 v\in[\mathrm{L}^2(\Omega)]^{d\times d}\},$$

are Hilbert spaces. Define the error $\eta=\tilde{u}-\hat{u}$. Using (6.134) and (6.135), we verify that for any $w\in\mathcal{V}(\Omega)$

$$\big(\partial_t\eta(t),w\big)_{\mathcal{S}}+A\big(\eta(t),w\big)=-\big((\varepsilon\check{a}^{12}+\varepsilon^2\check{a}^{22})\nabla\tilde{u},\nabla w\big)_{\mathrm{L}^2(\Omega)},\tag{6.138}$$

and $\eta(0)=\partial_t\eta(0)=0$. To estimates $\|\eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)}$, we need the following generalization of Lemma 4.2.1.

Lemma 6.4.2. *Let $\eta\in\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{V}(\Omega))$ with $\partial_t\eta\in\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{S}(\Omega))$ and $\partial_t^2\eta\in\mathrm{L}^2(0,T^\varepsilon;\mathcal{S}(\Omega))$ satisfies*

$$\begin{aligned}&\big(\partial_t^2\eta(t),w\big)_{\mathcal{S}}+A\big(\eta(t),w\big)=\big(r(t),\nabla w\big)_{\mathrm{L}^2}\qquad\forall w\in\mathcal{V}(\Omega)\quad \textit{for a.e. } t\in[0,T^\varepsilon],\\&\eta(0)=\partial_t\eta(0)=0,\end{aligned}\tag{6.139}$$

where $r\in\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))$. Then the following estimate holds

$$\|\eta\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))}\le C\varepsilon^{-2}\|r\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))},$$

where C depends only on λ and T.

Proof. For a.e. $t\in[0,T^\varepsilon]$, let $\hat{v}(t)\in\mathcal{V}(\Omega)$ be the solution of the elliptic problem

$$A\big(\hat{v}(t),w\big)=\big(\partial_t\eta(t),w\big)_{\mathcal{S}}\quad\forall w\in\mathcal{V}(\Omega).\tag{6.140}$$

Thanks to Lax–Milgram theorem, $\hat{v}(t)$ exists and is unique. Differentiating (6.140) with respect to t, we find that for all $w\in\mathcal{V}(\Omega)$, $A\big(\partial_t\hat{v}(t),w\big)=\big(\partial_t^2\eta(t),w\big)_{\mathcal{S}}$. Using the test function $w=\hat{v}(t)$ in (6.139), we thus get

$$\big(r(t),\nabla\hat{v}(t)\big)_{\mathrm{L}^2}=\big(\partial_t^2\eta(t),\hat{v}(t)\big)_{\mathcal{S}}+A\big(\eta(t),\hat{v}(t)\big)=A\big(\partial_t\hat{v}(t),\hat{v}(t)\big)+\big(\partial_t\eta(t),\eta(t)\big)_{\mathcal{S}}.$$

Thanks to the symmetry of the forms A and $(\cdot,\cdot)_{\mathcal{S}}$, this equality can be rewritten as

$$\tfrac{1}{2}\tfrac{\mathrm{d}}{\mathrm{d}t}\Big(A\big(\hat{v}(t),\hat{v}(t)\big)+\big(\eta(t),\eta(t)\big)_{\mathcal{S}}\Big)=\big(r(t),\nabla\hat{v}(t)\big)_{\mathrm{L}^2}.$$

Integrating over $[0,\xi]$, we get, for any $\xi\in[0,T^\varepsilon]$,

$$A\big(\hat{v}(\xi),\hat{v}(\xi)\big)+\big(\eta(\xi),\eta(\xi)\big)_{\mathcal{S}}=2\int_0^\xi\big(r(t),\nabla\hat{v}(t)\big)_{\mathrm{L}^2}.$$

Using the Cauchy–Schwartz, Hölder, and Young inequalities, we bound the term of the right hand side as

$$2\int_0^\xi\big(r(t),\nabla\hat{v}(t)\big)_{\mathrm{L}^2}\le 2\|r\|_{\mathrm{L}^1(\mathrm{L}^2)}\|\nabla\hat{v}\|_{\mathrm{L}^\infty(\mathrm{L}^2)}\le\tfrac{2}{\lambda}\|r\|^2_{\mathrm{L}^1(\mathrm{L}^2)}+\tfrac{\lambda}{2}\|\nabla\hat{v}\|^2_{\mathrm{L}^\infty(\mathrm{L}^2)}.$$

Combining the two last equations with the ellipticity of A, we obtain successively

$$\tfrac{\lambda}{2}\|\nabla\hat{v}\|^2_{\mathrm{L}^\infty(\mathrm{L}^2)}\le\tfrac{2}{\lambda}\|r\|^2_{\mathrm{L}^1(\mathrm{L}^2)},\qquad\|\eta\|^2_{\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{S})}\le\tfrac{4}{\lambda}\|r\|^2_{\mathrm{L}^1(\mathrm{L}^2)}.$$

Using Hölder's inequality gives $\|r\|_{\mathrm{L}^1(\mathrm{L}^2)}\le T\varepsilon^{-2}\|r\|_{\mathrm{L}^\infty(\mathrm{L}^2)}$ and we obtain the desired estimate. □

Proof of Theorem 6.4.1. Combining (6.138) with Lemma 6.4.2, we verify that $\eta=\tilde{u}-\hat{u}$ satisfies

$$\|\eta\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))}\le C\big(\varepsilon^{-1}\max_{ij}\|\check{a}^{12}_{ij}\|_{\mathcal{C}^0(\bar{\Omega})}+\max_{ij}\|\check{a}^{22}_{ij}\|_{\mathcal{C}^0(\bar{\Omega})}\big)|\tilde{u}|_{\mathrm{L}^\infty(\mathrm{H}^1)},$$

where C depends only on λ and T. The proof of the theorem is complete. □

6.5 Numerical experiments

In this section, we test the theoretical results obtained in this chapter in diverse numerical experiments. First, we consider a one-dimensional example in a pseudoinfinite domain and verify that the family of effective equations describes well the long time behavior of the heterogeneous wave. Second, we consider a two-dimensional example in a small domain and verify the long time validity of an effective solution. Finally, in a two-dimensional pseudoinfinite domain, we compare an effective solution with the classical homogenized solution. In particular, we provide a visualization of the long time dispersion effects.

6.5.1 One-dimensional example

We consider here a one-dimensional example in a pseudoinfinite locally periodic medium. Let us fix the initial data and the right hand side for the test problem as $g_0(x) = e^{-20x^2}$, $g_1 = 0$, $f = 0$. Let us first consider the locally periodic tensor given by

$$a\big(x, \tfrac{x}{\varepsilon}\big) = \tfrac{249}{419} + \tfrac{1}{6}\sin(2\pi x) + \tfrac{1}{6}\sin\big(2\pi\tfrac{x}{\varepsilon}\big), \tag{6.141}$$

with $\varepsilon = 1/20$ ($Y = (-0.5, 0.5)$). Denoting $c = \frac{249}{419}$ and

$$I(x,y) = \int^y \frac{1}{a(x,z)}\,\mathrm{d}z = \frac{6\sqrt{2}\operatorname{atan}\left(\frac{\sqrt{2}(\tan(\pi x)(6c+\sin(2\pi y))+1)}{\sqrt{72c^2+24c\sin(2\pi y)-\cos(4\pi y)-1}}\right)}{\pi\sqrt{72c^2+24c\sin(2\pi y)-\cos(4\pi y)-1}},$$

we verify that

$$a^0(x) = \frac{1}{I(x,1/2) - I(x,-1/2)}, \qquad \chi(x,y) = a^0(x)I(x,y) - y + C_0,$$

where C_0 is such that $\big\langle \chi(x)\big\rangle_Y = 0$. Furthermore, we have

$$\int_\Omega \sqrt{a^0(x)}\,\mathrm{d}x \approx 3/4, \quad \int_\Omega \big\langle \chi(x)^2\big\rangle_Y\,\mathrm{d}x \approx 1.1978\cdot 10^{-3}.$$

For these data, we compare the solution u^ε of (6.3), the homogenized solution u^0 and effective solutions $\tilde{u}$ in the family $\mathcal{E}$ (Definition 6.1.2) at the time $T^\varepsilon = \varepsilon^{-2} = 400$. For the waves not to reach the boundary, we set $\Omega = (-301, 301)$. We denote $\tilde{u}_r$ the solutions of the family $\mathcal{E}$ defined in Definition 6.1.2, where the subscript r specify the dependence on the parameter r. To approximate u^ε, we use a spectral method (Section 2.3) on a grid of size $h = \varepsilon/25$ and a leap-frog scheme for the time integration with a time step $\Delta t = h/50$ (Section A.5). To approximate u^0 and $\tilde{u}_r$, the same methods are used with $h = \varepsilon/4$ and $\Delta t = h/50$. Note that a gradient method is needed as the second order ODE for $\tilde{u}_r$ is implicit.

In Figure 6.1, we display the frontal wave of u^ε, u^0, and $\tilde{u}_r$ for some $r \in [0, 0.1]$ at $t = \varepsilon^{-2} = 400$. As expected, we observe that the macroscopic behavior of u^ε is not well described by u^0. On the contrary, $\tilde{u}_r$ describes well these effects, as predicted by Theorem 6.1.1. Let us now compare the L^2 error between $u^\varepsilon(t)$ and $u^0(t)$, $\tilde{u}_r(t)$. Let us denote the normalized error as

$$\mathrm{err}(v)(t) = \|(u^\varepsilon - v)(t)\|_{\mathrm{L}^2(\Omega)}/\|u^\varepsilon(t)\|_{\mathrm{L}^2(\Omega)}, \quad v \in \{u^0, \tilde{u}\},$$

In Figure 6.2, the computed errors for u^0 and $\tilde{u}_r$ are compared. First, we note that the error of the homogenized solution increases comparatively fast with respect to t. Next, we see that the error of $\tilde{u}_r$ increases notably as r increases. As Figure 6.1 showed, the frontal wave is well captured for all the values of r, hence the error is located elsewhere. In Figure 6.3, u^ε, u^0, and $\tilde{u}_r$

are displayed away from the frontal wave. We observe that as r increases, $\tilde{u}_r$ significantly drives away from u^ε. Indeed, for most of the values of r, $\tilde{u}_r$ is locally even worse than u^0. We conclude that the elements of the family of effective equations $\mathcal{E}$ capture well the long time dispersion effects at the frontal wave, while u^0 does not. However, as r increases, $\tilde{u}_r$ drifts away from the frontal wave. From this example, we can thus conclude that a too large increase of the parameter has negative repercussion on the accuracy of the effective solutions.

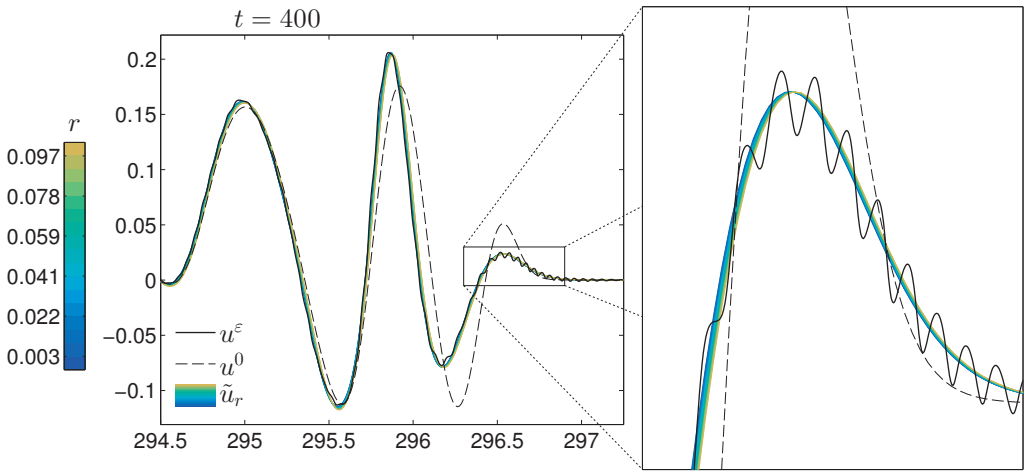

Figure 6.1: Comparison between the frontal waves of u^ε, u^0, and $\tilde{u}_r$ at time $t = 400$ and zoom on $x \in [296.3, 296.9]$.

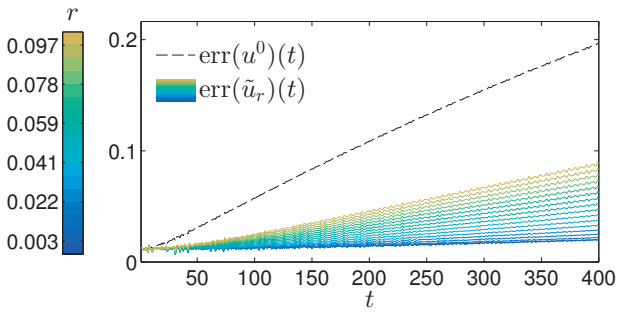

Figure 6.2: Comparison of the normalized L^2 error between u^ε and u^0, $\tilde{u}_r$ over the time interval $[0, 400]$.

In Section 6.4, we discussed the importance of the operators εL^1, $\varepsilon^2 L^{2,1}$ in the effective equations. In one dimension, we know that $L^1 = 0$. However, we verify in an example that the operator $\varepsilon^2 L^{2,1} = \varepsilon^2(b^{20}\partial_t^2 - \partial_x(a^{22}\partial_x \cdot))$ is important in certain situations. Let $\hat{u}_r$ be the solution of the equation

$$\partial_t^2 \hat{u}_r - \partial_x(a^0 \partial_x \hat{u}_r) + \varepsilon^2\big(\partial_x^2(a^{24}\partial_x^2 \hat{u}_r) - \partial_x(b^{22}\partial_x \partial_t^2 \hat{u}_r)\big) = f \quad \text{in } (0, T^\varepsilon] \times \Omega, \tag{6.142}$$

with periodic boundary conditions and initial conditions $\hat{u}_r(0) = g^0, \partial_t \hat{u}_r(0) = g^1$ (we write $w = w_r$ to specify the dependence on the parameter r). For the tensor (6.141), we verify

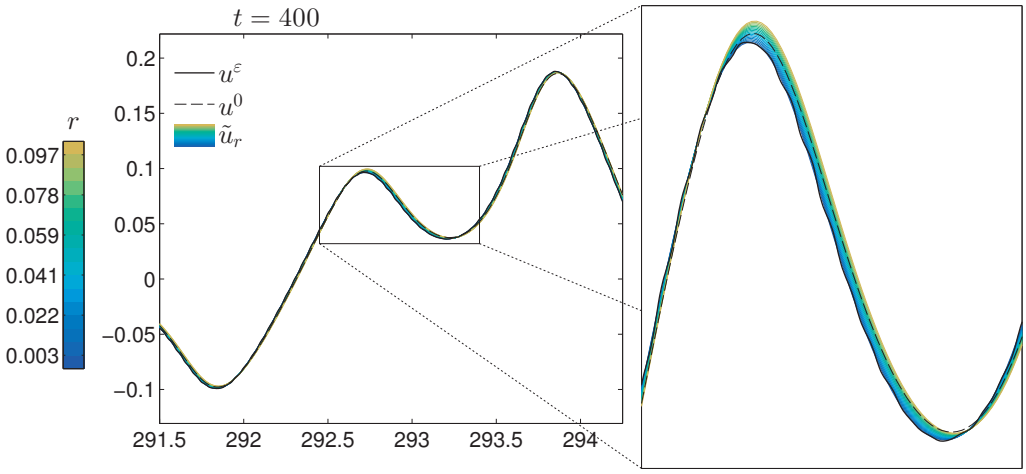

Figure 6.3: Comparison of u^ε, u^0, and $\tilde{u}_r$ at $t = 400$ away from the frontal wave ($x \in [291.5, 294.25]$) and zoom.

that $\hat{u}_r$ and $\tilde{u}_r$ are almost the same function. Indeed, for $r = 0.16$ we compute the difference $\|\hat{u}_r - \tilde{u}_r\|_{\mathrm{L}^\infty(\mathrm{L}^2)} = 2.681 \cdot 10^{-2}$ and $\hat{u}_r$ and $\tilde{u}_r$ can not be distinguished at the macroscopic scale. To obtain an example where the difference between $\hat{u}_r$ and $\tilde{u}_r$ is significant, we consider a tensor with a sharper variation in the slow variable. Let us then define the tensor

$$a\big(x, \tfrac{x}{\varepsilon}\big) = \tfrac{676}{1221} + \tfrac{1}{4}\mathrm{erf}\big(10(\{x\} - 1/4)\big) - \tfrac{1}{4}\mathrm{erf}\big(10(\{x\} - 3/4)\big) + \tfrac{1}{2}\cos\big(2\pi\tfrac{x}{\varepsilon}\big),$$

where $\{x\} = x - \lfloor x \rfloor$ is used to extend $x \mapsto a(x, y)$ from $[0, 1[$ to $\mathbb{R}$ by periodicity. We fix $\varepsilon = 1/20$. As for the previous tensor, $a^0(x)$ and $\chi(x, \cdot)$ can be computed analytically. We verify that $\int_\Omega \sqrt{a^0(x)}\,\mathrm{d}x \approx 0.75$ and $\int_\Omega \langle \chi(x)^2 \rangle_Y \,\mathrm{d}x \approx 1.0452 \cdot 10^{-2}$. For this tensor, the approximation of u^ε is more demanding. Indeed, if the grid is not sufficiently fine, the sharpness of the variation in x provokes the instability of the numerical method. We thus fix the domain $\Omega = (-4, 4)$ and apply the same method as for the previous example with mesh sizes $h = \varepsilon/40$ and $\Delta t = h/100$. To compute $\hat{u}_r$ and $\tilde{u}_r$, we use the same method with the mesh sizes $h = \varepsilon/4$ and $\Delta t = h/50$. In Figure 6.4, we compare the errors computed for $\tilde{u}_r$ and $\hat{u}_r$ for $r = 0.16$. We observe that $\mathrm{err}(\tilde{u}_r)$ remains small at all times $t \in [0, 400]$, while $\mathrm{err}(\hat{u}_r)$ becomes significantly large as t increases. From these examples, we conclude that $\varepsilon^2 L^{2,1}$ is necessary only for certain tensor. Furthermore, its importance is connected to the variation of the map $x \mapsto a(x, y)$.

6.5.2 Two-dimensional example

Let us now consider an example of locally periodic media in two dimensions. First, we compare the effective solution with the original wave and the homogenized solution in a small domain. Then, we compare the effective solution with the homogenized solution in a pseudoinfinite domain. Note that the effective solution is approximated using the numerical method defined in Chapter 7, Section 7.2.

Let us first define the locally periodic tensor describing the medium. We let the reference cell be $Y = (\text{-}1/2, 1/2)^2$. For given parameters $s, c \in \mathbb{R}$, let us define $\varphi[s, c] : \mathbb{R} \to \mathbb{R}$ as

$$\varphi[s, c](z) = \frac{1}{2}\big(\mathrm{erf}\big(s(z - c)\big) + 1\big), \qquad \mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-s^2}\,\mathrm{d}s.$$

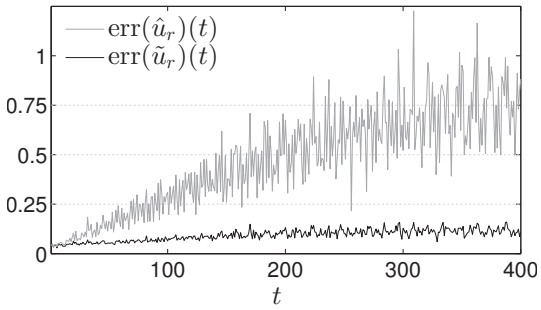

Figure 6.4: Comparison between the errors of the effective solution $\tilde{u}_r$ and the solution of (6.142) w_r for $r = 0.16$ over the time interval $[0, 400]$.

The function $\varphi[s, c]$ is a "smooth step function" between 0 and 1. The parameter s determines the slope of the step, and c is its center. We define the tensor $a(x, y)$ on the subdomain $(x, y) \in (0, 1)^2 \times Y$ and then extend it by periodicity to $\mathbb{R}^2 \times \mathbb{R}^2$. Let $R[\theta]$ be the matrix of rotation of angle θ. To each macro point $x \in (0, 1)^2$, we associate an angle $\alpha_x \in [0, \pi/2]$ and the associated matrix of rotation $R_x = R[\alpha_x]$ of $y \in \mathbb{R}^d$:

$$\begin{aligned}
&\alpha_x = \tfrac{\pi}{8}\big(-\cos(2\pi x_1) + 1\big)\big(-\cos(2\pi x_2) + 1\big),\\
&R_x y = \big((R_x y)_1, (R_x y)_2\big) = \big(y_1 \cos(\alpha_x) + y_2 \sin(\alpha_x), -y_1 \sin(\alpha_x) + y_2 \cos(\alpha_x)\big).
\end{aligned}$$

The tensor $a : (0, 1)^2 \times Y \to \mathbb{R}^{2\times 2}$ is then defined as

$$\begin{aligned}
a(x, y) &= \begin{pmatrix} \tilde{a}(x, y) & 0 \\ 0 & \tilde{a}(x, y) \end{pmatrix},\\
\tilde{a}(x, y) &= \frac{1}{2} + \prod_{i=1}^{2} \mu(y_i)\varphi[s, c_1^i]\big((R_x y)_i\big) - \varphi[s, c]\big((R_x y)_i\big),
\end{aligned} \tag{6.143}$$

where we set the parameters $s = 10$, $c^1 = (-1/4, 1/4)$, $c^2 = (-1/8, 1/8)$, and $\mu(y_i) = \varphi[50, -0.45](y_i) - \varphi[50, 0.45](y_i)$ is a cutoff function in the i-th direction. We then extend $a(x, y)$ by periodicity to $\mathbb{R}^2 \times \mathbb{R}^2$: $a^\sharp(x, y) = a\big(\{x\}_{[0,1[}, \{y\}_Y\big)$, where $\{x\}_{[0,1[} = x - \lfloor x \rfloor$ and $\{y\}_Y = \{y + 1/2\}_{[0,1[} - 1/2$. In Figure 6.5, we display the tensor $a^\varepsilon(x) = a\big(x, \frac{x}{\varepsilon}\big)$, where $a(x, y)$ is defined in (6.143), in $(0, 1)^2$ and for $\varepsilon = 1/10$, $1/20$, and $1/30$.

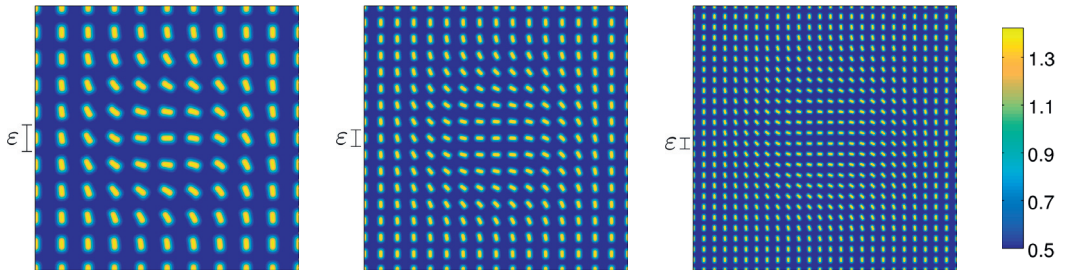

Figure 6.5: Tensor $a^\varepsilon(x) = a\big(x, \frac{x}{\varepsilon}\big)$ where $a(x, y)$ is defined in (6.143) displayed in $(0, 1)^2$ for, respectively from left to right, $\varepsilon = 1/10$, $1/16$, and $1/25$.

Example in a small domain

Let us consider the test problem given by the data $g_0(x) = e^{-40x^2}$, $g_1 = 0$, $f = 0$, and the tensor $a^\varepsilon(x) = a\big(x, \frac{x}{\varepsilon}\big)$, with $\varepsilon = 1/10$, where $a(x, y)$ is defined in (6.143). We are interested in the long time behavior of u^ε. In order to be able to approximate u^ε, we consider the bounded domain $\Omega = (-1, 1)^2$. We compare u^ε at $t = \varepsilon^{-2}$ with the homogenized solution u^0 and an effective solution $\bar{u}$ in the family $\mathcal{E}$ (see Section 6.2.1). We compute the oscillating wave u^ε as follows. The space discretization is done using P1 FEM on a mesh of size $h_{\text{ref}} = \varepsilon/32$. The leap frog scheme is used for the time integration with the timestep $\Delta t = h_{\text{ref}}/40$. To approximate $\bar{u}$, we use the spectral homogenization method given in Section 7.2.2 (an obvious simplification of the method is used to compute u^0). Note that for $\bar{u}$, a gradient method is needed as the obtained ODE is implicit. The settings of the spectral homogenization method are as follows. In Step 1, we approximate the effective tensors at the nodes of the grid of Ω of size $\Delta x = \varepsilon/8$ ($M = 160$ in both directions) with a $\mathcal{P}^2$-FEM on a mesh of size $h = 1/200$. As $x \mapsto a(x, y)$ is 1-periodic, we compute the tensors only at the points lying in the subdomain $(0, 1)^2$ and extend them by periodicity. In Step 2, the spectral method is used to approximate $\bar{u}$ and u^0 on the same grid, i.e., $N = M$. The leap frog scheme with timestep $\Delta t = \Delta x/50$ is used for the time integration. For $v \in \{u^0, \bar{u}\}$, we denote the normalized error $\text{err}(v)(t) = \|(u^\varepsilon - v)(t)\|_{\text{L}^2(\Omega)}/\|u^\varepsilon(t)\|_{\text{L}^2(\Omega)}$. In Figure 6.6, we observe that the error for u^0 increases notably with respect to t, while the error for $\bar{u}$ stays low. This example illustrates the result of Theorem 6.2.1 that establishes that the elements of the family of effective equations describes well the behavior of u^ε up to timescales $\mathcal{O}(\varepsilon^{-2})$. Visualizations of u^ε, u^0, and $\bar{u}$ are displayed in Figure 6.7 at $t = \varepsilon^{-2} = 100$. The macroscopic difference between the two surfaces u^ε and u^0 is clearly visible. On the contrary, $\bar{u}$ describes well u^ε up to the micro oscillations, as predicted by Theorem 6.2.1.

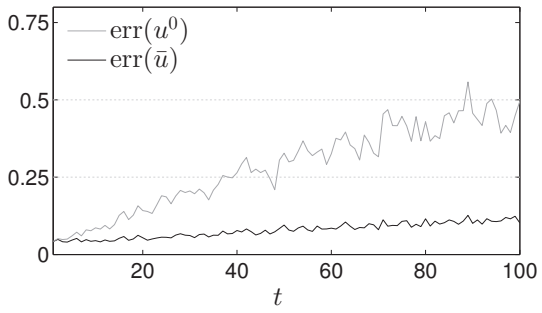

Figure 6.6: Comparison of the normalized errors of the effective solution $\bar{u}$ and the homogenized solution u^0 over the time interval $[0, 100]$.

Example in a pseudoinfinite domain

Let us now consider the same locally periodic medium in a pseudoinfinite domain. let the data be $g_0(x) = e^{-100x^2}$, $g_1 = 0$, $f = 0$. We are interested in approximating u^ε at time $T = 50$. We define the pseudoinfinite domain as

$$\Omega = (-L_1, L_1) \times (-L_2, L_2), \qquad L_i = \left\lfloor \sqrt{\langle a^0_{ii} \rangle_Y} T \right\rfloor + 2.$$

In such a large domain, approximating u^ε with reasonable accuracy is not possible. We are however able to approximate an effective solution $\bar{u}$ in the family defined in Definition 6.2.2 and Theorem 6.2.1 ensures that $\bar{u}$ is a good approximation of u^ε. We compare $\bar{u}$ with the homogenized solution u^0. To compute $\bar{u}$, we use the spectral homogenization method defined in Section 7.2.2

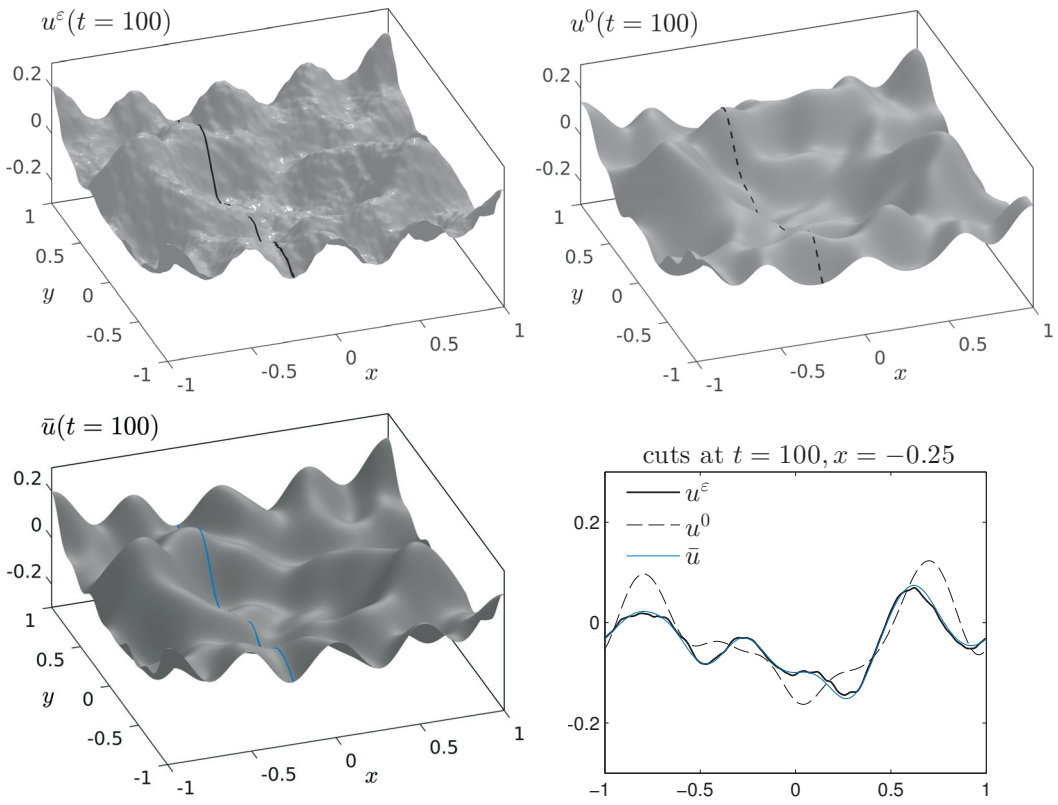

Figure 6.7: Comparison on Ω between u^ε (top-left), the homogenized solution u^0 (top-right) and the effective solution $\bar{u}$ (bottom-left) and cuts at $x = -0.25$ (bottom-right) at $t = \varepsilon^{-2} = 100$.

(an obvious simplification of the method is used to compute u^0). We re-use the effective tensors computed in the previous example on the grid of Ω of size $\Delta x = \varepsilon/8$, i.e., $M_i = 80 \cdot L_i$. In Step 2, we apply the spectral method on the subgrid with $N_i = M_i/2$ nodes in each directions. The leap frog scheme with timestep $\Delta t = \Delta x/16$ is used for the time integration. For $\bar{u}$, a gradient method is needed as the ODE is implicit. The obtained approximations are displayed on subdomains in Figure 6.8: $\bar{u}$ on the top-left plot, u^0 on the top-right plot, and the corresponding cuts along $y = 0$ in the bottom plot. We observe that both functions have variations at the macroscopic scale, which are due to the dependence of the tensors in the slow variable. The front waves of $\bar{u}$ and u^0 are clearly distinct. In particular, the amplitude of the front wave of $\bar{u}$ is notably smaller than that of u^0.

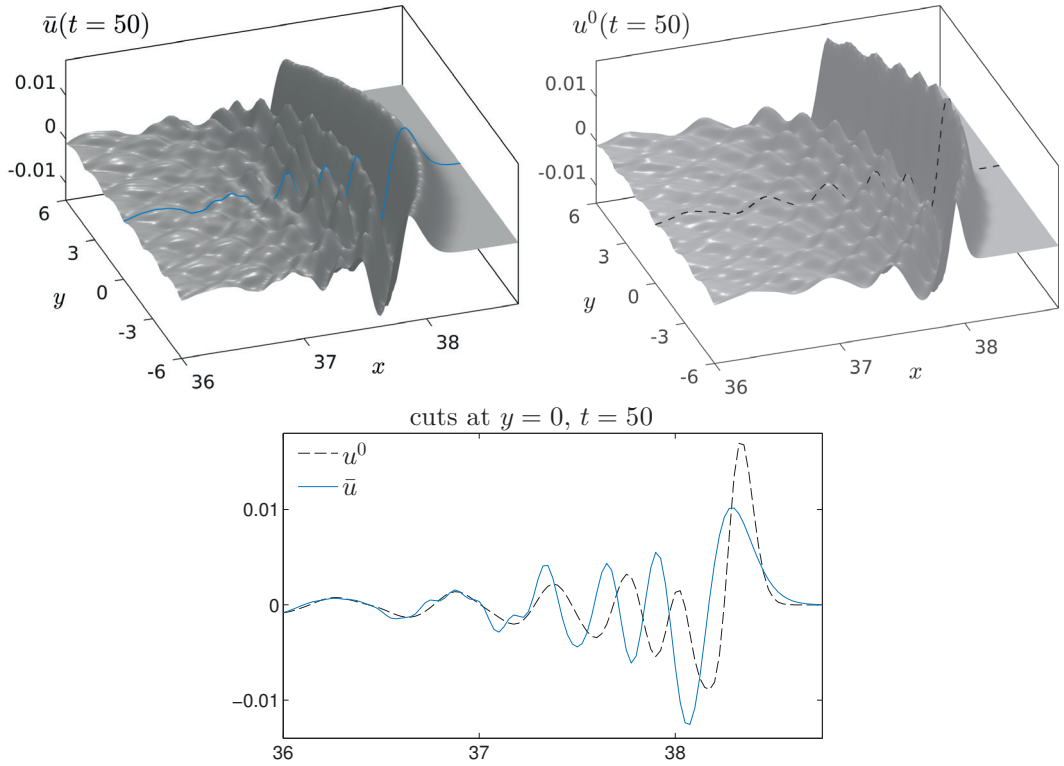

Figure 6.8: Top: Comparison of the effective solution $\bar{u}$ and the homogenized solution u^0 on the subdomain $[36, 38.75] \times [-6, 6]$ at $t = 50$. Bottom: Corresponding cuts along $y = 0$.

7 Analysis of numerical homogenization methods for long time wave propagation

In Chapter 6, we defined a family of effective equations for wave propagation in locally periodic media at timescales $\mathcal{O}(\varepsilon^{-2})$. In this chapter, we analyze numerical methods that are designed to approximate effective solutions. We consider an arbitrarily large hypercube $\Omega \subset \mathbb{R}^d$ and let $a^\varepsilon(x)$ be a locally periodic tensor: $a^\varepsilon(x) = a\big(x, \frac{x}{\varepsilon}\big)$, where $a(x,y)$ is Ω-periodic in x and Y-periodic in y (Y is a reference cell, e.g. $Y = (0,1)^d$). For $T^\varepsilon = \varepsilon^{-2}T$, we consider the wave equation: $u^\varepsilon : [0, T^\varepsilon] \times \mathbb{R}^d \to \mathbb{R}$ such that

$$\partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \big(a^\varepsilon(x)\nabla_x u^\varepsilon(t,x)\big) = f(t,x) \quad \text{in } (0,T^\varepsilon] \times \Omega, \tag{7.1}$$

where $u^\varepsilon(0,x)$, $\partial_t u^\varepsilon(0,x)$ are given and periodic boundary conditions are imposed.

In the first part of the chapter, we study a method designed specifically for the one-dimensional case. In that case, one effective equation in the family does not have a fourth order operator in space and reads

$$\partial_t^2 \bar{u}(t,x) - \partial_x\big(a^0(x)\partial_x \bar{u}(t,x)\big) - \varepsilon^2 \partial_x\big(b^2(x)\partial_x \partial_t^2 \bar{u}(t,x)\big) = f(t,x) \quad \text{in } (0,T^\varepsilon] \times \Omega,$$

where the coefficients a^0 and b^2 can be computed with the first corrector. We can thus easily modify the finite element heterogeneous multiscale method (FE-HMM), defined in Section 3.4, to capture the long time dispersive effects of u^ε. This method, called the FE-HMM-L, was introduced in [10, 9], and was fully analyzed for small domains in [13] (this analysis is presented in Section 7.1.3). The following error estimate is proved between the approximation of the FE-HMM-L u_H and u^ε:

$$\|u^\varepsilon - u_H\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} \leq C\left(\varepsilon + \left(\frac{h}{\varepsilon^2}\right)^2 + \frac{H^{\ell+1}}{\varepsilon^2}\right),$$

where h is the micro mesh size, H is the macro mesh size, and ℓ is the degree of the macro finite element space. We emphasize that the factor ε^{-2} in the macro error comes from the length of the time interval $T^\varepsilon = \varepsilon^{-2}T$. This error estimate holds if $\mathrm{diam}(\Omega) = \mathcal{O}(1)$. In addition, we provide a new priori error analysis of the FE-HMM-L that is valid for arbitrarily large domains. In particular, we prove the estimate

$$\|u^\varepsilon - u_H\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} \leq C\left(\varepsilon + \left(\frac{h}{\varepsilon^2}\right)^2 + \frac{H^{\ell}}{\varepsilon^2}\right), \tag{7.2}$$

where the norm $\|\cdot\|_W$ is defined as (see (A.4))

$$\|w\|_W = \inf_{\substack{w=w_1+w_2 \\ w_1,w_2 \in \mathrm{W}_{\mathrm{per}}(\Omega)}} \Big\{\|w_1\|_{\mathrm{L}^2(\Omega)} + \|\nabla w_2\|_{\mathrm{L}^2(\Omega)}\Big\} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(\Omega).$$

As in the first estimate, the factor ε^{-2} in the macro error comes from the timescale $T^\varepsilon = \varepsilon^{-2}T$. As the dependence on Ω of the constant in (7.2) is tracked (it is only present in the norms of the data), it can be used to apply the method in pseudoinfinite domains. To prove (7.2), the key is the definition of a new elliptic projection. In particular, this definition allows to avoid the use of the Poincaré inequality needed in the classical proof.

In the second part of the chapter, we provide a method suited for multidimensional problems. The method targets an effective solution in the family defined in Chapter 6, which reads

$$\partial_t^2 \bar{u}(t,x) - \partial_i\big(a_{ij}^0(x)\partial_j \bar{u}(t,x)\big) + \varepsilon L^1 \bar{u}(t,x) + \varepsilon^2 L^2 \bar{u}(t,x) = f(t,x) \quad \text{in } (0,T^\varepsilon]\times\Omega,$$

where the correction operators are

$$L^1 = -\partial_i\big(a_{ij}^{12}(x)\partial_j\,\cdot\big) + b^{10}\partial_t^2, \quad L^2 = \partial_{ij}^2\big(a_{ijkl}^{24}(x)\partial_{kl}^2\,\cdot\big) - \partial_i\big(b_{ij}^{22}(x)\partial_j\partial_t^2\,\cdot\big) - \partial_i\big(a_{ij}^{22}(x)\partial_j\,\cdot\big) + b^{20}\partial_t^2.$$

Let us present the method in a simple setting. Let G_N be a uniform grid of Ω, with N_ν points in the direction ν and denote Δx_ν the size of the grid in the direction ν. In the first step of the method, we approximate the effective tensors of L^1, L^2 at the nodes of the grid G_N. To do so, we use the FEM of degree q to approximate the solutions of the cell problems on a mesh of size h. This process is costly but can be parallelized. In the second step, we use the computed tensors to approximate $\bar{u}$ with a spectral method on the grid G_N. Assuming that the effective solution $\bar{u}$ and its time derivatives belongs to $\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^{s+2}(\Omega))$, we prove the following error estimate

$$\|u^\varepsilon - u_N\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} \le C\left(\varepsilon + \frac{|\Delta x|^s}{\varepsilon^2} + \left(\frac{h^q}{\varepsilon}\right)^2 + \frac{|\Delta x|}{\varepsilon} + \frac{h^q}{\varepsilon} + \frac{h^q|1/\Delta x|}{\varepsilon} + h^q|1/\Delta x|^2\right), \quad (7.3)$$

where $|1/\Delta x|^2 = \sum_\nu 1/\Delta x_\nu^2$. Note that if the effective solution is smooth and if the grid captures the wavelength of the initial data and of the source term, the term $\varepsilon^{-2}|\Delta x|^s$ is smaller than ε. We emphasize that in (7.3) h is the size of the mesh of Y, while in (7.2) h is the size of the mesh of εY. Again, the factors $\varepsilon^{-2}, \varepsilon^{-1}$ in (7.3) come from the timescale $T^\varepsilon = \varepsilon^{-2}T$. As the dependence of (7.3) on Ω is tracked, the estimate can be used in pseudoinfinite domains. We note that this result is the first a priori error analysis of a numerical homogenization method for the approximation of the wave equation in locally periodic media over long time $\mathcal{O}(\varepsilon^{-2})$.

The chapter is organized as follows. In Section 7.1, we present the FE-HMM-L for the long time approximation of the wave equation in one-dimension. In particular, we provide two a priori error analyses of the FE-HMM-L: the first one is valid for small domains and the second one holds for arbitrarily large domains. In Section 7.2, we present the spectral homogenization method for the approximation of the multidimensional wave equation over long time. In particular, we proceed to the a priori error analysis of the method and prove an error estimate that holds in arbitrarily large hypercubes.

7.1 One dimension : finite element heterogeneous multiscale method for long time wave propagation (FE-HMM-L)

In this section, we analyze the FE-HMM-L, a numerical homogenization method designed for the long time approximation of the wave equation in heterogeneous media in one dimension. The FE-HMM-L is a modification of the FE-HMM, defined in Section 3.4. The method was introduced in [10, 9]. The main results are two a priori error analyses over long time. The first one, published in [13] and presented in Section 7.1.3, is valid in small domains. The second one, presented in Section 7.1.4, is new and holds for arbitrarily large domains.

To define the FE-HMM-L in the same settings as the FE-HMM in Section 3.4, we consider a general tensor $a^\varepsilon(x)$. However, we emphasize that all the results are proved under the assumption

that the tensor is locally periodic (see assumption (7.28) below). Let then $a^\varepsilon \in \mathrm{L}^\infty(\Omega)$ be tensor that is elliptic and bounded, i.e., there exists $0 < \lambda \leq \Lambda$ such that

$$\lambda \leq a^\varepsilon(x) \leq \Lambda \qquad \text{for a.e. } x \in \Omega. \tag{7.4}$$

We consider the one-dimensional wave equation: $u^\varepsilon : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ such that

$$\begin{array}{ll} \partial_t^2 u^\varepsilon(t,x) - \partial_x\big(a^\varepsilon(x)\partial_x u^\varepsilon(t,x)\big) = f(t,x) & \text{in } (0,T^\varepsilon]\times\Omega, \\ x \mapsto u^\varepsilon(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,T^\varepsilon], \\ u^\varepsilon(0,x) = g^0(x), \quad \partial_t u^\varepsilon(0,x) = g^1(x) & \text{in } \Omega. \end{array} \tag{7.5}$$

The well-posedness of (7.5) is proved in Section 2.1.1. In particular, if we assume that the data satisfy $g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega)$, $g^1 \in \mathrm{L}^2_0(\Omega)$, $f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{L}^2_0(\Omega))$, then there exists a unique weak solution $u^\varepsilon \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{W}_{\mathrm{per}}(\Omega))$, $\partial_t u^\varepsilon \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2_0(\Omega))$, $\partial_t^2 u^\varepsilon \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{W}^*_{\mathrm{per}}(\Omega))$.

7.1.1 An appropriate effective model for numerical homogenization

To construct a numerical homogenization method for the approximation of the wave equation in heterogeneous media over long time, we need to select an effective model. We discuss here the selection of this model in the family of effective equations defined in Chapter 6.

In Chapter 6, we defined a family of effective equations $\mathcal{E}$ for u^ε in the case of a locally periodic tensor $a^\varepsilon(x) = a\big(x, \frac{x}{\varepsilon}\big)$ (Definition 6.1.2). We recall that the family $\mathcal{E}$ is composed of the equation of the form

$$\begin{array}{ll} \partial_t^2\tilde{u} - \partial_x(a^0\partial_x\tilde{u}) + \varepsilon^2\big(\partial_x^2(a^{24}\partial_x^2\tilde{u}) - \partial_x(b^{22}\partial_x\partial_t^2\tilde{u}) + b^{20}\partial_t^2\tilde{u} - \partial_x(a^{22}\partial_x\tilde{u})\big) = f & \text{in } (0,T^\varepsilon]\times\Omega, \\ x \mapsto \tilde{u}(t,x) \ \ \Omega\text{-periodic} & \text{in } [0,T^\varepsilon], \\ \tilde{u}(0,x) = g^0(x), \quad \partial_t\tilde{u}(0,x) = g^1(x) & \text{in } \Omega, \end{array} \tag{7.6}$$

where $a^0(x)$ is the homogenized tensor and

$$\begin{array}{ll} a^{24}(x) = r a^0(x)^2, & b^{22}(x) = \big\langle \chi(x)^2 \big\rangle_Y + r a^0(x), \\ b^{20} = r \max_{x\in\Omega}\{\partial_x^2 a^0(x)\}, & a^{22}(x) = -r a^0(x)\partial_x^2 a^0(x) + b^{20} a^0(x), \end{array}$$

for some parameter $r \geq 0$ ($\chi(x,\cdot)$ is the first corrector, see (7.7)). In order to define a numerical homogenization method, we first need to select an equation in $\mathcal{E}$. Among the equations in the family, one naturally distinguishes itself. For the choice of parameter $r = 0$, the coefficients a^{24}, a^{22} and b^{20} vanishes and the only remaining correction is $-\varepsilon^2\partial_x(b^{22}\partial_x\partial_t^2\tilde{u})$. Note that in the case of a uniformly periodic tensor, this choice corresponds to the natural choice of parameter $\langle\chi\rangle_Y = 0$ (see Section 4.3.1). The approximation of (7.6) is clearly easier in the case where the fourth order operator vanishes. Furthermore, compared to the homogenized equation the only additional coefficient is $b^{22}(x)$, which depends only on $\chi(x,\cdot)$. We can thus easily modify the FE-HMM, defined in Section 3.4.

Let us define explicitly the selected effective equation. For each $x \in \Omega$, define $\chi(x,\cdot) \in \mathrm{W}_{\mathrm{per}}(Y)$ as the unique solution of the cell problem

$$\big(a(x,\cdot)\partial_y\chi(x,\cdot), \partial_y w\big)_{\mathrm{L}^2(Y)} = -\big(a(x,\cdot), \partial_y w\big)_{\mathrm{L}^2(Y)} \qquad \forall w \in \mathrm{W}_{\mathrm{per}}(Y). \tag{7.7}$$

For $x \in \Omega$, let the tensors a^0, b^2 be defined as

$$a^0(x) = \big\langle a(x,\cdot)(1 + \partial_y\chi(x,\cdot))\big\rangle_Y, \qquad b^2(x) = \big\langle \big(\chi(x,\cdot)\big)^2 \big\rangle_Y. \tag{7.8}$$

We verify that $a^0(x)$ and $b^2(x)$ satisfy

$$\lambda \leq a^0(x) \leq \Lambda, \qquad 0 \leq b^2(x) \leq C \qquad \text{for a.e. } x \in \Omega. \tag{7.9}$$

where λ, Λ are given in (7.4) and C depends on $\|a\|_{\mathcal{C}^0(\bar{\Omega};\mathrm{L}^\infty(Y))}$ and λ (see (6.44)). The effective equation is then: $\bar{u} : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ such that

$$\begin{array}{ll} \partial_t^2 \bar{u} - \partial_x\big(a^0(x)\partial_x \bar{u}\big) - \varepsilon^2 \partial_x\big(b^2(x)\partial_x \partial_t^2 \bar{u}\big) = f & \text{in } (0, T^\varepsilon] \times \Omega, \\ x \mapsto \bar{u}(t, x) \ \ \Omega\text{-periodic} & \text{in } [0, T^\varepsilon], \\ \bar{u}(0, x) = g^0(x), \quad \partial_t \bar{u}(0, x) = g^1(x) & \text{in } \Omega. \end{array} \tag{7.10}$$

The well-posedness of (7.10) is proved in (2.1.2). Define the bilinear forms

$$A^0(v, w) = \big(a^0(x)\partial_x v, \partial_x w\big)_{\mathrm{L}^2(\Omega)}, \qquad B^2(v, w) = \big(b^2(x)\partial_x v, \partial_x w\big)_{\mathrm{L}^2(\Omega)}, \tag{7.11}$$

and the functional space

$$\mathcal{S}(\Omega) = \big\{v \in \mathrm{L}^2_0(\Omega) : \sqrt{b^2}\partial_x v \in \mathrm{L}^2(\Omega)\big\}.$$

Equipped with the inner product and corresponding norm

$$(v, w)_{\mathcal{S}} = (v, w)_{\mathrm{L}^2(\Omega)} + \varepsilon^2 B^2(v, w), \qquad \|v\|_{\mathcal{S}} = \sqrt{(v, v)_{\mathcal{S}}}, \tag{7.12}$$

$\mathcal{S}(\Omega)$ is a Hilbert space. If $g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega)$, $g^1 \in \mathcal{S}(\Omega)$ and $f \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{L}^2_0(\Omega))$, then there exists a unique $\bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{W}_{\mathrm{per}}(\Omega))$ with $\partial_t \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathcal{S}(\Omega))$ and $\partial_t^2 \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathcal{S}(\Omega))$, such that

$$\begin{array}{l} \big(\partial_t^2 \bar{u}(t), v\big)_{\mathcal{S}} + A^0\big(\bar{u}(t), v\big) = \big(f(t), v\big)_{\mathrm{L}^2} \quad \forall v \in \mathrm{W}_{\mathrm{per}}(\Omega) \quad \text{for a.e. } t \in [0, T^\varepsilon], \\ \bar{u}(0) = g^0, \quad \partial_t \bar{u}(0) = g^1. \end{array} \tag{7.13}$$

7.1.2 Definition of the FE-HMM-L

Following [10, 9], we define here the FE-HMM-L. We recall that the definition of the method is done for general tensors a^ε and the results are proved for locally periodic tensors (see assumption (7.28), below).

Let $\mathcal{T}_H$ be a partition of Ω. Denote by H_K the diameter of the element $K \in \mathcal{T}_H$ and define $H = \max_{K \in \mathcal{T}_H} H_K$. For a given $\ell \in \mathbb{N}_{>0}$, the macro finite element space is defined as

$$V_H(\Omega) = \{v_H \in \mathrm{W}_{\mathrm{per}}(\Omega) : v_H|_K \in \mathcal{P}^\ell(K) \ \forall K \in \mathcal{T}_H\}, \tag{7.14}$$

where $\mathcal{P}^\ell(K)$ is the space of polynomials on K of degree at most ℓ. Let $\hat{K}$ be the reference element and for every $K \in \mathcal{T}_H$ let F_K the unique continuous mapping such that $F_K(\hat{K}) = K$ with $\partial_x F_K > 0$. We are given a quadrature formula on $\hat{K}$ by a set of weights and quadrature points $\{\hat{\omega}_j, \hat{x}_j\}_{j=1}^J$. Note that it naturally induces a quadrature formula on K whose weights and quadrature points are given by $\{\omega_{K_j} = \partial_x F_K \hat{\omega}_j, \ x_{K_j} = F_K(\hat{x}_j)\}_{j=1}^J$. The following assumptions are required for the construction of the stiffness matrix to ensure the optimal convergence rate of FEM with numerical quadrature (see Appendix A.3.2 and [34, 33]):

$$\begin{array}{ll} \textit{(i)} & \hat{\omega}_j > 0, \qquad j = 1, \dots, J, \\ \textit{(ii)} & \int_{\hat{K}} \hat{p}(\hat{x}) \, \mathrm{d}\hat{x} = \sum_{j=1}^J \hat{\omega}_j \hat{p}(\hat{x}_j) \quad \forall \hat{p} \in \mathcal{P}^\sigma(\hat{K}), \ \sigma = \max\{2\ell - 2, 1\}. \end{array} \tag{7.15}$$

Furthermore, we assume that the quadrature formula $\{\hat{\omega}'_j, \hat{x}'_j\}_{j=1}^{J'}$, required for the computation of the mass matrix, fulfills the following hypothesis:

$$\textit{(iii)} \quad \sum_{j=1}^{J'} \hat{\omega}'_j |\hat{p}(\hat{x}'_j)|^2 \geq \hat{\lambda}' \|\hat{p}\|^2_{\mathrm{L}^2(\hat{K})} \quad \forall \hat{p} \in \mathcal{P}^\ell(\hat{K}) \text{ for a } \hat{\lambda}' > 0. \tag{7.16}$$

Thanks to (7.16), the quadrature formula $\{\hat{\omega}'_j, \hat{x}'_j\}_{j=1}^{J'}$ defines an inner product (and associated norm) on $V_H(\Omega) \times V_H(\Omega)$ equivalent to the standard L^2 inner product. For every macro element

$K \in \mathcal{T}_H$ and every $j \in \{1, \dots, J\}$, we define around the quadrature point x_{K_j} a sampling domain $K_{\delta j} = x_{K_j} + \delta Y$, where δ is a positive real number such that $\delta \geq \varepsilon$. Each sampling domain $K_{\delta j}$ is discretized in a partition $\mathcal{T}_h$, where $h = \max_{Q \in \mathcal{T}_h} h_Q$ is the maximal diameter of an element $Q \in \mathcal{T}_h$. For a $q \in \mathbb{N}_{>0}$, the micro finite element space is defined as

$$V_h(K_{\delta j}) = \{z_h \in \mathrm{W}_{\mathrm{per}}(K_{\delta j}) : z_h|_Q \in \mathcal{P}^q(Q)\ \forall Q \in \mathcal{T}_h\}. \tag{7.17}$$

Remark 7.1.1. Other finite element spaces for the micro scale are possible. For example, we can use $\mathring{V}_h(K_{\delta j}) = \{z_h \in \mathrm{H}^1_0(K_{\delta j}) : z_h|_Q \in \mathcal{P}^q(Q)\ \forall Q \in \mathcal{T}_h\}$. The formulation of the FE-HMM-L then has to be adapted accordingly, e.g., replacing the function v_h by $(v_h - \langle v_h \rangle_{K_{\delta j}})$ in the FE-HMM-L formulas below.

The FE-HMM-L

Let g^0_H, g^1_H be suitable approximations in $V_H(\Omega)$ of the initial conditions g^0, g^1. The FE-HMM-L is defined as follows: find $u_H : [0, T^\varepsilon] \to V_H(\Omega)$ such that

$$\begin{aligned} &\big(\partial_t^2 u_H(t), v_H\big)_Q + A_H\big(u_H(t), v_H\big) = \big(f(t), v_H\big)_{\mathrm{L}^2} \quad \forall v_H \in V_H(\Omega) \quad \text{for a.e. } t \in [0, T^\varepsilon], \\ &u_H(0) = g^0_H, \quad \partial_t u_H(0) = g^1_H. \end{aligned} \tag{7.18}$$

The bilinear forms are defined for $v_H, w_H \in V_H(\Omega)$ as

$$A_H(v_H, w_H) = \sum_{K \in \mathcal{T}_H} \sum_{j=1}^{J} \frac{\omega_{K_j}}{|K_{\delta j}|} \int_{K_{\delta j}} a^\varepsilon(x) \partial_x v_{h,K_j}(x) \partial_x w_{h,K_j}(x)\, \mathrm{d}x, \tag{7.19}$$

$$\big(v_H, w_H\big)_Q = \big(v_H, w_H\big)_H + \big(v_H, w_H\big)_M, \tag{7.20}$$

$$\big(v_H, w_H\big)_H = \sum_{K \in \mathcal{T}_H} \sum_{j=1}^{J'} \omega'_{K_j} v_H(x'_{K_j}) w_H(x'_{K_j}), \tag{7.21}$$

$$\big(v_H, w_H\big)_M = \sum_{K \in \mathcal{T}_H} \sum_{j=1}^{J} \frac{\omega_{K_j}}{|K_{\delta j}|} \int_{K_{\delta j}} \big(v_{h,K_j} - v^{\mathrm{lin}}_{H,K_j}\big)\big(w_{h,K_j} - w^{\mathrm{lin}}_{H,K_j}\big)(x)\, \mathrm{d}x, \tag{7.22}$$

where the piecewise linear approximation of v_H (resp. w_H) around x_{K_j} is given by

$$v^{\mathrm{lin}}_{H,K_j}(x) = v_H(x_{K_j}) + (x - x_{K_j}) \partial_x v_H(x_{K_j}),$$

and the micro functions v_{h,K_j} for v_H (resp. w_H) are the solutions of the following micro problems in $K_{\delta j}$: find v_{h,K_j} such that $(v_{h,K_j} - v^{\mathrm{lin}}_{H,K_j}) \in V_h(K_{\delta j})$ and

$$\big(a^\varepsilon(x) \partial_x v_{h,K_j}, \partial_x z_h\big)_{\mathrm{L}^2(K_{\delta j})} = 0 \qquad \forall z_h \in V_h(K_{\delta j}). \tag{7.23}$$

Useful reformulation of the FE-HMM-L

To proceed to the a priori analysis, let us reformulate the method. For every $(K, j) \in \mathcal{T}_H \times \{1, \dots, J\}$, define $\psi_{h,K_j} \in V_h(K_{\delta j})$ as the solution of the cell problem in the sampling domain $K_{\delta j}$:

$$\big(a^\varepsilon(x) \partial_x \psi_{h,K_j}, \partial_x z_h\big)_{\mathrm{L}^2(K_{\delta j})} = -\big(a^\varepsilon(x), \partial_x z_h\big)_{\mathrm{L}^2(K_{\delta j})} \qquad \forall z_h \in V_h(K_{\delta j}), \tag{7.24}$$

and define the approximated tensors a^0_K and b^2_K at the quadrature point x_{K_j} as

$$a^0_K(x_{K_j}) = \big\langle a^\varepsilon(x)\big(1 + \partial_x \psi_{h,K_j}\big)\big\rangle_{K_{\delta j}}, \quad b^2_K(x_{K_j}) = \varepsilon^{-2}\big\langle \big(\psi_{h,K_j}\big)^2 \big\rangle_{K_{\delta j}}. \tag{7.25}$$

We recall Lemma 3.4.1 from Section 3.4 (originally in [1, 3]).

Lemma 7.1.2. *The bilinear form A_H can be rewritten for $v_H, w_H \in V_H(\Omega)$ as*

$$A_H(v_H, w_H) = \sum_{K\in\mathcal{T}_H} \sum_{j=1}^{J} \omega_{K_j} a_K^0(x_{K_j}) \partial_x v_H(x_{K_j}) \partial_x w_H(x_{K_j}). \tag{7.26}$$

Furthermore, A_H is elliptic and bounded, i.e., for any $v_H, w_H \in V_H(\Omega)$,

$$A_H(v_H, v_H) \geq \lambda \|\partial_x v_H\|^2_{\mathrm{L}^2(\Omega)}, \qquad A_H(v_H, w_H) \leq \Lambda^2/\lambda \|\partial_x v_H\|_{\mathrm{L}^2(\Omega)} \|\partial_x w_H\|_{\mathrm{L}^2(\Omega)}.$$

Similarly, we prove the following result (originally in [9, 10]).

Lemma 7.1.3. *The product $(\cdot,\cdot)_M$ can be rewritten as $(v_H, w_H)_M = \varepsilon^2 B_H(v_H, w_H)$, where the bilinear form B_H is defined as*

$$B_H(v_H, w_H) = \sum_{K\in\mathcal{T}_H} \sum_{j=1}^{J} \omega_{K_j} b_K^2(x_{K_j}) \partial_x v_H(x_{K_j}) \partial_x w_H(x_{K_j}),$$

and is positive semidefinite and bounded, i.e., for any $v_H, w_H \in V_H(\Omega)$,

$$B_H(v_H, v_H) \geq 0, \qquad B_H(v_H, w_H) \leq C \|\partial_x v_H\|_{\mathrm{L}^2(\Omega)} \|\partial_x w_H\|_{\mathrm{L}^2(\Omega)}, \tag{7.27}$$

where C is a constant independent of H.

Proof. First, note that by definition, B_H satisfies $B_H(v_H, v_H) \geq 0$ for any $v_H \in V_H(\Omega)$. Let us then prove that $(v_H, w_H)_M = \varepsilon^2 B_H(v_H, w_H)$ and that B_H is bounded. As in Lemma 3.4.1, thanks the uniqueness of the solution of problem (7.24), we verify that the micro function v_{h,K_j} satisfies $v_{h,K_j} = v^{\mathrm{lin}}_{H,K_j} + \psi_{h,K_j} \partial_x v^{\mathrm{lin}}_{H,K_j}$ (and similarly for w_{h,K_j}). Plugging this equalities in (7.22), we obtain $(v_H, w_H)_M = \varepsilon^2 B_H(v_H, w_H)$. As $\|\psi_{h,K_j}\|_{\mathrm{L}^2}$ is bounded, so is B_H and the proof of the lemma is complete. □

Remark 7.1.4. We emphasize that although $b_K^2(x_{K_j})$ depends on ε, the product $(\cdot,\cdot)_M$ does not. In fact, ψ_{h,K_j} is an approximation of $\varepsilon\chi(x_{K_j}, \frac{\cdot}{\varepsilon})$, where χ is defined in (7.7) (see the proof of Lemma 7.1.14 for details). Hence, assuming $\delta = \varepsilon$, we have via the change of variable $x = \varepsilon y$

$$b_K^2(x_{K_j}) = \varepsilon^{-2} |K_{\delta j}|^{-1} \int_{K_{\delta j}} \big(\psi_{h,K_j}(x)\big)^2 \,\mathrm{d}x \approx |Y|^{-1} \int_Y \big(\chi(x_{K_j}, y)\big)^2 \,\mathrm{d}y = b^2(x_{K_j}),$$

where $b^2(x)$ is defined in (7.8). Consequently, B_H is obtained from B^2 by approximating the integral with numerical quadrature and approximating $b^2(x_{K_j})$ with $b_K^2(x_{K_j})$.

Remark 7.1.5. As a consequence of Lemmas 7.1.2 and 7.1.3, problem (7.18) is equivalent to a regular second order ordinary differential equation. Therefore, existence and uniqueness of a solution of (7.18) is given by classical theory for ordinary differential equations [38] and the FE-HMM-L is well-posed. Furthermore, the solution u_H satisfies the regularity $u_H \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{W}_{\mathrm{per}}(\Omega)), \partial_t u_H \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{L}^2_0(\Omega))$.

7.1.3 Long time a priori error analysis of the FE-HMM-L in small domains

In this section, we present the long time a priori error analysis of the FE-HMM-L in small domains, which was published in [13]. In particular, we prove error estimates in the $\mathrm{L}^\infty(\mathrm{L}^2)$

and $\mathrm{L}^\infty(\mathrm{H}^1)$ norms, valid for small domains Ω such that $\mathrm{diam}(\Omega) = \mathcal{O}(1)$. Indeed, the classical techniques for such analysis lead to error estimates with a constant depending on the domain (as done in [44, 21, 22], see Section 2.2). For the H^1 norm, the constant depends on the Poincaré constant, while for the L^2 norm, it depends in addition on the constant coming from elliptic regularity. In the next section, we prove an error estimate where the constant is independent of the size of Ω.

We make the assumption that the tensor is locally periodic and collocated in the slow variable, i.e.,

$$a^\varepsilon(x) = a\big(x_{K_j}, \tfrac{x}{\varepsilon}\big) \quad \text{for a.e. } x \in K_{\delta j} \quad \forall (K, j) \in \mathcal{T}_H \times \{1, \dots, J\}. \tag{7.28}$$

Provided $\delta/\varepsilon \in \mathbb{N}_{>0}$, this assumption ensures that the micro problems (7.24) match the cell problems for χ, i.e., $\psi_{h,K_j} = \varepsilon\chi\big(x_{K_j}, \frac{\cdot}{\varepsilon}\big)$ (see Lemma 7.1.14). At short time, if this assumption is not satisfied, the error is known to suffer only a small additional term of order ε (see Section 3.4.2). However, the impact of this error on timescales $\mathcal{O}(\varepsilon^{-2})$ is not conceivable. Therefore, if the tensor is not locally periodic, or its explicit form not known, the long time approximation provided by the FE-HMM-L might be of poor accuracy.

Let us first comment on our analysis. Let $\bar{u}_H$ be the FE approximation in $V_H(\Omega)$ of $\bar{u}$, defined in Section 2.2. Theorem 7.1.6 provides a priori error estimates for $\mathrm{e}^{\mathrm{FE}} = \|\bar{u} - \bar{u}_H\|$ in the H^1 and L^2 norms. In our analysis of the FE-HMM-L, the purpose is not to analyze e^{FE} but to estimate the error generated by the upscaling procedure $\mathrm{e}^{\mathrm{HMM}} = \|\bar{u}_H - u_H\|$. However, in order to formulation regularity requirements on $\bar{u}$ (and not on $\bar{u}_H$), we have to proceed to the full analysis and estimate $\|\bar{u} - u_H\|$.

Recall that ℓ is the degree of the macro finite element space $V_H(\Omega)$. Let I_H be an interpolation operator such that for $v \in \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^{s+1}(\Omega)$, where $1 \le s \le \ell$,

$$\Big(\sum_{K \in \mathcal{T}_H} \|v - I_H v\|^2_{\mathrm{H}^m(K)} \Big)^{1/2} \le C H^{s+1-m} \|v\|_{\mathrm{H}^{s+1}(\Omega)}, \quad 0 \le m \le s+1, \tag{7.29}$$

where C is a constant independent of H and v. For example, I_H can be the nodal interpolation operator introduced in Section A.3 (see also [33]). We recall the a priori error estimates for the FEM provided in Theorem 2.2.1, Section 2.2:

Theorem 7.1.6. *Assume that the quadrature formulas satisfy the assumptions* (7.15) *and* (7.16). *Let $\bar{u}$ denote the solution of* (7.13) *and let $\bar{u}_H$ be its FE approximation in $V_H(\Omega)$.*

i) Assume that $a^0, b^2 \in \mathrm{W}^{\ell,\infty}(\Omega)$ and $\partial_t^k \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^{\ell+1}(\Omega))$ for $0 \le k \le 4$. Then the error satisfies $\|\bar{u} - \bar{u}_H\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^1(\Omega))} \le \mathrm{e}^{\mathrm{FE}}_{\mathrm{H}^1}$, where

$$\begin{aligned} \mathrm{e}^{\mathrm{FE}}_{\mathrm{H}^1} = {} & C_1\big(\|g^1 - g^1_H\|_{\mathrm{H}^1(\Omega)} + \|g^0 - g^0_H\|_{\mathrm{H}^1(\Omega)}\big) \\ & + C_2\big(H^\ell + T^\varepsilon H^{\ell+1} + T^\varepsilon(1+\varepsilon)\varepsilon H^\ell\big) \textstyle\sum_{k=0}^4 \|\partial_t^k \bar{u}\|_{\mathrm{L}^\infty(\mathrm{H}^{\ell+1})}, \end{aligned}$$

where C_1, C_2 are independent of H and ε but depend on Ω.

ii) Assume that $a^0 \in \mathrm{W}^{\ell+1,\infty}(\Omega)$, $b^2 \in \mathrm{W}^{\ell,\infty}(\Omega)$ and $\partial_t^k \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^{\ell+1}(\Omega))$ for $0 \le k \le 3$. Then the error satisfies $\|\bar{u} - \bar{u}_H\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} \le \mathrm{e}^{\mathrm{FE}}_{\mathrm{L}^2}$, where

$$\begin{aligned} \mathrm{e}^{\mathrm{FE}}_{\mathrm{L}^2} = {} & C_1\big(\|g^0 - g^0_H\|_{\mathrm{L}^2(\Omega)} + \varepsilon\|g^0 - g^0_H\|_{\mathrm{H}^1(\Omega)} + \|g^1 - g^1_H\|_{\mathrm{L}^2(\Omega)} + \varepsilon\|g^1 - g^1_H\|_{\mathrm{H}^1(\Omega)}\big) \\ & + C_2(1 + T^\varepsilon)\big(H^{\ell+1} + \varepsilon H^\ell\big) \textstyle\sum_{k=0}^3 \|\partial_t^k \bar{u}\|_{\mathrm{L}^\infty(\mathrm{H}^{\ell+1})}, \end{aligned}$$

where C_1, C_2 are independent of H and ε but depend on Ω.

The two following theorems provide a priori error estimates for the FE-HMM-L. We start with an $\mathrm{L}^\infty(\mathrm{H}^1)$ estimate.

Theorem 7.1.7. *Assume that δ satisfies $\delta/\varepsilon \in \mathbb{N}_{>0}$, that the micro mesh size is $h \leq \varepsilon$ and that the degree of the micro finite element space is $q = 1$. Furthermore, assume that the tensor is locally periodic and collocated in the slow variable (assumption (7.28)). Finally, assume that $a \in \mathcal{C}^\ell(\bar{\Omega}; \mathrm{L}^\infty(Y)) \cap \mathcal{C}^0(\bar{\Omega}; \mathrm{W}^{1,\infty}(Y))$ and $\partial_t^k \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^{\ell+1}(\Omega))$ for $0 \leq k \leq 4$. Then the error $e = \bar{u} - u_H$ satisfies the estimate*

$$\|\partial_t e\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} + \|e\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^1(\Omega))} \leq C\left(\frac{h}{\varepsilon^2}\right)^2 + \mathrm{e}^{\mathrm{FE}}_{\mathrm{H}^1}, \tag{7.30}$$

where C is independent of H, h, ε, and δ, but depends on Ω, and $\mathrm{e}^{\mathrm{FE}}_{\mathrm{H}^1}$ is the standard FEM error estimate given in Theorem 7.1.6.

The next result is an $\mathrm{L}^\infty(\mathrm{L}^2)$ estimate.

Theorem 7.1.8. *As in Theorem 7.1.7, assume that $h \leq \varepsilon$, $q = 1$, that a^ε satisfies (7.28) where $a \in \mathcal{C}^0(\bar{\Omega}; \mathrm{W}^{1,\infty}(Y)) \cap \mathcal{C}^{\ell+1}(\bar{\Omega}; \mathrm{L}^\infty(Y))$. Furthermore, assume that $\partial_t^k \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^{\ell+1}(\Omega))$ for $0 \leq k \leq 3$. Then the error $e = \bar{u} - u_H$ satisfies the estimate*

$$\|e\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} \leq C\left(\frac{h}{\varepsilon^2}\right)^2 + \mathrm{e}^{\mathrm{FE}}_{\mathrm{L}^2}, \tag{7.31}$$

where C is independent of H, h, ε, and δ, but depends on Ω, and $\mathrm{e}^{\mathrm{FE}}_{\mathrm{L}^2}$ is the standard FEM error estimate given in Theorem 7.1.6.

Next, we combine (7.31) with Theorem 6.1.1. We obtain an estimate of the error between the oscillatory wave u^ε and the solution of the FE-HMM-L in the $\mathrm{L}^\infty(\mathrm{L}^2)$ norm, in the case of small domain Ω such that $\mathrm{diam}(\Omega) = \mathcal{O}(1)$.

Corollary 7.1.9. *Assume that Ω is a union of cells of volume $\varepsilon|Y|$ (assumption (4.25)), that the tensor is collocated in the slow variable (assumption (7.28)) and satisfies the regularity $a \in \mathcal{C}^1(\bar{\Omega}; \mathrm{W}^{1,\infty}(Y)) \cap \mathcal{C}^4(\bar{\Omega}; \mathrm{L}^\infty(Y))$. Also, assume that $g^0_H = I_H g^0$, $g^1_H = I_H g^1$, and let the settings of the FE-HMM-L be such that $\delta/\varepsilon \in \mathbb{N}_{>0}$, $h \leq \varepsilon$, $q = 1$ and $\ell = 1$. Finally assume that the following regularity holds:*

$$g^0 \in \mathrm{H}^4(\Omega),\ g^1 \in \mathrm{H}^3(\Omega),\ f \in \mathrm{L}^2(0, T^\varepsilon; \mathrm{H}^2(\Omega)),\ \partial_t^k \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^{5-k}(\Omega)),\ 0 \leq k \leq 3.$$

Then we have the following estimate:

$$\|u^\varepsilon - u_H\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} \leq C\left(\varepsilon + \left(\frac{h}{\varepsilon^2}\right)^2 + \frac{H^2}{\varepsilon^2} + \frac{H}{\varepsilon}\right),$$

where C independent of H, h, ε, and δ but depends on Ω.

Remark 7.1.10. Under suitable regularity of the initial conditions, the result of Corollary 7.1.9 can be generalized to obtain the error estimate

$$\|u^\varepsilon - u_H\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} \leq C\left(\varepsilon + \left(\frac{h}{\varepsilon^2}\right)^2 + \frac{H^{\ell+1}}{\varepsilon^2} + \frac{H^\ell}{\varepsilon}\right). \tag{7.32}$$

To appreciate the benefit of the FE-HMM-L, we compare its cost with the cost of a fine scale FEM applied to approximate u^ε. First, let us give the classical a priori error estimate for a fine scale FEM on a mesh of size h (see [21] and also the discussion in [12] or in Section 3.1):

$$\|u^\varepsilon - u_h\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} \leq C\frac{h}{\varepsilon}\Big(\|u^\varepsilon\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^1(\Omega))} + \|\partial_t u^\varepsilon\|_{\mathrm{L}^1(0,T^\varepsilon;\mathrm{H}^1(\Omega))}\Big) \leq C\frac{h}{\varepsilon^3}, \tag{7.33}$$

where a factor $\mathcal{O}(\varepsilon^{-2})$ comes from the time interval $T^\varepsilon = T/\varepsilon^2$ and a factor $\mathcal{O}(\varepsilon^{-1})$ comes from the elliptic projection. Here, we have used well prepared initial data in order to bound $\|\partial_t u^\varepsilon\|_{\mathrm{L}^1(\mathrm{H}^1)}$, for otherwise the standard FEM estimate would read $\|u^\varepsilon - u_h\|_{\mathrm{L}^\infty(\mathrm{L}^2(\Omega))} \leq Ch/\varepsilon^4$ (see [12] for details). We now fix an order of tolerance τ for the error and compute the cost of each method, based on the corresponding error estimate. For the sake of simplicity denote as $\mathrm{cost}(\Delta t, N)$ the cost per time-step of the time integration of a second order ODE of dimension N. Based on (7.33), the cost of the fine scale FEM is $\mathrm{cost}(\Delta t, \varepsilon^{-3}\tau^{-1})$. For the FE-HMM-L with linear elements ($\ell = 1$), from (7.32) we set $H = \varepsilon\tau$, $h/\varepsilon = \varepsilon\tau^{1/2}$. The cost of resolution of the micro problems is then $H^{-1}(\varepsilon/h) = \varepsilon^{-2}\tau^{-3/2}$, and the cost of the time integration is $\mathrm{cost}(\Delta t, \varepsilon^{-1}\tau^{-1})$. As we are integrating over a long time interval, note that the resolution of the micro problems is a negligible preprocessing step. We see that a significant reduction in computational cost is achieved by the FE-HMM-L for long time interval $\mathcal{O}(\varepsilon^{-2})$. Note also that in the FE-HMM-L, higher degree in the macro finite element ($\ell \geq 1$) is allowed in (7.32), obtaining then $H = \min\{(\varepsilon^2\tau)^{1/(\ell+1)}, (\varepsilon\tau)^{1/\ell}\}$. In that case, the cost of the preprocessing step is larger as the number of micro problems increases (because a higher order macro quadrature formula is required). For the fine scale FEM, using higher degree is not possible, as the error estimate involves higher space derivatives of u^ε and $\partial_t u^\varepsilon$, which brings ε^{-1} factors in the estimate. Finally, we notice that the cost of the time integration is also significantly smaller with the FE-HMM-L. If we use an explicit method (such as the leap-frog scheme), the stability constraint reads $\Delta t \sim \varepsilon^3$ for the fine scale integrator whereas it is only $\Delta t \sim \varepsilon$ for the FE-HMM-L. Of course, this could be avoided by using an implicit solver, but then the cost of solving the linear system is also significantly higher for the full fine scale solver due to the much larger system of ODEs.

Proof of the a priori error estimates

The proofs of Theorems 7.1.7 and 7.1.8 are divided into four lemmas. We split the error $\bar{u} - u_H$ as

$$\bar{u} - u_H = (\bar{u} - \pi_H\bar{u}) - (u_H - \pi_H\bar{u}) = \eta - \zeta_H, \tag{7.34}$$

where $\pi_H\bar{u}$ is the elliptic projection defined below. We first provide a priori estimates for η and ζ_H in Lemmas 7.1.11, 7.1.12, and 7.1.13. We then quantify the error made at the micro level by the FEM and the error coming from the upscaling procedure of the FE-HMM-L in Lemma 7.1.14.

In the whole proof, c and C represent generic constants independent of H, h, ε, δ, $\bar{u}$, e_{a^0}, e_{b^2} (defined below). Hypothesis (7.16) ensures that $\|v_H\|_H = (v_H, v_H)_H^{1/2}$ is a norm on $V_H(\Omega)$, equivalent to the L^2 norm independently of H. Hence, using the result of Lemma 7.1.3, the norm $\|v_H\|_Q = (v_H, v_H)_Q^{1/2}$ (where $(\cdot,\cdot)_Q$ is defined in (7.20)) satisfies

$$c\|v_H\|_{\mathrm{L}^2} \leq \|v_H\|_Q \leq C\big(\|v_H\|_{\mathrm{L}^2} + \varepsilon\|v_H\|_{\mathrm{H}^1}\big). \tag{7.35}$$

Let us introduce the following bilinear forms for $v_H, w_H \in V_H(\Omega)$:

$$A_H^0(v_H, w_H) = \sum_{K\in\mathcal{T}_H}\sum_{j=1}^{J}\omega_{K_j}a^0(x_{K_j})\partial_x v_H(x_{K_j})\partial_x w_H(x_{K_j}),$$
$$B_H^2(v_H, w_H) = \sum_{K\in\mathcal{T}_H}\sum_{j=1}^{J}\omega_{K_j}b^2(x_{K_j})\partial_x v_H(x_{K_j})\partial_x w_H(x_{K_j}),$$

where $a^0(x)$, $b^2(x)$ are the exact tensors defined in (7.8). The HMM errors are defined as

$$\mathrm{e}_{a^0} = \sup_{K\in\mathcal{T}_H, 1\le j\le J} |a^0(x_{K_j}) - a^0_K(x_{K_j})|, \qquad \mathrm{e}_{b^2} = \sup_{K\in\mathcal{T}_H, 1\le j\le J} \varepsilon^2 |b^2(x_{K_j}) - b^2_K(x_{K_j})|,$$

where $a^0_K(x_{K_j})$, $b^2_K(x_{K_j})$ are defined in (7.25). Using Lemmas 7.1.2 and 7.1.3, we verify that for any $v_H, w_H \in V_H(\Omega)$,

$$\begin{aligned} |A_H(v_H,w_H) - A^0_H(v_H,w_H)| &\le \mathrm{e}_{a^0} \|\partial_x v_H\|_{\mathrm{L}^2} \|\partial_x w_H\|_{\mathrm{L}^2}, \\ \varepsilon^2 |B_H(v_H,w_H) - B^2_H(v_H,w_H)| &\le \mathrm{e}_{b^2} \|\partial_x v_H\|_{\mathrm{L}^2} \|\partial_x w_H\|_{\mathrm{L}^2}. \end{aligned} \tag{7.36}$$

Finally, the broken norm on $V_H(\Omega)$ is defined as $\|v_H\|_{\bar{\mathrm{H}}^k(\Omega)} = \big(\sum_{K\in\mathcal{T}_H} \|v_H\|^2_{\mathrm{H}^k(K)}\big)^{1/2}$. Thanks to assumptions (7.15) and (7.16) and provided sufficient regularity of a^0, b^2, we have the following estimates for the numerical integration errors (see [33, 82] and Theorems A.3.6 and A.3.9):

$$\begin{aligned} |A^0(v_H,w_H) - A^0_H(v_H,w_H)| &\le CH^{\ell+\mu} \|a^0\|_{\mathrm{W}^{\ell+\mu,\infty}} \|v_H\|_{\bar{\mathrm{H}}^{\ell+1}} \|w_H\|_{\bar{\mathrm{H}}^{1+\mu}}, \\ |A^0(v_H,w_H) - A^0_H(v_H,w_H)| &\le CH \|a^0\|_{\mathrm{W}^{1,\infty}} \|v_H\|_{\mathrm{H}^1} \|w_H\|_{\mathrm{H}^1}, \\ |B^2(v_H,w_H) - B^2_H(v_H,w_H)| &\le CH^{\ell} \|b^2\|_{\mathrm{W}^{\ell,\infty}} \|v_H\|_{\bar{\mathrm{H}}^{\ell+1}} \|w_H\|_{\bar{\mathrm{H}}^{1}}, \\ |(v_H,w_H)_{\mathrm{L}^2} - (v_H,w_H)_H| &\le CH^{\ell+\mu} \|v_H\|_{\bar{\mathrm{H}}^{\ell+1}} \|w_H\|_{\bar{\mathrm{H}}^{1+\mu}}, \end{aligned} \tag{7.37}$$

for any $v_H, w_H \in V_H(\Omega)$ and $\mu = 0, 1$ (A^0, B^2 are defined in (7.11)). Note that in Theorem 7.1.7, as we assume $a \in \mathcal{C}^\ell(\bar{\Omega}; \mathrm{L}^\infty(Y))$, a^0 and b^2 satisfy the regularity $a^0, b^2 \in \mathcal{C}^\ell(\bar{\Omega})$ (see (6.44)). Similarly, in Theorem 7.31, we have $a^0, b^2 \in \mathcal{C}^{\ell+1}(\bar{\Omega})$. In the proof, we need the following estimates: for $v \in \mathrm{H}^{\ell+1}(\Omega) \cap \mathrm{W}_{\mathrm{per}}(\Omega)$ and $w_H \in V_H(\Omega)$, $\mu = 0, 1$,

$$\begin{aligned} |A^0(v,w_H) - A_H(I_H v, w_H)| &\le C\big(\mathrm{e}_{a^0} \|v\|_{\mathrm{H}^1} + H^\ell \|v\|_{\mathrm{H}^{\ell+1}}\big) \|w_H\|_{\bar{\mathrm{H}}^1}, \\ |(v,w_H)_{\mathcal{S}} - (I_H v, w_H)_Q| &\le C\big(\mathrm{e}_{b^2} \|v\|_{\mathrm{H}^1} + (H^{\ell+\mu} + \varepsilon^2 H^\ell) \|v\|_{\mathrm{H}^{\ell+1}}\big) \|w_H\|_{\bar{\mathrm{H}}^{1+\mu}}, \end{aligned} \tag{7.38}$$

where $(\cdot,\cdot)_{\mathcal{S}}$ is defined in (7.12). They are obtained by combining the triangle inequality, (7.29), (7.36), and (7.37).

Define the elliptic projection $\pi_H \bar{u} : [0, T^\varepsilon] \to V_H(\Omega)$, solution of

$$A_H\big(\pi_H \bar{u}(t), v_H\big) = \big(f(t), v_H\big)_{\mathrm{L}^2} - \big(I_H \partial_t^2 \bar{u}(t), v_H\big)_Q \quad \forall v_H \in V_H(\Omega) \quad \text{ for a.e. } t \in [0, T^\varepsilon]. \tag{7.39}$$

As A_H is elliptic and bounded $\pi_H \bar{u}(t)$ exists and is unique for a.e. $t \in [0, T^\varepsilon]$. Furthermore, using (7.13) we have $\big(f(t), v_H\big)_{\mathrm{L}^2} = A^0\big(\bar{u}(t), v_H\big) + \big(\partial_t^2 \bar{u}(t), v_H\big)_{\mathcal{S}}$ and we obtain the estimate

$$\|\pi_H \bar{u}(t)\|_{\mathrm{H}^1} \le C\big(\|\bar{u}(t)\|_{\mathrm{H}^1} + \|\partial_t^2 \bar{u}(t)\|_{\mathrm{H}^1}\big) \qquad \text{for a.e. } t \in [0, T^\varepsilon]. \tag{7.40}$$

Hence, provided $\partial_t^2 \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^1(\Omega))$, $\pi_H \bar{u}$ satisfies the regularity $\pi_H \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^1(\Omega))$.

We prove the following result for $\eta = \bar{u} - \pi_H \bar{u}$.

Lemma 7.1.11. *Assume that for $1 \le p \le \infty$, $\partial_t^k \bar{u}, \partial_t^{k+2} \bar{u} \in \mathrm{L}^p(0, T^\varepsilon; \mathrm{H}^{\ell+1}(\Omega))$ for $k \ge 0$. Then $\partial_t^k \pi_H \bar{u} \in \mathrm{L}^p(0, T^\varepsilon; \mathrm{H}^1(\Omega))$ and, provided $a^0, b^2 \in \mathrm{W}^{\ell,\infty}(\Omega)$, the following estimate holds for $\eta = \bar{u} - \pi_H \bar{u}$:*

$$\begin{aligned} \|I_H \partial_t^k \eta\|_{\mathrm{L}^p(\mathrm{H}^1)} + \|\partial_t^k \eta\|_{\mathrm{L}^p(\mathrm{H}^1)} \le\ & C\Big(\big(\mathrm{e}_{a^0} + \mathrm{e}_{b^2}\big)\big(\|\partial_t^k \bar{u}\|_{\mathrm{L}^p(\mathrm{H}^1)} + \|\partial_t^{k+2} \bar{u}\|_{\mathrm{L}^p(\mathrm{H}^1)}\big) \\ & + H^\ell\big(\|\partial_t^k \bar{u}\|_{\mathrm{L}^p(\mathrm{H}^{\ell+1})} + \|\partial_t^{k+2} \bar{u}\|_{\mathrm{L}^p(\mathrm{H}^{\ell+1})}\big)\Big). \end{aligned} \tag{7.41}$$

If in addition we assume $a^0 \in \mathrm{W}^{\ell+1,\infty}(\Omega)$*, then*

$$\begin{aligned}\|I_H\partial_t^k\eta\|_{\mathrm{L}^p(\mathrm{L}^2)} + \|\partial_t^k\eta\|_{\mathrm{L}^p(\mathrm{L}^2)} \le C\Big((1+\mathrm{e}_{a^0})(\mathrm{e}_{a^0}+\mathrm{e}_{b^2}) + H^{\ell+1} + \varepsilon^2 H^\ell\Big)\\ \times\big(\|\partial_t^k\bar u\|_{\mathrm{L}^p(\mathrm{H}^{\ell+1})} + \|\partial_t^{k+2}\bar u\|_{\mathrm{L}^p(\mathrm{H}^{\ell+1})}\Big).\end{aligned} \tag{7.42}$$

Proof. First, as the forms A^0, $(.,.)_{\mathcal{S}}$, A_H, and $(.,.)_Q$ are time independent, the time differentiation of (7.39) and (7.13) yields, similarly to (7.40), the estimate

$$\|\partial_t^k\pi_H\bar u(t)\|_{\mathrm{H}^1} \le C\big(\|\partial_t^k\bar u(t)\|_{\mathrm{H}^1} + \|\partial_t^{k+2}\bar u(t)\|_{\mathrm{H}^1}\big) \qquad \text{for a.e. } t\in[0,T^\varepsilon].$$

Hence in view of the assumption on $\partial_t^k\bar u, \partial_t^{k+2}\bar u$ we obtain $\partial_t^k\omega_H \in \mathrm{L}^p(0,T^\varepsilon;\mathrm{H}^1(\Omega))$. Second, we prove estimates (7.41) and (7.42) for $k=0$. The proof for $k>0$ is obtained in the same way by differentiating (7.39) and (7.13). Using (7.39) and (7.13) we have almost everywhere in $[0,T^\varepsilon]$,

$$A_H\big(I_H\eta, v_H\big) = A_H\big(I_H\bar u, v_H\big) - A^0\big(\bar u, v_H\big) + \big(\partial_t^2\bar u, v_H\big)_{\mathcal{S}} - \big(I_H\partial_t^2\bar u, v_H\big)_Q.$$

We make use of (7.38) to obtain for a.e $t\in[0,T^\varepsilon]$,

$$A_H\big(I_H\eta(t), v_H\big) \le C\Big(\big(\mathrm{e}_{a^0}+\mathrm{e}_{b^2}\big)\textstyle\sum_{k=0,2}\|\partial_t^k\bar u(t)\|_{\mathrm{H}^1} + H^\ell\sum_{k=0,2}\|\partial_t^k\bar u(t)\|_{\mathrm{H}^{\ell+1}}\Big)\|v_H\|_{\mathrm{H}^1}.$$

Letting now $v_H = I_H\eta(t)$, using the ellipticity of A_H and taking the L^p norm with respect to t, we obtain

$$\|I_H\eta\|_{\mathrm{L}^p(\mathrm{H}^1)} \le C\Big(\big(\mathrm{e}_{a^0}+\mathrm{e}_{b^2}\big)\textstyle\sum_{k=0,2}\|\partial_t^k\bar u\|_{\mathrm{L}^p(\mathrm{H}^1)}\big) + H^\ell\sum_{k=0,2}\|\partial_t^k\bar u\|_{\mathrm{L}^p(\mathrm{H}^{\ell+1})}\Big).$$

Note that $\eta = \bar u - I_H\bar u + I_H\eta$ and $\|\bar u - I_H\bar u\|_{\mathrm{L}^p(\mathrm{H}^1)} \le CH^\ell\|\bar u\|_{\mathrm{L}^p(\mathrm{H}^{\ell+1})}$ and we have proved estimate (7.41) for $k=0$. To prove (7.42), we use a standard Aubin–Nitsche argument. For a.e. $t\in[0,T^\varepsilon]$, note that $\|\eta(t)\|_{\mathrm{L}^2} = \sup_{g\in\mathrm{L}^2(\Omega)}\|g\|_{\mathrm{L}^2}^{-1}\big|\big(\eta(t),g\big)_{\mathrm{L}^2}\big|$. Let now $g\in\mathrm{L}^2(\Omega)$ and define φ_g as the solution of the elliptic problem $A^0(v,\varphi_g) = (g,v)_{\mathrm{L}^2}\ \ \forall v\in\mathrm{W}_{\mathrm{per}}(\Omega)$. The regularity of a^0 and the polygonal domain ensure that $\|\varphi_g\|_{\mathrm{H}^2}\le C\|g\|_{\mathrm{L}^2}$ (see [71]). Using (7.39) and (7.13), we verify that

$$\begin{aligned}A^0\big(\eta(t),\varphi_g\big) = A^0\big(\eta(t),\varphi_g - v_H\big) + \big(I_H\partial_t^2\bar u(t), v_H\big)_Q - \big(\partial_t^2\bar u(t), v_H\big)_{\mathcal{S}}\\ + A_H\big(\pi_H\bar u(t), v_H\big) - A^0\big(\pi_H\bar u(t), v_H\big)\end{aligned} \tag{7.43}$$

for any $v_H\in V_H(\Omega)$ and a.e. $t\in[0,T^\varepsilon]$. Note that we can rewrite the last two terms as

$$\begin{aligned}A_H\big(\pi_H\bar u(t), v_H\big) - A^0(\pi_H\bar u(t), v_H) = A^0\big(I_H\eta(t), v_H\big) - A_H\big(I_H\eta(t), v_H\big)\\ + A_H\big(I_H\bar u(t), v_H\big) - A^0\big(I_H\bar u(t), v_H\big).\end{aligned}$$

Hence, using the triangle inequality and (7.29), (7.36), and (7.37), we have

$$\big|A_H\big(\pi_H\bar u(t), v_H\big) - A^0(\pi_H\bar u(t), v_H)\big| \le C\big((\mathrm{e}_{a^0}+H)\|I_H\eta(t)\|_{\mathrm{H}^1} + (\mathrm{e}_{a^0}+H^{\ell+1})\|\bar u(t)\|_{\mathrm{H}^{\ell+1}}\big)\|v_H\|_{\bar{\mathrm{H}}^2}.$$

Now, as $(\eta(t),g)_{\mathrm{L}^2} = A^0(\eta(t),\varphi_g)$, from (7.43) with $v_H = I_H\varphi_g$, we use estimates (7.29) and (7.38) to obtain for a.e. t

$$\begin{aligned}\big|\big(\eta(t),g\big)_{\mathrm{L}^2}\big| \le C\Big(H\|\eta(t)\|_{\mathrm{H}^1} + (\mathrm{e}_{a^0}+H)\|I_H\eta(t)\|_{\mathrm{H}^1} + (\mathrm{e}_{a^0}+H^{\ell+1})\|\bar u(t)\|_{\mathrm{H}^{\ell+1}}\\ + \mathrm{e}_{b^2}\|\partial_t^2\bar u(t)\|_{\mathrm{H}^1} + (H^{\ell+1}+\varepsilon^2H^\ell)\|\partial_t^2\bar u(t)\|_{\mathrm{H}^{\ell+1}}\Big)\|\varphi_g\|_{\mathrm{H}^2}.\end{aligned}$$

Hence, recalling that $\|\eta(t)\|_{\mathrm{L}^2} \leq \|g\|_{\mathrm{L}^2}^{-1}\big|\big(\eta(t),g\big)_{\mathrm{L}^2}\big|$ and $\|\varphi_g\|_{\mathrm{H}^2} \leq C\|g\|_{\mathrm{L}^2}$, we obtain for a.e. $t \in [0,T^\varepsilon]$

$$\|\eta(t)\|_{\mathrm{L}^2} \leq C\Big((1+\mathrm{e}_{a^0}+H)\big(\mathrm{e}_{a^0}+\mathrm{e}_{b^2}\big)+\mathrm{e}_{a^0}H^\ell+H^{\ell+1}+\varepsilon^2H^\ell\Big)\textstyle\sum_{k=0,2}\|\partial_t^k\bar{u}(t)\|_{\mathrm{H}^{\ell+1}}.$$

Taking the L^p norm with respect to t and using estimate (7.41) brings

$$\|\eta\|_{\mathrm{L}^p(\mathrm{L}^2)} \leq C\Big((1+\mathrm{e}_{a^0}+H)\big(\mathrm{e}_{a^0}+\mathrm{e}_{b^2}\big)+\mathrm{e}_{a^0}H^\ell+H^{\ell+1}+\varepsilon^2H^\ell\Big)\textstyle\sum_{k=0,2}\|\partial_t^k\bar{u}\|_{\mathrm{L}^p(\mathrm{H}^{\ell+1})},$$

which yields estimate (7.42) for $\|\eta\|_{\mathrm{L}^p(\mathrm{L}^2)}$. Finally, note that $\|I_H\eta\|_{\mathrm{L}^p(\mathrm{L}^2)} \leq \|\bar{u}-I_H\bar{u}\|_{\mathrm{L}^p(\mathrm{L}^2)} + \|\eta\|_{\mathrm{L}^p(\mathrm{L}^2)}$ and use (7.29) to obtain (7.42) for $k=0$. That ends the proof of Lemma 7.1.11. □

Lemma 7.1.12. *The following estimate holds for $\zeta_H = u_H - \pi_H\bar{u}$:*

$$\begin{aligned}\|\partial_t\zeta_H\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \|\zeta_H\|_{\mathrm{L}^\infty(\mathrm{H}^1)} \leq C\Big(&\mathrm{e}_{\mathrm{H}^1}^{\mathrm{data}} + \|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|\partial_t\eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \varepsilon\|\partial_t\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}\\ &+ \|I_H\partial_t^2\eta\|_{\mathrm{L}^1(\mathrm{L}^2)} + \varepsilon\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(\mathrm{H}^1)}\Big),\end{aligned} \tag{7.44}$$

where $\mathrm{e}_{\mathrm{H}^1}^{\mathrm{data}} = \|g^0-g_H^0\|_{\mathrm{H}^1} + \|g^1-g_H^1\|_{\mathrm{L}^2} + \varepsilon\|g^1-g_H^1\|_{\mathrm{H}^1}$.

Proof. Using (7.18) and (7.39), we verify that for any $v_H \in V_H(\Omega)$ and a.e. $t \in [0,T^\varepsilon]$ it holds

$$\big(\partial_t^2\zeta_H(t),v_H\big)_Q + A_H\big(\zeta_H(t),v_H\big) = \big(I_H\partial_t^2\eta(t),v_H\big)_Q. \tag{7.45}$$

Set $v_H = \partial_t\zeta_H(t)$ and use the symmetry of the forms $(\cdot,\cdot)_Q$ and A_H to get for a.e. $t \in [0,T^\varepsilon]$

$$\tfrac{1}{2}\tfrac{\mathrm{d}}{\mathrm{d}t}\Big(\|\partial_t\zeta_H(t)\|_Q^2 + A_H\big(\zeta_H(t),\zeta_H(t)\big)\Big) = \big(I_H\partial_t^2\eta(t),\partial_t\zeta_H(t)\big)_Q.$$

Setting $E_H\zeta_H(t) = \|\partial_t\zeta_H(t)\|_Q^2 + A_H\big(\zeta_H(t),\zeta_H(t)\big)$, we integrate this equality and get

$$E_H\zeta_H(\xi) = E_H\zeta_H(0) + 2\int_0^\xi \big(I_H\partial_t^2\eta(t),\partial_t\zeta_H(t)\big)_Q\,\mathrm{d}t \quad \forall \xi \in [0,T^\varepsilon]. \tag{7.46}$$

We now apply the Cauchy–Schwartz, Hölder, and Young inequalities to bound the second term of the right hand side of (7.46) as

$$2\int_0^\xi \big(I_H\partial_t^2\eta(t),\partial_t\zeta_H(t)\big)_Q\,\mathrm{d}t \leq 2\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(Q)}^2 + \tfrac{1}{2}\|\partial_t\zeta_H\|_{\mathrm{L}^\infty(Q)}^2. \tag{7.47}$$

As $A_H\big(\zeta_H(\xi),\zeta_H(\xi)\big) \geq 0$, combining (7.46) and (7.47) and taking the L^∞ norm with respect to ξ, we obtain the estimate $\frac{1}{2}\|\partial_t\zeta_H\|_{\mathrm{L}^\infty(Q)}^2 \leq E_H\zeta_H(0) + 2\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(Q)}^2$. A similar bound can then be deduced for $\|\zeta_H\|_{\mathrm{L}^\infty(\mathrm{H}^1)}^2$ from (7.46), (7.47), and the ellipticity of A_H. Then, using the boundedness of A_H, we obtain

$$\tfrac{1}{2}\|\partial_t\zeta_H\|_{\mathrm{L}^\infty(Q)}^2 + \lambda\|\zeta_H\|_{\mathrm{L}^\infty(\mathrm{H}^1)}^2 \leq \|\partial_t\zeta_H(0)\|_Q^2 + \Lambda^2/\lambda\|\zeta_H(0)\|_{\mathrm{H}^1}^2 + 2\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(Q)}^2.$$

The first two terms satisfy (recall the splitting of the error (7.34))

$$\begin{aligned}&\|\partial_t\zeta_H(0)\|_Q \leq \|g_H^1-g^1\|_Q + \|\partial_t\eta(0)\|_Q \leq \|g_H^1-g^1\|_Q + \|\partial_t\eta\|_{\mathrm{L}^\infty(Q)},\\ &\|\zeta_H(0)\|_{\mathrm{H}^1} \leq \|g_H^0-g^0\|_{\mathrm{H}^1} + \|\eta(0)\|_{\mathrm{H}^1} \leq \|g_H^0-g^0\|_{\mathrm{H}^1} + \|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}.\end{aligned}$$

Finally, we make use of (7.35) to obtain estimate (7.44) and that concludes the proof of Lemma 7.1.12. □

Lemma 7.1.13. *The function $\zeta_H = u_H - \pi_H \bar{u}$ satisfies*

$$\|\zeta_H\|_{\mathrm{L}^\infty(\mathrm{L}^2)} \leq C\Big(\mathrm{e}^{\mathrm{data}}_{\mathrm{L}^2} + \|\eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \varepsilon\|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|I_H\partial_t\eta\|_{\mathrm{L}^1(\mathrm{L}^2)} + \varepsilon\|I_H\partial_t\eta\|_{\mathrm{L}^1(\mathrm{H}^1)}\Big), \quad (7.48)$$

where $\mathrm{e}^{\mathrm{data}}_{\mathrm{L}^2} = \|g^0 - g^0_H\|_{\mathrm{L}^2} + \varepsilon\|g^0 - g^0_H\|_{\mathrm{H}^1} + \|g^1 - g^1_H\|_{\mathrm{L}^2} + \varepsilon\|g^1 - g^1_H\|_{\mathrm{H}^1}$.

Proof. Rewriting (7.45) with $v_H = w_H(t)$, where $w_H \in \mathrm{H}^1(0, T^\varepsilon; V_H(\Omega))$, we have almost everywhere in $[0, T^\varepsilon]$

$$-\big(\partial_t\zeta_H, \partial_t w_H\big)_Q + A_H\big(\zeta_H, w_H\big) = \tfrac{\mathrm{d}}{\mathrm{d}t}\big(\partial_t(I_H\eta - \zeta_H), w_H\big)_Q - \big(\partial_t I_H\eta, \partial_t w_H\big)_Q.$$

For $\xi \in [0, T^\varepsilon]$, we define $\hat{w}_H(t) = \int_t^\xi \zeta_H(\tau)\mathrm{d}\tau$, which satisfies $\hat{w}_H \in \mathrm{H}^1(0, T^\varepsilon; V_H(\Omega))$, $\hat{w}_H(\xi) = 0$, and $\partial_t\hat{w}_H = -\zeta_H$. We set $w_H = \hat{w}_H$ in the previous equality and thanks to the symmetry of the forms A_H, $(\cdot,\cdot)_Q$ we get almost everywhere in $[0, T^\varepsilon]$

$$\tfrac{1}{2}\tfrac{\mathrm{d}}{\mathrm{d}t}\Big(\|\zeta_H\|^2_Q + A_H\big(\hat{w}_H, \hat{w}_H\big)\Big) = \tfrac{\mathrm{d}}{\mathrm{d}t}\big(\partial_t(I_H e), \hat{w}_H\big)_Q + \big(I_H\partial_t\eta, \zeta_H\big)_Q,$$

where we denoted $e = u - u_H = \eta - \zeta_H$. We integrate over $[0, \xi]$ and obtain $\forall \xi \in [0, T^\varepsilon]$,

$$\|\zeta_H(\xi)\|^2_Q + A_H\big(\hat{w}_H(0), \hat{w}_H(0)\big) = \|\zeta_H(0)\|^2_Q - 2\big(I_H\partial_t e(0), \hat{w}_H(0)\big)_Q + 2\int_0^\xi \big(I_H\partial_t\eta(t), \zeta_H(t)\big)_Q \,\mathrm{d}t. \quad (7.49)$$

The first term of the right hand side is bounded using the triangle inequality as

$$\|\zeta_H(0)\|_Q \leq \|\bar{u}(0) - u_H(0)\|_Q + \|\eta(0)\|_Q \leq \|g^0 - g^0_H\|_Q + \|\eta\|_{\mathrm{L}^\infty(Q)}.$$

The second term is bounded using Cauchy-Schwartz and Young inequalities as

$$2\big(I_H\partial_t e(0), \hat{w}_H(0)\big)_Q \leq \tfrac{2C^2}{\lambda_\Omega}\|I_H\partial_t e(0)\|^2_Q + \tfrac{\lambda_\Omega}{2C^2}\|\hat{w}_H(0)\|^2_Q \leq \tfrac{2C^2}{\lambda_\Omega}\|I_H\partial_t e(0)\|^2_Q + \tfrac{\lambda_\Omega}{2}\|\hat{w}_H(0)\|^2_{\mathrm{H}^1},$$

where C is the constant in (7.35) and $\lambda_\Omega = \lambda/(1 + C^2_\Omega)$, where λ is the ellipticity constant of a^0 and C_Ω is the Poincaré constant. For the third term we use Cauchy–Schwarz, Hölder, and Young inequalities to get

$$2\int_0^\xi \big(I_H\partial_t\eta(t), \zeta_H(t)\big)_Q \,\mathrm{d}t \leq 2\|I_H\partial_t\eta\|^2_{\mathrm{L}^1(Q)} + \tfrac{1}{2}\|\zeta_H\|^2_{\mathrm{L}^\infty(Q)}.$$

Thus, we obtain from the combination of (7.49) with the last three bounds and the ellipticity of $A_H(\cdot,\cdot)$:

$$\tfrac{1}{2}\|\zeta_H\|^2_{\mathrm{L}^\infty(Q)} + \tfrac{\lambda}{2}\|\hat{w}_H(0)\|^2_{\mathrm{H}^1} \leq C\big(\|g^0 - g^0_H\|^2_Q + \|I_H g^1 - g^1_H\|^2_Q + \|\eta\|^2_{\mathrm{L}^\infty(Q)} + \|I_H\partial_t\eta\|^2_{\mathrm{L}^1(Q)}\big).$$

Combined with (7.35) this estimate proves (7.48) and the proof of Lemma 7.1.13 is complete. □

Lemma 7.1.14. *Under the hypotheses of Theorem 7.1.7,* e_{a^0} *and* e_{b^2} *satisfy*

$$\mathrm{e}_{a^0} \leq C\left(\frac{h}{\varepsilon}\right)^2, \qquad \mathrm{e}_{b^2} \leq C\varepsilon\left(\frac{h}{\varepsilon}\right)^2. \quad (7.50)$$

Proof. The proof of the estimate for e_{a^0} can be found in [1]. We prove here the estimate for e_{b^2} in a similar way. For $(K, j) \in \mathcal{T}_H \times \{1, \ldots, J\}$, we introduce the exact solution of the cell problem in $K_{\delta j}$: $\psi_{K_j} \in \mathrm{W}_{\mathrm{per}}(K_{\delta j})$ is the solution of

$$\big(a^\varepsilon(x)\partial_x\psi_{K_j}, \partial_x z\big)_{\mathrm{L}^2(K_{\delta j})} = -\big(a^\varepsilon(x), \partial_x z\big)_{\mathrm{L}^2(K_{\delta j})} \qquad \forall z \in \mathrm{W}_{\mathrm{per}}(K_{\delta j}). \quad (7.51)$$

We define $\bar{b}_K^2(x_{K_j}) = \varepsilon^{-2}\langle \psi_{K_j}^2 \rangle_{K_{\delta j}}$ and split e_{b^2} as $\mathrm{e}_{b^2} \le \mathrm{e}_{b^2}^{\mathrm{mod}} + \mathrm{e}_{b^2}^{\mathrm{mic}}$, where

$$\mathrm{e}_{b^2}^{\mathrm{mod}} = \sup_{K,j} \varepsilon^2 \big| b^2(x_{K_j}) - \bar{b}_K^2(x_{K_j}) \big|, \qquad \mathrm{e}_{b^2}^{\mathrm{mic}} = \sup_{K,j} \varepsilon^2 \big| \bar{b}_K^2(x_{K_j}) - b_K^2(x_{K_j}) \big|.$$

We show that (i) $\mathrm{e}_{b^2}^{\mathrm{mod}} = 0$ and (ii) $\mathrm{e}_{b^2}^{\mathrm{mic}} \le C\varepsilon\big(h/\varepsilon\big)^2$. Fix $(K,j) \in \mathcal{T}_H \times \{1,\dots,J\}$ and write $n = \frac{\delta}{\varepsilon} \in \mathbb{N}_{>0}$, $K_{n\varepsilon} = K_{\delta j}$, $x_K = x_{K_j}$, $\psi = \psi_{K_j}$, $\psi_h = \psi_{h,K_j}$, $b^2 = b^2(x_{K_j})$, and similarly for $\bar{b}_K^2$ and b_K^2. We verify that for any $z \in \mathrm{W}_{\mathrm{per}}(K_{n\varepsilon})$,

$$\Big(a\big(x_K, \tfrac{x}{\varepsilon}\big)\big(\partial_x\big(\varepsilon\chi(x_K, \tfrac{x}{\varepsilon})\big) + 1\big), \partial_x z \Big)_{\mathrm{L}^2(K_{n\varepsilon})} = 0. \tag{7.52}$$

In order to do this, we split the integral over $K_{n\varepsilon}$ into n integral over subcells of size $\varepsilon|Y|$, make the change of variable $x = \varepsilon y$, and use the equation for χ in (7.7). We conclude from (7.52) that $\psi(x) = \varepsilon\chi\big(x_K, \frac{x}{\varepsilon}\big)$ a.e. on $K_{n\varepsilon}$. Similarly we show that

$$\bar{b}_K^2 = (n\varepsilon)^{-1}|Y|^{-1}\int_{K_{n\varepsilon}} \big(\chi\big(x_K, \tfrac{x}{\varepsilon}\big)\big)^2 \,\mathrm{d}x = (n\varepsilon)^{-1}|Y|^{-1}\sum_{k=1}^{n}\int_Y \big(\chi(x_K, y)\big)^2 \,\varepsilon \mathrm{d}y = b^2,$$

and that proves (i). We now show (ii). First, as $a^\varepsilon \in \mathrm{W}^{1,\infty}(\Omega)$ and $|a^\varepsilon|_{\mathrm{W}^{1,\infty}(\Omega)} \le C\varepsilon^{-1}$, elliptic H^2-regularity ensures that $|\psi|_{\mathrm{H}^2(K_\delta)} \le C\varepsilon^{-1}|K_\delta|^{1/2}$. Hence,

$$\|\psi - \psi_h\|_{\mathrm{L}^2(K_\delta)} \le Ch^2|\psi|_{\mathrm{H}^2(K_\delta)} \le Ch^2\varepsilon^{-1}|K_\delta|^{1/2}. \tag{7.53}$$

We then evaluate $|K_\delta|\varepsilon^2\big|\bar{b}_K^2 - b_K^2\big| = \big|\|\psi\|_{\mathrm{L}^2(K_\delta)}^2 - \|\psi_h\|_{\mathrm{L}^2(K_\delta)}^2\big|$ as

$$|K_\delta|\varepsilon^2\big|\bar{b}_K^2 - b_K^2\big| \le \|\psi - \psi_h\|_{\mathrm{L}^2(K_\delta)}\big(2\|\psi\|_{\mathrm{L}^2(K_\delta)} + \|\psi - \psi_h\|_{\mathrm{L}^2(K_\delta)}\big),$$

and using (7.53), we obtain

$$|K_\delta|\varepsilon^2|\bar{b}_K^2 - b_K^2| \le C\varepsilon\big(h/\varepsilon\big)^2|K_\delta|^{1/2}\Big(\|\psi\|_{\mathrm{L}^2(K_\delta)} + \varepsilon\big(h/\varepsilon\big)^2|K_\delta|^{1/2}\Big).$$

As we are in dimension 1, $\psi \in \mathrm{L}^\infty(K_\delta)$ and $\|\psi\|_{\mathrm{L}^2(K_\delta)} \le |K_\delta|^{1/2}\|\psi\|_{\mathrm{L}^\infty(K_\delta)}$, hence,

$$|K_\delta|\varepsilon^2|\bar{b}_K^2 - b_K^2| \le C|K_\delta|\Big(\varepsilon\big(h/\varepsilon\big)^2 + \varepsilon^2\big(h/\varepsilon\big)^4\Big).$$

As we assume $h \le \varepsilon$, (ii) is proved, and the proof of Lemma 7.1.14 is complete. □

Proof of Theorem 7.1.7. Let $e = \bar{u} - u_H$ and denote the norm $\|v\| = \|\partial_t v\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \|v\|_{\mathrm{L}^\infty(\mathrm{H}^1)}$. Recall the splitting (7.34): $e = \eta - \zeta_H$. We apply the triangle inequality and Lemma 7.1.12 and obtain

$$\begin{aligned}\|e\| \le \|\eta\| + \|\zeta_H\| \le C\Big(&\mathrm{e}_{\mathrm{H}^1}^{\mathrm{data}} + \|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|\partial_t\eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \varepsilon\|\partial_t\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} \\ &+ \|I_H\partial_t^2\eta\|_{\mathrm{L}^1(\mathrm{L}^2)} + \varepsilon\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(\mathrm{H}^1)}\Big),\end{aligned} \tag{7.54}$$

where $\mathrm{e}_{\mathrm{H}^1}^{\mathrm{data}} = \|g^1 - g_H^1\|_{\mathrm{H}^1} + \|g^0 - g_H^0\|_{\mathrm{H}^1}$. Using Hölder inequality, gives

$$\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(\mathrm{L}^2)} + \varepsilon\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(\mathrm{H}^1)} \le T^\varepsilon\big(\|I_H\partial_t^2\eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \varepsilon\|I_H\partial_t^2\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}\big).$$

Applying then Lemma 7.1.11, we obtain

$$\|e\| \le C_1\mathrm{e}_{\mathrm{H}^1}^{\mathrm{data}} + C_2\Big((1+T^\varepsilon)(\mathrm{e}_{a^0} + \mathrm{e}_{b^2}) + H^\ell + T^\varepsilon H^{\ell+1} + \varepsilon T^\varepsilon H^\ell\Big)\textstyle\sum_{k=0,4}\|\partial_t^k\bar{u}\|_{\mathrm{L}^\infty(\mathrm{H}^{\ell+1})}.$$

As $T^\varepsilon = \varepsilon^{-2}T$, Lemma 7.1.14 ensures that

$$(1+T^\varepsilon)(\mathrm{e}_{a^0}+\mathrm{e}_{b^2}) \leq C\varepsilon^{-2}(\mathrm{e}_{a^0}+\mathrm{e}_{b^2}) \leq C(h/\varepsilon^2)^2.$$

As all the other terms parts of $\mathrm{e}^{\mathrm{FE}}_{\mathrm{H}^1}$, we obtain estimate (7.30) and the proof of Theorem 7.1.7 is complete. □

Proof of Theorem 7.1.8. First, note that as we assume $h \leq \varepsilon$ Lemma 7.1.14 ensures that $(1+\mathrm{e}_{a^0})(\mathrm{e}_{a^0}+\mathrm{e}_{b^2}) \leq C(h/\varepsilon)^2$. The rest of the proof follows the same line as for Theorem 7.1.7: Using the triangle and Hölder inequalities and Lemma 7.1.13, we obtain

$$\begin{aligned}\|e\|_{\mathrm{L}^\infty(\mathrm{L}^2)} &\leq C\big(\mathrm{e}^{\mathrm{data}}_{\mathrm{L}^2} + \|\eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + \varepsilon\|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|I_H\partial_t\eta\|_{\mathrm{L}^1(\mathrm{L}^2)} + \varepsilon\|I_H\partial_t\eta\|_{\mathrm{L}^1(\mathrm{H}^1)}\big)\\ &\leq C\varepsilon^{-2}(h/\varepsilon)^2\textstyle\sum_{k=0}^3\|\partial_t^k\bar{u}\|_{\mathrm{L}^\infty(\mathrm{H}^{\ell+1})} + \mathrm{e}^{\mathrm{FE}}_{\mathrm{L}^2},\end{aligned}$$

where $\mathrm{e}^{\mathrm{data}}_{\mathrm{L}^2} = \big(\|g^0-g^0_H\|_{\mathrm{L}^2} + \varepsilon\|g^0-g^0_H\|_{\mathrm{H}^1} + \|g^1-g^1_H\|_{\mathrm{L}^2} + \varepsilon\|g^1-g^1_H\|_{\mathrm{H}^1}\big)$. That proves estimate (7.31) and the proof of Theorem 7.1.8 is complete. □

7.1.4 Long time a priori error analysis of the FE-HMM-L in arbitrarily large domains

In Section 7.1.3, we derived a priori error estimates for the FE-HMM-L in the $\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))$ and $\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^1(\Omega))$ norms in small domains of diameters $\mathcal{O}(1)$. As the constants in these estimates depend on the size of Ω, they can not be used in pseudoinfinite domains. In this section, we provide an a priori error analysis that is valid for arbitrarily large domains. In particular, we track the dependence of the estimate on the size of the domain. Specifically, the key point of the proof of the estimate is the use of a new elliptic projection that avoid the need of the Poincaré inequality. Hence, under suitable assumptions, this error estimate can be used in pseudoinfinite domains. This is the first a priori error analysis of a numerical homogenization method for long time wave propagation that holds for arbitrarily large domains.

Again, we assume that the tensor is locally periodic and collocated in the slow variable, i.e.,

$$a^\varepsilon(x) = a\big(x_{K_j}, \tfrac{x}{\varepsilon}\big) \quad \text{for a.e. } x \in K_{\delta j} \quad \forall (K,j) \in \mathcal{T}_H \times \{1,\dots,J\}. \tag{7.55}$$

Theorem 7.1.15. *Let $\bar{u}$ be the solution of (7.13), u_H the solution of the FE-HMM-L (7.18). Assume that δ satisfies $\delta/\varepsilon \in \mathbb{N}_{>0}$, that the micro mesh size is $h \leq \varepsilon$ and that the degree of the micro finite element space is $q=1$. Furthermore, assume that the tensor is locally periodic and collocated in the slow variable (assumption (7.55)). If $a \in \mathcal{C}^0(\bar{\Omega};\mathrm{W}^{1,\infty}(Y)) \cap \mathcal{C}^\ell(\bar{\Omega};\mathrm{L}^\infty(Y))$ and $\partial_t^k\bar{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^{\ell+1}(\Omega))$ for $0 \leq k \leq 4$, then the error $e = \bar{u}-u_H$ satisfies the estimate*

$$\|\partial_t e\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + |e|_{\mathrm{L}^\infty(\mathrm{H}^1)} \leq C\left(\mathrm{e}^{\mathrm{data}}_{\mathrm{H}^1} + \left(\frac{h}{\varepsilon^2}\right)^2 + \frac{H^\ell}{\varepsilon^2}\right)\left(\sum_{\sigma=1}^{\ell+1}|\bar{u}|_{\mathrm{L}^\infty(\mathrm{H}^\sigma)} + \sum_{k=1}^4\|\partial_t^k\bar{u}\|_{\mathrm{L}^\infty(\mathrm{H}^{\ell+1})}\right), \tag{7.56}$$

where $\mathrm{e}^{\mathrm{data}}_{\mathrm{H}^1} = |g^0-g^0_H|_{\mathrm{H}^1(\Omega)} + \|g^1-g^1_H\|_{\mathrm{H}^1(\Omega)}$ and $C = \tilde{C}\big(\|a\|_{\mathcal{C}^\ell(\bar{\Omega};\mathrm{L}^\infty(Y))} + \|a\|_{\mathcal{C}^0(\bar{\Omega};\mathrm{W}^{1,\infty}(Y))}\big)$ with $\tilde{C}$ independent of ε, H and Ω.

Remark 7.1.16. The term H^ℓ/ε^2 in (7.56) is a part of the standard error estimate for the FE approximation of $\bar{u}$ in $V_H(\Omega)$. In the proof, we verify that the factor ε^{-2} comes from the length of the time interval and can not be avoided.

Combining Theorems 6.1.1 and 7.1.15, we obtain the following estimate for $u^\varepsilon - u_H$ in the $\mathrm{L}^\infty(0,T^\varepsilon;W)$ norm.

Corollary 7.1.17. *Assume that Ω is a union of cells of volume $\varepsilon|Y|$ and that the tensor is locally periodic and collocated in the slow variable (assumption (7.55)) and satisfies the regularity $a \in \mathcal{C}^1(\bar{\Omega}; \mathrm{W}^{1,\infty}(Y)) \cap \mathcal{C}^{4\vee\ell}(\bar{\Omega}; \mathrm{L}^\infty(Y))$. Also, assume that $g_H^0 = I_H g^0$, $g_H^1 = I_H g^1$, and let the settings of the FE-HMM-L be such that $\delta/\varepsilon \in \mathbb{N}_{>0}$, $h \leq \varepsilon$, and $q = 1$. Finally assume that the following regularity holds:*

$$g^0 \in \mathrm{H}^4(\Omega),\ g^1 \in \mathrm{H}^3(\Omega),\ f \in \mathrm{L}^2(0,T^\varepsilon; \mathrm{H}^2(\Omega)),\ \partial_t^k \bar{u} \in \mathrm{L}^\infty(0,T^\varepsilon; \mathrm{H}^{(5-k)\vee(\ell+1)}(\Omega)),\ 0 \leq k \leq 4,$$

where we use the notation $m \vee n = \max\{m,n\}$. Then the following estimate holds

$$\|u^\varepsilon - u_H\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} \leq C\left(\varepsilon + \left(\frac{h}{\varepsilon^2}\right)^2 + \frac{H^\ell}{\varepsilon^2}\right)\left(\sum_{\sigma=1}^{5\vee(\ell+1)} |\bar{u}|_{\mathrm{L}^\infty(\mathrm{H}^\sigma)} + \sum_{k=1}^{4} \|\partial_t^k \bar{u}\|_{\mathrm{L}^\infty(\mathrm{H}^{(5-k)\vee(\ell+1)})}\right),$$

where $C = \tilde{C}\Big(\|a\|_{\mathcal{C}^1(\mathrm{W}^{1,\infty})} + \|a\|_{\mathcal{C}^{4\vee\ell}(\mathrm{L}^\infty)}\Big)$ and $\tilde{C}$ is independent of ε, H and Ω and we recall the definition of the norm (see (A.4))

$$\|w\|_W = \inf_{\substack{w=w_1+w_2 \\ w_1,w_2 \in \mathrm{W}_{\mathrm{per}}(\Omega)}} \Big\{\|w_1\|_{\mathrm{L}^2(\Omega)} + \|\nabla w_2\|_{\mathrm{L}^2(\Omega)}\Big\} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(\Omega).$$

We emphasize that the constant $\tilde{C}$ in Theorem 7.1.15 is independent of the domain Ω. Hence, for an arbitrarily large domain Ω, if the quantities

$$\|a\|_{\mathcal{C}^\ell(\bar{\Omega};\mathrm{L}^\infty(Y))}, \quad \sum_{\sigma=1}^{\ell+1} |\bar{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^\sigma(\Omega))}, \quad \sum_{k=1}^{4} \|\partial_t^k \bar{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^{\ell+1}(\Omega))},$$

are of order $\mathcal{O}(1)$, then the mesh size h, H can be set such that the error satisfies a given order of tolerance. This is the case for example if $\bar{u}$ and its time derivatives have a sufficiently small spatial support. Corollary 7.1.17 can then be used to set the parameters of the FE-HMM-L as follows. Let $\tau \geq \varepsilon$ be a desired order of tolerance. Then, setting

$$h = \varepsilon^2 \tau^{1/2}, \qquad H = (\varepsilon^2 \tau)^{1/\ell}, \tag{7.57}$$

the error $\|u^\varepsilon - u_H\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)}$ is at most of order τ.

Proof of the a priori error estimate

The proof of Theorem 7.1.15 follows the same structure as that of Theorems 7.1.7 and 7.1.8, in Section 7.1.3. We split the error as

$$\bar{u} - u_H = (\bar{u} - \pi_H \bar{u}) - (u_H - \pi_H \bar{u}) = \eta - \zeta_H. \tag{7.58}$$

The function $\pi_H \bar{u}$ is a new elliptic projection (defined in (7.62)). In particular, its definition allows to avoid the use of the Poincaré inequality to estimate $\|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}$ (see Remark 7.1.18).

Let us recall some basic estimates and notations used in Section 7.1.3. Note that we need to track the eventual dependence of the constants in the domain Ω. We denote the error in the coefficients as

$$\mathrm{e}_{a^0} = \sup_{K\in\mathcal{T}_H, 1\leq j\leq J} |a^0(x_{K_j}) - a_K^0(x_{K_j})|, \qquad \mathrm{e}_{b^2} = \sup_{K\in\mathcal{T}_H, 1\leq j\leq J} \varepsilon^2 |b^2(x_{K_j}) - b_K^2(x_{K_j})|,$$

where a^0, b^2 and $a_K^0(x_{K_j}), b_K^2(x_{K_j})$. The broken seminorm and norm on $V_H(\Omega)$ are

$$|v_H|_{\bar{\mathrm{H}}^k(\Omega)} = \Big(\sum_{K\in\mathcal{T}_H} |v_H|^2_{\mathrm{H}^k(K)}\Big)^{1/2}, \qquad \|v_H\|_{\bar{\mathrm{H}}^k(\Omega)} = \Big(\sum_{K\in\mathcal{T}_H} \|v_H\|^2_{\mathrm{H}^k(K)}\Big)^{1/2}.$$

As we assume $a \in \mathcal{C}^\ell(\bar\Omega;\mathrm{L}^\infty(Y))$, a^0 and b^2 satisfy $a^0, b^2 \in \mathcal{C}^\ell(\bar\Omega)$ (see (6.44)). Hence, we verify that for $v \in \mathrm{H}^{\ell+1}(\Omega)\cap \mathrm{W}_{\mathrm{per}}(\Omega)$ and $w_H \in V_H(\Omega)$, the following estimates hold

$$|A^0(v,w_H) - A_H(I_H v, w_H)| \le C\Big(\mathrm{e}_{a^0}|v|_{\mathrm{H}^1} + H^\ell \textstyle\sum_{\sigma=1}^{\ell+1}|v|_{\mathrm{H}^\sigma}\Big)|w_H|_{\bar{\mathrm{H}}^1}, \tag{7.59a}$$

$$|(v,w_H)_{\mathcal{S}} - (I_H v, w_H)_Q| \le C\Big(\mathrm{e}_{b^2}\|v\|_{\mathrm{H}^1} + H^\ell\|v\|_{\mathrm{H}^{\ell+1}}\Big)\|w_H\|_{\bar{\mathrm{H}}^1}. \tag{7.59b}$$

where the C depends only λ, Λ, Y and $\|a\|_{\mathcal{C}^\ell(\bar\Omega;\mathrm{L}^\infty(Y))}$ We emphasize that in (7.59), the only dependence on Ω lies in the norms of v, w_H, and a. To see it, let us recall how (7.59a) is obtained (it is similar for (7.59b)). The error is split into three parts : the part coming from the error on the coefficients, the part from the interpolation onto $V_H(\Omega)$ and the part coming from numerical integration :

$$\begin{aligned}|A^0(v,w_H) - A_H(I_H v, w_H)| \le{}& |A^0(v - I_H v, w_H)| + |A^0(I_H v, w_H) - A^0_H(I_H v, w_H)| \\ &+ |A^0_H(I_H v, w_H) - A_H(I_H v, w_H)| = \mathrm{e}_{\mathrm{I_H}} + \mathrm{e}_{\mathrm{int}} + \mathrm{e}_{\mathrm{coef}},\end{aligned} \tag{7.60}$$

where the form A^0_H is defined for $v_H, w_H \in V_H(\Omega)$ as

$$A^0_H(v_H,w_H) = \sum_{K\in\mathcal{T}_H}\sum_{j=1}^{J}\omega_{K_j}a^0(x_{K_j})\partial_x v_H(x_{K_j})\partial_x w_H(x_{K_j}).$$

The interpolation operator satisfies for $v \in \mathrm{W}_{\mathrm{per}}(\Omega)\cap\mathrm{H}^{s+1}(\Omega)$ where $1 \le s \le \ell$,

$$\Big(\sum_{K\in\mathcal{T}_H}|v - I_H v|^2_{\mathrm{H}^m(K)}\Big)^{1/2} \le CH^{s+1-m}|v|_{\mathrm{H}^{s+1}(\Omega)} \quad 0 \le m \le s+1, \tag{7.61}$$

where C is independent of Ω (see Section A.3 and [33]). Using the bound on A^0 and (7.61), we have

$$\mathrm{e}_{\mathrm{I_H}} \le \Lambda|v - I_H v|_{\mathrm{H}^1}|w_H|_{\mathrm{H}^1} \le CH^\ell|v|_{\mathrm{H}^{\ell+1}}|w_H|_{\mathrm{H}^1},$$

for a constant C that is independent of Ω. Next, standard results on numerical integration provide the estimate (see [33, 82] and Theorems A.3.6 and A.3.9)

$$\mathrm{e}_{\mathrm{int}} \le C\|a^0\|_{\mathrm{W}^{\ell,\infty}}H^\ell\textstyle\sum_{k=1}^{\ell}|I_H v|_{\bar{\mathrm{H}}^k}|w_H|_{\mathrm{H}^1}.$$

The proof is done locally for each $K \in \mathcal{T}_H$ and the constant depends only on the reference element $\hat K$ (for $d > 1$, it would additionally depend on the shape regularity of the mesh). Finally, using the definitions of A^0_H and A_H, we easily obtain $\mathrm{e}_{\mathrm{coef}} \le \mathrm{e}_{a^0}|I_H v|_{\mathrm{H}^1}|w_H|_{\mathrm{H}^1}$ and then use the stability of I_H in $\mathrm{H}^1(\Omega)$. Combining the estimates for $\mathrm{e}_{\mathrm{I_H}}$, $\mathrm{e}_{\mathrm{int}}$, and $\mathrm{e}_{\mathrm{coef}}$ with (7.60), we obtain (7.59a).

We define now the elliptic projection $\pi_H\bar u : [0,T^\varepsilon] \to V_H(\Omega)$, where for a.e. t, $\pi_H\bar u(t) \in V_H(\Omega)$ satisfies

$$\big(\pi_H\bar u(t), v_H\big)_Q + A_H\big(\pi_H\bar u(t), v_H\big) = \big(f(t), v_H\big)_{\mathrm{L}^2} - \big(I_H\partial_t^2\bar u(t), v_H\big)_Q + \big(I_H\bar u(t), v_H\big)_Q, \tag{7.62}$$

for any test function $v_H \in V_H(\Omega)$. Using the ellipticity of A_H and (7.35), we verify that the form $(\cdot,\cdot)_Q + A_H(\cdot,\cdot)$ is elliptic and bounded:

$$(v_H,v_H)_Q + A_H(v_H,v_H) \ge c\|v_H\|^2_{\mathrm{H}^1}, \quad (v_H,w_H)_Q + A_H(v_H,w_H) \le C\|v_H\|_{\mathrm{H}^1}\|w_H\|_{\mathrm{H}^1}, \tag{7.63}$$

where c, C are independent of ε and Ω. Furthermore, we verify that the right hand side of (7.62) defines an element of the dual of $V_H(\Omega)$. Hence, Lax–Milgram theorem ensures the existence and uniqueness of $\pi_H\bar u(t)$.

Remark 7.1.18. Observe that the elliptic projection defined in (7.62) is different from the elliptic projection (7.39) from Section 7.1.3. Indeed, the terms $\big(\pi_H\bar{u}(t),v_H\big)_Q$ and $\big(I_H\bar{u}(t),v_H\big)_Q$ have been added respectively on the left and the right hand side. The purpose of these terms is to prove (7.64), with a constant independent of the Poincaré constant. Indeed, the additional term $(\cdot,\cdot)_Q$ in the bilinear form ensures that the bounds (7.63) hold without dependence on the Poincaré constant.

The two following lemmas provide bounds for $\eta=\bar{u}-\pi_H\bar{u}$ and $\zeta_H=u_H-\pi_H\bar{u}$.

Lemma 7.1.19. *Assume that $\partial_t^k\bar{u},\partial_t^{k+2}\bar{u}\in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^{\ell+1}(\Omega))$ for $k\geq 0$. Then $\partial_t^k\pi_H\bar{u}\in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^1(\Omega))$ and the following estimate holds for $\eta=\bar{u}-\pi_H\bar{u}$,*

$$\begin{aligned}\|I_H\partial_t^k\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}+\|\partial_t^k\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}\\ \leq C\Big((\mathrm{e}_{a^0}+H^\ell)\textstyle\sum_{\sigma=1}^{\ell+1}|\partial_t^k\bar{u}|_{\mathrm{L}^\infty(\mathrm{H}^\sigma)}+(\mathrm{e}_{b^2}+H^\ell)\|\partial_t^{k+2}\bar{u}\|_{\mathrm{L}^\infty(\mathrm{H}^{\ell+1})}\Big),\end{aligned}\tag{7.64}$$

where $C=\tilde{C}\|a\|_{\mathcal{C}^\ell(\bar{\Omega};\mathrm{L}^\infty(Y))}$ with $\tilde{C}$ independent of H, ε, and Ω.

Proof. We prove the result for $k=0$. The proof for $k>0$ is obtained in the same way by differentiating equations (7.62) and (7.13) with respect to t. First, using (7.13) we rewrite (7.62) for all $v_H\in V_H(\Omega)$ as

$$\begin{aligned}\big(\pi_H\bar{u}(t),v_H\big)_Q+A_H\big(\pi_H\bar{u}(t),w_H\big)\\ =A^0\big(\bar{u}(t),v_H\big)+\big(\partial_t^2\bar{u}(t),v_H\big)_{\mathcal{S}}-\big(I_H\partial_t^2\bar{u}(t),v_H\big)_Q+\big(I_H\bar{u}(t),v_H\big)_Q.\end{aligned}\tag{7.65}$$

Using the test function $v_H=\pi_H\bar{u}(t)$ in (7.65), the ellipticity of A_H and the bound on A^0, we obtain the estimate

$$\|\pi_H\bar{u}(t)\|_{\mathrm{H}^1}\leq C\big(\|\bar{u}(t)\|_{\mathrm{H}^1}+\|\partial_t^2\bar{u}(t)\|_{\mathrm{H}^1}\big).$$

We take the L^∞ norm with respect to t in this inequality and the regularity of $\bar{u}$ ensures that $\pi_H\bar{u}\in\mathrm{L}^\infty(\mathrm{H}^1)$. Next, we prove estimate (7.64). Using (7.65) and (7.13), we verify that almost everywhere in $[0,T^\varepsilon]$,

$$\big(I_H\eta,v_H\big)_Q+A_H\big(I_H\eta,v_H\big)=A_H\big(I_H\bar{u},v_H\big)-A^0\big(\bar{u},v_H\big)-\big(\partial_t^2\bar{u},v_H\big)_{\mathcal{S}}+\big(I_H\partial_t^2\bar{u},v_H\big)_Q.$$

Thanks to (7.59), we obtain for a.e $t\in[0,T^\varepsilon]$,

$$\big(I_H\eta,v_H\big)_Q+A_H\big(I_H\eta(t),v_H\big)\leq C\Big((\mathrm{e}_{a^0}+H^\ell)\textstyle\sum_{\sigma=1}^{\ell+1}|\bar{u}(t)|_{\mathrm{H}^\sigma}+(\mathrm{e}_{b^2}+H^\ell)\|\partial_t^2\bar{u}(t)\|_{\mathrm{H}^{\ell+1}}\Big)\|v_H\|_{\mathrm{H}^1}.$$

We let $v_H=I_H\eta(t)$ and the ellipticity of the form $(\cdot,\cdot)_Q+A_H(\cdot,\cdot)$ gives

$$\|I_H\eta(t)\|_{\mathrm{H}^1}\leq C\Big((\mathrm{e}_{a^0}+H^\ell)\textstyle\sum_{\sigma=1}^{\ell+1}|\bar{u}(t)|_{\mathrm{H}^\sigma}+(\mathrm{e}_{b^2}+H^\ell)\|\partial_t^2\bar{u}(t)\|_{\mathrm{H}^{\ell+1}}\Big).$$

Taking the L^∞ norm with respect to t, we obtain (7.64) for $\|I_H\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}$. Finally, the triangle inequality yields $\|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}\leq\|\bar{u}-I_H\bar{u}\|_{\mathrm{L}^\infty(\mathrm{H}^1)}+\|I_H\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}$ and using (7.61) proves the (7.64) for $\|\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}$. The proof of Lemma 7.1.19 is complete. □

Lemma 7.1.20. *The following estimate holds for $\zeta_H=u_H-\pi_H\bar{u}$,*

$$\begin{aligned}\|\partial_t\zeta_H\|_{\mathrm{L}^\infty(\mathrm{L}^2)}+|\zeta_H|_{\mathrm{L}^\infty(\mathrm{H}^1)}\leq C\Big(\mathrm{e}_{\mathrm{H}^1}^{\mathrm{data}}+|\eta|_{\mathrm{L}^\infty(\mathrm{H}^1)}+\|\partial_t\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}\\ +\|I_H\eta\|_{\mathrm{L}^1(\mathrm{H}^1)}+\|I_H\partial_t^2\eta\|_{\mathrm{L}^1(\mathrm{H}^1)}\Big),\end{aligned}\tag{7.66}$$

where $\mathrm{e}_{\mathrm{H}^1}^{\mathrm{data}}=|g^0-g_H^0|_{\mathrm{H}^1}+\|g^1-g_H^1\|_Q$ and C is independent of H, ε and Ω.

Proof. Using equations (7.18) and (7.62), we verify that for any $v_H \in V_H(\Omega)$ and a.e. $t \in [0, T^\varepsilon]$ it holds

$$\big(\partial_t^2 \zeta_H(t), v_H\big)_Q + A_H\big(\zeta_H(t), v_H\big) = \big(I_H \partial_t^2 \eta(t) - I_H \eta(t), v_H\big)_Q. \tag{7.67}$$

We let $v_H = \partial_t \zeta_H(t)$ and use the symmetry of the forms $(\cdot,\cdot)_Q$ and A_H to get for a.e. $t \in [0, T^\varepsilon]$

$$\tfrac{1}{2}\tfrac{\mathrm{d}}{\mathrm{d}t}\Big(\|\partial_t \zeta_H(t)\|_Q^2 + A_H\big(\zeta_H(t), \zeta_H(t)\big)\Big) = \big(I_H \partial_t^2 \eta(t) - I_H \eta(t), v_H\big)_Q.$$

Denoting $E_H \zeta_H(t) = \|\partial_t \zeta_H(t)\|_Q^2 + A_H\big(\zeta_H(t), \zeta_H(t)\big)$, we integrate the last equality over $[0, \xi]$ and obtain for any $\xi \in [0, T^\varepsilon]$

$$E_H \zeta_H(\xi) = E_H \zeta_H(0) + 2\int_0^\xi \big(I_H \partial_t^2 \eta(t) - I_H \eta(t), \partial_t \zeta_H(t)\big)_Q \,\mathrm{d}t. \tag{7.68}$$

Applying Cauchy–Schwartz, Hölder, and Young inequalities, we bound the right hand side as

$$2\int_0^\xi \big(I_H \partial_t^2 \eta(t) - I_H \eta(t), \partial_t \zeta_H(t)\big)_Q \,\mathrm{d}t \le 4\|I_H \partial_t^2 \eta\|_{\mathrm{L}^1(Q)}^2 + 4\|I_H \eta\|_{\mathrm{L}^1(Q)}^2 + \tfrac{1}{2}\|\partial_t \zeta_H\|_{\mathrm{L}^\infty(Q)}^2. \tag{7.69}$$

As $A_H\big(\zeta_H(\xi), \zeta_H(\xi)\big) \ge 0$, combining (7.68) and (7.69) and taking the supremum with respect to ξ, we obtain the estimate

$$\tfrac{1}{2}\|\partial_t \zeta_H\|_{\mathrm{L}^\infty(Q)}^2 \le E_H \zeta_H(0) + 4\|I_H \partial_t^2 \eta\|_{\mathrm{L}^1(Q)}^2 + 4\|I_H \eta\|_{\mathrm{L}^1(Q)}^2.$$

A similar bound can then be deduced for $|\zeta_H|_{\mathrm{L}^\infty(\mathrm{H}^1)}^2$ from (7.68), (7.69) and the ellipticity of A_H. Using the bound on A_H, we then obtain

$$\|\partial_t \zeta_H\|_{\mathrm{L}^\infty(Q)} + \|\zeta_H\|_{\mathrm{L}^\infty(\mathrm{H}^1)} \le C\big(\|\partial_t \zeta_H(0)\|_Q + |\zeta_H(0)|_{\mathrm{H}^1} + \|I_H \partial_t^2 \eta\|_{\mathrm{L}^1(Q)} + \|I_H \eta\|_{\mathrm{L}^1(Q)}\big).$$

Thanks to (7.58), the first terms satisfy

$$\begin{aligned}
&\|\partial_t \zeta_H(0)\|_Q \le \|g_H^1 - g^1\|_Q + \|\partial_t \eta(0)\|_Q \le \|g_H^1 - g^1\|_Q + \|\partial_t \eta\|_{\mathrm{L}^\infty(Q)},\\
&|\zeta_H(0)|_{\mathrm{H}^1} \le |g_H^0 - g^0|_{\mathrm{H}^1} + |\eta(0)|_{\mathrm{H}^1} \le |g_H^0 - g^0|_{\mathrm{H}^1} + |\eta|_{\mathrm{L}^\infty(\mathrm{H}^1)}.
\end{aligned}$$

Using (7.35), we obtain (7.66) and that concludes the proof of the lemma. □

Proof of Theorem 7.1.15. Let $e = \bar{u} - u_H$ and recall that $e = \eta - \zeta_H$. Applying the triangle inequality and Lemma 7.1.20, we have

$$\begin{aligned}
\|\partial_t e\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + |e|_{\mathrm{L}^\infty(\mathrm{H}^1)} &\le \|\partial_t \eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + |\eta|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|\partial_t \zeta_H\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + |\zeta_H|_{\mathrm{L}^\infty(\mathrm{H}^1)}\\
&\le C\big(\mathrm{e}_{\mathrm{H}^1}^{\mathrm{data}} + |\eta|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|\partial_t \eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|I_H \eta\|_{\mathrm{L}^1(\mathrm{H}^1)} + \|I_H \partial_t^2 \eta\|_{\mathrm{L}^1(\mathrm{H}^1)}\big).
\end{aligned}$$

Hölder inequality implies $\|I_H \eta\|_{\mathrm{L}^1(\mathrm{H}^1)} + \|I_H \partial_t^2 \eta\|_{\mathrm{L}^1(\mathrm{H}^1)} \le \varepsilon^{-2} T\big(\|I_H \eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|I_H \partial_t^2 \eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}\big)$ and thus, applying Lemma 7.1.19, we obtain

$$\|\partial_t e\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + |e|_{\mathrm{L}^\infty(\mathrm{H}^1)} \le C\mathrm{e}_{\mathrm{H}^1}^{\mathrm{data}} + C\varepsilon^{-2}\big(\mathrm{e}_{a^0} + \mathrm{e}_{b^2} + H^\ell\big)\big(\textstyle\sum_{\sigma=1}^{\ell+1} |\bar{u}|_{\mathrm{L}^\infty(\mathrm{H}^\sigma)} + \sum_{k=1}^4 \|\partial_t^k \bar{u}\|_{\mathrm{L}^\infty(\mathrm{H}^{\ell+1})}\big).$$

Lemma 7.1.14 gives $\varepsilon^{-2}(\mathrm{e}_{a^0} + \mathrm{e}_{b^2}) \le C(h/\varepsilon^2)^2$ and that proves estimate (7.56). The proof of Theorem 7.1.15 is complete. □

7.1.5 Numerical experiments

In this section, we perform numerical experiments to illustrate the theoretical results that were obtained on the FE-HMM-L. In particular, we confirm the micro and macro convergence rates provided by Theorem 7.1.15 for arbitrarily large domains. We also compare the approximation of the FE-HMM-L with the heterogeneous wave u^ε at long time and in a pseudoinfinite medium.

First, we investigate the error estimate from Theorem 7.1.15. We consider the two model problems given by the sets of data

$$\begin{aligned} &g_0(x) = e^{-10x^2}, \quad g_1 = 0, \quad f = 0, \quad a^\varepsilon(x) = a\big(\tfrac{x}{\varepsilon}\big) = \sqrt{2} - \cos\big(2\pi\tfrac{x}{\varepsilon}\big), \\ &L_\varepsilon = \varepsilon^{-2} + 1, \ \Omega = (-L_\varepsilon, L_\varepsilon), \end{aligned} \tag{7.70a}$$

$$\begin{aligned} &g_0(x) = e^{-20x^2}, \quad g_1 = 0, \quad f = 0, \quad a^\varepsilon(x) = a\big(x, \tfrac{x}{\varepsilon}\big) = \tfrac{249}{419} + \tfrac{1}{6}\sin(2\pi x) + \tfrac{1}{6}\sin\big(2\pi\tfrac{x}{\varepsilon}\big), \\ &L_\varepsilon = 0.75\varepsilon^{-2} + 1, \ \Omega = (-L_\varepsilon, L_\varepsilon), \end{aligned} \tag{7.70b}$$

where we fix for now $\varepsilon = 1/10$. We refer respectively to Section 4.4.1 and 6.5.1 for the correctors and effective tensors corresponding to each tensors. In particular, note that for both examples the wave never reaches the boundary of Ω. We approximate u^ε with the FE-HMM-L, where we set $\delta = \varepsilon$, $q = 1$, $\ell = 4$, $H = \varepsilon/4$ and each micro mesh size in the sequence $\{h_n = 2^{-(n-1)}\varepsilon\}_{n=1}^8$. The reference effective solution is computed with a $\mathcal{P}^4$-FEM on a mesh of size $H_{\text{ref}} = \varepsilon/8$. The obtained $\mathrm{L}^\infty(\mathrm{L}^2)$ error for each micro mesh size is display in Figure 7.1. On the left, for model problem (7.70a) and on the right for model problem (7.70b). We observe that in both cases the error decreases with the rate $(h/\varepsilon^2)^2$ as predicted by Theorem 7.1.15. Next, for model problem (7.70b), the same experiment is performed including smaller micro mesh sizes $\{h_n = 2^{-(n-1)}\varepsilon\}_{n=1}^{12}$ and for the different macro mesh sizes $H_1 = 0.025$, $H_2 = 0.05$, $H_3 = 0.1$. In Figure 7.2, we observe that the error saturates when the macro error $\varepsilon^{-2}H^4$ becomes dominant. Indeed, we verify that the three saturation stages are of order $\mathcal{O}(\varepsilon^{-2}H_i^4)$.

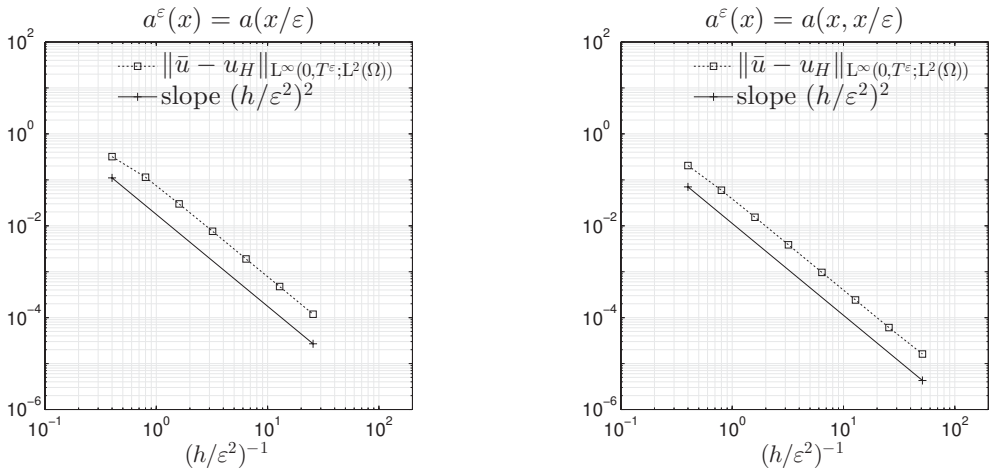

Figure 7.1: Loglog plot of the error $\bar{u} - u_H$ for a decreasing micro mesh size h. Left: model problem (7.70a). Right: model problem (7.70b).

Let us next use the error estimate and approximate u^ε with the FE-HMM-L. We fix $\varepsilon = 1/20$ and $T^\varepsilon = \varepsilon^{-2} = 400$) in the corresponding pseudo infinite domain $\Omega = (-L_\varepsilon, L_\varepsilon)$. For both examples

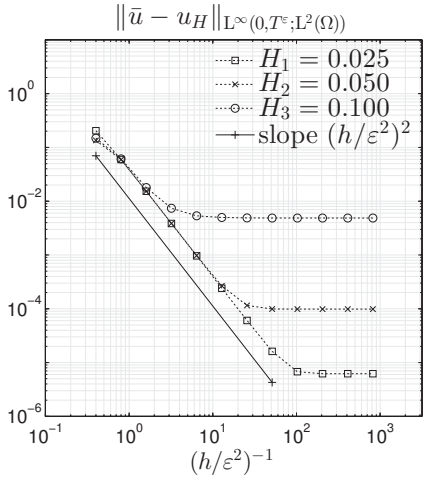

Figure 7.2: Loglog plot of the error $\bar{u} - u_H$ for a decreasing micro mesh size h for three different macro mesh sizes.

in (7.70), recall that the effective solution $\bar{u}$ and the other elements of the family $\mathcal{E}$ capture well the dispersive behavior of u^ε (see Figures 4.3 and 6.1). At the macro scale, we use finite elements of degree $\ell = 3$. We let the tolerance on the error be $\tau = \varepsilon$. Using (7.57), we thus set $h = \varepsilon^{5/2}$, $H = \varepsilon$. The obtained approximations are displayed with u^ε in Figure 7.3 for example (7.70a) and in Figure 7.4 for example (7.70b). In both examples, we observe that the approximation u_H captures well all the features of u^ε. In particular, it describes the long time dispersive effects.

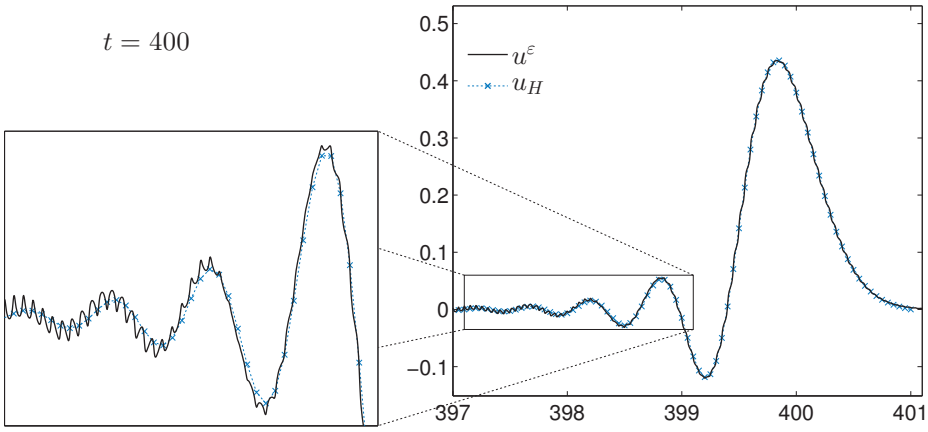

Figure 7.3: Comparison between the wave u^ε with approximation obtained with the FE-HMM-L u_H for example (7.70a) at time $t = \varepsilon^{-2} = 400$ with zoom on $x \in [397.1, 399.1]$.

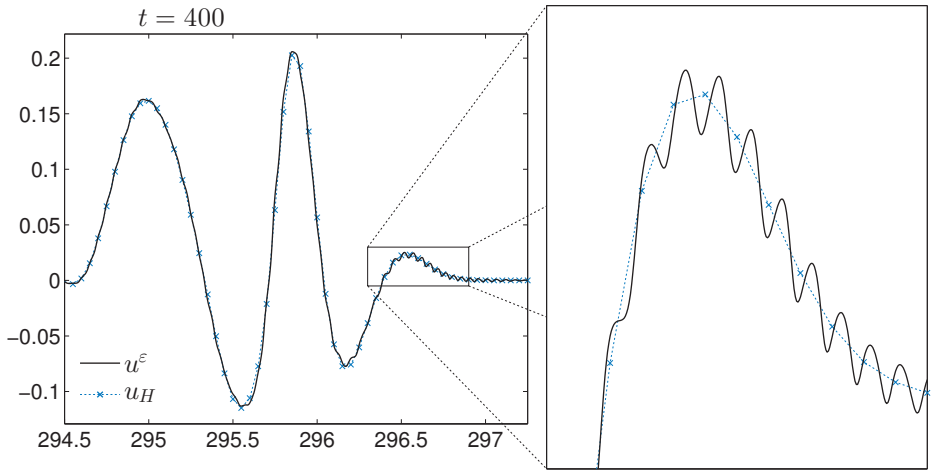

Figure 7.4: Comparison between the frontal waves of u^ε and the approximation obtained with the FE-HMM-L u_H for example (7.70b) at time $t = 400$ and with on $x \in [296.3, 296.9]$.

7.2 Several dimensions : a spectral homogenization method for long time wave propagation in locally periodic media

In this section, we define and analyze a spectral homogenization method for the approximation of multidimensional long time wave propagation in locally periodic media. The method is built to approximate an effective equation in the family of effective equations defined in Chapter 6. In particular, it is structured as follows. In a first step, we use the FEM to solve the cell problems and approximate the effective tensors at the nodes of a grid of the domain. In a second step, we use a spectral method to approximate the effective solution on the grid. The main result of this section is the a priori error analysis of the method, presented in Section 7.2.3. In particular, we prove an error estimate between the approximation and the heterogeneous wave that holds on long times $\mathcal{O}(\varepsilon^{-2})$ and in arbitrarily large periodic domains.

Let us introduce the settings. Let $a^\varepsilon(x) = a\big(x, \frac{x}{\varepsilon}\big)$ be a $d \times d$ locally periodic tensor, i.e., $a(x, y)$ is Y-periodic in y and Ω-periodic in x. The domain $\Omega \subset \mathbb{R}^d$ is an arbitrarily large hypercube, assumed to be the union of cells of length $\varepsilon|Y|$ (see assumption (4.25)). This assumption ensures that $a^\varepsilon(x)$ is Ω-periodic ($y \mapsto a(x, y)$ is extended by periodicity). For $T^\varepsilon = \varepsilon^{-2}T$, we consider the wave equation: find $u^\varepsilon : [0, T^\varepsilon] \times \Omega \to \mathbb{R}$ such that

$$\begin{aligned} &\partial_t^2 u^\varepsilon(t,x) - \nabla_x \cdot \big(a\big(x, \tfrac{x}{\varepsilon}\big)\nabla_x u^\varepsilon(t,x)\big) = f(t,x) && \text{in } (0, T^\varepsilon] \times \Omega, \\ &x \mapsto u^\varepsilon(t,x) \ \ \Omega\text{-periodic} && \text{in } [0, T^\varepsilon], \\ &u^\varepsilon(0,x) = g^0(x), \quad \partial_t u^\varepsilon(0,x) = g^1(x) && \text{in } \Omega, \end{aligned} \tag{7.71}$$

where g^0, g^1 are given initial conditions and f is a source. We assume that a is uniformly elliptic and bounded, i.e. there exists $\lambda, \Lambda > 0$ such that

$$\lambda|\xi|^2 \le a(x,y)\xi \cdot \xi \le \Lambda|\xi|^2 \ \text{ for a.e. } (x,y) \in \Omega \times Y. \tag{7.72}$$

The well-posedness of (7.71) is proved in Section 2.1.1. If $g^0 \in \mathrm{W}_{\mathrm{per}}(\Omega)$, $g^1 \in \mathrm{L}^2_0(\Omega)$, $f \in \mathrm{L}^2(0, T^\varepsilon; \mathrm{L}^2_0(\Omega))$, then there exists a unique weak solution $u^\varepsilon \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{W}_{\mathrm{per}}(\Omega))$ with $\partial_t u^\varepsilon \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{L}^2_0(\Omega))$ and $\partial_t^2 u^\varepsilon \in \mathrm{L}^2(0, T^\varepsilon; \mathrm{W}^*_{\mathrm{per}}(\Omega))$.

7.2.1 Selection of an effective equation for numerical homogenization

The first step on the derivation of a numerical homogenization method for the long time approximation of the wave equation in locally periodic media is to select an appropriate effective model. Recall that a family of effective equations capturing the effective behavior of u^ε at timescales $\mathcal{O}(\varepsilon^{-2})$ was parametrized in Definition 6.2.2, Section 6.2.1. In particular, in the multidimensional case, we have to approximate a fourth order differential equation. As no equation in the family has a specificity, we select the equation defined by the minimal value of the parameter. This is what should be done in practice. For the a priori error analysis, however, we slightly increase the value of the parameter (see below). Doing so ensures the coercivity of the corresponding bilinear form in $\mathrm{H}^2(\Omega)$, which is needed in the proof of the error estimates (see Remark 7.2.1).

Let us then recall the definitions of the tensors of the selected effective equation (as defined in (6.59) to (6.64)). For $x \in \Omega$, let $\{\chi_i(x)\}_{i=1}^d$, $\{\theta_{ij}^0(x)\}_{ij=1}^d$, $\{\theta_i^1(x)\}_{i=1}^d \subset \mathrm{W}_{\mathrm{per}}(Y)$ be the solutions of the local cell problems (6.72) and (6.75). Recall that the homogenized tensor is defined for all $x \in \Omega$ as

$$a_{ij}^0(x) = \big\langle a(x)(\nabla_y \chi_j(x) + e_j) \cdot (\nabla_y \chi_i(x) + e_i) \big\rangle_Y. \tag{7.73}$$

We recall the definition of the operator (see (6.60))

$$L^1 = -\partial_i\big(\bar{a}_{ij}^{12}(x)\partial_j \cdot\big) + b^{10}\partial_t^2,$$

and of the tensors

$$\begin{aligned}
&p_{ijk}^{13}(x) = \big\langle a(x)(\nabla_y \chi_k(x) + e_k) \cdot e_j \chi_i(x) \big\rangle_Y,\\
&q_{ij}^{12}(x) = \big\langle a(x)(\nabla_y \chi_j(x) + e_j) \cdot \nabla_x \chi_i(x) \big\rangle_Y,\\
&\check{a}_{ij}^{12}(x) = S_{ij}^2\Big\{ -\partial_r p_{rij}^{13}(x) + \partial_r p_{irj}^{13}(x) - \partial_r p_{irj}^{13}(x) + 2q_{ij}^{12}(x)\Big\},\\
&b^{10} = \max_{x\in\Omega}\left\{ -\frac{\lambda_{\min}(\check{a}^{12}(x))}{\lambda_{\min}(a^0(x))}\right\}_+,\\
&\bar{a}_{ij}^{12}(x) = \check{a}_{ij}^{12}(x) + b^{10}a_{ij}^0(x),
\end{aligned} \tag{7.74}$$

where $\{\cdot\}_+ = \max\{0, \cdot\}$. Before recalling the definition of the operator L^2, in (6.62), let us observe that the tensors $\bar{a}_{ijkl}^{24}$ and $\bar{a}_{ij}^{22}$, defined in (6.63) and (6.64), can be computed with the symmetrized cell function $\bar{\theta}_{ij}^0 = S_{ij}^2\{\theta_{ij}^0\}$. It is obvious for $\bar{a}_{ijkl}^{24}$ and for $\bar{a}_{ij}^{22}$ observe that we can rewrite

$$S_{ij}^2\big\{\partial_r(p_{jir}^{23} - p_{rij}^{23} - p_{irj}^{23})\big\} = \partial_r \bar{p}_{jir}^{23} - \partial_r S_{ij}^2\big\{p_{rij}^{23} + p_{irj}^{23}\big\} = \partial_r \bar{p}_{jir}^{23} - 2S_{ij}^2\big\{\partial_r \bar{p}_{rij}^{23}\big\},$$

where

$$\bar{p}_{ijk}^{23} = S_{ij}^2\big\{p_{ijk}^{23}\big\} = S_{ij}^2\big\{\big\langle a e_j \chi_i \cdot \nabla_x \chi_k \big\rangle_Y\big\} - \big\langle a\nabla_y \bar{\theta}_{ji}^0 \cdot \nabla_y \theta^1 \big\rangle_Y.$$

This observation leads to a considerable gain of computational time. Indeed, d^2 cell problems must be solved to obtain $\{\theta_{ij}^0(x)\}$, while only $\binom{d+1}{2}$ cell problems are required to compute $\{\bar{\theta}_{ij}^0(x)\}$. We now recall the definition of the operator

$$L^2 = \partial_{ij}^2\big(\bar{a}_{ijkl}^{24}(x)\partial_{kl}^2 \cdot\big) - \partial_i\big(b_{ij}^{22}(x)\partial_j\partial_t^2 \cdot\big) - \partial_i\big(\bar{a}_{ij}^{22}(x)\partial_j \cdot\big) + b^{20}\partial_t^2,$$

and the tensors

$$\begin{aligned}
&\check{a}_{ijkl}^{24}(x) = S_{ij,kl}^{2,2}\Big\{\big\langle a(x)\chi_i(x)e_j \cdot \chi_l(x)e_k\big\rangle_Y\Big\} - \big\langle a(x)\nabla_y\bar{\theta}_{ij}^0(x) \cdot \nabla_y\bar{\theta}_{kl}^0(x)\big\rangle_Y,\\
&A^{24}(x) = M\Big(\check{a}^{24}(x)\Big), \quad A^0(x) = M\Big(S_{ij,kl}^{2,2}\{a_{jk}^0(x)a_{il}^0(x)\}\Big),\\
&\delta = \max_{x\in\Omega}\left\{ -\frac{\lambda_{\min}(A^{24}(x))}{\lambda_{\min}(A^0(x))}\right\}_+ + \frac{\alpha}{\lambda^2},\\
&\bar{a}_{ijkl}^{24}(x) = \check{a}_{ijkl}^{24}(x) + \delta S_{ij,kl}^{2,2}\big\{a_{jk}^0(x)a_{il}^0(x)\big\},\\
&b_{ij}^{22}(x) = \big\langle \chi_i(x)\chi_j(x)\big\rangle_Y + \delta a_{ij}^0(x),
\end{aligned} \tag{7.75}$$

and

$$\begin{aligned}
\bar{p}^{23}_{ijk}(x) &= S^2_{ij}\Big\{\big\langle a(x)e_j\chi_i(x)\cdot\nabla_x\chi_k(x)\big\rangle_Y\Big\} - \big\langle a(x)\nabla_y\bar{\theta}^0_{ji}(x)\cdot\nabla_y\theta^1_k(x)\big\rangle_Y,\\
p^{22}_{ij}(x) &= \big\langle a(x)\nabla_x\chi_j(x)\cdot\nabla_x\chi_i(x)\big\rangle_Y - \big\langle a(x)\nabla_y\theta^1_i(x)\cdot\nabla_y\theta^1_j(x)\big\rangle_Y,\\
\check{a}^{22}_{ij}(x) &= \partial_r\bar{p}^{23}_{jir}(x) - 2S^2_{ij}\{\partial_r\bar{p}^{23}_{rij}(x)\} + p^{22}_{ij}(x)\\
&\quad + b^{10}\check{a}^{12}_{ij}(x) + \delta\partial_s a^0_{ri}(x)\partial_r a^0_{sj}(x) - \delta\partial_r\big(a^0_{rs}(x)\partial_s a^0_{ij}(x)\big),\\
b^{20} &= \max_{x\in\Omega}\left\{-\frac{\lambda_{\min}(\check{a}^{22}(x))}{\lambda_{\min}(a^0(x))}\right\}_+,\\
\bar{a}^{22}_{ij}(x) &= \check{a}^{22}_{ij}(x) + b^{20}a^0_{ij}(x).
\end{aligned} \tag{7.76}$$

Remark 7.2.1. In the definition of δ in (7.75), the parameter $\alpha \geq 0$ is a fixed real value. In practice, $\alpha = 0$ should be used. However, $\alpha > 0$ is used in the a priori error analysis, in Section 7.2.3. Indeed, this ensure the coercivity of the bilinear form A^h_N in $\mathrm{H}^2(\Omega)$ (see (7.96) below) as we verify that (see Lemmas 4.3.2 and 4.3.4)

$$\bar{a}^{24}_{ijkl}(x)\xi_{ij}\xi_{kl} \geq \frac{\alpha}{\lambda^2}S^{2,2}_{ij,kl}\{a^0_{jk}(x)a^0_{il}(x)\}\xi_{ij}\xi_{kl} \geq \alpha\|\xi\|^2_F.$$

The target effective equation for the numerical method is then the solution $\bar{u}$ of (6.65). Let us give the weak formulation for $\bar{u}$ (see Section 2.1.2 for the details). Define the bilinear forms $(\cdot,\cdot)_{\mathcal{S}}$ and $A(\cdot,\cdot)$ as

$$\begin{aligned}
\big(v,w\big)_{\mathcal{S}} &= \big((1+\varepsilon b^{10}+\varepsilon^2 b^{20})v,w\big)_{\mathrm{L}^2(\Omega)} + \big(\varepsilon^2 b^{22}\nabla v,\nabla w\big)_{\mathrm{L}^2(\Omega)},\\
A\big(v,w\big) &= \big((a^0+\varepsilon\bar{a}^{12}+\varepsilon^2\bar{a}^{22})\nabla v,\nabla w\big)_{\mathrm{L}^2(\Omega)} + \big(\varepsilon^2\bar{a}^{24}\nabla^2 v,\nabla^2 w\big)_{\mathrm{L}^2(\Omega)},
\end{aligned} \tag{7.77}$$

and define the spaces

$$\mathcal{S}(\Omega) = \big\{v\in\mathrm{L}^2_0(\Omega) : \sqrt{b^{22}}\nabla v\in[\mathrm{L}^2(\Omega)]^d\big\},\quad \mathcal{V}(\Omega) = \big\{v\in\mathrm{W}_{\mathrm{per}}(\Omega) : \sqrt{\bar{a}^{24}}\nabla^2 v\in[\mathrm{L}^2(\Omega)]^{d\times d}\big\}.$$

Equipped with the inner product $(\cdot,\cdot)_{\mathcal{S}}$ and $A(\cdot,\cdot)$, respectively, $\mathcal{S}(\Omega)$ and $\mathcal{V}(\Omega)$ are Hilbert spaces. If we assume the regularity

$$\begin{gathered}
a^0,\bar{a}^{12},\bar{a}^{22}\in\mathrm{W}^{1,\infty}(\Omega),\quad \bar{a}^{24}\in\mathrm{W}^{2,\infty}(\Omega),\\
g^0\in\mathcal{V}(\Omega)\cap\mathrm{H}^4(\Omega),\quad g^1\in\mathcal{S}(\Omega)\cap\mathrm{H}^2(\Omega),\quad f\in\mathrm{H}^1(0,T^\varepsilon;\mathrm{L}^2_0(\Omega)),
\end{gathered}$$

then there exists a unique weak solution $\bar{u}\in\mathrm{L}^\infty(0,T^\varepsilon;\mathcal{V}(\Omega))$, with $\partial_t\bar{u}\in\mathrm{L}^2(0,T^\varepsilon;\mathcal{S}(\Omega))$ and $\partial^2_t\bar{u}\in\mathrm{L}^2(0,T^\varepsilon;\mathcal{S}(\Omega))$, such that

$$\begin{aligned}
&\big(\partial^2_t\bar{u}(t),v\big)_{\mathcal{S}} + A\big(\bar{u}(t),v\big) = \big(f(t),v\big)_{\mathrm{L}^2(\Omega)}\quad \forall v\in\mathcal{V}(\Omega)\quad \text{for a.e. } t\in[0,T^\varepsilon],\\
&\bar{u}(0) = g^0,\quad \partial_t\bar{u}(0) = g^1.
\end{aligned} \tag{7.78}$$

7.2.2 Definition of the spectral homogenization method

In this section, we define the spectral homogenization method for the long time approximation of the wave equation in locally periodic media. We first describe the structure of the method and then provide all the details.

For $N\in\mathbb{N}^d_{>0}$, let G_N be a uniform grid of Ω (see (7.90)) and let $\mathring{V}_N(\Omega)\subset\mathcal{V}(\Omega)$ be the associated space of trigonometric polynomials of zero mean (defined in (7.94)). In Step 1, we approximate the effective tensors at the nodes of G_N by solving the cell problems with the FEM. In Step 2, using the tensors computed in Step 1, we define the bilinear forms $(\cdot,\cdot)_Q$ and $A^h_N(\cdot,\cdot)$ on

$\mathring{V}_N(\Omega) \times \mathring{V}_N(\Omega)$ to approximate the forms $(\cdot,\cdot)_{\mathcal{S}}$ and $A(\cdot,\cdot)$, defined in (7.77). The spectral homogenization method for the long time approximation of the wave equation in locally periodic media is then defined as: $u_N : [0, T^\varepsilon] \to \mathring{V}_N(\Omega)$ such that

$$\begin{aligned} &\big(\partial_t^2 u_N(t), v_N\big)_Q + A_N^h\big(u_N(t), v_N\big) = \big(f(t), v_N\big)_{\mathrm{L}^2(\Omega)} \quad \forall v_N \in \mathring{V}_N(\Omega) \ \text{ for a.e. } t \in [0, T^\varepsilon], \\ &u_N(0) = g_N^0, \quad \partial_t u_N(0) = g_N^1, \end{aligned} \tag{7.79}$$

where g_N^0 and g_N^1 are appropriate approximations of the initial conditions g^0, g^1 in $\mathring{V}_N(\Omega)$.

In the method, Step 1 is a preprocessing step that only involves the tensor $a(x, y)$. As the cell problems can be decomposed into independent subsets of equations, this process can be parallelized. Furthermore, the outcome of Step 1 can be reused for different initial data and source terms.

Let us now provide the details of Step 1 and Step 2.

Step 1 – Approximation of the effective tensors

In the first step, we approximate the effective tensors at the nodes of the grid G_N. To allow a maximal control on the precision of the method, the tensors are in fact approximated on a subgrid G_M of G_N. Indeed, recall that the definitions of the effective tensors involve space derivatives with respect to the slow variable. As these derivatives are approximated with central differences, considering the subgrid G_M allows to increase the accuracy of the approximation.

Let M be a multiple of N, i.e., for all ν, $M_\nu = \ell N_\nu$ for some $\ell \in \mathbb{N}_{>0}$. Let then $G_M = \{x_m\}$ be the uniform grid of $\Omega = (a_1, b_1) \times \cdots \times (a_d, b_d)$, defined by

$$x_m = (m_1 \Delta x_1, \ldots, m_d \Delta x_d)^T, \quad 0 \le m_1 \le 2M_1 - 1, \ldots, 0 \le m_d \le 2M_d - 1,$$

and the size in each direction is $\Delta x_\nu = (b_\nu - a_\nu)/(2M_\nu)$. For $v \in \mathcal{C}^0(\bar{\Omega}; \mathrm{L}^2(Y))$, let $D_k v(x_m)$ be the approximation of $\partial_{x_k} v(x_m)$ with a central difference, i.e.,

$$v \mapsto D_k v, \quad D_k v(x_m) = \frac{1}{2\Delta x_k}\Big(v(x_{[m+e_k]}) - v(x_{[m-e_k]})\Big) \in \mathrm{L}^2(Y) \quad \forall x_m \in G_M,$$

where $\{e_1, \ldots, e_d\}$ is the canonical basis of $\mathbb{R}^d$ and $[m] \in \mathbb{N}^d$ is defined as $\big([m]\big)_\nu = \big(\mathrm{mod}(m_\nu, 2M_\nu - 1)$ (Ω is a periodic domain). Furthermore, we denote the approximation of the operator ∇_x as $D_x = (D_1 \cdot, \ldots, D_d \cdot)^T$. Using Taylor expansion, we can show the following error estimate for $v \in \mathcal{C}^3(\bar{\Omega}; \mathrm{H}^s(Y))$

$$\|\partial_{x_k} v(x_m) - D_k v(x_m)\|_{\mathrm{H}^s(Y)} \le C \Delta x_k^2 \|v\|_{\mathcal{C}^3(\mathrm{H}^s(Y))}, \tag{7.80}$$

where $\mathrm{H}^0(Y) = \mathrm{L}^2(Y)$. The approximation of the second derivative of a function $v \in \mathcal{C}^0(\bar{\Omega}; \mathrm{L}^2(Y))$ is defined as $v \mapsto D_{kl}^2 v$, where for $x_m \in G_M$

$$D_{kl}^2 v(x_m) = \begin{cases} \dfrac{1}{4\Delta x_k \Delta x_l}\Big(v(x_{[n+e_k+e_l]}) - v(x_{[n+e_k-e_l]}) - v(x_{[n-e_k+e_l]}) + v(x_{[n-e_k-e_l]})\Big) & \text{if } k \neq l, \\ \dfrac{1}{\Delta x_k^2}\Big(v(x_{[n+e_k]}) - 2v(x_{[n]}) + v(x_{[n-e_k]})\Big) & \text{if } k = l. \end{cases}$$

Using Taylor expansion, we can show the following error estimate for $v \in \mathcal{C}^4(\bar{\Omega}; \mathrm{H}^s(Y))$

$$\|\partial_{x_{kl}}^2 v(x_m) - D_{kl}^2 v(x_m)\|_{\mathrm{H}^s(Y)} \le C\big(\Delta x_k^2 + \Delta x_l^2\big)\|v\|_{\mathcal{C}^4(\mathrm{H}^s(Y))}. \tag{7.81}$$

Let us define the finite element space for the approximation of the cell problems. Let $\mathcal{T}^h$ be a triangulation of Y, where h is the maximum diameter of an element in $\mathcal{T}^h$. For and integer $q \geq 1$, the finite element space is then

$$V^h(Y) = \big\{ w^h \in \mathrm{W}_{\mathrm{per}}(Y) : w^h|_K \in \mathcal{P}^q(K) \ \forall K \in \mathcal{T}^h \big\},$$

where $\mathcal{P}^q(K)$ is the set of polynomials of degree smaller or equal to q on K.

We are now able to define the approximations of the correctors and of the effective tensors. For all $x_m \in G_M$, let $\chi_i^h(x_m) \in V^h(Y)$ solves

$$\big(a(x_m)\nabla_y \chi_i^h(x_m), \nabla_y w^h\big)_Y = -\big(a(x_m)e_i, \nabla_y w^h\big)_Y \quad \forall w^h \in V^h(Y). \tag{7.82}$$

The approximation of the homogenized tensor at $x_m \in G_M$ is then defined as

$$a_{ij}^{0h}(x_m) = \big\langle a(x_m)(\nabla_y \chi_j^h(x_m) + e_j) \cdot (\nabla_y \chi_i^h(x_m) + e_i) \big\rangle_Y. \tag{7.83}$$

For all $x_m \in G_M$, let then $\bar{\theta}_{ij}^{0h}(x_m), \theta_i^{1h}(x_m) \in V^h(Y)$ solve $\forall w^h \in V^h(Y)$

$$\begin{aligned}
\big(a(x_m)\nabla_y \bar{\theta}_{ij}^{0h}(x_m), \nabla w_h\big)_Y =& S_{ij}^2 \Big\{ \big(a(x_m)e_i\chi_j^h(x_m), \nabla_y w^h\big)_Y \\
&+ \big(a(x_m)(\nabla_y \chi_j^h(x_m) + e_j) - a^{0h}(x_m)e_j, e_i w^h\big)_Y \Big\}, \qquad (7.84) \\
\big(a(x_m)\nabla_y \theta_i^{1h}(x_m), \nabla w^h\big)_Y =& -\big(a(x_m)D_x\chi_i^h(x_m), \nabla_y w^h\big)_Y \\
&+ \big(D_x \cdot (a(\nabla_y \chi_i^h + e_i))(x_m) - D_x \cdot (a^{0h}(x_m)e_i), w^h\big)_Y. \qquad (7.85)
\end{aligned}$$

We define the following tensors for all $x_m \in G_M$ as (compare to (7.74))

$$\begin{aligned}
p_{ijk}^{13h}(x_m) &= \big\langle a(x_m)(\nabla_y \chi_k^h(x_m) + e_k) \cdot e_j \chi_i^h(x_m) \big\rangle_Y, \\
q_{ij}^{12h}(x_m) &= \big\langle a(x_m)(\nabla_y \chi_j^h(x_m) + e_j) \cdot D_x \chi_i^h(x_m) \big\rangle_Y, \\
\check{a}_{ij}^{12h}(x_m) &= S_{ij}^2 \Big\{ - D_r p_{rij}^{13h}(x_m) + D_r p_{jir}^{13h}(x_m) - D_r p_{irj}^{13h}(x_m) + 2q_{ij}^{12h}(x_m) \Big\}, \\
b^{10h} &= \max_{x_m \in G_M} \left\{ - \frac{\lambda_{\min}(\check{a}^{12h}(x_m))}{\lambda_{\min}(a^{0h}(x_m))} \right\}_+, \\
\bar{a}_{ij}^{12h}(x_m) &= \check{a}_{ij}^{12h}(x_m) + b^{10} a_{ij}^{0h}(x_m).
\end{aligned} \tag{7.86}$$

Furthermore, for all $x_m \in G_M$,

$$\begin{aligned}
\check{a}_{ijkl}^{24h}(x_m) &= S_{ij,kl}^{2,2} \Big\{ \big\langle a(x_m)\chi_i^h(x_m)e_j \cdot \chi_l^h(x_m)e_k \big\rangle_Y \Big\} - \big\langle a(x_m)\nabla_y \bar{\theta}_{ij}^{0h}(x_m) \cdot \nabla_y \bar{\theta}_{kl}^{0h}(x_m) \big\rangle_Y, \\
A^{24h}(x_m) &= M\big(\check{a}^{24h}(x_m)\big), \qquad A^{0h}(x_m) = M\big(S_{ij,kl}^{2,2}\{a_{jk}^{0h}(x_m)a_{il}^{0h}(x_m)\}\big), \\
\delta^h &= \max_{x_m \in G_M} \left\{ - \frac{\lambda_{\min}(A^{24h}(x_m))}{\lambda_{\min}(A^{0h}(x_m))} \right\}_+ + \frac{\alpha}{\lambda^2}, \\
\bar{a}_{ijkl}^{24h}(x_m) &= \check{a}_{ijkl}^{24h}(x_m) + \delta^h S_{ij,kl}^{2,2}\{a_{jk}^{0h}(x_m)a_{il}^{0h}(x_m)\}, \\
b_{ij}^{22h}(x_m) &= \big\langle \chi_i^h(x_m)\chi_j^h(x_m) \big\rangle_Y + \delta^h a_{ij}^{0h}(x_m),
\end{aligned} \tag{7.87}$$

and the tensor $\bar{a}_{ij}^{22h}$ as (compare to (7.76))

$$
\begin{aligned}
\bar{p}_{ijk}^{23h}(x_m) &= S_{ij}^2\Big\{\big\langle a(x_m)e_j\chi_i^h(x_m)\cdot D_x\chi_k^h(x_m)\big\rangle_Y\Big\} - \big\langle a(x_m)\nabla_y\bar{\theta}_{ji}^{0h}(x_m)\cdot\nabla_y\theta_k^{1h}(x_m)\big\rangle_Y,\\
p_{ij}^{22h}(x_m) &= \big\langle a(x_m)D_x\chi_j^h(x_m)\cdot D_x\chi_i^h(x_m)\big\rangle_Y - \big\langle a(x_m)\nabla_y\theta_i^{1h}(x_m)\cdot\nabla_y\theta_j^{1h}(x_m)\big\rangle_Y,\\
\check{a}_{ij}^{22h}(x_m) &= D_r\bar{p}_{jir}^{23h}(x_m) - 2S_{ij}^2\big\{D_r\bar{p}_{rij}^{23h}(x_m)\big\} + p_{ij}^{22h}(x_m) + b^{10h}\check{a}_{ij}^{12h}(x_m)\\
&\quad + \delta^h\big(D_sa_{ri}^{0h}(x_m)D_ra_{sj}^{0h}(x_m) - D_ra_{rs}^{0h}(x_m)D_sa_{ij}^{0h}(x_m) - a_{rs}^{0h}(x_m)D_{rs}^2a_{ij}^{0h}(x_m)\big),\\
b^{20h} &= \max_{x_m\in G_M}\left\{-\frac{\lambda_{\min}(\check{a}^{22h}(x_m))}{\lambda_{\min}(a^{0h}(x_m))}\right\}_+,\\
\bar{a}_{ij}^{22h}(x_m) &= \check{a}_{ij}^{22h}(x_m) + b^{20h}a_{ij}^{0h}(x_m).
\end{aligned}
\tag{7.88}
$$

Recall that α in (7.87) has only a theoretical use and $\alpha = 0$ should be used in applications.

Step 2 - Spectral method for the approximation of the wave

In the second step of the method, we approximate the effective solution with a spectral method. We introduce here the space of approximation, i.e., the finite dimensional space of trigonometric polynomials on the grid G_N. We refer to Appendix A.4 for an introduction on the space of trigonometric polynomials and the corresponding interpolation theory. See also Section 2.3, where the spectral method for the wave equation is analyzed.

Let F_Ω be the bijective affine mapping defined by

$$
F_\Omega : (0,2\pi)^d \to \Omega, \quad \bar{x} \mapsto F_\Omega(\bar{x}) = B_\Omega\bar{x} + a, \tag{7.89}
$$

where B_Ω is the diagonal matrix defined by $(B_\Omega)_{jj} = (b_j - a_j)/(2\pi)$. Recall that $G_M = \{x_m\}$ is the grid of $\Omega = (a_1, b_1) \times \ldots \times (a_d, b_d)$ on which the coefficients are approximated in Step 1. In particular, we recall that $N \in \mathbb{N}_{>0}^d$ divide M, i.e., $M_\nu/N_\nu \in \mathbb{N}_{>0}$ for $\nu = 1, \cdots, d$. Let us define the size of the grid in each direction as $H_\nu = (b_\nu - a_\nu)/(2N_\nu)$. Let then $G_N = \{x_n\}$ be the uniform subgrid of G_M given by

$$
x_n = (n_1H_1, \ldots, n_dH_d)^T, \quad 0 \le n_1 \le 2N_1 - 1, \ldots, 0 \le n_d \le 2N_d - 1. \tag{7.90}
$$

We assume that the ratio $r(N) = \max_\nu N_\nu / \min_\nu N_\nu$ is bounded. We define the space of trigonometric polynomials of order N as

$$
\begin{aligned}
&V_N(\Omega) = \operatorname{span}(B_N),\\
&B_N = \big\{w_{k_1\cdots k_d}(x) = \textstyle\prod_{\nu=1}^d \bar{w}_{k_\nu}^\nu \circ F_\Omega^{-1}(x) : \bar{w}_{k_\nu}^\nu \in B_{N_\nu}^1\big\},\\
&\text{where} \quad B_{N_\nu}^1 = \big\{\bar{w}_{k_\nu}^\nu(\bar{x}) = e^{\mathrm{i}k_\nu\bar{x}} : |k_\nu| \le N_\nu - 1\big\} \cup \big\{\bar{w}_{N_\nu}^\nu(\bar{x}) = \tfrac{1}{2}\big(e^{\mathrm{i}N_\nu\bar{x}} + e^{\mathrm{i}N_\nu\bar{x}}\big)\big\}.
\end{aligned}
$$

We verify that $V_N(\Omega)$ is a vector space of dimension $\prod_{\nu=1}^d 2N_\nu$. On $V_N(\Omega)$, we define the inner product and corresponding norm

$$
(p,q)_N = H^{\mathbb{1}} \sum_{x_n\in G_N} p(x_n)\overline{q(x_n)} = H_1 \sum_{n_1=0}^{2N_1-1} \cdots H_d \sum_{n_d=0}^{2N_d-1} p(x_{n_1\cdots n_d})\overline{q(x_{n_1\cdots n_d})},
$$

where $H^{\mathbb{1}} = H_1 \cdots H_d$ and $\overline{z}$ denote the complex conjugate of $z \in \mathbb{C}$. The corresponding norm is denoted $\|\cdot\|_N = \sqrt{(\cdot,\cdot)_N}$. We verify that

$$
\big(p,q\big)_N = \big(p,q\big)_{\mathrm{L}^2(\Omega)} \qquad \forall p, q \in V_N(\Omega), \tag{7.91}
$$

and thus $p \in V_N(\Omega)$ is uniquely determined by its values on the grid G_N. Let $I_N : \mathrm{L}^2_{\mathrm{per}}(\Omega) \to V_N(\Omega)$ be the interpolation operator defined in (A.74). Theorem A.4.4 states that if $v \in \mathrm{L}^2_{\mathrm{per}}(\Omega) \cap \mathrm{H}^s(\Omega)$, for some $s \geq (d+1)/2$, then, for any $\sigma \leq s$,

$$\left|v - I_N v\right|_{\mathrm{H}^\sigma(\Omega)} \leq C \frac{r(N)^{s-\sigma}}{\left|B_\Omega^{-1} N\right|^{s-\sigma}} |v|_{\mathrm{H}^s(\Omega)}, \tag{7.92}$$

where B_Ω is the matrix in (7.89) and C is a constant depending only on d, s, and $r(N) = \max_\nu N_\nu / \min_\nu N_\nu$. Let us introduce the *convolution* of two trigonometric polynomials $p, q \in V_N(\Omega)$ as the unique trigonometric polynomial $p * q \in V_N(\Omega)$ such that $p * q(x_n) = p(x_n) q(x_n)$ for all $x_n \in G_N$ (the name comes from the fact that the coefficients of $p * q$ are obtained as the finite convolution of the coefficients of p and q). For $c \in \mathrm{L}^\infty_{\mathrm{per}}(\Omega), v \in \mathrm{L}^2_{\mathrm{per}}(\Omega)$, we verify that for all $x_n \in G_N$,

$$I_N c * I_N v(x_n) = I_N c(x_n) I_N v(x_n) = c v(x_n) = I_N(cv)(x_n), \tag{7.93}$$

which implies the equality $I_N c * I_N v = I_N(cv)$. We introduce the subspace

$$\mathring{V}_N(\Omega) = V_N(\Omega) \cap \mathrm{W}_{\mathrm{per}}(\Omega), \tag{7.94}$$

and the corresponding interpolation operator $\mathring{I}_N : \mathrm{L}^2_{\mathrm{per}}(\Omega) \to \mathring{V}_N(\Omega)$, defined in (A.82). Theorem A.4.5 ensures that if $v \in \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^s(\Omega)$, for some $s \geq (d+1)/2$, then, for any $\sigma \leq s$,

$$\left|v - \mathring{I}_N v\right|_{\mathrm{H}^\sigma(\Omega)} \leq C \frac{r(N)^{s-\sigma}}{\left|B_\Omega^{-1} N\right|^{s-\sigma}} |v|_{\mathrm{H}^s(\Omega)}, \tag{7.95}$$

where C is a constant depending only on d, s, and $r(N)$.

Let a^{0h}, b^{i0h}, $\bar{a}^{i2h}$, $\bar{a}^{24h}$, and b^{22h} be the tensors defined in Step 1 (see (7.86), (7.88), and (7.87)). As they are defined at each node of the grid $x_n \in G_N$, they define trigonometric polynomials in $V_N(\Omega)$ (b^{10h} and b^{20h} are in fact constant). We define the following bilinear forms on $\mathring{V}_N(\Omega)$ (approximations of the forms in (7.77))

$$\begin{aligned} \big(v_N, w_N\big)_Q &= \big((1 + \varepsilon b^{10h} + \varepsilon^2 b^{20h}) v_N, w_N\big)_N + \big(\varepsilon^2 b^{22h} * \nabla v_N, \nabla w_N\big)_N, \\ A_N^h\big(v_N, w_N\big) &= \big((a^{0h} + \varepsilon \bar{a}^{12h} + \varepsilon^2 \bar{a}^{22h}) * \nabla v_N, \nabla w_N\big)_N + \big(\varepsilon^2 \bar{a}^{24h} * \nabla^2 v_N, \nabla^2 w_N\big)_N. \end{aligned} \tag{7.96}$$

In (7.96), the matrix-vector convolution products are defined as $(a^{0h} * \nabla v_N)_i = a_{ij}^{0h} * \partial_j v_N \in V_N(\Omega)$ (and $\bar{a}^{24h} * \nabla^2 v_N$ similarly). Note that by construction, b^{10h}, b^{20h} and b^{22h} are positive semidefinite so that $(\cdot,\cdot)_Q$ is an inner product on $\mathring{V}_N(\Omega)$. We define the corresponding norm $\|v_N\|_Q = \sqrt{(v_N, v_N)_Q}$. Furthermore, the tensors $a^{0h}, \bar{a}^{12h}, \bar{a}^{22h}, \bar{a}^{24h}$ being symmetric, the form $A_N^h(\cdot,\cdot)$ is symmetric.

The spectral numerical homogenization method for long time wave propagation is then defined as the solution $u_N : [0, T^\varepsilon] \to \mathring{V}_N(\Omega)$ of (7.79). To prove the stability and well-posedness of the method, we prove the following lemma.

Lemma 7.2.2. *Assume that $a \in \mathcal{C}^0(\bar{\Omega}; \mathrm{W}^{q,\infty}(Y)) \cap \mathcal{C}^2(\bar{\Omega}; \mathrm{L}^\infty(Y))$, $\varepsilon h^q \leq C_{\mathrm{s},1} \Delta x_{\min}$, and $\varepsilon \leq C_{\mathrm{s},2} \Delta x_{\min}$. Then, there exists L and Γ such that for all $v_N, w_N \in V_N(\Omega)$*

$$\|v_N\|_{\mathrm{L}^2(\Omega)} \leq \|v_N\|_Q \leq L \|v_N\|_{\mathrm{H}^1(\Omega)}, \tag{7.97}$$

$$A_N^h(v_N, v_N) \geq \lambda |v_N|^2_{\mathrm{H}^1(\Omega)} + \varepsilon^2 \alpha |v_N|^2_{\mathrm{H}^2(\Omega)}, \qquad A_N^h(v_N, w_N) \leq \Gamma \|v_N\|_{\mathrm{H}^2(\Omega)} \|w_N\|_{\mathrm{H}^2(\Omega)}, \tag{7.98}$$

where L and Γ depends on λ, Λ, $C_{\mathrm{s},1}, C_{\mathrm{s},2}, \|a\|_{\mathcal{C}^0(\bar{\Omega}; \mathrm{W}^{q,\infty}(Y))}, \|a\|_{\mathcal{C}^2(\bar{\Omega}; \mathrm{L}^\infty(Y))}$ and d.

Lemma 7.2.2 ensures that (7.79) is equivalent to a well-posed second order ODE and we obtain the existence of a unique solution $u_N \in \mathcal{C}^1([0,T^\varepsilon];\mathring{V}_N(\Omega))$ (see e.g. [38]). Furthermore, using (7.97) and (7.98), we obtain the following stability estimate

$$\|\partial_t u_N\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} + \|\nabla u_N\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} \leq C\big(\|g^1\|_{\mathrm{H}^1(\Omega)} + \|g^0\|_{\mathrm{H}^2(\Omega)} + \|f\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))}\big),$$

where C depends on λ, Γ and L.

Proof of Lemma 7.2.2. The proof is structured as follows. First, to prove the ellipticity of A_N^h, we prove that the approximated homogenized tensor $a^{0h}(x_m)$ is positive definite and bounded. Second, we prove that all the approximated tensors are bounded, which ensures the bounds on A_N^h and $\|\cdot\|_Q$.
i) Let us prove the ellipticity of A_N^h. To that end, we first show that $a^{0h}(x_m)$ is positive definite and bounded. We follow the lines of the proof of Lemma 3.3.1 on the properties of a^0. For $\xi \in \mathbb{R}^d$, we have for all $x_m \in G_M$,

$$|Y| a^{0h}(x_m)\xi\cdot\xi = \big(a(x_m)(\nabla_y \chi_i^h(x_m) + e_i)\cdot(\nabla_y\chi_i^h(x_m)+e_i)\xi_i)\big)_Y \xi_i\xi_j = \big(a(x_n)F_\xi^{h,m}, F_\xi^{h,m}\big)_Y, \tag{7.99}$$

where we denoted the field $F_\xi^{h,m} = \sum_{i=1}^d (\nabla_y\chi_i^h(x_m)+e_i)\xi_i$. As $\chi_i^h(x_m)$ is Y-periodic, it satisfies $\big(\nabla_y\chi_i^h(x_m), e_j\big)_Y = \int_Y \partial_{y_j}\chi_i^h(x_m)\,\mathrm{d}y = 0$, and thus

$$\begin{aligned}\|F_\xi^{h,m}\|^2_{\mathrm{L}^2(Y)} &= \big(\nabla_y\chi_i^h, \nabla_y\chi_j^h\big)_Y\xi_i\xi_j + \big(\nabla_y\chi_i^h, e_j\big)_Y\xi_i\xi_j + \big(e_i,\nabla_y\chi_j^h\big)_Y\xi_i\xi_j + \big(e_i,e_j\big)_Y\xi_i\xi_j\\ &= \big\|\textstyle\sum_i \nabla_y\chi_i^h\xi_i\big\|^2_{\mathrm{L}^2} + |Y||\xi|^2 \geq |Y||\xi|^2.\end{aligned}$$

Using (7.99) and the ellipticity of a, this estimate implies $|Y|a^{0h}(x_m)\xi\cdot\xi \geq \lambda\|F_\xi^{h,m}\|^2_{\mathrm{L}^2(Y)} \geq |Y|\lambda|\xi|^2$, which proves the λ-ellipticity of a^{0h}. Using the cell problem for χ_j^h and the ellipticity of a, we have

$$\begin{aligned}\big(a(x_m)F_\xi^{h,m}, F_\xi^{h,m}\big)_Y &= \big(a\nabla_y\chi_i^h,\nabla_y\chi_j^h\big)\xi_i\xi_j + \big(a\nabla_y\chi_i^h, e_j\big)\xi_i\xi_j + \big(ae_i,\nabla_y\chi_j^h\big)\xi_i\xi_j + \big(ae_i,e_j\big)\xi_i\xi_j\\ &= -\big(a\nabla_y\chi_i^h,\nabla_y\chi_j^h\big)\xi_i\xi_j + \big(ae_i,e_j\big)\xi_i\xi_j\\ &= -\Big(a\big(\textstyle\sum_i\nabla_y\chi_i^h\xi_i\big)\cdot\big(\sum_i\nabla_y\chi_i^h\xi_i\big)\Big)_Y + (a\xi,\xi)_Y \leq (a(x_m)\xi,\xi)_Y.\end{aligned}$$

Using then (7.99) and the bound on a, we get $|Y|a^{0h}(x_m)\xi\cdot\xi \leq \big(a(x_m)\xi,\xi\big)_Y \leq |Y|\Lambda|\xi|^2$. This estimate proves that a^{0h} is bounded as $a^{0h}(x_m)\xi\cdot\xi \leq \Lambda|\xi|^2$. A similar argument as in Remark 7.2.1 ensures that $\big(\bar{a}^{24h}\nabla_x^2 v_N, \nabla_x^2 v_N\big)_N \geq \alpha|v_N|^2_{\mathrm{H}^2}$. Hence, the ellipticity of A_N^h in (7.98) is proved.
ii) We next prove that the forms $(\cdot,\cdot)_Q$ and $A_N^h(\cdot,\cdot)$ are bounded independently of N. Note that the regularity of $a(x,y)$ ensures (see (6.108))

$$\chi_i, \theta_{ij}^0 \in \mathcal{C}^0(\bar\Omega;\mathrm{H}^{q+1}(Y))\cap\mathcal{C}^2(\bar\Omega;\mathrm{H}^1(Y)), \quad \theta_{ij}^1 \in \mathcal{C}^0(\bar\Omega;\mathrm{H}^1(Y)), \quad a^0 \in \mathcal{C}^2(\bar\Omega).$$

Standard estimates in the analysis of the finite element method thus ensure (see e.g., [33] or Appendix A.3)

$$\|\chi_i^h(x_m)\|_{\mathrm{H}^1(Y)} \leq C, \quad \|\chi_i(x_m)-\chi_i^h(x_m)\|_{\mathrm{H}^1(Y)} \leq Ch^q, \quad \|\chi_i(x_m)-\chi_i^h(x_m)\|_{\mathrm{L}^2(Y)} \leq Ch^{q+1}, \tag{7.100}$$

$$\|\bar\theta_{ij}^{0h}(x_m)\|_{\mathrm{H}^1(Y)} \leq C, \quad \|\bar\theta_{ij}^0(x_m)-\bar\theta_{ij}^{0h}(x_m)\|_{\mathrm{H}^1(Y)} \leq Ch^q. \tag{7.101}$$

Furthermore, the condition $\varepsilon h^q \leq C_{s,1}\Delta x_{\min}$ and Taylor's theorem ensure

$$\begin{aligned}\varepsilon\|D_k\chi_i^h(x_m)\|_Y &\leq \varepsilon\|\partial_{x_k}\chi_i(x_m)\|_Y + \varepsilon\|\partial_{x_k}\chi_i(x_m) - D_k\chi_i(x_m)\|_Y + \varepsilon\|D_k(\chi - \chi_i^h)(x_m)\|_Y\\ &\leq 2\varepsilon\|\chi_i\|_{\mathcal{C}^1(\mathrm{L}^2)} + C\varepsilon h^{q+1}\Delta x_k^{-1}\|\chi_i\|_{\mathcal{C}^0(\mathrm{H}^{q+1})} \leq C\big(1+\varepsilon h^{q+1}\Delta x_{\min}^{-1}\|\chi\|_{\mathcal{C}^1(\mathrm{H}^{q+1})}\big) \leq C.\end{aligned}$$

Similarly, denoting $F_{ik} = e_k^T a(\nabla_y \chi_i + e_i)$, $F_{ik}^h = e_k^T a(\nabla_y \chi_i^h + e_i)$, we have

$$\begin{aligned}\varepsilon\|D_k F_{ik}^h(x_m)\|_Y \leq & \varepsilon\|\partial_{x_k} F_{ik}(x_m)\|_Y + \varepsilon\|\partial_{x_k} F_{ik}(x_m) - D_k F_{ik}(x_m)\|_Y + \varepsilon\|D_k(F_{ik} - F_{ik}^h)(x_m)\|_Y \\ & \leq C\varepsilon\|F_{ik}\|_{\mathcal{C}^1(\mathrm{L}^2)} + C\varepsilon h^q \Delta x_k^{-1}\|\chi_i\|_{\mathcal{C}^0(\mathrm{H}^{q+1})} \leq C\big(1 + \varepsilon h^q \Delta x_{\min}^{-1}\big) \leq C,\end{aligned}$$

and we prove in the same way $\varepsilon\|D_k a_{ki}^{0h}(x_m)\|_Y \leq C$. We thus get $\varepsilon\|\theta_i^{1h}(x_m)\|_{\mathrm{H}^1(Y)} \leq C$. Using the bounds on $\chi_i^h, \bar{\theta}_{ij}^{0h}$ and θ_i^{1h}, we verify that $a_{ij}^{0h},\ \varepsilon\bar{a}_{ij}^{12h},\ \varepsilon b^{10h},\ \bar{a}_{ijkl}^{24h},\ b_{ij}^{22h} \leq C$. We still have to prove that $\varepsilon^2\bar{a}_{ij}^{22}$ and $\varepsilon^2 b^{20}$ are bounded. To see it, first note that

$$\varepsilon^2 p_{ij}^{22}(x_m) \leq C\big(\varepsilon^2\|D_x\chi_i^h(x_m)\|_{\mathrm{L}^2(Y)}\|D_x\chi_j^h(x_m)\|_{\mathrm{L}^2(Y)} + \varepsilon^2\|\theta_i^{1h}(x_m)\|_{\mathrm{H}^1(Y)}\|\theta_j^{1h}(x_m)\|_{\mathrm{H}^1(Y)}\big) \leq C.$$

Next,

$$\varepsilon\bar{p}_{ijk}^{23}(x_m) \leq C\big(\varepsilon\|\chi_i^h(x_m)\|_{\mathrm{L}^2(Y)}\|D_x\chi_k^h(x_m)\|_{\mathrm{L}^2(Y)} + \varepsilon\|\bar{\theta}_{ji}^{0h}(x_m)\|_{\mathrm{H}^1(Y)}\|\theta_k^{1h}(x_m)\|_{\mathrm{H}^1(Y)}\big) \leq C,$$

and thus, thanks to the second condition, $\varepsilon \leq C_{\mathrm{s},2}\Delta x_{\min}$, we have

$$\varepsilon^2 D_r\bar{p}_{ijk}^{23}(x_m) = \varepsilon(2\Delta x_r)^{-1}\big(\varepsilon\bar{p}_{ijk}^{23}(x_{m+e_r}) - \varepsilon\bar{p}_{ijk}^{23}(x_{m-e_r})\big) \leq C\varepsilon\Delta x_{\min}^{-1} \leq C.$$

Finally, as $a^0 \in \mathcal{C}^2(\bar{\Omega})$ and thanks to the first condition,

$$\begin{aligned}\varepsilon^2 D_{rs}^2 a_{ij}^{0h}(x_m) \leq & \varepsilon^2|\partial_{x_r s}^2 a_{ij}^0(x_m)| + \varepsilon^2|\partial_{x_{rs}}^2 a_{ij}^0(x_m) - D_{rs}^2 a_{ij}^0(x_m)| + \varepsilon^2|D_{rs}^2(a_{ij}^0 - a_{ij}^{0h})(x_m)| \\ & \leq C\varepsilon^2\Delta x_{\min}^{-2} h^{2q} \leq C,\end{aligned}$$

and we obtain the bounds $\varepsilon^2\bar{a}_{ij}^{22}, \varepsilon^2 b^{20} \leq C$. As all the coefficients are bounded, we obtain the upper bounds in (7.97) and (7.98) and the proof of the lemma is complete. □

7.2.3 A priori error analysis of the spectral homogenization method

We present here the main result of this section: the a priori error analysis of the spectral homogenization method defined in the previous section. In particular, we provide an error estimate between the approximation and the effective solution that holds over long time and in arbitrarily large periodic domains. The proof of the result is presented in Section 7.2.4.

Let $\bar{u}$ be the effective solution (7.78) and let u_N be its approximation defined in (7.79). Recall that q is the degree of the finite element space $V^h(Y)$ used for the approximation of the correctors. We prove the following a priori error estimate for $\bar{u} - u_N$.

Theorem 7.2.3. *Assume that for some $s \geq (d+1)/2$, the tensor and the effective solution satisfy the regularity*

$$\begin{aligned}& a \in \mathcal{C}^0(\bar{\Omega}; \mathrm{W}^{q,\infty}(Y)) \cap \mathcal{C}^1(\bar{\Omega}; \mathrm{W}^{q-1,\infty}(Y)) \cap \mathcal{C}^{s+2}(\bar{\Omega}; \mathrm{L}^\infty(Y)), \\ & \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^{s+2}(\Omega)), \qquad \partial_t^k \bar{u} \in \mathrm{L}^\infty(0, T^\varepsilon; \mathrm{H}^{s+1}(\Omega)),\ 1 \leq k \leq 4.\end{aligned}$$

Furthermore, assume that the ratios $r(\Delta x) = \max_\nu \Delta x_\nu / \min_\nu \Delta x_\nu$ and $r(N) = \max_\nu N_\nu / \min_\nu N_\nu$ are bounded and that ε and Δx_ν are bounded independently of $\mathrm{diam}(\Omega)$. *Then the error $e = \bar{u} - u_N$ satisfies the following estimate*

$$\begin{aligned}& \|\partial_t e\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{L}^2(\Omega))} + |e|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^1(\Omega))} \\ & \quad \leq C\mathrm{e}_{\mathrm{data}} + C\bigg(\frac{1}{\varepsilon^2|B_\Omega^{-1}N|^s} + \mathrm{e}_1 + \mathrm{e}_2\bigg)\bigg(\sum_{\sigma=1}^{s+2}|\bar{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^\sigma(\Omega))} + \sum_{k=1}^{4}\|\partial_t^k\bar{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^{s+1}(\Omega))}\bigg),\end{aligned} \tag{7.102}$$

where $\mathrm{e}_{\mathrm{data}} = \|g^0 - g_N^0\|_{\mathrm{H}^2(\Omega)} + \|g^1 - g_N^1\|_{\mathrm{H}^1(\Omega)}$, B_Ω *is the matrix in the affine mapping* $F_\Omega : (0,2\pi)^d \to \Omega$ *in* (7.89), *and*

$$\begin{aligned} \mathrm{e}_1 &= \left(\frac{h^q}{\varepsilon}\right)^2 + \frac{|\Delta x|}{\varepsilon} + \frac{h^q}{\varepsilon}, \\ \mathrm{e}_2 &= \frac{h^q|1/\Delta x|}{\varepsilon} + h^q|1/\Delta x|^2, \qquad |1/\Delta x| = \sqrt{\textstyle\sum_{\nu=1}^d 1/\Delta x_\nu^2}, \end{aligned} \tag{7.103}$$

and C *depends only on* λ, α, Y, d, s, $r(N)$, $r(\Delta x)$, $\|a\|_{\mathcal{C}^0(\bar\Omega;\mathrm{W}^{q,\infty}(Y))}$, $\|a\|_{\mathcal{C}^1(\bar\Omega;\mathrm{W}^{q-1,\infty}(Y))}$, *and* $\|a\|_{\mathcal{C}^{s+2}(\bar\Omega;\mathrm{L}^\infty(Y))}$.

Let us discuss the terms e_1 and e_2 in (7.102). The error term e_1 originates from the approximations of the tensors a^0, b^{22} and a^{24}. Note that we obtain a linear rate in $|\Delta x|$ instead of the square rate expected by the use of the central difference scheme. This lower rate is due to the approximation of the maximum on the domain by a maximum on the grid in the definition of b^{22}, a^{24} (see Lemma 7.2.7). The error term e_2 comes from the approximations of a^{12}, b^{10}, a^{22}, and b^{20} and constrains the value of h with respect Δx. Let us explain why. Note that the accuracy of the approximated slope between two approximated points strongly depends on the accuracy of the points. In particular, to obtain an accurate value of the slope, the smaller the distance between the two points is, the more accurate the approximation of the points must be. Likewise, in the spectral homogenization method, if the correctors $\chi_i(x_n - e_k)$ and $\chi_i(x_n + e_k)$ are not approximated accurately enough, we can not expect the central difference scheme to provide an accurate approximation of $\partial_k \chi_i(x_n)$.

We verify that e_1 is connected to the second stability condition of Lemma 7.2.2. Indeed, if we enforce a tolerance τ for e_1, then the requirement $\varepsilon \le C_{\mathrm{s},2}\Delta x_{\min}$ holds for $C_{\mathrm{s},2} = \tau/d$. We see that e_2 is connected to the first stability condition of Lemma 7.2.2 in the following way. As

$$\mathrm{e}_2 \ge \sqrt{d}\frac{h^q}{\varepsilon\Delta x_{\max}} + d\frac{h^q}{\Delta x_{\max}^2} \ge \sqrt{d}r(\Delta x)\varepsilon^{-2}\frac{h^q}{\varepsilon\Delta x_{\min}}\left(1 + \frac{\varepsilon}{\Delta x_{\min}}\right),$$

if we enforce a tolerance τ for e_2, then the requirement $\varepsilon h^q \le C_{\mathrm{s},1}\Delta x_{\min}$ automatically holds for $C_{\mathrm{s},1} = \tau\varepsilon^2/(\sqrt{d}r(\Delta x))$.

Combining Theorems 6.2.1 and 7.2.3, we obtain the following estimate for $u^\varepsilon - u_N$.

Corollary 7.2.4. *Assume that for some* $s \ge 3$, *the data and the effective solution satisfy the regularity*

$$\begin{aligned} a &\in \mathcal{C}^0(\bar\Omega;\mathrm{W}^{q,\infty}(Y)) \cap \mathcal{C}^1(\bar\Omega;\mathrm{W}^{(q-1)\vee 2,\infty}(Y)) \cap \mathcal{C}^2(\bar\Omega;\mathrm{W}^{1,\infty}(Y)) \cap \mathcal{C}^{s+2}(\bar\Omega;\mathrm{L}^\infty(Y)), \\ &g^0 \in \mathrm{H}^{s+2}(\Omega), \quad g^1 \in \mathrm{H}^{s+1}(\Omega), \quad f \in \mathrm{L}^2(0,T^\varepsilon;\mathrm{H}^2(\Omega)), \\ \bar u &\in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^{s+2}(\Omega)), \qquad \partial_t^k \bar u \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^{s+1}(\Omega)),\ 1 \le k \le 4, \end{aligned} \tag{7.104}$$

where $m \vee n = \max\{m,n\}$. *Furthermore, assume that the ratios* $r(\Delta x)$ *and* $r(N)$ *are bounded and that* ε *and* Δx_ν *are bounded independently of* $\mathrm{diam}(\Omega)$. *Finally, let the initial condition in the method* (7.79) *be* $g_N^i = \mathring{I}_N g^i$. *Then the following estimate holds*

$$\begin{aligned} &\|u^\varepsilon - u_N\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)} \\ &\quad \le C\left(\varepsilon + \frac{1}{\varepsilon^2|B_\Omega^{-1}N|^s} + \mathrm{e}_1 + \mathrm{e}_2\right)\left(\sum_{\sigma=1}^{s+2}|\bar u|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^\sigma(\Omega))} + \sum_{k=1}^{4}\|\partial_t^k \bar u\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^{s+1}(\Omega))}\right), \end{aligned} \tag{7.105}$$

where $\mathrm{e}^1, \mathrm{e}^2$ *are defined in* (7.103) *and* C *depends only on* λ, α, Y, d, s, $r(N)$, $r(\Delta x)$, $\|a\|_{\mathcal{C}^0(\bar\Omega;\mathrm{W}^{q,\infty}(Y))}$, $\|a\|_{\mathcal{C}^1(\bar\Omega;\mathrm{W}^{(q-1)\vee 2,\infty}(Y))}$, $\|a\|_{\mathcal{C}^2(\bar\Omega;\mathrm{W}^{1,\infty}(Y))}$, *and* $\|a\|_{\mathcal{C}^{s+2}(\bar\Omega;\mathrm{L}^\infty(Y))}$ *and we recall*

the definition of the norm (see (A.4)*)*

$$\|w\|_W = \inf_{\substack{w=w_1+w_2 \\ w_1,w_2\in \mathrm{W}_{\mathrm{per}}(\Omega)}} \left\{ \|w_1\|_{\mathrm{L}^2(\Omega)} + \|\nabla w_2\|_{\mathrm{L}^2(\Omega)} \right\} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(\Omega).$$

We emphasize that the only dependence on the domain Ω of the constant C in (7.102) and (7.105) is in the norms of $a(x,y)$. In particular, for an arbitrarily large domain Ω, if $a(x,y)$ is sufficiently smooth and the quantities

$$\textstyle\sum_{\sigma=1}^{s+2} |\bar{u}|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^\sigma(\Omega))}, \quad \|\partial_t^k \bar{u}\|_{\mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^{s+1}(\Omega))}\ 1 \le k \le 4,$$

are of order $\mathcal{O}(1)$, then (7.105) ensures $\|u^\varepsilon - u_N\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)}$ to be of order $\varepsilon + \varepsilon^{-2}|B_\Omega^{-1}N|^{-s} + \mathrm{e}^1 + \mathrm{e}^2$. Estimate (7.105) can thus be used as follows. Note first that we expect the term $|B_\Omega^{-1}N|^{-s}$ to be negligible. Indeed, if N is reasonably large with respect to Ω, we can assume the existence of s, such that (7.104) holds and $|B_\Omega^{-1}N|^{-s} \le \varepsilon^3$. In practice, N must be set so that the corresponding grid captures the initial conditions (and the source). We then fix an order of tolerance τ for the error, where $\varepsilon \le \tau \le 1$. From the second term in e_1, we set $|\Delta x| = \tau\varepsilon$. For simplicity we set $\Delta x_1 = \ldots = \Delta x_d = d^{-1/2}\tau\varepsilon$ (this makes sense if the variation $x \mapsto a(x,y)$ is isotropic). The second term in e_2 then reads $h^q|1/\Delta x|^2 = h^q d^2(\tau\varepsilon)^{-2}$ and the imposition of the tolerance brings $h = (d^{-2}\tau^3\varepsilon^2)^{1/q}$. We verify that this value for h ensures the first and third terms of e_1 and the first term of e_2 to be of order τ. To summarize, if we set the parameters of the method as

$$h = (d^{-2}\tau^3\varepsilon^2)^{1/q}, \qquad \Delta x_\nu = d^{-1/2}\tau\varepsilon, \tag{7.106}$$

then (7.105) ensures the error $\|u^\varepsilon - u_N\|_{\mathrm{L}^\infty(0,T^\varepsilon;W)}$ to be of order τ.

Remark 7.2.5. In Section 6.4, we discussed the necessity of the operators $\varepsilon L^1 = \varepsilon b^{10} - \varepsilon\partial_i\big(a_{ij}^{21}\partial_j \cdot\big)$ and $\varepsilon^2 L^{2,1} = \varepsilon^2 b^{20}\partial_t^2 - \varepsilon^2\partial_i\big(a_{ij}^{22})\partial_j \cdot\big)$ in the effective equations. In particular, in the tested numerical examples, these operators are unnecessary to describe the observed long time effects. We note that if we drop the approximation of these effective tensors in the method, its cost is significantly reduced. Indeed, first, the approximations of the cell problems for $\{\theta_i^1\}_{i=1}^d$ are in this case unnecessary (see (7.88)). Then, we verify that if we drop εL^1 and $\varepsilon^2 L^{2,1}$, the term e_2 disappear from the error estimate (7.105). Hence, the severe restrictions on h imposed by e_2 are relaxed and the computational cost is reduced. Nevertheless, as no applicable criterion were found to determine whether the operators can be dropped, we are not able to provide a rigorous numerical procedure that benefits this gain.

7.2.4 Proof of the a priori error estimate (Theorem 7.2.3)

The proof of Theorem 7.2.3 is divided into three parts. In the first part, we estimate the error made in the approximation of the effective tensors (Step 1 of the method). In particular, we use standard FE error estimates for the approximated correctors and we provide an estimate for the error in the eigenvalues involved in the definition of the tensors (see Lemma 7.2.8). The second part consists in the estimation of the error made in the approximation of the bilinear forms $(\cdot,\cdot)_{\mathcal{S}}$ and A (Lemmas 7.2.9 and 7.2.10). Finally, in the third part, we derive the a priori error estimate for the spectral homogenization method. For this last step, we follow a similar process as in Section 7.1.4: we define an elliptic projection and split the error in two parts that we estimate separately (Lemmas 7.2.11 and 7.2.12). In particular, as the definition of the elliptic projection allows to avoid the use of the Poincaré inequality, we obtain an error estimate that can be applied in pseudoinfinite domains.

Part 1 – Error in the effective tensors

In the first part of the proof of Theorem 7.2.3, we provide error estimates for the approximated effective tensors (Lemma 7.2.6). In particular, the main difficulty is to estimate the error for the approximation of the terms of the form $\left\{\frac{\lambda_{\min}(B)}{\lambda_{\min}(A)}\right\}_+$ (see Lemma 7.2.8). Indeed, the evaluation of these terms involves two obstacles. First, the eigenvalues $\lambda_{\min}(\cdot)$ are approximated by the eigenvalues of approximated matrices. And second, the maximum on the whole domain $\{\cdot\}_+$ is approximated by a maximum on the grid G_M.

Lemma 7.2.6. *Assume that a satisfies the regularity*

$$a \in \mathcal{C}^0(\bar{\Omega}; \mathrm{W}^{q,\infty}(Y)) \cap \mathcal{C}^1(\bar{\Omega}; \mathrm{W}^{q-1,\infty}(Y)) \cap \mathcal{C}^3(\bar{\Omega}; \mathrm{L}^{\infty}(Y)),$$

and that $r(\Delta x) = \max_\nu \Delta x_\nu / \min_\nu \Delta x_\nu$ is bounded. Then, for any $1 \leq i,j,k,l \leq d$ and any $x_m \in G_M$, the following estimates hold

$$\left|a^0_{ij}(x_m) - a^{0h}_{ij}(x_m)\right| \leq Ch^{2q}, \tag{7.107a}$$
$$\left|b^{10} - b^{10h}\right| \leq C\big(|\Delta x| + h^q|1/\Delta x| + h^q\big), \tag{7.107b}$$
$$\left|\bar{a}^{12}_{ij}(x_m) - \bar{a}^{12h}_{ij}(x_m)\right| \leq C\big(|\Delta x| + h^q|1/\Delta x| + h^q\big), \tag{7.107c}$$
$$\left|\bar{a}^{24}_{ijkl}(x_m) - \bar{a}^{24h}_{ijkl}(x_m)\right| \leq C\big(|\Delta x| + h^q\big), \tag{7.107d}$$
$$\left|b^{22}_{ij}(x_m) - b^{22h}_{ij}(x_m)\right| \leq C\big(|\Delta x| + h^q\big), \tag{7.107e}$$
$$\left|b^{20} - b^{20h}\right| \leq C\Big(|\Delta x| + h^q\big(|1/\Delta x|^2 + |1/\Delta x|\big) + h^q\Big), \tag{7.107f}$$
$$\left|\bar{a}^{22}_{ij}(x_m) - \bar{a}^{22h}_{ij}(x_m)\right| \leq C\Big(|\Delta x| + h^q\big(|1/\Delta x|^2 + |1/\Delta x|\big) + h^q\Big), \tag{7.107g}$$

where $|1/\Delta x| = \sqrt{\sum_{\nu=1}^d 1/\Delta x_\nu^2}$ and C depends only on d, Y, λ, $r(\Delta x)$, $\|a\|_{\mathcal{C}^0(\bar{\Omega};\mathrm{W}^{q,\infty}(Y))}$, $\|a\|_{\mathcal{C}^1(\bar{\Omega};\mathrm{W}^{q-1,\infty}(Y))}$, and $\|a\|_{\mathcal{C}^3(\bar{\Omega};\mathrm{L}^{\infty}(Y))}$.

To derive the error estimates involving eigenvalues, we need the two following lemmas.

Lemma 7.2.7. *Let $A, \bar{A}$ be two symmetric positive definite matrices, and let $B, \bar{B}$ be two symmetric matrices. Then*

$$\left|\frac{\lambda_{\min}(B)}{\lambda_{\min}(A)} - \frac{\lambda_{\min}(\bar{B})}{\lambda_{\min}(\bar{A})}\right| \leq C\Big(\|B - \bar{B}\|_F + \|A - \bar{A}\|_F\Big),$$

where C depends on $\lambda_{\min}(A)^{-1}, \lambda_{\min}(\bar{A})^{-1}$ and $\lambda_{\min}(\bar{B})$ and $\|\cdot\|_F$ is the Frobenius norm.

Proof. First, note that for two symmetric matrices $D, \bar{D}$, the estimate $|\lambda_{\min}(D) - \lambda_{\min}(\bar{D})| \leq d\|D - \bar{D}\|_F$ holds. Denoting, $a = \lambda_{\min}(A)$, $\bar{a} = \lambda_{\min}(\bar{A})$, $b = \lambda_{\min}(b)$, and $\bar{b} = \lambda_{\min}(\bar{B})$, we then use this estimate in the equality

$$a^{-1}b - \bar{a}^{-1}\bar{b} = a^{-1}(b - \bar{b}) + \bar{b}(a\bar{a})^{-1}(\bar{a} - a),$$

and obtain the lemma. □

Lemma 7.2.8. *Let $A, B : \mathbb{R}^d \to \mathrm{Sym}^2(\mathbb{R}^d)$ be two matrix functions of class $\mathcal{C}^1$ and let $\big\{A^h(x_m), B^h(x_m)\big\}_{x_m \in G_M}$ be bounded matrix functions given on the grid. We assume that*

$A(x)$ and $A^h(x_m)$ are positive definite for any $x \in \Omega$ and $x_m \in G_M$. Then the following estimate holds

$$\Big| \sup_{x\in\Omega} \Big\{ -\frac{\lambda_{\min}(B(x))}{\lambda_{\min}(A(x))} \Big\}_+ - \max_{x_m\in G_M} \Big\{ -\frac{\lambda_{\min}(B^h(x_m))}{\lambda_{\min}(A^h(x_m))} \Big\}_+ \Big| \tag{7.108}$$
$$\leq C\Big(|\Delta x| + \max_{x_m\in G_M} \Big\{ \|B(x_m) - B^h(x_m)\|_F + \|A(x_m) - A^h(x_m)\|_F \Big\} \Big),$$

where the constant C depends on $\|\lambda_{\min}(A)^{-1}\|_{\mathrm{L}^\infty}$, $\|\lambda_{\min}(A^h)^{-1}\|_{\mathrm{L}^\infty}$, $\|\lambda_{\min}(B)\|_{\mathrm{L}^\infty}$, $\|\lambda_{\min}(B^h)\|_{\mathrm{L}^\infty}$, $\|A_{ij}\|_{\mathcal{C}^1}$, $\|B_{ij}\|_{\mathcal{C}^1}$ and d.

Proof. Let us introduce some notations. For any $x_m \in G_M$, let $K(x_m)$ be the element defined as $K(x_m) = \big\{x \in \Omega : x = x_m + t, t_\nu \in [0, \Delta x_\nu[\big\}$. We verify that the diameter of $K(x_m)$ is $|\Delta x|$ and $\Omega = \mathrm{int}\big(\cup_{x_m\in G_M} \overline{K(x_m)}\big)$. Using the shorthand notation $H = |\Delta x|$, we define the operator P_H onto the space of piecewise constant functions

$$v \in \mathrm{L}^\infty(\Omega) \mapsto P_H v, \qquad P_H v(x) = \sum_{x_m\in G_M} v(x_m) \mathbb{1}_{K(x_m)}(x),$$

where $\mathbb{1}_{K(x_m)}(x)$ is the indicator function, $\mathbb{1}_{K(x_m)}(x) = 1$ if $x \in K(x_m)$ and 0 otherwise. Note that $P_H v \in \mathrm{L}^\infty(\Omega)$ and it satisfies $P_H v(x_m) = v(x_m)$. For a function $v \in \mathrm{W}^{1,\infty}(K(x_m))$, we verify that for any $x \in K(x_m)$, $|v(x) - v(x_m)| \leq |\Delta x| |v|_{\mathrm{W}^{1,\infty}(K(x_m))}$. Hence, $P_H v$ satisfies the following properties for any $v \in \mathrm{W}^{1,\infty}(\Omega)$

$$\|P_H v\|_{\mathrm{L}^\infty(\Omega)} = \max_{x_m\in G_M} |v(x_m)|, \qquad \|v - P_H v\|_{\mathrm{L}^\infty(\Omega)} \leq |\Delta x| |v|_{\mathrm{W}^{1,\infty}(\Omega)}. \tag{7.109}$$

Denoting $R(x) = -\frac{\lambda_{\min}(B(x))}{\lambda_{\min}(A(x))}$ and $R^h(x_m) = -\frac{\lambda_{\min}(B^h(x_m))}{\lambda_{\min}(A^h(x_m))}$, the left hand side of (7.108) is split as

$$\mathrm{e} = \big| \|\{R\}_+\|_{\mathrm{L}^\infty(\Omega)} - \max_{x_m\in G_M} \{R^h(x_m)\}_+ \big| \leq \mathrm{e}_1 + \mathrm{e}_2,$$
$$\mathrm{e}_1 = \Big| \|\{R\}_+\|_{\mathrm{L}^\infty(\Omega)} - \|\{P_H R\}_+\|_{\mathrm{L}^\infty(\Omega)} \Big|, \qquad \mathrm{e}_2 = \Big| \max_{x_m\in G_M} \{R(x_m)\}_+ - \max_{x_m\in G_M} \{R^h(x_m)\}_+ \Big|.$$

In order to use (7.109) on e_1, we need to verify that $R \in \mathrm{W}^{1,\infty}(\Omega)$. It is sufficient to prove that R is Lipschitz continuous, which is done using Lemma 7.2.7: for any $x, \bar{x} \in \Omega$

$$|R(x) - R(\bar{x})| \leq C\Big(\|B(x) - \bar{B}(\bar{x})\|_F + \|A(x) - \bar{A}(\bar{x})\|_F \Big) \leq L|x - \bar{x}|,$$

where L depends on $\|\lambda_{\min}(A)\|_{\mathrm{L}^\infty}$, $\|\lambda_{\min}(B)\|_{\mathrm{L}^\infty}$, $\|A_{ij}\|_{\mathcal{C}^1}$, $\|B_{ij}\|_{\mathcal{C}^1}$ and d. Using the reverse triangle inequality, the fact that $|\{a\}_+ - \{b\}_+| \leq |a - b|$ and (7.109), we thus have

$$\mathrm{e}_1 \leq \|\{R\}_+ - \{P_H R\}_+\|_{\mathrm{L}^\infty} \leq \|R - P_H R\|_{\mathrm{L}^\infty} \leq |R|_{\mathrm{W}^{1,\infty}} |\Delta x| \leq C|\Delta x|.$$

Following a similar argument, we have

$$\mathrm{e}_2 \leq \max_{x_m\in G_M} \big| \{R(x_m)\}_+ - \{R^h(x_m)\}_+ \big| \leq \max_{x_m\in G_M} \big| R(x_m) - R^h(x_m) \big|$$
$$\leq \max_{x_m\in G_M} \Big\{ \|B(x_m) - B^h(x_m)\|_F + \|A(x_m) - A^h(x_m)\|_F \Big\},$$

where we used Lemma 7.2.7 in the last inequality. Combining the two last estimates gives (7.108) and the proof of the lemma is complete. □

We now have all the technical tools to prove the error estimates (7.107).

Proof of Lemma 7.2.6. Recall once and for all that the exact tensors are defined in (7.73), (7.74), (7.75), and (7.76), and their approximations in (7.83), (7.86), (7.87), and (7.88). The exact cell functions are defined in (6.72) and (6.75) and their approximations in (7.82), (7.84) and (7.85) The $\mathrm{L}^2(Y)$ inner product is simply denoted $(\cdot,\cdot)$ and the corresponding norm is denoted $\|\cdot\|_Y$. Furthermore, for the sake of clarity, let us assume that $|Y|=1$, so that $\langle vw\rangle_Y=(v,w)$ for any $v,w\in\mathrm{L}^2(Y)$. In the whole proof, C denotes a generic constant that depends only on d, Y, λ, $C_{r(\Delta x)}$ $\|a\|_{\mathcal{C}^0(\bar\Omega;\mathrm{W}^{q,\infty}(Y))}$, $\|a\|_{\mathcal{C}^1(\bar\Omega;\mathrm{W}^{q-1,\infty}(Y))}$, and $\|a\|_{\mathcal{C}^3(\bar\Omega;\mathrm{L}^{\infty}(Y))}$.

Using (6.108), the regularity of a ensures (at least) the following regularities:

$$\chi_i,\theta^0_{ij}\in\mathcal{C}^0(\bar\Omega;\mathrm{H}^{q+1}(Y))\cap\mathcal{C}^3(\bar\Omega;\mathrm{H}^1(Y)),\quad \theta^1_i\in\mathcal{C}^0(\bar\Omega;\mathrm{H}^{q+1}(Y)),$$
$$a^0_{ij}\in\mathcal{C}^4(\bar\Omega),\quad \check a^{12}_{ij},\check a^{24}_{ijkl},\check a^{22}_{ij}\in\mathcal{C}^1(\bar\Omega).$$

Let us fix an arbitrary grid point $x_m\in G_M$. From now on, all the tensors, cell functions and their approximations are evaluated at x_m.

We now prove the estimates in (7.107) one after another. We begin with the error estimate for a^0 (7.107a). Using the cell problem for χ_i and for χ^h_i, we verify that

$$a^0_{ij}-a^{0h}_{ij}=\big(ae_i,\nabla_y(\chi_j-\chi^h_j)\big)-\big(a\nabla_y\chi_i,\nabla_y(\chi_j-\chi^h_j)\big)=\big(a\nabla_y(\chi^h_i-\chi_i),\nabla_y(\chi_j-\chi^h_j)\big),$$

which, combined with (7.100), gives (7.107a). Next, we prove the error estimates for $\bar a^{24}_{ijkl}$ and b^{22}_{ij} in (7.107d) and (7.107e). Using (7.100) and (7.101), we have

$$\begin{aligned}\big|\check a^{24}_{ijkl}-\check a^{24h}_{ijkl}\big|=\big|S^{2,2}_{ij,kl}\big\{\big(a(\chi_i-\chi^h_i),\chi_l\big)+\big(a\chi^h_i,\chi_l-\chi^h_l\big)-\big(a\nabla_y(\bar\theta^{0h}_{ji}-\bar\theta^0_{ji}),\nabla_y\bar\theta^0_{kl}\big)\\ -\big(a\nabla_y\bar\theta^{0h},\nabla_y(\bar\theta^0_{kl}-\bar\theta^{0h}_{kl})\big)\big\}\big|\le Ch^q.\end{aligned}\tag{7.110}$$

Estimate (7.107a) ensures

$$\big|a^0_{jk}a^0_{il}-a^{0h}_{jk}a^{0h}_{il}\big|\le\big|a^0_{jk}(a^0_{il}-a^{0h}_{il})+(a^0_{jk}-a^{0h}_{jk})a^{0h}_{il}\big|\le Ch^{2q}.\tag{7.111}$$

Note that for any major and minor symmetric $q\in\mathrm{Ten}^4(\mathbb{R}^d)$, we have $\|M(q)\|_F^2\le C(d)\sum_{ijkl}q^2_{ijkl}$. Hence, applying Lemma 7.2.8, we get

$$|\delta-\delta^h|\le C\big(|\Delta x|+h^q\big).\tag{7.112}$$

Writing then

$$\bar a^{24}_{ijkl}-\bar a^{24h}_{ijkl}=\check a^{24}_{ijkl}-\check a^{24h}_{ijkl}+S^{2,2}_{ij,kl}\big\{(\delta-\delta^h)a^0_{jk}a^0_{il}+\delta^h(a^0_{jk}a^0_{il}-a^{0h}_{jk}a^{0h}_{il})\big\},$$

and using (7.110), (7.111), and (7.112), we obtain (7.107d). Similarly, writing

$$b^{22}_{ij}-b^{22h}_{ij}=\big(\chi_i,\chi_j-\chi^h_j\big)+\big(\chi_i-\chi^h_i,\chi^h_j\big)+\big(\delta-\delta^h\big)a^{0h}_{ij}+\delta\big(a^0_{ij}-a^{0h}_{ij}\big),$$

and using (7.100), (7.112) and (7.107a) proves (7.107e). Next, we prove the estimate for the error in the approximation of a^{12}_{ij} and b^{10} in (7.107c) and (7.107b). Using (7.100), we have

$$\big|p^{13}_{ijk}-p^{13h}_{ijk}\big|\le\big|\big(a\nabla_y(\chi_k-\chi^h_k),e_j\chi_i\big)\big|+\big|\big(a\nabla_y\chi^h_k,e_j(\chi_i-\chi^h_i)\big)\big|\le Ch^q.$$

Then, using this estimate and (7.80), we obtain

$$\big|\partial_r p^{13}_{rij}-D_rp^{13h}_{rij}\big|\le\big|\partial_rp^{13}_{rij}-D_rp^{13}_{rij}\big|+\big|D_rp^{13}_{mij}-D_rp^{13h}_{rij}\big|\le C\big(|\Delta x|^2+h^q|1/\Delta x|\big),\tag{7.113}$$

where we denoted $|1/\Delta x| = \sqrt{\sum_{\nu=1}^{d} 1/\Delta x_\nu^2}$. Next, using (7.80) and (7.100), we verify that

$$\begin{aligned}\|\nabla_x\chi_i - D_x\chi_i^h\|_Y &\le \|\nabla_x\chi_i - D_x\chi_i\|_Y + \|D_x\chi_i - D_x\chi_i^h\|_Y \\ &\le C\big(|\Delta x|^2 + h^{q+1}|1/\Delta x|\big),\end{aligned} \tag{7.114}$$

which, combined with (7.100) implies

$$\begin{aligned}\big|q_{ij}^{12} - q_{ij}^{12h}\big| &\le \big|\big(a(\nabla_y\chi_j^h + e_j), \nabla_x\chi_i - D_x\chi_i^h\big)\big| + \big|\big(a\nabla_y(\chi_j - \chi_j^h), \nabla_x\chi_i\big)\big| \\ &\le C\big(|\Delta x|^2 + h^{q+1}|1/\Delta x| + h^q\big).\end{aligned} \tag{7.115}$$

Combining (7.113) and (7.115) brings $\big|\check{a}_{ij}^{12} - \check{a}_{ij}^{12h}\big| \le C\big(|\Delta x|^2 + h^q|1/\Delta x| + h^q\big)$. Hence, using (7.107a), Lemma 7.2.8 ensures $\big|b^{10} - b^{10h}\big| \le C\big(|\Delta x| + h^q|1/\Delta x| + h^q\big)$, which proves (7.107b). We thus obtain

$$\big|a_{ij}^{12} - a_{ij}^{12h}\big| \le \big|\check{a}_{ij}^{12} - \check{a}_{ij}^{12h}\big| + \big|a_{ij}^0\big(b^{10} - b^{10h}\big)\big| + \big|\big(a_{ij}^0 - a_{ij}^{0h}\big)b^{10h}\big| \le C\big(|\Delta x| + h^q|1/\Delta x| + h^q\big),$$

and that proves (7.107c). In order to prove (7.107f) and (7.107g), we derive an error estimate for θ_i^{1h}. First, similarly as (7.114), we prove

$$\begin{aligned}\|\nabla_x\cdot(a(\nabla_y\chi_i + e_i)) - D_x\cdot(a(\nabla_y\chi_i^h + e_i))\|_Y &\le C\big(|\Delta x|^2 + h^q|1/\Delta x|\big), \\ |\nabla_x\cdot(a^0 e_i) - D_x\cdot(a^{0h}e_i)| &\le C\big(|\Delta x|^2 + h^{2q}|1/\Delta x|\big).\end{aligned} \tag{7.116}$$

Then, as we assume $\Delta x_\nu \le C$ to be bounded independently of $\operatorname{diam}(\Omega)$, using (7.114), we bound

$$\|D_x\chi_i^h\|_Y \le \|\nabla_x\chi_i\|_Y + \|\nabla_x\chi_i - D_x\chi_i^h\|_Y \le C,$$

and similarly, using (7.116), we show that $\|D_x\cdot(a(\nabla_y\chi_i^h + e_i))\|_Y \le C$ and $\|D_x\cdot(a^{0h}e_i)\|_Y \le C$. Hence, thanks to (7.116) and (7.114), standard FEM error estimates ensure

$$\|\theta_i^{1h}\|_{\mathrm{H}^1(Y)} \le C, \qquad \|\theta_i^1 - \theta_i^{1h}\|_{\mathrm{H}^1(Y)} \le C\big(h^q + |\Delta x|^2 + h^q|1/\Delta x|\big). \tag{7.117}$$

We now need to estimate individually the numerous terms of $\check{a}_{ij}^{22} - \check{a}_{ij}^{22h}$. First, using (7.100), (7.114), (7.101) and (7.117), we obtain

$$\begin{aligned}\big|\bar{p}_{ijk}^{23} - \bar{p}_{ijk}^{23h}\big| &\le \big|S_{ij}^2\big\{\big(ae_j(\chi_i - \chi_i^h), \nabla_x\chi_k\big) + \big(ae_j\chi_i^h, D_x\chi_k^h\big)\big\}\big| + \big|\big(a\nabla_y(\bar{\theta}_{ji}^{0h} - \bar{\theta}_{ji}^0), \nabla_y\theta_k^1\big)\big| \\ &\quad + \big|\big(a\nabla_y\bar{\theta}_{ji}^{0h}, \nabla_y(\theta_k^{1h} - \theta_k^1)\big)\big| \le C\big(|\Delta x|^2 + h^q|1/\Delta x| + h^q\big).\end{aligned}$$

Hence, we have

$$\begin{aligned}\big|\partial_m\bar{p}_{mij}^{23} - D_m\bar{p}_{mij}^{23h}\big| &\le \big|\partial_m\bar{p}_{mij}^{23} - D_m\bar{p}_{mij}^{23}\big| + \big|D_m\bar{p}_{mij}^{23} - D_m\bar{p}_{mij}^{23h}\big| \\ &\le C\big(|\Delta x|^2 + |\Delta x|^2|1/\Delta x| + h^q|1/\Delta x|^2 + h^q|1/\Delta x|\big).\end{aligned}$$

Note that the hypothesis $\Delta x_{\max}/\Delta x_{\min} \le C$ implies $|\Delta x|^2|1/\Delta x| \le |\Delta x|$ and thus

$$\big|\partial_m\bar{p}_{mij}^{23} - D_m\bar{p}_{mij}^{23h}\big| \le C\big(|\Delta x| + h^q|1/\Delta x|^2 + h^q|1/\Delta x|\big). \tag{7.118}$$

Next, (7.114) and (7.101) imply that

$$\begin{aligned}\big|p_{ij}^{12} - p_{ij}^{12h}\big| &\le \big|\big(a(\nabla_x\chi_j - D_x\chi_j^h), \nabla_x\chi_i\big)\big| + \big|\big(aD_x\chi_j^h, \nabla_x\chi_i - D_x\chi_i^h\big)\big| + \big|\big(a\nabla_y(\theta_i^{1h} - \theta_i^1), \nabla_y\theta_j^1\big)\big| \\ &\quad + \big|\big(a\nabla_y\theta_i^{1h}, \nabla_y(\theta_j^1 - \theta_j^{1h})\big)\big| \le C\big(|\Delta x|^2 + h^q|1/\Delta x| + h^q\big).\end{aligned} \tag{7.119}$$

Further, using (7.112) and (7.116), we have

$$\begin{aligned}\big|\delta\partial_s a_{ri}^0\partial_r a_{sj}^0 - \delta^h D_s a_{ri}^{0h} D_r a_{sj}^{0h}\big| &\le C\Big(\big|\delta - \delta^h\big| + \sum_{rs}\big|\partial_s a_{ri}^0 - D_s a_{ri}^{0h}\big| + \sum_{rs}\big|\partial_r a_{sj}^0 - D_r a_{sj}^{0h}\big|\Big) \\ &\le C\big(|\Delta x| + h^{2q}|1/\Delta x| + h^q\big),\end{aligned} \tag{7.120}$$

and a similar estimate holds for $\big|\delta\partial_r a^0_{rs}\partial_s a^0_{ij} - \delta^h D_r a^{0h}_{rs} D_s a^{0h}_{ij}\big|$. Thanks to (7.81), we verify in a similar way as (7.116) that $\sum_{rs}\big|\partial^2_{rs}a^0_{ij} - D^2_{rs}a^{0h}_{ij}\big| \le C\big(|\Delta x|^2 + h^{2q}|1/\Delta x|^2\big)$, and thus

$$\begin{aligned}\big|\delta a^0_{rs}\partial^2_{rs}a^0_{ij} - \delta^h a^{0h}_{rs} D^2_{rs}a^{0h}_{ij}\big| &\le C\Big(\big|\delta-\delta^h\big| + \sum_{rs}\big|a^0_{rs}-a^{0h}_{rs}\big| + \sum_{rs}\big|\partial^2_{rs}a^0_{ij} - D^2_{rs}a^{0h}_{ij}\big|\Big)\\ &\le C\big(|\Delta x| + h^{2q}|1/\Delta x|^2 + h^q\big). \end{aligned} \tag{7.121}$$

Combining now (7.118) (7.119), (7.120), (7.121) and the estimate for b^{10}, we obtain

$$\big|\check{a}^{22}_{ij} - \check{a}^{22h}_{ij}\big| \le C\Big(|\Delta x| + h^q\big(|1/\Delta x|^2 + |1/\Delta x|\big) + h^q\Big).$$

Applying Lemma 7.2.8, we obtain the bound (7.107f) for $|b^{20} - b^{20h}|$ and that proves (7.107g). The proof of Lemma 7.2.6 is complete. □

Part 2 – Error in the bilinear forms

In the second part of the proof of Theorem 7.2.3, we estimate the errors in the approximation of the forms $(\cdot,\cdot)_{\mathcal{S}}$ and A, defined in (7.77), by the forms $(\cdot,\cdot)_Q$ and A^h_N, defined in (7.96). In particular, we use the error estimates on the effective tensors obtained in the first part (Lemma 7.2.6). Let us define

$$\mathrm{e}_{a^0} = \max_{\substack{1\le ij\le d\\ x_n\in G_N}}\big|a^0_{ij}(x_n) - a^{0h}_{ij}(x_n)\big|,$$

and similarly $\mathrm{e}_{\bar{a}^{12}}$, $\mathrm{e}_{b^{10}}$, $\mathrm{e}_{\bar{a}^{24}}$, $\mathrm{e}_{b^{22}}$, $\mathrm{e}_{\bar{a}^{22}}$, and $\mathrm{e}_{b^{20}}$. Recall that F_Ω is the bijective affine mapping defined in (7.89) as

$$F_\Omega : (0,2\pi)^d \to \Omega, \quad \bar{x}\mapsto F_\Omega(\bar{x}) = B_\Omega\bar{x} + a,$$

where B_Ω is the diagonal matrix defined by $(B_\Omega)_{jj} = (b_j - a_j)/(2\pi)$.

Lemma 7.2.9. *Assume that for some $s \ge (d+1)/2$, $a(x,y)\in\mathcal{C}^s(\bar{\Omega};\mathrm{L}^\infty(Y))$. Then, for any $v\in\mathrm{W}_{\mathrm{per}}(\Omega)\cap\mathrm{H}^{s+1}(\Omega)$ and $w_N\in\mathring{V}_N(\Omega)$, the bilinear form $(\cdot,\cdot)_Q$ satisfies*

$$\big|\big(v,w_N\big)_{\mathcal{S}} - \big(\mathring{I}_N v, w_N\big)_Q\big| \le C\bigg(\frac{1}{|B_\Omega^{-1}N|^s} + \mathrm{e}_{b^i}\bigg)\|v\|_{\mathrm{H}^{s+1}(\Omega)}\|w_N\|_{\mathrm{H}^1(\Omega)},$$

where $\mathrm{e}_{b^i} = \varepsilon\mathrm{e}_{b^{10}} + \varepsilon^2\mathrm{e}_{b^{20}} + \varepsilon^2\mathrm{e}_{b^{22}}$ and the constant C depends on d, s, $r(N)$, λ, Y, and $\|a\|_{\mathcal{C}^s(\bar{\Omega};\mathrm{L}^\infty(Y))}$.

Proof. First, we verify thanks to (6.108) that the regularity of a ensures $b^{22}\in\mathcal{C}^s(\bar{\Omega})$. Hence, $(\cdot,\cdot)_{\mathcal{S}}$ is bounded. We denote $\rho = 1+\varepsilon b^{10}+\varepsilon^2 b^{20}$ and $\rho^h = 1+\varepsilon b^{10h}+\varepsilon^2 b^{20h}$, and split the error as

$$\big|\big(v,w_N\big)_{\mathcal{S}} - \big(\mathring{I}_N v, w_N\big)_Q\big| \le \mathrm{e}^1_N + \mathrm{e}^2_N + \mathrm{e}^1_h + \mathrm{e}^2_h,$$

where

$$\begin{aligned}
\mathrm{e}^1_N &= \big|\big(\rho v, w_N\big)_{\mathrm{L}^2} - \big(\rho\mathring{I}_N v, w_N\big)_N\big|,\\
\mathrm{e}^1_h &= \big|\big(\rho\mathring{I}_N v, w_N\big)_N - \big(\rho^h\mathring{I}_N v, w_N\big)_N\big|,\\
\mathrm{e}^2_N &= \big|\big(\varepsilon^2 b^{22}\nabla v, \nabla w_N\big)_{\mathrm{L}^2} - \big(\varepsilon^2 I_N b^{22} * \nabla\mathring{I}_N v, \nabla w_N\big)_N\big|,\\
\mathrm{e}^2_h &= \big|\big(\varepsilon^2 I_N b^{22} * \nabla\mathring{I}_N v, \nabla w_N\big)_N - \big(\varepsilon^2 b^{22h} * \nabla\mathring{I}_N v, \nabla w_N\big)_N\big|.
\end{aligned}$$

Using (7.91) and (7.95), we find

$$\mathrm{e}^1_N \le |\rho|\big|\big(v - \mathring{I}_N v, w_N\big)_{\mathrm{L}^2}\big| \le C\frac{1}{|B_\Omega^{-1}N|^s}|\rho||v|_{\mathrm{H}^s}\|w_N\|_{\mathrm{L}^2}.$$

Note that (7.93) and the definition of I_N imply that $I_N(b^{22}\nabla v) = I_N b^{22} * I_N(\nabla v)$. Furthermore, as $\mathring{I}_N v = I_N v - \langle I_N v\rangle_\Omega$, we have $\nabla \mathring{I}_N v = \nabla I_N v$. Hence, using (7.91), we bound

$$\begin{aligned}\mathrm{e}_N^2 &\le \big|\big(\varepsilon^2 b^{22}\nabla v, \nabla w_N\big)_{\mathrm{L}^2} - \big(\varepsilon^2 I_N(b^{22}\nabla v), \nabla w_N\big)_{\mathrm{L}^2}\big| \\ &\quad + \big|\big(\varepsilon^2 I_N b^{22} * I_N(\nabla v), \nabla w_N\big)_N - \big(\varepsilon^2 I_N b^{22} * \nabla I_N v, \nabla w_N\big)_N\big| =: \mathrm{e}_N^{2,1} + \mathrm{e}_N^{2,2}.\end{aligned}$$

Thanks to (7.92), we verify

$$\begin{aligned}\mathrm{e}_N^{2,1} &\le C\frac{1}{|B_\Omega^{-1}N|^s}\varepsilon^2 |b^{22}\nabla I_N v|_{\mathrm{H}^s}|w_N|_{\mathrm{H}^1} \le C\frac{\varepsilon^2\|b^{22}\|_{\mathcal{C}^s}}{|B_\Omega^{-1}N|^s}\|v\|_{\mathrm{H}^{s+1}}|w_N|_{\mathrm{H}^1}, \\ \mathrm{e}_N^{2,1} &\le C\|b^{22}\|_{\mathcal{C}^0}\big(\|I_N(\nabla v) - \nabla v\|_{\mathrm{L}^2} + |v - I_N v|_{\mathrm{L}^2}\big)|w_N|_{\mathrm{H}^1} \le C\frac{\varepsilon^2\|b^{22}\|_{\mathcal{C}^0}}{|B_\Omega^{-1}N|^s}\|v\|_{\mathrm{H}^{s+1}}|w_N|_{\mathrm{H}^1}.\end{aligned}$$

Then, using (7.91), we have

$$\begin{aligned}\mathrm{e}_h^1 &= H^{\mathbb{1}}\Big|\sum_{x_n\in G_N}(\rho-\rho^h)\mathring{I}_N v(x_n)\overline{w_N(x_n)}\Big| \\ &\le (\varepsilon \mathrm{e}_{b^{10}} + \varepsilon^2 \mathrm{e}_{b^{20}})\big(\mathring{I}_N v, w_N\big)_N \le C(\varepsilon \mathrm{e}_{b^{10}} + \varepsilon^2 \mathrm{e}_{b^{20}})\|v\|_{\mathrm{L}^2}\|w_N\|_{\mathrm{L}^2}.\end{aligned}$$

Similarly, $\mathrm{e}_h^2 \le C\varepsilon^2 \mathrm{e}_{b^{22}}|v|_{\mathrm{H}^1}|w_N|_{\mathrm{H}^1}$. Combining the estimates for e_N^1, e_N^2, e_h^1 and e_h^2, we obtain the desired estimate and the proof of the lemma is complete. □

Lemma 7.2.10. *Assume that for some $s \ge (d+1)/2$, $a \in \mathcal{C}^{s+2}(\bar\Omega; \mathrm{L}^\infty(Y))$. Then, for any $v \in \mathrm{W}_{\mathrm{per}}(\Omega)\cap \mathrm{H}^{s+2}(\Omega)$ and $w_N \in \mathring{V}_N(\Omega)$, the bilinear form A_N^h satisfies*

$$\big|A\big(v, w_N\big) - A_N^h\big(\mathring{I}_N v, w_N\big)\big| \le C\bigg(\frac{1}{|B_\Omega^{-1}N|^s} + \mathrm{e}_{a^i}\bigg)\bigg(\sum_{\sigma=1}^{s+2}|v|_{\mathrm{H}^\sigma(\Omega)}\bigg)\Big(|w_N|^2_{\mathrm{H}^1(\Omega)} + \varepsilon^2|w_N|^2_{\mathrm{H}^2(\Omega)}\Big)^{1/2}, \tag{7.122}$$

where $\mathrm{e}_{a^i} = \mathrm{e}_{a^0} + \varepsilon \mathrm{e}_{\bar a^{12}} + \varepsilon^2 \mathrm{e}_{\bar a^{22}} + \varepsilon \mathrm{e}_{\bar a^{24}}$, and the constant C depends on d, s, $r(N)$, λ, Y, and $\|a\|_{\mathcal{C}^{s+2}(\bar\Omega;\mathrm{L}^\infty(Y))}$.

Proof. As $\nabla\mathring{I}_N v = \nabla I_N v$, it is sufficient to prove the estimate for $A_N^h\big(I_N v, w_N\big)$. Thanks to (6.108), the regularity of a ensures that $a^0, \bar a^{12}, \bar a^{22}, \bar a^{24} \in \mathcal{C}^s(\bar\Omega)$. We denote the tensors $c = a^0 + \varepsilon\bar a^{12} + \varepsilon^2\bar a^{22}$, $c^h = a^{0h} + \varepsilon\bar a^{12h} + \varepsilon^2\bar a^{22h}$, and define the following bilinear form on $\mathring{V}_N(\Omega)\times\mathring{V}_N(\Omega)$:

$$A_N\big(v_N, w_N\big) = \big(I_N c * \nabla v_N, \nabla w_N\big)_N + \big(\varepsilon^2 I_N \bar a^{24} * \nabla^2 v_N, \nabla^2 w_N\big)_N.$$

Let us split the forms as $A = A^1 + A^2$, $A_N = A_N^1 + A_N^2$, and $A_N^h = A_N^{h,1} + A_N^{h,2}$, where

$$A^1\big(v, w\big) = \big(c\nabla v, \nabla w\big)_{\mathrm{L}^2}, \quad A^2\big(v, w\big) = \big(\varepsilon^2\bar a^{24}\nabla^2 v, \nabla^2 w\big)_{\mathrm{L}^2},$$

and A_N^1, A_N^2, $A_N^{h,1}$, $A_N^{h,2}$ are defined similarly. We split the error as

$$\big|A\big(v, w_N\big) - A_N^h\big(I_N v, w_N\big)\big| \le \mathrm{e}_N^1 + \mathrm{e}_N^2 + \mathrm{e}_h^1 + \mathrm{e}_h^2,$$

where, for $i = 1, 2$,

$$\mathrm{e}_N^i = \big|A^i\big(v, w_N\big) - A_N^i\big(I_N v, w_N\big)\big|, \quad \mathrm{e}_h^i = \big|A_N^i\big(v, w_N\big) - A_N^{h,i}\big(I_N v, w_N\big)\big|.$$

Note that (7.93) implies $I_N(c\nabla v) = I_N c * I_N(\nabla v)$. Hence, we bound

$$\begin{aligned}\mathrm{e}_N^1 &\le \big|\big(c\nabla v, \nabla w_N\big)_{\mathrm{L}^2} - \big(I_N(c\nabla v), \nabla w_N\big)_N\big| \\ &\quad + \big|\big(I_N c * I_N(\nabla v), \nabla w_N\big)_{\mathrm{L}^2} - \big(I_N c * \nabla(I_N v), \nabla w_N\big)_N\big| =: \mathrm{e}_N^{1,1} + \mathrm{e}_N^{1,2}.\end{aligned}$$

Using (7.92), we verify that

$$\mathrm{e}_N^{1,1} \leq C\frac{1}{|B_\Omega^{-1}N|^s}\|c\nabla v\|_{\mathrm{H}^s} \leq C\frac{\|c\|_{\mathcal{C}^s}}{|B_\Omega^{-1}N|^s}\sum_{\sigma=1}^{s+1}|v|_{\mathrm{H}^\sigma}|w_N|_{\mathrm{H}^1},$$
$$\mathrm{e}_N^{1,2} \leq C\|c\|_{\mathcal{C}^0}\big(\|I_N(\nabla v)-\nabla v\|_{\mathrm{L}^2}+|v-I_Nv|_{\mathrm{H}^1}\big)|w_N|_{\mathrm{H}^1} \leq C\frac{\|c\|_{\mathcal{C}^0}}{|B_\Omega^{-1}N|^s}|v|_{\mathrm{H}^{s+1}}|w_N|_{\mathrm{H}^1}.$$

We verify in a similar manner that

$$\mathrm{e}_N^2 \leq C\frac{\varepsilon\|\bar{a}^{24}\|_{\mathcal{C}^s}}{|B_\Omega^{-1}N|^s}\sum_{\sigma=2}^{s+2}|v|_{\mathrm{H}^\sigma\varepsilon}|w_N|_{\mathrm{H}^2}.$$

Furthermore, denoting $\mathrm{e}_c = \mathrm{e}_{a^0}+\varepsilon\mathrm{e}_{\bar{a}^{12}}+\varepsilon^2\mathrm{e}_{\bar{a}^{22}}$ and using (7.91), we have

$$\mathrm{e}_h^1 = \big|A_N^1\big(I_Nv,w_N\big)-A_N^{h,1}\big(I_Nv,w_N\big)\big| = H^{\mathbb{1}}\Big|\sum_{x_n\in G_N}(c-c^h)(x_n)\nabla I_Nv(x_n)\overline{\nabla w_N(x_n)}\Big|$$
$$\leq \mathrm{e}_c\|\nabla I_Nv\|_N\|\nabla w_N\|_N \leq C\mathrm{e}_c|v|_{\mathrm{H}^1}|w_N|_{\mathrm{H}^1}.$$

Similarly,

$$\mathrm{e}_h^2 = \big|A_N^2\big(I_Nv,w_N\big)-A_N^{h,2}\big(I_Nv,w_N\big)\big| \leq C\varepsilon\mathrm{e}_{\bar{a}^{24}}|v|_{\mathrm{H}^2\varepsilon}|w_N|_{\mathrm{H}^2}.$$

Combining the estimates for e_N^i, e_h^i and using the discrete Cauchy–Schwarz inequality proves the result. □

Part 3 – A priori error estimate

In the third and final part of the proof of Theorem 7.2.3, we prove the error estimate for $\bar{u}-u_N$, where $\bar{u}$ is the solution of (7.78) and u_N is the approximation of the spectral homogenization method defined in (7.79). To do so, we first split the error as

$$\bar{u}-u_N = (\bar{u}-\pi_N\bar{u})-(u_N-\pi_N\bar{u}) = \eta-\zeta_N,$$

where $\pi_N\bar{u}$ is the elliptic projection defined below. Then, we estimate η and ζ_N separately in the norm $\|\nabla\cdot\|_{\mathrm{L}^\infty(\mathrm{L}^2)}$ (Lemmas 7.2.11 and 7.2.12). In particular, the definition of the elliptic projection avoid the use of the Poincaré inequality in the estimate of $\|\nabla\eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)}$ and we obtain an error estimate valid in arbitrarily large domains.

Let us first define the elliptic projection. For almost every $t\in[0,T^\varepsilon]$, let $\pi_N\bar{u}:[0,T^\varepsilon]\to\mathring{V}_N$ be the solution of

$$\big(\pi_N\bar{u}(t),v_N\big)_Q + A_N^h\big(\pi_N\bar{u}(t),v_N\big) = \big(f(t),v_N\big)_{\mathrm{L}^2} - \big(\mathring{I}_N\partial_t^2\bar{u}(t),v_N\big)_Q + \big(\mathring{I}_N\bar{u}(t),v_N\big)_Q, \quad (7.123)$$

for all $v_N\in\mathring{V}_N(\Omega)$. Let us verify that (7.123) is well-posed. For notational convenience, let us define the following norm on $\mathrm{H}^2(\Omega)$:

$$\|v\|_{\mathrm{H}^2,\varepsilon} = \big(\|v\|_{\mathrm{H}^1}^2+\varepsilon^2|v|_{\mathrm{H}^2}^2\big)^{1/2}. \quad (7.124)$$

Thanks to Lemma 7.2.2, we verify that the bilinear form $(\cdot,\cdot)_Q+A_N^h(\cdot,\cdot)$ is coercive and bounded for the norm $\|\cdot\|_{\mathrm{H}^2,\varepsilon}$. Using (7.78) in (7.123), we verify that for all $v_N\in\mathring{V}_N(\Omega)$

$$\begin{aligned}&\big(\pi_N\bar{u}(t),v_N\big)_Q + A_N^h\big(\pi_N\bar{u}(t),v_N\big)\\ &\quad = A\big(\bar{u}(t),v_N\big) + \big(\partial_t^2\bar{u}(t),v_N\big)_{\mathcal{S}} - \big(\mathring{I}_N\partial_t^2\bar{u}(t),v_N\big)_Q + \big(\mathring{I}_N\bar{u}(t),v_N\big)_Q.\end{aligned} \quad (7.125)$$

Using Lax–Milgram theorem, we obtain the existence and uniqueness of $\pi_N\bar{u}(t) \in \mathring{V}_N(\Omega)$. Furthermore, using the test function $v_N = \pi_N\bar{u}(t)$ (7.125) and making use of the properties of $(\cdot,\cdot)_Q$, A_N^h, $(\cdot,\cdot)_{\mathcal{S}}$, and A, we obtain for a.e. $t \in [0, T^\varepsilon]$

$$\|\pi_N\bar{u}(t)\|_{\mathrm{H}^2,\varepsilon} \le C\big(\|\bar{u}(t)\|_{\mathrm{H}^2,\varepsilon} + \|\partial_t^2\bar{u}(t)\|_{\mathrm{H}^1}\big), \tag{7.126}$$

where C depends on λ, α, and $\|a\|_{\mathcal{C}^2(\mathrm{L}^\infty)}$.

The two following lemmas provide error estimates for η and ζ_N.

Lemma 7.2.11. *Assume that for $k \ge 0$, we have $\partial_t^k\bar{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^{s+2}(\Omega))$ and $\partial_t^{k+2}\bar{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathrm{H}^{s+1}(\Omega))$ for some $s \ge (d+1)/2$. Then $\partial_t^k\pi_H\bar{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathring{V}_N(\Omega))$ and, provided $a \in \mathcal{C}^{s+2}(\bar{\Omega};\mathrm{L}^\infty(Y))$, the following estimate holds for $\eta = \bar{u} - \pi_N\bar{u}$,*

$$\begin{aligned}\|\partial_t^k\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|\mathring{I}_N\partial_t^k\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \varepsilon|\partial_t^k\eta|_{\mathrm{L}^\infty(\mathrm{H}^2)} + \varepsilon|\mathring{I}_N\partial_t^k\eta|_{\mathrm{L}^\infty(\mathrm{H}^2)}\\ \le C\bigg(\frac{1}{|B_\Omega^{-1}N|^s} + \mathrm{e}_{a^i} + \mathrm{e}_{b^i}\bigg)\bigg(\sum_{\sigma=1}^{s+2}|\partial_t^k\bar{u}|_{\mathrm{L}^\infty(\mathrm{H}^\sigma)} + \|\partial_t^{k+2}\tilde{u}\|_{\mathrm{L}^\infty(\mathrm{H}^{s+1})}\bigg),\end{aligned} \tag{7.127}$$

where e_{a^i} and e_{b^i} are defined in Lemma 7.2.9 and 7.2.10 and C depends only on λ, Λ, α, $\|a\|_{\mathcal{C}^{s+2}(\mathrm{L}^\infty)}$, Y, d, s, and $r(N)$.

Proof. Applying ∂_t^k to (7.125) and using that A_N^h is coercive and A bounded, Lax–Milgram theorem ensures the existence and uniqueness of $\pi_N\partial_t^k\bar{u}(t) \in \mathring{V}_N(\Omega)$. With a similar argument as in (7.126), we prove that

$$\|\partial_t^k\pi_N\bar{u}(t)\|_{\mathrm{H}^2,\varepsilon} \le C\big(\|\partial_t^k\bar{u}(t)\|_{\mathrm{H}^2,\varepsilon} + \|\partial_t^{k+2}\bar{u}(t)\|_{\mathrm{H}^1}\big).$$

Taking the L^∞ norm with respect to t, we obtain the regularity $\partial_t^k\pi_N\bar{u} \in \mathrm{L}^\infty(0,T^\varepsilon;\mathring{V}_N(\Omega))$. Next, we prove the estimate (7.127) for $k = 0$. The general proof follows the same lines, starting with the time differentiation of (7.125). Using (7.125), we have, almost everywhere in $[0,T^\varepsilon]$ and for any $v_N \in \mathring{V}_N(\Omega)$,

$$\big(\mathring{I}_N\eta, v_N\big)_Q + A_N^h\big(\mathring{I}_N\eta, v_N\big) = A_N^h\big(\mathring{I}_N\bar{u}, v_N\big) - A\big(\bar{u}(t), v_N\big) + \big(\mathring{I}_N\partial_t^2\bar{u}(t), v_N\big)_Q - \big(\partial_t^2\bar{u}(t), v_N\big)_{\mathcal{S}}.$$

Applying Lemmas 7.2.9 and 7.2.10, we obtain

$$\big|\big(\mathring{I}_N\eta(t), v_N\big)_Q + A_N^h\big(\mathring{I}_N\eta(t), v_N\big)\big| \le C\mathrm{e}_{A,\mathcal{S}}\bigg(\sum_{\sigma=1}^{s+2}|\partial_t^k\bar{u}(t)|_{\mathrm{H}^\sigma} + \|\partial_t^{k+2}\tilde{u}(t)\|_{\mathrm{H}^{s+1}}\bigg)\|v_N\|_{\mathrm{H}^2,\varepsilon},$$

where $\mathrm{e}_{A,\mathcal{S}} = |B_\Omega^{-1}N|^{-s} + \mathrm{e}_{a^i} + \mathrm{e}_{b^i}$. We let $v_N = I_N\eta(t)$ and use the coercivity of the bilinear form to get

$$\|\mathring{I}_N\eta(t)\|_{\mathrm{H}^2,\varepsilon}^2 \le C\mathrm{e}_{A,\mathcal{S}}\bigg(\sum_{\sigma=1}^{s+2}|\partial_t^k\bar{u}(t)|_{\mathrm{L}^\infty(\mathrm{H}^\sigma)} + \|\partial_t^{k+2}\tilde{u}(t)\|_{\mathrm{H}^{s+1}}\bigg).$$

Taking the L^∞ norm with respect to t proves the estimate for $\mathring{I}_N\eta$ (recall the definition of $\|\cdot\|_{\mathrm{H}^2,\varepsilon}$ in (7.124)) The estimate for η is obtained with the equality $\eta = \bar{u} - \mathring{I}_N\bar{u} + \mathring{I}_N\eta$, the triangle inequality, and (7.92). The proof of the lemma is complete. □

Lemma 7.2.12. *The following estimate holds for $\zeta_N = u_N - \pi_N\bar{u}$,*

$$\begin{aligned}\|\partial_t\zeta_N\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + |\zeta_N|_{\mathrm{L}^\infty(\mathrm{H}^1)} \le C\Big(&\mathrm{e}^{\mathrm{data}} + |\eta|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \varepsilon|\eta|_{\mathrm{L}^\infty(\mathrm{H}^2)} + \|\partial_t\eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}\\ &+ \|\mathring{I}_N\eta\|_{\mathrm{L}^1(\mathrm{H}^1)} + \|\mathring{I}_N\partial_t^2\eta\|_{\mathrm{L}^1(\mathrm{H}^1)}\Big),\end{aligned} \tag{7.128}$$

where $\mathrm{e}^{\mathrm{data}} = |g^0 - g_N^0|_{\mathrm{H}^2,\varepsilon} + \|g^1 - g_N^1\|_Q$ and C depends only on λ, $\|a\|_{\mathcal{C}^2(\mathrm{L}^\infty)}$.

Proof. Using (7.79) and (7.123) we verify that for any $v_N \in \mathring{V}_N(\Omega)$ and a.e. $t \in [0, T^\varepsilon]$

$$\big(\partial_t^2 \zeta_N(t), v_N\big)_Q + A_N^h\big(\zeta_N(t), v_N\big) = \big(\mathring{I}_N \partial_t^2 \eta(t), v_N\big)_Q - \big(\mathring{I}_N \eta(t), v_N\big)_Q.$$

We let $v_N = \partial_t \zeta_N(t)$ and use the symmetry of A_N^h and $(\cdot,\cdot)_Q$ to obtain for a.e. $t \in [0, T^\varepsilon]$

$$\tfrac{1}{2}\Big(\|\partial_t \zeta_N(t)\|_Q^2 + A_N^h\big(\zeta_N(t), \zeta_N(t)\big)\Big) = \big(\mathring{I}_N \partial_t^2 \eta(t), \partial_t \zeta_N(t)\big)_Q - \big(\mathring{I}_N \eta(t), \partial_t \zeta_N(t)\big)_Q.$$

Denoting

$$E_N \zeta_N(t) = \|\partial_t \zeta_N(t)\|_Q^2 + A_N^h\big(\zeta_N(t), \zeta_N(t)\big), \quad R(t) = \big(\mathring{I}_N \partial_t^2 \eta(t), \partial_t \zeta_N(t)\big)_Q - \big(\mathring{I}_N \eta(t), \partial_t \zeta_N(t)\big)_Q,$$

we integrate over $[0, \xi]$ and get for any $\xi \in [0, T^\varepsilon]$

$$E_N \zeta_N(\xi) = E_N \zeta_N(0) + 2\int_0^\xi R(t)\,\mathrm{d}t.$$

Using Cauchy–Schwartz, Hölder, and Young inequalities, we bound the integral term as

$$\begin{aligned} 2\int_0^\xi R(t)\,\mathrm{d}t &\le 2\|\mathring{I}_N \partial_t^2 \eta\|_{\mathrm{L}^1(Q)} \|\partial_t \zeta_N\|_{\mathrm{L}^\infty(Q)} + 2\|\mathring{I}_N \eta\|_{\mathrm{L}^1(Q)} \|\partial_t \zeta_N\|_{\mathrm{L}^\infty(Q)} \\ &\le 4\|\mathring{I}_N \partial_t^2 \eta\|_{\mathrm{L}^1(Q)}^2 + 4\|\mathring{I}_N \eta\|_{\mathrm{L}^1(Q)}^2 + \tfrac{1}{2}\|\partial_t \zeta_N\|_{\mathrm{L}^\infty(Q)}^2. \end{aligned}$$

Combining the two last equations and using A_N^h ellipticity (7.98), we obtain successively

$$\begin{aligned} \tfrac{1}{2}\|\partial_t \zeta_N\|_{\mathrm{L}^\infty(Q)}^2 &\le E_N \zeta_N(0) + 4\|\mathring{I}_N \partial_t^2 \eta\|_{\mathrm{L}^1(Q)}^2 + 4\|\mathring{I}_N \eta\|_{\mathrm{L}^1(Q)}^2, \\ \lambda |\zeta_N|_{\mathrm{L}^\infty(\mathrm{H}^1)}^2 &\le 2E_N \zeta_N(0) + 8\|\mathring{I}_N \partial_t^2 \eta\|_{\mathrm{L}^1(Q)}^2 + 8\|\mathring{I}_N \eta\|_{\mathrm{L}^1(Q)}^2. \end{aligned} \tag{7.129}$$

Thanks to (7.98), we have $E_N \zeta_N(0) \le \|\partial_t \zeta_N(0)\|_Q^2 + C\|\zeta(0)\|_{\mathrm{H}^2,\varepsilon}^2$ and denoting $e = \bar{u} - u_N$, we bound the two terms as

$$\begin{aligned} &\|\partial_t \zeta_N(0)\|_Q \le \|\partial_t e(0)\|_Q + \|\partial_t \eta(0)\|_Q \le \|g^1 - g_N^1\|_Q + \|\partial_t \eta\|_{\mathrm{L}^\infty(Q)}, \\ &\|\zeta_N(0)\|_{\mathrm{H}^2,\varepsilon} \le \|e(0)\|_{\mathrm{H}^2,\varepsilon} + \|\eta(0)\|_{\mathrm{H}^2,\varepsilon} \le \|g^0 - g_N^0\|_{\mathrm{H}^2,\varepsilon} + C\big(|\eta|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \varepsilon |\eta|_{\mathrm{L}^\infty(\mathrm{H}^2)}\big). \end{aligned} \tag{7.130}$$

Combining (7.129) and (7.130), we obtain estimate (7.128) and the proof of the lemma is complete. □

Proof of Theorem 7.2.3. Let $e = \bar{u} - u_H$ and recall that $e = \eta - \zeta_N$, where $\eta = \bar{u} - \pi_N \bar{u}$, $\zeta_N = u_N - \pi_N \bar{u}$ and $\pi_N \bar{u}$ is the elliptic projection defined in (7.123). The triangle inequality and Lemma 7.2.12 ensure

$$\begin{aligned} \|\partial_t e\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + |e|_{\mathrm{L}^\infty(\mathrm{H}^1)} &\le \|\partial_t \eta\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + |\eta|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|\partial_t \zeta_N\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + |\zeta_N|_{\mathrm{L}^\infty(\mathrm{H}^1)} \\ &\le C\big(\mathrm{e}_{\mathrm{H}^1}^{\mathrm{data}} + |\eta|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \varepsilon |\eta|_{\mathrm{L}^\infty(\mathrm{H}^2)} + \|\partial_t \eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} \\ &\qquad + \|\mathring{I}_N \eta\|_{\mathrm{L}^1(\mathrm{H}^1)} + \|\mathring{I}_N \partial_t^2 \eta\|_{\mathrm{L}^1(\mathrm{H}^1)}\big). \end{aligned}$$

The Hölder inequality implies

$$\|\mathring{I}_N \eta\|_{\mathrm{L}^1(\mathrm{H}^1)} + \|\mathring{I}_N \partial_t^2 \eta\|_{\mathrm{L}^1(\mathrm{H}^1)} \le \varepsilon^{-2} T\big(\|\mathring{I}_N \eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)} + \|\mathring{I}_N \partial_t^2 \eta\|_{\mathrm{L}^\infty(\mathrm{H}^1)}\big),$$

and thus, applying Lemma 7.2.11, we obtain

$$\begin{aligned} &\|\partial_t e\|_{\mathrm{L}^\infty(\mathrm{L}^2)} + |e|_{\mathrm{L}^\infty(\mathrm{H}^1)} \\ &\quad \le C\mathrm{e}_{\mathrm{H}^1}^{\mathrm{data}} + C\varepsilon^{-2}\big(|B_\Omega^{-1} N|^{-s} + \mathrm{e}_{a^i} + \mathrm{e}_{b^i}\big)\big(\textstyle\sum_{\sigma=1}^{s+2} |\bar{u}|_{\mathrm{L}^\infty(\mathrm{H}^\sigma)} + \sum_{k=1}^4 \|\partial_t^k \bar{u}\|_{\mathrm{L}^\infty(\mathrm{H}^{s+1})}\big). \end{aligned}$$

Using the estimates from Lemma 7.2.6, we have

$$\varepsilon^{-2}(\mathrm{e}_{a^i} + \mathrm{e}_{b^i}) = \varepsilon^{-2}(\mathrm{e}_{a^0} + \varepsilon \mathrm{e}_{\bar{a}^{12}} + \varepsilon^2 \mathrm{e}_{\bar{a}^{22}} + \varepsilon \mathrm{e}_{\bar{a}^{24}} + \varepsilon \mathrm{e}_{b^{10}} + \varepsilon^2 \mathrm{e}_{b^{20}} + \varepsilon^2 \mathrm{e}_{b^{22}}) \leq C(\mathrm{e}_1 + \mathrm{e}_2),$$

where

$$\mathrm{e}_1 = \left(\frac{h^q}{\varepsilon}\right)^2 + \frac{|\Delta x|}{\varepsilon} + \frac{h^q}{\varepsilon}, \qquad \mathrm{e}_2 = \frac{h^q |1/\Delta x|}{\varepsilon} + h^q |1/\Delta x|^2,$$

are the error terms defined in (7.103). We thus obtain the error estimate (7.102) and the proof of Theorem 7.2.3 is complete. □

8 Conclusion and outlook

8.1 Conclusion

In this thesis, we have studied effective models for long time wave propagation in heterogeneous media. In particular, we have designed numerical homogenization methods for the approximation of the multiscale wave equation in periodic and locally periodic media over long time.

First, we considered periodic media. In particular, we defined a family of first order effective equations that describe the macroscopic behavior of the wave at timescales $\mathcal{O}(\varepsilon^{-2})$. The derivation was done using asymptotic expansions. Furthermore, an a priori error estimate that guarantees the validity of the family was proved. In addition, we provided a numerical procedure for the computation of first order effective tensors. In particular, the computational cost was significantly reduced compared to the earlier procedures. This led to an efficient numerical homogenization method for the approximation of wave propagation in periodic media at timescales $\mathcal{O}(\varepsilon^{-2})$.

Second, we constructed a family of effective equations for timescales of arbitrary order $\mathcal{O}(\varepsilon^{-\alpha})$, where $\alpha \in \mathbb{N}$. Furthermore, a numerical procedure for the computation of effective tensors of arbitrary order was also provided. In particular, the resulting homogenization method is also appropriate to approximate the wave equation in periodic media with high frequency initial data. Numerical tests confirm the validity of the theory and indicate possible improvements of the algorithm in several dimensions.

Third, the family of effective equations for timescales $\mathcal{O}(\varepsilon^{-2})$ was generalized from periodic to locally periodic media. In this case as well, an a priori error estimate corroborating the validity of the family was established. This result led to the design of a spectral homogenization method for the numerical approximation of the multidimensional wave equation in locally periodic media over long time. In particular, we provided an a priori error analysis of the method that guarantees the convergence of the approximation to an effective solution. As the dependence of the error estimate on the domain is explicit, it can be used in pseudoinfinite domains. Furthermore, we also performed the a priori error analysis of the FE-HMM-L for the one-dimensional approximation of the wave equation in locally periodic media over long time. In particular, we provided an a priori error estimate that ensures the convergence of the approximation to an effective solution of the family, over long time and in arbitrarily large domains.

8.2 Outlook

Some of the results and numerical methods of this thesis call for further investigations, both from the practical and the theoretical point of view. Let us comment on some possible future

directions of research.

8.2.1. Effective equations for arbitrary large timescales in periodic media

Let us discuss potential developments of the results obtained in Chapter 5, on the effective equations for arbitrary timescales.

Construct better effective equations in the family. As discussed earlier, the algorithm for the computation of effective tensors of arbitrary order could probably be improved. Indeed, we observed in a two-dimensional example that the obtained higher order approximation exhibits, locally, an undesired flattening of the dispersion. We think that this negative effect is connected to the particular effective equation that is constructed by the algorithm. And it is likely that other equations of the family would be more accurate. In particular, recall that the construction of the effective tensors of each order follows an algebraic argument. Specifically, to obtain a non-negative tensor, we add a positive contribution that relies on the minimal eigenvalues associated to the tensors. Note that this process relies on several choices, like the form of the added positive tensor or the construction of the matrix associated to the fourth order tensors. Hence, the influence of these choices on the approximation should be investigated more carefully. However, we believe that the right direction of research is to find an alternative to the algebraic procedure. Namely, an argument based on physical properties would probably lead to more accurate effective equations of the family. For example, we could attempt to minimize the energy associated to the error between the oscillating wave and the adaptation of the effective solution. Other optimization procedures could be designed in the attempt to obtain better effective equations in the family.

Link the order of the effective equation to the frequency of the initial data. To homogenize the wave equation with high frequency initial data, we have seen that higher order effective equations must be used. In particular, in order to capture the dispersion effects, the higher the frequencies are, the higher the order of the effective equation should be. In practice, it would be useful to have a criterion based on the frequency of the initial data in order to determine what order of equation should be used. The results obtained in the thesis, and in particular the formula for the effective tensors of arbitrary order, could be used in the attempt of deriving such criterion.

8.2.2. Spectral homogenization method for long time wave propagation in locally periodic media

Let us comment on possible improvements of the spectral homogenization method, defined in Chapter 7.

Reduced order modelling for the spectral homogenization method. The first step of the spectral homogenization method is time consuming. Indeed, even though it can be parallelized, the approximation of the effective tensors requires to solve numerous cell problems. Furthermore, we verify that to gain in accuracy, the number of cell problems must be increased and solved on finer mesh thus augmenting the cost. However, a similar issue has been addressed for the FE-HMM using a reduced order modeling technique. The reduced basis FE-HMM (RB-FE-HMM, see [5, 6]), was precisely developed to diminish the cost of approximation of the effective data (it was discussed in Section 3.4). The reduced basis technique is applicable to reduce the cost of the spectral homogenization method. However, investigation is needed in order to incorporate the additional tensors and correctors in the a posteriori error estimator involved in the greedy procedure of the offline stage.

Practical criterion to simplify the spectral homogenization method. Recall that there is a possibility to simplify the effective model targeted by the spectral homogenization method. In particular, for several examples we verified numerically that some of the operators in the effective equations are unnecessary. In such cases, a simplified homogenization method could thus be used. As the cost of this method is significantly lower, it would be profitable to have at our disposal a criterion to decide when to use the simplified method and when not. Note that the application of this criterion should be less expensive than computing all the effective tensors.

8.2.3. General prospects

Let us end this outlook by mentioning some general possibilities of research related to this thesis.

More general media. It would be interesting to apply the techniques that were developed for locally periodic media to more general media. Indeed, the theory has recently been started for almost periodic, quasiperiodic and random media (see [23]). However, these theoretical results need to be translated to numerical homogenization methods capable of handling these media. Furthermore, in the spirit of the generalization to locally periodic media, other types of media could also be considered.

Other physical problems. The theory and techniques that were developed could probably be applied to other physical problems. In particular, the extension to elastic waves should be relatively direct. In addition, we could also attempt to adapt the theory to the challenging case of electromagnetic waves. Further, the techniques could be used in other physical situations where high order effective models are needed.

Boundary conditions. Recall that in thesis we exclusively considered infinite media. It would be challenging to study what happens when we add boundary conditions. As the theory relies on the assumption that the domain is a hypercube and the union of reference cells, it certainly can not be applied easily to this case. In particular, a better understanding of the homogenization in the boundary layers is essential.

A Appendix

In this appendix, we discuss various aspects of the analysis and numerical analysis of the PDEs met in this thesis. First, in Sections A.1 and A.2, we introduce the fundamentals of functional analysis for the study of elliptic PDEs with periodic boundary conditions. In particular, we discuss the well-posedness and regularity of such problems, which are essential questions in the derivation of effective equations in Chapters 4 and 6. Then, in Section A.3, a short introduction on the finite element method (FEM) is given, which is used in many parts of the thesis. In addition, we discuss the main results of the analysis of FEM with numerical quadrature, which is at the center of the finite element numerical homogenization methods studied in this thesis. Next, in Section A.4, we present the basic theory for the interpolation by trigonometric polynomials. In particular, these results are at the foundation of the spectral method used in Sections 2.3 and 7.2. Finally, in Section A.5, we present the leap frog method, which is used in most of the numerical experiments.

A.1 Definition of the functional spaces

Let $\mathcal{O} \subset \mathbb{R}^d$ be an open set of $\mathbb{R}^d$. We denote $\mathcal{D}(\mathcal{O})$ the space of functions $\varphi : \mathcal{O} \to \mathbb{R}$ of class $\mathcal{C}^\infty$ which are compactly supported in $\mathcal{O}$. The dual space of $\mathcal{D}(\mathcal{O})$, denoted $\mathcal{D}'(\mathcal{O})$, is the space of distributions. The derivatives of $v \in \mathcal{D}'(\mathcal{O})$ are defined as $\langle \partial_{x_i} v, \varphi \rangle = -\langle v, \partial_{x_i} \varphi \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the dual evaluation in $\mathcal{D}'(\mathcal{O})$. Hence, for $\alpha \in \mathbb{N}^d$, we have $\langle \partial^\alpha v, \varphi \rangle = (-1)^\alpha \langle v, \partial^\alpha \varphi \rangle$.

For $p \in [1, \infty)$, the space of p-integrable functions $L_p(\mathcal{O})$ consists of measurable functions $v : \mathcal{O} \to \mathbb{R}$ such that $\int_{\mathcal{O}} |v(x)|^p \, \mathrm{d}x < \infty$. For $p = \infty$, $L_\infty(\mathcal{O})$ is the space of measurable functions $v : \mathcal{O} \to \mathbb{R}$ such that $\inf\{a \in \mathbb{R} : |v(x)| \leq a \text{ for a.e. } x \in \mathcal{O}\} < \infty$. For $p \in [1, \infty]$, two functions $v, w \in L_p(\mathcal{O})$ are equivalent if the set where they differ has (Lebesque) measure zero. We define then the space $\mathrm{L}^p(\mathcal{O}) = L_p(\mathcal{O}) / \sim$. Equipped with the norm

$$\|v\|_{\mathrm{L}^p(\mathcal{O})} = \begin{cases} \left(\int_{\mathcal{O}} |v(x)|^p \, \mathrm{d}x \right)^{1/p} & 1 \leq p < \infty, \\ \operatorname{ess\,sup}_{x \in \mathcal{O}} |v(x)| = \inf\{a \in \mathbb{R} : |v(x)| \leq a \text{ for a.e. } x \in \mathcal{O}\} & p = \infty. \end{cases}$$

$\mathrm{L}^p(\mathcal{O})$ is a Banach space. The space $\mathrm{L}^2(\mathcal{O})$ is a Hilbert space for the following inner product and corresponding norm:

$$(v, w)_{\mathrm{L}^2(\mathcal{O})} = \int_{\mathcal{O}} v(x) w(x) \, \mathrm{d}x, \qquad \|v\|_{\mathrm{L}^2(\mathcal{O})} = \sqrt{(v, v)_{\mathrm{L}^2(\mathcal{O})}}, \qquad v, w \in \mathrm{L}^2(\mathcal{O}).$$

For $k \in \mathbb{N}_{>0}$, $p \in [1, \infty]$, the Sobolev space $\mathrm{W}^{k,p}(\mathcal{O})$ is the set of functions $v \in \mathrm{L}^p(\mathcal{O})$ such that

$\partial^\alpha v \in \mathrm{L}^p(\mathcal{O})$ for all multi-index $\alpha \in \mathbb{N}^d$ such that $|\alpha| \le k$. Equipped with the norm

$$\|v\|_{\mathrm{W}^{k,p}(\mathcal{O})} = \Big(\sum_{|\alpha|\le k} \|\partial^\alpha v\|^p_{\mathrm{L}^p(\mathcal{O})} \Big)^{1/p},$$

$\mathrm{W}^{k,p}(\mathcal{O})$ is a Banach space. In the particular case $p = 2$, the space $\mathrm{W}^{k,2}(\mathcal{O})$ is denoted $\mathrm{H}^k(\mathcal{O})$. Equipped with the inner product and corresponding norm

$$(v,w)_{\mathrm{H}^k(\mathcal{O})} = \sum_{|\alpha|\le k} \big(\partial^\alpha v, \partial^\alpha w\big)_{\mathrm{L}^2(\Omega)} \qquad \|v\|_{\mathrm{H}^k(\mathcal{O})} = \sqrt{(v,v)_{\mathrm{H}^k(\mathcal{O})}}, \qquad v, w \in \mathrm{L}^2(\mathcal{O}),$$

$\mathrm{H}^k(\mathcal{O})$ is a Hilbert space.

The mean of a function $v \in \mathrm{L}^1(\mathcal{O})$ is

$$\langle v \rangle_{\mathcal{O}} = |\mathcal{O}|^{-1} \int_{\mathcal{O}} v(x)\,\mathrm{d}x.$$

We define the quotient space $\mathcal{L}^2(\mathcal{O}) = \mathrm{L}^2(\mathcal{O})/\mathbb{R}$ and denote by a bracket $[\![v]\!]$ the equivalence class in $\mathcal{L}^2(\mathcal{O})$ of $v \in \mathrm{L}^2(\mathcal{O})$. Equipped with the inner product

$$\Big([\![v]\!], [\![w]\!] \Big)_{\mathcal{L}^2(\mathcal{O})} = \Big(v - \langle v \rangle_{\mathcal{O}}, w - \langle w \rangle_{\mathcal{O}} \Big)_{\mathrm{L}^2(\mathcal{O})} = \big(v,w\big)_{\mathrm{L}^2(\mathcal{O})} - |\mathcal{O}| \langle v \rangle_{\mathcal{O}} \langle w \rangle_{\mathcal{O}} \quad \forall v, w \in \mathrm{L}^2(\mathcal{O}),$$

$\mathcal{L}^2(\mathcal{O})$ is a Hilbert space. Let $\mathcal{C}^\infty_{\mathrm{per}}(\mathcal{O})$ be the space of $\mathcal{O}$-periodic functions of $\mathcal{C}^\infty(\mathcal{O})$ and define the space $\mathrm{H}^1_{\mathrm{per}}(\mathcal{O})$ as the closure of $\mathcal{C}^\infty_{\mathrm{per}}(\mathcal{O})$ for the H^1 norm. We define the quotient space $\mathcal{W}_{\mathrm{per}}(\mathcal{O}) = \mathrm{H}^1_{\mathrm{per}}(\mathcal{O})/\mathbb{R}$ and denote by a bold face letter $\boldsymbol{v}$ the elements of $\mathcal{W}_{\mathrm{per}}(\mathcal{O})$. Equipped with the inner product

$$\big(\boldsymbol{v}, \boldsymbol{w}\big)_{\mathcal{W}_{\mathrm{per}}(\mathcal{O})} = \big([\![v]\!], [\![w]\!] \big)_{\mathcal{L}^2(\mathcal{O})} + \big(\partial_k v, \partial_k w\big)_{\mathrm{L}^2(\mathcal{O})}, \quad \forall v \in \boldsymbol{v}, w \in \boldsymbol{w},$$

and the induced norm $\|\boldsymbol{v}\|_{\mathcal{W}_{\mathrm{per}}(\mathcal{O})} = \sqrt{(\boldsymbol{v},\boldsymbol{v})_{\mathcal{W}_{\mathrm{per}}(\mathcal{O})}}$, $\mathcal{W}_{\mathrm{per}}(\mathcal{O})$ is a Hilbert space. Note that the k-th partial derivative of $\boldsymbol{v} \in \mathcal{W}_{\mathrm{per}}(\mathcal{O})$ is simply $\partial_k \boldsymbol{v} = \partial_k v \in \mathrm{L}^2(\mathcal{O})$ for all $v \in \boldsymbol{v}$. Thanks to the Poincaré–Wirtinger inequality, $\boldsymbol{v} \mapsto \|\nabla \boldsymbol{v}\|_{\mathrm{L}^2(\mathcal{O})}$ is also a norm on $\mathcal{W}_{\mathrm{per}}(\mathcal{O})$, equivalent to $\|\cdot\|_{\mathcal{W}_{\mathrm{per}}(\mathcal{O})}$. The dual space $\mathcal{W}^*_{\mathrm{per}}(\mathcal{O})$ is characterized as follows: for $F \in \mathcal{W}^*_{\mathrm{per}}(\mathcal{O})$, there exists $[\![f^0]\!] \in \mathcal{L}^2(\mathcal{O})$, $f^1_1, \dots, f^1_d \in \mathrm{L}^2(\mathcal{O})$ such that

$$\big\langle F, \boldsymbol{v} \big\rangle_{\mathcal{W}^*_{\mathrm{per}}(\mathcal{O}), \mathcal{W}_{\mathrm{per}}(\mathcal{O})} = \big([\![f^0]\!], \boldsymbol{v} \big)_{\mathcal{L}^2(\mathcal{O})} + \big(f^1_k, \partial_k \boldsymbol{v}\big)_{\mathrm{L}^2(\mathcal{O})}. \tag{A.1}$$

Furthermore,

$$\|F\|_{\mathcal{W}^*_{\mathrm{per}}(\mathcal{O})} = \inf \Big\{ \Big(\|[\![f^0]\!]\|^2_{\mathcal{L}^2(\mathcal{O})} + \|f^1\|^2_{\mathrm{L}^2(\mathcal{O})} \Big)^{1/2} : [\![f^0]\!] \in \mathcal{L}^2(\mathcal{O}), f^1 \in \mathrm{L}^2(\mathcal{O}) \text{ satisfies (A.1)} \Big\},$$

From characterization (A.1), we verify that a functional of $[\mathrm{H}^1_{\mathrm{per}}(\mathcal{O})]^*$ given by

$$w \mapsto (f^0, w)_{\mathrm{L}^2(\mathcal{O})} + (f^1_k, \partial_k w)_{\mathrm{L}^2(\mathcal{O})},$$

for some $f^0, f^1_1, \dots, f^1_d \in \mathrm{L}^2(\mathcal{O})$, belongs to $\mathcal{W}^*_{\mathrm{per}}(\mathcal{O})$ if and only if

$$\big(f^0, 1\big)_{\mathrm{L}^2(\mathcal{O})} = 0, \tag{A.2}$$

or equivalently f^0 has zero mean. Define $\mathrm{L}^2_0(\mathcal{O})$ (resp. $\mathrm{W}_{\mathrm{per}}(\mathcal{O})$) as the set constituted with the zero mean representative of $\mathcal{L}^2(\mathcal{O})$ (resp. of $\mathcal{W}_{\mathrm{per}}(\mathcal{O})$). Equipped with the standard L^2

inner product (resp. H^1), $\mathrm{L}^2_0(\mathcal{O})$ is a Hilbert space (resp. $\mathrm{W}_{\mathrm{per}}(\mathcal{O})$). Note that the following embeddings are dense $\mathrm{W}_{\mathrm{per}}(\mathcal{O}) \subset \mathrm{L}^2_0(\mathcal{O}) \subset \mathrm{W}^*_{\mathrm{per}}(\mathcal{O})$.

We define the following norm on $\mathcal{W}_{\mathrm{per}}(\mathcal{O})$

$$\|\boldsymbol{w}\|_{\mathcal{W}} = \inf_{\substack{\boldsymbol{w}=\boldsymbol{w}_1+\boldsymbol{w}_2 \\ \boldsymbol{w}_i=[w_i]\in\mathcal{W}_{\mathrm{per}}(\mathcal{O})}} \left\{ \| [w_1] \|_{\mathcal{L}^2(\mathcal{O})} + \|\nabla w_2\|_{\mathrm{L}^2(\mathcal{O})} \right\} \quad \forall \boldsymbol{w} \in \mathcal{W}_{\mathrm{per}}(\mathcal{O}), \tag{A.3}$$

and the corresponding norm on $\mathrm{W}_{\mathrm{per}}(\mathcal{O})$

$$\|w\|_W = \inf_{\substack{w=w_1+w_2 \\ w_1,w_2\in\mathrm{W}_{\mathrm{per}}(\mathcal{O})}} \left\{ \|w_1\|_{\mathrm{L}^2(\mathcal{O})} + \|\nabla w_2\|_{\mathrm{L}^2(\mathcal{O})} \right\} \quad \forall w \in \mathrm{W}_{\mathrm{per}}(\mathcal{O}). \tag{A.4}$$

Note that for all $w_1, w_2 \in \mathrm{H}^1_{\mathrm{per}}(\mathcal{O})$ it holds

$$\| [w_1] \|_{\mathcal{L}^2(\mathcal{O})} + \|\nabla w_2\|_{\mathrm{L}^2(\mathcal{O})} = \|w_1 - \langle w_1 \rangle_{\mathcal{O}}\|_{\mathrm{L}^2(\mathcal{O})} + \|\nabla(w_2 - \langle w_2 \rangle_{\mathcal{O}})\|_{\mathrm{L}^2(\mathcal{O})},$$

and thus, for $w \in \mathrm{W}_{\mathrm{per}}(\mathcal{O})$, we have the equality $\|w\|_W = \| [w] \|_{\mathcal{W}}$. Note that $\|\cdot\|_W$ is equivalent to the L^2 norm

$$\|w\|_W \le \|w\|_{\mathrm{L}^2(\mathcal{O})} \le \max\{1, C_{\mathcal{O}}\}\|w\|_W \quad \forall w \in \mathrm{W}_{\mathrm{per}}(\mathcal{O}), \tag{A.5}$$

where the second inequality follows from the Poincaré–Wirtinger inequality and $C_{\mathcal{O}}$ is the Poincaré constant.

For a Banach space X and $1 \le p \le \infty$, $\mathrm{L}^p(0,T;X)$ is the space of measurable functions $v : (0,T) \to X$ such that the map $t \mapsto \|v(t)\|_X$ belongs to $\mathrm{L}^p(0,T)$. Equipped with the norm

$$\|v\|_{\mathrm{L}^p(0,T;X)} = \begin{cases} \left(\int_0^T \|v(t)\|_X^p \, \mathrm{d}t \right)^{1/p} & 1 \le p < \infty, \\ \operatorname{ess\,sup}_{t\in(0,T)} \|v(t)\|_X & p = \infty, \end{cases}$$

$\mathrm{L}^p(0,T;X)$ is a Banach space. In the particular case $X = \mathrm{H}^k(\mathcal{O})$, we use the following notation for the seminorm in $\mathrm{L}^p(0,T;\mathrm{H}^k(\mathcal{O}))$:

$$|v|_{\mathrm{L}^p(0,T;\mathrm{H}^k(\mathcal{O}))} = \begin{cases} \left(\int_0^T |v(t)|^p_{\mathrm{H}^k(\mathcal{O})} \, \mathrm{d}t \right)^{1/p} & 1 \le p < \infty, \\ \operatorname{ess\,sup}_{t\in(0,T)} |v(t)|_{\mathrm{H}^k(\mathcal{O})} & p = \infty. \end{cases}$$

For an open set $\mathcal{O} \subset \mathbb{R}^d$, we define the space $\mathcal{C}^0(\bar{\mathcal{O}}; X)$ as the set of measurable functions $v : \bar{\mathcal{O}} \to X, x \mapsto v(x)$ that are continuous, i.e., for all $x \in \bar{\mathcal{O}}$ and for all $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that for $\|h\|_{\mathbb{R}^d} \le \delta$ we have $\|v(x+h) - v(x)\|_X \le \varepsilon$. Equipped with the norm $\|v\|_{\mathcal{C}^0(\bar{\mathcal{O}};X)} = \sup_{x\in\mathcal{O}} \|v(x)\|_X$, $\mathcal{C}^0(\bar{\mathcal{O}}; V)$ is a Banach space. For $m \ge 0$, the space $\mathcal{C}^m(\bar{\mathcal{O}}; V)$ is the set of functions $v \in \mathcal{C}^0(\bar{\mathcal{O}}; X)$ such that $\partial^\alpha v \in \mathcal{C}^0(\bar{\mathcal{O}}; X)$ for all the multi-index $\alpha \in \mathbb{N}^d$ such that $0 \le |\alpha| \le m$. Equipped with the norm $\|v\|_{\mathcal{C}^m(\bar{\mathcal{O}};X)} = \sum_{|\alpha|\le m} \|\partial^\alpha v\|_{\mathcal{C}^0(\bar{\mathcal{O}};X)}$, $\mathcal{C}^m(\bar{\mathcal{O}}; X)$ is a Banach space.

A.2 Important results in the theory of partial differential equations

In this section, we present some general results in functional analysis for the study of PDES. In particular, we apply these results in the periodic settings, used in most of the thesis.

We start with the following classical and essential result (we refer to [48] for the proof).

Theorem A.2.1. (Lax–Milgram) *Let V be a Hilbert space, $A : V \times V \to \mathbb{R}$ be a bilinear form, and $f \in V^*$ a linear functional. Assume that there exist $\alpha, \beta > 0$ such that*

$$A(v,v) \geq \alpha \|v\|_V^2, \qquad A(v,w) \leq \beta \|v\|_V \|w\|_V \qquad \forall v, w \in V. \tag{A.6}$$

Then there exists a unique $u \in V$ such that

$$A(u,v) = \langle f, v \rangle_{V^*,V} \qquad \forall v \in V.$$

Furthermore, u satisfies the estimate $\|u\|_V \leq \frac{1}{\alpha}\|f\|_{V^*}$.

Let us precise these results in the context of this thesis, where we encounter abundant elliptic PDEs with periodic boundary conditions. For an open hypercube $Y \subset \mathbb{R}^d$, let a be a Y-periodic $d \times d$ symmetric tensor that is elliptic and bounded, i.e., there exists $\lambda, \Lambda > 0$ such that

$$a(y)\xi \cdot \xi \geq \lambda |\xi|^2, \quad a(y)\xi \cdot \xi \leq \Lambda |\xi|^2 \quad \text{for a.e. } y \in Y.$$

Given a function f, we look for a Y-periodic u such that

$$-\nabla_y \cdot \big(a(y)\nabla_y u(y)\big) = f(y) \quad \text{in } Y. \tag{A.7}$$

The existence and uniqueness of a solution u of this equation is classical in functional analysis. It is proved using the Lax–Milgram theorem. We let V be the Hilbert space $\mathcal{W}_{\mathrm{per}}(Y)$ (or similarly $\mathrm{W}_{\mathrm{per}}(Y)$) equipped with the H^1 norm. Using the assumptions on $a(y)$, the Poincaré–Wirtinger and the Cauchy–Schwartz inequalities, we can prove that the bilinear form $A(\boldsymbol{v}, \boldsymbol{w}) = \big(a\nabla_y \boldsymbol{v}, \nabla_y \boldsymbol{w}\big)_{\mathrm{L}^2(Y)}$ satisfies (A.6). To apply Theorem A.2.1, we need f to belong to the dual $\mathcal{W}^*_{\mathrm{per}}(Y)$. According to (A.2), $f \in [\mathrm{H}^1_{\mathrm{per}}(Y)]^*$ given by

$$\big\langle f, w \big\rangle = \big(f^0, w\big)_{\mathrm{L}^2(Y)} + \big(f^1_k, \partial_k w\big)_{\mathrm{L}^2(Y)}.$$

for some $f^0, f^1_1, \ldots, f^1_d \in \mathrm{L}^2(Y)$ belongs to $\mathcal{W}^*_{\mathrm{per}}(Y)$ if and only if

$$\big(f^0, 1\big)_{\mathrm{L}^2(Y)} = 0, \tag{A.8}$$

Hence, provided f satisfies (A.8), Theorem (A.6) ensures that there exists a unique $\boldsymbol{u} \in \mathcal{W}^*_{\mathrm{per}}(Y)$ such that

$$A(\boldsymbol{u}, \boldsymbol{v}) = \big\langle f, \boldsymbol{v} \big\rangle \qquad \forall \boldsymbol{v} \in \mathcal{W}^*_{\mathrm{per}}(Y).$$

We verify that if $u \in \boldsymbol{u}$ is in $\mathrm{H}^2(Y)$, it satisfies (A.7) in the L^2 sense. The solvability condition (A.8) is essential for the derivation of effective equations in Chapters 4, 5, and 6.

Another classical result deals with the regularity of the weak solution $\boldsymbol{u}$ of (A.7). Let us state it for the zero mean solution $u \in \boldsymbol{u}$, $u \in \mathrm{W}_{\mathrm{per}}(Y)$. The following result, discussed in [26], gives sufficient conditions on a and f for u to belong to $\mathrm{H}^k(Y)$.

Theorem A.2.2. *Let $u \in \mathrm{W}_{\mathrm{per}}(Y)$ be the zero mean weak solution of* (A.7)*. If for some integer $m \geq 0$, $a(y)$ and f satisfy the regularity $a \in \mathrm{W}^{m+1,\infty}(Y)$ and $f \in \mathrm{H}^m(Y)$ (where $\mathrm{H}^0(Y) = \mathrm{L}^2(Y)$), then u satisfies the regularity $u \in \mathrm{H}^{m+2}(Y)$. Furthermore, the following estimate holds*

$$\|u\|_{\mathrm{H}^{m+2}} \leq C \|f\|_{\mathrm{H}^m},$$

where the constant C depends only on Y, $\|a\|_{\mathrm{W}^{m+1,\infty}(Y)}$, d, and m.

Theorem A.2.2 provides sufficient conditions for the solution u to belong to the Sobolev space of any order. The following theorem provides efficient condition for the solution to be continuous (we refer to [48] for the proof).

Theorem A.2.3. (Sobolev embeddings) *Let $\mathcal{O}$ be an open subset of $\mathbb{R}^d$ with a $\mathcal{C}^1$ boundary.*

i) If $k < \frac{d}{p}$, then for q satisfying $\frac{1}{q} = \frac{1}{p} - \frac{k}{d}$, we have $\mathrm{W}^{k,p}(\mathcal{O}) \hookrightarrow L^q(\mathcal{O})$.

ii) If $k > \frac{d}{p}$, then $\mathrm{W}^{k,p}(\mathcal{O}) \hookrightarrow \mathcal{C}^0(\bar{\mathcal{O}})$.

Let us show how Theorem A.2.3 can be used in the context of periodic functions. Assuming for simplicity that $d \leq 3$, we prove that the following embedding holds:

$$\mathrm{H}^2_{\mathrm{per}}(Y) \hookrightarrow \mathcal{C}^0_{\mathrm{per}}(\bar{Y}). \tag{A.9}$$

To see it, let $v \in \mathrm{H}^2_{\mathrm{per}}(Y)$ and denote $v^\sharp$ its extension to $\mathbb{R}^d$ by periodicity. Let $\{Y_i\}_{i=1}^{3^d-1}$ be neighbor copies of Y surrounding Y. and let U be a smooth domain containing $\bar{Y}$ and contained in $Y \cup \big(\cup_{i=1}^{3^d-1} \bar{Y}_i\big)$. Thanks to Theorem A.2.3 *ii*) and the periodicity of v, we have

$$\|v\|_{\mathcal{C}^0(\bar{Y})} \leq \|v^\sharp\|_{\mathcal{C}^0(\bar{U})} \leq C\|v^\sharp\|_{\mathrm{H}^2(U)} \leq 3^{d/2} C \|v\|_{\mathrm{H}^2(Y)}.$$

and (A.9) is verified.

We are now able to provide sufficient conditions for the solution of (A.7) to be, for example, continuous. Indeed, assuming $d \leq 3$, then if $a \in \mathrm{W}^1(Y)$ and $f \in \mathcal{L}^2(Y)$, Theorem A.2.2 ensures that $u \in \mathrm{H}^2(Y)$ and (A.9) implies $u \in \mathcal{C}^0(\bar{Y})$.

A.3 A short introduction on the finite element method for elliptic equations

In this section, we briefly introduce the finite element method for the approximation of elliptic problems. The purpose is to give an overview of the general theory and to describe the main tools used for the derivation of a priori error estimates. We first follow [33] to introduce the method for the approximation of elliptic equations. We prove the standard a priori error estimates in the H^1 and L^2 norm. Second, we give some details on the tools used to estimate the error caused by numerical integration [34, 33]. In particular, we derive conditions on the quadrature formula such that the optimal convergence rates of the method are maintained. As most of the results are classical, we refer to [33] for the missing proofs and for detailed explanations. For the implementation of the method, we refer to [47, 32].

A.3.1 The finite element method for elliptic equations

We follow here [33] and introduce the finite element method for the approximation of elliptic equations. The purpose is the general understanding of the method and we refer to [33, 47] for a more thorough introduction.

Let $\Omega \subset \mathbb{R}^d$ be a polygonal domain. Let $a(x)$ be a tensor function. Given some function $f : \Omega \to \mathbb{R}$, whose regularity will be specified, we want to approximate the solution $u : \Omega \to \mathbb{R}$ of the boundary value problem

$$\begin{aligned} &-\nabla \cdot \big(a(x)\nabla u(x)\big) = f(x) \quad x \in \Omega, \\ &\text{conditions on } u|_{\partial\Omega}. \end{aligned} \tag{A.10}$$

The boundary conditions can be of diverse nature. For simplicity, we focus on the two (simple) types of boundary conditions used in this thesis. First, homogeneous Dirichlet boundary conditions, i.e. $u|_{\partial\Omega} = 0$, in which case we define the functional space $V = \mathrm{H}^1_0(\Omega)$. Second, periodic boundary conditions, i.e., $x \mapsto u(x)$ is Ω-periodic (in this case, Ω is assumed to be a hypercube) and we define $V = \mathrm{W}_{\mathrm{per}}(\Omega)$. In both case, a Poincaré type inequality holds: $\|v\|_{\mathrm{L}^2(\Omega)} \leq C_\Omega \|\nabla v\|_{\mathrm{L}^2(\Omega)}$

$\forall v \in V$. To ensure the well-posedness of (A.10), we assume that a is uniformly elliptic and bounded, i.e., there exists $\alpha, \beta > 0$ such that

$$a(x)\xi \cdot \xi \geq \alpha|\xi|^2, \quad |a(x)\xi| \leq \beta|\xi|, \qquad \text{for a.e. } x \in \Omega.$$

These properties imply that the bilinear form

$$A : V \times V \to \mathbb{R}, \ (v, w) \mapsto A(v, w) = (a\nabla v, \nabla w)_{\mathrm{L}^2(\Omega)}, \tag{A.11}$$

is elliptic and bounded, i.e.,

$$A(v, v) \geq \tilde{\alpha}_\Omega \|v\|^2_{\mathrm{H}^1(\Omega)}, \quad A(v, w) \leq \beta \|v\|_{\mathrm{H}^1(\Omega)} \|w\|_{\mathrm{H}^1(\Omega)}, \tag{A.12}$$

where $\tilde{\alpha}_\Omega = \frac{\alpha}{C_\Omega^2 + 1}$ and C_Ω is the Poincaré constant. Then, for any $f \in V^*$, Lax–Milgram theorem ensures the well-posedness of the variational problem: $u \in V$ is the unique function such that

$$A(u, v) = \langle f, v \rangle \qquad \forall v \in V. \tag{A.13}$$

Riesz representation theorem provides the following characterization of V^*: $f \in V^*$ has the form

$$\big\langle f, v \big\rangle = \big(f^0, v\big)_{\mathrm{L}^2(\Omega)} + \big(f^1, \nabla v\big)_{\mathrm{L}^2(\Omega)}, \tag{A.14}$$

for some $f^0 \in \mathrm{L}^2(\Omega), f^1 \in [\mathrm{L}^2(\Omega)]^d$ if $V = \mathrm{H}^1_0(\Omega)$, or $f^0 \in \mathrm{L}^2_0(\Omega), f^1 \in [\mathrm{L}^2(\Omega)]^d$ if $V = \mathrm{W}_{\mathrm{per}}(\Omega)$. Let now V_H be a finite dimensional subspace of V. The space V_H can be defined in many ways, depending on the specific context. As our purpose is the general understanding of the analysis of the method, we will consider basic finite element spaces (defined later). Let us first define the (yet abstract) finite element approximation of the solution of (A.10): $u_H \in V_H$ is the solution of

$$A(u_H, v_H) = \langle f, v_H \rangle \qquad \forall v_H \in V_H. \tag{A.15}$$

Note that the well-posedness of (A.15) follows the Lax–Milgram theorem (Theorem A.2.1), using the properties of $a(x)$ and f. From (A.13) and (A.15), Galerkin orthogonality follows naturally:

$$A(u - u_H, v_H) = 0 \qquad \forall v_H \in V_H. \tag{A.16}$$

From (A.16), we obtain Céa's lemma [33, Thm 2.4.1]:

Lemma A.3.1. (Céa's lemma) *Let u and u_H be the solutions of respectively* (A.13) *and* (A.15). *Then, the following error estimate holds*

$$\|u - u_H\|_{\mathrm{H}^1(\Omega)} \leq \beta/\tilde{\alpha}_\Omega \inf_{v_H \in V_H} \|u - v_H\|_{\mathrm{H}^1(\Omega)},$$

where β and $\tilde{\alpha}_\Omega$ are given in (A.12).

The result of Lemma A.3.1 signifies that the FEM has the same order of accuracy as the best approximation of u in V_H, in the H^1 norm. In other words, the accuracy of the FEM is directly linked to the capacity of V_H to capture u and its gradient. This naturally leads to the question of interpolation of a function $v \in V$ onto V_H. Indeed, if $I_H : V \to V_H$ is an interpolation operator, Lemma A.3.1 ensures the estimate $\|u - u_H\|_{\mathrm{H}^1(\Omega)} \leq C\|u - I_H u\|_{\mathrm{H}^1(\Omega)}$. The challenge lies then in finding an interpolation operator I_H with optimal order of accuracy. There are several ways to define such I_H and we follow here [33] to define the nodal interpolation operator. To that end, let us introduce a conformal mesh $\mathcal{T}_H$ of Ω. We assume here that the elements $K \in \mathcal{T}_H$ are d-simplices (note that quadrilaterals could be used). Further, we assume that each K is affine-equivalent to a reference element $\hat{K} \subset \mathbb{R}^d$, i.e., there exists an invertible affine mapping

$$F_K : \mathbb{R}^d \to \mathbb{R}^d, \hat{x} \mapsto F_K(\hat{x}) = B_K \hat{x} + b,$$

such that $F_K(\hat{K}) = K$. For $K \in \mathcal{T}_H$, we define the quantities

$$H_K = \mathrm{diam}(K), \quad \rho_K = \sup\{\mathrm{diam}(S_K) : S_K \text{ is a ball contained in K}\},$$

and $\hat{H}, \hat{\rho}$ are defined similarly for $\hat{K}$. The parameter H is the size of the partition, $H = \max_{K\in\mathcal{T}_H} H_K$. We have the following theorem [33, Thms 3.1.2 & 3.1.3].

Theorem A.3.2. *Let $v \in \mathrm{W}^{m,p}(K)$ for some $m \geq 0$, $p \in [1,\infty]$. Then $\hat{v} = v \circ F_K \in \mathrm{W}^{m,p}(\hat{K})$ and*

$$|\hat{v}|_{\mathrm{W}^{m,p}(\hat{K})} \leq C \frac{H_K^m}{\hat{\rho}^m} \big|\det(B_K)\big|^{-1/p} |v|_{\mathrm{W}^{m,p}(K)},$$

where the constant C depends only on d and m. Similarly, for $\hat{v} \in \mathrm{W}^{m,p}(\hat{K})$, we have $v = \hat{v} \circ F_K^{-1} \in \mathrm{W}^{m,p}(K)$ and

$$|v|_{\mathrm{W}^{m,p}(K)} \leq C \frac{\hat{H}^m}{\rho_K^m} \big|\det(B_K)\big|^{1/p} |\hat{v}|_{\mathrm{W}^{m,p}(\hat{K})},$$

where the constant C depends only on d and m.

In this settings, we can rewrite $\|v - I_H v\|_{\mathrm{H}^1(\Omega)} = \big(\sum_{K\in\mathcal{T}_H} \|v - I_H v\|^2_{\mathrm{H}^1(K)}\big)^{1/2}$, and the interpolation operator can then be constructed locally for each element $K \in \mathcal{T}_H$. For notations convenience, we introduce the broken norm on V_H

$$\|v_H\|_{\bar{\mathrm{H}}^k(\Omega)} = \Big(\sum_{K\in\mathcal{T}_H} \|v_H\|^2_{\mathrm{H}^k(K)} \Big)^{1/2}.$$

The following result provides a tool for the construction of I_H. It establishes an error estimate for any polynomial preserving operator [33, Thm 3.1.4].

Theorem A.3.3. *For integers $k, m \geq 0$ and real numbers $p, q \in [1,\infty]$, let $\mathrm{W}^{k+1,p}(\hat{K})$ and $\mathrm{W}^{m,q}(\hat{K})$ be such that $\mathrm{W}^{k+1,p}(\hat{K}) \hookrightarrow \mathrm{W}^{m,q}(\hat{K})$. Furthermore, let $\hat{\Pi} \in \mathcal{L}(\mathrm{W}^{k+1,p}(\hat{K}); \mathrm{W}^{m,q}(\hat{K}))$ be a linear mapping such that*

$$\hat{\Pi}\hat{p} = \hat{p} \qquad \forall \hat{p} \in \mathcal{P}_k(\hat{K}),$$

and define $\Pi_K \in \mathcal{L}(\mathrm{W}^{k+1,p}(K); \mathrm{W}^{m,q}(K))$ as $v \mapsto \Pi_K v = \big(\hat{\Pi}(v \circ F_K)\big) \circ F_K^{-1}$. Then for any $v \in \mathrm{W}^{k+1,p}(K)$

$$|v - \Pi_K v|_{\mathrm{W}^{m,q}(K)} \leq C |K|^{1/q - 1/p} \frac{H_K^{k+1}}{\rho_K^m} |v|_{\mathrm{W}^{k+1,p}(K)},$$

where the constant C depends on $\hat{\Pi}$ and $\hat{K}$.

Theorem A.3.3 is an important tool for the design of finite element spaces. Recall that we introduce here a simple type of finite elements and refer to [33, 47] for a wider variety. In particular, we assume the elements $K \in \mathcal{T}_H$ to be d-simplices. Let us also assume that they are shape regular, i.e., there exist a constant σ such that

$$\frac{H_K}{\rho_K} \leq \sigma \quad \forall K \in \mathcal{T}_H. \tag{A.17}$$

For an integer $\ell \geq 1$, we define the finite element space

$$V_H = \{v_H \in V : v_H|_K \in \mathcal{P}_\ell(K)\ \forall K \in \mathcal{T}_H\}. \tag{A.18}$$

It can be verified that $V_H \subset \mathcal{C}^0(\bar{\Omega}) \cap \mathrm{H}^1(\Omega)$. Note that this finite element space is suited in our context of general understanding of the theory and analysis. However, it is rarely used in

applications for $k \geq 4$. For higher order elements, more sophisticated finite elements spaces should be used.

We are now able to define the interpolation operator. Denoting $\{a_j\}_{j=1}^{d+1}$ the vertices of the d-simplex K, we verify that a polynomial $p \in \mathcal{P}_\ell(K)$ is uniquely determined by its values on the set

$$L_\ell(K) = \big\{x = \textstyle\sum_{j=1}^{d+1} \lambda_j a_j : \sum_{j=1}^{d+1} \lambda_j = 1,\ \lambda_j \in \{0, 1/\ell, \ldots, (\ell-1)/\ell, 1\},\ 1 \leq j \leq d+1\big\}.$$

We define the interpolation of $\hat{v} \in \mathrm{H}^1(\hat{K})$ as the unique polynomial $\hat{I}\hat{v} \in \mathcal{P}_\ell(\hat{K})$ such that $\hat{v}(\hat{x}) = \hat{I}\hat{v}(\hat{x})$ for all nodes $\hat{x} \in L_\ell(\hat{K})$. The local interpolation operator is then given as $I_K : \mathrm{H}^1(K) \to \mathcal{P}_\ell(K)$, $v \mapsto I_K v = \big(\hat{I}(v \circ F_K)\big) \circ F_K^{-1}$. Combining Theorem A.3.3 and assumption (A.17), we obtain, for any $v \in V \cap \mathrm{H}^{k+1}(\Omega)$ and $0 \leq k \leq \ell$, the estimate

$$|v - I_K v|_{\mathrm{H}^m(K)} \leq C H_K^{k+1-m} |v|_{\mathrm{H}^{k+1}(K)}, \quad 0 \leq m \leq k+1, \tag{A.19}$$

where C depends only on $\hat{I}$, $\hat{K}$, and σ. In particular, for any $v \in V \cap \mathrm{H}^{\ell+1}(\Omega)$, we have

$$\|v - I_K v\|_{\mathrm{L}^2(K)} \leq C H_K^{\ell+1} |v|_{\mathrm{H}^{\ell+1}(K)}, \qquad |v - I_K v|_{\mathrm{H}^1(K)} \leq C H_K^{\ell} |v|_{\mathrm{H}^{\ell+1}(K)}, \tag{A.20}$$

where C depends only on $\hat{I}$, $\hat{K}$, and σ. The interpolation operator of $v \in V$ onto V_H is then defined as $I_H v|_K = I_K v$. Combining now Lemma A.3.1 and (A.20), we obtain the following a priori error estimate for the finite element approximation u_H (u_H is the solution of (A.15) and V_H is defined in (A.18)):

$$\|u - u_H\|_{\mathrm{H}^1(\Omega)} \leq C H^\ell \|u\|_{\mathrm{H}^{\ell+1}(\Omega)}, \tag{A.21}$$

where C is independent of H.

In view of (A.20), it is natural to ask whether u_H approximate u with order $\ell+1$ in the L^2 norm. This question is answered by the following result, known as the Aubin–Nitsche duality argument [33, Thm 3.2.4].

Theorem A.3.4. *Let u and u_H be the solutions of respectively* (A.13) *and* (A.15)*. Then the following estimate holds*

$$\|u - u_H\|_{\mathrm{L}^2(\Omega)} \leq \Lambda \|u - u_H\|_{\mathrm{H}^1(\Omega)} \left(\sup_{g \in \mathrm{L}^2(\Omega)} \left\{ \|g\|_{\mathrm{L}^2(\Omega)}^{-1} \inf_{\varphi_H \in V_H} \|\varphi_g - \varphi_H\|_{\mathrm{H}^1(\Omega)} \right\} \right),$$

where φ_g is the unique solution in V of the problem $A(v, \varphi_g) = (g, v)_{\mathrm{L}^2(\Omega)}$ $\forall v \in V$.

Tanks to Theorem A.3.4 and elliptic regularity (Theorem A.2.2), we can prove an error estimate in the L^2 norm. Note that if $V = \mathrm{H}_0^1(\Omega)$, the elliptic regularity holds provided $\partial\Omega$ is polygonal or sufficiently smooth. We then have

$$\inf_{\varphi_H \in V_H} \|\varphi_g - \varphi_H\|_{\mathrm{H}^1(\Omega)} \leq \|\varphi_g - I_H \varphi_g\|_{\mathrm{H}^1(\Omega)} \leq C H \|\varphi_g\|_{\mathrm{H}^2(\Omega)} \leq C H \|g\|_{\mathrm{L}^2(\Omega)},$$

where C depends on Ω. Hence, combining this estimate to Theorem A.3.4 and (A.21), we obtain the following error estimate in the L^2 norm:

$$\|u - u_H\|_{\mathrm{L}^2(\Omega)} \leq C H^{\ell+1} \|u\|_{\mathrm{H}^{\ell+1}(\Omega)}, \tag{A.22}$$

where C is independent of H.

A.3.2 Effect of the numerical integration in the finite element method

Note that in the FEM (A.15), in the previous section, we assumed that the forms could be computed exactly. In practice, except for special type of a and f, the forms $A(v_H, w_H)$ and $\langle f, v_H\rangle$, defined in (A.11) and (A.14) can not be evaluated exactly. To go further in the analysis, we have to take into account the error made in the approximation of the integrals. In this section, we follow [34, 33] and derive sufficient conditions on the quadrature formula for the optimal order of convergence to be preserved (Theorems A.3.6 and A.3.9). Note that this analysis is also performed in [15, 4], where the effect of numerical quadrature error is studied in the context of numerical homogenization.

Let $\{\hat\omega_j, \hat x_j\}_{j=1}^J$ be a quadrature formula on the reference element $\hat K$. Note that via F_K, it induces the quadrature formula $\{\omega_{K_j}, x_{K_j}\}_{j=1}^J$ on K, where $\omega_{K_j} = |\det B_K|\hat\omega_j$ and $x_{K_j} = F_K(\hat x_j)$. Let us define

$$A_H(v_H, w_H) = \sum_{K\in\mathcal{T}_H}\sum_{j=1}^J \omega_{K_j} a(x_{K_j})\nabla v_H(x_{K_j})\cdot\nabla w_H(x_{K_j}),$$
$$\langle f_H, v_H\rangle = \sum_{K\in\mathcal{T}_H}\sum_{j=1}^J \omega_{K_j} f^0(x_{K_j}) v_H(x_{K_j}) + \sum_{K\in\mathcal{T}_H}\sum_{j=1}^J \omega_{K_j} f_i^1(x_{K_j})\cdot\nabla v_H(x_{K_j}).$$

The finite element method is then to find $u_H \in V_H$ such that

$$A_H(u_H, v_H) = \langle f_H, v_H\rangle \qquad \forall v_H \in V_H. \tag{A.23}$$

The first question concerns the well-posedness of (A.23). In particular, we have to verify the ellipticity of the bilinear form $A_H(\cdot,\cdot)$. This question is addressed by [33, Thm 4.1.2], which ensures that if the quadrature formula $\{\hat\omega_j, \hat x_j\}_{j=1}^J$ has order $2\ell-1$, then the form A_H is elliptic on V_H. Explicitly, if $\{\hat\omega_j, \hat x_j\}_{j=1}^J$ satisfies

$$\int_{\hat K}\hat p(\hat x)\,\mathrm{d}\hat x = \sum_{j=1}^J \hat\omega_j \hat p(\hat x_j) \qquad \forall \hat p \in \mathcal{P}^{2\ell-2}(\hat K), \tag{A.24}$$

then there exits $\gamma > 0$ such that $A_H(v_H, v_H) \geq \gamma\|v_H\|^2_{\mathrm{H}^1(\Omega)}$ for any $v_H \in V_H$. Hence, Lax–Milgram theorem ensures the well-posedness of (A.23).

The next concern is the accuracy of the method. We look for a quadrature formula accurate enough so that the optimal order accuracy obtained in (A.21) and (A.22) are maintained. The first tool for the analysis of the accuracy is provided by the following theorem [33, Thm 4.1.1].

Theorem A.3.5. (First Strang lemma) *Let u and u_H be the solutions of respectively* (A.13) *and* (A.15)*. Then the following error estimate holds*

$$\|u-u_H\|_{\mathrm{H}^1(\Omega)} \leq C \inf_{v_H\in V_H}\left\{\|u-v_H\|_{\mathrm{H}^1(\Omega)} + \sup_{w_H\in V_H}\frac{|A(v_H,w_H)-A_H(v_H,w_H)|}{\|w_H\|_{\mathrm{H}^1(\Omega)}} + \sup_{w_H\in V_H}\frac{|\langle f, w_H\rangle - \langle f_H, w_H\rangle|}{\|w_H\|_{\mathrm{H}^1(\Omega)}}\right\},$$

where C depends only on Λ and γ.

Theorem A.3.5 indicates that to analyze the error can be analyzed independently for the numerical integration errors $|A(v_H, w_H) - A_H(v_H, w_H)|$ and $|\langle f, w_H\rangle - \langle f_H, w_H\rangle|$. These errors are studied

locally for each element $K \in \mathcal{T}_H$. Let us define the local quadrature error on K and $\hat{K}$, for $\varphi \in \mathrm{L}^1(K)$, $\hat{\varphi} \in \mathrm{L}^1(\hat{K})$, as

$$E_K(\varphi) = \int_K \varphi(x)\mathrm{d}x - \sum_{j=1}^{J} \omega_{K_j} \varphi(x_{K_j}), \quad \hat{E}(\hat{\varphi}) = \int_{\hat{K}} \hat{\varphi}(\hat{x})\mathrm{d}\hat{x} - \sum_{j=1}^{J} \hat{\omega}_j \varphi(\hat{x}_j).$$

Note that if for $\varphi \in \mathrm{L}^1(K)$, $E_K(\varphi) = |\det(B_K)|\hat{E}(\hat{\varphi})$, where $\hat{\varphi} = \varphi \circ F_K$. The following results [33, Thms 4.1.4 & 4.1.5] give sufficient conditions for the method (A.23) to converge with optimal order accuracy in the H^1 norm. Note that point iii) is not proved in [33], but the proof follows the same line as point ii) with minor modifications (as done in the proof of Theorem A.3.9 iii) below).

Theorem A.3.6. *Assume that the quadrature formula $\{\hat{\omega}_j, \hat{x}_j\}_{j=1}^J$ satisfies* (A.24).

i) If $a \in [\mathrm{W}^{\ell,\infty}(K)]^{d\times d}$, then, for any polynomials $q, p \in \mathcal{P}_\ell(K)$, the following estimate holds

$$E_K(a\nabla q \nabla p) \le C H_K^\ell \|a_{ij}\|_{\mathrm{W}^{\ell,\infty}(K)} \|\partial_i q\|_{\mathrm{H}^{\ell-1}(K)} \|\partial_j p\|_{\mathrm{L}^2(K)},$$

where C is independent of K.

ii) If for some $q \in [1,\infty]$ such that $\ell - d/q > 0$, we have $f \in \mathrm{W}^{\ell,q}(K)$, then, for any polynomial $p \in \mathcal{P}_\ell(K)$, the following estimate holds

$$E_K(fp) \le C H_K^\ell |K|^{1/2-1/q} \|f\|_{\mathrm{W}^{\ell,q}(K)} \|p\|_{\mathrm{H}^1(K)},$$

where C is independent of K.

iii) For any polynomials $q, p \in \mathcal{P}_\ell(K)$, the following estimate holds

$$E_K(qp) \le C H_K^\ell \|q\|_{\mathrm{H}^\ell(K)} \|p\|_{\mathrm{H}^1(K)},$$

where C is independent of K.

Thanks to Theorems A.3.5 and A.3.6 and using the interpolation operator I_H, we can prove the following optimal a priori error estimate in the H^1 norm.

Theorem A.3.7. *Assume that $d \le 3$ and that the data in* (A.13) *satisfy the regularity $a \in [\mathrm{W}^{\ell,\infty}(\Omega)]^{d\times d}$ and $f^0, f^1 \in \mathrm{H}^{m+\ell}(\Omega)$ for some $m \ge d/4$. Let u be the solution of* (A.13) *and u_H be the solution of* (A.23). *Then the following error estimate holds*

$$\|u - u_H\|_{\mathrm{H}^1(\Omega)} \le C H^\ell \Big(\max_{ij} \|a_{ij}\|_{\mathrm{W}^{\ell,\infty}(\Omega)} + \|f^0\|_{\mathrm{H}^{m+\ell}(\Omega)} + \|f^1\|_{\mathrm{H}^{m+\ell}(\Omega)} \Big) \|u\|_{\mathrm{H}^{\ell+1}(\Omega)}, \quad \text{(A.25)}$$

where C is independent of H.

Remark A.3.8. Note that $\mathrm{H}^{m+\ell}(\Omega) \hookrightarrow \mathrm{W}^{\ell,4}(K)$ and for $d \le 3$ we have $q = 4 > d/\ell$ for any $\ell \ge 1$, so that the regularity assumption in Theorem A.3.6 ii) is satisfied for f^0 and f^1. Notice also that the assumption on the quadrature formula to approximate $(f^1, \nabla v_H)_{\mathrm{L}^2(\Omega)}$ could be weakened, as $\partial_j v_H|_K \in \mathcal{P}_{\ell-1}(K)$.

We still need an estimate to ensure the optimal order of convergence in the L^2 norm. We prove the following theorem.

Theorem A.3.9. *Assume that the quadrature formula $\{\hat{\omega}_j, \hat{x}_j\}_{j=1}^J$ satisfies*

$$\int_{\hat{K}} \hat{p}(\hat{x})\,\mathrm{d}\hat{x} = \sum_{j=1}^{J} \hat{\omega}_j \hat{p}(\hat{x}_j) \qquad \forall \hat{p} \in \mathcal{P}_\sigma(\hat{K}),\ \sigma = \max\{2\ell - 2, 1\}. \quad \text{(A.26)}$$

i) If $a \in [\mathrm{W}^{\ell+1,\infty}(K)]^{d\times d}$, then, for any polynomials $q, p \in \mathcal{P}_\ell(K)$, the following estimate holds

$$E_K(a\nabla q\nabla p) \le C H_K^{\ell+1} \|a_{ij}\|_{\mathrm{W}^{\ell+1,\infty}(K)} \|\partial_i q\|_{\mathrm{H}^{\ell-1}(K)} \|\partial_j p\|_{\mathrm{H}^1(K)}, \tag{A.27}$$

where C is independent of K.

ii) If for some $q \in [1,\infty]$ such that $\ell - d/q > 0$, we have $f \in \mathrm{W}^{\ell+1,q}(K)$, then, for any polynomial $p \in \mathcal{P}_\ell(K)$, the following estimate holds

$$E_K(fp) \le C H_K^{\ell+1} |K|^{1/2-1/q} \|f\|_{\mathrm{W}^{\ell+1,q}(K)} \|p\|_{\mathrm{H}^2(K)}, \tag{A.28}$$

where C is independent of K.

iii) For any polynomials $q, p \in \mathcal{P}_\ell(K)$, the following estimate holds

$$E_K(qp) \le C H_K^{\ell+1} \|q\|_{\mathrm{H}^\ell(K)} \|p\|_{\mathrm{H}^2(K)}, \tag{A.29}$$

where C is independent of K.

These results are not proved in [33] and can only be found in [34] with different notations. We provide here a detailed proof. We first prove ii), and then i), iii). The proof relies on the following lemma [33, Thm 4.1.3].

Lemma A.3.10. (Bramble–Hilbert lemma) *For an integer $k \ge 0$ and a number $q \in [1,\infty]$, let $L \in [\mathrm{W}^{k+1,q}(\hat K)]^*$ be a continuous functional such that*

$$L(\hat p) = 0 \quad \forall \hat p \in \mathcal{P}_k(\hat K).$$

Then there exists a constant C that depends on $\hat K$ such that

$$|L(v)| \le C\|L\|_{[\mathrm{W}^{k+1,q}(\hat K)]^*} |\hat\psi|_{\mathrm{W}^{k+1,q}(\hat K)} \quad \forall \hat\psi \in \mathrm{W}^{k+1,q}(\hat K).$$

Proof of Theorem A.3.9 ii). First, via a change of variable we have

$$E_K(fp) = |\det(B_K)| \hat E(\hat f \hat p), \tag{A.30}$$

where $\hat f = f \circ F_K, \hat p = p \circ F_K$. Let $\hat\Pi$ be the L^2 projection onto $\mathcal{P}_1(\hat K)$, i.e., $\hat\Pi : \mathrm{L}^1(\hat K) \to \mathcal{P}_1(\hat K)$, $\hat v \mapsto \hat\Pi \hat v$ such that

$$\big(\hat\Pi \hat v, \hat p\big)_{\mathrm{L}^2(\hat K)} = \big(\hat v, \hat p\big)_{\mathrm{L}^2(\hat K)} \quad \forall \hat p \in \mathcal{P}_1(\hat K). \tag{A.31}$$

We split the error as

$$\hat E(\hat f \hat p) = \hat E(\hat f \hat\Pi \hat p) + \hat E(\hat f(\hat p - \hat\Pi \hat p)). \tag{A.32}$$

We first estimate the first term of the right hand side. Consider the linear functional $L : \hat\psi \mapsto L(\hat\psi) = \hat E(\hat\psi)$. We verify that L belongs to $[\mathrm{W}^{\ell,q}(\hat K)]^*$: note that the assumption $\ell - d/q > 0$ ensures the continuous embedding $\mathrm{W}^{\ell,q}(\hat K) \hookrightarrow \mathcal{C}^0(\hat K)$, hence for any $\hat\psi \in \mathrm{W}^{\ell,q}(\hat K)$,

$$|L(\hat\psi)| \le C\|\hat\psi\|_{\mathrm{L}^\infty(\hat K)} \le C\|\hat\psi\|_{\mathrm{W}^{\ell,q}(\hat K)}.$$

By assumption, L vanishes on $\mathcal{P}_\ell(\hat K)$ (indeed, as $\sigma = \max\{2\ell-2, 1\}$, it holds $\mathcal{P}_\ell(\hat K) \subset \mathcal{P}_\sigma(\hat K)$), hence applying Bramble–Hilbert lemma, we obtain $|\hat E(\hat\psi)| \le C|\hat\psi|_{\mathrm{W}^{\ell+1,q}(\hat K)}$. Applying that estimate to $\hat\psi = \hat f \hat\Pi \hat p$, we have

$$\begin{aligned} |\hat E(\hat f \hat\Pi \hat p)| &\le C \sum_{j=0}^{\ell+1} |\hat f|_{\mathrm{W}^{\ell+1-j,q}(\hat K)} |\hat\Pi \hat p|_{\mathrm{W}^{j,\infty}(\hat K)} \\ &\le C\big(|\hat f|_{\mathrm{W}^{\ell+1,q}(\hat K)} \|\hat\Pi \hat p\|_{\mathrm{L}^\infty(\hat K)} + |\hat f|_{\mathrm{W}^{\ell,q}(\hat K)} |\hat\Pi \hat p|_{\mathrm{W}^{1,\infty}(\hat K)}\big). \end{aligned}$$

Using that two norms are equivalent on the finite dimensional space $\mathcal{P}_1(\hat{K})$, we have $\|\hat{\Pi}\hat{p}\|_{\mathrm{L}^\infty(\hat{K})} \leq C\|\hat{\Pi}\hat{p}\|_{\mathrm{L}^2(\hat{K})}$ and thus $\|\hat{\Pi}\hat{p}\|_{\mathrm{L}^\infty(\hat{K})} \leq C\|\hat{p}\|_{\mathrm{L}^2(\hat{K})}$. Furthermore, using the norm equivalence and Theorem A.3.3 ($\hat{\Pi}$ leaves $\mathcal{P}_0(\hat{K})$ invariant), we have

$$|\hat{\Pi}\hat{p}|_{\mathrm{W}^{1,\infty}(\hat{K})} \leq C|\hat{\Pi}\hat{p}|_{\mathrm{H}^1(\hat{K})} \leq C\big(|\hat{p}|_{\mathrm{H}^1(\hat{K})} + |\hat{p} - \hat{\Pi}\hat{p}|_{\mathrm{H}^1(\hat{K})}\big) \leq C|\hat{p}|_{\mathrm{H}^1(\hat{K})},$$

and we obtain

$$|\hat{E}(\hat{f}\hat{\Pi}\hat{p})| \leq C\big(|\hat{f}|_{\mathrm{W}^{\ell+1,q}(\hat{K})}\|\hat{p}\|_{\mathrm{L}^2(\hat{K})} + |\hat{f}|_{\mathrm{W}^{\ell,q}(\hat{K})}|\hat{p}|_{\mathrm{H}^1(\hat{K})}\big). \tag{A.33}$$

Let us now estimate the second term of the right hand side of (A.32). Note that if $\ell = 1$, we have $(\hat{p} - \hat{\Pi}\hat{p}) = 0$, so from now on we assume that $\ell \geq 2$. In that case, let us show that there exists a number $\rho \in [0, \infty]$ such that the following embeddings hold

$$\mathrm{W}^{\ell,q}(\hat{K}) \hookrightarrow \mathrm{W}^{\ell-1,\rho}(\hat{K}) \hookrightarrow \mathcal{C}^0(\hat{K}). \tag{A.34}$$

First, we assume that $1 \leq q < d$. We let ρ be such that $1/\rho = 1/q - 1/d$ so that $\mathrm{W}^{1,q}(\hat{K}) \hookrightarrow \mathrm{L}^\rho(\hat{K})$ holds and thus the first embedding in (A.34) holds. The second embedding holds as we verify that $\ell - 1 - (d/\rho) = \ell - d/q > 0$. Second, assume that $q \geq d$. Then, for any $\rho \neq \infty$, the embedding $\mathrm{W}^{1,q}(\hat{K}) \hookrightarrow \mathrm{L}^\rho(\hat{K})$ holds and thus the first embedding in (A.34) holds. For the second embedding in (A.34) to hold, we choose ρ large enough so that $\ell - 1 - (d/\rho) > 0$. Let us now define the linear functional $L : \mathrm{W}^{\ell-1,\rho}(\hat{K}) \to \mathbb{R}$ as $\hat{\psi} \mapsto L(\hat{\psi}) = \hat{E}(\hat{\psi}(\hat{p} - \hat{\Pi}\hat{p}))$. Using (A.34) and the equivalence of norms in $\mathcal{P}_\ell(\hat{K})$, we verify that $L \in [\mathrm{W}^{\ell-1,\rho}(\hat{K})]^*$:

$$|L(\hat{\psi})| \leq C\|\hat{\psi}\|_{\mathrm{L}^\infty(\hat{K})}\|\hat{p} - \hat{\Pi}\hat{p}\|_{\mathrm{L}^\infty(\hat{K})} \leq C\|\hat{\psi}\|_{\mathrm{W}^{\ell-1,\rho}(\hat{K})}\|\hat{p} - \hat{\Pi}\hat{p}\|_{\mathrm{L}^2(\hat{K})}.$$

Furthermore, assumption (A.26) ensures that L vanishes over $\mathcal{P}_{\ell-2}(\hat{K})$. Bramble–Hilbert lemma thus implies that

$$\begin{aligned}|\hat{E}(\hat{f}(\hat{p} - \hat{\Pi}\hat{p}))| = |L(\hat{f})| &\leq C\|L\|_{[\mathrm{W}^{\ell-1,\rho}(\hat{K})]^*}|\hat{f}|_{\mathrm{W}^{\ell-1,\rho}(\hat{K})} \\ &\leq C\|\hat{p} - \hat{\Pi}\hat{p}\|_{\mathrm{L}^2(\hat{K})}\big(|\hat{f}|_{\mathrm{W}^{\ell-1,q}(\hat{K})} + |\hat{f}|_{\mathrm{W}^{\ell,q}(\hat{K})}\big),\end{aligned}$$

where for the second inequality we used the embedding $\mathrm{W}^{1,q}(\hat{K}) \hookrightarrow \mathrm{L}^\rho(\hat{K})$. As $\hat{\Pi}$ leaves $\mathcal{P}_1(\hat{K})$ invariant, using Theorem A.3.3, we obtain $\|\hat{p} - \hat{\Pi}\hat{p}\|_{\mathrm{L}^2(\hat{K})} \leq C|\hat{p}|_{\mathrm{H}^2(\hat{K})}$ and thus

$$|\hat{E}(\hat{f}(\hat{p} - \hat{\Pi}\hat{p}))| \leq C\big(|\hat{f}|_{\mathrm{W}^{\ell-1,q}(\hat{K})} + |\hat{f}|_{\mathrm{W}^{\ell,q}(\hat{K})}\big)|\hat{p}|_{\mathrm{H}^2(\hat{K})}. \tag{A.35}$$

Combining now (A.30), (A.32), (A.33) and (A.35) with the following bounds

$$\begin{aligned}&|\hat{f}|_{\mathrm{W}^{\ell+1-j,q}(\hat{K})} \leq CH_K^{\ell+1-j}|\det(B_K)|^{-1/q}|f|_{\mathrm{W}^{\ell+1-j,q}(K)} \quad j = 0, 1, 2, \\ &|\hat{p}|_{\mathrm{H}^j(\hat{K})} \leq CH_K^j|\det(B_K)|^{-1/2}|p|_{\mathrm{H}^j(K)} \quad j = 0, 1, 2,\end{aligned}$$

obtained thanks to Theorem A.3.2 (note that $|\det(B_K)| = |K|/|\hat{K}|$), the proof of (A.28) is complete. □

Proof of Theorem A.3.9 i). Let us prove an estimate for

$$E_K(bqp) = |\det(B_K)|\hat{E}(\hat{b}\hat{q}\hat{p}), \tag{A.36}$$

where $b \in \mathrm{W}^{\ell+1,\infty}(K)$, $q, p \in \mathcal{P}_{\ell-1}(\hat{K})$ and $\hat{b} = b \circ F_K$, $\hat{b} = b \circ F_K$, $\hat{b} = b \circ F_K$. Let $\hat{\Pi}$ be the L^2 projection onto $\mathcal{P}_1(\hat{K})$ (as defined in (A.31)) and split the error as

$$\hat{E}(\hat{b}\hat{q}\hat{p}) = \hat{E}(\hat{b}\hat{q}\hat{\Pi}\hat{p}) + \hat{E}(\hat{b}\hat{q}(\hat{p} - \hat{\Pi}\hat{p})). \tag{A.37}$$

Similarly as to obtain (A.33) (with $\hat{f} = \hat{b}\hat{q} \in \mathrm{W}^{\ell+1,\infty}(\hat{K})$), we obtain

$$|\hat{E}(\hat{b}\hat{q}\hat{\Pi}\hat{p})| \leq C\big(|\hat{b}\hat{q}|_{\mathrm{W}^{\ell+1,\infty}(\hat{K})}\|\hat{p}\|_{\mathrm{L}^2(\hat{K})} + |\hat{b}\hat{q}|_{\mathrm{W}^{\ell,\infty}(\hat{K})}|\hat{p}|_{\mathrm{H}^1(\hat{K})}\big).$$

As $\hat{q} \in \mathcal{P}_{\ell-1}(\hat{K})$, we have $|\hat{q}|_{\mathrm{W}^{\ell,\infty}(\hat{K})} = |\hat{q}|_{\mathrm{W}^{\ell+1,\infty}(\hat{K})} = 0$. Hence, using the equivalence of norms in $\mathcal{P}_{\ell-1}(\hat{K})$, we have

$$\begin{aligned}
|\hat{b}\hat{q}|_{\mathrm{W}^{\ell+1,\infty}(\hat{K})} &\leq C\sum_{j=0}^{\ell+1}|\hat{b}|_{\mathrm{W}^{\ell+1-j,\infty}(\hat{K})}|\hat{q}|_{\mathrm{W}^{j,\infty}(\hat{K})} \leq C\sum_{j=0}^{\ell-1}|\hat{b}|_{\mathrm{W}^{\ell+1-j,\infty}(\hat{K})}|\hat{q}|_{\mathrm{H}^{j}(\hat{K})},\\
|\hat{b}\hat{q}|_{\mathrm{W}^{\ell,\infty}(\hat{K})} &\leq C\sum_{j=0}^{\ell}|\hat{b}|_{\mathrm{W}^{\ell-j,\infty}(\hat{K})}|\hat{q}|_{\mathrm{W}^{j,\infty}(\hat{K})} \leq C\sum_{j=0}^{\ell-1}|\hat{b}|_{\mathrm{W}^{\ell-j,\infty}(\hat{K})}|\hat{q}|_{\mathrm{H}^{j}(\hat{K})}.
\end{aligned} \tag{A.38}$$

We obtain the following estimate for the first term of the right hand side of (A.37):

$$|\hat{E}(\hat{b}\hat{q}\hat{\Pi}\hat{p})| \leq C\Big(\sum_{j=0}^{\ell-1}|\hat{b}|_{\mathrm{W}^{\ell+1-j,\infty}(\hat{K})}|\hat{q}|_{\mathrm{H}^{j}(\hat{K})}\|\hat{p}\|_{\mathrm{L}^2(\hat{K})} + \sum_{j=0}^{\ell-1}|\hat{b}|_{\mathrm{W}^{\ell-j,\infty}(\hat{K})}|\hat{q}|_{\mathrm{H}^{j}(\hat{K})}|\hat{p}|_{\mathrm{H}^1(\hat{K})}\Big). \tag{A.39}$$

Let us now estimate the second term of the right hand side of (A.37). We define the linear functional $L : \mathrm{W}^{\ell,\infty}(\hat{K}) \to \mathbb{R}$ as $\hat{\psi} \mapsto L(\hat{\psi}) = \hat{E}(\hat{\psi}(\hat{p} - \hat{\Pi}\hat{p}))$. Using the embedding $\mathrm{W}^{\ell,\infty}(\hat{K}) \hookrightarrow \mathcal{C}^0(\hat{K})$, and the equivalence of norms in $\mathcal{P}_{\ell-1}(\hat{K})$, we verify that $L \in [\mathrm{W}^{\ell,\infty}(\hat{K})]^*$:

$$|L(\hat{\psi})| \leq C\|\hat{\psi}\|_{\mathrm{L}^\infty(\hat{K})}\|\hat{p} - \hat{\Pi}\hat{p}\|_{\mathrm{L}^\infty(\hat{K})} \leq C\|\hat{\psi}\|_{\mathrm{W}^{\ell,\infty}(\hat{K})}\|\hat{p} - \hat{\Pi}\hat{p}\|_{\mathrm{L}^2(\hat{K})}.$$

As by assumption L vanishes over $\mathcal{P}_{\ell-1}(\hat{K})$, Bramble–Hilbert lemma gives, for $\hat{\psi} = \hat{b}\hat{q}$,

$$\hat{E}(\hat{b}\hat{q}(\hat{p} - \hat{\Pi}\hat{p})) = |L(\hat{b}\hat{q})| \leq C\|L\|_{[\mathrm{W}^{\ell,\infty}(\hat{K})]^*}|\hat{b}\hat{q}|_{\mathrm{W}^{\ell,\infty}(\hat{K})} \leq C\|\hat{p} - \hat{\Pi}\hat{p}\|_{\mathrm{L}^2(\hat{K})}|\hat{b}\hat{q}|_{\mathrm{W}^{\ell,\infty}(\hat{K})}.$$

Using (A.38) and the bound $\|\hat{p} - \hat{\Pi}\hat{p}\|_{\mathrm{L}^2(\hat{K})} \leq C|\hat{p}|_{\mathrm{H}^1(\hat{K})}$ (Theorem A.3.3, $\hat{\Pi}$ leaves $\mathcal{P}_0(\hat{K})$ invariant), we get

$$\hat{E}(\hat{b}\hat{q}(\hat{q} - \hat{\Pi}\hat{q})) \leq C\sum_{j=0}^{\ell-1}|\hat{b}|_{\mathrm{W}^{\ell-j,\infty}(\hat{K})}|\hat{q}|_{\mathrm{H}^{j}(\hat{K})}|\hat{p}|_{\mathrm{H}^1(\hat{K})}. \tag{A.40}$$

Combining (A.36), (A.37), (A.39) and (A.40) with the estimates

$$\begin{aligned}
&|\hat{b}|_{\mathrm{W}^{k,\infty}(\hat{K})} \leq CH_K^k|b|_{\mathrm{W}^{k,\infty}(K)} \quad k = 0,\dots,\ell+1,\\
&|\hat{q}|_{\mathrm{H}^{j}(\hat{K})} \leq CH_K^j|\det(B_K)|^{-1/2}|q|_{\mathrm{H}^{j}(K)} \quad j = 0,\dots,\ell-1,\\
&\|\hat{p}\|_{\mathrm{L}^2(\hat{K})} \leq C|\det(B_K)|^{-1/2}\|p\|_{\mathrm{L}^2(K)}, \qquad |\hat{p}|_{\mathrm{H}^1(\hat{K})} \leq CH_K|\det(B_K)|^{-1/2}|p|_{\mathrm{H}^1(K)},
\end{aligned} \tag{A.41}$$

obtained thanks to Theorem A.3.2, we get the bound

$$E_K(bqp) \leq CH_K^{\ell+1}\|b\|_{\mathrm{W}^{\ell+1,\infty}(K)}\|q\|_{\mathrm{H}^{\ell-1}(K)}|p|_{\mathrm{H}^1(K)}.$$

To obtain (A.27), we apply this estimate to every term in $a\nabla q \cdot \nabla p = \sum_{ij} a_{ij}\partial_j q \partial_i p$, where $\partial_j q, \partial_i p \in \mathcal{P}_{\ell-1}(K)$ and that completes the proof. □

Proof of Theorem A.3.9 iii). The proof follows the same lines as the proof of ii) with $f = q \in \mathrm{W}^{\ell+1,\infty}(K)$ and the following changes. Using the equivalence of norms in $\mathcal{P}_\ell(\hat{K})$ in (A.33) and (A.35), we obtain respectively

$$\begin{aligned}
|\hat{E}(\hat{q}\hat{\Pi}\hat{p})| &\leq C\big(|\hat{q}|_{\mathrm{H}^{\ell+1}(\hat{K})}\|\hat{p}\|_{\mathrm{L}^2(\hat{K})} + |\hat{q}|_{\mathrm{H}^{\ell}(\hat{K})}|\hat{p}|_{\mathrm{H}^1(\hat{K})}\big),\\
|\hat{E}(\hat{q}(\hat{p} - \hat{\Pi}\hat{p}))| &\leq C\big(|\hat{q}|_{\mathrm{H}^{\ell-1}(\hat{K})} + |\hat{q}|_{\mathrm{H}^{\ell}(\hat{K})}\big)|\hat{p}|_{\mathrm{H}^2(\hat{K})},
\end{aligned}$$

where we note that $|\hat{q}|_{\mathrm{H}^{\ell+1}(\hat{K})} = 0$. The proof is then completed by combining these estimates with (A.30), (A.32) and bounds as in (A.41). □

We are now able to prove an optimal a priori error estimate in the L^2 norm.

Theorem A.3.11. *Assume that $d \leq 3$ and that the data in* (A.13) *satisfy the regularity $a \in [\mathrm{W}^{\ell,\infty}(\Omega)]^{d\times d}$ and $f^0, f^1 \in \mathrm{H}^{m+\ell+1}(\Omega)$ for some $m \geq d/4$. Let u be the solution of* (A.13) *and u_H be the solution of* (A.23). *Then the following error estimate holds*

$$\|u - u_H\|_{\mathrm{L}^2(\Omega)} \leq CH^{\ell+1}\Big(\max_{ij} \|a_{ij}\|_{\mathrm{W}^{\ell+1,\infty}(\Omega)} + \|f^0\|_{\mathrm{H}^{m+\ell+1}(\Omega)} + \|f^1\|_{\mathrm{H}^{m+\ell+1}(\Omega)}\Big)\|u\|_{\mathrm{H}^{\ell+1}(\Omega)}, \tag{A.42}$$

where C is independent of H.

Proof. First, as discussed for Theorem A.3.7, a Sobolev embedding ensures $f^0, f^1 \in \mathrm{H}^{m+\ell+1}(\Omega) \hookrightarrow \mathrm{W}^{\ell+1,4}(\Omega)$ and as $q = 4 > d/\ell$ for any $\ell \geq 1$, f^0 and f^1 satisfy the regularity assumption of Theorem A.3.9 i). To prove the estimate, we apply the Aubin–Nitsche argument (see Theorem A.3.4). We write the error in the L^2 norm as

$$\|u - u_H\|_{\mathrm{L}^2(\Omega)} = \sup_{g\in \mathrm{L}^2(\Omega)} \Big\{ \|g\|^{-1}_{\mathrm{L}^2(\Omega)} |(u - u_H, g)_{\mathrm{L}^2(\Omega)}| \Big\}. \tag{A.43}$$

Let us fix $g \in \mathrm{L}^2(\Omega)$ and define $\varphi_g \in V$ as the unique solution to the elliptic problem $A(v, \varphi_g) = (g, v)_{\mathrm{L}^2(\Omega)}\ \forall v \in V$. Elliptic regularity ensures that $\|\varphi_g\|_{\mathrm{H}^2(\Omega)} \leq C\|g\|_{\mathrm{L}^2(\Omega)}$. Next, we use the definition of φ_g and equations (A.13) and (A.23) to write for any $v_H \in V_H$

$$\begin{aligned} |(u - u_H, g)_{\mathrm{L}^2(\Omega)}| &= |A(u - u_H, \varphi_g)| \\ &\leq |A(u - u_H, \varphi_g - v_H)| + |\langle f, v_H\rangle - \langle f_H, v_H\rangle| + |A_H(u_H, v_H) - A(u_H, v_H)|. \end{aligned} \tag{A.44}$$

Let us estimate the three terms of the right hand side for $v_H = I_H\varphi_g$. Using (A.19) we have $\|\varphi_g - I_H\varphi_g\|_{\mathrm{H}^1(\Omega)} \leq C\|\varphi_g\|_{\bar{\mathrm{H}}^2(\Omega)}$, and thus the first term satisfies

$$A(u-u_H, \varphi - I_H\varphi_g) \leq \Lambda\|u-u_H\|_{\mathrm{H}^1(\Omega)}\|\varphi_g - I_H\varphi_g\|_{\mathrm{H}^1(\Omega)} \leq CH\|u-u_H\|_{\mathrm{H}^1(\Omega)}\|\varphi_g\|_{\mathrm{H}^2(\Omega)}. \tag{A.45}$$

To bound the second term, we use Theorem A.3.9 ii) to obtain

$$\begin{aligned} |\langle f, I_H\varphi_g\rangle - \langle f_H, I_H\varphi_g\rangle| &\leq CH^{\ell+1}\big(\|f^0\|_{\mathrm{H}^{m+\ell+1}(\Omega)} + \|f^1\|_{\mathrm{H}^{m+\ell+1}(\Omega)}\big)\|I_H\varphi_g\|_{\bar{\mathrm{H}}^2(\Omega)} \\ &\leq CH^{\ell+1}\big(\|f^0\|_{\mathrm{H}^{m+\ell+1}(\Omega)} + \|f^1\|_{\mathrm{H}^{m+\ell+1}(\Omega)}\big)\|\varphi_g\|_{\mathrm{H}^2(\Omega)}, \end{aligned} \tag{A.46}$$

where we also used (A.19) for the bound

$$\|I_H\varphi_g\|_{\bar{\mathrm{H}}^2(\Omega)} \leq \|\varphi_g - I_H\varphi_g\|_{\bar{\mathrm{H}}^2(\Omega)} + \|\varphi_g\|_{\bar{\mathrm{H}}^2(\Omega)} \leq C\|\varphi_g\|_{\mathrm{H}^2(\Omega)}.$$

To bound the third term, we first rewrite it as

$$A_H(u_H, v_H) - A(u_H, v_H) = A_H(u_H - I_Hu, v_H) - A(u_H - I_Hu, v_H) + A_H(I_Hu, v_H) - A(I_Hu, v_H),$$

and then use Theorem A.3.6 i) and Theorem A.3.9 i) to get

$$\begin{aligned} |A_H(u_H, I_H\varphi_g) - A(u_H, I_H\varphi_g)| \leq{}& CH \max_{ij} \|a_{ij}\|_{\mathrm{W}^{1,\infty}(\Omega)}\|u_H - I_Hu\|_{\mathrm{H}^1(\Omega)}\|\varphi_g\|_{\mathrm{H}^1(\Omega)} \\ &+ CH^{\ell+1} \max_{ij} \|a_{ij}\|_{\mathrm{W}^{\ell+1,\infty}(\Omega)}\|u\|_{\mathrm{H}^{\ell}(\Omega)}\|\varphi_g\|_{\mathrm{H}^2(\Omega)}, \end{aligned} \tag{A.47}$$

where again we used (A.19) to bound $\|I_H\varphi_g\|_{\bar{\mathrm{H}}^j(\Omega)} \leq C\|\varphi_g\|_{\mathrm{H}^j(\Omega)}$, $j = 1, 2$ and $\|I_Hu\|_{\bar{\mathrm{H}}^{\ell}(\Omega)} \leq C\|u\|_{\mathrm{H}^{\ell}(\Omega)}$. We now combine (A.44), (A.45), (A.46) and (A.47) with the triangle inequality

$\|u_H - I_H u\|_{\mathrm{H}^1(\Omega)} \leq \|u - I_H u\|_{\mathrm{H}^1(\Omega)} + \|u - u_H\|_{\mathrm{H}^1(\Omega)}$ and recall that $\|\varphi_g\|_{\mathrm{H}^2(\Omega)} \leq C\|g\|_{\mathrm{L}^2(\Omega)}$ to obtain

$$\begin{aligned}|(u-u_H, g)_{\mathrm{L}^2(\Omega)}| \leq\ & CH\|u-u_H\|_{\mathrm{H}^1(\Omega)}\|g\|_{\mathrm{L}^2(\Omega)} \\ & + CH^{\ell+1}\big(\|f^0\|_{\mathrm{H}^{m+\ell+1}(\Omega)} + \|f^1\|_{\mathrm{H}^{m+\ell+1}(\Omega)} + \max_{ij}\|a_{ij}\|_{\mathrm{W}^{\ell+1,\infty}(\Omega)}\big)\|u\|_{\mathrm{H}^{\ell+1}(\Omega)}\|g\|_{\mathrm{L}^2(\Omega)}.\end{aligned}$$

Finally, we use (A.43) and Theorem A.3.7 to prove (A.42) and the proof is complete. □

A.4 Trigonometric interpolation and spectral methods

The spectral method, is extremely accurately method to approximate smooth solutions. Indeed, under some high regularity requirements, the approximation is proved to reach so-called spectral accuracy as their rate of approximation is exponential. The analysis of the spectral method relies essentially on the study of interpolation by trigonometric polynomials, which are the Fourier basis functions. In this section, we prove error estimates for the interpolation of periodic functions by trigonometric polynomials. In particular, we define a Sobolev norm based on the Fourier coefficients that allow to track the dependence of the estimate on the domain. We refer to [59, 68, 69, 58, 89, 29, 25, 63] for the full theory on spectral method and to [91] for its implementation. Trigonometric polynomials can also simply be used to differentiate smooth functions. We also give simple Matlab implementations of the Fourier differencing method and the spectral method for the wave equation (introduced in Section 2.3). The code uses the Fast Fourier Transform (FFT) algorithm (see [62], [56]).

A.4.1 Basics of Fourier analysis for periodic functions

The fundamental question of Fourier analysis, is to ask what functions can be written as a linear combination of smooth trigonometric functions. This question has been studied extensively in the 19th and 20th centuries. Many advances were done until finally the following famous result was proved by Lennart Carleson in [31]: for $\Omega \subset \mathbb{R}$, any $v \in \mathrm{L}^2_{\mathrm{per}}(\Omega)$ coincides with its Fourier series in $\mathrm{L}^2(\Omega)$ (i.e. almost everywhere in Ω). This result was then generalized to several dimensions in [50, 49]. A considerable literature is available on Fourier analysis and its applications (for example [88, 55]). We introduce here only the objects and results needed in the scope of this thesis, which is a non exhaustive part of this vast topic.

Let $\Omega \subset \mathbb{R}^d$ be a periodic hypercube, $\Omega = (a_1, b_1) \times \cdots \times (a_d, b_d)$ and denote F_Ω the bijective affine mapping

$$F_\Omega : \mathbb{R}^d \to \mathbb{R}^d, \quad \hat{x} \mapsto F_\Omega(\hat{x}) = B_\Omega \hat{x} + a, \tag{A.48}$$

where B_Ω is the diagonal matrix defined as $(B_\Omega)_{jj} = (b_j - a_j)/(2\pi)$. We verify that $F_\Omega((0,2\pi)^d) = \Omega$. We consider the Fourier basis of $\mathrm{L}^2_{\mathrm{per}}(\Omega)$ denoted $\{w_k\}_{k\in\mathbb{Z}^d}$. Explicitly, for $k \in \mathbb{Z}^d$, $k \neq 0$

$$w_k(x) = C_\Omega \prod_{\nu=1}^{d} \exp\left(\mathrm{i}k_\nu \frac{2\pi(x_\nu - a_\nu)}{b_\nu - a_\nu}\right) = C_\Omega e^{\mathrm{i}k\cdot F_\Omega^{-1}(x)},$$

where the scaling $C_\Omega = |\Omega|^{-1/2} = \big(\prod_{\nu=1}^d b_\nu - a_\nu\big)^{-1/2}$, ensures that the basis $\{w_k\}_{k\in\mathbb{Z}^d}$ is orthonormal. Hence, for any $v \in \mathrm{L}^2_{\mathrm{per}}(\Omega)$ the equality

$$v = \sum_{k\in\mathbb{Z}^d} (v, w_k)_{\mathrm{L}^2} w_k \qquad \text{in } \mathrm{L}^2_{\mathrm{per}}(\Omega). \tag{A.49}$$

The right term in equality (A.49) is known as the *Fourier series* or *Fourier expansion* of v. It is

more commonly written as

$$v(x) \overset{\mathrm{L}^2}{=} \sum_{k\in\mathbb{Z}^d} \hat{v}(k) e^{\mathrm{i}k\cdot F_\Omega^{-1}(x)}, \qquad \hat{v}(k) = \frac{1}{|\Omega|}\int_\Omega v(x) e^{-\mathrm{i}k\cdot F_\Omega^{-1}(x)}\,\mathrm{d}x. \tag{A.50}$$

Remark A.4.1. As the index of the series in (A.49) and (A.50) belongs to $\mathbb{Z}^d$, we need to explain the meaning of the limit, i.e., we have to define what its partial sum is. As discussed in [49], (A.49) does not hold for any definition of partial sum (e.g. taking the limit successively with respect to each dimension leads to counter examples of (A.49)). We follow here [50, 49] to define the meaning of the series $\sum_{k\in\mathbb{Z}^d} A_k$, where $\{A_k\}_{k\in\mathbb{Z}^d}$ is a sequence indexed by a multi-index $k \in \mathbb{Z}^d$. Let P be an open polygon of $\mathbb{R}^d$ containing the origin and let $P_\lambda = \{\lambda x : x \in P\} \subset \mathbb{R}^d$ for some $\lambda > 0$. We define the *partial sum* S_λ and the corresponding series S as

$$S_\lambda = \sum_{k\in\lambda P\cap\mathbb{Z}^d} A_k, \qquad S = \lim_{\lambda\to\infty} S_\lambda. \tag{A.51}$$

The series S is denoted $\sum_{k\in\mathbb{Z}^d} A_k$.

From (A.49) and the orthogonality of the basis $\{w_k\}_{k\in\mathbb{Z}^d}$ follows Plancherel formula

$$\|v\|^2_{\mathrm{L}^2(\Omega)} = \sum_{k\in\mathbb{Z}^d} |(v, w_k)_{\mathrm{L}^2(\Omega)}|^2 = |\Omega| \sum_{k\in\mathbb{Z}^d} |\hat{v}(k)|^2. \tag{A.52}$$

Let us use this formula to define Sobolev norms that are convenient in the context of this thesis. For simplicity consider first the periodic hypercube $\Omega = \mathbb{T}^d = (0, 2\pi)^d$ (so that $F_\Omega = \mathrm{Id}$). Note that the Fourier coefficients of the derivatives of $v \in \mathrm{H}^s(\mathbb{T}^d)$ are given by (see (A.50)) $\widehat{\partial^\alpha v}(k) = (\mathrm{i}k)^\alpha \hat{v}(k)$, where $k^\alpha = k_1^{\alpha_1}\cdots k_d^{\alpha_d}$. We can thus write the H^s seminorm as

$$|v|^2_{\mathrm{H}^s(\mathbb{T}^d)} = \sum_{k\in\mathbb{Z}^d} \sum_{|\alpha|_1=s} k^{2\alpha} |\hat{v}(k)|^2.$$

Note that by convention the notation $|\cdot|$ denotes in general the 2-norm, i.e., $|x| = \sqrt{x_1^2 + \ldots + x_d^2}$ for $x \in \mathbb{R}^d$. For multiindices in $\mathbb{N}^d$, $|\cdot|_1$ denotes the 1-norm, i.e., $|\alpha|_1 = |\alpha_1| + \ldots + |\alpha_d|$ for $\alpha \in \mathbb{N}^d$. The multinomial formula gives

$$|k|^{2s} = (k_1^2 + \ldots + k_d^2)^s = \sum_{|\alpha|_1=s} \binom{s}{\alpha} k^{2\alpha} = \sum_{|\alpha|_1=s} \binom{s}{\alpha_1\cdots\alpha_d} k_1^{2\alpha_1}\cdots k_d^{2\alpha_d}, \qquad \binom{s}{\alpha} = \frac{s!}{\alpha_1!\cdots\alpha_d!}.$$

As $\binom{s}{\alpha} \geq 1$ for all $|\alpha|_1 = s$, we verify that

$$\sum_{|\alpha|_1=s} k^{2\alpha} \leq |k|^{2s} \leq C(d,s) \sum_{|\alpha|_1=s} k^{2\alpha}, \tag{A.53}$$

where $C(d,s) = \max_{|\alpha|_1=s} \binom{s}{\alpha}$. This relation ensures that the quantity

$$|v|^2_{\tilde{\mathrm{H}}^s(\mathbb{T}^d)} = \sum_{k\in\mathbb{Z}^d} |k|^{2s} |\hat{v}(k)|^2,$$

is a seminorm equivalent to $|\cdot|^2_{\mathrm{H}^s(\mathbb{T}^d)}$: $|v|_{\mathrm{H}^s(\mathbb{T}^d)} \leq |v|_{\tilde{\mathrm{H}}^s(\mathbb{T}^d)} \leq \sqrt{C(d,s)}|v|_{\mathrm{H}^s(\mathbb{T}^d)}$. For a general hypercube Ω, the same reasoning leads to the definition

$$|v|^2_{\tilde{\mathrm{H}}^s(\Omega)} = \sum_{k\in\mathbb{Z}^d} \left|B_\Omega^{-1} k\right|^{2s} |\hat{v}(k)|^2, \tag{A.54}$$

where B_Ω is the scaling matrix in F_Ω (A.48), i.e., $(B_\Omega^{-1}k)_\nu = 2\pi/(b_\nu - a_\nu)k_\nu$. Relation (A.53) ensures the equivalence

$$|v|_{\mathrm{H}^s(\Omega)} \leq |v|_{\tilde{\mathrm{H}}^s(\Omega)} \leq \sqrt{C(d,s)}|v|_{\mathrm{H}^s(\Omega)}, \tag{A.55}$$

where we emphasize that the constant $C(d,s)$ does not depend on Ω. Finally, we define the $\tilde{\mathrm{H}}^m$ norm as

$$\|v\|^2_{\tilde{\mathrm{H}}^s(\Omega)} = \sum_{m=0}^{s} |v|^2_{\tilde{\mathrm{H}}^m(\Omega)}, \tag{A.56}$$

and it is equivalent to the standard H^s norm. Note that the classical definition of the Sobolev norm of order m using Fourier analysis is $\|v\|^2 = \sum_{k\in\mathbb{Z}^d} \left(1+|k|^2\right)^m |\hat{v}(k)|^2$, which is equivalent to the $\tilde{\mathrm{H}}^m$ norm defined in (A.56). In particular, this definition allow to generalize the Sobolev space $\mathrm{H}^s(\Omega)$ of integer order to real orders. However, this is not the purpose here and (A.54) and (A.56) are more convenient in our analysis, as we want to track the dependence of our estimates in the domain Ω.

A.4.2 Interpolation of periodic functions by trigonometric polynomials in 1d

In this section, we define the space of trigonometric polynomials in one dimension. In particular, we define an interpolant for periodic functions and estimate the interpolation error. The proof of the result is inspired by [89]. Note that the theory is generalized to the multidimensional case and arbitrary hypercubes in Section A.4.4.

Let us consider a 2π-periodic function $v \in \mathrm{L}^2_{\mathrm{per}}(0,2\pi)$. The Fourier expansion (A.50) for v reads

$$v(x) \overset{\mathrm{L}^2}{=} \sum_{k\in\mathbb{Z}} \hat{v}(k) e^{\mathrm{i}kx}, \qquad \hat{v}(k) = \frac{1}{2\pi}\int_0^{2\pi} v(x) e^{-\mathrm{i}kx}\,\mathrm{d}x. \tag{A.57}$$

The basis functions $e^{\mathrm{i}kx}$ are called *trigonometric polynomials*. Thanks to the properties of the exponential function, we verify that $\widehat{\partial_x^m v}(k) = (\mathrm{i}k)^m \hat{v}(k)$. We would like to take advantage of this relation to approximate the derivatives of v. In that purpose, we need to define a convenient finite dimensional subspace of $\mathrm{L}^2_{\mathrm{per}}(0,2\pi)$. For a given integer $N \in \mathbb{N}_{>0}$, we consider the uniform grid of $(0,2\pi)$ of size $h = \pi/N$:

$$G_N = \{x_n = nh : 0 \leq n \leq 2N-1\}.$$

We verify that the set $\{(e^{\mathrm{i}kx_0}, \ldots, e^{\mathrm{i}kx_{2N-1}})^T : k = -N+1, \ldots, N\}$ is an orthogonal basis of $\mathbb{C}^{2N}$,

$$\sum_{n=0}^{2N-1} e^{\mathrm{i}kx_n}\overline{e^{\mathrm{i}\ell x_n}} = \sum_{n=0}^{2N-1} e^{\mathrm{i}\pi(k-l)n/N} = 2N\delta_{k\ell}. \tag{A.58}$$

Hence, we first consider the finite dimensional space $\tilde{V}_N(0,2\pi) = \mathrm{span}\{e^{\mathrm{i}kx} : k = -N+1, \ldots, N\}$. As $\tilde{V}_N(0,2\pi)$ is a vector space of dimension $2N$, its elements are uniquely determined by their values on the grid G_N. Note that in $\tilde{V}_N(0,2\pi)$, the wave number k is treated asymmetrically and a simple example illustrates why the symmetry should hold. Consider the sawtooth function $p \in \tilde{V}_N(0,2\pi)$, $p(x) = e^{\mathrm{i}Nx}$. The function p oscillate smoothly between the values $p(x_n) = (-1)^n$ and its derivative is zero at the grid points x_n. We thus expect that $\partial_x p = 0$ in $\tilde{V}_N(0,2\pi)$. However, we verify that $\partial_x p(x) = \mathrm{i}Ne^{\mathrm{i}Nx}$ in $\tilde{V}_N(0,2\pi)$. To solve this issue we need to symmetrize the higher wave number and we thus set

$$V_N(0,2\pi) = \mathrm{span}(B_N), \quad B_N = \left\{e^{\mathrm{i}kx} : |k| \leq N-1\right\} \cup \left\{\tfrac{1}{2}(e^{\mathrm{i}Nx} + e^{-\mathrm{i}Nx})\right\}. \tag{A.59}$$

The sawtooth function $p(x) = e^{\mathrm{i}Nx}$ does not belong to that space, but we will see that its interpolant (defined in (A.63)) has a zero derivative in $V_N(0, 2\pi)$. As $V_N(0, 2\pi)$ has dimension $2N$, $p \in V_N(0, 2\pi)$ is uniquely determined by its value on the grid G_N. Furthermore, using (A.58), we find that p is uniquely written in the basis B_N as

$$p(x) = \sum_{|k| \leq N-1} \hat{p}_k e^{\mathrm{i}kx} + \hat{p}_N \tfrac{1}{2}(e^{\mathrm{i}Nx} + e^{-\mathrm{i}Nx}), \qquad \hat{p}_k = \frac{1}{2N} \sum_{n=0}^{2N-1} p(x_n) e^{-\mathrm{i}kx_n} \quad k = -N+1, \ldots, N.$$

As $e^{\mathrm{i}Nx_n} = e^{\mathrm{i}\pi n} = e^{-\mathrm{i}Nx_n}$, we verify that the values of p on the grid are given by

$$p(x_n) = \sum_{k=-N+1}^{N} \hat{p}_k e^{\mathrm{i}kx_n} \qquad 0 \leq n \leq 2N-1. \tag{A.60}$$

If we extend the definition of $\hat{p}_k$ to $k = -N$, we verify that $\hat{p}_{-N} = \hat{p}_N$ and we can thus rewrite $p \in V_N(0, 2\pi)$ as

$$p(x) = \sum_{|k| \leq N}{}' \hat{p}_k e^{\mathrm{i}kx}, \qquad \hat{p}_k = \frac{1}{2N} \sum_{n=0}^{2N-1} p(x_n) e^{-\mathrm{i}kx_n} \quad |k| \leq N, \tag{A.61}$$

where the notation $\sum'$ indicates that the terms $k \in \{-N, N\}$ are halved. We define the following inner product and its corresponding norm on $V_N(0, 2\pi)$:

$$(p, q)_h = h \sum_{n=0}^{2N-1} p(x_n)\overline{q(x_n)}, \qquad \|p\|_h = \sqrt{(p, p)_h} \qquad \forall p, q \in V_N(0, 2\pi). \tag{A.62}$$

Using the orthogonality of the Fourier basis and the definition of $\hat{p}_k$ in (A.61), we verify that for any $p, q \in V_N(0, 2\pi)$, $(p, q)_{\mathrm{L}^2(0,2\pi)} = (p, q)_h$.

We have introduced the finite dimensional space of trigonometric polynomials $V_N(0, 2\pi)$. We now define an interpolant for $v \in \mathrm{L}^2_{\mathrm{per}}(0, 2\pi)$ onto $V_N(0, 2\pi)$. We define the *trigonometric interpolant* $I_N : \mathrm{L}^2_{\mathrm{per}}(0, 2\pi) \to V_N(0, 2\pi)$ as (compare to (A.61))

$$I_N v(x) = \sum_{|k| \leq N}{}' \hat{v}_k e^{\mathrm{i}kx}, \qquad \hat{v}_k = \frac{1}{2N} \sum_{n=0}^{2N-1} v(x_n) e^{-\mathrm{i}kx_n} \quad |k| \leq N. \tag{A.63}$$

As we verify that $\hat{v}_{-N} = \hat{v}_N$, $I_N v$ indeed belongs to $V_N(0, 2\pi)$. Note that $\hat{v}_k$ approximates the Fourier coefficients $\hat{v}(k)$ in (A.57) as

$$\hat{v}(k) = \frac{1}{2\pi} \sum_{n=0}^{2N-1} \int_{x_n}^{x_{n+1}} v(x) e^{-\mathrm{i}kx} \, \mathrm{d}x \approx \frac{1}{2\pi} \sum_{n=0}^{2N-1} \frac{\pi}{N} v(x_n) e^{-\mathrm{i}kx_n} = \hat{v}_k,$$

where the integrals are approximated with the forward Euler rule. Let us verify that the interpolant of the sawtooth function $v(x) = e^{\mathrm{i}Nx}$ seen earlier has a zero derivative in $V_N(0, 2\pi)$. Indeed, we compute $\hat{v}_k = 1$ if $k = \pm N$ and $\hat{v}_k = 0$ otherwise, and thus $\partial_x I_N v(x) = \mathrm{i}N(e^{\mathrm{i}Nx} - e^{-\mathrm{i}Nx})/2$, which vanishes on the grid $\{x_n\}$, so that $\partial_x I_N v(x) = 0$ in $V_N(0, 2\pi)$.

In [89], an a priori estimate for $\|v - I_N v\|_{\mathrm{H}^\sigma(0,2\pi)}$ is proved for any order σ. Using the same technique, we prove an estimate of the $\tilde{\mathrm{H}}^\sigma$ seminorm of the error (the $\tilde{\mathrm{H}}^\sigma$ seminorm is defined in (A.54)). This is indeed more convenient in the context of this thesis, as it can be generalized to an error estimate with an explicit dependence on the domain.

Theorem A.4.2. *Let v be a 2π-periodic function such that $v \in \mathrm{H}^s(0,2\pi)$, for some $s > 1/2$. Then, for any $\sigma \le s$, the trigonometric interpolant I_N defined in* (A.63) *satisfies the estimate*

$$|v - I_N v|_{\tilde{\mathrm{H}}^\sigma(0,2\pi)} \le C(s)\frac{1}{N^{s-\sigma}}|v|_{\tilde{\mathrm{H}}^s(0,2\pi)}, \tag{A.64}$$

where $C(s) = \left(1 + 2\sum_{\ell=1}^{\infty}(\ell-1)^{-2s}\right)^{1/2}$.

Note that the constant $C(s)$ in Theorem A.4.2 has a fast decay for $1/2 < s < 1$ and $C(s) \le C(1) = \sqrt{1+\pi^2/2}$ for $s \ge 1$. Estimate (A.64) implies that if $v \in \mathcal{C}^\infty_{\mathrm{per}}(0,2\pi)$, then $\|\partial_x^m v - \partial_x^m I_N v\|_{\mathrm{L}^2(0,2\pi)} = \mathcal{O}(N^{-s})$ for any s and thus $\partial_x^m I_N v$ converges to $\partial_x^m v$ with an arbitrarily large order of convergence. These are strong theoretical results. In practice however, making use of the error estimate is difficult as the quantity $\|v\|_{\mathrm{H}^s(0,2\pi)}$ might be difficult to estimate for large values of s. For more practical estimate, we refer to [89].

In order to prove Theorem A.4.2, we need the following lemma.

Lemma A.4.3. (Aliasing) *Let v be a 2π-periodic function such that $v \in \mathrm{H}^s(0,2\pi)$, with $s > 1/2$. Then*

$$\hat{v}_k = \sum_{\ell\in\mathbb{Z}} \hat{v}(k+2\ell N) \quad \textit{for } |k| \le N. \tag{A.65}$$

Proof. Using (A.57) in the definition of $\hat{v}_k$ in (A.63), we have

$$\hat{v}_k = \frac{1}{2N}\sum_{n=0}^{2N-1}\left(\sum_{j\in\mathbb{Z}}\hat{v}(j)e^{\mathrm{i}jx_n}\right)e^{-\mathrm{i}kx_n} = \sum_{j\in\mathbb{Z}}\hat{v}(j)\left(\frac{1}{2N}\sum_{n=0}^{2N-1}e^{2\pi\mathrm{i}n\frac{j-k}{2N}}\right).$$

We define the set of index $S = \{j \in \mathbb{Z} : j = k + 2\ell N \text{ for some } \ell \in \mathbb{Z}\}$ and $S^c = \mathbb{Z}\backslash S$. For $j \in S$, we have $\ell = \frac{j-k}{2N} \in \mathbb{Z}$ and thus $e^{2\pi\mathrm{i}n\frac{j-k}{2N}} = e^{2\pi\mathrm{i}n\ell} = 1$. For $j \in S^c$, we have $a = \frac{j-k}{2N} \notin \mathbb{Z}$ and thus $e^{2\pi\mathrm{i}na} \ne 1$. Hence, for $j \in S^c$ we have $\sum_{n=0}^{2N-1} e^{2\pi\mathrm{i}na} = \frac{1-e^{2\pi\mathrm{i}2Na}}{1-e^{2\pi\mathrm{i}a}} = 0$, as we verify that $2Na = j - k \in \mathbb{Z}$. Finally, we obtain

$$\hat{v}_k = \sum_{j\in S}\hat{v}(j)\left(\frac{1}{2N}\sum_{n=0}^{2N-1}e^{2\pi\mathrm{i}n\frac{j-k}{2N}}\right) = \sum_{j\in S}\hat{v}(j) = \sum_{\ell\in\mathbb{Z}}\hat{v}(k+2\ell N),$$

which proves (A.65) and completes the proof of the lemma. □

Proof of Theorem A.4.2. Using (A.63) and Lemma A.4.3, we write

$$I_N v(x) = {\sum_{|k|\le N}}' \hat{v}(k)e^{\mathrm{i}kx} + {\sum_{|k|\le N}}'\left(\sum_{\ell\ne 0}\hat{v}(k+2\ell N)\right)e^{\mathrm{i}kx}.$$

Using (A.57), we compute explicitly the error as

$$(v - I_N v)(x) = -\sum_{|k|<N}\left(\sum_{\ell\in\mathbb{Z}}\hat{v}(k+2\ell N)\right)e^{\mathrm{i}kx} + \sum_{|k|=N}\frac{1}{2}\left(\hat{v}(k) - \sum_{\ell\ne 0}\hat{v}(k+2\ell N)\right)e^{\mathrm{i}kx} + \sum_{|k|>N}\hat{v}(k)e^{\mathrm{i}kx}.$$

Then, the $\tilde{\mathrm{H}}^\sigma$ norm of the error satisfies (see (A.56))

$$|v - I_N v|^2_{\tilde{\mathrm{H}}^\sigma} \le {\sum_{|k|\le N}}' |k|^{2\sigma}|\hat{v}(k)|^2 + {\sum_{|k|\le N}}' |k|^{2\sigma}\left|\sum_{\ell\ne 0}\hat{v}(k+2\ell N)\right|^2 =: E_1 + E_2. \tag{A.66}$$

As $\sigma \le s$, the first term can be bounded as

$$E_1 \le \sum_{|k|\ge N}{}' \frac{1}{|k|^{2(s-\sigma)}} |k|^{2s} |\hat{v}(k)|^2 \le \frac{1}{N^{2(s-\sigma)}} \|v\|^2_{\mathring{\mathrm{H}}^s}. \tag{A.67}$$

Let us estimate the second term E_2. Using Cauchy–Schwartz inequality in ℓ^2, we have, for $|k| \le N$,

$$\left| \sum_{\ell\neq 0} \hat{v}(k+2\ell N) \right|^2 \le \left(\frac{1}{N^{2s}} \sum_{\ell\neq 0} \frac{1}{|k+2\ell N|^{2s} N^{-2s}} \right) \left(\sum_{\ell\neq 0} |k+2\ell N|^{2s} |\hat{v}(k+2\ell N)|^2 \right).$$

Furthermore, using the reverse triangle inequality, we verify that

$$|k+2\ell N| = |k-(-2\ell N)| \ge \big| |k| - 2|\ell| N \big| = 2|\ell| N - |k|,$$

where we used that for $\ell \neq 0$ and $|k| \le N$, we have $2|\ell|N \ge |k|$. As $|k|/N \le 1$, it holds $|k+2\ell N| N^{-1} \ge 2|\ell| - 1$, and thus,

$$E_2 \le \frac{2}{N^{2(s-\sigma)}} \left(\sum_{\ell=1}^{\infty} \frac{1}{(2\ell-1)^{2s}} \right) \sum_{|k|\le N}{}' \sum_{\ell\neq 0} |k+2\ell N|^{2s} |\hat{v}(k+2\ell N)|^2.$$

A careful study of the double sum reveals that the only indices appearing twice in the total sum correspond to $k = \pm N$: indeed, if $(k_1, \ell_1) = (N, \ell)$ and $(k_2, \ell_2) = (-N, \ell+1)$ then

$$k_1 + 2\ell_1 N = N + 2\ell N = (-N) + 2(\ell+1)N = k_2 + 2\ell_2 N.$$

These double terms are thus exactly removed by $\sum'$ and thus the double sum is bounded by $\|v\|^2_{\mathring{\mathrm{H}}^s}$. We thus obtain

$$E_2 \le \frac{2}{N^{2(s-\sigma)}} \sum_{\ell=1}^{\infty} \frac{1}{(2\ell-1)^{2s}} \|v\|^2_{\mathring{\mathrm{H}}^s},$$

which, combined with (A.66) and (A.67), proves estimate (A.64). The proof of the theorem is complete. □

A.4.3 The Fourier differencing method in one dimension and its implementation

One of the properties of the trigonometric polynomials is that they are easily differentiable. In particular, the trigonometric interpolant naturally leads to the Fourier differencing method. In this section, we introduce this method. Note that it is generalized in Section A.4.5.

Recall the definition of the trigonometric interpolant in (A.63). The spectral derivative of a function is defined as the derivative of its trigonometric interpolant, i.e., for $v \in \mathrm{H}^m_{\mathrm{per}}(0, 2\pi)$, we approximate

$$\partial_x^m v \approx \partial_x^m I_N v \in V_N(0, 2\pi).$$

For $v \in \mathrm{H}^{s+m}_{\mathrm{per}}(0, 2\pi)$ with $s > 1/2$, Theorem A.4.2 ensures the error estimate

$$\|v - I_N v\|_{\mathrm{H}^m(0,2\pi)} \le C \frac{1}{N^s} \|v\|_{\mathrm{H}^{s+m}(0,2\pi)},$$

and the method converges as $N \to \infty$.

Let us explain how to apply the method and actually compute the approximation of the derivatives of a given function. Let E be the map of evaluation on the grid G_N, i.e., $v \in \mathrm{L}^2_{\mathrm{per}}(0, 2\pi) \mapsto Ev =$

$(v(x_0),\dots,v(x_{2N-1}))^T$. We have seen in the previous section that $E|_{V_N(0,2\pi)} : V_N(0,2\pi) \to \mathbb{C}^{2N}$ is an isomorphism. We define the *discrete Fourier transform* (DFT) as the map $\mathcal{F}_h : \mathbb{C}^{2N} \to \mathbb{C}^{2N}$, $V \mapsto \mathcal{F}_h(V)$, where

$$\mathcal{F}_h(V)_k = \frac{1}{2N} \sum_{n=0}^{2N-1} V_n e^{-\mathrm{i}kx_n} \qquad k = -N+1,\dots,N.$$

The *inverse DFT* (iDFT) is the map $\mathcal{F}_h^{-1} : \mathbb{C}^{2N} \to \mathbb{C}^{2N}$, $\hat{V} \mapsto \mathcal{F}_h^{-1}(\hat{V})$, where

$$\mathcal{F}_h^{-1}(\hat{V})_n = \sum_{|k|\le N}{}' \hat{V}_k e^{\mathrm{i}kx_n} \qquad n = 0,\dots,2N-1.$$

With these definitions, equality (A.60) reads

$$(Ep)_n = \sum_{k=-N+1}^{N} \big(\mathcal{F}_h \circ E(p)\big)_k e^{\mathrm{i}kx_n} \qquad 0 \le n \le 2N-1.$$

Let us define the map of differentiation in the Fourier space as $\hat{D}^m : \mathbb{C}^{2N} \to \mathbb{C}^{2N}$, $\hat{V} \mapsto D\hat{V}$, where

$$\begin{aligned} &(\hat{D}^m\hat{V})_k = (\mathrm{i}k)^m \hat{V}_k \quad |k| \le N-1, \quad (\hat{D}^m\hat{V})_N = 0, && \text{if } m \text{ is odd,}\\ &(\hat{D}^m\hat{V})_k = (\mathrm{i}k)^m \hat{V}_k \quad k = -N+1,\dots,N, && \text{if } m \text{ is even.} \end{aligned}$$

Finally, we define the differentiation map as

$$D^m : \mathbb{C}^{2N} \to \mathbb{C}^{2N}, \quad V \mapsto D^m V = \mathcal{F}_h^{-1} \circ \hat{D}^m \circ \mathcal{F}_h(V).$$

For a function $v \in \mathrm{H}^m_{\mathrm{per}}(0,2\pi)$, the approximation of $\partial_x^m v$ is then defined as

$$\partial_x I_N v = (E|_{V_N(0,2\pi)})^{-1} \circ D^m \circ E(v),$$

and can be computed on the grid G_N as

$$\partial_x^m v(x_n) \approx \partial_x^m I_N v(x_n) = \big(\mathcal{F}_h^{-1} \circ \hat{D}^m \circ \mathcal{F}_h \circ E(v)\big)_n \quad 0 \le n \le 2N-1.$$

Note that the DFT $\mathcal{F}_h$ and the iDFT $\mathcal{F}_h^{-1}$ can be computed respectively by the fast Fourier transform algorithms (FFT) and inverse fast Fourier transform algorithms (iFFT), whose complexities are $\mathcal{O}(N\log(N))$ (possibly less, depending on the prime decomposition of N, see [62], [56]).

In Program A.1, we present an example of implementation of the Fourier differencing method using Matlab. Note that the Matlab implementation of iFFt require to shift the vectors in the Fourier space (see the variable `k` and Matlab's help on the function `fft`). Note that for data in $\mathbb{R}$, we have the relation $\hat{v}_{-k} = \overline{\hat{v}_k}$. Some implementations of FFT and iFFT take advantage of this symmetry to improve the performance by 2. While this feature is not available in the native Matlab functions, it can be used in the FFTW library (see [56]).

A.4.4 Interpolation of general periodic functions by trigonometric polynomials

In this section, we generalize the theory on trigonometric polynomials introduced in one dimension in Section A.4.2. We first define the space of trigonometric polynomials defined in the torus $\mathbb{T}^d$, where $\mathbb{T} = (0,2\pi)$, and then adapt it to any hypercube $\Omega \subset \mathbb{R}^d$. In particular, we prove an error estimate for the trigonometric interpolant. We emphasize that thanks to the definition of the

Program A.1: Matlab implementation of the Fourier differencing method in 1d.

```
% function v and its derivatives
v   = @(x) exp(sin(x));
d1v = matlabFunction( diff(sym(v),'x') );
d2v = matlabFunction( diff(sym(d1v),'x') );
% discretization
N = 8;
h = pi/N;
x = (0:h:(2*pi-h))';
% approximation of the derivatives
Ev = v(x);
k = fftshift((-N:N-1)');
ifOdd = (k~=N);
D1v = real(ifft( 1i*k.*ifOdd .*fft(Ev) ));
D2v = real(ifft( (1i*k).^2   .*fft(Ev) ));
% error and plots
fprintf('|d1v-D1v|=%g\n', max(abs(d1v(x)-D1v)));
fprintf('|d2v-D2v|=%g\n', max(abs(d2v(x)-D2v)));
figure;
xf = (0:1e-3:2*pi)';
subplot(1,2,1);    plot(x,D1v,'bo',xf,d1v(xf),'k');
subplot(1,2,2);    plot(x,D2v,'bo',xf,d2v(xf),'k');
```

particular Sobolev seminorm $|\cdot|_{\tilde{\mathrm{H}}^s}$, in (A.54), we are able to track the dependence of the estimate in Ω. For the sake of clarity, we use the convention that $\bar{x} \in \mathbb{T}^d$ and $x \in \Omega$.

We consider a periodic function on the torus $\mathbb{T}^d$, $v \in \mathrm{L}^2_{\mathrm{per}}(\mathbb{T}^d)$. The Fourier expansion (A.50) for v is

$$v(\bar{x}) \overset{\mathrm{L}^2}{=} \sum_{k \in \mathbb{Z}^d} \hat{v}(k) e^{\mathrm{i}k\cdot\bar{x}}, \qquad \hat{v}(k) = \frac{1}{|\mathbb{T}^d|} \int_{\mathbb{T}^d} v(\bar{x}) e^{-\mathrm{i}k\cdot\bar{x}} \, \mathrm{d}\bar{x}. \tag{A.68}$$

For a given $N \in \mathbb{N}^d_{>0}$, we consider a uniform grid of $\mathbb{T}^d$

$$\bar{G}_N = \big\{ \bar{x}_{n_1 \cdots n_d} = (n_1 \bar{h}_1, \ldots, n_d \bar{h}_d)^T : \ 0 \le n_1 \le 2N_1 - 1, \ldots, 0 \le n_d \le 2N_d - 1 \big\},$$

where the mesh size in each direction is $\bar{h}_\nu = \pi / N_\nu$. We define the finite dimensional space of trigonometric polynomials as

$$V_N(\mathbb{T}^d) = \mathrm{span} B_N, \quad B_N = \big\{ \textstyle\prod_{\nu=1}^d \bar{p}_\nu(\bar{x}) : \bar{p}_\nu \in B_{N_\nu} \big\}, \tag{A.69}$$

where B_{N_ν} is the basis of $V_{N_\nu}(\mathbb{T})$, the one-dimensional space of trigonometric polynomials of order N_ν defined in (A.59). Using (A.58), we verify that a trigonometric polynomial $p \in V_N(\mathbb{T}^d)$ can be written as (A.72) (with $\Omega = \mathbb{T}^d$, i.e., $F_\Omega = Id$). We thus define the trigonometric interpolant of $v \in \mathrm{L}^2_{\mathrm{per}}(\mathbb{T}^d)$ as

$$I_N v(\bar{x}) = \sum_{|k_1| \le N_1}{}' \cdots \sum_{|k_d| \le N_d}{}' \hat{v}_{k_1 \cdots k_d} e^{ik\cdot\bar{x}}, \tag{A.70}$$

$$\hat{v}_{k_1 \cdots k_d} = \frac{1}{2N_1} \sum_{n_1=0}^{2N_1-1} \cdots \frac{1}{2N_d} \sum_{n_d=0}^{2N_d-1} v(\bar{x}_{n_1 \cdots n_d}) e^{-ik_1 n_1 \bar{h}_1} \cdots e^{-ik_d n_d \bar{h}_d} \quad -N_\nu \le k_\nu \le N_\nu.$$

Let us generalize the space of trigonometric polynomials and of the interpolant to any hypercubes. Let $\Omega \subset \mathbb{R}^d$ be a hypercube given by $\Omega = (a_1, b_1) \times \cdots \times (a_d, b_d)$. Let F_Ω be the bijective affine mapping

$$F_\Omega : \mathbb{T}^d \to \Omega, \quad \bar{x} \mapsto F_\Omega(\bar{x}) = B_\Omega \bar{x} + a,$$

where B_Ω is the diagonal matrix defined as $(B_\Omega)_{jj} = (b_j - a_j)/(2\pi)$. For $N \in \mathbb{N}^d_{>0}$, let

$$G_N = \left\{ x_{n_1 \cdots n_d} = F_\Omega(\bar{x}_{n_1 \cdots n_d}) \forall \bar{x}_{n_1 \cdots n_d} \in \bar{G}_N \right\}.$$

be the uniform grid of Ω. We verify that the size of the grid G_N in each direction is $h_\nu = (b_\nu - a_\nu)/(2N_\nu)$. The finite dimensional space of trigonometric polynomials in $\mathrm{L}^2_{\mathrm{per}}(\Omega)$ is defined as

$$V_N(\Omega) = \mathrm{span}\{p = \bar{p} \circ F_\Omega^{-1} : \bar{p} \in B_N\}, \tag{A.71}$$

where B_N is defined in (A.69). We verify that $p \in V_N(\Omega)$, can be written as

$$\begin{aligned} p(x) &= \sum_{|k_1| \le N_1}{}' \cdots \sum_{|k_d| \le N_d}{}' \hat{p}_{k_1 \cdots k_d} e^{ik \cdot F_\Omega^{-1}(x)}, \\ \hat{p}_{k_1 \cdots k_d} &= \frac{1}{2N_1} \sum_{n_1=0}^{2N_1-1} \cdots \frac{1}{2N_d} \sum_{n_d=0}^{2N_d-1} p(x_{n_1 \cdots n_d}) e^{-ik_1 n_1 h_1} \cdots e^{-ik_d n_d h_d} \quad - N_\nu \le k_\nu \le N_\nu. \end{aligned} \tag{A.72}$$

We define the following inner product and corresponding norm on $V_N(\Omega)$

$$(p,q)_N = h_1 \sum_{n_1=0}^{2N_1-1} \cdots h_d \sum_{n_d=0}^{2N_d-1} p(x_{n_1 \cdots n_d}) \overline{q(x_{n_1 \cdots n_d})} \qquad \|p\|_N = \sqrt{(p,p)_h} \qquad \forall p, q \in V_N(\Omega). \tag{A.73}$$

Using the orthogonality of the Fourier basis and the definition of $\hat{p}_{k_1 \cdots k_d}$ in (A.72), we can show that for any $p, q \in V_N(\Omega)$, $(p,q)_{\mathrm{L}^2(\Omega)} = (p,q)_N$.

The trigonometric interpolant $\mathrm{L}^2_{\mathrm{per}}(\Omega) \to V_N(\Omega)$ is then defined as $I_N^\Omega v = I_N(v \circ F_\Omega) \circ F_\Omega^{-1}$, i.e.,

$$\begin{aligned} I_N^\Omega v(x) &= \sum_{|k_1| \le N_1}{}' \cdots \sum_{|k_d| \le N_d}{}' \hat{v}_{k_1 \cdots k_d} e^{\mathrm{i}k \cdot F_\Omega^{-1}(x)}, \\ \hat{v}_{k_1 \cdots k_d} &= \frac{1}{2N_1} \sum_{n_1=0}^{2N_1-1} \cdots \frac{1}{2N_d} \sum_{n_d=0}^{2N_d-1} v(x_{n_1 \cdots n_d}) e^{-ik_1 n_1 \bar{h}_1} \cdots e^{-ik_d n_d \bar{h}_d}, \end{aligned} \tag{A.74}$$

where we used the fact that $v \circ F_\Omega(\bar{x}_{n_1 \cdots n_d}) = v(x_{n_1 \cdots n_d})$. Note that this definition agrees with the Fourier expansion given in (A.50), where the change of variables $\bar{x} = F_\Omega^{-1}(x)$ leads to rewriting

$$v(x) \overset{\mathrm{L}^2}{=} \sum_{k \in \mathbb{Z}^d} \hat{v}(k) e^{\mathrm{i}k \cdot F_\Omega^{-1}(x)}, \qquad \hat{v}(k) = \frac{1}{|\mathbb{T}^d|} \int_{\mathbb{T}^d} v \circ F_\Omega(\bar{x}) e^{-\mathrm{i}k \cdot \bar{x}} \, \mathrm{d}\bar{x}.$$

Indeed, observe that $\hat{v}_{k_1 \cdots k_d} \approx \hat{v}(k)$.

We prove the following generalization of Theorem A.4.2, for $d \le 3$: an estimate of $\left| v - I_N^\Omega v \right|_{\tilde{\mathrm{H}}^\sigma}$ for any order σ (the $\tilde{\mathrm{H}}^\sigma$ seminorm is defined in (A.54)). In particular, we emphasize that the constant in (A.75) does not depend on the domain Ω.

Theorem A.4.4. *Let $d \le 3$ and assume that the ratio $r(N) = N_{\max}/N_{\min}$ is bounded. Let v be an Ω-periodic function such that $v \in \mathrm{H}^s(\Omega)$, for some $s \ge (d+1)/2$. Then, for any $\sigma \le s$, the trigonometric interpolant I_N^Ω defined in* (A.74) *satisfies the estimate*

$$\left| v - I_N^\Omega v \right|_{\tilde{\mathrm{H}}^\sigma(\Omega)} \le C \frac{r(N)^{s-\sigma}}{\left| B_\Omega^{-1} N \right|^{s-\sigma}} |v|_{\tilde{\mathrm{H}}^s(\Omega)}, \tag{A.75}$$

where C is a constant depending only on d and $r(N)$.

Proof. For simplicity, we prove the result for $\Omega = \mathbb{T}^d$. The proof for a general hypercube Ω follows the same line, with the basis functions $e^{\mathrm{i}k\cdot F_\Omega^{-1}(x)}$. First, using the properties of the exponential, we prove the aliasing relation (see Lemma A.4.3).

$$\hat{v}_{k_1\cdots k_d} = \sum_{\ell\in\mathbb{Z}^d} \hat{v}(k_1 + 2\ell_1 N_1, \dots, k_d + 2\ell_d N_d) \qquad \text{for } |k_1| \le N_1, \dots, |k_d| \le N_d. \tag{A.76}$$

Let us decompose the index set $\mathbb{Z}^d$ into the disjoint sets

$$\begin{aligned} K_> &= \{k \in \mathbb{Z}^d : |k_\nu| \ge N_\nu \text{ for at least one } \nu\},\\ K_= &= \{k \in \mathbb{Z}^d : |k_\nu| = N_\nu \text{ for at least one } \nu\},\\ K_< &= \{k \in \mathbb{Z}^d : |k_\nu| \le N_\nu \text{ for all } \nu\}, \end{aligned}$$

and define also $K_\ge = K_> \sqcup K_=$, $K_\le = K_< \sqcup K_=$. For $k \in K_=$, we denote $m(k) = |\{\nu : |k_\nu| = N_\nu\}|$ and verify that $m(k) \ge 1\ \forall k \in K_=$. Using the Fourier expansion of v in (A.50), the definition of I_N^Ω (A.70), and the aliasing relation (A.76), we write

$$\begin{aligned} (v - I_N v)(\bar{x}) = &- \sum_{k\in K_>} \Big(\sum_{\ell\in\mathbb{Z}} \hat{v}(k + 2\ell N)\Big) e^{\mathrm{i}k\bar{x}} \\ &+ \sum_{k\in K_=} \big(\tfrac{1}{2}\big)^{m(k)} \Big(\hat{v}(k) - \sum_{\ell\neq 0} \hat{v}(k + 2\ell N)\Big) e^{\mathrm{i}k\bar{x}} + \sum_{k\in K_<} \hat{v}(k) e^{\mathrm{i}k\bar{x}}. \end{aligned}$$

Using the inequality $(a+b)^2 \le 2(a^2+b^2)$ and the fact that $2(1/2)^{2m(k)} \le 1/2$, we obtain

$$|v - I_N v|^2_{\mathring{\mathrm{H}}^\sigma} \le {\sum_{k\in K_\ge}}' |k|^{2\sigma} |\hat{v}(k)|^2 + {\sum_{k\in K_\le}}' |k|^{2\sigma} \Big|\sum_{\ell\neq 0} \hat{v}(k + 2\ell N)\Big|^2 =: E_1 + E_2, \tag{A.77}$$

where we used the shortened notation $(k + 2\ell N)_\nu = k_\nu + 2\ell_\nu N_\nu$ and the multiindex ℓ is summed over $\mathbb{Z}^d\backslash\{0\}$. For $k \in K_\ge$ we have $|k| \ge N_{\min}$. Hence, as $\sigma \le s$, we estimate E_1 as

$$E_1 = {\sum_{k\in K_\ge}}' \frac{1}{|k|^{2(s-\sigma)}} |k|^{2s} |\hat{v}(k)|^2 \le \frac{1}{N_{\min}^{2(s-\sigma)}} |v|^2_{\mathring{\mathrm{H}}^s(\mathbb{T}^d)} \le \frac{(d^{1/2} r(N))^{2(s-\sigma)}}{|N|^{2(s-\sigma)}} |v|^2_{\mathring{\mathrm{H}}^s(\mathbb{T}^d)}, \tag{A.78}$$

where we used the bound $|N|^2 \le d N_{\max}^2$, which implies $N_{\min}^{-2} = N_{\max}^{-2} r(N)^2 \le |N|^{-1} d r(N)^2$. In order to bound E_2, we first use Cauchy–Schwartz in ℓ^2 to get

$$\Big|\sum_{\ell\neq 0} \hat{v}(k + 2\ell N)\Big|^2 \le \Big(\sum_{\ell\neq 0} |k + 2\ell N|^{2s} \big|\hat{v}(k + 2\ell N)\big|^2\Big) \Big(\frac{1}{|N|^{2s}} \sum_{\ell\neq 0} \frac{1}{|k + 2\ell N|^{2s} |N|^{-2s}}\Big), \tag{A.79}$$

where we need to show that the second series converges. Using the reverse triangle inequality, and as $k_\nu \le 2\ell_\nu N_\nu$ for $k \in K_\le$, we have

$$|k + 2\ell N| \ge d^{-1/2} \sum_{\nu=1}^d |k_\nu + 2\ell_\nu N_\nu| \ge d^{-1/2} \sum_{\nu=1}^d \big||k_\nu| - 2|\ell_\nu| N_\nu\big| = d^{-1/2} \sum_{\nu=1}^d 2|\ell_\nu| N_\nu - |k_\nu|.$$

Note that $N_\nu |N|^{-1} \ge N_{\min} d^{-1/2} N_{\max}^{-1} = d^{-1/2} r(N)^{-1}$ and $|k_\nu|/|N| \le |k_\nu|/N_\nu \le 1 \le \gamma$ (for $k \in K_\le$) for any $\gamma \ge 1$. Consequently,

$$|k + 2\ell N||N|^{-1} \ge d^{-1/2} \sum_{\nu=1}^d \big(2|\ell_\nu| d^{-1/2} r(N)^{-1} - \gamma\big) = d^{-1} r(N)^{-1} \big(2|\ell|_1 - \gamma d^{3/2} r(N)\big),$$

where $|\ell|_1 = \sum_{\nu=1}^{d} |\ell_\nu|$ is the 1-norm of the multiindex ℓ. Fixing $1 \le \gamma \le 2$ such that $\gamma d^{3/2} r(N)/2 \notin \mathbb{N}_{>0}$, and denoting the constant $c = \gamma d^{3/2} r(N)$, we obtain the bound

$$\sum_{\ell \neq 0} \frac{1}{|k + 2\ell N|^{2s} |N|^{-2s}} \le \sum_{\ell \neq 0} \frac{1}{(2|\ell|_1 - c)^{2s}} =: T(d, c, s),$$

where all the terms of the series $T(d, c, s)$ are well defined. Recall that the partial sum and its limit are defined in the sense (A.51). Let P_λ be the open ball in ℓ^1 of radius λ, $P_\lambda = \{x \in \mathbb{R}^d : |x|_1 < \lambda\}$, and define $L_d(n)$ as the cardinality of the set $\{\ell \in \mathbb{N}^d : |\ell|_1 = n\}$. We can then rewrite the partial sum of $T(d, s)$ as

$$T_\lambda(d, c, s) = \sum_{\ell \in P_\lambda} \frac{1}{(2|\ell|_1 - c)^{2s}} \le 2^d \sum_{n=0}^{\lfloor \lambda \rfloor} L_d(n) \frac{1}{(2n - c)^{2s}},$$

and we have $T(d, c, s) = \lim_{\lambda \to \infty} T_\lambda(d, c, s)$. Let us consider the case $d = 1, 2, 3$ independently. Assume that $d = 1$. As $L_1(n) = 1$ for all n, the series converges for $s > 1/2$ (see the discussion after Theorem A.4.2) and if $s \ge 1$ the limit is bounded independently of s. In the case $d = 2$, we count $L_2(n) = n + 1$ and the series converges for $s > 1$. If $s \ge (d + 1)/2 = 3/2$, then $T(d, c, s)$ is bounded independently of s. If $d = 3$, we count $L_3(n) = (n + 1)(n + 2)/2$ and thus the limit $\lim_{\lambda \to \infty} T_\lambda(d, c, s)$ converges if $s > 3/2$. If $s \ge (d + 1)/2 = 2$, then $T(d, c, s)$ is bounded independently of s. We thus denote $T(d, c) = T(d, c, s)$. Using (A.79), we obtain the following estimate for E_2

$$E_2 \le \frac{T(d, c)}{|N|^{2(s-\sigma)}} \sum_{k \in K_\le}{}' \sum_{\ell \neq 0} |k + 2\ell N|^{2s} \big|\hat{v}(k + 2\ell N)\big|^2.$$

As in the proof of Theorem A.64, we verify that the only multiindices appearing twice in the double sum correspond to $k_\nu = \pm N_\nu$. Hence, these double terms are exactly removed by $\sum'$ and thus the double sum is bounded by $|v|^2_{\tilde{\mathrm{H}}^s}$. We thus get

$$E_2 \le \frac{T(d, c)}{|N|^{2(s-\sigma)}} |v|^2_{\tilde{\mathrm{H}}^s(\mathbb{T}^d)},$$

which, combined with (A.78), proves estimate (A.75) and the proof of the theorem is complete. □

A.4.5 The Fourier differencing method in several dimensions and its implementation

In this section, we generalize the Fourier differencing method presented in one dimension in Section A.4.3. In particular, we adapt it to functions defined on any multidimensional hypercubes. The method is based on the theory on trigonometric polynomials derived in the previous section.

Let $\Omega = (a_1, b_1) \times \cdots \times (a_d, b_d)$ be a hypercube and let F_Ω be the bijective affine mapping defined as

$$F_\Omega : \mathbb{T}^d \to \Omega, \quad \bar{x} \mapsto F_\Omega(\bar{x}) = B_\Omega \bar{x} + a,$$

where B_Ω is the diagonal matrix defined as $(B_\Omega)_{jj} = (b_j - a_j)/(2\pi)$. For a given multi-index $\alpha \in \mathbb{N}^d$, with $|\alpha|_1 \le m$, the spectral derivatives of $v \in \mathrm{H}^m_{\mathrm{per}}(\Omega)$ are the derivatives of the trigonometric interpolant of v, defined in (A.74), i.e,

$$\partial^\alpha_x v(x) \approx \partial^\alpha_x I^\Omega_N v(x) = J^\alpha \partial^\alpha_{\bar{x}} \big(I_N(v \circ F_\Omega)\big)(F_\Omega^{-1} x) \in V_N(\Omega),$$

where $J_\nu = (2\pi)/(b_\nu - a_\nu)$ is the ν-th diagonal of B_Ω^{-1}. For $v \in \mathrm{H}^{s+m}(\Omega)$, where $s \ge (d + 1)/2$, Theorem A.4.4 gives the error estimate

$$\|\partial^\alpha_x v - \partial^\alpha_x I^\Omega_N v\|_{\mathrm{L}^2(\Omega)} \le C \frac{r(N)^s}{|B_\Omega^{-1} N|^s} |v|_{\tilde{\mathrm{H}}^{s+m}(\Omega)},$$

where $r(N) = N_{\max}/N_{\min}$. Consequently, the method converges if all $N_\nu \to \infty$ simultaneously (i.e. $r(N)$ stays bounded).

Let us explain how the method is implemented. For $N \in \mathbb{N}_{>0}$ let $G_N = \{x_{n_1\cdots n_d}\}_{n_\nu=0}^{2N_\nu-1}$ be the uniform grid of Ω, where the size in each direction is $h_\nu = (b_\nu - a_\nu)/(2N_\nu)$. Define the map of evaluation on the grid $E : \mathrm{L}^2_{\mathrm{per}}(\Omega) \to \mathbb{C}^{2N_1\times\cdots\times 2N_d}$ as $v \mapsto Ev$, where $Ev_{n_1\cdots n_d} = v(x_{n_1\cdots n_d})$. We verify that $E|_{V_N(\Omega)} : V_N(\Omega) \to \mathbb{C}^{2N_1\times\cdots\times 2N_d}$ is an isomorphism. We define the (multidimensional) DFT as the map $\mathcal{F}_h : \mathbb{C}^{2N_1\times\cdots\times 2N_d} \to \mathbb{C}^{2N_1\times\cdots\times 2N_d}$, $V \mapsto \mathcal{F}_h(V)$ given by

$$\mathcal{F}_h(V)_{k_1\cdots k_d} = \frac{1}{2N_1}\sum_{n_1=0}^{2N_1-1}\cdots\frac{1}{2N_d}\sum_{n_d=0}^{2N_d-1} V_{n_1\cdots n_d} e^{-ik_1n_1h_1}\cdots e^{-ik_dn_dh_d}, \tag{A.80}$$

for $k_\nu = -N_\nu+1,\ldots,N_\nu$, and the (multidimensional) inverse DFT as the map $\mathcal{F}_h^{-1} : \mathbb{C}^{2N_1\times\cdots\times 2N_d} \to \mathbb{C}^{2N_1\times\cdots\times 2N_d}$, $\hat{V} \mapsto \mathcal{F}_h^{-1}(\hat{V})$ given by

$$\mathcal{F}_h^{-1}(\hat{V})_{n_1\cdots n_d} = \sum_{k_1=-N_1+1}^{N_1}\cdots\sum_{k_d=-N_d+1}^{N_d} \hat{V}_{k_1\cdots k_d} e^{ik_1n_1h_1}\cdots e^{ik_dn_dh_d}.$$

Note that the multidimensional DFT and iDFT can be computed using FFT and iFFT algorithms. Let us also define the map of differentiation in the Fourier space $\hat{D}_\nu^m : \mathbb{C}^{2N_1\times\cdots\times 2N_d} \to \mathbb{C}^{2N_1\times\cdots\times 2N_d}$, $\hat{V} \mapsto \hat{D}_\nu^m\hat{V}$, where

$$\begin{aligned}
&(\hat{D}_\nu^m\hat{V})_{k_1\cdots k_d} = (\mathrm{i}k_\nu)^m\hat{V}_{k_1\cdots k_d} \quad |k_\nu| \le N_\nu - 1, \quad (\hat{D}_\nu^m\hat{V})_{k_1\cdots k_d} = 0 \quad k_\nu = N_\nu, && \text{if } m \text{ is odd,}\\
&(\hat{D}_\nu^m\hat{V})_{k_1\cdots k_d} = (\mathrm{i}k_\nu)^m\hat{V}_{k_1\cdots k_d} && \text{if } m \text{ is even.}
\end{aligned}$$

Then, the spectral differentiation map D_ν^m is

$$D_\nu^m : \mathbb{C}^{2N_1\times\cdots\times 2N_d} \to \mathbb{C}^{2N_1\times\cdots\times 2N_d}, \quad V \mapsto D_\nu^m V = \Big(\frac{2\pi}{b_\nu - a_\nu}\Big)^m \mathcal{F}_h^{-1}\circ\hat{D}_\nu^m\circ\mathcal{F}_h(V). \tag{A.81}$$

For a multi-index $\alpha \in \mathbb{N}_{>0}^d$, the derivative of $v \in \mathrm{L}^2_{\mathrm{per}}(\Omega)$ is then approximated on the grid G_N as

$$\partial_x^\alpha v(x_{n_1\cdots n_d}) \approx \partial_x^\alpha I_N^\Omega v(x_{n_1\cdots n_d}) = \big(D_1^{\alpha_1}\circ\cdots\circ D_d^{\alpha_d}(Ev)\big)_{n_1\cdots n_d}.$$

In Program A.2, we present an example of implementation of the Fourier differencing method using Matlab. Note that the Matlab implementation of iFFT require to shift the vectors in the Fourier space (see `k1`, `k2` and Matlab's help for the function `fft`). The implementation takes advantage of the fact that to compute D_ν^m, the maps DFT and iDFT can be performed only in the direction ν. To approximate the mixed derivative $\partial_{12}^2 v(x_n)$, the DFT has to be computed along both directions and for the iDFT, we use the fact that `ifft2(V) = ifft(ifft(V,[],1),[],2)`. As in 1d, Matlab does not allow to take advantage of the symmetry $\hat{v}_{-k} = \overline{\hat{v}_k}$ satisfied by real valued functions v and the FFTW library [56] can be used to speed up the computations.

A.4.6 Finite dimensional space for the approximation of periodic partial differential equations

The space of trigonometric polynomials, defined in Section A.4.4, is at the center of the definition of the spectral method. The use of spectral methods is particularly judicious for the approximation of smooth PDEs with periodic boundary conditions. In this section, we define the variational settings of the spectral method. In particular, we define the finite dimensional subspace of approximation and provide the corresponding error estimates for the corresponding interpolation

Program A.2: Matlab implementation of the Fourier differencing method in 2d.

```
% domain and jacobian
Om = [-1 1 ; -2 2];
J  = 2*pi./(Om(:,2)-Om(:,1));
% function v, its derivatives
v = @(x,y) exp(sin(pi*x)).*cos(pi*y);
d1v  = matlabFunction( diff(sym(v),'x') );
d2v  = matlabFunction( diff(sym(v),'y') );
d11v = matlabFunction( diff(sym(d1v),'x') );
d12v = matlabFunction( diff(sym(d2v),'x') );
d22v = matlabFunction( diff(sym(d2v),'y') );
% discretization
N = [5 ; 10];
h = (Om(:,2)-Om(:,1))./(2*N);
x1 = (Om(1,1):h(1):Om(1,2)-h(1))';
x2 = (Om(2,1):h(2):Om(2,2)-h(2))';
[X1,X2] = meshgrid(x1,x2); X1 = X1'; X2 = X2';
Eval = @(w) reshape( w(X1(:),X2(:)) , 2*N');
% spectral differentiation
k1 = repmat( fftshift((-N(1):N(1)-1)') ,[1,2*N(2)]);
k2 = repmat( fftshift((-N(2):N(2)-1) ) ,[2*N(1),1]);
ifOdd1 = (k1~=-N(1));
ifOdd2 = (k2~=-N(2));
Ev = Eval(v);
fft1_Ev  = fft(Ev,[],1);
fft2_Ev  = fft(Ev,[],2);
fft12_Ev = fft(fft1_Ev,[],2);
D1v  = J(1)  *real(ifft( 1i*k1.*ifOdd1.*fft1_Ev ,[],1));
D2v  = J(2)  *real(ifft( 1i*k2.*ifOdd2.*fft2_Ev ,[],2));
D11v = J(1)^2*real(ifft( (1i*k1).^2.*fft1_Ev ,[],1));
D22v = J(2)^2*real(ifft( (1i*k2).^2.*fft2_Ev ,[],2));
D12v = J(1)*J(2)*real(ifft2( (1i*k1).*ifOdd1.*(1i*k2).*ifOdd2.*fft12_Ev ));
% errors
Ed1v = Eval(d1v);   Ed2v = Eval(d2v);
fprintf('|Ed1v-D1v|=%g\n', max(abs(Ed1v(:)-D1v(:))));
fprintf('|Ed2v-D2v|=%g\n', max(abs(Ed2v(:)-D2v(:))));
Ed11v = Eval(d11v); Ed12v = Eval(d12v); Ed22v = Eval(d22v);
fprintf('|Ed11v-D11v|=%g\n', max(abs(Ed11v(:)-D11v(:))));
fprintf('|Ed12v-D12v|=%g\n', max(abs(Ed12v(:)-D12v(:))));
fprintf('|Ed22v-D22v|=%g\n', max(abs(Ed22v(:)-D22v(:))));
```

theory. These settings used are used in Sections 7.2 and 2.3, where spectral method for hyperbolic PDEs are analyzed.

Let us assume that the solution of interest belongs to $\mathrm{W}_{\mathrm{per}}(\Omega)$, where Ω is a periodic hypercube of $\mathbb{R}^d$ ($d \leq 3$), $\Omega = (a_1, b_1) \times \cdots \times (a_d, b_d)$. We denote F_Ω the bijective affine mapping

$$F_\Omega : \mathbb{T}^d \to \Omega, \quad \bar{x} \mapsto F_\Omega(\bar{x}) = B_\Omega \bar{x} + a,$$

where B_Ω is the diagonal matrix defined as $(B_\Omega)_{jj} = (b_j - a_j)/(2\pi)$. We define the finite dimensional subspace of $\mathrm{W}_{\mathrm{per}}(\Omega)$

$$\mathring{V}_N(\Omega) = V_N(\Omega) \cap \mathrm{W}_{\mathrm{per}}(\Omega),$$

where $V_N(\Omega)$ is the finite dimensional space of trigonometric polynomials defined in (A.71). Note that the condition $\langle v_N \rangle_\Omega = 0$ is satisfied if and only if the coefficient of the basis function $w_{0\cdots0}(x) = 1$ is zero, i.e., $\mathring{V}_N(\Omega) = \mathrm{span}(B_N \backslash \{w_{0\cdots0}\})$ (B_N is also defined in (A.71)). We define then the interpolant onto $\mathring{V}_N(\Omega)$, $\mathring{I}_N^\Omega : \mathrm{W}_{\mathrm{per}}(\Omega) \to \mathring{V}_N(\Omega)$, as

$$\mathring{I}_N^\Omega v = I_N^\Omega v - \langle I_N^\Omega v \rangle_\Omega, \tag{A.82}$$

where I_N^Ω is the trigonometric interpolant defined in (A.74). Now, we verify that for $v \in \mathrm{W}_{\mathrm{per}}(\Omega) \cap \in \mathrm{H}^s(\Omega)$, where $s \geq (d+1)/2$, we have

$$|\langle I_N^\Omega v \rangle_\Omega| \leq C \frac{r(N)^{s-\sigma}}{\left|B_\Omega^{-1} N\right|^{s-\sigma}} |v|_{\tilde{\mathrm{H}}^s(\Omega)},$$

where $r(N) = N_{\max}/N_{\min}$ and C is a constant depending only on d. The proof is similar to the first part of the proof of Theorem A.4.4 (it corresponds to the aliasing error of v, see (A.76)). Thus, combining this estimate with Theorem A.4.4 and (A.55), we obtain the following estimate for $\mathring{I}_N^\Omega$.

Theorem A.4.5. *Let $d \leq 3$ and assume that the ratio $r(N) = N_{\max}/N_{\min}$ is bounded. Let v be a zero mean Ω-periodic function such that $v \in \mathrm{W}_{\mathrm{per}}(\Omega) \cap \mathrm{H}^s(\Omega)$, for some $s \geq (d+1)/2$. Then, for any $\sigma \leq s$, the interpolant $\mathring{I}_N^\Omega$, defined in* (A.82)*, satisfies the estimate*

$$\left|v - \mathring{I}_N^\Omega v\right|_{\mathrm{H}^\sigma(\Omega)} \leq C \frac{r(N)^{s-\sigma}}{\left|B_\Omega^{-1} N\right|^{s-\sigma}} |v|_{\mathrm{H}^s(\Omega)}, \tag{A.83}$$

where C is a constant depending only on s, d and $r(N)$.

We emphasize that the constant C in (A.83) is independent of the domain Ω. Furthermore, the presence of B_Ω^{-1} confirms the instinctive idea that the number of point in the grid must be increased if the domain grows.

A.4.7 Implementations of the spectral method and of the Fourier method

In this section, we list two codes for the approximation of the wave equation and of the Boussinesq equation with constant coefficients. First, Program A.3 presents an implementation of the spectral method for the wave equation, analyzed in Section 2.3. Second, Program A.4 presents an implementation of the Fourier method for constant coefficients PDEs, defined in Section 2.4.

Program A.3: Matlab implementation of the spectral method defined in Section 2.3.

```
% time, epsilon & initial data
T  = 5;    epsilon = 0.1;
g0 = @(x,y) exp( -(x.^2 +y.^2)/0.05 );
% constant coefficents operators
a11 = @(x,y) 1-0.5*sin(pi*x).*sin(pi*y);
a22 = @(x,y) 1-0.5*sin(pi*x).*sin(pi*y);
% domain & jacobian
Om = [-2,2; -2,2];      J  = 2*pi./(Om(:,2)-Om(:,1));
% discretization
N = ceil((Om(:,2)-Om(:,1))./(1*epsilon));   h = (Om(:,2)-Om(:,1))./(2*N);
x1 = (Om(1,1):h(1):Om(1,2)-h(1))';          x2 = (Om(2,1):h(2):Om(2,2)-h(2))';
[X1,X2] = meshgrid(x1,x2);                  X1 = X1'; X2 = X2';
Eval = @(w) reshape( w(X1(:),X2(:)) , 2*N');
% differentiation map & operator A
k1 = J(1)*repmat( fftshift((-N(1):N(1)-1)') ,[1,2*N(2)]);
k2 = J(2)*repmat( fftshift((-N(2):N(2)-1) ) ,[2*N(1),1]);
ifOdd1  = (k1~=-N(1));     ifOdd2 = (k2~=-N(2));
D1 = @(V) J(1)*real(ifft( 1i*k1.*ifOdd1.*fft(V,[],1) ,[],1));
D2 = @(V) J(2)*real(ifft( 1i*k2.*ifOdd2.*fft(V,[],2) ,[],2));
A11 = Eval(a11);    A22 = Eval(a22);
apply_A = @(V) D1(A11.*D1(V)) + D2(A22.*D2(V));
% time integration with the leap frog method
U = Eval(g0);   V = U + 0.5*dt^2*apply_A(U);
dt =min(h)/20;  Ntime =ceil(T/dt);  DTshow =0.01;  nshow =ceil(DTshow/dt);
fig =figure;    axlim =[Om(1,:),Om(2,:),[-1 1]];
for n=1:Ntime
    V = V + dt*apply_A(U)*(1-0.5*(n==1));
    U = U + dt*V;
    if mod(n,nshow)==0;
        if ~ishandle(fig);  fprintf('closed\n');  break;  end;
        mesh(x2,x1,U,'edgecolor','k'); axis(axlim); drawnow(); pause(DTshow);
    end
end
```

Program A.4: Matlab implementation of the method given in Section 2.4.

```
% time, epsilon & initial data
t  = 100;    epsilon = 0.1;
g0 = @(x,y) exp( -(x.^2 +y.^2)/0.05 );
% constant coefficents operators
opa0 = [1; 0; sqrt(3)/2];           % [a0_11 2a0_12 a0_22]
opb0 = [6.3404e-03; 0; 1.0045e-02]; % [b0_11 2b0_12 b0_22]
opa2 = [2.9468e-03; 0; 2.2074e-02; 0; 5.4910e-03];
    % [a2_1111 4a2_1112 6a2_1122 4a2_1222 a2_2222]
% domain (based on a0 and t & jacobian
Om = [ [-1,1]*sqrt(opa0(1))*t+[-1,1] ; [-1,1]*sqrt(opa0(3))*t+[-1,1] ];
J  = 2*pi./(Om(:,2)-Om(:,1));
% discretization
N = ceil((Om(:,2)-Om(:,1))./epsilon);
h = (Om(:,2)-Om(:,1))./(2*N);
x1 = (Om(1,1):h(1):Om(1,2)-h(1))';
x2 = (Om(2,1):h(2):Om(2,2)-h(2))';
[X1,X2] = meshgrid(x1,x2); X1 = X1'; X2 = X2';
Eval = @(w) reshape( w(X1(:),X2(:)) , 2*N');
% Fourier space indices k
k1 = J(1)*repmat( fftshift((-N(1):N(1)-1)') ,[1,2*N(2)]);
k2 = J(2)*repmat( fftshift((-N(2):N(2)-1) ) ,[2*N(1),1]);
a0kk = opa0(1)*k1.^2 +opa0(2)*k1.*k2 +opa0(3)*k2.^2;
b0kk = opb0(1)*k1.^2 +opb0(2)*k1.*k2 +opb0(3)*k2.^2;
a2kkTkkT = opa2(1)*k1.^4 +opa2(2)*k1.^3.*k2 + opa2(3)*k1.^2.*k2.^2 ...
          +opa2(4)*k1.*k2.^3 + opa2(5)*k2.^4;
rk = (a0kk +epsilon^2*a2kkTkkT)./(1 +epsilon^2*b0kk);
% approximation of u(t)
uN = real(ifft2( fft2(Eval(g0)).*cos(sqrt(rk)*t) ));
% display cut
cut1  = sqrt(opa0(1))*t +[-3,0.9];
icut1 = 1:2*N(1);  icut1 = icut1(cut1(1)<=x1 & x1<=cut1(2));
cut2  = [max(Om(2,1),-4), min(Om(2,2),4)];
icut2 = 1:2*N(1);  icut2 = icut2(cut2(1)<=x2 & x2<=cut2(2));
figure;  surf(x2(icut2),x1(icut1),uN(icut1,icut2));  shading interp;
  axis([cut2(1:2),cut1(1:2)]); view(2);
```

A.5 Leap frog integration in time

The leap frog method is a simple and accurate integrator for the approximation of second order ordinary differential equations. The advantages of this scheme are diverse. The main one is its symplecticity, which ensures the (modified) energy associated to the dynamical system is conserved by the approximation. In addition, the method is explicit when applied to a system of the form (A.84), and has a second order accuracy. In this section, we prove a stability condition and the second order of convergence of the method. We refer to [60, 61] for the general theory (the method is also called the Störmer-Verlet method). For more general scheme with similar properties, we refer to [57].

Let us consider the second order ODE in $\mathbb{R}^d$

$$\begin{aligned} &\ddot{u}(t) = f(u(t)) \quad \text{for a.e. } t \in (0,T], \\ &u(0) = u^0, \quad \dot{u}(0) = v^0, \end{aligned} \tag{A.84}$$

where $f : \mathbb{R}^d \to \mathbb{R}^d$ is a field and $u^0, v^0 \in \mathbb{R}^d$ are given initial conditions. Denoting $y(t) = (u(t), \dot{u}(t))^T \in \mathbb{R}^{2d}$, $F(y(t)) = (\dot{u}(t), f(u(t)))^T$ and $y^0 = (u^0, v^0)^T$, we verify that (A.84) is equivalent to the first order ODE $\mathbb{R}^{2d}$

$$\begin{aligned} &\dot{y}(t) = F(y(t)) \quad \text{for a.e. } t \in (0,T) \\ &y(0) = y^0. \end{aligned} \tag{A.85}$$

If f is Lipschitz continuous, then standard theory ensures the existence and uniqueness of a solution $u \in \mathcal{C}^1([0,T];\mathbb{R}^n)$ to (A.84) (see e.g., [38]).

Let us discretized (A.84) with the leap frog method. For $N \in \mathbb{N}_{>0}$, let $t^n = n\Delta t$ $\Delta t = T/N$ be the uniform discretization of the interval $[0,T]$. If u is sufficiently regular, Taylor expansion gives

$$\ddot{u}(t^n) = \frac{u(t^{n+1}) - 2u(t^n) + u(t^n)}{\Delta t^2} + \mathcal{O}(\Delta t^2).$$

Using this approximation in (A.84) leads to the scheme

$$u^{n+1} = 2u^n - u^{n-1} + \Delta t^2 f(u^n). \tag{A.86}$$

Let us define

$$v^{n+1/2} = \frac{u^{n+1} - u^n}{\Delta t}, \qquad v^n = \frac{u^{n+1} - u^{n-1}}{2\Delta t} = v^{n+1/2} + v^{n-1/2}.$$

Using this notation, (A.86) can be rewritten as a one step method

$$\begin{aligned} &v^{n+1/2} = v^{n-1/2} + \Delta t f(u^n), \\ &u^{n+1} = u^n + \Delta t v^{n+1/2}. \end{aligned} \tag{A.87}$$

Using then (A.87) and the definition of v^n leads to the relation $2v^{n+1/2} = 2v^n + \Delta t f(u^n)$ which is used for the initialization of the scheme (A.87), $v^{1/2} = v^0 + \frac{\Delta t}{2} f(u^0)$.

Let us derive a stability condition for the leap frog method (A.86). Consider the linear scalar equation

$$\begin{aligned} &\ddot{u}(t) = \lambda u(t) \quad \text{for a.e. } t > 0, \\ &u(0) = u^0, \ \dot{u}(0) = u^1, \end{aligned}$$

where λ is an eigenvalue of the jacobian f'. As the solution of this equation is $u(t) = C_1 e^{\sqrt{\lambda}t} + C_2 e^{\sqrt{\lambda}t}$, where C_1, C_2 depend on u^0, u^1, the stability domain is $\{\lambda \leq 0\}$. Applying the method (A.86), we obtain the recursive relation

$$u^{n+1} - (2 + \lambda\Delta t^2)u^n + u^{n-1} = 0, \tag{A.88}$$

where u^0, u^1 are given. Let us find an explicit formula for u^n. Making the ansatz that u^n has the form $u^n = \zeta^n$, with $\zeta \in \mathbb{C}$, we find that ζ must be a root of the polynomial

$$\rho(\zeta) = \zeta^2 - (2 + \lambda\Delta t^2)\zeta + 1.$$

As the two roots of ρ ζ_1, ζ_2 satisfy $\zeta_1\zeta_2 = 1$, they can be written as $\zeta_1 = \zeta, \zeta_2 = \zeta^{-1}$. By linearity, $u^n = A\zeta^n + B\zeta^{-n}$ satisfies the recursive relation (A.88). If the roots of ρ are not simple, i.e., $\zeta = 1$ or -1, we can chose u^0, u^1 such that $|u^n| = n$ and the method is unstable ($u^n = n$ if $\zeta = 1$ and $u^n = (-1)^n n$ if $\zeta = -1$). Hence, the roots have to be simple, i.e., $\zeta \neq \pm 1$. The solution of (A.88) is thus $u^n = A\zeta^n + B\zeta^{-n}$, where A, B are such that $A + B = u^0$ and $A\zeta + B\zeta^{-1} = u^1$. The method is thus stable if and only if $|u^n| = \left|A\zeta^n + B\zeta^{-n}\right|$, is bounded for any $A, B \in \mathbb{R}$. This is equivalent to the condition $|\zeta| = 1$, i.e., ζ must lie on the unit circle, i.e., $\zeta = e^{\mathrm{i}\theta}$. In that case,

$$\lambda\Delta t^2 = \zeta^{-1}\rho(\zeta) - 2 = e^{\mathrm{i}\theta} + e^{\mathrm{i}\theta} - 2 = 2(\cos\theta - 1),$$

and the method is stable if and only if $-4 < \lambda\Delta t^2 < 0$, i.e., $0 < \Delta t < 2/\sqrt{-\lambda}$.

Let us now prove that under sufficient regularity of f, the approximation $y^n = (u^n, v^n)^T$ of (A.85) has order 2. First, we prove that it has local order 2, i.e., there exists a constant C depending on $f, |u^0|$ and $|v^0|$ such that

$$|y(t^1) - y^1| \leq C\Delta t^3. \tag{A.89}$$

To prove (A.89), we use the relation $2v^{n+1/2} = 2v^n + \Delta t f(u^n)$ to rewrite (A.87) as $y^1 = y^0 + \Delta t G(y^0)$, where

$$G(y^0) = \begin{pmatrix} v^0 + \frac{1}{2}\Delta t f(u^0) \\ \frac{1}{2}\big(f(u^0) + f(u^1)\big) \end{pmatrix},$$

so that

$$y(t^1) - y^1 = \int_0^{t^1} F(y(t))\,\mathrm{d}t - \Delta t G(y^0).$$

Using Taylor expansion and (A.84), we compute for $t \in (0, t_1)$

$$\begin{aligned}
&G_2(y^0) = f(u^0) + \tfrac{1}{2}\Delta t f'(u^0)v^0 + \mathcal{O}(\Delta t^2), \\
&\int_0^{t^1} F_1(y(t))\,\mathrm{d}t = \int_0^{t^1} v(t)\,\mathrm{d}t = \Delta t v^0 + \tfrac{1}{2}\Delta t^2 f(u^0) + \mathcal{O}(\Delta t^3), \\
&\int_0^{t^1} F_2(y(t))\,\mathrm{d}t = \int_0^{t^1} f(u(t))\,\mathrm{d}t = \Delta t f(u^0) + \tfrac{1}{2}\Delta t^2 f'(u^0)v^0 + \mathcal{O}(\Delta t^3),
\end{aligned}$$

where f' denote the Jacobian matrix of f. That proves (A.89). Let us now prove that the local order (A.89) ensures a global order 2 for the method. For each $n = 1, \ldots, N$, we define

$$\hat{y}(t^n) = y^{n-1} + \int_{t^{n-1}}^{t^n} F(y(t))\,\mathrm{d}t,$$

where y^{n-1} is the approximation at the step $n-1$. Observe then that if the method is stable, i.e., $|u^n|, |v^n|$ are bounded, then the estimate (A.89) applies to $\hat{y}(t^n) - y^n$. Using (A.89), we thus verify that the global error satisfies

$$\begin{aligned}
\left|y(T) - y^N\right| &= \left|y_0 + \sum_{n=1}^{N}\int_{t^{n-1}}^{t^n} F(y(t))\,\mathrm{d}t - \left(y^0 + \sum_{n=1}^{N} y^n - y^{n-1}\right)\right| = \sum_{n=1}^{N}\left|\hat{y}(t^n) - y^n\right| \\
&\leq CN\Delta t\Delta t^2 = CT\Delta t^2,
\end{aligned}$$

and the method has order 2.

Bibliography

[1] A. Abdulle, *On a priori error analysis of fully discrete heterogeneous multiscale FEM*, Multiscale Model. Simul., 4 (2005), pp. 447–459.

[2] ——, *The finite element heterogeneous multiscale method: a computational strategy for multiscale PDEs*, in Multiple scales problems in biomathematics, mechanics, physics and numerics, vol. 31 of GAKUTO Internat. Ser. Math. Sci. Appl., Gakkōtosho, Tokyo, 2009, pp. 133–181.

[3] ——, *Discontinuous Galerkin finite element heterogeneous multiscale method for elliptic problems with multiple scales*, Math. Comp., 81 (2012), pp. 687–713.

[4] ——, *The role of numerical integration in numerical homogenization*, ESAIM: Proceedings, 50 (2015), pp. 1–20.

[5] A. Abdulle and Y. Bai, *Reduced basis finite element heterogeneous multiscale method for high-order discretizations of elliptic homogenization problems*, J. Comput. Phys., 231 (2012), pp. 7014–7036.

[6] A. Abdulle, Y. Bai, and T. Pouchon, *Reduced basis numerical homogenization method for the multiscale wave equation*, in Numerical Mathematics and Advanced Applications-ENUMATH 2013, Springer, 2015, pp. 397–405.

[7] A. Abdulle, W. E, B. Engquist, and E. Vanden-Eijnden, *The heterogeneous multiscale method*, Acta Numer., 21 (2012), pp. 1–87.

[8] A. Abdulle and M. J. Grote, *Finite element heterogeneous multiscale method for the wave equation*, Multiscale Model. Simul., 9 (2011), pp. 766–792.

[9] A. Abdulle, M. J. Grote, and C. Stohrer, *FE heterogeneous multiscale method for long-time wave propagation*, C. R. Math. Acad. Sci. Paris, 351 (2013), pp. 495–499.

[10] ——, *Finite element heterogeneous multiscale method for the wave equation: long-time effects*, Multiscale Model. Simul., 12 (2014), pp. 1230–1257.

[11] A. Abdulle and P. Henning, *Chapter 20 - Multiscale methods for wave problems in heterogeneous media*, in Handbook of Numerical Methods for Hyperbolic Problems Applied and Modern Issues, R. Abgrall and C.-W. Shu, eds., vol. 18 of Handbook of Numerical Analysis, North-Holland, Amsterdam, 2017, pp. 545–576.

[12] ——, *Localized orthogonal decomposition method for the wave equation with a continuum of scales*, Math. Comp., 86 (2017), pp. 549–587.

[13] A. Abdulle and T. Pouchon, *A priori error analysis of the finite element heterogeneous multiscale method for the wave equation over long time*, SIAM J. Numer. Anal., 54 (2016), pp. 1507–1534.

[14] ——, *Effective models for the multidimensional wave equation in heterogeneous media over long time and numerical homogenization*, Math. Models Methods Appl. Sci., 26 (2016), pp. 2651–2684.

[15] A. Abdulle and G. Vilmart, *The effect of numerical integration in the finite element method for nonmonotone nonlinear elliptic problems with application to numerical homogenization methods*, C. R. Acad. Sci. Paris, Ser. I, 349 (2011), pp. 1041–1046.

[16] G. Allaire, *Homogenization and two-scale convergence*, SIAM J. Math. Anal., 23 (1992), pp. 1482–1518.

[17] G. Allaire, *Shape Optimization by the Homogenization Method*, Applied Mathematical Sciences, 146, Springer-Verlag New York, 2002.

[18] G. Allaire, M. Briane, and M. Vanninathan, *A comparison between two-scale asymptotic expansions and Bloch wave expansions for the homogenization of periodic structures*, SeMA J., 73 (2016), pp. 237–259.

[19] D. Arjmand and O. Runborg, *Analysis of heterogeneous multiscale methods for long time wave propagation problems*, Multiscale Model. Simul., 12 (2014), pp. 1135–1166.

[20] ——, *Estimates for the upscaling error in heterogeneous multiscale methods for wave propagation problems in locally-periodic media.* ArXiv e-print 1605.02386, 2016.

[21] G. A. Baker, *Error estimates for finite element methods for second order hyperbolic equations*, SIAM J. Numer. Anal., 13 (1976), pp. 564–576.

[22] G. A. Baker and V. Dougalis, *The effect of quadrature errors on finite element approximations for second order hyperbolic equations*, SIAM J. Numer. Anal., 13 (1976).

[23] A. Benoit and A. Gloria, *Long-time homogenization and asymptotic ballistic transport of classical waves*, arXiv preprint arXiv:1701.08600, (2017).

[24] A. Bensoussan, J.-L. Lions, and G. Papanicolaou, *Asymptotic analysis for periodic structures*, North-Holland Publishing Co., Amsterdam, 1978.

[25] C. Bernardi and Y. Maday, *Approximations spectrales de problèmes aux limites elliptiques*, vol. 10 of Mathématiques & Applications (Berlin) [Mathematics & Applications], Springer-Verlag, Paris, 1992.

[26] L. Bers, F. John, and M. Schechter, *Partial differential equations*, vol. Proceedings of the summer seminar of Lectures in Applied Mathematics, Proceeding of the Summer Seminar, Boulder, CO, 1957.

[27] S. Brahim-Otsmane, G. A. Francfort, and F. Murat, *Correctors for the homogenization of the wave and heat equations*, J. Math. Pures Appl., 71 (1992), pp. 197–231.

[28] J. T. Bushberg and J. M. Boone, *The essential physics of medical imaging*, Lippincott Williams & Wilkins, 2011.

[29] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral methods in fluid dynamics*, Springer Series in Computational Physics, Springer-Verlag, New York, 1988.

[30] Y. Capdeville, L. Guillot, and J.-J. Marigo, *1-d non-periodic homogenization for the seismic wave equation*, Geophys. J. Int., 181 (2010), pp. 897–910.

[31] L. Carleson, *On convergence and growth of partial sums of Fourier series*, Acta Mathematica, 116 (1966), pp. 135–157.

[32] L. Chen, *ifem: an integrated finite element methods package in matlab*, University of California at Irvine, (2009).

[33] P. G. Ciarlet, *The finite element method for elliptic problems*, vol. 4 of Studies in Mathematics and its Applications, North-Holland, 1978.

[34] P. G. Ciarlet and P. A. Raviart, *The combined effect of curved boundaries and numerical integration in isoparametric finite element methods*, in The mathematical foundations of the finite element method with applications to partial differential equations, 1972, pp. 409–474.

[35] D. Cioranescu, A. Damlamian, and G. Griso, *Periodic unfolding and homogenization*, C. R. Math. Acad. Sci. Paris, 335 (2002), pp. 99–104.

[36] D. Cioranescu, A. Damlamian, and G. Griso, *The periodic unfolding method in homogenization*, SIAM J. Math. Anal., 40 (2008), pp. 1585–1620.

[37] D. Cioranescu and P. Donato, *An introduction to homogenization*, vol. 17 of Oxford Lecture Series in Mathematics and its Applications, Oxford University Press, New York, 1999.

[38] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations*, International series in pure and applied mathematics, McGraw-Hill, 1955.

[39] C. Conca, R. Orive, and M. Vanninathan, *On burnett coefficients in periodic media*, J. Math. Phys., 47 (2006).

[40] C. Conca and M. Vanninathan, *Homogenization of periodic structures via bloch decomposition*, SIAM J. Appl. Math., 57 (1997), pp. 1639–1659.

[41] E. De Giorgi and S. Spagnolo, *Sulla convergenza degli integrali dell'energia per operatori ellittici del secondo ordine*, Boll. Un. Mat. Ital., 4 (1973), pp. 391–411.

[42] T. Dohnal, A. Lamacz, and B. Schweizer, *Bloch-wave homogenization on large time scales and dispersive effective wave equations*, Multiscale Model. Simul., 12 (2014), pp. 488–513.

[43] ——, *Dispersive homogenized models and coefficient formulas for waves in general periodic media*, Asymptot. Anal., 93 (2015), pp. 21–49.

[44] T. Dupont, *L^2-estimates for Galerkin methods for second order hyperbolic equations*, SIAM J. Numer. Anal., 10 (1973), pp. 880–889.

[45] B. Engquist, H. Holst, and O. Runborg, *Multi-scale methods for wave propagation in heterogeneous media*, Commun. Math. Sci., 9 (2011).

[46] ——, *Multiscale methods for wave propagation in heterogeneous media over long time*, in Numerical analysis of multiscale computations, Springer, 2012, pp. 167–186.

[47] A. Ern and J. Guermond, *Theory and practice of finite elements*, vol. 159 of Applied Mathematical Sciences, Springer-Verlag, New york, 2004.

[48] L. C. Evans, *Partial Differential Equations*, Graduate studies in mathematics, American Mathematical Society, 1998.

[49] C. Fefferman, *On the convergence of multiple Fourier series*, Bulletin of the American Mathematical Society, 77 (1971), pp. 744–745.

[50] ——, *On the divergence of multiple Fourier series*, Bulletin of the American Mathematical Society, 77 (1971), pp. 191–195.

[51] J. Fish, W. Chen, and G. Nagai, *Non-local dispersive model for wave propagation in heterogeneous media: multi-dimensional case*, Internat. J. Numer. Methods Engrg., 54 (2002), pp. 347–363.

[52] ——, *Non-local dispersive model for wave propagation in heterogeneous media: one-dimensional case*, Internat. J. Numer. Methods Engrg., 54 (2002), pp. 331–346.

[53] S. Fliss and P. Joly, *Exact boundary conditions for time-harmonic wave propagation in locally perturbed periodic media*, Appl. Numer. Math., 59 (2009), pp. 2155–2178.

[54] ——, *Wave propagation in locally perturbed periodic media (case with absorption): numerical aspects*, J. Comput. Phys., 231 (2012), pp. 1244–1271.

[55] G. B. Folland, *Fourier analysis and its applications*, vol. 4, American Mathematical Society, 1992.

[56] M. Frigo and S. G. Johnson, *The design and implementation of FFTW3*, Proceedings of the IEEE, (216–231). Special issue on "Program Generation, Optimization, and Platform Adaptation".

[57] J.-C. Gilbert and P. Joly, *Higher order time stepping for second order hyperbolic problems and optimal cfl conditions*, Partial Differential Equations : Modeling and Numerical Simulation, 16 (2008), pp. 67–93.

[58] D. Gottlieb, M. Y. Hussaini, and S. A. Orszag, *Theory and applications of spectral methods*, in Spectral methods for partial differential equations (Hampton, Va., 1982), SIAM, Philadelphia, PA, 1984, pp. 1–54.

[59] D. Gottlieb and S. A. Orszag, *Numerical analysis of spectral methods: theory and applications*, Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1977. CBMS-NSF Regional Conference Series in Applied Mathematics, No. 26.

[60] E. Hairer, C. Lubich, and G. Wanner, *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*, Springer-Verlag, Berlin and New York, 2002.

[61] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations I. Nonstiff Problems*, vol. 8, Springer Verlag Series in Comput. Math., Berlin, 1993.

[62] P. Henrici, *Fast Fourier methods in computational complex analysis*, SIAM Rev., 21 (1979), pp. 481–527.

[63] J. S. Hesthaven, S. Gottlieb, and D. Gottlieb, *Spectral methods for time-dependent problems*, vol. 21 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2007.

[64] L. Jiang and Y. Efendiev, *A priori estimates for two multiscale finite element methods using multiple global fields to wave equations*, Numer. Methods Partial Differential Equations, 28 (2012), pp. 1869–1892.

[65] L. Jiang, Y. Efendiev, and V. Ginting, *Analysis of global multiscale finite element methods for wave equations with continuum spatial scales*, Appl. Numer. Math., 60 (2010), pp. 862–876.

[66] V. V. Jikov, S. M. Kozlov, and O. A. Oleinik, *Homogenization of differential operators and integral functionals*, Springer-Verlag, Berlin, Heidelberg, 1994.

[67] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*, SIAM, 2001.

[68] H.-O. Kreiss and J. Oliger, *Comparison of accurate methods for the integration of hyperbolic equations*, Tellus, 24 (1972), pp. 199–215.

[69] ——, *Stability of the Fourier method*, SIAM J. Numer. Anal., 16 (1979), pp. 421–433.

[70] J. M. L. Guillot, Y. Capdeville, *2-d non periodic homogenization for the sh wave equation*, Geophys. J. Int., 182 (2010), pp. 1438–1454.

[71] O. A. Ladyzhenskaya, *The boundary value problems of mathematical physics*, vol. 49 of Applied Mathematical Sciences, Springer-Verlag, New York, 1985.

[72] A. Lamacz, *Dispersive effective models for waves in heterogeneous media*, Math. Models Methods Appl. Sci., 21 (2011), pp. 1871–1899.

[73] T. Lay and T. C. Wallace, *Modern global seismology*, vol. 58, Academic press, 1995.

[74] J.-L. Lions and E. Magenes, *Problèmes aux limites non homogènes et applications*, vol. 1 of Travaux et recherches mathématiques, Dunod, Paris, 1968.

[75] A. Målqvist and D. Peterseim, *Localization of elliptic multiscale problems*, Math. Comp., 83 (2014), pp. 2583–2603.

[76] F. Murat and L. Tartar, *H-convergence*, in Topics in the mathematical modelling of composite materials, vol. 31 of Progr. Nonlinear Differential Equations Appl., Birkhäuser Boston, Boston, MA, 1997, pp. 21–43.

[77] A. H. Nayfeh, *Wave propagation in layered anisotropic media: With application to composites*, vol. 39, Elsevier, 1995.

[78] O. A. Oleinik, A. Shamaev, and G. Yosifian, *Mathematical Problems in Elasticity and homogenization*, North-Holland, Amsterdam, 1992.

[79] H. Owhadi and L. Zhang, *Numerical homogenization of the acoustic wave equations with a continuum of scales*, Comput. Methods Appl. Mech. Engrg., 198 (2008), pp. 397–406.

[80] ——, *Localized bases for finite-dimensional homogenization approximations with nonseparated scales and high contrast*, Multiscale Model. Simul., 9 (2011), pp. 1373–1398.

[81] N. Panasenko and N. Bakhvalov, *Homogenization: Averaging Processes in Periodic Media: Mathematical Problems in the Mechanics of Composite Materials*, Kluwer Academic, 1989.

[82] P. A. Raviart, *The use of numerical integration in finite element methods for solving parabolic equations*, in Topics in numerical analysis. Proceedings of the Royal Irish Academy, Conference on Numerical Analysis, 1972, J. J. H. Miller, ed., Academic Press, 1973, pp. 233–264.

[83] G. Rozza, D. B. P. Huynh, and A. T. Patera, *Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations*, Arch. Comput. Methods. Eng., 15 (2008), pp. 229–275.

[84] E. Sánchez-Palencia, *Non-homogeneous media and vibration theory*, vol. 127 of Lecture Notes in Phys., Springer, 1980.

[85] F. Santosa and W. Symes, *A dispersive effective medium for wave propagation in periodic composites*, SIAM J. Appl. Math., 51 (1991), pp. 984–1005.

[86] H. Sato, M. C. Fehler, and T. Maeda, *Seismic wave propagation and scattering in the heterogeneous earth*, vol. 484, Springer, 2012.

[87] S. Spagnolo, *Sulla convergenza di soluzioni di equazioni paraboliche ed ellittiche*, Ann. Sc. Norm. Super. Pisa Cl. Sci., 22 (1968), pp. 571–597.

[88] E. Stein and R. Shakarchi, *Fourier Analysis: An Introduction*, Princeton lectures in analysis, Princeton University Press, 2011.

[89] E. Tadmor, *The exponential accuracy of Fourier and Chebyshev differencing methods*, SIAM Journal on Numerical Analysis, 23 (1986), pp. 1–10.

[90] L. Tartar, *The general theory of homogenization*, vol. 7 of Lecture Notes of the Unione Matematica Italiana, Springer-Verlag, Berlin; UMI, Bologna, 2009. A personalized introduction.

[91] L. N. Trefethen, *Spectral methods in MATLAB*, vol. 10, SIAM, 2000.

[92] C. H. Wilcox, *Theory of Bloch waves*, J. Analyse Math., 33 (1978), pp. 146–167.

[93] Ö. Yilmaz, *Seismic data analysis: Processing, inversion, and interpretation of seismic data*, Society of exploration geophysicists, 2001.

[94] K. Yosida, *Functional analysis*, Classics in Mathematics, Springer-Verlag, Berlin, 1995. Reprint of the sixth (1980) edition.

Curriculum Vitae

Personal data

Name	Timothée Pouchon
Date of birth	December 14, 1987
Nationality	Swiss

Education

2013 – 2017	**PhD in Mathematics** École Polytechnique Fédérale de Lausanne, Switzerland. Thesis advisor: Prof. A. Abdulle.
2011 – 2012	**Master of Science in Mathematics** École Polytechnique Fédérale de Lausanne, Switzerland. Thesis advisor: Prof. M. Picasso, Prof. E. Burman.
2007 – 2010	**Bachelor of Science in Mathematics** École Polytechnique Fédérale de Lausanne, Switzerland.

PhD publications

[1] A. Abdulle, Y. Bai, and T. Pouchon, *Reduced basis numerical homogenization method for the multiscale wave equation*, in Numerical Mathematics and Advanced Applications-ENUMATH 2013, Springer, 2015, pp. 397–405.

[2] A. Abdulle and T. Pouchon, *A priori error analysis of the finite element heterogeneous multiscale method for the wave equation over long time*, SIAM J. Numer. Anal., 54 (2016), pp. 1507–1534.

[3] ——, *Effective models for the multidimensional wave equation in heterogeneous media over long time and numerical homogenization*, Math. Models Methods Appl. Sci., 26 (2016), pp. 2651–2684.

In preparation

- *Effective models for wave propagation in periodic media for arbitrary timescales.*
- *Effective models for long time wave propagation in locally periodic media.*
- *Analysis of a spectral homogenization method for long time wave propagation in locally periodic media.*

Other publications

[4] S. Coutu, T. Pouchon, P. Queloz, and N. Vernaz, *Integrated stochastic modeling of pharmaceuticals in sewage networks, Stoch. Environ. Res. Risk Assess.*, 30 (2016) : pp. 1087–1097.

Presentations

- MATHICSE Retreat (Leysin, Switzerland, 13–15 July 2015);
  Talk: *Effective models for the wave equation in heterogeneous media over long time.*
- Swiss Numerical Analysis Day (Fribourg, Switzerland, 22 April 2016);
  Talk: *Multiscale method for the wave equation over long time.*
- 7th European Congress of Mathematics (Berlin, Germany, 18–22 July 2016);
  Talk: *Effective models and multiscale method for long time wave propagation in heterogeneous media.*